Multichannel Dereverberation Theorems and Robustness Issues

Hüseyin Hacıhabiboğlu*, *Member, IEEE*, and Zoran Cvetković, *Senior Member, IEEE*

***Abstract*—Multichannel dereverberation amounts to the inversion of a multiple-input/multiple-output linear time-invariant system. In this paper necessary and sufficient conditions for perfect dereverberation using stable and FIR filters are established. It is then shown that the inverse system given by the pseudoinverse of the original transfer function matrix exhibits a noise reduction property. A necessary and sufficient condition under which this pseudoinverse system is FIR is also given. Further, an FIR approximation to the pseudoinverse system is considered and the effects of the length of this approximation on the dereverberation accuracy are investigated. Finally, an analytical and numerical assessment of the dependence of the dereverberation accuracy on the accuracy of the acquisition of room impulse responses is provided.**

***Index Terms*—Multichannel dereverberation, MIMO systems, room acoustics**

I. Introduction

FOR a single source and single microphone recording setup, a room acts as a single-input/single-output linear, time-invariant system. The audio signal recorded by the microphone is the signal emitted by the source convolved with the corresponding room impulse response (RIR). An RIR consists of the direct path, early reflections, and late reverberation components, and can be modeled as a long FIR filter. Room acoustics may have detrimental effects on the quality of audio recordings by smearing the original audio source signal in time and changing its spectral properties. The effects of room acoustics can be compensated for or eliminated by means of signal processing techniques [1], [2]. Based on the availability of the knowledge of the underlying acoustical system, these methods can be classified into two categories: blind and non-blind methods. Blind methods are generally based on *a priori* statistical hypotheses and not on the explicit knowledge on the acoustics of the room [3]. In contrast, non-blind methods are based on the use of empirical data from acoustical measurements or estimation. In this paper, we consider non-blind multichannel dereverberation.

What makes the dereverberation problem challenging even when underlying impulse responses are known is that the RIRs are very long, which makes their inversion computationally demanding. They are at the same time generally non-minimum phase [4], which means that they do not have stable exact inverse systems which are FIR or causal [5]. Early attempts, in the context of the mathematically equivalent problem of room equalization, concentrated on the equalization of recorded signals via correcting the spectral notches in magnitude responses of room transfer functions either by spatial averaging of the outputs of multiple microphones [6] or by correcting the magnitude response [7]. Such methods reduce spectral coloration but are not effective in reducing reverberation in the time domain. It was shown that stable but non-causal FIR filters can be designed to model the inverse responses for equalization and dereverberation [8], [9]. The problem becomes much more tractable if multiple recordings of a single source signal are available. In such a situation, perfect dereverberation using finite impulse response (FIR) filters is possible as long as impulse responses at microphone positions have no zeros in common [10].

When multiple sources are recorded using multiple microphones, the signal recorded by each microphone is a convolutive mixture of all source signals. Early work on multichannel dereverberation used phase aligned addition at different channels to increase the level of the original signal and to reduce reverberation [11]. Another non-blind multichannel dereverberation method used frequency-domain inversion with regularization to obtain FIR inverse filters [12]. A recent work approaches the problem by optimizing inverse filters to minimize a combination of the dereverberation error and perturbation caused by additive noise with particular known statistics [13].

In this paper a necessary and sufficient condition for perfect dereverberation using stable filters is established. When the number of microphones is larger than the number of source signals, the stable inverse system is not unique and may not be causal. It is therefore of interest to know when perfect dereverberation using FIR filters is possible, and a necessary and sufficient condition for FIR dereverberation is established. While all inverse systems perform exact dereverberation under ideal conditions, they have a different effect on additive noise. It is then shown that the inverse system given by the pseudoinverse of the room transfer function matrix has a desirable noise reduction property, and a necessary and sufficient condition for

Manuscript received September 16, 2010, revised May 25, 2011. This work was supported by Engineering and Physical Sciences Research Council (EPSRC) Research Grant, "Perceptual Sound Field Reconstruction and Coherent Emulation" (EP/F001142/1). The work reported in this article was carried out while H. Hacıhabiboğlu was with King's College London. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick Naylor.

H. Hacıhabiboğlu is with Department of Modeling and Simulation, Informatics Institute, Middle East Technical University, Ankara, Turkey. (e-mail: huseyin@ii.metu.edu.tr).

Z. Cvetkovic is with the Centre for Telecommunications Research (CTR), King's College London, London, WC2R 2LS, United Kingdom. (e-mail: zoran.cvetkovic@kcl.ac.uk)

Digital Object Identifier 10.1109/TASL.2011.XXXXXX

the pseudoinverse to be FIR is also derived. The filters of the pseudo inverse system are in general IIR and specified by a very large number of numerator and denominator coefficients, hence it is of interest to find FIR approximations and compute these approximations in a numerically efficient manner. A DFT-based method for finding an FIR approximation to the pseudoinverse proposed by Kirkeby *et al.* [12] is then reviewed and a numerical assessment of the effects of the size of the DFT employed and the length of the approximate filters on the accuracy of the dereverberation is provided.

Non-blind methods assume the knowledge of room impulse responses, however these are acquired only within a certain limited accuracy. Furthermore rooms are only weakly stationary systems [14], and variations of RIRs, due to temperature changes for instance, cause problems in echo cancellation [15] and dereverberation [16]. Complex smoothing [16], [17], [18] of RIRs was proposed to alleviate these problems. It was also shown that even small changes of positions of sources and microphones can cause significant performance problems in room response inversion [19] and dereverberation [20], [21]. Spatial average equalization was proposed as a possible solution to this problem [22]. In this article we present a theoretical assessment of the effects of RIR perturbations on the dereverberation accuracy for the inversion using the pseudoinverse system. Then we focus on perturbations which can be modelled as delay modulation and additive noise and present some experimental results which suggest that the dereverberation using the pseudoinverse system is not very sensitive to RIR perturbations of this kind.

Notation: A polynomial matrix $\mathbf{H}(z)$ is said to be unimodular if its determinant is a constant. $\tilde{\mathbf{H}}(z)$ denotes the matrix obtained by transposing $\mathbf{H}(z)$, conjugating all the coefficients and replacing z by z^{-1}.

II. Multichannel Dereverberation Theorems

Let us consider M microphones recording signals emitted by L sources, where $L \leq M$. The signal $Y_m(z)$ captured by the m^{th} microphone is a convolutive mixture of source signals $X_l(z)$,

$$Y_m(z) = \sum_{l=1}^{L} H_{ml}(z) X_l(z) \tag{1}$$

where $H_{ml}(z)$ is the room transfer function between the m^{th} microphone and the l^{th} source. The problem addressed in this article is the inversion of this MIMO system, that is, the dereverberation of the recorded signals $Y_1(z), \ldots, Y_M(z)$ to obtain source signals $X_1(z), \ldots, X_L(z)$. We will assume that room impulse responses are FIR, albeit very long FIR filters. It should be noted that there exist methods to model multiple RIRs as IIR filters with a set of common poles [23], however, in this paper we pursue modelling using long FIR filters.

The convolutive mixture in (1) can be represented in matrix form as $\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$, where

$$[\mathbf{H}(z)]_{ij} = H_{ij}(z),\ i = 1, \ldots, M,\ j = 1, \ldots, L\ ,$$

$$\mathbf{X}(z) = [X_1(z) \ldots X_L(z)]^T \text{ and } \mathbf{Y}(z) = [Y_1(z) \ldots Y_M(z)]^T\ .$$

The deconvolution requires finding an equalization filter matrix $\mathbf{G}(z)$,

$$[\mathbf{G}(z)]_{ij} = G_{ij}(z),\ i = 1, \ldots, L,\ j = 1, \ldots, M\ ,$$

such that the combined transfer function of the cascade of the system matrix and the equalization filter matrix $\mathbf{G}(z)$ is an identity,

$$\mathbf{G}(z)\mathbf{H}(z) = \mathbf{I}\ .$$

Note that impulse responses corresponding to transfer functions $H_{ml}(z)$, due to non-perfectly reflective room boundaries, exhibit exponential decay and are therefore square-summable. We restrict considerations to finite energy source signals.

A necessary and sufficient condition for the existence of a stable dereverberation system is given in the following theorem.

Theorem 1: Perfect deconvolution using stable filters is possible if and only if $\mathbf{H}(z)$ is of full rank everywhere on the unit circle.

Proof: Assume that $\mathbf{H}(e^{j\omega})$ is singular at a frequency $\omega = \omega_0$. Then there exists a unit-norm vector $\mathbf{c} = [c_1 \ \ldots \ c_L]^T$, $\|\mathbf{c}\|^2 = \sum_{l=1}^{L} |c_l|^2 = 1$, such that $\mathbf{H}(e^{j\omega_0})\mathbf{c} = \mathbf{0}$. Denote by $H_1^c(e^{j\omega}), \ldots, H_M^c(e^{j\omega})$ filters $[H_1^c(e^{j\omega}) \ldots H_M^c(e^{j\omega})]^T = \mathbf{H}(e^{j\omega})\mathbf{c}$, and let $a_k(n)$, $k = 1, 2, \ldots$ be a sequence of signals given in the Fourier domain by $A_k(e^{j\omega}) = \sqrt{k\pi}$, $\omega \in \Omega_k$, where $\Omega_k = \{\omega : \omega_0 - \frac{1}{2k} \leq |\omega| \leq \omega_0 + \frac{1}{2k}\}$, and $A_k(e^{j\omega}) = 0$, $\omega \notin \Omega_k$. These signals satisfy $\frac{1}{2\pi}\int_{-\pi}^{\pi} |A_k(e^{j\omega})|^2 d\omega = 1$ for all k. Consider the sequence of vector signals $\mathbf{X}_k(z) = [X_{k,1}(z) \ldots X_{k,L}(z)]^T = \mathbf{c}A_k(z)$. If $\mathbf{X}_k(z)$ is the input to $\mathbf{H}(z)$ then the corresponding output $\mathbf{Y}_k(z) = [Y_{k,1}(z) \ldots Y_{k,M}(z)]^T$ is given by $[Y_{k,1}(e^{j\omega}) \ldots Y_{k,M}(e^{j\omega})]^T = [H_1^c(e^{j\omega}), \ldots, H_M^c(e^{j\omega})]^T A_k(e^{j\omega})$ and satisfies $|Y_{k,m}(e^{j\omega})| = 0$, $\omega \notin \Omega_k$, for all m, $1 \leq m \leq M$, and all k. Note further that filters $H_m^c(z)$ have finite impulse responses, and that therefore all frequency responses $H_m^c(e^{j\omega})$ are continuous in ω. This continuity implies that since $H_m^c(e^{j\omega_0}) = 0$, for all $m = 1, \ldots, M$, and we consider finitely many filters $H_m^c(e^{j\omega})$, for any $\delta > 0$, there exists an ϵ_δ such that $|H_m^c(e^{j\omega})| < \delta$ for $\omega_0 - \epsilon_\delta \leq |\omega| \leq \omega_0 + \epsilon_\delta$ for all $m = 1, \ldots, M$. Hence, given a δ, for any $k > 1/(2\epsilon_\delta)$,

$$\sum_{m=1}^{M} \frac{1}{2\pi}\int_{-\pi}^{\pi} |Y_{k,m}(e^{j\omega})|^2 d\omega = \sum_{m=1}^{M} \frac{1}{2\pi}\int_{\omega\in\Omega_k} |Y_{k,m}(e^{j\omega})|^2 d\omega \leq \sum_{m=1}^{M} \frac{1}{2\pi}\int_{\omega\in\Omega_k} k\pi\delta^2 d\omega = M\delta^2\ .$$

This proves that $\lim_{k\to\infty} \frac{1}{2\pi}\sum_{m=1}^{M}\int_{-\pi}^{\pi} |Y_{k,m}(e^{j\omega})|^2 d\omega = 0$. At the same time the corresponding sequence of input signals satisfies $\frac{1}{2\pi}\sum_{l=1}^{L}\int_{-\pi}^{\pi} |X_{k,l}(e^{j\omega})|^2 d\omega = \|\mathbf{c}\|^2 = 1$ for all k. Assume that there exists a system $\mathbf{G}(z)$ which performs perfect reconstruction of signals $\mathbf{X}(z)$ from their convolutive mixtures $\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$ as $\mathbf{X}(z) = \mathbf{G}(z)\mathbf{Y}(z)$. Then, in order to have the combined energy of signals $\hat{\mathbf{X}}_k(z) =$

$\mathbf{G}(z)\mathbf{Y}_k(z)$ equal to the combined energy of signals $\mathbf{X}_k$, which is a necessary condition for $\hat{\mathbf{X}}_k(z) = \mathbf{X}_k(z)$, while the energy of signals $\mathbf{Y}_k(z)$ tends to zero, at least one of $G_{lm}(e^{j\omega})$ must be unbounded at ω_0, *i.e.* at least one of the filters of $\mathbf{G}(z)$ must have a pole on the unit circle. This proves that the condition of the theorem is necessary for stable dereverberation. To prove the sufficiency, assume that $\mathbf{H}(e^{j\omega})$ is of full rank for every ω. Then, $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$ is also of full rank for all z on the unit circle. The pseudoinverse $\mathbf{H}^{\dagger}(z) = \left[\tilde{\mathbf{H}}(z)\mathbf{H}(z)\right]^{-1}\tilde{\mathbf{H}}(z)$ of $\mathbf{H}(z)$, therefore, has no poles on the unit circle, hence it is a stable inverse of $\mathbf{H}(z)$. □

Stable inversion stated by Theorem 1 is not necessarily FIR or causal. If the stable inverse is IIR or not causal, it needs to be approximated by an FIR system to perform stable approximate dereverberation with some delay. It is therefore of interest to know when an exact FIR inverse exists. A condition for FIR dereverberation will be established here based on the Smith form [24] of $\mathbf{H}(z)$. An $M \times L$ ($M \geq L$) polynomial matrix $\mathbf{H}(z)$ can be decomposed as

$$\mathbf{H}(z) = \mathbf{U}_H(z)[\mathbf{D}_H(z)\ \mathbf{0}]^T\mathbf{V}_H(z),$$

where $\mathbf{U}_H(z)$ and $\mathbf{V}_H(z)$ are $M \times M$ and $L \times L$ unimodular matrices, respectively, $\mathbf{D}_H(z)$ is a diagonal polynomial matrix, $\mathbf{D}_H(z) = \mathrm{diag}(d_1(z), d_2(z), \ldots, d_L(z))$, and $\mathbf{0}$ is the $L \times (M-L)$ zero matrix. $\mathbf{U}_H(z)$ and $\mathbf{V}_H(z)$ are products of matrices which correspond to elementary row and column operations, respectively, and have one of the following three forms: i) a permutation matrix; ii) a diagonal matrix with elements on the diagonal equal to 1, except for one which is a different nonzero constant; iii) a matrix with 1s on the main diagonal and one polynomial entry off the diagonal. $\mathbf{U}_H(z)$ and $\mathbf{V}_H(z)$ can further be selected such that polynomials $d_i(z)$ are monic and $d_i(z)$ is a factor of $d_{i+1}(z)$. Such a matrix $[\mathbf{D}_H(z)\ \mathbf{0}]^T$ is referred to as the Smith normal form of $\mathbf{H}(z)$. A sufficient condition for FIR dereverberation in the case when $L = 1$, and in the case of arbitrary L with $M = L+1$ has been given by Miyoshi and Kaneda [10]. The next theorem establishes a necessary and sufficient condition for FIR dereverberation for arbitrary L and M, $M \geq L$.

Theorem 2: Perfect dereverberation using FIR filters is possible if and only if the polynomials in the Smith form of $\mathbf{H}(z)$ are monomials.

Proof: Let $\mathbf{G}(z)$ be an FIR inverse of $\mathbf{H}(z)$, and let $\mathbf{H}(z) = \mathbf{U}_H(z)[\mathbf{D}_H(z)\ \mathbf{0}]^T\mathbf{V}_H(z)$ and $\mathbf{G}(z) = \mathbf{U}_G(z)[\mathbf{D}_G(z)\ \mathbf{0}]\mathbf{V}_G(z)$ be the Smith form decompositions of $\mathbf{H}(z)$ and $\mathbf{G}(z)$. Since $\mathbf{G}(z)\mathbf{H}(z) = \mathbf{I}$, it follows that

$$\begin{aligned}\det\mathbf{G}(z)\det\mathbf{H}(z) &= \det\mathbf{U}_G(z)\\ &\cdot\det([\mathbf{D}_G(z)\ \mathbf{0}]\mathbf{V}_G(z)\mathbf{U}_H(z)[\mathbf{D}_H(z)\ \mathbf{0}]^T)\\ &\cdot\det\mathbf{V}_H(z) = 1\ .\end{aligned}$$

This equality and the fact that $\mathbf{U}_G(z)$ and $\mathbf{V}_H(z)$ are unimodular further imply that

$$\det([\mathbf{D}_G(z)\ \mathbf{0}]\mathbf{V}_G(z)\mathbf{U}_H(z)[\mathbf{D}_H(z)\ \mathbf{0}]^T) = const.$$

On the other hand

$$[\mathbf{D}_G(z)\ \mathbf{0}]\mathbf{V}_G(z)\mathbf{U}_H(z)[\mathbf{D}_H(z)\ \mathbf{0}]^T = \mathbf{D}_G(z)\mathbf{C}(z)\mathbf{D}_H(z)\ ,$$

where $\mathbf{C}(z)$ is the $L \times L$ upper left corner submatrix of $\mathbf{V}_G(z)\mathbf{U}_H(z)$. Hence

$$\det\mathbf{D}_G(z)\det\mathbf{C}(z)\det\mathbf{D}_H(z) = const.$$

Since $\det\mathbf{D}_G(z)$ is a polynomial, by the definition of the Smith form, $\det\mathbf{C}(z)$ is a polynomial because $\mathbf{C}(z)$ is a submatrix of a polynomial matrix, $\det\mathbf{D}_G(z)\det\mathbf{C}(z)\det\mathbf{D}_H(z) = const.$ is possible only if $\det\mathbf{D}_H(z)$, $\det\mathbf{C}(z)$ and $\det\mathbf{D}_G(z)$ are all monomials. This holds because the product of three polynomials (in z and z^{-1}) can be a constant only if all three polynomials are monomials. Finally, this implies that all polynomials in $\det\mathbf{D}_H(z)$ are also monomials. To prove that the condition is sufficient, assume that all polynomials $d_i(z)$ in the Smith form of $\mathbf{H}(z)$ are monomials. Then

$$\mathbf{G}(z) = \mathbf{V}_H^{-1}(z)\left[\mathrm{diag}\left(\frac{1}{d_1(z)}, \ldots, \frac{1}{d_L(z)}\right)\ \mathbf{0}\right]\mathbf{U}_H^{-1}(z)$$

is an inverse of $\mathbf{H}(z)$ and is an FIR matrix since $\mathbf{U}_H(z)$ and $\mathbf{V}_H(z)$ are unimodular and $d_i(z)$ are monomial. □

When a stable inverse system exists and $M > L$, that inverse is not unique. While all inverse systems perform perfect deconvolution in noise-free conditions, they affect additive noise differently. To gain intuition about the impact of the choice of the inverse to noise robustness, consider a multiple-input/multiple-output linear time-invariant system described by a transfer matrix $\mathbf{H}(z)$. This system is an operator which maps a vector $\mathbf{X}(z)$ of L finite energy signals to a vector $\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$ of M finite energy signals. If the operator is invertible, than this mapping is one-to-one from $\mathcal{X}$, the space of vectors of L finite energy signals, to $\mathcal{R}\{\mathbf{H}(z)\}$, the range of $\mathbf{H}(z)$. When $M > L$, $\mathcal{R}\{\mathbf{H}(z)\}$ is a proper subspace of $\mathcal{Y}$, the space of vectors of M finite energy signals. Consider now reconstructing a vector of input signals $\mathbf{X}(z)$ from the corresponding output vector $\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$ degraded by some additive error/noise $\mathbf{E}_Y(z)$ by applying an inverse $\mathbf{G}(z)$ of $\mathbf{H}(z)$. This reconstruction of $\mathbf{X}(z)$ from noisy $\mathbf{Y}(z)$ gives $\mathbf{X}(z) + \mathbf{E}_X(z) = \mathbf{G}(z)\mathbf{Y}(z) + \mathbf{G}(z)\mathbf{E}_Y(z)$ where $\mathbf{E}_X(z) = \mathbf{G}(z)\mathbf{E}_Y(z)$. Note that $\mathbf{Y}(z) + \mathbf{E}_Y(z)$ is not necessarily in the range of $\mathbf{H}(z)$, and if that is the case, the image of the reconstructed signal $\mathbf{X}(z) + \mathbf{E}_X(z)$ under $\mathbf{H}(z)$ is $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$, where $\hat{\mathbf{E}}_Y(z) = \mathbf{H}(z)\mathbf{E}_X(z)$, which is different from $\mathbf{E}_Y(z)$. Hence, $\mathbf{G}(z)$ implicitly first projects $\mathbf{Y}(z) + \mathbf{E}_Y(z)$ onto the range of $\mathbf{H}(z)$, which gives $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$, and then finds the vector in $\mathcal{X}$ whose image under $\mathbf{H}(z)$ is $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$. Hence, the reconstruction effectively maps $\mathbf{E}_Y(z)$ to $\hat{\mathbf{E}}_Y(z)$, which may reduce it, or amplify it, or leave it unchanged (the latter happens when $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$ is in the range of $\mathbf{H}(z)$), and that is the way in which the choice of the inverse affects the error. The mapping from $\hat{\mathbf{E}}_Y(z)$ to $\mathbf{E}_X(z)$ is completely determined by the original operator $\mathbf{H}(z)$. The effects of the choice of inverse system on $\hat{\mathbf{E}}_Y(z)$ and $\mathbf{E}_X(z)$ are illustrated by the following simple example.

Example 1: Let us consider the case of a single input, $X(z)$, and M outputs, $\mathbf{Y}(z) = [Y_1(z)\ \ldots\ Y_M(z)]$, and assume $\mathbf{H}(z) = [1\ \ldots\ 1]^T$, such that $Y_m(z) = X(z)$, $m = 1, \ldots, M$. This is an extremely well conditioned system and

any system of filters $\mathbf{G}(z) = [G_1(z) \ \ldots \ G_M(z)]$ *such that* $\sum_{m=1}^{M} G_m(z) = 1$ *will be an inverse of* $\mathbf{H}(z)$*. One possible inverse is given by* $\mathbf{G}_o(z) = [\frac{1}{M} \ \frac{1}{M} \ \ldots \ \frac{1}{M}]$ *and another inverse by* $\mathbf{G}_s(z) = [M \ -1 \ \ldots \ -1]$.

Consider reconstructing $X(z)$ *from* $\mathbf{Y}(z) + \mathbf{E}_Y(z)$*, where* $\mathbf{E}_Y(z) = [E_{Y,1}(z) \ \ldots \ E_{Y,M}(z)]$ *is a vector of noise signals. If* $\mathbf{G}_o(z)$ *is used for the dereverberation, then the reconstructed signal is* $X(z) + E_X(z)$*, where* $E_X(z) = \left(\sum_{m=1}^{M} E_{Y,m}(z)\right)/M$*. Consequently*

$$\int_{-\pi}^{\pi} |E_X(e^{j\omega})|^2 d\omega \leq \frac{1}{M} \sum_{m=1}^{M} \int_{-\pi}^{\pi} |E_{Y,m}(e^{j\omega})|^2 d\omega \ , \quad (2)$$

i.e. the energy of the error in the reconstructed signal is smaller than the energy of the error in the recorded signals. $X(z)$ *is effectively reconstructed from signals* $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$*, where* $\hat{\mathbf{E}}_Y(z) = \mathbf{H}(z)E_X(z)$*, i.e.* $\mathbf{G}_o(z)(\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)) = X(z) + E_X(z)$*. This effective error is in this case* $\hat{E}_{Y,1}(z) = \hat{E}_{Y,2}(z) = \ldots = \hat{E}_{Y,M}(z) = \left(\sum_{m=1}^{M} E_{Y,m}(z)\right)/M$*, and it satisfies:*

$$\sum_{m=1}^{M} \int_{-\pi}^{\pi} |\hat{E}_{Y,m}(e^{j\omega})|^2 d\omega \leq \sum_{m=1}^{M} \int_{-\pi}^{\pi} |E_{Y,m}(e^{j\omega})|^2 d\omega \ , \quad (3)$$

i.e. its energy is less than or equal to the original error energy. These last two inequalities follow immediately from elementary vector norm inequalities and hold for every error of bounded energy.

Consider now reconstructing $X(z)$ *using* $\mathbf{G}_s(z)$ *and assume that noise is present only in* $Y_1(z)$*, i.e.* $E_{Y,2}(z) = \ldots = E_{Y,M}(z) = 0$*. The reconstructed signal is then given by* $X(z) + E_X(z)$*, where* $E_X(z) = ME_{Y,1}(z)$*. For this inverse and this noise we then have*

$$\begin{aligned} \int_{-\pi}^{\pi} |E_X(e^{j\omega})|^2 d\omega &= M^2 \int_{-\pi}^{\pi} |E_{Y,1}(e^{j\omega})|^2 d\omega \\ &= M^2 \sum_{m=1}^{M} \int_{-\pi}^{\pi} |E_{Y,m}(e^{j\omega})|^2 d\omega \ , \end{aligned}$$

i.e. the energy of the reconstruction error is significantly higher than the energy of the error in the recorded signals. In this case, the signal is effectively reconstructed from $\mathbf{Y}(z) + \hat{\mathbf{E}}_Y(z)$*, where again* $\hat{\mathbf{E}}_Y(z) = \mathbf{H}(z)E_X(z)$*, which develops as* $\hat{E}_{Y,1}(z) = \hat{E}_{Y,2}(z) = \ldots = \hat{E}_{Y,M}(z) = ME_{Y,1}(z)$*. This effective error satisfies:*

$$\sum_{m=1}^{M} \int_{-\pi}^{\pi} |\hat{E}_{Y,m}(e^{j\omega})|^2 d\omega = M^3 \sum_{m=1}^{M} \int_{-\pi}^{\pi} |E_{Y,m}(e^{j\omega})|^2 d\omega \ , \quad (4)$$

i.e. its energy is significantly higher than the original error energy.

Ideally we would like to use an inverse of $\mathbf{H}(z)$ which has the property that the energy of the effective noise $\hat{\mathbf{E}}_Y(z)$ is always smaller or equal than the energy of the original noise $\mathbf{E}_Y(z)$. An inverse which has this property is provided by the left pseudo-inverse of $\mathbf{H}(z)$,

$$\mathbf{H}^{\dagger}(z) = \left[\tilde{\mathbf{H}}(z)\mathbf{H}(z)\right]^{-1} \tilde{\mathbf{H}}(z) \ ,$$

where $\tilde{\mathbf{H}}(z)$ denotes the matrix obtained by transposing $\mathbf{H}(z)$, conjugating all the coefficients and replacing z by z^{-1}. This result is established by the following theorem.

Theorem 3: Let $\mathbf{X}(z)$ *be a vector of bounded energy signals at the input of a linear time-invariant system* $\mathbf{H}(z)$*, and let* $\mathbf{Y}(z)$ *be the vector of corresponding output signals,* $\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$*. Assume that* $\mathbf{H}(z)$ *allows for inversion using a system of stable filters and let* $\mathbf{X}(z) + \mathbf{E}_X(z)$ *be obtained by applying* $\mathbf{H}^{\dagger}(z)$ *to* $\mathbf{Y}(z)$ *degraded by some additive error* $\mathbf{E}_Y(z)$*,* $\mathbf{X}(z) + \mathbf{E}_X(z) = \mathbf{H}^{\dagger}(z)\mathbf{Y}(z) + \mathbf{H}^{\dagger}(z)\mathbf{E}_Y(z)$*. Then, the effective error signal* $\hat{\mathbf{E}}_Y(z) = \mathbf{H}(z)\mathbf{E}_X(z)$ *satisfies*

$$\sum_{m=1}^{M} \int_{-\pi}^{\pi} |\hat{E}_{Y,m}(e^{j\omega})|^2 d\omega \leq \sum_{m=1}^{M} \int_{-\pi}^{\pi} |E_{Y,m}(e^{j\omega})|^2 d\omega \quad (5)$$

Proof: When $\mathbf{H}(z)$ has a stable inverse as specified in Theorem 1, then $\mathbf{H}(z)$ is a *frame operator* [25], [26] that maps the space of finite energy vector signals $\mathbf{X}(z) = [X_1(z) \ \ldots \ X_L(z)]^T$ onto a subspace $\mathcal{R}\{\mathbf{H}(z)\}$ of the space of finite energy vector signals $\mathbf{Y}(z) = [Y_1(z) \ \ldots \ Y_M(z)]^T$. The pseudoinverse system $\mathbf{H}^{\dagger}(z)$ is then the operator which is the minimal dual of the frame operator $\mathbf{H}(z)$ [25], [26], and therefore when applied to an arbitrary finite energy vector signal $\mathbf{Y}(z) + \mathbf{E}_Y(z)$, $\mathbf{H}^{\dagger}(z)$ implicitly performs orthogonal projection of $\mathbf{Y}(z) + \mathbf{E}_Y(z)$ onto $\mathcal{R}\{\mathbf{H}(z)\}$ [25]. Hence $\hat{\mathbf{E}}_Y(z)$ is the orthogonal projection of $\mathbf{E}_Y(z)$ onto $\mathcal{R}\{\mathbf{H}(z)\}$, and therefore its norm, and consequently energy, is smaller or equal than the energy of $\mathbf{E}_Y(z)$. □

This result does not mean that the pseudoinverse system performs maximal reduction of every noise. However, if an inverse system is optimized to maximally reduce noise of a given type, it may amplify the effect of other types of perturbation. The psudoinverse system is guaranteed to reduce, or at least not amplify, any type of finite energy error. An alternative approach to achieving noise robustness would be to optimize a balance between dereverberation error and error caused by noise for a given filter length and noise statistics. That approach has been recently investigated in the context of speech dereverberation in the case of one source signal [13].

An arbitrary FIR inverse of $\mathbf{H}(z)$, provided it exists, is not necessarily its pseudoinverse, and the pseudoinverse may not be FIR. A condition under which the pseudoinverse is FIR is established in the following theorem.

Theorem 4: The pseudoinverse of $\mathbf{H}(z)$ *is an FIR system if and only if* $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$ *is a unimodular matrix, that is, if and only if* $\det(\tilde{\mathbf{H}}(z)\mathbf{H}(z)) = const.$

Proof: Assume that $\mathbf{H}^{\dagger}(z)$ is a polynomial matrix. Let $\mathbf{H}(z) = \mathbf{U}(z)[\mathbf{D}(z) \ \mathbf{0}]^T\mathbf{V}(z)$ be the Smith form decomposition of $\mathbf{H}(z)$, where $\mathbf{D}(z) = \mathrm{diag}(d_1(z), \ldots, d_L(z))$. Then $\mathbf{H}^{\dagger}(z) = \mathbf{V}^{-1}(z)[\mathbf{D}^{-1}(z)\mathbf{A}^{-1}(z) \ \mathbf{0}]\tilde{\mathbf{U}}(z)$, where $\mathbf{A}(z)$ is the upper-left corner $L \times L$ submatrix of $\tilde{\mathbf{U}}(z)\mathbf{U}(z)$. Hence, $\mathbf{V}(z)\mathbf{H}^{\dagger}(z)\tilde{\mathbf{U}}^{-1}(z) = [\mathbf{D}^{-1}(z)\mathbf{A}^{-1}(z) \ \mathbf{0}]$. Notice that since $\mathbf{U}(z)$ is unimodular, the left hand side of the last expression is a polynomial matrix, and therefore, $\mathbf{D}^{-1}(z)\mathbf{A}^{-1}(z)$ is also a polynomial matrix. Since both $\mathbf{D}^{-1}(z)\mathbf{A}^{-1}(z)$ and $\mathbf{A}(z)\mathbf{D}(z)$ are polynomial matrices and $\det(\mathbf{A}(z)\mathbf{D}(z)) \det\left(\mathbf{D}^{-1}(z)\mathbf{A}^{-1}(z)\right) = 1$, $\det(\mathbf{A}(z)\mathbf{D}(z))$

must be a monomial, and that is only possible if all polynomials on the diagonal of $\mathbf{D}(z)$ are monomials. Further, using some elementary matrix manipulations it can be shown that

$$\mathbf{H}^{\dagger}(z)\tilde{\mathbf{U}}^{-1}(z)[(\tilde{\mathbf{D}}^{-1}(z)\tilde{\mathbf{V}}^{-1}(z))^{T}\ \mathbf{0}]^{T} = (\tilde{\mathbf{H}}(z)\mathbf{H}(z))^{-1}\ .$$

Since polynomials on the diagonal of $\mathbf{D}(z)$ are monomials, the left hand side of this equation is a polynomial matrix, and therefore, $(\tilde{\mathbf{H}}(z)\mathbf{H}(z))^{-1}$ must also be a polynomial matrix. Because both $(\tilde{\mathbf{H}}(z)\mathbf{H}(z))^{-1}$ and $(\tilde{\mathbf{H}}(z)\mathbf{H}(z))$ are polynomial matrices the determinant of $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$ must be a monomial, $\det(\tilde{\mathbf{H}}(z)\mathbf{H}(z)) = cz^{k}$. However, if $\det(\tilde{\mathbf{H}}(z)\mathbf{H}(z)) = P(z)$ where $P(z)$ is a polynomial in z, then $P(z)$ must satisfy $\tilde{P}(z) = P(z)$. The only monomial which satisfies this is $P(z) = const$. This proves that if $\mathbf{H}^{\dagger}(z)$ is FIR, then $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$ must be unimodular. To prove the sufficiency of the condition note that $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$ is a polynomial matrix. If furthermore $\det(\tilde{\mathbf{H}}(z)\mathbf{H}(z)) = const$, then $(\tilde{\mathbf{H}}(z)\mathbf{H}(z))^{-1}$ must be a polynomial matrix since its entries are products of polynomials. Hence, $\mathbf{H}^{\dagger}(z) = \left[\tilde{\mathbf{H}}(z)\mathbf{H}(z)\right]^{-1}\tilde{\mathbf{H}}(z)$ is a product of two polynomial matrices, and therefore it must be a polynomial matrix. □

In the case of one source signal, $L = 1$, perfect reconstruction using FIR filters is possible if and only if filters $H_{m1}(z),\ m = 1, \ldots, M$ have no zeros in common, and then the solution for inverse FIR filters of order smaller than the order of filters $H_{m1}(z)$ is unique. This result is known in the context of deconvolution of audio signals as multiple-input/output inverse theorem (MINT) [10]. The pseudoinverse is given by [27]

$$G_{m1}(z) = \frac{H_{m1}(z^{-1})}{\sum_{m=1}^{M} H_{m1}(z)H_{m1}(z^{-1})}\ ,$$

and these filters are FIR if and only if room impulse responses are power complementary

$$\sum_{m=1}^{M} H_{m1}(z)H_{m1}(z^{-1}) = const.$$

These theorems indicate that the deconvolution problem is nontrivial, that solutions, when they exist, come in different flavors, and point in the direction of the pseudoinverse of $\mathbf{H}(z)$ or some FIR approximation thereof as a numerically efficient and robust inverse solution.

III. Numerical Issues

The previous section provided an in-depth analysis of theoretical aspects of multichannel dereverberation. In this section we focus on practical aspects. In particular, numerically efficient inversion of the room transfer function is considered and the sensitivity of dereverberation to RIR acquisition errors is investigated.

A. Fast inversion algorithm

Considerations in the previous section point out that not every stable inverse of a room response necessarily exhibits desired noise reduction behavior. The inverse that always reduces additive noise is given by the pseudoinverse of $\mathbf{H}(z)$, and under certain conditions the pseudoinverse is FIR. When RIRs are very long, the exact computation of the pseudoinverse of $\mathbf{H}(z)$, and the verification of the conditions for FIR inversion may not be computationally tractable. Hence a fast algorithm to find an FIR approximation of the pseudoinverse is of a great importance. Kirkeby *et al.* [12] propose the following DFT approach. Consider the N-point discrete Fourier transform (DFT) of the system matrix $\mathbf{H}(z)$: $\mathbf{H}(e^{j\frac{2\pi}{N}k}),\ k = 0, ..., N-1$. The N-point DFT of the pseudoinverse system is given by

$$\mathbf{H}^{\dagger}(e^{j\frac{2\pi}{N}k}) = \left[\tilde{\mathbf{H}}(e^{j\frac{2\pi}{N}k})\mathbf{H}(e^{j\frac{2\pi}{N}k})\right]^{-1}\tilde{\mathbf{H}}(e^{j\frac{2\pi}{N}k})\ ,$$

and an N-tap FIR approximation of the filters of the pseudoinverse can be obtained by applying the N-point inverse DFT to $\mathbf{H}^{\dagger}(e^{j\frac{2\pi}{N}k}),\ k = 0, ..., N-1$, giving filters

$$\hat{h}^{\dagger}_{ml}(n) = \mathrm{IDFT}\{[\mathbf{H}^{\dagger}(e^{j\frac{2\pi}{N}k})]_{ml},\ k = 0, \ldots, N-1\}\ . \quad (6)$$

Impulse responses $h^{\dagger}_{ml}(n),\ m = 1, .., M,\ l = 1, .., L$, of the exact pseudoinverse and their approximations $\hat{h}^{\dagger}_{ml}(n)$ are related according to [5]:

$$\hat{h}^{\dagger}_{ml}(n) = \sum_{i\in\mathbb{Z}} h^{\dagger}_{ml}(n - iN),\ n = 0, \ldots, N-1\ . \quad (7)$$

If N is shorter than the length of the exact pseudoinverse filters, which is always the case when the actual pseudoinverse is IIR, the FIR approximation obtained in this manner is a time-aliased version of the desired inverse system. The length N of the DFT is an important design parameter, which should be set so as to achieve a satisfactory compromise between the accuracy, which requires low aliasing and hence large N, and low implementation complexity and system delay, which require small N.

As a benchmark, we propose the DFT size N equal to the order of the polynomial in the denominator of $\tilde{\mathbf{H}}(z)\mathbf{H}(z)$, so that the discretization in frequency preserves all the information about the denominator of filters in $\mathbf{H}^{\dagger}(z)$. This rule suggests the DFT size

$$N = 2L(L_h - 1) + 1 \quad (8)$$

where L_h is the length of room impulse responses. The length of the corresponding inverse filters would therefore also be $L_g = N$. Note that the DFT size suggested here is one half of the DFT size proposed by Kirkeby *et al.* [12]. If a lower inversion accuracy is acceptable, smaller DFT sizes may be used to decrease computational requirements both for filter design and real-time implementation. Alternatively, FIR inverse filters obtained using a given large N may be symmetrically truncated to decrease the run time computational requirements of the system. The effects of these two approaches to shortening inverse filters are assessed numerically in Section IV.

B. Sensitivity to errors in room impulse responses

Non-blind deconvolution methods implicitly assume a perfect knowledge of room impulse responses. RIRs are however acquired with a limited accuracy. In this section we provide a theoretical assessment of the impact of these errors on the dereverberation accuracy; an experimental assessment is presented in the next section.

Let $\hat{H}_{ij}(z)$ be the acquired room transfer functions used to compute the inverse system $\hat{\mathbf{G}}(z) = \left[\tilde{\hat{\mathbf{H}}}(z)\hat{\mathbf{H}}(z)\right]^{-1}\tilde{\hat{\mathbf{H}}}(z)$, and let the actual transfer functions be $H_{ij}(z)$. The deconvolution using $\hat{\mathbf{G}}(z)$ gives signals $\hat{X}_i(z)$ which differ from the desired signals $X_i(z)$. Corresponding errors $\epsilon_i(e^{j\omega}) = \hat{X}_i(e^{j\omega}) - X_i(e^{j\omega})$ are given in the Fourier domain by

$$\epsilon_i(e^{j\omega}) = \sum_{m=1}^{M}\sum_{l=1}^{L}\hat{G}_{im}(e^{j\omega})\Delta H_{ml}(e^{j\omega})X_l(e^{j\omega})$$

where $\Delta H_{ml}(e^{j\omega}) = H_{ml}(e^{j\omega}) - \hat{H}_{ml}(e^{j\omega})$. To separate the effects of input signals, inverse filters and RIR inaccuracies on the deconvolution error, consider the following upper bound

$$\left|\epsilon_i(e^{j\omega})\right| \leq \|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}\left(\sum_{m=1}^{M}\left|\hat{G}_{im}(e^{j\omega})\right|^2\right)^{\frac{1}{2}}\mathcal{E}_X(e^{j\omega}) \ , \tag{9}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, $\Delta\mathbf{H}(e^{j\omega}) = \mathbf{H}(e^{j\omega}) - \hat{\mathbf{H}}(e^{j\omega})$ is the matrix of room transfer function errors, and $\mathcal{E}_X(e^{j\omega}) = \left(\sum_{l=1}^{L}\left|X_l(e^{j\omega})\right|^2\right)^{\frac{1}{2}}$. Note that this upper bound is obtained by successive application of Cauchy-Schwarz inequality, and it is therefore a tight bound, *i.e.* it is theoretically achievable. The root mean square (RMS) error across all input channels then has the following tight upper bound:

$$\begin{aligned}
&\sqrt{\frac{1}{L}\sum_{i=1}^{L}\left|\epsilon_i(e^{j\omega})\right|^2} \\
&\quad\leq \sqrt{\frac{1}{L}}\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}\left[\sum_{i=1}^{L}\sum_{m=1}^{M}\left|\hat{G}_{im}(e^{j\omega})\right|^2\right]^{\frac{1}{2}}\mathcal{E}_X(e^{j\omega}) \\
&\quad= \sqrt{\frac{1}{L}}\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}\|\hat{\mathbf{G}}\left(e^{j\omega}\right)\|_{\mathrm{F}}\mathcal{E}_X(e^{j\omega}) \ .
\end{aligned}$$

Observe that $\hat{\mathbf{G}}(e^{j\omega})\tilde{\hat{\mathbf{G}}}(e^{j\omega}) = \left[\tilde{\hat{\mathbf{H}}}(e^{j\omega})\hat{\mathbf{H}}(e^{j\omega})\right]^{-1}$, hence $\|\hat{\mathbf{G}}\left(e^{j\omega}\right)\|_{\mathrm{F}} = \sqrt{\sum_{l=1}^{L}1/\left(\sigma_i(e^{j\omega})\right)^2}$, where $\sigma_i(e^{j\omega})$ are singular values of $\hat{\mathbf{H}}(e^{j\omega})$. This identity finally gives the following upper bound for the root mean square deconvolution error:

$$\begin{aligned}
&\sqrt{\frac{1}{L}\sum_{i=1}^{L}\left|\epsilon_i(e^{j\omega})\right|^2} \\
&\quad\leq \sqrt{\frac{1}{L}}\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}\sqrt{\sum_{i=1}^{L}\frac{1}{\sigma_i^2(e^{j\omega})}}\mathcal{E}_X(e^{j\omega}) \\
&\quad\leq \frac{\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}}{\sigma_{min}(e^{j\omega})}\mathcal{E}_X(e^{j\omega}) \ ,
\end{aligned} \tag{10}$$

where $\sigma_{min}(e^{j\omega}) = \min_i \sigma_i(e^{j\omega})$. A quantitative assessment of the inversion accuracy under RIR acquisition errors independent of input signals can be obtained by considering the above error bound when all input signals are equal to the unit Dirac pulse, *i.e.* $X_i(z) = 1,\ i = 1,\ldots,L$. In that case for all $X_i(z)$ and all frequencies $X_i(e^{j\omega}) = 1$, and therefore $\mathcal{E}_X(e^{j\omega}) = \sqrt{L}$. Hence the inversion error in (10) develops as

$$\sqrt{\frac{1}{L}\sum_{i=1}^{L}\left|\epsilon_i(e^{j\omega})\right|^2} \leq \frac{\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}}{\sigma_{min}(e^{j\omega})}\sqrt{L} \ . \tag{11}$$

Recall that error bounds in (10) and (11) are tight and they demonstrates that the extent to which measurement errors affect the deconvolution accuracy depends strongly on how well conditioned the transfer function matrix $\mathbf{H}(e^{j\omega})$ is. If $\mathbf{H}(e^{j\omega})$ is ill-conditioned at some frequency, then even if it is possible to find numerically stable inverse filters, this will cause the upper bound of the RMS error due to measurement errors to be very high. Such a case may occur if the sources or the microphones are positioned very close to each other. Similarly, sources or microphones positioned on or near strong acoustical nodes or anti-nodes will cause this upper bound to be higher at corresponding frequencies. When the approximate pseudoinverse as considered in the previous subsection is used, then such a $\hat{\mathbf{G}}(z)$ is the pseudoinverse of a system $\hat{\mathbf{H}}(z)$ which has an FIR, and therefore stable, pseudoinverse. The approximation procedure, therefore, implicitly performs regularisation of $\mathbf{H}(z)$ which prevents $\sigma_{min}(e^{j\omega})$ from being very low. The approximation, on the other hand increases the RIR error $\Delta\mathbf{H}\left(e^{j\omega}\right)$ due to the time aliasing described by (7), and that has the opposite effect on the overall dereverberation accuracy. Regularization can be also introduced explicitly as $\hat{\mathbf{G}}(z) = (\tilde{\hat{\mathbf{H}}}(z)\hat{\mathbf{H}}(z) + \beta\mathbf{I})^{-1}\hat{\mathbf{H}}(z)$, for some $\beta > 0$, and then $\sigma_{min}(e^{j\omega})$ is lower-bounded by $\sqrt{\beta}$. High level of regularization on the other hand again increases the $\|\Delta\mathbf{H}\left(e^{j\omega}\right)\|_{\mathrm{F}}$ factor in the above bounds. Finding optimal β, or an optimal level or regularization in general, is an interesting problem beyond the scope of the present paper. The problem has been recently investigated experimentally by Hikichi *et al.* [13] for the case of one source and perturbations due to source displacement.

C. Room impulse response errors

It is now of interest to investigate the distortion of room transfer functions,

$$\Delta H_{ml}(e^{j\omega}) = H_{ml}(e^{j\omega}) - \hat{H}_{ml}(e^{j\omega}) \ ,$$

caused by errors in the acquisition of the corresponding impulse responses. A room impulse response $h(n)$ is approximately a series of impulses,

$$h(n) = \sum_{k=0}^{\infty}a_k\delta(n - t_k) \ , \tag{12}$$

arriving at time instants t_k with amplitudes a_k. We focus on errors which manifest as modulations τ_k of arrival instants t_k, perturbations α_k in the observed amplitudes a_k, and additive noise $\eta(n)$. Due to these effects, the acquired impulse response becomes

$$\hat{h}(n) = \sum_{k}(a_k + \alpha_k)\delta(n - t_k - \tau_k) + \eta(n) \ . \tag{13}$$

Note that the discrete-time notation of (12) and (13) implicitly restricts t_k and τ_k to integer values. Nevertheless, the Fourier-domain analysis presented in this subsection holds also for

non-integer delay and delay modulation parameters, and in the simulations reported in the paper $\delta(n - t_k - \tau_k)$ are replaced by impulse responses of corresponding fractional delay filters. Amplitude perturbation factors α_k and noise $\eta(n)$ can be jointly considered just as additive noise, *i.e.* one can write

$$\hat{h}(n) = \sum_k a_k \delta(n - t_k - \tau_k) + \xi(n) \ ,$$

where $\xi(n) = \eta(n) + \sum_k \alpha_k \delta(n - t_k - \tau_k)$.

Towards establishing perturbation bounds, let us first focus on the delay modulation only, that is, consider

$$\hat{h}(n) = \sum_k a_k \delta(n - t_k - \tau_k).$$

The corresponding perturbation of the frequency response

$$\Delta H(e^{j\omega}) = H(e^{j\omega}) - \hat{H}(e^{j\omega}) = \sum_{k=0}^{\infty} a_k e^{-j\omega t_k} \left(1 - e^{-j\omega\tau_k}\right) \ ,$$

can be upper bounded as

$$\left|\Delta H(e^{j\omega})\right| \leq 2 \sum_{k=0}^{\infty} \left| a_k \sin\left(\frac{\omega\tau_k}{2}\right)\right| \ . \tag{14}$$

This upper bound is tight, and therefore theoretically achievable. For low to moderate frequencies, *i.e.* such that $\omega\tau_{max} < \pi$, where $\tau_{\max} = \max\{|\tau_k|\}$, a simpler, albeit looser bound can be established as

$$\left|\Delta H(e^{j\omega})\right| \leq 2 \left|\sin\left(\frac{\omega\tau_{max}}{2}\right)\right| \sum_{k=0}^{\infty} |a_k| \ , \tag{15}$$

and further as

$$\left|\Delta H(e^{j\omega})\right| \leq \omega\tau_{\max} \sum_{k=0}^{\infty} |a_k| \ . \tag{16}$$

The bound in (16) exhibits an increase with frequency at the rate of 3 dB per octave, which is also observed in simulations shown in Fig. 1. The last two upper bounds are valid in the whole frequency range for $\tau_{max} \leq 1$, which is at 44.1 kHz sampling equivalent to 22.7 μs. Another simpler bound, that holds for all frequencies and does not depend on τ_{max}, develops as

$$\left|\Delta H(e^{j\omega})\right| \leq 2 \sum_{k=0}^{\infty} |a_k| \ . \tag{17}$$

It should be noted that the upper bounds in (16) and (17) are very conservative, and while the bound in (14) is theoretically achievable, the error can be expected to fall significantly below all these bounds. Fig. 1 shows two examples of error magnitude response in comparison with the upper bound established by (15) for $\tau_{max} = 0.1323$ and $\tau_{max} = 1.323$, which at 44.1 kHz sampling corresponds to 3 μs and 30 μs, respectively. It can be observed that the actual error is very close to the upper bound in (15) at low frequencies, that it is significantly below the bound for higher frequencies, and that its dependence on frequency and τ_{max} obeys the same laws as the upper bound in (16). These examples were obtained by introducing delay modulation to simulated room impulse responses using variable fractional delay filters as described in the next section.

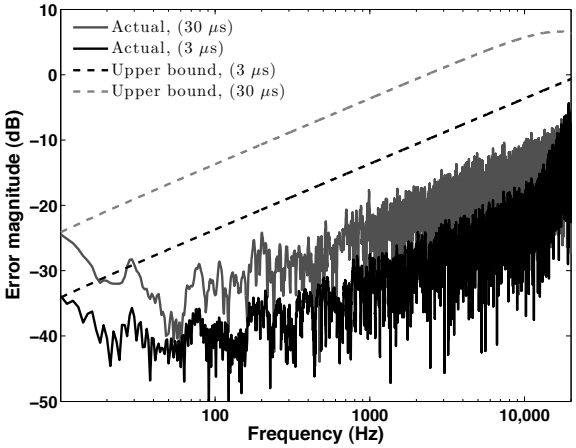


Fig. 1. Error magnitude due to delay modulation. Solid curves are obtained by simulating some typical scenarios with 3 μs and 30 μs delay modulation. Theoretical upper bounds are shown as dashed lines.

The presence of additive error $\xi(n)$ introduces also an additive term in error bounds in (14-17), so the bound in (14), for example, becomes

$$\left|\Delta H(e^{j\omega})\right| \leq 2 \sum_{k=0}^{\infty} \left| a_k \sin\left(\frac{\omega\tau_k}{2}\right)\right| + \left|\Xi(e^{j\omega})\right| ,$$

where $\Xi(e^{j\omega})$ is the Fourier transform of $\xi(n)$. The level of the additive perturbation will be quantified in this paper as a signal-to-noise ratio (SNR) with the following meaning:

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_n h(n)^2}{\sum_n \xi(n)^2} \right) \ . \tag{18}$$

This definition is used in the next section where experimental results are reported.

In conclusion of this section, Fig. 2 shows the magnitude of the error spectrum between the RIR of the example shown in Fig. 1 with $\tau_{max} = 1.323$, and SNRs of 10 dB and 30 dB, which represent severe levels of additive error. It can be observed that the perturbation due to additive error is more pronounced at low frequencies than at high frequencies where the overall perturbation is dominated by the effects of delay modulation.

IV. Experimental Assessment

In this section we present a numerical assessment of the impact of the FIR approximation of the pseudo-inverse and RIR acquisition errors on dereverberation accuracy. For that purpose we measured room impulse responses for $L = 4$ sources and $M = 5$ microphones positioned at random in a rectangular acoustic isolation booth of dimensions 6.52 m $\times$ 4.56 m $\times$ 2.1 m with a reverberation time of RT60 $\approx$ 200 ms. Five AKG C417 miniature microphones were used to measure the room impulse responses due to four MACKIE HR824 loudspeakers using the maximum-length sequences (MLS) method [28]. The sampling frequency used in the measurements was $F_s = 44.1$ kHz, and measured RIRs were truncated to a length of $L_h = 15358$ samples. The positions of sources and microphones are shown in Fig. 3. We

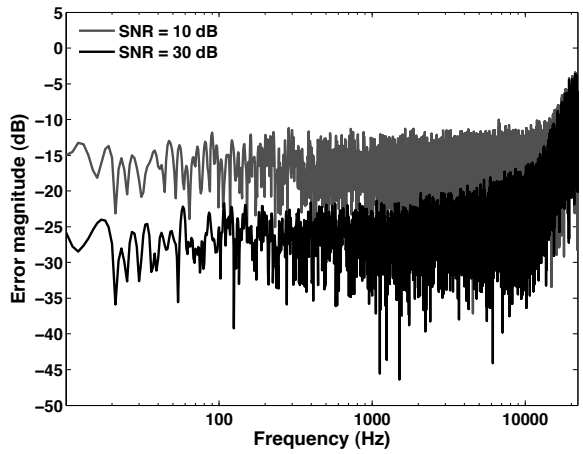


Fig. 2. Error magnitude due to the combined effect of the delay modulation, amplitude errors, and background noise.

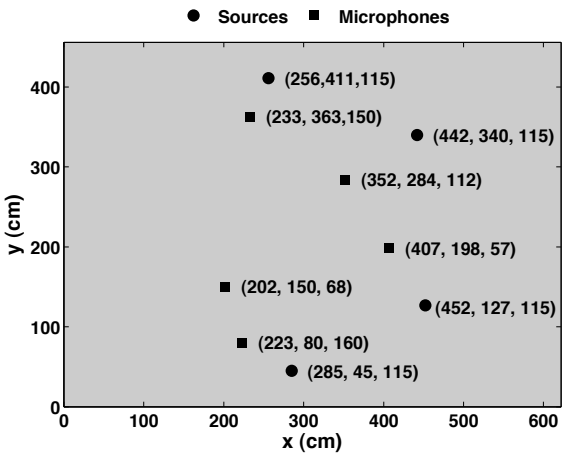


Fig. 3. Top view of the acoustic isolation booth and the coordinates of the sources and microphones used in the measurements. The circles represent sources and the squares represent microphones.

preferred a random placement of microphones and sources so as to avoid a situation in which the transfer matrix of the system would be particularly well conditioned by design, hence in applications with properly designed microphone arrays the inversion might perform better than suggested by results reported in this section. Random positioning is also beneficial as it prevents geometrical symmetries (e.g. a linear microphone array positioned at the centre of the room) which could decrease system order.

The delay modulation was simulated by filtering the RIRs with a 10^{th} order tunable allpass fractional delay filter [29]. The effective delay modulation of the RIR sample at a time instant n achieved in this manner is $\tau(n) = \rho(n)\tau_{\max}$ where $\rho(n)$ is a modulating function. To avoid non-stationary transient errors due to coefficient update in the employed fractional delay filter, the modulating function $\rho(n)$ was selected as a smooth function, in particular

$$\rho(n) = \sin\left(\frac{2\pi T_s}{T_{mod}} n + \phi\right) \ ,$$

where ϕ is the random phase offset, T_s is the sampling interval, and T_{mod} is the period which was set to 0.02 s.

In numerical assessments of the method, the following three error metrics are used:

i) *Dereverberation error energy* [16]. It measures the deviation of the equalized impulse response from the ideal, Dirac impulse, in the time domain:

$$J_i = 10\log_{10}\left(\frac{1}{L_q}\sum_n \left(\delta(n) - q_{ii}(n)\right)^2\right) \ , \qquad (19)$$

where $q_{ii}(n)$ is the impulse responses corresponding to $Q_{ii}(z)$, the i-th diagonal element of the inverted transfer function $\mathbf{Q}(z) = \mathbf{G}(z)\mathbf{H}(z)$, and L_q is its length.

ii) *Deverberation spectral deviation* [1]. It is a measure of the flatness of the equalized response calculated as:

$$V_i = \left(\frac{1}{L_q}\sum_{k=0}^{L_q-1}\left(10\log_{10}\left|Q_{ii}(e^{j\omega_k})\right| - \bar{Q}_i\right)^2\right)^{\frac{1}{2}} \ , \qquad (20)$$

where $\bar{Q}_i = \frac{1}{L_q}\sum_{k=0}^{L_q-1} 10\log_{10}\left|Q_{ii}(e^{j\omega_k})\right|$ and $\omega_k = 2\pi k/L_q$. If the equalized response is the ideal unit impulse, its spectrum is flat and then $V_i = 0$. This metric provides information about the effectiveness of the method in reducing the spectral coloration due to room acoustics.

iii) *Signal-to-interference ratio (SIR)* [30]. It is generally used in assessing the performance of source separation algorithms in terms of the level of leakage from unwanted sources. While the proposed algorithm is not directly aimed at separation of mixtures, separation is inherent in the algorithm. There are many different definitions for SIR in the source separation literature (see [31] for a review) and we have adopted the following:

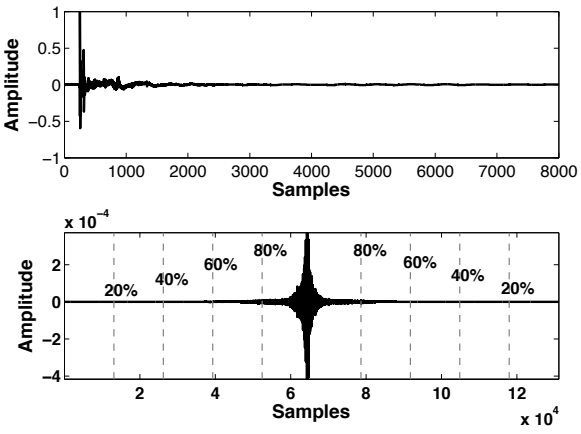
$$\text{SIR}_i = 10\log_{10}\frac{\sum_n |q_{ii}(n)|^2}{\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_n |q_{ij}(n)|^2} \qquad (21)$$

where $q_{ij}(n)$ is the impulse responses corresponding to $Q_{ij}(z)$ entry of $\mathbf{Q}(z) = \mathbf{G}(z)\mathbf{H}(z)$.

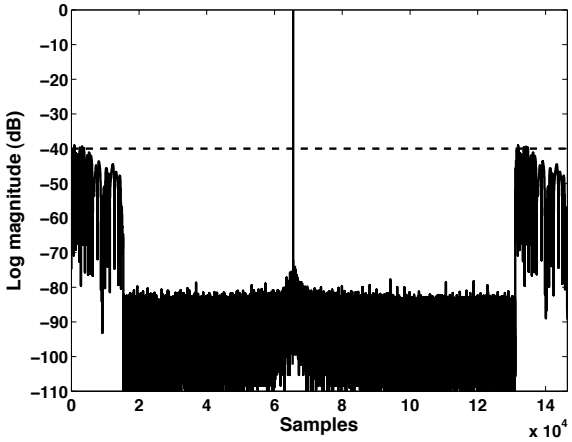
A. Effects of the DFT size and length of inverse filters

Let us consider the effects of the size of the DFT used in the approximate inversion and the effects of the truncation of inverse filters on the dereverberation accuracy. For an illustration, the inverse filter matrix was first computed for the system of $L = 4$ sources and $M = 5$ microphones. For $L_h = 15358$, the DFT size suggested in Section III-A is $2L(L_h - 1) + 1 = 122857$. The DFT size $N = 2^{17}$ is the smallest radix-2 larger than this value.

The top panel of Fig. 4(a) shows the first 8000 samples (using $F_s = 44.1$ kHz) of the normalized RIR, $h_{3,1}(n)$, obtained in the measurements. A typical inverse filter, $g_{2,1}(n)$, is shown in the bottom panel. It may be observed that the filter has many coefficients which are negligibly small. This suggests that it might be possible to symmetrically truncate filters in



(a)



(b)

Fig. 4. A typical room impulse response, inverse filter and equalized response. (a) Top panel: normalized RIR, $h_{3,1}(n)$, for the first source and the third microphone. Bottom panel: the inverse filter, $g_{2,1}(n)$, with different levels of symmetric truncation denoted by dashed lines, and (b) Log magnitude of the equalized response, $10\log_{10}|q_{3,3}(n)|$, for the third channel.

order to reduce the computational requirements. Truncation levels of 20%, 40%, 60%, and 80% are also shown on the plot. The magnitude of the impulse response of the equalized transfer function for the third source, $10\log_{10}|q_{3,3}(n)|$, is shown in 4(b). The most pronounced departures from the Dirac appear at the beginning and at the end of the equalized response and are caused by the aliasing due to finite DFT size. It may be observed that the equalized response is a delayed Dirac. This system delay is determined by the length of the inverse filters, which depends on the DFT size and the extent of the filter truncation. The initial delay for an equalized impulse response obtained with a DFT size of N, and a truncation level of $P\%$ is $\lfloor N(1-\frac{P}{100})/2 \rfloor$.

1) Effects of the DFT size: The DFT size is an important design parameter, having an impact on the time-domain aliasing and the accuracy of the method on one side, and the length of inverse filters, and consequently the complexity of their computation and implementation, on the other. Effects of the DFT size on the accuracy of the dereverberation are assessed experimentally in this section.

Fig. 5 shows the dereverberation error energy, dereverberation spectral deviation, and SIR for DFT sizes between $N = 2^{14}$ and $N = 2^{17}$, for different number of sources, L, with a fixed number of microphones, $M = 5$. Here, the error metrics were calculated by obtaining the inverse system response for all ${}_LC_k$, $k = 1 \cdots L$ combinations of $L = 4$ sources. The error bars in this and other figures in the paper denote maximum and minimum values of the respective metric. Results corresponding to a given fixed L are not placed on the same vertical line, but are displaced horizontally for visualization clarity.

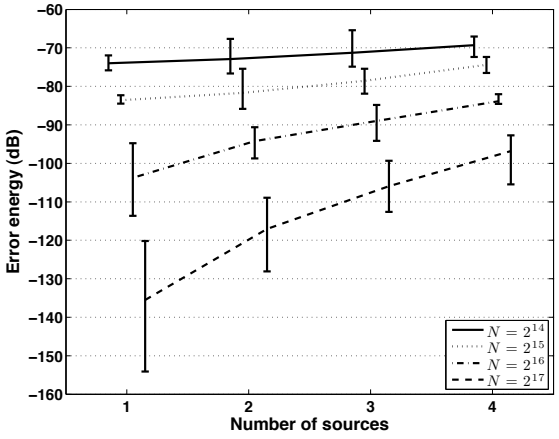
It can be observed that, as expected, larger DFT sizes cause error energy and spectral deviation metrics to be smaller and SIR to be higher. Note that for $L = 1$ and $L = 4$ sources the corresponding DFT sizes (closest radix-2 numbers) according to the rule proposed in Section III-A are $N = 2^{15}$ and $N = 2^{17}$, respectively. Values of the three error metrics reported in the figure indicate that the proposed rule for deciding on the DFT size ensures very high accuracy, but that smaller sizes could also suffice, as well as that by increasing the size beyond that value improves the accuracy further.

An important side effect of reducing DFT size is possible occurrence of pre-echoes and spurious echoes in the equalized response due to aliasing. Two equalized responses for $L = 4$ and $M = 5$ obtained with DFT sizes $N = 2^{12}$ and 2^{17} are shown in Fig. 6. The pre-echo and the spurious echo in the equalized response for $N = 2^{12}$ are marked in the plot with an arrow and a circle, respectively. The higher level of the error in the equalized response for shorter DFT size is also evident in the figure. Note that the scales of the two plots are different.

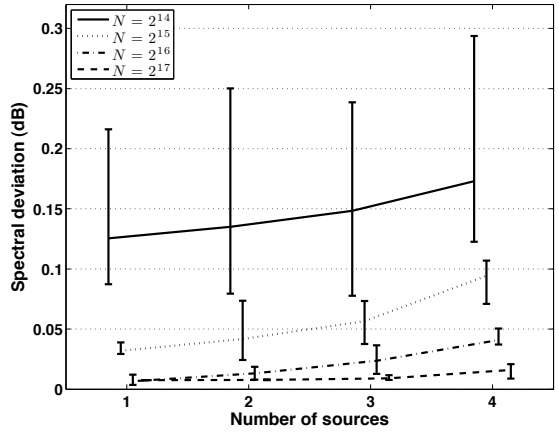
While the experimental results presented in this section can only represent the performance of the method for a specific measurement condition, the results indicate that computational savings may be possible by reducing the DFT size. It should be noted that the DFT size should not be decreased below a certain value so as not to cause audible echoes or pre-ringing in the equalized responses. In the example that we provided, the DFT size can be reduced to $N = 2^{14}$ without decreasing the overall performance of the system significantly.

2) Effects of filter length: The bottom panel of Fig. 4(a) shows that the energy of the inverse filters is concentrated near the middle and that their impulse responses exhibit rapid decay away from the centre. This suggests that it might be possible to truncate these filters in order to reduce computational requirements of the dereverberation. The truncation would however degrade dereverberation accuracy. Here, we assess this accuracy impairment numerically for different levels of filter truncation.

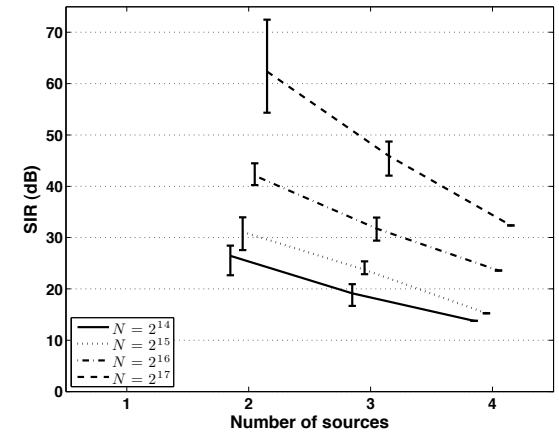
The DFT size used in the inversion is $N = 2^{17}$, and therefore, the length of the inverse filters is $L_g = N = 2^{17}$. Fig. 7 shows the error metrics for different truncation levels between 20% and 80%. The error metrics for equalized responses using filters of the original length $L_g = N$ are also given for comparison. It may be observed that although the error metrics show progressive degradation of the deconvolution accuracy as the filter-length shortens, good results are obtained even when the inverse filters are truncated to 20% of their original length. It should be noted that although we only consider



(a)



(b)



(c)

Fig. 5. Effects of the DFT size, N, on deconvolution accuracy for $M = 5$ microphones and $L = 1, \ldots, 4$ sources. The curves represent mean values, averaged over considered channels, and the error bars represent maximum and minimum values of the corresponding metric. (a) Dereverberation error energy, J, and (b) Dereverberation spectral deviation, V, and (c) SIR.

simple truncation in this article these inverse filters can also be shortened by smoothing [17], [18].

Fig. 8 shows the equalized responses for a single channel obtained with truncated filters for $M = 5$ and $L = 4$.

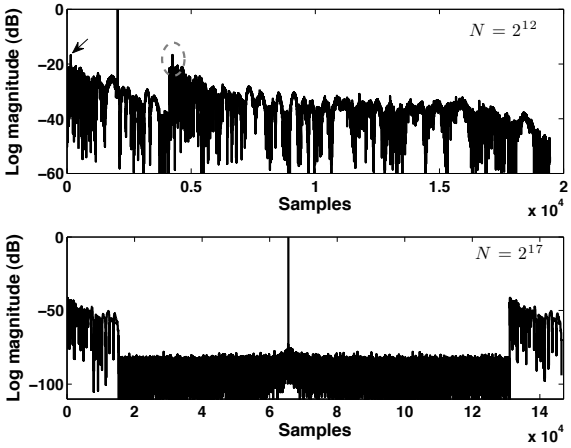


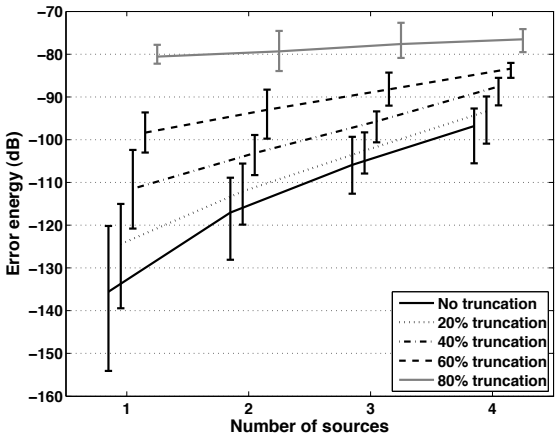
Fig. 6. Pre-echo and spurious echo due to short DFT size. The top plot shows an equalized response obtained using inverse filters computed via the $N = 2^{12}$-point DFT. Pre-echo is denoted by an arrow and the spurious echo is circled. The bottom plot shows the equalized response for $N = 2^{17}$.

It may be observed that even at 80% truncation the error level is still around -40 dB and no spurious echoes or pre-echoes are observed. A comparison between these results and those shown in Section IV-A1 suggests that truncation of inverse filters is a better way of reducing computational requirements than reducing the DFT size both in terms of inversion accuracy and for avoiding spurious echoes and pre-echoes. This observation could be expected considering that a larger DFT size decreases time-domain aliasing, as it follows from (7), so that the samples retained after the truncation are closer to the original impulse responses than those obtained by decreasing the DFT size to the desired filter length.

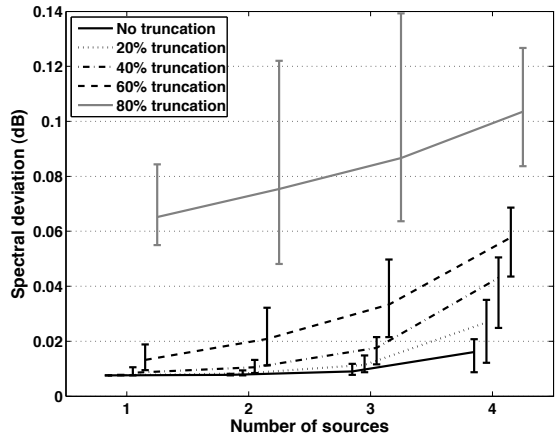
B. Robustness to RIR errors

Now we consider the impact of errors in the acquisition of RIRs on the accuracy of the deconvolution, focusing on additive errors and the delay modulation as discussed in Section III-C. We consider the effect of not knowing the RIRs with infinite accuracy, but with a certain additive error, and then using these inaccurate impulse responses to compute the pseudoinverse system according to (6). The inverse filters thus suffer from two forms of error: one which is introduced by applying additive error and delay modulation to RIRs and the other by using the FIR approximations of the inverse filters. The effects of the FIR approximation are discussed in general terms in Section III-B, however the plots in Fig. 5 corresponding to $N = 2^{17}$, which is the DFT size and filter length used in these experiments, show that this FIR approximation practically has a negligible effect on the error of the inversion of the simulated system.

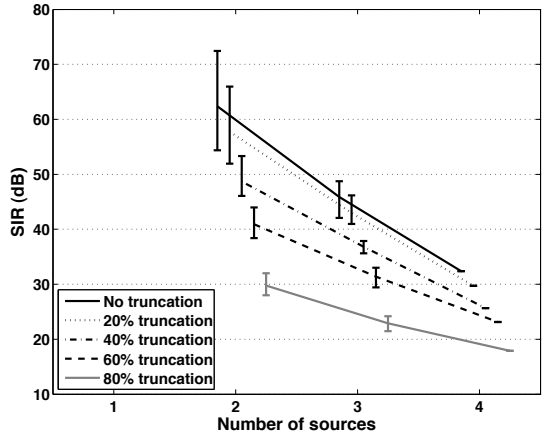
Fig. 9(a) shows an example of log magnitude plots of equalized responses for SNR levels of 10, 20, 30, and 40 dB (as defined in (18)). At this point we consider only errors due to ambient noise which we model as pink noise. It can be observed that, not surprisingly, the error floor in the equalized response increases proportionally with the level of the additive error, however it is always low in comparison



(a)



(b)



(c)

Fig. 7. Effects of inverse filter truncation on deconvolution accuracy in the case of $M = 5$ microphones and $L = 1, \ldots, 4$ sources. The curves represent mean values, averaged over considered channels. The error bars represent maximum and minimum values of the corresponding metric. (a) Deconvolution error energy, J, (b) Spectral deviation, V, and (c) SIR.

with the SNR level. A subjective evaluation of the effects of these perturbations is beyond the scope of this paper, however in preliminary listening tests no preringing effects were audible for the tested conditions. Fig. 9(b) shows the

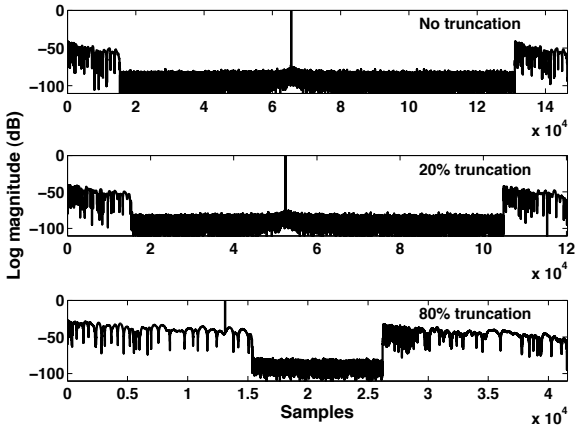


Fig. 8. Effects of inverse filter truncation on equalized responses in the time domain. Log magnitude of equalized responses $q_{3,3}(n)$ for different levels of filter truncation.

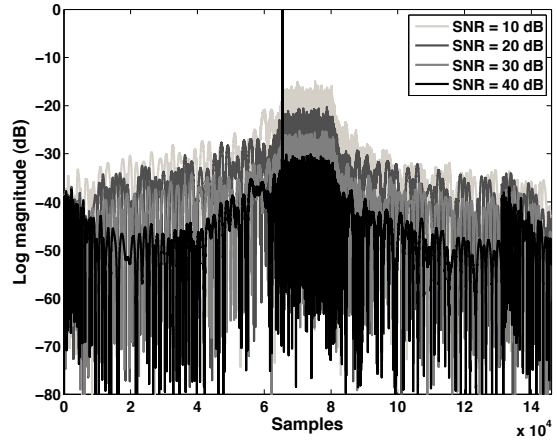
effect of increasing the maximum delay modulation. The equalized responses shown in the figure were obtained for maximum delay modulations of $\tau_{max} = 3$ μs, $\tau_{max} = 10$ μs and $\tau_{max} = 30$ μs. It can be observed that even if the level of delay modulation is increased by an order of magnitude, the error floor of the equalized response does not increase significantly.

Fig. 10 shows the dependence of the three metrics of the inversion quality on RIR perturbation levels. As expected, the deconvolution error energy metric, J, and spectral deviation V systematically increase with the perturbation. Still, the error energy J is less than about -50 dB even when the RIR error reaches 10 dB SNR. Spectral deviation decreases significantly when SNR increases from 10 to 40 dB, attaining 0.1 dB level at 40 dB SNR. At 30 dB SNR, the SIR ranges from around 8 dB for 4 sources to close to 20 dB for 2 sources, which presents a good level of separation.

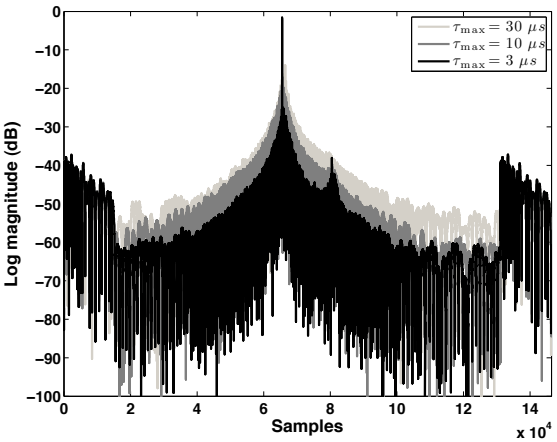
Finally, Fig. 11 shows the same error metrics for different delay modulations. It can be observed that the increase in the delay modulation affects the error energy more severely than spectral deviation. Increasing the level of delay modulation degrades the SIR, but still a reasonable separation is achieved if the number of sources is not too close to the number of microphones.

Note again that as it was shown in Section III-B, the robustness of system inversion to the accuracy of RIR acquisition depends critically on how well conditioned the transfer function, $\mathbf{H}(z)$, of the system is. The example presented in this section is intended merely for providing some quantitative intuition about error levels in an environment where no special arrangement of microphones and sources is used to ensure that the transfer function of the system is particularly well conditioned.

In order to control perturbation levels, in the examples provided up to this point additive noise and delay modulation were added to the measured RIRs synthetically. Next, we measured a second set of impulse responses after a time interval of 30 minutes after acquiring the first set of impulse responses. While the RIRs obtained were similar as expected,



(a)

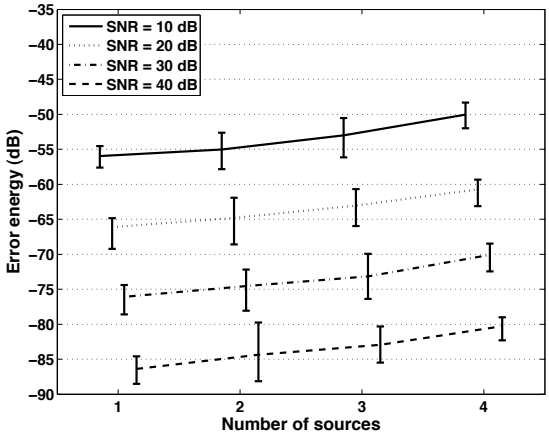


(b)

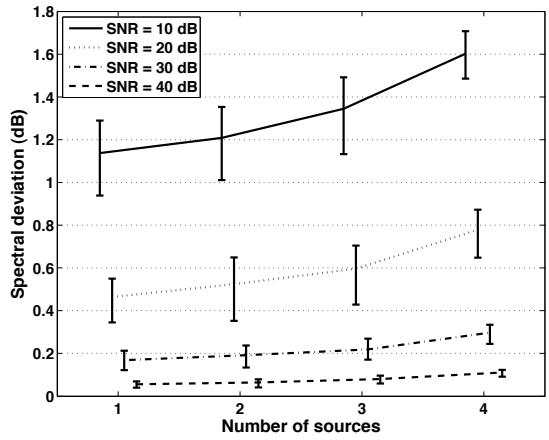
Fig. 9. Effects of delay modulation and additive RIR errors on the responses of inverted systems. The equalized responses of the first channel, $10\log_{10}|q_{11}(n)|$, for $M = 5$, and $L = 4$. (a) Different SNR levels, and (b) delay modulations of $\tau_{max} = 3$ μs, $\tau_{max} = 10$ μs, and $\tau_{max} = 30$ μs.

there were small but significant differences between the two sets of RIRs. These differences occur due to many factors, including local temperature fluctuations, background noise, small displacements of the microphones, finite precision of the measurement, etc. The average error between two sets of RIRs normalized with respect to the first set of RIRs was -21.70 dB with a maximum error of -15.63 dB and a minimum error of -33.42 dB.

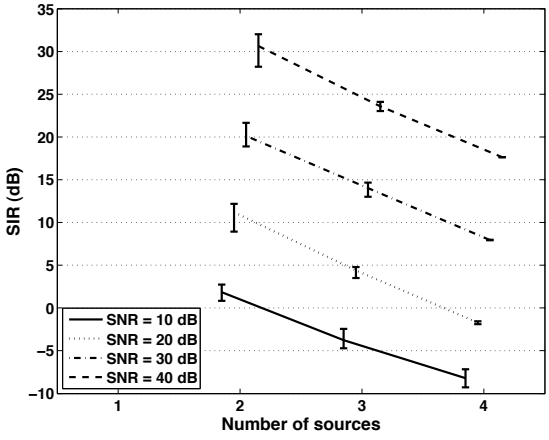
The inverse system designed using the first set of RIRs with $N = 2^{17}$ was then used to invert the system with the new set of RIRs. The same metrics as obtained for the earlier examples were also obtained for all source combinations from $L = 1$ to $L = 4$ and $M = 5$ microphones. The results are summarized in Fig. 12. It can be observed that error energy and spectral deviation are very low. The lowest SIR is observed for $L = 4$ case, and still presents an acceptable level of separation. Furthermore, these results, obtained with real measurement discrepancies, are better than expected based on simulations which involved artificially added perturbations. This can be observed by comparing results in Fig. 12 with the



(a)



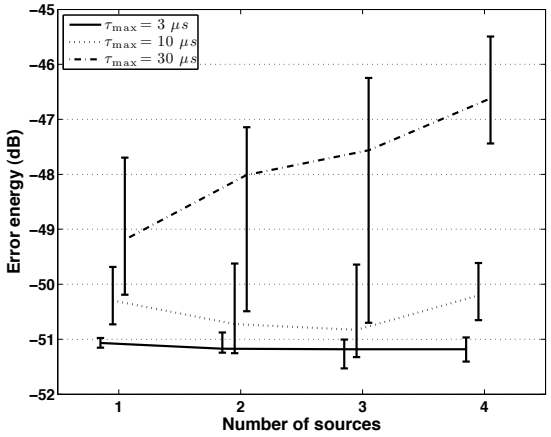
(b)



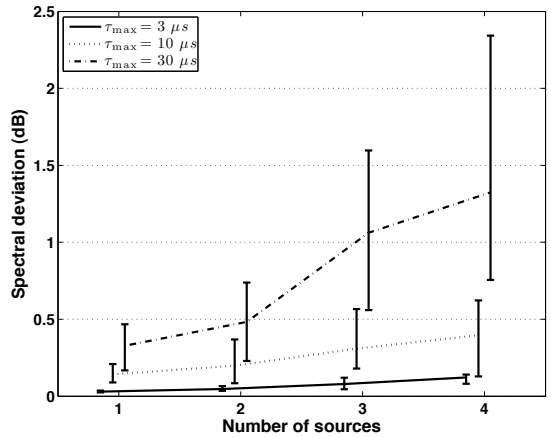
(c)

Fig. 10. Effects of additive noise on deconvolution accuracy parameterized by the SNR level. The plots represent averages over individual channels parameterized by SNR due to additive pink noise. The error bars represent maximum and minimum values of the corresponding metric. (a) Deconvolution error energy, J, (b) Spectral deviation, V, and (c) SIR.

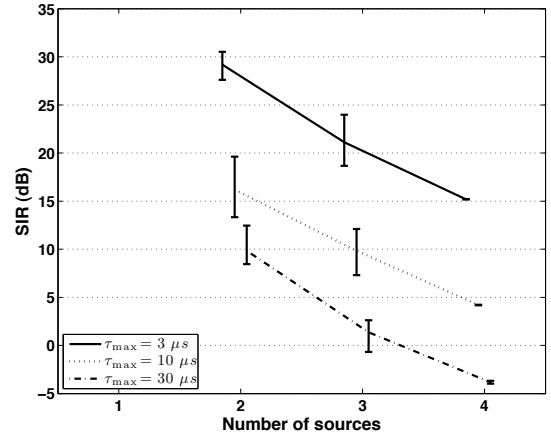
results in Fig. 10 corresponding to 20 dB SNR. Note that neither the experiment with measured impulse responses, nor the simulations with pink noise model properly the additive error caused by RIR truncation, which is known to have



(a)



(b)



(c)

Fig. 11. Effects of delay modulation on dereverberation accuracy. The plots represent averages over individual channels parameterized by delay modulation. The error bars represent maximum and minimum values of the corresponding metric. (a) Deconvolution error energy, J, (b) Spectral deviation, V, and (c) SIR.

a more complex spectral shape than white or pink noise. An informal subjective listening test also confirmed that the investigated algorithm provides a good level of dereverberation and separation without significant preringing artifacts.

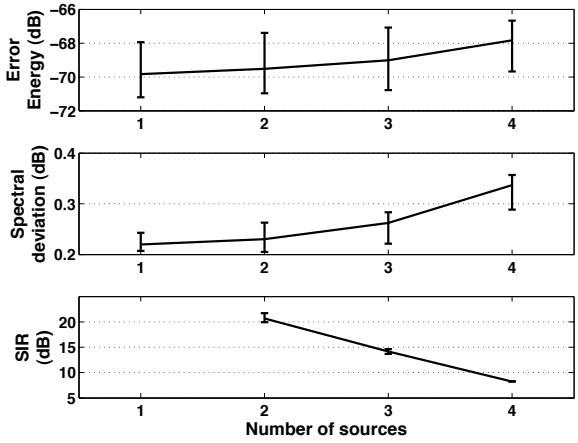


Fig. 12. Error metrics for a realistic measurement error scenario. The plots represent averages over individual channels. The error bars represent maximum and minimum values of the corresponding metric. (a) Deconvolution error energy, J, (b) Spectral deviation, V, and (c) SIR.

V. Conclusion

This paper presents a study of several aspects of non-blind multichannel dereverberation. Conditions for perfect dereverberation using stable or FIR filters are established. When perfect deconvolution is possible, the inverse system is not unique. The solution which is optimal in terms of robustness to additive noise is provided by the pseudoinverse of the system transfer function. A necessary and sufficient condition for the pseudoinverse to be FIR is established. The sensitivity of multichannel dereverberation to perturbations of acquired room impulse responses is also investigated theoretically and numerically. Simulation results suggest that multichannel dereverberation of a well conditioned system using its pseudoinverse, or an FIR approximation thereof, is robust to perturbations which are expressed as delay modulation and additive error.

Acknowledgment

The authors would like to thank the anonymous reviewers of this article for the valuable feedback and insightful comments.

References

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, November 1994.

[2] M. Karjalainen, T. Paatero, J. N. Mourjopoulos, and P. D. Hatziantoniou, "About room response equalization and dereverberation," in *Proc. 2005 IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA'05)*, New Paltz, NY, USA, 16-19 October 2005, pp. 183–186.

[3] P. A. Naylor and N. D. Gaubitch, Eds., *Speech dereverberation*. Springer-Verlag, 2010.

[4] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, July 1979.

[5] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing*, 2nd ed. Prentice-Hall, 1999.

[6] J. L. Flanagan and R. C. Lummis, "Signal processing to reduce multipath distortion in small rooms," *J. Acoust. Soc. Am.*, vol. 47, no. 6, pp. 1475–1481, June 1970.

[7] R. B. Schulein, "In situ measurement and equalization of sound reproduction systems," *J. Audio Eng. Soc.*, vol. 23, no. 3, pp. 178–186, April 1975.

[8] J. Mourjopoulos, P. M. Clarkson, and J. K. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed-phase signals," in *Proc. 1982 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP'82)*, 1982, pp. 1858–1861.
[9] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–585, July/August 1999.
[10] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 2, pp. 145–152, February 1988.
[11] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912–915, October 1977.
[12] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña-Bustamente, "Fast deconvolution of multichannel systems using deconvolution," *IEEE Trans. on Speech and Audio Process.*, vol. 6, no. 2, pp. 189–194, March 1998.
[13] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuation," *EURASIP J. Adv. Signal Process.*, vol. Article ID 34013, 2007.
[14] U. P. Svensson and J. L. Nielsen, "Errors in MLS measurements caused by time variance in acoustic systems," *J. Audio Eng. Soc.*, vol. 47, no. 11, pp. 907–927, November 1999.
[15] G. W. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. 2003 Int. Workshop on Acoust. Echo and Noise Cont. (IWAENC2003)*, Kyoto, Japan, Septemper 2003, pp. 67–70.
[16] P. D. Hatziantoniou and J. N. Mourjopoulos, "Errors in real-time room acoustics dereverberation," *J. Audio Eng. Soc.*, vol. 52, no. 9, pp. 883–899, September 2004.
[17] ——, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, April 2000.
[18] S. Bharitkar, C. Kyriakakis, and T. Holman, "Variable-octave complex smoothing for loudspeaker-room response equalization," in *Proc. IEEE Int. Conf. Consumer Elect.*, Las Vegas, NV, USA, January 2008, pp. 6.1–3.
[19] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound Vib.*, vol. 102, no. 2, pp. 217–228, 1985.
[20] L. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 3–26, January/February 2003.
[21] B. D. Radlović, 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.
[22] S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Robustness of spatial average equalization: A statistical reverberation model approach," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3491–3497, December 2004.
[23] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. on Speech and Audio Process.*, vol. 2, no. 2, pp. 320–328, April 1994.
[24] H. J. S. Smith, "On systems of linear indeterminate equations and congruences," *Philos. Trans. R. Soc. London, Ser. A*, vol. 151, pp. 293–326, 1861.
[25] I. Daubechies, *Ten Lectures on Wavelets*, 1st ed. SIAM, 1992.
[26] Z. Cvetković and M. Vetterli, "Oversampled filter banks," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1245–1255, May 1998.
[27] B. S. Olswang and Z. Cvetković, "Separation of audio signals into direct and diffuse soundfields for surround sound," in *Proc. 2006 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP'06)*, vol. 5, Toulouse, France, 14-19 May 2006, pp. 357–360.
[28] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, June 1989.
[29] H. Hacıhabiboğlu, B. Günel, and A. M. Kondoz, "Analysis of root displacement interpolation method for tunable allpass fractional-delay filters," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4896–4906, October 2007.
[30] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation*, no. 261-266, Aussois, France, January 1999.
[31] K. E. Hild II, D. Erdogmus, and J. Principe, "Experimental upper bound for the performance of convolutive source separation methods," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 627–635, February 2006.

Hüseyin Hacıhabiboğlu (S'96-M'00) received the B.Sc. (honors) degree from the Middle East Technical University (METU), Ankara, Turkey, in 2000, the M.Sc. degree from the University of Bristol, Bristol, U.K., in 2001, both in electrical and electronic engineering, and the Ph.D. degree in computer science from Queen's University Belfast, Belfast, U.K., in 2004. He held research positions at University of Surrey, Guildford, U.K. (2004-2008) and King's College London, London, U.K. (2008-2011). Currently, he is a lecturer at Informatics Institute, Middle East Technical University, Ankara, Turkey. His research interests include audio signal processing, room acoustics modeling and simulation, multichannel audio systems, psychoacoustics of spatial hearing, and microphone arrays. Dr. Hacıhabiboğlu is a member of the IEEE Signal Processing Society, Audio Engineering Society (AES), Turkish Acoustics Society (TAD), and the European Acoustics Association (EAA).

Zoran Cvetković (SM'04) received the Dipl.Ing.El. and Mag.El. degrees from the University of Belgrade, Belgrade, Yugoslavia, in 1989 and 1992, respectively, the M.Phil. degree from Columbia University, New York, in 1993, and the Ph.D. degree in electrical engineering from the University of California, Berkeley, in 1995. He held research positions at EPFL, Lausanne, Switzerland (1996), and at Harvard University, Cambridge, (2002-2004). From 1997 to 2002, he was a Member of Technical Staff at AT&T Shannon Laboratory. He is now a Reader in Signal Processing at King's College London, London, U.K. His research interests are in the broad area of signal processing, ranging from theoretical aspects of signal analysis to applications in source coding, telecommunications, and audio and speech technology.