

RESEARCH

Open Access



Improving the conditioning of the optimization criterion in acoustic multi-channel equalization using shorter reshaping filters

Ina Kodراسي*  and Simon Doclo

Abstract

In acoustic multi-channel equalization techniques, such as complete multi-channel equalization based on the multiple-input/output inverse theorem (MINT), relaxed multi-channel least-squares (RMCLS), and partial multi-channel equalization based on MINT (PMINT), the length of the reshaping filters is generally chosen such that perfect dereverberation can be achieved for perfectly estimated room impulse responses (RIRs). However, since in practice the available RIRs typically differ from the true RIRs, this reshaping filter length may not be optimal. This paper provides a mathematical analysis of the robustness increase of equalization techniques against RIR perturbations when using a shorter reshaping filter length than conventionally used. Based on the condition number of the (weighted) convolution matrix of the RIRs, a mathematical relationship between the reshaping filter length and the robustness against RIR perturbations is established. It is shown that shorter reshaping filters than conventionally used yield a smaller condition number, i.e., a higher robustness against RIR perturbations. In addition, we propose an automatic non-intrusive procedure for determining the reshaping filter length based on the L-curve. Simulation results confirm that using a shorter reshaping filter length than conventionally used yields a significant increase in robustness against RIR perturbations for MINT, RMCLS, and PMINT. Furthermore, it is shown that PMINT using an optimal intrusively determined reshaping filter length outperforms all other considered techniques. Finally, it is shown that the automatic non-intrusively determined reshaping filter length in PMINT yields a similar performance as the optimal intrusively determined reshaping filter length.

Keywords: Speech dereverberation, Condition number, Reshaping filter length, L-curve

1 Introduction

The microphone signals recorded in many hands-free speech communication applications, such as teleconferencing, voice-controlled systems, or hearing aids, do not only contain the desired speech signal but also attenuated and delayed copies due to reverberation. While early reverberation may be desirable [1–3], late reverberation may degrade the perceived speech quality and intelligibility [4–6] as well as the performance of automatic speech recognition systems [7, 8]. In order to mitigate these detrimental effects of reverberation, several single-channel and

multi-channel dereverberation techniques have been proposed [9], with multi-channel techniques being generally preferred since they are able to exploit both the spectro-temporal and the spatial characteristics of the received microphone signals. Existing multi-channel dereverberation techniques can be broadly classified into spectro-temporal enhancement techniques [10–14], probabilistic modeling-based techniques [15–18], and acoustic multi-channel equalization techniques [19–26]. Acoustic multi-channel equalization techniques aim to reshape the available room impulse responses (RIRs) between the speaker and the microphone array. Since in theory they can achieve perfect dereverberation [19], they represent an attractive approach to speech dereverberation.

*Correspondence: ina.kodراسي@uni-oldenburg.de

Department of Medical Physics and Acoustics, University of Oldenburg, 26111 Oldenburg, Germany

A well-known complete multi-channel equalization technique aiming at acoustic system inversion is the multiple-input/output inverse theorem (MINT)-based technique [19], which however suffers from drawbacks in practice. Since the available RIRs typically differ from the true RIRs due to fluctuations (e.g., temperature or position variations [27]) or due to the sensitivity of blind system identification (BSI) and supervised system identification (SSI) methods to near-common zeros or interfering noise [28–30], MINT generally fails to invert the true RIRs, possibly leading to severe distortions in the output signal [22–24, 26]. In order to increase the robustness against RIR perturbations, partial multi-channel equalization techniques, such as relaxed multi-channel least-squares (RMCLS) [23] and partial multi-channel equalization based on MINT (PMINT) [24], have been proposed. Since early reflections tend to improve speech intelligibility [1–3] and late reflections are the major cause of speech intelligibility degradation [4–6], the objective of partial equalization techniques is to shorten the overall impulse response by suppressing only the late reflections. While RMCLS imposes no constraints on the remaining early reflections, PMINT has been shown to be more perceptually advantageous since it also aims to control the remaining early reflections. Although partial equalization techniques can be significantly more robust than MINT, their performance still remains rather susceptible to RIR perturbations [23, 24, 26]. As a result, several methods have been proposed to further increase the robustness against RIR perturbations. In [22, 24], it has been proposed to incorporate regularization, such that the distortion energy due to RIR perturbations is decreased. In [26], it has been proposed to use a signal-dependent penalty function to promote sparsity in the output signal and reduce artifacts generated by non-robust techniques. In [31, 32], it has been proposed to relax the constraints on the filter design by constructing approximate reshaping filters in the subband domain. In [33], it has been proposed to relax the constraints on the filter design by using a shorter reshaping filter length than conventionally used. The objective of this paper is to provide a mathematical analysis of the robustness increase when using a shorter reshaping filter length as well as to propose an automatic non-intrusive procedure for selecting an optimal shorter reshaping filter length.

The length of the reshaping filters in MINT, RMCLS, and PMINT is conventionally chosen such that perfect dereverberation can be achieved for perfectly estimated RIRs. As already mentioned, since in practice the available RIRs typically differ from the true RIRs, this choice of the reshaping filter length yields a high sensitivity to RIR perturbations. In [33], it has been analytically shown that decreasing the reshaping filter length increases the robustness for MINT and PMINT only if the

multi-channel convolution matrix of the RIRs is a square matrix. In this paper, it is analytically shown that decreasing the reshaping filter length increases the robustness of MINT, RMCLS, and PMINT independently of the dimension of the (weighted) multi-channel convolution matrix of the RIRs. A mathematical relationship between the reshaping filter length and the condition number of the (weighted) multi-channel convolution matrix of the available RIRs, hence, the sensitivity of equalization techniques to RIR perturbations, is derived. We show that shorter reshaping filters than conventionally used yield a smaller condition number, i.e., a higher robustness against RIR perturbations.

In general, the reshaping filter length yielding optimal performance can only be determined intrusively (i.e., using a clean reference signal), obviously limiting the practical applicability. Hence, we also propose and investigate an automatic non-intrusive selection procedure for the reshaping filter length based on the L-curve [34–36].

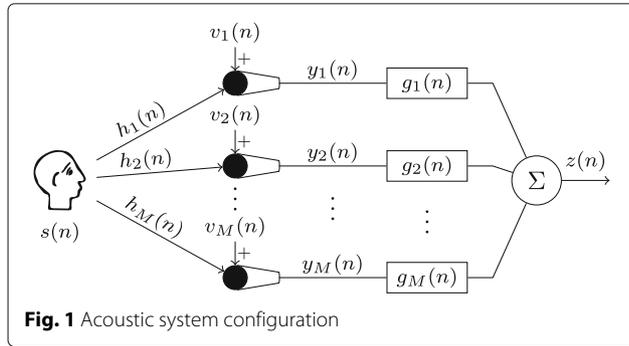
Simulation results for several acoustic systems and RIR perturbations show by means of instrumental performance measures that using shorter reshaping filters in MINT, RMCLS, and PMINT significantly increases the robustness against RIR perturbations. In addition, it is demonstrated that PMINT using the optimal intrusively determined reshaping filter length outperforms the other considered equalization techniques, yielding a larger reverberant energy suppression and perceptual speech quality improvement. Furthermore, it is shown that the non-intrusively determined reshaping filter length yields a nearly optimal performance for PMINT.

The paper is organized as follows. In Section 2, the considered acoustic configuration and the used notation is introduced. In Section 3, state-of-the-art acoustic multi-channel equalization techniques, i.e., MINT, RMCLS, and PMINT, are briefly reviewed. In Section 4, the sensitivity of these equalization techniques to RIR perturbations is evaluated by means of the condition number of the (weighted) convolution matrix and analytical insights on increasing the robustness by decreasing the reshaping filter length are provided. In Section 5, the automatic non-intrusive procedure for determining the reshaping filter length is discussed. Using instrumental performance measures, the dereverberation performance of all considered techniques is compared in Section 6.

2 Configuration and notation

We consider an acoustic system with a single speech source and M microphones, as depicted in Fig. 1. The m -th microphone signal $y_m(n)$, $m = 1, 2, \dots, M$, at discrete time index n , is given by

$$y_m(n) = \underbrace{h_m(n) * s(n)}_{x_m(n)} + v_m(n), \quad (1)$$



where $*$ denotes convolution, $s(n)$ is the clean speech signal, $h_m(n)$ is the RIR between the speech source and the m -th microphone, $x_m(n)$ is the reverberant speech component, and $v_m(n)$ is the noise component. Since acoustic multi-channel equalization techniques generally design reshaping filters without taking the additive noise into account, in the following it is assumed that $v_m(n) = 0$, and hence, $y_m(n) = x_m(n)$.

Using the filter-and-sum structure in Fig. 1, the output signal $z(n)$ is equal to the sum of the filtered microphone signals, i.e.,

$$z(n) = \sum_{m=1}^M x_m(n) * g_m(n) \quad (2)$$

$$= s(n) * \underbrace{\sum_{m=1}^M h_m(n) * g_m(n)}_{c(n)}, \quad (3)$$

where $g_m(n)$ is the filter applied to the m -th microphone signal and $c(n)$ denotes the equalized impulse response (EIR) between the speech source and the output of the system. In vector notation, the RIR \mathbf{h}_m and the filter \mathbf{g}_m are given by

$$\mathbf{h}_m = [h_m(0) \ h_m(1) \ \dots \ h_m(L_h - 1)]^T, \quad (4)$$

$$\mathbf{g}_m = [g_m(0) \ g_m(1) \ \dots \ g_m(L_g - 1)]^T, \quad (5)$$

where L_h and L_g denote the RIR length and the reshaping filter length, respectively. Using the ML_g -dimensional stacked filter vector $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \dots \ \mathbf{g}_M^T]^T$, the EIR vector \mathbf{c}

of length $L_c = L_h + L_g - 1$, i.e., $\mathbf{c} = [c(0) \ c(1) \ \dots \ c(L_c - 1)]^T$ can be expressed as

$$\mathbf{c} = \mathbf{H}\mathbf{g}, \quad (6)$$

where \mathbf{H} denotes the $L_c \times ML_g$ -dimensional multi-channel convolution matrix of the RIRs, i.e., $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_M]$, with

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \dots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h - 1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h - 1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_m(L_h - 1) \end{bmatrix}. \quad (7)$$

Using the L_c -dimensional clean speech vector $\mathbf{s}(n) = [s(n)s(n-1) \ \dots \ s(n-L_c+1)]^T$, the output signal in (3) can be expressed as

$$z(n) = \mathbf{c}^T \mathbf{s}(n) = \mathbf{g}^T \mathbf{H}^T \mathbf{s}(n). \quad (8)$$

The reshaping filter \mathbf{g} can then be constructed based on different design objectives for the EIR \mathbf{c} .

3 Acoustic multi-channel equalization

Acoustic multi-channel equalization techniques aim at speech dereverberation by designing the reshaping filter \mathbf{g} such that the (weighted) EIR \mathbf{c} in (6) is equal to a (weighted) target EIR \mathbf{c}_d . For the equalization techniques considered in this paper, i.e., MINT [19], RMCLS [23], and PMINT [24], the definition of the target EIR \mathbf{c}_d is presented in Table 1, where τ denotes a delay, L_d denotes the length of the direct path and early reflections, and $p \in \{1, \dots, M\}$. The delay τ is incorporated in order to relax the causality constraints on the filter design [22]. The length of the direct path and early reflections L_d in the RMCLS and PMINT is typically considered to be between 10–50 ms [23, 24]. It should be realized that in practice, only the perturbed RIRs \hat{h}_m are available, i.e., $\hat{h}_m = h_m + e_m$, where e_m represents the RIR perturbations due to fluctuations (e.g., temperature or position fluctuations [27]) or due to the sensitivity of BSI and

Table 1 Definition of the target EIR \mathbf{c}_d and weighting matrix \mathbf{W} for MINT, RMCLS, and PMINT

Technique	Target EIR \mathbf{c}_d	Weighting matrix \mathbf{W}
MINT	$[0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$ τ	\mathbf{I}
RMCLS	$[0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$ τ	$\text{diag}\{[1 \ \dots \ 1 \ 1 \ 0 \ \dots \ 0 \ 1 \ \dots \ 1]^T\}$ τ L_d
PMINT	$[0 \ \dots \ 0 \ \hat{h}_p(0) \ \dots \ \hat{h}_p(L_d - 1) \ 0 \ \dots \ 0]^T$ τ L_d	\mathbf{I}

SSI methods to near-common zeros or interfering noise [28–30]. Hence, for the filter design, the perturbed convolution matrix $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ is used, where \mathbf{E} represents the convolution matrix of the RIR perturbations. The considered equalization techniques compute the filter \mathbf{g} as the solution of the system of equations

$$\mathbf{W}\hat{\mathbf{H}}\mathbf{g} = \mathbf{W}\mathbf{c}_d, \quad (9)$$

with \mathbf{W} an $L_c \times L_c$ -dimensional diagonal weighting matrix. The definition of the weighting matrix \mathbf{W} for MINT, RMCLS, and PMINT is presented in Table 1, where \mathbf{I} denotes the $L_c \times L_c$ -dimensional identity matrix. Based on these definitions of \mathbf{W} and \mathbf{c}_d , it can be observed that on the one hand, MINT and PMINT do not use a weighting matrix and constrain all taps of the EIR (i.e., $\mathbf{W} = \mathbf{I}$), while on the other hand, RMCLS uses a weighting matrix and does not constrain all taps of the EIR (i.e., $\mathbf{W} = \text{diag}\{\underbrace{[1 \dots 1]^\tau}_{\tau} \underbrace{[1 \ 0 \dots 0]}_{L_d} \ 1 \dots 1\}^T$). It has been

experimentally validated in [23, 24, 26] that by constraining all taps of the EIR, MINT and PMINT may result in a good perceptual speech quality but a high sensitivity to RIR perturbations, whereas by not constraining all taps of the EIR, RMCLS may result in a lower sensitivity to RIR perturbations but a decreased perceptual speech quality.

For all considered equalization techniques, the reshaping filter solving (9) is computed by minimizing the (weighted) least-squares cost function

$$J_{LS} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{c}_d)\|_2^2. \quad (10)$$

As shown in [19, 23, 24], assuming that the RIRs do not share common zeros and using a reshaping filter length

$$L_g \geq \left\lceil \frac{L_h - 1}{M - 1} \right\rceil, \quad (11)$$

with $\lceil \cdot \rceil$ the ceiling operator, the reshaping filter minimizing (10) to 0, and hence solving (9), is given by [19, 23, 24]

$$\mathbf{g}_{LS} = (\mathbf{W}\hat{\mathbf{H}})^+(\mathbf{W}\mathbf{c}_d), \quad (12)$$

where $\{\cdot\}^+$ denotes the matrix pseudo-inverse. When the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the reshaping filter of length L_g according to (11) yields perfect dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{g}_{LS} = \mathbf{W}\mathbf{c}_d$. However, in the presence of RIR perturbations, i.e., $\hat{\mathbf{H}} \neq \mathbf{H}$, this filter typically fails to achieve perfect dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{g}_{LS} \neq \mathbf{W}\mathbf{c}_d$, possibly even causing severe distortions in the output signal [24, 26]. The sensitivity of the reshaping filter to RIR perturbations can be evaluated by analyzing the condition number of the matrix $\mathbf{W}\hat{\mathbf{H}}$.

4 Robust acoustic multi-channel equalization

In this section, the Wedin theorem [37] relating the condition number of the matrix being inverted to the sensi-

tivity of the solution to perturbations is briefly reviewed. In addition, it is analytically shown that using shorter reshaping filters than conventionally used decreases the condition number of the matrix $\mathbf{W}\hat{\mathbf{H}}$, hence increasing the robustness against RIR perturbations.

Wedin theorem [37]: Consider the system of equations $\mathbf{A}\mathbf{q} = \mathbf{b}$, where the matrix \mathbf{A} has dimensions $u \times v$ and rank $r \leq \min\{u, v\}$. Let \mathbf{A} be perturbed to $\mathbf{A} + \Delta\mathbf{A}$. The pseudo-inverse solution $\mathbf{q} = \mathbf{A}^+\mathbf{b}$ is then perturbed to $\mathbf{q} + \Delta\mathbf{q} = (\mathbf{A} + \Delta\mathbf{A})^+\mathbf{b}$, where $\Delta\mathbf{q}$ is the deviation between the true and the perturbed solution. The condition number χ_A of the matrix \mathbf{A} is defined as

$$\chi_A = \frac{\|\mathbf{A}\|_2}{\|\mathbf{A}^+\|_2} = \frac{\sigma_A(1)}{\sigma_A(r)}, \quad (13)$$

with $\sigma_A(i)$ the i -th singular value of the matrix \mathbf{A} , ordered such that $\sigma_A(1) \geq \sigma_A(2) \geq \dots \geq \sigma_A(r) > 0$. Using χ_A and defining the variable ξ as

$$\xi = \frac{\|\Delta\mathbf{A}\|_2}{\|\mathbf{A}\|_2}, \quad (14)$$

the norm of the deviation between the true and the perturbed solution is bounded by

$$\|\Delta\mathbf{q}\|_2 \leq \frac{\chi_A \xi \|\mathbf{q}\|_2}{1 - \chi_A \xi} + \|(\mathbf{A}\mathbf{A}^T)^+\mathbf{b}\|_2 \|\mathbf{A}\|_2, \quad (15)$$

where it is assumed that $\chi_A \xi < 1$.

The relation in (15) shows that a large condition number χ_A can result (and typically does) in a large deviation between the true and the perturbed solution [37, 38].

For clarity of presentation, the notation summarized in Table 2 is used in the following. In order to satisfy (11), reshaping filters in acoustic multi-channel equalization

Table 2 Notation for different reshaping filter lengths and the corresponding matrices

Variable	Denotes
$L_g^t = \left\lceil \frac{L_h - 1}{M - 1} \right\rceil$	Reshaping filter length conventionally used in acoustic multi-channel equalization techniques
$\mathbf{W}_t \hat{\mathbf{H}}_t$	Matrix when using the reshaping filter length L_g^t
$p_t = L_h + L_g^t - 1$	Number of rows in $\mathbf{W}_t \hat{\mathbf{H}}_t$
$q_t = M L_g^t \geq p_t$	Number of columns in $\mathbf{W}_t \hat{\mathbf{H}}_t$
$r_t \leq p_t$	Rank of $\mathbf{W}_t \hat{\mathbf{H}}_t$
$L_g^s < L_g^t$	Reshaping filter length smaller than L_g^t
$\mathbf{W}_s \hat{\mathbf{H}}_s$	Matrix when using the reshaping filter length L_g^s
$p_s = L_h + L_g^s - 1$	Number of rows in $\mathbf{W}_s \hat{\mathbf{H}}_s$
$q_s = M L_g^s < p_s$	Number of columns in $\mathbf{W}_s \hat{\mathbf{H}}_s$
$r_s = q_s$	Rank of $\mathbf{W}_s \hat{\mathbf{H}}_s$

techniques are conventionally designed using the filter length $L_g^t = \left\lceil \frac{L_h - 1}{M - 1} \right\rceil$, i.e., based on the $p_t \times q_t$ -dimensional matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ with $p_t \leq q_t$ and rank $r_t \leq p_t$ (cf. Table 2). However, reshaping filters can also be designed using a shorter filter length $L_g^s < L_g^t$, i.e., based on the $p_s \times q_s$ -dimensional matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$ (cf. Table 2). Considering that $L_g^s < \frac{L_h - 1}{M - 1}$, the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$ is a full column-rank matrix with fewer columns than rows, i.e., $q_s < p_s$, since

$$(M - 1)L_g^s < L_h - 1 \Rightarrow \underbrace{ML_g^s}_{q_s} < \underbrace{L_h + L_g^s - 1}_{p_s}. \quad (16)$$

As schematically illustrated in Fig. 2, the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$ is a sub-matrix of $\mathbf{W}_t \hat{\mathbf{H}}_t$, constructed by deleting $L_g^t - L_g^s$ rows and $M(L_g^t - L_g^s)$ columns from $\mathbf{W}_t \hat{\mathbf{H}}_t$. Aiming at establishing a relation between the condition numbers of the matrices $\mathbf{W}_t \hat{\mathbf{H}}_t$ and $\mathbf{W}_s \hat{\mathbf{H}}_s$, i.e.,

$$\chi_{\mathbf{W}_t \hat{\mathbf{H}}_t} = \frac{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1)}{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t)}, \quad (17)$$

$$\chi_{\mathbf{W}_s \hat{\mathbf{H}}_s} = \frac{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1)}{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s)}, \quad (18)$$

we consider the following interlacing inequalities between the singular values of a matrix and its sub-matrices [39].

Interlacing inequalities [39]: Given a matrix \mathbf{A} of dimensions $u \times v$ and a sub-matrix \mathbf{B} obtained by deleting l rows and/or columns from \mathbf{A} , the singular values of \mathbf{A} and \mathbf{B} interlace as

$$\sigma_{\mathbf{A}}(i) \geq \sigma_{\mathbf{B}}(i) \geq \sigma_{\mathbf{A}}(i + l), \quad (19)$$

for $i = 1, \dots, \min\{u - l, v - l\}$.

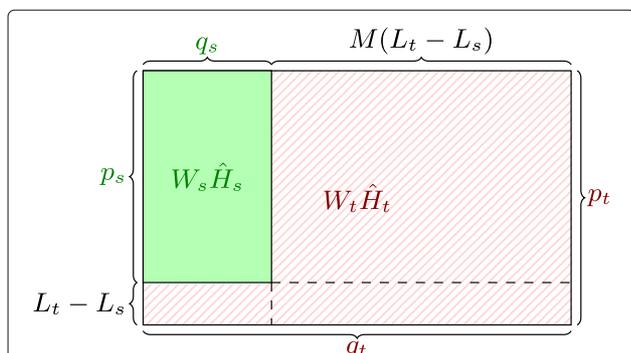


Fig. 2 Schematic illustration of the construction of the $p_s \times q_s$ -dimensional sub-matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$ from the $p_t \times q_t$ -dimensional matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$

Using (19), in Appendix A we derive the following inequalities relating the largest and the smallest non-zero singular values of $\mathbf{W}_t \hat{\mathbf{H}}_t$ and $\mathbf{W}_s \hat{\mathbf{H}}_s$:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (20)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (21)$$

It readily follows from (20) and (21) that the condition number of $\mathbf{W}_s \hat{\mathbf{H}}_s$ is smaller than or equal to the condition number of $\mathbf{W}_t \hat{\mathbf{H}}_t$, i.e.,

$$\chi_{\mathbf{W}_s \hat{\mathbf{H}}_s} = \frac{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1)}{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s)} \leq \frac{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1)}{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t)} = \chi_{\mathbf{W}_t \hat{\mathbf{H}}_t}. \quad (22)$$

Hence, using a shorter reshaping filter than conventionally used in equalization techniques can result (and based on simulations, it always does) in a lower condition number of the matrix being inverted.

Figure 3 depicts the singular values of the matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ for PMINT, constructed using the conventional reshaping filter length $L_g^t = 1947$. The used acoustic system is system S_1 described in Section 6.1, with $M = 4$ microphones and $L_h = 5840$. The singular values of two sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$, constructed using $L_g^s = 1000$ and $L_g^s = 300$, are also depicted. The largest and the smallest non-zero singular values of each matrix are marked in order to illustrate the inequalities presented in (20) and (21). Using these singular values, the condition numbers of the different matrices are presented in Table 3, where it is illustrated that using a shorter reshaping filter length than conventionally used decreases the condition number of the matrix $\mathbf{W}\hat{\mathbf{H}}$.

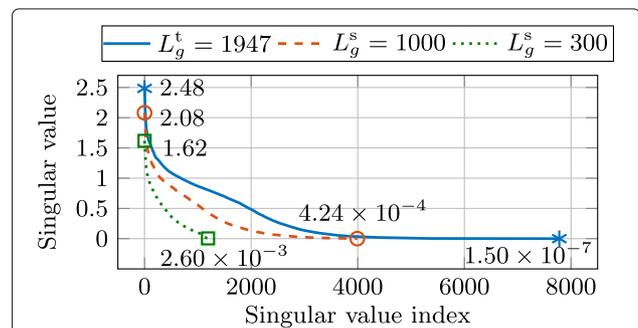


Fig. 3 Singular values of the matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ ($L_g^t = 1947$) and of two sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$ ($L_g^s = 1000$ and $L_g^s = 300$) for PMINT. The largest and the smallest non-zero singular values of each matrix are explicitly marked. The considered acoustic system is system S_1 described in Section 6.1

Table 3 Condition number of the matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ ($L_g^t = 1947$) and of two sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$ ($L_g^s = 1000$ and $L_g^s = 300$) for PMINT

Filter length	Condition number
$L_g^t = 1947$	$\chi_{\mathbf{W}_t \hat{\mathbf{H}}_t} = 1.65 \times 10^7$
$L_g^s = 1000$	$\chi_{\mathbf{W}_s \hat{\mathbf{H}}_s} = 4.91 \times 10^3$
$L_g^s = 300$	$\chi_{\mathbf{W}_s \hat{\mathbf{H}}_s} = 6.23 \times 10^2$

The considered acoustic system is system S_1 described in Section 6.1

In summary, decreasing the reshaping filter length in acoustic multi-channel equalization techniques decreases the condition number of the matrix being inverted, increasing the robustness against RIR perturbations. However, decreasing the reshaping filter length also reduces the equalization performance with respect to the true RIRs, resulting in a trade-off between equalization performance for perfectly estimated RIRs and robustness in the presence of RIR perturbations. Using a shorter reshaping filter is not only desirable to increase the robustness against RIR perturbations, but also because of the lower computational complexity of the reshaping filter design.

5 Automatic non-intrusive reshaping filter length

The optimal reshaping filter length L_g^{opt} yielding the highest dereverberation performance obviously depends on the acoustic system and the RIR perturbation level. In simulations, L_g^{opt} can be *intrusively* determined by exploiting a clean reference signal. Reshaping filters for several reshaping filter lengths can be computed and applied to the received microphone signals such that different output signals are generated. The optimal reshaping filter length L_g^{opt} can then be selected by comparing the different output signals to the clean reference signal. Since one typically does not have access to the clean reference signal, an automatic *non-intrusive* procedure is required in practice.

Motivated by the simplicity and the applicability of the L-curve to automatically determine a regularization parameter in regularized (weighted) least-squares solutions [24, 34, 35], in this section, we propose to use the L-curve to automatically determine the reshaping filter length L_g^{auto} in acoustic multi-channel equalization techniques.

Using a shorter reshaping filter introduces a trade-off between the condition number $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and the (weighted) least-squares error $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$. An appropriate filter length should incorporate knowledge about $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$, such that preferably both quantities are kept as small as possible. Due to the arising trade-off between these quantities, the parametric plot of the condition number versus the (weighted)

least-squares error for several reshaping filter lengths has an L-shape. The corner of the L-curve, i.e., the point of maximum curvature, is located where the filter changes from being dominated by a large condition number to being dominated by a large (weighted) least-squares error. Hence, we propose to automatically select the reshaping filter length L_g^{auto} as the filter length corresponding to the corner of the parametric plot of the condition number $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ versus the (weighted) least-squares error $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$.

Figure 4 depicts a typical L-curve obtained using PMINT for acoustic system S_1 described in Section 6.1. As illustrated in this figure, decreasing the reshaping filter length decreases $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ but at the same time increases $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$. Although from such a curve it seems straightforward to determine the reshaping filter length corresponding to the corner of the L-curve (i.e., $L_g^s = 1000$ in this example), numerical problems may occur, and hence, a numerically stable algorithm is required. Similarly as in [24], in this paper we use the triangle method [36] for locating the corner of the L-curve.

6 Simulation results and discussion

In this section, we investigate the influence of the reshaping filter length on the dereverberation performance of all considered acoustic multi-channel equalization techniques. In Section 6.1, the considered acoustic systems, instrumental performance measures, and algorithmic settings are introduced. In Section 6.2, the increase in robustness when using shorter reshaping filter lengths is investigated. In Section 6.3, the performance of all considered equalization techniques using the intrusively determined reshaping filter length L_g^{opt} is compared for several acoustic systems and RIR perturbation levels. In Section 6.4, the performance of PMINT using the automatic non-intrusively determined reshaping filter length L_g^{auto} is investigated.

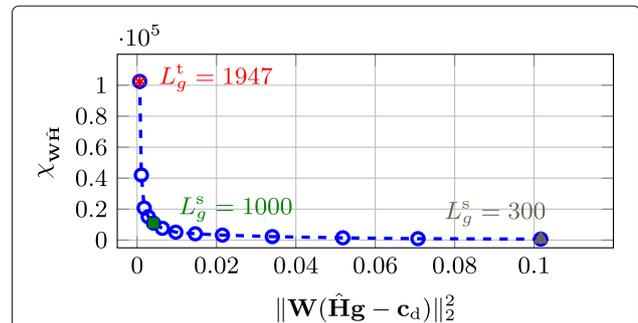


Fig. 4 Typical L-curve obtained using PMINT for several reshaping filter lengths ranging from $L_g^t = 1947$ to $L_g^s = 300$. The considered acoustic system is system S_1 described in Section 6.1

6.1 Acoustic systems, instrumental performance measures, and algorithmic settings

Acoustic systems. We consider three different reverberant acoustic systems with a single speech source placed at a distance of 2 m from $M = 4$ omni-directional microphones. The RIRs between the speech source and the microphones are measured using the swept-sine technique [40] and the reverberant signals are generated by convolving 10 sentences (approximately 17 s long) from the HINT database [41] with measured RIRs. For each acoustic system, Table 4 presents the reverberation time T_{60} of the room, the length of the RIRs L_h at a sampling frequency $f_s = 8$ kHz, and the input direct-to-reverberant-ratio (iDRR). The iDRR is computed using the RIR of the first microphone $h_1(n)$ and is defined as

$$\text{iDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} h_1^2(n)}{\sum_{n=n_d}^{L_h-1} h_1^2(n)}, \quad (23)$$

where the first n_d samples (corresponding to 3 ms) of $h_1(n)$ represent the direct path propagation and the remaining samples represent reflections [9]. In order to simulate RIR perturbations, the measured RIRs are perturbed by proportional Gaussian distributed errors as proposed in [42], such that a desired normalized projection misalignment (NPM), i.e.,

$$\text{NPM} = 20 \log_{10} \frac{\left\| \mathbf{h} - \frac{\mathbf{h}^T \hat{\mathbf{h}}}{\hat{\mathbf{h}}^T \hat{\mathbf{h}}} \hat{\mathbf{h}} \right\|_2}{\|\mathbf{h}\|_2}, \quad (24)$$

is generated. Introducing proportional Gaussian distributed errors is a widely used technique to systematically simulate RIR perturbations arising from system identification methods. The considered NPMs for each acoustic system are

$$\text{NPM} \in \{-33 \text{ dB}, -30 \text{ dB}, \dots, -15 \text{ dB}\}, \quad (25)$$

with -33 dB a moderate perturbation level and -15 dB a larger perturbation level. It should be noted that the

NPMs in (25) represent realistic NPMs achieved by state-of-the-art BSI methods (for relatively short RIRs in the order of 300 – 500 taps) [30].

Instrumental performance measures. The performance of the equalization techniques is evaluated in terms of the reverberant energy suppression and perceptual speech quality improvement. The reverberant energy suppression is evaluated using the improvement in direct-to-reverberant ratio (ΔDRR), i.e., $\Delta\text{DRR} = \text{oDRR} - \text{iDRR}$, with

$$\text{oDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} c^2(n)}{\sum_{n=n_d}^{L_c-1} c^2(n)}, \quad (26)$$

and iDRR defined in (23). Although ΔDRR exactly describes the reverberant energy suppression, it cannot be solely used to evaluate the dereverberation performance of equalization techniques, since it does not provide any insight on the reverberant energy decay rate. To evaluate the reverberant energy decay rate, the energy decay curve (EDC) [9] of the EIR $c(n)$ is compared to the energy decay curve of the true first RIR $h_1(n)$. The EDC of the EIR \mathbf{c} is computed as

$$\text{EDC}_{\mathbf{c}}(n) = \frac{1}{\|\mathbf{c}\|_2^2} \sum_{j=n}^{L_c-1} c^2(j), \quad n = 0, 1, \dots, L_c - 1, \quad (27)$$

and the EDC of the RIR $h_1(n)$ is computed similarly. The perceptual speech quality is evaluated using the frequency-weighted segmental signal-to-noise-ratio (fSNR) and the cepstral distance (CD) [43]. In [44], it has been shown that measures such as fSNR and CD can exhibit a high correlation with subjective listening tests when evaluating the overall quality and the perceived amount of reverberation for a wide range of state-of-the-art dereverberation (and noise reduction) techniques. These signal-based measures are intrusive measures, generating a similarity score between a test signal and a reference signal. The reference signal employed here is obtained by convolving the clean speech signal with the direct path and early reflections (considered to be 10 ms long) of $h_1(n)$. The improvement in fSNR, i.e., ΔfSNR , is computed as the difference between the fSNR of the output signal $z(n)$ and the fSNR of the first microphone signal $x_1(n)$. Similarly, the improvement in CD, i.e., ΔCD , is computed as the difference between the CD of the output signal $z(n)$ and the CD of the first microphone signal $x_1(n)$. Note that a positive ΔfSNR and a negative ΔCD indicate a performance improvement.

Table 4 Characteristics of the considered measured acoustic systems

Acoustic system	T_{60} [ms]	L_h	iDRR [dB]
S_1	730	5840	-2.79
S_2	610	4880	-0.87
S_3	360	2880	1.43

Algorithmic settings. For the acoustic systems described in Table 4 and for all considered equalization techniques, the conventionally used filter length is $L_g^t = \left\lceil \frac{L_R - 1}{M - 1} \right\rceil$, i.e., $L_g^t = 1947$ for system S_1 , $L_g^t = 1627$ for system S_2 , and $L_g^t = 960$ for system S_3 . The delay is set to $\tau = 90$ and the length of the direct path and early reflections is set to $L_d = 0.01 \times f_s$, corresponding to 10 ms (cf. Table 1). The target EIR c_d for PMINT is constructed using the first RIR, i.e., $p = 1$ (cf. Table 1).

We consider several shorter reshaping filter lengths for all equalization techniques, i.e.,

$$L_g^s \in \begin{cases} \{300, 330, \dots, 1920\} & \text{for system } S_1 \\ \{300, 330, \dots, 1620\} & \text{for system } S_2 \\ \{300, 330, \dots, 930\} & \text{for system } S_3 \end{cases} \quad (28)$$

For each acoustic system, each NPM, and each equalization technique, the optimal filter length L_g^{opt} is selected from (28) as the filter length yielding the lowest CD. It should be noted that using the CD for determining the optimal reshaping filter length is an intrusive procedure which cannot be applied in practice, since knowledge of the clean reference signal is required. In Section 6.4, the performance when using the automatic non-intrusive procedure for selecting the reshaping filter length is investigated.

6.2 Increasing robustness using shorter reshaping filters

In this section, the robustness of MINT, RMCLS, and PMINT against RIR perturbations is investigated when using the conventional reshaping filter length L_g^t and the optimal (shorter) reshaping filter length L_g^{opt} . Although similar results are obtained for all considered acoustic systems, in this section, only results for acoustic system S_1 are presented. For completeness, the intrusively determined optimal reshaping filter length L_g^{opt} for each considered technique and NPM is presented in Table 5.

MINT using L_g^t and L_g^{opt} . Figure 5 depicts the performance of MINT when using the filter lengths L_g^t

and L_g^{opt} in terms of ΔDRR , EDC, ΔfSNR , and ΔCD . As expected, the ΔDRR values presented in Fig. 5a show that using the conventional filter length L_g^t fails to suppress the reverberant energy, even significantly worsening the DRR by about 20 dB on average in comparison to \mathbf{h}_1 . Furthermore, it can be observed that using the shorter filter length L_g^{opt} (cf. Table 5) significantly increases the robustness of MINT for all NPMs, improving the DRR by about 4 dB on average in comparison to \mathbf{h}_1 . These results are confirmed in Fig. 5b, which depicts the EDC of \mathbf{h}_1 and the EDCs of the EIRs obtained using MINT with L_g^t and L_g^{opt} for an NPM of -33 dB. It can be observed that while using L_g^t completely fails to achieve dereverberation and results in a slower reverberant energy decay rate than \mathbf{h}_1 , using L_g^{opt} yields a significantly faster reverberant energy decay rate. However, using L_g^{opt} yields only a slight improvement of the reverberant energy decay rate when compared to \mathbf{h}_1 , even for the moderate NPM of -33 dB. The ΔfSNR and ΔCD values depicted in Fig. 5c, d show that using the conventional filter length L_g^t in MINT yields a significantly worse quality than the unprocessed microphone signal $x_1(n)$ for all NPMs. While an increase in robustness is obtained for all NPMs using L_g^{opt} , for most considered NPMs, the performance in terms of ΔfSNR is still worse than for the unprocessed microphone signal $x_1(n)$.

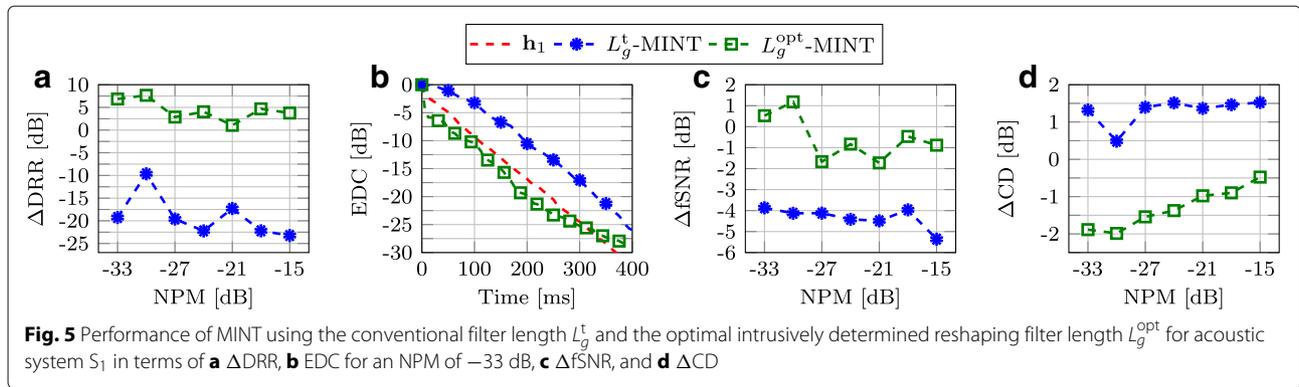
In summary, as expected from the theoretical analysis in Section 4, these simulation results demonstrate that using an optimal intrusively determined shorter reshaping filter length than conventionally used in MINT is advantageous to increase the robustness against RIR perturbations. However, since acoustic system inversion using MINT is very sensitive to RIR perturbations, these results indicate that even a shorter reshaping filter is not sufficient to make MINT robust enough against RIR perturbations.

RMCLS using L_g^t and L_g^{opt} . Figure 6 depicts the performance of RMCLS using the filter lengths L_g^t and L_g^{opt} in terms of ΔDRR , EDC, ΔfSNR , and ΔCD . The ΔDRR values presented in Fig. 6a show that using the conventional filter length L_g^t improves the DRR for moderate NPMs, whereas for NPMs larger than -21 dB, the DRR is worsened in comparison to \mathbf{h}_1 . In addition, it can be observed that using the shorter filter length L_g^{opt} (cf. Table 5) significantly increases the reverberant energy suppression for all NPMs, on average yielding a 6 dB larger ΔDRR in comparison to the ΔDRR obtained using L_g^t . To evaluate the reverberant energy decay rate, Fig. 6b depicts the EDC of \mathbf{h}_1 and the EDCs of the EIRs obtained using RMCLS with L_g^t and L_g^{opt} for an NPM of -33 dB. It can be observed that for this moderate NPM, a very similar reverberant energy decay rate is obtained for RMCLS

Table 5 Optimal intrusively determined reshaping filter length L_g^{opt} for MINT, RMCLS, and PMINT for acoustic system S_1 and all considered NPMs

NPM [dB]	MINT	RMCLS	PMINT
-33	1140	1200	1170
-30	1200	1200	1230
-27	930	1140	1200
-24	1050	1020	1050
-21	870	840	900
-18	780	780	900
-15	510	660	510

The conventionally used reshaping filter length is $L_g^t = 1947$



when using L_g^t and L_g^{opt} . Since RMCLS using the conventional filter length L_g^t is relatively robust for moderate NPMs and yields a fast reverberant energy decay rate, a shorter reshaping filter does not yield any improvement in the reverberant energy decay rate, but instead leads to a significant improvement in the perceptual speech quality. This is illustrated in Fig. 6c, d, which shows that using L_g^{opt} in RMCLS significantly improves the Δ fSNR and Δ CD values for all NPMs.

In summary, as expected from the theoretical analysis in Section 4, these simulation results demonstrate that using an optimal intrusively determined shorter reshaping filter length than conventionally used in RMCLS is advantageous and increases the robustness against RIR perturbations.

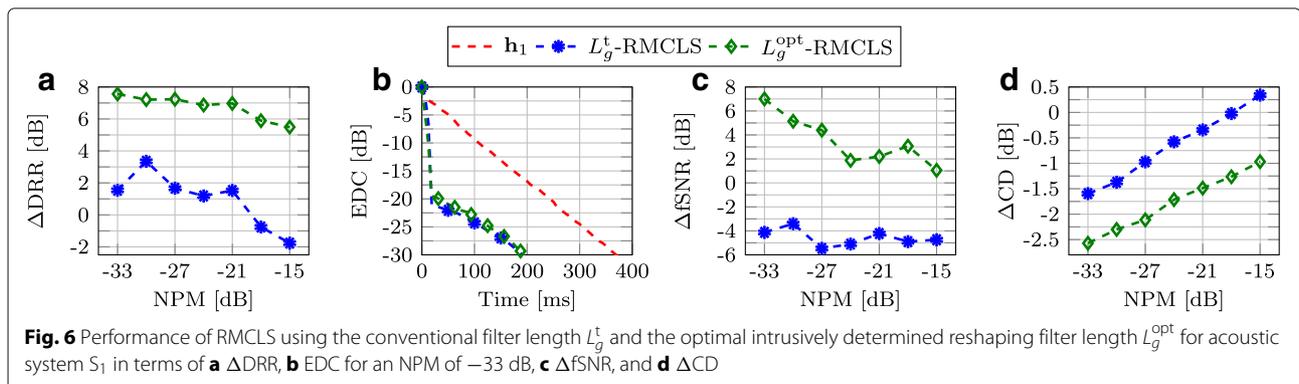
PMINT using L_g^t and L_g^{opt} . Figure 7 depicts the performance of PMINT using the filter lengths L_g^t and L_g^{opt} in terms of Δ DRR, EDC, Δ fSNR, and Δ CD. As expected, the Δ DRR values presented in Fig. 7a show that using the conventional filter length L_g^t in PMINT fails to suppress the reverberant energy, even worsening the DRR in comparison to h_1 . Furthermore, it can be observed that using L_g^{opt} (cf. Table 5) significantly increases the robustness for all NPMs, on average improving the DRR by about 7 dB in comparison to h_1 . These results are further confirmed in Fig. 7b, which depicts the EDC of h_1 and the

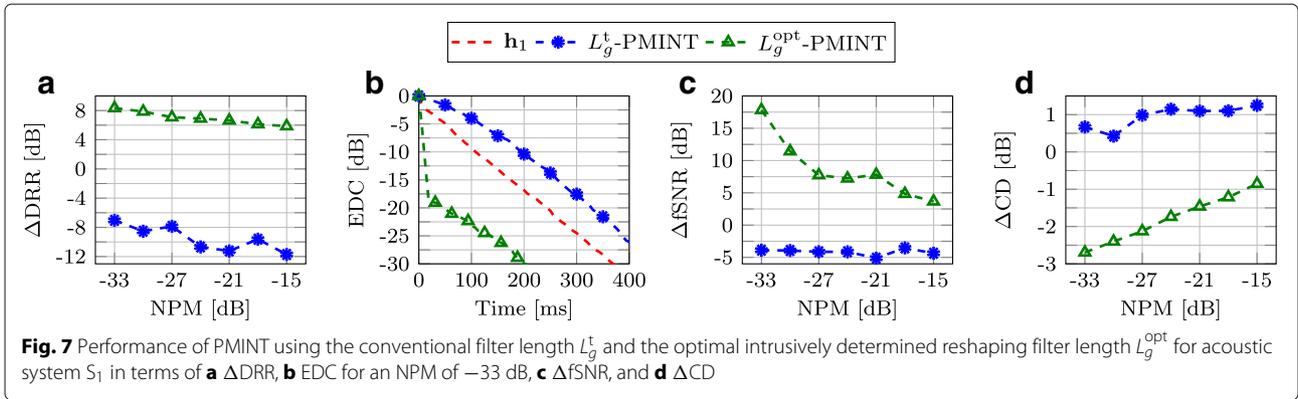
EDCs of the EIRs obtained using PMINT with L_g^t and L_g^{opt} for an NPM of -33 dB. It can be observed that PMINT using L_g^t completely fails to achieve dereverberation and results in a slower reverberant energy decay rate than h_1 . Using the optimal reshaping filter length L_g^{opt} yields a significant increase in robustness, resulting in a much faster reverberant energy decay rate than h_1 . Furthermore, the Δ fSNR and Δ CD values depicted in Fig. 7c, d show that while PMINT using L_g^t worsens the perceptual speech quality in comparison to the unprocessed microphone signal $x_1(n)$, using L_g^{opt} results in a significantly better performance.

In summary, as expected from the theoretical analysis in Section 4, these simulation results demonstrate that using an optimal intrusively determined shorter reshaping filter length than conventionally used in PMINT results in a significant increase in robustness against RIR perturbations, both in terms of reverberant energy suppression and perceptual speech quality improvement.

6.3 Performance of equalization techniques when using the optimal intrusive reshaping filter length

In the previous section, it was shown that using a shorter reshaping filter than conventionally used increases the robustness of all considered equalization techniques against RIR perturbations. In this section, the





performance of MINT, RMCLS, and PMINT using the optimal intrusively determined reshaping filter length L_g^{opt} is extensively compared for all acoustic systems in Table 4 and all NPMs in (25). The performance of the different techniques is evaluated in terms of ΔDRR , ΔfSNR , and ΔCD , where the presented performance measures are averaged over all considered NPMs.

Table 6 presents the obtained ΔDRR , ΔfSNR , and ΔCD values for all considered techniques¹. First, it can be observed that MINT using L_g^{opt} results in the lowest performance in terms of all performance measures, often worsening the perceptual speech quality in comparison to the unprocessed microphone signal $x_1(n)$. Since MINT is very sensitive to RIR perturbations (cf. Fig. 5), the robustness increase that can be obtained by using a shorter reshaping filter length is also limited. Second, it can be observed that RMCLS and PMINT using L_g^{opt} result in a high reverberant energy suppression in terms of ΔDRR , with PMINT outperforming RMCLS for systems S_2 and S_3 whereas a similar performance is obtained for system S_1 . Finally, it can be observed that for all considered acoustic systems, PMINT using the reshaping filter length L_g^{opt} yields the highest perceptual speech quality improvement, outperforming RMCLS in terms of ΔfSNR and ΔCD . While PMINT always improves the perceptual speech quality in comparison to the unprocessed microphone signal $x_1(n)$, RMCLS sometimes fails to yield an improvement, as indicated by the negative ΔfSNR for systems S_2 and S_3 . The advantage of PMINT lies in its control of the early reflections in the EIR, hence better preserving the perceptual speech quality of the output signal.

In summary, based on instrumental measures, it can be said that PMINT using the optimal intrusively determined reshaping filter length L_g^{opt} is a robust and perceptually advantageous equalization technique, yielding a high reverberant energy suppression and outperforming all other considered equalization techniques in terms of perceptual speech quality. Informal listening tests further support this conclusion.

6.4 Performance of PMINT when using the automatic non-intrusive reshaping filter length

In this section, we investigate the performance of PMINT when using the automatic non-intrusively determined reshaping filter length L_g^{auto} (cf. Section 5) instead of the optimal intrusively determined reshaping filter length L_g^{opt} . For completeness, the obtained values of L_g^{auto} are also compared to the values of L_g^{opt} . Similarly as in Section 6.3, we consider all acoustic systems in Table 4 and all NPMs in (25). In order to generate the parametric L-curve, the matrix $\mathbf{W}\hat{\mathbf{H}}$ is constructed for all reshaping filter lengths in (28), the PMINT reshaping filter is computed, and the quantities $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$ are calculated. Using the triangle method [36], the automatic reshaping filter length L_g^{auto} corresponding to the point of maximum curvature of the L-curve is determined. The performance of PMINT using L_g^{auto} is evaluated in terms of ΔDRR , ΔfSNR , and ΔCD , where the presented performance measures are averaged over all considered NPMs.

Table 7 presents the values of L_g^{opt} and L_g^{auto} for the acoustic system S_1 and all considered NPMs². It can be

Table 6 Average performance of MINT, RMCLS, and PMINT using the optimal intrusively determined reshaping filter length L_g^{opt}

	S_1			S_2			S_3		
	ΔDRR [dB]	ΔfSNR [dB]	ΔCD [dB]	ΔDRR [dB]	ΔfSNR [dB]	ΔCD [dB]	ΔDRR [dB]	ΔfSNR [dB]	ΔCD [dB]
L_g^{opt} -MINT	4.41	-0.55	-1.31	1.66	-2.07	0.07	1.20	-3.89	-0.22
L_g^{opt} -RMCLS	6.75	3.53	-1.77	1.76	-0.81	-0.39	1.31	-0.61	-0.52
L_g^{opt} -PMINT	6.98	8.65	-1.78	4.42	2.58	-0.66	2.40	1.88	-0.57

Table 7 Intrusively and non-intrusively determined reshaping filter lengths for PMINT for acoustic system S_1 and all considered NPMs

NPM [dB]	-33	-30	-27	-24	-21	-18	-15
L_g^{opt}	1170	1230	1200	1050	900	900	510
L_g^{auto}	1200	1170	1170	1230	1230	1170	1170

observed that for low NPMs, the non-intrusively determined reshaping filter length is very similar to the optimal intrusively determined one. As the NPM increases beyond -21 dB, the reshaping filter length obtained using the proposed non-intrusive procedure is larger than the optimal intrusively determined one.

Table 8 presents the ΔDRR , $\Delta fSNR$, and ΔCD obtained using PMINT with L_g^{auto} for all considered acoustic systems. It can be observed that using the automatic non-intrusively determined reshaping filter length in PMINT yields a high dereverberation performance, both in terms of reverberant energy suppression and perceptual speech quality improvement. In addition, when comparing the performance measures presented in Table 8 to the ones presented in Table 6, it can be observed that in general, the performance of PMINT when using L_g^{auto} is similar to the performance when using L_g^{opt} . More precisely, the average performance degradation over all considered acoustic systems when using the automatic non-intrusively determined reshaping filter length L_g^{auto} instead of the optimal intrusively determined reshaping filter length L_g^{opt} is only 0.58 dB in terms of ΔDRR , 1.15 dB in terms of $\Delta fSNR$, and 0.24 dB in terms of ΔCD .

In summary, the presented results show that the automatic non-intrusively determined reshaping filter length in PMINT yields a high performance in the presence of RIR perturbations, making PMINT when using this shorter reshaping filter length a robust and perceptually advantageous acoustic multi-channel equalization technique.

7 Conclusions

In this paper, we have analyzed the use of a shorter reshaping filter length than conventionally used in order to increase the robustness of acoustic multi-channel equalization techniques. We have analytically shown that using a shorter reshaping filter decreases the condition number of the (weighted) convolution matrix, increasing as a result the robustness against RIR perturbations. In addition, we have proposed to automatically determine the reshaping filter length as the point of maximum curvature

of the parametric plot of the condition number versus the (weighted) least-squares error, such that both quantities are simultaneously kept small. Using instrumental performance measures, it has been shown that using shorter reshaping filters indeed increases the robustness of MINT, RMCLS, and PMINT against RIR perturbations. In addition, it has been shown that PMINT using the optimal intrusively determined reshaping filter length outperforms MINT and RMCLS. Finally, it has been shown that the automatic non-intrusive procedure for selecting the reshaping filter length in PMINT yields a nearly optimal performance, confirming the practical applicability of using shorter reshaping filters in acoustic multi-channel equalization.

Endnotes

¹ It should be noted that the performance measures presented for system S_1 are an average of the results already presented in Section 6.2.

² It should be noted that presented L_g^{opt} values are the same as the ones presented in Table 5.

Appendix A

In order to construct the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$ from the matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ (cf. Fig. 2), we first create an intermediate $[p_t - (L_g^t - L_g^s)] \times [q_t - (L_g^t - L_g^s)]$ -dimensional sub-matrix \mathbf{T} by deleting $L_g^t - L_g^s$ rows and $L_g^t - L_g^s$ columns from $\mathbf{W}_t \hat{\mathbf{H}}_t$. The interlacing inequalities in (19) for the matrices $\mathbf{W}_t \hat{\mathbf{H}}_t$ and \mathbf{T} can then be written as

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(i) \geq \sigma_{\mathbf{T}}(i) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(i + (L_g^t - L_g^s)), \quad (29)$$

for $i = 1, \dots, r_t - (L_g^t - L_g^s), \dots, p_t - (L_g^t - L_g^s)$. Using $i = 1$ and $i = r_t - (L_g^t - L_g^s)$ in (29), the following inequalities between the singular values of the matrices $\mathbf{W}_t \hat{\mathbf{H}}_t$ and \mathbf{T} can be established:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{T}}(1), \quad (30)$$

$$\sigma_{\mathbf{T}}(r_t - (L_g^t - L_g^s)) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (31)$$

Table 8 Average performance of PMINT using the automatically non-intrusively determined reshaping filter length L_g^{auto}

	S_1			S_2			S_3		
	ΔDRR [dB]	$\Delta fSNR$ [dB]	ΔCD [dB]	ΔDRR [dB]	$\Delta fSNR$ [dB]	ΔCD [dB]	ΔDRR [dB]	$\Delta fSNR$ [dB]	ΔCD [dB]
L_w^{auto} -PMINT	6.68	7.90	-1.68	3.41	1.62	-0.33	1.97	0.14	-0.29

In order to construct the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$, $(M-1)(L_g^t - L_g^s)$ columns are now deleted from the matrix \mathbf{T} . The interlacing inequalities in (19) for the matrices \mathbf{T} and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can then be written as

$$\sigma_{\mathbf{T}}(i) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(i) \geq \sigma_{\mathbf{T}}(i + (M-1)(L_g^t - L_g^s)), \quad (32)$$

for $i = 1, \dots, r_s$. Using $i = 1$ and $i = r_s$ in (32), the following inequalities between the singular values of the matrices \mathbf{T} and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can be written:

$$\sigma_{\mathbf{T}}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (33)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{T}}(r_s + (M-1)(L_g^t - L_g^s)). \quad (34)$$

Since the number of columns in $\mathbf{W}_t \hat{\mathbf{H}}_t$ is greater than or equal to its rank, i.e.,

$$q_t = ML_g^t \geq p_t \geq r_t, \quad (35)$$

the index of the singular value in the right hand side of (34) can be written as

$$\begin{aligned} r_s + (M-1)(L_g^t - L_g^s) &= ML_g^s + (M-1)(L_g^t - L_g^s) \\ &= ML_g^t - (L_g^t - L_g^s) \geq r_t - (L_g^t - L_g^s). \end{aligned} \quad (36)$$

Based on (36) and the fact that the singular values of the matrices are sorted in descending order, it can be said that

$$\sigma_{\mathbf{T}}(r_s + (M-1)(L_g^t - L_g^s)) \geq \sigma_{\mathbf{T}}(r_t - (L_g^t - L_g^s)), \quad (37)$$

such that the inequality in (34) can also be written as

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{T}}(r_t - (L_g^t - L_g^s)). \quad (38)$$

Finally, combining (30), (31), (33), and (38), the following inequalities relating the largest and the smallest non-zero singular values of $\mathbf{W}_t \hat{\mathbf{H}}_t$ and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can be established:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (39)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (40)$$

Funding

This work was supported in part by the Cluster of Excellence 1077 "Hearing4All," funded by the German Research Foundation (DFG) and the joint Lower Saxony-Israeli Project ATHENA, funded by the State of Lower Saxony.

Authors' contributions

The contribution of the first author consists in developing the main algorithmic idea, deriving the mathematical analysis, performing simulations, analyzing the simulation results, and drafting the article. The contribution of the second author consists in critically discussing the mathematical analysis and the simulation results and in proofreading and revising the article. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 11 July 2017 Accepted: 30 January 2018

Published online: 09 February 2018

References

1. JS Bradley, H Sato, M Picard, On the importance of early reflections for speech in rooms. *J. Acoust. Soc. Am.* **113**(6), 3233–3244 (2003)
2. I Arweiler, JM Buchholz, The influence of spectral characteristics of early reflections on speech intelligibility. *J. Acoust. Soc. Am.* **130**(2), 996–1005 (2011)
3. A Warzybok, J Rennies, T Brand, S Doclo, B Kollmeier, Effects of spatial and temporal integration of a single early reflection on speech intelligibility. *J. Acoust. Soc. Am.* **133**(1), 269–282 (2013)
4. R Beutelmann, T Brand, Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* **120**(1), 331–342 (2006)
5. S Goetze, E Albertin, J Rennies, EAP Habets, K-D Kammeyer, in *Proc. AES International Conference on Sound Quality Evaluation*. Speech quality assessment for listening-room compensation, (Pitea, 2010), pp. 11–20
6. A Warzybok, I Kodrasi, JO Jungmann, EAP Habets, T Gerkmann, A Mertins, S Doclo, B Kollmeier, S Goetze, in *Proc. International Workshop on Acoustic Echo and Noise Control*. Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithms, (Antibes, 2014), pp. 333–337
7. T Yoshioka, A Sehr, M Delcroix, K Kinoshita, R Maas, T Nakatani, W Kellermann, Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition. *IEEE Signal Process. Mag.* **29**(6), 114–126 (2012)
8. F Xiong, BT Meyer, N Moritz, R Rehr, J Anemüller, T Gerkmann, S Doclo, S Goetze, Front-end technologies for robust ASR in reverberant environments—spectral enhancement-based dereverberation and auditory modulation filterbank features. *EURASIP J. Adv. Signal Process.* **2015**(1), 1–18 (2015)
9. PA Naylor, ND Gaubitch (eds.), *Speech Dereverberation* (Springer, London, 2010)
10. K Lebart, JM Boucher, A new method based on spectral subtraction for speech dereverberation. *Acta. Acoustica.* **87**(3), 359–366 (2001)
11. EAP Habets, S Gannot, I Cohen, Late reverberant spectral variance estimation based on a statistical model. *IEEE Sig. Process Lett.* **16**(9), 770–774 (2009)
12. A Kuklasinski, S Doclo, SH Jensen, J Jensen, Maximum likelihood PSD estimation for speech enhancement in reverberation and noise. *IEEE/ACM Trans. Audio Speech Lang. Process.* **24**(9), 1595–1608 (2016)
13. I Kodrasi, S Doclo, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Late reverberant power spectral density estimation based on an eigenvalue decomposition, (New Orleans, 2017), pp. 611–615
14. I Kodrasi, S Doclo, in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Multi-channel late reverberation power spectral density estimation based on nuclear norm minimization, (New York, 2017). (accepted for publication)
15. T Nakatani, T Yoshioka, K Kinoshita, M Miyoshi, B-H Juang, Speech dereverberation based on variance-normalized delayed linear prediction. *IEEE Trans. Audio Speech Lang. Process.* **18**(7), 1717–1731 (2010)
16. D Schmid, G Enzner, S Malik, D Kolossa, R Martin, Variational Bayesian inference for multichannel dereverberation and noise reduction. *IEEE/ACM Trans. Audio Speech Lang. Process.* **22**(8), 1320–1335 (2014)
17. B Schwartz, S Gannot, EAP Habets, Online speech dereverberation using Kalman filter and EM algorithm. *IEEE/ACM Trans. Audio Speech Lang. Process.* **23**(2), 394–406 (2015)
18. A Jukić, T Van Waterschoot, T Gerkmann, S Doclo, Multi-channel linear prediction-based speech dereverberation with sparse priors. *IEEE/ACM Trans. Audio Speech Lang. Process.* **23**(9), 1509–1520 (2015)
19. M Miyoshi, Y Kaneda, Inverse filtering of room acoustics. *IEEE Trans. Acoust. Speech Signal Process.* **36**(2), 145–152 (1988)
20. M Kallinger, A Mertins, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Multi-channel room impulse response shaping - a study, (Toulouse, 2006), pp. 101–104
21. JO Jungmann, R Mazur, M Kallinger, M Tiemin, A Mertins, Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping. *IEEE Trans. Audio Speech Lang. Process.* **20**(6), 1829–1842 (2012)
22. T Hikichi, M Delcroix, M Miyoshi, Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations. *EURASIP J. Adv. Signal Process.* **2007** (2007)
23. F Lim, W Zhang, EAP Habets, PA Naylor, Robust multichannel dereverberation using relaxed multichannel least squares. *IEEE/ACM Trans. Audio Speech Lang. Process.* **22**(9), 1379–1390 (2014)

24. I Kodrasi, S Goetze, S Doclo, Regularization for partial multichannel equalization for speech dereverberation. *IEEE Trans. Audio Speech Lang. Process.* **21**(9), 1879–1890 (2013)
25. RS Rashobh, AWH Khong, D Liu, Multichannel equalization in the KLT and frequency domains with application to speech dereverberation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **22**(3), 634–646 (2014)
26. I Kodrasi, S Doclo, Signal-dependent penalty functions for robust acoustic multi-channel equalization. *IEEE Trans. Audio Speech Lang. Process.* **25**(7), 1512–1525 (2017)
27. BD Radlovic, RC Williamson, RA Kennedy, Equalization in an acoustic reverberant environment: robustness results. *IEEE Trans. Speech Audio Process.* **8**(3), 311–319 (2000)
28. AWH Khong, L Xiang, PA Naylor, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization, (Las Vegas, 2008), pp. 229–232
29. MA Haque, T Hasan, Noise robust multichannel frequency-domain LMS algorithms for blind channel identification. *IEEE Signal Process. Lett.* **15**, 305–308 (2008)
30. M Hu, ND Gaubitch, PA Naylor, DB Ward, in *Proc. European Signal Processing Conference*. Noise robust blind system identification algorithms based on a Rayleigh quotient cost function, (Nice, 2015)
31. ND Gaubitch, PA Naylor, Equalization of multichannel acoustic systems in oversampled subbands. *IEEE Trans. Audio Speech Lang. Process.* **17**(6), 1061–1070 (2009)
32. F Lim, PA Naylor, in *Proc. European Signal Processing Conference*. Robust speech dereverberation using subband multichannel least squares with variable relaxation, (Marrakech, 2013)
33. I Kodrasi, S Doclo, in *Proc. European Signal Processing Conference*. The effect of inverse filter length on the robustness of acoustic multichannel equalization, (Bucharest, 2012)
34. PC Hansen, Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* **34**(4), 561–580 (1992)
35. PC Hansen, DP O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* **14**(6), 1487–1503 (1993)
36. JL Castellanos, S Gómez, V Guerra, The triangle method for finding the corner of the L-curve. *Appl. Numer. Math.* **43**(4), 359–373 (2002)
37. P Wedin, Perturbation theory for pseudo-inverses. *BIT Numer. Math.* **13**(2), 217–232 (1973)
38. G Golub, C Van Loan, *Matrix Computations*. (The John Hopkins University Press, Baltimore, 1996)
39. RA Horn, CR Johnson, *Topics in matrix analysis*. (Cambridge University Press, Cambridge, 1999)
40. A Farina, in *Proc. AES Convention*. Simultaneous measurement of impulse response and distortion with a swept-sine technique, (Pitea, 2000), pp. 18–22
41. M Nilsson, SD Soli, A Sullivan, Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* **95**(2), 1085–1099 (1994)
42. W Zhang, PA Naylor, An algorithm to generate representations of system identification errors. *Res. Lett. Signal Process.* **2008** (2008)
43. Y Hu, PC Loizou, Evaluation of objective quality measures for speech enhancement. *IEEE Trans. Audio Speech Lang. Process.* **16**(1), 229–238 (2008)
44. K Kinoshita, M Delcroix, S Gannot, EAP Habets, R Haeb-Umbach, W Kellermann, V Leutnant, R Maas, T Nakatani, B Raj, A Sehr, T Yoshioka, A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. *EURASIP J. Adv. Signal Process.* **2016**(1), 1–19 (2016)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com