A PERCEPTUALLY CONSTRAINED CHANNEL SHORTENING TECHNIQUE FOR SPEECH DEREVERBERATION

Ina Kodrasi[1], Stefan Goetze[2], Simon Doclo[1,2]

[1]University of Oldenburg, Department of Medical Physics and Acoustics, Oldenburg, Germany
[2]Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, Oldenburg, Germany
`ina.kodrasi@uni-oldenburg.de`

ABSTRACT

The objective of acoustic multichannel equalization is to design a reshaping filter that reduces reverberation, improves the perceptual speech quality, and is robust to errors in the estimated room impulse responses (RIRs). Although the channel shortening (CS) technique has been shown to be effective in achieving dereverberation, it may fail to preserve the natural shape of an RIR leading to speech quality degradation. Furthermore, CS yields multiple reshaping filters that satisfy its optimization criterion but result in a different perceptual speech quality.

In this paper, we propose a robust perceptually constrained channel shortening technique (PeCCS) that resolves the selection ambiguity of CS and leads to joint dereverberation and speech quality preservation. Simulation results for erroneously estimated RIRs show that PeCCS preserves the perceptual speech quality and results in a higher reverberant tail suppression than other state-of-the-art techniques, such as CS and the regularized partial multichannel equalization technique based on the multiple-input/output inverse theorem (P-MINT).

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

1. INTRODUCTION

Acoustic multichannel equalization techniques aim at achieving speech dereverberation by reshaping the estimated room impulse responses (RIRs) between the source and the microphone array. Although in theory perfect dereverberation can be obtained when multiple microphones are available [1, 2], such an approach poses the practical challenge of achieving robustness to errors in the estimated RIRs [3, 4, 5]. Since the estimated RIRs typically differ from the true RIRs [6], acoustic multichannel equalization techniques may fail to achieve dereverberation and may even lead to distortions in the output speech signal [7]. In order to address this robustness issue, techniques such as regularized partial multichannel equalization based on the multiple-input/output inverse theorem (P-MINT) [5] and channel shortening (CS) [4, 8] have been proposed.

The regularized P-MINT technique aims at simultaneously suppressing the reverberant tail and controlling the perceptual speech quality of the output signal. Experimental investigations in [5] have shown that regularized P-MINT outperforms state-of-the-art techniques such as CS in terms of perceptual speech quality. However, the performance in terms of reverberant tail suppression is generally lower than CS.

The objective of CS is to design a reshaping filter which minimizes the reverberant energy while maximizing the energy of the direct path and early reflections. Such an energy-based optimization technique typically achieves a high level of reverberant tail suppression, but may fail to preserve the general shape of a natural RIR, hence leading to a degraded speech quality [5]. Furthermore, CS yields multiple reshaping filters which all satisfy its optimization criterion. However, these reshaping filters result in a different perceptual speech quality [4], raising the question of how to obtain a perceptually advantageous solution in practice.

In this paper, a method for resolving the selection ambiguity of a perceptually advantageous solution in CS is proposed. The reshaping filter is designed as a linear combination of the multiple CS reshaping filters that leads to a reshaped overall response which is similar to the first part of one of the estimated RIRs. As a result, a robust perceptually constrained channel shortening technique (PeCCS) is established, which jointly achieves reverberant tail suppression as well as perceptual speech quality preservation.

2. ACOUSTIC MULTICHANNEL EQUALIZATION

2.1. Configuration and notation

Consider the M-channel acoustic system depicted in Fig. 1. The m-th microphone signal $x_m(n)$ at time index n is given by

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

where $*$ 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

$$\mathbf{h}_m = [h_m(0)\ h_m(1)\ \ldots\ h_m(L_h - 1)]^T, \quad (2)$$

with L_h being the RIR length and $[\cdot]^T$ denoting the transpose operation. Applying reshaping filters $\mathbf{g}_m$ of length L_g, i.e.

$$\mathbf{g}_m = [g_m(0)\ g_m(1)\ \ldots\ g_m(L_g - 1)]^T, \quad (3)$$
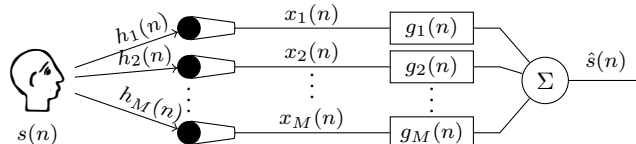

Fig. 1. Multichannel equalization system

the output signal $\hat{s}(n)$ is given by the sum of the filtered microphone signals, i.e.

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

where $c(n)$ denotes the equalized impulse response (EIR) of length $L_c = L_h + L_g - 1$, which can be described in vector notation as $\mathbf{c} = [c(0)\ c(1)\ \ldots\ c(L_c - 1)]^T$. Using the stacked ML_g-dimensional reshaping filter $\mathbf{g}$, i.e.

$$\mathbf{g} = [\mathbf{g}_1^T\ \mathbf{g}_2^T\ \ldots\ \mathbf{g}_M^T]^T, \quad (5)$$

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

$$\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2\ \ldots\ \mathbf{H}_M], \quad (6)$$

with $\mathbf{H}_m$ being the $L_c \times L_g$-dimensional convolution matrix of $\mathbf{h}_m$, i.e.

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \ldots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h - 1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h - 1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & h_m(L_h - 1) \end{bmatrix}, \quad (7)$$

the EIR can be expressed as

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

The convolution matrix $\mathbf{H}$ is assumed to be a full row-rank matrix [9], and the length of the reshaping filters is assumed to be $L_g \geq \lceil \frac{L_h - 1}{M - 1} \rceil$ [1]. The reshaping filter $\mathbf{g}$ can then be constructed based on different design objectives for the EIR $\mathbf{c}$.

However, since the true RIRs are generally not available in practice, acoustic multichannel equalization techniques design $\mathbf{g}$ using the estimated multichannel convolution matrix $\hat{\mathbf{H}}$ (constructed from the estimated RIRs $\hat{\mathbf{h}}_m$). Hence, the estimated EIR $\hat{\mathbf{c}} = \hat{\mathbf{H}}\mathbf{g}$ is typically optimized instead of the true EIR in (8). Several techniques which aim at decreasing the sensitivity of the reshaping filter $\mathbf{g}$ to RIR estimation errors have been proposed, of which the regularized P-MINT and the CS techniques are briefly reviewed in the following.

2.2. Regularized partial multichannel equalization based on the multiple-input/output inverse theorem (P-MINT)

The objective of the regularized P-MINT technique [5] is to suppress the reverberant tail of the EIR while also preserving the general shape of a natural RIR. This is achieved by defining the desired EIR to be $\hat{\mathbf{h}}_p^d$, which denotes the first part (i.e. direct path and early reflections) of one of the estimated RIRs, i.e.

$$\hat{\mathbf{h}}_p^d = [\underbrace{0 \ldots 0}_{\tau}\ \underbrace{\hat{h}_p(0) \ldots \hat{h}_p(L_d - 1)}_{L_d}\ 0 \ldots 0]^T, \quad (9)$$

with τ being a delay in number of samples, L_d denoting the length of the direct path and early reflections in number of samples, and $p \in$

$\{1,\ \ldots,\ M\}$. In order to decrease the sensitivity of the reshaping filter to errors in the estimated RIRs, a regularization term needs to be incorporated, leading to the regularized P-MINT cost function

$$J_{\text{P-MINT}}^{\text{R}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^d\|_2^2 + \delta\|\mathbf{g}\|_2^2 \quad (10)$$

with δ being a regularization parameter. The regularized P-MINT reshaping filter minimizing (10) can be computed as

$$\mathbf{g}_{\text{P-MINT}}^{\text{R}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{h}}_p^d \quad (11)$$

with $\mathbf{I}$ being the $ML_g \times ML_g$-dimensional identity matrix. In [5] it has been experimentally validated that such a technique is preferable in terms of perceptual speech quality. However, its performance in terms of reverberant tail suppression is typically lower than other state-of-the-art multichannel equalization techniques.

2.3. Channel shortening (CS)

CS has been extensively investigated in the context of digital communication applications [10] and has recently been applied to acoustic system equalization [4, 8]. The objective of CS is to maximize the energy in the first L_d taps (i.e. direct path and early reflections) of the EIR, while minimizing the energy in the remaining taps (i.e. reverberant tail). This objective can be expressed in terms of a generalized Rayleigh quotient maximization problem, i.e.

$$J_{\text{CS}} = \frac{\|\mathbf{W}_d\hat{\mathbf{H}}\mathbf{g}\|_2^2}{\|\mathbf{W}_u\hat{\mathbf{H}}\mathbf{g}\|_2^2} = \frac{\mathbf{g}^T\hat{\mathbf{B}}\mathbf{g}}{\mathbf{g}^T\hat{\mathbf{A}}\mathbf{g}} \quad (12)$$

where $\mathbf{W}_d$ and $\mathbf{W}_u$ are diagonal matrices representing the desired and undesired windows respectively, i.e.

$$\mathbf{W}_d = \text{diag}\{[\underbrace{0 \ldots 0}_{\tau}\ \underbrace{1 \ldots 1}_{L_d}\ 0 \ldots 0]\}, \quad (13)$$

$$\mathbf{W}_u = \text{diag}\{[\underbrace{1 \ldots 1}_{\tau}\ \underbrace{0 \ldots 0}_{L_d}\ 1 \ldots 1]\}, \quad (14)$$

and

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T\mathbf{W}_d^T\mathbf{W}_d\hat{\mathbf{H}}, \quad (15)$$

$$\hat{\mathbf{A}} = \hat{\mathbf{H}}^T\mathbf{W}_u^T\mathbf{W}_u\hat{\mathbf{H}}. \quad (16)$$

Maximizing (12) 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 $\lambda_{\max}$, i.e.

$$\hat{\mathbf{B}}\mathbf{g}_{\text{CS}} = \lambda_{\max}\hat{\mathbf{A}}\mathbf{g}_{\text{CS}} \quad (17)$$

In [4] it has been shown that there exist L_d linearly independent generalized eigenvectors satisfying (17), which however yield a different perceptual speech quality. Furthermore, also a selection criterion has been proposed in [4], namely the generalized eigenvector leading to the minimum l_2-norm EIR. However, this criterion is based on observations using informal listening tests for perfectly estimated RIRs.

In the following, we propose a method for deriving a perceptually advantageous solution in the presence of RIR estimation errors, which is a linear combination of all generalized eigenvectors satisfying (17). The experimental results in Section 4 validate the advantages of the proposed method.

3. PERCEPTUALLY CONSTRAINED CHANNEL SHORTENING (PECCS)

Consider the $ML_g \times L_d$-dimensional matrix $\mathbf{G}_{\text{CS}}$ whose columns are the generalized eigenvectors $\mathbf{g}_{\text{CS}}^k$, $k = 1, \ldots, L_d$, satisfying (17), i.e.

$$\mathbf{G}_{\text{CS}} = [\mathbf{g}_{\text{CS}}^1 \ \mathbf{g}_{\text{CS}}^2 \ \cdots \ \mathbf{g}_{\text{CS}}^{L_d}]. \tag{18}$$

Then, any linear combination of these generalized eigenvectors, i.e.

$$\mathbf{g} = \mathbf{G}_{\text{CS}} \boldsymbol{\alpha}, \tag{19}$$

with $\boldsymbol{\alpha}$ being an L_d-dimensional vector of scalar coefficients, also maximizes the generalized Rayleigh quotient in (12). In order to obtain a reshaping filter that also yields a high perceptual speech quality, we propose to compute the vector $\boldsymbol{\alpha}$ minimizing the cost function

$$J_{\text{PeCCS}} = \|\hat{\mathbf{H}}\mathbf{G}_{\text{CS}}\boldsymbol{\alpha} - \hat{\mathbf{h}}_p^{\text{d}}\|_2^2 + \|\mathbf{G}_{\text{CS}}\boldsymbol{\alpha}\|_2^2 \tag{20}$$

The first term in (20) aims to obtain a reshaping filter that preserves the first taps of one of the estimated RIRs whereas the second term controls the reshaping filter energy in order to decrease its sensitivity to errors in the estimated RIRs. Clearly, the weight given to the minimization of the energy of the reshaping filter in (20) can be controlled by incorporating a regularization parameter as done in the regularized P-MINT technique (cf. (10)). However, the analysis of such a regularized technique is beyond the scope of this paper.

By setting the derivative of (20) with respect to $\boldsymbol{\alpha}$ equal to $\mathbf{0}$, the vector $\boldsymbol{\alpha}_{\text{PeCCS}}$ minimizing J_{PeCCS} can be computed as

$$\begin{aligned}\boldsymbol{\alpha}_{\text{PeCCS}} &= [(\hat{\mathbf{H}}\mathbf{G}_{\text{CS}})^T(\hat{\mathbf{H}}\mathbf{G}_{\text{CS}}) + \mathbf{G}_{\text{CS}}^T\mathbf{G}_{\text{CS}}]^{-1} \\ &\times (\hat{\mathbf{H}}\mathbf{G}_{\text{CS}})^T(\hat{\mathbf{h}}_p^{\text{d}})\end{aligned} \tag{21}$$

leading to the PeCCS reshaping filter

$$\mathbf{g}_{\text{PeCCS}} = \mathbf{G}_{\text{CS}} \boldsymbol{\alpha}_{\text{PeCCS}} \tag{22}$$

The proposed PeCCS technique is hence a two-step multichannel equalization technique, which exploits the CS multiple solutions that already achieve reverberant tail suppression in order to obtain a robust and perceptually advantageous reshaping filter.

4. EXPERIMENTAL RESULTS

In this section, the proposed PeCCS technique is compared to the CS [4] and regularized P-MINT [5] techniques.

4.1. Simulation parameters and performance measures

We have considered an acoustic scenario with a single speech source and $M = 2$ microphones in a room with reverberation time $T_{60} \approx 600$ ms. The RIRs have been measured using the swept-sine technique with $L_h = 4800$ at a sampling frequency $f_s = 8$ kHz. In order to simulate estimation errors, the measured RIRs have been perturbed by adding scaled white noise as proposed in [11], i.e.

$$\hat{h}_m(n) = h_m(n)[1 + e(n)], \tag{23}$$

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

$$E_m = 10 \log_{10} \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}, \tag{24}$$

is generated. Although other normalized channel mismatch values have been investigated, due to space constraints only $E_m = -33$ dB is considered in this paper. Furthermore, the simulation parameters are set to $L_g = 4799$, $\tau = 0$, $p = 1$, and several desired window lengths have been considered, i.e.

$$L_s = L_d \frac{1000}{f_s} \in \{20 \text{ ms}, \ 30 \text{ ms}, \ 40 \text{ ms}, \ 50 \text{ ms}\}. \tag{25}$$

The suppression of the reverberant tail is evaluated using the energy decay curve (EDC) of the EIR defined as

$$\text{EDC}(n) = 10 \log_{10} \frac{1}{\|\mathbf{c}\|_2^2} \sum_{i=n}^{L_c-1} c^2(i), \ n = 0, \ldots, L_c - 1, \tag{26}$$

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

A commonly used perceptually motivated objective criterion which indicates how much of the non-direct path energy will be perceived as coloration instead of reverberation is the clarity index C_{L_s} [12], defined as

$$C_{L_s} = 10 \log_{10} \frac{\sum_{n=0}^{L_d-1} c^2(n)}{\sum_{n=L_d}^{L_c-1} c^2(n)}. \tag{27}$$

Furthermore, it has been shown in [13] that measures relying on auditory models exhibit the highest correlation with subjective listening tests when evaluating the quality of dereverberated speech. Hence, to evaluate the perceptual speech quality we have used the objective speech quality measure PESQ [14], where the reference signal employed in PESQ is $s(n) * h_1^{\text{d}}(n)$, 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_s.

For the CS technique, the reshaping filter is selected as the generalized eigenvector leading to the minimum l_2-norm estimated EIR $\hat{\mathbf{c}}$ as proposed in [4]. Furthermore, for the regularized P-MINT technique the regularization parameter δ in (11) is determined as in [5], i.e. a set of regularization parameters $\delta \in \{10^{-9}, \ 10^{-8}, \ \ldots, \ 10^{-1}\}$ is considered, and the optimal parameter is selected as the one leading to the highest PESQ score. It should be noted that such a procedure is intrusive and inapplicable in practice, since knowledge of the true RIRs is required to compute the true EIR $\mathbf{c}$ and to compute the reference signal $s(n) * h_1^{\text{d}}(n)$ for PESQ.

4.2. Results

Fig. 2 depicts the EDC of the true RIR $\mathbf{h}_1$ and the EDCs of the EIRs obtained using all considered techniques for the desired window length $L_s = 50$ ms. As can be observed in this figure, the generalized CS eigenvector leading to the minimum l_2-norm estimated EIR yields an EDC that is only slightly lower than the EDC of $\mathbf{h}_1$. On the other hand, the regularized P-MINT technique achieves a higher level of reverberant tail suppression and the artificial tail introduced after 0.5 s is not audible. However, the proposed PeCCS technique yields the highest level of reverberant tail suppression throughout the duration of the RIR, with its reverberant tail being significantly below the reverberant tail of $\mathbf{h}_1$. Fig. 3 shows the clarity index of $\mathbf{h}_1$ and the clarity index of the EIRs obtained using all considered techniques for different L_s. The depicted C_{L_s} values validate the previous conclusion based on the EDCs, i.e. the proposed PeCCS technique yields the highest performance, followed by the regularized P-MINT technique. The generalized eigenvector leading to the
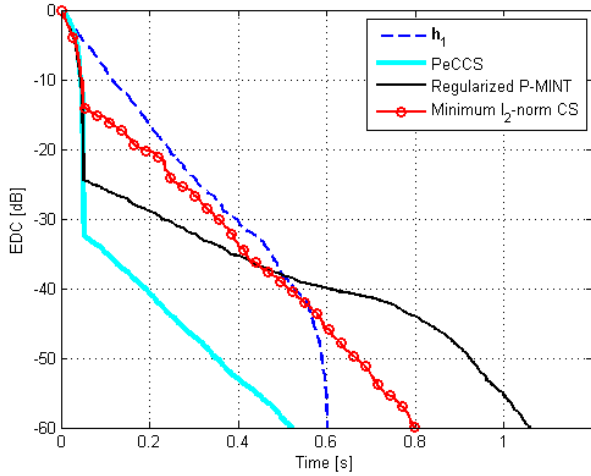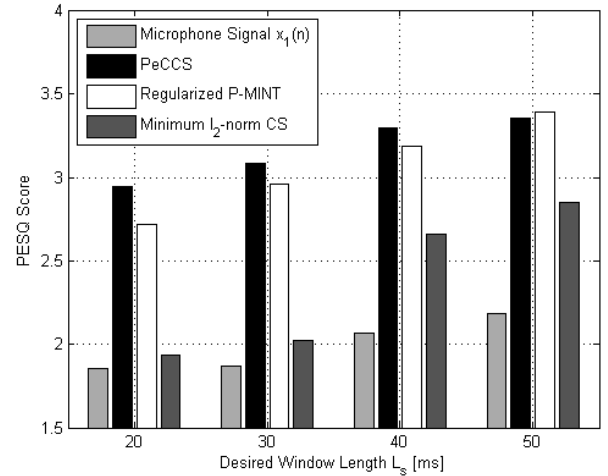
Fig. 2. EDC of the true RIR $\mathbf{h}_1$ and EDC of the EIR obtained using PeCCS, regularized P-MINT, and the generalized CS eigenvector leading to the minimum l_2-norm estimated EIR ($L_s = 50$ ms, $E_m = -33$ dB)

minimum l_2-norm estimated EIR yields the lowest clarity index for all considered desired window lengths L_s.

Finally, we have evaluated the perceptual speech quality using PESQ for the different considered desired window lengths L_s. The PESQ score of the first microphone signal $x_1(n)$ is also computed in order to determine the effectiveness of applying the considered dereverberation techniques. The results are presented in Fig. 4, where it can be seen that the generalized CS eigenvector leading to the minimum l_2-norm estimated EIR yields the lowest perceptual speech quality for all considered L_s. Furthermore, the proposed PeCCS technique yields a higher performance than the regularized P-MINT
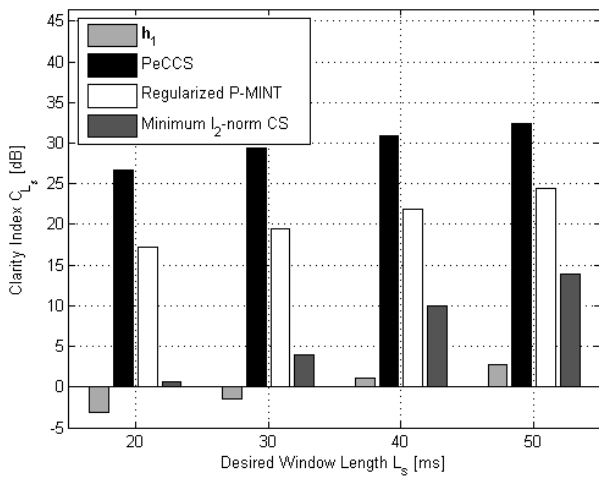
Fig. 3. Clarity index of the true RIR $\mathbf{h}_1$ and clarity index of the EIR obtained using PeCCS, regularized P-MINT, and the generalized CS eigenvector leading to the minimum l_2-norm estimated EIR ($E_m = -33$ dB)

Fig. 4. PESQ score of the first microphone signal $x_1(n)$ and PESQ score of the system's output $\hat{s}(n)$ obtained using PeCCS, regularized P-MINT, and the generalized CS eigenvector leading to the minimum l_2-norm estimated EIR ($E_m = -33$ dB)

technique for $L_s = 20$ ms, $L_s = 30$ ms, and $L_s = 40$ ms, whereas a similar performance as the regularized P-MINT technique is obtained for $L_s = 50$ ms.

In conclusion, the presented simulation results demonstrate that applying the proposed PeCCS technique leads to a significantly higher level of reverberant tail suppression as well as speech quality preservation.

5. RELATED WORK

To the best of our knowledge, the presented PeCCS technique is the first technique that considers the selection ambiguity of a perceptually advantageous solution in CS [4, 8] in the presence of RIR estimation errors. In contrast to the selection criterion proposed in [4], PeCCS takes into account RIR estimation errors and uses a novel optimization criterion which leads to a linear combination of the multiple CS solutions instead of a single generalized eigenvector. Furthermore, unlike the regularized P-MINT technique proposed in [5], PeCCS is a two-step multichannel equalization technique which exploits reshaping filters that already achieve reverberant tail suppression in order to derive a perceptually advantageous linear combination.

6. CONCLUSION

In this paper, we have presented a robust perceptually constrained acoustic multichannel equalization technique for speech dereverberation based on CS (PeCCS). Experimental results have shown that PeCCS achieves a significantly higher level of reverberant tail suppression than other state-of-the-art techniques such as CS and regularized P-MINT, while preserving the perceptual speech quality.

7. REFERENCES

[1] 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.

[2] H. Hacihabibouglu and Z. Cvetkovic, "Multichannel dereverberation theorems and robustness issues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 676–689, Feb. 2012.

[3] 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.

[4] 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 (IWAENC)*, Tel Aviv, Israel, Sept. 2010.

[5] I. Kodrasi, S. Goetze, and S. Doclo, "Increasing the robustness of acoustic multichannel equalization by means of regularization," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012, pp. 161–164.

[6] 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.

[7] 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.

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

[9] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 215–218, Jan. 1998.

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

[11] 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.

[12] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer Publishing Company, 2010.

[13] S. Goetze, E. Albertin, J. Rennies, E. A. P. Habets, and K.-D. Kammeyer, "Speech quality assessment for listening-room compensation," in *Proc. 38th AES International Conference on Sound Quality Evaluation*, Pitea, Sweden, June 2010, pp. 11–20.

[14] 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.