

# REGULARIZED SUBSPACE-BASED ACOUSTIC MULTICHANNEL EQUALIZATION FOR SPEECH DEREVERBERATION

Ina Kodrasi, Simon Doclo

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

## ABSTRACT

Acoustic multichannel equalization techniques aim at designing robust reshaping filters which reduce the reverberant energy and preserve the perceptual speech quality. Although the recently proposed channel shortening (CS) technique can achieve reverberant energy suppression, its optimization criterion results in a subspace of solutions where each solution yields a different perceptual speech quality.

In this paper, we propose a perceptually advantageous subspace-based equalization technique, where the reshaping filter is constrained to lie in the subspace of the multiple CS solutions. Since it has been shown that regularization increases the robustness of equalization techniques to errors in the estimated room impulse responses (RIRs), a regularization term is incorporated in order to control the energy of the reshaping filter. Simulation results for erroneously estimated RIRs demonstrate the advantages of the proposed approach.

**Index Terms**— acoustic multichannel equalization, speech dereverberation, channel shortening

## 1. INTRODUCTION

Enhancement of speech quality and intelligibility in reverberant rooms is an important research topic for several applications such as hearing aids and teleconferencing applications. In order to mitigate the detrimental effects of reverberation, several acoustic multichannel equalization techniques have been investigated [1, 2, 3, 4]. Such techniques aim at speech dereverberation by designing filters which reshape the estimated room impulse responses (RIRs) between the source and the microphone array. While in theory the RIRs can be perfectly reshaped to a desired target response [2, 4], in practice large speech distortion can arise due to RIR estimation errors [5].

The regularized partial multichannel equalization technique based on the multiple-input/output inverse theorem (P-MINT) [2] and the relaxed multichannel least squares technique with constrained initial taps (RMCLS-CIT) [3] aim at

designing robust reshaping filters which simultaneously suppress the reverberant energy and control the perceptual speech quality. It has been experimentally validated that the regularized P-MINT technique yields a high perceptual speech quality, however, its performance in terms of reverberant energy suppression is generally lower than other state-of-the-art techniques. On the other hand, the RMCLS-CIT technique achieves a higher level of reverberant energy suppression, but only partly controls the perceptual speech quality.

The recently proposed channel shortening (CS) technique [1, 6] aims at suppressing the reverberant energy without imposing any further constraints on the perceptual speech quality. Furthermore, its optimization criterion results in a subspace of solutions, which all lead to a different performance [1, 2], posing an ambiguity problem in selecting the optimal reshaping filter in practice. In order to overcome the selection ambiguity of CS, a perceptually constrained channel shortening technique (PeCCS) has recently been proposed [7].

The objective in this paper is to extend PeCCS and establish a general framework for a subspace-based acoustic multichannel equalization technique constructed from the multiple CS solutions. Using the perceptually motivated desired target responses in P-MINT and RMCLS-CIT, we aim at finding linear combinations of the multiple CS solutions which lead to such target responses. In addition, since it has been shown that regularization is effective for increasing the robustness of equalization techniques to RIR estimation errors [2], a regularization term is also incorporated in order to decrease the energy of the reshaping filter. Simulation results for a measured acoustic system and different channel estimation errors illustrate the advantages of the proposed approach.

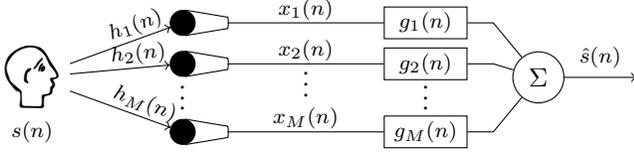
## 2. ACOUSTIC MULTICHANNEL EQUALIZATION

Fig. 1 depicts an  $M$ -channel acoustic system, where the  $m$ -th microphone signal  $x_m(n)$  at time index  $n$  is given by the convolution of the clean speech signal  $s(n)$  with the RIR between the source and the  $m$ -th microphone  $h_m(n)$ , i.e.

$$x_m(n) = s(n) * h_m(n), \quad m = 1, \dots, M. \quad (1)$$

The RIR can be described in vector notation as  $\mathbf{h}_m = [h_m(0) \ h_m(1) \ \dots \ h_m(L_h - 1)]^T$ , with  $L_h$  being the RIR

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).



**Fig. 1:** Multichannel equalization system

length and  $[\cdot]^T$  denoting the transpose operation. Equalization techniques design and apply reshaping filters  $\mathbf{g}_m$  of length  $L_g$ , i.e.  $\mathbf{g}_m = [g_m(0) \ g_m(1) \ \dots \ g_m(L_g - 1)]^T$ , yielding the output signal

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

where  $c(n)$  denotes the equalized impulse response (EIR) of length  $L_c = L_h + L_g - 1$ , i.e.  $\mathbf{c} = [c(0) \ c(1) \ \dots \ c(L_c - 1)]^T$ . Using the stacked  $ML_g$ -dimensional reshaping filter  $\mathbf{g}$  and the  $L_c \times ML_g$ -dimensional multichannel convolution matrix  $\mathbf{H}$ , i.e.

$$\begin{aligned} \mathbf{g} &= [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \dots \ \mathbf{g}_M^T]^T, \\ \mathbf{H} &= [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_M], \end{aligned} \quad (3)$$

with  $\mathbf{H}_m$  being the  $L_c \times L_g$ -dimensional convolution matrix of  $\mathbf{h}_m$ , the EIR can be expressed as

$$\mathbf{c} = \mathbf{H}\mathbf{g} \quad (5)$$

The reshaping filter  $\mathbf{g}$  can then be designed based on different optimization criteria for the EIR  $\mathbf{c}$ . When the RIRs do not share any common zeros and the reshaping filter length is  $L_g \geq \lceil \frac{L_h - 1}{M - 1} \rceil$ , the EIR can be perfectly reshaped to a desired target response [2, 4]. Since the true RIRs are however generally not available in practice [5], acoustic multichannel equalization techniques design  $\mathbf{g}$  using the estimated multichannel convolution matrix  $\hat{\mathbf{H}}$  (constructed from the estimated RIRs  $\hat{\mathbf{h}}_m$ ). Several techniques which aim at decreasing the sensitivity of  $\mathbf{g}$  to RIR estimation errors have been proposed, of which the P-MINT, RMCLS-CIT, CS, and PeCCS techniques are briefly reviewed in the following.

*Partial Multichannel Equalization Based on the Multiple-Input/Output Inverse Theorem* [2]. P-MINT aims at designing a reshaping filter that simultaneously sets the reverberant tail of the EIR to  $\mathbf{0}$  and preserves the perceptual speech quality. This objective is formulated as the minimization of the least-squares cost function

$$J_{\text{P-MINT}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2 \quad (6)$$

where the target EIR is chosen to be equal to  $\hat{\mathbf{h}}_p^d$ , which denotes the first part (i.e. direct path and early reflections) of one of the estimated RIRs, i.e.

$$\hat{\mathbf{h}}_p^d = \underbrace{[0 \ \dots \ 0]_{\tau}}_{\tau} \underbrace{[\hat{h}_p(0) \ \dots \ \hat{h}_p(L_d - 1)]}_{L_d} \underbrace{[0 \ \dots \ 0]}_{L_c - L_d - L_d}, \quad (7)$$

with  $\tau$  a delay,  $L_d$  the length of the direct path and early reflections, and  $p \in \{1, \dots, M\}$ . In order to increase the robustness of P-MINT to errors in the estimated RIRs, a regularization term has been incorporated, leading to the regularized P-MINT cost function

$$J_{\text{P-MINT}}^R = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2 + \delta \|\mathbf{g}\|_2^2 \quad (8)$$

with  $\delta$  being a regularization parameter. The regularized P-MINT reshaping filter minimizing (8) can be computed as

$$\mathbf{g}_{\text{P-MINT}}^R = (\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \delta \mathbf{I})^{-1} \hat{\mathbf{H}}^T \hat{\mathbf{h}}_p^d \quad (9)$$

with  $\mathbf{I}$  being the  $ML_g \times ML_g$ -dimensional identity matrix.

*Relaxed Multichannel Least Squares with Constrained Initial Taps* [3]. RMCLS-CIT aims at setting the reverberant tail of the EIR to  $\mathbf{0}$ , while partly controlling the perceptual speech quality. This can be achieved by introducing a weighting matrix  $\mathbf{W}_r$  in the least-squares cost function in (6), i.e.

$$J_{\text{RMCLS-CIT}} = \|\mathbf{W}_r(\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d)\|_2^2 \quad (10)$$

with

$$\mathbf{W}_r = \text{diag}\left\{\underbrace{[1 \ \dots \ 1]_{\tau}}_{\tau} \underbrace{[1 \ \dots \ 1]_{L_r}}_{L_r} \underbrace{[0 \ \dots \ 0]_{L_d - L_r}}_{L_d - L_r} [1 \ \dots \ 1]\right\}, \quad (11)$$

where  $L_r$  denotes the number of initial taps to be constrained within the desired window length. Since it has been shown that regularization increases the robustness of equalization techniques to RIR estimation errors [2], we introduce a regularization term in the RMCLS-CIT cost function in (10), leading to the regularized RMCLS-CIT cost function

$$J_{\text{RMCLS-CIT}}^R = \|\mathbf{W}_r(\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d)\|_2^2 + \delta \|\mathbf{g}\|_2^2 \quad (12)$$

The reshaping filter minimizing the regularized least-squares cost function in (12) can be computed as

$$\mathbf{g}_{\text{RMCLS-CIT}}^R = [(\mathbf{W}_r \hat{\mathbf{H}})^T (\mathbf{W}_r \hat{\mathbf{H}}) + \delta \mathbf{I}]^{-1} \times (\mathbf{W}_r \hat{\mathbf{H}})^T (\mathbf{W}_r \hat{\mathbf{h}}_p^d) \quad (13)$$

*Channel Shortening* [1, 6]. The objective of CS is to maximize the energy ratio between the first  $L_d$  taps (i.e. direct path and early reflections) and the remaining taps (i.e. reverberant tail) of the EIR without imposing further constraints on the perceptual speech quality. This objective is expressed in terms of a generalized Rayleigh quotient maximization problem, i.e.

$$J_{\text{CS}} = \frac{\|\mathbf{W}_d \hat{\mathbf{H}}\mathbf{g}\|_2^2}{\|\mathbf{W}_u \hat{\mathbf{H}}\mathbf{g}\|_2^2} = \frac{\mathbf{g}^T \hat{\mathbf{B}}\mathbf{g}}{\mathbf{g}^T \hat{\mathbf{A}}\mathbf{g}} \quad (14)$$

with

$$\mathbf{W}_d = \text{diag}\left\{\underbrace{[0 \ \dots \ 0]_{\tau}}_{\tau} \underbrace{[1 \ \dots \ 1]_{L_d}}_{L_d} \underbrace{[0 \ \dots \ 0]}_{L_c - L_d - L_d}\right\}, \quad (15)$$

$$\mathbf{W}_u = \text{diag}\left\{\underbrace{[1 \ \dots \ 1]_{\tau}}_{\tau} \underbrace{[0 \ \dots \ 0]_{L_d}}_{L_d} \underbrace{[1 \ \dots \ 1]}_{L_c - L_d - L_d}\right\}, \quad (16)$$

and

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T \mathbf{W}_d^T \mathbf{W}_d \hat{\mathbf{H}}, \quad (17)$$

$$\hat{\mathbf{A}} = \hat{\mathbf{H}}^T \mathbf{W}_u^T \mathbf{W}_u \hat{\mathbf{H}}. \quad (18)$$

Maximizing (14) is equivalent to solving the generalized eigenvalue problem  $\hat{\mathbf{B}}\mathbf{g} = \lambda\hat{\mathbf{A}}\mathbf{g}$ , where the optimal reshaping filter  $\mathbf{g}_{\text{CS}}$  is the generalized eigenvector corresponding to the largest generalized eigenvalue  $\lambda_{\text{max}}$ , i.e.

$$\hat{\mathbf{B}}\mathbf{g}_{\text{CS}} = \lambda_{\text{max}}\hat{\mathbf{A}}\mathbf{g}_{\text{CS}} \quad (19)$$

In [1] it has been shown that there exist  $L_d$  linearly independent generalized eigenvectors  $\mathbf{g}_{\text{CS}}^1, \mathbf{g}_{\text{CS}}^2, \dots, \mathbf{g}_{\text{CS}}^{L_d}$ , satisfying (19), which however yield a different performance in terms of reverberant energy suppression and perceptual speech quality. Following similar arguments as in [1] it can be shown that the reshaping filters computed by P-MINT and RMCLS-CIT also satisfy  $\mathbf{g}^T \hat{\mathbf{A}}\mathbf{g} = 0$  and  $\mathbf{g}^T \hat{\mathbf{B}}\mathbf{g} \neq 0$ , such that they also maximize the generalized Rayleigh quotient in (14). Hence, the solutions to the P-MINT and the RMCLS-CIT cost functions in (6) and (10) can be expressed as linear combinations of the multiple CS generalized eigenvectors<sup>1</sup>.

*Perceptually Constrained Channel Shortening* [7]. PeCCS aims at resolving the selection ambiguity of CS by finding a linear combination of the multiple CS reshaping filters that leads to the P-MINT target response. The linear combination  $\alpha$  is found by minimizing the cost function

$$J_{\text{PeCCS}} = \|\hat{\mathbf{H}}\mathbf{G}\alpha - \hat{\mathbf{h}}_p^d\|_2^2 + \|\mathbf{G}\alpha\|_2^2 \quad (20)$$

with  $\mathbf{G} = [\mathbf{g}_{\text{CS}}^1 \ \mathbf{g}_{\text{CS}}^2 \ \dots \ \mathbf{g}_{\text{CS}}^{L_d}]$  being an  $ML_g \times L_d$ -dimensional matrix whose columns span the solution space of CS. Minimizing (20) yields the linear combination vector

$$\alpha_{\text{PeCCS}} = [(\hat{\mathbf{H}}\mathbf{G})^T (\hat{\mathbf{H}}\mathbf{G}) + \mathbf{G}^T \mathbf{G}]^{-1} (\hat{\mathbf{H}}\mathbf{G})^T \hat{\mathbf{h}}_p^d \quad (21)$$

leading to the PeCCS reshaping filter

$$\mathbf{g}_{\text{PeCCS}} = \mathbf{G}\alpha_{\text{PeCCS}} \quad (22)$$

In the following, we propose extending the PeCCS technique through incorporating regularization and a weighting matrix in order to further increase robustness to channel estimation errors.

### 3. REGULARIZED SUBSPACE-BASED ACOUSTIC MULTICHANNEL EQUALIZATION

In [2] it has been experimentally validated that there exist reshaping filters in the solution space of CS which can achieve a high level of reverberant tail suppression. In order to obtain a robust reshaping filter in this solution space which also

yields perceptual speech quality preservation, we propose the regularized subspace-based cost function (SuB)

$$J_{\text{SuB}}^{\text{R}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{G}\alpha - \hat{\mathbf{h}}_p^d)\|_2^2 + \delta\|\mathbf{G}\alpha\|_2^2 \quad (23)$$

with  $\mathbf{W}$  being an  $L_c \times L_c$ -dimensional weighting matrix. The first term in (23) aims at finding a linear combination of the generalized eigenvectors that yields the perceptually advantageous desired target response  $\hat{\mathbf{h}}_p^d$ . Using a weighting matrix  $\mathbf{W} \neq \mathbf{I}$  allows to relax the constraints on the reshaping filter design, which may increase robustness in the presence of RIR estimation errors. The second term in (23) controls the reshaping filter energy by means of the regularization parameter  $\delta$ , in order to avoid distortions in the output signal due to errors in the estimated RIRs.

The vector  $\alpha_{\text{SuB}}^{\text{R}}$  minimizing (23) is given by

$$\alpha_{\text{SuB}}^{\text{R}} = [(\mathbf{W}\hat{\mathbf{H}}\mathbf{G})^T (\mathbf{W}\hat{\mathbf{H}}\mathbf{G}) + \delta\mathbf{G}^T \mathbf{G}]^{-1} \times (\mathbf{W}\hat{\mathbf{H}}\mathbf{G})^T (\mathbf{W}\hat{\mathbf{h}}_p^d) \quad (24)$$

leading to the reshaping filter

$$\mathbf{g}_{\text{SuB}}^{\text{R}} = \mathbf{G}\alpha_{\text{SuB}}^{\text{R}} \quad (25)$$

Depending on the choice of  $\mathbf{W}$  and  $\delta$ , the cost function in (23) gives rise to different reshaping filters. When regularization is omitted, i.e.  $\delta = 0$ , two special cases can be considered: 1) when  $\mathbf{W} = \mathbf{I}$ , the obtained reshaping filter minimizes the P-MINT cost function in (6); 2) when  $\mathbf{W} = \mathbf{W}_r$ , the obtained reshaping filter minimizes the RMCLS-CIT cost function in (10). Furthermore, with  $\delta = 1$  and  $\mathbf{W} = \mathbf{I}$ , minimizing (23) yields the PeCCS reshaping filter in (22).

In the following, the performance of the regularized SuB technique for  $\delta \neq 0$  is investigated for two different scenarios: 1) using no weighting matrix, i.e.  $\mathbf{W} = \mathbf{I}$ ; 2) using the RMCLS-CIT weighting matrix, i.e.  $\mathbf{W} = \mathbf{W}_r$  defined in (11).

## 4. EXPERIMENTAL RESULTS

In order to investigate the performance of the proposed technique, we have considered a measured acoustic system with  $M = 2$  microphones in a room with reverberation time  $T_{60} \approx 700$  ms as the true system to be equalized. The sampling frequency is  $f_s = 8$  kHz, the RIR length is  $L_h = 3000$ , and 5 different desired window lengths are investigated, i.e.  $L_w \in \{10 \text{ ms}, 20 \text{ ms}, 30 \text{ ms}, 40 \text{ ms}, 50 \text{ ms}\}$ , with  $L_w = \frac{L_d \times 10^3}{f_s}$  being the desired window length in ms. Furthermore, the simulation parameters are set to  $\tau = 0$ ,  $L_g = 2999$ ,  $p = 1$ , and  $L_r = \lfloor \frac{L_d}{3} \rfloor$ .

In order to simulate estimation errors, the true acoustic system is perturbed as in [8], i.e.

$$\hat{h}_m(n) = h_m(n)[1 + e_m(n)], \quad (26)$$

<sup>1</sup>The proof of this statement is omitted due to space constraints.

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

$$E_m = 10 \log_{10} \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}, \quad (27)$$

is generated. The considered normalized channel mismatch values are  $E_m = -33$  dB and  $E_m = -24$  dB.

The achieved reverberant energy suppression is evaluated using the energy decay curve (EDC) of the true EIR  $\mathbf{c}$ , defined as

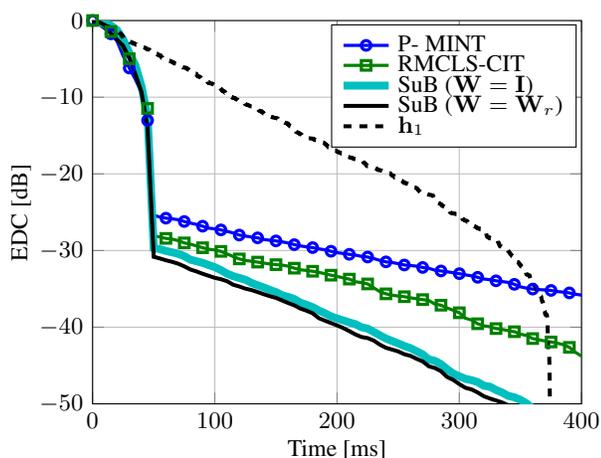
$$\text{EDC}(n) = 10 \log_{10} \frac{1}{\|\mathbf{c}\|_2^2} \sum_{j=n}^{L_c-1} c^2(j), \quad n = 0, \dots, L_c - 1, \quad (28)$$

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

The perceptual speech quality of the output signal  $\hat{s}(n)$  is evaluated using the objective speech quality measure PESQ [9] which outputs a similarity score between the reference and output signal in the range 1 – 4.5. The reference signal employed in PESQ is  $s(n) * h_1^d(n)$ , i.e. the clean speech signal convolved with the first part of the true first RIR for each value of the desired window length  $L_w$ .

Finally, the set of considered regularization parameters is  $\delta \in \{10^{-9}, 10^{-8}, \dots, 10^{-1}\}$  and the optimal regularization parameter used in each scenario is intrusively selected as the one leading to the highest PESQ score.

Fig. 2 depicts the EDC of the true RIR  $\mathbf{h}_1$  and the EDCs of the EIRs obtained using all the considered techniques for  $E_m = -33$  dB and  $L_w = 50$  ms. It can be seen that the regularized SuB technique (using both  $\mathbf{W} = \mathbf{I}$  and  $\mathbf{W} = \mathbf{W}_r$ ) leads to the largest reverberant energy suppression, with the reverberant tail being well below the reverberant tail of the

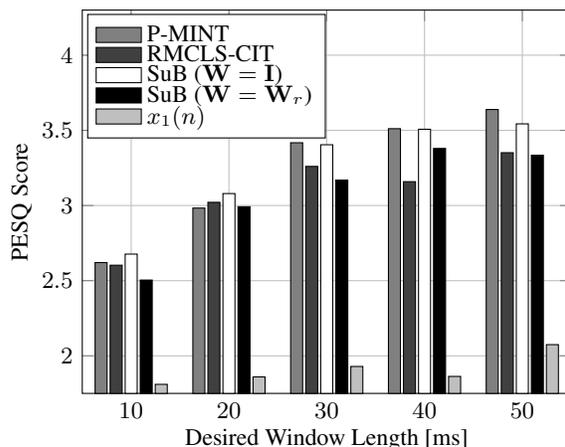


**Fig. 2:** EDC of the true RIR  $\mathbf{h}_1$  and EDC of the EIR obtained using regularized P-MINT, regularized RMCLS-CIT, regularized SuB with  $\mathbf{W} = \mathbf{I}$ , and regularized SuB with  $\mathbf{W} = \mathbf{W}_r$  ( $E_m = -33$  dB,  $L_w = 50$  ms)

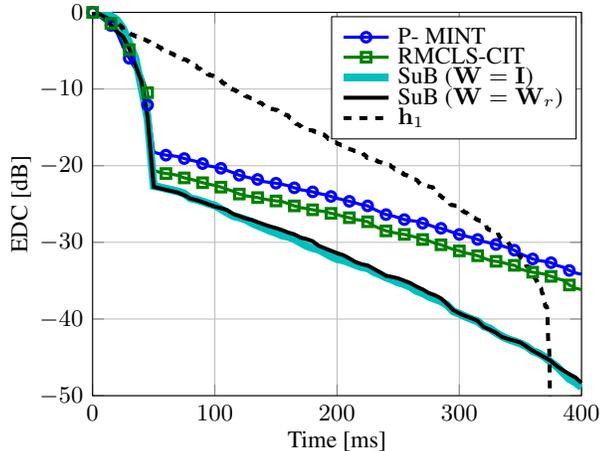
original RIR. The decay rates of the reverberant energy obtained using regularized P-MINT and regularized RMCLS-CIT are lower, with the regularized P-MINT technique leading to the lowest reverberant energy suppression. Further, Fig. 3 depicts the perceptual speech quality evaluated using PESQ for all considered desired window lengths. It can be noticed that all techniques lead to a significant performance improvement compared to the original unprocessed microphone signal  $x_1(n)$ . The regularized RMCLS-CIT technique and the regularized SuB technique with  $\mathbf{W} = \mathbf{W}_r$  typically lead to the lowest perceptual speech quality. Furthermore, the performance of the regularized P-MINT technique and regularized SuB technique with  $\mathbf{W} = \mathbf{I}$  is similar, providing the highest PESQ scores for all considered desired window lengths. Based on the analysis for the normalized channel mismatch  $E_m = -33$  dB, it can be concluded that the regularized SuB technique with  $\mathbf{W} = \mathbf{I}$  leads to a high level of reverberant energy suppression as well as a high perceptual speech quality preservation.

In order to evaluate the performance for increasing normalized channel mismatch values, Fig. 4 depicts the EDC of the true RIR  $\mathbf{h}_1$  and the EDCs of the EIRs obtained for  $E_m = -24$  dB and  $L_w = 50$  ms. As in the previous simulation, the performance of the regularized SuB technique with  $\mathbf{W} = \mathbf{I}$  and  $\mathbf{W} = \mathbf{W}_r$  is very similar, providing the highest level of reverberant energy suppression among all considered equalization techniques. Also a similar conclusion as in the previous simulation can be derived with respect to the performance in terms of PESQ depicted in Fig. 5, i.e. the perceptual speech quality obtained using regularized P-MINT and regularized SuB with  $\mathbf{W} = \mathbf{I}$  is again similar and typically higher than the perceptual speech quality obtained using the other equalization techniques.

Summarizing these simulation results, we conclude that



**Fig. 3:** PESQ score of  $x_1(n)$  and PESQ score of the output signal  $\hat{s}(n)$  for several  $L_w$  obtained using regularized P-MINT, regularized RMCLS-CIT, regularized SuB with  $\mathbf{W} = \mathbf{I}$ , and regularized SuB with  $\mathbf{W} = \mathbf{W}_r$  ( $E_m = -33$  dB)



**Fig. 4:** EDC of the true RIR  $h_1$  and EDC of the EIR obtained using regularized P-MINT, regularized RMCLS-CIT, regularized SuB with  $\mathbf{W} = \mathbf{I}$ , and regularized SuB with  $\mathbf{W} = \mathbf{W}_r$  ( $E_m = -24$  dB,  $L_w = 50$  ms)

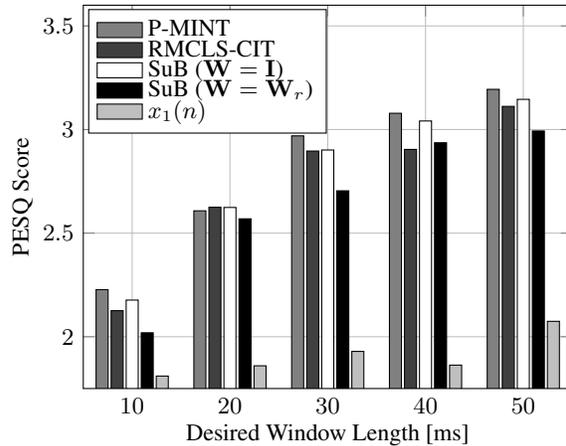
the regularized subspace-based equalization technique using  $\mathbf{W} = \mathbf{I}$  and  $\mathbf{W} = \mathbf{W}_r$  leads to the highest level of reverberant energy suppression among all considered techniques. Furthermore, the perceptual speech quality obtained when no weighting matrix is used, i.e.  $\mathbf{W} = \mathbf{I}$ , is higher than the perceptual speech quality obtained when the RMCLS-CIT weighting matrix is used, i.e.  $\mathbf{W} = \mathbf{W}_r$ .

## 5. CONCLUSION

In this paper, we have presented a robust regularized subspace-based equalization technique for speech dereverberation (SuB), where the reshaping filter is constrained to lie in the subspace spanned by the multiple CS solutions. We have compared the performance between using a weighting matrix to relax the constraints on the filter design and using no weighting matrix. Experimental results for several normalized channel mismatch values and desired window lengths demonstrate that using no weighting matrix leads to a high performance, both in terms of reverberant energy suppression and perceptual speech quality preservation.

## 6. REFERENCES

- [1] 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.
- [2] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for Partial Multichannel Equalization for Speech Dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, in Press.



**Fig. 5:** PESQ score of  $x_1(n)$  and PESQ score of the output signal  $\hat{s}(n)$  for several  $L_w$  obtained using regularized P-MINT, regularized RMCLS-CIT, regularized SuB with  $\mathbf{W} = \mathbf{I}$ , and regularized SuB with  $\mathbf{W} = \mathbf{W}_r$  ( $E_m = -24$  dB)

- [3] F. Lim and P. A. Naylor, "Relaxed Multichannel Least Squares with Constrained Initial Taps for Multichannel Dereverberation," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Aachen, Germany, Sep. 2012.
- [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] 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*, Florence, Italy, Sep. 2006.
- [6] 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.
- [7] I. Kodrasi, S. Goetze, and S. Doclo, "A perceptually constrained channel shortening technique for speech dereverberation," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, May 2013, pp. 151–155.
- [8] J. Cho, D. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 718–724, Nov. 1999.
- [9] 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.