Best Wavelet Filter for a Wavelet Neural Fricatives Recognition System

Dr. Ahmed Maamoon Alkababji
Lecturer
Computer Engineering Department, Collage of Engineering,
University of Mosul, Mosul, Iraq.

Abstract

Direct recognition of phonemes in speaker independent speech recognition systems still cannot guarantee good enough recognition results. But grouping phonemes at first then trying to recognize the phoneme itself is a promising field. On the other hand wavelets are widely used in speech and speaker recognition systems, this is motivated by the ability of wavelet coefficients to capture important time and frequency features. In this work the effect of the wavelet filter type on the efficiency of a phoneme recognition system is investigated (specifically fricatives). The Probabilistic neural network was used as a pattern matching stage for its well known and power full ability in solving classification problems. It was found that the Daubechies wavelet family (generally from db15 to db23) is a good candidate for a fricatives phoneme recognition system that is based on wavelets as a feature extraction stage.

Keywords: Phoneme recognition, Fricatives, Wavelet, Probabilistic neural network.

مرشح التحويل المويجي الأفضل لنظام تمييز المقاطع الصوتية الاحتكاكية باعتماد التحويل المويجي والشبكات العصبية

د.احمد مأمون فاضل
مدرس
قسم هندسة الحاسبات, كلية الهندسة, جامعة الموصل, الموصل, العراق

الملخص

التمييز المباشر للمقطع الصوتي في أنظمة تمييز الكلام غير المعتمدة على الشخص لا تستطيع ضمان نسبة تمييز جيدة. لكن تقسيم المقاطع الصوتية إلى مجاميع (حسب النوع) ثم التمييز ضمن المجموعة كمرحلة لاحقة هو من المجالات الواعدة. من جهة أخرى فان التحويل المويجي له استخدامات واسعة في أنظمة تمييز المتكلم أو الكلام هذا بسبب قدرته العالية على استخلاص خصائص للزمن والتردد. في العمل الحالي تم دراسة تأثير نوع المرشح المويجي على أداء نظام تمييز للمقاطع الصوتية (الاحتكاكية بشكل خاص). تم استخدام الشبكة العصبية الاحتمالية كمرحلة مطابقة للهياكل وذلك لقدرتها العالية في حل مشاكل التصنيف .أظهرت النتائج أن المرشح من نوع دوبيجي (تحديدا من 15 إلى 23) هو من أفضل المرشحات للاستخدام في مرحلة استخلاص الخواص في أنظمة تمييز المقاطع الصوتية المبنية باستخدام التحويل المويجي.

Received: 5 – 1 - 2011　　　　Accepted: 4 – 7 - 2011

1. Introduction

Automatic speech recognition (ASR) is a process by which a machine identifies speech. The machine takes a human utterance as an input and returns a string of words , phrases or continuous speech in the form of text as output. As ASR technology matures, the range of possible applications increases. However, a domain and speaker independent system able to correctly decode all speech found in communication between people into strings of words is not realistic with the current state of technology[1].

Each speaker differs from others with individual voice tract characteristics. So the acoustical realization of the same word or utterance pronounced by different speakers could differ very much. Even the same speaker can‘t pronounce the same word or phrase identically several times. So phonemic speech recognition should confront with big variation of the same phoneme and this causes degradation in phoneme recognition accuracy[2].

Currently, the majority of speech recognition systems are based on template or pattern recognition principles and methods. The main idea of these methods is that at first we prepare templates of those phonemic units that we want to recognize and later they are compared with tested feature vectors to find the closest match during recognition stage. Phoneme recognition is a problem which aims to find the class of phoneme to whom belongs part of speech signal. The simplest algorithm for template based phoneme classification is to compare features describing part of speech signal with template parameters of each phoneme and after that to prescribe to the class of phoneme that is closest under some selected criteria. Such recognition requires relatively long time and the errors inside of similar phonemes and outside such group are different. These drawbacks could be partly lessened by hierarchical phoneme recognition structure. Here recognition process is divided into two steps:

1. It is identified dependence of analyzed speech signal to the one of the main groups of phonemes (vowel, semivowel, consonant, etc.).

2. Then recognition inside this group of phonemes is carried on to make a final decision.

Theory of phonetics interprets phonemes as tree type phonetic hierarchy where nodes of the tree represents phonemes and are grouped into the some groups (vowels, consonants, etc.) example of such tree (for American English) is presented in Figure 1[3].

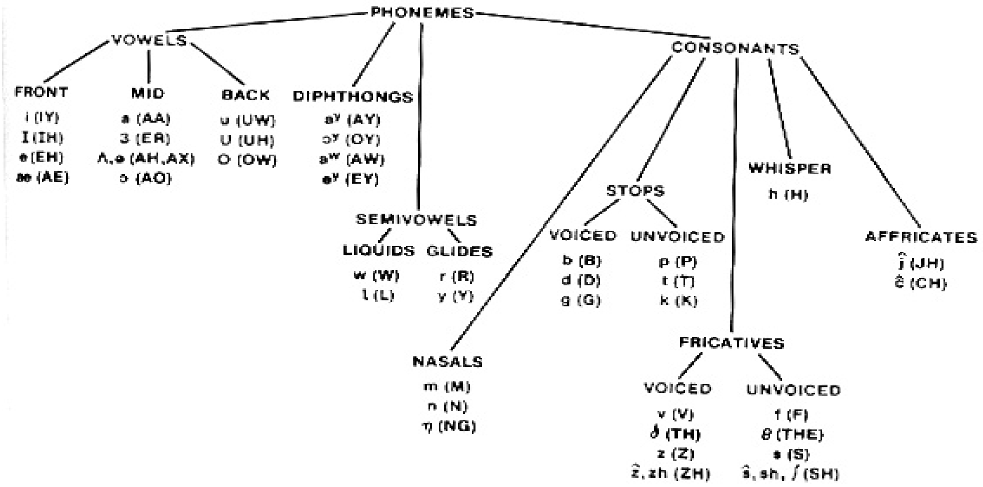


Figure 1: The phoneme classification[3].

In this work a phoneme recognition system is built based on wavelet as a feature extracting front end stage and a neural network for matching. The main aim of this work is to explore the influence of the type of the wavelet filter on the recognition of phonemes, hopefully to find the filter type that is best suitable for this application. Experiments were performed on the largest set of consonants in the English language which is fricatives.

2.Motivation and previous work

Although many speech processing tasks, like speech and speaker recognition, reached satisfactory performance levels on specific applications, many problems remain an open research area.

Koizumi T. and others [4] used a structural phoneme recognition system. Feature vectors were obtained by filtering short-term speech signal spectrum with 16 filters evenly spaced in the Bark scale. Classifier has been realized using Multilayered Neural Networks or RNN (Recurrent. Neural Networks). During experiments phonemes were brought into 6 groups – voiced and unvoiced plosive consonants, voiced and un-voiced fricative consonants, nasal consonants and vowels.

Abdelatty A. and others [5,6] implemented a structural consonant recognition system. In this system classification is based on the logical rules obtained from the analysis of such phonetic – acoustic properties as spectrum, magnitude, place of articulation, voiciness/unvoiciness and duration. Experiments were performed using TIMIT speech corpora. Phonemes were brought into such groups as plosive and fricative consonants, affricates. Further they were brought into voiced and unvoiced and even further into labials, palatals, alveolar, etc.

Juneja V. and Espy W. [7] performed experiments with the recordings from TIMIT corpora. In these experiments they compared performance of HMM based approach and hierarchical classification methods. Speech signal was classified into 5 classes: silence, vowels, sonorants, fricative and plosive con-sonants.

As could be seen, phoneme recognition based on the phonetic – acoustic knowledge is sufficiently applicable and perspective method that could allow achieving higher general speech recognition accuracy level.

Historically, LPC, LPCC, and MFCC speech features dominated the speech and speaker recognition areas in consequent periods. Other features like, PLP, ACW, wavelet-based features, did not gain widespread practical use, often due to their relatively more sophisticated computation. Nowadays many earlier computational limitations are overcome that opens possibilities for revaluation of the traditional solutions when speech features are selected for a specific task[8].

wavelet transform as a promising non-linear tool for signal analysis that has been used widely in phoneme recognition systems. The indications are that the Wavelet Transform and its variants are useful in speech recognition due to their good feature localization but furthermore because more accurate (non-linear) speech production models can be assumed. The analysis of the power in different frequency bands offers potential for the distinguishing of phonemes[9].

The main algorithm (wavelet) dates back to the work of Stephane Mallat in 1988. Since then, research on wavelets has become international. Wavelets and wavelet packets have been widely used in speaker ,speech and phoneme recognition this is seen in several past works as in [8],[10],[11]and[12].

In the conclusion of a comparative study in [13] between the use of wavelet and the traditional well known Mel Frequency Cepstral Coefficients (MFCC) it is mentioned that using wavelet may bring potential in automatic speech recognition.

Another comparative work [14] consider the following relatively less studied speech parameterization techniques: SBC of Sarikaya & Hansen, WPF of Farooq & Datta, WPSR of Siafarikas et al., OWPF of Siafarikas et al. and HFCC-E of Skowronsky & Harris. In addition, the well-known LFCC, MFCC and PLP, whose performance is well studied, were employed as reference points.

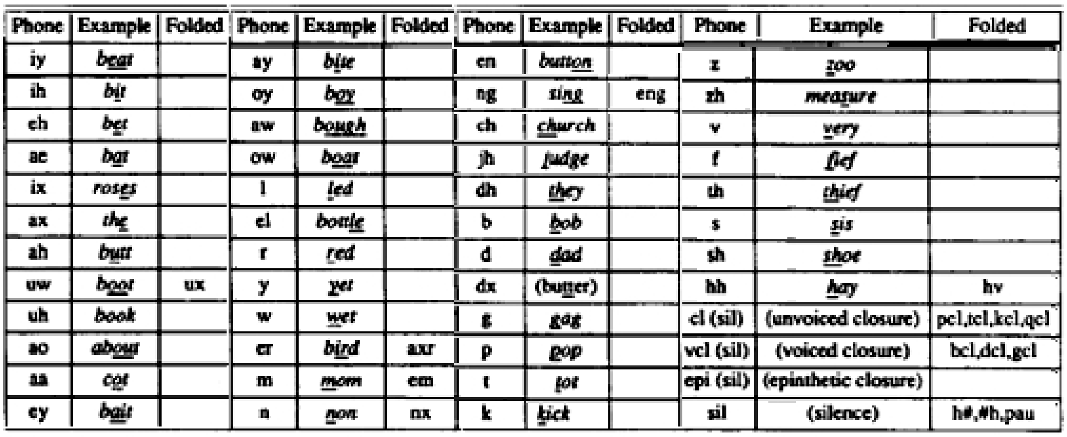
As a conclusion from all the above the filter-bank design is an open study point in the field of feature extraction as a front end of any speech/phoneme recognition system.

On the other hand the use of Artificial Neural Network (ANN) in general and specifically the Probabilistic Neural network as a decision, template matching or a classification stage is found in many past work convolving speaker, speech and phoneme recognition systems[1],[15],[16],[17] and [18].

3.Phoneme Classification

In this section the acoustic phonetic classification is discussed in general with a special concentration on fricatives. The recently developed TIMIT database [19] is ideal for evaluating phone recognizers. It consists of a total of 6300 sentences recorded from 630 speakers. Most of the sentences have been selected to achieve phonetic balance, and have been labeled at MIT. Lee K.& Hon H. [20] studied this data and labeled a total of 64 possible phonetic labels. From this set, 48 phones were selected. All “Q” (glottal stops) were removed from the labels. Also 15 allophones were identified, and folded them into the corresponding phones. Table 1 enumerates the list of 48 phones, along with examples, and the allophones folded into them. Among these 48 phones, there are seven groups where within-group confusions are not counted: {sil, cl, vcl, epi}, {el, l}, {en, n}, {sh, zh}, {ao, aa}, {ih, ix), {ah, ax}. Thus, there are effectively 39 phones that are in separate categories. This folding was performed to conform to CMU/MIT standards. It was found that folding closures together was necessary (and appropriate) for good performance, but folding the other categories only led to small improvements.

Table 1: List of phones used in phoneme recognition [20].



Phone	Example	Folded	Phone	Example	Folded	Phone	Example	Folded	Phone	Example	Folded
iy	*beat*		ay	*bite*		en	*button*		z	*zoo*	
ih	*bit*		oy	*boy*		ng	*sing*	eng	zh	*measure*	
eh	*bet*		aw	*bough*		ch	*church*		v	*very*	
ae	*bat*		ow	*boat*		jh	*judge*		f	*fief*	
ix	*roses*		l	*led*		dh	*they*		th	*thief*	
ax	*the*		el	*bottle*		b	*bob*		s	*sis*	
ah	*butt*		r	*red*		d	*dad*		sh	*shoe*	
uw	*boot*	ux	y	*yet*		dx	(butter)		hh	*hay*	hv
uh	*book*		w	*wet*		g	*gag*		cl (sil)	(unvoiced closure)	pcl,tcl,kcl,qcl
ao	*about*		er	*bird*	axr	p	*pop*		vcl (sil)	(voiced closure)	bcl,dcl,gcl
aa	*cot*		m	*mom*	em	t	*tot*		epi (sil)	(epinthetic closure)	
ey	*bait*		n	*non*	nx	k	*kick*		sil	(silence)	h#,#h,pau

The fricatives form the largest set of consonants in the English language which has nine standard fricative consonants, namely: the voiceless fricatives which include the labio-dental /f/ as in leaf, the linguo-dental /th/ as in teeth, the alveolar /s/ as in lease and the palatal /sh/ as in leash and their voiced cognates /v/ as in leave, /dh/ as in seethe, /z/ as in Lee’s and /zh/ as in azure. The ninth fricative is the /h/ which is considered also a semivowel. These consonants can be distinguished by English speaking listeners in identical phonetic contexts, regardless of whether these contexts are meaningful utterances or nonsense syllables. Therefore, the features needed for such discrimination can only reside in the acoustical signal[21].

4.Wavelet and Wavelet Packets

In the very most of ASR solutions, filter banks are used for parameterization of speech into acoustic features. Spectral analysis of the speech signal is the most appropriate method for extracting information from speech signals. DWT has been successfully used in many signal processing applications including speech for the spectral analysis of data.[8]

According to the multi-resolution theory, any wavelet ψ that generates an orthogonal basis of L^2(R) is characterized, by means of a filter bank construction, by a pair of discrete filters consisting of a high-pass (HPF) and a low-pass one (LPF) followed by sub-sampling by two to reduce redundancy. These filters belong to a particular class of filters, called *conjugate mirror filters,* cascading these filters produces a fast discrete wavelet transform.

Wavelet packet functions generalize the filter bank tree that relates wavelets and conjugate mirror filters. In the decomposition with the wavelet packet transform, the lower, as well as the higher frequency bands are decomposed giving a balanced binary tree structure. Such a tree is illustrated in Figure 2. To each node in the tree, a wavelet packet space $W^p{}_j$ is associated, where j is the depth, and p is the number of the nodes to the left of this particular node at the same depth.

Figure 2 illustrates 8 wavelet packets $W^p{}_j$ at the depth j=3[8].

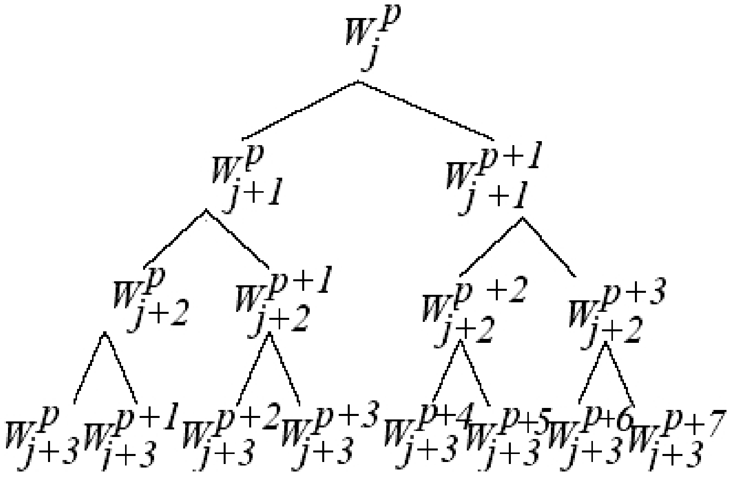


Figure 2: Binary tree of wavelet packet spaces

5. The Probabilistic Neural Network

Artificial neural networks ("ANN") are adaptive models with a network-like structure consisting of a large number of processing units, called neurons.

In the present work a special type of neural network is used called Probabilistic Neural Network (PNN) (see [22] for details) .

The use of the probabilistic neural network (PNN)in this work is motivated by its well known power full classification characteristics. So it is used in this work to classify the input phoneme segment (after extracting its features).

Figure 3 shows the architecture of the probabilistic neural network used in this work.

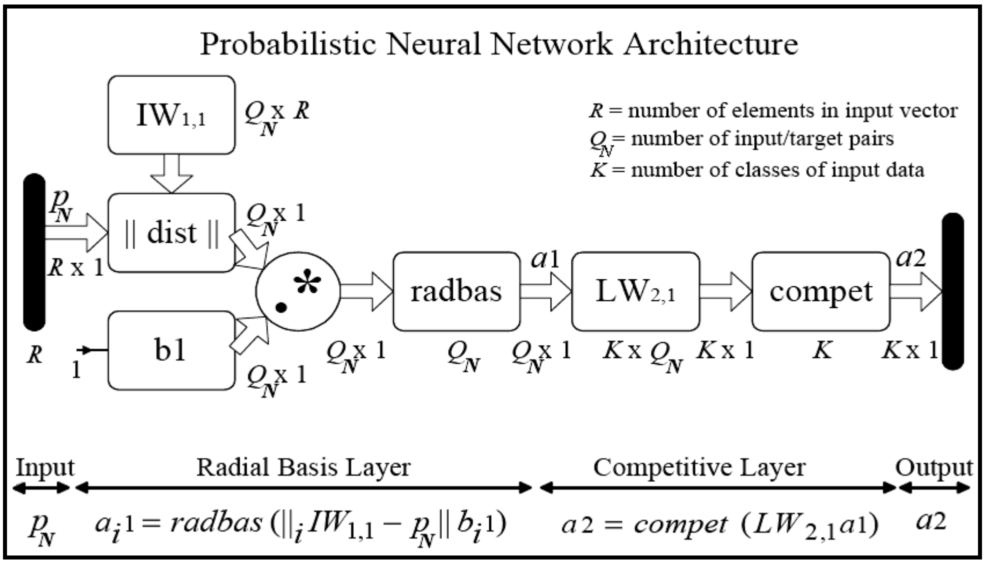

Probabilistic Neural Network Architecture
IW1,1
Q_N x R
R = number of elements in input vector
Q_N = number of input/target pairs
K = number of classes of input data
p_N
R x 1
|| dist ||
Q_N x 1
.*
radbas
a1
LW2,1
compet
a2
R
1
b1
Q_N x 1
Q_N x 1
Q_N
Q_N x 1
K x Q_N
K x 1
K
K x 1
Input
Radial Basis Layer
Competitive Layer
Output
p_N
a_i1 = radbas (||_i IW1,1 − p_N || b_i1)
a2 = compet (LW2,1 a1)
a2

Figure 3: Architecture of the Probabilistic Neural Network

6. Speech Corpus

The speech corpus used to find the best type of wavelet filter in the proposed phoneme recognition system is the standard American English TIMIT provided by Linguistic Data Consortium [19]. TIMIT is an acoustic-phonetic database including 6300 sentences and 630 speakers who speak English. The audio format is PCM, the audio samples are quantized in 16 bit, the recordings are single-channel, the mean duration is 3.28 sec and the standard deviation (st. dev.) is 1.52sec. From all the available data in the TIMIT corpus two arbitrary subsets of speakers are used in this work. The male speaker's subset contained 70 speakers and the female speaker's subset contained 70 speakers too. There are 10 speech files for each speaker; two of the files have the same linguistic content for all speakers, whereas the remaining eight files are phonetically diverse.

For the evaluation of the proposed system 10 speakers were selected arbitrary from the TIMIT corpus, six of them were used for training and the other four for testing. First phonemes were extracted from each speech file and grouped according to its type, as mentioned earlier in this work we are interested in fricatives(/f/,/th/,/s/,/sh/,/v/,/dh/ and/z/), according to [20] /zh/ is grouped with /sh/ so it was not include in this work.

7.System Architecture for Phoneme Recognition

The proposed system has two main stages (as any recognition system). But in this work (differing from any known phoneme recognition system) Firstly a preprocessing and feature extracting stage which is the wavelet packet. Followed by the classification stage and that is the Probabilistic Neural Network (PNN). Next is the procedure used to extract the features of the fricative phonemes, train the neural network and finally test the system.

The feature extraction as a procedure is the same for the training phase and for the testing phase. Each phoneme file is applied to a wavelet packet tree of a depth of seven (*j=7*) this provide a total of 128 frequency sub bands. Due to the compact support of wavelet, no Hamming window or other window is required and there is a single output from the wavelet tree every 8 msec. This is because the down sampling by two at every stage in the wavelet packet. The frequency resolution in accordance of the wavelet tree is 125Hz (16000Hz, which the sampling frequency of the input speech signal, divided by 2^7), see Figure 4.

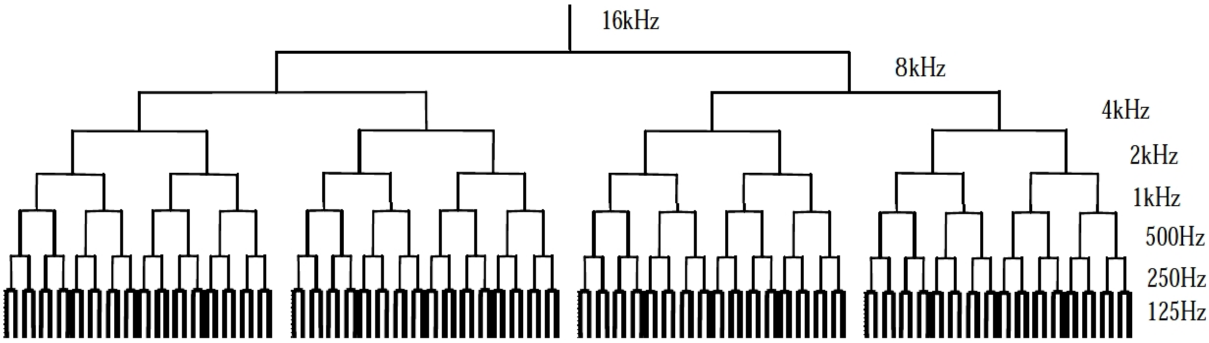


Figure 4: The proposed wavelet packet tree.

Next, the normalized energy vector for all the frequency band is computed:

$$E(p)=\frac{\left[W_{j+7}^{p}f(i)\right]^{2}}{\sum_{p=1}^{128}\left[W_{j+7}^{p}f(i)\right]^{2}},\ p=1,2,\ldots,128 \qquad (1)$$

where $W^{p}_{j+7}f\ (i)$ is the i-th coefficient of the wavelet packet transform of a signal f at node W^{p}_{j+7} of the wavelet packet.

As a result, a matrix of 128 rows by N columns is obtained for each phoneme, where N depends on the duration of the phoneme file (N= duration in seconds /8msec.). Each vector (column) of this matrix is a feature vector representing this phoneme. This N vectors feature matrix has a redundancy in it. This redundancy is removed using clustering processes. The clustering processes can be performed using any clustering algorithm. However, the most popular and the simplest clustering algorithm (the generalized Lloyd algorithm) (GLA) is used. The algorithm is also known as Linde-Buzo-Gray algorithm (LBG) according to its inventors or the K-mean clustering algorithm. The K-mean clustering algorithm reduces the size of this matrix to (128*32). This overall processes is repeated seven times for all the phonemes (/f/,/th/,/s/,/sh/,/v/,/dh/ and/z/). The clustering algorithm is used in the training phase only. At this point seven matrices (one for each phoneme) are obtained. These matrices are then concatenated to form one matrix which is used to train the PNN which has an input of 128 nodes and an output of 7 which is the number of classes of phonemes (Fricatives) used. At this point the training process is completed.

In the testing phase the phoneme speech file (7 files, one for each phoneme /f/,/th/,/s/,/sh/,/v/,/dh/ and/z/) is passed through the same stages mentioned above but not the clustering stage to extract the features. After the feature matrix is obtained it is entered to the neural network which produce an output for each vector (column) in this matrix. Then the recognition rate is found by dividing the correct recognitions by the total number of input vectors. Figure 5 illustrate the architecture of the proposed system for both the training phase and the testing phase.

To this point the system is completed. But this system is designed for a particular type of wavelet filter which was used in building the wavelet packet tree. Keeping in mind that the main purpose of this work is to find the best wavelet filter to be used in phoneme recognition systems the all above procedure is repeated for all the filters under examination.

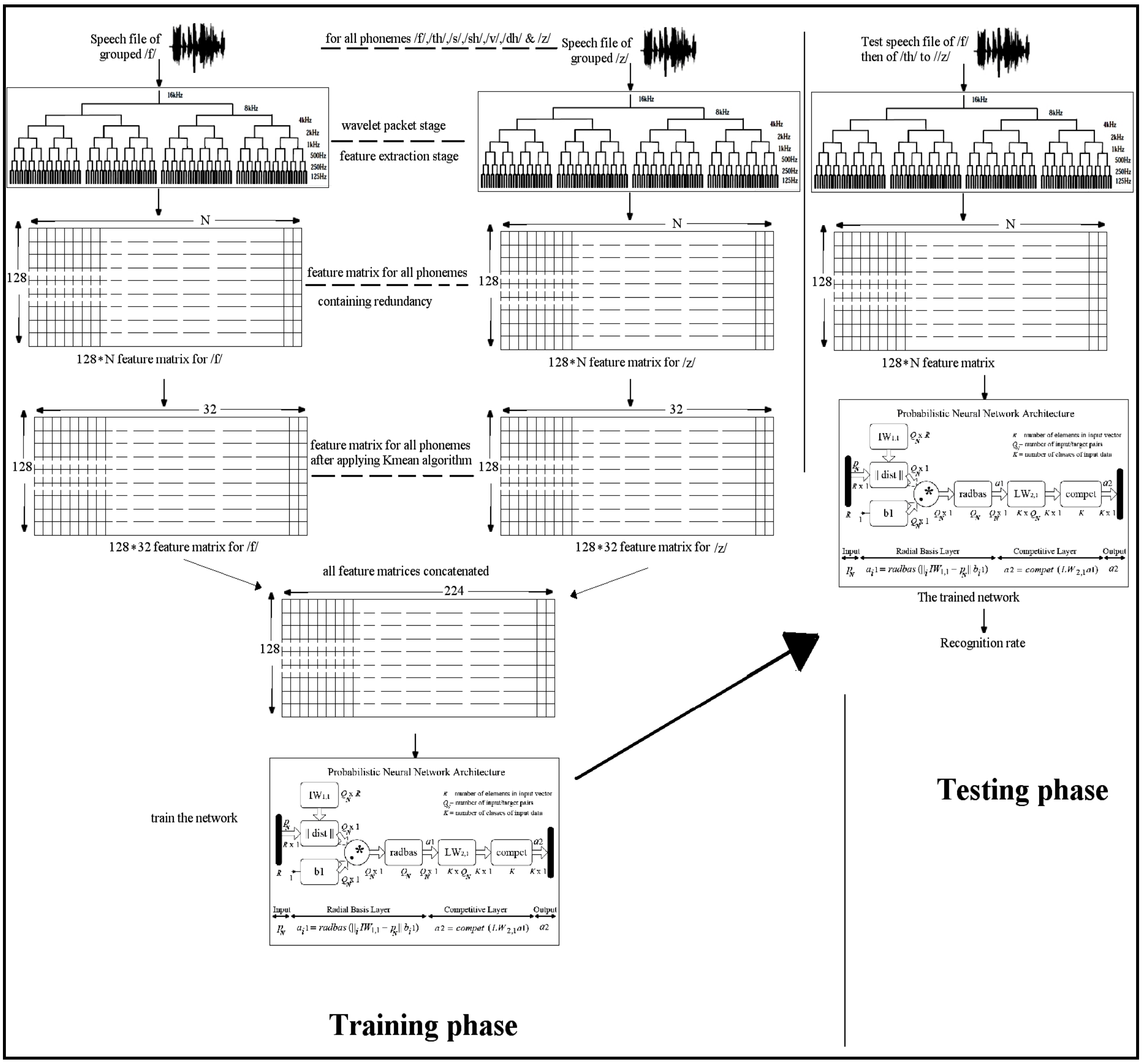

Speech file of grouped /f/
for all phonemes /f/,/th/,/s/,/sh/,/v/,/dh/ & /z/
Speech file of grouped /z/
Test speech file of /f/ then of /th/ to //z/
16kHz
8kHz
4kHz
2kHz
1kHz
500Hz
250Hz
125Hz
wavelet packet stage
feature extraction stage
N
128
feature matrix for all phonemes
containing redundancy
128*N feature matrix for /f/
128*N feature matrix for /z/
128*N feature matrix
32
feature matrix for all phonemes
after applying Kmean algorithm
128*32 feature matrix for /f/
128*32 feature matrix for /z/
all feature matrices concatenated
224
Probabilistic Neural Network Architecture
IW1,1
|| dist ||
b1
radbas
LW2,1
compet
Input
Radial Basis Layer
Competitive Layer
Output
train the network
The trained network
Recognition rate
Testing phase
Training phase

Figure 5: Architecture of the system used for phoneme recognition (training phase and testing phase)

8.Experiments and Results

After training the PNN, the network is tested with the same training data to check the system. It was found that the recognition rates were between 99.11% and 70.54% as shown in Table 2. This procedure is done for every type of wavelet filter, as a result, the training and testing phase is repeated for 85 times. The types of the wavelet filters that were examined are: Daubechies 1-25,27,31,35,40 and 45, Coiflet 1-5, Symlet 2-15,17,19,21,23 and 27, Discrete Meyer, Biorthogonal 1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5 and 6.8, Reverse Biothogonal 1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5 and 6.8.

Table2: The recognition rates for the same training data

Filter Type	Recognition Rate for z (%)	Recognition Rate for v (%)	Recognition Rate for th (%)	Recognition Rate for sh (%)	Recognition Rate for s (%)	Recognition Rate for f (%)	Recognition Rate for dh (%)	**Total Recognition Rate (%)**
bior35	100.00	100.00	100.00	96.88	96.88	100.00	100.00	**99.11**
db23	100.00	100.00	100.00	96.88	96.88	100.00	96.88	**98.66**
db35	100.00	100.00	100.00	100.00	100.00	96.88	93.75	**98.66**
db13	93.75	96.88	100.00	100.00	100.00	100.00	96.88	**98.21**
bior37	100.00	100.00	100.00	93.75	93.75	100.00	100.00	**98.21**
bior39	100.00	100.00	96.88	100.00	90.63	100.00	100.00	**98.21**
db15	96.88	100.00	100.00	96.88	100.00	96.88	93.75	**97.77**
db17	93.75	100.00	100.00	100.00	100.00	100.00	90.63	**97.77**
db19	93.75	100.00	100.00	100.00	100.00	100.00	90.63	**97.77**
db9	93.75	100.00	96.88	100.00	100.00	93.75	96.88	**97.32**
db27	96.88	100.00	100.00	100.00	100.00	93.75	90.63	**97.32**
db14	87.50	100.00	100.00	96.88	100.00	100.00	93.75	**96.88**
db21	93.75	100.00	100.00	100.00	100.00	96.88	87.50	**96.88**
db25	96.88	100.00	100.00	100.00	100.00	93.75	87.50	**96.88**
db31	96.88	100.00	100.00	100.00	100.00	90.63	90.63	**96.88**
coif5	90.63	100.00	100.00	100.00	96.88	93.75	96.88	**96.88**
db10	93.75	100.00	96.88	96.88	100.00	93.75	93.75	**96.43**
db12	87.50	100.00	96.88	100.00	96.88	100.00	93.75	**96.43**
db40	90.63	100.00	100.00	100.00	100.00	90.63	93.75	**96.43**
db45	93.75	100.00	100.00	100.00	100.00	90.63	90.63	**96.43**
bior33	96.88	100.00	100.00	93.75	90.63	100.00	93.75	**96.43**
db7	90.63	100.00	93.75	96.88	100.00	96.88	90.63	**95.54**
db11	87.50	100.00	96.88	96.88	100.00	93.75	93.75	**95.54**
db6	90.63	100.00	87.50	96.88	100.00	100.00	90.63	**95.09**
sym17	96.88	100.00	93.75	93.75	96.88	100.00	84.38	**95.09**
rbio37	90.63	100.00	93.75	87.50	100.00	96.88	96.88	**95.09**
sym7	93.75	100.00	90.63	93.75	96.88	100.00	87.50	**94.64**
sym10	84.38	100.00	93.75	93.75	100.00	100.00	90.63	**94.64**
sym19	84.38	100.00	90.63	93.75	96.88	100.00	96.88	**94.64**
bior24	93.75	100.00	93.75	93.75	93.75	96.88	90.63	**94.64**
rbio26	93.75	100.00	93.75	93.75	90.63	96.88	93.75	**94.64**
db5	81.25	100.00	93.75	96.88	100.00	100.00	87.50	**94.20**
db8	90.63	100.00	93.75	93.75	96.88	96.88	87.50	**94.20**
coif3	90.63	100.00	84.38	93.75	100.00	100.00	90.63	**94.20**
sym14	90.63	100.00	90.63	93.75	96.88	96.88	87.50	**93.75**
rbio33	90.63	100.00	96.88	96.88	90.63	87.50	93.75	**93.75**
sym15	90.63	100.00	87.50	93.75	96.88	93.75	90.63	**93.30**
sym23	87.50	100.00	90.63	93.75	93.75	100.00	87.50	**93.30**
dmey	81.25	100.00	96.88	90.63	96.88	100.00	87.50	**93.30**
bior28	84.38	100.00	96.88	96.88	87.50	100.00	87.50	**93.30**
rbio28	90.63	100.00	90.63	84.38	93.75	100.00	93.75	**93.30**
rbio35	100.00	100.00	93.75	87.50	84.38	93.75	93.75	**93.30**
rbio39	87.50	100.00	96.88	87.50	84.38	100.00	96.88	**93.30**
coif4	84.38	100.00	87.50	96.88	96.88	96.88	87.50	**92.86**
sym11	78.13	100.00	90.63	93.75	100.00	100.00	87.50	**92.86**
sym27	87.50	100.00	87.50	90.63	93.75	100.00	90.63	**92.86**
rbio55	93.75	100.00	87.50	93.75	93.75	100.00	81.25	**92.86**
db4	75.00	100.00	87.50	96.88	96.88	100.00	90.63	**92.41**
sym21	81.25	100.00	84.38	96.88	96.88	100.00	87.50	**92.41**
rbio24	87.50	100.00	90.63	100.00	84.38	96.88	87.50	**92.41**
rbio31	87.50	100.00	96.88	93.75	87.50	90.63	90.63	**92.41**
sym9	78.13	100.00	87.50	93.75	100.00	100.00	84.38	**91.96**
sym8	84.38	100.00	87.50	90.63	100.00	100.00	78.13	**91.52**
sym12	84.38	100.00	87.50	90.63	96.88	93.75	87.50	**91.52**
sym13	93.75	100.00	81.25	90.63	93.75	96.88	81.25	**91.07**
bior31	78.13	100.00	96.88	90.63	93.75	84.38	93.75	**91.07**
bior44	84.38	100.00	84.38	93.75	93.75	100.00	81.25	**91.07**
bior68	75.00	100.00	87.50	96.88	100.00	93.75	84.38	**91.07**
coif2	81.25	100.00	78.13	93.75	93.75	100.00	87.50	**90.63**
bior26	93.75	100.00	93.75	87.50	75.00	100.00	84.38	**90.63**
bior55	90.63	100.00	93.75	84.38	87.50	96.88	81.25	**90.63**
rbio22	84.38	100.00	90.63	96.88	84.38	96.88	81.25	**90.63**
rbio44	87.50	100.00	78.13	93.75	90.63	100.00	84.38	**90.63**
db3	78.13	100.00	75.00	96.88	96.88	100.00	84.38	**90.18**
sym5	75.00	100.00	81.25	96.88	100.00	100.00	75.00	**89.73**
rbio68	78.13	100.00	90.63	87.50	93.75	96.88	81.25	**89.73**
sym3	81.25	100.00	78.13	90.63	93.75	100.00	81.25	**89.29**
sym6	81.25	100.00	75.00	90.63	96.88	100.00	81.25	**89.29**
sym4	68.75	100.00	81.25	93.75	93.75	100.00	75.00	**87.50**
sym2	59.38	100.00	71.88	96.88	93.75	100.00	78.13	**85.71**
coif1	71.88	100.00	65.63	96.88	90.63	93.75	78.13	**85.27**
rbio15	65.63	100.00	62.50	93.75	100.00	96.88	71.88	**84.38**
db2	68.75	100.00	62.50	90.63	93.75	100.00	65.63	**83.04**
bior13	68.75	100.00	59.38	90.63	90.63	100.00	71.88	**83.04**
bior15	56.25	100.00	62.50	93.75	90.63	100.00	75.00	**82.59**
rbio13	68.75	100.00	53.13	87.50	87.50	100.00	75.00	**81.70**
db18	62.50	71.88	87.50	84.38	87.50	75.00	100.00	**81.25**
db16	46.88	84.38	87.50	87.50	78.13	81.25	96.88	**80.36**
bior22	62.50	68.75	81.25	87.50	81.25	81.25	87.50	**78.57**
db24	59.38	75.00	78.13	87.50	81.25	71.88	93.75	**78.13**
db22	43.75	78.13	81.25	87.50	78.13	84.38	90.63	**77.68**
db20	46.88	62.50	78.13	96.88	71.88	84.38	96.88	**76.79**
rbio11	53.13	100.00	43.75	93.75	71.88	96.88	53.13	**73.21**
bior11	50.00	100.00	43.75	87.50	75.00	100.00	50.00	**72.32**
db1(HAAR)	37.50	100.00	46.88	90.63	75.00	96.88	46.88	**70.54**

Table3: The recognition rates for the testing phase.

Filter Type	Recognition Rate for z (%)	Recognition Rate for v (%)	Recognition Rate for th (%)	Recognition Rate for sh (%)	Recognition Rate for s (%)	Recognition Rate for f (%)	Recognition Rate for dh (%)	**Total Recognition Rate(%)**
db21	18.04	45.26	70.27	76.03	74.70	37.98	37.14	**51.35**
db23	46.68	37.59	61.74	67.09	51.98	57.63	36.49	**51.31**
db22	21.79	43.88	76.11	68.97	75.99	37.31	33.33	**51.06**
db18	28.27	38.93	46.67	74.35	58.14	59.13	50.00	**50.78**
db15	31.38	43.20	35.35	62.77	61.77	69.11	50.00	**50.51**
db16	31.38	43.20	35.35	62.77	61.77	69.11	50.00	**50.51**
db17	26.58	43.41	48.54	74.21	71.18	50.80	38.71	**50.49**
db19	22.40	40.60	45.79	66.41	68.59	46.85	59.09	**49.96**
db11	11.68	41.03	31.52	70.19	78.82	72.69	42.00	**49.0**
db14	14.97	25.20	41.24	70.59	80.28	61.89	51.79	**49.42**
db10	16.39	39.13	12.22	71.39	73.15	71.61	56.25	**48.59**
db20	23.32	37.78	41.28	68.13	65.22	59.77	42.65	**48.31**
coif5	22.87	37.60	9.09	76.33	64.60	68.29	58.62	**48.20**
db13	28.76	23.97	34.38	72.31	62.57	63.22	51.85	**48.15**
db24	20.81	35.66	74.36	77.66	74.44	24.24	28.95	**48.02**
db5	9.27	28.57	11.25	70.59	77.98	82.30	52.63	**47.51**
db9	17.58	42.48	14.77	66.58	75.59	68.38	45.65	**47.29**
db12	19.46	32.77	20.21	70.62	69.23	72.08	46.15	**47.22**
db6	19.27	30.84	15.85	71.31	69.47	80.70	42.50	**47.14**
db25	31.06	30.34	64.71	78.54	60.34	27.07	32.05	**46.30**
db8	13.26	41.44	10.47	61.71	69.33	76.72	50.00	**46.13**
db40	26.76	33.33	55.70	75.35	66.99	27.12	35.19	**45.78**
db27	18.00	33.56	72.36	66.00	74.87	28.15	26.83	**45.68**
coif3	20.05	34.51	4.55	57.26	72.69	80.34	50.00	**45.63**
db31	21.08	30.13	64.12	75.00	73.87	20.14	34.44	**45.54**
db45	19.72	35.87	53.46	70.64	71.04	24.59	40.68	**45.14**
sym19	9.64	30.83	13.08	66.67	71.55	77.17	45.45	**44.91**
db7	16.94	31.19	9.52	68.42	71.77	72.17	42.86	**44.70**
coif4	13.78	40.34	3.19	56.33	70.13	77.08	51.92	**44.68**
sym15	18.09	24.80	8.08	53.99	70.09	77.24	55.17	**43.92**
db4	11.58	23.30	5.13	74.08	70.77	83.04	38.89	**43.83**
sym17	31.32	31.78	13.59	50.00	49.91	82.40	46.77	**43.68**
sym14	20.59	26.02	8.25	54.55	57.19	83.20	53.57	**43.34**
sym7	18.89	40.37	4.76	52.91	65.03	83.04	38.10	**43.30**
bior28	18.68	30.97	12.50	67.67	53.53	73.08	45.65	**43.15**
sym11	13.32	32.48	7.61	57.99	69.84	76.89	42.00	**42.88**
sym4	12.71	30.10	6.41	67.32	69.49	83.04	30.56	**42.80**
bior26	33.06	31.19	10.71	65.10	32.24	79.13	47.62	**42.72**
sym12	15.14	25.21	7.45	61.19	66.55	82.08	40.38	**42.57**
db35	38.22	19.51	61.87	71.15	52.89	21.33	32.65	**42.52**
coif2	17.04	28.04	6.10	61.84	58.68	87.28	37.50	**42.35**
sym13	28.23	23.97	7.29	44.35	53.12	85.54	53.70	**42.31**
rbio44	21.63	26.67	3.75	66.39	49.72	83.63	42.11	**41.98**
sym3	15.91	20.79	6.58	59.21	66.05	86.04	35.29	**41.41**
bior68	12.09	28.32	4.55	56.71	73.60	77.35	36.96	**41.37**
sym10	19.67	28.70	3.33	59.67	64.14	81.36	31.25	**41.16**
Dmey	9.62	23.98	20.47	45.09	69.54	72.87	46.51	**41.15**
sym5	16.29	24.76	8.75	57.14	69.72	77.88	31.58	**40.88**
sym23	12.50	24.82	16.52	57.91	66.95	74.05	32.43	**40.74**
rbio24	29.49	31.43	2.50	62.18	44.04	67.26	47.37	**40.61**
rbio55	32.40	30.84	10.98	59.33	35.47	85.09	30.00	**40.59**
coif1	12.50	22.77	11.84	66.29	68.82	80.63	20.59	**40.49**
db3	13.35	25.74	6.58	56.09	67.34	86.94	26.47	**40.36**
sym21	18.30	29.20	5.41	50.26	66.38	72.87	40.00	**40.34**
rbio15	12.64	20.00	2.50	55.74	73.03	76.11	42.11	**40.30**
sym2	6.00	20.20	2.70	66.67	71.30	85.91	28.13	**40.13**
sym8	25.69	26.13	4.65	55.92	62.43	76.29	29.55	**40.09**
sym6	9.78	25.23	7.32	56.55	72.21	83.77	25.00	**39.98**
rbio13	15.34	27.72	2.63	70.82	60.15	82.43	20.59	**39.96**
sym9	7.97	26.55	7.95	47.95	77.03	80.77	30.43	**39.81**
bior11	8.33	28.87	4.17	69.91	47.77	89.45	30.00	**39.79**
bior44	12.92	30.48	3.75	63.87	66.24	73.89	26.32	**39.64**
bior15	6.46	20.95	3.75	64.99	66.24	85.84	28.95	**39.60**
rbio68	19.78	23.01	2.27	53.15	60.76	80.77	36.96	**39.53**
db1(HAAR)	4.89	28.87	5.56	70.77	59.85	86.70	20.00	**39.52**
rbio11	2.87	26.80	4.17	72.21	59.67	86.24	23.33	**39.33**
bior13	15.06	16.83	3.95	62.61	55.72	88.74	32.35	**39.32**
bior39	25.41	37.39	15.56	53.68	40.72	63.98	33.33	**38.58**
bior55	27.65	24.30	7.32	48.19	41.50	75.44	42.50	**38.13**

The above table shows that the training is sufficient. Know the system is tested with new data not used for training. This testing was carried out for all the types of wavelet filters to find the wavelet filter that gives the best recognition rate. The results of the testing phase is shown in Table 3.

Table3: The recognition rates for the testing phase (continued).

Filter Type	Recognition Rate for z (%)	Recognition Rate for v (%)	Recognition Rate for th (%)	Recognition Rate for sh (%)	Recognition Rate for s (%)	Recognition Rate for f (%)	Recognition Rate for dh (%)	Total Recognition Rate(%)
db2	5.14	20.20	2.70	59.83	62.41	88.18	25.00	**37.64**
rbio26	18.89	27.52	5.95	63.71	48.09	66.96	28.57	**37.10**
sym27	23.00	25.50	8.13	43.50	53.31	71.85	28.05	**36.19**
bior22	48.58	22.77	9.21	51.27	18.45	79.28	23.53	**36.16**
bior24	25.56	29.52	10.00	57.14	37.80	72.12	18.42	**35.80**
rbio28	34.34	22.12	1.14	42.74	50.45	63.68	30.43	**34.99**
bior35	28.77	34.58	7.32	58.50	30.71	70.18	12.50	**34.65**
bior37	38.67	25.23	12.79	53.99	28.86	67.67	13.64	**34.41**
rbio22	21.88	19.80	3.95	58.07	43.36	72.52	14.71	**33.47**
bior33	39.27	28.16	5.13	54.37	24.45	62.05	19.44	**33.27**
rbio31	25.43	48.48	9.46	40.17	33.15	50.91	25.00	**33.23**
rbio35	39.66	30.84	8.54	43.45	30.35	50.44	25.00	**32.61**
rbio37	20.44	21.62	3.49	38.57	44.28	56.03	40.91	**32.19**
rbio39	27.60	20.87	4.44	43.60	26.85	49.15	39.58	**30.30**
rbio33	20.06	28.16	10.26	46.76	33.09	39.29	25.00	**28.94**
bior31	16.29	26.26	14.86	33.62	50.00	30.91	3.13	**25.01**
rbio11	2.87	26.80	4.17	72.21	59.67	86.24	23.33	**39.33**
bior13	15.06	16.83	3.95	62.61	55.72	88.74	32.35	**39.32**
bior39	25.41	37.39	15.56	53.68	40.72	63.98	33.33	**38.58**
bior55	27.65	24.30	7.32	48.19	41.50	75.44	42.50	**38.13**
db2	5.14	20.20	2.70	59.83	62.41	88.18	25.00	**37.64**
rbio26	18.89	27.52	5.95	63.71	48.09	66.96	28.57	**37.10**
sym27	23.00	25.50	8.13	43.50	53.31	71.85	28.05	**36.19**
bior22	48.58	22.77	9.21	51.27	18.45	79.28	23.53	**36.16**
bior24	25.56	29.52	10.00	57.14	37.80	72.12	18.42	**35.80**
rbio28	34.34	22.12	1.14	42.74	50.45	63.68	30.43	**34.99**
bior35	28.77	34.58	7.32	58.50	30.71	70.18	12.50	**34.65**
bior37	38.67	25.23	12.79	53.99	28.86	67.67	13.64	**34.41**
rbio22	21.88	19.80	3.95	58.07	43.36	72.52	14.71	**33.47**
bior33	39.27	28.16	5.13	54.37	24.45	62.05	19.44	**33.27**
rbio31	25.43	48.48	9.46	40.17	33.15	50.91	25.00	**33.23**
rbio35	39.66	30.84	8.54	43.45	30.35	50.44	25.00	**32.61**
rbio37	20.44	21.62	3.49	38.57	44.28	56.03	40.91	**32.19**
rbio39	27.60	20.87	4.44	43.60	26.85	49.15	39.58	**30.30**
rbio33	20.06	28.16	10.26	46.76	33.09	39.29	25.00	**28.94**
bior31	16.29	26.26	14.86	33.62	50.00	30.91	3.13	**25.01**

It is seen in the results that the first best five wavelet filters are Daubechies 21,23,22,18 and 15. Another point to notice is that the recognition rate is rather low. So the results of the first best five filters is further examined as seen in Table 4.

Table 4: The recognition rates of the first best five filters.

Filter		**z**	**v**	**th**	**sh**	**s**	**f**	**dh**	**Filter**	**z**	**v**	**th**	**sh**	**s**	**f**	**dh**
db21 **51.35**	Recognized as dh	0.00	2.86	55.71	0.00	0.00	4.29	**37.14**	**db18** **50.78**	0.00	1.56	40.63	0.00	0.00	7.81	**50.00**
	Recognized as f	1.16	0.78	42.64	8.14	1.94	**37.98**	7.36		1.19	1.98	19.05	9.52	0.79	**59.13**	8.33
	Recognized as s	19.41	0.00	2.43	2.95	**74.70**	0.35	0.17		33.6	0.00	1.40	4.20	**58.14**	2.28	0.35
	Recognized as sh	4.12	0.00	0.77	**76.03**	9.79	9.28	0.00		3.93	0.00	0.52	**74.35**	7.33	13.87	0.00
	Recognized as th	0.90	1.80	**70.27**	3.60	3.60	7.21	12.61		0.95	0.00	**46.67**	4.76	6.67	25.71	15.24
	Recognized as v	0.00	**45.26**	10.22	0.73	0.00	13.87	29.93		0.00	**38.93**	18.32	0.00	0.00	19.08	23.66
	Recognized as z	**18.04**	0.26	2.06	3.09	73.97	1.80	0.77		**28.3**	0.00	1.57	3.66	63.35	2.36	0.79
		z	v	th	sh	s	f	dh		z	v	th	sh	s	f	dh
db23 **51.31**	Recognized as dh	0.00	0.00	51.35	0.00	0.00	12.16	**36.49**	**db15** **50.51**	0.00	0.00	37.93	0.00	0.00	12.07	**50.00**
	Recognized as f	1.53	0.76	29.77	4.20	0.76	**57.63**	5.34		0.81	2.03	12.20	4.07	1.22	**69.11**	10.57
	Recognized as s	40.79	0.00	0.86	3.27	**51.98**	2.93	0.17		29.3	0.00	0.18	4.07	**61.77**	4.25	0.35
	Recognized as sh	3.57	0.00	0.51	**67.09**	8.67	20.15	0.00		4.26	0.00	0.27	**62.77**	7.71	25.00	0.00
	Recognized as th	3.48	0.00	**61.74**	3.48	1.74	20.87	8.70		3.03	0.00	**35.35**	1.01	8.08	36.36	16.16
	Recognized as v	0.00	**37.59**	12.77	0.71	0.00	21.28	27.66		0.00	**43.20**	10.40	0.00	0.00	11.20	35.20
	Recognized as z	**46.68**	0.00	3.57	1.02	45.92	2.04	0.77		**31.3**	0.53	0.00	1.86	62.50	2.39	1.33
		z	v	th	sh	s	f	dh								
db22 **51.06**	Recognized as dh	0.00	1.39	58.33	0.00	0.00	6.94	**33.33**								
	Recognized as f	1.54	5.77	43.85	5.77	1.15	**37.31**	4.62								
	Recognized as s	17.27	0.00	2.25	3.63	**75.99**	0.69	0.17								
	Recognized as sh	4.10	0.00	2.56	**68.97**	9.23	15.13	0.00								
	Recognized as th	3.54	0.00	**76.11**	1.77	7.08	4.42	7.08								
	Recognized as v	0.00	**43.88**	12.23	0.00	0.00	11.51	32.37								
	Recognized as z	**21.79**	0.51	2.31	3.08	70.77	0.51	1.03								

From the previous table, there are three major problems, the phoneme /dh/ is falsely recognized as /th/, the phoneme /f/ is falsely recognized as /th/ and the phoneme /z/ is falsely recognized as/s/.

9. Conclusions

The effect of the type of the wavelet filter on phoneme recognition in a phoneme recognition system based on wavelet and neural network was examined. From the results it is noticed that the Daubechies wavelet family is a good candidate for phoneme recognition system that are based on wavelets as a feature extraction stage, generally from db15 to db23. For the proposed system there was a problem of a false recognition of a phoneme specifically as another one (/dh/ as/th/,/f/ as /th/ and /z/ as /s/) this led to a degradation in the total recognition rate of the system. But keeping in mind that the main goal of this work is to find the best wavelet filter the results is still very useful in building any wavelet based phoneme recognition system. On the other hand these false recognitions are between similar pronounced fricatives and in most word are easily pronounced as each other according to the person. Therefore, if this taken into consideration and by adding this false values to the true values, for example for the db21 case, the total recognition rate can reach as high as 75.29% which is an acceptable value compared to recent phoneme recognition systems. For example 77% to 80% as in [21]

10.References

[1] Uma M., Kabilan A. and Venkatesh R."Speaker Independent Phoneme Recognition Using Neural Networks" Journal of Theoretical and Applied Information TechnologI, 2009, 230-235.

[2] Driaunys K., Rudžionis V. and Žvinys P. "Analysis Of Vocal Phonemes And Fricative Consonant Discrimination Based On Phonetic Acoustics Features", ISSN 1392 – 124X Information Technology And Control, Vol.34, No.3, 2005, 257-262.

[3] Rabiner L., Juang B., "Fundamentals of Speech Recognition", Prentice Hall, Englewood Cliffs, New Jersey, (1993).

[4] Koizumi T., Mori M., Taniguchi S. and Maruya M., "Recurrent Neural Networks for Phoneme Recognition" ICSLP 96, Proceedings, Fourth International Conference, Vol. 1, 3-6 October 1996, 326 –329.

[5] Abdelatty A., Van D. and Mueller P. "An Acoustic-Phonetic Feature-based System for Automatic Phoneme Recognition in Continuous Speech" IEEE ISCAS, May 1999, Proc. Vol. III, 118-121.

[6] Abdelatty A., Van D. and Mueller P. "Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants" IEEE Transactions on Speech and Audio Processing, Vol. 9, 2001, 833-741.

[7] Juneja V. and Epsy W., "Speech segmentation using Probabilistic Phonetic Feature Hierarchy and Support Vector Machines" Proceedings of International Joint Conference on Neural Networks, Portland, Oregan, 2003, 675-679.

[8] Siafarikas M., Todor G. and Nikos F., "Wavelet Packet Based Speaker Verification", Odyssey Speaker and Language Recognition Work Shop, Toledo, Spain, (May 31-June 3 2004), 257-264.

[9] Zi´ołko B., Manandhar S., and Wilson R., "Phoneme segmentation of speech" Proceedings of 18th International Conference on Pattern Recognition (2006).

[10] Long, C. and Datta, S."Wavelet based feature extraction for phoneme recognition", Proc. of. 4th Int. Conf. of Spoken Language Processing, Philadelphia, USA, Vol. 1 (1996) 264-267.
[11] Kadambe S. and Srinivasan P. " Adaptive wavelet based phoneme recognition" Proceedings of 40th Midwest Symposium Circuits And Systems, 1997.
[12] Tan B., Minyue F., Spray A. and Dermody P."The use of wavelet transforms in phoneme recognition", The Fourth International Conference on Spoken Language Processing, Philidelphia, USA, 1996.
[13] Modic R., Lindberg B. and petek B." Comparative Wavelet and MFCC Speech Recognition Experiments on the Slovenian and English SpeechDat2" An ISCA Tutorial And Research Workshop On Non-Linear Speech Processing, Le Croisic, France Demuth H., Beale M, 20-23 May 2003.
[14] Mporas I., Ganchev T.,M. Siafarikas M. and Fakotakis N." Comparison of Speech Features on the Speech Recognition Task" Journal of Computer Science , 2007, 608-616.
[15] Elenius K. and Tråvén H. "Multi-layer perceptrons and probabilistic neural networks for phoneme recognition" Proceedings of Eurospeech ,1993, 1237-1240.
[16] Cosi P., Mian G. and Contolini M."Speaker Independent Phonetic Recognition Using Auditory Modeling and Recurrent Neural Networks" Proceedings of ICANN-94, International Conference on Artificial Neural Networks, Sorrento, Italy, 26-29 May, 1994, 925-928.
[17] El-wakdy M., El-sehely E., El-tokhy M.," Speech Recognition Using A Wavelet Transform To Establish Fuzzy Inference System Through Subtractive Clustering And Neural Network (ANFIS)" 12th WSEAS International Conference on Systems, Heraklion, Greece, July 22-24, 2008, 381-386.
[18] Uma M., Kabilan A. and Venkatesh R.," Speech Recognition System Based On Phonemes Using Neural Networks " IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.7, July 2009, 148-153.
[19] Garofolo J., Lamel L., Fisher W., "Darpa TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM Manual", National Institute of Standards and Technology (NIST), (1993).
[20] Lee K., Hon H. "Speaker-Independent Phone Recognition Using Hidden Markov Models" Proceedings of IEEE Transaction on ASSP, Vol. 37, No. 11, (1989), pp. 1641-1648.
[21] Abdelatty A., Van D. and Mueller P. "An acoustic-phonetic feature-based system for the automatic recognition of fricative consonants" Proceedings of IEEE ICASSP'98, vol. 2, 1998, 961–964.
[22] Gorunescu F."Benchmarking Probabilistic Neural Network Algorithms" International Conference on Artificial Intelligence and Digital Communication, Research Center for Artificial Intelligence, (2006).

The work was carried out at the college of Engineering. University of Mosul