CHAPTER 4

ADAPTIVE WAVELET NEURO FUZZY INFERENCE SYSTEM BASED SPEAKER VERIFICATION SYSTEM

4.1 ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

Fuzzy systems are used in a wide range of industrial and scientific applications with the main application areas being fuzzy control, data analysis and knowledge based systems. Fuzzy controllers model the control strategy of a human expert to control a system for which no mathematical or physical model is required. They use a set of linguistic rules to describe the human behavior. They model the technical system and interpret the system description by using linguistic rules. However, the implementation of a fuzzy system is time consuming because there are no mathematical equations to determine its parameters (fuzzy sets and fuzzy rules). Therefore it requires algorithms through which fuzzy systems can automatically be learnt from data. Further, prior knowledge given by already existing fuzzy rules is not easy to introduce into the learning process. These drawbacks of fuzzy systems can be eliminated by the use of Neuro-fuzzy systems. They can be used to learn fuzzy rules and fuzzy sets, and also to optimize fuzzy systems derived by fuzzy clustering algorithms.

Neuro-Fuzzy Systems (NFS) are fuzzy systems that are trained by a learning algorithm derived from neural network. It learns on local information and makes the changes to the parameters of the fuzzy system. But the learning
process is not knowledge based, it is data driven. Adaptive learning is the important characteristics of neural networks. The main objective of ANFIS is to integrate the best features of fuzzy systems and neural networks. Representation of prior knowledge into a set of constraints to reduce the optimization search space as in Fuzzy systems and adaptation of back propagation network structure to automate the tuning of fuzzy system parameters are integrated. Adaptive Neuro Fuzzy Inference System (ANFIS) is used for system identification based on the available data. Modelling a system, based on conventional mathematical tools is not well suited, when the system is ill-defined and uncertain. But, a fuzzy inference system employing fuzzy if then rules can model the qualitative aspects of human knowledge and reasoning processes without using precise quantitative analyses (Jang 2009). They are good at tasks such as pattern matching and classification, function approximation, optimization and data clustering, while traditional computers are inefficient at these tasks, especially pattern-matching tasks because of their architecture (Avci et al 2005).

In this perspective, the aim of this work is to design a novel ANFIS architecture for speaker verification system. Here, an efficient method is proposed to design ANFIS for speaker verification by tuning fuzzy membership functions and neural networks. The structure of ANFIS is based on the basis of fuzzy rules including wavelet functions in the consequent parts of rules. In order to improve the verification rate and general capability of the ANFIS system, an efficient Genetic Algorithm (GA) approach is used to adjust the parameters of dilation, translation, weights, and membership functions. The performance of our proposed method is superior to that of existing methods.

Basically a Fuzzy Inference System (FIS) is composed of five functional blocks (Jang 1993) (Figure 4.1): a fuzzification interface which
transforms the crisp inputs into degrees of match with linguistic values; a **database** which defines the Input membership functions of the fuzzy sets used in the fuzzy rules; a **rule base** containing a number of fuzzy if-then rules; a **decision-making unit** which performs the inference operations on the rules; a **de-fuzzification interface** which transforms the fuzzy results of the inference into a crisp output.

![Fuzzy Inference System Diagram](image.png)

**Figure 4.1 Fuzzy Inference System**

In fuzzification, the input MFCC features are projected on the input membership functions. The membership value of each linguistic input in the feature vector is determined. The membership values are combined through a multiplication operator and firing strength of each rule is calculated. Then the consequent parameters of each rule are determined using the firing strength of the rule. At last, the qualified consequents are aggregated to produce a crisp output.

### 4.2 ARCHITECTURE OF ANFIS

Adaptive Neuro Fuzzy Inference System (ANFIS) was first proposed by Jang (1993). ANFIS can be easily implemented for a given input/output task and hence it is attractive for many application purposes. It consists of five layers (as in Figure 4.2) to perform a task. The rule set for the ANFIS shown in Figure 4.2 is given by
\[ IF(x_i \text{ is } A_i) \text{ AND } (x_2 \text{ is } B_i) \text{ THEN } f_1 = p_1x_1 + q_1x_2 + r_i \]
\[ IF(x_i \text{ is } A_2) \text{ AND } (x_2 \text{ is } B_2) \text{ THEN } f_2 = p_2x_1 + q_2x_2 + r_2 \]
\[ w_i = \max(MF_{A_i}(x_1), MF_{B_i}(x_2)) \]
\[ w_2 = \max(MF_{A_2}(x_1), MF_{B_2}(x_2)) \]  
(4.1)
\[ f = \frac{w_1f_1 + w_2f_2}{w_1 + w_2} \]

where \( A_i \) and \( B_i \) are fuzzy sets in the antecedent, \( MF \) is the Membership Function of the fuzzy sets, while ‘f’ is a crisp function in the consequent. The functions \( f_1 \) and \( f_2 \) are the polynomials in the input variables \( x_1 \) and \( x_2 \) and \( p_i \), \( q_i \), and \( r_i \) are the coefficients of the polynomial. The node functions in the same layer are of the same family as described below:

**Layer 1**  Every node \( i \) in this layer has a node function

\[ O_i^1 = MF_{A_i}(x_i) \]  
(4.2)

where \( x \) is the input to node \( i \), and \( A_i \) is the linguistic label associated with \( x_1, x_2 \) (components of feature vectors), etc. associated with the node function.

In other words, \( O_i \) is the membership function of \( A_i \) and it specifies the degree to which the given input satisfies the quantifier \( A_i \). Usually we choose \( \mu_{A_i}(x) \) to be bell-shaped with maximum equal to 1 and minimum equal to 0, such as the generalized bell function

\[ MF_{A_i}(x_i) = \frac{1}{1 + \left(\frac{x_i - c_i}{a_i}\right)^b} \]  
(4.3)
or the Gaussian function $MF_{A_i}(x_i) = \exp\left[-\left(\frac{x_i - c_i}{a_i}\right)^2\right]$ (4.4)

where $\{a_i, b_i, c_i\}$ is the parameter set. As the values of these parameters change, the Gaussian functions vary accordingly, thus exhibiting various forms of membership functions on linguistic label $A_i$. Parameters in this layer are referred to as *premise parameters*.

![Figure 4.2 Architecture of ANFIS equivalent to two input first order Sugeno fuzzy model with two rules](image)

**Layers**  
1 2 3 4 5

**Figure 4.2** Architecture of ANFIS equivalent to two input first order Sugeno fuzzy model with two rules

**Layer 2** Every node in this layer multiplies the incoming signals and sends the product out. For instance,

$$w_i = \mu_{A_i}(x_i) \times \mu_{B_i}(x_2) \cdots \times \mu_{C_i}(x_{26}) \quad i = 1, \ldots, 12$$

where in1,in2,...,in26 specifies the 26 components of input feature vector & $i$ – number of rules (4.5)
Each node output represents the firing strength of a rule.

**Layer 3** Every node in this layer calculates the ratio of the i-th rule’s firing strength to the sum of all rules’ firing strengths:

\[
\overline{w_i} = \frac{w_i}{w_1 + w_2 + \ldots + w_{12}}, \quad i = 1, \ldots, 12
\]  

(4.6)

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

**Layer 4** Every node in this layer is a node with a node function

\[
O_i^4 = w_i f_i = w_i \left( \sum_{j=1}^{26} p_{ij} x_j + q_i \right) \quad i = 1, \ldots, 12
\]  

(4.7)

where \( w_i \) is the output of layer 3, and \( \{p_{ij}, q_i\} \) is the consequent parameter set.

**Layer 5** The single node in this layer is a node that computes the overall output as the summation of all incoming signals, i.e,

\[
O_i^5 = \sum_i \overline{w_i f_i} = \frac{\sum_i w_i f_i}{\sum_i w_i} = \text{overall output}
\]  

(4.8)

### 4.3 ANFIS LEARNING ALGORITHM

Though the gradient Descent algorithm is enough to identify the parameters in an adaptive network, the method is generally slow and likely to become trapped in local minima. Therefore hybrid learning rule which combines the gradient method and the least squares estimate (LSE), is used in ANFIS to identify the parameters.
The hybrid learning procedure involves two steps (Jang 1993). The first step is the forward pass, in which, functional signals go forward till layer 4 and the consequent parameters are identified by the least squares estimate. In the second step—backward pass, the error rates propagate backward and the premise parameters are updated by 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:

\[
f = \frac{w_1}{w_1 + \ldots + w_{12}} f_1 + \frac{w_2}{w_1 + \ldots + w_{12}} f_2
\]

\[= \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}
\]

where

\[
f_1 = p_1 x_1 + q_1 x_2 + r_1
\]

\[
f_2 = p_2 x_1 + q_2 x_2 + r_2
\]

which is linear in consequent parameters \(p_1, q_1, r_1; p_2, q_2\) and \(r_2\). The consequent parameters determined in the forward pass are optimal under the condition that the premise parameters are fixed. Consequently, the hybrid approach converges much faster since it reduces the search space dimensions of the original pure back propagation method. There are several ways to combine the gradient descent and least square method. As stated in (Jang 1993), the tuning of membership functions by learning algorithm is decided based on the available size of the data. If the data is too large then fine tuning of membership functions is recommended; whereas if the data size is small then membership functions should be kept fixed throughout the learning process. Because the user defined membership functions represent important information that may not be described by data set. Hence, based on the size of input data, the tuning of premise and/or consequent parameters is chosen. However, the performance of ANFIS is better when both premise and consequent parameters are updated through hybrid learning algorithm. Since
these parameters are decoupled in the hybrid learning rule, the speed of learning can be increased by using variants of gradient method or other optimization techniques on the premise parameters.

4.4 PROPOSED SPEAKER VERIFICATION SYSTEM USING AWFIS

In this work, an Automatic Speaker Verification System using Adaptive Wavelet Neuro-Fuzzy Inference System (AWNFIS) is proposed. The block diagram that describes this work is shown in Figure 4.3. A combination of wavelet bases functions and AWFIS is used to efficiently extract the features from pre-processed real speech signals for the purpose of automatic speaker verification. In this work, the novelty presented is the development of an effective Adaptive Wavelet based Neuro Fuzzy Inference System (AWNFIS) and it is used to model the speaker in speaker verification. Development of this AWFIS increases the correct speaker verification rate and provides an efficient classification method in automatic speaker verification area.

![Figure 4.3 Speaker Verification System using AWFIS](image-url)
The performance of the speaker verification system is evaluated on TIMIT database. The speech signals are pre processed and MFCC features are extracted. The Mel Frequency Cepstral Coefficients are used as the input for AWNFIS model. AWNFIS model is trained with the training feature vectors to produce the target speaker model. As shown in Figure 4.4, initially, input membership functions are generated from the input features of each speaker. Then the rules are framed from the membership functions. The Rule base of Fuzzy Inference System (FIS) is a model that maps input characteristics to input membership functions (premise parameters), input membership function to rules, rules to a set of output membership functions (consequent parameters), and the output membership function to a single-valued output which makes a decision. When these model parameters are updated according to the error (output-target), optimum consequent parameters can be obtained. Once the training is completed, the input MFs and the corresponding output MF of the AWNFIS approximate the training vectors learnt. In the inference/testing phase, the testing feature vectors that are to be identified are applied as input to the trained AWNFIS. Now, AWNFIS recognizes the input vector by finding the firing strength of the feature vector through the input MFs and produces the consequent value as output. The decision is made according to the output obtained by the network.

The speaker model is developed using Sugeno fuzzy reasoning mechanism (Sugeno & Kang 1988; Takagi & Sugeno 1985), whose parameters are tuned off-line through hybrid learning algorithm (Loganathan, & Girija 2013). A comparison is made between the proposed system and the system using conventional BPN with the same input features.
Figure 4.4 Proposed systems with AWNFIS architecture

4.5 WAVELET BASES IN AWNFIS

Wavelet bases functions are rapidly evolving in fields as diverse as speech recognition, speaker recognition, telecommunications and radar target recognition. Because of their suitability for analyzing non-stationary signals, they have become a powerful alternative to Fourier methods in many speaker recognition applications (Nava & Taylor 1996a, b). One of the key features of wavelet bases is the smooth characterization of various function spaces in terms of wavelet coefficients (Shung-Yung Lung 2008). It uses a varying window size, being wide for slow frequencies and narrow for high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges. In addition, due to the fact that windows are adapted to the transients of each scale, wavelets lack of the requirement of stationary (Nava & Taylor 1996a, b).
Controlling and modeling systems which are complex and nonlinear often challenge the control engineers in design. AWNFIS provides good solution to this problem. However, these systems require formal synthesis techniques that provide global stability and acceptable performance. Here we have introduced the wavelet basis functions in the input membership functions and consequent part of AWNFIS in order to improve the performance of the system. AWNFIS not only offers a framework for combining linguistic information and numerical data but also takes advantages of rigorous approximation of wavelet basis function on the feature space.

Given a mother wavelet function $\psi \in L^2 (\mathbb{R}^n)$, consider the sequence functions $\psi_{j,k}$ generated by dilating and transferring mother wavelet function in the following form

$$\psi_{j,k}(x) = \det(D_j^{1/2}\psi(D_j x - \Lambda_k))$$

(4.10)

where $j = [j_1, ..., j_n]^T \in \mathbb{Z}^n, k \in \mathbb{Z}^n$;

Dilation matrix $D_j = \text{diag}(a^{j_1}, ..., a^{j_n})$,

(4.11)

Translation matrix $\Lambda_k = \text{diag}(b_1, ..., b_n), a > 1, a \in \mathbb{R}, b = (b_1, ..., b_n) \in \mathbb{R}^n$

(4.12)

Conditions on $\psi$, $a$ and $b$ to guarantee that a multi scaling wavelet frame for $L^2(\mathbb{R}^n)$ have been obtained. Let the multidimensional wavelet functions be the generalizations of one-dimensional wavelet functions, i.e.,

$$\psi(x) = \psi_1(x_1) ... \psi_n(x_n)$$

(4.13)
That is, applying one dimensional wavelet transform separately in
each of \( n \) orthogonal directions, we have

\[
\hat{\psi}(\omega) = \hat{\psi}_1(\omega_1) \ldots \hat{\psi}_n(\omega_n)
\]  
(4.14)

where \( \hat{\psi}(\omega), \hat{\psi}_1(\omega_1) \) etc., are the Fourier transform of \( \psi(\omega) \) and \( \psi_1(x_i) \)
respectively, which must satisfy the admissibility condition

\[
\int \left| \frac{\hat{\psi}_j(\omega)}{\omega} \right|^2 d\omega_j < \infty
\]  
(4.15)

The fuzzy model proposed by Takagi & Sugeno (1985), which is
described by fuzzy IF- THEN rules, is of the following form:

Rule m: IF \( x_i \) is \( \tilde{A}_{m1} \) and...and \( x_n \) is \( \tilde{A}_{mn} \) THEN \( y = d_{m0} + d_{m1}x_1 + \ldots + d_{mn}x_n \)

(4.16)

where \( \tilde{A}_{m1}, \tilde{A}_{m2}, \ldots and \tilde{A}_{mn} \) are fuzzy sets, \( y \) is the output and
\( x = x_1, x_2, \ldots, x_n \) \( \top \) is the input vector. THEN part of Sugeno’s fuzzy model is a
linear combination of premise variables to represent local linear input –output
relationship of nonlinear systems. The proposed AWNFIS has a rule in the
form of

Rule m: IF \( x_i \) is \( \tilde{\psi}_{m1} \) and...and \( x_n \) is \( \tilde{\psi}_{mn} \) THEN \( y = d_{m0} + d_{m1}x_1 + \ldots + d_{mn}x_n \)

(4.17)

where \( \tilde{\psi}_{m1}, \tilde{\psi}_{m2}, \ldots and \tilde{\psi}_{mn} \) are wavelet basis functions used as input
membership functions and \( \psi(x_i), \ldots, \psi(x_n) \) are shifted and scaled functions of
its mother wavelet. The mother wavelet used in this work is Daubechies wavelet. Thus for this AWNFIS architecture, the output $y_i$ is

$$y_i = \frac{\sum_m [d_m \prod_{n=1}^n \mu_{\psi_{m1}}(x_i)\psi(x_i)]}{\sum_m \prod_{n=1}^n \mu_{\psi_{m1}}(x_i)}$$

(4.18)

where $\mu_{\psi_{m1}}(x_i)$ is the membership function of $\psi_{m1}$ and $d_m$ are the weights of the inputs. Use of wavelet basis function as membership function in the input layer provides decomposition of the input signal into multilevel. If the firing strength of each rule of the proposed AWNFIS is the wavelet basis function, it is obvious that the THEN part is the linear combinations of coefficient $d_m$ and functions of consequent variables. The aim of this modification is to combine the AWNFIS and the wavelet theory such that the AWNFIS model can share the advantages of wavelet transforms.

### 4.6 IMPLEMENTATION OF PROPOSED SYSTEM

An automatic Speaker Verification system based on adaptive wavelet neuro fuzzy inference system is developed. Figure 4.3 shows the proposed system. It consists of two important parts: (a) data pre processing and feature extraction and (b) modelling and verification using Adaptive Wavelet based Neuro Fuzzy Inference System (AWNFIS).

#### 4.6.1 Database

TIMIT Acoustic-Phonetic Continuous Speech data base is used in this work. TIMIT contains broadband recordings of 630 speakers from eight major dialects of American English, each reading ten phonetically rich sentences. The TIMIT corpus includes time-aligned orthographic, phonetic
and word transcriptions as well as a 16-bit, 16 kHz speech waveform file for each utterance. The system’s performance is evaluated for 50 female speakers from the same dialect region (New England) separately.

4.6.2 Feature Extraction

The purpose of this module is to convert the speech waveform to some type of parametric representation. The TIMIT database is used to evaluate the performance of the system. The input speech signal is divided into overlapping frames and hamming windowed. For each frame, 13 Mel Frequency Cepstral Coefficients and its 13 derivatives are calculated. Thus, totally 26 coefficients per frame are obtained. These feature vectors are used as inputs for the system. 80% of the input data are used as training data from which the fuzzy rules are formed and the remaining 20% data are used as test data. To obtain the speech features as mentioned above the Voice box toolbox of MATLAB software is used.

4.6.3 Subtractive Clustering

The features extracted from all speakers will have to be clustered before modelling and training. Clustering algorithms are used extensively, not only to organize and categorize data, but also for data compression and model construction (Muhammad Ridwan et al, 2009). The most representative off-line clustering techniques used in conjunction with RBFN and Fuzzy modelling are C-means (hard) clustering, fuzzy C-means clustering, mountain clustering and subtractive clustering.

AWNFiS training makes use of the subtractive clustering technique (Jang 1993), Subtractive clustering uses scatter partitioning to cluster the input data. For a collection of n data points \(\{x_1, x_2, x_3...x_n\}\) in an M-dimensional space, a density measure is calculated at each data point. The
data point with the highest density measure is selected as the first cluster center. Before finding the next cluster center the stopping criterion is verified by setting threshold. Again the density measure of each data point is revised according to the first cluster center. Then the next cluster center is selected and the density measures of all data points are revised again. This process is repeated until a sufficient number of cluster centers are generated.

The entire training data set has been classified into 12 classes by means of subtractive clustering. Subtractive clustering finds the cluster centres based on the density of surrounding data points. For that it requires the range of influence (=0.5) which specifies the size of cluster in each of the data dimension.

4.6.4 Fuzzy Logic Model Adopted

The Sugeno fuzzy model (as in Figure 4.5) for the proposed speaker verification system is generated for the input MFCC vectors of 26 dimensions. An initial Fuzzy Inference system is generated for AWNFIS training. This is accomplished by extracting a set of rules that models the data behaviour. The model is generated with 26 input parameters (the size of feature vector) with 12 (rules) wavelet membership functions/input parameter and one output node. From the antecedent membership functions of the rules, the consequent equations are derived. The solution of the 12 consequent equations gives the output of the FIS.

The cluster centers generated using subtractive clustering is initialized as the centers for fuzzy rules’ premise in Sugeno fuzzy model. Once the premise part has been determined, the consequent part can be estimated by the least squares method.
Figure 4.5 Architecture of Sugeno AWFIS for the proposed system

Figure 4.6 Hybrid Learning Procedure for ANFIS
The AWNFIS consists of five layers (as in figure 4.5) to perform the task. The node functions in the same layer are of the same family as described below:

**Layer 1** Every node $i$ in this layer is a node with a node function

$$O_i^1 = \mu_{A_i}(x)$$

(4.19)

where $x$ is the input to node $i$, and $A_i$ is the linguistic label $In1$, $In2$(components of feature vectors), etc. associated with the node function.

In other words, $O_i$ is the membership function of $A_i$ and it specifies the degree to which the given input satisfies the quantifier $A_i$. Usually we choose $\mu_{A_i}(x)$ to be bell-shaped with maximum equal to 1 and minimum equal to 0, such as the generalized bell function

$$\mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x - c_i}{a_i}\right)^2}$$

(4.20)

or the Gaussian function

$$\mu_{A_i}(x) = \exp\left[-\left(\frac{x - c_i}{a_i}\right)^2\right]$$

(4.21)

where $\{a_i, b_i, c_i\}$ is the parameter set. As the values of these parameters change, the Gaussian functions vary accordingly, thus exhibiting various forms of membership functions on linguistic label $A_i$. Parameters in this layer are referred to as premise parameters.

**Layer 2** Every node in this layer multiplies the incoming signals and it is given to the next layer. For instance,

$$w_i = \mu_{A_i}(in1) \times \mu_{A_i}(in2) \times \ldots \times \mu_{A_i}(in26) \quad i = 1\ldots12$$

where $in1, in2, \ldots, in26$ specifies the 26 components of input feature vector & $i$ – number of rules

(4.22)

Each node output represents the firing strength of a rule.

**Layer 3** Every node in this layer calculates the ratio of the $i$-th rule’s firing strength to the sum of all rules’ firing strengths

$$\bar{w}_i = \frac{w_i}{w_1 + w_2 + \ldots w_{12}}, \quad i = 1,\ldots,12$$

(4.23)
For convenience, outputs of this layer will be called normalized firing strengths.

**Layer 4** Every node in this layer is a node with a node function

\[ O^4_i = w_i f_i = w_i \left( \sum_{j=1}^{26} p_{ij}i + q_i \right) \quad i = 1, \ldots, 12 \]  

(4.24)

where \( w_i \) is the output of layer 3, and \( \{p_{ij}, q_i\} \) is the consequent parameter set.

**Layer 5** The single node in this layer is a circle node that computes the overall output as the summation of all incoming signals, i.e,

\[ O^5_i = \sum_j w_j f_j = \frac{\sum_j w_j f_j}{\sum w_j} = \text{overall output} \]  

(4.25)

Thus an adaptive network which is functionally equivalent to a type-3 fuzzy inference system is constructed. The initial AWNFIS is trained by the neural network to obtain the least possible error between the desired output (target) and the FIS output. The result of training is the final FIS.

The AWNFIS is trained using hybrid learning algorithm, which combines the gradient method and the least squares estimate (LSE) to identify parameters. The parameters of the neuro fuzzy architectures define the shape of membership functions of the fuzzy sets in the IF-THEN rules. Tuning these parameters thus optimizes the form of the rules. Moreover, the number of rules in the ‘rule base’ of the fuzzy systems can be determined using a subtractive clustering method.

The hybrid learning algorithm as shown in figure 4.6, involves two steps. The first step is the forward pass, in which, functional signals go forward till layer 4 and the consequent parameters are identified by the least squares estimate. In the second step-backward pass, the error rates propagate backward and the premise parameters are updated by the gradient descent. AWNFIS uses a hybrid learning algorithm to identify the membership function parameters of single-output, Sugeno type Fuzzy Inference Systems (FIS). A combination of least-squares and back propagation gradient descent
methods are used for training FIS membership function parameters to model a given set of input/output data.

4.7 SUMMARY

An automatic text independent speaker verification system was developed using Adaptive Wavelet Neuro-Fuzzy Inference System. This system is constructed based on the wavelet transform theory in ANFIS architecture. Since the wavelet transform is a powerful tool for non stationary signal analysis, the performance of AWNFIS is better than the ANFIS.