

Subjective Investigations of Inverse Filtering*

SCOTT G. NORCROSS, *AES Member*, GILBERT A. SOULODRE, *AES Fellow*, AND MICHEL C. LAVOIE

Communications Research Centre, Ottawa, Ont. K2H 8S2, Canada

Inverse filtering is the concept that one can “undo” the filtering caused by a system such as a loudspeaker or room. This approach strives to correct both the magnitude and the phase of the system. Inverse filtering has been proposed for numerous applications in audio and telecommunications, such as loud speaker equalization, virtual source creation, and room deconvolution. When inverting the impulse response (IR), undesired audible artifacts may be produced. The severity of these artifacts is affected by the characteristics of the IR of the system, and the method used to compute the inverse filter. When the IR is nonminimum phase, the artifacts tend to be more severe and become distinctly audible. The artifacts produced by the inverse-filtering process can actually degrade the overall signal quality rather than improve it. Formal subjective tests were conducted to investigate and highlight potential limitations associated with several inverse-filtering techniques. Time-domain and frequency-domain methods were implemented, along with several types of regularization and complex smoothing to help reduce the level of audible artifacts. The results of the subjective tests show that the various inverse-filtering techniques can sometimes improve the subjective quality and in other cases degrade the audio quality.

0 INTRODUCTION

Equalization techniques have long been used to correct loudspeaker and room responses so that a flat spectrum could be achieved for a desired listening area. Traditional techniques involve graphic or parametric equalizers that shape the spectrum of the signal using minimum-phase filters. These techniques have limitations due to the frequency resolution of the filters. Moreover, these techniques do not attempt to equalize the phase response. A more complete approach is to use deconvolution or inverse filtering. This approach is based on the concept that one can “undo” the filtering caused by the loudspeaker or room by convolving the measured impulse response (IR) with its inverse filter. This approach strives to correct both the magnitude and the phase of the system.

Different methods to calculate the inverse of a filter exist [1]–[4], but the authors are not aware of any formal study to subjectively evaluate their performance. Even though one method may create a more mathematically correct inverse filter, it may not be the best perceptually. It is possible for the inverse-filtering process to make the signal perceptually poorer than if the inverse filter were not applied. The authors investigated several inverse-filtering techniques that were included in a plug-in for a commercial audioediting package, and the results were not as anticipated. In trying to correct the response of several

commercially available loudspeakers, various distortions, time-domain artifacts, and other filtering effects were clearly audible. This prompted the present study of the perceptual aspects of inverse filtering and equalization of loudspeakers.

Loudspeaker and room IRs are typically nonminimum phase [6], and hence a true causal inverse may not exist. A modeling delay can be used in the inversion process to produce a causal inverse filter, though preringing in the inverse filter can result. This can lead to audible artifacts being produced. A way to avoid potential preringing is to decompose the IR its minimum-phase and all-pass components and invert only the minimum-phase component of the IR.

When inverting the IR of a loudspeaker, which has a typical rolloff at the low and high frequencies, the inverse filter will attempt to correct this rolloff. This will cause the inverse filter to have large boosts at these frequencies, which could overload the system and could also produce a long filter. Regularization can be used to limit the amount of effort the inverse filter will apply [2]–[4], thereby limiting the magnitude of the boost. An alternative approach is to smooth [7] the transfer function before calculating the inverse to reduce large dips and peaks in the frequency response so that the inverse filter does not have to work as hard to correct them.

Another concern with inverse filtering is that measuring the IR of a loudspeaker or a room results in one IR or transfer function between the loudspeaker and a measurement microphone located some distance away. Changes in

*Manuscript received 2003 December 1; revised 2004 July 28 and August 18.

the location of the microphone or loudspeaker in the room will result in a corresponding change in the IR due to the directional characteristics of the loudspeaker and/or the acoustical characteristics of the room. Therefore an inverse filter created to correct the response at one location in the room will not be accurate for some other location. This is a known problem with single-channel inverse filtering, and multichannel techniques have been suggested as a means to create a larger equalized listening area [8].

Two methods of inverse filtering are examined in this paper: time-domain least-squares and frequency-domain deconvolution. The frequency-domain method is more desirable due to its computational speed, but due to time-domain aliasing, blocking or wrapping effects occur. The severity of these blocking effects is dependent upon the characteristics of the IR being inverted and can be minimized by increasing the length of the inverse filter. Increasing the length of the inverse filter reduces the magnitude of the blocking effects but distributes the artifacts in time such that they may become more audible due to perceptual unmasking.

Formal subjective tests were conducted to evaluate the performance of various inverse-filtering strategies. The time-domain and frequency-domain techniques were evaluated as well as the effects of inverting only the minimum-phase component of the IR. Correcting the off-axis response with an inverse filter created from the on-axis IR is also evaluated. The time-domain method proved to be more subjectively robust but is more computationally intensive than the frequency-domain method. Therefore regularization and complex smoothing were used with the frequency-domain method to evaluate their effectiveness at reducing audible artifacts introduced by that inversion method.

1 INVERSE FILTER THEORY

The concept of inverse filtering originates from the linear filtering or convolution operation

$$d(n) = c(n) \otimes h(n) \quad (1)$$

where $d(n)$ is the result of convolving (denoted by \otimes) by the filter $c(n)$ with some correction filter $h(n)$. For example, $c(n)$ might be the IR of a loudspeaker while $h(n)$ might be a correction filter designed to produce a desired response $d(n)$. If one assumes that the desired “ideal” frequency response of a loudspeaker should be a flat spectrum with zero phase response, then Eq. (1) will become

$$\delta(n) = c(n) \otimes h(n) \quad (2)$$

where $\delta(n)$ is the Kronecker delta function or unit impulse function and $h(n)$ is the inverse filter of $c(n)$. The Kronecker delta function is defined as

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \quad (3)$$

Therefore the problem of inverse filtering, also referred to as a deconvolution problem, is to calculate $h(n)$ from Eq. (2).

In many real-world applications, physical constraints are such that a true inverse filter does not exist. For example, loudspeaker and room IRs are typically nonminimum phase, and so a true inverse does not actually exist [6]. Therefore one is left with the problem of trying to identify a suitable approximate solution for Eq. (2). To avoid producing an noncausal inverse filter, a modeling delay of m samples is employed in the inversion process so that the delta function in Eq. (2) becomes $\delta(n - m)$.

1.1 Least-Squares Time Domain

The first method to be considered is a least-squares (LS) time-domain filter design approach as described in Kirkeby and Nelson [2]. The optimal LS filter is derived in matrix form so that the convolution calculation is a matrix multiplication. One advantage of this form is that it lends itself well to the multichannel case. If $c(n)$ is the filter to be inverted, then one can construct \mathbf{C} , the convolution matrix of $c(n)$, as

$$\mathbf{C} = \begin{bmatrix} c(0) & & & 0 \\ \vdots & \ddots & & \vdots \\ c(N_c - 1) & \ddots & c(0) & \\ \vdots & \ddots & \vdots & \\ 0 & & c(N_c - 1) \end{bmatrix} \quad (4)$$

where N_c is the length of the filter $c(n)$. The number of columns of \mathbf{C} is equal to the length of the sequence h with which it is being convolved. In this case the sequence is h and the length is N_h . The number of rows is equal to the sum of the lengths of each sequence minus 1 ($N_h + N_c - 1$). In the notation, uppercase variable names (for example, \mathbf{A}) indicate a vector, and uppercase bold variable names (for example, \mathbf{A}) are used to indicate a matrix.

Using a deterministic LS approach and a desired response of $\delta(n - m)$ as defined, we have the expression

$$h(n) = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{a}_m \quad (5)$$

where $h(n)$ is the LS optimal inverse filter of $c(n)$ and $\mathbf{a}_m(n)$ is a column vector of zeros with a 1 in the m th position to create the modeling delay. The convolution matrix \mathbf{C} is of Toeplitz form (that is, the elements along a diagonal are identical) and the product $\mathbf{C}^T \mathbf{C}$ produces a symmetric matrix. By exploiting these properties one may use a Levinson–Durbin algorithm to compute a solution to Eq. (5) and speed up the computation of the inverse [9].

1.2 Frequency-Domain Deconvolution

Although fast algorithms exist for calculating the time-domain solution given by Eq. (5), it is still time-consuming for long filter sizes. As an alternative, a fast frequency-domain deconvolution method can be used to derive an inverse filter [3]. This approach is based on the fact that a time-domain convolution becomes a multiplication in the frequency domain via the discrete Fourier transform (DFT), and the deconvolution process can thus be written as

$$H(k) = \frac{D(k)}{C(k)} \quad (6)$$

where $C(k)$ and $H(k)$ are the DFTs of $c(n)$ and $h(n)$, respectively, and $D(k)$ is the DFT of $\delta(n - m)$. The required modeling delay is incorporated into $D(k)$, resulting in a delayed pulse.

A potential problem inherent to this method of filter inversion is that a multiplication or division in the frequency domain is a circular operation due to the periodicity of the DFT, and blocking or wrapping artifacts will result. This is similar to the problem that arises when one wishes to perform a linear convolution in the frequency domain without the proper zero padding that is needed to avoid the circular wrapping effects. The deconvolution process described in Eq. (6) can also benefit from zero padding, though here the blocking effects will always exist to some extent. Since real-world filtering (such as playback through a loudspeaker) is a linear convolution, any filter that was optimally designed from a circular process such as the deconvolution process of Eq. (6) will create blocking effects. The blocking effects can be reduced in some instances by using longer DFTs, which will result in a longer inverse filter being computed. By using longer DFTs and creating longer inverse filters one can push the blocking effects further out in time and also attenuate them.

1.3 Minimum-Phase Decomposition

As stated, a true inverse may not exist due to the IR being nonminimum phase. To avoid this issue altogether one can decompose the IR into its minimum-phase and all-pass components [5] and only invert the minimum-phase component. For the IR $c(n)$ and its corresponding transfer function $C(k)$, this can be expressed as

$$C(k) = M(k)A(k) \quad (7)$$

where $M(k)$ is the minimum-phase component and $A(k)$ is the all-pass component. An efficient method of computing the minimum-phase component is presented in [6], and it will not be elaborated here.

2 SUBJECTIVE TEST METHOD

2.1 Measurement of the Impulse Responses

The IRs of two different types of loudspeakers were measured, on-axis (0°) and off-axis (45°) in an anechoic environment. The first loudspeaker, which will be referred to as loudspeaker A, was a conventional two-driver loudspeaker where the tweeter and the woofer are physically located at two separate locations on the front baffle of the loudspeaker. The second loudspeaker, which will be referred to as loudspeaker B, was a dual-concentric type, where the tweeter is located in the center of the low-frequency driver. The layout of the measurement setup is shown in Figs. 1 and 2. To achieve a more diverse range of IRs, loudspeaker A was placed horizontally so that the woofer and the tweeter are side by side. Therefore a total of three IRs were measured on-axis, at 45° on the tweeter side, and at 45° on the woofer side. Due to the dual-concentric loudspeaker (loudspeaker B) having both vertical and horizontal symmetry, only two IRs were measured for it, on-axis and at 45° .

The IRs were measured in an anechoic environment using the CRC-MARS (Multichannel Audio Research System) software developed at the Communications Research Centre (CRC). The source signal was a maximum-length sequence that was captured using an omnidirectional measurement microphone. The length of the sequence was 32767 samples and was sampled at 44.1 kHz. 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 [10]. The IRs were then truncated to 1024 samples (23.2 ms). Fig. 3 shows IRs, whereas Fig. 4 shows the on-axis and off-axis (45° on the woofer side and 45° on the tweeter side) magnitude responses for loudspeaker A.

Fig. 5 shows the IRs, whereas Fig. 6 shows the on-axis and off-axis magnitude responses for loudspeaker B. In

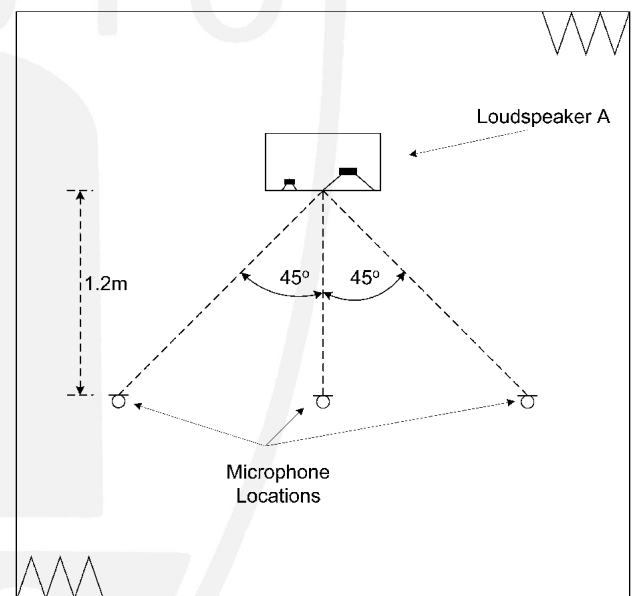


Fig. 1. IR measurement setup for two-driver loudspeaker A.

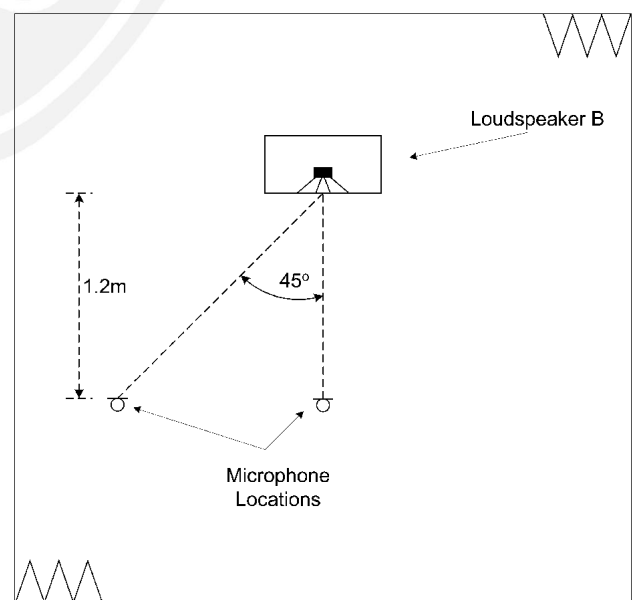


Fig. 2. IR measurement setup for dual-concentric loudspeaker B.

total, five loudspeaker IRs were measured and used for the calculation of the correction filters and for the subjective tests in this paper.

2.2 Subjective Test Strategy

To evaluate the performance of the computed inverse filters, double-blind subjective tests were conducted using the multiple stimulus with hidden reference anchors (MUSHRA) (ITU-R BS.1534) test method over headphones [11]. The multistimulus methodology allows the subject to instantly compare several test items in order to

derive a score for each item. Using a slightly modified version of the ITU-R BS.1116-1 impairment scale (see Fig. 8) [12], subjects were asked to compare each test item to a reference signal and rate the severity of any artifacts introduced by the processing. The reference signal consisted of the audio sequence without any processing. A score of 5.0 indicated that the subject could hear no perceptible difference between the test item and the reference signal. Conversely, a score of 1.0 indicated that there were large differences and that the artifacts in the test item were unacceptable.

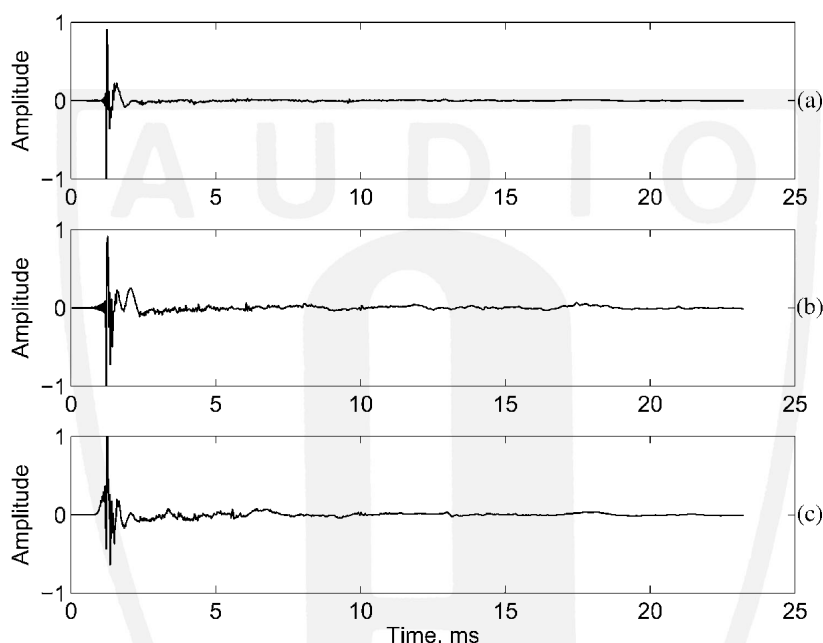


Fig. 3. Three IRs of conventional two-driver loudspeaker: loudspeaker A. (a) On-axis. (b) 45° off-axis on tweeter side. (c) 45° off-axis on woofer side.

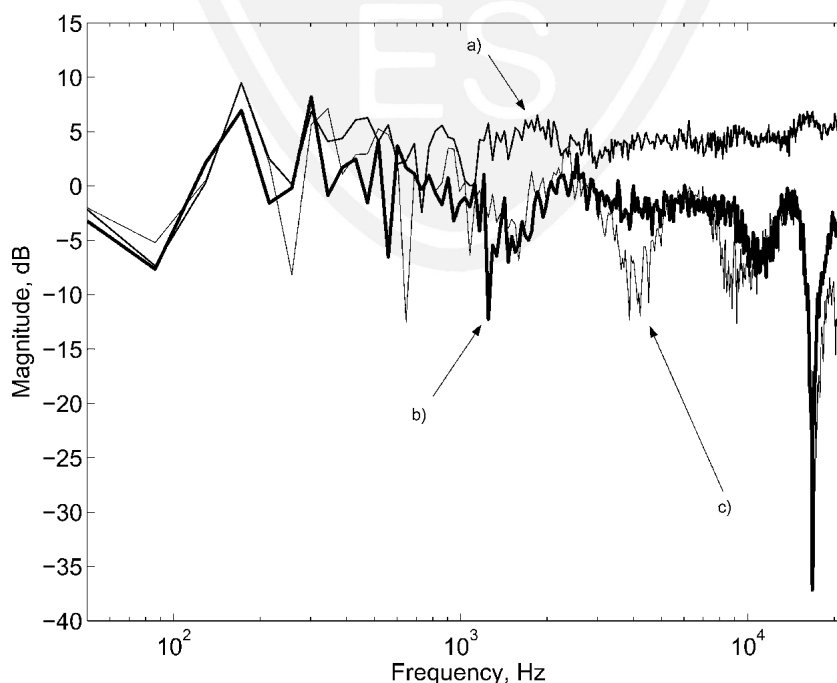


Fig. 4. Magnitude of frequency response for loudspeaker A. (a) On-axis. (b) 45° off-axis on tweeter side. (c) 45° off-axis by on woofer side.

The test files were created from a mono recording of a castanets passage at a sample rate of 44.1 kHz. The test strategy was to filter the audio source signal with each of the five measured IRs and then process these filtered audio signals with an inverse filter to correct for the loudspeaker's response. Ideally, if the inverse filter were perfect, the result of this process should yield the same audio file as the original audio source signal. Test files that were filtered with the loudspeaker IRs only were also included in the experiments to allow a direct comparison between the corrected and uncorrected loudspeaker responses. Fig. 7 shows an overview of how the subject, while listening over headphones (STAX LAMBA PRO), was able to

switch between the unprocessed reference signal and the reference signal processed by a loudspeaker response $C_i(k)$, alone or in combination with an inverse or a correction filter $H_j(k)$. The terms inverse and correction will be used synonymously throughout this paper.

All filtering of the reference signal was done off-line using double-precision floating-point arithmetic. During the tests a computer-based switching system was used to play the audio files to the subject. Subjects were presented with the computer interface shown in Fig. 8. The subjects were able to switch between the different processed files simply by moving the mouse and clicking on one of the lettered buttons. The reference audio file could be selected

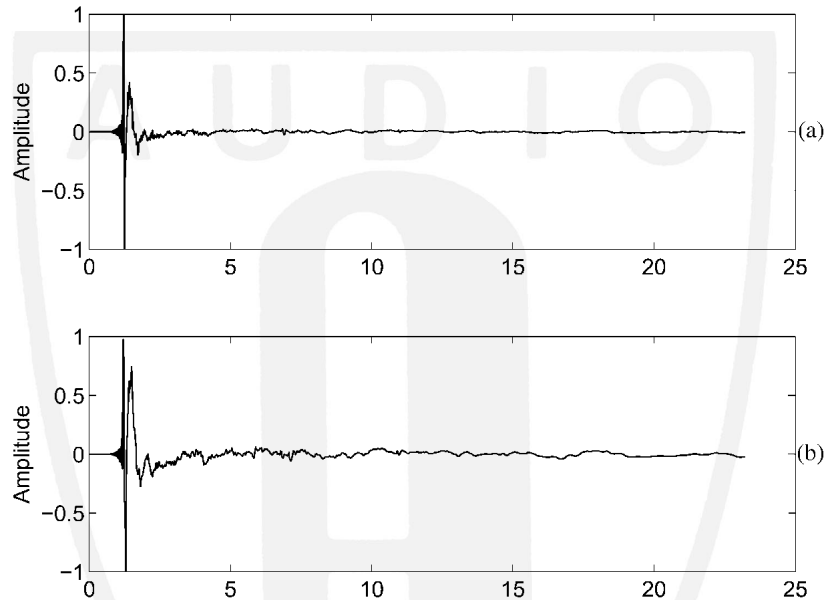


Fig. 5. IRs of dual-concentric loudspeaker, loudspeaker B. (a) On-axis. (b) 45° off-axis.

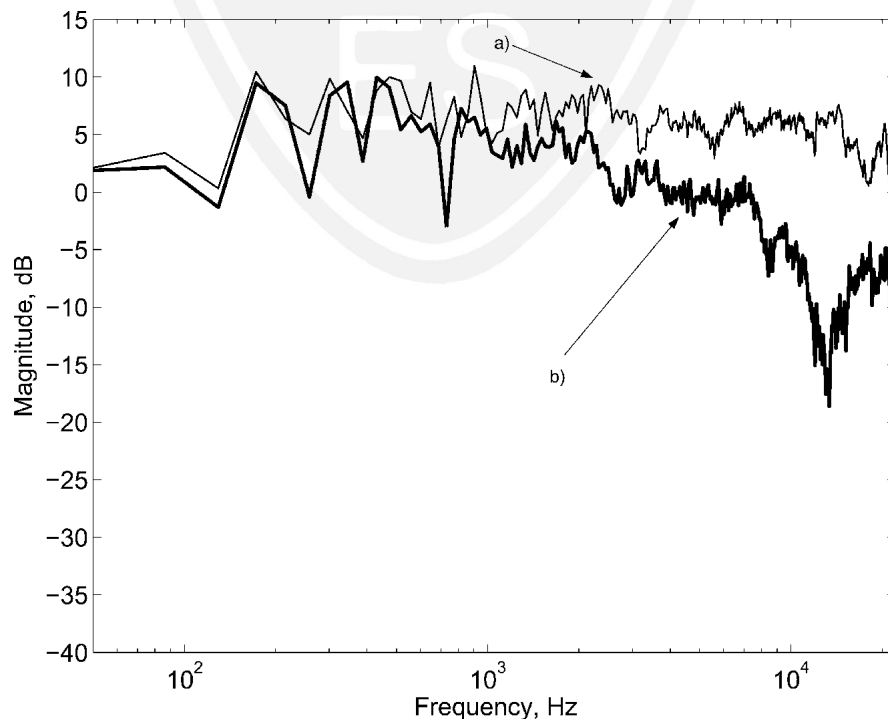


Fig. 6. Magnitude of frequency response for loudspeaker B. (a) On-axis. (b) 45° off-axis.

at any time by clicking on the REF button. The music played continuously while the subject switched between audio files, and audible switching artifacts were avoided by using a rapid cross-fade. Subjects were allowed to compare the audio files as much as they needed in order to make their ratings.

Before each formal test the subjects completed a training session where they were exposed to the full range of test conditions that they would later encounter and score in the blind test. During the formal test each subject was alone and allowed to complete the experiment at their own pace.

3 EXPERIMENT 1

In the first experiment inverse filters were derived using the LS time-domain and frequency-domain deconvolution

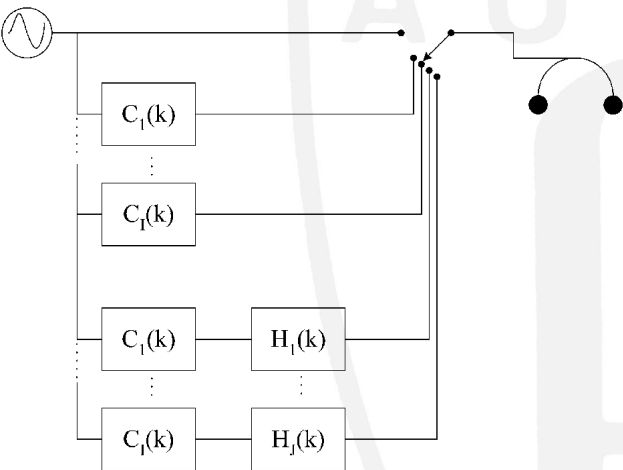


Fig. 7. Schematic diagram showing concept of subjective test setup. $C_l(k)$ —response of loudspeaker; $H_l(k)$ —inverse filter used.

methods described earlier. Inverse filters for the five measured IRs were computed. As stated, the loudspeaker IRs were 1024 samples long while the resulting inverse filters were 2048 samples long. The modeling delay was set to 1024 samples to produce causal inverse filters. Minimum-phase inverses were also computed for comparison, and Table 1 lists the combinations of IRs and inverse filters that were linearly convolved for loudspeaker A. In the

Table 1. Loudspeaker IRs and the inverse filters that were linearly convolved with the audio source to create the audio files for the subjective test.

Loudspeaker	Inverse	Method
A-0°	—	—
A-0°	0°	Time
A-0°	0°	Freq.
A-0°	Min ϕ -0°	Time
A-0°	Min ϕ -0°	Freq.
A-45°T	—	—
A-45°T	ϕ -45° T	Time
A-45°T	ϕ -45° T	Freq.
A-45°T	Min ϕ -45° T	Time
A-45°T	Min ϕ -45° T	Freq.
A-45°W	—	—
A-45°W	ϕ -45° W	Time
A-45°W	ϕ -45° W	Freq.
A-45°W	Min ϕ 45° W	Time
A-45°W	Min ϕ 45° W	Freq.
A-45°T	ϕ -0°	Time
A-45°T	ϕ -0°	Freq.
A-45°T	Min ϕ -0°	Time
A-45°T	Min ϕ -0°	Freq.
A-45°W	ϕ -0°	Time
A-45°W	ϕ -0°	Freq.
A-45°W	Min ϕ -0°	Time
A-45°W	Min ϕ -0°	Freq.

T/W—side of loudspeaker, tweeter (T) or woofer (W), Min ϕ —only the minimum-phase component was inverted.

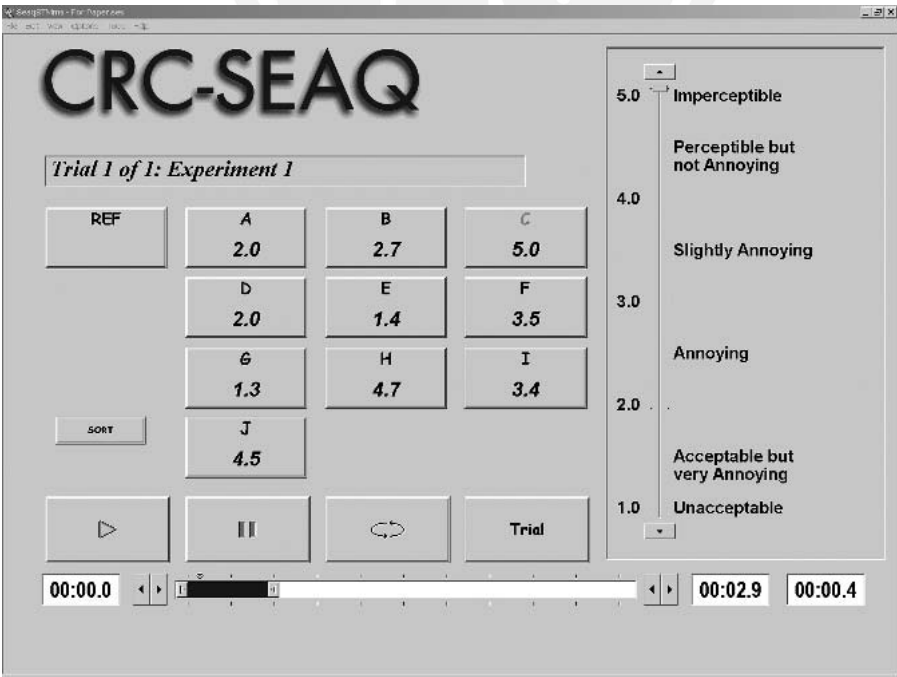


Fig. 8. User interface of computer-based switching system used for subjective tests.

loudspeaker column, T indicates the tweeter side and W indicates the woofer side. Min ϕ indicates that only the minimum-phase component of the IR was inverted. Also included in Table 1 are the situations where the audio source was linearly convolved with only the loudspeaker IR (that is, no inverse filter was included), and in that case “inverse” and “method” are shown as —. The same combination of IRs and inverse filters was used for loudspeaker B, except that only one group of 45° IRs was used due to the symmetry of the loudspeaker. Therefore a total of 38 test files, including a hidden reference, were graded by the subjects. A total of 17 subjects participated in this experiment.

An analysis of variance (ANOVA) [13], [14], was performed on the subjective test data, and the results for the first part of experiment 1 (excluding the off/on-axis corrections) are shown in Fig. 9. The results of the ANOVA show highly significant main effects ($p < 0.001$) due to the loudspeaker type and the inverse-filtering method. The error bars in Fig. 9 represent the critical difference for the experiment derived using a t -test [13], [14]. As such any two data points are statistically different ($p < 0.05$) if their error bars do not overlap, while overlapping error bars indicate that the data points must be considered to be statistically identical. Fig. 9 plots the mean subjective grade versus five different correction methods (including no correction) for the five measured loudspeaker IRs. Also included in the figure is the data point representing the subjective score given to the hidden reference. Any data point in the figure whose error bars overlap with the hidden reference can be considered to be subjectively transparent, and thus the inverse filter can be considered to have worked perfectly.

The subjective scores for the uncorrected loudspeaker responses vary over a wide range. As expected, the off-axis responses for the two loudspeakers received lower scores due to the reduction in high-frequency content. The results show that the time-domain correction always provides a subjective improvement whereas the performance of the frequency-domain method is quite variable. The frequency-domain correction method provides a consistent improvement for loudspeaker B, but seriously degrades the performance of loudspeaker A for the on-axis case.

The effect of minimum-phase corrections is also quite varied. For loudspeaker A correcting only the minimum-phase portion of the IR results in lower subjective performance than if the entire IR were corrected. For loudspeaker B correcting only the minimum-phase portion of the IR improved the subjective performance for the frequency-domain correction method, but degrades the performance for the time-domain method.

The magnitude response for loudspeaker B and three corrected magnitude responses are shown in Fig. 10 with the curves offset for clarity of presentation. Fig. 10(a) is the uncorrected on-axis response of loudspeaker B. The next curves are the result of using a 2k time-domain (TD) LS inverse filter, a 2k minimum-phase time-domain LS inverse filter, and a 2k frequency-domain (FD) inverse filter, respectively. The full time-domain correction has the flattest magnitude response, and this is reflected in the results of the subjective test. The minimum-phase time-domain correction performed better subjectively than the frequency-domain method, and this is again reflected in terms of the “flatness” of the responses.

The corrected IRs (that is, the IR convolved with the inverse filters) for the correction filters shown in Fig. 10

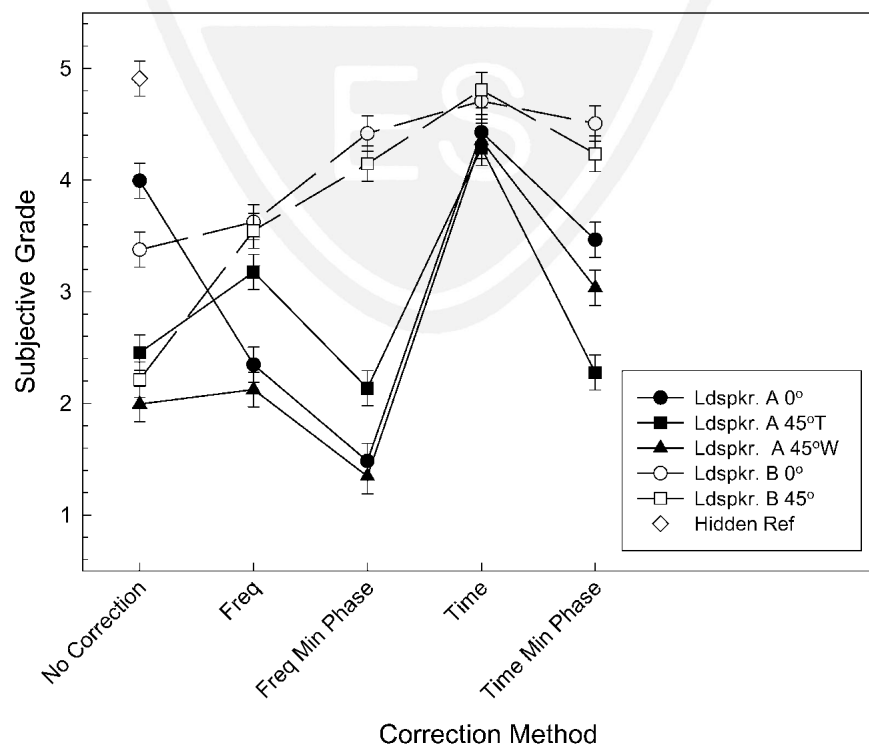


Fig. 9. Mean subjective grades versus correction method from first part of experiment 1. Curves represent five different IR configurations, three for loudspeaker A and two for loudspeaker B.

are plotted in Fig. 11. Ideally, the corrected IR would result in a Dirac pulse and indeed, Fig. 11(a) and (c) appears to achieve this. Fig. 11(b) is the only one showing large differences from a perfect pulse. However, as pointed out by Fielder [15], this type of plot can be deceiving. Fielder plotted the magnitude of the corrected responses in dB, as shown in Fig. 12, where the differences between a perfect pulse and the corrected responses are more easily seen. It can now be seen that the error in the corrected IR resulting from the time-domain

inverse filter [Fig. 12(a)] is down almost 60 dB and is distributed uniformly over time. Conversely the error in the minimum-phase correction [Fig. 12 (b)] occurs only after the delayed delta function. The plot for the frequency-domain correction [Fig. 12 (c)] reveals artifacts above -50 dB that are pushed out in time (both before and after the delayed delta function). These artifacts produced delay-type effects that were clearly audible to the subjects. Also of interest in Fig. 12 (c) is the region immediately after the delta function. In this region the error

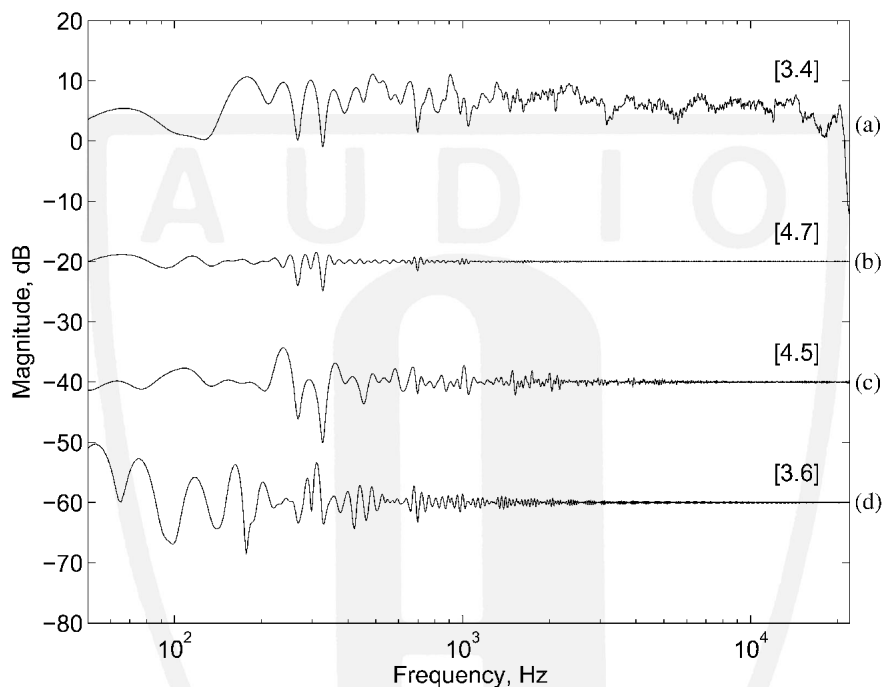


Fig. 10. On-axis magnitude response and corrections for loudspeaker B. (a) Uncorrected response. (b) Corrected with LS time-domain 2k filter. (c) Corrected with minimum-phase LS time-domain 2k filter. (d) Corrected with FD 2k filter. Curves are offset for clarity of presentation; subjective grades are shown in square brackets.

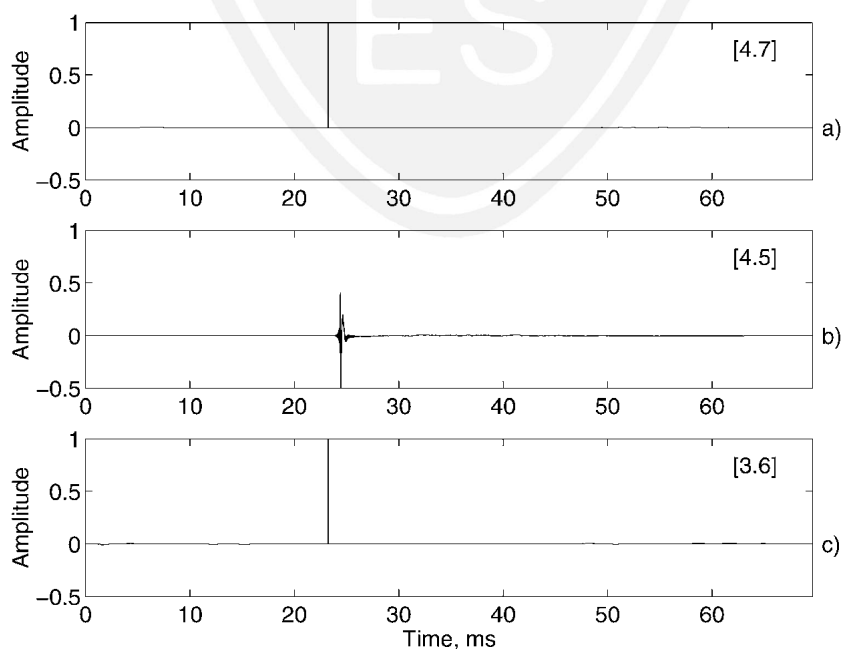


Fig. 11. On-axis corrected time response for loudspeaker B with the amplitude plotted on a linear scale. (a) 2k LS time-domain correction. (b) 2k minimum-phase LS time-domain correction. (c) 2k FD correction. Subjective grades are shown in square brackets.

is zero due to the circular processing used to compute this inverse filter.

Overall the subjective results as shown in Fig. 9 indicate poorer performance when correcting loudspeaker A as compared with loudspeaker B. This poorer performance may be due to some characteristic of this loudspeaker's IR, which makes it more difficult to invert. It may be that a longer inverse filter is required for better performance. One reason for needing a longer inverse filter could be due to the transfer function having zeros close to the unit

circle. When inverting such a filter, those zeros will become poles. Since these poles are close to the unit circle, the inverse filter will have a longer decay time. This problem can occur due to the rolloff of the antialiasing filter in the analog-to-digital converter. This would result in an inverse filter with a large boost at or near Nyquist to correct for this rolloff. Fig. 13 shows the magnitude response of two inverse filters for loudspeaker A [(a) on-axis and (b) 45° off-axis (tweeter side)], both calculated with the frequency deconvolution method. It can be seen that

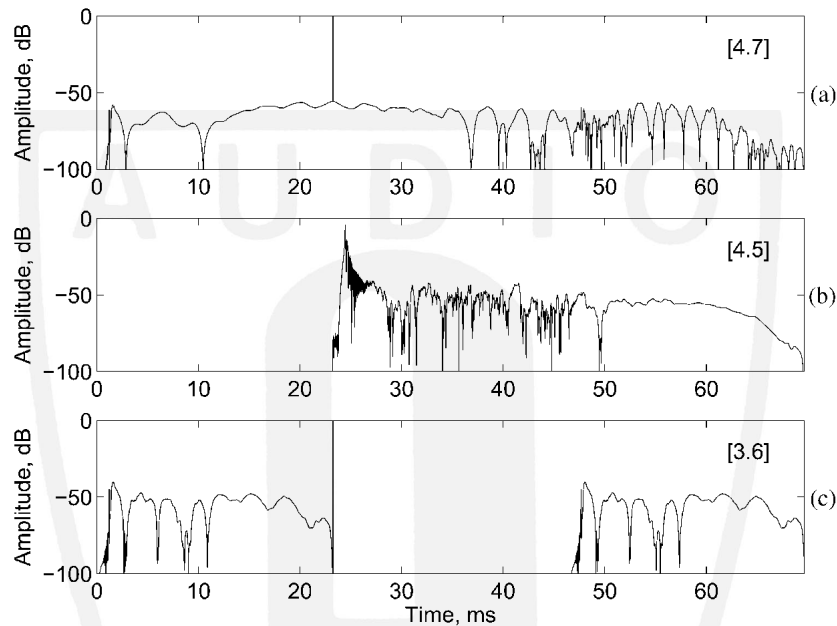


Fig. 12. On-axis corrected time response for loudspeaker B with the amplitude plotted in dB. (a) 2k LS time-domain correction. (b) 2k minimum-phase LS time-domain correction. (c) 2k FD correction. Subjective grades are shown in square brackets.

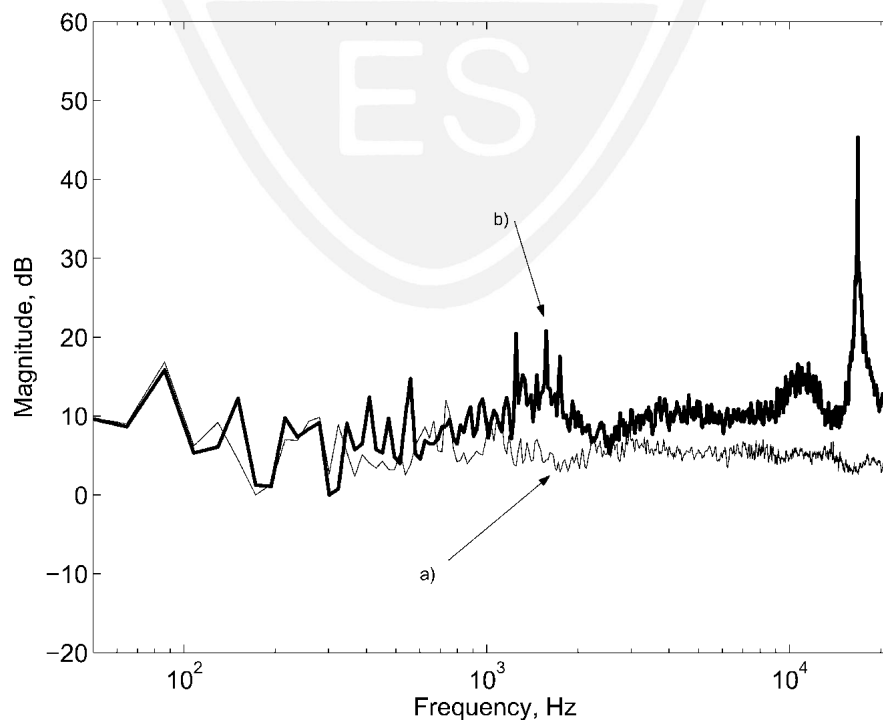


Fig. 13. Magnitude of frequency response of two inverse filters for loudspeaker A, calculated using the frequency deconvolution method with a length of 2048 samples. (a) On-axis. (b) 45° off-axis tweeter side.

the boost at or near Nyquist is not very much, and in fact, the off-axis response has a larger boost not at Nyquist. [One reviewer commented on that a solution to this would be to use the response of the antialiasing filter as the desired signal $D(k)$.] The fact that excessive boosts can cause the inverse filter to ring can be a problem, and solutions to that will be explored further in the next section.

The results of the first experiment suggest that frequency deconvolution is not as effective as the time-domain LS approach, as mentioned in [16]. For both loudspeakers the time-domain inverse method performed better than the frequency-domain method.

Fig. 14 shows the magnitude response for loudspeaker A (on-axis) as well as the corrected magnitude responses using the time-domain and frequency-domain deconvolution approaches. The curves are offset for clarity of presentation. As was found for loudspeaker B, it can be seen that the time-domain inverse filter method results in a flatter magnitude response. This fact is reflected in the subjective results where the time-domain method obtained a higher score than the frequency-domain method.

Fig. 15 shows the corrected IRs (that is, the IR convolved with the inverse filters) for loudspeaker A. The error in the corrected IR resulting from the time-domain inverse filter is again down by almost 60 dB and is distributed evenly over time. Conversely, the error for the frequency-domain method is above -40 dB. The resulting artifacts were easily audible to the subjects.

3.1 Off-Axis with On-Axis Inverse Filter

Fig. 16 shows the subjective results from experiment 1 for the cases where the on-axis inverse filter was used to correct the off-axis response of the loudspeakers. Again the error bars represent the critical difference for this experiment. In

all cases there was no improvement in the subjective rating as compared to not having any inverse filter. The time-domain corrections have the highest scores of all the correction scenarios. This result highlights a fundamental problem with single-channel inverse filtering. That is, the response can only be inverted correctly at one point. This demonstrates that trying to correct a loudspeaker response based on only the on-axis response may not be effective.

The magnitude responses for loudspeaker B and two corrected magnitude responses are shown in Fig. 17 with the curves offset for clarity of presentation. Fig. 17(a) is the uncorrected loudspeaker response, whereas Fig. 17(b) is the result of using a time-domain inverse filter based on the on-axis IR. Fig. 17(c) is the result of using a frequency-domain inverse filter.

It is clear from the figure that neither inverse filter based on the on-axis response is effective at correcting the off-axis response of the loudspeaker. This is not surprising since the on-axis response does not exhibit the high-frequency rolloff found in the off-axis response. In this case the inverse filter actually degrades the quality of the signal somewhat.

Fig. 18 shows the corrected IRs for the off-axis correction with an on-axis inverse filter for loudspeaker B. For the time-domain inverse the errors are down by about -40 dB, while again the frequency-domain inverse has peaks down by about -35 dB. The delay between these peaks and the main delta function is significant and caused artifacts that were readily audible to the subjects.

4 EXPERIMENT 2

In experiment 1 it was found that the frequency-domain inverse filters did not perform very well. A likely reason

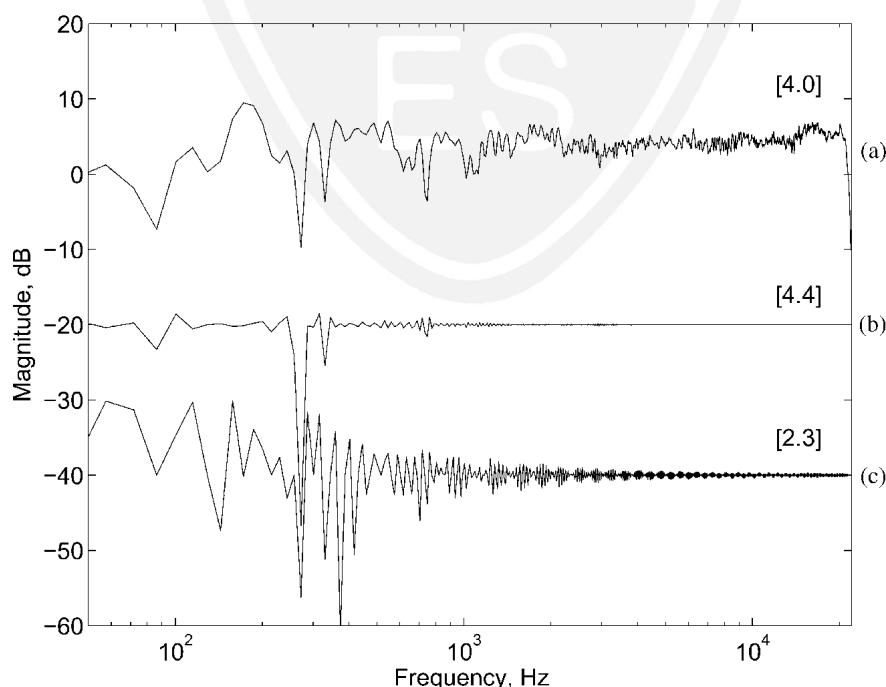


Fig. 14. On-axis magnitude response and corrections for loudspeaker A. (a) Uncorrected response. (b) Corrected with LS time-domain 2k filter. (c) Corrected with minimum-phase frequency-domain 2k filter. Curves are offset for clarity of presentation; subjective grades are shown in square brackets.

for this is that the inverse filters were not long enough. As one increases the length of the filter, the block processing artifacts should be pushed out further in time and be lower in amplitude. The purpose of experiment 2 was to examine the subjective effect of the length of the inverse filter created via frequency deconvolution. The inverse filters computed for this experiment were of length 2k, 4k, 8k, 16k, 32k, and 64k samples. In order to compare the relative performance, time-domain correction filters (length

2k) were also included in this experiment. In all cases the full IR (minimum plus excess phase) was corrected. A total of 36 test items, including a hidden reference, were graded by 17 subjects in this experiment.

An ANOVA was conducted on the results of experiment 2, and highly significant main effects were found due to the type of loudspeaker as well as the correction method. The results are plotted in Fig. 19, with the error bars once again representing the critical difference. The

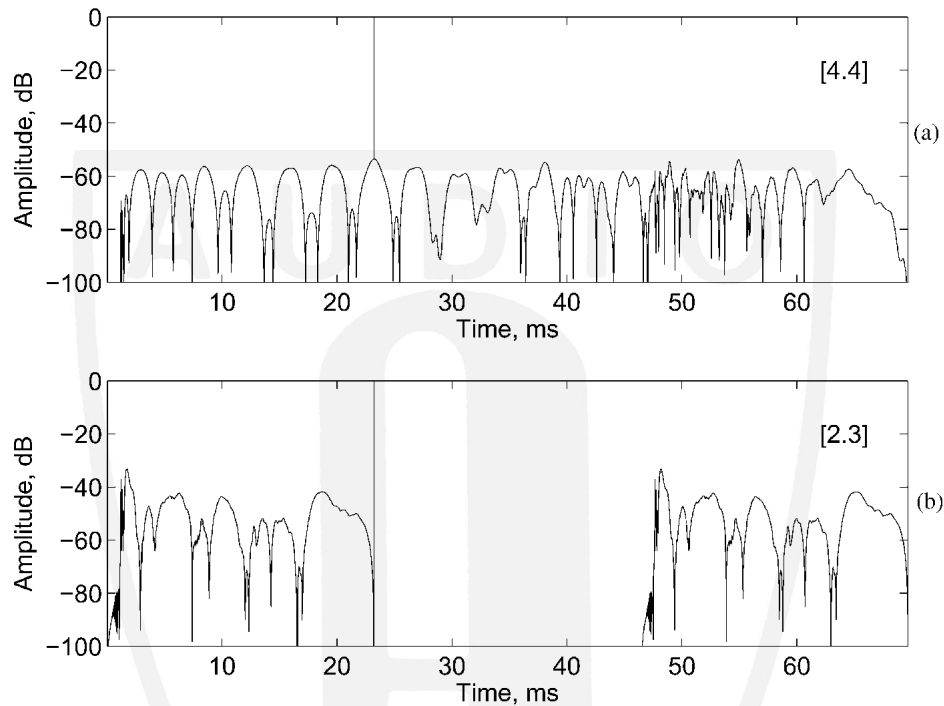


Fig. 15. On-axis corrected time response for loudspeaker A. (a) 2k LS time-domain correction. (b) 2k FD correction. Subjective grades are shown in square brackets.

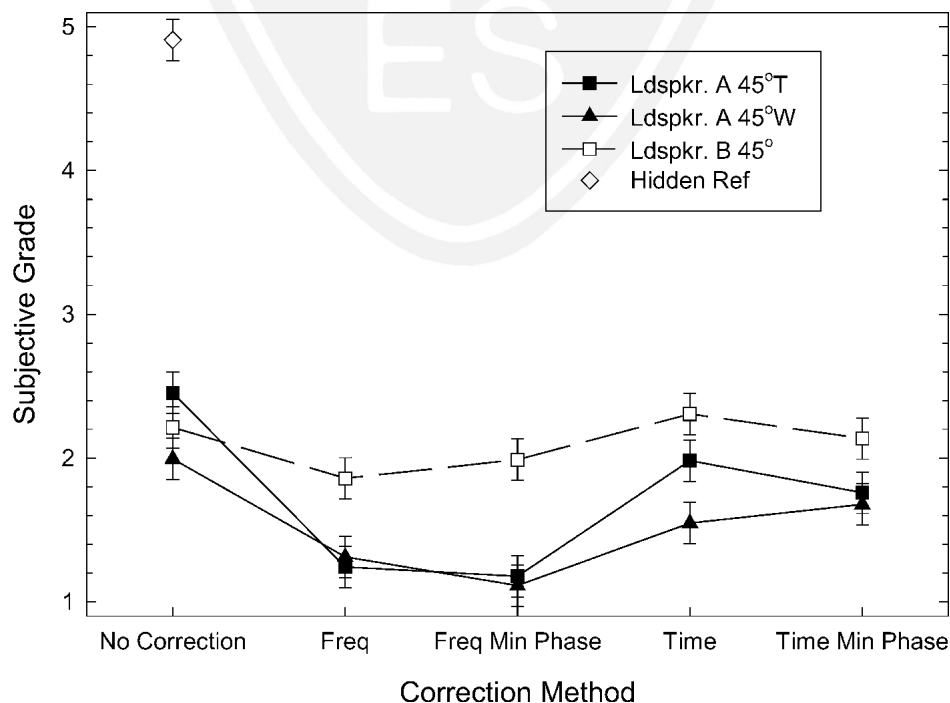


Fig. 16. Mean subjective grade versus correction method for off-axis responses corrected with on-axis inverse filter.

figure plots the mean subjective grades versus the method used to derive the inverse filter. The mean grade given to the hidden reference is also shown in the plot.

The figure shows that, in the absence of any correction filters, the on-axis response of both loudspeakers scores higher than their corresponding off-axis responses. Also, the time-domain correction filter always provides a significant subjective improvement. This is consistent with the results of experiment 1. Also, as anticipated, the performance of the frequency-domain inverse filter improves

as its length is increased. However, the length of inverse filter required in order to obtain a subjectively “perfect” inverse filter is not consistent.

For loudspeaker B the performance improves systematically as the length of the inverse filter is increased. This is true for both the on-axis and the off-axis responses. Moreover, the length of inverse filter required to obtain a dramatic improvement in subjective quality is not very long. Conversely, for loudspeaker A the relation between the length of the frequency-domain inverse filter and the corre-

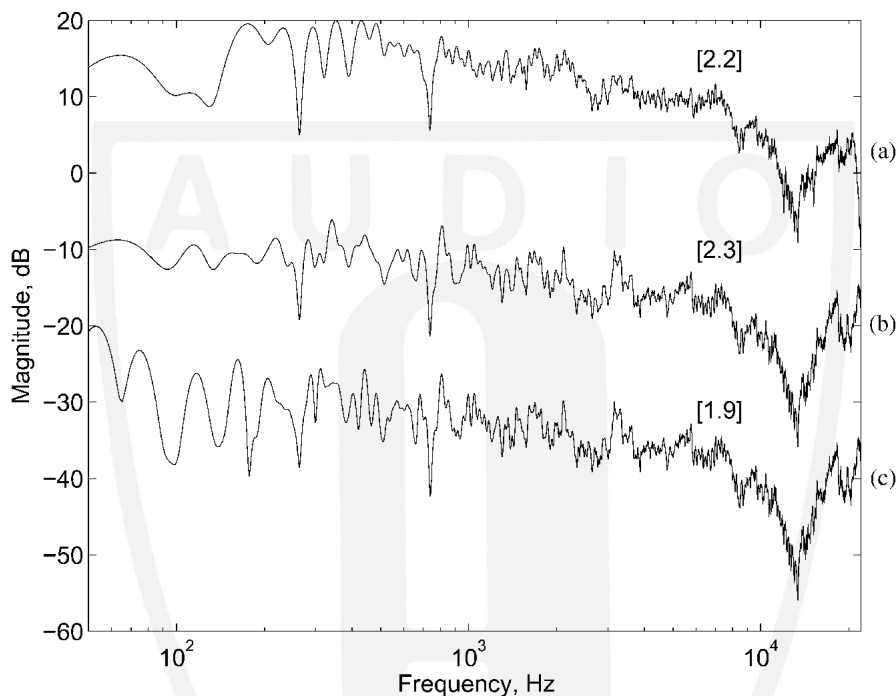


Fig. 17. Off-axis magnitude response and correction using on-axis inverse filter for loudspeaker B. (a) Uncorrected response. (b) Corrected with LS time-domain 2k filter. (c) Corrected with frequency-domain 2k filter. Subjective grades are shown in square brackets.

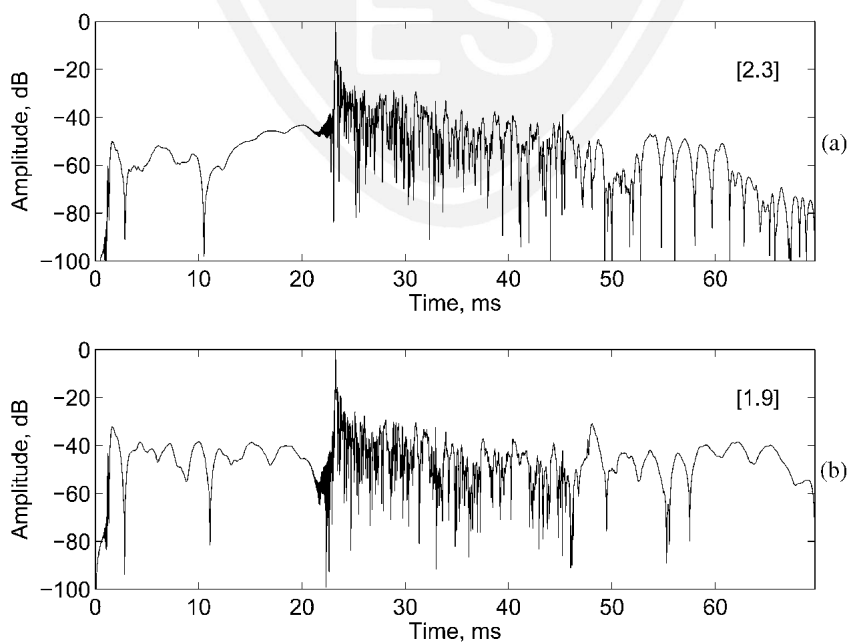


Fig. 18. Off-axis corrected, with on-axis inverse filter, IR for loudspeaker B. (a) 2k LS time-domain correction. (b) 2k frequency-domain correction. Subjective grades are shown in square brackets.

sponding subjective performance is quite varied. Increasing the length of the inverse filter does not yield any subjective improvement unless a sufficiently long filter is used. For the on-axis response of loudspeaker A, none of the frequency-domain inverse filters was long enough to provide any subjective improvement. For this case the frequency-domain inverse filter always caused a degradation in sub-

jective quality, and so it is better to remove the inverse filter altogether. The subjective quality of the off-axis responses (both tweeter and woofer side) could be improved with sufficiently long frequency-domain inverse filters.

Fig. 20 shows the off-axis (tweeter) magnitude response and the corrected responses for loudspeaker A for three different-length frequency-domain inverse filters. The

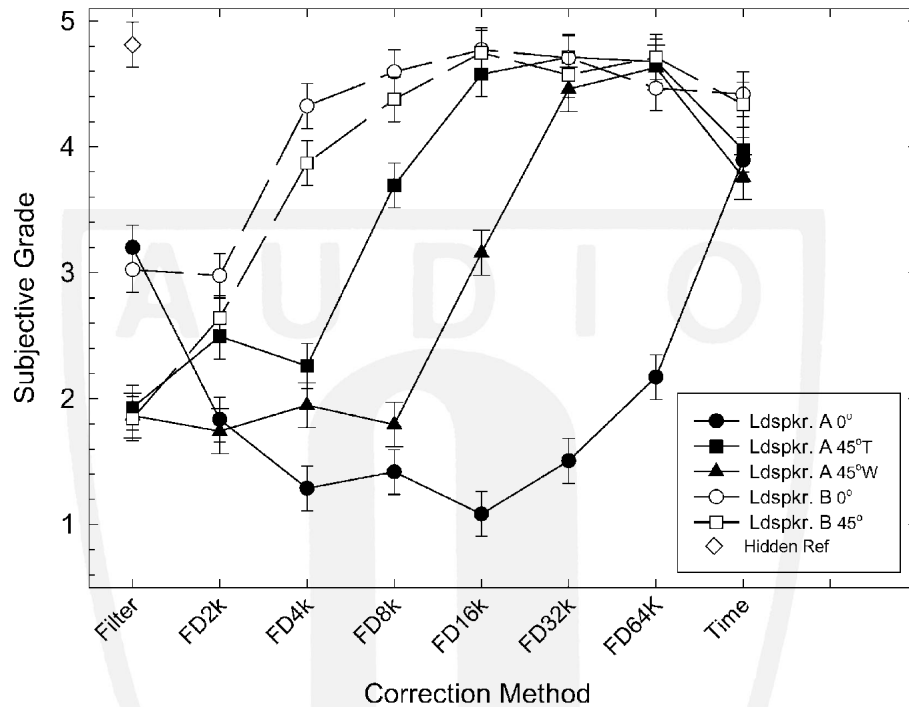


Fig. 19. Mean subjective grades versus length of frequency-domain inverse filter for five different loudspeaker configurations. Lengths of inverse filters are included in label of frequency-domain (FD) correction method. FD2k means a 2k or 2048 length filter was used. Time-domain filter was 2048 samples long.

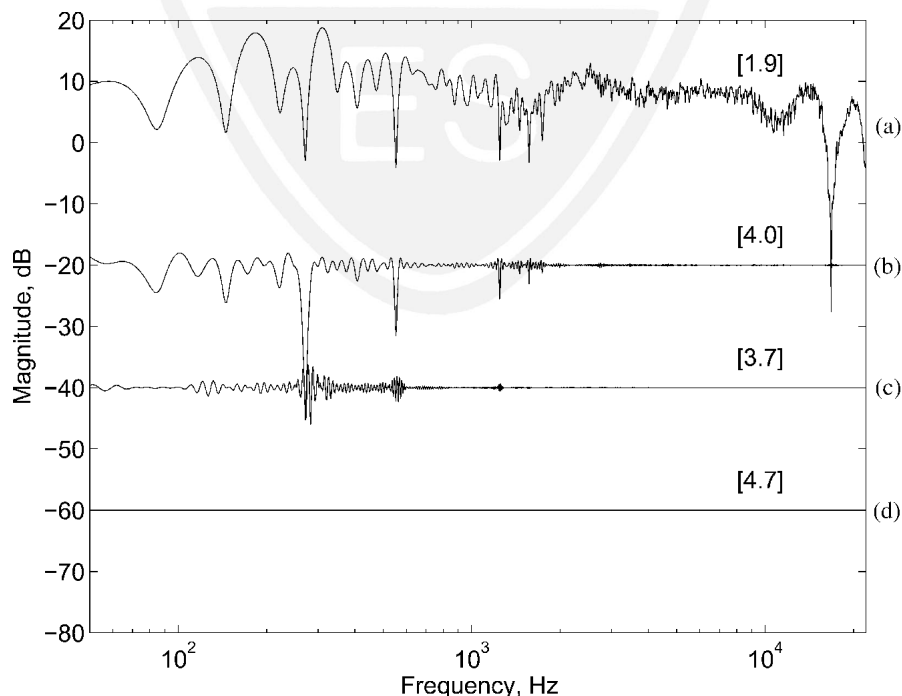


Fig. 20. Off-axis (tweeter side) magnitude response and corrections for loudspeaker A. (a) Uncorrected response. (b) Corrected with LS time-domain 2k filter. (c) Corrected with 8k FD inverse filter. (d) Corrected with 32k FD inverse filter. Subjective grades are shown in square brackets.

curves are offset for clarity of presentation. It can be seen that the magnitude response becomes flatter as the length of the inverse filter is increased. It should be noted that the performance of the time-domain filter should also improve as its length is increased, but this was not tested in this experiment.

The result for the 32k filter shown in Fig. 20(d) shows a perfectly flat magnitude response, and in the experiment subjects rated it as perceptually identical to the hidden reference signal.

The corrected IRs (that is, the IR convolved with the inverse filters) for the correction filters for loudspeaker A are shown in Fig. 21. Fig. 21(a) plots the corrected IR when a 2k time-domain correction filter is used. Fig. 21(b) and (c) plots the corrected IRs when 8k and 32k frequency-domain inverse filters are used, respectively. It can be seen that increasing the length of the inverse filter pushes the error further away from the delta function and also reduces the level of the error. It is interesting to note that the errors shown in Fig. 21(b) were clearly audible to the listeners (see Fig. 19), even though they are at a level of about -60 dB. This is due to the relatively long pre-delay (approximately 90 ms) between the error component and the delta function.

5 EXPERIMENT 3

Earlier experiments showed mixed results when only the minimum-phase portion of the IR was corrected. As described earlier, a modeling delay is usually employed when calculating the inverse filter so that the result is causal. This modeling delay should not be needed if the filter being inverted is minimum phase. Therefore the goal of this experiment was to examine the effect of the modeling delay when calculating the inverse of the minimum-phase component of the IRs of the loudspeakers.

Only the on-axis IRs from each loudspeaker were used in this experiment since the results of the first two experiments indicated that these IRs represented the range of results adequately. The IRs were 1024 samples long, and the inverse filters that were calculated were 2048 and 4096 samples in length. The frequency-domain deconvolution method was used for the calculation of the inverse filters. For each of the two loudspeakers a total of six inverses were calculated: two full inverses (2k and 4k), two minimum-phase inverses with a modeling delay (2k and 4k), and two minimum-phase inverses without a modeling delay. The uncorrected response was also included for comparison to see whether the correction filters actually degraded the subjective performance from when no correction was applied. A summary of these filter conditions is provided in Table 2, along with the corresponding figure label abbreviations. A total of 11 subjects participated in the formal subject test.

An ANOVA was performed on the subjective test data and revealed highly significant main effects ($p < 0.001$) due to the loudspeaker type and inverse filtering method. A plot of the overall mean subjective grades versus correction filter conditions is shown in Fig. 22.

Table 2. Correction filter conditions for experiment 1, explaining figure labels.

Figure Label	Length	Correction Method
Filt	0	No correction
2k = Freq	2k	Full inverse
4k = Freq	4k	Full inverse
2k = Freqmin	2k	Minimum = phase inverse
4k = Freqmin	4k	Minimum = phase inverse
2k = Freqmin_nd	2k	Minimum phase with no modeling delay
4k = Freqmin_nd	4k	Minimum phase with no modeling delay

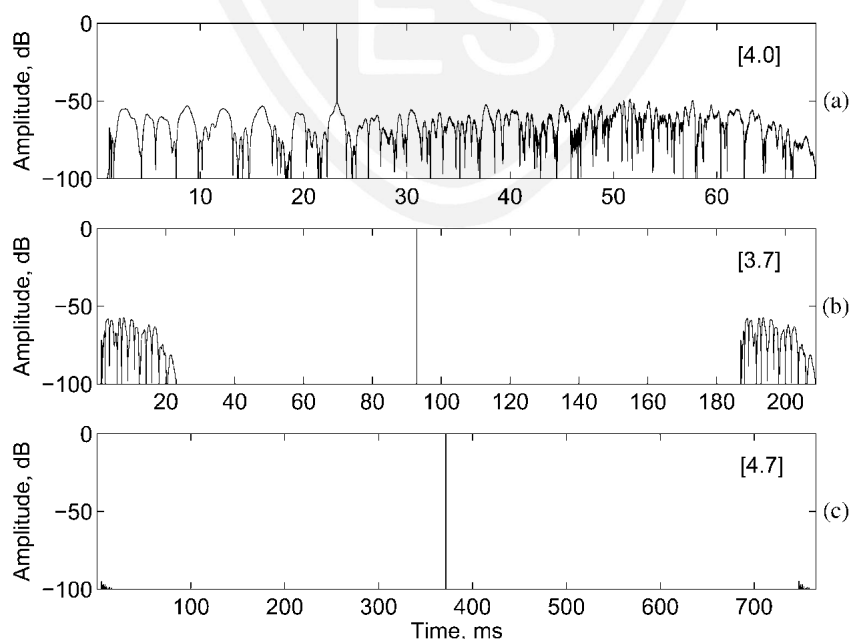


Fig. 21. Off-axis (tweeter side) corrected time response for loudspeaker A. (a) 2k LS time-domain correction. (b) 8k FD correction. (c) 32k FD correction. Note different time scales. Subjective grades are shown in square brackets.

The error bars in Fig. 22 represent the critical difference for the experiment using a *t*-test. It can be seen that the correction filters using the entire IR (2k-Freq and 4k-Freq) scored the same or lower than the uncorrected condition. Also, the performance of the minimum-phase inverse filters improves significantly when the modeling delay is removed.

Since the ANOVA showed that there was a highly significant ($p < 0.001$) effect due to the loudspeaker type, one can look at the loudspeakers separately. Fig. 23 shows the mean subjective grade versus seven different correction methods for each of the two loudspeakers.

It can be seen that the 4k full inverse improved the performance with loudspeaker B, but did not for loud-

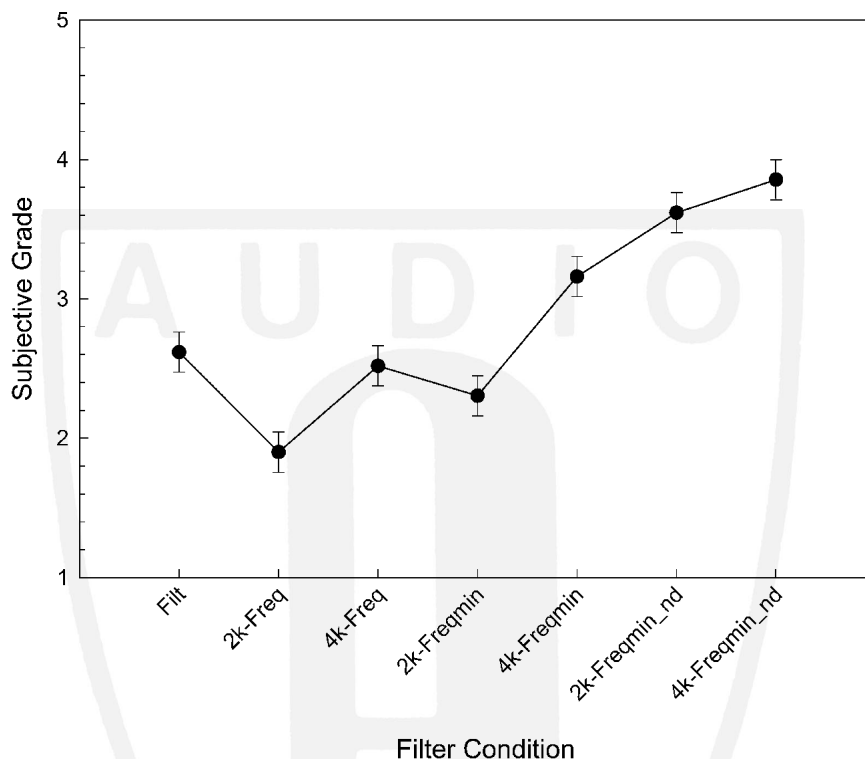


Fig. 22. Overall mean subjective grades versus correction filter condition (both loudspeakers).

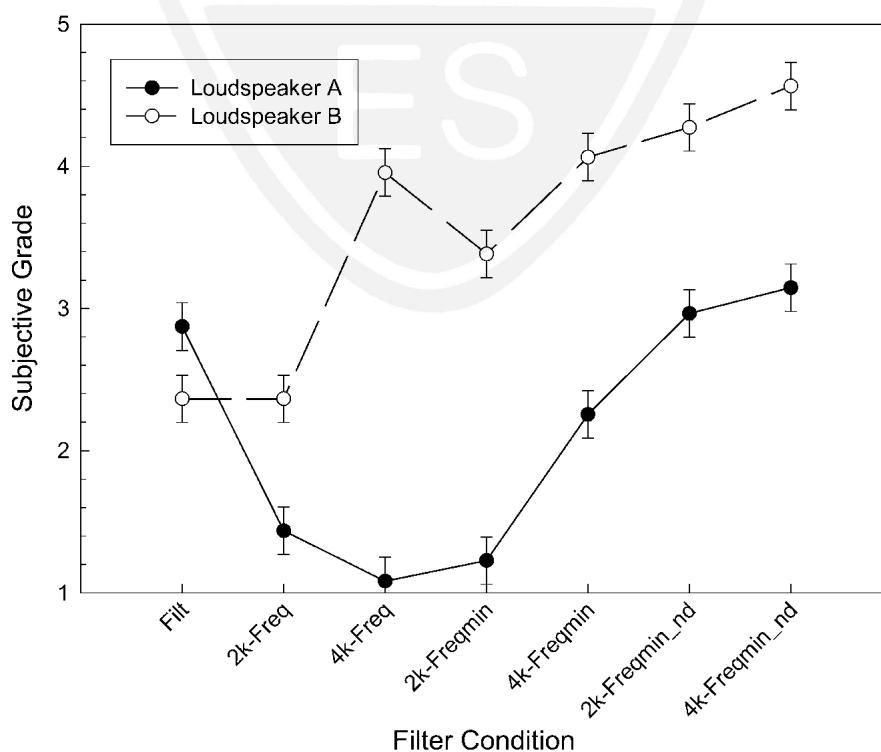


Fig. 23. Mean subjective grades versus correction filter condition for two loudspeakers of experiment 3.

speaker A. This agrees with previous results [17] that the correction filter can actually degrade the subjective performance. Fig. 23 also shows that removing the modeling delay improved the performance statistically for both cases of loudspeaker A, but for only one case of loudspeaker B, where the increase is not statistically significant. [4k-Freqmin (~4.0) to 4k-Freqmin_nd (~4.5)].

Fig. 24(a) shows the original uncorrected magnitude response. The corrected magnitude responses are shown in Fig. 24(b) with modeling delay and in Fig. 24(c) without

modeling delay for the on-axis case of loudspeaker A. The modeling delay correction has more midfrequency ripple than without a modeling delay, which could account for the difference in the subjective grade. The corrected IRs, shown in Fig. 25 show another possible reason for the difference in the subjective grades. Fig. 25(a) is the corrected IR with a modeling delay of 1024 samples being added, which is seen by the delay of the pulse occurring at approximately 23 ms. When adding a delay, artifacts from the block processing can be produced before the pulse and

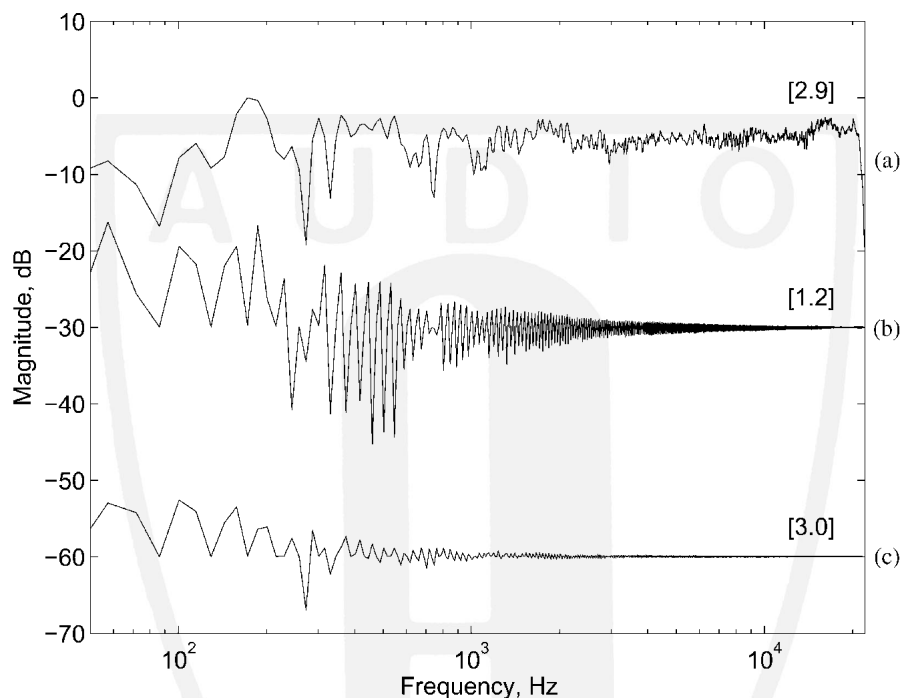


Fig. 24. On-axis magnitude response and corrections for loudspeaker A. (a) Uncorrected response. (b) Corrected with 2k FD filter with modeling delay. (c) Corrected with 2k FD filter without modeling delay. Subjective grades are shown in square brackets.

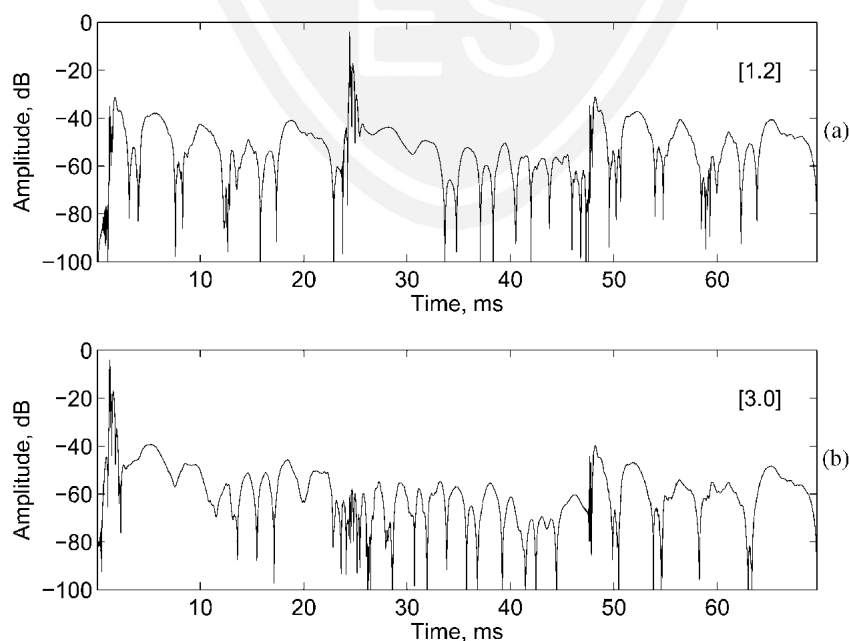


Fig. 25. On-axis corrected time response for loudspeaker A. (a) 2k FD correction with modeling delay. (b) 2k FD correction without modeling delay. Subjective grades are shown in square brackets.

can create preringing effects. Without a modeling delay being added, the preringing does not exist since there is no delay or time for artifacts to be present.

6 EXPERIMENT 4

The frequency-domain calculation of the inverse filter given by Eq. (6) provides an intuitive look at a potential problem with inverse filtering. For every frequency where there is a dip in the filter $C(k)$, the inverse filter $H(k)$ will have a corresponding peak at the same frequency. This can be a problem if the magnitude of the dip is large since the inverse filter will compensate by creating a large peak at that frequency. When correcting loudspeakers, this problem can become especially severe at low frequencies and at frequencies near the Nyquist frequency of the inverse filter. At these extremes, the frequency response of the loudspeaker often rolls off and the resulting excessive boosts in the inverse filter could overload the loudspeaker. Perceptually these excessive narrow-band boosts are undesirable and should therefore be avoided.

One solution to the excessive boosting at certain frequencies is to use regularization in the inversion process. This limits the effort used to correct the IR so that large dips or peaks are not produced in the inverse filter.

In the LS time-domain approach introduced in Section 2.1 regularization adds an effort term to the cost function and is given by

$$J = E + \beta V \quad (8)$$

where E is the performance error term, V is the effort term, and β is a scaling factor used to vary the amount of regularization [2].

The effort term appears in the optimal LS solution as a regularization filter $b(n)$ of length N_b . The convolution matrix can be formed from the sequence $b(n)$ in a similar way as was the convolution matrix C given by Eq. (4). Therefore the convolution matrix B will have the dimensions of N_h columns by $(N_h + N_b - 1)$ rows and is then given by

$$B = \begin{bmatrix} b(0) & & & 0 \\ \vdots & \ddots & & \vdots \\ b(N_b - 1) & & \ddots & b(0) \\ \vdots & & \ddots & \vdots \\ 0 & & & b(N_b - 1) \end{bmatrix}. \quad (9)$$

The LS solution becomes then

$$h = (C^T C + \beta B^T B)^{-1} C^T a_m. \quad (10)$$

The regularization filter appears in the frequency-domain deconvolution method in a similar fashion, and Eq. (6) becomes

$$H(k) = \frac{D(k)C^*(k)}{C(k)C^*(k) + \beta B(k)B^*(k)} \quad (11)$$

as given in [3], [15]. When defining the regularization filter $b(n)$, it is only the energy of the filter that is impor-

tant and not the phase response. In defining a regularization filter, the regularization will have the most effect in the passband of the filter. That is, in the passband it will limit the effort applied by the correction filter. For example, if a high-pass filter were used as the regularization filter, then the regularization process would have its greatest effect in the high frequencies. This is clear when looking at the frequency-domain inverse filter of Eq. (11), where the regularization term $\beta B(k)B^*(k)$ is at a maximum for the passband of the filter $b(n)$ and the regularization will thus have a greater effect. Likewise in the stopband of the regularization filter the regularization term will be near zero and it will have a minimal effect. Therefore in this example the inverse filter would spend less effort correcting dips and peaks in the high frequencies as compared to the low frequencies.

One can also use a frequency-independent (scalar) form of regularization, and in this case, the term $B^T B$ becomes the identity matrix and β scales the amount of regularization used.

In the present study three types of regularization were implemented: a scalar and two vector methods. In the first vector approach the low (80 Hz and below) and high (18 kHz and above) frequencies were regularized to a value of 1, whereas the midfrequencies were regularized by different amounts. For the second vector approach the regularization was set relative to the one-third-octave spectrum of the IR, as in [15]. Fig 26 shows the three regularization types that were implemented.

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 between 10^{-5} and 7×10^{-1} were used for both the scalar and the vector types of regularization. All of the inverse filters were 2048 samples in length, and were calculated using Eq. (11).

Preliminary subjective experiments were carried out with various levels of scalar and vector regularization using the five measured IRs in order to explore the ranges of β that would be suitable for a larger formal subjective test.

6.1 Preliminary Tests

Due to the large number of combinations of IRs, regularization types, and levels, preliminary subjective tests were conducted (with only three subjects) to explore the effects of regularization and narrow down the number of test cases for the formal subjective test.

An ANOVA was performed on the subjective test data, and the results for one part of the preliminary tests are shown in Fig 27. This preliminary test used the first vector-based regularization method in computing the inverse filter to correct the off-axis (tweeter side) response of loudspeaker A.

The results of Fig 27 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 (that is, $\beta = 0$).

There is a range of regularization values ($\beta = 10^{-5}$ to 5×10^{-3}), which is acceptable in that it gives the same audio quality as with no regularization. The plot also

shows that using higher values of regularization ($\beta \geq 0.1$) actually degrades the audio quality to the point where using an inverse filter is worse than having no correction at all. Therefore great care must be taken in setting the amount of regularization. A similar result was also found with the scalar regularization, although the range of acceptable values was even smaller.

Fig. 28 shows the corrected off-axis (tweeter side) time response, (a) without regularization and (b) with the

simple vector form, $\beta = 10^{-3}$. As was seen before, it is not a perfect delta function but rather a pulse with a significant amount of energy arriving both before and after the pulse. If this residual energy is too high in level or is far enough away from the delta function, then it will cause audible artifacts.

It can be seen from Fig. 28(b) that the effect of regularization is to push the residual energy closer to the main pulse, which helps to reduce the audibility of echoes.

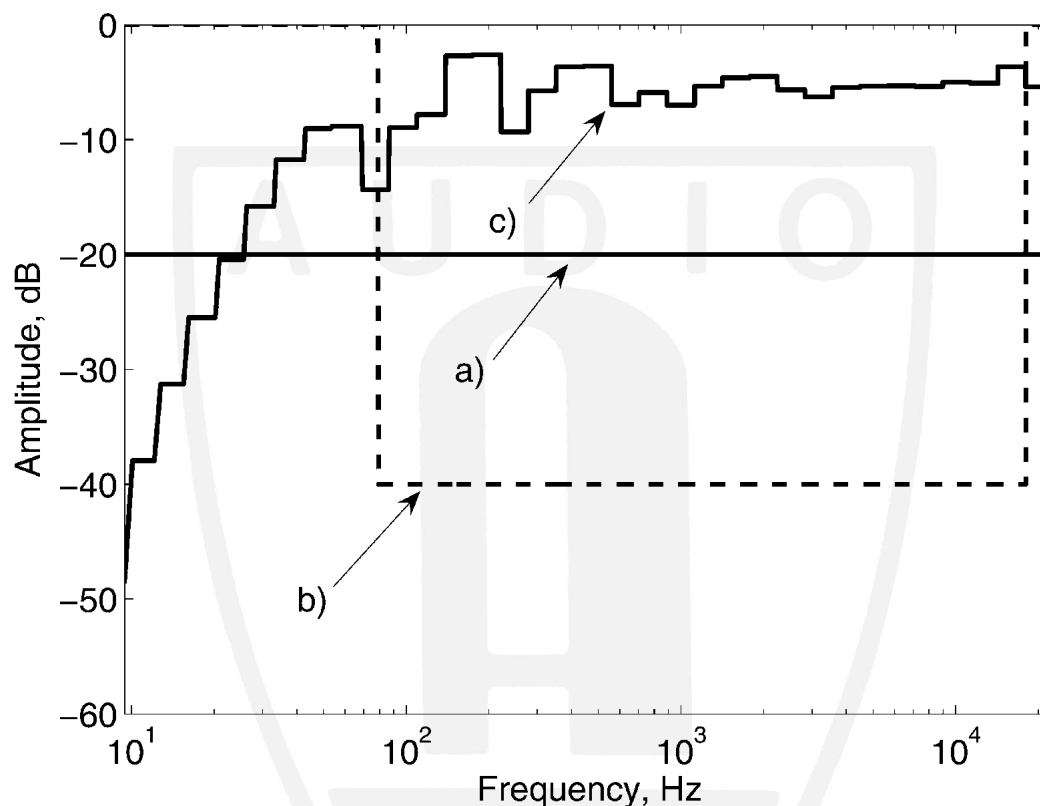


Fig. 26. Three regularization types implemented. (a) Scalar. (b) Vector type 1. (c) One-third-octave vector method.

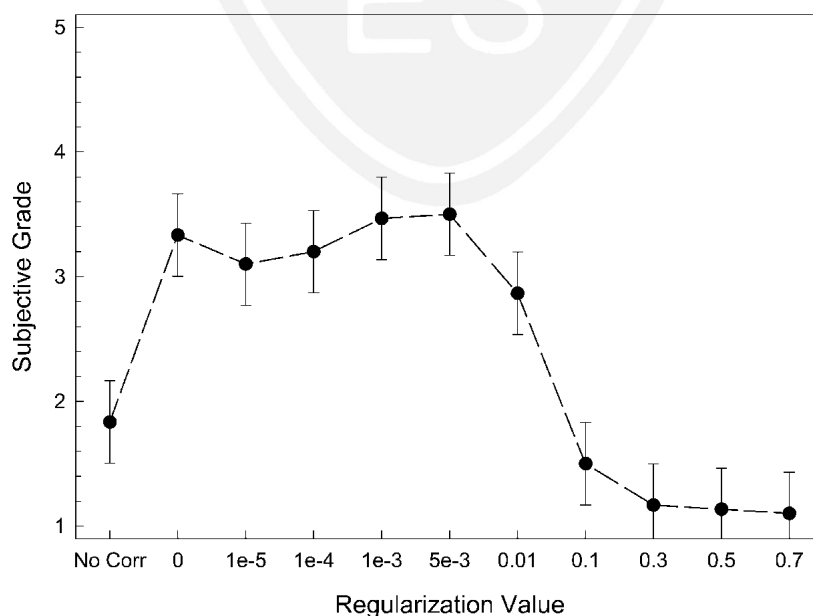


Fig. 27. Mean subjective grades versus amount of regularization used for off-axis (tweeter side) correction for loudspeaker A. First vector-based regularization was used in inverse-filter calculation.

However, by doing so the regularization causes a widening of the pulse or delta function, which in turn creates audible artifacts that are similar to the pre-echo artifacts found in some perceptual coders.

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

6.2 Formal Test

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 levels of regularization for each of the three regularization types were also selected based on the findings of the preliminary tests. These were selected to demonstrate the range of performance of the regularization with the two different IRs. For comparison, the uncorrected filtered versions and a full correction with no regularization 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 Fig. 29, with the error bars once again representing 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 .

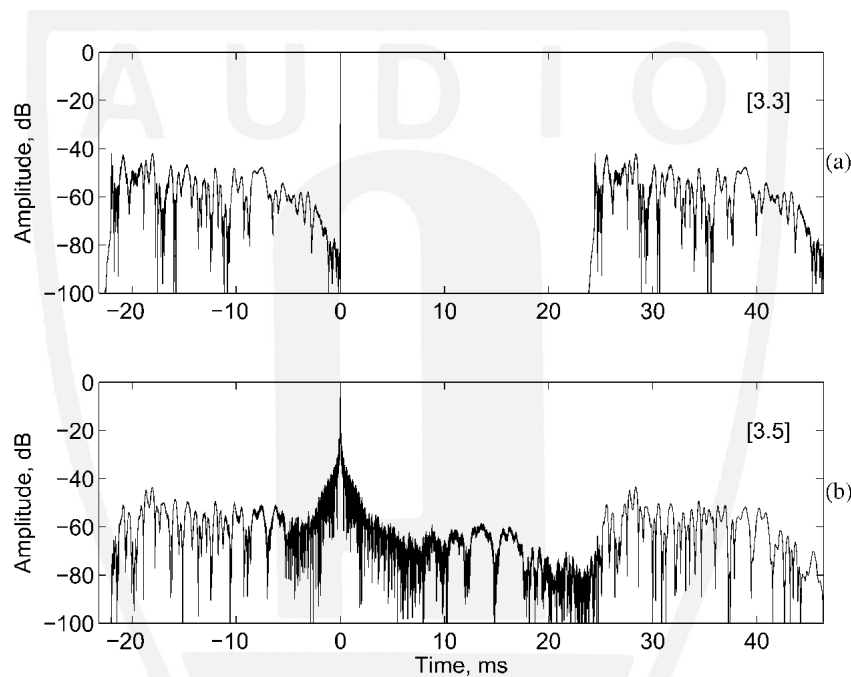


Fig. 28. Off-axis (tweeter side) correction for loudspeaker A. (a) 2k frequency-domain inverse with no regularization. (b) 2k frequency-domain inverse using vector regularization type 1 with $\beta = 10^{-3}$. Subjective grades are shown in square brackets.

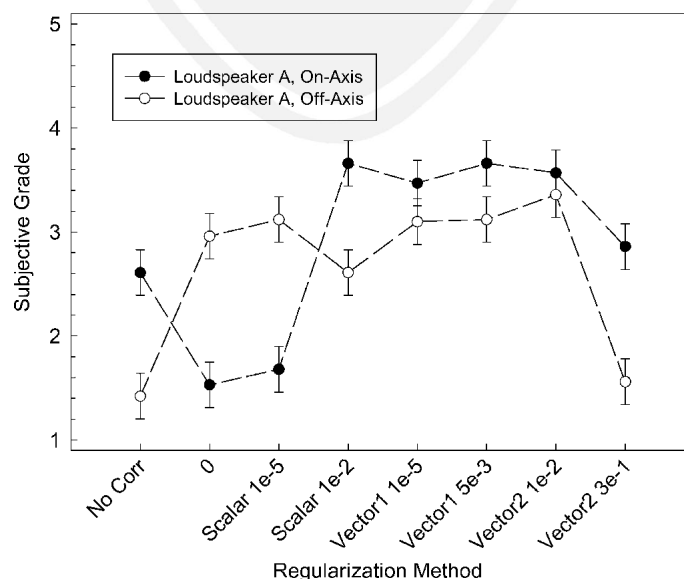


Fig. 29. Mean subjective grade versus regularization type and amount for loudspeaker A, on/off-axis correction.

It can be seen that for the on-axis response, frequency-domain inverse filtering without regularization actually degrades the audio quality. This confirms our earlier findings in [17]. Adding a very small amount of scalar regularization (10^{-5}) does not provide a significant increase in performance. However, increasing the amount of scalar regularization to 10^{-2} provides a dramatic improvement in the subjective performance of the inverse filter. For the off-axis response, however, the opposite is true. Very good performance is achieved without regularization, and a scalar regularization of 10^{-2} degrades the performance of the inverse filter slightly. Thus the optimal value of β for scalar regularization depends on the characteristics of the IR being inverted.

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. These wrapping effects are reduced when a small regularization term is added.

The first vector-based 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 result of the preliminary subjective tests. As such, these values of regularization are the result of a hand-tuning process. Nonetheless, the first vector-based approach provided more robust results than the scalar method.

The performance of the second vector-based (one-third-octave) regularization method was not as robust. Smaller amounts of regularization (10^{-2}) gave good subjective results for both the on-axis and the 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 deconvolution with a length of 2k samples. Fig. 30(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 delta function. Since this residual energy is relatively far away from the delta function (20 ms before and 50 ms after), it is readily audible as time-domain artifacts.

Fig. 30(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 Fig. 30(a). It can be seen that in this case the residual energy is always below 50 dB.

Fig. 30(c) shows the corrected response using the first vector-based 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.

Kirkeby et al. [3] showed the effect of the 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

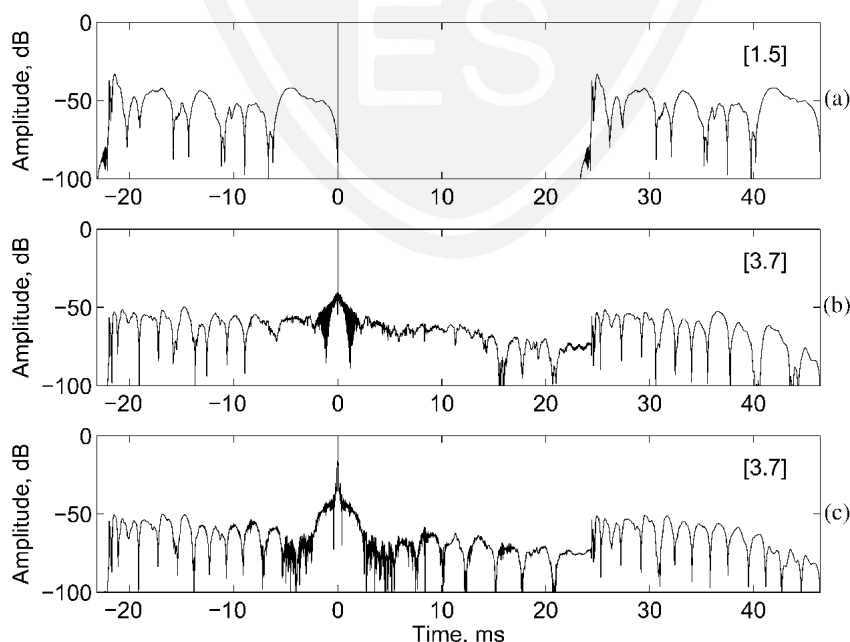


Fig. 30. On-axis correction for loudspeaker A. (a) 2k frequency-domain inverse with no regularization. (b) 2k frequency-domain inverse using scalar regularization, $\beta = 10^{-2}$. (c) 2k frequency-domain inverse using vector regularization type 1 with $\beta = 5 \times 10^{-3}$. Subjective grades are shown in square brackets.

Fielder [15], 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 preringing effects.

7 EXPERIMENT 5

Another way of controlling how much the inverse filter has to work is to smooth the transfer function, to reduce the severity of the peaks and dips. Traditional spectral smoothing operations only use the power spectra, and no method of identifying the smoothed phase is specified, which makes the process nonreversible with respect to recovering a “smoothed” impulse response. In some cases the time responses are derived from the smoothed magnitude spectrum and the zero-phase component [18]. Let $H(k)$ be a loudspeaker or room transfer function, where k is the discrete frequency index ($0 \leq k \leq N-1$). The smoothing operation can be described as a circular convolution,

$$C_{ts}(k) = \sum_{i=0}^{N-1} |C[(k-i) \bmod N]|^2 \cdot W_{sm}(m, i) \quad (12)$$

where $C_{ts}(k)$ is the traditional smoothed response of $C(k)$ and $W_{sm}(m, k)$ is a zero-phase spectral smoothing window function. The windowing function has the shape of a low-pass filter with the sample index m corresponding to the cutoff frequency f_c .

To overcome the uncertainty with the phase, complex smoothing has been suggested by Hatziantoniou and Mourjopoulos [7]. It also has been suggested as a method to overcome some of the problems with inverse filtering, such as long inverse filter lengths and position dependence of the IR being inverted.

The complex smoothing as defined in [19] is a more generalized version of the power spectra smoothing shown in Eq. (12), in that it involves a convolution of the complex frequency response C with a weighting function W , and is given by

$$C_{cs}(k) = \sum_{i=0}^{N-1} C[(k-i) \bmod N] \cdot W_{sm}(m, i) \quad (13)$$

The discrete variable m is a function of k , and $m(k)$ can be considered a bandwidth function so that a fractional octave or other nonuniform frequency smoothing can be achieved. The approach used in this paper is detailed by Hatziantoniou and Mourjopoulos [19].

In previous experiments regularization was used to reduce the amount of work done by the inverse filter and was shown to help reduce some of the block effects associated with the frequency deconvolution method and could be used to improve the subjective performance. In this experiment the effect of complex smoothing of the transfer function before calculating the inverse was studied. One-third-octave complex smoothing, as given by Eq. (13), was performed on the loudspeaker IR, prior to being inverted using the frequency-domain deconvolution

method. The smoothing or weighting function $W_{sm}(m, k)$ is given by

$$W_{sm}(m, k) = \begin{cases} \frac{b - (b-1)\cos[(\pi/m)k]}{2b(m+1) - 1}, & k = 0, 1, \dots, m \\ \frac{b - (b-1)\cos[(\pi/m)(k-N)]}{2b(m+1) - 1}, & k = N-m, N-(m-1), \dots, N-1 \\ 0, & k = m+1, \dots, N-(m+1) \end{cases} \quad (14)$$

where m , in samples, is a function of k to provide nonuniform smoothing and b determines the rolloff of the smoothing window. For example, $b = 1$ would be a rectangular frequency smoothing function. In this experiment b was selected to be 0.7 to provide a smoother transition between frequency bands.

Fig. 31 shows the original on-axis IR for loudspeaker A and a one-third-octave complex smoothed version. It is easily seen that the smoothing reduces the variation of the IR more and more as the time increases.

Fig. 32 shows the on-axis magnitude and phase response for loudspeaker A and the one-third-octave complex smoothed version of each. As with the IR the effect of the smoothing is quite apparent.

A subjective test was conducted to evaluate the performance of the complex smoothing operation. Inverse filters employing complex smoothing were compared directly with nonsmoothed versions. Four IRs were used in this experiment. They included the on-axis and off-axis (woofer side) IRs of loudspeaker A and both IRs of loudspeaker B. Therefore a total of 12 filtering scenarios were evaluated, three inverse-filtering methods for each of the four loudspeaker IRs. This included an uncorrected version and two corrected items with a 2k frequency-domain deconvolution inverse filter. 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 of experiment 5 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 Fig. 33 where, Filter indicates the situation with no correction, 2k-Freq means a 2k frequency deconvolution inverse was used, and 2k-Freq-Sm indicates a 2k frequency deconvolution inverse from the complex smoothed IR was used.

From Fig. 33 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 Fig. 34. It can be seen that the loudspeaker corrections tended to perform better with the smoothed correction. 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. It is interesting to note that for the on-axis response of loudspeaker A, the use of a complex smoothed frequency-domain inverse filter appears to degrade the audio quality somewhat. While this degradation

was not statistically significant in this test, it does suggest that inverse filtering using complex smoothing may suffer from problems similar to those shown for regularization.

Fig. 35 shows the corrected off-axis IRs for loudspeaker B, without and with complex smoothing. The main pulse

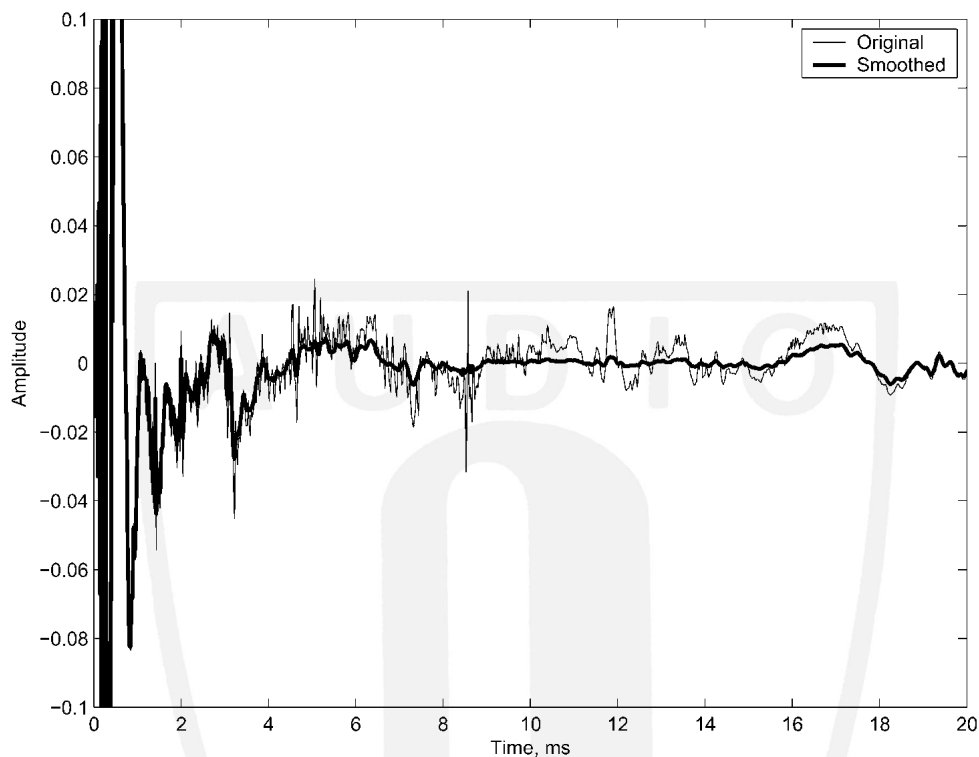


Fig. 31. IR of loudspeaker A (on-axis) showing original and one-third-octave complex smoothed version.

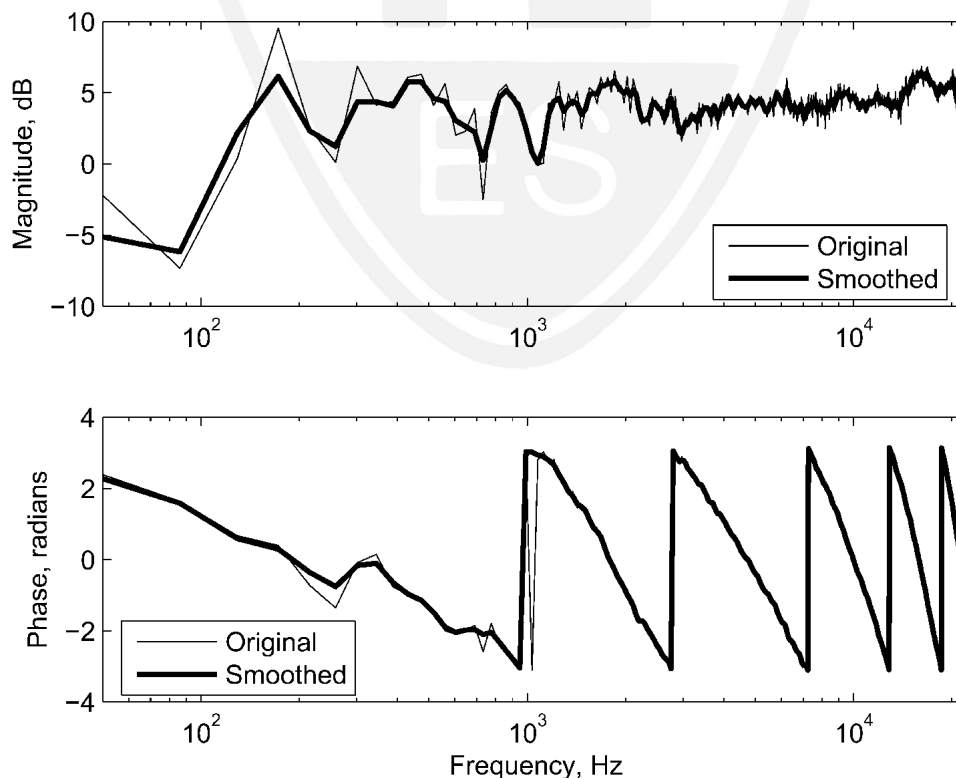


Fig. 32. Magnitude and phase of frequency response of loudspeaker A (on-axis) showing original and one-third-octave complex smoothed version.

occurring at about 23 ms is still very narrow, as opposed to the regularization case shown in Fig. 30, in which it was broadened. The blocking artifacts that are present near 0 ms and 47 ms are lowered by 10 dB in the 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.

8 CONCLUSIONS

Two methods of inverse filtering—time-domain least-squares and frequency deconvolution—were presented.

A subjective test strategy was devised to formally evaluate the performance of the inverse-filtering methods. The test strategy not only allowed the listener to compare the different inverse-filtering methods but also let them compare the methods to the case where no correction was used. This provided a way to see whether the inverse filtering provided any improvement or whether it actually degraded the audio signal. The formal subjective tests were conducted using the MUSHRA-based method.

In the first experiment the two methods of calculating the inverse of an IR were compared. The time-domain

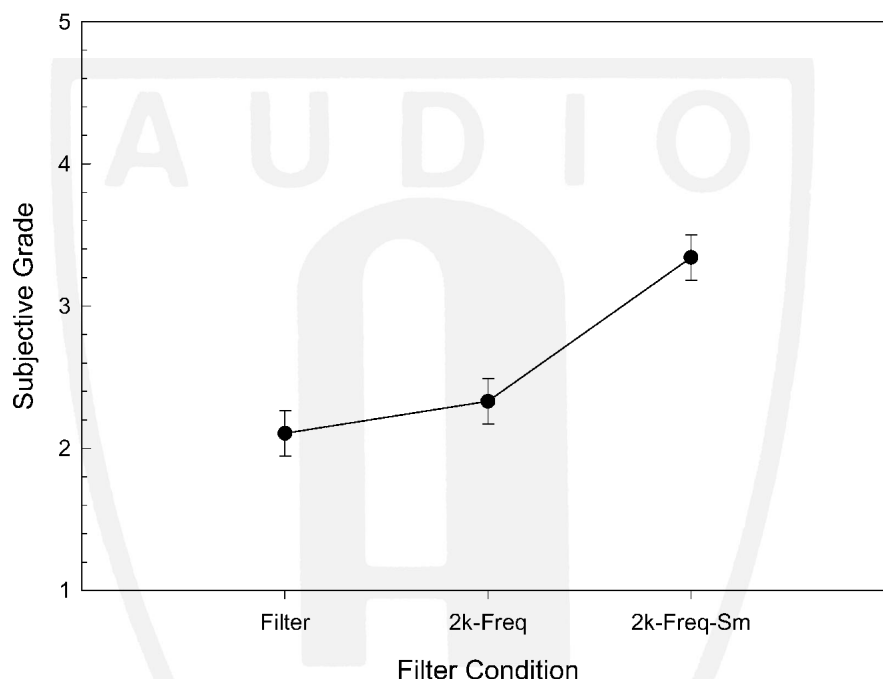


Fig. 33. Mean subjective grades versus correction filter conditions for combination of all loudspeakers in experiment 5.

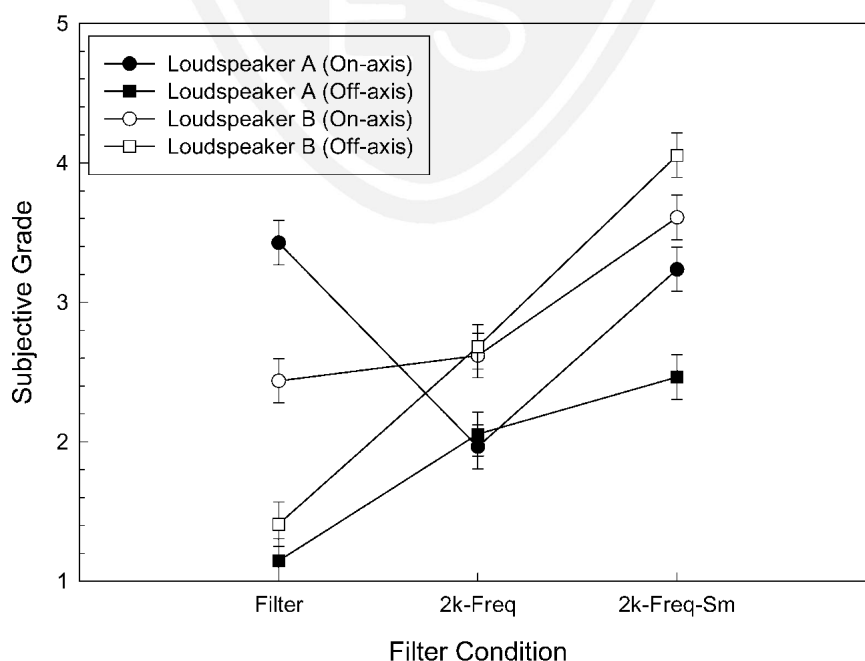


Fig. 34. Mean subjective grades versus correction filter conditions for four loudspeakers in experiment 5.

LS approach proved to be more robust than the frequency-domain deconvolution method in identical conditions. It was also shown that the subjective performance was highly dependent on the IR being inverted and in some cases, most notably the frequency-domain method, the correction filter degraded the audio quality compared to when no correction filter was added. Correcting the off-axis response with an inverse-filter calculated from an on-axis IR did not provide any improvement over the case with no correction being done. This was true for both loudspeakers. The frequency-domain method suffered from wrapping effects due to the block processing nature of the method. This produced time-domain artifacts (low-level delays) that were clearly audible.

In the second experiment frequency-domain correction filters of longer lengths were created and evaluated. Inverse filters with lengths up to 32 768 samples were created, and it was shown that by increasing the length of the inverse filter, the frequency-domain method performed subjectively better. However, the subjective performance was still dependent on the IR being inverted. This was quite apparent with one IR (on-axis IR for loudspeaker A), which only showed some subjective improvement when increasing the length of the inverse filter to 32k. The wrapping artifacts produced by the frequency-domain method were pushed out farther in time from the Dirac pulse, but were attenuated in level. This resulted in a better subjective performance. Therefore creating longer inverse filters would reduce the blocking effects, but using such long filters has practical limitations such as being more computationally demanding.

A modeling delay is used to make a causal inverse filter when the IR being inverted is not minimum phase. When only the minimum-phase component of the IR was used in calculating the inverse filter, the first experiment showed

that this situation did not perform very well. One possible reason was the use of the modeling delay. In this situation, since only the minimum-phase component is being used in the inversion process, no modeling delay is required to create a causal inverse filter. Experiment 3 showed that when inverting a minimum-phase IR, a modeling delay actually degraded the subjective performance of the correction filter.

To overcome some of the shortcomings of inverse filtering in the frequency-domain such as the blocking and wrapping effects, various forms of regularization were used in the creation of the inverse filters. Varying amounts of frequency-independent and -dependent forms of regularization were used with the frequency-domain method when calculating the inverse filters. The regularization did improve the subjective performance over the case when no regularization was used, but the amount of regularization needed was very dependent on the loudspeaker IR being corrected. Applying the incorrect amount of regularization actually degraded the audio quality, and resulted in a subjective performance poorer than if no regularization were used.

Complex smoothing was also implemented to improve the subjective performance of the frequency-domain method. With complex smoothing one creates a smoothed version of the transfer function and the phase is well defined, as opposed to traditional smoothing where there is no specific formula for the phase.

One-third-octave complex smoothing was carried out on the transfer function prior to calculating the inverse filters. It was shown that complex smoothing improved the subjective performance of the inverse filters in most cases, and never degraded their performance. The corrected responses also showed a very narrow Dirac pulse as compared to regularization, which tended to broaden the pulse.

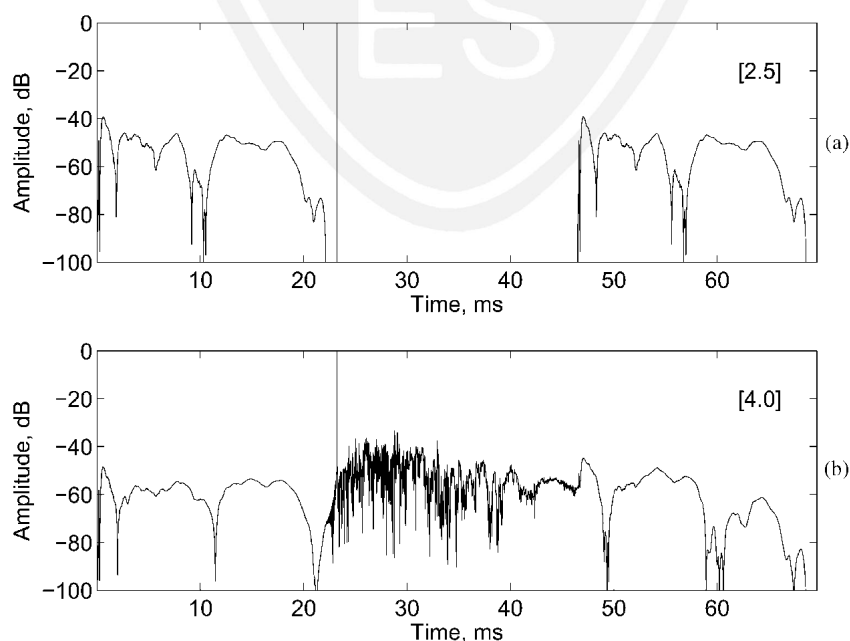


Fig. 35. Off-axis corrected IR for loudspeaker B. (a) 2k FD correction. (b) 2k FD correction from complex smoothed IR. Subjective grades are shown in square brackets.

9 ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

10 REFERENCES

- [1] J. N. Mourjopoulos, "Digital Equalization of Room Acoustics," *J. Audio Eng. Soc.*, vol. 42, pp. 884–900 (1994 Nov.).
- [2] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," *J. Audio Eng. Soc.*, vol. 47, pp. 583–595 (1999 July/Aug.).
- [3] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña-Bustamante, "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Trans. Speech and Audio Process.*, vol. 6, pp. 189–194 (1998 Mar.).
- [4] P. G. Craven and M. A. Gerzon, "Practical Adaptive Room and Loudspeaker Equaliser for Hi-Fi Use," presented at the AES UK DSP Conference (1992 Sept.).
- [5] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
- [6] S. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *J. Acoust. Soc. Am.*, vol. 66, pp. 165–169 (1979 July).
- [7] P. D. Hatziantoniou and J. N. Mourjopoulos, "Results for Room Acoustics Equalization Based on Smoothed Responses," presented at the 114th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 51, p. 422 (2003 May), convention paper 5779.
- [8] P. A. Nelson, F. Orduña-Bustamante, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Trans. Speech and Audio Process.*, vol. 3, pp. 185–192 (1995 May).
- [9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, MD, 1996).
- [10] D. D. Rife and J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences," *J. Audio Eng. Soc.*, vol. 37, pp. 419–444 (1989 June).
- [11] ITU-R BS.1534, "Method for the Subjective Assessment of Intermediate Audio Quality," Geneva, Switzerland (2001).
- [12] ITU-R BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," Geneva, Switzerland (1994).
- [13] J. L. Bruning and B. L. Kintz, *Computational Handbook of Statistics* (Addison Wesley Longman, Reading, MA, 1997).
- [14] G. Keppel and S. Zedeck, *Data Analysis for Research Designs* (W. H. Freeman, New York, 1989).
- [15] L. D. Fielder, "Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization," *J. Audio Eng. Soc.*, vol. 51, pp. 3–26 (2003 Jan./Feb.).
- [16] P. M. Clarkson, J. Mourjopoulos, and J. K. Hammond, "Spectral, Phase, and Transient Equalization for Audio Systems," *J. Audio Eng. Soc.*, vol. 33, pp. 127–132 (1985 Mar.).
- [17] 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. Audio Eng. Soc. (Abstracts)*, vol. 50, p. 961 (2002 Nov.), convention paper 5662.
- [18] S. P. Lipshitz, T. C. Scott, and J. Vanderkooy, "Increasing the Audio Measurement Capability of FFT Analyzers by Microcomputer Postprocessing," *J. Audio Eng. Soc.*, vol. 33, pp. 626–648 (1985 Sept.).
- [19] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses," *J. Audio Eng. Soc.*, vol. 48, pp. 259–280 (2000 Apr.).

THE AUTHORS



S. G. Norcross



G. A. Soulodre



M. C. Lavoie

Scott G. Norcross received a B.Sc. degree in physics from McGill University, Montreal, Quebec, Canada, in 1993. He joined the Audio Research Group at the University of Waterloo, Canada, and received an M.Sc. degree in physics in 1996, under the supervision of Professors John Vanderkooy and Stanley Lipshitz. His thesis was on "The Effects of Nonlinearity on Impulse Response Measurements."

Mr. Norcross then spent the next year at the American University in Washington, DC, where he taught acoustics, electronics, and audio technology for the Audio Technology Program in the Department of Physics. In 1997 he joined the acoustics group at the National Research Council (NRC) of Canada in Ottawa, under the supervision of Dr. John S. Bradley. There he worked on acoustical measurement systems for concert halls, rooms, airplanes, and

offices; and designed and conducted subjective tests on speech intelligibility and open office acoustics. He is now a research engineer in the Advanced Audio Systems Group at the Communications Research Centre in Ottawa, working under the supervision of Dr. Gilbert Soulodre, doing research on DSP techniques and subjective aspects of multichannel audio and inverse filtering for room and loudspeaker equalization. The latter is the research area for his Ph.D. degree in electrical engineering at the University of Ottawa, which he is currently working on under the supervision of Prof. Martin Bouchard.

●

Gilbert A. Soulodre received B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba in Winnipeg, Canada. In 1987 he joined the Audio and Acoustics department of Bell-Northern Research (now Nortel) as a member of the scientific staff. There he was involved in the development of digital audio systems for telecommunications. In 1990 he began research on the development of adaptive DSP algorithms for removing noise from audio signals for his Ph.D. degree. From 1991 to 1994 he was an assistant professor in the Graduate Program in Sound Recording at McGill University, Montreal, Quebec, Canada.

Dr. Soulodre is currently a researcher with the Advanced Audio Systems Group at the Communications Research Centre, Ottawa, Canada. There his main focus is in the areas of audio processing and sound perception. He participates in the ITU-R audio standards committees and was heavily involved in the development of the BS.1116 and MUSHRA standards for subjective testing. He was also a research adjunct professor of psychology at Carleton University, where he examined the subjective components of sound fields in concert halls and multichannel surround systems.

In 1996, Dr. Soulodre was recognized for his work on spatial impression and listener envelopment by the Acoustical Society of America and the American Institute of Physics. He is a fellow of the AES.

●

Michael C. Lavoie received a bachelor's degree in electrical engineering from the University of Manitoba, Canada, in 1984. For the next two years he worked for the Canadian Broadcasting Corporation, where he occupied various positions in radio and television production. From 1986 until 1995 he worked as an independent contractor in live and recorded audio production. In 1995 he joined the Signal Processing and Psychoacoustics Group at the Communications Research Centre, Ottawa, Canada, where he conducts research in subjective testing.