DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

NEW SPEECH PROCESSING STRATEGIES BASED ON WAVELET PACKET TRANSFORM IN COCHLEAR IMPLANTS

by

Yahya ÖZTÜRK

September, 2009

İZMİR

NEW SPEECH PROCESSING STRATEGIES BASED ON WAVELET PACKET TRANSFORM IN COCHLEAR IMPLANTS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in
Electrical and Electronics Engineering Program

by
Yahya ÖZTÜRK

September, 2009
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **NEW SPEECH PROCESSING STRATEGIES BASED ON WAVELET PACKET TRANSFORM IN COCHLEAR IMPLANTS** completed by **YAHYA ÖZTÜRK** under supervision of **ASST. PROF. DR. GÜLDEN KÖKTÜRK** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

…………………………………………

Asst. Prof. Dr. Gülden KÖKTÜRK

Supervisor

……………………………………..

Asst. Prof. Dr. Nalan ÖZKURT

(Jury Member)

…………………………………….

Asst. Prof. Dr. Barış BOZKURT

(Jury Member)

Prof. Dr. Cahit HELVACI
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor Assoc. Prof. Dr. Gülden KÖKTÜRK for her guidance and support in every stage of my research. The technique background and the research experience I have gained under her care will be valuable asset to me in the future.

I am grateful Mehmet Akif Kılıç regarding to prepare balance words lists that were used in the intelligibility experiments.

I would like to thank Deniz Başkent for cochlear implant N of M speech strategy simulation MATLAB codes.

Finally, I am grateful to my wife and parents for their patience and never ending support throughout my life.

Yahya ÖZTÜRK

NEW SPEECH PROCESSING STRATEGIES BASED ON WAVELET PACKET TRANSFORM IN COCHLEAR IMPLANTS

ABSTRACT

Cochlear implants (CI) improve partial hearing to profoundly deaf people. Many investigators from various disciplines made combined efforts for progression on these implants. The speech processing strategy in modern CI's extracts and encodes amplitude information in a number of frequency bands. In thesis study, we proposed an approach to improve the performance of speech enhancement techniques based on wavelet packet (WP) algorithm. This algorithm has better results on speech intelligibility than other existing algorithm and this result has been proved by the intelligibility experiments. The WP algorithm was modified to effectiveness of the strategies and then an entropy based modification was applied for electrode selection, thus this modification increases noise resistance of the new speech processing algorithm that proposed in thesis study.

Keywords: Cochlear implant, wavelet transform, wavelet packet transform, entropy

PARÇACIK PAKET DÖNÜŞÜMÜ BAZLI YENİ KOKLEAR IMPLANT STRATEJİSİ

ÖZ

Koklear implant sağırlık derecesindeki duyma kaybı olan insanların duyma seviyesini artırmaktadır. Farklı disiplinlerdeki çok sayıda araştırmacı kullanılan bu implant üzerinde çalışmalar yapmaktadır. Implant içerisinde kullanılan ses işleme algoritmaları genel olarak farklı frekans bandlarındaki sinyal gücünü açığa çıkarmak ve bunları kodlamak sureti ile çalışır. Bu çalışmamızda, daha iyi ses işleme kapasitesine sahip ve Parçacık Paket Dönüşümü bazlı yeni bir model ve yaklaşım öneriyoruz. Önerdiğimiz algoritma, ses anlaşırlığını mevcut algoritmalara göre artırmıştır ve bu sonuç yapılan anlaşılma deneyleri ile kanıtlanmıştır. Ayrıca elektrot seçiminde Parçacık Paket Dönüşümü entropi yaklaşımı kullanılmış ve bu sayede algoritmanın gürültüye karşı dayanımı artırılmıştır.

Anahtar Sözcükler: Koklear Implant, parçacık dönüşümü,parçacık paket dönüşümü, entropi

CONTENTS

CHAPTER ONE

INTRODUCTION

A particular percentage of the populations in developed countries encounter hearing impairment. Cochlear Implant (CI) has been developed to increase the hearing capacity for these people. In recent years, adults and children have benefited by usage of CI and they affected from improvement of implant techniques as well. Although these devices permit increased performance, a significant gap in speech recognition still remain between CI listener and people which possess normal listening capability.

The CI prosthesis is an electronic device intended to directly stimulate the auditory nerve in deaf people who have lost the receptor cells in the cochlea (Wilson B., 1993). The clinical research for these devices began in the mid 1960's by most researchers and the prosthesis would assist a limited number of patients accomplish mitigate levels of hearing rehabilitation. Key developments have been achieved in the implanted stimulating system, signal processing strategies, and patient fitting techniques (G. Loeb, 1990). Continued development in these areas, especially signal processing strategies, may produce near complete restoration of hearing in a large number of patients.

In most deaf people the auditory transducers have been destroyed. The networks of neural connections between the cochlea and the brain have significant functional capacity. Multichannel cochlear implants have an important role for damaged hair cells by activating the remaining frequency-specific neural pathways in the cochlea and central auditory system (J. Millar, Y. Tong & G. Clark, 1984).

CI system often consists of the following modules: a microphone, a speech processor, a transmitter, a receiver and an electrode array as shown Figure 1.1 and Figure 1.2 (C. Parkins & S. Anderson 1983).

The fundamental part of CI is speech processor which provides acceptable stimulation parameters. The characterization of cochlea can be modeled with the assistance of time scale analysis of wavelets. Therefore, this study investigates a new wavelet based method to apply extraction of these features and proposes to improve the interface between the stimulating electrode arrays for N-of-M strategy in cochlear implants.

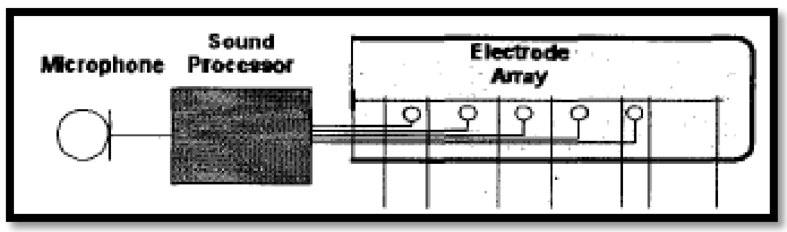

Figure 1.1 Block diagram of Cochlear implant

(Loizou Phillip, 1998)

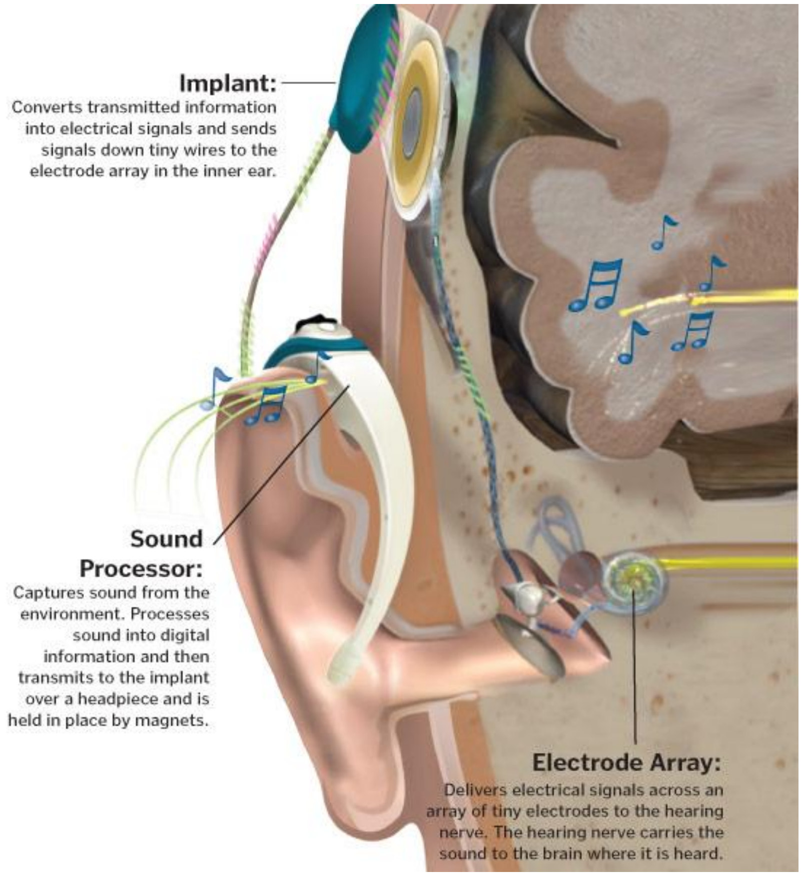

Figure 1.2 Detail view of Cochlear Implant

(Illustration Courtesy of Advanced Bionics, LLC
Graphic: The Washington Post - April 13, 2008)

Since William House developed the first single channel implant, it responds to coarse temporal fluctuations as much as frequency characteristic (W. House & J. Urban, 1973; W. House & K. Berliner, 1982). Furthermore, speech recognition was restricted to transmitted frequency information and it was inadequate in comprehensibility. When multi-channel implants were introduced in the 1980s, several questions were raised regarding multi channel stimulation. Most important question was: “What kind of information should be transmitted to each electrode?” Depending on how researchers tried to address these questions, different types of signal processing techniques were developed. The various signal processing strategies developed for multi-channel cochlear prosthesis, can be divided into three categories:

waveform strategies, feature-extraction strategies and “N-of-M” strategies (B. Wilson, 1993). These strategies differ in the way that information, is extracted from the speech signal and presented to the electrodes.

Speech strategies play an extraordinarily important role to maximize the complete communicative potential of user. In addition, different strategies developed over the past two decades intend to improve intelligibility of deaf people as naturally as possible. N-of-M strategy divides the speech signal into M subbands and extracts the envelope information from each band of signal. N bands which have the largest amplitude are then selected for stimulation (N out of M) (W. Nogueira, A. Giese, B. Edler & A. Buchner, 2006).

In this study, the proposed method is different from traditional N of M speech strategies. It selects active electrodes by using wavelet entropy changes which are determined best tree function on wavelet packet (WP) theory.

In the literature, there have been various reported studies but there is still significant research to be done investigation on wavelet packet transform for speech processing applications. A generalization type of the discrete wavelet transform (DWT) called as WP analysis enables subband analysis without the constraint of dyadic decomposition. Basically, the discrete WP transform performs an adaptive decomposition in frequency axis. This particular discrimination may be doned with optimization criterions (L. Brechet, M.F. Lucas at all. 2007).

A new strategy based on the wavelet transform in speech processing might enhance the exactness of the cochlear implant in coding speech features. The wavelet transform which provides good resolution both in time and frequency is most suitable tool to analyze non-stationary signals such as speech signals. Moreover, the power of the wavelet transform in analyzing speech strategies of CIs is the fact that the cochlea seems to be behaving as parallel with the wavelet transform filter banks.

The wavelet theory guarantees a unified framework for various signal processing applications such as signal and image denoising, compression, analysis of non-stationary signal, etc. In speech processing applications, the wavelet transform has been intended to improve the speech enhancement quality of classical methods. The suggested method in this work is tested on recorded noisy speech from real environments.

WPs were first investigated by Coifman and Meyer as a orthogonal bases for L2(R). Realization of a desired signal with a best basis selection method involves the introduction of an adequate cost function which provides energy localization to a decrising operation (R.R.Coifman & M.V. Wickerhauser, 1992). The cost function selection is directly related to the fixed structure of the application. Consequently if signal compression, identification or classifications are the interests as an application, entropy may reveal desired basis functions. Then, the statistical analysis of coefficients taking from these basis functions may be used indicating the original signal. Therefore, the WP analysis is effective to the signal localization in time and frequency.

1.1 Main Contribution

This thesis study will give detailed cochlear implant information, cochlear implant companies, speech processing strategies and especially a description of the N-of-M strategy and the basis of its development in the chapter two. This section will helps to you understand cochlear implant concept and its details. Then the chapter three will cover wavelet transformation, thresholding, wavelet packet (WP) algorithm, best tree and entropy of the cochlear implant to auditory models. It is core section for literatures study because this study lies on new speech processing approach as wavelet packet transform and wavelet entropy.

This is followed by literatures study which is based on the new structure of the electrode selection and a more detailed characterization of this new speech processing method in the chapter four. Chapter five has results that are based on human experiment of intelligibility and signal processing simulations on sample speech signals.

Finally, conclusion chapter which is chapter six will cover advantages and disadvantages of this work, significant points in this study and future works that can be handling with another study.

CHAPTER TWO

COCHLEAR IMPLANT

2.1 What is a Cochlear Implant

A cochlear implant (CI) is a surgically implanted electronic device that provides a sense of sound to a person who is profoundly deaf or severely hard of hearing. The cochlear implant is often referred to as a bionic ear (Cochlear Implant, 2009).

The implant consists of an external portion that sits behind the ear and a second portion that is surgically placed under the skin as shown Figure 2.1 (National Institute on Deafness Other Communication Disorders, 2009). An implant has the following parts:

- A microphone, which picks up sound from the environment.
- A speech processor, which selects and arranges sounds picked up by the microphone.
- A transmitter and receiver/stimulator, which receive signals from the speech processor and convert them into electric impulses.
- An electrode array, which is a group of electrodes that collects the impulses from the stimulator and sends them to different regions of the auditory nerve.

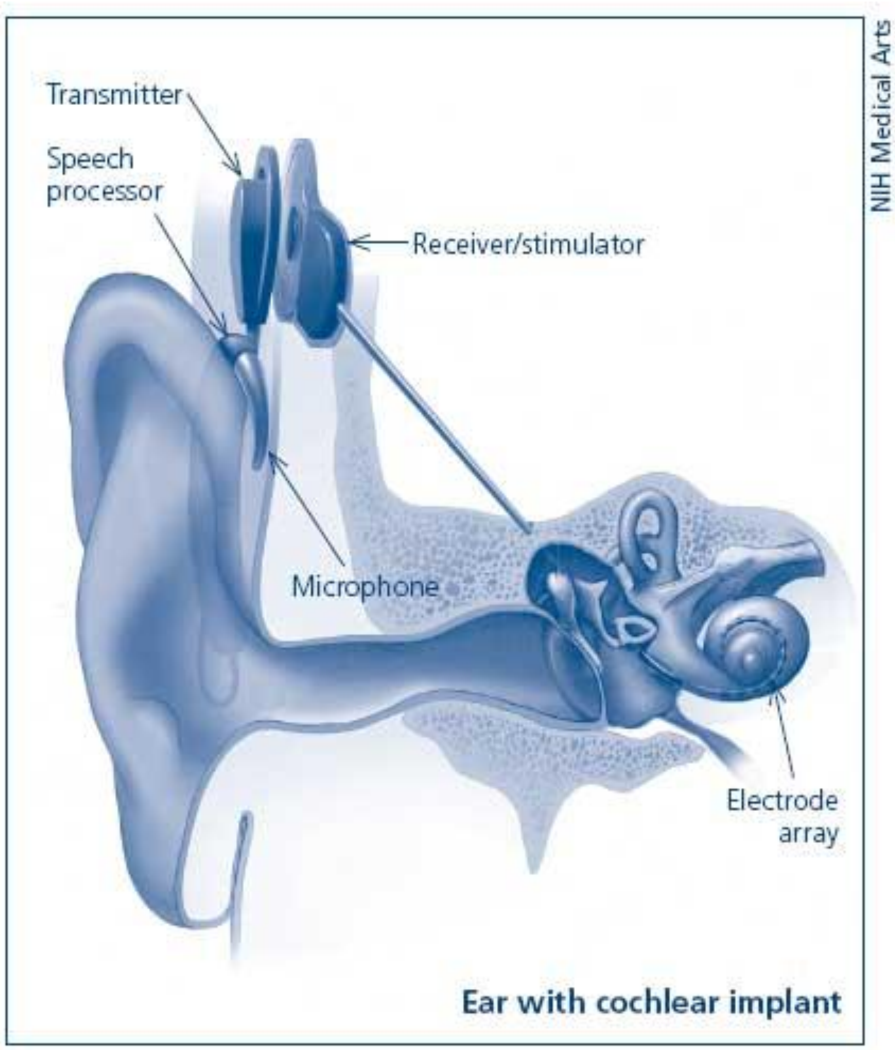

Figure 2.1 Cochlear Implant

(Medical illustrations by NIH, Medical Arts & Photography Branch)

An implant does not restore normal hearing. Instead, it can give a deaf person a useful representation of sounds in the environment and help him or her to understand speech.

2.2 Single Channel Implants

Single-channel implants provide electrical stimulation in the cochlea using a single electrode. These implants are simple in design and their cost lower than multi-channel implants. They are also preferred because they do not require much hardware and conceivably all the electronics could be packaged into a behind-the-ear device.

Single-channel implants were first implanted in human subjects in the early 1970s. At the time, there was a lot of skepticism about whether single-channel stimulation could really work (W. House, 1985). These early efforts led to, among other devices, the House/3M single-channel implant and the Vienna/3M single-channel implant (Loizou Phillip, 1998)

2.3 Multi Channel Implants

Multi-channel implants provide electrical stimulation in the cochlea using an array of electrodes. An electrode array is used so that different auditory nerve fibers can be stimulated at different places in the cochlea, thereby exploiting the place mechanism for coding frequencies. Electrodes are responsible for each the frequency of the signal. Electrodes near the base of the cochlea are stimulated with high frequency signals, while electrodes near the apex are stimulated with low frequency signals.

When multi-channel implants were developed, researchers faced several questions regarding multi-channel stimulation:

1. How many electrodes should be used? If one channel of stimulation is not sufficient for speech perception, then how many channels are needed to obtain high levels of speech understanding?
2. Since more than one electrode will be stimulated, what kind of information should be transmitted to each electrode? Should it be some type of spectral feature or attribute of the speech signal that is known to be important for speech perception (e.g.,

first and second formants). or some type of waveform derived by filtering the original speech signal into several frequency bands?

Researchers experimented with different number of electrodes. Some devices used a large number of electrodes (22) but only stimulated a few, while other devices used a few electrodes (4-8) and stimulated all of them. The answer to the question on how many channels are needed to obtain high levels of speech understanding is still the subject of argument (R. Shannon, F. Zeng, V. Kamath, J. Wygonski & M. Ekelid, 1995; M. Dorman, P. Loizou, & D. Rainey, 1997).

The various signal processing strategies developed for multi-channel cochlear can be collected into two main categories:

1. Waveform strategies
2. Feature-extraction strategies

These strategies extract the speech information from the speech signal and present to the electrodes. The waveform strategies use some type of waveform (in analog or pulsatile form) derived by filtering the speech signal into different frequency bands. The feature extraction strategies use some type of spectral features, such as formants, derived using feature extraction algorithms.

2.4 Cochlear Implant Companies

There are several different manufacturers of cochlear implants that have been approved by the FDA for use in the United States and Turkey (Ashley Nicole Norkus, 2007). These are 3M/House Cochlear Implant, Advanced Bionics, Med El Corporation and Cochlear Corporation (Chute, P.M., Nevins & M.E., 2002; Christiansen, J.B, Leigh & I. W., 2002).

3M/House CI were a single channel cochlear implant and were the first approved by the FDA for use in postlingually deaf adults. Advanced Bionics is located in

California and started producing implants in 1995. They have been through several generations of cochlear implants including both internal and external devices. Med El Corporation is based out of Australia and they have been producing cochlear implants since the early 1980s. Cochlear is located in Australia and they were the first to produce multi channel cochlear implants in the world in the early 1980s.

2.5 External Components

The microphone, speech processor, transmitter and power supply are all parts of the external devices of the cochlear implant (Moore, J.A., Teagle & H.F.B., 2002; Ashley Nicole Norkus 2007; Nevins, M.E., Chute & P. M. 1996).

Batteries are the power supply for the cochlear implant. They can be either rechargeable or alkaline depending on the type of device that is used. Cables deliver the sound from the microphone to the speech processor. Coils contain magnets that hold the implant to the head and transfer the signals from the speech processor via radio waves through the skin into the internal device. Microphone picks up the incoming signals. It is important part of cochlear implant because it affects quality of speech signals. The speech processor is an electronic device that filters the input signal from the microphone and converts it into a series of electrical signals to be delivered to the internal device within the cochlea, it keeps speech processing algorithm in its memory and behind the Ear (BTE) Processor a speech processor that sits on the ear and is much smaller than the body-worn processor. Body-worn Processor is a speech processor that is worn on the belt or in a special harness and is pager sized.

2.6 Internal Components

The parts of the cochlear implant are placed under the skin with surgery operation (Chute, P.M., Nevins & M.E., 2002; Ashley Nicole Norkus 2007).

A channel is a single electrode that responsible for appropriate frequency range on cochlea. An electrode actively delivers the signal to the cochlear nerve endings, it puts into inner ear with surgery operation. Electrode positioning system guides the electrodes into the cochlea (Advanced Bionics Corporation, 2000). Internal receiver part of the implant is placed under the skin behind the ear that includes that magnet and antenna.

2.7 Speech Processing Strategies in Cochlear Implant

2.7.1 Compressed-Analog (CA) approach

The compressed-analog (CA) approach was originally used in the Ineraid device manufactured by Symbion, Inc. (Eddington, D., 1980). The signal is first compressed using an automatic gain control and then filtered into four contiguous frequency bands, with center frequencies at 0.5, 1, 2 and 3.4 kHz. The filtered waveforms go through adjustable gain controls and then sent directly through a percutaneous connection to four intracochlear electrodes. The filtered waveforms are delivered simultaneously to four electrodes in analog form. The CA approach, used in the Ineraid device, was very successful because it enabled many patients to obtain open-set speech understanding (Dorman, M., M. Hannley, K. Dankowski, L. Smith & G. McCandless, 1989).

2.7.2 Continuous Interleaved Sampling (CIS)

Researchers at the Research Triangle Institute (RTI) developed the Continuous Interleaved Sampling (CIS) approach (Wilson, B., C. Finley, D. Lawson, R. Wolford, D. Eddington & W. Rabinowitz, 1991) which addressed the channel interaction issue by using non-simultaneous, interleaved pulses. Trains of biphasic pulses are delivered to the electrodes in a non-overlapping fashion, in a way such that only one electrode is stimulated at a time. The amplitudes of the pulses are derived by extracting the envelopes of bandpassed waveforms. The signal is first pre-emphasized and passed through a bank of band pass filters (Figure 2.2). The envelopes of the filtered waveforms are then extracted by full-wave rectification and low-pass filtering. The envelope outputs are finally compressed and then used to modulate biphasic pulses. A non-linear compression function (e.g., logarithmic) is used to ensure that the envelope outputs fit the patient's dynamic range of electrically evoked hearing. The rate at which the pulses are delivered to the electrodes has been found to have a major impact on speech recognition (intelligibility). High pulse-rate stimulation typically yields better performance than low pulse rate stimulation. Comparison between the CA and CIS approach revealed higher levels of speech recognition with the CIS approach (Wilson B., C. Finley, D. Lawson, R. Wolford, D. Eddington & W. Rabinowitz, 1991).

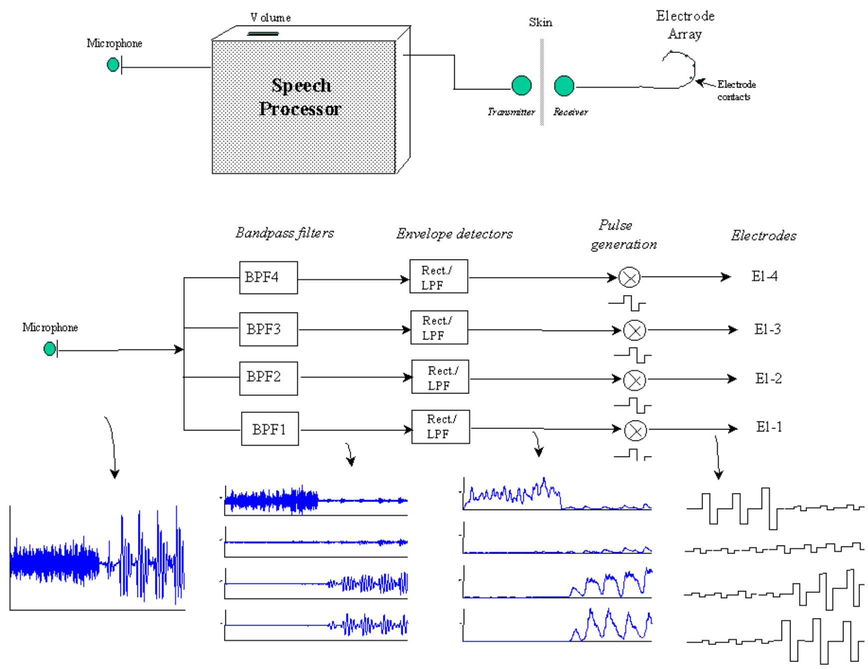

Figure 2.2 Detailed block diagram of CIS speech strategy in cochlear implant (Loizou Phillip, 1998).

2.7.3 N-Of-M Speech Processor for Cochlear Implants

In these strategies, the signal is filtered into m frequency bands, and the processor selects, out of m envelope outputs, the n (n<m) envelope outputs with the largest energy (Figure 2.3). Only the electrodes corresponding to the n selected outputs are stimulated at each cycle. For example, in a 6-of-22 strategy, from a maximum of twenty two channel outputs, only the six channel outputs with the largest amplitudes are selected for stimulation at each cycle. The "N-of-M" strategy can be considered to be a hybrid strategy in that it combines a feature representation with a waveform representation as shown Figure 2.4.

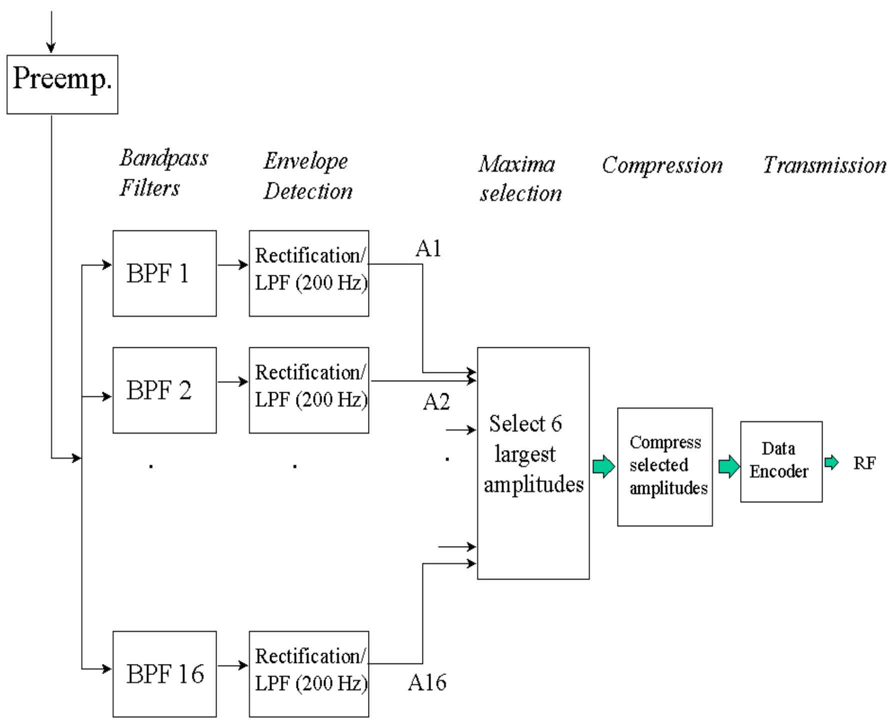

Figure 2.3 Detailed block diagram of N of M speech strategy in cochlear implant (Loizou, P., 1998).

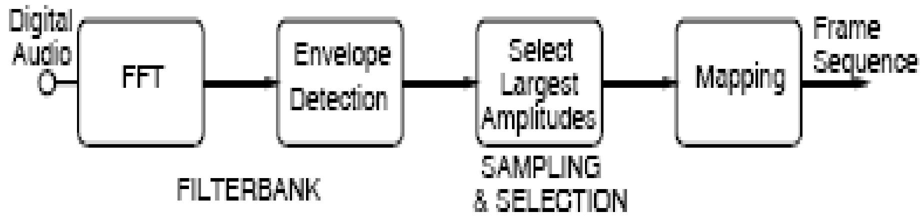

Figure 2.4 Block diagram of N of M speech strategy in cochlear implant

(Loizou, P., 1998).

CHAPTER THREE

WAVELET BASED METHODS

Many speech enhancement techniques discussed are based on the spectral information obtained through the short time Fourier transform analysis of the signal (Xiaolong Yuan, 2003). These are all frequency-based methods intending to preserve the slow-varying short time spectral characteristics of the speech such as the low-frequency harmonics of vowels, which is still not enough to maintain speech quality after the processing. We also wish the speech enhancement algorithm to preserve instantaneous properties such as the attack of the plosives (i.e., the stop consonants like b, d, g, p, t, k. that are transient, non-continuant sounds produced by building up pressure behind a total constriction somewhere along the vocal tract, and suddenly releasing this pressure (Deller, J. R., Proakis, J.G., Hansen & J.H.L., 1994). As a powerful time-frequency tool, the wavelet transform has established a reputation as a tool for signal analysis: having high frequency-resolution (and low time-resolution) for the low frequency content of the signal while having low frequency- resolution (and high time-resolution) for the high frequency content of the signal. The wavelet transform can be regarded as a bank of band-pass filters with constant Q factor (the ratio of the bandwidth and the central frequency). Through appropriate choice of a mother wavelet that both has finite effective support width in the time domain and concentrating property in the frequency domain, the wavelet analysis has a distinct ability to detect local features of the signal in both time and frequency, such as the plosive fine structures of the speech and other transient, instantaneous and dynamic speech components that contribute significantly to the quality of the speech (Quatieri ,2001). We will first introduce the basic concepts of the classic wavelet transform and its relationship to the Fourier transform.

3.1 Continuous Wavelet Transform

The Fourier transform has long been the most important underpinning for frequency-domain signal processing. The theory on wavelet transform, which originated as a branch of applied mathematics in the 1980's, was first introduced into the signal processing field thanks to the efforts of French mathematicians I. Daubechies and S. Mallat (Mallat, S., 1998; Daubechies ,1992). Today, intertwined with multi- resolution and filter bank theory, wavelets analysis plays an important role in time-frequency analysis.

3.1.1 Wavelet Introduction

The word "wavelet" literally means "a small wave". A wavelet is a function that has finite energy and zero mean. It is a powerful tool for the analysis of transient, non-stationary characteristics such as drift, trends, abrupt changes, beginning and ends of events, breakdown points, and discontinuities in higher derivatives and self-similarity (Xiaolong Yuan, 2003). We have available many kinds of wavelets: Haar, Mortlet, Daubeshies, etc.; they look different and have different properties: orthogonal, bi-orthogonal, normalized etc. For example, the Morlet wavelet is illustrated in Figure 3.1, with a solid line as its real part and a dashed line as its imaginary part.

It is a complex exponential function at frequency ω_0 with Gaussian envelope

$$\varphi(t) = e^{\frac{-t}{2}} e^{j\omega_0 t} \tag{3-1}$$

Wavelet analysis is one way to localize events in time (or space) and frequency. The goal of wavelet analysis is to create a set of basis functions (i.e., expansion functions) so that the transform will give an informative, efficient and useful description of the target signal. In a nutshell, the continuous wavelet transform (CWT) is nothing but a set of the inner products of the observed signal $f(t)$ with the shifted and scaled mother wavelets $\varphi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-\tau}{a}\right)$ where τ and a represent the time shift and scale variables.

$$\langle f(t), \varphi_{a,\tau}(t)\rangle = WT_f(a,\tau) = \frac{1}{\sqrt{a}}\int f(t)\varphi^*\left(\frac{t-\tau}{a}\right)dt \tag{3-2}$$

If $\varepsilon = \int |\varphi(t)|^2 dt$ is the energy of the basic mother wavelet, the shifted and dilated wavelets $\varphi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-\tau}{a}\right)$ maintaining the same energy due to the scaling factor$\frac{1}{\sqrt{a}}$:

$$\varepsilon' = \int \left|\frac{1}{\sqrt{a}}\varphi\left(\frac{t-\tau}{a}\right)\right|^2 dt = \frac{1}{a}\int \left|\varphi\left(\frac{t-\tau}{a}\right)\right|^2 dt = \varepsilon \tag{3-3}$$

In order to have an inverse transform, any mother wavelet chosen must satisfy the admissibility condition that means:

$$c_\varphi = \int_0^{+\infty} \frac{|\Gamma(\omega)|^2}{\omega} d\omega < +\infty \tag{3-4}$$

where $\Gamma(\omega)$denotes the mother wavelet in the frequency domain. This condition implies at least two things about a valid mother wavelet:

1. $\Gamma(\omega)$ has band-pass property
2. $\varphi(t)$ has an oscillatory characteristic

After satisfying the admissibility condition, the inverse transform is given by:

$$f(t) = \frac{1}{C_\varphi}\int_0^{+\infty}\int_{-\infty}^{+\infty} WT_f(a,\tau)\varphi(t)d\tau\frac{da}{a^2} \tag{3-5}$$

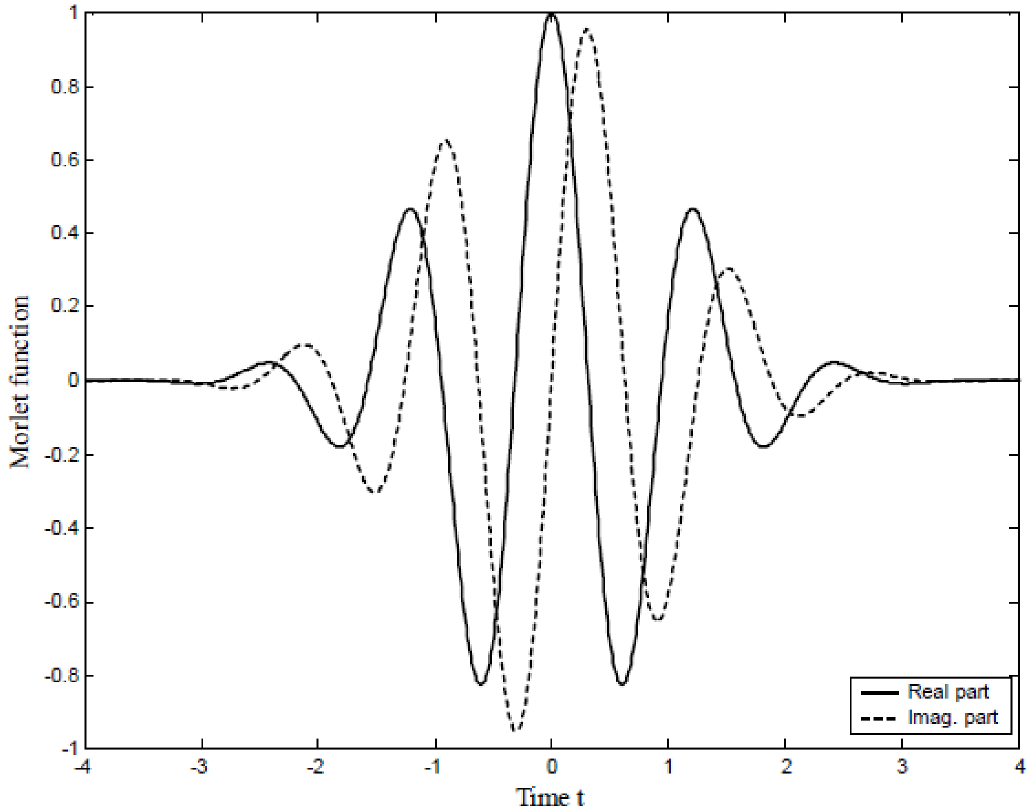

Figure 3.1 The Morlet wavelet in the time domain

3.1.2 Comparison with Short Time Fourier Transform (STFT)

To understand the major advantages of wavelet transforms, let us first review the short time Fourier transform (STFT) that is the most used spectral analysis method in speech signal processing.

$$F(\omega,\tau) = \int_{-\infty}^{+\infty} f(t)w(t-\tau)e^{-j\omega t}\,dt \tag{3-6}$$

where $f(t)$ is the target signal and $w(t-\tau)$ is the moving window. The limitation of the standard Fourier transform is that it extracts the frequency content of the signal only but not the frequency changes with respect to time. This is partially solved through the STFT by using sliding analysis windows. However the STFT uses a fixed window length and still cannot always simultaneously resolve short-lived events and closely spaced long-duration tones in speech (Quatieri, 2001). This drawback is rooted in the well-known uncertainty principle that limits time-frequency resolution: $D(x)B(x) > \frac{1}{4}$ where the product of time duration $D(x)$ and bandwidth $B(x)$ of a signal x must exceed a constant.

The wavelet transform minimizes the limitation of the uncertainty principle by varying the length of the moving window with variant scaling factor. Ideally, long windows are employed on low frequency parts of the speech signal for good frequency resolution and short windows are employed on high frequency components of the speech signal, say the attack of the glottal pulse and plosives of speech, for good time resolution. The wavelet transform succeeds in adjusting time and frequency resolution without defeating the uncertainty principle.

3.1.3 Implementation of Continuous Wavelet Transform

To calculate the inner product of the CWT, normally we need to resort to numerical integration using computers. The simplest way is to discretize time and shift as follows: $t = nT_s$ and $\tau = kT_s$ and T_s is the sampling interval. Then Eq. 3.2 becomes:

$$WT_f(a, kT_s) = \frac{T_s}{\sqrt{a}} \sum_n f(nT_s) \varphi \left[\frac{(n-k)T_s}{a}\right] \qquad (3\text{-}7)$$

For each value of the scale, we obtain a set of wavelet coefficients under this specific scale. There are some other existing fast algorithms for the continuous wavelet transform such as algorithm a’trous (Holschneider M., 1989). chirp-z transforms (Jones D., 1991). Mellin transform (Bertrand J., 1990). a under the admissibility condition:

$$c_\varphi = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < +\infty \qquad (3\text{-}8)$$

the two-dimensional wavelet coefficients $WT_f(a, kT_s)$ are a complete, stable yet redundant representation of the one dimensional signal. In order to speed up computation and save memory, we wish to discretize the scale a and shift τ in an efficient way to form a new set of wavelet coefficients.

3.2 Discrete Wavelet Transform

One drawback of the CWT is that the representation of the signal is often redundant.

Unlike the continuous wavelet transform, which can operate on every scale, the discrete wavelet transform (DWT) chooses a subset of scales and positions to calculate. A sample version of the wavelet coefficients $WT_f(a, \tau)$ can reconstruct the original signal in an efficient way if the family of dilated and shifted mother wavelets of selected a and τ constitute an orthonogonal and complete basis (Daubechies, 1992). A common sampling practice is that for each scale $a_m = a_0^m$ for m = 0, 1, 2, 3…N, the sampling interval is $\tau_m = \tau_0 a_0^m$ for m=0, 1, 2, 3…N. One particular natural case is when $a_0 = 2$ so that the sampling rate of the shift decreases by a factor of two as the scale increases by a factor of two (Quatieri, 2001). This is so called *dyadic or octave* sampling and it allows the implementation of a fast dyadic wavelet transform and its inverse with filter banks. High-pass filter removes the low-frequency components of the signal and the corresponding filter parameters become the *detailing* part of the wavelet coefficients. Low-pass filter removes the high frequency components of the signal and the corresponding filter parameters become the *smoothing* part of the wavelet coefficients. Partly due to the efficient implementation and auditory and visual cortex-like properties of dyadic wavelets, a large part of wavelet theory has involved finding dyadic wavelet bases that are orthogonal and that are useful in a variety of applications (Mallat S., 1998).

3.3 Multi-resolution Analysis of Discrete Wavelet Transform

The multi-resolution analysis concept was initiated by Meyer (Meyer Y., 1992) and Mallat (Mallat S., 1989) and provides a natural framework for the understanding of wavelet bases. In the dyadic wavelet transform, the basis functions are two parts: the scaling functions $\Psi(t)$ and the wavelet functions $\varphi(t)$.

$$\Psi_{m,\tau}(t) = 2^{-\frac{m}{2}}\Psi^0(2^{-m}t - \tau)\ where\ m\ \epsilon Z, \tau = n * 2^m \epsilon Z \qquad (3\text{-}9)$$

$$\varphi_{m,\tau} = a_0^{-\frac{m}{2}} \qquad (3\text{-}10)$$

The scaling function can be obtained as a sum of copies (dilated, shifted, scaled versions) of itself as illustrated in Eq.3.11,

$$\Psi^0(t) = \sum_{\tau=0}^{L} C_\tau \Psi(2t - \tau) \qquad (3\text{-}11)$$

and the wavelet function $\varphi^0(t)$can be then obtained from the scaling function $\Psi^0(t)$ as follows:

$$\varphi^0 = \sum_{-\infty}^{+\infty}(-1)^\tau C_{1-\tau}\Psi^0(2t - \tau) \qquad (3\text{-}12)$$

Where C_τ can be seen as the low-pass filter coefficients and $C_{1-\tau}$ can be seen as the high-pass filter coefficients and where L–1 is related to the number of vanishing moments in the scaling function $\Psi^0(t)$. They two together constitute a quadrature mirror filter (QMF) and an extensive study of the QMF can be found in (Monzon, 1994). The simple relation of two filter coefficients is as follows:

$$C_\tau(\tau) = (-1)^\tau C_{1-\tau}(L - 1 - \tau) \qquad (3\text{-}13)$$

Having the basis for decomposition, we can write the dyadic wavelet transform as follows:

$$f(t) = \sum_\tau c_{j,\tau}\varphi^0\,(t - \tau) + \sum_\tau \sum_{m=1}^{J} d_{j,k}\varphi^0(a_0^m * t - \tau) \qquad (3\text{-}14)$$

Where ϕ is the scaling function and φ is the wavelet function, $a_0 = 2, m = 1,2,3 \dots N$ and $\tau = \tau_0 * a_0^m$. The above equation shows how a signal can be decomposed into the summation of approximations (low frequency components of the signal) and details (high frequency components of the signal) at different resolutions.

3.4 Wavelet Thresholding

As wavelet analysis has its basis emulating the front-end auditory periphery (Mallat S., 1998). efforts have been made to take advantage this signal-processing tool for speech enhancement. The most used approach is based on the non-linear thresholding of the wavelet coefficients (Donoho D. L., 1995). which bridges the multi-resolution analysis and non-linear filtering.

3.4.1 Principle

Donoho proposed this powerful wavelet-based approach as follows (Donoho D. L., 1995):

Let y be a finite length observation sequence of the signal x that is corrupted by zero-mean white Gaussian noise n with variance σ^2:

$$y = x + n \qquad (3\text{-}15)$$

In the wavelet domain, this gives:

$$W_y = W_x + W_n \qquad (3\text{-}16)$$

The clean signal x can be estimated in the following way:

$$x = W^{-1} X_{estimation} = W^{-1} Y_{threshold} \qquad (3\text{-}17)$$

where $Y_{threshold}$ represents the wavelet coefficients after thresholding and W^{-1} denotes the inverse wavelet transform. The approach capitalizes on the fact that an appropriate transform (i.e., wavelet transform) projects the signal onto the transformed domain where the signal energy is concentrated in a small number of coefficients, while the noise is evenly distributed across the transformed domain. There are generally two ways of thresholding: one is called hard thresholding (Eq.3.18) and the other is called soft thresholding (Eq.3.19). Figure 3.2 is an illustration of this technique.

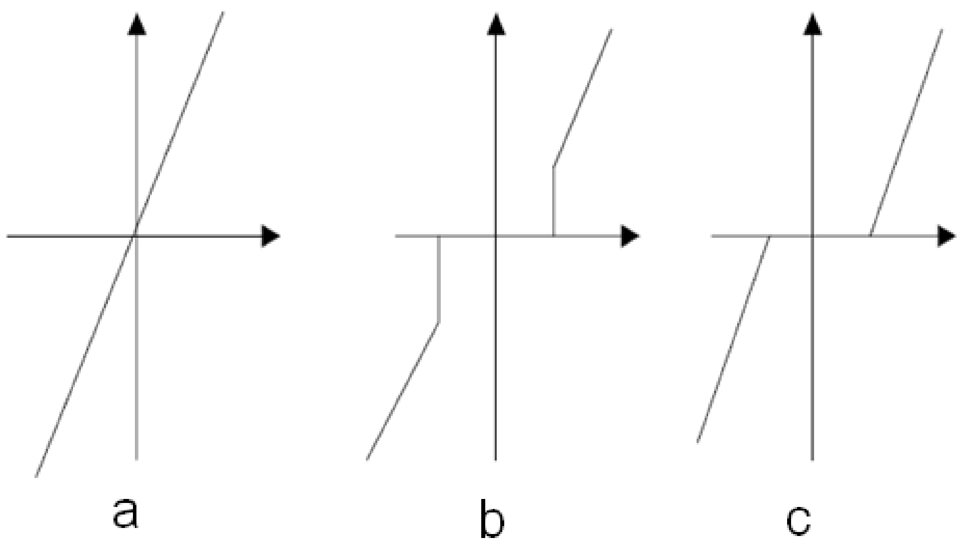

Figure 3.2 Wavelet thresholding a) No threshold, b) Hard Threshold, c) Soft Theshhold

$$T_{hard}(X,T) = \begin{Bmatrix} X & |X| > T \\ 0 & |X| < T \end{Bmatrix} \qquad (3\text{-}18)$$

$$T_{soft}(X,T) = \begin{Bmatrix} sgn(X)(X-T) & |X| > T \\ 0 & |X| < T \end{Bmatrix} \qquad (3\text{-}19)$$

Where X represents the wavelet coefficients before thresholding and T is the threshold. Both of these two methods suffer from distortion of the speech because they set coefficients to zero that may carry useful information, resulting in observable sharp time frequency discontinuities in the speech spectrogram. Various modifications have been made. For example, Sheikhzadeh (Sheikhzadeh, 2001) proposed using an exponential function to attenuate coefficients that are smaller than the threshold value in a nonlinear manner to avoid creating abrupt changes. Other data compression functions can also be chosen such as the μ-law:

$$T_{hard}(X,T) = \begin{Bmatrix} X & |X| > T \\ T\left(\frac{\left[(1+\mu)^{X/T}\right]}{\mu} sgn(X)\right) & |X| < T \end{Bmatrix} \tag{3-20}$$

Where X is the wavelet coefficients and T is the threshold value.

3.4.2 How to Choose the Threshold

The choosing of the threshold value can be determined in many ways. Donoho derived the following formula based on white Gaussian noise assumption:

$$T = \sigma\sqrt{2\log(N)} \tag{3-21}$$

where T is the threshold value, N is the length of the noisy signal y, and σ=MAD/0.6745, with MAD denoting the absolute median estimated on the first scale of the wavelet coefficients.

Johnstone and Silverman (Johnstone & Silverman, 1997) proposed the level dependent threshold method to deal with correlated noise, where for each frequency interval the threshold is proportional to the standard deviation of the noise in that interval.

$$\lambda_a = \sigma_a \sqrt{2\log(N_a)} \tag{3-22}$$

with $\sigma_a = \frac{MAD_a}{0.6745}$, N_a is the number of samples in scale a, and MAD_a is the absolute median estimated at scale a.

3.4.3 Four Types of Threshold Selection Rules

1. Threshold selection rule based on Stein's unbiased estimate of the risk

Different estimation rules could be compared on the basis of their resulting mean-square error (MSE) or more formally, the risk

$$R(s,T) = E\left\{\left\|s - \hat{s}\right\|^2\right\} \tag{3-23}$$

(Stein, S. M., 1981) has, under quite general conditions, derived an unbiased estimator of such a risk for a Gaussian estimator.

2. Heuristic threshold selection rule

This is a heuristic variant of the first option (Mathworks, 1998).

3. Fixed form threshold selection rule

This rule uses the universal threshold shown by Eq.3.21.

4. Minimax performance threshold selection rule

The minimax rule uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The derived formula is as follows (Guo, 2000) : $T = 0.3936 + 0.1829 * \frac{\log(N)}{long\ (2)}$ where N is the length of the signal.

3.5 Wavelet Packet Algorithm

The wavelet packet (WP) transform is a direct expansion of the traditional discrete wavelet transform. Most importantly, it has well localization both in time and frequency domain. WP decomposition was first introduced by Coifman, Meyer and Wickerhauser (C. Herley & M. Vetterly, 1994).

Orthonormal basis which best represents the function under a definite criterion, is available for WP representation. It is emphasize that WP expansion is signal dependent. An algorithm for a given signal invents the best set of basis functions so that the decomposition of the signal. Choosing a basis implies choosing a tree structure of a dydic filter bank which obtains the transform coefficients (R.R. Coifman & M.V. Wickerhauser). Therefore, the demonstration of the decomposition is simple a computationally efficient.

WP analysis for a time series can be summarized as follows (S. Mallat, 1999). A space V_j of a multiresolution analysis in L2(R) is analyzed in a lower resolution space V_{j+1} added a detail spaceW_{j+1}. Dividing the orthogonal basis $\{\emptyset_j(t-2^{j_n})\}_{n\epsilon Z}$ of in to new orthogonal basis constitutes $\{\emptyset_{j+1}(t-2^{j+1_n})\}_{n\epsilon Z}$ of V_j and $\{\psi_{j+1}(t-2^{j+1_n})\}_{n\epsilon Z}$ of W_{j+1} .

The decompositions of ϕ_{j+1} and ψ_{j+1} are denoted by a pair of conjugate mirror filter h[n] and $g[n]=(-1)^{1-n}h[1-n]$.

Theorem 1:

Let $\{\emptyset_j(t-2^{j_n})\}_{n\epsilon Z}$ be an orthonormal basis of a space U_j. Let h and g a pair of conjugate mirror filters. This relation is defined by

$$\theta_{j+1}^0(t)=\sum_{n=-\infty}^{+\infty} h[n]\theta_j(t-2^{j_n}) \qquad (3\text{-}24)$$

and

$$\theta_{j+1}^{1}(t) = \sum_{n=-\infty}^{+\infty} g[n]\theta_j\left(t - 2^{j}n\right) \tag{3-25}$$

The family

$$\left\{\theta_{j+1}^{0}\left(t - 2^{j+1}n\right), \theta_{j+1}^{1}\left(t - 2^{j+1}n\right)\right\}_{n\epsilon Z} \tag{3-26}$$

is an orthonormal basis of U_j .

This theorem proves that we can set $U_j = W_j$ and divide these detail spaces to create new bases. The recursive slicing of vector spaces is evidenced in a binary tree. If the signal is approximated at the scale 2L, it is associated the approximation space VL to the root of the tree. This space permits an orthonormal basis of scaling functions $\left(\emptyset_L(t - 2^{L}n)\right)_{n\epsilon Z}$ with $\emptyset_L(t) = 2^{-L/2}\emptyset(2^{-L}t)$.

Any node of the binary tree is labeled by (j, k). where j-L ≥ 0 is the depth of the node on the tree, and k is the number of nodes. A space W_j^k allowing an orthonormal basis $\left\{\psi_j^k\left(t - 2^{j}n\right)\right\}_{n\epsilon Z}$ is associated to each node (j, k) by going down the tree. At the root, it has $W_L^0 = V_L$ and $\psi_L^0 = \phi_L$. The WP orthogonal bases at the nodes are defined by

$$\psi_{j+1}^{2k}(t) = \sum_{n=-\infty}^{+\infty} h[n]\psi_j^k\left(t - 2^{j}n\right) \tag{3-27}$$

and

$$\psi_{j+1}^{2k+1}(t) = \sum_{n=-\infty}^{+\infty} g[n]\psi_j^k\left(t - 2^{j}n\right) \tag{3-28}$$

because of $\left\{\psi_j^k\left(t - 2^{j}n\right)\right\}_{n\epsilon Z}$ is orthonormal

$$h[n] = \langle\psi_{j+1}^{2k+1}(t), \psi_j^k\left(t - 2^{j}n\right)\rangle \tag{3-29}$$

and

$$g[n] = \langle \psi_{j+1}^{2k+1}(t), \psi_j^k\left(t - 2^{j_n}\right)\rangle \quad (3\text{-}30)$$

Therefore, this recursive splitting determines a binary tree of wavelet packet spaces which defined as

$$W_{j+1}^{2k} \oplus W_{j+1}^{2k+1} = W_j^k \quad (3\text{-}31)$$

We illustrate $\tilde{x}[n] = x[-n]$ and the signal $\tilde{x}[n]$ is given by injecting a zero between each sample. Respectively, the decomposition and reconstruction coefficients are constituted by

$$d_{j+1}^{2k}[t] = d_j^k * \tilde{h}[2t] \text{ and } d_{j+1}^{2k+1}[t] = d_j^k * \tilde{g}[2t] \quad (3\text{-}32)$$

$$d_j^k[t] = \tilde{d}_{j+1}^{2k} * h[2t] + \tilde{d}_{j+1}^{2k+1} * g[2t] \quad (3\text{-}33)$$

To sub sampling the convolution of d_j^k with $\tilde{h}$ and$\tilde{g}$, the coefficients can be obtained. Iterating these equations the all branches of the tree are computed by WP coefficients. This is given in Figure 3.3.

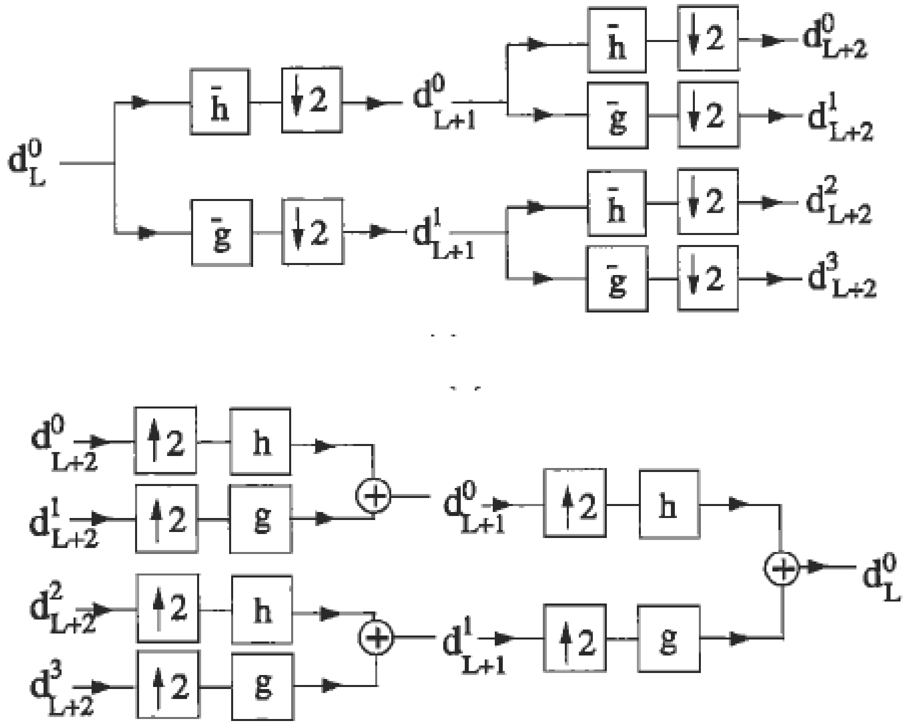

Figure 3.3 (a) Wavelet packet decomposition with down sampling, (b) Wavelet packet reconstruction with up sampling

3.5.1 Best Tree

Best tree (Coifman, R.R. & M.V. Wickerhauser, 1992) function is a one- or two-dimensional wavelet packet analysis function that computes the optimal sub tree of an initial tree with respect to an entropy type criterion. The resulting tree may be much smaller than the initial one. Following the organization of the wavelet packets library, it is natural to count the decompositions issued from a given orthogonal wavelet. A signal of length N = 2L can be expanded in α different ways, where α is the number of binary sub trees of a complete binary tree of depth L where$a \geq 2^{N/2}$. This number may be very large, and since explicit enumeration is generally intractable, it is interesting to find an optimal decomposition with respect to a convenient criterion, computable by an efficient algorithm. We are looking for a minimum of the criterion.

3.5.2 Algorithm

Consider the one-dimensional case. Starting with the root node, the best tree is calculated using the following scheme. A node N is split into two nodes N1 and N2 if and only if the sum of the entropy of N1 and N2 is lower than the entropy of N. This is a local criterion based only on the information available at the node N. Several entropy type criteria can be used. If the entropy function is an additive function along the wavelet packet coefficients, this algorithm leads to the best tree. Starting from an initial tree T and using the merging side of this algorithm, we obtain the best tree among all the binary sub trees of T (Mathworks, 1998).

3.5.3 Entropy

Entropy provides a complexity measure of a time series, such as discretized speech signal.

3.5.4 Shannon Entropy

The Shannon entropy equation provides a way to estimate the average minimum number of bits needed to encode a string of symbols, based on the frequency of the symbols (Schneier, Shannon & Claude E., January, 1951).

$$H(X) = -\sum_{i=0}^{N-1} p_i \log_2 p_i \qquad (3\text{-}34)$$

In the Shannon entropy equation, pi is the probability of a given symbol. To calculate log2 from another log base (e.g., log10 or loge):

CHAPTER FOUR

NEW SPEECH PROCESSING STRATEGIES METHODS

4.1 Speech Processing

4.1.1 Windowing

In signal processing, a window function (or apodization function) is a function that is zero-valued outside of some chosen interval. For speech processing the signal is assumed which is short-time stationary and perform a Fourier transform on these small blocks. Solution: multiple the signal by a window function that is zero outside some defined range (Eric W. Weisstein, 2003).

The Hanning window (Blackman, R. B. & Tukey, J. W., 1959; W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T., 1992) is a general purpose window for the analysis of continuous signals and should be used in most cases, because it has the best overall filter characteristic. We separate the signal with windowing process.

The Hann function, named after the Austrian meteorologist Julius von Hann, is a discrete probability mass function given by

$$\omega(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \qquad (4\text{-}1)$$

4.1.2 Noise Theory and Performance Criteria

Assuming that the speech signal, X, and the noise, N, are additive, the noisy speech, y, is modeled as

$$Y = X + N \qquad (4\text{-}2)$$

It is generally adopted that the speech is not correlated with noise; this is a reasonable assumption in most cases when the signal and noise are generated by independent sources. Noise equation can be write easily as

$$N = Y - X \tag{4-3}$$

The performance criteria is SNR value which is estimated by this formula

$$SNR(Y,\hat{Y}) = 10\log\left[\frac{\|Y\|_2^2}{\|Y-\hat{Y}\|_2^2}\right][dB] \tag{4-4}$$

where Y input signal, $\hat{Y}$ output signal and related transfer block as shown Figure 4.1.

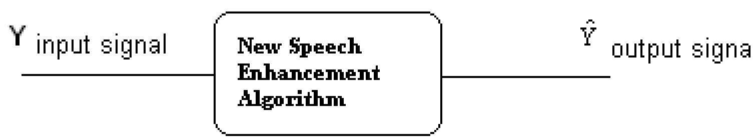

Figure 4.1 Block diagram of the transfer function of SNR enhancement

We assume $\hat{Y}$ approximately equals original signal X therefore Y – $\hat{Y}$ equals N. Generally, the form of noise is classified as white noise and colored noise.

4.1.2.1 White Noise

Pure white noise (Saeed V. Vaseghi, 2000; Bell D.A., 1960; Bennett W.R, 1960) is a theoretical concept, since it would need to have infinite power to cover an infinite range of frequencies. Furthermore, a discrete-time signal by necessity has to be band-limited, with its highest frequency less than half the sampling rate. A more practical concept is band-limited white noise, defined as a noise with a flat spectrum in a limited bandwidth. The spectrum of band-limited white noise with a bandwidth of B Hz is given by

$$P_{NN}(f) = \begin{cases} \sigma^2, & |f| \le B \\ 0, & otherwise \end{cases} \qquad (4\text{-}5)$$

4.1.2.2 Cloured Noise

Although the concept of white noise provides a reasonably realistic and mathematically convenient and useful approximation to some predominant noise processes encountered in telecommunications systems, many other noise processes are nonwhite. The term 'coloured noise' (Saeed V. Vaseghi, 2000; Bell D.A., 1960; Bennett W.R., 1960) refers to any broadband noise with a non-white spectrum. For example most audio frequency noise, such as the noise from moving cars, noise from computer fans, electric drill noise and people talking in the background, has a nonwhite predominantly low frequency spectrum. Also, a white noise passing through a channel is 'coloured' by the shape of the frequency response of the channel. Two classic varieties of coloured noise are so-called 'pink noise' and 'brown noise', shown in Figure 4.2 and Figure 4.3.

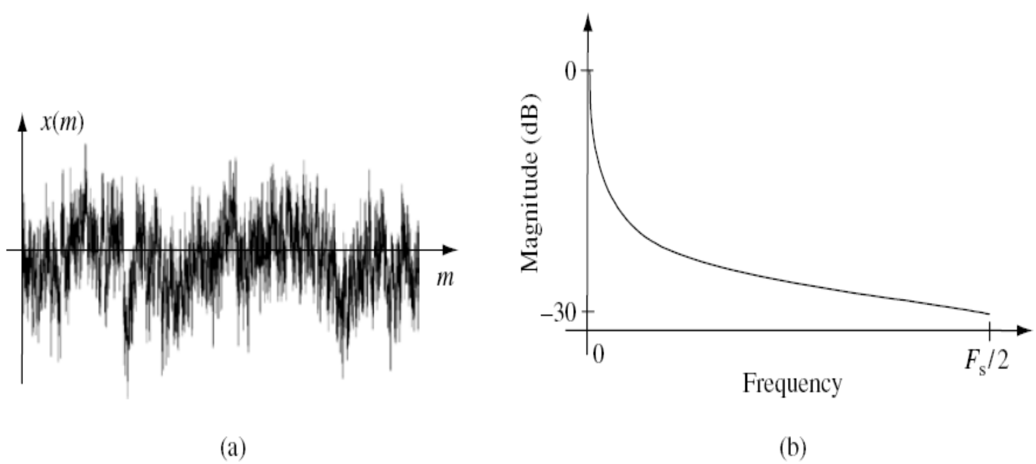

Figure 4.2 (a) A pink noise signal and (b) its magnitude spectrum

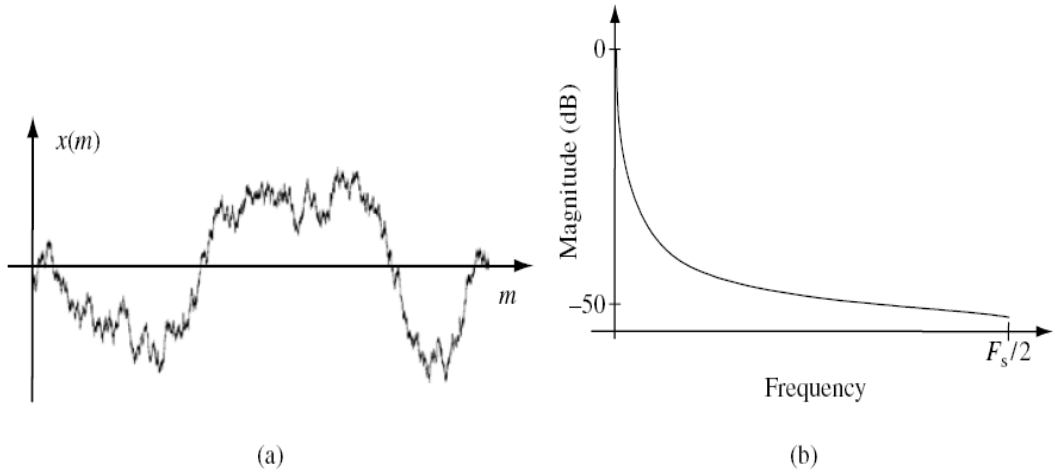

Figure 4.3 (a) A brown noise signal and (b) its magnitude spectrum

4.2 New Speech Processing Strategies Methods

4.2.1 Algorithm

New speech processing method (Figure 4.4) is constituted five blocks which are windowing, wavelet packet transform, construct optimum tree, and determine channels outputs, electrodes selection, stimuli (constructed signal). Each block is explained step by step below. Each step is handled with MATLAB environment and simulation codes are provided in Appendix section.

$$S_n = S * W_n \gg C_{i,j} = \psi_n[S_n] \gg C'_{i,j} = \Lambda[C_{i,j}] \gg E_k = M[C'_{i,j}] \qquad (4\text{-}6)$$

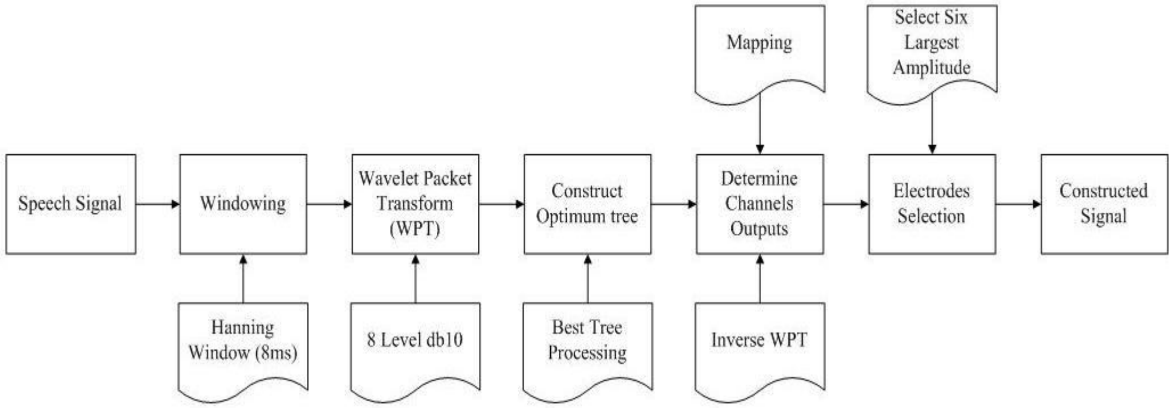

Figure 4.4 Block diagram of New Speech Processing method

4.2.2 Windowing

Windowing is useful operator in order to eliminate the sparks from signal. In our study the speech signal is separated for speech processing by Hanning window and its window length is 8 ms.

$$S_n = S * W_n \tag{4-7}$$

where S speech signal and W_n is windowing operator.

4.2.3 Wavelet Packet Transform

A wavelet transform iterates the decomposition of the smooth part into a smooth part and details while leaving the details intact. In wavelet packet transforms, the details are further decomposed into a "smooth" part plus "details". This block is very important because selected mother wavelet and processing level change our resolution results in signal processing. Mother wavelet is selected with experimentally then decided to use db10 wavelet all process and processing level is 8.

$$C_{i,j} = \psi_n[S_n] \tag{4-8}$$

where ψ_n wavelet packet operator and after this operation we have wavelet coefficients $C_{i,j}$.

4.2.4 Determine Optimum Tree

In this block which is first process to clean noise components from speech signals. Optimum tree is decided by using Shannon entropy into WPT to clean unnecessary nodes from wavelet packet tree, in this way noise parts are eliminated from speech signal and reduce channel interaction between neighbor channels owing to determine clean channel outputs.

This process is most innovation in speech strategies in cochlear implant because noise very big problem for success of cochlear implant strategy. This parts helps to us select more accurately cochlear implant electrode during process.

This part can simply present as by

$$C'_{i,j} = \Lambda[C_{i,j}] \tag{4-9}$$

where Λ is bets tree operator and this operator rearrange our wavelet coefficients. $C'_{i,j}$ is new wavelet coefficients after Λ operation.

4.2.5 Determine Channels Outputs and Mapping

The mapping applied to optimum tree and channels outputs are determined from mapping function. Mapping function refers relations of between electrodes and wavelet packet transforms outputs nodes. You can find relation of between electrodes and nodes in the Table 5.1, where ***number of electrode*** is cochlear implant electrode identifier ; this study has 22 electrodes for stimuli simulation, ***F1 and F2 electrodes*** are defined cochlear implant electrodes cut-off frequencies that derived by equation (5-6). ***F1 and F2 wavelet nodes*** are defined WPT nodes cut-off frequencies that derived by WPT tree. Finally ***number of node*** is WPT node that matches with band-width values between cochlear implant electrodes. This node can be combination of several WPT nodes such as number 8 in the table.

Mapping operation is defined as

$$E_k = M\left[C'_{i,j}\right] \tag{4-10}$$

where E_k is electrode output and M is mapping operator (Table 5.1).

We calculate channels bandwidths, F1 and F2 for human cochlea; the frequency-position function can be described as the following equation

$$f = A(10^{ax} - k) \tag{4-11}$$

Where f represents frequency in Hz, x is expressed as a proportion of basilar length (from 0 to l) A=165.4 and a=2.1. Then we map all channel to node or nodes group for determine channels outputs.

Table 4.1 Cochlear Implant electrodes and wavelet packet transform node mapping list. **First column** indicates cochlear implant electrode number, **second column** indicates cut – off frequencies of cochlear implant electrodes, **sixth column** indicates: cut – off frequencies of wavelet nodes, **seventh column** indicates Wavelet Packet Transform nodes

Number of Electrode	**F1 (electrodes)**	**F2 (electrodes)**	**Bandwidth**	**Number of Node**	**F1 (wavelet nodes)**	**F2 (wavelet nodes)**
1	150	201.53	51.533	257	125	187.5
2	201.53	262.05	60.518	258	187.5	250
3	262.05	333.12	71.07	259	250	312.5
4	333.12	416.58	83.462	260	312.5	375
5	416.58	514.6	98.015	130	375	500
6	514.6	629.7	115.11	131	500	625
7	629.7	764.88	135.18	132	625	750
8	764.88	923.62	158.74	267-268-269	750	937.5
9	923.62	1110	186.42	270-271-272	937.5	1125
10	1110	1329	218.93	136-137	1125	1375
11	1329	1586.1	257.1	138-139	1375	1625
12	1586.1	1888	301.93	140-141	1625	1875
13	1888	2242.6	354.57	142-143-144	1875	2250
14	2242.6	2659	416.4	72-73	2250	2750
15	2659	3148	489	74-75	2750	3250
16	3148	3722.3	574.27	37	3000	3500
17	3722.3	4396.6	674.4	38-39	3500	4500
18	4396.6	5188.6	791.99	80-81-82-83	4250	5250
19	5188.6	6118.7	930.08	20	5000	6000
20	6118.7	7211	1092.2	43-44-45	6000	7500
21	7211	8493.7	1282.7	22-23	7000	9000
22	8493.7	10000	1506.3	11	8000	10000

4.2.6 Electrodes Selection

Electrodes selection phase is same as traditional N of M strategy. In our study six channels are selected for stimuli using largest amplitudes in channel outputs. Therefore in our study N = 6, M =22.

4.2.7 Stimuli (constructed signal)

The six amplitudes of the spectral maxima are finally logarithmically compressed to fit the patient's electrical dynamic range, and transmitted to the six selected electrodes through a radio-frequency link. In our study this process is simulated by adding operation. In order to construct the signal from output channels selected output signal are added respectively.

CHAPTER FIVE

RESULTS

5.1 Process Output and Selected Electrodes

In this research, the output waveforms are constructed using N of M selection approach. New speech processing strategy waveform, as shown the Figure 5.2, looks like similar original signal Figure 5.1 than traditional N of M method waveform Figure 5.3. Both signals are produced by using MATLAB simulation codes that are given in the Appendix chapter and graphical illustrations are prepared by SFS 4/Windows.

As shown from the graphs, traditional N of M removes some high frequency component that are between 25ms and 75ms at wide-band spectrogram, high frequency components are very important for intelligibility and consonant recognition such as “s”, “ş”, “f”, etc. New method keeps high frequency component using WPT because WPT analyzes high-frequency component as well as low frequency component. Another effect is mother wavelet selection; db10 is more effective high frequency analysis that figured out by experimentally.

“Determine Optimum Tree” block eliminate noise and unnecessary components in speech signal therefore, we can obtain better result than N of M for electrode selection. New strategy output electrodes (channels) more accurate than N of M and it conduces to reduce interaction between neighbor channels, selection result for the word “good” is shown below Figure 5.4 and Figure 5.5. New method electrodes have high-frequency presenting electrodes than traditional N of M method; this is parallel with spectrogram results for each signal.

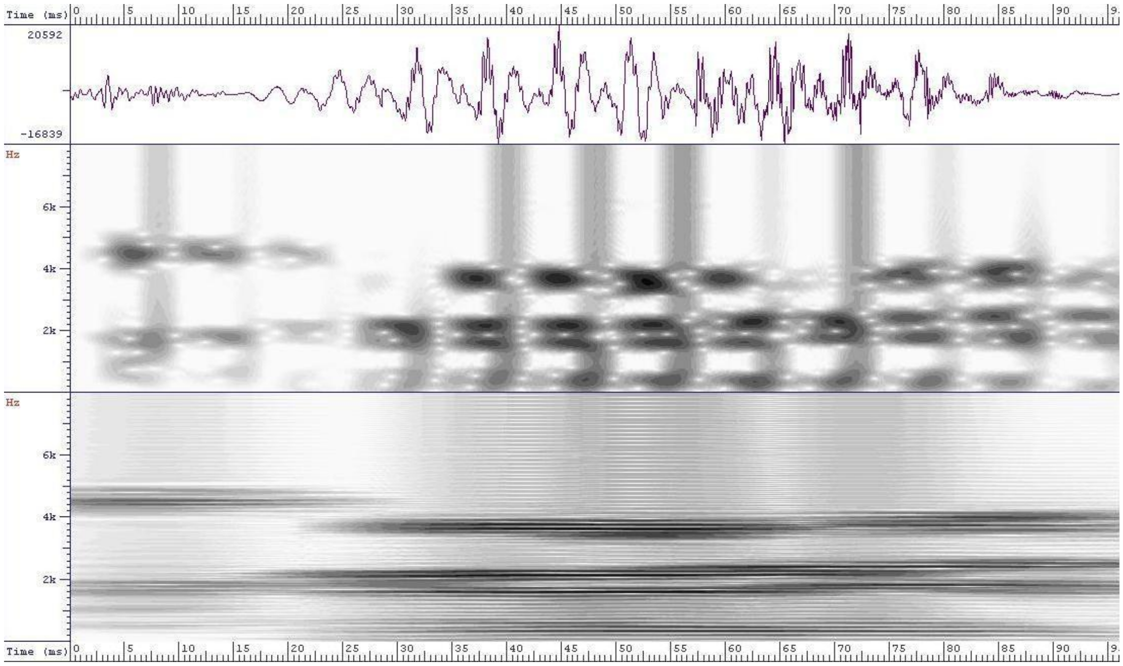

Figure 5.1 Signal waveforms, wide-band spectrogram and narrow-band spectrogram for original signal.

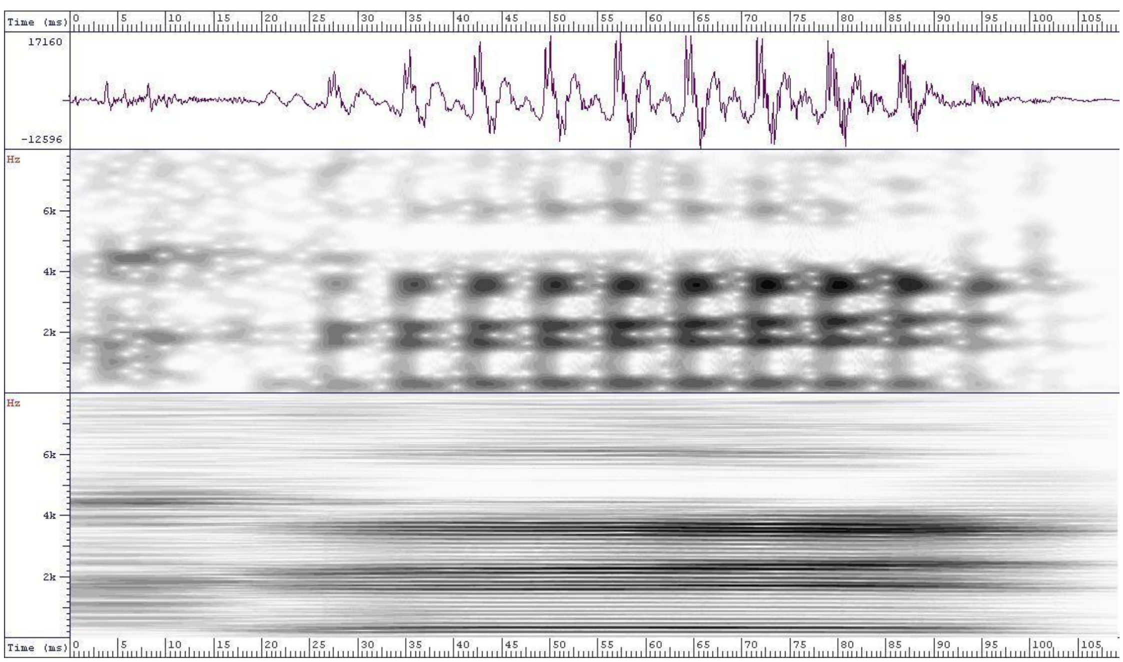

Figure 5.2 Signal waveforms, wide-band spectrogram and narrow-band spectrogram for New Speech Processing method.

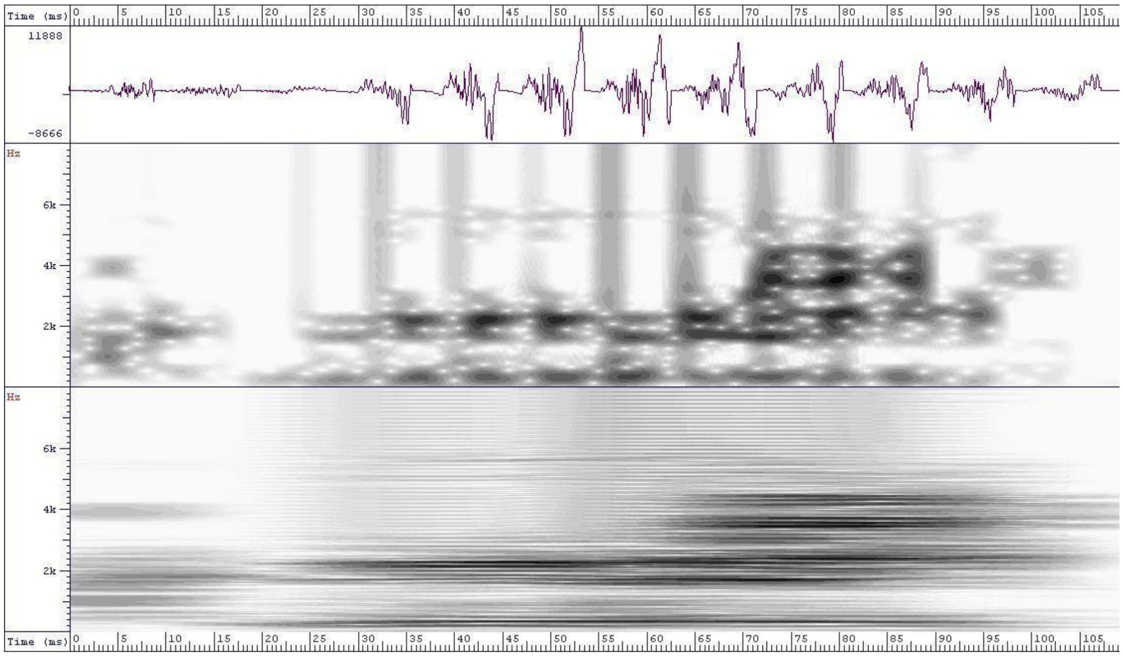

Figure 5.3 Signal waveforms, wide-band spectrogram and narrow-band spectrogram N of M strategy.

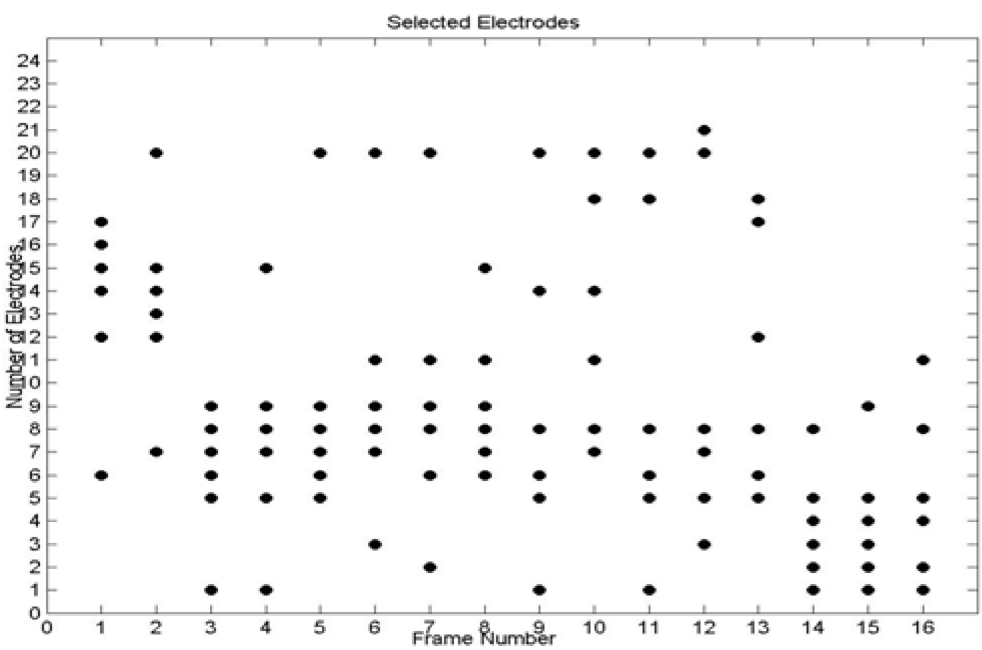

Figure 5.4 Frame number vs. Cochlear Implant electrodes mapping for new method. Each frame has six selected electrodes.

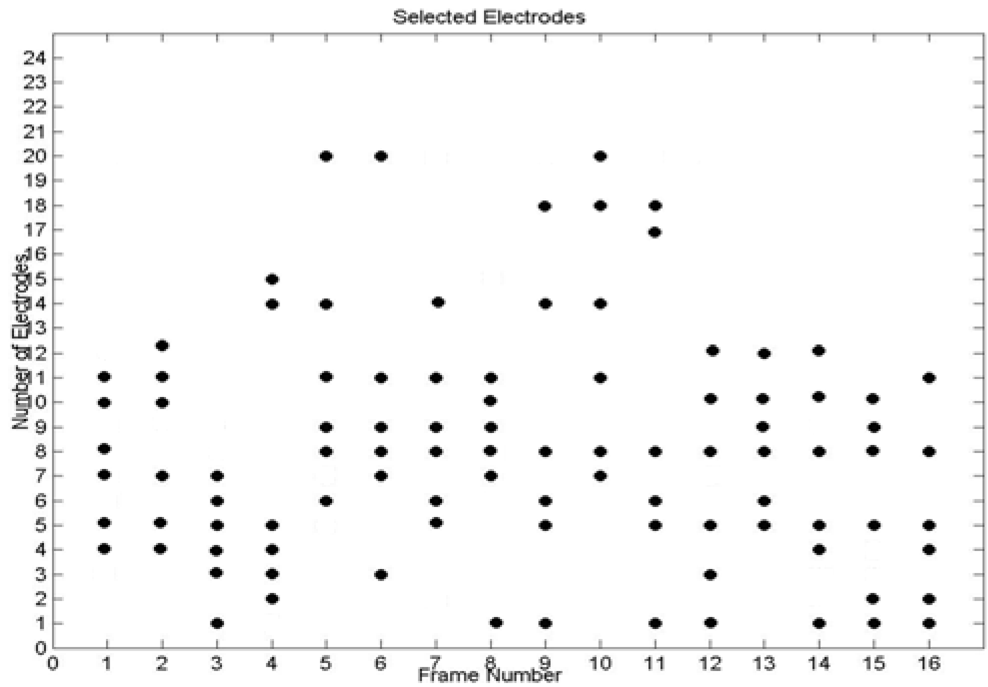

Figure 5.5 Frame number vs. Cochlear Implant electrodes mapping for traditional N of M method. Each frame has six selected electrodes.

5.2 Intelligibility

Twenty normal-hearing listeners between the ages of 23 to 33, with an average of 24.3 years, participated in the experiment. All subjects were native speakers of Turkish and had air conduction thresholds better than 20 dB HL at octave frequencies ranging from 250 to 6000 Hz bilaterally. The immittance test results from tympanograms and acoustic reflex thresholds were consistent with normal middle ear function in both ears.

For the practice session, twenty-five words were used for each algorithm (Table 5.1). purpose of usage two different lists during intelligibility test is avoiding from *recall effect* that appears if listeners listen same list during experiment. These words are phonetically balanced and difficult level is adjusted same for both list. In order to prepare these lists, vowel and consonant usage frequencies in Turkish Language and

Turkish Language characteristics were considered, number of vowel and consonant in each list as defined at frequency tables Table 5.2 and Table 5.3. This information extracts from all Turkish words that has three letters in Turkish Language Association Dictionary.

All words as per list are simulated by appropriate method which are new method and traditional N of M strategy. Simulated words are listened to listeners directly by head-set and requested to type to excel sheet from listener what understood when they was listening words. Intelligibility criteria were calculated as

$$Intteligibility = \left(\frac{Number\ of\ correct\ letter}{All\ letter\ in\ the\ list}\right)\ X\ 100$$

For example, each list has 75 letters (25 words and each word has 3 letters) and if listener understands 50 correct letter from whole list intelligibility should be % 66.66.

Test results showed us that new method has better intelligibility from traditional N of M strategy as a result of practice session on Table 5.4 and as shown Figure 5.6. The values which are average percentage of intelligibility for male, female listeners per algorithm are given in the Table 5.4.

Table 5.1 List of intelligibility test samples

List of New Speech Method	**List of N of M Method**
ben	bir
bin	bor
cin	cıs
çığ	çim
dar	dal
der	dev
dur	din
dür	dul
fay	fiş
giz	göl
hat	hız
kal	kal

kil	kan
kol	kin
muş	mey
nal	nem
nem	pul
pas	ret
sil	sar
sön	ser
şık	şık
tel	tan
tim	tar
yan	yağ
yer	yün

Table 5.2 Vowel frequency table

Vowel	Frequency
a	7
e	5
ı	3
i	5
o	1
ö	1
u	2
ü	1

Table 5.3 Consonant frequency table

Consonant	First position frequency	End position frequency
b	2	0
c	1	0
ç	1	0
d	4	0
f	0	1
g	1	0
ğ	0	1
h	1	0

j	0	0
k	2	2
l	0	5
m	1	2
n	1	5
p	0	1
r	1	5
s	2	1
ş	1	1
t	2	1
v	0	1
y	2	1
z	0	1

Table 5.4 Average values of intelligibility test result for each algorithm

	Sex	**Age Range**	**Number of Attendees**	**Intelligibility (%)**
New	Male	23 – 29	8	81.93
	Female	23 – 33	12	79.17
	Total	23 – 33	20	80.55
N of M	Male	23 – 29	8	77.60
	Female	23 – 33	12	75.21
	Total	23 – 33	20	76.40

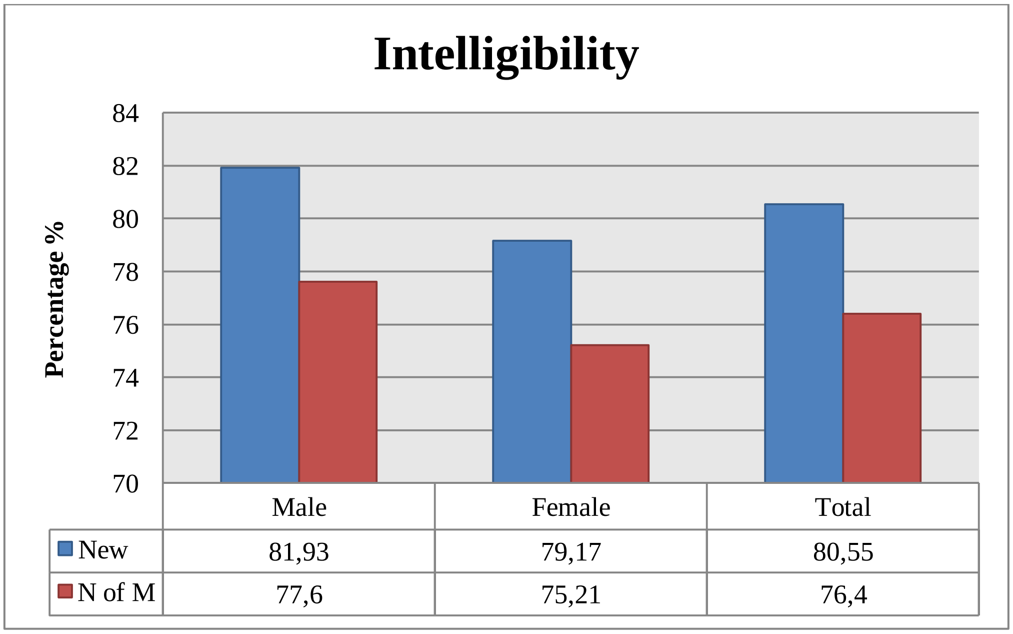

Figure 5.6 Graphical presentation of intelligibility test result

5.3 Noise Resistance Comparison

Another test is SNR enhancement test. In our test the samples are contaminated with different noise types which are pink, F-16, factory and volvo noise and the noise level is 5 DB. Then we applied new selection method and traditional N of M method to whole samples and compared SNR changes by using SNR enhancement method. As shown Figure 5.7 new selection method gives better result than traditional N of M method.

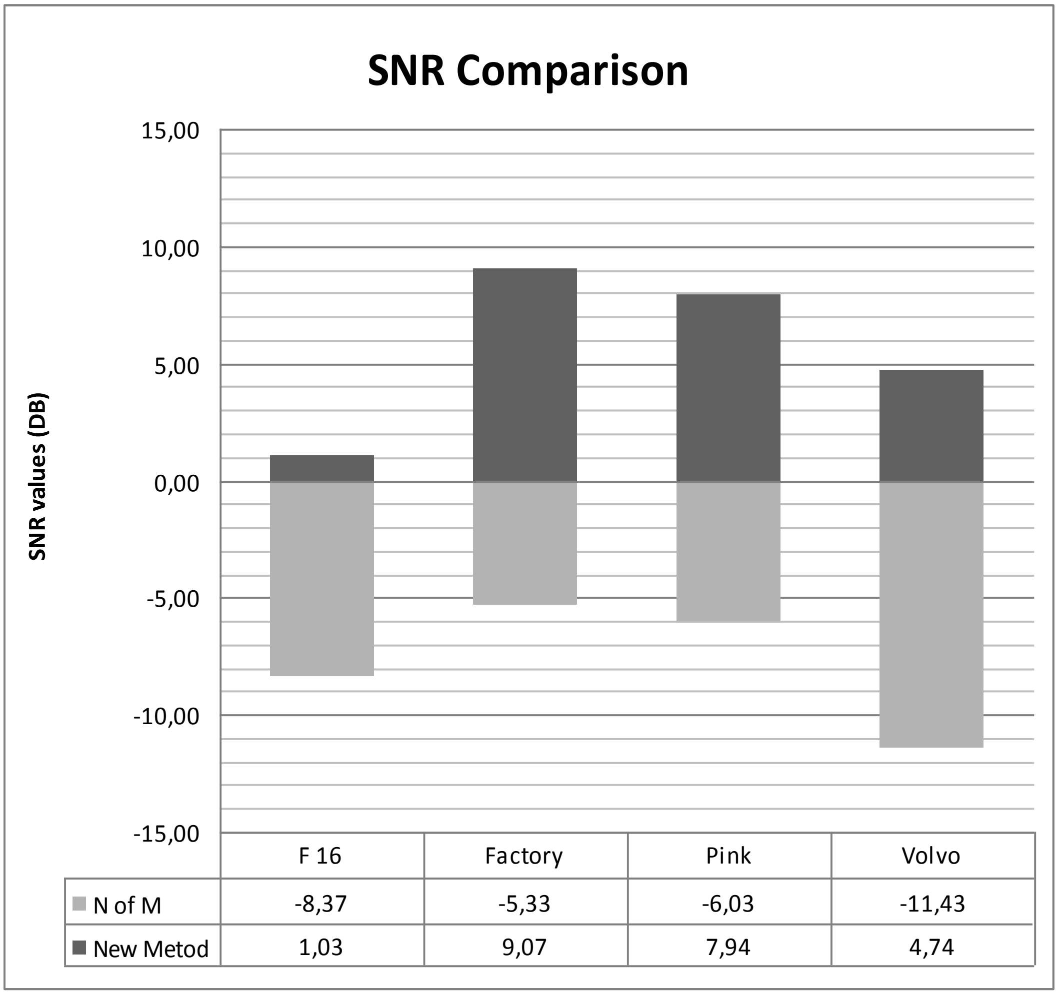

Figure 5.7 SNR comparison for “good” word. The sample is contaminated with different type real noises which are “F 16 cockpit”, “Factory”, “Pink”, “Volvo cockpit”.

CHAPTER SIX

CONCLUSION

6.1 General Results

In this study, an improved speech processing system that works in wavelet domain was proposed for digital hearing aid applications. The core of the system is based on the WPT and also used the energy of the wavelet coefficients. By applying several different tests, we investigated on the effect of intelligibility and noise resistance for the suggested speech processing method. Then, we presented a new electrode selection algorithm which depends on wavelet entropy distribution. The proposed electrode selection increased the noise performance and intelligibility. Additionally, the performance of the proposed methods is better than traditional and recent published methods. Further studies can be done on the improving intelligibility in the speech enhancement systems.

"Determine optimum tree" by using best three function is significant part of this study because this part eliminates noise and unnecessary components from speech signal. It helps to improve intelligibility of speech in noisy environments such as roads, train stations, conference halls, etc…

Unfortunately, using wavelet packet transform and best tree function increase speech processing time and it is not sufficient real-time application yet. This study cannot be use into current cochlear implant speech processors.

During the human experiments session normal-hearing people are used and all result based on only normal-hearing people as well, patients who are using cochlear implant should use for more accurate result for intelligibility. This might give us more accurate results.

6.2. Future Plan

For this thesis study, three topics below might be considered for future study. First of all is using hybrid mother wavelet during wavelet decomposition process. Daubechies family could be use for low-pass filter decomposition and Symlet family for high-pass filter decomposition. Second one is deciding mother wavelet due to speech signal characteristic at run-time. It might be give better results for speech intelligibility. Last topic is bionic wavelet usage instead of wavelet packet transform in entire speech processing. Bionic wavelet concept is new and it has better time-frequency resolution then wavelet packet transform.

REFERENCES

Advanced Bionics Corporation (2000). *How a cochlear implant works*. U.S.A.

Ashley N. N. (2007). *The History of Cochlear Implants Comparing the Current Cochlear Implant Manufacturers*. Department of Exceptionality Programs College of Professional Studies Bloomsburg University, Bloomsburg, Pennsylvania

Bell D.A. (1960). *Electrical Noise and Physical Mechanism*. Van Nostrand, London.

Bennett W.R (1960). *Electrical Noise*. McGraw-Hill, NewYork.

Bertrand, J. (1990). Discrete mellin transform for signal analysis. *International conference on audio, speech and signal processing, IEEE*.

Blackman, R. B. and Tukey, J. W. (1959). Particular Pairs of Windows In The Measurement of Power Spectra. From the *Point of View of Communications Engineering*, New York: Dover.

C. Herley, M. Vetterly (1994). Orthonormal time varying filterbanks and wavelet packets. *IEEE Trans. on Signal Processing*, vol. 41, no. 10.

C. Parkins and S. Anderson(eds.) (1983). *Cochlear prostheses: An international symposium New York*. New York Academy of Sciences.

Christiansen, J.B, Leigh, I. W. (2002). Cochlear *implants in children: Ethics and choices*. Washington, DC. Gallaudet University Press.

Chute, P.M., Nevins, M.E. (2002). *The parents' guide to cochlear implants*. Washington, DC. Gallaudet University Press.

Cochlear Inc (2005). Retrieved September 2009 from Cochlear Implant website. http://www.cochlear.com.

Cochlear, Inc (2005). Nucleus freedom with smart sound. Retrieved September 2009 from Cochlear Implant website. http://www.cochlearamericas.com/Products/22.asp.

Coifman, R.R.; M.V. Wickerhauser (1992). Entropy-based algorithms for best basis selection, *IEEE Trans. on Inf. Theory*, vol. 38, 2, pp. 713-718.

Daubechies I. (1992). *Ten lectures on wavelets Philadelphia*. USA, Society for Industrial and Applied Mathematics.

Deller, J. R., Proakis, J.G., Hansen, J.H.L. (1994). *Discrete-time processing of speech signals*. New York, Macmillan Publishing Company.

Donoho L. (1995). Denosing by soft thresholding. *IEEE Trans. on Information Theory* 41(3): 613-627.

Eric W. Weisstein (2003). *CRC Concise Encyclopedia of Mathematics*. CRC Press. ISBN 1584883472

G. Loeb (1990). Cochlear prosthetics. *Annual Review in Neuroscience*, vol. 13, pp. 357-371.

Guo, D. (2000). A study of wavelet thresholding denoising. *IEEE International conference on signal processing*.

Holschneider, M. (1989). *Wavelet, time frequency methods and phase space*. First International conference on wavelet, Springer-Verlag.

J. Millar, Y. Tong, and G. Clark (1984). Speech processing for cochlear implant prostheses. *Journal of Speech and Hearing Research*, vol. 27, pp. 280-296.

Johnstone, I. M. a. S., B.W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Royal Statistical Society* 59(2): 319-351.

Jones, D. (1991). Efficient approximation of continuous wavelet transform. *Electronic letters, IEEE* 27(9): 748-750.

L. Brechet, M.F. Lucas at all. (2007). Compression of biomedical signal with mother wavelet optimization and best basis wavelet packet selection. *IEEE Trans. On Biomedical Eng.*, vol. 54, no. 12, September.

Loizou, P. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, 15(5). 101-130.

M. Dorman, P. Loizou, and D. Rainey (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America*, vol. 102, pp. 2403-2411.

Mallat, S. (1989). A theory for multiresolution signal decomposition. *IEEE TRans. On Pattern Anal. Machine Intell.* (11): 674-693.

Mallat, S. (1998). *A wavelet tour of signal processing*, New York, Academic Press.

Mathworks (1998). *Matlab Wavelet Toolbox User's Guide*, V5.12.

McCormick, B., Archbold, S. (2003). *Cochlear implants for young children (2nd*

Edition). London and Philadelphia: Whurr Publishers.

Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University.

Monzon. L. A. (May 1994). *Ph.D dissertation: Constructive Multiresolution Analysis and the Structure of Quadrature Mirror Filters*, Yale University.

Moore, J.A. & Teagle, H.F.B. (2002). *An introduction to cochlear implant technology, activation, and programming Language, Speech, and Hearing Services in Schools*, 153-161.

National Institute on Deafness Other Communication Disorders (2009). Retrieved September 2009 from NIDCD website http://www.nidcd.nih.gov/health/hearing/coch.asp

Nevins, M.E., Chute, P. M. (1996).*Children with cochlear implants in educational settings*. San Diego and London: Singular Publishing Group Inc.

Quatieri (2001). *Discrete-Time Speech Signal Processing, Principles and Practice*. PrenticeHall PTR.

R. Shannon, F. Zeng, V. Kamath, J. Wygonski, and M. Ekelid (1995). Speech recognition with primarily temporal cues, *Science*, vol. 270, pp. 303-304.

R.R. Coifman, M.V. Wickerhauser, B*est adapted wave packet bases, Numerical Algorithm*. Research Group, Dept. of Mathematics, Yale University, New Haven, Connecticut

R.R.Coifman, M.V. Wickerhauser (1992). Entropy based algorithms for best basis selection, *IEEE Trans. on Information Theory*, vol. 38, part 2.

S. Mallat (1999). *A tour of Signal processing*. Academic Press.

Saeed V. Vaseghi (2000). *Advanced Signal Processing and Noise Reduction, 2nd Edition*, John Wiley & Sons.

Schneier, B. *Applied Cryptography, Second edition*, page 234. John Wiley and Sons.

Shannon, Claude E. (January 1951). Prediction and entropy of printed English, *The Bell System Technical Journal*, 30:50-64.

Sheikhzadeh (2001). *An improved wavelet-based speech enhancement system*. Eurospeech.

Stein, S. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist*. 9(6): 1135-1151.

Tye-Murray, N. (2004). *Foundations of aural rehabilitation (2nd Edition)*. United States: Thomson Delmar Learning, Inc.

Wilson B. (1993). *Signal processing in Cochlear Implants: Audiological Foundations. (R. Tyler, ed.)*. pp. 35-86, Singular Publishing Group, Inc.

W. House (1985). *A personal perspective on cochlear implants, in Cochlear implants (R. Schindler and M. Merzenich, eds.)*. pp. 13-16, New York: Raven Press.

W. House and J. Urban (1973). Long term results of electrode implantation and electronic stimulation of the cochlea in man. *Annals of Otology, Rhinology and Laryngology*, vol. 82, pp. 504-517.

W. House and K. Berliner (1982). Cochlear implants: Progress and perspectives. *Annals of Otology, Rhinology and Laryngology*, vol. (Suppl. 91). pp. 1-124.

W. Nogueira, A. Giese, B. Edler, A. Buchner (2006). *Wavelet Packet Filterbank for Speech Processing Strategies in Cochlear Implants*, ICASSP.

W. H.; Flannery, B. P.; Teukolsky, S. A. and Vetterling, W. T (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed*. Cambridge, England: Cambridge University Press, pp. 554-556.

Cochlear Implant, 2009, Retrieved September 2009 from Wikipedia http://en.wikipedia.org/wiki/Cochlear_implant

Xiaolong Yuan (2003). *Auditory Model-based Bionic Wavelet Transform*. Speech and Signal Processing Lab Milwaukee, Wisconsin

APPENDIX

Matlab Code for New Speech Processing Method

```
clear all;
phrase=[''];
[file,path]=uigetfile('*.*',phrase);
infile=[path file];
[wavedata,fs,bits] = wavread(infile);
subband = 22;
dlength = length(wavedata);
block=8e-3*fs;
x=0;
z=floor(block);
S=1;
csdata = zeros(dlength,1);
while(S<=ceil(dlength/z));
   if (z+x) > dlength
        xbuffer=wavedata(x+1:dlength);
      [A,B] = ProcessSignal(xbuffer, fs);
      csdata(x+1:dlength) = A;
      numberOfElectode(S,:) = B;
   else
     xbuffer=wavedata(x+1:z+x);
      [A,B] = ProcessSignal(xbuffer, fs);
      csdata(x+1:z+x) = A;
      numberOfElectode(S,:) = B;
   end
     x=x+z;
     S=S+1;
```

```
end %end of while loop
subplot(2,1,1). plot(wavedata); title('Original Signal')
ylabel('Magnitude');
subplot(2,1,2). plot(csdata); title('Constructed Signal')
xlabel('Discrete Index')
sound(csdata,fs)
wavwrite(csdata, fs, 'C:\ \deneme.wav');
```

```
function [rsig,numberOfElectode]= ProcessSignal(wavedata,fs)
T = wpdec(wavedata,8,'db10','log energy');
[T,E,N] = besttree(T);

MAP = GenerateMAP;
N = allnodes(T);
electrodes = zeros(22,length(wavedata));

for i = 1 : length(MAP)
    for j = 1 : length(N)
        if MAP(i,2) == N(j)
            temp = wprcoef(T,N(j));
            electrodes(MAP(i,1).:) = electrodes(MAP(i,1).:)+temp';
        end
    end
end

for i = 1 : 22
   Ent(i) = mean(electrodes(i,:).^2);
end
```

```
rsig = zeros(1,length(wavedata));
p = 1;
numberOfElectode = 0;
for k=1:6
    if (max(Ent) ~= 0)
        index = find(Ent == max(Ent));
        rsig = rsig + electrodes(index,:);
        Ent(index) = 0;
        numberOfElectode(p) = index;
        p = p+1;
    end
end
```

Matlab Code for the Traditional N of M speech Processing Method

```
clear;                                  % clear all variables
                                        % sound sampling rate

channel=22;                             % number of channels
phrase='Load *.wav File';
[file,path]=uigetfile('*.*',phrase);
infile=[path file];
[wavedata,srate,bits]=wavread(infile);
wavedata=wavedata-mean(wavedata);
srate=22050;
N=length(wavedata);                     % length of sound record
dt=1/srate;                             % sampling interval
df=1./(dt.*N);                          % frequency interval
fmax=df.*N./2;                          % maximum frequency
d=.5*srate;                             % frequency scalar
t=(1: N)*dt-dt;                         % time array
f=(1: N)*df-df;                         % frequency array
f1=(0:511)*d/512;                       % frequency array for sampled spectrum
Tmax=N*dt;                              % time length of sound
record

block=6e-3*srate;                       %4ms time block
tfout(1:N)=0;                           %array used for output
%------------------------PLOT------------
figure(1)
subplot(2,1,1).plot(t, wavedata);
title('Original Time Waveform');
xlabel('Time (sec)');           % label x axis
```

```
ylabel('Amplitude');                    % label y axis
axis([0 Tmax min(wavedata) max(wavedata)]);   % set axis limits

subplot(2,1,2).specgram(wavedata,[],srate,[],[]);
title('Spectrogram of Original Sound');

%-----------------------PLOT-----------

sound(wavedata, srate);                        % First time

[b1, a1]=cheby2(3,30,[50/d 450/d]);                         % First channel
[b2, a2]=cheby2(3,30,[250/d 650/d]);               % Second channel
[b3, a3]=cheby2(3,30,[450/d 850/d]);               % Third channel
[b4, a4]=cheby2(3,30,[650/d 1050/d]);              % Fourth channel
[b5, a5]=cheby2(3,30,[850/d 1250/d]);              % Fifth channel
[b6, a6]=cheby2(3,30,[1050/d 1450/d]);             % Sixth channel
[b7, a7]=cheby2(3,30,[1250/d 1650/d]);             % Seventh channel
[b8, a8]=cheby2(3,30,[1450/d 1900/d]);             % Eighth channel
[b9, a9]=cheby2(3,30,[1650/d 2150/d]);             % Ninth channel
[b10, a10]=cheby2(3,30,[1900/d 2500/d]);           % Tenth channel
[b11, a11]=cheby2(3,30,[2150/d 2900/d]);           % Eleventh channel
[b12, a12]=cheby2(3,30,[2500/d 3300/d]);           % Twelfth channel
[b13, a13]=cheby2(3,30,[2900/d 3800/d]);           % Thriteenth channel
[b14, a14]=cheby2(3,30,[3300/d 4500/d]);           % Fourteenth channel
[b15, a15]=cheby2(3,30,[3800/d 5400/d]);           % Fifteenth channel
[b16, a16]=cheby2(3,30,[4500/d 6300/d]);           % Sixteenth channel
[b17, a17]=cheby2(3,30,[5000/d 7000/d]);           % Seventeenth channel
```

```
[b18, a18]=cheby2(3,30,[6300/d 7500/d]);        % Eightteenth channel
[b19, a19]=cheby2(3,30,[7000/d 7800/d]);        % Nineteenth channel
[b20, a20]=cheby2(3,30,[7500/d 8300/d]);        % Twentyteenth channel
[b21, a21]=cheby2(3,30,[7800/d 8700/d]);        % Twenty-first channel
[b22, a22]=cheby2(3,30,[8300/d 10000/d]);       % Twenty-second channel



[blow, alow]=cheby2(8,30,200/d); %envelope filter



x=0;
z=floor(block);
S=1;
while(S<floor(N/z));
    xnbuffer(x+1:z+x)=wavedata(x+1:z+x).*hanning(z);

    out1n=filter(b1, a1, xnbuffer(x+1:z+x));
    out2n=filter(b2, a2, xnbuffer(x+1:z+x));
    out3n=filter(b3, a3, xnbuffer(x+1:z+x));
    out4n=filter(b4, a4, xnbuffer(x+1:z+x));
    out5n=filter(b5, a5, xnbuffer(x+1:z+x));
    out6n=filter(b6, a6, xnbuffer(x+1:z+x));
    out7n=filter(b7, a7, xnbuffer(x+1:z+x));
    out8n=filter(b8, a8, xnbuffer(x+1:z+x));
    out9n=filter(b9, a9, xnbuffer(x+1:z+x));
    out10n=filter(b10, a10, xnbuffer(x+1:z+x));
    out11n=filter(b11, a11, xnbuffer(x+1:z+x));
    out12n=filter(b12, a12, xnbuffer(x+1:z+x));
    out13n=filter(b13, a13, xnbuffer(x+1:z+x));
```

```
  out14n=filter(b14, a14, xnbuffer(x+1:z+x));
  out15n=filter(b15, a15, xnbuffer(x+1:z+x));
  out16n=filter(b16, a16, xnbuffer(x+1:z+x));
  out17n=filter(b16, a16, xnbuffer(x+1:z+x));
  out18n=filter(b16, a16, xnbuffer(x+1:z+x));
  out19n=filter(b16, a16, xnbuffer(x+1:z+x));
  out20n=filter(b16, a16, xnbuffer(x+1:z+x));
  out21n=filter(b16, a16, xnbuffer(x+1:z+x));
  out22n=filter(b16, a16, xnbuffer(x+1:z+x));

rout1n=abs(out1n);
low1n=filter(blow, alow, rout1n);
rout2n=abs(out2n);
low2n=filter(blow, alow, rout2n);
rout3n=abs(out3n);
low3n=filter(blow, alow, rout3n);
rout4n=abs(out4n);
low4n=filter(blow, alow, rout4n);
rout5n=abs(out5n);
low5n=filter(blow, alow, rout5n);
rout6n=abs(out6n);
low6n=filter(blow, alow, rout6n);
rout7n=abs(out7n);
low7n=filter(blow, alow, rout7n);
rout8n=abs(out8n);
low8n=filter(blow, alow, rout8n);
rout9n=abs(out9n);
low9n=filter(blow, alow, rout9n);
rout10n=abs(out10n);
low10n=filter(blow, alow, rout10n);
```

```
rout11n=abs(out11n);
low11n=filter(blow, alow, rout11n);
rout12n=abs(out12n);
low12n=filter(blow, alow, rout12n);
rout13n=abs(out13n);
low13n=filter(blow, alow, rout13n);
rout14n=abs(out14n);
low14n=filter(blow, alow, rout14n);
rout15n=abs(out15n);
low15n=filter(blow, alow, rout15n);
rout16n=abs(out16n);
low16n=filter(blow, alow, rout16n);
rout17n=abs(out17n);
low17n=filter(blow, alow, rout16n);
rout18n=abs(out18n);
low18n=filter(blow, alow, rout16n);
rout19n=abs(out19n);
low19n=filter(blow, alow, rout16n);
rout20n=abs(out20n);
low20n=filter(blow, alow, rout16n);
rout21n=abs(out21n);
low21n=filter(blow, alow, rout16n);
rout22n=abs(out22n);
low22n=filter(blow, alow, rout16n);



En(1)=sum(low1n.^2);
  En(2)=sum(low2n.^2);
  En(3)=sum(low3n.^2);
```

```
   En(4)=sum(low4n.^2);
   En(5)=sum(low5n.^2);
   En(6)=sum(low6n.^2);
   En(7)=sum(low7n.^2);
   En(8)=sum(low8n.^2);
   En(9)=sum(low9n.^2);
   En(10)=sum(low10n.^2);
   En(11)=sum(low11n.^2);
   En(12)=sum(low12n.^2);
   En(13)=sum(low13n.^2);
   En(14)=sum(low14n.^2);
   En(15)=sum(low15n.^2);
   En(16)=sum(low16n.^2);
En(17)=sum(low17n.^2);
En(18)=sum(low18n.^2);
En(19)=sum(low19n.^2);
En(20)=sum(low20n.^2);
En(21)=sum(low21n.^2);
En(22)=sum(low22n.^2);

Ebuf = En; %buffuring En



   kl = 1;
    for k=1:6   % select max six output
            find=max(En);
            for p=1:channel
                    if find==En(p)
                    En(p)=0;
         index(S,k) = p;
```

```
            kl = kl+1;
              end
              end
       end


   fout1n=0;fout2n=0;fout3n=0;fout4n=0;fout5n=0;fout6n=0;fout7n=0;fout8n=0;
%clear output
   fout9n=0;fout10n=0;fout11n=0;fout12n=0;fout13n=0;fout14n=0;fout15n=0;fout16
n=0;


      if En(1)==0; %construct processed speech
        fout1n=out1n;
        end

        if En(2)==0;
          fout2n=out2n;
      end

        if En(3)==0;
          fout3n=out3n;
      end

        if En(4)==0;
          fout4n=out4n;
      end

        if En(5)==0;
          fout5n=out5n;
```

```
end

if En(6)==0;
    fout6n=out6n;
end

if En(7)==0;
    fout7n=out7n;
end

if En(8)==0;
    fout8n=out8n;
end

if En(9)==0;
    fout9n=out9n;
end

if En(10)==0;
  fout10n=out10n;
end

if En(11)==0;
    fout11n=out11n;
end

if En(12)==0;
    fout12n=out12n;
end
```

```
if En(13)==0;
      fout13n=out13n;
end

if En(14)==0;
      fout14n=out14n;
end

if En(15)==0;
      fout15n=out15n;
end

if En(16)==0;
     fout16n=out16n;
end

if En(17)==0;
   fout17n=out17n;
end

if En(18)==0;
   fout18n=out18n;
end

if En(19)==0;
   fout19n=out19n;
end

if En(20)==0;
   fout20n=out20n;
```

```
        end

        if En(21)==0;
          fout21n=out21n;
        end

        if En(22)==0;
          fout22n=out22n;
        end

          foutn=fout1n+fout2n+fout3n+fout4n+fout5n+fout6n+fout7n+fout8n+fout9n+f
out10n+fout11n+fout12n+fout13n+fout14n+fout15n+fout16n+fout17n+fout18n+fout
19n+fout20n+fout21n+fout22n;%add process

    tfoutn(x+1:z+x)=foutn;
          x=x+z;
          S=S+1;
    end
    index

    amp1=sum(wavedata.^2);
    amp2=sum(tfoutn.^2);
    tfoutn=tfoutn*sqrt(amp1/amp2);

    subplot(2,1,1).plot(tfoutn);
    title('Synthesized Time Waveform');
    xlabel('Time (sec)');
    ylabel('Amplitude');

    subplot(2,1,2).specgram(tfoutn,[],srate,[],[]);
```

```
title('Spectrogram of Synthesized Sound');

sound(tfoutn, srate); %play output speech

wavwrite(tfoutn, srate, 'Please provide directory information here\f16sample_NofM.wav');
```