uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Scott G. Norcross

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Electrical Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Evaluation and Alternative Regularization of Audio Equalization Techniques

TITRE DE LA THÈSE / TITLE OF THESIS

Martin Bouchard

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

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

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Miodrag Bolic

Wail Gueaieb

Rafik Goubran

Roch Lefebvre

Gary W. Slater

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

Evaluation and Alternative Regularization of Audio Equalization Techniques

by

Scott G. Norcross

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Doctor of Philosophy degree in
Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-59532-9
Our file Notre référence
ISBN: 978-0-494-59532-9

Abstract

Inverse-filtering techniques are used in audio equalization of rooms and loudspeakers to correct or enhance the listening environment and to create virtual reality audio systems. The inverse-filtering concept is based on the concept that one can undo a system (loudspeaker, room, or some audio/acoustic path) completely or modify it to some desired/target response. Regularization is commonly used in the calculation of inverse filters to reduce the effort done by the inverse filter. The regularization is needed when the system is ill-conditioned and/or when zeros are present in the transfer function. Current regularization methods applied can create an audible pre-response to the corrected time-domain impulse responses, thus are not always suitable to high-quality audio systems. This thesis reviews some known FIR single and multichannel inverse filtering methods, and evaluates them objectively and subjectively, highlighting problems with relating to audible artifacts. New attempts are investigated for reducing the pre-response caused by regularization in single and multichannel inverse filtering systems, by adjusting the phase of the regularization terms to be minimum or partially minimum phase. The minimum-phase regularization method is shown to reduce the pre-response and therefore improve the audio quality of the inverse filters.

Acknowledgements

The completion of this thesis has been a long undertaking and I would like to thank very deeply my supervisor Martin Bouchard who took me on as a graduate student while I was still working full-time. Without his patience, dedication and expertise I would not have be able to carry out this research and complete this thesis. Thank you so very much Martin!

A thesis takes lots of time and energy to complete and lots of help directly and indirectly from co-workers, family and friends. First I would like to thank my colleagues at the CRC for the input and help in carrying out the research and subjective tests. To all my friends who were there along on this long journey whether they knew it or not. A special thanks to Andrew, who could always cheer me up with some long talks about music, computers or anything at all (and to his wife Bronwyn who endured some of them as well!). Whether it was in Woodstock, Ottawa, Toronto or even Moree, he could always get put me on the right path. A thanks to Dr. Jason Hinek, who provided moral support in the final stretch of writing this thesis and could always provide some for of music recommendation!

I owe a big thank you to my Aunt (or Auntie) Lorna, who always encouraged me to continue on with my academic career.

A very special thanks to my sister-in-law Sylvia and brother Kevin for having those late night discussions from lac beetles to marsh mallow root and of course live-long continuous support of my journeys. A special thank you to my niece Lilly, who always brings me joy and I can only hope she gets to have the opportunities I have had.

My parents Greg and Muriel, who always give me support and encouraged me throughout my life long pursuits. No words can express how how thankful I am Mom and Dad!!!

I thank you to Alexander, who always provided me with much needed distractions such as playing video games or hockey, and for just being around during my long working hours. I hope you also get the opportunities I have had in my life.

Finally, I would like to thank my wife Tara, who has endured my long working hours on research and this thesis, and was always there to help me through it in some form on another. There is really no words that can be written to show how much you are apart of this thesis too!

Dedication

I was lucky to have four living grandparents while I was growing up. They always provided me with love and support in all my endeavors . Unfortunately, they will not see me finish this thesis and receive Ph.D. but they are always in my memory and heart. Thank you Grandma and Grandpa Norcross, and Grandpa and Grandma Neufeld and I dedicate this thesis to them.

Without my parents I would not have had the opportunities I had in terms of education and life and would not have be able to complete this thesis. Therefore, I also dedicate this thesis to my parents Greg and Muriel Norcross.

Finally I dedicate this thesis to my late brother Gary....

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

1.1 Equalization and Correction Filters

An equalizer is defined as a "device designed to compensate for an undesired amplitude-frequency or phase-frequency characteristic, or both" of a system [2]. It usually consists of a filter or series of filters to adjust the frequency dependent variable either up or down.

The first use of equalizers in the audio field was to improve the sound quality of motion pictures in the 1930s [4]. John Volkman, who worked for RCA, is credited as being the first person to use a variable equalizer to improve reproduced sound. Volkman worked at Bell Labs where he used fixed equalizers for correcting audio transmission losses.

Fixed equalizers, where the characteristics of the filters do not change, are commonly used in preemphasis/deemphasis schemes in audio. For example, the RIAA (Recording Industry Association of Amercia) equalization curve for phonograph preamplifiers. The purpose of the preemphasis/deemphasis scheme is to improve the effective signal-to-noise ratio of an audio system. Noise introduced in systems is typically white in spectral shape, and will result in more high frequency noise to the human listener. If the signal is boosted in the high-frequencies before the addition of the noise from the system and then reduced by the same amount after the audio system, but before the listener hears it, the high frequency noise added by the system will be reduced in level.

Variable equalizers allow the user to change certain parameters such as frequency, bandwidth or gain. Variable equalizers are traditionally used in audio applications, room and loudspeaker equalization where user control of parameters is needed. Two common types of variable equalizers used in audio applications are graphic and parametric. A

graphic equalizer breaks the frequency spectrum into banks or bands and the user can adjust the individual gain on each of the frequency bands. Typically, the bands are broken up into octave or third-octaves and cover the audible spectrum, 20 Hz to 20 kHz. A parametric equalizer provides more precise control of the equalization by allowing the user to alter the centre frequency and bandwidth of the filter banks/bands.

The graphic and parametric types of equalizers that are described above only adjust the magnitude of the signal and not the phase. They typically use minimum phase filters to adjust the amplitude of the signal and are not concerned with any phase correction. As mentioned above, these equalizers are typically used for loudspeakers or for equalization in sound mixing or recording. With the advent of digital signal processing (DSP), the concept of equalizing the magnitude and phase became possible. The basic premiss is to use an inverse filter calculated from a measured or modeled system, such as a loudspeaker or room. This inverse filter "undoes" or corrects the magnitude and phase response errors of the system to give a desired result. This type of equalization is commonly referred to as "digital" equalization, though this can be misleading since there are digital graphic and parametric equalizers available. The term "correction" is also used to describe this type of equalization or filtering as it attempts to remove or correct any of the linear distortion that the room or loudspeaker (or some system) adds to the signal.

One of the first uses of inverse filtering was with acoustic crosstalk cancelation by Atal and Schroeder which is described in their patent "Apparent Sound Source Translator" [1]. An acoustic crosstalk canceler pre-filters the signal to a loudspeaker to remove the acoustic signal at the listening position from other loudspeakers in the playback setup. Typically, as in Atal and Schroeder's case, a two loudspeaker setup is used which is commonly referred to as a "stereo-dipole" [24]. The reason for removing the acoustic path from the other loudspeakers is to ensure that the signal that is sent to the loudspeakers is what arrives to the listener's ears. This also allows for a controlled playback system that can be used as virtual source imaging systems or for binaural playback.

The Atal and Schroeder [1] acoustic crosstalk canceler consisted of two loudspeakers placed 30° to the left and right of the centre line and was initially used for enhanced stereo hi-fi applications. Schroeder [41] later describes in more detail the filters that are required for the acoustics crosstalk canceler and one use of the system was for subjective evaluation of concert halls. Earlier subjective evaluations were carried out in the actual halls, and there could be days or weeks between comparisons, resulting in highly varied results. By using this two-loudspeaker layout, one could instantly switch between different recordings made from two different concert halls.

Instead of trying to recreate a different soundfield from the current location of the loudspeaker setup, Mourjopoulos [30], Bean and Craven [3], Craven and Gerzon [8] and Howe and Hawksford [16] proposed methods and strategies for equalizing or correcting an audio system and/or room for regular stereo listening. The term correcting refers to modifying the system response in such a way to remove unwanted characteristics of it, such as room reflections or frequency-response deviations. Mourjopoulos [30] uses the inverse filter in correcting an impulse response of a room or audio system. He defines the inverse of an impulse response as a filter which results in a delta function (a function that is "non-spreading in time and has a flat spectrum and linear phase") when linearly convolved with the impulse response. Due to the non-minimum phase nature of many systems, including rooms as pointed out by Neely et al. [31] and crossovers in loudspeakers, a true causal inverse is not realizable. A delay is introduced in the inversion process so that the result is causal. A least-squares filter design technique can be used to best approximate the inverse filter of such systems.

Craven and Gerzon [8] present the detail of a room and loudspeaker equalizer scheme highlighting the mixed phase nature of the responses. Their method equalizes the room and loudspeaker separately since the room is the dominant non-minimum phase part of the system. Therefore, only the minimum phase part of the room is inverted, to avoid pre-response artifacts that occur when inverting a non-minimum phases system. Pre-echo, pre-ringing or any pre-response is more audible than any post-echo due to the masking characteristics of human hearing. Thus, one tends to avoid any processing that creates any pre-echo and Genereux [10] suggests limiting any pre-responses to less than 2 ms. A causal minimum phase filter has no pre-response and the inverse of a causal minimum phase filter is also causal and minimum phase, thus avoiding any pre-response.

A potential problem with any equalization method is the application of too much gain by the equalizing; for example, if a correction is to be made to a loudspeaker that has a large dip in the frequency response. The inverse filter will try and compensate for the dip by creating a large boost. This can have numerous negative effects on the performance. The first thing is that the inverse filter could have a very long time response leading to a large inverse filter length, which can in turn lead to time wrapping effects and pre-echo when frequency-domain inverse methods are used. Secondly, a large boost in an audio system could overload some part of the audio system and cause it to distort. Finally, boosts are more audible [44] than dips in the frequency response. If the dip is caused by certain reflections in a room the location of the dip in frequency could be position dependent. At the one point in space, the equalization will correct for the dip, but away

from that point an audible boost could be present and be audible. Another potential limitation of the inverse filter is when the entire frequency range does not need to be corrected. For example, a loudspeaker has a low-frequency and high-frequency roll-off where it cannot output any or very little energy. It seems pointless to try and equalize (or correct) below or above the natural roll-offs of a device. By doing so one could potentially overdrive the loudspeaker and cause it to distort. Therefore it is desirable to limit the amount of work the inverse filter must do to try and avoid some of these potential problems.

One approach to limit the work the inverse filter must do is to use regularization. Regularization has been used by Craven [8] and Kirkeby [23] to limit the effort done by the inverse filter. Craven [8] uses "amplitude" regularization based around the mean value of the spectrum, and a regularization constant that can be selected depending on the degree of regularization needed. Kirkeby [23] defines the regularization based on trial and error so that optimum audio performance is achieved. Both methods can be defined to be frequency dependent to account for any frequency-dependency that is desired. In Fielder [9] and in other previous studies [37], it has been shown that the regularization must be selected carefully and that the regularization can actually degrade the audio quality. It is therefore important to use these techniques with care to avoid audible degradation in the audio signal.

Another method of limiting the amount the inverse filter works with respect to correcting dips and peaks is to smooth the frequency response. By smoothing the frequency response one can eliminate large peaks and dips that could potentially cause problems in the inversion process. A one-third octave bandwidth is a typical width for smoothing as it is similar to the critical bandwidth of the human ear. For smoothing impulse responses, a magnitude smoothing technique does not smooth the phase. A phase must be defined if one wants to retrieve the time-domain response to create an inverse filter. Lipshitz [27] suggests that a zero-phase component can be added to the smoothed magnitude. Hatziantoniou et al. [12] suggested a complex method of smoothing where the complex spectrum is smoothed and they have shown success when using this technique with room impulse responses [14] [13]. Panzer [38] utilizes a continuous phase smoothing method that claims to be successful for loudspeaker equalization.

Inverse filtering, or rather cancelation filters, are often used in noise control applications [26]. These applications usually involve adaptive filters that adjust the filter coefficients to a changing soundfield so that there is maximum cancelation according to some error criteria. These methods are usually not associated with high quality audio

applications.

1.2 Thesis Outline

This thesis will review (Chapter 2) common FIR methods of inverse filtering for single and multichannel scenarios that will include time-domain solutions as well as more efficient frequency-domain methods. A general Wiener filter approach is shown to be equivalent to various deterministic methods found in the literature [23] [25], such as the time-domain least squares approach when the system is purely deterministic.

The results of formal subjective tests using single-channel inverse filtering techniques that have been previously published are compared and reviewed in Chapter 3. These include a time-domain least-squares method and a frequency-domain deconvolution method, along with various regularization and smoothing methods. The formal subjective tests highlighted some of the problems with the various methods of inverse filtering. The frequency-domain deconvolution method is used because of its computational efficiency, but is prone to artifacts from time-aliasing due to the block processing nature of frequency-domain processing. These block effects can be eased by limiting the effort done by the inverse filter in trying to correct the system that is being inverted. One such way of doing this is to add regularization to the inverse filter, but how much regularization to add is dependent on the impulse response or system being corrected. Also, the regularization itself can lead to audible artifacts such as some pre-response.

Chapter 4 introduces and develops methods of reducing the pre-response created by regularization for single and multichannel systems. The proposed methods are based on the principle of forcing the regularization terms to be minimum phase or a mix of minimum- and zero-phase terms, to reduce the pre-response that is created by the regularization. The methods are motivated from the time-domain masking characteristics of the human ear.

Examples and results for the methods proposed in Chapter 4 are included in Chapter 5, illustrating the new method to effectively reduce the pre-response of regularization.

A summary is included in the chapter 6, and further ideas for future work are also included.

1.3 Contributions

The main contributions of this thesis are:

- Showing the links between the different methods, from stochastic Wiener solutions to deterministic least-squares, and data-matrix formulations. (Chapter 2)
- Design of a method for subjectively evaluating different inverse filtering methods, and then carrying out a study on previous published methods to understand their short comings in terms of audio quality. (Chapter 3)
- Development of a new mono-channel equalization method that helps reduce the pre-response created by classic regularization. (Chapter 4)
- Development of a new multichannel equalization method, based on a similar approach as the mono-channel method, but adapted for multichannel. (Chapter 4)
- Evaluation of the new proposed methods. (Chapter 5)

1.3.1 Publications

Referred Journal Papers

- S. G. Norcross, G.A. Soulodre and M.C. Lavoie, "Subjective Investigations of Inverse Filtering", *J. Audio Eng. Soc.*, Vol. 52, No. 10, (October 2004).
- G. A. Soulodre, M. C. Lavoie and S. G. Norcross, "Objective Measures of Listener Envelopment in Multichannel Surround Systems", *J. Audio Eng. Soc.*, vol. 51, no. 9 (September 2003).
- M. Bouchard and S. Norcross, "Computation Load Reduction of Fast Convergence Algorithms for Multichannel Active Noise Control", *Signal Processing, Vol.*, vol. 83, Issue 1, Jan 2003, pp.121-134.

Conference Proceedings and Abstracts

- S. G. Norcross and M. Bouchard, "Multichannel Inverse Filtering With Minimal-Phase Regularization", 123^{th} Convention of the Audio Engineering Society, Preprint 7265.
- S. G. Norcross, M. Bouchard, and G. A. Soulodre, "Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization", 121^{th} Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 54 , No. 12, (Dec. 2006), Preprint 6929.

- S. G. Norcross, M. Bouchard, and G. A. Soulodre, "Adaptive Strategies for Inverse Filtering", 119th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 53, No. 12, (Dec. 2005) Preprint 6563.

- S. G. Norcross, "Aspects of Inverse Filtering for Loudspeakers", Canadian Acoustics Conference (Ottawa), ct. 2004), *Canadian Acoustics*, Vol. 32, No. 3, September 2004.

- S. G. Norcross, G.A. Soulodre and M.C. Lavoie, "Distortion Audibility in Inverse Filtering", 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 53 , No. 1/, (Jan/Feb 2005), Preprint 6311.

- Scott G. Norcross, Gilbert A. Soulodre and Michel C. Lavoie, "Further Investigations of Inverse Filtering", 115th Audio Engineering Society Convention, , *J. Audio Eng. Soc. (Abstracts)*, vol. 51 , No. 12, (Dec. 2003), Preprint 5923.

- Gilbert A. Soulodre and Scott G. Norcross, "Objective Measures of Loudness", 115th Audio Engineering Society Convention, , *J. Audio Eng. Soc. (Abstracts)*, vol. 51 , No. 12, (Dec. 2003), Preprint 5896.

- Gilbert A. Soulodre, Michel C. Lavoie and Scott G. Norcross, "The Subjective Loudness of Typical Program Material", 115th Audio Engineering Society Convention, , *J. Audio Eng. Soc. (Abstracts)*, vol. 51 , No. 12, (Dec. 2003), Preprint 5892.

- Gilbert A. Soulodre, Michel C. Lavoie and Scott G. Norcross, "Objective Measures of Listener Envelopment in Multichannel Surround Systems", 24th Audio Engineering Society Conference:Multichannel Audio, The New Reality, June 2003.

- Scott G. Norcross, Gilbert A. Soulodre and Michel C. Lavoie, "Subjective Effects of Regularization on Inverse Filtering", 114th Audio Engineering Society Convention, *J. Audio Eng. Soc. (Abstracts)*, vol. 51, No. 5, (May 2003), Preprint 5848.

- Gilbert A. Soulodre,Michel C. Lavoie and Scott G. Norcross, "Temporal Aspects of Listener Envelopment in Multichannel Surround Systems", 114th Audio Engineering Society Convention, *J. Audio Eng. Soc. (Abstracts)*, vol. 51, No. 5, (May 2003), Preprint 5803.

- Scott G. Norcross, Gilbert A. Soulodre and Michel C. Lavoie, "Subjective Evaluation of Inverse Filtering Techniques for Room/Speaker Equalization", 113th Audio Engineering Society Convention, *J. Audio Eng. Soc. (Abstracts)*, vol. 50, No. 11, (Nov. 2002), Preprint 5662.

- Gilbert A. Soulodre, Michel C. Lavoie and Scott G. Norcross, "Investigation of Listener Envelopment in Multichannel Surround Systems", 113th Audio Engineering Society Convention, *J. Audio Eng. Soc. (Abstracts)*, vol. 50, No. 11, (Nov. 2002), Preprint 5676.

- M. Bouchard and S. Norcross, "A Comparison of Algorithms and The Development of a Fast Convergence and Reduced Computational Load Algorithm for Multichannel Active Noise Control", 143^{rd} Acoustical Society of America Meeting, Pittsburgh, Pennsylvania, June 2002.

Chapter 2

Inverse Filtering Background

In this chapter the inverse filtering concept is introduced and time- and frequency-domain methods of inverse filtering methods are explored. It is shown that the general Wiener filtering approach can be expressed in purely deterministic terms when the system is deterministic. This is shown to be true not only in the single-channel case but also in a general multichannel case. The use of regularization and smoothing is also introduced in the context of inverse filtering.

2.1 Inverse Filtering Concept

The basic premise of inverse filtering is to undo any unwanted filtering imposed by a loudspeaker, room or linear system. Linear filtering can be viewed as a convolution and can be expressed as

$$c(n) \otimes h(n), \tag{2.1}$$

where $c(n)$ is some measured or known filter, $h(n)$ is some correction filter and n is the discrete-time index. For example, $c(n)$ might be the impulse response of a loudspeaker while $h(n)$ might be a correction filter designed to produce desired response. If one assumes that the desired "ideal" frequency response of a loudspeaker should be a flat spectrum with zero-phase, then equation (2.1) would become

$$\delta(n - n_D) \;=\; c(n) \otimes h(n), \tag{2.2}$$

where $\delta(n - n_D)$ is the discrete time delta function with a delay of n_D. The inverse filter $h(n)$ can then be found by solving for $h(n)$ in equation (2.2). If a flat spectrum is not

desired, but rather some other desired spectrum or response is wanted, equation (2.2) can be written as

$$a(n) = c(n) \otimes h_a(n), \tag{2.3}$$

where $a(n)$ is the desired or target response. One must solve for $h_a(n)$ in equation (2.3) given $c(n)$ the measured or modeled impulse response and $a(n)$ a desired time response. Figure 2.1 illustrates how an inverse filter might be used in correcting the response of a loudspeaker or room. Assuming $a(n) = \delta(n)$, the desired signal $d(n)$ is first filtered through an inverse filter $h(n)$, that has been computed beforehand to correct for the response of $c(n)$, before it is sent to the system $c(n)$. Ideally, the result of the filtering, $u(n)$, would be equal to the desired signal $d(n)$, though that is not always the case due to limitations in the inverse filter.

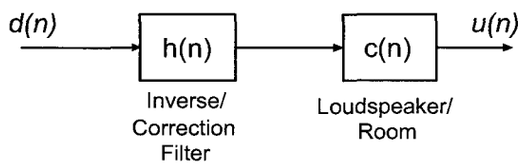


Figure 2.1: The implementation of an inverse filter, where $h(n)$ is the inverse filter that is correcting the system $c(n)$.

If the system that is to be corrected or inverted is not minimum phase as is typical with room impulse responses [31], then a delay must be added in the desired response $a(n)$ so that the result is causal. It should also be noted that due to the fact that convolution is a linear process, the inverse filter can and will only correct linear systems. Any non-linearities in the system will not be affected by the inverse filter.

In the next few sections, several known methods to calculate the inverse filter of a system will be presented.

2.1.1 Time Domain Solutions

Figure 2.2 shows a general inverse filtering structure for an input signal $x(n)$, output signal $u(n)$ and error at time n, $e(n)$. If we define the input/output and desired signal vectors as

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

$$\mathbf{y}(n) = [y(n), y(n-1), y(n-N_c+1)]^T \tag{2.5}$$

and the vectors for the system, target and inverse filters by

$$\mathbf{c} = [c(0), c(1), ..., c(N_c - 1)] \tag{2.6}$$

$$\mathbf{a} = [a(0), a(1), ..., a(N_a - 1)] \tag{2.7}$$

$$\mathbf{h} = [h(0), h(1), ..., h(N_h - 1)]. \tag{2.8}$$

N_c, N_a and N_h are the lengths of the input signal, the desired signal and the inverse filters respectively. Note that depending on the context, the vector $x(n)$ will also be defined as having a length N_a, for example as in Equation 2.9 below.

The error at time n can be written as

$$\begin{aligned} e(n) &= d(n) - \mathbf{cy}(n) \\ &= \mathbf{ax}(n) - \mathbf{cy}(n). \end{aligned} \tag{2.9}$$

If ones assumes that h is time invariant, the order of h and c does not matter and they can be switched. This is commonly used in active noise control [32] and referred to as the filtered-X method, where the input signal $x(n)$ is first filtered by c. We can define

$$\mathbf{v}(n) = [v(n), v(n-1), ..., v(n - N_h + 1)] \tag{2.10}$$

where

$$v(n) = \mathbf{cx}(n), \tag{2.11}$$

or

$$v(n) = \sum_{n_c=0}^{N_c-1} x(n - n_c) c_{n_c}. \tag{2.12}$$

Now the error at time n can then be written as

$$e(n) = \mathbf{ax}(n) - \mathbf{v}(n)\mathbf{h}^T, \tag{2.13}$$

where $\mathbf{h}^T$ is the transpose of $\mathbf{h}$. The transpose is used to keep the vector/matrix notation more compact.

If we use a Wiener filter approach which minimizes the mean-squared error between $d(n)$ and $u(n)$, as defined in [15], the cost function is given by,

$$J = \mathrm{E}\left[e^T(n)e(n)\right], \tag{2.14}$$

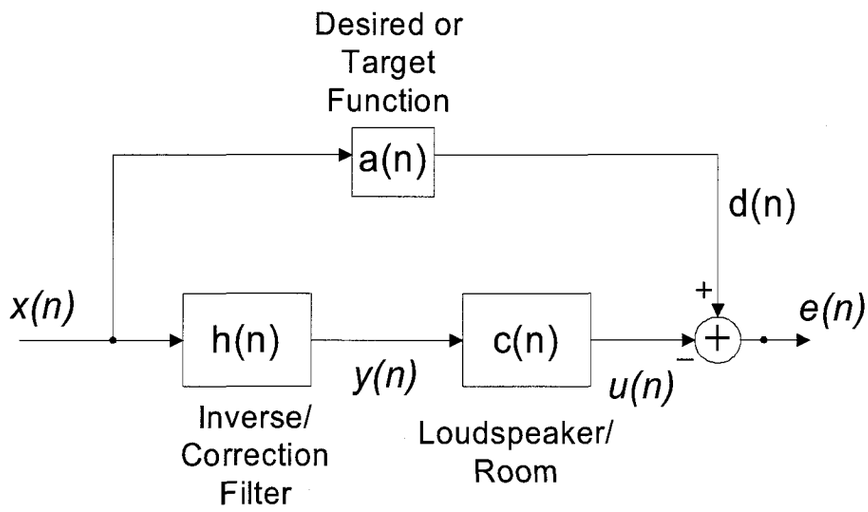


Figure 2.2: Block diagram of the inverse filtering method.

where the E signifies the statistical expectation value. After substituting in the expression for the error given by Equation 2.13, the cost function is given by,

$$J = \mathrm{E}\left[\left(\mathbf{a}\mathbf{x}(n) - \mathbf{v}(n)\mathbf{h}^T\right)^T \left(\mathbf{a}\mathbf{x}(n) - \mathbf{v}(n)\mathbf{h}^T\right)\right]. \tag{2.15}$$

Differentiating the cost function J with respect to h^T gives,

$$\frac{\partial J}{\partial h^T} = \mathrm{E}\left[-\mathbf{v}^T\mathbf{a}\mathbf{x} - \mathbf{v}^T\mathbf{a}\mathbf{x} + (\mathbf{v}^T\mathbf{v} + \mathbf{v}^T\mathbf{v})\mathbf{h}^T\right] \tag{2.16}$$

and setting the result to zero yields,

$$\mathrm{E}\left[\mathbf{v}^T\mathbf{v}\mathbf{h}^T\right] = \mathrm{E}\left[\mathbf{v}^T\mathbf{a}\mathbf{x}\right] \tag{2.17}$$

Solving for the inverse filter $\mathbf{h}^T$ gives

$$\mathbf{h}^T = \mathrm{E}\left[\mathbf{v}^T\mathbf{v}\right]^{-1}\mathrm{E}\left[\mathbf{v}^T\mathbf{a}\mathbf{x}\right]. \tag{2.18}$$

Now with a Wiener filter, it is assumed that the signals, $x(n)$ and $d(n)$ are all stationary and zero-mean, [39], therefore the expectations values are correlations. We define the statistical cross-correlation as

$$\begin{aligned} \gamma_{yx}(l) &= \mathrm{E}\left[y(n+l)x(n)\right] \\ &= \mathrm{E}\left[x(n-l)y(n)\right] \end{aligned} \tag{2.19}$$

for stochastic signals.

For the expectations in Equation 2.18 we want the following autocorrelation and cross-correlation functions, γ_{vv} and γ_{dv}. These two functions can be expanded as follows:

$$\gamma_{vv}(l) = \gamma_{xx}(l) * c(l) * c(-l) \tag{2.20}$$

and

$$\gamma_{dv}(l) = \gamma_{xx}(l) * a(l) * c(-l). \tag{2.21}$$

Therefore, from Equations 2.20 and 2.21 it can be seen that the deterministic terms, $a(l)$ and $c(l)$, can be isolated from the stochastic terms.

The Wiener solution given in Equation 2.18 can then be expressed in matrix form as,

$$\mathbf{\Gamma}_{vv}\mathbf{h}^T = \mathbf{\Gamma}_{dv}, \tag{2.22}$$

and be expanded as

$$\begin{bmatrix} \gamma_{vv}(0) & \gamma_{vv}(1) & \dots & \gamma_{vv}(L-1) \\ \gamma_{vv}(1) & \gamma_{vv}(0) & \dots & \gamma_{vv}(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{vv}(L-1) & \gamma_{vv}(L-2) & \dots & \gamma_{vv}(0) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{L-1} \end{bmatrix} = \begin{bmatrix} \gamma_{dv}(0) \\ \gamma_{dv}(1) \\ \vdots \\ \gamma_{dv}(L-1) \end{bmatrix}. \tag{2.23}$$

By using Equations 2.20 and 2.21 for the stochastic terms in the Wiener solution, Equation 2.23 can be divided out leaving only the deterministic terms. For brevity, we define the following short-hand for the convolutions which also correspond to correlation functions defined for finite energy deterministic sequences [39].

$$cc(l) = c(l) * c(-l) \tag{2.24}$$

and

$$ac(l) = a(l) * c(-l). \tag{2.25}$$

Therefore, the solution for the inverse filter h can be expressed solely by the deterministic terms. The final result, in matrix form, is given by,

$$\begin{bmatrix} cc(0) & cc(1) & \dots & cc(L-1) \\ cc(1) & cc(0) & \dots & cc(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ cc(L-1) & cc(L-2) & \dots & cc(0) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{L-1} \end{bmatrix} = \begin{bmatrix} ac(0) \\ ac(1) \\ \vdots \\ ac(L-1) \end{bmatrix} \tag{2.26}$$

Equation 2.26 contains only the autocorrelation of the system filter c and the cross-correlation between c and the target-function a. This can be expressed as the following,

$$\mathbf{Rh^T} = \mathbf{g} \tag{2.27}$$

and is the common solution for least-squares error filter design [15], where the matrix $\mathbf{R}$ is the autocorrelation matrix of the system c, $\mathbf{g}$ is the cross-correlation between the system c and the target response a and $\mathbf{h}$ is the inverse filter. Therefore, we show that in the case where all the filters, a, h, c are deterministic, that the Wiener case simplifies to a filter design solution that does not involve the statistics of $x(n)$ or $d(n)$.

Instead of looking over an infinite signal length for $v(n)$ and $x(n)$, as in the Wiener method, a least-squares approach can be used where the time samples are taken over a finite length. The cost function for the least-squares approach is given by,

$$J = \sum_{n=0}^{m} \lambda^{m-n} e^T(n)e(n) \tag{2.28}$$

$$= \sum_{n=0}^{m} \lambda^{m-n} \left[\mathbf{ax}(n) - \mathbf{v}(n)\mathbf{h}^T\right]^T \left[\mathbf{ax}(n) - \mathbf{v}(n)\mathbf{h}^T\right] \tag{2.29}$$

where λ is the forgetting factor to ensure that the data in the past is weighted less, which is needed for non-stationarity in the input and desired signals. The summation is taken over a length of $m+1$ input samples (the observation window).

Differentiating the cost function J with respect to h^T gives

$$\frac{\partial J}{\partial \mathbf{h}^T} = \sum_{n=0}^{m} \lambda^{m-n} \left[-\mathbf{v}^T(n)\mathbf{ax}(n) - \mathbf{v}^T(n)\mathbf{ax}(n) + (\mathbf{v}^T(n)\mathbf{v}(n) + \mathbf{v}^T(n)\mathbf{v}(n))\mathbf{h}^T\right] \tag{2.30}$$

and setting $\frac{\partial J}{\partial h}$ equal to zero results in

$$\mathbf{h}^T = \left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{v}^T(n)\mathbf{v}(n)\right]^{-1} \left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{v}^T(n)\mathbf{ax}(n)\right] \tag{2.31}$$

The summation form for the inverse filter h^T(or h), given by equation 2.31 and if $\lambda = 1$, is equivalent to the data matrix form that is commonly used, as in Haykin [15]. The summation,

$$\sum_{n=0}^{m} \mathbf{v}^T(n)\mathbf{v}(n) \tag{2.32}$$

can be written as a product of the two vectors as in

$$[\mathbf{v}(0), \mathbf{v}(1), ..., \mathbf{v}(m)]^T [\mathbf{v}(0), \mathbf{v}(1), ..., \mathbf{v}(m)], \tag{2.33}$$

as well as matrices such as

$$\mathbf{V}^T(m)\mathbf{V}(m). \tag{2.34}$$

where the matrix $\mathbf{V}^T(m)$ is given by

$$\mathbf{V}^T(m) = \begin{bmatrix} v(0) & v(1) & \dots & v(m) \\ v(-1) & v(0) & \dots & v(m-1) \\ \vdots & \vdots & \ddots & \vdots \\ v(-N_h+1) & v(-N_h+2) & \dots & v(m-N_h+1) \end{bmatrix} \tag{2.35}$$

The inverse filter $\mathbf{h}^T$ would then be given by

$$\mathbf{h}^T(m) = \left[\mathbf{V}^T(m)\mathbf{V}(m)\right]^{-1} \mathbf{V}^T(m)\mathbf{d}(m) \tag{2.36}$$

where the target signal $\mathbf{d}(m)$ is given by

$$\mathbf{d}(m) = [d(0), d(1), \dots, d(m)]^T \tag{2.37}$$

This is equivalent to the data matrix method expressed in [15] and is commonly used with $(m+1) = N_a$.

Similarly, we can use the data matrix approach to re-write the deterministic solution developed and which resulted in Equation 2.26. This was done in Kirkeby [23],

$$\mathbf{C} = \begin{bmatrix} c(0) & & & \\ \vdots & \ddots & & \\ c(N_c-1) & \ddots & \ddots & c(0) \\ & \ddots & \ddots & \vdots \\ 0 & & & c(Nc-1) \end{bmatrix} \tag{2.38}$$

leading to a new formulation for 2.26,

$$\mathbf{h}^T = (\mathbf{C}^T\mathbf{C})^{-1} \cdot \mathbf{C}^T\mathbf{a}, \tag{2.39}$$

where $\mathbf{C}$ is called the $N_a \times N_h$ convolution matrix of $c(n)$. Equation 2.39 is what is used by Kirkeby et al. in [23] for a least-squares inverse filter calculation, which is equivalent to Equations 2.27 and 2.36 in the deterministic case. In Kirkeby et al. [23], the target function $\mathbf{a}$ that is used is a delta function with a delay of $N_h/2 + 1$ samples.

2.1.2 Frequency Domain Deconvolution

Frequency-domain methods are usually more efficient and are frequently used to reduce computational load requirements in comparison to time-domain methods where a direct convolution is involved. Using the same approach as was done for the time-domain in Section 2.1.1, but in the frequency-domain, all the terms will be computed on a frequency-by-frequency basis. Therefore for each transformed frame of the time signal, the error given in Equation 2.9 would be,

$$E(k) = A(k)X(k) - C(k)Y(k). \tag{2.40}$$

where k is the frequency unit.

With the similar substitutions the error at a given frequency index k is given by,

$$E(k) = A(k)X(k) - V(k)H(k). \tag{2.41}$$

where the error, $E(k)$ is computed for each frequency index k from the frequency-transformed time-domain signals and filters.

The cost function for the Wiener solution approach would be given by,

$$J = \mathrm{E}\left[E^H(k)E(k)\right], \tag{2.42}$$

and substituting in the above expressions for the error term give,

$$J = \mathrm{E}\left[(AX - VH)^H(AX - VH)\right]. \tag{2.43}$$

Calculating the derivative with respect to H and setting it to zero, leading to,

$$\frac{\partial J}{\partial H} = \mathrm{E}\left[-V^HAX - V^HAX + V^HV + V^HV)\right] = 0. \tag{2.44}$$

and is true if,

$$\mathrm{E}\left[V^HVH\right] = \mathrm{E}\left[V^HAX\right] \tag{2.45}$$

and a solution for H is,

$$H = \mathrm{E}\left[V^HV\right]^{-1}\mathrm{E}\left[V^HAX\right]. \tag{2.46}$$

The deterministic components in Equation 2.46 can be isolated by substituting in the expression of V, and this results in,

$$C^HCH\mathrm{E}\left[X^HX\right] = C^HA\mathrm{E}\left[X^HX\right]. \tag{2.47}$$

The expectation values can be canceled out which gives an expression for the inverse filter solely based on the deterministic terms

$$H(k) = \left(C^H(k)C(k)\right)^{-1} C^H(k)A(k) \quad (2.48)$$
$$= C^{-1}(k)A(k) \quad (2.49)$$
$$= \frac{A(k)}{C(k)}. \quad (2.50)$$

Thus it can be seen that in the frequency domain the stochastic Wiener solution once again can be formulated without explicitly using the statistics of $x(n)$ or $d(n)$.

It should be noted that the property of turning a convolution in the time-domain to a multiplication in the frequency-domain does not yield identical results because of the discretization of the Fourier transform. The result that is computed in the frequency-domain is actually a circular-convolution in the time-domain and zero-padding cannot fully prevent the time-aliasing or "wrap-around" effects in this case, because the ideal inverse filter $h(n)$ is theoretically of infinite length. In the deconvolution situation as given by Equation 2.50, time-aliasing thus can never be totally avoided. By choosing a length of FFT large enough, the effect of the time-aliasing can be reduced to a desired level, though it may require a substantial increase in filter size.

If Equation 2.46 is to be estimated in practice by a least-squares method, the solution would have the form

$$H(m) = \left[\sum_{n=0}^{m} \lambda^{m-n} V^*(n)V(n)\right]^{-1} \left[\sum_{n=0}^{m} \lambda^{m-n} V^*(n)AX(n)\right] \quad (2.51)$$

where m and n are the frame index counters and the inverse filter $H(m)$ is computed over m frequency-domain frames. λ is a frame weighting parameter (or forgetting factor) that can be used to weight the recent frames more than older frames thus reducing the dependence on past data.

2.2 Minimum-Phase Decomposition

A causal discrete linear time-invariant system (LTI) is referred to being minimum phase if it has all its poles and zeros inside the unit circle, $|z| = 1$. This produces a stable and causal system in which the inverse of the system is also stable and causal. One desirable property of minimum phase systems is that there is no pre-response or signal before the most energetic part of the time response. Psychoacoustics tells us that a pre-response

could create an audible artifact that would not be desirable. It is for this reason that inverting only the minimum-phase part of the impulse response has been used in the past [31].

It is known that room impulse responses [31] are non-minimum phase, and loudspeaker crossovers are the usual reason why loudspeakers are non-minimum phase. To avoid the introduction of a pre-response the impulse responses can be broken down into their minimum-phase and all-pass components. Therefore, an inverse that is causal and stable could be calculated only for the minimum phase portion of the IR, thus avoiding the introduction of a pre-response.

Any mixed phase system or signal can be decomposed into a minimum-phase component and an all-pass component as follows. If we denote the minimum-phase component of $c(n)$ as $c_{min}(n)$ and the all-pass component as $c_{all}(n)$, and

$$C(k) = C_{min}(k)C_{all}(k) \tag{2.52}$$

$$|C(k)| = |C_{min}(k)| \tag{2.53}$$

and

$$|C_{all}(k)| = 1. \tag{2.54}$$

The minimum-phase component can be determined by zeroing the cepstrum for the negative quefrencies [31]. This can be described as

$$\hat{c}(n) = \begin{cases} c_p(n), & n = 0, N/2, \\ 2c_p(n), & 1 \leq n < N/2, \\ 0, & N/2 < n \leq N-1, \end{cases} \tag{2.55}$$

where

$$c_p(n) = \mathrm{DFT}^{-1}\left[\log(|C(k)|)\right]. \tag{2.56}$$

The minimum phase frequency response is then given by

$$C_{min}(k) = \exp[\mathrm{DFT}[\hat{c}(n)]]. \tag{2.57}$$

The all-pass component $C_{all}(k)$ can then be computed by

$$C_{all}(k) = \frac{C(k)}{C_{min}(k)}, \tag{2.58}$$

and the time-domain version as

$$c_{min}(n) = \text{DFT}^{-1}(C_{min}(k)) \quad (2.59)$$

$$c_{all}(n) = \text{DFT}^{-1}(C_{all}(k)). \quad (2.60)$$

Another way of finding the minimum-phase component is noting that all the frequency magnitude information of $c(n)$ is contained in $C_{min}(k)$ which is given by equation (2.53). Therefore, what is left to calculate is the phase of the minimum-phase component. It is known that the phase of a minimum-phase signal is related to the Hilbert transform of the log-magnitude of the signal, given by

$$\phi_{C_{min}}(k) = -\text{img}(\ln(\text{hilbert}(|C_{min}(k)|))). \quad (2.61)$$

2.3 Regularization

In Equations 2.27 and 2.39, the matrix $\mathbf{R}$ or $(\mathbf{C}^T\mathbf{C})$ could be ill-conditioned or singular and therefore a useable matrix inverse may not be found. Another way to view this problem is to look at the frequency-domain inverse filter expression given by Equation 2.50 and the effect when there is a zero or near-zero term in the denominator, $C(k)$. For some value of k, a small $C(k)$ would result in a large boost in the inverse filter $H(k)$. This could result in a very long inverse filter that could potentially ring on, and therefore would wrap around if calculated in the frequency-domain due to circular convolution effects. Implementing an inverse filter with a large boost is not always desirable as it could cause excessive gains in the audio chain and overload components and cause distortions. A method to avoid such problems is to regularize the system, or to add a bias to the system so that it is not ill-conditioned or that the denominator is not very small.

Figure 2.3 shows a block diagram of an inverse filtering scheme where regularization has been added. The regularization filter is given by

$$\mathbf{b}(n) = [b(n), b(n-1), b(n-N_b+1)]^T \quad (2.62)$$

and the regularized signal $y_R(n)$ is defined by,

$$y_R(n) = \mathbf{by}(n) \quad (2.63)$$

with

$$\mathbf{y}(n) = [y(n), y(n-1), y(n-N_b+1)]^T, \quad (2.64)$$

or in summation form,

$$y_R(n) = \sum_{n_c=0}^{N_b-1} y(n-n_c)b_{n_c}. \quad (2.65)$$

Note that the phase response of the regularization filter is irrelevant, as it will be canceled out in the cost function to be defined.

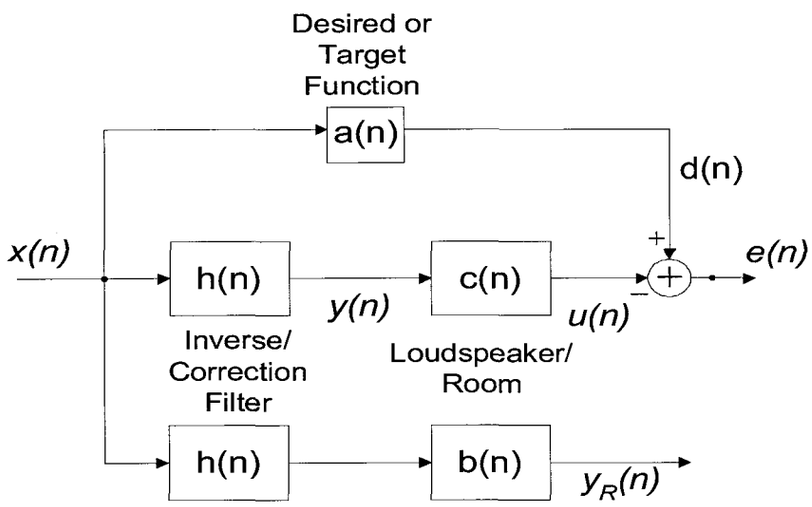


Figure 2.3: Block diagram of the inverse filtering with regularization.

For the Wiener case, given by Equation 2.14, a regularization term is added to the cost function to give,

$$J = \mathrm{E}\left[e^T(n)e(n) + \beta y_R(n)^T y_R(n)\right], \quad (2.66)$$

where β is scalar to adjust the amount of regularization. By following the same notation and method from Section 2.1.1, the Wiener solution for the inverse filter h can be expressed as,

$$\mathbf{h}^T = \mathrm{E}[\mathbf{v}^T\mathbf{v} + \beta\mathbf{y}(n)^T\mathbf{b}^T\mathbf{b}\mathbf{y}(n)]^{-1}\mathrm{E}[\mathbf{v}^T\mathbf{a}\mathbf{x}] \quad (2.67)$$

Similarly, the least-squares solution will be given by

$$\mathbf{h}^T = \left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{v}^T(n)\mathbf{v}(n) + \beta\mathbf{y}(n)^T\mathbf{b}^T\mathbf{b}\mathbf{y}(n)\right]^{-1}\left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{v}^T(n)\mathbf{a}\mathbf{x}(n)\right]. \quad (2.68)$$

In both cases, the regularization adds another term to the part of the inverse filter expression that is inverted, so that the denominator does not become ill-conditioned or zero.

In the same way as in Section 2.1.1 the least-squares solution given by Equation 2.68 can be written in matrix form using only deterministic terms as in [23],

$$\mathbf{h}^T = (\mathbf{C}^T\mathbf{C} + \beta\mathbf{B}^T\mathbf{B})^{-1} \cdot \mathbf{C}^T\mathbf{a}, \tag{2.69}$$

where $\mathbf{B}$ is the $N_a \times N_h$ regularization filter matrix that has the form,

$$\mathbf{B} = \begin{bmatrix} b(0) & \ddots & \\ \vdots & \dots & b(0) \\ b(N_b-1) & \ddots & \vdots \\ & \dots & b(N_b-1) \end{bmatrix}, \tag{2.70}$$

where $b(n)$ is the regularization filter vector. Equation 2.69 expresses the inverse filter h in terms of the impulse response c of the system, and the target function a.

Just as in the time-domain inverse filtering methods, a regularization term can be added in the frequency-domain method. The cost function given by Equation 2.42 would be given by

$$J = \mathrm{E}\left[E^H(k)E(k) + \beta Y_R(k)Y_R(k)\right], \tag{2.71}$$

where a scalar term β is used to control the overall amount of regularization added.

Using the same methods as in the case without regularization, the inverse filter with regularization is given by,

$$H(k) = \frac{A(k)C^*(k)}{C(k)C^*(k) + \beta B(k)B^*(k)}. \tag{2.72}$$

This result is identical to the results in Kirkeby et al. [25].

The regularization adds a bias or noise to the resulting inverse filter, and the amount is controlled with the variable β. The frequency-dependent regularization vector $B(k)$ is best described in the frequency-domain, in term of how it would be used. Setting $B(k) = 0$ would be the minimum amount of regularization or none at all and Equation 2.72 would reduce to Equation 2.50. There is no real upper limit to the value of the regularization Though at some point when the regularization is larger than the $C(k)$ term, the inverse filter will tend to zero. The frequency dependence of the term $B(k)$

lets one vary the amount of regularization across the frequency spectrum. For example, if one is attempting to calculate the inverse filter of a loudspeaker where the frequency response of the loudspeaker rolls-off at the low and high ends, one does not want to boost the low and high frequencies since the loudspeaker does not output very much energy there. Therefore, one would apply larger regularization at those frequencies so that the inverse filter will not affect those frequencies since the inverse filter will be very small in amplitude.

Figure 2.4 shows three different forms that the regularization term $[\beta B(k)B^*(k)]$ could take. Figure 2.4 a) shows the regularization being constant across the frequency range, i.e. $|B(k)| = 1$ and the value β would determine the amount added. The b) curve shows a regularization as what is described above where maximum values are applied to frequencies where the impulse response rolls-off. The final type, c) is based on the $\frac{1}{3}$-octave spectrum of $C(k)$ such as proposed by Fielder [9].

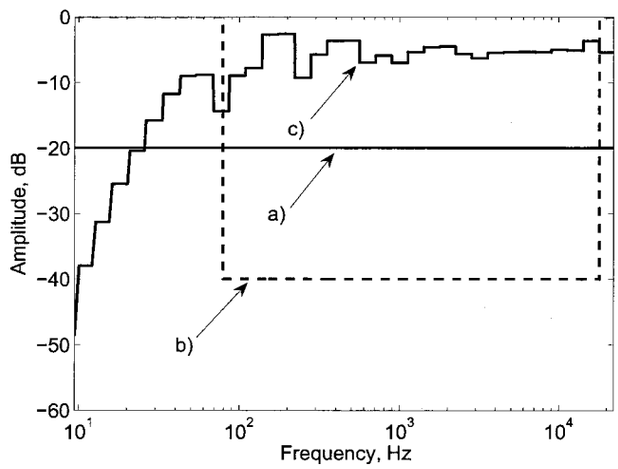


Figure 2.4: Three different regularization types: a) scalar where $B(k)$ is unity and β controls amount, b) vector type where $B(k)$ is inversely shaped compared to $|C(k)|$, and c) where $B(k)$ is based on the $\frac{1}{3}$-octave spectrum of $|C(k)|$

Using the spectrum of a system to define a regularization response provides a way to limit the effort done by the inverse filter relative to localized frequencies as opposed to a general frequency independent or constant regularization.

Craven and Gerzon [8] implement a similar regularization method, based on the spectrum of the system, in the frequency-domain, given by the following expression, in

a slightly different notation,

$$\hat{S}_r(k) = \frac{S_r(k) + l(k)^2}{l_0 + l(k)^2} \tag{2.73}$$

where $S_r(k)$ is the spectrum of the filter $C(k)$ and l_0 is chosen such that the average value of $\hat{S}_r(k)$ equals 1. l is chosen to provide "a desired degree of trough filling" and can control the frequency range of the regularization. It has the form of

$$l(k)^2 = \left(\frac{k_{low}}{k}\right)^4 + e + \left(\frac{k}{k_{high}}\right)^4 \tag{2.74}$$

where k_{low} and k_{high} are the frequency index limits between which the regularization is to be implemented. The constant e determines the "basic degree" or overall amount of regularization. This implementation is solely based on the spectrum of the signal, $\hat{S}_r(k)$ and was used to derive a minimum-phase compensation filter.

2.4 Windowing and Smoothing

Smoothing is the process of removing short-term variations or the "detail" of a signal. It is common practice to smooth the frequency response curves for loudspeakers, microphones and other electronic equipment to reduce the variation of the response, to make it look more uniform and "clean". Smoothing can actually hide peaks and dips in frequency responses so that they look more flat. A simple type of smoothing is a moving-average where the current sample is replaced by the average of itself and adjacent samples. For a time-domain sequence, the moving average type of smoothing can be expressed as

$$x_s(n) = \frac{1}{2M+1} \sum_{m=-M}^{M} x(n+m) \tag{2.75}$$

where M samples on either side of the current sample are averaged together with the current sample. The smoothing is a type of low-pass filtering process and can be implemented as a convolution. In the above example, Equation 2.75 can be expressed as a convolution of the signal $x(n)$ with a smoothing window $w(n)$ that is a "box-shaped" pulse of length $2M+1$ samples with all samples being equal to $\frac{1}{2M+1}$ and the result is given by,

$$x_s(n) = \sum_{m=-\infty}^{\infty} w(m)x(n-m). \tag{2.76}$$

With the above definition of smoothing, the shape of the smoothing function can be easily changed from a rectangular type to others by varying the weighting of adjacent samples.

A convolution, or more specifically a circular convolution, in the time-domain can be written as a multiplication in the DFT frequency domain. Therefore, the time-domain smoothing can be written as a multiplication of the DFT of the signal by the DFT of the smoothing function and is given by

$$X_s(k) = W(k)X(k). \tag{2.77}$$

Similarly, one could smooth a signal in the frequency-domain, which would be equivalent to windowing the signal in the time time-domain. We will be referring to windowing as a multiplication operation, where the signal or response is multiplied by a windowing function. Smoothing in the frequency-domain for discrete time signals could be defined as

$$\begin{aligned} C_{ps}(k) &= \mid C(k) \mid^2 \otimes W_{sm}(k) \\ &= \sum_{i=0}^{N-1} \mid C[(k-i)\mathrm{mod}N] \mid^2 \cdot W_{sm}(i), \end{aligned} \tag{2.78}$$

where $|C(k)|^2$ is the power spectrum of $c(n)$, $W_{sm}(k)$ is the spectral smoothing function that defines the bandwidth of the smoothing and $\otimes$ indicates circular convolution. The bandwidth of the smoothing function pertains to the range of values of $C(k)$ that will contribute to the smoothed version $C_{ps}(k)$. It should be noted that the mod operator is necessary if $C(k)$ is only defined for $0 \leq k \leq N-1$, otherwise it can be removed if $C(k)$ is periodic. This method is the traditional way of smoothing the frequency response or power spectrum. There is no smoothing of the phase, which generally means that the corresponding impulse response or time-domain function cannot be recovered from the smoothed power spectrum. Recovering the time response is important if one is trying to correct the magnitude and phase response with an inverse filter. It has been pointed out in Lipshitz [27], that a smoothed impulse response can be recovered from a smoothed power spectrum using a zero-phase component with an inverse DFT.

A smoothing method for acoustic/audio impulse responses, where the smoothing window is applied to the real and imaginary components of the frequency-response,

introduced by Hatziantoniou and Mourjopoulos, [12], is given by

$$\begin{aligned} H_{cs}(k) &= H(k) \otimes W_{sm}(k) \\ &= \sum_{i=0}^{N-1} H[(k-i)\mathrm{mod}N] \cdot W_{sm}(i). \end{aligned} \tag{2.79}$$

The smoothed time-domain signal can be obtained from the inverse DFT of the complex frequency-response. Instead of smoothing the real and imaginary parts of the frequency response, smoothing the magnitude and phase can also be done [38].

In general a smoothing window can be defined in the time-domain by a Fourier series [11] as in

$$w_{sm}(n) = \sum_{i=0}^{L-1} b_i \cdot \cos\left(\frac{2\pi i n}{N}\right), \quad 0 \leq n \leq N-1, \tag{2.80}$$

where i is an integer index, L is the number of one-sided Fourier coefficients and N is the overall length of the window and is even. If $L = 1$ and $b_0 = 1$, $w(n)$ given by Equation 2.80 defines a rectangular window. With $L = 2$, the well known Hann ($b_0 = 0.5$ and $b_1 = -0.5$) and Hamming ($b_0 = 0.54$ and $b_1 = -0.46$) windows can be also be defined by Equation 2.80. The time-window given by Equation 2.80 can be expressed in the frequency-domain as

$$W_{sm}(k) = b - (1-b)\cos\left(\frac{2\pi k}{N}\right), \quad 0 \leq k \leq N-1 \tag{2.81}$$

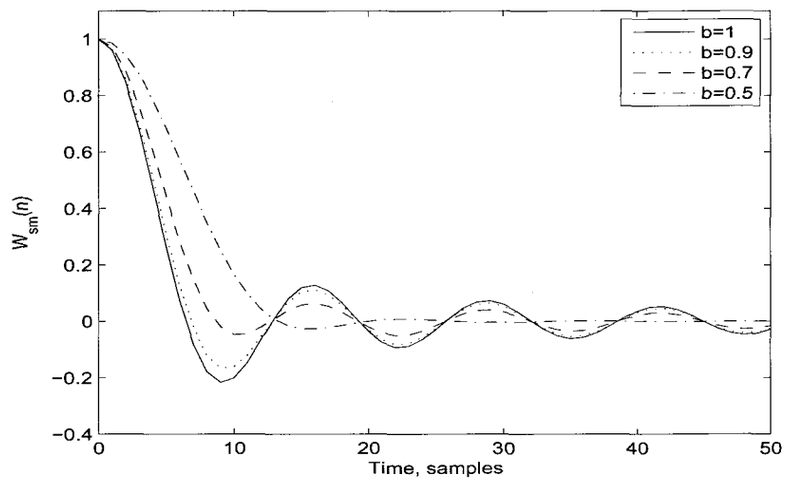
The window or smoothing function can be defined in general as the one-sided function by

$$W_{sm}(k) = \begin{cases} \frac{b-(b-1)\cos[(\pi/m)k]}{2b(m+1)-1}, & k = 0, 1, \ldots, m \\ \frac{b-(b-1)\cos[(\pi/m)(k-N)]}{2b(m+1)-1}, & k = N-m, N-(m-1), \ldots, N-1 \\ 0, & k = m+1, \ldots, N-(m+1) \end{cases} \tag{2.82}$$

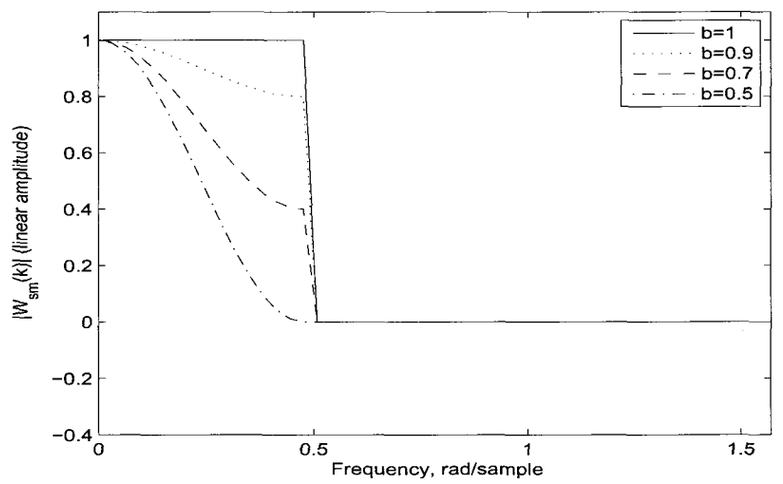
where b is a weighting coefficient from the Fourier definition of a data window [12] [11] and m is the smoothing index and corresponds to the length (in samples) of the half-window. When $b = 1$, Equation 2.82 becomes a rectangular frequency smoothing function given by,

$$W_{sm}(k) = \begin{cases} \frac{1}{2m+1}, & k \in \{0, 1, \ldots, m\} \cup \{N-m, N-(m-1), \ldots, N-1\} \\ 0, & k \in \{k = m+1, \ldots, N-(m+1)\} \end{cases} \tag{2.83}$$

Figure 2.5 shows the time-domain and frequency-domain forms of smoothing windows derived from Equations 2.80 and 2.81 for different values of b.



(a)



(b)

Figure 2.5: General smoothing window derived from Equation 2.81 for various values of the parameter b. Shown are (a) the time-domain function and (b) the frequency-domain function.

In the preceding discussion the smoothing window $w(n)$ in the time-domain or $W(k)$ in the frequency-domain have all been of constant bandwidth or have produced a uniform smoothing in time or frequency respectively. Traditional smoothing methods use fractional-octave windows, typically one-third-octave, in the frequency-domain which requires a non-uniform smoothing function. For fractional-octave windows, the upper and lower frequency limits of the band, given by f_U and f_L respectively as a function of centre frequency f, are given by,

$$f_U = 2^{0.5(\text{octave fraction})} f \tag{2.84}$$

and

$$f_L = 0.5^{0.5(\text{octave fraction})} f. \tag{2.85}$$

The *octave fraction* would be 1 for octave, $\frac{1}{3}$ for third octave and so forth. Third-octave bandwidths are commonly used since the size resembles the bandwidth for the critical bands of the ear at the mid-frequencies.

To accommodate variable bandwidth smoothing, for example fractional-octave as stated above, m which is introduced in Equation 2.82 must become a function of the the frequency variable k and be expressed as $m(k)$. As stated at the beginning of this section, fractional-octave smoothing is a common method of smoothing, namely one-third octave, since that size of bandwidth resembles the bandwidth for the critical bands at mid-frequencies. A critical band is defined in psychoacoustic masking terms as the maximum bandwidth of noise which can be perceived by humans as the same loudness as a sinewave of similar power at the band centre. One critical bandwidth represents one Bark, and hence this type of smoothing is sometimes referred to as Bark smoothing. The critical bandwidth can be calculated by the following expression,

$$BW_{Bark} = 94 + 71 f_{kHz}^{3/2} \tag{2.86}$$

The equivalent rectangular bandwidth (ERB) by Moore and Glasberg [29] is a replacement of the Bark Scale that is based on a revised Zwicker model of hearing. The ERB bandwidth can be calculated from,

$$BW_{ERB} = 6.23 f_{Hz}^2 + 93.39 f_{Hz} + 28.52. \tag{2.87}$$

Figure 2.6 plots the bandwidth as a function of frequency for ERB, Bark and third-octave bandwidths.

By smoothing the frequency response prior to calculating the inverse filter, one hopes to remove the larger extremes in the frequency response so that the inverse filter does not

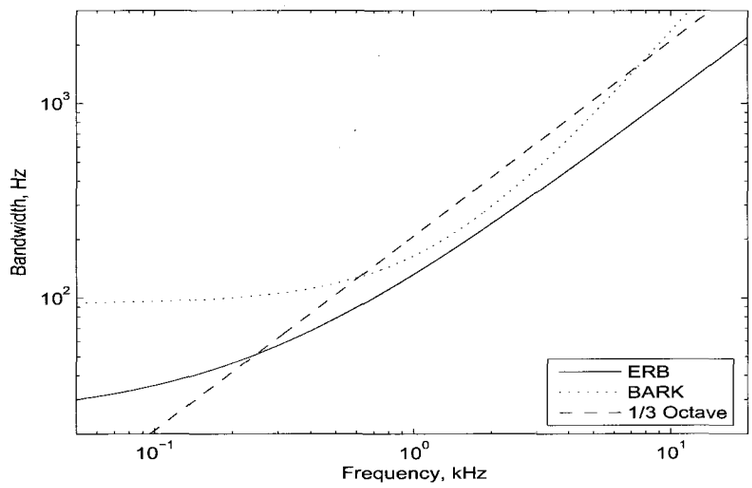


Figure 2.6: The bandwidths plotted versus frequency for three perceptually based smoothing methods: Bark scale, ERB and third-octave.

have to correct them. For example, an impulse response with a large dip in the frequency response would imply that it has a zero or zeros close to the unit circle. The inverse filter for this impulse response would have a pole or poles close to the unit circle, thus making a longer inverse filter in time. Also more fine detail in the frequency response indicates the inverse filter's frequency response would have to have a similar amount of detail. The more fine detail would mean the inverse filter would have a longer time response. Therefore, by smoothing the frequency response of the impulse response, one hopes to remove enough detail and variation to make the inverse filter shorter, but not too much to make the smoothing audible.

Windowing can also be used to reduce effects of time-aliasing when block processing techniques such as DFT frequency-domain deconvolution is used. By applying a window to the time-domain inverse filter that was calculated, one can remove energy at the ends of the filter that can cause audible artifacts. As noted above, this windowing in the time-domain is equivalent to a convolution in the frequency-domain, therefore this has a smoothing effect. A flap-top style of window can be used where a fade-in is created from the first half of the window function and a fade-out is created from the second half of the window function. Ones are inserted between the fade-in and fade-out parts such that the proper length of window is achieved. Fielder [9] used a Kaiser-Bessel window for the fade-in and fade-out each of length 2048 samples, or 42.7 ms when sampled at

48kHz.

2.5 Multichannel Filtering

2.5.1 General Description

Multichannel systems are more common nowadays and are used in surround audio systems or binaural simulators or crosstalk cancelers. Figure 2.7 shows a general multichannel setup with I signals (independent audio sources), K receivers (microphones) and J loudspeakers.

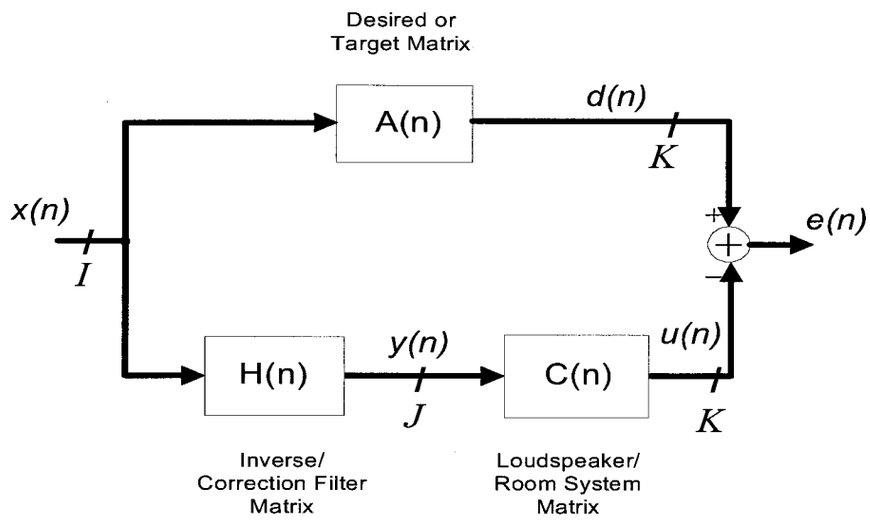


Figure 2.7: Block diagram of general multichannel inverse filtering scheme.

The multichannel solution can be arrived at in a similar way as the single-channel approach given in section 2.1.1, except instead of the filters being vectors, they become matrices. The inverse filters solution for each source I can be computed separately and the total result can be arrived at from a superposition of the I sources. This also simplifies the notation. Therefore, we define the target function matrix $\mathbf{A}_i$, size $K \times N_a$, the system matrix $\mathbf{C}$ size $K \times JN_c$ and inverse filter matrix $\mathbf{H}_i$ size $1 \times JN_h$. The signal vectors $X_i(n)$ and $Y_i(n)$ are $N_A \times 1$ and $JN_c \times 1$ respectively.

The multichannel error, size $K \times 1$ is given by

$$\mathbf{E}_i = \mathbf{A}_i\mathbf{X}_i - \mathbf{C}\mathbf{Y}_i, \tag{2.88}$$

and using a filtered-x approach as done with the single-channel version, the error can be written as

$$\mathbf{E}_i = \mathbf{A}_i\mathbf{X}_i - \mathbf{V}_i\mathbf{H}_i^T \tag{2.89}$$

where $\mathbf{V}_i(n)$ of size $K \times JN_h$ is now in the form of,

$$\mathbf{V}_i(n) = \begin{bmatrix} v_{i,k=1,j=1}(n) & \dots & v_{i,k=1,j=J}(n) & \dots & v_{i,k=1,j=1}(n-N_H+1) & \dots & v_{i,k=1,j=J}(n-N_H+1) \\ & \dots & & & \dots & & \\ v_{i,k=K,j=1}(n) & \dots & v_{i,k=K,j=J}(n) & \dots & v_{i,k=K,j=1}(n-N_H+1) & \dots & v_{i,k=K,j=J}(n-N_H+1) \end{bmatrix}, \tag{2.90}$$

where $v_{i,k,j}(n)$ is given by

$$v_{i,k,j} = \sum_{n_c=0}^{N_c-1} x_i(n-n_c)c_{k,j,n_c} \tag{2.91}$$

As with the single-channel case, the cost function is given by

$$J = \mathrm{E}\left[\mathbf{E}_i^T\mathbf{E}_i\right]. \tag{2.92}$$

Substituting the expressions for the error terms, differentiating with respect to $\mathbf{H}_i^T$, setting the result to zero and solving for $\mathbf{H}_i^T$, in the same way as the single-channel case, will give,

$$\mathbf{H}_i^T = \mathrm{E}\left[\mathbf{V}_i^T\mathbf{V}_i\right]^{-1}\mathrm{E}\left[\mathbf{V}_i^T\mathbf{A}_i\mathbf{X}_i\right]. \tag{2.93}$$

Following the same method as in the single-channel case, we want the following autocorrelation and cross-correlation functions $\gamma_{v_jv_{j'}}(l)$ and $\gamma_{dv_{j'}}(l)$ for the expectations in Equation 2.93. These functions can be defined as,

$$\gamma_{v_jv_{j'}}(l) = \mathrm{E}\left[\sum_k v_{i,k,j}(n+l)v_{i,k,j'}(n)\right] \tag{2.94}$$

and

$$\gamma_{dv_{j'}}(l) = \mathrm{E}\left[\sum_k d_{i,k}(n+l)v_{i,k,j'}(n)\right]. \tag{2.95}$$

Equation 2.93 can then be expanded in matrix form to give,

$$\begin{bmatrix} \gamma_{v_1v_1}(0) & \cdots & \gamma_{v_1v_J}(0) & \cdots & \gamma_{v_1v_1}(N_h-1) & \cdots & \gamma_{v_1v_J}(N_h-1) \\ \vdots & & \vdots & \cdots & \vdots & & \vdots \\ \gamma_{v_Jv_1}(0) & \cdots & \gamma_{v_Jv_J}(0) & \cdots & \gamma_{v_Jv_1}(N_h-1) & \cdots & \gamma_{v_Jv_J}(N_h-1) \\ & \vdots & & \vdots & & \vdots & \\ & \vdots & & \vdots & & \vdots & \\ \gamma_{v_1v_1}(N_h-1) & \cdots & \gamma_{v_1v_J}(N_h-1) & \cdots & \gamma_{v_1v_1}(0) & \cdots & \gamma_{v_1v_J}(0) \\ \vdots & & \vdots & \cdots & \vdots & & \vdots \\ \gamma_{v_Jv_1}(N_h-1) & \cdots & \gamma_{v_Jv_J}(N_h-1) & \cdots & \gamma_{v_Jv_1}(0) & \cdots & \gamma_{v_Jv_J}(0) \end{bmatrix} \begin{bmatrix} h_{i,1,0} \\ \vdots \\ h_{i,J,0} \\ \vdots \\ \vdots \\ h_{i,1,N_h-1} \\ \vdots \\ h_{i,J,N_h-1} \end{bmatrix} =$$

$$\begin{bmatrix} \gamma_{dv_1}(0) \\ \vdots \\ \gamma_{dv_J}(0) \\ \vdots \\ \vdots \\ \gamma_{dv_1}(N_h-1) \\ \vdots \\ \gamma_{dv_J}(N_h-1) \end{bmatrix}. \tag{2.96}$$

The expectation values can be expressed as a product of the stochastic component and deterministic component similar to the single-channel case,

$$\gamma_{v_jv_{j'}}(l) = \left[\sum_k c_{kj}(l) * c_{kj'}(-l)\right] * \gamma_{x_ix_i}(l) \tag{2.97}$$

and

$$\gamma_{dv_{j'}}(l) = \left[\sum_k a_{ik}(l) * c_{kj'}(-l)\right] * \gamma_{x_ix_i}(l). \tag{2.98}$$

For compactness, similar to what was done in the single-channel case, we define the following short-hand notation for the convolutions which correspond to the correlation

functions defined for finite energy deterministic sequences [39],

$$c_j c_{j'} = \sum_k c_{j,k}(l) * c_{j',k}(-l) \tag{2.99}$$

$$a_i c_{j'} = \sum_k a_{i,k}(l) * c_{j',k}(-l). \tag{2.100}$$

Therefore Equation 2.5.1 can be expressed solely by deterministic terms, and the result is given by,

$$\begin{bmatrix} c_1c_1(0) & \dots & c_1c_J(0) & \dots & c_1c_1(N_h-1) & \dots & c_1c_J(N_h-1) \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ c_Jc_1(0) & \dots & c_Jc_J(0) & \dots & c_Jc_1(N_h-1) & \dots & c_Jc_J(N_h-1) \\ & \vdots & & \vdots & & \vdots & \\ & \vdots & & \vdots & & \vdots & \\ c_1c_1(N_h-1) & \dots & c_1c_J(N_h-1) & \dots & c_1c_1(0) & \dots & c_1c_J(0) \\ \vdots & & \vdots & \dots & \vdots & & \vdots \\ c_Jc_1(N_h-1) & \dots & c_Jc_J(N_h-1) & \dots & c_Jc_1(0) & \dots & c_Jc_J(0) \end{bmatrix} \begin{bmatrix} h_{i,1,0} \\ \vdots \\ h_{i,J,0} \\ \vdots \\ \vdots \\ h_{i,1,N_h-1} \\ \vdots \\ h_{i,J,N_h-1} \end{bmatrix} =$$

$$\begin{bmatrix} a_ic_1(0) \\ \vdots \\ a_ic_J(0) \\ \vdots \\ \vdots \\ a_ic_1(N_h-1) \\ \vdots \\ a_ic_J(N_h-1) \end{bmatrix}. \tag{2.101}$$

As with the single-channel case in Equation 2.31, the inverse filters can be computed from a least-squares estimation giving,

$$\mathbf{H}_i^T(m) = \left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{V}_i^T(n)\mathbf{V}_i(n)\right]^{-1}\left[\sum_{n=0}^{m} \lambda^{m-n}\mathbf{V}_i^T(n)\mathbf{A}_i\mathbf{X}_i(n)\right] \tag{2.102}$$

Instead of Equation 2.102, the form of Equation 2.35 making use of data matrices could also be extended to the multichannel case. We prefer to present here the data matrix form of Equation 2.101, where the correlation functions are replaced by equivalent "convolution

matrices", as in Kirkeby [23]. The result is the multichannel extension of Equation 2.39. We define the multichannel convolution matrix $\mathbf{C}$ as the following $KN_a \times JN_h$ matrix,

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}(0) & \ddots & \\ \vdots & \ddots & \mathbf{c}(0) \\ \mathbf{c}(N_c-1) & \ddots & \vdots \\ & & \mathbf{c}(N_c-1) \end{bmatrix} \tag{2.103}$$

where each matrix $\mathbf{c}(l)$ is of size K by J and is defined as

$$\mathbf{c}(l) = \begin{bmatrix} c_{k=1,j=1}(l) & \dots & c_{k=1,j=J}(l) \\ \vdots & & \vdots \\ c_{k=K,j=1}(l) & \dots & c_{k=K,j=J}(l) \end{bmatrix}. \tag{2.104}$$

We also define the multichannel $\mathbf{A}_i$ vector as the following $KN_a \times 1$ vector,

$$\mathbf{A}_i = [\mathbf{a}_i(0), \dots, \mathbf{a}_i(N_a-1)]^T \tag{2.105}$$

where each vector $\mathbf{a}_i(l)$ is of size $K \times 1$,

$$\mathbf{a}_i(l) = [a_{i,k=1}(l), \dots, a_{i,k=K}(l)]^T. \tag{2.106}$$

Note that the definition of the matrix $\mathbf{C}$ and vector $\mathbf{a}_i$ are different from the previous definitions as shown in Figure 2.7.

The multichannel solution can then be expressed as,

$$\mathbf{H}_i^T = \left[\mathbf{C}^T\mathbf{C}\right]^{-1}\mathbf{C}^T\mathbf{A}_i. \tag{2.107}$$

Next, we choose this formulation to present the solution with a regularization term. As in the mono channel case with Equation 2.69 and illustrated in Figure 2.3, a regularization filter can be used, and is included in the cost-function to be minimized. This regularization vector is a frequency dependent function to control the effort from the loudspeaker signals. The multichannel regularization filter matrix is then defined as the following $N_a \times JN_h$ matrix,

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}(0) & \ddots & \\ \vdots & \ddots & \mathbf{b}(0) \\ \mathbf{b}(N_b-1) & & \vdots \\ & & \mathbf{b}(N_b-1) \end{bmatrix}, \tag{2.108}$$

where each element $\mathbf{b}(i)$ in $\mathbf{B}$ is defined as the $1 \times J$ vector,

$$\mathbf{b}(i) = [b(i), b(i), b(i)]. \tag{2.109}$$

The multichannel regularization vector given by Equation 2.109 is the case where each channel has the same regularization response, $b(i)$.

The multichannel least-squares solution with regularization then becomes,

$$\mathbf{H}_i^T = \left(\mathbf{C}^T\mathbf{C} + \beta\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{C}^T\mathbf{A}_i, \tag{2.110}$$

where β again is a scalar controlling the overall level of regularization. Note that a regularization term could also be added to other formulations, such as in Equations 2.93, 2.101 and/or 2.102.

2.5.2 Frequency-domain Multichannel Inverse filtering

The results of Equation 2.107 and Equation 2.110 can also be formulated in the frequency domain for the multichannel solution with and without regularization. This provides the following results, which are the extension of the mono channel results from Equation 2.50 and Equation 2.72,

$$\mathbf{H}_i(k) = \left[\mathbf{C}^H(k)\mathbf{C}(k)\right]^{-1}\mathbf{C}^H(k)\mathbf{A}_i(k) \tag{2.111}$$

and

$$\mathbf{H}_i(k) = \left[\mathbf{C}^H(k)\mathbf{C}(k) + \beta\mathbf{B}^H(k)\mathbf{B}(k)\right]^{-1}\mathbf{C}^H(k)\mathbf{A}_i(k), \tag{2.112}$$

where $\mathbf{H}_i(k)$ is of size $J \times 1$, $\mathbf{C}(k)$ is of size $K \times J$, $\mathbf{A}_i(k)$ is of size $K \times 1$ and $\mathbf{B}(k)$ is of size $1 \times J$. Note that other formulations would be possible, such as those involving statistics of $x(n)$ and $d(n)$ as in Equation 2.93 and Equation 2.102. It should be noted the difference between the k_{th} receiver and k the frequency index.

2.6 Chapter Summary

In this chapter the inverse filtering concept was introduced along with some techniques such as regularization and smoothing to improve on the general concept. It was shown that the general Wiener filtering case for single-channel and multichannel scenarios can be solely expressed deterministically if the system that is being inverted is deterministic.

Chapter 3

Inverse Filter Evaluation and Tests on Known Methods

In this chapter a subjective evaluation method is introduced and developed to test the various inverse filtering methods introduced in Chapter 2. Formal subjective tests are carried out to evaluate the various methods and the results are reported.

3.1 Measurement of Impulse Responses

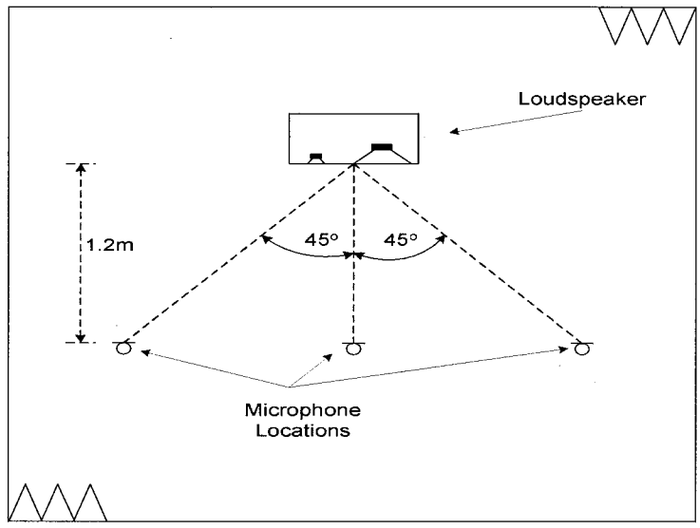
To evaluate the various inverse filtering methods the IRs of two different types of loudspeakers were measured, on-axis (0^o) and off-axis (45^o) in an anechoic environment. The choice of using loudspeaker impulse responses measured in an anechoic environment was to generate "simple" IRs that are short in length and potentially easy to calculate an inverse for. Even though shorter impulse responses are used in the comparisons in this chapter, the methods used to calculate the inverse filters are still suitable for longer more "complex" cases.

The first loudspeaker, a Mackie HR824 powered studio monitor (Loudspeaker A), is a conventional two-driver loudspeaker where the tweeter and woofer are physically located at two separate locations on the front baffle of the loudspeaker. The second loudspeaker, a Tannoy 800A powered studio monitor (Loudspeaker B) is a dual-concentric type, where the tweeter is located in the centre of the low frequency driver. The layout of the measurement setup is shown in Figure 3.1 . To achieve a more diverse range of IRs, Loudspeaker A was placed horizontally so that the woofer and tweeter are side-by-side. Therefore, a total of three IRs were measured, on-axis, at 45^o on the tweeter side, and at 45^o on

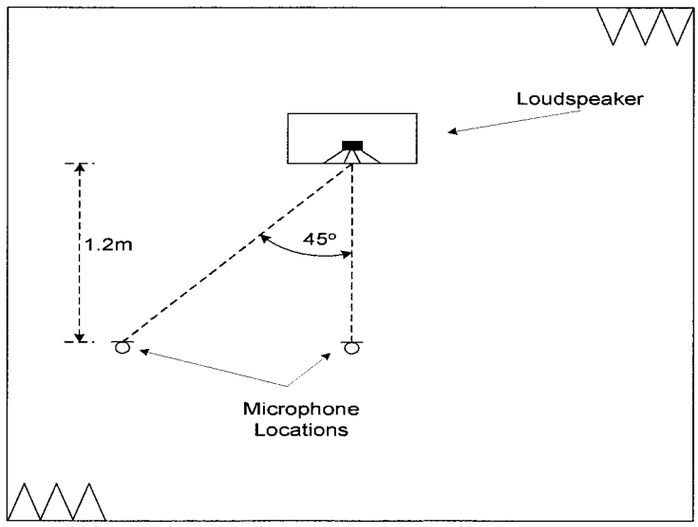
the woofer side. Due to the horizontal symmetry of the dual-concentric loudspeaker (loudspeaker B) only two impulse responses were measured for it, on-axis and at 45^o.

The IRs were measured in an anechoic environment using [6] the CRC-MARS (Multichannel Audio Research System) software developed at the Communications Research Centre (CRC). The excitation signal was a maximum-length sequence, [40], that was captured using an omni-directional measurement microphone. The length of the sequence was 32767 samples and was sampled at 44.1kHz. Synchronous averaging was carried out to improve the signal-to-noise ratio of the measurements. The IRs were computed from the circular cross correlation of the input with the output of the microphone signal [40]. The IRs were then truncated to 1024 samples (23.2ms).

The measurements, (I) impulse responses and (II) magnitude responses for the Mackie HR824 loudspeaker (Loudspeaker A) at the three microphones positions are shown in Figure 3.2. Similarly, Figure 3.3, shows the impulse response and magnitude response measurements for the Tannoy 800A loudspeaker (Loudspeaker B).

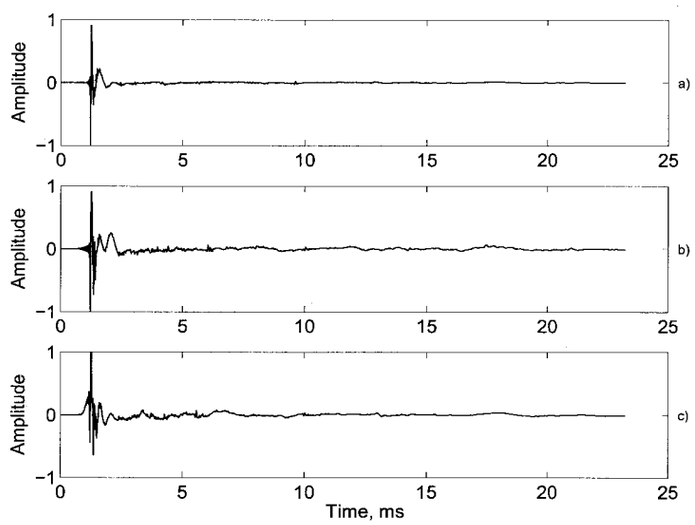


(a)

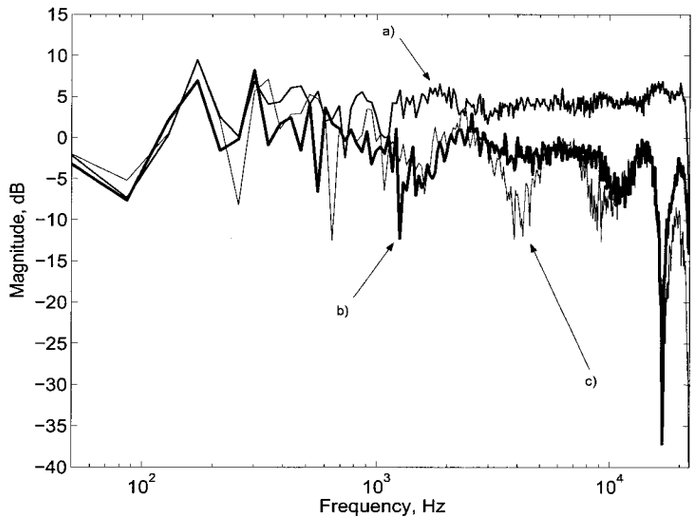


(b)

Figure 3.1: Loudspeaker and microphone layout for the impulse response measurements, a) for the Mackie HR824 (Loudspeaker A) and b) the Tannoy 800A (Loudspeaker B) .



(I)



(II)

Figure 3.2: Measurements of the Mackie HR824 (Loudspeaker A) loudspeaker. I) Impulse responses measurements for a) 0^o on-axis, b) 45^o on the tweeter side and c) 45^o on the woofer side. II) Magnitude response measurements for a) 0^o on-axis, b) 45^o on the tweeter side and c) 45^o on the woofer side.

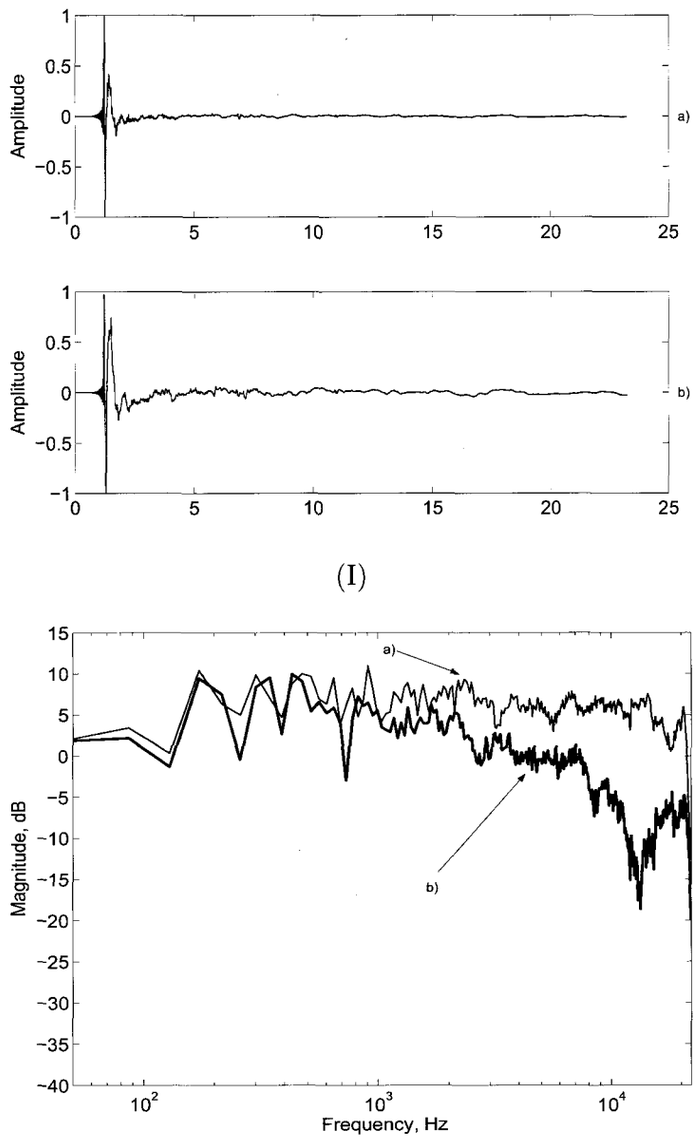


Figure 3.3: Measurements of the Tannoy 800A (Loudspeaker A) loudspeaker. I) Impulse responses measurements for a) 0^o and b) 45^o II) Magnitude response measurements for a) 0^o and b) 45^o.

3.2 Objective Evaluation

A common method of evaluating the difference in signals is to look at the root-mean-squared (RMS) error given by

$$error_{RMS} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(y_{reference}(n) - y_{calculated}(n))^2\right]^{\frac{1}{2}} \tag{3.1}$$

where $y_{reference}(n)$ is the reference signal and $y_{calculated}(n)$ is the test signal that has been calculated and is to be similar to the reference signal.

This method can be applied to the case of evaluating inverse filtering by comparing the desired signal to that of what is yielded with the inverse filter. In fact there are two RMS errors that could be calculated with inverse filtering. The first RMS error can be calculated from the difference between the desired filtered response $a(n)$ and the linear convolution of the impulse response $c(n)$ and inverse filter $h(n)$. This comes directly from Equation 2.3 where the definition of the inverse filter is introduced. Equation 3.1 would then become,

$$error_{RMS} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(a(n)) - (c(n) \otimes h(n))^2\right]^{\frac{1}{2}} \tag{3.2}$$

$$\tag{3.3}$$

A second RMS error can be calculated directly from the audio signal that a listener would hear. This would seem more appropriate to calculate and compare with subjective results. With this error one would need a reference audio signal $y_{reference}$ that is what the inverse filter should strive for when correcting an uncorrected audio signal. Using the notation from before for the impulse response, $c(n)$ and the inverse filter $h(n)$, the following is obtained:

$$error_{RMS} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(y_{reference}(n)) - (y_{reference} \otimes c(n) \otimes h(n))^2\right]^{\frac{1}{2}} \tag{3.4}$$

It is common to express the RMS error in decibels by the following formula,

$$error_{dB} = 20\log_{10}(error_{RMS}) \tag{3.5}$$

Using a flat RMS error might not seem like the best way of evaluating the various inverse filtering methods. Since these methods are for audio applications, an error that

is reflective on human hearing would seem appropriate. The most common frequency weighting method is A-weighting which approximates the 40 phon equal loudness contour from the Fletcher-Munson, [46], equal loudness plots. Two other common weightings are the B- and C-weighting which are approximations based on the 70 and 100 phon contour respectively. A final weighting based on recent loudness studies [43] is also shown in Figure 3.4.

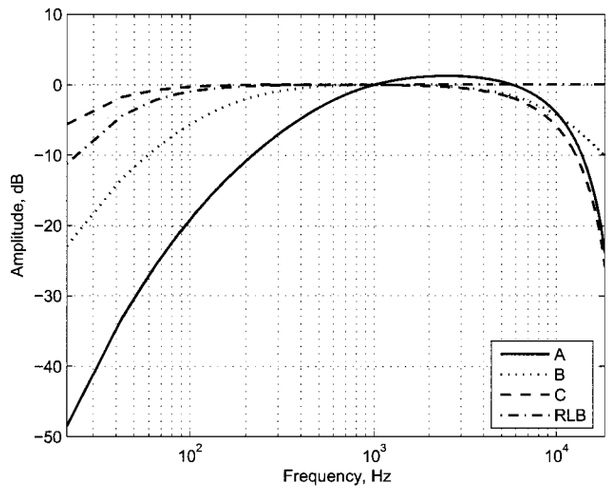


Figure 3.4: Frequency weighting used in the objective RMS measurements. The common A,B and C weightings based on equal loudness contours are included plus one based on recent loudness tests referred to as RLB.

The RMS error measures are potentially simple and easy ways of evaluating the various inverse filtering methods, but they may not be accurate enough when evaluating the audio performance. An objective measurement of high quality audio is PEAQ [17] which is an ITU-R standard. However PEAQ was developed for the evaluation of small impairments in high quality audio codecs.

3.3 Subjective Evaluation

The ultimate method of evaluating the performance of any audio system or algorithm is to carry out subjective tests. Proper former subjective tests are very time costly and time consuming. There are ITU-R standards [19] and [18] that outline the different subjective testing methods and strategies for evaluating high quality audio.

Most types of subjective tests are conducted in a "double-blind" scenario, where the subject as well as the observer/experimenter do not know the items being tested. This is to avoid any bias from knowledge of exactly what the subject is listening to. A standard audio subjective test, ITU-R-BS.1116 [19] compares two audio files, one being the unknown reference, to a known reference. The subject must decide on which of the two test files is the reference and score the quality of the other test file in relation to the known reference. This type of test is very long and can sometimes lead to large variances in test scores, so another test called Multiple Stimulus with Hidden Reference Anchors (MUSHRA) [18] was introduced. In this type of test the subject can instantly compare several test items to a known reference in order to derive a score for each item. The MUSHRA style of test has been shown to reduce the variance in the subject's scores [42].

To test the various inverse filtering methods that have been introduced in the previous section, a test procedure has been devised based on inverting the measured loudspeaker impulse responses. It consists of first filtering an audio file with an impulse response to arrive at a filtered audio file. Now a correction or inverse filter can be applied to this filtered audio file to try and "remove" the effects of the initial filtering. If the inverse filtering process is perfect, the corrected audio file should be identical to the initial audio file. Any degradation made by the inverse filtering will appear in the inverse-filtered audio file and can then be evaluated based on its audibility. By conducting tests in this manner, one can evaluate the effect of the inverse filtering by comparing the original reference audio file to the filtered audio file. When conducted in a controlled environment, one can formally evaluate the various inverse filtering methods.

Figure 3.5 shows a schematic of the listening test strategy with N system filters and M inverse or correction filters. The subject is able to switch between the original reference audio file to various filtered and corrected audio files for immediate comparison. By using this type of strategy, the subject can listen to the audio file at the various stages of being corrected: the reference, the filtered file (to simulate a loudspeaker's effect on the audio signal) and a loudspeaker corrected audio file that tries to remove the effects of the loudspeaker. The subject can then score each of the different audio files, and one can see the effects and improvements, if any, that the correction filters provide.

The audio playback system consisted of a standard PC running Microsoft Windows 2000 equipped with an RME Digi 96/52 audio card using Steinberg Nuendo A-to-D convertors. The audio was then played through STAX LAMBA PRO headphones.

Figure 3.6 shows the computer screen of the CRC testing software SEAQ [7] that the

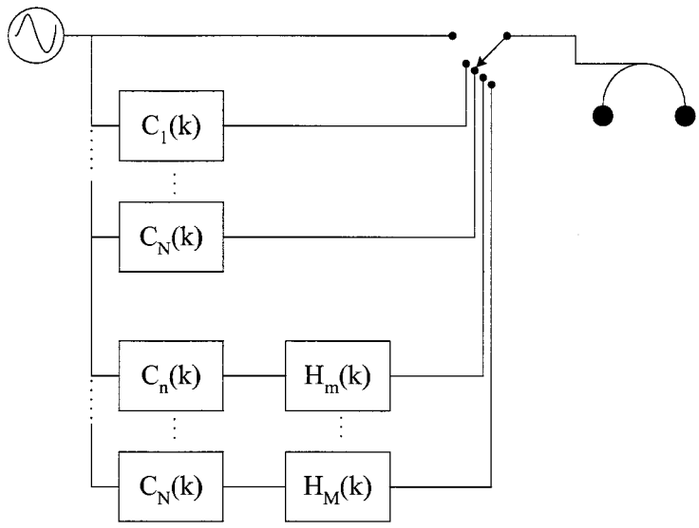


Figure 3.5: Schematic diagram showing the concept of the subjective test set-up. $C_n(k)$ is the response of the loudspeaker and $H_m(k)$ is the inverse filter used.

subject will see when a MUSHRA test is being conducted.

The subject clicks on any of the lettered buttons (ref, A - J) to listen to various unknown audio sequences, and assigns a score, by moving the slider on the right, to each one of them. The scores are saved on the computer for further analysis after the tests have be completed.

3.4 Initial Investigations

To get a better understanding of the inverse filtering techniques that are described in Chapter 2, single-channel versions of the methods have been implemented and evaluated. To test the various methods/techniques and measure their performance, formal subjective tests were conducted with a pool of 10 subjects. The complete subjective evaluation of the various inverse filtering methods are included in the paper [37] which includes results from the conference papers,[34][36][35]. An overview highlighting the results will be presented in the following sections.

An analysis of variance (ANOVA)[22] [5] was carried out on the data from all the subjective tests, so that meaningful results can be obtained. On the plots presented the error bars represent the critical difference for the experiment using a t-test. As such, any

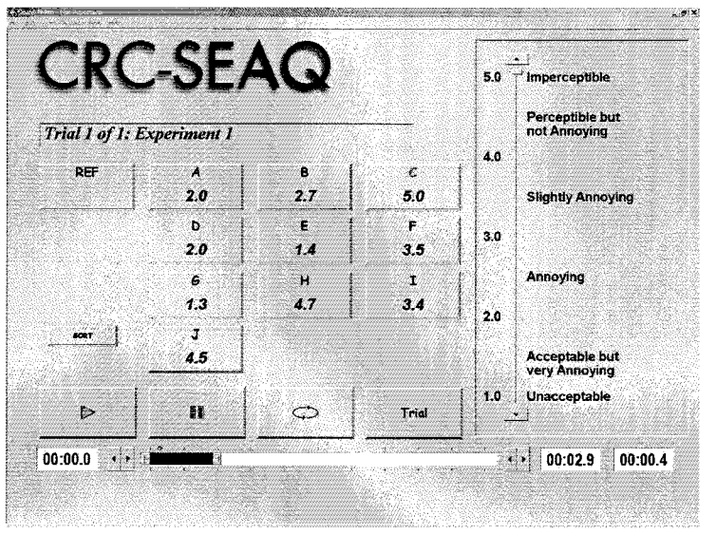


Figure 3.6: User interface of computer-based switching system used for the subjective tests.

two data points are statistically different ($p < 0.05$) if their error bars do not overlap.

3.5 Time- and Frequency-domain Comparison

The first methods to be examined were the time-domain least-squares and frequency-domain methods given by Equation 2.39 and Equation 2.50 respectively. The impulse responses (IRs) that were used were from the measurements of the loudspeakers that are shown in Section 3. The length of the IRs were 1024 samples and sampled at 44.1kHz. (total time of $\approx$ 23ms). Figure 3.7 shows the inverse filter with a length (N_h) of 2048 samples, calculated using Equation 2.39 with a modeling delay of 1024 samples. It can be seen that the inverse filter has pre-ringing or energy before the maximum main peak which would indicate that it is non-minimum phase. Also this could lead to the creation of audible artifacts that may degrade the audio quality.

The magnitude of the frequency response of the inverse filter in Figure 3.7 is shown in Figure 3.8. The magnitude of the frequency response of the original IR is shown by the thicker curve above the inverse filter magnitude of the frequency response. Figures 3.9 and 3.10 show the time response and magnitude response for the inverse filter calculated from the same IR but using the frequency-domain method.

The inverse filters shown in Figures 3.7-3.10 both look very similar and do correct

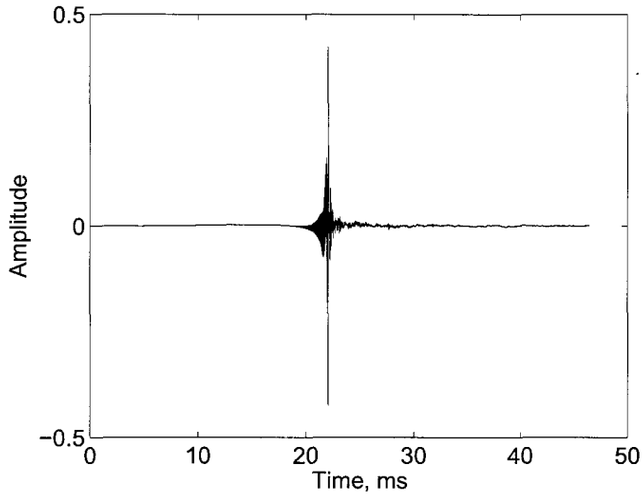


Figure 3.7: Inverse filter calculated using the time-domain least squares method calculated from the on-axis response of loudspeaker B. The IR had a length of 1024 samples, the inverse filter had 2048 samples and a modeling delay of 1024 samples was used.

the magnitude of the frequency response reasonably well as seen in Figures 3.8 and 3.10.

Similar inverse filters were calculated for the other loudspeaker IRs described in Chapter 3. A formal subjective test that was detailed in Chapter 3 was carried out to compare the inverse filtering methods. Along with the 5 loudspeaker IRs, the minimum-phase component of each of the IRs was also used to generate an inverse filter using the same lengths and modeling delay.

Figure 3.11 shows the mean subjective grades versus correction methods for a comparison of time- and frequency domain least-squares techniques along with a minimum phase correction. In the minimum phase correction case, an inverse filter based only on the minimum phase component of the impulse response was used. Also plotted in Figure 3.11 is the result of a hidden reference. A subjective grade of 5 indicates that the subject could not distinguish between the reference audio file and the test audio file. A subjective grade of 0 indicates the largest difference between the reference audio file and test audio file. The error bars on the plot represent the critical difference for the experiment derived using a t-test [22] [5]. As such, any two data points are statistically different ($p < 0.05$) if their error bars do not overlap.

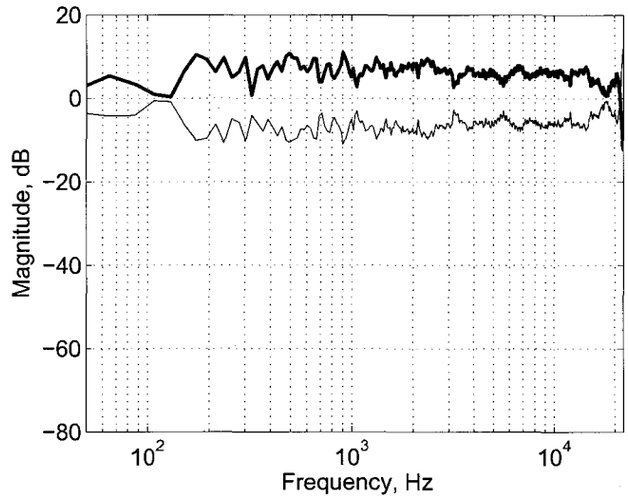


Figure 3.8: The magnitude of the frequency response of the inverse filter shown in Figure 3.7. Plotted above by the thicker curve is the magnitude of the frequency response of the original IR of loudspeaker B.

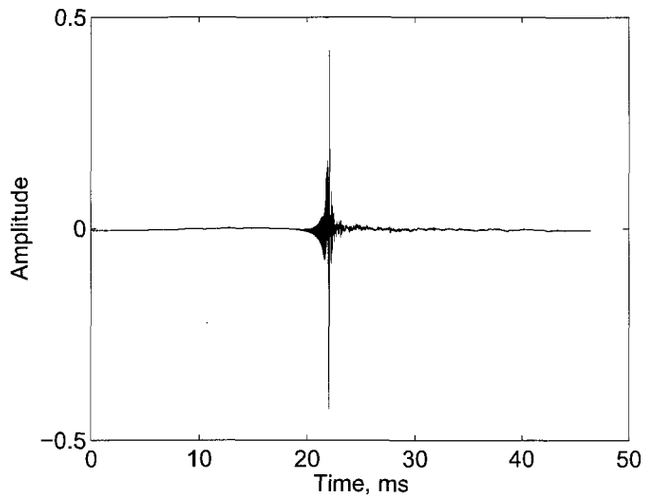


Figure 3.9: Inverse filter calculated using the frequency-domain method calculated from the on-axis response of loudspeaker B. The IR had a length of 1024 samples, the inverse filter was 2048 samples and a modeling delay of 1024 samples was used.

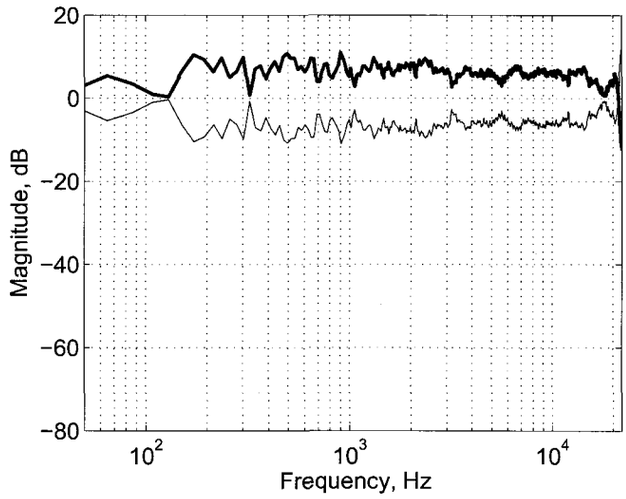


Figure 3.10: The magnitude of the frequency response of the inverse filter shown in Figure 3.9. Plotted above by the thicker curve is the magnitude of the frequency response of the original IR of loudspeaker B.

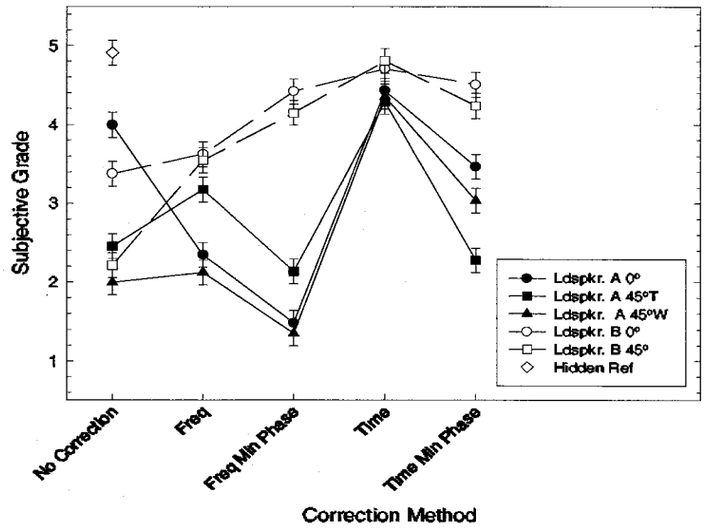


Figure 3.11: Mean subjective grades versus correction method from the time- and frequency domain method comparison. The 5 curves represent the 5 different IR configurations, 3 for loudspeaker A and 2 for loudspeaker 3.

From Figure 3.11 it can be seen that the time-domain least squares approach is more robust than the frequency-domain method. For all 5 conditions, the time-domain inverse filters always scored above 4, whereas the frequency-domain inverse filters were spread over a broader range. It should be noted that the minimum-phase inversion case had a modeling delay added which is not necessary and could actually cause more audible artifacts, especially in the frequency-domain deconvolution case. By adding a delay, this shortens the length of the filter that follows the main peak potentially causing more energy to be wrapped around before the peak and therefore producing an audible artifact.

Figure 3.12 shows the on-axis corrected time responses for loudspeaker B. The inverse filters were designed so that the desired or corrected time-responses should be perfect Dirac delta functions with a known delay. Even though the time-domain corrected inverse filter, Figure 3.12 a) scored better (with a score of 4.7) than the frequency-domain correction Figure 3.12 c) (with a score of 3.6); it is difficult to see the difference by comparing these figures since they both look like "perfect" delta functions. Fielder [9] mentions that viewing the corrected responses this way can be misleading and suggests plotting the time-responses on a logarithmic scale.

Figure 3.13 shows the same corrected time-responses of Figure 3.12 but plotted on a logarithmic scale. It is now more clear why the frequency-domain inverse filter performed

more poorly than the time-domain one. At the beginning of the corrected time-responses there is an artifact that stands out and it is -60dB down in the time-domain version whereas in the frequency-domain inverse filter it is only -40dB down. Listening to the frequency-domain inverse filter one can detect an echo that is created by this artifact. The time-response of the frequency-domain inverse filter also has a large area of "zeros", starting at around 23ms, after the main pulse. This is due to the circular deconvolution that was used to calculate the inverse filter and then being linearly convolved with the original impulse response.

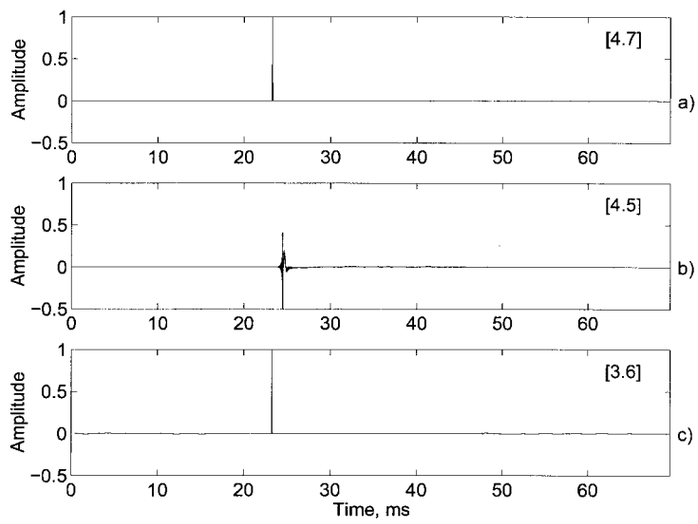


Figure 3.12: On-axis corrected time response for loudspeaker B: a) 2048 LS time-domain correction, b) 2048 minimum-phase LS time-domain correction, c) 2048 frequency-domain correction. The subjective grades are shown in square brackets []

Figure 3.14 show the corrected magnitudes of the frequency-response for the conditions shown in Figure 3.13, plus the uncorrected response. It can be seen that the time-domain method, curve b) is closest to a flat line, which was the target. The other curves show more variation, especially at the lower frequencies. Longer inverse filter lengths should improve the lower frequency response, and also help with the circular deconvolution effects.

A further test was then carried out the explore the circular deconvolution effects when the length of the inverse filter is increased. Therefore, inverse filters were calculated for lengths of 2048, 4096, 8192, 16384, 32768 and 65536 samples in length.

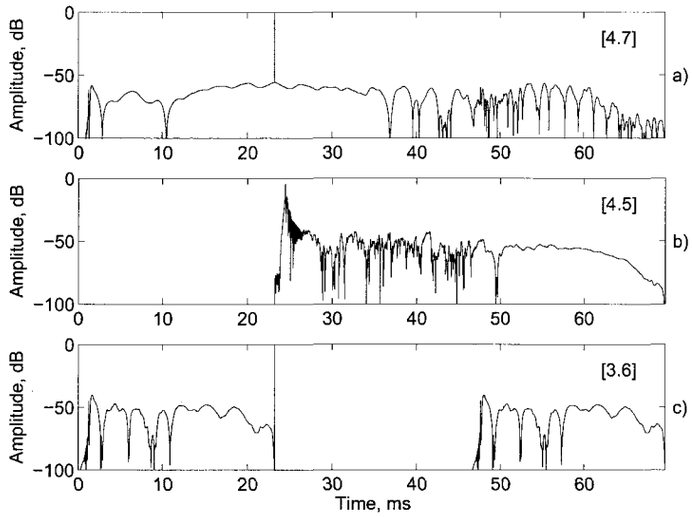


Figure 3.13: On-axis corrected time response for loudspeaker B plotted on a logarithmic scale: a) 2048 LS time-domain correction, b) 2048 minimum-phase LS time-domain correction, c) 2k frequency-domain correction. The subjective grades are shown in square brackets []

Figure 3.15 shows the subjective results when various length of frequency-domain inverse filters are used. When increasing the length of the inverse filtering, the subjective grades tended to go up, though at different rates. Thus, when the length of the frequency-domain inverse filters is increased the circular deconvolution effects are minimized as expected. Though effect is filter dependent due to the characteristics of the original impulse response.

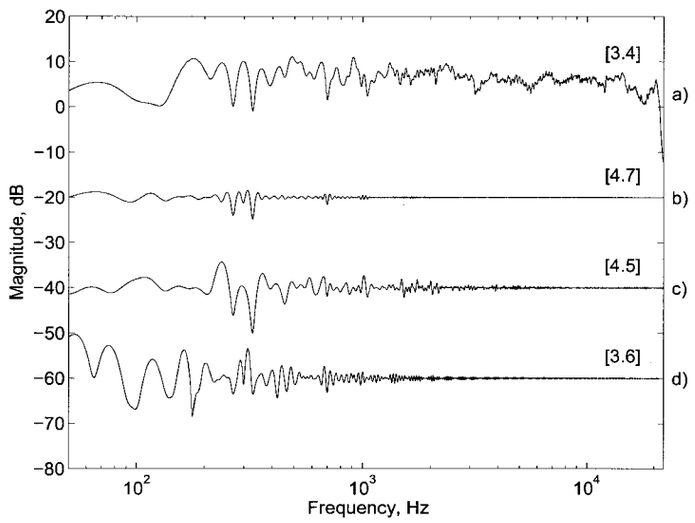


Figure 3.14: On-axis magnitude frequency response and corrections for loudspeaker B: a) uncorrected response, b) corrected with LS time-domain 2048 filter, c) corrected with minimum phase LS time-domain 2048 filter, d) corrected with frequency-domain 2048 filter. The curves are offset for clarity of presentation. The subjective grades are shown in square brackets []

Figure 3.16 shows the corrected time-responses for loudspeaker A, off-axis (tweeter side). Figure 3.16 a) is the LS time-domain corrected response, where b) and c) are frequency-domain corrected responses with inverse filter lengths of 8192 and 32768 respectively. The effect of increasing the length of the frequency-domain inverse filter is shown. As the length increases, the artifacts are lowered in amplitude, but are also pushed out farther in time from the main peak. Being pushed out in time can make it more audible since it may not be masked. The masking characteristics are not only dependent on the location of the error with respect to the main peak, but also on the relative amplitude. Therefore, by lowering the amplitude of the error, it helps to make the error less audible.

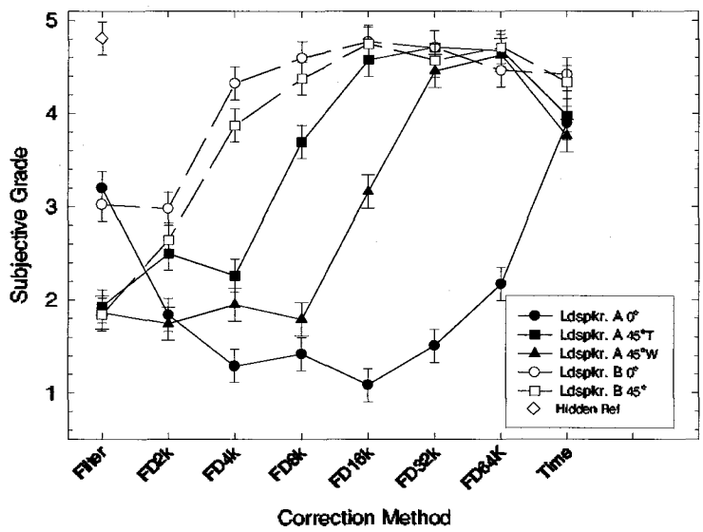


Figure 3.15: Mean subjective grades versus length of the frequency-domain inverse filter for the 5 different loudspeaker configurations. The length of the inverse filters are included in the label of the frequency-domain (FD) method: FD2k means a 2048 length filter was used. The time-domain filter was 2048 samples long.

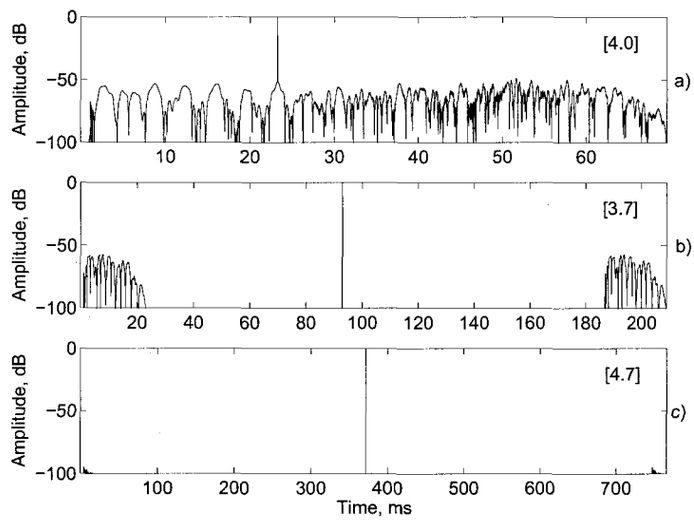


Figure 3.16: Off-axis (tweeter side), corrected time response for loudspeaker A: a) 2048 LS time-domain correction, b) 8192 FD correction, c) 32768 FD correction. The subjective grades are shown in square brackets []

3.6 Minimum-phase comparisons

In the previous section, minimum-phase correction filters were used with a modeling delay, but a delay should not be necessary. The inverse of a minimum phase impulse response will also be minimum phase and will be causal without any modeling delay added. By only using minimum-phase correction filters, one can eliminate the occurrence of pre-echo. A minimum-phase inverse filter can be created by only using the minimum-phase component of the impulse response to calculate an inverse filter. In the derivations of calculating an inverse filter that have been given, a modeling delay is included so that the inverse filter will be causal. Therefore, what are the consequences of including a modeling delay when the inverse filter is minimum phase. Figure 3.17 shows the subjective results when two different lengths of frequency-domain inverse filters are used and minimum-phase inverse filters with and without a modeling delay. By adding a modeling delay, one allows the error signal to occur before the main peak and thus can degrade the audio performance.

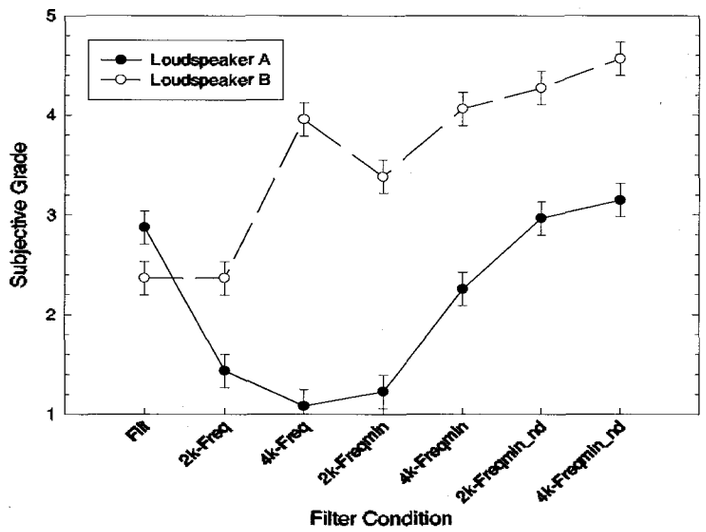


Figure 3.17: Mean subjective grades versus correction filter condition for the 2 loudspeakers and two filter lengths, 2048 and 4192 are used. Full impulse response inversion and minimum(min) phase component inversion with and without(nd) a modeling delay are included.

Another observation from Figure 3.17 is the large difference between the subjective scores for the two loudspeakers. Loudspeaker A always scores lower than Loudspeaker B

in all the inverse filtering cases. This difference implies that Loudspeaker A is more difficult invert than Loudspeaker B, which could be due to zero or near-zero terms in the Loudspeaker A response. The near zero terms would mean that a longer inverse filter would be needed to correct the impulse response.

3.7 Regularization

To test the effects of regularization, three types were implemented as illustrated in Figure 2.4 and applied to Equation 2.72, the frequency-domain deconvolution method. Due to the large number of combinations of IRs, regularization types, and levels, smaller preliminary subjective tests were conducted (with only 3 subjects) to explore the effects of regularization and narrow down the number of test cases for a formal subjective test.

As stated before, the IRs were normalized so that the maximum value of the magnitude response $|C(k)|^2$ was 0 dB (or a value of 1). Values of β ranged from 10^{-5} to 7×10^{-1} for all three types of regularization. With the scalar method, $B(k)$ was simply set to unity for all frequencies. The first vector or frequency-dependent method of regularization mimicked curve b) from Figure 2.4 with the low-frequency cutoff set at 80Hz and the high-frequency cutoff set at 18kHz, meaning that $B(k)$ was set to 0 between 80Hz and 18kHz and set to 1 above and below those frequencies respectively. The second frequency-dependent regularization method was based on the $\frac{1}{3}$-octave spectrum of $C(k)$ as was implemented in [9].

As with the previous formal subjective tests, an ANOVA was performed on the subjective test data, and the results for one part of the preliminary tests are shown in 3.18. This result shows the use of the first frequency-dependent based regularization method in computing the inverse filter to correct the off-axis response of the loudspeaker A.

The results shown in 3.18 indicate that for this loudspeaker there is no advantage in using regularization. That is, there is no value of regularization that gives a subjective grade that is higher than the grade obtained with no regularization ($\beta = 0$). Regularization values of $10^{-5} \leq \beta \leq 7 \times 10^{-1}$, yield acceptable results in that it gives the same result as with no regularization. The plot also shows that adding too much regularization, $\beta \geq 10^{-1}$ actually degrades the audio quality to the point where using an inverse filter is worse than having no correction at all. Therefore care must be taken in setting the amount of regularization. A similar result was also found with scalar regularization , although the range of acceptable values was even smaller.

Figure 3.19 shows the corrected off-axis time response for loudspeaker A, (a) without regularization and (b) with the first frequency-dependent form of regularization, $\beta = 10^{-3}$. It can be seen that the effect of regularization is to push the residual energy closer to the main pulse, which helps to reduce the audibility of echoes. However, by doing so the regularization causes a widening of the pulse or delta function, which in turn can create audible artifacts that are similar to the pre-echo artifacts found in some perceptual

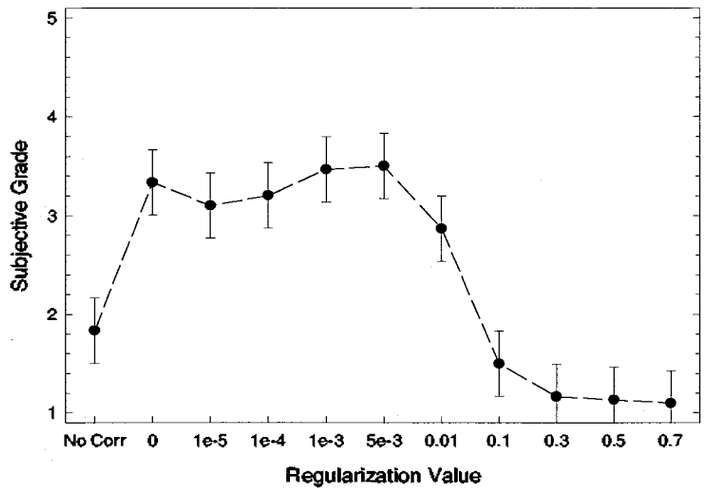


Figure 3.18: Mean subjective grades versus amount of regularization used for off-axis (tweeter side) correction for loudspeaker A. The first vector based regularization was used in the inverse filter calculation.

coders.

The preliminary tests revealed a major limitation of simple scalar and frequency-dependent regularization methods. The optimal regularization value is dependent on the IR that is being inverted. Using these methods therefore, one needs to hand tune the regularization when designing an inverse filter.

From the preliminary tests two IRs were chosen for a formal subjective test: on-axis and off-axis(tweeter side) IRs from loudspeaker A. Two values of β were chosen for each of the three regularization types based on the findings from the preliminary tests. These were selected to demonstrate the range of performance of the regularization with the two different IRs. For comparison, uncorrected and full correction with no regularization inverses were also included. A total of 10 subjects conducted the formal subjective test. An ANOVA was conducted on the results of the formal test, and the results are plotted in Figure 3.20, with the error bars representating the critical difference. The figure shows the mean subjective grades versus the regularization condition, which includes the type of regularization, if used and the value of β.

It can be seen that for the on-axis response, the frequency-domain inverse filtering without regularization actually degrades the audio quality. This is in agreement with previous results. Adding a small amount of regularization (10^{-5}) does not provide any improvement. However, increasing the amount of scalar regularization to 10^{-2} provides

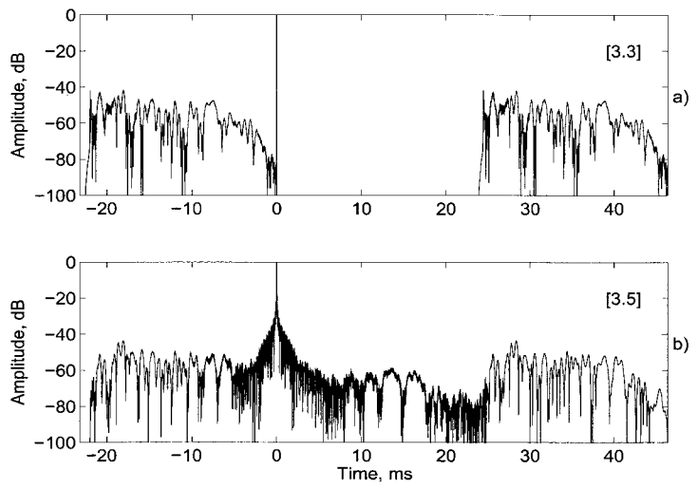


Figure 3.19: Off-axis(tweeter side) corrected time response for loudspeaker A: a) 2048 frequency-domain inverse with no regularization, b) 2048 frequency-domain inverse using vector regularization type 1 with $\beta = 10^{-3}$. The subjective grades are shown in the square brackets [].

a dramatic improvement in the subjective performance of the inverse filter. The off-axis response yields very different results. Good performance is achieved with an inverse filter using no regularization, and there is not much change when a small amount of regularization, 10^{-5} is added. With the larger amount of regularization, 10^{-2}, the subjective performance is degraded some. Thus the optimal value of β for scalar regularization depends on the characteristics of the IR being inverted. It should be noted that the mean level of the off-axis inverse response is 6 dB higher than the on-axis response. This seem counter intuitive, since the off-axis response needs less regularization than the on-axis response.

The on-axis response has a zero near dc that causes the inverse filter to ring for a long period of time, which in turn produces audible wrapping effects in the frequency-domain deconvolution method. The regularization limits the ringing of the inverse filter, hence reducing the wrapping effects.

The first frequency-dependent regularization method appears to be more robust in the selected range. That is, the two levels of regularization give the same perceptual benefits for the two IRs being corrected. It should be recalled, however, that these two levels of regularization were chosen as a results of the preliminary subjective tests. As such, these

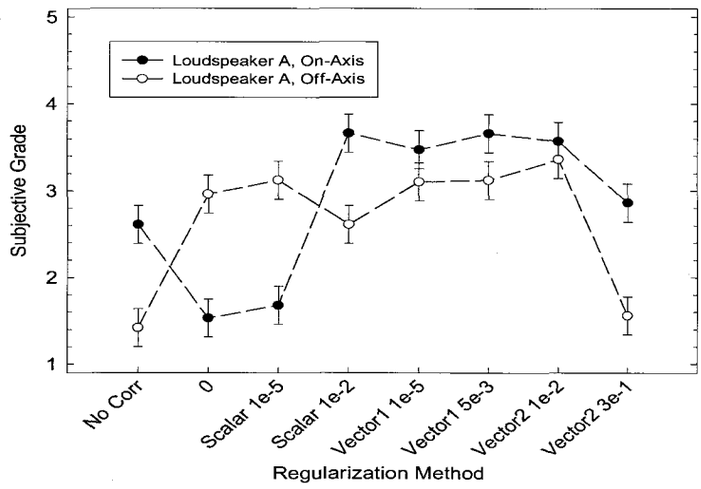


Figure 3.20: Mean subjective grades versus regularization type and amount used for loudspeaker A, on/off axis correction.

values of regularization are the result of a hand-tuning process. Nonetheless, the first frequency-dependent approach provided more robust results than the scalar method.

The performance of the second frequency-dependent ($\frac{1}{3}$-octave based) regularization method was not as robust. Smaller amounts of regularization (10^{-2}) gave good subjective results for both the on- and off-axis responses. However, the higher regularization level (0.3) resulted in a significant drop in audio quality. In this case no subjective benefit was gained from the inverse-filtering process.

Overall, correctly chosen regularization provided a significant subjective improvement when correcting the on-axis response. Conversely, for the off-axis case, regularization did not provide any subjective benefit as compared to not having any regularization.

In previous experiments it was found that the on-axis IR for loudspeaker A did not invert very well using the frequency-domain method with a length of 2048 samples. Figure 3.21(a) shows the corrected response without regularization, which received a subjective grade of 1.5. It can be seen that the level of the residual(uncorrected) energy is only about 30 dB below the level of the main peak. Since this residual energy is relatively far way from the main peak (20 ms before and 50 ms after), it is readily audible as time-domain artifacts.

Figure 3.21(b) shows the corrected response when a small amount ($\beta = 10^{-2}$) of scalar regularization is added. This received a subjective grade of 3.7 and is therefore

perceptually much better than the response in Figure 3.21(a). It can be seen that in this case the residual energy is always below 50 dB.

Figure 3.21(c) shows the corrected response using the first frequency-dependent regularization method with $\beta = 5 \times 10^{-3}$. This also received a subjective grade of 3.7. Again, it can be seen that, except for the area very near the delta function, the level of the residual energy is always below -50 dB.

These results suggest that the role of correctly chosen regularization is to "shape" the uncorrected energy in such a way that it is less perceptible. It should also be noted that none of the frequency-domain filter inversion methods performed as well as the time-domain LS approach tested earlier.

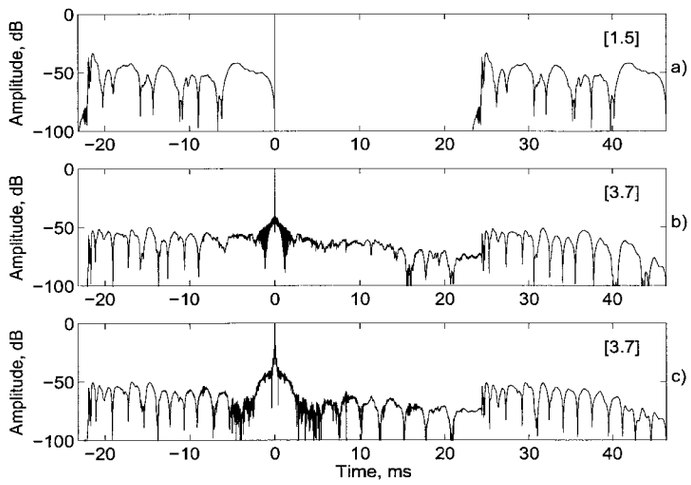


Figure 3.21: Off-axis(tweeter side) corrected time response for loudspeaker A: a) 2048 frequency-domain inverse with no regularization, b) 2048 frequency-domain inverse using scalar regularization with $\beta = 10^{-2}$ and c) 2048 frequency-domain inverse using vector regularization type 1 with $\beta = 5 \times 10^{-3}$. The subjective grades are shown in the square brackets [].

Kirkeby et al. [25] showed the effect of regularization on $H(z)$ to be that it replaces a z-domain pole near the unit circle with a pair of poles and a zero. As pointed out by Fielder [9], one of the newly created poles is often outside the unit circle. This would create a nonminimum-phase element and be acausal and would generate undesirable pre-ringing effects.

3.8 Complex Smoothing

Another way to control the performance of an inverse filter is to smooth the frequency-response function, to reduce the severity of the peaks and dips. Traditional spectral smoothing operations only use the power spectra, whereas Hatziantoniou and Mourjopoulos [12] have proposed a method of complex smoothing as a way of smoothing a transfer function and recovering a time-response. Using equation 2.79 and a smoothing window defined by 2.82 with $b = 0.7$ and a width defined as a $\frac{1}{6}$ of an octave.

Figure 3.22 shows the on-axis magnitude and phase response for loudspeaker A and the $\frac{1}{3}$-octave complex smoothed version of each. As with the IR the effect of the smoothing is quite apparent.

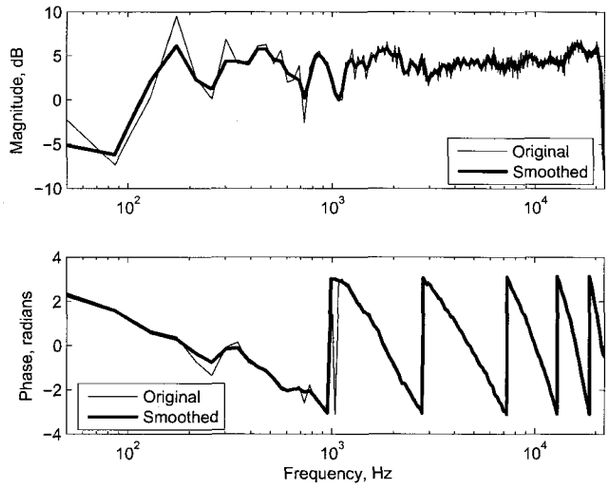


Figure 3.22: Magnitude and phase of frequency response of loudspeaker A (on-axis) showing the original and $\frac{1}{3}$-octave complex smoothed version.

For this subjective test, inverse filters employing complex smoothing were compared directly with nonsmoothed versions. Four IRs were used in this experiment. They included the on- and off-axis (woofer side) IRs of loudspeaker A and both IRs of loudspeaker B. Therefore a total of 12 filtering scenarios were evaluated, two inverse-filtering methods for each of the four loudspeakers IRs, plus an uncorrected version. The frequency-domain deconvolution method was used with a inverse filter length of 2048 samples. One inverse filter was calculated from the original IR and the other was calculated from the complex smoothed IR.

An ANOVA was conducted on the results and showed highly significant main effects ($p < 0.001$) due to loudspeaker type and inverse-filtering method. The overall mean subjective grade versus filter condition results are shown in Figure 3.23 where "Filter" indicates the situation with no correction, "2k-Freq" means a 2048 frequency-domain deconvolution inverse was used, and "2k-Freq-Sm" indicates a 2048 frequency-domain deconvolution inverse from the complex smoothed IR was used.

From Figure 3.23 it appears that the complex smoothing provides a statistically significant improvement in the performance of the inverse filter. A plot of the individual loudspeaker means versus filter conditions is shown in Figure 3.24. It can be seen that the loudspeaker corrections tended to perform better with the smoothed correction, as seen before from the combined results. The exception is the on-axis response of loudspeaker A, which showed some improvement, although not statistically meaningful. On the positive side, the smoothing did not degrade the performance in comparison to the non-smoothed versions.

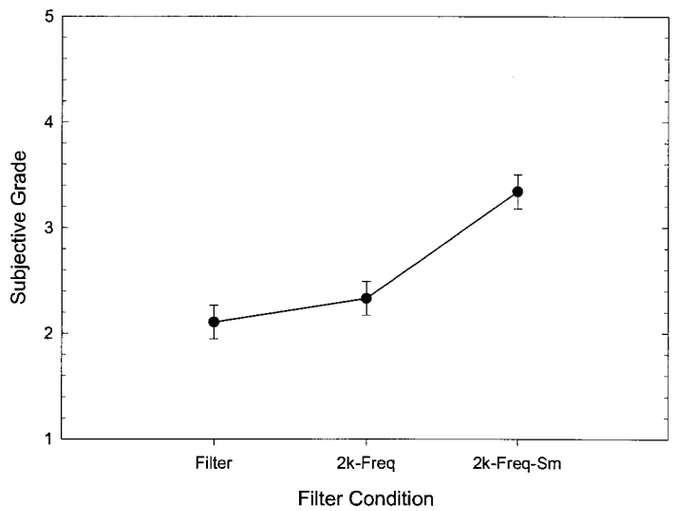


Figure 3.23: Mean subjective grade versus filter condition, not smoothed and complex smoothed (sm) for the average of four of the loudspeaker configurations.

Figure 3.25 shows the corrected off-axis IRs for loudspeaker B, without and with complex smoothing. The main pulse occurring at about 23 ms is still very narrow, as opposed to the regularization case shown in Figure 3.19, in which it was broadened. The blocking artifacts that are present near 0 ms and 47 ms are reduced by 10 dB in the

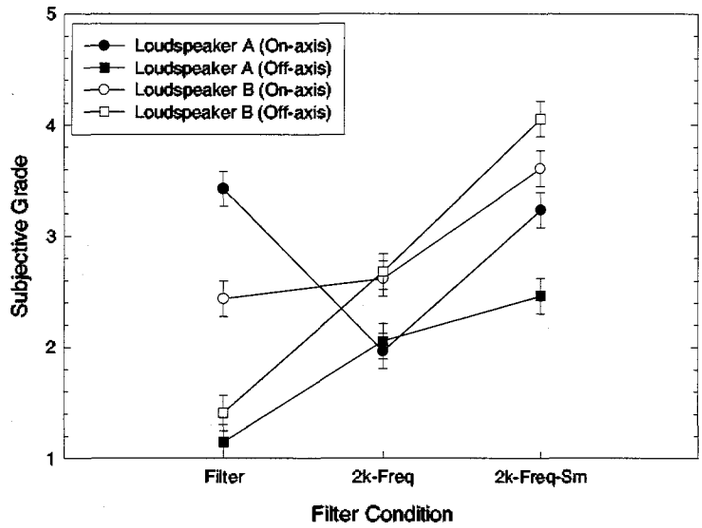


Figure 3.24: Mean subjective grade versus filter condition, not smoothed and complex smoothed (sm) for four of the loudspeaker configurations.

smoothed case. In all cases the smoothing never degraded the subjective performance of the inverse filter over the case when no correction filter was added.

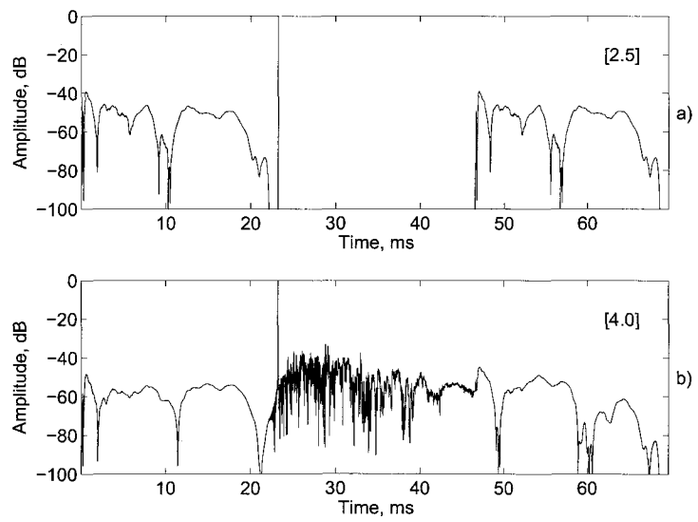


Figure 3.25: Off-axis corrected impulse response for loudspeaker B: a) 2048 FD correction, b) 2048 FD correction from the complex-smoothed IR. The subjective grades are shown in square brackets [].

Though it cannot be statistically shown here, it would appear that the complex smoothing may perform better than the regularization methods that were evaluated. Both techniques help and reduce the time-aliasing effects associated with the frequency-domain deconvolution method, though it may appear that the complex smoothing might perform better. One reason is due to the fact that the regularization tends to broaden or smear the main peak of the response, i.e. produce some pre-response, whereas the complex smoothing tends not to effect the main peak as much. This can be seen if one compares Figures 3.25 b) and 3.19 b), where the later case, with regularization, has a main peak that is spread out some.

3.9 Objective Evaluations

In the above section, various inverse filtering methods were implemented and evaluated by means of formal subjective tests. Even though subjective testing is the best way to evaluate such inverse filtering methods, it is very time consuming and is not always practical. Objective methods of evaluation are usually employed to evaluate the performance of filter, and as stated previously the RMS error metric is a common way of expressing the error. Using Equations 3.4 and 3.5, the RMS errors for all the audio sequences that were used in the subjective tests in the previous section were calculated. Various RMS errors were calculated using different frequency-weightings as shown in Figure 3.4. Also included are two results from PEAQ, an objective audio evaluation piece of software for high quality audio [7]. In order to compare these objective errors, they were correlated with the subjective results. The correlation r is given by,

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2(y-\bar{y})^2}} \tag{3.6}$$

where, $\bar{x}$ indicates the average value of x. For the comparisons here, a vector of the average subjective grade was correlated with a vector of the RMS error for each audio sequence. A correlation coefficient of 1 indicates a perfect correlation while a correlation coefficient closer to 0 means less correlation. Table 3.1 shows the results.

Experiment	dB	dBA	dBB	dBC	dBrlb	PEAQ-basic	PEAQ-Advanced
1	0.596	0.619	0.596	0.627	0.600	0.907	0.944
2	0.639	0.636	0.639	0.652	0.632	0.911	0.806
3	0.595	0.682	0.595	0.690	0.596	0.947	0.927
4	0.102	0.153	0.102	0.165	0.112	0.845	0.855
5	0.328	0.345	0.328	0.342	0.339	0.765	0.740

Table 3.1: Table showing absolute correlation values of the various objective measures with subjective test scores.

It can be seen that the simple RMS error methods do not correlate vary well to the subjective test results and that PEAQ seem to correlate better. There are some variations in the correlations between tests, for example, experiment 4 has very lower correlations across all objective metrics, though especially with the RMS errors. Experiment 5 shows the lowest correlation for PEAQ, though it was not the lowest for the RMS errors. These last two tests, experiments 4 and 5 dealt with regularization and smoothing, two processes

that PEAQ may not be "trained" to deal with. Even with that being said, PEAQ could be used as a preliminary method to evaluate the inverse filtering methods, and a final subjective would be needed to confirm final results.

3.10 Chapter Summary

The first set of tests compared the time-domain and frequency-domain inverse filtering methods, and showed that the time-domain method outperformed the frequency-domain method in all cases, for the considered impulse responses and under these filter lengths. The frequency-domain method was hindered by time-aliasing or blocking effects associated with frequency-domain block processing which produced audible artifacts such as delays and pre-echoes. By increasing the length of the inverse filter, the time-aliasing effects can be reduced and in theory the performance will approach that of the time-domain least-squares method. Regularization and complex smoothing were used to help the performance of the frequency-domain method. By using regularization, one can better control the amount the work the inverse filter does at specific frequencies with respect to correcting the original response. The required amount of regularization needed must be fine tuned to achieve the best perceptually sounding inverse filtering, and if too much regularization is added, the result could be worse than if no inverse filter was used at all. One effect of regularization on the corrected impulse response is that it tends to broaden the main peak, thus creating a pre-response that can be audible. Complex smoothing was used and showed promise as it does not tend to broaden the main peak as much.

Chapter 4

Minimum-phase Regularization Techniques

In this chapter some psychoacoustic concepts are discussed and then a new technique to remove the pre-response created by the regularization technique is introduced. This new technique is developed for single and multichannel scenarios.

4.1 General Psychoacoustics

The purpose of inverse filtering is to correct and/or improve an audio system. By "improve" one means to make the system sound better or show some form of perceptual improvement in audio quality. With this being said, one must understand some of the characteristics of the human auditory system. Without going into a physiological description of the actual auditory system, the following is a brief review of some basic principles.

Not only do humans have a limited range of hearing, from 20 Hz to 20Khz, but in that frequency range we do not perceive all frequencies to be equally loud. For example we are less sensitive to sounds at the lower frequencies than at middle frequencies. This frequency sensitivity variation is also dependent on the absolute level of the sound. Figure 4.1 shows what is commonly referred to as Fletcher-Munson equal loudness contours for various sound levels. On the plot each line is an equal loudness contour, meaning that any sound on the line is equally loud and represents a certain level in phons. The unit of phons is normalized to the 1000 Hz SPL level, so that a 100 dB SPL sound would have a loudness of 100 phons. Therefore, going along the 100 phon contour, a sound at 50 Hz

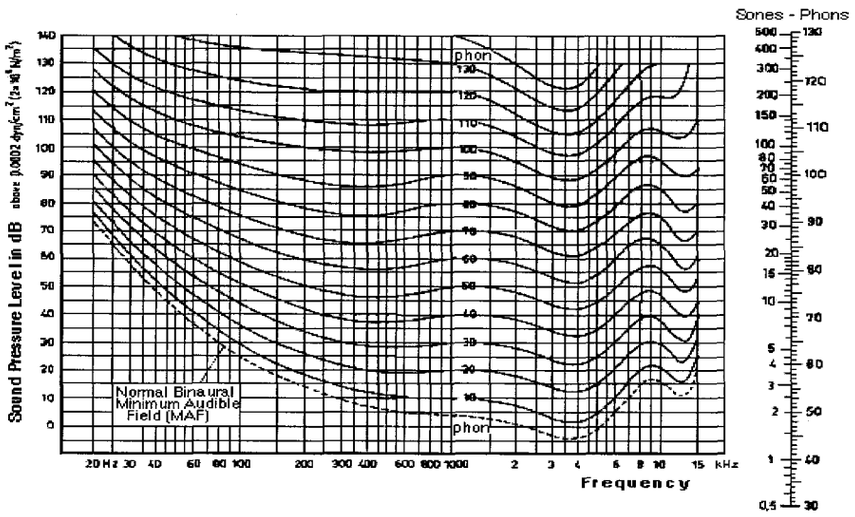


Figure 4.1: Equal loudness curves for the human ear, taken from [45]

would have to be at level of 110 dB SPL to be perceived as the same loudness as sound at 1000 Hz at 100 dB SPL. The very lowest curve in Figure 4.1 represents the level of audibility across the frequency spectrum. Any sound below this curve would therefore be inaudible by humans. This curve can be approximated by,

$$T_q(f) = -3.64\left(\frac{f}{1000}\right)^{-.08} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 dBSPL. \tag{4.1}$$

An important property of the human auditory system is the frequency selectivity and masking characteristics that it exhibits. First introduced by Fletcher, the concept of critical bands helps to explain them. A critical band is best defined as the range of frequencies around a pure tone where the auditory system does perceive other tones being present when played simultaneously. The other tones that are played are masked by the original tone. This frequency range is commonly referred to as the critical bandwidth, and is frequency dependent and Figure 2.6 shows the frequency bandwidth dependency for critical bands (or Bark scale), where the bandwidth increases with frequency.

The concept of critical bands is the basis of simultaneous masking, where a louder tone or noise signal, the masker, can make another tone or signal, the maskee, become inaudible. The masker will mask anything that falls below the masking threshold for that signal. Figure 4.2 illustrates this principle with a 1 kHz sine wave masker.

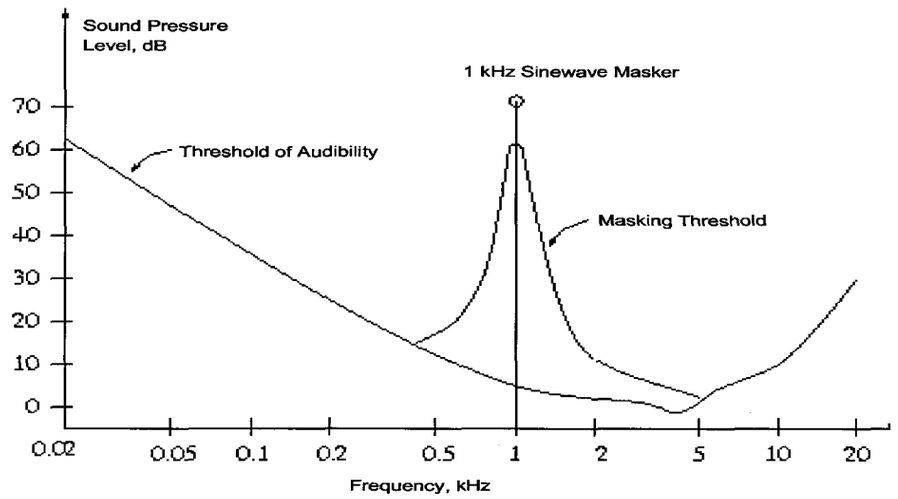


Figure 4.2: A masking tone and its masking threshold where a signal below the threshold will not be heard.

A masking threshold can be seen and signals that fall below it will not be heard. Masking is level and frequency dependent, and also depends on the type of sound, such as a pure tone, noise-like or transient.

Another type of masking, referred to as non-simultaneous masking (forward- and backward-masking or pre- and post-masking) pertains to masking in the time domain. Similar in concept to simultaneous masking in the frequency-domain in that one sound will mask or hide another sound, non-simultaneous masking is best understood in the time-domain. Non-simultaneous masking can be broken down into two components: forward-masking and backward masking. Forward-masking occurs when the masker occurs before the masked sound, whereas backward-masking occurs when the masker comes after the masked sound. Forward-masking is the dominant of the two and can provide significant amount of masking up to 150 to 200 ms after the stop of the masker. Backward-masking provides far less masking, in the range of 15 ms, and studies have even shown that trained listeners can exhibit almost no backward masking [28].

Figure 4.3 illustrates the concept of time-domain masking. Here there is a signal at time 0 ms, and it will create a masking threshold that ranges from about -15 ms to about 100 ms after. Any signal underneath this masking threshold will not be heard. For example, the signal labeled a), would be heard whereas signal b) would not be heard. These signals occur after the masker, but as can be seen from the masking threshold,

signals occurring before the masker can also be masked, though not to a large extent.

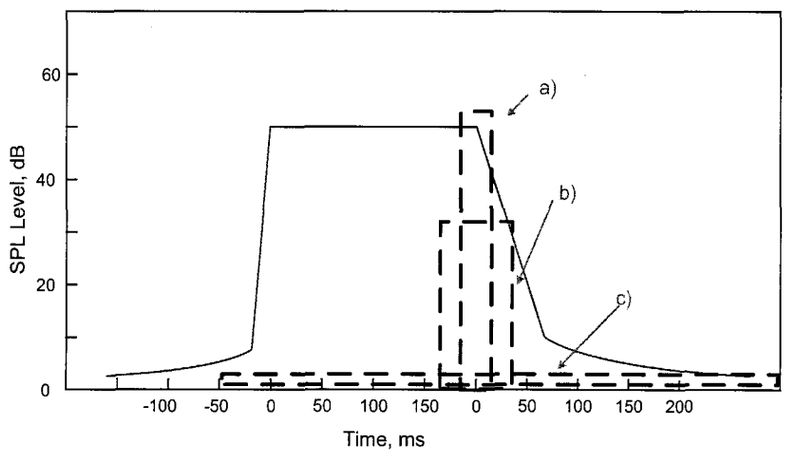


Figure 4.3: Masking signal at 0 ms creating a masking threshold before and after it in time. A signal above the threshold, a) would be heard whereas one below, b) would not be.

An aspect of non-simultaneous masking that is not illustrated in Figure 4.3 is the frequency-dependence. It has has been known that the amount of masking is level-dependent as well as frequency dependent [28]. A study by Jesteadt *et* al. [21] showed a frequency and level dependence, and derived a relationship between masking level, frequency and delay.

4.2 Minimum-phase Regularization

One of the issues that arose in the inverse-filtering techniques that were evaluated in Chapter 3, was the introduction of a pre-response in the corrected impulse response. This was especially true in the case when regularization was added. This was due to the regularization being zero-phase, thus adding a symmetric response in the time-domain around the main peak. In this chapter a method of forcing the regularization to be minimum-phase will be introduced.

4.2.1 Single-Channel Target-function Magnitude from Regularization

If we start with the frequency-domain inverse filter given by Equation 2.72 for ease of notation, and compare two cases: one with no regularization, i.e. $\beta = 0$, and the other with regularization while the target function is a delay, T_d or e^{-jkT_d}. For the case with no regularization Equation 2.72 simplifies to,

$$H(k) = \frac{A(k)C^*(k)}{|C(k)|^2}. \tag{4.2}$$

Setting the target function to a simple delay, e^{-jkT_d}, and adding the classic regularization term in, Equation 2.72 becomes,

$$H(k) = \frac{e^{-jkT_d}C^*(k)}{|C(k)|^2 + \beta|B(k)|^2}. \tag{4.3}$$

In comparing Equations 4.2 and 4.3 it can be observed that one cannot choose a regularization term $B(k)$ that will duplicate the effect of any choice of the target function $A(k)$. This is due to the regularization term being squared and being a real-valued magnitude. But the other way around is possible, meaning that one would be able to choose an $A(k)$ that would be able to duplicate the same behaviour as $B(k)$. By equating the two Equations, 4.2 and 4.3 and solving for $A(k)$, one gets

$$A(k) = \frac{e^{-jkT_d}}{1 + \beta\frac{|B(k)|^2}{|C(k)|^2}} \tag{4.4}$$

or

$$A_{eq}(k) = \frac{1}{1 + \beta\frac{|B(k)|^2}{|C(k)|^2}} \tag{4.5}$$

if the delay in the numerator is removed. Equation 4.4 gives us the target function which would be needed to duplicate the effect of the regularization. The target function $A_{eq(k)}$ as defined by Equation 4.5 is a real function with zero-phase. This implies that the time-domain response of the target-function will be even and will be symmetric about its maximum value. Having a symmetric time-response is not always desirable and could cause audible artifacts, as described in Section 4.1. By using the existing form of the equivalent target function, $A_{eq}(k)$, as the magnitude, an expression for a new equivalent target-function $\hat{A}_{eq}(k)$, can be given by,

$$\hat{A}_{eq}(k) = A_{eq}(k)e^{-\hat{\phi}(k)}, \tag{4.6}$$

where we have defined a new phase function $\hat{\phi}(k)$. One of the causes of audible artifacts is the pre-response that is generated from the zero-phase regularization. It is therefore advantageous to minimize the pre-response that is generated. One such way is to define the new target function to be minimum phase, as

$$\phi_{min}(k) = -\text{imag}\left[\text{ Hilbert}\{\ln|A_{eq}(k)|\}\right], \tag{4.7}$$

from the relationship of the Hilbert transform, the magnitude and the minimum phase of the function. Therefore if one sets $\hat{\phi}(k) = \phi_{min}(k)$, and uses the target function,

$$\hat{A}_{eq}(k) = A_{eq}(k)e^{-\phi_{min}(k)}, \tag{4.8}$$

and substitute it into Equation 4.2 to get the inverse filter,

$$H(k) = \frac{\hat{A}_{eq}(k)e^{-jkT_d}C^*(k)}{|C(k)|^2}. \tag{4.9}$$

It should be noted that this ensures that the target-function is minimum phase, but not the inverse filter $H(k)$ itself. Therefore, the corrected impulse response, $h(n)$ convolved with $c(n)$, will aim to be minimum-phase with a pure delay.

By choosing a target-function that is minimum phase, one is attempting to remove all the pre-response created from the "classic" regularization case. By doing so, one is adding energy to the post-response of the signal. Adding too much post-response might cause audible artifacts, therefore one could mix the minimum- and zero-phase functions together with a mixing parameter α and define the phase for the equivalent target function as,

$$\hat{\phi}(k) = \alpha\phi_{min}(k) + (1-\alpha)\phi_{zero}(k). \tag{4.10}$$

It is also known that time-domain masking is frequency dependent [21] [46], there by making the mixing parameter, α, frequency-dependent, one can vary the pre/post-response of the regularization according the frequency-dependence of time-domain masking. The resulting mixing of the phases would then take the form of,

$$\hat{\phi}(k) = \alpha(k)\phi_{min}(k) + (1-\alpha(k))\phi_{zero}(k). \tag{4.11}$$

By operating on the phase of the target-function, one can directly relate it to time-domain masking characteristics of the human ear. The target-function is what one desires or tries to achieve with the use of the inverse filter, therefore it seemed appropriate to operate

on this function. In the next section for multichannel systems, operating on the target-function cannot be done directly, but rather one needs to operate on the regularization functions of the inverse filtering calculation. Note that a time-domain implementation of the proposed method would be possible. We delay the introduction of the time-domain formulation to the next sub-section, which deals with the more general multichannel case.

4.2.2 Multichannel

In the previous section, Section 4.2.1, a method of reducing the pre-response caused by "classic" regularization was introduced for the mono-channel case, in the frequency-domain. In that method, an equivalent target-function was calculated based on the regularization weighting function. In doing so, one could then modify the phase of the new equivalent target-function, and customize it based on psychoacoustic characteristics of the human ear. In the multichannel case, the concept of creating an equivalent based on the regularization cannot be done directly. One must operate on the regularization functions alone.

If we start at the frequency-domain inverse method given by Equation 2.112, eliminating the frequency variable k for clarity,

$$\mathbf{H}_i = \left[\mathbf{C}^H\mathbf{C} + \beta\mathbf{B}^H\mathbf{B}\right]^{-1}\mathbf{C}^H\mathbf{A}_i, \tag{4.12}$$

where the definition and dimensions of matrices are the same as before and are summarized in the following table:

I	Number of Sources
J	Number of Loudspeakers
K	Number of Receiver (microphone) positions
H_i	J×1 Matrix of inverse filters for the i^{th} source signal
C	K×J Matrix of transfer functions
B	1×J Matrix of regularization filters
A_i	K×1 Matrix of target functions for the i^{th} source signal

Table 4.1: Multichannel matrix definitions and dimensions

The proposed multichannel frequency-domain method with minimal phase regularization can be described by the following steps:

1. Compute an "ideal" non-regularized solution using Equation 4.12 with $\beta = 0$. If the inverse of $\mathbf{C}^H\mathbf{C}$ is too ill-conditioned, then some small amount of regularization maybe used.

2. Compute a zero-phase regularized solution using Equation 4.12 with $\beta \neq 0$.

3. Computer the ratio $\mathbf{R}_i(k)$ between the ideal filter and the zero-phase regularized filter, for each frequency k and each filter i ($1 \leq i \leq J$). This ratio is the regularization term. Note that this ratio is well conditioned ($0 \leq |\mathbf{R}_i(k)| \leq 1$).

4. Make each regularization filter $\mathbf{R}_i(k)$ minimal phase to produce $\mathbf{R}_i^{\phi min}(k)$.

5. Compute the minimal phase regularized inverse filter as,

$$\mathbf{H}_i^{\phi min}(k) = \mathbf{R}_i^{\phi min}(k) \cdot \mathbf{H}_i^{\phi 0}(k) \tag{4.13}$$

A time-domain version of the multichannel minimum-phase regularization method is also possible. Time-domain methods are more costly as they involve inverting larger matrices, however they are not subject to circular convolution or wrap-around effects which can lead to pre-echo effects. For short filter lengths, inverse filters computed from time-domain methods usually perform better.

The time-domain version can be based from Equation 2.110,

$$\mathbf{H}_i^T = \left(\mathbf{C}^T\mathbf{C} + \beta\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{C}^T\mathbf{A}_i, \tag{4.14}$$

where $\mathbf{H}_i$,$\mathbf{C}$,$\mathbf{B}$ and $\mathbf{A}_i$ are of size $1\times JN_h$, $KN_a\times JN_h$, $N_a\times JN_h$ and $KN_a\times 1$ respectively.

The steps are then:

1. An ideal solution is computed from Equation 4.14 with $\beta = 0$, i.e. no regularization.

2. A solution with zero-phase regularization (classic regularization) is computed from Equation 4.14 with $\beta \neq 0$. For each of the J inverse filters $\mathbf{H}_i$ the next steps can be performed in the frequency-domain.

3. For each frequency, compute the ratios $\mathbf{R}_i(k)$ between the ideal solution and the zero-phase regularization solution. This ratio is the regularization term in the frequency-domain.

4. Make each regularization filter $\mathbf{R}_i(k)$ minimum-phase to produce $\mathbf{R}_i^{\phi min}(k)$.

5. Compute the minimum-phase regularized inverse filter as,

$$\mathbf{H}_i^{\phi_{min}}(k) \;=\; \mathbf{R}_i^{\phi_{min}}(k) \cdot \mathbf{H}_i^{\phi_0}(k), \tag{4.15}$$

   and convert back to the time-domain.

While it could be argued that this is a frequency-domain procedure just like the previous one, and as such it will suffer from the same wrap-around and pre-echo effects, this method behaves actually differently. The reason is that the deconvolution which requires regularization is done in the time-domain, reducing the pre-echo effects. The computation of $\mathbf{R}_i$ in the frequency domain can also be seen as a deconvolution, however it is a very well conditioned one, i.e. $|\mathbf{R}_i| < 1$ with no peak in the resulting frequency response. Therefore this part will not suffer from severe wrap-around or pre-echo effects in the time response.

4.3 Chapter Summary

This chapter provided a short overview of some psychoacoustic principles like non-simultaneous masking. In addition a new method of regularization was introduced and developed for the single- and multichannel cases. This new regularization method removes the pre-response created by the "classic" regularization that has been previously used in the literature.

Chapter 5

Evaluation of Minimum-Phase Regularization Techniques

In this chapter the minimum-phase techniques introduced and developed in Chapter 4 are evaluated. For the single-channel case, the inverse filter of a listening impulse response is calculated using the minimum-phase regularization method. It is compared to the "classic" inverse case as well as the ideal inverse case. Informal listening tests are done as a quick subjective evaluation as well as some objective measures are compared. These objective measure are PEAQ scores, as well as the energy before and after the main peak. The later measure is useful in showing how the signal is moved to after the main peak when the minimum-phase regularization technique is used.

In the multichannel case, a simple cross-talk canceler setup is used to evaluate the new method, and the results are shown.

5.1 Examples - Minimum Phase Target Function/Regularization

5.1.1 Single-Channel Minimum-phase Target Function

To see the effects of using a minimum-phase target function as defined by Equation 4.5, that incorporates the effects of the defined regularization term, some example will be presented. The impulse response that is used is shown in Figure 5.1 and was measured in the ITU standard [20] listening room at the Communications Research Centre (CRC). The impulse response was measured using a swept sine stimulus from the CRC-MARS

measurement software [6] with a sample rate of 44.1 kHz. The length of the IR used is 32768 samples (~743 ms). A longer room response was chosen over any of the anechoic impulse responses used previously in Chapter 3, since the room response would be less minimum-phase [31]. Using a longer response that is less minimum-phase should force the inverse filter to have more pre-echo effects from regularization and hence provide a good test with the new proposed method of using an equivalent target-function that is minimum-phase.

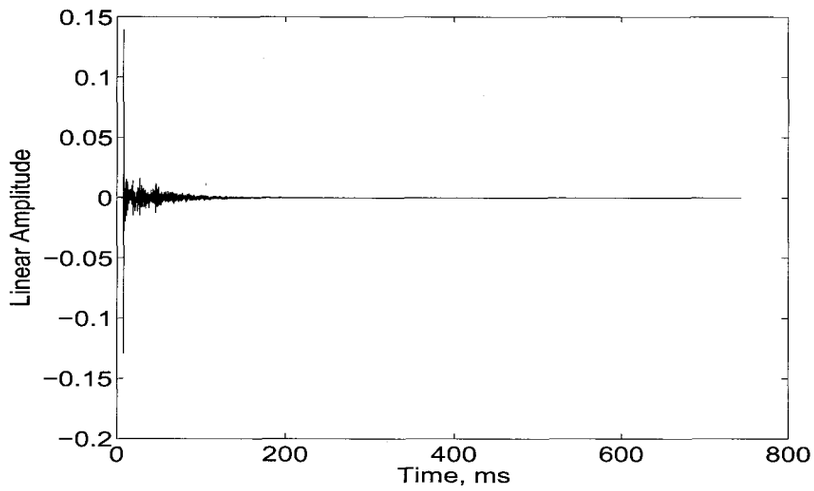


Figure 5.1: Impulse response of the ITU listening room at CRC used for the examples presented.

The magnitude of the frequency-response of the CRC listening room is shown in Figure 5.2. There is significant amplitude variation across frequency, therefore there is something significant to correct and make uniform in the frequency-response by means of an inverse filter. The magnitude of the frequency-response gives an indication to the frequency range of the regularization. Trying to correct the response outside the frequency range of the loudspeaker is virtually pointless since the loudspeaker cannot produce much energy in that range.

A frequency-dependent regularization response, along with the magnitude response of the CRC listening room are shown in Figure 5.3. The regularization response in Figure 5.3 shows cut-off frequencies around 40 Hz and 17000 Hz, which correspond to the cut-off frequencies of the frequency-response. In the pass-band region, 40 Hz to 17000 Hz, a regularization based on the $\frac{1}{3}$-octave spectrum of the magnitude response is

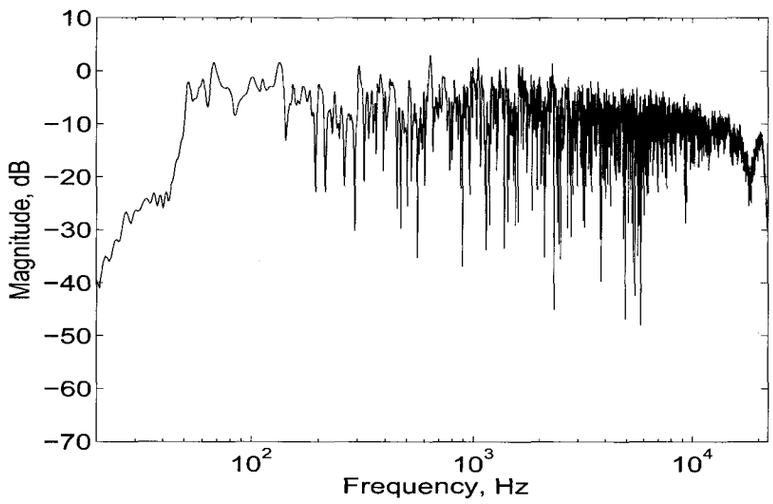


Figure 5.2: Magnitude of the frequency-response of the ITU listening room at the CRC used for the examples.

used. This is based on the regularization scheme used by Fielder, in [9]. The purpose of the $\frac{1}{3}$-octave spectrum weighting in the pass band is to provide a frequency dependent regularization weighting that will limit the effort that the inverse filter will do.

The regularization weighting shown in Figure 5.3 includes the exact inner ($\beta = 0.3$) and outbound ($\beta = 10$) weighting.

In all cases the length of the inverse filter is 65536 samples. First the time-domain method was used to calculate the inverse filter using Equation 2.39, where a delayed delta-function was used as the desired response. The resulting inverse filter is shown in Figure 5.5 a). Using the regularization response shown in Figure 5.3, from Equation 2.69 the inverse filter shown in Figure 5.5 b) was calculated. To test the minimum-phase target function approach, Equation 4.5 was used to create an equivalent target function, which was then made minimum-phase and then Equation 4.9 was used to calculate the inverse filter. The resulting inverse filter is shown in Figure 5.5 c). A time-domain window, shown in Figure 5.4 was applied to the inverse filters to reduce any edge effects. The top plot, Figure 5.5 a) does appear different from the other two in that it has a low frequency ripple before the main peak. The other two inverse filters are very much alike, though the minimum-phase target one, Figure 5.5 c), appears to have slightly less of a pre-response.

By convolving the individual inverse filters with the original impulse response c, a

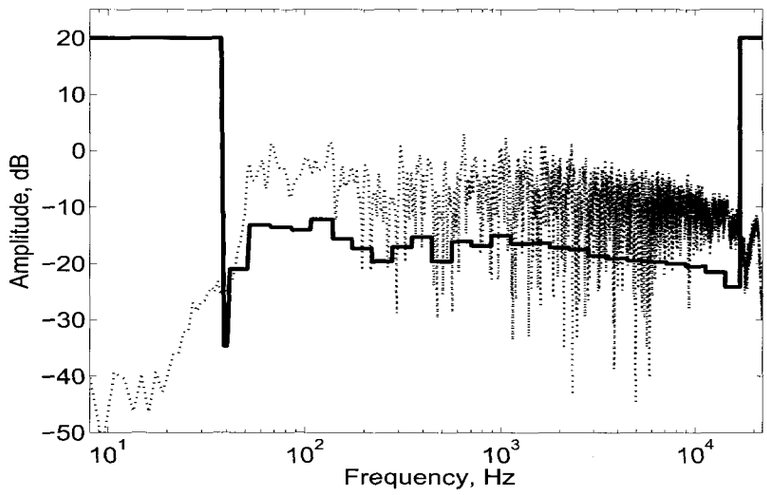


Figure 5.3: Regularization filter used in the evaluation.

corrected impulse response which ideally should be the target function used, is calculated. Figure 5.6 shows all three corrected impulse responses in the same order as the previous figure, with the ideal inverse on top (a)), the classic regularization in the middle (b)) and the minimum-phase target function one at the bottom (c)).

There are large differences between them all. The ideal inverse has a fairly uniform error spread throughout time, where the ends are tapered due to the effect of the time window on the inverse filter. The classic regularization case clearly illustrates the effect of the regularization term on the corrected response, it be symmetric around the main peak. This pre-response is clearly audible as a ramp up of the signal, similar to a reverse reverberant tail. In contrast, the minimum-phase target function case has very little pre-response. One issue with the minimum-phase case is the energy in the pre-response has been moved to the post-response. Even though from the time-domain (non-simultaneous) masking models, forward masking is greater than backwards masking, there will be a point where it will become audible. This may appear as a slight timbre change and to the extreme as extra reverberation in the sound. There will be a trade-off with respect to audibility, in that which is more detrimental or bothersome to the listener, the addition of pre- or post-response to the corrected sound.

Figure 5.7 shows the magnitudes of the corrected frequency-responses for the 3 different inverse filtering methods along with the original uncorrected response. The curves have been offset for clarity, therefore only a comparison between methods is valid. The

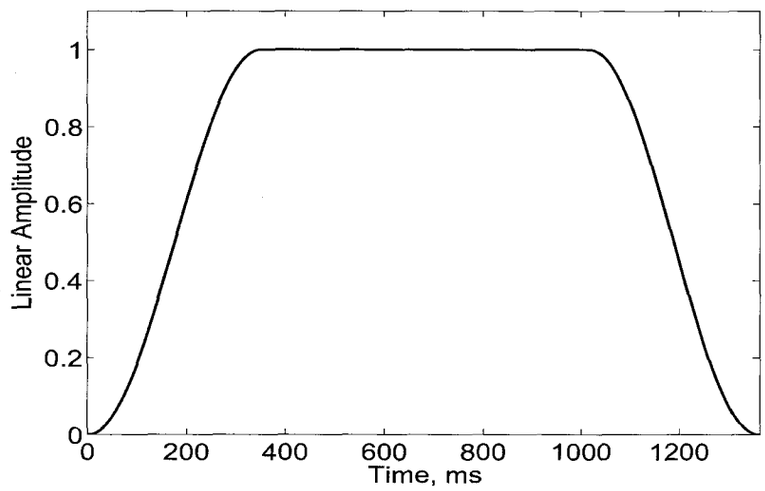


Figure 5.4: Time-domain window applied to inverse filters to reduce edge effects.

ideal inverse filter, shown in Figure 5.7 a), attempts to correct the entire frequency and is successful except for a few areas. The two other corrected impulse responses that used regularization do not correct the entire frequency range as expected. The regularization is doing its job with respect to the low and high cut-off frequencies. In the pass-band region between the cut-off frequencies, the two inverse filters correct most of the range. Where there are larger variations, this coincides with larger variations of the original response. The regularization that was used in this region was based on the $\frac{1}{3}$-octave spectrum of the original IR, therefore the amount that the inverse filter will correct is related to how the original spectrum is behaving in that frequency-range. It should also be noted that the spectrums of the two regularization cases are identical.

Another method of looking at the corrected impulse responses is by a spectrogram, which calculates the FFT of the corrected impulse response for specific times and window sizes. This results in a combination of time and frequency information that is useful to look at especially when one is concerned with audibility of artifacts in time which may be frequency-dependent. The data is plotted as frequency versus time, where the amplitude is represented by intensity, with brighter being a larger value. Figure 5.8 shows the spectrogram for corrected impulse response with classic regularization, where frequency is displayed on the horizontal axis and time on the vertical axis and the lighter the area, the larger the amplitude. The main peak is represented by the horizontal line at around 0.75 seconds, anything below that is pre-response and anything after is post-response.

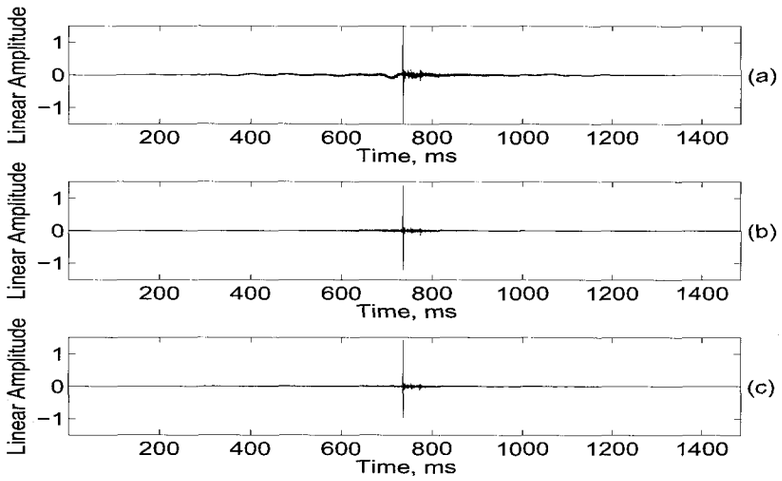


Figure 5.5: Inverse filters calculated from time-domain method, for CRC listening room: a) Ideal inverse, b) classic regularization, c) minimum-phase equivalent target function.

In this case, most of the energy is symmetrical around the main peak as was observed in the Figure 5.6, but now one can distinguish in time, the frequency content of specific artifacts. The low frequencies tend to ring on longer than the high frequencies. This is characteristic of the acoustics in the room. Therefore, in the low frequency area, the corrected impulse response is spread over a longer time frame. One unique feature is the very high frequency components that seem to be spread out over a long period of time.

Figure 5.8 shows the spectrogram of the corrected impulse response for the case with the minimum-phase target function. It is quite apparent that the majority of the pre-response has been reduced except for some at the low frequencies. Again, the inverse filter is not trying to correct there, so there should not be very much of a difference between the classic regularization case and the minimum-phase target function case. There does seem to be an increase of post-response energy though, which may start to become audible.

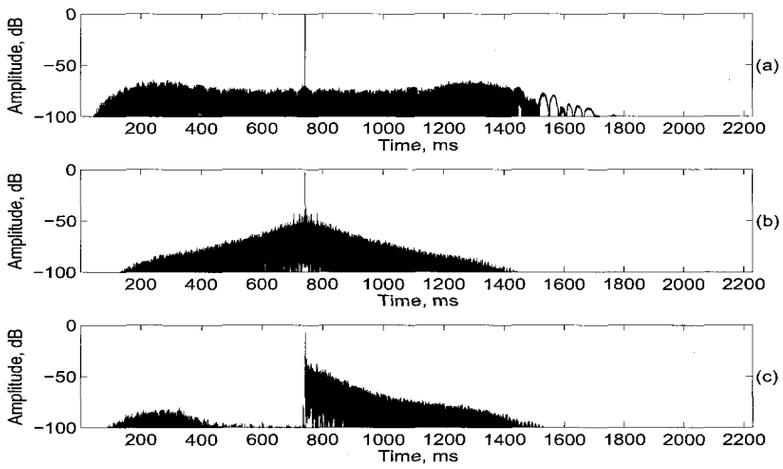


Figure 5.6: Corrected impulse responses for CRC listening room: a) ideal inverse, b) classic regularization and c) minimum-phase equivalent target function.

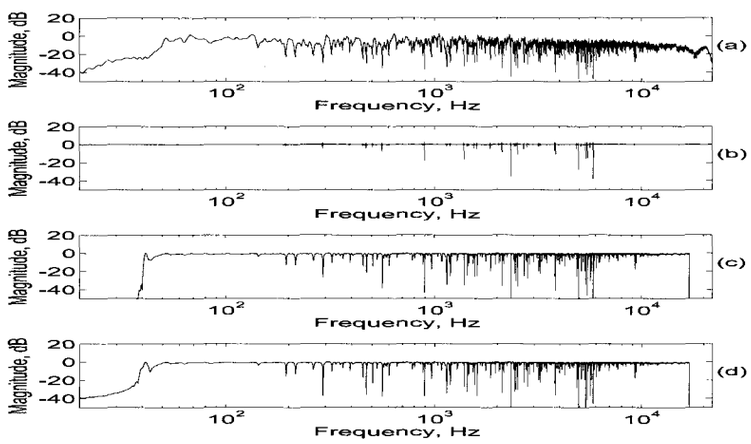


Figure 5.7: Magnitude of the frequency-responses of a) uncorrected loudspeaker and corrected with b) Ideal inverse, c) classic regularization, d) minimum-phase equivalent target function.

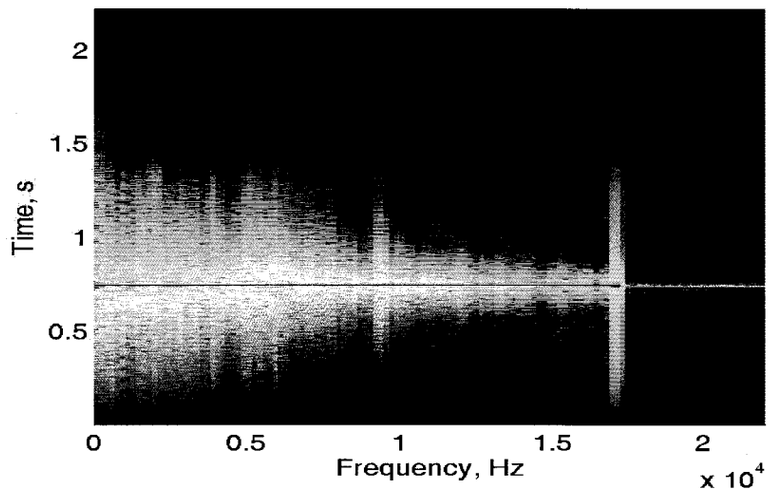


Figure 5.8: Spectrogram of the corrected impulse response using time-domain classic regularization method.

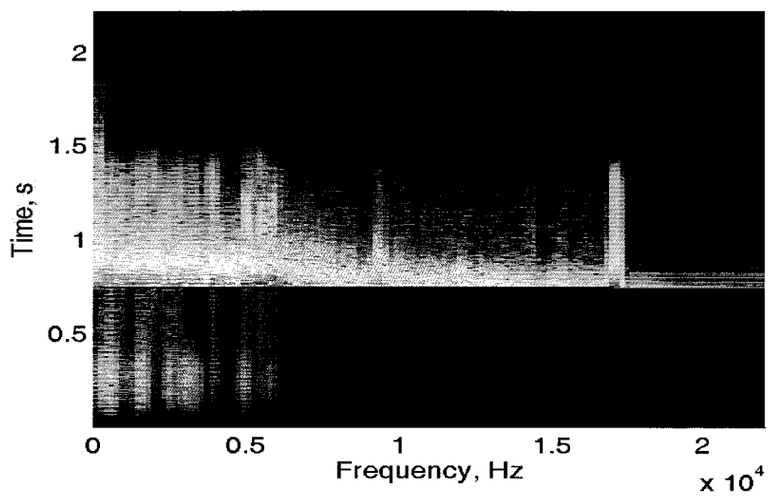


Figure 5.9: Spectrogram of the corrected impulse response using time-domain minimum-phase target function technique.

In the previous example, the inverse filter was calculated via the time-domain least squares method described by Equation 2.39 and it showed how the minimum-phase target function helps in reducing the pre-response created by the regularization. In Chapter 3, it was highlighted that the frequency-domain deconvolution method suffered more from time-aliasing, thus needed regularization, therefore using the minimum-phase target method should be beneficial. The frequency-domain deconvolution method given by Equation 2.50 and with classic regularization 2.72 were tested with the same IR and regularization parameters as with the time-domain example presented above. Figure 5.10 shows the inverse filters for: a) the ideal inverse, b) using classic regularization and c) using the minimum-phase target function approach. Similar to the time-domain method, the ideal inverse filter is most different from the other two, in that there is more low frequency ripple throughout it. The two other inverse filters are more compact with respect to time, and the minimum-phase target filter does appear to have less pre-response.

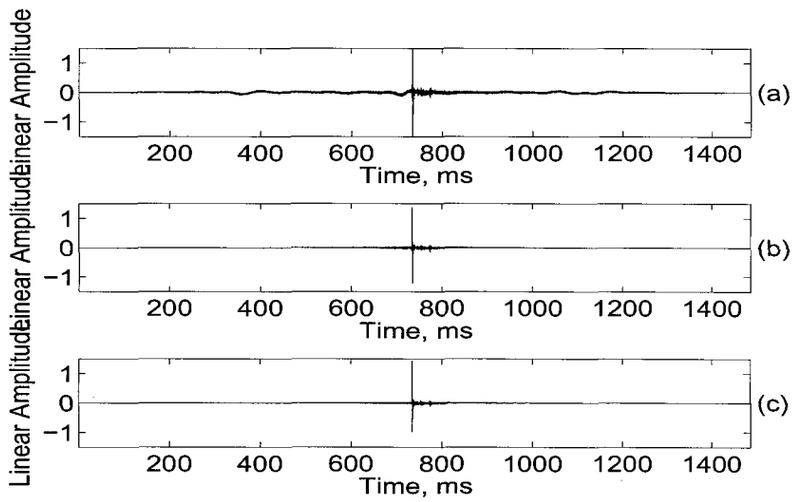


Figure 5.10: Inverse filters calculated from frequency-domain method, for CRC listening room: a) ideal inverse, b) classic regularization, c) minimum-phase equivalent target function.

Figure 5.11 shows the corrected impulse responses, again plotted with a dB amplitude, for the 3 inverse filters shown in Figure 5.10. The ideal inverse filter case shown in the top plot, a), has the characteristics of the frequency-domain deconvolution method highlighted in Chapter 3. There are large time-aliasing or wrap-around artifacts which have been reduced somewhat at the beginning and end by applying the time-window to

the inverse filter. The time-window rounds off the time-response at the extremes, though there are still large amounts of energy near the beginning and end peaking at a level of about -60 dB. This can be audible and will be perceived as distinct delays in the audio signal.

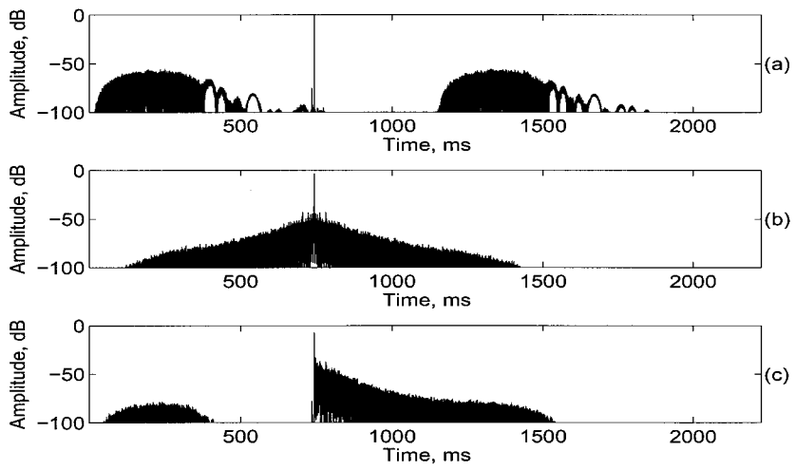


Figure 5.11: Corrected impulse responses for CRC listening room: a) ideal inverse, b) classic regularization, c) minimum-phase equivalent target function.

Right after the main peak there is a region of very low level caused by the circular deconvolution process. The classic regularization case, shown in Figure 5.11 b), resembles the time-domain results. In this case, the time-aliasing effects have been greatly reduced, though there is the expected symmetric energy around the main peak. The minimum-phase target function method, shown in c) has reduced the pre-response caused by the regularization. There is still a hump at the beginning, similar to what was left in the time-domain method, though the frequency-deconvolution method is at a larger amplitude. This could be time-aliasing artifacts that were hidden by the classic regularization artifacts.

Figure 5.12 shows the magnitude of the frequency-responses for all the corrected impulse responses along with the original response of the room. Once again the plots are offset for clarity. The ideal inverse case corrects as best as it can the entire frequency range as expected, and shows a similar result to the time-domain case. The classic regularization and minimum-phase target function cases show a similar response, just like the time-domain method case.

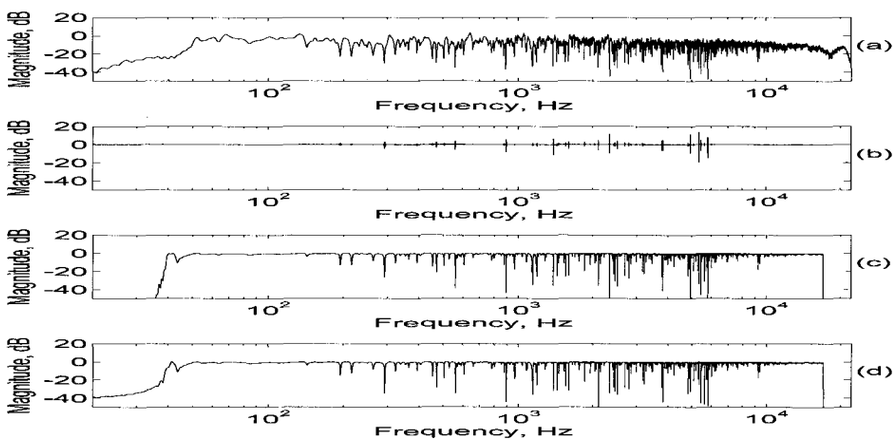


Figure 5.12: Magnitude of the frequency-responses of a) uncorrected loudspeaker and corrected with b) Ideal inverse, c) classic regularization, d) minimum-phase equivalent target function.

Once again the use of the spectrogram is used to view the corrected impulse responses in the time/frequency domain to get a better understanding of the shape of the response. Figure 5.13 shows the spectrogram for the classic regularization case which is shown in Figure 5.11 b). The energy is fairly symmetric around the main peak, which is the effects of the regularization.

Whereas Figure 5.14 shows the spectrogram when the minimum-phase target function is used. The pre-response has been reduced except for the lower frequencies, but that will be due to the fact that the inverse filter is not altering those frequencies very much.

The above examples calculated inverse filters that were just 2 times the length of the original impulse response. In previous studies [9], longer lengths of inverse filters were examined and showed improvement with respect to reducing artifacts. Therefore, inverse filters were calculated for length of 4 times or 131072 samples. By increasing the length of the inverse filter, the result should be a inverse filter with lower amplitude artifacts from time-aliasing.

To show the effects of a longer filter length on the minimum-phase target function method, the time-domain method was used, and the corrected impulse responses are shown in Figure 5.15, in the same order as previously displayed.

In all cases the amplitude of the error (from the ideal target-function) is lower in

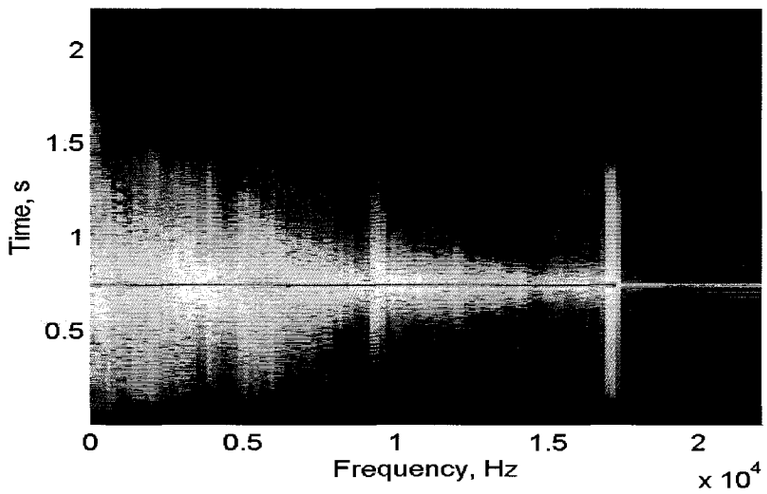


Figure 5.13: Spectrogram of the corrected impulse response using time-domain classic regularization method.

amplitude. With the ideal case shown in a), the level of the error has dropped about 10dB, and about 20dB in the minimum-phase target function method. With the later case, the artifact in the pre-response is close to -90 down and not audible.

Figures 5.16 and 5.17 show the spectrograms for the regularization and minimum-phase target function cases with the longer inverse filter length. The classic regularization case shows a symmetric spread of energy around the main peak whereas the minimum-phase target function shows the pre-response energy being reduced.

In informal listening tests where the same castanets audio sequence was used as the audio source just like in Chapter 3, the corrected impulse responses for the three methods, ideal, classic regularization and minimum-phase regularization, were convolved with the source audio clip. This provided a simulated corrected system, where a listener could compare the three methods. Time- and frequency-domain inversion methods were used. It was observed that the classic regularization method had a large pre-response that was clearly audible and that the minimum-phase method removed that pre-response. The ideal case had some lower level signal throughout the audio sequence, though not as audible as the classic regularization case, it was more audible than the minimum-phase case. One negative comment about the minimum-phase case was that it extended the tail, or the post-response of the audio signal. This was expected, and could be solved with a mix-phased regularization method where the zero-phase and minimum-

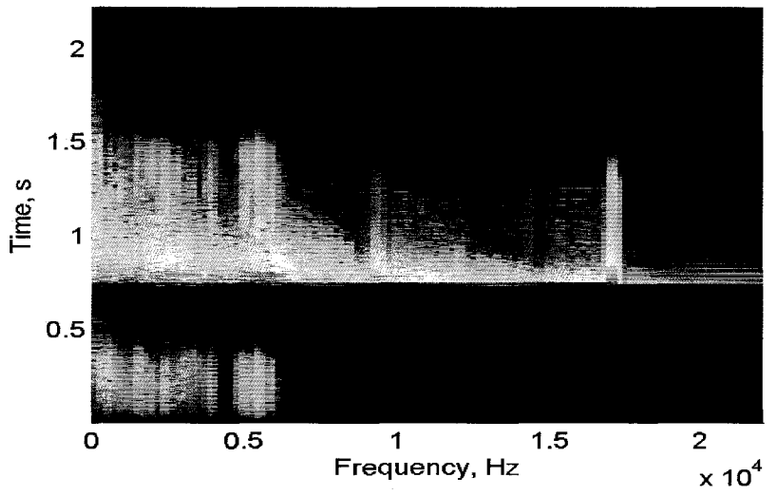


Figure 5.14: Spectrogram of the corrected impulse response using frequency-domain using minimum-phase target-function technique.

phase regularization responses are mixed. It was observed that the time-domain method produced less audible artifacts than the frequency-domain method with the same inverse filter length, similar to what was seen in Chapter 3.

To evaluate this new method objectively only PEAQ was used, as it had the best correlation with the formal subjective tests results over the other RMS error methods used in Chapter 3. Column 1 in Table 5.1 shows the ODG scores from the basic model of PEAQ. A score of 0 would indicate no audible difference. The PEAQ scores show a slight improvement when the minimum-phase regularization method is used over the classic regularization approach, but very close to the ideal case. Another objective metric to illustrate what the minimum-phase method is doing in comparison with the classic regularization method, is a pre- and post-peak energy comparison. In this metric, the main peak is determined in the corrected impulse response (and taken to be 3 samples wide) and then the energy is computed before and after the peak. These values are given in Table 5.1 for the same three methods and each value is normalized with the total energy for that specific corrected impulse reseponse.

It can be seen that the minimum-phase regularization method reduces the energy by a large amount before the main peak in comparison to the classic regularization case. The energy before the main peak goes from -11.5 dB to -50.4 dB and this result agrees with what can be seen in Figure 5.6. In comparing with the ideal case, the minimum-phase

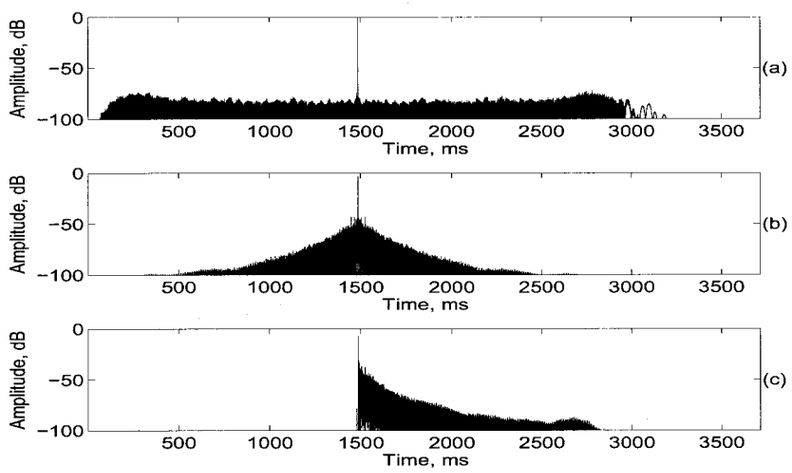


Figure 5.15: Corrected impulse responses for CRC listening room: a) ideal inverse, b) classic regularization and c) minimum-phase equivalent target function.

Method	PEAQ(basic) ODG	Energy Before (dB)	Energy After (dB)
Ideal	-0.841	-34.5	-34.2
Classic Regularization	-1.103	-11.5	-11.5
Minimum-phase Regularization	-0.872	-50.4	-3.9

Table 5.1: PEAQ scores and energy before and after main peak of corrected impulse response for the time-domain inverse with inverse filter twice the length of the original IR.

regularization method does appear to have less energy before the main peak, -50.4 dB compared to -34.5 dB.

Table 5.2 shows the before and after energies for the three methods, but when a frequency-domain inverse filtering method is used. Again the minimum-phase regularization method reduces the before energy in comparison with the classic regularization.

Table 5.3 gives the before and after energies for the time-domain inverse technique for the 3 different methods and again shows the effect that the minimum-phase regularization has on the before energy with an inverse filter four times the length of the original IR.

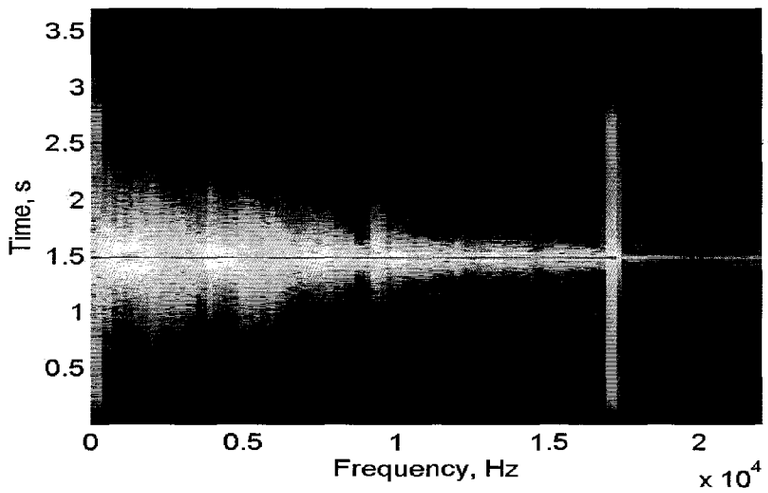


Figure 5.16: Spectrogram of the corrected impulse response using time-domain using classic regularization with length of inverse filter 4x that of IR.

Method	Energy Before (dB)	Energy After (dB)
Ideal	-26.6	-26.4
Classic Regularization	-11.5	-11.5
Minimum-phase Regularization	-47.7	-3.9

Table 5.2: Energy before and after main peak of corrected impulse response for the frequency-domain inverse with inverse filter twice the length of the original IR.

Method	Energy Before (dB)	Energy After (dB)
Ideal	-40.6	-40.2
Classic Regularization	-11.5	-11.5
Minimum-phase Regularization	-63.4	-3.9

Table 5.3: Energy before and after main peak of corrected impulse response for the time-domain inverse with inverse filter four times the length of the original IR.

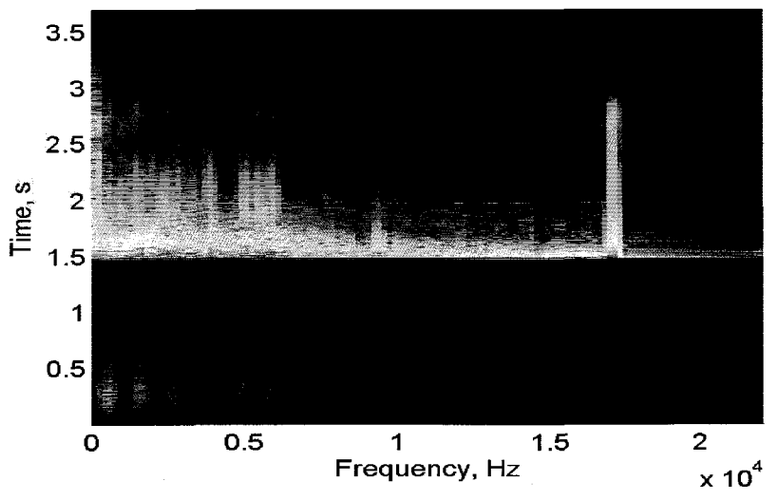


Figure 5.17: Spectrogram of the corrected impulse response using time-domain using minimum-phase target-function technique with length of inverse filter 4x that of IR.

5.1.2 Multichannel - Minimum Phase Regularization

To evaluate the multichannel minimum-phase regularization method, a system with dimensions $I = J = K = 2$ was used, and the impulse responses shown in Figure 5.18 were used. The impulse responses were measured in the CRC listening room, from two loudspeaker positions to a pair of spherical microphones that mimic the ear of a human head. The setup is similar to a crosstalk canceler system[41]. The impulse responses were measured using the CRC-MARS measurement software [6] and sampled at 48 kHz.

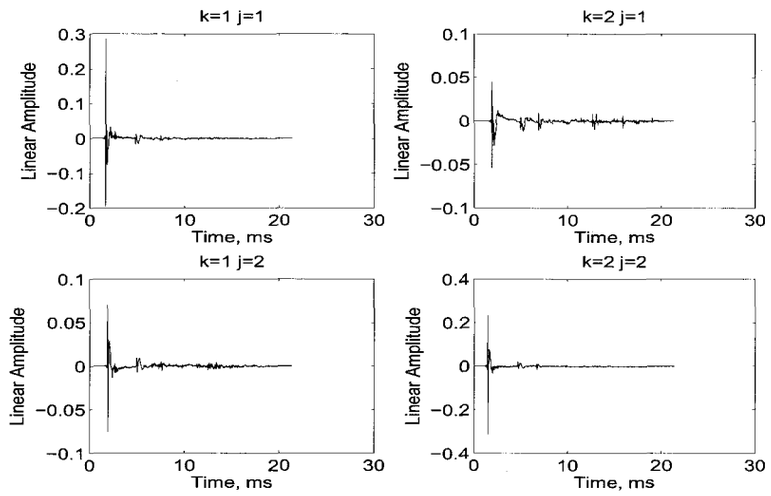


Figure 5.18: Multichannel impulse responses of loudspeakers and listening room at CRC used to evaluate the minimum-phase regularization technique.

Figure 5.19 shows the magnitude of the frequency-responses for all four impulse responses.

The regularization weighting that was used was that of a stop-band filter with the time-response and magnitude of the frequency response as shown in Figure 5.20.

This will generate four inverse filters which will correct the system such that it will cancel the output of the system on the non-diagonal terms, while reinforcing the diagonal terms. For example this will be observed in Figure 5.24, to be introduced later. Figure 5.21 shows the inverse filters calculated with the time-domain method, Equation 2.107. The inverse filters are not very compact and have a large amount of pre-response. When regularization is applied, Equation 2.110, the corresponding inverse filters become more compact as seen in Figure 5.22.

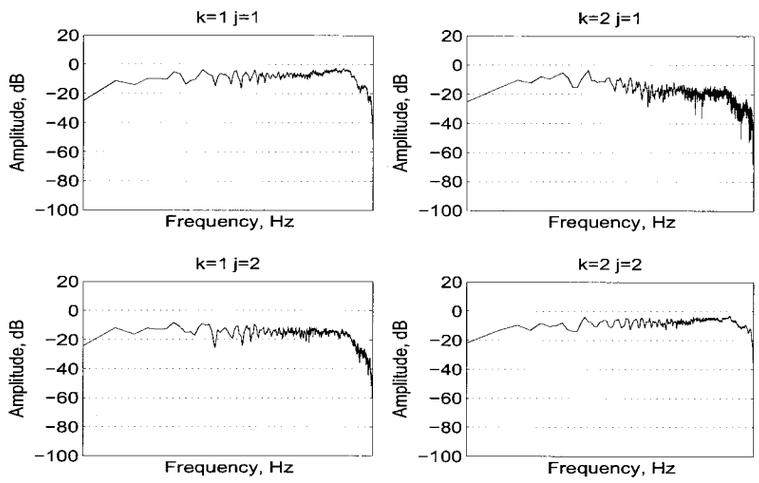


Figure 5.19: Magnitude of the frequency-responses of loudspeaker and listening room at CRC used to evaluate the minimum-phase regularization technique.

Figure 5.23 shows the inverse filters when a minimum-phase regularization method is used as outlined in section 4.2.2. The inverse filters do not look very different from the classic regularization case.

The corrected impulse responses for the ideal inverse case are shown in Figure 5.24. The diagonal responses, when $j = k$, show a sharp peak at time 0 ms, whereas the off-diagonal show a cancelation with the error around -60 dB down.

Figure 5.24 shows the corrected impulse responses with the classic regularization case. As with the mono channel case, the classic regularization adds a symmetric component to the corrected responses, even to the off-diagonal ones. The cancelation in the off-diagonal responses appears not good as in the ideal case, which is expected since the regularization is limiting the inverse filtering effort.

Figure 5.26 shows the corrected impulse responses when the minimum-phase regularization method, as described in Section 4.2.2, is used. The pre-response in all the responses is greatly reduced, in fact it appears lower than in the ideal case shown in Figure 5.24. The off-diagonal responses have a similar peak error as the classic regularization case, about -40 dB. The trade-off with this error and the regularization can be adjusted to what is needed and required.

For the multichannel scenario, objective measures are more difficult to derive due to the multiple paths.

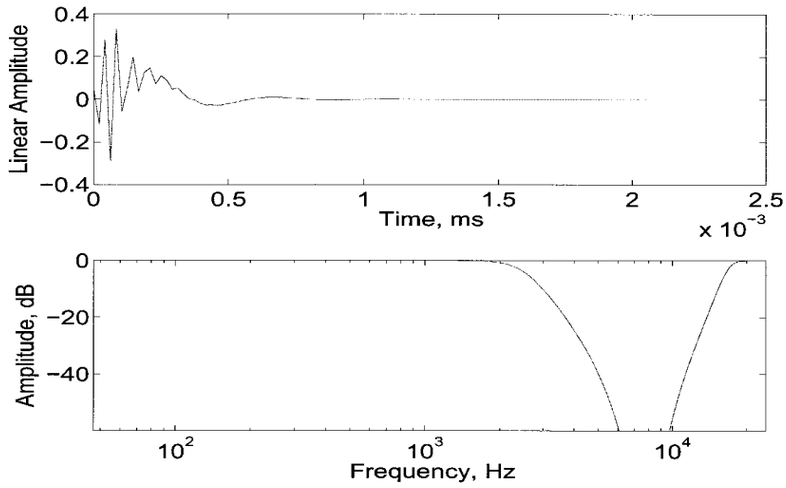


Figure 5.20: The regularization time-response (top) and magnitude of the frequency-response of the regularization filter used in the multichannel inverse filtering evaluation.

5.2 Chapter Summary

The results presented in this chapter have successfully demonstrated the expected behavior of the methods introduced in Chapter 4. The use of minimum-phase regularization does reduce the pre-response of the corrected impulse responses.

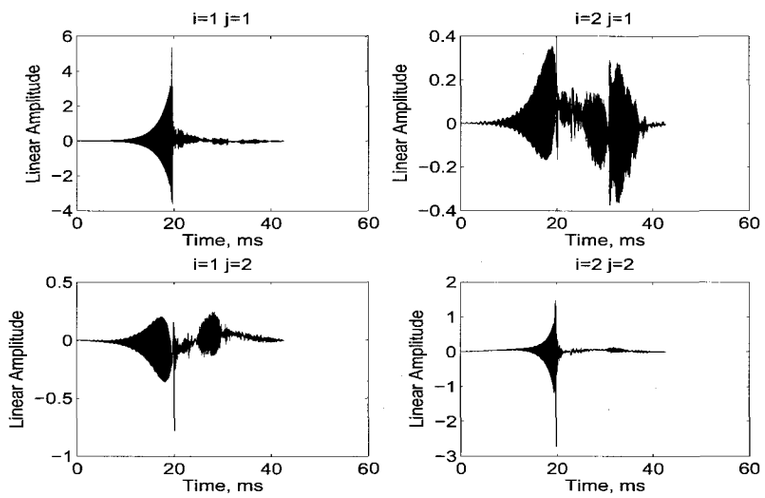


Figure 5.21: Multichannel inverse filters calculated in the the ideal case.

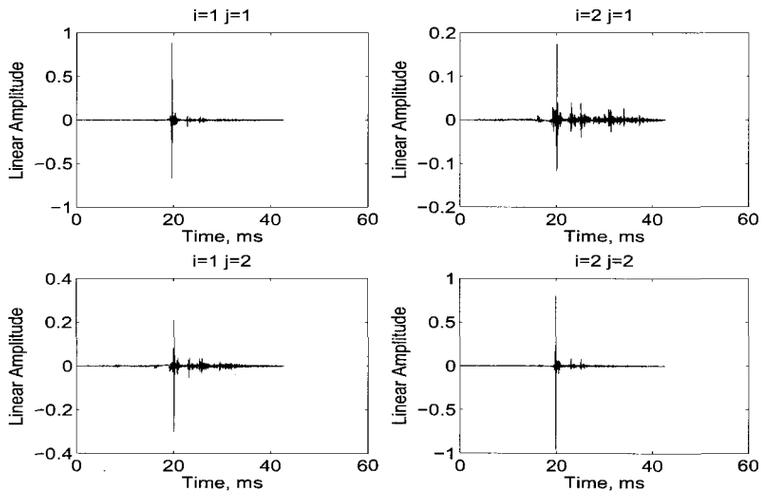


Figure 5.22: Multichannel inverse filters calculated with classic regularization.

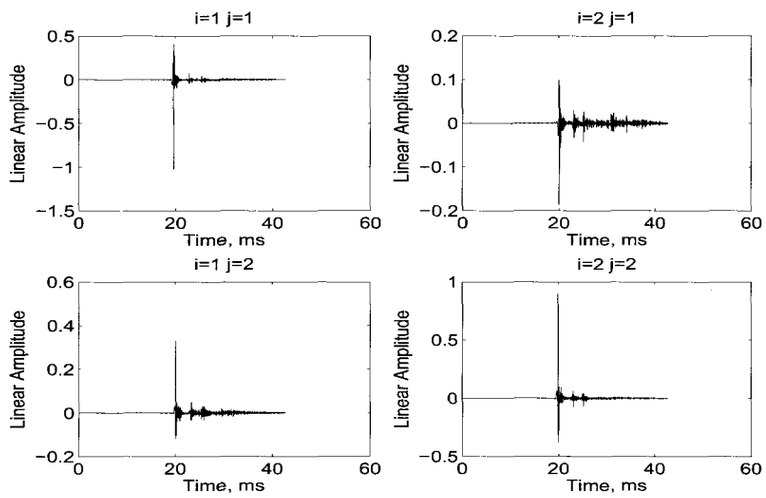


Figure 5.23: Multichannel inverse filters calculated with minimum-phase regularization.

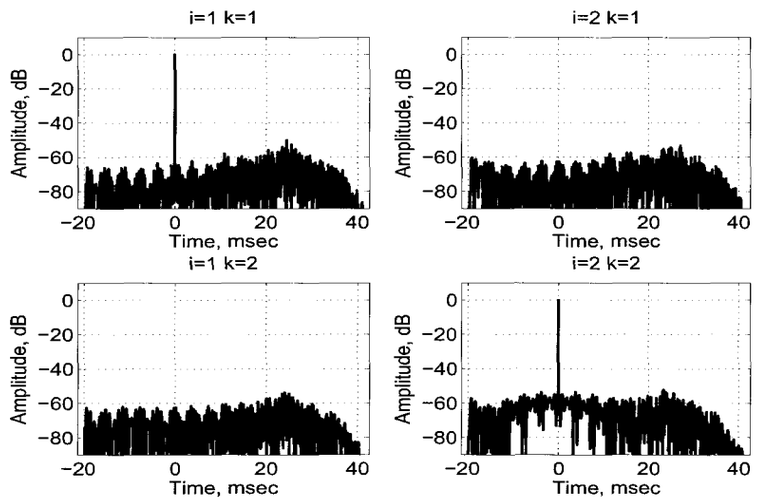


Figure 5.24: Corrected multichannel inverse filters calculated in the ideal case, i.e. no regularization.

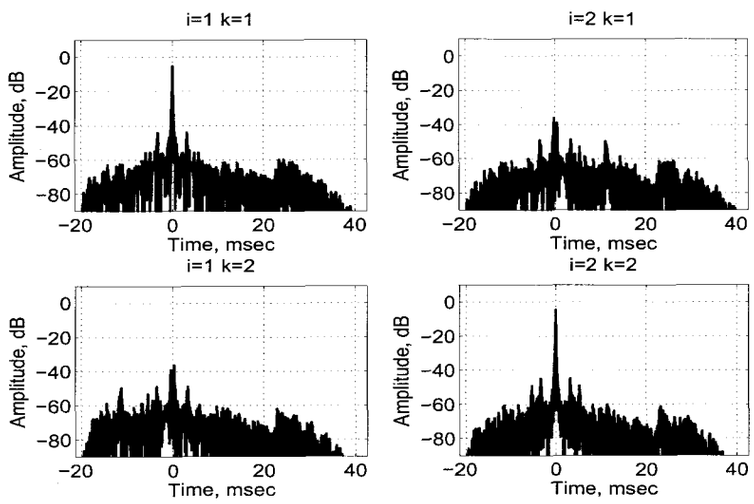


Figure 5.25: Corrected multichannel inverse filters calculated using classic regularization.

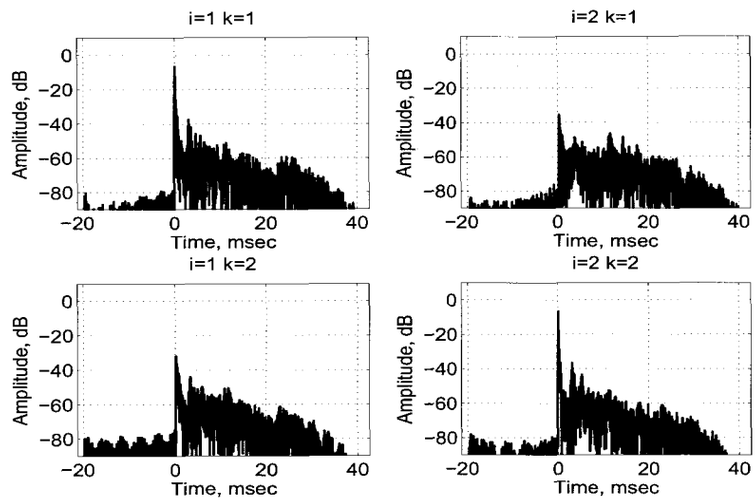


Figure 5.26: Corrected multichannel inverse filters calculated using classic minimum-phase regularization.

Chapter 6

Conclusions

6.1 Summary of Thesis

This thesis provided a review of audio equalization and with the advent of digital signal processing the concept of equalizing or correcting the magnitude and phase became possible. One of the first uses of inverse filtering was with cross-talk cancelation by Atal and Schroeder. A cross-talk canceler pre-filters the audio signal to two or more loudspeakers by inverse filtering to control the acoustic signal arriving at the two ears.

Chapter 2 introduced the inverse filtering concept in more detail and reviewed time- and frequency-domain methods for calculating the inverse filter of a system. In reviewing the various inverse filtering methods, it was shown that the stochastic Wiener methods can be expressed solely in term of deterministic terms if the system is deterministic. This was shown for the single-channel case as well as for the multichannel case. Regularization and smoothing were also discussed in this chapter and the single- and multichannel were derived with regularization for both time- and frequency-domain methods.

A subjective test method was developed in Chapter 3 that provides a way of evaluating inverse filtering methods. It allows the subject to listen and evaluate various inverse filtering methods and compare them to a reference signal. Formal subjective tests were carried out to evaluate previously published single-channel inverse filtering methods. Those included time-domain least-squares and frequency-domain deconvolution methods with and without regularization. The time-domain least-squares method proved to be more robust compared to frequency-domain deconvolution with the same inverse filter length. Regularization did help but created a pre-response which could degrade the audio quality. Complex smoothing the impulse response before it was inverted was also

evaluated and showed to improve the performance over the case where it was not used. Objective evaluation methods such as RMS errors, weighted RMS errors and PEAQ were investigated and correlated against the results of the subjective tests. The RMS errors did not correlate very well with the subjective results compared to the results of PEAQ. PEAQ did not always predict close outcomes but it could be used as a general tool for evaluation, though one must be careful since PEAQ was developed for evaluation of high-quality audio codecs. If it encounters artifacts that it is not familiar with, the results will not be accurate. Therefore is does not replace formal subjective testing in cases where it might encounter artifacts different from those it was trained for.

Chapter 4 provided a brief review of psychoacoustics, namely on non-simultaneous masking. A new method of minimum-phase regularization was then introduced and developed for single and multichannel cases. With this method, the pre-response of the regularization can be avoided, or rather moved to after the main peak of the corrected impulse response. This is preferred due to the masking characteristics of the human ear where forward masking is far greater than backward masking. The minimum-phase regularization method can be implemented in the time-domain least-squares method or in the frequency-domain deconvolution method.

Chapter 5 implements the new minimum-phase regularization method for the single- and multichannel cases. In the single-channel case an impulse response from a listening room at the CRC is used as the system. In informal listening tests, the minimum-phase regularization approach is audibly superior to the "classic" regularization approach, and to the ideal inverse. When evaluating the method with PEAQ, the minimum-phase regularization method performs slightly better than the "classic" regularization method, -0.872 compared to -1.103 when the time-domain least-squares method was used. In terms of comparing the effect of the minimum-phase regularization method, looking at the energy before and after the main peak would be a better comparison. In this case the energy before the main peak dropped from -11.5 dB with the "classic" regularization case to -50.4 dB with the minimum-phase regularization method. The frequency-domain method showed similar results. Finally, a cross-talk canceler was implemented to evaluate the multichannel minimum-phase regularization method. Again, plots of the corrected impulse responses showed a reduction in the pre-response when the minimum-phase regularization method was used.

6.2 Future Work

6.2.1 Signal-dependent Inverse Filtering

In the previous derivations of inverse filtering, some of the final expressions, for example Equations 2.39 and 2.69 have been signal independent, formulations not involving x, d or v, while others such as Equation 2.36 express the inverse filter in terms of the signal x, d, v. As was presented in [33], some simple strategies for adaptive, or rather signal-dependent inverse filtering will be shown here.

Figure 6.1 shows a diagram of a general strategy of an adaptive inverse filtering scheme where the input audio signal, $x(n)$ along with knowledge or model ($\hat{C}(\omega)$), of the impulse response/transfer function $C(\omega)$ of the system to be corrected, can be used to help construct the inverse filter $H(\omega)$.

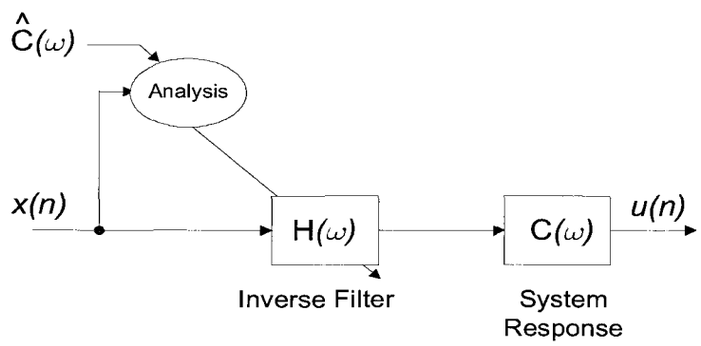


Figure 6.1: Diagram of the adaptive strategy for adapting the inverse filter $H(\omega)$ from the incoming audio signal.

The key to this approach is to develop a successful strategy for designing the inverse filter on a signal-dependent basis. The spectrum or masking threshold of the incoming signal could be one of the characteristics that could be used in the design of the inverse filter. The inverse filter could be modified with each input audio block or it could be selected from a table of available correction filters. If the block contains very little energy, it could be decided that no inverse filter should be applied at all.

The frequency-domain inverse filter, as described above, is desirable to use in an adaptive manner since it is computationally efficient and can be easily implemented. It lends itself well to a block processing structure where the input is partitioned into blocks to be analyzed. Figure 6.2 shows the inverse filtering structure with a weighting term

$A_w(\omega)$ on the error plus a regularization term $A_R(\omega)$. $C(\omega)$ is the transfer function that is to be inverted and $H(\omega)$ is the inverse filter of $C(\omega)$. $A(\omega)$ is the desired function when $C(\omega)$ is multiplied by $H(\omega)$. $x(n)$ is the input signal, with $d(n)$ and $e_A(n)$ being the desired signal and weighted-error signal respectively.

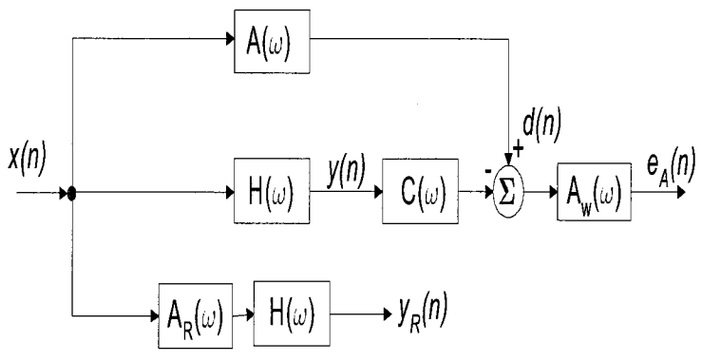


Figure 6.2: Diagram for inverse filtering with a weighting $A_w(\omega)$ on the error term and a regularization term $A_r(\omega)$

By using the commutative property of linear systems, the weighting term on the error, $A_w(\omega)$, can be moved to the left of the diagram as shown in Figure 6.3. This new

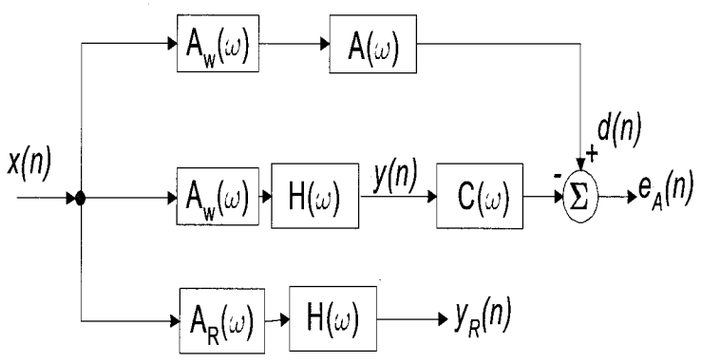


Figure 6.3: Modified structure of the one shown in Figure 6.2 where the weighting on the error term, $A_w(\omega)$, has been moved to the beginning of the structure.

form simplifies the analysis and the cost function, $J(\omega)$, in the frequency-domain, for this structure is given by

$$J(\omega) = \frac{1}{2}E\left[E_A(\omega)E_A^*(\omega)\right] + \beta E\left[\frac{1}{2}Y_R(\omega)Y_R^*(\omega)\right], \tag{6.1}$$

where the error term $E_A(\omega)$ is given by

$$E_A(\omega) = X(\omega)A_w(\omega)A(\omega) - X(\omega)C(\omega)H(\omega)A_w(\omega), \tag{6.2}$$

for each tranformed frame of data $x(n)$.

The regularization term $Y_R(\omega)$ is given by

$$Y_R(\omega) = X(\omega)A_r(\omega)H(\omega), \tag{6.3}$$

and β is a constant to control the amount of regularization. Taking the derivative of the cost function $J(\omega)$ with respect to $H(\omega)$ and setting it to zero, the solution for the inverse filter $H(\omega)$ can be found. The solution is

$$H(\omega) = \frac{A(\omega)C^*(\omega)}{|C(\omega)|^2 + \beta\frac{|A_R(\omega)|^2}{|A_w(\omega)|^2}} E\left[\frac{|X(\omega)|^2}{|X(\omega)|^2 + \alpha}\right], \tag{6.4}$$

where a constant α has been inserted in the denominator to account for when the source $X(\omega)$ is zero. It is interesting to note that the regularization weighting term, $A_R(\omega)$ and the error weighting, $A_w(\omega)$, appear in the same term. They are simply reciprocals of each other and therefore can be mathematically interchangeable.

Another way of incorporating a signal-dependent term is to make the regularization term, $A_R(\omega)$ dependent on the incoming audio signal, $x(n)$. One such approach is given by

$$A_R(\omega) = \left[\frac{1}{|C_s(\omega)|} E\left[\frac{1}{|X_s(\omega)|}\right]\right] \tag{6.5}$$

where $C_s(\omega)$ and $X_s(\omega)$ as smoothed or fractional-octave band spectra. With this approach when either of the spectra $C_s(\omega)$ or $X_s(\omega)$ get very small, the regularization will be increased thereby limiting the inversion of $C(\omega)$.

These two methods shown above represent only two approaches of signal-dependent inverse filtering. Both were evaluated using PEAQ [17] and the results varied with improvements in some cases [37]. Further investigations on these and other signal-dependent system inverse methods are required in order to fully assess their performance and potential.

6.2.2 Adaptive Filtering

The use of adaptive filtering in loudspeaker or room equalization is a more complex task than just calculating the inverse filter. For example, if a listener is in a room and moves

from one equalized position to another, there must be a means to relay that movement, such that a new correction filter can be calculated. To simplify the computation, a look up table of pre-calculated corrections filters could be undertaken.

6.2.3 Methods to Increase "Sweet-Spot"

In the inverse filtering examples illustrated in the thesis, a mono channel case and a multichannel ($2 \times 2 \times 2$) system, the correction was being done at one or two positions. This may or may not help the correction of the sound field at positions around those points. In Chapter 3, one of the inverse test evaluations was conducted using an off-angle impulse response to correct an on-axis impulse response. The results were not too successful. It is therefore an area to further explore with the use of a multichannel setup, either with more receiver positions than source positions, or using some sort of impulse response averaging, so that an inverse filter is calculated such that it will equalize an area as opposed to a single point.

Bibliography

[1] B. S. Atal and M. R. Schroeder. "Apparent Sound Source Translator". US Patent 3,236,949 (1966).

[2] G. M. Ballou, editor. *Handbook for Sound Engineers: The New Audio Cyclopedia.* Howard W. Sams & Co., 2nd edition, 1991.

[3] C. Bean and P. Craven. "Loudspeaker and Room Correction Using Digital Signal Processing" presented at the 86th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 50(7/8), 1989.

[4] D. Bohn. "Operator Adjustable Equalizers: An Overview". In *Proceedings of the AES 6th International Conference: Sound Reinforcement*, (1988 May).

[5] J. L. Bruning and B.L. Kintz. *Computational Handbook of Statistics.* Addison Wesley Longgman, Reading MA., 1997.

[6] Commnications Research Centre. CRC-MARS (multichannel audio research system. Software package for multichannel audio/acoustic measurements.

[7] Communications Research Centre. CRC-SEAQ (system for the evaluation of audio quality. Commercial software package for subjective and objective audio testing.

[8] P. G. Craven and M. A. Gerzon. "Practical Adaptive Room and Loudspeaker Equaliser for Hi-Fi Use". *J. of the Audio Eng. Soc. (Abstracts)*, 40(5), 1992.

[9] L. D. Fielder. "Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization". *J. Audio Eng. Soc.*, 51(1/2):3–26, 2003.

[10] R. P. Genereux. "Adaptive Loudspeaker Systems: Correcting for the Acoustic Environment" presented at the 8th International Conference of the Audio Engineering Society. In *Audio Eng. Soc.*, 1990.

[11] F. J. Harris. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform". *Proceedings of the IEEE*, 66(1):51–83, 1978.

[12] P. D. Hatziantoniou and J. N. Mourjopoulos. "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses". *J. Audio Eng. Soc.*, 48(9):259–280, 2000.

[13] P. D. Hatziantoniou and J. N. Mourjopoulos. "Errors in Real-Time Room Acoustics Devreverberation". *J. Audio Eng. Soc.*, 52(9):883–899, 2004.

[14] P. D. Hatziantoniou and J. N. Mourjopoulos. "Real-Time Room Equalization based on Complex Smoothing: Robustness Results" presented at the 116th Convention of the Audio Engineering Society. *J. Audio Eng. Soc. (Abstracts)*, 52(7/8), 2004.

[15] S. Haykin. *Adaptive Filter Theory.* Prentice Hall Inc., third edition, 1996.

[16] R.M. Howe and M.O.J. Hawksford. "Methods of Local Room Equalization and Their Effect Over the Listening Area" presented at the 91th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 50(7/8), 1991.

[17] *Method for Objective Measurement of Perceived Audio Quality.* ITU-R Recommendation BS.1387, Geneva.

[18] *Method for the Subjective Assessment of Intermediate Audio Quality.* ITU-R Recommendation BS.1534, Geneva.

[19] *Method for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems.* ITU-R Recommendation BS.1116, Geneva.

[20] *Method for the Subjective Assessment of Small Impariments in Audio Systems Including Multichannel Sounds Systems.* ITU-R Recommendation BS.1116, Geneva.

[21] W. Jesteadt, S.P. Bacon, and J.R. Lehman. "Forward Masking as a Function of Frequency, Masker Level and signal Delay". *J. Acoust. Soc. Am.*, 71:950–062, 1982.

[22] G. Keppel and S. Zedeck. *"Data Analysis for Research Designs".* W.H. Freeman and Company, New York, NY, 1989.

[23] O. Kirkeby and P. A. Nelson. "Digital Filter Design for Inversion Problems in Sound Reproduction". *J. Audio Eng. Soc.*, 47(7/8):583–595, 1999.

[24] O. Kirkeby, P.A. Nelson, and H. Hamada. "The "Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers". *J. Audio Eng. Soc.*, 46(5):387–395, 1998.

[25] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante. "Fast Deconvolution of Multichannel Systems Using Regularization". *IEEE Trans. Speech and Audio Proc.*, 6(2):189–195, 1998.

[26] S. M. Kuo and D. R. Morgan. "Active Noise Control: A Tutorial Review". *Proceedings of teh IEEE*, 87(6):941–973, June 1999.

[27] S.P. Lipsthiz, T. C. Scott, and J. Vanderkooy. "Increasing the Audio Measurement Capability of FFT Analyzers by Microcomputer Postprocessing". *J. Audio Eng. Soc.*, 33(9):626–648, 1985.

[28] B. C. J. Moore. *"Introduction to the Psychology of Hearing"*. Academic Press, London, 1997.

[29] B. C. J. Moore and B. R. Glasberg. "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excition Patterns". *J. Acoust. Soc. Am.*, 74(3):750–753, 1983.

[30] J. Mourjopoulos. "Digital Equalization Methods for Audio Systems". *J. of the Audio Eng. Soc. (Abstracts)*, 36(5), 1988.

[31] S.T. Neely and J. B. Allen. "Invertibility of a Room Impulse Response". *J. Acoust. Soc. Am.*, 66(1):165–169, 1979.

[32] P. A. Nelson and S. J. Elliot. *"Active Control Of Sound"*. London Academic Press, 1992.

[33] S. G. Norcross, M. Bouchard, and G. A. Soulodre. "Adaptive Strategies for Inverse Filtering", presented at the 119th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 53(12), 2005.

[34] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie. "Evaluation of Inverse Filtering Techniques for Room/Speaker Equalization" presented at the 113th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 50(11), 2002.

[35] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie. "Further Investigations of Inverse Filtering", presented at the 115th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 51(12), 2003.

[36] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie. "Subjective Effects of Regularization on Inverse Filtering" presented at the 114th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 51(10), 2003.

[37] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie. "Subjective Investigations of Inverse Filtering". *J. Audio Eng. Soc.*, 52(10):1003–1028, 2004.

[38] J. Panzer and L. Ferekidis. "The use of Continuous Phase for Interpolation, Smoothing and Forming Mean Values of Complex Frequency Response Curves", presented at the 116th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 52(7/8), 2004.

[39] John G. Proakis. *Digital Signal Processing, Algorithms, and Applications.* Prentice Hall, Upper Saddle River, N.J., 3rd edition, 1996.

[40] Douglas D. Rife and John Vanderkooy. Transfser-Function Measurement with Maximum-Length Sequences. *J. of the Audio Eng. Soc.*, 37(6), 1989.

[41] M. R. Schroeder. "Models of Hearing". *Proc. of the IEEEE*, 63(9):1332–1350, September 1975.

[42] G. A. Soulodre and M. C. Lavoie. "Subjective Evaluation of Large and Small Impairments in Audio Codecs" presented at the AES 17th Conference: High Quality Audio Coding. *J. of the Audio Eng. Soc. (Abstracts)*, 47(11), 1999.

[43] G. A. Soulodre and S. G. Norcross. "Objective Measures of Loudness" presented at the 115th Convention of the Audio Engineering Society. *J. of the Audio Eng. Soc. (Abstracts)*, 51(12), 2003.

[44] F. E. Toole and S. E. Olive. "The Modification of Timbre by Resonances: Perception and Measurement". *J. Audio Eng. Soc.*, 36(3):122–142, 1988.

[45] B. Truax. http://www.sfu.ca/sonic-studio/handbook/Phon.html (Sept 14, 2008).

[46] E. Zwicker and H. Fastl. *"Psychoacoustics: Facts and Models".* Springer-Verlag, 1990.