INCORPORATING THE NOISE STATISTICS IN ACOUSTIC MULTI-CHANNEL EQUALIZATION

Ina Kodrasi and Simon Doclo*

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

Correspondence should be addressed to Ina Kodrasi (ina.kodrasi@uni-oldenburg.de)

ABSTRACT

Acoustic multi-channel equalization techniques, such as the regularized partial multi-channel equalization technique based on the multiple-input/output inverse theorem (RP-MINT), are able to achieve a high dereverberation performance in the presence of room impulse response perturbations but may lead to additive noise amplification. This paper proposes to directly extend the RP-MINT technique by incorporating the noise statistics in the reshaping filter design, such that joint dereverberation and noise reduction is achieved. In addition to the regularization parameter used in the RP-MINT technique, a weighting parameter is introduced to trade off between dereverberation and noise reduction. To automatically determine the regularization and weighting parameters, a novel non-intrusive procedure based on the L-hypersurface is proposed. Simulation results using instrumental performance measures show that the proposed technique maintains the high dereverberation performance of the RP-MINT technique, while improving the noise reduction performance.

1. INTRODUCTION

Speech signals recorded in an enclosed space by microphones placed at a distance from the source are often corrupted by reverberation and additive noise, which typically degrade speech quality, impair speech intelligibility, and decrease the performance of automatic speech recognition systems [1–3]. With the continuously growing demand for high-quality hands-free speech communication, speech enhancement techniques aiming at joint dereverberation and noise reduction have become indispensable. In this paper, we focus on the effective integration of the dereverberation and noise reduction tasks using acoustic multi-channel equalization techniques.

Acoustic multi-channel equalization techniques [4–8] aim to reshape the measured or estimated room impulse responses (RIRs) between the speech source and the microphone array. These techniques comprise in principle an attractive approach to speech dereverberation since in theory perfect dereverberation can be achieved [4]. In practice however, such techniques suffer from several drawbacks. Since the available (measured or estimated) RIRs typically differ from the true RIRs due to fluctuations (e.g., temperature or position variations [9]) or due to the sensitivity of supervised and blind system identification methods to near-common zeros and interfering noise [10, 11], such techniques can fail to achieve dereverberation and possibly cause additional speech distortion in the output signal [6, 7]. Furthermore, acoustic multi-channel equalization techniques typically design reshaping filters aiming only at speech dereverberation, without taking the presence of the additive noise into account. Applying such dereverberation filters may result in a large noise amplification [7].

To increase the robustness against RIR perturbations, several techniques have been proposed [6–8], with the regularized partial multi-channel equalization technique based on the multiple-input/output inverse theorem (RP-MINT) shown to yield a high dereverberation performance [7]. By incorporating regularization in the RP-MINT technique, the energy of the reshaping filter is decreased, reducing the distortions in the output signal due to RIR perturbations, and hence, increasing the dereverberation performance. While the regularization parameter introduced in the RP-MINT technique is also effective in partly avoiding the additive noise amplification, the noise reduction performance is limited since the actual noise statistics are not explicitly taken into account.

*This work was supported in part by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development, the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG), and the Marie Curie Initial Training Network DREAMS (Grant no. 316969).

In this paper we propose to directly extend the RP-MINT technique by explicitly taking the actual noise statistics into account such that joint dereverberation and noise reduction is achieved. In addition to the regularization parameter used in the RP-MINT technique, a weighting parameter is introduced, which enables to trade off between dereverberation and noise reduction. Furthermore, a novel procedure for the joint automatic selection of the regularization and weighting parameters is also proposed. Simulation results show that the proposed technique maintains the high dereverberation performance of the RP-MINT technique, while improving the noise reduction performance.[1]

2. CONFIGURATION AND NOTATION

Consider the acoustic system depicted in Fig. 1, consisting of a single speech source, M microphones, and additive noise. Each microphone signal $y_m(n)$, $m = 1, \ldots, M$, at discrete-time index n, consists of a filtered version of the clean speech signal $s(n)$ and a noise component $v_m(n)$, i.e.,

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

where $h_m(n)$ is the RIR between the speech source and the m-th microphone, $x_m(n)$ is the reverberant speech component at the m-th microphone, and $*$ denotes convolution. Since the RIR $h_m(n)$ consists of a direct path and early reflections component $h_{e,m}(n)$ and a late reflections component $h_{r,m}(n)$, i.e., $h_m(n) = h_{e,m}(n) + h_{r,m}(n)$, the microphone signal in (1) can also be written as

$$y_m(n) = \underbrace{h_{e,m}(n) * s(n)}_{x_{e,m}(n)} + \underbrace{h_{r,m}(n) * s(n)}_{x_{r,m}(n)} + v_m(n), \quad (2)$$

with $x_{e,m}(n)$ the early reverberation component and $x_{r,m}(n)$ the late reverberation component at the m-th microphone. Using the filter-and-sum structure in Fig. 1, the enhanced output signal $z(n)$ is equal to the sum of the filtered microphone signals, i.e.,

$$z(n) = \underbrace{\sum_{m=1}^{M} x_m(n) * w_m(n)}_{z_x(n)} + \underbrace{\sum_{m=1}^{M} v_m(n) * w_m(n)}_{z_v(n)}, \quad (3)$$

where $w_m(n)$ is the 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

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

with $c(n)$ the equalized impulse response (EIR) between the clean speech signal $s(n)$ and the output speech component $z_x(n)$. Furthermore, the early reverberation output speech component $z_{e,x}(n)$ and the late reverberation output speech component $z_{r,x}(n)$ are defined as

$$z_{e,x}(n) = \sum_{m=1}^{M} x_{e,m}(n) * w_m(n), \quad (5)$$

$$z_{r,x}(n) = \sum_{m=1}^{M} x_{r,m}(n) * w_m(n). \quad (6)$$

In vector notation, the RIR $\mathbf{h}_m$ and the filter $\mathbf{w}_m$ are given by

$$\mathbf{h}_m = [h_m(0)\ h_m(1)\ \ldots\ h_m(L_h-1)]^T, \quad (7)$$

$$\mathbf{w}_m = [w_m(0)\ w_m(1)\ \ldots\ w_m(L_h-1)]^T, \quad (8)$$

with L_h and L_w the RIR and the filter length respectively. Using the ML_w-dimensional stacked filter vector $\mathbf{w}$, i.e., $\mathbf{w} = [\mathbf{w}_1^T\ \mathbf{w}_2^T\ \ldots\ \mathbf{w}_M^T]^T$, the EIR vector $\mathbf{c}$ of length L_c, i.e., $\mathbf{c} = [c(0)\ c(1)\ \ldots\ c(L_c-1)]^T$, is equal to

$$\mathbf{c} = \mathbf{H}\mathbf{w}, \quad (9)$$

with $\mathbf{H}$ the $L_c \times ML_w$-dimensional multi-channel convolution matrix. Using the ML_w-dimensional stacked vector of the received microphone signals

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

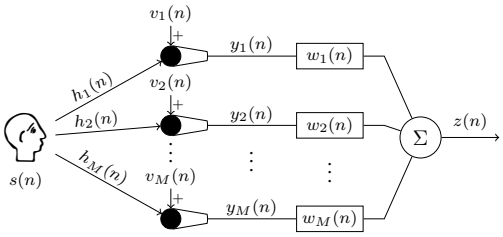

Fig. 1: Schematic illustration of a typical time-domain multi-channel speech enhancement system.

[1] An extensive theoretical and experimental analysis of the technique proposed here is provided in [12].

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, the enhanced output signal $z(n)$ can be expressed as

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

with $\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-L_c-1)]^T$ and $\mathbf{x}(n) = \mathbf{H}^T\mathbf{s}(n)$. For conciseness, the time index n will be omitted when possible in the remainder of this paper.

3. ACOUSTIC MULTI-CHANNEL EQUALIZATION TECHNIQUES

Acoustic multi-channel equalization techniques typically disregard the presence of the additive noise $\mathbf{v}$ and design the reshaping filter $\mathbf{w}$ such that only the EIR $\mathbf{c}$ is optimized. Since the presence of the additive noise is disregarded, such techniques can result in a large noise amplification [7] (cf. Section 6.3). Furthermore, since in practice only the perturbed RIRs $\hat{h}_m$ are available, the perturbed convolution matrix $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ is used for the reshaping filter design, with $\mathbf{E}$ the convolution matrix of the RIR perturbations.

In this paper we will focus on the partial multi-channel equalization technique based on the multiple-input/output inverse theorem (P-MINT) proposed in [7], which aims at suppressing the late reverberation and preserving the perceptual speech quality. To this purpose, the late reflection taps of the target EIR $\mathbf{c}_t$ are set equal to $\mathbf{0}$, while the remaining taps are set equal to the direct path and early reflections of one of the available RIRs, i.e.,

$$\mathbf{c}_t = [\underbrace{\hat{h}_p(0)\ \ldots\ \hat{h}_p(L_d-1)}_{L_d}\ 0\ \ldots\ 0\ ]^T, \quad (14)$$

where L_d corresponds to the length of the direct path and early reflections and $p \in \{1,\ 2,\ \ldots,\ M\}$. The P-MINT filter is computed by minimizing the least-squares cost function

$$J_{\text{P}} = \|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2. \quad (15)$$

As shown in [4], assuming that the available RIRs do not share any common zeros and using $L_w \geq \left\lceil \frac{L_h-1}{M-1} \right\rceil$, the P-MINT filter minimizing the least-squares cost function in (15) is equal to

$$\mathbf{w}_{\text{P}} = \hat{\mathbf{H}}^+\mathbf{c}_t, \quad (16)$$

where $\{\cdot\}^+$ denotes the matrix pseudo-inverse. When the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the P-MINT filter yields perfect dereverberation performance, i.e., $\mathbf{H}\mathbf{w}_{\text{P}} = \mathbf{c}_t$. However, in the presence of RIR perturbations, i.e., $\hat{\mathbf{H}} \neq \mathbf{H}$, applying the P-MINT filter to the true convolution matrix yields

$$\mathbf{H}\mathbf{w}_{\text{P}} = \hat{\mathbf{H}}\mathbf{w}_{\text{P}} - \mathbf{E}\mathbf{w}_{\text{P}} = \mathbf{c}_t - \mathbf{E}\mathbf{w}_{\text{P}}. \quad (17)$$

The first term in (17) is the target EIR, whereas the second term represents distortions due to RIR perturbations. If the energy of the reshaping filter $\mathbf{w}$ is small, then the distortions caused by RIR perturbations are also small. To decrease the reshaping filter energy, and hence, to increase the robustness of the P-MINT technique against RIR perturbations, the RP-MINT technique has been proposed in [7, 13]. The RP-MINT cost function is given by

$$J_{\text{RP}} = \underbrace{\|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2}_{\varepsilon_c} + \delta\underbrace{\mathbf{w}^T\mathbf{w}}_{\varepsilon_w}, \quad (18)$$

where ε_c denotes the dereverberation error energy, ε_w denotes the reshaping filter energy, and δ is a regularization parameter providing a trade-off between both terms. Minimizing (18) yields the RP-MINT filter

$$\mathbf{w}_{\text{RP}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t, \quad (19)$$

with $\mathbf{I}$ the $ML_w \times ML_w$-dimensional identity matrix. While the P-MINT filter fails to achieve dereverberation in the presence of RIR perturbations, it has been shown in [7] that the RP-MINT filter yields a significantly better dereverberation performance. Furthermore, the RP-MINT filter is able to partly avoid the additive noise amplification at the output of the system [7] (cf. Section 6.3), however, the noise reduction performance is limited since the actual noise statistics are not explicitly taken into account.

Clearly, the dereverberation performance of the RP-MINT technique depends on the regularization parameter δ which enables to trade off between the dereverberation error energy ε_c and the reshaping filter energy ε_w, with

$$\varepsilon_c = \|\hat{\mathbf{H}}\mathbf{w}_{\text{RP}} - \mathbf{c}_t\|_2^2, \quad (20)$$

$$\varepsilon_w = \mathbf{w}_{\text{RP}}^T\mathbf{w}_{\text{RP}}. \quad (21)$$

An appropriate regularization parameter should incorporate knowledge about both the dereverberation error energy and the reshaping filter energy, such that both terms are low. In order to automatically compute the regularization parameter in the RP-MINT technique, it has been proposed in [7] to use a parametric plot of the reshaping filter energy $\mathcal{E}_w$ versus the dereverberation error energy $\mathcal{E}_c$ 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 reshaping filter $\mathbf{w}_{\text{RP}}$ in (19) changes from being dominated by over-regularization to being dominated by under-regularization. It has therefore been proposed in [7] to automatically select the regularization parameter δ as the point of maximum curvature of this L-curve. Experimental results in [7] have shown that this automatic parameter selection procedure yields a very similar robustness against RIR perturbations as intrusively selecting the regularization parameter.

4. INCORPORATING THE NOISE STATISTICS IN ACOUSTIC MULTI-CHANNEL EQUALIZATION

Since acoustic multi-channel equalization techniques design reshaping filters aiming only at speech dereverberation without taking the presence of the additive noise into account, the output noise power is not explicitly controlled and may even be amplified compared to the noise power in the microphone signals. The output noise power $\mathcal{E}_v$ is given by

$$\mathcal{E}_v = \mathscr{E}\{(\mathbf{w}^T\mathbf{v})^2\} = \mathbf{w}^T\mathbf{R}_{\mathbf{v}}\mathbf{w}, \tag{22}$$

with $\mathscr{E}$ the expected value operator and $\mathbf{R}_{\mathbf{v}}$ the additive noise correlation matrix.

Aiming at controlling the dereverberation error energy $\mathcal{E}_c$, the reshaping filter energy $\mathcal{E}_w$, as well as the output noise power $\mathcal{E}_v$, we propose to extend the RP-MINT cost function in (18) such that the actual noise statistics are explicitly taken into account. The RP-MINT cost function for joint dereverberation and noise reduction (RP-DNR) can then be written as

$$J_{\text{RP-DNR}} = J_{\text{RP}} + \mu\mathcal{E}_v \tag{23}$$

$$= \underbrace{\|\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t\|_2^2}_{\mathcal{E}_c} + \delta\underbrace{\mathbf{w}^T\mathbf{w}}_{\mathcal{E}_w} + \mu\underbrace{\mathbf{w}^T\mathbf{R}_{\mathbf{v}}\mathbf{w}}_{\mathcal{E}_v}, \tag{24}$$

with δ the regularization parameter controlling the weight given to the reshaping filter energy and μ an additional weighting parameter controlling the weight given to the output noise power. The RP-DNR filter minimizing (24) is equal to

$$\mathbf{w}_{\text{RP-DNR}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I} + \mu\mathbf{R}_{\mathbf{v}})^{-1}\hat{\mathbf{H}}^T\mathbf{c}_t. \tag{25}$$

Clearly, the dereverberation and noise reduction performance of the RP-DNR filter in (25) depend on the regularization and weighting parameters δ and μ. Increasing the regularization parameter δ results in a lower reshaping filter 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. While in simulations the optimal values for the parameters δ and μ can be intrusively determined, in practice an automatic non-intrusive procedure is required. In Section 5 we propose a novel procedure for the joint automatic selection of the regularization and weighting parameters in the RP-DNR technique.

5. AUTOMATIC SELECTION OF THE REGULARIZATION AND WEIGHTING PARAMETERS

As already mentioned, different regularization and weighting parameters δ and μ obviously result in different RP-DNR filters in (25), which yield different dereverberation error energy $\mathcal{E}_c$, reshaping filter energy $\mathcal{E}_w$, and output noise power $\mathcal{E}_v$, with

$$\mathcal{E}_c = \|\hat{\mathbf{H}}\mathbf{w}_{\text{RP-DNR}} - \mathbf{c}_t\|_2^2, \tag{26}$$

$$\mathcal{E}_w = \mathbf{w}_{\text{RP-DNR}}^T\mathbf{w}_{\text{RP-DNR}}, \tag{27}$$

$$\mathcal{E}_v = \mathbf{w}_{\text{RP-DNR}}^T\mathbf{R}_{\mathbf{v}}\mathbf{w}_{\text{RP-DNR}}. \tag{28}$$

Similarly as for the RP-MINT technique, appropriate parameters δ and μ should incorporate knowledge about the dereverberation error energy, the reshaping filter energy, and the output noise power, such that all three terms are low. Motivated by the simplicity and the applicability of the L-curve for regularizing least-squares techniques [14], the so-called L-hypersurface has been proposed in [15] 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 (cf. Section 3), 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 $\mathcal{E}_v$ versus the dereverberation error energy $\mathcal{E}_c$ and the reshaping filter

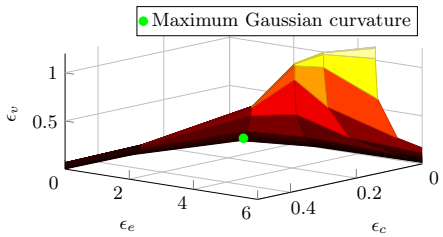

Fig. 2: Exemplary parametric surface of the output noise power ε_v versus the dereverberation error energy ε_c and the distortion energy ε_w for the RP-DNR technique.

energy ε_w for several parameters δ and μ.
Fig. 2 depicts an exemplary L-hypersurface obtained by plotting ε_v versus ε_c and ε_w for several regularization and weighting parameters δ and μ, with the circle 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 [16], such that a numerically stable algorithm is required. In this paper, the minimum distance method proposed in [16] has been used to compute the point of maximum Gaussian curvature.

6. SIMULATIONS

In this section the dereverberation and noise reduction performance when using the P-MINT filter in (16), the automatically regularized RP-MINT filter in (19), and the automatically parametrized RP-DNR filter in (25) will be evaluated.

6.1. Acoustic system and algorithmic settings

We have considered an acoustic scenario with a single speech source placed in broadside direction to a linear microphone array with $M = 4$ equidistant microphones. The room reverberation time is $T_{60} \approx 610$ ms [17], the RIR length is $L_h = 4880$, and the sampling frequency is $f_s = 8$ kHz. The distance between the microphones is 4 cm and the distance between the speech source and the microphone array is 2 m. The speech components in the microphone signals are generated by convolving clean speech signals from the HINT database [18] with the measured RIRs. The noise components consist of a directional interference and spatially diffuse noise which is simulated using [19]. 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 5 dB and the broadband input speech-to-diffuse-noise-ratio is set to 10 dB. The "speech plus noise" signal is 13 s long and is preceded by a 7 s long "noise only" signal, which is not taken into account during evaluation.
In order to simulate RIR perturbations, the measured RIRs are perturbed by adding scaled white noise as proposed in [20], such that a desired level of normalized projection misalignment (NPM), i.e.,

$$\text{NPM} = 20\log_{10}\frac{\|\mathbf{h} - \frac{\mathbf{h}^T\hat{\mathbf{h}}}{\hat{\mathbf{h}}^T\hat{\mathbf{h}}}\hat{\mathbf{h}}\|_2}{\|\mathbf{h}\|_2}, \tag{29}$$

is generated. The considered NPM values are

$$\text{NPM} \in \{-33\text{ dB},\ -27\text{ dB},\ -21\text{ dB},\ -15\text{ dB}\}. \tag{30}$$

For all considered techniques the filter length is set equal to $L_w = \left\lceil \frac{L_h - 1}{M-1} \right\rceil = 1627$, the desired window length is set equal to $L_d = 0.01 \times f_s$, and the target EIR $\mathbf{c}_t$ is chosen as the direct path and early reflections of the perturbed first RIR $\hat{\mathbf{h}}_1$. In order to generate the L-curve and the L-hypersurface required for the automatic selection of the regularization and weighting parameters, the considered regularization and weighting parameters are

$$\delta, \mu \in \{10^{-6},\ 10^{-5},\ \ldots,\ 10^{-1},\ 1,\ 3,\ 5,\ 7\}. \tag{31}$$

Furthermore, the noise correlation matrix is estimated during the "noise only" period as

$$\mathbf{R}_{\mathbf{v}} = \frac{1}{L_v}\sum_{l=1}^{L_v}\mathbf{v}_l\mathbf{v}_l^T, \tag{32}$$

with L_v denoting the number of available "noise only" signal vectors. To avoid other sources of errors, we have assumed that a perfect voice-activity-detector is used.

6.2. Instrumental performance measures

The *dereverberation performance* is evaluated in terms of the reverberant energy suppression and perceptual speech quality improvement. As commonly done in the evaluation of acoustic multi-channel equalization techniques, the reverberant energy suppression is evaluated as the improvement in direct-to-reverberant ratio (ΔDRR) [21] between the EIR $c(n)$ and the input RIR $h_1(n)$, i.e., $\Delta\text{DRR} = \text{oDRR} - \text{iDRR}$, with

$$\text{oDRR} = 10\log_{10}\frac{\sum_{n=0}^{n_d-1} c^2(n)}{\sum_{n=n_d}^{L_c-1} c^2(n)}, \tag{33}$$

$$\text{iDRR} = 10\log_{10}\frac{\sum_{n=0}^{n_d-1} h_1^2(n)}{\sum_{n=n_d}^{L_h-1} h_1^2(n)}, \tag{34}$$

where the first n_d taps of the EIR and RIR represent the direct-path propagation. The perceptual speech quality is evaluated using the instrumental quality measure PESQ [22], which generates a similarity score between a test signal and a reference signal in the range of 1 to 4.5. The reference signal employed here is $x_{e,1}(n) = s(n) * h_{e,1}(n)$, i.e., the early reverberation 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 $x_1(n)$.

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

$$\eta_{\text{NR}} = 10\log_{10}\frac{\mathscr{E}\{v_1^2(n)\}}{\mathscr{E}\{z_v^2(n)\}}, \tag{35}$$

with $v_1(n)$ the noise component in the first microphone and $z_v(n)$ the output noise component defined in (3).

The *joint dereverberation and noise reduction performance* is evaluated in terms of the improvement in signal-to-reverberation-and-noise-ratio (ΔSRNR), i.e., ΔSRNR = oSRNR - iSRNR, with

$$\text{iSRNR} = 10\log_{10}\frac{\mathscr{E}\{x_{e,1}^2(n)\}}{\mathscr{E}\{x_{r,1}^2(n)\} + \mathscr{E}\{v_1^2(n)\}}, \tag{36}$$

$$\text{oSRNR} = 10\log_{10}\frac{\mathscr{E}\{z_e^2(n)\}}{\mathscr{E}\{z_r^2(n)\} + \mathscr{E}\{z_v^2(n)\}}, \tag{37}$$

where $x_{e,1}(n)$ and $x_{r,1}(n)$ are the early and late reverberation speech components in the first microphone defined in (2) and $z_e(n)$ and $z_r(n)$ are the early and late reverberation output speech components defined in (5) and (6).

6.3. Results

Performance of the P-MINT technique

To illustrate that the P-MINT technique generally fails to achieve dereverberation and results in additive noise amplification, in this section we only investigate the performance of the P-MINT technique. The presented performance measures are averaged over the different considered NPM values in (30). Table 1 presents the ΔDRR, ΔPESQ, η_{NR}, and ΔSRNR values obtained using the P-MINT technique for the different considered input SIRs. Since the P-MINT reshaping filter is independent of the input SIR, the obtained ΔDRR and ΔPESQ values are the same for all considered input SIRs. As illustrated, the P-MINT technique fails to achieve dereverberation, worsening the DRR and the PESQ score by 10.8 dB and 0.4 respectively. Furthermore, the noise reduction factors presented in Table 1 shows that the additive noise is significantly amplified, which is to be expected since the P-MINT reshaping filter is designed without taking the noise statistics into account. Since the P-MINT technique fails to achieve dereverberation and amplifies the additive noise, it results in a significantly worse SRNR value than in the input signal, as illustrated by the large negative ΔSRNR values presented in Table 1. These simulation results confirm that when the RIR perturbations are not taken into account, acoustic multi-channel equalization techniques fail to achieve dereverberation. Furthermore, they confirm that designing reshaping filters for speech dereverberation without taking the presence of the additive noise into account results in a large noise amplification.

Performance of the RP-MINT and RP-DNR techniques

In this section the performance of the automatically regularized and parametrized RP-MINT and RP-DNR techniques is investigated. Similarly as before, the presented performance measures are averaged over the different considered NPM values in (30). Fig. 3 depicts the performance of considered techniques in terms of the ΔDRR, ΔPESQ, η_{NR}, and ΔSRNR measures. It can be observed in Fig. 3a that the ΔDRR obtained using the RP-MINT and RP-DNR techniques is very similar, with an insignificant difference of at most 0.05 dB for low input SIRs. Furthermore, Fig. 3b shows that also the PESQ score obtained using the RP-MINT and RP-DNR techniques is very similar, with an insignificant difference of at most 0.02 for low input SIRs. These results show that the dereverberation performance of the proposed RP-DNR technique is very similar to the dereverberation performance of the RP-MINT technique. Although one would expect the dereverberation performance of the RP-DNR technique to be lower than that of the RP-MINT technique, this is not the case in these simulation results. This occurs due to the automatic selection of the regularization parameter in the RP-MINT technique, which does not yield the best dereverberation performance one would otherwise obtain by intrusively selecting the regulariza-

Table 1: Average performance of the P-MINT technique for several input SIR.

Input SIR [dB]	−5	−2.5	0	2.5	5
ΔDRR [dB]			−10.8		
ΔPESQ			−0.4		
η_{NR} [dB]	−28.8	−28.7	−28.4	−28.3	−28.0
ΔSRNR [dB]	−13.1	−12.7	−12.1	−11.2	−10.1

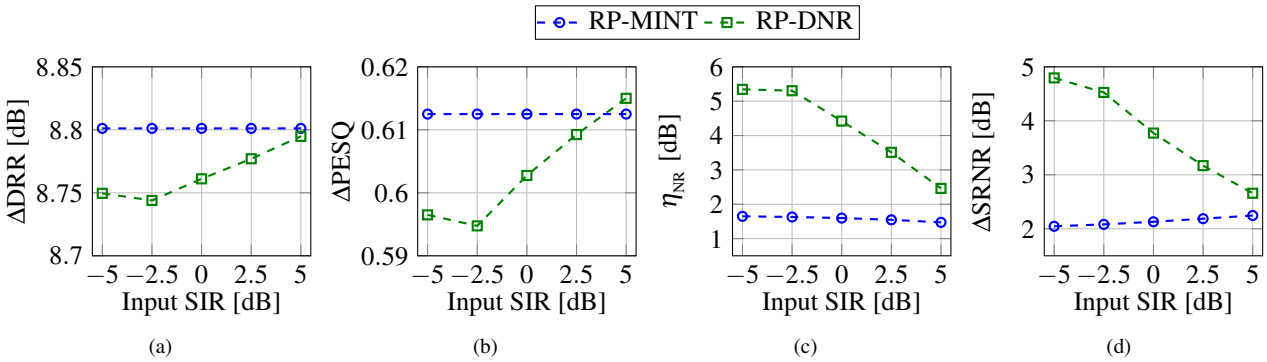

Fig. 3: Average performance of the automatically regularized P-MINT and the automatically parametrized RP-DNR techniques for several input SIR in terms of (a) ΔDRR, (b) ΔPESQ, (c) η_{NR}, and (d) ΔSRNR.

tion parameter. While the dereverberation performance of both techniques is very similar, it can be observed in Fig. 3c that the noise reduction factor obtained using the RP-DNR technique is up to 4 dB higher than the noise reduction factor obtained using the RP-MINT technique. Furthermore, Fig. 3c also shows that the incorporation of regularization in acoustic multi-channel equalization techniques avoids the significantly large noise amplification one would otherwise obtain (cf. Table 1). This occurs due to the decrease in the reshaping filter energy when regularization is incorporated, which is also partly effective in reducing the distortions in the output signal arising due to the additive noise. However, it should be noted that there is no guarantee that any noise reduction can be achieved using the RP-MINT technique, since the actual noise statistics are not explicitly taken into account. The very similar dereverberation performance but higher noise reduction performance of the RP-DNR technique in comparison to the RP-MINT technique is also reflected in the ΔSRNR values depicted in Fig. 3d, with the RP-DNR technique yielding a higher ΔSRNR of up to 3 dB when compared to the RP-MINT technique.

In summary, these simulation results demonstrate the importance of taking the noise statistics into account in order to achieve joint dereverberation and noise reduction. Furthermore they show that the proposed RP-DNR technique does not sacrifice the high dereverberation performance of the RP-MINT technique, but improves the noise reduction as well as the joint dereverberation and noise reduction performance.

7. CONCLUSION

In this paper we have proposed the RP-DNR technique which aims at joint dereverberation and noise reduction based on acoustic multi-channel equalization. The RP-DNR technique directly extends the RP-MINT technique by explicitly taking the noise statistics into account. In addition to the regularization parameter used in the RP-MINT technique, the RP-DNR technique introduces an additional weighting parameter, enabling to trade off between dereverberation and noise reduction. We have also proposed an automatic non-intrusive procedure based on the L-hypersurface for selecting the regularization and weighting parameters. Simulation results demonstrate that the RP-DNR technique maintains the high dereverberation performance of the RP-MINT technique, while

improving the noise reduction as well as the joint dereverberation and noise reduction performance.

8. REFERENCES

[1] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, Jul. 2006.

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

[3] R. Maas, E. A. P. Habets, A. Sehr, and W. Kellermann, "On the application of reverberation suppression to robust speech recognition," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 297–300.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[5] M. Kallinger and A. Mertins, "Multi-channel room impulse response shaping - a study," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006, pp. 101–104.

[6] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Tel Aviv, Israel, Sep. 2010.

[7] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.

[8] F. Lim, W. Zhang, E. A. P. Habets, and P. A. Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1379–1390, Jun. 2014.

[9] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

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

[11] 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 Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 888–899, Mar. 2012.

[12] I. Kodrasi and S. Doclo, "Joint dereverberation and noise reduction based on acoustic multi-channel equalization," *IEEE Transactions on Audio, Speech, and Language Processing*, 2015, submitted.

[13] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[14] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.

[15] 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*, vol. 328, Sep. 1998.

[16] ——, "Efficient determination of multiple regularization parameters in a generalized L-curve framework," *Inverse Problems*, vol. 18, pp. 1161–1183, Jul. 2002.

[17] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sep. 2014, pp. 313–317.

[18] 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," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[19] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

[20] W. Zhang and P. A. Naylor, "An algorithm to generate representations of system identification errors," *Research Letters in Signal Processing*, vol. 2008, Jan. 2008.

[21] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. London, UK: Springer, 2010.

[22] 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*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.