Regularization for Partial Multichannel Equalization for Speech Dereverberation

Ina Kodrasi, *Student Member, IEEE*, Stefan Goetze, *Member, IEEE*, and Simon Doclo, *Member, IEEE*

***Abstract*—Acoustic multichannel equalization techniques such as the multiple-input/output inverse theorem (MINT), which aim to equalize the room impulse responses (RIRs) between the source and the microphone array, are known to be highly sensitive to RIR estimation errors. To increase robustness, it has been proposed to incorporate regularization in order to decrease the energy of the equalization filters. In addition, more robust partial multichannel equalization techniques such as relaxed multichannel least-squares (RMCLS) and channel shortening (CS) have recently been proposed. In this paper, we propose a partial multichannel equalization technique based on MINT (P-MINT) which aims to shorten the RIR. Furthermore, we investigate the effectiveness of incorporating regularization to further increase the robustness of P-MINT and the aforementioned partial multichannel equalization techniques, i.e., RMCLS and CS. In addition, we introduce an automatic non-intrusive procedure for determining the regularization parameter based on the L-curve. Simulation results using measured RIRs show that incorporating regularization in P-MINT yields a significant performance improvement in the presence of RIR estimation errors, whereas a smaller performance improvement is observed when incorporating regularization in RMCLS and CS. Furthermore, it is shown that the intrusively regularized P-MINT technique outperforms all other investigated intrusively regularized multichannel equalization techniques in terms of perceptual speech quality (PESQ). Finally, it is shown that the automatic non-intrusive regularization parameter in regularized P-MINT leads to a very similar performance as the intrusively determined optimal regularization parameter, making regularized P-MINT a robust, perceptually advantageous, and practically applicable multichannel equalization technique for speech dereverberation.**

***Index Terms*—Acoustic multichannel equalization, automatic regularization, speech dereverberation.**

Manuscript received August 24, 2012; revised December 21, 2012; accepted April 14, 2013. Date of publication April 30, 2013; date of current version July 11, 2013. This work was supported in part by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development and the Cluster of Excellence 1077 "Hearing4All," funded by the German Research Foundation (DFG). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Nobutaka Ono.

I. Kodrasi is with the Department of Medical Physics and Acoustics, Signal Processing Group, University of Oldenburg, D-26111 Oldenburg, Germany (e-mail: ina.kodrasi@uni-oldenburg.de).

S. Goetze is with the Fraunhofer Institute Digital Media Technology (IDMT)—Project group, Hearing, Speech, and Audio, 26129 Oldenburg, Germany (e-mail: s.goetze@idmt.fraunhofer.de).

S. Doclo is with the Department of Medical Physics and Acoustics, Signal Processing Group, University of Oldenburg, D-26111 Oldenburg, Germany, and also with the Fraunhofer Institute Digital Media Technology (IDMT)—Project group, Hearing, Speech, and Audio, 26129 Oldenburg, Germany (e-mail: simon.doclo@uni-oldenburg.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2013.2260743

I. Introduction

Speech signals recorded in an enclosed space by microphones placed at a distance from the source are often corrupted by reverberation, which arises from the superposition of delayed and attenuated copies of the anechoic speech signal. Reverberation causes signal degradation, typically leading to decreased speech intelligibility [1], [2] and performance deterioration in speech recognition systems [3]–[5]. Hence, many speech communication applications such as teleconferencing applications, voice-controlled systems, or hearing aids, require effective dereverberation algorithms [4]–[6].

In the last decades, several dereverberation approaches have been developed, which can be broadly classified into speech enhancement and acoustic channel equalization approaches [7]. While both single and multichannel dereverberation techniques have been investigated, multichannel techniques are generally preferred since they enable the use of both spectro-temporal and spatial processing of the received microphone signals. Well-known multichannel speech enhancement techniques for dereverberation are either based on spectral subtraction [8], [9] or on linear prediction [10]–[12]. Furthermore, acoustic multichannel equalization techniques [13]–[19] aim to reshape the estimated room impulse responses (RIRs) between the source and the microphone array. Such techniques comprise an attractive approach to speech dereverberation since in theory perfect channel equalization can be achieved [13], [20].

A widely known multichannel equalization technique that aims at complete equalization is the multiple-input/output inverse theorem (MINT) [13], which however suffers from several drawbacks in practice. Since the estimated RIRs typically differ from the true RIRs due to fluctuations (e.g., temperature or position variations [21]) or estimation errors (e.g., due to the sensitivity of blind system identification (BSI) methods to near-common zeros [22] or interfering noise [23]), MINT fails to equalize the true RIRs, possibly leading to severe distortions in the output signal. In an attempt to increase the robustness of MINT, it has been proposed to incorporate regularization in order to decrease the energy of the equalization filters [15].

In addition, more robust partial multichannel equalization techniques such as relaxed multichannel least-squares (RMCLS) [17] and channel shortening (CS) [14] have recently been proposed. Since early reflections tend to improve speech intelligibility [24], [25] and late reverberation (typically defined as the part of the RIR after 50–80 ms) is the major cause of speech intelligibility degradation, the objective of such techniques is to shorten the RIR by suppressing only the reverberant tail. It has been experimentally validated that partial equalization techniques lead to a significant increase in robustness in the presence of RIR estimation errors as compared to complete equalization [17]. However, by not imposing any constraints on the remaining early reflections of the shortened RIR, RMCLS and CS may lead to undesired perceptual effects.

In this paper, we first introduce a partial multichannel equalization technique based on MINT (P-MINT), which aims to shorten the RIR and to directly control the perceptual speech quality [19]. Furthermore, since incorporating regularization is also expected to further increase the robustness of partial multichannel equalization techniques, the effectiveness of incorporating regularization in all aforementioned techniques is investigated. To this end, a regularization term proportional to the energy of the reshaping filters is added to the cost functions for P-MINT, RMCLS, and CS. Whereas a closed-form solution exists for minimizing the regularized cost functions for P-MINT and RMCLS, an iterative approach is required for minimizing the regularized cost function for CS.

In general, the optimal regularization parameter yielding the highest perceptual speech quality needs to be determined intrusively (i.e., using a dereverberated reference signal and knowledge of the true RIRs), limiting the practical applicability of the regularized techniques. In this paper, we also propose and extensively investigate an automatic non-intrusive selection procedure for the regularization parameter based on the L-curve [26].

Using simulations with a realistic acoustic system in the presence of estimation errors, it is shown that a significant performance increase is obtained for P-MINT when regularization is incorporated, whereas a smaller improvement is observed for RMCLS and CS. In addition, it is demonstrated that the intrusively regularized P-MINT technique outperforms the intrusively regularized RMCLS and CS techniques, typically leading to the highest robustness and perceptual speech quality. Furthermore, it is shown that the non-intrusively determined regularization parameter yields a nearly optimal perceptual speech quality in regularized P-MINT, making it a robust, perceptually advantageous, and practically applicable multichannel equalization technique for speech dereverberation.

The paper is organized as follows. In Section II the acoustic multichannel equalization problem is introduced as well as several state-of-the-art multichannel equalization techniques for designing reshaping filters. A mathematical relation between the P-MINT solution and the multiple possible CS solutions is provided, showing that the P-MINT solution can be expressed as a linear combination of the CS solutions. Furthermore, the incorporation of a regularization term in all multichannel equalization techniques is discussed in Section III, whereas in Section IV an automatic non-intrusive procedure for computing the regularization parameter is proposed. Using simulations, the reverberant tail suppression and the perceptual speech quality of all considered equalization techniques is extensively compared in Section V.

II. Acoustic Multichannel Equalization

In this section, complete and partial acoustic multichannel equalization techniques are discussed. First, the general problem is stated and some notational conventions are given. Then the cost functions of several multichannel equalization techniques are discussed.

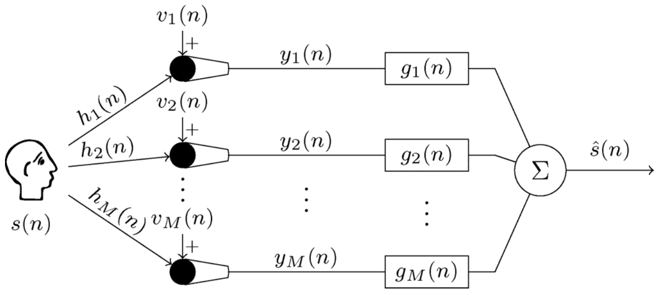


Fig. 1. Multichannel equalization system.

A. Problem Formulation and Notation

Consider an acoustic system with a single speech source and M microphones as depicted in Fig. 1. The m-th microphone signal, $m = 1, \ldots, M$, at time index n is given by

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

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

The RIR can be described in vector notation as $\mathbf{h}_m = [h_m(0) \quad h_m(1) \quad \ldots \quad h_m(L_h - 1)]^T$, with L_h being the RIR length and $[\cdot]^T$ denoting the transpose operation. Given reshaping filters $\mathbf{g}_m$ of length L_g, i.e., $\mathbf{g}_m = [g_m(0) \quad g_m(1) \quad \ldots \quad g_m(L_g - 1)]^T$, the output signal $\hat{s}(n)$ of the multichannel equalization system is given by the sum of the filtered microphone signals, i.e.,

$$\hat{s}(n) = \sum_{m=1}^{M} x_m(n) * g_m(n) = s(n) * \underbrace{\sum_{m=1}^{M} h_m(n) * g_m(n)}_{c(n)}, \quad (2)$$

where $c(n)$ is the equalized impulse response (EIR) between the source and the output of the system. The EIR can be described in vector notation as $\mathbf{c} = [c(0) \quad c(1) \quad \ldots \quad c(L_c - 1)]^T$, with $L_c = L_h + L_g - 1$ being the EIR length. Using the ML_g–dimensional stacked filter vector $\mathbf{g}$, i.e.,

$$\mathbf{g} = [\mathbf{g}_1^T \quad \mathbf{g}_2^T \quad \ldots \quad \mathbf{g}_M^T]^T, \quad (3)$$

and the $L_c \times ML_g$–dimensional multichannel convolution matrix $\mathbf{H}$, i.e.,

$$\mathbf{H} = [\mathbf{H}_1 \quad \mathbf{H}_2 \quad \ldots \quad \mathbf{H}_M], \quad (4)$$

with

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \dots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h-1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h-1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_m(L_h-1) \end{bmatrix}, \tag{5}$$

the output signal can be expressed as

$$\hat{s}(n) = \sum_{m=1}^{M} \mathbf{g}_m^T \mathbf{H}_m^T \underbrace{\begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s(n-L_c+1) \end{bmatrix}}_{\mathbf{s}(n)} \tag{6}$$

$$= \mathbf{g}^T \mathbf{H}^T \mathbf{s}(n) = \mathbf{c}^T \mathbf{s}(n). \tag{7}$$

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

$$\boxed{\mathbf{c} = \mathbf{H}\mathbf{g}} \tag{8}$$

Since the true RIRs are typically not available in practice, acoustic multichannel equalization techniques design the reshaping filter $\mathbf{g}$ using the estimated multichannel convolution matrix $\hat{\mathbf{H}}$ constructed from the estimated RIRs

$$\hat{h}_m(n) = h_m(n) + e_m(n), \tag{9}$$

with $e_m(n)$ representing the estimation error. The sensitivity of several multichannel equalization techniques to these estimation errors will be investigated in Section V.

B. Complete Multichannel Equalization

The objective of complete multichannel equalization techniques such as MINT is to invert the acoustic system up to a delay, such that the output of the system is a shifted version of the clean speech signal.

Multiple-input/output inverse theorem [13]: MINT aims to recover the delayed anechoic speech signal by designing a filter $\mathbf{g}$ such that

$$\hat{\mathbf{H}}\mathbf{g} = \mathbf{d}, \tag{10}$$

where $\mathbf{d}$ is the desired EIR defined as a delayed impulse, i.e.,

$$\mathbf{d} = [\underbrace{0 \, \dots \, 0}_{\tau} \; 1 \, 0 \, \dots \, 0]^T, \tag{11}$$

with τ being the delay in number of samples. The inverse filter is then computed by minimizing the least-squares cost function

$$\boxed{J_{\text{MINT}}(\mathbf{g}) = \|\hat{\mathbf{H}}\mathbf{g} - \mathbf{d}\|_2^2} \tag{12}$$

As shown in [13], assuming that

- the estimated RIRs do not share any common zeros in the z-plane, and
- $L_g \geq \lceil (L_h - 1)/(M - 1) \rceil$,

where $\lceil \cdot \rceil$ denotes the ceiling function, the filter that inverts the multichannel acoustic system can be computed as

$$\boxed{\mathbf{g}_{\text{MINT}} = \hat{\mathbf{H}}^+ \mathbf{d}} \tag{13}$$

with $\{\cdot\}^+$ denoting the Moore-Penrose pseudo-inverse. Since the estimated convolution matrix $\hat{\mathbf{H}}$ is assumed to be a full row-rank matrix [27], its pseudo-inverse can be computed as $\hat{\mathbf{H}}^+ = \hat{\mathbf{H}}^T (\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^{-1}$.

When the RIRs are perfectly estimated, MINT achieves perfect equalization. However, when the estimated RIRs differ from the true RIRs, the resulting EIR $\mathbf{c} = \mathbf{H}\hat{\mathbf{H}}^+\mathbf{d}$ not only differs from the desired response $\mathbf{d}$, but usually causes large distortions in the output signal [15], [21].

C. Partial Multichannel Equalization

Whereas MINT is very sensitive to estimation errors, partial multichannel equalization techniques which aim at reshaping the EIR instead of complete equalization, are significantly more robust. The recently proposed partial multichannel equalization techniques such as RMCLS and CS aim at suppressing the late reverberation only, while imposing no constraints on the early reflections, which may lead to undesired perceptual effects. Therefore we also introduce a partial multichannel equalization technique based on MINT, which aims at simultaneously suppressing the reverberant tail as well as directly controlling the perceptual speech quality of the output signal.

Relaxed multichannel least-squares [17]: RMCLS achieves partial equalization by introducing a weighting vector $\mathbf{w}$ in the least-squares cost function in (12), i.e., the RMCLS cost function is defined as

$$\boxed{J_{\text{RMCLS}}(\mathbf{g}) = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2} \tag{14}$$

with $\mathbf{W} = \text{diag}\{\mathbf{w}\}$, and the weighting vector $\mathbf{w}$ equal to

$$\mathbf{w} = [\underbrace{1 \, \dots \, 1}_{\tau} \; \underbrace{1 0 \dots 0}_{L_s} 1 \, \dots \, 1]^T, \tag{15}$$

where L_s denotes the length of the direct path and early reflections in number of samples. The minimization of (14) aims at setting the reverberant tail of the EIR to $\mathbf{0}$, while the first taps corresponding to the early reflections are not constrained. Similarly to the MINT solution in (13), the reshaping filter minimizing the RMCLS cost function in (14) can be computed as

$$\boxed{\mathbf{g}_{\text{RMCLS}} = (\mathbf{W}\hat{\mathbf{H}})^+ \mathbf{W}\mathbf{d}} \tag{16}$$

Channel shortening [14]: CS has been extensively investigated in the context of digital communication applications [28] and has recently been applied to acoustic system equalization in [14], [17]. CS is achieved by maximizing the energy in the first L_s taps of the EIR (i.e., direct path and early reflections), while minimizing the energy in the remaining taps (i.e., reverberant tail). This optimization problem is expressed as the maximization of a generalized Rayleigh quotient, i.e.,

$$\boxed{J_{\mathrm{CS}}(\mathbf{g}) = \frac{\|\mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{H}}\mathbf{g}\|_2^2}{\|\mathrm{diag}\{\mathbf{w}_u\}\hat{\mathbf{H}}\mathbf{g}\|_2^2} = \frac{\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g}}{\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g}}} \tag{17}$$

where $\mathbf{w}_d$ and $\mathbf{w}_u$ represent the desired and undesired window respectively, i.e.,

$$\mathbf{w}_d = [\underbrace{0\ \dots\ 0}_{\tau}\ \underbrace{1\ \dots\ 1}_{L_s}\ 0\ \dots\ 0]^T \tag{18}$$

$$\mathbf{w}_u = [\underbrace{1\ \dots\ 1}_{\tau}\ \underbrace{0\ \dots\ 0}_{L_s}\ 1\ \dots\ 1]^T = \mathbf{1} - \mathbf{w}_d, \tag{19}$$

and

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T\mathrm{diag}\{\mathbf{w}_d\}^T\mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{H}} \tag{20}$$

$$\hat{\mathbf{A}} = \hat{\mathbf{H}}^T\mathrm{diag}\{\mathbf{w}_u\}^T\mathrm{diag}\{\mathbf{w}_u\}\hat{\mathbf{H}}. \tag{21}$$

Maximizing (17) is equivalent to solving the generalized eigenvalue problem $\hat{\mathbf{B}}\mathbf{g} = \lambda\hat{\mathbf{A}}\mathbf{g}$, where the optimal reshaping filter $\mathbf{g}_{\mathrm{CS}}$ is the generalized eigenvector corresponding to the largest generalized eigenvalue $\lambda_{\max}$, i.e.,

$$\boxed{\hat{\mathbf{B}}\mathbf{g}_{\mathrm{CS}} = \lambda_{\max}\hat{\mathbf{A}}\mathbf{g}_{\mathrm{CS}}} \tag{22}$$

Designing the reshaping filter using such an energy-based optimization technique however imposes no other, e.g., perceptually relevant, constraints on the remaining filter taps of the EIR, which may lead to undesired perceptual effects (cf. Section V-B). Furthermore, multiple solutions to (22) exist (cf. Section II-D), and each of these solutions will lead to a perceptually different EIR. In [17] it has been proposed to select the generalized eigenvector leading to the minimum l_2-norm estimated EIR. In this paper, the intrusively selected generalized eigenvector leading to the highest perceptual speech quality has been used (cf. Section V-A).

Partial multichannel equalization based on MINT [19]: In order to directly control the perceptual quality of the output signal, we recently proposed the P-MINT technique, where the direct path and early reflections of the EIR are controlled by using the first part of one of the estimated RIRs as the desired EIR in (10), i.e.,

$$\hat{\mathbf{H}}\mathbf{g} = \hat{\mathbf{h}}_p^d, \tag{23}$$

where

$$\hat{\mathbf{h}}_p^d = [\underbrace{0\dots0}_{\tau}\underbrace{\hat{h}_p(0)\dots\hat{h}_p(L_s-1)}_{L_s}0\dots0]^T, \tag{24}$$

with $p \in \{1, 2, \dots, M\}$. Without loss of generality, also other desired EIRs could be used instead of (24), as long as they are perceptually close to the true RIRs. The least-squares cost function to be minimized in P-MINT is hence defined as

$$\boxed{J_{\mathrm{P-MINT}}(\mathbf{g}) = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2} \tag{25}$$

Assuming that the same conditions as for MINT are satisfied, the reshaping filter minimizing (25) can be computed as

$$\boxed{\mathbf{g}_{\mathrm{P-MINT}} = \hat{\mathbf{H}}^+\hat{\mathbf{h}}_p^d} \tag{26}$$

D. Relation Between P-MINT and CS

Following similar arguments as in [17], a mathematical relation between the P-MINT solution and the multiple possible CS solutions can be derived.

The maximization of the CS cost function in (17) can be reformulated as computing a filter $\mathbf{g}$ belonging to the null space of $\hat{\mathbf{A}}$ but not belonging to the null space of $\hat{\mathbf{B}} + \hat{\mathbf{A}}$, i.e., satisfying the system of equations

$$\begin{cases} \mathbf{g}^T(\hat{\mathbf{B}} + \hat{\mathbf{A}})\mathbf{g} \neq 0 \\ \mathbf{g}^T\hat{\mathbf{A}}\mathbf{g} = 0, \end{cases} \tag{27}$$

with $\hat{\mathbf{B}} + \hat{\mathbf{A}} = \hat{\mathbf{H}}^T\hat{\mathbf{H}}$. Since the convolution matrix $\hat{\mathbf{H}}$ is assumed to be a full row-rank matrix with $\mathrm{rank}(\hat{\mathbf{H}}) = L_c$, also $\mathrm{rank}(\hat{\mathbf{B}} + \hat{\mathbf{A}}) = L_c$. Exploiting the relationship between the rank and the dimension of the null space of a matrix [29], the dimension of the null space of $\hat{\mathbf{B}} + \hat{\mathbf{A}}$ is equal to

$$\dim[\mathrm{Nullspace}(\hat{\mathbf{B}} + \hat{\mathbf{A}})] = ML_g - L_c, \tag{28}$$

where $\dim[\cdot]$ denotes the dimension of the considered space. In addition, since $\mathrm{rank}(\hat{\mathbf{A}}) = \mathrm{rank}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}) - L_s = L_c - L_s$, the dimension of the null space of $\hat{\mathbf{A}}$ is equal to

$$\dim[\mathrm{Nullspace}(\hat{\mathbf{A}})] = ML_g - (L_c - L_s). \tag{29}$$

Hence, the number of linearly independent vectors satisfying (27) and therefore maximizing the generalized Rayleigh quotient in (17) is $[ML_g - (L_c - L_s)] - [ML_g - L_c] = L_s$.

In order to derive a mathematical relation between the P-MINT solution and the multiple possible CS solutions, consider that the desired EIR in P-MINT can be expressed as

$$\hat{\mathbf{h}}_p^d = \mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{h}}_p, \tag{30}$$

with $\mathbf{w}_d$ defined in (18). For the filter $\mathbf{g}_{\mathrm{P-MINT}}$ in (26), the denominator of the Rayleigh quotient in (17) is equal to

$$\mathbf{g}_{\mathrm{P-MINT}}^T\hat{\mathbf{A}}\mathbf{g}_{\mathrm{P-MINT}} = \|\mathrm{diag}\{\mathbf{w}_u\}\hat{\mathbf{H}}\hat{\mathbf{H}}^+\hat{\mathbf{h}}_p^d\|_2^2 \tag{31}$$

$$= \|\mathrm{diag}\{\mathbf{w}_u\}\mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{h}}_p\|_2^2 = 0, \tag{32}$$

whereas the nominator in (17) is equal to

$$\mathbf{g}_{\mathrm{P-MINT}}^T\hat{\mathbf{B}}\mathbf{g}_{\mathrm{P-MINT}} = \|\mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{H}}\hat{\mathbf{H}}^+\hat{\mathbf{h}}_p^d\|_2^2 \tag{33}$$

$$= \|\mathrm{diag}\{\mathbf{w}_d\}\hat{\mathbf{h}}_p^d\|_2^2 \neq 0. \tag{34}$$

Therefore since the P-MINT filter satisfies (27), it is also in the solution space of the CS optimization problem. As a result, the P-MINT reshaping filter can be expressed as a linear combination of the L_s generalized eigenvectors maximizing the generalized Rayleigh quotient in (17).

III. Regularization in Acoustic Multichannel Equalization

As previously mentioned, the estimated RIRs $\hat{h}_m(n)$ generally differ from the true RIRs (cf. (9)). Since the reshaping filters $\mathbf{g}_m$ are designed using the estimated RIRs, the output signal of the multichannel equalization system is given by

$$\hat{s}(n) = s(n) * \sum_{m=1}^{M} h_m(n) * g_m(n) + \sum_{m=1}^{M} v_m(n) * g_m(n) \tag{35}$$

$$= s(n) * \sum_{m=1}^{M} \left[\hat{h}_m(n) - e_m(n)\right] * g_m(n) + \sum_{m=1}^{M} v_m(n) * g_m(n) \tag{36}$$

$$= s(n) * \sum_{m=1}^{M} \hat{h}_m(n) * g_m(n) \tag{37}$$

$$- s(n) * \sum_{m=1}^{M} e_m(n) * g_m(n) + \sum_{m=1}^{M} v_m(n) * g_m(n), \tag{38}$$

where the term in (37) represents the clean speech signal convolved with the desired EIR and the remaining terms in (38) may (and typically do) give rise to large signal distortions due to RIR estimation errors and the additive noise. However, if the energy of the filters $g_m(n)$ is small, then the value of these distortion terms is also small. To increase the robustness of MINT, it has therefore been proposed to add a regularization term

$$J_{\text{reg}} = \delta \|\mathbf{g}\|_2^2 \tag{39}$$

to the cost function in (12), with the aim of decreasing the energy of the filter $\mathbf{g}$. The regularization parameter δ controls the weight given to the minimization of the energy of the inverse filter. In this paper, we will investigate the effectiveness of incorporating the regularization term J_{reg} in all partial multichannel equalization techniques discussed in Section II. Moreover, in Section IV the computation of the regularization parameter δ is discussed, where both an optimal intrusive computation procedure as well as an automatic non-intrusive procedure is proposed. As previously mentioned, acoustic multichannel equalization techniques generally design reshaping filters disregarding the presence of noise, hence in the following it is again assumed that $v_m(n) = 0$.

Regularized MINT [15]: In the regularized MINT technique, the least-squares cost function in (12) is extended to

$$\boxed{J^{\text{R}}_{\text{MINT}}(\mathbf{g}) = \|\hat{\mathbf{H}}\mathbf{g} - \mathbf{d}\|_2^2 + \delta\|\mathbf{g}\|_2^2} \tag{40}$$

such that the regularized MINT filter minimizing this cost function is equal to

$$\boxed{\mathbf{g}^{\text{R}}_{\text{MINT}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\mathbf{d}} \tag{41}$$

with $\mathbf{I}$ being the $ML_g \times ML_g$-dimensional identity matrix. In [15] it has been shown that incorporating regularization in MINT is useful in reducing the distortions in the output signal due to fluctuations of the RIRs.

Regularized RMCLS: Since RMCLS is a least-squares technique, incorporating the regularization term J_{reg} can be done similarly as for MINT. The regularized RMCLS cost function to be minimized is defined as

$$\boxed{J^{\text{R}}_{\text{RMCLS}}(\mathbf{g}) = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2 + \delta\|\mathbf{g}\|_2^2} \tag{42}$$

and the regularized RMCLS filter minimizing this cost function can be calculated as

$$\boxed{\mathbf{g}^{\text{R}}_{\text{RMCLS}} = [(\mathbf{W}\hat{\mathbf{H}})^T(\mathbf{W}\hat{\mathbf{H}}) + \delta\mathbf{I}]^{-1}(\mathbf{W}\hat{\mathbf{H}})^T\mathbf{W}\mathbf{d}} \tag{43}$$

Regularized P-MINT: Similarly to the regularized least-squares technique for MINT and RMCLS, the regularized P-MINT cost function is defined as

$$\boxed{J^{\text{R}}_{\text{P-MINT}}(\mathbf{g}) = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2 + \delta\|\mathbf{g}\|_2^2} \tag{44}$$

Minimizing (44) yields the regularized P-MINT filter

$$\boxed{\mathbf{g}^{\text{R}}_{\text{P-MINT}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{h}}_p^d} \tag{45}$$

Regularized CS: In order to incorporate the regularization term J_{reg} in CS, the maximization problem in (17) is first reformulated in terms of a generalized Rayleigh quotient *minimization* problem, such that the regularized CS cost function to be minimized can be defined as

$$\boxed{J^{\text{R}}_{\text{CS}}(\mathbf{g}) = \frac{\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g}}{\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g}} + \delta\|\mathbf{g}\|_2^2} \tag{46}$$

However, since no analytical solution to minimize (46) exists, an iterative optimization technique for minimizing this nonlinear cost function will be used in the following. In order to improve the numerical robustness and the convergence speed of the optimization technique, the gradient

$$\frac{\partial J^{\text{R}}_{\text{CS}}(\mathbf{g})}{\partial \mathbf{g}} = 2\left[\frac{(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})\hat{\mathbf{A}}\mathbf{g} - (\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g})\hat{\mathbf{B}}\mathbf{g}}{(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})^2} + \delta\mathbf{g}\right], \tag{47}$$

and the Hessian

$$\frac{\partial^2 J^{\text{R}}_{\text{CS}}(\mathbf{g})}{\partial^2 \mathbf{g}} = 2\left[\frac{(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})\hat{\mathbf{A}} - (\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g})\hat{\mathbf{B}} + 2(\hat{\mathbf{A}}\mathbf{g}\mathbf{g}^T\hat{\mathbf{B}} - \hat{\mathbf{B}}\mathbf{g}\mathbf{g}^T\hat{\mathbf{A}})}{(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})^2} - 4\frac{[(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})\hat{\mathbf{A}}\mathbf{g} - (\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g})\hat{\mathbf{B}}\mathbf{g}]\mathbf{g}^T\hat{\mathbf{B}}}{(\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g})^3} + \delta\mathbf{I}\right], \tag{48}$$

can be provided.

Since the non-linear cost function in (46) typically contains local minima, it should be noted that this technique is sensitive to the initial vector provided to the numerical optimization algorithm. In an attempt to find the global minimum, we have considered different initial vectors, i.e., one of the generalized eigenvector $\mathbf{g}_{\text{CS}}$ solving (22), the P-MINT solution $\mathbf{g}_{\text{P-MINT}}$ in (26), and the vector $[\,1 \quad 0 \quad \ldots \quad 0\,]^T$. The optimal solution is then selected as the one leading to the highest perceptual speech quality (cf. Section V-A).

IV. Non-Intrusive Selection of the Regularization Parameter

Increasing the regularization parameter δ in all regularized equalization techniques presented in Section III decreases the norm of the reshaping filter $\mathbf{g}$, increasing the robustness to RIR estimation errors. However, increasing this parameter 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 estimation errors.

Obviously, different values of the regularization parameter δ lead to different performance. The optimal value δ_{opt} that yields the highest perceptual speech quality depends on the acoustic system to be equalized, the RIR estimation errors, as well as the equalization technique being used. While in simulations δ_{opt} can be intrusively determined exploiting the known true RIRs (cf. Section V-B), an automatic non-intrusive procedure is required in practice.

For conciseness, the automatic non-intrusive procedure for selecting the regularization parameter in acoustic multichannel equalization techniques is discussed only for the regularized P-MINT technique. However, the procedure proposed here can be extended to any regularized least-squares technique, such as regularized MINT and regularized RMCLS.[1]

Incorporating regularization in P-MINT introduces a trade-off between minimizing the residual energy $\|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2$ and minimizing the filter energy $\|\mathbf{g}\|_2^2$ (cf. (44)). A good regularization parameter should hence incorporate knowledge about both the residual energy and the filter energy, such that both energies are kept small. In order to automatically compute a regularization parameter for regularized least-squares problems, it has been proposed in [26] to use a parametric plot of the solution norm versus the residual norm for several values of δ. This plot always has an L-shape with the corner (i.e., the point of maximum curvature) located exactly where the regularized least-squares solution changes in nature from being dominated by over-regularization to being dominated by under-regularization.

We therefore propose selecting the regularization parameter δ_{auto} in the regularized P-MINT technique as the one corresponding to the corner of the parametric plot of the filter norm $\|\mathbf{g}^{\text{R}}_{\text{P-MINT}}\|_2$ versus the residual norm $\|\hat{\mathbf{H}}\mathbf{g}^{\text{R}}_{\text{P-MINT}} - \hat{\mathbf{h}}_p^d\|_2$. As is experimentally validated in Section V, such a regularization parameter also leads to a nearly optimal perceptual speech quality.

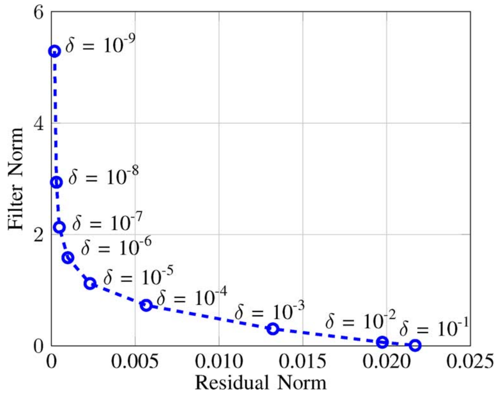


Fig. 2. Typical L-curve obtained using regularized P-MINT for an erroneously estimated acoustic system.

The L-curve can be generated by computing the reshaping filter $\mathbf{g}^{\text{R}}_{\text{P-MINT}}$ in (45) for several values of the regularization parameter δ and then calculating the required norms. However, in order to reduce the computational complexity, it is beneficial to generate the L-curve using the singular value decomposition (SVD) of the estimated convolution matrix $\hat{\mathbf{H}}$. Consider the SVD of $\hat{\mathbf{H}}$, i.e.,

$$\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^T, \tag{49}$$

where $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are orthogonal matrices and $\hat{\mathbf{S}}$ is a diagonal matrix containing the singular values $\hat{\sigma}_k$ of $\hat{\mathbf{H}}$ in descending order, i.e., $\hat{\mathbf{S}} = \operatorname{diag}\{[\,\hat{\sigma}_1 \quad \hat{\sigma}_2 \quad \ldots \quad \hat{\sigma}_{L_c}\,]\}$. Using (45) and (49), the regularized P-MINT filter can be expressed as

$$\mathbf{g}^{\text{R}}_{\text{P-MINT}} = \sum_{k=1}^{L_c} \frac{\hat{\sigma}_k \hat{\mathbf{u}}_k^T \hat{\mathbf{h}}_p^d}{\hat{\sigma}_k^2 + \delta} \hat{\mathbf{v}}_k, \tag{50}$$

where $\hat{\mathbf{u}}_k$ and $\hat{\mathbf{v}}_k$ denote the k-th column of $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ respectively. Hence, for a given δ, the filter norm and the residual norm can be expressed in terms of the singular values/vectors as

$$\boxed{\|\mathbf{g}^{\text{R}}_{\text{P-MINT}}\|_2 = \sqrt{\sum_{k=1}^{L_c} \frac{\hat{\sigma}_k^2 (\hat{\mathbf{u}}_k^T \hat{\mathbf{h}}_p^d)^2}{(\hat{\sigma}_k^2 + \delta)^2}}} \tag{51}$$

$$\boxed{\|\hat{\mathbf{H}}\mathbf{g}^{\text{R}}_{\text{P-MINT}} - \hat{\mathbf{h}}_p^d\|_2 = \sqrt{\sum_{k=1}^{L_c} \frac{\delta^2 (\hat{\mathbf{u}}_k^T \hat{\mathbf{h}}_p^d)^2}{(\hat{\sigma}_k^2 + \delta)^2}}} \tag{52}$$

Therefore, once the SVD is computed, the complete L-curve can be readily generated using (51) and (52).

Fig. 2 depicts a typical L-curve obtained using regularized P-MINT for equalizing an estimated acoustic system (cf. Section V-A). As illustrated in this figure, increasing the value of δ decreases the filter norm but at the same time increases the residual norm. Although from such a curve it seems easy to determine the regularization parameter that corresponds to the maximum curvature, numerical problems due to small singular values may occur and hence, a numerically stable algorithm is required. In this work, the triangle method [30] is used for locating the point of maximum curvature of the L-curve.

[1] The presented approach cannot be used for the regularized CS technique. Automatic selection of the regularization parameter in CS remains a topic for future investigation.

V. Simulations

In this section, simulation results for a scenario with a single speech source and 2 microphones are presented. In Section V-A, the acoustic systems and the used performance measures are introduced. In Section V-B, the performance of all equalization techniques and their regularized counterparts with the intrusively determined regularization parameter δ_{opt} is compared in the presence of channel estimation errors. In Section V-C, the performance of regularized P-MINT when using the automatic non-intrusive procedure for determining the regularization parameter δ_{auto} instead of using δ_{opt} is extensively investigated. Finally, in Section V-D, the performance of P-MINT and automatically regularized P-MINT in the presence of both channel estimation errors and additive noise will be investigated. Sound samples from each simulation can be found at www.sigproc.uni-oldenburg.de/audio/dereverb/pmint.html.

A. Acoustic System and Performance Measures

We have considered an acoustic scenario with a single speech source and $M = 2$ omni-directional microphones placed at a distance of 2.3 m from the source in a room with reverberation time $T_{60} \approx 550$ ms (in Section V-C, also rooms with reverberation times $T_{60} \approx 450$ ms and $T_{60} \approx 750$ ms have been considered). The RIRs between the source and the microphones have been measured using the swept-sine technique [31] and the RIR length has been set to $L_h = 4400$ at a sampling frequency $f_s = 16$ kHz. In order to simulate estimation errors, the measured RIRs have been perturbed by adding scaled white noise as proposed in [32], i.e.,

$$\hat{h}_m(n) = h_m(n) + \underbrace{e(n)h_m(n)}_{e_m(n)} \tag{53}$$

with $e(n)$ being an uncorrelated Gaussian noise sequence with zero mean and an appropriate variance, such that a desired *normalized channel mismatch* E_m, defined as

$$E_m = 10\log_{10}\frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}, \tag{54}$$

is generated. In practice, BSI methods [18], [23] should be used to directly estimate the acoustic system. However, to the best of our knowledge the performance of state-of-the-art BSI methods highly depends on the considered acoustic system and no model has been established to systematically describe the estimation errors that such methods yield. Therefore, (53) and (54) are used to generate the considered estimation errors in the following simulations.

The simulation parameters for all considered multichannel equalization techniques are set to $L_g = 4399$ and $\tau = 0$. Furthermore, 5 different desired window lengths for the partial equalization techniques are investigated, i.e., $L_d \in \{10\,\text{ms}, 20\,\text{ms}, 30\,\text{ms}, 40\,\text{ms}, 50\,\text{ms}\}$, with $L_d = (L_s \times 10^3)/f_s$ being the desired window length in ms. The desired EIR in P-MINT and regularized P-MINT is chosen as the direct path and early reflections of the estimated first RIR, i.e., $\hat{\mathbf{h}}_1^d$.

The performance of all considered equalization techniques is evaluated both in terms of reverberant tail suppression and perceptual speech quality. The reverberant tail suppression is evaluated using the *energy decay curve* (EDC) [7] of the EIR defined as

$$\text{EDC}(n) = 10\log_{10}\frac{1}{\|\mathbf{c}\|_2^2}\sum_{i=n}^{L_c-1} c^2(i), \quad n = 0, \dots, L_c - 1, \tag{55}$$

where $\mathbf{c} = \mathbf{H}\mathbf{g}$ and the reshaping filter $\mathbf{g}$ is designed using the estimated RIRs $\hat{\mathbf{h}}_m$.

The perceptual speech quality of the output signal $\hat{s}(n)$ is evaluated using the objective speech quality measure PESQ [33], which generates a similarity score between the output signal and a reference signal in the range of 1 to 4.5. It has been shown in [34] that measures relying on auditory models such as PESQ exhibit the highest correlation with subjective listening tests when evaluating the quality of dereverberated speech. The reference signal employed here is $s(n) * h_1^d(n)$, i.e., the clean speech signal convolved with the first part of the true first RIR (which is different for each value of the desired window length L_d). It should be noted that with increasing L_d, the reference signal becomes more similar to the unprocessed microphone signal.

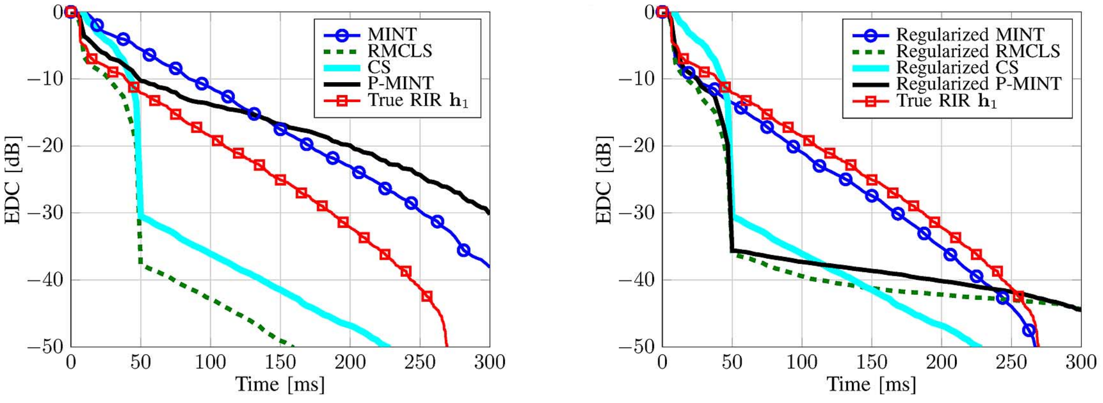
As already mentioned in Section II-C, for the CS technique multiple possible solutions exist. Out of these solutions, we have intrusively selected the generalized eigenvector leading to the highest PESQ score.

Furthermore, in order to evaluate the effectiveness of incorporating regularization in all multichannel equalization techniques, a set of regularization parameters have been considered, i.e., $\delta \in \{0, 10^{-9}, 10^{-8}, \dots, 10^{-1}\}$, and the optimal parameter δ_{opt} is selected as the one leading to the highest perceptual speech quality, i.e., PESQ score. It should be noted that the computation of the PESQ score for selecting the optimal regularization parameter is an intrusive procedure that is not applicable in practice, since knowledge of the true RIRs is required in order to compute the reference signal $s(n) * h_1^d(n)$ and the true EIR $\mathbf{c} = \mathbf{H}\mathbf{g}$. However, with the aim of illustrating the full potential of incorporating regularization in acoustic multichannel equalization techniques, the results presented in Section V-B are generated using such an optimal regularization parameter, whereas in Section V-C the performance when using the automatic non-intrusive procedure for the selection of the regularization parameter will be investigated.

B. Optimal Regularization in the Presence of Channel Estimation Errors

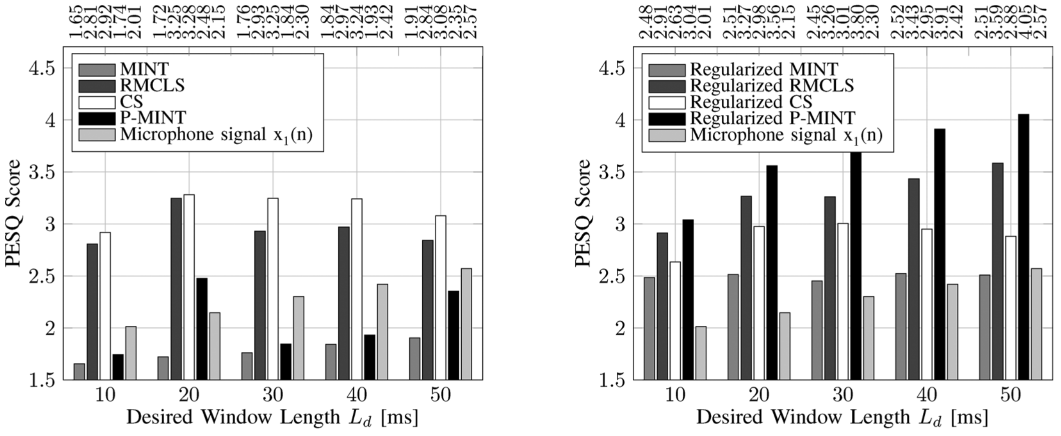
For the sake of clarity and in order to avoid overcrowded plots, these simulations are structured into two parts, with a different normalized channel mismatch in each part. In the first simulation, a moderate mismatch $E_m = -33$ dB is considered, whereas in the second simulation a larger mismatch $E_m = -15$ dB is considered.

Simulation 1 ($E_m = -33$ dB): Fig. 3(a) depicts the EDCs of the EIRs obtained using MINT, RMCLS, CS, and P-MINT for $L_d = 50$ ms. It can be seen that both MINT and P-MINT fail to equalize the acoustic system, leading to an EDC that is higher than the EDC of the true RIR $\mathbf{h}_1$. On the other hand, RMCLS and CS are more robust, with their reverberant tails



(a) (b)

Fig. 3. EDC of the true RIR $\mathbf{h}_1$ and EDC of the EIR obtained using MINT, RMCLS, CS, and P-MINT (a) without regularization and (b) with optimal regularization ($E_m = -33$ dB, $L_d = 50$ ms).

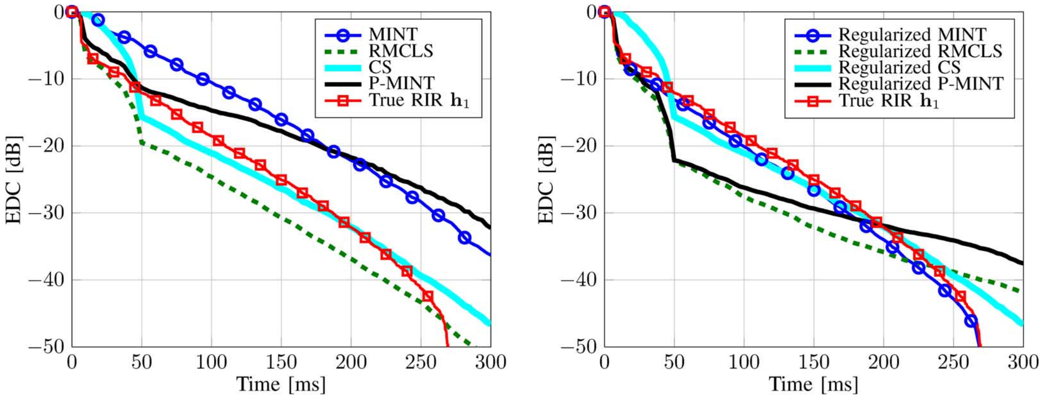


(a) (b)

Fig. 4. PESQ score of the first microphone signal $x_1(n)$ and PESQ score of the system's output $\hat{s}(n)$ obtained for several L_d using MINT, RMCLS, CS, and P-MINT (a) without regularization and (b) with optimal regularization ($E_m = -33$ dB).

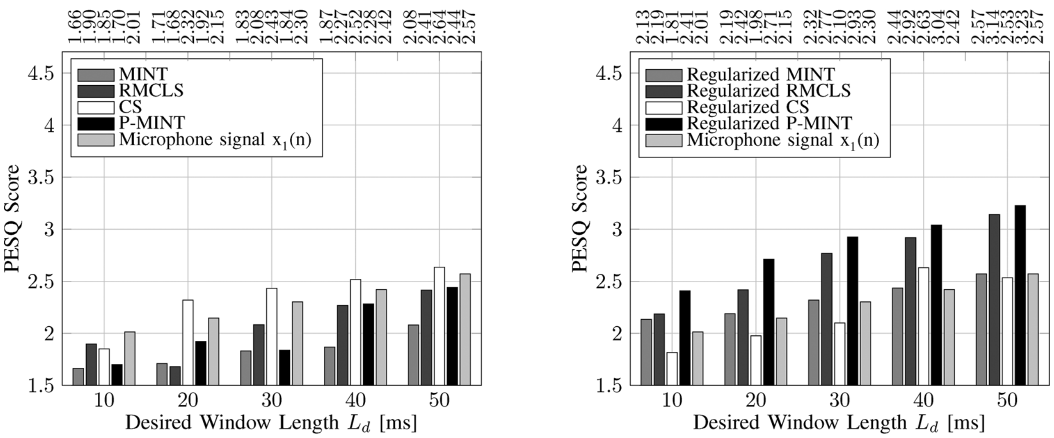
being below −30 dB. In order to evaluate the effectiveness of incorporating regularization in all equalization techniques, Fig. 3(b) depicts the EDCs obtained using regularized MINT, regularized RMCLS, regularized CS, and regularized P-MINT with the optimal intrusively determined regularization parameter δ_{opt}. As illustrated in this figure, the regularized MINT technique still fails to equalize the acoustic system. On the contrary, all regularized partial multichannel equalization techniques are significantly more robust, providing a similar performance in terms of reverberant tail suppression. Comparing Fig. 3(a) and (b), it can be noticed that a significant improvement is obtained when incorporating regularization in P-MINT, while even a slight performance deterioration can be observed for RMCLS. This performance deterioration can be explained by the fact that δ_{opt} is selected such that the PESQ score is maximized, imposing no other constraints on the reverberant tail suppression. Furthermore, the performance of CS does not change when regularization is incorporated, since the exhaustive comparison of the PESQ scores that each regularization parameter yields favors the intrusively selected generalized eigenvector obtained from the CS solution, hence, $\delta_{\text{opt}} = 0$.

Since different EIRs leading to different perceptual speech quality may have very similar EDCs (which is the case for all regularized partial multichannel equalization techniques in this simulation), we have also evaluated the perceptual speech quality using PESQ for different desired window lengths L_d ranging from 10 ms to 50 ms. The PESQ score of the first microphone signal $x_1(n)$ is also computed in order to determine the effectiveness of applying such dereverberation techniques to the system. Fig. 4(a) depicts the PESQ scores obtained for all considered equalization techniques without regularization, whereas Fig. 4(b) depicts the PESQ scores when regularization is incorporated. From Fig. 4(b) it can be seen that the regularized P-MINT technique outperforms all other investigated regularized techniques leading to the highest PESQ score for all considered L_d. Comparing the results in Fig. 4(a) and (b), it is clear that incorporating regularization yields no performance improvement for CS, whereas the performance of MINT, RMCLS, and P-MINT is increased. For a precise numerical comparison, the exact change in the PESQ score for all considered techniques and desired window lengths when incorporating regularization is presented in Table I, with the maximum improvement for each desired window length indicated in bold. Additionally, the average improvement over all considered L_d values is also presented in the last column. As previously observed, regularization is particularly useful for P-MINT, leading to an average PESQ score improvement



(a) (b)

Fig. 5. EDC of the true RIR $\mathbf{h}_1$ and EDC of the EIR obtained using MINT, RMCLS, CS, and P-MINT (a) without regularization and (b) with optimal regularization ($E_m = -15$ dB, $L_d = 50$ ms).



(a) (b)

Fig. 6. PESQ score of the first microphone signal $x_1(n)$ and PESQ score of the system's output $\hat{s}(n)$ obtained for several L_d using MINT, RMCLS, CS, and P-MINT (a) without regularization and (b) with optimal regularization ($E_m = -15$ dB).

TABLE I
PESQ SCORE IMPROVEMENT WHEN INCORPORATING REGULARIZATION IN MINT, RMCLS, CS, AND P-MINT FOR SEVERAL L_d ($E_m = -33$ dB)

L_d [ms]	10	20	30	40	50	Average
MINT	0.83	0.79	0.69	0.68	0.60	0.72
RMCLS	0.11	0.02	0.33	0.47	0.74	0.33
CS	0.00	0.00	0.00	0.00	0.00	0.00
P-MINT	**1.30**	**1.08**	**1.95**	**1.98**	**1.70**	**1.60**

of 1.60. Furthermore, also the regularized MINT and the regularized RMCLS techniques lead to a higher performance as compared to MINT and RMCLS respectively, whereas no improvement is observed in CS.

Simulation 2 ($E_m = -15$ dB): In this simulation, similar analysis as in Simulation 1 will be conducted for the normalized channel mismatch $E_m = -15$ dB. The EDCs obtained for $L_d =$ 50 ms using MINT, RMCLS, CS, and P-MINT without regularization are depicted in Fig. 5(a) whereas the EDCs obtained when regularization is incorporated are depicted in Fig. 5(b). From Fig. 5(a) it can be seen that similarly as in Simulation 1, MINT and P-MINT yield higher EDCs than the EDC of the true RIR $\mathbf{h}_1$, whereas RMCLS and CS are more robust but still fail to entirely suppress the reverberant tail. Furthermore, Fig. 5(b) shows that regularized MINT still fails to equalize the acoustic system as in Simulation 1. However, also the robustness of the regularized partial multichannel equalization techniques decreases with increasing estimation errors, with all techniques achieving a low level of reverberant tail suppression. In order to evaluate the perceptual speech quality, Fig. 6(a) and (b) depict the PESQ scores obtained using the different techniques without and with regularization. As shown in Fig. 6(b), the regularized P-MINT technique again leads to the highest perceptual speech quality as compared to all other investigated techniques for all considered L_d. Therefore even when the reverberant tail suppression is not satisfactory and might lead to audible levels of reverberation, the regularized P-MINT technique still yields the highest perceptual speech quality. Furthermore, comparing Fig. 6(a) and (b) shows that the incorporation of regularization yields a performance increase for MINT, RMCLS, and P-MINT. It can also be noticed that similarly to Simulation 1, the performance of CS does not significantly improve when incorporating regularization. The relative change in the PESQ scores when regularization is incorporated as presented in Table II shows that regularization is again particularly useful for P-MINT, leading to an average improvement in the PESQ score of 0.83.

It should be noted that the regularized CS technique does not outperform the CS technique only if the CS solution is intrusively selected as the generalized eigenvector leading to the

TABLE II
PESQ SCORE IMPROVEMENT WHEN INCORPORATING REGULARIZATION IN MINT, RMCLS, CS, AND P-MINT FOR SEVERAL L_d ($E_m = -15$ dB)

L_d [ms]	10	20	30	40	50	Average
MINT	0.47	0.48	0.49	0.57	0.49	0.50
RMCLS	0.29	0.74	0.68	0.65	0.73	0.62
CS	0.00	0.00	0.00	0.11	0.00	0.02
P-MINT	**0.71**	**0.79**	**1.09**	**0.76**	**0.79**	**0.83**

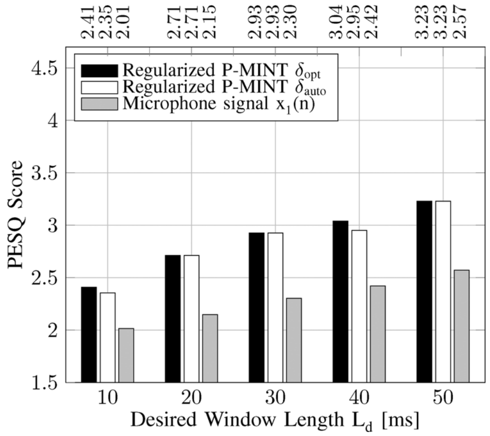


Fig. 7. PESQ score of the first microphone signal $x_1(n)$ and PESQ score of the system's output $\hat{s}(n)$ obtained for several L_d using regularized P-MINT with δ_{opt} and regularized P-MINT with δ_{auto} ($E_m = -15$ dB).

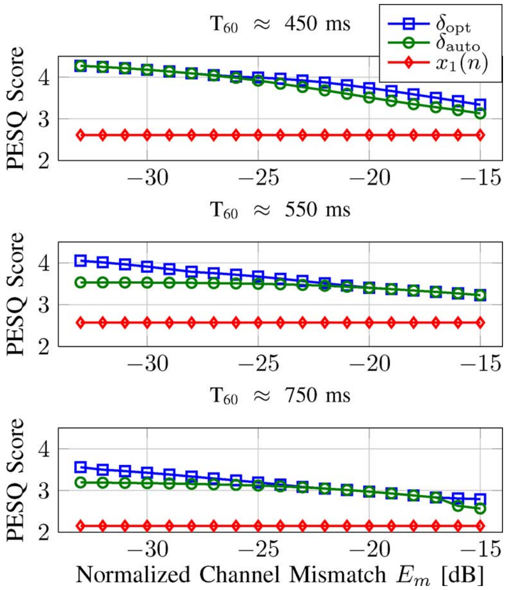


Fig. 8. PESQ score of the first microphone signal $x_1(n)$ and PESQ score of the system's output $\hat{s}(n)$ obtained for different acoustic systems and several E_m using regularized P-MINT with δ_{opt} and regularized P-MINT with δ_{auto} ($L_d = 50$ ms).

highest PESQ score (which is inapplicable in practice). When the eigenvector leading to the minimum l_2-norm estimated EIR is used (as suggested in [17]), regularized CS yields a higher performance than CS. However, due to space constraints, these results are not presented here.

Summarizing the simulation results, we conclude that regularized P-MINT is a robust and perceptually advantageous equalization technique, outperforming all other considered equalization techniques in terms of perceptual speech quality. The large performance improvement obtained for P-MINT when regularization is incorporated can be explained by the significantly higher reverberant tail suppression that is achieved. The remaining advantage that leads to regularized P-MINT outperforming state-of-the-art techniques lies in the direct control of the early reflections.

C. Automatic Regularization in the Presence of Channel Estimation Errors

In this section we will investigate the performance degradation for the regularized P-MINT technique when using the non-intrusive and practically applicable procedure for determining the regularization parameter δ_{auto} (discussed in Section IV) instead of δ_{opt}. In the following, the filter norm $\|\mathbf{g}^{\mathrm{R}}_{\mathrm{P-MINT}}\|_2$ and the residual norm $\|\hat{\mathbf{H}}\mathbf{g}^{\mathrm{R}}_{\mathrm{P-MINT}} - \hat{\mathbf{h}}^d_p\|_2$ are computed using (51) and (52) for the regularization parameters $\delta \in \{10^{-9}, 10^{-8}, \ldots, 10^{-1}\}$. The parametric L-curve is then constructed and the regularization parameter δ_{auto} corresponding to the point of maximum curvature is determined using the triangle method [30]. The PESQ scores obtained using the regularized P-MINT technique with the optimal and automatic regularization parameters for several desired window lengths and $E_m = -15$ dB are depicted in Fig. 7. As illustrated in this figure, the performance when using δ_{auto} is generally similar to the performance obtained when using δ_{opt}. The average performance degradation over all considered L_d is only 0.03, implying that the automatic selection procedure for the regularization parameter provides a nearly optimal performance. Furthermore, the normalized mean square error between the optimal and automatic regularization parameter over all considered L_d is 0.03, where the normalized error is defined as $(\log_{10}\delta_{\mathrm{opt}} - \log_{10}\delta_{\mathrm{auto}})/\log_{10}\delta_{\mathrm{opt}}$.

Since the optimal regularization parameter heavily depends on the channel mismatch and the considered acoustic system, we have also evaluated the performance when using δ_{auto} for different E_m, i.e., $E_m \in \{-33\ \mathrm{dB}, -32\ \mathrm{dB}, \ldots, -15\ \mathrm{dB}\}$, and for different acoustic systems ($T_{60} \approx 450$ ms, $T_{60} \approx 550$ ms, $T_{60} \approx 750$ ms). The desired window length in this simulation is set to $L_d = 50$ ms. Fig. 8 depicts the PESQ scores obtained using the optimal and the automatic selection procedure for the different normalized channel mismatch values and acoustic systems. It can be seen that for all considered acoustic systems and most channel mismatch values, very similar performance is achieved by both regularization parameters, whereas for some scenarios a small performance degradation can be noticed when using δ_{auto}. The average performance degradation over all channel mismatch values when using δ_{auto} is 0.11 ($T_{60} \approx 450$ ms), 0.18 ($T_{60} \approx 550$ ms), and 0.12 ($T_{60} \approx 750$ ms). Furthermore, the normalized mean square error between the optimal and automatic regularization parameter over all channel mismatch values is 0.64 ($T_{60} \approx 450$ ms), 0.05 ($T_{60} \approx 550$ ms), and 0.05 ($T_{60} \approx 750$ ms).

Therefore, the presented results show that the automatic regularization parameter in P-MINT leads to a nearly optimal performance, making regularized P-MINT not only a robust and perceptually advantageous equalization technique, but practically applicable as well.

TABLE III
SNR IMPROVEMENT OBTAINED FOR SEVERAL L_d USING P-MINT AND REGULARIZED P-MINT WITH δ_{auto}

L_d [ms]	10	20	30	40	50
P-MINT	-11.24	-23.11	-25.20	-31.22	-19.59
Reg P-MINT	-3.22	4.39	2.89	0.39	2.14

D. Robustness in the Presence of Channel Estimation Errors and Additive Noise

In this section we will investigate the performance of P-MINT and regularized P-MINT when using the non-intrusive regularization parameter δ_{auto} in the presence of both channel estimation errors and additive noise $v_m(n)$, with $v_m(n)$ consisting of microphone self-noise and diffuse background noise recorded for the same acoustic scenario as in Section V-B. More in particular, we will investigate the possible amplification of this additive noise when using reshaping filters that are designed without taking the presence of noise into account.

The noisy and reverberant signals $y_m(n)$ are generated by convolving the clean speech signal $s(n)$ with the true measured RIRs $h_m(n)$ and adding the recorded noise $v_m(n)$. The input broadband SNR averaged over all channels is $\text{SNR}_{\text{av}} = 15$ dB, with

$$\text{SNR}_{\text{av}} = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\left(\frac{\sum\limits_{n}[x_m(n)]^2}{\sum\limits_{n}[v_m(n)]^2}\right). \quad (56)$$

Furthermore, the considered normalized channel mismatch between the true and estimated RIRs is $E_m = -15$ dB. Hence, the same reshaping filters computed in the presence of only channel estimation errors using P-MINT and regularized P-MINT with δ_{auto} (cf. Sections V-B and V-C respectively) are applied to the received microphone signals $y_m(n)$. The performance at the output of the equalization system is evaluated in terms of the SNR improvement as well as in terms of the perceptual speech quality using PESQ. Table III presents the SNR improvement for different values of L_d for both considered techniques. The presented negative SNR improvement values show that P-MINT significantly amplifies the additive noise. However, when regularization is incorporated, the robustness of P-MINT to additive noise significantly increases, which can be explained by the fact that the energy of the reshaping filters is decreased. When the energy of the reshaping filters is lower, the amplification of the undesired noise term in the output signal is also smaller (cf. (38)). Furthermore, Table IV depicts the obtained PESQ scores, where it can be seen that P-MINT leads to a lower perceptual speech quality than the received microphone signal $y_1(n)$. However, the automatically regularized P-MINT technique yields a higher perceptual speech quality, with a significant improvement over $y_1(n)$.

These simulation results show that regularization is effective in P-MINT not only to increase robustness to channel estimation errors, but also to avoid amplification of the additive noise present at the microphones.

TABLE IV
PESQ SCORE OF THE FIRST MICROPHONE SIGNAL $y_1(n)$ AND PESQ SCORE OF THE SYSTEM'S OUTPUT $\hat{s}(n)$ OBTAINED FOR SEVERAL L_d USING P-MINT AND REGULARIZED P-MINT WITH δ_{auto}

L_d [ms]	10	20	30	40	50
P-MINT	1.65	1.83	1.78	2.03	2.29
Reg P-MINT	2.25	2.60	2.79	2.81	3.07
$y_1(n)$	1.99	2.11	2.27	2.39	2.52

VI. CONCLUSION

In this paper we introduced the partial multichannel equalization technique P-MINT, which aims to suppress the reverberant tail of the RIR as well as to directly control the perceptual speech quality. Furthermore, we presented a robust extension to P-MINT and other recently proposed multichannel equalization techniques, i.e., RMCLS and CS, by incorporating a regularization term in the reshaping filter design. In addition, we proposed an automatic non-intrusive selection procedure for the regularization parameter which leads to a nearly optimal perceptual speech quality.

We have extensively investigated the effectiveness of regularization both in terms of reverberant tail suppression and perceptual speech quality for all considered equalization techniques. Simulation results show that the regularized P-MINT technique with the intrusively determined regularization parameter leads to the highest robustness and perceptual speech quality. Furthermore, simulation results demonstrate that regularization is particularly important for P-MINT, whereas for RMCLS and CS a smaller performance improvement is achieved. Finally, it is shown that the automatic non-intrusive procedure for the selection of the regularization parameter yields a nearly optimal perceptual speech quality in regularized P-MINT.

REFERENCES

[1] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Amer.*, vol. 77, no. 3, pp. 1069–1077, Mar. 1985.

[2] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Amer.*, vol. 120, no. 1, pp. 331–342, Jul. 2006.

[3] M. Omologo, P. Svaizer, and M. Matassoni, "Environmental conditions and acoustic transduction in hands-free speech recognition," *Speech Commun.*, vol. 25, no. 1–3, pp. 75–95, Aug. 1998.

[4] A. Sehr, "Reverberation modeling for robust distant-talking speech recognition," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürenberg, Erlangen, Germany, 2009.

[5] R. Maas, E. A. P. Habets, A. Sehr, and W. Kellermann, "On the application of reverberation suppression to robust speech recognition," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 297–300.

[6] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1732–1745, Sep. 2010.

[7] P. A. Naylor and N. D. Gaubitch*, Speech Dereverberation*, 1st ed. New York, NY, USA: Springer, 2010.

[8] E. A. P. Habets, S. Gannot, I. Cohen, and P. C. W. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1433–1451, Nov. 2008.

[9] E. A. P. Habets and S. Gannot, "Dual-microphone speech dereverberation using a reference signal," in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 901–904.

[10] N. D. Gaubitch, D. B. Ward, and P. A. Naylor, “Statistical analysis of the autoregressive modeling of reverberant speech,” *J. Acoust. Soc. Amer.*, vol. 120, no. 6, pp. 4031–4039, Dec. 2006.

[11] M. Delcroix, T. Hikichi, and M. Miyoshi, “Precise dereverberation using multichannel linear prediction,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 430–440, Feb. 2007.

[12] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and J. Biing-Hwang, “Speech dereverberation based on variance-normalized delayed linear prediction,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1717–1731, Sep. 2010.

[13] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[14] M. Kallinger and A. Mertins, “Multi-channel room impulse response shaping – A study,” in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 101–104.

[15] T. Hikichi, M. Delcroix, and M. Miyoshi, “Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations,” *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.

[16] A. Mertins, T. Mei, and M. Kallinger, “Room impulse response shortening/reshaping with infinity- and p-norm optimization,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

[17] W. Zhang, E. A. P. Habets, and P. A. Naylor, “On the use of channel shortening in multichannel acoustic system equalization,” in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Sep. 2010.

[18] M. A. Haque, T. Islam, and M. K. Hasan, “Robust speech dereverberation based on blind adaptive estimation of acoustic channels,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 775–787, May 2011.

[19] I. Kodrasi and S. Doclo, “Robust partial multichannel equalization techniques for speech dereverberation,” in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 537–540.

[20] H. Hacihabibouglu and Z. Cvetkovic, “Multichannel dereverberation theorems and robustness issues,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 676–689, Feb. 2012.

[21] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, “Equalization in an acoustic reverberant environment: Robustness results,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 311–319, May 2000.

[22] L. Xiang, A. W. H. Khong, and P. A. Naylor, “A forced spectral diversity algorithm for speech dereverberation in the presence of near-common zeros,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 888–899, Mar. 2012.

[23] K. Hasan and P. A. Naylor, “Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: Robustness issue,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Florence, Italy, Sep. 2006.

[24] I. Arweiler and J. M. Buchholz, “The influence of spectral characteristics of early reflections on speech intelligibility,” *J. Acoust. Soc. Amer.*, vol. 130, no. 2, pp. 996–1005, Aug. 2011.

[25] A. Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier, “Effects of spatial and temporal integration of a single early reflection on speech intelligibility,” *J. Acoust. Soc. Amer.*, vol. 133, no. 1, pp. 269–282, Jan. 2013.

[26] P. C. Hansen and D. P. O’Leary, “The use of the L-curve in the regularization of discrete ill-posed problems,” *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.

[27] G. Harikumar and Y. Bresler, “FIR perfect signal reconstruction from multiple convolutions: Minimum deconvolver orders,” *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 215–218, Jan. 1998.

[28] R. K. Martin, K. Vanbleu, M. Ding, G. Ysebaert, M. Milosevic, B. L. Evans, M. Moonen, and C. R. Johnson, “Unification and evaluation of equalization structures and design algorithms for discrete multitone modulation systems,” *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3880–3894, Oct. 2005.

[29] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: John Hopkins Univ. Press, 1996.

[30] J. L. Castellanos, S. Gómez, and V. Guerra, “The triangle method for finding the corner of the L-curve,” *Appl. Numerical Math.*, vol. 43, no. 4, pp. 359–373, Dec. 2002.

[31] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Proc. 108th AES Convent.*, Paris, France, Feb. 2000, pp. 18–22.

[32] J. Cho, D. Morgan, and J. Benesty, “An objective technique for evaluating doubletalk detectors in acoustic echo cancelers,” *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[33] ITU-T, Perceptual evaluation of speech quality (PESQ), An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862 Int. Telecomm. Union (ITU-T) Rec., 2001.

[34] S. Goetze, E. Albertin, J. Rennies, E. A. P. Habets, and K.-D. Kammeyer, “Speech quality assessment for listening-room compensation,” in *Proc. 38th AES Int. Conf. Sound Qual. Eval.*, Pitea, Sweden, Jun. 2010, pp. 11–20.

Ina Kodrasi (S’11) received the Master of Science degree in Communications, Systems and Electronics in 2010 from Jacobs University Bremen, Bremen, Germany. Since 2010 she has been a Ph.D. student at the Signal Processing Group of the University of Oldenburg, Germany. Her research interests are in the area of signal processing for speech and audio applications.

From 2010 to 2011 she was with the Fraunhofer Institute for Digital Media Technology (IDMT), Project group Hearing, Speech and Audio Technology in Oldenburg where she worked on microphone-array beamforming.

Stefan Goetze is head of Audio System Technology for Assistive Systems at the Fraunhofer Institute for Digital Media Technology (IDMT), Project group Hearing, Speech and Audio (HSA) in Oldenburg, Germany. He received his Dipl.-Ing. in 2004 at the University of Bremen, Germany, where he worked as a research engineer from 2004 to 2008. His research interests are assistive technologies, sound pick/up and enhancement, such as noise reduction, acoustic echo cancellation and dereverberation, as well as detection and classification of acoustic events and automatic speech recognition. He is lecturer at the University of Bremen and project leader of national and international projects in the field of ambient assisted living (AAL). He is member of IEEE and AES.

Simon Doclo (S’95–M’03) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003. From 2003 to 2007 he was a Postdoctoral Fellow with the Research Foundation – Flanders at the Electrical Engineering Department (Katholieke Universiteit Leuven) and the Adaptive Systems Laboratory (McMaster University, Canada). From 2007 to 2009 he was a Principal Scientist with NXP Semiconductors at the Sound and Acoustics Group in Leuven, Belgium. Since 2009 he is a full professor at the University of Oldenburg, Germany, and scientific advisor for the project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical applications, more specifically microphone array processing, active noise control, acoustic sensor networks and hearing aid processing. Prof. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with Marc Moonen) and the IEEE Signal Processing Society 2008 Best Paper Award (with Jingdong Chen, Jacob Benesty, Arden Huang). He is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008–2013). He has been secretary of the IEEE Benelux Signal Processing Chapter (1998–2002), and has served as a guest editor for the EURASIP Journal on Applied Signal Processing.