Robust Equalization of Multichannel Acoustic Systems

Wancheng Zhang

A Thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy of Imperial College London

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London
2010

Abstract

In most real-world acoustical scenarios, speech signals captured by distant microphones from a source are reverberated due to multipath propagation, and the reverberation may impair speech intelligibility. Speech dereverberation can be achieved by equalizing the channels from the source to microphones. Equalization systems can be computed using estimates of multichannel acoustic impulse responses. However, the estimates obtained from system identification always include errors; the fact that an equalization system is able to equalize the estimated multichannel acoustic system does not mean that it is able to equalize the true system. The objective of this thesis is to propose and investigate robust equalization methods for multichannel acoustic systems in the presence of system identification errors.

Equalization systems can be computed using the multiple-input/output inverse theorem or multichannel least-squares method. However, equalization systems obtained from these methods are very sensitive to system identification errors. A study of the multichannel least-squares method with respect to two classes of characteristic channel zeros is conducted. Accordingly, a relaxed multichannel least-squares method is proposed. Channel shortening in connection with the multiple-input/output inverse theorem and the relaxed multichannel least-squares method is discussed.

Two algorithms taking into account the system identification errors are developed. Firstly, an optimally-stopped weighted conjugate gradient algorithm is

proposed. A conjugate gradient iterative method is employed to compute the equalization system. The iteration process is stopped optimally with respect to system identification errors. Secondly, a system-identification-error-robust equalization method exploring the use of error models is presented, which incorporates system identification error models in the weighted multichannel least-squares formulation.

Acknowledgment

I am very much obliged to my supervisor Dr. Patrick Naylor. Patrick has been providing me great guidance throughout my research work. This thesis would not have been completed without his insightful suggestions and feedback. It is truly an honor for me to have worked with him.

It was a pleasure to be a member of the Communications and Signal Processing Group at Imperial College London. The very diverse and interesting members stimulated an enjoyable atmosphere. I wish to express my special thanks to Dr. Emanuël Habets, Prof. Andy Khong, Dr. Nikolay Gaubitch and Dr. Jimi Wen for their comments and suggestions in my research.

I am very grateful to Dr. Pier Luigi Dragotti and Dr. Stephan Weiss for willing to take extra burden of reading and refereeing this thesis.

Finally, this thesis is as much of my effort as it is to my family. I am indebted to my parents for their endless love and support during my study. Thanks to my Jiaqi.

Contents

List of Figures

List of Tables

Glossary

Acronyms

BSI	blind system identification
CG	conjugate gradient
CR	cross-relation
CRLB	Cramér-Rao lower bound
CS	channel shortening
DRR	direct-to-reverberant ratio
EDC	energy decay curve
EIR	equalized impulse response
ELR	early-to-late reverberation ratio
FIR	finite impulse response
HOS	higher order statistics
IIR	infinite impulse response
LS	least-squares
MARDY	multichannel acoustic reverberation database at York
MCLS	multichannel least-squares
MINT	multiple-input/output inverse theorem
MSE	mean square error
MVU	minimum variance unbiased

NMCFLMS	normalized multichannel frequency-domain least-mean-squares
NMSE	normalized mean square error
NPM	normalized projection misalignment
OS-WCG	optimally-stopped weighted conjugate gradient
PRN	psuedo-random noise
RIR	room impulse response
RMCLS	relaxed multichannel least-squares
SIE	system identification error
SIEREM	system-identification-error-robust equalization method
SNR	signal-to-noise ratio
SOS	second order statistics
SSI	supervised system identification
TIR	target impulse response
WCG	weighted conjugate gradient
WDRR	weighted direct-to-reverberant ratio
WLS	weighted least-squares
WMCLS	weighted multichannel least-squares
WMSE	weighted mean square error

Operators

$\{\cdot\}^T$	transpose
$*$	linear convolution
$\mathrm{E}\{\cdot\}$	expectation operator
$\mathrm{var}(\cdot)$	variance operator
$\|\cdot\|_2$	ℓ_2-norm
$\{\cdot\}^+$	Moore-Penrose pseudo-inverse
$\lceil\cdot\rceil$	ceil operator
$\dim(\cdot)$	dimension of linear space
$\mathrm{null}(\cdot)$	null space of matrix
$\setminus$	exclusion operator

Notations

x	scalar quantity
$x(n)$	function of discrete variable n
$x(i)$	function of finite discrete variable i
$\mathbf{x}$	vector formed by stacking $x(i)$
$\mathbf{x}(n)$	vector formed by stacking a segment of $x(n)$ till n
$X(z)$	z-transform of $x(i)$
$X(e^{j\omega})$	transfer function of $x(i)$
$\hat{x}$	estimate of x
$\mathrm{diag}\{\mathbf{x}\}$	diagonal matrix having $x(i)$ on diagonal
$\mathbf{Y}$	matrix quantity

Symbols and Variables

$b(i)$	equalized impulse response
$\tilde{b}(i)$	equivalent single-channel least-squares inverse filter relating to common zeros
$b'(i)$	scaled equalized impulse response
C_{50}	early-to-late reverberation ratio
$\mathbf{C}_e$	covariance matrix of system identification error
$d(i)$	(delayed) delta function
$e_m(i)$	mth-channel system identification error function
$\tilde{e}_m(i)$	generated substitute of error, error model
$\mathbf{e}$	error vector, projection error vector
$\mathbf{e}_p$	parallel component of projection error vector
$\mathbf{e}_v$	normal component of projection error vector
$\mathbf{E}$	multichannel convolution matrix formed by error
$\tilde{\mathbf{E}}$	multichannel convolution matrix formed by error model
$\mathscr{E}(i)$	energy decay curve
f_s	sampling frequency
$\mathbf{f}_m$	cross-correlation vector of source signal and noise at mth-microphone
$\mathbf{f}_m(N)$	cross-correlation vector of source signal and noise at mth-microphone relating to data sequences of length N
$g_m(i)$	impulse response of mth-channel component of equalization system
$g_0(i)$	single-channel least-squares inverse filter relating to common zeros
$\mathbf{g}$	equalization system
$\mathbf{g}[k]$	equalization system after the kth iteration
$h_m(i)$	mth-channel impulse response
$\hat{h}_m(i)$	estimate of mth-channel impulse response

$h_0(i)$	impulse response relating to common zeros
$\tilde{h}_m(i)$	mth-channel impulse response excluding common zeros
$\mathbf{h}$	acoustic system
$\hat{\mathbf{h}}$	estimated acoustic system
$H(\omega)$	sum of squared multichannel magnitude responses
$\mathbf{H}_m$	mth-channel convolution matrix
$\mathbf{H}_0$	convolution matrix relating to common zeros
$\mathbf{H}$	multichannel convolution matrix
$\mathbf{I}$	identity matrix
J	cost function
k	index of iteration
k^{opt}	peak index of weighted direct-to-reverberant ratio
L	channel length
L_c	critical length of equalization system components for existence of inverse system
L_i	length of equalization system components
L_w	length of relaxed window, target length for channel shortening
m	index of channel
M	number of microphones
n_e :	transition index from early reflections to late reflections
N	length of data sequence
N_0	number of common zeros
$\mathcal{N}(\cdot)$	Gaussian distribution
pi	circumference ratio
P	speech power level
r	modulus of zeros

$\mathbf{r}_m(N)$	cross-correlation vector of source signal and mth-microphone signal relating to data sequences of length N
R_a	autocorrelation coefficient
R_c	cross-correlation coefficient
$\mathbf{R}_{xy}$	correlation matrix of signal x and y
$\mathbf{R}_{xy}(N)$	correlation matrix of signal x and y relating to data sequences of length N
$s(n)$	source signal
t_0	time delay in milliseconds
T_{60}	reverberation time
T_{30}	time interval for energy decay curve to drop by 30 dB
$u(i)$	weighting function
$v_m(n)$	noise at the mth microphone
$w(i)$	weighting function
$\mathbf{W}$	diagonal matrix having $w(i)$ on diagonal
$x_m(n)$	mth-channel output
$y_m(n)$	mth-microphone signal
$z_{m,k}$	the kth zero of the mth channel
α	decay rate of reverberation level
β	multiplicative factor relating to level of error
γ	scaling factor relating to blind system identification
δ	near-common-zero vicinity character
Δ	slope of energy decay curve
$\varepsilon_m(i)$	white sequence, white Gaussian sequence
$\varepsilon_{ml}(n)$	error signal between mth and lth channel
$\zeta(i)$	zero-mean stationary Gaussian noise
θ	angle relating to misalignment, angle of zeros

ι	square matrix size
λ	generalized eigenvalue
σ^2	variance of noise
σ_s^2	variance of source signal
τ	integer delay
$\mathbf{0}$	vector containing zeros
$\mathbf{1}$	vector containing ones

Chapter 1

Introduction

1.1 Context of work

Speech communication is a fast growing industry and is closing the distance between people due to advances in technology, decreasing costs and increasing capabilities. The ideal case scenario is to be able to speak to target audience at a different place from you while your voice sounds just like you were sitting in front of the target audience. However, there are many types of degradation along a signal path that impair our ability to decipher voice transmissions, such as background noise, reverberation, and other interferences.

In many hands-free devices such as mobile phones, PDAs, voice over IP, hearing aids or teleconferencing equipments, where the microphone or array of microphones is not placed close to the mouth of the speaker, the reflected paths cannot be neglected compared with the direct sound. In these applications, the reflected paths or reverberation in the talker's environment can be an important factor in degrading the overall speech quality as perceived at the listener's end.

Reverberation is the process of multipath propagation of an acoustic signal from its source to the microphone as shown in Figure 1.1. The received signal

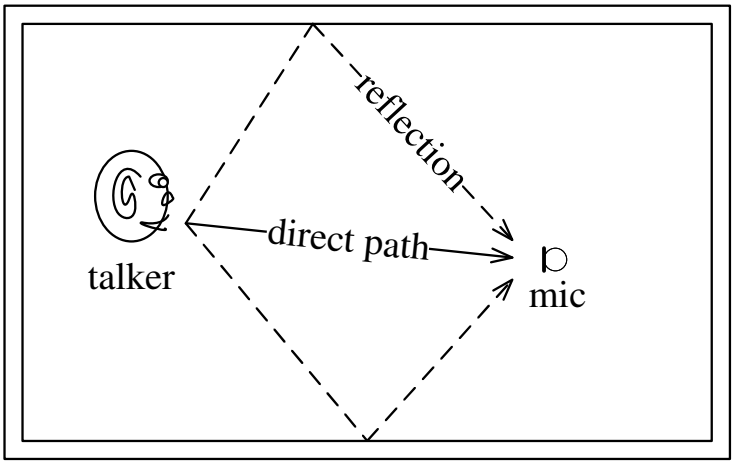

Figure 1.1: Reverberation in a room.

generally consists of a direct sound, reflections that arrive shortly after the direct sound (commonly called early reflections), and reflections that arrive after the early reflections (commonly called late reflections). The effect of reverberation on speech is to cause it to sound distant and spectrally modified which can reduce naturalness and intelligibility. Early reflections are not perceived as separate sound events but instead cause a spectral distortion called colouration. Late reverberation often forms a background ambience which is distinct from the foreground sound and may impair speech intelligibility.

The problem of reverberation can be resolved by utilizing a headset, where the microphone is kept close to the mouth. Nevertheless, this imposes restrictions on the flexibility and comfort of the user, which are the main desired features in the use of the aforementioned hands-free devices. Therefore, signal processing approaches enhancing the reverberant speech are desired.

Recent research has produced various algorithms for speech dereverberation, which can be divided broadly into three main categories:

1. Spatial processing - the signals received at the different microphones are delayed, weighted and summed, so as to form a beam in the direction of the

desired source and to attenuate sounds from other directions. Contributions falling in this category can be found in [1, 2, 3].

2. Speech enhancement - the reverberant speech signal is modified so that some features of it are closer to those of the clean speech signal according to *a priori* models of the speech waveform or spectrum. Contributions can be found in [4, 5, 6, 7].

3. Acoustic system equalization - the inverse system of either a single-channel or multichannel acoustic system is estimated, where an acoustic channel refers to the multiple propagation paths from the source to a microphone and an acoustic system refers to the single or multiple channels. The inverse system is either estimated directly from the reverberant speech signals [8,9,10] or from the estimates of channel impulse responses obtained from system identification [11, 12, 13, 14].

1.2 Research aim and thesis structure

The objective of the research presented in this thesis is to make hands-free speech sound as similar as possible to that of a closely located microphone. The focus of this research is on equalization of multichannel acoustic systems robust to errors included in the estimates of channel impulse responses, which are referred to as system identification errors (SIEs).

Acoustic channels are usually modeled as finite impulse response (FIR) filters. Since a single acoustic channel is generally nonminimum phase [15], its infinite impulse response (IIR) causal inverse is an unstable system which is not useful in practice. When multiple microphones are employed, the multichannel acoustic system can be exactly inverted by a set of FIR filters, which is referred to as multichan-

nel inverse system, using the multiple-input/output inverse theorem (MINT) [12]. However, the estimates of channel impulse responses always include SIEs and the multichannel inverse system is very sensitive to the SIEs, as will be shown later in this thesis. The fact that the multichannel inverse system is able to equalize the estimated acoustic system does not mean that it is able to equalize the true acoustic system, where an estimated acoustic system refers to the system formed by the estimates of multichannel impulse responses. When the inverse system of the estimated acoustic system is employed to equalize the true acoustic system, the impulse response from the source to the output of the inverse system will deviate from the delta function due to the aforementioned SIEs. Since in general the inverse of the multichannel acoustic system cannot be obtained based on the estimated system, we will use the term 'equalization system' rather than the more strict 'inverse system'. In this thesis, our aim is to find robust equalization system design methods, using which a multichannel equalization system computed based on the estimated multichannel system can equalize the true system, resulting in reduced reverberation and improved intelligibility.

The remaining chapters of this thesis are organized as follows.

Chapter 2 provides a review of system identification and equalization literature and serves as the technical foundation of this thesis. Firstly, a formulation of identification and equalization of multichannel acoustic systems is provided. Then, both the supervised (non-blind) system identification (SSI) and blind system identification (BSI) techniques are reviewed and performance measures for them are presented. Next, system equalization techniques including MINT, least-squares (LS), weighted least-squares (WLS), and channel shortening (CS) are reviewed. Some of them were traditionally applied in the single-channel scenarios or with particular restrictions. These restrictions and new problems that arise when they are applied to multichannel systems are discussed. After

this, characteristics of room acoustics and psychoacoustics are presented. This is helpful for understanding the characteristics of acoustic system equalization. Accordingly, performance measures used through this thesis for evaluation of equalization algorithms are defined. Finally, the problem of equalization in the presence of SIEs is discussed. Simulation examples are presented and characteristics of acoustic system equalization are summarized.

Chapter 3 provides some new insights into the multichannel least-squares (MCLS) method from the point of view of channel zeros, and presents a relaxed multichannel least-squares (RMCLS) method. The performance of MCLS when common zeros among multiple channels are present is shown by an experiment. Next, two classes of characteristic zeros causing strong peaks in the frequency responses of the filters of MCLS equalization system, which lead to the high sensitivity of MCLS to SIEs, are defined. The performance of RMCLS with respect to these two classes of characteristic zeros is studied.

Chapter 4 investigates the use of channel shortening technique in equalization of acoustic systems. Firstly a mathematical link between the MINT and the traditional CS is derived. Next, a criterion for developing a perceptually advantageous equalization system from the multiple solutions to CS is provided. In multichannel scenarios, the CS can provide multiple solutions but not all of them are useful in terms of speech perception.

Chapter 5 investigates the use of conjugate gradient (CG) iterative methods for the equalization of acoustic systems, the channel estimates of which are obtained from SSI. An optimally-stopped weighted conjugate gradient (OS-WCG) algorithm is presented. In the presence of SSI errors, firstly a peak of the weighted direct-to-reverberant ratio (WDRR) in the iterative process is defined. Next, a method to estimate the iteration index of the peak is pro-

vided. Then, a condition for stopping the iteration is proposed. Finally, the OS-WCG is evaluated.

Chapter 6 presents a system-identification-error-robust equalization method (SIEREM) which uses models of SSI error and BSI error. An SSI error model is obtained using information about SSI available in literature. An experimental study of BSI error is conducted and a BSI error model is developed. Then, the SIEREM which incorporate the error models in equalization formulation is derived. Finally, SIEREM is evaluated.

Chapter 7 summarizes the work presented in this thesis and provides a comparative summary of the acoustic system equalization algorithms developed in this thesis. Finally, the thesis is concluded with guidelines for further developments of the herein presented ideas.

1.3 Scope and original contributions

To the best knowledge of the author, the following aspects of this thesis are believed to be original contributions:

1. Derivation of an estimator (2.54) for normalized mean square error (NMSE). (Chapter 2, Section 2.2.3)

2. Derivation of mean power of distortion in the equalized impulse response (2.81) in relation to mean squared ℓ_2-norm of equalization systems. (Chapter 2, Section 2.5)

3. Study of MCLS with respect to common zeros and characteristic zeros. (Chapter 3, Section 3.1.1, Section 3.2.2)

4. Development of RMCLS, and study of RMCLS with respect to common zeros and characteristic zeros. (Chapter 3, Section 3.1.2, Section 3.2.3)

5. Derivation of a mathematical link between MINT and CS. (Chapter 4, Section 4.1)

6. Work that provides a criterion for selecting a perceptually advantageous equalization system from the multiple solutions to CS. (Chapter 4, Section 4.2)

7. Development and evaluation of OS-WCG algorithm. (Chapter 5)

8. Study of BSI error. (Chapter 6, Section 6.1.2)

9. Development of an algorithm to generate representations of BSI errors. (Chapter 6, Section 6.3.1)

10. Derivation and evaluation of SIEREM. (Chapter 6, Section 6.2, Section 6.3)

Chapter 2

Literature Review

2.1 Formulation of identification and equalization of acoustic systems

Consider a source signal $s(n)$ propagating through an M-channel acoustic system $\mathbf{h} = [\mathbf{h}_1^T \ \cdots \ \mathbf{h}_m^T \ \cdots \ \mathbf{h}_M^T]^T$, where n denotes the discrete time index, as illustrated in Figure 2.1. The acoustic channel between the source and the mth microphone is characterized by its impulse response $\mathbf{h}_m = [h_m(0) \ h_m(1) \ \ldots \ h_m(i) \ \ldots \ h_m(L-1)]^T$, $m = 1, \ldots, M$, where $\{\cdot\}^T$ denotes the transpose operation. In our work we assume that the acoustic channels are finite in length and time-invariant.

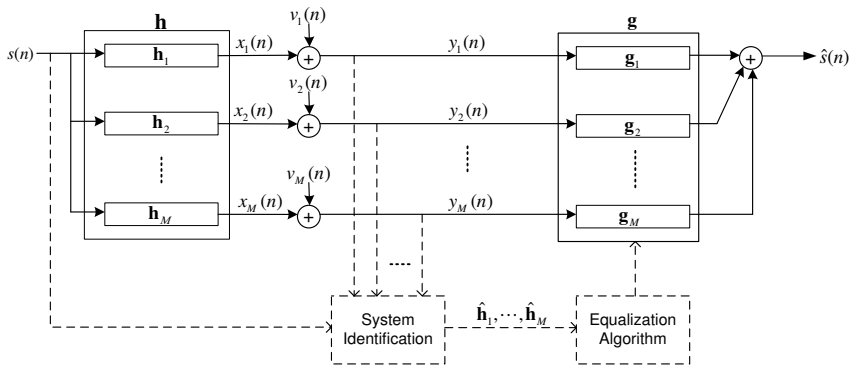

Figure 2.1: Illustration of identification and equalization of an acoustic system.

The impulse responses $\mathbf{h}_m$ can be identified blindly or non-blindly. In supervised (non-blind) system identification (SSI) [16,17,18], using the source signal $s(n)$ and the reverberant signals

$$y_m(n) = x_m(n) + v_m(n) \quad \text{for } m = 1, \ldots, M, \tag{2.1}$$

where

$$x_m(n) = s(n) * h_m(n), \tag{2.2}$$

estimates of the room impulse responses (RIRs) $\mathbf{h}_m$ can be obtained, where $*$ denotes linear convolution and $v_m(n)$ denotes additive noise at the mth microphone. In blind system identification (BSI) [8,13,19,20,21,22,23,24,25,26,27,28,29], the estimates $\hat{\mathbf{h}}_m = [\hat{h}_m(0) \ \ldots \ \hat{h}_m(i) \ \ldots \ \hat{h}_m(L-1)]^T$ can be obtained using only the reverberant signals $y_m(n)$.

Generally speaking, an equalization system $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \cdots \ \mathbf{g}_M^T]^T$, where $\mathbf{g}_m = [g_m(0) \ g_m(1) \ \ldots \ g_m(i) \ \ldots \ g_m(L_i - 1)]^T$ is the mth-channel component, can be computed using the estimated system $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T \ \cdots \ \hat{\mathbf{h}}_m^T \ \cdots \ \hat{\mathbf{h}}_M^T]^T$. The $\hat{\mathbf{h}}$ includes some error due to finite data, the existence of the additive noise, and under-/over-modeling of channel order. As a result, the response from the source to the output of the equalization system may still distort the speech signal severely due to the SIEs. In this thesis, as is common practice in the current literature, we assume the channel orders are known or can be correctly estimated.

For the sake of discussion, we define the following vectors:

$$\mathbf{s}(n) = [s(n) \ s(n-1) \ \ldots \ s(n-L+1)]^T, \tag{2.3}$$

$$\mathbf{x}_m(n) = [x_m(n) \ x_m(n-1) \ \ldots \ x_m(n-L+1)]^T, \tag{2.4}$$

$$\mathbf{v}_m(n) = [v_m(n) \ v_m(n-1) \ \ldots \ v_m(n-L+1)]^T, \tag{2.5}$$

$$\mathbf{y}_m(n) = [y_m(n) \ y_m(n-1) \ \ldots \ y_m(n-L+1)]^T. \tag{2.6}$$

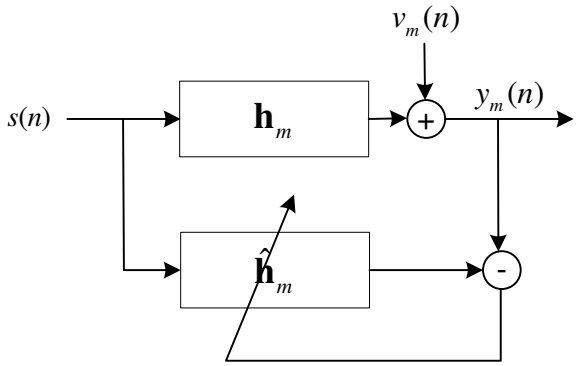

Figure 2.2: Schematic of supervised system identification.

2.2 System identification

2.2.1 Supervised system identification

Figure 2.2 depicts the schematic of supervised system identification. For the mth channel, the cost function under the least-squares criterion is [16]

$$J = \frac{1}{N}\sum_{n=0}^{N-1}(y_m(n) - \mathbf{s}(n)^T\hat{\mathbf{h}}_m)^2. \tag{2.7}$$

Introducing

$$\mathbf{R}_{ss}(N) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{s}(n)\mathbf{s}^T(n) \tag{2.8}$$

and

$$\mathbf{r}_m(N) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{s}(n)y_m(n), \tag{2.9}$$

the estimate $\hat{\mathbf{h}}_m$ minimizing (2.7) can be obtained with, provided the indicated inverse exists,

$$\hat{\mathbf{h}}_m = \mathbf{R}_{ss}^{-1}(N)\mathbf{r}_m(N). \tag{2.10}$$

In the case that all channels are driven by the same input signal $s(n)$, in matrix form the channel estimates can be obtained from

$$[\hat{\mathbf{h}}_1 \ \hat{\mathbf{h}}_2 \ \cdots \ \hat{\mathbf{h}}_M] = \mathbf{R}_{ss}^{-1}(N)[\mathbf{r}_1(N) \ \mathbf{r}_2(N) \ \cdots \ \mathbf{r}_M(N)]. \tag{2.11}$$

Channel identifiability

Introducing

$$\mathbf{f}_m(N) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{s}(n)v_m(n), \tag{2.12}$$

it is found that

$$\begin{aligned} \lim_{N\to\infty} \mathbf{h}_m - \hat{\mathbf{h}}_m &= -\lim_{N\to\infty} \mathbf{R}_{ss}^{-1}(N)\mathbf{f}_m(N) \\ &= -\mathbf{R}_{ss}^{-1}\mathbf{f}_m, \end{aligned} \tag{2.13}$$

where

$$\mathbf{R}_{ss} = \mathrm{E}\{\mathbf{s}(n)\mathbf{s}^T(n)\} \tag{2.14}$$

and

$$\mathbf{f}_m = \mathrm{E}\{\mathbf{s}(n)v_m(n)\} \tag{2.15}$$

provided $s(n)$ and $v_m(n)$ are quasi-stationary, where $\mathrm{E}\{\cdot\}$ denotes expectation operator. For $\hat{\mathbf{h}}_m$ to converge to $\mathbf{h}_m$, it is required that [16]

1. $\mathbf{R}_{ss}$ is non-singular.
2. $\mathbf{f}_m = 0$. This will be the case if either:
   - $v_m(n)$ is a sequence of independent random variables with zero mean values (white noise).
   - $s(n)$ is independent of the zero mean sequence $v_m(n)$.

Supervised system identification error

It is found that if $v_m(n)$ is white Gaussian noise, (2.10) is a minimum variance unbiased (MVU) estimator [17]. Assuming the variance of $v_m(n)$ is σ^2, the covariance matrix of

$$\mathbf{e}_m = [e_m(0) \ \dots \ e_m(i) \ \dots \ e_m(L-1)]^T \tag{2.16}$$

$$= \mathbf{h}_m - \hat{\mathbf{h}}_m \tag{2.17}$$

is [17]

$$\mathbf{C}_e = \frac{\sigma^2}{N}\mathbf{R}_{ss}^{-1}(N), \tag{2.18}$$

and

$$\mathbf{e}_m \sim \mathcal{N}(\mathbf{0}_{L\times 1}, \mathbf{C}_e), \tag{2.19}$$

where $\mathcal{N}(\cdot)$ denotes Gaussian distribution.

When pseudo-random noise (PRN) [18,30] is used as the probing signal $s(n)$ [17], it is approximately realized that

$$\mathbf{R}_{ss}(N) = \sigma_s^2\mathbf{I}, \tag{2.20}$$

where $\mathbf{I}$ is identity matrix. Hence, the variance of the error approximates

$$\mathrm{var}(e_m(i)) = \frac{\sigma^2}{N\sigma_s^2}, \tag{2.21}$$

where σ_s^2 is the variance of $s(n)$ and $\mathrm{var}(\cdot)$ denotes variance operator.

2.2.2 Blind identification of multichannel systems

Blind system identification uses only the microphone signals $y_m(n)$ to estimate the system. Single-channel BSI needs to use higher order statistics (HOS) of the micro-

phone signal [19], whereas multichannel systems can be blindly identified using only second order statistics (SOS) [25]. A multichannel system can be identified using the cross-relation (CR) method [23]. The CR between two channels is expressed as

$$x_m(n)*h_l(n) = s(n)*h_m(n)*h_l(n) = x_l(n)*h_m(n),\ m = 1, 2, \ldots, M, l \neq m. \quad (2.22)$$

In the presence of noise, an error can be formed

$$\varepsilon_{ml}(n) = y_m(n) * h_l(n) - y_l(n) * h_m(n). \quad (2.23)$$

A cost function is formed [23]

$$J = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=1}^{M-1} \sum_{l=m+1}^{M} \varepsilon_{ml}^2(n). \quad (2.24)$$

Introducing

$$\mathbf{R}_{yy}(N) = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}y_m y_m(N) & -\mathbf{R}y_2 y_1(N) & \cdots & -\mathbf{R}y_M y_1(N) \\ -\mathbf{R}y_1 y_2(N) & \sum_{m\neq 2} \mathbf{R}y_m y_m(N) & \cdots & -\mathbf{R}y_M y_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}y_1 y_M(N) & -\mathbf{R}y_2 y_M(N) & \cdots & \sum_{m\neq M} \mathbf{R}y_m y_m(N) \end{bmatrix} \quad (2.25)$$

with

$$\mathbf{R}y_m y_l(N) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{y}_m(n)\mathbf{y}_l^T(n), \quad (2.26)$$

the estimate $\hat{\mathbf{h}}$ minimizing (2.24) would be the eigenvector of $\mathbf{R}_{yy}(N)$ corresponding to its smallest eigenvalue [23]. If the multichannel system is blindly identifiable, $\mathbf{h}$ can be uniquely determined up to a scaling factor.

Introducing

$$\mathbf{R}_{xx}(N) = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}x_m x_m(N) & -\mathbf{R}x_2 x_1(N) & \cdots & -\mathbf{R}x_M x_1(N) \\ -\mathbf{R}x_1 x_2(N) & \sum_{m\neq 2} \mathbf{R}x_m x_m(N) & \cdots & -\mathbf{R}x_M x_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}x_1 x_M(N) & -\mathbf{R}x_2 x_M(N) & \cdots & \sum_{m\neq M} \mathbf{R}x_m x_m(N) \end{bmatrix}, \tag{2.27}$$

$$\mathbf{R}_{xv}(N) = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}x_m v_m(N) & -\mathbf{R}x_2 v_1(N) & \cdots & -\mathbf{R}x_M v_1(N) \\ -\mathbf{R}x_1 v_2(N) & \sum_{m\neq 2} \mathbf{R}x_m v_m(N) & \cdots & -\mathbf{R}x_M v_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}x_1 v_M(N) & -\mathbf{R}x_2 v_M(N) & \cdots & \sum_{m\neq M} \mathbf{R}x_m v_m(N) \end{bmatrix} \tag{2.28}$$

and

$$\mathbf{R}_{vv}(N) = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}v_m v_m(N) & -\mathbf{R}v_2 v_1(N) & \cdots & -\mathbf{R}v_M v_1(N) \\ -\mathbf{R}v_1 v_2(N) & \sum_{m\neq 2} \mathbf{R}v_m v_m(N) & \cdots & -\mathbf{R}v_M v_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}v_1 v_M(N) & -\mathbf{R}v_2 v_M(N) & \cdots & \sum_{m\neq M} \mathbf{R}v_m v_m(N) \end{bmatrix}, \tag{2.29}$$

with

$$\mathbf{R}x_m x_l(N) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_m(n)\mathbf{x}_l^T(n), \tag{2.30}$$

$$\mathbf{R}x_m v_l(N) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_m(n)\mathbf{v}_l^T(n) \tag{2.31}$$

and

$$\mathbf{R}v_m v_l(N) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{v}_m(n)\mathbf{v}_l^T(n), \tag{2.32}$$

and using (2.1), (2.25) can be simply expressed as

$$\mathbf{R}_{yy}(N) = \mathbf{R}_{xx}(N) + \mathbf{R}_{xv}(N) + \mathbf{R}_{xv}^T(N) + \mathbf{R}_{vv}(N). \tag{2.33}$$

Channel identifiability

Channel identifiability is concerned with the existence of a unique solution to the unknown channel impulse responses with respect to BSI techniques using SOS. The identifiability conditions are summarized as follows:

1. Channel diversity [23]: the use of multichannel techniques introduces channel diversity and enables the exploration of SOS of the system outputs for blind identification of multichannel systems. Channel diversity in this context refers to channels being coprime, that is, multichannel transfer functions do not share any common zero. If one or more common zeros exist across all channels then these channels are not coprime. When common zeros exist BSI techniques cannot distinguish the common zeros due to the unknown system from ones due to the source signal.

2. Condition for the input signals [23]: the autocorrelation matrix of the input signal $\mathbf{R}_{s's'} = \mathrm{E}\{\mathbf{s}'\mathbf{s}'^T\}$, where $\mathbf{s}'(n) = [s(n)\ s(n-1)\ \ldots\ s(n-2L+2)]^T$, is of full rank (such that the multichannel system can be fully exited).

3. It is found that

$$\lim_{N\to\infty} \mathbf{R}_{yy}(N) = \mathbf{R}_{yy} \tag{2.34}$$

$$= \mathbf{R}_{xx} + \mathbf{R}_{xv} + \mathbf{R}_{xv}^T + \mathbf{R}_{vv}, \tag{2.35}$$

where

$$\mathbf{R}_{yy} = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}_{y_m y_m} & -\mathbf{R}_{y_2 y_1} & \cdots & -\mathbf{R}_{y_M y_1} \\ -\mathbf{R}_{y_1 y_2} & \sum_{m\neq 2} \mathbf{R}_{y_m y_m} & \cdots & -\mathbf{R}_{y_M y_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{y_1 y_M} & -\mathbf{R}_{y_2 y_M} & \cdots & \sum_{m\neq M} \mathbf{R}_{y_m y_m} \end{bmatrix}, \tag{2.36}$$

$$\mathbf{R}_{xx} = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}_{x_m x_m} & -\mathbf{R}_{x_2 x_1} & \cdots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{m\neq 2} \mathbf{R}_{x_m x_m} & \cdots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_M} & -\mathbf{R}_{x_2 x_M} & \cdots & \sum_{m\neq M} \mathbf{R}_{x_m x_m} \end{bmatrix}, \tag{2.37}$$

$$\mathbf{R}_{xv} = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}_{x_m v_m} & -\mathbf{R}_{x_2 v_1} & \cdots & -\mathbf{R}_{x_M v_1} \\ -\mathbf{R}_{x_1 v_2} & \sum_{m\neq 2} \mathbf{R}_{x_m v_m} & \cdots & -\mathbf{R}_{x_M v_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 v_M} & -\mathbf{R}_{x_2 v_M} & \cdots & \sum_{m\neq M} \mathbf{R}_{x_m v_m} \end{bmatrix} \tag{2.38}$$

and

$$\mathbf{R}_{vv} = \begin{bmatrix} \sum_{m\neq 1} \mathbf{R}_{v_m v_m} & -\mathbf{R}_{v_2 v_1} & \cdots & -\mathbf{R}_{v_M v_1} \\ -\mathbf{R}_{v_1 v_2} & \sum_{m\neq 2} \mathbf{R}_{v_m v_m} & \cdots & -\mathbf{R}_{v_M v_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{v_1 v_M} & -\mathbf{R}_{v_2 v_M} & \cdots & \sum_{m\neq M} \mathbf{R}_{v_m v_m} \end{bmatrix} \tag{2.39}$$

with

$$\mathbf{R}_{y_m y_l} = \mathrm{E}\{\mathbf{y}_m(n)\mathbf{y}_l^T(n)\}, \tag{2.40}$$

$$\mathbf{R}_{x_m x_l} = \mathrm{E}\{\mathbf{x}_m(n)\mathbf{x}_l^T(n)\}, \tag{2.41}$$

$$\mathbf{R}_{x_m v_l} = \mathrm{E}\{\mathbf{x}_m(n)\mathbf{v}_l^T(n)\} \tag{2.42}$$

and

$$\mathbf{R}_{v_m v_l} = \mathrm{E}\{\mathbf{v}_m(n)\mathbf{v}_l^T(n)\}. \tag{2.43}$$

From (2.22), we know that

$$\mathbf{R}_{xx}\mathbf{h} = \mathbf{0}_{ML\times 1}. \tag{2.44}$$

Therefore, for $\hat{\mathbf{h}}$ to converge to $\mathbf{h}$, it is required that $v_m(n)$ is white noise,

incoherent or uncorrelated [31], so that $\mathbf{R}_{yy}$ can be expressed as the sum of $\mathbf{R}_{xx}$ and a scaled identity matrix, and the eigenvector of $\mathbf{R}_{yy}$ corresponding to the smallest eigenvalue is in the 1-dimension null space of $\mathbf{R}_{xx}$.

2.2.3 Performance measures for system identification

Mean square error (MSE) is a frequently used measure of the differences between the estimate $\hat{\mathbf{h}}_m$ and the true RIR $\mathbf{h}_m$ for SSI, which is defined as [17]

$$\mathrm{MSE}(\hat{h}_m(i)) = \mathrm{E}\{(h_m(i) - \hat{h}_m(i))^2\}. \tag{2.45}$$

In our work, to measure the relative level of the error with respect to the true RIR, we use normalized mean square error (NMSE)

$$\mathrm{NMSE} = \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}, \tag{2.46}$$

where 'mean' refers to time average rather than the ensemble average in the definition of MSE in (2.45), and $\|\cdot\|_2$ denotes ℓ_2-norm.

As for BSI, since scaling ambiguity is always introduced in $\hat{\mathbf{h}}$, using the NMSE can produce misleading results. A widely used measure for BSI is the normalized projection misalignment (NPM) [32]

$$\mathrm{NPM} = \min_{\gamma} \frac{\|\mathbf{h} - \gamma\hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2}, \tag{2.47}$$

where the minimum is attained when

$$\gamma = \frac{\hat{\mathbf{h}}^T\mathbf{h}}{\hat{\mathbf{h}}^T\hat{\mathbf{h}}}. \tag{2.48}$$

The geometric meaning of NPM is shown in Figure 2.3. In the remainder of this

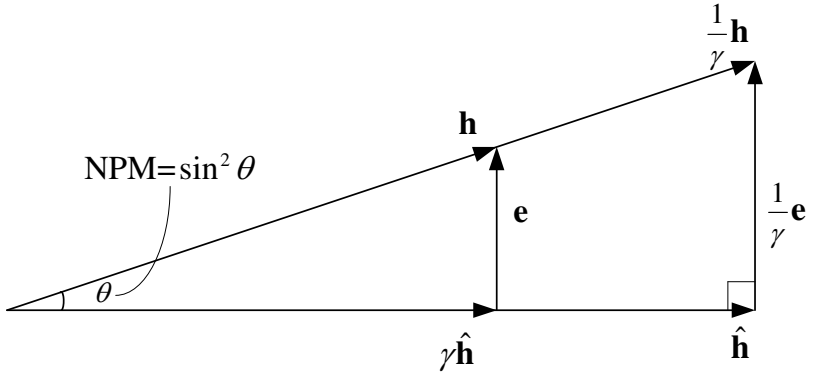

Figure 2.3: Illustration of misalignment and scaling ambiguity introduced by BSI.

thesis, the vector

$$\mathbf{e} = \mathbf{h} - \gamma\hat{\mathbf{h}} \tag{2.49}$$

shown in Figure 2.3 will be referred to as projection error vector, where $\mathbf{e} = [\mathbf{e}_1^T \ \cdots \ \mathbf{e}_m^T \ \cdots \ \mathbf{e}_M^T]^T$ with

$$\begin{aligned} \mathbf{e}_m &= [e_m(0) \ \ldots \ e_m(i) \ \ldots \ e_m(L-1)]^T & (2.50)\\ &= \mathbf{h}_m - \gamma\hat{\mathbf{h}}_m. & (2.51) \end{aligned}$$

Estimation of NMSE and NPM

Since the SSI error sequence $e_m(i) = h_m(i) - \hat{h}_m(i)$ is a white sequence, as presented in Section 2.2.1, and the error sequence is very long, which is usually thousands of taps, it can be assumed that

$$\|\mathbf{e}_m\|_2^2 = L \cdot \mathrm{var}(e_m(i)). \tag{2.52}$$

On the other hand, assuming the probing signal $s(n)$ is white noise, we have [33]

$$\sum_{n=0}^{N-1}(s(n) * h_m(n))^2 = N\sigma_s^2\|\mathbf{h}_m\|_2^2. \tag{2.53}$$

Assuming (2.20) is true, using (2.21), (2.46), (2.52) and (2.53), the NMSE of the SSI error can be estimated using

$$\widehat{\mathrm{NMSE}} = \frac{L}{N \cdot \mathrm{SNR}}, \tag{2.54}$$

where SNR (the signal-to-noise ratio) is defined as

$$\mathrm{SNR} = \frac{\sum_{n=0}^{N-1}(s(n) * h_m(n))^2}{N\sigma^2}. \tag{2.55}$$

As for BSI, an asymptotic variance of the CR method is derived and is compared with its Cramér-Rao Lower Bound (CRLB) in [34]. However, the CRLB derived in [34] corresponds to a normalization different from the normalization which leads to the NPM, and cannot be applied to the NPM [32]. What is more, the computing of either the variance or the CRLB uses the source signal and the RIRs, which are unknown in the blind scenarios. The blind estimation of NPM is still an open question.

2.2.4 Simulation examples

In this section, simulation examples for both SSI and BSI are presented. In this thesis, without loss of generality, we always assume that the direct-path propagation time l_1 from the source to the 1st microphone of the microphone array is the shortest. The propagation time l_1 is trimmed for all channels in all experiments in this thesis.[1]

In the first example, a 2-channel system, the RIRs of which are from the MARDY database [35], is identified using the SSI. The length of the channels is truncated to $L = 2000$ corresponding to 0.25 s with a sampling frequency of $f_s = 8$ kHz. The channels are driven by white Gaussian noise. The SNR is set to 20 dB.

[1]The propagation time l_1 represents only a bulk propagation delay and is not significant to either the system identification or the equalization system design.

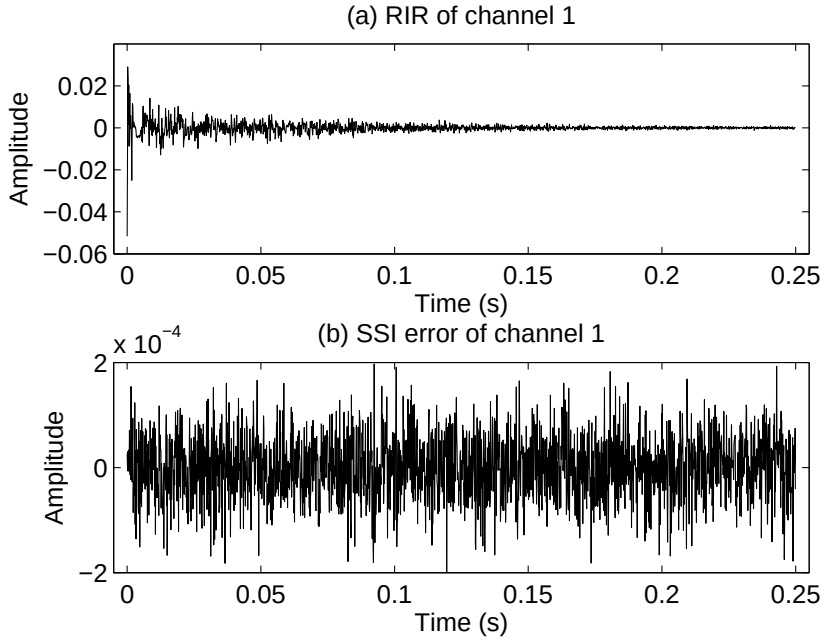

Figure 2.4: The RIR of channel 1 and the SSI error of channel 1.

$N = 40000$ samples are used. The RIR $\mathbf{h}_1$ and the error vector $\mathbf{e}_1 = \mathbf{h}_1 - \hat{\mathbf{h}}_1$ are shown in Figure 2.4. The error vector $\mathbf{e}_2$ has similar shape as $\mathbf{e}_1$ and is therefore omitted. It can be seen that the error sequence resembles a white Gaussian sequence. The NMSE is -32.4 dB and -32.6 dB for $\mathbf{e}_1$ and $\mathbf{e}_2$ respectively. The $\widehat{\text{NMSE}}$ computed from (2.54), where it is assumed that (2.20) is true, is -33.0 dB for both channels. The true NMSE is very close to the estimate of it, which shows that (2.54) is a good estimator of the NMSE.

In the second example, a 6-channel system is identified using normalized multichannel frequency-domain least-mean-squares (NMCFLMS) algorithm [28], which is an adaptive approach of BSI based on the CR method. The RIRs are generated using the image method [36]. The room dimensions are 6.4 m × 5 m × 3.6 m (length × width × height), the distance between the speaker and the center of the microphone array is set to 1 m, the inter-microphone distance is 5 cm, the reverberation time T_{60} [37], which will be elaborated in Section 2.4, is set to 0.6 s and the

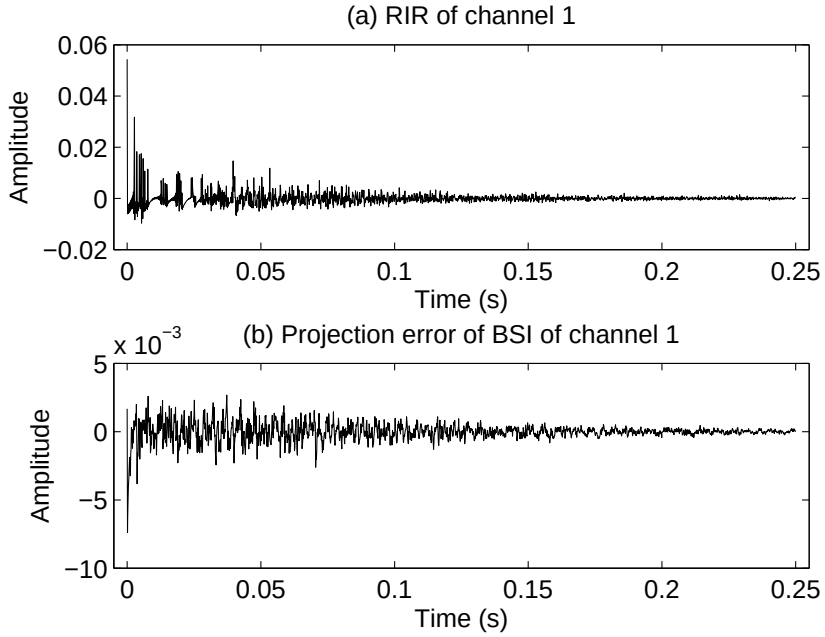

Figure 2.5: The RIR of channel 1 and the projection error of channel 1.

SNR is set to 25 dB. The length of the RIRs used for the experiment are $L = 2000$. White Gaussian noise is used as the source signal. The RIR $\mathbf{h}_1$ and the error vector $\mathbf{e}_1 = \mathbf{h}_1 - \gamma\hat{\mathbf{h}}_1$ are shown in Figure 2.5. The error vectors $\mathbf{e}_m$ for $m \in \{2, \ldots, M\}$ have similar overall temporal shape as $\mathbf{e}_1$ and are therefore omitted. It can be seen that the error sequence is damping with time. The NPM between $\hat{\mathbf{h}}$ and $\mathbf{h}$ is -10.0 dB.

2.3 System equalization

In general, for a given multichannel system $\mathbf{h}$, an equalization system $\mathbf{g}$ can be computed that satisfies

$$\sum_{m=1}^{M} h_m(i) * g_m(i) = d(i) \quad \text{for } i = 0, \ldots, L + L_i - 2, \tag{2.56}$$

where $d(i)$ defines the target impulse response (TIR), which in most cases is desired to equal the delta function.

2.3.1 Least-squares (LS) and MINT

An equalization system $\mathbf{g}$ can be obtained by solving the system of equations (2.56), where the TIR is given by

$$d(i) = \begin{cases} 0 & \text{if } 0 \leq i < \tau; \\ 1 & \text{if } i = \tau; \\ 0 & \text{otherwise,} \end{cases} \tag{2.57}$$

with an integer delay τ. In matrix form, (2.56) can be written as

$$\mathbf{Hg} = \mathbf{d}, \tag{2.58}$$

where $\mathbf{H} = [\mathbf{H}_1 \ \cdots \ \mathbf{H}_M]$ and $\mathbf{d} = [d(0) \ \ldots \ d(i) \ \ldots \ d(L + L_i - 2)]^T$, with $\mathbf{H}_m$ the $(L + L_i - 1) \times L_i$ convolution matrix of $\mathbf{h}_m$:

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \cdots & 0 \\ h_m(1) & h_m(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_m(L-1) & \cdots & \vdots & \vdots \\ 0 & h_m(L-1) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & h_m(L-1) \end{bmatrix}.$$

When only a single microphone is deployed, (2.58) is an over-determined system of equations, and the traditional single-channel LS inverse filter that minimizes the cost function

$$J = \|\mathbf{Hg} - \mathbf{d}\|_2^2, \tag{2.59}$$

has been used for acoustic system equalization [11]. The LS solution is given by

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{d}, \tag{2.60}$$

where $\{\cdot\}^{+}$ denotes Moore-Penrose pseudo-inverse [38].

When two or more microphones are deployed, exact solution(s) to (2.58) exist when the following two conditions are both satisfied:

C-1 There is no zero common to $H_m(z)$, $m = 1, \ \ldots, \ M$, the z-transforms of the multichannel RIRs $h_m(i)$ [12].

C-2 $L_i \geq L_c$ [39][2], where

$$L_c = \left\lceil \frac{L-1}{M-1} \right\rceil \tag{2.61}$$

with $\lceil \kappa \rceil$ denotes the smallest integer larger than or equal to κ.

If both conditions are satisfied, (2.60) gives the minimum ℓ_2-norm solution to (2.58). If any one or both conditions are violated, (2.60) gives a multichannel least-squares solution.

2.3.2 Weighted least-squares (WLS)

The WLS method [40] has been used for over-determined cases. For multichannel systems, over-determination happens when the condition **C-2** is violated. The WLS solution is obtained by minimizing the following cost function

$$J = \|\mathbf{W}(\mathbf{H}\mathbf{g} - \mathbf{d})\|_2^2, \tag{2.62}$$

[2]It should be noted that the exact solution(s) to (2.58) always exist when $L_i \geq L - 1$ [22]. Unfortunately, this cannot be guaranteed when $L_i \geq L_c$. However, it has been proved in [39] that exact solution(s) exist for almost all cases.

where $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ denotes the diagonal weighting matrix with $\mathbf{w} = [w(0) \ \ldots \ w(i) \ \ldots \ w(L+L_i-2)]^T, \ w(i) \neq 0$ for $\forall \ i$.

The solution is given by

$$\mathbf{g} = (\mathbf{WH})^{+}\mathbf{Wd}. \tag{2.63}$$

When conditions **C-1** and **C-2** are both satisfied and $w(i) \neq 0$, the solution given by (2.63) is same as that given by (2.60) [41]. This means that the use of the weighting function is ineffective. This problem will be addressed in Chapter 5, where a method enables the use of weighting function is developed.

2.3.3 Channel Shortening (CS)

Channel shortening techniques [42, 43, 44, 45, 46, 47, 48], which were developed for the equalization of digital communication channels, have also been used for acoustic system equalization [49, 50]. The CS aims to maximize a generalized Rayleigh quotient:

$$\mathbf{g} = \arg\max_{\mathbf{g}} \frac{\mathbf{g}^T\mathbf{B}\mathbf{g}}{\mathbf{g}^T\mathbf{A}\mathbf{g}}, \tag{2.64}$$

where

$$\begin{aligned} \mathbf{B} &= \mathbf{H}^T\text{diag}\{\mathbf{w}_d\}^T\text{diag}\{\mathbf{w}_d\}\mathbf{H} \\ \mathbf{A} &= \mathbf{H}^T\text{diag}\{\mathbf{w}_u\}^T\text{diag}\{\mathbf{w}_u\}\mathbf{H} \end{aligned}$$

with

$$\begin{aligned} \mathbf{w}_d &= [\underbrace{0 \ \cdots \ 0}_{\tau} \ \underbrace{1 \ \cdots \ 1}_{L_w} \ 0 \ \cdots \ 0]^T_{(L+L_i-1)\times 1} \\ \mathbf{w}_u &= \mathbf{1}_{(L+L_i-1)\times 1} - \mathbf{w}_d, \end{aligned}$$

where L_w defines the region that it is desired to maximize.

The solution is the eigenvector corresponding to the largest eigenvalue of the generalized eigenvalue problem [48]

$$\mathbf{Bg} = \lambda \mathbf{Ag}. \tag{2.65}$$

However, at this point of time, the use of CS in acoustic system equalization has not been well established. For multichannel system shortening, for specific design parameters (such as $L_i = L_c$), $\mathbf{A}$ can be rank deficient. For a rank deficient $\mathbf{A}$, (2.64) has multiple solutions and the computation load for solving the generalized eigenvalue problem is extremely high. Although all solutions maximize the Rayleigh quotient, the response from the source to the output of the equalization system might be different from the viewpoint of speech perception. It is not clear in previous work which of these solutions provides perceptually advantageous speech signals. This issue will be addressed in Section 4.2.

2.4 Room acoustics and performance measures for equalization

In this section, some basic properties of room acoustics are introduced, which are important for understanding the characteristics of acoustic system equalization. Secondly, performance measures used throughout this thesis for acoustic system equalization are presented.

2.4.1 Room acoustics

In principle, any complex sound field can be considered as a superposition of numerous simple sound waves. Acoustic wave equation which governs the propagation of acoustic waves through a material medium in addition with source function and boundary conditions is a basic form to characterize the room acoustics. However, practically the wave equation is seldom directly used to analyze the room acoustics. The acoustic properties of room are modeled with various models [51, 52], i.e. pole-zero, all-zero, all-pole, and common pole-zero. On the other hand, statistical room acoustics provides a statistical description of the room impulse response. A well known time-domain model developed by Polack [53] describes the RIR as one realization of an exponentially decaying stochastic process:

$$h_m(i) = \zeta(i)e^{-\alpha i} \tag{2.66}$$

where $\zeta(i)$ is zero-mean stationary Gaussian noise, and α is the decay rate. In time-domain this model is valid after the time interval it takes for the reverberation process to become diffuse after the emission of a sound pulse by the source. This time interval is a reasonable approximation for the transition time between early reflections and late reflections [54, 55], which is somewhere in the range from 50 to 100 ms [37].

As is common practice in the current literature dealing with identification and equalization of room acoustic systems, the all-zero model (FIR filter in time-domain) of the RIR is employed throughout this thesis, as has been defined in Section 2.1.

Reverberation time T_{60}

Reverberation time T_{60} is a measurement of the severity of reverberation within a room. T_{60} is defined as the time interval it takes for the sound energy level to drop

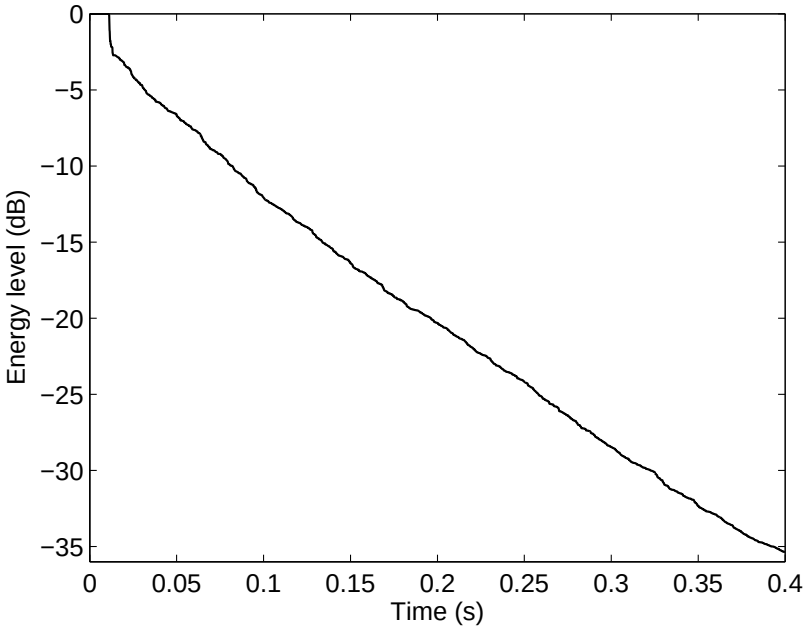

Figure 2.6: An example of energy decay curve.

by 60 dB following the sudden cessation of a broadband sound of sufficient duration (long enough to create a steady-state sound field) [37, 56].

An approach to estimate the T_{60} is through using the energy decay curve (EDC), which is defined as [37]

$$\mathscr{E}(i) = \frac{1}{\|\mathbf{h}_m\|_2^2} \sum_{j=i}^{L-1} h_m^2(j). \tag{2.67}$$

As an example, the EDC of an RIR from the MARDY database [35] is shown in Figure 2.6. It can be seen that the late reflections can be described by a realization of an exponentially decaying process, since the energy level in dB against time in the late part of the EDC is approximately linear. The gradient of the linear function fitted to the late part of the EDC can be estimated. Denoting this gradient as Δ, the T_{60} is approximated as [57]

$$\mathrm{T}_{60} = 60\Delta^{-1}. \tag{2.68}$$

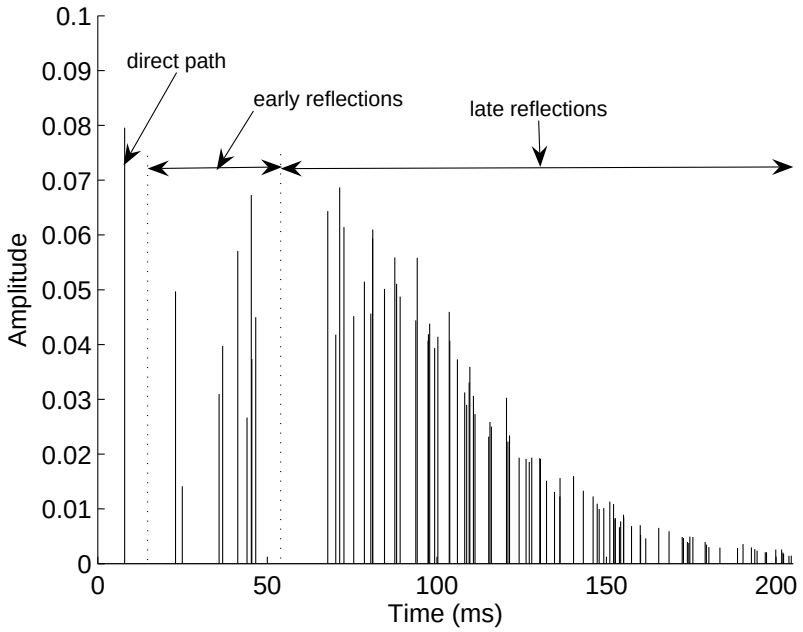

Figure 2.7: Schematic representation of an RIR.

This leads to the relationship between T_{60} and the decay rate α in (2.66) [37]

$$\alpha = \frac{3\ln(10)}{\mathrm{T}_{60} \cdot f_s}, \tag{2.69}$$

where f_s is the sampling frequency.

Subjective room acoustics

Perceptually, an RIR can be divided into three segments, the direct path, early reflections, and late reflections, as illustrated in Figure 2.7. As stated in [37], early reflections are not perceived as something separate from the direct sound. Effects caused by early reflection mainly include two aspects: increased loudness of the direct sound and changed characteristic of timbre, i.e. colouration. Since the early reflections give support to a sound source they are considered useful. On the other hand, late reflections are noticed as echoes and impair the intelligibility of speech because they blur its time structure and mix up the spectral characteristics of suc-

cessive phonemes or syllables [37]. Late reflections are considered to be detrimental from the viewpoint of speech perception [37].

For the sake of discussion, here we define the equalized (or processed) impulse response (EIR), which is the impulse response from the source to the output of the equalization system:

$$b(i) = \sum_{m=1}^{M} h_m(i) * g_m(i). \tag{2.70}$$

Ideally, $b(i)$ is expected to be equal to the delta function. However, in the presence of SIEs, the delta function cannot be achieved.

It should be noted that the above descriptions of the subjective room acoustics are applicable to an RIR, but not always to an EIR. Suppose we have an equalization system leading to an EIR which keeps exactly the direct-path and early reflections of the RIR of one channel, and completely suppresses the late reflections. In this case it is hard to say if the early reflections are still helpful. Since the late reflections impairing the intelligibility no longer exist, the loudness increase attributed to the early reflections is not that important. An EIR which has enhanced direct-path but not any reflection can also increase the loudness, and at the same time does not cause any colouration. If an EIR has only direct-path and early reflections of the RIR can be obtained, it should be better than the RIR. However, if an EIR which only has enhanced direct-path is achievable, then it might be preferable to the one that has both direct-path and early reflections; which is preferred depends on the preferences of different listeners.

Another issue should be mentioned here is about the pattern of the early reflections. In above discussion, we talked about an EIR that has early reflections exactly the same as those of an RIR. In applications, an EIR is possibly obtained with early reflections pattern not the same as that of an RIR. For EIRs having only direct-path and early reflections, the fact that an EIR with early reflections pattern

that is the same as that of an RIR is perceptively acceptable does not mean that an EIR with any pattern is acceptable. We found by informally listening to a few examples that the colouration caused by some patterns is not perceptually likable. Literature on the correlation between the early reflections pattern and subjective opinion of the caused colouration is lacking. However, studies on perceptibility of only one reflection have been conducted and results are summarized in [37]. Useful results for equalization system design are represented below:

- For speech with a sound pressure level of 70 dB, the audibility threshold of the reflected sound is

$$P = -0.575t_0 - 6 \text{ dB} \tag{2.71}$$

  where P is the pressure level of the reflected sound signal relative to the sound pressure of the direct sound and t_0 is its time delay in milliseconds. Reflected sound below this level is not perceivable at all; neither colouration nor separate echo would be perceived [37].

- If the power level of reflection is 10 times the power of direct sound, reflection within time delay of 20 ms would not be perceived as separate echo. This finding is frequently referred to as 'Haas effect' [37, 58].

- For direct sound and one reflection of equal power level, the critical delay time for the reflection to be perceived as a separate echo is between 40 and 80 ms [37, 58].

2.4.2 Performance measures for equalization

In order to evaluate the performance of equalization systems different measures can be used. In this section we define performance measures that are used in this thesis. In the simulations in this thesis, RIRs from MARDY database [35] or generated

using the image method [36] are used as true channel impulse responses $\mathbf{h}_m$. In both system identification and equalization the true RIRs are assumed to be unknown, but for purpose of evaluation the true RIRs are used to compute the EIRs such that performance measures based on the EIRs can be used. The performance measures listed below have been selected to provide a sufficiently comprehensive measurement without being exhaustive.

The first performance measure is the energy decay curve (EDC). The EDC of an EIR can be obtained with

$$\mathscr{E}(i) = \frac{1}{\|\mathbf{b}\|_2^2} \sum_{j=i}^{L+L_i-2} b^2(j), \tag{2.72}$$

where $\mathbf{b} = [b(0)\ b(1)\ \ldots\ b(L+L_i-2)]^T$. EDC is not a quantitative measure but it shows the whole decaying process.

The second performance measure is the early-to-late reverberation ratio (ELR), also known as the Clarity Index, which is defined as [37, 59]

$$C_{50} = \frac{\mathscr{E}(0) - \mathscr{E}(n_e)}{\mathscr{E}(n_e)}, \tag{2.73}$$

where $n_e = 50\ \text{ms} \cdot f_s$. This is the ratio of the energy contained in direct-path and early reflections to that contained in late reflections, which is highly correlated to the intelligibility of reverberant speech [37].

The third performance measure is T_{30} which is defined as the time interval it takes for the EDC $\mathscr{E}(i)$ to drop by 30 dB. Since the computation load for computing an equalization system for full-length RIRs is too high to afford, we usually have to truncate the RIRs in experiments. Therefore, it is not always possible to determine properly the time at which the EDC has decreased by 60 dB. Therefore, we use the T_{30} rather than the more conventional T_{60} measure. The T_{30} is an important

complement to the C_{50}. An EIR decaying fast in the first 50 ms but very slowly after 50 ms might have good C_{50} but poor T_{30}, which is not desired.

The fourth performance measure is the weighted mean square error (WMSE) between the equalized impulse response $b(i)$ and the target impulse response $d(i)$ which is defined as

$$\text{WMSE} = \sum_{i=0}^{L+L_i-2} \left(u(i) \left[\frac{1}{\gamma} b(i) - d(i) \right] \right)^2, \tag{2.74}$$

where $u(i)$ is a weighting function. Since in the measurement the weighting function is not necessarily equal to the weighting function used in the cost function (2.62), it is denoted by $u(i)$ to be distinguished from the $w(i)$. Since in speech perception, the samples in the EIR are not equally important (such as the late samples are more important in that samples of small amplitudes can cause reduced speech intelligibility), we use a weighting function $u(i)$ whose samples in the early part normally have lower amplitudes than those in the late part.

The fifth performance measure is the direct-to-reverberant ratio (DRR) [37] which is defined as

$$\text{DRR} = \frac{b(\tau)^2}{\sum_{i \neq \tau} b(i)^2}. \tag{2.75}$$

Again, since the samples in the EIR are not equally important, we can introduce a weighting function and use the weighted direct-to-reverberant ratio (WDRR)

$$\text{WDRR} = \frac{(u(\tau) \cdot b(\tau))^2}{\sum_{i \neq \tau} (u(i) \cdot b(i))^2}, \tag{2.76}$$

where $u(i)$ is the weighting function.

2.5 Equalization in the presence of system identification errors

When only $\hat{\mathbf{h}}$ is available (which is always the case in practice), an equalization system $\mathbf{g}$ can be computed that satisfies

$$\sum_{m=1}^{M} \hat{h}_m(i) * g_m(i) = d(i) \quad \text{for } i = 0, \ldots, L + L_i - 2, \tag{2.77}$$

which leads to an EIR:

$$\begin{aligned} b(i) &= \sum_{m=1}^{M} h_m(i) * g_m(i) \\ &= \gamma \sum_{m=1}^{M} \hat{h}_m(i) * g_m(i) + \sum_{m=1}^{M} e_m(i) * g_m(i) \\ &= \gamma d(i) + \sum_{m=1}^{M} e_m(i) * g_m(i). \end{aligned} \tag{2.78}$$

Since the BSI process usually introduces an unknown scaling factor γ, the first term on the right hand side of (2.78) includes the γ, which means the system $\mathbf{h}$ is equalized up to γ. As for SSI, $\gamma = 1$. In the presence of SIEs, the EIR $b(i)$ may still distort the speech signal severely due to the second term in (2.78).

The equalization approaches reviewed in Section 2.3 are generally designed without the consideration of SIEs. Illustrative examples for the performance of the MCLS are presented for both SSI and BSI. The acoustic systems used and their estimates obtained in Section 2.2.4 are employed here. The MCLS equalization systems are computed based on $\hat{\mathbf{h}}$ and then used to equalize the true systems $\mathbf{h}$ for both SSI and BSI.

For SSI, the 2-channel MCLS equalization system is computed with parameters set to $L_i = L_c$ and $\tau = 0$. The EIR and its EDC are shown in Figure 2.8. For

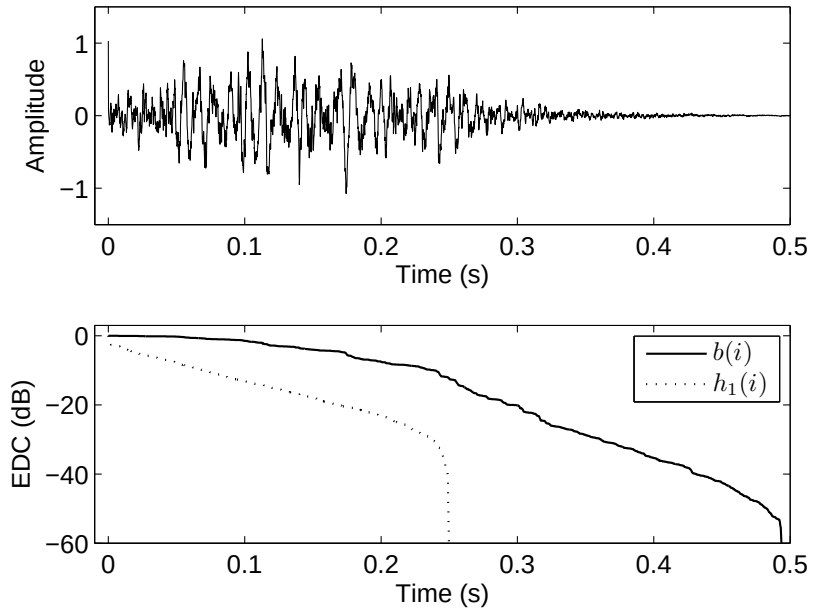

Figure 2.8: The EIR and EDCs of the EIR and $\mathbf{h}_1$ for SSI.

BSI, we only use the first two channels of the 6-channel system used in Section 2.2.4 to form a 2-channel system since the computation load and memory requirement for computing an MCLS equalization system of a 6-channel system is too high. The 2-channel MCLS equalization system is computed based on $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ with parameters set to $L_i = L_c$ and $\tau = 0$. The EIR and its EDC are shown in Figure 2.9. These two 2-channel acoustic systems for BSI and SSI respectively will be used throughout following chapters to compare the performance of different equalization algorithms developed in this thesis. It can be seen that in the presence of SIEs, the MCLS fails to equalize the acoustic system for both SSI and BSI. Although the EDC for BSI shows equalization in the early part of the EIR, the reverberation energy level in the late part of the EIR (the 'tail') is increased.

Recalling the definition of EIR in (2.78), the first term is desired, and the

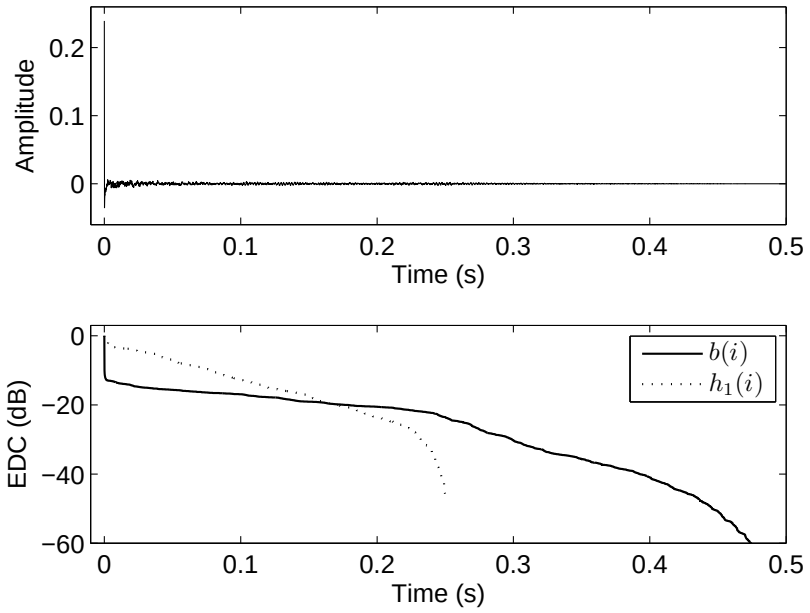

Figure 2.9: The EIR and EDCs of the EIR and $\mathbf{h}_1$ for BSI.

second term forms distortion. The mean power of the distortion is

$$\mathrm{E}\left\{\sum_{i=0}^{L+L_i-2}\left(\sum_{m=1}^{M} e_m(i) * g_m(i)\right)^2\right\} = \mathrm{E}\left\{\sum_{m=1}^{M}\sum_{i=0}^{L+L_i-2}(e_m(i) * g_m(i))^2\right\}. \tag{2.79}$$

Assuming the error sequences are white (which is true for SSI but might not be true for BSI), we have [33]

$$\sum_{i=0}^{L+L_i-2}(e_m(i) * g_m(i))^2 = \|\mathbf{e}_m\|_2^2 \cdot \|\mathbf{g}_m\|_2^2. \tag{2.80}$$

Assuming the power of the error uniformly spreads among channels, and using (2.46), (2.79) and (2.80), we have the mean power of the distortion equal to

$$\mathrm{E}\left\{\sum_{i=0}^{L+L_i-2}\left(\sum_{m=1}^{M} e_m(i) * g_m(i)\right)^2\right\} = \frac{\|\mathbf{h}\|_2^2 \cdot \mathrm{NMSE}}{M} \cdot \mathrm{E}\{\|\mathbf{g}\|_2^2\}, \tag{2.81}$$

which is proportional to the mean squared ℓ_2-norm of $\mathbf{g}$. This means that the

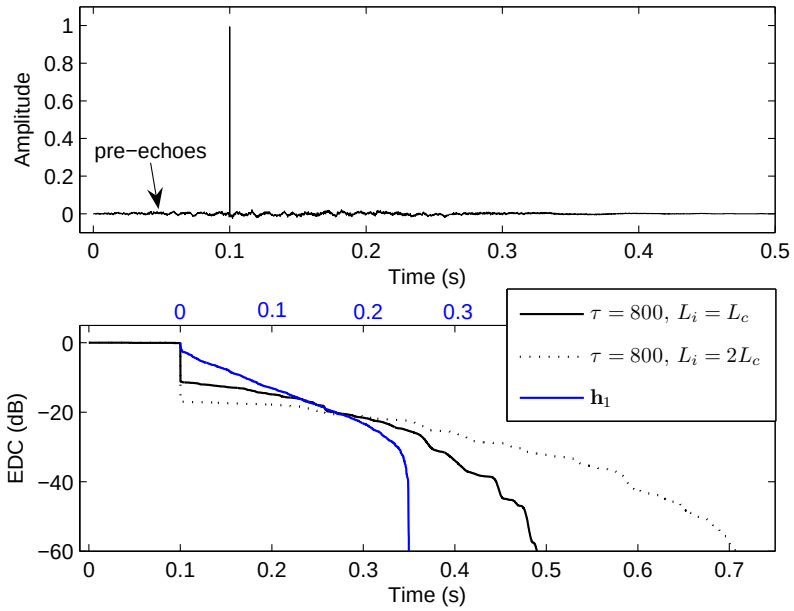

Figure 2.10: The EIR obtained with $\tau = 800$ and $L_i = L_c$, and EDCs of $\mathbf{h}_1$ and the EIRs obtained with $\tau = 800$ and $L_i = L_c$, and $\tau = 800$ and $L_i = 2L_c$.

ℓ_2-norm of the equalization system is very important for its robustness to SIEs.

It is shown in [60] that introducing delay and increasing the length of the components of the equalization system can reduce the ℓ_2-norm of the equalization system, where the components refer to the filters $\mathbf{g}_m$. In Figure 2.10, the equalization results obtained with delay introduced and L_i increased for the above SSI system are shown. For the sake of comparison, the EDC of $\mathbf{h}_1$ is plotted in accordance with the scale shown in blue color. It can be seen from the EDCs that with delay and increased length, the overall level of the distortion is reduced and the DRR is increased. However, introducing delay at the same time causes pre-echoes [61, 62], which are annoying in speech perception; increasing the length of the components introduces a longer tail in the EIR.

Unlike in digital communications where an EIR with enhanced DRR and reverberation power uniformly distributed along time in reflections is acceptable

(though the relevant terms in digital communications are not called DRR and reverberation), equalization of acoustic systems for speech dereverberation has different requirements. We summarize them below:

- Reducing the power of the distortion is not the sole objective, though it is very important. The power level of the EIR is also desired to decay with time.
- Long delay, which causes pre-echoes, cannot be introduced since the pre-echoes are annoying.
- Very long equalization system components cannot be used since the long filters introduce a long tail in the EIR.

Chapter 3

Relaxed Multichannel Least-Squares Equalization of Acoustic Systems

It is revealed by Bezout's identity [63] that for a group of polynomials $H_m(z),\ m = 1, \ldots, M$, there exists polynomials $G_m(z),\ m = 1, \ldots, M$, such that

$$\sum_{m=1}^{M} H_m(z)G_m(z) = H_0(z), \tag{3.1}$$

where $H_0(z)$ is the greatest common factor of $H_m(z)$. Traditionally, the polynomials $G_m(z)$ can be determined with the extended Euclidean algorithm [64], though they are not unique.

When $H_m(z)$ are coprime, i.e., they do not have any common zero, according to Bezout's identity, there exists $G_m(z)$ such that

$$\sum_{m=1}^{M} H_m(z)G_m(z) = 1. \tag{3.2}$$

Regarding $H_m(z)$ and $G_m(z)$ as the transfer functions of RIRs $h_m(i)$ and the equal-

ization system components $g_m(i)$, we see the same principle as that given by Miyoshi and Kaneda in their famous paper [12] published in 1988, which is called multiple-input/output inverse theorem (MINT). More importantly, in their paper, the authors provide an approach to compute the $g_m(i)$, which is via solving the system of equations (2.58):

$$\mathbf{Hg} = \mathbf{d}, \tag{3.3}$$

where the forms of $\mathbf{H}$ and $\mathbf{g}$ are given in Section 2.3, and

$$\mathbf{d} = [1\ 0\ \ldots\ 0]^T_{(L+L_i-1)\times 1}. \tag{3.4}$$

It is indicated in [12] that, for the cases of $M = 2$, (3.3) could have a unique solution under the requirement that the length of $\mathbf{g}_m$ is less than that of $\mathbf{h}_m$. However, under what requirement for the length of $\mathbf{g}_m$ the equation (3.3) has solution(s) is not clearly indicated in the paper. In [22], it is indicated that solution(s) to (3.3) always exist when $L_i \geq L - 1$. In [39], Harikumar and Bresler proved that exact solution(s) to (3.3) exist for almost all cases if

$$L_i \geq L_c = \left\lceil \frac{L-1}{M-1} \right\rceil. \tag{3.5}$$

Hereafter, we call L_c the critical length. These two conditions for the existence of solution(s) to (3.3) are summarized in **C-1** and **C-2** in Section 2.3.1. For the sake of discussion, we repeated them here:

C-1 $H_m(z)$, the z-transforms of the multichannel RIRs $h_m(i)$, do not share any common zero.

C-2 $L_i \geq L_c = \lceil \frac{L-1}{M-1} \rceil$.

In [60], following MINT, the authors formulated the inverse filtering problem into a multichannel least-squares (MCLS) form. The $\mathbf{g}$ is computed by minimizing

the cost function

$$J = \|\mathbf{H}\mathbf{g} - \mathbf{d}\|_2^2 \tag{3.6}$$

and $\mathbf{d}$ is generalized to include an arbitrary integer delay τ:

$$\mathbf{d} = [\underbrace{0 \ \ldots \ 0}_{\tau} \ 1 \ 0 \ \ldots \ 0]^T_{(L+L_i-1)\times 1}. \tag{3.7}$$

The equalization system $\mathbf{g}$ is computed using

$$\mathbf{g} = \mathbf{H}^+\mathbf{d}. \tag{3.8}$$

In this formulation, when multiple solutions exist, (3.8) gives the minimum ℓ_2-norm solution.

It is revealed in [60] that increasing L_i and introducing τ are both important for reducing the ℓ_2-norm of the resulting $\mathbf{g}$, which is important for the robustness of $\mathbf{g}$ to SIEs, as we shown in Section 2.5.

In Section 3.1, it is studied the performance of the MCLS when common zeros are present. Then a relaxed multichannel least-squares (RMCLS) algorithm is proposed for computing the $G_m(z)$ satisfying (3.1) when common zeros are present ($H_0(z) \neq 1$). After this, in Section 3.2, we study the reasons for which the equalization system obtained with (3.8) is of high ℓ_2-norm from the viewpoint of channel zeros. Two classes of characteristic zeros that cause high ℓ_2-norm of $\mathbf{g}$ are defined and how increasing L_i and introducing τ reduces the effect of these classes of zeros is investigated. Finally, we propose the use of the RMCLS in the equalization of the acoustic systems in the presence of SIEs.

3.1 RMCLS when common zeros are present

3.1.1 Performance of MCLS when common zeros are present

When common zeros are present, the condition **C-1** is violated and (3.3) does not have any solution. However, a $\mathbf{g}$ can still be obtained from (3.8). In this section, we study the performance of the $\mathbf{g}$ obtained from (3.8), which we call the MCLS solution. An earlier version of this study was presented in [P-6].

Two synthetic impulse responses $h_1(i)$ and $h_2(i)$, the transfer functions of which have $N_0 = 2$ common zeros are used in this experimental study. These two zeros are a pair of conjugate zeros. The length of $\mathbf{h}_1$ and $\mathbf{h}_2$ is $L = 128$. The transfer functions $H_1(z)$ and $H_2(z)$ can be written as

$$H_1(z) = \tilde{H}_1(z)H_0(z), \tag{3.9}$$

$$H_2(z) = \tilde{H}_2(z)H_0(z), \tag{3.10}$$

where $\tilde{H}_1(z)$ and $\tilde{H}_2(z)$ are coprime, and $H_0(z)$ is the greatest common factor.

Consider an equalization system $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T]^T$ obtained using (3.8), where $L_i = L_c$ and $\tau = 0$. Applying $\mathbf{g}$ to $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T]^T$, the EIR is obtained,

$$\begin{aligned} b(i) &= g_1(i) * h_1(i) + g_2(i) * h_2(i) \\ &= (g_1(i) * \tilde{h}_1(i) + g_2 * \tilde{h}_2(i)) * h_0(i) \\ &= \tilde{b}(i) * h_0(i), \end{aligned} \tag{3.11}$$

Where $h_0(i)$, $\tilde{h}_1(i)$, and $\tilde{h}_2(i)$ are the inverse z-transforms of $H_0(z)$, $\tilde{H}_1(z)$, and $\tilde{H}_2(z)$. On the other hand, a single-channel LS inverse filter of $h_0(i)$ with a design

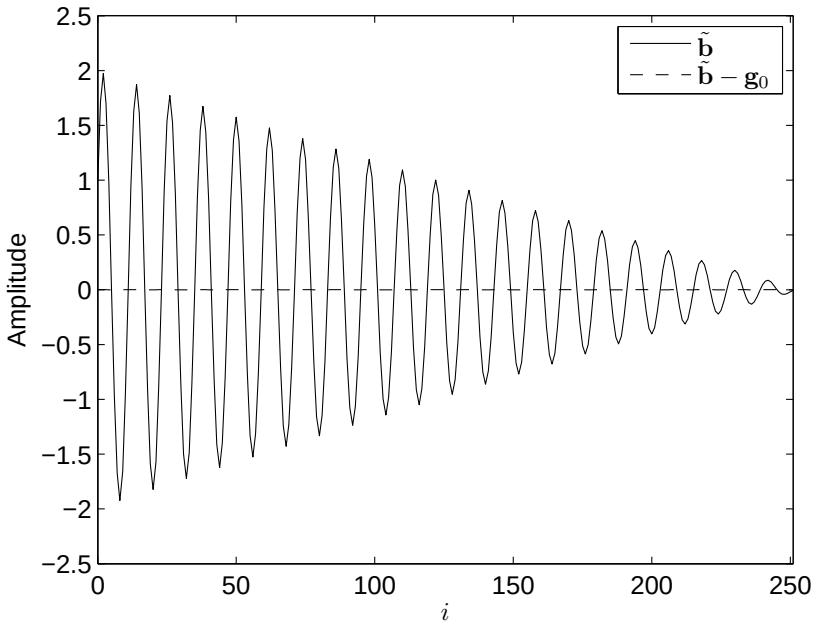

Figure 3.1: $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{b}} - \mathbf{g}_0$.

length as same as that of $\tilde{\mathbf{b}} = [\tilde{b}_0(0) \ \ldots \ \tilde{b}_0(L + L_i - N_0 - 2)]^T$ can be obtained using

$$\mathbf{g}_0 = \mathbf{H}_0^+ \mathbf{d}, \tag{3.12}$$

where $\mathbf{g}_0 = [g_0(0) \ \ldots \ g_0(L+L_i-N_0-2)]^T$ and $\mathbf{H}_0$ is the $(L+L_i-1)\times(L+L_i-N_0-1)$ convolution matrix of $\mathbf{h}_0 = [h_0(0) \ \ldots \ h_0(N_0)]^T$.

Figure 3.1 indicates that $\tilde{\mathbf{b}}$ and $\mathbf{g}_0$ are almost identical. In the experiment, the ℓ_2-norm distance between $\tilde{\mathbf{b}}$ and $\mathbf{g}_0$ is actually 1.2×10^{-11}, which is due to the rounding error in the computation. This means that when common zeros exist, the overall effect of MCLS is equivalent to performing single-channel LS inversion of the inverse z-transform of the common factor $H_0(z)$. Experiments with 100 generated systems which have two or four common zeros indicate the same property of the MCLS.

3.1.2 Performance of RMCLS when common zeros are present

By Bezout's identity, when common zeros are present, there exists $\mathbf{g}$ such that

$$\sum_{m=1}^{M} g_m(i) * h_m(i) = h_0(i). \tag{3.13}$$

In this section, we propose an approach to compute such $\mathbf{g}$, which we call relaxed multichannel least-squares (RMCLS) method. Instead of minimizing (3.6), we minimize

$$J = \|\mathbf{W}(\mathbf{H}\mathbf{g} - \mathbf{d})\|_2^2, \tag{3.14}$$

where $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ with

$$\mathbf{w} = [\underbrace{1 \ \dots \ 1}_{\tau} \ \underbrace{1 \ 0 \ \dots \ 0}_{L_w} \ 1 \ \dots \ 1]^T_{(L+L_i-1)\times 1}, \tag{3.15}$$

where L_w defines the 'relaxed window'. The difference of the RMCLS from the traditional WLS reviewed in Section 2.3.2 is that in the weighting function $w(i)$ of RMCLS, a segment of entries equal to 0 is included, whereas in the traditional WLS, $w(i)$ has non-zero entries. In the minimization process of (3.14), the amplitudes of the samples in the relaxed window is unconstrained. The first entry in the relaxed window is set to 1 rather than 0 to avoid the trivial solution. The solution $\mathbf{g}$ can be obtained using

$$\mathbf{g} = (\mathbf{W}\mathbf{H})^{+}\mathbf{W}\mathbf{d}. \tag{3.16}$$

By employing this relaxed window, we expect that the $\mathbf{h}_0$ can be manifested in the resulting EIR within the relaxed region, rather than is inverted in the single-channel LS sense. To achieve this, the number of common zeros N_0 is needed.

Firstly, we assume N_0 is known, and then L_w can be set to $L_w = N_0 + 1$,

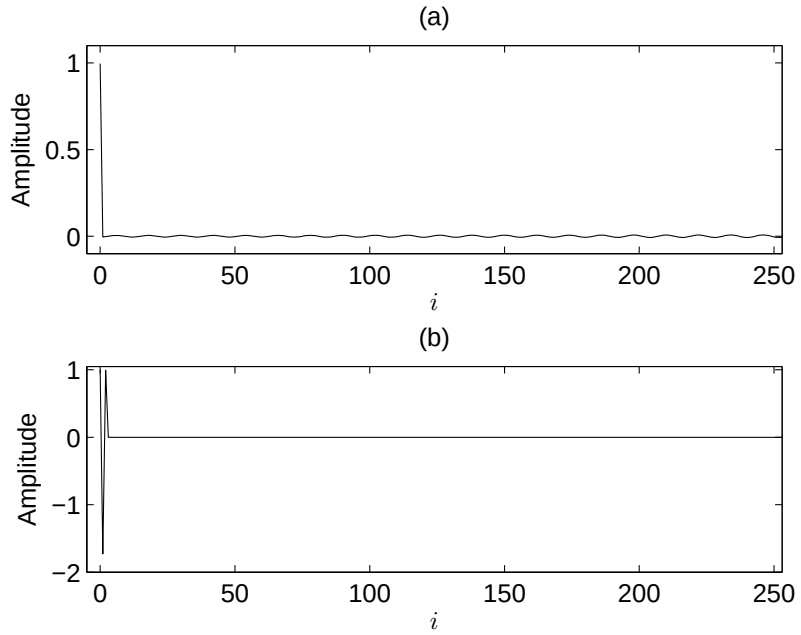

Figure 3.2: (a) EIR obtained with MCLS and (b) EIR obtained with RMCLS.

which is equivalent to the length of $\mathbf{h}_0$. The system $\mathbf{h}$ used in Section 3.1.1 is used in this experimental study. Parameters are set to $L_i = L_c$ and $\tau = 0$. The EIR obtained using $\mathbf{g}$ computed from (3.16) is shown in Figure 3.2(b). Figure 3.2(a) shows the EIR obtained using the $\mathbf{g}$ computed from (3.8). It can be seen in Figure 3.2(a) that the EIR has a non-zero tail exhibiting ripple. As has been discussed in Section 3.1.1, the EIR is equal to an LS inversion of $h_0(i)$. In Figure 3.2(b), since the 'relaxed window' is employed and no attempt is made to equalize $\mathbf{h}_0$, the equalization tail is completely suppressed with no evidence of ripple.

The number of common zeros N_0 is not always known *a priori*. However, in the application to acoustic system equalization of interest in this thesis, the exact value of N_0 is not required. Using any $L_w \geq N_0 + 1$, the equalization tail can be completely suppressed and the common part $\mathbf{h}_0$ is included as a convolutional factor in the EIR (the EIR can be expressed as the convolution of $h_0(i)$ with another function).

3.2 RMCLS method in the presence of characteristic zeros

Practically, common zeros do not exist and the significance of the study carried out in Section 3.1 only shows up when zeros from different channels are too close such that the numerical precision of the employed computations cannot discriminate them. In this case, the zeros are regarded as common zeros by the computing system and in this thesis we do not distinguish these 'too-close' zeros from the true common zeros and just refer to both of them as common zeros. Actually, the effect of common zeros on the MCLS is not very important since though common zeros are present in some acoustic systems, the number N_0 is not large. As a result, the total effect of MCLS is to perform single-channel LS inversion to a very short system $h_0(i)$ with a long filter $\tilde{b}(i)$, which gives rise to ripple with negligible amplitude in the tail of the EIR.

On the other hand, two classes of characteristic zeros are commonly present for acoustic systems, which are harmful for the robustness of the MCLS to SIEs. One class is near-common zeros. The near-common zeros are different from the above mentioned 'too-close' zeros. Near-common zeros are also close, but can be discriminated within the given numerical precision. A more strict definition for near-common zeros is given in Section 3.2.1. The other class is zeros far outside the unit circle. The performance of MCLS and RMCLS with respect to these two classes of characteristic zeros is studied in this section. An earlier version of this study was presented in [P-5].

3.2.1 Features of zeros of acoustic channels

In this section, we summarize the features of zeros of the room channels. For a typical RIR of thousands taps, the zeros of it have the following features:

1. The angles θ of the zeros are approximately uniformly distributed on $(-\pi, \pi]$ and the modulus r of most of the zeros is close to 1 [65]. Our studies have shown that the modulus r of most of the zeros lies in the interval (0.995, 1.002).

2. The modulus of a few zeros is evidently smaller than 1.

3. The modulus of a few zeros is evidently greater than 1. The modulus of such zeros can be $r > 1.03$.

4. Among multiple channels, near-common zeros usually exist.

Here we use the definition in [66] that a cluster of near-common zeros is defined when M zeros from M different RIRs are located in the same vicinity in the z-plane, the vicinity being characterized by a small 'tolerance' δ .

In Figure 3.3 an example is provided showing the distribution of the zeros of an RIR from the MARDY database [35]. The length of this RIR is truncated to $L = 2000$. In Figure 3.3, the zeros whose angles are in $[0, \pi]$ are shown; complex conjugates of these zeros are omitted for clarity. It can be seen in Figure 3.3 that there are two zeros of $r > 1.05$. In Figure 3.4, we show the number of clusters of near-common zeros against the tolerance δ between two MARDY channels. The clusters are identified with a clustering algorithm described in [66]. It can be seen that between these two $L = 2000$ channels, near-common zeros that are within vicinities of $\delta > 10^{-5}$ commonly exist.

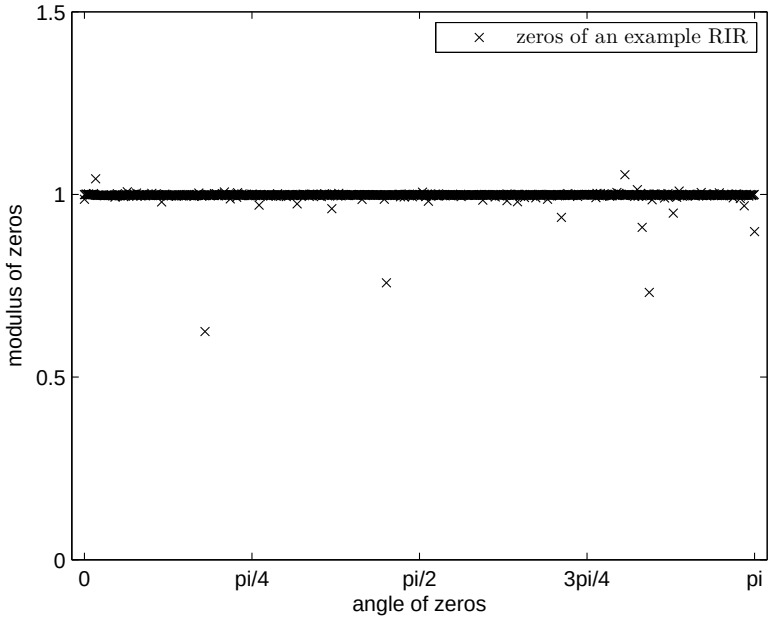

Figure 3.3: An example showing the distribution of the zeros of a typical RIR.

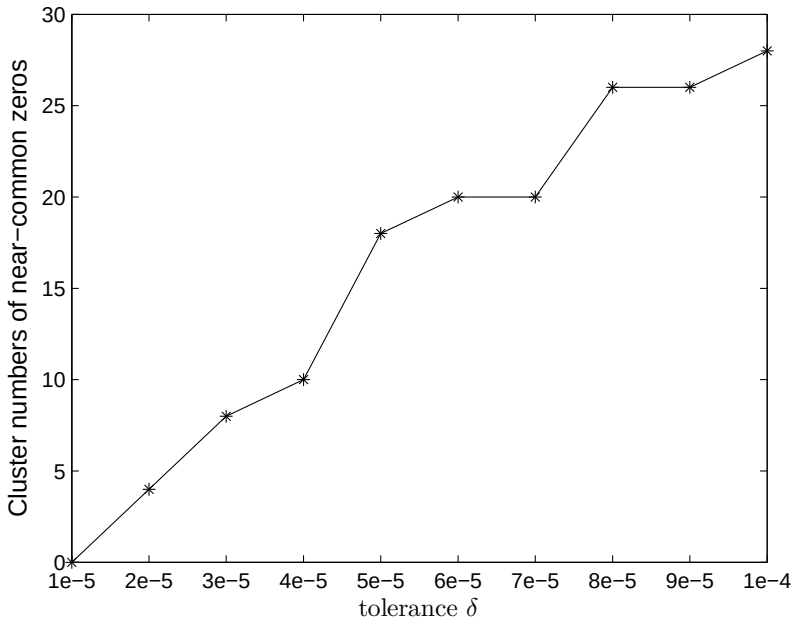

Figure 3.4: An example showing the number of clusters of near-common zeros against the tolerance δ.

3.2.2 Frequency responses of the components of equalization system in the presence of characteristic zeros

In [67], for the cases of 2 channels and with $L_i = L_c$ employed, the transfer functions $G_m(z)$ are expressed based on $H_m(z)$ and the the zeros of $H_m(z)$. Assuming the zeros of the mth channel are $z_{m,1}, \ \ldots, \ z_{m,L-1}$, $G_1(z)$ and $G_2(z)$ can be expressed as

$$G_1(z) = \sum_{k=1}^{L-1} \frac{z_{2,k}^{-\tau-1}}{H_1(z_{2,k})H_2'(z_{2,k})} \frac{H_2(z)}{(1 - z_{2,k}z^{-1})} \tag{3.17}$$

$$G_2(z) = \sum_{k=1}^{L-1} \frac{z_{1,k}^{-\tau-1}}{H_2(z_{1,k})H_1'(z_{1,k})} \frac{H_1(z)}{(1 - z_{1,k}z^{-1})}, \tag{3.18}$$

where $H_m'(z_{m,k}) = \frac{d}{dz}H_m(z)|_{z=z_{m,k}}$.

$G_1(z)$ is a weighted sum of $H_2(z)/(1 - z_{2,k}z^{-1})$ with weights $z_{2,k}^{-\tau-1}/H_1(z_{2,k})H_2'(z_{2,k})$, for $k = 1, \ldots, L-1$. When $H_2(z)$ has zeros approximately uniformly distributed around the unit circle, the frequency response of $H_2(z)/(1 - z_{2,k}z^{-1})$ has a peak at the frequency corresponding to the angle of $z_{2,k}$. Therefore, if the modulus of the weight $z_{2,k}^{-\tau-1}/H_1(z_{2,k})H_2'(z_{2,k})$ for any particular k is great, then there would be a great peak at the frequency corresponding to the angle of $z_{2,k}$ in the frequency response of $G_1(z)$, and likewise for $G_2(z)$. Recalling the features of the channel zeros listed in Section 3.2.1, it is seen that firstly, if $z_{2,k}$ and $z_{1,l}$ are a cluster of near-common zeros, $z_{2,k}$ would lead to $H_1(z_{2,k})$ of very small modulus and therefore result in a great peak in the frequency response of $G_1(z)$ at the frequency corresponding to the angle of $z_{2,k}$, and likewise for $G_2(z)$. Secondly, if $z_{2,k}$ is a zero far outside the unit circle, $z_{2,k}$ would lead to $H_1(z_{2,k})$ and $H_2'(z_{2,k})$ of very small modulus, and therefore in low delay cases (where τ has small value and the modulus of $z_{2,k}^{-\tau-1}$ is not small) result in a great peak in the frequency response of $G_1(z)$ at the frequency corresponding to the angle of $z_{2,k}$, and likewise for $G_2(z)$;

however, we can see from (3.17) and (3.18) that the strong peaks caused by the zeros far outside the unit circle can be reduced by introducing delay which leads to $z_{2,k}^{-\tau-1}$ of small modulus.

In remainder of this section, it will be shown by experiments that the near-common zeros and zeros far outside the unit circle cause strong peaks in the frequency responses of the components of MCLS equalization systems obtained from (3.8). Hereafter, these two classes of zeros will be referred to as characteristic zeros. The effect of delay in reducing the level of the peaks caused by zeros far outside the unit circle will also be shown. Since we cannot analyze the effect of increasing the length L_i using (3.17) and (3.18), we will show by experiments that increasing L_i can reduce the level of the peaks caused by the near-common zeros.

The RIRs used in the experiments are all from MARDY database [35] and are truncated to $L = 128$. For the sake of comparison, all $\mathbf{h}$ are normalized to unit ℓ_2-norm in the experiments.

Clusters of near-common zeros (Case 1)

The zeros of $\mathbf{h}_1$ and $\mathbf{h}_2$ used in this experiment are shown in Figure 3.5. A zero of $\mathbf{h}_2$ at the angle of $\theta = 0.6812\pi$ is manually moved towards a zero of $\mathbf{h}_1$ at $\theta = 0.6823\pi$ to make these two zeros within the vicinity of $\delta = 3 \times 10^{-5}$. The modulus of this pair of near-common zeros is $r = 0.983$. The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ are shown in Figure 3.6. It can be seen in Figure 3.6 that, the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained with $L_i = L_c$ and $\tau = 0$ have a strong peak at the normalized frequencies around 0.68π. It can also be seen that introducing delay does not help to reduce the level of peaks caused by the common-zeros. On the other hand, increasing the length L_i reduces the level of the peaks.

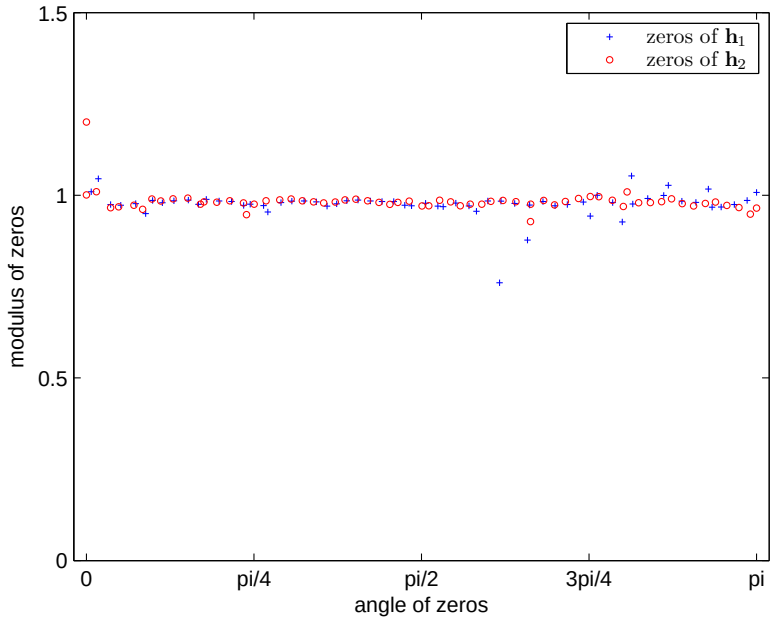

Figure 3.5: Zeros of the RIRs used to show the effect of near-common zeros.

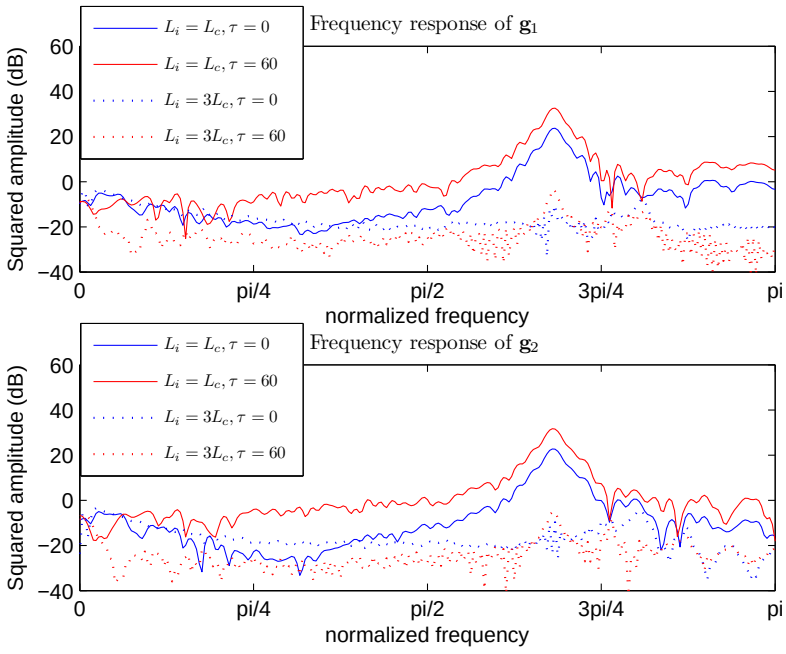

Figure 3.6: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ showing the effect of near-common zeros.

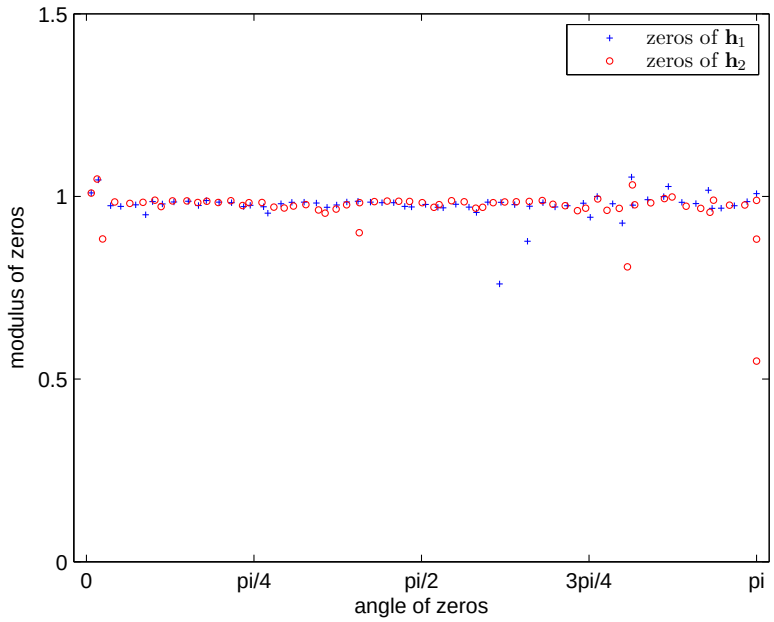

Figure 3.7: Zeros of the RIRs used to show the effect of zeros far outside the unit circle.

Zeros far outside the unit circle (Case 2)

The zeros of the RIRs used to show the effect of zeros far outside the unit circle are shown in Figure 3.7. It can be seen from Figure 3.7 that at angles about 0.02π and 0.81π, both of the two channels have zeros of $r > 1.03$. The modulus of the zeros at 0.02π is $r = 1.045$ and $r = 1.048$; the zeros at angle 0.81π are of modulus $r = 1.053$ and $r = 1.032$. The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ are shown in Figure 3.8. It can be seen in Figure 3.8 that there are two peaks in the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained with $L_i = L_c$ and $\tau = 0$. One peak is at low frequencies corresponding to the zeros at 0.02π and the other is at frequencies around 0.81π corresponding to the zeros at 0.81π. It can also be seen that only increasing the length L_i helps little to reduce the level of the peaks caused by the zeros far outside the unit circle. On the other hand, introducing delay reduces the level of the peaks.

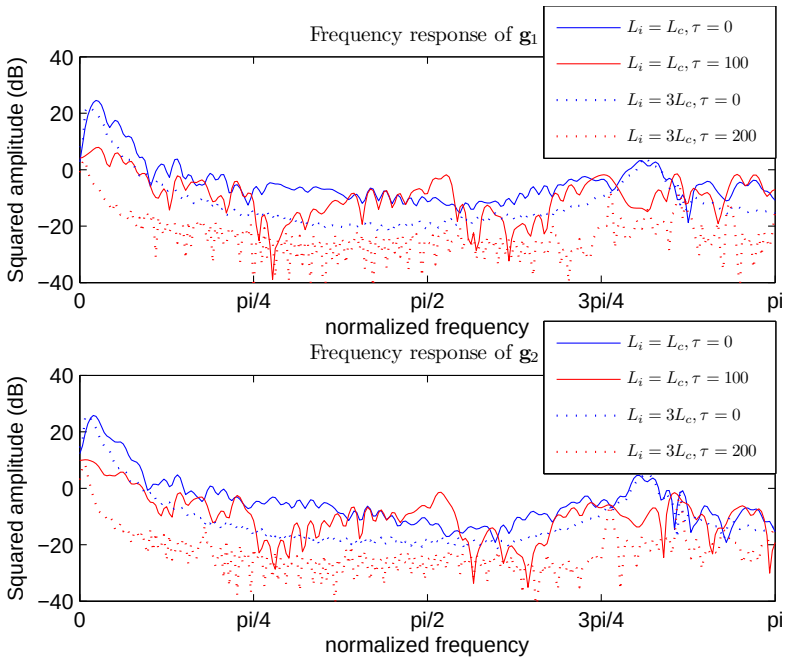

Figure 3.8: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ showing the effect of zeros far outside the unit circle.

Discussion

It can be seen in Figure 3.8 that even if with $\tau = 200$ and $L_i = 3L_c$ employed, the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ still have some peaks. (The frequency responses are repeatedly shown in Figure 3.9(a) for clearness.) It is shown in Figure 3.9(b) that these peaks are related to the notches in

$$H(\omega) = \sqrt{\sum_{m=1}^{M} |H_m(e^{j\omega})|^2}. \tag{3.19}$$

We can see in Figure 3.9 that the peaks in the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ appear at the frequencies where $H(\omega)$ has notches. $H(\omega)$ indicates the sum of the magnitude responses of the multiple channels. For $\mathbf{g}$ to be inverse system of $\mathbf{h}$ its components $\mathbf{g}_m$ must have peaks in their magnitude responses to compensate these notches, which is what has been shown in Figure 3.9. These peaks cannot be seen in

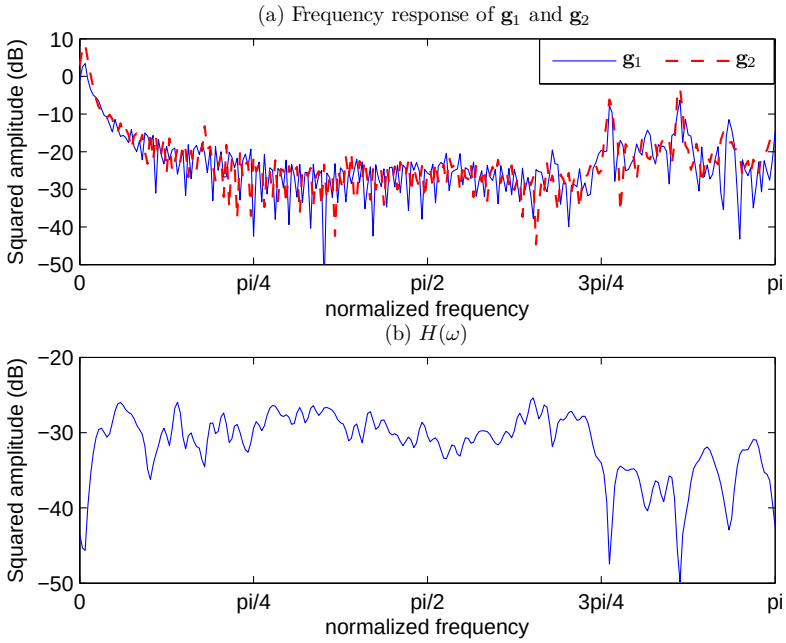

Figure 3.9: (a) The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$, and (b) squared $H(\omega)$.

the magnitude responses of $\mathbf{g}_m$ obtained with $\tau = 0$ and $L_i = L_c$ because the strong peaks caused by the characteristic zeros mask them. With $\tau = 200$ and $L_i = 3L_c$ employed, the effect of the characteristic zeros is reduced, and therefore the other peaks are more clearly visible.

3.2.3 Performance of the RMCLS in the presence of characteristic zeros

In this section, it will be shown that using (3.16) with a weighting function

$$\mathbf{w} = [\underbrace{1 \ \cdots \ 1}_{\tau} \ \underbrace{1 \, 0 \ \cdots \ 0}_{L_w} \ 1 \ \cdots \ 1]^T_{(L+L_i-1)\times 1}$$

can reduce the level of the peaks caused by the characteristic zeros without employing delay or increasing the length L_i.

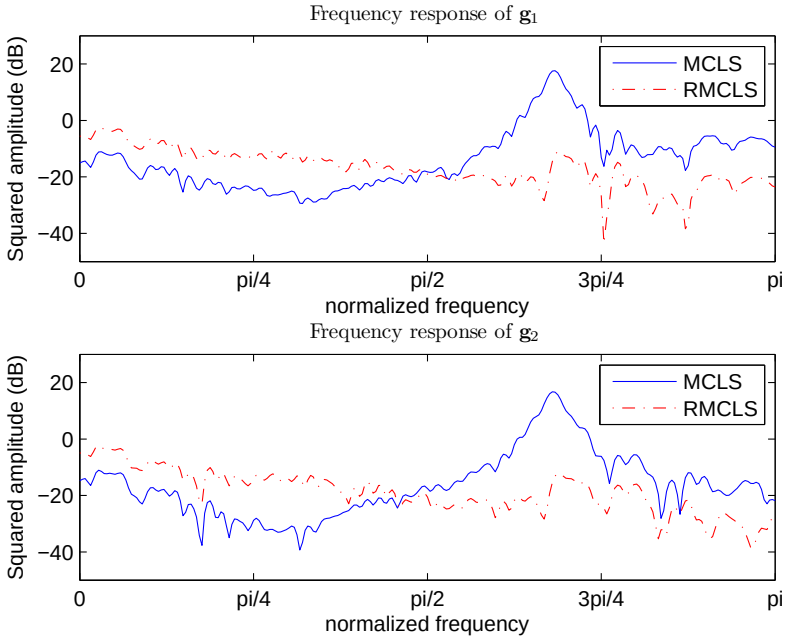

Figure 3.10: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained using the RMCLS in the presence of near-common zeros.

Clusters of near-common zeros

In Case 1 in Section 3.2.2, the channels have a cluster of near-common zeros at the angle $\theta = 0.6823\pi$. Counting in their conjugates, there are two clusters of near-common zeros in all. Therefore, we experiment with $L_w = 3$. Figure 3.10 shows the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained from (3.16) with $L_i = L_c$ and $\tau = 0$. It can be seen that without using of delay nor increased length, the peak at about $\theta = 0.68\pi$ does not exist.

The resulting EIR has accordingly from its first nonzero tap to its last nonzero tap $L_w = 3$ taps . Factorizing these 3 taps, we obtain zeros at ($r = 0.9706$, $\theta = \pm 0.6781\pi$), which can be regarded as replacements of the cluster of near-common zeros at ($r = 0.983$, $\theta = \pm 0.6823\pi$).

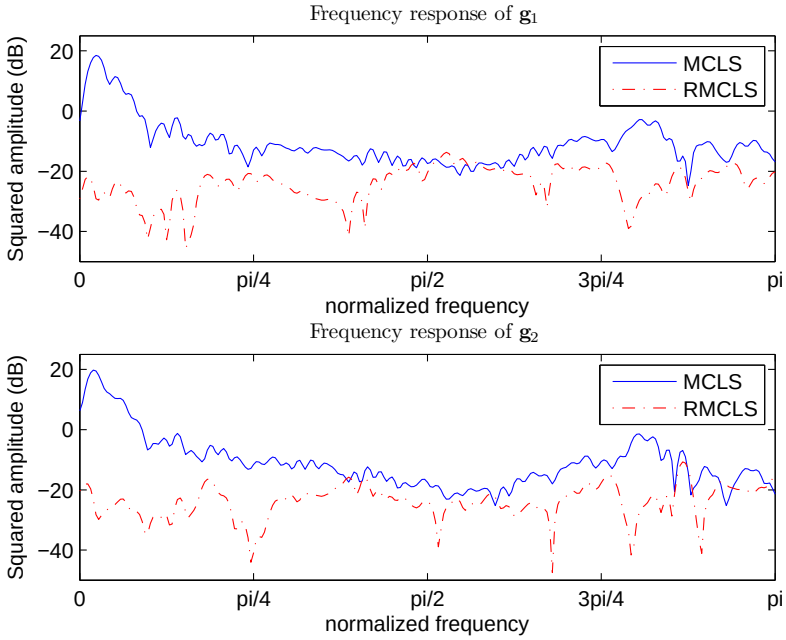

Figure 3.11: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained using the RMCLS in the presence of zeros far outside the unit circle.

Zeros far outside the unit circle

In Case 2 in Section 3.2.2, the channels have zeros far outside the unit circle at the angles about 0.02π and 0.81π. Counting in their conjugates, there are four clusters of zeros far outside the unit circle. Therefore, we experiment with $L_w = 5$. Figure 3.11 shows the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained from (3.16) with $L_i = L_c$ and $\tau = 0$. It can be seen that with neither introduced delay nor increased length, the peaks corresponding to these zeros do not exist, which shows the advantage of employing the relaxed window.

The resulting EIR has accordingly from its first nonzero tap to its last nonzero tap $L_w = 5$ taps . Factorizing these 5 taps, we obtain the zeros at ($r = 1.0237$, $\theta = \pm 0.0266\pi$) and zeros at ($r = 1.0182$, $\theta = \pm 0.8114\pi$), which correspond to the zeros far outside the unit circle at 0.02π and 0.81π respectively.

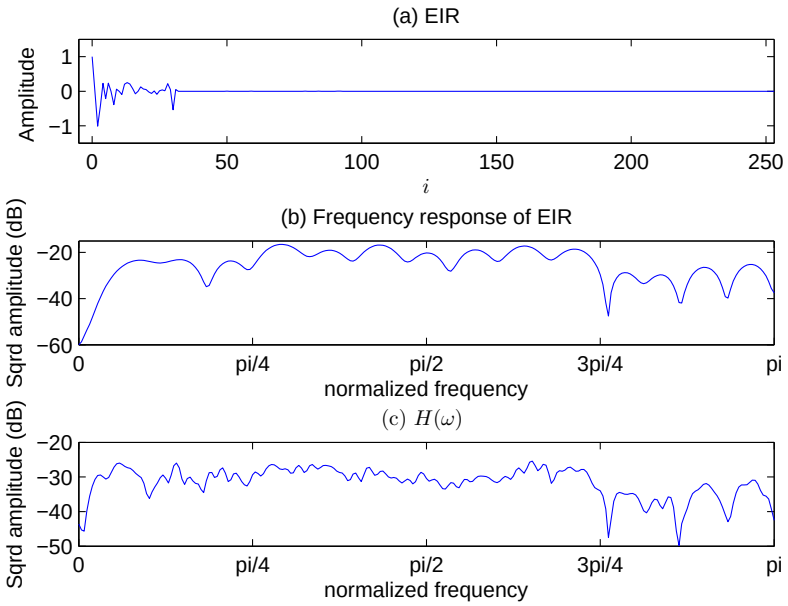

Figure 3.12: (a) the EIR, (b) the frequency response of the EIR and (c) squared $H(\omega)$.

Discussion

We experiment with the **h** used in Case 2 with $L_i = L_c$ and $\tau = 0$, but with a larger $L_w = 32$. The EIR $b(i)$, its frequency response and $H(\omega)$ are shown in Figure 3.12(a), (b) and (c) respectively.

It can be seen in Figure 3.12(b) that with $L_w = 32$ taps relaxed, the frequency response of the EIR also manifests the notches of $H(\omega)$. This shows to us that, using the RMCLS, the notches in $H(\omega)$ are left without compensation. This further reduces the ℓ_2-norm of **g**.

3.2.4 Summary of Section 3.2

We have shown in (2.81) in Section 2.5 that the mean power of the distortion in the EIR is proportional to the mean squared ℓ_2-norm of the equalization system

Table 3.1: Summary of squared ℓ_2-norm of $\mathbf{g}$ in different cases.

		$L_i = L_c$		$L_i = 3L_c$	
Case 1	MCLS	$\tau = 0$	3032.6	$\tau = 0$	22.3
		$\tau = 60$	23354.2	$\tau = 60$	7.2
	RMCLS	$\tau = 0$	4.7	$\times$	
Case 2	MCLS	$\tau = 0$	4194.5	$\tau = 0$	2634.4
		$\tau = 100$	249.4	$\tau = 200$	31.5
	RMCLS	$\tau = 0$	4.5	$\times$	

$\mathbf{g}$, which shows that the ℓ_2-norm of $\mathbf{g}$ is very important for its robustness in the presence of SIEs. The squared ℓ_2-norms of $\mathbf{g}$ obtained in all above experiments in Section 3.2 are summarized in Table 3.1. It can be seen that the squared ℓ_2-norm of $\mathbf{g}$ obtained with MCLS can be thousands times higher than that obtained from RMCLS, especially when neither delay nor increased length is employed. Introducing delay and increasing the length L_i can greatly reduce the ℓ_2-norm of $\mathbf{g}$ obtained from MCLS. On the other hand, using the RMCLS, the resulting ℓ_2-norm of $\mathbf{g}$ is very small. We have shown above when RMCLS is used, the characteristic zeros and the notches in $H(\omega)$ do not cause high ℓ_2-norm of $\mathbf{g}$, which is because they are left in the EIR, without compensation.

3.3 Simulations

It has been shown that in the presence of characteristic zeros, the relaxed window employed in RMCLS can reduce the level of the peaks in the frequency responses of $\mathbf{g}_m$ caused by these characteristic zeros, and the characteristic zeros are manifested in the resulting EIR. We have also shown that with the RMCLS, the notches in $H(\omega)$ are left without compensation, which suppresses the peaks in $|G_m(e^{j\omega})|$ obtained from MCLS, which are caused by these notches.

In this section, we apply the RMCLS to the estimates of the acoustic systems

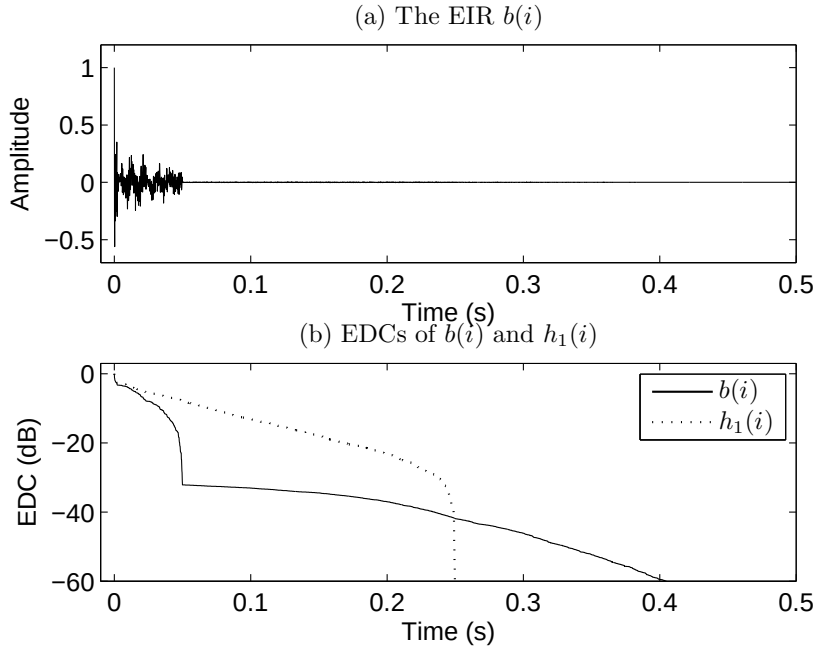

Figure 3.13: The EIR and EDCs of the EIR obtained using RMCLS and $\mathbf{h}_1$ for SSI. (A comparison of the EDC obtained from RMCLS with the EDCs obtained from MCLS and other algorithms developed in this thesis can be found in Figure 7.1.)

used in Section 2.5 and use the resulting RMCLS equalization system to equalize the acoustic systems. Parameters are set to $\tau = 0$, $L_i = L_c$ and $L_w = 400$ (corresponding to 50 ms for sampling frequency $f_s = 8000$ Hz, which is a typical transition time between early reflections and late reflections). The L_w is set according to the duration of early reflections such that the reflections in the relaxed window in the resulting EIR do not impair the intelligibility of the equalized speech. The EIR and its EDC for the SSI is shown in Figure 3.13. The EIR and its EDC for the BSI is shown in Figure 3.14. It can be seen that compared with the EIRs obtained using the MCLS, which are shown in Figure 2.8 and Figure 2.9, the equalization results obtained from RMCLS is better. Although the early reflections are not suppressed much, the late reflections are greatly suppressed. The EDCs shows more than 15 dB reduction compared with those of the RIRs at any time in the late parts for both the SSI and BSI.

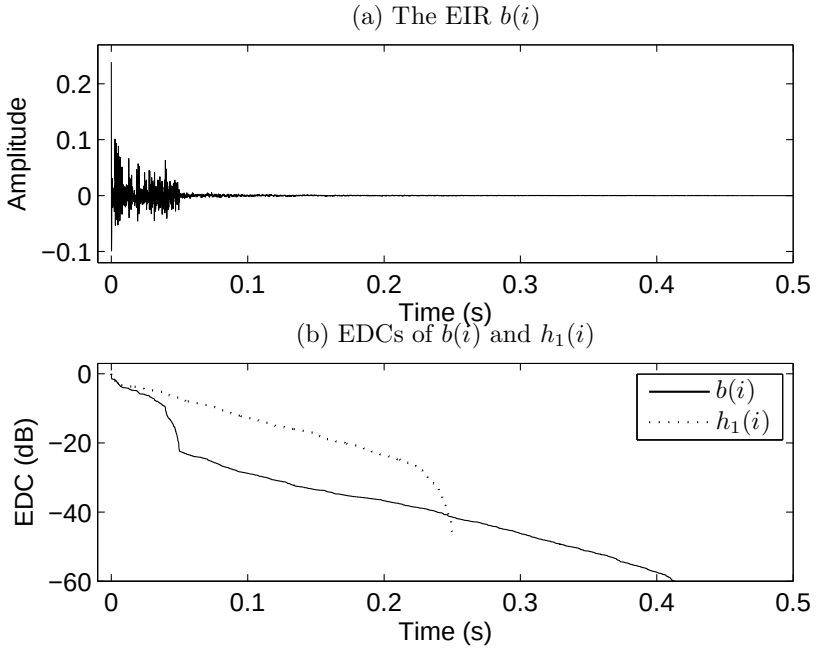

Figure 3.14: The EIR and EDCs of the EIR obtained using RMCLS and $\mathbf{h}_1$ for BSI. (A comparison of the EDC obtained from RMCLS with the EDCs obtained from MCLS and other algorithms developed in this thesis can be found in Figure 7.2.)

3.4 Summary

In this chapter, a relaxed multichannel least-squares (RMCLS) equalization method is proposed. The performance of both MCLS and RMCLS when common zeros among multiple channels are present is studied. It is shown that the overall effect of MCLS is equivalent to performing single-channel least-squares inversion of the common factor, and the RMCLS avoids equalizing the common factor. Then, two classes of characteristic zeros causing strong peaks in the frequency responses of the MCLS equalization system components, which lead to the high sensitivity of the MCLS to SIEs, are defined. The performance of both MCLS and RMCLS with respect to these two classes of characteristic zeros is studied and it is shown that since the RMCLS reduces the level of the peaks caused by the characteristic zeros, it is more robust to the SIEs than MCLS.

Due to the high computational complexity of RMCLS, which needs singular value decomposition for large matrix, we were not able to produce systematic evaluation of RMCLS. This could be carried out in future work.

Chapter 4

Channel Shortening for Use in Acoustic System Equalization

Channel shortening (CS) techniques have been developed in the context of digital communications to mitigate inter-symbol and inter-carrier interference. Both closed form [42] and adaptive [43,45,46] methods have been well studied. These techniques have been extended to the multiple-input/multiple-output (MIMO) systems in [44, 47]. A common frame work for CS can be found in [48]. Channel shortening has been used for acoustic system equalization in [49] and [50].

In an unified form [48], which is adopted in [50], the CS aims to maximize a generalized Rayleigh quotient in (2.64), which is repeated here:

$$\mathbf{g} = \arg\max_{\mathbf{g}} \frac{\mathbf{g}^T\mathbf{B}\mathbf{g}}{\mathbf{g}^T\mathbf{A}\mathbf{g}}, \tag{4.1}$$

where

$$\begin{aligned} \mathbf{B} &= \mathbf{H}^T\mathrm{diag}\{\mathbf{w}_d\}^T\mathrm{diag}\{\mathbf{w}_d\}\mathbf{H} \\ \mathbf{A} &= \mathbf{H}^T\mathrm{diag}\{\mathbf{w}_u\}^T\mathrm{diag}\{\mathbf{w}_u\}\mathbf{H} \end{aligned}$$

with

$$\begin{aligned}\mathbf{w}_d &= [\underbrace{0 \cdots 0}_{\tau}\ \underbrace{1 \cdots 1}_{L_w}\ 0 \cdots 0]^T_{[L+L_i-1]} \\ \mathbf{w}_u &= \mathbf{1}_{[(L+L_i-1)\times 1]} - \mathbf{w}_d,\end{aligned}$$

where L_w defines the region of the EIR that it is desired to maximize.

For single-channel shortening, since the target shortening length L_w is always desired to be smaller than the channel order L, $\mathbf{A}$ is of full rank. For multichannel shortening, for specific design parameters (such as $L_i = L_c$), $\mathbf{A}$ can be rank deficient. For a rank deficient $\mathbf{A}$, (4.1) has multiple solutions. Any of these solutions leads to an EIR of zero late reflections, but different solutions lead to EIRs of different early reflections patterns. In digital communications, the main issue of concern is that the quotient in (4.1) is maximized; the pattern of the impulse response after shortening is not important. In acoustic system equalization however, although all solutions maximize the Rayleigh quotient, the resulting EIRs are different from a perceptual point of view. This issue has not been considered in previous work for use of channel shortening in acoustic system equalization. In this chapter, a mathematical link between the CS and the MINT is derived. Then, a criterion for developing a perceptually advantageous equalization system from the multiple solutions to CS is provided. An earlier version of the content of this chapter was presented in [P-1].

4.1 A link between MINT and channel shortening

It is shown by MINT that, when both conditions **C-1** and **C-2** in Section 2.3.1, i.e. multiple channels do not have common zeros and $L_i \geq L_c$, are satisfied, a multichannel system can be exactly inverted. When $L_i \geq L_c$, $\mathbf{A}$ is rank deficient.

Therefore, $\mathbf{g}$ that satisfies

$$\begin{cases} \mathbf{g}^T\mathbf{A}\mathbf{g} &= 0 \\ \mathbf{g}^T\mathbf{B}\mathbf{g} &\neq 0 \end{cases} \tag{4.2}$$

maximizes the Rayleigh quotient in (4.1). Equivalently, since $\mathbf{F} \doteq \mathbf{H}^T\mathbf{H} = \mathbf{B} + \mathbf{A}$, $\mathbf{g}$ that satisfies

$$\begin{cases} \mathbf{g}^T\mathbf{A}\mathbf{g} &= 0 \\ \mathbf{g}^T\mathbf{F}\mathbf{g} &\neq 0 \end{cases} \tag{4.3}$$

is a solution to (4.1).

When multiple channels do not have common zeros, $\mathbf{H}$ is full row-rank [39], and the rank of $\mathbf{F}$ is $(L + L_i - 1)$. Since $\text{null}(\mathbf{F}) \subset \text{null}(\mathbf{A})$, where

$$\dim(\text{null}(\mathbf{F})) = ML_i - (L + L_i - 1) \doteq L_F \tag{4.4}$$

and

$$\dim(\text{null}(\mathbf{A})) = ML_i - (L + L_i - 1 - L_w), \tag{4.5}$$

we can assume vectors $\mathbf{q}_1,\ \mathbf{q}_2,\ \ldots,\ \mathbf{q}_{L_w},\ \mathbf{p}_1,\ \mathbf{p}_2,\ \ldots,\ \mathbf{p}_{L_F}$ to be a basis of $\text{null}(\mathbf{A})$, with $\mathbf{p}_1,\ \mathbf{p}_2,\ \ldots,\ \mathbf{p}_{L_F}$ to be a basis of $\text{null}(\mathbf{F})$, where $\dim(\cdot)$ denotes the dimension and $\text{null}(\cdot)$ denotes the null space. Any solution to (4.1), which is in the space $\text{null}(\mathbf{A})\backslash\text{null}(\mathbf{F})$, where $\backslash$ denotes exclusion operator, can be expressed as

$$\mathbf{g} = [\mathbf{q}_1\ \mathbf{q}_2\ \cdots\ \mathbf{q}_{L_w}]\mathbf{t}_{[L_w\times 1]} + [\mathbf{p}_1\ \mathbf{p}_2\ \ldots\ \mathbf{p}_{L_F}]\mathbf{r}_{[L_F\times 1]}, \tag{4.6}$$

with $\mathbf{t}_{[L_w\times 1]} \neq \mathbf{0}$.

In [48], the $\mathbf{g}$ maximizing the quotient in (4.1) is found by solving the generalized eigenvalue problem

$$\mathbf{B}\mathbf{g} = \lambda\mathbf{A}\mathbf{g}. \tag{4.7}$$

The solution to (4.1) is then obtained by computing the eigenvector relating to

the largest eigenvalue. For rank deficient $\mathbf{A}$, (4.7) can be solved using the QZ algorithm [68]. In this work, the MATLAB function *eig(**B**,**A**,'qz')* which employs the QZ algorithm is used to solve (4.7). L_w vectors relating to $\lambda = \infty$ can be obtained. Since the equalization system(s) obtained from MINT satisfy (4.3) as well, there must be one MINT solution which can be expressed as a linear combination of these L_w vectors.

4.2 A criterion for selecting perceptually advantageous equalization system

The MATLAB function *eig(**B**,**A**,'qz')* provides L_w independent vectors corresponding to $\lambda = \infty$. Using these vectors as equalization systems, some of them result in EIRs which lead to improved perceptual speech quality after equalization compared to the original received signal, but some do not.

As an illustrative example, Figure 4.1 shows two EIRs resulting from two equalization systems of the L_w solutions obtained by solving the generalized eigenvalue problem, for which the acoustic system is the 2-channel system from MARDY database used in Section 2.5 and parameters are set to $L_i = L_c$, $\tau = 0$, and $L_w = 400$ (corresponding to 50 ms for sampling frequency $f_s = 8000$ Hz, which is a typical transition time between early reflections and late reflections). As has been discussed in Section 2.4.1, although these two EIRs only have early reflections, they may not be both satisfactory for perception. It can be seen that the EIR shown in Figure 4.1(a) has a decaying pattern, whereas the one shown in Figure 4.1(b) has a non-decaying pattern. We conducted informal listening tests to investigate the sound quality of speech relating to the different EIRs. In the tests, speech segments were listened using a Sennheiser HD 650 headphone by 8 subjects. The following questions were investigated in the tests: is the speech warm or not warm, thin or

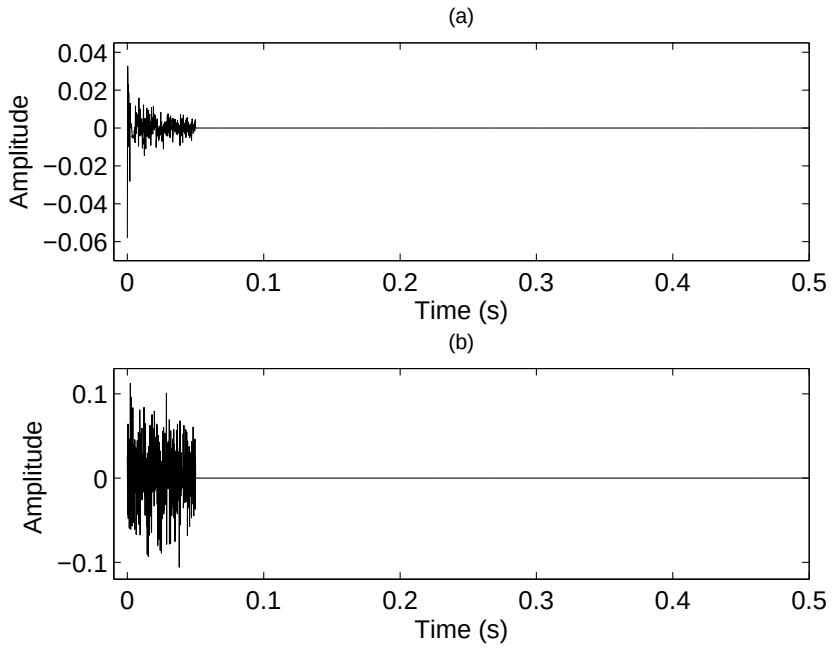

Figure 4.1: (a) An EIR resulting in perceptually improved speech, and (b) an EIR resulting in perceptually degraded speech, which are obtained from CS.

not thin, perceptually close to or not close to the anechoic speech? The obtained results indicated that speech relating to the EIR in Fig. 4.1(b) is perceived as thin and harsh, whereas speech relating to the EIR in Fig. 4.1(a) is perceived as warm, closer to the anechoic speech, and is perceptually preferred. The speech resulted from some other solutions obtained from the generalized eigenvalue decomposition sounds similar to that relating to the EIR in Figure 4.1(b) and the solution leading to the EIR in Figure 4.1(a) is more likeable than these solutions. Our proposal is that exhaustive comparison of all the L_w solutions is not necessary; a method with which the solution leading to the EIR in Figure 4.1(a) can be picked out is adequate. We found that the solution leading to the EIR in Figure 4.1(a) has the following characteristic: among the EIRs resulted from all L_w solutions, the EIR resulted from it has minimum ℓ_2-norm.

Experiments with 30 different acoustic systems show that an equalization system leading to an EIR similar (in terms of leading to similar sounding equalized

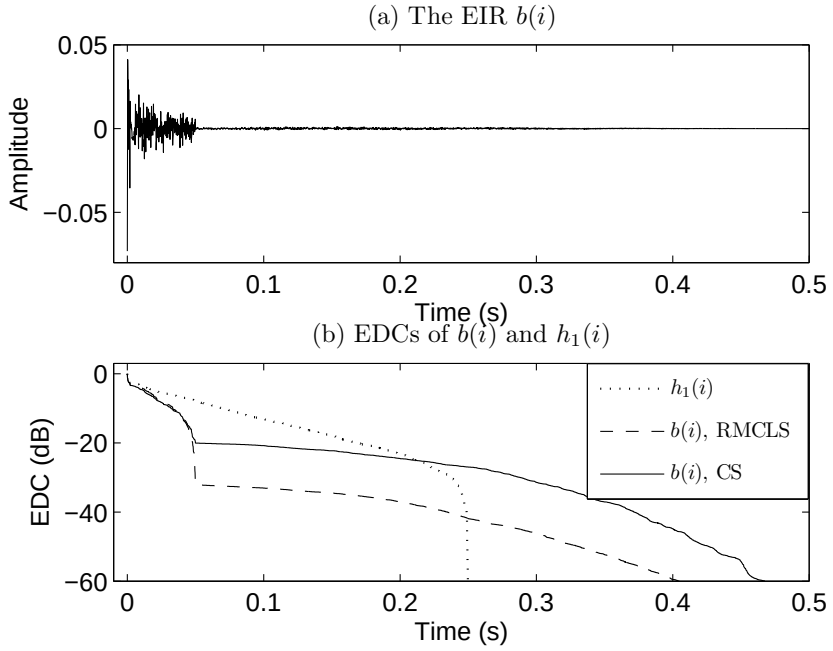

Figure 4.2: The EIR obtained from CS, and EDCs of the EIRs obtained using RMCLS and CS, and $\mathbf{h}_1$ for SSI.

speech) to the one in Figure 4.1(a) always exists for different systems, and it always retains the above ℓ_2-norm characteristic. Therefore, we propose the following criterion for selecting a solution that will result in improved perceptual quality of the speech: among the multiple solutions obtained by solving the generalized eigenvalue problem (4.7), we choose the one which results in the minimum ℓ_2-norm EIR. In the remainder of this thesis, when CS equalization system is mentioned, it always refers to the one selected with the above criterion.

4.3 Simulations and discussion

We show the performance of CS equalization system in the presence of SIEs by simulation examples. The two systems used in Section 2.5 for SSI and BSI are used again to test the robustness of the CS equalization system. Parameters are set to $\tau = 0$, $L_i = L_c$ and $L_w = 400$. The EIR and its EDC for the SSI is shown in Figure

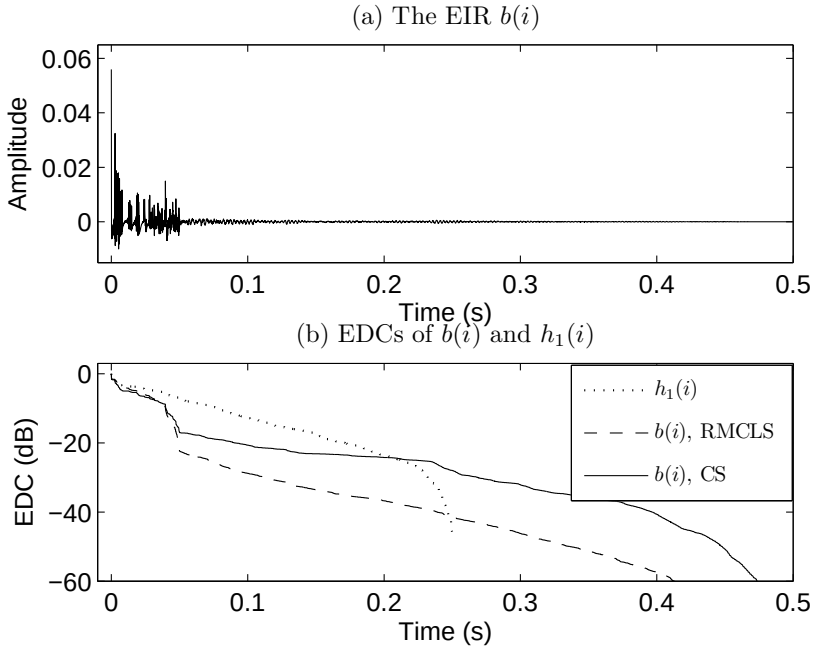

Figure 4.3: The EIR obtained from CS, and EDCs of the EIRs obtained using RMCLS and CS, and $\mathbf{h}_1$ for BSI.

4.2. The EIR and its EDC for the BSI is shown in Figure 4.3. It can be seen that in the presence of SIEs, the part of the EDC before 0.2 s is below the EDC of $\mathbf{h}_1$ for both SSI and BSI. However, the decay rate of the EDC after 0.05 s is smaller than that of $\mathbf{h}_1$, and a deleterious tail in the EIR is introduced.

Generally speaking, the RMCLS method discussed in Chapter 3 can also be classified under the concept of channel shortening. The RMCLS equalization system satisfies (4.3) as well and is in the solution space of (4.1). The early reflections pattern given by the RMCLS, which is also decaying, can be seen in Figure 3.13 and Figure 3.14. In informal listening tests, the sound of speech resulting from RMCLS is thinner than the selected CS equalization system, but is warmer and preferred to the other CS solutions. On the other hand, the RMCLS equalization system is superior in robustness to SIEs due to its characteristics discussed in Chapter 3. The EDCs obtained from RMCLS are also plotted in Figure 4.2 and Figure 4.3. It can be seen that the RMCLS is more robust than the CS. The EDCs for RMCLS

shows more than 15 dB reduction compared with $\mathbf{h}_1$ at any time after 0.05 s and no artificial tail is introduced for both SSI and BSI.

4.4 Summary

The use of channel shortening technique in equalization of acoustic systems is investigated in this chapter. A mathematical link between MINT and CS is derived. Multiple solutions to CS can be obtained and one MINT solution can be expressed as a linear combination of the CS solutions. Then, a criterion for selecting a perceptually advantageous equalization system from the multiple solutions to CS is provided. The results of informal listening tests showed that equalization using the solution corresponding to the EIR with minimum ℓ_2-norm is perceptually preferred. Extended listening tests are required to reach scientifically significant conclusions and could be conducted in future work. The performance of CS is compared with RMCLS and simulations show that RMCLS outperforms CS in robustness to SIEs.

It might be possible to formulate the maximization of the generalized Rayleigh quotient in (4.1) and finding the CS solution which leads to the minimum ℓ_2-norm EIR in one optimization problem. This could be investigated in future work.

Chapter 5

Equalization System Construction Using an Iterative Method

The weighted multichannel least-squares (WMCLS) method, with which the solution is obtained by minimizing the cost function

$$J = \|\mathbf{W}(\mathbf{Hg} - \mathbf{d})\|_2^2, \tag{5.1}$$

where the weighting function $w(i) \neq 0$ for $\forall\ i$, is used in the over-determined circumstances in [40] for acoustic system equalization. The over-determination happens when the condition **C-2** in Section 2.3.1 is violated, i.e. when $L_i < L_c$.

The solution is given by

$$\mathbf{g} = (\mathbf{WH})^{+}\mathbf{Wd}. \tag{5.2}$$

However, when conditions **C-1** and **C-2** in Section 2.3.1 are both satisfied and the weights $w(i) \neq 0$ for $\forall\ i$, the solution given by (5.2) is equal to that given by (2.60) [41], which means the WMCLS makes no difference from the MCLS, or in other words, the weighting function $w(i)$ is not effective. In Chapter 3, it has

been shown that with a relaxed window L_w, where the entries $w(i) = 0$ in this window, the RMCLS exhibits its characteristic difference from the MCLS, and the robustness of the resulting equalization system to SIEs is improved. Using RMCLS, the equalization in the relaxed region is totally unconstrained, and the pattern of the early reflections in the EIR is out of control. However, we prefer an approach with which the pattern of the early reflections in the EIR can be to some extent shaped. When conditions **C-1** and **C-2** are both satisfied, an approach which enables the weighting function with weights $w(i) \neq 0$, for $\forall\ i$, is desired.

In this chapter, we use an iterative method to approach a solution of the system of equations

$$\mathbf{WHg} = \mathbf{Wd}, \tag{5.3}$$

where $w(i) \neq 0$ for $\forall\ i$. The iterative method of course can also be applied to the MINT equations (2.58):

$$\mathbf{Hg} = \mathbf{d}. \tag{5.4}$$

Although the solutions to (5.3) and (5.4) are equivalent, the iteration processes converging to the solutions are different. In our approach, a particular initialization is assigned for the iterative method and the adaptation for (5.3) is stopped at some 'proper point', rather than after final 'convergence'. The system obtained at the 'proper point' is used as the equalization system, which is different from any of the closed-form solutions.

Iterative methods such as steepest descent [69,70] and conjugate gradient [70] can be used. In this work, we employ the conjugate gradient (CG) method. When using the CG method to solve (5.4), we refer to it simply as CG, and when we use it to solve (5.3), we refer to it as weighted conjugate gradient (WCG). After every iteration, a system $\mathbf{g}$ can be obtained and results in a corresponding EIR $b(i)$.

In this work, we only consider the use of the iterative method for robust

equalization to SSI error. Firstly, we show the different iteration processes of the CG and WCG in terms of the resulting EIRs after every iteration. Secondly, it is shown that in the presence of SSI error, the WDRR (see Section 2.4.2) of the EIR has a peak in the iteration process of the WCG, where the iteration index of the peak was referred to as 'proper point' above. Thirdly, we present an approach to estimate the iteration index of the peak of WDRR in the presence of SSI error such that the $\mathbf{g}$ after this iteration is recorded and adaptation is properly stopped. The $\mathbf{g}$ obtained from such approach provides the optimal WDRR, so we call the whole algorithm optimally-stopped weighted conjugate gradient (OS-WCG). Fourthly, the OS-WCG algorithm is evaluated.

5.1 The iteration process of the WCG

The WCG is given in Table 5.1, where k in the square brackets $[\cdot]$ denotes the index of iteration, and $\mathbf{d}$ is the vector form of the delayed delta function defined in (2.57). An earlier demonstration of this algorithm was presented in [P-4]. When implementing the CG, the line labeled 2 in Table 5.1 should be replaced by $\mathbf{f} = \mathbf{H}^T\mathbf{d},\ \mathbf{A} = \mathbf{H}^T\mathbf{H}$. We use an example to show the iteration processes of CG and WCG. In this example, the 2-channel acoustic system $\mathbf{h}$ used in Section 2.5 for SSI, the RIRs $\mathbf{h}_m$ of which are from the MARDY database [35], is employed. As discussed in Section 2.4.1, suppression of late reflections of the RIR is more important than the suppression of early reflections for improving the intelligibility of the speech signal. Therefore, for WCG, it is natural to use some weighting function $w(i)$ for which the amplitudes relating to the late reflections are larger than those relating to the early reflections. We adopt the following weighting function:

$$w(i) = \begin{cases} 1 & \text{if } 0 \le i \le \tau; \\ e^{\alpha(i-\tau)} - 1 & \text{if } i > \tau. \end{cases} \tag{5.5}$$

Table 5.1: WCG algorithm.

1	$\mathbf{g}[0] = \mathbf{0}_{ML_i}$
2	$\mathbf{f} = (\mathbf{WH})^T\mathbf{Wd},\ \mathbf{A} = (\mathbf{WH})^T(\mathbf{WH})$
3	$\mathbf{r}_a = \mathbf{f} - \mathbf{Ag}[0],\ \mathbf{p}_a = \mathbf{r}_a,\ \mu = (\mathbf{r}_a^T\mathbf{r}_a)/(\mathbf{p}_a^T\mathbf{Ap}_a)$
4	$\mathbf{g}[1] = \mathbf{g}[0] + \mu\mathbf{p}_a,\ \mathbf{r}_b = \mathbf{r}_a - \mu\mathbf{Ap}_a$
5	for $k = 1, 2, \ldots$
6	$\beta = (\mathbf{r}_b^T\mathbf{r}_b)/(\mathbf{r}_a^T\mathbf{r}_a)$
7	$\mathbf{p}_b = \mathbf{r}_b + \beta\mathbf{p}_a$
8	$\mathbf{q} = \mathbf{Ap}_b$
9	$\mu = (\mathbf{r}_b^T\mathbf{r}_b)/(\mathbf{p}_b^T\mathbf{q})$
10	$\mathbf{g}[k+1] = \mathbf{g}[k] + \mu\mathbf{p}_b$
11	$\mathbf{r}_a = \mathbf{r}_b$
12	$\mathbf{r}_b = \mathbf{r}_b - \mu\mathbf{q}$
13	$\mathbf{p}_a = \mathbf{p}_b$
14	end for

The parameters are set to $\tau = 0$, $L_i = L_c$, and $\alpha = 0.0014$ for the weighting function. The EIR $\mathbf{b}[k]$ at different iterations, where k denotes the index of iteration, are shown in Figure 5.1 and Figure 5.2 for CG and WCG respectively. It can be seen that in the iteration process, the amplitude of $b(0)$ approaches 1, and the amplitudes of $b(i)$ for $i \neq 0$ decrease. In the iteration process of CG, $b(i)$ decrease in similar rates for different i. On the other hand, the decreasing rates of $b(i)$ for different i are different in the iteration process of WCG. This means that unlike in the closed-form approach, the weighting function in the iterative approach is effective in that if we stop the adaptation at some iteration before convergence, we have different equalization systems $\mathbf{g}$ for different weighting functions used. This is important for controlling the shape of both the early part and the tail of the EIR in the presence of SSI error.

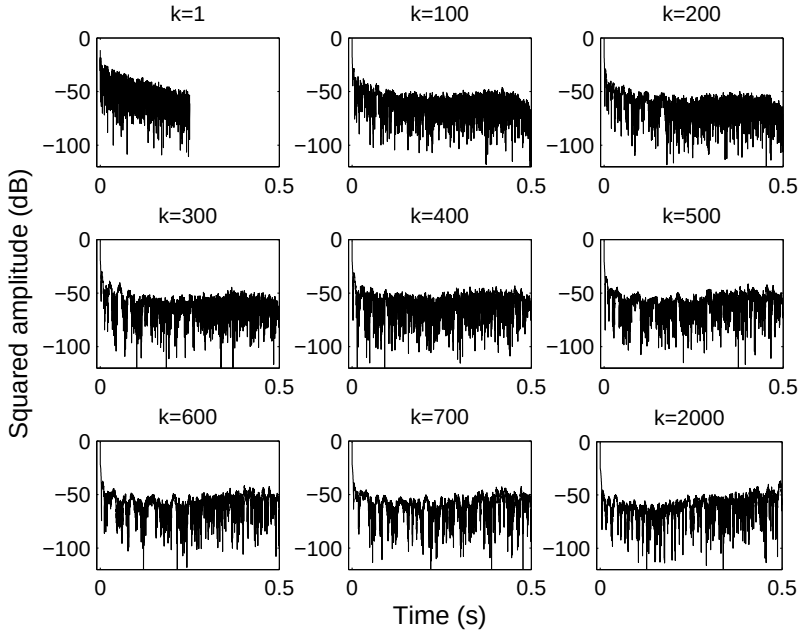

Figure 5.1: The EIR $\mathbf{b}[k]$ at different iterations showing the iteration process of CG.

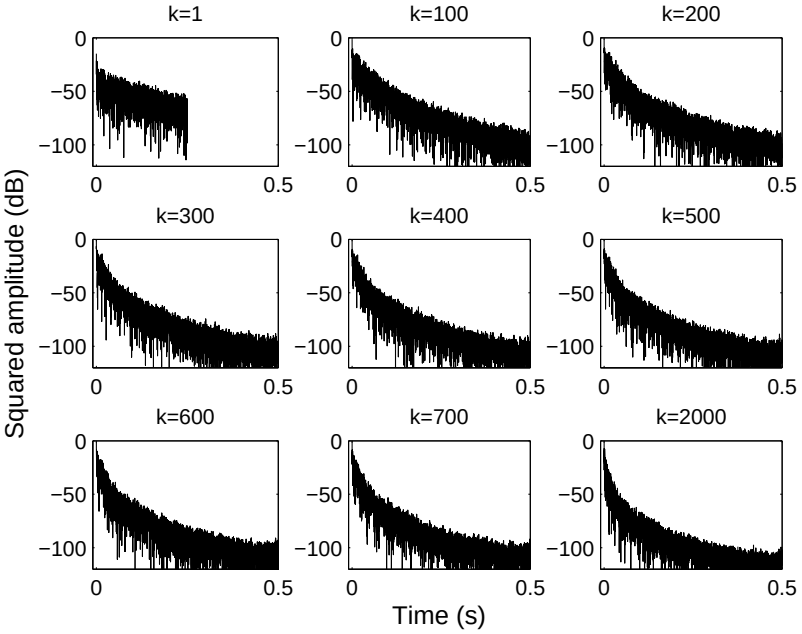

Figure 5.2: The EIR $\mathbf{b}[k]$ at different iterations showing the iteration process of WCG.

5.2 The peak of WDRR in the iteration process

In the presence of SSI error, the EIR $\mathbf{b}[k]$ can be written as

$$\mathbf{b}[k] = \mathbf{p}[k] + \mathbf{q}[k], \tag{5.6}$$

where

$$\mathbf{p}[k] = [p(0,k) \ \cdots \ p(i,k) \ \cdots \ p(L+L_i-2,k)]^T \tag{5.7}$$

with

$$p(i,k) = \sum_{m=1}^{M} \hat{h}_m(i) * g_m(i,k) \tag{5.8}$$

and

$$\mathbf{q}[k] = [q(0,k) \ \cdots \ q(i,k) \ \cdots \ q(L+L_i-2,k)]^T \tag{5.9}$$

with

$$q(i,k) = \sum_{m=1}^{M} e_m(i) * g_m(i,k). \tag{5.10}$$

In the WCG, $\mathbf{g}$ is initialized as $\mathbf{g}[0] = \mathbf{0}_{ML_i \times 1}$. With the iteration progressing (the number k increasing), $p(\tau,k)$ increases to 1, and the amplitudes of $p(0,k), \cdots, p(\tau-1,k),\ p(\tau+1,k), \cdots, p(L+L_i-2,k)$ decrease to 0, though the convergence rates of different samples are not the same. The ratio

$$\frac{p(\tau,k)^2}{\sum_{i\neq\tau} p(i,k)^2} \tag{5.11}$$

correspondingly increases. Meanwhile, since $\mathbf{g}$ is initialized as $\mathbf{g}[0] = \mathbf{0}_{ML_i \times 1}$, generally the amplitudes of the coefficients $g_m(i,k)$ increase; as a result, the amplitudes of $q(i,k)$ increase. Therefore, it can be expected that

$$\mathrm{WDRR}[k] = \frac{(u(\tau,k) \cdot b(\tau,k))^2}{\sum_{i\neq\tau}(u(i,k) \cdot b(i,k))^2} \tag{5.12}$$

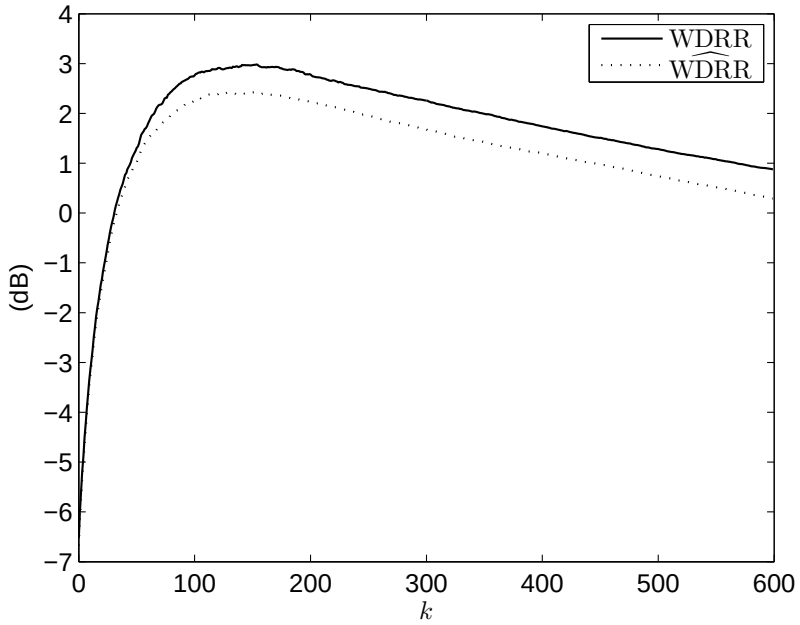

Figure 5.3: Trajectory of WDRR and $\widehat{\text{WDRR}}$.

would undergo a process of firstly increasing and then dropping, where the increasing is due to the increase of (5.11) and the dropping is due to the increase of the amplitudes of $q(i,k)$. It should be noted that in (5.12), as well as in the definition of WDRR in (2.76), weighting function is denoted by $u(i)$ to distinguish it from $w(i)$, since the weighting function used for computing the WDRR is not necessarily equal to the one used for constructing the equalization system. The turning point from increasing to dropping, i.e. the peak, would give us the largest WDRR. Accordingly, $\mathbf{g}[k^{opt}]$, where k^{opt} is the index of the peak, would be the optimal equalization system in the sense of WDRR.

The solid curve in Figure 5.3 shows an example of the trajectory of WDRR. The acoustic system $\mathbf{h}$ and the estimates $\hat{\mathbf{h}}_m$ for SSI in Section 2.5 are used in this example. The parameters are set to $\tau = 0$ and $L_i = L_c$. The weighting functions $u(i)$ is set to $u(i) = w(i)$ with $\alpha = 0.0014$, where $w(i)$ is defined in (5.5). We see in Figure 5.3 that the WDRR has a peak in the iteration process.

Therefore, to obtain the optimal $\mathbf{g}$, the k^{opt} should be found. However, in (5.12), where $b(i,k) = p(i,k) + q(i,k)$, $p(i,k)$ is observable in every iteration, but $q(i,k)$ is unobservable since $\mathbf{e}$ is unknown, so WDRR as a total is unobservable. Therefore, we need to find a way to estimate the peak index k^{opt}. In next section, the estimation of the k^{opt} is discussed.

5.3 Estimation of the iteration index of the peak

We discuss the estimation of k^{opt} for the SSI. We know from Section 2.2.1 that the SSI error is a white Gaussian process. We can also estimate the NMSE of the SSI error using (2.54). Using the above information, we propose the following approach to estimate k^{opt}.

The idea is to use a substitute of the unknown $\mathbf{e}$ to compute the WDRR after each iteration. As presented in Section 2.2.1, the SSI error $\mathbf{e}_m = \mathbf{h}_m - \hat{\mathbf{h}}_m$ satisfies

$$\mathbf{e}_m \sim \mathcal{N}(\mathbf{0}_{L\times 1}, \mathbf{C}_e), \tag{5.13}$$

where $\mathcal{N}(\cdot)$ denotes Gaussian distribution, and $\mathbf{C}_e$ is the covariance matrix of $\mathbf{e}_m$ which is asymptotically a scaled identity matrix. Therefore, we use

$$\tilde{e}_m(i) = \beta \cdot \varepsilon_m(i) \tag{5.14}$$

to substitute $e_m(i)$ in (5.10) to compute an estimate of $q(i,k)$, where $\varepsilon_m(i)$ is a generated unit-variance white Gaussian sequence, and

$$\beta = \|\mathbf{h}_m\|_2 \cdot \sqrt{\frac{\text{NMSE}}{L}} \tag{5.15}$$

with NMSE defined in (2.46). The value of β given by (5.15) ensures that the generated sequence has the same power as the true error sequence $\mathbf{e}_m$. Then an

estimate of the WDRR can be computed

$$\widehat{\text{WDRR}}[k] = \frac{(u(\tau) \cdot (p(\tau,k) + \hat{q}(\tau,k)))^2}{\sum_{i \neq \tau}(u(i) \cdot (p(i,k) + \hat{q}(i,k)))^2}, \tag{5.16}$$

where

$$\hat{q}(i,k) = \sum_{m=1}^{M} \tilde{e}_m(i) * g_m(i,k). \tag{5.17}$$

Then, $\hat{k}^{opt}$, the peak index of $\widehat{\text{WDRR}}[k]$ is used as the estimate of k^{opt}.

To obtain an estimate of k^{opt} using the above approach, the NMSE of $\mathbf{e}_m$ is necessary. In practice, the NMSE can be estimated using (2.54). However, the estimate may not be accurately equal to the true NMSE. In this section we assume the NMSE is known. An evaluation of the performance of the proposed algorithm when the NMSE is not exactly known is conducted later in Section 5.5.3.

The dashed line in Figure 5.3 shows the $\widehat{\text{WDRR}}$ obtained with a random generation of $\tilde{\mathbf{e}} = [\tilde{\mathbf{e}}_1^T \ \cdots \ \tilde{\mathbf{e}}_M^T]^T$ with $\tilde{\mathbf{e}}_m = [\tilde{e}_m(0) \ \ldots \ \tilde{e}_m(L-1)]^T$. The value of the $\widehat{\text{WDRR}}$ is not important; we only use the $\widehat{\text{WDRR}}$ to obtain an estimate of the peak index. What is important is that $\hat{k}^{opt}$, the peak index of $\widehat{\text{WDRR}}$, is a good estimate of k^{opt} and $\text{WDRR}[\hat{k}^{opt}]$ only has 0.01 dB degradation compared with $\text{WDRR}[k^{opt}]$.

5.4 Stopping condition for the WCG

In practical implementation, a stopping condition is needed to ensure $\mathbf{g}[\hat{k}^{opt}]$ is recorded and the adaptation well stopped. Since the trajectory of $\widehat{\text{WDRR}}$ in Figure 5.3 is not smooth but at the same time does not ripple dramatically, k that satisfies the following condition would be found to be $\hat{k}^{opt}$ and is recorded as the estimate of the peak index of WDRR:

$$\widehat{\text{WDRR}}[k] > \widehat{\text{WDRR}}[k+j], \text{ for } j = 1, \cdots, 100.$$

Table 5.2: Stopping condition for OS-WCG.

```
g ⇐ 0_{ML_i×1}, WDRR̂_max ⇐ 0, j ⇐ 0
while j < 100
update g
calculate WDRR̂ using (5.16)
    if WDRR̂_max < WDRR̂
        g^o ⇐ g, WDRR̂_max ⇐ WDRR̂, j ⇐ 0
    else
        j ⇐ j + 1
    end if
end while
```

Accordingly, $\mathbf{g}[\hat{k}^{opt}]$ is recorded as the optimal equalization system

$$\mathbf{g}^{\mathrm{o}} = \mathbf{g}[\hat{k}^{opt}]. \tag{5.18}$$

This stopping condition can be realized with the codes given in Table 5.2.

5.5 Performance evaluation

In this section, simulations are delivered to evaluate the performance of the OS-WCG. Firstly, we evaluate the accuracy of the peak index estimation. Secondly, the performance of the OS-WCG is evaluated using some of the performance measures given in Section 2.4.2. Thirdly, the sensitivity of the OS-WCG to NMSE estimation is evaluated. In the simulations, 2-channel systems with RIRs generated using the image method [36] are used as the acoustic systems $\mathbf{h}$. The OS-WCG can be applied to systems with more channels but it requires more memory and the computation load could be increased. Another reason for which we employ 2-channel systems is that if good performance can be achieved for 2-channel systems, systems with more channels are not necessary. Since it has been shown in Section 2.2.1 that the SSI

error is white Gaussian sequences, instead of truly identifying the acoustic systems we generate white Gaussian sequences to represent the error and subtract them from the RIRs to obtain the $\hat{\mathbf{h}}$. The room dimensions are 6.4 m $\times$ 5 m $\times$ 3.6 m; the distance between the speaker and the center of the microphone array is set to 1 m, 2 m and 3 m, and the inter-microphone distance is 10 cm; the reverberation time T_{60} is set to 0.6 s and 0.8 s; the NMSE of generated errors are -30 dB and -40 dB. The sampling frequency is $f_s = 8000$. The RIRs are of $L = 2000$. Parameters for equalization system design are set to $L_i = L_c$, $\tau = 0$. In each run of the simulation, the relative geometry of the speaker and microphone array is kept constant, but the speaker is relocated at a random position in the room, and new error is randomly generated. The evaluation results are obtained by averaging 100 runs.

The weighting function (5.5) is used, but with α set to different values for different room settings. The value of α is set relating to the T_{60} [37]

$$\alpha = \frac{3\ln(10)}{\mathrm{T}_{60} \cdot f_s}. \tag{5.19}$$

The weighting function $u(i)$ for computing the WDRR is set to $u(i) = w(i)$.

5.5.1 The accuracy of the peak index estimation

The accuracy of the peak index estimation is evaluated by comparing the mean of WDRR$[\hat{k}^{opt}]$ and WDRR$[k^{opt}]$, and the standard deviation of WDRR$[k^{opt}]$ $-$ WDRR$[\hat{k}^{opt}]$ obtained in the 100 runs. The mean of WDRR$[\hat{k}^{opt}]$ and WDRR$[k^{opt}]$ are shown in Table 5.3. It can be seen that compared with WDRR$[k^{opt}]$, WDRR$[\hat{k}^{opt}]$ only has a degradation of $0.02 - 0.04$ dB, which means that the $\mathbf{g}$ obtained after the $\hat{k}^{opt}$th iteration is almost the optimal equalization system. The standard deviation of WDRR$[k^{opt}]$ $-$ WDRR$[\hat{k}^{opt}]$ is shown in Table 5.4. We can see that the values of the standard deviation are at least 21 dB smaller than the mean values, which

Table 5.3: Mean of WDRR$[\hat{k}^{opt}]$ and WDRR$[k^{opt}]$.

NMSE (dB)	T_{60} (s)	dist (m)	WDRR$[k^{opt}]$ (dB)	WDRR$[\hat{k}^{opt}]$ (dB)
-30	0.6	1	-0.36	-0.40
		2	-5.90	-5.93
		3	-9.12	-9.16
	0.8	1	3.42	3.39
		2	-2.14	-2.16
		3	-5.81	-5.84
-40	0.6	1	7.12	7.09
		2	1.17	1.14
		3	-2.09	-2.11
	0.8	1	10.87	10.85
		2	4.79	4.76
		3	1.08	1.06

Table 5.4: Standard deviation of WDRR$[k^{opt}]$ – WDRR$[\hat{k}^{opt}]$.

NMSE (dB)	T_{60} (s)	dist (m)	standard deviation (dB)
-30	0.6	1	-21.89
		2	-26.83
		3	-30.33
	0.8	1	-18.86
		2	-24.63
		3	-29.03
-40	0.6	1	-15.30
		2	-21.26
		3	-23.04
	0.8	1	-12.81
		2	-19.05
		3	-22.65

Table 5.5: The C_{50} of $\mathbf{h}_1$ and the EIR obtained with OS-WCG.

NMSE (dB)	T_{60} (s)	dist (m)	C_{50} of EIR (dB)	C_{50} of $\mathbf{h}_1$ (dB)
-30	0.6	1	15.70	5.25
		2	13.57	3.70
		3	13.15	3.59
	0.8	1	16.90	2.91
		2	14.48	1.54
		3	14.22	1.38
-40	0.6	1	27.15	as above
		2	23.83	
		3	23.26	
	0.8	1	27.51	
		2	24.04	
		3	23.35	

means that the OS-WCG performs well in each run.

5.5.2 Performance evaluation of OS-WCG

In this section, we present the improvement in C_{50} and T_{30} achieved by the OS-WCG. Table 5.5 shows the averaged C_{50} of the EIR and $\mathbf{h}_1$. It can be seen that for all parameter settings, it shows more than 10 dB improvement in C_{50} compared with $\mathbf{h}_1$ for NMSE $= -30$ dB. For NMSE $= -40$ dB, the OS-WCG provides about 20 dB improvements. Table 5.6 shows the averaged T_{30} of the EIR and $\mathbf{h}_1$. It can be seen that for all parameter settings, the OS-WCG improves the T_{30}.

Apply the OS-WCG to the estimate of the acoustic system for SSI used in Section 2.5, the EIR and its EDC is shown in Figure 5.4. Parameters are set to $\tau = 0$, $L_i = L_c$. It can be seen that compared with $\mathbf{h}_1$, the EDC of the EIR shows more than 10 dB improvement at any time after 0.1 s. Compared with the EDC obtained from RMCLS shown in Figure 3.13, the OS-WCG achieves more suppression of the reflections before 0.05 s, and unlike the RMCLS whose EDC shows a sudden drop at 0.05 s, the EDC of OS-WCG shows a smooth transition from early part to late

Table 5.6: The T_{30} of $\mathbf{h}_1$ and the EIR obtained with OS-WCG.

NMSE (dB)	T_{60} (s)	dist (m)	T_{30} of EIR (s)	T_{30} of $\mathbf{h}_1$ (s)
-30	0.6	1	0.151	0.234
		2	0.167	0.237
		3	0.171	0.238
	0.8	1	0.202	0.245
		2	0.218	0.246
		3	0.223	0.246
-40	0.6	1	0.063	as above
		2	0.076	
		3	0.079	
	0.8	1	0.066	
		2	0.092	
		3	0.099	

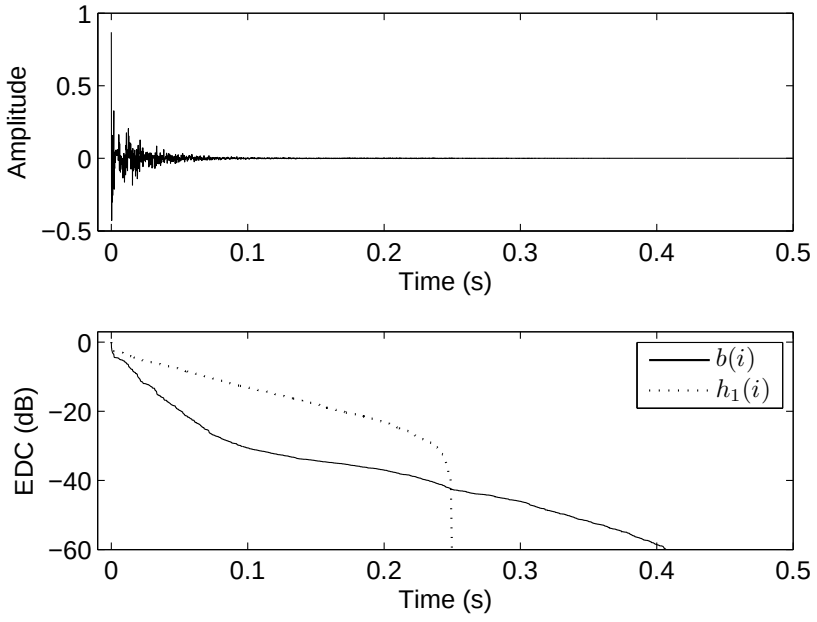

Figure 5.4: The EIR and EDCs of the EIR obtained using OS-WCG and $\mathbf{h}_1$ for SSI.

Table 5.7: The WDRR$[\hat{k}^{opt}]$ of EIR obtained with different NMSE estimates.

NMSE (dB)	$\widehat{\text{NMSE}}$ (dB)	WDRR$[\hat{k}^{opt}]$ (dB)	WDRR of $\mathbf{h}_1$ (dB)
-30	-20	-9.25	-14.94
	-30	-5.94	
	-40	-7.42	
-40	-30	-1.61	
	-40	1.14	
	-50	-0.29	

part.

5.5.3 Sensitivity to NMSE estimation

As stated in Section 5.3, the OS-WCG algorithm needs the NMSE. In practice, the NMSE needs to be estimated. The NMSE can be estimated using (2.54), which is repeated here:

$$\widehat{\text{NMSE}} = \frac{L}{N \cdot \text{SNR}}. \tag{5.20}$$

Firstly, recalling the derivation of (5.20) in Section 2.2.3, it is assumed that (2.20) is true, i.e. the autocorrelation matrix of the probing signal $s(n)$ is a scaled identity matrix. However, it cannot be ensured the true variance of the error is equivalent to that approximated by (2.21). There must be a difference between the approximated variance and the true variance, though the difference is small. Secondly, to compute $\widehat{\text{NMSE}}$, the SNR is needed, which usually needs to be estimated. Therefore, it is very possible that the estimated NMSE is not equal to the true NMSE. As a necessary part for the evaluation of the OS-WCG, we investigate the robustness of OS-WCG to NMSE estimation in this section. The WDRR$[\hat{k}^{opt}]$ for $\text{T}_{60} = 0.6$ s and distance between speaker and center of the microphone array 2 m are shown in Table 5.7.

We can see that with error in NMSE estimates, OS-WCG shows performance degradation in the WDRR of EIR. However, even with ± 10 dB error in NMSE

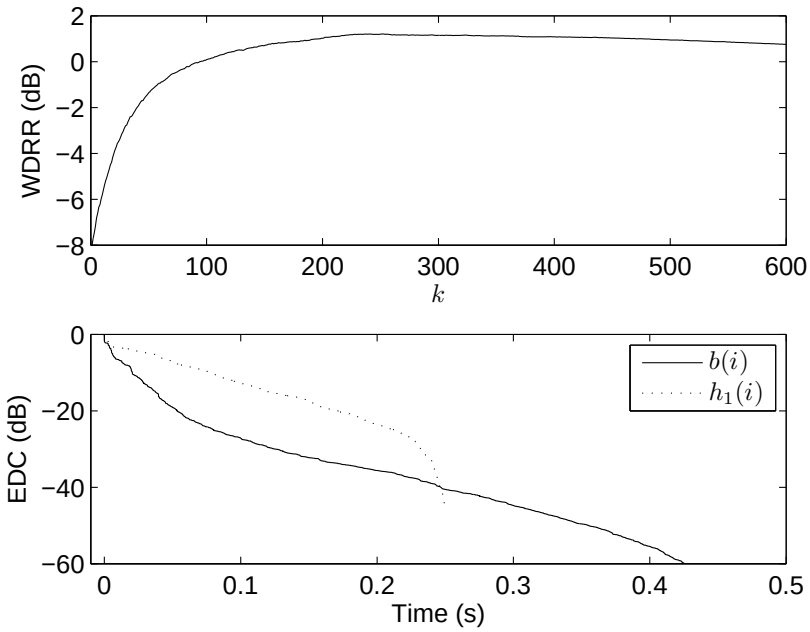

Figure 5.5: The trajectory of the WDRR and EDCs of the EIR at the iteration of the peak and $\mathbf{h}_1$ for BSI.

estimates, the resulting WDRR of the EIR still shows at least 5 dB improvement compared with that of $\mathbf{h}_1$ for NMSE $= -30$ dB. This means that the OS-WCG is not sensitive to the NMSE estimation. We also see that under-estimation of NMSE is better than over-estimation.

5.6 Discussion

Above we only discussed the application of WCG to SSI. A stopping condition for the WCG is provided and results in the OS-WCG algorithm. The WCG can also be applied to acoustic systems with BSI error and a peak of the WDRR in the iteration process can also be seen. We apply WCG to the $\hat{\mathbf{h}}$ used in Section 2.5 for the BSI, and the trajectory of the WDRR is shown in Figure 5.5, where we can see the peak. The EDC of the EIR obtained at the iteration of the peak is also shown in Figure 5.5. We see that the trajectory curve of the WDRR around its peak is very flat,

which means that if we have an approach to roughly estimate the peak index, the acoustic system with BSI error can be well equalized using WCG. However, we do not have such an approach now and it could be further explored in future work.

5.7 Summary

This chapter investigates the use of conjugate gradient iterative method for the equalization of acoustic systems, the channel estimates of which are obtained from SSI. An optimally-stopped weighted conjugate gradient (OS-WCG) algorithm is presented. In the presence of SSI error, firstly a peak of WDRR in the iterative process is shown. After this, a method to estimate the iteration index of the peak is provided. Then, a stopping condition for the iteration is proposed. Evaluation results show that using OS-WCG, WDRR at the estimated peak index only shows 0.02-0.04 dB drop compared with it at true peak index. The OS-WCG improves both the C_{50} and T_{30}. For NMSE $= -40$ dB, the OS-WCG provides about 20 dB improvement in C_{50}. Meanwhile, the OS-WCG is not sensitive to the NMSE estimation. With ± 10 dB error in NMSE estimates, the resulting WDRR of the EIR only shows less than 4 dB degradation for NMSE $= -30$ dB.

Similar peak in the WDRR trajectory for BSI can be seen. However, we do not have a method to estimate the peak index for BSI. This could be further explored in future work.

Chapter 6

Equalization of Acoustic Systems Using Models of System Identification Error

In foregoing work, except for Chapter 5 where we use the statistics of the SSI error to estimate the optimal iteration index in the OS-WCG algorithm, information of the SIEs is not used in any of the equalization method. In contrast, we explore the use of SIE models for system equalization in this chapter. The statistics for the SSI error are well known from textbooks [17] and were reviewed in Section 2.2.1. A model of SSI error can be immediately obtained from the statistics and is presented in Section 6.1.1. On the other hand, statistics of BSI error that can be straightforwardly used for system equalization are not yet available. Therefore, we conduct an experimental study of the BSI error in Section 6.1.2 and then investigate methods for modeling the error based on the statistics of the error obtained in the study. Then, in Section 6.2, the models of both the SSI and BSI error will be incorporated in the formulation of the proposed equalization method, which we call System-Identification-Error-Robust Equalization Method (SIEREM). The SIEREM is evaluated in Section 6.3.

6.1 Modeling of system identification error

It has been shown in Section 2.2 that when the channel identifiability conditions are satisfied, the channels can be identified using either the supervised or blind system identification techniques. However, in practice, even if the identifiability conditions are satisfied, we can never use infinitely long data sequences in the identification. As a result, the channel estimates always include errors. In this section, we model the SIEs.

6.1.1 Supervised system identification error

As presented in Section 2.2.1, the SSI error $\mathbf{e}_m = \mathbf{h}_m - \hat{\mathbf{h}}_m$ satisfies

$$\mathbf{e}_m \sim \mathcal{N}(\mathbf{0}_{L\times 1}, \mathbf{C}_e), \tag{6.1}$$

where $\mathcal{N}(\cdot)$ denotes Gaussian distribution, and $\mathbf{C}_e$ is the covariance matrix of $\mathbf{e}_m$ which is asymptotically a scaled identity matrix. An immediate model of the SSI error based on this is a white Gaussian sequence

$$\tilde{e}_m(i) = \beta \cdot \varepsilon_m(i), \tag{6.2}$$

where $\varepsilon_m(i)$ is a unit-variance white Gaussian sequence, and

$$\beta = \|\mathbf{h}_m\|_2 \cdot \sqrt{\frac{\text{NMSE}}{L}} \tag{6.3}$$

with

$$\text{NMSE} = \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}, \tag{6.4}$$

as defined in (2.46).

6.1.2 Blind system identification error

Using the cross-relation (CR) method [23], the estimate $\hat{\mathbf{h}}$ of the multichannel system $\mathbf{h}$ is found by computing the eigenvector corresponding the the smallest eigenvalue of $\mathbf{R}_{yy}(N)$ in (2.33):

$$\mathbf{R}_{yy}(N) = \mathbf{R}_{xx}(N) + \mathbf{R}_{xv}(N) + \mathbf{R}_{xv}^{T}(N) + \mathbf{R}_{vv}(N). \tag{6.5}$$

Compared with $\mathbf{h}$, which is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{R}_{xx}(N)$, $\hat{\mathbf{h}}$ is misaligned due to the matrix $\mathbf{R}_{xv}(N)+\mathbf{R}_{xv}^{T}(N)+\mathbf{R}_{vv}(N)$. In [34], an asymptotic variance of the CR method is derived and is compared with its CRLB. However, these analytical results cannot be used in system equalization. Firstly, the normalized projection misalignment (NPM), as defined in (2.47):

$$\text{NPM} = \frac{\|\mathbf{h} - \gamma\hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2} \tag{6.6}$$

with $\gamma = \hat{\mathbf{h}}^T\mathbf{h}/(\hat{\mathbf{h}}^T\hat{\mathbf{h}})$, is regarded as the most consistent measure for BSI but the CRLB derived in [34] corresponds to a normalization different from the normalization which leads to the NPM, and cannot be applied to the NPM [32]. Secondly, the computation of either the variance or the CRLB uses the source signal and the RIRs, which are unknown in blind scenarios. Therefore, instead, we conduct an experimental study of the BSI error and propose a model based on the study.

We use the normalized multichannel frequency-domain least-mean-squares (NMCFLMS) algorithm, which is an adaptive algorithm based on the CR method, to identify RIRs which are generated using the image method [36], and study the projection error vector which is defined in (2.49):

$$\mathbf{e} = \mathbf{h} - \gamma\hat{\mathbf{h}}. \tag{6.7}$$

It is well known that the NMCFLMS algorithm suffers misconvergence under noisy conditions [71]. In the adaptation process, the NPM of the estimate $\hat{\mathbf{h}}$ would firstly decrease and then diverge. In this work, we study the error relating to the $\hat{\mathbf{h}}$ which achieves the minimum of the NPM in the adaptation process. In the first experiment, the room dimensions are set to 6.4 m $\times$ 5 m $\times$ 3.6 m and the reverberation time T_{60} is set to 0.6 s, which represent the geometry and reverberation of a typical conference room. A linear microphone array with $M = 6$ microphones and inter-microphone distance 5 cm is deployed and the distance between the speaker and the center of the microphone array is set to 1 m. The sampling frequency is $f_s = 8000$ Hz. The RIRs are $L = 2000$ taps. The SNR is set to 25 dB. The RIRs are driven by white Gaussian noise. In each run of the simulation, the speaker is relocated at a random position in the room. The speaker and microphone arrays are avoided being too close to the walls. We employ 1000 runs and the statistical characteristics of the resulting BSI error are studied.

The 1000 error vectors obtained in the experiment are of different NPM level (and different power level). To consider the overall temporal shape of the error regardless of the power level, firstly all the error vectors are normalized to unit norm. The power level of the error relating to the corresponding NPM will be considered secondly.

The squared mean and the variance of the error of the first channel, $e_1(i)$, are shown in Figure 6.1. It can be seen that firstly, except for a few taps in the range $i < 200$, the mean of $e_1(i)$ is very small compared with its standard deviation, and secondly, the temporal shape of the variance (or standard deviation) of the error, especially for $i \geq 200$, is approximately exponentially decaying. An exponential function $\beta e^{-\alpha i}$ is fitted to the standard deviation of the error for $i \geq 200$, and the fitted value of α is $\alpha = 0.0013$. The fitted value of β is not important at this stage; it only relates to the power level of the error and will be elaborated later. With the

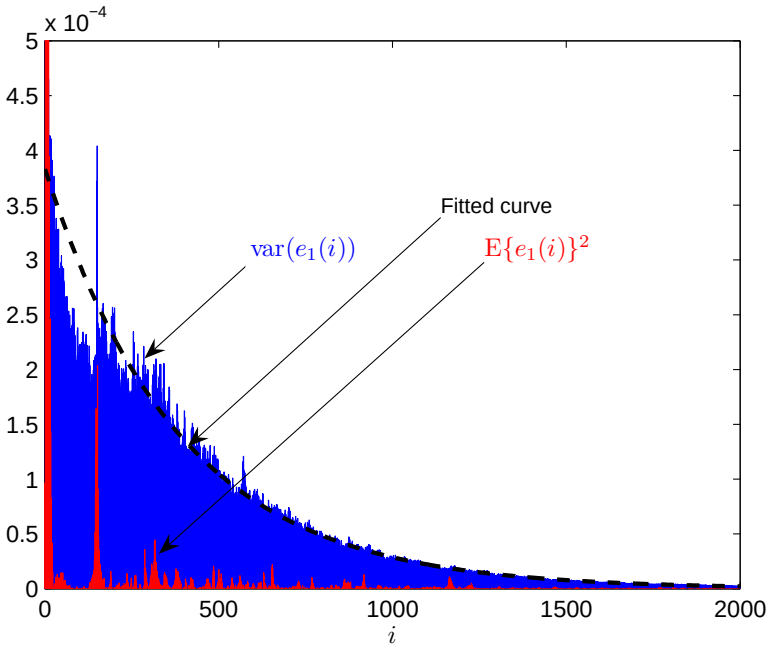

Figure 6.1: Squared mean and variance of the error of the first channel, and the exponential decay curve fitted to the variance of the error.

relation [37]

$$\alpha = \frac{3\ln(10)}{\mathrm{T}_{60} \cdot f_s}, \tag{6.8}$$

the value of α corresponds to $\mathrm{T}_{60} = 0.67$ s, which is about 12% higher than the reverberation time of the room set to $\mathrm{T}_{60} = 0.6$ s in the experiment.

Figure 6.2 shows the autocorrelation coefficients of the error $e_1(i)$. It can be seen that the coefficients $R_a(i,j) = 1$ for $i = j$ and is very small for $i \neq j$.

The cross-correlation coefficients of the error of two different channels are shown in Figure 6.3. The distance between the microphones corresponding to these two channels is 25 cm. It can be seen that the cross-correlation coefficients $R_c(i,j)$ are smaller than 0.4 for $i = j$ and is smaller than 0.1 for most of $i \neq j$. Our study also finds that for channels with smaller inter-microphone distance, the $R_c(i,j)$ for some $i = j$ can be greater, though for $i \neq j$ it remains very small.

In the second experiment, the room dimensions are set to 6.4 m $\times$ 5 m $\times$ 3.6 m;

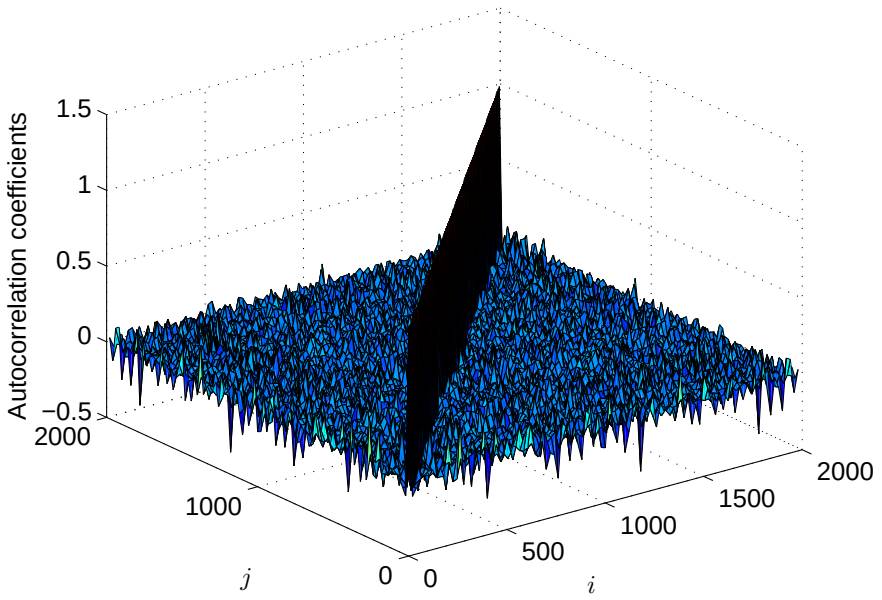

Figure 6.2: Autocorrelation coefficients of $e_1(i)$.

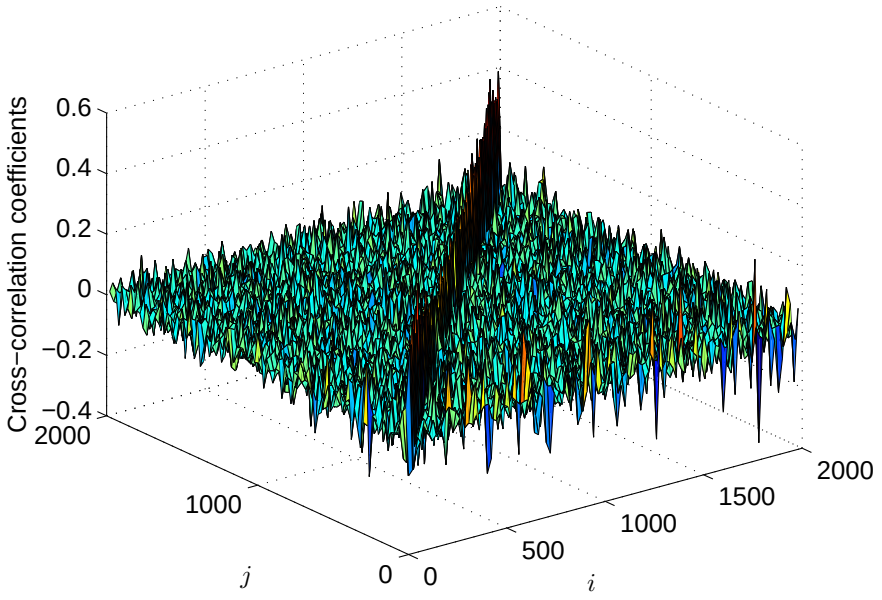

Figure 6.3: Inter-channel cross-correlation coefficients of the error.

linear microphone array with $M = 4$ microphones and inter-microphone distance 10 cm is deployed and the distance between the speaker and the centre of the microphone array is set to 2 m; the reverberation time T_{60} is set to 0.8 s. The sampling frequency is $f_s = 8000$ Hz. The RIRs are truncated to $L = 2800$ taps. We again employ 1000 runs.

The statistical characteristics of the error in this experiment are similar to those in the first experiment. The only obvious difference is that the fitted $\alpha = 0.00098$, which corresponds to $\mathrm{T}_{60} = 0.88$ s. This is about 10% higher than the reverberation time of the room set to $\mathrm{T}_{60} = 0.8$ s in the experiment.

With all these observations, the BSI error can be approximately modeled by a random sequence with an exponential decay rate α:

$$\tilde{e}_m(i) = \beta \cdot \varepsilon_m(i) \cdot e^{-\alpha i} \tag{6.9}$$

for $m = 1, \ldots, M$ and $i = 0, \ldots, L-1$, where $\varepsilon_m(i)$ is a white sequence with unit variance, β a multiplicative factor relating to the power level of $\mathbf{e}$, and the decay rate α equal to the decay rate of the RIRs, which can be estimated from $\hat{\mathbf{h}}$. The following modeling approximations have been made: firstly, the amplitudes of the errors are assumed to have zero mean; secondly, the errors are assumed to have an exponential decay rate equal to that of the RIRs; and thirdly, the errors for different channels are assumed to be uncorrelated.

The error model (6.9) will be used in next section to derive an equalization system. Although some approximations are made in modeling the error, it will be shown that with this model used, the equalization method is robust to the BSI error.

Now we derive the relationship between the multiplicative factor β and the NPM of the estimate $\hat{\mathbf{h}}$. A version of the derivation was presented in [P-3]. The BSI usually introduces an unknown scaling factor γ, and our equalization system $\mathbf{g}$ can

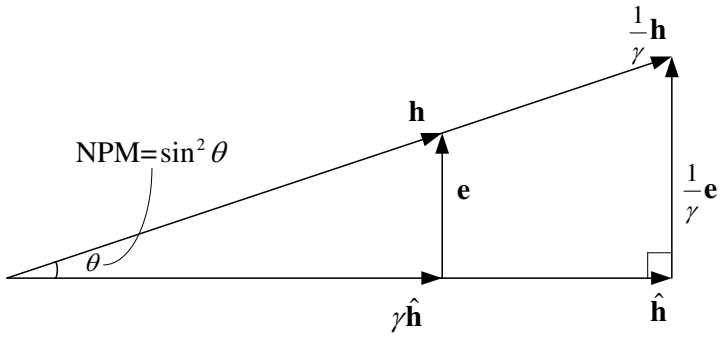

Figure 6.4: Illustration of misalignment and scaling ambiguity introduced by BSI.

only be designed based on $\hat{\mathbf{h}}$ to equalize $(1/\gamma)\mathbf{h}$ (see Figure 6.4, which is a repetition of Figure 2.3), rather than $\mathbf{h}$. However, assuming that equalizing $(1/\gamma)\mathbf{h}$ gives

$$b'(i) = \sum_{m=1}^{M} \frac{1}{\gamma} h_m(i) * g_m(i), \tag{6.10}$$

using (2.78) we see that the resulting equalized impulse response (EIR) $b(i) = \gamma b'(i)$, i.e., the acoustic system is equalized only up to a scaling factor γ.

Therefore, we need to model the $(1/\gamma)\mathbf{e}$ rather than $\mathbf{e}$. The NPM can be expressed as (see Figure 6.4),

$$\text{NPM} = \frac{\|\mathbf{e}\|_2^2}{\|\mathbf{h}\|_2^2} = \sin^2(\theta), \tag{6.11}$$

where θ is the angle between $\mathbf{h}$ and $\hat{\mathbf{h}}$ as defined in Figure 6.4. Alternatively, we have

$$\frac{\|\mathbf{e}\|_2^2}{\|\gamma\hat{\mathbf{h}}\|_2^2} = \tan^2(\theta). \tag{6.12}$$

In order to make the NPM caused by $\tilde{\mathbf{e}}$ on average equal to the true NPM, where $\tilde{\mathbf{e}} = [\tilde{\mathbf{e}}_1^T \ \ldots \ \tilde{\mathbf{e}}_M^T]^T$ with $\tilde{\mathbf{e}}_m = [\tilde{e}_m(0) \ \ldots \ \tilde{e}_m(L-1)]^T$, we require that

$$\mathrm{E}\{\|\tilde{\mathbf{e}}\|_2^2\} = \left\|\frac{1}{\gamma}\mathbf{e}\right\|_2^2. \tag{6.13}$$

Using (6.9), (6.11), (6.12) and (6.13), we can express β as

$$\beta = \frac{\tan\left[\arcsin(\sqrt{\text{NPM}})\right] \cdot \sqrt{e^{-2\alpha}-1}}{\sqrt{M \cdot (e^{-2\alpha L}-1)}} \cdot \|\hat{\mathbf{h}}\|_2. \tag{6.14}$$

6.2 The system-identification-error-robust equalization method

In this section, the system-identification-error-robust equalization method (SIEREM) using the above error models is derived. A version of the SIEREM was presented in [P-3].

We aim to obtain

$$\mathbf{g}^{\text{o}} = \arg\min_{\mathbf{g}} J \tag{6.15}$$

where

$$J = \left\| \mathbf{W} \left[(\hat{\mathbf{H}} + \frac{1}{\gamma}\mathbf{E})\mathbf{g} - \mathbf{d} \right] \right\|_2^2, \tag{6.16}$$

where $\mathbf{E}$ is formed from $\mathbf{e}$ and has the same form as $\mathbf{H}$, and $\gamma = 1$ for SSI. Compared with the cost function for WMCLS in (2.62), it can be seen a term $(1/\gamma)\mathbf{E}$ relating to SIE is incorporated. Since $\mathbf{e}$ is unknown, $\mathbf{E}$ is also unknown. In order to find $\mathbf{g}$ that minimizes (6.16), we replace $(1/\gamma)\mathbf{E}$ by $\tilde{\mathbf{E}}$ giving

$$\mathbf{g}^{\text{o}} = \arg\min_{\mathbf{g}} \left\| \mathbf{W} \left[(\hat{\mathbf{H}} + \tilde{\mathbf{E}})\mathbf{g} - \mathbf{d} \right] \right\|_2^2, \tag{6.17}$$

where $\tilde{\mathbf{E}}$ is formed by realizations of (6.2) and (6.9) for SSI and BSI respectively. However, with different realizations of the sequence $\varepsilon_m(i)$, (6.2) or (6.9) provides good and bad replacements for $(1/\gamma)\mathbf{E}$, and the performance of $\mathbf{g}$ obtained from (6.17) with different realizations of the sequence $\varepsilon_m(i)$ varies much. Our proposal is to compute a $\mathbf{g}$ which performs well on average for all realizations. Therefore, the

$\mathbf{g}$ that minimizes

$$J = \mathrm{E}\left\{ \left\| \mathbf{W}\left[(\hat{\mathbf{H}} + \tilde{\mathbf{E}})\mathbf{g} - \mathbf{d} \right] \right\|_2^2 \right\} \tag{6.18}$$

is computed. Although the $\mathbf{g}$ minimizing (6.18) is not designed for the particular error vector relating to the $\hat{\mathbf{h}}$ in question, it is desired that the $\mathbf{g}$ is more robust than an equalization system designed without any consideration of the SIE.

Expanding the right hand side of (6.18) gives

$$\begin{aligned} J &= (\mathbf{W}\hat{\mathbf{H}}\mathbf{g} - \mathbf{W}\mathbf{d})^T(\mathbf{W}\hat{\mathbf{H}}\mathbf{g} - \mathbf{W}\mathbf{d}) \\ &\quad +(\mathbf{W}\hat{\mathbf{H}}\mathbf{g} - \mathbf{W}\mathbf{d})^T\mathbf{W}\mathrm{E}\{\tilde{\mathbf{E}}\}\mathbf{g} \\ &\quad +(\mathbf{W}\mathrm{E}\{\tilde{\mathbf{E}}\}\mathbf{g})^T(\mathbf{W}\hat{\mathbf{H}}\mathbf{g} - \mathbf{W}\mathbf{d}) \\ &\quad +\mathbf{g}^T\mathrm{E}\{\tilde{\mathbf{E}}^T\mathbf{W}^T\mathbf{W}\tilde{\mathbf{E}}\}\mathbf{g}, \end{aligned} \tag{6.19}$$

where $\mathrm{E}\{\tilde{\mathbf{E}}\}$ is a zero matrix. The $\mathbf{g}$ that minimizes (6.19) can be obtained by computing the derivative of J with respect to $\mathbf{g}$ and subsequently solving

$$\frac{\partial J}{\partial \mathbf{g}} = \mathbf{0}_{ML_i \times 1}. \tag{6.20}$$

Using (6.19) and (6.20), we can obtain

$$\mathbf{g} = (\hat{\mathbf{H}}^T\mathbf{W}^T\mathbf{W}\hat{\mathbf{H}} + \mathrm{E}\{\tilde{\mathbf{E}}^T\mathbf{W}^T\mathbf{W}\tilde{\mathbf{E}}\})^{-1}\hat{\mathbf{H}}^T\mathbf{W}^T\mathbf{W}\mathbf{d}. \tag{6.21}$$

The matrix $\mathbf{R} = \mathrm{E}\{\tilde{\mathbf{E}}^T\mathbf{W}^T\mathbf{W}\tilde{\mathbf{E}}\}$ is a diagonal matrix with $r(j)$ on its diagonal, where

$$r((m-1)\cdot L_i + j) = \beta^2 \sum_{i=0}^{L-1} w^2(i+j-1) \tag{6.22}$$

for $m = 1, \ldots, M$ and $j = 1, \ldots, L_i$ for SSI and

$$r((m-1)\cdot L_i + j) = \beta^2 \sum_{i=0}^{L-1} w^2(i+j-1)e^{-2\alpha i} \tag{6.23}$$

Table 6.1: Summary of SIEREM.

1	Estimate NMSE for SSI, or NPM and T_{60} for BSI.
2	Compute β using (6.3) for SSI or (6.14) for BSI.
3	Compute α using (6.8) for BSI (not applicable to SSI).
4	Compute $r(j)$ using (6.22) for SSI or (6.23) for BSI.
5	Compute equalization system $\mathbf{g}$ using (6.21).

for BSI. Since the error models for SSI and BSI in Section 6.1 are different, we obtain different $r(j)$ for SSI and BSI.

Following the above derivation, we conclude that, given the NMSE of the SSI error, or the α of the RIRs and the NPM of the BSI error, we are able to design equalization systems that takes into account the SIE. Clearly the NMSE, or the decay rate and NPM are not known *a priori* and therefore need to be estimated. The NMSE can be estimated using (2.54):

$$\widehat{\mathrm{NMSE}} = \frac{L}{N \cdot \mathrm{SNR}}; \tag{6.24}$$

the decay rate (or equivalently the T_{60} of the room) can be estimated from $\hat{\mathbf{h}}$ or other methods [72, 73], but an approach to blind estimation of NPM is not yet available and blind NPM estimation is still an open question. We will therefore investigate later in this chapter the sensitivity in performance of our equalization method to the accuracy of the estimates of the NMSE, and the decay rate and NPM.

The SIEREM is summarized in Table 6.1.

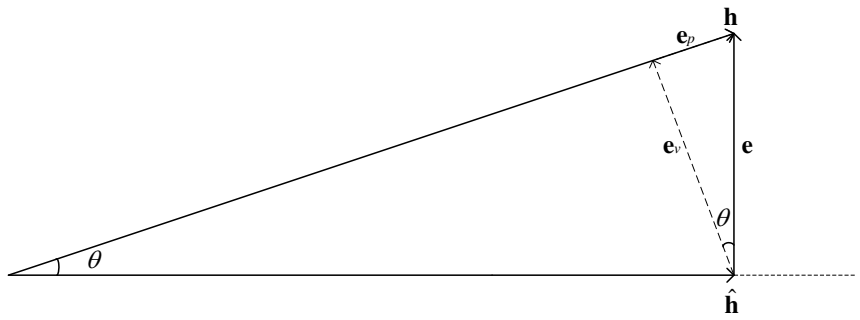

Figure 6.5: Illustration of the error representation

6.3 Evaluation

In this section, the performance of the proposed SIEREM is evaluated. The weighting function (5.5)

$$w(i) = \begin{cases} 1 & \text{if } 0 \leq i \leq \tau; \\ e^{\alpha(i-\tau)} - 1 & \text{if } i > \tau, \end{cases} \tag{6.25}$$

is used again for SIEREM, where α is directly related to T_{60} with (6.8).

Because existing BSI techniques are not yet able to provide channel estimates with very low NPM, we evaluate the performance at different NPM by subtracting a generated error vector $\mathbf{e}$ from $\mathbf{h}$. An algorithm to generate representations of BSI error of desired NPM is presented firstly.

6.3.1 Generation of BSI error

In this section, an algorithm to generate representations of BSI error, which enables systematic testing of the performance of system equalization, is presented. With this algorithm, the NPM of the generated error representation can be chosen to suit any desired level.

We formulate the problem in Figure 6.5. For illustration, we reduce this problem to 2 dimensions although extension to higher dimensionality is straightforward.

Since the scaling factor γ in the true estimates does not influence the equalization quality, we neglect the γ in error generation.

The NPM corresponds, in terms of Figure 6.5, only to the angle θ between $\mathbf{h}$ and $\hat{\mathbf{h}}$. The error vector $\mathbf{e}$ can be decomposed into two components, of which one is parallel to $\mathbf{h}$ and the other is normal to $\mathbf{h}$. The length of the parallel component $\mathbf{e}_p$ is $\|\mathbf{e}_p\|_2 = \sin^2\theta\|\mathbf{h}\|_2$. The component $\mathbf{e}_v$ normal to $\mathbf{h}$ is constrained by

$$\mathbf{h}^T\mathbf{e}_v = 0 \tag{6.26}$$

$$\|\mathbf{e}_v\|_2 = \|\mathbf{h}\|_2 \sin\theta\cos\theta. \tag{6.27}$$

Substituting (6.11) into (6.27) gives

$$\|\mathbf{e}_v\|_2 = \|\mathbf{h}\|_2\sqrt{\mathrm{NPM}(1-\mathrm{NPM})}. \tag{6.28}$$

It can be seen that the direction of $\mathbf{e}_v$ is constrained by (6.26) and its length is determined from (6.28).

An ensuing procedure is first to generate a random vector orthogonal to $\mathbf{h}$, and then adjust it to the desired length. The error vector can be generated following the steps below:

1. Generate

$$a_m(i) = \epsilon_m(i)e^{-\alpha i} \tag{6.29}$$

   to form $\mathbf{a} = [\mathbf{a}_1^T \ \cdots \ \mathbf{a}_M^T]^T$ with $\mathbf{a}_m = [a_m(0) \ \ldots \ a_m(L-1)]^T$, where $\epsilon_m(i)$ is a white Gaussian sequence with unit variance.

2. Apply Gram-Schmidt orthogonalization [74] to $\mathbf{h}$ and the random vector $\mathbf{a}$ to obtain a new vector $\mathbf{a}_v$ which is orthogonal to $\mathbf{h}$.

3. Adjust the length of $\mathbf{a}_v$ according to (6.28) to obtain $\mathbf{e}_v$.

4. Generate $\mathbf{e}_p = \sin^2\theta\mathbf{h}$.

5. Sum $\mathbf{e}_v$ and $\mathbf{e}_p$ to obtain $\mathbf{e}$.

We have presented an algorithm to generate representations of BSI error. A version of this algorithm is presented in [P-7]. The proposed error generation algorithm facilitates repeatable testing of system equalization.

6.3.2 Performance evaluation

Firstly, we present the evaluation results of SIEREM for SSI error. In the simulations, 2-channel systems with RIRs generated using the image method [36] are used as the acoustic systems $\mathbf{h}$. Since it has been shown in Section 2.2.1 that the SSI error is white Gaussian sequences, white Gaussian sequences are generated and subtracted from the RIRs to obtain the $\hat{\mathbf{h}}$.

The room dimensions are 6.4 m $\times$ 5 m $\times$ 3.6 m; the distance between the speaker and the center of the microphone array is set to 1 m, 2 m and 3 m, and the inter-microphone distance is 10 cm; the reverberation time T_{60} is set to 0.6 s and 0.8 s; the NMSE of generated error is -30 dB and -40 dB. The RIRs are truncated to $L = 2000$. Parameters for equalization are set to $L_i = L_c$, $\tau = 0$. In each run of the simulation, the relative geometry of the speaker and microphone array is kept constant, but the speaker is relocated at a random position in the room, and new error is randomly generated. The evaluation results are obtained by averaging 100 runs.

Table 6.2 shows the averaged C_{50} of the EIR and $\mathbf{h}_1$. It can be seen that for all room and source-microphone settings, the SIEREM improves the C_{50} by more than 10 dB for NMSE $= -30$ dB and more than 20 dB for NMSE $= -40$ dB. Table 6.3 shows the averaged T_{30} of the EIR and $\mathbf{h}_1$. It can be seen that for all room and source-microphone settings, the SIEREM also improves the T_{30}.

Table 6.2: The C_{50} of $\mathbf{h}_1$ and the EIR obtained with SIEREM for SSI.

NMSE (dB)	T_{60} (s)	dist (m)	C_{50} of EIR (dB)	C_{50} of $\mathbf{h}_1$ (dB)
-30	0.6	1	16.55	5.34
		2	14.45	3.79
		3	14.15	3.61
	0.8	1	16.96	3.02
		2	14.58	1.50
		3	14.23	1.40
-40	0.6	1	26.17	as above
		2	23.12	
		3	22.53	
	0.8	1	26.68	
		2	23.32	
		3	22.68	

Table 6.3: The T_{30} of $\mathbf{h}_1$ and the EIR obtained with SIEREM for SSI.

NMSE (dB)	T_{60} (s)	dist (m)	T_{30} of EIR (s)	T_{30} of $\mathbf{h}_1$ (s)
-30	0.6	1	0.153	0.234
		2	0.174	0.237
		3	0.175	0.238
	0.8	1	0.194	0.245
		2	0.214	0.246
		3	0.217	0.246
-40	0.6	1	0.065	as above
		2	0.080	
		3	0.083	
	0.8	1	0.068	
		2	0.094	
		3	0.102	

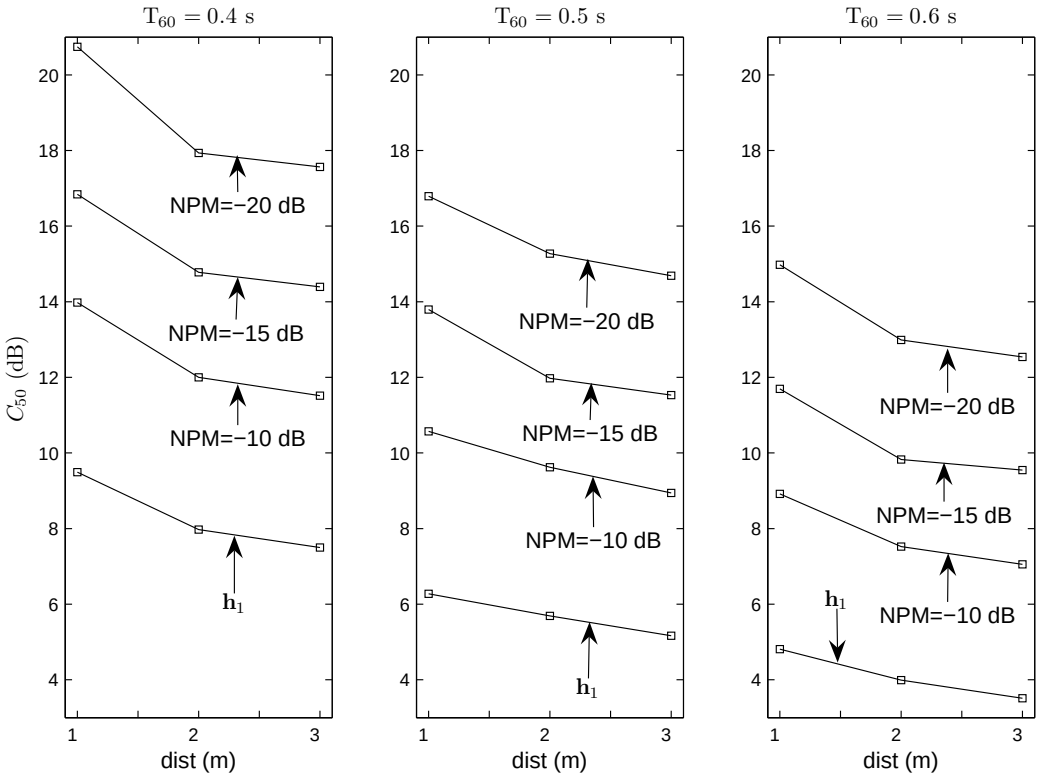

Figure 6.6: The C_{50} of $\mathbf{h}_1$ and the EIRs obtained with SIEREM for BSI.

Secondly, we present the evaluation results for BSI error. 2-channel systems with RIRs generated using the image method [36] are used as the acoustic systems $\mathbf{h}$. The room dimensions are 6.4 m $\times$ 5 m $\times$ 3.6 m; the distance between the speaker and the center of the microphone array is set to 1 m, 2 m and 3 m, and the inter-microphone distance is 5 cm; the reverberation time T_{60} is set to 0.4 s, 0.5 s and 0.6 s; the error vectors are generated with the algorithm presented in Section 6.3.1 and the NPM is set to -10 dB, -15 dB and -20 dB. Parameters for equalization are set to $L_i = L_c$, $\tau = 0$. In each run of the simulation, the relative geometry of the speaker and microphones is kept constant, but the speaker is relocated at a random position in the room. The averaged C_{50} of the resulting EIRs, which are obtained using 100 randomly chosen speaker positions and generated error vectors, are plotted in Figure 6.6.

We also use estimates $\hat{\mathbf{h}}$ obtained from actual BSI experiments using adaptive estimation of $\hat{\mathbf{h}}$ from NMCFLMS. In the 1000 $\hat{\mathbf{h}}$ obtained in the 1000 runs of the NMCFLMS in Section 6.1.2 for $T_{60} = 0.6$ s, we use the $\hat{\mathbf{h}}$ in the range of -9.95 dB $<$ NPM $<$ -10.05 dB, and set NPM $= -10$ dB to compute the corresponding equalization systems for each $\hat{\mathbf{h}}$ using SIEREM. In the simulations, 2-channel systems which include the $\mathbf{h}_m$ and $\hat{\mathbf{h}}_m$ of the first two channels are used. There are 52 $\hat{\mathbf{h}}$ in this NPM range, and we average the C_{50} of the 52 resulting EIRs. The averaged value of the C_{50} is 10.29 dB. The C_{50} obtained with generated error for the same room and source-to-microphone distance, which is already shown in Figure 6.6, is 8.92 dB. Comparing these two figures, we see that the C_{50} obtained with generated error is not substantially different from the one obtained with the true error. It can be seen in Figure 6.6 that SIEREM can always equalize the acoustic systems to good effect.

Finally, we apply SIEREM to the estimates of the acoustic systems used in Section 2.5 and use the resulting equalization system to equalize the true acoustic systems. These systems have been used in foregoing chapters to show the performance of different algorithms developed in this thesis and are employed here again. Parameters are set to $\tau = 0$, $L_i = L_c$. The EIR and its EDC for the SSI is shown in Figure 6.7. The EIR and its EDC for the BSI is shown in Figure 6.8. It can be seen that SIEREM can equalize the acoustic system for both SSI and BSI. The EDC for the SSI shows the effect of equalization in both the early part and the late part. In late part, the EDC is improved by up to 18 dB. The EDC for BSI in Figure 6.8 shows that the early part is less suppressed than the late part, which is due to the high NPM of the BSI error.

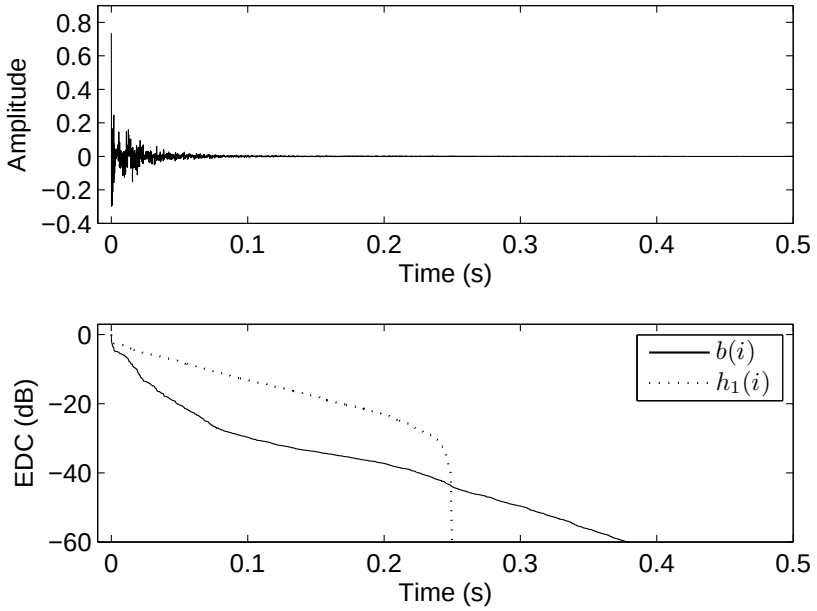

Figure 6.7: The EIR and EDCs of the EIR obtained using SIEREM and h_1 for SSI.

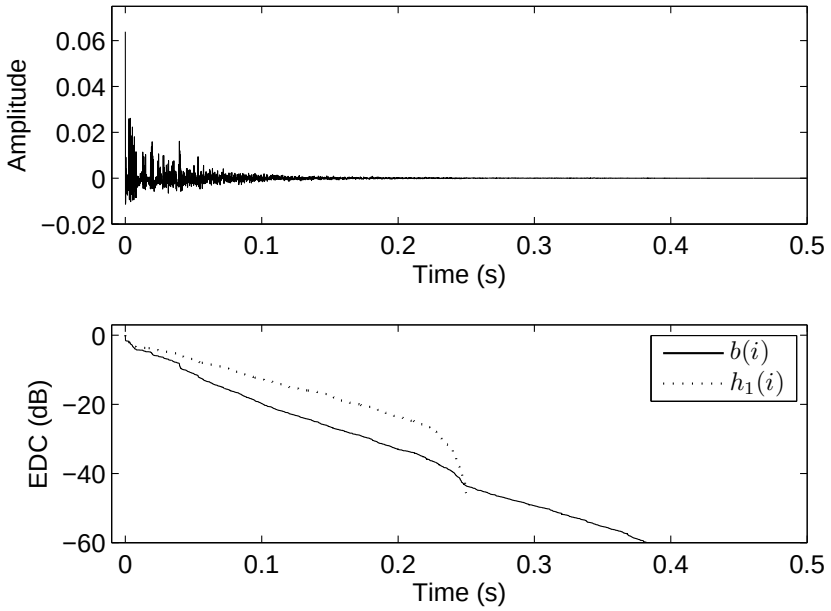

Figure 6.8: The EIR and EDCs of the EIR obtained using SIEREM and h_1 for BSI.

Table 6.4: The WMSE of EIR obtained with different NMSE estimates.

NMSE (dB)	$\widehat{\text{NMSE}}$ (dB)	WMSE (dB)	WMSE of $\mathbf{h}_1$ (dB)
-30	-20	-0.57	15.14
	-30	-1.42	
	-40	2.41	
-40	-30	-2.15	
	-40	-4.34	
	-50	-1.44	

6.3.3 Sensitivity of SIEREM to model parameters

As was mentioned in Section 6.2, the equalization system $\mathbf{g}$ in (6.21) depends on the NMSE for SSI, and NPM and the decay rate for BSI. However, in practice, the exact NMSE or the NPM and decay rate are not typically known. Therefore, to compute $\mathbf{g}$, these model parameters need to be estimated. In this section, the sensitivity of the SIEREM to the accuracy to which these parameters can be estimated is studied.

Firstly, we study the sensitivity of SIEREM to NMSE estimation for SSI. The WMSE, which is defined in (2.74) and equal to cost function (6.16) when the weighting function $u(i)$ is equal to $w(i)$, is computed for different NMSE estimates. In this work, to compute the WMSE, the $u(i)$ is set to $u(i) = w(i)$. The WMSE of $\mathbf{h}_1$ is computed after $\mathbf{h}_1$ is normalized by the amplitude of its direct-path response. The WMSE results for $\text{T}_{60} = 0.6$ s and distance between speaker and center of the microphone array 2 m are shown in Table 6.4. It can be seen that even with ± 10 dB error in NMSE estimates, the SIEREM still provides at least 12 dB improvements in WMSE for NMSE $= -30$ dB.

Secondly, we study the sensitivity of SIEREM to NPM and T_{60} (which is directly related to the decay rate α) estimation for BSI. 2-channel systems with RIRs generated using the image method [36] are used in the experiment. The room dimensions are 6.4 m $\times$ 5 m $\times$ 3.6 m; the distance between the speaker and the

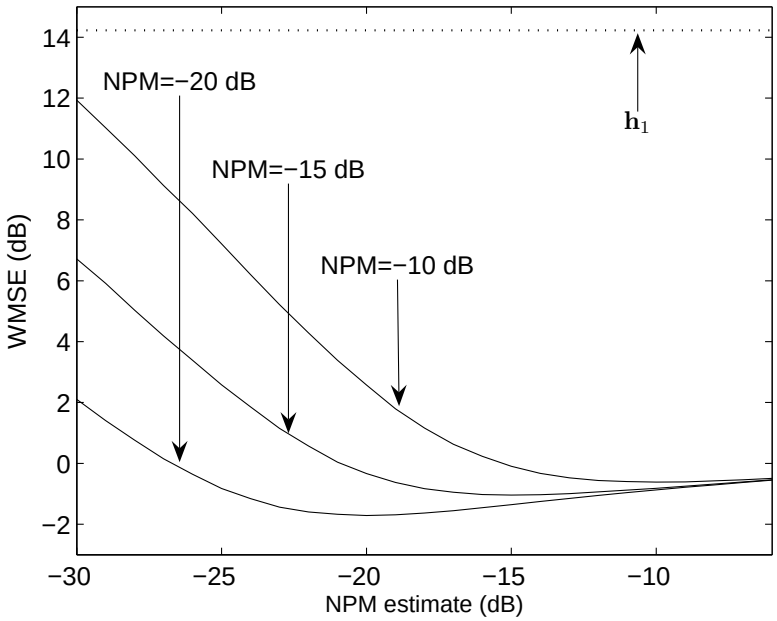

Figure 6.9: The WMSE as a function of the estimated NPM, demonstrates the sensitivity of the SIEREM to the errors in NPM estimates.

center of the microphone array is set to 2 m, and the inter-microphone distance is 5 cm. In the study of the sensitivity of SIEREM to NPM estimation, T_{60} is set to 0.6 s; the error vectors are generated with the algorithm presented in Section 6.3.1 and the NPM is set to -10 dB, -15 dB and -20 dB. The WMSE of the EIRs obtained with SIEREM for NPM estimates from -30 dB to -6 dB is calculated. Parameters for equalization are set to $L_i = L_c$, $\tau = 0$. In each run of the simulation, the relative geometry of the speaker and microphones is kept constant, but the speaker is relocated at a random position in the room. The averaged WMSE over 100 randomly chosen speaker positions and generated error vectors are plotted in Figure 6.9. It can be seen in Figure 6.9 that the SIEREM is not sensitive to NPM estimation. For example, for NPM $= -20$ dB, even with ± 10 dB error in NPM estimates, the SIEREM shows less than 4 dB performance degradation in WMSE.

In the study of the sensitivity of SIEREM to T_{60} estimation, the reverberation time is set to $T_{60} = 0.4$ s, $T_{60} = 0.5$ s and $T_{60} = 0.6$ s, and the NPM of generated

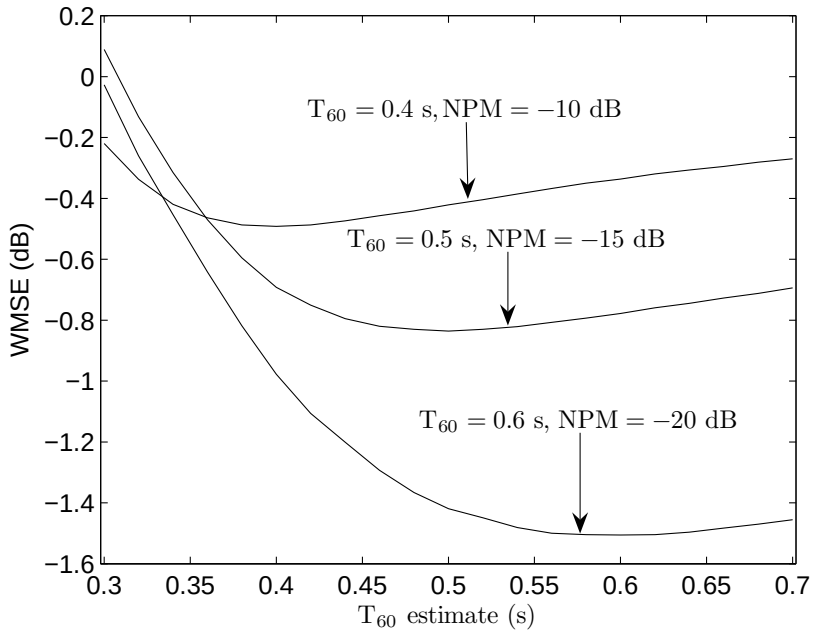

Figure 6.10: The WMSE as a function of the estimated T_{60}, demonstrates the sensitivity of the SIEREM to the error in T_{60} estimates.

error vectors is set to -10 dB, -15 dB and -20 dB respectively. The WMSE of the EIRs obtained with SIEREM for T_{60} estimates from 0.3 s to 0.7 s is calculated. Parameters for equalization are set to $L_i = L_c$, $\tau = 0$. The averaged WMSE over 100 randomly chosen speaker positions and generated error vectors are plotted in Figure 6.10. We can see in Figure 6.10 that the SIEREM is not sensitive to T_{60} estimation. $\pm 20\%$ error in T_{60} estimates causes less than 1 dB degradation in WMSE.

6.4 Summary

In this chapter, system identification error is modeled for both the SSI and BSI. A system-identification-error-robust equalization method exploring the use of the error models is presented. Evaluation results show that the SIEREM can give significantly beneficial equalization results for both SSI and BSI. SIEREM requires as input estimate of the NMSE for SSI, and estimates of T_{60} associated with the acoustic

system and an estimate of the NPM level for BSI, however, it has been shown that SIEREM is not significantly sensitive to the accuracy of these estimates.

Chapter 7

Conclusions

7.1 Summary and discussion

The aim of this thesis was to propose and investigate robust equalization methods for acoustic systems in the presence of system identification errors, in order for speech dereverberation. In Chapter 2, the fundamentals of system identification and equalization were reviewed. Characteristics of room acoustics were discussed and performance measures for equalization were accordingly defined. In Chapter 3, the MCLS method was investigated from the viewpoint of channel zeros. Two classes of characteristic zeros resulting in sensitive MCLS equalization systems were defined. The RMCLS method was proposed and investigated. The RMCLS is more robust than traditional MCLS. In Chapter 4, channel shortening for use in acoustic system equalization was investigated. A link between MINT, CS, and RMCLS was derived. A criterion for selecting a perceptually advantageous equalization system from the multiple solutions to channel shortening was provided. In Chapter 5, the OS-WCG algorithm was proposed for robust equalization of acoustic systems estimated using SSI. A conjugate gradient iterative method was employed and the peak index of WDRR in the iteration process was estimated, which led to an optimal equalization

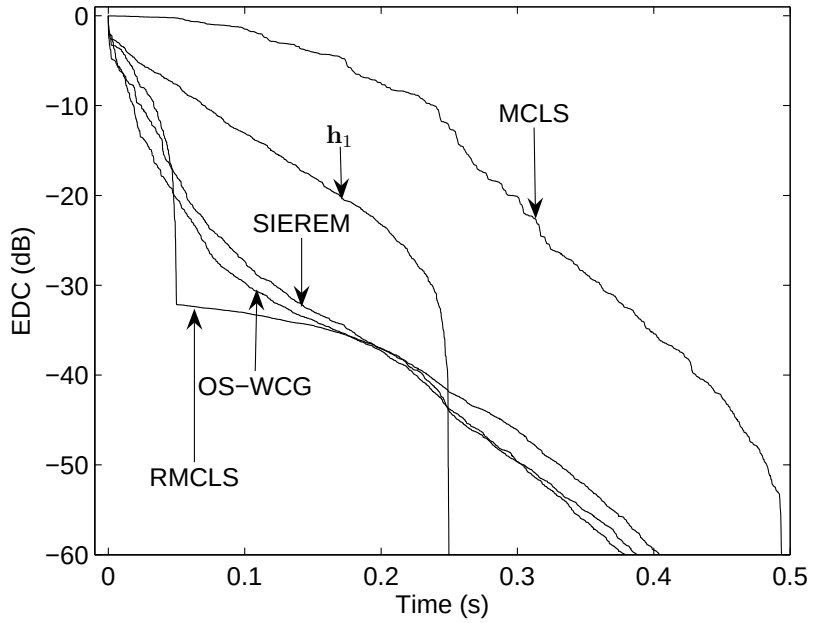

Figure 7.1: Comparison of the EDCs obtained using MCLS, RMCLS, OS-WCG and SIEREM for SSI.

system. In Chapter 6, an equalization method using models of SIEs was discussed. A study of BSI error relating to the BSI performance measure NPM was conducted. The SSI and BSI error models were incorporated in the least-squares formulation and the SIEREM was obtained.

As a summary, we compare these methods in terms of performance and computational complexity. The EDCs obtained by applying different methods to the systems and their estimates used in Section 2.5 are reproduced together in Figure 7.1 and Figure 7.2 for SSI and BSI respectively. It was found that for SSI, OS-WCG and SIEREM produce comparable results, which was also reflected by the evaluation results in C_{50} and T_{30} presented in Section 5.5.2 and Section 6.3.2. The RMCLS, OS-WCG, and SIEREM can all equalize the acoustic system. A difference of the EDC of RMCLS from those of OS-WCG and SIEREM is that it shows a sudden drop at 0.05 s. For BSI, the level of the EDC of RMCLS is about 10 dB less than that of SIEREM in the period shortly after 0.05 s, but the difference between the

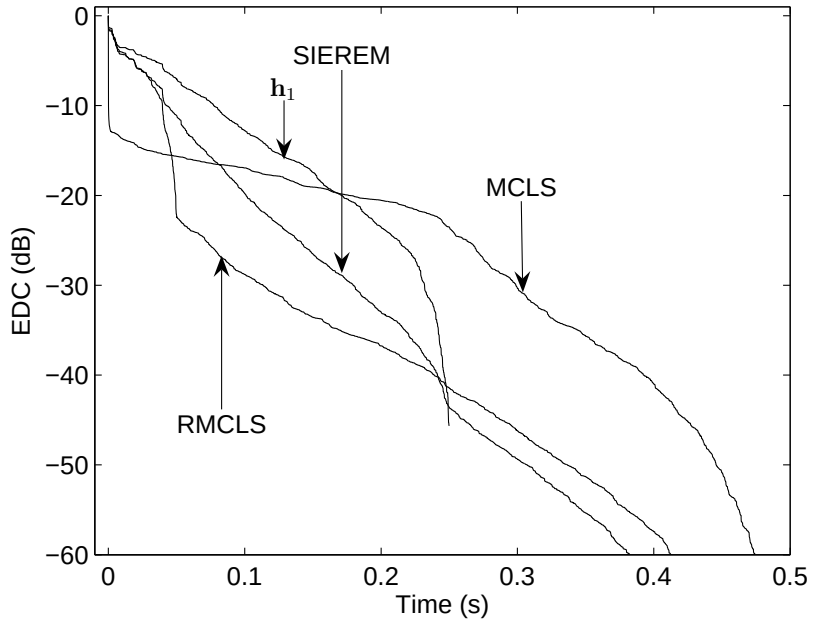

Figure 7.2: Comparison of the EDCs obtained using MCLS, RMCLS and SIEREM for BSI.

level of the EDCs of RMCLS and SIEREM decreases with time. Compared with RMCLS for BSI, speech resulted from SIEREM sounds more natural. (It sound like speech captured in a room which is less reverberant than the original room, whereas the RMCLS introduces colouration due to the sudden drop at 0.05 s.)

As for computational complexity, both MCLS and RMCLS need singular value decomposition (SVD), CS needs QZ decomposition, which result in high computation load for these algorithms. On the other hand, OS-WCG only needs one multiplication of a matrix by a vector in each iteration, and SIEREM only needs a matrix inversion. Considering a simple case, in which $\mathbf{H}$ is a square matrix (which can be achieved by, for example, $M = 2$ and $L_i = L_c$) with size $\iota \times \iota$, the computational complexity of these algorithms in terms of floating point operations (flops) [70] is shown in Table 7.1. The results in Table 7.1 are obtained using the computational complexity analysis of SVD, QZ, and matrix inversion via Gaussian elimination given in [70]. The complexity of OS-WCG depends on the NMSE of the SSI error. Lower

Table 7.1: Computational complexity comparison.

	computational complexity (flops)
MCLS	$25\iota^3$
RMCLS	$25\iota^3$
CS	$48\iota^3$
OS-WCG	$2\iota^3 + 2\iota^2 \times number\ of\ iterations$
SIEREM	$(7/3)\iota^3$

NMSE needs more iterations. As an example, a typical number of iterations needed for NMSE $= -30$ dB is 130. It can be seen that the computational complexities of OS-WCG and SIEREM are 10-20 times lower than RMCLS or CS.

7.2 Future directions

In this section, we discuss some future work directions arising from this thesis.

- In Chapter 3, it was shown that the characteristic zeros causing strong peaks in the frequency responses of the components of equalization systems obtained from MCLS did not cause peaks in the RMCLS counterparts. Replacements of the characteristic zeros, which were near the characteristic zeros, could be found in the zeros of the EIR resulted from RMCLS. These results were shown by experiments. Mathematical explanations for them are desired.

- In Chapter 4, the generalized eigenvalue problem was solved and multiple solutions that maximize the quotient in (4.1) were obtained. Then the solution which results in the minimum ℓ_2-norm EIR was regarded as a perceptually advantageous equalization system and was selected for use. It might be possible to formulate the maximization of the generalized Rayleigh quotient in (4.1) and finding the CS solution which leads to the minimum ℓ_2-norm EIR in one optimization problem.

- The OS-WCG algorithm was developed for SSI in Chapter 5. One characteristic of OS-WCG is that the iterative method enables the weighting function, which is not effective in the closed-form WMCLS. Another characteristic is that the peak index of WDRR in the iteration process is estimated, which results in an optimal equalization system. Applying the iterative method to BSI, the WDRR also showed a peak. An approach to estimate the peak index for BSI is not yet available and is to be explored.

- In Chapter 6, the SIEREM needs NPM of the BSI error as an input parameter. However, method for blind estimation of NPM is not yet available and the blind estimation of NPM is still an open question.

- In our work, NPM was used as the measure for BSI quality. This measure characterizes the angle between the true acoustic system and the estimated system and is considered to be a consistent measure among all measures base on ℓ_2-norm distance. However, no evidence shows that an ℓ_2-norm distance based measure can capture all the characteristics of BSI. Therefore, other measures are desired to be explored. This would be helpful for both the peak index estimation for BSI in OS-WCG and BSI error model improvement for SIEREM.

- Due to the high computational complexity of RMCLS and CS, which need singular value decomposition and QZ decomposition respectively for large matrix, we were not able to produce systematic evaluation of RMCLS and CS. This could be carried out in future work.

- For both OS-WCG and SIEREM, weighting function (5.5) was used. Since it was desired that the decay rate of the EIR was not larger than that of the RIR, the parameter α controlling $w(i)$ in (5.5) was set according to (5.19). However, α is not necessarily related to the T_{60}, and (5.5) is not necessarily

used. Different α or different weighting functions can result in EIRs leading to speech with different characteristics. Performance of OS-WCG and SIEREM with different values of α or with different weighting functions is interesting. In this thesis, the performance of the developed algorithms was evaluated with limited design parameter settings. Their performance for parameters such as τ and L_i set to different values is relevant.

- The main objective of this work was to improve the intelligibility of reverberant speech. The performance of the algorithms was evaluated with objective measures, which are considered to be correlated to subjective perception. Evaluation confirmed the improvements made by the algorithms. However, it is known that objective measures have not yet been able to characterize the whole picture of speech perception. Therefore, formal listening tests could be organized.

Bibliography

[1] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, vol. 78, no. 5, pp. 1508–1518, 1985.

[2] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 425–437, 1997.

[3] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationary with application to speech," *IEEE Trans. Signal Processing*, vol. 49, pp. 1614–1626, 2001.

[4] T. Langhans and H. Strube, "Speech enhancement by nonlinear multiband envelope filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 7, 1982, pp. 156–159.

[5] H. Attias and L. Deng, "Speech denoising and dereverberation using probabilistic models," in *Proc. Conf. Neural Information Processing Systems*, vol. 13, 2001, pp. 758–764.

[6] E. A. P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 4, 2005, pp. 173–176.

[7] T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity based blind dereverberation for single channel speech signals," vol. 15, pp. 80–95, 2007.

[8] J. R. Hopgood and P. J. W. Rayner, "Blind single channel deconvolution using nonstationary signal processing," *IEEE Trans. Speech Audio Processing*, vol. 11, no.5, pp. 476–488, 2003.

[9] T. S. Bakir, "Blind adaptive dereverberation of speech signals using a microphone array," Ph.D. dissertation, Georgia Institute of Technology, 2004.

[10] K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans. Speech Audio Processing*, vol. 15, pp. 1579–1591, 2007.

[11] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1982.

[12] M. Miyoshi and K. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 145–152, 1988.

[13] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

[14] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Processing*, vol. 13, pp. 882–895, 2005.

[15] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, pp. 165–169, 1979.

[16] L. Ljung, *System identification, second edition.* Prentice-Hall, Inc, 1999.

[17] S. M. Kay, *Fundamentals of statistical signal processing.* Prentice Hall PTR, 1993, vol. 1.

[18] J. Vanderkooy, "Aspects of MLS measuring systems," *J. Audio Eng. Soc.*, vol. 42, pp. 219–231, 1994.

[19] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 7360–377, 1989.

[20] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conference on Signals, Systems, and Computers*, vol. 2, 1991, pp. 856–860.

[21] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 1994.

[22] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, 1995.

[23] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, 1995.

[24] R. Liu, "Blind signal processing: An introduction," in *IEEE Int. Symposium Circuits and Systems*, vol. 2, 1996, pp. 81–83.

[25] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, pp. 83–99, 1996.

[26] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, 1998.

[27] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Processing*, vol. 82, pp. 1127–1138, 2002.

[28] ——, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Processing*, vol. 51, pp. 11–24, 2003.

[29] Y. Huang, J. Benesty, and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Processing*, vol. 86, pp. 1278–1295, 2006.

[30] F. J. MacWilliams and N. J. A. Sloane, "Pseudo-random sequences and arrays," *Proc. IEEE*, vol. 64, pp. 1715–1729, 1976.

[31] J. Benesty, M. M. Sondhi, and Y. Huang, Eds., *Springer handbook of speech processing*. Springer, 2008.

[32] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Lett.*, vol. 5, pp. 174–176, 1998.

[33] P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *Research Letters in Signal Processing*, vol. 2009, 2009.

[34] H. H. Yang and Y. Hua, "On performance of cross-relation method for blind-channel identification," *Signal processing*, vol. 62, pp. 187–205, 1997.

[35] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2006.

[36] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[37] H. Kuttruff, *Room acoustics*, 4th ed. Taylor & Frances, 2000.

[38] R. Rado, "Note on generalized inverse of matrices," *Proc. Cambridge Philos. Soc.*, vol. 52, pp. 600–601, 1956.

[39] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. Signal Processing*, vol. 46, pp. 215–218, 1998.

[40] M. Hofbauer, "Optimal linear separation and deconvolution of acoustical convolutive mixtures," Ph.D. dissertation, Swiss Federal Institute of Technology, 2005.

[41] T. N. E. Greville, "Note on the generalized inverse of a matrix product," *SIAM Review*, vol. 8, pp. 518–521, 1966.

[42] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, 1996.

[43] M. Nafie and A. Gatherer, "Time-domain equalizer training for ADSL," in *IEEE Int. Conf. Commun.*, 1997.

[44] N. Al-Dhahir, "FIR channel shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, pp. 213–218, 2001.

[45] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Processing Lett.*, vol. 9, pp. 341–343, 2002.

[46] R. Nawaz and J. A. Chambers, "Blind adaptive channel shortening by single lag autocorrelation minimisation," *Electronics letters*, vol. 40, no. 25, 2004.

[47] R. K. Martin, J. M. Walsh, and C. R. Johnson Jr., "Low-complexity MIMO blind, adaptive channel shortening," *IEEE Trans. Signal Processing*, vol. 53, pp. 1324–1334, 2005.

[48] R. K. Martin, K. Vanbleu, M. Ding, G. Ysebaert, M. Milosevic, B. L. Evans, M. Moonen, and C. R. Johnson Jr., "Unification and evaluation of equalization structures and design algorithms for discrete multitone modulation systems," *IEEE Trans. Signal Processing*, vol. 53, pp. 3880–3894, 2005.

[49] M. Kallinger and A. Mertins, "Impulse response shortening for acoustic listening room compensation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2005.

[50] ——, "Multi-channel room impulse response shaping - a study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2006.

[51] J. N. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vib.*, vol. 146, no. 2, pp. 281–302, 1991.

[52] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 320–328, 1994.

[53] J. D. Polack, "La transmission de l'energie sonore dans les salles," Ph.D. dissertation, Universite du Maine, 1988.

[54] W. Reichardt and U. Lehmann, "Raumeidruck als oberbegriff von raumlichkeit und halligkeit, erlauterungen des raumeindrucksmasses," *Acustica*, vol. 40, pp. 174–183, 1978.

[55] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, 2007.

[56] W. C. Sabine, in *The American Architect 1900, Collected Papers on Acoustics No. 1.* Harvard University Press, Cambridge, 1923.

[57] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[58] H. Haas, *Acustica*, vol. 1, p. 49, 1951.

[59] W. Reichardt, A. O. Abdel, and W. Schmidt, "Abhangigkeit der grenzen zwischen brauchbarer und unbrauchbarer durchsichtigkeit von der art des musikmotivs, der nachhallzeit und der nachhalleinsatzzeit," *Appl. Acoustics*, vol. 7, p. 243, 1974.

[60] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[61] P. Jinachitra and R. E. Prieto, "Towards speech recognition oriented dereverberation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2005.

[62] C. Tantibundhit, F. Pernkopf, and G. Kubin, "Speech enhancement based on joint time-frequency segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2009.

[63] G. A. Jones and J. M. Jones, *Elementary Number Theory.* Springer-Verlag, 1998.

[64] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition.* MIT Press and McGraw-Hill, 2001.

[65] X. Lin, N. D. Gaubitch, and P. A. Naylor, "Two-stage blind identification of SIMO systems with common zeros," in *Proc. European Signal Processing Conf.*, 2006.

[66] A. W. H. Khong, X. Lin, and P. A. Naylor, "Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2008.

[67] R. Casas, F. Lopez de Victoria, I. Fijalkow, P. Schniter, and T. J. Endrea, "On MMSE fractionally-spaced equalizer design," in *Proc. 13th Int. Conf. Digital Signal Processing*, 1997.

[68] C. B. Moler and G. W. Stewart, "An algorithm for generalized matrix eigenvalue problems," *SIAM J. Number. Anal.*, vol. 10, no. 2, pp. 241–256, 1973.

[69] S. S. Haykin, *Adaptive filter theory, 4th edition.* Prentice Hall, 2002.

[70] G. H. Golub and C. F. van Loan, *Matrix computations*, 3rd ed. London: John Hopkins University Press, 1996.

[71] R. Ahmad, N. D. Gaubitch, and P. A. Naylor, "A noise-robust dual filter approach to multichannel blind system identification," in *Proc. European Signal Processing Conf.*, 2007.

[72] R. Ratnam, D. L. Jones, B. C. Wheeler, J. William, D. O'Brien, C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Amer.*, vol. 114, no. 5, pp. 2877–2892, 2003.

[73] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2008.

[74] J. H. Wilkinson and C. Reinsch, *Linear algebra*. Springer-Verlag, 1971.

List of Publications

[P-1] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," to appear in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2010.

[P-2] W. Zhang, E. A. P. Habets, and P. A. Naylor, "Equalization of multichannel acoustic systems in the presence of system identification errors," in preparation for *IEEE Trans. Audio, Speech, Language Processing*.

[P-3] W. Zhang, E. A. P. Habets, and P. A. Naylor, "A system-identification-error-robust method for equalization of multichannel acoustic systems," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2010.

[P-4] W. Zhang, A. W. H. Khong, and P. A. Naylor, "Acoustic system equalization using channel shortening techniques for speech dereverberation," in *Proc. 17th European Signal Processing Conf.*, 2009.

[P-5] W. Zhang, and P. A. Naylor, "An experimental study of the robustness of multichannel inverse filtering systems to near-common zeros," in *Proc. 17th European Signal Processing Conf.*, 2009.

[P-6] W. Zhang, A. W. H. Khong, and P. A. Naylor, "Adaptive inverse filtering of room acoustics," in *Proc. 42nd Asilomar Conf. Signals, Systems, and Computers*, 2008.

[P-7] W. Zhang and P. A. Naylor, "An algorithm to generate representations of system identification errors," *Research Letters in Signal Processing*, 2008.

[P-8] W. Zhang, N. D. Gaubitch, and P. A. Naylor, "A computationally effective inverse filtering algorithm for speech dereverberation robust to system estimation errors," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, 2008.