uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Rana M. Rahal

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical and Computer Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Automatic Volume Setting for Environment Sensitive Hearing Aids

TITRE DE LA THÈSE / TITLE OF THESIS

W. Gueaieb

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

C. Giguère

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

R. Goubran

H. Dajani

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Automatic Volume Settings for Environment Sensitive Hearing Aids

by

Rana M. Rahal

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.A.Sc. degree in
Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-65485-9
Our file Notre référence
ISBN: 978-0-494-65485-9

Abstract

The development of intelligent devices is becoming a popular trend in the hearing aid industry. Such devices aim at making the user's listening experience more natural and at improving customer satisfaction. One focus of interest in this dissertation is the automatic adjustment of the hearing aid control settings to minimize the need for manual user interventions. The proposed system is based on computational intelligence tools, namely artificial neural networks and neurofuzzy systems, which have the ability to learn the dynamics of highly nonlinear systems without the need for the explicit knowledge of their mathematical models. Such techniques are adopted here to map the acoustic features (input space) to the desired volume setting (output space) of the hearing aid user. Two computational intelligence tools, a multilayer perceptron and an adaptive network-based fuzzy inference system were analyzed on three simulated users with moderate, severe, and profound hearing losses. A hearing aid simulation system provided target volume settings to train and test the learning networks, selected to optimize the speech intelligibility index in each acoustic situation. The performances of both soft computing models obtained from over 2000 recordings demonstrated a high efficiency of the adopted approach in automatically optimizing volume settings for the three simulated users. In worst case scenario 95% of the testing patterns obtained 0.06 SII error or less, over 400 audio files. A future step is to extend to an online adaptation and eventually the proposed system would be integrated into a trainable self-learning hearing aid.

Acknowledgements

I would like to thank my thesis supervisors Dr. Wail Gueaieb and Dr. Christian Giguère for their ongoing support in guiding me to the proper direction of this thesis. As well their guidance and wisdom provided me the tools which I will carry with me to the professional world. I would also like to thank Dr. Tyseer Aboulnasr and Dr. Hisham Othman for their mentorship and helping me through obstacles which I faced during my research.

A special thank you goes out to my family, my parents Samih and Therese Rahal, my twin sister Lara, sisters Sherine and Nissreen and my brother Ramsey, whose continuous love and prayers gave me the strength to accomplish my goals.

Contents

List of Tables

List of Figures

Acronyms

BTE: Behind The Ear.
ITE: In The Ear.
CIC: Completely In the Canal.
SFS: Sequential Forward Search.
SBS: Sequential Backward Search.
UCL: Uncomfortable Listening Levels.
SII: Speech Intelligibility Index.
SPL: Sound Pressure Level.
HL: Hearing Loss.
AGC: Automatic Gain Control.
MF: Membership Function.
FIS: Fuzzy Inference System.
SFS: Sequential Forward Search.
PR%: Percentile Percentage.
STE: Short Time Energy.
LSTER: Low Short Time Energy Ratio.
avSTE: The average STE.
sgn(): The sign function.
ZCR: Zero Crossing Rate.
avZCR: The average ZCR.
HZCRR: High Zero Crossing Rate Ratio.
VDR: Volume Dynamic Ratio.
RMS: Root Mean Square.
RMSE: Root Mean Square Error.
MSE: Mean Square Error.
SNR: Signal-to-Noise Ratio.
MLP: Multi-Layer Perceptron.
ANFIS: Adaptive Network-based Fuzzy Inference System.
HN: Hidden Nodes.
IN: Input Nodes.
ON: Output Nodes.
L-M: Levenberg-Marquardt backpropagation learning algorithm.
SIIE: Speech Intelligibility Index Error.

VE: Absolute Value of the Volume Error in dB.

Symbols

p: Rule index.

t: Frame index.

N: The total number of frames.

M: The number of samples in a frame.

m: The number of patterns.

K: Total number of frequency bins in a frame.

k: Frequency bin index.

i: Pattern index.

η**:** Learning Rate.

$\forall t$**:** Across all frames.

Chapter 1

Introduction

1.1 Hearing Loss and Hearing Device Development

Within a generation, it is expected that the hearing loss population will grow by one-third and reach up to 40 million people in the United States. In 2005, more than one million adults report having a hearing-related disability and compared to the two past generations, people are losing their hearing 20 years earlier [33]. Kochkin in [33] conducted a survey on 3000 random hearing aid owners and 3000 random individuals with hearing loss who have not purchased a hearing aid. This survey determined that the rate of purchase of hearing aid devices in the United States decreased from 23.8% to 20.4% in 1984 to 1997. In 2004, the purchase rate only increased back to 23.5%. Additionally, from 1989 to 2004, the number of active hearing aid owners increased to 6.2 million, while the population of hearing loss individuals who do not own a hearing aid grew from 1.8 million to 24.1 million.

Many hearing loss individuals from the survey report not wearing their hearing aid due to the degradation of the hearing instrument [33]. Furthermore, a large segment of the population with hearing losses do not wear hearing instruments to improve their hearing performance due to several stigmas. Such stigmas include the reputation that hearing instruments are intended for senior citizens, and that hearing instruments do not effectively compensate for hearing loss [49].

Thus, there is significant pressure on hearing instrument manufacturers to combat the stigma placed on hearing instruments, and many ameliorative strategies have been proposed to enhance the performance of these devices and draw less attention to the fact that the user is wearing an assistive device. Such strategies include the use of automatic volume control settings for hearing aids, potentially providing users more freedom with a touch-less hearing aid capable of maintaining good audibility and user satisfaction during daily use despite changes in the environment or listening situations.

1.2 Thesis Motivation

A study performed by Kochkin in [31] investigates the root causes of customer dissatisfaction with hearing aids, and surveys hearing loss individuals who decided not to wear their hearing aid. Reasons for dissatisfaction included that the benefit of the hearing aid was minimal, e.g. a hearing aid would perform amplification but would not help users distinguish words. Volume control adjustments was another concern. Customers were frustrated with the constant manipulation of volume control and preferred a smart or automatic hearing aid. Research conducted in [48] also suggests another issue which results in users' dissatisfaction with hearing devices, is not achieving optimal calibration of signal processing parameters.

Kochkin [31] states that in order to improve customer satisfaction, the hearing aid industry must uniquely understand the essential function of a hearing aid to enhance the user's speech intelligibility in their daily environment. In fact, Kochkin mentions that a study demonstrated that customer satisfaction can be increased by 30% using advanced programmable technology. The move towards automatic or smart hearing aid devices is becoming a popular trend, a shift which aim to make the user's experience more natural, and improve customer satisfaction.

Büchler in [8] conducted a study on the feasibility and acceptance of performing automatic switching for hearing aid settings. Büchler determined that 75% of the test subjects found automatic switching useful, even if the performance was not perfect. However, further research performed in [32], [49] and [8] demonstrated that it is best to not completely remove volume control for some patients and use the volume control as a manual override to the automatic volume hearing aid. As such, Dreschler in [14] and Zakis in [57] showed that clients can make reliable adjustments of control setting to express their personal preferences. Personal preferences can vary widely across individuals with similar hearing loss [30]. Wagener in [54] shows that different users go about in a very wide range of acoustical environments in their daily lives and that the frequency of occurance of specific environments vary widely across users in practice. This provides good support for a new class of trainable hearing aids [30], [12] and [57], whereby manual adjustments by the user are tracked in different environments with the purpose of minimizing the frequency of interventions overtime. As well, Chalupper in [9] demonstrates that trainable hearing aids can be used to allow users to train different parameters of the hearing aid (e.g. compression, frequency shaping, gain) in order to customize the hearing aid parameters to users' preferences.

Mueller in [42] shows that when trainable hearing aids are used, the gain preferences expressed by the users are influenced by the initial gains provided to them, thereby demonstrating the importance of supplying good initial gains. One current topic of interest of this dissertation aims at supplying good initial volume settings in a wide range of listening situations for the users, prior to further user-specific training that would be taking place online.

The primary motivation for this thesis is to provide initial volume settings aimed at optimizing hearing aid users' speech intelligibility by performing automatic volume settings instead of relying on manual adjustments. This thesis will use computational intelligence tools, such as neural networks, in order to supply optimal initial volume settings on the basis of input acoustic features characterizing various listening situations.

1.3 Thesis Objectives and Contributions

A selection of amplification schemes is typically desired by hearing aid users when faced with different hearing environments. Generally, this is achieved by an automatic sound classification system which provides the user with an optimal volume setting based on the type of environment the user is in (for example, speech in quiet)[3]. A drawback of performing a priori classification of the environment is that the input features used to classify the environment may not be sufficient to predict the user's preference volume gains, for example the volume gain requirement for loud speech in quiet can be different than soft speech in quiet even though both cases belong to the same class. It is important to use input features that are correlated to the user's preferred settings, instead of class type per se. Hence, an alternative approach is directly mapping environment sounds to user preference volume gains. This approach is typically accomplished using computational intelligence tools, such as neural networks and neurofuzzy systems. Hearing aids which perform adaptation can improve the user's speech intelligibility in changing environments, thus increasing their comfort level.

The main objective of this thesis is to optimize a user's speech intelligibility in different environments by performing automatic volume settings. More precisely, this thesis also aims to contribute the following to the existing body of research on hearing device development:

- Develop a system to set initial volume settings in a hearing aid which can optimally match user's preferences.
- Provide a method to select the best acoustic features to directly map features to user preferences in volume settings.
- Demonstrate the importance of either using acoustic features which are specifically correlated to the user's audiogram or using acoustic features unified to all types of user audiograms.
- Compare two different mapping algorithms to perform automatic volume settings, thereby eliminating the sound environment classification process for this application.
- Test the automatic volume control system with real sound recordings.

By achieving the above objectives may help minimize users' intervention with the hearing aid, thus eliminating stigmas and allowing users to lead a normal life.

1.4 Thesis Outline

The remainder of the thesis is organized as follows:

Chapter 2 briefly surveys the main concepts and techniques used to compensate for a specific hearing loss. It also reviews the various volume control techniques used in today's hearing aids. The chapter ends by providing a section describing various automatic volume control methods.

Chapter 3 provides a review of common computational intelligence tools, such as neural networks and neurofuzzy systems. The chapter gives an overview of their architecture, learning paradigm and function. The chapter is concluded with a comparison of the systems reviewed.

Chapter 4 proposes the automatic volume setting framework. The system's layout and the experimental data used to train and test the system. A description of the features extracted is discussed and the design requirements are also covered. Lastly, the chapter concludes with the measures used to evaluate the system's performances.

Chapter 5 presents the results and analyzes two computational intelligence tools used to perform the automatic volume settings. Results are evaluated using performance measures which demonstrate the system's success in optimizing a specific user's volume gains. Training and testing results are demonstrated for three user profiles with moderate, severe and profound hearing losses. The chapter also presents results and discusses the relevance of either using features strictly correlated to the user's audiogram or using features unified to all three hearing losses (moderate, severe, profound). The chapter is concluded by testing the systems with real sound recordings.

Chapter 6 conveys a conclusion of the proposed automatic volume setting framework. The chapter discusses the work and its findings as well as the contributions of the proposed solution. Finally, the thesis ends with a brief discussion of the potential future research directions.

Chapter 2

Hearing Aid Technology

This chapter begins by introducing the evolution of hearing aid technology and briefly discussing hearing loss and the available assistive devices. The chapter introduces the process of hearing aid fitting and the available prescriptive formulas. As well, it provides an overview of the various volume control methods that could potentially be used in hearing aids. An understanding of the anatomy of the ear and the auditory pathway is essential in evaluating and treating hearing loss. The following sections provide an overview of the causes of hearing loss and discuss several hearing aid technologies.

2.1 Hearing Loss

The human peripheral auditory system, presented in figure 2.1, includes three sections: the outer ear, the middle ear and the inner ear. Sound waves enter the outer ear, also known as the pinna, and travel down the ear canal, causing the eardrum to vibrate. In the middle ear, the vibration of the eardrum causes movement of the middle ear bones, also known as the ossicles. Here, the ossicles transfer the vibration of the eardrum to the oval window, and the movement of the oval window shifts fluid within the cochlea, a fluid-filled structure in the inner ear that contains approximately 20,000 hair cells. The shifting of the fluid in the cochlea ultimately causes hair cells to vibrate. When the hair cells vibrate, impulses in the nerve fibers are generated. The nerve fibers connect the cochlea with the brainstem, a portion of the central nervous system found above the spinal cord but below the brain. Their impulses reach the auditory cortex of the brain, allowing the perception of sounds [15]. Figure 2.1 also displays the auditory system in engineering terms. From an engineering perspective, impedance matching occurs at the middle ear between the air filled external ear and the fluid filled cochlea at the inner ear [24]. Furthermore, the inner ear can be considered a signal-processing device that performs filtering, amplification and nonlinear compression.

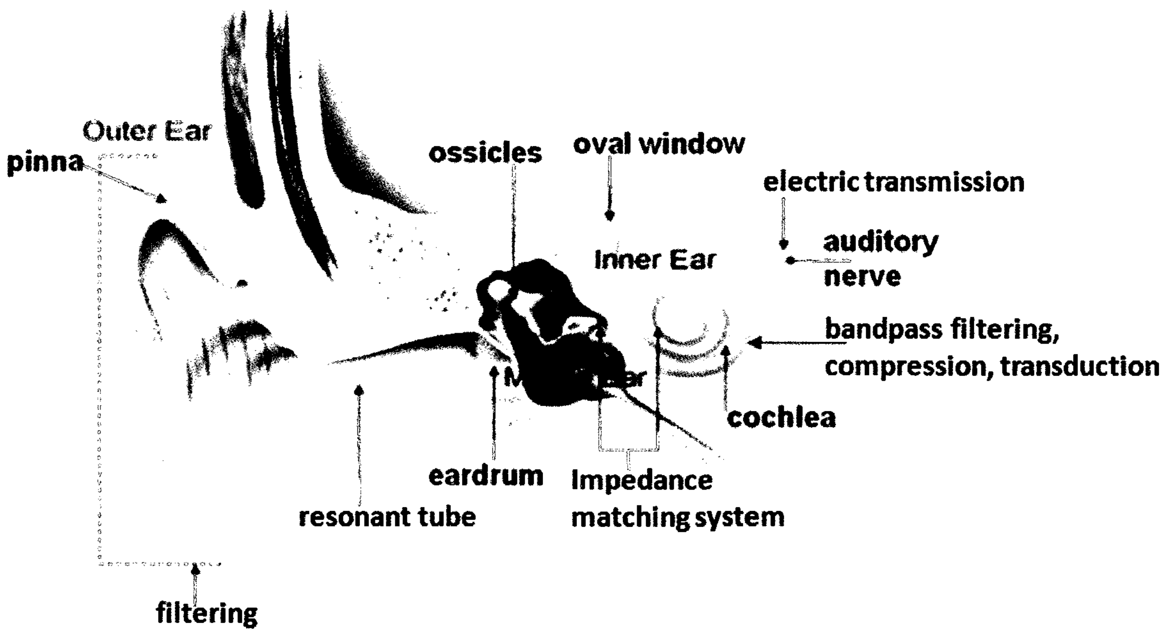


Figure 2.1: Cross-section of the human peripheral auditory system [2]

Having knowledge of the basic principles of the auditory system and some of the main types of hearing losses can help engineers develop appropriate assistive devices. The various types of hearing losses include conductive, sensorineural, mixed and central losses. Conductive hearing loss occurs when there is a problem in the outer and/or the middle ear. The hearing loss results when sound waves cannot be transmitted effectively to the inner ear. The causes include damage to the eardrum or to the ossicles in the middle ear. Sensorineural hearing loss occurs when there is a problem in the cochlea or the inner ear. This type of hearing loss may result for example from damage to the hair cells in the cochlea or to the auditory nerve. Sensorineural hearing loss can be further divided into sensory hearing loss, due to problems in the cochlea, and neural hearing loss, due to problems in the auditory nerve. Mixed hearing loss occurs when there are problems in both the inner ear or auditory nerve and the middle or outer ears [15]. Depending on the amount of elevation of auditory threshold in dB compared to a normal ear, the degree of hearing losses are usually classified as mild, moderate, moderately severe, severe, or profound. Individuals with sensorineural hearing loss also suffer from a reduced dynamic range in their hearing [11], compared to an individual with normal hearing.

The final type of hearing loss is the central hearing loss, which results from lesions or disorders within the pathways of the central auditory nervous system. The selection of assistive technology for an individual with hearing impairment depends on the type and degree of hearing loss and other characteristics [19].

2.2 Overview of Hearing Aids

2.2.1 Function of Hearing Aids

Sound is an important source of information which supports our ability to communicate [53]. The majority of face-to-face communication is by sound signals in the form of speech. However, hearing impaired individuals have difficulty effectively making use of sound signals for communication. Thus, assistive technology plays a major role in help hearing impaired individuals communicate effectively. One of the earliest hearing instrument developed was the ear trumpets. Ear trumpets were used to amplify acoustic signals by means of acoustic resonances. In the late nineteenth century, carbon microphones were also used as a hearing instrument. The carbon microphone provided electrical amplification in addition to acoustic amplification; however, the microphones produced excessive noise [19]. In 1910, vacuum tubes were used as a body instrument to construct hearing instruments and provided much greater acoustic amplification than was possible with carbon hearing instruments. In the 1940s, the invention of transistors replaced vacuum tubes, making it possible for users to wear the hearing instrument on their heads [24]. Today, the development of hearing instruments is at an advanced stage; body hearing instruments have been displaced by hearing aids which evolved in parallel with fully implanted devices.

Hearing aids are electronic devices with key components including a microphone, signal processor, receiver, controls and battery, as shown in figure 2.2. From figure 2.2, we see that the microphone converts incident sound into an electrical signal to the hearing aid body. The signal processor modifies the electrical signal to compensate for the user's hearing loss, such as by amplifying the sound frequencies to which the user has low sensitivity. The receiver converts the processed signal back into sound, an operation similar to that of a loudspeaker [7]. Control buttons allow the user to adjust the hearing aid and influence the output, amplification, and frequency response of the hearing instrument, for example, by adjusting the volume of the hearing aid. The battery provides power to all the units and the vent allows for a passage of air to avoid complete blockage of the ear. The vent also reduces the occlusion effect, which occurs when the hearing aid fully blocks the ear canal and the user hears its voice as being unpleasantly hollow and loud [24].

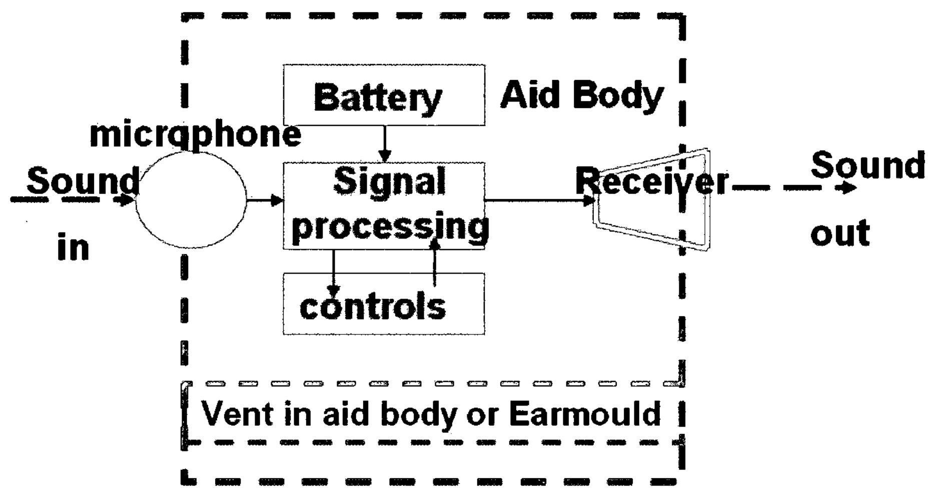


Figure 2.2: Typical hearing aid schematic diagram [24]

2.2.2 Categories of Hearing Aids

Hearing aids are categorized by their size and position on the user's ear. The most common hearing aids are behind-the-ear (BTE), in-the-ear (ITE) and completely-in-the-canal (CIC), as presented in figure 2.3. Since 2007, BTE hearing aids were the most commonly prescribed hearing aid in North America [53]. The receiver, microphone, signal processor, battery and controls are built in the hearing aid, which is worn behind the ear and the sound is carried through soft plastic tubing to the ear mold in the ear canal. The volume control is typically adjusted manually using rotary controls or wirelessly with a remote control.

ITE hearing aids have all the electronic and acoustic components built inside the hearing aid, which fits completely inside the outer ear, and are usually prescribed for those with mild to severe hearing loss. ITE hearing aids typically use wireless remote controls for volume adjustments. ITE hearing aids that cover a small portion of the cavum concha of the ear are known as in-the-canal (ITC) hearing aids.

Figure 2.4 displays the components typically located in an ITE and a BTE hearing aid. Hearing aids that fit entirely within the ear canal are known as completely-in-the-canal hearing aids [11]. In 1993, CIC hearing aids were introduced to the market, and in 1995, they grew quickly in popularity in North America due to the cosmetic appeal [11]. CIC aids are typically used for those with mild to moderately severe hearing loss because the battery used with the CIC aids is much smaller compared to that of the BTE aids. Thus, a smaller amplification gain is achievable, which is only satisfactory for those with mild to moderately severe hearing loss. A wireless remote control is also typically used to achieve amplification gain adjustments. BTE hearing aids are typically more reliable for those with mild to profound hearing loss than ITE and CIC aids. Since the battery is larger for BTE aids, more processing and amplification to the signal can be performed. BTEs and ITEs can also be fitted with directional microphones, in contrast to CICs.

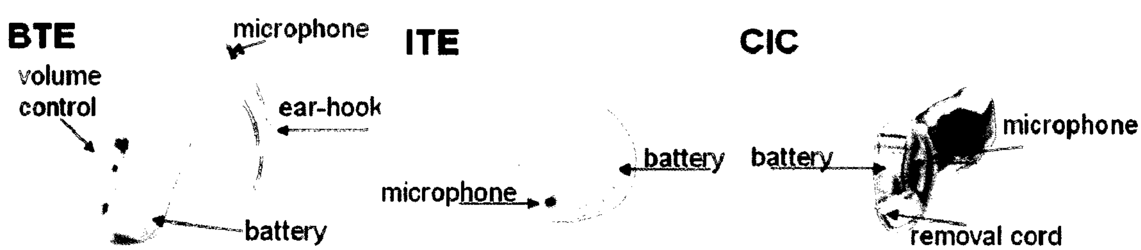


Figure 2.3: Typical BTE, ITE and CIC hearing aids [1]

2.2.3 Hearing Aid Fitting

Once a hearing aid is selected, a fitting procedure is performed to adjust the amplification characteristics of the device in order to compensate the hearing loss of the patient. First, the audiogram of the patient is measured at frequencies of 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz. Next, a target gain is calculated at each frequency based on how much gain the patient requires to compensate for their hearing loss. The target gains at those frequencies are then programmed into the hearing aid. Typically, audiologists use prescriptive methods to determine initial settings of gain based on the hearing loss of the patient. Prescriptive methods use a formula to determine the patient's gains from the patient's audiological data. There are several prescriptive procedures. For linear hearing

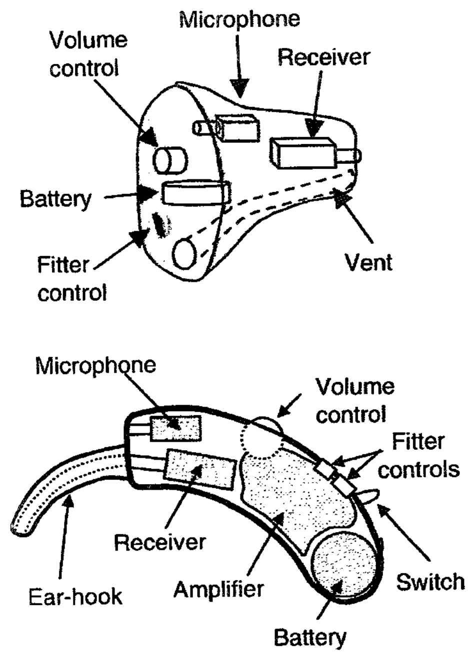


Figure 2.4: Typical components located in an ITC and a BTE hearing aid [11]

aids, typical methods are the POGO (Prescription Of Gain and Output), NAL (National Acoustic Laboratories of Australia), NAL-R (a revised formula of the NAL) and NAL-RP (a revised formula of the NAL and including the correction for profound hearing loss) prescription formulas, which aims to maximize speech intelligibility, audibility and comfort. Figure 2.5 shows an insertion frequency gain response prescribed by a NAL-R formula for a user with a flat 40 dB hearing loss. For nonlinear hearing aids, a common prescription is the NAL-NL1 which normalizes the overall loudness in order to maximize speech intelligibility [11]. Thus, there are many techniques which can be used to set the initial gains for a patient.

The prescriptive formulas above will provide the same gain in all listening situations and for all users who have the same hearing losses. Several researchers are seeking to customize the fitting process. Gao et al. [20] uses a combination of neural networks and fuzzy logic to achieve optimal gain requirements that enhance users' satisfaction. A neural network is used to generate initial target gains for a specific patient. The target gains generated by the neural network are the initial target gains provided to the user to be evaluated. To fine-tune the target gains to the user's preference, a fuzzy logic system is employed. The neurofuzzy system is helpful for off-line hearing aid fitting process.

The combination of both computational tools, fuzzy logic and neural networks, is a useful concept for automatic volume settings systems proposed for hearing aids. As well, Takagi and Ohsaki [50] introduced an interactive evolutionary computation hearing aid fitting method. Takagi and Ohsaki's hearing aid fitting technique uses evolutionary computation such as genetic algorithms to optimize a hearing aid based on the user's evaluation of their hearing. Despite those advanced fitting strategies, people want to have some control over their overall hearing, e.g. mean of a volume control button [41]. It is also well known that different users with similar hearing profiles have different preferences [30]. Therefore, hearing aid manufacturers have provided different means of allowing volume control. The next section covers several volume control techniques used in hearing aid technology.

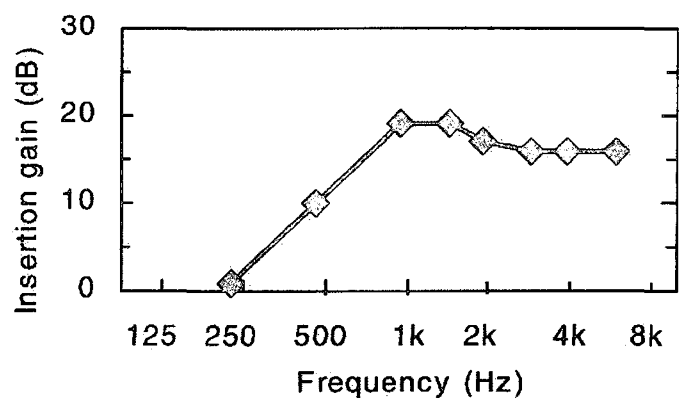


Figure 2.5: Typical frequency gain response using the NAL-R prescribed formula for a flat 40 dB hearing loss [11]

2.3 Volume Control in Hearing Aid Technology

2.3.1 Manual Volume Control

Typically, a potentiometer or a digital up/down counter are used to manually adjust the gain in order to vary the volume level of the input signal. There are several disadvantages of manually controlling the volume level, firstly the gain is fixed and there is no learning involved. Conversations also may be interrupted if the user needs to adjust the volume level setting and that valuable time is wasted selecting the appropriate volume setting.

As well if the patient uses two hearing aids, typically the user will have difficulty to balance the volume level in both hearing aids independently.

In most hearing aids, an Automatic Gain Controller (AGC) is typically built in along with the manual volume control. The Automatic Gain Control (AGC) is simply a compression amplifier, which decreases its gain if the output signal level is high or above an uncomfortable level for the user [11],[53].

2.3.2 Wireless Volume Control

Using a remote control to wirelessly adjust the volume levels manually is one method of minimizing the direct user's intervention with a hearing aid. Reyes et al. [46] designed a wireless volume control receiver for hearing aids. An advantage of using a wireless volume control is that it provides comfort to the user by using a remote control that gives the user the flexibility to adjust the volume and switch between programs of the hearing aid. As well, a wireless volume control is helpful for a patient with two hearing aids, by adjusting the volume level of both hearing aids simultaneously, rather then independently with the manual volume controller.

2.3.3 SMART Volume Control

Smart volume control settings have not only become popular in the hearing aid industry, the technique has also grown vastly in popularity in the field of consumer electronics. A smart acoustic volume controller, as defined by Kumar [34], is a controller which can "intelligently adjust the volume levels." As suggested by Kumar, advantages of using computational intelligence tools for the implementation of smart volume controllers are that these techniques are robust, computationally inexpensive, and low cost. The potential benefits of using smart volume controllers include that they minimize the user's interaction with the hearing aid, thus avoid drawing attention to the fact that the user is wearing a hearing aid. Smart volume controllers may also improve the user's speech quality by considering the user's hearing profile and the user's volume setting preference. These are promising statements for automatic volume setting systems for hearing aids. However, a drawback with smart volume controllers is that they may at times automatically provide an uncomfortable volume level to the user. Kumar solves this problem by suggesting to use a smart volume controller with a manual control option. This will allow the user to override the automatic volume control option with the manual control.

Several methods have been proposed for the development of smart volume control in order to enhance users' hearing satisfaction. Alexandre et al. [3] developed an automatic sound classification algorithm that provides users their preference hearing aid settings (e.g. volume level, directional microphones) for different listening environments. The automatic sound classification operates by performing feature extraction on the input signal. Several feature vectors considered were the percentage of low energy frames, loudness, mel-frequency cepstral coefficients and spectral flatness measure. Alexandre et al. used three classifiers to map the input feature vector into one of the four classes (speech in quiet, speech in noise, stationary noise or nonstationary noise). Lamarche [35] also developed an adaptive environmental classification system for hearing aids. Lamarche environmental classification system performed feature extraction on the input signal, where the features were dependent on the class and then using one classifier to distinguish the environment into three classes (speech, noise, and music). As well, Ravindran and Anderson [45] introduced an audio classification system that could be used to automatically switch between different hearing aid settings based on the user's environments. Büchler [8] reports that automatic switching for hearing aids using sound classification is favored, however further refinement in the selection of input features is critical in order to achieve a desirable performance with the classifier.

Another alternative to perform automatic volume control is to learn actual user preferences online with trainable hearing aids. Ypma et al. [56] developed a volume control algorithm that learns online a hearing aid user's volume gain preference. Two learning algorithms are implemented using the normalized least mean square and the Kalman filtering techniques to personalize the hearing aid. Feature extraction is performed on the input signal and the feature vectors are sent as inputs to an automatic volume controller. The feature vectors carry information such as short-term RMS and SNR estimates of the input signal. The parameters of the automatic volume controller are adjusted using the learning algorithms, providing the user with preferred volume gains. Ypma et al.'s learning volume control algorithms is advantageous because of its online ability to adaptively track a user's volume gain preferences.

2.4 Summary

This chapter provided background on hearing loss and the assistive devices available for compensating for specific hearing losses. A discussion on the hearing aid fitting process was also presented. Target gains for a hearing aid user are determined using prescribed formulas. Since these initial target gains are generated from a formula, they may not completely satisfy a specific patient's desired gains. As well, there exists a nonlinear relationship between the audiogram of the patient and the target gains, thus a computational intelligence tool such as a neural network model may be useful to determine target gains for hearing aids. Gao et al. [20] uses a neural network to generate initial target gains and applies a fuzzy logic system to fine tune the gains to the desire of the hearing aid user. As well, Takagi and Ohsaki [50] uses evolutionary computation for hearing aid fitting. Several volume control techniques were also introduced such as the manual volume control (rotary controls), wireless volume control (remote control) and SMART volume controllers. SMART volume controllers are becoming a popular technique in hearing aids. Techniques introduced which performed automatic hearing aid settings are [3], [35] and [45] who used an environment classification system to provide user preference hearing aid settings in various listening environments. However, as mentioned by Büchler [8], an appropriate set of features is essential in order to obtain satisfactory classification performance with the classifier. The drawback of performing environment classification in order to provide user's desired volume settings is that the input features used to classify the environments are correlated to the characteristics of the input signal rather than being correlated to the user's hearing aid settings (e.g. volume settings). Ypma et al. [56] developed an intelligent volume control algorithm that learns and adaptively tracks a user's volume setting preferences online. Ypma et al. excludes the process of performing sound classification to automatically adjust hearing aid settings. Ypma used learning algorithms such as the normalized least square and the Kalman filtering techniques to personalize the hearing aid. Thus, in this thesis, the step of performing environment classification is also excluded and features which are correlated to the volume setting are used and are directly mapped to the user's target volume settings using computational intelligence tools. The following chapter provides a review of the common computational intelligence tools used in an automatic volume controller.

Chapter 3

Computational Intelligence Tools

In this dissertation, computational intelligence tools were used to develop an automatic volume control system for hearing aids. Computational intelligence tools are considered intelligent machines because of their ability to interpret, learn and make decisions from incomplete information; in fact, many of these tools have been developed as decision support systems in engineering, medical, business, educational and agricultural fields. Artificial neural networks, fuzzy logic and genetic algorithms are typical computational intelligence tools used to mimic the decision-making of humans. Hybrid techniques, which combine the advantages of two or more of these tools, have been proven to be especially effective. This chapter provides an overview of the computational intelligence tools used to perform automatic volume settings in this thesis.

3.1 Artificial Neural Networks

3.1.1 Overview

Artificial neural networks were developed based on the biological architecture of neurons in the human brain. The advantage of neural networks is their ability to approximate arbitrary nonlinear functions and learn from incomplete information. Neural networks are adaptive models, which have a parallel processing structure of neurons (or nodes) organized into layers (input, hidden, output) and are connected by weights [23]. Figure 3.1 presents a typical representation of an artificial neural network. Neurons behave as simple processors, which take the weighted sum of their inputs from other neurons and apply nonlinear mapping called an activation function to them, generating the output of that neuron. A bias (or threshold value) is summed along with the inputs and network weights to shift up or down the output of the neurons. The operation of a neuron is depicted in figure 3.2. Only neurons in the hidden layers and the output layer are made up of an activation function (which differ depending on which layer the neuron is located) and a sum. Common activation functions used include the sigmoid function, signum function, step function, and linear function [28].

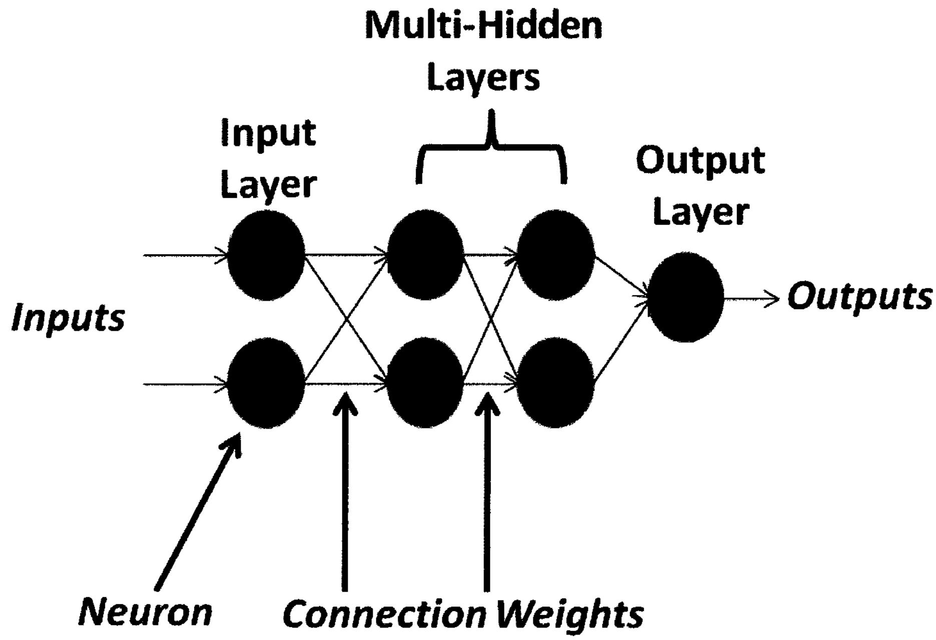


Figure 3.1: Typical representation of an artificial neural network [28]

A common type of artificial neural network topology is the feedforward network. Feedforward architecture steers information flow among the network neurons in the forward direction. Common neural networks which use the feedforward topology are multilayer perceptrons. These feedforward networks employ "supervised learning," which is discussed further in the next section [23].

3.1.2 Learning Paradigm

Neural networks interpret information through relationships found in the data provided to them; therefore, it is necessary that a relationship or correlation exists between the input and output data in order for the neural network to learn. A neural network can be trained to perform a particular function by adjusting the values of the connection weights between neurons. First, the neural network is presented with a priori known

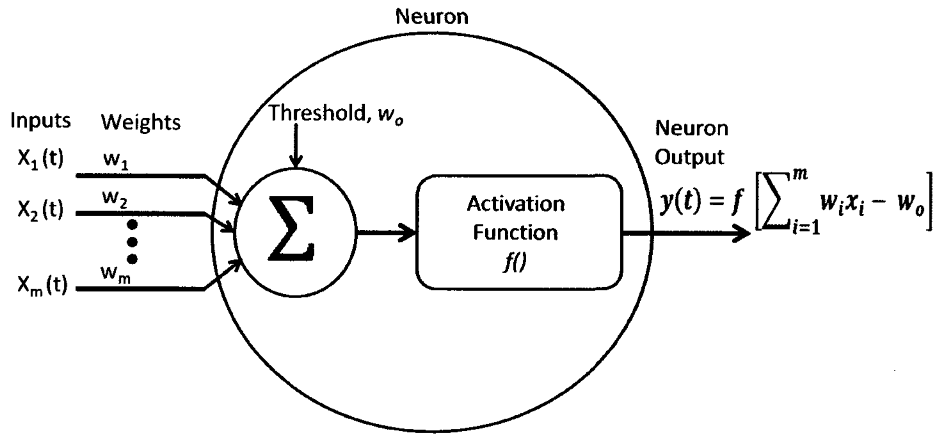


Figure 3.2: Operation at a neuron of an artificial neural network [28]

set of input and output data, to map a specific input to its appropriate output. In the case of automatic volume control settings, both input and output patterns (the features and the optimal volume settings) are a priori measured, thus supervised learning is used. Moreover, the learning mechanism of the supervised learning algorithm is an optimization process. During training, the output resulting from the neural network is being continuously compared with the target signal. The training algorithm uses the mean squared error (MSE) between the predicted volume setting from the neural network and the optimal volume setting to update the connection weights. The MSE is calculated as follows,

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left[optimal(i) - predicted(i)\right]^2 \tag{3.1}$$

where m is the number of training patterns, i is the pattern index, *optimal* refers to the target volume signal and *predicted* is the neural network's actual volume output signal [28].

The algorithm aims at minimizing the mean square error, such that there is a close match between the target output and the network's output. This reaction is depicted in figure 3.3.

A common supervise learning algorithm is the backpropagation algorithm. The back-propagation algorithm is based on the gradient descent techniques which minimizes the network error. The algorithm updates the weights in the direction of the gradient descent.

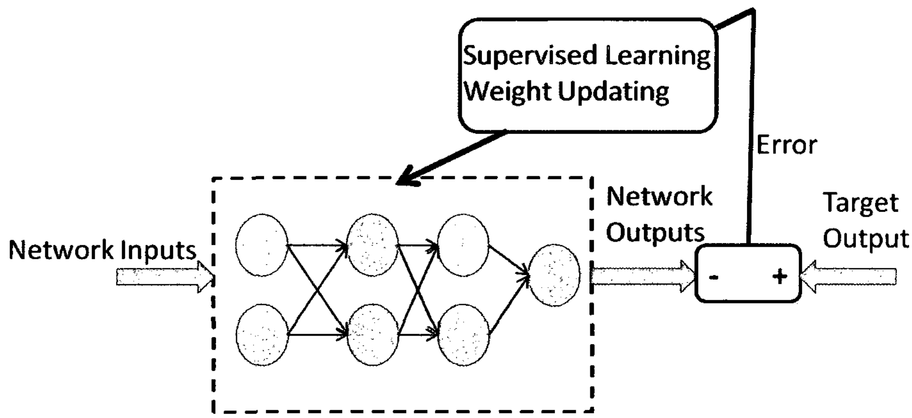


Figure 3.3: Typical representation of supervised learning [28]

3.2 Multilayer Perceptrons (MLP)

3.2.1 Overview

The multilayer perceptron is a very popular computational intelligence technique used in diverse fields such as medicine, mechanics, economics and engineering. Chang et al. [10] used a multilayer perceptron to improve vowel recognition performance for cochlea implant patients. Guler and Ubeyli [21] used a multilayer perceptron as an automatic diagnostic system to differentiate between ophthalmic arterial and internal carotid arterial Doppler ultrasound signals. Hadad et al. [22] employed a multilayer perceptron as a dynamic fault identification system for a nuclear power plant; the MLP provides plant operators with appropriate information that may necessitate corrective actions during hazardous situations. Another interesting application is described in Wang et al. [55], which used a multilayer perceptron for handwritten Chinese character recognition.

Multilayer perceptrons belong to a class of feedforward neural networks. The architecture of a MLP is typically made up of an input layer, a hidden layer and an output layer. Depending on the application, the neurons in the output layers can either be linear activation functions or nonlinear activation functions; however, for neurons in the hidden layer, the activation functions must be differentiable, such as by use of the sigmoid function or the tan-hyperbolic function [28]. This constraint is due to the MLP using the backpropagation algorithm as the learning algorithm.

3.2.2 Learning Process

Training

The training process requires a set of known network inputs and target outputs for the MLP to learn. During training, the weights and biases are iteratively adjusted to minimize the mean square error between the network outputs and the target outputs. The MLP uses a backpropagation algorithm as a mechanism for its learning process. However, many backpropagation variants have been implemented to improve the neural network's performance.

Validation

The validation process is a useful technique to improve the network's generalization during testing and avoid the network from "overfitting" the training data. After each training epoch, the MLP is given input and output validation data sets that the MLP was not trained with, and the validation error (MSE) is monitored. During the initial phase of training, both the error of the training set and the validation set decrease. Once the validation error increases, this is an indication that the trained network is no longer generalizing. Thus, the validation process can be used to end training. The final values of the weights and biases are set when the validation error was at its minimum.

Testing

Finally, the testing stage provides the MLP with unfamiliar data and determines whether the MLP has the capability to make accurate decisions. This testing stage will determine the model's robustness and ability to make generalizations on unfamiliar data. The following section describes another common computation intelligence tool considered for learning volume settings.

3.3 Adaptive Network-based Fuzzy Inference System (ANFIS)

3.3.1 Overview

Neural networks and fuzzy logic are decision-making tools. Fuzzy logic performs decision-making through *if-then* rules of linguistic, fuzzy descriptors (e.g., large, slow, far, moderate), which depends on expert knowledge and is not always available. Neural networks differ from fuzzy logic by representing knowledge within computational units and does not rely on expert knowledge, unlike fuzzy logic. Neurofuzzy systems such as the Adaptive Network-based Fuzzy Inference System (ANFIS) combine the advantages of neural network's ability to learn and fuzzy logic's accuracy in knowledge representation of a system. ANFIS has a hybrid architecture which integrates a neural network and a fuzzy logic system. An illustration of the ANFIS structure is presented in figure 3.4. ANFIS differs from the MLP because it uses fuzzy reasoning to map an input space to an output space. Decision-making performed by ANFIS is done by fuzzifying the crisp inputs (non-fuzzy value) to the system, applying fuzzy logic operations (rule evaluation) and then defuzzifying the output back into crisp values [27]. ANFIS are fuzzy Sugeno

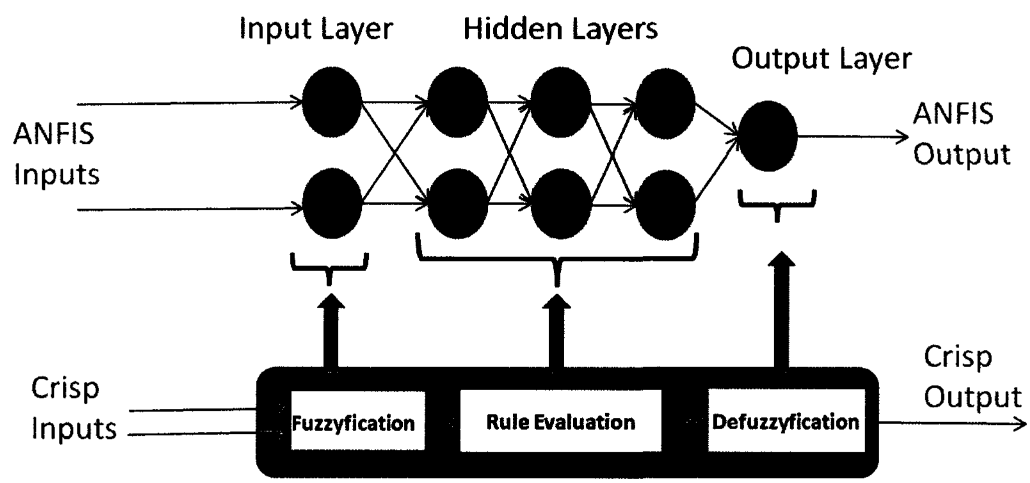


Figure 3.4: Typical representation of ANFIS system [28]

models [28] that have the ability to adapt and learn. Sugeno models are one method of fuzzy inferencing, which map a given input to a given output using fuzzy logic. The difference with the Sugeno models compared to other fuzzy inferencing techniques is that the Sugeno output membership functions are either linear or constant [6]. ANFIS is a five-layer system, where each layer can only be in the form of a first or zero order Sugeno-type systems with linear or constant output membership functions respectively, and unity weights for each rule [27]. During the rule evaluation, the fuzzy rules used for Sugeno-type fuzzy system are of the form:

$$Rule_p : If\ x_1\ is\ A_1^{(p)}\ and\ x_2\ is\ A_2^{(p)} ...and\ x_n\ is\ A_n^{(p)}\ Then\ O_p = \alpha_0^{(p)} + \alpha_1^{(p)} x_1 + ... + \alpha_n^{(p)} x_n \quad (3.2)$$

where x_i is the i^{th} input linguistic variable of the p^{th} rule with $i = 1, ..., n$. x_i refers to the antecedent part of the rule which corresponds to the input space. $A_i^{(p)}$ is a fuzzy set, which has a membership function of $\mu_{A_i^{(p)}}$. A membership function is used to represent the linguistic variable and this function gives the degree of possibility that the variable x_i belongs to the fuzzy set A_i. O_p is the consequent output (the decision) of the p-th rule and $\alpha_0^{(p)}$, $\alpha_1^{(p)}$,..., $\alpha_n^{(p)}$ are its consequent parameters. The tasks for each layer in the ANFIS structure are organized as follows:

(i) Layer one performs the fuzzification, where crisp inputs are fuzzified through mapping into membership functions. In this work, Gaussian membership functions are used to represent the linguistic variable x_i.

$$\mu_{A_i^{(p)}}(x_i) = e^{\frac{-(x_i - C_p)^2}{2a_p^2}} \quad (3.3)$$

The Gaussian function depends on two parameters a_p and C_p, where p is the p-th rule. The parameter C_p locates the center of curve and a_p is the width of the curve [38].

(ii) The rule evaluation begins at Layer two. Layer two is the rule nodes layer, where each node represents a fuzzy *if-then* rule and is connected to the nodes in the previous layer to form the antecedent of the rule (*if part*). The degree to which the antecedent part is fulfilled is known as the firing strength (w_p).

(iii) Layer three normalizes the firing strengths of the fuzzy rules in order to determine the ratio of the p-th rule firing strength to the sum of all rules' firing strengths,

$$\overline{w}_p = \frac{w_p}{\sum_p w_p} \quad (3.4)$$

where $\overline{w}_p$ is the normalized firing strength for the p-th rule.

(iv) Layer four is the consequent layer, which computes the consequent (*then* part) of the fuzzy rule. The normalized firing strengths from the third layer are multiplied with the variables from the consequent part,

$$\overline{O}_p = \overline{w}_p O_p = \overline{w}_p(\alpha_0^{(p)} + \alpha_1^{(p)} x_1 + ... + \alpha_n^{(p)} x_n) \tag{3.5}$$

(v) Finally, Layer five computes the overall output as a summation of the incoming signals:

$$O = \sum_p \overline{O}_p \tag{3.6}$$

3.3.2 Learning Process

The learning process of the ANFIS model is similar to the MLP model technique: training, validation and testing are performed. During the learning process, the parameters of the membership functions, such as the center and the width, change so as to minimize an error measure. This determines how well the ANFIS is modeling the input and output data set. Thus, comparing the MLP and ANFIS systems in terms of self-learning, the MLP applies the backpropagation gradient descent method to update the weights of the neural network and mimic a given data set. On the other hand, ANFIS uses the backpropagation gradient descent method for tuning the parameters of the input membership functions and the least-squares error technique to optimize those of the output membership functions [27].

3.4 Summary

This chapter provided an overview of the computational intelligence tools used in the system proposed in this thesis, the MLP and ANFIS models. The functionality and theory of both models were discussed. The learning process of both models involves (1) training to allow the model to become familiar with every possible outcome, (2) validation to prevent the model from overfitting, and (3) testing to ensure the model can perform generalizations on unfamiliar data. The primary distinction between the MLP and ANFIS is that the latter is a hybrid algorithm, combining both neural network and fuzzy logic theory. The ANFIS model uses a neural network structure but operates in a fuzzy logic manner. Nevertheless, both models are potential candidates to learn automatic volume settings. The following chapter discusses the framework of the proposed system and presents the role of the computational intelligence tools adopted.

Chapter 4

Proposed Automatic Volume Setting Framework

An objective of this thesis is to provide optimal initial volume setting of the hearing instrument to optimize speech intelligibility in all acoustic environments. A proposed solution is to perform automatic volume control, by mapping acoustic features to the user's preference volume setting. The general framework to perform automatic volume settings is described in this chapter. The experimental data collected and used for training and testing for the system is depicted. As well, the design requirements and the performance measures are presented.

4.1 System Layout

The automatic volume control system proposed contains the following components: a large data set of environmental audio files, a feature extraction module, a process to select the most influential features, a mapping of features to their optimal volume setting and finally a smoothing of the output with a post-processor. The experimental procedure is depicted in figure 4.1.

First, a database of audio files is run through a hearing aid simulator to obtain target volume settings for a specific user. Next, feature vectors are extracted from each file via a dedicated feature extraction module, shown in figure 4.1. A module then selects the most influential features which are highly correlated to the user's target volume settings. This process allows to build training and testing sets for the networks to learn target volume settings for a given hearing aid user from the input acoustic features. The network then maps the selected features to the user's optimal volume settings and post-processing is performed to smoothen the output. The system's components are described in further detail throughout this chapter.

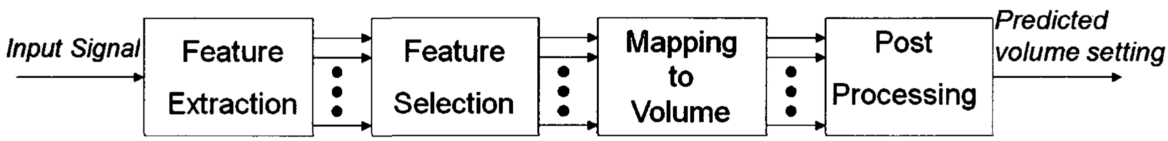


Figure 4.1: Experimental procedure

4.2 Experimental Data

4.2.1 Audio Files

It is desirable to have data which mimics speech and environmental noises that a user would typically face such that the system will be valid in various environmental conditions. Twenty pure speech files provided by SIEMENS were used to generate the noisy speech files, which are made up of ten female and ten male pure speech files (shown in table 4.1). Table 4.1 defines the set of acoustic conditions used to generate the noisy speech files. Each pure speech file was used to generate a hundred different noisy audio files using various combinations of acoustic conditions, thus generating two thousand audio files in all.

Various types and magnitudes of noise files were mixed with the pure speech signals. Twelve types of noise files were used, four were artificial noises (white noise, low frequency noise, pink noise and Brownian noise) and eight were real environmental noises provided by SIEMENS (cafeteria, automobile, music, airplane, street, radio, television series and kindergarten)[51]. Linear and nonlinear distortions were also introduced to the signal, such as reverberation and channel distortion (e.g. signal clipping and band-limiting distortion), as well as conditions without distortion. Reverberation distortion was applied by selecting a typical size of a room, source signal location, noise location, listener's location and observation angle. From these parameters the room impulse response was calculated and generated the reverberated signal. As well, the volume of the overall output signal for the two thousand noisy speech files ranged from 40 to 85 dB SPL, which covers soft and loud speech levels. An environment simulator was used to generate the distorted speech files. Shown in figure 4.2 the environment simulator requires the noise file and the pure speech file to generate a noisy audio file. The environment simulator can also emulate speech distortion and noise mixing in real environments, thus creating realistic training and testing data for the system. The sound database used for the experiments consisted of a total of two thousand noisy speech files, which mimic typical environmental conditions. The audio files were limited to only speech conditions and not music. Each file is 30 seconds in length and sampled at a frequency of 20000 Hz.

20 pure speech files (10 female + 10 male)
12 noise files (4 artificial noises + 8 real environmental noises)
Linear Distortion (with/without reverberation)
Channel Distortion (clipping, band-limiting, no channel distortion)
Volume of Overall Output Signal (40 to 85 dB SPL, covers soft to loud speech)

Table 4.1: Set of acoustic conditions

4.3 Hearing Aid Simulator

4.3.1 Target Volume Settings

Typically, the optimal volume settings should be obtained from hearing aid users to select their volume level preference when listening to the audio files. However, since this data is not available, the target volume settings are obtained through a simulated hearing aid user assumed to adjust its hearing aid to optimize intelligibility at all times [44]. This set of targets, defined as the volume settings for maximum Speech Intelligibility Index (SII) [44] in all environment conditions, is required to train the networks to map features into optimal volume. In figure 4.2 the hearing aid simulator can emulate interactions between acoustic environments (e.g. input sound) and the user's adjustments of hearing aid control settings. The hearing aid simulator uses objective measures such as the SII to predict likely user hearing aid interactions over time for a specific user profile (e.g. audiogram, uncomfortable listening levels (UCLs)) and input sound. Deriving the optimal volume setting with the hearing aid simulator requires access to both the output signal of the hearing aid and the pure input speech signal, as shown in figure 4.2. Owing to nonlinear processing stages in hearing aids, the coherence speech intelligibility index (CSII) [29] is used to estimate the SII. The time required to obtain the optimal volume setting for each file of 30 seconds took approximately 30 minutes per file. Initially, the hearing aid simulator used an exhaustive search over a possible range of volume settings in order to determine the optimal volume setting. This was very time consuming to search a range of volume extending from -60 to 60 dB with a 1 dB-step which involved 120 cases to evaluate. The hearing aid simulator was then modified from the exhaustive search to a fast search technique which measured the speech intelligibility at non-uniformly spaced points. However, using the fast search technique required approximately 2 minutes per file to obtain the optimal volume setting. The time required to obtain the optimal volume setting and the requirement for the pure input speech signal in the SII calculations are such that the hearing aid simulator is only usable offline and is not applicable for real-time hearing aids. Instead, it is used in this research to generate realistic training data for fast learning algorithms of volume control, in lieu of subjective data from real hearing aid users.

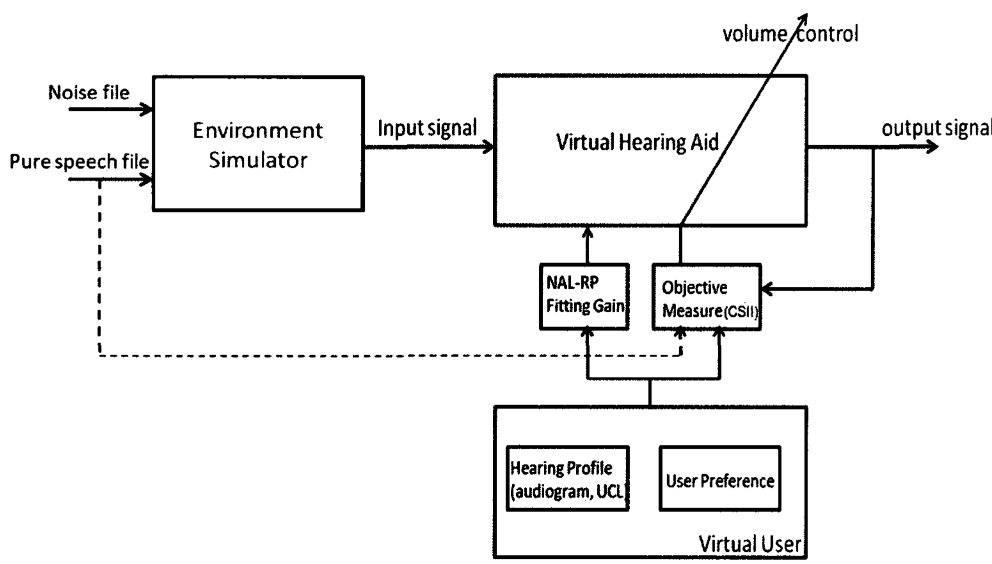


Figure 4.2: Generation of target volume settings

4.3.2 Virtual Hearing Aid

The virtual hearing aid shown in figure 4.2 is a linear hearing aid and uses the NAL-RP method to provide initial target gains for the hearing aid user [44]. NAL-RP uses the audiogram of a user to provide gain targets at frequencies 0.25, 0.5, 1, 2, 3, 4 and 6 kHz. The virtual hearing aid also limits the output signal by peak clipping the signal when the signal is higher than the user's UCL. As well, the virtual hearing aid has a volume control (amplifier) and a spectral tilt control for the user to adjust.

The profile of the virtual user also shown in figure 4.2, includes the user's audiogram and UCL. Users with moderate, severe and profound hearing losses (HL) and different UCLs were simulated. For each subject, the target volume level optimizing the SII is obtained at a 1 dB resolution for each of the two thousand audio files.

Table 4.2 presents the subjects' profiles simulated to obtain the set of optimal volume settings. Hearing levels are measured at the given frequencies presented in the table and are taken from real subjects.

Hearing Loss [dB HL]	**Frequency [kHz]**							**UCL [dB SPL]**
	0.25	**0.5**	**1**	**2**	**3**	**4**	**6**	
Moderate	10	15	10	20	10	45	55	120
Severe	20	15	15	30	70	80	80	105
Profound	35	60	70	90	93	95	98	110

Table 4.2: Subjects' profiles

4.3.3 Speech Intelligibility Index

One of the objectives of this thesis is to learn preferred volume settings. In the hearing aid simulator [44], preferred volume settings is assumed to be the volume setting (in dB) that maximizes the SII. Thus, obtaining the maximum SII and the associated gain corresponding to each of the two thousand audio files for a particular subject is an essential task to perform. The SII provides a measure of understanding speech for a particular subject in different environmental conditions and is "highly correlated with intelligibility of speech" [25]. The SII is typically calculated by summing the signal-to-noise ratio (in dB) in each frequency band and taking into account auditory thresholds and frequency-domain masking effects [25]. Calculating the SII is described in detail in the ANSI S3.5 standard [5]. The SII is directly related to the amount of audible speech information available to the user [26]. As well, the SII is mostly applicable for stationary noises. The value of SII ranges from 0.0 to 1.0. An SII value of 0.0 indicates that the speech information is not audible or usable for speech understanding for a particular user. Likewise, an SII value of 1.0 indicates that all of the speech information is audible and usable for speech understanding for a particular user [25]. Thus, if the SII is high for a specific audio file, it indicates that the subject can comprehend the speech. Fletcher and Galt [18] describes the relationship between the SII index and speech intelligibility scores. In the most sensitive portion of the relationship, a 0.04 increase in SII typically incurs to a 10% improvement in word intelligibility scores.

The hearing aid simulator was used to obtain the volume settings for optimal SII for the subjects listed in table 4.2, for each of the two thousand audio files. A study conducted by Othman [44] investigated the relationship between the SII and the volume settings for specific user profile and input sound. Figures 4.3, 4.4 and 4.5 present the estimated speech intelligibility curves for three simulated users with moderate, severe and profound hearing losses, respectively.

The speech intelligibility curves are plotted for a moderately noisy input at 65 dB SPL with different SNR levels (-10 dB to 30 dB) and different volume settings (-60 dB to 60 dB). The squares on the intelligibility curves present the location of the optimal volume gain and optimal SII. Figures 4.3, 4.4 and 4.5 shows that the size of the plateau of the intelligibility curves depends on the user's hearing loss. For instance, a moderate hearing loss user would have a larger saturation plateau (as shown in figure 4.3) compared to a profound hearing loss user whose plateau would be narrower (figure 4.5). Thus, for a moderate hearing loss, due to the wide saturation plateau, a large error in volume will not necessarily affect intelligibility. For a profound hearing loss, a small error in volume may lead to a large intelligibility decrease due to a narrower saturation plateau.

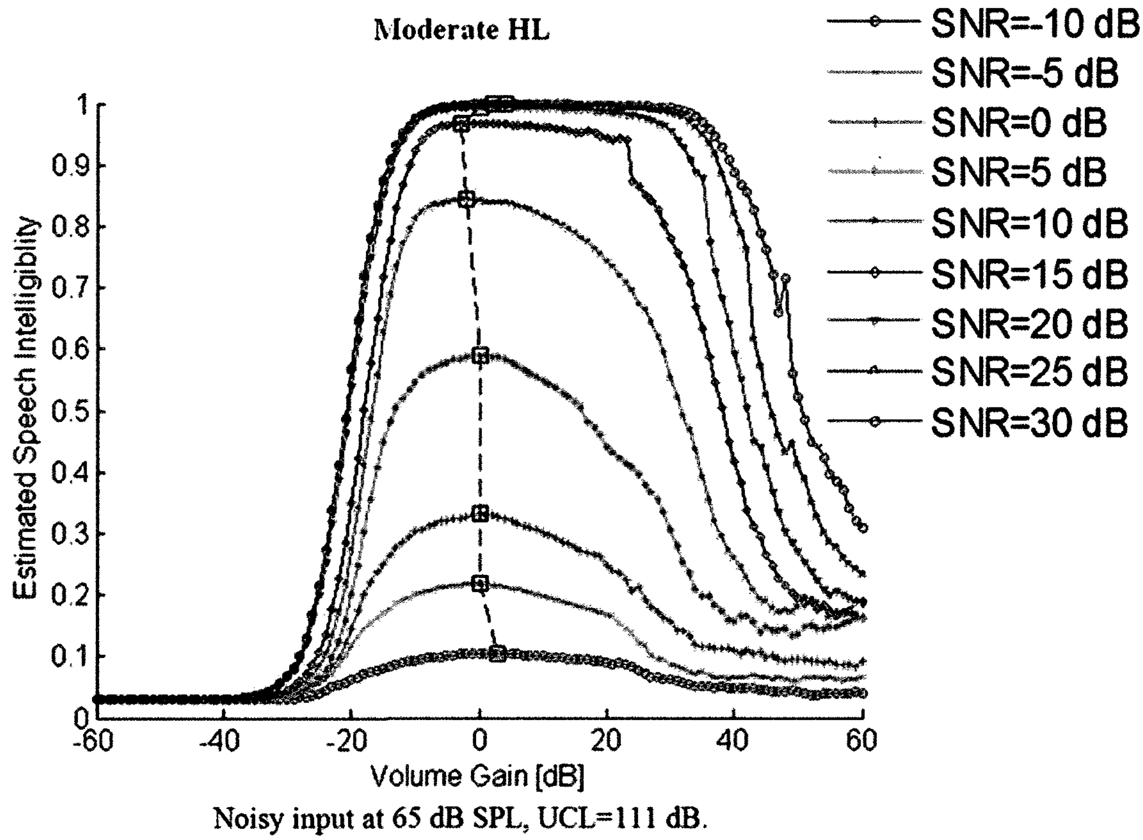


Figure 4.3: Estimated speech intelligibility (SII index) curves for moderate hearing loss, noisy input at 65 dB SPL [44]

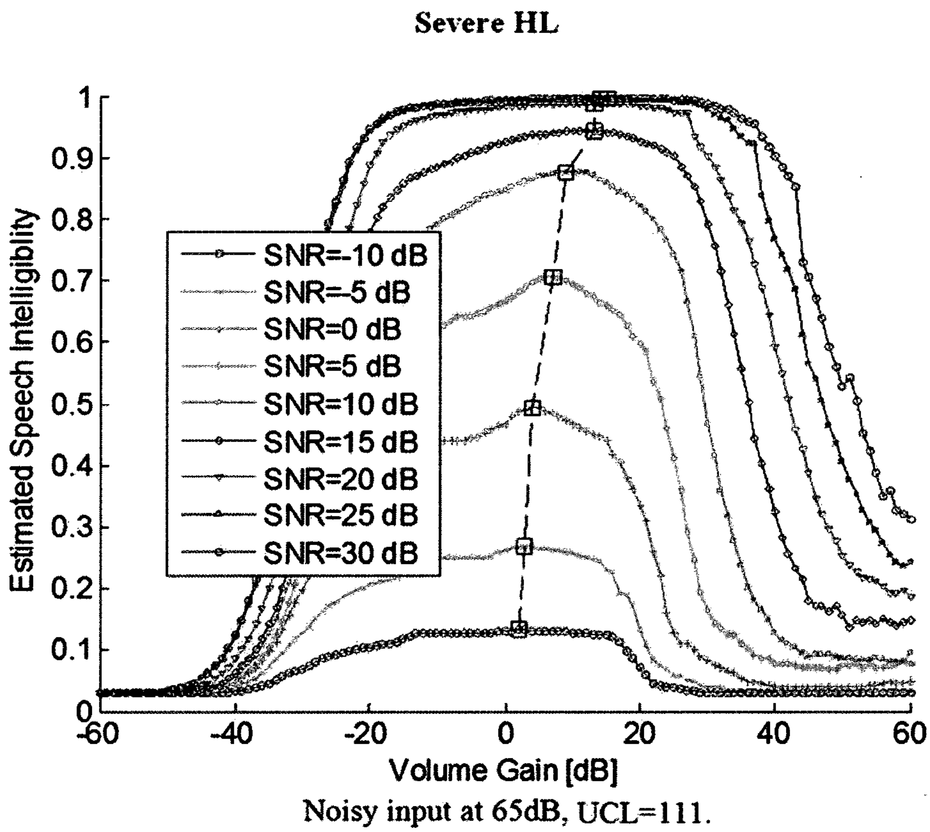


Figure 4.4: Estimated speech intelligibility (SII index) curves for severe hearing loss, noisy input at 65 dB SPL [44]

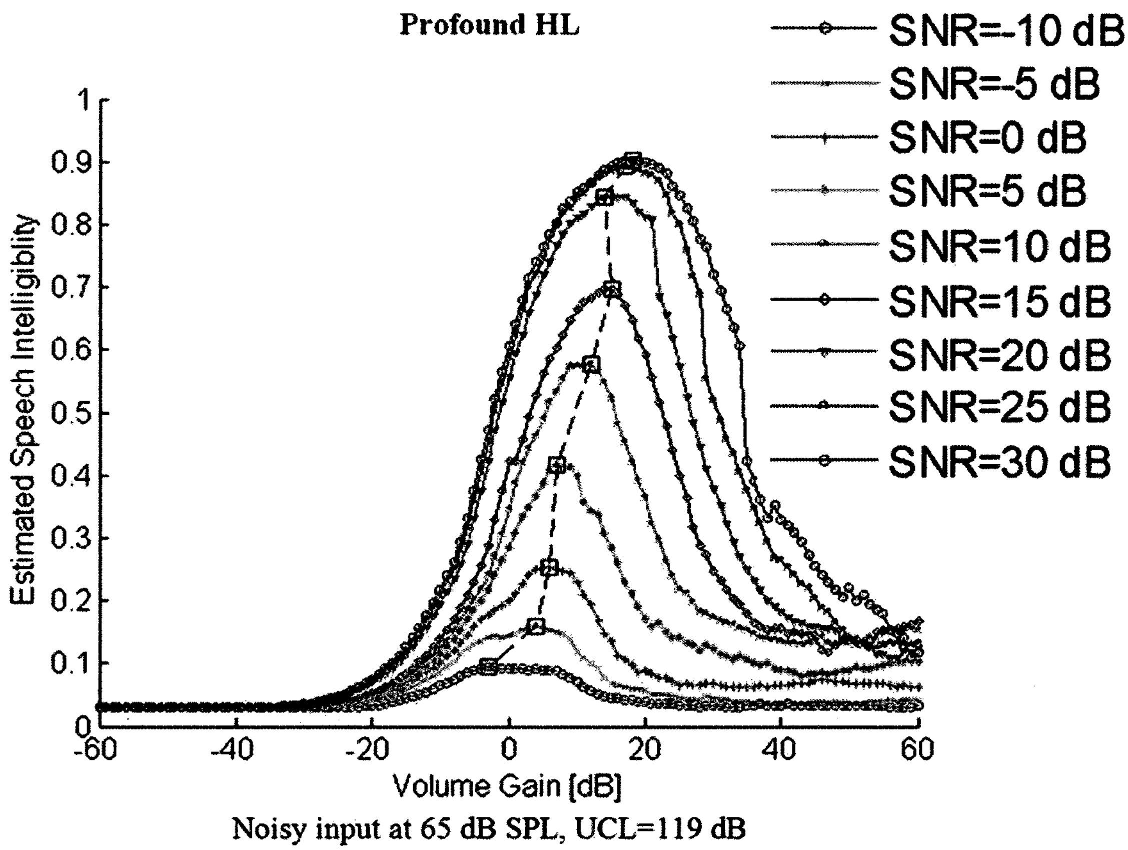


Figure 4.5: Estimated speech intelligibility (SII index) curves for profound hearing loss, noisy input at 65 dB SPL [44]

4.4 Feature Extraction

Feature extraction is a method of obtaining a set of characteristic vectors from a source [17]. To perform mapping of the feature input space to the optimum volume, it is desirable to extract a set of feature vectors which preserves information that is highly correlated to the output (e.g. volume level). Doing so will improve the performance of the network and decrease processing time. There is a wide range of features to select from depending on the application, thus energy-related features were investigated in this thesis. Features are typically expressed in the time or frequency domain. Alexander et al. [3] used a combination of energy related features in the time and frequency domains for sound classification, such as the zero crossing rate (ZCR) and the spectral centroid, respectively. As well, Liu et al. [36] used the time-domain feature, volume dynamic ratio (VDR) for content-based audio classification, which is also related to the energy of the signal. Features are typically extracted from the frames of a sampled audio signal, where the frames are usually either overlapped by 50% or not overlapped at all.

In this application, feature extraction is performed on each audio file which are 30 seconds in length and sampled at a frequency of 20 kHz. Frame by frame processing is performed during feature extraction and the input audio signal is segmented into frames of 1000 samples each (50 ms considering the sampling frequency), and with no overlap between adjacent frames. Features are averaged over the length of the audio file (30 seconds) and obtained an input matrix of the size (number of audio files)x(number of features). For example, a database of 2000 audio files and extracting 24 features, gave an input size of 2000x24.

4.4.1 Features

Initially, a set of 24 time-domain and frequency-domain features was considered. The features are listed in table 4.3. The first eight features presented in the table are energy-related features implemented for this project whereas the other 16 features were obtained from a module provided by SIEMENS [51]. The eight energy-related features are described in the following sections.

Spectral Centroid

The spectral centroid is a frequency-domain feature, which indicates where the center of mass of the spectrum. The spectral centroid of a frame is calculated by multiplying each

Feature Number	**Features**
1	Spectral Centroid
2	Spectral Rolloff Point
3	Short-Time Energy
4	Low Short-Time Energy Ratio
5	Zero Crossing Rate
6	High Zero Crossing Rate Ratio
7	Volume Dynamic Ratio
8	Spectral Flux
9	Amplitude Modulation at 0-4Hz
10	Amplitude Modulation at 4-16Hz
11	Amplitude Modulation at 16-64Hz
12	Average Of Spectral Center Of Gravity
13	Fluctuation Of Spectral Center Of Gravity
14-18	Amplitude Distribution Percentiles 10%, 30%, 50%, 70%, 90%
19	Symmetry
20	Skewness
21	Lower Half
22	Temporal Onsets
23	Onset Mean
24	Onset Variance

Table 4.3: List of features

frequency bin by its magnitude spectrum, taking the sum and then normalizing by the sum of its magnitude spectrum. The spectral centroid is computed as follows,

$$Centroid_t = \frac{\sum_{k=1}^{K} (|X_t[k]| \cdot k)}{\sum_{k=1}^{K} |X_t[k]|} \tag{4.1}$$

where K is the total number of frequency bins in a frame t and $X_t[k]$ is the k-th frequency bin of the spectrum at frame t [3]. The spectral centroid is also a measure of brightness of a sound [13] and perceptual research on musical timber states that brightness is one of the perceptually strongest distinctions between sounds [40].

Spectral Rolloff Point

The spectral rolloff point is another common frequency-domain feature, which is related to the shape of the power spectral distribution. It is the frequency ($RollOff_t$) below which a given percentage (typically 85%) of the magnitude distribution is concentrated [3]. The spectral rolloff point of a frame is calculated by multiplying the sum of its magnitude spectrum with the percentile percentage, which is presented by,

$$\sum_{k=1}^{RollOff_t} |X_t[k]| = PR \cdot \sum_{k=1}^{K} |X_t[k]| \tag{4.2}$$

where K is the total number of frequency bins in a frame t and $X_t[k]$ is the k-th frequency bin of the spectrum at frame t [3]. The spectral rolloff point can be used to describe the shape of the spectrum [13], which may be useful to distinguish low-frequency sounds from high-frequency sounds.

Short-Time Energy

Speech is time-varying and non-stationary and can be considered nearly stationary if speech is split into short segments. The short-time energy (STE) of a frame is computed as follows,

$$STE_t = \sum_{n=1}^{M} |x_t(n)|^2 \tag{4.3}$$

where M is the number of samples in a frame t and $x_t(n)$ is the discrete time audio signal [3]. Short-time energy is a common time-domain feature used in many voice activity detections, where it is well known to be effective [16]. The short-time energy may be useful to identify loud sounds, which have higher energy content, from softer sounds. As well, the short-time energy is dependent on the energy, which may be highly correlated to the volume level.

Low Short-Time Energy Ratio

Low short-time energy ratio (LSTER) is obtained from the short-time energy and is defined as the ratio of frames whose STE is 0.5 times below the mean STE. The low short-time energy ratio is computed as follows,

$$LSTER = \frac{1}{2N}\sum_{t=1}^{N}[sgn(0.5avSTE - STE_t) + 1] \tag{4.4}$$

where:

$$avSTE = \frac{1}{N}\sum_{t=1}^{N} STE_t \tag{4.5}$$

$$sgn(y) = 1, y \geq 0 \quad , sgn(y) = -1, y < 0 \tag{4.6}$$

STE_t is the short-time energy at frame t, where t is the frame index, N is the total number of frames, $avSTE$ is the average STE and sgn() is the sign function. The LSTER is a time-domain feature and is useful to identify sounds with silent frames, where LSTER is high for sounds with more silent frames and lower for sounds with less silent frames [37].

Zero Crossing Rate

Zero crossing rate (ZCR) is the number of times a signal changes sign in a frame and is defined as the number of time-domain crossings in a frame. The zero crossing rate for a frame is defined as follows,

$$ZCR_t = \frac{1}{2}\sum_{n=1}^{M}|sgn(x_t[n]) - sgn(x_t[n-1])| \tag{4.7}$$

where M is the number of samples in frame t and *sgn()* is the sign function [3]. ZCR is a time-domain feature and it may be helpful to distinguish sounds that have successive samples with different signs. Thus, a low ZCR may indicate sounds rich in low frequencies.

High Zero Crossing Rate Ratio

The high zero crossing rate ratio (HZCRR) is derived from the ZCR feature and is defined as the ratio of frames whose ZCR is 1.5 times above the mean ZCR. The HZCRR is computed as follows,

$$HZCRR = \frac{1}{2N} \sum_{t=1}^{N} [sgn(ZCR_t - 1.5avZCR) + 1] \tag{4.8}$$

where:

$$avZCR = \frac{1}{N} \sum_{t=1}^{N} ZCR_t \tag{4.9}$$

and t is the frame index, ZCR_t is the zero crossing rate at frame t, N is the total number of frames, *avZCR* is the average ZCR and *sgn()* is the sign function [37]. HZCRR is a popular time-domain feature used to identify sounds with alternating high and low frequency content [3].

Volume Dynamic Ratio

The volume dynamic ratio (VDR) is another common time-domain feature and is defined as the difference between maximum and minimum frame root mean square (RMS), normalized by the maximum RMS. The VDR is presented in the following equation:

$$VDR = \frac{max_{\forall t}(RMS_t) - min_{\forall t}(RMS_t)}{max_{\forall t}(RMS_t)} \tag{4.10}$$

where:

$$RMS_t = \sqrt{(1/M) \sum_{n=1}^{M} (x_t(n)^2)} \tag{4.11}$$

and $max_{\forall t}$ (RMS_t) is the maximum RMS across all frames of a signal and $min_{\forall t}$ (RMS_t) is the minimum RMS across all frames of a signal. RMS_t is the root mean square for a frame t and M is the number of samples in frame t [37]. The volume dynamic ratio is a useful feature since it is a measure of energy fluctuation. Thus, highly-fluctuating sounds will have a larger *VDR* than continuous or steady-state signals.

Spectral Flux

The spectral flux is the amount of spectral local changes between two adjacent frames and is computed as follows,

$$flux_t = \sum_{k=1}^{K} (|X_t[k]| - |X_{t-1}[k]|)^2 \tag{4.12}$$

where K is total number of frequency bins in a frame t and $X_t[k]$ is the k-th frequency bin of the spectrum at frame t [3]. The spectral flux is a common frequency-domain feature which is a measure of how quickly the power spectrum of a signal is varying.

4.5 Design Requirements

4.5.1 Feature Selection

Hearing aids have limited resources for digital signal processing (DSP) due to the small size of the device. This leads to complexity constraints on computation capacity and memory. Thus, it is essential to reduce the number of input features to the optimization algorithm as each must be programmed on the DSP chip. Furthermore, irrelevant features may act as noise to the system and degrade the system's performance. Alexandre et al. [4] used a genetic algorithm to perform feature selection and found a subset from a set of features which maximizes the probability of correct classification. There are also feature selection algorithms such as the sequential forward search (SFS) and the sequential backward search (SBS) algorithms, which can use the Multi-Layer Perceptron (MLP).

Romeo and Sopena [47] performed feature selection using the SBS algorithm. The SBS algorithm performs feature selection by initially training the MLP with the complete set of available features. After each epoch, a feature is removed based on the lowest evaluation criterion and the weights of the MLP are adjusted. The best subset of features is determined when further removal of a feature causes the performance of the MLP to degrade. Onnia et al. [43] uses the SFS technique to perform feature selection. Unlike the SBS, the SFS trains the MLP with one input feature at a time from the set of available features. The SFS selects an influential feature based on the feature which obtains the highest classification accuracy for the test data. The process is repeated for all the available features and the subset of influential features is determined once an evaluation criterion has been achieved. Verma and Zakos [52] also performed feature selection based on a neural network classification and the features which obtained the best classification rate were selected. Selecting features which are highly correlated to the output space is a challenging topic. Nevertheless, feature selection can provide a near optimal set of features and reduce the dimensionality of the feature vector. In this thesis, the SFS feature selection method was used and applied on a starting set of 24 features (table 4.3) which selected a subset of highly correlated features.

The SFS method applied in this thesis uses Adaptive Network-based Fuzzy Inference System (ANFIS) modeling to perform feature selection. The method selects the B most influential features from the set of A available features, by processing (2*A-B+1)*B/2 ANFIS models [39]. The method performs feature selection by generating an ANFIS model for each available feature and determines the root mean squared error for each model. The best single feature is selected first, based on the model that led to the lowest error. The next influential feature is selected by combining each feature input with the features already selected and again determining the model with the least error. This process is repeated until all the B influential features are selected.

The SFS method was used to select three sets of influential features. Each set corresponds to one of the simulated users (moderate, severe and profound hearing loss) shown in tables 4.4, 4.5 and 4.6, respectively. The SFS was also used to find one set of unified influential features for all three simulated users, presented in table 4.7, hence the features are not biased towards a specific user's hearing loss profile.

2 Features	**4 Features**	**6 Features**	**8 Features**
Spectral Rolloff Point	Spectral Rolloff Point	Spectral Rolloff Point	Spectral Rolloff Point
Symmetry	Symmetry	Symmetry	Symmetry
	Short time energy	Short time energy	Short time energy
	Percentile 50%	Percentile 50%	Percentile 50%
		Low short time energy ratio	Low short time energy ratio
		Skewness	Skewness
			Percentile 70%
			Onset variance

Table 4.4: List of most influential features for moderate HL

4.5.2 Mapping to Volume

Neural networks have been found to be highly successful in the area of pattern recognition [28] and one goal of this dissertation is to directly map environmental sounds to a user's target volume setting, without an explicit stage of sound classification. Two prediction models were used to perform the mapping task, the Multilayer Perceptron (MLP) and the Adaptive Network-based Fuzzy Inference System (ANFIS). Both models are trained in a supervised manner where the input and target data are a priori known. After training, the models are tested by analyzing their performance with new patterns in

2 Features	**4 Features**	**6 Features**	**8 Features**
Skewness	Skewness	Skewness	Skewness
Onset variance	Onset variance	Onset variance	Onset variance
	Symmetry	Symmetry	Symmetry
	Onset Mean	Onset Mean	Onset Mean
		Spectral Rolloff Point	Spectral Rolloff Point
		Modulation at 4 Hz	Modulation at 4 Hz
			Percentile 50%
			Lower Half

Table 4.5: List of most influential features for severe HL

2 Features	**4 Features**	**6 Features**	**8 Features**
Percentile 50%	Percentile 50%	Percentile 50%	Percentile 50%
Onset variance	Onset variance	Onset variance	Onset variance
	Percentile 30%	Percentile 30%	Percentile 30%
	Onset Mean	Onset Mean	Onset Mean
		Modulation at 16 Hz	Modulation at 16 Hz
		Skewness	Skewness
			Modulation at 64 Hz
			Symmetry

Table 4.6: List of most influential features for profound HL

2 Features	**4 Features**	**6 Features**	**8 Features**
Symmetry	Symmetry	Symmetry	Symmetry
Onset variance	Onset variance	Onset variance	Onset variance
	Skewness	Skewness	Skewness
	Onset Mean	Onset Mean	Onset Mean
		Spectral Rolloff Point	Spectral Rolloff Point
		Temporal Onsets	Temporal Onsets
			Percentile 30%
			Percentile 90%

Table 4.7: List of unified most influential features to all three hearing losses

order to evaluate their capability to generalize. The models are trained to map the most influential features to the user's target volume setting. The performance measure of each model in terms of mapping accuracy is discussed in detail in section 4.6.

4.5.3 Post-Processing

The MLP and ANFIS models were trained with discrete volume steps of a 1 dB resolution. However, the output of the MLP and ANFIS models are linearly scaled and continuous. Thus, a post-processor is required to discretize the output to a 1 dB resolution, in order to be consistent with the resolution of the target volume settings. For example, an output of 4.8 dB will be post-processed to 5 dB. As a result, the post-processor specifies the final predicted volume settings to the hearing aid.

4.6 Performance Measures

Performances of the adopted models is measured by determining how accurate the predicted outputs are to the target ones. The performance measure is evaluated by the absolute value of the volume error (VE) and the speech intelligibility index error (SIIE), presented in the following equations:

$$VE(i) = |V_{opt}(i) - V_{pred}(i)| \tag{4.13}$$

$$SIIE(i) = SII_{opt}(i) - SII_{pred}(i) \tag{4.14}$$

$$mean_{|V|} = \frac{1}{m}\sum_{i=1}^{m} VE(i) \tag{4.15}$$

$$mean_{SII} = \frac{1}{m}\sum_{i=1}^{m} SIIE(i) \tag{4.16}$$

$$STD_{|V|} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(VE(i) - mean_{|V|})^2} \tag{4.17}$$

$$STD_{SII} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(SIIE(i) - mean_{SII})^2} \tag{4.18}$$

where m is the number of patterns and i is a pattern index. The subscripts *opt* and *pred* refer to the optimal and predicted volume. The *mean* indicates the average pattern error and the *STD* is the standard deviation which indicates how disperse the pattern errors are from the mean. Thus, a small *STD* implies that the pattern errors tend to be very close to the mean.

4.7 Summary

This chapter described the system's layout to perform automatic volume settings for a hearing aid user and the components involved in the system's implementation. Three simulated user profiles with a moderate, severe and profound hearing threshold are investigated in this research. The chapter also described eight energy related features that were added to a set of given features from SIEMENS. As a design requirement it is essential to perform feature selection and determine the most influential features to improve the mapping performance of the network. The sequential forward search algorithm was used to achieve this task and was used to select three sets of influential features, where each set corresponds to a certain simulated user (moderate, severe and profound). Furthermore, the SFS method was used to determine one set of unified features for all the three simulated users.

Two prediction models are undertaken to perform the mapping of the feature input space to the output space (optimal volume setting), which are the MLP and ANFIS. Finally, the performance measures to assess the system's performance are presented. The following chapter will present the results of the two proposed computational intelligence tools to perform the automatic volume settings framework.

Chapter 5

Automatic Volume Settings using Prediction Models

Two prediction models are investigated to perform the mapping of a feature input space to the output space (optimal volume setting): the multilayer perceptron (MLP), and the adaptive network-based fuzzy inference system (ANFIS). For both the MLP and ANFIS models preprocessing are performed on the network's inputs and targets. Before training, the inputs and targets are normalized to fall in the range [-1,1] and to simplify the training of the network. After training, the output of the networks are scaled to produce outputs in the range [-1,1]. Thus, the normalized network outputs are unnormalized to be converted back into the original target units. A MLP and ANFIS are analyzed in this chapter. The following sections contrast the performances of both models.

5.1 MLP Architecture

Four MLP models are implemented using the two, four, six and eight influential feature sets (presented in section 4.5.1) as inputs to the MLP models and all four MLP models are examined for each of the simulated user profiles (moderate HL, severe HL, profound HL) listed in table 4.2. Thus, a total of twelve MLP model performances are evaluated. During experimentation, the data set is partitioned into three groups, 1200 files (60% of the total files) for training, 400 files (20% of the total files) for validation and 400 (20% of the total files) for testing. The files in each subset are selected randomly, ensuring that the relative proportion of files of each category is preserved for each set. All the experiments are performed with cross validation, where each group (training, validation and testing) are randomly shuffled to ensure that the MLP model is independent on the order of the data set and is repeated three times. The MLP's architecture has an input layer, a hidden layer and an output layer. The number of neurons in the input layer is equal to the number of influential features (two, four, six or eight). For the output layer, only one neuron is used which represents the predicted volume setting and the activation function selected is a linear activation function in order to cover the whole range of the output space (approximately -60 to 60 dB).

The number of neurons in the hidden layer varies from model to model. Determining the number of neurons in the hidden layer is an architectural challenge for MLPs. There is no direct approach in determining the optimum number of neurons in the hidden layer. The number of neurons at the hidden layer was determined by an empirical study explained further in the following section.

5.1.1 System's Structure

Selecting the number of neurons in the hidden layer is based on the MLP's lowest volume error average over the training root mean square error (RMSE) and testing RMSE. The number of hidden neurons is varied until the lowest possible average volume error (dB) has been reached. Table 5.1 demonstrates the selection process for the MLP architecture with the six most influential features as inputs and for a user with moderate hearing loss profile. Note that the entries for the training RMSE in table 5.1 are an average of three trials and the last column of the table refers to the average of the training and testing RMSEs. To determine the optimal number of hidden nodes, the number of hidden nodes was varied until the minimum average error has been reached. It is evident in table 5.1 that the 24 hidden nodes provided the lowest average error. Thus, 24 hidden neurons are chosen for the MLP with the six most influential features set and moderate hearing loss user profile. The same process was repeated to determine the MLP architectures for the two, four and eight most influential features using the moderate hearing loss user profile. As well, the same process was used to obtain the MLP structures for the severe and profound hearing loss cases. Table 5.2 summarizes the network parameters (i.e, number of neurons in the input, hidden, output layers) determined for the MLP models for all three simulated user profiles.

Neurons in the hidden layer used tan-sigmoid activation functions, since differential activation functions must be used for the backpropagation learning algorithm as mentioned in chapter 3. The tan-sigmoid activation function is a common activation function

Number of Hidden Neurons	Training RMSE (dB)	Testing RMSE (dB)	Average RMSE (dB)
3	19.0	17.1	18.1
6	18.7	17.3	18.1
9	18.0	17.4	17.7
12	19.4	16.2	17.8
15	18.8	16.5	17.6
18	19.2	16.0	17.6
21	18.3	16.7	17.5
23	19.2	15.7	17.4
24	**18.9**	**15.8**	**17.3**
25	19.2	15.5	17.5
26	18.8	16.2	17.6
27	18.7	16.2	17.5
28	18.8	16.2	17.5

Table 5.1: Determining the number of hidden nodes for the 6-feature MLP model, moderate HL

Influential Features	Moderate HL	Severe HL	Profound HL
2	2 IN - 27 HN - 1 ON	2 IN - 24 HN - 1 ON	2 IN - 34 HN - 1 ON
4	4 IN - 27 HN - 1 ON	4 IN - 30 HN - 1 ON	4 IN - 23 HN - 1 ON
6	6 IN - 24 HN - 1 ON	6 IN - 24 HN - 1 ON	6 IN - 40 HN - 1 ON
8	8 IN - 21 HN - 1 ON	8 IN - 34 HN - 1 ON	8 IN - 35 HN - 1 ON

Table 5.2: Network parameters for the MLP models

used for backpropagation; however, other differential activation functions may be used. The output layer used the linear activation function in order to cover the whole output space, the full range of the desired volume settings is from -60 dB to 60 dB. These activation functions are used on all the MLP models for each user profile.

5.2 ANFIS Architecture

Four ANFIS models are also implemented using the two, four, six and eight influential feature sets as inputs to the ANFIS models and all four models are examined for each of the simulated user profiles (moderate HL, severe HL, profound HL) listed in table 4.2. Thus, a total of twelve ANFIS model performances are evaluated. Applying a similar set up as the MLP, all the experiments performed on ANFIS are done with cross validation. During experimentation, the data set is partitioned into three groups: 1200 files (60% of the total files) for training, 400 files (20% of the total files) for validation and 400 files (20% of the total files) for testing and each group are randomly shuffled to ensure that ANFIS is independent on the order of the data set. Each experiment is also repeated three times.

5.2.1 System's Structure

The number of inputs used for ANFIS is 2, 4, 6 and 8 depending on the number of influential features. A single output with a linear membership function is used, which represents the optimal volume setting.

Similar to the MLP on determining the number of hidden nodes, the number of input membership functions per input is selected based on the volume error average over the training RMSE and testing RMSE. The number of input membership functions is varied until the lowest possible average volume error (dB) has been reached.

Table 5.3 presents the selection process of the number of input membership functions per input for the 2 most influential features ANFIS model for the moderate HL user. The error decreases until the number of input membership function increases and reaches a limit of 5 input membership functions. This limit is due to the curse of dimensionality which is typically seen with the grid partitioning technique. The grid partitioning technique generates rules by enumerating all possible combinations of membership functions of all inputs [28]. Thus, the curse of dimensionality occurs when the number of fuzzy rules generated grows exponentially with the number of inputs.

It is shown in table 5.3 that 3 input membership functions provided the lowest average error. Thus, for the 2-feature ANFIS model three membership functions per input are used. The same procedure is repeated for the 4-feature ANFIS model.

To overcome the curse of dimensionality for cases where more than five inputs are required, the clustering or scattering partitioning methods can be used [38]. Moreover, the subtractive clustering method partitions the data into clusters and generates a minimum number of fuzzy rules required. For the ANFIS models with the 6 and 8 most influential features the subtractive clustering technique is used.

Since the subtractive clustering technique determines the optimum approach for dimension reduction of the fuzzy rules, the number of membership functions assigned to each input is determined by this technique. Table 5.4 shows the network parameters (input nodes, input membership functions, output node) for all twelve ANFIS models.

Number of MF/Input	Training RMSE (dB)	Testing RMSE (dB)	Average RMSE (dB)
2	19.0	16.4	17.8
3	**18.8**	**16.5**	**17.6**
4	18.7	17.3	18.0
5	18.6	17.8	18.2

Table 5.3: Determining the number of input MF per input for the 2-feature ANFIS model, moderate HL

Influential Features	Moderate HL	Severe HL	Profound HL
2	2 IN - 3 MF/input - 1 ON	2 IN - 5 MF/input - 1 ON	2 IN - 5 MF/input - 1 ON
4	4 IN - 3 MF/input - 1 ON	4 IN - 3 MF/input - 1 ON	4 IN - 3 MF/input - 1 ON
6	6 IN - 2 MF/input - 1 ON	6 IN - 2 MF/input - 1 ON	6 IN - 2 MF/input - 1 ON
8	8 IN - 3 MF/input - 1 ON	8 IN - 3 MF/input - 1 ON	8 IN - 3 MF/input - 1 ON

Table 5.4: Network parameters for the ANFIS models

After determining the architecture of both the MLP and ANFIS models, the next objective is to determine which influential feature set (2, 4, 6 or 8) is the optimal number of features as inputs to the network. This is achieved by evaluating the performances of all the networks and selecting the network which has the best performance. This is discussed further in the next section.

5.3 Simulation and Results

The experimental data described in section 4.2 are used to evaluate the MLP and ANFIS models to perform automatic volume settings. Training and testing results are presented below.

5.3.1 Training

Training begins once the input and output patterns from the training data set are provided. During training, the stopping criterion is controlled by the validation process as described in section 3.2.2. To analyze the performances of the MLP and ANFIS during training, the performance measures of volume error and SII error (section 4.6) were evaluated from the networks' outputs.

Volume Error

After three trials of shuffling the training set, the average of the network outputs are taken and the algebraic difference between the optimal and predicted volume for each training pattern are calculated and plotted shown in figure 5.1. As well the volume errors are ordered from lowest to highest error. Figure 5.1 presents the MLP's and ANFIS's training performances for the moderate HL user and using the 6 most influential features specific to this user profile as inputs (presented in table 4.4). The plots show a symmetrical distribution, indicating that both the MLP and ANFIS models are not bias towards overestimating or underestimating the optimal volume. From this data, the average volume error (equation (4.15)) and the standard deviation of the volume errors (equation (4.17)) yielded an error of 12.6±11.2 dB for the MLP and 10.2±9.7 dB for ANFIS. These error results seem relatively high at first glance; as seen in figure 5.1, there is a number of volume errors greater than 20 dB. However, a large volume error does not necessarily lead to a large intelligibility error as discussed in section 4.3.3. This concept is explained further in the following section.

SII Error

The SII error is another performance measure used to evaluate how well the trained MLP and ANFIS has optimized the speech intelligibility for a particular hearing loss. It is important to stress out that the goal of this research is to minimize the patient's manual fiddling with the hearing aid's knobs by automatically setting the hearing aid's

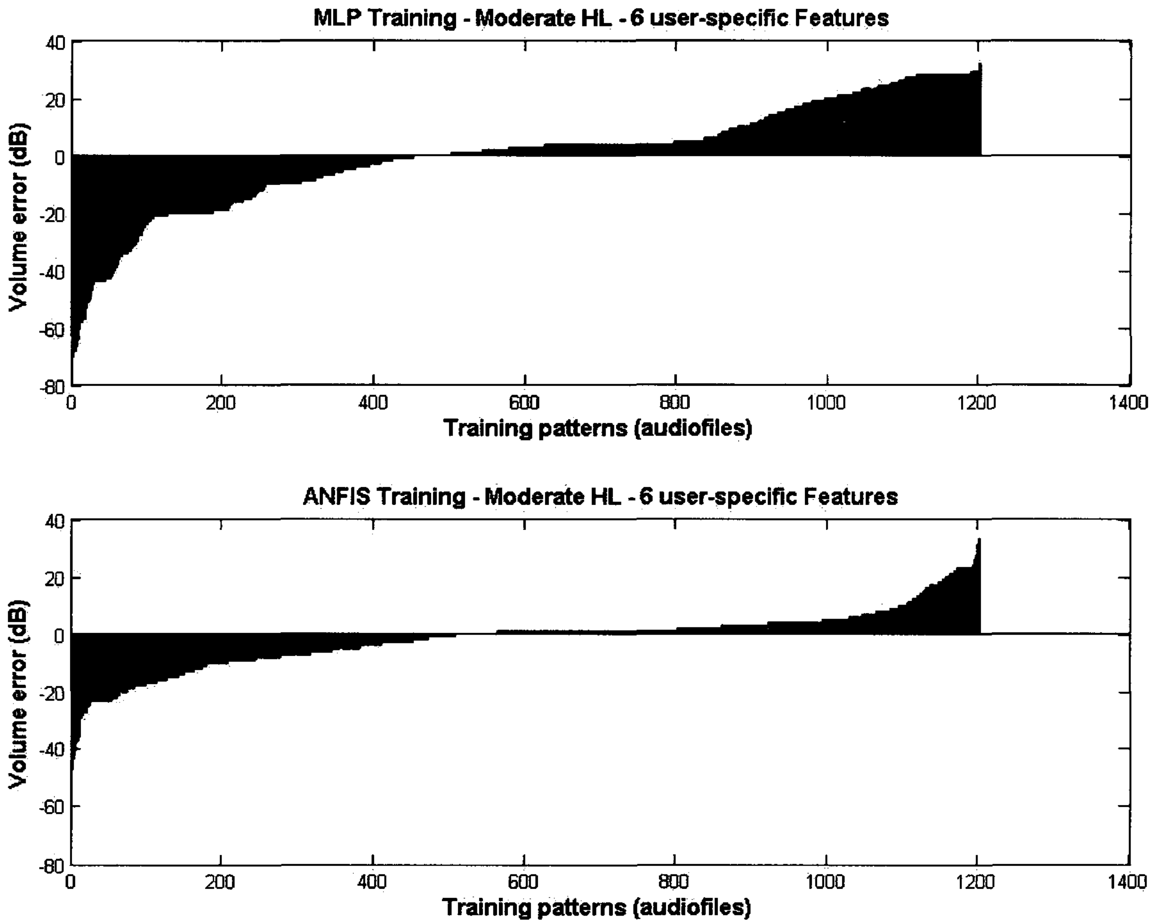


Figure 5.1: Volume error plots for MLP and ANFIS training performance using 6 user-specific features for moderate HL user

volume to an estimated desired setting, as to maximize the SII value. Since the SII is directly correlated to the volume (as shown in figures 4.3, 4.4 and 4.5) the goal can be achieved by selecting a value setting that maximizes the SII. In other words, the ultimate goal is to optimize the SII value; where volume tuning is only a means to achieve this goal. It is worth mentioning that the volume to SII mapping is a many-to-one mapping. In other words, optimizing the SII does not have a unique volume solution (as depicted in figures 4.3, 4.4 and 4.5). Therefore, the volume errors are far less important here than the SII errors.

Figure 5.2 present the SII errors obtained from the training patterns of MLP model and ANFIS model for the moderate HL user using 6 influential features as inputs. The SII errors are ordered from lowest to highest error. Figure 5.2 demonstrates that the great majority of the audio files resulted in a low SII error (equation (4.14)). The MLP yielded a $mean_{SII}$ and STD_{SII} of 0.004$\pm$0.02 while ANFIS yielded an error of 0.003$\pm$0.002. It is clear that the ANFIS model outperformed the MLP model.

The percentage of the training patterns leading to a low SII error is more evident in the cumulative distribution curve presented in figure 5.3. Figure 5.3 demonstrates that, for the MLP, 88% of the training patterns obtained an error of less than 0.01 and 95% of the patterns obtained an error of less than 0.03. The low SII error demonstrates the MLP ability in optimizing the SII and provides confidence for the patterns which have a high volume error. Only 4% of the patterns obtained an SII error higher than 0.08. For ANFIS, the model led to an SII error of less than 0.003 for 90% of the training patterns and an SII error of less than 0.004 for 95% of the training patterns. The percentage of the patterns with an SII error larger than 0.007 is negligible. Again, the performance of ANFIS is advantageous compared to the MLP model for a moderate HL user, demonstrating ANFIS greater accuracy in optimizing the SII for a moderate HL user compared to the MLP.

The training performance for the severe hearing loss user is relatively similar to the performance of the moderate hearing loss user. This is due to the wide saturation plateau of the speech intelligibility curves for the severe hearing loss, which is depicted in figure 4.4. The SII error yielded from the training patterns for the MLP and ANFIS are plotted in figure 5.4, as well the SII errors are ordered from lowest to highest error. For the MLP, the $mean_{SII}$ as well as the STD_{SII} led to be 0.005$\pm$0.01, which is relatively close to the moderate hearing loss user. As well, in figure 5.4, ANFIS yielded a SII error with a $mean_{SII}\pm STD_{SII}$ of 0.002$\pm$0.003. This demonstrates MLP and ANFIS effectiveness in optimizing the speech intelligibility for the severe hearing loss user.

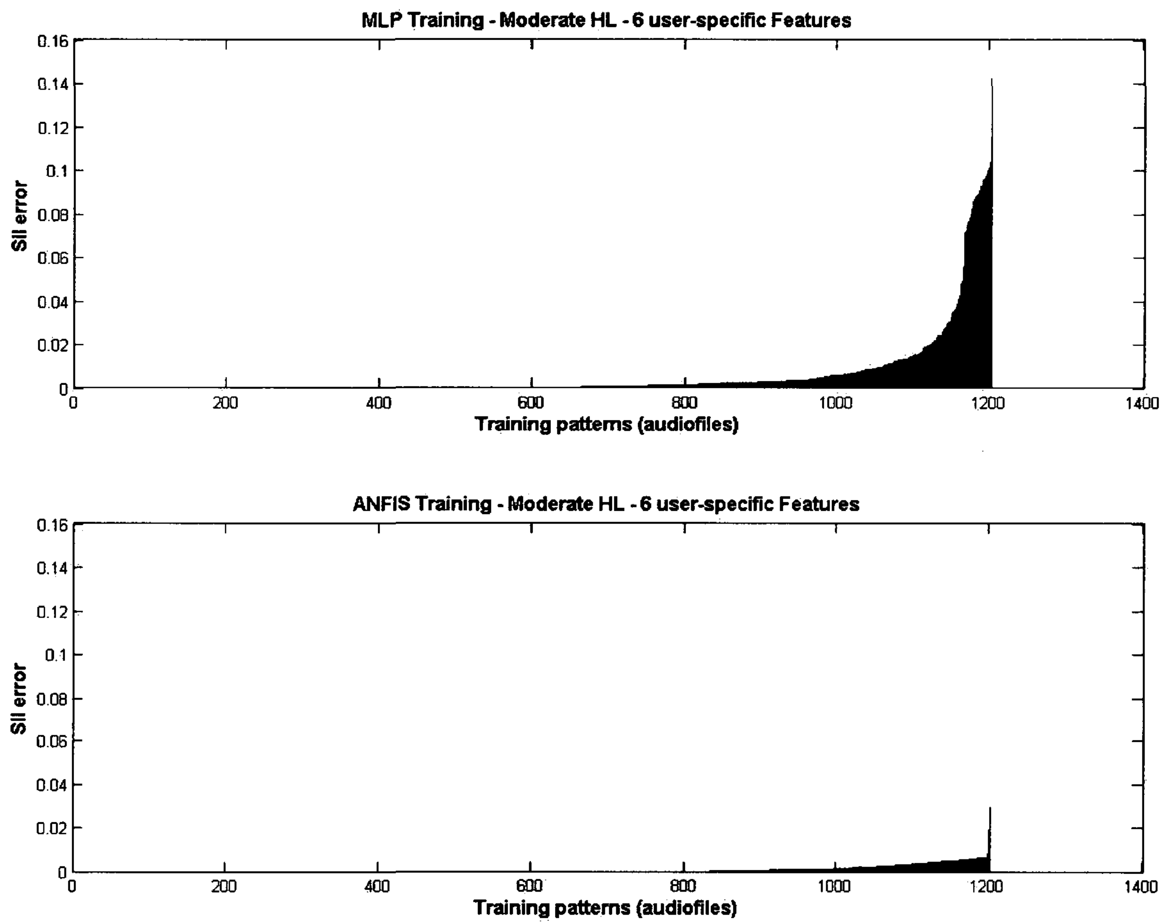


Figure 5.2: SII error plots for MLP and ANFIS training performance using 6 user-specific features for moderate HL user

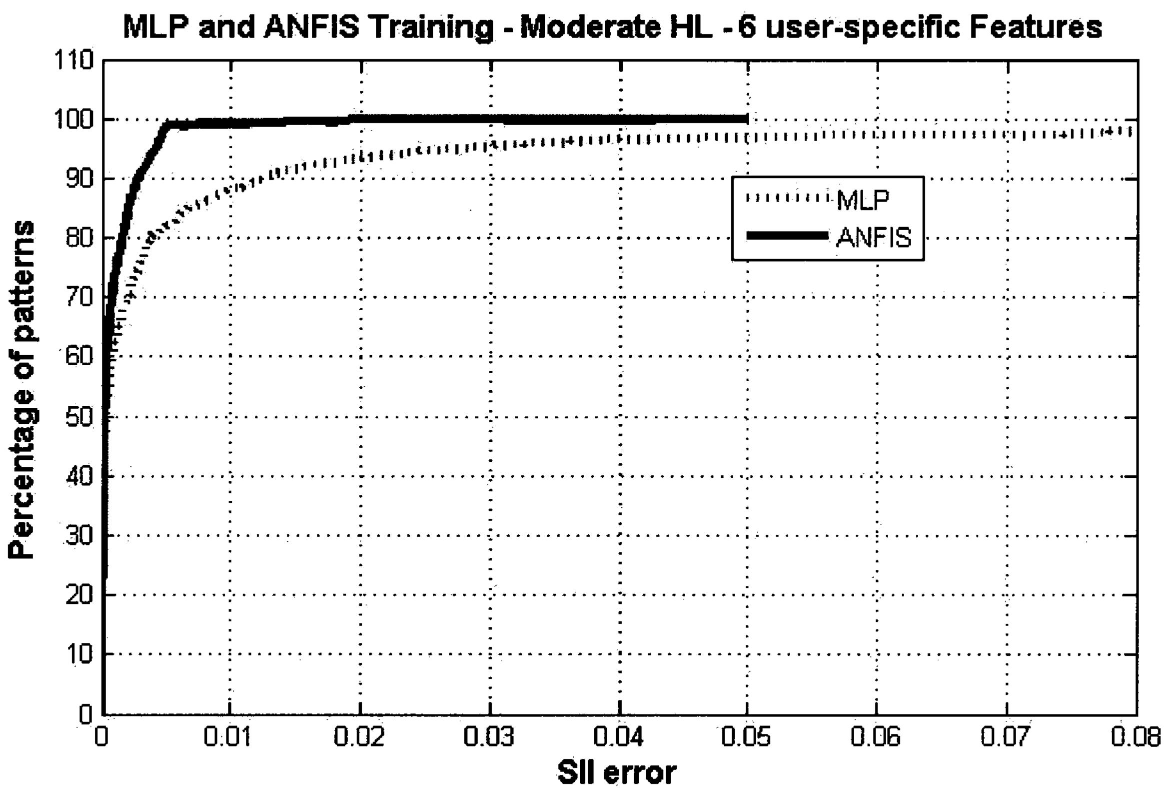
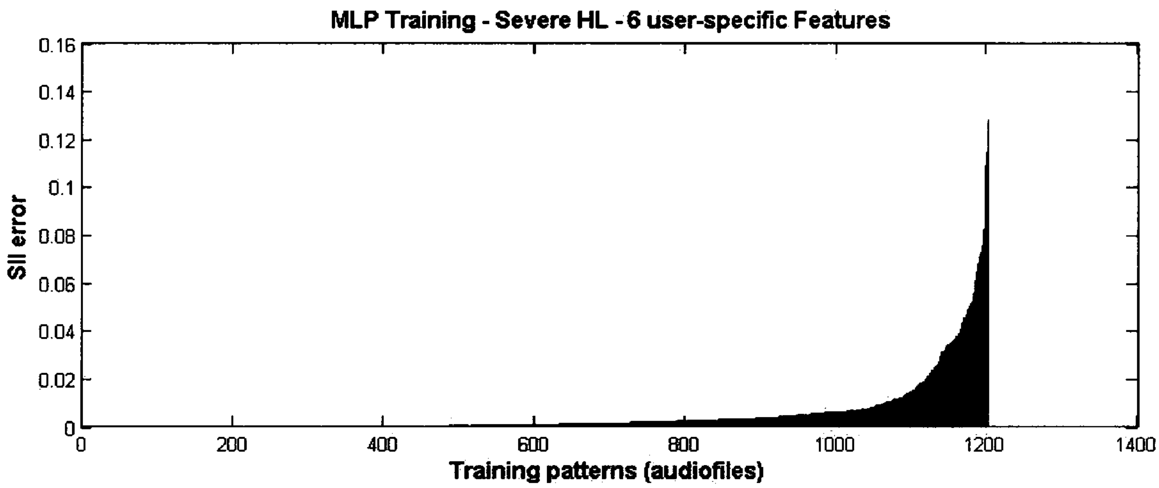


Figure 5.3: SII error plot by cumulative distribution for MLP and ANFIS training performances using 6 user-specific features for moderate HL user

This is more evident in figure 5.5, which shows that for the MLP 90% of the training patterns yielded an SII error of less than 0.01 and 95% of the training patterns yielded an SII error of less than 0.03. However, figure 5.5 also presents that the ANFIS model yielded a much lower error in SII for the severe user compared to the MLP, where 90% of the training patterns led to a SII error of less than 0.003 and 95% of the training patterns led to a SII error of less than 0.004. This result demonstrates ANFIS improvement in optimizing the SII for the severe user compared to MLP. Again, the number of patterns yielding an SII error greater than 0.005 is negligible.



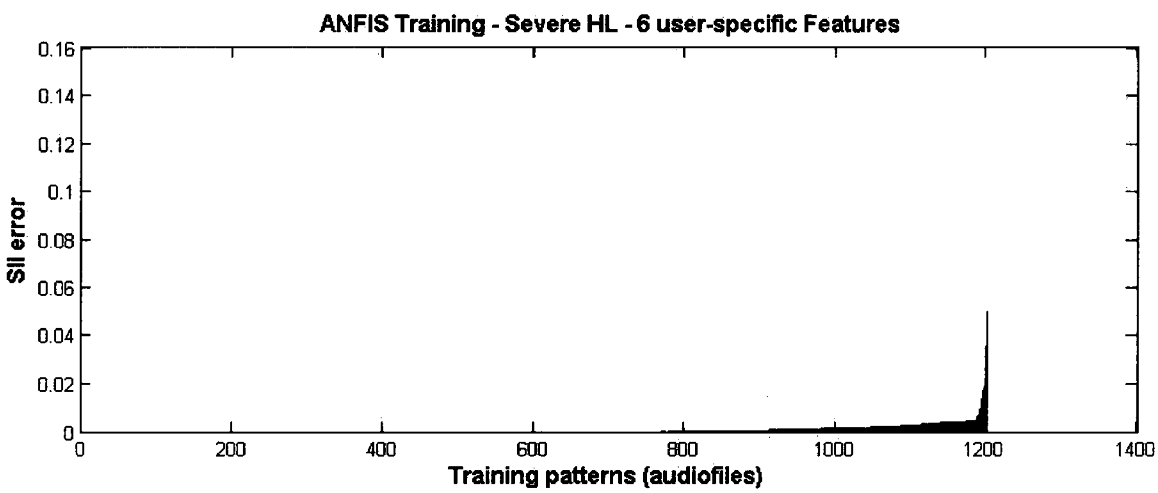


Figure 5.4: SII error plots for MLP and ANFIS training performance using 6 user-specific features for severe HL user

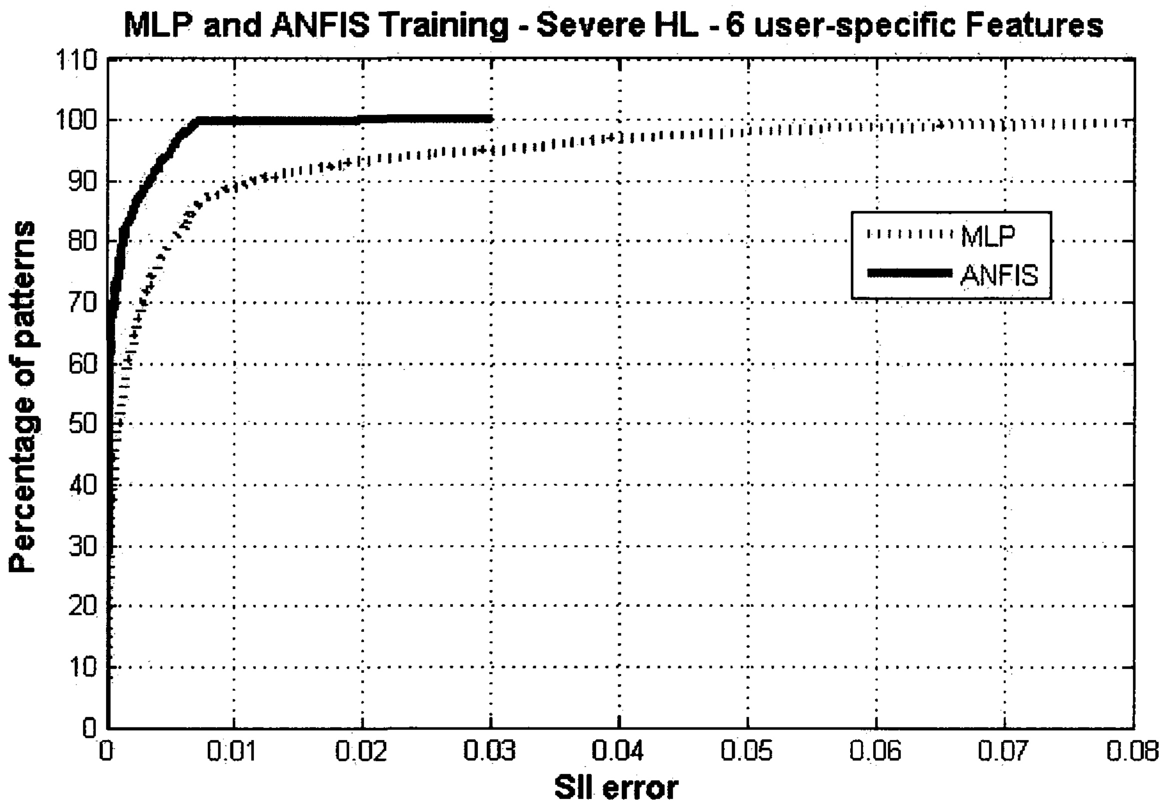


Figure 5.5: SII error plot by cumulative distribution for MLP and ANFIS training performances using 6 user-specific features for severe HL user

As for the profound hearing loss user, it is expected to yield a larger SII error due to its narrower saturation plateau presented in the speech intelligibility curves in figure 4.5. For the MLP model, figure 5.6 demonstrates that several training patterns obtained SII errors higher than 0.2. The $mean_{SII}$ as well as the STD_{SII} led to 0.01±0.06, which are only slightly higher compared to the moderate and severe users. In such a case where both the SII error and the volume error are high, the user can manually override the automatically set volume. As for the ANFIS model, figure 5.6 presents the reduction in SII error compared to the MLP and yielding a $mean_{SII} \pm STD_{SII}$ of 0.005±0.02. The reduction in SII error is more apparent in figure 5.7. For the MLP, 90% of the training patterns led to an error of less than 0.01 and 95% of the training patterns led to an error of less than 0.03. The ANFIS model achieved a SII error of less than 0.004 for 90% of its training patterns and a SII error of less than 0.006 for 95% of the training patterns.

Tables 5.5 and 5.6 present the training performance measures ($mean_{|V|} \pm STD_{|V|}$ in

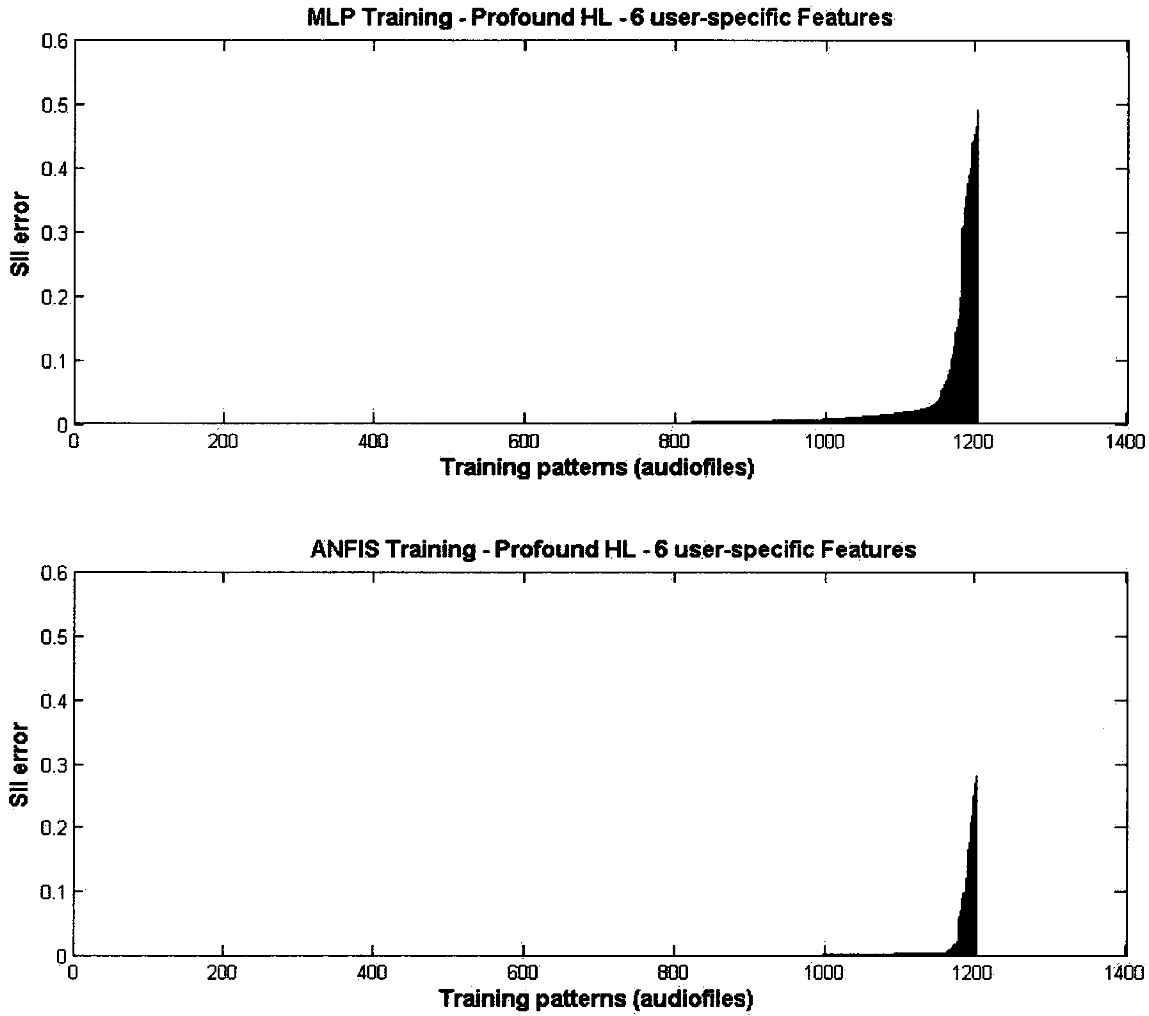


Figure 5.6: SII error plots for MLP and ANFIS training performance using 6 user-specific features for profound HL user

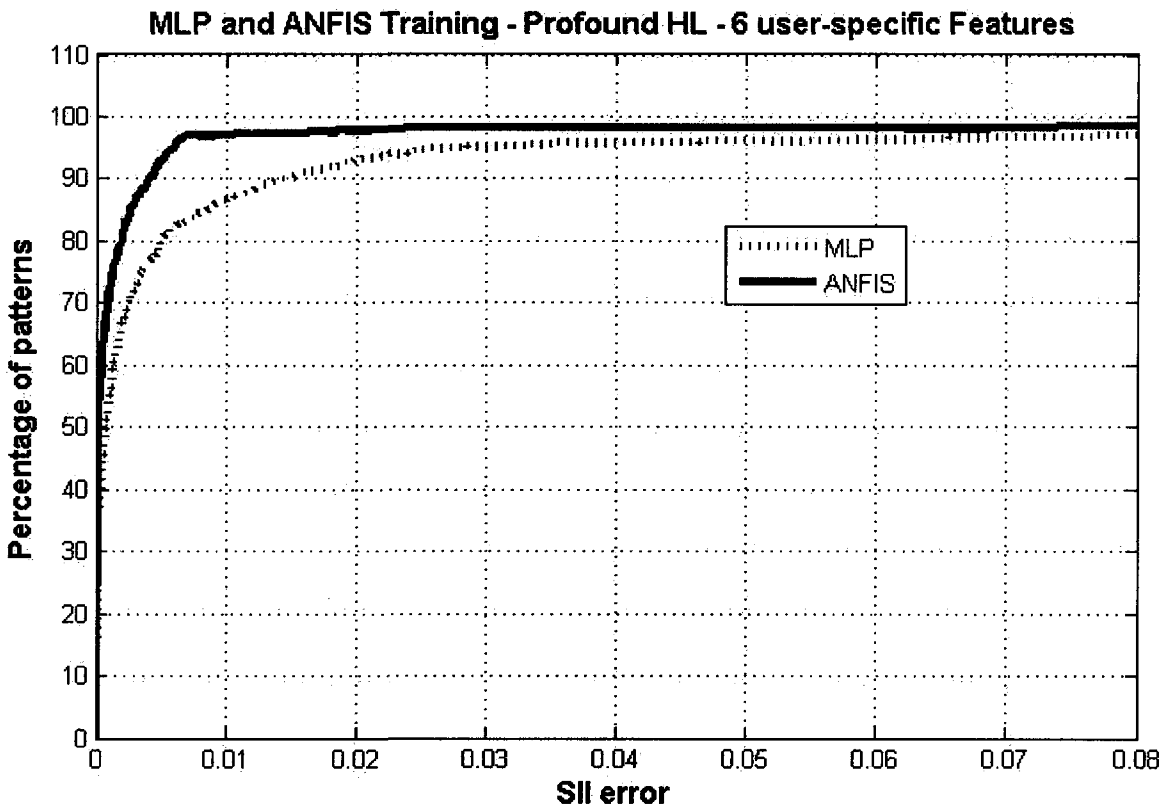


Figure 5.7: SII error plot by cumulative distribution for MLP and ANFIS training performances using 6 user-specific features for profound HL user

dB and $mean_{SII} \pm STD_{SII}$) for the 2, 4, 6 and 8 most influential feature MLP models for all three simulated users (moderate HL, severe HL and profound HL). In table 5.5 it is shown that, for training, the 8-feature MLP model generally yielded the lowest $mean_{|V|}$ for all user profiles. This may be due to the fact that more features are being used which are highly correlated to the output volume settings. In table 5.6, the majority of the time the 6 and 8 influential features obtained the lowest SII error and options are very comparable. Thus, this is an indication that the 6 and 8 features are optimal features for the ANFIS and MLP models. However, due to computational complexity, it is best to select the 6 influential features. Therefore, the 6 influential feature set was determined to be the optimal number of feature inputs to both the MLP and ANFIS.

	Moderate User		**Severe User**		**Profound User**	
	MLP	**ANFIS**	**MLP**	**ANFIS**	**MLP**	**ANFIS**
2 Features	12.9±11.8	12.8±12.7	13.1±11.5	13.0±11.4	13.2±12.0	12.4±9.9
4 Features	12.9±11.9	11.8±9.5	12.8±11.4	10.1±10.1	12.6±9.9	11.1±9.8
6 Features	12.6±11.2	10.2±9.7	12.6±11.4	10.0±9.0	12.4±9.9	9.9±8.9
8 Features	12.4±11.3	10.9±9.2	12.1±10.9	9.9±9.1	11.1±9.5	9.8±9.9

Table 5.5: Volume error ($mean_{|V|} \pm STD_{|V|}$ (dB)) performance measures for MLP and ANFIS, trained on user-specific features

	Moderate User		**Severe User**		**Profound User**	
	MLP	**ANFIS**	**MLP**	**ANFIS**	**MLP**	**ANFIS**
2 Features	0.006±0.02	0.006±0.02	0.006±0.02	0.006±0.01	0.02±0.1	0.007±0.03
4 Features	0.006±0.02	0.006±0.02	0.006±0.02	0.005±0.002	0.01±0.05	0.01±0.02
6 Features	0.004±0.02	0.003±0.002	0.005±0.01	0.002±0.003	0.01±0.06	0.005±0.02
8 Features	0.004±0.02	0.002±0.003	0.005±0.02	0.002±0.002	0.01±0.05	0.003±0.02

Table 5.6: SII error ($mean_{SII} \pm STD_{SII}$) performance measures for MLP and ANFIS, trained on user-specific features

5.3.2 Testing

After the network has been trained. The testing process is used to validate the network's performance to generalize with unfamiliar data. The testing performance of the MLP and ANFIS was evaluated by also using the performance measures volume error and speech intelligibility error, (section 4.6).

Volume Error

The volume error for each testing pattern is calculated by taking the algebraic difference between the optimal and predicted volumes and plotted, shown in figure 5.8. As well, the volume errors are ordered from lowest to highest error. Figure 5.8 presents the testing patterns' volume error for the moderate hearing loss user with the 6-feature MLP model and the 6-feature ANFIS model. In figure 5.8, the MLP yielded a volume error $mean_{|V|} \pm STD_{|V|}$ of 11.0±9.0 dB. Compared to the training performance of 12.6±11.2 dB, the testing performance is slightly better for the MLP. As well, in figure 5.8, the ANFIS model volume error is 9.0±9.0 dB during testing, a slightly lower value than the 10.2±9.7 dB error during training. The decrease in testing error confirms that the connectionist models could capture the system's nonlinear input-output relationships during training without overfitting.

In figure 5.8, there are audio files with volume errors as high as 20 dB; however, this may have little effect on speech intelligibility if those particular audio files have a small SII error, which is showed further in the next section.

SII Error

The SII error was also used to evaluate the testing performance of the MLP and ANFIS. The SII error is calculated using equation (4.14). It is desirable to obtain a low SII error for the majority of the audio files irrespective of the volume error. Figure 5.9 presents the SII error plot for moderate HL user with the 6-feature MLP model, where the SII errors are ordered from lowest to highest error. The $mean_{SII} \pm STD_{SII}$ error is 0.006±0.02 and the majority of the testing patterns achieved a very low SII error. This is vivid in figure 5.10 where 90% of the testing patterns obtained an SII error of less than 0.01 and 95% of the testing patterns obtained an SII error of less than 0.02.

As well, figure 5.9 presents the SII error obtained during testing for the moderate HL user with the 6-feature ANFIS model. It can be seen that the error is smaller than MLP

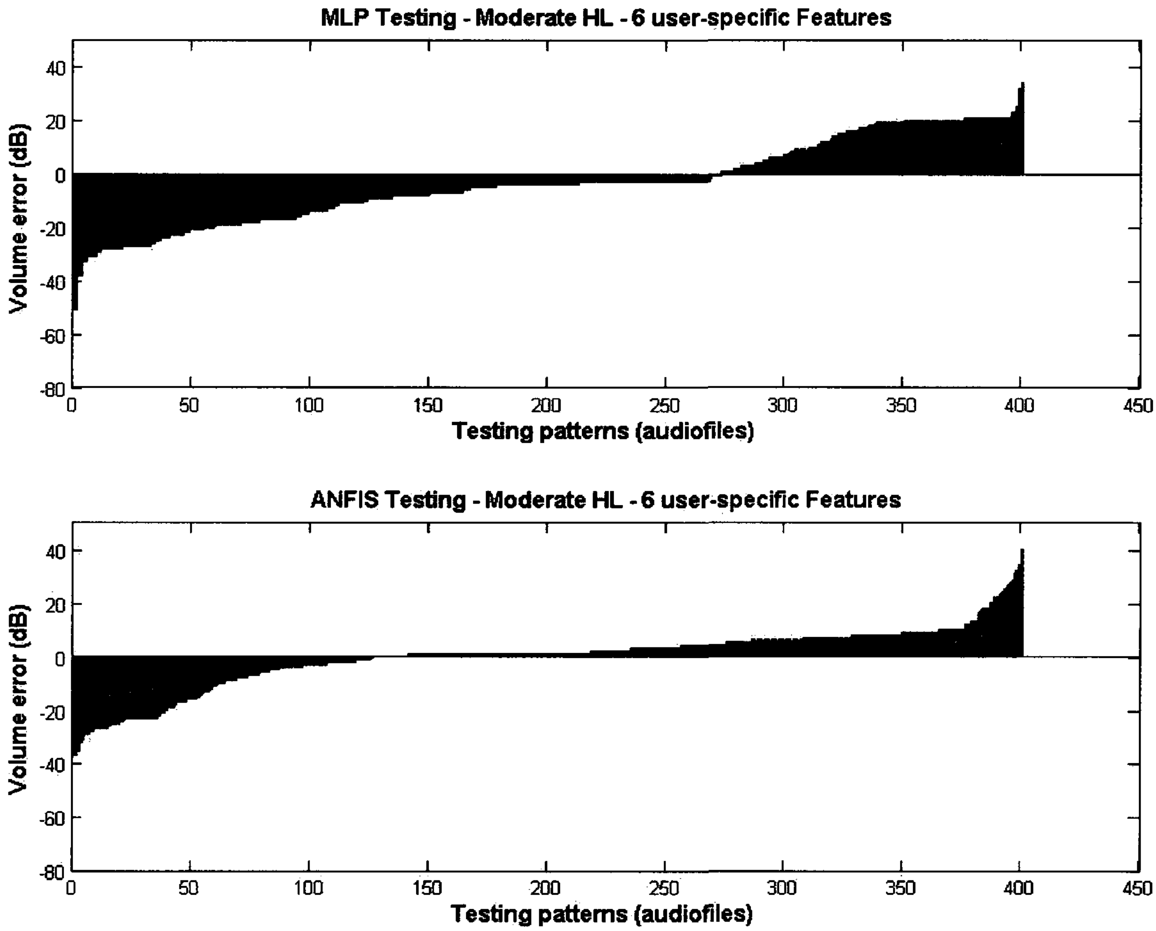


Figure 5.8: Volume error plot for MLP and ANFIS testing performance using 6 user-specific features for moderate HL user

in figure 5.9. The performance error measure, $mean_{SII} \pm STD_{SII}$, is determined to be 0.004±0.002, which is slightly smaller than MLP testing performance. Figure 5.10 shows ANFIS performance where 90% of the testing patterns had a SII error of less than 0.003 and 95% of the testing patterns had a SII error of less than 0.005. The low value in SII error demonstrates ANFIS ability to generalize better with unfamiliar data compared to the MLP.

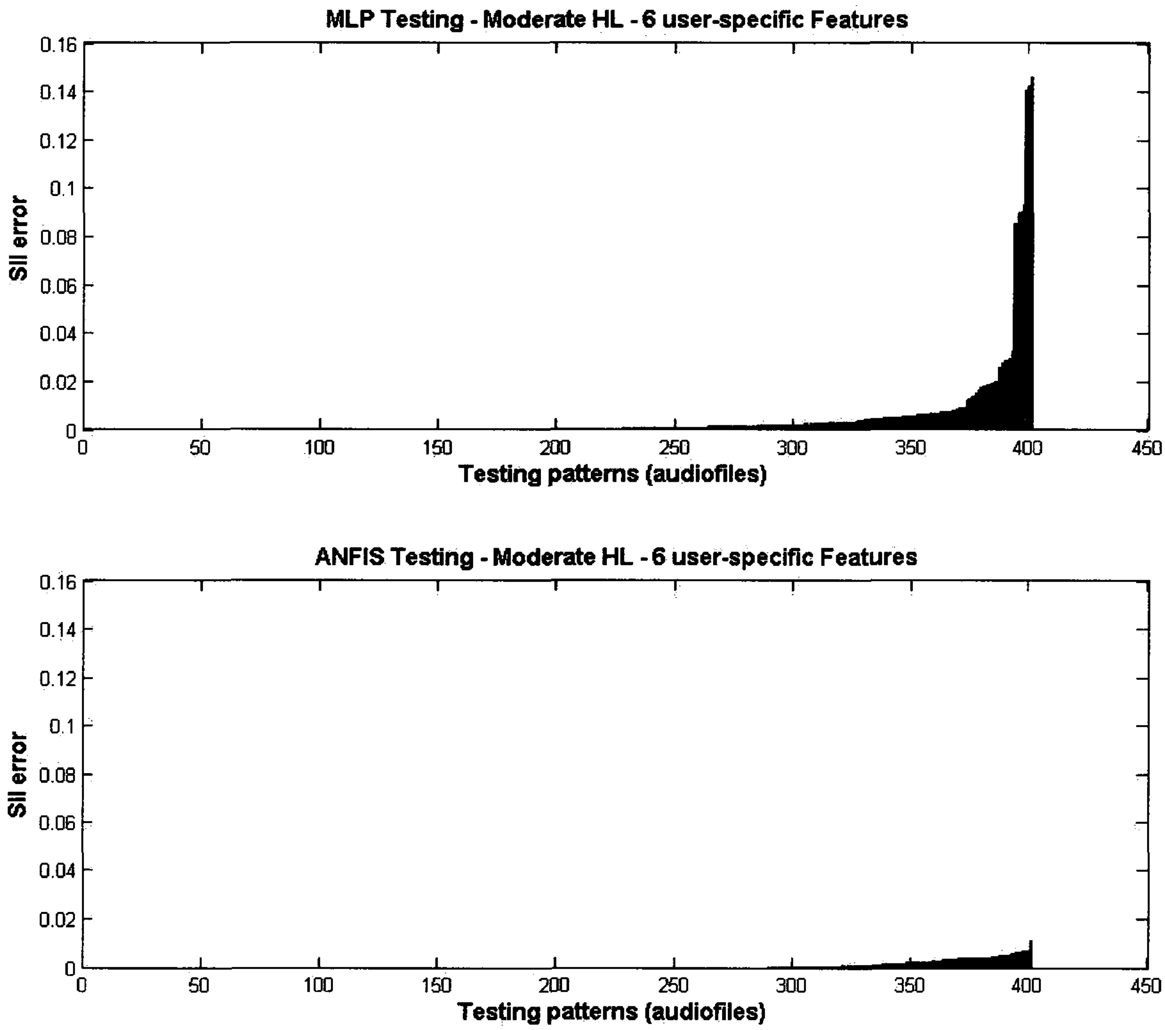


Figure 5.9: SII error plots for MLP and ANFIS testing performance using 6 user-specific features for moderate HL user

For the severe hearing loss user, the SII errors when testing the 6-feature MLP and ANFIS models are present in figure 5.11. For the MLP, the $mean_{SII} \pm STD_{SII}$ led to an

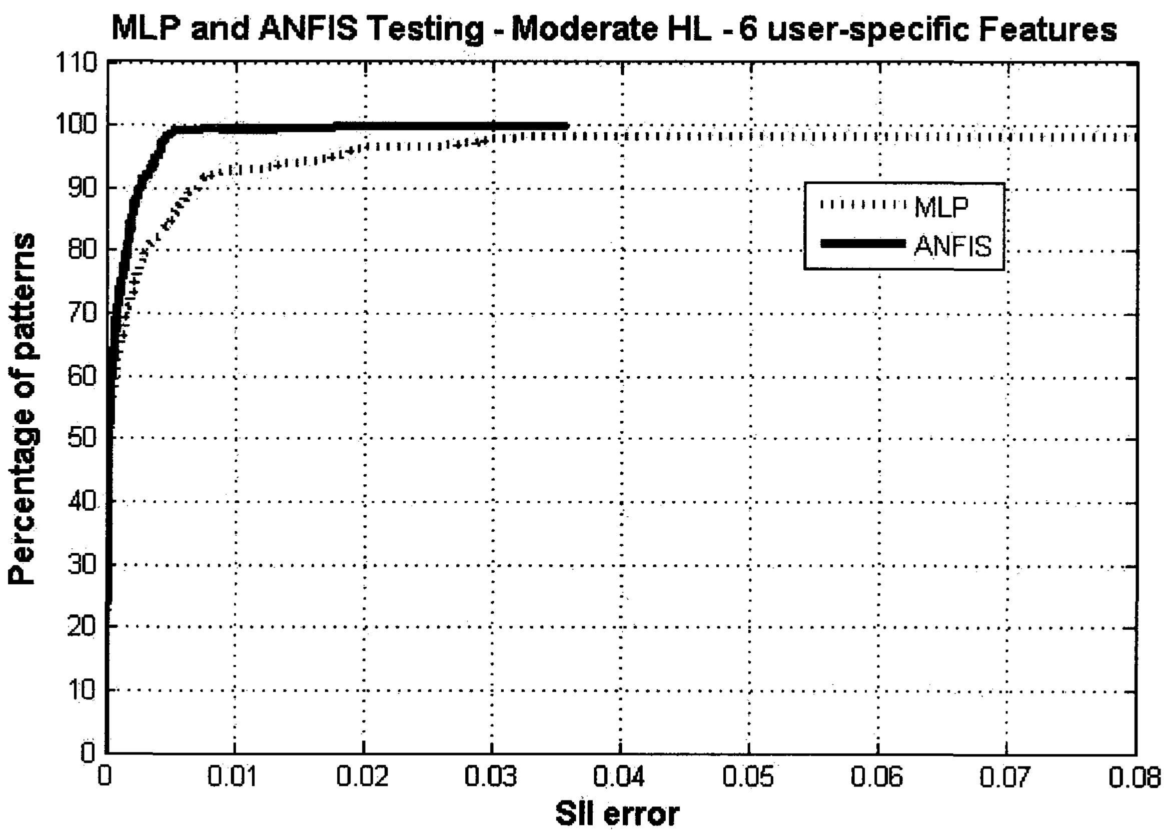


Figure 5.10: SII error plot by cumulative distribution for MLP and ANFIS testing performance using 6 user-specific features for moderate HL user

error of 0.006±0.01. ANFIS yielded a $mean_{SII} \pm STD_{SII}$ of 0.003±0.002, a very small error similar to the one obtained for the moderate hearing loss. Figure 5.12 compares the MLP and ANFIS testing performances for the severe hearing loss user. For the MLP, 90% of the testing patterns yielded an error of less than 0.02 and 95% of the testing patterns yielded an error of less than 0.03. The low error in SII demonstrates the MLP ability to optimize the SII for a severe hearing loss user. For ANFIS, 90% of the testing patterns obtain a SII error of less than or equal to 0.003 and 95% of patterns obtained a SII error of less than or equal to 0.005, a superior performance compared to MLP.

For the profound hearing loss user, the SII errors obtained when testing the 6-user specific features MLP and ANFIS are present in figure 5.13. The MLP yielded a $mean_{SII} \pm STD_{SII}$ of 0.009±0.03 and ANFIS led to a $mean_{SII} \pm STD_{SII}$ error of 0.007±0.03. These results are slightly higher compared to the moderate and severe hearing loss cases. However, this is expected due to the narrow saturation plateau of the speech intelligibility curves for the profound hearing loss user (figure 4.5). Nevertheless, figure 5.14 demonstrates that the majority of the patterns achieved a very low SII. For the MLP, 90% of the testing patterns obtained an error of less than 0.02 and 95% of the testing patterns received an error of less than 0.06. As well, for ANFIS 90% of the testing patterns led to an SII error of less than 0.005 and 95% of the testing patterns yielded an SII error of less than 0.03. The low error in SII demonstrates the ANFIS and MLP ability to optimize the SII even for a profound hearing loss user.

Tables 5.7 and 5.8 present the testing performance measures ($mean_{|V|} \pm STD_{|V|}$ in dB and $mean_{SII} \pm STD_{SII}$) for the 2, 4, 6 and 8-feature MLP and ANFIS models for all three simulated users (moderate HL, severe HL and profound HL). In tables 5.7 and 5.8, for the majority of the cases the 6 and 8 features obtained the lowest volume error and SII error for both the MLP and ANFIS models, thus confirming the observation that the 6 and 8 most influential features are a good choice for the number of feature inputs to the models. However, again due to computational constraints, 6 features are favored. Thus, 6 features are selected as the optimal inputs for the MLP and ANFIS models.

As well, from tables 5.7 and 5.8 it is evident that the ANFIS model is advantageous compared to the MLP model in terms of volume and SII errors for all three simulated user profiles.

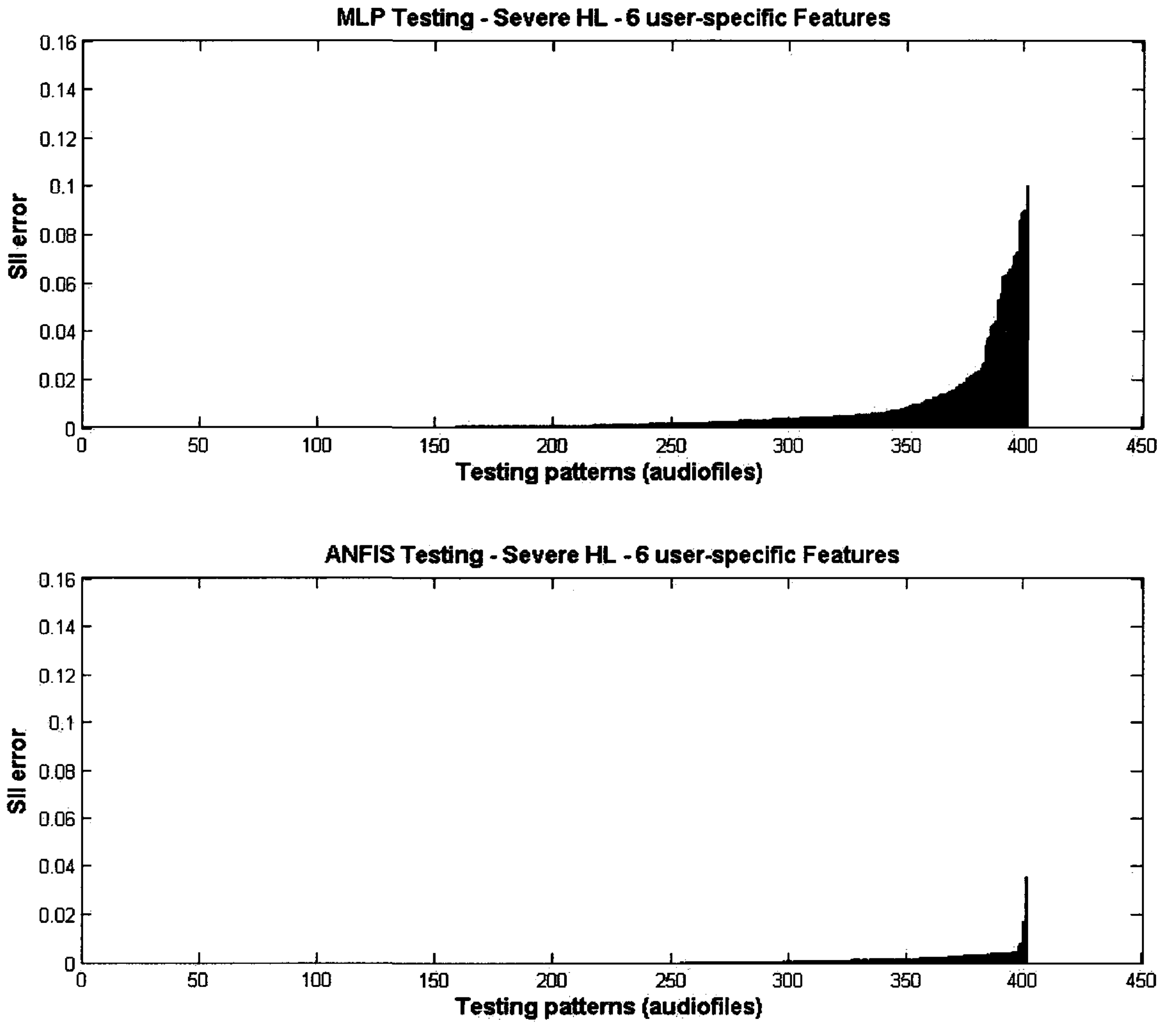


Figure 5.11: SII error plots for MLP and ANFIS testing performances using 6 user-specific features for severe HL user

	Moderate HL		**Severe HL**		**Profound HL**	
	MLP	**ANFIS**	**MLP**	**ANFIS**	**MLP**	**ANFIS**
2 Features	11.9±9.2	11.8±10.4	13.1±10.1	11.8±10.0	12.4±10.0	12.2±9.3
4 Features	11.9±9.1	10.6±9.4	12.8±9.8	11.4±11.0	13.4±9.4	11.1±9.8
6 Features	11.0±9.0	9.0±9.0	12.1±9.8	9.7±9.2	12.4±9.9	9.0±9.5
8 Features	11.0±9.5	9.2±9.1	12.4±9.9	9.5±9.0	12.4±9.6	9.1±9.0

Table 5.7: Volume error ($mean_{|V|} \pm STD_{|V|}$ (dB)) performance measures for MLP and ANFIS, tested on user-specific features

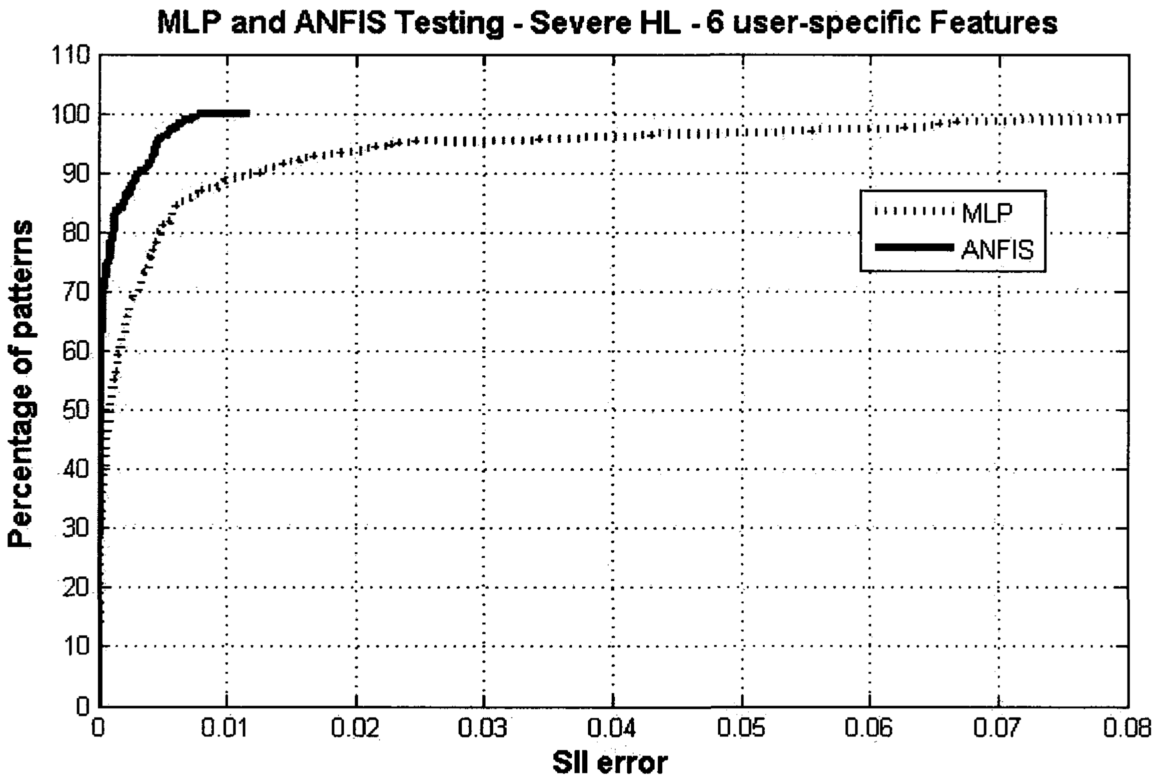
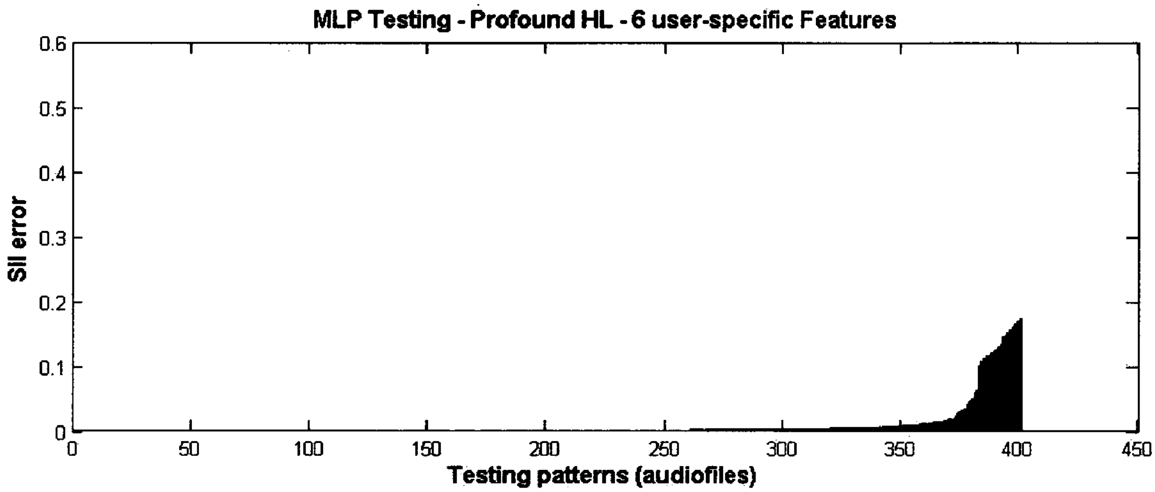


Figure 5.12: SII error plot by cumulative distribution for MLP and ANFIS testing performance using 6 user-specific features for severe HL user

	Moderate HL		**Severe HL**		**Profound HL**	
	MLP	**ANFIS**	**MLP**	**ANFIS**	**MLP**	**ANFIS**
2 Features	0.009±0.04	0.005±0.02	0.007±0.01	0.007±0.02	0.02±0.1	0.01±0.05
4 Features	0.009±0.04	0.005±0.02	0.007±0.02	0.007±0.002	0.02±0.1	0.01±0.02
6 Features	0.006±0.02	0.004±0.002	0.006±0.01	0.003±0.002	0.009±0.03	0.007±0.03
8 Features	0.006±0.02	0.003±0.001	0.006±0.02	0.003±0.001	0.008±0.07	0.005±0.03

Table 5.8: SII error ($mean_{SII} \pm STD_{SII}$) performance measures for MLP and ANFIS, tested on user-specific features



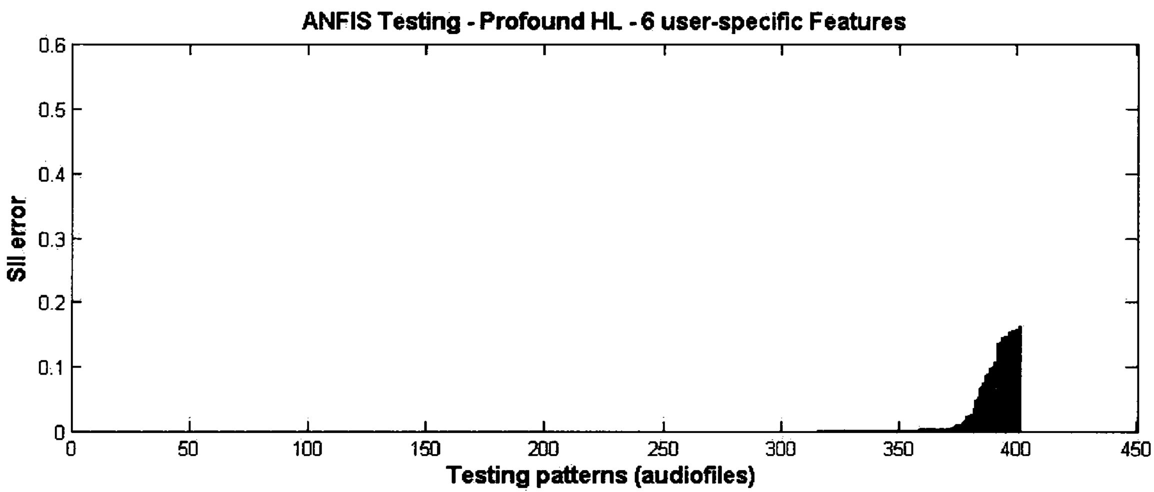


Figure 5.13: SII error plots for MLP and ANFIS testing performances using 6 user-specific features for profound HL user

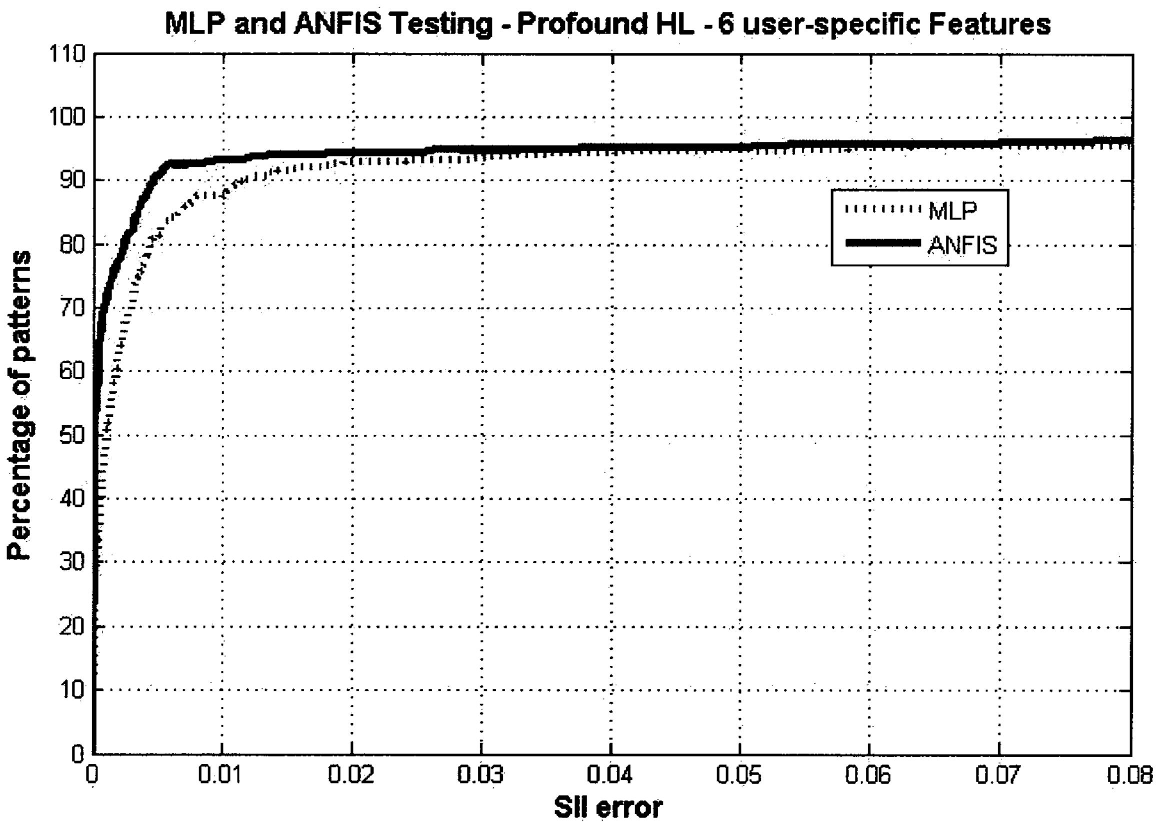


Figure 5.14: SII error plot by cumulative distribution for MLP and ANFIS testing performances using 6 user-specific features for profound HL user

5.4 Unified Influential Features

From the point of a view of a hearing aid manufacturer, it is typically desirable to reduce hardware complexity of the hearing aid. This section explores the possibility of maintaining an exceptional performance with the MLP and ANFIS, while using a generic set of features for various types of hearing losses than a user-specific set of features. This will simplify the design of the hearing aid since each user will be fitted with the same set of features.

Four sets of the most influential features are determined during feature selection and presented in tables 4.4, 4.5, 4.6 and 4.7. Three sets are specific to a certain simulated user (moderate, severe and profound hearing losses) and are used in section 5.3.1 and 5.3.2. The fourth set, is the unified set of influential features for all three hearing losses.

It is interesting to compare the performances of the MLP and ANFIS models using influential features specific to the type of simulated user (tables 4.4-4.6) versus a unified set of influential features for all three types of hearing losses (table 4.7).

Training and testing is performed as before in sections 5.3.1 and 5.3.2. The 6 unified features are inputs to the MLP and ANFIS models and their performances are analyzed for the three simulated user profiles over the 2000 audio recordings.

Table 5.9 presents the testing performance of the MLP comparing the SII error when using the six user-specific influential features or the six unified influential features as inputs. From table 5.9, it is shown that the error has slightly increased using the unified set of features as inputs compared to using user-specific features. However, the error still remains relatively low, typically a mean SII error of 0.01 is obtained for all three hearing losses.

Table 5.10 presents the testing performance for ANFIS and it also compares the SII error when using the six user-specific features or the six unified influential features as inputs. Table 5.10 shows a slight increase in error when using the set of unified features as inputs. However, the mean SII is 0.007 or lower for all three hearing losses.

From tables 5.9 and 5.10, both ANFIS and the MLP performs better when using influential features specific to each user at the expense of a more complex design process for the hearing aid where up to 24 different features need to be coded. Using one set of unified features lead to a slight increase in error; however, the error still remains relatively low and the design of the hearing aid is simplified by having a reduced set of features to implement.

MLP Model	Moderate HL		Severe HL		Profound HL	
	User-Specific Features	**Unified Features**	**User-Specific Features**	**Unified Features**	**User-Specific Features**	**Unified Features**
SII Error	0.006±0.02	0.01±0.02	0.006±0.01	0.01±0.02	0.009±0.03	0.01±0.06
SII Error: 95% testing patterns	<0.02	<0.07	<0.03	<0.07	<0.06	<0.05
SII Error: 90% testing patterns	<0.01	<0.03	<0.02	<0.04	<0.02	<0.03

Table 5.9: SII error ($mean_{SII} \pm STD_{SII}$) performance measures for MLP, tested on user-specific and unified features

ANFIS Model	Moderate HL		Severe HL		Profound HL	
	User-Specific Features	**Unified Features**	**User-Specific Features**	**Unified Features**	**User-Specific Features**	**Unified Features**
SII Error	0.004±0.002	0.006±0.02	0.003±0.002	0.005±0.01	0.007±0.03	0.01±0.02
SII Error: 95% testing patterns	<0.005	< 0.07	≤0.005	<0.03	<0.03	<0.06
SII Error: 90% testing patterns	<0.003	< 0.03	≤0.003	<0.01	<0.005	<0.02

Table 5.10: SII error ($mean_{SII} \pm STD_{SII}$) performance measures for ANFIS, tested on user-specific and unified features

5.5 Testing the Proposed System in Varying Background Environments

This section tests the trained ANFIS and MLP models discussed in section 5.3.1 in varying background environments such as cafeteria noise, traffic noise, etc. The test involves new user profiles which were not used to train the connectionist models. This is important to observe how these models would generalize to different users and noises.

The trained MLP and ANFIS models are tested with a 14.45 minute sound recording, which was generated using a small Linear PCM Olympus recorder. The Olympus digital recorder was carried and walked around into different environments. The recording began outdoors at the University of Ottawa campus and then moved into a pharmacy store. The recorder was then carried into a coffee shop, to record crowd noise. Next, traffic noise was recorded at a bus station and finally a quieter environment was sampled in a laboratory at the University of Ottawa. Thus, the sound recording was generated with varying background environments. Once the sound recording was completed, the sound file was scaled to ensure a SPL between 40 to 85 dB SPL. After scaling the sound file, it was mixed with clean speech files provided by SIEMENS, different than the same clean speech files used in table 4.1. The clean speech files also consist of female and male speech files. After mixing the environment sound with the clean speech files, the noisy audio file is segmented for every 5 seconds. Thus, the SII is measured every 5 seconds. The new user profiles simulated to test the ANFIS and MLP models are presented in table 5.11, the table shows the frequencies at which the hearing loss measurement was taken.

Hearing Loss [dB HL]	**Frequency [kHz]**							**UCL [dB SPL]**
	0.25	**0.5**	**1**	**2**	**3**	**4**	**6**	
User A	15	15	25	30	35	40	55	96
User B	15	10	0	5	40	65	75	106
User C	15	25	30	60	60	70	70	101

Table 5.11: New user profiles

The target data is obtained by measuring a user's optimal SII for every 5 second segment of the noisy audio file (14.45 minutes in length), thus, generating 174 testing patterns. This data set is used to test the performance of the trained MLP and ANFIS models. After three trials of randomly shuffling the testing set, the average of the system outputs are taken. As well, a fixed gain at 0 dB is analyzed with each user to determine the significance of adapting the volume gain. The performances of the models and the fixed gain are evaluated using the performance measures presented in equations 4.14, 4.16 and 4.18.

5.5.1 User A

Figure 5.15 demonstrates the performances of the MLP and the ANFIS models in optimizing the SII for user A, over the 14.45 minute (867 seconds) sound recording. User A has a slightly moderate hearing loss, thus, the MLP and ANFIS models are tested using the 6-user specific features for a moderate hearing loss listed in table 4.4. As well, presented on the plot is the predicted SII from the fixed volume gain at 0 dB (no adaptation).

From figure 5.15 it is evident that both the MLP and the ANFIS model tracks the target SII effectively. Whereas the fixed volume gain does not track the target SII as effectively throughout the sound recording. This demonstrates the importance of adaptation (MLP and ANFIS), in order to preserve a user's speech intelligibility. This is more evident with the performance measures. The MLP yielded an SII error $mean_{SII}\pm STD_{SII}$ of 0.008±0.03, ANFIS led to an SII error of $mean_{SII}\pm STD_{SII}$ of 0.004±0.03 and the fixed volume gain (0 dB) yielded a $mean_{SII}\pm STD_{SII}$ of 0.1±0.2. Thus, from these results both the MLP and ANFIS models outperforms the fixed volume gain and the ANFIS model is slightly advantageous compared to the MLP model.

The MLP and ANFIS models also tested User A using the 6 unified features listed in table 4.7. Figure 5.16 demonstrate the MLP and ANFIS testing performances, as well as the fixed gain (0 dB) performance. From figure 5.16 the MLP model does not track the target SII as effectively as the ANFIS model. This is more apparent from the performance measures, the MLP led to a $mean_{SII}\pm STD_{SII}$ of 0.01±0.05 and ANFIS yielded a $mean_{SII}\pm STD_{SII}$ of 0.006±0.03. For the fixed volume gain, there is no change in error, $mean_{SII}\pm STD_{SII}$ of 0.1±0.2.

5.5.2 User B

The performances of the MLP and the ANFIS models tracking the SII for user B are presented in figure 5.17. User B possess a slightly severe hearing loss, thus, the MLP and ANFIS models are tested using the 6-user specific features for the severe hearing loss listed in table 4.5. From figure 5.17 it is shown that both the MLP and the ANFIS

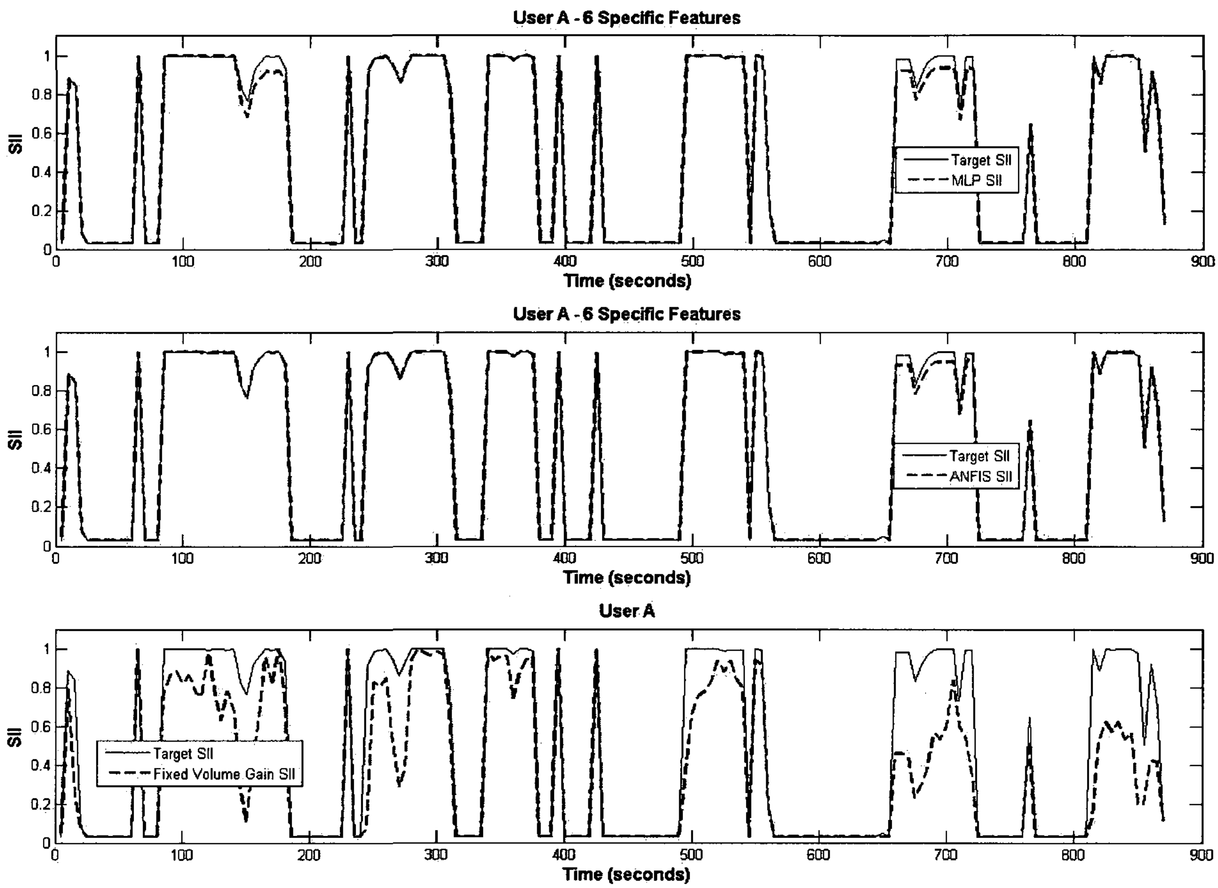


Figure 5.15: SII from MLP, ANFIS and fixed volume gain for User A using 6 user-specific features compared to target SII

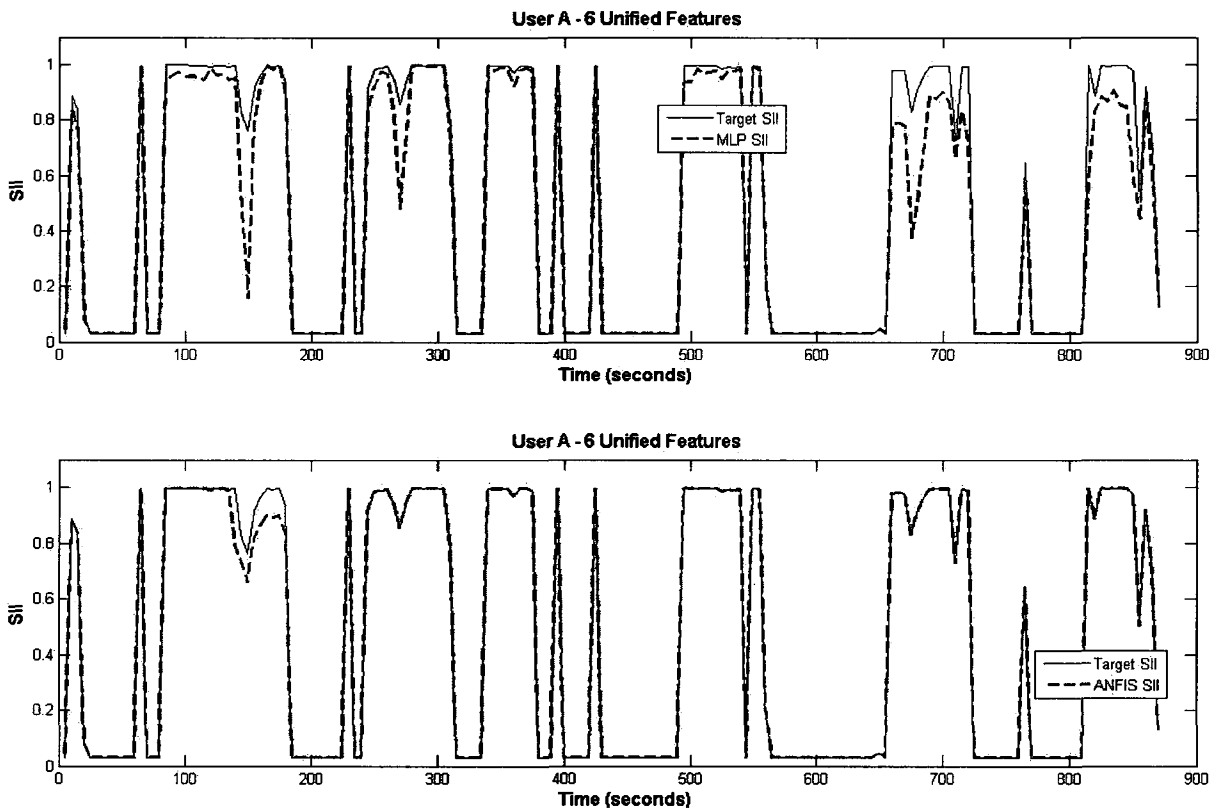


Figure 5.16: SII from MLP and ANFIS for User A using 6 unified features compared to target SII

models track the target SII for user B relatively well, unlike the fixed volume gain (no adaptation). The MLP showed an SII error $mean_{SII} \pm STD_{SII}$ of 0.005±0.03 which is slightly higher than ANFIS which gave a $mean_{SII} \pm STD_{SII}$ of 0.002±0.03. As well, the fixed volume gain obtained a much higher error result $mean_{SII} \pm STD_{SII}$ of 0.02±0.05.

Figure 5.18 presents the performances of the MLP and ANFIS model tracking the target SII for user B and using the unified set of features listed in table 4.7. From figure 5.18, it is seen that the MLP model is not effectively tracking the target SII for user B as compared to the ANFIS model. This is apparent by the SII error, where the MLP obtained a $mean_{SII} \pm STD_{SII}$ error of 0.04±0.06 and ANFIS yielded a lower result with a $mean_{SII} \pm STD_{SII}$ of 0.008±0.01, demonstrating that the ANFIS model is better compared to the MLP.

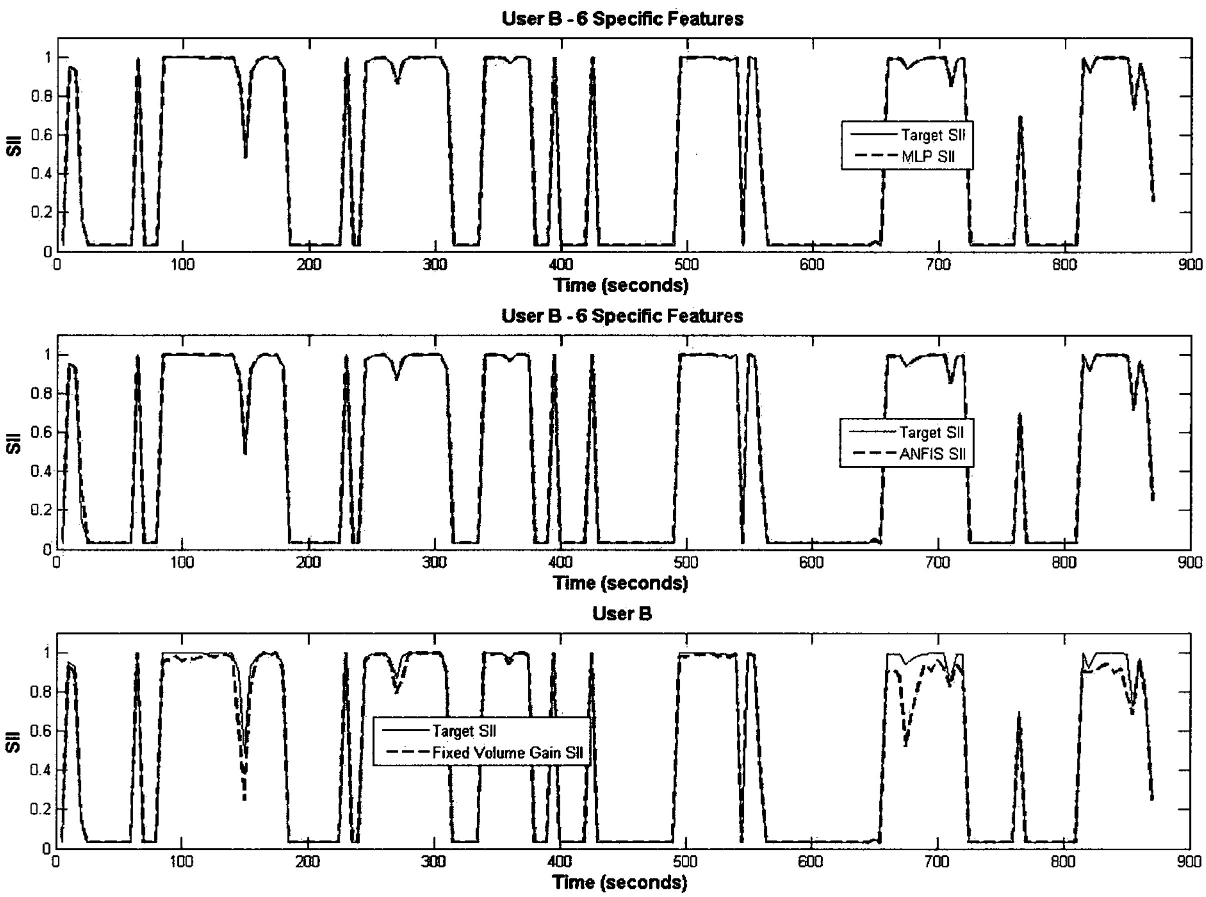


Figure 5.17: SII from MLP, ANFIS and fixed volume gain for User B using 6 user-specific features compared to target SII

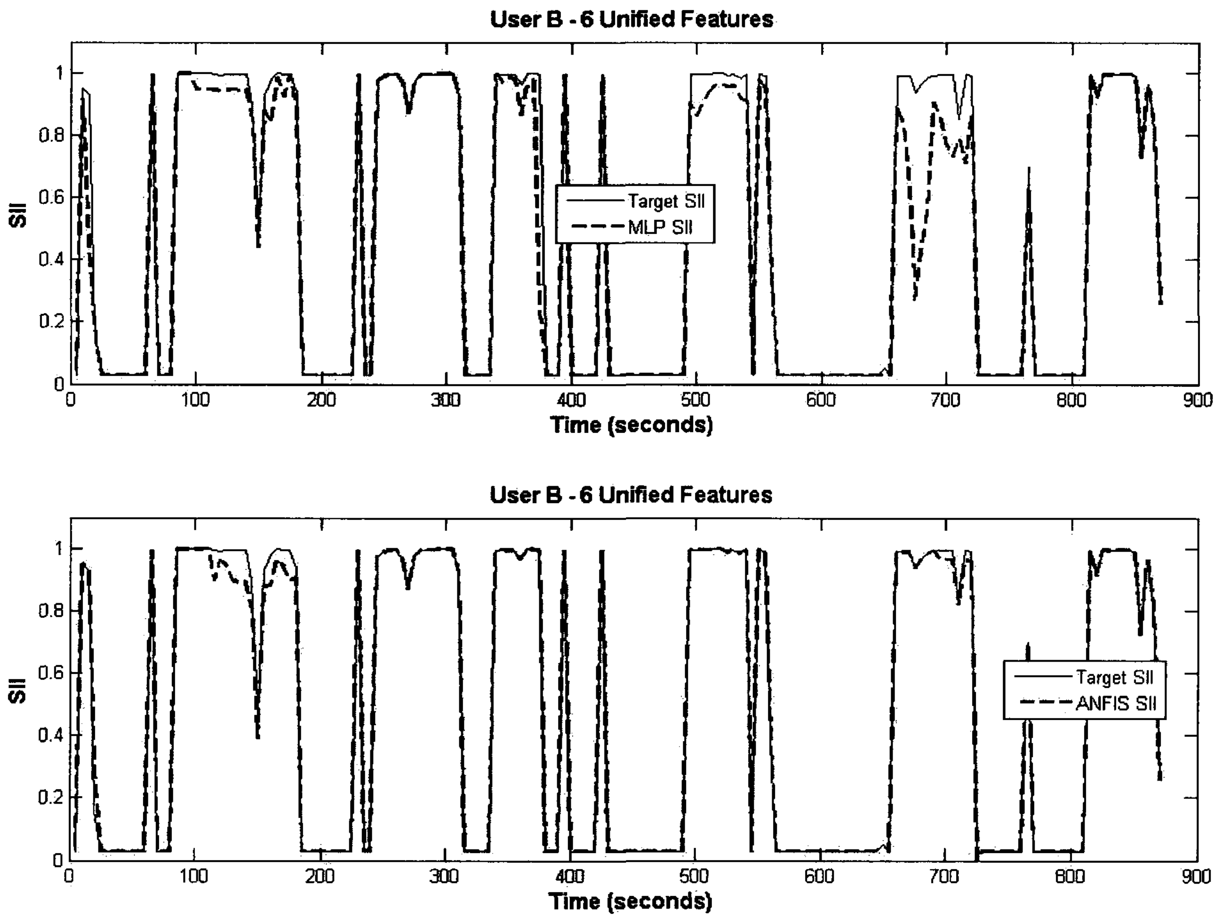


Figure 5.18: SII from MLP and ANFIS for User B using 6 unified features compared to target SII

5.5.3 User C

User C hearing loss is slightly profound, thus, the MLP and the ANFIS model are tested using the 6-user specific features for a profound hearing loss listed in table 4.6. Figure 5.19 also demonstrates the MLP and ANFIS models effectiveness in tracking the target SII for user C throughout the sound recording, compared to the fixed volume gain. This is shown in the SII error, the MLP obtained a $mean_{SII} \pm STD_{SII}$ of 0.01±0.03 and ANFIS led to a $mean_{SII} \pm STD_{SII}$ of 0.008±0.03. As well, the fixed volume gain yielded a much higher error result of $mean_{SII} \pm STD_{SII}$ 0.2±0.2.

Figure 5.20 present the performances of the MLP and the ANFIS models using the 6 unified features listed in table 4.7. From figure 5.20, both the MLP and the ANFIS models track the target SII for user C relatively well. The MLP obtained a $mean_{SII} \pm STD_{SII}$ error of 0.03±0.04 and the ANFIS yielded a $mean_{SII} \pm STD_{SII}$ error of 0.01±0.03, these errors are slightly higher compared to user A and user B, this is expected due to the narrow intelligibility curves that profound hearing loss users posses (figure 4.5).

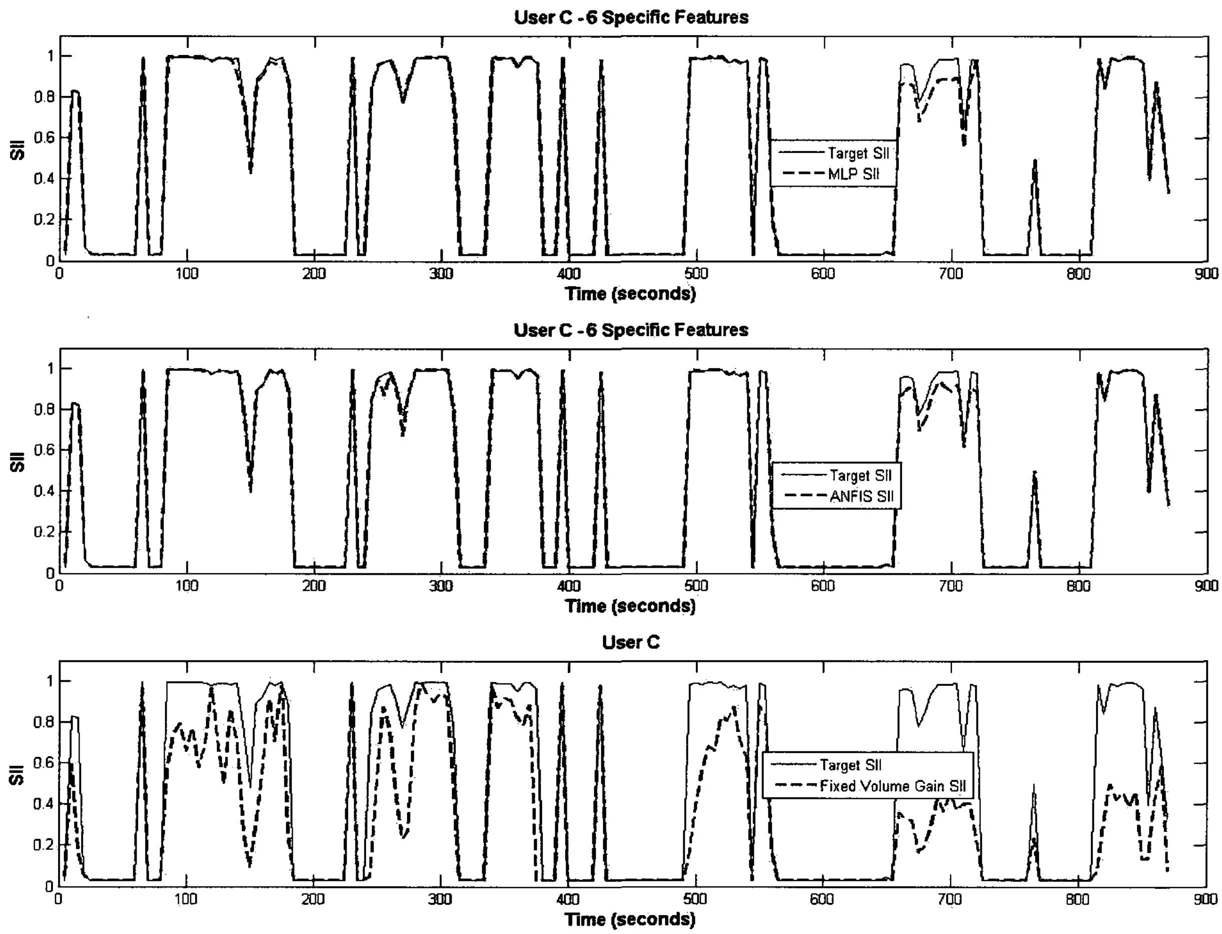


Figure 5.19: SII from MLP, ANFIS and fixed volume gain for User C using 6 user-specific features compared to target SII

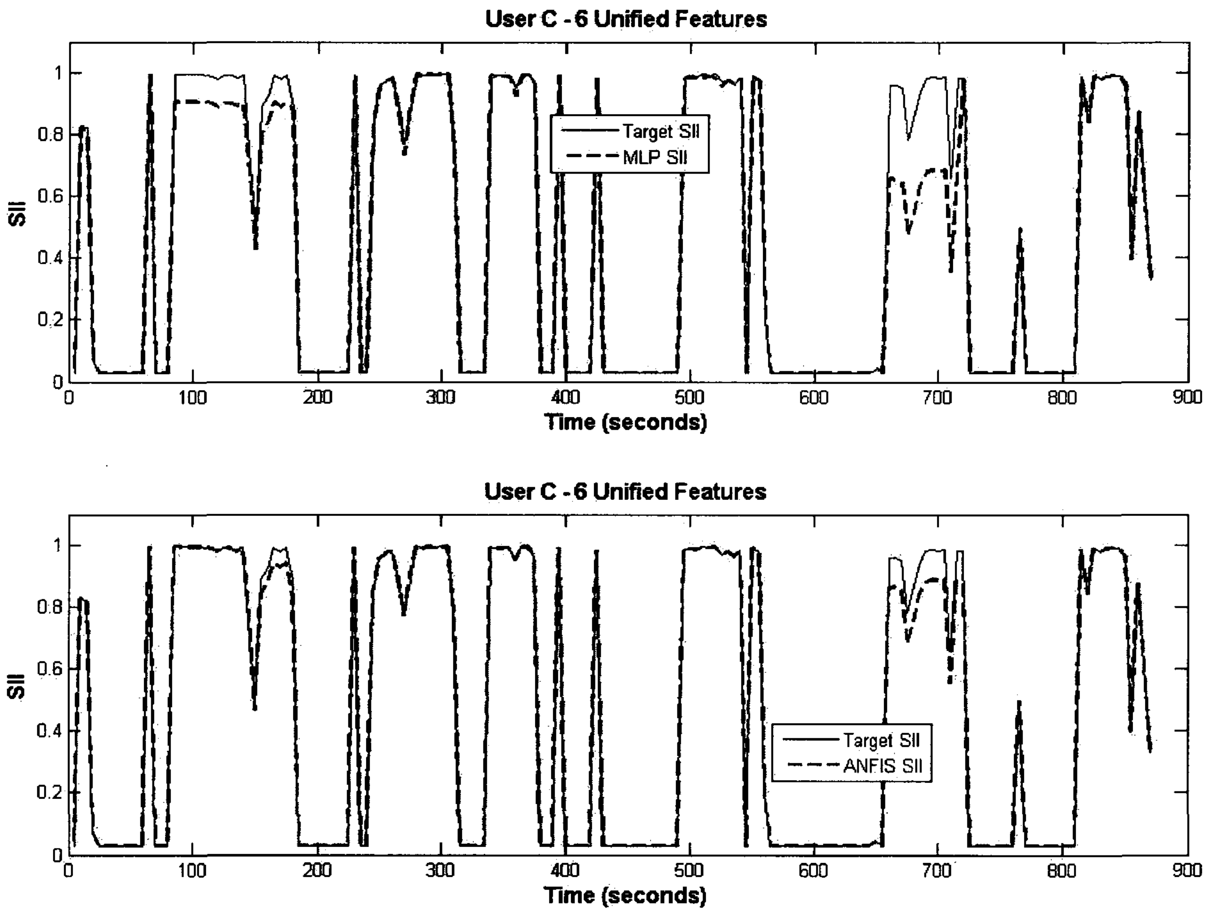


Figure 5.20: SII from MLP and ANFIS for User C using 6 unified features compared to target SII

5.6 Summary and Comparison

In this chapter, two prediction models MLP and ANFIS were introduced to perform automatic volume settings to optimize speech intelligibility for specific simulated hearing loss users. Four MLP models were implemented for three simulated user profiles (moderate, severe and profound hearing losses). The architecture of the MLP model was designed having three layers: an input layer, hidden layer and output layer. Inputs to the MLP model was one of the 2, 4, 6 or 8 influential features and the number of neurons in the input layer corresponded to the number of inputs. The output layer of the MLP model had one output neuron which was the optimal volume setting. The number of neurons in the hidden layer were determined based on the amount of hidden neurons which obtained the lowest average volume error between the training RMSE and testing RMSE. Four ANFIS models were also implemented for each user having the 2, 4, 6 or 8 influential features as inputs, Gaussian membership functions per input and 1 output with a linear membership function.

The performances of the MLP and ANFIS models were evaluated using the performance measures: volume error ($mean_{|V|} \pm STD_{|V|}$) and SII error ($mean_{SII} \pm STD_{SII}$). Comparing the training and testing performances of all the MLP and ANFIS models from tables 5.5, 5.6, 5.7 and 5.8, the 6 influential feature inputs stood out as a good candidate for the proposed automatic volume setting systems.

Both the MLP and ANFIS showed high accuracy in optimizing SII for all three simulated user profiles (moderate, severe and profound). However, ANFIS performance was better compared to MLP. For the MLP, 95% of the testing patterns had a SII error of 0.06 or less in worst case scenario over 400 audio files, whereas for ANFIS, 95% of the testing patterns obtained an SII error of 0.03 or less over 400 audio files. These SII error results seem to be satisfactory since a 0.04 change in SII will lead to approximately a 10% decrease in word intelligibility [18].

As well, the performances of the MLP and ANFIS were compared using 6-user specific features versus 6 unified features for all three types of hearing losses. Shown in section 5.4, it was determined that both ANFIS and the MLP perform better at an expense of using features specific to the type of hearing loss, rather than a unified set of features. However, using one unified set of features for the three simulated users only lead to a slight increase in error. Using a generic set of features also reduces the hardware complexity of the hearing aid.

Section 5.5 showed the performances of the MLP and ANFIS models tested with a sound recording with varying background environments different from the original database. The test showed that the MLP and ANFIS models can track the optimal SII for a simulated user more accurately using the 6-user specific features than using the 6 unified features. However, the decrement in performance was minimal with ANFIS. As well, the test demonstrated that having adaptation preserves the speech intelligibility for a specific user. This was shown with the fixed volume gain at 0 dB, which poorly tracked the target SII throughout the sound recording.

Therefore, from the results seen in this chapter it is evident that both the MLP and the ANFIS models are potential candidates to perform automatic volume settings and optimize a specific user's speech intelligibility. However, ANFIS is slightly advantageous compared to the MLP.

Chapter 6

Conclusions

6.1 Summary of Findings

The objective of this dissertation was to examine methods to automatically control the volume settings of a hearing aid to maintain optimal listening at all times. In this thesis, two prediction models were utilized to perform automatic volume settings for three simulated users, representing individuals with moderate, severe and profound hearing losses. The goal of the prediction model was to maximize speech intelligibility in all situations.

The first prediction model, the multilayer perceptron (MLP), is a feed-forward neural network that performs automatic volume settings by adaptively updating its weights to reach a target volume setting for a particular user. The targets are obtained using a hearing aid simulator, which provides the target volume settings at maximum SII for a simulated user. The second prediction model, the Adaptive Network-based Fuzzy Inference System (ANFIS) is also a feed-forward network; however, it uses fuzzy reasoning to perform the automatic volume settings for the user.

The MLP and ANFIS systems differ in terms of self-learning. The MLP applies the backpropagation gradient descent method to update the weights of the neural network in order to predict a given output data set. On the other hand, the ANFIS uses the backpropagation gradient descent method and the least-squares method for tuning the parameters of the input membership functions and output membership functions, respectively, in order to predict a given output data set.

Feature selection is a useful technique to perform in order to select input features which are correlated to the output volume settings, thus improving the system's performance. The feature selection method SFS was used in this thesis to select three sets of influential features. Each set corresponded to one of the simulated users with moderate, severe and profound hearing losses. As well, the SFS technique was used to select a unified set of influential features for all three simulated users.

Based on the previous simulation and results, several conclusions can be drawn. Both the MLP and ANFIS predict the optimal volume setting for the majority of the testing patterns for the three simulated users. Section 5.3.2 showed that for the three simulated users, large volume errors did not lead to large SII errors for the majority of the testing patterns. This was seen particularly with the moderate hearing loss user due to the wide plateau on the intelligibility curves presented in figure 4.3. In contrast, the profound hearing loss user obtained slightly higher SII error compared to the moderate and severe hearing loss users, due to a narrower intelligibility curves shown in figure 4.5.

ANFIS performance was better than MLPs performance in terms of SII error. In the worst case scenario, ANFIS obtained a SII error of 0.03 or less for 95% of the testing patterns over 400 audio files. The MLP showed 95% of the testing patterns had a SII error of 0.06 or less over 400 audio files. From the data of Fletcher and Galt [18], a 0.04 change in SII leads to a change in word intelligibility of 10% or less. Accordingly, both systems are effective in maintaining near optimal intelligibility for the simulated user.

As well, using a unified set of input features for various types of hearing losses may be helpful for the hearing aid manufacturer, thus reducing the complexity of the hardware. In section 5.4, tables 5.9 and 5.10 shows that the MLP and ANFIS models perform better using three sets of influential features related to the three types of hearing losses, rather than using one set of unified features for the three simulated users. However the error yielded by using the unified features is still relatively low. Furthermore, the robustness of the MLP and ANFIS models were demonstrated by testing the models with a real sound recording that mimics a daily routine with varying background noises. For each simulated user, the majority of the time both the MLP and ANFIS models effectively tracked the simulated user's optimal SII throughout the sound recording, compared to a fixed gain. This demonstrates the importance of adapting the volume gain, where adaptation in hearing aids can improve the user's speech intelligibility in changing environments, thus possibly increasing their comfort level.

6.2 Contributions

This study opens doors for new research and development of hearing aids. Integrating the proposed automatic volume settings system as a new hearing aid feature may improve users' speech intelligibility, and facilitate hearing for a broad base of users. In addition, this dissertation makes the following contributions:

- Provides initial volume settings to a user's hearing aid in a wide range of listening situations (over 2000 audio files), which may optimally match a user's volume setting preference;
- Uses prediction models to ascertain user volume setting preferences directly from the input sound or features, without a priori classification of the environment;
- Optimizes a simulated user's speech intelligibility by performing automatic volume settings; and
- Shows that the ANFIS prediction model is slightly advantageous compared to the MLP prediction model in performing automatic volume settings and optimizing a simulated user's speech intelligibility
- Demonstrates the improvement in performance of the ANFIS and MLP models using features which are specifically correlated to the user's audiogram, rather than using a unified set of features for all three types of audiograms (moderate HL, severe HL, profound HL)
- Shows that adapting the volume gain using MLP and ANFIS models preserves the speech intelligibility for a specific user.

6.3 Potential Future Research Directions

In the future, this research may explore the selection of specific acoustic features to ascertain a user's volume setting preference. Determining features which are correlated to a user's volume settings is an essential task to ensure that the features selected are not classified as noise to the prediction model, and so that the model accurately maps the input space (features) to the output space (target volume settings).

Furthermore, although the suggested prediction models stated in this thesis were only applied to learn the volume setting preferences, they can be easily applied to learn the other hearing aid settings that the user can control, such as treble or bass and turning the directional microphone or noise reduction algorithm on or off. As well, an extension of the current approach can be done by incorporating multiple systems trained to learn optimal volume settings under different types of distortion or reverberation conditions, instead of using a single MLP or ANFIS system to handle all conditions.

Lastly, future research may also seek to obtain training data from the preferred volume settings of real human subjects in different acoustic environments. Testing the proposed system on real patients can be done by determining whether the trained prediction models do provide initial volume settings preferred by the patient. The next stage will be to integrate the proposed system into a trainable self-learning hearing aid to learn actual user preferences.

Bibliography

[1] Audibel. `http://www.audibel.com/product/styles.html`, November 2008.

[2] Hearing aids. `http://www.lhsc.on.ca/cochlear/anatomy.htm`, November 2008.

[3] E. Alexandre, L. Cuadra, L. Alvarez, M. Rosa-Zurera, and F. Lopez-Ferreras. Automatic sound classification for improving speech intelligibility in hearing aids using a layered structure. In *Springer-Verlag*, 2006.

[4] E. Alexandre, L. Cuadra, M. Rosa, and F. Lopez-Ferreras. Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(8):2249–2256, November 2007.

[5] ANSI-S3.5-1997. American national standard: Methods for the calculation of the speech intelligibility index. In *American National Standards Institute*, New York, 1997.

[6] K. Assaleh. Extraction of fetal electrocardiogram using adaptive neuro-fuzzy inference systems. *IEEE Transactions on Biomedical Engineering*, 54(1):59–68, January 2007.

[7] G.M. Ballou. Handbook for sound engineers. *Butterwoth-Heinemann*, 2001.

[8] M. Büchler. Algorithms for sound classification in hearing instruments. *PhD thesis, Swiss Federal Institute of Technology*, 19(14498), 2002.

[9] J. Chalupper, D. Junius, and T. Powers. Algorithm lets users train aid to optimize compression, frequency shape, and gain. *The Hearing Journal*, 62(8):26–30, 2009.

[10] C. H. Chang, G. T. Anderson, and P. C. Loizou. A neural network model for optimizing vowel recognition by cochlear implant listeners. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(1):42–48, March 2001.

[11] H. Dillon. *Hearing Aids*. Boomerang Press, 2001.

[12] H. Dillon, A. Justin, H. McDermott, and G. Keidser. The trainable hearing aid: What will it do for clients and clinicians? *The Hearing Journal*, 59(4):30–36, 2006.

[13] Q. Ding and M. Zhang. Classification of recorded musical instruments sounds based on neural networks. In *Computational Intelligence in Image and Signal Processing, 2007. CIISP 2007. IEEE Symposium on*, pages 157–162, Honolulu, HI, April 2007.

[14] W. A. Dreschler, G. Keidser, E. Convery, and H. Dillon. Client-based adjustments of hearing aid gain: The effect of different control configurations. *Ear and Hearing*, 29:214–227, 2008.

[15] B. Edwards. Speech processing in the auditory system: Hearing aids and hearing impairment. *Springer New York*, pages 339–421, 2004.

[16] D. Enqing, L. Guizhong, Z. Yatong, and C. Yu. Voice activity detection based on short-time energy and noise spectrum adaptation. In *Signal Processing, 2002 6th International Conference on*, volume 1, pages 464–467, August 2002.

[17] C. Euisun and L. Chulhee. Optimizing feature extraction for multiclass problems. In *IEEE Transactions on Geoscience and Remote Sensing*, volume 39, October 2008.

[18] H. Fletcher and R.H. Galt. The perception of speech and its relation to telephony. *Journal of The Acoustical Society of America*, 22(2):89–151, 1950.

[19] J. Galvin and M. Scherer. Evaluating, selecting, and using appropriate assistive technology. *Aspen Publishers*, pages 144–161, 1996.

[20] R. Gao, S. Basseas, D. T. Bargiotas, and L. H. Tsoukalas. Next-generation hearing prosthetics. *IEEE Robotics & Automation Magazine*, 10(1):21–25, March 2003.

[21] I. Guler and E. D. Ubeyli. Automated diagnostic systems with diverse and composite features for doppler ultrasound signals. *IEEE Transactions on Biomedical Engineering*, 53(10):1934–1942, October 2006.

[22] K. Hadad, M. Mortazavi, M. Mastali, and A. A. Safavi. Enhanced neural network based fault detection of a VVER nuclear power plant with the aid of principal component analysis. *IEEE Transactions on Nuclear Science*, 55:3611–3619, December 2008.

[23] M.T. [Hagan and H.B. Demuth. Neural networks for control. In *Proceedings of the 1999 American Control Conference*, pages 1642–1656, 1999.

[24] M. Hersh and M. Johnson. Assistive technology for the hearing-impaired, deaf and deafblind. *Springer London*, 39(1):1–39, 2003.

[25] B. W. Y. Hornsby. The speech intelligibility index: What is it and what's it good for? In *The Hearing Journal*, volume 57, pages 10–17, October 2004.

[26] T. Houtgast and H. Steeneken. Past, present and future of the speech transmission index. In *TNO Human Factors*, pages 3–29. S. V. Wijngaarden, Ed, October 2002.

[27] J. S. R. Jang. ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3):665–685, May/June 1993.

[28] F. O. Karray and C. De Silva. *Soft Computing and Intelligent Systems Design: Theory, Tools and Applications.* Pearson Education Limited, England, 2004.

[29] J.M. Kates and K.H. Arehart. Coherence and the speech intelligibility index. *Journal of The Acoustical Society of America*, 117:2224–2237, 2005.

[30] G. Keidser. Many factors are involved in optimizing environmentally adaptive hearing aids. *The Hearing Journal*, 62(1):26–29, 2009.

[31] S. Kochkin. Marketrak v: "why my hearing aids are in the drawer": The consumers' perspective. *The Hearing Journal*, 53(2):34–41, 2000.

[32] S. Kochkin. Marketrak vi: Isolating the impact of the volume control on customer satisfaction. *Hearing Review*, pages 1–6, 2003.

[33] S. Kochkin. Marketrak vii: Hearing loss population tops 31 million people. *Hearing Review*, 12(1):16–19, 2005.

[34] S. Kumar. A review of smart volume controllers for consumer electronics. *IEEE Transactions on Consumer Electronics*, 51(2):600–605, May 2005.

[35] L. Lamarche. Adaptive environmental classification system for hearing aids. Master's thesis, University of Ottawa, 2008.

[36] M. Liu, C. Wan, and L. Wang. Content-based audio classification and retrieval using a fuzzy logic system: Towards multimedia search engines. In *Springer-Verlag*, 2002.

[37] L. Lu, H. J. Zhang, and H. Jiang. Content analysis for audio classification and segmentation. *IEEE Transactions on Speech and Audio Processing*, 10(7):504–516, October 2002.

[38] MATLAB. *Fuzzy Logic Toolbox*, 7.4.0 edition, 2007.

[39] MATLAB. *Neural Network Toolbox*, 7.4.0 edition, 2007.

[40] S. McAdams. Perspectives on the contribution of timbre to musical structure. In *Computer Music Journal*, volume 23, October 2008.

[41] R. Meredith. Volume control preferences with wdrc digital hearing aids. *The Hearing Review*, pages 1–9, 2006.

[42] H. G. Mueller, B. W.Y. Hornsby, and J. E. Weber. Using trainable hearing aids to examine real-world preferred gain. *Journal of the American Academy of Audiology*, 19(10):758–773, 2008.

[43] V. Onnia, M. Tico, and J. Saarinen. Feature selection method using neural network. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 513–516, Thessaloniki, October 2001.

[44] H. Othman, T. Aboulnasr, C. Gigure, W. Gueaieb, L. Lamarche, R. Rahal, A. Cole, L. Zhou, N. Ellaham, and J. Zhou. Learning systems. Version 1.8, May 2009.

[45] S. Ravindran and D. V. Anderson. Audio classification and scene recognition and for hearing aids. In *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pages 860–863, May 2005.

[46] A. H. Reyes, E. Sanchez-Sinencio, and J. F. Duque-Carrillo. A wireless volume control receiver for hearing aids. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 42(1):16–23, January 1995.

[47] E. Romero and J. M. Sopena. Performing feature selection with multilayer perceptrons. *IEEE Transactions on Neural Networks*, 19(3):431–441, March 2008.

[48] M. Ross and B. E. Walden. Digital hearing aids: Magic, marketing, and clinical trials. *The Hearing Journal*, 54(3):56–63, 2001.

[49] R. K. Surr, M. T. Cord, and B. E. Walden. User-operated volume control: Response of hearing aid wearers to the absence of a user-operated volume control. *The Hearing Journal*, 54(4):32–36, 2001.

[50] H. Takagi and M. Ohsaki. Interactive evolutionary computation-based hearing aid fitting. *IEEE Transactions on Evolutionary Computation*, 11(3):414–427, June 2007.

[51] Siemens Audiologische Technik. *GmbH SAT AEGS Gebbertstr.* 125 91058 Erlangen.

[52] B. Verma and J. Zakos. A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Transactions on Information Technology in Biomedicine*, 5(1):46–54, March 2001.

[53] A. Vonlanthen and H. Arndt. *Hearing Instrument Technology: for the Hearing Healthcare Professional Third Edition.* Thomson Delmar Learning, 2007.

[54] K. C. Wagener, M. Hansen, and C. Ludvigsen. Recording and classification of the acoustic environment of hearing aid users. *Journal of the American Academy of Audiology*, 19(4):348–370, 2008.

[55] C. H. Wang, B. H. Xiao, and R. W. Dai. A new integration scheme with multilayer perceptron networks forhandwritten chinese character recognition. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 961–964, Barcelona, Spain, 2000.

[56] A. Ypma, B. de Vries, and J. Geurts. Robust volume control personalisation from on-line preference feedback. In *Machine Learning for Signal Processing, 2006. Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on*, pages 441–446, Arlington, VA, September 2006.

[57] J. A. Zakis, H. Dillon, and H. J. McDermott. The design and evaluation of a hearing aid with trainable amplification parameters. *Ear and Hearing*, 28(6):812–830, 2007.