

RESEARCH

Improving the Conditioning of the Optimization Criterion in Acoustic Multi-Channel Equalization

Ina Kodrası* and Simon Doclo

Abstract

In least-squares acoustic multi-channel equalization techniques, i.e., 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 is achieved for perfectly estimated room impulse responses (RIRs). Since in practice however the available RIRs typically differ from the true ones, this choice of the reshaping filter length may not be optimal. This paper proposes to increase the robustness of equalization techniques against RIR perturbations by using a shorter reshaping filter length than conventionally used, such that a better conditioned optimization criterion is obtained. Based on the condition number of the convolution matrix of the RIRs, a mathematical link 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 a novel automatic non-intrusive procedure for determining the reshaping filter length based on the L-curve. Simulation results using measured RIRs confirm that using a shorter reshaping filter length than conventionally used yields a significant increase in robustness against RIR perturbations for the MINT, RMCLS, and PMINT techniques. Furthermore, it is shown that the PMINT technique using an optimal intrusively determined reshaping filter length outperforms all other considered techniques in terms of perceptual speech quality. Finally, it is shown that the automatic non-intrusively determined reshaping filter length in the PMINT technique 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 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. Reverberation degrades the intelligibility of the speech signal [1,2] as well as the performance of automatic speech recognition systems [3,4].

In order to mitigate these detrimental effects of reverberation, several single- and multi-channel dereverberation techniques have been proposed [5], with multi-channel techniques being generally preferred since they exploit both the spectro-temporal and the spatial characteristics of the received microphone signals. Existing multi-channel dereverberation techniques can be broadly classified into spectral enhancement techniques [6–9], probabilistic modelling-based

techniques [10–13], and acoustic multi-channel equalization techniques [14–20]. Acoustic multi-channel equalization techniques aim to reshape the available room impulse responses (RIRs) between the speaker and the microphone array. They can in theory achieve perfect dereverberation [14,21], and hence, comprise an attractive approach to speech dereverberation.

A well-known multi-channel equalization technique aiming at acoustic system inversion is the multiple-input/output inverse theorem (MINT) technique [14], 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 [22]) or due to the sensitivity of blind system identification (BSI) and supervised system identification (SSI) methods to near-common zeros or interfering noise [23–25], MINT fails to invert the true RIRs, possibly leading to severe distortions in the output signal [17–19]. In order to increase the robustness against RIR perturbations, partial multi-channel

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

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, University of Oldenburg, Oldenburg, Germany

equalization techniques such as relaxed multi-channel least-squares (RMCLS) [18] and partial multi-channel equalization based on MINT (PMINT) [19] have been recently proposed. Since early reflections tend to improve speech intelligibility [26, 27] and late reflections are the major cause of speech intelligibility degradation, the objective of these 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 is more perceptually advantageous since it aims to also 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 [18, 19].

The length of the reshaping filters in MINT, RMCLS, and PMINT has been conventionally chosen such that perfect dereverberation can be achieved for perfectly estimated RIRs. Since in practice the available RIRs typically differ from the true ones, this choice of the reshaping filter length yields a high sensitivity of acoustic multi-channel equalization techniques to RIR perturbations. In this paper we propose to increase the robustness by decreasing the reshaping filter length, such that a better conditioned optimization criterion is obtained. A mathematical link between the reshaping filter length and the condition number of the 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. Although this result is rather intuitive, to the best of our knowledge it has not been mathematically analysed or investigated in the context of acoustic multi-channel equalization. In general, the optimal reshaping filter length yielding the highest performance needs to be determined intrusively (i.e., using a dereverberated reference signal and knowledge of the true RIRs), limiting the practical applicability of using shorter reshaping filter lengths in acoustic multi-channel equalization techniques. In this paper, we also propose and investigate an automatic non-intrusive selection procedure for the reshaping filter length based on the L-curve [28]. Some preliminary results on using shorter reshaping filters in acoustic multi-channel equalization techniques are presented in [29].

Simulation results for several measured acoustic systems and RIR perturbations show by means of instrumental measures that using shorter reshaping filters in MINT, RMCLS, and PMINT significantly increases the robustness against RIR perturbations. In addition, it is demonstrated that the P-MINT technique using the optimal intrusively determined reshaping filter

length outperforms the other considered techniques, yielding a high 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 in P-MINT, making this technique a practically applicable multi-channel equalization technique for speech dereverberation.

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 least-squares acoustic multi-channel equalization techniques, i.e., MINT, RMCLS, and PMINT, are briefly reviewed. In Section 4 their sensitivity to RIR perturbations is evaluated by means of the condition number and analytical insights on increasing the robustness by decreasing the reshaping filter length are provided. In Section 5 the automatic procedure for determining the reshaping filter length is discussed. Using instrumental 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 Figure 1. The m -th microphone signal, $m = 1, 2, \dots, M$, at time index n is given by

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

$$= x_m(n) + v_m(n), \quad (2)$$

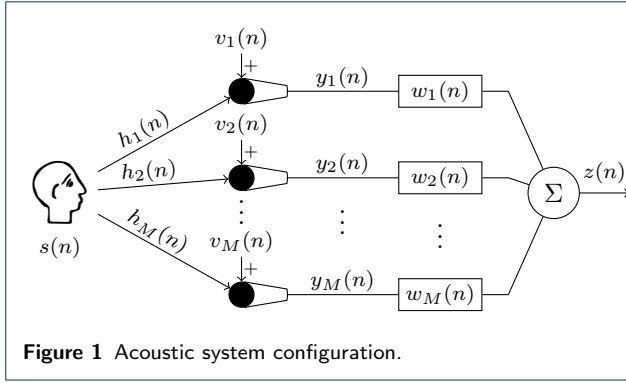
where $*$ denotes convolution, $s(n)$ is the clean speech signal, $h_m(n)$ is the RIR between the source and the m -th microphone, $x_m(n)$ is the reverberant speech component, and $v_m(n)$ is the noise component. Since acoustic multi-channel equalization techniques generally design dereverberation filters disregarding the presence of the additive noise, in the following it is assumed that $v_m(n) = 0$, hence, $y_m(n) = x_m(n)$.

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

$$z(n) = \sum_{m=1}^M x_m(n) * w_m(n) \quad (3)$$

$$= s(n) * \underbrace{\sum_{m=1}^M h_m(n) * w_m(n)}_{c(n)}, \quad (4)$$

where $w_m(n)$ is the filter applied to the m -th microphone signal and $c(n)$ denotes the equalized impulse



response (EIR) between the source and the output of the system. In vector notation, the RIR \mathbf{h}_m and the filter \mathbf{w}_m are given by

$$\mathbf{h}_m = [h_m(0) \ h_m(1) \ \dots \ h_m(L_h - 1)]^T, \quad (5)$$

$$\mathbf{w}_m = [w_m(0) \ w_m(1) \ \dots \ w_m(L_w - 1)]^T, \quad (6)$$

where L_h and L_w denote the RIR length and the filter length, respectively. Using the ML_w -dimensional stacked filter vector $\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \dots \ \mathbf{w}_M^T]^T$, the EIR vector \mathbf{c} of length $L_c = L_h + L_w - 1$, i.e., $\mathbf{c} = [c(0) \ c(1) \ \dots \ c(L_c - 1)]^T$, can be expressed as [14, 19]

$$\mathbf{c} = \mathbf{H}\mathbf{w}, \quad (7)$$

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

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \dots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h-1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h-1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_m(L_h-1) \end{bmatrix}. \quad (8)$$

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 (4) can be expressed as

$$z(n) = \mathbf{w}^T \mathbf{H}^T \mathbf{s}(n) = \mathbf{c}^T \mathbf{s}(n). \quad (9)$$

The reshaping filter \mathbf{w} can 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 a filter \mathbf{w} such that the EIR \mathbf{c} in (7) is equal to a dereverberated target EIR \mathbf{c}_d . It should be realized that in practice however, 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 [22]) or due to the sensitivity of BSI and SSI methods to near-common zeros or interfering noise [23–25]. 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.

State-of-the-art least-squares acoustic multi-channel equalization techniques seek to compute the filter \mathbf{w} as the solution to the system of equations

$$\mathbf{W}\hat{\mathbf{H}}\mathbf{w} = \mathbf{W}\mathbf{c}_d, \quad (10)$$

with \mathbf{W} an $L_c \times L_c$ -dimensional diagonal weighting matrix. The definition of the weighting matrix \mathbf{W} and target EIR \mathbf{c}_d for the equalization techniques considered in this paper, i.e., MINT [14], RMCLS [18], and PMINT [19], is presented in Table 1, where \mathbf{I} denotes the $L_c \times L_c$ -dimensional identity matrix, 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 [17]. The length of the direct path and early reflections L_d in the RMCLS and PMINT techniques is typically considered to be between 10–50 ms [18, 19]. Based on these definitions of \mathbf{W} and \mathbf{c}_d it can be said that on the one hand, the MINT and PMINT techniques constrain all taps of the EIR (i.e., $\mathbf{W} = \mathbf{I}$), which results in a good perceptual speech quality but high sensitivity to RIR perturbations [17, 19], while on the other hand, the RMCLS technique does not constrain all taps of the EIR (i.e., $\mathbf{W} \neq \mathbf{I}$), which results in a lower sensitivity to RIR perturbations but decreased perceptual speech quality [18, 19].

For all considered equalization techniques, the filter solving (10) is computed by minimizing the least-squares cost function

$$J_{\text{LS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2. \quad (11)$$

As shown in [14, 18, 19], assuming that the RIRs do not share any common zeros and using a reshaping filter length

$$L_w \geq \left\lceil \frac{L_h - 1}{M - 1} \right\rceil, \quad (12)$$

Table 1 Definition of the weighting matrix \mathbf{W} and target EIR \mathbf{c}_d for state-of-the-art least-squares equalization techniques.

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

with $\lceil \cdot \rceil$ the ceiling operator, the least-squares filter solving (10) and hence, minimizing (11) to 0, can be computed as [14, 18, 19]

$$\mathbf{w}_{\text{LS}} = (\mathbf{W}\hat{\mathbf{H}})^+(\mathbf{W}\mathbf{c}_d), \quad (13)$$

where $\{\cdot\}^+$ denotes the matrix pseudo-inverse. When the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the least-squares filter of length according to (12) yields perfect dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{w}_{\text{LS}} = \mathbf{W}\mathbf{c}_d$. However, in the presence of RIR perturbations, i.e., $\hat{\mathbf{H}} \neq \mathbf{H}$, this least-squares filter typically fails to achieve dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{w}_{\text{LS}} \neq \mathbf{W}\mathbf{c}_d$, possibly causing severe distortions in the output signal [19]. In this paper, the sensitivity of the least-squares filter to RIR perturbations is evaluated by the so-called condition number of the matrix $\mathbf{W}\hat{\mathbf{H}}$.

4 Robust acoustic multi-channel equalization

In this section, the Wedin theorem [30] relating the condition number of the matrix being inverted to the sensitivity of least-squares solutions 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 $\mathbf{W}\hat{\mathbf{H}}$, increasing the robustness of acoustic multi-channel equalization techniques against RIR perturbations. For clarity of presentation, the notation summarized in Table 2 is used in the following.

Wedin theorem [30]: 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 $\chi_{\mathbf{A}}$ of the matrix \mathbf{A} is defined as

$$\chi_{\mathbf{A}} = \frac{\|\mathbf{A}\|_2}{\|\mathbf{A}^+\|_2} = \frac{\sigma_{\mathbf{A}}(1)}{\sigma_{\mathbf{A}}(r)}, \quad (14)$$

with $\sigma_{\mathbf{A}}(i)$ the i -th singular value of the matrix \mathbf{A} , ordered such that $\sigma_{\mathbf{A}}(1) \geq \sigma_{\mathbf{A}}(2) \geq \dots \geq \sigma_{\mathbf{A}}(r) > 0$.

Using $\chi_{\mathbf{A}}$ and defining the variable ξ as

$$\xi = \frac{\|\Delta\mathbf{A}\|_2}{\|\mathbf{A}\|_2}, \quad (15)$$

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

$$\|\Delta\mathbf{q}\|_2 \leq \frac{\chi_{\mathbf{A}}\xi\|\mathbf{q}\|_2}{1 - \chi_{\mathbf{A}}\xi} + \|(\mathbf{A}\mathbf{A}^T)^+\mathbf{b}\|_2\|\mathbf{A}\|_2, \quad (16)$$

where it is assumed that $\chi_{\mathbf{A}}\xi < 1$.

The relation in (16) shows that a large condition number $\chi_{\mathbf{A}}$ can result (and typically does) in a large deviation between the true and the perturbed solution [30, 31].

In order to satisfy (12), reshaping filters in acoustic multi-channel equalization techniques are conventionally designed using the filter length L_w^t , 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 $r_t \leq p_t$. However, reshaping filters can also be designed using a shorter filter length L_w^s , i.e., based on the $p_s \times q_s$ -dimensional matrix $\mathbf{W}_s\hat{\mathbf{H}}_s$. Considering that $L_w^s < \left\lceil \frac{L_h-1}{M-1} \right\rceil$, which implies that $L_w^s < \frac{L_h-1}{M-1}$, 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

$$(M-1)L_w^s < L_h-1 \Rightarrow \underbrace{ML_w^s}_{q_s} < \underbrace{L_h + L_w^s - 1}_{p_s}. \quad (17)$$

In addition, the matrix $\mathbf{W}_s\hat{\mathbf{H}}_s$ is a sub-matrix of $\mathbf{W}_t\hat{\mathbf{H}}_t$, constructed by deleting $L_w^t - L_w^s$ rows and $M(L_w^t - L_w^s)$ 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 $\mathbf{W}_s\hat{\mathbf{H}}_s$, with

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

$$\chi_{\mathbf{W}_s\hat{\mathbf{H}}_s} = \frac{\sigma_{\mathbf{W}_s\hat{\mathbf{H}}_s}(1)}{\sigma_{\mathbf{W}_s\hat{\mathbf{H}}_s}(r_s)}, \quad (19)$$

we consider the interlacing inequalities [32] between the singular values of a matrix and its sub-matrices presented in the following.

Table 2 Notation for different reshaping filter lengths and the corresponding least-squares matrices.

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

Interlacing inequalities [32]: 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

$$\sigma_{\mathbf{A}}(i) \geq \sigma_{\mathbf{B}}(i) \geq \sigma_{\mathbf{A}}(i + l), \quad (20)$$

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

Using (20), in Appendix A we derive the following inequalities relating the largest and smallest non-zero singular values of $\mathbf{W}_t \hat{\mathbf{H}}_t$ and $\mathbf{W}_s \hat{\mathbf{H}}_s$:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (21)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (22)$$

It readily follows from (21) and (22) that the condition number of $\mathbf{W}_s \hat{\mathbf{H}}_s$ is smaller or equal than the condition number of $\mathbf{W}_t \hat{\mathbf{H}}_t$, i.e.,

$$\chi_{\mathbf{W}_s \hat{\mathbf{H}}_s} = \frac{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1)}{\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s)} \leq \frac{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1)}{\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t)} = \chi_{\mathbf{W}_t \hat{\mathbf{H}}_t}. \quad (23)$$

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

Figure 2 depicts the singular values of an exemplary matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ for the PMINT technique, constructed using the conventional reshaping filter length $L_w^t = 1947$. 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 exemplary sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$, constructed using $L_w^s = 1000$ and $L_w^s = 300$, are also depicted. The largest and smallest non-zero singular values of each matrix are marked in order to illustrate the inequalities presented in (21) and (22). 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 $\mathbf{W} \hat{\mathbf{H}}$.

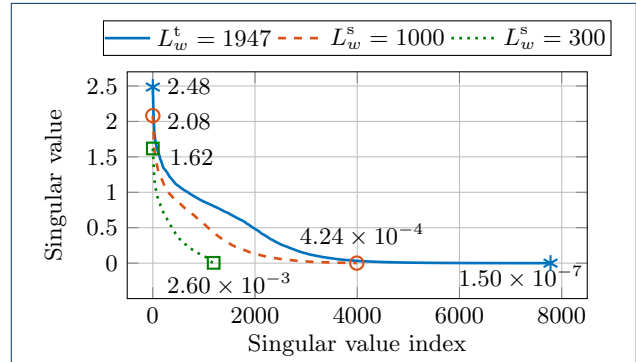


Figure 2 Singular values of an exemplary matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ ($L_w^t = 1947$) and of two sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$ ($L_w^s = 1000$ and $L_w^s = 300$) for the PMINT technique. The largest and smallest non-zero singular values of each matrix are explicitly denoted. The considered acoustic system is system S_1 described in Section 6.1.

Table 3 Condition number of an exemplary matrix $\mathbf{W}_t \hat{\mathbf{H}}_t$ ($L_w^t = 1947$) and of two sub-matrices $\mathbf{W}_s \hat{\mathbf{H}}_s$ ($L_w^s = 1000$ and $L_w^s = 300$) for the PMINT technique. The considered acoustic system is system S_1 described in Section 6.1.

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

Using a shorter reshaping filter can be considered as a method of regularizing the least-squares solution, yielding a trade-off between equalization performance when the true RIRs are available and robustness in the presence of RIR perturbations. Designing shorter reshaping filters is not only desirable to increase the robustness of least-squares equalization techniques against RIR perturbations, but also because of the lower computational complexity of the reshaping filter design.

5 Automatic non-intrusive reshaping filter length

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. Obviously, different reshaping filter lengths yield a different performance in acoustic multi-channel equalization techniques. The optimal length L_w^{opt} yielding the highest performance depends on the acoustic system being equalized and the RIR perturbation level. While in simulations L_w^{opt} can be intrusively determined exploiting the true known RIRs (cf. Section 6.1), 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 least-squares techniques [19, 33, 34], in this section we propose to use the L-curve to automatically determine the reshaping filter length L_w^{auto} in acoustic multi-channel equalization techniques.

As previously mentioned, using a shorter reshaping filter introduces a trade-off between the condition number $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and the least-squares error $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$. An appropriate filter length should incorporate knowledge about both $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$, such that both quantities are kept small. Due to the arising trade-off between these quantities, the parametric plot of the condition number versus the least-squares error for several reshaping filter lengths has an L-shape with the corner (i.e., point of maximum curvature) located where the least-squares filter changes from being dominated by a large condition number to being dominated by a large error. Hence, we propose to automatically select the reshaping filter length L_w^{auto} in acoustic multi-channel equalization techniques as the one corresponding to the corner of the parametric plot of the condition number $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ versus the least-squares error $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$.

Figure 3 depicts a typical L-curve obtained using PMINT to equalize a typical acoustic system S_1 described in Section 6.1. As illustrated in this figure, decreasing the reshaping filter length decreases the condition number but at the same time increases the least-squares error. Although from such a curve it seems easy to determine the reshaping filter length that corresponds to the corner of the L-curve (i.e., $L_w^s = 1000$ in this example), numerical problems may occur, and hence, a numerically stable algorithm is required. As in [19], in this

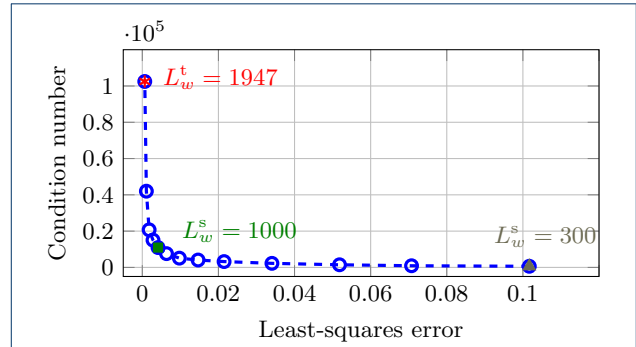


Figure 3 Typical L-curve obtained using PMINT with several reshaping filter lengths ranging from $L_w^t = 1947$ to $L_w^s = 300$. The considered acoustic system is system S_1 described in Section 6.1.

work we use the triangle method [28] for locating the corner of the L-curve.

6 Simulation results

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

6.1 Acoustic systems, instrumental measures, and algorithmic settings

Acoustic systems

We have considered 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 [35] and the reverberant signals are generated by convolving 10 sentences (approximately 17 s long) from the HINT database [36] with the measured RIRs. For each acoustic system, Table 4 presents the reverberation time T_{60} of the room and the length of the RIRs L_h at a sampling frequency $f_s = 8$ kHz. In order to simulate RIR perturbations, the measured RIRs are perturbed by proportional Gaussian distributed errors

Table 4 Characteristics of the considered measured acoustic systems.

Acoustic system	T_{60} [ms]	L_h
S_1	730	5840
S_2	610	4880
S_3	360	2880

as proposed in [37], such that a desired level of 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. 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 achieved by state-of-the-art BSI methods (for relatively short RIRs) range between -10 dB and -20 dB, hence, the values in (25) can be considered as rather realistic [25].

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) [5], i.e., $\Delta\text{DRR} = \text{oDRR} - \text{iDRR}$, with

$$\text{oDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} c^2(n)}{\sum_{n=n_d}^{L_c-1} c^2(n)}, \quad \text{iDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} h_1^2(n)}{\sum_{n=n_d}^{L_h-1} h_1^2(n)}, \quad (26)$$

where the first $n_d = 0.008 \times f_s$ samples of $c(n)$ and $h_1(n)$ represent the direct path propagation and the remaining samples represent reflections. Although the 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 insights on the reverberant energy decay rate. To evaluate the reverberant energy decay rate, the energy decay curve (EDC) [5] of the EIR $c(n)$ is compared to the energy decay curve of the true RIR $h_1(n)$. The EDC of the EIR is computed

as

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

and the EDC of the RIR is computed as

$$\text{EDC}_{h_1}(n) = \frac{1}{\|\mathbf{h}_1\|_2^2} \sum_{j=n}^{L_h-1} h_1^2(j), \quad n = 0, 1, \dots, L_h-1. \quad (28)$$

The perceptual speech quality is evaluated using the frequency weighted segmental signal-to-noise-ratio (fSNR) and the cepstral distance (CD) [38]. In [39] 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 the direct path and early reflections speech signal in the first microphone, i.e., the signal obtained by convolving the clean speech signal with the direct path and early reflections (considered to be 10 ms long) of \mathbf{h}_1 . 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.

Algorithmic settings

For all considered equalization techniques the conventionally used filter length is $L_w^t = \left\lceil \frac{L_h-1}{M-1} \right\rceil$, i.e., $L_w^t = 1947$ for system S_1 , $L_w^t = 1627$ for system S_2 , and $L_w^t = 960$ for system S_3 . The delay is set to $\tau = 90$ and the length of the direct path and early reflections is set to $L_d = 0.01 \times f_s$, i.e., 10 ms (cf. Table 1). The target EIR \mathbf{c}_d for the PMINT technique 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.,

$$L_w^s \in \begin{cases} \{300, 330, \dots, 1920\} & \text{for system } S_1 \\ \{300, 330, \dots, 1620\} & \text{for system } S_2 \\ \{300, 330, \dots, 930\} & \text{for system } S_3 \end{cases} \quad (29)$$

For each acoustic system, each NPM, and each equalization technique the optimal filter length L_w^{opt} is selected from (29) as the one yielding the lowest CD. It should be noted that using the CD for determining the optimal reshaping filter length is an intrusive procedure not applicable in practice, since knowledge of the clean speech signal and of the true RIRs 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 the MINT, RMCLS, and PMINT techniques is investigated when using the conventional reshaping filter length L_w^t and the optimal reshaping filter length L_w^{opt} . Although similar results are obtained for all considered acoustic systems, in this section only results for acoustic system S_1 are presented.

MINT using L_w^t and L_w^{opt} . Figure 4 depicts the performance of the MINT technique when using the filter lengths L_w^t and L_w^{opt} in terms of ΔDRR , EDC, ΔfSNR , and ΔCD . As expected, the ΔDRR values presented in Figure 4(a) show that using the conventional filter length L_w^t in MINT fails to suppress the reverberant energy, even worsening the DRR by an average of 15 dB in comparison to \mathbf{h}_1 . Furthermore, it can be observed that using the shorter filter length L_w^{opt} significantly increases the robustness of MINT for all NPMs, improving the DRR by an average of 4 dB in comparison to \mathbf{h}_1 . These results are confirmed in Figure 4(b), which depicts the EDC of \mathbf{h}_1 and the EDCs of the EIRs obtained using MINT with L_w^t and L_w^{opt} for an exemplary NPM of -33 dB. It can be observed that while using L_w^t completely fails to achieve dereverberation and results in a slower reverberant energy decay rate than \mathbf{h}_1 , using L_w^{opt} yields a significantly faster reverberant energy decay rate. However, using L_w^{opt} yields only a slight improvement of the reverberant energy decay rate when compared to \mathbf{h}_1 , even for the moderate NPM value of -33 dB. Again as expected, the ΔfSNR and ΔCD values depicted in Figures 4(c) and 4(d) show that using the conventional filter length L_w^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_w^{opt} , the performance in terms of the ΔfSNR is still worse than for the unprocessed microphone signal $x_1(n)$.

Therefore as expected from the theoretical analysis in Section 4, these simulation results validate that using a shorter reshaping filter 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, even a shorter reshaping filter is not sufficient to make the MINT technique fully robust against RIR perturbations.

RMCLS using L_w^t and L_w^{opt} . Figure 5 depicts the performance of the RMCLS technique using the filter lengths L_w^t and L_w^{opt} in terms of ΔDRR , EDC, ΔfSNR , and ΔCD . The ΔDRR values presented in Figure 5(a) show that for moderate NPMs using the conventional filter length L_w^t in RMCLS slightly suppresses the reverberant energy, whereas for NPMs larger than -21 dB using L_w^t in RMCLS fails to achieve dereverberation and worsens the DRR in comparison to \mathbf{h}_1 . Furthermore, it can be observed that using the optimal reshaping filter length L_w^{opt} significantly increases the reverberant energy suppression, yielding on average a 6 dB larger ΔDRR in comparison to the ΔDRR obtained using L_w^t . To evaluate the reverberant energy decay rate, Figure 5(b) depicts the EDC of \mathbf{h}_1 and the EDCs of the EIRs obtained using RMCLS with L_w^t and L_w^{opt} for an exemplary NPM of -33 dB. It can be observed that for this moderate NPM, a very similar reverberant energy decay rate is obtained for the RMCLS technique when using L_w^t and L_w^{opt} . Since the RMCLS technique using the conventional filter length L_w^t is rather robust yielding 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 Figures 5(c) and 5(d), which shows that using L_w^{opt} in RMCLS significantly improves the ΔfSNR and ΔCD for all NPMs.

Therefore as expected from the theoretical analysis in Section 4, these simulation results demonstrate that using a shorter reshaping filter than conventionally used in RMCLS is advantageous and increases the robustness against RIR perturbations.

PMINT using L_w^t and L_w^{opt} . Figure 6 depicts the performance of the PMINT technique using the filter lengths L_w^t and L_w^{opt} in terms of ΔDRR , EDC, ΔfSNR , and ΔCD . As expected, the ΔDRR values presented in Figure 6(a) show that the PMINT technique using the conventionally used filter length L_w^t fails to suppress the reverberant energy, even worsening the DRR in comparison to \mathbf{h}_1 . Furthermore, it can be observed that using L_w^{opt} in PMINT significantly increases the robustness for all NPMs, improving the DRR by an average of 11 dB in comparison to \mathbf{h}_1 . These results are further confirmed in Figure 6(b), which depicts the EDC of \mathbf{h}_1 and the EDCs of the EIRs obtained using PMINT with L_w^t and L_w^{opt} for an exemplary NPM of -33 dB. It can be observed that the PMINT technique using L_w^t completely fails to achieve dereverberation

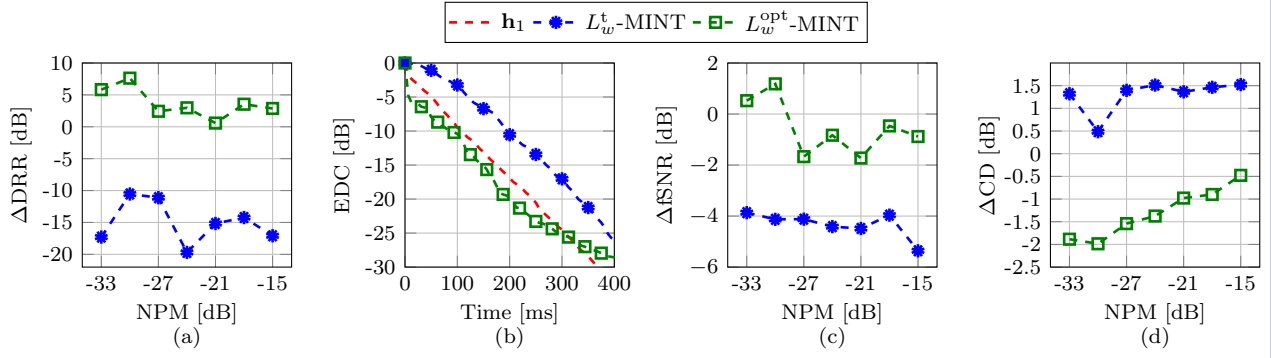


Figure 4 Performance of the MINT technique using the conventional filter length L_w^t and the intrusively determined optimal filter length L_w^{opt} for acoustic system S_1 in terms of (a) ΔDRR , (b) EDC for an exemplary NPM of -33 dB, (c) ΔfSNR , and (d) ΔCD .

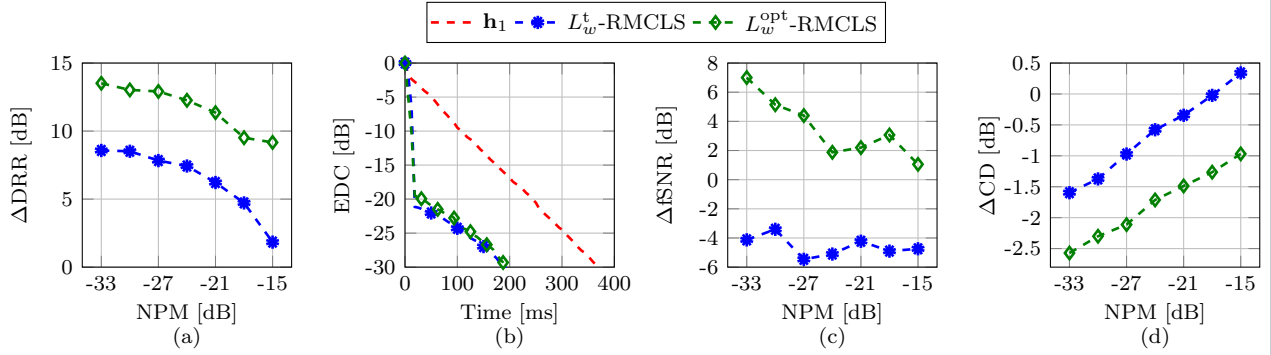


Figure 5 Performance of the RMCLS technique using the conventional filter length L_w^t and the intrusively determined optimal filter length L_w^{opt} for acoustic system S_1 in terms of (a) ΔDRR , (b) EDC for an exemplary NPM of -33 dB, (c) ΔfSNR , and (d) ΔCD .

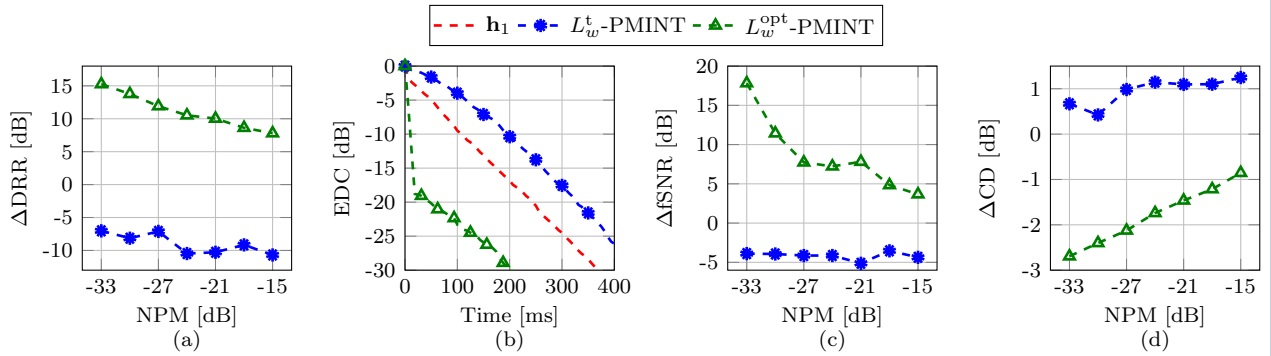


Figure 6 Performance of the PMINT technique using the conventional filter length L_w^t and the intrusively determined optimal filter length L_w^{opt} for acoustic system S_1 in terms of (a) ΔDRR , (b) EDC for an exemplary NPM of -33 dB, (c) ΔfSNR , and (d) ΔCD .

and even results in a slower reverberant energy decay rate than h_1 . Using the optimal reshaping filter length L_w^{opt} yields a significant increase in robustness, resulting in a much faster reverberant energy decay rate than h_1 . Furthermore, the ΔfSNR and ΔCD values depicted in Figures 6(c) and 6(d) show that while PMINT using

L_w^t worsens the perceptual speech quality in comparison to the unprocessed microphone signal $x_1(n)$, using L_w^{opt} results in a significantly better performance.

Therefore as expected from the theoretical analysis in Section 4, decreasing the reshaping filter length in PMINT results in a significant increase in robustness

Table 5 Average performance of the MINT, RMCLS, and PMINT techniques using the optimal reshaping filter length L_w^{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_w^{opt} -MINT	3.69	-0.55	-1.31	1.25	-2.07	0.07	0.42	-3.89	-0.22
L_w^{opt} -RMCLS	11.68	3.53	-1.77	3.57	-0.81	-0.39	3.96	-0.61	-0.52
L_w^{opt} -PMINT	11.13	8.65	-1.78	5.99	2.58	-0.66	4.33	1.88	-0.57

Table 6 Average performance of the PMINT technique using the automatically non-intrusively determined reshaping filter length L_w^{auto} .

	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_w^{auto} -PMINT	10.58	7.90	-1.68	4.41	1.62	-0.33	3.34	0.14	-0.29

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

Since it was shown in the previous section that using a shorter reshaping filter than conventionally used increases the robustness of the considered equalization techniques in terms of all performance measures, in this section the performance of the MINT, RMCLS, and PMINT techniques using the optimal intrusively determined filter length L_w^{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 , and the presented performance measures are averaged over all considered NPMs.

Table 5 presents the obtained ΔDRR , ΔfSNR , and ΔCD for all considered techniques^[1]. First it can be observed that the MINT technique using L_w^{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 the MINT technique is very sensitive to RIR perturbations (cf. Figure 4), the robustness increase that can be obtained by using a shorter reshaping filter length is also limited. Second it can be observed that the RMCLS and PMINT techniques using L_w^{opt} result in a high reverberant energy suppression in terms of ΔDRR , with the PMINT technique outperforming the RMCLS technique for system S_2 whereas a similar performance is obtained for systems S_1 and S_3 . Finally it can be observed that for all considered acoustic systems the PMINT technique using the reshaping filter length L_w^{opt} yields the highest perceptual

speech quality improvement, outperforming the RMCLS technique in terms of ΔfSNR . While the PMINT technique always improves the perceptual speech quality in comparison to the unprocessed microphone signal $x_1(n)$, the RMCLS technique might fail to yield an improvement, as indicated by the negative ΔfSNR for systems S_2 and S_3 . The advantage of using the PMINT technique lies in its control of the early reflections in the EIR, hence better preserving the perceptual speech quality of the output signal.

In summary, it can be said that the PMINT technique using the optimal intrusively determined reshaping filter length L_w^{opt} is a robust and perceptually advantageous equalization technique, yielding a high reverberant energy suppression and outperforming all other considered techniques in terms of perceptual speech quality improvement.

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

In this section we investigate the performance of the PMINT technique when using the automatic non-intrusive and practically applicable procedure for determining the reshaping filter length L_w^{auto} (discussed in Section 5) instead of the optimal intrusively determined reshaping filter length L_w^{opt} . Similarly as in Section 6.3, we consider all acoustic systems in Table 4 and all NPMs in (25). For all reshaping filter lengths in (29) the matrix $\mathbf{W}\hat{\mathbf{H}}$ is constructed and the PMINT reshaping filter is designed. Furthermore, in order to generate the parametric L-curve the condition number $\chi_{\mathbf{W}\hat{\mathbf{H}}}$ and the least-squares error $\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_d)\|_2^2$ is computed. Using the triangle method [28], the automatic reshaping filter length L_w^{auto} corresponding to the point of maximum curvature of the L-curve is determined. The performance of the PMINT technique using L_w^{auto} is evaluated in terms of ΔDRR , ΔfSNR , and ΔCD , and the presented performance measures are averaged over all considered NPMs.

Table 6 presents the ΔDRR , ΔfSNR , and ΔCD obtained using PMINT with L_w^{auto} for all considered

^[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.

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. When comparing the performance measures presented in Table 6 to the ones presented in Table 5, it can be said that the performance of PMINT when using L_w^{auto} is generally similar to the performance obtained when using L_w^{opt} . More precisely, the average performance degradation over all considered acoustic systems when using the automatic non-intrusively determined reshaping filter length L_w^{auto} instead of the optimal intrusively determined reshaping filter length L_w^{opt} is only 1.04 dB in terms of ΔDRR , 2.29 dB in terms of ΔfSNR , 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 the PMINT technique using a shorter reshaping filter length a robust, perceptually advantageous, and practically applicable technique.

7 Conclusion

In this paper we have proposed to use a shorter reshaping filter length than conventionally used in acoustic multi-channel equalization techniques, such that a better conditioned optimization criterion is obtained. We have analytically shown that using a shorter reshaping filter decreases the condition number of the 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 least-squares error, such that both quantities are kept small. Using instrumental 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.

Appendix A

In order to construct the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$, we first create an intermediate $[p_t - (L_w^t - L_w^s)] \times [q_t - (L_w^t - L_w^s)]$ -dimensional sub-matrix \mathbf{T} by deleting $L_w^t - L_w^s$ rows

and $L_w^t - L_w^s$ columns from $\mathbf{W}_t \hat{\mathbf{H}}_t$. The interlacing inequalities in (20) for the matrices $\mathbf{W}_t \hat{\mathbf{H}}_t$ and \mathbf{T} can be written as

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(i) \geq \sigma_{\mathbf{T}}(i) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(i + (L_w^t - L_w^s)), \quad (30)$$

for $i = 1, \dots, r_t - (L_w^t - L_w^s), \dots, p_t - (L_w^t - L_w^s)$. Based on (30), the following inequalities between the singular values of the matrices $\mathbf{W}_t \hat{\mathbf{H}}_t$ and \mathbf{T} can be established:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{T}}(1), \quad (31)$$

$$\sigma_{\mathbf{T}}(r_t - (L_w^t - L_w^s)) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (32)$$

In order to construct the matrix $\mathbf{W}_s \hat{\mathbf{H}}_s$, $(M-1)(L_w^t - L_w^s)$ columns are now deleted from the matrix \mathbf{T} . The interlacing inequalities in (20) for the matrices \mathbf{T} and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can be written as

$$\sigma_{\mathbf{T}}(i) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(i) \geq \sigma_{\mathbf{T}}(i + (M-1)(L_w^t - L_w^s)), \quad (33)$$

for $i = 1, \dots, r_s$. Using (33), the following inequalities between the singular values of the matrices \mathbf{T} and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can be written:

$$\sigma_{\mathbf{T}}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (34)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{T}}(r_s + (M-1)(L_w^t - L_w^s)). \quad (35)$$

The index of the singular value in the right hand side of (35) can be written as

$$\begin{aligned} r_s + (M-1)(L_w^t - L_w^s) &= ML_w^s + (M-1)(L_w^t - L_w^s) \\ &= ML_w^t - (L_w^t - L_w^s) \geq r_t - (L_w^t - L_w^s), \end{aligned} \quad (36)$$

with (36) clearly holding since the number of columns in $\mathbf{W}_t \hat{\mathbf{H}}_t$ is greater or equal than its rank, i.e.,

$$q_t = ML_w^t \geq p_t \geq r_t. \quad (37)$$

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

$$\sigma_{\mathbf{T}}(r_s + (M-1)(L_w^t - L_w^s)) \geq \sigma_{\mathbf{T}}(r_t - (L_w^t - L_w^s)). \quad (38)$$

Using (38), the inequality in (35) can also be written as

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{T}}(r_t - (L_w^t - L_w^s)). \quad (39)$$

Finally, combining (31), (32), (34), and (39) the following inequalities relating the largest and smallest

non-zero singular values of $\mathbf{W}_t \hat{\mathbf{H}}_t$ and $\mathbf{W}_s \hat{\mathbf{H}}_s$ can be established:

$$\sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(1) \geq \sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(1), \quad (40)$$

$$\sigma_{\mathbf{W}_s \hat{\mathbf{H}}_s}(r_s) \geq \sigma_{\mathbf{W}_t \hat{\mathbf{H}}_t}(r_t). \quad (41)$$

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG), the Marie Curie Initial Training Network DREAMS (Grant no. 316969), and the joint Lower Saxony-Israeli Project ATHENA, funded by the State of Lower Saxony.

References

- Beutelmann, R., Brand, T.: Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* **120**(1), 331–342 (2006)
- Warzybok, A., Kodrasi, I., Jungmann, J.O., Habets, E.A.P., Gerkmann, T., Mertins, A., Doclo, S., Kollmeier, B., Goetze, S.: Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithm. In: *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, pp. 333–337 (2014)
- Yoshioka, T., Sehr, A., Delcroix, M., Kinoshita, K., Maas, R., Nakatani, T., Kellermann, W.: Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition. *IEEE Signal Processing Magazine* **29**(6), 114–126 (2012)
- Xiong, F., Meyer, B.T., Moritz, N., Rehr, R., Anemüller, J., Gerkmann, T., Doclo, S., Goetze, S.: Front-end technologies for robust ASR in reverberant environments—spectral enhancement-based dereverberation and auditory modulation filterbank features. *EURASIP Journal on Advances in Signal Processing* **2015**(70) (2015)
- Naylor, P.A., Gaubitch, N.D. (eds.): *Speech Dereverberation*. Springer, London, UK (2010)
- Lebart, K., Boucher, J.M.: A new method based on spectral subtraction for speech dereverberation. *Acta Acoustica* **87**(3), 359–366 (2001)
- Habets, E.A.P., Gannot, S., Cohen, I.: Late reverberant spectral variance estimation based on a statistical model. *IEEE Signal Processing Letters* **16**(9), 770–774 (2009)
- Braun, S., Habets, E.A.P.: A multichannel diffuse power estimator for dereverberation in the presence of multiple sources. *EURASIP Journal on Audio, Speech, and Music Processing* **2015**(24) (2015)
- Kuklasinski, A., Doclo, S., Jensen, S.H., Jensen, J.: Maximum Likelihood PSD estimation for speech enhancement in reverberation and noise. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**(9), 1595–1608 (2016)
- Nakatani, T., Yoshioka, T., Kinoshita, K., Miyoshi, M., Juang, B.-H.: Speech dereverberation based on variance-normalized delayed linear prediction. *IEEE Transactions on Audio, Speech, and Language Processing* **18**(7), 1717–1731 (2010)
- Schmid, D., Enzner, G., Malik, S., Kolossa, D., Martin, R.: Variational Bayesian inference for multichannel dereverberation and noise reduction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**(8), 1320–1335 (2014)
- Schwartz, B., Gannot, S., Habets, E.A.P.: Online speech dereverberation using Kalman filter and EM algorithm. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**(2), 394–406 (2015)
- A. Jukić, T. Van Waterschoot, T. Gerkmann, and S. Doclo: Multi-channel linear prediction-based speech dereverberation with sparse priors. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **23**(9), 1509–1520 (2015)
- Miyoshi, M., Kaneda, Y.: Inverse filtering of room acoustics. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**(2), 145–152 (1988)
- Kallinger, M., Mertins, A.: Multi-channel room impulse response shaping - a study. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, pp. 101–104 (2006)
- Jungmann, J.O., Mazur, R., Kallinger, M., Tiemin, M., Mertins, A.: Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping. *IEEE Transactions on Audio, Speech, and Language Processing* **20**(6), 1829–1842 (2012)
- Hikichi, T., Delcroix, M., Miyoshi, M.: Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations. *EURASIP Journal on Advances in Signal Processing* **2007** (2007)
- Lim, F., Zhang, W., Habets, E.A.P., Naylor, P.A.: Robust multichannel dereverberation using relaxed multichannel least squares. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**(9), 1379–1390 (2014)
- Kodrasi, I., Goetze, S., Doclo, S.: Regularization for partial multichannel equalization for speech dereverberation. *IEEE Transactions on Audio, Speech, and Language Processing* **21**(9), 1879–1890 (2013)
- Rashobh, R.S., Khong, A.W.H., Liu, D.: Multichannel equalization in the KLT and frequency domains with application to speech dereverberation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**(3), 634–646 (2014)
- Hacihabiboglu, H., Cvetkovic, Z.: Multichannel dereverberation theorems and robustness issues. *IEEE Transactions on Audio, Speech, and Language Processing* **20**(2), 67–689 (2012)
- Radlovic, B.D., Williamson, R.C., Kennedy, R.A.: Equalization in an acoustic reverberant environment: robustness results. *IEEE Transactions on Speech and Audio Processing* **8**(3), 311–319 (2000)
- Khong, A.W.H., Xiang, L., Naylor, P.A.: Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, USA, pp. 229–232 (2008)
- Haque, M.A., Hasan, T.: Noise robust multichannel frequency-domain LMS algorithms for blind channel identification. *IEEE Signal Processing Letters* **15**, 305–308 (2008)
- Hu, M., Gaubitch, N.D., Naylor, P.A., Ward, D.B.: Noise robust blind system identification algorithms based on a Rayleigh quotient cost function. In: *Proc. European Signal Processing Conference*, Nice, France (2015)
- Arweiler, I., Buchholz, J.M.: The influence of spectral characteristics of early reflections on speech intelligibility. *Journal of the Acoustical Society of America* **130**(2), 996–1005 (2011)
- Warzybok, A., Rennie, J., Brand, T., Doclo, S., Kollmeier, B.: Effects of spatial and temporal integration of a single early reflection on speech intelligibility. *Journal of the Acoustical Society of America* **133**(1), 269–282 (2013)
- Castellanos, J.L., Gómez, S., Guerra, V.: The triangle method for finding the corner of the L-curve. *Applied Numerical Mathematics* **43**(4), 359–373 (2002)
- Kodrasi, I., Doclo, S.: The effect of inverse filter length on the robustness of acoustic multichannel equalization. In: *Proc. European Signal Processing Conference*, Bucharest, Romania (2012)
- Wedin, P.: Perturbation theory for pseudo-inverses. *BIT Numerical Mathematics* **13**(2), 217–232 (1973)
- G. Golub and C. Van Loan: *Matrix Computations*. The John Hopkins University Press, Baltimore, USA (1996)
- Horn, R.A., Johnson, C.R.: *Topics in matrix analysis*. Cambridge University Press, Cambridge, United Kingdom (1999)
- Hansen, P.C.: Analysis of discrete ill-posed problems by means of the L-curve. *SIAM review* **34**(4), 561–580 (1992)
- Hansen, P.C., O'Leary, D.P.: The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing* **14**(6), 1487–1503 (1993)
- Farina, A.: Simultaneous measurement of impulse response and distortion with a swept-sine technique. In: *Proc. AES Convention*,

- Paris, France, pp. 18–22 (2000)
36. Nilsson, M., Soli, S.D., Sullivan, A.: Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America* **95**(2), 1085–1099 (1994)
 37. Zhang, W., Naylor, P.A.: An algorithm to generate representations of system identification errors. *Research Letters in Signal Processing* **2008** (2008)
 38. Hu, Y., Loizou, P.C.: Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing* **16**(1), 229–238 (2008)
 39. Kinoshita, K., Delcroix, M., Gannot, S., Habets, E.A.P., Haeb-Umbach, R., Kellermann, W., Leutnant, V., Maas, R., Nakatani, T., Raj, B., Sehr, A., Yoshioka, T.: A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. *EURASIP Journal on Advances in Signal Processing* **2016**(7) (2016)