# 國 立 交 通 大 學

# 電信工程學系

# 碩 士 論 文

## 利用多喇叭實現強健串音消除系統

## Robust Crosstalk Cancellation for 3D Sound Using Multiple Loudspeakers


研究生：蘇大中

指導教授：謝世福 博士


中 華 民 國 九十五 年 九 月

# 利用多喇叭實現強健串音消除系統

# Robust Crosstalk Cancellation for 3D Sound Using Multiple Loudspeakers


研究生 ：蘇大中　　Student：D. J. Su

指導教授：謝世福　　Advisor：S. F. Hsieh

國 立 交 通 大 學
電信工程學系碩士班


碩士論文

A Thesis
Submitted to Department of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Electrical Engineering
September, 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年九月

# 利用多喇叭實現強健串音消除系統

學生：蘇大中　　　　　　指導教授：謝世福

國立交通大學電信工程學系碩士班

## 摘要

傳統串音消除技術是針對一對喇叭來實現。而串音消除器主要的問題來自於容易受到擾動的影響。由文獻可得知，當喇叭數增加則串音消除器消除串音的能力亦隨之提升。同時，多喇叭架構對於擾動而造成雙耳音訊失真的效應較一對喇叭要來得輕微。在本篇論文中，我們將呈現各種不同形態的串音消除器在三個喇叭架構下，並且能夠達到節省成本及計算複雜度的目標。我們也會針對兩個喇叭和三個喇叭架構的串音消除器作擾動分析，以及透過模擬數據來作為分析的依據。

# Robust Crosstalk Cancellation for 3D Sound Using Multiple Loudspeakers

**Student : D. J. Su**　　　**Advisor : S. F. Hsieh**

**Department of Communication Engineering**

**National Chiao Tung University**

## Abstract

Traditional crosstalk cancellation is for two loudspeakers. The most critical problem bothering us is the effect of perturbation. In several literatures, we know that multiple loudspeakers arrangement results in good performance of crosstalk cancellation. In addition, it is more robust for perturbation than two loudspeakers setup. In this thesis, we show different types of crosstalk canceller of multiple loudspeakers arrangement and economical realizations with less computational complexity is achievable. Perturbation analysis is proposed in comparing two and three loudspeakers setup. The simulation results are used to compare their performance numerically in this thesis.

# Acknowledgement

I would like to express my deepest gratitude to my advisor, Dr. S. F. Hsieh, for his enthusiastic guidance and great patience, especially the autonomy in research. I also appreciate my friends for their inspiration and help. Finally, I would like to show my thanks to my parents for their unceasing encouragement and love.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Using immersive audio techniques it is possible to render virtual sound sources in 3D space via a set of loudspeakers or headphones. The goal of such systems is to reproduce the same sound pressure level at the listener's eardrums that would be present if a real sound source was placed in the location of the virtual sound source. However, reproducing the 3D sound via two or more loudspeakers will suffer from several factors to degrade the performances, such as room reverberance, crosstalk disturbance and imperfection of the loudspeakers. In this thesis, we will focus on crosstalk cancellation.

Previous work [13] has investigated that the design of conventional crosstalk cancellation systems which deliver binaural audio to a listener has the serious constraint that the listener may not move. In many literatures, [2, 3, 4, 5, 6], have demonstrated that the performance of crosstalk canceller suffered from head movement and have mathematically calculated the robustness of crosstalk cancellation system. The approach proposed in [1] is to track the listener and adjust the loudspeaker signals to maintain the binaural transmission so that a more robust crosstalk canceller is possible. Recent work [9, 23] has demonstrated that if a number of loudspeakers is used which exceeds the number of points in the listening space, the performance of such a system can be improved. In such case, the reproduction is sought and has better immunity of the head movement. Therefore, multi-channel sound reproduction is our main focus in this thesis.

In this thesis, we will discuss how to synthesize the aural virtual reality environment via multiple loudspeakers. We will focus on how to reduce the crosstalk and design different types of crosstalk canceller so that the 3D sound can be reproduced at listener's ear precisely. Also, we will discuss how to factorize the crosstalk canceller matrix so that economical realizations and less computation are possible. In addition, we will discuss the robustness of crosstalk canceller by perturbation analysis. The perturbation analysis of different structures of crosstalk canceller for multiple loudspeakers arrangement is shown in chapter 4.

This thesis is organized as follows. In chapter 2, we will introduce the properties of HRTFs and discuss the problem of sound reproduction via headphones and loudspeakers. Chapter 3 focuses on the design of crosstalk canceller using three loudspeakers. Chapter 4 is the main part of this thesis; we will investigate the robustness of crosstalk canceller and derive the optimum loudspeakers position for different number of loudspeakers arrangement. In chapter 5, we will use computer simulations to compare performances of different crosstalk cancellers and discuss the robustness of crosstalk canceller. In chapter 6, we will make a conclusion to summarize the results of the simulations.

# Chapter 2

# Transaural Stereo System

A transaural stereo system uses the binaural sound recording and reproduces them. The system may have more than two loudspeakers. We will focus on the transaural system with multiple loudspeakers. In this chapter, we will first introduce how the human localize the sound source by the principle of ITD (interaural time differences) and ILD (interaural level difference).

In section 2.2, we will illustrate how HRTFs (head-related transfer functions) can aid in distinguishing sound location from one position to listener. Section 2.3 will investigate the problems encountered in binaural reproduction via loudspeakers and the idea of crosstalk canceller will be presented in detail. In the last section, we will investigate the design of the transaueal stereo system and two types of layout reformatters will be presented in this section.

## 2.1 Spatial audio

The human hearing process is based on the analysis of input signals to the two ears for differences in intensity, time of arrival, and directional filtering by the outer ear. [11], [12] identified two basic mechanisms as being responsible for sound location which are ITD (interaural time differences) and ILD (interaural level difference). As described in [21], ITD and ILD cues that operated in different wavelength. For short wavelengths (corresponding to frequencies in the range of about 4–20 kHz), the listener's head casts an acoustical shadow giving rise to a lower sound level at the ear farthest from the sound source (ILD) as shown in Figure 2.1. (b). At long wavelengths (corresponding to frequencies in the range of about 20 Hz–1 kHz), the head is very small compared to the wavelength, and localization is based on perceived differences in the time of arrival of sound at the two ears (ITD) as shown in Figure 2.1. (a). The two mechanisms of interaural time and level differences formed the basis of what became known as the *duplex* theory of sound localization. In the frequency range between approximately 1 and 4 kHz, both of these mechanisms are active, which results in several conflicting cues that tend to cause localization errors.

While time or intensity differences provide source direction information in the horizontal plane, in the median plane, time differences are constant and localization is based on spectral filtering. The reflection and diffraction of sound waves from the head, torso, shoulders, and pinna, combined with resonances caused by the ear canal, form the physical basis of the head-related transfer functions (HRTFs).

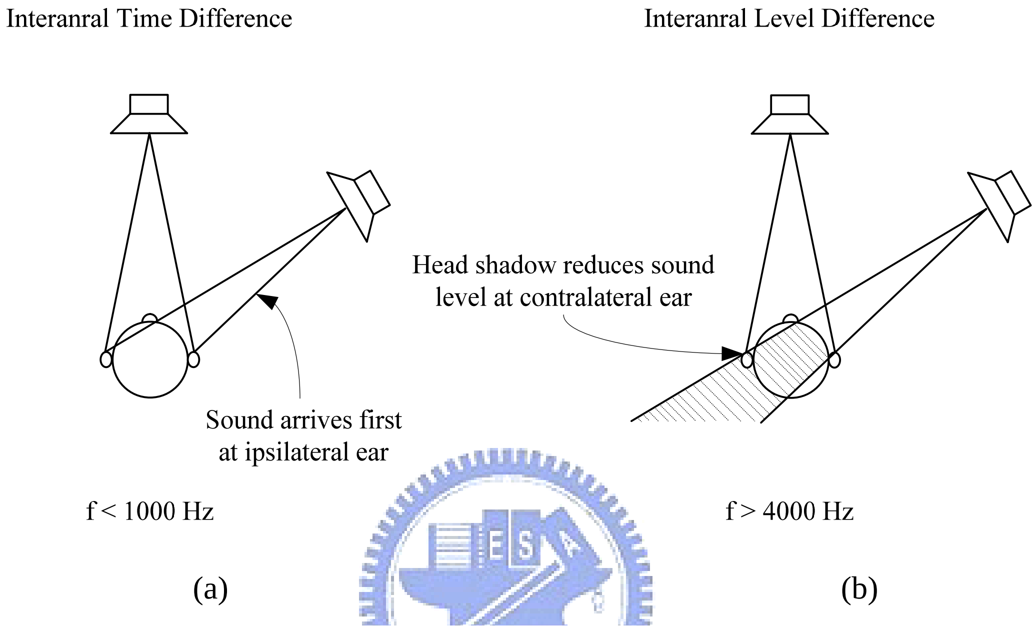


Figure 2.1 The ITD and ILD. (a) In the low-frequency regime, sound is localized based on differences in the time of arrival at each ear. (b) At higher frequencies, sound localization is based on perceived level differences caused by head shadowing.

## 2.2 HRTFs

3D audio rendering systems are based on digital implementations of such head-related transfer functions (HRTFs). In principle, it is possible to achieve excellent reproduction of 3D sound fields using such methods, however, this requires precise measurement of each listener's individual HRTFs. In fact, the magnitude and phase of these head-related transfer functions vary significantly not only for each sound direction, but also from person to person. Current research [7] in this area is focused on achieving good localization performance while using nonindividualized HRTFs derived through averaging or modeling or based on the HRTFs of subjects that have been determined to be “good localizer”. In [26], Begault found that there are currently three major barriers in 3D audio implementations: 1) psychoacoustic errors such as front-back reversals typical in headphone-based systems, 2) large amounts of data required to represent measured HRTFs accurately, and 3) frequency and phase response errors that arise from mismatches between nonindividualized and measured HRTFs.

The HRTFs database we use is from MIT Multi-Media Lab in [27], each HRTFs has 512 samples and with 44.1 kHz sample rate. The pickup of the HRTFs is a pair of microphones embedded in the ears of a dummy head to simulate the ears of the human head. The setup of the dummy head is shown in Figure 2.2. The spherical space around the dummy head was sampled at elevations from -40 degrees (40 degrees below the horizontal plane) to +90 degrees (directly overhead). At each elevation, a full 360 degrees of azimuth was sampled in equal sized increments. The increment sizes were chosen to maintain approximately 5 degree great-circle increments. In total, 710 different positions were sampled at elevations from -40 degrees to +90 degrees.

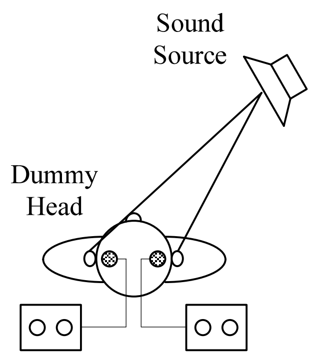


Figure 2.2 The setup of the dummy head.

## 2.3 Crosstalk Cancellation

A simple way to reproduce the 3D sound is using headphones. Listening the binaural signals via headphones can avoid the crosstalk interference, since the audio signals are sent solely to their own destination. Because the listener would feel the virtual sound image exists inside the head and be unable to provide sufficient realistic perceptual feeling. Hence, using headphone for 3D sound might not be proper. To get rid of the phenomenon, we would replace the headphones with two or more loudspeakers.

In the case of conventional crosstalk canceller, this situation states playbacking the binaural signals via a pair loudspeakers for one listener. However, the left ear of listener received the sound not only from the left loudspeaker but also the right loudspeaker and vice versa. This phenomenon is called crosstalk. These crosstalks would disturb the directional perception in 3D sound reproduction since it would change the spectrum, conveying the direction perception cues of the audio signals. In order to eliminate the crosstalk, the crosstalk cancellation should be introduced.

### 2.3.1 Problem Formulation

The algebraic structure of transaural stereo will be developed with the aid of Figure 2.3. In that figure, $N$ program signals $x_1, x_2, \cdots, x_N$ are to be used to create $M$ loudspeaker signals $y_1, y_2, \cdots, y_M$, which in turn result in $L$ ear signals $s_1, s_2, \cdots, s_L$. Let these three sets of signals be represented by the vectors

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}^T \tag{2.1}$$

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_M \end{bmatrix}^T \tag{2.2}$$

$$\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_L \end{bmatrix}^T \tag{2.3}$$

where T denotes matrix transposition. Next, define three matrices of transfer functions. First let **G** be an $L \times M$ matrix, the *acoustic* matrix, such that element $g_{ij}$ is the transfer function to the *i*th ear from the *j*th speaker. Similarly, let **C** be an $M \times N$ matrix, the *crosstalk canceller* matrix, for which $c_{ij}$ is the transfer function of the

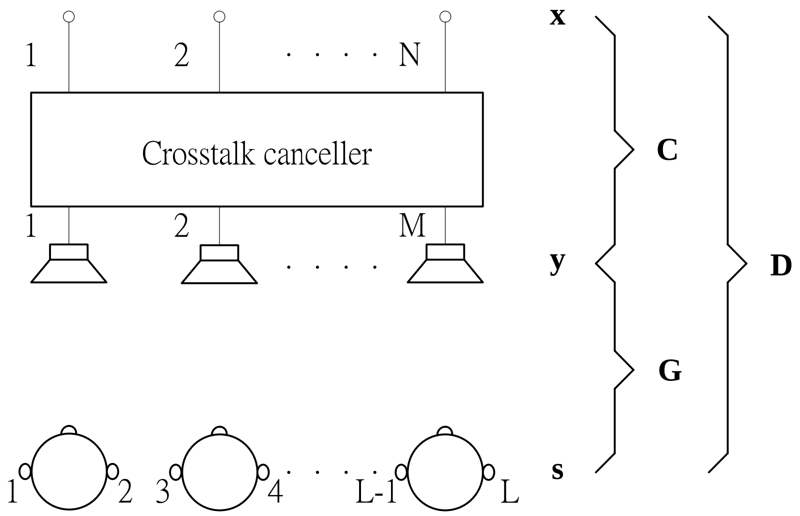


Figure 2.3 Signal and transfer function definitions for transaural stereo.

crosstalk canceller from *j*th input to the *i*th output of the network, and let **D** be the $L \times N$ matrix of *desired* transfer functions describing the overall transfer of signal from the inputs of the crosstalk canceller to the ears, for which $d_{ij}$ is the transfer function to the *i*th ear from the *j*th crosstalk canceller input. With this notation in place, the acoustic propagation can be written as the matrix form

$$\mathbf{s} = \mathbf{G}\mathbf{y}, \tag{2.4}$$

the action of the crosstalk canceller can be written

$$\mathbf{y} = \mathbf{C}\mathbf{x}, \tag{2.5}$$

and the desired transfer of signals is

$$\mathbf{s} = \mathbf{D}\mathbf{x}. \tag{2.6}$$

From these, a solution is found by solving

$$\mathbf{GC} = \mathbf{D} \tag{2.7}$$

for **C**. In this thesis, where $N$ = 2, $M$ = 2, 3, 4 and $L$ = 2 so that **G** is a $2 \times M$ matrix, **C** is an $M \times 2$ and **GC** results in a $2 \times 2$ desired transfer matrix **D**. In order to deliver the binaural signals to listener precisely so that **D** must be an identity matrix. Thus, we will find the crosstalk canceller by solving $\mathbf{G_{2\times M}C_{M\times 2} = I_{2\times 2}}$ in chapter 3.

# Chapter 3

# Crosstalk cancellers

To achieve good reproduction of 3D audio it is necessary to precisely control the acoustic signals at the listener's ear. The simplest way to do this is to deliver binaural signals through headphones. However, in many applications, e.g., home entertainment environment, spatialized video-conferencing, it is preferable that the listener is not required to wear headphones. If loudspeakers are used to deliver binaural signals, the crosstalk signal that arrives at each ear from the other loudspeakers must be canceled. This is achieved by pre-filtering binaural signals before send to loudspeakers and is shown in Figure 3. 1.

In this chapter, we will discuss difference structures of crosstalk canceller of three loudspeakers arrangement. In addition, economical realizations, less computations and performance comparison are our main focus.

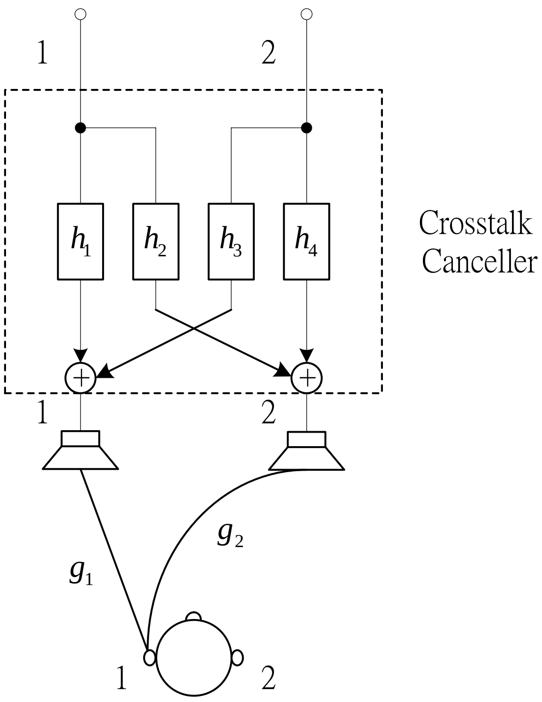


Figure 3.1 Crosstalk cancellation system (Atal and Schroeder in 1963).

## 3.1 Conventional Two Loudspeakers Arrangement

First, we consider the typical crosstalk canceller. Assume a symmetrical case here. This case has two channels, two loudspeakers, and two ears, as shown in Figure 3.1. The acoustic transfer function matrix $\mathbf{G}(Z)$ is

$$\mathbf{G}(z) = \begin{bmatrix} g_1(z) & g_2(z) \\ g_2(z) & g_1(z) \end{bmatrix}. \tag{3.1}$$

In order to reproduce the binaural signal at each ear, $\mathbf{C}(z)$ must be the inverse of the acoustic transfer function matrix $\mathbf{G}(z)$, so that

$$\mathbf{C}(z)=\frac{1}{g_1^2(z)-g_2^2(z)}\begin{bmatrix} g_1(z) & -g_2(z) \\ -g_2(z) & g_1(z) \end{bmatrix} \tag{3.2}$$

From Eq. (3.2), it requires four filters to realize crosstalk cancellation which has heavy computation load. However, some simplification is possible in implementation. In 1989, Cooper and Bauck proposed shuffler structure which $\mathbf{C}(z)$ can be factored by the standard diagonalization technique of finding its eigenvectors and eigenvalues. This results in the shuffler form.

$$\mathbf{C}(z)=\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{2\left(g_1(z)+g_2(z)\right)} & 0 \\ 0 & \frac{1}{2\left(g_1(z)-g_2(z)\right)} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.3}$$

The structure of the shuffler form crosstalk canceller is shown in Figure 3. 2. We can see only two filters $\Sigma$ and $\Delta$ are needed, unlike the conventional four filters in Figure 3.1.

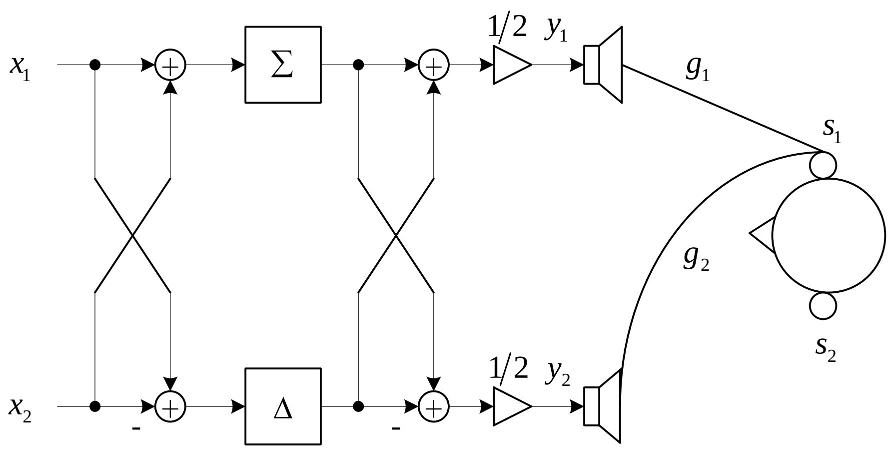


Figure 3.2 Shuffler filter structure (Cooper and Bauck in 1989).

In our approach, we adopt the FIR filter and use the least square error method to find out the finite impulse responses of the filters in Eq. (3.3). Let

$$\Sigma(z)=\frac{1}{g_1(z)+g_2(z)} \tag{3.4}$$

and

$$\Delta(z)=\frac{1}{g_1(z)-g_2(z)} \tag{3.5}$$

then we can obtain two equation as follows:

$$\Sigma(z)\cdot g_1(z)+\Sigma(z)\cdot g_2(z)=d_1(z) \tag{3.6}$$

$$\Delta(z)\cdot g_1(z)-\Delta(z)\cdot g_2(z)=d_2(z) \tag{3.7}$$

where $d_1(z)$ and $d_2(z)$ are the signal that we want to approximate . In the matrix form, these two equations can be expressed as follows:

$$\mathbf{G}\cdot\underline{\mathbf{c}}=\mathbf{D} \tag{3.8}$$

$$\begin{bmatrix}\mathbf{G_1} & 0\\ 0 & \mathbf{G_2}\end{bmatrix}\begin{bmatrix}\underline{\Sigma}\\ \underline{\Delta}\end{bmatrix}=\begin{bmatrix}\underline{\mathbf{d}}_1\\ \underline{\mathbf{d}}_2\end{bmatrix} \tag{3.9}$$

where $G_1$ and $G_2$ represent the convolution matrices of $g_1[n]+g_2[n]$ and $g_1[n]-g_2[n]$, respectively. The vector $\underline{\Sigma}=\{\Sigma[0]\quad\Sigma[1]\quad\cdots\quad\Sigma[J-1]\}$ and $\underline{\Delta}=\{\Delta[0]\quad\Delta[1]\quad\cdots\quad\Delta[J-1]\}$ represent the inverse FIR filters with $J$ taps. The desired response $\underline{\mathbf{d}}_i=\{d_i[0]\quad d_i[1]\quad\cdots\quad d_i[K+J-2]\}=\{1\quad 0\quad 0\quad\cdots\}$ for $i$ =1, 2 represents the impulse response in time domain, where $K$ is the number of taps of the transfer functions.

In this method, we want the convolution sum approaches $\mathbf{D}[i]$ for $i$ = 0, 1, 2 …

$K$+$J$-2 such that the following square error as small as possible.

$$e^2 = \left\| \mathbf{D} - \mathbf{G} \cdot \underline{\mathbf{c}} \right\|^2 \tag{3.10}$$

The least square error solution can be shown as

$$\left( \mathbf{G}^T \mathbf{G} \right) \cdot \underline{\mathbf{c}} = \mathbf{G}^T \mathbf{D} \tag{3.11}$$

$$\underline{\mathbf{c}} = \left( \mathbf{G}^T \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{D} \tag{3.12}$$

Hence, the FIR filter would be

$$\underline{\Sigma} = \{ c[0] \quad c[1] \quad \cdots \quad c[J-1] \} \tag{3.13}$$

$$\underline{\Delta} = \{ c[J] \quad c[J+1] \quad \cdots \quad c[2J-1] \} \tag{3.14}$$

It should be mentioned that the desired signal can not be the pure impulse. It requires an extra delay to decrease the least square error while approximating the desired signal. The extra delay we choice is typically the half of the inverse FIR filter order.

## 3.2 Three Loudspeakers Setup

Next, we look at the situation depicted in Figure 3.3, a symmetrical arrangement of three loudspeakers and one listener. We have

$$\mathbf{G}(z) = \begin{bmatrix} g_1(z) & g_3(z) & g_2(z) \\ g_2(z) & g_3(z) & g_1(z) \end{bmatrix} \tag{3.15}$$

It should be mentioned that the solution of crosstalk canceller is not unique. Since there are three variables to be solved from only two equations. Thus, there exist an infinite number of solutions of $C_i(z)$. In the following sections, we will discuss some possible solutions of crosstalk canceller of three loudspeakers geometric and illustrate how to find the crosstalk canceller.

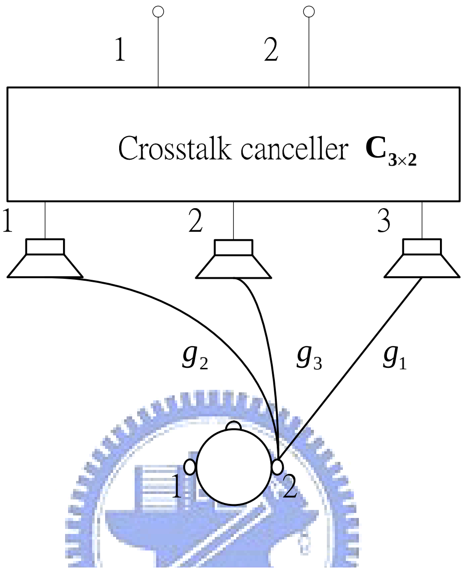


Figure 3.3 Geometry and transfer functions for three loudspeakers.

### 3.2.1 Direct Forward Type

This structure has proposed by Cooper and Bauck in [10]. In order to reproduce the binaural signal at both ear accurately, the general solution is that $\mathbf{C}(z)$ must be the pseudoinverse of $\mathbf{G}(z)$ as shown below

$$\mathbf{G}^{+}(z)=\mathbf{G}^{H}(z)\left(\mathbf{G}(z)\mathbf{G}^{H}(z)\right)^{-1} \tag{3.16}$$

hence, the crosstalk canceller $\mathbf{C}(z)$ is

$$\mathbf{C}(z)=\mathbf{G}^{+}(z) \tag{3.17}$$

$$= \begin{bmatrix} C_1(z) & C_3(z) \\ C_2(z) & C_2(z) \\ C_3(z) & C_1(z) \end{bmatrix} \tag{3.18}$$

in which

$$C_1(z) = \frac{B_1(z)g_1^*(z) - B_2(z)g_2^*(z)}{B_1^2(z) - B_2^2(z)} \tag{3.19}$$

$$C_2(z) = \frac{B_1(z)g_3^*(z) - B_2(z)g_3^*(z)}{B_1^2(z) - B_2^2(z)} \tag{3.20}$$

$$C_3(z) = \frac{B_1(z)g_2^*(z) - B_2(z)g_1^*(z)}{B_1^2(z) - B_2^2(z)} \tag{3.21}$$

and

$$B_1(z) = |g_1(z)|^2 + |g_3(z)|^2 + |g_2(z)|^2 \tag{3.22}$$

$$B_2(z) = g_1(z)g_2^*(z) + |g_3(z)|^2 + g_2(z)g_1^*(z) \tag{3.23}$$

.

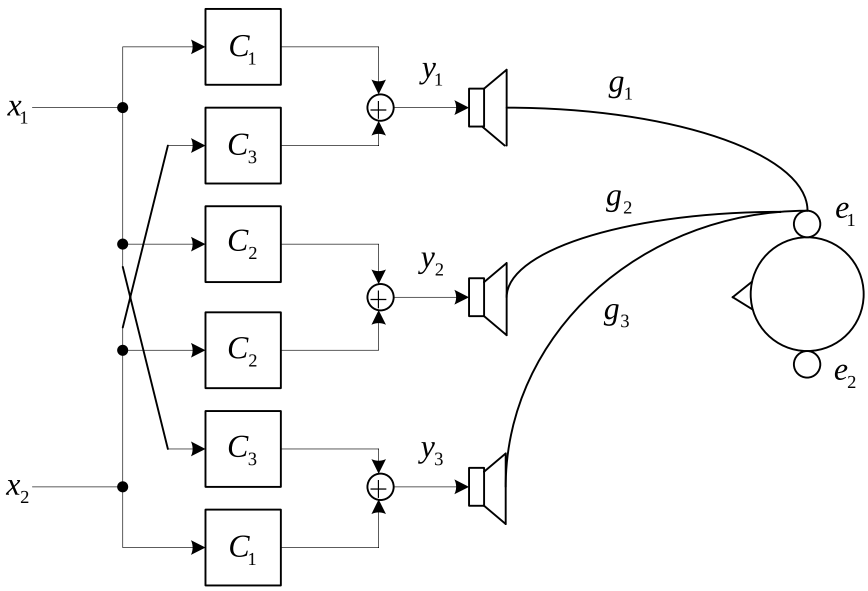


Figure 3.4 The structure of the direct forward type crosstalk canceller.

The structure of the direct forward type crosstalk canceller is shown in Figure 3. 4. To perform these filters directly causes $C_i(z)$ unstable due to the denominator of $C_i(z)$ has non-minimum phase. Thus, in order to avoid the stability problem we can find these filters by using the least square method

### 3.2.2 Least Square Forward Type

In order to overcome the problem of stability, we can find the crosstalk cancellation filters by using the least square method. The structure of this type crosstalk canceller is same as Figure 3. 4. Our goal is that $\mathbf{C}(z)$ multiplied by $\mathbf{G}(z)$ equals an identity matrix $\mathbf{I}_{2\times 2}$ with some extra delay. Since symmetric arrangement, the four equations need to solve become two equations as shown below

$$g_1(z)C_1(z)+g_3(z)C_2(z)+g_2(z)C_3(z)=d(z) \tag{3.24}$$

$$g_2(z)C_1(z)+g_3(z)C_2(z)+g_1(z)C_3(z)=0 \tag{3.25}$$

In the matrix form:

$$\mathbf{G}\underline{\mathbf{c}}=\mathbf{D}_{\mathbf{s}} \tag{3.26}$$

$$\begin{bmatrix} \mathbf{G_1} & \mathbf{G_3} & \mathbf{G}_2 \\ \mathbf{G}_2 & \mathbf{G_3} & \mathbf{G_1} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{c}}_1 \\ \underline{\mathbf{c}}_2 \\ \underline{\mathbf{c}}_3 \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{d}} \\ \underline{\mathbf{0}} \end{bmatrix} \tag{3.27}$$

where $\mathbf{G_1}$, $\mathbf{G}_2$ and $\mathbf{G_3}$ represent the convolution matrices of $g_1[n]$, $g_2[n]$ and $g_3[n]$, respectively. The vector $\underline{\mathbf{c}}_\mathbf{i}$ for $i$ = 1, 2, 3 represent the FIR filters with $J$ taps.

The desired response $\underline{\mathbf{d}} = \{0 \quad 0 \quad \cdots \quad 1 \quad 0 \quad \cdots\}$ represents the impulse response with $K$+$J$-1 taps and $J/2$ modeling delay in time domain. $\underline{\mathbf{0}}$ represents the zero vector with $K$+$J$-1 taps. Thus, the least square error solution can be obtained by Eq. (3.11) and Eq. (3.12) then the FIR filter would be

$$\underline{\mathbf{c}}_1 = \{c[0] \quad c[1] \quad \cdots \quad c[J-1]\} \tag{3.28}$$

$$\underline{\mathbf{c}}_2 = \{c[J] \quad c[J+1] \quad \cdots \quad c[2J-1]\} \tag{3.29}$$

$$\underline{\mathbf{c}}_3 = \{c[2J] \quad c[2J+1] \quad \cdots \quad c[3J-1]\} \tag{3.30}$$

This type crosstalk canceller requires six FIR filters and causes complex calculation. In the next section, we will focus on factoring the crosstalk canceller matrix $\mathbf{C}(z)$ so that economical realizations are possible.

### 3.2.3 Shuffler Structure

This structure has proposed by Cooper and Bauck in [10]. Eq. (3.18) is useful in showing that six filters are required, but of the six there are only three specifications. Noticing that if we ignore the middle row of Eq. (3.18), the remaining elements have the same symmetric form as Eq. (3.2), and we can benefit from the factorization of Eq. (3.3). With this aid, we can write the factorization of Eq. (3.18) by inspection, reinserting the middle row:

$$\mathbf{C}(z)=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}\begin{bmatrix}\frac{C_1(z)+C_3(z)}{2} & 0\\ C_2(z) & 0\\ 0 & \frac{C_1(z)-C_3(z)}{2}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{3.31}$$

$$=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}\begin{bmatrix}\frac{\Sigma_2(z)}{2} & 0\\ C_2(z) & 0\\ 0 & \frac{\Delta_2(z)}{2}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{3.32}$$

The shuffler topology of three loudspeakers setup is shown in Figure 3. 5. It can be shown that of the three loudspeaker signals $y_1$, $y_2$ and $y_3$, $y_1$ and $y_3$ are created from the symmetric combination of $x_1$ and $x_2$, one signal can be found from the other by replacing $x_1$ by $x_2$ and vice versa. In addition, $y_2$ is formed of only a filtered $x_1+x_2$ signal. If $\mathbf{G}(z)$ are real-valued, then

$$\Sigma_2(z)=\frac{g_1(z)+g_2(z)}{\left(g_1(z)+g_2(z)\right)^2+2g_3^2(z)} \tag{3.33}$$

$$C_2(z)=\frac{g_3(z)}{\left(g_1(z)+g_2(z)\right)^2+2g_3^2(z)} \tag{3.34}$$

$$\Delta_2(z)=\frac{1}{g_1(z)-g_2(z)} \tag{3.35}$$

Note that these filters become identical to the corresponding filters of crosstalk canceller of Eq. (3.3) when $g_3(z)=0$. The impulse response of these filters are shown in Figure 3.6, Figure 3.7 and Figure 3.8.

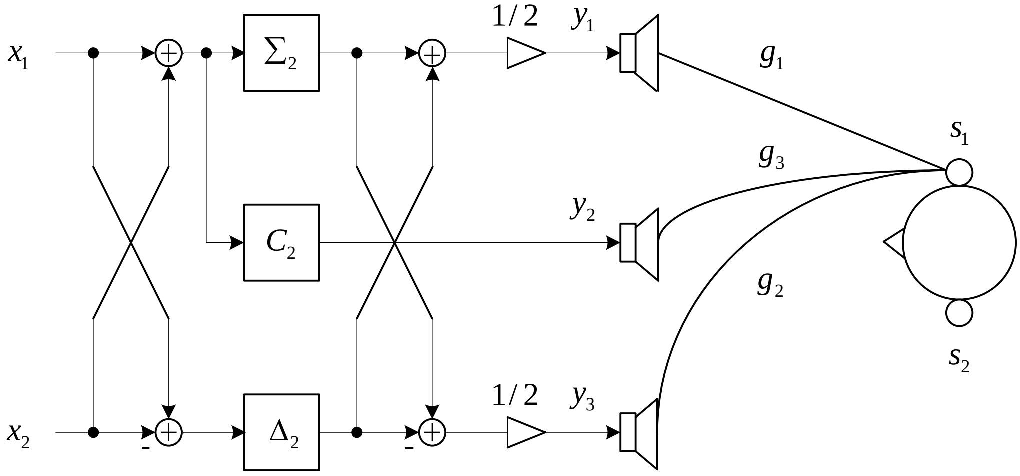


Figure 3.5 The shuffler topology of three loudspeakers geometric (Cooper and Bauck in 1996).

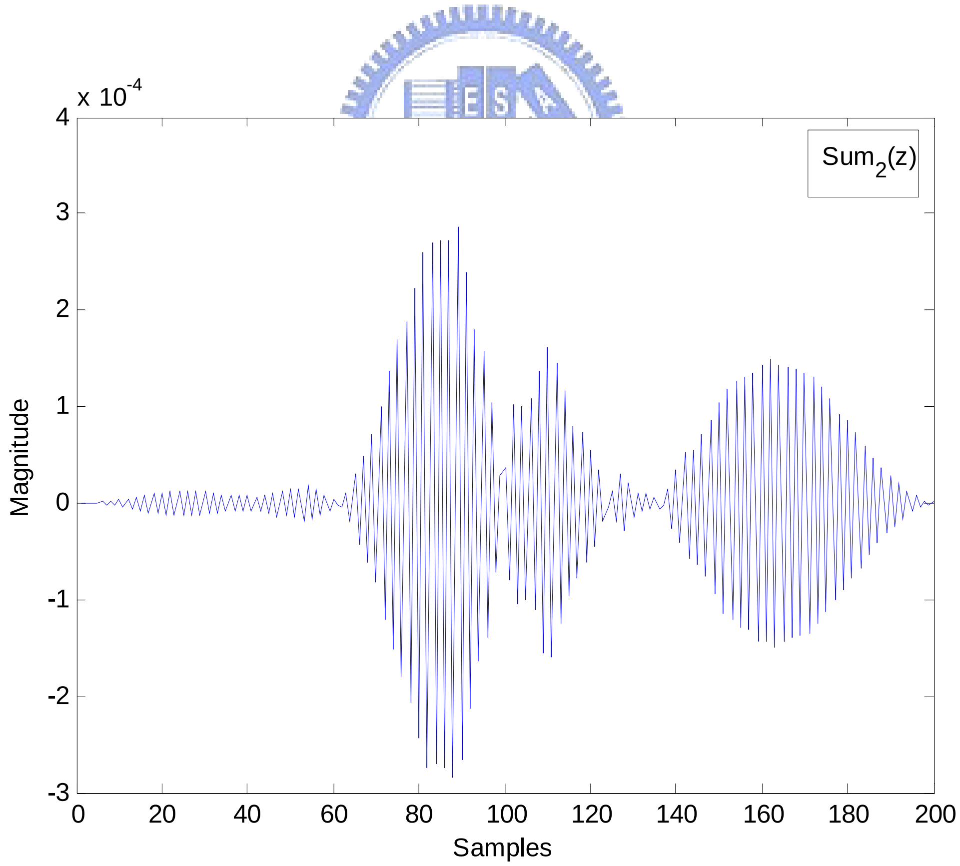


Figure 3.6 The impulse response of the filter $\sum_2$.

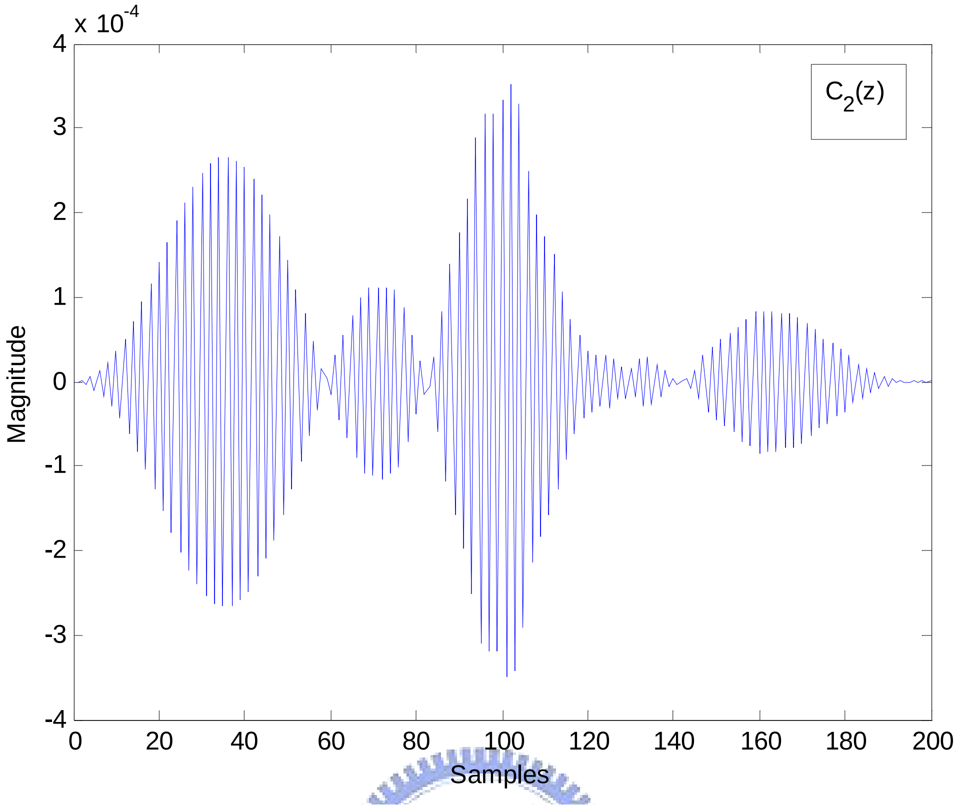


Figure 3.7 The impulse response of the filter $C_2$.

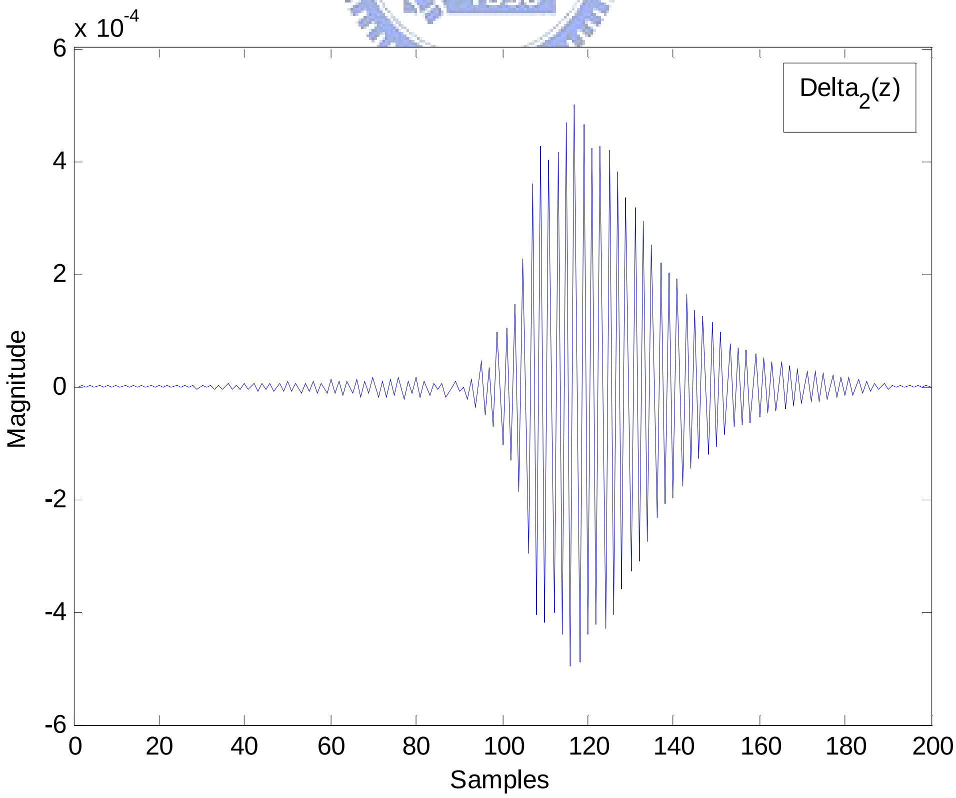


Figure 3.8 The impulse response of the filter $\Delta_2$.

### 3.2.4 A Simplified Shuffler Form

As described above, the direct forward type requires six filters to realize crosstalk cancellation and has heavy computation. Although the shuffler structure requires fewer filters than direct implementation, the crosstalk cancellation filters of such structure are not easy to realize. In this section, we derive one possible solution of crosstalk canceller, that is, a simplified shuffler form, in which only two filters are needed to realize crosstalk cancellation. In addition, the crosstalk cancellation filters of this structure are easier to implement than those in shuffler form.

Now, we illustrate how to find $\mathbf{C}$. Ideally, $\mathbf{G}_{2\times 3}\mathbf{C}_{3\times 2} = \mathbf{I}_{2\times 2}$. Since symmetric arrangement so that $\mathbf{G}_{2\times 3}\mathbf{C}_{3\times 2}$ results in two equations as follows:

$$C_1(z)g_1(z)+C_2(z)g_3(z)+C_3(z)g_2(z)=1 \tag{3.36}$$

$$C_1(z)g_2(z)+C_2(z)g_3(z)+C_3(z)g_1(z)=0 \tag{3.37}$$

From Eq. (3.36) and Eq. (3.37), there are three variables to be solved by two equations so that there exist infinite solutions of $C_i(z)$. The $C_i(z)$ we find here is one possible solution. First, by subtracting and adding Eq. (3.36) and Eq. (3.37), we have

$$C_1(z)-C_3(z)=\frac{1}{g_1(z)-g_2(z)} \tag{3.38}$$

$$\left(C_1(z)+C_3(z)\right)\left(g_1(z)+g_2(z)\right)+2C_2(z)g_3(z)=1 \tag{3.39}$$

In order to cancel the right-hand-side of Eq. (3.39), we then choose

$$C_2(z)=\frac{1}{2g_3(z)} \tag{3.40}$$

In addition, $g_1(z)+g_2(z)\neq 0$ so that

$$C_1(z)+C_3(z)=0 \tag{3.41}$$

From Eq. (3.38) and Eq. (3.41), we can obtain:

$$C_1(z)=\frac{1}{2\left(g_1(z)-g_2(z)\right)} \tag{3.42}$$

$$C_3(z)=\frac{-1}{2\left(g_1(z)-g_2(z)\right)} \tag{3.43}$$

Thus, one possible solution of crosstalk canceller can be:

$$\mathbf{C}(z)=\frac{1}{2}\begin{bmatrix} \frac{1}{g_1(z)-g_2(z)} & \frac{-1}{g_1(z)-g_2(z)} \\ \frac{1}{g_3(z)} & \frac{1}{g_3(z)} \\ \frac{-1}{g_1(z)-g_2(z)} & \frac{1}{g_1(z)-g_2(z)} \end{bmatrix} \tag{3.44}$$

Also, we can utilize the factorization as described in section 3.2.3. If we ignore the middle row of Eq. (3.44), the remaining elements have the same symmetric form as Eq. (3.2), then we can benefit from the factorization of Eq. (3.3) and we can write the factorization of Eq. (3.44) by reinserting the middle row as shown below

$$\mathbf{C}(z)=\frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ \frac{1}{g_3(z)} & 0 \\ 0 & \frac{1}{g_1(z)-g_2(z)} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.45}$$

$$=\frac{1}{2}\begin{bmatrix}0 & 1\\ 1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}H_3(z) & 0\\ 0 & \Delta_3(z)\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{3.46}$$

The structure of this type crosstalk canceller is shown in Figure 3. 9 and the impulse response of these filters are shown in Figure 3.10 and Figure 3.11. Obviously, this structure is easier to implement than direct forward type and shuffler form. In addition, it requires only two filters so that economical realization and less computation can be achievable.

It should be mentioned that only the sum signal $x_1 + x_2$ is fed to the center loudspeaker, and only the difference signal $\pm(x_1 - x_2)$ is fed to the side loudspeakers. One side gets the same signal as the other side loudspeaker, but inverted in sign. Feeding the side loudspeakers with opposite sign signals is similar to a dipole, especially when the side loudspeakers are placed close together as described in [17].

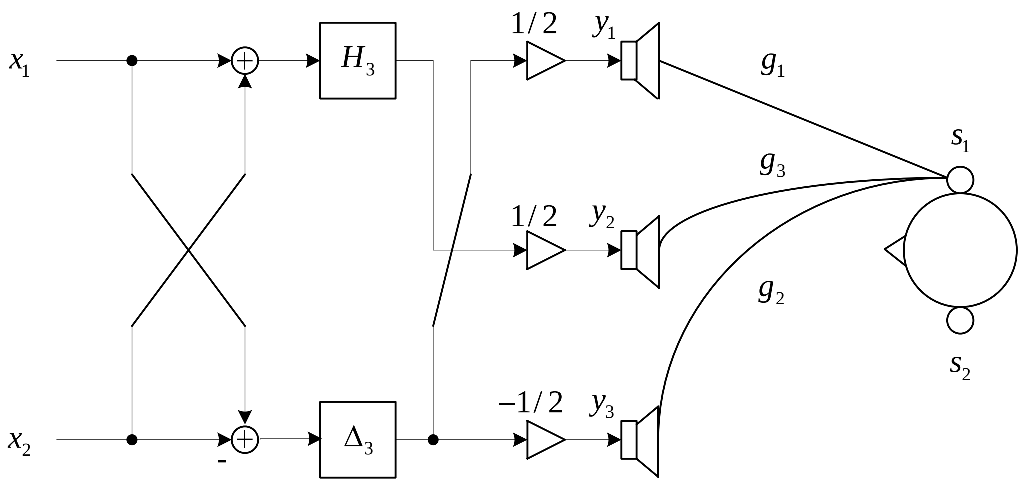


Figure 3.9 The simplified shuffler structure of crosstalk canceller for three loudspeakers arrangement.

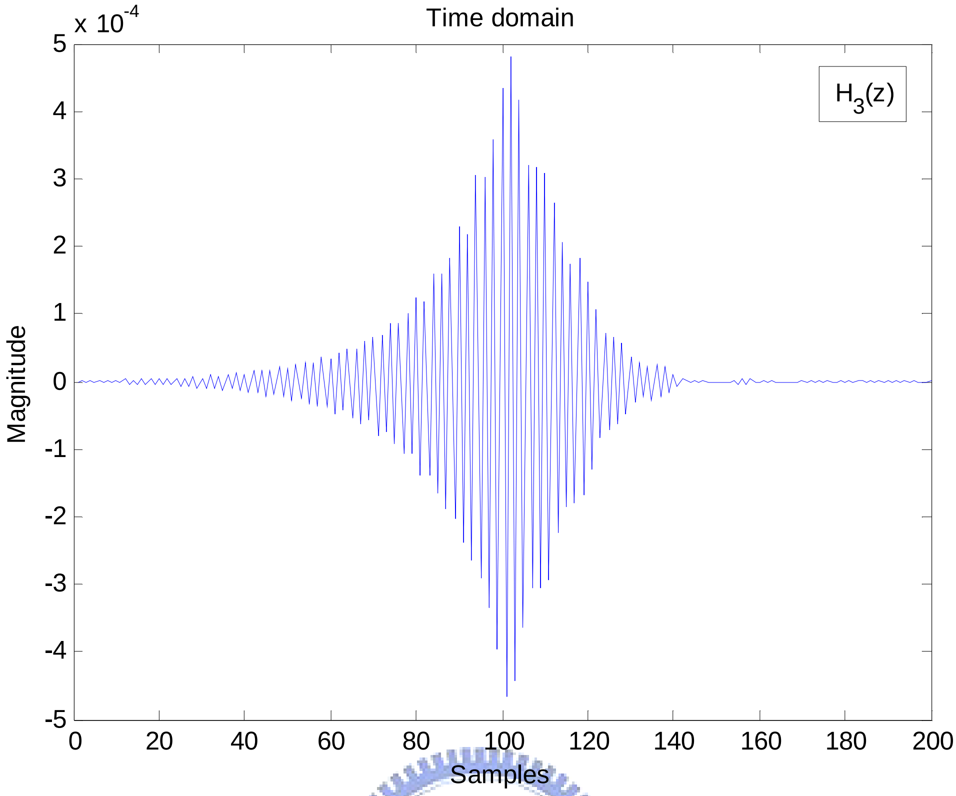


Figure 3.10 The impulse response of the filter $H_3$.

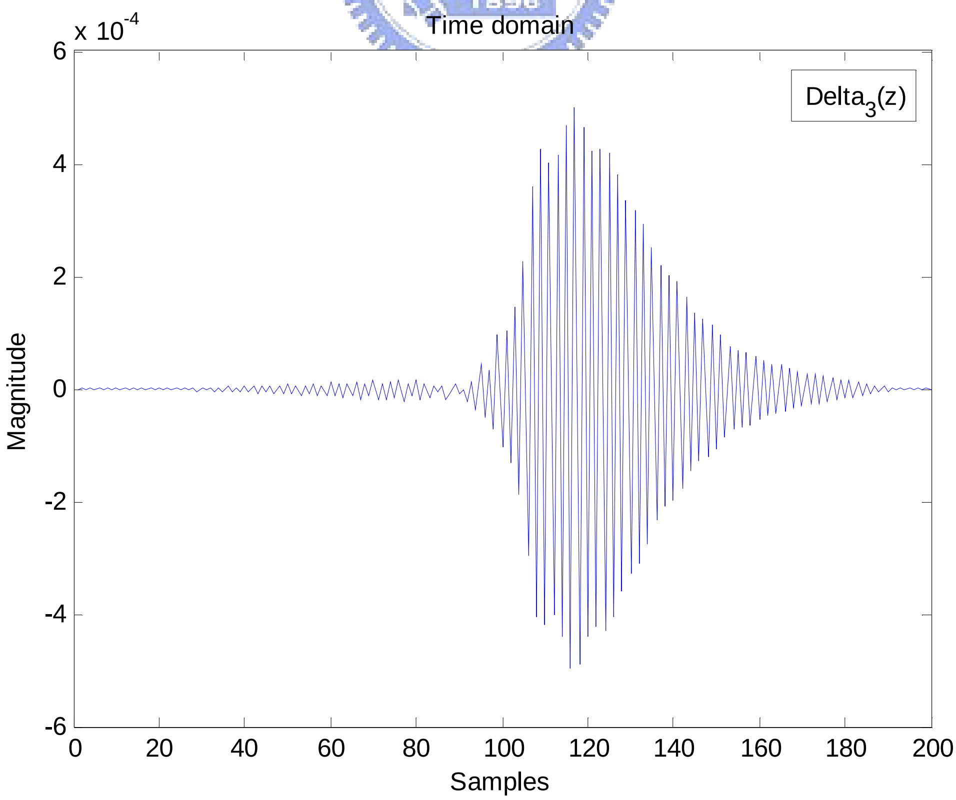


Figure 3.11 The impulse response of the filter $\Delta_3$.

### 3.2.5 A Simplified Shuffler Form of Four Loudspeakers Arrangement

Similarly, we can extend to four loudspeakers and derive the simplified shuffler form. In this case, we deliver the binaural signals to listener via four loudspeakers as shown in Figure 3.12. Under symmetric arrangement, the acoustic matrix **G** is a $2\times4$ matrix and the crosstalk canceller matrix $\mathbf{C}$ is a $4\times2$ matrix as shown below

$$\mathbf{G}(z)=\begin{bmatrix} g_1(z) & g_2(z) & g_3(z) & g_4(z) \\ g_4(z) & g_3(z) & g_2(z) & g_1(z) \end{bmatrix} \tag{3.53}$$

$$\mathbf{C}(z)=\begin{bmatrix} C_1(z) & C_4(z) \\ C_2(z) & C_3(z) \\ C_3(z) & C_2(z) \\ C_4(z) & C_1(z) \end{bmatrix} \tag{3.54}$$

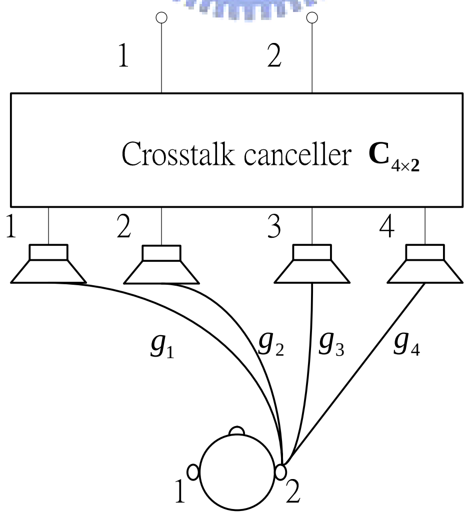


Figure 3.12 Geometry and transfer functions for four loudspeakers.

In addition, our desired transfer function matrix **D** is a $2\times 2$ identity matrix. Thus, we can obtain two equations as follows

$$C_1(z)g_1(z)+C_2(z)g_2(z)+C_3(z)g_3(z)+C_4(z)g_4(z)=1 \tag{3.55}$$

$$C_1(z)g_4(z)+C_2(z)g_3(z)+C_3(z)g_2(z)+C_4(z)g_1(z)=0 \tag{3.56}$$

By subtracting and adding Eq. (3.55) and Eq. (3.56), we have

$$\begin{cases}\left(g_1(z)-g_4(z)\right)\left(C_1(z)-C_4(z)\right)+\left(g_2(z)-g_3(z)\right)\left(C_2(z)-C_3(z)\right)=1\\ \left(g_1(z)+g_4(z)\right)\left(C_1(z)+C_4(z)\right)+\left(g_2(z)+g_3(z)\right)\left(C_2(z)+C_3(z)\right)=1\end{cases} \tag{3.57}$$

Next, we choose $C_2(z)=C_3(z)$ and $C_2(z)+C_3(z)=1/\left(g_2(z)+g_3(z)\right)$ so that

$$\begin{cases}\left(g_1(z)-g_4(z)\right)\left(C_1(z)-C_4(z)\right)=1\\ \left(g_1(z)+g_4(z)\right)\left(C_1(z)+C_4(z)\right)=0\end{cases} \tag{3.58}$$

We can further obtain

$$C_1(z)=-C_4(z)=\frac{1}{2}\frac{1}{g_1(z)-g_4(z)} \tag{3.59}$$

and

$$C_2(z)=C_3(z)=\frac{1}{2}\frac{1}{g_2(z)+g_3(z)} \tag{3.60}$$

then one possible solution of crosstalk canceller can be

$$\mathbf{C}(z)=\frac{1}{2}\begin{bmatrix}\dfrac{1}{g_1(z)-g_4(z)} & \dfrac{-1}{g_1(z)-g_4(z)}\\ \dfrac{1}{g_2(z)+g_3(z)} & \dfrac{1}{g_2(z)+g_3(z)}\\ \dfrac{1}{g_2(z)+g_3(z)} & \dfrac{1}{g_2(z)+g_3(z)}\\ \dfrac{-1}{g_1(z)-g_4(z)} & \dfrac{1}{g_1(z)-g_4(z)}\end{bmatrix} \tag{3.61}$$

Using the factorization as described in section 3.2.3, we have

$$\mathbf{C}(z)=\begin{bmatrix}0 & 1\\ 1 & 0\\ 1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\dfrac{1}{2\left(g_2(z)+g_3(z)\right)} & 0\\ 0 & \dfrac{1}{2\left(g_1(z)-g_4(z)\right)}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

$$=\frac{1}{2}\begin{bmatrix}0 & 1\\ 1 & 0\\ 1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\Sigma_4(z) & 0\\ 0 & \Delta_4(z)\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} \tag{3.62}$$

The structure of this type crosstalk canceller is shown in Figure 3.13. Same as the simplified shuffler form of three loudspeakers setup, it requires only the inversed-sum filter $\Sigma_4(z)$ and the inversed-difference filter $\Delta_4(z)$ to realize the crosstalk canceller.

It should be mentioned that the approach we propose here is only one possible solution. There must be lots of different implementation of crosstalk canceller for four loudspeakers arrangement.

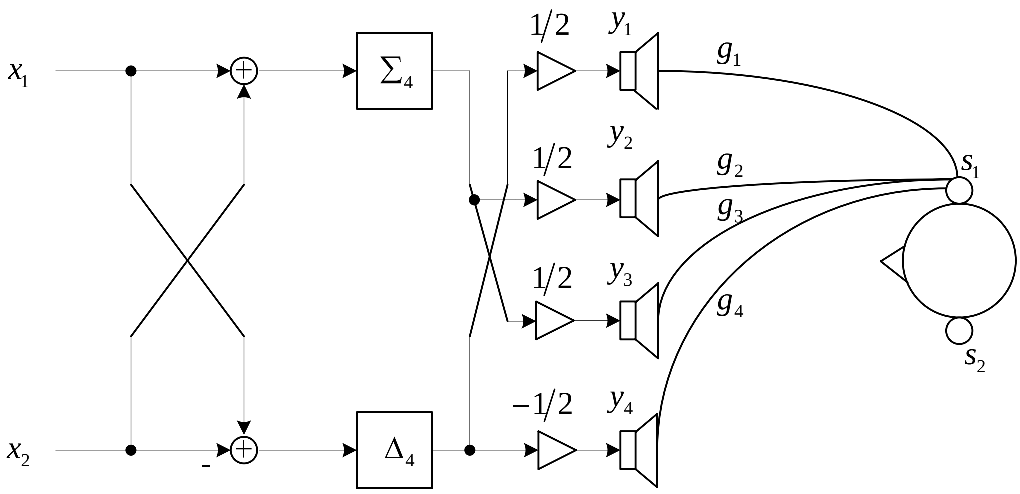


Figure 3.13 The simplified shuffler structure of crosstalk canceller for four loudspeakers arrangement.

## 3.2.6 Reduced-Order Modeling

In order to efficiently reduce the orders of the model, we use pole-zero models to approximate the impulse response of FIR filters. Here, we use Prony appoximation method to model. Either model can be expressed by the system function:

$$C(z)=\frac{B_q(z)}{A_p(z)}=\frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1+\sum_{k=1}^{p} a_p(k) z^{-k}} \tag{3.63}$$

The Prony approximation wants to minimize the error:

$$E(z)=A_p(z)x(z)-B_q(z) \tag{3.64}$$

where $x(z)$ is the system's impulse response that we want to model. Since $b_q(n)=0$ for $n>q$, we may write the error explicitly for each n as follows:

$$e(n)=\begin{cases} x[n]+\sum_{l=1}^{p} a_p[l]x[n-l]-b_q[n] & ; \quad n=0,1,\ldots,q \\ x[n]+\sum_{l=1}^{p} a_p[l]x[n-l] & ; \quad n>q \end{cases} \tag{3.65}$$

Prony approximation begins by finding the coefficients $a_p(k)$ that minimize the square error

$$e_{prony}=\sum_{n=q+1}^{\infty}\left|e[n]\right|^2=\sum_{n=q+1}^{\infty}\left|x[n]+\sum_{l=1}^{p} a_p[l]x[n-l]\right|^2 \tag{3.66}$$

Form this equation, we know the modeled impulse response $\hat{x}[n]$ approximates $x[n]$ without error over the interval $[0,q]$. However, there is modeling error between $x[n]$ and $\hat{x}[n]$. We will discuss how the modeling error affects the performance of crosstalk cancellation in chapter 4.

# Chapter 4

# Perturbation Analysis

The main disadvantage of crosstalk canceller system is that it is critically dependent on the listener's head being in a fixed design position, the so called "sweet-spot". Many studies have shown that the lateral movement away from the design position of as little as a few centimeters results in loss of the 3D audio effect. In this chapter, we will discuss the robustness of two and three loudspeakers arrangement. The analysis of shuffler form and simplified shuffler form are our main focus here.

In addition, we will focus on the performances of crosstalk suppression, and see how perturbations or modeling errors of these inverse filters will affect the crosstalk suppression in the crosstalk canceller. In the following section, we will discuss which type of crosstalk canceller has better immunity to perturbation.

# 4.1 Perturbation Analysis for Filter Modeling

First, we will investigate the robustness of crosstalk canceller due to filter modeling error as shown in Figure 4.1. Ideally, the product of $\mathbf{C_0}$ and $\mathbf{G}$ results in an identity matrix. After modeling, $\mathbf{C}_0$ becomes $\mathbf{C}_+$ which means there are modeling errors in $\mathbf{C}_+$. Hence, the product of $\mathbf{C}_+$ and $\mathbf{G}$ is no longer an identity matrix, since the modeling error affects the performance of equalization and crosstalk cancellation. In this section, we will analyze how the modeling errors affect the performance of two and three loudspeakers shuffler forms and simplified shuffler form, since their structures are similar.

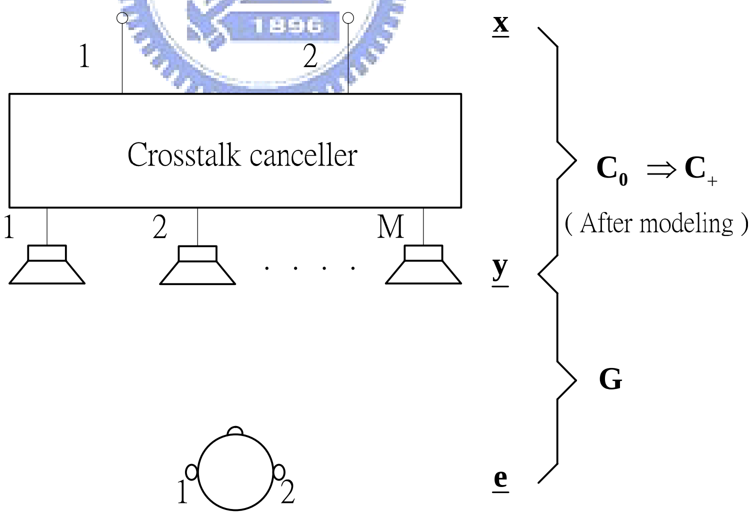


Figure 4.1 The geometry of crosstalk cancellation system for multiple loudspeakers after filter modeling

### 4.1.1 Analysis on Shuffler Form of Two Loudspeakers Setup

From Figure 3.2, we can derive the relation between $Y$ and $X$ in a matrix form.

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \Sigma+\Delta & \Sigma-\Delta \\ \Sigma-\Delta & \Sigma+\Delta \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{4.1}$$

After modeling, the filters $\Sigma$ and $\Delta$ will be changed to be $\Sigma_+$ and $\Delta_+$ with modeling error $e_{s,1}$ and $e_{d,1}$, respectively, so that the relation between $Y_+$ and $X$ can be written as follows:

$$\begin{aligned}\begin{bmatrix} Y_{+,1} \\ Y_{+,2} \end{bmatrix} &= \frac{1}{2}\begin{bmatrix} \Sigma_+ + \Delta_+ & \Sigma_+ - \Delta_+ \\ \Sigma_+ - \Delta_+ & \Sigma_+ + \Delta_+ \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} \Sigma+e_{s,1}+\Delta+e_{d,1} & \Sigma+e_{s,1}-\Delta-e_{d,1} \\ \Sigma+e_{s,1}-\Delta-e_{d,1} & \Sigma+e_{s,1}+\Delta+e_{d,1} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\end{aligned} \tag{4.2}$$

where $Y_{+,1}$ and $Y_{+,2}$ represent the perturbed signals sent to the loudspeaker pair. The modeling errors after crosstalk cancellation can be expressed as follows:

$$E_{2,Shuffler} = \frac{1}{2}\begin{bmatrix} e_{s,1}+e_{d,1} & e_{s,1}-e_{d,1} \\ e_{s,1}-e_{d,1} & e_{s,1}+e_{d,1} \end{bmatrix}\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \tag{4.3}$$

Now, we begin the analysis. First, we factorize the filters $\Sigma$ and $\Delta$ as shown below

$$\Sigma = \frac{1}{g_1+g_2} = \frac{1}{g_1}\cdot\frac{1}{1+F} \tag{4.4}$$

$$\Delta = \frac{1}{g_1-g_2} = \frac{1}{g_1}\cdot\frac{1}{1-F} \tag{4.5}$$

In Eq. (4.4) and Eq. (4.5), we let

$$F = \frac{g_2}{g_1} \tag{4.6}$$

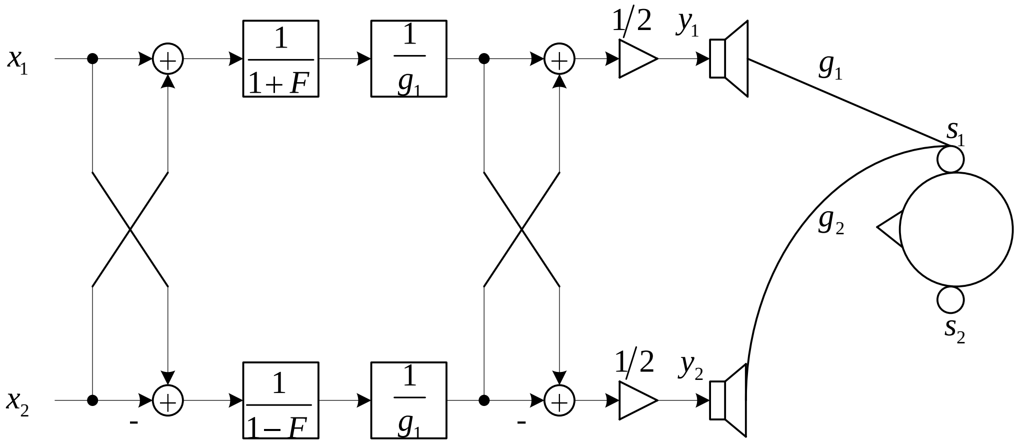


Figure 4.2 The factorized two loudspeakers shuffler form.

Thus, we can obtain the structure as shown in Figure 4.2. It should be mentioned that we focus on the perturbation of $F$ due to filter modeling so that we assume there are no modeling errors in $1/g_1$. By utilizing Taylor series, we have

$$\Sigma = \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} (-F)^n \right] \tag{4.7}$$

$$\Delta = \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} F^n \right] \tag{4.8}$$

After low order filter design, $F$ becomes $F_+$ with filter modeling error $e$, i.e.

$$e = F_+ - F \tag{4.9}$$

Therefore, we have

$$\begin{aligned} \Sigma_+ &= \frac{1}{g_1} \cdot \frac{1}{1+F_+} \\ &= \frac{1}{g_1} \cdot \left[ 1 - (F+e) + (F+e)^2 - (F+e)^3 + \cdots \right] \\ &= \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} (-1)^n (F+e)^n \right] \end{aligned} \tag{4.10}$$

Using binomial series, we can obtain

$$\Sigma_{+} = \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} (-1)^n \sum_{k=0}^{n} \binom{n}{k} e^k F^{n-k} \right]$$

$$= \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} (-F)^n + \sum_{n=1}^{\infty} \sum_{k=1}^{n} (-1)^n \binom{n}{k} e^k F^{n-k} \right] \quad (4.11)$$

Similarly,

$$\Delta_{+} = \frac{1}{g_1} \cdot \frac{1}{1 - F_{+}}$$

$$= \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} F^n + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k} \right] \quad (4.12)$$

We then have

$$e_{s,1} = \Sigma_{+} - \Sigma$$

$$= \frac{1}{g_1} \cdot \sum_{n=1}^{\infty} \sum_{k=1}^{n} (-1)^n \binom{n}{k} e^k F^{n-k}$$

$$= \frac{1}{g_1} \cdot \left( -e + 2eF + e^2 - 3eF^2 + 3e^2 F - e^3 + \cdots \right) \quad (4.13)$$

Assume $\left(2eF - e\right) \gg \left(e^2 - 3eF^2 + 3e^2 F - e^3 + \cdots\right)$ so that

$$e_{s,1} \cong \frac{1}{g_1} \cdot \left(2F - 1\right) e \quad (4.14)$$

and

$$e_{d,1} = \Delta_{+} - \Delta$$

$$= \frac{1}{g_1} \cdot \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k}$$

$$= \frac{1}{g_1} \cdot \left( e + 2eF + e^2 + 3eF^2 + 3e^2 F + e^3 + \cdots \right) \quad (4.15)$$

Assume the common error e much small so that

$$e_{d,1} \cong \frac{1}{g_1} \cdot (2F+1)e \tag{4.16}$$

Thus, the modeling error vector can be written as follows

$$E_{2,Shuffler} \cong \frac{1}{g_1} \begin{bmatrix} 2Fe & -e \\ -e & 2Fe \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \tag{4.17}$$

To further simplify this analysis, we will assume vectors are reduced to scalars, $g_1 = 1$ and $g_2 = r$ to account for interaural intensity difference. Hence, $F = r$ and the error vector can be rewritten as

$$\begin{aligned} E_{2,Shuffler} &\cong \begin{bmatrix} 2re & -e \\ -e & 2re \end{bmatrix} \begin{bmatrix} 1 \\ r \end{bmatrix} \\ &= \begin{bmatrix} er \\ e(2r^2 - 1) \end{bmatrix} \end{aligned} \tag{4.18}$$

The first element of this error vector indicates the error of equalization part and the second element represents the error of crosstalk part due to the filter modeling. In [20], the author has investigated the perturbation analysis on direct forward type 2. The error vector can be derived as

$$\begin{aligned} E_2 &= \frac{1}{1-r^2} \begin{bmatrix} 0 & e \\ e & 0 \end{bmatrix} \begin{bmatrix} 1 \\ r \end{bmatrix} \\ &= \frac{1}{1-r^2} \begin{bmatrix} er \\ e \end{bmatrix} \end{aligned} \tag{4.19}$$

It is difficult to compare which structure has better immunity of the perturbation. We will make more detailed comparison from simulated results as shown in chapter 5.

### 4.1.2 Analysis on Shuffler Form of Three Loudspeakers Arrangement

Also, we take the same analysis on shuffler form of three loudspeakers arrangement and derive the modeling error vector for comparison. In the same way, we derive the relation between $Y$ and $X$ in the shuffler form crosstalk canceller of three loudspeakers arrangement as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\Sigma_2 + \Delta_2) & \frac{1}{2}(\Sigma_2 - \Delta_2) \\ C_2 & C_2 \\ \frac{1}{2}(\Sigma_2 - \Delta_2) & \frac{1}{2}(\Sigma_2 + \Delta_2) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{4.20}$$

After modeling, the filters $\Sigma_2$, $C_2$ and $\Delta_2$ will be changed to be $\Sigma_{+,2}$, $C_{+,2}$ and $\Delta_{+,2}$ with modeling error $e_{s,2}$, $e_{c,2}$ and $e_{d,2}$, respectively, so that the relation between $Y_+$ and $X$ can be written as follows:

$$\begin{aligned} \begin{bmatrix} Y_{+,1} \\ Y_{+,2} \\ Y_{+,3} \end{bmatrix} &= \begin{bmatrix} \frac{1}{2}(\Sigma_{+,2} + \Delta_{+,2}) & \frac{1}{2}(\Sigma_{+,2} - \Delta_{+,2}) \\ C_{+,2} & C_{+,2} \\ \frac{1}{2}(\Sigma_{+,2} - \Delta_{+,2}) & \frac{1}{2}(\Sigma_{+,2} + \Delta_{+,2}) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}(\Sigma_2 + e_{s,2} + \Delta_2 + e_{d,2}) & \frac{1}{2}(\Sigma_2 + e_{s,2} - \Delta_2 - e_{d,2}) \\ C_2 + e_{c,2} & C_2 + e_{c,2} \\ \frac{1}{2}(\Sigma_2 + e_{s,2} - \Delta_2 - e_{d,2}) & \frac{1}{2}(\Sigma_2 + e_{s,2} + \Delta_2 + e_{d,2}) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \end{aligned} \tag{4.21}$$

where $Y_{+,1}$, $Y_{+,2}$ and $Y_{+,3}$ represent the perturbed signals sent to the loudspeakers. Then the modeling error after crosstalk cancellation can be expressed as follows:

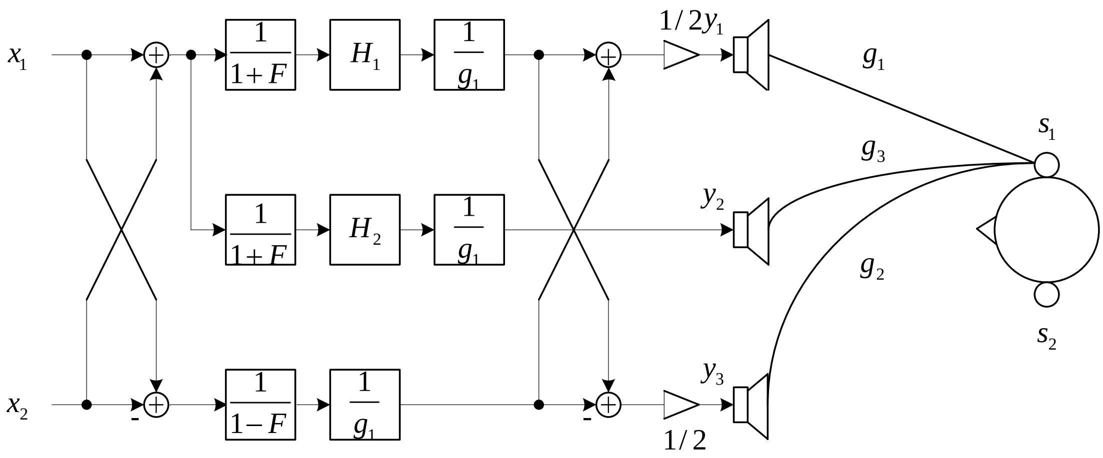


Figure 4.3 The factorized three loudspeakers shuffler form.

$$E_{3,shuffler} = \begin{bmatrix} \frac{1}{2}\left(e_{s,2}+e_{d,2}\right) & e_{c,2} & \frac{1}{2}\left(e_{s,2}-e_{d,2}\right) \\ \frac{1}{2}\left(e_{s,2}-e_{d,2}\right) & e_{c,2} & \frac{1}{2}\left(e_{s,2}+e_{d,2}\right) \end{bmatrix} \begin{bmatrix} g_1 \\ g_3 \\ g_2 \end{bmatrix} \tag{4.22}$$

Now, we begin the analysis. First, we factorize the filters $\Sigma_2$, $C_2$ and $\Delta_2$ as follows:

$$\Sigma_2 = \frac{g_1+g_2}{\left(g_1+g_2\right)^2+2g_3^2} = \frac{1}{g_1}\cdot\frac{1}{1+2\left(\frac{g_3}{g_1+g_2}\right)^2}\cdot\frac{1}{1+F} \tag{4.23}$$

$$C_2 = \frac{g_3}{\left(g_1+g_2\right)^2+2g_3^2} = \frac{1}{g_1}\cdot\frac{\left(g_1+g_2\right)g_3}{\left(g_1+g_2\right)^2+2g_3^2}\cdot\frac{1}{1+F} \tag{4.24}$$

$$\Delta_2 = \frac{1}{g_1-g_2} = \frac{1}{g_1}\cdot\frac{1}{1-F} \tag{4.25}$$

where $F$ is described in section 4.1.1. Thus, we can obtain the structure as shown in Figure 4.3. The filters $H_1$ and $H_2$ in Figure 4.3 are shown below

$$\begin{cases} H_1 = \dfrac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2} \\ H_2 = \dfrac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2} \end{cases} \tag{4.26}$$

By utilizing Taylor series, we have

$$\Sigma_2 = \frac{1}{g_1}\cdot\frac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2}\cdot\left[1+\sum_{n=1}^{\infty}(-F)^n\right] \tag{4.27}$$

$$C_2 = \frac{1}{g_1}\cdot\frac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2}\cdot\left[1+\sum_{n=1}^{\infty}(-F)^n\right] \tag{4.28}$$

$$\Delta_2 = \frac{1}{g_1}\cdot\left[1+\sum_{n=1}^{\infty}F^n\right] \tag{4.29}$$

After low order filter design, $F$ becomes $F_+$ with filter modeling error $e$ as described in Eq. (4.9) so that we have

$$\begin{aligned}\Sigma_{+,2} &= \frac{1}{g_1}\cdot\frac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2}\cdot\frac{1}{1+F_+} \\ &= \frac{1}{g_1}\cdot\frac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2}\cdot\left[1-(F+e)+(F+e)^2-(F+e)^3+\cdots\right] \\ &= \frac{1}{g_1}\cdot\frac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2}\cdot\left[1+\sum_{n=1}^{\infty}(-1)^n(F+e)^n\right]\end{aligned} \tag{4.30}$$

Using binomial series, we can obtain

$$\Sigma_{+,2} = \frac{1}{g_1}\cdot\frac{1}{1+2\left(\dfrac{g_3}{g_1+g_2}\right)^2}\cdot\left[1+\sum_{n=1}^{\infty}(-F)^n+\sum_{n=1}^{\infty}\sum_{k=1}^{n}(-1)^n\binom{n}{k}e^kF^{n-k}\right] \tag{4.31}$$

Similarly,

$$C_{+,2} = \frac{1}{g_1} \cdot \frac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2} \cdot \frac{1}{1+F_+}$$

$$= \frac{1}{g_1} \cdot \frac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2} \cdot \left[1+\sum_{n=1}^{\infty}(-F)^n + \sum_{n=1}^{\infty}\sum_{k=1}^{n}(-1)^n \binom{n}{k} e^k F^{n-k}\right] \quad (4.32)$$

$$\Delta_{+,2} = \frac{1}{g_1} \cdot \frac{1}{1-F_+}$$

$$= \frac{1}{g_1} \cdot \left[1+\sum_{n=1}^{\infty} F^n + \sum_{n=1}^{\infty}\sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k}\right] \quad (4.33)$$

We then have

$$e_{s,2} = \Sigma_{+,2} - \Sigma_2$$

$$= \frac{1}{g_1} \cdot \frac{1}{1+2\left(\frac{g_3}{g_1+g_2}\right)^2} \cdot \sum_{n=1}^{\infty}\sum_{k=1}^{n}(-1)^n \binom{n}{k} e^k F^{n-k}$$

$$= \frac{1}{g_1} \cdot \frac{1}{1+2\left(\frac{g_3}{g_1+g_2}\right)^2} \cdot \left(-e+2eF+e^2-3eF^2+3e^2F-e^3+\cdots\right) \quad (4.34)$$

Assume $\left(2eF-e\right) \gg \left(e^2-3eF^2+3e^2F-e^3+\cdots\right)$ so that

$$e_{s,2} \cong \frac{1}{g_1} \cdot \frac{1}{1+2\left(\frac{g_3}{g_1+g_2}\right)^2} \cdot \left(2F-1\right)e \quad (4.35)$$

Similarly

$$e_{c,2} = C_{+,2} - C_2$$

$$= \frac{1}{g_1} \cdot \frac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2} \cdot \sum_{n=1}^{\infty}\sum_{k=1}^{n}(-1)^n \binom{n}{k} e^k F^{n-k}$$

$$\cong \frac{1}{g_1} \cdot \frac{(g_1+g_2)g_3}{(g_1+g_2)^2+2g_3^2} \cdot (2F-1)e \tag{4.36}$$

$$\begin{aligned} e_{d,2} &= \Delta_{+,2} - \Delta_2 \\ &= \frac{1}{g_1} \cdot \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k} \\ &\cong \frac{1}{g_1} \cdot (2F+1)e \end{aligned} \tag{4.37}$$

To further simplify this analysis, we will assume $g_1 = 1$, $g_2 = r$ and $g_3 = \alpha$ where $r < \alpha < 1$. Hence, the modeling error vector can be rewritten as

$$\begin{aligned} E_{3,shuffler} &= \begin{bmatrix} \frac{1}{2}(e_{s,2}+e_{d,2}) & e_{c,2} & \frac{1}{2}(e_{s,2}-e_{d,2}) \\ \frac{1}{2}(e_{s,2}-e_{d,2}) & e_{c,2} & \frac{1}{2}(e_{s,2}+e_{d,2}) \end{bmatrix} \begin{bmatrix} 1 \\ \alpha \\ r \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}(e_{s,2}+e_{d,2}) + \alpha e_{c,2} + \frac{1}{2} r (e_{s,2}-e_{d,2}) \\ \frac{1}{2}(e_{s,2}-e_{d,2}) + \alpha e_{c,2} + \frac{1}{2} r (e_{s,2}+e_{d,2}) \end{bmatrix} \end{aligned} \tag{4.38}$$

From the first element of Eq. (4.37), that is, the equalized error:

$$\|eq.\| \cong \left\| \frac{1}{2}(1+r)\frac{1}{1+2\left(\frac{\alpha}{1+r}\right)^2}(2r-1)e + \alpha \cdot \frac{1}{1+2\left(\frac{\alpha}{1+r}\right)^2} \frac{\alpha(2r-1)e}{1+r} + \frac{1}{2}(1-r)(1+2r)e \right\| \tag{4.39}$$

Assume $2r-1>0$ so that the coefficients of common modeling error $e$ are positive.

Thus, the norm of equalized error can be rewritten as

$$\begin{aligned} \|eq.\| &< \left\| \frac{1}{2}(1+r)(2r-1)e + \alpha \cdot \frac{\alpha(2r-1)e}{1+r} + \frac{1}{2}(1-r)(1+2r)e \right\| \\ &= \left\| e\left[ \frac{1}{2}(1+r)(2r-1) + \frac{1}{2}(1-r)(1+2r) + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right] \right\| \end{aligned}$$

$$= \left\| e \left[ r + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right] \right\|$$

$$= \|e\| \cdot \left| r + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right| \tag{4.40}$$

If $r \gg \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i$, thus

$$\|eq.\| < \|e\| \cdot |r| \tag{4.41}$$

Also, the second element of Eq. (4.37), that is, the crosstalk error:

$$\|cr.\| \cong \left\| \frac{1}{2}(1+r) \frac{1}{1+2\left(\frac{\alpha}{1+r}\right)^2} (2r-1)e + \alpha \cdot \frac{1}{1+2\left(\frac{\alpha}{1+r}\right)^2} \frac{\alpha(2r-1)}{1+r} e - \frac{1}{2}(1-r)(1+2r)e \right\| \tag{4.42}$$

Assume $2r-1>0$ and $r$ closes to $\alpha$. Thus, the norm of crosstalk error can be rewritten as

$$\|cr.\| \cong \left\| \frac{1}{2}(1+r)(2r-1)e + \alpha \cdot \frac{\alpha(2r-1)e}{1+r} - \frac{1}{2}(1-r)(1+2r)e \right\|$$

$$= \left\| e \left[ \frac{1}{2}(1+r)(2r-1) - \frac{1}{2}(1-r)(1+2r) + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right] \right\|$$

$$= \left\| e \left[ (2r^2-1) + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right] \right\|$$

$$= \|e\| \cdot \left| (2r^2-1) + \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right| \tag{4.43}$$

If $|2r^2-1| \gg \left| \alpha^2 (2r-1) \sum_{i=0}^{\infty} (-r)^i \right|$, then

$$\|cr.\| \cong \|e\| \cdot |2r^2-1| \tag{4.44}$$

The result in Eq. (4.40) indicates that the norm of equalized error is smaller than that of Eq. (4.17). It means that three loudspeakers shuffler form is not sensitive to equalized error. Eq. (4.43) denotes that the norm of crosstalk error is similar to that of Eq. (4.17) which means three loudspeakers shuffler has the same immunity as two loudspeakers shuffler to crosstalk error. Thus, we can see that three loudspeakers shuffler form has better immunity of filter modeling than two loudspeakers shuffler form. But, the filters of three loudspeakers shuffler form is more complex.

### 4.1.3 Analysis on Simplified Shuffler Form

In this section, we will take the perturbation analysis on filter modeling of the simplified shuffler form. From Figure 3.6, we can derive the relation between $Y$ and $X$ as shown below

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \Delta_3 & -\Delta_3 \\ H_3 & H_3 \\ -\Delta_3 & \Delta_3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{4.45}$$

After modeling, the filters $H_3$ and $\Delta_3$ will be changed to be $H_{+,3}$ and $\Delta_{+,3}$ with modeling error $e_{H,3}$ and $e_{d,3}$, respectively, so that the relation between $Y_+$ and $X$ can be written as follows:

$$\begin{bmatrix} Y_{+,1} \\ Y_{+,2} \\ Y_{+,3} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \Delta_{+,3} & -\Delta_{+,3} \\ H_{+,3} & H_{+,3} \\ -\Delta_{+,3} & \Delta_{+,3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \Delta_3 + e_{d,3} & -\Delta_3 + e_{d,3} \\ H_3 + e_{H,3} & H_3 + e_{H,3} \\ -\Delta_3 + e_{d,3} & \Delta_3 + e_{d,3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{4.46}$$

It should be mentioned that the filter $H_3$ includes the information of $g_3$ only and the filter $\Delta_3$ does not include $g_3$ as depicted in Eq. (3.45) and Eq. (3.46). Therefore, the modeling error $e_{H,3}$ and $e_{d,3}$ are uncorrelated. Then the modeling error after crosstalk cancellation can be expressed as follows:

$$E_{3,simplified} = \frac{1}{2} \begin{bmatrix} e_{d,3} & e_{H,3} & -e_{d,3} \\ -e_{d,3} & e_{H,3} & e_{d,3} \end{bmatrix} \begin{bmatrix} g_1 \\ g_3 \\ g_2 \end{bmatrix} \tag{4.47}$$

Now, we begin the analysis. Since the filter $H_3$ without the information of $g_1$ and $g_2$ so that we factorize filter $\Delta_3$ as follows

$$\Delta_3 = \frac{1}{g_1 - g_2} = \frac{1}{g_1} \cdot \frac{1}{1 - F} \tag{4.48}$$

where $F$ is described in section 4.1.1. Thus, we can obtain the structure as shown in Figure 4.4. By utilizing Taylor series, we can obtain

$$\Delta_3 = \frac{1}{g_1} \cdot \left[ 1 + \sum_{k=1}^{\infty} F^k \right] \tag{4.49}$$

After low order filter design, $F$ becomes $F_+$ with filter modeling error $e$ as described in Eq. (4.9) so that we have

$$\begin{aligned} \Delta_{+,3} &= \frac{1}{g_1} \cdot \frac{1}{1 - F_+} \\ &= \frac{1}{g_1} \cdot \left[ 1 + (F + e) + (F + e)^2 + (F + e)^3 + \cdots \right] \\ &= \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} (F + e)^n \right] \end{aligned} \tag{4.50}$$

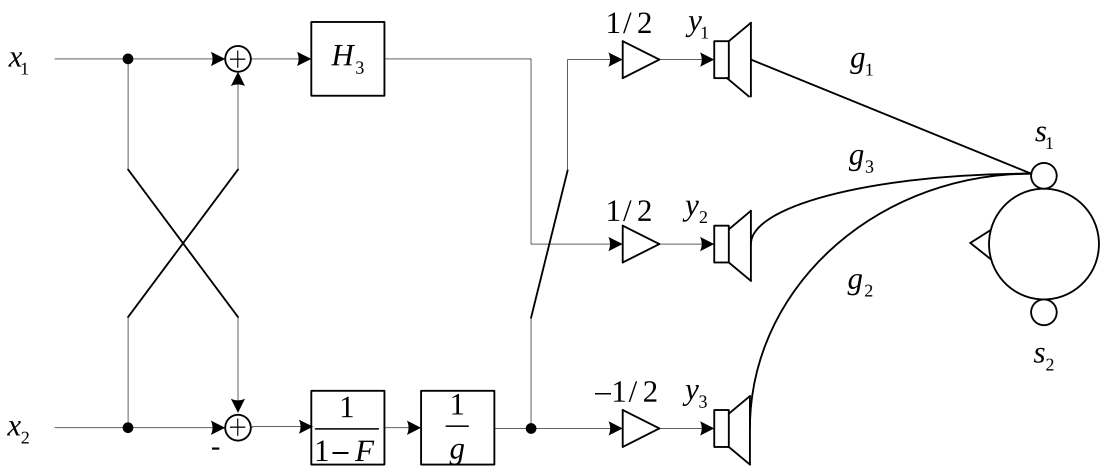


Figure 4.4 The factorized three loudspeakers simplified shuffler form.

As before, using binomial series so that

$$\Delta_{+,3} = \frac{1}{g_1} \cdot \left[ 1 + \sum_{n=1}^{\infty} F^n + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k} \right] \tag{4.51}$$

We than have

$$\begin{aligned} e_{d,3} &= \Delta_{+,3} - \Delta_3 \\ &= \frac{1}{g_1} \cdot \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} e^k F^{n-k} \\ &= \frac{1}{g_1} \left( e + 2eF + e^2 + 3eF^2 + 3e^2 F + e^3 + \cdots \right) \\ &\cong \frac{1}{g_1} \cdot (2F + 1) e \end{aligned} \tag{4.52}$$

In addition, the modeling error of $H_3$ is

$$e_{H,3} = H_{+,3} - H_3 \tag{4.53}$$

and the modeling error vector can be written as follows

$$E_{3,simplified} \cong \frac{1}{2}\begin{bmatrix} \frac{1}{g_1}\cdot(2F+1)e & e_{H,3} & -\frac{1}{g_1}\cdot(2F+1)e \\ -\frac{1}{g_1}\cdot(2F+1)e & e_{H,3} & \frac{1}{g_1}\cdot(2F+1)e \end{bmatrix}\begin{bmatrix} g_1 \\ g_3 \\ g_2 \end{bmatrix} \tag{4.54}$$

To further simplify this analysis, we will assume $g_1 = 1$, $g_2 = r$ and $g_3 = \alpha$ where $r < \alpha < 1$. Hence, the error vector can be rewritten as

$$E_{3,simplified} \cong \frac{1}{2}\begin{bmatrix} (2r+1)e & e_{H,3} & -(2r+1)e \\ -(2r+1)e & e_{H,3} & (2r+1)e \end{bmatrix}\begin{bmatrix} 1 \\ \alpha \\ r \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} (1-r)(2r+1)e + \alpha e_{H,3} \\ -(1-r)(2r+1)e + \alpha e_{H,3} \end{bmatrix} \tag{4.55}$$

From the above result, it is difficult to make theoretical comparison between Eq. (4.55) and Eq. (4.18) or Eq. (4.38). Since the modeling error $e_{H,3}$ is uncorrelated with $e$ in Eq. (4.55). In chapter 5, we will discuss which structure is more robust for filter modeling error from simulated results.

## 4.2 Perturbation Analysis for Head movement

In this section, we will investigate the robustness of crosstalk canceller due to head movement. In addition, we are interested in which loudspeaker positions can result in a robust crosstalk canceller. Head movement, changes the default acoustic matrix $\mathbf{G}$ to a perturbed acoustic matrix $\mathbf{G}_+$ as shown in Figure 4.5. Thus, the product of $\mathbf{C}$ and $\mathbf{G}_+$ is no longer an identity matrix. The perturbation analysis of head movement will be shown in following sections.

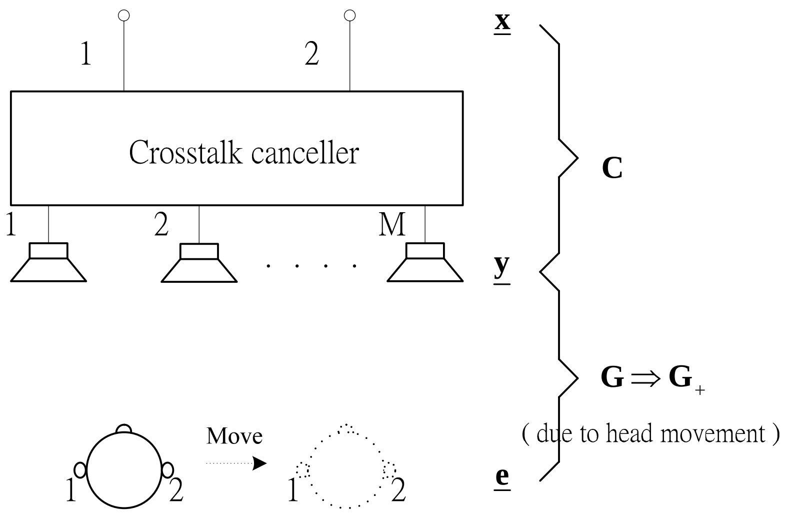


Figure 4.5 The geometry of crosstalk cancellation system for multiple loudspeakers due to head movement.



First, we consider the shuffler form of two loudspeakers arrangement as shown in Figure 3.2. The crosstalk canceller of this structure is shown in Eq. (3.3). Ideally, $\mathbf{GC}=\mathbf{I}$. If the head moves, $\mathbf{G}$ becomes $\mathbf{G}_+$ so that $\mathbf{G}_+\mathbf{C}\neq\mathbf{I}$. It is because the error exists between $\mathbf{G}$ and $\mathbf{G}_+$. By defining the acoustic channel error $\delta\mathbf{G}_2$ as

$$\delta\mathbf{G}_2=\begin{bmatrix}\delta g_{11} & \delta g_{12}\\ \delta g_{21} & \delta g_{22}\end{bmatrix} \tag{4.56}$$

The acoustic channel error after crosstalk cancellation can be written as

$$\mathbf{E}_{2,Shuffler}=\delta\mathbf{G}_2\mathbf{C}_{2,Shuffler}$$

$$=\frac{1}{2}\begin{bmatrix}\Sigma(\delta g_{11}+\delta g_{12})+\Delta(\delta g_{11}-\delta g_{12}) & \Sigma(\delta g_{11}+\delta g_{12})-\Delta(\delta g_{11}-\delta g_{12})\\ \Sigma(\delta g_{21}+\delta g_{22})+\Delta(\delta g_{21}-\delta g_{22}) & \Sigma(\delta g_{21}+\delta g_{22})-\Delta(\delta g_{21}-\delta g_{22})\end{bmatrix} \tag{4.57}$$

The diagonal elements of this error matrix denote the equalization error and the off-diagonal elements represent the crosstalk error.

Now, we consider the simplified shuffler form of three loudspeakers arrangement as shown in Figure 3.6. The crosstalk canceller of this structure is shown in Eq. (3.46). If the acoustic channel error $\delta\mathbf{G}_3$ is

$$\delta\mathbf{G}_3 = \begin{bmatrix} \delta g_{11} & \delta g_{12} & \delta g_{13} \\ \delta g_{21} & \delta g_{22} & \delta g_{23} \end{bmatrix} \tag{4.58}$$

then the acoustic channel error after crosstalk cancellation can be written as

$$\begin{aligned} \mathbf{E}_{3,simplified} &= \delta\mathbf{G}_3\mathbf{C}_{3,simplified} \\ &= \frac{1}{2}\begin{bmatrix} H_3\delta g_{12} + \Delta_3\left(\delta g_{11} - \delta g_{13}\right) & H_3\delta g_{12} - \Delta_3\left(\delta g_{11} - \delta g_{13}\right) \\ H_3\delta g_{22} + \Delta_3\left(\delta g_{21} - \delta g_{23}\right) & H_3\delta g_{22} - \Delta_3\left(\delta g_{21} - \delta g_{23}\right) \end{bmatrix} \end{aligned} \tag{4.59}$$

Next, we consider the simplified shuffler form of four loudspeakers arrangement as shown in Figure 3.8.. The crosstalk canceller of this structure is shown in Eq. (3.62). Denote the acoustic channel error $\delta\mathbf{G}_4$ as

$$\delta\mathbf{G}_4 = \begin{bmatrix} \delta g_{11} & \delta g_{12} & \delta g_{13} & \delta g_{14} \\ \delta g_{21} & \delta g_{22} & \delta g_{23} & \delta g_{24} \end{bmatrix} \tag{4.60}$$

then the acoustic channel error after crosstalk cancellation can be written as

$$\mathbf{E}_{4,simplified} = \delta\mathbf{G}_4\mathbf{C}_{4,simplified}$$

$$= \frac{1}{2}\begin{bmatrix} \Sigma_4\left(\delta g_{12} + \delta g_{13}\right) + \Delta_4\left(\delta g_{11} - \delta g_{14}\right) & \Sigma_4\left(\delta g_{12} + \delta g_{13}\right) - \Delta_4\left(\delta g_{11} - \delta g_{14}\right) \\ \Sigma_4\left(\delta g_{22} + \delta g_{23}\right) + \Delta_4\left(\delta g_{21} - \delta g_{24}\right) & \Sigma_4\left(\delta g_{22} + \delta g_{23}\right) - \Delta_4\left(\delta g_{21} - \delta g_{24}\right) \end{bmatrix} \tag{4.61}$$

It should be mentioned that the above analysis is not complete yet. Since we can't find a common channel error exists so that there is no relationship between these acoustic channel errors. Therefore, it is difficult to compare which structure is more robust for channel error from Eq. (4.57), Eq. (4.59) and Eq. (4.61). Therefore, we will discuss how the changed channel affects the performance of crosstalk cancellation by calculating the condition number of **G**.

## 4.3 Optimum Loudspeakers Position for Robust Crosstalk Canceller

As shown in Figure 4.5, $\mathbf{G}_{+}$ indicates the perturbed acoustic transmission path which can be represented as

$$\mathbf{G}_{+} = \mathbf{G} + \delta\mathbf{G} \tag{4.62}$$

where $\delta\mathbf{G}$ denotes the channel error, thus

$$\mathbf{G}_{+}\mathbf{C} = \left(\mathbf{G} + \delta\mathbf{G}\right)\mathbf{C} \neq \mathbf{I}_{2\times 2} \tag{4.63}$$

If **C** can be adjusted by $\delta\mathbf{C}$ so that

$$\left(\mathbf{G} + \delta\mathbf{G}\right)\left(\mathbf{C} + \delta\mathbf{C}\right) = \mathbf{I}_{2\times 2} \tag{4.64}$$

From linear systems theory in [26], we can obtain that

$$\frac{\left\|\delta\mathbf{C}\right\|}{\left\|\mathbf{C}\right\|} \leq \left\|\mathbf{G}^{-1}\right\| \cdot \left\|\delta\mathbf{G}\right\|$$

$$= \left\| \mathbf{G}^{-1} \right\| \left\| \mathbf{G} \right\| \cdot \frac{\left\| \delta \mathbf{G} \right\|}{\left\| \mathbf{G} \right\|}$$

$$= cond\left\{ \mathbf{G} \right\} \cdot \frac{\left\| \delta \mathbf{G} \right\|}{\left\| \mathbf{G} \right\|} \tag{4.65}$$

Eq. (4.65) shows that the adjustment of **C** can be bounded by the condition number of **G**. The larger condition number of **G** indicates that crosstalk canceller **C** is sensitive to the channel error $\delta\mathbf{G}$. Thus, we can call such **C** is nonrobust for channel error. Conversely, the smaller condition number of **G** represents the crosstalk canceller **C** has better immunity to channel error so that we call such **C** is robust for channel error. Here, the condition number of **G** is defined as

$$cond\left\{ \mathbf{G} \right\} = \frac{\sigma_{\max}\left( \sqrt{\mathbf{G}\mathbf{G}^H} \right)}{\sigma_{\min}\left( \sqrt{\mathbf{G}\mathbf{G}^H} \right)} \tag{4.66}$$

where $\sigma_{\min}(\cdot)$ and $\sigma_{\max}(\cdot)$ represent the smallest and largest singular values, respectively. When $\mathbf{G}$ is ill-conditioned, the crosstalk canceller designed according to $\mathbf{C} = \mathbf{G}^{-1}$ or $\mathbf{C} = \mathbf{G}^{+}$ is inherently nonrobust. Thus, Eq. (4.66) serves as a useful robustness measure for the crosstalk canceller. In following section, we will use Eq. (4.66) to calculate the condition of multiple loudspeakers arrangement.

### 4.3.1 Two Loudspeakers Arrangement

In the analysis in [3, 4, 5], the author let the acoustic transmission functions between loudspeakers and the listener's ears be

$$\mathbf{G_2} = \begin{bmatrix} g_1 & g_2 \\ g_2 & g_1 \end{bmatrix}$$

$$= \begin{bmatrix} e^{j2\pi\lambda^{-1}d_1} & e^{j2\pi\lambda^{-1}d_2} \\ e^{j2\pi\lambda^{-1}d_2} & e^{j2\pi\lambda^{-1}d_1} \end{bmatrix} \tag{4.67}$$

where $\lambda$ is the wavelength, and $d_i$ is the distance from the *i*th loudspeaker to the ear. Then the condition number of matrix $\mathbf{G_2}$ as shown below

$$cond\{\mathbf{G}_2\} = \frac{\sigma_{\max}\left(\sqrt{\mathbf{G}_2\mathbf{G}_2^H}\right)}{\sigma_{\min}\left(\sqrt{\mathbf{G}_2\mathbf{G}_2^H}\right)}$$

$$= \frac{\sqrt{1+\sqrt{\cos 2\pi\lambda^{-1}\beta}}}{\sqrt{1-\sqrt{\cos 2\pi\lambda^{-1}\beta}}} \tag{4.68}$$

where $\beta = d_1 - d_2$ indicates the interaural path difference which can be used to determine the angle of the loudspeakers. In [3], the interaural difference is given by

$$\beta = 2r_H \sin\theta \tag{4.69}$$

where $r_H$ is the radius of the head. Usually, $r_H$ is 0.0875 m. Since our goal is to find the optimum $\beta$ so that $cond\{\mathbf{G}_2\}$ approaches 1. It means that under such loudspeaker arrangement the crosstalk canceller is more robust for perturbation of head movement. The result in [3, 4, 5] as follows

$$cond\{\mathbf{G}_2\} = \begin{cases} 1, & robust \\ \infty, & nonrobust \end{cases} \tag{4.70}$$

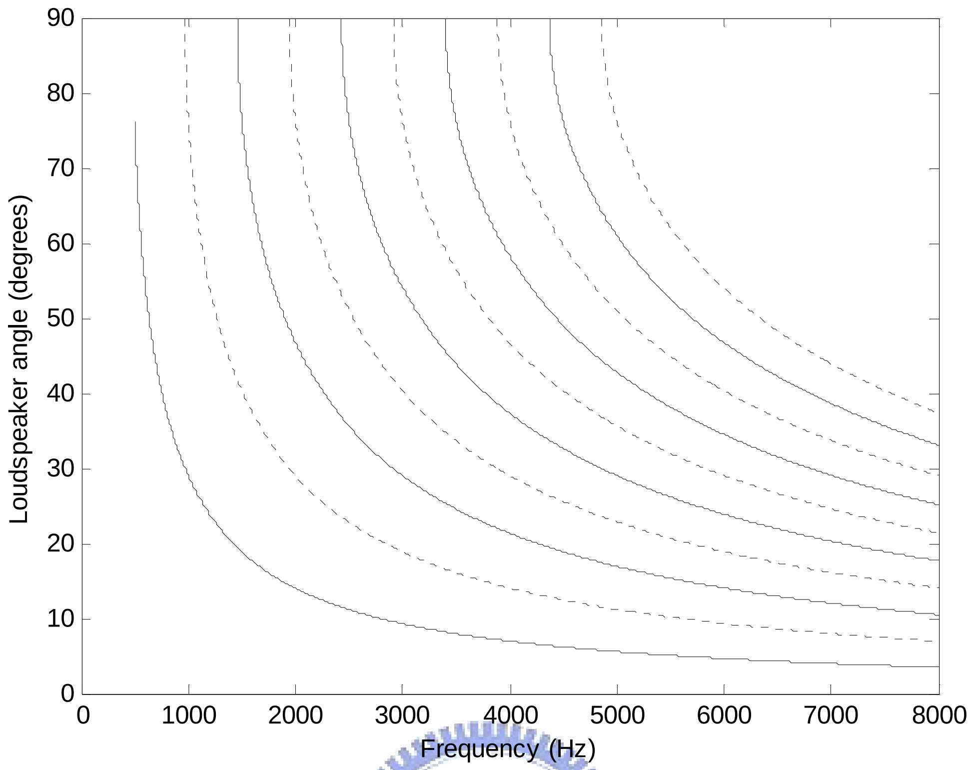


Figure 4.6 The condition numbers of $\mathbf{G}_2$ with different loudspeakers position. The solid line indicates the loudspeakers position of the side loudspeakers for condition number equals to 1 and the dotted line indicates condition numbers close to infinitity.

so that

$$\begin{cases} \arg\min\limits_{\beta} cond\left\{\mathbf{G}_2\right\} = \dfrac{\lambda}{4} + k\dfrac{\lambda}{2}, & robust \\ \arg\max\limits_{\beta} cond\left\{\mathbf{G}_2\right\} = i\dfrac{\lambda}{2}, & nonrobust \end{cases} \tag{4.71}$$

where $i, k \in \mathbb{Z}$, then

$$\begin{cases} \sin\theta = \dfrac{\beta}{2r_H} = \dfrac{\lambda}{0.7} + k\dfrac{\lambda}{0.35}, & robust \\ \sin\theta = \dfrac{\beta}{2r_H} = i\dfrac{\lambda}{0.35}, & nonrobust \end{cases} \tag{4.72}$$

The loudspeaker positions obtained from Eq. (4.72) are shown in Figure 4.6. The solid curve indicates the condition number is 1 and results in a robust crosstalk canceller while loudspeakers pair located at these positions for different frequencies. The dotted line should be avoided since they produce a nonrobust crosstalk canceller. Ideally, the loudspeaker spacing should vary with frequency so that a more robust crosstalk canceller is possible. In following sections, we can take the similar analysis to derive the optimum loudspeakers position for multi-loudspeakers arrangement.

## 4.3.2 Three Loudspeakers Arrangement

Different from two loudspeakers arrangement, there is an additional loudspeaker located at $0°$ between the side loudspeakers in the case of three loudspeakers. Also, we can let the acoustic transmission functions between loudspeakers and the listener's ears as shown in Eq. (4.73) to simplify analysis.

$$\mathbf{G}_3 = \begin{bmatrix} g_1 & g_3 & g_2 \\ g_2 & g_3 & g_1 \end{bmatrix}$$

$$= \begin{bmatrix} e^{j2\pi\lambda^{-1}d_1} & e^{j2\pi\lambda^{-1}d_3} & e^{j2\pi\lambda^{-1}d_2} \\ e^{j2\pi\lambda^{-1}d_2} & e^{j2\pi\lambda^{-1}d_3} & e^{j2\pi\lambda^{-1}d_1} \end{bmatrix} \qquad (4.73)$$

We then can derive the condition number of matrix $\mathbf{G}_3$ as shown below

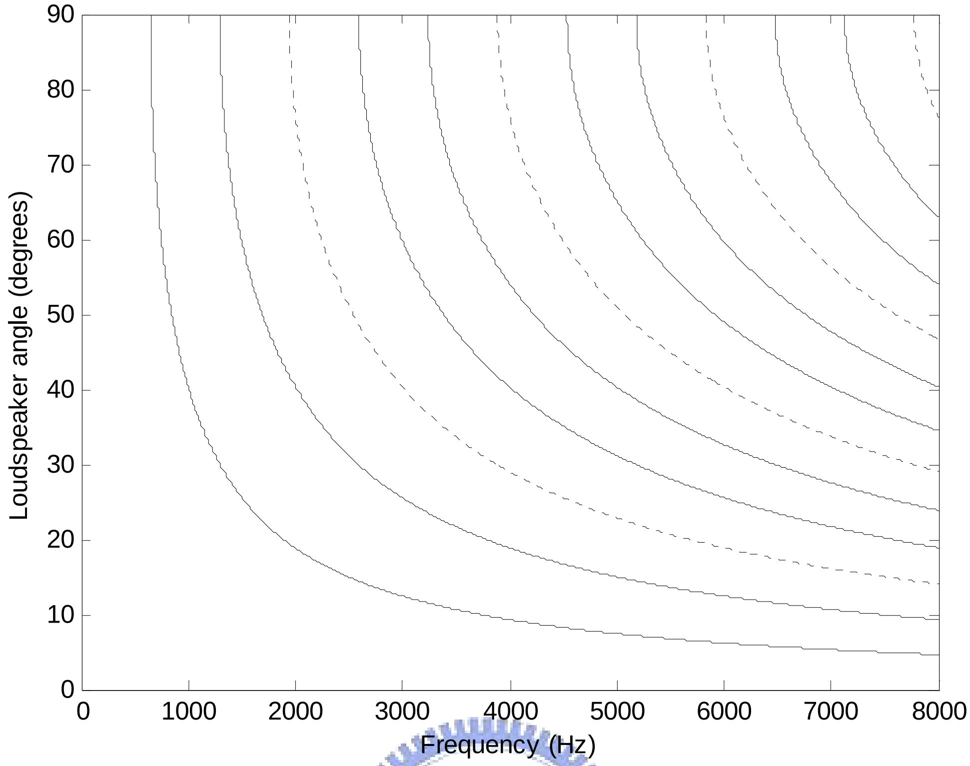


Figure 4.7 The condition numbers of $\mathbf{G}_3$ with different loudspeakers position. The solid line indicates the loudspeakers position of the side loudspeakers for condition number equals to 1 and the dotted line indicates condition numbers close to infinitity.

$$cond\left\{\mathbf{G}_3\right\} = \frac{\sqrt{\sqrt{3}+\sqrt{1+2\cos 2\pi\lambda^{-1}\beta}}}{\sqrt{\sqrt{3}-\sqrt{1+2\cos 2\pi\lambda^{-1}\beta}}} \tag{4.74}$$

so that we can obtain the result as shown below

$$\begin{cases} \arg\min_{\beta} cond\left\{\mathbf{G}_3\right\} = \dfrac{\lambda}{3}+k\lambda, & robust \\ \arg\max_{\beta} cond\left\{\mathbf{G}_3\right\} = i\lambda, & nonrobust \end{cases} \tag{4.75}$$

where $i,k \in \mathbb{Z}$, then

$$\begin{cases} \sin\theta = \dfrac{\beta}{2r_H} = \dfrac{\lambda}{0.525} + k\dfrac{\lambda}{0.175}, & robust \\ \sin\theta = \dfrac{\beta}{2r_H} = i\dfrac{\lambda}{0.175}, & nonrobust \end{cases} \tag{4.76}$$

The loudspeaker positions obtained from Eq. (4.76) are shown in Figure 4.7. The solid curve indicates the optimum positions and the dotted line should be avoided since they produce a nonrobust crosstalk canceller.

### 4.3.3 Four Loudspeakers Arrangement

In this section, we will discuss the robustness of four loudspeakers arrangement due to head movement. In this case, there are additional closer loudspeakers pair between wider loudspeakers pair. Thus, there are two loudspeaker positions should be considered. To simplify the analysis, we let the acoustic transmission functions between loudspeakers and the listener's ears be

$$\mathbf{G}_4 = \begin{bmatrix} g_1 & g_2 & g_3 & g_4 \\ g_4 & g_3 & g_2 & g_1 \end{bmatrix}$$

$$= \begin{bmatrix} e^{j2\pi\lambda^{-1}d_1} & e^{j2\pi\lambda^{-1}d_2} & e^{j2\pi\lambda^{-1}d_3} & e^{j2\pi\lambda^{-1}d_4} \\ e^{j2\pi\lambda^{-1}d_4} & e^{j2\pi\lambda^{-1}d_3} & e^{j2\pi\lambda^{-1}d_2} & e^{j2\pi\lambda^{-1}d_1} \end{bmatrix} \tag{4.77}$$

We then can derive the condition number of matrix $\mathbf{G}_4$ as shown below

$$cond\{\mathbf{G}_4\} = \frac{\sqrt{2+\sqrt{2\cos 2\pi\lambda^{-1}\beta_1 + 2\cos 2\pi\lambda^{-1}\beta_2}}}{\sqrt{2-\sqrt{2\cos 2\pi\lambda^{-1}\beta_1 + 2\cos 2\pi\lambda^{-1}\beta_2}}} \tag{4.78}$$

where $\beta_1 = d_1 - d_4$ and $\beta_2 = d_2 - d_3$ indicates the interaural path difference of wider loudspeakers pair and closer loudspeakers pair, respectively. To be robust, the $\mathbf{G}_4$ is well conditioned and we can obtain that

$$\arg\min_{\beta_1} cond\left\{\mathbf{G}_4\right\} = \beta_2 + \frac{k}{2}\lambda \tag{4.79}$$

so that

$$\sin\theta_1 = \sin\theta_2 + \left(2k+1\right)\frac{\lambda}{0.35} \tag{4.80}$$

As to nonrobustness, we can obtain that

$$\begin{cases} \arg\max_{\beta_1} cond\left\{\mathbf{G_4}\right\} = (i+1)\lambda \\ \arg\max_{\beta_2} cond\left\{\mathbf{G_4}\right\} = i\lambda \end{cases} \tag{4.81}$$

and

$$\begin{cases} \sin\theta_1 = \dfrac{\lambda}{0.175}(i+1) \\ \sin\theta_2 = \dfrac{\lambda}{0.175}i \end{cases} \tag{4.82}$$

From the results in Eq. (4.80) and Eq. (4.82), angle between two loudspeakers should be considered. As shown in Eq. (4.80), the position of wider loudspeaker can be obtained while the position of closer loudspeaker is determined for fixed frequency. For $f = 2KHz$, we let the closer loudspeaker locate at $\pm 5°$ then we can obtain the position of wider loudspeaker is located at $\pm 35°$ and results in a robust crosstalk canceller.

Here, we defined the parameter, that is, the *robust bandwidth* to indicate that the condition number is below a specified value in the lowest continuous frequency range. Figure 4.8 has shown that the robust bandwidth of different numbers of loudspeakers while the condition number is under a specified value, say 3. The robust bandwidth of four loudspeakers is wider than three and two loudspeakers in this figure. It means that four loudspeakers arrangement can result in a more robust crosstalk canceller than less loudspeakers arrangement. With the result as described above, we make a conclusion about crosstalk canceller of multi-loudspeakers arrangement.

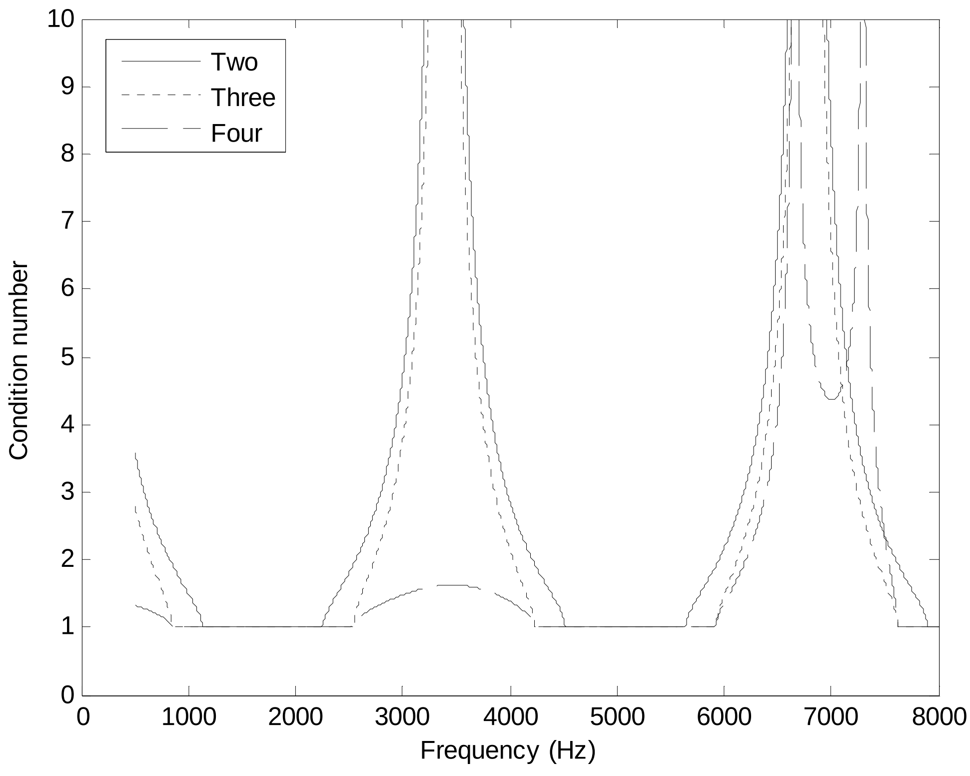


Figure 4.8 The robust bandwidth of different number of loudspeakers with the side loudspeakers located at $\pm 30°$.

## 4.4 Conclusions

From the description in section 4.1, with the common filter modeling error *e*, the three loudspeakers shuffler form has better immunity of the perturbation than the two loudspeakers shuffler form. From analysis on head movement, the result shows that three loudspeakers arrangement for fixed position of the side loudspeakers has wider robust bandwidth than two loudspeakers arrangement. In addition, four loudspeakers can result in a more robust crosstalk canceller than fewer loudspeakers arrangement. In chapter 5, the simulation results will demonstrate that the multiple loudspeakers setup is more robust.

# Chapter 5

# Computer Simulations

In this chapter, we will simulate different crosstalk cancellers to see their performance. Numerous simulations were conducted to compare the performances of different implementations of the crosstalk canceller. Our simulations will focus on delivering the binaural signals via three loudspeakers as shown in Figure 2.3. with $N = 2$, $M = 3$ and $L = 2$. An impulse signal is used as input of $X_1$ and a zero signal as input of $X_2$. The listener will receive no crosstalk and distortionless desired signal if perfect crosstalk cancellation is achieved.

## 5.1 Assumptions

In our simulation, we have the following assumptions:

1. The structure of the crosstalk canceller is based on the symmetric loudspeaker

arrangement so that the system performance is shown for one side only.

2. We use the HRTFs from MIT Media Lab as the acoustic channel data of the simulation.

In order to evaluate the performance of crosstalk cancellers, we will define some parameters. First, as described in Chapter 2, Eq. (2.7) has shown

$$\begin{aligned}\mathbf{GC} &= \mathbf{D} \\ &\cong \begin{bmatrix} z^{-d} & 0 \\ 0 & z^{-d} \end{bmatrix}\end{aligned} \tag{5.1}$$

where $\mathbf{G}$ is an $L \times M$ matrix, the *acoustic* matrix, $\mathbf{C}$ is an $M \times N$ matrix, the *crosstalk canceller* matrix and $\mathbf{D}$ is an $L \times N$ matrix of *desired* transfer functions. In time domain, the errors between the desired and the actual transfer function can be shown below

$$\underline{\mathbf{e}}_1 = \mathbf{H}\underline{\mathbf{c}}_1 - \begin{bmatrix} \underline{\mathbf{d}} \\ \underline{\mathbf{0}} \end{bmatrix} \tag{5.2}$$

$$\underline{\mathbf{e}}_2 = \mathbf{H}\underline{\mathbf{c}}_2 - \begin{bmatrix} \underline{\mathbf{0}} \\ \underline{\mathbf{d}} \end{bmatrix} \tag{5.3}$$

where $\mathbf{H}$ represents the convolution matrix of the acoustic matrix $\mathbf{G}$, then we define the total error as

$$\varepsilon = \sqrt{\|\underline{\mathbf{e}}_1\|_2^2 + \|\underline{\mathbf{e}}_2\|_2^2} \tag{5.4}$$

which denotes the sum of the distance of two received vector and desired vector. The better performance of crosstalk cancellers, the smaller of $\varepsilon$.

Next, we define another parameter, that is, the crosstalk suppression factor (*CSF*)

$$CSF = 10\log\left\|\frac{h_o}{h_x}\right\|^2\Bigg|_{withC's} - 10\log\left\|\frac{h_o}{h_x}\right\|^2\Bigg|_{withoutC's} \tag{5.5}$$

Now, we illustrate how to define the *CSF*. The product of **GC** as

$$\mathbf{GC} = \begin{bmatrix} h_o & h_x \\ h_x & h_o \end{bmatrix} \tag{5.6}$$

where the diagonal elements of this matrix are equalization part and the off-diagonal are crosstalk part. Therefore, the first term of Eq. (5.5) is the signal power ratio of equalization part and crosstalk part after crosstalk cancellation, denote as $R_c = 10\log\left\|\frac{h_o}{h_x}\right\|^2\Bigg|_{withC's}$, and the second term is the one without crosstalk cancellation.

To justify the performance of the crosstalk canceller, we not only consider the crosstalk suppression, but also the equalization result. Here, we use the equalization factor (*EQ*) to see the equalization performance in frequency domain. First, we define the average deviation $\delta$ of $S_1[n]$ as

$$\delta = \left[\frac{1}{M}\sum_{m=0}^{M-1}\left(20\log\left|S_1(m)\right| - AV\right)^2\right]^{\frac{1}{2}} \tag{5.7}$$

where

$$AV = \frac{1}{M}\sum_{m=0}^{M-1} 20\log\left|S_1(m)\right| \tag{5.8}$$

and $S_1(m)$ is the M-point DFT of $S_1[n]$, then the equalization factor can be defined as the average deviation with crosstalk cancellation subtracted by that without crosstalk cancellation.

$$EQ = \delta - \delta_{withC's} \tag{5.9}$$

In the simulation, the larger the *CSF* and *EQ* be, the better performance crosstalk cancellers achieve.

It should be mentioned that without crosstalk cancellation, in such case the **C** might be an identity matrix for $M = 2$. For $M = 3$ (i.e. three loudspeakers), the problem exists that the degree of freedom of the center loudspeaker can be used. Thus, we can let the matrix **C** be

$$\mathbf{C} = \begin{bmatrix} 1-q & 0 \\ q & q \\ 0 & 1-q \end{bmatrix} \tag{5.10}$$

where $0 < q < 1$. If $q = 0$, the center loudspeaker is turned off. For $q = 1$, the binaural signals are delivered by the center loudspeaker only. In following section, we will use $\varepsilon$, *CSF* and *EQ* to compare the performance of different structures of crosstalk cancellers.

## 5.2 Two Loudspeakers Geometric

The typical crosstalk canceller of two loudspeakers setup is shown in Eq. (3.2). In order to achieve economical realizations, we factorize the crosstalk canceller matrix **C** in Eq. (3.2) and results in shuffler form as described in Eq. (3.3).

From Eq. (3.3), it requires only two filters, that is, $\Sigma(z)$ and $\Delta(z)$. First, we implement these two FIR filters by using least square error method. In our simulation, we will try different number of filter taps to examine the performance of crosstalk canceller.

Figure 5.1 and Figure 5.2 show the results of crosstalk cancellation with loudspeaker pair located at $\pm 30^{\circ}$. In this example, the order of FIR filter is 200 taps with 100 samples modeling delays. Since the maximum delay between $g_1[n]$ and $g_2[n]$ is 40 samples, therefore we choose the extra delay as 140 samples. In this scheme, the crosstalk cancellation performance can reach $CSF = 15.499dB$ and $EQ = 13.046dB$. From Figure 5.1, the equalization is worst at very low and high frequencies and almost 8 kHz. Since the HRTFs have deep notches at very low and high frequencies.

Table 5.1 lists the result that the performance of crosstalk canceller versus different length of FIR filters. Obviously, the $CSF$ and $EQ$ is smaller while the order of FIR filters decreases. In addition, $\varepsilon$ is smaller with increased filter order. Thus, the larger order of FIR filters, the performance of crosstalk canceller can be better.

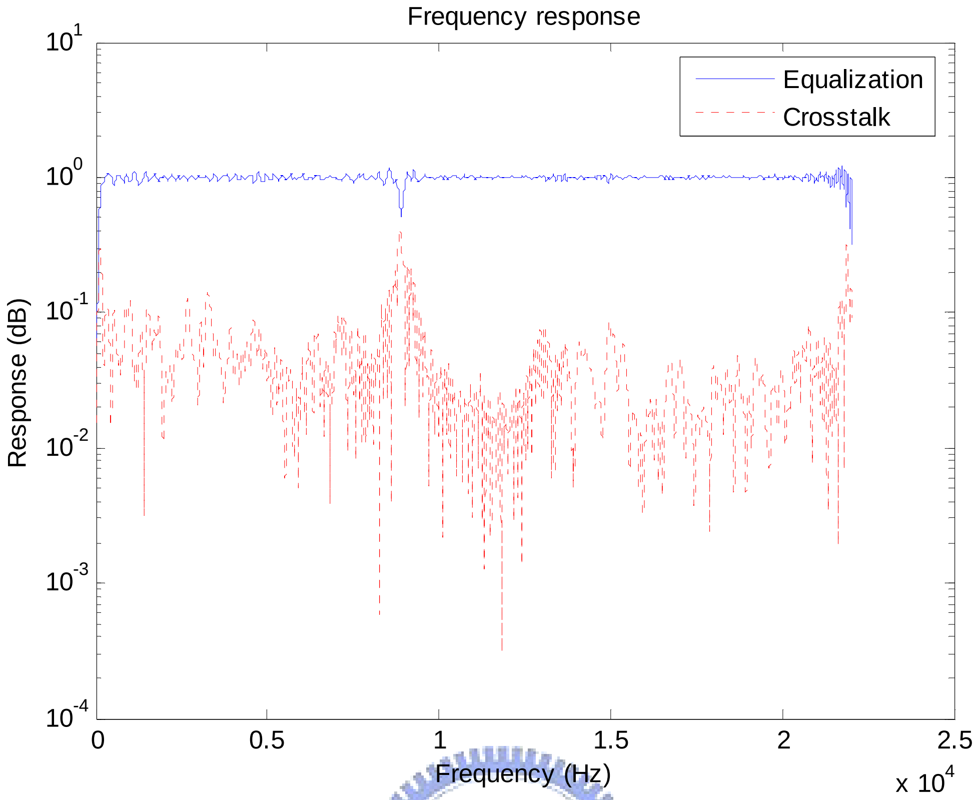


Figure 5.1 The crosstalk cancellation results in frequency domain by using least square error method with loudspeaker located at $\pm 30^{\circ}$.

| **Parameter** / **FIR Filter Order** | $EQ(dB)$ | $CSF\,(dB)$ | $\varepsilon$ |
|---|---|---|---|
| 50 | 9.9652 | 9.6569 | 0.47646 |
| 100 | 12.369 | 13.274 | 0.26971 |
| 200 | 13.046 | 15.499 | 0.17652 |
| 500 | 13.372 | 20.381 | 0.12575 |

Table 5.1 The performance of shuffler form for two loudspeakers setup with different order of FIR filters.

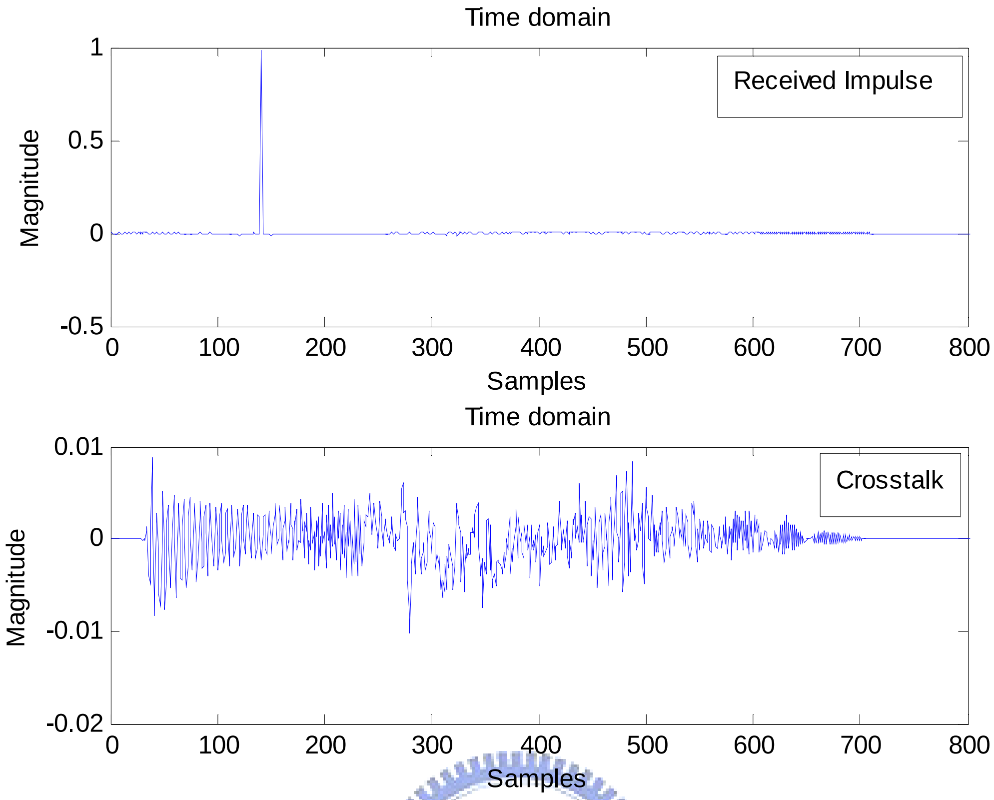


Figure 5.2 The crosstalk cancellation results in time domain by using least square method with loudspeaker located at $\pm 30^{\circ}$.

## 5.3 Three Loudspeakers Geometric

### 5.3.1 Least Square Forward Type

In this section, we will design the crosstalk canceller generated by the least square method. Using the least square method can avoid the stability problem. In our simulation, the side loudspeakers are located at $\pm 30^{\circ}$ and the center one is located at $0^{\circ}$. The FIR filter order is 200 taps with 140 modeling delay and the simulation results are shown in Figure 5.3 and Figure 5.4. In this scheme, the crosstalk cancellation performance can reach $CSF = 22.021dB$ and $EQ = 11.722dB$ as shown in Table 5.3. As described in section 5.1, the degree of freedom of using the third loudspeaker should be considered. We will examine the different $q$ as described in Eq. (5.10) versus $CSF$ and the result is listed in Table 5.2. This table shows that the $CSF$ is larger while only using the center loudspeaker (i.e. $q = 1$) to deliver the binaural signals without crosstalk cancellation. In the following simulation, we will select $q$ as 0.5.

Also, we can using least square method to find the crosstalk canceller of four loudspeakers arrangement. Table 5.4 lists the performance comparison of different number of loudspeakers by using least square method to realize crosstalk canceller.

Next, we summarize the results of simulations for two, three and four loudspeakers arrangements. From Table 5.4, we can obtain that the performance of crosstalk cancellation with the same FIR filter order can be improved and the total error is smaller while the number of loudspeakers is increased. Figure 5.5 and Figure 5.6 has shown the results from Table 5.4.

| $q$ | 0 | 0.5 | 0.8 | 1 |
|---|---|---|---|---|
| $CSF\,(dB)$ | 18.474 | 22.021 | 26.403 | 26.583 |

Table 5.2 The $CSF$ with different degree of freedom of the center loudspeaker.

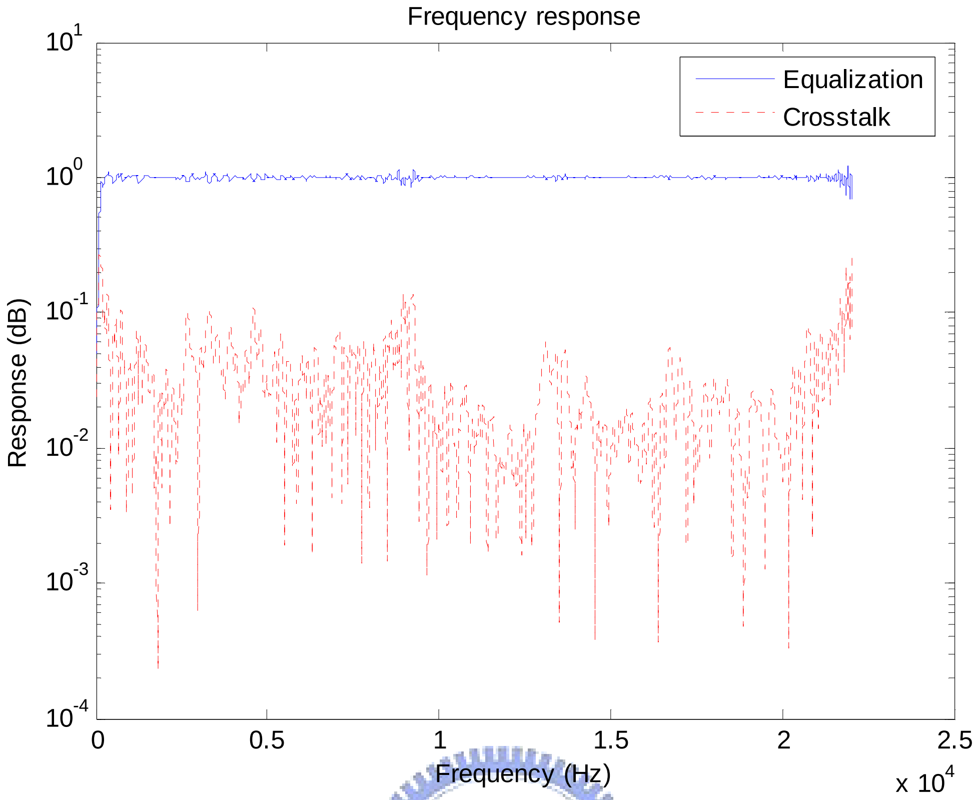


Figure 5.3 The crosstalk cancellation results in frequency domain of least square forward type with the side loudspeakers located at $\pm 30^\circ$ and the center one located at $0^\circ$.

| Parameter / FIR Filter Order | $EQ(dB)$ | $CSF\,(dB)$ | $\varepsilon$ |
|---|---|---|---|
| 50 | 9.9411 | 11.426 | 0.36601 |
| 100 | 10.933 | 18.078 | 0.19728 |
| 200 | 11.722 | 22.021 | 0.13812 |
| 500 | 12.589 | 22.435 | 0.089119 |

Table 5.3 The performance of least forward type crosstalk canceller for three loudspeakers setup with different order of FIR filters.

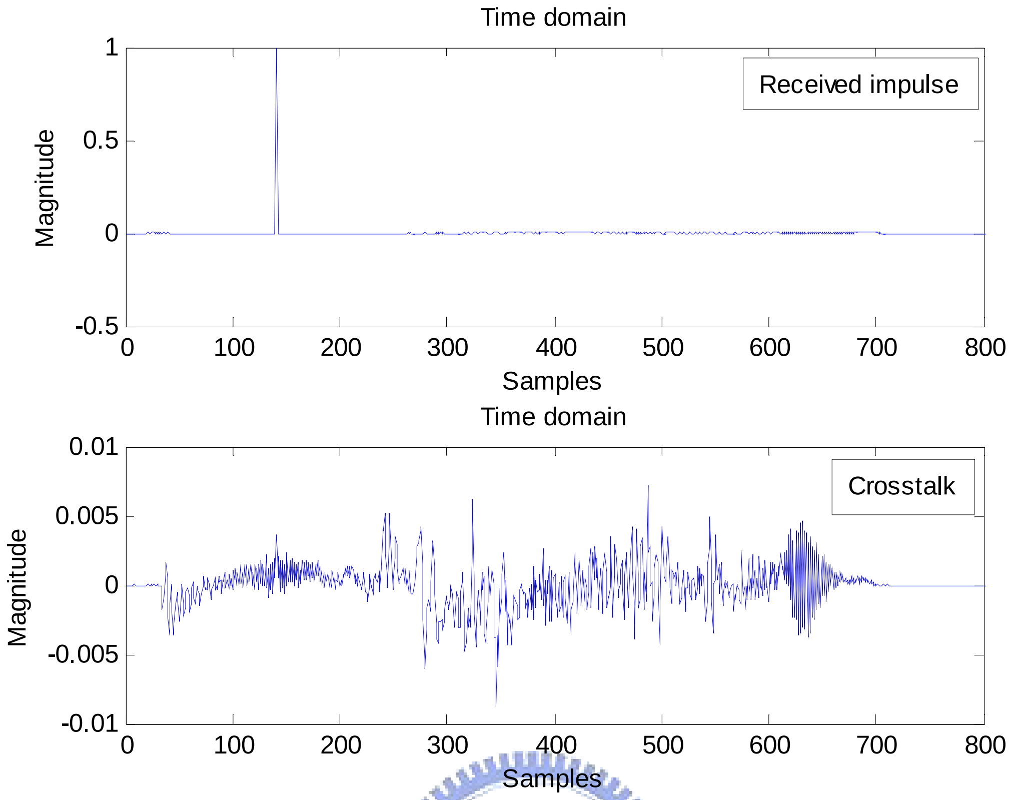


Figure 5.4 The crosstalk cancellation results in time domain of least square forward type with the side loudspeakers located at $\pm 30^{\circ}$ and the center one located at $0^{\circ}$.

| Parameter / FIR Filter order | Two | | Three | | Four | |
|---|---|---|---|---|---|---|
| | $R_c$ (dB) | $\varepsilon$ | $R_c$ (dB) | $\varepsilon$ | $R_c$ (dB) | $\varepsilon$ |
| 50 | 16.1686 | 0.47646 | 18.9882 | 0.36601 | 21.3454 | 0.25154 |
| 100 | 20.3828 | 0.26971 | 22.6403 | 0.19728 | 25.2167 | 0.14184 |
| 200 | 22.6076 | 0.17652 | 26.5829 | 0.13812 | 27.5995 | 0.093848 |
| 500 | 27.4901 | 0.12575 | 26.9977 | 0.089119 | 42.187 | 0.01562 |

Table 5.4 The performance comparison of crosstalk cancellation by using least square method for different numbers of loudspeakers arrangement.

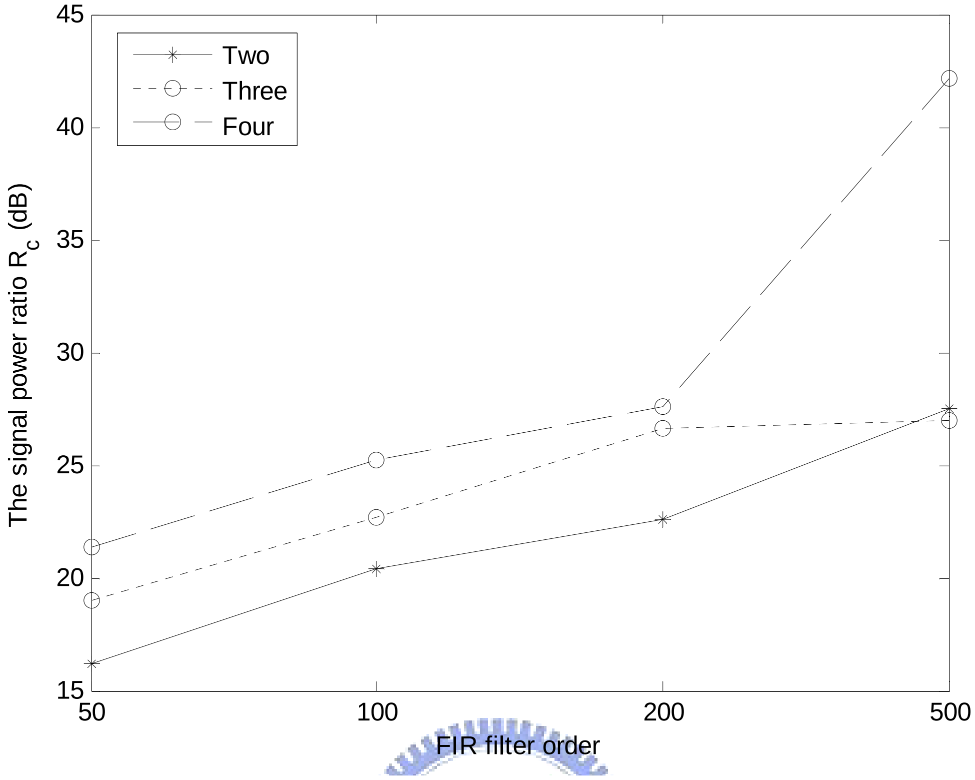


Figure 5.5 The signal power ratio $R_c$ of different numbers of loudspeakers with different filter orders.

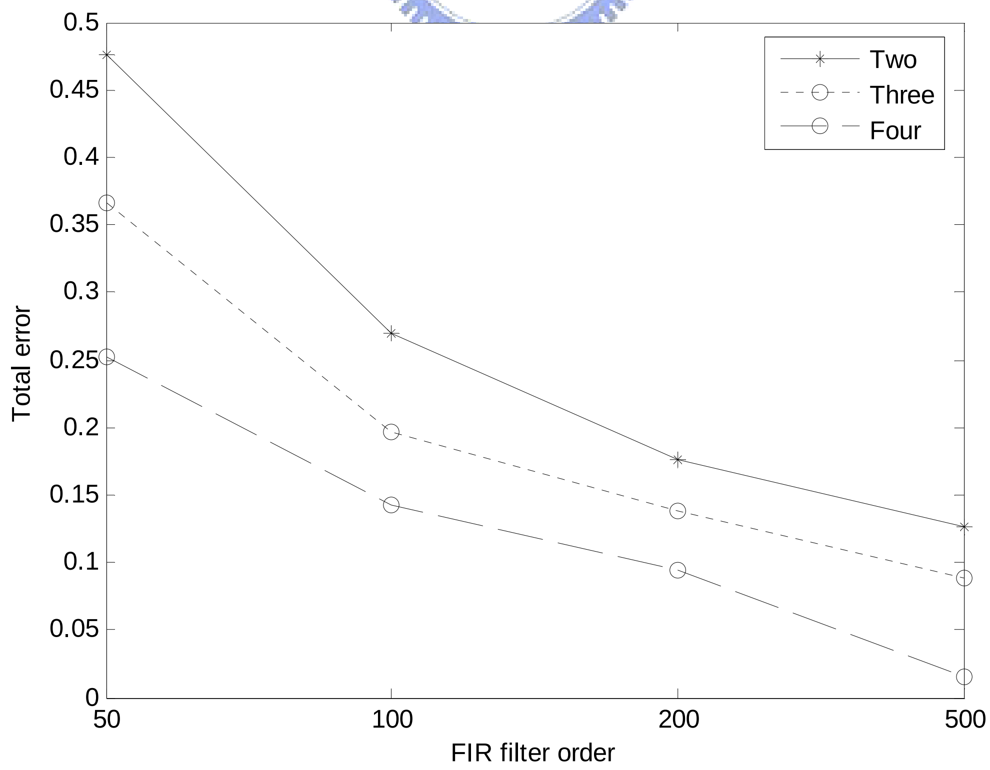


Figure 5.6 The total error $\varepsilon$ of different number of loudspeakers with different filter orders.

## 5.3.2 Shuffler Structure

The shuffler form crosstalk canceller can be obtained by factorizing Eq. (3.18) and results in Eq. (3.32). In this case, the side loudspeakers are located at $\pm 30^\circ$ and the center one is located at $0^\circ$. Obviously, it requires only three filters to form crosstalk canceller and has less computational complexity than direct forward type. In our simulation, we approximate these crosstalk cancellation filters by using FIR filter. Figure 5.7 and Figure 5.8 show the results of crosstalk cancellation in frequency and time domain; the FIR filter order is 200 taps with 140 modeling delays.

Table 5.5 lists the performance of crosstalk canceller with different FIR filter order. This table has shown that the performance of crosstalk cancellation is better while filter order is increasing. It should be mentioned that *CSF* is negative when filter order was under 100 taps. It is because that the main part of impulse response of ideal $\Sigma_2(z)$ and $\Delta_2(z)$ is behind 100 taps so that the approximated FIR filter is not precise enough. Therefore, the performance of crosstalk cancellation is worst. Next, we make the comparison between shuffler form and least forward type crosstalk canceller. The shuffler structure requires fewer filters than directly implement, but the performance of shuffler form is worst. Thus, the tradeoff between the performance and the number of the filter order must be considered while realizing the crosstalk canceller.

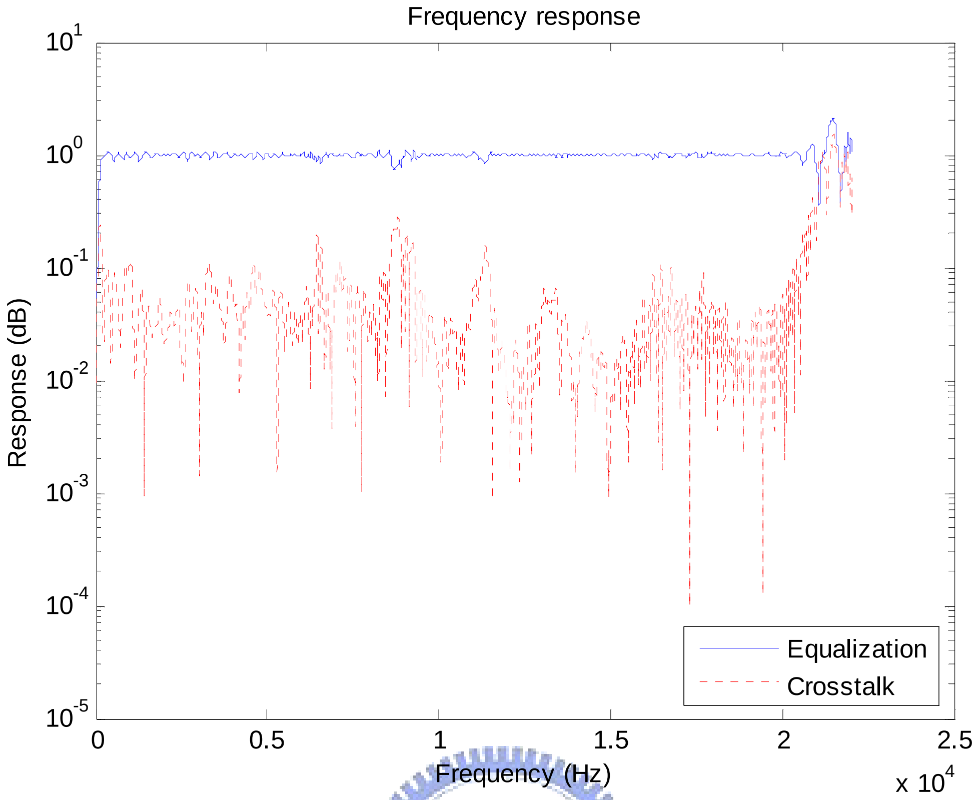


Figure 5.7 The crosstalk cancellation results in frequency domain of shuffler structure with the side loudspeakers located at $\pm 30^{\circ}$ and the center one located at $0^{\circ}$.

| Parameter / FIR Filter Order | $EQ(dB)$ | $CSF\,(dB)$ | $\varepsilon$ |
|---|---|---|---|
| 50 | 8.3822 | -0.7878 | 1.5602 |
| 100 | 10.6386 | -0.74746 | 1.4318 |
| 200 | 11.3793 | 9.3135 | 0.42041 |
| 500 | 11.8342 | 11.635 | 0.32014 |

Table 5.5 The performance of shuffler form for three loudspeakers setup with different FIR filter order.

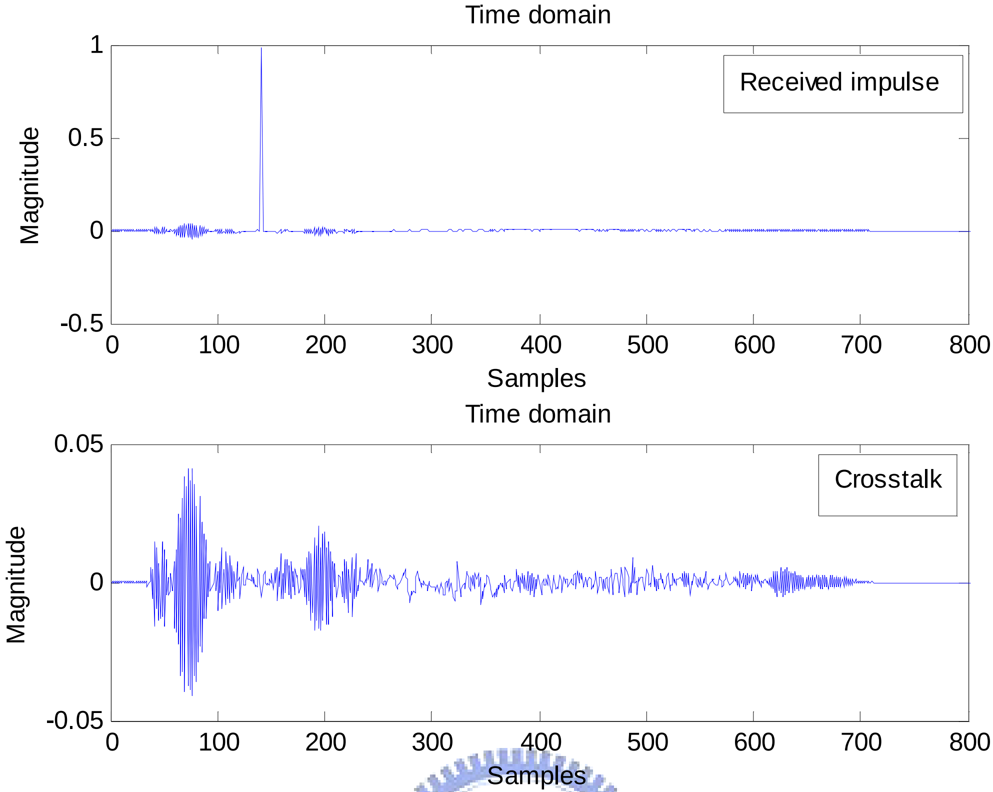


Figure 5.8 The crosstalk cancellation results in time domain of shuffler structure with the side loudspeakers located at $\pm 30^{\circ}$ and the center one located at $0^{\circ}$.

## 5.3.3 A Simplified Shuffler Form

As described in section 3.2.4, the crosstalk canceller of this structure is shown in Eq. (3.46). It requires only two filters for implementation and has less computation than the other structures introduced in previous section. In our simulation, we will use the first 200 taps of these two filters for realization since the later terms of them are smaller and can be ignored.

In addition, our case states that the side loudspeakers are located at $\pm 30^{\circ}$ and the center one is located at $0^{\circ}$. Figure 5.9 and Figure 5.10 show the results of crosstalk cancellation in frequency and time domain, respectively. Table 5.6 lists the performance of crosstalk canceller with different filter orders. Same as previous structures, which performance is better while the filter order increasing. It can be seen that the *CSF*, *EQ* and $\varepsilon$ of this structure are close to that of direct forward type and only requires two filters so that the economical realization is achievable.

Figure 5.11 and Figure 5.12 have shown the performance comparison of least square forward type, shuffler form and simplified shuffler form crosstalk canceller. It can be seen that *CSF* and *EQ* of simplified shuffler form is close to least square forward type and is better than shuffler form in Figure 5.11 and Figure 5.12. Also, we can obtain the same result for total error as shown in Figure 5.13. In summary, the shuffler form has less computation than least square forward type but its error performance is worse. Our proposed structure, the simplified shuffler form, which requires two filters ( $2\times 200$ multiplications) than shuffler form ( $3\times 200$ multiplications) and least square forward type ($6\times 200$ multiplications) to realize crosstalk cancellation and its performance is close to least square forward type.

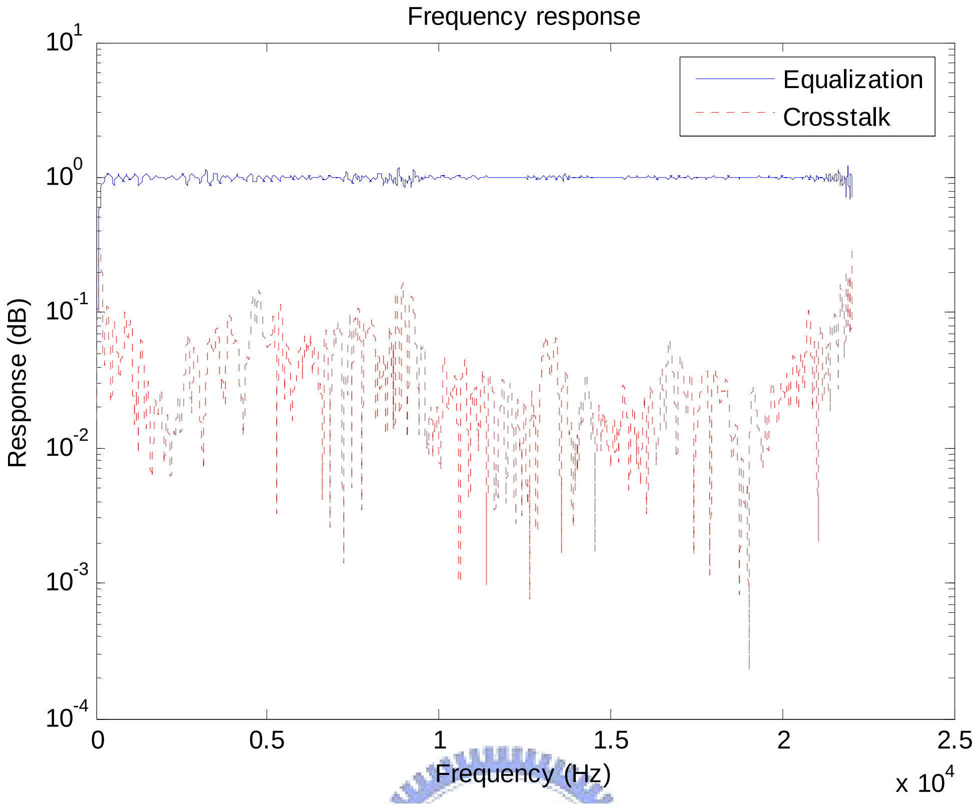


Figure 5.9 The crosstalk cancellation results in frequency domain of simplified shuffler form with the side loudspeakers located at $\pm 30^\circ$ and the center one located at $0^\circ$.

| Parameter / FIR Filter Order | $EQ(dB)$ | $CSF\,(dB)$ | $\varepsilon$ |
|---|---|---|---|
| 50 | 9.7496 | 10.196 | 0.41702 |
| 100 | 11.277 | 16.534 | 0.2152 |
| 200 | 11.709 | 20.969 | 0.14937 |
| 500 | 11.937 | 24.138 | 0.11537 |

Table 5.6 The performance of simplified shuffler form for three loudspeakers setup with different orders of FIR filters.

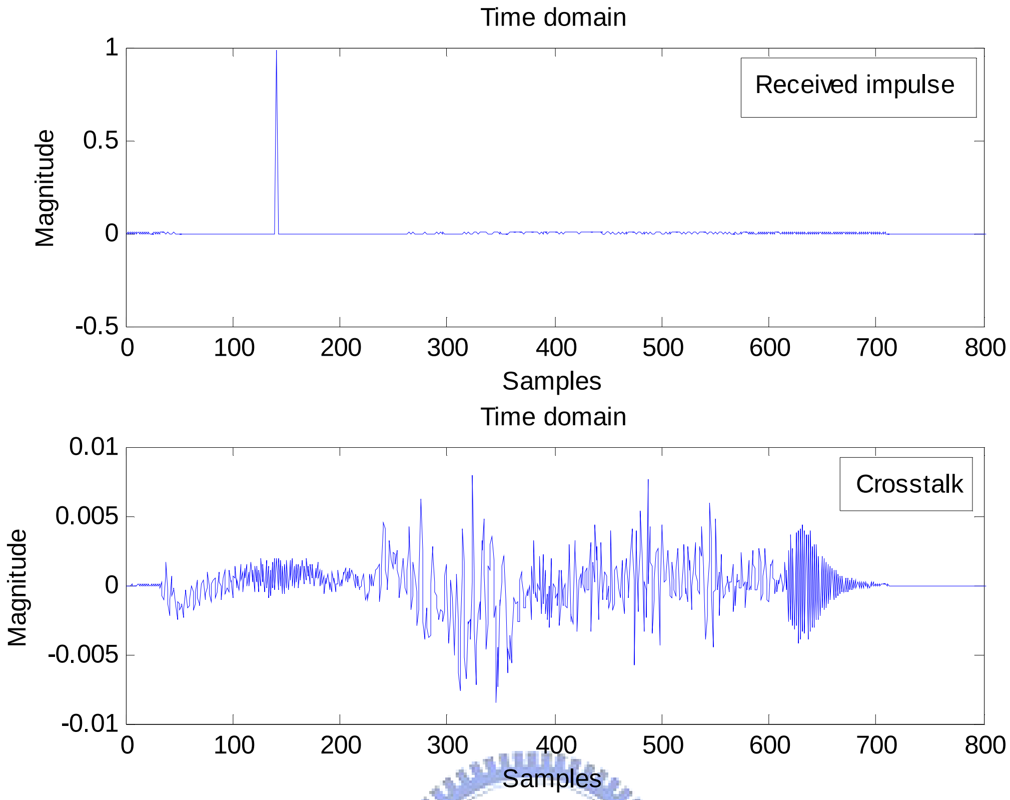


Figure 5.10 The crosstalk cancellation results in time domain of simplified shuffler form.

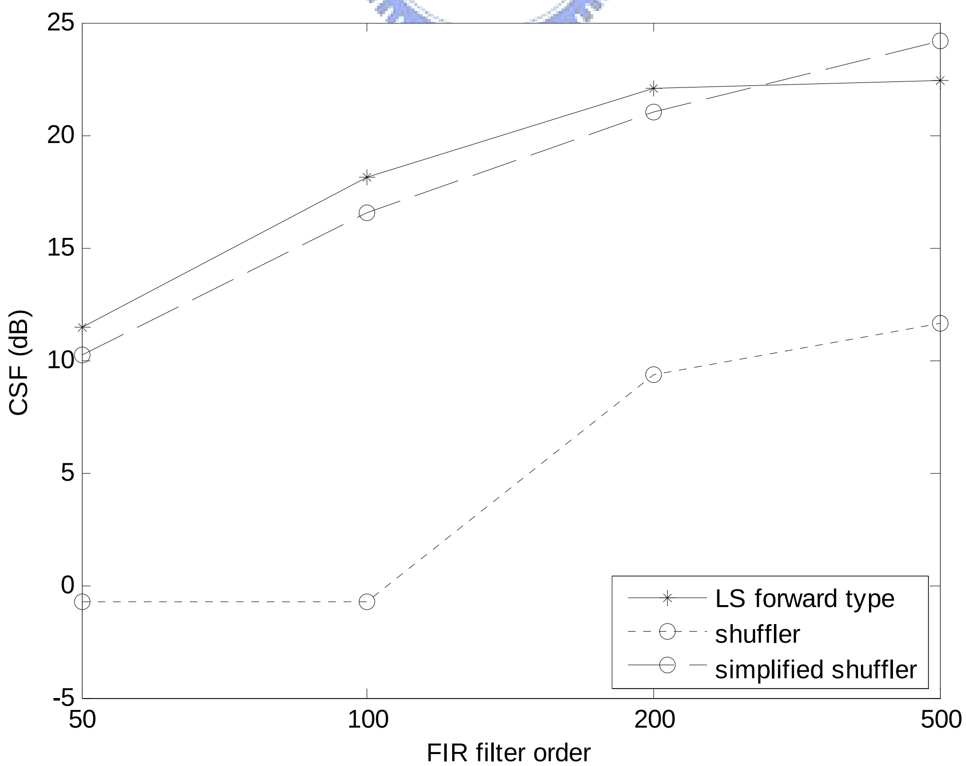


Figure 5.11 The *CSF* of different structure crosstalk canceller of three loudspeakers arrangement.

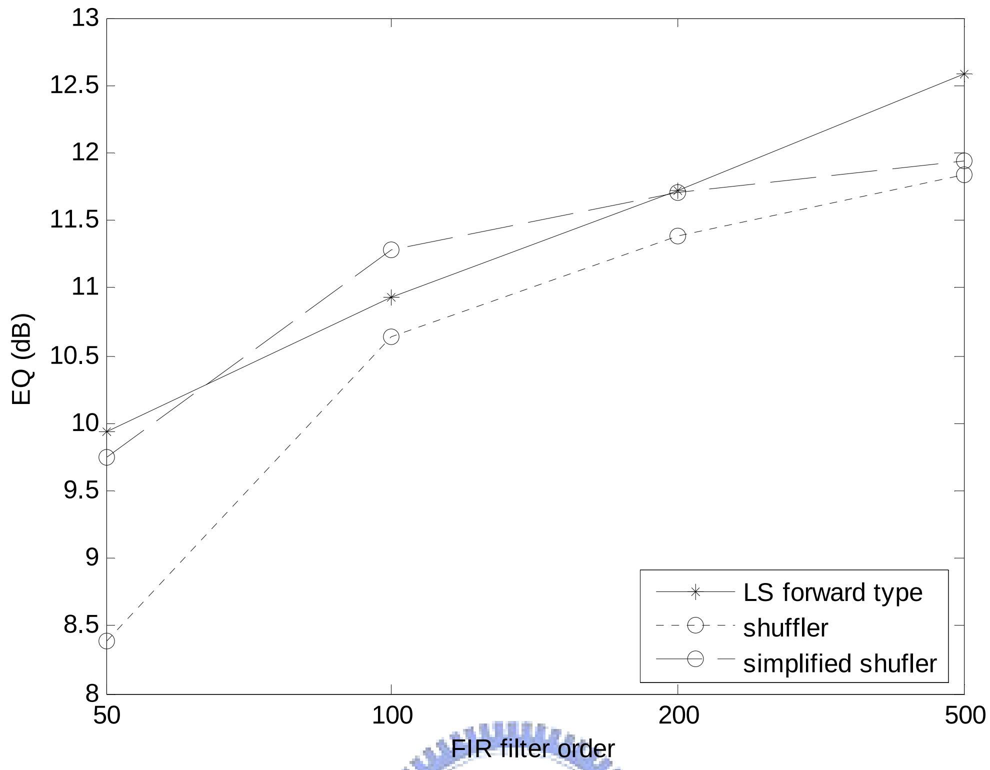


Figure 5.12 The *EQ* of different structure crosstalk canceller of three loudspeakers arrangement.

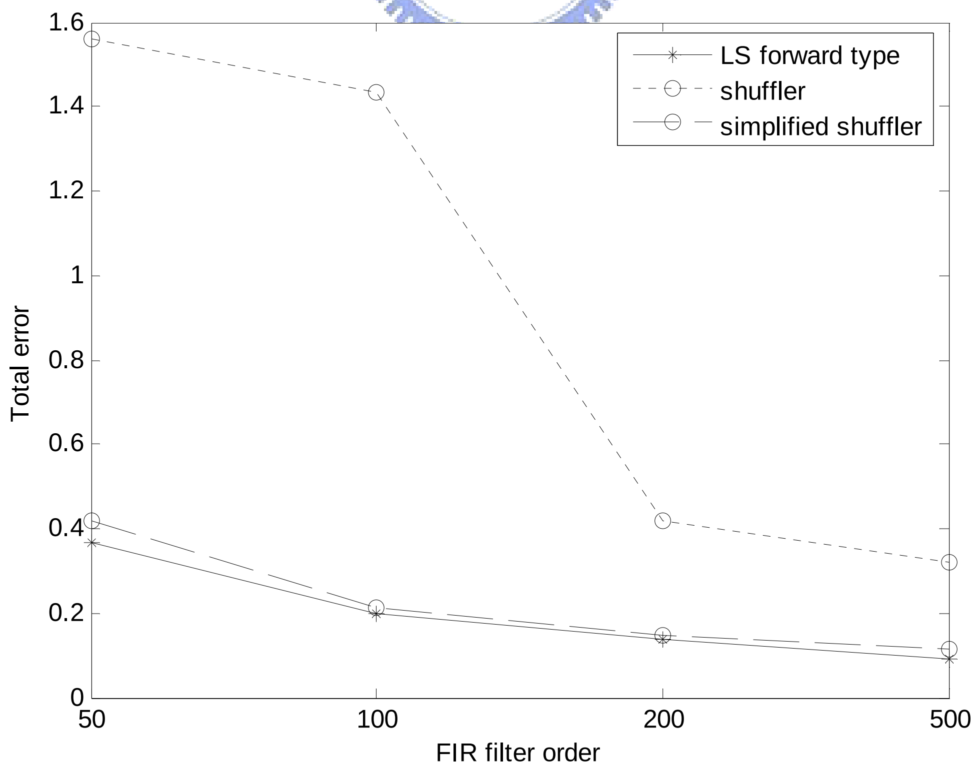


Figure 5.13 The total error of different structure crosstalk canceller of three loudspeakers arrangement.

### 5.3.4 A Simplified Shuffler Form of Four Loudspeakers Arrangement

In this section, we will discuss the crosstalk cancellation of four loudspeakers arrangement. In our simulation, the wider loudspeakers pair are located at $\pm 30°$ and the closer loudspeakers pair are located at $\pm 15°$. A simplified shuffler form crosstalk canceller is considered here. The structure of this type crosstalk canceller is shown in Eq. (3.62). Also, it requires only two filters to realize crosstalk canceller, that is, $\Sigma_4(z)$ and $\Delta_4(z)$. In this case, we will use FIR filter with 200 taps and 140 modeling delays to form these two filters.

Figure 5.14 and Figure 5.15 show the results of crosstalk cancellation in frequency and time domain, respectively. Table 5.7 lists the performance of crosstalk canceller with different filter orders. We can obtain that the performance of this type crosstalk canceller is better while the filter order increasing. In addition, this structure has less computation ( $2 \times 200$ multiplications) than directly implementation ( $8 \times 200$ multiplications) for four loudspeakers arrangement.

Figure 5.16 and Figure 5.17 show the performance comparison between two loudspeakers shuffler form and three and four simplified shuffler form. It can be seen the signal power ratio after crosstalk cancellation $R_c$ of three and four loudspeakers simplified shuffler form is similar and is larger then two loudspeakers shuffler form.

Also, the total error of four loudspeakers simplified shuffler is smaller then the other two structures. From the results described above, we can obtain a good performance crosstalk canceller while using multi-loudspeakers.

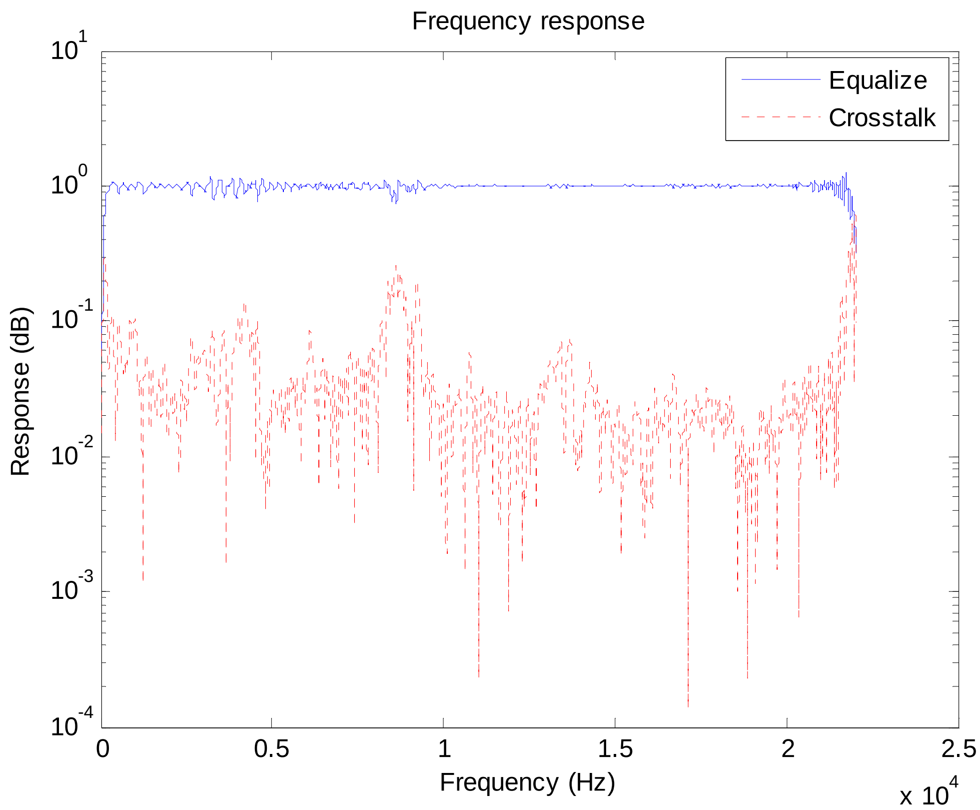


Figure 5.14 The crosstalk cancellation results in frequency domain of simplified shuffler form of four loudspeakers setup with the wider loudspeakers pair located at $\pm 30^{\circ}$ and the closer loudspeakers pair located at $\pm 15^{\circ}$.

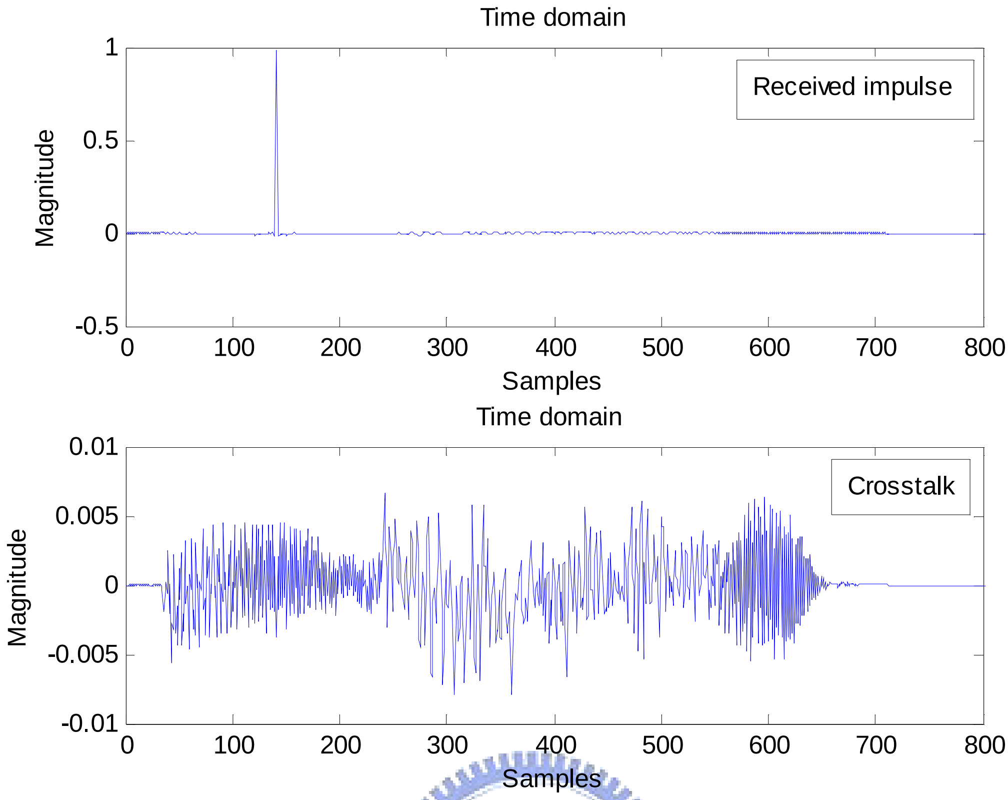


Figure 5.15 The crosstalk cancellation results in time domain of simplified shuffler form of four loudspeakers setup with the wider loudspeakers pair located at $\pm 30^\circ$ and the closer loudspeakers pair located at $\pm 15°$.

| **Parameter** / **FIR Filter Order** | $EQ(dB)$ | $CSF\,(dB)$ | $\varepsilon$ |
|---|---|---|---|
| 50 | 8.8797 | 21.369 | 0.37428 |
| 100 | 10.577 | 26.049 | 0.17729 |
| 200 | 10.982 | 29.515 | 0.11532 |
| 500 | 11.264 | 32.733 | 0.075 |

Table 5.7 The performance of simplified shuffler form for four loudspeakers setup with different orders of FIR filters.

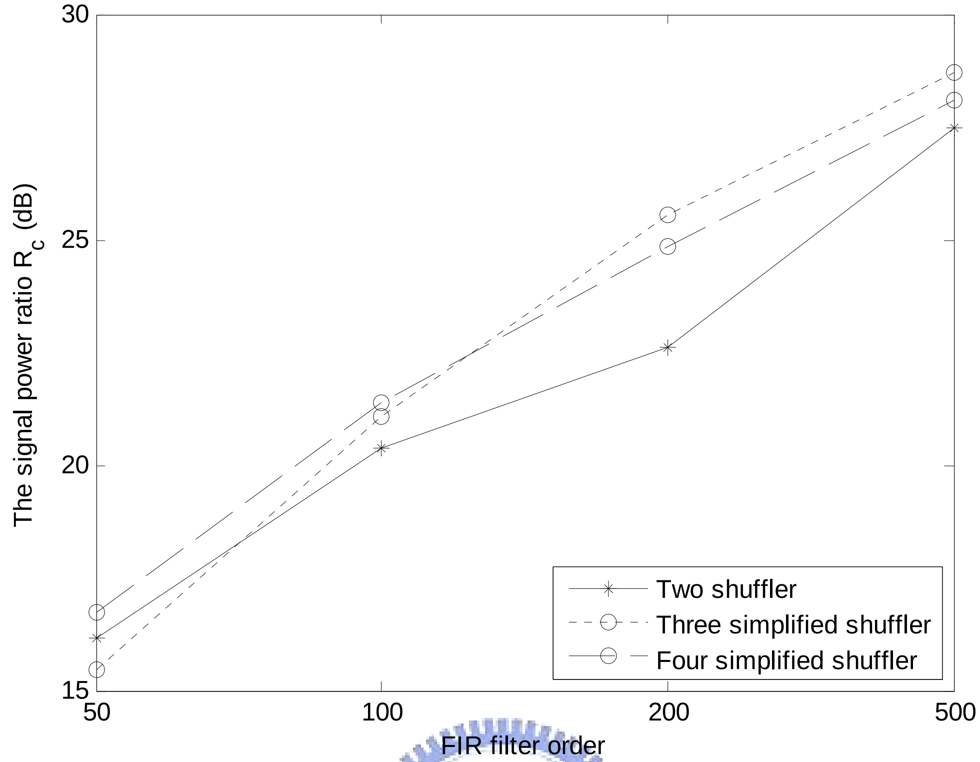


Figure 5.16 The signal power ratio after crosstalk cancellation $R_c$ of two loudspeakers shuffler form and three and four loudspeakers simplified shuffler form with different FIR filter orders.

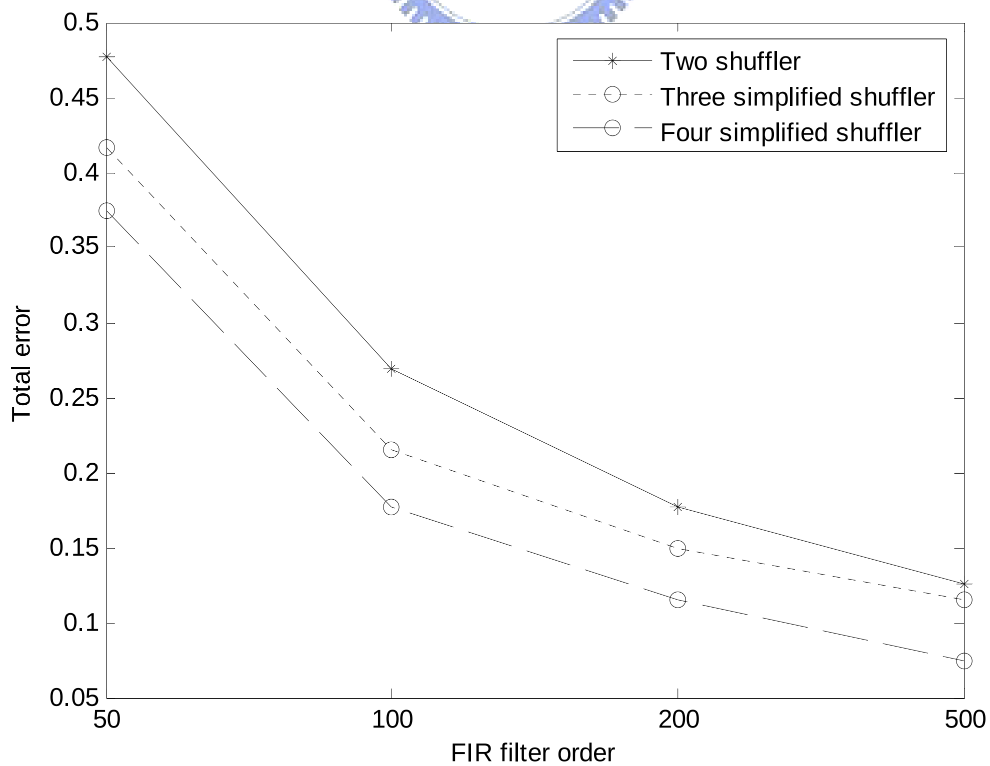


Figure 5.17 The total error of two loudspeakers shuffler form and three and four loudspeakers simplified shuffler form with different FIR filter orders.

## 5.4 Perturbation Analysis

The previous works realize the crosstalk canceller for fixed position of listener's head. In this section, we will investigate the robustness of crosstalk canceller after filter modeling and listener's head movement. As shown in Figure 5.18 and Figure 5.19, the performance of crosstalk canceller is degrades due to perturbation. In our case, the *CSF* reduces from 15.499 dB to 11.159 dB and the *EQ* from 13.046 dB to 9.601 dB and the total error from 0.17652 to 0.96565.

In addition, we will compare two loudspeakers arrangement with multiple loudspeakers arrangement. From the simulation results, it can be demonstrated that multiple loudspeakers setup is more robust than two loudspeakers setup after perturbation. Let's begin the analysis on filter modeling.

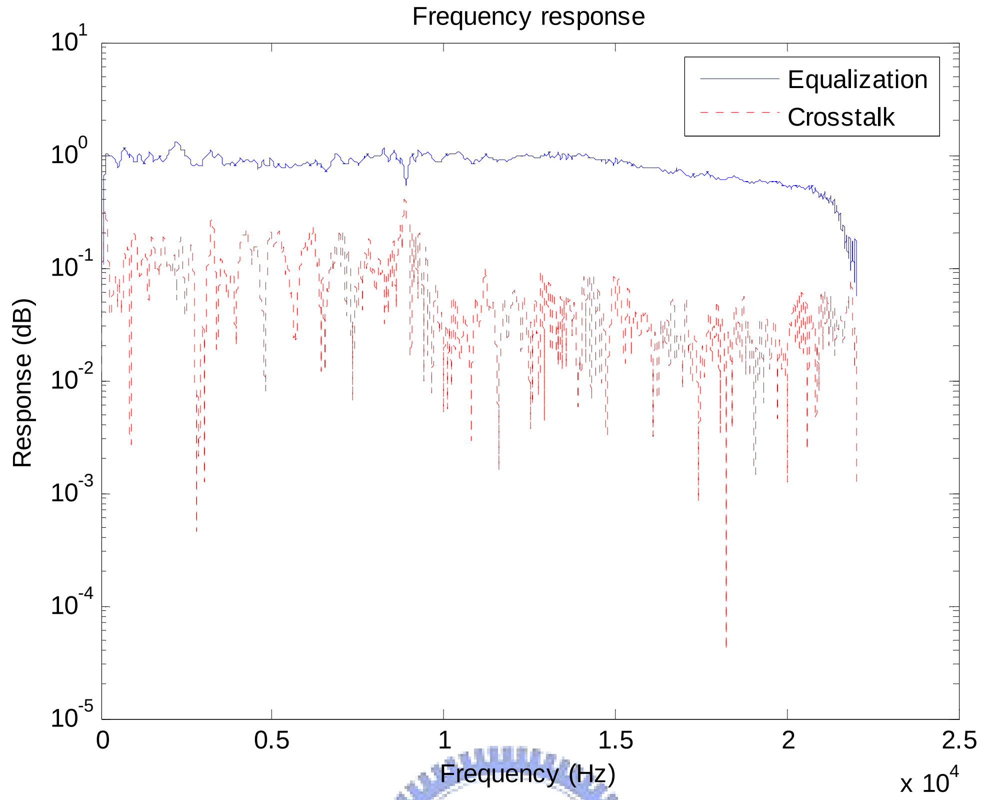


Figure 5.18 The crosstalk cancellation result in frequency domain due to perturbation.

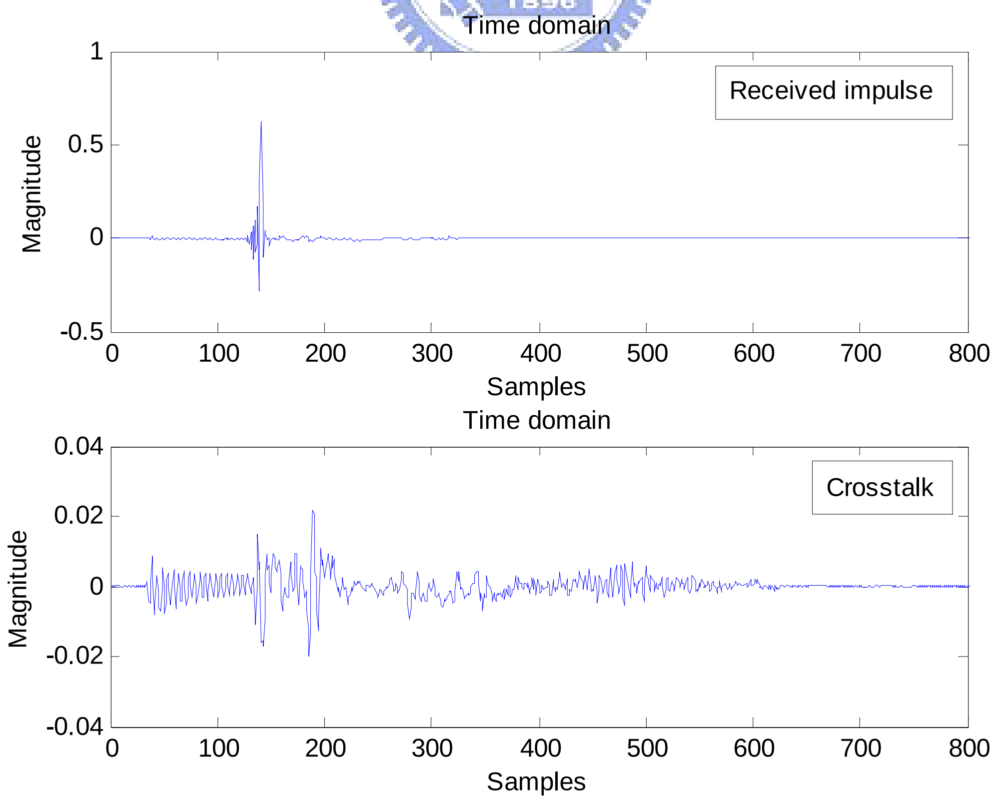


Figure 5.19 The crosstalk cancellation result in time domain due to perturbation.

### 5.4.1 Simulations on Filter Modeling

In our simulation, we will approximate the FIR filter by using pole-zero modeling such as *prony* approximation method. We will discuss how the filter modeling error affects the performance of crosstalk cancellation. First, we consider the two loudspeakers shuffler form. As described in section 4.3.1, the modeling error vector after crosstalk cancellation can be written as follows

$$E_{2,Shuffler} = \frac{1}{2}\begin{bmatrix} e_{s,1}(g_1+g_2)+e_{d,1}(g_1-g_2) \\ e_{s,1}(g_1+g_2)-e_{d,1}(g_1-g_2) \end{bmatrix} = \begin{bmatrix} Eq. \\ Cr. \end{bmatrix} \tag{5.11}$$

where $e_{s,1}$ and $e_{d,1}$ represent the error after perturbation of modeling filter $\Sigma$ and $\Delta$, respectively. The first element of Eq. (5.11) indicates the equalized error and the second element represents the crosstalk error. The theoretical $e_{s,1}$ and $e_{d,1}$ are described in Eq.(4.14) and Eq. (4.16) so that the theoretical modeling error vector of Eq. (5.11) is

$$E_{2,Shuffler} \cong \begin{bmatrix} eF \\ (2F^2-1)e \end{bmatrix} \tag{5.12}$$

in which $F = g_2/g_1$ with modeling error *e*. In our simulation, the theoretical equalized and crosstalk error are obtained from Eq. (5.12).

Next, we illustrate how to obtain the experimental modeling error. Here we model the FIR filters $\Sigma$ and $\Delta$ by using *Prony* approximation method and results in the

IIR filters $\Sigma_+$ and $\Delta_+$. Thus, the modeling error $e_{s,1}$ and $e_{d,1}$ are

$$e_{s,1} = \Sigma_+ - \Sigma \tag{5.13}$$

$$e_{d,1} = \Delta_+ - \Delta \tag{5.14}$$

We then can obtain the simulated equalized and crosstalk error from Eq. (5.11), Eq. (5.13) and Eq. (5.14).

Table 5.8 lists the result of theoretical and simulated error norm of Eq. (5.11). In this table, the filter tap $N_n/N_p$ indicate the IIR filter with the tap of numerator and denumerator, respectively. In addition, $\|Eq.\|$ and $\|Cr.\|$ denote the modeling error norm of equalization and crosstalk part which means that the degree of modeling error affects the performance of equalization and crosstalk. In this table, there are lots of differences between the theoretical and the experimental error norm as shown in Figure 5.20 and Figure 5.21. The actual $e_{s,1}$ and $e_{d,1}$ is found by approximating the filter $\Sigma$ and $\Delta$ directly. The theoretical $e_{s,1}$ and $e_{d,1}$ is derived by extracting the common part of $\Sigma$ and $\Delta$ so that there is a common modeling error exists $e_{s,1}$ and $e_{d,1}$. But we can obtain the same result that the modeling error norm is larger while IIR modeling from theoretical analysis and numerical simulation. It is because the modeling error is introduced and the performance of crosstalk cancellation is worse. In addition, the error vector of direct forward type 2 derived in [20] is shown below

$$E_2 = \frac{g_1}{g_1^2 - g_2^2}\begin{bmatrix} g_2 e \\ g_1 e \end{bmatrix}$$

$$= \frac{1}{1-F^2}\begin{bmatrix} eF \\ e \end{bmatrix} \tag{5.15}$$

where $F$ is as same as above. Figure 5.22 and Figure 5.23 show the equalized error and the crosstalk error of shuffler structure and direct forward type 2. From Figure 5.22 and Figure 5.23, direct forward type 2 has better immunity of the perturbation than shuffler form. But shuffler structure has less computation than direct forward type two. In our case, the shuffler form only requires two FIR filters (i.e. 400 multiplications) and direct forward type two requires four FIR filters (i.e. 800 multiplications).

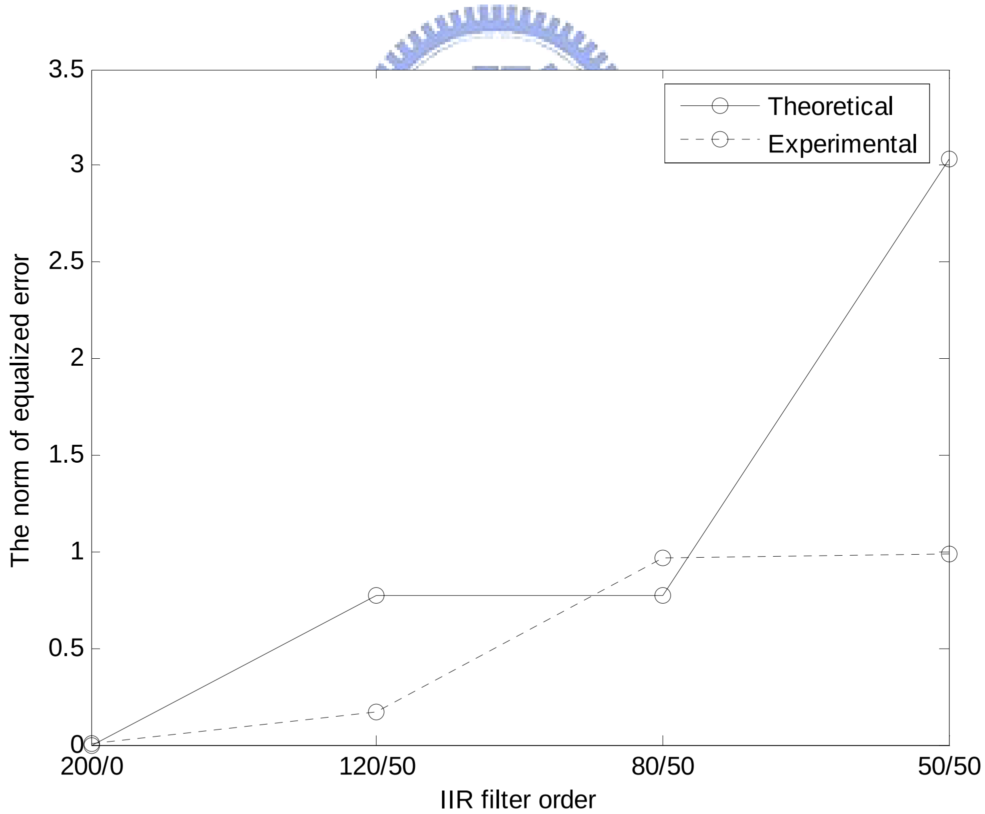


Figure 5.20 The norm of equalized error of two loudspeakers shuffler form with different orders of IIR filter. The solid line indicates the theoretical value and the doted line denotes the experiment result.

| Filter order | Theoretical | | Experimental | |
|---|---|---|---|---|
| | $\|Eq.\|$ | $\|Cr.\|$ | $\|Eq.\|$ | $\|Cr.\|$ |
| 200/0 | $9.8410\times10^{-18}$ | $3.9690\times10^{-17}$ | $4.8783\times10^{-15}$ | $5.1821\times10^{-15}$ |
| 120/50 | 0.7692 | 4.8939 | 0.1639 | 0.0746 |
| 80/50 | 0.7764 | 5.1283 | 0.9722 | 0.0805 |
| 50/50 | 3.037 | 20.5832 | 0.9879 | 0.5094 |

Table 5.8 The error norm of equalization and crosstalk part for two loudspeakers shuffler form crosstalk canceller.

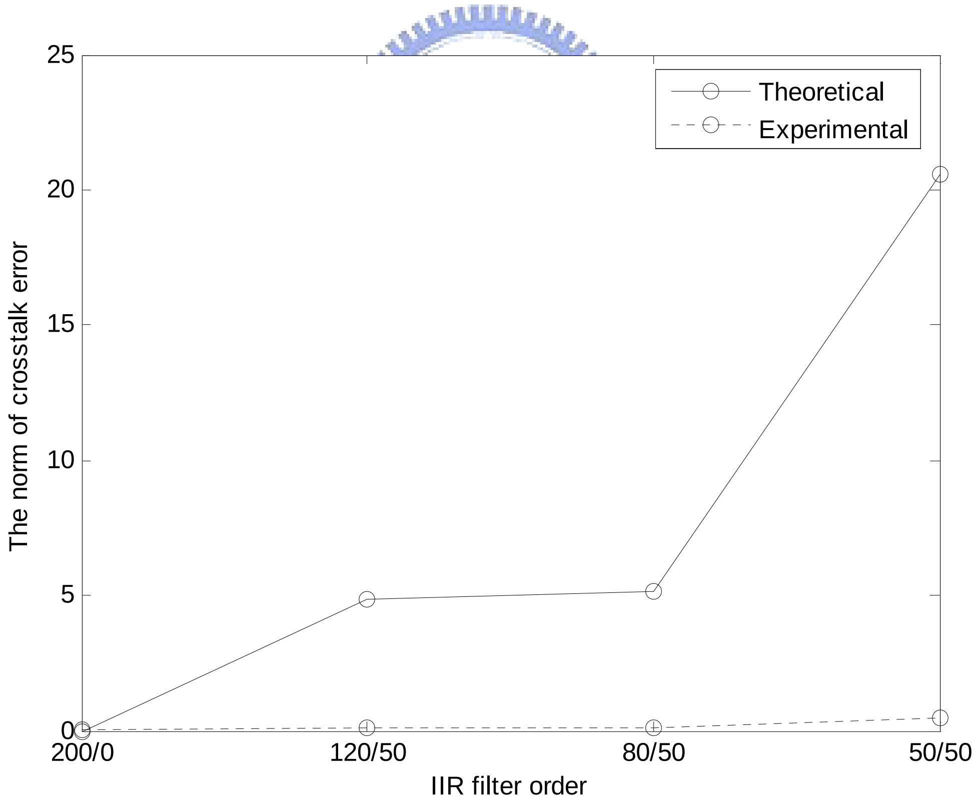


Figure 5.21 The norm of crosstalk error of two loudspeakers shuffler form with different orders of IIR filter. The solid line indicates the theoretical value and the doted line denotes the experiment result.

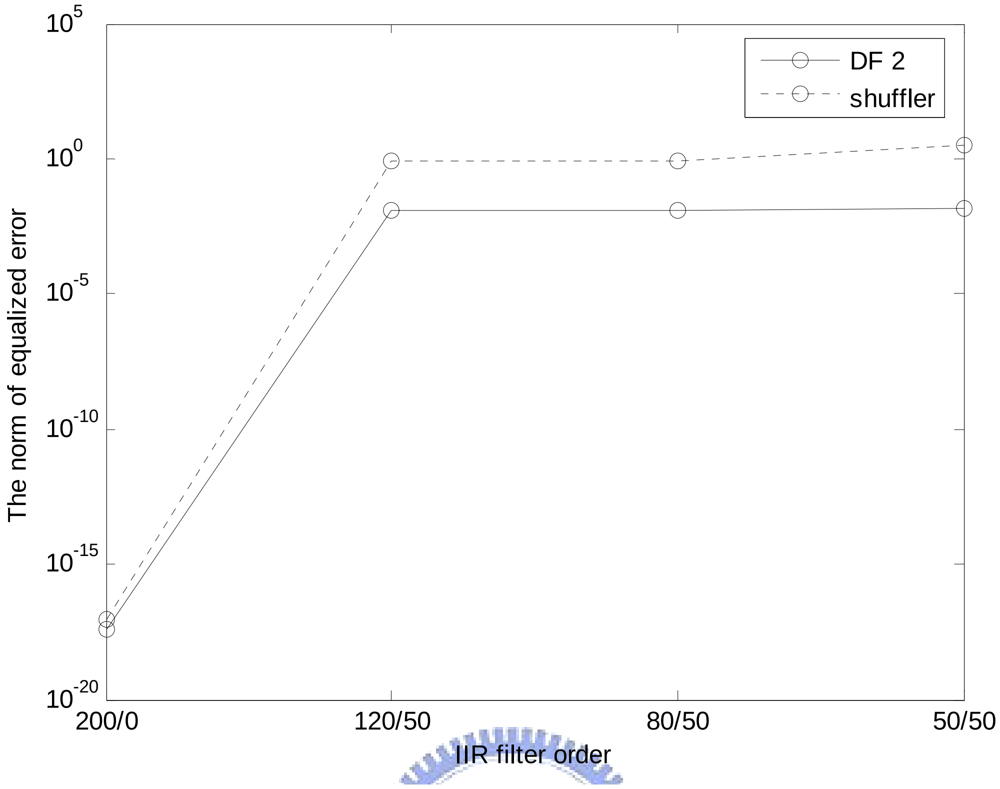


Figure 5.22 The equalized error of shuffler form versus direct forward type 2.

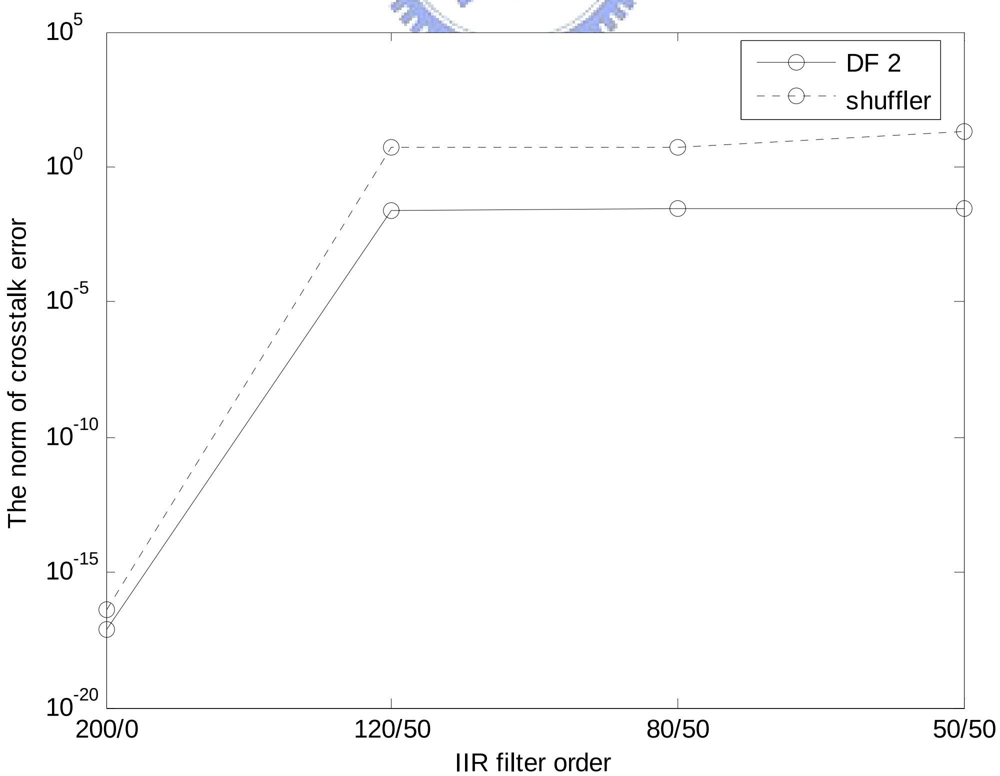


Figure 5.23 The crosstalk error of shuffler form versus direct forward type 2.

Next, we consider the three loudspeakers arrangement. The filter modeling error vector of shuffler form as described in Eq. (4.22) and that of simplified shuffler form as described in Eq. (4.47). Also, we can derive the modeling error vector of four loudspeakers simplified shuffler form and can be written as

$$E_{4,simplified} = \frac{1}{2}\begin{bmatrix} e_{d,4}\left(g_1 - g_4\right) + e_{s,4}\left(g_2 + g_3\right) \\ -e_{d,4}\left(g_1 - g_4\right) + e_{s,4}\left(g_2 + g_3\right) \end{bmatrix} \tag{5.16}$$

where $e_{s,4}$ and $e_{d,4}$ represent the filter modeling error of the summed filter $\Sigma_4$ and the differenced filter $\Delta_4$.

Figure 5.24 and Figure 5.25 show the equalized error norm and the crosstalk error norm of different structures of crosstalk canceller. In Figure 5.24, it can be seen that equalized error norm of four loudspeakers simplified shuffler form is smaller then the other structures which means four loudspeakers simplified shuffler form has better immunity of modeling error for equalization. From Figure 5.25, the crosstalk error norm of three and four loudspeakers simplified shuffler form is smaller then the other two structures. But the crosstalk error norm of three loudspeakers shuffler form is worse than two loudspeakers shuffler form.

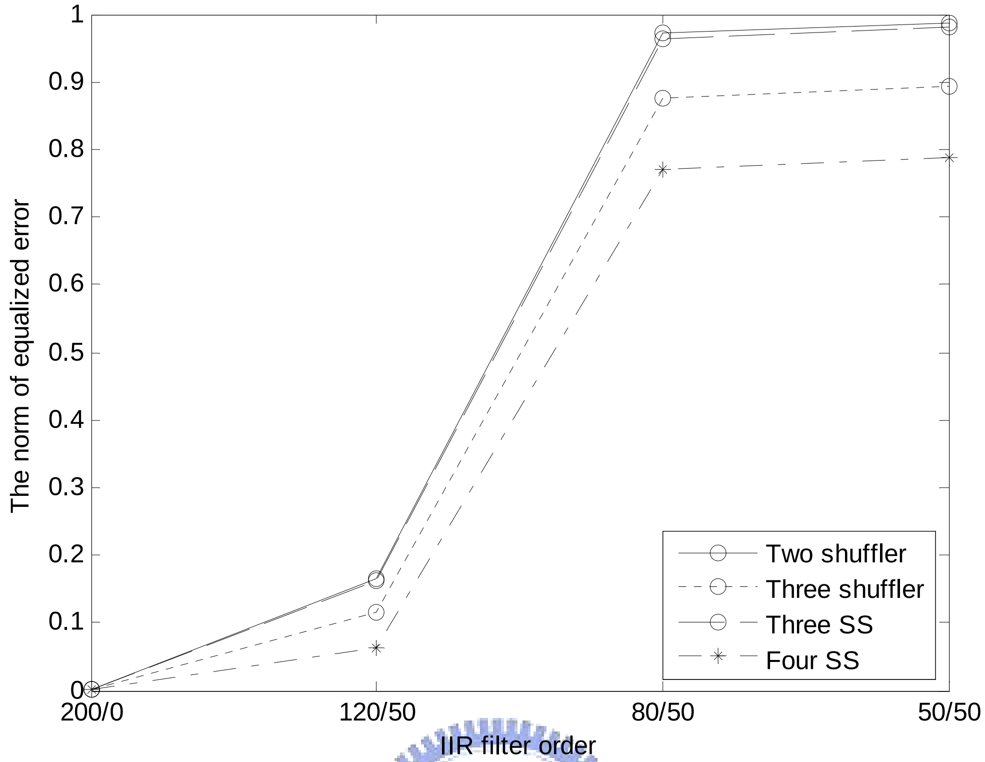


Figure 5.24 The equalized error of different structures of crosstalk canceller with different IIR filter orders.

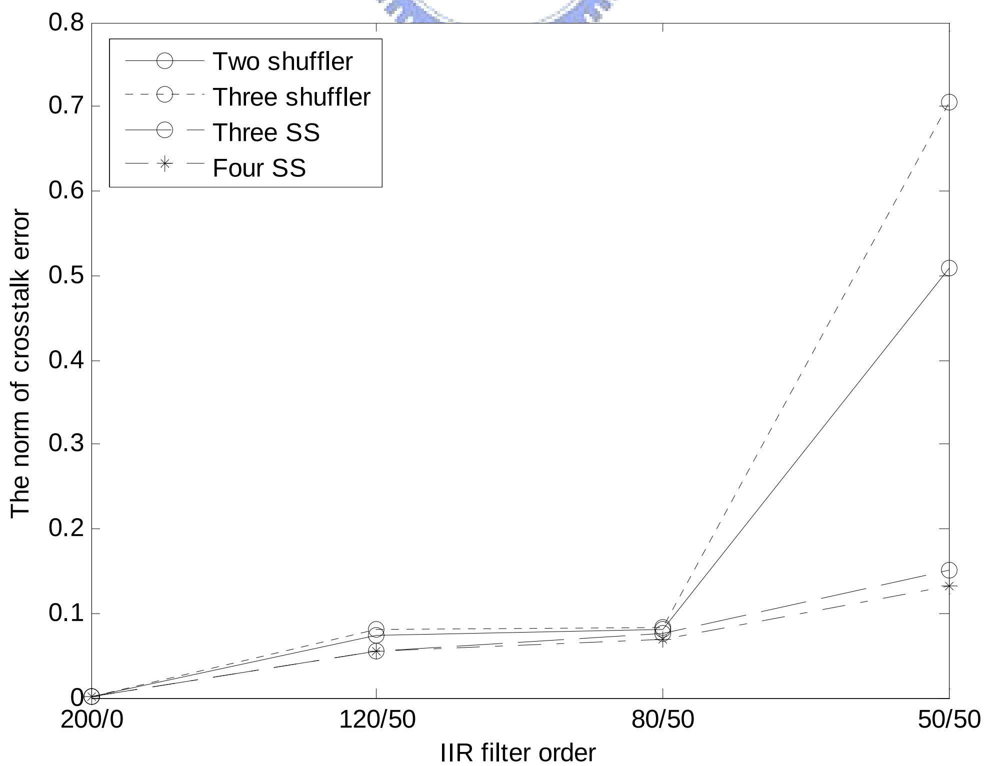


Figure 5.25 The crosstalk error of different structures of crosstalk canceller with different IIR filter orders.

### 5.4.2 Simulations on Head movement

Since the head movement will change the acoustic transmission path and worsen the performance of crosstalk canceller for the default position. In this section, we will discuss how the head movement affects the performance of crosstalk canceller for difference numbers of loudspeakers.

In literature [2], it has shown that crosstalk canceller is sensitive for lateral movement. Thus, the lateral head movement is our focus in the simulation. First, we consider the shuffler structure of two loudspeakers arrangement. In our simulation, the default position of loudspeaker pair are located at $\pm 30^{\circ}$. Table 5.9 lists the performance of crosstalk canceller after head moving. It can be seen the performance is worse while head moving. In Table 5.9, $R$ and $L$ represent the right side loudspeaker and the left side loudspeaker, respectively. The first two rows of this table indicate the head move to right and the last two rows indicate the head move to left.

Now, we consider the three loudspeakers arrangement. The structure of crosstalk canceller we discuss here is the simplified shuffler form. In this case, we insert the center loudspeaker at $0^{\circ}$ facing the listener and the side loudspeakers are located at $\pm 30^{\circ}$ for default position. The simulated results are listed in Table 5.10. In Table 5.10, $R$ and $L$ as described above and $C$ represent the center loudspeaker. Also, the performance of crosstalk canceller is worst while head moving in this case.

Next, we discuss the robustness of head movement for four loudspeakers arrangement. We let the wider loudspeaker pair located at $\pm 30°$ and the closer loudspeaker pair located at $\pm 15°$ for default position. The simulated results are listed in Table 5.11. Also, we can obtain the same result as described in Table 5.9 and Table 5.10. To compare these three tables. Same as analysis on filter modeling, it can be shown that the crosstalk canceller of multiple loudspeakers is more robust while listener's head moving as shown in Figure 5.26 and Figure 5.27. In Figure 5.26, where *A* indicates the default position of the side loudspeakers pair, *B*, *C* represent the position corresponding to the listener's head movement to right and *D*, *E* represent the listener movement to left.

| **Parameter** / **Position** | | $R_c(dB)$ | $EQ(dB)$ | Total error $\varepsilon$ |
|---|---|---|---|---|
| $R:30°$ | $L:30°$ | 22.6076 | 13.046 | 0.17652 |
| $R:25°$ | $L:35°$ | 12.224 | 11.06 | 1.7705 |
| $R:20°$ | $L:40°$ | 8.577 | 9.9881 | 2.4567 |
| $R:35°$ | $L:25°$ | 10.478 | 11.371 | 1.7705 |
| $R:40°$ | $L:20°$ | 6.1287 | 11.279 | 2.4567 |

Table 5.9 The performance of crosstalk cancellation for two loudspeakers shuffler form after head movement.

| Position \ Parameter | | | $R_c(dB)$ | $EQ(dB)$ | Total error $\varepsilon$ |
|---|---|---|---|---|---|
| $R:30°$ | $C:0°$ | $L:30°$ | 25.532 | 11.709 | 0.14937 |
| $R:25°$ | $C:-5°$ | $L:35°$ | 12.489 | 11.124 | 1.826 |
| $R:20°$ | $C:-10°$ | $L:40°$ | 6.9758 | 10.347 | 2.328 |
| $R:35°$ | $C:5°$ | $L:25°$ | 13.582 | 10.436 | 1.826 |
| $R:40°$ | $C:10°$ | $L:20°$ | 9.1144 | 9.5206 | 2.328 |

Table 5.10 The performance of crosstalk cancellation for three loudspeakers simplified shuffler form after head movement.

| Position \ Parameter | | $R_c(dB)$ | $EQ(dB)$ | Total error $\varepsilon$ |
|---|---|---|---|---|
| **Wider** | **Closer** | | | |
| $R:30°, L:30°$ | $R:15°, L:15°$ | 24.8598 | 10.982 | 0.11532 |
| $R:25°, L:35°$ | $R:10°, L:20°$ | 14.629 | 10.633 | 1.769 |
| $R:20°, L:40°$ | $R:5°, L:25°$ | 9.7383 | 9.9711 | 2.3139 |
| $R:35°, L:25°$ | $R:20°, L:10°$ | 14.417 | 10.56 | 1.769 |
| $R:40°, L:20°$ | $R:25°, L:5°$ | 10.396 | 9.697 | 2.3139 |

Table 5.11 The performance of crosstalk cancellation for four loudspeakers simplified shuffler form after head movement.

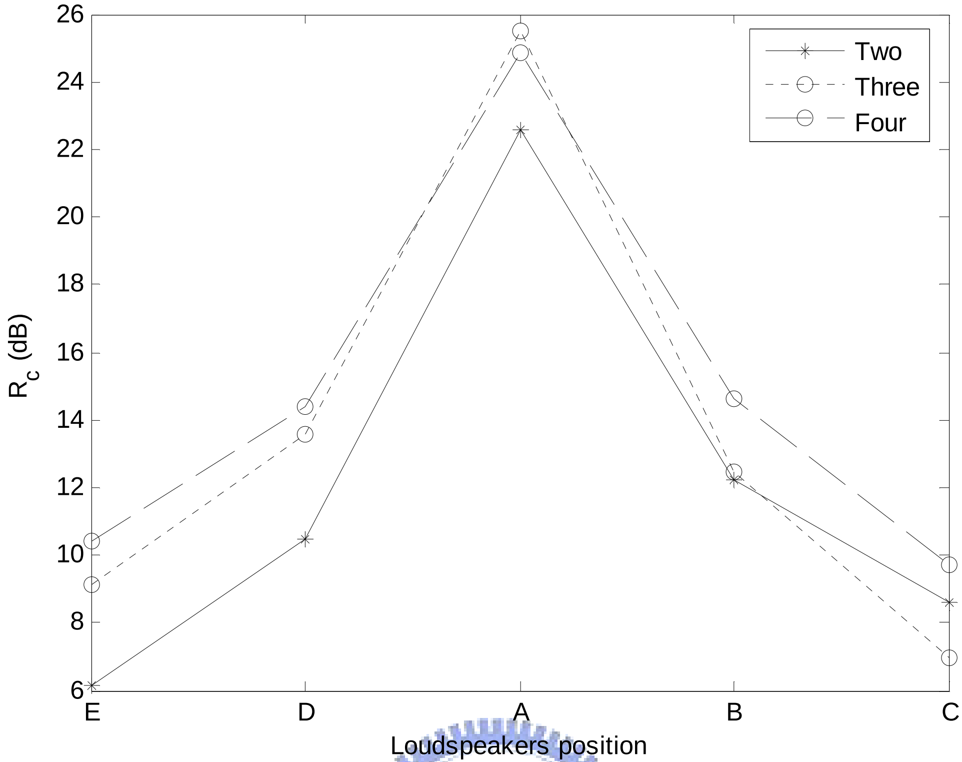


Figure 5.26 The signal power ratio $R_c$ of crosstalk cancellation after head moving for different numbers of loudspeakers.

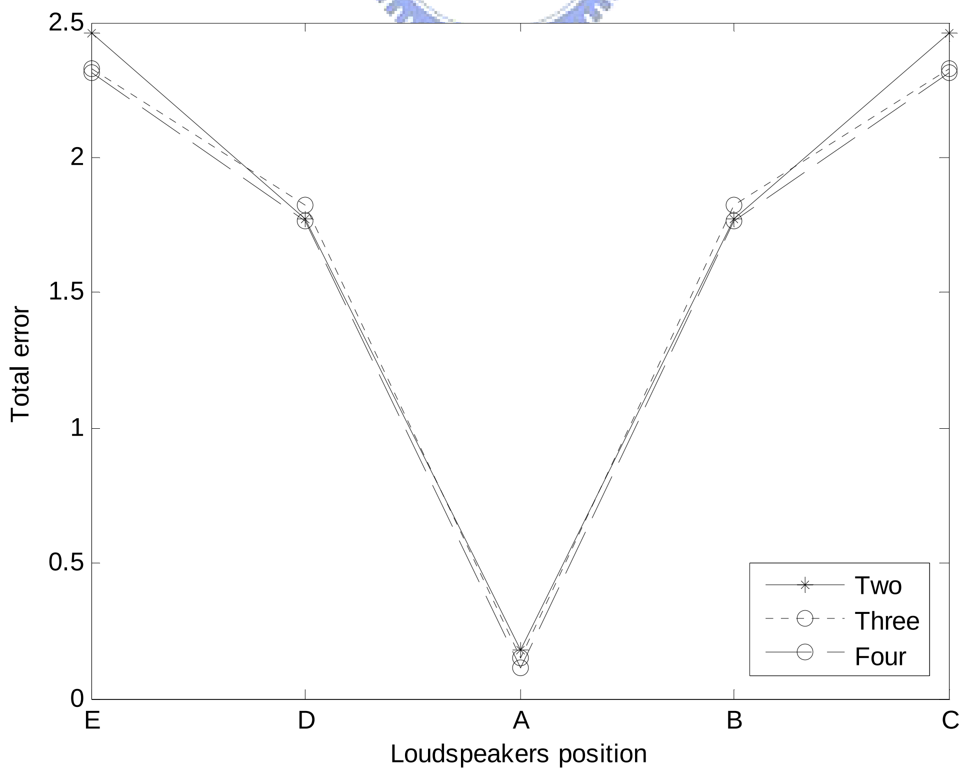


Figure 5.27 The total error of crosstalk cancellation after head moving for different numbers of loudspeakers.

# Chapter 6

# Conclusions

In this thesis, we investigate the different structures of crosstalk canceller of three loudspeakers arrangement and we make the performance comparison such as the signal power ratio after crosstalk cancellation $R_c$, $EQ$ and total error between conventional two loudspeakers and multiple loudspeakers arrangement. From the computer simulated results, we can see that $R_c$ and total error of crosstalk cancellation is better while using multiple loudspeakers for 3D sound reproduction.

In addition, we develop a simplified shuffler form which only requires $2\times200$ multiplications and is less then least square forward type ($6\times200$ multiplications) and shuffler form ($3\times200$ multiplications). In addition, simplified shuffler form is easer to realize than the other structures of three loudspeakers setup such structure can achieve the performance as well as least square forward type crosstalk canceller. Also, we can use the same procedure as described in section 3.2.4 to find the simplified shuffler form of four loudspeakers arrangement.

In chapter 4, we have discussed how the perturbation of filter modeling and head movement affects the performance of crosstalk canceller. The theoretical analysis results that the crosstalk canceller of multiple loudspeakers arrangement is more robust for perturbation. We can demonstrate this fact from the computer simulations as described in chapter 5. In addition, we derive the optimum loudspeakers position for different loudspeakers arrangement.

Summarily, our proposed structure can reduce the cost of required filters to realize crosstalk canceller for 3D sound reproduction while using multiple loudspeakers which has better performance of crosstalk cancellation and better immunity for perturbation.

# Bibliography


[1] W. G.. Gardner, *3D audio using loudspeakers*, Kluwer Academic Pub: Boston, 1998.

[2] D. B. Ward and G. W. Elko, “Optimum loudspeaker spacing for robust crosstalk cancellation,” in *Proc. IEEE ICASSP’*98, Seattle, WA, vol. 6, pp. 3541-3544, May 1998.

[3] D. B. Ward and G. W. Elko, “Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,” *IEEE Signal Processing Lett.*, vol. 6, pp. 106-108, May 1999.

[4] D. B. Ward and G. W. Elko, “A robustness analysis of 3D audio using loudspeakers,” in *Proc. IEEE Workshop on Applicat. of Signal Processing to Audio and Acoust. (WASPAA-99)*, New Paltz, NY, USA, Oct. 1999.

[5] D. B. Ward, “Joint least squares optimization for robust acoustic crosstalk cancellation,” *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 211-215,

Feb. 2000.

[6] D. B. Ward and G. W. Elko, “A new robust system for 3D audio using loudspeakers,” *IEEE Acoustics, Speech and Signal Processing*, vol. 2, pp. 1781-1784, June 2000.

[7] H. S. Kim, P. M. Kim and H. B. Kim, “Crosstalk cancellation algorithm for 3D sound reproduction,” *ETRI Journal*, vol. 22, no. 2, pp. 11-19, June 2000.

[8] J. J. Lopez and A. Gonzalez, “Experimental evaluation of crosstalk cancellation regarding loudspeakers’ angle of listening,” *IEEE signal Processing Lett.*, vol. 8, no. 1, Jan. 2001.

[9] J. Yang, W. S. Gan and S. E. Tang, “Development of virtual sound imaging system using triple elevated speakers,” *IEEE Trans. Consumer Electron.*, vol. 50, no. 3, pp. 916-922, Aug. 2004.

[10] J. L. Bauck and D. H. Cooper, “Generalized transaural stereo and applications,” *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683-685, Sept. 1996.

[11] C. Kyriakakis, “Fundamental and technological limitations of immersive audio systems,” *Proc. IEEE*, vol. 86, no. 5, pp. 941-951, May 1998.

[12] C. Kyriakakis, P. Tskalides and T. Holman, “Surrounded by sound,” *IEEE Signal Processing Magazine*, vol. 16, pp. 55-66, Jan. 1999.

[13] M. R. Schroeder, “Models of hearing,” *Proc. IEEE*, vol. 63, no. 9, pp. 1332-1350,

Sept. 1975.

[14] A. Mouchtaris, P. Reveliotis and C. Kyriakakis, “Inverse filter design for immersive audio rendering over loudspeakers,” *IEEE Trans. Multimedia*, vol. 2, no. 2, pp. 77-87, June 2000.

[15] M. M. Mansour and A. Mehrotra, “Model-order reduction based on Prony’s method,” *IEEE Design, Automation and Test in Europe Conference and Exhibition*, pp. 530-535, 2003.

[16] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 2, pp. 145-152, Feb. 1988.

[17] O. Kirkeby, P. A. Nelson and H. Hamada, “The stereo dipole-A virtual source imaging system using two closely spaced loudspeakers,” *J. Audio Eng. Soc.*, vol. 46, pp. 387-395, May 1998.

[18] P. A. Nelson, H. Hamada and S. J. Elliott, “Adaptive inverse filters for stereophonic sound reproduction,” *IEEE Trans. Signal Processing*, vol. 40, no. 7 pp. 1621-1632, July 1992.

[19] P. A. Nelson, F. Orduña-Bustamante and H. Hamada, “Inverse filter design and equalization zones in multichannel sound reproduction,” *IEEE Trans. Speech, Audio Processing*, vol. 3, no. 3, pp. 185-192, May 1995.

[20] P. Y. Wang, “3D sound reproduction via two loudspeakers,” *NCTU Master These*,

1997.

[21] F. A. Everest and R. Streicher, *The New Stereo Soundbook, McGraw-Hill*, 1992.

[22] W. S. Gan, S. E. Tan, M. H. Er and Y. K. Chong, “Elevated speaker projection for digital home entertainment system,” *IEEE Trans. Consumer Electron.*, vol. 47, no. 3, pp. 631-637, Aug. 2001.

[23] A. V. Oppenheim and R.W. Shafer, *Discrete Time Signal Processing, $2^{nd}$ ed.*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[24] S. Haykin, *Adaptive Filter Theory, $4^{th}$ ed.*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations, $2^{nd}$ ed*. Johns Hopkins, 1989.

[26] Begault, D. R., “*Challenges to the Successful Implementation of 3-D sound*”, *J. Audio Eng. Soc.*, vol. 39, pp. 895-904, 1990.

[27] B. Gardner and K. Martin, “*HRTF Measurement of a KEMAR Dummy-Head Microphone*”, *MIT Media Lab Perceptual Computing*.