1-1-2005

Noise reduction in hearing aids using spectral subtraction techniques

Fatos Myftari
Ryerson University

NOISE REDUCTION IN HEARING AIDS USING SPECTRAL SUBTRACTION TECHNIQUES.

by

Fatos Myftari, BSc

A project

presented to Ryerson University

in partial fulfillment

of the requirements for the degree

of

Master of Engineering

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2005

I hereby declare that I am the sole author of this thesis.

I authorize the Ryerson University to lend this project to other institutions or individuals for the purpose of the scholarly research.

Signature

I further authorize Ryerson University to lend this project by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature

Ryerson University requires the signature of all the persons using or photocopying this thesis. Please sign below and give the address and date.

Abstract.

This thesis is concerned with noise reduction in hearing aids. Hearing – impaired listeners and hearing – impaired users have great difficulty understanding speech in a noisy background. This problem has motivated the development and the use of noise reduction algorithms to improve the speech intelligibility in hearing aids. In this thesis, two noise reduction algorithms for single channel hearing instruments are presented, evaluated using objective and subjective tests. The first noise reduction algorithm, conventional Spectral Subtraction, is simulated using MATLAB 6.5, R13.

The second noise reduction algorithm, Spectral Subtraction in wavelet domain is introduced as well. This algorithm is implemented off line, and is compared with conventional Spectral Subtraction. A subjective evaluation demonstrates that the second algorithm has additional advantages in speech intelligibility, in poor listening conditions relative to conventional Spectral Subtraction. The subjective testing was performed with normal hearing listeners, at Ryerson University. The objective evaluation shows that the Spectral Subtraction in wavelet domain has improved Signal to Noise Ratio compared to conventional Spectral Subtraction.

Acknowledgements

I would like to express many thanks to my supervisor, Prof. Sridhar Krishnan, for entertaining my interests in speech signal processing, and giving me the opportunity to explore speech enhancement research. This work would not have been possible without his support. I would like to thank Dr. Henry Luo, Manager of DSP Applications, at Unitron Hearing, for his help and guidance in this project. Special thanks go to SAR members, Karthi and Jim who kindly helped me out in software issues and evaluation of the algorithm.

Fatos Myftari,
Ryerson University, 2005

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The human population of the earth is growing and the age distribution is shifting toward higher ages in the developed countries. According to many surveys, one out of ten people suffers from hearing loss and would benefit from using hearing aids [5]. Hearing loss and treatment of hearing loss are important research topics. Most of the people with hearing loss have mild or moderate hearing loss that can be treated with hearing aids. The hearing aids users would like to have improved speech intelligibility for listening to speech in quiet and noisy environment, in telephone and in different background noise environment.

A major part of the interaction between humans takes place via speech communication. The speech processing systems used to communicate or store speech are usually designed for a noise free environment [10], but in real word environment, the presence of the background interference in the form of the additive background and channel noise drastically degrades the performance of these systems, causing inaccurate information exchange and listener fatigue. Restoring the desired speech signal from the different background interference is amongst the oldest, still elusive goals in speech processing research [8]. The main objective of the noise reduction algorithms is to improve one or more perceptual aspects of speech, such as the speech quality or intelligibility. So far, many researchers and engineers have developed a number of methods to address this problem, but due to complexities of the speech signals, this area of research still poses a considerable challenge [8]. It is difficult to reduce noise without

distorting the speech and thus, the trade off between speech distortion and noise reduction limits the performance of speech enhancement systems.

The noise reduction algorithms in Hearing Aids, single channel system, is the most common scenario and very complex to deal with [9]. The complexity and ease of implementation of any proposed scheme is another important criterion especially, since the majority of the speech enhancement and noise reduction algorithms find applications in real-time portable systems, like cellular phone, hearing aids and hands free kits [28].

The ability to remove noise from speech is reliant upon differences between characteristics of the speech and unwanted noise. A hearing-impaired individual more typically finds that the noisy environment he encounters is of competing conversations or babble speech [27]. The single channel methods of speech enhancement generally attempt to minimize measures of mean square error to improve the ratio of speech signal power to noise power. The spectral subtraction method has been one of the best – known techniques for the noise reduction [4]. Spectral Subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum [4]. This approach generally produces a residual noise commonly called *musical noise* [4].

In this research a Spectral Subtraction in wavelet domain algorithm, is implemented and is compared with a general Spectral Subtraction. The proposed method gives much better results than the conventional method of Spectral Subtraction, based on objective and subjective evaluation. In the proposed algorithm, the SNR (signal to noise ratio) is increased, and on the same time the processed speech is less distorted. The

residual noise, that is very disturbing in conventional SS, is eliminated in the proposed algorithm.

The following is a block diagram of the proposed method:

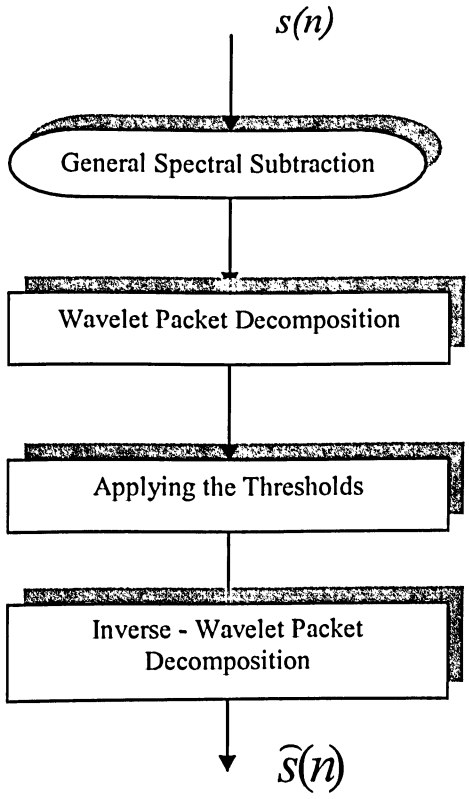


Figure 1.1. *Block Diagram of the Proposed Method.*

The thesis is organized as follows: Chapter 2 gives an overview of hearing aids, various aspects of hearing which are critical in developing hearing aids, and noise reduction algorithms based on human perception along with a review of various noise reduction techniques. Chapter 3 gives an introduction to Wavelet Theory. Chapter 4 discusses the proposed Noise Reduction technique using Spectral Subtraction in wavelet domain. Chapter 5 considers the summary along with the suggested directions for future work.

CHAPTER 2

LITERATURE REVIEW

The noise reduction algorithms for speech communications have been a challenging area for researchers for more than three decades now. The goal of the speech enhancement is to eliminate the additive noise present in speech signal and restore the speech signal to its original form [8]. Most of the methods, known so far, have been developed with some or other auditory model, perceptual or statistical constraints placed on the speech and noise signals [1] [12] [20].

In real word situations, it is very difficult to reliably predict the characteristics of the speech waveform, so as a result the speech enhancement methods are sub-optimal and can only reduce the amount of noise in the signal to some extent, and in the other side some of the speech signal can be distorted during this process. The effectiveness of the speech enhancement system can therefore be measured based on how well it performs the trade off between noise reduction and speech distortion [4] [20].

The speech enhancement systems include improving the performance of: - Digital mobile radiotelephony systems, that suffer from background noise in the environment as well as from channel noise [6]. – Hands free telephone systems suffering from car noise. – Ground air communications where the cockpit/engine noise corrupts the received signal. Hearing Aids applications and cochlear implants in a noisy environment [20].

This chapter starts with an overview of hearing and the physiology of the human ear that are very important in developing Speech Enhancement algorithms for hearing aids. An introduction to the development of hearing aids is presented, and different

speech enhancements methods, with greater emphasis on single channel subtractive type algorithms, are described.

2.1 Hearing.

The human auditory system has an unsurpassed capability to adapt noise and is based on a time-frequency analysis [14]. The information received by the human ears can be described most conveniently as non-linear auditory responses to frequency selectivity and perceived loudness [20]. The general properties of frequency selectivity are related to the concepts of critical band by the assumption that incoming sounds are preprocessed by the peripheral auditory system through a bank of band pass filters [27].

The critical band has a perceptual and a physical relationship with the auditory system, and represents the first approximation of the ear's ability to discriminate different frequencies [20]. Experiments have shown that 25 critical bands exist over the frequency range of human hearing from 20 Hz to 20 kHz as shown in *Table 2.1*.

Critical band analysis is the basis for almost all the models based on the auditory system, and it is the first stage for analysis performed by the inner ear. This analysis is a frequency-domain transformation, which can be seen as a filterbank with bandpass filters. A critical filterbank gives an equal weight to portions of speech with the same perceptual importance [20]. The notion of critical band is related to the phenomenon of masking. The hearing threshold level in quiet environment is a representative of average among the values obtained from different people, and below this level the human ear cannot perceive sound [28].

Masking is a fundamental aspect of the human auditory system and is a basic element of perceptual coding systems [9]. Masking can be defined as either the process

by which the thresholds of audibility of one sound is raised by presence of another sound or the amount by which that threshold is increased.

Sub Band:	Lower edge: (Hz):	Center (Hz):	Upper edge (Hz):
1	*0*	**50**	*100*
2	*100*	**150**	*200*
3	*200*	**250**	*300*
4	*300*	**350**	*400*
5	*400*	**450**	*510*
6	*510*	**570**	*630*
7	*630*	**700**	*770*
8	*770*	**840**	*920*
9	*920*	**1000**	*1080*
10	*1080*	**1170**	*1270*
11	*1270*	**1370**	*1480*
12	*1480*	**1600**	*1720*
13	*1720*	**1850**	*2000*
14	*2000*	**2150**	*2320*
15	*2320*	**2500**	*2700*
16	*2700*	**2900**	*3150*
17	*3150*	**3400**	*3700*
18	*3700*	**4000**	*4400*
19	*4400*	**4800**	*5300*
20	*5300*	**5800**	*6400*
21	*6400*	**7000**	*7700*
22	*7700*	**8500**	*9500*
23	*9500*	**10500**	*12000*
24	*12000*	**13500**	*15500*
25	*15500*	**18000**	*20000*

Table 2.1 *Critical Bands.*

Masking occurs because the auditory system is not able to differentiate two signals close in frequency or in time [14]. Loudness is another important attribute of auditory perception in terms of which sounds can be ranked on a scale extending from quiet to loud. These aspects of human auditory system such as critical band structure, masking,

absolute threshold and excitation patterns have been applied in speech coding, speech recognition and speech quality evaluation [4] [8].

2.2 Noise Characteristics

Noise may be defined as any unwanted signal that interferes with the communication, measurement or processing of an information – bearing signal. Noise is present in various degrees in almost all environments. The success of a noise processing method depends on its ability to characterize and model the noise process, and to use the noise characteristics advantageously to differentiate the signal from noise [4].

Noise can be classified depending on its source or depending on its frequency characteristics.

Source classification of noise is as follow:

- Acoustic noise: noise caused from moving, vibrating, or colliding sources and is the most familiar type of the noise present in various degrees in everyday environment.
- Electromagnetic noise: present at all frequencies and in particular at the radio frequencies. All electric devices, such as radio and television transmitter and receivers generate electromagnetic noise.
- Electrostatic noise: generated by the presence of a voltage with or without current flow. Fluorescent lighting is one of the more common sources of the electrostatic noise.
- Channel distortion, echo and fading: due to non-ideal characteristics of the communication channel. Radio channels, such as those at microwave frequencies

used by cellular mobile phone operators, are particularly sensitive to the propagation characteristics of the channel environment.

- Processing noise: the noise that results from the digital/analog processing of the signals, e.g. quantisation noise in digital coding of speech or image signals.

Depending on its frequency or time characteristics, a noise can be classified as [4]:

- Narrowband noise: a noise process with a narrow bandwidth such as 50/60 Hz 'hum' from the electricity supply.
- White noise: purely random noise that has a flat power spectrum. White noise theoretically contains all the frequencies in equal intensity. White noise is defined as an uncorrelated noise process with equal power at all frequencies.
- Band – limited white noise: a noise with a flat spectrum and a limited bandwidth that usually covers the limited spectrum of the device or the signal of the interest.
- Colored noise: non-white noise or any wideband noise whose spectrum has a non-flat shape; examples are pink noise, brown noise and autoregressive noise.
- Impulsive noise: consists of short duration pulses of random amplitude and random duration
- Transient noise pulses: consists of relatively long duration noise pulses

To model noise accurately, we need a structure for modeling both the temporal and the spectral characteristics of the noise. Accurate modeling of noise statistics is the key to high-quality noisy signal classification and enhancement.

2.3 The peripheral auditory system.

A healthy hearing system organ is an impressive sensory organ that can detect tones between 20 Hz and 20 kHz and has a dynamic range of about 100 dB between the

hearing threshold and the level of the discomfort [27]. The hearing organ comprises the outer ear, the middle ear and the inner ear (Figure 2.1). The outer ear including the ear canal works as a passive acoustic amplifier. The sound pressure level at the eardrum is 5 – 20 dB higher than the free field sound pressure level outside the outer ear (2 kHz -5 kHz). Due to the shape of the outer ear, sounds coming approximately from in front of the head are amplified more than sound coming from other directions.

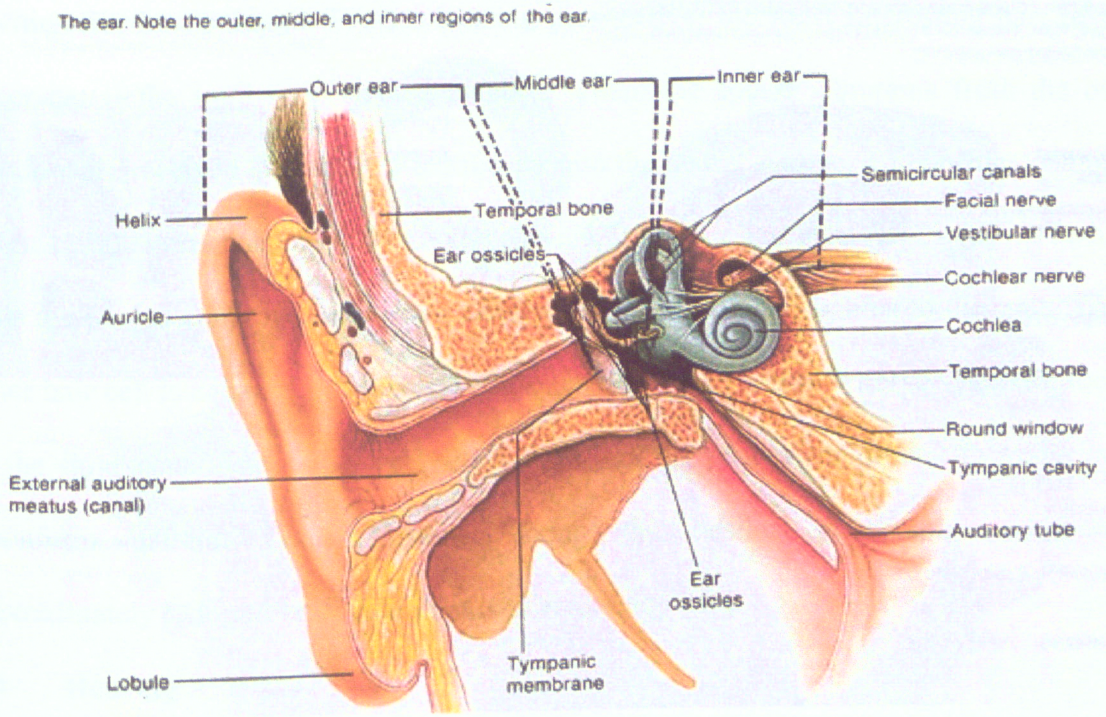


Figure 2.1. *Overview of the Peripheral auditory system (Courtesy of Tissues and Organs: A Text – Atlas of Scanning Electron Microscopy, by R.G Kessel and R.H Kardon, 1979)*

The task of the *middle ear* is to work as an impedance converter and transfer vibrations in the air to the liquid in the inner ear. The middle ear consists of three bones; malleus (that is attached to eardrum), incus and stapes, (that are attached to the oval

window). Acting primarily as levers performing an impedance matching transformation (from the air outside the eardrum to the fluid cochlea), they also protect against very strong transient sounds. The acoustic reflex activates the middle ear muscles, to change the type of motion of the ossicles when low-frequency sounds with Sound Pressure Level (SPL) above 85-90 dB reach the eardrum [27]. Ossicle acts as a low-pass filter with a cutoff frequency around 1000 Hz.

The inner ear is a system called the osseus, or the bony labyrinth, with canal and cavities filled with liquid. From a hearing perspective, the most interesting part of the inner ear is the snail shell shaped cochlea, where the sound vibrations from the oval window are received and transmitted further into the neural system. The cochlea contains about 12,000 outer hair cells (OHC) and about 3,500 inner hair cells (IHC) [27]. The hair cells are placed along a 35 mm long area from the base of the cochlea to the apex. Each inner hair cell is connected to several neurons in the main auditory nerve. The vibrations in the fluid generated at the oval window causes the basilar membrane to move in a waveform pattern. Linear distance along the basilar membrane corresponds approximately to logarithmic increments of frequency [27].

2.4 Hearing Aids.

There are two types of hearing loss: conductive, sensorineural or mixed hearing loss. They can appear isolated or simultaneously. A problem that causes a hearing loss outside the cochlea is called a conductive hearing loss, and damage to the cochlea or the auditory nerve is referred to as a sensorineural hearing loss. A conductive loss causes a deteriorated impedance conversion between the eardrum and the oval window in the middle ear. This non-normal attenuation in the middle ear is linear and frequency

dependant and can be treated automatically. A more problematic impairment is the sensorineural hearing loss. This includes damage to the inner and outer hair cells or the abnormalities of the auditory nerve.

Damage to the outer hair cells causes changes in the input /output characteristics of the basilar membrane movement resulting in a smaller dynamic range. This is the main reason for using automatic gain control in the hearing aids. According to the statistics, 40 percent of the adults between 65 and 75 years have a hearing loss, and 45 to 50 percent of the people at age of 75 and older have a hearing loss [10]. Hearing aids are used to help people with hearing loss issues. The hearing aid detects the sound signals in the acoustical environment through one or more microphones.

The acoustic sound signals consist of the speech signal uttered by the speaker to whom the hearing aid user is listening, the interference background noises (e.g. train in a subway station, restaurant environment, traffic noise) as well as possible signal leakage from the hearing aid loudspeaker to the microphone, also referred to as acoustic feedback. The recorded microphone signals are processed by some signal processing algorithms, with the aim of compensating for the hearing impairment of the hearing aid user. The processed sound is then applied to the ear canal through a loud speaker known as receiver. Hearing aids assume that the inner hair cell function is still in tact. The goal in fitting a hearing aid is an improvement in the overall quality of life. This is a general block diagram of a digital hearing aid [8]:

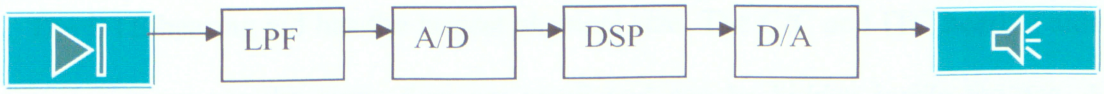


Figure 2.2 – *Digital Hearing Aid*

The most important requirements for the modern high performance of the hearing aids are:

- Low physical area consumption (size).
- Low power dissipation.

There are four common types of hearing aids:

- in the canal (ITC)
- completely in the canal (CIC)
- in the ear (ITE)
- behind the ear (BTE)

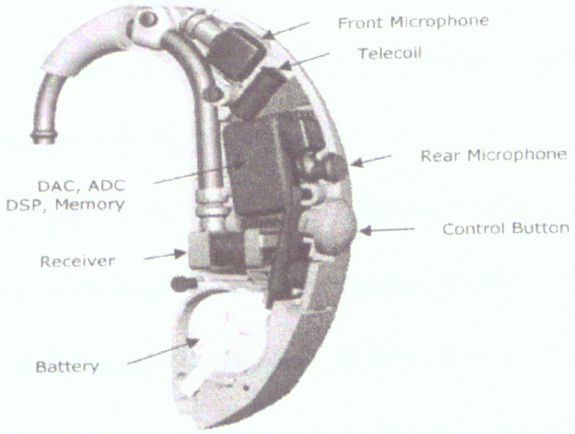


Figure 2.3 *Overview of the components included in a modern BTE digital hearing aid (Courtesy of Unitron Hearing).*

The BTE hearing aid has the largest physical size. The CIC and ITC hearing aids are becoming more popular since they are small and can be hidden inside the ear. A hearing aid comprises at least a microphone, an amplifier, a receiver, an ear mold and a battery. Modern hearing aids are digital also, and include an analog to digital converter (ADC), a digital signal processing (DSP), a digital to analog converter (DAC). Larger models as BTE, ITE and ITC include a telecoil in order to work in locations with induction-loop support (telephone) [10]. Some hearing aids have directionality features, which mean that a directional microphone or several microphones are used. Various signal processing techniques are used in hearing aids.

Automatic gain control or compression is a central technique for compensating a sensorineural hearing loss. The modern hearing aids use an improved technique called Wide Dynamic Range Compression (WDRC) – where more gain is applied to low intensity signals than to high intensity signals [8]. Nowadays, hearing aids usually include feedback suppression, multi - band automatic gain control, adaptive beam forming, and other noise reduction algorithms.

The general opinion is that single channel noise reduction, where only one microphone is used, mainly improves sound quality, while effects on speech intelligibility are usually negative. A multi channel noise reduction approach where two or more microphones are used, can improve both sound quality and speech intelligibility. In this research single channel noise reduction was considered, since in the small size hearing aids as ITC, CIC that are more in demand, it is difficult to employ two microphones.

2.5 Acoustic feedback in hearing aids.

In addition to the difficulty of understanding the speech in noise, the occurrence of acoustic feedback poses a major problem to hearing aid users. Acoustic feedback refers to the acoustical coupling between the loudspeaker and the microphone of the hearing aid. As a result of this coupling, the amplified sound sent through the loudspeaker is fed back into the microphones, and the hearing aid produces severe distortion of the desired signal and an annoying howling sound when the gain is increased.

The use of an amplification that brings the system close to instability indicates that the hearing aid user desires more amplification from the hearing aid than actually provided, and as a result of this gain limitation, low-energy signals fall below the hearing threshold so the instrument does not compensate for hearing loss in the patient. The acoustic feedback stems from the vent, i.e., the hole in the earmold of the hearing aid: depending on the size of the vent, venting establishes an acoustic feedback path that limits the maximum stable gain in a hearing aid to 40 dB and often even less [9]. This effect refers to the increase in loudness of the own voice and the low frequency boost that hearing aid users experience when the ear canal is completely blocked with an earmold.

The unnatural perception of their own voice while talking is disturbing to most hearing aid users and is often a reason to stop wearing their hearing aids. Eliminating the vent or considerably reducing its size is not an acceptable solution for the acoustic feedback problem. Actually, in hearing aid industry there is even a growing tendency towards hearing aids with an open fitting (i.e., without an earmold) to improve listening comfort and binaural hearing (e.g., Phonak, Unitron, GN Resound).

Beside the vent size, also the geometric configuration of the hearing aid, the ear canal and the acoustic outside the ear determine the feedback path. Because of the shorter distance between the loudspeaker (receiver) and microphones, the attenuation of the feedback path is smaller for in- the -ear (ITE) and in-the-canal (ITC) hearing aids than for the behind-the -ear (BTE) model. Since the ear canal shape differs among hearing users, the feedback path is user-dependent [33]. People with a moderate hearing loss suffer from an increase in hearing threshold of 40 dB to 70 dB, while the maximum stable gain in a hearing aid is limited to at most 40 dB and often even less. As a result, there is a strong need for efficient feedback suppression techniques.

To reduce the negative effects introduced by acoustic feedback (i.e., howling and the limited maximum possible amplification), several techniques have been proposed in the literature. In *Feedforward suppression techniques*, the regular signal-processing path of the hearing aid is modified in such a way that is stable in conjunction with the feedback path as shown in Figure 2.4.

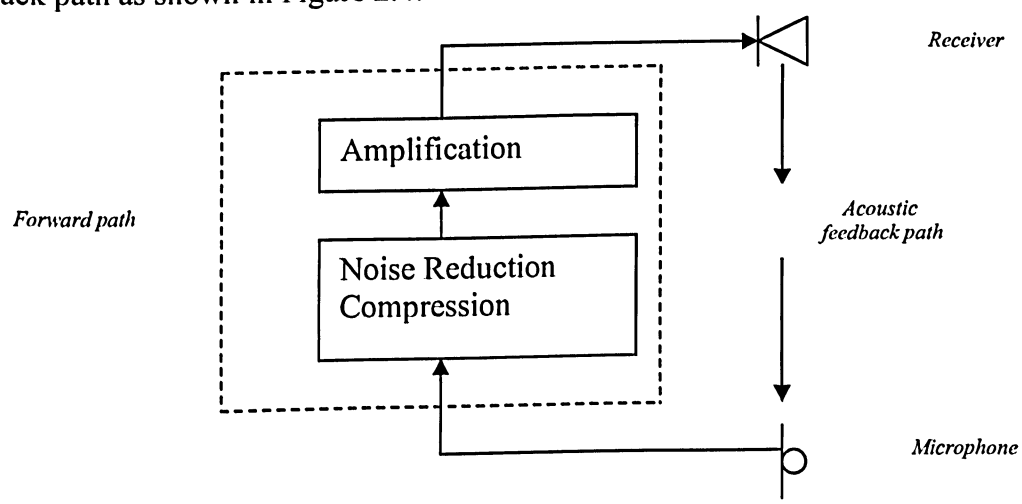


Figure 2.4: *Feedforward Suppression Algorithm.*

The most common technique is the use of a notch filter, where the gain is reduced in a narrow frequency band around the critical frequencies, whenever feedback occurs. Other examples include equalizing the phase of the open-loop response, and using time-varying elements (such as frequency shifting, delay and phase modulation) in the forward path.

The increase in maximum stable gain with *feedforward* suppression techniques has generally been found limited, and on the other side, all compromise the basic frequency response of the hearing aid, so it may seriously affect the sound quality. A more promising solution for acoustic feedback is the use of a *feedback cancellation* algorithm, which is shown in Figure 2.5.

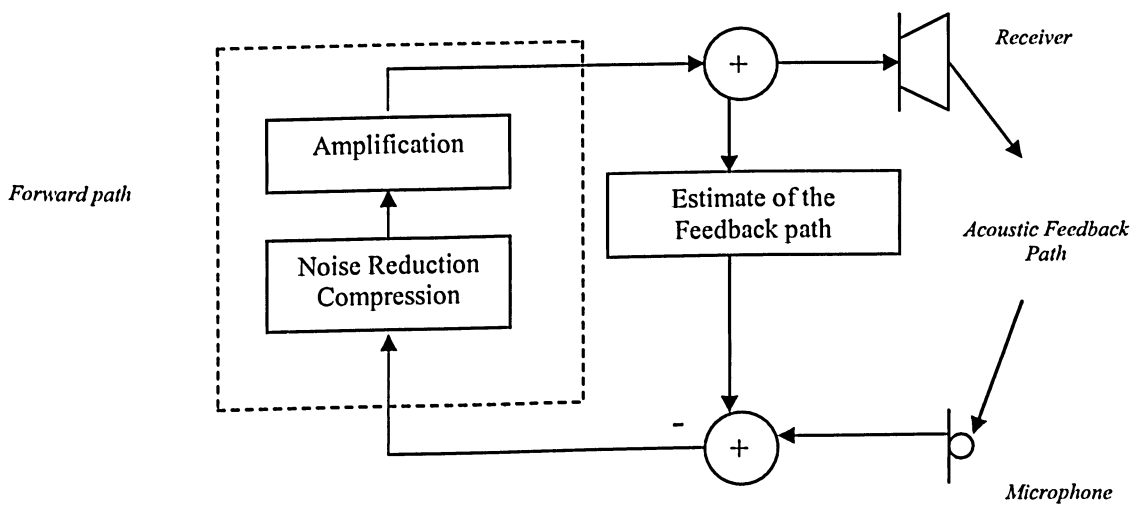


Figure 2.5: *Feedback Cancellation Algorithm.*

The feedback canceller estimates the feedback path signal and subtracts this estimate from the microphone signal, so that ideally, only the desired signal is preserved at the input of the forward path. Since the acoustic path between the loudspeaker and the microphone can vary significantly depending on the acoustical environment, the feedback canceller must be adaptive.

2.6 Noise Reduction Algorithms in Hearing Aids.

The development of noise reduction strategies in hearing aids is guided by the technical limitations of these devices, such as the size, the calculation power of the DSPs and the number of microphones available. The first and the simplest noise reduction algorithm implemented in hearing aids was a highpass filter [45]. It was assumed that the energy of environmental noise was mainly at low frequencies (20 Hz – 300 Hz).

The hearing aid used a fixed highpass filter, which could be manually switched ON or OFF as required. Afterwards, variable high pass filters were used. These filters were adjusted to increasingly reject energy of the low frequencies when this energy increases. In a non-noisy environment, the highpass filter became an all pass filter. Plomp argued that the speech intelligibility for hearing - impaired persons could be improved by a selective expansion or compression of the temporal modulation envelope of the signals under certain conditions [28]. Based on this, a method that uses *filter bank* to separate the input signal of the microphone into different frequency channels was proposed by Clarkson and Bahgat [29].

The modulation frequency in each channel is analyzed to decide whether the signals are more likely to be speech or noise and each frequency range is amplified or attenuated accordingly. High-pass and multiple bandpass filters work in the frequency

domain. On the other hand, directional microphones work on the spatial characteristics of the sounds.

Directional microphones are designed to be most sensitive to sound arriving from the front and try to cancel (or null) noise sounds coming from a specific direction [14]. The directional microphone has two entry ports that allow sound to enter both the front and the rear cavities and arrive on the either side of the microphone diaphragm. If the delayed version of the sound at rear port reaches the diaphragm at the same time as the sound coming from the front port, a cancellation of the sounds occur. The spatial characteristics of the directional microphone depends on this formula:

$$D(\alpha) = 1 + \gamma \cos(\alpha), \qquad (2.1)$$

where γ is the ratio between the external delay (the travel time of the sound due to distance between the two ports of the microphone) and the internal delay. The spatial characteristics of the directional microphone for different values of the γ are illustrated in Figures 2.6 and 2.7.

Based on the hardware directional microphone strategy, *software directional microphones* have been developed [30]. This method uses two omni-directional microphones as the front and the rear entry port of the directional microphone. The directional microphone signal is computed as the difference between the signals from the front microphone and the delayed – weighted signal of the rear microphone, resulting in a response comparable to a hardware directional microphone. The microphone parameter is the frequency dependent weight for the rear microphone and the spatial characteristics of the directional microphone varies as a function of the internal delay time and the weight.

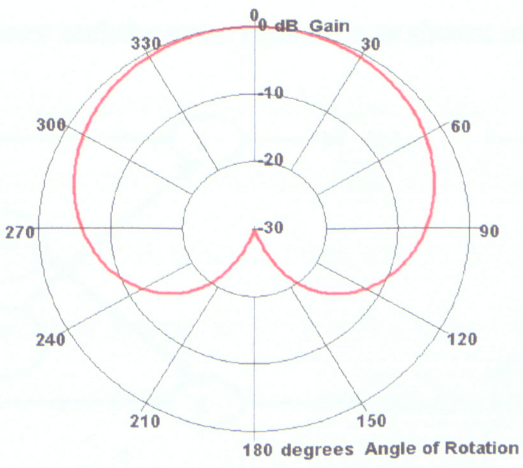


Figure 2.6 *Cardioid spatial characteristics of the hardware directional microphone ($\gamma = 1$).*

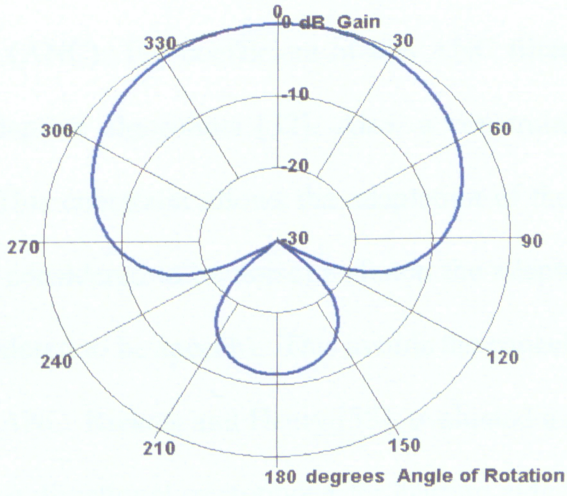


Figure 2.7 *Hypercardioid spatial characteristics of the hardware directional microphone ($\gamma = 3$).*

The *adaptive directional microphone*, currently the state-of-the- art solution in modern, commercial hearing aids, is used at Unitron Liaison, GN Resound Canta,

Phonak Claro. [31]. Two software directional microphones provide reference signals, namely the *speech reference* and the *noise reference* as shown in Figure 2.8.

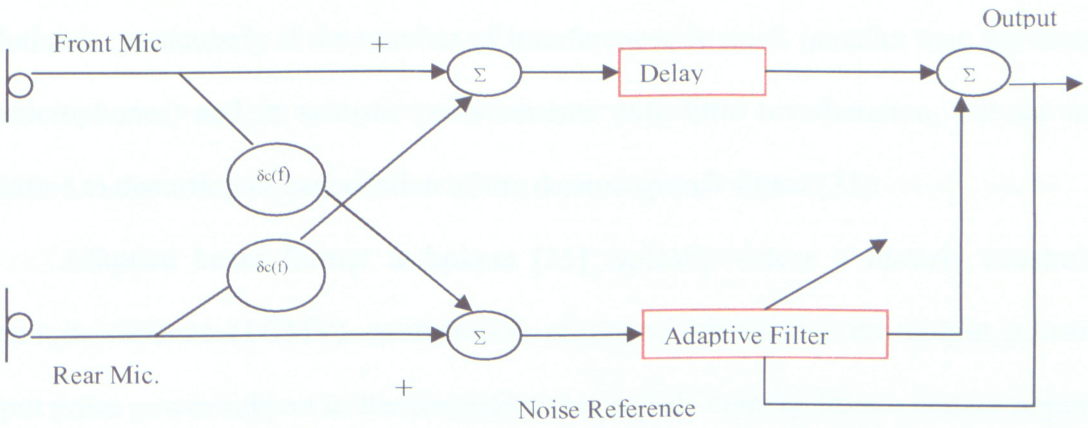


Figure 2.8. *Representation of the adaptive directional microphone.*

The speech reference is created with a front spatial cardioid, (null at 180°) and the noise reference is created with a rear spatial cardioid (null at 0°). The signals of the software directional microphones, speech and noise reference are connected at to an Adaptive Noise Canceller (ANC). The coefficient of this ANC filter can be updated by the means of classical adaptive algorithms [32]. Also, a constraint is applied on the coefficients of the ANC. This constraint allows the adaptation of the coefficient when a source is at the back (it is considered to be noise) and stop the adaptation when a source is at the front (this is considered to be speech). This avoids the cancellation of the speech signal at the output of the ANC. Rickets and Henry [33], evaluated a software directional microphone and an adaptive directional microphone for hearing aids. In this study, it was shown that the advantage of the adaptive, over the fixed directional microphone, was prominent when noise sources are situated on the side of the listener.

The very advanced technique for noise reduction in hearing aids is the so-called *beamformer*. The beamformer is designed to receive a signal radiating from a specific

direction and attenuate the signals radiating from other directions of no interest. Adaptive techniques generally have a better noise reduction performance than fixed beam forming techniques, particularly if the number of interferences is small (smaller than the number of microphones) and in acoustic environments with little reverberation, but are more sensitive to distortion or cancellation of the desired speech signal [33].

Adaptive beam former technique [35] typically solves a linearly constrained minimum variance (LCMV) optimization criterion, minimizing the output power or output noise power subject to the constraint that signals coming from a certain region or direction (i.e., ideally the direction of the desired speech) are preserved. An efficient realization of the LCMV is the generalized side - lobe canceller (GSC) [34]. The GSC transforms the constrained LCMV optimization criterion into an unconstrained criterion through a combination of a fixed spatial pre – processor, i.e., a fixed beam former and a blocking matrix are designed to avoid so-called speech leakage into the noise references.

A physical evaluation of three fixed and two adaptive beamformers was carried out [34]. This evaluation provided useful information for selecting an array – processing algorithm as function of performance. The two first fixed beamformers were the delay-and-sum beamformer and an over steered array. The latter was similar to a delay-and-sum beamformer but the time delays used in combining the microphone output signals were greater than acoustic propagation times between the microphones. The third fixed technique was the optimal super directive beamformer [32]. The two adaptive beamformers were the scaled projection algorithm [35] and an extension of this technique using a composite structured correlation matrix.

The beam former techniques were evaluated for adverse listening conditions in two different anechoic chambers. It appeared that the single cardioid microphone was more effective than the delay-sum beamformer. The adaptive beamformers results in a better SNR improvement than the fixed beamformer. The adaptive beamformer is successfully used at the latest products developed from Unitron Hearing (Liaison) and Phonak (Claro).

Another relatively new signal processing strategy for noise reduction in hearing aids is *blind source separation (BSS)*. The goal of BSS [34] is to recover independent sources given only sensor observations that are linear mixtures of the independent source signals. The term 'blind' indicates that both the source signals and the way the signals are mixed are unknown. There are two different methods to solve the BSS problem. First one is based on second order statistics, called Principal Component Analysis (PCA), and the second one called Independent Component Analysis (ICA), does not only de-correlate the signals, such as PCA, but also reduces higher-order statistical independences. In speech processing, only PCA approaches were developed or evaluated. Two different types of signal separation can be distinguished, the scalar separation and the convolute separation.

The scalar separation assumes that the mixture between the sources is performed by a static mixing matrix and typical applications areas are the small band signal processing. The convolute separation assumes that the mixture between the signals is done by a transfer function and the sources are broadband like in speech and audio processing. Nguyen-Thi (1992) evaluated both types of the algorithms, scalar and convolute separation, based on a PCA approach. The evaluation was carried out in anechoic condition and the sources, speech and noise, were close to the microphones. The

SNR measure was used to evaluate the algorithms and the strategies based on the convolute separation gave the best separation. Stability and non-unique solutions can occur with the separation algorithms.

A recent research done by [34], introduces a BSS algorithm using frequency domain. This algorithm was implemented in real time on a PC–platform, and the performance of the algorithm was evaluated both by simulations and experimentally, including the separation of a moving and a fixed speaker in an anechoic chamber.

2.7. Spectral Subtraction Algorithm.

Spectral Subtraction (**SS**) is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present [4]. In many applications the only signal that is available is the noisy signal, so it is impossible to cancel out the random noise, but it may be possible to reduce the average affects of the noise on the signal spectrum.

If we assume that $y(n)$, the discrete noise corrupted input signal, is composed of the clean speech signal $s(n)$ and the uncorrelated additive noise signal $d(n)$, then the noisy signal could be represented as:

$$y(n) = s(n) + d(n). \tag{2.2}$$

The assumption that speech is stationary is used in this equation. The process is carried out on a short – time basis (frame by frame), therefore a time – limited window $w(n)$, multiplies the original speech, noise and the speech signal as well.

So the windowed signal could be represented as:

$$y_w(n) = s_w(n) + d_w(n) \tag{2.3}$$

In the frequency domain, with their respective Fourier transforms, the power spectrum of the noisy signal can be represented as:

$$|Y_w(\omega)|^2 = |S_w(\omega)|^2 + |D_w(\omega)|^2 + S_w(\omega) \bullet D^*_w(\omega) + S^*_w(\omega) \bullet D_w(\omega)\,, \tag{2.4}$$

where $D^*_w(\omega)$ and $S^*_w(\omega)$ represent the complex conjugates of $D_w(\omega)$ and $S_w(\omega)$ respectively. The function $|S_w(\omega)|^2$ is referred to as the Short Time Power Spectrum of speech. The DFT of $Y_w(\omega)$ is given by:

$$Y_w(\omega) = \sum_{n=0}^{N-1} y(n) \cdot e^{-j\frac{2\pi\omega n}{N}} = |Y_w(\omega)| e^{j\phi(\omega)}\,, \tag{2.5}$$

where $\varphi(\omega)$ is the phase of the corrupted noisy signal and N is the number of samples in the windowed speech signal. In (2.4) the terms $|D_w(\omega)|^2$, $S_w(\omega) \bullet D^*_w(\omega)$, $S^*_w(\omega) \bullet D_w(\omega)$ cannot be obtained directly and are approximated as $E[|D_w(\omega)|^2]$, $E[S_w(\omega) \bullet D^*_w(\omega)]$ and $E[S^*_w(\omega) \bullet D_w(\omega)]$ where $E[\,]$ denotes the expectation operator. Typically, $E[|D_w(\omega)|^2]$, is estimated during the silence periods, and is calculated as $\left|\hat{D}(\omega)\right|^2$. If we assume that $d(n)$ is zero mean and uncorrelated with $s(n)$, then the terms $E[S_w(\omega) \bullet D^*_w(\omega)]$ and $E[S^*_w(\omega) \bullet D_w(\omega)]$ are reduced to zero. Thus from the above based assumptions, the estimate of the clean speech can be given as:

$$\left|\hat{S}(\omega)\right|^2 = |Y(\omega)|^2 - E\left[|D(\omega)|^2\right] \tag{2.6}$$

From (2.6) can be seen that the spectral subtraction process involves the subtraction of an averaged estimate of the noise from the instantaneous spectrum of the corrupted speech. The estimate $\left|\hat{S}(\omega)\right|^2$ cannot be guaranteed to be non-negative, as the right side can

become negative due to errors estimating the noise. These negative values can either be made positive by changing the sign (full-wave rectification) or can be set to zero (half wave rectification) which is implemented in this algorithm. Once the estimate of the clean speech is obtained in the spectral domain, the enhanced speech signal is obtained according to:

$$\hat{s}(n) = IDFT[|\hat{S}(\omega)| \cdot e^{j\phi(\omega)}] \tag{2.7}$$

The phase information from the corrupted signal is used to reconstruct the time domain signal by taking the IDFT.

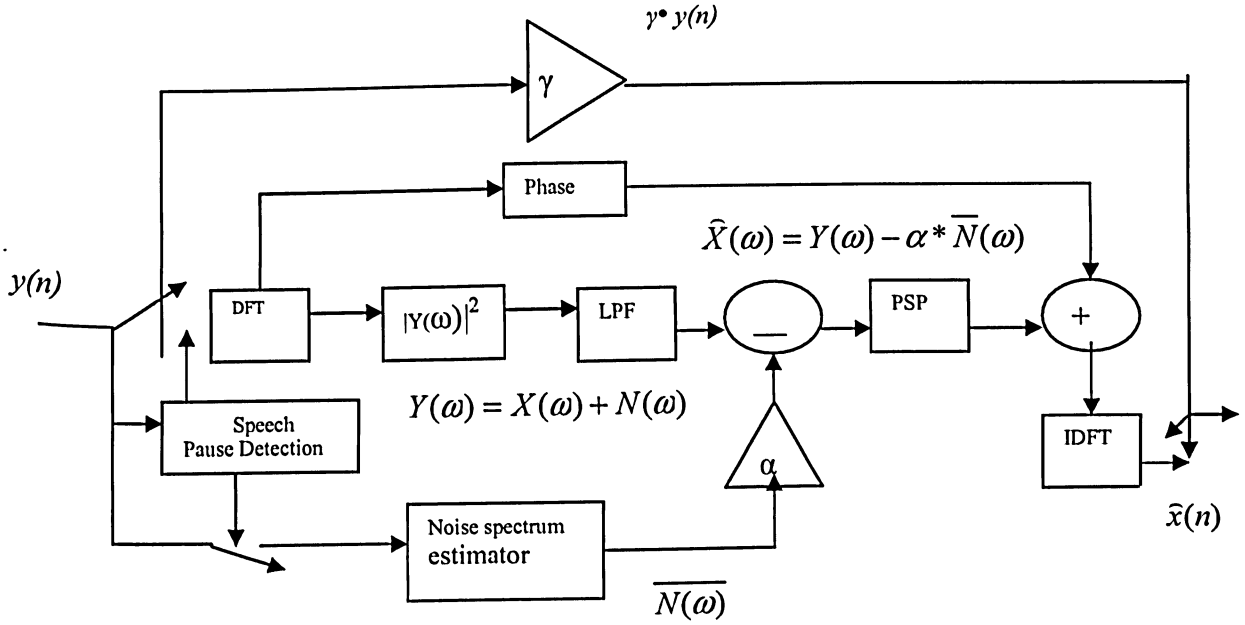


Figure 2.9. *Block diagram of a Spectral Subtraction System. (PSP is a Post Spectral Subtraction Processing).*

By generalizing the exponent in (2.6) can be further written as below:

$$\left|\hat{S}(\omega)\right|^k = \left|Y(\omega)\right|^k - E\left[\left|D(\omega)\right|^k\right] \quad , \tag{2.8}$$

where $\left|\hat{D}(\omega)\right|$ can be obtained via a running average the number of frames taken into account depends on the stationary of the noise, $k = 1$ for magnitude spectrum and $k=2$ for power spectrum. Figure 2.9 presents the Spectral Subtraction implemented in this research.

The DFT - based spectral subtraction is a block-processing algorithm. The incoming audio signal is buffered and divided into overlapping blocks of N samples as shown in Figure 2.9. Each block is Hamming windowed, and then transformed via a DFT to the frequency domain. After spectral subtraction, the magnitude spectrum is combined with the phase of the noisy signal, and transformed back to the time domain.

Each signal block is then overlapped and added to the preceding and succeeding blocks to form the final output [4]. This algorithm was implemented offline, using MATLAB 6.5 R13, and the database from Speech Enhancement and Assessment Resource (SpEAR), from Oregon Health and Science University. The most commonly window length of 20 mS was used. First 100 mS were considered to be noise; the smoothing factor α was selected 0.05 as the optimal one.

One advantage of the single-microphone noise reduction algorithm, compared to multi-microphone methods, is their robustness against the number of noise sources and the level of reverberation. The main disadvantage in this method is the presence of processing distortions caused by the random variations of the noise.

CHAPTER 3

WAVELET THEORY

Wavelet transforms and multiresolution are topics that have been studied by mathematicians, scientists, and engineers. However, it was not until the early 1980's that connections were drawn across the fields. Most of the speech enhancement techniques discussed so far, are based on the spectral information obtained through the short time Fourier analysis of the speech signal [36]. These are all frequency-based methods intending to preserve the slow-varying short time spectral characteristics of the speech quality after the processing.

The wavelet transforms, a time-frequency analysis, 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 (high time resolution) for the high frequency content of the signal. Wavelets are mathematical function that cut up data into different frequency components, and then study each component with a resolution matched to its scale. The wavelet transform can be regarded as a bank of band-pass filter with constant Q factor (the ratio of the bandwidth and the central frequency) [7] [18] [25].

Wavelet analysis has a distinct ability to detect local features of the signal in time and frequency, such as fine structures of the speech signal and other transient, instantaneous and dynamic speech components that contribute significantly to the quality of speech. This chapter starts with a short explanation of the continuous wavelet transform and its relation to the filter banks, wavelet packet decomposition are

introduced, along with fundamentals of wavelet threshold used for speech de-noising techniques.

3.1 The 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, was first introduced into the signal processing field thanks to the efforts of French mathematicians I. Daubechies and S. Mallat. Analysis with wavelets involves breaking up a signal into shifted and scaled versions of the original (mother) wavelet, and it uses a time-scale region rather than a time-frequency region. The wavelet de-noising techniques, intertwined with multi-resolution and filter bank theory, have been a hot research topic in recent years. The word "wavelet" literally means " a small wave".

A wavelet is a function $\psi(t) \in L^2(\mathrm{R})$ that satisfies the following [7] three conditions:

1. It has an average of zero:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0.$$

2. It is normalized such as:

$$\|\psi(t)\| = 1.$$

3. It is centered in the neighborhood of $t = 0$.

The wavelet function $\psi(t)$ can be scaled by s and translated by u such that:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \tag{3.1}$$

Thus, changes in u cause the wavelet to slide to different points along the time axes, and changes in s stretch on wavelet in time domain. This stretching is often called 'dilating' or 'scaling' because it allows varying resolution at different frequency and time scales. (Note that the factor of $\frac{1}{\sqrt{s}}$ on the right-hand side of Equation 3.1 ensures that the dilated wavelet satisfies the normalization condition $\|\psi(t)\| = 1$). Essentially, s and u determine the time and frequency support of the wavelet function. The time-frequency properties of a wavelet due to translation and dilation are shown in Figure 3.1, where ξ is defined as:

$$\xi = \frac{1}{2\pi}\int_0^{+\infty} \omega \left|\hat{\psi}(\omega)\right|^2 d\omega \tag{3.2}$$

is the center frequency of the wavelet with $s = 1$.

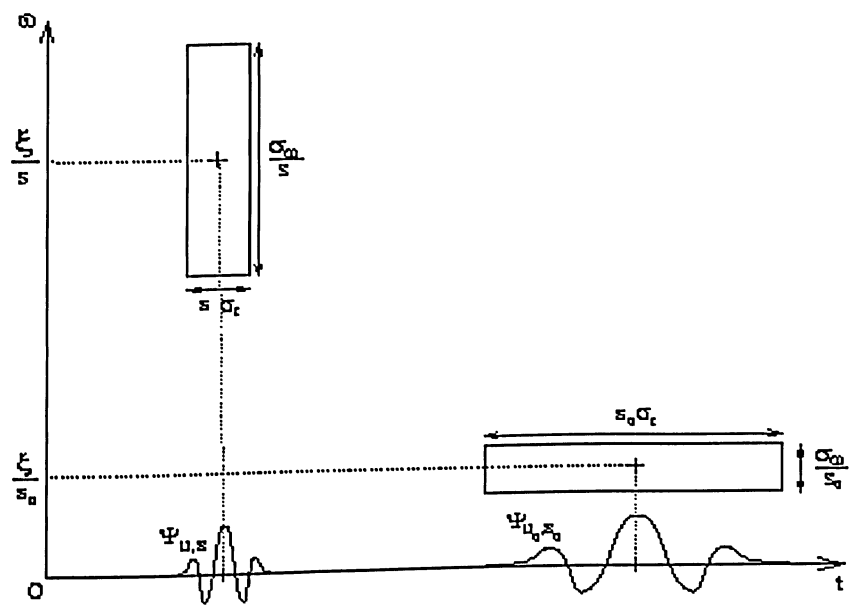


Figure 3.1. *Time-Frequency Effects of Dilating and Translating Wavelets [7]*

When the wavelet is scaled (s is made smaller), then the following happened:

- Its center frequency ξ is shifted higher.
- Its time support decreases, increasing the resolution in the time domain.
- Its frequency support increases, decreasing the resolution in the frequency domain.

Logically, when the wavelet is stretched (s is made larger), then the opposite occurs:

- Its center frequency ξ is shifted lower.
- Its time support increases, decreasing the resolution in the time domain.
- Its frequency support decreases, increasing the resolution in the frequency domain.

The wavelet transform W of a function f, is the inner product of $\psi_{u,s}$, and the function f:

$$Wf(u,s) = < f, \psi_{u,s} > = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt \qquad (3.3)$$

This allows some frequency components of a portion of f to be examined, depending upon the time-frequency spread of the dilated and translated wavelet $\psi_{u,s}$. Wavelet transform is the equivalent of a convolution operation, or linear filtering. Convolution is defined as:

$$h * g(y) = \int_{-\infty}^{+\infty} g(x)h(y-x)dx = \int_{-\infty}^{+\infty} h(x)g(y-x)dx \qquad (3.4)$$

Now, if:

$$\overline{\psi}_s(v) = \psi_s^*(-v) = \frac{1}{\sqrt{s}} \psi^*(-\frac{v}{s}),$$

then Equation (3.3) may be written as:

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-s}{s})dt = \int_{-\infty}^{+\infty} f(t)\overline{\psi}_s(u-t)dt = f * \overline{\psi_s}(u) \qquad (3.5)$$

However, a function's wavelet transform alone is an incomplete representation of the signal. What is needed is a complementary transform, which takes some sort of the average, much as the wavelet transform examines the differences or the details.

The function that accomplishes this transform is the scaling function, $\varphi(t)$. The average, or low frequency, approximation of a function f is:

$$Vf(u,s) = < f, \varphi_{u,s} > = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\varphi^*(\frac{t-u}{s})dt \qquad (3.6)$$

The same logic could be used as above, to prove that averaging transform could as well be a convolution operation. If:

$$\overline{\varphi}_s(v) = \varphi_s^*(-v) = \frac{1}{\sqrt{s}}\varphi^*(-\frac{v}{s}),$$

then

$$Vf(u,s) = < f, \varphi_{u,s} > = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\varphi^*(\frac{t-u}{s})dt = \int_{-\infty}^{+\infty} f(t)\overline{\varphi}_s(u-t)dt = f * \overline{\varphi_s}(u) \qquad (3.7).$$

Mallat proves that a function can be fully reconstructed using the wavelet transform up to including a certain level s_o, and the low frequency approximation from the same level s_o [7].

$$f(t) = \frac{1}{C\psi}\int_0^{so} Wf(\cdot,s) * \psi_s(t)\frac{ds}{s^2} + \frac{1}{C_\psi s_o} Vf(\cdot,s_o) * \varphi_{so}(t), \qquad (3.8)$$

where

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega \qquad (3.9)$$

and $\bar{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. Equation (3.8) implies that a function's space is simply the sum of the spaces occupied by its wavelet transform and low-frequency transform:

$$\{f(t)\} = \{Vf(\cdot, s)\} \oplus \{Wf(\cdot, s)\}, \tag{3.10}$$

where $\oplus$ is the direct sum of two vector spaces.

An example of a "Mexican hat" wavelet is shown in Figure 3.2.

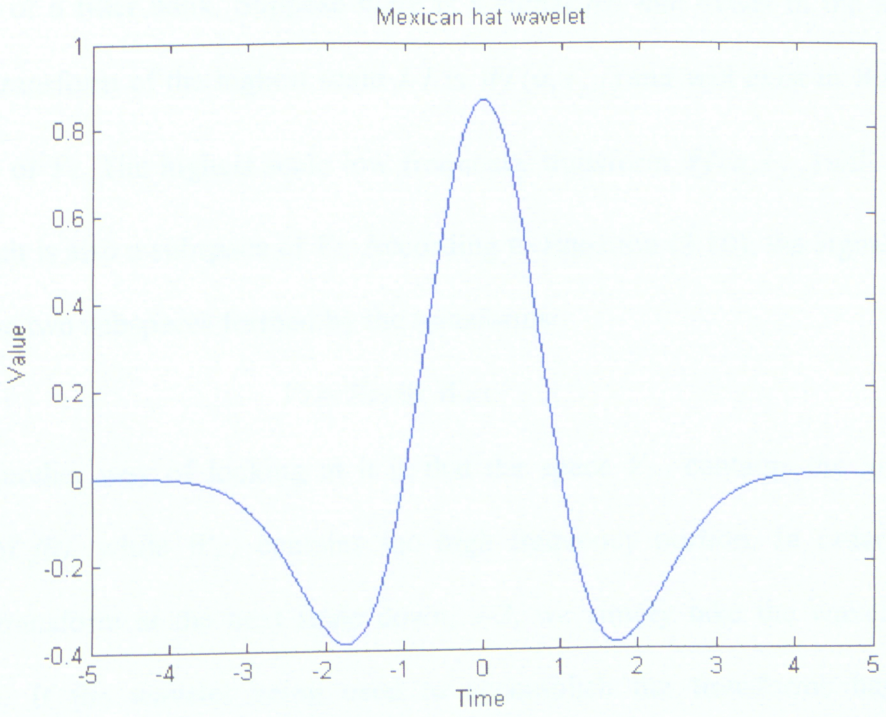


Figure 3.2. *Example wavelet: The Mexican Hat*

It is also important to mention that some wavelets have the useful property of orthogonality. If a wavelet has this property, then the wavelet of one scale will be orthogonal to itself at all other scales, provided the scales used cover different frequency intervals. That is, the wavelet at each scale occupies a space that does not intersect with the spaces of the wavelet at any other scale. This is expressed as:

$$\{\psi_{u,so}(t)\} \not\subset \{\psi_{u,s\neq so}\} \tag{3.11}$$

Spaces occupied by orthogonal wavelet transforms of a function at different levels can also be shown to be orthogonal. This is expressed as:

$$\{Wf(u,s_0)\} \not\subset \{Wf(u,s \neq s_0)\} \tag{3.12}$$

3.2 Multiresolution Filter Banks.

This section describes implementing the wavelet transform in the discrete domain in the form of a filter bank. Suppose there is a signal *f(t)* that exists in the space V_J. The wavelet transform of the highest scale *J-1* is $Wf(u,s_{J-1})$ and will exist in W_{J-1}, which is a subspace of V_J. The highest scale low frequency transform $Vf(u,s_{J-1})$ will than exist in V_{J-1}, which is also a subspace of V_J. According to Equation (3.10), the signal space is the sum of the two subspaces formed by the transforms:

$$V_J = V_{J-1} \oplus W_{J-1}. \tag{3.13}$$

Another way of looking at it is that the space V_{J-1} contains the low frequency portion of *f(t)*, while W_{J-1} contains the high frequency portion. In order to take the wavelet transform at the next scale down, *J-2*, we simply take the wavelet transform $Wf(u,s_{J-2})$. If the wavelet being used to accomplish the transforms happens to be orthogonal, then it is known that the wavelet transform of the function *f(t)* at different scales occupy completely different spaces (Equation 3.12). The space V_{J-1} always contains the remainder of the signal, whatever is not represented by W_{J-1}. So, the *J-2* wavelet transform of the signal *f(t)* is equivalent to the *J-2* wavelet transform of the *J-1* low frequency portion of *f(t)* :

$$Wf(u,s_{J-2}) = Wf_{J-1}(u,s_{J-2}), \tag{3.14}$$

where $f_{J-1}(t) = Vf(u,s_{J-1})$.

The same logic could be used for the wavelet transform at scale s_{J-3}, s_{J-4} and so on. This makes the decomposition of the signal $f(t)$ into wavelet and low frequency spaces at different levels recursive in nature. The recursive decomposition as a binary tree of low frequency spaces V_j and wavelet spaces W_j is shown in Figure 3.3.

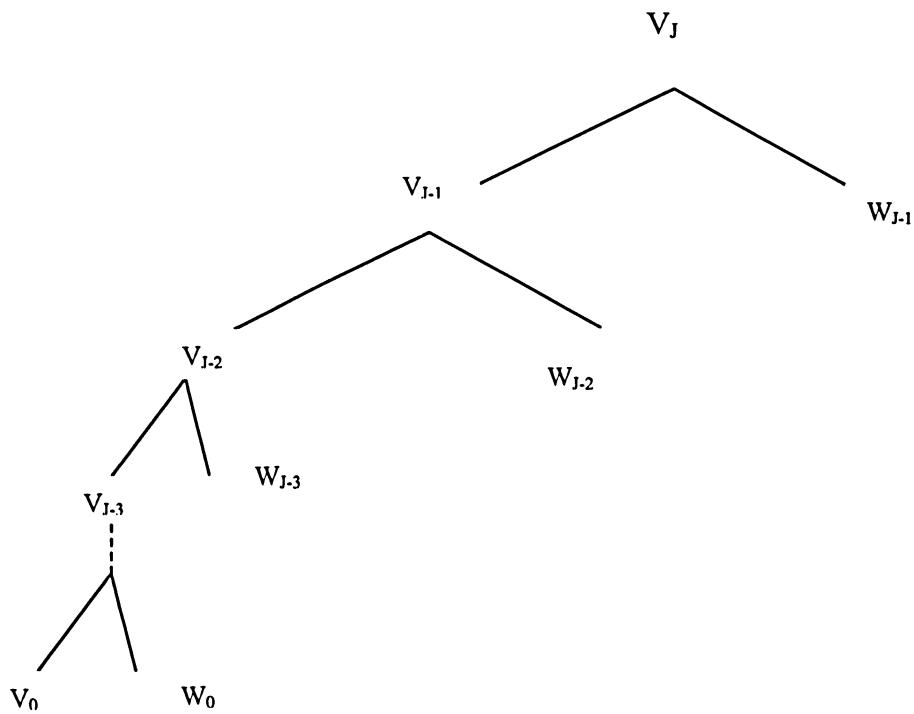


Figure 3.3. *Binary tree of the Wavelet Decomposition Transform.*

The binary tree decomposition follows Equation. 3.8. A signal space is fully represented by the sum of the wavelet spaces at all levels and low frequency space at the lowest level. Thus, the sum of all these spaces are said to be a complete basis for the signal:

$$V_J = V_0 \oplus W_{J-1} \oplus W_{J-2} \oplus W_{J-3} \oplus \ldots\ldots \oplus W_0 = V_0 \oplus \sum_{j=0}^{J-1} Wj \qquad (3.15)$$

In digital domain, this recursive feature proves to be rather convenient. As implied by Equations 3.5 and 3.7, it is possible to perform the wavelet and low frequency transforms via convolution, or linear filtering in the digital domain.

Let *c[k]* and *d[k]* be the digital filters associated with $\varphi(t)$ and $\psi(t)$, respectively; *c[k]* is used to perform the low frequency transforms, and *d[k]* is used to perform the wavelet (high-frequency) transform. Strang [36] gives the relation between filters and the continuous time functions:

$$\varphi(t) = \sum_k c[k]\varphi(2t-k) \tag{3.16}$$

$$\psi(t) = \sum_k c[k]\varphi(2t-k)$$

According to these equations, the filter coefficients depend on the wavelet and scaling functions at a given time resolution t and the wavelet and scaling functions at the next highest time resolution $2t$. This implies that the filter coefficients would have to be re-calculated for each level of the wavelet decomposition. In the digital domain this is avoided by down - sampling the signal at each level.

Down sampling, also known as 'decimation', involves removing every other sample, so that the resulting vector is one half the length of the original. Down-sampling the signal at each level decreases the scale of the signal, which is the equivalent of the decreasing the scale of the filters. This allows the same filters *c[h]* and *d[h]* to be used at each level. The implementation of this recursive filtering and downsampling are known as a filter bank. The schematic of the filter bank shown in Figure 3.5, corresponds to the binary tree of spaces shown earlier in the Figure 3.3. The down sampling operation is shown as $\downarrow 2$ (Figure 3.4)

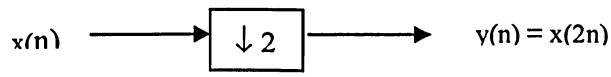


Figure 3.4 - *Downsampler*

The vectors $\breve{v}_j$ and $\breve{w}_j$ indicate the results of the downsampled low frequency and wavelet transforms at level j, and are known as wavelet coefficients. This kind of filter bank is used in digital signal processing systems as a fast implementation of the wavelet transform, and is known as the fast biorthogonal wavelet transform. It is a recursive process of low- and high pass- filtering downsampled signals, and the entire process requires *O(N)* operations, where N is the length of the signal [7]. The fast biorthogonal wavelet transform generates N wavelet and low-frequency coefficients, so that the wavelet representation of a discrete signal occupies more space in computer memory than the original signal.

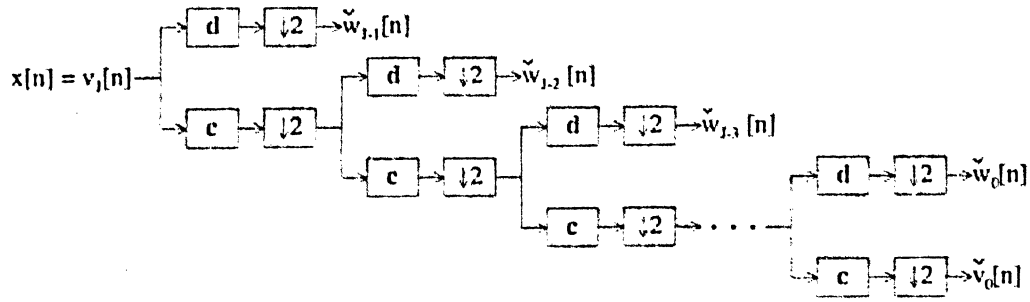
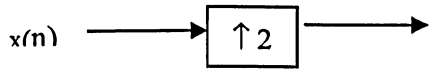


Figure 3.5, *Analysis Filter Bank* (*Courtesy "An introduction to wavelets", Amara Graphs, IEEE, 1995*)

Provided the filters *c* and *d* satisfy conditions for perfect reconstruction, then the original signal can be recovered from the wavelet coefficients. This is accomplished by first upsampling (Figure 3.6) and then low- and high- pass filtering the coefficients at each level.



y(n) = x(n/2) if n even and 0 o.w.

Figure 3.6 *Upsampler.*

Upsampling a vector is done by inserting a zero after every element, making the vector twice as long. Figure 3.7 shows a synthesis filter bank, which would reconstruct the signal, decomposed by the analysis filter bank of Figure 3.5, where *g* and *h* are low- and high- pass filters respectively and ↑*2* denotes upsampling.

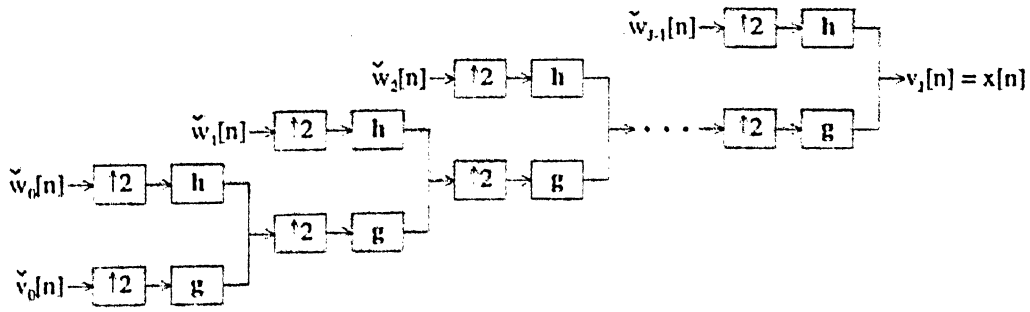


Figure 3.5, *Analysis Filter Bank* (*Courtesy "An introduction to wavelets", Amara Graphs, IEEE, 1995*)

In order to recover the signal perfectly, however, the filters *c, d, g,* and *h* must be specially designed. Strang [36] outlines the requirements as

$$G(z)C(z) + H(z)D(z) = 2z^{-1},$$

$$G(z)C(-z) + G(z)D(-z) = 0, \qquad (3.17)$$

where *C(z), D(z), G(z)* and *H(z)* are the representations of the filters *c, d, g,* and *h* in the *z*-domain.

A convenient way of determining the filters, is by first calculating the low frequency filter *c* from the scaling function $\varphi(t)$, and then using *c* to determine *d, g* and *h*. This section was indented to give an overview of how filter banks are used to implement wavelet transforms of discrete signals, and to help the reader obtain an intuitive feel for how wavelet coefficients represent time –domain signals.

3.3 Wavelet Packets.

The wavelet transform described earlier effectively decomposes the signal into a set of wavelet (or high frequency) bases and one low pass basis. Wavelet packets are considered as a method for representing natural grains. They differ from wavelets in that at every decomposition step, the difference and average coefficients are further broken down. The results are particular linear combinations or superposition of wavelets. They form bases, which retain many of the orthogonality, smoothness, and localization properties of their wavelets [22].

Wavelet packets seem to have lot of potential: depending on the strategy used to find a suitable basis from the over-complete set of packets in a full wavelet transform, you can find the basis which represents the signal with the fewest non-zero coefficients. Wickerhauser, explain various basis-finding algorithms. While efficiency of representation is important, there are some key problems that cannot be easily overcome [22]. Comparison between different regions of the wavelet packets is extremely difficult because the wavelet packets are not shift-invariant [24]. Another difficulty with comparison has to do with packet levels. Every packet is denoted by the order in which

the average or difference coefficients have been further broken down, as shown in Figure 3.8.

There is no guarantee that all portions of the signal will be represented on the same packet level. If we have a packet that is denoted by *VWVWVV* in Figure 3.8, it is not trivial to change it with another packet as denoted by *VWV*. They are different sizes and have to interact with different packets in order to be properly put through the inverse wavelet decomposition. This makes switching positions of packets very difficult. In order for a wavelet packet basis to be complete and non - redundant, it must be composed from the lowermost nodes of an admissible binary tree of bases. An admissible binary tree is one whose nodes have either zero or two children.

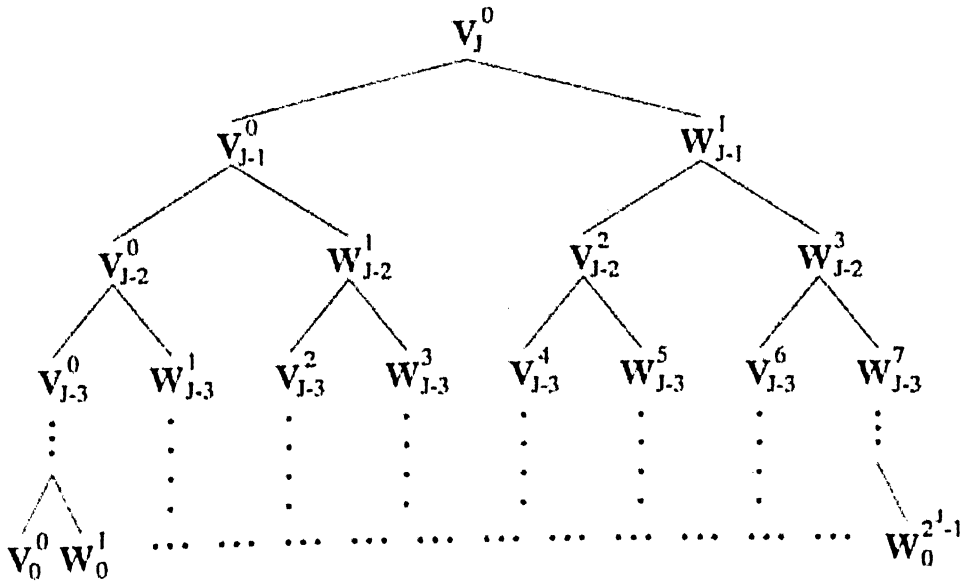


Figure 3.8. *Wavelet Packet Spaces.*

The number of operation required to decompose a signal into its full wavelet packet tree is $O(K N \log_2 N)$, where N is the length of the signal and K is the number of non-zero filter coefficients in c and d [7]. The storage of all coefficients from all the nodes in a full wavelet packet binary tree requires $N \log_2 N$ locations in memory. However, once a

particular wavelet packet has been chosen, the redundant coefficients can be dropped, reducing the storage requirements to N memory locations.

3.3 Wavelet Thresholding.

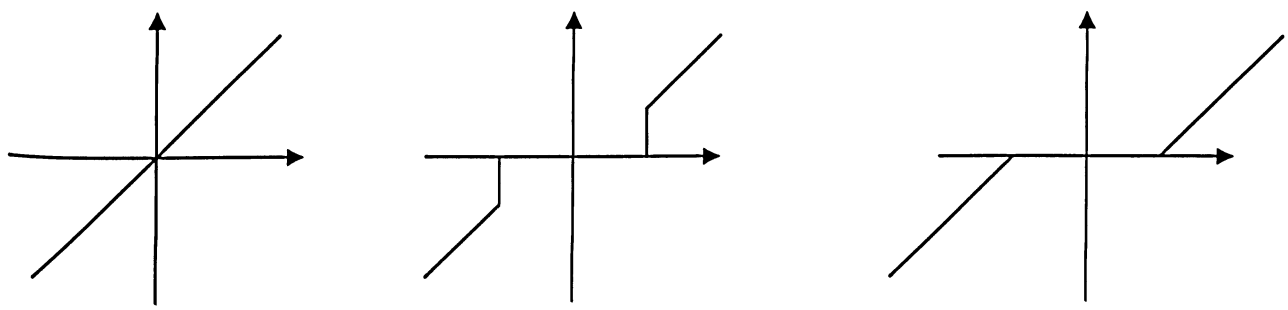
As wavelet has its basis emulating the front-end auditory periphery, efforts have been made to take advantage of this signal-processing tool for speech enhancement. The most used approach is based on the non-linear thresholding of the wavelet coefficients as it is explained by Donoho [11], which bridges the multi-resolution analysis and non-linear filtering. Donoho proposed this powerful wavelet-based approach as follows:

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:

$$y = x + n \tag{3.18}$$

In the wavelet domain, this will result on:

$$Wy = Wx + Wn. \tag{3.19}$$



(a) (b) (c)

Figure 3.9. *Without Thresholding (a), Hard Thresholding (b), Soft Thresholding (c).*

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

$$X = W^{-1}X_{estimation} = W^{-1}Y_{thresh} \tag{3.20}$$

where Y_{thresh} represents the wavelet coefficients after thresholding and W^{-1} denotes 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: soft and hard thresholding shown in Figure 3.9, and mathematically illustrated as:

$$Thr_{Hard}(X,T) = \begin{cases} X & |X| > T \\ 0 & |X| < T \end{cases} \tag{3.21}$$

$$Thr_{Soft}(X,T) = \begin{cases} Sgn(X)\,(|X| - T) & |X| > T \\ 0 & |X| < T, \end{cases} \tag{3.22}$$

where X represents the wavelet coefficients before thresholding and T is the threshold.

Major problems arise when the basic wavelet thresholding method is applied, to a speech degraded by noises. Most of the practical situations the environment noise is colored noise, non-stationary noise, and using the universal thresholds will result is speech distortion. Various modifications have been made. For example, Sheikhzadeh ,

DSP factory Waterloo, ON, [37], 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:

$$Thr(X,T) = \begin{cases} X & |X| > T \\ T * \left[\dfrac{\left[(1+\mu)^{|X/T|} - 1\right]}{\mu} \operatorname{sgn}(X) \right] & |X| < T, \end{cases} \tag{3.23}$$

where $0 < \mu < 1$ is the amplification constant for $X < T$.

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

$$T = \sigma \sqrt{2\log(N)} \ , \tag{3.24}$$

where, N, is the length of the noisy signal, and $\sigma = MAD/0.6745$, with MAD denoting the absolute median estimated on the first scale of the wavelet coefficients. Johnstone and Silverman, proposed the level depended threshold method to deal with correlated noise, where for each frequency interval the threshold is proportional to the standard deviation in that interval:

$$\lambda_a = \sigma_a \sqrt{2\log(N_a)}, \tag{3.25}$$

where $\sigma_a = MAD_a/0.6745$, N_a is the number of samples in scale a, and $MADa$ is the absolute median estimate at scale a [38].

CHAPTER 4

DE - NOISING ALGORITHM

4.1 Introduction.

Single channel speech enhancement is still an important issue when the single channel speech is the only available source as in Hearing Aids. As explained earlier, wavelet analysis can be easily used for noise reduction algorithms in image processing, audio and speech coding. The corrupting white noise can be removed, if a wavelet coefficient threshold is adequately selected, and by subtracting this threshold from the noisy wavelet coefficients. This method, suffers from residual noise and speech distortion in the noise reduction applications.

The performance of speech enhancement algorithm will be improved, if the threshold is adaptively updated according to corrupting noise level. Sheikhzadeh and Abutalebi [37], suggested the adaptive threshold based on a voiced frame and unvoiced frame. So, the threshold is increased for high-bands in the voiced frames, and decreased for unvoiced frames in the high-bands. A level dependent threshold was suggested by Rouat [39], where a Tiger energy operator is utilized to improve the discriminability for determining the speech-noise – dominated segments. Besides reducing the noise and good quality of speech enhancement, another advantage of this method is that it does not need a segmental SNR estimation.

Method proposed by Virag [40], where the threshold was derived with the assumption that the noise spectrum was white and stationary, is a trade off between the amount of noise reduction, the speech distortion, and the level of musical residual noise. In this research, a novel approach for noise reduction in Hearing Aids is proposed. In this

algorithm, the noisy speech is first preprocessed using a Spectral Subtraction with an MMSE (minimum mean square error) as proposed by Ephraim and Malah [3], to initially lower noise level with negligible speech distortion. A perceptual wavelet transform is then used to decompose the resulting speech into critical bands.

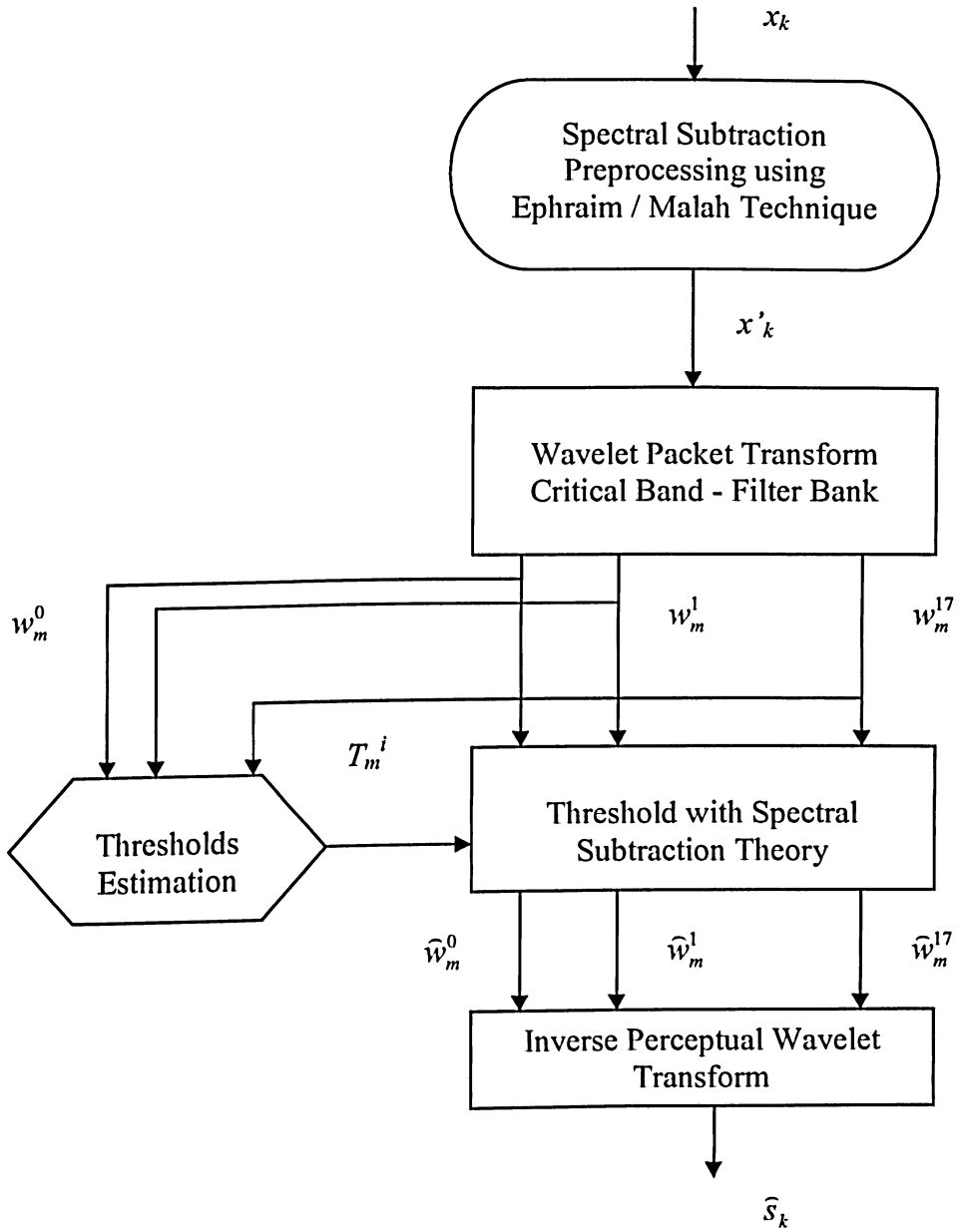


Figure 4.1. *The proposed algorithm structure.*

Threshold estimation is implemented that is both time and frequency dependent, providing robustness to non-stationary and correlated noise environments. Figure 4.1 shows the algorithm structure for this research:

4.2 Description and Analysis of the algorithm.

4.2.1 Spectral Subtraction preprocessing using Ephraim / Malah Technique.

The purpose of this preprocessing block, is to initially lower the noise level of the noisy speech x_k, while keeping the distortion level at minimum. In 1984, Ephraim and Malah [41] proposed an optimum MMSE STSA (short term spectral amplitude) estimator. This method calculated a gain function based on the a priori and a posteriori SNR-s. The following equations describe the method:

$$\hat{S}(k) = H(k)Y(k), \tag{4.1}$$

$$H(k) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{\lambda_n}}\cdot\frac{\gamma_A}{1+\gamma_A}F[\gamma_N\frac{\gamma_A}{1+\gamma_A}\ ], \tag{4.2}$$

where γ_A is the a priori SNR, which is calculated as:

$$\gamma_A = \beta\cdot\left(\frac{|\hat{S}_{i-1}(k)|^2}{|\hat{D}_i(k)|^2}\right)+(1-\beta)\cdot P(\gamma_{N,i}-1), \tag{4.3}$$

β – is a weight coefficient ($\beta = 0.98$), i is the frame index with $P(x) = x$ if $x \geq 0$, and $P(x) = 0$ otherwise. γ_N is the a posteriori SNR and F is a function representing:

$$F(x) = e^{\frac{-x}{2}}\left[(1+x)I_0\left(\frac{x}{2}\right)+I_1\left(\frac{x}{2}\right)\right], \tag{4.4}$$

where $I_0(y)$ and $I_1(y)$ are zero and first order modified Bessel function respectively. Unlike the magnitude averaging where averaging is performed irrespective of whether frame contains speech or noise, this method performs non-linear smoothing only when the SNR is low, i.e. when the frame predominantly contains noise.

The residual noise present due to this technique has been observed to be colorless. This method reduces the distortions in the speech parts due to averaging. Ephraim – Malah algorithm is well suited for noise reduction in hearing aids.

4.2.2 Critical – Band wavelet packet decomposition.

Auditory masking is a phenomenon related to the hearing perception of neighboring signal components [20]. If a strong signal A (the masker) masks a weak signal B (the masker) the weak signal B, is not even heard, even though it is present. Two main categories of masking, depending on the time and frequency location of A and B, may be considered. When both signal occur at the same time, masking is considered simultaneous, and is modeled in the frequency domain. If B either precedes or succeeds A, masking is termed *temporal* or *non simultaneous.*

Modeling is strongly based in this kind of masking. Several studies have highlighted the non-uniform temporal and spectral resolution of the human ear. Frequency components of sounds are integrated into critical bands (as mentioned earlier) whose centers and bandwidths have been measured. The center frequency location of these sub- bands is known as the *critical band rate z* and is expressed as:

$$z = 13\arctan(7.6*10^{-4} f) + 3.5 \arctan(1.33 * 10^{-4} f) \text{ [Bark]} \qquad (4.4)$$

The critical bandwidth of these filters is of approximately 100 Hz below 500 Hz. Beyond 500 Hz this bandwidth corresponds to 20 % of the center frequency value. The distance from one critical band center to the center of the next one is known as 1 Bark. The human auditory system covers approximately 25 Barks. The absolute threshold of hearing (AHT), or threshold in quiet, is the average sound pressure level (SPL) below

which the human ear does not detect any stimulus. This threshold is frequency dependent and their relation is expressed as:

$$AHT_{SPL}(f) = 3.64 f^{0.8} - 6.5\, e^{-0.6(f-3.3)^2} + 0.0001 f^4 \; [dBSPL] \qquad (4.5)$$

In this algorithm, we are trying to approximate the critical band analysis using the overlapped block orthogonal transform, previously introduced, and to estimate the auditory masking threshold. In the 0 – 4 kHz bandwidth there are only 18 critical bands (Table 2.1). Using a five stage DWPT ($p = 5$, $N = 32$), a frequency resolution of 125 Hz can be achieved. The choice of the prototype filter of the transform, as well as its length, influences the separation of the sub band signals and determines the maximal window length. The temporal analysis of the human ear requires that the analysis windows be limited to 5 – 10 mS, toward higher frequencies, and they could spread up to 100 mS in the lower frequencies.

In this algorithm I have considered the bandwidth of input signal as 4 kHz. Filter banks that perceptually divide whole bands in filter banks are illustrated in Figure 4.2. Looking at these filter banks, the locations of high- and low- pass filters are symmetry in each node. The branches that contain decomposed levels less than five, in the synthesis stage, must be delayed, to synchronize with the branches that contain five levels.
The filters proposed by Daubechies are the one that best preserve frequency selectivity as the number of stages m, of the DWPT. The energy in each critical band is summed as :

$$Bi = \sum_{\omega=bli}^{bhi} P(\omega), \qquad (4.6)$$

where bl, is the lower boundary of the critical band i, bh is the upper boundary of critical band i, and $P(\omega)$ is the power spectrum.

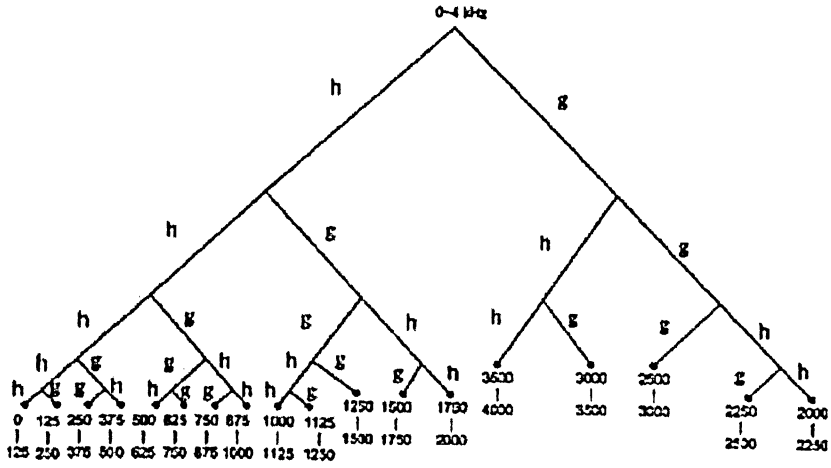


Figure 4.2. *Decomposition tree of the filter bank (Courtesy of Ta Lu, Ch. , Wang, H. "Enhancement of single channel speech based on masking property and wavelet transform", Speech Communications, 2003)*

Bandwidth and subbands index are described in Table 4.1.

Level	Subband Index	Bandwidth (Hz)
m = -5	i = 1,2,3.....10	125
m = -6	i = 11,12.....15	250
m = -7	i = 16,.....18	500

Table 4.1 *Bandwidth and subband index*

The relation among critical – band index i, the level index m, and the wavelet subband k, could be described as follows[13]:

$$i = \begin{cases} k & 1 \le k \le 10 \quad for\ m = -5 \\ k + 10 & 1 \le k \le 5 \quad for\ m = -6 \\ k + 15 & 1 \le k \le 3 \quad for\ m = -7 \end{cases} \tag{4.7}$$

Energy distribution along basilar membrane of the inner ear $E_m(i)$, could be obtained by convolving subband energy $B_m(i)$, with spreading function $S_m(i)$ as proposed by Schroder [19]:

$$E_m(i) = B_m(i) \oplus S_m(i) \tag{4.8}$$

The spreading function $S_m(i)$, is expressed in (4.9), and shown in Figure 4.3:

$$10 \log_{10} S_m(i) = 15.81 + 7.5 \bullet (i + 0.474) - 17.5 \bullet (1 + (i + 0.474)^2)^{0.5} \tag{4.9}$$

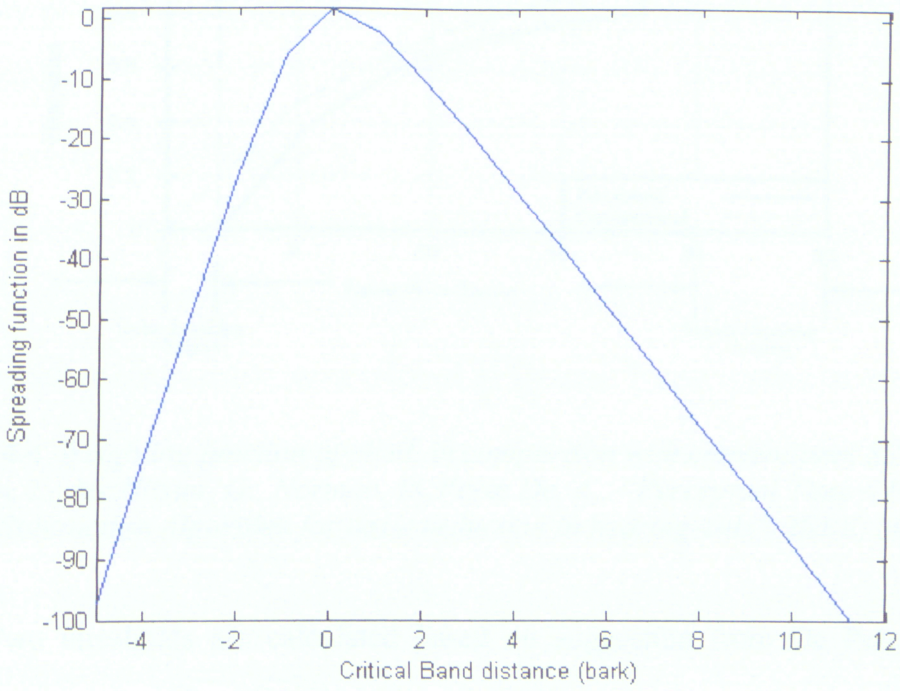


Figure 4.3. *Spreading Function.*

4.2.3 Threshold Estimation.

As explained earlier the hard thresholding has two values for the weighting function: zero if SNR is lower than threshold T, and unchanged if it is bigger. When noise and speech are presented together within a critical band and the excitation produced by

the weaker signal is not large, then the strong signal will dominate the auditory filter. Figure 4.4, shows the weighting functions applied in this research. If the signal is much stronger than noise, we are in signal - masking region, and the noise is rendered inaudible. Otherwise if the noise is much stronger than signal, than we are in noise-masking region. To distinguish these regions, a frame SNR is calculated, and this is applied in threshold criteria.

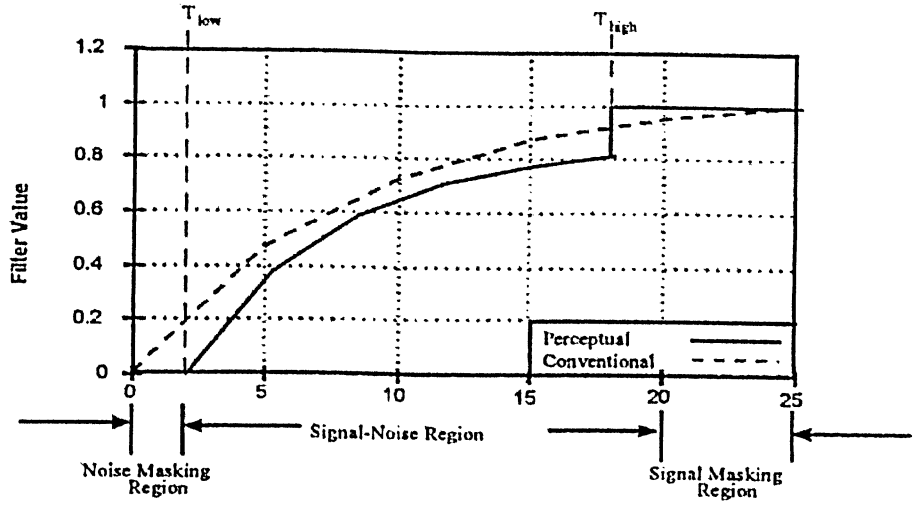


Figure 4.4. *Weighting function applied, in comparison with conventional SS. (Courtesy of Min, L., McAllister, G., Norman, D.,Perez De, A., "Perceptual Time – Frequency Subtraction Algorithm for noise reduction in hearing aids", IEEE, 2001)*

Two thresholds are calculated based on suggestion from De Perez [20], low threshold T_{low} and high threshold T_{high}. In order to determine T_{low} and T_{high}, an estimation of the subjective loudness, of both speech and noise is required. Since we have not at a present time any literature to perform this task, some subjective and objective results were used to determine the lower and higher threshold. Lower threshold was determined as:

$$T_{low} = \sigma\sqrt{2\log N}, \qquad (4.10)$$

where N is the frame length, and the $\sigma = MAD/0.6745$ (explained earlier), and

$$T_{high} = 2.5\ T_{low}. \tag{4.11}$$

These values were used in this research, and using these thresholds did not introduce any noticeable distortion in the processed speech. These determinations were primarily based on the auditory perceptual features. This method could reduce the noise in different SNR noisy speech, and could reduce the amount of the musical noise produced by the conventional spectral subtraction. The wavelet packet decomposition used makes it easy to estimate the signal from noise. When the residual noise is present, even after processing, the speech in low SNR does not contribute significantly toward intelligibility.

4.2.4 *Applying the thresholds.*

This algorithm is a subtraction of noise from noisy speech in the wavelet packet domain and incorporates the perceptual auditory features. It has a similar form with conventional spectral subtraction explained in Chapter 2, and applied in this research as well. Knowing that every wavelet coefficients, contributes noise of variance σ, but only a few of them contribute to the signal, help us to determine speech/noise segments using the wavelet transform. The filtering used in spectral subtraction method, is adopted here as well to calculate speech spectrum:

$$\hat{s}_{i,k}(m) = h(i,k)s_{i.k}(m), \tag{4.12}$$

where $h\ (i,\ k)$ is a zero-phase magnitude response filter, computed according to the SNR in each critical band. It is proved, that the auditory system perceives sounds based on SNR or SMR (signal-to-masker ratio) in each critical band.

As explained [2.6], in conventional spectral subtraction, the estimated filter is based on the SNR in each frequency band and is given by:

$$H(\omega) = 1 - \frac{E[|N(\omega)|]}{|X(\omega)|} \tag{4.13}$$

In wavelet domain, this filter is estimated by considering the linear model [20]. The proposed filtering is as follow:

$$h(i,k) = 1 - \frac{\beta(E[w^2_{i,k}(n)])^{1/2}}{|w_{i,k}(m)|}, \tag{4.14}$$

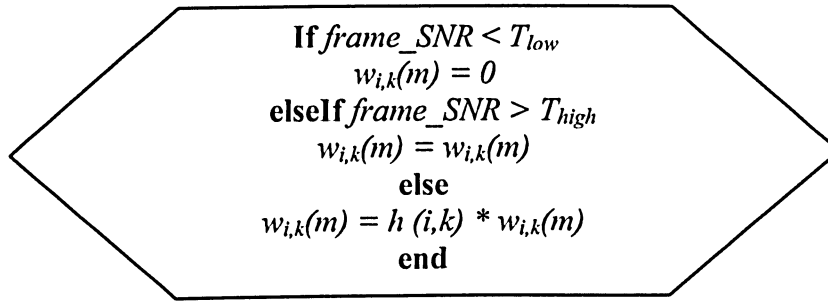
where:

β – constant and calculated based on [20]: $\beta = \sqrt{2\log 10 N \log 2N}$ (N – frame length)

$(E[w^2_{i,k}(n)])^{1/2}$ - standard deviation.

$w_{i,k}(m)$ – wavelet packet coefficients.

This filter is convenient for including the non-linear properties as well, and it can incorporate the masking behavior as shown in Figure 4.4. So, in this algorithm this is the pseudo - code for threshold evaluation:



This algorithm is an optimization of the wavelet subtraction using both hard and soft thresholding.

4.3 Implementation.

To evaluate these algorithms a speech database provided from Speech Enhancement and Assessment Resource (SpEAR), Oregon Health and Science University, was used.

Please refer to Table 4.2, for the speech evaluated, and their SNR's. An IBM PC, using Windows XP, and MATLAB 6.5 R13 was used. Off line simulations were conducted, in order to check the validity and feasibility of the algorithms.

#	File	SNR (dB)
1	BigTips_At_Min_5_93_SNR.wav	F16 Noise at: SNR = –5.93 dB
2	BigTips_At_Min_0_67_SNR.wav	F16 Noise at: SNR = –0.67 dB
3	BigTips_At_3_26_SNR.wav	F16 Noise at: SNR = 3.26 dB
4	BigTips_At_10_11_SNR.wav	F16 Noise at: SNR = 10.11 dB
	BigTips_Clean.wav	"Good Service should be rewarded by big tips"
5	Female_Vega_m_4_93_SNR.wav	Pink Noise at SNR = - 4.93 dB
6	Female_Vega_2_97_SNR.wav	Pink Noise at SNR = 2.97 dB
7	Female_Vega_ 8_96_SNR.wav	Pink Noise at SNR = 8.96 dB
8	Female_Vega_ 13_65_SNR.wav	Pink Noise at SNR = 13.65 dB
	Female_Vega_Clean.wav	"I am sitting in the morning at the dinner in the corner"

Table 4.2. *Sample speeches.*

Implementation on a workstation permits modifications and changes without constraints of time, memory or computational power.

4.3.1 Conventional Spectral Subtraction.

The speech signal is first Hamming windowed using a 20-mS window and a 50 percent overlap (10 mS). The amount of overlap between consecutive frames is also associated with the frame size, and is required to prevent discontinuities at frame boundaries. The windowed speech frame is then analyzed using the Fast Fourier Transform.

Since noise is estimated during non-speech periods, in this algorithm, a robust speech detector was implemented [1]. This is how the VAD is implemented:

1. **Buffer** data into m^{th} frame, $x(n.m)$ and transform to the frequency domain:

$$X(\omega,m) = FFT(x(n,m)) \tag{4.15}$$

2. **Initialize** the noise spectrum and noise mean for $m = 1$.

$$N(\omega) = X(\omega,m) \tag{4.16}$$

$$\mu_N = \frac{1}{L}\sum_{\omega=0}^{L-1} N(\omega) \tag{4.17}$$

3. **If VAD = 0,** then update the noise spectrum, mean and standard deviation for frame. Frames two through ten are assumed to be noise in order to get a good initial average of the stationary noise in the environment.

$$N(\omega) = a * N(\omega) + (1-a) * X(\omega,m) \tag{4.18}$$

$$\mu_N(m) = \frac{1}{L}\sum_{\omega=0}^{L-1} N(\omega) \tag{4.19}$$

$$\mu_N = b * \mu_N + (1-b) * \mu_N(m) \tag{4.20}$$

$$\sigma_N = (b * \sigma^2 N + (1-b)\mu N(m)^2)^{\frac{1}{2}} \tag{4.21}$$

where μ_N, σ_N *and* σ_N^2 are the mean, standard deviation and variance of the noise estimate.

4. **Update thresholds** if a frame does not contain speech, using the mean and variance of the noise estimate where threshold settings are adjusted using the multipliers a_S (speech) and a_N (noise), which can be adapted experimentally.

$$Thresh_S = \mu_N + a_S * \sigma_N \tag{4.22}$$

$$Thresh_N = \mu_N + a_N * \sigma_N \tag{4.23}$$

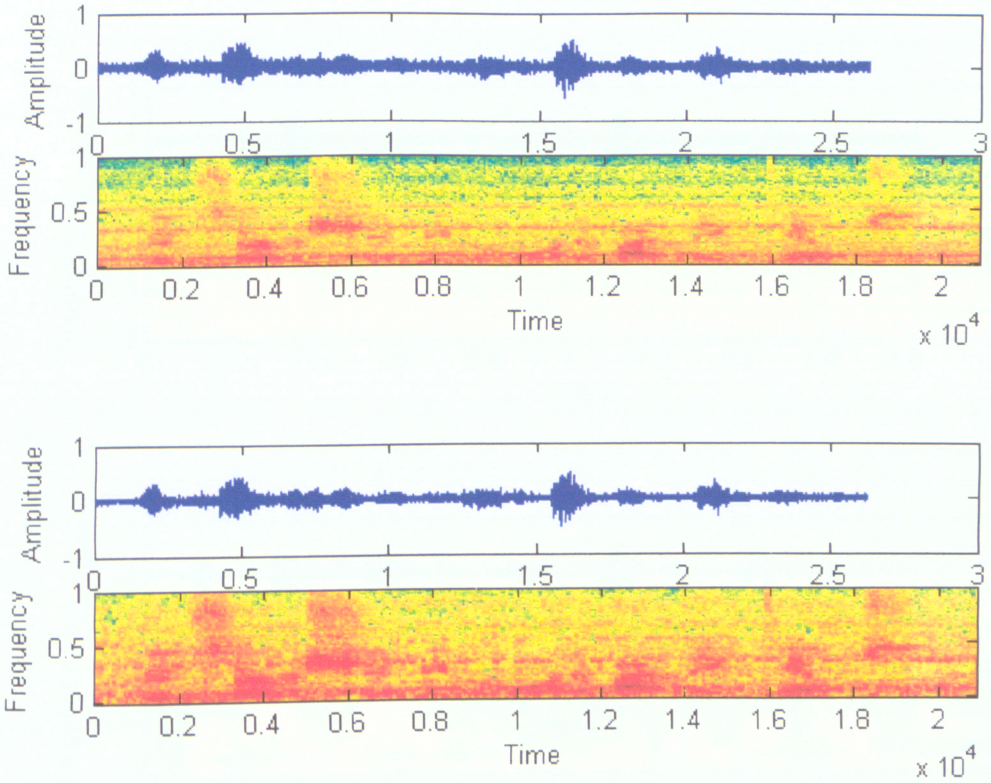


Figure 4.5. *a), b) Frequency Response and Spectrogram of the Noisy Speech (2); c), d) Frequency Response and Spectrogram of the enhanced speech using Spectral Subtraction.*

The Figure 4.5 shows frequency response of the File 2 (refer to Table 4.2) before and after processing using Spectral Subtraction. The SNR is improved, but as it is mentioned, in this algorithm there is always a trade-off between the noise reduction and speech distortion.

4.3.2 Spectral Subtraction in wavelet domain.

The first part of this algorithm is the same as the conventional Spectral Subtraction. The only difference is that here I am using Ephraim / Malah Technique (4.2.1), to reduce the noise. Then the de - noised speech is processed using the wavelet packets transform.

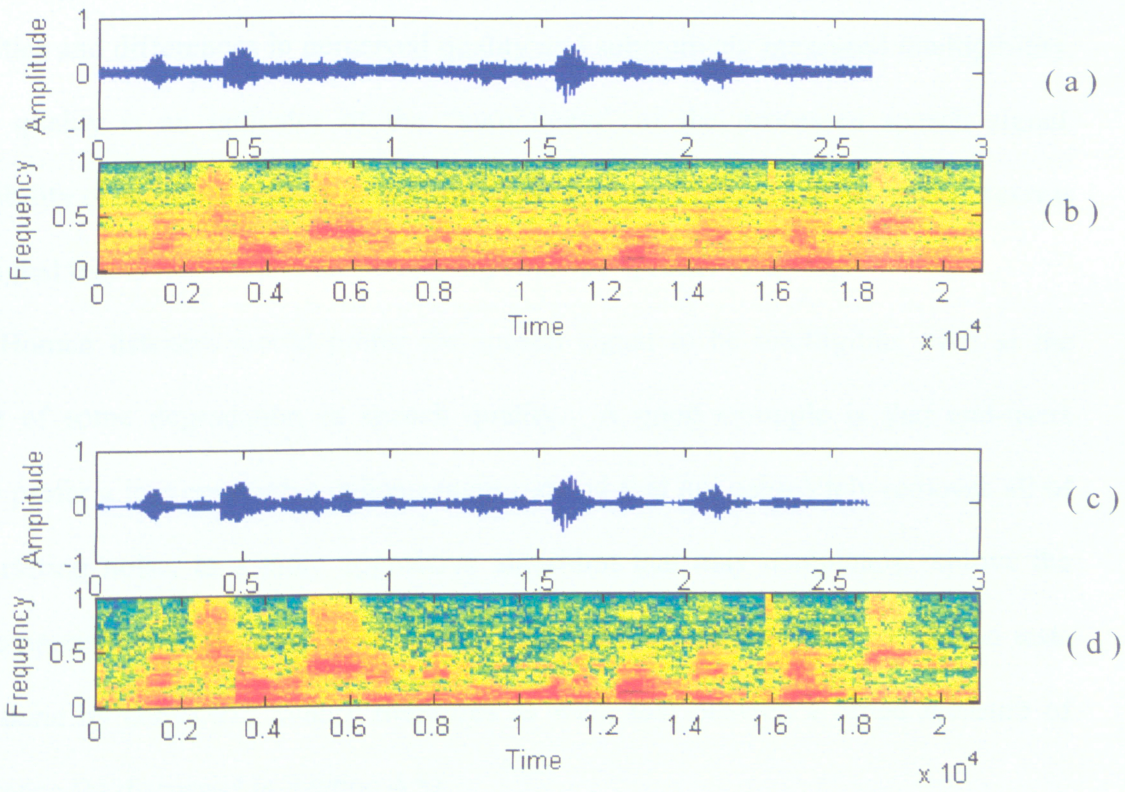


Figure 4.6. *a), b) Frequency Response and Spectrogram of the Noisy Speech (2); c), d) Frequency Response and Spectrogram of the enhanced speech using Spectral Subtraction in wavelet domain.*

Eighteen wavelet sub - bands are employed to implement the critical-band wavelet packet decomposition. Figure 4.6 shows the frequency response and the spectrogram of the same noisy speech (2) that was implemented in the conventional SS. The SNR has improved by 2.58 dB compare to the conventional SS, and the speech distortion is lower as evaluated perceptually.

The noise level was estimated from first 100 mS, and frame length used is 16 mS. This algorithm is evaluated using objective measures (SNR), and subjective tests based on mean – opinion score studies (MOS).

4.4 Objective measures for performance evaluation.

In the evaluation of speech enhancement algorithms, it is necessary to identify the similarities and differences in perceived quality and subjectively measured intelligibility. Speech quality is an indicator of the "naturalness" of the processed speech signal. Intelligibility of speech signals is a measure of the amount of speech information presents in the signal that is responsible for conveying what the speaker is saying.

Human listeners would prefer the speech signal to be intelligible, even at the expense of some degradation in speech quality. A good example is that end-users actually prefer a less aggressive enhancement method that not completely removes all of the interfering noise, to a more aggressive algorithm that may completely remove the noise component but also reduce the speech intelligibility. Performance evaluation tests can be done by subjective quality measures as well, that provide a broad measure of performance (as discussed in Section 4.5).

Objective measures, provide a measure that can be easily implemented and reliably produced. It is based on mathematical comparison of the original and processed

speech signals. The majority of objective quality measures quantify speech quality in terms of a numerical distance measure or a model of the perception of speech quality by the human auditory system. It is desired that the objective measure with the judgment of the human perception of the speech. However, it has been seen that the correlation between the results obtained by objective measures are not highly correlated with those obtained by subjective measures. The signal-to-noise ratio (SNR) and the Itakura-Saito (IS) measure are some of the most widely used objective measures. The SNR is a popular method to measures speech quality.

It is calculated as the ratio of the signal to noise power in decibels:

$$SNR_{dB} = 10 \cdot \log_{10} \left(\frac{\sum_{m} y^2(m)}{\sum_{m} [s(m) - \hat{y}(m)]^2} \right), \tag{4.24}$$

where $y(m)$ is the clean speech, $s(m)$ is the noisy speech and $\hat{y}(m)$ is the enhanced speech.

Nr.	Speech Samples	SNR after applying SS (dB)	SNR after applying SS in WD (dB)
1	Male speech at: SNR = – 5.93 dB	*3.52*	*4.25*
2	Male speech at: SNR = – 0.67 dB	*4.36*	*6.94*
3	Male speech at: SNR = 3.26 dB	*9.31*	*11.32*
4	Male speech at: SNR = 10.16 dB	*15.34*	*16.61*
5	Female singing at: SNR= – 4.93 dB	*2.41*	*3.26*
6	Female singing at: SNR = 2.97 dB	*8.76*	*10.23*
7	Female singing at: SNR = 8.96dB	*15.69*	*17.52*
8	Female singing at: SNR = 13.65 dB	*19.68*	*21.35*

Table 4.3. *The SNR improvement after applying SS in time domain and SS in wavelet domain.*

If the summation is performed over the whole signal length, the operation is called *global* SNR, and if the summation is performed on a frame basis, this operation is referred to *segmental* SNR. An average of the segmental *SNR's*, over the whole speech length can be performed, and this method has better correlation to subjective results than the global SNR. Table 4.3 shows the SNR improvement for the speech samples at Table 4.1, in different noise levels.

The results from Table 4.3 where obtained using global SNR's. Even the SNR for both algorithms are almost the same, the de - noising algorithm reduces the noise as much as conventional SS, but the speech with SS in wavelet domain is less distorted, based on MOS results.

4.5 Subjective evaluation of the algorithm.

The best speech quality measurement requires a subjective judgment by a listener as to how "good" speech material sounds. Subjective speech quality measures have varied forms. One method is to ask a group of listeners to rank the quality of speech along a predetermined scale after the comparisons of original and processed speech data [2]. The mean opinion score (MOS) is most widely used subjective quality measure [2]. In this method, listeners rate the speech on a five – point scale, where a listener's subjective impressions are assigned a numerical value as shown in Table 4.4. Although the number of subjects may not be sufficient for a formal assessment of the algorithm, it gives a good idea for its performance. Another method used for intelligibility is the diagnostic rhyme test (DRT) that requires listeners to circle the world spoken among a pair of rhyming words.

Scale	Meaning
5	*Imperceptible*
4	*Perceptible*
3	*Slightly annoying*
2	*Annoying*
1	*Very annoying*

Table 4.4. *Rating Scale used for MOS.*

The original noisy speeches were played first, and then the enhanced speeches from both algorithms were played as well. The results of MOS are plotted and shown in Figure 4.7:

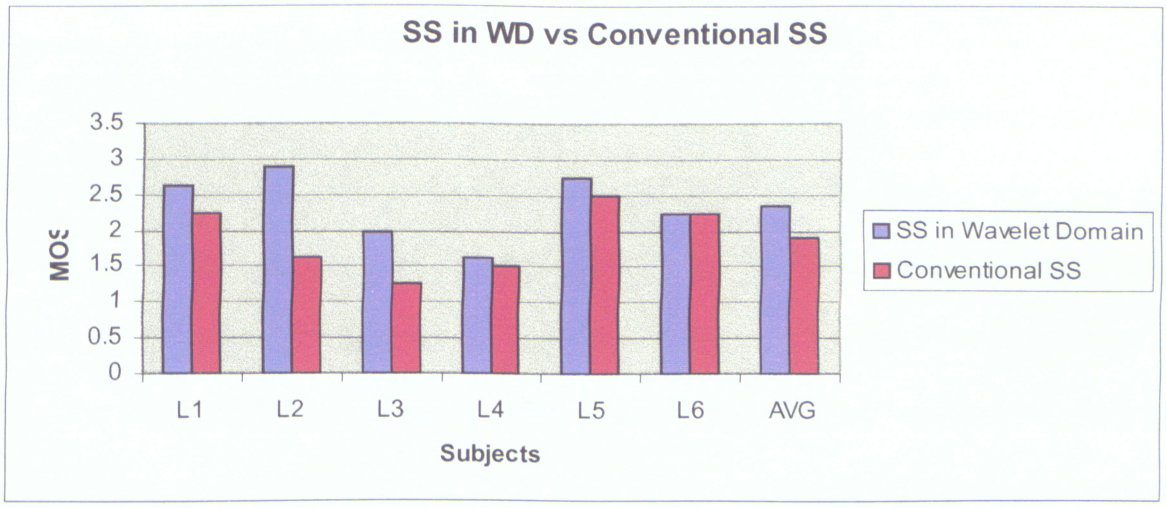


Figure 4.7. *MOS Test results for six different subjects.*

Average (AVG), denotes the average score for six listeners. The test shows better performance of the Spectral Subtraction in wavelet domain compared to conventional SS. Looking at MOS it is obvious that the speech enhancement in different noise level (male

speech) was rated higher than the female speech (female singing). MATLAB code is available in a CD or by contacting the author (fmyftari@ee.ryerson.ca). The simulation is available on line at: www.alkospace.com/fmyftari.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

The proposed speech enhancement algorithm, based on speech perception of normal and impaired auditory systems, developed principally for use in hearing aids, could be optimized for developing new speech enhancement systems. The project was concentrated on the design of the algorithm based on the features of the auditory perception. Wavelet packet transforms were used to model the human auditory system, that every stage of critical bands is consistent with psychoacoustics properties of human ear. This algorithm, was intended for noise-reduction in single channel systems, and is based on the concept of conventional spectral subtraction algorithm.

The spectral subtraction algorithm in wavelet domain was compared with the conventional spectral subtraction, and the results show that de-noising using wavelet packets are more efficient than conventional SS, and the speech enhancement is less distorted. The spectral subtraction in wavelet domain provides a definite improvement over the conventional spectral subtraction method, and does not suffer from musical noise. Most of the researchers for noise reduction in hearing aids, suggest that greater benefit for the hearing-impaired listeners is possible with nonlinear processing, and to achieve it in this algorithm the masking phenomenon are incorporated.

This algorithm produced more noise reduction than the conventional spectral subtraction, because of the wavelet packet auditory filtering, leading to better speech estimates with time and frequency, and the masking phenomenon being exploited

accurately by identifying the three perceptually different SNR regions. Future research could be conducted to adaptively update the noise segments, gain function, and threshold estimations. It can be expected that better psychophysical modeling of the cochlea, will provide a more precise description of the auditory system and the interactions between the signal level, masking and loudness.

The future digital hearing aids should be developed with a nonlinear processor that can adapt to conditions of noise and speech, to maintain optimal performance with regard to masking, loudness and speech perception over a wide range of conditions. Hearing aid development requires knowledge of acoustics, transducers, signal processing, auditory physiology and psychoacoustics, as well as low power semiconductor technology. Advances in digital technology and speech processing have been rapid and are continuing. The complexity of circuitry that can be fabricated in silicon is increasing, and the cost and power consumption is decreasing. In designing hearing aids it is easy to become focused on only one aspect of the problem, such as speech communication under ideal conditions.

Auditory experience encompasses a large range of conditions, from eavesdropping in on a quiet distant conversation, to talking in a noisy area, to listen the music. The hearing-impaired listener will greatly benefit if the hearing aids cover the conversational range of sounds and more completely if the normal auditory entire range of sounds is taken care of.

References

[1] Cho, Y.D., Al - Naimi, K. and Kondoz, A., "Improved Voice Activity Detection based on a Smoothed Statistical Likelihood Ratio," *IEEE ICASSP*, Salt Lake City, USA, May 2001.

[2] Deller, J., Hansen, J.H.L. and Proakis, J., "Discrete – Time Processing of Speech Signals", *Wiley – IEEE Press,* pp. 145 – 178, September 1999.

[3] Ephraim, Y. and Van Trees, H.L., "A signal subspace approach for speech enhancement", *IEEE Trans. Speech and Audio Processing,* vol. 3, pp. 251 – 266, July 1995.

[4] Vaseghi V. S., "Advanced Signal Processing and Noise Reduction", *John Wiley & Sons,* pp. 333- 352, 2000.

[5] Hear it AISBL www.hear-it.org

[6] Liao, L. and Gregory, M.A, "Algorithms for speech classifications", *International Symposium on Signal Processing and its Applications (ISSPA),* pp. 623 – 627, August 1999.

[7] Mallat, S. "A Wavelet Tour of Signal Processing", *Academic Press*, pp. 79 – 91, 2001.

[8] Davis, M. "Noise Reduction in Speech Applications", *CRC Press*, pp. 379 – 391, 2002.

[9] Levitt, H "Noise Reduction in Hearing Aids", *Journal of Rehabilitation and Research,* vol 38, pp. 1 – 12, February 2001.

[10] Ross, M. "Hearing Assistance technology", *The Hearing Journal*, vol. 57, November 2004.

[11] Donoho, D.L and Johnstone. I. M, "Ideal spatial adaption via wavelet shrinkage", *Biometrika*, 81, pp. 425 – 455, September 1994.

[12] Bahoura, M. and Roaut, J.., "Wavelet speech enhancement based on the Teager energy operator", *IEEE Signal Processing Letters*, vol. 8, No. 1, pp 10-12, January 2001.

[13] Ching, L., Chuan, W. "Enhancement of single channel speech based on masking property of wavelet transform", *Speech Communication*, vol 41, pp. 409 – 427, 2003

[14] Venema, T. "Compressions for Clinicians", *Singular Publishing Group*, pp. 91-113, 1999.

[15] Boll, S.F., "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. on Acoust., and Audio Processing,* vol. 27, No.2, pp. 113 – 120, April 1979.

[16] Gong, Y., "Speech recognition in noisy environments: A survey", *Speech Communications*, Vol. 16, No. 3, pp 261 – 291, April 1995.

[17] Erdol, N "VAD over Multiresolution Subspaces", *IEEE Sensor Array Multichannel workshop .*, pp 217 – 200, July 2000

[18] R.R. Coifman and N. Saito, "Construction of local orthonormal bases for classification and regression", *Comptes Rendus Acad. Sci. Paris, Serie I*, Vol. 319, 1994.

[19] Shroder, M.,R., Atal, B.S., Hall,J.L., "Optimizing digital speech coders by exploiting masking properties of the human ear", *J. Acoustic. Soc. Amer.* Vol 66, pp. 1647 – 1652, 1979.

[20] Min, L., McAllister, G., Norman, D.,Perez De, A., "Perceptual Time – Frequency Subtraction Algorithm for noise reduction in hearing aids", *IEEE Trans Biomed Eng*, September 2001.

[21] K. Hesarmaskan, "Optimally weighted Local Discriminant Bases – Theory and Applications in Statistical Signal and Image Processing", *MASc Thesis, Ryerson University*, Toronto, 2002

[22] M.V. Wickerhauser," Adaptive Wavelet Analysis from Theory to Software", *A.K Peters, Ltd. Wesllesley, MA*, 1994.

[23] Wavelet Software, Rice University, http://www.dsp.rice.edu/software/rwt.shtml

[24] Wavelet in Matlab, http://www.control.auc.dk/~alc/html/wavelets_in_matlab.html

[25] R. Polikar, "The engineer's ultimate guide to wavelet analysis", *Rowan University*, Wavelet Tutorial, 2001.

[26] WaveLab v.8.02, http://www.stat-stanfords.edu/~wavelab.

[27] Ulehlova, L., Voldrich, L., and Janisch, R., Correlative study of sensory cell density and cochlear length in humans", *Hearing Research,* Vol. 28, pp 149 – 151, 1987.

[28] Plomp, R. "Speech – reception threshold for sentences as a function of age and noise level", *Journal of Acoustical Society of America*, Vol. 66, pp 533 – 549, 1978.

[29] Clarkson,P., and Bahgat, S., "Envelope expansion methods for speech enhancement", *Journal of Acoustical Society of America*, Vol. 89, pp 1378 – 1382, 1991.

[30] Bitzer, J., Simmer, K.U., "Multi – microphone noise reduction techniques as front – end devices for speech recognition", *Speech Communication*, Vol. 34, pp 3 –12, 2001.

[31] Cox, H., Zeskind, R.M., and Owen, M.M., "Robust adaptive beam-forming", *IEEE Trans. On Speech and Audio Processing*, pp 1365 – 1376, 1987.

[32] Hykin, S., "Adaptive Filter Theory", *Prentice Hall*, pp. 35 – 46, 2002.

[33] Rickets, T. and Henry, P., "Evaluation of an adaptive, directional microphone hearing aid", *International Journal of Audiology*, Vol. 41, pp 100 – 112.

[34] Rombouts, G "Adaptive filtering algorithms for acoustic echo and noise reduction", *PhD thesis*, Katholieke Universiteit Leuven, 2003.

[35] Aichner, R., Herbodt, W., "Least – squares error beam forming using minimum statistics and multi - channel frequency domain adaptive filtering", *International Workshop on Acoustic Echo and Noise Control - IWAENC*, pp. 223 – 226, 2003.

[36] Starng, G., and Nguyen, T., "Wavelets and filter banks", *Wellesley-Cambridge Press*, Boston, Massachusets, pp. 35 – 62, 1996.

[37] Sheikzadeh, H., and Abutalebi, H., "An improved wavelet – based speech enhancement system", *IEEE Signal Processing Letters*, 2001.

[38] Johnstone, I. M., and Silverman, B. W., "Wavelet threshold estimators for data with correlated noise", *J. Roy. Statist. Soc. B*, Vol. 59, pp 319 – 351, 1997.

[39] Rouat, J., and Bahoura, M., " Wavelet speech enhancement based on the teager energy operator", *IEEE Signal Processing Letters*, 2001.

[40] Virag. N "Single Channel speech enhancement based on masking properties of the human auditory system", *IEEE Trans. On Speech and Audio Processing,* 1999.

[41] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean – square error short term spectral amplitude estimator", *IEEE Trans. On Acoust., and Audio Processing,* Dec. 1984.

[42] Black, M... and Zeytinoglu, M., "Computationally Efficient Wavelet Packet Coding of Wide-band Stereo Audio Signals", *ICASSP*, pp. 3075 – 3078, May 1995.

[43] Kamath, S. and Loizou, P., "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *ICASSP* 2002.

[44] Cohen. A, Laura, J. "Wavelet Analysis in recruitment of loudness Compensation", *IEEE Trans. On Acoust., and Audio Processing,,* 1993.

[45] Sandlin, R. "Hearing Aid Amplification", *Singular Publishing Group,* pp. 37 – 107, 2000.

[46] Naoki Saito, "Local Feature extraction and its applications using a library of bases", *PhD thesis*, Yale University, 1994.

[47] Flogeras, D., Kaye.M.E "A real time spectral subtraction based on speech enhancement scheme", *CCECE*, vol1. pp. 1071-1075, 2003.