Speech recognition using a wavelet transform to establish fuzzy inference system through subtractive clustering and neural network (ANFIS)

First A. M. ELWAKDY, Second B. E. ELSEHELY, Third C. M. ELTOKHY, and Fourth D. A. ELHENNAWY

Abstract— In this paper, a proposed speech recognition algorithm is presented. This paper deals with the combination of a feature extraction by wavelet transform, subtractive clustering and adaptive neuro-fuzzy inference system (ANFIS). The feature extraction is used as input of the subtractive clustering to put the data in a group of clusters. Also it is used as an input of the neural network in ANFIS. The initial fuzzy inference system is trained by the neural network to obtain the least possible error between the desired output (target) and the fuzzy inference system (FIS) output to get the final FIS. The performance of the proposed speech recognition algorithm (SRA) using a wavelet transform and ANFIS is evaluated by different samples of speech signals-isolated words- with added background noise. The proposed speech recognition algorithm is tested using different isolated words obtaining a recognition ratio about 99%.

Key-words—Adaptive neuro-fuzzy inference system, noise cancellation, subtractive clustering.

I. Introduction

Speech /voice recognition is a very difficult task to be performed by a computer system [1]. Many speech /voice processing tasks, like speech and word recognition, reached satisfactory performance levels on specific applications, and although a variety of commercial products were launched in the last decade, many problems remain an open research area, and absolute solutions have not been found out yet [2]. Wavelets which bring a new tool to the speech signal classification [1]. It can be said that when using wavelets [3]-[6], the event is connected to the time when it occurs. Wavelet and wavelet packet analysis have been proven effective signal processing techniques for a variety of signal processing problems [2].
There are some basic methods of fuzzy clustering. One of them is called a subtractive clustering. The basic characteristic of this method is illustrated by experiment.

First A. M.ELWAKDY is with the Department of Electronics Technology, Faculty of Industrial Education, Helwan University, Ameria, EGYPT (Email: mohamedelwakdy2004@yahoo.com)
Second B. E.ELSEHELY is with the Avionics Research and Development Center, Salah Salim, EGYPT. (Email: elsehely@yahoo.com)
Third C. M.ELTOKHY is with the Department of Electronics Technology, Faculty of Industrial Education, Helwan University, America, , EGYPT (Email: mostafaeltokhy@hotmail.com)
Fourth D. A.ELHENNAWY, 3 Faculty of Engineering, Ain Shams University, Abasia, EGYPT (Email: a.hennawy@marsem.com)

Artificial neural network performance is depending on the size and quality of training samples [6]. When the number of training data is small, not representative of the possibility space, standard neural network results are poor [7], [8]. Incorporation of neuro-fuzzy techniques can improve performance in this case [7], [8]. Fuzzy theory has been used successfully in many applications [9]. This applications show that fuzzy theory can be used to improve neural network performance. Artificial neural networks (ANN) are known to be excellent classifiers, but their performance can be prevented by the size and quality of the training set [1].By incorporation some fuzzy techniques and neural networks, a more efficient network results, one which is extremely effective for a class of problems [1].
A combination of wavelet packet signal processing and ANFIS to efficiently extract the features from pre-processed real speech signals for the purpose of speech recognition among variety English language words [1].
In this study, a proposed algorithm of speech recognition using a wavelet transform to establish fuzzy inference system through subtractive clustering and neural network (ANFIS) is displayed as an efficient algorithm in speech recognition area
In this paper, English language speech signals-isolated words- were used. The speech recognition experiment set used in this study, speech signals are transmitted to the computer by using a microphone and an audio card which has 22 KHZ sampling frequencies .
The paper is organized as follows: Section 2 stated the following issues: reading signals and noise cancellation, preparation of data using wavelet transform (wavelet packet decomposition), subtractive clustering and ANFIS. The experimental applications including database of this study, data acquisition, ANFIS architecture and training parameters are described in section 3. The effectiveness of the proposed method for classification of speech signals in the proposed SRA through ANFIS is demonstrated in section 4. Finally, in section 5 discussion and conclusion are presented.

2 Proposed Algorithm

The proposed algorithm of speech recognition consists of four parts that are noise cancellation, wavelet transform, subtractive clustering and ANFIS. The proposed algorithm is used to carry the goal of this research.

2.1 English word discrimination using a wavelet transform, subtractive clustering and ANFIS.

The proposed algorithm is illustrated in Fig.1.

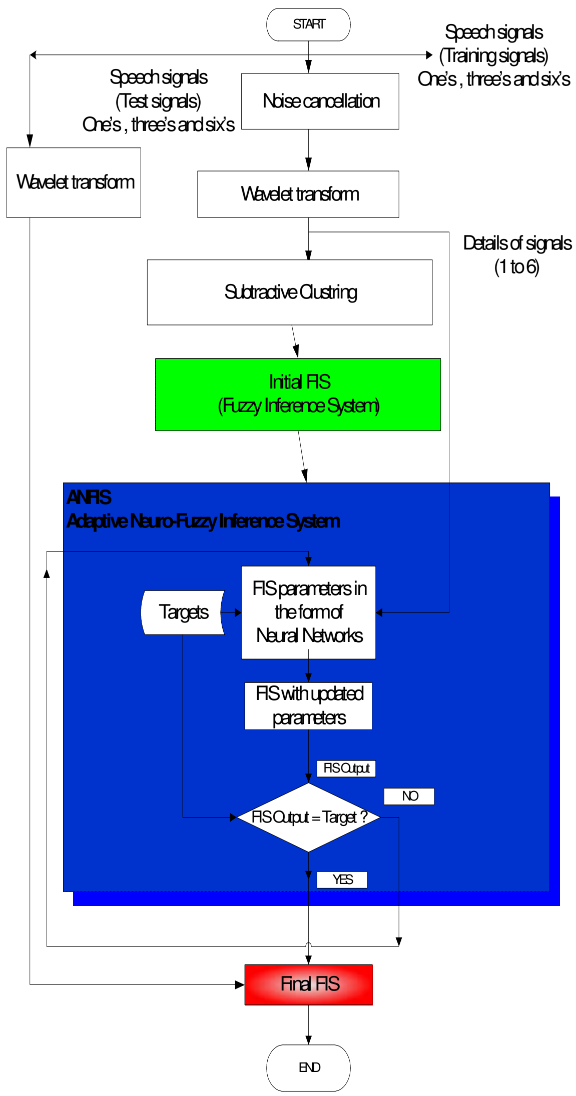


Fig. 1 the proposed algorithms of the speech recognition using a wavelet transform, subtractive clustering and ANFIS.

The following subsections are clarified in the above system's block diagram.

2.1.1. Reading signals and noise cancellation

The training signals and checking signals are read before multiplying them by 10. After that, background noise signal is added as illustrated in Eq. (1):

$$n(t) = \mathrm{randn}(\mathrm{size}(x(t)))*0.01 \quad (1)$$

Where n (t) is background noise signal and x (t) is training and checking signals.

The added background noise signal on training and checking (test) signals, and microphone noise are cancelled by noise cancellation as in Fig.2. The total number of input samples- one, three and six- is 28000 samples of training and checking signals. If the number of samples is less than 28000 samples, the difference in the number of samples is zeros.

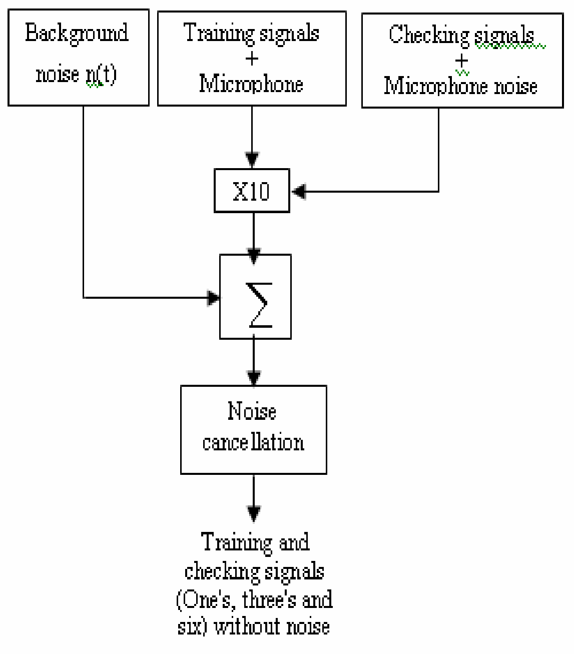


Fig .2 reading signals and noise cancellation.

2.1.2 Preparation of data

The preparation of data of the training and checking signals is done using wavelet packet decomposition. The wavelet packet transform rapidly surfaces in fields as diverse as speech recognition, telecommunications and radar target recognition [1].The main advantages of wavelets are that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time–frequency resolution in all frequency ranges. Wavelet decomposition uses the fact that it is possible to resolve high frequency components within a small time window, and only low frequency components need large time windows [1].This is because a low frequency component completes a cycle in a large time interval whereas a high frequency component completes a cycle in a much shorter interval [1].The wavelet decomposition functions at level m and time location t_m can be expressed as Eq. (2):

$$d_m(t_m) = x(t)\psi_m\left(\frac{t - t_m}{2^m}\right) \quad (2)$$

Where t_m is the decomposition filter at frequency level m .The effect of the decomposition filter is scaled by the factor 2m at stage m, but otherwise the shape is the same at all stages. The synthesis of the signal from its time–frequency coefficients given in Eq. (3) can be rewritten to express the composition of the signal x[n] from its wavelet coefficients.

$$d[n] = x[n]h[n] \quad , \quad c[n] = x[n]g[n] \qquad (3)$$

Where h[n] is the impulse response of the high pass filter and g[n] is the impulse response of the low pass filter [9].
One can choose between smooth wavelets, compactly supported wavelets, wavelets with simple mathematical expressions, wavelets with simple associated filters, etc. the most simple is the Haar [10].The Daubeachies-8, the most nearly symmetric wavelet [11], was used as a mother wavelet based on Horgan, who suggested that Daubechies-8 gave better results in removing noise from rapidly varying signal than Haar or Daubechies-4 wavelets [12].
Through a wavelet packet transform, the speech features (details) are extracted from training and checking data and then each detail represents certain frequency band after that, the standard deviation is taken of each detail as in Fig.3.

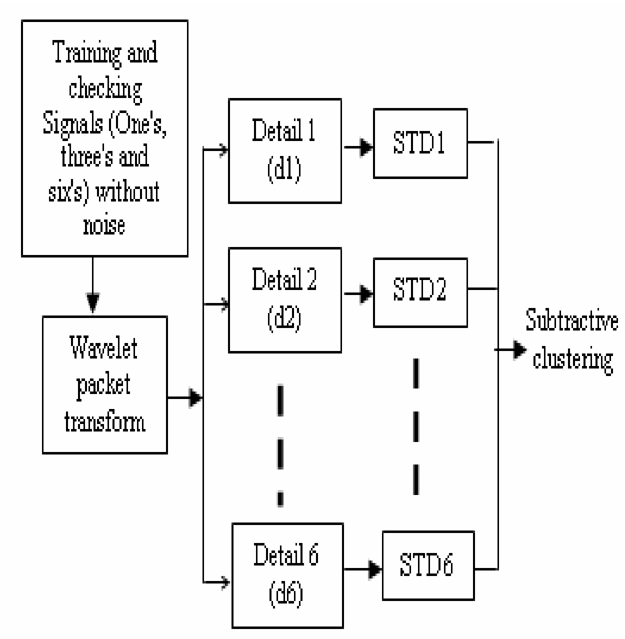


Fig.3 analysis framework

2.1.3 Subtractive Clustering

In order to obtain a set of Nr rules and avoiding the problems inherent in grid partitioning based clustering techniques, (rule base explosion), subtractive clustering is applied [13].this technique is employed since it allowed a scatter input-output space partitioning.

The Subtractive clustering is one-pass algorithm for estimating the number of clusters and the cluster centers through the training data. The subtractive clustering method partitioned the training data into groups called clusters [14].

The subtractive clustering is, essentially, a modified from of the Mountain Method [13]. The procedure for grouping 30 data point clusters {Z1, Z2, Z3,…. Zn=30} in the training set is described below [15].

1. Compute the initial potential value for each data point (Zi) as in Eq (4) [15].

$$P_i = \sum_{j-1}^{n} e^{-\alpha\left(z_i - z_j\right)} \qquad (4)$$

$\alpha = 4/r_a^2$ where r_a is a positive constant representing a normalized neighborhood data radius for each cluster.
Any point falls outside this encircling region will have little influence to the potential point. The point with the highest potential value is selected as the first cluster center. This tentatively defines the first cluster center [15].

2. A point is qualified as the first center if its potential value (P (1)) is equal to the maximum of initial potential value (P (1)*) as in Eq. (5) [15].

$$P^{(1)*} = \max_i \left(P^{(1)}(z_i)\right) \qquad (5)$$

3. Define a threshold δ as the decision to continue or stop the cluster center search. This process will continue if the current maximum potential remains greater than δ
δ = (reject ratio) × (potential value of the first cluster center)
Where η is the rejection ratio (Will be explained) later and P (1)* is the potential value of the first cluster center [15].

4. Remove the previous cluster center from further consideration [15].

5. Revise the potential value of the remaining points according to the Eq. (6) [15].

$$P_i = P_i - P_k^* e^{-\beta\left(z_i - z_k^*\right)^2} \qquad (6)$$

Where z_k^* is the point of the kth cluster center, P_k^* is its potential value, and $\beta = 4/r_b^2$ and $r_b > 0$ represent the radius of the neighborhood for which significant potential reduction will occur [16].

This procedure is repeated to generate the cluster centers until the maximum potential value in the current iteration is equal to or less than the threshold δ. After applying subtractive clustering, we get different cluster center numbers from 30 training patterns depending on rejection ratio [15].

By the end of clustering, a set of fuzzy rules will be obtained. The FIS is generated with minimum number of rules [14] .the clustering is carried out in a multidimensional space; the related fuzzy sets must be obtained. As each axis refers to a variable, the centers of the member ship functions are obtained by projecting the center of each cluster in the corresponding axis. The widths are obtained on the basis of the radius [16].
Following subsections clarify the parameters for clustering.

2.1.3.1. Range Of Influence

It indicates the radius of a cluster when the data space is considered as a unit hypercube [18].A small cluster radius will usually yield many small clusters in the data, resulting in many rules and vice versa [18]. The value of 0.15 was used for each cluster here.

2.1.3.2. Squash Factor

This is the factor used to multiply the radii values that determine the neighborhood of a cluster center, so as to squash the potential for outlying points to be considered as part of that cluster [18]. The squash factor of 1.0 was used in here.

2.1.3.3. Accept Ratio

This ratio sets the potential, as a fraction of the potential of the first cluster center, above which another data point will be accepted as a cluster center. High values are used to accept data points that have a very strong potential for being cluster centers [18].An accept ratio of 0.8 was used here.

2.1.3.4. Reject Ratio

Rejection ratio is the condition to reject a data point to be a cluster center, which is obtained from fractions of the potential first cluster center, below which a data point will be rejected as a cluster center. A reject ratio of 0.6 was used here [15].

The criteria for cluster center consideration are based on acceptance and rejection ratios [15].
Based on the range of inference, squash factor, accept ratio and reject ratio is the construction of initial FIS as in Fig.1.

2.1.4. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fuzzy modeling [17] is a new branch of system identification which deals with the construction of a fuzzy inference system 'or fuzzy model' that can predict and explain the behavior of an unknown system described by a set of sample data. The FIS structure contains all the FIS information [14].All the information for a given FIS is contained in the FIS structure; including variable names, membership function definitions [1].

A fuzzy inference system employing fuzzy 'if-then rules' can model the qualitative aspects of human knowledge and reasoning processes without employing precise quantitative analyses [18].

The ANFIS is limited to using Sugeno systems. Both artificial neural network and fuzzy logic are used in ANFIS's architecture. An adaptive neural network is a network structure consisting of a number of nodes connected through directional links. Each node is characterized by a node function with fixed or adjustable parameters. Learning or training phase of a neural network is a process to determine parameter values to sufficiently fit the training data [18].

Considering a first-order Takagi, Sugeno and Kang (TSK) fuzzy inference system, a fuzzy model contains two rules [19]:

Rule 1: If x is A1 and y is B1; then f1 ¼ p1x q1y r1
Rule 2: If x is A2 and y is B2; then f2 ¼ p2x q2y r2

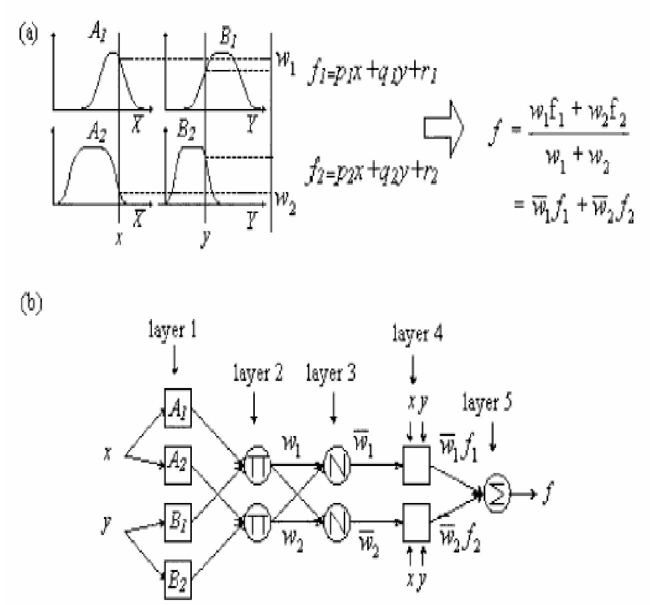


Fig.4 (a) the first- order of TSK fuzzy model; (b) corresponding ANFIS architecture.

If f1 and f2 are constants instead of linear equations, then we have zero-order TSK fuzzy model. Fig. 4(a) and (b) illustrates the fuzzy reasoning mechanism and the corresponding ANFIS architecture, respectively. Node functions in the same layer are of the same function family, as described below. Note that Oji denotes the output of the ith node in layer j:

Layer 1: Each node i in this layer generates a membership grades of a linguistic label. For instance, the node function of the ith node might be as in Eq. (7):

$$o_i^1 = \mu A_i(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}} \qquad (7)$$

Where x is the input to node i; and Ai is the linguistic label (small, large, etc.) associated with this node; and {ai; bi; ci} (or $\{a_i; c_i\}$ in the latter case) is the parameter set that changes the shapes of the membership function. Parameters in this layer are referred to as the 'premise parameters'.

Layer 2: Each node in this layer calculates the 'firing strength' of each rule via multiplication as in Eq. (8):

$$o_i^2 = w_i = \mu A_i(x) x \mu B_i(y),\ i = 1,2. \tag{8}$$

Layer 3: The ith node of this layer calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strengths as in Eq. (9):

$$o_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2,}\ i = 1,2 \tag{9}$$

For convenience, outputs of this layer will be called 'normalized firing' strengths.

Layer 4: Every node i in this layer is a square node with a node function as in Eq. (10):

$$o_i^4 = \overline{w}_i f_i = \overline{w}(p_i x + q_i y + r_i) \tag{10}$$

Where $\overline{w}_i$ is the output of layer 3, and is the parameter set. Parameters in this layer will be referred to as 'consequent parameters'.

Layer 5: The single node in this layer is a circle node labeled P that computes the 'overall output' as the summation of all incoming signals, i.e. as in Eq. (11):

$$o_i^5 = overalloutput = \sum_i \overline{w}_i f_i = \frac{\sum_i \overline{w}_i f_i}{\sum_i \overline{w}_i} \tag{11}$$

Thus, an adaptive network is presented in Fig.4 (b), which is functionally equivalent to a fuzzy inference system in Fig.4 (a). The basic learning rule of ANFIS is the back-propagation gradient descent [20], which calculates error signals (defined as the derivative of the squared error with respect to each node's output) recursively from the output layer backward to the input nodes. This learning rule is exactly the same as the back-propagation learning rule used in the common feed-forward neural-networks [21]. From the ANFIS architecture in Fig. 4, it is observed that given the values of premise parameters, the overall output f can be expressed as a linear combination of the consequent parameters. Based on this observation, a hybrid-learning rule is employed here, which combines the gradient descent and the least-squares method to find a feasible set of antecedent and consequent parameters [22, 23]. The details of the hybrid rule are given by Jang et al. [17], where it is also claimed to be significantly faster than the classical back-propagation method.
There are two passes in the hybrid learning procedure for ANFIS. In the forward pass of the hybrid learning algorithm, functional signals go forward till layer 4 and the consequent parameters are identified by the least squares estimate [18]. In the backward pass, the error rates propagate backward and the premise parameters are updated by the gradient descent. When the values of the premise parameters are fixed, the overall output can be expressed as a linear combination of the consequent parameters [18]:

$$\begin{aligned} f &= \frac{w_1}{w_1+w_2} f_i + \frac{w_2}{w_1+w_2} f_2 = \overline{w}_1 f_2 + \overline{w}_2 f_2 \\ &= (\overline{w}_1 x) p_1 + (\overline{w}_1 y) q_1 + (\overline{w}_1) r_1 + (\overline{w}_2 x) p_2 + (\overline{w}_1 y) q_2 \\ &\quad + (\overline{w}_1) r_2 \end{aligned} \tag{12}$$

Which is linear in the consequent parameters p1; q1; r1; p2; q2 and r2 .A flowchart of hybrid learning procedure for ANFIS is shown schematically Fig.5[18].

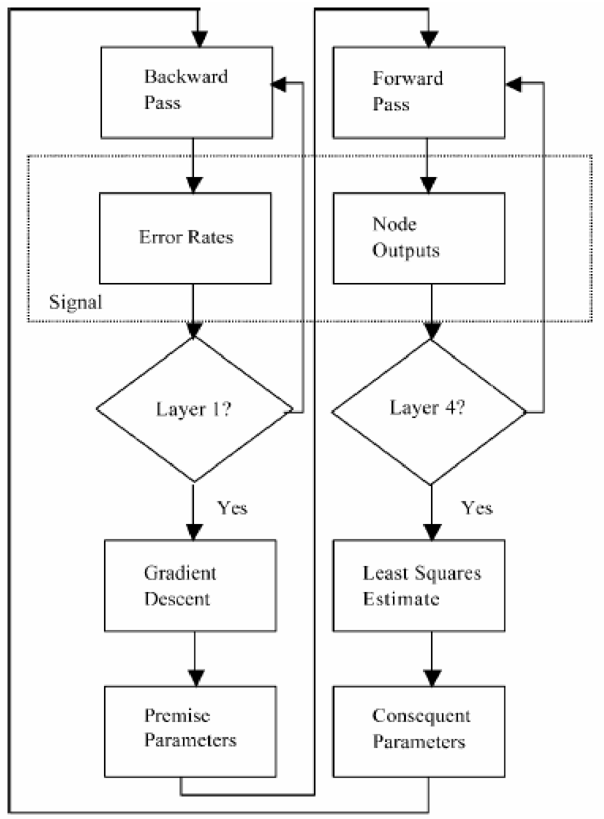


Fig.5 hybrid learning procedure of ANFIS

By the neural network, the initial FIS is trained to access the least possible error between the desired output (target) and the FIS output through the data set (details) which has been defined to the neural network to obtain the final FIS . The final FIS is identified the efficiency of ANFIS for the speech recognition by taking the test (checking) signals as input of the final FIS as in Fig.1.

The schematic of the architecture of ANFIS based on Sugeno fuzzy model is shown in Fig.6.

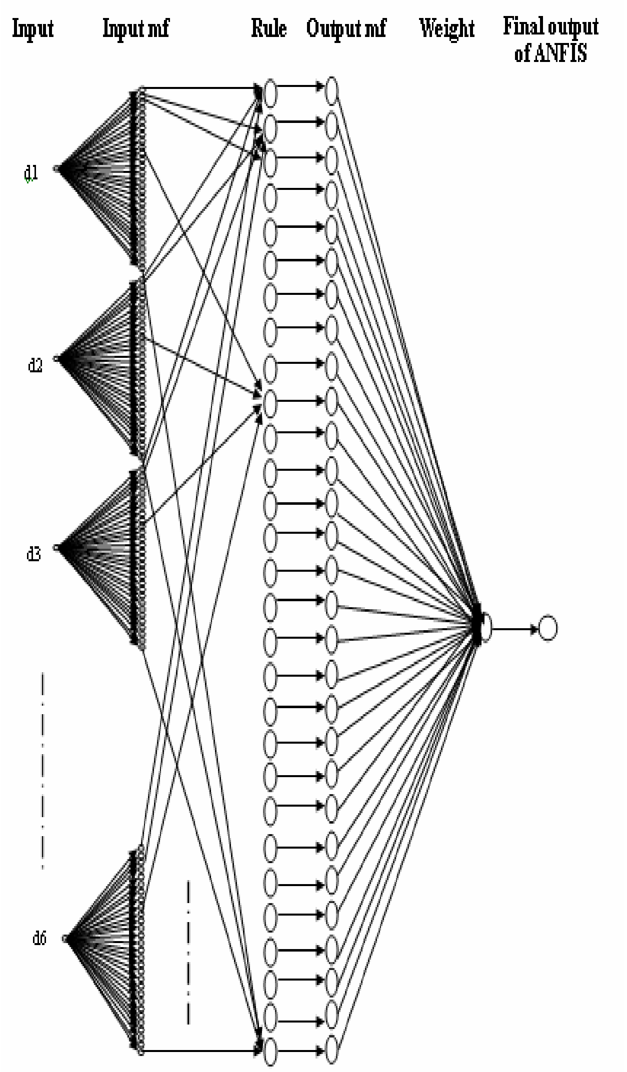


Fig.6 the schematic of ANFIS architecture based on Sugeno fuzzy model.

Multi-input one-output (MIOO) modeling indicated the six standard deviations (d1, d2, d3, d4, d5 and d6) of all sample training – one's, three's and six's- that loaded to ANFIS where each sample training is repeated ten times.
In MIOO modeling, these input data were trained and the output data were predicted.
In the other side, the number of coefficients in the linear equation in this paper is seven coefficients, because the six coefficients are multiplied with six input standard deviations in addition to the constant.

For simplicity, if the FIS has the input standard deviation x1, x2, x3, x4, x5 and x6 that are the first order Sugeno-type fuzzy inference model of six input standard deviations can be expressed with base fuzzy if-then rules as in Eq.(13):

If x1A1x2B1x3C1x4D1x5E1x6F1
Then f1=px1+rx2+qx3 +Sx4+SSx5+PPx6+ u (13)

Where p, r, q, S, SS, PP, u are linear output parameters. The ANFIS's architecture which has six input standard deviations of all sample training and one output.

A typical input–output (input standard 1(in 1), input standard 2(in 2) and output of ANFIS) surface of the training phase of samples –one's, three's and six's- is plotted in Fig.7.
The membership functions (MF) plots of all standard deviations (d1, d2, d3, d4, d5andd6) are shown in Fig.8.

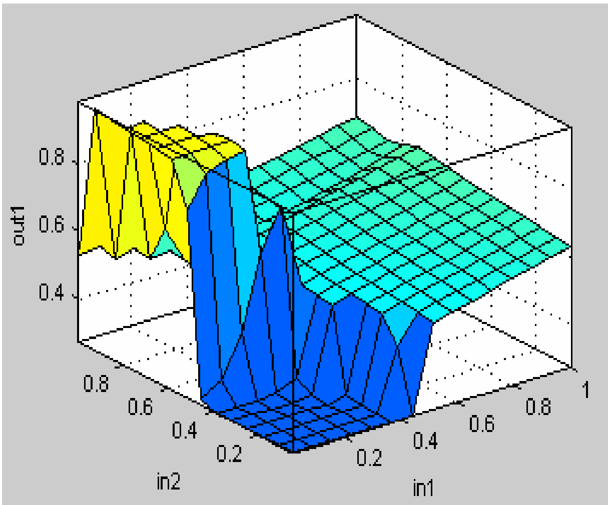
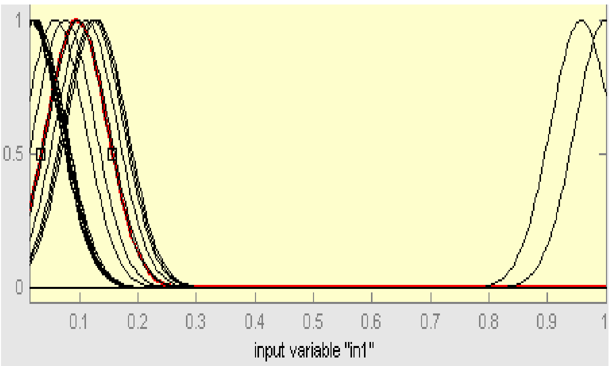
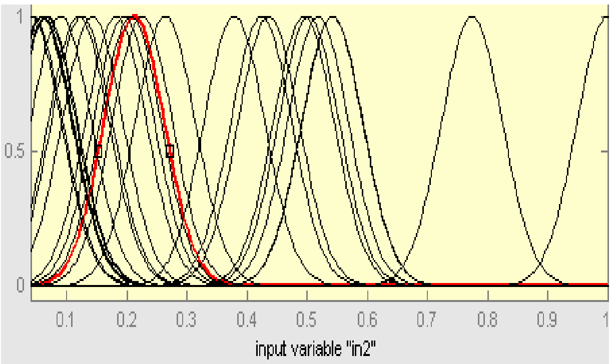


Fig.7 overall input–output surface of standard deviation input1, standard deviation input2 and the output of ANFIS

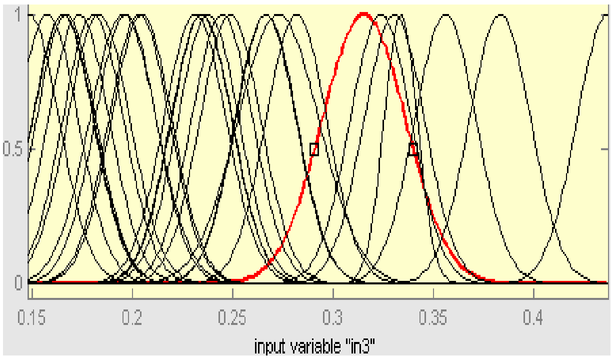
The membership functions (MF) plots of all standard deviations (d1, d2, d3, d4, d5 and d6) are shown in Fig.8.



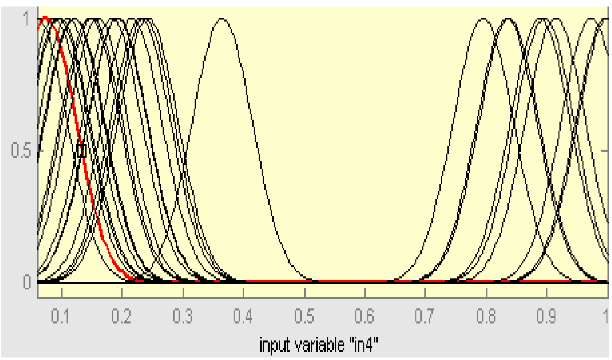
(a)



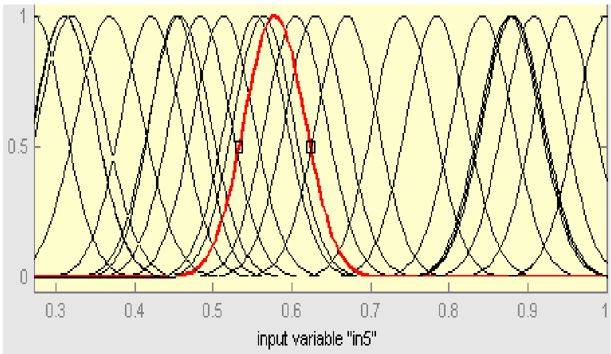
(b)



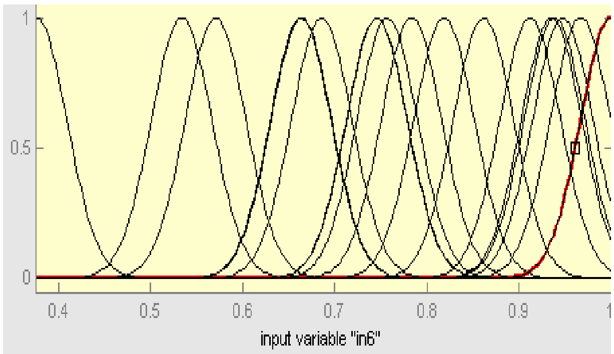
(C)



(D)



(e)



(f)

Fig.8 membership function plots of input standard deviations (d1, d2, d3, d4, d5 and d6). (a) A typical initial MF setting, where input standard deviation (in 1) range is between 0.014331 and 1, (b) a typical initial MF setting, where input standard deviation detail (in 2) range is between 0.03979 and 1, (c) a typical initial MF setting, where input standard deviation (in 3) range is between 0.01485 and 0.437, (d) a typical initial MF setting, where input standard deviation (in 4) range is between 0.0591 and 1, (e) a typical initial MF setting, where input standard deviation (in 5) range is between 0.27141 and 1, (f) a typical initial MF setting, where input standard deviation (in 6) range is between 0.3764 and 1.

3 Experimental Applications

A complete speech recognition system based on wavelet transform, subtractive clustering and ANFIS were used to carry out the goals of this research. The experimental applications consist of two parts: (a) data base used in this experimental study and (b) data acquisition.

3.1 Database used in this experimental study

A database was created from 3 English words. These 3 English words are given Table I. One speaker who spoke these 3 English words for training and checking phases ten times separately. The training and checking signals are Multiplied by 10 before background noise was added on it. The total number of tokens considered for training was 30 (1 speaker x 3 English noiseless words x 10 times repetition of a single word). The total number of tokens considered for checking was 30 (1 speaker x 3 English noiseless words x 10 times repetition of a single word).

Table I English words used in this study

.

English word	Pronunciation	No. of syllables
One	wan	1
Three	Thri	1
six	Siks	1

3.2 Data acquisition

All the original speech signals were acquired from the experimental set whose block diagram is showed in fig .3. The recode sounds are by the computer through a microphone. A sound file's size depends on the sample rate (frequency), the number of bits per sample, and whether the sound is mono or stereo. The sampling rate was used to as 22 KHZ with 16-bit resolution for all recorded speech tokens. The system parameters are illustrated in table II.

Table II the system parameters

Parameter	**Value**
Sampling rate	22KHZ , 16bits
Database	Isolated 3 English words
Speaker	1 Male
Bit rate	352Kbps
Number of Channels	1 channel (Mono)
Audio format	PCM

3.3 ANFIS architecture and training Parameters

ANFIS is intelligent classification using features from wavelet packet layer. The training parameters and the structure of the ANFIS used in this study are shown in table III.

Table III ANFIS architecture and training parameters

Architecture	**Training parameters**
Number of inputs	six input details
Number of outputs	One output
Rules number	29
Type of input membership function	Gauss-shaped
Optimization method(learning rule)	Hybrid optimization method
Error tolerance	0
Epoch	Once or many times for least error

4 Experimental Results

A Sample of 60 English isolated word is applied to the proposed algorithm. A background noise is added to the training and checking samples. A sample of 30 isolated words is used as training samples(signals) and the other 30 isolated words sample is used as unknown speech samples(signals) for the proposed algorithm. In these experiments, nearly 100 % correct classification was obtained at the ANFIS training with the training data among the 3 different English isolated words using one speaker mode as shown in Fig. 9.

Nearly 99% correct classification was obtained at the ANFIS training with checking data as shown in Fig.10. The training data appears as circles and the checking data appears as plusses.

The average error for testing the training and checking data against the trained FIS was shown in table IV.

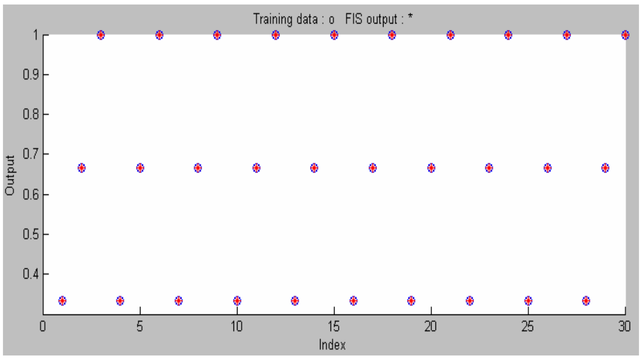


Fig.9 trained FIS against the training data" one's, three's and six's".

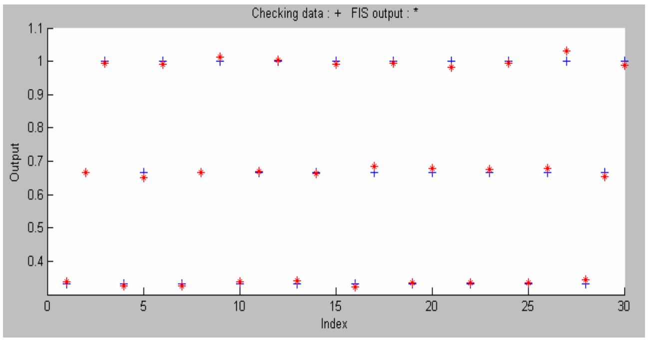


Fig.10 Trained FIS against the checking data "one's, three's and six's".

Table IV the average error for testing training and checking data against the trained FIS.

The average error for testing training	**Value**
The average error for testing training data against the trained FIS	2.6872e-7
The average error for testing checking data against the trained FIS	0.010641

These results are better than the reported work [1].

II. CONCLUSION

In this study, pattern recognition used the speech/voice signals of one speaker. The tasks of feature extraction and classification were performed using wavelet transform, subtractive clustering and ANFIS. The average error for testing training and checking data against the trained FIS indicated the performance of the speech recognition system. The wavelet packet transform proved to be very useful features for characterizing the speech signal, furthermore, the information obtained from the wavelet packet transform is related to the energy and consequently with the amplitude of signal. These features make the system suitable for speech recognition. This means that with this method, new information can be accessed with an approach different from the traditional analysis of amplitude of speech signal.
The subtractive clustering demonstrated to be an effective tool to take the details of the training signals and putting them in a group of clusters.
Through the subtractive clustering, each data point assumed as a potential center cluster. The likelihood that each data point would define the cluster center based on the density of surrounding data is calculated by the subtractive clustering.
The subtractive clustering generates an FIS with minimum number rules required to recognize the training data associated with each clusters.
The most important aspect of ANFIS is the ability to recognize on the words (one, three and six) efficiently through the work of training for the FIS by the neural network.
The recognition performance of this study shows the advantages of recognition speech system using ANFIS because it is rapid and gives the ease of use.

III. REFERENCES:

[1] E. Avci, and Z.H. Akpolat, Speech recognition using a wavelet packet adaptive network based fuzzy inference system, SinceDirect, vol.31, no. 3, 2006, pp 495- 503.
[2] M.Siafarikas, T.Ganchev, and N.Fakotakis, Wavelet packet based speaker verification," Odyssey, in Toledo, Spain, 2004, pp 257-264.
[3] Saito, and Naoki, Local feature extraction and its application using a library of bases, Proquest, in Yale University, 1994.
[4] J.B.Buckheit, and D.L.Donoho, Wavelet and reproducible research, CiteSeer, in Stanford University, New York: Springer, 1995, pp 55-82.
[5] E. Wesfreid, and M.V Wickerhauser, Adapted local trigonometric transforms and speech processing, IEEE SP, vol.41, no. 12, 1993, pp 3396-3600,.
[6] E. Visser, M. Otsuka, and T.Lee, A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments, Ingentaconnect, Speech Communication, vol.41, no. 2,, 2003, pp 393-407.
[7] P.A. Nava, and J.M. Taylor, Speaker independent Voice recognition with a fuzzy neural network, Proceedings of the Fifth IEEE international conference on fuzzy systems, vol.3, 1996, pp. 2049–2052.
[8] P. Nava, and J. Taylor, the Optimization of Neural Network Performance through Incorporation of Fuzzy Theory, Proceedings of the Eleventh International Conference on Systems Engineering, pp. 897-901.
[9] L. A. Zadeh, R. Yager, S. Ovchinnokov, R. Tong, and H. Nguyen, Fuzzy Sets and Applications, New York, 1987.
[10] G. Strang, Wavelets and filter banks, Wellesley–Cambridge Press, 1996, pp 1-34, pp 53-60, pp 155-172.
[11] M.S.Crouse, R.D.Nowrak, and R.G.Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, IEEE Tran. Signal processing, vol.41, no. 4, 2006, pp 886-902.
[12] G.W. Horgan, Using wavelets for data smoothing: a simulation study, Journal of Applied Statistics, vol.26, no. 8, 1999, pp 923-932.
[13] S.L.Chiu, Fuzzy model identification based on cluster estimation, J.Intell.Fuzzy systems, vol.2, no.3, 1994, pp.267-278.
[14] Mathworks, Fuzzy logic Matlab toolbox, ver.7, copyright 1984-2004.
[15] S. Nakkrasae, P.sophatsathit, W.R. Edwards, Fuzzy subtractive clustering based indexing approach for software components classification,

[16] M.Eftekhari, S.D.Katebi, Extracting compact fuzzy rules for nonlinear system modeling using subtractive clustering, GA and unscented filter,2007.
[17] J.S.R Jang, C.T Sun, E. Mizutani, neuro-fuzzy and soft computing: approach to learning and machine intelligence,London:prentice-Hall international, vol.8,no.8.1997.
[18] S. Akbuluta, A.S. Hasiloglub, and S. Pamukcu, Data generation for shear modulus and damping ratio in reinforced sends using adaptive neuro-fuzzy inference system, ScienceDirect, vol.24, no. 11, 2004, pp 805-814.
[19] M. Sugeno, G.T. Kang, structure identification of fuzzy model, fuzzy Sets Syst, vol.28, no.1, 1988, pp15-33,.
[20] P. Werbos, beyond regression: new tools for prediction and analysis in the behavioral sciences, PhD thesis, Harvard University, 1974.
[21] D.E. Rumelhart, G.E. Hinton, R.J. William, Learning internal representations by error propagation, In: D.E. Rumelhart, J.L. McCelland, ed Fuzzy modeling using generalized neural networks and Kalman filter algorithm editors, Parallel distributed processing: explorations in the microstructure of cognition , vols.1-8.the MIT Press ,1986, pp.62-318.
[22] J.R.S. Jang, ANFIS: adaptive-network-based fuzzy inference systems, IEEE Trans Syst Man Cybern, 1993, pp.85-665.
[23] J.R.S. Jang, Fuzzy modeling using generalized neural networks and Kalman filter algorithm, Proc Ninth Natl Conf Artif Intell (AAA-91), 1991, PP.7-762.