JOINT DEREVERBERATION AND NOISE REDUCTION BASED ON ACOUSTIC MULTICHANNEL EQUALIZATION

Ina Kodrasi, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics,
and Cluster of Excellence Hearing4All, Oldenburg, Germany
{ina.kodrasi, simon.doclo}@uni-oldenburg.de

ABSTRACT

In many speech communication applications, the recorded microphone signals are often corrupted by both reverberation and noise, which can significantly impair speech quality and intelligibility. While acoustic multichannel equalization techniques can achieve a high dereverberation performance, they may lead to amplification of the additive noise, since the equalization filters are typically designed without taking the presence of noise into account.

This paper presents a novel approach to joint dereverberation and noise reduction based on acoustic multichannel equalization (DeNoREq). DeNoREq produces a weighted minimum mean-square error estimate of the clean speech signal, where the weighting parameter trades off between dereverberation and noise reduction performance. Furthermore, an automatic procedure for the selection of the weighting parameter is established. Experimental results for perfectly and erroneously estimated room impulse responses illustrate the effectiveness of the proposed technique in achieving a high dereverberation and noise reduction performance.

Index Terms— dereverberation, noise reduction, acoustic multichannel equalization, robustness, P-MINT

1. INTRODUCTION

Speech signals recorded in an enclosed space by microphones placed at a distance from the source are often corrupted by reverberation and background noise, which typically degrade speech quality, impair speech intelligibility, and decrease the performance of automatic speech recognition systems [1, 2, 3]. In order to mitigate these detrimental effects, algorithms aiming at joint dereverberation and noise reduction have been proposed [4, 5, 6, 7, 8, 9]. In this paper, we focus on the effective integration of the dereverberation and noise reduction tasks using acoustic multichannel equalization.

Acoustic multichannel equalization techniques [10, 11, 12, 13], which are based on estimating and reshaping the room impulse responses (RIRs) between the source and the microphone array, comprise an attractive approach to speech dereverberation, since in theory perfect dereverberation can be achieved [10, 11]. However, in practice such techniques suffer from several drawbacks. Firstly, acoustic multichannel equalization techniques typically design reshaping filters aiming only at speech dereverberation, without taking the presence of additive background noise into account. Applying such reshaping filters may result in a large noise amplification [10], hindering the high dereverberation performance potential. Secondly, the estimated RIRs generally differ from the true ones [14, 15, 16],

such that reshaping filters based on these estimates may even fail to achieve dereverberation, yielding speech distortion in the processed output signal [10, 12, 17].

In order to increase robustness to RIR estimation errors, several techniques have been proposed [10, 12, 17, 18], with the regularized partial multichannel equalization technique based on the multiple-input/output inverse theorem (P-MINT) shown to yield a high reverberant tail suppression and perceptual speech quality preservation [10]. By incorporating regularization in P-MINT, the energy of the reshaping filter is decreased, reducing the distortions in the output signal due to RIR estimation errors and increasing the dereverberation performance [10]. While the regularization parameter introduced in P-MINT is also effective in partly avoiding noise amplification, the noise reduction performance is limited since the noise statistics are not explicitly taken into account.

In this paper, a novel technique is proposed which aims at joint Dereverberation and Noise Reduction based on acoustic multichannel Equalization (DeNoREq). DeNoREq produces a weighted minimum mean-square error (MMSE) estimate of the dereverberated speech, where the weighting parameter allows to trade off between dereverberation error energy and output noise energy. The weighting parameter is automatically computed using the L-curve based procedure proposed in [10]. Furthermore, in order to avoid the estimation of the clean speech correlation matrix, the reshaping filter designed from equalization techniques is exploited. Experimental results illustrate the effectiveness of the proposed technique in achieving joint dereverberation and noise reduction.

2. CONFIGURATION AND NOTATION

Consider an acoustic system with a single speech source and M microphones. The m-th microphone signal, $m = 1, \ldots, M$, at time index n is given by $y_m(n) = x_m(n) + v_m(n) = s(n) * h_m(n) + v_m(n)$, where $x_m(n)$ and $v_m(n)$ denote the reverberant speech and noise components respectively, $*$ denotes convolution, $s(n)$ is the clean speech signal, and $h_m(n)$ denotes the RIR between the source and the m-th microphone. The RIR can be described in vector notation as $\mathbf{h}_m = [h_m(0) \; h_m(1) \; \ldots \; h_m(L_h - 1)]^T$, with L_h the RIR length and $[\cdot]^T$ denoting the transpose operation. Applying filters $g_m(n)$ of length L_g, i.e., $\mathbf{g}_m = [g_m(0) \; g_m(1) \; \ldots \; g_m(L_g - 1)]^T$, the system output signal $z(n)$ is given by

$$z(n) = \sum_{m=1}^{M} y_m(n) * g_m(n) \tag{1}$$

$$= s(n) * \underbrace{\sum_{m=1}^{M} h_m(n) * g_m(n)}_{c(n)} + \sum_{m=1}^{M} v_m(n) * g_m(n), \tag{2}$$

where $c(n)$ denotes the equalized impulse response (EIR) between the clean speech signal and the output speech component. The EIR can be described in vector notation as $\mathbf{c} = [c(0) \ c(1) \ \ldots \ c(L_c - 1)]^T$, with $L_c = L_h + L_g - 1$ the EIR length. Using the ML_g–dimensional stacked filter vector $\mathbf{g}$, i.e., $\mathbf{g} = \left[\mathbf{g}_1^T \ \mathbf{g}_2^T \ \ldots \ \mathbf{g}_M^T\right]^T$, the output signal can be expressed as $z(n) = \mathbf{g}^T \mathbf{y}(n)$, with

$$\mathbf{y}(n) = [\mathbf{y}_1^T(n) \ \mathbf{y}_2^T(n) \ \ldots \ \mathbf{y}_M^T(n)]^T, \tag{3}$$

$$\mathbf{y}_m(n) = [y_m(n) \ y_m(n-1) \ \ldots \ y_m(n - L_g + 1)]^T. \tag{4}$$

Defining the reverberant speech and noise vectors $\mathbf{x}(n)$ and $\mathbf{v}(n)$ similarly as $\mathbf{y}(n)$ in (3) and (4), $z(n)$ can also be expressed as

$$\boxed{z(n) = \mathbf{g}^T \mathbf{x}(n) + \mathbf{g}^T \mathbf{v}(n) = \underbrace{(\mathbf{Hg})^T}_{\mathbf{c}} \mathbf{s}(n) + \mathbf{g}^T \mathbf{v}(n)} \tag{5}$$

with $\mathbf{H}$ the $L_c \times ML_g$–dimensional multichannel convolution matrix, $\mathbf{s}(n) = [s(n) \ s(n-1) \ \ldots \ s(n - L_c + 1)]^T$, and $\mathbf{x}(n) = \mathbf{H}^T \mathbf{s}(n)$.

As described in the following section, equalization techniques disregard the presence of the additive noise $\mathbf{v}(n)$ and design $\mathbf{g}$ such that only the EIR $\mathbf{c}$ is optimized. Furthermore, since the true RIRs are typically not available in practice, the estimated multichannel convolution matrix $\hat{\mathbf{H}}$ (constructed from the estimated RIRs $\hat{h}_m(n)$) is used for the reshaping filter design.

3. ACOUSTIC MULTICHANNEL EQUALIZATION

In this paper, we will focus on the partial multichannel equalization technique based on MINT which aims at setting the reverberant tail of the EIR to $\mathbf{0}$, while controlling the remaining taps corresponding to the direct path and early reflections [10]. To accomplish this objective, the least-squares cost function

$$\boxed{J_{\mathrm{P}} = \|\hat{\mathbf{H}}\mathbf{g} - \mathbf{c}_t\|_2^2} \tag{6}$$

is minimized, where the direct path and early reflections of the target EIR $\mathbf{c}_t$ are set to the direct path and early reflections of one of the estimated RIRs, i.e., $\mathbf{c}_t = [\hat{h}_p(0) \ \ldots \ \hat{h}_p(L_d - 1) \ 0 \ \ldots \ 0]^T$, with $p \in \{1, \ldots, M\}$ and L_d denoting the number of the EIR taps to be preserved. Assuming that the estimated RIRs do not share any common zeros and that $L_g \geq \lceil \frac{L_h - 1}{M - 1} \rceil$, the P-MINT reshaping filter minimizing (6) is equal to [10]

$$\boxed{\mathbf{g}_{\mathrm{P}} = \hat{\mathbf{H}}^+ \mathbf{c}_t} \tag{7}$$

with $\{\cdot\}^+$ denoting the Moore-Penrose pseudo-inverse [19]. For perfectly estimated RIRs, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the P-MINT reshaping filter yields perfect dereverberation when applied to the received microphone signals [10], i.e.,

$$\boxed{\mathbf{c}_t = \mathbf{Hg}_{\mathrm{P}}} \tag{8}$$

However, for erroneously estimated RIRs the P-MINT reshaping filter may result in large speech distortion. In order to increase the robustness of P-MINT to RIR estimation errors, the automatically regularized P-MINT technique has been proposed [10], which minimizes the regularized least-squares cost function

$$\boxed{J_{\mathrm{RP}} = \|\hat{\mathbf{H}}\mathbf{g} - \mathbf{c}_t\|_2^2 + \delta \|\mathbf{g}\|_2^2} \tag{9}$$

with δ being a regularization parameter. The filter minimizing (9) is equal to

$$\boxed{\mathbf{g}_{\mathrm{RP}} = (\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \delta \mathbf{I})^{-1} \hat{\mathbf{H}}^T \mathbf{c}_t} \tag{10}$$

where δ is automatically computed using the procedure proposed in [10]. Introducing a regularization parameter reduces the energy of the reshaping filter, hence, reducing speech distortion in the output signal due to RIR estimation errors. While the P-MINT filter typically fails to achieve dereverberation in the presence of estimation errors, i.e., $\mathbf{c}_t \neq \mathbf{Hg}_{\mathrm{P}}$, the regularized P-MINT filter results in a significantly better dereverberation performance [10], i.e.,

$$\boxed{\mathbf{c}_t \approx \mathbf{Hg}_{\mathrm{RP}}} \tag{11}$$

Furthermore, decreasing the energy of the reshaping filter by means of regularization is also effective in partly avoiding the (otherwise large) noise amplification at the output of the system [10] (cf. Section 5). However, the noise reduction performance of the automatically regularized P-MINT filter in (10) is limited since the noise statistics are not explicitly taken into account. In the following, a novel technique is proposed which aims at estimating the dereverberated and denoised speech signal $\mathbf{c}_t^T \mathbf{s}(n)$.

4. DEREVERBERATION AND NOISE REDUCTION BASED ON MULTICHANNEL EQUALIZATION (DENOREQ)

Aiming at joint dereverberation and noise reduction, an MMSE estimate of the desired signal $\mathbf{c}_t^T \mathbf{s}(n)$ can be obtained by minimizing the cost function

$$J = \mathcal{E}\{[z(n) - \mathbf{c}_t^T \mathbf{s}(n)]^2\} \tag{12}$$

$$= \mathcal{E}\{[\mathbf{g}^T \mathbf{x}(n) + \mathbf{g}^T \mathbf{v}(n) - \mathbf{c}_t^T \mathbf{s}(n)]^2\}, \tag{13}$$

with $\mathcal{E}$ denoting the expected value operator. The cost function in (13) is similar to the well-known multichannel Wiener filter cost function for noise reduction [20], with the difference consisting in the desired signal being the denoised and dereverberated speech signal. Assuming that the speech and noise components are uncorrelated and introducing a weighting parameter μ to trade off between dereverberation and noise reduction, the cost function of the proposed dereverberation and noise reduction technique based on equalization (DeNoREq) can be written as

$$\boxed{J_{\mathrm{DeNoREq}} = \mathcal{E}\{[\mathbf{g}^T \mathbf{x}(n) - \mathbf{c}_t^T \mathbf{s}(n)]^2\} + \mu \mathcal{E}\{[\mathbf{g}^T \mathbf{v}(n)]^2\}} \tag{14}$$

Increasing the parameter μ in (14) increases the noise reduction performance at the expense of decreased dereverberation performance. The filter minimizing (14) is equal to

$$\mathbf{g}_{\mathrm{DeNoREq}} = (\mathbf{R}_{\mathbf{x}} + \mu \mathbf{R}_{\mathbf{v}})^{-1} \mathcal{E}\{\mathbf{x}(n)\mathbf{s}^T(n)\}\mathbf{c}_t \tag{15}$$

$$= (\mathbf{R}_{\mathbf{x}} + \mu \mathbf{R}_{\mathbf{v}})^{-1} \mathbf{H}^T \mathbf{R}_{\mathbf{s}} \mathbf{c}_t, \tag{16}$$

with $\mathbf{R}_{\mathbf{s}} = \mathcal{E}\{\mathbf{s}(n)\mathbf{s}^T(n)\}$ the $L_c \times L_c$–dimensional correlation matrix of the clean speech signal $\mathbf{s}(n)$ and

$$\mathbf{R}_{\mathbf{x}} = \mathcal{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\} = \mathbf{H}^T \mathbf{R}_{\mathbf{s}} \mathbf{H}, \tag{17}$$

$$\mathbf{R}_{\mathbf{v}} = \mathcal{E}\{\mathbf{v}(n)\mathbf{v}^T(n)\}, \tag{18}$$

the $ML_g \times ML_g$–dimensional correlation matrices of the reverberant speech and noise respectively. Hence, in order to compute the filter in (16), an estimate of the clean speech, reverberant speech, and noise correlation matrices is required. While the estimation of $\mathbf{R}_{\mathbf{x}}$ and $\mathbf{R}_{\mathbf{v}}$ can in practice be done e.g., using a voice activity detector (VAD) [21], estimating the clean speech correlation matrix $\mathbf{R}_{\mathbf{s}}$ is

not at all trivial. However, as described in Section 3, for perfectly estimated RIRs the P-MINT reshaping filter yields perfect dereverberation performance (cf. (8)). Using (8) and (17), the following equality

$$\mathbf{H}^T \mathbf{R_s} \mathbf{c}_t = \mathbf{R_x} \mathbf{g}_{\mathrm{P}} \tag{19}$$

can be derived. Exploiting the equality in (19), the filter minimizing the DeNoREq cost function in (14) can hence be computed as

$$\mathbf{g}_{\mathrm{DeNoREq}}^{\mathrm{P}} = (\mathbf{R_x} + \mu \mathbf{R_v})^{-1} \mathbf{R_x} \mathbf{g}_{\mathrm{P}} \tag{20}$$

which does not require an estimate of the clean speech correlation matrix $\mathbf{R_s}$. In the presence of RIR estimation errors, the P-MINT reshaping filter however fails to achieve dereverberation. As a result, the DeNoREq filter in (20) is expected to inherit the sensitivity of P-MINT to estimation errors (cf. Section 5), since the equality in (19) used for computing $\mathbf{g}_{\mathrm{DeNoREq}}^{\mathrm{P}}$ does not hold. On the other hand, the regularized P-MINT filter provides a better approximation to $\mathbf{c}_t$ (cf. (11)). Hence, to increase robustness in the presence of RIR estimation errors we propose to compute the DeNoREq filter as[1]

$$\mathbf{g}_{\mathrm{DeNoREq}}^{\mathrm{RP}} = (\mathbf{R_x} + \mu \mathbf{R_v})^{-1} \mathbf{R_x} \mathbf{g}_{\mathrm{RP}} \tag{21}$$

Clearly, the dereverberation and noise reduction performance of the proposed technique depends on the weighting parameter μ, which introduces a trade off between the dereverberation error energy $\epsilon_s^2 = \mathcal{E}\{[\mathbf{g}^T\mathbf{x}(n) - \mathbf{c}_t^T\mathbf{s}(n)]^2\}$ and the noise output energy $\epsilon_v^2 = \mathcal{E}\{[\mathbf{g}^T\mathbf{v}]^2\}$, with $\mathbf{g}$ being the filter computed in (20) or (21). In order to automatically select a weighting parameter μ, we use the procedure proposed in [10] for the automatic selection of the regularization parameter in multichannel equalization techniques, which requires a parametric plot of ϵ_v^2 versus ϵ_s^2 for a set of weighting parameters μ. Since this parametric plot has an L-shape, the weighting parameter μ can be automatically selected as the one corresponding to the point of maximum curvature, i.e., the corner of the L-curve, such that both the dereverberation error and the noise output energies are kept small. The noise output energy can be computed as $\epsilon_v^2 = \mathbf{g}^T \mathbf{R_v} \mathbf{g}$, whereas the computation of ϵ_s^2 requires an estimate of the cross-correlation between the reverberant and clean speech, which is not available in practice. Instead, we propose computing the dereverberation error energy based on the deviation of the resulting EIR from the target EIR $\mathbf{c}_t$. Since in practice only the estimated multichannel convolution matrix $\hat{\mathbf{H}}$ is available, the dereverberation error energy is computed as $\epsilon_s^2 = \|\hat{\mathbf{H}}\mathbf{g} - \mathbf{c}_t\|_2^2$.

Fig. 1 depicts a typical L-curve obtained using DeNoREq based on regularized P-MINT for a perfectly estimated acoustic system. As illustrated in this figure, increasing the value of μ decreases the output noise energy, hence increasing the noise reduction performance. However, increasing μ increases the dereverberation error energy, hence decreasing the dereverberation performance. Although from such a curve it seems intuitively easy to determine the weighting parameter that corresponds to the maximum curvature, a numerically stable algorithm is needed to detect it automatically. In this work, the triangle method [22] is used.

5. SIMULATION RESULTS

In the following, the dereverberation and noise reduction performance when using the P-MINT filter in (7), the regularized P-MINT filter in (10), the DeNoREq filter based on P-MINT in (20), and the DeNoREq filter based on regularized P-MINT in (21) will be evaluated.

---

[1]The DeNoREq filter can be computed using any robust acoustic multichannel equalization technique, as long as the reshaping filter resulting from the equalization technique provides a high dereverberation performance.
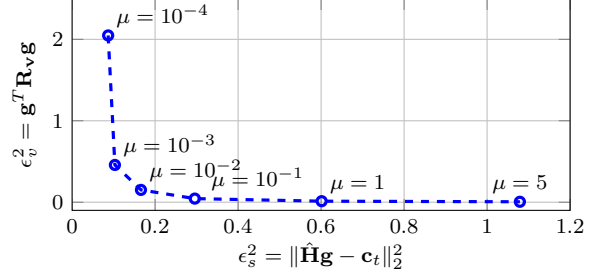
Fig. 1: Typical L-curve obtained using DeNoREq based on regularized P-MINT for a perfectly estimated acoustic system

5.1. Simulation parameters and performance measures

We have considered an acoustic scenario with a single speech source, a directional noise source, and a linear microphone array with $M = 4$ equidistant microphones in a room with reverberation time $T_{60} \approx 450$ ms. The distance between the microphones is 5 cm and the distance between the sources and the microphone array is 1 m. The speech source is located in front of the microphone array and the noise source is positioned at an angle of about $70°$, i.e, on the left of the microphone array. Measured RIRs from the MARDY database [23] have been used, with $L_h = 3600$ at a sampling frequency $f_s = 8$ kHz. In order to simulate RIR estimation errors, the RIRs have been perturbed by adding scaled white noise as proposed in [24], such that a normalized projection misalignment (NPM) defined as

$$\mathrm{NPM} = 10 \log_{10} \frac{\|\mathbf{h} - \frac{\mathbf{h}^T \hat{\mathbf{h}}}{\hat{\mathbf{h}}^T \hat{\mathbf{h}}}\hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2} \tag{22}$$

is generated. Although several NPMs have been investigated, only NPM $= -\infty$ dB (i.e., perfectly estimated RIRs) and NPM $= -33$ dB have been considered in this paper due to space constraints. The simulation parameters are set to $L_g = 1200$, $p = 1$, $L_d = 0.03 \times f_s$ (i.e., 30 ms), and the input signal-to-noise ratio (SNR$_i$) of the noisy first microphone signal is 0 dB, i.e.,

$$\mathrm{SNR_i} = 10 \log_{10} \frac{\sum_n [x_1(n)]^2}{\sum_n [v_1(n)]^2} = 0 \text{ dB}. \tag{23}$$

To avoid other sources of errors, we have put aside the influence of the VAD in this paper and estimated the reverberant speech and noise correlation matrices as long-term sample averages of $\mathbf{x}(n)$ and $\mathbf{v}(n)$ respectively. In order to generate the parametric L-curve required for the automatic selection of μ, the set of considered weighting parameters is $\mu \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 5\}$.

The dereverberation performance of the considered techniques is evaluated using the energy decay curve (EDC) of the resulting EIR [25] defined as

$$\mathrm{EDC}(n) = 10 \log_{10} \frac{1}{\|\mathbf{c}\|_2^2} \sum_{i=n}^{L_c - 1} c^2(i), \; n = 0, \ldots, L_c - 1, \tag{24}$$

where $\mathbf{c} = \mathbf{H}\mathbf{g}$ and the filter $\mathbf{g}$ is designed using the estimated multichannel convolution matrix $\hat{\mathbf{H}}$. In order to evaluate the noise reduction performance, the SNR improvement (ΔSNR) is computed, i.e.,

$$\Delta\mathrm{SNR} = \mathrm{SNR_o} - \mathrm{SNR_i}, \quad \mathrm{SNR_o} = 10 \log_{10} \frac{\sum_n [z_x(n)]^2}{\sum_n [z_v(n)]^2}, \tag{25}$$

with $z_x(n)$ and $z_v(n)$ the speech and noise components of the output signal. It should be noted that when aiming at dereverberation,
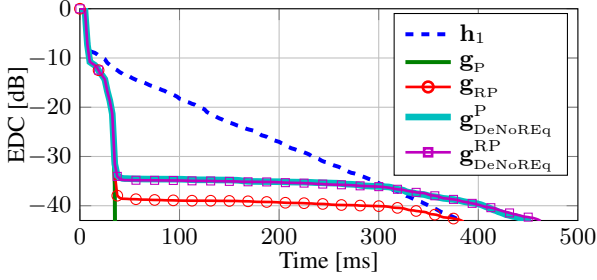
Fig. 2: EDC of $\mathbf{h}_1$ and EDCs of the EIRs obtained using P-MINT, regularized P-MINT, DeNoREq based on P-MINT, and DeNoREq based on regularized P-MINT for NPM $= -\infty$ dB

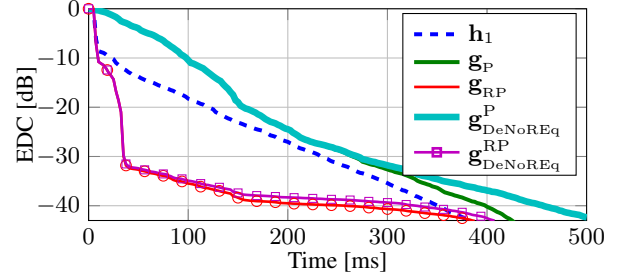| Measure [dB]	$\mathbf{g}_\mathrm{P}$	$\mathbf{g}_\mathrm{RP}$	$\mathbf{g}_\mathrm{DeNoREq}^\mathrm{P}$	$\mathbf{g}_\mathrm{DeNoREq}^\mathrm{RP}$
ΔSNR	-53.0	-1.4	28.3	**28.9**
η_NR	-52.8	-1.2	28.5	**29.1**
η_SR	**0.2**	**0.2**	**0.2**	**0.2**

Table 1: Performance of P-MINT, regularized P-MINT, DeNoREq based on P-MINT, and DeNoREq based on regularized P-MINT for NPM $= -\infty$ dB

some speech reduction is intrinsically introduced. In order to separate the influence of noise reduction and speech reduction on the SNR improvement, also the noise reduction factor η_NR and speech reduction factor η_SR are computed, with

$$\eta_\mathrm{NR} = 10 \log_{10} \frac{\sum_n [v_1(n)]^2}{\sum_n [z_v(n)]^2}, \quad \eta_\mathrm{SR} = 10 \log_{10} \frac{\sum_n [x_1(n)]^2}{\sum_n [z_x(n)]^2}. \quad (26)$$

5.2. Results

To evaluate the dereverberation performance for perfectly estimated RIRs, Fig. 2 depicts the EDC of the true RIR $\mathbf{h}_1$ and the EDCs of the EIRs obtained using all considered techniques for NPM $= -\infty$ dB. For both DeNoREq filters, the automatically determined weighting parameter is $\mu = 10^{-2}$. As expected, P-MINT yields perfect dereverberation, with the reverberant tail fully suppressed, whereas regularized P-MINT achieves a high dereverberation performance but only partly suppresses the reverberant tail. Furthermore, also the proposed DeNoREq technique achieves a high level of dereverberation, since the reverberant tail is suppressed by more than 35 dB and the artificial tail introduced after about 300 ms is not audible. Both DeNoREq filters yield a similar dereverberation performance in this scenario, with the reverberant tails being approximately 5 dB higher than the reverberant tail obtained using the regularized P-MINT filter. Some performance degradation in terms of dereverberation when using DeNoREq in comparison to P-MINT is expected, since DeNoREq aims at simultaneously suppressing the additive noise as well. Furthermore, Table 1 presents the SNR improvement, noise reduction, and speech reduction factors of the considered techniques. It can be seen that the P-MINT filter significantly amplifies the noise, decreasing the SNR at the output by 53.0 dB. Although the regularized P-MINT technique partly avoids the noise amplification in comparison to P-MINT, the SNR at the output is still decreased by 1.4 dB. On the other hand, the proposed DeNoREq technique achieves a high level of noise reduction, with the filter computed based on regularized P-MINT, i.e., $\mathbf{g}_\mathrm{DeNoREq}^\mathrm{RP}$, yielding the highest SNR improvement of 28.9 dB. Based on these simulation results it can be said that the proposed DeNoREq technique simultaneously achieves dereverberation and noise reduction, whereas the P-MINT-based equalization techniques achieve dereverberation but amplify the noise.

Fig. 3: EDC of $\mathbf{h}_1$ and EDCs of the EIRs obtained using P-MINT, regularized P-MINT, DeNoREq based on P-MINT, and DeNoREq based on regularized P-MINT for NPM $= -33$ dB

Measure [dB]	$\mathbf{g}_\mathrm{P}$	$\mathbf{g}_\mathrm{RP}$	$\mathbf{g}_\mathrm{DeNoREq}^\mathrm{P}$	$\mathbf{g}_\mathrm{DeNoREq}^\mathrm{RP}$
ΔSNR	-22.9	-0.4	24.2	**26.0**
η_NR	-32.2	-0.1	14.9	**26.3**
η_SR	-9.3	**0.3**	-9.3	**0.3**

Table 2: Performance of P-MINT, regularized P-MINT, DeNoREq based on P-MINT, and DeNoREq based on regularized P-MINT for NPM $= -33$ dB

In order to evaluate the performance in the presence of RIR estimation errors, Fig. 3 depicts the EDC of the true RIR $\mathbf{h}_1$ and the EDCs of the EIRs obtained using all considered techniques for an NPM $= -33$ dB. The automatically determined weighting parameter is $\mu = 10^{-2}$ for $\mathbf{g}_\mathrm{DeNoREq}^\mathrm{P}$ and $\mu = 10^{-3}$ for $\mathbf{g}_\mathrm{DeNoREq}^\mathrm{RP}$. As expected, the P-MINT technique completely fails to achieve dereverberation, whereas the regularized P-MINT technique is significantly more robust, resulting in a high level of reverberant tail suppression. Since $\mathbf{g}_\mathrm{DeNoREq}^\mathrm{P}$ is computed using the P-MINT filter, it inherits its sensitivity to RIR estimation errors and hence also fails to achieve dereverberation. On the other hand, the DeNoREq filter computed using the regularized P-MINT technique is robust and achieves a high dereverberation performance, with a very similar reverberant tail suppression as the regularized P-MINT technique. To evaluate the noise reduction performance, Table 2 also depicts the SNR improvement, noise reduction factor, and speech reduction factor obtained by all considered techniques. Similarly as in the case of perfectly estimated RIRs, the P-MINT and regularized P-MINT techniques result in noise amplification. Furthermore, the DeNoREq filter computed using P-MINT, i.e., $\mathbf{g}_\mathrm{DeNoREq}^\mathrm{P}$, improves the SNR at the output by 24.2 dB but significantly distorts the speech as shown by the speech reduction factor of -9.3 dB. On the other hand, the DeNoREq filter based on regularized P-MINT is robust and achieves a high level of noise reduction also in the presence of RIR estimation errors, with an SNR improvement of 26.0 dB. Summarizing these simulation results, it can be observed that the DeNoREq technique using the robust regularized P-MINT filter results in a high dereverberation and noise reduction performance, both for perfectly estimated RIRs as well as in the presence of RIR estimation errors.

6. CONCLUSION

In this paper, we have presented a novel approach to joint dereverberation and noise reduction based on acoustic multichannel equalization (DeNoREq). DeNoREq produces a weighted MMSE estimate of the dereverberated speech signal, with the weighting parameter automatically selected such that both a high dereverberation and noise reduction performance is achieved. Experimental results demonstrate that unlike equalization techniques which amplify the noise, DeNoREq yields a high dereverberation performance while significantly suppressing the additive noise.

7. 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, July 2006.

[2] A. Sehr, *Reverberation Modeling for Robust Distant-Talking Speech Recognition*, Ph.D. thesis, 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] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Darmstadt, Germany, Sept. 2001, pp. 31–34.

[5] M. Delcroix, T. Hikichi, and M. Miyoshi, "Dereverberation and denoising using multichannel linear prediction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1791–1801, Aug 2007.

[6] T. Yoshioka, T. Nakatani, and M. Miyoshi, "Integrated speech enhancement method using noise suppression and dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 231–246, Feb 2009.

[7] E. A. P. Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 945–958, May 2013.

[8] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sept. 2013.

[9] 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," in *Proc. REVERB Workshop*, Florence, Italy, May 2014.

[10] 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, Sept. 2013.

[11] 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.

[12] 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, Sept. 2010.

[13] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity- and p-norm optimization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

[14] 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.

[15] K. Hasan and P. A. Naylor, "Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: robustness issue," in *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sept. 2006.

[16] 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.

[17] F. Lim and P. A. Naylor, "Robust speech dereverberation using subband multichannel least squares with variable relaxation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sept. 2013.

[18] F. Lim, M. R. P. Thomas, and P. A. Naylor, "MINTFormer: A spatially aware channel equalizer," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, Oct. 2013.

[19] G. Golub and C. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, Maryland, 3 edition, 1996.

[20] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1218–1234, July 2006.

[21] J. Ramírez, J. C. Segura, C. Benítez, Á. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," *Speech Communication*, vol. 42, no. 3-4, pp. 271–287, Oct. 2004.

[22] J. L. Castellanos, S. Gómez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Applied Numerical Mathematics*, vol. 43, no. 4, pp. 359–373, Dec. 2002.

[23] J. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Paris, France, Sept. 2006.

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

[25] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer, New York, NY, USA, 2010.