Joint Dereverberation and Noise Reduction Based on Acoustic Multi-Channel Equalization

Ina Kodrasi, *Student Member, IEEE*, and Simon Doclo, *Senior Member, IEEE*

***Abstract*—Regularized acoustic multi-channel equalization techniques, such as regularized partial multi-channel equalization based on the multiple-input/output inverse theorem (RPMINT), are able to achieve a high dereverberation performance in the presence of room impulse response perturbations but may lead to amplification of the additive noise. In this paper, two time-domain techniques aiming at joint dereverberation and noise reduction based on acoustic multi-channel equalization are proposed. The first technique, namely RPMINT for joint dereverberation and noise reduction (RPM-DNR), extends RPMINT by explicitly taking the noise statistics into account. In addition to the regularization parameter used in RPMINT, the RPM-DNR technique introduces an additional weighting parameter, enabling a trade-off between dereverberation and noise reduction. The second technique, namely multi-channel Wiener filter for joint dereverberation and noise reduction (MWF-DNR), takes both the speech and the noise statistics into account and uses the RPMINT filter to compute a dereverberated reference signal for the multi-channel Wiener filter. The MWF-DNR technique also introduces an additional weighting parameter, which now provides a trade-off between speech distortion and noise reduction. To automatically select the regularization and weighting parameters, for the RPM-DNR technique a novel procedure based on the L-hypersurface is proposed, whereas for the MWF-DNR technique two decoupled optimization procedures based on the L-curve are used. Extensive simulations demonstrate using instrumental measures that the RPM-DNR technique maintains the dereverberation performance of the RPMINT technique while improving its noise reduction performance. Furthermore, it is shown that the MWF-DNR technique yields a significantly better noise reduction performance than the RPM-DNR technique at the expense of a worse dereverberation performance.**

***Index Terms*—Acoustic multi-channel equalization, automatic parameter selection, dereverberation, L-hypersurface, noise reduction.**

I. Introduction

In MANY hands-free speech communication applications, such as teleconferencing, voice-controlled systems, or hearing aids, the recorded microphone signals do not only contain the desired speech signal, but also attenuated and delayed copies of the desired speech signal due to reverberation, as well as additive noise. Reverberation and noise cause the recorded signals to sound distant and spectrally distorted, and typically result in a degradation of speech intelligibility and a performance deterioration of automatic speech recognition systems [1], [2]. With the continuously growing demand for high-quality hands-free communication, speech enhancement techniques aiming at joint dereverberation and noise reduction have become indispensable. In the last decade, both single- as well as multi-channel techniques have been proposed, with multi-channel techniques being generally preferred since they enable to exploit both the spectro-temporal and the spatial characteristics of the received microphone signals. Existing dereverberation techniques can be broadly classified into spectral enhancement techniques [3]–[6], probabilistic modeling-based techniques [7]–[9], and acoustic multi-channel equalization techniques [10]–[15].

Spectral enhancement techniques aim to suppress the late reverberation in the spectral domain by estimating the late reverberant power spectral density, e.g., based on an exponentially decaying room impulse response (RIR) model [3]–[5] or a diffuse sound field model [6]. Such techniques have been extended to achieve joint dereverberation and noise reduction typically by using a two-stage approach. A commonly used two-stage approach is based on the decomposition of the multi-channel Wiener filter (MWF) into a minimum variance distortionless response (MVDR) beamformer and a single-channel postfilter [16]. The MVDR beamformer is applied to reduce the noise and some reverberation, whereas the single-channel postfilter is used to suppress the residual noise and reverberation at the MVDR output [17], [18]. In [19] another two-stage beamforming approach to joint dereverberation and noise reduction was proposed, which does not explicitly model and estimate the late reverberant power spectral density but nevertheless relies on the assumption of a diffuse sound field model for the late reverberation. Based on this assumption, in the first stage a superdirective beamformer is applied to generate a dereverberated reference signal, whereas in the second stage the MWF is used to achieve noise reduction.

Probabilistic modeling-based techniques generally model the acoustic transfer function either as an auto-regressive process [7], [8] or using the convolutive transfer function model [9], whereas the clean speech spectral coefficients are modeled using, e.g., a Gaussian [7] or a Laplacian distribution [8]. Dereverberation is then performed by maximum likelihood estimation of all unknown model parameters. In addition, by modeling the noise spectral coefficients, probabilistic

Manuscript received June 02, 2015; revised January 12, 2016; accepted January 12, 2016. Date of publication January 18, 2016; date of current version March 02, 2016. This work was supported in part by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development, in part by the Cluster of Excellence 1077 Hearing4All funded by the German Research Foundation (DFG), and in part by the Marie Curie Initial Training Network DREAMS under Grant 316969. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mads Christensen.

The authors are with the Signal Processing Group, Department of Medical Physics and Acoustics, and Cluster of Excellence Hearing4All, University of Oldenburg, 26111 Oldenburg, Germany (e-mail: ina.kodrasi@uni-oldenurg.de; 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/TASLP.2016.2518804

modeling-based techniques have also been proposed for joint dereverberation and noise reduction [20], [21].

Acoustic multi-channel equalization techniques aim to reshape the available RIRs (measured or estimated) between the speech source and the microphone array. These techniques in principle comprise an attractive approach to speech dereverberation since in theory perfect dereverberation can be achieved [10], [22]. A well-known acoustic multi-channel equalization technique which aims at perfect dereverberation is the multiple-input/output inverse theorem (MINT) technique [10], which however suffers from several drawbacks in practice. Since the available RIRs typically differ from the true RIRs due to fluctuations (e.g., temperature or position variations [23]), due to the sensitivity of blind system identification (BSI) methods to near-common zeros or interfering noise [24]–[26], or due to the sensitivity of supervised system identification (SSI) methods to interfering noise [14], MINT fails to invert the true RIRs, possibly leading to severe distortions in the output signal [13], [14]. In order to increase the robustness against RIR perturbations, partial multi-channel equalization techniques such as channel shortening (CS) [11], relaxed multi-channel least-squares (RMCLS) [14], and partial multi-channel equalization based on MINT (PMINT) [13] have been proposed. Since early reflections tend to improve speech intelligibility [27] and late reverberation is the major cause of speech intelligibility degradation, the objective of these techniques is to relax the constraints on the reshaping filter design by suppressing only the late reverberation. To additionally increase the robustness against RIR perturbations, regularization has been incorporated into these techniques in [13], where the regularization parameter enables a trade-off between dereverberation accuracy and robustness against RIR perturbations. In [13] it has been experimentally validated that the regularized PMINT (RPMINT) technique outperforms the regularized CS and regularized RMCLS techniques in terms of dereverberation performance. However, even though partial acoustic multi-channel equalization techniques are able to achieve a high dereverberation performance in a noiseless scenario, in the presence of additive noise this noise may even be amplified, since the noise statistics are not explicitly taken into account in the reshaping filter design [13], [28].

In this paper, we propose two time-domain techniques to achieve joint dereverberation and noise reduction based on acoustic multi-channel equalization. The first technique, namely RPMINT for joint dereverberation and noise reduction (RPM-DNR), extends RPMINT by explicitly taking the noise statistics into account. In addition to the regularization parameter used in RPMINT, the RPM-DNR technique introduces an additional weighting parameter, enabling a trade-off between dereverberation and noise reduction performance. The second technique, namely MWF for joint dereverberation and noise reduction (MWF-DNR), takes both the speech and the noise statistics into account and uses the RPMINT filter to compute a dereverberated reference signal for the MWF. The reason behind using the RPMINT filter to compute a dereverberated reference signal for the MWF-DNR technique lies in the high dereverberation performance of the RPMINT technique, as has been experimentally validated in [13]. Similarly to the RPM-DNR technique, the MWF-DNR technique also introduces a weighting parameter, now enabling a trade-off between speech distortion and noise reduction, with speech distortion being the deviation of the output speech signal from the dereverberated reference signal. Some preliminary results for the MWF-DNR technique have been presented in [29]. The optimal regularization and weighting parameters yielding the best performance for the proposed RPM-DNR and MWF-DNR techniques can only be determined intrusively, i.e., exploiting knowledge of the true RIRs and the true speech and noise statistics, limiting their practical applicability. In this paper we therefore also propose a novel automatic procedure based on the L-hypersurface [30] for jointly selecting the regularization and weighting parameters in the RPM-DNR technique. To automatically select the regularization and weighting parameters in the MWF-DNR technique it is proposed to use two decoupled optimization procedures based on the L-curve [31]. Extensive simulations for different levels of additive noise, RIR perturbations, and correlation matrix estimation errors show by means of instrumental measures that the RPM-DNR technique can achieve a better noise reduction performance while not degrading the dereverberation performance of the RPMINT technique. Furthermore, it is shown that the MWF-DNR technique yields a significantly better noise reduction performance than the RPM-DNR technique at the expense of a worse dereverberation performance (depending on the correlation matrix estimation errors).

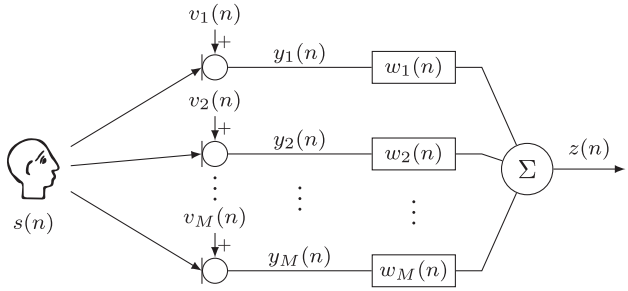

Fig. 1. Acoustic system configuration.

The paper is organized as follows. In Section II the considered acoustic configuration and the used notation is introduced. In Section III state-of-the-art acoustic multi-channel equalization techniques are briefly reviewed. In Section IV two novel time-domain techniques for joint dereverberation and noise reduction based on acoustic multi-channel equalization are proposed, for which automatic procedures for selecting the regularization and weighting parameters are proposed in Section V. The dereverberation and noise reduction performance of all considered techniques is extensively compared in Section VI using instrumental measures.

II. Configuration and Notation

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

$$y_m(n) = \sum_{l=0}^{L_h-1} h_m(l)s(n-l) + v_m(n) \quad (1)$$

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

where $h_m(l), l = 0, \ldots, L_h - 1$, are the coefficients of the L_h-taps long RIR between the speech source and the m-th microphone, $s(n)$ is the clean speech signal, $x_m(n)$ is the reverberant speech component, and $v_m(n)$ is the additive noise component. Denoting the direct path and early reflections of the RIR by $h_{\mathrm{d},m}(l)$ and the late reverberant tail by $h_{\mathrm{r},m}(l)$, the m-th microphone signal in (1) can also be expressed as

$$\begin{aligned} y_m(n) &= \sum_{l=0}^{L_\mathrm{d}-1} h_{\mathrm{d},m}(l)s(n-l) \\ &\quad + \sum_{l=0}^{L_h-L_\mathrm{d}-1} h_{\mathrm{r},m}(l)s(n-l) + v_m(n) \\ &= x_{\mathrm{d},m}(n) + x_{\mathrm{r},m}(n) + v_m(n), \end{aligned} \quad (3)$$

where L_d denotes the length of the direct path and early reflections, $x_{\mathrm{d},m}(n)$ is the m-th direct speech component, and $x_{\mathrm{r},m}(n)$ is the m-th late reverberant speech component. Using the filter-and-sum structure in Fig. 1, the output signal $z(n)$ is equal to the sum of the filtered microphone signals, i.e.,

$$\begin{aligned} z(n) &= \sum_{m=1}^{M} \sum_{l=0}^{L_w-1} w_m(l)x_m(n-l) \\ &\quad + \sum_{m=1}^{M} \sum_{l=0}^{L_w-1} w_m(l)v_m(n-l) \\ &= z_x(n) + z_v(n), \end{aligned} \quad (4)$$

where $w_m(l), l = 0, \ldots, L_w - 1$, are the coefficients of the L_w-taps long filter applied to the m-th microphone, $z_x(n)$ is the output speech component, and $z_v(n)$ is the output noise component. The output speech component can also be written as

$$\begin{aligned} z_x(n) &= \sum_{m=1}^{M} \sum_{l=0}^{L_w-1} w_m(l)x_{\mathrm{d},m}(n-l) \\ &\quad + \sum_{m=1}^{M} \sum_{l=0}^{L_w-1} w_m(n)x_{\mathrm{r},m}(n-l) \\ &= z_\mathrm{d}(n) + z_\mathrm{r}(n), \end{aligned} \quad (5)$$

with $z_\mathrm{d}(n)$ the output direct speech component and $z_\mathrm{r}(n)$ the output late reverberant speech component.

In vector notation, the RIR $\mathbf{h}_m$ and the filter $\mathbf{w}_m$ can be described as

$$\mathbf{h}_m = [h_m(0) \;\; h_m(1) \;\; \ldots \;\; h_m(L_h-1)]^T, \quad (6)$$

$$\mathbf{w}_m = [w_m(0) \;\; w_m(1) \;\; \ldots \;\; w_m(L_w-1)]^T. \quad (7)$$

Using the ML_w-dimensional stacked filter vector $\mathbf{w} = [\mathbf{w}_1^T \;\; \mathbf{w}_2^T \;\; \ldots \;\; \mathbf{w}_M^T]^T$, the equalized impulse response (EIR) vector $\mathbf{c}$ of length $L_c = L_h + L_w - 1$, i.e., $\mathbf{c} = [c(0) \;\; c(1) \;\; \ldots \;\; c(L_c-1)]^T$, can be expressed as

$$\mathbf{c} = \mathbf{H}\mathbf{w}, \quad (8)$$

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

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \ldots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h-1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h-1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & h_m(L_h-1) \end{bmatrix}. \quad (9)$$

Using the ML_w-dimensional stacked vector of the microphone signals $\mathbf{y}(n)$, i.e.,

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n), \quad (10)$$

with

$$\mathbf{y}(n) = [\mathbf{y}_1^T(n) \;\; \mathbf{y}_2^T(n) \;\; \ldots \;\; \mathbf{y}_M^T(n)]^T, \quad (11)$$

$$\mathbf{y}_m(n) = [y_m(n) \;\; y_m(n-1) \;\; \ldots \;\; y_m(n-L_w+1)]^T, \quad (12)$$

and $\mathbf{x}(n)$ and $\mathbf{v}(n)$ similarly defined as in (11) and (12), the output signal can be expressed in vector notation as

$$z(n) = \mathbf{w}^T\mathbf{x}(n) + \mathbf{w}^T\mathbf{v}(n) = (\underbrace{\mathbf{H}\mathbf{w}}_{\mathbf{c}})^T\mathbf{s}(n) + \mathbf{w}^T\mathbf{v}(n), \quad (13)$$

with $\mathbf{s}(n) = [s(n) \;\; s(n-1) \;\; \ldots \;\; s(n-L_c+1)]^T$ and

$$\mathbf{x}(n) = \mathbf{H}^T\mathbf{s}(n). \quad (14)$$

The correlation matrices of $\mathbf{x}(n)$, $\mathbf{v}(n)$, and $\mathbf{y}(n)$ are defined as

$$\mathbf{R}_\mathbf{x} = \mathcal{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}, \quad (15)$$

$$\mathbf{R}_\mathbf{v} = \mathcal{E}\{\mathbf{v}(n)\mathbf{v}^T(n)\}, \quad (16)$$

$$\mathbf{R}_\mathbf{y} = \mathcal{E}\{\mathbf{y}(n)\mathbf{y}^T(n)\}, \quad (17)$$

with $\mathcal{E}$ denoting the expected value operator. Using (14), the reverberant speech correlation matrix $\mathbf{R}_\mathbf{x}$ can be expressed as

$$\mathbf{R}_\mathbf{x} = \mathcal{E}\{\mathbf{H}^T\mathbf{s}(n)\mathbf{s}^T(n)\mathbf{H}\} = \mathbf{H}^T\mathbf{R}_\mathbf{s}\mathbf{H}, \quad (18)$$

with $\mathbf{R}_\mathbf{s}$ being the clean speech correlation matrix. Assuming that the speech and the noise components are uncorrelated, $\mathbf{R}_\mathbf{y} = \mathbf{R}_\mathbf{x} + \mathbf{R}_\mathbf{v}$. For conciseness, the time index n will be omitted when possible in the remainder of this paper.

III. Acoustic Multi-channel Equalization

Acoustic multi-channel equalization techniques aim only at speech dereverberation by designing a reshaping filter $\mathbf{w}$ such that the resulting EIR in (8) equals a target dereverberated EIR $\mathbf{c}_t$, where typically the presence of the additive noise $\mathbf{v}(n)$ is completely disregarded. In practice, only 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 [23]) or due to the sensitivity of BSI and SSI methods to near-common zeros or interfering noise [14], [24]–[26]. Hence, for the reshaping 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. It should be noted that BSI methods result in a convolutive RIR perturbation instead of an additive one [32]. However, the discussion in the remainder of this paper is independent of the type of perturbations, as long as a model can be developed to characterize these perturbations.

In this paper we will focus on the PMINT technique proposed in [13], which aims at suppressing the late reverberation and preserving the perceptual speech quality. To this purpose, the late reverberant taps of the target EIR $\mathbf{c}_t$ are set equal to 0, while the remaining taps are set equal to the direct path and early reflections of one of the available RIRs. Without loss of generality, the RIR of the first microphone, i.e., $\hat{\mathbf{h}}_1$, is used to define the target EIR as

$$\mathbf{c}_t = [\underbrace{0 \dots 0}_{\tau}\ \hat{h}_1(0)\ \dots\ \hat{h}_1(L_{\mathrm{d}}-1)\ 0\ \dots\ 0]^T, \quad (19)$$

with τ a delay introduced to relax the causality constraints on the filter design [33]. The PMINT filter is computed by minimizing the least-squares cost function

$$J_{\mathrm{P}} = \|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2. \quad (20)$$

As shown in [10], assuming that the available RIRs do not share any common zeros and using $L_w \geq \lceil \frac{L_h-1}{M-1} \rceil$, the PMINT filter minimizing the least-squares error in (20) to 0 is equal to

$$\mathbf{w}_{\mathrm{P}} = \hat{\mathbf{H}}^{+}\mathbf{c}_t, \quad (21)$$

where $\{\cdot\}^+$ denotes the matrix pseudoinverse. Since the perturbed convolution matrix $\hat{\mathbf{H}}$ is assumed to be a full row-rank matrix, its pseudoinverse can be computed as $\hat{\mathbf{H}}^+ = \hat{\mathbf{H}}^T(\hat{\mathbf{H}}\hat{\mathbf{H}}^T)^{-1}$.

When the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the PMINT filter yields perfect dereverberation, i.e., $\mathbf{H}\mathbf{w}_{\mathrm{P}} = \mathbf{c}_t$ [13]. However, in the presence of RIR perturbations, applying the PMINT filter to the true convolution matrix yields

$$\mathbf{H}\mathbf{w}_{\mathrm{P}} = \hat{\mathbf{H}}\mathbf{w}_{\mathrm{P}} - \mathbf{E}\mathbf{w}_{\mathrm{P}} = \mathbf{c}_t - \mathbf{E}\mathbf{w}_{\mathrm{P}}. \quad (22)$$

The first term in (22) is the target EIR, whereas the second term represents distortions due to RIR perturbations. In order to increase the robustness of acoustic multi-channel equalization techniques against RIR perturbations, regularized techniques such as the RPMINT technique have been proposed [13]. As shown in [33], when taking the RIR perturbations into account, an optimal reshaping filter in the minimum mean-square error sense can be computed by minimizing the cost function

$$J = \|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2 + \mathbf{w}^T\mathcal{E}\{\mathbf{E}^T\mathbf{E}\}\mathbf{w}, \quad (23)$$

where it is assumed that $\mathcal{E}\{\mathbf{E}\} = \mathbf{0}$. The matrix $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\}$ in (23) obviously depends on the energy and the type of RIR perturbations, e.g., perturbations arising due to microphone position fluctuations [12], [23], or perturbations arising from BSI or SSI methods [24]–[26]. While statistical models can be developed for the correlation structure of different types of perturbations, the exact $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\}$ cannot be known in practice. To account for inaccuracies in modeling $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\}$, regularized acoustic multi-channel equalization techniques introduce a regularization parameter δ and use $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\} = \delta\mathbf{R}_e$, with $\mathbf{R}_e$ constructed based on a perturbation model [12], [26]. When no knowledge about the perturbations is available, they are often assumed to be spatially and temporally white, i.e., $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\} = \delta\mathbf{I}$, with $\mathbf{I}$ denoting the $ML_w \times ML_w$-dimensional identity matrix [13], [33]. This assumption has been used for the experimental results in Section VI.

Using $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\} = \delta\mathbf{R}_e$ in (23), the RPMINT cost function is given by [13]

$$J_{\mathrm{RP}} = \|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2 + \delta\mathbf{w}^T\mathbf{R}_e\mathbf{w} \quad (24)$$

$$= \epsilon_c + \delta\epsilon_e, \quad (25)$$

where ϵ_c denotes the dereverberation error energy and ϵ_e denotes the distortion energy due to RIR perturbations. Clearly, the dereverberation performance, i.e., the deviation of the resulting EIR from the target EIR $\mathbf{c}_t$, depends on both the dereverberation error and distortion energies (cf. (22)), and the regularization parameter δ provides a trade-off between the two. Minimizing (24) yields the RPMINT filter

$$\mathbf{w}_{\mathrm{RP}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{R}_e)^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t, \quad (26)$$

where δ can be automatically computed using the procedure based on the L-curve proposed in [13] (cf. Section V). While the PMINT filter fails to achieve dereverberation in the presence of RIR perturbations, it has been shown in [13] that the RPMINT filter yields a significantly better dereverberation performance, i.e.,

$$\mathbf{w}_{\mathrm{RP}}^T\mathbf{x} \approx \mathbf{c}_t^T\mathbf{s}, \quad (27)$$

outperforming other regularized techniques such as regularized CS and regularized RMCLS. Furthermore, the RPMINT technique is able to partly avoid the noise amplification at the output of the system [13] (cf. Section VI), however, its noise reduction performance is limited since the actual noise statistics are not explicitly taken into account.

IV. Joint Dereverberation and Noise Reduction based on Acoustic Multi-Channel Equalization

Since acoustic multi-channel equalization techniques design reshaping filters for dereverberation without taking the presence of additive noise into account, the output noise power ϵ_v, i.e.,

$$\epsilon_v = \mathcal{E}\{(\mathbf{w}^T\mathbf{v})^2\} = \mathbf{w}^T\mathbf{R}_{\mathbf{v}}\mathbf{w}, \quad (28)$$

is not explicitly controlled and may even be amplified compared to the noise power in the microphone signals. In this section, two time-domain techniques aiming at joint dereverberation and noise reduction based on acoustic multi-channel equalization are proposed, namely RPMINT for joint dereverberation and noise reduction taking the noise statistics into

account (cf. Section IV-A) and MWF for joint dereverberation and noise reduction taking both the speech and the noise statistics into account (cf. Section IV-B).

A. RPMINT for Joint Dereverberation and Noise Reduction (RPM-DNR)

Aiming at controlling the dereverberation error energy ϵ_c, the distortion energy ϵ_e, as well as the output noise power ϵ_v, we propose to extend the RPMINT cost function in (24) by explicitly taking the noise statistics into account. The RPMINT cost function for joint dereverberation and noise reduction (RPM-DNR) can then be written as

$$J_{\text{RDNR}} = J_{\text{RP}} + \mu\epsilon_v \tag{29}$$

$$= \|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2 + \delta\mathbf{w}^T\mathbf{R}_e\mathbf{w} + \mu\mathbf{w}^T\mathbf{R}_{\mathbf{v}}\mathbf{w}, \tag{30}$$

with δ the regularization parameter controlling the weight given to the distortion energy and μ an additional weighting parameter controlling the weight given to the output noise power. The RPM-DNR filter minimizing (30) is equal to

$$\mathbf{w}_{\text{RDNR}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{R}_e + \mu\mathbf{R}_{\mathbf{v}})^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t. \tag{31}$$

As is experimentally validated in Section VI-B, the dereverberation and noise reduction performance of the RPM-DNR filter in (31) depend on the regularization and weighting parameters δ and μ. Increasing the regularization parameter δ results in a higher suppression of the distortion energy at the expense of a higher dereverberation error energy and a larger output noise power. Increasing the weighting parameter μ results in a better noise reduction performance at the expense of a worse dereverberation performance, which simultaneously depends on the dereverberation error and distortion energies. While in simulations the optimal values for the parameters δ and μ can be intrusively determined using knowledge of the true RIRs and of the true noise statistics, in practice an automatic non-intrusive procedure is required. In Section V a novel procedure is proposed for the joint automatic selection of both parameters.

B. MWF for Joint Dereverberation and Noise Reduction (MWF-DNR)

The RPM-DNR technique proposed in Section IV-A aims at joint dereverberation and noise reduction by considering only the perturbed RIRs and the noise statistics. Taking also the reverberant speech statistics into account, we propose a second technique to achieve joint dereverberation and noise reduction by minimizing the mean-square error between the output signal and a dereverberated reference signal s_{ref}, i.e.,

$$J = \mathcal{E}\{(\mathbf{w}^T\mathbf{y} - s_{\text{ref}})^2\}. \tag{32}$$

The cost function in (32) is the well-known MWF cost function [34], where the reference signal now is the dereverberated speech signal. The estimation of several reference signals has been considered for the MWF, e.g., the clean speech signal, the reverberant speech component at an arbitrarily chosen microphone, or a spatially pre-processed reference signal [19], [35], [36]. Considering the high and robust dereverberation performance of the time-domain RPMINT technique (cf. (27)), in this paper we propose to use the RPMINT filter to generate the dereverberated reference signal in (32), i.e., $s_{\text{ref}} = \mathbf{w}_{\text{RP}}^T\mathbf{x} \approx \mathbf{c}_t^T\mathbf{s}$. Assuming that the speech and the noise components are uncorrelated and using a weighting parameter μ to enable a trade-off between speech distortion and noise reduction, the cost function of the proposed MWF for joint dereverberation and noise reduction (MWF-DNR) can be written as

$$J_{\text{MDNR}} = \mathcal{E}\{(\mathbf{w}^T\mathbf{x} - \mathbf{w}_{\text{RP}}^T\mathbf{x})^2\} + \mu\mathcal{E}\{(\mathbf{w}^T\mathbf{v})^2\} \tag{33}$$

$$= \epsilon_x + \mu\epsilon_v, \tag{34}$$

with ϵ_x being the speech distortion, which refers to the deviation of the output speech component from the dereverberated reference signal $\mathbf{w}_{\text{RP}}^T\mathbf{x}$. The MWF-DNR filter minimizing (33) is equal to

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}})^{-1}\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\text{RP}}. \tag{35}$$

Using the RPMINT filter from (26) in (35), the MWF-DNR filter can also be written as

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{R}_{\mathbf{x}} + \mu\mathbf{R}_{\mathbf{v}})^{-1}\mathbf{R}_{\mathbf{x}}(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{R}_e)^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t. \tag{36}$$

As is experimentally validated in Section VI-B, the dereverberation and noise reduction performance of the MWF-DNR filter in (36) depend on the regularization and weighting parameters δ and μ. The regularization parameter δ affects the dereverberation performance of the RPMINT filter $\mathbf{w}_{\text{RP}}$, hence, the dereverberation performance of the MWF-DNR reference signal $\mathbf{w}_{\text{RP}}^T\mathbf{x}$. The weighting parameter μ affects the speech distortion ϵ_x (as a result, the dereverberation performance of the MWF-DNR filter) as well as the noise reduction performance. While in simulations the optimal values for the parameters δ and μ can be intrusively determined using knowledge of the true RIRs and of the true speech and noise statistics, in practice an automatic non-intrusive procedure is required. In Section V we propose to automatically select the regularization and weighting parameters δ and μ using two decoupled optimization procedures based on the L-curve.

C. Insights on the RPM-DNR and MWF-DNR Techniques

The main difference between the RPM-DNR and MWF-DNR filters in (31) and (36) consists in the fact that the MWF-DNR filter uses the reverberant speech correlation matrix $\mathbf{R}_{\mathbf{x}}$, which implicitly depends on the true convolution matrix $\mathbf{H}$ and on the clean speech correlation matrix $\mathbf{R}_{\mathbf{s}}$, cf. (18), whereas the RPM-DNR filter uses only the perturbed convolution matrix $\hat{\mathbf{H}}$. Substituting (18) in (36), the MWF-DNR filter can be written as

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{H}^T\mathbf{R}_{\mathbf{s}}\mathbf{H} + \mu\mathbf{R}_{\mathbf{v}})^{-1}\mathbf{H}^T\mathbf{R}_{\mathbf{s}}\mathbf{H}(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{R}_e)^{-1} \times \hat{\mathbf{H}}^T\mathbf{c}_t. \tag{37}$$

As can be seen in (37), unlike the RPM-DNR filter, the MWF-DNR filter indirectly incorporates knowledge of the true convolution matrix $\mathbf{H}$ and of the clean speech correlation matrix $\mathbf{R}_{\mathbf{s}}$.

In the following, it is shown that only when assuming that i) the clean speech signal is uncorrelated, ii) the true RIRs are available, and iii) the regularization parameter δ approaches 0, i.e., $\delta \to 0$, the RPM-DNR and MWF-DNR filters are equivalent.

First, assuming that the clean speech signal is uncorrelated, i.e., $\mathbf{R_s} = \sigma_s^2\mathbf{I}$, with σ_s^2 the clean speech variance, the MWF-DNR filter is equal to

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{H}^T\mathbf{H} + \frac{\mu}{\sigma_s^2}\mathbf{R_v})^{-1}\mathbf{H}^T\mathbf{H}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}+\delta\mathbf{R}_e)^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t. \tag{38}$$

Hence, even for an uncorrelated clean speech signal (which is generally not the case in practice), the MWF-DNR filter in (38) differs from the RPM-DNR filter in (31) by indirectly incorporating the true convolution matrix $\mathbf{H}$.

Second, assuming that the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the RPM-DNR filter in (31) and the MWF-DNR filter in (38) can be written as

$$\mathbf{w}_{\text{RDNR}} = (\mathbf{H}^T\mathbf{H} + \delta\mathbf{R}_e + \mu\mathbf{R_v})^{-1}\mathbf{H}^T\mathbf{c}_t, \tag{39}$$

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{H}^T\mathbf{H} + \frac{\mu}{\sigma_s^2}\mathbf{R_v})^{-1}\mathbf{H}^T\mathbf{H}(\mathbf{H}^T\mathbf{H}+\delta\mathbf{R}_e)^{-1}\mathbf{H}^T\mathbf{c}_t. \tag{40}$$

Finally, assuming that the regularization parameter δ approaches 0, i.e., $\delta \to 0$, the RPM-DNR filter in (39) and the MWF-DNR filter in (40) can be written as

$$\mathbf{w}_{\text{RDNR}} = (\mathbf{H}^T\mathbf{H} + \mu\mathbf{R_v})^{-1}\mathbf{H}^T\mathbf{c}_t, \tag{41}$$

$$\mathbf{w}_{\text{MDNR}} = (\mathbf{H}^T\mathbf{H} + \frac{\mu}{\sigma_s^2}\mathbf{R_v})^{-1}\mathbf{H}^T\mathbf{c}_t, \tag{42}$$

where (42) is derived from (40) using the fact that $\lim_{\delta\to 0}(\mathbf{H}^T\mathbf{H} + \delta\mathbf{R}_e)^{-1}\mathbf{H}^T\mathbf{c}_t = \mathbf{H}^+\mathbf{c}_t$. Comparing (41) and (42), it can be observed that under the assumptions of an uncorrelated clean speech signal, knowledge of the true RIRs, and $\delta \to 0$, the RPM-DNR and MWF-DNR filters are equivalent (up to the scaling of the weighting parameter μ by the clean speech variance σ_s^2). However, in practice the clean speech signal is not uncorrelated, i.e., $\mathbf{R_s} \neq \sigma_s^2\mathbf{I}$, and most importantly, the true RIRs are not known. As is experimentally validated in Section VI-D, by incorporating the true speech statistics $\mathbf{R_x}$ in the MWF-DNR technique, the noise reduction and the overall joint dereverberation and noise reduction performance can be significantly improved in comparison to the RPM-DNR technique. The importance of incorporating the true reverberant speech correlation matrix $\mathbf{R_x}$ is further validated in Section VI-D by the performance degradation of the MWF-DNR technique in the presence of estimation errors.

V. Automatic Selection of Regularization and Weighting Parameters

The optimal value of the regularization and weighting parameters in the RPM-DNR and MWF-DNR techniques depends on the acoustic system, the RIR perturbations, the additive noise, as well as on what is more important for the considered application, i.e., dereverberation or noise reduction performance. While in simulations these parameters can be determined intrusively, i.e., using knowledge of the true RIRs and of the speech and noise statistics, in practice an automatic non-intrusive procedure is required. In [13] an automatic procedure has been proposed for selecting the regularization parameter in RPMINT. This procedure will be reviewed in Section V-A and further adapted to the automatic selection of the regularization and weighting parameters for the MWF-DNR technique in Section V-C. Furthermore, in Section V-B a novel procedure is proposed for the joint automatic selection of the regularization and weighting parameters in the RPM-DNR technique.

D. Automatic Parameter Selection in RPMINT

As mentioned in Section III, the regularization parameter δ in RPMINT enables a trade-off between the dereverberation error energy ϵ_c and the distortion energy ϵ_e, with

$$\epsilon_c = \|\hat{\mathbf{H}}\mathbf{w}_{\text{RP}} - \mathbf{c}_t\|_2^2, \tag{43}$$

$$\epsilon_e = \mathbf{w}_{\text{RP}}^T\mathbf{R}_e\mathbf{w}_{\text{RP}}. \tag{44}$$

An appropriate regularization parameter should incorporate knowledge about both the dereverberation error energy and the distortion energy, such that both terms are appropriately controlled. In order to automatically compute the regularization parameter in RPMINT, it has been proposed in [13] to use a parametric plot of the distortion energy ϵ_e versus the dereverberation error energy ϵ_c for different values of the regularization parameter δ. Due to the arising trade-off, this parametric plot has an L-shape, with the corner (i.e., the point of maximum curvature) located where the RPMINT filter $\mathbf{w}_{\text{RP}}$ in (26) changes from being dominated by over-regularization to being dominated by under-regularization. It has therefore been proposed in [13] to automatically select the regularization parameter δ in RPMINT as the point of maximum curvature of this L-curve. The curvature κ of the parametric plot of the distortion energy versus the dereverberation error energy can be analytically computed as [31]

$$\kappa = \frac{\epsilon_c'\epsilon_e'' - \epsilon_c''\epsilon_e'}{(\epsilon_c' + \epsilon_e')^{\frac{3}{2}}}, \tag{45}$$

with $\{\cdot\}'$ and $\{\cdot\}''$ denoting the first- and second-order derivatives with respect to δ. The first- and second-order derivatives can also be analytically computed and substituted in (45), such that the curvature κ can be analytically expressed as a function of the regularization parameter δ. In order to determine the unique point of maximum curvature, a one-dimensional optimization procedure can then be used. The analytical derivative and curvature expressions have been omitted in this paper since using an optimization procedure to maximize the curvature not only results in a high computational complexity, but is also prone to numerical errors. Therefore the triangle method proposed in [37], which is a numerically robust geometric procedure, has been used to determine the point of maximum curvature of the L-curve, similarly as in [13]. Experimental results in [13] have shown that this automatic parameter selection procedure yields a very similar robustness against RIR perturbations as intrusively selecting the regularization parameter.

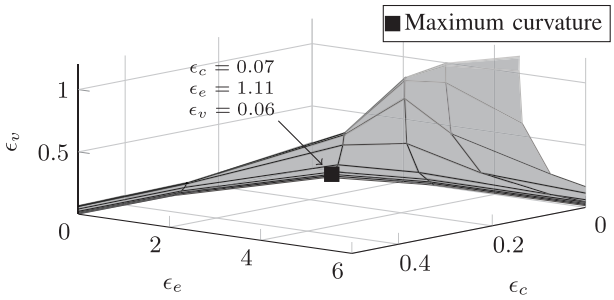

Fig. 2. Exemplary parametric surface of the output noise power ϵ_v versus dereverberation error energy ϵ_c and distortion energy ϵ_e for the RPM-DNR technique.

B. Automatic Parameter Selection in RPM-DNR

Different regularization and weighting parameters δ and μ obviously result in different RPM-DNR filters in (31), which yield different dereverberation error energy ϵ_c, distortion energy ϵ_e, and output noise power ϵ_v, with

$$\epsilon_c = \|\hat{\mathbf{H}}\mathbf{w}_{\mathrm{RDNR}} - \mathbf{c}_t\|_2^2, \tag{46}$$

$$\epsilon_e = \mathbf{w}_{\mathrm{RDNR}}^T \mathbf{R}_e \mathbf{w}_{\mathrm{RDNR}}, \tag{47}$$

$$\epsilon_v = \mathbf{w}_{\mathrm{RDNR}}^T \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\mathrm{RDNR}}. \tag{48}$$

Similarly to the RPMINT technique, appropriate parameters δ and μ should incorporate knowledge about the dereverberation error energy, the distortion energy, and the output noise power, such that all three terms are appropriately controlled. Motivated by the simplicity and the applicability of the L-curve for regularizing least-squares techniques [31], the so-called L-hypersurface has been proposed in [30] as a multi-parameter generalization of the L-curve. Similarly to the L-curve procedure where the optimal parameter is selected as the point of maximum curvature, we propose to select the regularization and weighting parameters δ and μ as the point of maximum Gaussian curvature of the L-hypersurface, obtained by plotting the output noise power ϵ_v versus the dereverberation error energy ϵ_c and the distortion energy ϵ_e for several values of the parameters δ and μ. Fig. 2 depicts an exemplary L-hypersurface obtained by plotting ϵ_v versus ϵ_c and ϵ_e for several regularization and weighting parameters δ and μ for the RPM-DNR technique, with the square denoting the point of maximum Gaussian curvature. Although the Gaussian curvature of a surface can be analytically computed, numerical inaccuracies due to the manipulation of large-dimensional matrices can occur when maximizing it [38], such that a numerically stable algorithm is required. In this paper, the minimum distance method proposed in [38] has been used to compute the point of maximum Gaussian curvature.

C. Automatic Parameter Selection in MWF-DNR

Similarly to the RPM-DNR technique, different regularization parameters δ and μ result in different MWF-DNR filters in (36), which obviously yield different dereverberation error energy ϵ_c, distortion energy ϵ_e, speech distortion ϵ_x, and output noise power ϵ_v. To automatically select the regularization and weighting parameters δ and μ for the MWF-DNR technique, we propose to use two decoupled optimization procedures based on the L-curve. First, in order to obtain a dereverberated reference signal $\mathbf{w}_{\mathrm{RP}}^T\mathbf{x}$, the parameter δ is automatically computed using the L-curve procedure described in Section V-A. For a fixed regularization parameter δ, i.e., a fixed $\mathbf{w}_{\mathrm{RP}}$, changing the parameter μ in the MWF-DNR technique yields a different speech distortion ϵ_x and output noise power ϵ_v, i.e.,

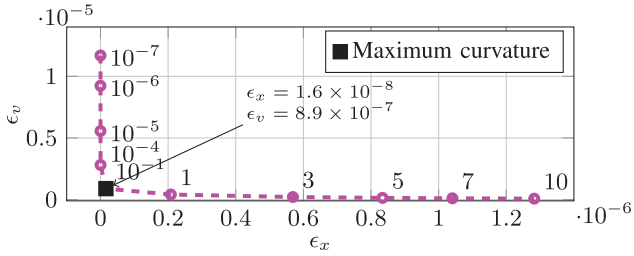

Fig. 3. Exemplary parametric plot of the output noise power ϵ_v versus speech distortion ϵ_x for the MWF-DNR technique. The marked points show the value of μ.

$$\epsilon_x = \mathbf{w}_{\mathrm{MDNR}}^T \mathbf{R}_{\mathbf{x}} \mathbf{w}_{\mathrm{MDNR}} - 2\mathbf{w}_{\mathrm{MDNR}}^T \mathbf{R}_{\mathbf{x}} \mathbf{w}_{\mathrm{RP}} + \mathbf{w}_{\mathrm{RP}}^T \mathbf{R}_{\mathbf{x}} \mathbf{w}_{\mathrm{RP}}, \tag{49}$$

$$\epsilon_v = \mathbf{w}_{\mathrm{MDNR}}^T \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\mathrm{MDNR}}, \tag{50}$$

with (49) derived by expanding $\epsilon_x = \mathcal{E}\{(\mathbf{w}^T\mathbf{x} - \mathbf{w}_{\mathrm{RP}}^T\mathbf{x})^2\}$ from (33). Similarly to before, an appropriate weighting parameter μ should incorporate knowledge about the speech distortion and the output noise power, such that both terms are appropriately controlled. Fig. 3 depicts an exemplary parametric plot of the output noise power versus speech distortion for a set of parameters μ. This parametric plot has an L-shape, with the point of maximum curvature, i.e., the corner of the L-curve, located where the MWF-DNR filter changes from being dominated by large speech distortion to being dominated by large output noise power. Hence, we propose to select the weighting parameter μ in the MWF-DNR technique as the point of maximum curvature of this parametric plot. Although from the exemplary plot in Fig. 3 it may seem straightforward to determine the point of maximum curvature, numerical problems typically occur as described in Section V-A, such that a numerically stable algorithm is required. In this paper, the triangle method proposed in [37] has been used to determine the point of maximum curvature of the L-curve.

VI. Experimental Results

In this section the dereverberation and noise reduction performance of the proposed RPM-DNR and MWF-DNR techniques is investigated using instrumental measures. In Section VI-A the considered acoustic systems, algorithmic settings, and instrumental measures are introduced. In Section VI-B the influence of the regularization and weighting parameters on the performance of the proposed techniques is investigated. In Section VI-C, the automatically parametrized RPM-DNR and MWF-DNR techniques are compared to acoustic multi-channel equalization techniques. In Section VI-D the performance of

the automatically parametrized RPM-DNR and MWF-DNR techniques is extensively investigated for different noise levels, RIR perturbations, and correlation matrix estimation errors using simulated acoustic systems and simulated RIR perturbations. Finally, in Section VI-E the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques is investigated using recorded acoustic systems with different noise levels and RIR perturbations arising from the least-squares SSI [14].

A. Acoustic Systems, Algorithmic Settings, and Performance Measures

1) Simulated Acoustic System and Algorithmic Settings: To be able to directly control the level of RIR perturbations, we have considered a simulated acoustic system with $M = 4$ equidistant microphones and a speech source placed in broadside direction at a distance of 2 m from the microphones. The distance between the microphones is 4 cm and the room reverberation time is $T_{60} \approx 610$ ms [39]. The RIRs between the source and the microphones are measured using the swept-sine technique and the RIR length is $L_h = 5100$ at a sampling frequency $f_s = 8$ kHz. The speech components in the microphone signals are generated by convolving clean speech signals from the HINT database [40] with the measured RIRs. The noise consists of a directional interference and spatially diffuse noise which is simulated using [41]. The directional interference is located in endfire direction at a distance of 2 m from the microphones. The broadband input speech-to-interference ratio (SIR) is varied between -5 dB and 10 dB and the broadband input speech-to-diffuse noise ratio is 10 dB. The speech-plus-noise signal is 18 s long and is preceded by a 14 s long noise-only signal, such that the noise statistics can be estimated during speech absence (cf. (57)). The noise-only signal is not taken into account during evaluation. In order to simulate RIR perturbations, the measured RIRs are perturbed by proportional Gaussian distributed errors as proposed in [42], such that a desired level of normalized projection misalignment (NPM), i.e.,

$$\mathrm{NPM} = 10\log_{10}\frac{\left\|\mathbf{h}_m - \frac{\mathbf{h}_m^T\hat{\mathbf{h}}_m}{\hat{\mathbf{h}}_m^T\hat{\mathbf{h}}_m}\hat{\mathbf{h}}_m\right\|_2^2}{\|\mathbf{h}_m\|_2^2}[\mathrm{dB}], \tag{51}$$

is generated. The considered NPMs are

$$\mathrm{NPM} \in \{-33\mathrm{dB}, -27\mathrm{dB}, \ldots, -3\mathrm{dB}\}. \tag{52}$$

For all considered techniques the filter length is set to $L_w = \lceil\frac{L_h-1}{M-1}\rceil = 1700$ (cf. Section III), the length of the direct path and early reflections is set to $L_\mathrm{d} = 0.01 f_s$, and the delay is set to $\tau = 192$, cf. (19). The matrix $\mathbf{R}_e$ modeling the perturbations is set to $\mathbf{R}_e = \mathbf{I}$ as in [13], [33], which results in controlling the reshaping filter energy since $\epsilon_e = \mathbf{w}^T\mathbf{I}\mathbf{w} = \|\mathbf{w}\|_2^2$.

2) Recorded Acoustic System and Algorithmic Settings: To investigate the performance of the different techniques in a more realistic acoustic scenario, we have considered a recorded acoustic system with reverberation time $T_{60} \approx 450$ ms and $M = 4$ equidistant microphones with an inter-microphone distance of 2.5 cm. A speech source is placed at an angle $\theta \approx 10^\circ$ and a distance of 2 m from the microphone array. The noise consists of ambient noise, microphone self-noise, and a directional interferer placed at an angle $\theta \approx 70^\circ$ and a distance of 2.5 m from the microphone array. The speech and the noise components in the microphone signals are recorded by playing back a clean speech signal from HINT the database and a noise signal over loudspeakers. The broadband input SIR is again varied between -5 dB and 10 dB and the broadband input speech-to-ambient noise ratio is 25 dB. The speech-plus-noise signal is 18 s long and is preceded by a 14 s long noise-only signal.

The true RIRs $\mathbf{h}_m$ are measured using the swept-sine technique, with RIR length $L_h = 4000$ at a sampling frequency $f_s = 8$ kHz.[1] For each input SIR, RIR estimates are obtained using SSI by minimizing the least-squares cost function [14]

$$J = \|\mathbf{S}\mathbf{h}_m - \mathbf{y}_m\|_2^2, \quad m = 1, \ \ldots, \ M, \tag{53}$$

where $\mathbf{S}$ denotes the $L_s \times L_h$-dimensional convolution matrix of the clean speech signal, with $L_s = 18 f_s$. The minimization of (53) yields

$$\hat{\mathbf{h}}_m = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{y}_m. \tag{54}$$

Due to the noise in $\mathbf{y}_m$, the estimated RIRs $\hat{\mathbf{h}}_m$ differ from the true RIRs $\mathbf{h}_m$ [14], i.e.,

$$\hat{\mathbf{h}}_m = \mathbf{h}_m + \underbrace{(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{v}_m}_{\mathbf{e}_m}. \tag{55}$$

As illustrated by (55), the RIR perturbations $\mathbf{e}_m$ depend on the clean speech and noise statistics, such that they are typically not Gaussian distributed and also depend on the SIR.

The simulation parameters for all considered techniques are $L_w = \lceil\frac{L_h-1}{M-1}\rceil = 1333$, $L_\mathrm{d} = 0.01 f_s$, $\tau = 192$, and $\mathbf{R}_e = \mathbf{I}$.

3) Correlation Matrix Computation: The correlation matrices are computed as follows:

i. Perfectly estimated from the speech and noise signals in order to evaluate the full potential of the proposed techniques by avoiding correlation matrix estimation errors (Sections VI-B and VI-C), i.e.,

$$\mathbf{R}_\mathbf{x} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^T, \quad \mathbf{R}_\mathbf{v} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{v}_k\mathbf{v}_k^T, \tag{56}$$

with K denoting the number of available speech-plus-noise signal vectors.

ii. Erroneously estimated as $\mathbf{R}_\mathbf{x} = \mathbf{R}_\mathbf{y} - \mathbf{R}_\mathbf{v}$, with $\mathbf{R}_\mathbf{y}$ estimated during the speech-plus-noise period and $\mathbf{R}_\mathbf{v}$ estimated during the noise-only period in order to achieve a realistic evaluation of the proposed techniques (Section VI-D and VI-E), i.e.,

$$\mathbf{R}_\mathbf{y} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}_k\mathbf{y}_k^T, \quad \mathbf{R}_\mathbf{v} = \frac{1}{K_v}\sum_{k=1}^{K_v}\mathbf{v}_k\mathbf{v}_k^T,$$
$$\mathbf{R}_\mathbf{x} = \mathbf{R}_\mathbf{y} - \mathbf{R}_\mathbf{v}, \tag{57}$$

[1]Please note that referring to the measured RIRs as the true RIRs is not entirely correct. The RIRs are measured at a different time instant than the recorded speech, hence, environmental conditions (such as temperature) may have changed, possibly yielding a different true RIR.

with K_v denoting the number of available noise-only signal vectors. Due to the fact that the speech and noise signals are not perfectly uncorrelated and the noise is temporally nonstationary, computing the reverberant speech correlation matrix as $\mathbf{R_x} = \mathbf{R_y} - \mathbf{R_v}$ may not yield a positive semi-definite matrix, particularly at low input SIR. The estimated $\mathbf{R_x}$ is forced to be a positive semi-definite matrix by computing its eigenvalue decomposition and setting the negative eigenvalues to 0.

4) Performance Measures: The *dereverberation performance* is evaluated in terms of the reverberant energy suppression and the perceptual speech quality improvement. As is commonly done in acoustic multi-channel equalization techniques, the reverberant energy suppression is evaluated as the improvement in direct-to-reverberant ratio (ΔDRR) [43], i.e., $\Delta\mathrm{DRR} = \mathrm{oDRR} - \mathrm{iDRR}$ [dB], with

$$\mathrm{oDRR} = 10\log_{10}\frac{\sum_{n=0}^{L_\mathrm{d}-1} c^2(n)}{\sum_{n=L_\mathrm{d}}^{L_\mathrm{c}-1} c^2(n)}[\mathrm{dB}], \tag{58}$$

$$\mathrm{iDRR} = 10\log_{10}\frac{\sum_{n=0}^{L_\mathrm{d}-1} h_1^2(n)}{\sum_{n=L_\mathrm{d}}^{L_h-1} h_1^2(n)}[\mathrm{dB}], \tag{59}$$

where $c(n)$ is the resulting EIR defined in (8). The perceptual speech quality is evaluated using the instrumental measure PESQ [44], which generates a similarity score between a test signal and a reference signal in the range of 1 to 4.5. It has been shown in [45] that instrumental measures relying on auditory models such as PESQ exhibit the highest correlation with subjective listening tests when evaluating the quality of dereverberated (noiseless) speech. The reference signal employed here is $x_{\mathrm{d},1}(n) = s(n) * h_{\mathrm{d},1}(n)$, i.e., the direct path and early reflections speech component in the first microphone. The improvement in perceptual speech quality ΔPESQ is computed as the difference between the PESQ score of the output speech component $z_x(n)$ and the PESQ score of the reverberant speech component in the first microphone signal $x_1(n)$.

The *noise reduction performance* is evaluated in terms of the noise reduction factor ψ_{NR}, i.e.,

$$\psi_{\mathrm{NR}} = 10\log_{10}\frac{\mathcal{E}\{v_1^2(n)\}}{\mathcal{E}\{z_v^2(n)\}}[\mathrm{dB}], \tag{60}$$

with $v_1(n)$ the noise component in the first microphone and $z_v(n)$ the output noise component defined in (4).

The *joint dereverberation and noise reduction performance* is evaluated in terms of the improvement in signal-to-reverberation-and-noise ratio (ΔSRNR), i.e., $\Delta\mathrm{SRNR} = \mathrm{oSRNR} - \mathrm{iSRNR}$ [dB], with

$$\mathrm{iSRNR} = 10\log_{10}\frac{\mathcal{E}\{x_{\mathrm{d},1}^2(n)\}}{\mathcal{E}\{x_{\mathrm{r},1}^2(n)\} + \mathcal{E}\{v_1^2(n)\}}[\mathrm{dB}], \tag{61}$$

$$\mathrm{oSRNR} = 10\log_{10}\frac{\mathcal{E}\{z_\mathrm{d}^2(n)\}}{\mathcal{E}\{z_\mathrm{r}^2(n)\} + \mathcal{E}\{z_v^2(n)\}}[\mathrm{dB}], \tag{62}$$

where $x_{\mathrm{d},1}(n)$ and $x_{\mathrm{r},1}(n)$ are the direct and the late reverberant speech components in the first microphone defined in (3) and $z_\mathrm{d}(n)$ and $z_\mathrm{r}(n)$ are the direct and the late reverberant output speech components defined in (5). In addition, in order to evaluate the overall quality of the dereverberated and denoised signal, the frequency-weighted segmental SNR (fwSSNR) [46] is used, with $x_{\mathrm{d},1}(n)$ as reference signal. The improvement in overall quality $\Delta\mathrm{fwSSNR}$ [dB] is computed as the difference between the fwSSNR of the output signal $z(n)$ and the fwSSNR of the first microphone signal $y_1(n)$.

B. Influence of the Regularization and Weighting Parameters on the Performance of RPM-DNR and MWF-DNR

In this section the influence of the regularization and weighting parameters δ and μ on the performance of the RPM-DNR and MWF-DNR techniques is investigated using the simulated acoustic system for an exemplary scenario of $\mathrm{SIR} = 0$ dB and $\mathrm{NPM} = -33$ dB. The considered regularization and weighting parameter values are

$$\delta, \mu \in \{10^{-7}, 10^{-6}, \ldots, 10^{-1}, 1, 10\}, \tag{63}$$

and the speech and noise correlation matrices are perfectly estimated from the speech and noise signals as in (56).

Figs. 4(a) and 4(b) depict the DRR improvement and the noise reduction factor for the RPM-DNR technique. The following observations can be made:

i. For small values of the regularization and weighting parameters δ and μ (e.g., $\delta = 10^{-7}$ and $\mu = 10^{-7}$), the dereverberation performance is high whereas the noise is amplified. Since the RIR perturbation level is relatively low, i.e., $\mathrm{NPM} = -33$ dB, also the optimal value of the regularization parameter δ required for a high dereverberation performance is small (e.g., $\delta = 10^{-7}$). In addition, a small value of the weighting parameter μ (e.g., $\mu = 10^{-7}$), i.e., (almost) disregarding the noise, leads to noise amplification.
ii. For a fixed value of the weighting parameter μ (e.g., $\mu = 10^{-5}$), increasing the parameter δ initially yields a slight increase in ΔDRR (not visible in Fig. 4(a)), however, as the regularization parameter δ is increased beyond 10^{-5}, the ΔDRR values decrease. This is to be expected since for a relatively low RIR perturbation level, i.e., $\mathrm{NPM} = -33$ dB, the optimal value of δ required for a high dereverberation performance is small.
iii. For a fixed value of the weighting parameter μ (e.g., $\mu = 10^{-5}$), increasing the regularization parameter δ also increases the noise reduction factor. This can be explained by the fact that for increasing values of δ the energy of the resulting RPM-DNR filter decreases (since $\delta\mathbf{I}$ is used as the regularization term), which results in a smaller output noise power.
iv. For a fixed value of the regularization parameter δ (e.g., $\delta = 10^{-5}$), increasing the weighting parameter μ results in a trade-off between dereverberation and noise reduction performance, as can be seen by the (slight) decrease in DRR improvement and the increase in noise reduction

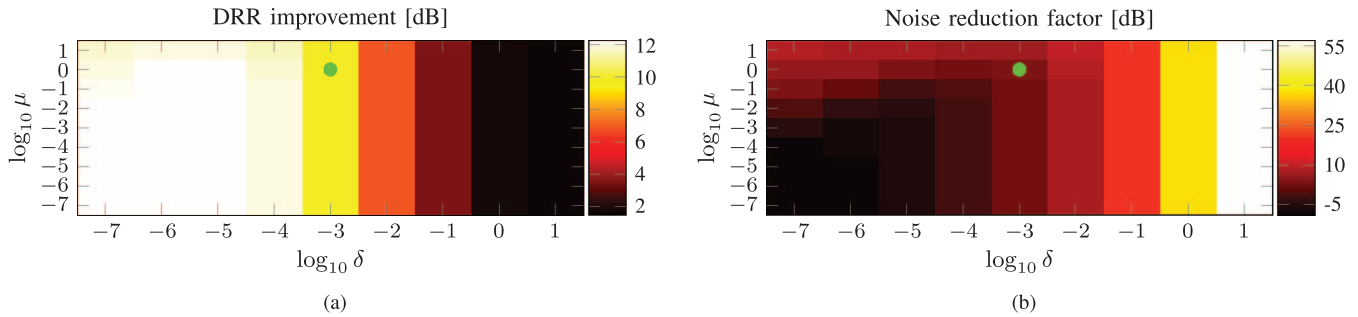

Fig. 4. Performance of the RPM-DNR technique for several regularization and weighting parameters δ and μ in terms of (a) DRR improvement and (b) noise reduction factor. The circles denote the automatically selected parameters (simulated acoustic system, SIR $= 0$ dB, NPM $= -33$ dB).

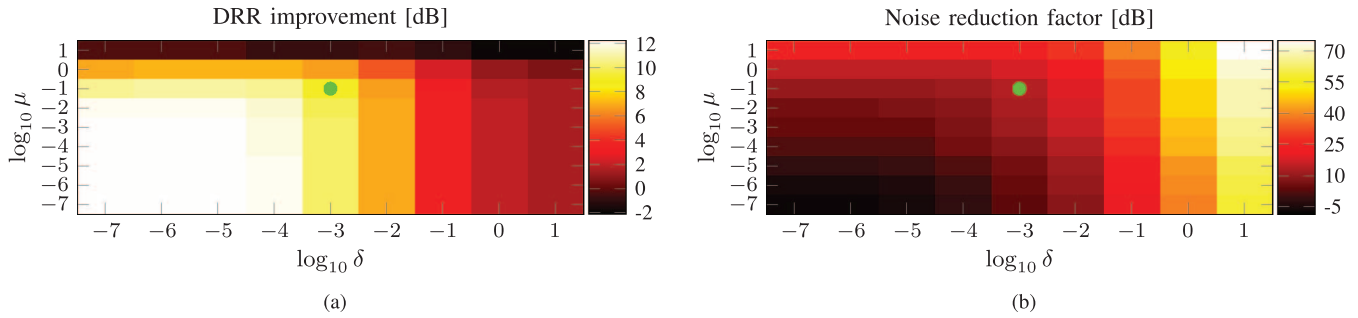

Fig. 5. Performance of the MWF-DNR technique for several regularization and weighting parameters δ and μ in terms of (a) DRR improvement and (b) noise reduction factor. The circles denote the automatically selected parameters (simulated acoustic system, SIR $= 0$ dB, NPM $= -33$ dB).

factor. However, for large values of the regularization parameter δ (e.g., $\delta = 1$), increasing the weighting parameter μ hardly has any effect on the dereverberation or the noise reduction performance, since the resulting RPM-DNR filter has very low energy.

For the considered example, the procedure proposed in Section V-B for automatically selecting the weighting parameters based on the L-hypersurface yields $\delta = 10^{-3}$ and $\mu = 1$, which are denoted by the circles in Figs. 4(a) and 4(b). While it is not possible to judge upon the optimality of a set of parameters, it can be said that the automatic procedure yields parameters resulting in a reasonable trade-off between dereverberation and noise reduction performance. This is also confirmed in Section VI-D for other NPMs and SIRs.

Figs. 5(a) and 5(b) depict the DRR improvement and the noise reduction factor for the MWF-DNR technique. Similarly to the RPM-DNR technique, the following observations can be made:

i. For small values of the regularization and weighting parameters δ and μ (e.g., $\delta = 10^{-7}$ and $\mu = 10^{-7}$), the dereverberation performance is high whereas the noise is amplified.
ii. For a fixed value of the weighting parameter μ (e.g., $\mu = 10^{-5}$), increasing the parameter δ initially yields a slight increase in ΔDRR (not visible in Fig. 5(a)), however, as the regularization parameter δ is increased beyond 10^{-5}, the ΔDRR values decrease.
iii. For a fixed value of the weighting parameter μ (e.g., $\mu = 10^{-5}$), increasing the parameter δ also increases the noise reduction factor.
iv. For a fixed value of the regularization parameter δ (e.g., $\delta = 10^{-5}$) increasing the parameter μ results in a trade-off between dereverberation and noise reduction performance, as can be seen by the decrease in DRR improvement and the increase in noise reduction factor.

For the considered example, the automatic procedure for selecting the regularization parameter δ in RPMINT yields $\delta = 10^{-3}$. Using this RPMINT filter, the automatic procedure for selecting the weighting parameter μ in MWF-DNR yields $\mu = 10^{-1}$. These parameter values are denoted by the circles in Figs. 5(a) and 5(b). It can be observed that using the two decoupled L-curve procedures for automatically selecting the regularization and weighting parameters in MWF-DNR yields parameters resulting in a reasonable trade-off between dereverberation and noise reduction performance. This is also confirmed in Section VI-D for other NPMs and SIRs.

C. Comparison of the Automatically Parametrized RPM-DNR and MWF-DNR Techniques to Acoustic Multi-channel Equalization Techniques

To illustrate the importance of taking the RIR perturbations and the noise statistics into account, in this section the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques is compared to the performance of the PMINT and the automatically regularized PMINT techniques using the simulated acoustic system for an exemplary scenario of SIR $= 0$ dB and NPM $= -33$ dB. The speech and noise correlation matrices for the RPM-DNR and MWF-DNR techniques are perfectly estimated as in (56).

TABLE I
PERFORMANCE OF PMINT AND AUTOMATICALLY PARAMETRIZED RPMINT, RPM-DNR, AND MWF-DNR (SIMULATED ACOUSTIC SYSTEM, SIR $= 0$ dB, NPM $= -33$ dB)

Measure	PMINT	RPMINT	RPM-DNR	MWF-DNR
ΔDRR [dB]	-3.3	**9.9**	9.8	9.1
ΔPESQ	-0.4	**0.7**	**0.7**	0.6
ψ_{NR} [dB]	-26.8	1.9	3.2	**13.0**
ΔSRNR [dB]	-11.5	2.2	3.0	**7.1**
ΔfwSSNR [dB]	-3.0	0.9	1.1	**3.2**

Table I presents the obtained $\Delta\mathrm{DRR}$, $\Delta\mathrm{PESQ}$, ψ_{NR}, $\Delta\mathrm{SRNR}$, and $\Delta\mathrm{fwSSNR}$ values for all considered techniques. As shown by the negative $\Delta\mathrm{DRR}$ and $\Delta\mathrm{PESQ}$ values, PMINT fails to achieve dereverberation, introducing more reverberant energy than in the microphone signal. By taking the RIR perturbations into account, RPMINT achieves a high reverberant energy suppression and perceptual speech quality improvement. The proposed RPM-DNR technique achieves a very similar dereverberation performance as RPMINT, whereas the proposed MWF-DNR technique yields only a slightly worse dereverberation performance. Even though one would expect the dereverberation performance of the RPM-DNR technique to be worse than the dereverberation performance of RPMINT, in this scenario the dereverberation performance of both techniques is very similar. This occurs due to the automatic selection of the regularization parameter, which does not yield the best dereverberation performance one would otherwise obtain by intrusively selecting the regularization parameter in the RPMINT technique. Furthermore, as discussed in Section III and as illustrated by the negative noise reduction factor, PMINT leads to a large noise amplification. Due to the decrease in the reshaping filter energy by incorporating a regularization parameter, RPMINT avoids the noise amplification and reduces the noise by 1.9 dB. By taking the noise statistics explicitly into account, the proposed RPM-DNR technique improves the noise reduction factor to 3.2 dB, whereas by taking also the reverberant speech statistics into account the proposed MWF-DNR technique yields a significantly larger noise reduction factor of 13.0 dB. The high dereverberation and noise reduction performance of the proposed techniques in comparison to acoustic multi-channel equalization techniques is also illustrated by the higher $\Delta\mathrm{SRNR}$ and $\Delta\mathrm{fwSSNR}$ values presented in Table I, where the MWF-DNR technique outperforms the RPM-DNR technique in terms of both instrumental measures. Summarizing these results, it can be said that the RIR perturbations and the noise statistics should be taken into account in order to avoid noise amplification and to achieve joint dereverberation and noise reduction. By taking also the reverberant speech statistics into account, an overall better performance can be achieved.

D. Performance of the Automatically Parametrized RPM-DNR and MWF-DNR Techniques for Simulated Acoustic Systems

In this section the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques is extensively investigated using the simulated acoustic system for different noise levels, RIR perturbation levels, and correlation matrix estimation errors. The considered NPMs are given in (52) and the presented performance measures for each SIR value are averaged over the different considered NPMs. The performance of the proposed RPM-DNR and MWF-DNR techniques is investigated for perfectly estimated correlation matrices as in (56) and for erroneously estimated correlation matrices as in (57).

Fig. 6 depicts the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques for perfectly estimated speech and noise correlation matrices. As shown by the $\Delta\mathrm{DRR}$ and $\Delta\mathrm{PESQ}$ values in Figs. 6(a) and 6(b), the dereverberation performance of both techniques is very similar, with the RPM-DNR technique yielding a slightly better performance. However, as shown by the noise reduction factor in Fig. 6(c), the MWF-DNR technique achieves a significantly better noise reduction performance. The similar dereverberation performance but better noise reduction performance of the MWF-DNR technique is reflected in the higher $\Delta\mathrm{SRNR}$ and $\Delta\mathrm{fwSSNR}$ values achieved by the MWF-DNR technique, as depicted in Figs. 6(d) and 6(e). Hence, it can be said that by also taking the true reverberant speech statistics into account, the MWF-DNR technique outperforms the RPM-DNR technique, since it yields a similarly high dereverberation performance but a significantly higher noise reduction performance.

Fig. 7 depicts the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques for erroneously estimated correlation matrices as in (57). Since the RPM-DNR technique only requires the noise correlation matrix $\mathbf{R}_{\mathbf{v}}$ and since estimating this matrix from a long enough spatially stationary noise-only period does not yield a significantly different estimate from the previous experiment, the performance of the RPM-DNR technique for erroneously estimated correlation matrices is very similar to the performance for perfectly estimated correlation matrices (compare Figs. 6 and 7). However, as shown in Figs. 7(a) and 7(b) the dereverberation performance of the MWF-DNR technique significantly decreases. Due to the fact that the speech and noise signals are not perfectly uncorrelated and the noise is temporally nonstationary, estimation errors occur in the estimate of the speech correlation matrix $\mathbf{R}_{\mathbf{x}} = \mathbf{R}_{\mathbf{y}} - \mathbf{R}_{\mathbf{v}}$, especially for low input SIR. These estimation errors result in a worse dereverberated reference signal $\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{RP}}$ for the MWF-DNR technique, hence, significantly decreasing the dereverberation performance. However, the noise reduction performance for the MWF-DNR technique is significantly better than for the RPM-DNR technique as depicted in Fig. 7(c), resulting in higher overall $\Delta\mathrm{SRNR}$ and $\Delta\mathrm{fwSSNR}$ values as depicted in Figs. 7(d) and 7(e). Furthermore, the noise reduction performance of the MWF-DNR technique for erroneously estimated correlation matrices can be better than for perfectly estimated correlation matrices (compare Figs. 6c and 7c for $\mathrm{SIR} = 5$ dB and $\mathrm{SIR} = 10$ dB). This occurs due to the automatic selection of the weighting parameter μ in the MWF-DNR technique, which for erroneously estimated correlation matrices may result in a higher parameter value, hence a better noise reduction performance (at the expense of a worse dereverberation performance).

In summary, when the speech and noise correlation matrices can be directly estimated from the speech and noise signals, the

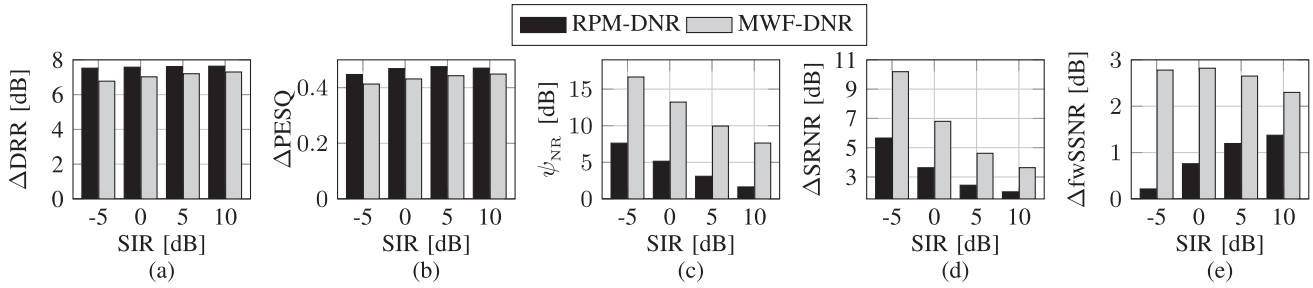

Fig. 6. Average performance of the automatically parametrized RPM-DNR and MWF-DNR techniques in terms of (a) $\Delta\mathrm{DRR}$, (b) $\Delta\mathrm{PESQ}$ of the output speech component, (c) ψ_{NR}, (d) $\Delta\mathrm{SRNR}$, and (e) $\Delta\mathrm{fwSSNR}$ (simulated acoustic system, perfectly estimated correlation matrices).

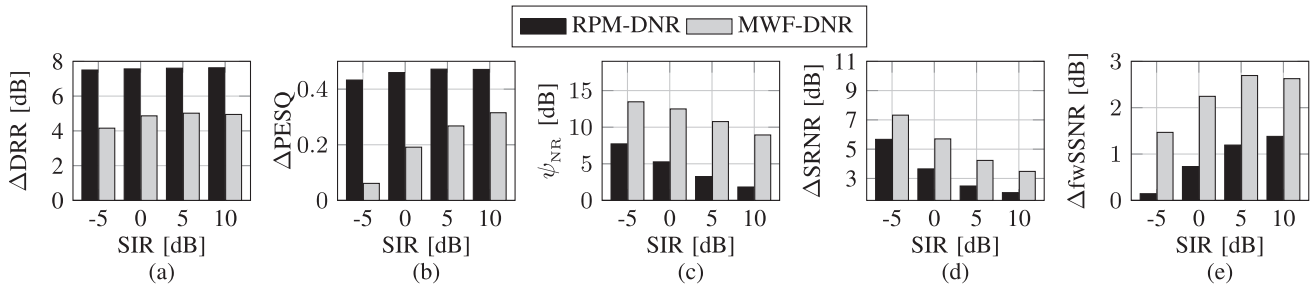

Fig. 7. Average performance of the automatically parametrized RPM-DNR and MWF-DNR techniques in terms of (a) $\Delta\mathrm{DRR}$, (b) $\Delta\mathrm{PESQ}$ of the output speech component, (c) ψ_{NR}, (d) $\Delta\mathrm{SRNR}$, and (e) $\Delta\mathrm{fwSSNR}$ (simulated acoustic system, erroneously estimated correlation matrices).

TABLE II
PERFORMANCE OF THE AUTOMATICALLY PARAMETRIZED RPM-DNR AND MWF-DNR TECHNIQUES (RECORDED ACOUSTIC SYSTEM, ERRONEOUSLY ESTIMATED CORRELATION MATRICES)

		ΔDRR [dB]		ΔPESQ		η_{NR} [dB]		ΔfwSSNR [dB]	
SIR [dB]	NPM [dB]	RPM-DNR	MWF-DNR	RPM-DNR	MWF-DNR	RPM-DNR	MWF-DNR	RPM-DNR	MWF-DNR
−5	−3.0	**6.0**	3.9	**0.2**	−0.1	5.9	**16.4**	0.6	**3.8**
0	−4.5	**6.8**	5.2	**0.5**	0.0	4.4	**12.7**	1.1	**4.7**
5	−5.2	**7.4**	6.0	**0.7**	0.2	3.3	**9.9**	1.6	**5.7**
10	−5.5	**7.8**	6.8	**0.8**	0.4	2.4	**8.0**	1.7	**6.6**
10	$-\infty$	**8.4**	6.8	**0.8**	0.3	1.5	**8.0**	1.0	**5.6**

MWF-DNR technique outperforms the RPM-DNR technique since it yields a similarly high dereverberation performance and a significantly better noise reduction performance. However, when the required correlation matrices are prone to estimation errors, the RPM-DNR technique yields a significantly better dereverberation performance but still a worse noise reduction performance than the MWF-DNR technique. The technique to be used should be chosen depending on what is more important for the application under consideration, i.e., dereverberation or noise reduction performance.

E. Performance of the Automatically Parametrized RPM-DNR and MWF-DNR Techniques for Recorded Acoustic Systems

In this section the performance of the automatically parametrized RPM-DNR and MWF-DNR techniques is investigated using the recorded acoustic system for different noise levels, and hence, different RIR perturbations, cf. (55). The correlation matrices are erroneously estimated as in (57).

Table II presents the obtained $\Delta\mathrm{DRR}$, $\Delta\mathrm{PESQ}$, η_{NR}, and $\Delta\mathrm{fwSSNR}$ values for both proposed techniques. It should be noted that the presented $\Delta\mathrm{PESQ}$ values do not only reflect the improvement in dereverberation performance, but also the reduction of ambient noise and microphone self-noise (since for the recorded acoustic system, the individual speech, ambient noise or microphone self-noise components are not available). For each considered SIR, the NPM arising between the measured RIRs and the estimated RIRs is also presented in Table II. As a baseline, the performance of the RPM-DNR and MWF-DNR techniques using the measured RIRs (i.e., $\mathrm{NPM} = -\infty$) for the exemplary scenario of $\mathrm{SIR} = 10$ dB is also presented.

As illustrated by the $\Delta\mathrm{DRR}$ and $\Delta\mathrm{PESQ}$ values and as expected from the results of Section VI-D, the RPM-DNR technique yields a better dereverberation performance than the MWF-DNR technique. Furthermore, as illustrated by the η_{NR} values, the MWF-DNR technique yields a better noise reduction performance than the RPM-DNR technique. The better noise reduction performance of the MWF-DNR technique also results in a better overall joint dereverberation and noise reduction performance, as illustrated by the $\Delta\mathrm{fwSSNR}$ values. Most importantly, it can be observed that at $\mathrm{SIR} = 10$ dB the performance of both techniques for $\mathrm{NPM} = -5.5$ dB is very similar to the performance for $\mathrm{NPM} = -\infty$ dB, illustrating the robustness of the proposed techniques against RIR perturbations arising due to least-squares SSI.

In summary, these simulation results confirm the results of Section VI-D, i.e., in the presence of correlation matrix estimation errors, the RPM-DNR technique yields a significantly better dereverberation performance whereas the MWF-DNR technique yields a significantly better noise reduction and joint dereverberation and noise reduction performance. Most importantly, these simulation results show the applicability of the proposed techniques to more realistic acoustic scenarios, with realistic RIR perturbations arising from SSI methods.

VII. Conclusion

In this paper we have proposed two techniques for joint dereverberation and noise reduction based on acoustic multi-channel equalization. The RPM-DNR technique can be seen as an extension of the RPMINT technique by explicitly taking the noise statistics into account. The MWF-DNR technique takes also the reverberant speech statistics into account and uses the dereverberated output signal of the RPMINT technique as the reference signal for the MWF. In addition, we proposed an automatic non-intrusive procedure based on the L-hypersurface for selecting the regularization and weighting parameters in the RPM-DNR technique, whereas two decoupled procedures based on the L-curve were used for the automatic selection of the parameters in the MWF-DNR technique. Simulation results demonstrate that the RPM-DNR technique maintains the high dereverberation performance of acoustic multi-channel equalization techniques while improving the noise reduction performance. Furthermore, it is shown that the MWF-DNR technique yields a significantly better noise reduction performance than the RPM-DNR technique at the expense of a worse dereverberation performance, depending on the amount of estimation errors in the speech correlation matrix.

References

[1] M. Omologo, P. Svaizer, and M. Matassoni, "Environmental conditions and acoustic transduction in hands-free speech recognition," *Speech Commun.*, vol. 25, 1–3pp. 75–95, Aug. 1998.

[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] K. Lebart and J. M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoust.*, vol. 87, no. 3, pp. 359–366, May–Jun. 2001.

[4] E. A. P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, USA, Mar. 2005, pp. 173–176.

[5] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sep. 2009.

[6] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," *Proc. Eur. Signal Process. Conf.*, Lisbon, Portugal, Sep. 2014, pp. 61–65.

[7] 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.

[8] A. Jukić and S. Doclo, "Speech dereverberation using weighted prediction error with Laplacian model of the desired signal," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 5172–5176.

[9] B. Schwartz, S. Gannot, and E. A. P. Habets, "Online speech dereverberation using Kalman filter and EM algorithm," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 2, pp. 394–406, Feb. 2015.

[10] 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.

[11] M. Kallinger and A. Mertins, "Multi-channel room impulse response shaping - a study," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, May 2006, pp. 101–104.

[12] J. O. Jungmann, R. Mazur, M. Kallinger, M. Tiemin, and A. Mertins, "Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, pp. 1829–1842, Aug. 2012.

[13] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.

[14] F. Lim, W. Zhang, E. A. P. Habets, and P. A. Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 9, pp. 1379–1390, Sep. 2014.

[15] R. S. Rashobh, A. W. H. Khong, and D. Liu, "Multichannel equalization in the KLT and frequency domains with application to speech dereverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 3, pp. 634–646, Mar. 2014.

[16] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," *Microphone Arrays*, M. Brandstein and D. Ward, Eds, Berlin, Germany: Springer, 2001.

[17] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," *Proc. Eur. Signal Process. Conf.*, Marrakech, Morocco, Sep. 2013.

[18] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Joint dereverberation and noise reduction using beamforming and a single-channel speech enhancement scheme," *Proc. Reverb Challenge Workshop*, Florence, Italy, May 2014.

[19] E. A. P. Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 945–958, May 2013.

[20] T. Yoshioka, T. Nakatani, and M. Miyoshi, "Integrated speech enhancement method using noise suppression and dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 2, pp. 231–246, Feb. 2009.

[21] N. Ito, S. Araki, and T. Nakatani, "Probabilistic integration of diffuse noise suppression and dereverberation," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 5167–5171.

[22] 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.

[23] 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.

[24] M. A. Haque and T. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 305–308, Feb. 2008.

[25] 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.

[26] F. Lim and P. Naylor, "Statistical modelling of multichannel blind system identification errors," *Proc. Int. Workshop Acoust. Echo Noise Control*, Antibes, France, Sep. 2014, pp. 119–123.

[27] 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.

[28] M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Application of channel shortening to acoustic channel equalization in the presence of noise and estimation error," *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2011, pp. 113–116.

[29] I. Kodrasi and S. Doclo, "Joint dereverberation and noise reduction based on acoustic multichannel equalization," *Proc. Int. Workshop Acoust. Echo Noise Control*, Antibes, France, Sep. 2014, pp. 139–143.

[30] M. Belge, M. Kilmer, and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet transform domain," *SPIE 3459, Bayesian Inference for Inverse Problems*, Sep. 1998, vol. 328.

[31] 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.

[32] M. R. P. Thomas, N. D. Gaubitch, E. A. P. Habets, and P. A. Naylor, "An Insight into Common Filtering in Noisy SIMO Blind System Identification," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 521–524.
[33] 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.
[34] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Commun.*, vol. 49, no. 7–8, pp. 636–656, Jul. 2007.
[35] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.
[36] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," *Proc. Int. Workshop Acoust. Echo Noise Control*, Darmstadt, Germany, Sep. 2001, pp. 31–34.
[37] J. L. Castellanos, S. Gómez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Appl. Numer. Math.*, vol. 43, no. 4, pp. 359–373, Dec. 2002.
[38] M. Belge, M. Kilmer, and E. L. Miller, "Efficient determination of multiple regularization parameters in a generalized L-curve framework," *Inverse Problems*, vol. 18, pp. 1161–1183, Jul. 2002.
[39] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," *Proc. Int. Workshop Acoust. Echo Noise Control*, Antibes, France, Sep. 2014, pp. 313–317.
[40] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.
[41] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Amer.*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.
[42] W. Zhang and P. A. Naylor, "An algorithm to generate representations of system identification errors," *Res. Lett. Signal Process.*, vol. 2008, Jan. 2008.
[43] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, London, U.K.: Springer, 2010.
[44] 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., Feb. 2001.
[45] S. Goetze, E. Albertin, J. Rennies, E. A. P. Habets, and K.-D. Kammeyer, "Speech quality assessment for listening-room compensation," *Proc. 38th AES Int. Conf. Sound Quality Eval.*, Pitea, Sweden, Jun. 2010, pp. 11–20.
[46] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

Ina Kodrasi (S'11) received the Master of Science degree in communications, systems and electronics in 2010 from Jacobs University Bremen, Germany. Currently she is 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 also 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.

Simon Doclo (S'95–M'03–SM'13) 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 has been 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 was member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008–2013) and Technical Program Chair for the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Doclo has served as guest editor for several special issues (*IEEE Signal Processing Magazine*, *Elsevier Signal Processing*) and is associate editor for the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH,, AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.