Multiple-point statistical room correction for audio reproduction: Minimum mean squared error correction filtering

Fredrik Lingvall[a] and Lars-Johan Brännmark[b]
Signals and Systems Group, Department of Engineering Sciences, Uppsala University, P.O. Box 528, SE-751 20 Uppsala, Sweden

(Received 17 October 2008; accepted 30 December 2008)

This paper treats the problem of correction of loudspeaker and room responses using a single source. The objective is to obtain a linear correction filter, which is robust with respect to listener movement within a predefined region-of-interest. The correction filter is based on estimated impulse responses, obtained at several positions, and a linear minimum mean squared error criteria. The impulse responses are estimated using a Bayesian approach that takes both model errors and measurement noise into account, which results in reliable impulse response estimates and a measure of the estimation errors. The correction filter is then constructed by using information from both the estimated impulse response coefficients and their associated estimation errors. Furthermore, in the optimization criteria a time-dependent reflection filter is introduced, which attenuates the high frequency parts of the reflected responses, that is, the parts of the responses that cannot be compensated with a single source system. The resulting correction filter is shown to significantly improve both the temporal and spectral properties of the responses compared to the uncorrected system, and, furthermore, the obtained correction filter has a low level of pre-ringing.
 [DOI: 10.1121/1.3075615]

PACS number(s): 43.60.Uv, 43.55.Br, 43.60.Pt [EJS] Pages: 2121–2128

I. INTRODUCTION

The aim of digital room correction is, in general, to improve audio reproduction in, for example, domestic or automotive sound systems. Improve means here to compensate, or at least reduce, the distortions of the signal due to both a non-perfect source and acoustic diffraction and room effects. In digital room correction this is normally accomplished by first acquiring data to estimate the room impulse responses and then construct a correction filter based on the estimated impulse responses.[1]

It is known that direct inversion of acoustic impulse responses acquired at a single listening point results in excessive boost at some frequencies due to interference induced deep spectral notches in the impulse responses, which results in high noise amplification,[2,3] and, furthermore, the correction performance quickly deteriorates when the listener moves away from the reference correction point.[4,5] In this area of application, spatial robustness is essential since the position of the listener cannot be expected to be fixed. In the literature, several approaches to robust design of correction filters have been studied. Examples of such approaches include complex smoothing,[6] multiple-point equalization of common poles,[7] derivative constrained equalization,[8] spatial averaging with fuzzy clustering of room responses,[9] time-domain inversion with frequency-dependent regularization,[2] and robust multiple-point designs.[4,10,11]

Another important issue in correction filter design is to try to avoid non-causal preresponses of the correction filter; non-causal means here the filter effects that occur before the main correction filter delay. Allowing a delay in the system normally improves the correction performance,[12] but it also implies a mixed phase character of the correction filters. It is known that mixed phase designs generate a pre-response in the filter, which may generate an audible pre-ringing in the corrected system.[3,13]

This paper is concerned with compensation of loudspeaker and room effects, using a single source, by means of deconvolution of acoustic impulse responses based on data measured at more than one listening point. The objective is to obtain correction filters, with a low amount of pre-ringing, that are robust in the sense that the position of the listener is allowed to vary within a predefined region-of-interest while still obtaining a good, or at least reasonable, compensation for room and loudspeaker characteristics. The reason for choosing such an objective is that we must realize that the one cannot perfectly reproduce the input signal at every point in a region-of-interest. This is mainly for two reasons: (i) we are using a single source so we cannot control (beamform) the sound field spatially and (ii) we do not know the impulse responses at every point in advance and we must, therefore, estimate them based on data acquired at a finite set of observation points. Hence, we only have knowledge of the characteristics of the sound reproduction system at some points, and, furthermore, the estimated impulse responses have a finite length and are based on noisy measurements resulting in limited accuracy of the impulse response estimates.

In this paper we will discuss a statistical inferential[14] method for estimating acoustic impulse responses as well as a method to design robust compensation filters based on (i)

[a]Electronic mail: fredrik.lingvall@gmail.com
[b]Electronic mail: lars-johan.brannmark@angstrom.uu.se

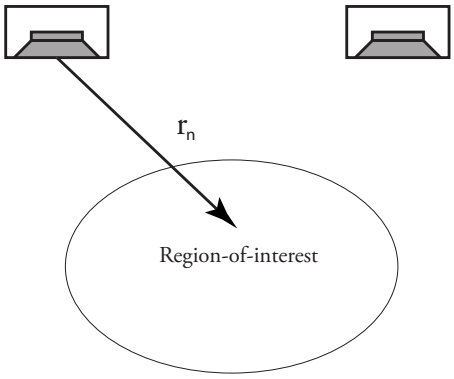

FIG. 1. Typical measurement setup.

impulse responses estimated from data acquired at several observation points and (ii) a linear minimum mean squared criteria. Note that we are seeking a linear finite impulse response filter that can be used for real time compensation.

This paper is organized as follows: in Sec. II a linear matrix-based impulse response model is introduced, which then is used in the estimation of the acoustic impulse responses, which is discussed in Sec. III. Furthermore, In Sec. IV the robust multiple-point design method for correction filters are described as well as how room reflections can be treated by means of a non-stationary filtering approach. In Sec. V the experiments are discussed, and, finally, in Sec. VI a discussion and the conclusions are given.

II. A LINEAR IMPULSE RESPONSE MODEL

Consider the loudspeaker setup shown in Fig. 1. We measure the acoustic pressure wave form at an observation point $\mathbf{r}_n=(x_n,y_n,z_n)$ using a microphone. If we let the loudspeaker be located at origo, $\mathbf{r}=(0,0,0)$, then we can express the received signal at the nth position as

$$y(\mathbf{r}_n,t)=h(\mathbf{r}_n,t)*u(t)+e(t), \tag{1}$$

where $*$ denotes temporal convolution. Note that the impulse response $h(\mathbf{r}_n,t)$ is a function of position due to diffraction effects of the loudspeaker and room effects, such as reflections and room modes; the impulse response, $h(\mathbf{r}_n,t)$, is comprised of several parts, such as the amplifier response, the electro-acoustical response of the loudspeaker, the acoustic (diffraction) response of the loudspeaker, the room impulse response, and the response of the microphone.

Here we are concerned with digital processing of the acoustic data; hence we need a discrete model of the system. This can readily be obtained by sampling the impulse responses, input signals, and the acquired data; here we represent, by convention, the data $\mathbf{y}_n$, the input signal $\mathbf{u}$, and the impulse responses $\mathbf{h}_n$ by column vectors. The discrete form of Eq. (1) can then be expressed as

$$\mathbf{y}_n=\mathbf{h}_n*\mathbf{u}+\mathbf{e}, \tag{2}$$

where $*$ now denotes discrete temporal convolution. Furthermore, a discrete-time convolution of two vectors $\mathbf{u}$ and $\mathbf{h}$ can be expressed as a matrix-vector multiplication of the form

$$\mathbf{h}*\mathbf{u}=\mathbf{H}\mathbf{u}=\mathbf{u}*\mathbf{h}=\mathbf{U}\mathbf{h}, \tag{3}$$

where the matrices $\mathbf{H}$ and $\mathbf{U}$ have a Toeplitz form consisting of delayed versions of $\mathbf{h}_n$ and $\mathbf{u}$, respectively.[12,15] If $\mathbf{h}_n$ is of length L_h and $\mathbf{u}$ is of length L_u, then the length of the discrete-time convolution, $\mathbf{h}_n*\mathbf{u}$, will be L_h+L_u-1. We will assume, for the remaining part of the paper, that when a matrix convolution, of the form (3), is used then the Toeplitz matrices involved have a dimension such that the length of the resulting vector is the same as for the normal form of the discrete-time convolution. Now, by using the matrix-vector notation in Eq. (3), the model (2) becomes

$$\mathbf{y}_n=\mathbf{H}_n\mathbf{u}+\mathbf{e}=\mathbf{U}\mathbf{h}_n+\mathbf{e}. \tag{4}$$

III. ESTIMATING ACOUSTIC IMPULSE RESPONSES

Before we can design the correction filter, we need to obtain the impulse responses, $\mathbf{h}_n$, for each observation point $\mathbf{r}_n$. To do this we play a suitable identification signal to the system and measure the acoustic waveform using a microphone at $\mathbf{r}_n$, as discussed in Sec. II. We must, however, acknowledge that the measurements are never error free; we will always have quantization errors, thermal noise, acoustic disturbances, etc. Also, we are using a discrete model of finite length, which never can describe the system perfectly. We have, therefore, chosen a Bayesian approach to estimate the impulse responses, which takes these uncertainties into account. More precisely, we have used the optimal linear estimator—also known as the linear minimum mean squared error (LMMSE) estimator or the maximum *a posteriori* estimator under Gaussian assumptions.

A. Prior information

A powerful feature in Bayesian analysis is that prior information is naturally incorporated, which significantly can improve our estimates. In this case we are interested in estimating acoustic impulse responses, and we know, for example, *a priori* that it takes a certain amount of time for the sound to travel from the active area of the loudspeaker to the observation point. The impulse response must consequently have a zero amplitude for that time interval, and data for this time interval can, therefore, be removed since it carries no information relevant to the design of the correction filter. For the response corresponding to the remaining time interval, which consists of a direct response and room responses (reflections and room modes), the direct response is often the strongest. However, if we have no knowledge from, for example, previous measurements, then it is difficult to say how much stronger the direct response is compared to the responses due to the room. Furthermore, we normally do not know if the impulse response amplitudes are more likely to be positive or negative. We can, however, say something regarding the mean signal variation since, in a typical measurement situation, the output power and the microphone gain is adjusted so that a sufficient acquired signal variation is obtained. This mean signal variation will correspond to a mean variation in the impulse responses, which we denote as

σ_h. Based on probability theory,[16,17] we can then assign a zero-mean Gaussian (prior) probability density function (pdf) to the impulse responses

$$p(\mathbf{h}_n|I) = \frac{1}{(2\pi)^{L_h/2}|\mathbf{C}_h|^{1/2}} e^{-(1/2)\mathbf{h}_n^T \mathbf{C}_h^{-1} \mathbf{h}_n}, \tag{5}$$

where $\mathbf{C}_h = \sigma_h^2 \mathbf{I} \forall n$.

B. The optimal linear estimator for acoustic impulse responses

To estimate the impulse responses $\mathbf{h}_n$, we first perform N experiments, one for each observation point, where we obtain the data $\mathbf{y}_n$, $n=1,2,\ldots,N$. Using the linear model (4), for the nth measurement point, the optimal linear estimate of $\mathbf{h}_n$ can be found by minimizing the mean squared error (MSE) criteria,

$$J(\mathbf{K}) = \arg\min_{\mathbf{K}} E\{\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|^2\} = \arg\min_{\mathbf{K}} E\{\|\mathbf{h}_n - \mathbf{K}\mathbf{y}_n\|^2\}, \tag{6}$$

which has the closed form solution given by[18]

$$\hat{\mathbf{h}}_n = \mathbf{K}\mathbf{y}_n = \mathbf{C}_h\mathbf{U}^T(\mathbf{U}\mathbf{C}_h\mathbf{U}^T + \mathbf{C}_e)^{-1}\mathbf{y}_n, \tag{7a}$$

where $E\{\cdot\}$ is the expectation operator, $\mathbf{C}_e$ is the noise covariance matrix, $\mathbf{C}_h$ is the impulse response covariance matrix given by Eq. (5), and T is the transpose operator.

The covariance matrix $\mathbf{C}_h$ allows us to incorporate *a priori* knowledge regarding the impulse responses. A conservative setting is, as discussed in Sec. III A, to let $\mathbf{C}_h = \sigma_h^2\mathbf{I}$; that is, we have no *a priori* knowledge of correlations and time varying amplitude variations of the individual magnitudes of the impulse response coefficients—if such correlations and amplitude variations do exist, then they must solely be given by the data.

The noise term, $\mathbf{e}$, in the model (4) above accounts for all uncertainties that we may have. As previously discussed we typically have model errors due to the finite length of the impulse responses, as well as the electronic measurement and the acoustic interference that cannot be avoided. Here we have also used a conservative model of the noise and let the covariance matrix have the form $\mathbf{C}_e = \sigma_e^2\mathbf{I}$. The value of σ_e is, however, in practice seldom known beforehand. By using probability theory we could assign a prior for σ_e and integrate (marginalize) it out.[16,17] This would, however, lead to a more complicated non-Gaussian posterior. An often good approximation is to estimate σ_e from data and then use the estimated value in Eq. (7a), which is the method chosen in this paper (the procedure for estimating the noise variance is described in Appendix A). The estimator (7a) can be rewritten in the equivalent form

$$\hat{\mathbf{h}}_n = (\mathbf{U}^T\mathbf{C}_e^{-1}\mathbf{U} + \mathbf{C}_h^{-1})^{-1}\mathbf{U}^T\mathbf{C}_e^{-1}\mathbf{y}_n, \tag{7b}$$

which can be practically useful since the size of the matrix inverse may then be smaller. The latter form (7b) can also be obtained by assuming a Gaussian distributed $\mathbf{h}_n$ and $\mathbf{e}$ and then choosing the estimate $\hat{\mathbf{h}}_n$ where the *a posteriori* pdf has its maximum.

In summary, we have assumed that $\mathbf{h}$ and $\mathbf{e}$ are zero mean and mutually independent with covariance matrices $\mathbf{C}_h = E\{\mathbf{h}_n\mathbf{h}_n^T\} = \sigma_h^2\mathbf{I}$ and $\mathbf{C}_e = E\{\mathbf{e}\mathbf{e}^T\} = \sigma_e^2\mathbf{I}$, respectively.

C. The impulse response error

An important feature of Bayesian analysis is that we obtain a measure of the accuracy of our estimates. This is given by the posterior distribution, which for the estimator in Eq. (7a) and (7b) is Gaussian. The estimation error, $\boldsymbol{\epsilon} = \hat{\mathbf{h}} - \mathbf{h}$, is then also (zero-mean) Gaussian where the covariance matrix for $\boldsymbol{\epsilon}$ is given by[18]

$$\begin{aligned}\mathbf{C}_\epsilon &= \mathbf{C}_h - \mathbf{C}_h\mathbf{U}^T(\mathbf{U}\mathbf{C}_h\mathbf{U}^T + \mathbf{C}_e)^{-1}\mathbf{U}\mathbf{C}_h \\ &= (\mathbf{C}_h^{-1} + \mathbf{U}^T\mathbf{C}_e^{-1}\mathbf{U})^{-1},\end{aligned} \tag{8}$$

where we have dropped $\mathbf{C}_e$'s dependence on $\mathbf{r}_n$ to simplify the notation.

IV. THE CORRECTION FILTER

The objective with the design of the correction filter is that the filter, as mentioned in Sec. I, should be robust in the sense that the listener should be able to move within a predefined region while still obtaining a reasonable correction. Since we are using a single source we cannot compensate for all (position-dependent) loudspeaker and room effects. That is, in particular, the reflected responses vary significantly over the region-of-interest,[5] and such effects, therefore, cannot be compensated—at least not without using a multiple-channel system. However, at lower frequencies the sound pressure varies more slowly with respect to listener movement, and if the position change is within a half wavelength then the sound wave will have a similar phase for the corresponding frequency. Thus, the correction filter can therefore compensate for the lower frequency room effects, and the cut-off frequency, where compensation is possible, is directly related to the size of our chosen correction region.[19] This information can be included in the design criteria, which will be further discussed below.

A. The linear minimum mean squared error correction filter

The objective is now to reproduce the input signal, $\mathbf{u}$, using linear pre-filtering according to

$$\hat{\mathbf{u}}_n * \boldsymbol{\delta}_{k-d_n} = \mathbf{h}_n * \mathbf{u} * \mathbf{f} + \mathbf{e} = \mathbf{H}_n\mathbf{U}\mathbf{f} + \mathbf{e}, \tag{9}$$

where $\mathbf{f}$ denotes the correction filter and where $\boldsymbol{\delta}_{k-d_n} = 1$ for $k = d_n$ and otherwise zero. Note that $\hat{\mathbf{u}}_n$ will be delayed relative to $\mathbf{u}$ due to the filtering delays associated with $\mathbf{h}_n$ and the filter delay due to the compensation filter $\mathbf{f}$ (i.e., it takes some time for the signal to pass through the system and the correction filter). We must, therefore, introduce a delay, d_n, in Eq. (9) in order to obtain a realistic estimator.

To find the linear filter we choose the filter, $\mathbf{f}$, that minimizes the mean squared sound reproduction error at a set of N observation points according to

$$\mathbf{f}_{\mathrm{OPT}} = \sum_{n=1}^{N} \arg\min_{\mathbf{f}} E\{\|\mathbf{u} * \boldsymbol{\delta}_{k-d_n} - \hat{\mathbf{u}}_n\|^2\}$$

$$= \sum_{n=1}^{N} \arg\min_{\mathbf{f}} E\{\|\boldsymbol{\Lambda}_{k-d_n}\mathbf{u} - \hat{\mathbf{u}}_n\|^2\}, \quad (10)$$

where $\boldsymbol{\Lambda}_{k-d_n}$ is the Toeplitz convolution matrix that corresponds to $\boldsymbol{\delta}_{k-d_n}$. We have again used a conservative approach and assumed that $E\{\mathbf{u}\mathbf{u}^T\} = \mathbf{C}_u = \sigma_u^2\mathbf{I}$.

Now, by using the estimated impulse responses in Eqs. (9) and (10) and equating the derivative with respect to $\mathbf{f}$ to zero, the correction filter which minimizes Eq. (10) becomes

$$\mathbf{f}_{\mathrm{LMMSE}} = \left(\sum_n \hat{\mathbf{H}}_n\hat{\mathbf{H}}_n^T\right)^{-1}\sum_n \hat{\mathbf{H}}_n^T\boldsymbol{\delta}_{k-d_n}, \quad (11)$$

where the dependence of the prior for $\mathbf{u}$ disappears since σ_u cancels out.

B. The LMMSE correction filter for room impulse responses with errors

Since we must estimate the room impulse responses from data, there will be deviations in our estimates from the true impulse responses, and, as discussed in Sec. III C, these deviations can be treated as Gaussian. A model that takes these uncertainties into account can then be expressed as

$$\mathbf{y}_n = (\hat{\mathbf{h}}_n + \Delta\mathbf{h}_n) * \mathbf{u} + \mathbf{e} = (\hat{\mathbf{H}}_n + \Delta\mathbf{H}_n)\mathbf{u} + \mathbf{e}, \quad (12)$$

where $\Delta\mathbf{h}_n \sim N(\mathbf{0}, \mathbf{C}_\epsilon^{(n)})$. Following the procedure used in Sec. IV A and using Eq. (12) in Eq. (10) results in

$$\mathbf{f}_{\mathrm{LMMSE}} = \left(\sum_n \hat{\mathbf{H}}_n^T\hat{\mathbf{H}}_n + \tilde{\mathbf{C}}_\epsilon^{(n)}\right)^{-1}\sum_n \hat{\mathbf{H}}_n^T\boldsymbol{\delta}_{k-d_n}, \quad (13)$$

where $\tilde{\mathbf{C}}_\epsilon^{(n)} = E\{\Delta\mathbf{H}_n^T\Delta\mathbf{H}_n\}$, which approximately is given by $\tilde{\mathbf{C}}_\epsilon^{(n)} \approx \text{Ł}_h\tilde{\sigma}_\epsilon^2\mathbf{I}$ (see Appendix B).

C. Reflection response filtering

Recall from the discussion in Sec. IV that we cannot compensate for reflections and high frequency room modes using a single source. Note, however, that since the reflection responses always arrive after the direct response, we can apply a suitable time varying filter to the impulse responses to attenuate the reflection effects in the impulse responses while keeping the low frequency room modes. Thus, by this procedure we can reduce the influence of the reflections in the design of the reconstruction filter.

The non-stationary filtering can be obtained by using a filter matrix where the columns are not delayed versions of the same filter, as for a normal Toeplitz convolution matrix, but now the filters are changing with time. This can be expressed in matrix form by multiplying the impulse response matrices $\mathbf{H}_n$ with a reflection filter matrix $\mathbf{F}_R$, that is, we replace $\mathbf{H}_n$ in the criteria (10) with $\mathbf{F}_R\mathbf{H}_n$, which results in the new criteria,

$$\mathbf{f}_{\mathrm{OPT}} = \arg\min_{\mathbf{f}} \sum_n E\{\|\mathbf{u} * \boldsymbol{\delta}_{k-d_n} - \mathbf{U}\mathbf{F}_R\mathbf{H}_n\mathbf{f}\|^2\}. \quad (14)$$

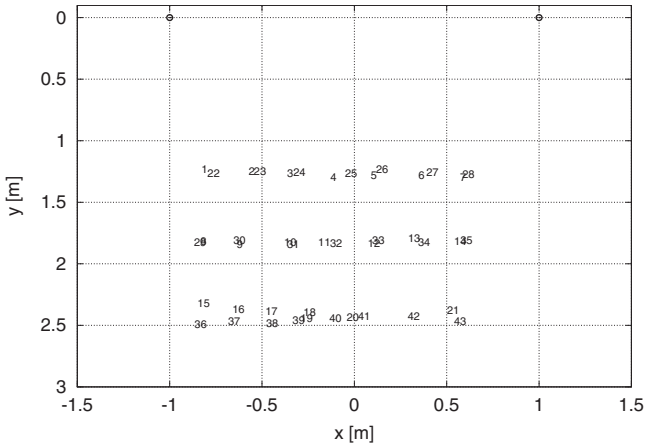

FIG. 2. The measurement setup. The circles indicate the loudspeaker positions and the numbers, 1–43, indicate the microphone measurement positions.

V. EXPERIMENTS

To verify the presented theoretical results above, experiments have been performed in a 5.8 m long, 4.5 m wide, and 2.6 m high room, which is designed after the live-end/dead-end principle; the wall behind the loudspeakers has sound absorbing draperies, and the wall at the opposite side to the loudspeakers has a sound diffuser. The data have been acquired using a DPA 4004 Reference microphone and an RME Multiface 24-bit sound card. The acoustic data were acquired at 43 measurement points, where measurements 1–21 were acquired at a height of 1.09 m and measurements 22–43 were acquired at 1.32 m, respectively. Two Genelec 8040A loudspeakers were used, and the layout of the experiments is shown in Fig. 2. In the results for the left loudspeaker presented here, the data from the 10th observation point, at $(-0.38, 1.83, 1.09)$ m, was used for evaluation, and the correction filter is based on data from the remaining 42 measurements. Furthermore, an impulse response (and correction filter) vector length of 16 384 elements was used here, which corresponds to a time duration of 371 ms. We have used white input signals of a duration of 3 s, sampled at a sampling frequency of 44.1 kHz, in the experiments, and the acoustic impulse responses were estimated using Eq. (7a) and (7b). We have given all estimated impulse responses an equal weight in the construction of the correction filter, and the data have, therefore, been normalized. We have also used a lower limit of 20 Hz and an upper limit 20 kHz for the working range of the correction filter.

In Fig. 3(a) the first part of the estimated impulse response for measurement point $(-0.38, 1.83, 1.09)$ m is shown. The response seen just before 2 ms is the direct path response, and the first (floor) reflection is the response seen at approximately 5 ms. One can also see that the frequency response, shown in Fig. 3(b), has many dips, which is known to cause problems in direct inversion approaches since then a division by zero nearly occurs.

Note that the arrival times of the same reflections but at different measurement points will not be the same since the propagation distances to the objects that reflect the sound will be different at different points. This is shown in Fig. 4(a)

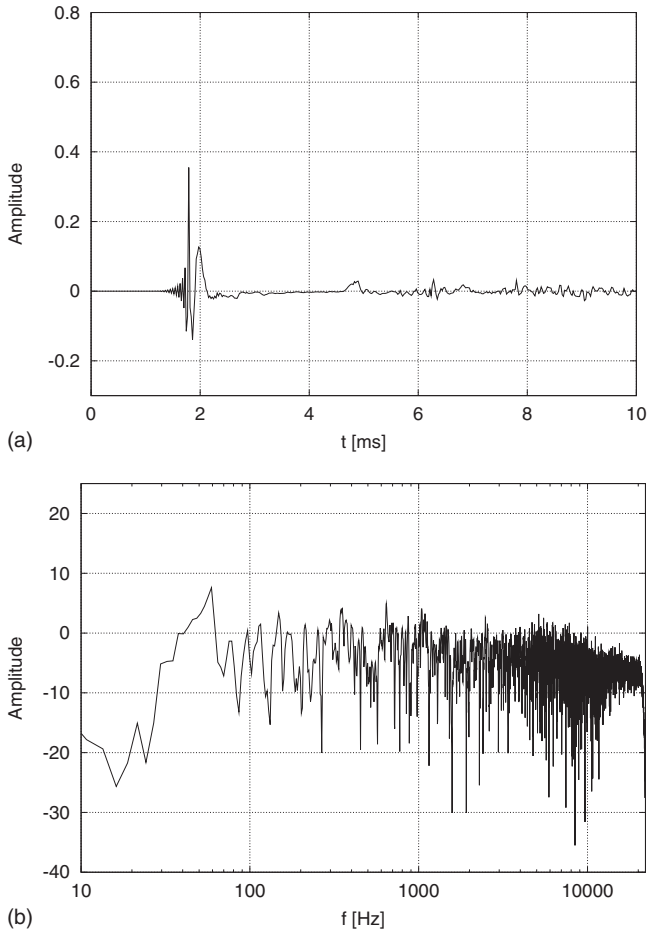

FIG. 3. The estimated acoustic impulse response for position (−0.38, 1.83, 1.09) (m).

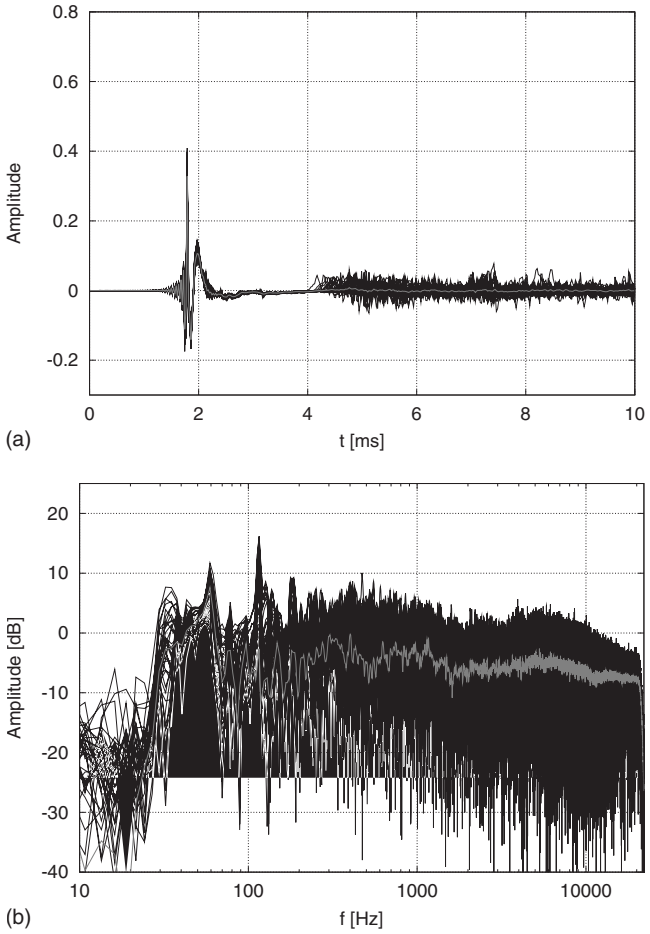

FIG. 4. The estimated acoustic impulse responses.

where all estimated impulse responses have been plotted in the same graph. As can be seen in the figure, the direct path impulse responses have similar, but not identical, shapes for all measurement points, but the reflections, which all arrive after 4 ms, vary significantly between different measurement points. This can also be seen by observing the average of the estimated impulse responses, shown as the gray solid line in Fig. 4(b); here it can be observed that the reflection part of the average impulse is close to zero since negative and positive responses does not arrive coherently for different points, and these parts will, therefore, almost cancel out in the average.

To further study the variability between estimated impulse responses, the spectra at each point have been plotted in the same graph in Fig. 4(b). Here we can see that the spectra have peaks and dips at roughly the same frequencies up to approximately 100–200 Hz, but above 200 Hz the peaks and dips occur at different frequencies for different points. This can also be seen in the average frequency response, shown as the gray solid line, which is relatively flat (with a falling trend) above 200 Hz with no strong isolated peaks or dips.

The LMMSE correction filter, based on the 42 "training" measurements, was then computed using Eq. (13), and the resulting filter is shown in Fig. 5. As seen in the frequency response in Fig. 5(b), the correction filter does not have excessive filter gain at any frequency. There is a peak at approximately 25 Hz since the loudspeakers do not transmit much energy below 30 Hz (see Fig. 4). One can also observe that there is three, or four, relatively well separated common resonance peaks in the spectra shown in Fig. 4(b) between 50 and 200 Hz. These peaks have corresponding dips in the correction filter's spectrum. One can also note that the correction filter compensates for the falling trend above 2–3 kHz seen in the average impulse response spectra in Fig. 4(b). The falling trend of the frequency responses can partly be explained by the diffraction effects of the acoustic source, which depend on the size of the source and the position of the measurement point; the fall-off trend for high frequencies increases when the distance from center-axis of the source increases (see, for example, Ref. 20).

To further show the properties of the LMMSE correction filter, the estimated impulse response for measurement point (−0.38, 1.83, 1.09) was filtered using the correction filter (shown in Fig. 5), which is displayed in Fig. 6. First one can observe, in Fig. 6(a), that the corrected system has a significantly shorter direct response compared to the uncorrected system. Second, the pre-ringings are roughly 50 dB lower than the maximum direct response [see Fig. 6(b)]. Third, the frequency response of the corrected system, shown in Fig. 6(c), is significantly closer to a flat response compared to uncorrected system. The correction is, as expected, not perfect since there are, in particular, still some dips in the spec-

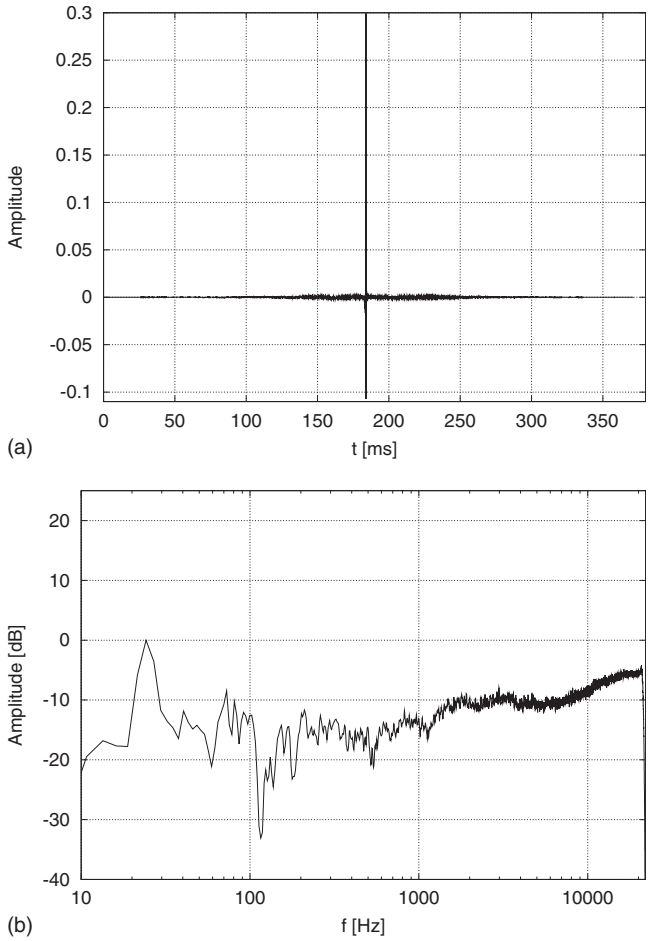

FIG. 5. The LMMSE correction filter.

tra for the corrected system. The high frequency part of the correction filter is also somewhat "noisy" [see Fig. 5(b)].

To see if the robustness of the correction can be improved by incorporating more prior information, we have also performed experiments using the (slightly heuristic) reflection filter approach described in Sec. IV C. Figure 7 shows the results where a non-stationary reflection filter with a cut-off frequency of 90 Hz was used when designing the correction filter. The start time for the non-stationary reflection filter was 3.6 ms, and the filter length was gradually increased until 5.6 ms where the full length (of four periods at the cut-off frequency) was reached. By comparing the results in Figs. 6(a) and 7(a), one can see that the correction of the direct response is similar both with and without reflection filtering, which was expected since the direct response was not affected by the filter. However, if we compare the results presented in logarithmic scale in Figs. 6(b) and 7(b), we can see that the reflection filter approach significantly has reduced the amplitudes of the pre-ringings. Furthermore, the high frequency parts of the correction filter is significantly smoother when reflection filtering is used, which can be seen by comparing Figs. 5(b) and 7(c).

VI. SUMMARY AND CONCLUSIONS

This paper treats the particular problem of the design of robust correction filters based on several impulse response

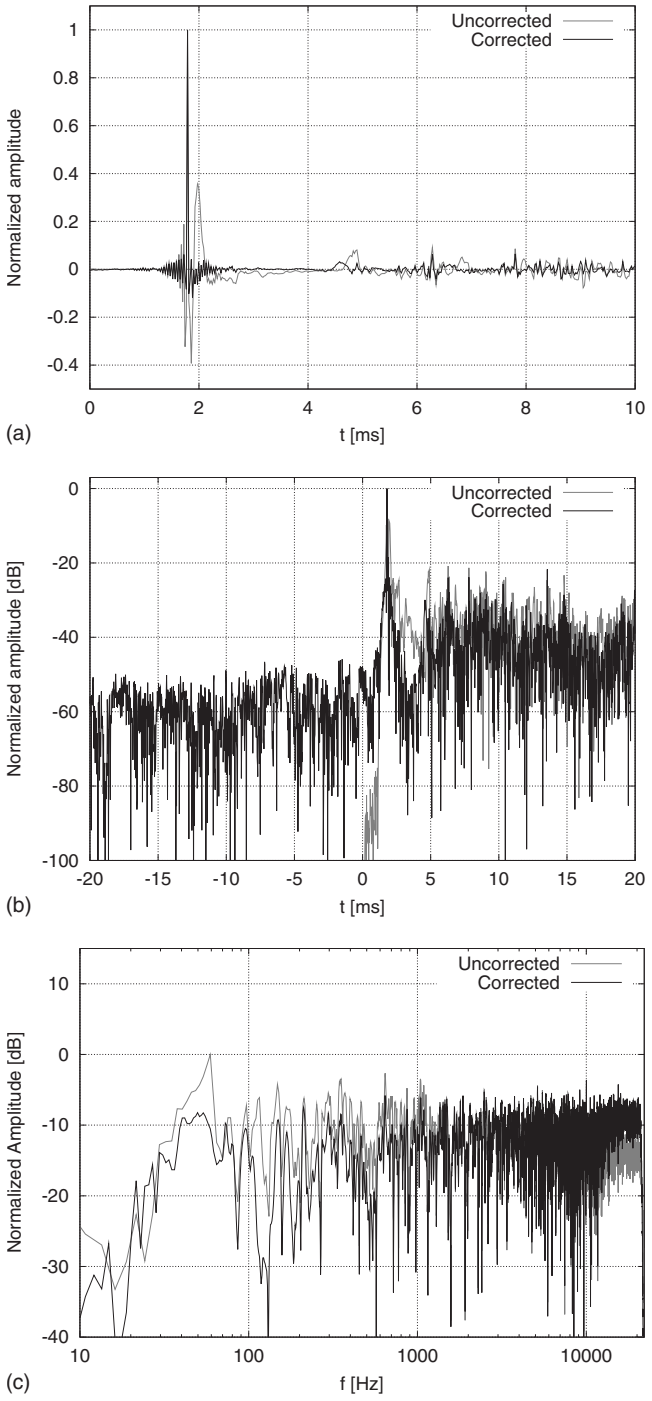

FIG. 6. The uncorrected and corrected impulse responses for measurement position $(-0.38, 1.83, 1.09)$ (m).

measurements. Clearly, the objective is to obtain the best correction possible, but we must first acknowledge the limitations of the experimental conditions and specify a realistic goal for our filter design. Here we are considering correction using a single source, and we are using a finite impulse response model where we must estimate the coefficients in the model from a limited number of noisy data. Due to the uncertainties, we have used a statistical approach in the design of the correction filter. The approach consists of two steps where we first estimate the impulse responses at a finite number of observation points in our region-of-interest, and then we construct a correction filter based on the estimated

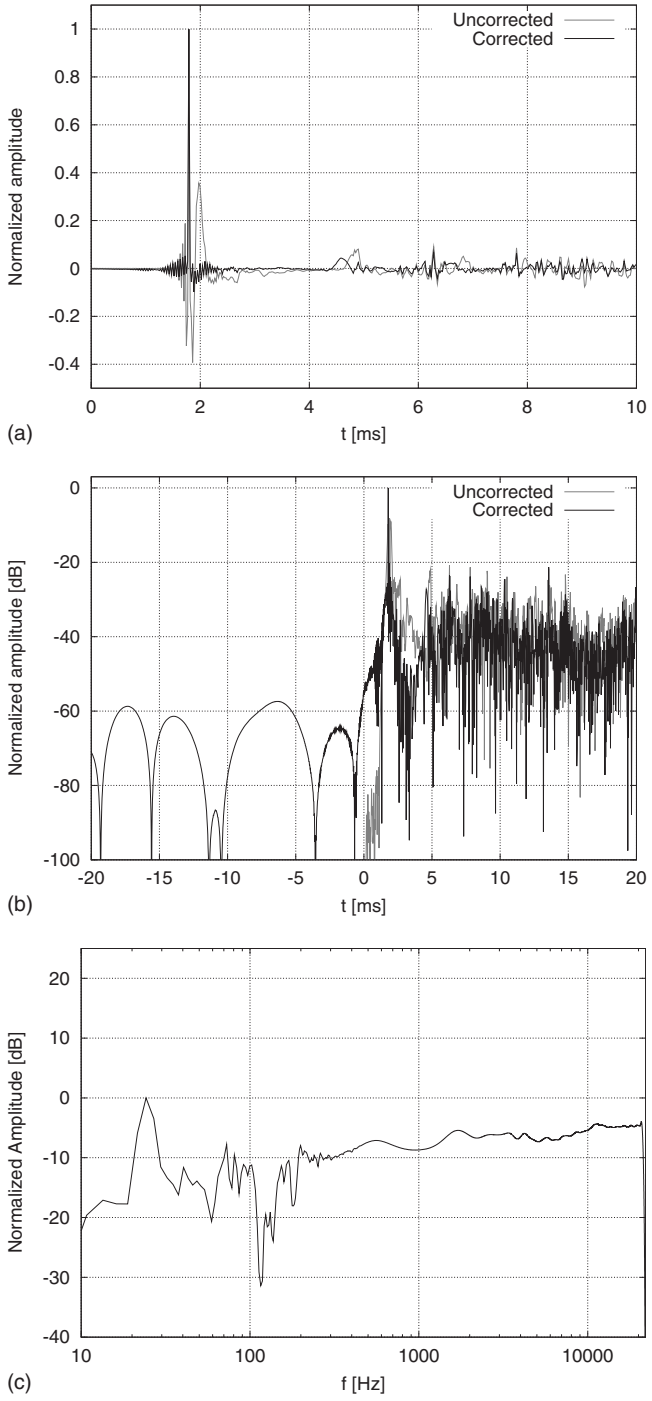

FIG. 7. The uncorrected and corrected impulse responses for measurement position (−0.38, 1.83, 1.09) (m) using a correction filter based on reflection filtered impulse responses.

impulse responses. By taking the uncertainties (i.e., the model error and measurement noise) into account, we have obtained reliable estimates of the impulse responses, which is vital for the design of the correction filter. The correction filter is then constructed by using a linear minimum mean squared criteria. We have, furthermore, introduced a time-dependent reflection filter in criteria, which attenuates the high frequency parts of the reflected responses, that is, the parts of the responses that we cannot compensate with a single source system.

As can be seen in the results from the experiments presented in this paper, the statistical multiple-point correction method introduced here is capable of significantly improving both the temporal and spectral responses compared to the uncorrected system. Furthermore, the obtained correction filter has a low level of pre-ringing, which was in the order of 60 dB lower than the direct response for the data used in this paper.

APPENDIX A: ESTIMATING THE NOISE VARIANCE

Here we are interested to estimate the noise variance σ_e^2 only and we, therefore, marginalize over the impulse response coefficients according to

$$p(\sigma_e|\mathbf{y}_n) \propto p(\sigma_e|I)\int_{\mathbf{h}_n} p(\mathbf{h}_n|I)p(\mathbf{y}_n|\mathbf{h}_n,\sigma_e,I)d\mathbf{h}_n = p(\sigma_e|I)\mathcal{L}(\sigma_e). \tag{A1}$$

If we, for convenience, use a uniform prior for $\mathbf{h}_n$, then the integral in Eq. (A1) becomes

$$\mathcal{L}(\sigma_e) = \int_{\mathbf{h}_n} \frac{1}{R_h}\frac{1}{(2\pi)^{L/2}\sqrt{|\sigma_e^2\mathbf{I}|}} \times \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}_n-\mathbf{U}\mathbf{h}_n)^T(\mathbf{y}_n-\mathbf{U}\mathbf{h}_n)\right\}d\mathbf{h}_n, \tag{A2}$$

where $R_h=\mathbf{h}_{\max}-\mathbf{h}_{\min}$ is large enough so that the integral, in practice, can be evaluated from $-\infty$ to ∞. Then, after evaluating the Gaussian integral, Eq. (A2) becomes,

$$\mathcal{L}(\sigma_e) = \frac{1}{R_h}\frac{1}{(2\pi)^{L/2}\sqrt{|\sigma_e^2\mathbf{I}|}}\sqrt{\left(\frac{(\sigma_e^2 2\pi)^{L_h}}{|\mathbf{U}^T\mathbf{U}|}\right)} \times \exp\left\{-\frac{1}{2\sigma_e^2}\mathbf{y}_n^T(\mathbf{I}-\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{y}_n\right\}. \tag{A3}$$

Furthermore, since the noise variance is a scale parameter, we use a (normalized) Jeffreys prior for σ_e,[17]

$$p(\sigma_e|I) = \frac{1}{\log(\sigma_{\max}/\sigma_{\min})}\frac{1}{\sigma_e}, \tag{A4}$$

where $\sigma_{\min}$ is a small value but larger than zero (typically determined by the bit resolution of the analog-to-digital converter) and $\sigma_{\max}$ is a large but finite value. Using Eqs. (A4) and (A3) in Eq. (A1), and equating the derivative of the log of the posterior (with respect to σ_e) to zero, then results in

$$\hat{\sigma}_e = \sqrt{\frac{\mathbf{y}_n^T(\mathbf{I}-\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{y}_n}{L-L_h+1}}; \tag{A5}$$

that is, Eq. (A5) is the maximum *a posteriori* estimate of σ_e for a uniform prior of $\mathbf{h}_n$.

Recall from Sec. II that here the noise describes the model misfit, and different impulse response lengths will, therefore, result in different values of the estimated noise variance. As an example, the estimated σ_e for six impulse response lengths are given in Table I for the observation point at (−0.38, 1.83, 1.09). One can note that estimated σ_e becomes smaller when the impulse response length in-

TABLE I. Estimated σ_c for six different impulse response lengths, L_h, for the observation point (−0.38, 1.83, 1.09) (m).

L_h	512	1024	2048	4096	8192	16 384
σ_c	0.1644	0.1241	0.0585	0.0226	0.0126	0.0119

creases. This trend can also be seen in Fig. 8 where the normalized log likelihood for σ_e has been plotted for three different impulse response lengths. Here, one can also observe that the value of the most likely σ_e decreases when the impulse response length increases since a longer impulse response can describe the data better, and, furthermore, the posteriors are so narrow that they are essentially delta functions; hence, the actual limits of the prior for σ_e have negligible influence of the posterior.

APPENDIX B: AN APPROXIMATE EXPRESSION FOR THE ERROR COVARIANCE MATRIX

The error covariance matrix $\mathbf{C}_e$ was given in Eq. (8). This expression can be simplified by noting that since we are using a white excitation signal the matrix $\mathbf{U}^T\mathbf{U}$ will be nearly diagonal and can, therefore, be approximated with $\mathbf{u}^T\mathbf{u}\mathbf{I}$. By making this approximation, Eq. (8) reduces to

$$\mathbf{C}_\epsilon \approx \left(\frac{1}{\sigma_h^2}\mathbf{I} + \frac{\mathbf{u}^T\mathbf{u}}{\sigma_e^2}\mathbf{I}\right)^{-1} = \frac{\sigma_h^2\sigma_e^2}{\sigma_e^2 + \sigma_h^2\mathbf{u}^T\mathbf{u}}\mathbf{I}. \tag{B1}$$

If we use this approximation in the correction filter discussed in Sec. IV B, then the matrix $\tilde{\mathbf{C}}_e^{(n)} = E\{\Delta\mathbf{H}_n^T\Delta\mathbf{H}_n\}$ reduces to

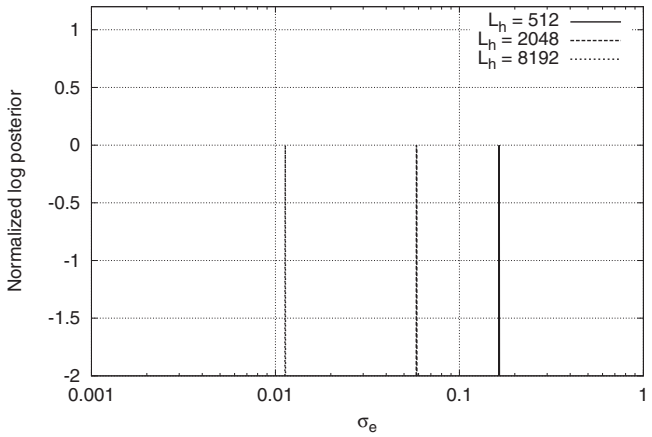

FIG. 8. The normalized log posterior for σ_e for three different impulse response lengths at observation point (−0.38, 1.83, 1.09) (m). The solid line corresponds to an impulse response length of 512 (right curve), the dashed line to 2048 (middle curve), and the dotted line to 8192 (left curve), respectively.

$$\tilde{\mathbf{C}}_\epsilon^{(n)} \approx \mathrm{L}_h\frac{\sigma_h^2\sigma_e^2}{\sigma_e^2 + \sigma_h^2\mathbf{u}^T\mathbf{u}}\mathbf{I} = \mathrm{L}_h\tilde{\sigma}_\epsilon^2\mathbf{I}, \tag{B2}$$

where we have dropped the dependence of n for σ_e^2 to simplify the notation.

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," J. Audio Eng. Soc. **42**, 884–900 (1994).
[2] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," J. Audio Eng. Soc. **47**, 583–595 (1999).
[3] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," J. Audio Eng. Soc. **51**, 3–26 (2003).
[4] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," J. Audio Eng. Soc. **37**, 899–907 (1989).
[5] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," IEEE Trans. Speech Audio Process. **8**, 311–319 (2000).
[6] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," J. Audio Eng. Soc. **48**, 259–280 (2000).
[7] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," IEEE Trans. Speech Audio Process. **5**, 325–333 (1997).
[8] F. Asano, Y. Suzuki, and T. Sone, "Sound equalization using derivative constraints," Acust. Acta Acust. **82**, 311–320 (1996).
[9] S. Bharitkar and C. Kyriakakis, "A cluster centroid method for room response equalization at multiple locations," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001), pp. 55–58.
[10] R. Wilson, "Equalization of loudspeaker drive units considering both on- and off-axis responses," J. Audio Eng. Soc. **39**, 127–139 (1991).
[11] N. Stefanakis, J. Sarris, G. Cambourakis, and F. Jacobsen, "Power-output regularization in global sound equalization," J. Acoust. Soc. Am. **123**, 33–36 (2007).
[12] O. Kirkeby, P. A. Nelson, and H. Hamada, "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am. **100**, 1584–1593 (1996).
[13] L.-J. Brännmark and A. Ahlén, "Robust loudspeaker equalization based on position-independent excess phase modeling," in IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, Las Vegas, NV (March 30–April 4, 2008).
[14] E. T. Jaynes, "Prior information and ambiguity in inverse problems," Inverse Probl. **14**, 151–166 (1984).
[15] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).
[16] E. T. Jaynes, *Probability Theory: The Logic of Science*, 1st ed. (Cambridge University Press, Cambridge, UK, 2003).
[17] P. C. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, Cambridge, UK, 2005).
[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1993), Vol. 1.
[19] A. O. Santillán, "Spatially extended sound equalization in rectangular rooms," J. Acoust. Soc. Am. **110**, 1989–1997 (2001).
[20] F. Lingvall and T. Olofsson, "Statistically motivated design of input signals for modern ultrasonic array systems," J. Acoust. Soc. Am. **123**, 2620–2630 (2008).