

MULTI-CHANNEL DYNAMIC-RANGE COMPRESSION TECHNIQUES
FOR HEARING DEVICES

by

Ziyan Zou



APPROVED BY SUPERVISORY COMMITTEE:

Dr. Issa M. S. Panahi, Chair

Dr. P. K. Rajasekaran

Dr. Mehrdad Nourani

Copyright 2018

Ziyan Zou

All Rights Reserved

I dedicate this thesis to my family and friends for their partnership and love

MULTI-CHANNEL DYNAMIC-RANGE COMPRESSION TECHNIQUES
FOR HEARING AID APPLICATIONS

by

ZIYAN ZOU, BS

THESIS

Presented to the Faculty of
The University of Texas at Dallas
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN
ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2018

ACKNOWLEDGMENTS

One year ago, I made a decision to join a research lab in my master's program. It was a great journey that dramatically influenced my attitudes towards life, study and research. There are many people who have contributed for making this journey towards my Master's degree a memorable one. I am grateful to every person who has impacted my life in a positive way.

First, I would like to express my sincere gratitude to my advisor, Dr. Issa M. S. Panahi, for providing an opportunity to study and do research in the Statistical Signal Processing Research Laboratory (SSPRL) and for his valuable mentorship throughout my research study.

I am thankful to Dr. P. K. Rajasekaran and Dr. Mehrdad Nourani for serving on my committee. I appreciate their time and efforts.

I thank my parents for their unlimited love and support, without which my accomplishments would have been impossible.

I also thank all the members of SSPRL: Dr. Anshuman Ganguly, Dr. Chandan Karadagur Ananda Reddy, Yiya Hao, Abdullah Kucuk, Nikhil Shankar, Gautam Shreedhar Bahat, Serkan Tokgoz and Parth Mishra for all the valuable discussions and collaborative efforts.

I would like to thank my friends in the US, Jingchen, Yoyo, Xinyi, and Taotao for their understanding and support during tough times.

This work was supported by NIDCD Institute of the National Institutes of Health under award number 5R01DC015430. The content of this thesis does not necessarily represent the official views of the National Institutes of Health

August 2018

MULTI-CHANNEL DYNAMIC-RANGE COMPRESSION TECHNIQUES FOR HEARING AID APPLICATIONS

Ziyan Zou, MSEE
The University of Texas at Dallas, 2018

Supervising Professor: Dr. Issa M. S. Panahi

Dynamic-range compression (DRC) is an important component in hearing aid devices (HADs). Research of multi-channel DRC design and real-time implementation has been carried out in the last few decades. The trade-offs of every DRC system include the frequency resolution, computational complexity and processing time delay. In this thesis, a crossover filter bank based nine-channel DRC with an optimized structure is proposed. A polyphase extension of this approach is then applied and a compensation filter is proposed to reduce the distortion. Then a subband filter bank is designed and implemented for multi-channel DRC to reduce the computational complexity. The final method to optimize a multi-channel DRC is to use Equalization technique in frequency domain, which further reduces the computational complexity. In this work, all the methods are implemented on a smartphone to work as an assistive device to hearing aids. Objective and subjective evaluation of the developed methods show the improvement in quality and intelligibility.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	v
ABSTRACT.....	vi
LIST OF FIGURES.....	ix
 CHAPTER 1 INTRODUCTION	 1
1.1 Motivation.....	1
1.2 Dynamic-range Compression.....	2
1.2.1 Single-channel Compression	2
1.2.2 Multi-channel DRC	5
1.2.3 Hearing Aid Devices & Smartphone Platform	6
1.3 Overview of the proposed solutions	7
 CHAPTER 2 MULTI-CHANNEL CROSSOVER FILTER BANK BASED DYNAMIC- RANGE COMPRESSION RUNNING ON SMARTPHONE.....	 9
2.1 Overview.....	9
2.2 Proposed Method.....	10
2.2.1 9-channel Filter Bank.....	11
2.2.2 Dynamic-range Compressor	12
2.2.3 Delay Aligner.....	13
2.2.4 Volume Limiter.....	14
2.3 Real Time Implementation.....	14
2.4 Test Result.....	16
2.4.1 Computational Time Measurement.....	16
2.4.2 Objective Evaluations	17
2.4.3 Subjective Test.....	18
2.5 Summary	20
 CHAPTER 3 COMPENSATED MULTI-CHANNEL DYNAMIC-RANGE COMPRESSION WITH POLYPHASE IMPLEMENTATION.....	 21
3.1 Overview.....	21
3.2 Proposed Method.....	22

3.2.1	9-channel filter bank.....	23
3.2.2	Compensation Model	24
3.2.3	Polyphase Implementation	26
3.2.4	Other Parts	27
3.3	Test Results	28
3.3.1	Objective Evaluation.....	28
3.3.2	Subjective test results.....	29
3.3.3	Computational Time measurement.....	31
3.4	Summary	31
CHAPTER 4 NON-UNIFORM COSINE MODULATED FILTER BANK USING FREQUENCY RESPONSE MASKING APPROACH.....		33
4.1	Overview	33
4.2	Proposed Method.....	34
4.2.1	Prototype filter	34
4.2.2	Uniform and non-uniform CMFB.....	37
4.3	Real-time Implementation.....	40
4.4	Test Results	42
4.4.1	Objective Test.....	42
4.4.2	Subjective Test.....	44
4.5	Summary	45
REFERENCES.....		46
BIOGRAPHICAL SKETCH.....		49
CURRICULUM VITAE		

LIST OF FIGURES

Fig. 1.1. Input/output relationship for a typical hearing-aid single channel compression.....	3
Fig. 1.2. The relationship of changes in input, gain and output in a compression system	4
Fig. 2.1. Block diagram of the <i>Crossover FB DRC</i>	10
Fig. 2.2. 9-channel crossover filter bank.....	11
Fig. 2.3. Block diagram of dynamic-range compressor.....	12
Fig. 2.4. Block diagram of frame-based processing	15
Fig. 2.5. Time Measurement Results	16
Fig. 2.6. Objective evaluations of PESQ and STOI under three noise type at different SNRs.....	18
Fig. 2.7. Subjective test setup.....	19
Fig. 2.8. MOS test results	20
Fig. 2.9. Word Recognition test results	20
Fig. 3.1. Block diagram of the proposed method	22
Fig. 3.2. 9-channel filter bank at sampling frequency of 48kHz.....	23
Fig. 3.3. white noise before and after the filter bank	24
Fig. 3.4. Compensation Model	25
Fig. 3.5. Polyphase realization of FIR filter	26
Fig. 3.6. Objective evaluation of speech quality and intelligibility.....	29
Fig. 3.7. Subjective test results.....	30
Fig. 4.1. Block diagram of the proposed method	34
Fig. 4.2. The principle of frequency response masking approach.....	35
Fig. 4.3. Proposed prototype filter bank.....	36

Fig. 4.4. M-channel decimated uniform CMFB	37
Fig. 4.5. M' -channel non-uniform CMFB	38
Fig. 4.6. 16-channel uniform CMFB and 9-channel non-uniform CMFB.....	39
Fig. 4.7. Distortion model of the <i>9-channel</i> FB and non-uniform CMFB.....	40
Fig. 4.8. Audio processing of android operating system using Superpowered SDK	41
Fig. 4.9. Objective results of PESQ and STOI.....	43
Fig. 4.10. Results of MOS subjective test.....	44

CHAPTER 1

INTRODUCTION

1.1 Motivation

Normal hearing people have hearing thresholds of 25 dB or better in both ears. People who are not able to hear as well as someone with normal hearing are regarded to have hearing loss. According to the World Health Organization (WHO), there are around 466 million people around the world having disabling hearing loss. And this number estimates to reach 900 million by 2050 [1]. Recent statistics reported by the National Institute on Deafness and other Communication Disorders (NIDCD) show that in the U.S., 15% of adults over the age of 18 (37.5 million people) have hearing problems, 25% of adults from aged from 65 to 74 report disabling hearing loss, and 50% of adults over the age of 75 have disabling hearing loss [2]. There are different reasons resulting in hearing problems. In addition to genetic issues, complications at birth and the use of particular drugs, people may gradually lose their hearing as they age. Moreover, hearing loss can also be caused by exposure to loud noise and the increased usage of headsets and earphones. The degree of hearing loss can be mild, moderate, severe or profound.

To improve hearing, people with hearing loss can use hearing devices, such as hearing aids (HAs), Cochlear implants (CIs) and other assistive devices [3]. HAs are small electronic devices consisting of a microphone, amplifier and speaker. The microphone picks up the sounds around and converts the sounds to electrical signals. The electrical signals are amplified and then delivered to the ear. HAs nowadays are able to adjust the sound amplification based on users' hearing needs. CIs tremendously benefit people with severe to profound hearing loss. There are two parts of CIs:

a microphone/processor module that locates behind the ear, and a receiver which requires surgery to implant under skin. CIs work directly with the auditory nerve and brain.

1.2 Dynamic-range Compression

Modern digital hearing aid devices (HADs) have speech enhancement, noise reduction, echo cancellation and dynamic-range compression (DRC) in the pipeline [3].

1.2.1 Single-channel Compression

The input and output relationship for a typical hearing-aid compression amplifier [4] is shown in Fig. 1.1. The point where the input and output relationship changes is called the threshold kneepoint (TK). For lower TK is usually 40-50 dB and the upper TK is usually 85-100 dB. If the input sound pressure level (SPL) is below the lower TK, the system is of a linear amplifier. If the input SPL is above the upper TK, the output SPL will remain the same no matter how the input SPL changes, this is called compression limiting. If the input SPL is between the lower TK and upper TK, the system will provide the dynamic-range compression (DRC).

The compression ratio (CR) reflects how much compression will applied is a change when given the change in the input sound pressure level (SPL). CR can be given as:

$$CR = \frac{\Delta Input}{\Delta Output}$$

Where, $\Delta Input$ is the change of input SPL and $\Delta Output$ is the change of output SPL. For example, in Fig. 1.1 the CR is 1, which means that the input SPL increase from 40 to 60 dB, the output SPL will increase from 80 to 100 dB.

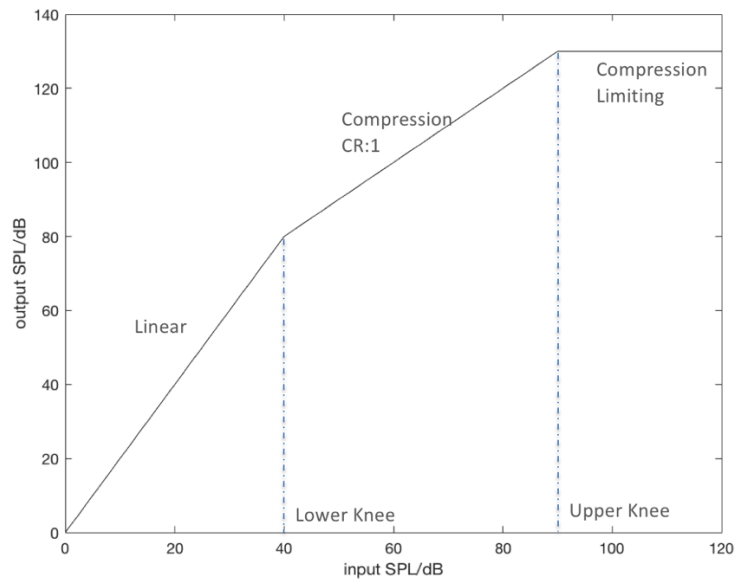


Fig. 1.1. Input/output relationship for a typical hearing-aid single channel compression

Another pair of characteristics of compression is attack time (AT) and release time (RT). When the SPL of the input signal suddenly changes, the DRC is unable to change the gain immediately. The AT and RT reflects the amount of time that the system requires to respond to a sudden change in the input SPL. Fig. 1.2 gives a brief representation of how the changes in input SPL will influence the gain and output SPL. AT represents the delay for time between when the input SPL changes from below the TK to above it to activate the compression and the gain reduces to its target level, and RT is the time delay between the time when the input SPL reduces to a level below the TK and the time that gain increases to its target value. The American National Standards Institute (ANSI) gives the definition of AT as the time between an abrupt increase in input SPL from 55 to 90 dB and the point that the output SPL reaches and stays within 3dB of the steady value for an input SPL of 90 dB. Also ANSI defines that RT as the interval between an abrupt drop in the input SPL from 90 to 55 dB and the point that the output SPL reaches and stays within 4dB of the steady value for the input SPL of 55 dB.

We can see from Fig. 1.2 that at T1, when the input SPL increase from a level below TK to a level above TK, the compression is activated, but the gain of the system does not change instantaneously so that there is a overshoot in the output SPL. When the gain gradually reduces to reach its target, the output approaches within 3dB of the target value. At time T2, the input SPL decreases to a level below TK, but the gain of the compression system does not change instantaneously so that there is a undershoot in the output SPL at time T2. When the gain gradually increases to reach its target, the output reaches within 4dB of its final level.

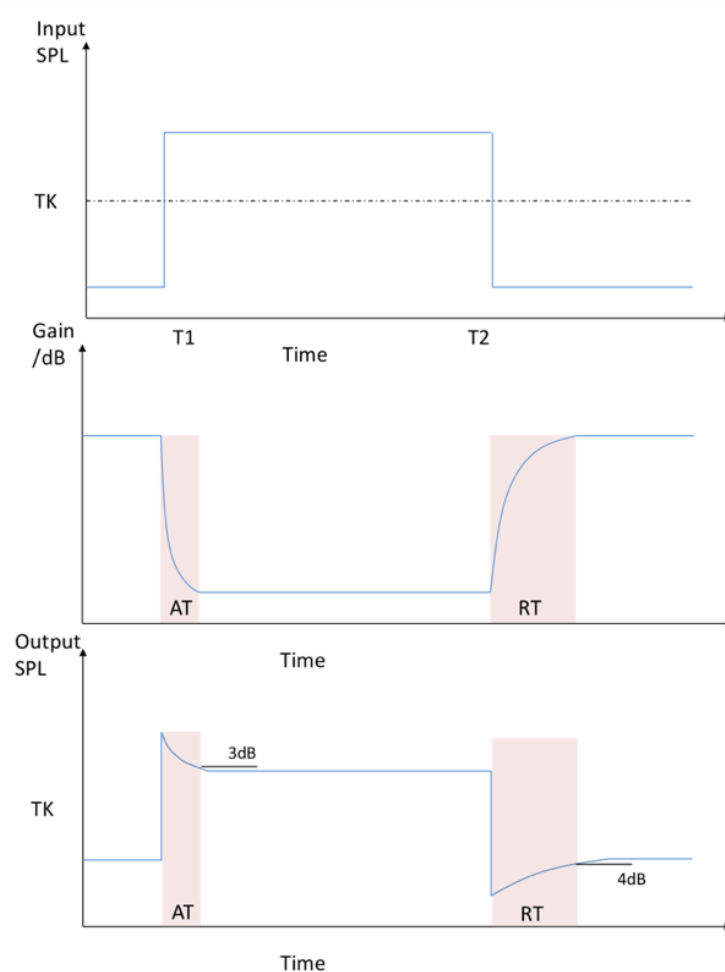


Fig. 1.2. The relationship of changes in input, gain and output in a compression system

Single-channel compression limiters restrict high-level sound to be lower than the uncomfortable loudness (UCL) of hearing impaired people [5]. Since single-channel compression systems calculate the amount of gain at a particular frequency based on the entire spectral content, the spectral shape of the input signal is supposed to be of long-term. Unfortunately, speech signals are mainly of short-term spectral shape and dominated by low-frequency components. Thus, single-channel compression could not provide effective compression ratio for high-frequency speech components. Additionally, these compressors fail to provide various compression levels in different frequency bands, hence, they can negatively influence the quality and intelligibility of output signals. Most HADs use multi-channel compression to offer effective solutions for those shortcomings.

1.2.2 Multi-channel DRC

A multi-channel compressor should be designed to match the system frequency resolution to the resolution of the auditory system of people with hearing loss. Multi-channel compressors calculate compression parameters such as CR, AT, and RT independently in each frequency band [6]. These parameters are controlled by the audio level of corresponding frequency band and could be adjusted based on the audiogram of people with different hearing losses, which would make customization of HADs possible.

One easy way to realize multi-channel system is using filter bank. Individual filters in the filter bank separate the input signal into multiple frequency bands. A N -channel filter bank [7] is a combination of N filters. The first one is a low-pass filter for the lowest frequencies, the last one is a high-pass filter for the highest frequencies, and $N-2$ band-pass filters locate in the middle. For an ideal filter bank, each filter is supposed to have the same phase response so that the output of

each filter would be in phase the output of any other filters at every frequency. In this case, without the phase interactions, we can obtain the composite output by simply adding the magnitudes of the outputs of all channels. Linear-phase FIR filters are widely used to realize this ideal. In a Linear-phase FIR filter, the output does not depend on the past output samples and the coefficients for all zeros, it only depends on the input samples. The FIR filter can be given as:

$$y(n) = \sum_{k=0}^K b_k x(n - k) \quad (1.1)$$

Where, k is the delayed samples and b_k is the weight.

A linear-phase FIR requires the set of coefficients $\{b\}$ has to be even, for example, $b_0 = b_K$, $b_1 = b_{K-1}$, etc. The Filter delay is then $K/2$ samples at all frequencies.

The FIR filters may have a large amount of ripples in their stop bands because of the limited filter length. Increase in filter length results in reduction in stop band ripples and the reduction of the transition width at filter band edges.

An ideal multi-channel filter bank is supposed to not introduce distortion to the system. In other word, if the separate bands are to be mixed back together, then the ideal audio crossover would split the input audio signal into separate bands that do not overlap or interact with each other to make sure the output signal unchanged in frequency, relative levels and phase response. However, the ideal condition is hard to realize in real situations.

1.2.3 Hearing Aid Devices & Smartphone Platform

HADs are expensive, especially those with additional hardware and advanced technology. Most of the hearing-impaired people need two devices: one hearing devices and an assistant tool. And depending on the customer demands, the price of HADs could vary from \$ 1000 to \$4000 for each

device, which means a great financial burden even for normal families with median income. Nowadays, some HADs manufactures use a pen or necklace to provide the additional microphone for capturing audio with higher signal-to-noise ratio (SNR) and wirelessly transmit the processed signal to HADs. However, these external devices are auxiliary and not portable.

Most people today have personal smartphone and carry them around. Smartphone could be a great platform for audio capture, complex computation and wireless data transmission. Using smartphone as an assistant tool for HADs is a good idea and no additional expense will be required for the hardware. Recently, there are many new HA applications coming up in both Apple and Android system based smartphones, for example, the Live Listen in iPhone. Most of these smartphone-based HA applications use single microphone to avoid latencies between input audio and output audio.

1.3 Overview of the proposed solutions

In this thesis, we will first introduce a multi-channel crossover filter bank based dynamic-range compression and its real time implementation in a smartphone in Chapter 2. And the 9-channel crossover filter bank, compressor, aligner and the volume limiter will be introduced in detail. We will refer this method as *Crossover FB DRC*. In Chapter 3, a compensation model is built to compensate the distortion caused by the multi-channel filter bank and the polyphase implementation will be applied to the *Crossover FB DRC* for reducing processing time. In Chapter 4, a subband based 16-channel uniform filter bank and 9-channel non-uniform filter bank will be introduced and use for multi-channel DRC. The 9-channel non-uniform filter bank will be implemented in a smartphone.

My contribution to this thesis are: In the project in Chapter 2, I designed and conducted the subjective test and analyzed the test results; In the project in Chapter 3, I developed the algorithm, realized the frame-by-frame processing in Matlab, obtained objective results and conducted subjective tests; In the project in Chapter 4, I developed the algorithms, realized the frame-by-frame processing in both Matlab and C, measured the objective evaluation and conducted subjective tests.

CHAPTER 2

MULTI-CHANNEL CROSSOVER FILTER BANK BASED DYNAMIC-RANGE COMPRESSION RUNNING ON SMARTPHONE

2.1 Overview

A multi-channel system could be developed using a filter bank, which consists of multiple filter of different frequency bands. A filter bank can separate an audio signal into several filter bank as each filter passes a specific frequency band of input audio signal. The multi-channel filter bank and compressor are supposed be designed to match the system frequency resolution to the auditory resolution of people with hearing loss. The N -channel filter bank is a combination of N FIR filters. The first one is a low-pass FIR filter, the last one is a high-pass FIR filter, and $N-2$ band-pass filters locate in middle. Multi-channel compressors calculate and adjust compression parameters such as CR, AT, and RT independently in each frequency band [8] according to the band-specific compressor requirement and the audiogram of people with different hearing losses. For example, for a hearing-impaired listener with hearing loss at lower frequency bands and normal hearing at the high frequency area, the multi-channel DRC provides greater CR value in lower frequency bands to avoid over compression and normal CR at high frequency area.

In this work, a 9-channel dynamic-range audio compression system in real-time is proposed. We refer to the 9-channel crossover filter bank as *9-channel FB*, and the 9-channel dynamic-range audio compression system as *Crossover FB DRC*. The *Crossover FB DRC* consists of three modules. The first module contains Kaiser filter based nine-channel crossover filter ban. In the second module, nine compressors with different parameters are developed to squeeze the dynamic range of input audio signal in each frequency band. The third module compensates the group delay

of each channel by using an aligner, sums all the outputs up together, and then uses a limiter to keep the volume of audio signal under a hard threshold.

The proposed *Crossover FB DRC* is frame-based and running in real-time for practical application. In this implementation, the frame size is 10 milliseconds, which equals to 160 samples based on 16 kHz sampling frequency. It is implemented on an Android smartphone, Google Pixel One under Android 6.0.1 operating system. Objective tests were conducted for machinery noise, traffic noise and multi-talker babble noise mixed with the clean speech at SNR levels of -10dB, -5dB, 0dB, 5dB and 10dB. During the subjective tests, the subjects provide commendable response in same noise types at SNR of -5dB and 0dB. The proposed method was compared with a commercial audio compression & limiter provided by Hotto Engineering [9] running on a laptop. The scores show significant improvements in both quality and intelligibility with the proposed method.

2.2 Proposed Method

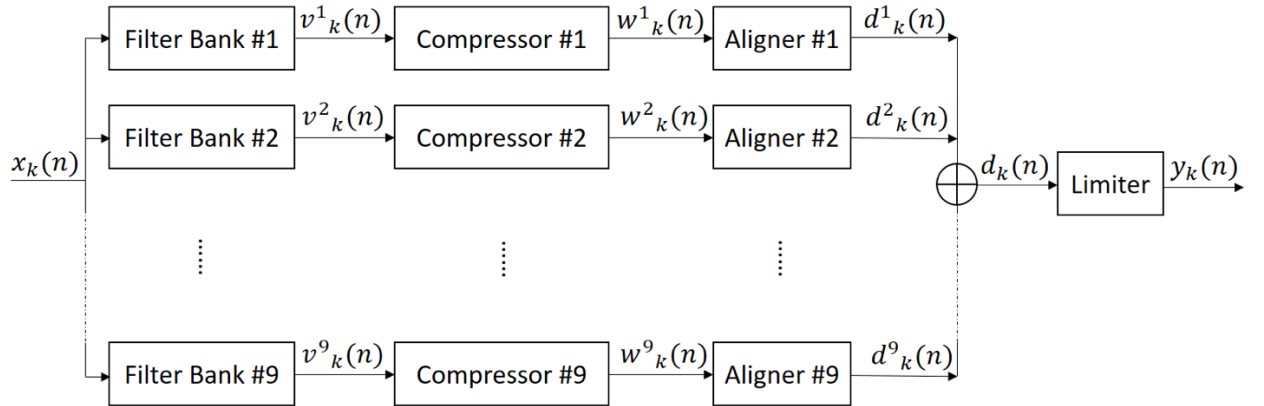


Fig. 2.1. Block diagram of the *Crossover FB DRC*

The block diagram of the proposed *Crossover FB DRC* shows in Fig. 2.1. The system runs in real time by frame-based processing. $x_k(n)$, the k^{th} input frame with length of L , first convolves with each filter in the filter bank. The outputs from filter bank are separated audio in different frequency bands, represented by $v_k^i(n)$, where $i = 1, 2, \dots, 9$, representing the i^{th} channel. Then the signals from nine channels process with nine compressors with different parameters. And the outputs from compressors, $w_k^i(n)$, are aligned up by the aligners to compensate the group delay of each filter to get $d_k^i(n)$. To get the complete audio signal back $d_k(n)$, the outputs of each channel are simply summed up. At last, a limiter is added to control or limit the maximum output of the system.

2.2.1 9-channel Filter Bank

The *9-channel FB* consists of one low pass FIR filter, one high pass FIR filter, and seven band pass FIR filters locating between the low pass filter and the high pass filter. All the filters are linear phase. The details of the filter bank are shown in the following Table 2.1 and Fig. 2.2.

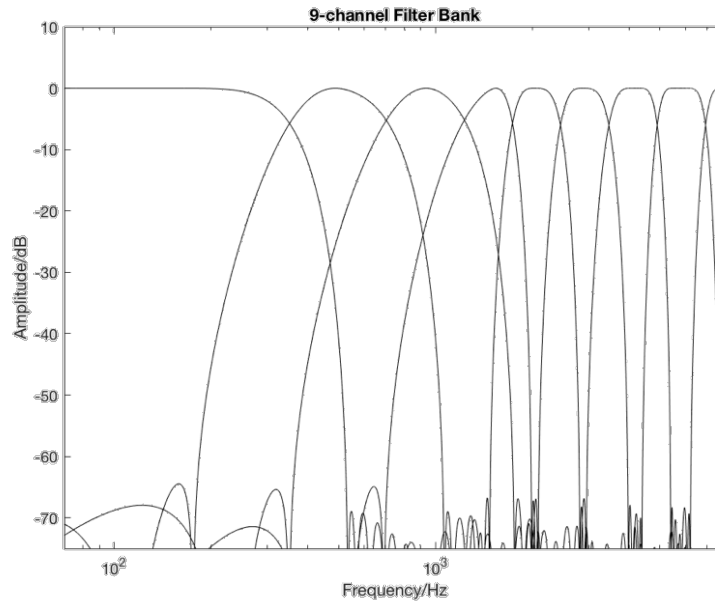


Fig. 2.2. 9-channel crossover filter bank

Table 2.1. 9-channel Filter Bank									
Channel	1	2	3	4	5	6	7	8	9
Center Frequency/Hz	/	500	1000	1500	2000	3000	4000	6000	/
Bandwidth/Hz	350	270	410	390	565	815	1220	1790	/
Filter Order	576	604	338	416	478	338	284	246	146

All the FIR filters are windowed with Kaiser window (1) to remove the ripples.

$$w(m) = \begin{cases} \frac{I_0[\alpha \sqrt{1 - (\frac{2t}{M-1} - 1)^2}]}{I_0[\alpha]}, & 0 \leq m \leq M-1 \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

Where, M is the length of the window, I_0 is the 0th-order modified Bessel function of the first kind. α is a non-negative real number that determines the shape of the window by playing with the trade-off between main-lobe width and side lobe level in the frequency domain.

2.2.2 Dynamic-range Compressor

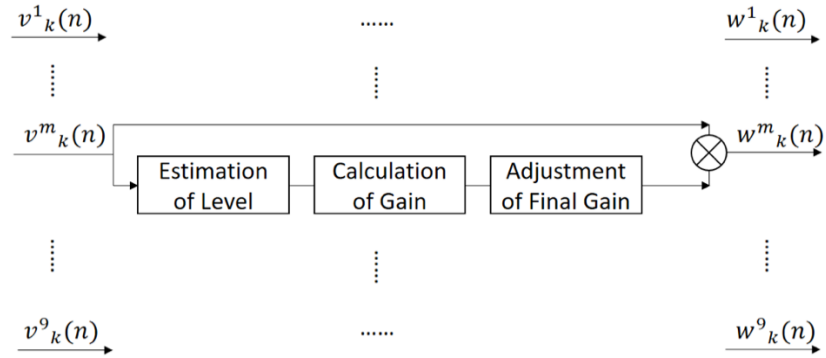


Fig. 2.3. Block diagram of dynamic-range compressor

Fig. 2.3 shows the block diagram of the dynamic-range compressor [10]. The $v^m_k(n)$ is the k^{th} output frame of the m^{th} channel of the filter bank. In each channel, a specific compressor is designed and applied after the filter. The compressor initially estimates the power of input signal

per sample, and then calculates the gain according to the signal level and the compressor ratio. If the signal level is greater than the preset threshold knee, the compression ratio is relatively larger, while if the signal level is smaller than the threshold knee, the compression ratio is smaller. The gain then will be adjusted based on the preset attack time and release time. And the final gain is determined to be a scaling factor. Finally, the input, $v^m_k(n)$, will be multiplied by the final gain to generate the compressed audio of m_{th} frequency band, $w^m_k(n)$.

2.2.3 Delay Aligner

After the compressor, the outputs $w^m_k(n)$ of all nine channels signals need be combined together to get the output audio signal. However, each FIR filter produces group delays and different filter have different amount of group delay [11].

$$V^m_k(\omega) = X_k(\omega)H^m(\omega) = X_k(\omega)A^m(\omega)e^{-jD_m\omega} \quad (2.2)$$

Where, $X_k(\omega)$ is the k^{th} input frame in frequency domain, $V^m_k(\omega)$ is the k^{th} output frame of m^{th} channel. $H^m(\omega) = A^m(\omega)e^{-jD_m\omega}$ represents the m_{th} FIR filter. D_m stands for the group delay of the m^{th} FIR filter. The group delay could be presented as:

$$D_m = \frac{L_m - 1}{2} \quad (2.3)$$

Where, L_m is the length of the m_{th} FIR filter. But since the FIR filters of the 9 channels are different from each other, the group delay varies. Before combining the outputs of all channels, the aligner will compensate the group delay in each channel by simply shifting the output samples. More details will be given in real-time implementation in Section 2.3.

2.2.4 Volume Limiter

Before the complete audio signal $d_k(n)$ is playing out, a volume limiter is applied to remove the audio samples with unexpected high levels. The compressor could reduce most of the high levels audio samples, but when DRC fails or the samples are of extremely intense level, the volume limiter will sharply cut the volume down to a safe range. The volume limiter works as:

$$\begin{cases} y_k(n) = d_k(n), & d_k(n) \leq \beta \\ y_k(n) = \beta, & d_k(n) > \beta \end{cases} \quad (2.4)$$

Where, β is the volume threshold. As long as the output $d_k(n)$ is below threshold β , the output $y_k(n)$ is equal to $d_k(n)$. But when $d_k(n)$ exceed the maximum output of the system, the peaks of the output are clipped to be β .

2.3 Real Time Implementation

For real time implementation, frame-based processing is required. In this work, the frame size is chosen to be 160 samples, which is 10 milliseconds when the sampling frequency is 16 kHz. This algorithm is implemented in an Android smartphone, Google Pixel 1 with Android 6.0.1 operating system. Since each frame is 10ms, the processing time for one frame is also supposed to be within 10ms. If the processing time runs beyond 10ms, the output audio signal will be distorted because of skipping frames. Java Native Interface (JNI) is used to meet the requirement of fast processing on Android smartphone. JNI uses C instead of Java as program language. There are two reasons for C works better than Java: firstly, C locates at lower layer in Android system so that it saves additional time of data transmission crossing different layers; secondly, C processes faster than Java in Android application layer. Moreover, an optimized rapid convolution function written in

C is used to further reduce the processing time. Fig. 2.4 shows the block diagram of the frame-based processing in detail.

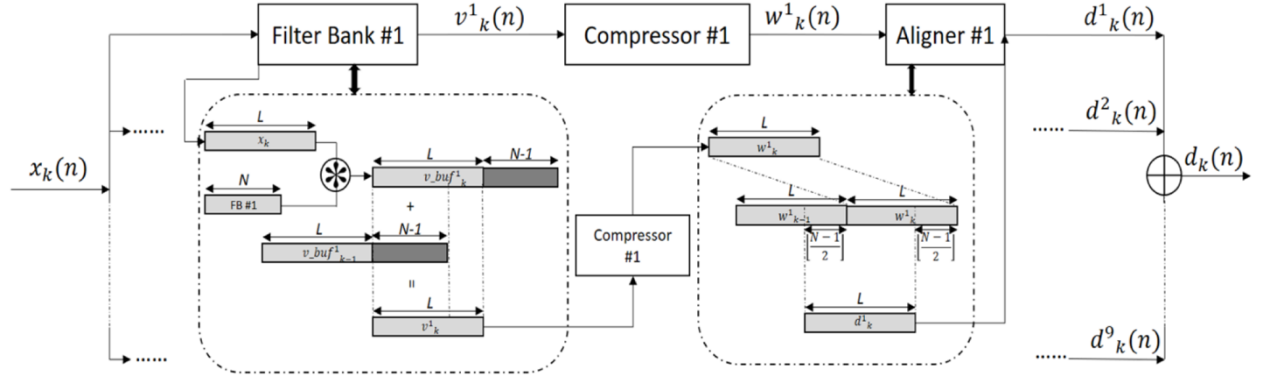


Fig. 2.4. Block diagram of frame-based processing

For frame-based convolution, we can take the first channel as example. When the k^{th} input frame x_k with the length of L convolves with the first filter with the length of N , the output is of length $L+N-1$ and will be save in buffer $v_buf_k^1$. The first L samples from the k^{th} buffer $v_buf_k^1$ then combine with the last $N-1$ samples, which is called tail, from the $(k-1)^{th}$ buffer $v_buf_{k-1}^1$, to get the final output v_k^1 of the current k^{th} frame.

The group delay compensation is different for real-time processing when comparing to offline processing. In the case of offline simulation, the whole input audio samples are known. The group delay can be easily compensated by shifting back $\left\lfloor \frac{N_m-1}{2} \right\rfloor$ samples and padding zeros to the end. However, in the case of real-time processing, only the previous and current audio samples are known. To achieve the aligner, one more buffer is added to make the length of $2L$. For example, the output of k^{th} frame after compressor stores in buffer w_k^1 with the length of L , and the output of the previous frame saves in buffer w_{k-1}^1 . Two buffers work in series. To compensate the delays

from the first FIR filter with length N , $\left\lfloor \frac{N-1}{2} \right\rfloor$ samples are shifted back to get current aligned frame output d_k^1 .

2.4 Test Result

In this section, we measure the time of the real-time processing of the proposed method on Android smartphone. Objective evaluations and subjective tests are conducted to measure the quality and intelligibility of the output audio signal processed by the proposed method.

2.4.1 Computational Time Measurement

Two kinds of processing times are measured: round-trip latency and algorithms processing time. Larsen test is used to measure the round-trip latency [12]. The round-trip latency T_1 is the time between microphone receiving signal and wired earbuds getting signal. This whole process includes analog to digital converting (A/D), DRC system processing and digital to analog converting (D/A). The algorithms processing time T_2 was measured in C using library Time.

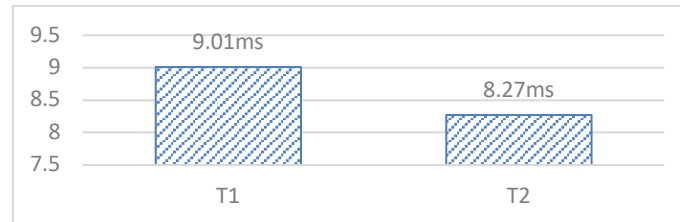


Fig. 2.5. Time Measurement Results

Fig. 2.5 shows the processing time measurements results. The average processing time T_2 is 8.27 ms, which is within the frame size 10ms. The average round-trip latency T_1 is 9.01ms. So we can tell that the processing time satisfies the requirement of real-time implementation.

2.4.2 Objective Evaluations

We evaluate the performance of the proposed *Crossover FB DRC* method by comparing with one method of single-channel DRC. We choose a laptop-based commercial audio compression & limiter provided by Hotto Engineering [9] as comparison.

The evaluations are carried out for 10 sentences of 3 seconds long from HINT [12] database. The objective evaluations are performed for 3 different noise types: multi-talker babble, machinery and traffic noises. These noises were recorded in real-world conditions in several different scenarios. The presented results are average of them all. The microphone signal is generated by adding the speech and noise together at SNR of -10, -5, 0, 5, and 10dB.

The perceptual evaluation of speech quality (PESQ) [13,14] is used as objective evaluation criteria for speech quality and short time objective intelligibility (STOI) [15, 16] is used to measure speech intelligibility. STOI is more recent and widely used because of its accuracy in many speech distortion types. PESQ ranges between 0.5 and 4, with 4 being high perceptual quality while 0.5 represents the worst quality. The higher the score of STOI, the better the speech intelligibility.

Fig. 2.6 shows the results of PESQ and STOI comparison between the proposed method and the comparison Hotto DRC. According to the PESQ results, the proposed *Crossover FB DRC* works better in three noise type under 5 and 10dB. At lower SNR levels, the PESQ scores for both methods are similar. The STOI curve of machinery and traffic noise are similar, under -5 and 0dB the proposed method provides better results. However, for babble noise, the STOI score for proposed method is slightly higher than Hotto DRC.

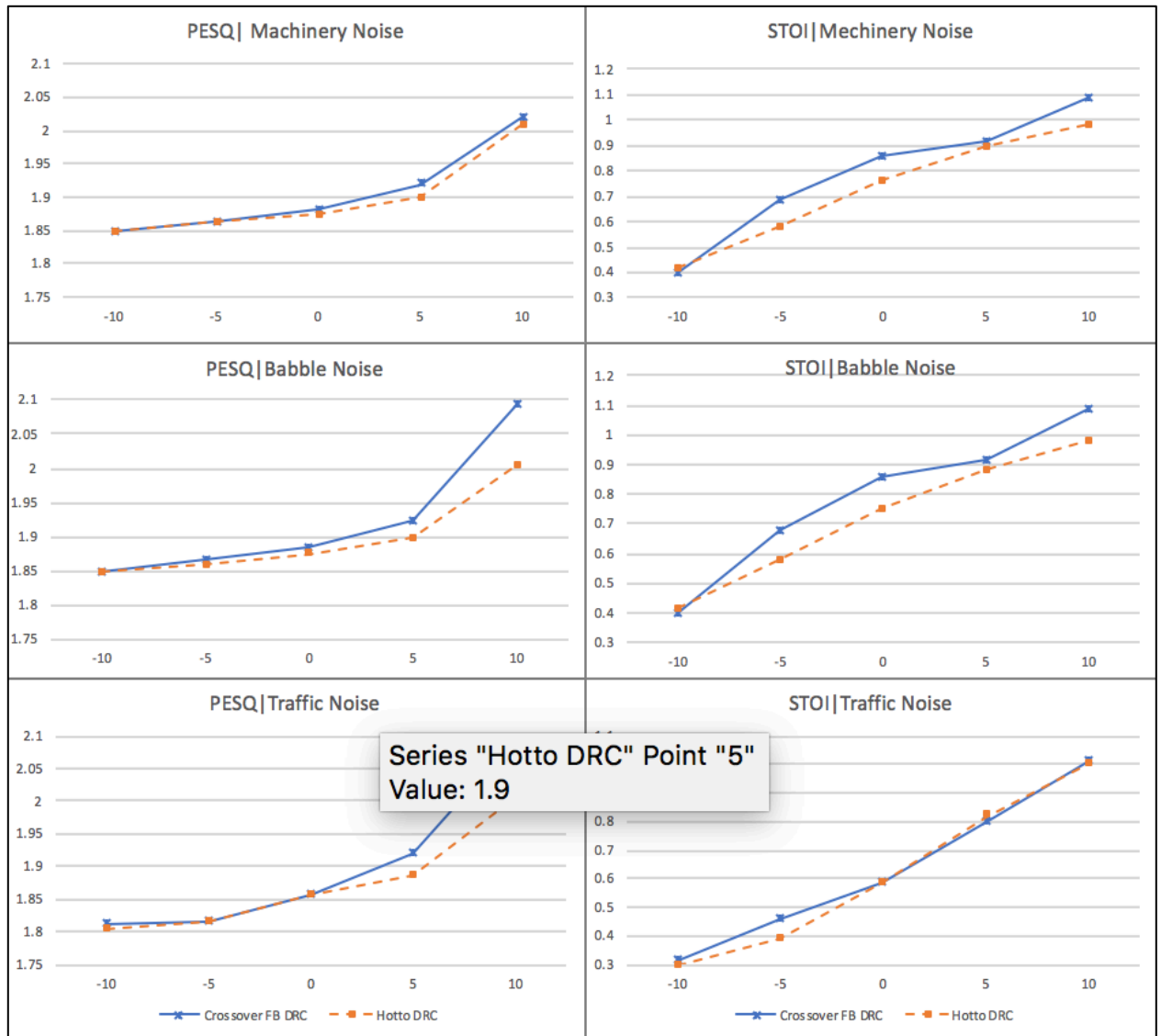


Fig. 2.6. Objective evaluations of PESQ and STOI under three noise type at different SNRs

2.4.3 Subjective Test

We performed subjective tests 10 normal hearing participants aged from 19 to 28 years old to illustrate the usefulness of the developed application in the real-world noisy environmental conditions. We emulated the noisy restaurant condition in our lab. The setup is shown in Fig. 2.7. The subject sits between two loud speakers, with one plays the clean speech in the foreground and

another plays restaurant noise in the background. The smartphone running the proposed method is set to be 3ft away from the front speaker and 4.5ft away from the back speaker. The subject can hear the output audio signal from smartphone via Bose Soundsport earphone.

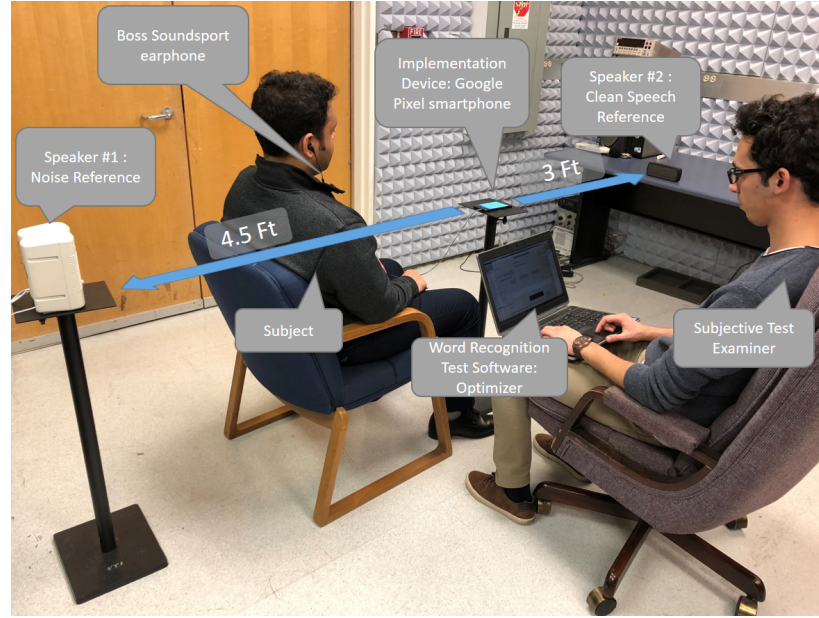


Fig. 2.7. Subjective test setup

Mean Opinion Score (MOS) conducts by presenting the audio signal processed by Hotto DRC and the propose *Crossover FB DRC*. The dataset is clean speeches from 3 lists of 10 sentences each from TIMIT database in the presence of restaurant noise at SNR levels of -5 and 0 dB. The subjects are supposed to give scores range from 1 to 5, where 5 stands for excellent and 1 stands for poor [17,18]. Fig. 2.8 reflects the average scores of 10 normal hearing subjects. The results show that at SNR of both -5 and 0 dB, the propose *Crossover FB DRC* provides better quality according to the subjects.

Word Recognition (WR) test is conducted under the same condition as MOS tests. A GUI based interface is used for the WR, more details of this tool and WR test procedures are presented in [19]. The score reflects the percentage of words that the subject recognizes correctly over the total

words in the sentence. The experimental result based on the average of 10 subjects presents in Fig. 2.9. The proposed *Crossover FB DRC* shows WR rate of almost 70% while Hotto DRC provides only 52% at SNR of -5 dB. At SNR of 0 dB, proposed *Crossover FB DRC* show WR rate over 90%.

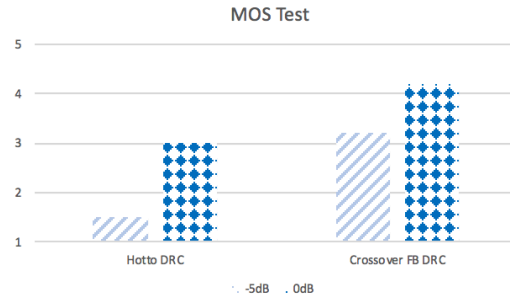


Fig. 2.8. MOS test results

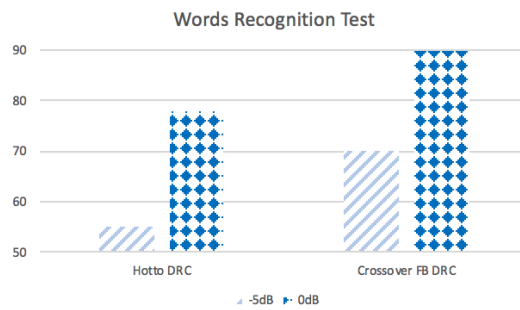


Fig. 2.9. Word Recognition test results

2.5 Summary

In this work, the crossover filter bank based *Crossover FB DRC* is proposed and implemented in Smartphone running in real time. Comparing to a single-channel Hotto DRC, which is running in laptop, the propose *Crossover FB DRC* provides better performance in both quality and intelligibility according to objective and subjective tests. Proposed method also provides a portable audio framework, which is not just limited in current version of DRC, but also could be extended or upgraded for further research study.

CHAPTER 3

COMPENSATED MULTI-CHANNEL DYNAMIC-RANGE COMPRESSION WITH POLYPHASE IMPLEMENTATION¹

3.1 Overview

An ideal multi-channel filter bank is supposed to not introduce distortion to the system. In other words, if the separate bands are to be mixed back together, then the ideal audio crossover would split the input audio signal into separate bands that do not overlap or interact with each other to make sure the output signal unchanged in frequency, relative levels and phase response. However, in real world, this ideal performance can only be approximated. In the crossover 9-channel FB proposed in Chapter 2, the frequency bands are overlapped with adjacent ones. The ripples of the filters and the overlaps of the filters frequency bands both contribute to undesirable distortion to the output audio. To measure and attenuate the distortion caused by the *9-channel FB*, a compensation model is built in this work. White noise with zero-mean and unit-variance is used to build the model. The input audio signal, instead of directly convolving with the *9-channel FB*, will go through the compensation model to reduce the distortion of the final output audio. The attenuation of distortion is supposed to improve the quality and intelligibility of output audio of the DRC system.

The trade-offs among frequency resolution, computational complexity, processing time delay and system distortion are inevitable in every compression system. The increase of the number of

¹ © 2018 IEEE. Adapted, with permission, from Z. Zou, Y. Hao, I. M. S. Panahi, “Design of Compensated Multi-Channel Dynamic-Range Compressor for Hearing Aid Devices Using Polyphase Implementation,” 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 2018.

channels also results in the computational complexity, which limits the sampling rate and results in systems latency. It makes real-time realization more difficult. Polyphase implementation [20] is a multi-rate signal processing technique. It can be used to deal with constraints of having high sampling rate and low computational complexity. In this proposed work, the FIR filter of each channel is decomposed into 4 sub-channels for polyphase implementation. This approach benefits the real-time operation of the filter bank since it simplifies the computation in each band and reduces filter bank delay.

3.2 Proposed Method

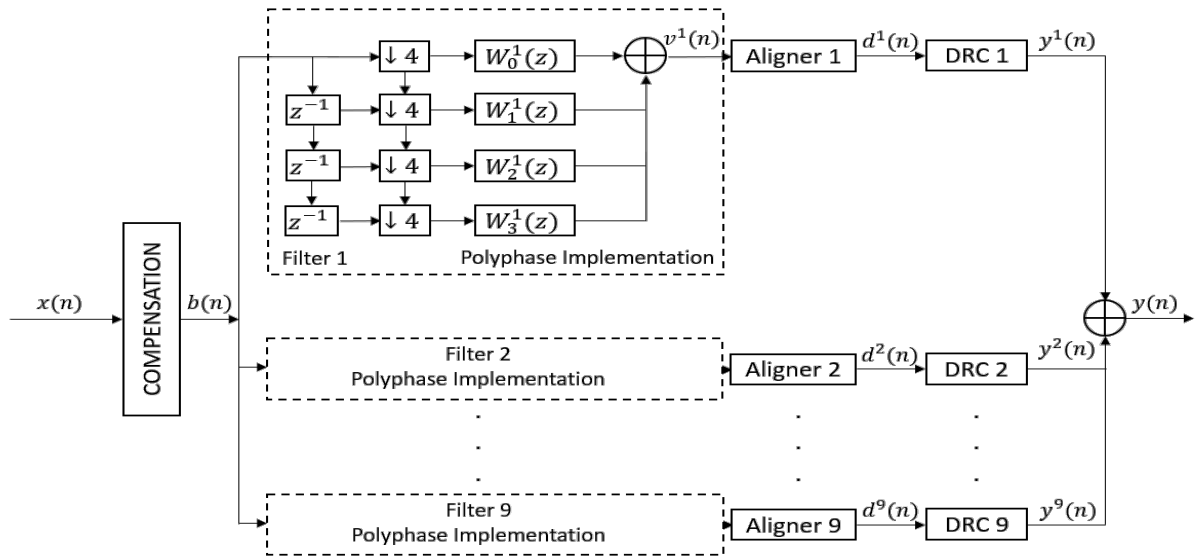


Fig. 3.1. Block diagram of the proposed method

For this work, we are based on the *Crossover FB DRC* proposed in Chapter 2. A filter based on compensation model is added before the filter bank to reduce the distortion caused by a filter bank. Furthermore, we realize each filter using polyphase approach in order to reduce the processing time. The block diagram of the proposed method shown in Fig. 3.1. The input signal $x_i(n)$ go

through the compensation model to get $b_i(n)$. Then in each channel, we first decompose the signal by 4:1 and get $b_i^{jp}(n)$, where j represents the j^{th} channel of the 9-channel FB, in this case, $j=1, 2 \dots 9$. p represents the p^{th} sub-channel based on polyphase implementation, here $p=1, 2, 3, 4$. After convolve with the subband filter, the outputs of subbands are combined together by simply sum up. Then the output $v_i^j(n)$ is proposed by the aligner and DRC to obtain $y_i^j(n)$. The final output $d_k(n)$ is the combination of the outputs from nine channels.

3.2.1 9-channel filter bank

The 9-channel crossover filter bank is similar to the *9-channel FB* proposed in Chapter 2, shown in Fig. 3.2. The center frequencies of these filters are: 250 Hz, 500 Hz, 1 kHz, 1.5 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz and 8 kHz. The orders of filters are: 579, 607, 339, 419, 479, 339, 287, 247, 147.

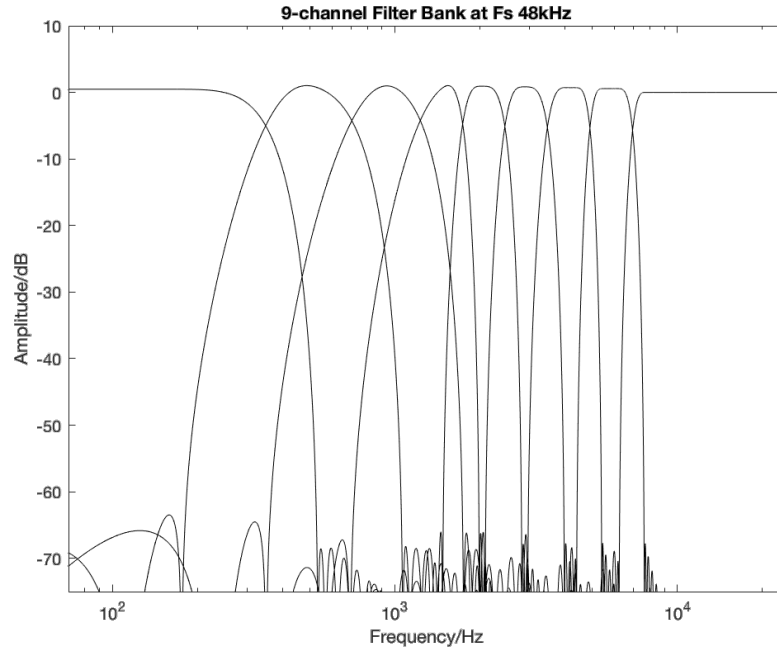


Fig. 3.2. 9-channel filter bank at sampling frequency of 48kHz

3.2.2 Compensation Model

The *9-channel FB* bring distortion because of the ripples of filter in each channel and overlap between filter bands. To measure the distortion, white noise with zero-mean and unit variance is used to go through the filter bank. According to the property of white noise, the difference between the spectrum of the output signal and the spectrum of the input noise could represent the filter bank distortion. Fig. 3.3 shows the difference between the white noise before and after the 9-channel filter bank.

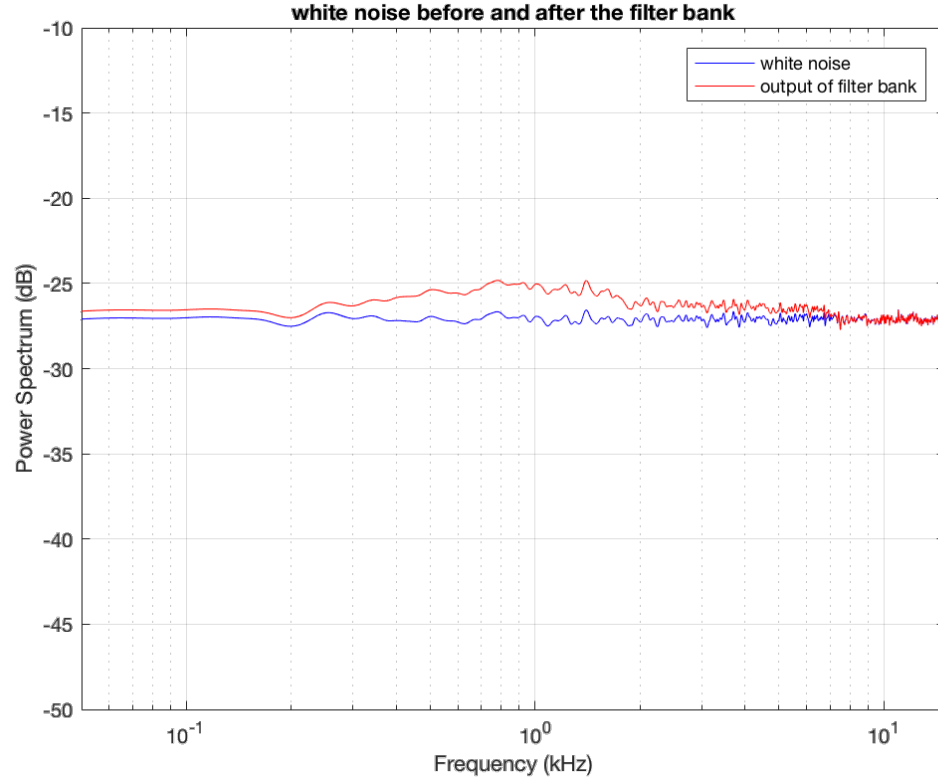


Fig. 3.3. White noise before and after the filter bank

Fig. 3.4 shows the compensation model, which is represented by the ratio of input and output amplitude in frequency domain. The model can be written as

$$R(f) = \frac{|N_{input}(f)|}{|N_{output}(f)|} \quad (3.1)$$

where $|N_{input}(f)|$ is the amplitude of input white noise in frequency domain, and $|N_{output}(f)|$ is the amplitude of the output white noise in frequency domain. For example, at 500 Hz, the amplitude of the input signal is 1 while the amplitude of the output signal is 1.075, so that we need to bring down the amplitude of input signal by ratio 0.93.

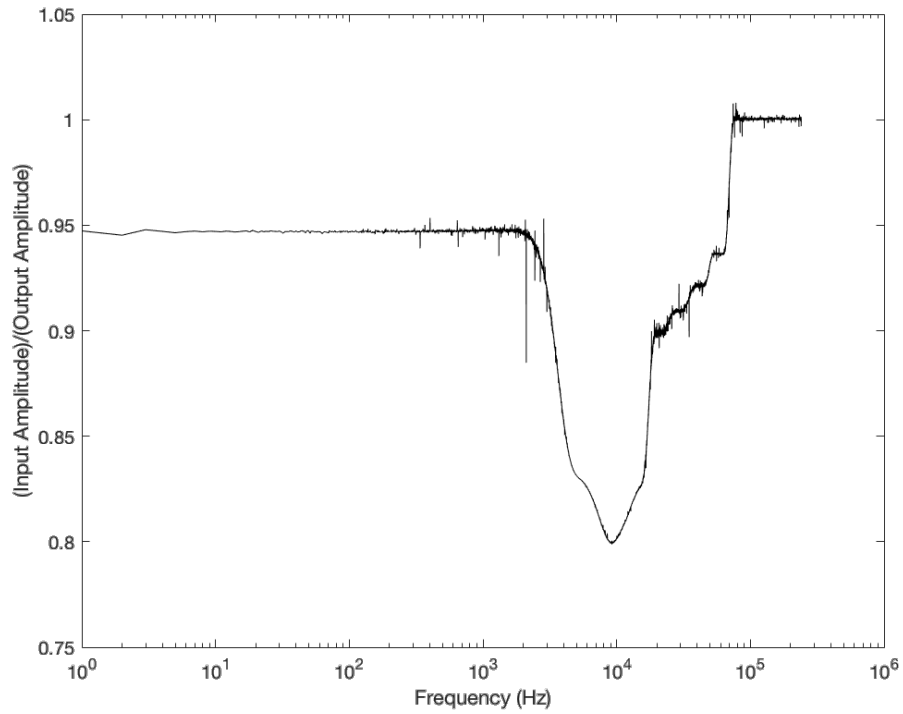


Fig. 3.4. Compensation Model

Because the *9-channel FB* decomposes the input signal spectrum into different frequency bands, it seems not efficient to compensate each band of signal after the filter bank. Thus, the input audio signal is passed through the compensation filter we have determined before it goes to the filter bank as shown in Fig. 3.4. It can be presented as

$$B(f) = R(f) \times X(f) \quad (3.2)$$

where $B(f)$ is the amplitude of the output audio signal in frequency domain, and $X(f)$ is the amplitude of the input audio signal in frequency domain. The phase of the signal is not changed.

3.2.3 Polyphase Implementation

The FIR filter output from the i^{th} -channel filter bank is,

$$v_i(k) = w^T(k)b_i(k), \text{ for } i = 0, 1, \dots, M - 1 \quad (3.3)$$

$b_i(k)$ is the input signal at every time instant k . The filter outputs are computed once for every N input samples. N is the number of sub-channels after polyphaser implementation.

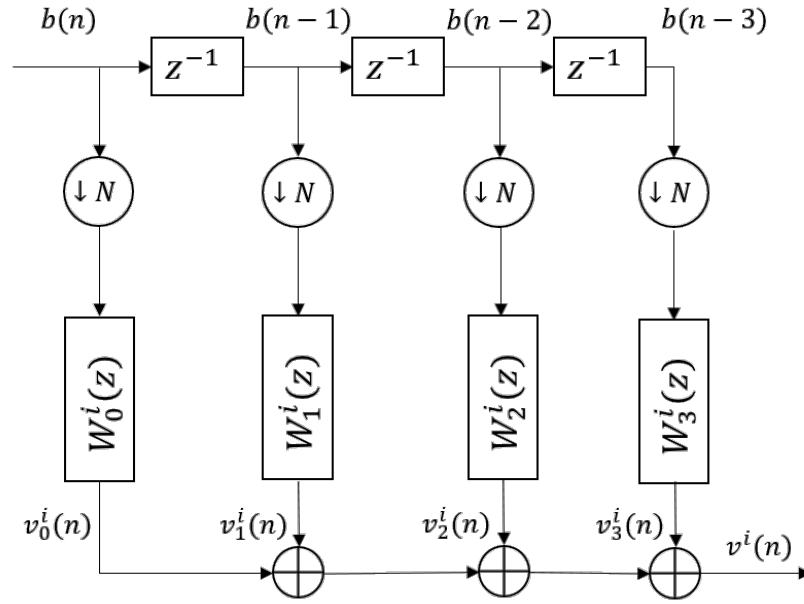


Fig. 3.5. Polyphase realization of FIR filter

Based on polyphase decomposition, an alternative realization of FIR filter $W(k, z)$, was proposed in paper [21]. According to Noble Identity, the polyphase realization is as shown in Fig. 3.5. In this structure, $W(k, z)$, is represented in polyphase form as,

$$W(k, z) = \sum_{p=0}^{N-1} W_p(k, z^N) z^{-p} \quad (3.4)$$

where $W_p(k, z)$ is the p^{th} polyphase filter of $W(k, z)$, given by

$$W_p(k, z) = \sum_{n=0}^{K-1} w_{nN+p}(k) z^{-n}, \text{ for } p = 0, 1, 2, 3 \quad (3.5)$$

where K is the length of the polyphase filters and the FIR filter is assumed to be of length $L = NK$.

3.2.4 Other Parts

The other parts of the DRC system are same or similar to these proposed in Chapter 2.

An aligner is added in each channel to deal with the group delay $\frac{L_m-1}{2}$ produced by the FIR filter of m^{th} channel, where L_m is the length of the m^{th} filter. Before going through the compressor, the $d_i^m(k)$ signal of each channel is shifted to be aligned with each other.

DRC modifies the amplitude of the input signal and reduces the sound pressure level (SPL) range. Level detection and gain stage are two fundamental blocks in the DRC algorithm [23]. Level detector helps in estimating the input signal level and convert it to logarithmic scale. The gain stage compares the estimated signal level with a predefined threshold level known as Compression Threshold (CT) and determines the compression gain according to the static gain curve. The static characteristics of a DRC also includes CR, defined as the ratio of change in input SPL to change in output SPL [24].

The compressor initially estimates the level of the input $d^j(n)$, then the estimation is compared with a fixed threshold key (TK). If it is higher than the TK, a gain with larger CR will be generated, else, if the level is lower than the TK, a gain with smaller CR will be created. Later, the gain will be adjusted based on the attack time and release time. AT and RT decide how fast the compressor should react to the change in input SPL. Finally, a gain will be generated based on a scaling factor

and multiplied with the input, i.e. $d^j(n)$ will be multiplied by the final gain to generate the compressed output signal $y^j(n)$, of the j^{th} frequency band.

3.3 Test Results

The proposed method is an improved extension of the Crossover FB DRC [25]. Their performances are evaluated and compared each other. Objective and subjective tests are conducted to measure the quality and intelligibility of output audio signal. The evaluations are carried out for 10 sentences of 3 seconds long from HINT [12] database. The speech files are combined with 3 different noise types: machinery, babble and traffic noise for 3 different types of SNRs; -5, 0, and 5 dB, and are sampled at 48 kHz.

3.3.1 Objective Evaluation

We choose the perceptual evaluation of speech quality (PESQ) for speech quality measurement. Coherence Speech Intelligibility Index (CSII) [26, 27] is used to measure the intelligibility of speech. PESQ ranges between 0.5 and 4.5, with 4.5 being highest perceptual quality. CSII ranges between 0 and 1, with 1 being high intelligibility.

Fig. 3.6 shows the values of PESQ and CSII versus SNR for the 3 background noise types. The objective measures show that the proposed method archives considerable better PESQ scores and similar CSII scores when compared to Crossover FB DRC. The PESQ scores show some variations for machinery noise at very low SNR.

3.3.2 Subjective test results

Along with objective measures, we perform mean opinion score (MOS) tests [17] on 16 normal hearing subjects, including 8 males and 8 females. The audio files were played through the headphones from the computer. The audio files were generated using *Crossover FB DRC* method and proposed method which is presented as *Polyphased DRC*.

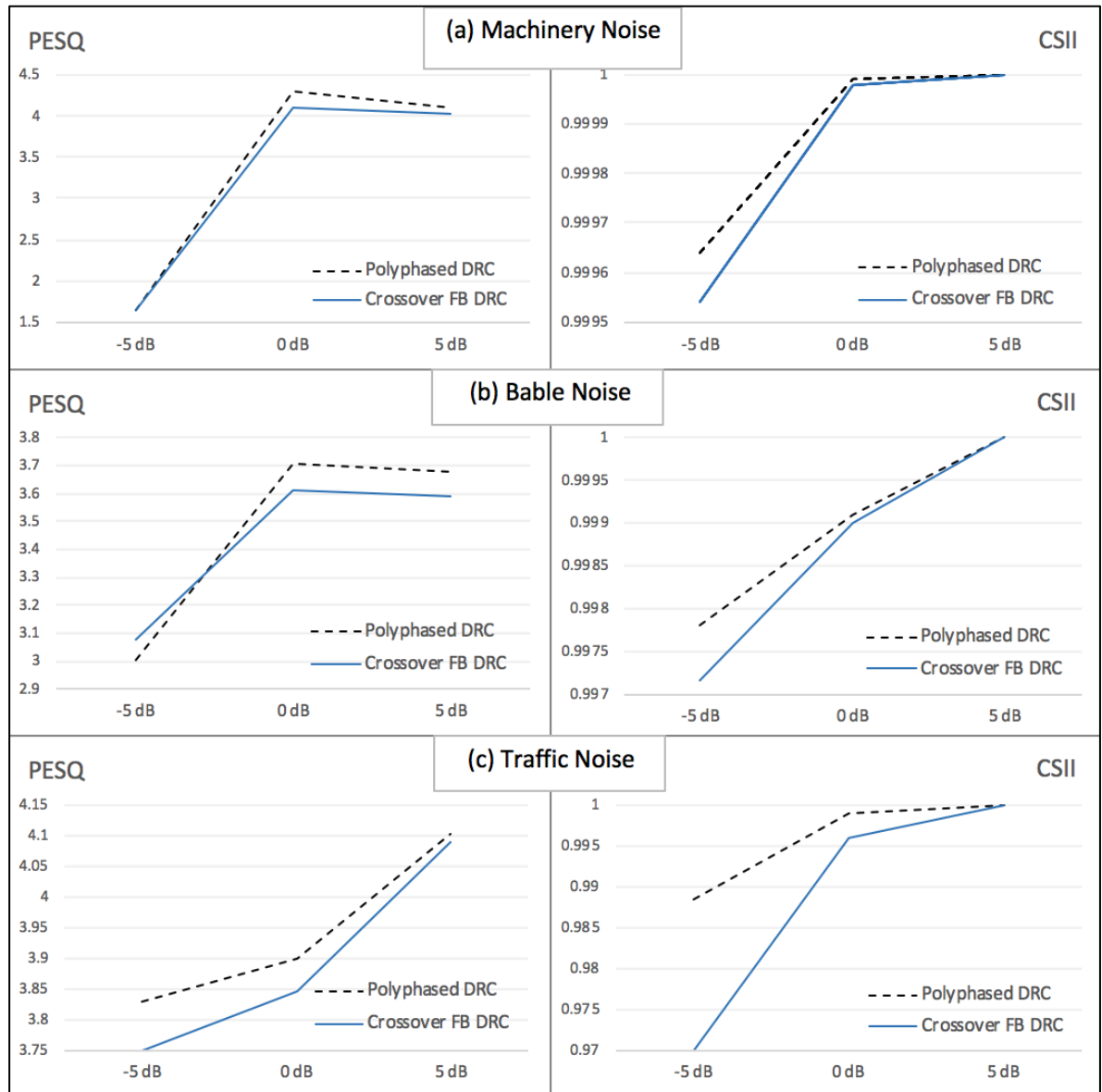


Fig. 3.6. Objective evaluation of speech quality and intelligibility

Each subject was instructed to score in the range 1 to 5 for the audio files heard, based on the following criteria: 5 being excellent speech quality and imperceptible level of distortion. 4 for good speech quality with a perceptible level of distortion. 3 for fair speech quality with a mediocre level of distortion. 2 for poor speech quality with a lot of uneven distortions. 1 having the least quality of speech and intolerable distortion level. The results of subjective test are shown in Fig. 3.7, which illustrates the effectiveness of the proposed method, maintaining the speech quality and intelligibility.

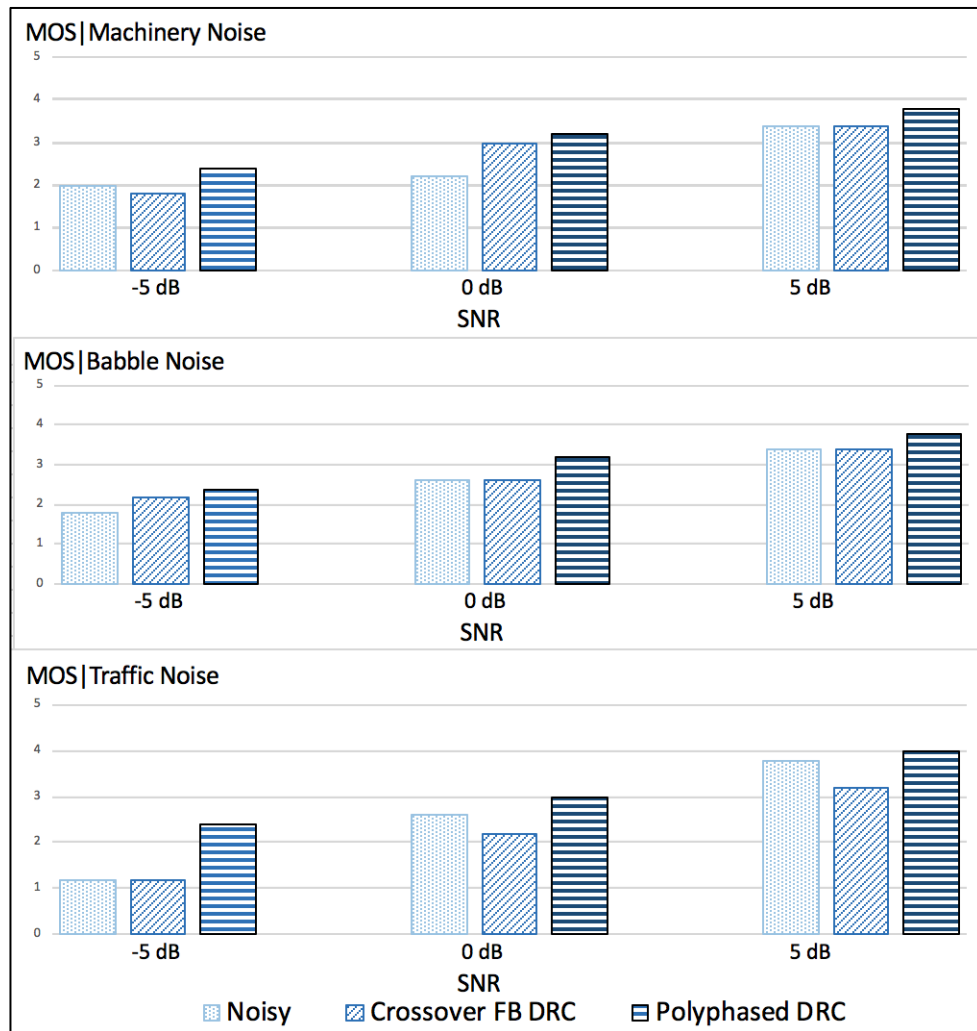


Fig. 3.7. Subjective test results

3.3.3 Computational Time measurement

To compare the computational complexity, the processing time of the proposed method and *Crossover FB DRC* are measured using Matlab R2017b on MacOS with processor 3.1 GHz Intel Core i7. The computational time of 10 sentences were measured individually and the average value and deviation was taken.

For the *Crossover FB DRC*, after compensation, the input audio signals took an average of 48.8 ms to go through the filter bank, aligner and DRC processing. And the deviation was about 1.29×10^{-6} . For the proposed method, we measured the computational time (τ^{jp}) between the input ($b^{jp}(n)$) and output ($v^{jp}(n)$) of each sub-filter as shown in Fig. 3.1. j represents the j^{th} channel of the 9-channel FB, in this case, $j=1, 2 \dots 9$. p represents the p^{th} sub-channel based on polyphase implementation, here $p=0, 1, 2, 3$. τ represents the average computational time of all filters. The total computational time is measured as summation of τ and the computational time of aligner and DRC processing. After this, the average value of computational time for 10 sentences was 39.7 ms, the deviation was 1.04×10^{-6} . The proposed method requires 9.1 ms less time than the *Crossover FB DRC* for processing. This result proves that applying polyphase implementation diminishes the complexity and reduces the computational time, benefiting real time operation.

3.4 Summary

In this work, a compensation filter is developed in order to attenuate the distortion introduced by the filter-bank. The proposed polyphaser implementation reduces the computational complexity of the algorithm. Objective measurements PESQ and CSII were conducted to evaluate the quality

and intelligibility of speech respectively. The results show that in most cases, the performance of the proposed method is better. And the results of MOS tests also show the improvement.

CHAPTER 4

NON-UNIFORM COSINE MODULATED FILTER BANK USING FREQUENCY RESPONSE MASKING APPROACH

4.1 Overview

Multi-channel dynamic-range compression is used by most hearing aid devices (HADs) to match the frequency resolution of the compression system to human audiogram. A compression system involves trade-offs between frequency resolution and computational complexity. A filter bank could be used to build a multi-channel system which separates the input signal into different frequency bands. Polyphase DFT filter banks are efficiently used to extract signal with equal bandwidth [28]. However, when we need to fit the human diagram, the subbands are supposed to be of different bandwidths, so that uniform filter banks are not suitable. Cosine modulated filter bank (CMFB) is one kind of M-Channel maximally decimated filter bank [29]. Its subband filters are derived from a prototype filter by cosine modulation. A prototype filter with appropriate characteristics is initially designed and then all the analysis and synthesis filters are efficiently generated from the cosine modulation of this prototype filter. In a uniform CMFB, all the analysis and synthesis filters are of equal bandwidths. A non-uniform CMFB can be derived from a corresponding uniform CMFB by simply combining a number of subbands together [30, 31]. To improve the frequency resolution of the compression system, the bandwidth and transition width of the prototype filter are supposed to be as narrow as possible. Frequency response masking (FRM) filter can be used to design filters with very narrow transition width [32, 33]. In this paper, one nine-channel dynamic-range compression system with non-uniform CMFB using FRM filter

as prototype filter is proposed and implemented on smartphone (iphone). The objective tests are conducted to measure the quality and intelligibility of the output audio.

4.2 Proposed Method

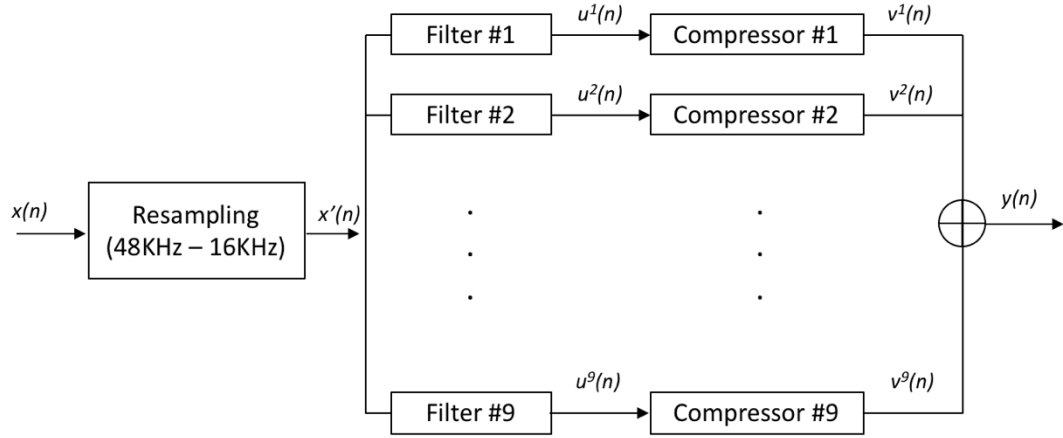


Fig. 4.1. Block diagram of the proposed method

As shown in Fig. 4.1, in this work, first we decimate the input signal $x(n)$. After go through the filter bank, the output of each channel directly goes through the compressor, which is specific designed for each frequency bands. One difference of this method from previous work is that aligner is not required to align up the output of each channel since each filter of the CMFB is of same order.

4.2.1 Prototype filter

The goal of the prototype filter design is narrow bandwidth and transition width, and the number of filter coefficients is supposed to be as small as possible. To achieve this goal, we use frequency response masking (FRM) approach to design the prototype filter.

The principle of FRM approach is as follow [34, 35]:

The model filter is a low-pass filter with z-transform function $H_a(z)$. The bandwidth of $H_a(z)$ is Δb and the transition width is Δa . The frequency response $H_a(e^{j\omega})$ is shown in Fig. 4.2. Replacing each delay by M delays so that we can get filter with z-transform function $H_b(z) = H_a(z^M)$. $H_c(z)$ is the z-transform function of the masking filter whose frequency response $H_c(e^{j\omega})$ is shown in Fig. 4.2. Using the masking filter to mask $H_b(z)$, then we can get resulting filter $H_d(z)$ with frequency response $H_d(e^{j\omega})$ showing in the figure. The bandwidth of filter $H_d(z)$ is $\Delta b/M$ and the transition is $\Delta a/M$. If $H_b(z)$ is masked by a band-pass masking filter $H_e(z)$, the resulting frequency response is $H_f(e^{j\omega})$.

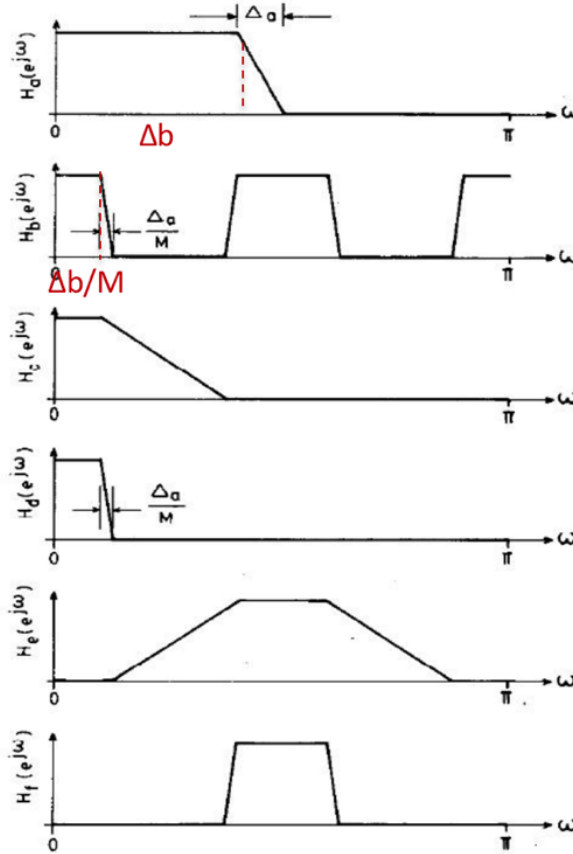


Fig. 4.2. The principle of frequency response masking approach

In our case, we choose a half-band filter as our model filter as shown in Fig. 4.3 (a). The number of coefficients of the model filter is 23 and the number of non-zero coefficients is only 12. As shown in Fig. 4.3 (b), then we replay each delay by $M=16$ delays to get the periodic model filter. Next in Fig. 4.3 (c), as low pass masking filter is generated and the length of the filter is 90. Finally, we get the prototype filter showing in Fig. 4.3 (d).

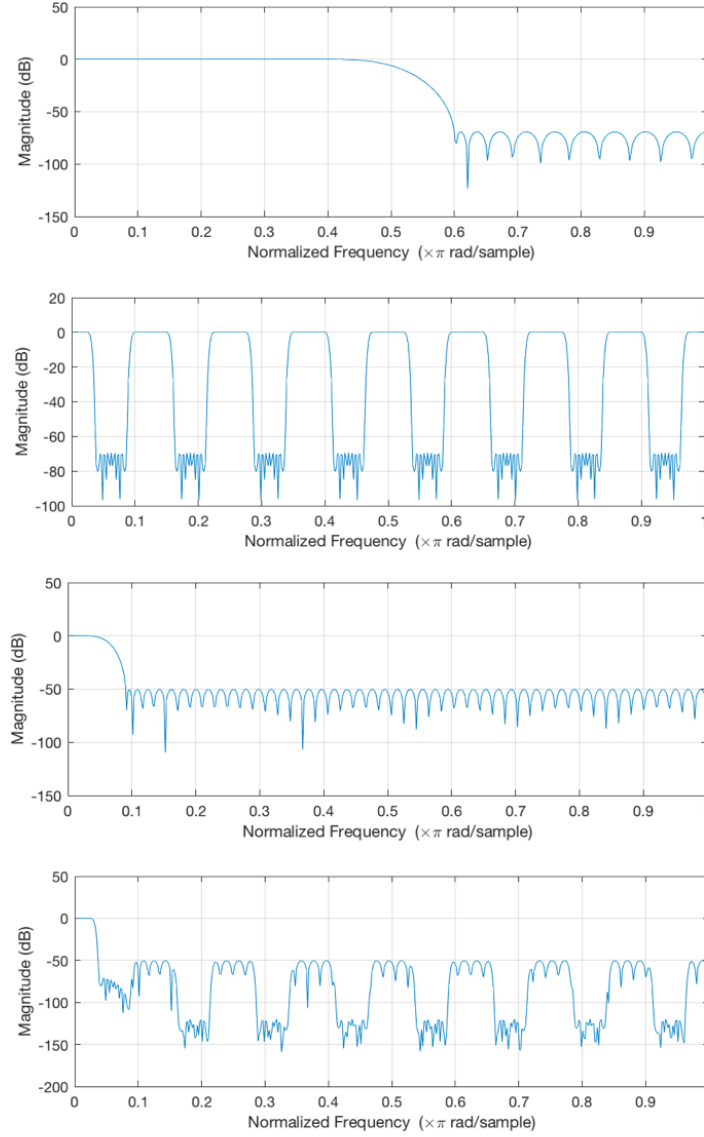


Fig. 4.3. Proposed prototype filter bank

The final length of the prototype filter is 458 and its bandwidth is around 250 Hz. And the transition band is very narrow. With the FRM technique, the resulting filter has very sparse coefficients, hence reduce the implementation complexity [36]. Furthermore, in conventional FIR filters, to reduce the transition width, the price needs to pay is to increase the filter order, which results in increasing implementation complexity. The length of a linear-phase FIR filter can be calculated by Bellanger's equation [36]:

$$N = \frac{-2\log(10\delta_1\delta_2)}{3\Delta f} - 1 \quad (4.1)$$

Where, δ_1 and δ_2 are the peak magnitudes of pass-band and stop-band ripples and Δf is the normalized transition width. In order to achieve the similar features such as pass-band ripples and stop-band ripples, the coefficient number of direct equiripple FIR filter needs to be at least 1126.

4.2.2 Uniform and non-uniform CMFB

Fig. 4.4 shows the structure of an M -channel decimated uniform CMFB. The input signal is decomposed into M subbands signal with equal bandwidth by a set of analysis filters $H_k(Z)$, where $0 \leq k \leq M - 1$. Then a set of synthesis filter $F_k(Z)$, $0 \leq k \leq M - 1$, reconstructs the signals from M subbands after interpolation of a factor of M in each channel.

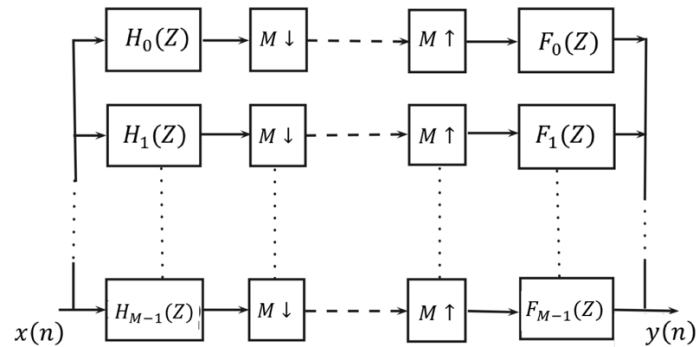


Fig. 4.4. M-channel decimated uniform CMFB

The analysis filters and synthesis filters are generated from a prototype filter cosine modulation.

The coefficients of the analysis filters and synthesis filters can be given by:

$$h_k(n) = 2p_0(n) \cos \left[\frac{\pi}{M} \left((k + 0.5) \left(n - \frac{N}{2} \right) + (-1)^k \frac{\pi}{4} \right) \right] \quad (4.2)$$

$$f_k(n) = 2p_0(n) \cos \left[\frac{\pi}{M} \left((k + 0.5) \left(n - \frac{N}{2} \right) - (-1)^k \frac{\pi}{4} \right) \right] \quad (4.3)$$

Where, $k = 0, 1, 2, \dots, M - 1$ representing k^{th} channel, $n = 0, 1, 2, \dots, N - 1$ with N representing the length of the filter, and $p_0(n)$ is the coefficients of the prototype filter with order N .

Fig. 4.5 shows the structure of a cosine modulated non-uniform filter bank of M' channels. The non-uniform filter bank separates the input signal into subbands of different bandwidths. The input signal is decomposed into M' subbands signal with unequal bandwidth by a set of analysis filters $H'_k(Z)$, where $0 \leq k \leq M' - 1$. Then a set of synthesis filter $F'_k(Z)$, $0 \leq k \leq M' - 1$, reconstructs the signals from M' subbands. The M' -channel non-uniform filter bank can be generated from the M -channel uniform filter bank by simply combining appropriate channels [38].

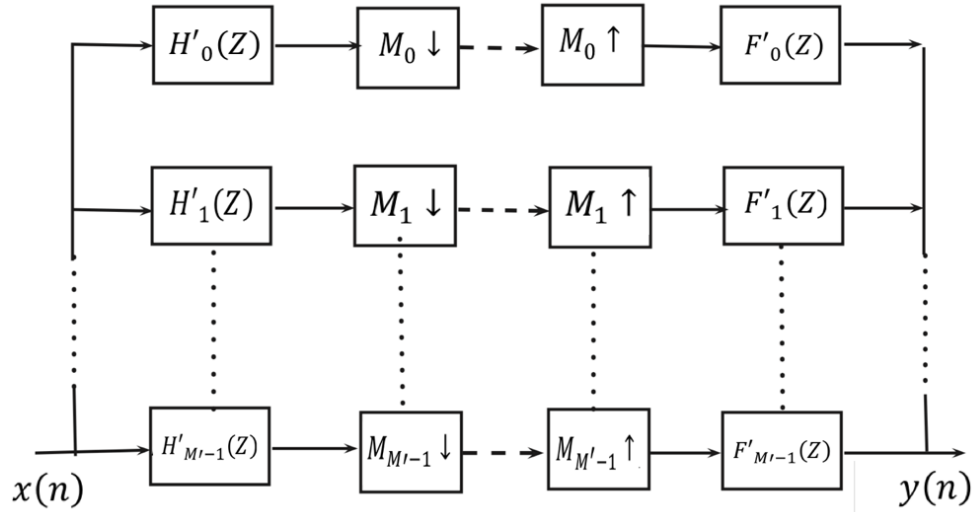


Fig. 4.5. M' -channel non-uniform CMFB

The analysis filters $H'_j(z)$ can be obtained by merging l_j adjacent analysis filters as:

$$H'_j(z) = \sum_{k=n_j}^{n_j+l_j-1} H_k(z) \quad (4.4)$$

Where, $j = 0, 1, 2 \dots, M' - 1$, n_j is the lower band edge and l_j is the number of adjacent channels to be combined.

The synthesis filters can be obtained similarly:

$$F'_j(z) = \frac{1}{l_j} \sum_{k=n_j}^{n_j+l_j-1} F_k(z) \quad (4.5)$$

Where the corresponding decimation factor is $M_i = \frac{M}{l_j}$.

In our case, we first realize a 16-channel uniform CMFB with the prototype filter proposed in Section 4.2.1. The 16-channel uniform CMFB is shown in Fig. 4.6 (a), and the 9-channel non-uniform CMFB is shown in Fig. 4.6 (b). The center frequency of the 9 non-uniform filters are: 250, 750, 1250, 1750, 2250, 3000, 4000, 5500, 7000 Hz.

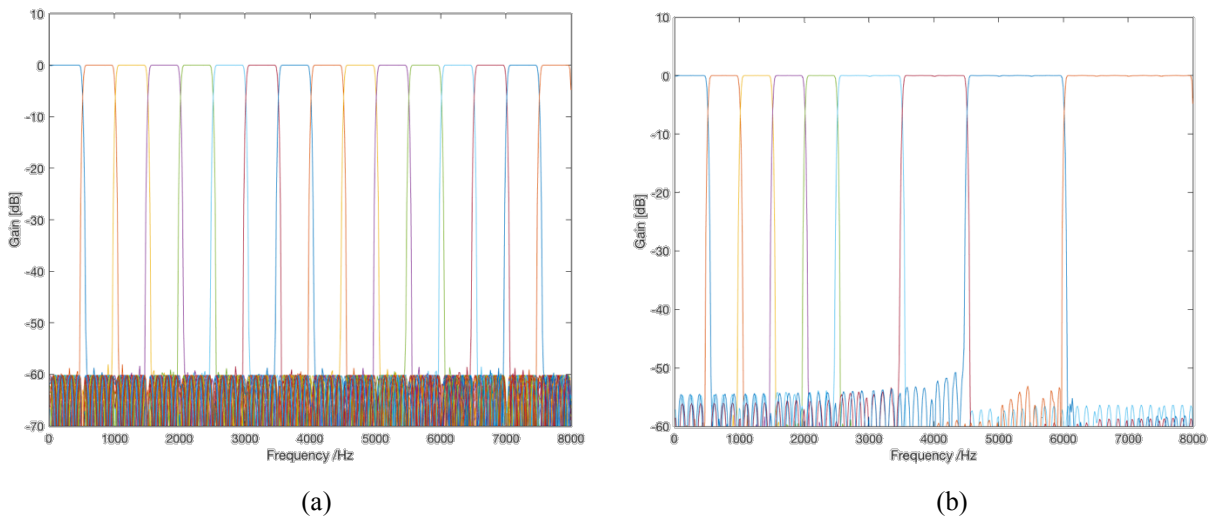


Fig. 4.6. 16-channel uniform CMFB and 9-channel non-uniform CMFB

To measure the distortion that caused by the filter bank, white noise with zero-mean and unit variance is used to go through the filter bank. According to the property of white noise, the difference between the spectrum of the output signal and the spectrum of the input noise could represent the filter bank distortion. In Fig. 4.7, we compare the distortion model of the proposed non-uniform CMFB (b) and the *9-channel FB* (a) proposed in Chapter 2. We can see that the compared the *9-channel FB*, the distortion caused by the non-uniform CMFB focuses on higher frequency areas above 2kHz. Moreover, unlike the distorted broad frequency area in *9-channel FB*, the frequency bin with big noticeable distortion is narrow enough to be ignored.

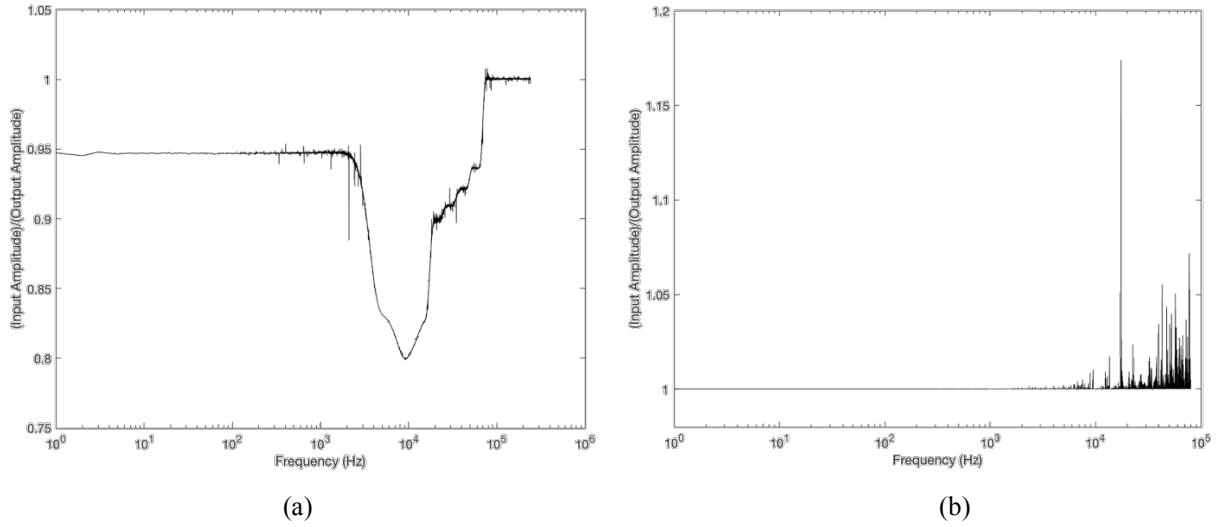


Fig. 4.7. Distortion model of the *9-channel FB* and non-uniform CMFB

4.3 Real-time Implementation

Since the length of the proposed prototype filter is 458, which is still too large to realize the real-time implementation. To further reduce the computational complex, we take advantage of the sparse coefficients of filters. In Section 4.2.1, the model filter is a half-band filter with 23

coefficients, 11 of which are zeros. Then the periodic model filter is generated from replacing the delay of model filter by $M=16$ delays. So, in the coefficients of the periodic filter, there are 363 zeros. We can use simple multiplication and summation to replace convolution, which saves a large amount of computation. To use these zeros, we shift the masking filter instead of the proposed prototype filter in Section 4.2.1, which means that to build a CMFB with masking filter as prototype filter. The input signal could first go through the periodic model filter using simple multiplication and summation, and then convolve with the masking filter bank.

The input/output latency is the main concern for DRC implementation in smartphone. In HADs, developer is able to access to lower layer hardware such as DSP core and audio Codec. However, developer could only program in higher layers in smartphone, which results in additional latency. Android media library [39] provides audio application programming interface (API) for android audio applications. With audio application programming interface (API) provided by Android media library, developers are able to access to audio data in application layer. However, the latency of such API is too high for multi-channel DRC.

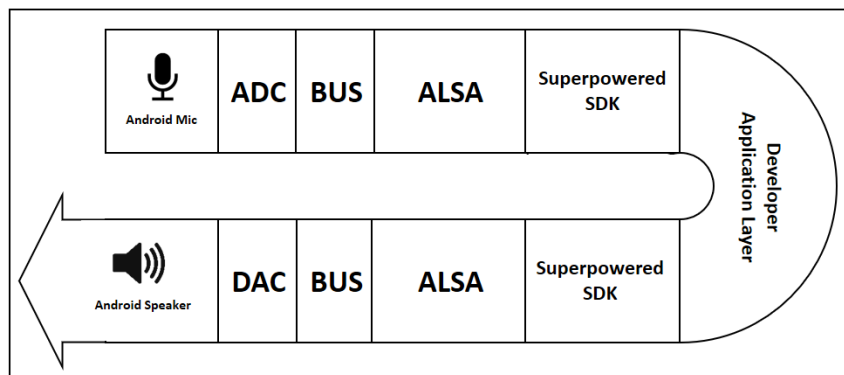


Fig. 4.8. Audio processing of android operating system using Superpowered SDK

Superpowered SDK is a good solution to deal with the audio latency challenge [40]. As shown in Fig. 4.8, Android Superpowered SDK builds a rapid bridge between advanced Linux sound

architecture (ALSA) and application layer, so that signal data could transfer quickly crossing layers. Furthermore, in application layer programming, C/C++ language processes faster than Java. Superpowered SDK allows developers to program in C/C++. Compared the other Android audio libraries which provide a solid foundation, Superpowered SDK could further reduce the audio latency for multi-channel DRC processing in Android smartphone.

The propose 9-channel non-uniform CMFB based DRC has been implemented on an android smartphone, Google Pixel 2, with Android operating system version 6.0.1. The input signal data is coming continually via the microphone of smartphone, then converting to digital data and stored in buffer. The microphone of smartphone captures audio input frame by frame continuously, and then the signal is converted to digital data and stored in buffers. After multi-channel DRC processing, the output frame goes through D/A convert and then plays out.

4.4 Test Results

We evaluate the performance of the propose non-uniform CMFB based DRC by comparing with a laptop-based commercial audio compression & limiter provided by Hotto Engineering [7] and the Crossover FB DRC proposed in Chapter 2.

4.4.1 Objective Test

The evaluations are carried out for 10 sentences of 3 seconds long from HINT [11] database. The objective evaluations are performed for 3 different noise types: multi-talker babble, machinery and traffic noises. These noises were recorded in real-world conditions in several different scenarios. The presented results are average of them all. The microphone signal is generated by adding the

speech and noise together at SNR of -10, -5, 0, 5, and 10dB. PESQ is used for speech quality measurement and STOI is used for intelligibility evaluation.

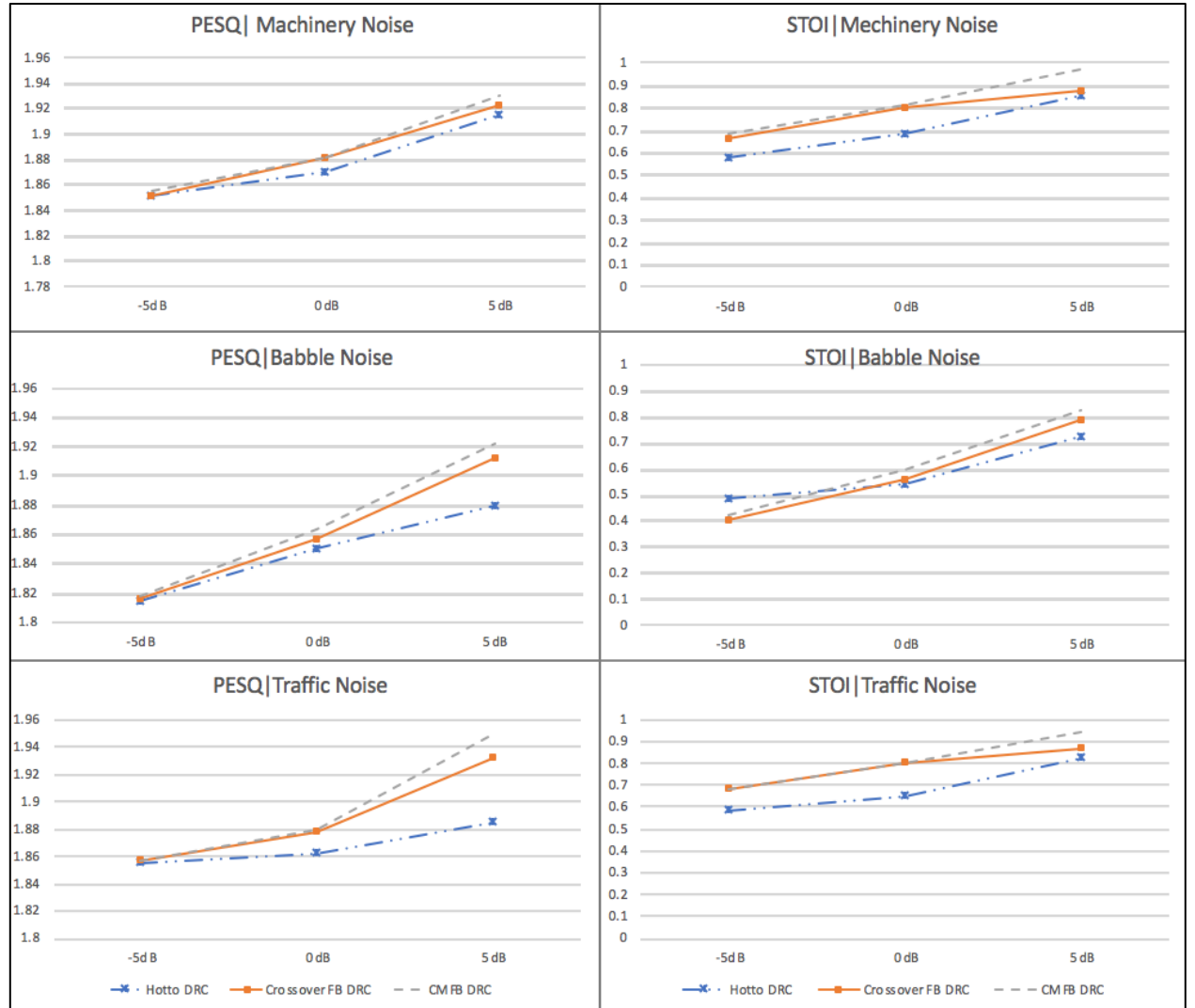


Fig. 4.9. Objective results of PESQ and STOI

Fig. 4.9 shows the results of PESQ and STOI comparison between the proposed method and the comparisons. According to the experimental scores above, we can tell that proposed non-uniform CMFB based DRC has the best results both in PESQ and STOI, which means that the quality and

the intelligibility of the signal processed by the proposed method are comparably better than the other two DRCs.

4.4.2 Subjective Test

We performed subjective tests on 10 normal hearing with 5 females and 5 males. The test setup is same as Section 2.4.3. MOS conducts by presenting the audio signal processed by Hotto DRC, *Crossover FB DRC* and proposed non-uniform CMFB based DRC. The dataset is clean speeches from 3 lists of 10 sentences each from TIMIT database in the presence of restaurant noise at SNR levels of -5 and 0 dB. The subjects are supposed to give scores range from 1 to 5, where 5 stands for excellent and 1 stands for poor. Fig. 4.10 reflects the average scores of 10 normal hearing subjects. The results show that at SNR of both -5 and 0 dB, the proposed non-uniform CMFB based DRC provides the best result according to the subjects.

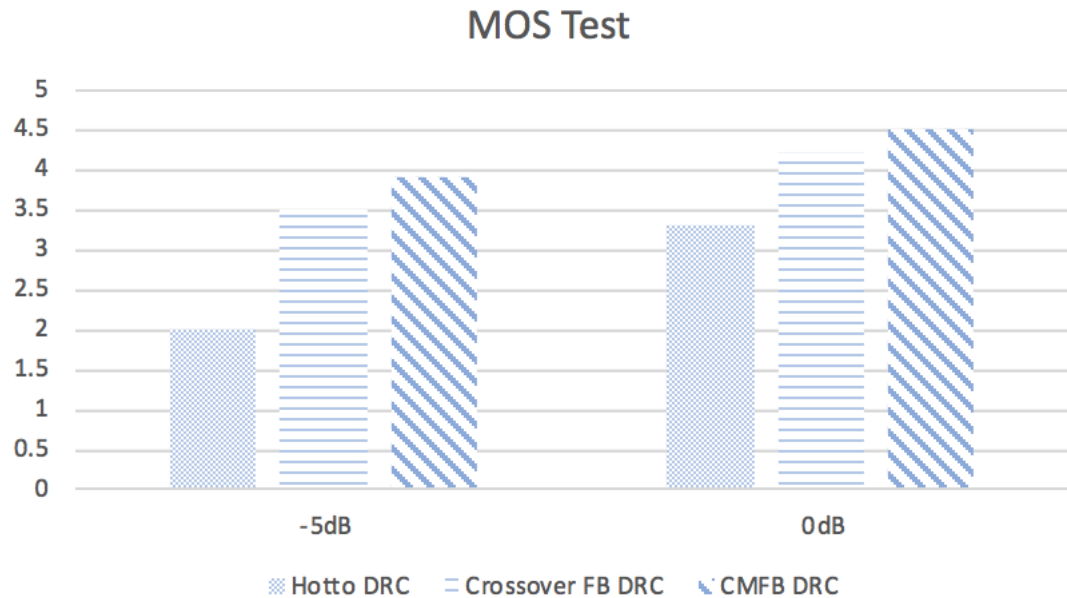


Fig. 4.10. Results of MOS subjective test

WR test is conducted under the same condition as MOS tests. A GUI based interface is used for the WR, more details of this tool and WR test procedures are presented in [18]. The experimental result based on the average of 10 subjects presents in Fig. 4.10. The proposed non-uniform CMFB based DRC shows best test in both WR rate and MOS.

4.5 Summary

In this work, a non-uniform CMFB based DRC is proposed and implemented in Smartphone running in real time. Comparing to a single-channel Hotto DRC, which is running in laptop and the *Crossover FB DRC* proposed in Chapter 2, the proposed non-uniform CMFB based DRC provides better performance in both quality and intelligibility according to objective and subjective tests.

REFERENCES

- [1] World Health Organization, “Deafness and Hearing Loss,” 2018; <http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
- [2] National Institute on Deafness and Other Communication Disorders, National Institute of Health, “Hearing, Ear Infections, and Deafness”, 2018; <https://www.nidcd.nih.gov/health/hearing-ear-infections-deafness>.
- [3] R.E. Sandlin, *Text Book of Hearing Aid Amplification*. Singular Publishing Group, 2000, pp.37-55.
- [4] J.M. Kates, “Principles of Digital Dynamic-Range Compression.” *Trends in Amplification* 9.2 (2005): 45–76. PMC. Web. 8 July 2018.
- [5] T. Schneider and R. Brennan, "A multichannel compression strategy for a digital hearing aid," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, 1997, pp. 411-414 vol.1.
- [6] B.C.J. Moore, B.R. Glasberg, and M.A. Stone, “Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CANEQ2-HF ,” in *International Journal of Audiology*, London, UK: Informa, 2010, pp.216-227.
- [7] K. Ngo et al., A combined multi-channel Wiener filter-based noise reduction and dynamic range compression in hearing aids, *Signal Processing*, v.92 n.2, p.417-426, February, 2012
- [8] B.C.J. Moore, “Hearing Aid”, *Cochlear Hearing Loss*, 2nd ed., Hoboken: Wiley-Interscience, 2007, ch. nine, sec. III, pp. 233-268.
- [9] Hotto Engineer, 2018; <http://www.hotto.de/>
- [10] A.V. Oppenheim, R.W. Schaffer, J.R. Buck, “Discrete-time signal processing” Upper Saddle River, N.J: Prentice Hall. 1999.
- [11] Android Source, “Audio Latency Measurements”, 2018, https://source.android.com/devices/audio/latency_measurements
- [12] Statistic Signal Processing Research Laboratory, The University of Texas at Dallas, “Hearing Aid Project, Database,” 2017, <http://www.utdallas.edu/ssprl/hearing-aid-project/database/>.
- [13] A.W. Rix et al., “Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs,” *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2, pp. 749-752., May 2001.
- [14] N Shiran, I.D. Shallom, “Enhanced PESQ algorithm for objective assessment of speech quality at a continuous varying delay,” *IEEE Int. Workshop, Quality of Multimedia Experience*, pp. 157-162, Jul 2009.

- [15] C.H. Tall et al., "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE trans. Audio, Speech, Lang. Process.* 19(7), pp. 2125-2136., Feb 2011.
- [16] J.M. Kates, H.A. Kathryn, "A model of speech intelligibility and quality in hearing aids," IEEE Workshop, Appl. of Signal Processing to Audio and Acoustics, pp. 53-56, Oct. 2005.
- [17] R.C. Streijl, S. Winkler, D.S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," in *Multimedia Systems* vol.22.2, pp. 213-227, 2016.
- [18] ITU-T Rec. P.830, "Subjective performance assessment of telephone- band and wideband digital codecs," 1996.
- [19] S. Tokgoz, Y. Hao, I.M.S. Panahi, "A Hearing Test Simulator GUI for Clinical Testing," *Acoustical Society of America (ASA)*, May 2018.
- [20] M. Bellanger, G. Bonnerot, and M. Coudreuse, "Digital filtering by polyphase network: Application to Sample Rate Alteration Filter Banks," *IEEE Transl. on Acoust. Speech and Signal Proc.*, vol.24, 1976, pp.109-114.
- [21] R.W Bäuml, W. Sörgel, "Uniform polyphase filter banks for use in hearing aids: Design and constraints," in *proc. European Signal Processing conf.*, 2008.
- [22] T. Schneider and R. Brennan, "A multichannel compression strategy for a digital hearing aid," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, 1997, pp. 411-414 vol.1.
- [23] S.P. Deepu, D.S. Sumam and K.M. Ramesh, "Estimation of attack time constant for dynamic range compressors in hearing aids," 2016 IEEE International Conference on Digital Signal Processing (DSP), Beijing, 2016, pp. 20-24.
- [24] H. Dillon(1999). NAL-NL1: A new prescriptive procedure for fitting non-linear hearing aids. *Hearing Journal*, 52(4), 10, 12, 14, 16
- [25] Y. Hao, Z. Zou, and I.M.S. Panahi, "A robust smartphone based multi-channel dynamic-range audio compressor for hearing aids," *175th Acoustic Society of America*, 2018.
- [26] P. Loizou, "Speech Enhancement: Theory and Practice", Boca Raton, FL: CRC Press, 2007.
- [27] M.C. Killion, S. Fikret-Pasa (1993). The 3 types of sensorineural hearing loss: loudness and intelligibility considerations. *Hearing Journal*, 46(11), 31-36
- [28] K.C. Zangi, R.D. Koilpillai, Software radio issues in cellular base stations, *IEEE Trans. Selected areas Commn.* 17 (4) (1999) 561–573.
- [29] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [30] J.-J. Lee, B.G. Lee, A design of nonuniform cosine modulated filter banks, *IEEE Trans. Circ. Syst. II: Anal. Dig. Signal Process.* 42 (11) (1995) 732–737.

- [31] K. Shaeen, E. Elias, Non-uniform cosine modulated filter banks using metaheuristic algorithms in CSD space, *J. Adv. Res.* (2014) doi:10.1016/j.jare.2014.06.008.
- [32] M.B. Furtado Jr., P.S. Diniz, S.L. Netto, Optimized prototype filter based on the FRM approach for cosine-modulated filter banks, *Circ. Syst. Sign. Process.* 22 (2) (2003) 193–210.
- [33] L. Rosenbaum, P. Lowenborg, M. Johansson, An approach for synthesis of modulated M-Channel FIR filter banks utilizing the frequency-response masking technique, *EURASIP J. Adv. Signal Process.* 2007 (2007) 068285.
- [34] K. Shaeen, E. Elias, Design of multiplier-less cosine modulated filter banks with sharp transition using evolutionary algorithms, *Int. J. Comput. Appl.* 68 (25) (2013) 1–9.
- [35] W. Zhang et al., "Unified FRM-based complex modulated filter bank structure with low complexity", *Electronics Letters*, vol. 54, no. 1, pp. 18-20, 2018.
- [36] A.T. Fuller, B. Nowrouzian, F. Ashrafzadeh, Optimization of FIR digital filters over the canonical signed-digit coefficient space using genetic algorithms, in: *Circuits and Systems, 1998. Proceedings. 1998 Midwest Symposium on*, IEEE, pp. 456–459, 1998.
- [37] J. M. Bellanger, On computational complexity in digital filters, in: *Proc. of the European conference on circuit theory and design*, pp. 58–63, 1981.
- [38] Z. Zijiang, Y. Yun, A simple design method for nonuniform cosine modulated filter banks, in: *Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, 2007 International Symposium on*, IEEE, pp. 1052–1055, 2007
- [39] Android Developers, “Android. Media”, 2018, <https://developer.android.com/reference/android/media/package-summary.html>
- [40] Superpowered, 2018, <http://superpowered.com/>.

BIOGRAPHICAL SKETCH

Ziyan Zou was born in Anhui, China in 1994. She completed her Bachelor of Engineering in Vehicle Engineering at Tongji University in July 2016. She then entered the Master's program in Electrical Engineering at UT Dallas in August 2016. She started working in the Statistical Signal Processing Research Laboratory (SSPRL) from June 2017. Her research includes, dynamic-range compression and beamforming. She has been working as a Student Worker in the Department of Electrical Engineering at UT Dallas from July 2017. She is working as an Acoustic System Engineer Intern in the company Cirrus Logic, where she is working on noise perception study for active noise cancellation evaluation. She has a very good understanding of Digital and Statistical Signal Processing concepts.

CURRICULUM VITAE

Ziyan Zou

CONTACT INFORMATION:

Department of Electrical and Computer Engineering

The University of Texas at Dallas

800 W. Campbell Rd.

Richardson, TX 75080-3021, U.S.A

EDUCATION:

MSEE, The University of Texas at Dallas, USA (CGPA 3.84 /4.0) Aug 2018

BE, Tongji University, Shanghai, China, P.R. (CGPA: 87.6/100) July 2016

PUBLICATIONS:

- 1) Y. Hao, Z. Zou, I. M.S. Panahi, "Smartphone-based Multi-Channel Dynamic-range Compression with Equalization for Hearing Study", (submitted) IEEE Signal Processing Letters
- 2) Z. Zou, Y. Hao, I. M.S. Panahi, "Design of Compensated Multi-Channel Dynamic-Range Compressor for Hearing Aid Devices using Polyphase Implementation," 40th International Engineering in Medicine and Biology Conference , 2018.
- 3) Y. Hao, Z. Zou, I. M.S. Panahi, "A Robust Smartphone Based Multi-Channel Dynamic-Range Audio Compressor for Hearing Aids," 175th Acoustic Society of America Conference , 2018.