University of Wisconsin Milwaukee
UWM Digital Commons

Theses and Dissertations

December 2018

Improvement of Speech Perception for Hearing-Impaired Listeners

Fang Du
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Electrical and Electronics Commons, and the Speech Pathology and Audiology Commons

IMPROVEMENT OF SPEECH PERCEPTION FOR HEARING-IMPAIRED

LISTENERS

by

Fang Du

A Dissertation Submitted in

Partial Fulfillment of the

Requirement of the Degree of

Doctor of Philosophy

in Engineering

at

The University of Wisconsin-Milwaukee

December 2018

ABSTRACT

IMPROVEMENT OF SPEECH PERCEPTION FOR HEARING-IMPAIRED
LISTENERS

by

Fang Du

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Yi Hu

Hearing impairment is becoming a prevalent health problem affecting 5% of world adult populations. Hearing aids and cochlear implant already play an essential role in helping patients over decades, but there are still several open problems that prevent them from providing the maximum benefits. Financial and discomfort reasons lead to only one of four patients choose to use hearing aids; Cochlear implant users always have trouble in understanding speech in a noisy environment.

In this dissertation, we addressed the hearing aids limitations by proposing a new hearing aid signal processing system named Open-source Self-fitting Hearing Aids System (OS SF hearing aids). The proposed hearing aids system adopted the state-of-art digital signal processing technologies, combined with accurate hearing assessment and machine learning based self-fitting algorithm to further improve the speech perception and comfort for hearing aids users. Informal testing with hearing-impaired listeners showed that the testing results from the proposed system had less than 10 dB (by average) difference when compared with those results obtained from clinical audiometer. In addition, Sixteen-channel filter banks with adaptive differential microphone array

ii

provides up to six-dB SNR improvement in the noisy environment. Machine-learning based self-fitting algorithm provides more suitable hearing aids settings.

To maximize cochlear implant users' speech understanding in noise, the sequential (S) and parallel (P) coding strategies were proposed by integrating high-rate desynchronized pulse trains (DPT) in the continuous interleaved sampling (CIS) strategy. Ten participants with severe hearing loss participated in the two rounds cochlear implants testing. The testing results showed CIS-DPT-S strategy significantly improved (11%) the speech perception in background noise, while the CIS-DPT-P strategy had a significant improvement in both quiet (7%) and noisy (9%) environment.

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

x

ACKNOWLEDGMENTS

During the five years of my studies in the University of Wisconsin-Milwaukee, many people have provided me their tremendous help.

First and foremost, I would like to express my special thanks to my major Ph.D. advisor, Dr. Yi Hu. I sincerely appreciate his outstanding support, guidance, patience, and help in my academic career. I am so honored and lucky to be one of his Ph.D. students. I would also like to express my gratitude to my co-advisor, Dr. Christina L. Runge for her patience and immense clinical knowledge. Her brilliant comments and suggestions lifted my research efforts to a higher level.

I was fortunate to receive fantastic feedback from my Ph.D. committee: Dr. Seyed H Hosseini, Dr. Chiu Law, and Dr. Guangwu Xu. I'd like to express my sincere thanks for their encouragement and insightful comments.

Furthermore, I gratefully acknowledge MED-EL that provided funding to make my Ph.D. work possible. I want to thank Sarah Mleziva for the subject recruitment and data collection, Dr. Josh Stohl for the MEL-EL device technical support. I want to thank my lab mate Santhosh Yegnaraman for his friendship, good advices, and collaborations.

I also received a lot of help from the professors in the University of Wisconsin-Milwaukee. I'd like to wholeheartedly thank Dr. Robert Turney for offering me an internship opportunity in Johnson Controls; I also really enjoyed the time working with Dr. George Hanson and Dr. David Yu as their TAs.

At last but not least, I would like to thank my parents and my friends for their warm love, continued patience, and endless support.

Chapter 1

Introduction

1.1 Background and Motivation

Hearing impairment or hearing loss is a partial or complete loss of hearing ability [1]. From the definition by World Health Organization [2], if a person cannot hear the 25dB pure tone, this person can be diagnosed with hearing impairment or hearing loss.

Depending on the patients' hearing thresholds, hearing impairment can be categorized into four different levels: mild, moderate, severe, and profound as shown in Figure 1-1. There are two treatments that could help hearing impaired listeners. Hearing aids work best for people with mild, moderate and sometimes severe hearing loss. Cochlear implant can benefit people with severe and profound hearing loss.

```
              ┌─────────────────────────────┐
              │    Hearing Impairment        │
              │  hearing threshold of 25dB or│
              │            worse             │
              └─────────────────────────────┘
       ┌──────────┬──────────┴──────────┬──────────┐
┌──────────┐ ┌──────────┐ ┌──────────┐ ┌──────────┐
│   Mild   │ │ Moderate │ │  Severe  │ │ Profound │
│ hearing  │ │ hearing  │ │ hearing  │ │ hearing  │
│threshold │ │threshold │ │threshold │ │threshold │
│ 25 – 40dB│ │ 40 – 70dB│ │ 70 – 95dB│ │  >95dB   │
└──────────┘ └──────────┘ └──────────┘ └──────────┘
```
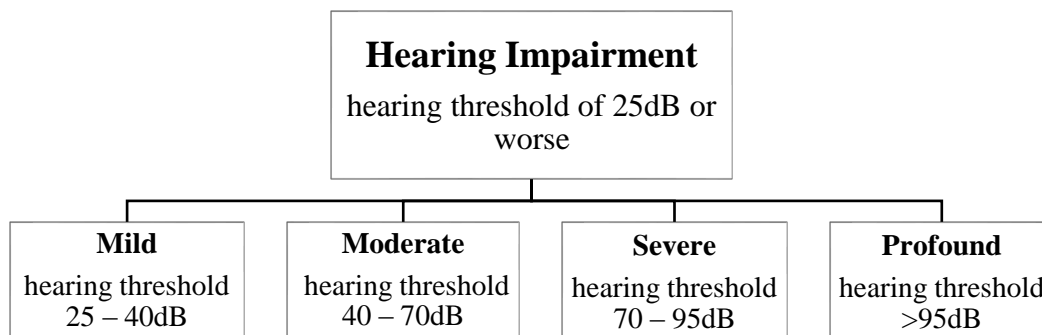
Figure 1-1 Hearing impairment category

According to World Health Organization, overall 5% of the world's population - around 432 million adults have disabling hearing loss [2]. Some studies [3][4] estimated that hearing loss population would double by 2050. From the summary produced by The National Institute on Deafness and Other Communication Disorders, approximately 15% of American adults reported some levels of hearing loss by 2014 [5]. The majority of the hearing-impaired listeners were over an age of 55. In this research study, we focus on the adult population of hearing-impaired listeners.

Figure 1-2 shows a typical hearing aid. It consists of four major parts: microphone, battery, amplifier, and receiver/speaker. The microphone picks up the sound and sends to amplifiers, the amplifier analyzes the microphone signal and makes the sound louder and clearer. Then the speaker built in the earmold sends the stronger sound into the ear canal.
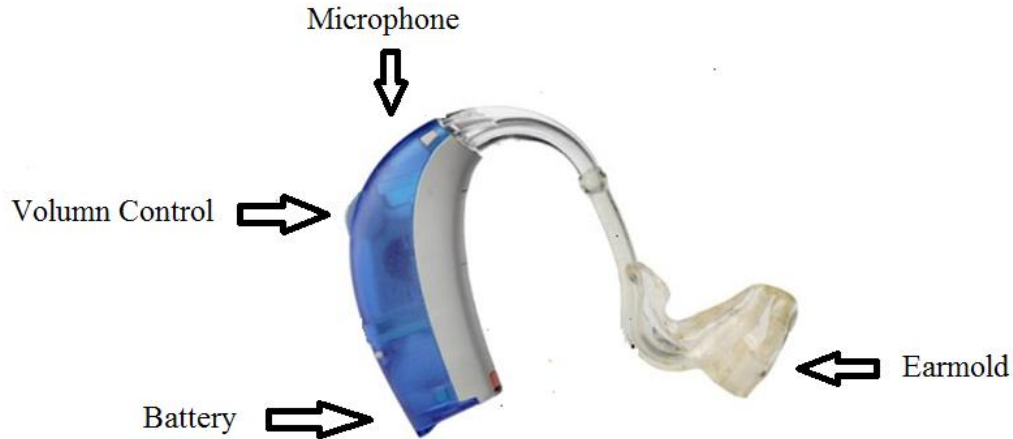
Figure 1-2 Structure of hearing aids

Cochlear implant is a device that benefits people who are profoundly deaf. Figure 1-3 illustrates the structure of a cochlear implant device. A cochlear implant device consists of two parts: the part composed of microphone, sound processor and transmitter that is worn outside of the body (Figure 1-3 on the right), and the part composed of receiver and

electrodes that are surgically placed inside the skull (Figure 1-3 on the left). When the device is on, the microphone picks up the sound; the speech processor converts the acoustic signals into electrical current signals; finally the transmitter, receiver and electrodes deliver the electric pulses to the auditory neurons.

Figure 1-3 Structure of cochlear implant

It has been reported that nearly 30 million US adults could benefit from hearing aids, but only 20% of them actually used it [2]. McCormack [6] analyzed the data and concluded that it is due to two primary potential reasons:

- Financially: The average price of a single hearing aid is 2,300 dollars [7]. If patients have bilateral hearing loss, the cost will double. Heading aid is a digital device similar to a smartphone or a laptop, and the lifespan is about 3-5 years. In other words, patients need to spend 2,000 dollars every three years to acquire the hearing aid device. Additionally, the battery of hearing aids is typical a Zinc-Air battery, and since it is unchangeable, it needs to be replaced every week.

- Practically: Patients need to see a doctor several times for hearing assessment and device fitting to get the best hearing aids setting. It is inconvenient for patients living in rural area. Some studies [8]-[10] noticed that some patients could not get the preferred setting even with several clinical visits.

The first aim of this research is to develop a new hearing aids solution to deal with financial and practical issues. The new hearing aids need to be a low-cost, rechargeable, self-fitting (integrated with a self-hearing assessment) device that can benefit more patients.

Cochlear implant has been a groundbreaking treatment option for hearing-impaired patients who are unable to benefit from hearing aids. Improvements in technology and expansions in candidacy criteria have helped many cochlear implant recipients regain satisfactory speech perception in quiet during the past three decades [11]. But there remains variability in performance across recipients, and speech perception in noise remains challenging [12]-[14]. Many signal processing techniques such as single- or multiple-microphone noise-reduction and speech-enhancement methods have been developed to address this issue, but in general, they increase speech distortion as well as battery power consumption with limited benefits [15]. Primary reasons for the challenges with cochlear implants have been identified as the limited number of spectral channels and channel interaction, which are particularly detrimental for complex listening situations [15].

The second aim of this study is to focus on developing a new method that can further improve on human speech perception in background babble noise without additional speech distortion.

4

1.2 Contribution

This dissertation makes contributions to both hearing aids and cochlear implant research. In the area of hearing aids, this research demonstrated the possibility and strength of a complete hearing aid system with self-testing and self-fitting. The proposed open-source system combined the latest hearing aids digital signal processing technologies (adaptive microphone array, 16 channels of filter banks, and multi-band dynamic compression) along with machine learning algorithm to improve speech perception for hearing aids users.

In our contribution to the cochlear implant research, this dissertation proposed and implemented a new cochlear implant coding strategy based on a combination of continuous interleaved strategy (CIS) and desynchronized pulse train (DPT). Both sequential version and parallel version of the new strategy show significant improvements in speech perception in quiet and in noise.

1.3 Dissertation Organization

This dissertation is organized as shown below:

- Chapter 2 describes the algorithms used in open-source self-fitting hearing aids system.

- Chapter 3 covers specific details in both hardware and software implementation of the proposed hearing aids systems.

- Chapter 4 describes the proposed of new cochlear implant coding strategy.

- Chapter 5 presents as well as analyzes the system testing results and subject testing results of the proposed open-source self-fitting hearing aids systems and the proposed CIS-DPT coding strategy.

- Supplemental information and references are provided in the appendices and bibliography.

6

Chapter 2

Algorithms in Open-Source Self-Fitting Hearing Aids System

2.1 Introduction

The concept of self-fitting hearing aids was firstly introduced in 2011 [17]. It is a device that enables the users to perform threshold measurements and fine-tuning of the hearing aids setting without audiological support [18].

Table 2-1 Currently available commercial self-fitting hearing aids

Company	Price per device	Hearing testing	Fine-turning
iHear Medical	$199	App for PC and Phone	App for PC and Phone
SoundWorld Solutions	$349-499	App for PC and Phone	App for PC and Phone
SoundHawk	$350	Not applicable	App for Phone
America Hears	$699-1499	Professional audiogram	App for PC
HearSource	$995	Professional audiogram	App for PC
Blamey Saunders Hears	$1050-1650	Online speech- based or professional audiogram	App for PC and Phone

Table 2-1 summarizes the currently available user-programmable devices on the market. Some of the products have received patents [19][20], but most of them can be considered

as partly self-fitting since the fine-tuning process only provide an interface to let users change the amplification gains. This functionality may not be efficient because most of the patients are non-experts, and they still need audiologist to assist them to decide the best setting. Although the cost of the devices reduced compared to professional hearing aids, they are still relatively high.

Some researchers are working on open-source hearing aids platform. Hendrik Kayser [21] and his collages developed an open community platform called openMHA for hearing aids algorithm research. The system is Raspberry Pi-based which has a powerful CPU. Some hearing aids algorithms that are found in commercial devices were implemented. But overall the size as presented and the price point (Over $300) does not meet patient's requirement.

William Audette [22] and his collages developed a real-time audio processing system (defined as ASA system) based on Teensy 3.6 board. A user-friendly software development interface allowed people to program the device even without any knowledge in digital signal processing (DSP). But so far, the hearing aids algorithm running in this system cannot compete with other commercial products, while the price is around $200.

The proposed hearing aids system aims to be a low-cost device (less than $100) but contains self-testing, self-fitting, and advanced hearing aids signal processing algorithms. It consists of three major parts: hearing assessment, fitting and the sound amplification hardware device. The block diagram is illustrated in Figure 2-1. In this chapter, each part of the proposed system was introduced.
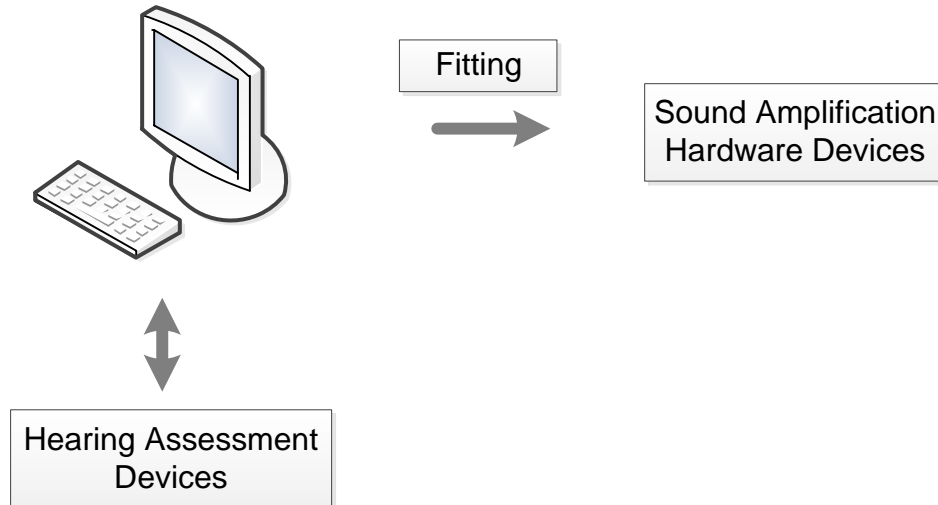
Figure 2-1 Open-source Self-Fitting Hearing Aids System Structure

2.2 Hearing Assessment

Hearing assessment is a means of examing an individual's overall hearing functionality. The goal of hearing assessment is to determine the lowest acoustic pressure level (in dB) a person can hear at frequencies varying from low pitches to high pitches. The hearing assessment will result in an audiogram, which shows a person's hearing thresholds at different frequencies.

The primary hearing testing method is the modified Hughson-Westlake procedure [23][24] which was proposed in 1986. This testing starts with playing a pure tone signal with the most comfortable level at frequency 1 KHz. Then the amplitude of the sound will be reduced in a fixed size until the listener cannot hear, whereafter the volume of the pure tone will be continually increased until listener responds again. The iteration will end if two out of three times the threshold was recorded at one frequency. The

9

incremental size usually is smaller than the decrement to protect listeners. The frequency of pure tone varies to cover 250 Hz – 8000 Hz.
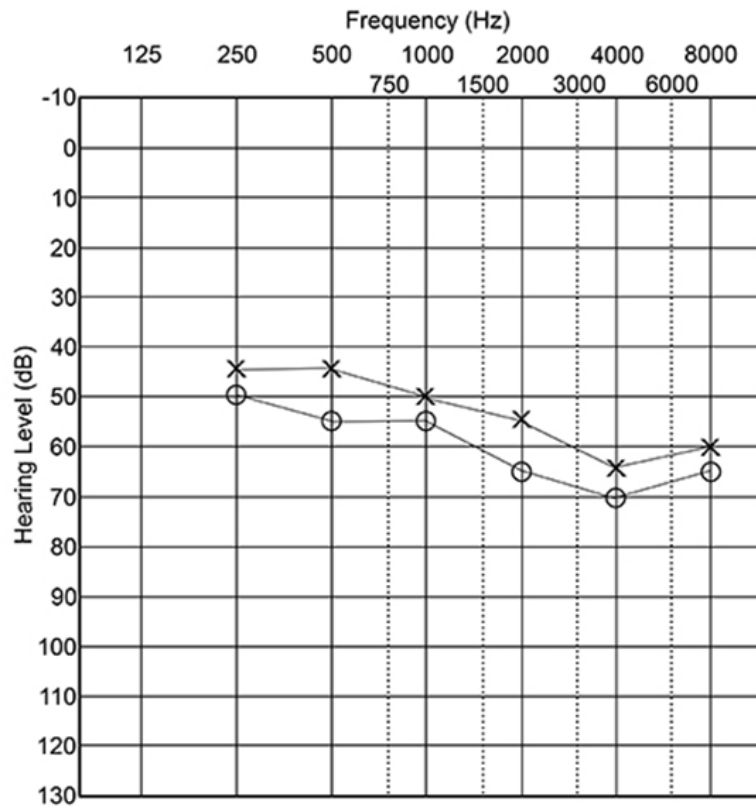
Figure 2-2 Audiogram Example

The results of the hearing assessment are recorded on a graph or in a table. The graph form is called audiogram. An example audiogram is illustrated in Figure 2-2. The audiogram visually explains the testing results. The horizontal and vertical axis represent the testing frequency in Hz and hearing threshold in hearing level decibel (dB HL), respectively. Different marks are used in audiogram (circle represents right ear, cross represents left ear) to distinguish testing result between the different ears for a single person.

The table form is called tabular audiogram. Table 2-2 displays an example associated with Figure 2-2. Instead of using a graph, the hearing threshold is presented in numerical values, which may be clearer to read and record.

Table 2-2 Tabular Audiogram Example

Frequency (Hz)	250	500	1000	2000	4000	8000
Right Ear(dB)	50	55	55	65	70	65
Left Ear(dB)	45	45	50	55	65	60

Patients usually need to work with an audiologist to receive an accurate audiogram. It may be difficult for patients who live far from an audiologist. Self-assessment audiometry will be an optimal solution for those patients. Self-recording audiometry [26][27] employs the use of an instrument to perform the hearing testing. The device can either increase or decrease the signal intensity at a fixed rate of decibels per second (dB/sec) based on listener's response. The self-recording audiometry requires the instrument to provide precise sound intensity output at individual frequencies; hence a calibrated device is needed to perform self-administered testing. The implementation details will be discussed in the following chapter.

2.3 Fitting

2.3.1 Amplification Prescription

Hearing impaired listeners have different levels of hearing loss, and the amplification parameters must be adjusted to fit each patient's requirements. The process that an audiologist helps patients adjust the devices to obtain satisfactorily amplified sound is called fitting. The amplification can be prescribed using some formulas that convert the

results obtained from hearing testing to the target amplification characteristics [28]. The hearing aids amplification prescription was categorized into linear and non-linear amplification.

Linear amplification has the same gain-frequency curve for all input levels. The linear amplification usually are suitable for severe hearing loss. The NAL-RP (National Acoustic Laboratories of Australia) formula is one of the most famous linear gain prescription formulas [29]. The NAL-RP aims to maximize speech intelligibility at the listening level preferred by the hearing aids user [30]. The NAL-RP formula made two modifications to the previous NAL formulas: a) adding additional gains at low frequencies and reduces high-frequency emphasis; b) supporting severe hearing impairment.

The gain calculations of NAL-RP were conducted as given in the equation below combined with the Table 2-3 and Table 2-4.

$$H_{3FA} = \frac{\left(H_{500} + H_{1000} + H_{2000} \right)}{3} \tag{2.1}$$

$$\begin{cases} X = 0.15H_{3FA} & for \ H_{3FA} \leq 60 \\ X = 0.15H_{3FA} + 0.2\left(H_{3FA} - 60\right) & for \ H_{3FA} \geq 60 \end{cases} \tag{2.2}$$

$$IG_i = X + 0.31H_i + k_i + PC \tag{2.3}$$

, where H_i represents the hearing threshold at frequency i Hz. H_{3FA} is the three-frequency(500Hz, 1000Hz, and 2000Hz) average of the threshold of hearing. IG_i is the insert gain at frequency i Hz. k_i is the correction factor at frequency i Hz and PC is adjustment factor when the hearing threshold at 2 KHz exceeds 90 dB.

12

Table 2-3 NAL k_i and Frequency Relationship

Frequency (Hz)	250	500	1000	2000	3000	4000	6000
k_i (dB)	-17	-8	1	-1	-2	-2	-2

Table 2-4 NAL PC and Frequency Relationship

H_{2kHz}	Frequency (Hz)						
	250	500	1000	2000	3000	4000	6000
90	0	0	0	0	0	0	0
95	4	3	0	-2	-2	-2	-2
100	6	4	0	-3	-3	-3	-3
105	8	5	0	-5	-5	-5	-5
110	11	7	0	-6	-6	-6	-6
115	13	8	0	-8	-8	-8	-8
120	15	9	0	-9	-9	-9	-9

The non-linear prescription can be viewed as specifying the gain-frequency response depending on sound input levels. In other words, this prescription is an input-output curve for several input levels. The non-linear prescription formula usually provides more accurate gains. The Fig 6 procedure is one of the non-linear prescription formula that aims at normalizing the speech loudness, at least for medium and high-level input signals [31]. Fig 6 got its name from Figure 6 of the article in which the underlying data were first presented. This formula was an empirical formula summarized based on a large number of people with similar degrees of threshold loss. According to different speech input levels (40dB SPL, 65 dB SPL, and 90 dB SPL), the FIG 6 formula was demonstrated as following:

For 40 dB SPL input levels:

$$\begin{cases} IG_i = 0 & \text{for } H_i < 20dB \\ IG_i = H_i - 20 & \text{for } 20 \le H_i < 60dB \\ IG_i = 0.5H_i + 10 & \text{for } H_i > 60dB \end{cases} \quad (2.4)$$

13

For 65 dB SPL input levels:

$$\begin{cases} IG_i = 0 & \text{for } H_i < 20dB \\ IG_i = 0.6(H_i - 20) & \text{for } 20 \leq H_i < 60dB \\ IG_i = 0.8H_i - 23 & \text{for } H_i > 60dB \end{cases} \quad (2.5)$$

For 95 dB SPL input levels:

$$\begin{cases} IG_i = 0 & \text{for } H_i \leq 40dB \\ IG_i = 0.1(H_i - 40)^{1.4} & \text{for } H_i > 60dB \end{cases} \quad (2.6)$$

DSL (Desired Sensation Level) is another widely used non-linear fitting prescription formula [32]. DSL aims to normalize the speech loudness similar to normal hearing listeners. The input of this formula is audiogram and the upper limit sound level of hearing-impaired patients. The DSL was formalized as follows:

$$\begin{cases} O = TH_{hi} - \left[(Th_n - SFt) - I \right] & \text{for } I \leq I_{min} \\ O = \dfrac{I + SFt - TH_n}{UL_{hi} - TH_n} \times (UL_{hi} - TH_{hi}) + TH_{hi} & \text{for } I_{min} < I \leq I_{max} \\ O = UL_{hi} & \text{for } I > I_{max} \end{cases} \quad (2.7)$$

where $I_{min} = TH_n - SFt$ and $I_{max} = UL_{hi} - SFt$

Table 2-5 DSL Compensation Parameter

Frequency (Hz)	250	500	1000	2000	3000	4000	6000
SFt	1.0	1.8	2.6	12.0	15.3	14.3	7.3
TH_n	17.5	12.7	9.6	16.0	15.7	13.2	17.8
UL_n	94.0	102.0	99.0	100.0	100.0	98.0	97.0

SFt is the sound field to ear canal transform, TH_n and UL_n are the threshold and upper limit of the normal hearing listeners.

14

2.3.2 Fitting Process

Obtaining the amplification prescription gain for each frequency band is only the first step of the fitting. Since most fitting formulas are empirically obtained, they may not be perfect for every listener. The audiologist needs to adjust the amplification parameters based on their clinical experience and patients' particularity, such as patients' age, sex, hearing loss type and level, etc.

A verification test comes next. The audiologist uses a thin tube to measure the impact of the sound with different loudness levels on patients' eardrum when patients wear hearing aids. By playing several sounds and directly talking to the patients in a different environment, the audiologist gather the feedback of patients and modified the gains until the patient felt the most comfortable.

It still needs to take a few weeks to finalize the gain for patients. The most comfortable setting in the verification test may change. The audiologist advises patients to wear them for just a few hours on the first day. In the following weeks, patients need to increase the length of time to wear their hearing aids until wearing them all day. The audiologist suggests patients wearing hearing aids to adapt to the amplified sound start with quiet environment until they have a conversation with other people. After all those adaptive processes, if patients still feel unconformable, then they need to see the audiologist again to repeat the process. It usually takes about 2-3 months and 3-4 clinical visits before the configuration is finalized.

2.3.3 Self-Fitting Algorithm

The traditional hearing aids fitting process as discussed above is a time-consuming process that heavily rely on professionals, and the outcome may not be optimal to patients as well. More recently, several self-fitting algorithms were proposed to solve this problem. A self-fitting algorithm based on machine learning was proposed in 2015 [33]. The machine learning algorithm involved Gaussian processes and active learning, but this algorithm didn't consider patients history data. In 2017, another self-fitting algorithm based on IEC (Interactive Evolutionary Computation) was proposed [34]. The IEC algorithm is based on genetic algorithms. It is commonly used to find the optimal solution to some artificial intelligence problems, especially in the cases that rely on any subjective feedback.

IEC algorithm uses $M-1$ historical data and one of the new measured gains as input. The $M-1$ historical data can be the actual gain for the patients or obtained from other patients whose information is closest to the current patient. The process contains encoding, fitness calculation, selection, crossover, mutation, and decoding.

Encoding is to convert the gain of each frequency band to a seven-bit binary number. The binary number acts as a gene in the chromosome. The fitness of the chromosome is a calculation based on user's feedback. The feedback is divided into five levels: bad, poor, medium, good and excellent. The corresponding score is:

$$f_n = \frac{1}{\sum_{k=1}^{5} \frac{1}{k^{0.5}} \times n^{0.5}} \tag{2.8}$$

, where n is 1 to 5 corresponding to excellent to poor.

16

The purpose of chromosome selection is to make sure higher fitness chromosome in the current population has higher possibility to be transmitted to the next generation. The roulette selection rule was applied to facilitate that the transmitted probability is proportional to the individual fitness. This rule also makes certain each chromosome has an opportunity to be selected, which is very important to maintain the diversity of the population. The probability is

$$p_i = \frac{f_i}{\sum\limits_{i=1}^{M} f_i} \tag{2.9}$$

, where f_i is the i-th chromosome fitness score.

After we select the chromosome that is transmitted to the next generation, crossover operation starts. The crossover is the exchange of genes in order to obtain two new chromosomes with probability P_c. Before the crossover, a pre-selection chromosome-paring checking needs to be done. This is done to avoid crossover of two similar chromosomes, which may lead to a locally optimized solution. The uncorrelated exponents of chromosomes are defined as:

$$r(x_l, x_m) = \sum_{k=1}^{N} g_{lk} \oplus g_{mk} \tag{2.10}$$

, where g_{lk} and g_{mk} are the gene in the chromosome $x_l = \{g_{l1}, g_{l2}, \cdots, g_{lN}\}$ and $x_m = \{g_{m1}, g_{m2}, \cdots, g_{mN}\}$.

Then the probability $P(y_i \mid x)$ with which the individual y_i is selected to pair with the individual x is described as

17

$$P(y_i \mid x) = \frac{1}{S}\left(1 + \lambda \frac{r(x, y_i) - r_{avg}}{r_{max} - r_{min}}\right) \qquad (2.11)$$

, where $\lambda \in [0,1]$ is set as a constant of 0.8. $r_{avg} = \frac{1}{S}\sum_{i=1}^{S} r(x, y_i)$, r_{min} and r_{max} are the

average, minimum and maximum values of $r(x, y_i)$.

The crossover position is derived in the range of $[n_{min} \; n_{max}]$.

$$\begin{cases} n_{min} = \min\{k \mid g_{lk} \neq g_{mk}, k = 1, 2, ..., N\} \\ n_{max} = \max\{k \mid g_{lk} \neq g_{mk}, k = 1, 2, ..., N\} \end{cases} \qquad (2.12)$$

An adaptive crossover probability was adapted as shown below.

$$P_c = \begin{cases} \zeta \dfrac{P_n^x - P_n^{avg}}{P_n^{max} - P_n^{avg}}, P_n^x > P_n^{avg} \\ P_c^{max} \qquad\qquad , P_n^x \leq P_n^{avg} \end{cases} \qquad (2.13)$$

, where ζ is a constant of $e^{-0.382}$, P_c^{max} is the upper limit of crossover probability, P_n^x is

the larger fitness value of the two individual chromosomes to be crossed. P_n^{max} is the

maximum fitness value of the population and P_n^{avg} is the average fitness value.

Mutation is an operator used to maintain diversity. The mutation probability is given by

$$P_m = \begin{cases} (1-\zeta) \dfrac{P_n^x - P_n^{avg}}{P_n^{max} - P_n^{avg}}, P_n^x > P_n^{avg} \\ P_m^{max} \qquad\qquad , P_n^x \leq P_n^{avg} \end{cases} \qquad (2.14)$$

, where P_m^{max} is the upper limit of the mutation probability.

The self-fitting IEC algorithm demonstrated improved fitting outcome [34]. But it

requires eight historical fitting data points which sometimes may not be accessible easily.

Besides, it requires feedback from a large number of training sets with multiple judgment

criteria (bad - excellent), which may cause users to lose patience during the process. The

18

hearing aids patients, especially older people, prefer procedures that can be done in a short time. The A/B testing method with fewer training sets and only three judgment options is widely used in the clinical field. A modified version of the IEC algorithm is proposed to address those issues.

The modified IEC algorithms starts with a single fitting result. If several amplification gain configurations are available, patients need to select the most comfortable setting with good speech perception. We can assume that this initial gain setting is very close to the optimal fitting. Define the gain step size as m dB and the possible state change for each frequency band is only from seven cases: $-3m$, $-2m$, $-m$, 0, $+m$, $+2m$ and $+3m$ dB. We can simplify the gene size from 7 bits to 3 bits (000_2 and 100_2 both represent 0 dB), and reduce the possible states from 128^n to 7^n if the number of the filter bank is n.

Instead of directly converting the filter bank gain to a gene, the proposed modified IEC used gain variation based on the best historical fitting as a gene. Since the EC algorithm needs to have at least two chromosomes to do the crossover, a random gain variation is generated (within the seven stages) as the initial input along with the best historical fitting gain (no variation). The proposed modified IEC algorithm made the following major changes:

a) Encoding and decoding function rely on the initial gain, and the number of gene representation bits is reduced from 7 to 3.

b) Each time only two sentences are played, and user feedback only has three levels: good, bad and the same.

2.4 Adaptive Differential Microphone Array

Improving the Speech-to-Noise Ratio (SNR) is the most common way to increase intelligibility for hearing impaired patients. For hearing aids user, there are two proven ways that can increase SNR: moving the hearing aids microphone closer to the source and using directional microphones [35]. Moving a microphone closer to the sound source will directly increase the speech level but may not be easy to achieve. Directional microphone now is widely used in commercial hearing aids. In 1984, Hawkins & Yacullo found directional microphone hearing aids can improve 3-4 dB SNR for hearing impaired patients compared to using a single omnidirectional microphone [36]. A directional microphone can be categorized into two types based on the polar pattern: Fix polar pattern and adaptive polar pattern.
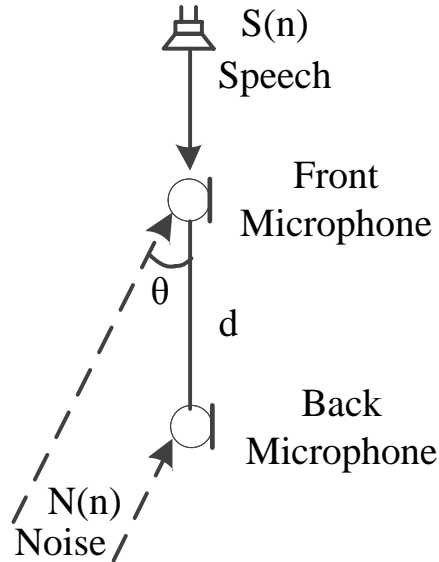
Figure 2-3 Differential microphone array structure

The most common configuration of microphones in hearing aids is a single directional microphone or dual omnidirectional microphones with fixed polar patterns because of the

limit of hardware size, computational resources, and power consumption [37]. The fixed

polar pattern may lead to an issue that only noise from a specific direction can be

suppressed. The adaptive differential microphone array structure was proposed in Figure

2-3. The two microphones are defined as a front microphone and a back microphone

based on the distance to the speech signal. Since in hearing aids applications, we can

always assume that the speech comes from the front of the hearing aids user. The

directions of noise signal may vary. In this case, we assume that the angle of noise can

only change from 90° to 270°.

Suppose the signal picked up from the two microphones are:

$$x_f(n) = S(n) + N_f(n) \tag{2.15}$$

$$x_b(n) = S(n-r) + N_b(n) \tag{2.16}$$

, where $x_f(n)$ and $x_b(n)$ are the signal received at the front and back microphone

respectively. $S(n)$ is the expected speech signal. $N_f(n)$ and $N_b(n)$ are the noise signal

received at the front and the back microphone respectively. As shown in Figure 2-3, r is

the sample phase difference, and can be represented as

$$r = \frac{d}{c} \cdot f_s \tag{2.17}$$

, where d is the distance between two microphones, c is the speed of sound in air, and

f_s is the sampling frequency.

The difference between $N_f(n)$ and $N_b(n)$ is described as:

$$N_f(n) = N_b\left(n - \frac{d\cos\theta}{c} f_s\right) \tag{2.18}$$

21

Combine Equation (2.17) and (2.18), Equation (2.15) and (2.16) can be revised as:

$$x_f(n) = S(n) + N_b\left(n - \frac{dcos\theta}{c}f_s\right) \qquad (2.19)$$

$$x_b(n) = S\left(n - \frac{d}{c} \cdot f_s\right) + N_b(n) \qquad (2.20)$$

The first-order differential microphone array scheme is illustrate as Figure 2-4.

$$x(n) = x_f(n) - x_b(n - D_s) \qquad (2.21)$$

$$y(n) = x_f(n - D_s) - x_b(n) \qquad (2.22)$$

$$z(n) = x(n) - \beta(n) \cdot y(n) \qquad (2.23)$$

, where D_s is an additional delay sample. $D_s = r = \frac{d}{c} \cdot f_s$. Equation (2.21), (2.22) can be

expanded as:

$$
\begin{aligned}
x(n) &= S(n) + N_b\left(n - \frac{dcos\theta}{c} \cdot f_s\right) - \left(S\left(n - \frac{2d}{c} \cdot f_s\right) + N_b\left(n - \frac{d}{c} \cdot f_s\right)\right) \\
&= S(n) - S\left(n - \frac{2d}{c} \cdot f_s\right) + N_b\left(n - \frac{dcos\theta}{c}f_s\right) - N_b\left(n - \frac{d}{c} \cdot f_s\right)
\end{aligned} \qquad (2.24)
$$

$$
\begin{aligned}
y(n) &= S\left(n - \frac{d}{c} \cdot f_s\right) + N_b\left(n - \frac{dcos\theta}{c}f_s - \frac{d}{c} \cdot f_s\right) - \left(S\left(n - \frac{d}{c} \cdot f_s\right) + N_b(n)\right) \\
&= N_b\left(n - \frac{dcos\theta}{c}f_s - \frac{d}{c} \cdot f_s\right) - N_b(n)
\end{aligned} \qquad (2.25)
$$
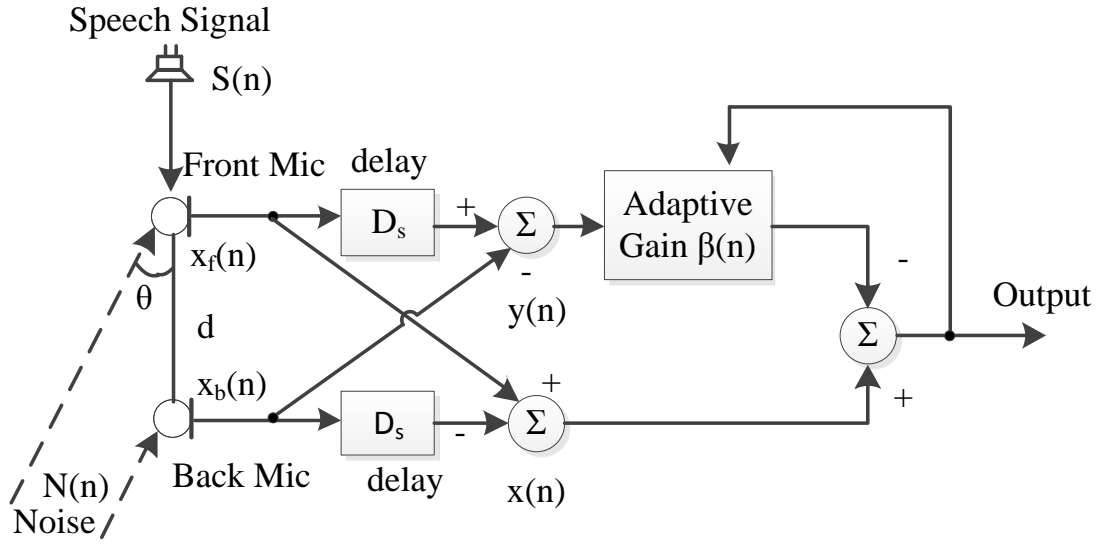
Figure 2-4 First-Order Differential Microphone

From Equation (2.25), $y(n)$ only contains noise information. Plus, the speech signal and noise signal are not correlated. To maximize the SNR is equivalent to minimizing the power of $z(n)$ as presented in Equation (2.26).

$$
\begin{aligned}
E\left[z^2(n)\right] &= E\left[\left(x(n) - \beta(n) \cdot y(n)\right)^2\right] \\
&= E\left[x^2(n) - 2\beta(n) \cdot x(n)y(n) + y^2(n)\right] \\
&= R_{xx} - 2\beta(n)R_{xy} + \beta^2(n)R_{yy}
\end{aligned}
\tag{2.26}
$$

Since Equation (2.26) is a quadratic function, there certainly is a unique β to minimize the power of the output[38].

$$
\beta_{opt} = \frac{R_{xy}}{R_{yy}}
\tag{2.27}
$$

Instead of using a fixed-polar-pattern microphone array, the adaptive microphone array can reduce noise from a broader direction. From Equation (2.24), $x(n)$ equals the highpass filtered speech signal, and the low-frequency information is suppressed in the

23

final output $z(n)$ [39][40]. A lowpass filter was designed to compensate for the energy loss of $z(n)$. The cutoff frequency of the compensator filter is

$$\omega_c = \frac{\pi}{2T} = \frac{\pi c}{2d} \tag{2.28}$$

And the ideal compensation filter is

$$H_c(\omega) = \begin{cases} \dfrac{1}{2\sin\dfrac{\pi \cdot \omega}{c \cdot \omega_c}} & 0 < \omega \leq \omega_c \\ \dfrac{1}{2} & otherwise \end{cases} \tag{2.29}$$

2.5 Filter Banks

2.5.1 Basic

In the past two decades, the research in filter banks and subband processing has been advancing at a fast pace. Subband processing of signals is widely used in applications such as telecommunication, audio, and image processing [41]-[43]. Compared to the equivalent full band processing, subband signal processing often leads to a performance increase due to the frequency band decomposition and potentially a reduction in computation due to the lower sampling rate of the subband signals [44][45].

24
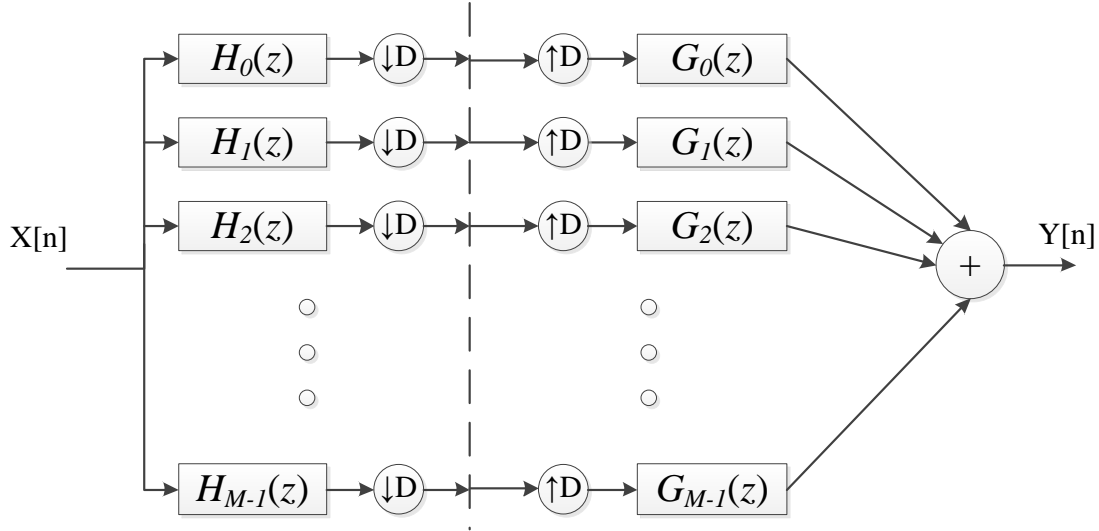
Figure 2-5 General filter bank structure

Figure 2-5 is a basic structure of the filter bank. Filter banks contain an analysis filter

bank and a synthesis filter bank [46]. The analysis filter bank decomposes a signal $x[n]$

by partitioning its spectrum into different frequency bands using a set of bandpass filters.

$$X_i^a[z] = X[z]H_i[z] \tag{2.30}$$

, where $X[z]$ is the frequency response of input signal $x[n]$, $H_i[z]$ is the frequency

response of each bandpass filter $h_i[n]$. $X_i^a[z]$ is the resulted signal after the input signal

passes i-th band filter.

Each subband signal is downsampled by a factor of D due to their reduced bandwidth.

This operation is called decimation. The time-domain operation is expressed as

following:

$$x_i^d[n] = x_i^a[Dn] \tag{2.31}$$

In the frequency domain, the relationship can be expressed as:

25

$$X_i^d[z] = \sum_{n=-\infty}^{\infty} x_i^d[n]z^{-n} = \sum_{k=-\infty}^{\infty} x_i^d\left[\frac{k}{D}\right]z^{-k/D}$$

$$= \sum_{k=-\infty}^{\infty} x_i^a[k]z^{-k/D} \tag{2.32}$$

$$= X_i^a[z^{1/D}]$$

The subband signals are synthesized into full band signal by first upsampling to the original sampling rate. This operation is also called expansion.

The time-domain operation can be expressed as following:

$$x_i^e[n] = x_i^d\left[\frac{n}{D}\right] \tag{2.33}$$

In the frequency domain, the relationship can be represented by:

$$X_i^e[z] = \sum_{n=-\infty}^{\infty} x_i^e[n]z^{-n} = \sum_{k=-\infty}^{\infty} x_i^e[kD]z^{-kD}$$

$$= \sum_{k=-\infty}^{\infty} x_i^d[k]z^{-kD} \tag{2.34}$$

$$= X_i^d[z^D]$$

And then the upsampled signal will be pass through a bandpass filter to remove spectral images introduced by upsampling. The final output is obtained by:

$$Y[z] = \frac{1}{D}\sum_{i=0}^{D-1} X_i^e[z]G_i[z] \tag{2.35}$$

The computational complexity of the filter bank implementation is

$$O(2D \cdot N\log(M) + 2DM) = O(n^2\log n) \tag{2.36}$$

The general filter bank structure's computation complexity limits its application to the embedded system. A polyphase representation helps to increase the computational efficiency. The term polyphase refers to the decimation of a signal- if it is appropriately bandlimited and can be decimated by a factor M.

26

The analysis filterbank transfer function $H(z)$ can be decomposed into the forms:

$$H(z) = \sum_{n=-\infty}^{\infty} h(nM)z^{-nM} + z^{-1}\sum_{n=-\infty}^{\infty} h(nM+1)z^{-nM} + \cdots + z^{-(M-1)}\sum_{n=-\infty}^{\infty} h(nM+M-1)z^{-nM}$$

(2.37)

Equation (2.37) can be simplified as:

$$H(z) = \sum_{k=0}^{M-1} z^{-k} E_{k,l}(z^M)$$
(2.38)

where

$$E_{k,l}(z) = \sum_{l=-\infty}^{\infty} h(Ml+k)z^{-l}, \quad 0 \le k \le M-1$$
(2.39)

Equation (2.38) is rewritten in a matrix format as:

$$
\begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix}
=
\begin{bmatrix}
E_{0,0}(z^M) & E_{0,1}(z^M) & \cdots & E_{0,M-1}(z^M) \\
E_{1,0}(z^M) & E_{1,1}(z^M) & \cdots & E_{1,M-1}(z^M) \\
\vdots & \vdots & \ddots & \vdots \\
E_{M-1,0}(z^M) & E_{M-1,1}(z^M) & \cdots & E_{M-1,M-1}(z^M)
\end{bmatrix}
\begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(M-1)} \end{bmatrix}
\quad (2.40)
$$

Similarly, synthesis filter bank is expressed as:

$$\begin{bmatrix} G_0(z) & G_1(z) & \cdots & G_{M-1}(z) \end{bmatrix} =$$

$$
\begin{bmatrix} z^{-(M-1)} & z^{-(M-2)} & \cdots & 1 \end{bmatrix}
\begin{bmatrix}
R_{0,0}(z^M) & R_{0,1}(z^M) & \cdots & R_{0,M-1}(z^M) \\
R_{1,0}(z^M) & R_{1,1}(z^M) & \cdots & R_{1,M-1}(z^M) \\
\vdots & \vdots & \ddots & \vdots \\
R_{M-1,0}(z^M) & R_{M-1,1}(z^M) & \cdots & R_{M-1,M-1}(z^M)
\end{bmatrix}
\quad (2.41)
$$

where

$$R_{k,l}(z) = \sum_{l=-\infty}^{\infty} g(Ml+k)z^{-l}, \quad 0 \le k \le M-1$$
(2.42)

Figure 2-6 and Figure 2-7 show the polyphase representation of the analysis filter and the synthesis filter bank corresponding to Equation (2.40) and (2.41).

27
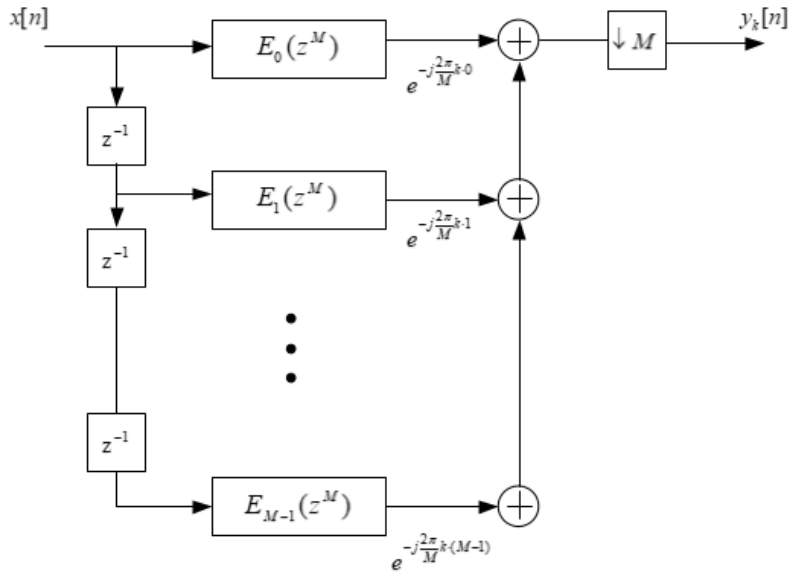
Figure 2-6 Polyphase representation of analysis filter banks
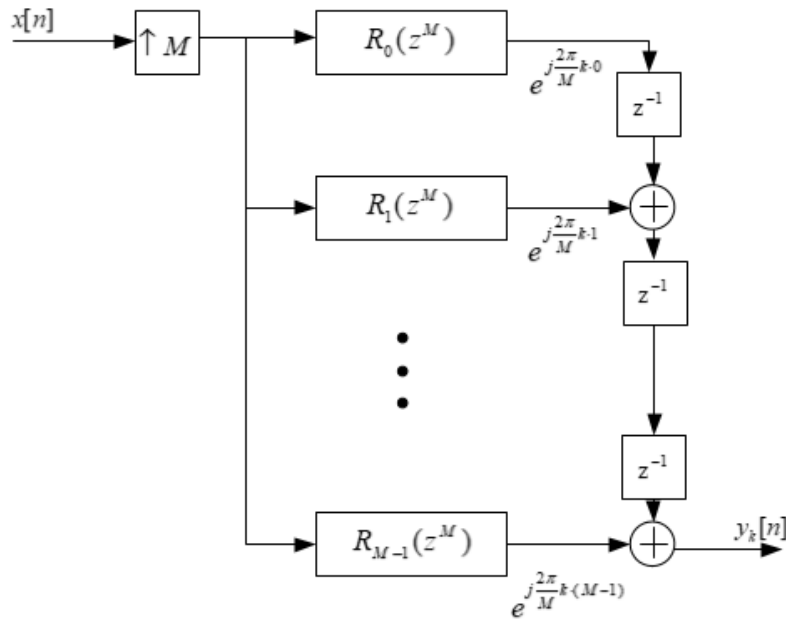
Figure 2-7 Polyphase representation of synthesis filter banks

Although the polyphase representation can reduce the computation complexity, it still contains redundant downsampling computation. It has been proven that downsampling the signal after the filter is equal to downsample the input signal before the filtering [46]. By moving the order of the downsampling operation, some redundant filter computations

can be further reduced. A more efficient filter bank structure is illustrated in Figure 2-8

and Figure 2-9. Compared to the original filter banks diagrams show in Figure 2-5, the

efficient polyphase structure can achieve a complexity reduction of a factor of M^2.
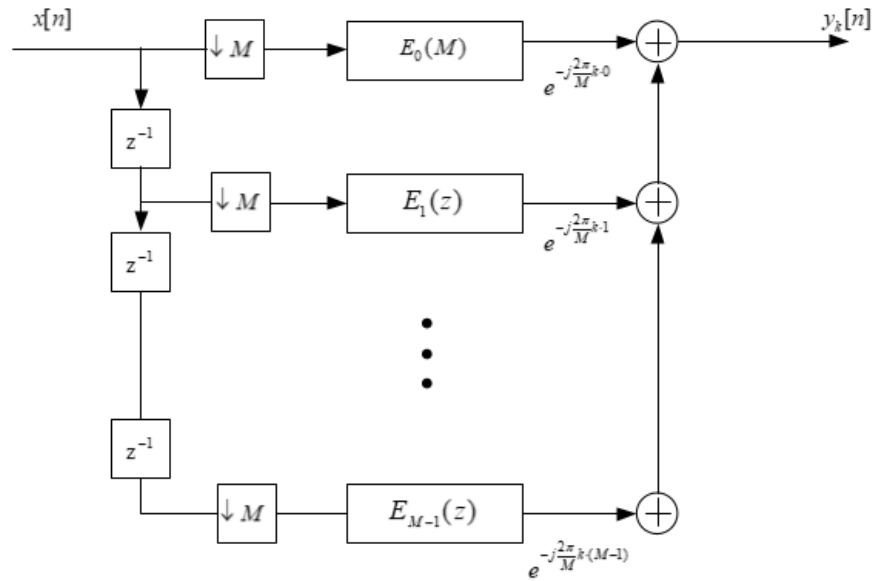
Figure 2-8 Efficient polyphase representation of analysis filter bank
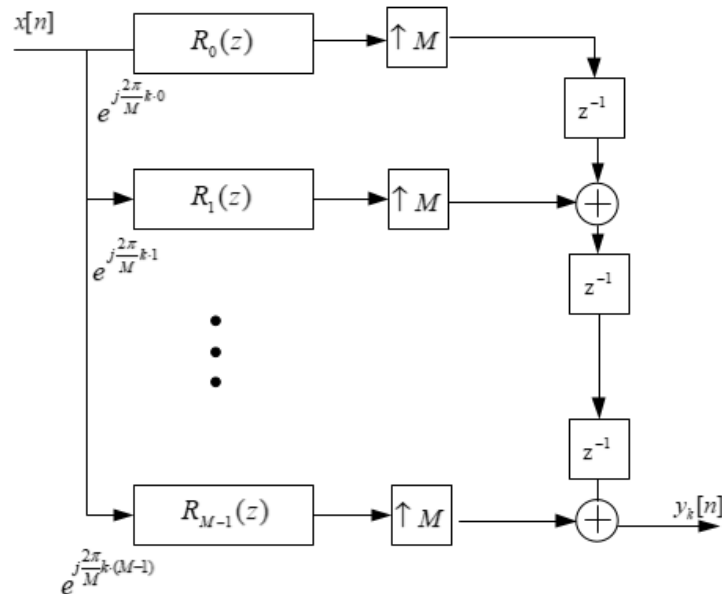
Figure 2-9 Efficient polyphase representation of synthesis filter bank

29

When the analysis filter bank and the synthesis filter bank are connected in series, the output of the filter bank is only a delayed version of the input signal, this kind of system is called a perfect reconstruction (PR) system.

$$y[n] = cx[n-n_0] \tag{2.43}$$

In other words, a PR system introduces no signal distortion except a phase delay. In general, the necessary and sufficient conditions for perfect reconstruction [46] are

$$\mathbf{R}(z)\mathbf{E}(z) = cz^{-m_0} \begin{bmatrix} 0 & \mathbf{I}_{M-r} \\ z^{-1}\mathbf{I}_r & 0 \end{bmatrix} \tag{2.44}$$

where $0 \le r \le M-1$, m_0 is an integer and c is a non-zero constant.

The general form of $\mathbf{R}(z)\mathbf{E}(z)$ can be regarded as a square root of a vector unit delay since

$$\begin{bmatrix} 0 & \mathbf{I}_{M-r} \\ z^{-1}\mathbf{I}_r & 0 \end{bmatrix}^2 = z^{-1}\mathbf{I}_M \tag{2.45}$$

Thus, the general case is approximated as

$$\mathbf{R}(z)\mathbf{E}(z) = cz^{-m_0}\mathbf{I}_M \tag{2.46}$$

except for some channel swapping and additional samples of lagging.

If the number of upsampling or downsampling coefficients equals the number of filter bank channels, it is called critically sampled filter bank or maximally decimated filter bank. If the number of upsampling or downsampling coefficients is smaller than the number of filter bank channels, it is called oversampled filter banks. The main advantage for oversampled filter banks over critically filters banks is that oversampling provides design freedom in obtaining perfect reconstruction, better noise immunity in each subband but it will also increase redundancy and decrease the computation efficiency.

2.5.2 DFT Filter Banks

Designing several different filters to meet M different transfer function that provides perfect reconstruction is computationally expensive and technically challenging as well [47]. Instead, a DFT filter bank is used that employs a single low-pass filter prototype to derive all the other filters across all bands. According to the frequency shifting property of Laplace transform, the M analysis filters can be derived from a low-pass prototype filter by applying a frequency-shifting.

$$H(e^{j(\Omega-\Omega_0)}) = \mathcal{L}(h(n) \cdot e^{jn\Omega_0}) \tag{2.47}$$

For M - channel filter bank, the M center-frequencies of the analysis filter are determined by equally dividing up the normalized frequency domain between 0 and 2π. Figure 2-10 and Figure 2-11 show an example of frequency shifting.
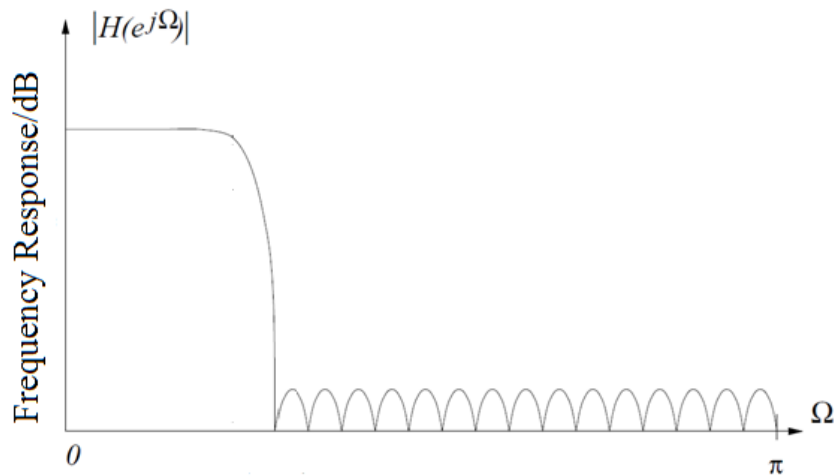
Figure 2-10 Prototype filter design
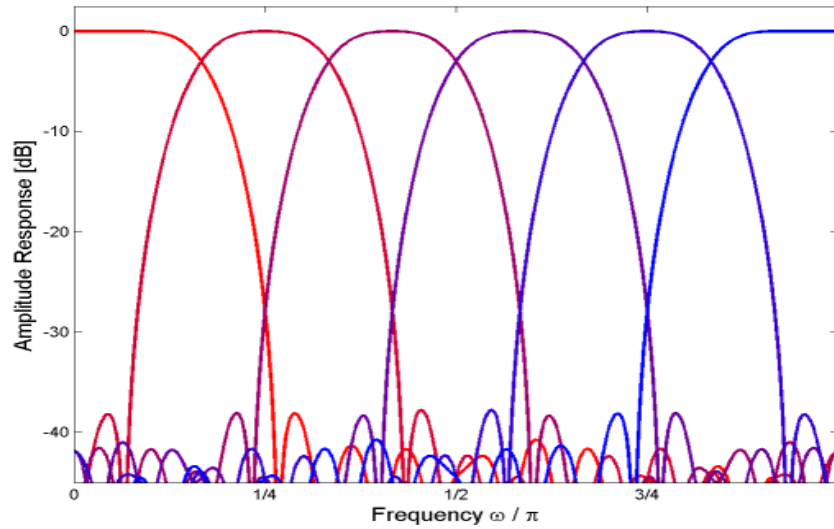
Figure 2-11 Frequency shifting of prototype filter

The frequency response of the M analysis filter is obtained from the Equation (2.47), with $\Omega = i2\pi/M$

$$H_i(e^{j\Omega}) = H(e^{j(\Omega - 2\pi i/M)}) \tag{2.48}$$

The synthesis filter also can be obtained by DFT modulation of an FIR prototype lowpass filter $g(k)$. In the proposed work, we are using the same prototype for both analysis filter bank and synthesis filter bank.

2.5.3 Oversampled GDFT Filter Banks

Oversampled generalized DFT (GDFT) filter banks are an important family of filter banks due to that FFT can further increase computational efficiency [48]-[51].

The GDFT filter bank analysis filter $h_k[n]$ and synthesis filter $g_k[n]$ can be derived from a real-valued prototype lowpass filter $p[n]$ of even length L_p that is modulated by a generalized DFT.

32

$$h_k[n] = v_{k,n} \cdot p[n] \qquad (2.49)$$

$$g_k[n] = h_k[L_p - n + 1] = v_{k,L_p-n+1} \cdot p[L_p - n + 1] \qquad (2.50)$$

$$v_{k,n} = e^{j\frac{2\pi}{K}(k+k_0)(n+n_0)} \qquad (2.51)$$

The prototype design is crucial for the filter bank. The prototype design here is based on iterative least squares design. This method minimizes both stopband energy and power complementarity [48].

The standard method to determine stopband energy is

$$E_s = \frac{1}{\pi} \int_{\pi/N}^{\pi} \left| P(e^{j\Omega}) \right|^2 d\Omega \qquad (2.52)$$

where $P(e^{j\Omega})$ is frequency response of the stopband power. The real part of the power can be express as follows:

$$\mathrm{Re}\{P(e^{j\Omega_i})\} = \mathbf{T}_{DCT} \cdot \mathbf{p} \qquad (2.53)$$

$\mathbf{T}_{DCT}$ is the DCT transform, while $\mathbf{p}$ here is the prototype filter,

$$\mathbf{p} = [p[0], p[1], \cdots, p[L_p - 1]]^T \qquad (2.54)$$

An approximation of the stopband energy is now given by

$$E_s \approx \frac{1}{\pi} \left\| \mathbf{T}_{DCT} \cdot \mathbf{p} \right\|_2^2 \qquad (2.55)$$

The error of stopband can be expressed as the following:

$$\begin{aligned}
\xi_e &= \left\| \mathbf{T}_{DCT} \cdot \mathbf{p} \right\|_2^2 \\
&= \left\| \begin{bmatrix} 1 & \cos(\omega_0 \cdot 1) & \cdots & \cos(\omega_0 \cdot (L_p - 1)) \\ 1 & \cos(\omega_1 \cdot 1) & \cdots & \cos(\omega_1 \cdot (L_p - 1)) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos(\omega_N \cdot 1) & \cdots & \cos(\omega_N \cdot (L_p - 1)) \end{bmatrix} \begin{bmatrix} p[0] \\ p[1] \\ \vdots \\ p[L_p - 1] \end{bmatrix} \right\|_2^2
\end{aligned} \qquad (2.56)$$

33

Denote the impulse response of the overall filter bank in the time domain as $t(k)$, which can be written as a convolution of the analysis and the synthesis filters

$$t(k) = \sum_{i=0}^{M-1} \frac{1}{S_k} \left(h_k[i] * g_k[i] \right) \tag{2.57}$$

, where the factor $\dfrac{1}{S_k}$ takes into account of the power loss in the channels due to decimation. Since both analysis filter and synthesis filter are both derived from the prototype filter $\mathbf{p}$, the impulse response of filter bank $\mathbf{t}$ can be rewrite as:

$$\mathbf{t} = \mathbf{V} \cdot \mathbf{S}_2 \cdot \mathbf{p} \tag{2.58}$$

, where $\mathbf{V}$ is a convolutional matrix to holds k th polyphase components of prototype filter $p[n]$ on k th row, and matrix $\mathbf{S}_2$ is $L_p \times L_p$ mapping matrix to covert synchronously order prototype filter into polyphase vector order.

To minimize the reconstruction error, the distance between the impulse response t and a desired response must reach the least value. The reconstruction error is derived as

$$\xi_r = \|t - v\|_2^2 = \|\mathbf{V} \cdot \mathbf{S}_2 \cdot \mathbf{p} - v\|_2^2 \tag{2.59}$$

Overall, the total error contains the reconstruction error ξ_r and the stopband energy error ξ_r. In order to obtain an optimal filter bank, the total error needs to be minimized as shown in Equation (2.60).

$$\begin{aligned}
\xi &= \xi_e + \lambda \cdot \xi_r \\
&= \|\mathbf{T}_{DCT} \cdot \mathbf{p}\|_2^2 + \lambda \cdot \|\mathbf{V} \cdot \mathbf{S}_2 \cdot \mathbf{p} - v\|_2^2 \\
&= \left\| \begin{bmatrix} \mathbf{T}_{DCT} \\ \mathbf{V} \cdot \mathbf{S}_2 \end{bmatrix} \mathbf{p} - \begin{bmatrix} 0 \\ v \end{bmatrix} \right\|_2^2
\end{aligned} \tag{2.60}$$

34

2.6 Multichannel Dynamic Range Compression

Dynamic range compression is frequently used in audio processing. The general functionality of compression is to map the dynamic range of the input signal into the target dynamic range for the output signal. Hearing aids wearers usually need the compression function to decrease the dynamic range of the sound in the environment in order to match the dynamic range of their auditory system [53]. In other words, loud sound needs to be amplified less than quite sound, and the gain needs to be varied to decrease the dynamic range. This process is called dynamic-range compression. The auditory system of normal-hearing listeners has a wide dynamic range, which means they can hear sound with intensities ranging from soft to loud. But for hearing-impaired listeners, they usually cannot perceive soft sound, and the dynamic range of the input signal needs to be reduced [53].
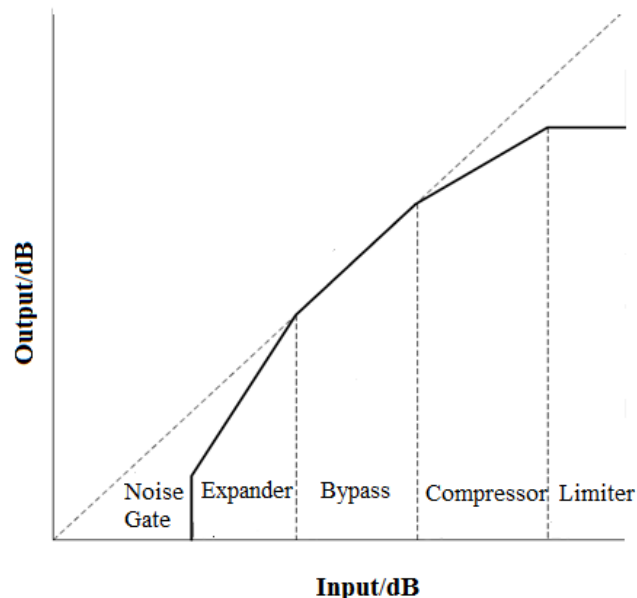
Figure 2-12 Input and output relationship of dynamic compression

Figure 2-12 demonstrates a general relationship between the dynamic range from the input signal and that from the output signal. Compression usually consists of four components: expander, bypass, compression, and limiter. The purpose of the expander is to prevent hearing aids from amplifying soft sound that hearing aids wearers are not interested in. The compressor is to compress the dynamic range of the sound into a limited dynamic range of the auditory system of hearing aids users. The limiter is applied to avoid the potential risk that the output of hearing aids exceeds the comfortable hearing level of patients [54].

For each of the four stages, it has several static characteristics (threshold, compression ratio) and dynamic characteristic (attack/release time). Using compressor as an example, the compression threshold (CT) is the sound level at which hearing aids start compressing. In practice, measurement standards define the compression threshold as the point at which the output deviates by 2 dB from the output where linear amplification has occurred. Noise threshold (NT), expander threshold (ET) and limiter threshold (LT) all have a similar definition.

The compression ratio (CR) describes how much the output intensity increases when the input level increases.

$$CR = \frac{\Delta I}{\Delta O} \tag{2.61}$$

For the compressor, the compression ratio is larger than 1. For hearing aids, the compression ratio is commonly in the range of 1.5-3. For expander, the compression ratio is less than 1, which means every 1 dB increase in the input level results in less than 1 dB increase in the output level. By that way, the level of background noise will be reduced at the same time. For bypass, the compression is 1, and the input and output have the same

36

level. From the Equation (2.61), if the slope of the limiter equals to 0, the compression in the limiter will go to infinite. In practice, the compression ratio of the limiter is higher than 10.

Compression is a dynamic process: it changes the gain depending on the changes in the input sound level. From the perspective of hearing aid signal processing, it needs to amplify the sound if it is soft and suppresses the sound if it is too loud. When the incoming signal level changes from below the compression threshold to above it, the compressor is unable to change the gain instantaneously. The time taken for the compressor to compress the increased sound level is referred to as the attack time (AT). The attack time is defined as the time taken for the output level to stabilize to within 2-3 dB of its final level after the level of the input to the hearing aids increases from 55 to 90 dB SPL.

Release time (RT) is the delay that the compressor amplifies the soft sound with appropriate gain to increase to the higher sound level. Release time is defined as the interval between the instant of the abrupt drop in the input level from 90 dB to 55 dB SPL and the instant that the output level has stabilized within 4 dB of the final value.

The attack and release time have a significant effect on how the compressor affects the levels of different syllables of speech. AT and RT setting varies depending on hearing aids user preference. The shorter the RT, the more distorted the output. If the attack time is too short, it may amplify the sound to a level louder than desired. The attack time usually is smaller than 10 ms, and the release time generally is longer than 20 ms.

This section presents a multichannel dynamic range compression algorithm. This algorithm takes the output of the filter bank as input and can be applied to the different

frequency band. It has been shown that using multichannel compression can increase the intelligibility [55].
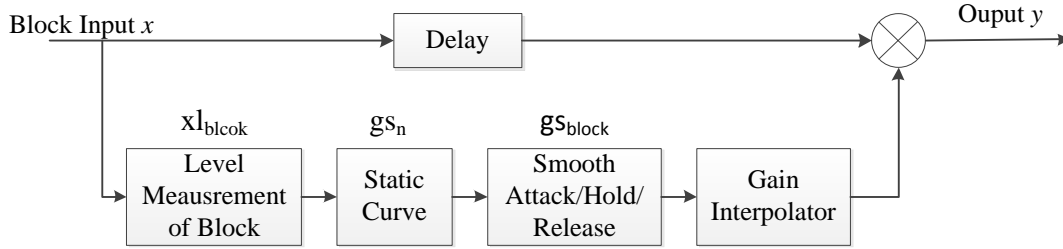
Figure 2-13 Scheme of signal channel dynamic range compression

Figure 2-13 shows the scheme of single-channel dynamic range compression. Multichannel dynamic compression needs to be applied to individual single-channel output. The algorithm starts with measuring the root-mean-square (RMS).

$$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} x^2(i)} \qquad (2.62)$$

, where N represents the length of each frame and $x(i)$ shows the input sample sound level. This result needs to be averaged with a first-order low-pass filter. The coefficient is

$$\text{TAV} = 1 - e^{\frac{-2.2T_s \cdot 1000}{t_M}} \qquad (2.63)$$

, where t_M is the average time in milliseconds, T_s is the sampling period. The RMS of the sound level at each frame is obtained by:

$$xl_{block} = (1 - \text{TAV}) \cdot xl_{block-1} + \text{TAV} \cdot x_{rms} \qquad (2.64)$$

After Equation (2.64), the input signal is switched to dB SPL using

$$xl_{dB} = 20\log_{10}(xl_{block}) \qquad (2.65)$$

38
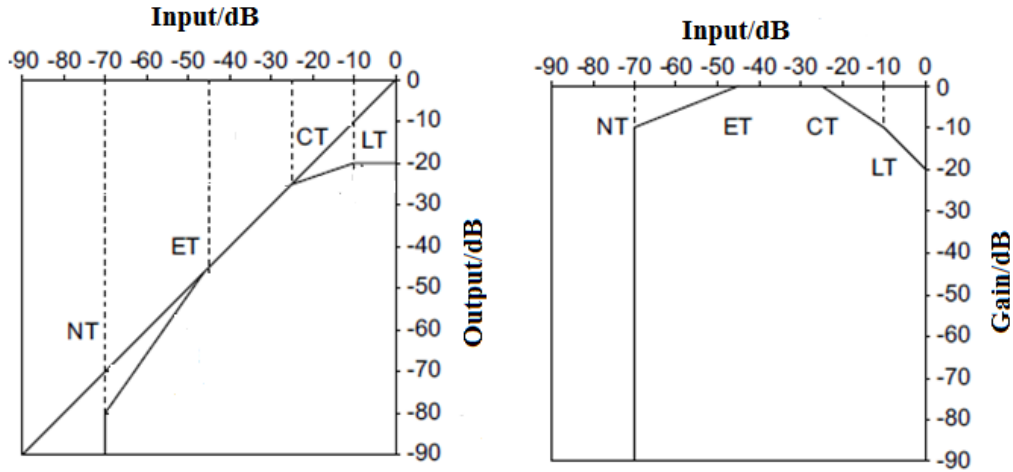
Figure 2-14 I/O Static curve and Gain static relationship

The I/O relationship shown in the left panel of Figure 2-14 cannot be directly applied in our application. One more step needs to be done to convert the y-axis to gain (output - input) as shown in the right panel of Figure 2-14. The gain relationship is stored as a lookup table to improve efficiency. This table may vary depending on resolution and fitting map. The lookup table is defined as function $\mathcal{L}$

$$gs_n = \mathcal{L}\left(xl_{dB}\right) \tag{2.66}$$

The gain gs_n needs to be smoothed and interpolated by the attack and release time.

$$gs_{block} = \left(1-k\right) \cdot gs_{block-1} + k \cdot gs_n \tag{2.67}$$

$$k = \begin{cases} 1 - e^{\frac{-2.2T_s \cdot 1000}{t_a}} & if \ gs_{block} < gs_{block-1} \\ 1 - e^{\frac{-2.2T_s \cdot 1000}{t_r}} & if \ gs_{block} \geq gs_{block-1} \end{cases} \tag{2.68}$$

, where t_a and t_r is the attack time and release time in milliseconds respectively. Once the gain gs_{block} for the frame is obtained, it needs to be converted back to the linear scale and applied to the input vector to obtain the output vector.

Chapter 3

Open-Source Self-Fitting Hearing Aids System Implementation

3.1 Hearing Assessment Implementation

3.1.1 Introduction

Hearing assessment is conducted in order to know a patient's hearing loss level across different frequency bands. Hearing assessment mostly occurs in a sound booth to keep out the environment noises. Patients need to wear a headphone with wires connected to a calibrated instrument called audiometer which is operated by an audiologist to conduct the testing. This type of pure-tone audiometry requires access to specialized medical equipment and audiologists. Although a hearing assessment conducted by a professional staff with calibrated equipment yields an accurate audiogram, it also raises cost and inconvenience. The related cost of the professional performing the procedures, time, and necessary equipment makes it inaccessible to the people who live far away from the urban testing center that evaluates and diagnoses hearing loss.

There are several studies that developed computer-based or smartphone-based pure-tone audiometer [56]-[58]. This type of testing system overcomes some of the difficulties mentioned above, but still, their results showed at least 10 dB of differences compared to the audiogram obtained by an audiologist. Recently Kelly and her colleagues [59] showed that one particular tablet-based screening app could yield an accuracy of 92% of the obtained thresholds falling within 10 dB of the clinical audiograms, but it only works on iPhone and with particular headsets.

In this section, we will propose a web-based hearing assessment system with a calibrated USB audio card. This design is an inexpensive but accurate solution that can solve the inconvenience of a hearing assessment conducted by a professional audiologist.

3.1.2 Hardware Design

Computers and internet access now are widely available across the world; a web-based hearing assessment could pose no limitations or restrictions for everyone. A calibrated USB audio card combined with a web-based audiometry service will be an optimal solution for hearing assessment.

The hearing assessment hardware is based on the TI PCM2912 IC [61]. This IC is a USB stereo audio codec with USB 2.0 full-speed protocol controller. The output dynamic range of PCM2912 is 90 dB. The schematic and the picture of USB audio cards are shown in Figure 3-1 and Figure 3-2 below.
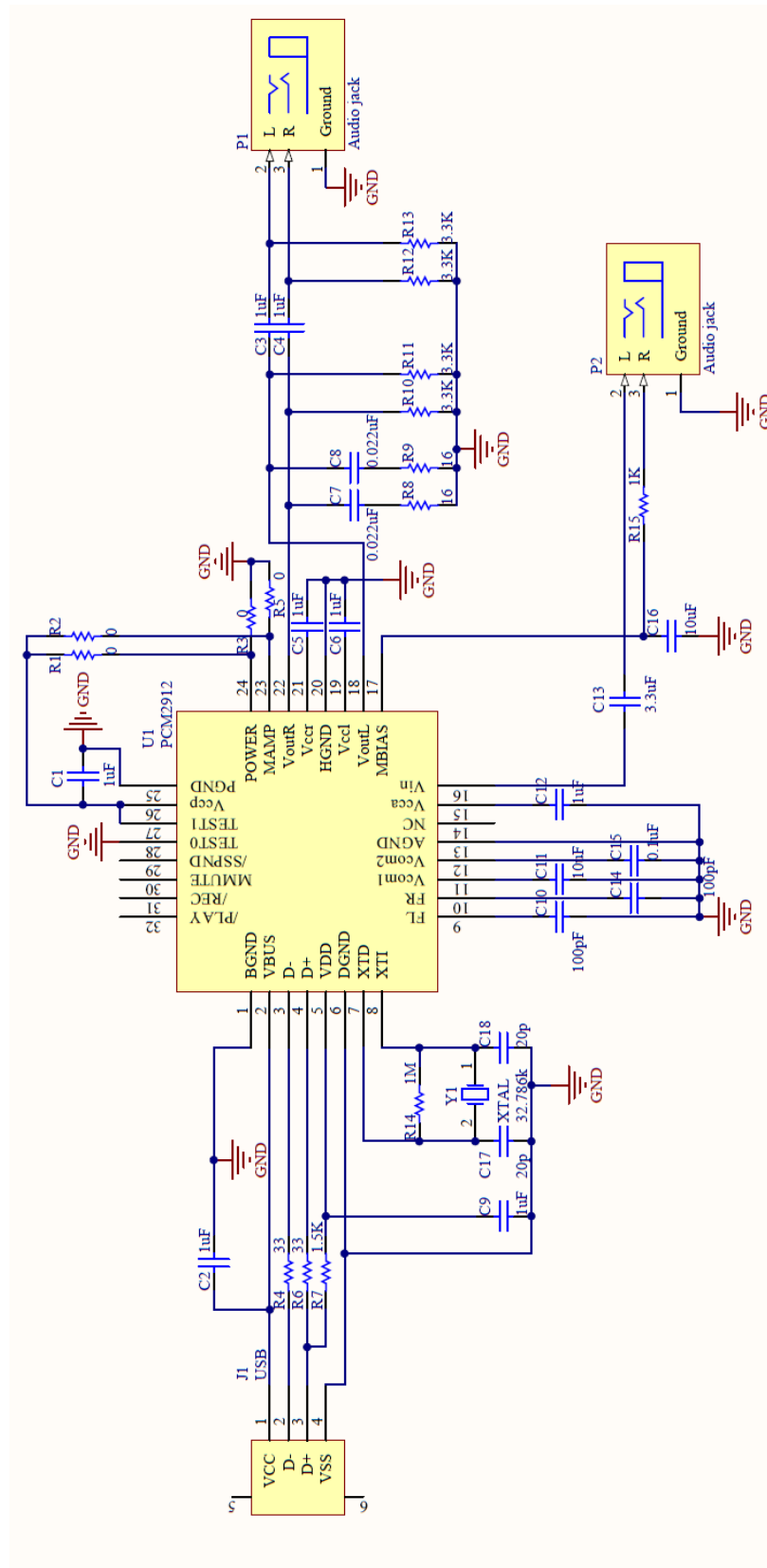
Figure 3-1 USB Audio card schematic
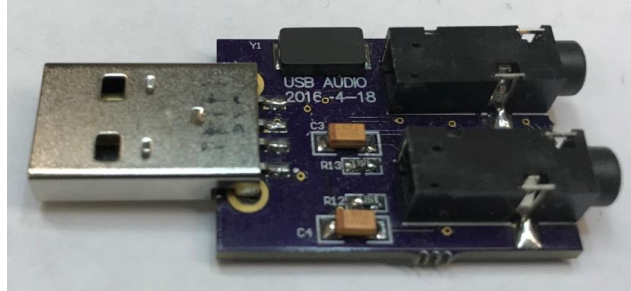
Figure 3-2 Picture of USB audio card

3.1.3 Software Design

The web-based audiometry service was implemented in Javascript. The hearing

assessment software consists of mainly three functionalities: regular hearing testing,

high-frequency hearing testing, and historical testing results. In the regular hearing

testing, a standard hearing assessment is conducted at pure-tone frequency of 250, 500,

1000, 2000, 3000, 4000, 6000 and 8000 Hz. High-frequency hearing assessment further

covers the pure-tone frequencies of 8000, 9000, 10,000, 12,000, 12,500, 14,000, 16,000

Hz. The testing result usually has a positive correlation with age and hearing condition so

that it can be an indicator that could point to potential hearing problems [60]. The web

server keeps records of all the historical testing results and patients' audio card

calibration information to help the users track their hearing conditions. Figure 3-3 and

Figure 3-4 show the hearing assessment user interface of the regular testing and the high-

frequency testing.
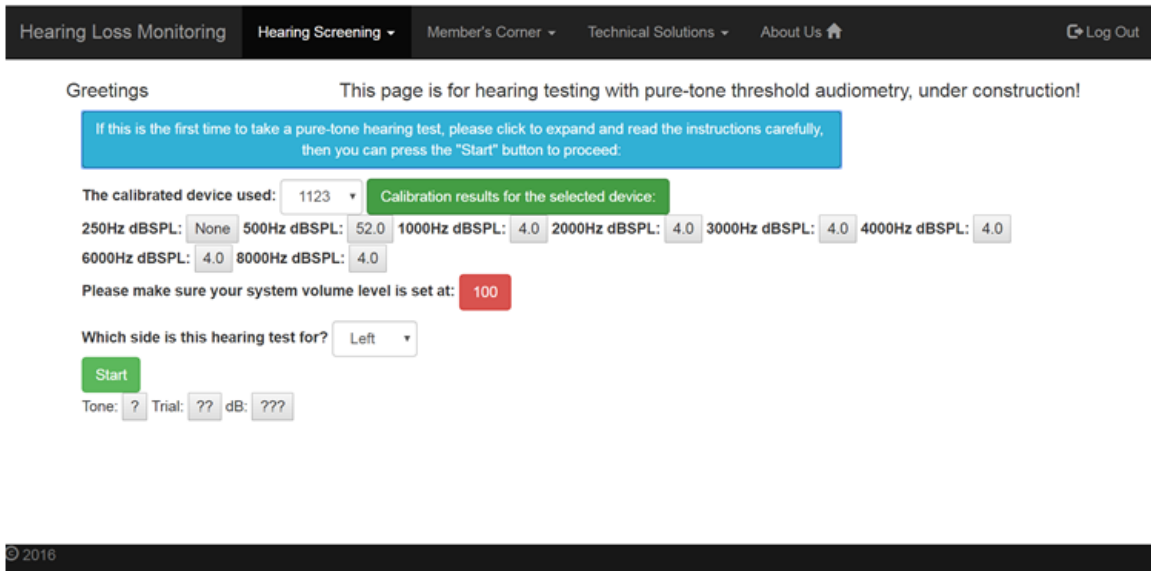
Figure 3-3 Hearing assessment regular testing user interface
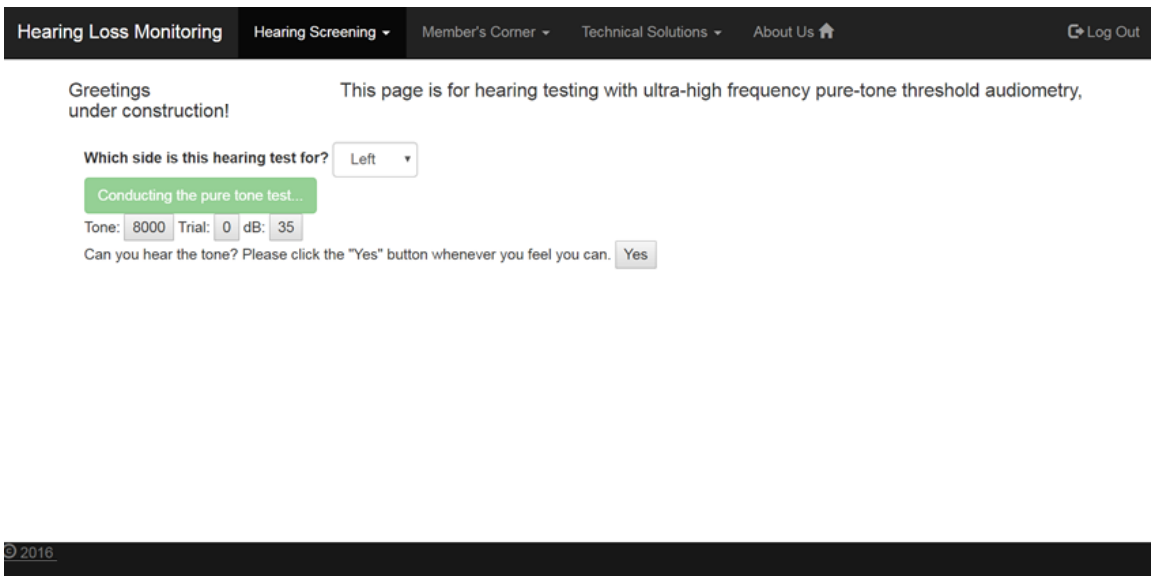
Figure 3-4 Hearing assessment high-frequency testing user interface

3.2 Fitting Software Implementation

The hearing-aid fitting involves calculating the amplification gain of each band. The

fitting software was implemented in MATLAB 2016b. The overall flowchart is shown as

Figure 3-5. As discussed in section 2.3, The three existing fitting methods NAL-NL,

FIG6 and DSL help to generate an initial fitting setting based on hearing assessment result. If the initial setting is already good enough, the fitting process will stop and select the best one as the hearing aids setting. Otherwise, patients will use self-fitting to obtain the fitting gain by themselves.
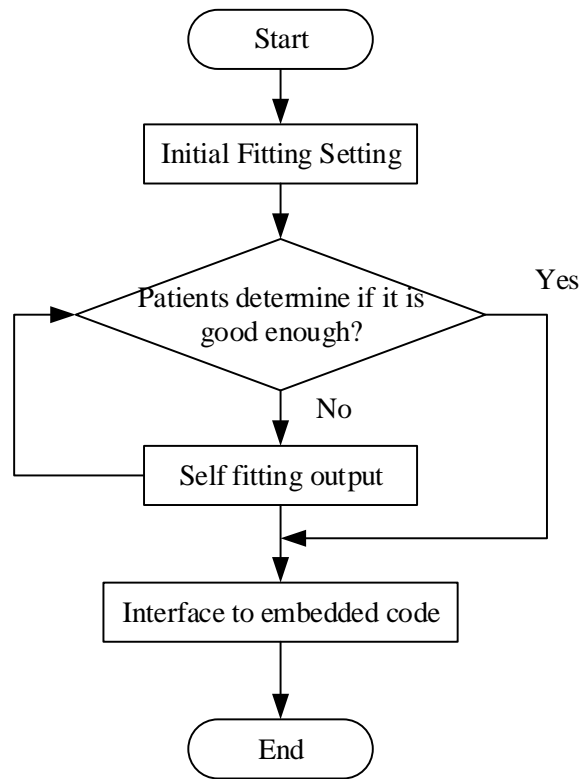
Figure 3-5 Hearing aids fitting code flowchart

Figure 3-6 illustrates the flowchart of the self-fitting algorithm. Self-fitting uses the fitting gain computed from established fitting formulas as one of the first chromosomes. The other chromosome is randomly generated. After the crossover and mutation operations as discussed in section 2.3.3 are applied, the user evaluates the outcome and determines the next step. Because the gain step size is limited, it may lead to the algorithm running into local optimal solutions. If this algorithm obtains three identical solutions, this process would start over. If the optimal solution is generated or the

maximum number of iterations is reached, the algorithm will exit and return the best evaluated fitting gain.
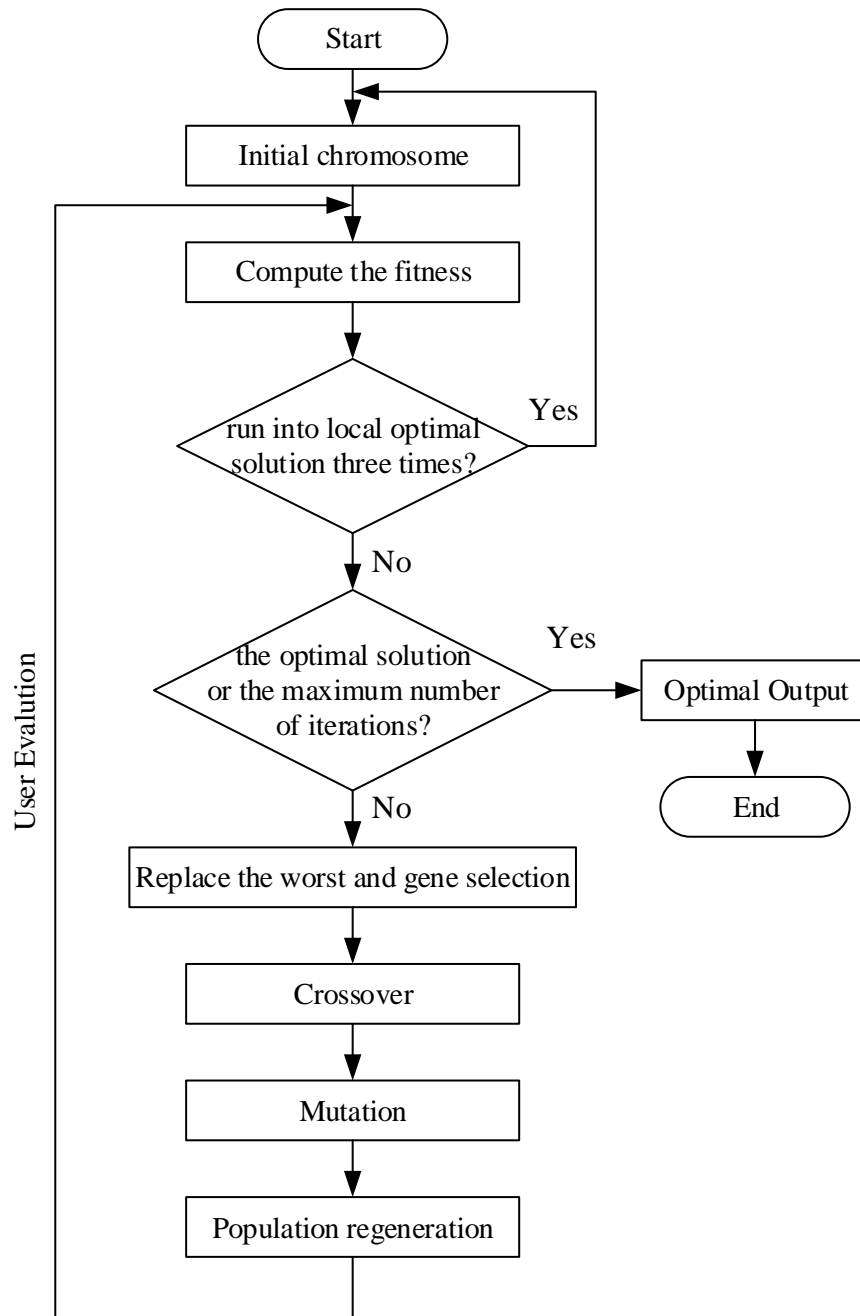
Figure 3-6 Self-fitting algorithm flowchart

After we obtain the amplification gains from the self-fitting algorithm, two more steps are needed to convert the gain into the gain programmed into the firmware. First, the

hardware calibration error needs to be taken into consideration. The calibration error was measured at different frequency band at different sound level stored and stored in a table. Second, the gain calculated in MATLAB is in individual frequency band(250, 500, 1000, 2000, 4000, and 8000 Hz) with octave spacing, but the filter bank in embedded system is in linear spacing (1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 Hz). Spline interpolation was adopted to convert the gains with octave spacing to those with linear spacing. Figure 3-7 presents the flowchart of the firmware gain calculation.
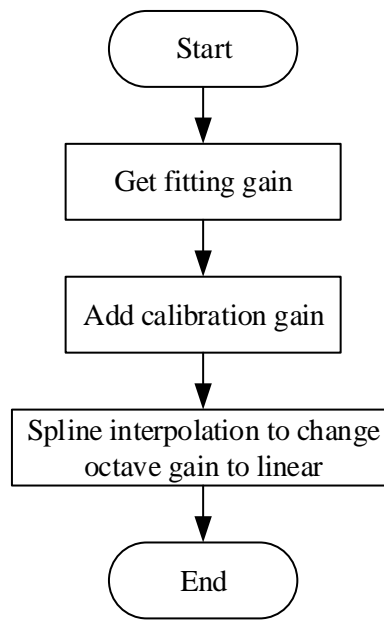
Figure 3-7 Firmware gain calculator flowchart

3.3 Amplification Device Implementation

3.3.1 Hardware Design

3.3.1.1 Overall Structure

The proposed self-fitting open-source hearing aids system hardware consists of four major parts: power management, codec, microcontroller, and other peripherals circuit. Figure 3-8 and Figure 3-9 provide the overall diagram of the hearing aids hardware.
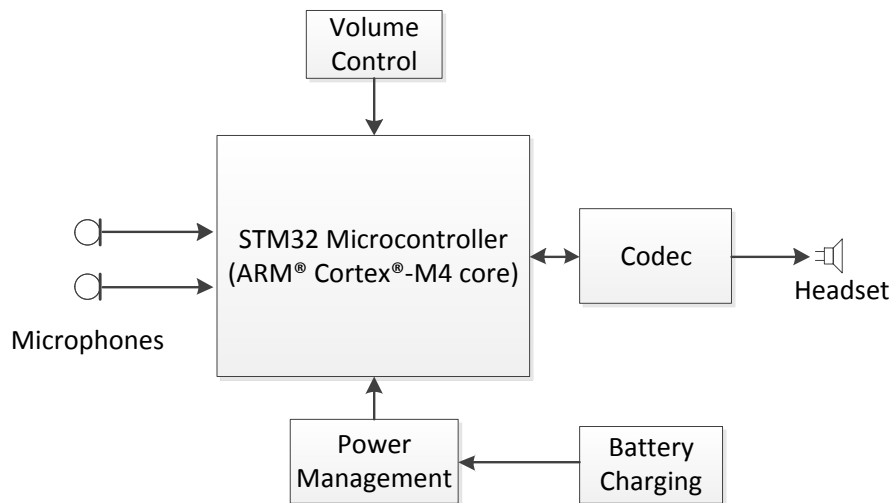
Figure 3-8 Self-Fitting Open Source Hearing aids system hardware structure
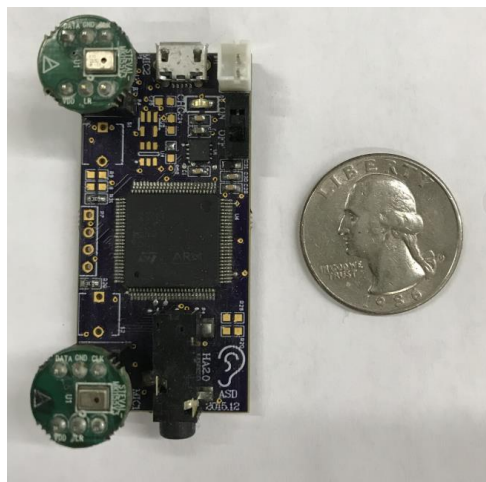
Figure 3-9 Picture of hearing aids system hardware

3.3.1.2 Power Management

Power management circuit is an import part of the hearing aids hardware. The battery selection is highly essential to the system operations. The traditional Zinc-air battery widely used in current hearing aids did not meet eco-friendly requirement. A 200 mAh lithium battery was a selection with good trade-off between system running time and hardware size. The Li-ion battery also provides the recharging ability. The charging circuit is designed as Figure 3-10 below.
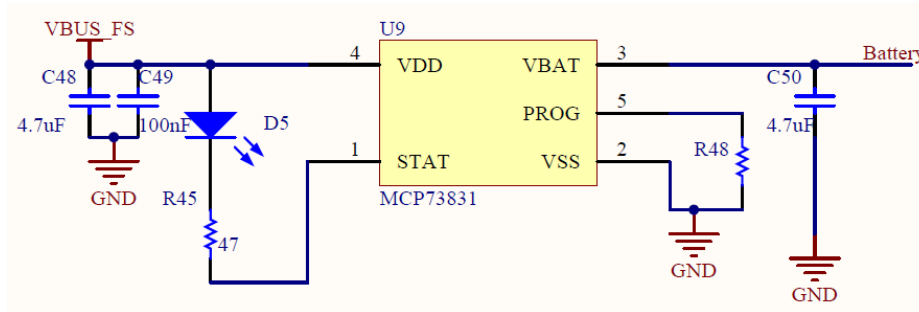
Figure 3-10 Schematic of charging circuit

MCP73831 is a highly advanced linear charge management controller for Li-ion and Li-Polymer battery [62]. It can provide programmable charge current from 15 mA to 500 mA by changing the resistor R$_{48}$.

$$I_{\text{charge}} = \frac{1000V}{R_{48}} \tag{3.1}$$

, where the unit of I_{charge} and R_{48} are in units of mA and kΩ. Charge and discharge rate of the battery is measured using Coulomb(C). 1 C means that a fully charged battery rated at 1 Ah provide 1 A for one hour. The recommended charge rate of the battery is usually between 0.5 C to 1 C, which means in this application, the charge current should be in the range of 100-200 mA. To maintain the reasonably good capacity of the battery,

49

a 10 kΩ resistor is used to achieve 100 mA charge current. The LED D5 will be red when it is charging. It will turn to green when all the charging is done.

It is well known that using the battery while charging will cause excessive heating of the battery, which affects the battery life and increases the charging time. To avoid this issue, a power selection circuit is designed as Figure 3-11 below.
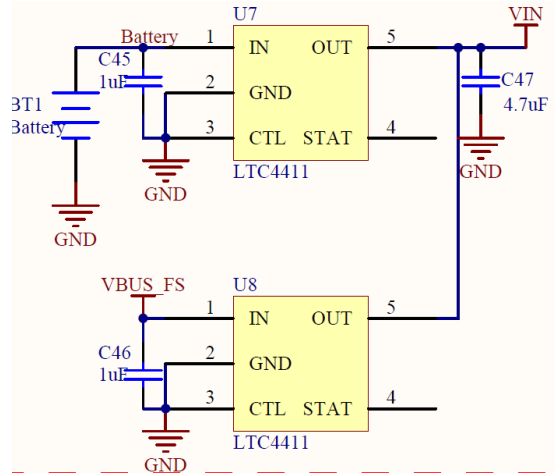
Figure 3-11 Schematic of power selection

The essence of the schematic above is an OR gate circuit constructed by ideal diodes. If a USB is connected (VBUS_FS = 5 V), the bottom circuit has higher input than the top part, and USB provides the power for the entire system; otherwise, if there is no USB power, the battery is the power source for the proposed system. Table 3-1 shows the truth table.

Table 3-1 The truth table of the power selection circuit

VBUS_FS voltage	Battery voltage	VIN Voltage
5 V	4.1 V	5 V
5 V	0 V(disconnected)	5 V
0 V(disconnected)	4.1 V	4.1 V
0 V(disconnected)	0 V(disconnected)	0 V

The LTC4411 is a low-loss ideal diode IC [63]. The regulated forward voltage is only 28 mV, and the loss current is at μA level. Those features make certain the energy loss of battery at this part is barely zero. The battery provides voltage for microcontroller and codec. Microcontroller and codec need 3.3 V and 2.5 V respectively. Two voltage regulation circuits are designed as shown in Figure 3-12 and Figure 3-13.
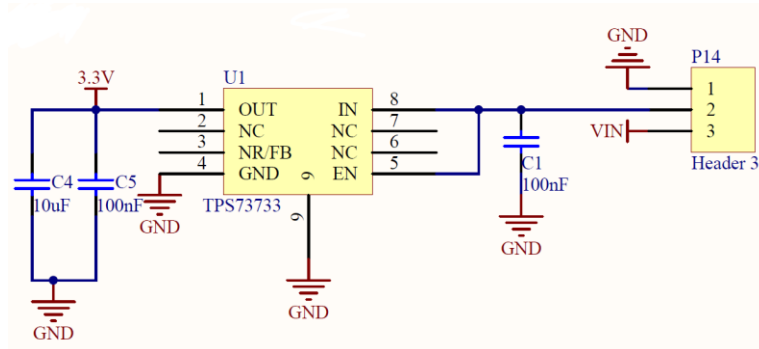
Figure 3-12 3.3 V voltage regulator
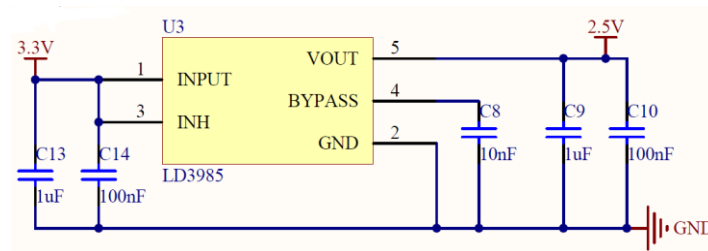
Figure 3-13 2.5 V voltage regular

The 3.3 V regulation circuit's input is either Li-ion battery 4.1 V or USB 5 V, and the output is constant 3.3 V. General voltage regulator didn't support such a low voltage drop (4.1 V to 3.3 V). TPS73733 is a low-dropout regulator with reverse current protection [64]. The dropout voltage of TSP7377 is 130 mV. It can accept the input voltage from 2.2 V to 5.5 V and provide at most 1 A current at 3.3 V with 1% accuracy. The use of C4, C5, and C1 is to reduce the noise. LD3985 is an ultra-low drop and low noise voltage

regulators [65]. It can provide up to 150 mA, and an input voltage between 2.5 V and 6 V. The low noise makes it suitable for battery-powered systems.

3.3.1.3 Codec

The audio codec is the CS43L22, a low power digital/analog converter with a class D power amplifier [66]. It provides 4 – 96 KHz sampling rate, 96 dB dynamic range, and 75 dB SNR. The class D power amplifier provides enough power to amplify sound to meet the subject's requirement. Figure 3-14 shows the schematic of the digital-to-analog circuit in the sound amplification device.
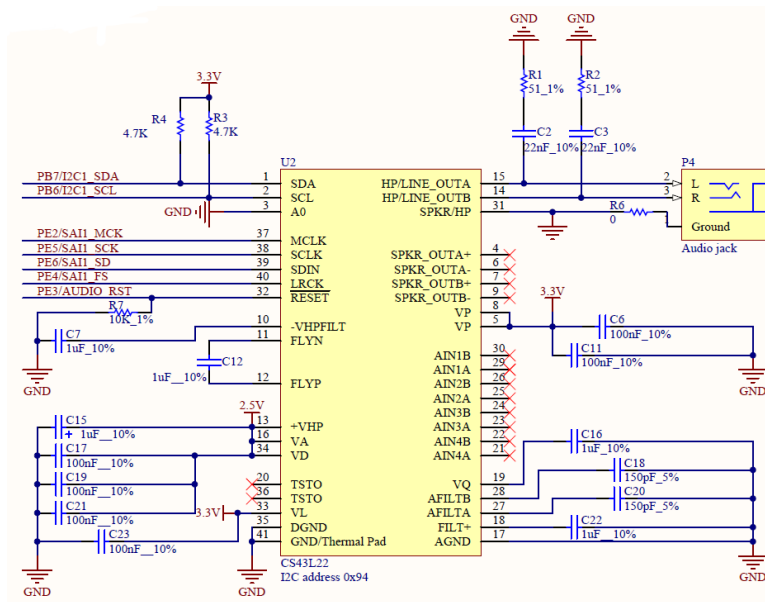
Figure 3-14 Schematic of codec CS43L22

The proposed hearing aids hardware need two omni-microphones to implement the differential microphone array algorithm. The MEMS audio sensor omnidirectional digital microphone MP45DT02 was adopted [67]. It is of low power consumption with 61 dB SNR and -26 dB sensitivity. The distance between the two microphones is set at 2 cm. Figure 3-15 is the schematic of two microphones connection.

www.manaraa.com
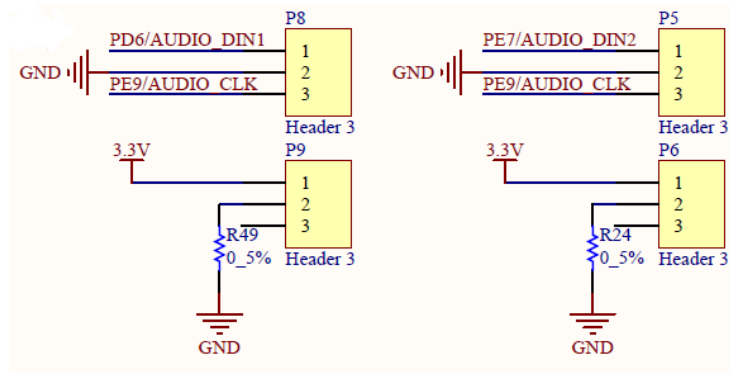
Figure 3-15 Schematic of two microphones connection

3.3.1.4 Microcontroller

The microcontroller is the brain of the hearing aids systems. It processes microphone data, performs signal processing and sends data to the codec to play the sound in real time. It must be a powerful and low-power consumption device. STM32L476 is an ultra-low-power ARM Cortex-M4 32-bit microcontroller [68]. It can run at up to 80 MHz clock speed, 100 DMIPS, and power consumption is 8 mA when it runs at full speed. The STM32L476 IC also provides DSP instruction set which will further reduce the running time of all DSP algorithms. Its built-in digital filter for sigma-delta modulators is a hardware feature that further reduces the time to convert digital microphone PDM signal to PCM signal. The STM32L476 has a rich set of analog peripherals that support all user interface needed, such as volume control and display.
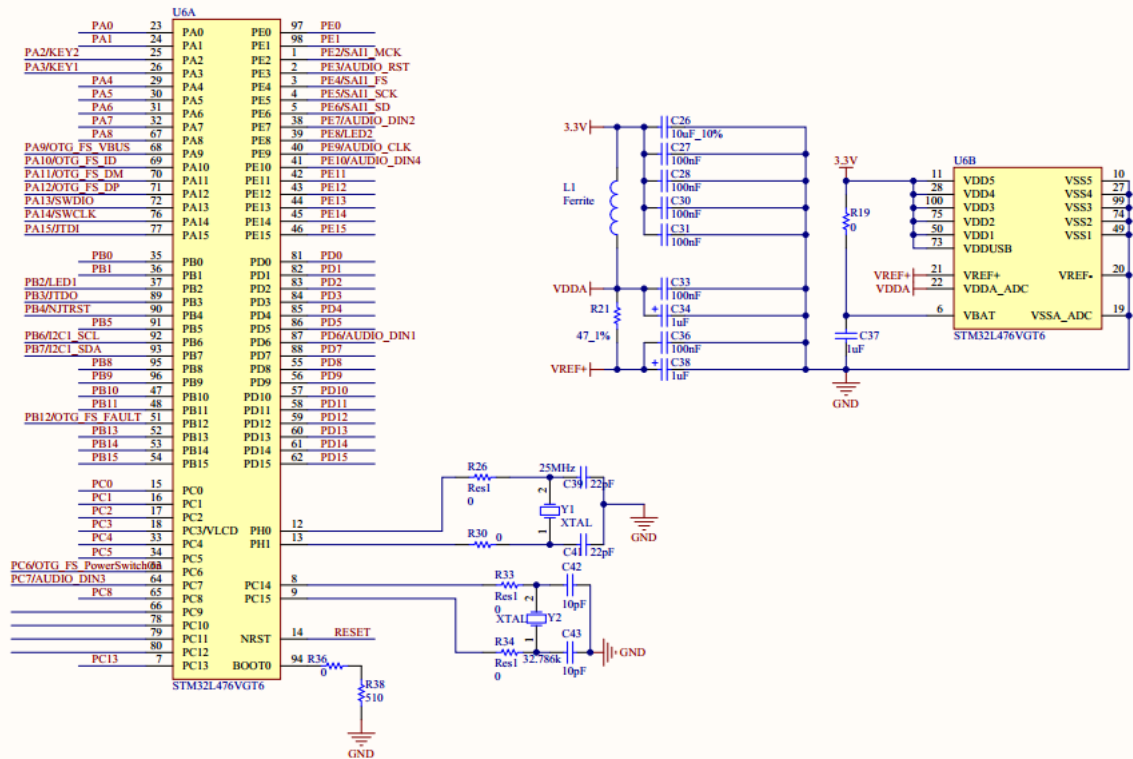
Figure 3-16 Schematic of microcontroller

3.3.1.5 User Interface
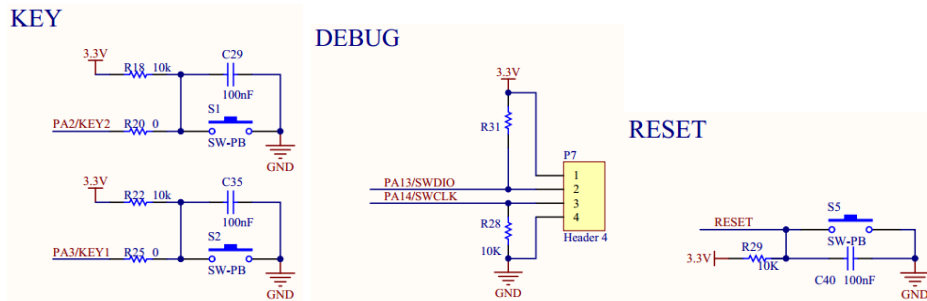
Figure 3-17 Schematic of keys and debugger

The purpose of the user interface is to adjust the volume of the sound and reset the whole system. Volume adjustment has two buttons: volume up and down.

54

3.3.2 Embedded Software Design

3.3.2.1 Overall Structure

Figure 3-18 shows the overall block diagram of the amplification device software design. All the code is based on embedded C programming language.
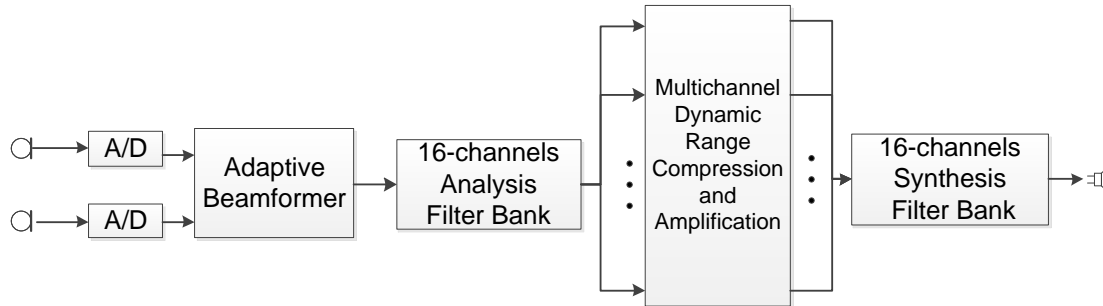
Figure 3-18 The amplification device software structure

3.3.2.2 Real-Time Audio Processing

The audio signal needs to be processed in real-time in hearing aids. Since the properties of the speech signal change relatively slow in time, short-time processing is commonly used [69]. To implement all the algorithms in the microcontroller, it is important to determine how to buffer the incoming blocks of data from ADC as well as the outgoing blocks of data to DAC. A typical solution is to use a form of double buffer (so-called ping-pong buffer) [71] as shown in Figure 3-19.
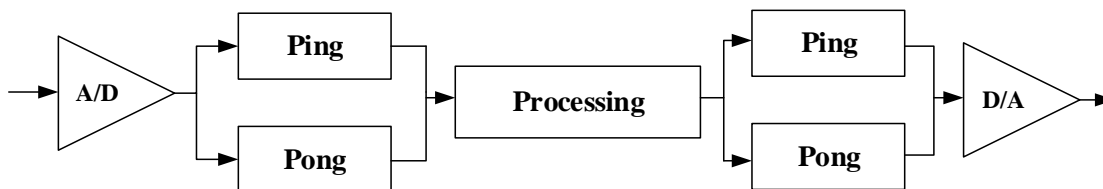
Figure 3-19 Double buffering mechanism

While the ping buffer is being filled with data from ADC, the DSP algorithm will be processing the data stored in the pong buffer. Once the ADC reaches the end of the ping buffer, it will start to store data into the pong buffer, and process new data in ping buffer. Although double buffer minimizes the latency, the algorithm needs to run more frequently. In our application, the sampling rate is 16 KHz, and the buffer size is 32 samples. Therefore, the data collecting time is 32 * 1/16000 = 2 ms, that means DSP processing time for each frame cannot excessed 2 ms.

After each frame is processed, a reconstruction method is needed to build the final signal from its short-time spectrum. The overlap-add method is applied to get the final audio signal output as displayed as the Figure 7.3.2 in [70].

3.3.2.3 Adaptive Microphone Array

All the equations in Section 2.4 were sample-by-sample based. To translate them to frame-based, R_{xy} and R_{yy} in Equation (2.27) need to be modified as follows

$$\hat{R}_{xy}(m) = \frac{1}{M} \sum_{n=1}^{M} x(n) y(n) \tag{3.2}$$

$$\hat{R}_{yy}(m) = \frac{1}{M} \sum_{n=1}^{M} y^2(n) \tag{3.3}$$

$$\hat{R}_{xy}(m) = \frac{0.9}{M} \sum_{n=1}^{M} x(n) y(n) + 0.1 \hat{R}_{xy}(m-1) \tag{3.4}$$

$$\hat{R}_{yy}(m) = \frac{0.9}{M} \sum_{n=1}^{M} y^2(n) + 0.1 \hat{R}_{yy}(m-1) \tag{3.5}$$

Figure 3-20 shows the flowchart of the adaptive microphone array algorithm.

56
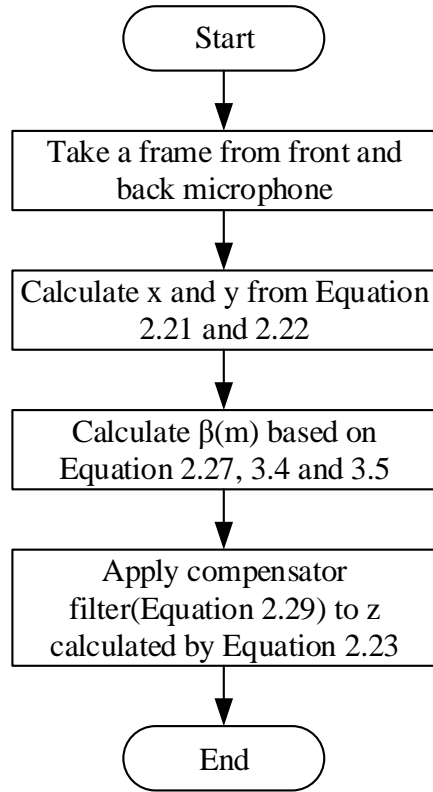
Figure 3-20 Adaptive microphone code flowchart

The STM32L476 has a built-in hardware-based digital filter for sigma-delta modulators (DFSDM) interface. The MEMS microphones provide pulse density modulation (PDM) signal. It is effortless to use DFSDM to capture more than two microphones' data at the same time with less CPU usage and less power consumption. x and y are calculated as Equation (2.21) and (2.22) by adding two samples. The optimal gain β is recursively updated using Equation (2.27). Finally, the compensation filter is applied to obtain the final output z.

The adaptive microphone array testing results are illustrated as Figure 3-21. A white noise signal was added to the clear speech to make a noisy speech testing signal. The Freq-SNRseg was improved 6dB when proposed adaptive microphone array was used

57

[72]. Figure 3-22 demonstrates the spectrogram of the differential microphone array output.
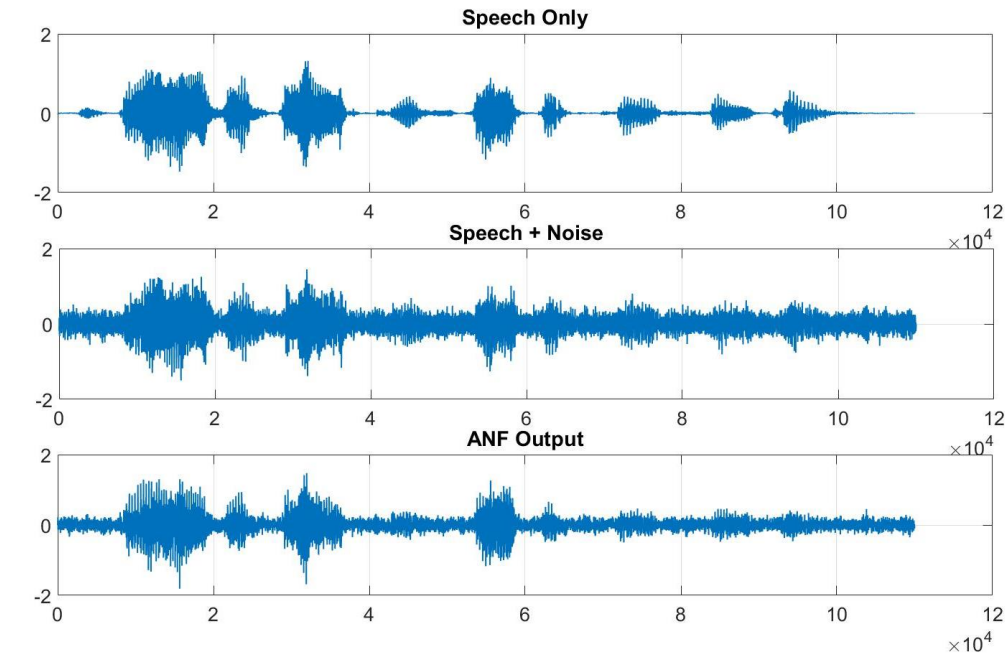
Figure 3-21 Testing result of adaptive differential microphone array
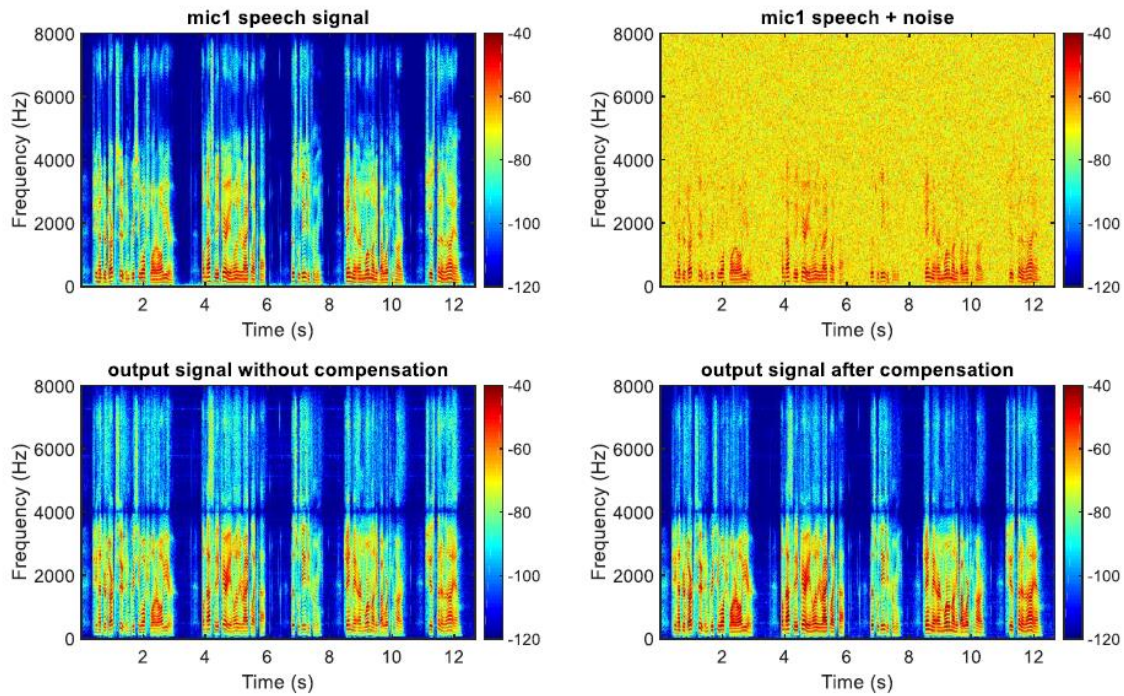
Figure 3-22 Spectrogram of differential microphone array output

3.3.2.4 Filter Bank

Although a lot of optimization has been done for filter banks algorithm, it is still the part that takes most of the processing time in microcontroller since it involves a number of matrix calculations. STM32L476 is supported by the CMSIS ARM DSP library, which is a DSP library optimized for the Cortex-M processor. It contains commonly used signal processing function such as FFT and filters. Additional optimizations need to be applied to reduce the running time further.

Since the input frame size and prototype filter length are both 32, based on Equation (2.33) and (2.39), the input is decomposed as a 4×8 matrix and the filter coefficients are converted to a 32×8 matrix. Each row of the filter matrix only contains a non-zero number with seven zeros. Instead of applying filtering operations, a vector multiplication and vector shift will improve the performance by ten times.

From Equation (2.49) and (2.50), the GDFT matrix needs to multiply with the filtered results. The size of GDFT matrix is 8×32, and the size of the analysis filter bank output is 32×16. The large-size matrix multiplication takes a lot of CPU cycles. The Equation (2.50) can be rewritten as

$$t_{k,n} = e^{j\frac{2\pi}{K}(k+k_0)(n+n_0)} = \underbrace{e^{j\frac{2\pi}{K}kn_0}}_{D_1} \cdot \underbrace{e^{j\frac{2\pi}{K}kn}}_{DFT} \cdot \underbrace{e^{j\frac{2\pi}{K}k_0(n+n_0)}}_{D_2} \tag{3.6}$$

, where D_1 and D_2 are constant diagonal matrices, $k \in [0, K-1]$ and $n \in [0, 2K-1]$. The Equation (3.6) can be further rewritten as:

$$\mathbf{T}_{GDFT} = \mathbf{D}_1 \mathbf{T}_{DFT} \mathbf{D}_2 \begin{bmatrix} \mathbf{I}_K & -\mathbf{I}_K \end{bmatrix} \tag{3.7}$$

, where $\mathbf{T}_{DFT}$ is a DFT matrix, $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown as the equation below.

59

$$\mathbf{D}_1 = e^{j\frac{2\pi}{K}kn_0} \cdot \mathbf{I}_k \tag{3.8}$$

$$\mathbf{D}_2 = \text{diag}\left[e^{j\frac{2\pi}{K}k_0(n+n_0)} \right] \tag{3.9}$$

By decomposing the GDFT matrix by two diagonal matrices ($\boldsymbol{D}_1$ and $\boldsymbol{D}_2$) and the DFT matrix $\mathbf{T}_{GDFT}$ (can be implemented by FFT), the filter bank code can be further optimized. The two optimization methods can also be applied to synthesis filter banks. The overall running time was reduced by five times compared to the original structure.
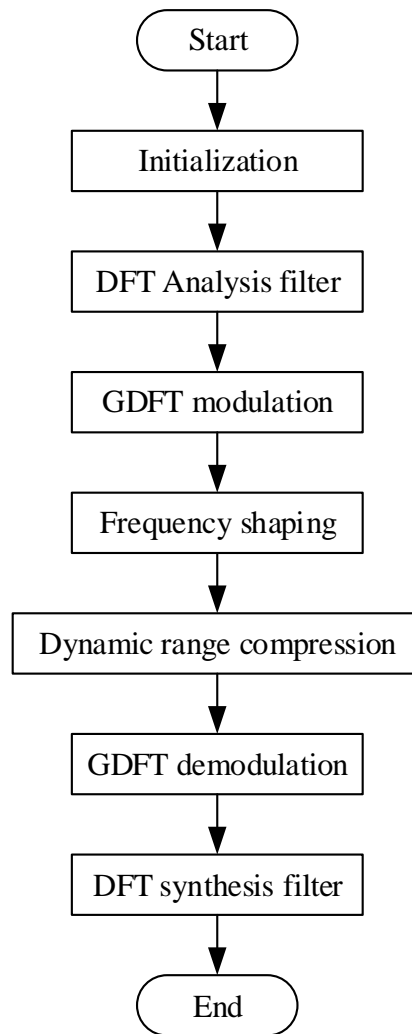
Figure 3-23 Flowchart of filter banks

Figure 3-23 illustrates the flowchart of the filter banks implementation. The initialization is to segment input into 32-sample frames, and preload the prototype filter coefficients as well as the diagonal matrix. The frequency shaping part is to apply different fitting gains after the analysis filter banks, afterward, the signal is passed through the synthesis filter banks to obtain the amplified sound. Adaptive dynamic range compression follows the frequency shaping in filter banks. Figure 2-13 presents the detailed compression implementation process.

Figure 3-24 presents the design result of 16 channels filter banks with the unity fitting gain. The prototype filters have an SNR of 80 dB. The distortion of the filter bank is within 0.1 dB.
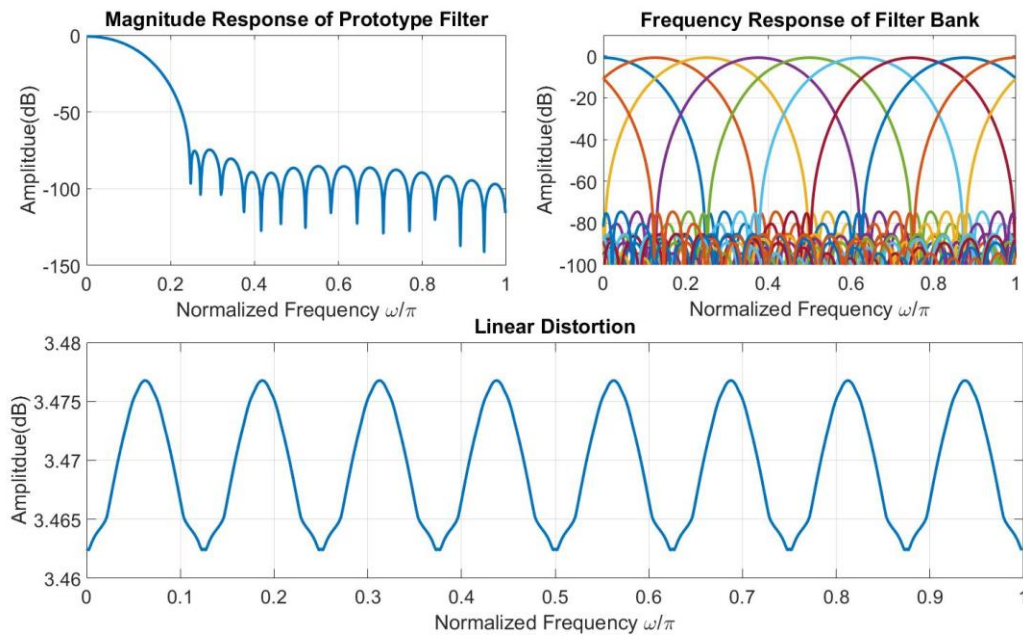
Figure 3-24 Filter banks design result

Chapter 4

Improving Speech Perception for Cochlear Implant Using Desynchronize Pulse Train

4.1 Introduction and Fundamentals of Electric Stimulation

Hearing-impaired listeners can usually benefit from hearing aids or cochlear implant. Hearing aids are easy to use, but cochlear implants require both a surgery and an adaptation process. Surgical implantations are always very costly and need significant efforts to adapt. The criteria for selecting hearing aids, or cochlear implants, is as dependent on hearing loss level [73]. The patient has mild, moderate hearing loss can benefit from hearing aids; patient with severe or profound hearing loss needed to use a cochlear implant.

The mechanisms of hearing aids and cochlear implant are different. Hearing aids improve speech comprehension by increasing sound intensities. Cochlear converts the vibrations into electrical signals that stimulate auditory neurons. This is called acoustic hearing. Acoustic hearing usually has a wide dynamic range (> 60 dB). The inner hair cell nerve synapse can have spontaneous activity and properly encode sound level changes; while the cochlear implant will bypass the cochlear processing and covert the acoustic sound

into electrical impulses that directly stimulate the auditory nerve. This is called electrical hearing. For electrical hearing, there is no spontaneous activity in inner hair cell nerve and limited dynamic range (between 5 and 15 dB) [122].

In 1970, William House and Jack Urban created the first single electrode cochlear implant – 3M implant. This implant has only a single filter band that provides little speech intelligibility. Due to the more in-depth studies on cochlea and the improvement of digital signal processing technology, the multichannel cochlear implant was developed in 1979. Later, people found the importance of high frequencies to speech intelligibility [74]. From then on, multichannel cochlear implant replaced the single electrode devices.
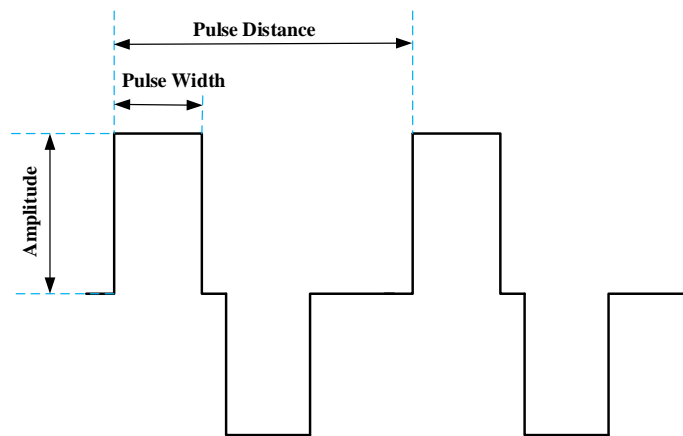
Figure 4-1 Biphasic electric impulse signal

Figure 4-1 illustrates the electric pulses generated by CI processor in two cycles. Since it has both positive and negative components, this type of pulse is called a biphasic or bipolar stimulation pulse. Biphasic stimulation is a common stimulation mode for all major cochlear implant manufacturers. It contains several important parameters that may affect the speech perception for cochlear implant user: amplitude, pulse width, pulse distance. Stimulus level is sometimes called stimulus amplitude, expressed in milliampere (mA). It is the peak value of the stimulation pulse. Pulse width is the

duration of the peak value, expressed in microseconds (ms). Bipolar pulse usually has the same pulse width for both positive and negative phases. The stimulus level and pulse width together determine the amount of charge, which affects the loudness of the sound.

$$\text{Charge of each pulse} = \text{Stimulus level} \times \text{Pulse Width} \qquad (4.1)$$

Stimulation rate is an essential parameter for pulses [75][76]. It defines how many biphasic electrical pulses generated at each electrode per second, expressed in pulse per second (pps). The optimal rate may vary from 250 pps to 10000 pps depending on the hardware limitation. Theoretically, the higher the stimulation rate, the better speech perception for patients. There is an inverse relationship between stimulation rate and pulse distance: higher stimulation rate leads to lower pulse distance. Also, the high stimulation rate will require shorter pulse width.
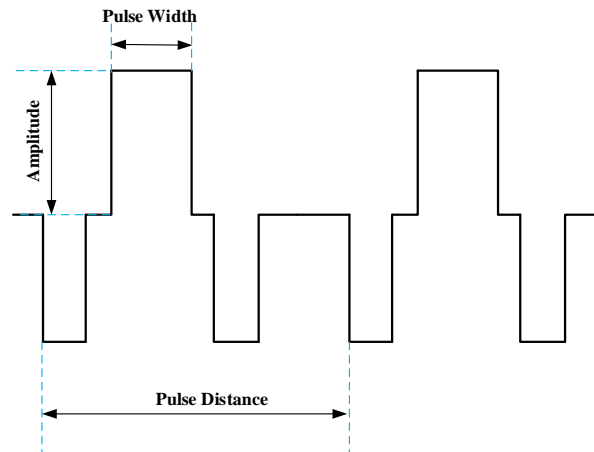
Figure 4-2 Triphasic stimulation pulse

Figure 4-2 demonstrates the triphasic stimulation pulse. It is also one of widely used electric pulse types in commercial CI products. Compared to biphasic pulse, it has a third anodic peak. There were some studies that showed triphasic pulse could benefit speech recognition for cochlear implant users [77], and some publications showed that triphasic

pulses could reduce some undesired somatic response (facial nerve stimulation) caused by using biphasic pulse [78][79].

4.2 Overview of Sound Coding Strategies

The CI sound coding strategies basically can be divided into two categories: Feature-extraction strategies and waveform strategies [74]. The feature-extraction strategy is based on extracting spectral features of the speech signal; while waveform strategy extracts the information by using filters.

4.2.1 Feature-Extraction Strategies

4.2.1.1 F0/F2 Strategy

The F0/F2 strategy was the first strategy developed for Nucleus device in early 1980s [80]. In this strategy, the fundamental frequency (F0) and the second formant (F2) are extracted from the speech signal using zero crossing detectors. The fundamental frequency of the speaker (F0 < 280 Hz) is used in stimulation. The selection of the stimulation electrodes depends on the frequency of the second format (between 800 and 4000 Hz). The amplitude of the pulse is set in proportion to the amplitude of F2. The initial result with F0/F2 strategy was encouraging as it enabled some patients to obtain open-set speech understanding. Figure 4-3 shows the F0/F2 strategy structure.

65
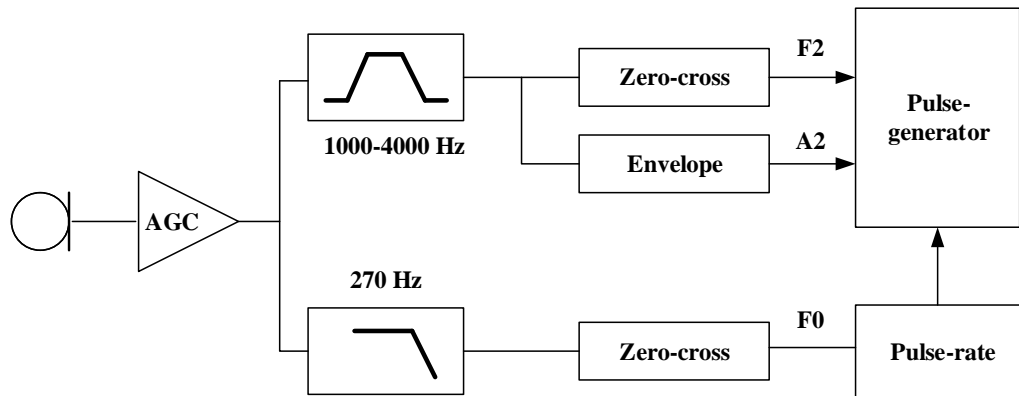
Figure 4-3 F0/F2 strategy structure

4.2.1.2 F0/F1/F2 Strategy

The F0/F1/F2 strategy was based on the F0/F2 strategy, and it encodes information about the first formant frequency [81]. The block diagram of the F0/F1/F2 processor is shown in Figure 4-4. The five most apical electrodes were dedicated to F1, while the remaining 15 electrodes were dedicated to F2. There are two electrodes that are stimulated in each processing cycle. The corresponding current is proportional to the F1 and F2 formant. The additional F1 information improved the speech recognition performance of patients wearing Nucleus cochlear implant, Compared to F0/F2, F0/F1/F2 increase 30% in terms of word recognition [82]. But the F0/F1/F2 strategy did not yield significant improvements on the consonant-recognition score since consonants require high-frequency information.
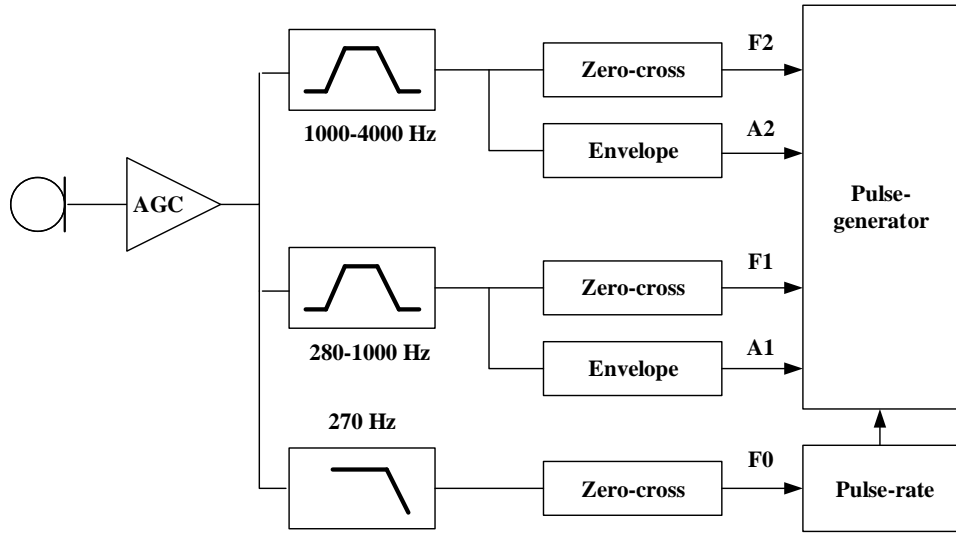
Figure 4-4 F0/F1/F2 strategy structure

4.2.1.3 MPEAK Strategy

MPEAK strategy is short for multi-peak strategy [83], and it was developed in the

University of Melbourne based on the F0/F1/F2 coding strategy. The improvement

included hardware improvement and introduced high-frequency (up to 6 KHz) pitch. The

purpose of including high-frequency information is to help patients understand

consonants. Some studies showed MPEAK has a significant improvement in sentence

recognition over F0/F1/F2 strategy on open-set speech recognition [84][85], but it had its

limitation in terms of poor performance in noisy environment. Figure 4-5 illustrates the

block diagram of the MPEAK strategy.

67
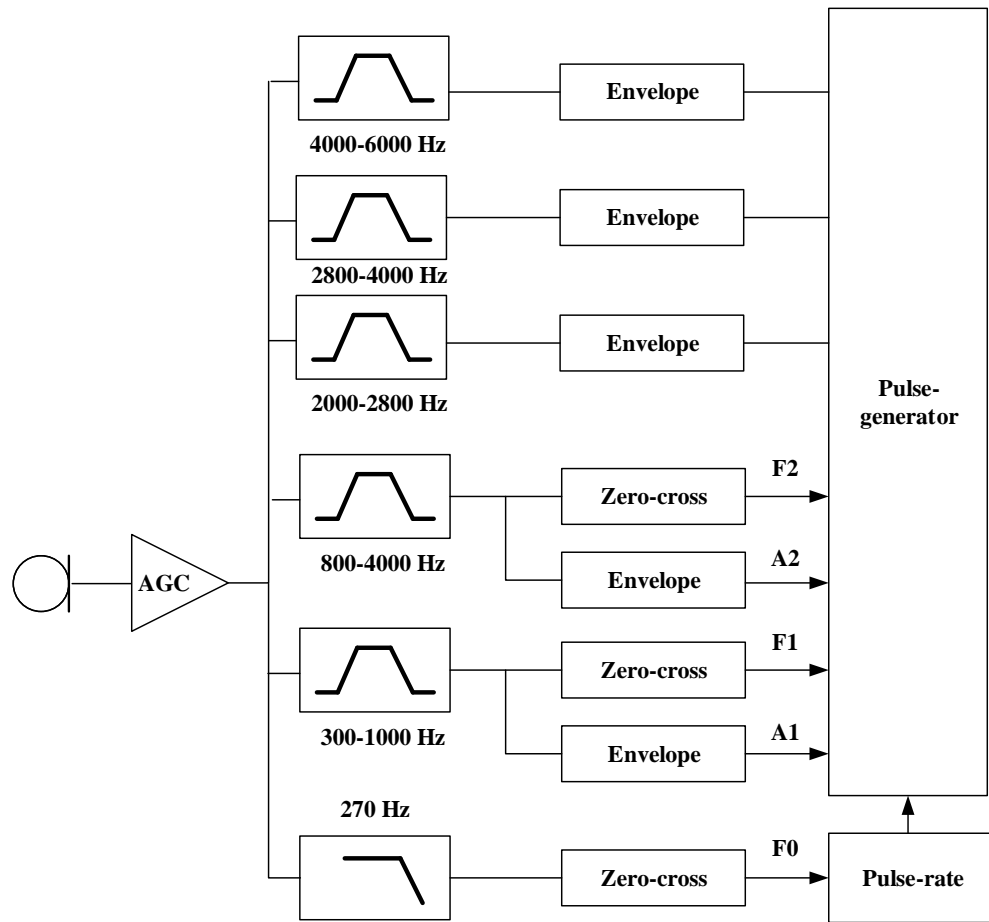
Figure 4-5 MPEAK strategy structure

4.2.1.4 SPEAK Strategy

SPEAK strategy is short for Spectral Peak strategy. It was developed by Cochlear Pty.

Limited, Australia in the early 1990s. The SPEAK coding strategy analyzes the speech

signals with 20 bandpass filters spanning from 250 Hz to 10 KHz [86]. Five to ten

spectral electrodes with maximum energies will be selected for stimulation dependent on

the spectral composition of the speech signal. That means only information on those

channels will be transmitted. The stimulation rate also can vary between 180 and 300 Hz

to preserve spectral and temporal information. It is a successful coding strategy

developed for the Nucleus implant since it has a higher frequency resolution as well as an

68

adaptive stimulation rate. When the SPEAK strategy and the MPEAK strategy were compared, SPEAK showed better performance on vowel, consonant, monosyllabic word and sentence recognition, and large improvement in noise [84][87]. Figure 4-6 provides the block diagram of SPEAK.
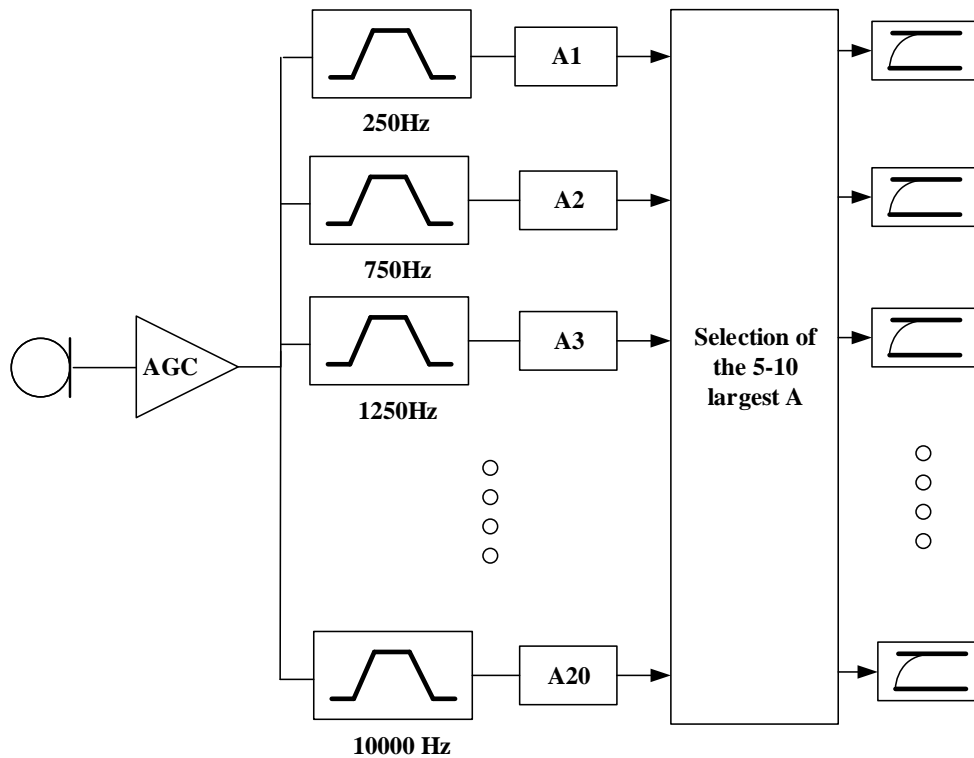
Figure 4-6 SPEAK strategy structure

4.2.1.5 ACE Strategy

ACE (Advanced Combination Encoder) [88][89] is a modified SPEAK strategy that has two more channels and a higher stimulation rate. Eight-ten channels with spectral maxima are selected to be stimulated. Experiment data showed that this strategy has better speech intelligibility performance in quiet than SPEAK [90][91]. This strategy is currently applied into Nucleus cochlear products. Figure 4-7 illustrates the ACE strategy block diagram.

69
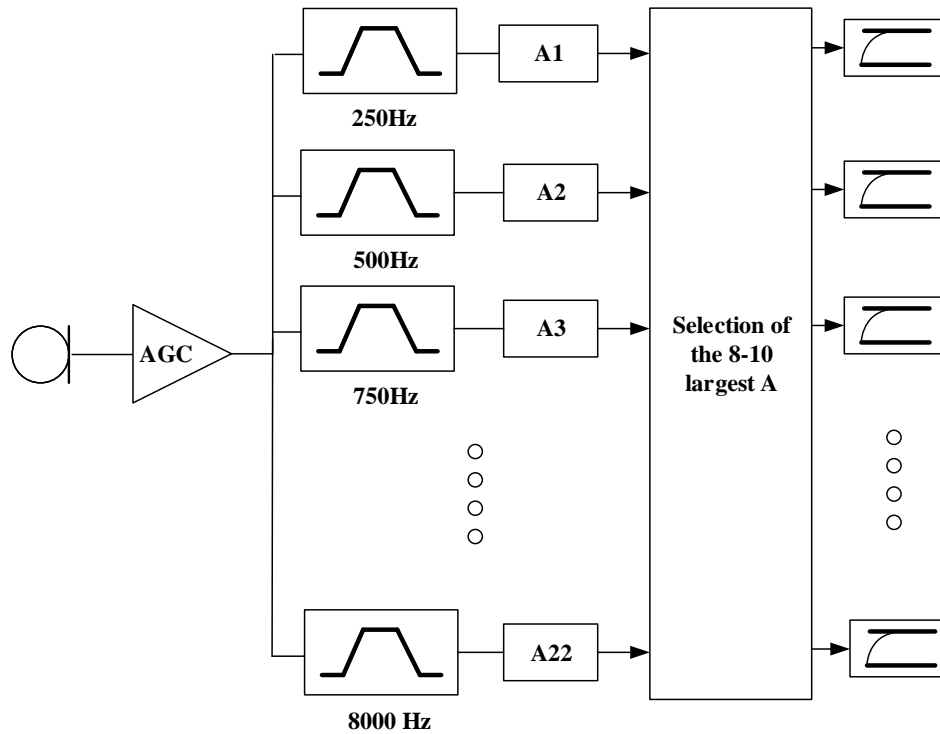
Figure 4-7 ACE strategy structure

4.2.2 Waveform Strategies

4.2.2.1 CA strategy

Compressed Analog (CA) was used in Ineraid device and UCSF/Stirz device [92]. It has

four band-pass filters between 100 Hz to 5000 Hz continuously. The filtered waveforms

are delivered simultaneously to four electrodes in analog form. CA strategy had better

performance than F0/F2 strategy, but it may lead to interactions between electrodes that

result in speech distortion and poor speech perception [93]. Figure 4-9 shows the block

diagram of the CA strategy.
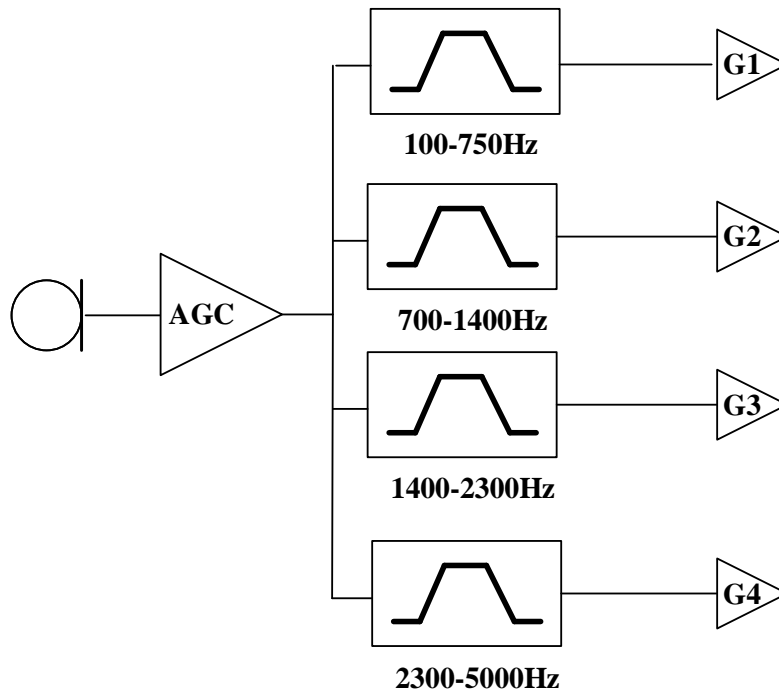
Figure 4-8 CA strategy structure

4.2.2.2 CIS strategy

Continuous Interleaved Sampling (CIS) strategy was developed by the Research Triangle

Institute in 1993 [94]. The design purpose of CIS strategy is to address the electrode

interaction issue in the CA coding strategy. In the CIS strategy, each stimulation pulses

are delivered to one electrode each time, and the impact of electrode interaction can be

minimized at the greatest extent. CIS strategy filtered the speech signals into eight

frequency bands ranging from 250 to 5500 Hz. The envelope outputs are compressed and

then used to modulate biphasic pulses. The stimulation rate has a significant impact on

speech recognition. High pulse-rate typically yields better performance than lower rate.

Compared to CA strategy, the CIS strategy significantly improved the performance in

monosyllabic word and sentences [74]. It has been shown that the following factors make

CIS strategy more successful than the CA strategy:

a) Non-simultaneous stimulations that minimized the channel interaction [95].

b) Using more frequency bands that cover more high-frequency information.

CIS is still a widely used coding strategy nowadays in commercial cochlear implants. There are two particular features designed in CIS strategy: stimulation order and compression function. Since in CIS strategy, the electrodes deliver the pulses in a sequential fashion, the stimulation order may lead to performance variations. One solution is to stimulate the electrodes from the apex channel (low frequency electrode) to the base channel (high frequency electrode). Alternatively, the electrodes can be stimulated in the stagger order [74]. Compression is applied to convert acoustical amplitudes into electrical amplitudes after filter bank processing. The dynamic range of the cochlear implant user is about 5 - 15 dB, while the input acoustic sound may vary over 30 - 60 dB. Currently, two non-linear functions were used for compression: Logarithmic function and Power-law function. Logarithmic function $Y = A\log(x) + B$ is typically used in MED-EL devices. The power-law function $Y = Ax^p + B$ is typically used in Advanced Bionic device. The electrical signal amplitude Y must fit within the most comfortable level (MCL) and the threshold level (THR). MCL and THR values vary across electrodes and the parameter A, B or p can also change across electrodes. Figure 4-9 illustrates the CIS strategy structure.
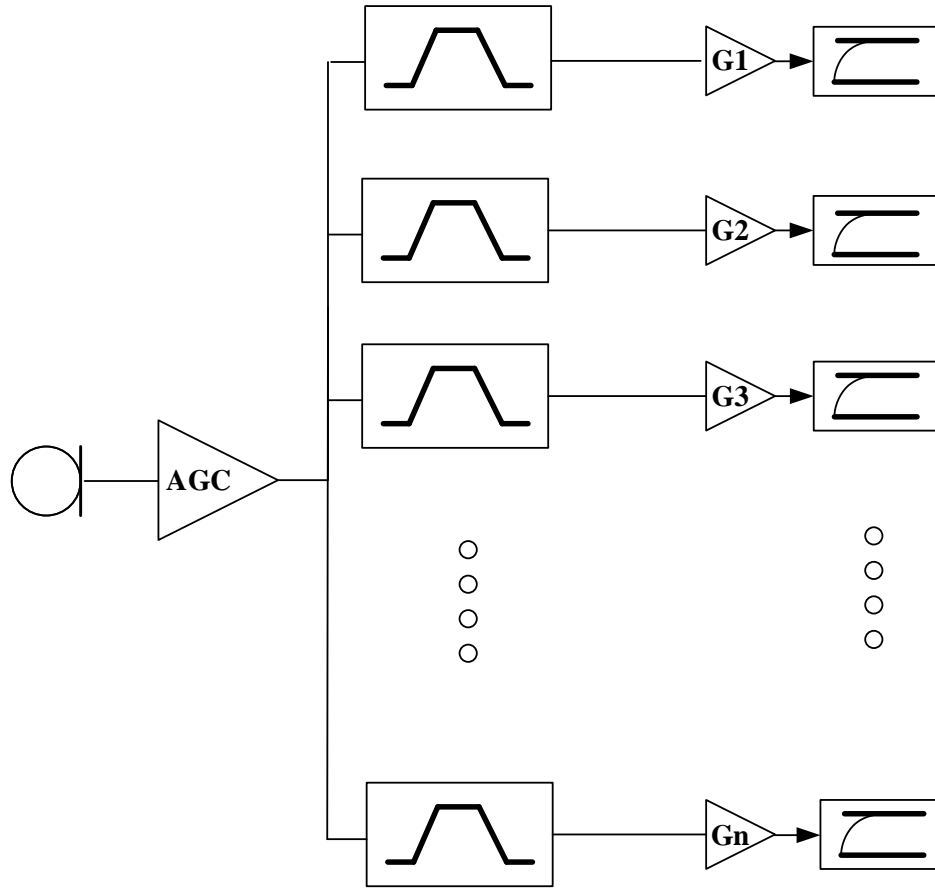
Figure 4-9 CIS strategy structure

4.2.3 Current Sound Coding Strategies

4.2.3.1 HiRes, HiRes 120 and HiRes Optima

HiRes strategy is the short notation for high-resolution strategy that is currently used in some Advanced Bionics (AB) devices [96][97]. It has the same signal processing strategy as in CIS but HiRes analyzes acoustic signals with high temporal resolution (sampled at 17.4 KHz) and delivers high stimulation rate (as high as 5000 pulses per second). However, with limited number of electrodes (16 electrodes), the high spectral resolution strategy has a hard time to show a better performance.

73

The improved version HiRes 120 attempts to increase the number of stimulation channels by current steering. Eight additional stimulation sites between each adjacent pair of electrodes lead to a maximum of 120 potential stimulation places for the entire electrode array. Combined with the basic idea from SPEAK, HiRes 120 uses the spectral maxima of the eight virtual channels between two electrodes to determine the stimulation amplitude. HiRes 120 uses fast Fourier transform (FFT) in the filter bank processing in order to achieve current steering and efficient computation [98].

Although acoustic simulation with normal-hearing listeners showed that HiRes 120 resulted in higher speech recognition in quiet and in noise, with implant user, the benefits were very limited compared to HiRes [99].

HiRes Optima is the currently available coding strategy used in AB devices [100]. Clinical testing result showed that speech perception performance of HiRes Optima is similar to HiRes 120, but it could improve the battery life from 25% to 109% for individual harmony processors [101][102]. It saves power consumption by

1) reducing the supply voltage from 8 V to 4 V;

2) limiting the current steering to only half of the area between two physical electrode contacts [102].

Figure 4-10 presents the block diagram of HiRes 120 strategy.
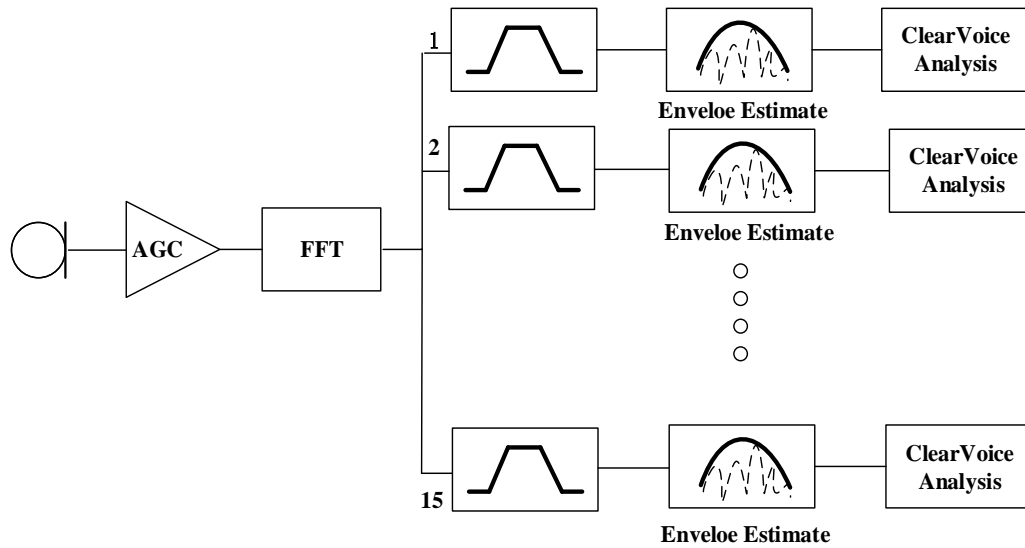
Figure 4-10 The block diagram of HiRes 120

4.2.3.2 FSP

FSP is short for fine structure processing. It has been shown that any signal can be decomposed into a slowly varying envelope signal (i,e, amplitude modulation) and a high-frequency carrier of constant amplitudes [103][104]. The high-frequency carrier was referred to as the term of 'fine structure'. For CI coding, the envelope signal is more important for speech perception, while the fine structure is more important for music perception, sound localization, speech perception in noise and tonal language perception [105].

Unlike other fixed-rate envelope-based coding strategies, FSP uses high-rate pulses with varying timing across electrodes. The timing information is used to construct the temporal structure of sound in low- to mid-frequency [106]. FSP strategy achieves timing based on channel-specific sampling sequences (CSSS), which are pulse package triggered by every other zero crossing of the filter band output. CSSS is usually used on the lower four channels below 1 KHz. The amplitude of each pulse is scaled to the instantaneous

75

amplitude of the corresponding Hilbert envelope. FSP particularly emphasizes temporal coding in low-to-mid frequency bands. For high-frequency bands, FSP realizes fine tonotopic structure using virtual channels, which shift energy between two electrodes to a third location where an electrode exists. Clinical studies showed FSP had significantly better performance on vowel, monosyllabic word perception as well as music perception than CIS [107]. Figure 4-11 shows the block diagram of FSP strategy.
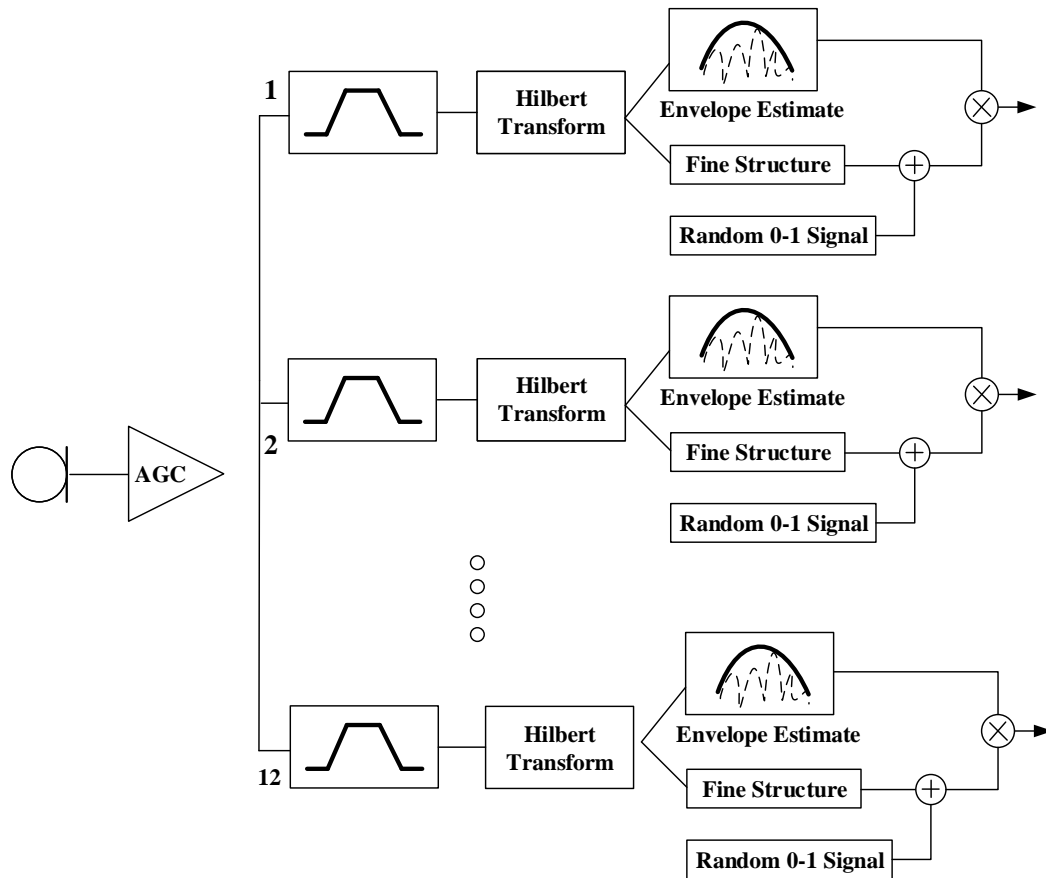
Figure 4-11 FSP structure

Different coding strategies have their own features, but the main goal of the coding strategy is the same – helping patients to achieve better speech recognition. Fig.3 in [109] shows the sentence recognition testing result in quiet. The black bars indicate the feature-extraction strategy score, while the white bars indicate the waveform strategy score. From

the trend shown in Fig.3 in [109], the modern coding strategies offer a higher score than the earlier generations. The existing commercially available strategies (such as HiRes, FSP, etc.) can reach about 80%, but no greater than 90% in terms of sentence recognition. To further improve the speech perception in quiet is very difficult since it will involve more DSP computations using high-performance speech processors. But the current 80% sentence recognition score already is a good enough one for all CI user daily use. Surprisingly, CIS is a relatively old coding strategy, but it still offers a competitive sentence recognition score when compared with current brand-new strategies.

4.2.4 Improving Speech Perception in Noise

Some studies [110]-[112] used techniques based on microphone arrays to improve speech perception in noise. An average 5 dB improvement can be achieved by using a two-microphone adaptive beamforming algorithm. The 5 dB improvement has been shown to translate to roughly a 50% improvement in sentence intelligibility [113][114].

Ying-Hui [115] and his colleagues used deep learning-based methods to help cochlear implant patients improve speech perception. A deep denoising autoencoder (DDAE) and noise classifier (NC) were developed. Compared with two classical noise reduction algorithms- logMMse and KLT, the speech perception improved about 30 perception at 0 dB SNR and 5 dB SNR using NC-DDAE model.

Srinivasan has shown that using tripolar stimulation can improve 30% in terms of sentence recognition in noise relative to monopolar stimulation. Compared to monopolar stimulation that is widely used today, Tripolar stimulation reduces current spread in the cochlea, which will lead to the improvement in speech perception [77][116].

77

Some researchers proposed some new coding strategies to improved speech perception in noise. Hu modified the ACE strategy by using the channel selection criterion based on SNR. Result showed a 24% improvement [117]. A more recent study further modified the ACE strategy by using environment-specific noise suppression. A 42% improvement was achieved at 5 dB SNR [118].

4.3 The Proposed Coding Strategy

The main reason cochlear implant user has poor performance in noise is rooted in the mechanism of electrical stimulation of the auditory system. Electrical stimulation of a damaged cochlea elicits abnormally high neural synchrony within and across auditory nerve fibers and is exacerbated by the absence of random neural activity generated by spontaneous release of neurotransmitter from inner hair cells [122]-[124]. The consequences of high neural synchrony include poor frequency selectivity and the necessity for compression of the wide range of acoustic levels into the narrow range of electric levels during pre-processing, which ultimately pose significant challenges for optimal speech perception with cochlear implants [125][126].

Rubinstein and colleagues proposed a technique in presenting constant-amplitude high-rate pulse trains designed to elicit independent, desynchronized neural response behaviors [119]. Since the high-rate pulses can desynchronize the neural responses, the pulses are called Desynchronized Pulse Train (DPT). The stimuli presented at a sufficiently high rate (i.e., ≥ 3000 pps) were predicted to drive auditory nerve fibers into maintaining a relative-refractory state within and across the neural population, thereby simulating

spontaneous, stochastic neural responses found in the normal auditory system [121]. Subsequent results from computer modeling and in vivo investigations of DPT stimuli have demonstrated the predicted desynchronizing effects [127]. Optimal DPT levels have shown improved neural sensitivity to electrical stimuli in cochlear implant recipients, with decreased thresholds and increased dynamic range [128]. In addition, DPT stimuli have been shown to improve electrode discrimination, which leads to enhanced spectral resolution that is critical for speech understanding [129]. These findings suggest that improved speech perception could be achieved for cochlear implant recipients with optimal DPT stimuli.

Based on previous findings [120], [127]-[129], a new coding strategy is proposed that combines DPT with the CIS strategy. Figure 4-12 illustrates a pulse train in one period generated by the CIS stimulus strategy. An apex to base stimulation order is used since it minimizes channel interactions. The experiment is designed using a pulse rate of 1000 pps for the CIS strategy, and a pulse rate of 3000 pps for the strategy involving the DPT.
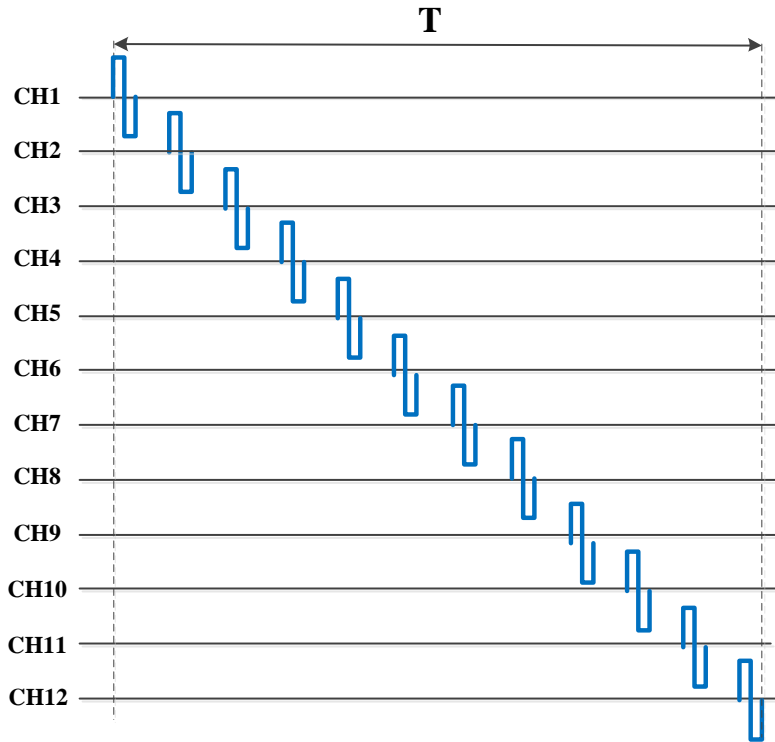
Figure 4-12 Regular CIS strategy stimulation order
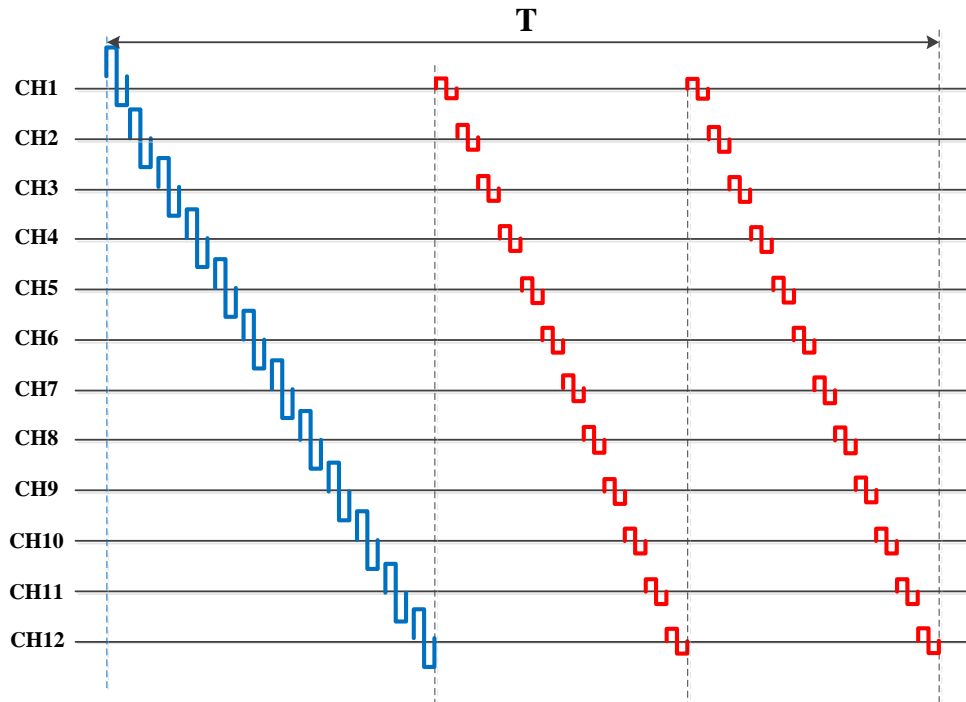
Figure 4-13 Optimal CIS-DPT-S strategy

Figure 4-13 presents the modified CIS strategy that integrates the DPT. The higher amplitude biphasic pulse presents the speech signal at different electrodes, and the lower amplitude pulse presents the additional unmodulated desynchronized pulse train. To obtain the best performance, the DPT is presented on all electrodes. But it also imposes a very harsh constraint. If the durations of even and odd channels in one period are defined as D_{even} and D_{odd} respectively, in order to achieve the stimulus pattern above, the following inequality needs to be satisfied:

$$\left(3D_{even} + 3D_{odd}\right) < \frac{1s}{1000} = 1000\mu s \qquad (4\text{-}2)$$

, which can be rewritten as:

$$\left(D_{even} + D_{odd}\right) < \frac{1000}{3}\mu s = 333.33\mu s \qquad (4\text{-}3)$$

Taking into consideration the delay introduced by the hardware as well, the effective duration of each channel will be less than 8 μs on average.

81
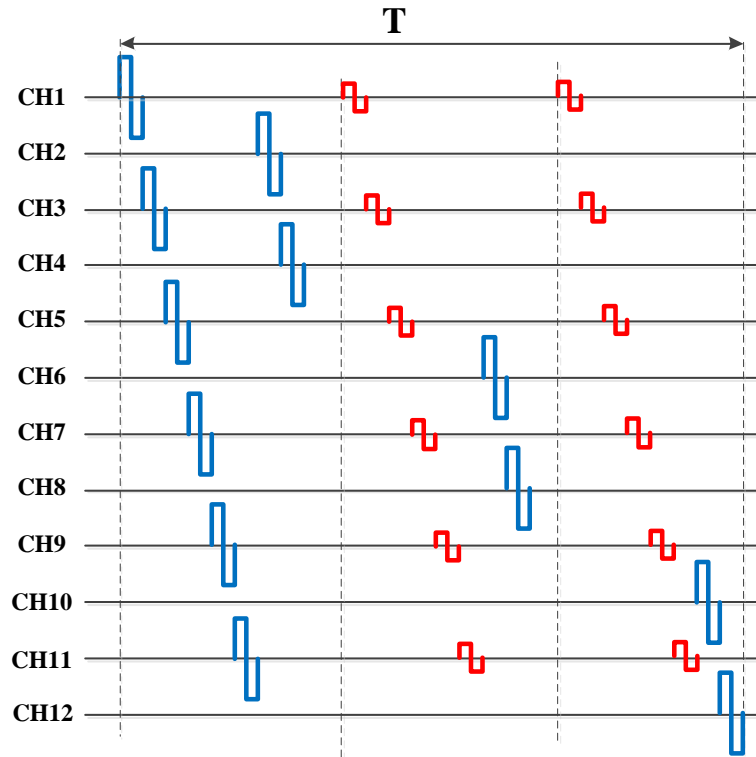
Figure 4-14 CIS-DPT-S strategy

Instead of presenting the pulses on every channel, inserting the DPT on every other channel can be used to prevent the abnormally high synchrony in the auditory nerve. Figure 4-14 shows the scenario that the DPT is only present on even channels. This coding strategy is denoted as CIS-DPT-S, and S here stands for sequential.

The speech pulse durations were given by the subject map reported from the Maestro software. The phase duration of DPT is identical to the speech pulse duration in the same electrode. Instead of stimulating from electrode 1 to 12 sequentially, the stimulation order of the proposed strategy is changed to stimulate the even electrodes first (2,4,6,8,10,12), and the odd-numbered channels are stimulated later. The additional DPT were presented to all subjects at amplitude ranging from 0 to 400 μA (in 100 μA steps). Now the

constraint of this strategy is changed to that the pulse period of all electrodes plus twice of periods on even-numbered channels should be smaller than the period 1/1000:

$$\left(3D_{even} + D_{odd}\right) < \frac{1s}{1000} = 1000\mu s \qquad (4\text{-}4)$$

The delay introduced by the hardware also needs to be taken into consideration. The criteria above will help us identify qualified CI patients. Sometimes a delay needs to be introduced to make the distance between two dashed lines in Figure 4-14 equal to T/3 if the phase duration is too small.

CIS with zero amplitude of DPT was used as the baseline test condition. Then whether the high-rate conditioner was present is the only different factor in this experiment [130]. To make sure the baseline is accurate, CIS with zero amplitude conditioner was tested three times, while the other conditioners were tested twice only.

Although the constraint in Equation (4-4) is more accessible compared to Equation (4-2), it still needs a large effort to find qualified patients. After screening nearly 150 cochlear implants user, only 23 of them satisfied this requirement. In addition to the sequential strategy, a parallel strategy was also proposed. Figure 4-15 illustrates the frame structure of CIS-DPT-P.
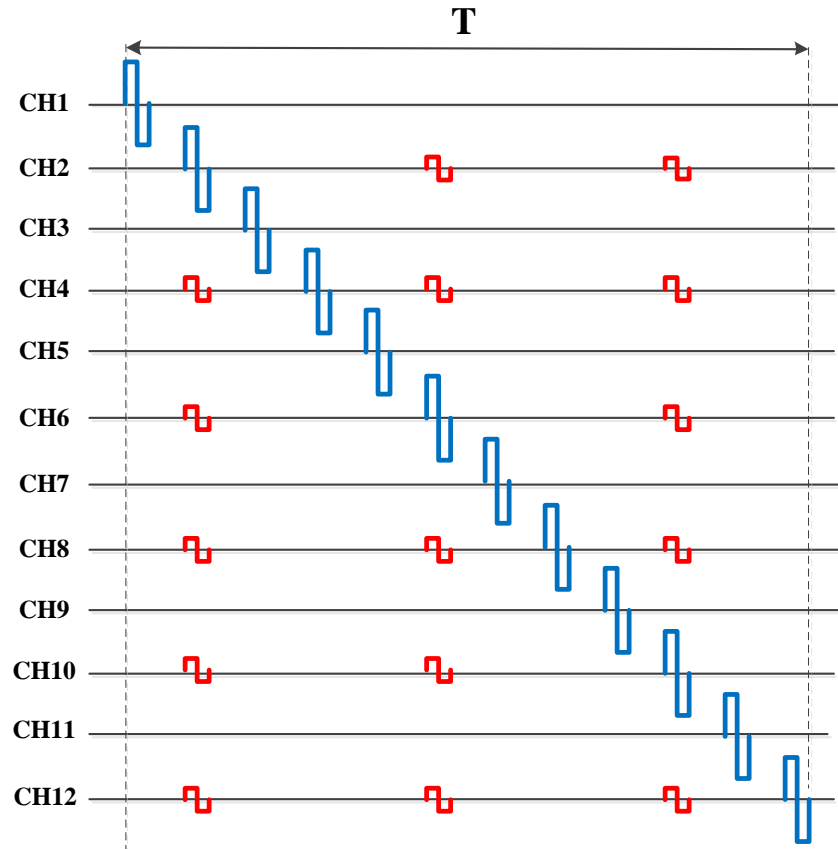
Figure 4-15 CIS-DPT-P strategy

The modified CIS with high-rate DPT in parallel (CIS-DPT-P) stimulates the speech

signal in a sequential order from apex to base. But DPT will be stimulated synchronously

in even-numbered channels. Several changes were made compared with the sequential

case:

1) Only electrodes 2, 6, 10 are stimulated with DPT to keep the rate of high rate pulse

equal to 3000 pps;

2) The effective duration of each electrode is the same due to the hardware limitation;

3) To have the same amount of charge for high-rate pulse train, the amplitude of DPT for

the parallel stimulation strategy will be reduced by five times compared to the sequential

one. For example, if the current of DPT is 200 µA for each channel in the sequential strategy, the amplitude for the parallel strategy is changed to 40 µA.

The criterion in the parallel strategy is simplified to

$$\left(D_{even} + D_{odd} \right) < \frac{1s}{1000} = 1000 \mu s \qquad (4\text{-}5)$$

4.4 Implementation

4.4.1 Hardware Setup

All stimulation strategies were implemented in the Research Interface Box II (RIB II, University of Innsbruck) platform that can be used in acute experiments for MED-EL recipients.

RIB II platform gives researchers an opportunity that tests new coding strategies for MEL-EL patients. On the hardware side of data transfer and acquisition, a PCIe-6351 data acquisition card was used (National Instruments Corporation, Austin, TX). For stimulation, a RIB detector Box i100 is used to act as electrodes. An oscilloscope can measure the 12-electrodes stimulation output signal by connecting the hooks on the i100 box. For subject testing, RIB II can directly connect the user's implant, and stream the modified speech electrical pulse signal.

4.4.2 Software Implementation

All the stimulus files were generated in MATLAB 2016b. A graphic user interface (GUI) was developed as shown in Figure 4-16 to help generate the stimulus on the fly.
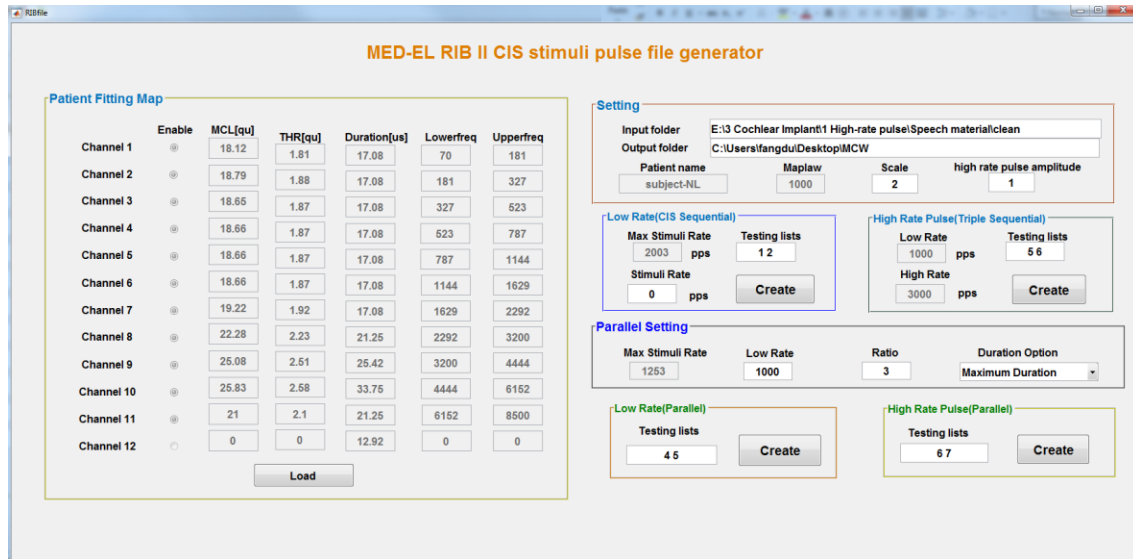
Figure 4-16 Cochlear implant stimulation files generator user interface

By clicking the load button, the user can select the patient's map, and the result will be displayed on the left panel. At the same time, the maximum stimulation rate for the modified sequential and parallel strategies can also be calculated. On the top of the right panel, the path for the input speech sound file and the generated stimulus files will be selected. By clicking the four create buttons, the user can generate low-rate sequential, high-rate sequential, low-rate parallel and high-rate parallel stimuli respectively. The callback function of the create button has the following function structure displays in Figure 4-17.
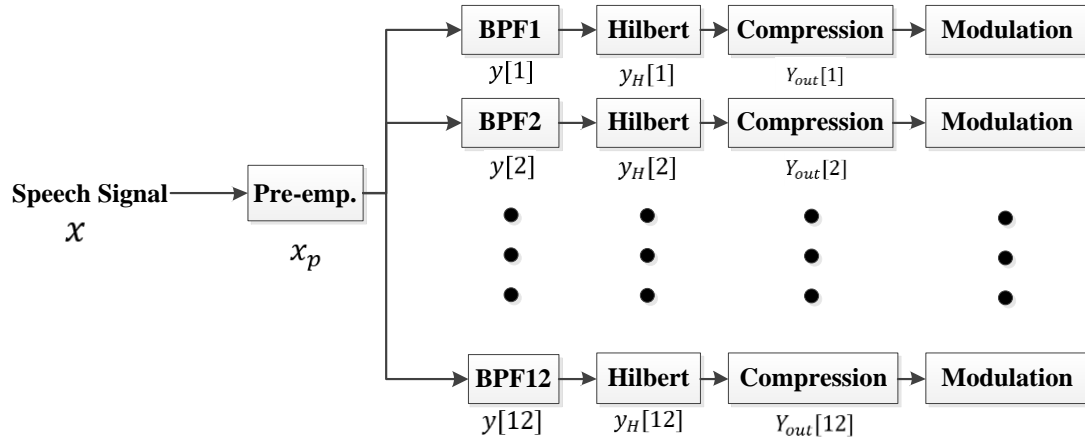
Figure 4-17 Structure of CIS-DPT signal processing

The speech signal is first passed through a pre-emphasis filter. The original speech signal usually has too much lower frequency energy, and increasing the amplitude of high frequency while decreasing the amplitude of lower frequency are needed [131]. If the original signal is x and the output of the pre-emphasized filter is $x_p[m], m = 1, 2, \cdots$

$$x_p[m] = x[m] - \alpha \cdot x[m-1] \tag{4-6}$$

, where α is a number between 0.9 and 1.

The processed signal x_p then is passed through a bank of bandpass filters. The filter bank was individually customized based on the patient's corner frequency (lower cutoff frequency and higher cutoff frequency). A 6th-order bandpass Butterworth filter was designed for each band. The filter bank separated the pre-emphasized signal x_p to 12 bandpass outputs $y[i], i = 1, 2, \cdots, 12$.

The speech information is delivered to the electrodes using a set of narrow biphasic pulses. The amplitudes of these pulses are extracted from the envelopes of the filtered bandpass waveforms $y[i]$. Hilbert Transform was used to precisely extract the overall shape, or envelope, of an incoming sound $y[i]$ with a high degree of accuracy [132].

87

Figure 4-18 illustrates an example of envelope extraction using Hilbert transform. The blue curve is one of passband output, and the red curve is its envelope.
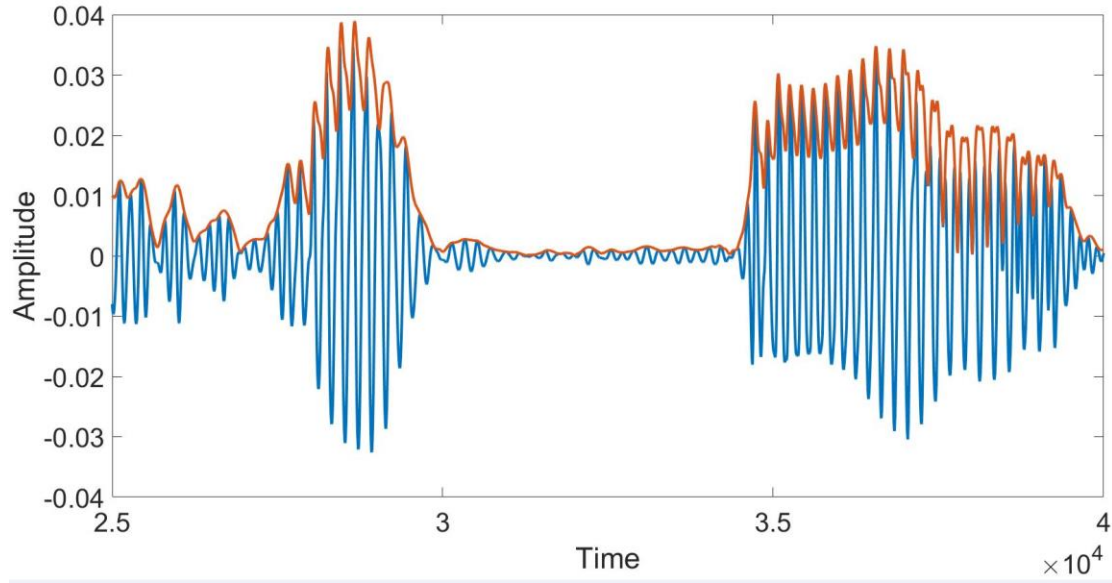
Figure 4-18 Hilbert transform envelope extraction

The output of Hilbert Transform $y_H[i]$ is down-sampled to meet the requirement of pulse rate. The sampling frequency of the speech is f_s, the pulse rate is p_s. The decimation factor is $M = f_s/p_s$. The downsampling result in:

$$y_D[i] = y_H[i \cdot M] \tag{4-7}$$

The logarithmic compression function was used to ensure that the envelope output fits the patient's map. If $MCL[i]$ and $THR[i]$ are the most comfortable level and threshold level of passband i. The compression output of pass band i is

$$Y_{out}[i] = (MCL[i] - THR[i]) \cdot \frac{\log(1 + y_D[i] \cdot c)}{\log(1 + c)} + THR[i] \tag{4-8}$$

, where c is the compression coefficient. Figure 4-19 presents the cochlear implant compression curve.

88
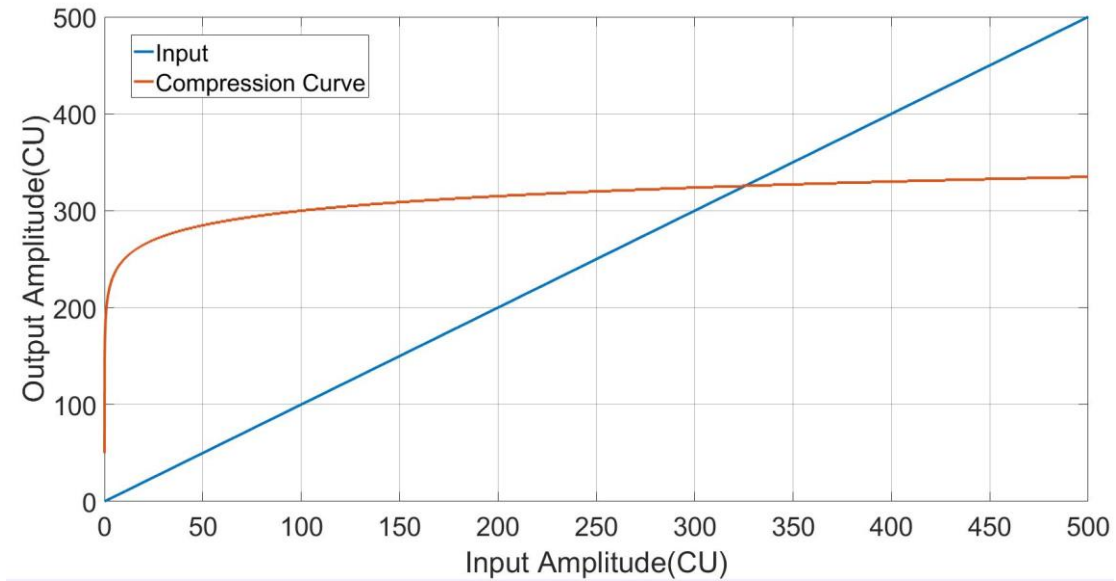
Figure 4-19 Logarithmic compression curve

Finally, modulating the compression output $Y_{out}[i]$ to an integer from 0 to 127 is needed to convert the speech signal amplitude to an electrical signal sent to CI patients. In RIB II box, there are four levels of current stimulation unit (step), the corresponding maximum current amplitude (current range) is $127 \times$ current unit. The proper current unit was selected to fit the compression output Y_{out} within the dynamic range. By organizing the pulse stimulation order and including DPT, the CIS-DPT strategy stimulation files are created based on MED-EL RIB II's source code format.

89

4.4.3 Simulation Result

Figure 4-20 RIB II system setup

Figure 4-20 illustrates the RIB II system setup in the experiments. The purpose of

simulation testing is to verify if the electrical pulse amplitude, stimulation order, and

pulse duration meet the requirements before subject testing.

Figure 4-21 RIB II speech simulation output

Figure 4-21 is an example display of the electrical signal without DPT using biphasic

pulse on one electrode. In subject testing, this signal will directly be streamed to the

patient's implant in our baseline conditions [130].

Figure 4-22 CIS-DPT-S simulation output

Figure 4-22 demonstrates a screenshot of CIS-DPT-S stimulation output on electrode 2

and 4. The yellow line indicates the pulse on electrode 2, while the cyan line indicates the

pulse on electrode 4. The high amplitude biphasic pulse represents speech signal, and the

relatively lower amplitude pulse is a DPT pulse. As seen from the pulse position provided

in Figure 4-22, pulses on electrode 4 follow electrode 2 in the sequential mode.

An example of CIS-DPT-P strategy on electrode 2 and 6 was shown in Figure 4-23. The

amplitude of the DPT pulse is smaller than in the sequential mode. And the speech and

DPT pulses on different electrodes are present at the same time, which follows the

parallel pattern.

Figure 4-23 CIS-DPT-P simulation output

Chapter 5

System Testing Result and Analysis

5.1 Introduction

The primary objective of this dissertation is to help hearing-impaired listeners improve their speech perception. To verify our proposed hearing aids system and the modified cochlear implant coding strategy, system level testing and subject testing are necessary. Testing method, materials, and result were described in detail in this chapter.

5.2 Open-Source Self-Fitting Hearing Aids System Testing

The testing of open-source self-fitting hearing aids systems is to verify the necessary electroacoustic parameters of the system, such as frequency response and battery life. The testing followed ANSI S3.22-2003 standard [133]. All the experiments were to demonstrate that the proposed hearing systems had the ability to accurately measure and improve the speech perception.

5.2.1 Testing Environment and Setup

The audiometric booth in the lab has the following technical parameters:

Table 5-1 Technical parameters of audiometric booths

Size	244 cm W x 234 cm L x 248 cm H						
Noise	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Reduction	24 dB	32 dB	42 dB	49 dB	55 dB	57 dB	55 dB
Sound	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000Hz
Absorption	0.02	1.08	1.10	1.02	1.04	1.00	1.05

The audiometric booth is equipped with the Madsen Astera 2 system. It is a PC-based

clinical audiometer used to measure patient's audiogram with high precision. In this

testing, this equipment was used as a pure tone generator. It provided single tones at

frequencies from 125 – 20000 Hz. The speaker connected with Madsen Astera 2 system

can produce up to 130 dB SPL sound with $\pm$ 3 dB accuracy.

Table 5-2 Technical specifications of EXTECH 407768 sound meter

Frequency Band	31.5 Hz – 8 KHz
Microphone	0.5'' Electret condenser Microphone
Measurement ranges	30-130 dB
Accuracy	$\pm$1.5 dB
Resolution	0.1 dB
Response time	125 ms
Applicable standards	ANSI S1.4: 1983 Type 2, / IEC 61672 Class 2
Communication	RS-232 interface to PC

A calibrated sound meter EXTECH 407768 was used to measure the sound pressure

levels. The detailed technical specification is listed in Table 5-2.

94

Figure 5-1 Hearing aids system testing setup in the sound booth

Figure 5-1 shows the testing setup of the hearing aids system. The OS-SF hearing aids system hardware was placed in the middle of the two speakers. The speakers produce speech signals and noise signals, respectively. The sound meter microphone was very close to the sound amplification device microphone to make sure they pick up similar sound pressure levels. All the data from the sound meter and proposed device is transmitted to the outside PC by USB.

5.2.2 Fitting Prescription Testing

Amplification prescription testing is to validate the implementation of existing prescription formula. Table 5-3 illustrates the errors between our device with the result published on [134] by giving the same audiograms. Flat and ski-slope audiograms were selected since those two types of hearing loss are common. From the Table 5-3, the error between our system and publication was within 1 dB in average. In other words, it validates our implementation.

95

Table 5-3 Amplification prescription testing result

Frequency	250	500	1000	2000	4000	6000
Flat audiogram						
NAL-RP	0	0	0	0	0	0
Fig6	1	-0.5	0	0	0	0
DSL	-1	-1	0.5	0	0	0
Ski-slope audiogram						
NAL-RP	0	0	0	0	0	0
Fig6	2	1	1	0	0	-1
DSL	1	0.5	1	0	0	0

5.2.3 Sound Amplification Device Testing

Electroacoustic characteristics describe the hearing aids system acoustic performance. In the latest ANSI standard[133], three most crucial parameters need to be tested: sound level range, frequency response, and gain.

The sound level range is the minimum and maximum input and output sound level that the sound amplification device can accept and provide. The dynamic range also can be calculated according to this range. For an input sound signal, the minimum sound level is the level that sound amplification device can detect; the maximum input level is the sound level needed to be limited for safety purposes. For output sound signal, the minimum output is the noise floor of hearing aids, while the maximum output is the peak sound level of hearing aid. The testing result was summarized as the Table 5-4.

Table 5-4 Sound level range of proposed hearing aids system

Input sound range	39 – 110 dB SPL
Input Referred Dynamic range	71 dB SPL
Output sound range	40 – 115 dB SPL
Output dynamic range	75 dB SPL

Frequency response is an indicator that characterizes the dynamic of hearing aids. A constant sound level (60 dB SPL) pure tone in different frequencies is adopted as a testing input signal. By setting the amplification gain to 1, the output of the amplification device is shown in Figure 5-2. The output is a relatively flat curve, and the variance is less than 3 dB compared to the input. Results demonstrated the proposed hearing aids system accurately captures the amplitude and frequency of the input sound signal.
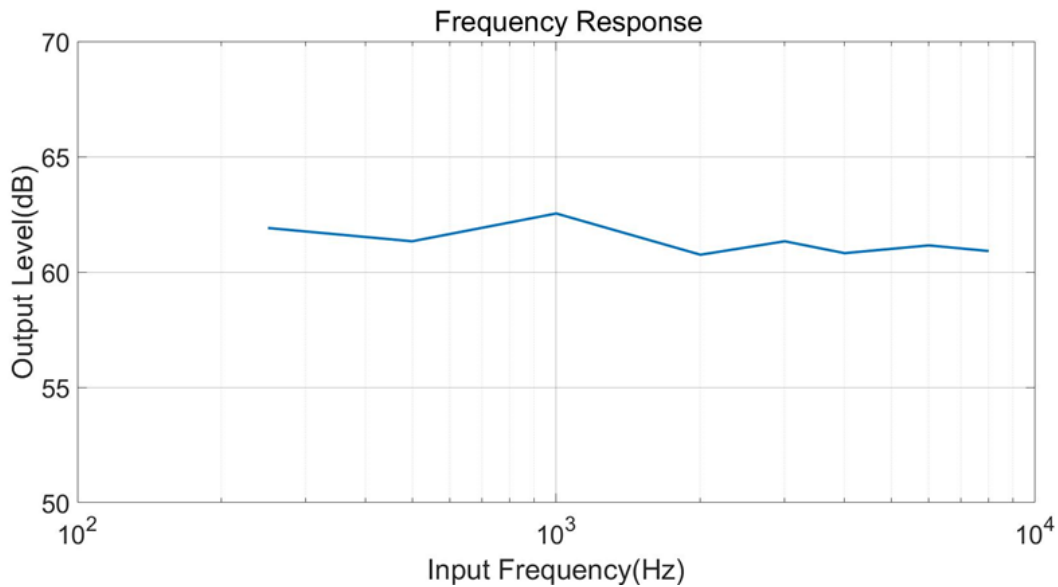
Figure 5-2 Testing result of the amplification device frequency response

Digital hearing aids can provide a different gain on the different frequency band, frequency-gain relationship reflects the main functionality of hearing aids. Both linear and non-linear amplifications were tested. First, the fitting calculator calculated the amplification gains and dynamic compression parameters in different frequency bands. Second, all those settings were programmed in the proposed sound amplification device. Figure 5-3 and Figure 5-4 illustrate the linear fitting setting and the related testing result; while Figure 5-5 and Figure 5-6 show the non-linear fitting setting details and the related testing result.

In linear fitting the gain is a constant number at different sound levels, and the input-output relationship is a straight line (the red line) shown in Figure 5-4. By adding dynamic compression, the gain is increased below 50 dB to suppress background noise, and the gain starts to decrease above 70 dB to keep the sound within the dynamic range. Between 50-70 dB, the device can provide a constant 18 dB gain.

```
Selection linear NAL-NL fitting alrigthm

Fitting =

            Band1     Band2     Band3     Band4     Band5     Band6     Band7     Band8
          _____  _____  _____  _____  _____  _____  _____  _____

IN50dB         7        19        17        16        16        16        16        16
IN65dB         7        19        17        16        16        16        16        16
IN75dB         7        19        17        16        16        16        16        16
Limiter       90        90        90        90        90        90        90        90
TAV       0.0271    0.0271    0.0271    0.0271    0.0271    0.0271    0.0271    0.0271
AT        0.0069    0.0069    0.0069    0.0069    0.0069    0.0069    0.0069    0.0069
AR        0.0011    0.0011    0.0011    0.0011    0.0011    0.0011    0.0011    0.0011
CT            60        67        60        70        67        75        71        65
CS           0.5       0.5      0.67      0.75         0       0.5      0.67       0.5
```

Figure 5-3 Linear fitting settings
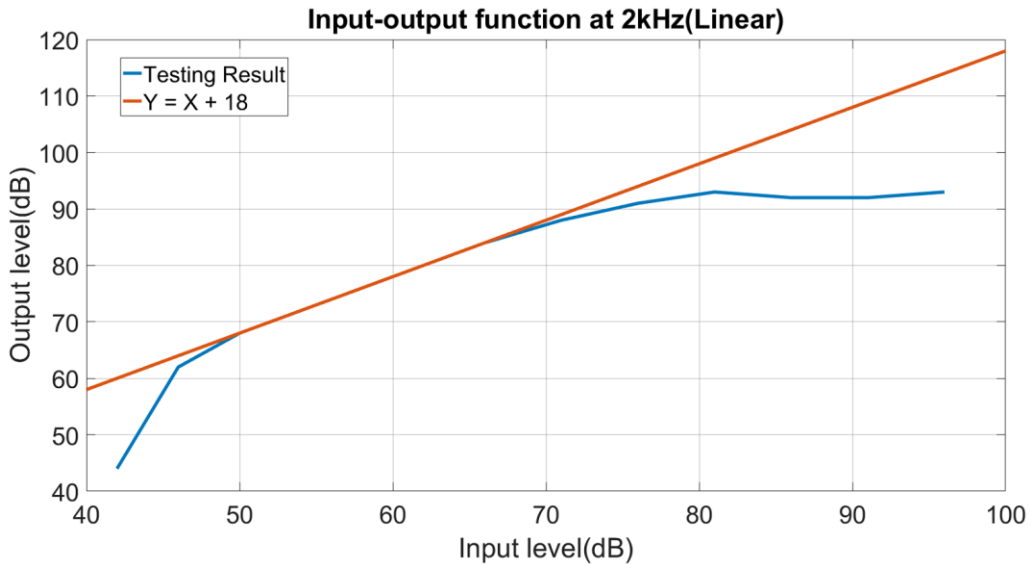
Figure 5-4 Linear fitting frequency-gain testing result

Non-linear fitting gain depends on the input sound level. The input-output relationship has a shape similar to an exponential curve. Only the fitting gain corresponding to 50 dB, 65 dB and 75 dB input signal were calculated. The other gains were generated by linear interpolation.

```
Selection non_linear DSL4 fitting alrigthm

Fitting =
```

	Band1	Band2	Band3	Band4	Band5	Band6	Band7	Band8
IN50dB	20	18	15	13	13	13	21	21
IN65dB	12	10	7	5	5	5	13	13
IN75dB	7	5	1	0	0	0	7	7
Limiter	90	90	90	90	90	90	90	90
TAV	0.0271	0.0271	0.0271	0.0271	0.0271	0.0271	0.0271	0.0271
AT	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069
AR	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
CT	60	67	60	70	67	75	71	65
CS	0.5	0.5	0.67	0.75	0	0.5	0.67	0.5

Figure 5-5 Non-linear fitting settings
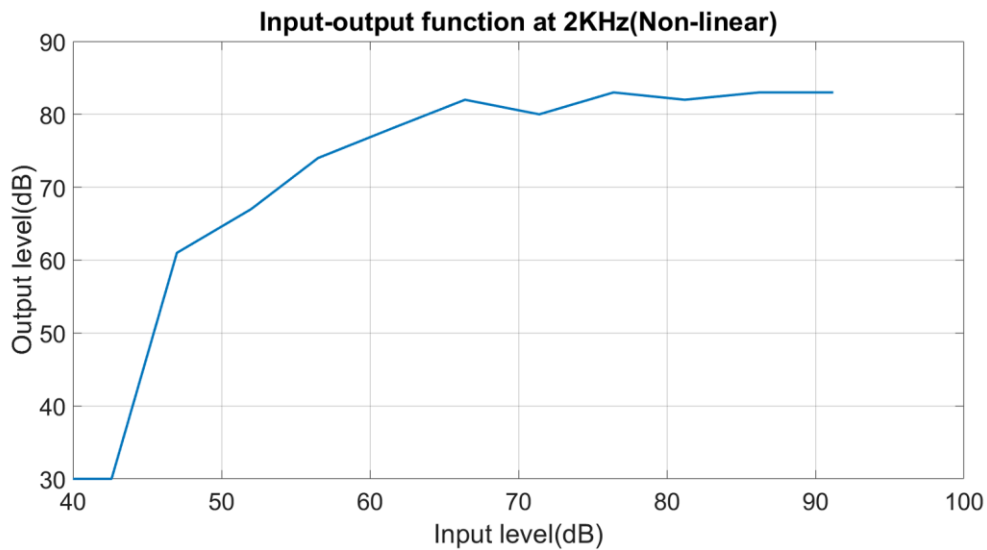
Figure 5-6 Non-linear fitting frequency-gain testing result

Table 5-5 Battery testing result

Battery Capacity	200 mAh
Peck battery current	22.9 mA
Minimum battery current	10 mA
Average battery life	9 hours
Average charge time	2 hours

Battery testing is also required in ANSI standard. Table 5-5 showed the detailed battery

information and the related testing result. 200 mAh Li-ion rechargeable battery can

sustain almost 9 hours of daily use. Table 5-6 lists current consumption for the major

components in the proposed sound amplification device.

Table 5-6 The current consuming of proposed device

MCU (STM32L476)	< 8 mA
Codec (CS43L22)	< 10 mA
Microphone and other peripheral	< 4 mA
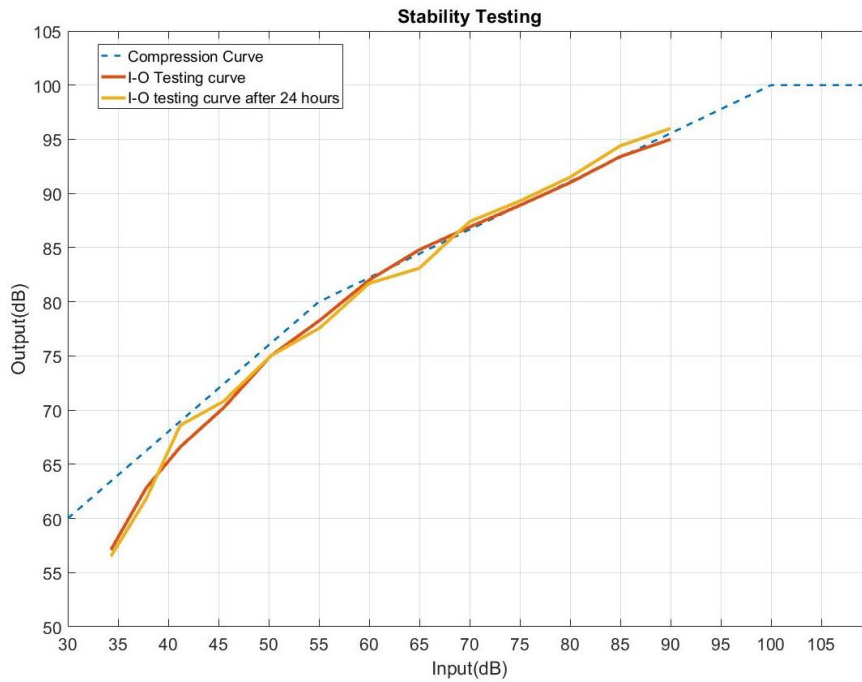
Figure 5-7 Proposed device longevity stability testing result

Hearing aids is a clinical device, the stability of the system needs to be tested to make certain that the device is stable and safe to use. The fitting gain shift before and after proposed device worked 24 hours was evaluated in Figure 5-7. The input-output relationship curves showed that the output of the proposed sound amplification device is consistent (less than 3 dB gain shifting after 24 hours use). The reason we only tested 24 hours is that the device need to restart every day after charging.

When the battery is low, the sound amplification devices still need to work correctly until the low-battery warning was sent (a voice reminder). The low-power stability testing was conducted to compare the frequency gain when battery voltage changes between 4.1 V and 3.6 V. The low-power voltage selected as 3.6 V is because it is equivalent to 90% discharge capacity according to manufacturer data shown in discharge curve in [135].
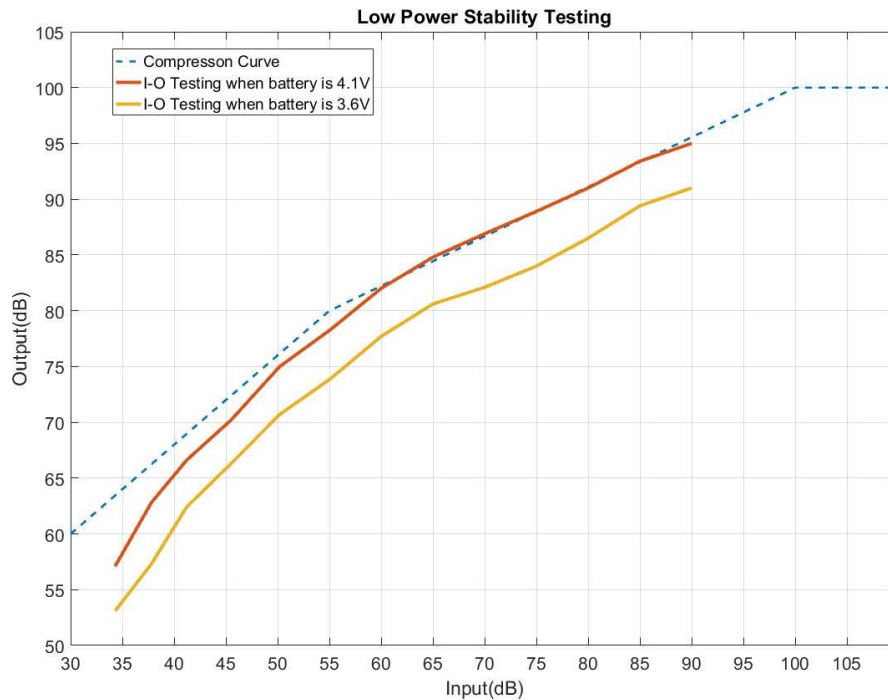
Figure 5-8 Low power stability testing result

Figure 5-8 illustrated the low power stability testing result. There is an average of 5 dB offset between the low voltage and the high voltage. The proposed device will start warning patients when running into this situation.

5.3 Cochlear Implant CIS-DPT Strategy Testing

5.3.1 Testing Material and Methods

The testing was conducted in the Medical College of Wisconsin Department of Otolaryngology Lab. Ten adult MED-EL (Innsbruck, Austria) cochlear implant recipients participated in the study. Demographic data are listed in Table 5-7.

Table 5-7 Cochlear implant testing demographic

Subject #	Length of Deafness	Age at testing	Length of CI Use	Internal Device and Electrode	Etiology
S1	3	59	3.0	Concert FLEX28	Meniere
S2	41	76	5.0	Concert FLEX28	Idiopathic progressive
S3	27	69	4.0	Concert FLEX28	Idiopathic progressive
S4	2	62	5.0	Sonatati100 Standard	Idiopathic progressive
S5	1	55	6.0	Sonatati100 Standard	Idiopathic progressive
S6	1	85	4.0	Concert FLEX28	Idiopathic progressive
S7	29	71	6.0	Sonatati100 Standard	Idiopathic sudden
S8	29	41	7.0	Sonatati100 Medium	Idiopathic sudden
S9	25	69	5.0	Concert Standard	Meniere
S10	24	55	9.0	Pulsarci100 FLEX24	Genetic

Length of Deafness = age onset deafness – age at implant;

Length of CI Use = age at test – age at implant

CIS-DPT-S testing was done in August 2017 and CIS-DPT-P testing was conducted in July 2018. Subject 1, 6 and 9 only participated in sequential mode testing because of scheduling conflicts. Study inclusion criteria included scores of clinical AzBio in quiet ≥30%, and sufficiently narrow clinical map pulse widths to satisfy DPT constraints as defined in Table 5-7.

Speech perception materials were IEEE sentence lists (IEEE Recommended Practice for Speech Quality Measurements in: IEEE Transactions on Audio and Electroacoustics, Volume: 17, Issue: 3, September 1969) presented in quiet and in 4-talker speech babble with +10 dB signal-to-noise ratio (SNR). IEEE sentence stimuli were created using a modified Continuous Interleaved Strategy (CIS) via the Research Interface Box II (RIB II, University of Innsbruck) platform designed for MED-EL cochlear implant (CI) recipients.

Subjects were directly connected to the RIB II system output via a DIB (Diagnostic Interface Box) Pulsar coil (MED-EL Corp., Innsbruck, Austria). Sentences in quiet and in noise were tested without DPT (DPT=0 µA) and with DPT at levels 100, 200, 300, and 400 µA. All stimulation conditions and sentence lists were randomized across trials. Prior to testing, subjects listened to a practice sentence list in quiet using the novel CIS-DPT strategy with no DPT present. For the conditions with DPT, subjects were asked to report subjective loudness of the stimuli at initial presentation and periodically as the perception adapted. Sentence testing was initiated once DPT loudness was reported below a soft-but-comfortable level.

103

5.3.2 Subject Testing Result

5.3.2.1 CIS-DPT-S Testing Result

The sequential mode speech perception scores across DPT level was tested and results published in Runge-Samuelson's Paper [136][137].

5.3.2.2 CIS-DPT-P Testing Result

The parallel mode speech perception scores across DPT level are shown for example subjects S7 and S4 (Figure 5-9). Compared to no-DPT, S7 had maximal speech perception improvement in noise for DPT levels of 47 µA and 66 µA, which is equivalent as CIS-DPT-S testing result. But unlike sequential mode, the speech perception improvement only happens when DPT level is 28 µA in quiet. S4 showed better performance with adding DPT in level of 28 µA compared to no-DPT in both quiet and noise.

Figure 5-10 illustrated CIS-DPT-P strategy individual (left) and group (right) analyses for no-DPT and max-DPT in quiet and in noise. In quiet, all individual data showed both improvements and decrements in performance with DPTs ranging from 3 to 15%. In noise, all subjects showed improved performance with DPT (range 0.4 to 21%). Group analyses of the effects of DPT showed a statically significant improvement for speech perception in quiet (no-DPT mean 69.4%, max-DPT mean 76.7%; paired t-test, t=-5.1, df=6, p=.002), and in noise (no-DPT mean 47.3%, max-DPT mean 56%; paired t-test, t=-3.02, df=6, p=.02). Parallel DPT shows to be less beneficial in noise compared to sequential mode.
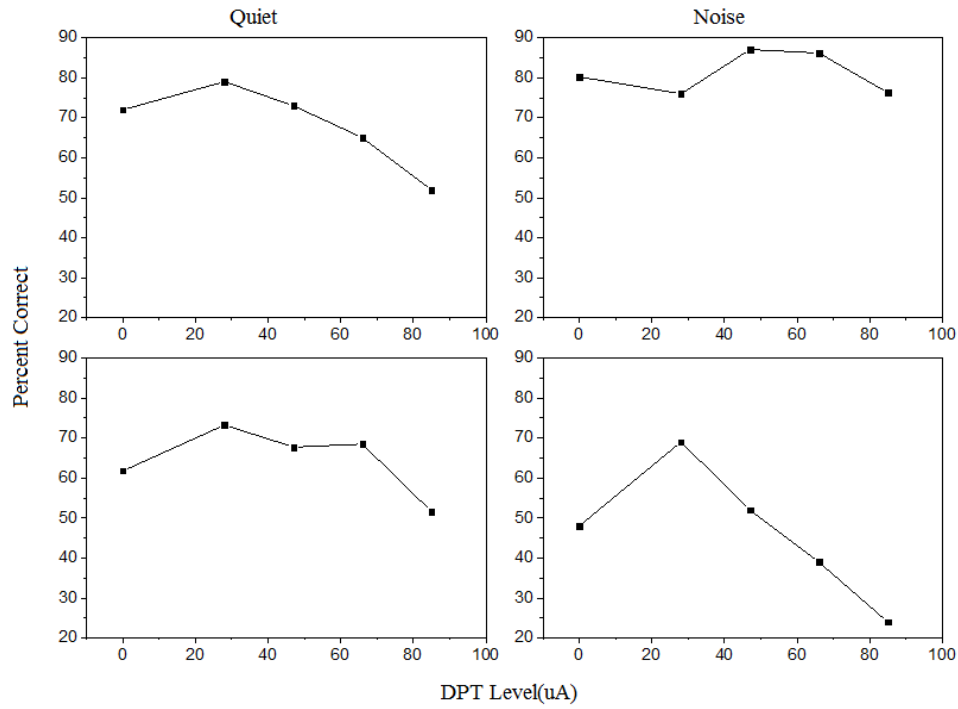
Figure 5-9 CIS-DPT-P Speech Perception testing result for subject S7 and S4
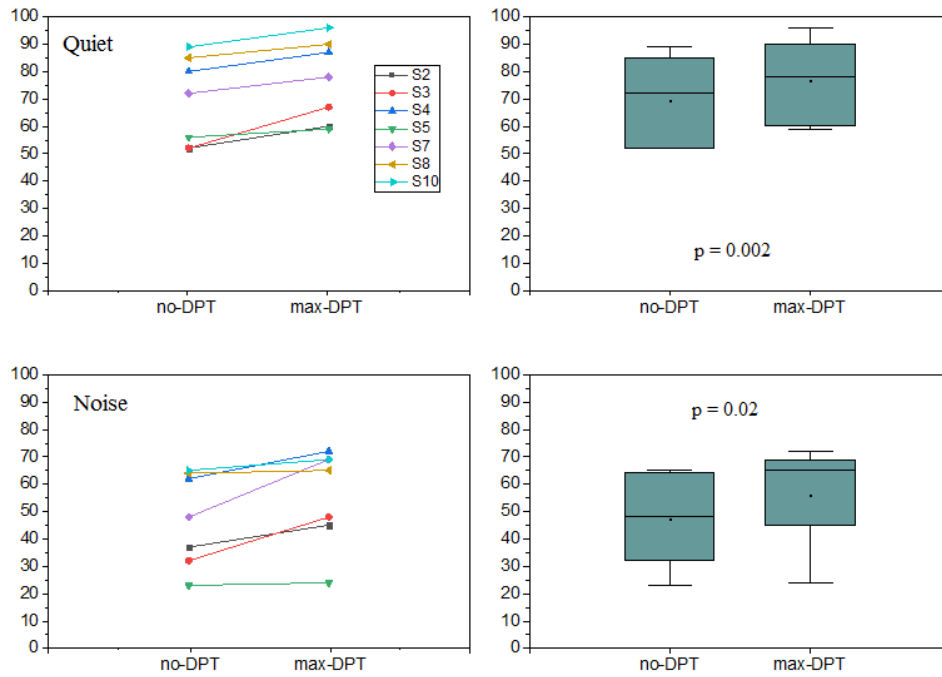
Figure 5-10 Analysis CIS-DPT-P result for no-DPT and max-DPT in quiet and noise

References

[1]. Hearing loss. (2018, September 28). Retrieved from https://en.wikipedia.org/wiki/Hearing_loss

[2]. Deafness and hearing loss. (n.d.). Retrieved from http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[3]. Kochkin, S. (2005). MarkeTrak VII: Hearing loss population tops 31 million people. *Hearing review*, *12*(7), 16-29.

[4]. Sprinzl, G. M., & Riechelmann, H. (2010). Current trends in treating hearing loss in elderly people: a review of the technology and treatment options–a mini-review. *Gerontology*, *56*(3), 351-358.

[5]. Quick Statistics About Hearing. (2017, December 20). Retrieved from https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing#3

[6]. McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them?. *International Journal of Audiology*, *52*(5), 360-368.

[7]. Hearing Aids Prices. (2014, November 09). Retriellonved from https://www.audicus.com/the-prices-of-hearing-aids-climbing-mount-everest/

[8]. Bertoli, S., Staehelin, K., Zemp, E., Schindler, C., Bodmer, D., & Probst, R. (2009). Survey on hearing aid use and satisfaction in Switzerland and their determinants. *International journal of audiology*, *48*(4), 183-195.

[9]. Cox, R. M. (2003). Assessment of subjective outcome of hearing aid fitting: getting the client's point of view. *International Journal of Audiology*, *42*(sup1), 90-96.

[10]. Caposecco, A., Hickson, L., & Meyer, C. (2011). Assembly and insertion of a self-fitting hearing aid: Design of effective instruction materials. *Trends in Amplification*, *15*(4), 184-195.

[11]. Krueger, B., Joseph, G., Rost, U., Strau-Schier, A., Lenarz, T., & Buechner, A. (2008). Performance groups in adult cochlear implant users: speech perception results from 1984 until today. *Otology & Neurotology*, 29(4), 509-512.

[12]. Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., ... & Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and hearing*, 25(4), 375-387.

[13]. Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., ... & Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and hearing*, 34(3), 342.

[14]. Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant future. *Hearing research*, *242*(1), 3-21.

[15]. Fang, D., CL Rung, & Yi H. (2018, April). Effects of High-Rate Pulse Trains on Speech Perception for Cochlear Implant Recipients. *Poster session presented at 10th Annal Student Research Poseter Competition*, Milwaukee, WI.

[16]. Mehta, A. H., & Oxenham, A. J. (2017). Vocoder simulations explain complex pitch perception limitations experienced by cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *18*(6), 789-802.

[17]. Convery, E., Keidser, G., Dillon, H., & Hartley, L. (2011). A self-fitting hearing aid: Need and concept. *Trends in amplification*, *15*(4), 157-166.

[18]. Keidser, G., & Convery, E. (2016). Self-fitting hearing aids: Status quo and future predictions. *Trends in hearing*, *20*, 2331216516643284.

[19]. Shennib, A. (2017). *U.S. Patent Application No. 15/368,342*.

[20]. Shennib, A. (2016). *U.S. Patent Application No. 15/220,292*.

[21]. Herzke, T., Kayser, H., Loshaj, F., Grimm, G., & Hohmann, V. (2017). Open signal processing software platform for hearing aid research (openMHA). In *Proceedings of the Linux Audio Conference* (pp. 35-42).

[22]. William E.A, Odile C. Creare LLC, etc.. (2017, March). Development of an Open Source Hearing Aids Platform. Poster session presented at the Annual Scientific and technology conference of the American auditory Scociet, Scottsdale, AZ

[23]. Hughson W, Westlake H. Manual for program outline for rehabilitation of aural casualties both military and civilian. Trans Am Acad Ophthalmol Otolaryngol. 1944; 48(Suppl):1–15.

[24]. ANSI S3.21-1978 (R-1992). American National Standard for Manual Audiometry

[25]. What is an Audiogram? (n.d.). Retrieved from https://www.babyhearing.org/what-is-an-audiogram

[26]. Barczik, J., & Serpanos, Y. C. (2018). Accuracy of Smartphone Self-Hearing Test Applications Across Frequencies and Earphone Styles in Adults. *American journal of audiology*, 1-11.

[27]. Poling, G. L., Kunnel, T. J., & Dhar, S. (2016). Comparing the Accuracy and Speed of Manual and Tracking Methods of Measuring Hearing Thresholds. *Ear and hearing*, *37*(5), e336.

[28]. Hearing Aids, Second Edition, Harvey Dillon; Thieme, New York, 2012, 286-336

[29]. D. Byrne & Wm. Tonisson (1976) Selecting the Gain of Hearing Aids for Persons with Sensorineural Hearing Impairments, Scandinavian Audiology, 5:2, 51-59

[30]. D. Byrne, A Parkinson, P Newall (1991) Modified hearing aids selection procedures for severe/profound hearing loss. In the Vanderbilt hearing aids report II, Studebaker GA, 295-300

[31]. Killion, M. C., & Fikret-Pasa, S. E. L. D. A. (1993). The 3 types of sensorineural hearing loss: Loudness and intelligibility considerations. *Hearing journal*, *46*, 31-31.

[32]. Cornelisse, L. E., Seewald, R. C., & Jamieson, D. G. (1995). The input/output formula: A theoretical approach to the fitting of personal amplification devices. *The Journal of the Acoustical Society of America*, *97*(3), 1854-1864.

[33]. Nielsen, J. B. B., Nielsen, J., & Larsen, J. (2015). Perception-based personalization of hearing aids using Gaussian processes and active learning. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, *23*(1), 162-173.

[34]. Liang, R., Guo, R., Xi, J., Xie, Y., & Zhao, L. (2017). Self-Fitting Algorithm for Digital Hearing Aid Based on Interactive Evolutionary Computation and Expert System. *Applied Sciences*, *7*(3), 272.

[35]. Smriga, D. (2015, October). The Verifit directional mic test: Evaluating modern directional microphone technologies. *AudiologyOnline*, Article 15371. Retrieved from http://www.audiologyonline.com

[36]. Ricketts, T. A. (2001). Directional hearing aids. *Trends in Amplification*, *5*(4), 139-176.

[37]. Elko, G. W. (1999). *A simple adaptive first-order differential microphone*. LUCENT TECHNOLOGIES INC MURRAY HILL NJ.

[38]. Luo, F. L., Yang, J., Pavlovic, C., & Nehorai, A. (2002). Adaptive null-forming scheme in digital hearing aids. *IEEE Transactions on signal processing*, *50*(7), 1583-1590.

[39]. Zhang, H., Fu, Q., & Yan, Y. (2006). Adaptive null-forming algorithm with auditory sub-bands. In *Chinese Spoken Language Processing* (pp. 248-257). Springer, Berlin, Heidelberg.

[40]. Messner, E. (2013). Differential microphone arrays.

[41]. Berner, S., & DeLeon, P. (1999). FPGA-Based Filterbank Implementation for Parallel Digital Signal Processing.

[42]. Kumar, A., Singh, G. K., & Anurag, S. (2015). An optimized cosine-modulated nonuniform filter bank design for subband coding of ECG signal. *Journal of King Saud University-Engineering Sciences*, *27*(2), 158-169.

[43]. Agrawal, S. K., & Sahu, O. P. (2013). Two-channel quadrature mirror filter bank: an overview. *ISRN Signal Processing*, *2013*.

[44]. Noor, A. O. A., Samad, S. A., & Hussain, A. (2013). A review of advances in subband adaptive filtering. *World Applied Sciences Journal*, *21*(1), 113-124.

[45]. Weiß, S., Lampe, L., & Stewart, R. W. (1997). Efficient subband adaptive filtering with oversampled GDFT filter banks.

[46]. Vaidyanathan, P. P. (1993). *Multirate systems and filter banks*. Pearson Education India.

[47]. Crochiere, R. E., & Rabiner, L. R. (1983). *Multirate digital signal processing* (Vol. 18). Englewood Cliffs, NJ: Prentice-Hall.

[48]. Weiss, S., & Stewart, R. W. (2000). Fast implementation of oversampled modulated filter banks. *Electronics Letters*, *36*(17), 1502-1503.

[49]. Weib, S. (1998). *On adaptive filtering on oversampled subbands* (Doctoral dissertation, Ph. D. thesis, Signal Processing Division, University of Strathclyde, Glasgow).

[50]. Wilbur, M. R., Davidson, T. N., & Reilly, J. P. (2004). Efficient design of oversampled NPR GDFT filterbanks. *IEEE transactions on signal processing*, *52*(7), 1947-1963.

[51]. Dumitrescu, B., Bregović, R., & Saramäki, T. (2006). Simplified design of low-delay oversampled NPR GDFT filterbanks. *EURASIP Journal on Applied Signal Processing*, *2006*, 100-100.

[52]. Kollmeier, B., Peissig, J., & Hohmann, V. (1993). Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *Journal of Rehabilitation Research and Development*, *30*, 82-82.

[53]. Chung, K. (2007). Effective compression and noise reduction configurations for hearing protectors. *The Journal of the Acoustical Society of America*, *121*(2), 1090-1101.

[54]. Zlzer, Udo, & Zölzer, Udo. (n.d.). Digital Audio Signal Processing (Second ed.). Chichester, U.K. Wiley. 225-239

[55]. Ramos, G. BLOCK PROCESSING STRATEGIES FOR COMPUTATIONALLY EFFICIENT DYNAMIC RANGE CONTROLLERS.

[56]. Horn, G. (2002). *U.S. Patent No. 6,379,314*. Washington, DC: U.S. Patent and Trademark Office.

[57]. Handzel, O., Ben-Ari, O., Damian, D., Priel, M. M., Cohen, J., & Himmelfarb, M. (2013). Smartphone-based hearing test as an aid in the initial evaluation of unilateral sudden sensorineural hearing loss. *Audiology and Neurotology*, *18*(4), 201-207.

[58]. Swanepoel, D. W., Mngemane, S., Molemong, S., Mkwanazi, H., & Tutshini, S. (2010). Hearing assessment—reliability, accuracy, and efficiency of automated audiometry. *Telemedicine and e-Health*, *16*(5), 557-563.

[59]. Kelly, E. A., Stadler, M. E., Nelson, S., Runge, C. L., & Friedland, D. R. (2018). Tablet-based Screening for Hearing Loss: Feasibility of Testing in Nonspecialty Locations. *Otology & Neurotology*, *39*(4), 410-416.

[60]. Vielsmeier, V., Lehner, A., Strutz, J., Steffens, T., Kreuzer, P. M., Schecklmann, M., ... & Kleinjung, T. (2015). The relevance of the high frequency audiometry in tinnitus patients with normal hearing in conventional pure-tone audiometry. *BioMed research international*, *2015*.

[61]. Texas Instruments, "Audio CODEC with USB interface" PCM2912 datasheet, Feb. 2008

[62]. Microchip, "Fully Integrated Li-Ion, Li-Polymer Charge Management Controllers". MCP73831/2 datasheet, 2014

[63]. Linear Technology, "2.6A low Loss Ideal Diode in ThinSOT", LTC4411, 2003

[64]. Texas Instruments, "Low-Dropout Regulator with Reverse Current Protection". TSP737xx datasheet, Jan 2006[Revised May 2015]

[65]. STMicroelectronics, "Ultra low drop-low noise BiCOMS voltage regulators low ESR capacitor compatible", LD3985 datasheet, July 2017.

[66]. Cirrus Logic, "Portable Audio DAC with integrated Class D speaker Driver", 2010.

[67]. STMicroelectronics, "MEMS audio sensor omnidirectional digital microphone", MP45DT02-M datasheet, June 2016.

[68]. STMicroelectronics, "Ultra-low-power Arm Cortex-M4 32-bit MCU+FCU". STM32L476xx datasheet, May 2018.

[69]. Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals* (Vol. 100, p. 17). Englewood Cliffs, NJ: Prentice-hall,117-130.

[70]. Proakis, J. G. (2001). *Digital signal processing: principles algorithms and applications*. Pearson Education India.

[71]. Dieffenderfer, J. N., & Kalla, R. N. (1993). *U.S. Patent No. 5,224,213*. Washington, DC: U.S. Patent and Trademark Office.

[72]. Hu, Y., & Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on audio, speech, and language processing*, *16*(1), 229-238.

[73]. Haumann, S., Hohmann, V., Meis, M., Herzke, T., Lenarz, T., & Büchner, A. (2012). Indication criteria for cochlear implants and hearing aids: impact of audiological and non-audiological findings. *Audiology research*, *2*(1).

[74]. Loizou, P. C. (1999). Introduction to cochlear implants. *IEEE Engineering in Medicine and Biology Magazine*, *18*(1), 32-42.

[75]. Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current opinion in otolaryngology & head and neck surgery*, *12*(5), 444-448.

[76]. Jain, S., & Vipin Ghosh, P. G. (2018). Acoustic simulation of cochlear implant hearing: Effect of manipulating various acoustic parameters on intelligibility of speech. *Cochlear implants international*, *19*(1), 46-53.

[77]. Srinivasan, A. G., Padilla, M., Shannon, R. V., & Landsberger, D. M. (2013). Improving speech perception in noise with current focusing in cochlear implant users. *Hearing research*, *299*, 29-36.

[78]. Bonnet, R. M., Frijns, J. H., Peeters, S., & Briaire, J. J. (2004). Speech recognition with a cochlear implant using triphasic charge-balanced pulses. *Acta oto-laryngologica*, *124*(4), 371-375.

[79]. Sürth, W., & Schatzer, R. (2017). *U.S. Patent No. 9,545,516*. Washington, DC: U.S. Patent and Trademark Office.

[80]. Seligman, P. M., Patrick, J. F., Tong, Y. C., Clark, G. M., Dowell, R. C., & Crosby, P. A. (1984). A signal processor for a multiple-electrode hearing prosthesis. *Acta Oto-Laryngologica*, *98*(sup411), 135-139.

[81]. Blamey, P. J., Dowell, R. C., Clark, G. M., & Seligman, P. M. (1987). Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant. *The Journal of the Acoustical Society of America*, *82*(1), 38-47.

[82]. Dowell, R. C., Seligman, P. M., Blamey, P. J., & Clark, G. M. (1987). Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. *Annals of Otology, Rhinology & Laryngology*, *96*(1_suppl), 132-134.

[83]. Patrick, J. F., & Clark, G. M. (1991). The Nucleus 22-channel cochlear implant system. *Scientific publications, vol. 5, 1989-1990, no. 370*.

[84]. Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., ... & Antogenelli, T. (1994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *Scientific publications, vol. 8, 1994-1995, no. 685*.

[85]. Von Wallenberg, E. L., & Battmer, R. D. (1991). Comparative speech recognition results in eight subjects using two different coding strategies with the Nucleus 22 channel cochlear implant. *British journal of audiology*, *25*(6), 371-380.

[86]. Seligman, P., & McDermott, H. (1995). Architecture of the Spectra 22 speech processor. *Scientific publications, vol. 8, 1994-1995, no. 761*.

[87]. Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., & Demorest, M. E. (1999). Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies II. Consonants. *Ear and hearing*, *20*(6), 443-460.

[88]. Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. In *Cochlear and brainstem implants* (Vol. 64, pp. 109-143). Karger Publishers.

[89]. Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and hearing*, *21*(6), 608-624.

[90]. Holden, L. K., Vandali, A. E., Skinner, M. W., Fourakis, M. S., & Holden, T. A. (2005). Speech recognition with the advanced combination encoder and transient emphasis spectral maxima strategies in nucleus 24 recipients. *Journal of Speech, Language, and Hearing Research*, *48*(3), 681-701.

[91]. Skinner, M. W., Arndt, P. L., & Staller, S. J. (2002). Nucleus® 24 Advanced Encoder conversion study: Performance versus preference. *Ear and Hearing*, *23*(1), 2S-17S.

[92]. Eddington, D. K. (1980). Speech discrimination in deaf subjects with cochlear implants. *The Journal of the Acoustical Society of America*, *68*(3), 885-891.

[93]. Dorman, M. F., Hannley, M. T., Dankowski, K., Smith, L., & McCandless, G. (1989). Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear and Hearing*, *10*(1), 44-49.

[94]. Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., & Zerbi, M. (1993). Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants. *Journal of rehabilitation research and development*, *30*, 110-110.

[95]. Loizou, P. C., Stickney, G., Mishra, L., & Assmann, P. (2003). Comparison of speech processing strategies used in the Clarion implant processor. *Ear and hearing*, *24*(1), 12-19.

[96]. Buechner, A., Brendel, M., Krüeger, B., Frohne-Büchner, C., Nogueira, W., Edler, B., & Lenarz, T. (2008). Current steering and results from novel speech coding strategies. *Otology & Neurotology*, *29*(2), 203-207.

[97]. Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., ... & Xu, L. (2009). Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear and hearing*, *30*(2), 169.

[98]. Brendel, M., Buechner, A., Krueger, B., Frohne-Buechner, C., & Lenarz, T. (2008). Evaluation of the Harmony soundprocessor in combination with the speech coding strategy HiRes 120. *Otology & Neurotology*, *29*(2), 199-202.

[99]. Firszt, J. B., Holden, L. K., Reeder, R. M., & Skinner, M. W. (2009). Speech recognition in cochlear implant recipients: comparison of standard HiRes and HiRes 120 sound processing. *Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *30*(2), 146.

[100]. de Jong, M. A., Briaire, J. J., & Frijns, J. H. (2017). Take-Home Trial Comparing Fast Fourier Transformation-Based and Filter Bank-Based Cochlear Implant Speech Coding Strategies. *BioMed research international*, *2017*.

[101]. HiRes Optima™ Sound Processing From Advanced Bionics. (n.d.). Retrieved from https://www.audiologyonline.com/releases/hires-optima-sound-processing-from-11740

[102]. Advanced Bionics. *HiRes™ Optima Clinical Results*. Sylmar, Calif, USA: Advanced Bionics White Paper; 2012.

[103]. Moon, I. J., & Hong, S. H. (2014). What is temporal fine structure and why is it important?. *Korean journal of audiology*, *18*(1), 1.

[104]. Nie, K., Stickney, G., & Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE transactions on biomedical engineering*, *52*(1), 64-73.

[105]. Li, X., Nie, K., Imennov, N. S., Won, J. H., Drennan, W. R., Rubinstein, J. T., & Atlas, L. E. (2012). Improved perception of speech in noise and Mandarin tones with acoustic simulations of harmonic coding for cochlear implants. *The Journal of the Acoustical Society of America*, *132*(5), 3387-3398.

[106]. Caldwell, M. T., Jiam, N. T., & Limb, C. J. (2017). Assessment and improvement of sound quality in cochlear implant users. *Laryngoscope investigative otolaryngology*, *2*(3), 119-124.

[107]. Müller, J., Brill, S., Hagen, R., Moeltner, A., Brockmeier, S. J., Stark, T., ... & Nopp, P. (2012). Clinical trial results with the MED-EL fine structure processing coding strategy in experienced cochlear implant users. *ORL*, *74*(4), 185-198.

[108]. Zeng, F. G. (2004). Trends in cochlear implants. *Trends in amplification*, *8*(1), 1-34.

[109]. Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear implants: system design, integration, and evaluation. *IEEE reviews in biomedical engineering*, *1*, 115-142.

[110]. Johnstone, P. M., Mills, K. E., Humphrey, E., Yeager, K. R., Jones, E., McElligott, K., ... & Little, J. P. (2018). Using Microphone Technology to Improve Speech Perception in Noise in Children with Cochlear Implants. *Journal of the American Academy of Audiology*.

[111]. Honeder, C., Liepins, R., Arnoldner, C., Šinkovec, H., Kaider, A., Vyskocil, E., & Riss, D. (2018). Fixed and adaptive beamforming improves speech perception in noise in cochlear implant recipients equipped with the MED-EL SONNET audio processor. *PloS one*, *13*(1), e0190718.

[112]. Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and multiple microphone noise reduction strategies in cochlear implants. *Trends in amplification*, *16*(2), 102-116.

[113]. Ma, J., & Loizou, P. C. (2011). SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech. *Speech Communication*, *53*(3), 340-354.

[114]. Gifford, R. H., & Revit, L. J. (2010). Speech perception for adult cochlear implant recipients in a realistic background noise: effectiveness of preprocessing strategies and external options for improving speech recognition in noise. *Journal of the American Academy of Audiology*, *21*(7), 441-451.

[115]. Lai, Y. H., Tsao, Y., Lu, X., Chen, F., Su, Y. T., Chen, K. C., ... & Lee, C. H. (2018). Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients. *Ear and hearing*, *39*(4), 795-809.

[116]. Arenberg, J. G., Parkinson, W. S., Litvak, L., Chen, C., Kreft, H. A., & Oxenham, A. J. (2018). A Dynamically Focusing Cochlear Implant Strategy Can Improve Vowel Identification in Noise. *Ear and hearing*.

[117]. Hu, Y., & Loizou, P. C. (2008). A new sound coding strategy for suppressing noise in cochlear implants. *The Journal of the Acoustical Society of America*, *124*(1), 498-509.

[118]. Hu, Y., & Loizou, P. C. (2010). Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *The Journal of the Acoustical Society of America*, *127*(6), 3689-3695.

[119]. Arenberg, J. G., Parkinson, W. S., Litvak, L., Chen, C., Kreft, H. A., & Oxenham, A. J. (2018). A Dynamically Focusing Cochlear Implant Strategy Can Improve Vowel Identification in Noise. *Ear and hearing*.

[120]. Rubinstein, J. T., Wilson, B. S., Finley, C. C., & Abbas, P. J. (1999). Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing research*, *127*(1-2), 108-118.

[121]. Hughes, M. L., Baudhuin, J. L., & Goehring, J. L. (2014). The relation between auditory-nerve temporal responses and perceptual rate integration in cochlear implants. *Hearing research*, *316*, 44-56.

[122]. Kiang, N. Y., & Moxon, E. C. (1972). Physiological considerations in artificial stimulation of the inner ear. *Annals of Otology, Rhinology & Laryngology*, *81*(5), 714-730.

[123]. Sewell, W. F. (1984). The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats. *Hearing research*, *14*(3), 305-314.

[124]. Hartmann, R., Topp, G., & Klinke, R. (1984). Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea. *Hearing research*, *13*(1), 47-62.

[125]. Kiang, N. Y. S., Eddington, D. K., & Delgutte, B. (1979). Fundamental considerations in designing auditory implants. *Acta oto-laryngologica*, *87*(3-6), 204-218.

[126]. Mitchell-Innes, A., Saeed, S. R., & Irving, R. (2018). The Future of Cochlear Implant Design. In *Advances in Hearing Rehabilitation* (Vol. 81, pp. 105-113). Karger Publishers.

[127]. Hong, R. S., & Rubinstein, J. T. (2003). High-rate conditioning pulse trains in cochlear implants: dynamic range measures with sinusoidal stimuli. *The Journal of the Acoustical Society of America*, *114*(6), 3327-3342.

[128]. Hong, R. S., Rubinstein, J. T., Wehner, D., & Horn, D. (2003). Dynamic range enhancement for cochlear implants. *Otology & neurotology*, *24*(4), 590-595.

[129]. Runge-Samuelson, C. L. (2009). Effects of high-rate pulse trains on electrode discrimination in cochlear implant users. *Trends in amplification*, *13*(2), 76-86.

[130]. Fang D, Yi H and CL. Runge (Feb 2018), "*Evaluation of a Research Sound Coding Strategy Using the Research Interface Box II*", Poster presented at meeting of the Association for Research in Otolaryngology 41th Annual Midwinter Meeting, San Diego, CA.

[131]. Ives, D. T., Calcus, A., Kalluri, S., Strelcyk, O., Sheft, S., & Lorenzi, C. (2013). Effects of noise reduction on AM and FM perception. *Journal of the Association for Research in Otolaryngology*, *14*(1), 149-157.

[132]. Hahn, S. L. (1996). *Hilbert transforms in signal processing*. Artech House.

[133]. ANSI, A. (2003). S3. 22-2003, Specification of hearing aid characteristics. *New York: American National Standards Institute*.

[134]. Rajkumar, S., Muttan, S., Jaya, V., & Vignesh, S. S. (2013). Comparative analysis of different prescriptive formulae used in the evaluation of real ear insertion gain for digital hearing aids. *Universal Journal of Biomedical Engineering*, *1*(2), 32-41.

[135]. Li-Ion & LiPoly Batteries. (n.d.). Retrieved from https://learn.adafruit.com/li-ion-and-lipoly-batteries/voltages

[136]. Runge, C. L., Du, F., & Hu, Y. (2018). Improved Speech Perception in Cochlear Implant Users With Interleaved High-Rate Pulse Trains. *Otology & Neurotology*, *39*(5), e319-e324.

[137]. RUNGE CL, Du F, Hu Y, Interleaved high-rate pulse trains improve speech perception in MED-EL cochlear implant users, 15th International Conference on Cochlear Implants and Other Implantable Auditory Technology, Antwerp, Belgium, 06/2018

Curriculum Vitae

Fang Du

EDUCATION

Ph.D. Electrical Engineering, University of Wisconsin-Milwaukee, WI, USA 2018

M.S. Electrical Engineering, Harbin University of Science and Technology, China 2014

B.S. Electrical Engineering, Harbin University of Science and Technology, China, 2011

RESEARCH INTERESTS

- Development of high-rate-pulse stochastic stimulation for cochlear implants.
- Development of open source hearing aid signal processing platform
- Embedded system design
- Digital signal processing

PROFESSIONAL EXPERIENCE

Post Graduate Intern at Johnson Control   Jan 2018- Present

Instructor   2017

Department of Electrical Engineering, University of Wisconsin- Milwaukee

- EE305 Electrical Circuit II

Graduate Teaching Assistant   2015-Present

Department of Electrical Engineering, University of Wisconsin- Milwaukee

- EE301 Electrical Circuit I
- EE305 Electrical Circuit II
- EE310 Signals and Systems
- EE335 Electronics II

PUBLICATION

- Runge, C. L., **Du, F**., & Hu, Y. (2018). Improved Speech Perception in Cochlear Implant Users With Interleaved High-Rate Pulse Trains. *Otology & Neurotology*, *39*(5), e319-e324.

- **F Du**, Y Hu, C L. Runge (2018). "*Evaluation of a Research Sound Coding Strategy Using the Research Interface Box II*", 41st annual midwinter meeting of Association for Research in Otolaryngology, San Diego, USA.

- Runge, C. L, **Du, F**., & Hu, Y (2018), "*Interleaved High-Rate Pulse Trains Improve Speech Perception in MED-EL Cochlear Implant Users*", 15th International Conference on Cochlear Implants and other Implantable Auditory Technology, Antwerp, Belgium, 2018

- YB Shi, **F Du**, M Wang, R Wang, D Xiu. (2013). "An Optimization of the Chlorine Gas Wireless Sensor Network Coverage and Leakage Model". *Sensors & Transducers*, 154(7), 62-70.

- M Wang, D Xiu, R Wang, **F Du**, YB Shi. (2013). "Data Mining Research in Wireless Sensor Network Based on Generic BP Algorithm". Proceeding of 2013 2nd International Conference on Measurement, information and Control, Harbin, China.

- Q Feng, Y Shi, M Wang, **F Du**. (2013). "Preparation and Test of Fe/Cr/Al-Ni/Cr composite Film Micro-heater". Sensor & Transducers, 156(9), 317-323.