

Spatially Robust Audio Compensation Based on SIMO Feedforward Control

Lars-Johan Brännmark, *Student Member, IEEE*, and Anders Ahlén, *Senior Member, IEEE*

Abstract—This paper introduces a single-input multiple-output (SIMO) feedforward approach to the single-channel loudspeaker equalization problem. Using a polynomial multivariable control framework, a spatially robust equalizer is derived based on a set of room transfer functions (RTFs) and a multipoint mean-square error (MSE) criterion. In contrast to earlier multipoint methods, the polynomial approach provides analytical expressions for the optimum filter, involving the RTF polynomials and certain spatial averages thereof. However, a direct use of the optimum solution is questionable from a perceptual point of view. Despite its multipoint MSE optimality, the filter exhibits similar, albeit less severe, problems as those encountered in nonrobust single-point designs. First, in the case of mixed phase design it is shown to cause residual “pre-rings” and undesirable magnitude distortion in the equalized system. Second, due to insufficient spatial averaging when using a limited number of RTFs in the design, the filter is overfitted to the chosen set of measurement points, thus providing insufficient robustness. A remedy to these two problems is proposed, based on a constrained MSE design and a method for clustering of RTF zeros. The outcome is a mixed phase compensator with a time-domain performance preferable to that of the original unconstrained design.

Index Terms—Acoustic signal processing, compensation, equalizers, optimal control, polynomials, robustness, transient response.

I. INTRODUCTION

THE problem of single-channel loudspeaker equalization by the use of digital filters has been extensively studied for about two decades, with an increasing concern in recent years about spatial robustness. In a broad sense, the aim of all audio channel equalization schemes is to remove undesired convolutional distortions introduced by the electroacoustical signal path of a sound system. In the literature, the work on robustness of equalization essentially falls into three categories. In the first category, the goal of filter design is a complete signal dereverberation at a single position in a room. The subsequent robustness analysis then investigates equalizer performance at other spatial positions, or under slightly modified acoustical circumstances. It is well known that this kind of filter design is highly

non-robust and causes severe signal degradation when the receiver position changes [1], and even for fixed receiver position, due to the “weak nonstationarity” of the acoustical paths in the room [2]. In the second category, the design objective is not a complete dereverberation, but rather a reduction of linear distortions, under the constraint that audio performance should not be degraded by changes of listener position. The standard approach in this category is to design a filter based on averaging and/or smoothing of one or several transfer functions and then perform a robustness analysis of the filter [3]. The third category imposes robustness directly on the design by employing a multipoint error criterion to optimize sound reproduction in a number of spatial positions, either by using measured room transfer functions (RTFs) [4] or by direct adaptation of the inverse [5]. We mention here parenthetically a fundamentally different multipoint scenario, where signals are filtered on the receiver side by a unique equalizer at each receiver point. Spatial robustness in this setting has been studied in [6] and [7]. This approach is, however, not applicable in the precompensation setting, where a single filter operates on the input to the system.

In the present paper, the problem formulation relates closest to the third of the above categories. We shall start by defining a multipoint mean-square error (MSE) criterion for spatially robust filter design in a single-input multiple-output (SIMO) setting. Using a polynomial approach to the multivariable feedforward control problem [8], a linear filter is designed to minimize the multipoint MSE criterion. The arising equations allow for mixed phase as well as minimum phase inverse design. In contrast to the Wiener–Levinson and adaptive least-mean-squares (LMS) approaches used in e.g., [4], [5], the polynomial approach imposes no restrictions on filter order or structure, and the analytical form of the solution is amenable to interpretation in terms of certain spatial averages of the RTFs. MSE optimality does, however, not necessarily imply a good perceptual behavior, which calls for a solution based on refined perceptual considerations. By lack of degrees of freedom in the SIMO setting, ideal equalization in all measurement positions is not possible. Consequently, there will be an equalization error in every position, contributing to the difference between the reconstructed signal and the desired signal. Correlations between this error and the desired signal for negative time lags should be limited, as they will be identified by a listener as “pre-rings” in the equalized system. By a further analysis of the design equations we develop a method for avoiding the pre-ringing problem, without resorting to a pure minimum phase inversion. An early version of this approach was introduced by the authors in [9].

The filter design and analysis presupposes an arbitrarily large number of available RTF measurements. For a practical filter design, a spectral smoothing operation has shown to be a valuable

Manuscript received March 24, 2008; accepted November 26, 2008. First published January 23, 2009; current version published April 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus. This research was supported in part by Dirac Research AB, Uppsala, Sweden.

L.-J. Brännmark is with the Signals and Systems Group, Department of Engineering Sciences, Uppsala University, SE-751 21 Uppsala, Sweden, and also with Dirac Research AB, SE-754 50 Uppsala, Sweden (e-mail: ljb@signal.uu.se).

A. Ahlén is with the Signals and Systems Group, Department of Engineering Sciences, Uppsala University, SE-751 21 Uppsala, Sweden (e-mail: aa@signal.uu.se).

Digital Object Identifier 10.1109/TSP.2009.2013893

complement to the insufficient spatial averaging that arises from using a limited number of RTF measurements. Furthermore, if the sound system subject to equalization has limited bandwidth, some limitation on the filter gain may be necessary in order not to boost frequencies outside the working range of the loudspeaker. Perceptual issues of more intricate nature such as desired tonal coloration etc. can be straightforwardly included in the design. To keep the discussion focused, such issues will, however, not be considered here.

The paper is organized as follows. Section II formulates the robust audio compensation problem in our SIMO feedforward setting. In Section III, the problem is stated and solved mathematically, and the special cases of minimum and mixed phase inversion with ideal target dynamics are studied. In Section IV, qualitative aspects of the filters are investigated for different design scenarios, and some perceptual problems are pointed out. In Section V, these problems are analyzed and remedies are proposed. In Section VI, the methods of previous sections are evaluated using RTFs acquired in a real room. Finally, Section VII concludes the paper and points out some directions for further research.

Notation and Terminology: Throughout this paper, we shall use the following notation and terminology: Scalar and vector valued discrete-time signals are denoted by normal and boldface italic letters, like $s(k)$ and $\mathbf{s}(k)$, respectively. In the style of [10], transfer functions are represented by polynomial and rational matrices in the backward shift operator q^{-1} , defined by $q^{-1}s(k) = s(k-1)$, corresponding to z^{-1} in the frequency domain.

Constant matrices are denoted by boldface capital letters as, for example, \mathbf{P} . Scalar polynomials are denoted by capital letters in italic as $P(q^{-1}) = p_0 + p_1q^{-1} + \dots + p_{n_p}q^{-n_p}$. Polynomial matrices are denoted by boldface capital letters in italic as $\mathbf{P}(q^{-1}) = \mathbf{P}_0 + \mathbf{P}_1q^{-1} + \dots + \mathbf{P}_{n_p}q^{-n_p}$. Rational matrices are denoted by boldface calligraphic letters as $\mathcal{G}(q^{-1})$, and are represented on right matrix fraction description (MFD) form [11]: $\mathcal{G} = \mathbf{Q}\mathbf{P}^{-1}$ which for SIMO systems is equivalent to the common denominator form $\mathcal{G}(q^{-1}) = \mathbf{Q}(q^{-1})/P(q^{-1})$, where $\mathbf{Q}(q^{-1})$ is a polynomial matrix and the scalar monic polynomial $P(q^{-1})$ is the least common denominator of all rational elements in $\mathcal{G}(q^{-1})$. For scalar rational functions, normal calligraphic letters are used, like $\mathcal{G}(q^{-1})$. The arguments q^{-1} and z^{-1} will often be omitted, unless there is a risk for confusion. All signals and polynomial coefficients are assumed to be real valued. For any polynomial matrix $\mathbf{P}(q^{-1})$, or scalar polynomial $P(q^{-1})$, we define their conjugates as $\mathbf{P}_*(q) = \mathbf{P}^T(q) = \mathbf{P}_0^T + \mathbf{P}_1^Tq + \dots + \mathbf{P}_{n_p}^Tq^{n_p}$, or $P_*(q) = P(q) = p_0 + p_1q + \dots + p_{n_p}q^{n_p}$. Two polynomials P and Q are said to be *coprime* if they have no common factors, i.e., $\gcd(P, Q) = 1$. A system or a transfer function having m inputs and p outputs is said to be of dimension $p|m$. $\Re(z)$ and $\Im(z)$ denote the real and imaginary parts respectively of a complex number z .

An RTF is a linear time-invariant model of the signal path between the source (sound system input) and the receiver (microphone output) in a room. In the general case, the RTF between the system input and receiver position i is represented in discrete time by a scalar rational transfer function

$\mathcal{H}_i(z^{-1}) = B_i(z^{-1})/A_i(z^{-1})$; $i \in \{1, \dots, p\}$, where p is the number of receiver positions. In the sequel, the receiver positions are referred to as *control points*. We will frequently use transfer operators, e.g., $\mathcal{H}_i(q^{-1})$ as a representation of RTFs. For simplicity we will, however, refer to both as RTFs, or simply transfer functions, and when $\mathcal{H}_i(q^{-1})$ is used in this context we mean that q^{-1} is substituted for the complex variable z^{-1} . For finite-impulse-response (FIR) models (i.e., $A_i(q^{-1}) = 1$ above), the polynomial notation $B_i(q^{-1})$ is used instead of $\mathcal{H}_i(q^{-1})$. The time-domain impulse response related to a transfer function $\mathcal{H}(z^{-1})$ is denoted by $h(k)$. The *complex spatial average* model $B_0(q^{-1})$ refers to the polynomial obtained by taking the coefficient-wise sum of the FIR transfer functions B_1, \dots, B_p :

$$B_0(q^{-1}) = \sum_{i=1}^p B_i(q^{-1}). \quad (1)$$

The *root-mean-square (RMS) spatial average* model $\beta(q^{-1})$ refers to the minimum phase polynomial obtained by spectral factorization of the coefficient-wise sum of the power responses $B_{1*}B_1, \dots, B_{p*}B_p$ associated with the FIR models B_1, \dots, B_p :

$$\beta_*(q)\beta(q^{-1}) = \sum_{i=1}^p B_{i*}(q)B_i(q^{-1}). \quad (2)$$

The *minimum phase equivalent* $\beta_i(q^{-1})$ of an FIR transfer function $B_i(q^{-1})$ is the minimum phase polynomial obtained by spectral factorization of the power response $B_{i*}B_i$. The *excess phase part* of the same transfer function is the all-pass response obtained as $B_i(q^{-1})/\beta_i(q^{-1})$. A *zero cluster* is a set of polynomial zeros $\{z_1, \dots, z_p\}$, located within a small neighborhood $N_\epsilon(z_0)$ in the complex plane, where each zero z_i belongs to exactly one RTF $B_i(z^{-1})$. If the region $N_\epsilon(z_0)$ is sufficiently small, then the RTFs are said to have a *near-common zero* at z_0 . Zeros outside the unit circle are referred to as *excess phase zeros*.

II. THE ROBUST AUDIO COMPENSATION PROBLEM

We consider a single-channel setting, where the equalizer filter $\mathcal{R}(q^{-1})$ is assumed to operate on a scalar input signal $w(k)$ (see Fig. 1). The filtered signal is emitted by a loudspeaker and is received by a listener in one out of (infinitely) many locations in a room. Each receiver location is associated with an individual RTF, and the filter should be designed so as to improve sound reproduction over a whole set of control points. The control points are selected so as to cover a spatial region of hypothetical listener positions, henceforth referred to as the *listening region*. In deriving the equations, the number of control points p is assumed large but finite. Theoretically, a finite p imposes no essential restriction, since by the limited range of wavelengths a discrete grid of points is sufficient to represent the complete sound field within the region of interest. However, the dense spatial sampling required for such a complete representation is infeasible in a practical situation, and the optimization will in general be based on a rather low number of RTFs. As we shall see, this restriction can be quite problematic and calls for

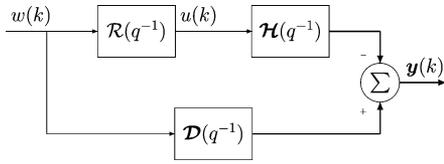


Fig. 1. Block diagram of the SIMO feedforward control problem. The thin lines represent scalar signals, and the thick lines represent vector-valued signals of dimension p .

a solution, if true robustness within the whole listening region is to be obtained.

Now, with each RTF described as a rational function $\mathcal{H}_i(q^{-1})$, the signals at the control points can be viewed as the p -dimensional output of a SIMO linear system of dimension $p|1$, having transfer function matrix $\mathcal{H}(q^{-1})$. Similarly, the desired responses $\mathcal{D}_i(q^{-1})$ can be stacked in a $p|1$ matrix $\mathcal{D}(q^{-1})$. If the criterion to be minimized is chosen as the sum of the mean squared errors $E\{|y_i(k)|^2\}$, with $y_i(k)$ being the difference between the received filtered signal and the desired signal, $y_i(k) = \mathcal{D}_i(q^{-1})w(k) - \mathcal{H}_i(q^{-1})\mathcal{R}(q^{-1})w(k)$, then the problem is equivalent to a SIMO linear-quadratic (LQ) feedforward control problem as depicted in the block diagram of Fig. 1. The sound propagation to the control points is affected by propagation delays of Δ_i samples. While the “true” RTF in position i is $q^{-\Delta_i}\mathcal{H}_i(q^{-1})$, we shall assume that the individual acoustic delay $q^{-\Delta_i}$ associated with each $\mathcal{H}_i(q^{-1})$ is removed prior to the filter design, so that all impulse responses $h_i(k)$ are aligned and start at $k = 0$. An equivalent but notationally more cumbersome approach would be to include the delays $q^{-\Delta_i}$ also in the desired responses $\mathcal{D}_i(q^{-1})$.¹

III. SIMO LQ FEEDFORWARD CONTROL

A. The SIMO Optimum Controller Equations

It is assumed that $w(k)$ is a scalar stationary white noise sequence with zero mean and covariance $E\{w^2(k)\} \triangleq \psi$. The stable rational matrices $\mathcal{H}(q^{-1})$ and $\mathcal{D}(q^{-1})$, representing respectively the original RTFs and the desired system responses at p spatial control points, are described by right MFD models

$$\mathcal{H} = \mathbf{B}/A = \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix} \frac{1}{A}; \quad \mathcal{D} = \mathbf{D}/E = \begin{bmatrix} D_1 \\ \vdots \\ D_p \end{bmatrix} \frac{1}{E} \quad (3)$$

so that

$$\mathbf{y}(k) = \frac{\mathbf{D}}{E}w(k) - \frac{\mathbf{B}}{A}\mathcal{R}w(k) \quad (4)$$

with A and E being stable monic polynomials. The robust SIMO compensator is defined as the filter which minimizes the sum of the powers of the signals in $\mathbf{y}(k)$. That is, the scalar rational filter $\mathcal{R}(q^{-1})$ is to be designed so that the criterion

$$J = E\{\|\mathbf{y}(k)\|_2^2\} = E\{\text{tr}(\mathbf{y}(k)\mathbf{y}^T(k))\} \quad (5)$$

¹With propagation delays $q^{-\Delta_i}$ included both in the RTFs \mathcal{H}_i and in the desired responses \mathcal{D}_i , the time-alignment operation would be performed anyway later on in the design, in the product $\mathbf{B}_*\mathbf{D}$ of (8).

is minimized, under the constraints of stability and causality of $\mathcal{R}(q^{-1})$. Formulated as above, this problem is readily seen to be a special case of the general multiple-input multiple-output (MIMO) feedforward problem treated in [8, Sec. 3.3]. Following that derivation and using our specialization of the problem, the optimum causal compensator filter is given by

$$\mathcal{R}(q^{-1}) = \frac{Q(q^{-1})A(q^{-1})}{\beta(q^{-1})E(q^{-1})} \quad (6)$$

where $\beta(q^{-1})$ is the minimum phase polynomial defined by

$$\beta_*(q)\beta(q^{-1}) = \mathbf{B}_*(q)\mathbf{B}(q^{-1}) = \sum_{i=1}^p B_{i*}(q)B_i(q^{-1}) \quad (7)$$

and $Q(q^{-1}) = q_0 + q_1q^{-1} + \dots + q_{n_Q}q^{-n_Q}$ along with the polynomial $L_*(q) = l_0 + l_1q + \dots + l_{n_L}q^{n_L}$ constitute the unique solution to the scalar polynomial Diophantine equation

$$\mathbf{B}_*(q)\mathcal{D}(q^{-1}) = \beta_*(q)Q(q^{-1}) + qL_*(q)E(q^{-1}) \quad (8)$$

with polynomial degrees

$$n_Q = \max\{n_D, n_E - 1\}; n_L = n_B - 1. \quad (9)$$

For a complete derivation of (6)–(9) in the general MIMO case, see [8].

B. Optimum Mixed and Minimum Phase Designs

In this subsection we study the filter defined by (6)–(8) more closely for the two important special cases of minimum and mixed phase inversion using ideal target dynamics. For clarity of presentation and ease of interpretation we assume that the system $\mathcal{H}(q^{-1})$ and the target dynamics $\mathcal{D}(q^{-1})$ are polynomial matrices of dimension $p|1$, containing the RTFs $B_i(q^{-1})$ and target responses $D_i(q^{-1})$, respectively. Hence, $A(q^{-1}) = E(q^{-1}) = 1$ in (3) and subsequent equations. This restriction is of no practical importance, since the FIR models $B_i(q^{-1})$ and $D_i(q^{-1})$ are allowed to be of arbitrarily high degree².

We begin the analysis by concluding from (7) that the polynomial $\beta(q^{-1})$ in the denominator of (6) is identical to that of the RMS average model (2). Further, we note that if $D_i(q^{-1}) = q^{-d}, \forall i \in \{1, \dots, p\}$, i.e., the desired response at position i is a pure delay of d samples (in addition to the acoustic delay $q^{-\Delta_i}$ discussed in Section II), then (8) can be rewritten as

$$\sum_{i=1}^p B_{i*}q^{-d} = \beta_*Q + qL_* \implies q^{-d}B_{0*} = \beta_*Q + qL_* \quad (10)$$

where $B_{0*}(q^{-1})$ is the complex spatial average (1). The delay q^{-d} represents the number of “future” input signal samples to be used by the filter. Exchanging q^{-1} for q in (10) and dividing by $\beta(q^{-1})$ gives the equivalent equation

$$q^d \frac{B_{0*}(q^{-1})}{\beta(q^{-1})} = Q_*(q) + q^{-1} \frac{L(q^{-1})}{\beta(q^{-1})}. \quad (11)$$

²Note that in some situations, it may be more efficient to include $A(q^{-1})$ and $E(q^{-1})$, for example in the modeling of very large or undamped rooms. However, to keep the discussion as clear as possible we use $A = E = 1$.

Since $\beta(q^{-1})$ is minimum phase, we can define the power series $\Gamma(q^{-1})$ and $\Lambda(q^{-1})$:

$$\begin{aligned}\Gamma(q^{-1}) &\triangleq \frac{B_0(q^{-1})}{\beta(q^{-1})} \\ &= \sum_{k=0}^{\infty} \gamma_k q^{-k}; |\gamma_k| < c_\gamma r_{\max}^k\end{aligned}\quad (12)$$

$$\begin{aligned}\Lambda(q^{-1}) &\triangleq q^{-1} \frac{L(q^{-1})}{\beta(q^{-1})} \\ &= \sum_{k=1}^{\infty} \lambda_k q^{-k}; |\lambda_k| < c_\lambda r_{\max}^k\end{aligned}\quad (13)$$

where c_γ and c_λ are positive constants and $r_{\max} < 1$ is the maximum radius for any zero of $\beta(z^{-1})$. Equation (11) can then be written

$$\begin{aligned}q^d \Gamma(q^{-1}) &= Q_*(q) + \Lambda(q^{-1}) \\ \Leftrightarrow q^d \sum_{k=0}^{\infty} \gamma_k q^{-k} &= Q_*(q) + \sum_{k=1}^{\infty} \lambda_k q^{-k} \\ \Leftrightarrow \sum_{k=0}^d \gamma_{d-k} q^k + \sum_{k=1}^{\infty} \gamma_{d+k} q^{-k} \\ &= Q_*(q) + \sum_{k=1}^{\infty} \lambda_k q^{-k}.\end{aligned}\quad (14)$$

Since $Q_*(q)$ is a polynomial in nonnegative powers of q only, identifying the coefficients for positive and negative powers of q in (14) yields

$$\sum_{k=1}^{\infty} \gamma_{d+k} q^{-k} = \sum_{k=1}^{\infty} \lambda_k q^{-k} = \Lambda = q^{-1} \frac{L(q^{-1})}{\beta(q^{-1})}\quad (15)$$

$$Q(q^{-1}) = \sum_{k=0}^d \gamma_{d-k} q^{-k}.\quad (16)$$

We know, however, from (12) that γ_k is an exponentially decaying sequence, so by increasing the delay d , the coefficients of Λ can be made arbitrarily small. Let the left-hand side of (14) be denoted $\tilde{Q}_*(q, q^{-1})$. Then, for the special case when d is very large,

$$\begin{aligned}Q(q^{-1}) &\approx \tilde{Q}(q^{-1}, q) = \sum_{k=0}^d \gamma_{d-k} q^{-k} + \sum_{k=1}^{\infty} \gamma_{d+k} q^k \\ &= q^{-d} \sum_{k=0}^{\infty} \gamma_k q^k = q^{-d} \frac{B_{0*}(q)}{\beta_*(q)}.\end{aligned}\quad (17)$$

With $Q(q^{-1}) \approx \tilde{Q}(q^{-1}, q)$ we mean that $Q(q^{-1})$, which is a polynomial in q^{-1} only, has almost the same impulse response as does $\tilde{Q}(q^{-1}, q)$, which is a rational function in q and q^{-1} . In fact, $Q(q^{-1})$ is the causal part of $\tilde{Q}(q^{-1}, q)$. The impulse response of Q , when approximated as above, is seen to be the time-reversed and delayed impulse response of the ratio between the complex (B_0) and RMS (β) spatial averages. Although technically the correct expression for Q is (16), we shall

be using the approximation (17) in the following, since it allows the pre-ringing part of the inverse filter to be interpreted as a non-causal filter containing excess phase poles. Using the approximation (17) for Q in (6), and assuming $A = E = 1$, the optimal compensator filter can be written

$$\mathcal{R}(q^{-1}) \approx q^{-d} \frac{B_{0*}(q)}{\beta_*(q)} \frac{1}{\beta(q^{-1})}\quad (18)$$

and the equalized system response $\mathcal{H}_i^{\text{eq}}(q^{-1})$ at position i becomes

$$\mathcal{H}_i^{\text{eq}} = B_i \mathcal{R} \approx q^{-d} \frac{B_{0*}}{\beta_*} \frac{1}{\beta} B_i.\quad (19)$$

Note that B_{0*}/β_* can be expressed as a decaying series in positive powers of q and therefore its impulse response has a non-causal decay.

A second special case of particular interest occurs when $d = 0$. Equation (9) with degree $n_D = 0$ then gives that Q must have zero degree and $Q = \gamma_0$, with γ_0 obtained from (12), so that

$$\mathcal{R} = \frac{\gamma_0}{\beta}.\quad (20)$$

The equalized system at position i can then be expressed as

$$\mathcal{H}_i^{\text{eq}} = B_i \mathcal{R} = \gamma_0 \frac{B_i}{\beta}\quad (21)$$

whose impulse response decays casually only. We shall follow the common terminology of the field and refer to the filters in (18) and (20) as the *mixed phase* and *minimum phase* inverse filters, respectively.

IV. QUALITATIVE ASPECTS

Based on the analysis in the previous section, we now state some qualitative properties of the optimum filter for different scenarios, some of which are perceptually very important. The system and target dynamics are modeled as in Section III-B. That is, $\mathcal{H}_i(q^{-1}) = B_i(q^{-1})$, and $\mathcal{D}_i(q^{-1}) = D_i(q^{-1}) = q^{-d}$, with d being either zero or very large.

A. Single-Point or Anechoic Mixed Phase Design

Consider a situation where all p transfer functions are equal, so that $B_i = B_C$, $\forall i \in \{1, \dots, p\}$, for some ‘‘common’’ FIR transfer function polynomial $B_C(q^{-1})$. This is trivially true in a single-point design, i.e., $p = 1$. It may, however, also be approximately true for $p > 1$ if, for example, the transfer functions are acquired in the far-field of a loudspeaker in an anechoic chamber. In order to include this case, we assume $p \geq 1$. Given these assumptions, by (1) we have

$$B_0 = \sum_{i=1}^p B_C = p B_C\quad (22)$$

and from (7) we obtain

$$\beta_* \beta = \sum_{i=1}^p B_{i*} B_i = p B_{C*} B_C \implies \beta = \sqrt{p} \beta_C\quad (23)$$

where β_C is the minimum phase spectral factor of $B_{C*}B_C$, i.e., $\beta_{C*}\beta_C = B_{C*}B_C$. Thus, from (17)–(18) we obtain

$$Q \approx q^{-d} \sqrt{p} \frac{B_{C*}}{\beta_{C*}} \quad (24)$$

$$\mathcal{R} \approx q^{-d} \frac{B_{C*}}{\beta_{C*}} \frac{1}{\beta_C} \quad (25)$$

$$\begin{aligned} \mathcal{H}_i^{\text{eq}} &= B_i \mathcal{R} = B_C \mathcal{R} \\ &\approx q^{-d} \frac{B_{C*} B_C}{\beta_{C*} \beta_C} = q^{-d} = D_i. \end{aligned} \quad (26)$$

We note from (24) that Q approximates an all-pass filter (since the magnitude responses of B_C and β_C are identical) scaled by a constant \sqrt{p} , and the equalization in (26) is perfect. We recognize \mathcal{R} as the time-reversed and delayed excess phase part of B_C in series with the minimum phase inverse $1/\beta_C$. This case is in general of little practical interest and will not be further considered.

B. Multipoint Mixed Phase Design

In a multipoint design ($p \gg 1$) in a normal room, perfect equalization cannot be expected in any point due to the phase and magnitude variability among the RTFs. This variability causes the optimal filter to behave differently from the single-point/anechoic case, and its behavior can be quite problematic from a perceptual perspective. First, the filter polynomial $Q(q^{-1})$ no longer has all-pass character, as was the case in (24), because the magnitude responses $|B_0(e^{-j\omega})|$ and $|\beta(e^{-j\omega})|$ differ by more than a constant factor. To see this, suppose that at two separate frequencies ω_0 and ω_1 , the magnitudes of all RTFs are equal to one, $|B_i(e^{-j\omega_0})| = |B_i(e^{-j\omega_1})| = 1$, while the phases are equal at ω_0 , $\angle B_i(e^{-j\omega_0}) = \phi$, but random and uniformly distributed at ω_1 , $\angle B_i(e^{-j\omega_1}) = \phi_i \sim \mathcal{U}[0, 2\pi]$. Then $|\beta(e^{-j\omega_0})| = |\beta(e^{-j\omega_1})| = \sqrt{p}$, and $|B_0(e^{-j\omega_0})| = p$. However, due to phase cancellations at ω_1 we have $|B_0(e^{-j\omega_1})| \ll p$. Therefore, $|Q(e^{-j\omega_0})| = \sqrt{p}$, but $|Q(e^{-j\omega_1})| \ll \sqrt{p}$, and $|\mathcal{R}(e^{-j\omega_0})| = 1 \gg |\mathcal{R}(e^{-j\omega_1})|$. Hence, at frequencies where phase variability among the RTFs is large, the MSE optimal filter will attenuate the signal, resulting in a magnitude distortion not suitable for e.g., music listening. Second, the equalized responses $\mathcal{H}_i^{\text{eq}}$ of (19) will contain residual pre-rings, since the impulse response of B_{0*}/β_* decays noncausally. In Section V, we show that the two problems above are interconnected, and a remedy is proposed.

C. Minimum Phase Design

For the case $d = 0$, the filter $\mathcal{R} = \gamma_0/\beta$ is minimum phase and has the same character regardless of any possible similarities or dissimilarities among the RTFs. Perfect equalization is obtained only if all B_i are minimum phase and identical. By the strict causality of $\mathcal{H}_i^{\text{eq}}$ in (21), the minimum phase filter is guaranteed to generate no pre-ringing artifacts. Therefore, it has become common practice in loudspeaker equalizer design to use variants of this filter, with more or less sophisticated processing of the RMS average prior to inversion. It should be noted that there is a significant risk of introducing artificial post-rings

with this type of filter, since all notches in the average frequency response are inverted by minimum phase poles, regardless of whether they were caused by minimum phase zeros or not. We shall be using this filter type for comparison purposes in the experiments in Section VI.

V. TREATMENT OF THE PRE-RINGING PROBLEM

As stated in Section IV-B, the optimum multipoint mixed phase inverse causes residual pre-rings in the equalized system, due to the noncausal component B_{0*}/β_* in (19). We now analyze this further, and propose a remedy to alleviate the pre-rings. A key issue is the possible existence of common excess phase zeros, as is shown next.

A. The Origin of Pre-Ringing

Suppose that all RTFs share a common factor which is independent of spatial position. Each RTF $B_i(q^{-1})$ can then be decomposed into a common factor $B^c(q^{-1})$ and a non-common factor $B_i^n(q^{-1})$ as

$$B_i(q^{-1}) = B^c(q^{-1}) B_i^n(q^{-1}) \quad (27)$$

where the zeros of $B^c(z^{-1})$ are insensitive to spatial movements of the receiver, and the zeros of $B_i^n(z^{-1})$ are varying between different spatial positions. The corresponding decompositions of the complex and RMS spatial averages then become

$$B_0 = \sum_{i=1}^p B^c B_i^n = B^c \sum_{i=1}^p B_i^n = B^c B_0^n \quad (28)$$

$$\begin{aligned} \beta_* \beta &= \sum_{i=1}^p B_*^c B_{i*}^n B^c B_i^n \\ &= B_*^c B^c \sum_{i=1}^p B_{i*}^n B_i^n = \beta_*^c \beta^c \beta_*^n \beta^n \end{aligned} \quad (29)$$

where B_0^n is the complex spatial average of B_1^n, \dots, B_p^n , β^c is the minimum phase equivalent of B^c , and β^n is the RMS spatial average of B_1^n, \dots, B_p^n . Using (28) and (29) in (17) and (19) yields

$$Q \approx q^{-d} \frac{B_{0*}}{\beta_*} = q^{-d} \frac{B_*^c B_0^n}{\beta_*^c \beta_*^n} \quad (30)$$

$$\mathcal{H}_i^{\text{eq}} \approx q^{-d} \frac{B_*^c B_{0*} B^c B_i^n}{\beta_*^c \beta_*^n \beta^c \beta^n} = q^{-d} \frac{B_{0*} B_i^n}{\beta_*^n \beta^n}. \quad (31)$$

The part of (31) that causes the pre-rings is seen to be B_{0*}/β_*^n , which is a factor of $\tilde{Q}(q^{-1}, q)$ and therefore approximately contained in $Q(q^{-1})$ [cf. (17)]. Note that the noncausally decaying B_{0*}/β_*^n will always occur in the equalized system as soon as the RTFs have noncommon zeros, whether they be minimum phase, excess phase or both.

B. A Proposed Improvement

We now present a modified design of the compensator \mathcal{R} , by which the noncausally decaying factor B_{0*}/β_*^n is prevented from appearing in the equalized system. The modified compensator is derived by adding a constraint that the pre-ringing error at all outputs of the compensated system must be zero. For generality, we represent the system and desired response on the gen-

eral MFD form, and it is assumed that all subsystems \mathcal{H}_i share a common factor B^c , so that \mathcal{H} can be written

$$\begin{aligned} \mathcal{H} &= \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix} \frac{1}{A} = \begin{bmatrix} B_1^n B^c \\ \vdots \\ B_p^n B^c \end{bmatrix} \frac{1}{A} \\ &= \begin{bmatrix} B_1^n \beta^c \\ \vdots \\ B_p^n \beta^c \end{bmatrix} \frac{1}{A} \frac{B^c}{\beta^c}. \end{aligned} \quad (32)$$

Note that at least one of the polynomials B_i^n must have a nonzero leading coefficient, because otherwise q^{-1} would be a common factor of B_1^n, \dots, B_p^n and therefore of B_1, \dots, B_p . Since β^c is the minimum phase equivalent of B^c , all minimum phase zeros of B^c are cancelled by those of β^c in the rational function B^c/β^c . Therefore, we define a lower-order all-pass system $\mathcal{F}(q^{-1})$ as

$$\frac{B^c(q^{-1})}{\beta^c(q^{-1})} = \frac{\bar{F}(q^{-1})}{F(q^{-1})} \triangleq \mathcal{F}(q^{-1}) \quad (33)$$

where $F(q^{-1}) = f_0 + f_1 q^{-1} + \dots + f_{n_F} q^{-n_F}$ is a polynomial constructed by reflecting all excess phase zeros of $B^c(z^{-1})$ to the inside of the unit circle, and $\bar{F}(q^{-1})$ is the reciprocal of $F(q^{-1})$, defined as $\bar{F}(q^{-1}) = f_{n_F} + f_{n_F-1} q^{-1} + \dots + f_0 q^{-n_F}$. The desired response \mathcal{D} is represented as

$$\mathcal{D} = D/E = \begin{bmatrix} D_1 \\ \vdots \\ D_p \end{bmatrix} \frac{1}{E} = \begin{bmatrix} q^{-d} \tilde{D}_1 \\ \vdots \\ q^{-d} \tilde{D}_p \end{bmatrix} \frac{1}{E} \quad (34)$$

where $d \geq 0$ is the lowest power of q^{-1} appearing in any of the individual responses \mathcal{D}_i , and at least one of the polynomials \tilde{D}_i has a nonzero leading coefficient. In the following we shall assume that d is very large, so that the analysis can be performed on general noncausal filters, as in previous sections.

Definition 1: A stable, possibly noncausal, precompensator $\mathcal{R}(q^{-1}, q)$ is said to generate *pre-ringing errors* in the SIMO system $\mathcal{H}(q^{-1})$, if the impulse response $h_i^{\text{eq}}(k)$ at any of the p outputs of the compensated system $\mathcal{H}^{\text{eq}}(q^{-1}, q) = \mathcal{H}(q^{-1})\mathcal{R}(q^{-1}, q)$ is nonzero for any $k < d$, where d is as in (34).

By this definition, a compensated system has zero pre-ringing error if and only if the equalized system response $\mathcal{H}_i^{\text{eq}}$ at any output can be written as $\mathcal{H}_i^{\text{eq}}(q^{-1}) = q^{-d} X_i(q^{-1})/Y_i(q^{-1})$, where $X_i(q^{-1})/Y_i(q^{-1})$ is an arbitrary causal and stable transfer function.

Lemma 1: A stable, possibly noncausal, filter $\mathcal{R}(q^{-1}, q)$ generates no pre-ringing errors when applied to the SIMO system $\mathcal{H}(q^{-1})$ if and only if it can be written on the form

$$\begin{aligned} \mathcal{R}(q^{-1}, q) &= q^{-d} \frac{\bar{F}_*(q)}{F_*(q)} \mathcal{R}_1(q^{-1}) \\ &= q^{-d} \mathcal{F}_*(q) \mathcal{R}_1(q^{-1}) \end{aligned} \quad (35)$$

where $\mathcal{F}_*(q)$ is defined by (33), and $\mathcal{R}_1(q^{-1})$ is a stable and causal filter.

Proof: Sufficiency is easily proven by applying the suggested filter to an arbitrary subsystem $\mathcal{H}_i(q^{-1})$ and verifying

that the resulting transfer function $\mathcal{H}_i^{\text{eq}}(q^{-1})$ contains no powers of q^{-1} less than d :

$$\mathcal{H}_i^{\text{eq}}(q^{-1}) = q^{-d} \frac{B_i^n(q^{-1})\beta^c(q^{-1})}{A(q^{-1})} \mathcal{R}_1(q^{-1}) \quad (36)$$

The right-hand side of (36) has the form $q^{-d} X_i(q^{-1})/Y_i(q^{-1})$ and thus the pre-ringing error is zero. Necessity is proven as follows. First note that a general stable filter $\mathcal{S}(q^{-1}, q)$ can be written as

$$\mathcal{S}(q^{-1}, q) = q^{-d} \mathcal{F}_*(q) \mathcal{R}_2(q^{-1}, q) \quad (37)$$

where $\mathcal{R}_2(q^{-1}, q)$ is an arbitrary stable filter, and the all-pass factor $q^{-d} \mathcal{F}_*(q)$ may be cancelled, either completely or in part, by some factors of \mathcal{R}_2 . If \mathcal{R}_2 is causal then (37) is equivalent to (35), so the proof is complete if we can show that \mathcal{S} generates no pre-ringing error only if \mathcal{R}_2 is causal. Therefore, suppose that \mathcal{S} in (37) generates no pre-ringing error when applied to (32). We decompose \mathcal{R}_2 into noncausal and causal parts as

$$\mathcal{R}_2(q^{-1}, q) = \frac{qR_{1*}(q)}{S_{1*}(q)} + \frac{R_2(q^{-1})}{S_2(q^{-1})} \quad (38)$$

where $S_1(q^{-1})$ and $S_2(q^{-1})$ are monic and stable polynomials, and qR_{1*} and S_{1*} are coprime, as are R_2 and S_2 . At output i , the compensated system $\mathcal{H}^{\text{eq}}(q^{-1}) = \mathcal{H}(q^{-1})\mathcal{S}(q^{-1}, q)$ is then

$$\mathcal{H}_i^{\text{eq}} = \frac{B_i^n \beta^c}{A} \frac{B^c}{\beta^c} q^{-d} \mathcal{F}_* \left[\frac{qR_{1*}}{S_{1*}} + \frac{R_2}{S_2} \right] = q^{-d} \frac{X_i}{Y_i} \quad (39)$$

where $X_i(q^{-1})/Y_i(q^{-1})$ is a causal and stable transfer function. The last equality in (39) comes from the requirement of zero pre-ringing error. Simplifying and rearranging (39) yields

$$B_i^n \beta^c [qR_{1*}S_2 + R_2S_{1*}]Y_i = X_i S_{1*} S_2 A. \quad (40)$$

In order for (40) to hold, S_{1*} must be a factor of $B_i^n \beta^c qR_{1*}S_2Y_i$. But S_{1*} has all its zeros outside the unit circle while $\beta^c S_2 Y_i$ has all its zeros on the inside, and qR_{1*} and S_{1*} are by definition coprime. The remaining alternative is that S_{1*} is a factor of B_i^n . But if (40) is to hold for all $i \in \{1, \dots, p\}$, B_i^n cannot contain S_{1*} , since then S_{1*} would be a common factor and thus belong to B^c . Therefore, $S_{1*} = 1$. Substitution of $S_{1*} = 1$ in (40) yields

$$B_i^n \beta^c qR_{1*}S_2Y_i = X_i S_2 A - B_i^n \beta^c R_2 \quad (41)$$

Since the polynomials β^c , S_2 , Y_i , and at least one of the p polynomials B_i^n have a nonzero leading coefficient, the left-hand side of (41) is a polynomial in q and q^{-1} . The right-hand side of (41) is, however, a polynomial in q^{-1} only and therefore R_{1*} must be zero, which proves that \mathcal{R}_2 is causal. ■

The constrained MSE optimal compensator is now given by the following theorem.

Theorem 1: If the SIMO system $\mathcal{H}(q^{-1})$ is given by (32) and the desired responses are given by (34), then the noncausal precompensator $\mathcal{R}(q^{-1}, q)$ that minimizes (5) under the constraint of zero pre-ringing error in $\mathcal{H}^{\text{eq}}(q^{-1}) = \mathcal{H}(q^{-1})\mathcal{R}(q^{-1}, q)$, is given by

$$\mathcal{R}(q^{-1}, q) = q^{-d} \mathcal{F}_*(q) \frac{Q(q^{-1})A(q^{-1})}{\beta(q^{-1})E(q^{-1})} \quad (42)$$

where $\mathcal{F}_*(q)$ is defined by (33), $\beta(q^{-1})$ is defined by (7), and $Q(q^{-1})$ along with a polynomial $L_*(q^{-1})$ are given by the Diophantine equation

$$\beta_*^c(q) \sum_{i=1}^p B_{i*}^n(q) \tilde{D}_i = \beta_*(q) Q(q^{-1}) + q L_*(q) E(q^{-1}). \quad (43)$$

Proof: By Lemma 1, the class of all stable and linear filters which do not generate pre-ringing errors can be written on the form (35). Allowing filters from this class only, and recalling the representation (32) of \mathcal{H} , the error signal in (4) can be written as

$$\begin{aligned} \mathbf{y}(k) &= (\mathcal{D} - \mathcal{H}q^{-d}\mathcal{F}_*\mathcal{R}_1)w(k) \\ &= (\mathcal{D} - \tilde{\mathcal{H}}\mathcal{R}_1)w(k) \end{aligned} \quad (44)$$

where $\tilde{\mathcal{H}}$ is the modified SIMO system

$$\begin{aligned} \tilde{\mathcal{H}}(q^{-1}) &= \tilde{\mathbf{B}}/A \\ &= \begin{bmatrix} q^{-d}B_1^n(q^{-1})\beta^c(q^{-1}) \\ \vdots \\ q^{-d}B_p^n(q^{-1})\beta^c(q^{-1}) \end{bmatrix} \frac{1}{A(q^{-1})}. \end{aligned} \quad (45)$$

Now with \mathcal{D} as in (34), the constrained mixed phase equalization problem consists in finding the causal filter \mathcal{R}_1 in (44) which minimizes the criterion (5). A reuse of the results in Subsection III-A gives

$$\mathcal{R}_1(q^{-1}) = \frac{Q(q^{-1})A(q^{-1})}{\beta(q^{-1})E(q^{-1})} \quad (46)$$

where $\beta_*\beta = \tilde{\mathbf{B}}_*\tilde{\mathbf{B}} = \mathbf{B}_*\mathbf{B}$ as before [cf. (7), and (29)]. Q and L are given by the Diophantine equation

$$\tilde{\mathbf{B}}_*(q)\mathcal{D}(q^{-1}) = \beta_*(q)Q(q^{-1}) + qL_*(q)E(q^{-1}) \quad (47)$$

which is equivalent to (43). The optimal constrained compensator of (42) is obtained by substituting (46) in the representation (35) of \mathcal{R} . ■

C. The Constrained Mixed Phase Compensator in Practice

Of course, one cannot expect it to occur in practice that all B_i share a truly common excess phase part $\mathcal{F} = B^c/\beta^c$. Nevertheless, in [9] it was demonstrated by the authors that an approximately common excess phase part can be found by detection of zero clusters in the set of RTFs. With the clusters represented by *nominal zeros* located at the cluster centra, a cluster is classified as invertible if the pre-ringing that results from placing a pole at the nominal zero location is kept below a pre-defined envelope constraint. We now relate this concept to the present work by using the excess phase zeros of B_0 in (28) as nominal zeros. In order to construct the part \mathcal{F}_* of the constrained compensator (42), the near-common all-pass factor B^c/β^c has to be found. This is equivalent to finding the excess phase zeros of B^c . In the case of exactly common zeros, this can be accomplished by discarding all zeros of B_0 which are not common to all B_i . For this argument to be transferable to the case when zeros are only near-common, we need to know whether the zero clusters of B_1, \dots, B_p are represented by zeros in B_0 which in some sense are close to the zero clusters. An empirical verifi-

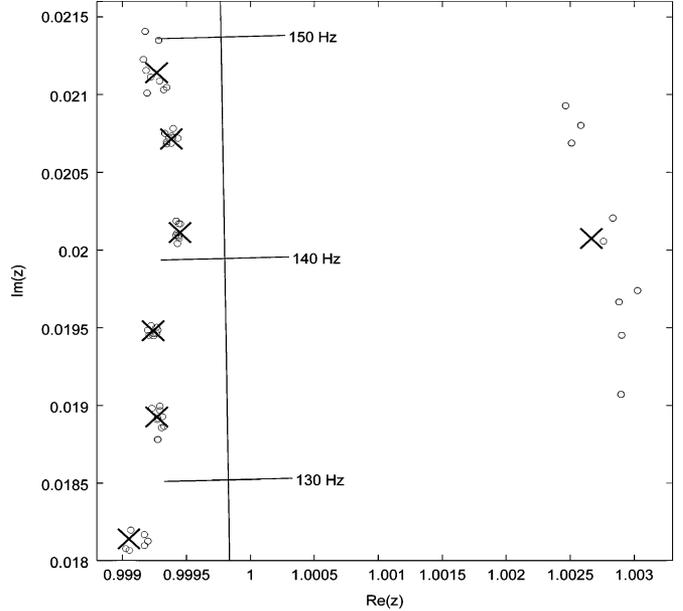


Fig. 2. Segment of the complex plane near the unit circle, showing the zeros of nine RTFs B_1, \dots, B_9 , marked as “o”, and of their complex spatial average B_0 , marked as “x”. The grouping of the RTF zeros into well separated clusters (each cluster containing one zero from each B_i ; $i = 1, \dots, 9$) clearly has the averaging effect on the zeros of B_0 , as predicted by (49).

cation of this property is provided in Fig. 2 where the zeros of B_0 are located approximately at the center of each zero cluster. While a rigorous proof of this property may be quite involved, we motivate it here by a heuristic argument as follows. Let $B_i(z^{-1})$, $i \in \{1, \dots, p\}$ represent the individual RTFs, and let $B_0(z^{-1})$ be the coefficient-wise sum of all $B_i(z^{-1})$. Suppose that there is a complex number z_0 and a small neighborhood $N_\epsilon(z_0)$ around it, such that each $B_i(z^{-1})$ has a zero z_i within $N_\epsilon(z_0)$. Define the polynomials G_i by factoring out the zero z_i as $B_i = (z - z_i)G_i$. Then

$$\begin{aligned} B_0(z^{-1}) &= \sum_{i=1}^p B_i = \sum_{i=1}^p (z - z_i)G_i \\ &= z \sum_{i=1}^p G_i - \sum_{i=1}^p z_i G_i \\ &= \left(z - \frac{\sum_{i=1}^p z_i G_i}{\sum_{i=1}^p G_i} \right) \sum_{i=1}^p G_i. \end{aligned} \quad (48)$$

Suppose further that the zero cluster contained in $N_\epsilon(z_0)$ is well separated from all other zeros of B_1, \dots, B_p so that the polynomials G_i do not contain zeros in the vicinity of $N_\epsilon(z_0)$. Then each G_i can be approximated by a constant for all $z \in N_\epsilon(z_0)$, so that $G_i(z^{-1}) \approx g_i$, $\forall z \in N_\epsilon(z_0)$. This corresponds to replacing $G_i(z^{-1})$ with the first term of its Taylor expansion around z_0 . We then obtain

$$B_0(z^{-1}) \approx \left(z - \frac{\sum_{i=1}^p z_i g_i}{\sum_{i=1}^p g_i} \right) \sum_{i=1}^p g_i, \quad \forall z \in N_\epsilon(z_0) \quad (49)$$

i.e., the polynomial $B_0(z^{-1})$ has a zero which is a weighted average of the zero locations z_i of the individual polynomials

$B_i(z^{-1})$. The near-common excess phase zeros of B_1, \dots, B_p can hence be found by inspecting each excess phase zero of B_0 and requiring it to be located within a cluster containing one zero of each B_i . It is intuitively clear that if a zero cluster is small enough, the corresponding zero of B_0 should be regarded as belonging to B^c in (32). Upon inversion of B^c , the remaining mismatch between B^c and the true zeros of B_1, \dots, B_p then causes pre-rings with negligible amplitudes. In the next subsection we establish a relation between zero cluster size and pre-ringing amplitude.

D. Quantification of Pre-Ringing Error

Suppose that a noncausal filter with transfer function $\mathcal{R}(z^{-1}, z)$ has been designed to be the inverse of a system $\mathcal{H}(z^{-1})$, but with a small mismatch, so that the excess phase poles of $\mathcal{R}(z^{-1}, z)$ do not completely cancel the excess phase zeros of $\mathcal{H}(z^{-1})$. The residual pre-ringing that results can be quantified as follows.

Let a zero of $\mathcal{H}(z^{-1})$ be represented by $z_0 = r_0 e^{j\omega_0}$ and a perturbation to this zero by $\epsilon = \rho e^{j\theta}$ where $r_0 > 1$; $0 < \rho \ll 1$; $0 < \omega_0 < \pi$; $-\pi \leq \theta \leq \pi$. Suppose that $\mathcal{H}(z^{-1})$ contains a complex conjugate pair of zeros at $z_0 + \epsilon$ and $\overline{z_0 + \epsilon}$, so that

$$\mathcal{H}(z^{-1}) = \mathcal{H}_1 \mathcal{H}_2 = (z - (z_0 + \epsilon))(z - \overline{(z_0 + \epsilon)}) \mathcal{H}_2. \quad (50)$$

Furthermore, suppose that the compensator $\mathcal{R}(z^{-1}, z)$ contains the pole pair z_0 and $\overline{z_0}$,

$$\mathcal{R}(z^{-1}, z) = \mathcal{R}_1 \mathcal{R}_2 = \frac{1}{(z - z_0)(z - \overline{z_0})} \mathcal{R}_2. \quad (51)$$

The total transfer function $\mathcal{H}^{\text{eq}}(z^{-1}, z)$ of the equalized system thus becomes [9]

$$\begin{aligned} \mathcal{H}^{\text{eq}}(z^{-1}, z) &= \frac{(z - (z_0 + \epsilon))(z - \overline{(z_0 + \epsilon)})}{(z - z_0)(z - \overline{z_0})} \mathcal{R}_2 \mathcal{H}_2 \\ &= \left(1 - \frac{(\epsilon + \bar{\epsilon})z - (\epsilon \overline{z_0} + \bar{\epsilon} z_0 + \epsilon \bar{\epsilon})}{(z - z_0)(z - \overline{z_0})} \right) \mathcal{R}_2 \mathcal{H}_2 \\ &= \left(1 - 2\Re(\epsilon) \frac{z - \frac{|z_0 + \epsilon|^2 - |z_0|^2}{2\Re(\epsilon)}}{(z - z_0)(z - \overline{z_0})} \right) \mathcal{R}_2 \mathcal{H}_2. \end{aligned} \quad (52)$$

Applying the inverse z -transform on each factor in the last line of (52) yields the total impulse response,

$$h^{\text{eq}}(k) = [\delta(k) + C r_0^k \cos(-\omega_0 k + \Phi) u(-k)] * r_2(k) * h_2(k) \quad (53)$$

where $*$ denotes convolution, $\delta(k)$ is the Kronecker delta function, $u(k)$ is the unit step function, and

$$\Phi = \arctan \left(\frac{2\Re(\epsilon) |z_0|^2}{|z_0 + \epsilon|^2 - |z_0|^2} - \Re(z_0) \right) \quad (54)$$

$$C = \frac{|z_0 + \epsilon|^2 - |z_0|^2}{|z_0|^2 \cos \Phi}. \quad (55)$$

In (52) and (54), we have used the assumptions that $\Re(\epsilon) \neq 0$, and $|z_0 + \epsilon|^2 \neq |z_0|^2$, which are reasonable for measured data. Equation (53) clearly shows how the pole/zero mismatch ϵ

between $\mathcal{H}(z^{-1})$ and $\mathcal{R}(z^{-1}, z)$ has created a noncausal ringing which affects the total system in a convolutive way.

Suppose now that $B_i(z^{-1})$; $i \in \{1, \dots, p\}$ represent the set of p RTFs in $\mathcal{H}(z^{-1})$, each containing M zeros z_{im} ; $m \in \{1, \dots, M\}$. Furthermore suppose that these zeros are expressed as perturbations, $z_{im} = z_{0m} + \epsilon_{im}$, of the nominal zeros z_{0m} ; $m \in \{1, \dots, M_0, M_0 + 1, \dots, M\}$, where the first M_0 nominal zeros are located outside the unit circle in the upper half plane. Once the nominal zeros $z_{0m} = r_{0m} e^{j\omega_{0m}}$ and their perturbations $\epsilon_{im} = \rho_{im} e^{j\theta_{im}}$ have been determined, (53)–(55) with obvious modifications can be used to determine the maximum amplitudes C_1, \dots, C_{M_0} of the residual pre-rings caused by placing poles at the nominal zero locations z_{01}, \dots, z_{0M_0} and their conjugated counterparts $\overline{z_{01}}, \dots, \overline{z_{0M_0}}$. We saw in the previous subsection that the excess phase zeros of B_0 may be used as nominal zeros z_{0m} . Given that all excess phase zeros are available, it remains to associate each z_{0m} with a zero cluster of as small a size a possible.

E. Extraction of Excess Phase Zeros

Suppose that a set of p RTFs $B_i(q^{-1})$; $i \in \{1, \dots, p\}$ has been acquired within the listening region. In order to apply the method of the previous subsection, the excess phase zeros of all B_i and of the complex average model B_0 are required. Considering that the polynomial degree is typically on the order of 10 000–20 000 for FIR models representing full-bandwidth RTFs, finding their zeros is a nontrivial task. However, since only the excess phase zeros of B_0 and B_1, \dots, B_p are sought, they can be found indirectly by identifying the poles of the all-pass sequences $\tilde{b}_0(k), \tilde{b}_1(k), \dots, \tilde{b}_p(k)$ defined by the excess phase parts of B_0, B_1, \dots, B_p as

$$\begin{aligned} \tilde{b}_0(k) &= \frac{B_0(q^{-1})}{\beta_0(q^{-1})} \delta(k); \\ \tilde{b}_i(k) &= \frac{B_i(q^{-1})}{\beta_i(q^{-1})} \delta(k). \end{aligned} \quad (56)$$

The excess phase zeros are then found as the conjugate reciprocals of the pole positions. Note that in $\tilde{b}_0(k)$ and $\tilde{b}_i(k)$ the minimum phase factors of B_0 and B_i are cancelled by corresponding factors in β_0 and β_i respectively, and the number of poles in $\tilde{b}_0(k)$ and $\tilde{b}_i(k)$ is therefore low compared to the polynomial degrees of B_0 and B_i . The polynomials β_0 and β_i in (56) can be computed with a suitable spectral factorization algorithm [12], and the poles of $\tilde{b}_0(k)$ and $\tilde{b}_i(k)$ are then found by performing a model reduction on the systems B_0/β_0 and B_i/β_i ; see, e.g., [13].

F. A Pre-Ringing Constraint

With all excess phase zeros given, the next step is to see whether the nominal zeros of B_0 can be associated with zero clusters of sufficiently small size. ‘‘Sufficiently small’’ here means that the pre-ringing caused by inverting the cluster with a pole at the nominal zero location should not exceed a pre-specified envelope at any control point. If q^{-d} is the desired system delay included in $\mathcal{D}(q^{-1})$, pre-rings are defined as nonzero values in the equalized system impulse response,

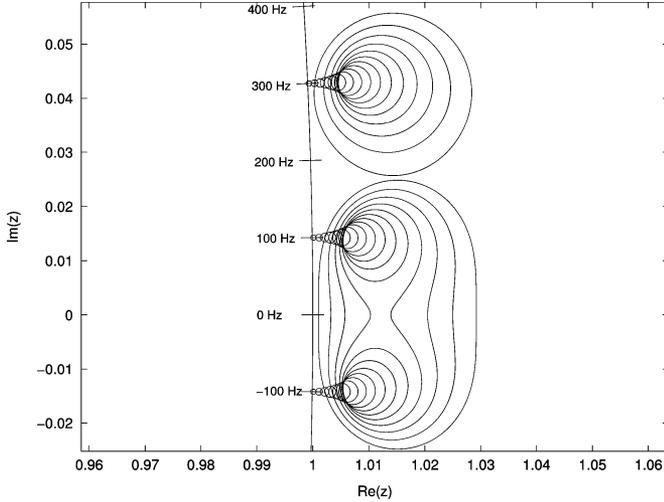


Fig. 3. Regions in the complex plane defining the maximum tolerable zero cluster size for different nominal zero locations, given the pre-ringing envelope constraint (57) with $L_{\max} = -60$ dB and $\kappa = 220$ samples. Each circle has an associated nominal zero z_0 located approximately at its center.

$|h^{\text{eq}}(k)| > 0$, for time indexes $k < d$. We define the maximum tolerable pre-ringing by an exponential envelope constraint as

$$20 \log_{10}(Cr_0^{-\kappa}) < L_{\max} \quad (57)$$

where C and r_0 are as in (53), $\kappa \in \mathbb{Z}^+$ is a time constant and L_{\max} is the maximum tolerable pre-ringing level in dB, in the equalized response $h^{\text{eq}}(k)$ at time index $k = d - \kappa$. This constraint ensures that the pre-ringing level is at most L_{\max} dB at all time instants prior to $k = d - \kappa$. Given a nominal zero z_0 , (54)–(55) along with the pre-ringing envelope constraint (57) implicitly define a region around z_0 (and \bar{z}_0) within which a zero cluster (and its conjugated counterpart) must be contained in order for z_0 to be considered a common, safely invertible zero of all RTFs. Fig. 3 shows the contours of such regions for different values of z_0 . We note from Fig. 3 that the zero clusters are allowed to be larger if the nominal zero z_0 is located far away from the unit circle, than when z_0 is close to the unit circle.

G. Clustering of Near-Common Excess Phase Zeros

We will now describe an algorithm for sorting the excess phase zeros of B_1, \dots, B_p into separated clusters, centered around the excess phase zeros of B_0 . The requirement that each cluster must contain exactly one zero from each B_i makes this problem somewhat different from the typical clustering problems encountered in image analysis, data mining etc. No standard off-the-shelf method has been found to be applicable, so the algorithm has been constructed with this specific application in mind. We start with some preliminaries. Suppose that B_0 contains M_o zeros outside the unit circle in the upper half plane, and that each B_i contains K_o^i such zeros. Further, assume that $M_o \leq K_o^i \quad \forall i$. Now arrange these zeros into the sets denoted $\tilde{\mathcal{Z}}_0$ and \mathcal{Z}_i respectively:

$$\tilde{\mathcal{Z}}_0 = \{z_{01}, \dots, z_{0M_o}\} \quad (58)$$

$$\mathcal{Z}_i = \{z_{i1}, \dots, z_{iK_o^i}\}, \quad i \in \{1, \dots, p\}. \quad (59)$$

The aim of the clustering algorithm is to associate each nominal zero $z_{0m} \in \tilde{\mathcal{Z}}_0$ from B_0 with one zero $z_{ik} \in \mathcal{Z}_i$ from each B_i . Thereby the zeros are sorted into clusters \mathcal{C}_m , defined as

$$\mathcal{C}_m = \{z_{1k_m^1}, z_{2k_m^2}, \dots, z_{pk_m^p}\}, \quad m \in \{1, \dots, M_o\} \quad (60)$$

where the indexes k_m^i determine which of the zeros $z_{i1}, \dots, z_{iK_o^i}$ in \mathcal{Z}_i is to be associated with a certain nominal zero z_{0m} . We will also make use of a set $\tilde{\mathcal{Z}}_0$, along with an index set μ , defined as

$$\mu = \{\mu_1, \dots, \mu_{\tilde{M}}\} \subset \{1, \dots, M_o\} \quad (61)$$

$$\tilde{\mathcal{Z}}_0 = \{z_{0\mu_1}, \dots, z_{0\mu_{\tilde{M}}}\} \subset \{z_{01}, \dots, z_{0M_o}\} = \tilde{\mathcal{Z}}_0 \quad (62)$$

where μ is always ordered, i.e., $\mu_j < \mu_{j+1}$, $j = 1, \dots, \tilde{M} - 1$. Note that \tilde{M} is the number of elements in μ and $\tilde{\mathcal{Z}}_0$, which varies between different passes through the algorithm. The algorithm is greedy in the sense that, by a principle of “mutually nearest neighbors”, it prioritizes dense and well separated clusters instead of minimizing a global criterion based on average distances, as is often the case with other clustering algorithms. The algorithm is described in pseudo code as follows.

Zero Clustering Algorithm

```

for  $m = 1$  to  $M_o$  do
   $\mathcal{C}_m \leftarrow \emptyset$ ;
end for
for  $i = 1$  to  $p$  do
   $\tilde{\mathcal{Z}}_0 \leftarrow \tilde{\mathcal{Z}}_0$ ;  $\mathcal{X}_0 \leftarrow \emptyset$ ;  $\mu \leftarrow \{1, \dots, M_o\}$ ;  $\xi \leftarrow \emptyset$ ;
  repeat
    for  $j = 1$  to  $\tilde{M}$  do
       $m \leftarrow \mu_j$ 
      Let  $z_{ik_m^i}$  be the zero in  $\mathcal{Z}_i$  closest to  $z_{0m}$ ;
      Let  $z_{0m_k^i}$  be the zero in  $\tilde{\mathcal{Z}}_0$  closest to  $z_{ik_m^i}$ ;
      if  $z_{0m_k^i} = z_{0m}$ 
        Add  $z_{ik_m^i}$  to  $\mathcal{C}_m$ :  $\mathcal{C}_m \leftarrow \mathcal{C}_m \cup \{z_{ik_m^i}\}$ ;
        Remove  $z_{ik_m^i}$  from  $\mathcal{Z}_i$ :  $\mathcal{Z}_i \leftarrow \mathcal{Z}_i \setminus \{z_{ik_m^i}\}$ ;
      else
        Add  $z_{0m}$  to  $\mathcal{X}_0$ :  $\mathcal{X}_0 \leftarrow \mathcal{X}_0 \cup \{z_{0m}\}$ ;
        Add  $m$  to  $\xi$ :  $\xi \leftarrow \xi \cup \{m\}$ ;
      end if
    end for
     $\tilde{\mathcal{Z}}_0 \leftarrow \mathcal{X}_0$ ;
     $\mu \leftarrow \xi$ ;
  until  $\tilde{\mathcal{Z}}_0 = \emptyset$ ;
end for

```

With the zeros $z_{ik_m^i}$ of each cluster \mathcal{C}_m expressed as perturbed nominal zeros, $z_{ik_m^i} = z_{0m} + \epsilon_{k_m^i}$, one can employ (54), (55) and (57) to decide which zeros in $\tilde{\mathcal{Z}}_0$ should be included in the inverted common all-pass factor \mathcal{F}_* of (42).

H. Smoothing of the RMS Spatial Average

In a practical filter design, the number p of transfer function measurements will be limited. Therefore, the RMS spatial average β as defined in (7) will not represent the true RMS average for all possible listener positions, and the filter will be op-

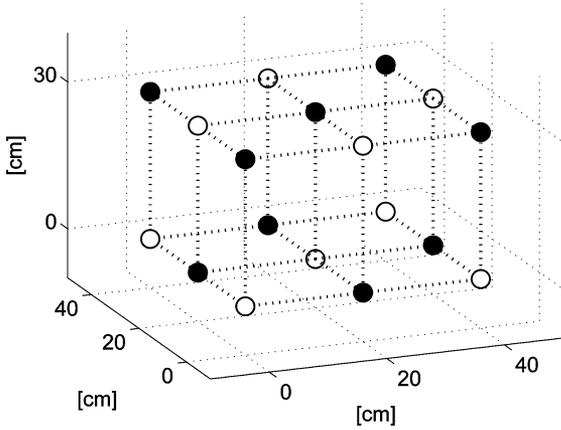


Fig. 4. Geometry of microphone positions for filter design (white) and validation (black).

timal only with respect to the actual measurement positions. For the filter to be truly robust, a method is needed which estimates the true RMS average from a limited number of RTFs. In the present work, this problem has been treated by a smoothing of the frequency response of the finite-sample RMS average, using a 1/6th octave resolution. We motivate this operation by the fact that local irregularities in the RMS frequency response are expected to be smoothed out as the number of RTFs tends to infinity. In Section VI, the benefit of such smoothing is confirmed. A further improved performance is, however, anticipated with a refined design of the smoothing operation. Another practical issue with influence on the filter design is the bandlimited nature of most loudspeakers. This can be treated by an “amplitude regularization” [14] at the extreme ends of the frequency spectrum, in order to prevent the inverse filter from boosting frequencies outside the working range of the loudspeaker. In our feedforward control setting, such regularization can be included in the design by introducing an extra penalty term in the criterion (5) as

$$J' = E \{ \|\mathbf{y}(k)\|_2^2 + \|Wu(k)\|_2^2 \} \quad (63)$$

where $W(q^{-1})$ is a weighting polynomial, and $u(k)$ is the control signal, $u(k) = \mathcal{R}(q^{-1})w(k)$. This leads to a change in the expression (7) for $\beta(q^{-1})$ (see [8] for details)

$$\beta_*\beta = \mathbf{B}_*\mathbf{B} + A_*W_*WA. \quad (64)$$

VI. A DESIGN EXAMPLE

In this section, we compare the performance of six different equalizer filters designed using the methods of the previous sections. The target dynamics was in all cases set to $\mathbf{D} = [q^{-d} \dots q^{-d}]^T$, with either $d = 0$ or $d = 4096$, depending on whether the filter is to be minimum or mixed phase. The filters will be referred to with letters from A to F, and they were designed as follows.

- A) The MSE optimal mixed phase ($d = 4096$ samples) filter of (18), without any smoothing or regularization of the RMS spatial average.
- B) The constrained mixed phase ($d = 4096$ samples) filter of (42), without smoothing or regularization.

- C) The minimum phase ($d = 0$) filter of (20), without smoothing or regularization.
- D) Same as filter A, but with smoothing and regularization of the frequency response of the RMS spatial average prior to computing β . Smoothing resolution was 1/6th octave, and regularization was used below 30 Hz and above 20 kHz.
- E) Same as filter B but with the same smoothing and regularization as filter D.
- F) Same as filter C but with the same smoothing and regularization as filter D. Filter F represents the “standard” minimum phase approach to loudspeaker equalization.

A. Methods for Evaluation

The performance of a filter will be assessed by studying simulated responses of the equalized system at different control points. These responses are obtained by applying the filter $\mathcal{R}(q^{-1})$ to the impulse responses of the RTFs in question:

$$h_n^{\text{eq}}(k) = B_n(q^{-1})\mathcal{R}(q^{-1})\delta(k), \quad n \in \{1, \dots, N\}. \quad (65)$$

Hence, here we rely on the assumption of linearity and time-invariance of the true system, i.e., that the simulated equalized response $h_n^{\text{eq}}(k)$ is equal to that obtained by a real RTF measurement of the system at position n , using a test signal pre-filtered with $\mathcal{R}(q^{-1})$. Robustness is assessed by comparing the performance for two different sets of RTFs. The first set is the *design set* containing p RTFs which represent the control points that were used for filter design. The second set is the *validation set*, representing control points within the listening region, but spatially separated from the design set (see Fig. 4). Such a comparison indicates to what extent the filters are over-fitted to the design points. Since our proposed modified design is based primarily on a time domain argument (avoidance of pre-rings), the assessment will focus on the time domain behavior of the filters. We will, however, start by presenting the RMS average frequency responses of the system, before and after equalization. The RMS average frequency response is a frequency domain representation of the RMS spatial average model, as defined in (2). The property of interest in the frequency domain is the amount of spectral flattening achieved by the different filters, and it should apply in both the design and validation points. Whenever spectral flattening in the design points comes at the expense of increased spectral distortion in the validation points, the filter is regarded as overfitted to the design points and therefore nonrobust. For graphic evaluation of the time-domain properties, we shall use the average Schroeder decay sequence $\bar{D}(k)$, the average energy step response, or energy build-up, $S(k)$ and the impulse response maximum level envelope $L(k)$

$$\bar{D}(k) = 10 \log_{10} \left(\frac{1}{N} \sum_{n=1}^N \sum_{l=k}^{M-1} \frac{h_n^2(l)}{\sum_{m=0}^{M-1} h_n^2(m)} \right) \quad (66)$$

$$S(k) = \frac{1}{N} \sum_{n=1}^N \sum_{l=0}^k \frac{h_n^2(l)}{\sum_{m=0}^{M-1} h_n^2(m)} \quad (67)$$

$$L(k) = 20 \log_{10} \left(\max_n |h_n(k)| \right) \quad (68)$$

defined in (66), (67), and (68), respectively. The Schroeder decay and energy build-up curves were introduced in [15] and [16], respectively. Here $h_n(k)$; $k \in \{0, \dots, M-1\}$ is an impulse response of length M in microphone position

n ; $n \in \{1, \dots, N\}$. Prior to computation of $D(k)$, $S(k)$, and $L(k)$, all responses are time-aligned and normalized so that $\max |h_n(k)| = |h_n(k_0)| = 1$ for some time instant $k = k_0$. While $L(k)$ is useful as a worst case presentation of pre- or post-ringing problems, $S(k)$ and $D(k)$ indicate how good are the transient properties of the system. In order for a comparison of systems with different pre-ringing behavior to be feasible, a further alignment of the curves $D(k)$ and $S(k)$ is needed. We have chosen to define the starting time, $k = 0$, of $S(k)$ so that $k = 1$ occurs at the sample where $S(k)$ for the first time reaches above 5% of its steady state value. For $D(k)$, we define $k = 0$ so that $k = 1$ occurs at the sample where the decay for the first time reaches below -0.5 dB. It is sometimes instructive to see how the curves $D(k)$ and $S(k)$ behave in narrow frequency bands, and we shall therefore complement the full frequency band presentations with low pass filtered versions, with a cutoff frequency of 320 Hz.

B. Experimental Conditions

In a room of dimensions $4.5 \times 6 \times 2.6$ m and an average distance between loudspeaker and microphones equal to 2.5 m, nine measurement positions for filter design ($p = 9$), and nine positions for validation were selected according to Fig. 4. This microphone configuration was designed to cover typical head movements of a normal listener. The RTFs were acquired using a pink-colored random phase multisine signal [17, Ch. 13] with a period time of 3 s. The FIR models so obtained were truncated to a length of 0.408 s, or 18 000 coefficients at a sampling frequency of 44 100 Hz. This model order is motivated by the reverberation time T_{60} of the room which is slightly less than 0.4 s. Filters A to F were then designed as described in the beginning of this section. The parameters in the pre-ringing constraint (57) were set to $L_{\max} = -60$ dB and $\kappa = 220$ samples. This value of κ corresponds to 5 ms, and it was chosen because pre-rings are unlikely to be audible for shorter delays. The minimum phase polynomials $\beta(q^{-1})$, $\beta_0(q^{-1})$ and $\beta_i(q^{-1})$ were obtained by spectral factorization [12], and the poles of the sequences $\tilde{b}_0(k)$ and $\tilde{b}_i(k)$ in (56) were identified using a Hankel matrix based model reduction technique [13]. The accuracy of this method for finding excess phase zeros has been found to be reasonably good when compared to a brute-force polynomial rooting approach. A deeper study of the accuracy of this method is unfortunately beyond the scope of the present paper.

C. Results

In this subsection, we present graphically the time and frequency domain performance of the filters A to F. We begin by stating some properties that are evident from the frequency responses of Fig. 5.

- Filters A–C perform as desired in the design points, but not in the validation points. Although the general trends in the frequency responses are corrected in the validation points also, the filters A–C seem to cause an increased jaggedness of the curves at high frequencies. Filters D–F do not introduce such artifacts.
- The “attenuation property” of the MSE optimal filter, discussed in Subsection IV-B, is evident in the frequency responses of filters A and D. The deep notches at 190, 280,

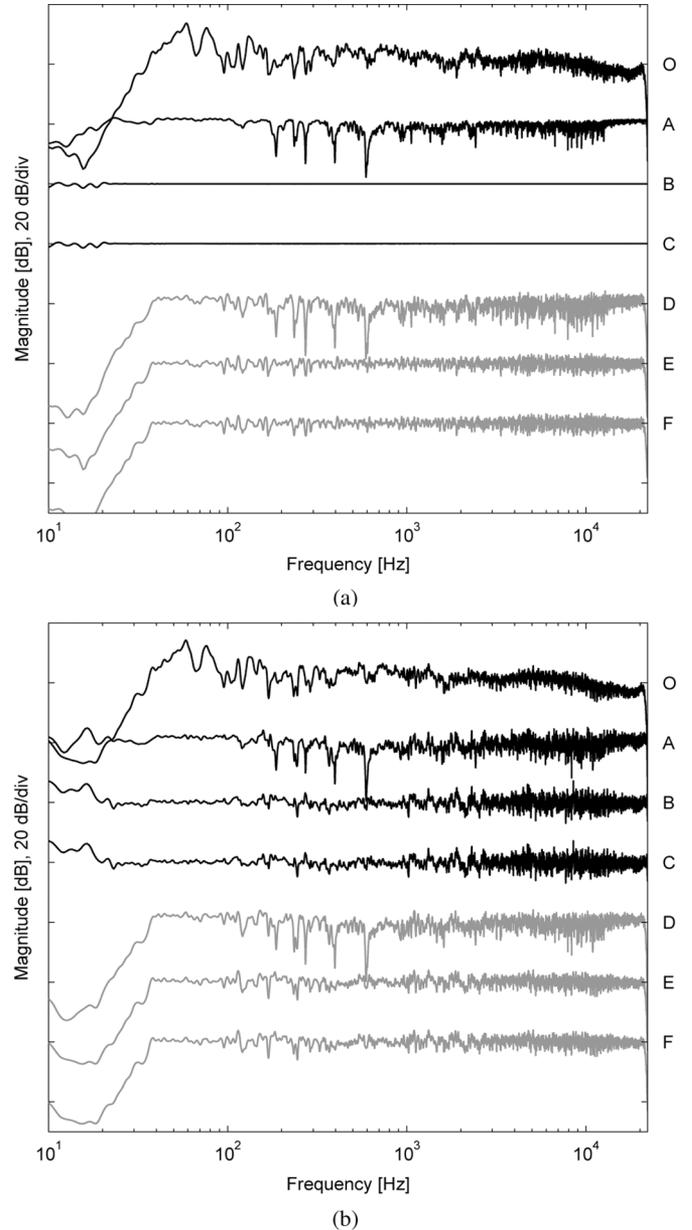


Fig. 5. RMS average frequency responses of original and equalized system for filters A–F. (a) Performance in design points. (b) Performance in validation points. Original responses are marked with the letter “O”.

400, and 600 Hz indicate a large phase variability at those frequencies among the RTFs in the listening region.

- In the frequency region between about 30 and 200 Hz, the “unsmoothed” filters A–C achieve a greater amount of spectral flattening, even in the validation points, than do the smoothed and regularized filters D–F. This suggests that the 1/6th octave smoothing that was applied prior to computing β in the design of filters D–F may be too coarse at those frequencies. That is, the smoothing has removed too much of the low-frequency details in β . A better performance at low frequencies could thus be expected with a more flexible smoothing operation.
- The most desirable overall frequency domain performance is exhibited by filters E and F, which flatten out the response without adding any strange properties to the curves.

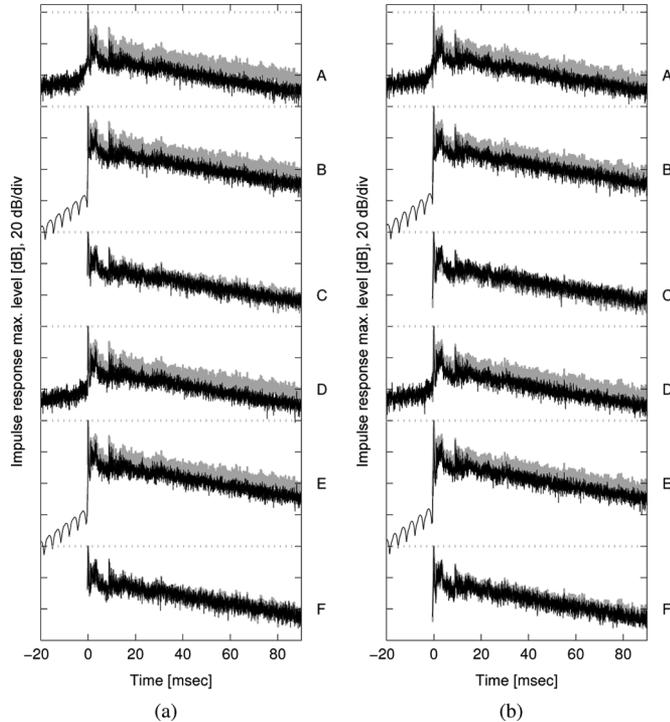


Fig. 6. Maximum level envelopes $L(k)$ of original (gray) and equalized (black) impulse responses, for filters A–F. The dotted gray lines indicate the 0 dB-level of each pair of original and equalized responses. The markings on the vertical axes indicate intervals of 20 dB relative to the 0-dB levels. (a) Performance in design points. (b) Performance in validation points.

A further discrimination between filters E and F is not possible based on Fig. 5, since they differ only by an all-pass factor.

Next, we turn to studying the time domain properties of the filters. The curves $L(k)$ in Fig. 6 obviously reveal some important properties not visible in Fig. 5. We summarize the details provided by Fig. 6 as follows.

- The pre-rings caused by filters A and D are unacceptably high, both in the design and validation points. (–40 dB at 20 ms before the maximum peak).
- The ratio between the maximum peak and the lower levels seems to be improved, both in the design and validation points, by all filters except filter C.
- Best overall performance is exhibited by filters B and E, which cause only a very low level of pre-ringing (about –60 dB immediately before the maximum peak, and rapidly decaying to –80 dB at 20 ms before the maximum peak), while substantially improving the ratio between the maximum peak and the lower levels in the responses.

So far, our graphical evaluation suggests filters E and F as the best candidates for a perceptually acceptable loudspeaker compensation, since they are the only filters without any immediately objectionable properties. However, provided that its low-level pre-rings can be tolerated, filter E seems to possess the most preferable time domain properties. This is confirmed by a study of the Schroeder decays and energy step responses in Figs. 7 and 8. We conclude this section by commenting on the behavior observed in these figures. It should be noted that the scales on the axes of the diagrams in Figs. 7 and 8 have been selected so as to display the most interesting parts of the

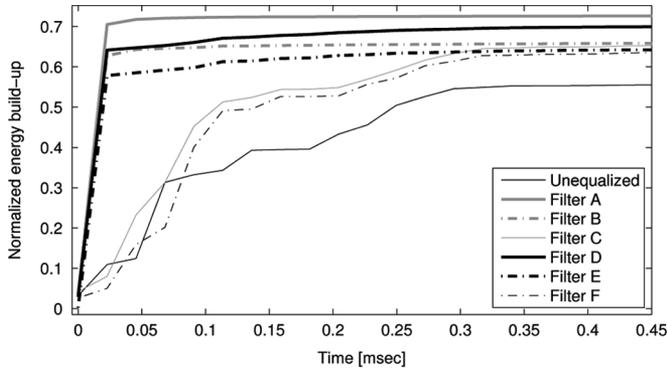
responses in a reasonable resolution. For example, in the full-band responses (0–22 050 Hz), the most interesting differences among the filters appear within the first 0.5 ms of the equalized responses. In the low-frequency band (0–320 Hz), this time frame is up to 100 ms long.

- Fig. 7(a) suggests an intuitively appealing ranking of the filters A–F: All of the filters A–E seem to improve the original system, with the fastest energy build-up being provided by filters A and D, closely followed by B and E, while C causes only a moderate improvement. Filter F degrades the step response in its earliest part. However, this ranking of the filters is not maintained in the other diagrams. Fig. 7(c) and (d) shows that it no longer holds at low frequencies, where the rise from 5% to 10% of the total energy takes about 17 ms in the design points and about 22 ms in the validation points for filters A and D. The pre-ringing error introduced by filters A and D thus contributes to a considerable part of the total low-frequency energy in the equalized responses, slowing down the early part of energy step response. In the validation points, the pre-ringing problem of filter A is evident also in the full bandwidth case. Moreover, in the validation points the unequalized response has, at times, better performance than the equalized responses of all filters except filter E. Particularly, in Fig. 7(d) at about 23 ms, the original response “catches up” on the step responses produced by filters B and C. Thus, by increasing the sound energy in the late parts of the impulse responses, the filters have caused artificial post-rings in the validation points.
- Fig. 8 provides essentially the same information as Fig. 7, although the post-ringing problems introduced by filters B and C at low frequencies are even more evident here.
- Based on Fig. 7 and 8, filter D can be ruled out due to severe pre-ringing at low frequencies. Filter F improves on the original response everywhere except in the first few samples of the fullband case. Filter E improves the original response everywhere. It is considerably better than filter F in the earliest parts (0.0–0.3 ms) of the fullband responses, and throughout the low-frequency responses.

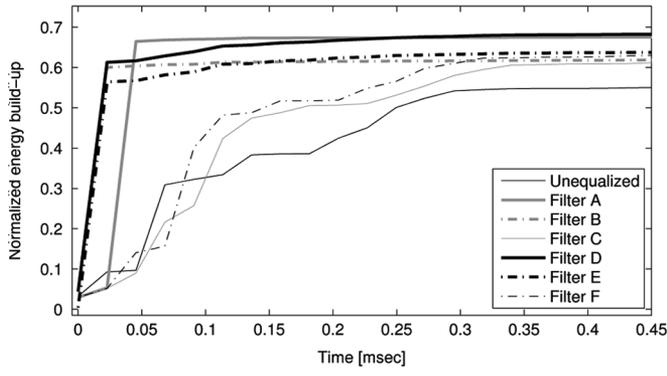
VII. CONCLUSION

A new method for robust mixed phase audio compensation has been presented. By the use of polynomial multivariable control techniques and a SIMO MSE criterion, analytical expressions for a spatially robust filter were obtained. It was shown that the optimum mixed phase MSE solution involves two kinds of spatial averages, here named the complex and RMS averages respectively, of which the latter is commonly used in minimum phase equalizer design. Due to perceptual shortcomings of the optimum mixed phase MSE filter, a constrained mixed phase design was proposed and experimentally shown to possess time domain qualities preferable to those of the MSE optimal mixed and minimum phase filters in the original unconstrained design. It is our opinion that this result motivates a revision of the widespread conclusion that excess phase properties of the RTFs must be neglected in a robust equalizer design.

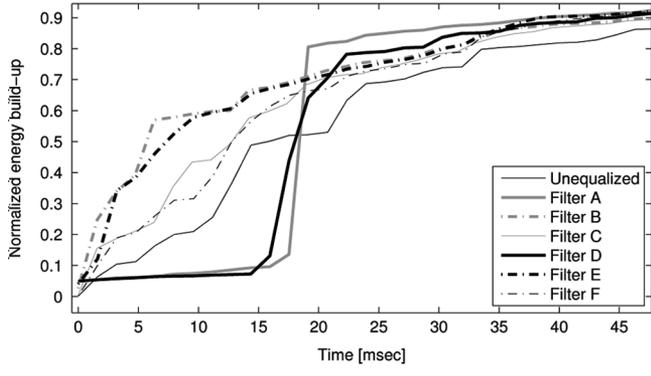
In order to keep the presentation transparent, RTFs were represented with FIR models $\mathcal{H}_i(q^{-1}) = B_i(q^{-1})$ in most of the analysis in Sections III–V and in the evaluations in Section IV.



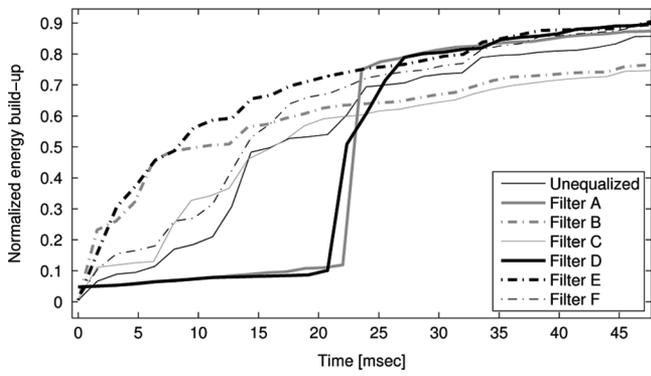
(a)



(b)

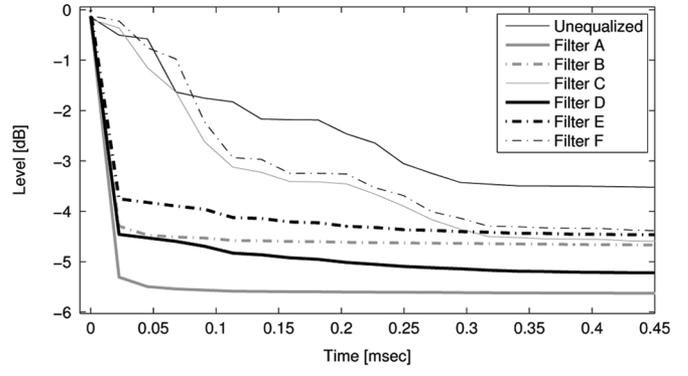


(c)

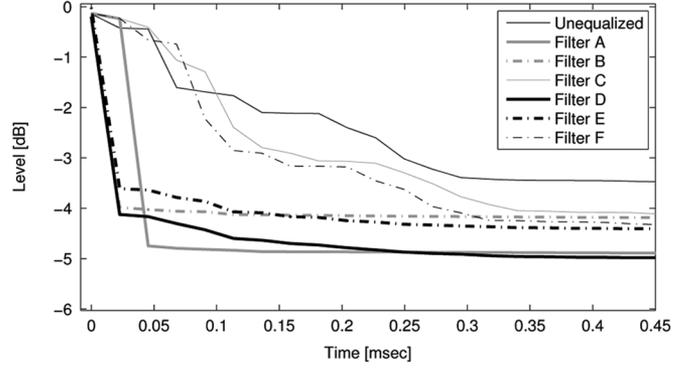


(d)

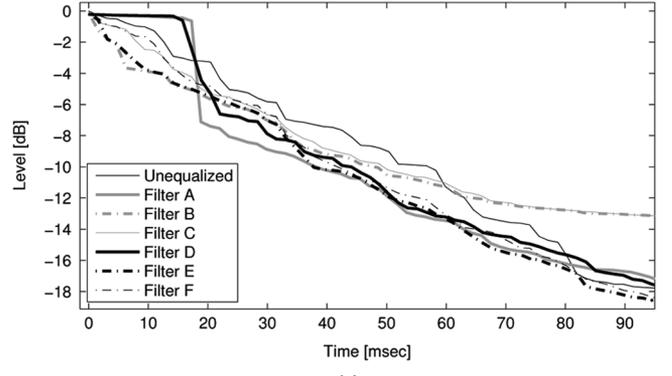
Fig. 7. Average energy step responses $S(k)$ of original and equalized responses for filters A–F. (a) Full bandwidth responses in design points. (b) Full bandwidth responses in validation points. (c) Responses below 320 Hz in design points. (d) Responses below 320 Hz in validation points.



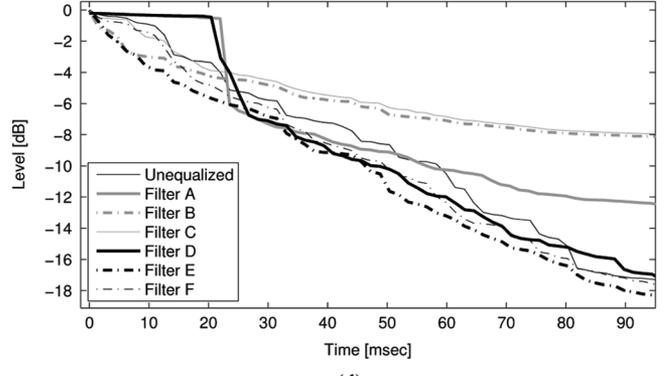
(a)



(b)



(c)



(d)

Fig. 8. Average Schroeder decay sequences $D(k)$ of original and equalized responses for filters A–F. (a) Full bandwidth responses in design points. (b) Full bandwidth responses in validation points. (c) Responses below 320 Hz in design points. (d) Responses below 320 Hz in validation points.

The results and interpretations regarding, e.g., spatial averages and clustering of near-common zeros can, however, be shown to be valid for the general IIR model $\mathcal{H}_i(q^{-1}) = B_i(q^{-1})/A_i(q^{-1})$. In particular, the results hold for the common acoustical pole and zero model (CAPZ) [18], where $\mathcal{H}_i(q^{-1}) = B_i(q^{-1})/A(q^{-1})$, for a common pole polynomial A . The inverse filters derived in the present work, however, differ significantly from that proposed in [18], which consists only of the common denominator, $\mathcal{R}(q^{-1}) = A(q^{-1})$.

We argued in Subsection V-H that an accurate estimate of the true power response average in the region, based only on a few measurements, is required for a practically useful filter design. We also saw in Section VI that our solution—a 1/6th octave smoothing of the RMS spectrum—was helpful, although probably far from optimal. A more flexible smoothing operation, taking more acoustical information into account, would probably further improve the filter performance.

Finally, we emphasize that the applicability of our proposed mixed phase method, and its superiority to a standard minimum phase design, heavily depends on the existence of a near-common excess phase part among the RTFs. In an arbitrary acoustic environment, there is of course nothing that guarantees the existence of such a common part. Our experience so far has, however, indicated that it may exist under quite general circumstances. It is an interesting topic for further research to reach a better understanding of the conditions for its existence, and to better quantify the properties of the noncommon and near-common factors among B_1, \dots, B_p . Recent results from the field of approximate greatest common divisors (AGCDs) of polynomials [19] may prove fruitful here.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers, whose efforts greatly helped to improve the quality of the paper.

REFERENCES

- [1] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 311–319, May 2000.
- [2] P. Hatziantoniou and J. Mourjopoulos, "Errors in real-time room acoustics dereverberation," *J. Audio Eng. Soc.*, vol. 52, no. 9, pp. 883–899, Sep. 2004.
- [3] P. Hatziantoniou and J. Mourjopoulos, "Real-time room equalization based on complex smoothing: Robustness results," in *Proc. 116th AES Convention*, Berlin, Germany, May 8–11, 2004.
- [4] R. Wilson, "Equalization of loudspeaker drive units considering both on-and off axis responses," *J. Audio Eng. Soc.*, vol. 39, no. 3, pp. 127–139, 1991.
- [5] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, 1989.
- [6] F. Talantzis and D. B. Ward, "Robustness of multichannel equalization in an acoustic reverberant environment," *J. Acoustic. Soc. Amer.*, vol. 114, no. 2, pp. 833–841, 2003.
- [7] F. Talantzis and L. Polymenakos, "Robustness of non-exact multi-channel equalization in reverberant environments," in *Artificial Intelligence and Innovations 2007: From Theory to Applications*, C. Boukis, A. Pnevmatikakis, and L. Polymenakos, Eds. Boston, MA: Springer, 2007, vol. 247 of Int. Fed. for Information Processing (IFIP), pp. 315–321.

- [8] M. Sternad and A. Ahlén, "LQ controller design and self-tuning control," in *Polynomial Methods in Optimal Control and Filtering*, K. Hunt, Ed. London, U.K.: Peter Peregrinus, 1993, pp. 56–92.
- [9] L.-J. Brännmark and A. Ahlén, "Robust loudspeaker equalization based on position-independent excess phase modeling," in *Proc. 2008 IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Las Vegas, NV, Mar. 30–Apr. 4, 2008, pp. 385–388.
- [10] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems: Theory and Design*. Upper Saddle River, NJ: Prentice-Hall, 1997.
- [11] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall.
- [12] A. H. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numer. Linear Algebra With Appl.*, vol. 8, no. 6–7, pp. 467–496, 2001.
- [13] S. Kung, "A new identification and model reduction algorithm via singular value decomposition," in *Proc. 12th Asilomar Conf. Circuits, Systems, Computing*, 1978, pp. 705–714.
- [14] P. Craven and M. Gerzon, "Practical adaptive room and loudspeaker equaliser for hi-fi use," in *Proc. AES DSP U.K. Conf.*, 1992, pp. 121–153, AES.
- [15] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoustic. Soc. Amer.*, vol. 37, no. 3, pp. 409–412, 1965.
- [16] E. A. Robinson, *Statistical Communication and Detection*. London, U.K.: Griffin, 1967.
- [17] L. Ljung, *System Identification—Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [18] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 325–333, Jul. 1997.
- [19] R. M. Corless, S. M. Watt, and L. Zhi, "QR factoring to compute the GCD of univariate approximate polynomials," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3394–3402, Dec. 2004.



Lars-Johan Brännmark (S'07) was born in Luleå, Sweden, in 1974. He received the Diploma degree in sound engineering from the Piteå School of Music, Luleå University of Technology, in 1996 and has studied electrical engineering at Purdue University, West Lafayette, IN, and musicology at Uppsala University, for one year each. He received the M.S. degree from Uppsala University, Uppsala, Sweden, in 2004. He is currently working towards the Ph.D. degree in signal processing at the Department of Engineering Sciences, Uppsala University.

He has a professional background as a Sound Engineer at the Swedish National Radio and as a Research Engineer at Dirac Research AB, Uppsala, Sweden. His research interests are in the field of signal processing for audio applications.



Anders Ahlén (S'80–M'84–SM'90) was born in Kalmar, Sweden. He received the Ph.D. degree in automatic control from Uppsala University, Uppsala, Sweden, in 1986.

He was with the Systems and Control Group, Uppsala University, from 1984 to 1989 as an Assistant Professor and from 1989 to 1992 as an Associate Professor in automatic control. During 1991 he was a visiting researcher at the Department of Electrical and Computer Engineering, The University of Newcastle, Australia. In 1992, he was appointed Associate Professor of Signal Processing at Uppsala University, where he is a Full Professor and holds the Chair in Signal Processing and is also the head of the Signals and Systems Group. From 2001 to 2004, he was the CEO of Dirac Research AB, Uppsala, Sweden, a company offering state-of-art audio signal processing solutions. He is currently the chairman of the board of the same company. His research interests, which include signal processing, communications and control, are currently focused on signal processing for wireless communications, wireless systems beyond 3G, wireless sensor networks, and audio signal processing.

From 1998 to 2004, he was the Editor of Signal and Modulation Design for the IEEE TRANSACTIONS ON COMMUNICATIONS.