Speech Recognition and Retrieving using Fuzzy Logic System

Arjuwan M. Abduljawad Al-Jawadi
Technical College , Foundation of Technical , Engineering , Mosul , Iraq
(Received 21 / 4 / 2009 , Accepted 1 / 11 / 2009)

Abstract
Speech Analysis is one of the most interesting fields in Digital Signal Processing in which many researches have been done on it based on different materials tools and scientific programs to produce an analysis that start from speech production, processing, coding and recognition, Chester, F.J .Taylor and M.Doyle were the first to apply the analysis of speech signal [1]. In this research, females and males speech samples of the word, 'Close', were used to build a system in Neural Network and Fuzzy Logic to recognize the male from female speech voice and compared between the results of the two systems, then the system of the fuzzy logic was developed based on three features of the speaker voice which are energy value of the signal ,power spectrum of the signal and vowel sound “O” in the word close in the speech samples to increase its ability in recognizing an individual speaker and to increase system security against intruders by making the system recognizes the speech of a one person giving a voice acceptance authority to that person and make an access denied to others to prevent accessing the system. The system shows good results during testing operation using samples of one person against others female's and male's samples.

Introduction
The speech signal is a high redundant and non-stationary signal. This attributes causes the speech signal to be very challenging to work [2]. The speech recognition systems fall into two categories according to [3]:

1. Speaker dependent systems that are used and often trained by one person.
2. Speaker independent systems that can be used by anyone.

In general, speaker recognition can be subdivided into speaker identification (who is speaking?) and speaker verification (Is the speaker who we think he or she is?) .In addition, speaker identification can be closed-set (The speaker is always one of a closed set used for training.) or open-set (speakers from outside the training set may be examined.). Also, each variant may be implemented as text-depended (The speaker must utter one of a closed set of words.) or text-independent (The speaker may utter any type of speech) as in [4].

Human speech can be separated into two distinct sections: sound production and sound shaping. Sound production is caused by air passing across the vocal cords (as in "a", "e", and "o") or from a construction in the vocal tract (as in "sss", "p", or "sh"). Sound production using the vocal chords is called voice speech and it has been taken as a feature that used in tracking the vowel sound during speech in male and female samples in my research; unvoiced speech is produced by the tongue, lips, teeth, and mouth .In signal processing sound producing is called excitation. Sound shaping is a combination of the vocal tract, the placement of the tongue, lips, teeth and the nasal passages. For each fundamental sound or phoneme, of English, the shape of the vocal tract is somewhat different which may lead to a different sound. In signal processing, sound shaping is called filtering [3]. MATLAB program was used in the work which a very useful program in processing non-stationary signals such as speech signal and many systems was designed based on its tool as in [5] .

Materials and Methods

1. Recording the Speech Samples
The first strep in the research was building a data base of speech samples, this step requires to collect both females and males samples and an individual voice speaker samples for a female the selected word for recording the speaker's voices was saying,' Close', word as the vowel sound can be heard clearly . Dell Laptop computer was used for recording the speech signals for both males and females samples and using Sound Recorder software of windows (XP) and a Microphone (Creative) model HS-350.

2. Low-Pass Filter
It is very important to use low-pass filter because the recording speech signals samples have noise that produces during speaking and usually that noise lies at the higher frequencies while the speech data lies at the lower frequencies so this filter was used twice, first when reading the speech signal in MATLAB program environment as shown in Fig (1) and second when calculating the power of the Discrete - Fourier Transform to separate the data of the power amplitudes for the speech signal from the noise. As shown in Fig (2).

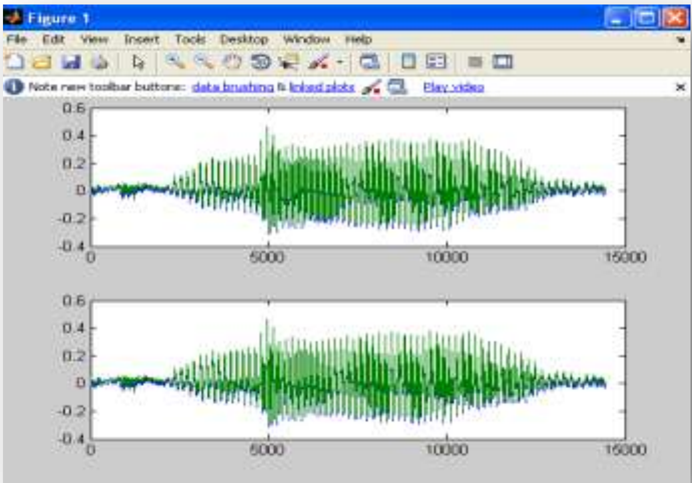

Fig (1) the Effect of the Low-Pass Filter on a sample of the Speech Signal

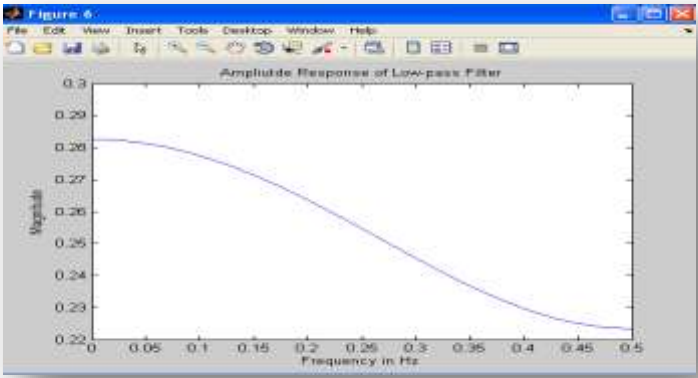

Fig (2) the Effect of the Low-Pass Filter on the Power of a sample of the Speech Signal

3. Features Extraction

1.3- Pitch

Pitch is the most distinctive difference between male and female speakers. A person's pitch originates in the vocal cords, and the rate at witch the vocal cords /folds vibrate is the frequency of the pitch. The reason pitch differs between sexes is the size, mass and the tension of the laryngeal trade which includes the vocal folds and the glottis (the spaces between and behind the vocal folds).The fundamental frequency or pitch of the human voice is about 250 Hz and the fold length is about 10.4mm . After puberty the human body grows to its full adult size, changing the dimension of the larynx area. The vocal fold length in males increases to about 15-25mm. While female's vocal fold length increases to about 13-15mm. The average pitch falls between 60 and 120 Hz ,and the range of female's pitch can be found between 120 and 200 Hz. Females have a higher pitch range than males because the size of their larynx is smaller [4]. This feature was very helpful in recognizing between the male and the female voice in many researches, one of them was proved using fuzzy logic system as in [6]. Based on the pitch feature I found in the research that the amplitudes of the power of the speech signal is another feature that can be used in making recognition between the male and the female voice By applying the Fast Fourier transform algorithm.

2.3- Fast Fourier Transform (FFT)

The discrete Fourier transform (DFT) with a million points are common in many applications. Modern signal and image processing applications would be impossible without an efficient method for computing the discrete Fourier transform which transform time or space – based data into frequency – based data [7]. The DFT was used as a features extractor because the frequency magnitude does contain information about the pitch and the formants. Beside the spectral magnitude also holds a great deal of other information beside the pitch and the formants magnitude .The DFT of the vector of length n is another vector y of length n.

$$X(k) = \sum_{j=1}^{N} x(j) W_N^{(j-1)(k-1)}$$

$$\boldsymbol{X(j) = (1/N) \sum_{k=1}^{N} X(k) W_N^{(j-1)(k-1)}}$$

$$W_{N=e^{(-2\pi i)/N}}$$

Where, X *(k)*: The discrete input signal in time domain

X (j): The discrete input signal in frequency domain

N: The number of the samples

j: Sample no. of discrete time domain

K: Sample no. of frequency domain

The last equation is an Nth root of unity. When using FFT algorithms, a distinction is made between the window length and the transform length. The window length is the length of the input data vector. It is determined by, for example, the size of an external buffer. The transform length is the length of the output, the computed DFT. An FFT algorithm pads or chops the input to achieve the desired transform length. Fig (3) illustrates the two lengths.

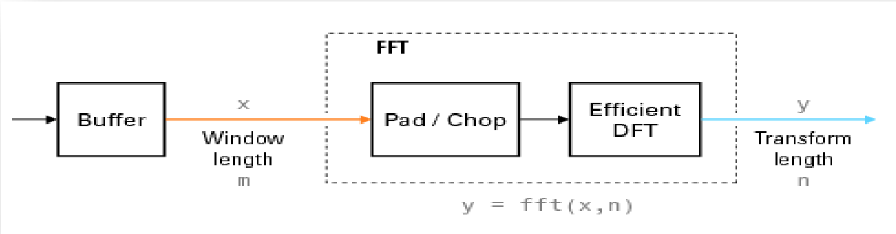

Fig (3) The Fast Fourier Transform algorithm chaps or pads the input

A program in MATLAB 7.6.0 (R2008a) was written using its commands to calculate the window length of the input signal which is the speech signal. The execution time of the fast Fourier transform algorithm depends on the transform length .It is fastest when the transform length is a power of two. The algorithm allows estimating component frequencies in data from a discrete set of values sampled at a fixed rate. From calculating the window length of the speech signal, the transform length of the signal was calculated by finding the smallest power of two that is greater than or equal to the absolute value of the window length for the speech signal [7]. The transform signal was used to calculate the power of the DFT for the speech signal .The result was shown in the plot of power versus frequency which is called a periodogram as shown in Fig (4).

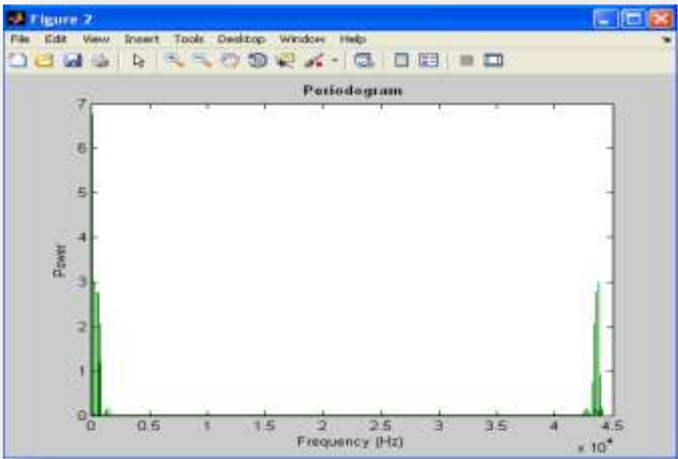

Fig (4) The Periodogram Plot of the speech signal

3.3- Spectrogram

Signals such as speech are composed by many different ranges of frequencies. Thus frequency representation is necessary in the interpretation of a speech signal. Spectrogram is one of well-known frequency representation of the original speech. The vertical axis corresponds to time .The intensity of the pattern at any instant of time corresponds to the energy level. Spectrogram allows users to know the amount of energy a speech, might have in terms of frequency scale. This is useful tool to detect voiced and unvoiced areas, and identifying the relevant frequency that is composed in this speech. There are other varieties of researches that can be performed using spectrogram .The amount of information a spectrogram can give is enormous and many speech researches can identify plain English text from spectrogram [3] . In the research spectrogram tool was used in MATALB language because the power spectrum of a signal represents the contribution of every frequency of the spectrum to the power of the overall signal. Beside it is very helpful in noise cancellation and system identification.

The process started Using MATLAB Signal Processing Block set, by loading a speech signal from MATLAB workspace for the word,' Close', for each sample. Then separating the speech signal to number of segments called frames with given sample time (1/8000) and 80 samples per frame. The output buffer size per channel (128) with (48) buffer overlap. Based on theses sitting parameters ,the first output frame contains 48 initial condition values followed by the first input frame .The second output frame contains the last 48 values from the previous frame followed by the second 80 samples from the second input frame and so on. The buffering of the input signal into an output signal with 128 samples per frame was to minimize the estimating noise added to the speech signal. After this step was completed, the Fourier transform was taken for the signal using Periodogram block which calculate a nonparametric estimation of the power spectrum of the speech signal. This operation is depending on windowing process like Hamming and the purpose of it is to make intense in some area while irrelevant in other areas. It was applied periodically to the speech signal and average two spectra at one time. The length of FFT was assumed 128, which is the number of samples per frame. Fig (5) shows the spectrogram of the speech signal system that was designed in MATLAB signal processing Block set [7].

The Only changes that were made in the designed system is how to load the input speech signal from MATLAB work space by using Simulink signal processing sources ,signal from workspace block and since the speech signal was produces as two dimension signal so another change was made through the selector block. The Vector Scope block was used to display the power spectrum of the speech signal as shown in Fig (6). While the spectrogram of the speech signal was viewed using Matrix Viewer as shown in Fig (7). The speech signal represents speech sample for female's voice saying "Close".

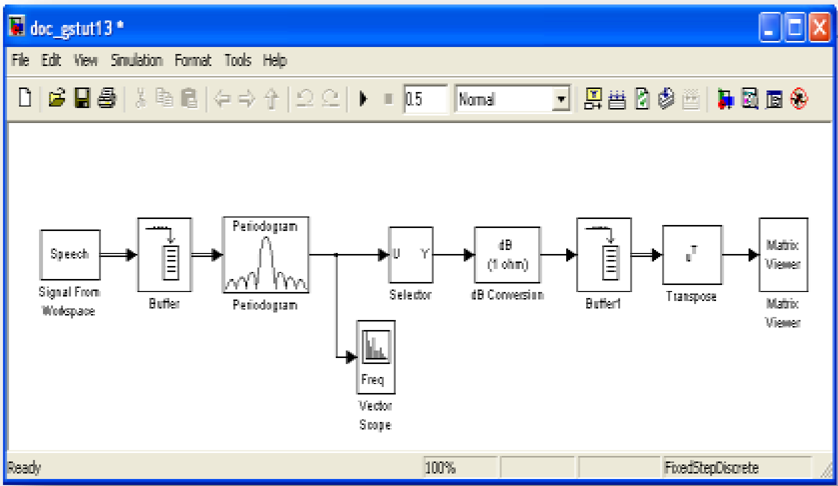

Fig (5) Viewing Spectrogram of the Speech Signal using MATLAB Signal Processing Block set

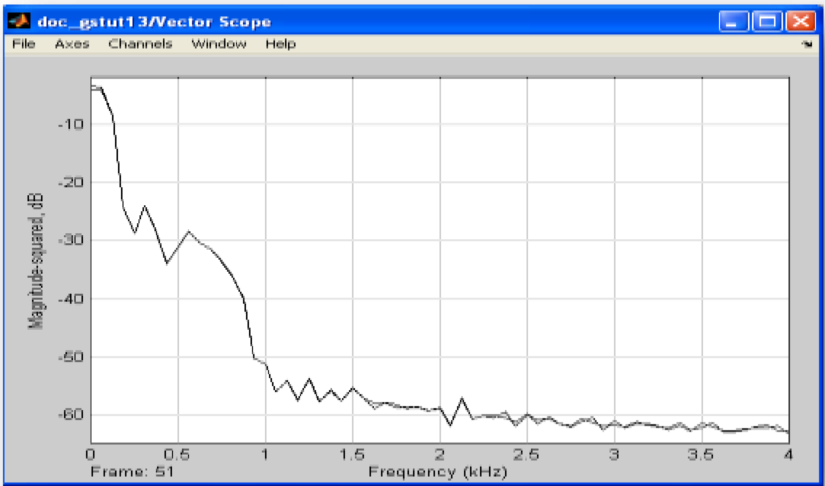

Fig (6) The Vector Scope Window displaying a Sequence of Power Spectrums, one for each Window of the Original Speech Signal

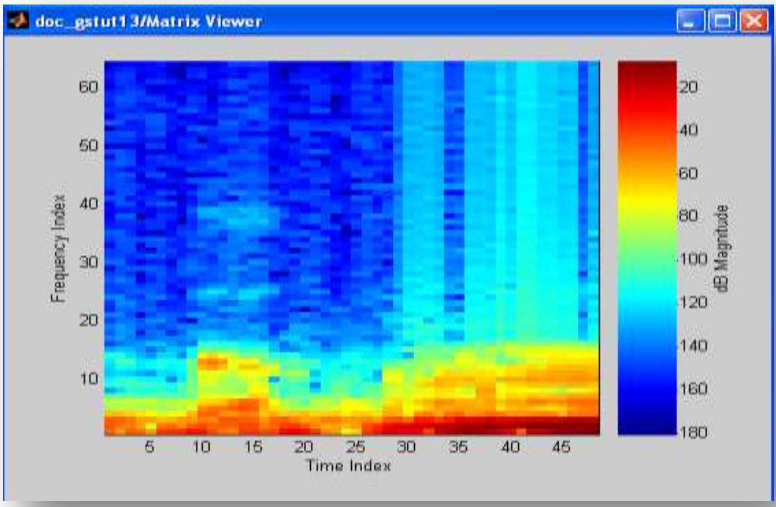

Fig (7) The Matrix Viewer Window displaying the Spectrogram of the Speech Signal

From Fig (7), we can notice the harmonics that are visible in the speech signal when the vowel, "o" is spoken in the word, "Close", where most of the speech signal's energy is concentrated. AS seen in the Fig (7) the Spectrograms represent the color-based visualizations of the evolution of the power spectrum of a speech signal as this signal is swept through time that was calculated using periodogram power spectrum estimation method.

4. Training Neural Networks for Speech Recognition

An Artificial Neural Networks (ANNs) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANNs are configured for a specific application, such as pattern recognition or data classification, through a learning process [8]. The best-known example of a neural network training algorithm is back propagation (Patterson, 1996; Haykin, 1994; Fausett, 1994). Modern second-order algorithms such as conjugate gradient descent and Levenberg-Marquardt (Bishop, 1995; Shepherd, 1997) are substantially faster (e.g., an order of magnitude faster) for many problems, but back propagation still has advantages in some circumstances, and is the easiest algorithm to understand. In back propagation, the gradient vector of the error surface is calculated. This vector points along the line of steepest descent from the current point, so we know that if we move along it a "short" distance, we will decrease the error. A sequence of such moves (slowing as we near the bottom) will eventually find a minimum of some sort. The difficult part is to decide how large the steps should be.

Large steps may converge more quickly, but may also overstep the solution or (if the error surface is very eccentric) go off in the wrong direction. A classic example of this in neural network training is where the algorithm progresses very slowly along a steep, narrow, valley, bouncing from one side across to the other. In contrast, very small steps may go in the correct direction, but they also require a large number of iterations. In practice, the step size is proportional to the slope (so that the algorithms settle down in a minimum) and to a special constant: the learning rate. The correct setting for the learning rate is application-dependent, and is typically chosen by experiment; it may also be time-varying, getting smaller as the algorithm progresses. The algorithm therefore progresses iteratively, through a number of epochs. On each epoch, the training cases are each submitted in turn to the network, and target and actual outputs compared and the error calculated. This error, together with the error surface gradient, is used to adjust the weights, and then the process repeats. The initial network configuration is random and training stops when a given number of epochs elapse, or when the error reaches an acceptable level, or when the error stops improving. In the research, Back propagation Neural Network has been used to build a network that able to recognize male from female's samples based on the power amplitudes response of the DFT for the speech signals samples where the results shows that the power amplitude response for males' samples were higher than the amplitude response for females' samples. The steps to create a Neural Network based recognizer are [9]:

1. Specify the phonetic categories that the network will recognize .In the research it was the power amplitudes for the DFT for both males and females samples.
2. Find many samples of each of these categories in the speech data. In the research the best female and male samples depending on its power amplitudes were used. The smallest power value for the female sample assigned target [0 1 0 1] and the largest power value for the male sample assigned target as [0 0 1 1] .
3. Train a network to recognize females and males samples. In the research the Back Propagation Neural Network (BPNN) was trained for 462 epochs, with 50-hidden unit as a start but it was not enough number to satisfy the recognition process so the number of

hidden units was increased to 100-hidden unit to enhance the results. Fig (8) shows the trained neural network with the sitting parameters using MATLAB nntool environment. Fig (9) shows the training state of the Neural Network.

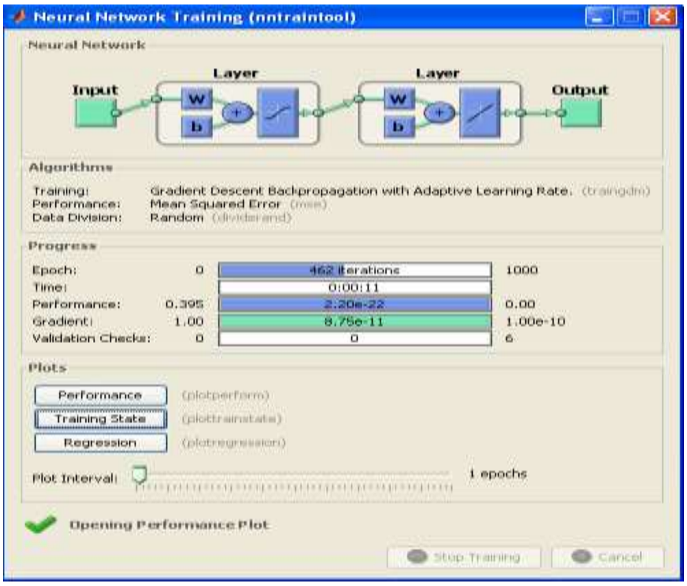

Fig (8) The Back propagation nntool Neural Network

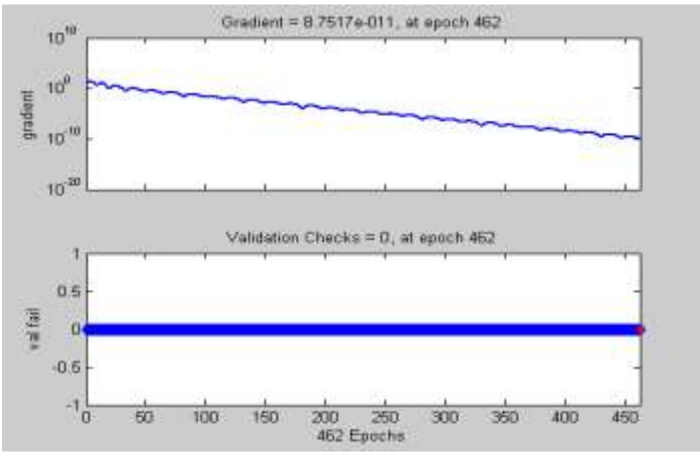

Fig (9) The Training State of the Back propagation Neural Network

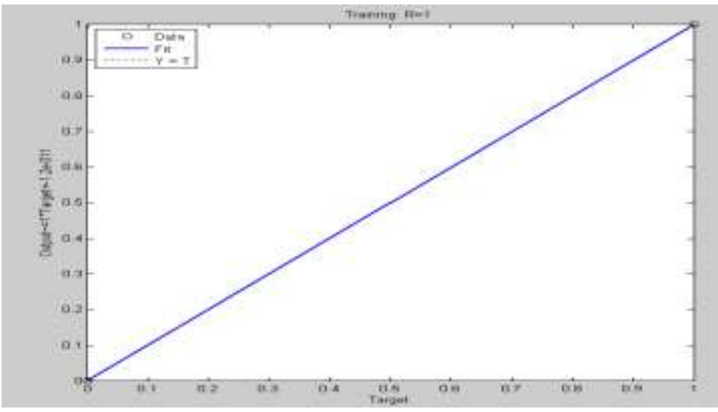

Fig (10) The Regression State of the Back propagation Neural Network

4. Evaluate the network performance using a test set. In the research the network was tested using 45 – sample for males' and females' samples after the training of the BPNN for the speech samples. Fig (11) shows the performance of the network.

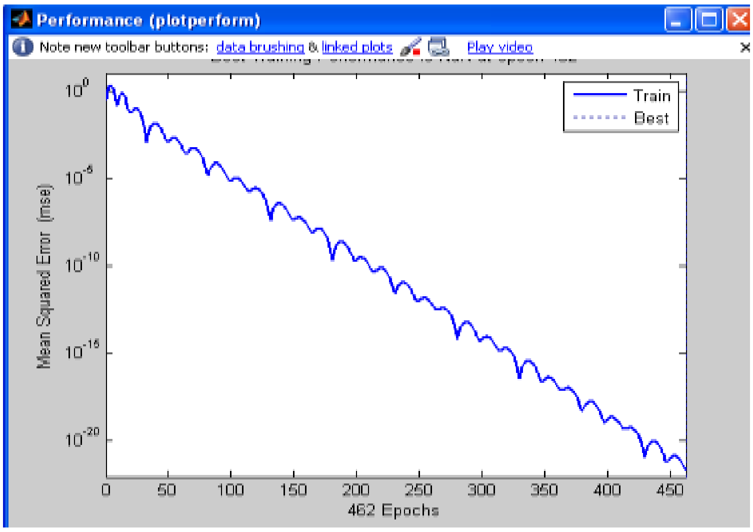

Fig (11) The Performance of the Back propagation Neural Network

5. Fuzzy Logic

Fuzzy set theory and fuzzy logic were conceived in 1965 by Lotfi Zadeh as a way of allowing uncertainly or vagueness to be represented mathematically .Fuzzy sets are a super –set of classical sets .Each element in a fuzzy set is associated with real number which represents the degree of membership of the element in the set. Fuzzy sets are usually expressed as a set of elements that having degree of membership for the truth values in closed unit interval [0, 1]. The idea behind a fuzzy set represents a concept and having a context is a further expanded by linguistic variables. A linguistic variable is assigned to a fuzzy region, a set of fuzzy sets that represent a complete concept. And it is also a super-set of classical logic that deals with prepositions which required being either true or false. Fuzzy logic allows highly nonlinear, poorly understood or mathematically complex systems to be modeled reliably and efficiently. And it deals well with noise data. These characteristics as in [10] suggested in the researches that fuzzy logic might be an effective tool for speech recognition. The fuzzy logic toolbox for use with MATLAB is a tool for solving problems. It helps to create and edit fuzzy inference systems by using graphical tools or command-line functioning [7]. In the research the first fuzzy system program was built to recognize between males and females samples based on two rules:

1- If the power amplitude of the speech signal is small value then female speaks

2- If the power amplitude of the speech signal is large value then male speaks

The two rules helped the system to recognize the speaker as female and as a male; the system is shown in fig (12) that was built using mamdani MATLAB toolbox7.6.0 (R2008a) and fig (13) shows the rule viewer of the system.

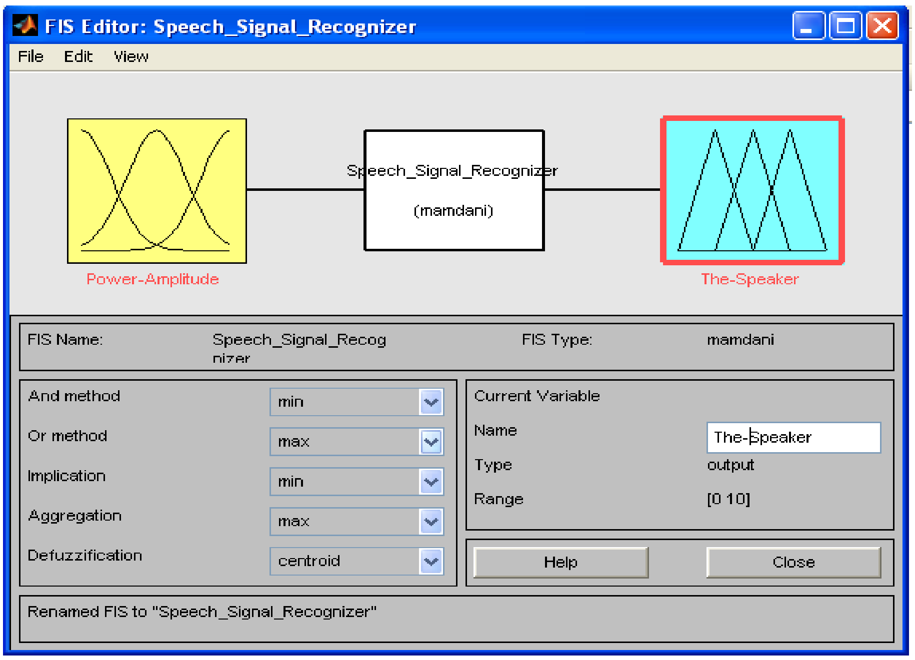

Fig (12) The Speech Signal Recognizer System

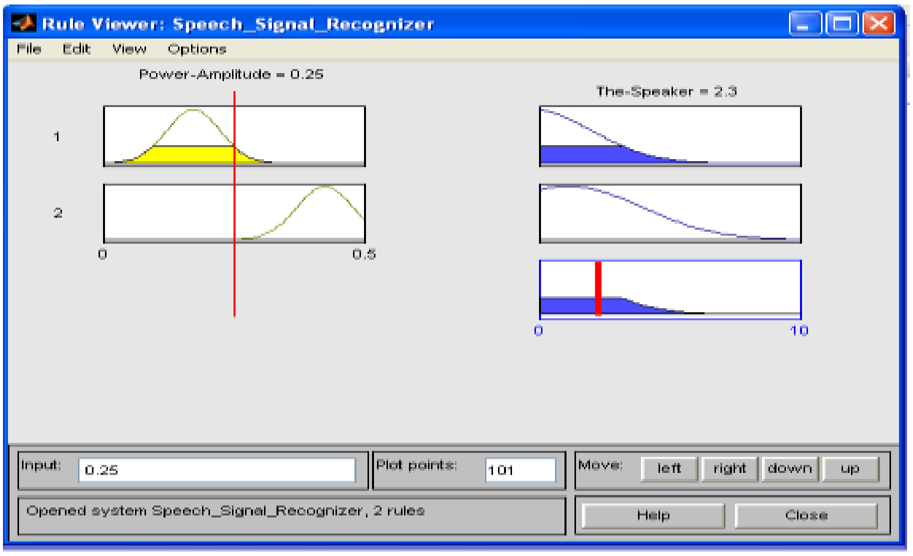

Fig (13) The Rule Viewer Speech Signal Recognizer System

Based on the extracted features, the power-amplitude, power spectrum and vowel-sound, the above system was developed to be able to recognize an individual speaker a female samples against other males and females samples. Testing the system through several cases and changing the rules .A system was built of 24- rules based on the values of the three extracted features, with 3-inputs represent the three extracted features for the speech samples, using the power amplitude of the DFT for the speech signal feature, The power Spectrogram of the speech signal feature and the signal energy feature that was concentrated in the harmonics of the vowel sound 'o' in the word,"Close". The system was able to retrieve the required speech signal for the required speaker. The system is shown in Fig (14) and the rule viewer is shown in Fig (15). While Fig (16) shows the 24- rules that used in the voice acceptance process.

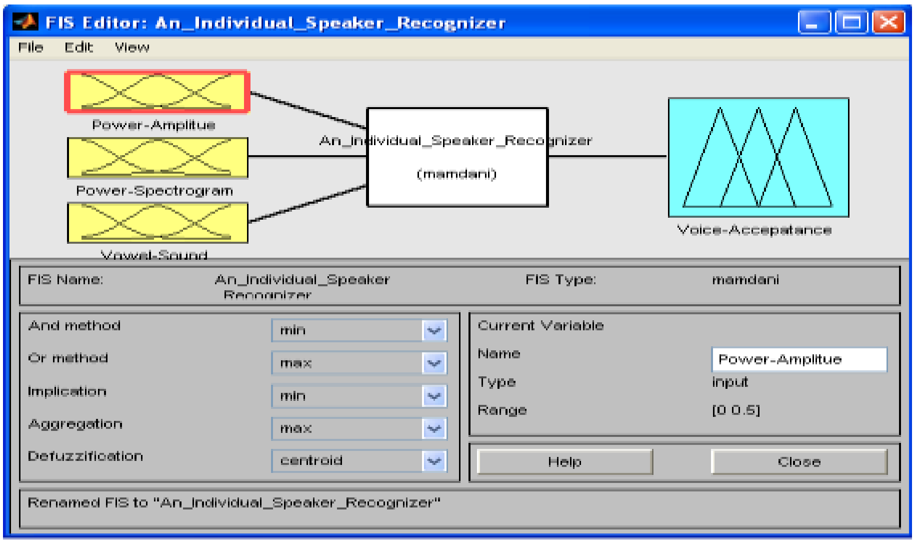

Fig (14): The Individual Speech Signal Recognizer

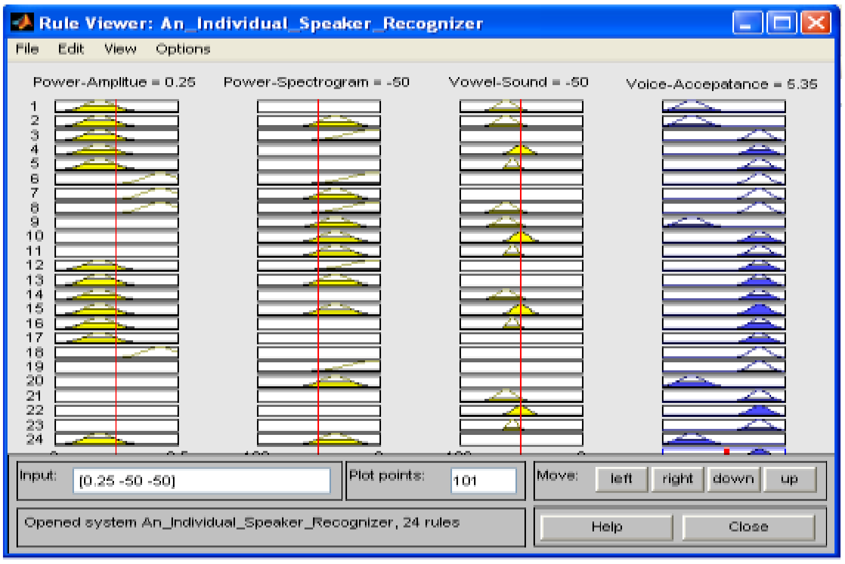

Fig (15) The Rule Viewer of the Individual Speech Signal Recognizer

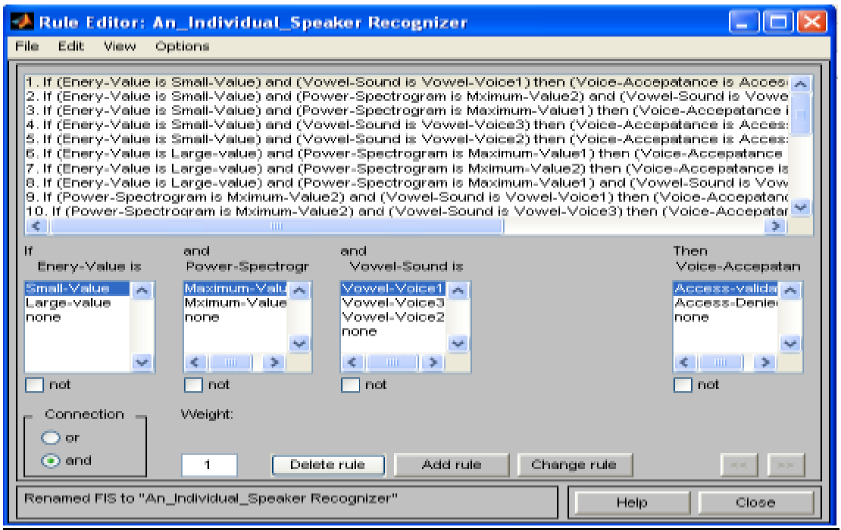

Fig (16) The Rule Editor of the Individual Speech Signal Recognizer

The Results

From collecting the speech samples for both male's and female's samples, the back propagation Neural Network (BPNN) was tested to recognize the speech signal, using 50-hidden units, The target was considered [0 0 1 1] for different males samples (M) and [0 1 0 1] for different females samples (F). The output of training [0 -1.5252e-011 1 1 ; 0 1 -3.0755e-011 1] , the results are shown in table (1).

Table (1): The Testing Results of the BPNN with 50-Hidden Units

Recorded Speech	Female Speech Output [0 1 0 1]	Recorded Speech	Male Speech Output [0 0 1 1]
F1	0 0.9668 0.7927 1	M1	0 -0.4829 3.0105 1
F2	0 1.8415 0.0727 1	M2	0 1.5231 0.2324 1
F3	0 1.0240 0.7514 1	M3	0 0.1036 0.9397 1
F4	0 1.9650 -0.0614 1	M4	0 1.7554 0.1526 1
F5	0 0.23 -3.2551 1	M5	0 0.1681 0.9723 1
F6	0 1.0343 0.1921 1	M6	0 -0.2679 1.6114 1
F7	0 1.4096 0.0081 1	M7	0 1.8479 0.0667 1
F8	0 -0.0137 1.0178 1	M8	0 0.5563 1.0399 1
F9	0 0.2674 1.0538 1	M9	0 -0.4514 3.0156 1
F10	0 -0.092 2.8621 1	M10	0 1.6042 0.2876 1
F11	0 0.4892 1.0666 1	M11	0 -0.0309 1.0415 1
F12	0 0.4693 1.0444 1	M12	0 -0.3433 1.8869 1
F13	0 0.7087 0.9610 1	M13	0 0.0847 0.9392 1
F14	0 -0.2339 2.5984 1	M14	0 1.0315 -0.0808 1
F15	0 -0.5156 2.6285 1	M15	0 -0.3244 1.8158 1

Table (2): The Testing Results of the BPNN with 100-Hidden Units

Recorded Speech	Female Speech Output [0 1 0 1]	Recorded Speech	Male Speech Output [0 0 1 1]
F1	0 0.1377 0.2221 1	M1	0 -0.4829 3.0105 1
F2	0 -0.1339 0.0883 1	M2	0 1.145 0.4904 1
F3	0 0.1695 0.1702 1	M3	0 -0.1462 1.0112 1
F4	0 -0.1741 0.0207 1	M4	0 -0.0167 0.0333 1
F5	0 1.1421 0.2155 1	M5	0 -0.2039 1.0091 1
F6	0 1.4022 0.3869 1	M6	0 0.3437 1.0299 1
F7	0 0.4446 -0.2962 1	M7	0 -0.1422 0.0913 1
F8	0 -0.2209 0.9423 1	M8	0 -0.0849 0.6085 1
F9	0 0.9996 -0.0001 1	M9	0 0.9182 0.2003 1
F10	0 0.3235 0.5692 1	M10	0 0.1735 -0.0632 1
F11	0 -0.1221 0.6829 1	M11	0 0.0473 0.9996 1
F12	0 0.5046 0.9067 1	M12	0 0.0407 1.015 1
F13	0 -0.0033 0.5 1	M13	0 -0.1239 1.0089 1
F14	0 0.2133 -0.2771 1	M14	0 1.2357 0.2661 1
F15	0 0.5603 0.7935 1	M15	0 0.3921 1.0214 1

The BPNN was trained using the female and male samples for specified number of epochs as mentioned before, the speech samples were taken from the energy values extracted feature in the speech processing operation, the target values was assigned as [0 0 1 1] for male samples and [0 1 0 1] for females samples, then the BPNN was tested against others females and males speech samples, in BPNN the results are approximately values as can seen in the above tables, the BPNN was trained and tested using 50-hidden units ,then with 100-hidden units . Table (3) shows the recognition success using BPNN with different numbers of hidden units. As can seen in table (3) as the number of hidden units increased, the performance of the BPNN was increased too , (True) samples refers to the number of samples that were tested and produced the required output ,while the (False) samples refers to the number of samples that were tested and did not produce the required results.

Table (3): The Recognition Success Rates using BPNN

Recognition Success (50-hidden units)		**Recognition Success (100-hidden units)**	
Female (True)	40%	**Female (True)**	33.3%
Female (False)	60%	**Female (False)**	66.7%
Male (True)	60%	**Male (True)**	80%
Male (False)	40%	**Male (False)**	20%

The fuzzy interface system was built using MATLAB7.6 (R2008) as a recognizer speech system for males and females samples. Two rules were used in the system to recognize the male from the female samples as mentioned above.

1- If the power amplitude of the speech signal is small value then female speaks. With range specified (0 – 0.5).
2- If the power amplitude of the speech signal is large value the male speaks. With range specified (0 – 0.5).

The output range for the speaker was (0 – 10), female speaker resulted in range (0 – 2.5), while male speaker resulted in range (2.6 and above). Like every system, there is an error value may result, the results are shown in table (4) , The small power amplitude values resulted for a females speaker , while the large power amplitude values resulted for males speaker , while table (5) shows the recognition success rate of the system.

Table (4): Speech Recognizer Fuzzy Interface System

Recorded Speech	Power Amplitude (Watt)	Fuzzy output Females Samples	Recorded Speech	Power Amplitude (Watt)	Fuzzy output Males Samples
F1	0.2826	2.27	M1	0.3368	4.07
F2	0.2647	2.07	M2	0.3209	3.47
F3	0.2816	2.26	M3	0.2673	2.1
F4	0.2573	2	M4	0.3126	3.06
F5	0.2056	1.72	M5	0.3935	2.86
F6	0.1132	1.83	M6	0.2645	2.09
F7	0.2337	1.84	M7	0.2911	2.41
F8	0.3033	4.01	M8	0.7108	3.74
F9	1.4565	2.72	M9	0.2709	2.13
F10	0.2929	3.74	M10	0.3420	3.94
F11	0.4648	2.45	M11	0.4163	2.68
F12	0.2875	2.97	M12	0.3235	3.61
F13	0.2424	2.35	M13	0.1728	1.69

F14	0.2426	1.89	M14	0.4102	2.7
F15	0.4980	3.69	M15	0.3262	3.77

Table (5): Speech Recognizer Fuzzy Interface System

Recognition Success (Females)		**Recognition Success (Males)**	
True %	80%	**True %**	73.3%
False %	20%	**False %**	26.7%

But this system was not enough to recognize an individual speaker to be tested against others speakers, females and males. 15- Samples were collected for an individual speaker all belong to one female file named (I) and tested against the other samples. The system was developed by using 24- rules based on the three extracted features as shown in Fig (16), 3-inputs ,the extracted features for each speech sample and one output represent the voice acceptance which based on two cases, First case produce," Access-Validate”, in range (0-5) for the required authority speaker to access a work area or security system ,which is the required speech sample (I) and second case to produce," Access- not Validate", in range (6-10) to give a rejection for non-required or authorized voice ,which is female speech sample (F) or male speech sample (M). The three inputs that represent the extracted features from the speech signal process are:

- Power Amplitude range (0 – 0.5) watt.
- Power Spectrogram range (-100 – 0) db
- Vowel Sound,’O’, (- 100 – 0) db.

Table (6) shows the results of the fuzzy logic interface system that was designed to recognize one authorized speaker only (I) against other females speakers (F) and males speakers (M), each speech samples passed through the speech signal processing operation for the features extraction ,power amplitude value produced from the power detection algorithm using DFT ,the power spectrum value produced from the power spectrogram in MATLAB processing block set and the vowel sound which resulted from reading the harmonics for the vowel sound, ’O’, in the word ‘Close’, the speech samples were recognized into females and males samples in the first fuzzy logic system using only two rules ,the power amplitude values ,then more rules were added to the fuzzy system ,using the three extracted features ,to make the system able to recognize only one specified speaker (I) , without the three grouped features ,it will be difficult for the system to recognize the required speakers, in BPNN this operation might be difficult and limited so fuzzy logic gives more flexibility in work . The using of the 24- rules produced a robust fuzzy logic system in recognizing an individual speaker. From the table, F1 represent a female sample with power amplitude (0.2826), power spectrum value (-4) and vowel sound value (-55), its voice acceptance is (6.05) ,it lays in the range of (6 - 10) ,which is “Access not – Validate”, this is because F1 is not the authorized speaker for the system , the same is true for sample M1 , while testing sample I1 , it’s power amplitude (0.1853) , power spectrum (-5) and it’s vowel sound value (-50) ,the voice acceptance result was (5.91) , according to the 24 - rules, it’s voice acceptance results in the range (0 – 5) ,which is “Access Validate” ,and from its result ,it is recognized as the required authorized speaker for the system . The others samples values are all shown in table (6). Table (7); show the success rate for recognizing the required speaker (I).

Table (6) Speech Recognizer Fuzzy Interface System for an Individual Speaker

Recorded Speech	Power Amplitude	Spectrogram Power	Vowel Sound	Voice Acceptance
F1	0.2826	-4 db	-55 db	6.05
F2	0.2647	-15 db	-70 db	5.35
F3	0.2816	-5 db	-48 db	6.38
F4	0.2573	-30 db	-60 db	4.85
F5	0.2056	-15 db	-70 db	5.16
F6	0.1132	-10 db	-50 db	5.26
F7	0.2337	-5 db	-55 db	5.07
F8	0.3033	-5 db	-60 db	7.56
F9	1.4565	-11 db	-61 db	6.68
F10	0.2929	-3 db	-60 db	7.75
F11	0.4648	-5 db	-70 db	6.36
F12	0.2875	-6 db	-46 db	7.54
F13	0.2424	-18 db	-61 db	5.84
F14	0.2426	-18 db	-60 db	5.13
F15	0.4980	-20 db	-61 db	5.58
M1	0.3368	-2 db	-60 db	7.52
M2	0.3209	-5 db	-60 db	7.17
M3	0.2673	-10 db	-69 db	5.63
M4	0.3126	-8 db	-60 db	6.95
M5	0.3935	-5 db	-61 db	7.63
M6	0.2645	-3 db	-44 db	7.29
M7	0.2911	-10 db	-56 db	6.31
M8	0.7108	0 db	-61 db	7.08
M9	0.2709	-1 db	-51 db	5.73
M10	0.3420	0 db	-62 db	6.47
M11	0.4163	-5 db	-66 db	7.86
M12	0.3235	-5 db	-64 db	7.24
M13	0.1728	-11 db	-50 db	4.81
M14	0.4102	-5 db	-55 db	7.63
M15	0.3262	0 db	-61 db	7.3
I1	0.1853	-5 db	-50 db	5.91
I2	0.2795	-40 db	-65 db	4.81
I3	0.2009	-40 db	-79 db	4.81
I4	0.3798	-25 db	-80 db	4.94
I5	0.4574	-45 db	-71 db	4.93
I6	0.3030	-12 db	-75 db	6.66
I7	0.2527	-25 db	-75 db	4.81
I8	0.2567	-40 db	-75 db	4.81
I9	0.1030	-35 db	-75 db	4.81
I10	0.1851	-50 db	-80 db	5.35
I11	0.1223	-44 db	-75 db	4.81
I12	0.1191	-38 db	-75 db	4.81
I13	0.1811	-40 db	-65 db	4.82
I14	0.1144	-30 db	-75 db	4.81
I15	0.1756	-40 db	-80 db	4.82

Table (7): Speech Recognizer Fuzzy Interface System for an Individual Speaker

Recognition Success (Individual Speaker)	
True %	91.1 %
False %	8.8 %

Discussion

Speech analysis is still a challenging area in recognition, classification and retrieval .many researches have done in speech analysis, .As seen in the results section, Fuzzy logic was proved to be a successful system in both recognition and retrieval operations of an individual speaker, compared with BPNN which suffered from several problems in giving a convincing recognition rates. Fuzzy logic is a flexible system where true and false decisions are taken based on a membership functions. The first system was used in recognizing between males and females samples and it shows a good success rate but to increase the recognition ability of this system ,more features from speech signal were required to be define ,so spectrogram analysis which used in speech analysis researches as in [4] ,[11] was very helpful analysis in defining two other robust features that define additional rules to the fuzzy interface system using MATLAB 7.6 (R2008a) tool and increases the system security and authority against unauthorized speakers that are not allowed to enter the system or to specified work area so as seen from table (6) ,the three features were the fuzzy interface system inputs and based on 24- rules a decision was made to give a voice acceptance to the speaker ,the system was named an," Individual Speaker Recognizer", as it was able to recognize a specified individual speaker and give it ,"Access- Validate", while give ,"Access- Denied ", to other unauthorized speakers whether they were males or females , the individual speaker in the research was taken form female's samples .The results of table (7) ,shows the recognition success rates of the system. Finally recent studies have shown that fuzzy systems and neural networks are both part of a class of universal approximations of continuous functions. This similarity means that a fuzzy system can be replaced by some form of neural network and vice versa. Fig (17) shows the model assumption for the recognizing and retrieving of speech signals samples .As a future work ,this fuzzy logic system can be developed by connecting it to a microcontroller device ,designing a complete recognizer system for an authorized speaker , performing orders ,like open system ,close system and so on.

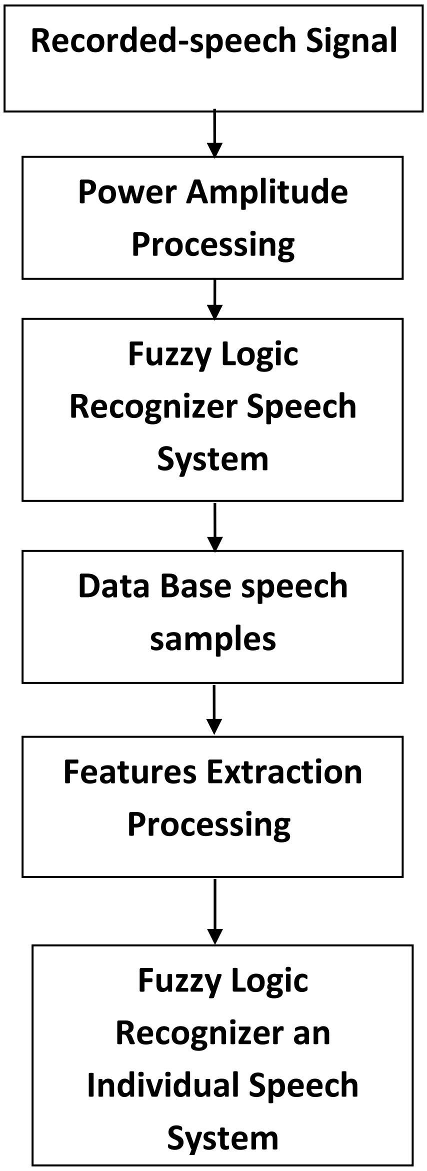

Fig (17) Model Assumption for Speech Recognition and Retrieval Operations

References

[1]. Chester ,F.J ,Taylor ,and M.Doyle ,"The Winger Distribution In Speech Processing Applications", J.Franklin Inst.,vol.318 ,pp. 415 – 420 ,1984.
[2]. Ho, K. Lai, and, Octavian, C.," Speech Processing Workshop", Department of Electrical and Electronic Engineering, Part IV Project Report 2003.
[3]. Jere, B. ," Digital Signal Processing Application using the ADSP-2100 Family ", Application Engineering staff of Analog Devices DSP Division Edited ,PTR ,Prentice Hall ,New Jersey ,p.330-p.331.
[4]. Brain, J. Love, and Jennifer, V. and, Xuening, S.,"Automatic Speaker Recognition Neural Networks", Electrical and Computer Engineering Department, the University of Texas at Austin, Spring 2004.
[5]. Abdul-Bary, R., S., Soad, T. A., "MATLAB-Based Design and Implementation of Time-Frequency Analyzer", College of Electrical Eng., Technical Institute.
[6]. M.Liu, C.Wan, L.Wang," Content-Based Audio Classification and Retrieval using a Fuzzy Logic System towards Multimedia Search Engines".
[7]. The Math Work, Inc. "MATLAB the Language of Technical Computing ", Version .7.6.0(R2008a), USA.
[8]. Christos, S. and, Dimitrios, S., "Neural Networks", http://www.emsl.pnl.gov:2080/docs/cie/neural/neural.homepage.html
[9]. John-Paul, H., and Ron, C. and Mark, F. and John , S. , and Yonghong ,Y. and Wei ,W. , "Training Neural Networks for Speech Recognition", Center for spoken language understanding ,(CSLU), Organ Graduate Institute of Science and Technology,January29,1998. http://speech.bme.ogi.edu/tutordemos/nnet_training/tutorial.html.
[10]. Patrick.M. Mills,"Fuzzy Speech Recognition", Bachelor of Science in Engineering, Swathmore College, 1994, University of South Carolina, Thesis 1996.
[11]. Dunia, J.,Jamma, "Low Cost Wavel et Analyzer", Computer Engineering Department, Technical College/ Mosul, Foundation of Technical College, Thesi , January /2005.

الملخص

تمييزصوت المتكلم هو واحد من أكثر المواضيع المثيرة للأهتمام في مجال معالجة الأشارة الرقمية حيث أنجزت العديد من البحوث في هذا المجال بأستخدم المواد والأدوات والبرامج العلمية المختلفة لتوليد ،معالجة ،تشفير وتمييز صوت المتكلم. في هذا البحث تم أستخدام عينات من أصوات الأناث و الذكور المتكلمين لبناء نظام قادرعلى التمييز بينهما بأستخدام الشبكات العصبية ونظام المنطق المضبب . تم أستخدام نظام المنطق المضبب بأعتماد ثلاث ميزات لصوت المتكلم وهي قيمة الطاقة للأشارة ، طيف الأشارة و صوت حرف العلة في الكلمة المذكورة ليكون قادر على التعرف على هوية الشخص وتمييز صوت متكلم محدد لشخص واحد دون أصوات المتكلمين الأخرين و تقنية النظام تعتمد على أمكانية أعطاء قبول وتخويل للشخص المتكلم بالدخول الى نظام أو منطقة عمل محددة ومنع الأشخاص غير المخولين من أختراق النظام أو منطقة العمل . خلال مرحلة الأختبار أعطى النظام نتائج جيدة بأستخدام عينات لصوت متكلم محدد مقارنة مع عينات أصوات متكلمين أخرين من أناث وذكور.