This is to certify that the
thesis entitled

AUGMENTING INFORMATION CHANNELS TO IMPROVE
COCHLEAR IMPLANT PATIENTS PERFORMANCE UNDER
ADVERSE CONDITIONS

presented by

ESRAA MUSTAFA AL-SHAROA

has been accepted towards fulfillment
of the requirements for the

MASTER OF SCIENCE degree in ELECTRICAL ENGINEERING

Major Professor's Signature

02/24/07
Date

MSU is an affirmative-action, equal-opportunity employer

Augmenting Information Channels to Improve Cochlear Implant Patients Performance Under Adverse Conditions

By

Esraa Mustafa Al-sharoa

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER'S OF SCIENCE

Electrical Engineering

2007

ABSTRACT

Augmenting Information Channels to Improve Cochlear Implant Patients Performance Under Adverse Conditions

By

Esraa Mustafa Al-sharoa

Cochlear Implant patients perform reasonably well in acoustically pristine environments. However, performance degrades significantly under adverse conditions, particularly within speech-like noisy surroundings. This thesis addresses the problem of resolving starts and ends of spoken words to improve speech intelligibility by CI patients under severe adverse conditions.

We propose a new approach based on sparse representation of speech signals. The approach is based on the discrete wavelet packet decomposition stemming from its excellent ability to capture transient signals. The obtained sparse representation is parameterized using a Gaussian mixture model to yield a feature set for classification and clustering purposes. We compare two methods to classify the start and end segments of spoken words in a noisy environment. The first relies on exploiting second order statistics of the sparsely represented signals, while the second relies on an expectation maximization approach to the Gaussian Mixture Model. We test the performance of both methods under various signal and noise conditions. Our preliminary results demonstrate that the sparse representation can capture eminent features in the spoken words that are indicative of start and end of the word. The proposed approach can be useful in driving cochlear implant signal transduction mechanisms with new features that are more robust to adverse conditions than the classical filter bank approach commonly used in current technology.

To my beloved husband, son, and parents

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Karim Oweiss, who without his guidance and valuable comments and inputs this thesis would not have been possible. In addition, I am very grateful to the committee members Dr. Hayder Radha, and Dr. Rong Jin for their useful comments. I would like also to thank my husband Mahmood Al-khassaweneh, my son Ahmad , and the coming daughter. With their help and support I was able to finish this thesis. Finally, I would like to thank my parents Mustafa Al-sharoa and my Mother Shama Al-kofahi, all my brothers and sisters, and my family in-law. I thank them their for support, kindness, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Cochlear implant has brought profound changes in the life of deaf people. For the past few decades scientists were able to partially restore hearing to deaf people by electrical stimulation of auditory nerve. The availability of this device and different signal processing strategies played a very important role in developing different techniques for deriving electrical stimuli from the speech signal.

Cochlear implants offer an easy way for deaf people to communicate with others. Signal processing techniques, mainly aimed to extract features from the speech signals to be transduced for driving implanted electrodes, help patients better communicate. The ability of the patient to understand what the speaker says in the existence of competing speakers, especially the recognition of the beginning and end of the words requires a robust signal processing algorithms particularly in the presence of noise.

Research on hearing aid and cochlear implant has made a considerable progress in the last two decades and attracted attention from different sides including speech science, signal processing, bioengineering, otolaryngology and physiology.

The focus of this thesis is on the recognition of the beginning and end of spoken words in a noisy environment.

This chapter is organized as follows. In section 1.1, different concepts that helps in understanding how cochlear implant and hearing aid work are discussed. Definition, characteristics, types of cochlear implants are discussed in sections 1.2 and 1.3. Section 1.4 gives an overview about what have been done in the field of cochlear implant and hearing aid in the past few years. Finally, section 1.5 summarizes the contribution of this thesis.

1.1 General Concepts

In order to understand how the cochlear implant is used to aid patients with damaged auditory system, the normal system should be studied first. In this section a brief description of the normal hearing, deafness and speech signal characteristics is presented.

1.1.1 Normal hearing

Human ear is divided into three main parts, the outer ear,the middle ear, and the inner ear. The acoustic stimulus travels from the environment through the outer ear to the middle ear where it is converted to mechanical vibrations towards the inner ear. In the inner ear the major part is the cochlea, which is a small shell-shaped cavity filled with fluid,transforms the received mechanical vibrations to the fluid to cause displacement in the basilar membrane. The hair cells attached to the basilar membrane are affected depending on the resultant displacement and stimulate neurons in the auditory nerve to transmit information about the signals to higher brain regions in the auditory cortex[1].

1.1.2 Deafness

Deafness occurs when mechanical energy of sound vibrations is not transmitted appropriately to neural information through the hair cells, so any damage of this part either fully or partially results in complete or partial loss of hearing. In the case of profound deafness a large number of hair cells or auditory neurons are damaged. However, there is usually an unknown number of cells that survive, and these can be electrically stimulated in an attempt to restore partial hearing[2]. [2]

1.1.3 Speech signal characteristics

When designing a CI, preserving certain information in the speech signal is a very important factor in the ability of the patients to understand the spoken words. The speech production process can be represented by the source-filter model[3]

The lungs can be considered as the source that produce different types of excitations, where there are two types. The first is the voiced periodic sounds that are produced by forcing the air through an opening between the vocal folds and the frequency related to this type of excitation is called the fundamental frequency F_0. The second is the unvoiced sounds generated constraining the pathway along the vocal tract and then forcing the air through it. The filter part of the speech production model is the vocal tract where the frequency response of this filter changes with changing the shape of the vocal tract, different shapes produce different sounds and frequencies , where the broad spectral peaks in the spectrum are called the formant frequencies and it was found that the formant frequencies carry significant information about the speech signal [4, 5, 3].

1.2 Cochlear Implants

The principle operation of CI is shown in Figure 1.1. The microphone picks up the acoustic signal and sends it to the speech processor that converts the received signal into electrical one, then the electrical signal is transmitted using a transmission link to an electrode array. Current cochlear implant uses the filter bank approach, as shown in Figure 1.1. A four channel implant, where the sound that is picked up by the microphone is processed through a set of bandpass filters that divide the signal into four channels, then the envelope of the signal is extracted using the envelope detector. The relative amplitude of the current pulses delivered to the electrodes reflect the spectral contents of the input signal. The basic assumption underlying the operation of CIs is that there are enough number of surviving auditory neurons that can be stimulated. The cochlear implant transmit different information about the received signal to the brain like sound pitch which is a function of the place that is stimulated in the cochlear implant and the sound loudness which is a function of the amplitude of the stimulus current.

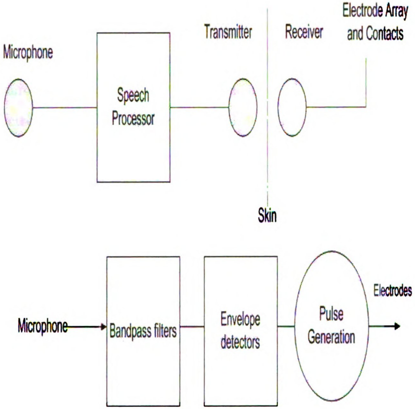

Figure 1.1. Schematic of the cochlear implant[6].

1.3 Characteristics and types of cochlear implant

1.3.1 Characteristics of cochlear implant

Different cochlear implant devices has different characteristics[7, 8], it can be summarized as follows:

1. Electrode design, multiple factors affect the electrode design, first is the location where the electrode is placed, the most common case is placing the electrodes in the scala tympani, and this kind of placement preserve the place mechanism of the

normal cochlea for coding frequencies, where this mechanism depends on the input frequency. For low frequencies, the electrodes near the apex are stimulated, which causes the auditory neurons that are tuned for low frequencies to be stimulated. For high frequencies, the electrodes near the base are stimulated to stimulate the auditory neurons that is tuned for high frequencies. Second factor is the spacing between electrodes and their number. The third is the configuration like monopolar and bipolar. Finally the orientation with respect to the excitable tissue[8, 9, 10].

2. Transmission link, there are two types of links that is used to transmit the signal from the speech processor to the electrodes, they are the transcutaneous connection and the percutaneous connection[8, 11].

3. Type of stimulation, where it can be analog or pulsatile. In the analog type, the acoustic signal is transmitted in electrical analog form to the electrodes while in the pulsatile stimulation the acoustic signal is transmitted in a narrow pulses form to the electrodes[8, 11].

4. Signal processing, different devices of cochlear implant use different signal processing strategies and this is one of the most challenging parts in the design of the cochlear implants where it transforms the speech signal to electrical stimuli[7, 11, 12].

1.3.2 Types of cochlear implant

There are two main types of cochlear implants, they are the single channel implants and the multichannel implants, where there are many different signal processing techniques that are used with these types of implants.

In the single channel implants the electrical stimulation is done at a single location on the cochlea, where it has only one electrode, while in the multichannel implants the electrical stimulation is provided to different locations on the cochlea using multiple electrodes or electrode array [12].Multichannel implants were introduced in 1980s, where the electrode array that is used in this type of implants allows stimulating different electrodes depending on the signal frequency. The number of electrodes in the electrode array varies from a device to another, where this matter is still under investigation. Different signal processing strategies have been used in multichannel implants, the main two categories that have been mentioned in literature are the waveform strategies and the feature extraction strategies. The waveform strategies can be summarized as presenting the analog or pulsatile waveform that is derived by filtering the speech signal into different frequency bands. On the other hand, feature extraction strategies use different feature extraction algorithms to present the speech signals such as the formants. An example of the waveform strategies is the Compressed-Analog (CA) approach, where it uses an automatic gain control to compress the signal and then filtering the signal into four frequency bands. The filtered waveforms are delivered simultaneously to four electrodes in analog form[11, 12]. An experiment on cochlear implants patients using this kind of signal processing is presented in [13], the problem of this strategy is the channel interaction, where the stimuli from one electrode might be destroyed the stimuli of another electrode and this will affect the neural responses and distort the spectral information of the speech signals[14]. Another example is the Continuous Interleaved Sampling(CIS) approach that was developed to resolve the problem of channel interaction by using nonsimultaneous, interleaved pulses, where only one electrode is stimulated at a time [12, 15]. Different factors affect the performance of the CIS approach like the pulse rate and duration, stimulation order and compression function[8, 16, 17]. In the Feature extraction approach the main device that uses these signal processing strategies is the Nucleus Multi-Electrode Implant that was introduced in the early 1980s, where many developments have been done for this device since that time[18]. Different strategies have been used in this device includes extracting the fundamental frequency(F_0)and the second formant frequency(F_2) and this is called the F_0/F_2 strategy[18, 19]. The F_0/F_2 strategy was developed by adding the extracting of the first formant and it was called the $F_0/F_1/F_2$ strategy[20]. More development where done in [22]. A comparison between the performance of the two methods has been presented in[21, 23].

1.4 Previous Work

During the last two decades numerous research has been done on cochlear implants and hearing aids and related topics. In [8]a review of the cochlear implant types, devices, characteristics and different signal processing strategies has been done.

In [24] the wavelet transform is used to implement the Continuous Interleaved Sampling(CIS) speech strategy in the cochlear implants, where it was found that using The WT will provide fast calculations in the CIS processor. A speech signal processing using Bionic wavelet transform and neural network simulation in [25] is proposed, where it was found that this approach reduces the number of the required channels and highly tolerates noise in addition to the reduction the stimulation duration for words and improving the recognition of vowels an consonants. A comparison between cochlear implants and acoustic hearing has been done in [26], where it measures the speech recognition in noise as a function of the number of the spectral channels and signal to noise ratio. Two different cochlear implant devices and three signal processing strategies have been used in this experiment, it was found that the performance of normal hearing people is better than the cochlear implants patients, and when comparing the CI patients performance with each other it was found that most of them were unable to make full use of the spectral information provided by the number of electrodes placed in their implants. A study about the effects of channel interactions of electrical pulse rate on the cochlear implants was done in [27]. Different experiments have been done in this paper to study this issue. The authors found that the channel interaction limits the number of information channels that are transmitted to the brain.

Since electrodes get activated if there is high noise between spoken words, many researchers studied the problem of noise reduction to improve the speech recognition.

Different methods and techniques were proposed for this goal. In [28],noise reduction algorithm in dual microphone behind the ear hearing aid is presented using the singular value decomposition based optimal filtering in combination with a voice activity detector. The algorithm helped in improving the ability to discriminate between the noisy signal periods and the noise-only periods.

Most signal processing strategies in the current cochlear implants focus on extraction of the amplitude modulation where it is sufficient in the speech recognition in noise free environment [29].However, in [30],a different speech processing strategy is proposed, where it encodes both amplitude and frequency modulations to improve the performance of the cochlear implants in a noisy environment. It was found that extracting and encoding frequency modulation is sufficient in the speech recognition in noisy environment.

Another speech processing strategy that encodes the envelop and the fundamental frequency of the speech signal to be used in the modulation of the amplitude and frequency of the electrical pulses that are transmitted to the electrodes was implemented in [31] on the tonal language. Moreover, a new signal processing strategies were proposed in [32]. Two methods of signal separation and noise suppression were presented to improve the performance of hearing aids and cochlear implants in adverse conditions. The authors found that the method performs better than the common band-pass filtering techniques used currently in HA and CI and it captures the rapid dynamics of the speech signal and minimizes the effect of noise.

In [33], the authorsdescribe some possible improvements that can be done for the cochlear implants to achieve better performance especially in multi-talker environment. They include the combination of electrical and acoustic stimulation of the auditory system when the patients have significant residual hearing. The authors also discuss the factors that affect the difficult cases of the cochlear implant patients.

1.5 Contributions of this thesis

In this thesis the problem of resolving starts and ends of spoken words is addressed to improve speech intelligibility by CI patients under severe adverse conditions. Two methods are implemented to study the characteristics of the transient periods of words. The first method is the Power method which uses signal subspace decomposition through low rank approximation to study the effect of the energy that is contained in the beginning and end of the words on the eigen vectors that spans the space of the target signal, and how do the competing signals affect the pattern of the target signal eigen vector. This method is explained in details in Chapter 2.

The second method is the parametrization method, where it parameterize a sparse representation of the speech signal with a Gaussian mixture model to obtain a feature set that can be used for classification and clustering. A detailed study of this method is presented in Chapter 3.

Finally, Chapter 4 concludes this thesis with a summary of contributions and future work.

CHAPTER 2

THE POWER METHOD

In this chapter the power of the speech signals is studied to see if it can help in the recognition of the beginning and end of the words.

This chapter is organized as follows. Section 3.1 gives an introduction about the power method. Section 3.2 describes the implementation of the proposed method. Section 3.3 provides simulation results to demonstrate the performance of the proposed method. Finally, Section 4.4 summarizes this chapter.

2.1 Introduction

Cochlear implants (CI) patients face a serious problem in the recognition of the beginning and end of the words in a noisy environment, where the desired signal undergoes considerable temporal modulation over the transitional time interval. Roughly speaking, the word has three parts: beginning, middle and end. The problem in the beginning and the end of the word is that their signals have aperiodic form,that is more transient like.

2.2 Implementation

Figure 2.1 shows the block diagram of the power method. Simple sinusoidal signals multiplied by exponential have been used to test this algorithm, considering them as the simplest case of the real speech signals, where these signals have different features that can be used in our study like frequency and temporal variation, increasing sinusoids were considered as the beginning of the words and decaying ones as the end of the words. This method can be summarized as follows:

1. Framing the received signal into frames of length N and calculating the power related to each frame signal.

2. The frames undergoes a $DWPT$ up to L levels, then finding the "best" subband, defined as the subband where the principal signal in each frame lives.

3. Singular value decomposition is computed for the wavelet coefficients found in the previous step to find the eigen vectors that span the signal subspace.

4. A comparison between the power calculations in step 1 with the eigen vector of the best subband , here we check if this eigen vector follows the same power pattern of the target signal over time.

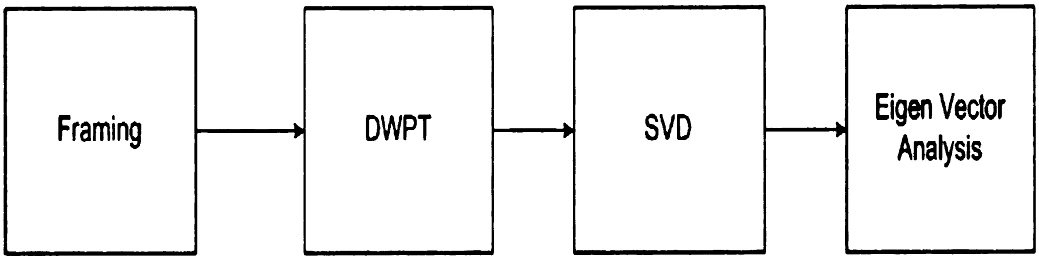

Figure 2.1. Block diagram for the Power method.

2.2.1 Framing

In studying speech signals it is necessary to transform the signal from its actual form into frames to guarantee local stationarity. Assuming that $s = [s[0]s[1]............s[N-1]]$ denotes the speech signal that will be transduced through the device over a time frame with length N. In a multi-speaker environment, the received signal is a mixture of speech signals and noise. Assuming that we have P independent speakers, the noise-free speech signal model is:

$$x_m = \sum_{p=1}^{P} a_{mp} s_p, \tag{2.1}$$

where, a_{mp} denotes the weight of the speaker p, s_p, in the m^{th} speech frame. the observations can be expressed in matrix form as,

$$Y_M = X_M + Z_M. \tag{2.2}$$

Framing the signal into M frames with frame length N according to,

$$Y_M = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_M \end{bmatrix} = \begin{bmatrix} a_1 s1(1)+ & ... & +a_p s_p(1) \\ a_1 s1(2)+ & ... & +a_p sp(2) \\ . & . & . \\ . & . & . \\ . & . & . \\ a_1 s1((M)+ & ... & +a_p sp(M) \end{bmatrix}, \tag{2.3}$$

Equation (2.3) represents time-invariant mixing, where the mixing of speakers across frames is fixed.

The other type of mixing is the time-variant mixing, where mixing of speakers across frames is dependent on the frame index. The signal is framed according to,

$$Y_M = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_M \end{bmatrix} = \begin{bmatrix} a_1(1) s1(1)+ & ... & +a_p(1) s_p(1) \\ a_1(2) s1(2)+ & ... & +a_p(2) sp(2) \\ . & . & . \\ . & . & . \\ . & . & . \\ a_1(M) s1((M)+ & ... & +a_p(M) sp(M) \end{bmatrix}, \tag{2.4}$$

where for $a_i(m)$, i is the speaker index and m is the frame index.

We consider the more challenging case of time-variant mixing. For the beginning of the word case the scaling factors are assumed to have the pattern, $a_M > a_{M-1} > ... > a_1$, while for the end of the word case, we assume $a_M < a_{M-1} < ... < a_1$. These assumptions are based on the definition of a start and end of a spoken word and characteristics of phonemes typically encountered in this segments of the word. The next step is to calculate the power of every frame, using Parseval's formula(2.5),

$$\text{Power} = \sum_{-\infty}^{\infty} |x(n)|^2. \tag{2.5}$$

2.2.2 Discrete wavelet packet decomposition

The m^{th} frame undergoes a Discrete Wavelet Packet Decomposition ($DWPT$) up to L levels (where the number of subbands is $J = 2^{L+1} - 1$ subbands). The $DWPT$ is an extension to the discrete wavelet transform, in other words the DWT is a subtree of the $DWPT$[34].

Figure 2.2 shows the wavelet packet filter-bank decomposition with successive filtering and down-sampling up to three levels.

In the DWT each level is calculated by decomposing the approximate coefficients through high and low pass filters, while in the $DWPT$ both approximate and details coefficients are decomposed using high and low pass filters and then down sampled, where this increases the frequency and time resolution.

After finding the $DWPT$ for every frame, the coefficients are rearranged in matrix form where each matrix contains the coefficients of all the frames at a certain subband, as described in equation (2.6).

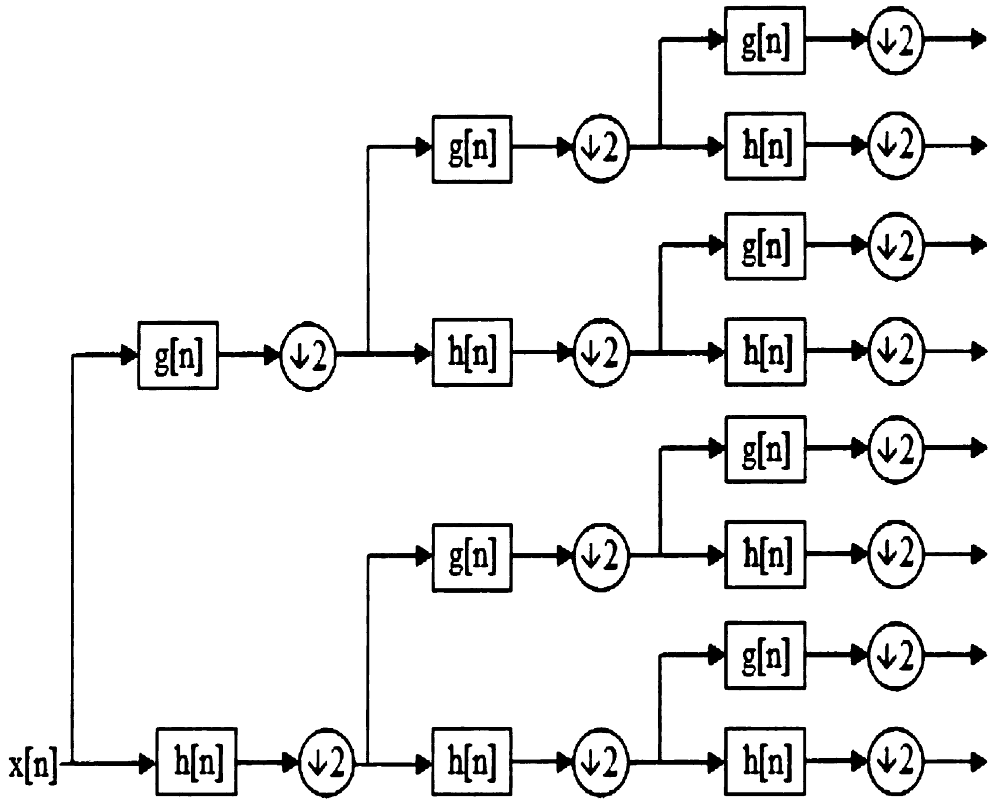

Figure 2.2. wavelet packet filter-bank decomposition up to three levels.

$$Coef(i,j) = \begin{bmatrix} y_1^{(i,j)} \\ y_2^{(i,j)} \\ \cdot \\ \cdot \\ \cdot \\ y_M^{(i,j)} \end{bmatrix}, \tag{2.6}$$

where $Coef(i,j)$ denotes for the wavelet coefficients at subband (i,j) and $y_k^{(i,j)}$ denotes the coefficients of the frame k at subband (i, j).

Now we can write,

$$y_m^j = x_m^j + z_m^j, \tag{2.7}$$

Where $y_m^j = (y_m^j(0), y_m^j(1), ..., y_m^j(N_j - 1))$ denotes the $DWPT$ of Y_m in the j^{th} subband. In doing so, one obtains an over complete representation of the observations in the form of a dictionary basis ΔJ to choose from. The transformed observations can be expressed in a matrix form as,

$$Y^j = X^j + Z^j. \tag{2.8}$$

When the energy of every subband is calculated, one can determine those with the highest energy , presumably where the signal lives.

It is important that the objective here is to detect the beginning and the end of the words in the target speaker speech signal in the observed mixture where the desired signal does not necessarily have the highest energy.

2.2.3 Singular value decomposition

Singular value decomposition(SVD) is a fundamental technique in many matrix analysis and computation[35]. By using SVD the matrix is decomposed into several component matrices that shows many useful and interesting properties of the original matrix. One more benefit of using the SVD of a matrix in computations is that it reduces the numerical error. In this method, our interest in using the SVD, comes from its ability to split a vector space into lower-dimensional subspaces.

This can be done using factorization of a rectangular real matrix. Assuming Y ia a $k * l$ matrix, then it can be expressed as,

$$Y = U_Y D_Y V_Y{}^*, \tag{2.9}$$

Where U_Y is $k * k$ unitary matrix that contains the orthonormal eigen vectors of YY^T.

D_Y is $k * l$ matrix with non-negative numbers on the diagonal and zeros every where else, where these numbers are the singular values in a descending order.

And $V_Y{}^*$ is the conjugate transpose of V_Y which is a unitary matrix with dimension $l * l$ and contains a set of orthonormal eigen vectors of Y^TY

In the proposed method SVD is used to find the eigen vector that spans the subspace of the target speaker, by calculating the factorization of the matrix(2.8) the resultant representation of this matrix is,

$$Y^j = U_Y^j D_Y^j {U_Y^j}^T, \tag{2.10}$$

After finding the SVD for every subband, then we find the eigen vector that is related to the subband that has the highest energy and compare the pattern of this eigen vector with the power pattern of the frames calculated in(2.5).

2.3 Simulations and Results

The Power method was applied to two sample signals, where the first was obtained using single sinusoids is multiplied by exponentially rising exponentials. In the first sample, the input signal was framed with frame length of $25ms$, then the power of every frame is calculated and normalized for the purposes of comparison.

Next step is the $DWPT$, where every frame is transformed from time domain to the time-frequency domain. In this step the power of every subband is calculated and by comparing the power of all subbands and choosing the one that has the highest power where the signal lives.

The third step is to find the SVD for the wavelet coefficients matrix found in the previous step, and check the principle eigen vector that is related to the subband where the signal lives and comparing the normalized power(Pn) with normalized eigen vector(Un).

Equation 2.11 describes the transformed observations of the speech signal to the time-frequency domain. Equation 2.12 describes the SVD factorization of 2.11,

$$Y^j = X^j + Z^j. \tag{2.11}$$

$$Y^j = U_Y^j D_Y^j {U_Y^j}^T, \tag{2.12}$$

Figure 2.4 and Figure 2.6 show a bar plots that presents a comparison between (Pn) and (Un) for two different signals $x1$ and $x2$, the first one fades in and the second one fades out. Figure 2.3, Figure 2.5 show the time domain representation of the two signals simultaneously. Where,

$$x1(t) = \exp(4t) cos(2\pi 400t) \tag{2.13}$$

$$x2(t) = 10 \exp(-2t) cos(2\pi 120t + \pi/5) \tag{2.14}$$

Table 2.1 through Table 2.10 present the details that is related to every figure, where it contains the frame number, the subband where the principal signal in each frame lives, the normalized power values and the normalized eigen vector.

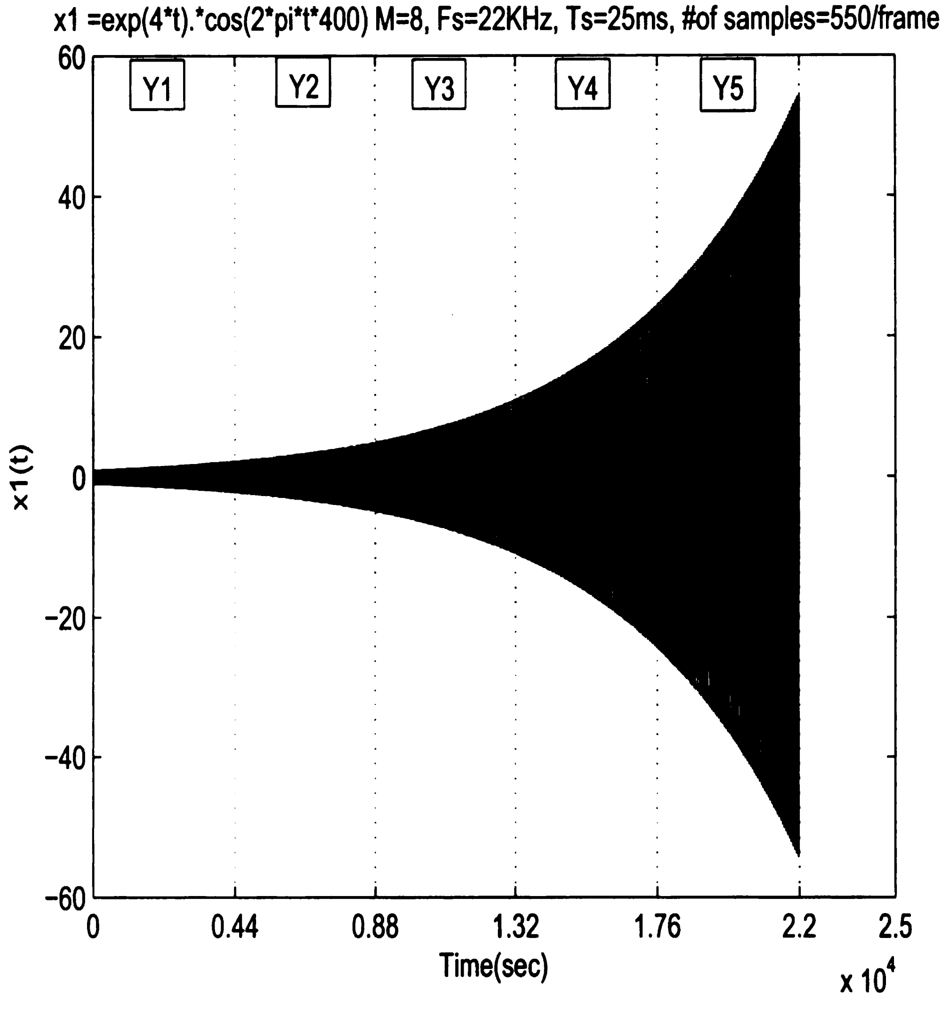

Figure 2.3. x1(t):Exponentially increasing signal

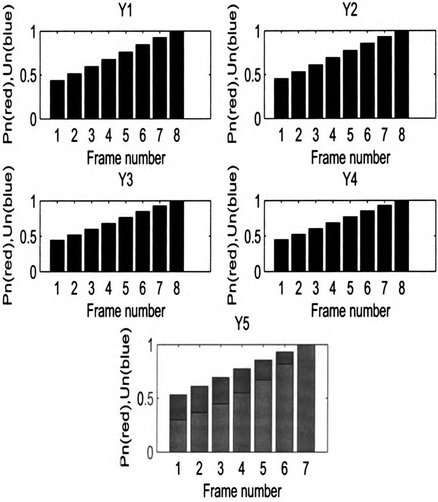

Figure 2.4. A comparison between Pn and Un for x1

Table 2.1. A comparison between Pn and Un at subband(5,2) for Y1 in xl

Frame	Subband	Pn	Un
1	(5,2)	0.2460	0.4351
2	(5,2)	0.3007	0.5117
3	(5,2)	0.3675	0.5922
4	(5,2)	0.4491	0.6755
5	(5,2)	0.5487	0.7600
6	(5,2)	0.6703	0.8438
7	(5,2)	0.8187	0.9247
8	(5,2)	1.0000	1.0000

Table 2.2. A comparison between Pn and Un at subband(4,1) for Y2 in xl

Frame	Subband	Pn	Un
1	(4,1)	0.2468	0.4536
2	(4,1)	0.3014	0.5297
3	(4,1)	0.3680	0.6092
4	(4,1)	0.4494	0.6908
5	(4,1)	0.5488	0.7728
6	(4,1)	0.6703	0.8534
7	(4,1)	0.8187	0.9301
8	(4,1)	1.0000	1.0000

From these figures and tables we can say that the eigen vector of the best subband follows the expected power pattern. In the second example, combination of the two signals $x1$ and $x2$ is used, where the mixing expression is presented in 2.15.

$$X(t) = \exp(4t)cos(2\pi 400t) + 10\exp(-2t)cos(2\pi 120t + \pi/5) \qquad (2.15)$$

Figure 2.7 shows the signal X in the time domain, and Figure 2.8 through Figure

Table 2.3. A comparison between Pn and Un at subband(4,1) for Y3 in x1

Frame	Subband	Pn	Un
1	(4,1)	0.2455	0.4401
2	(4,1)	0.2999	0.5166
3	(4,1)	0.3665	0.5968
4	(4,1)	0.4479	0.6796
5	(4,1)	0.5474	0.7635
6	(4,1)	0.6692	0.8464
7	(4,1)	0.8180	0.9262
8	(4,1)	1.0000	1.0000

Table 2.4. A comparison between Pn and Un at subband(4,1) for Y4 in x1

Frame	Subband	Pn	Un
1	(4,1)	0.2454	0.4451
2	(4,1)	0.3000	0.5214
3	(4,1)	0.3668	0.6014
4	(4,1)	0.4483	0.6838
5	(4,1)	0.5480	0.7669
6	(4,1)	0.6697	0.8490
7	(4,1)	0.8184	0.9276
8	(4,1)	1.0000	1.0000

2.12 show a comparison between the normalized power and the normalized eigen vectors. Table 2.11 through Table 2.15 show the details of these comparisons. The objective here is to examine if the eigen vector of every signal alone has the same pattern that is not affected by mixing.

Table 2.5. A comparison between Pn and Un at subband(4,1) for Y5 in x1

Frame	Subband	Pn	Un
1	(4,1)	0.3013	0.5316
2	(4,1)	0.3681	0.6110
3	(4,1)	0.4496	0.6924
4	(4,1)	0.5491	0.7742
5	(4,1)	0.6706	0.8544
6	(4,1)	0.8189	0.9306
7	(4,1)	1.0000	1.0000

Table 2.6. A comparison between Pn and Un at subband(5,1) for Y1 in x2

Frame	Subband	Pn	Un
1	(5,1)	1.0000	1.0000
2	(5,1)	0.9045	0.9525
3	(5,1)	0.8182	0.9061
4	(5,1)	0.7401	0.8610
5	(5,1)	0.6694	0.8172
6	(5,1)	0.6055	0.7747
7	(5,1)	0.5477	0.7335
8	(5,1)	0.4955	0.6937

Table 2.7. A comparison between Pn and Un at subband(5,1) for Y2 in x2

Frame	Subband	Pn	Un
1	(5,1)	1.0000	1.0000
2	(5,1)	0.9046	0.9525
3	(5,1)	0.8183	0.9062
4	(5,1)	0.7403	0.8611
5	(5,1)	0.6697	0.8173
6	(5,1)	0.6058	0.7748
7	(5,1)	0.5481	0.7336
8	(5,1)	0.4959	0.6938

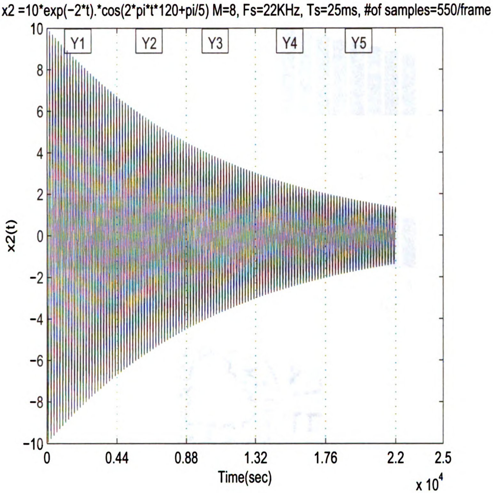

Figure 2.5. x2(t):Exponentially decaying signal

Table 2.8. A comparison between Pn and Un at subband(5,1) for Y3 in x2

Frame	Subband	Pn	Un
1	(5,1)	1.0000	1.0000
2	(5,1)	0.9047	0.9529
3	(5,1)	0.8185	0.9070
4	(5,1)	0.7406	0.8623
5	(5,1)	0.6700	0.8188
6	(5,1)	0.6062	0.7766
7	(5,1)	0.5485	0.7357
8	(5,1)	0.4963	0.6962

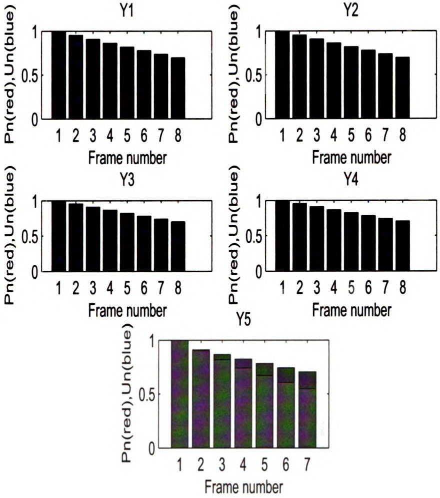

Figure 2.6. A comparison between Pn and Un for x2

Table 2.9. A comparison between Pn and Un at subband(5,1) for Y4 in x2

Frame	Subband	Pn	Un
1	(5,1)	1.0000	1.0000
2	(5,1)	0.9048	0.9537
3	(5,1)	0.8187	0.9085
4	(5,1)	0.7408	0.8644
5	(5,1)	0.6703	0.8215
6	(5,1)	0.6065	0.7798
7	(5,1)	0.5488	0.7393
8	(5,1)	0.4966	0.7001

Table 2.10. A comparison between Pn and Un at subband(5,1) for Y5 in x2

Frame	Subband	Pn	Un
1	(5,1)	1.0000	1.0000
2	(5,1)	0.9049	0.9099
3	(5,1)	0.8188	0.8664
4	(5,1)	0.7409	0.8241
5	(5,1)	0.6704	0.7828
6	(5,1)	0.6066	0.7428
7	(5,1)	0.5489	0.7040

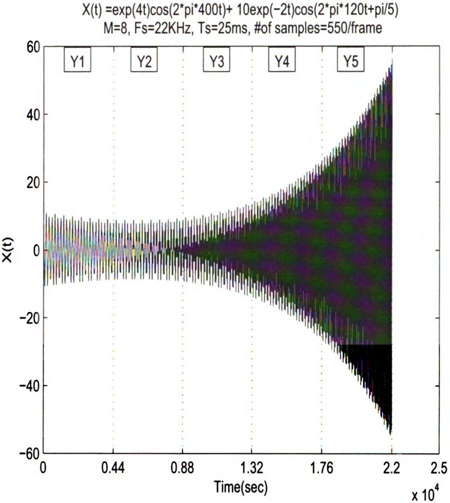

Figure 2.7. X(t):Combination of two signals:one feeds in and the other feeds out

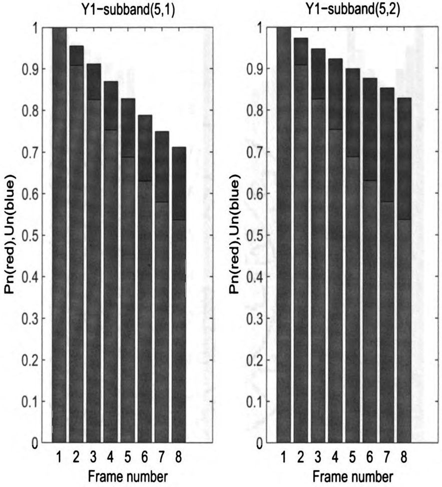

Figure 2.8. A comparison between Pn and Un for X

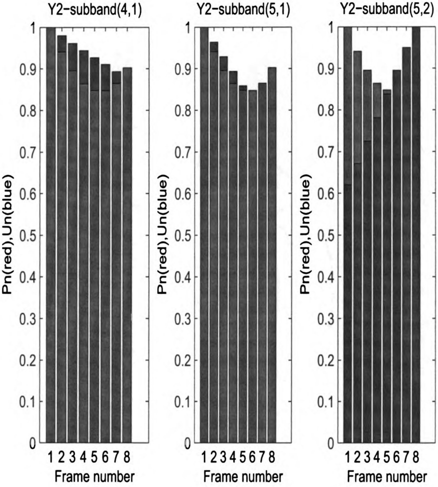

Figure 2.9. A comparison between Pn and Un for X

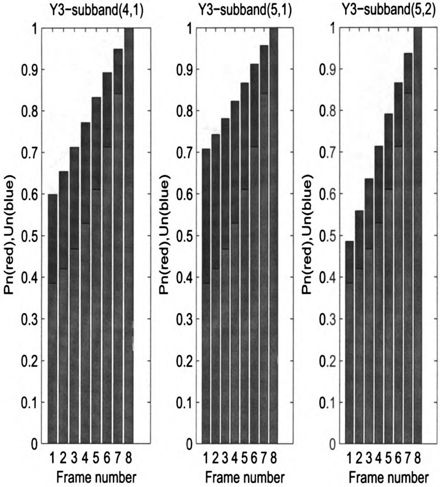

Figure 2.10. A comparison between Pn and Un for X

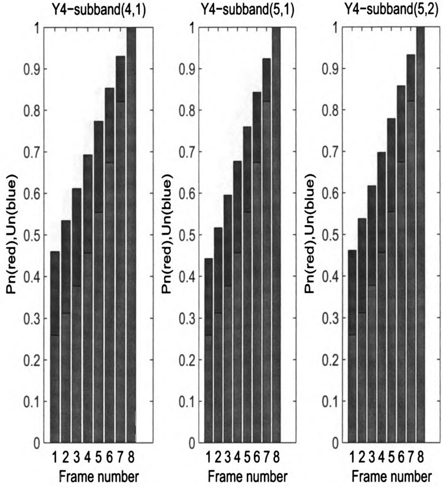

Figure 2.11. A comparison between Pn and Un for X

When applying this algorithm, it was expected that in the case of the combination of different exponential sinusoids, the dominant signal will have the main effect on the eigen vectors subspaces as we are using eigen decomposition which follows second order statistics and therefore will only track the highest power in a given frame and will not track the speaker specific signals. These expectation have been confirmed by applying this algorithm on multiple signals, where it was found that the wavelets offer no exception as compared to Fourier. The eigen vector that was assumed to span the

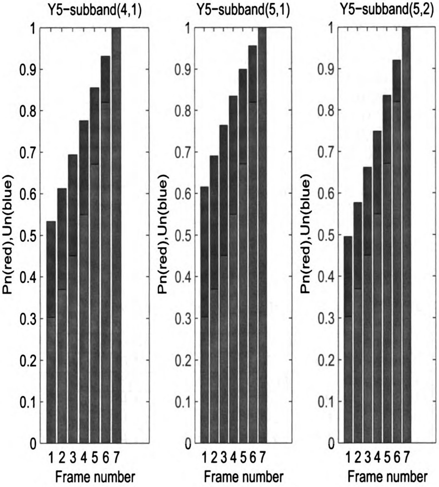

Figure 2.12. A comparison between Pn and Un for X

subspace of the target signal follows the power pattern of the dominant signal.

Table 2.11. A comparison between Pn and Un for Y1 in X

Frame	Subband(X)	Pn	Un(5,1)	Un(5,2)
1	(5,1)	1.0000	1.0000	1.0000
2	(5,1)	0.9083	0.9551	0.9729
3	(5,1)	0.8261	0.9114	0.9472
4	(5,1)	0.7528	0.8689	0.9227
5	(5,1)	0.6877	0.8278	0.8989
6	(5,1)	0.6303	0.7878	0.8755
7	(5,1)	0.5802	0.7489	0.8520
8	(5,1)	0.5370	0.7112	0.8278

Table 2.12. A comparison between Pn and Un for Y2 in X

Frame	Subband(X)	Pn	Un(4,1)	Un(5,1)	Un(5,2)
1	(5,1)	1.0000	1.0000	1.0000	0.6210
2	(5,1)	0.9409	0.9799	0.9638	0.6711
3	(4,1)	0.8954	0.9612	0.9283	0.7247
4	(4,1)	0.8640	0.9437	0.8933	0.7808
5	(4,1)	0.8475	0.9269	0.8586	0.8381
6	(4,1)	0.8472	0.9102	0.8237	0.8950
7	(4,1)	0.8648	0.8929	0.7883	0.9497
8	(4,1)	0.9025	0.8741	0.7520	1.0000

Table 2.13. A comparison between Pn and Un for Y3 in X

Frame	Subband(X)	Pn	Un(4,1)	Un(5,1)	Un(5,2)
1	(4,1)	0.3848	0.5989	0.7072	0.4854
2	(4,1)	0.4198	0.6536	0.7419	0.5589
3	(4,1)	0.4675	0.7115	0.7805	0.6353
4	(4,1)	0.5301	0.7714	0.8223	0.7132
5	(4,1)	0.6106	0.8321	0.8664	0.7908
6	(4,1)	0.7126	0.8918	0.9117	0.8662
7	(4,1)	0.8406	0.9486	0.9568	0.9369
8	(4,1)	1.0000	1.0000	1.0000	1.0000

Table 2.14. A comparison between Pn and Un for Y4 in X

Frame	Subband(X)	Pn	Un(4,1)	Un(5,1)	Un(5,2)
1	(4,1)	0.2589	0.4597	0.4423	0.4619
2	(4,1)	0.3116	0.5341	0.5165	0.5377
3	(4,1)	0.3765	0.6120	0.5949	0.6166
4	(4,1)	0.4563	0.6923	0.6764	0.6973
5	(4,1)	0.5541	0.7733	0.7596	0.7783
6	(4,1)	0.6739	0.8532	0.8427	0.8574
7	(4,1)	0.8206	0.9297	0.9237	0.9323
8	(4,1)	1.0000	1.0000	1.0000	1.0000

Table 2.15. A comparison between Pn and Un for Y5 in X

Frame	Subband(X)	Pn	Un(4,1)	Un(5,1)	Un(5,2)
1	(4,1)	0.3028	0.5334	0.6157	0.4954
2	(4,1)	0.3693	0.6125	0.6901	0.5768
3	(4,1)	0.4506	0.6936	0.7636	0.6615
4	(4,1)	0.5499	0.7752	0.8340	0.7482
5	(4,1)	0.6711	0.8551	0.8988	0.8350
6	(4,1)	0.8192	0.9309	0.9553	0.9198
7	(4,1)	1.0000	1.0000	1.0000	1.0000

2.4 Summary and Discussion

In this chapter, we proposed the Power method to study how the power of the signal affects the behavior of the beginning and end of the words in the wavelet domain. The proposed Power method used the sparse representation of the signal in combination with the singular value decomposition where the beginning and end of the words are considered as the target signal in a noisy environment, even if they don't have the highest energy.

The singular value decomposition of the sparse representation of the signals has been used to study the eigen vectors that spans the subspace of the desired signal when it is found in combination with other signals. The assumption was that the eigen vector that spans the subspace of the target signal might keep the same power pattern even in the existence of other signals, but it was found that the competing signals might be dominant and this affects the behavior of the eigen space of the whole signal, where the eigen vector of the target signal follows the power pattern of the dominant signal.

CHAPTER 3

THE PARAMETRIZATION METHOD

In the previous chapter, we introduced the power method. The results of this method were not satisfactory. Therefore in this chapter, we statistically analyze the distribution of the wavelet coefficients using a parametric model to improve the classification task.

This chapter is organized as follows. Section 3.1 gives an introduction about the proposed method. Section 3.2 describes the implementation of the parametrization method. Section 3.3 provides simulation results to demonstrate the performance of the proposed method. Finally, Section 4.4 summarizes the major contributions of this chapter.

3.1 Introduction

In this method a parametric model of the probability distribution of the wavelet coefficients is used rather than the raw wavelet coefficients to create a feature space that can be used in classification and clustering purposes.

3.2 Implementation

Figure 3.1 shows the block diagram for the parametrization method. As seen, there are seven steps to implement the proposed method. In the first step the words are

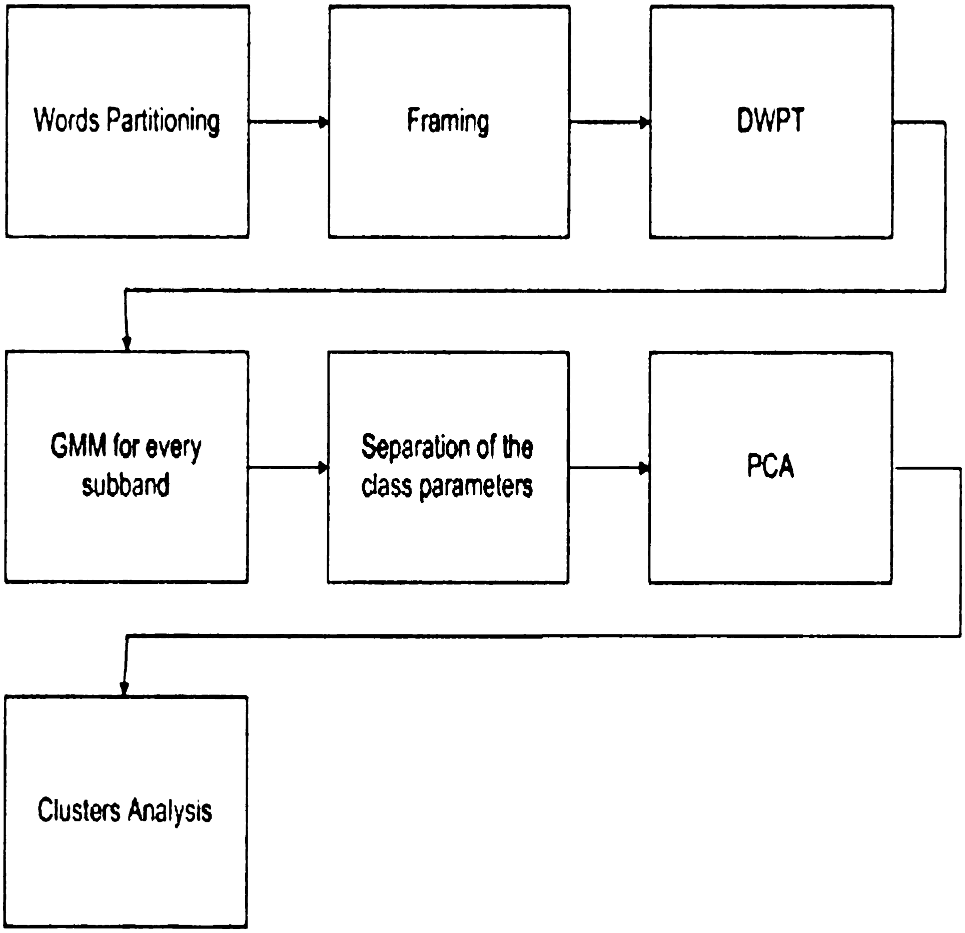

Figure 3.1. Block diagram for the Parametrization method.

divided into 3 main classes:start, middle and end. In the second step the the new form of the signal is framed with a frame length of N. The third step transforms the signals to the wavelet domain using the discrete wavelet packet decomposition. The forth step fits the probability distribution of the wavelet coefficients with a Gaussian mixture model(GMM), where the coefficients of every subband can be modeled as more than one Gaussian with different parameters(mean and variance)[36]. The next step classifies the parameters of the GMM depending on the frames class.

3.2.1 Words Partitioning

the Speech signals used in this method with babble noise added at different signal to noise ratios [37]. In this step the words are divided manually into three parts: start, middle and end. This knowledge of the parts of the word will help in studying the characteristics of each part alone.

3.2.2 Framing

The input to the algorithm is a vector that contains the three parts of the words. The frame length that is used in this approach is 25ms. Same way of framing that is presented in 2.2.1.

3.2.3 Discrete Wavelet Packet Decomposition

Here the frames undergoes a discrete wavelet packet decomposition up to L levels as explained in 2.2.2.

3.2.4 Gaussian Mixture Model

In this step a Gaussian mixture model(GMM) is used to fit the probability distribution of the wavelet coefficients found in 3.2.3, where each set of coefficients is represented with two Gaussian's with two parameters(mean and variance)and the parameters arranged into two main matrices, the mean matrix and the variance matrix. Each matrix has the parameters of all the frames at a certain subband as described in equations (3.1) and (3.2) respectively.

$$\mu_{i,j} = \begin{bmatrix} \mu_{11} & \mu_{12} & .... & \mu_{1M} \\ \mu_{21} & \mu_{22} & .... & \mu_{2M} \end{bmatrix}, \tag{3.1}$$

where $\mu_{i,j}$ denotes the mean of distributions characterizing the coefficients at subband (i, j) using the two Gaussian's.

$$\sigma^2_{i,j} = \begin{bmatrix} \sigma^2_{11} & \sigma^2_{12} & .... & \sigma^2_{1M} \\ \sigma^2_{21} & \sigma^2_{22} & .... & \sigma^2_{2M} \end{bmatrix}, \tag{3.2}$$

where $\sigma_{i,j}$ denotes the variance of distributions characterizing the coefficients at subband (i, j) for the two Gaussian mixtures.

3.2.5 Separation of Parameters

After estimating the parameters of the Gaussian mixture model that represents the distribution of the wavelet coefficients, this step comes to separate these parameters depending on the class that they belong to according to. This can be done by determining the parts of equations (3.1) and (3.2) that belong to the start, the middle and the end frames.

3.2.6 Principle Component analysis

In this step the algorithm starts to go into two directions to analyze the problem, the first is the classification and the second is the clustering. In the classification problem, the (PCA) is done for the parameters of each class separately, where every class is projected on its own principle components.

In the clustering problem the PCA is done for 3.1 and 3.2, which means finding the principle components of the whole signal regardless the type of the class and using them for the projection step. The expectation here is that if the features obtained are distinct among these classes, we will see separate clusters in the feature space[38].

3.2.7 Clusters Analysis

In cluster analysis the observed data set can be organized into groups depending on the degree of associations between the objects in each group.

There are different methods to examine how objects in a cluster are related to each other, which can be the degree of separation of these clusters or how compact is every cluster. One approach is computing distance between the objects that is forming the different clusters. One example of a distance measure is Euclidean distance, where this type is the geometric distance in a multidimensional space, and it can be computed according to,

$$E_d(v_1, v_2) = \sqrt{\left(\sum_i (v_{1i} - v_{2i})^2\right)}. \tag{3.3}$$

Equation (3.3) can be written as,

$$E_d^2(v_1, v_2) = \sum_i (v_{1i} - v_{2i})^2, \tag{3.4}$$

which is called the squared Euclidean distance.

The separability factor ρ is computed at each subband, defined as,

$$\rho = \frac{\text{average distance across clusters}}{\text{average distance within clusters}} \tag{3.5}$$

ρ can be used to study the clusters separation at every subband, where the subband that has a large value of ρ implies a high degree of separation and more compact clusters.

3.3 Simulation and Results

In the words partitioning step, we have considered two cases the first one divides the words into two classes, where it considers the transient periods of speech as one class(start and end of the words)and the steady state periods as the second class(middle of the words), this is referred to as the two-class problem. The second approach divides the word into three classes where it considers the start and the end as separable classes and the third class is the middle of the word. This is referred to as the three-class problem. Figure 3.2 shows an example of the input signal used in this algorithm in the time domain "Sky that morning was clear and bright blue".

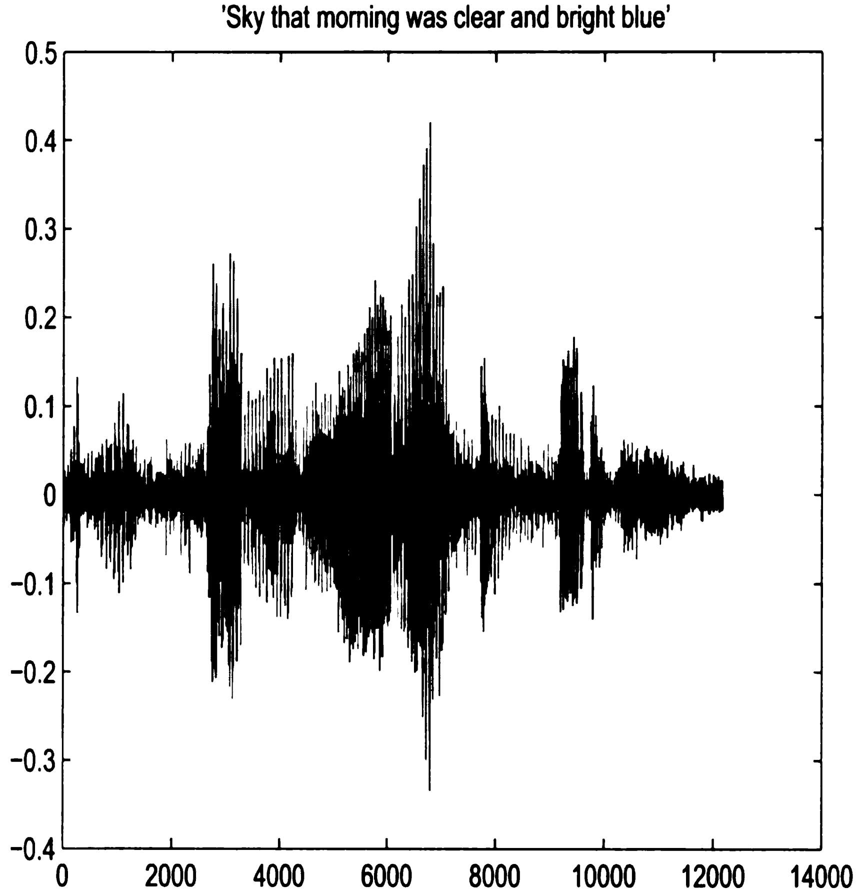

Figure 3.2. The input signal: "Sky that morning was clear and bright blue".

In the framing step the signal is framed with frame length of $25ms$, and then these frames undergoes a $DWPT$ and the wavelet coefficients are rearranged as mentioned in 3.2.3 to find the $DWPT$, different Daubechies and Symlets filters have been used. After computing the $DWPT$, the Gaussian mixture model is used to fit the probability distribution of the wavelet coefficients as shown in the figures. Figure 3.3 shows the histogram of the wavelet coefficients that represents frame1 at subband $(1, 0)$ and the Gaussian mixture that is used to fit the probability distribution of these coefficients and the related parameters.

Figure 3.4 show the histograms of frame1 at subband(2,0) with the related GMM.

In the next step the parameters of the different classes are separated as explained in 3.2.5 in the classification problem and then a principle components analysis is done for each class separately and the resultant principal components are used to compute the new representation of the GMM parameters computed before. By projecting each class on its own principle components, the clusters analysis starts takeing place to study the separability between different classes. In this method the metric that is used to study the clusters is the distance between clusters by using the factor ρ mentioned in equation (3.5).

Figure 3.5 through Figure 3.8 show the result by applying the method as two classes problem for different speech signals.

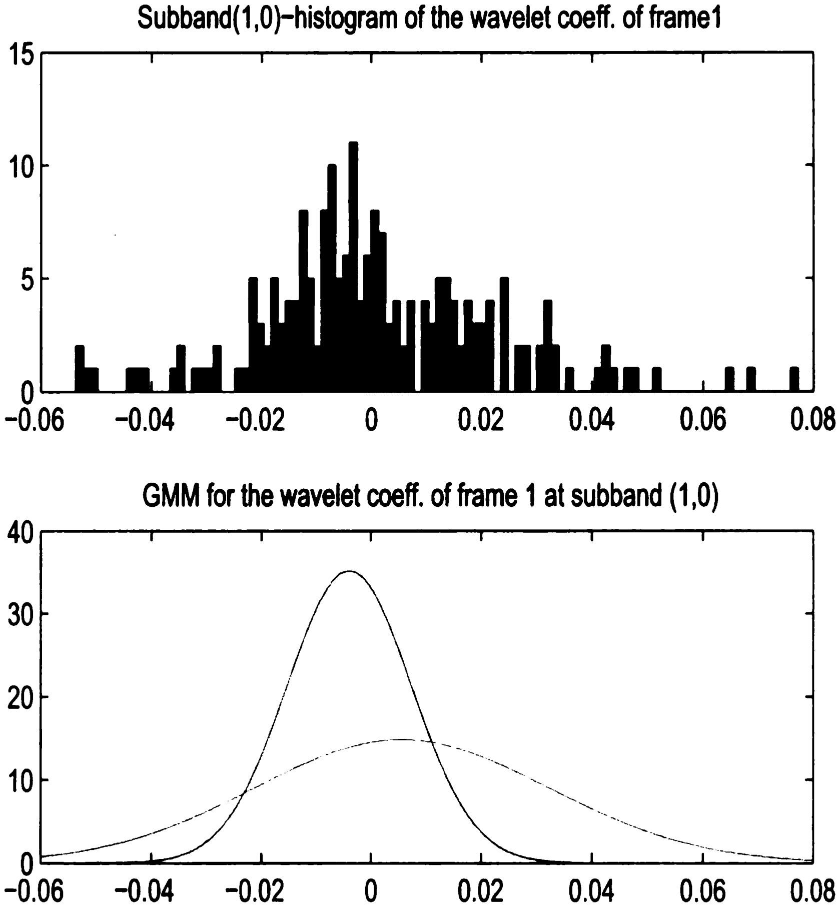

Figure 3.3. The histogram of the wavelet coefficients that represents frame1 at subband (1, 0) and its Gaussian mixture model.

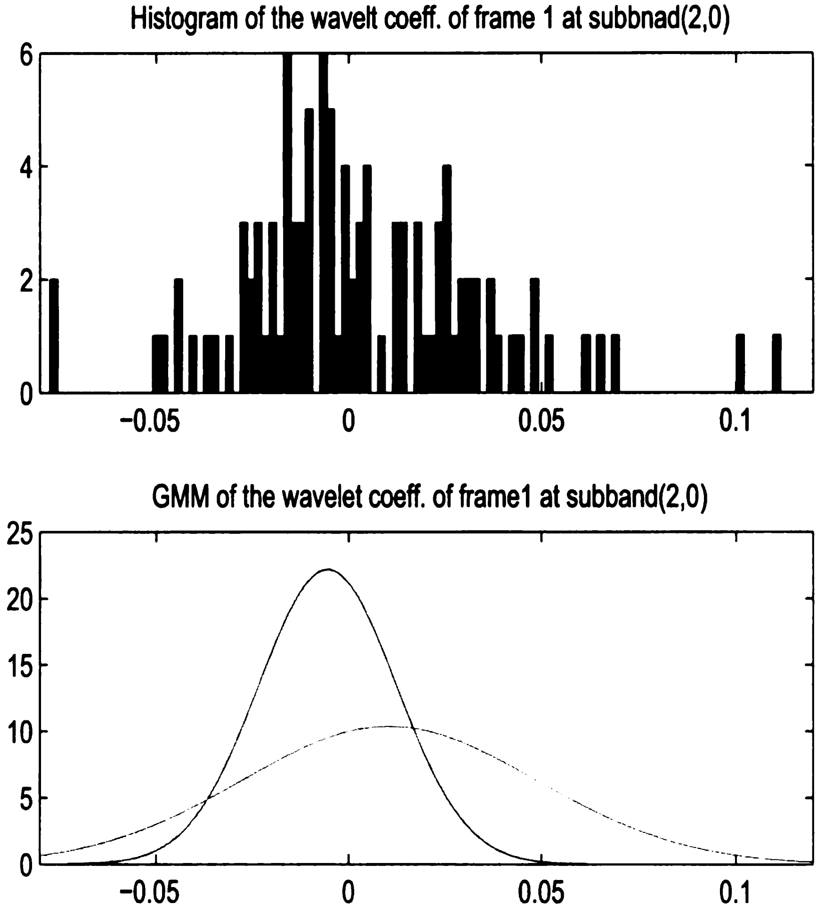

Figure 3.4. The histogram of the wavelet coefficients that represents frame1 at subband $(2, 0)$ and its Gaussian mixture model.

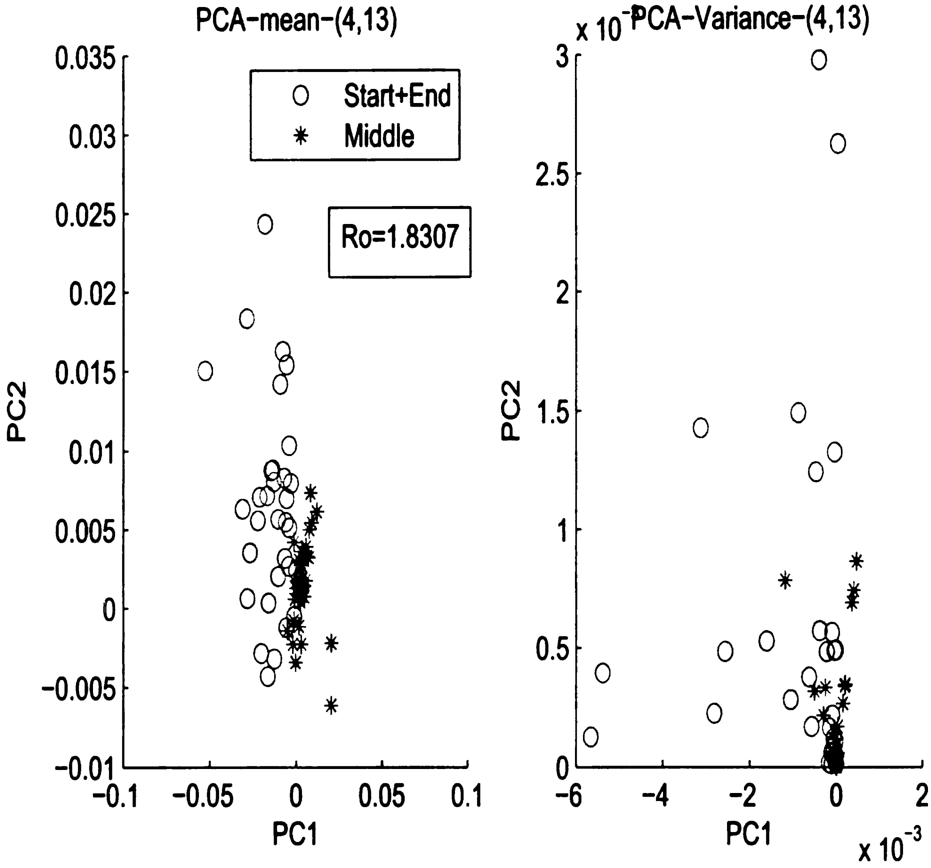

Figure 3.5. PCA for the start/end and middle of the words at subband $(4, 13)$ and SNR=15dB.

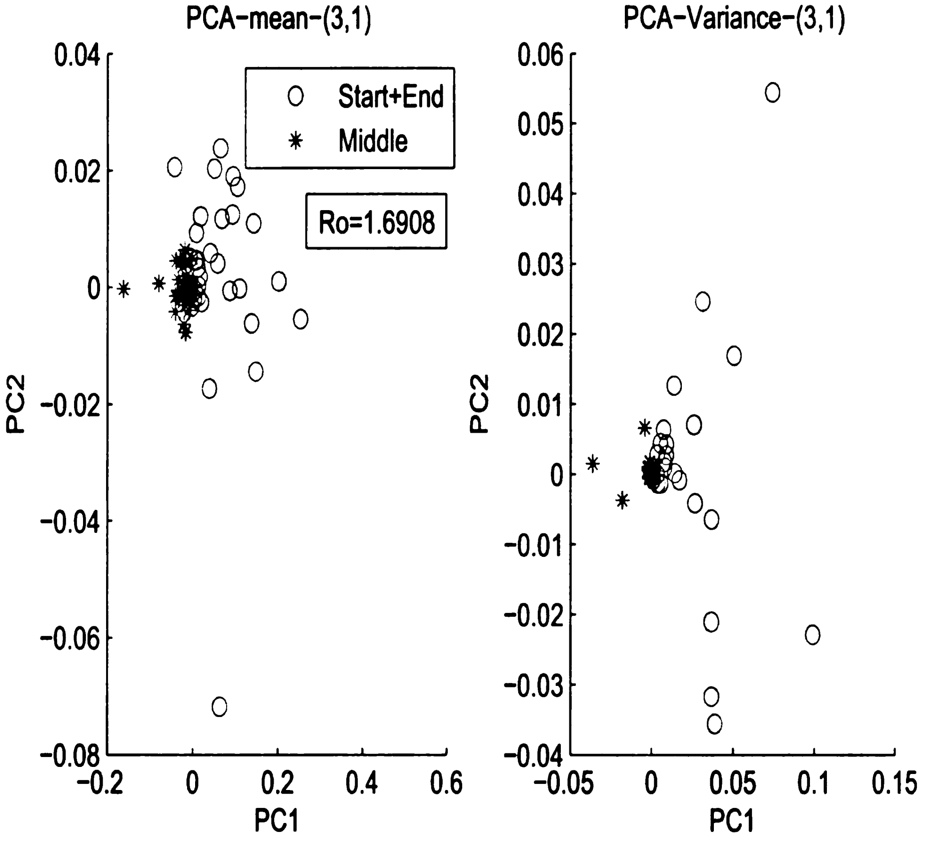

Figure 3.6. PCA for the start/end and middle of the words at subband $(3, 1)$ and SNR=15dB.

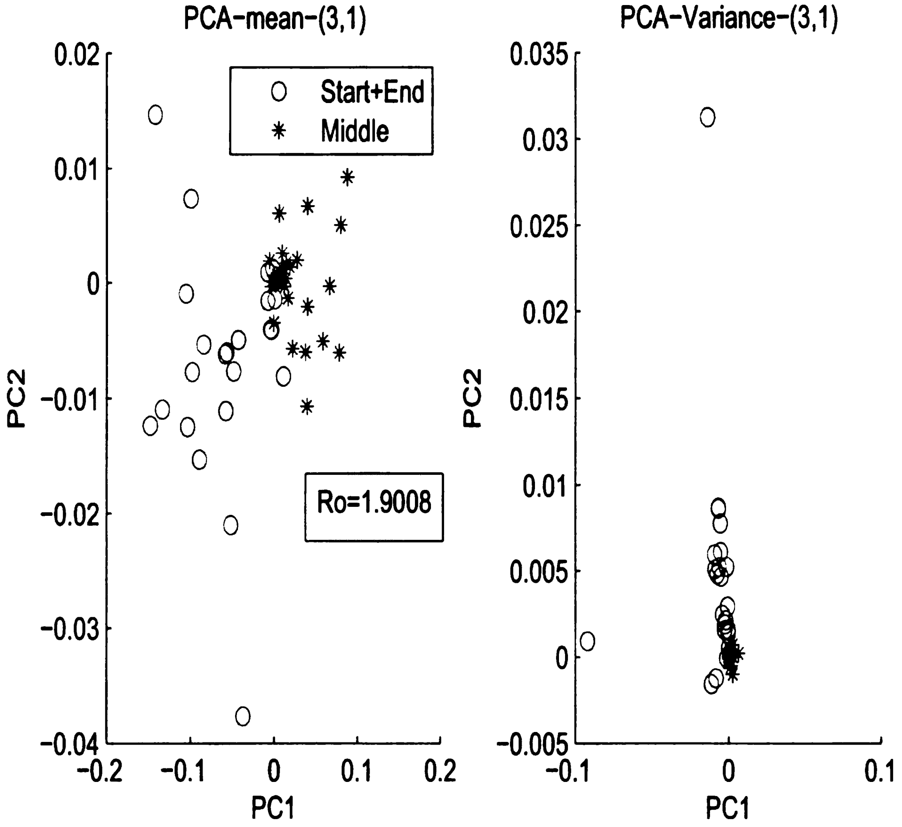

Figure 3.7. PCA for the start/end and middle of the words at subband (3, 1) and SNR=15dB.

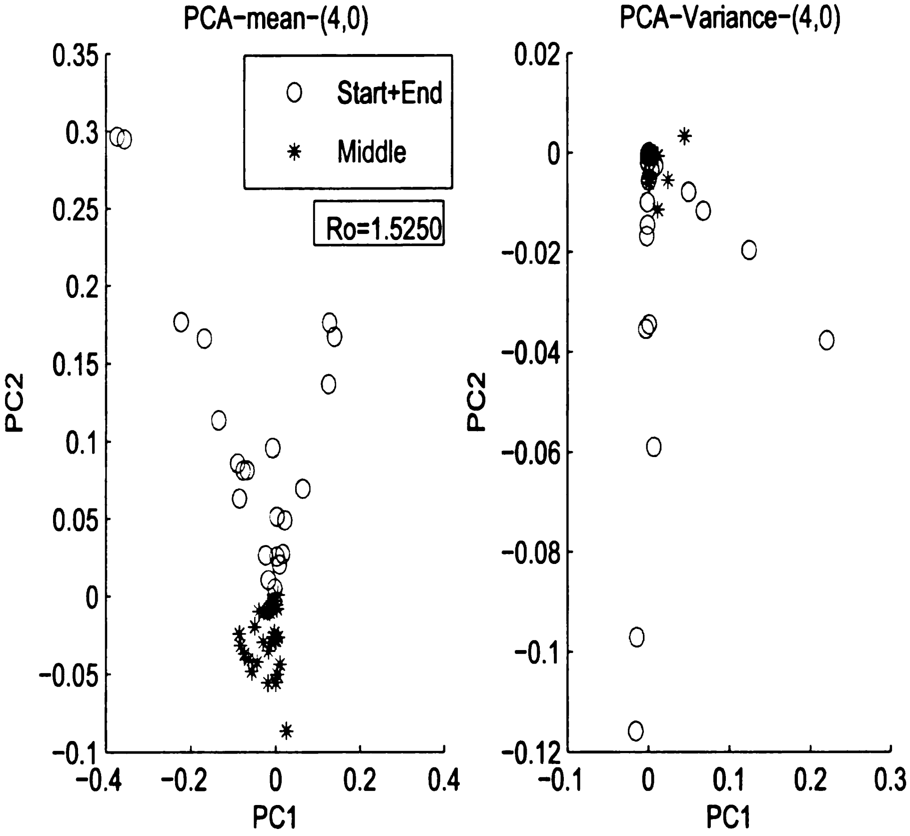

Figure 3.8. PCA for the start/end and middle of the words at subband (4,0) and SNR=15dB.

In the classification problem, we are trying to see if the eigen vectors that is produces by a ceratin signal can be used as a basis for other signals, which means if we have a frame and we need to classify this frame, then these frame undergos the same steps until the *PCA* step and then the already computed eigen vectors for the different classes are used to project this frame and to get the new coordinates for it. The next step is to measure the distance between this projected point and the centroid of every cluster, where the decision will be taken that the frame belongs to one of these clusters where the computed distance is minimum, The following ratio measure is used to find which subband performs better , 3.6,

$$\text{Ratio} = \frac{\text{Distance between the test frame and the centroid of the transient cluster}}{\text{Distance between the test frame and the centroid of the steady-state cluster}} \tag{3.6}$$

Where if this Ratio is less than 1, the frame is classified as transient. If the Ratio is more than 1, the frame is classified as steady-state. Figure 3.9 shows an example of the classification problem, where in this example a speech signal that has in the end of one of its words the phoneme 'n'. This signal is used to classify three frames that came from another word that has 'n' in the end. The subband that has the smallest Ratio value is shown in Figure 3.9, where the Ratio is less than 1, which means that the eigen vectors that was found for a certain signal can be used to classify the unknown frames if they belong to transient periods of the speech signals or steady-state.

In Figure 3.9 the left side shows the *PCA* of parameter 1, and the right side shows the *PCA* of parameter 2, where the blue bubbles refer to the parameters of the transient frames, asterisks refers to the parameters of the steady-state frames, red bubbles refer to the test frames projected on the eigen vectors of the transient frames, and the black bubbles refer to the test frames projected on the eigen vectors of the steady-state frames.

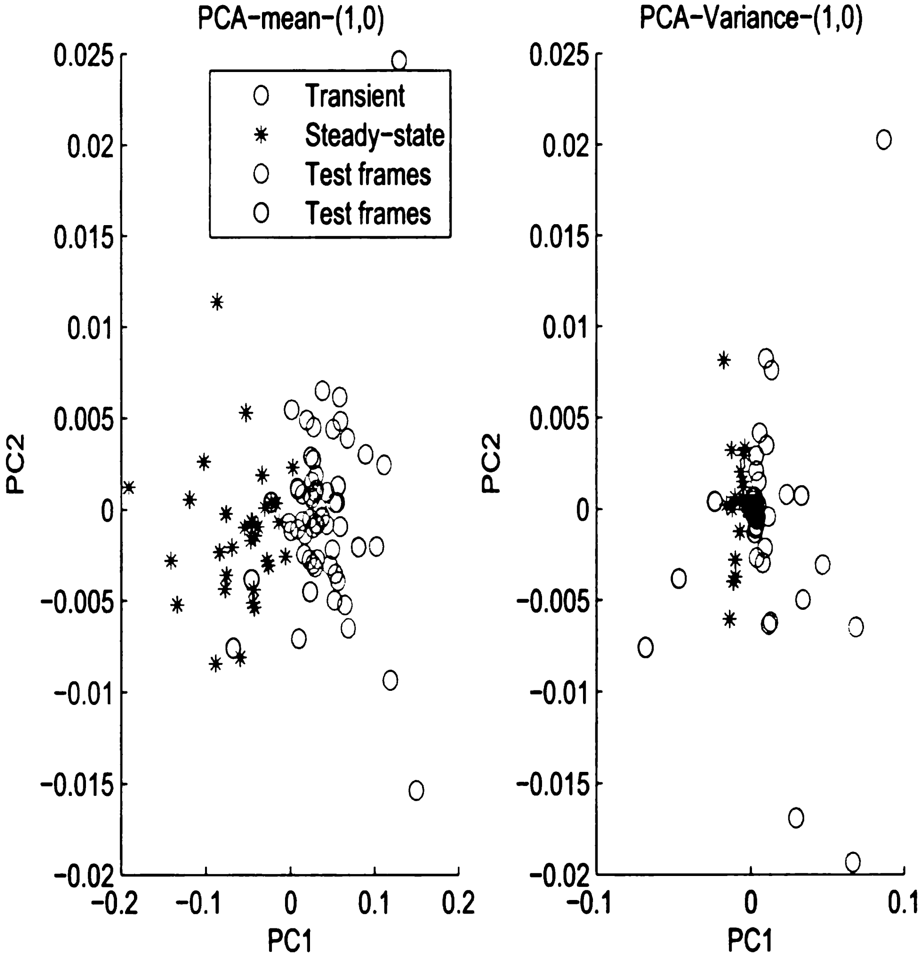

Figure 3.9. PCA for the transient and steady-state parts of the words at subband $(1, 0)$ and for the test frames.

As mentioned before in the clustering problem, the principal components of the whole signal are computed Figure 3.10 through Figure 3.16 show the clustering of the different classes across subbands.

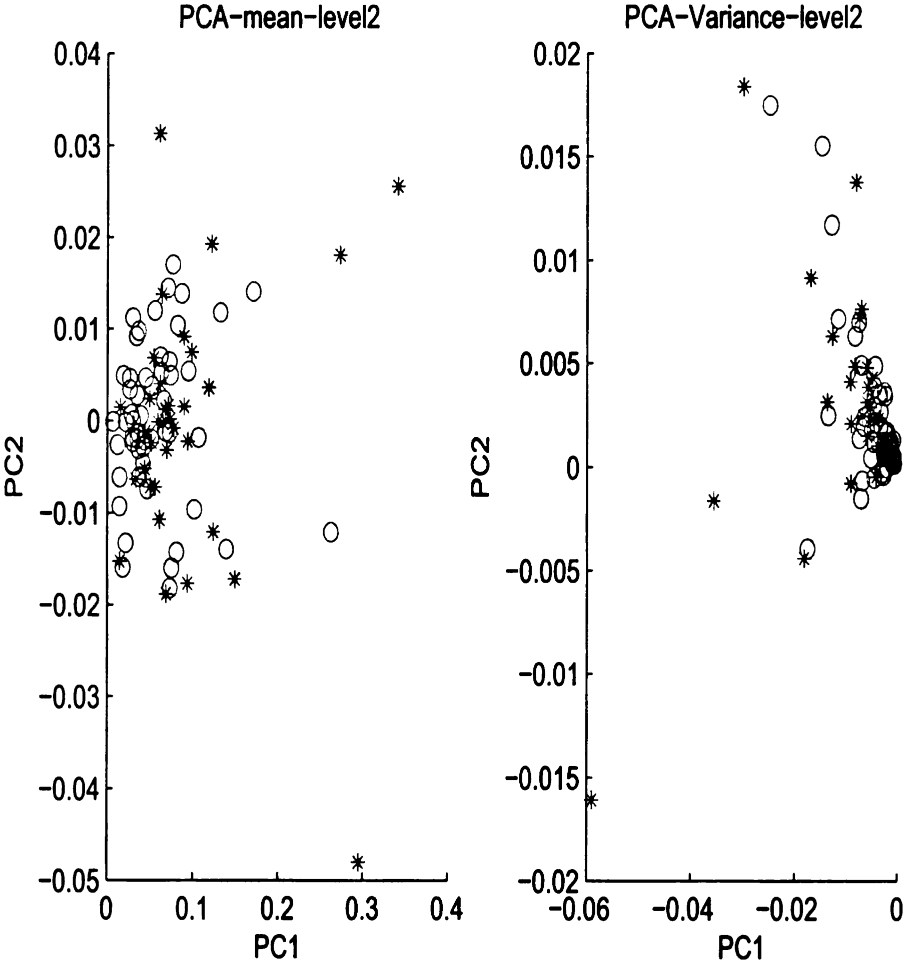

Figure 3.10. Clustering of different classes at subband(2,0).

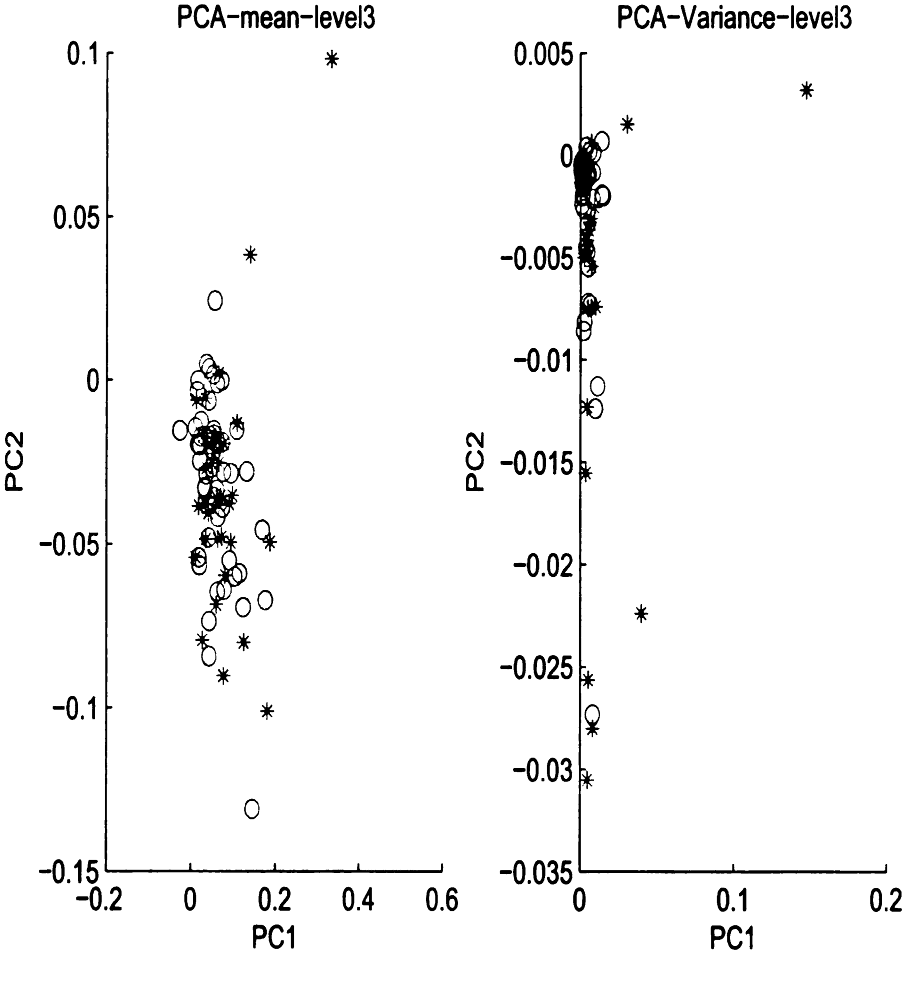

Figure 3.11. Clustering of different classes at subband(3,0).

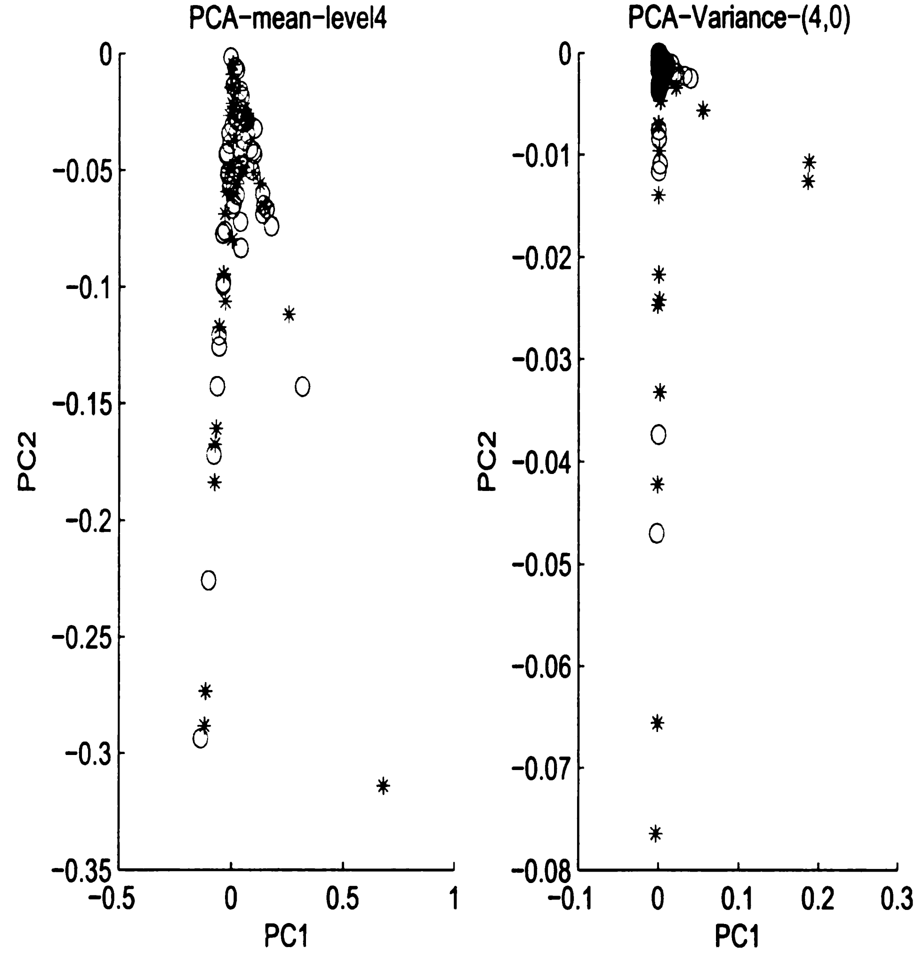

Figure 3.12. Clustering of different classes at subband(4,0).

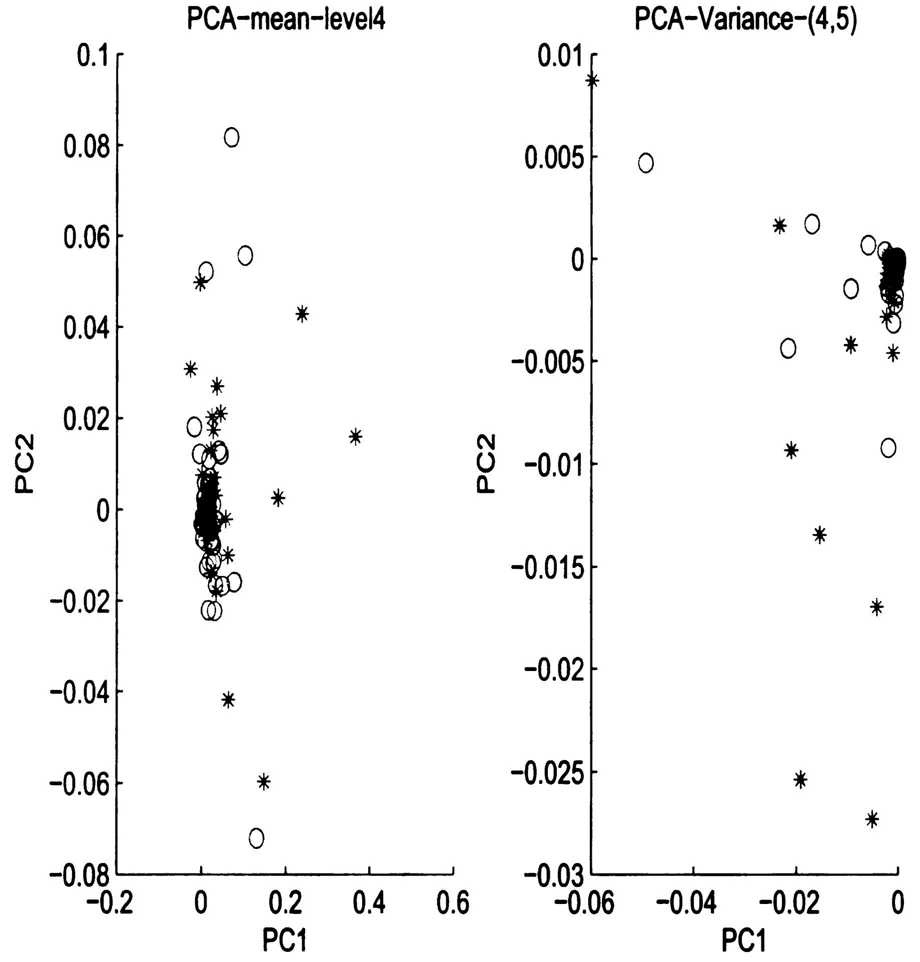

Figure 3.13. Clustering of different classes at subband(4,5).

Expectation maximization algorithm (EM), is a technique for separating clusters[39]. We used it here to separate the transient cluster from the steady-state cluster. For the example given in Figure 3.9, it was found that the subband that shows separability between the clusters very close to the expected clusters is (1,0), Figure 3.17 shows the estimated cluster of the transient periods of the speech signal by EM in the ellipse, where the expected clusters are presented in different styles, where the blue bubbles refer to the parameters of the transient frames, black bubbles

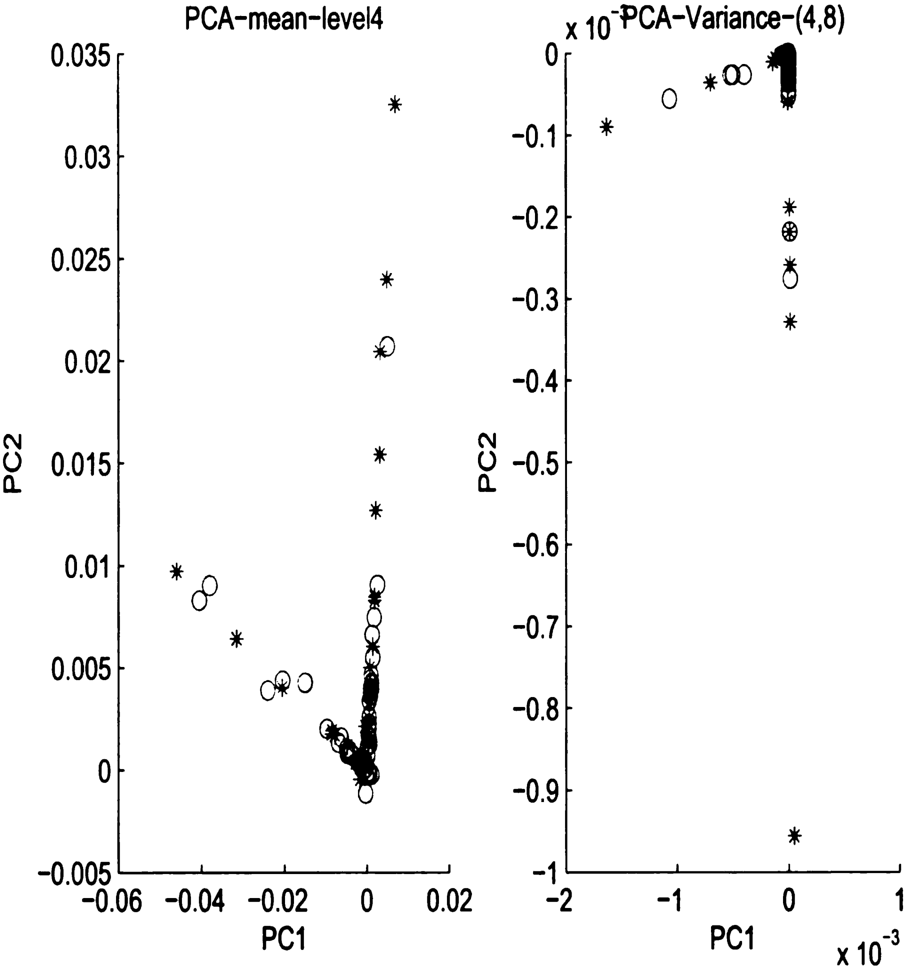

Figure 3.14. Clustering of different classes at subband(4,8).

refer to the parameters of the steady-state frames, Table 3.1 shows the number of correctly classified transient frames by EM, unclassified transient frames and false alarm frames.

Figure 3.18 shows the clustering of the classes for another signals with the test frames, and Figure 3.19 shows the estimated clusters by EM at that subband, where we were able to estimate a cluster that mainly captures the steady-state points. Table 3.2 shows the number of correctly classified steady-state frames, unclassified frames and false alarm frames.

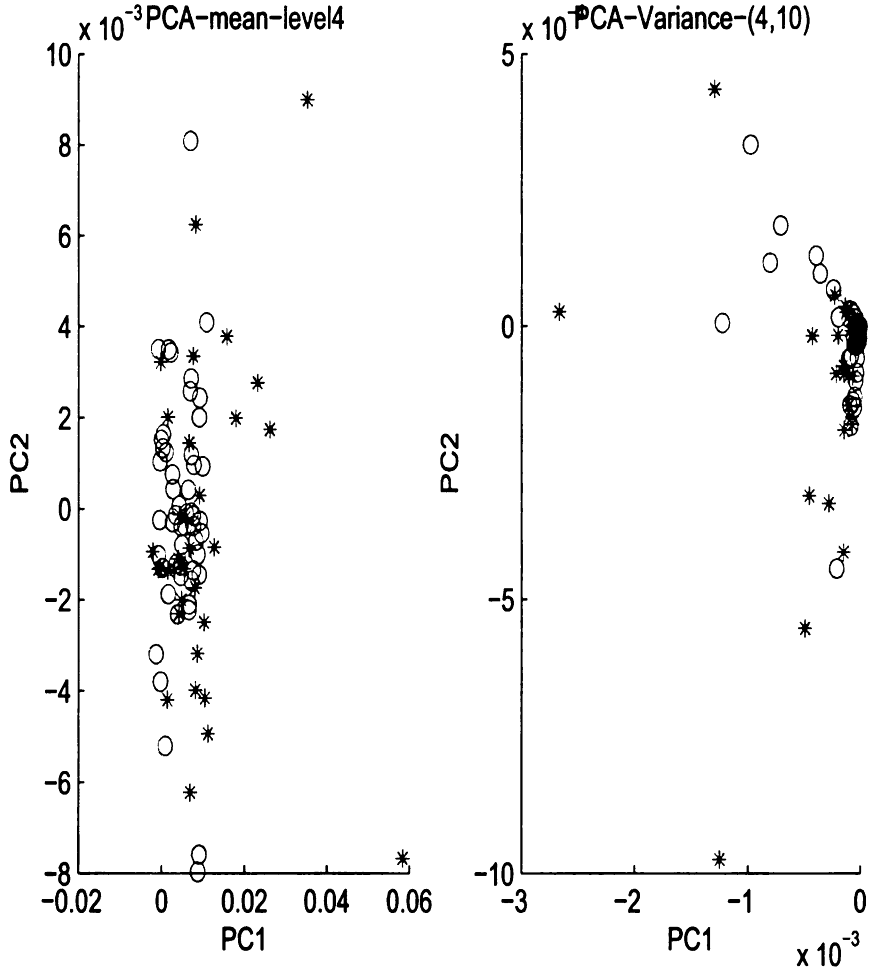

Figure 3.15. Clustering of different classes at subband(4,10).

Table 3.1. Details of the frames in the estimated transient cluster by EM at subband(1,0)

Frames	Transient(Total # of Frames=52)
Correctly classified	43
Unclassified	9
False alarm	23

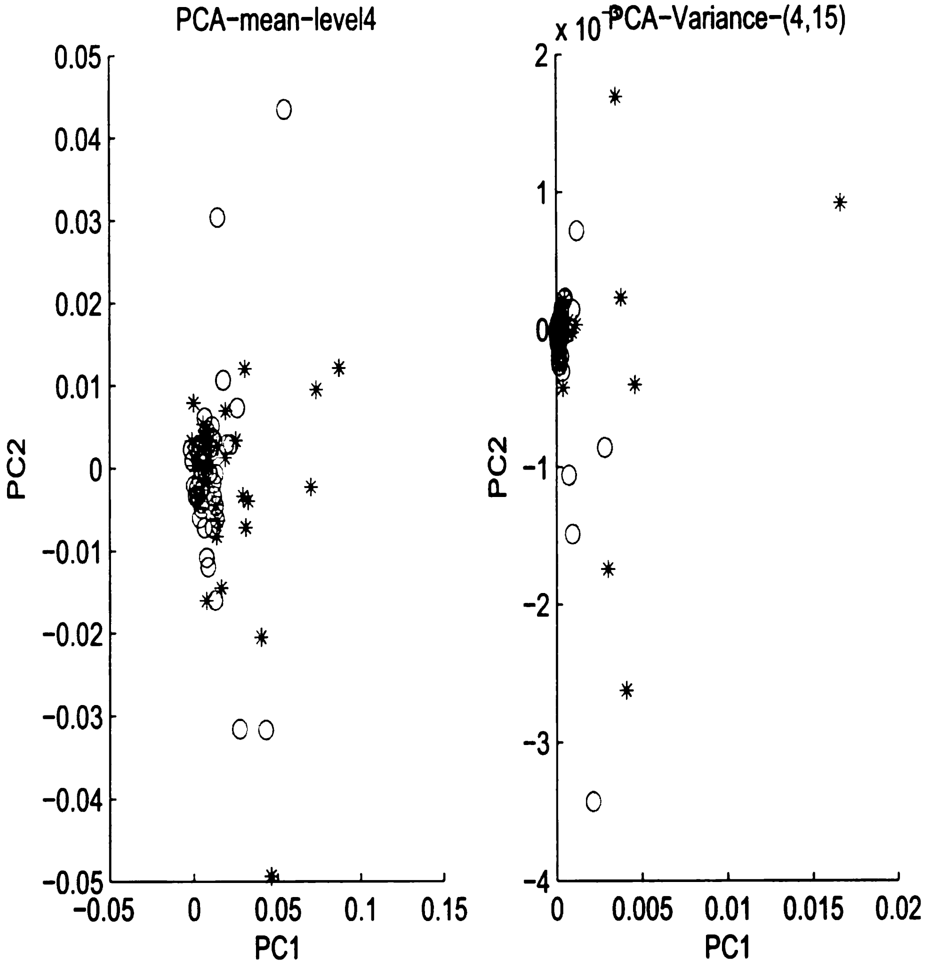

Figure 3.16. Clustering of different classes at subband(4,15).

Table 3.2. Details of the frames in the estimated steady-state cluster by EM at subband(4,13)

Frames	Steady-state(Total # of Frames=30)
Correctly classified	29
Unclassified	1
False alarm	16

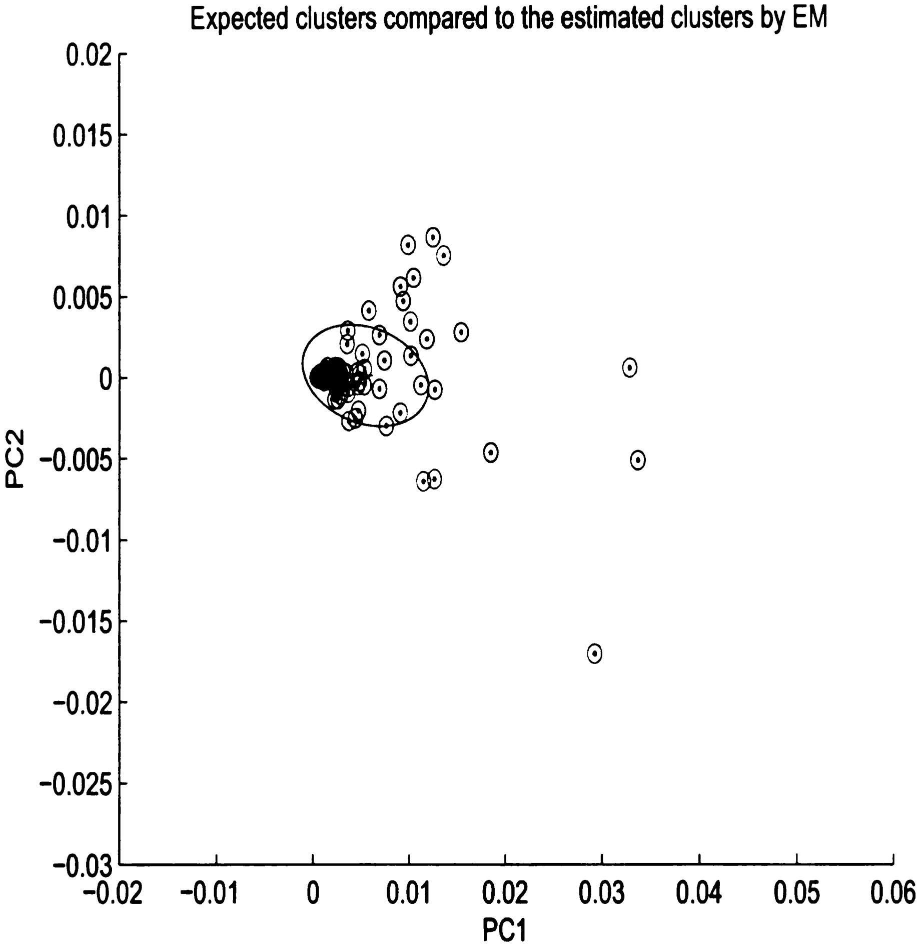

Figure 3.17. Comparison between the expected clusters and the estimated clusters using EM.

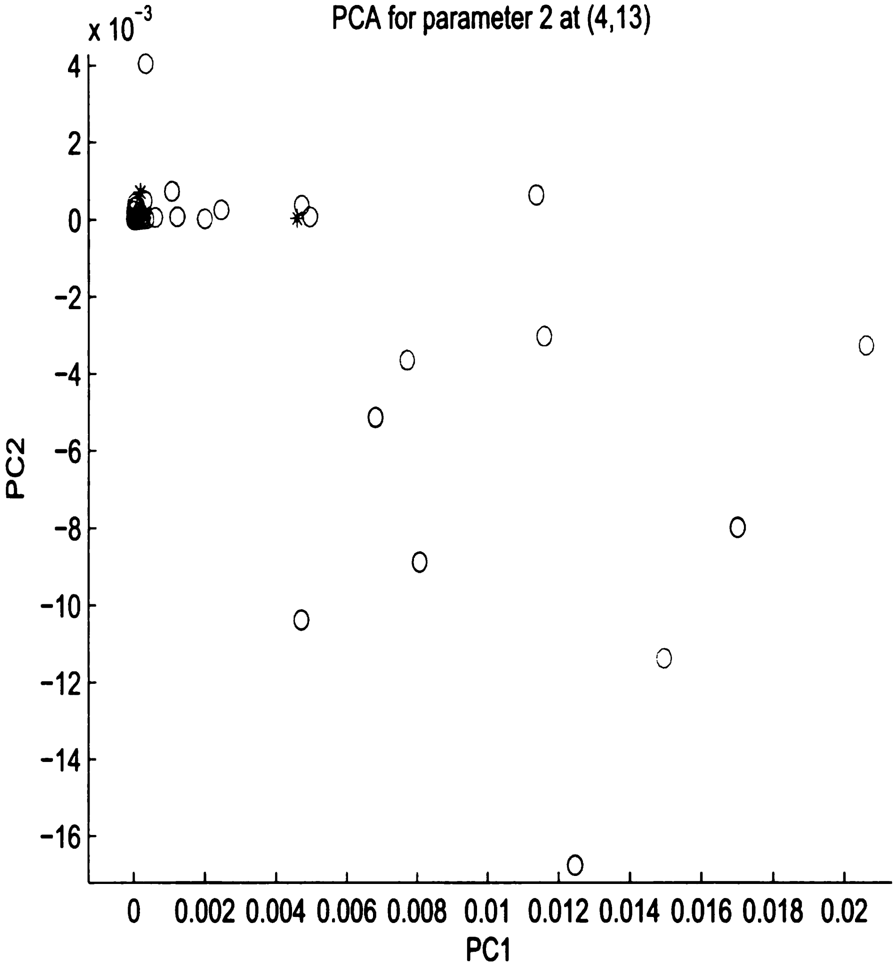

Figure 3.18. PCA for the transient and steady-state parts of the words at subband (4, 13) and for the test frames.

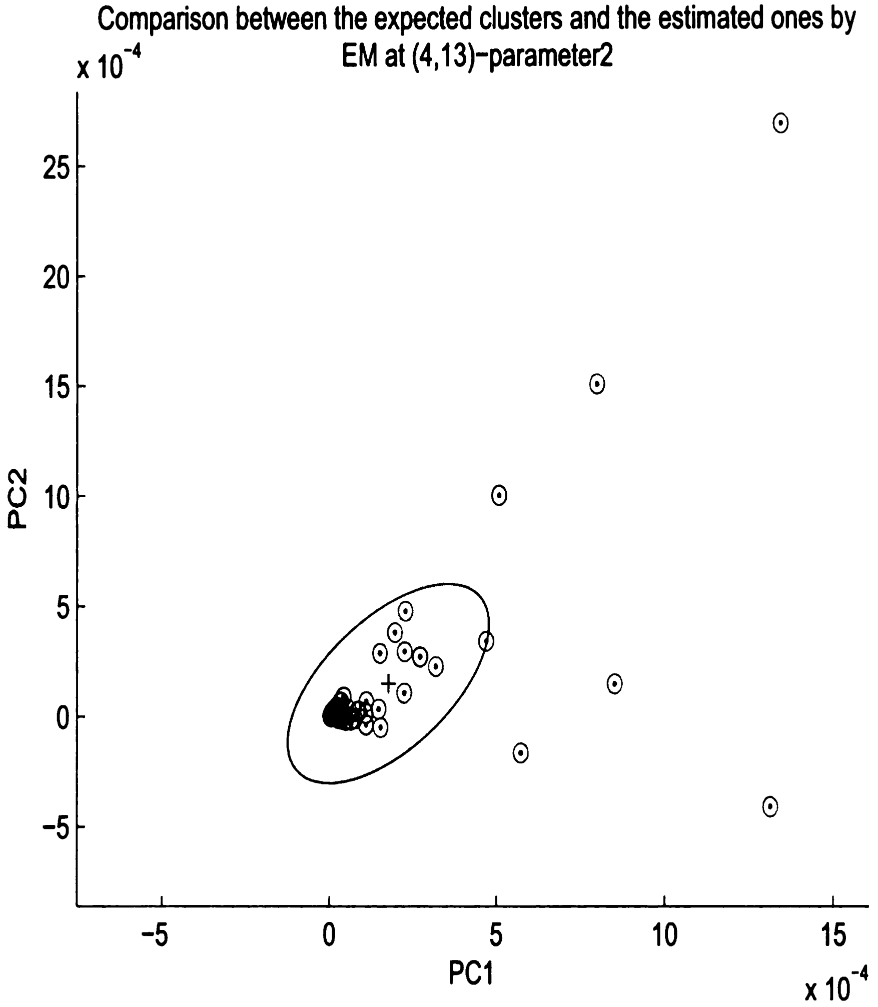

Figure 3.19. Comparison between the expected clusters and the estimated clusters using *EM* at (4,13) for parameter 2.

3.3.1 The effect of using the GMM model with the PCA

In this part of the chapter we are trying to investigate the effect of the GMM that have been used to fit the probability distribution of the wavelet coefficients and the effect of the principal component analysis.

In the first test we will try to see how parameterizing the wavelet coefficients using GMM affects the method. After finding the $DWPT$ for the framed signal, the mean and the variance of the wavelet coefficients are computed and used to check the clustering of these factors in different subbands. Figure 3.20 shows the clustering of the mean and the variance of the wavelet coefficients for the different classes, while Figure 3.21 shows the clustering of the parameters that is estimated using GMM in combination with the PCA. The separability factor related to every figure is shown also.

The second test tries to figure out the effect of the PCA of the GMM parameters. Figure 3.22 shows the clustering of the different classes parameters that were estimated by the GMM without doing the PCA step, whereas, Figure 3.23 shows the clustering of different classes using GMM in combination with PCA.

From these results we can conclude that our algorithm, that parameterize the wavelet coefficients using GMM model and then uses PCA to analyze the signals, gives the best separability between the different classes.

Moreover, Figure 3.24 shows the relationship between the sample size and the separability factor ρ under different signal to noise ratios. It was found that there isn't a clear pattern between ρ and the sample size, which means the relationship is inconsistent.

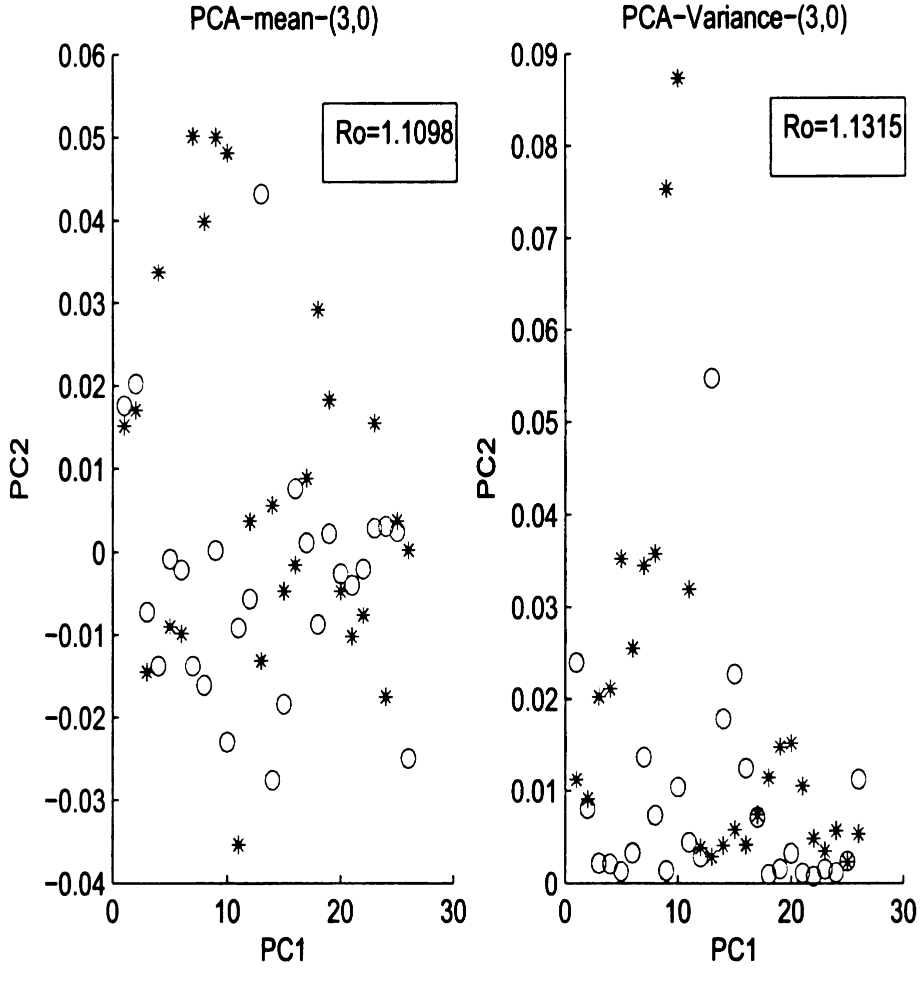

Figure 3.20. Clustering of different classes mean and variance of the wavelet coefficients at subband (3,0).

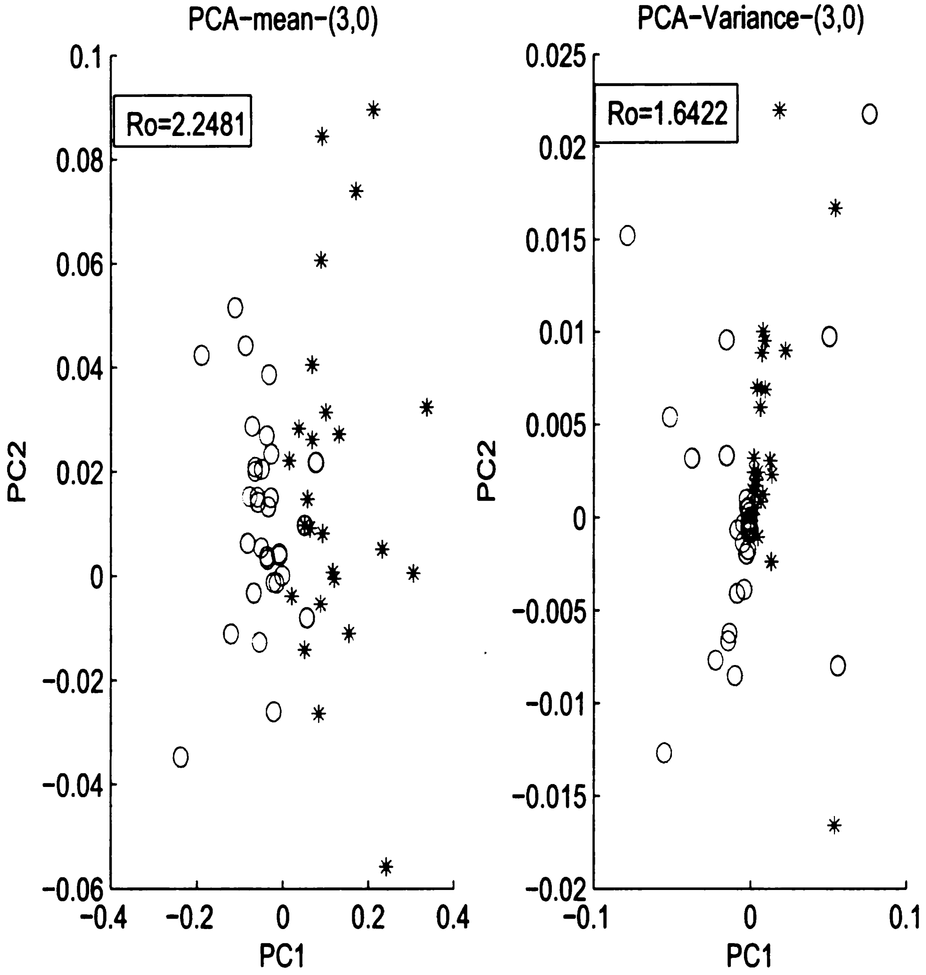

Figure 3.21. Clustering of different classes parameters that is estimated by the *GMM* at subband (3,0).

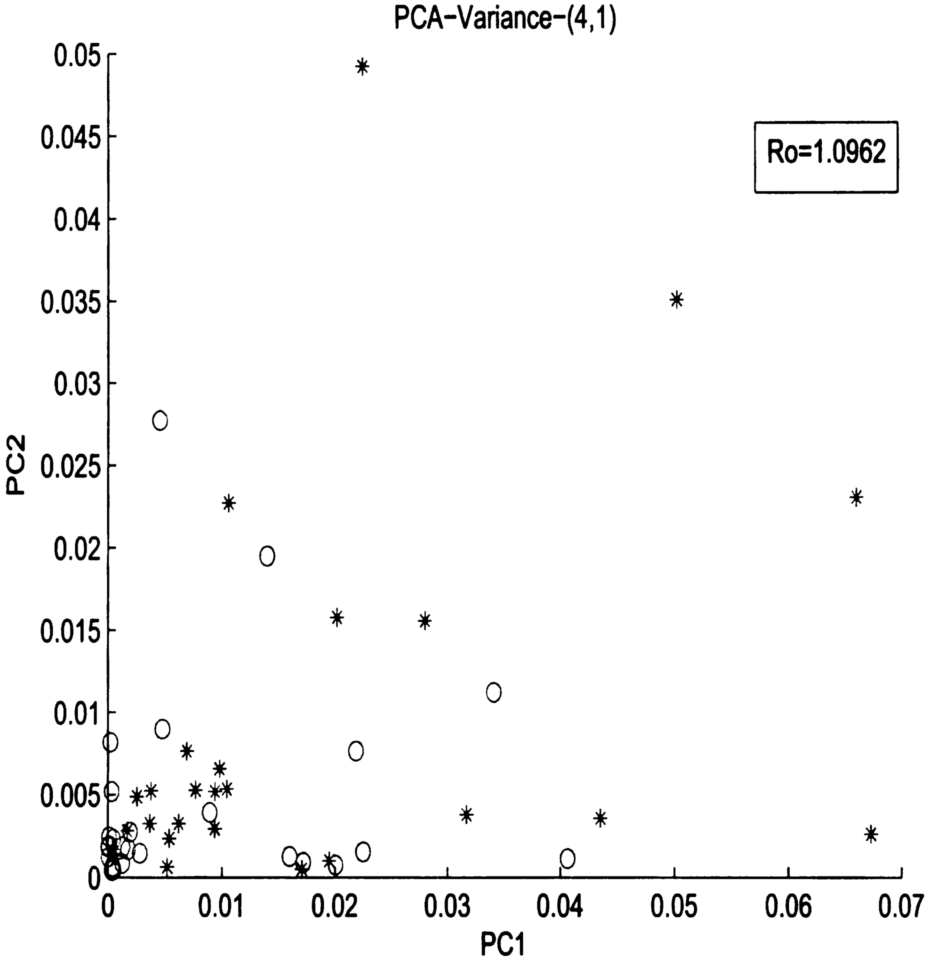

Figure 3.22. Clustering of different classes parameters found using GMM without doing PCA at subband (4,1).

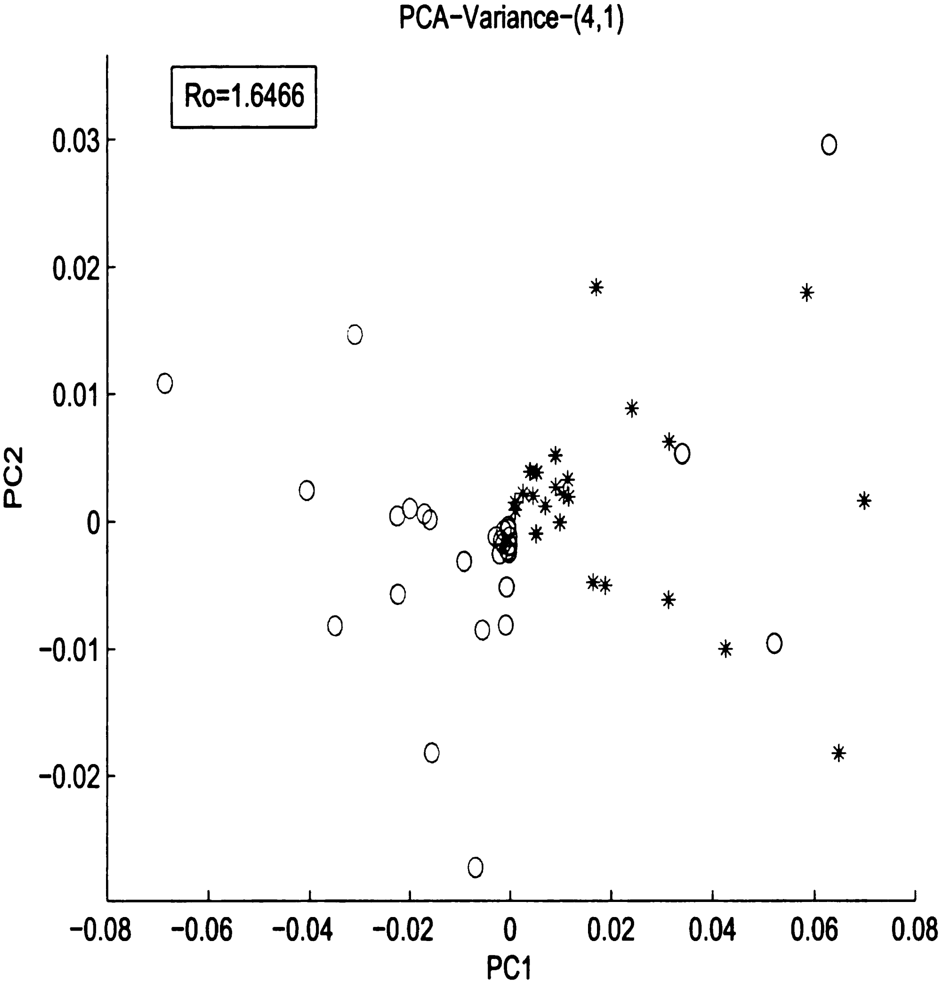

Figure 3.23. Clustering of different classes parameters that is estimated by the GMM using PCA at subband (4,1).

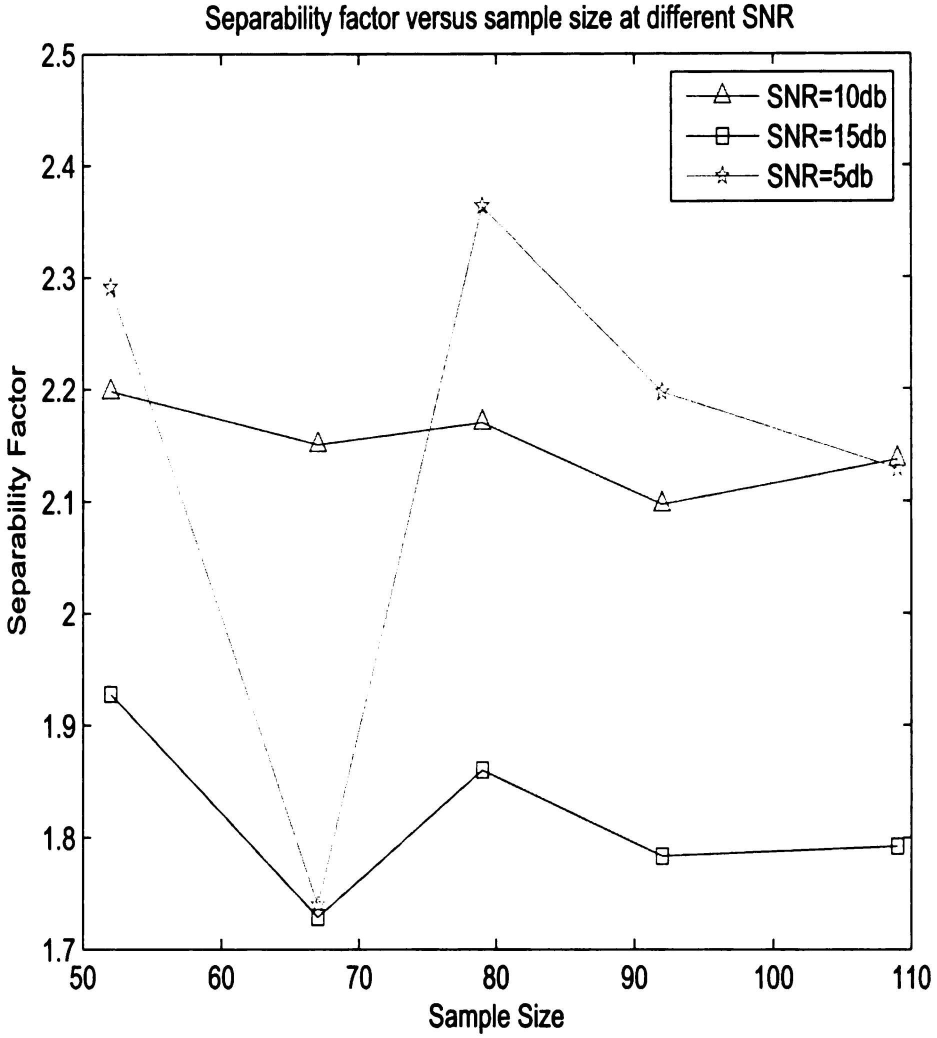

Figure 3.24. relationship between the separability factor and the sample size under different SNR.

3.4 Summary and Discussion

A Gaussian mixture model has been used to characterize the probability distribution of the wavelet coefficients. A feature set was obtained for classification and clustering. The classification problem relies on the prior knowledge of the principal eigne vectors of the samples from each class, while the clustering problem relies only on a separability measure in the feature space between individual classes. Expectation maximization algorithm has been used for cluster analysis, where the subbands that show the best classification also were best suited for cluster analysis using EM. The proposed method showed preliminary results on how sparse representation can be useful in capturing the eminent features of the speech signals that indicates the transient periods of the spoken words in a noisy environment. This can help the cochlear implant patients to better communicate under adverse conditions.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

4.1 Summary of the Dissertation

In this thesis, two algorithms based on the sparse representation of the speech signals have been introduced. Both methods rely on sparse representation, implemented using discrete wavelet packet decomposition of the speech signals, due to its excellent ability in capturing transient signals. The Power method uses this representation in combination with second order statistics, while the parametric method uses a Gaussian Mixture Model to characterize the distribution of the sparse representation of the speech signal. In the Power method, it has been shown that the method is affected by the dominant signals in the surrounding environment that may not necessarily be the ones of interest. The Gaussian mixture model characterizes the probability distribution of the wavelet coefficients and therefore uses higher order statistics. Using this parametrization we were able to get a feature set that can be used in classification and clustering. Implementing this method showed that there are some subbands that better separate the different classes. In order to use this result to improve cochlear implant patients performance in a noisy environment, these subbands can be given higher weights than the rest of the subbands to activate the implanted electrodes during the transient periods of the spoken words.

4.2 Future Work

The proposed Parametrization method assumes that the wavelet coefficients can be parameterized using a Gaussian mixture model of two Gaussians, and this may not be always the case. Multiple Gaussians can be used for the parametrization step. The technique that have been used to separate the different classes is the expectation maximization technique. However, there are many different techniques that can be used in cluster analysis and more experiments should be carried out to see which technique performs better.

The proposed technique uses the sparse representation of the speech signal through the wavelet packets. Other representations of the speech signal could be used if they reveal new characteristics of the speech signal that can help in detecting the start and the end of the words. The algorithm needs to be implemented and tested by cochlear implant patients to compare the performance to the classical filter-bank techniques under adverse conditions.

BIBLIOGRAPHY

[1] W. A. Yost, *Fundamentals of Hearings an Introduction*, Elsevier, 2007.

[2] R. Hinojosa and M. Marion, *Histopathology of Profound Sensorineural Deafness*, Ann. New York Acd. of Sci, 1983.

[3] J. Deller, J. Hansen, and J. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE press, 2000.

[4] G. Borden, K. Harris, and L. Raphael, *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*, Baltimore, MD: Williams and Ailkins, 1994.

[5] F. Cooper, P. Delattre, A. Liberman, J. Bost, and L. Gerstman, "Some Experiments on the Perception of Synthetic Speech Sounds," *Acoust. Soc. Amer.*, vol. 24, pp. 597-606, November 1952.

[6] S. B. Waltzman, J. Thomas Roland, *Cochlear Implants*, Thieme Medical Publishers, 2006.

[7] B. Wilson, *Signal Processing in Cochlear Implants:Audilogical Foundations*, Singular Publishing Group, 1993.

[8] P. C. Loizou, "Mimicking the Human Ear," *IEEE Signal Processing Magazine*, vol. 15, pp. 101-130, September 1998.

[9] I. Hochmair-Desoyer, E. Hochmair, and K. Burian , "Design and Fabrications of Multiwire Scala Tympani Electrodes," *Annals of New York Academy of Scienses*, vol. 405, pp. 173-182, 1983.

[10] G. Clark, R. Shepherd, R. Black, and Y. Tong , "Design and Fabrications of the Banded Electrode Array," *Annals of New York Academy of Scienses*, vol. 405, pp. 191-201, 1983.

[11] P. C. Loizou, "Introduction to Cochlear Implants," *IEEE Engineering in Medicine and Biology*, vol. 18, pp. 32-42, January/February 1999.

[12] P. C. Loizou, "Signal-processing Techniques for Cochlear Implants," *IEEE Engineering in Medicine and Biology*, pp. 34-46, May/June 1999.

[13] M. Dorman, M. Hannley, K. Dankowski, L. Smith, and G. McCandlless, "Word Recognition by 50 Patients Fitted with the Symbion Multichannel Cochlear Implant," *Ear and Hearing*, vol. 10, pp. 44-49, 1989.

[14] M. White, M. Merzenich, and J. Gardi, "Multichannel Cochlear Implants: Channel Interactions and Processor Design," *Archives of Otolaryngology*, vol. 110, pp. 493-501, 1984.

[15] B. Wilson, C. Finley, D. Lawson, R. Wolford, D. Eddington, and W.Rabinowitz, "Better Speech Recognition with Cochlear Implants," *Nature*, vol. 352, pp. 236-238, July 1991.

[16] B. Wilson, D. Lawson, and M. Zerbi, "Advances in Coding Strategies for Cochlear Implants," *Advances in Otolaryngology-Head and Neck Surgery*, vol. 9, pp. 105-129, 1995.

[17] M. Dorman, and P. Loizou, "Changes in Speech intelligibility as a Function of Time and Signal Processing Strategy for an Interaid Patient Fitted With Continuous Interleaved Sampling(CIS)processors," *Ear and Hearing*, vol. 18, pp. 147-155, 1997.

[18] G. Clark, "The University of Melbourne-Nucleus Multi-electrode Cochlear Implant," *Advances in Oto-Rhino-Laryngology*, vol. 38, pp. 1-189, 1987.

[19] P. Seligman, J. Patrick, Y. Tong, G. Clark, R. Dowell, and P. Crosby, "A Signal Processor for a Multiple-Electrode Hearing Prothesis," *Acta Otolaryngologica*, pp. 135-139, 1984.

[20] P. Blamey, R. Dowell, and G. Clark, "Acoustic Parameters Measured by a Formant-Estimating Speech Processor for a Multiple Channel Cochlear Implant," *Journal of the Acoustical Society of America*, vol. 82, pp. 38-47, 1987.

[21] R. Dowell, P. Seligman, P. Blamey, , and G. Clark, "Evaluation of a Two-Formant Speech Procesing Strategy for a Multichannel Cochlear Prothesis," *Annals of Otology, Rhinology and laryngology*, vol. 96, pp. 132-134, 1987.

[22] J. Patrick, and G. Clark, "The Nucleus 22-Channel Cochlear Implant Syastem," *gy, Ear and Hearing*, vol. 12, pp. 3-9, 1991.

[23] N. Tye-Murray, M. Lowder, and R. Tyler, "Comparison of the F_0/F_2 and $F_0/F_1/F_2$ Processing Strategies for the Cochlear Corporation Cochlear Implant ," *gy, Ear and Hearing*, vol. 11, pp. 195-200, 1990.

[24] k. Nie, N. Lan, and S. Gao, "Implementation of CIS Speech Processing Strategy For Cochlear Implants by Using Wavelet Transform," *Proceedings of ICSP*, pp. 1395-1398, 1998.

[25] J. Yao, and Y. Zhang, "The Application of Bionic Wavelet Transform to Speech Signal Processing in Cochlear Implants Using Neural Network Simulations," *IEEE Transactions on Biomedival Engineering*, vol. 49, pp. 1299-1309, November 2002.

[26] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech Recognition in Noise as a Function of the Number of Spectral Channels: Comparison of Acoustic hearing and Cochlear implants," *Acoust. Soc. Amer.*, vol. 110, pp. 1150-1163, August 2001.

[27] J. C. Middlebrooks, "Effects of Cochlear Implant Pulse Rate and Inter-Channel Timing on Channel Interaction and Thresholds," *Acoust. Soc. Amer.*, vol. 116, pp. 452-468, July 2004.

[28] J. Maj, L. Royackers, M. Moonen, and J. Wouters , "SVD-Based Optimal Filtering for Noise Reduction in Dual Microphone Hearing Aids: A Real Time Implementation and Perceptual Evaluation," *IEEE Transactions on Biomedival Engineering*, vol. 52, pp. 1563-1573, September 2005.

[29] D. K. Eddington, W. H. Dobelle, D. E. Brackmann, M. G. Mladevosky, and J. L. Parkin, "Auditory Prothesis Research with Multiple Channel Intracochlear Stimulation in Man ," *Ann. Otol. Rhinol. Laryngol.*, vol. 87, pp. 1-39, 1978.

[30] K. Nie, G. Stickney, and F. Zeng , "Encoding Frequency Modulation to Improve Cochlear Implant Performance in Noise," *IEEE Transactions on Biomedival Engineering*, vol. 52, pp. 64-73, January 2005.

[31] N. Lan, K. B. Nie, S. K. Gao, and F. G. Zeng , "A Novel Speech-Processing Strategy Incorporating Tonal Information for Cochlear Implants," *IEEE Transactions on Biomedival Engineering*, vol. 51, pp. 752-760, May 2004.

[32] Y. Suhail, and K. Owiess , "Augmenting Information Channels in Hearing Aids and Cochlear Implants Under Sever Conditions," *ICASSP Conference*, pp. 889-892, 2006.

[33] B. S. Wilson, D. T. Lawson, J. M. Muller, R. S. Tyler, and J. Kiefer , "Cochlear Implants: Some Likely Next Steps," *Annual Reviews Biomedical Engineering*, vol. 5, pp. 207-249, 2003.

[34] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[35] K. Baker , "Singular Value Decomposition Tutorial," , 2005

[36] http://en.wikipedia.org/wiki/Gaussianmixturemodel

[37] www.utdallas.edu/ loizou/speech/noizeus

[38] J. Shlens , "A Tutorial on Principal Component Analysis," , December 2005.

[39] J. A. Bilmes , "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," , April 1998.