# ROBUST PARTIAL MULTICHANNEL EQUALIZATION TECHNIQUES FOR SPEECH DEREVERBERATION

*Ina Kodrasi, Simon Doclo*

University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany
{ina.kodrasi,simon.doclo}@uni-oldenburg.de

## ABSTRACT

This paper presents a novel approach for partial multichannel equalization using the multiple-input/output inverse theorem with the first part of one of the estimated channels as the target response (P-MINT). In order to further increase the robustness against channel estimation errors, two extensions are proposed, i.e. the incorporation of a regularization parameter in the inverse filter design and a truncated singular value decomposition approach. Experimental results for speech dereverberation show that the regularized P-MINT method outperforms state-of-the-art techniques such as channel shortening and the relaxed multichannel least-squares method in terms of robustness to channel estimation errors.

***Index Terms*—** acoustic channel equalization, robustness, channel estimation errors, dereverberation

## 1. INTRODUCTION

In many speech communication applications the microphone signals are corrupted by reverberation, causing the speech to sound distant and spectrally distorted. With the continuously growing demand for high-quality hands-free speech communication in teleconferencing applications, voice-controlled systems and hearing aids, speech enhancement techniques aimed at dereverberation have become indispensable. One particular class of speech dereverberation techniques is acoustic channel equalization, which is based on estimating and inverse filtering the room impulse responses (RIRs), e.g. using the multiple-input/output inverse theorem (MINT) technique that aims to recover the anechoic speech signal [1].

However, since in practice the estimated acoustic system typically differs from the real system due to fluctuations of the RIRs (e.g. temperature or position variations [2]) or estimation errors (e.g. due to the sensitivity of blind system identification methods to interfering noise [3]), it is well known that MINT fails to equalize the true acoustic system, possibly leading to severe distortions in the output signal. In order to increase the robustness of MINT against estimation errors, it has been proposed to incorporate a regularization parameter in the filter design [4] or to use a truncated singular value decomposition (TSVD) approach [5].

However, since late reverberation (typically defined as the part of the room impulse response after 50-80 ms) is known to be the major cause of sound quality degradation, designing inverse filters to partially equalize the channel such that only the late reverberation effect is removed (e.g. using channel shortening [6]), is sufficient for speech dereverberation and has even been shown to be more robust than designing exact inverse filters that recover the anechoic speech signal. Aiming for robust partial multichannel equalization for speech dereverberation, we propose an alternative approach to channel shortening, which uses MINT and the first part of one of the estimated channels as the target response. In order to further enhance its robustness to channel estimation errors, we propose to incorporate regularization and truncation procedures similarly to [4] and [5], and investigate the robustness of all considered approaches against various channel estimation errors.

## 2. ACOUSTIC MULTICHANNEL EQUALIZATION TECHNIQUES

Consider an acoustic system with a single source and $M$ microphones as depicted in Fig. 1. The $m$-th microphone signal at time index $n$ is given by $x_m(n) = h_m(n) * s(n)$, where $*$ denotes the convolution operation, $s(n)$ is the clean speech signal, and $h_m(n)$ indicates the room impulse response (RIR) between the source and the $m$-th microphone, which can be described in vector notation as

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

with $L_h$ its length and $\{\cdot\}^T$ the transpose operation.

In order to achieve perfect channel equalization up to a delay $\tau$, inverse filters $\mathbf{g}_m$ of length $L_g$ need to be computed, with $\mathbf{g}_m = [g_m(0)\ g_m(1)\ \ldots\ g_m(L_g-1)]^T$, such that

$$\hat{s}(n) = \sum_{m=1}^{M} x_m(n) * g_m(n) = s(n-\tau). \tag{2}$$

This convolution operation can be expressed in matrix/vector notation as

$$\mathbf{Hg} = \mathbf{d}, \tag{3}$$

with

$$\begin{aligned} \mathbf{H} &= [\mathbf{H}_1\ \mathbf{H}_2\ \ldots\ \mathbf{H}_M]_{(L_h+L_g-1)\times(ML_g)} \\ \mathbf{g} &= \left[\mathbf{g}_1^T\ \mathbf{g}_2^T\ \ldots\ \mathbf{g}_M^T\right]^T_{(ML_g)\times(1)} \\ \mathbf{d} &= [\underbrace{0\ \ldots\ 0}_{\tau}\ 1\ 0\ \ldots\ 0]^T_{(L_h+L_g-1)\times(1)}, \end{aligned} \tag{4}$$

where $\mathbf{H}_m$ is the $(L_h+L_g-1)\times(L_g)$ convolution matrix of $\mathbf{h}_m$ and $\{\cdot\}_{(m)\times(n)}$ denotes the size of the matrix/vector under consid-

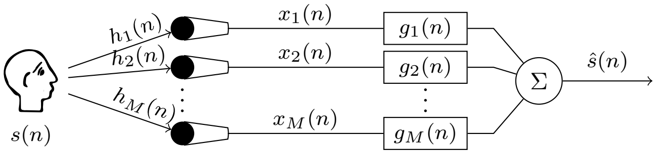

**Fig. 1**: Multichannel equalization system

eration. Assuming that the RIRs do not share any common zeros and that $L_g \geq \lceil \frac{L_h-1}{M-1} \rceil$, exact inverse filters can be computed using the multiple-input/output inverse theorem [1] as

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{d}, \tag{5}$$

where $\{\cdot\}^{+}$ denotes the Moore-Penrose pseudo-inverse. It has been shown in [7] that the matrix $\mathbf{H}$ is full row-rank, therefore its pseudo-inverse can be computed as $\mathbf{H}^{+} = \mathbf{H}^{T}(\mathbf{H}\mathbf{H}^{T})^{-1}$.

Since the least-squares solution given in (5) is very sensitive to channel estimation errors, regularization procedures for the inverse filter design have been proposed, such as the regularized least-squares and truncated singular value decomposition approach described below.

**Regularized MINT.** In the regularized least-squares approach presented in [4] the inverse filter is computed as

$$\mathbf{g} = (\mathbf{H}^{T}\mathbf{H} + \delta\mathbf{I})^{-1}\mathbf{H}^{T}\mathbf{d}, \tag{6}$$

where $\delta$ is a regularization parameter and $\mathbf{I}$ is the $(ML_g) \times (ML_g)$ identity matrix. Increasing the parameter $\delta$ in (6) decreases the norm of $\mathbf{g}$, which makes the inverse filter less sensitive to fluctuations of the RIRs. On the other hand, increasing this parameter reduces the accuracy of the inverse filters with respect to the true RIRs, resulting in a trade-off between robustness and equalization performance.

**Truncated MINT.** Another approach for increasing the robustness of MINT has been proposed in [5], where the pseudo-inverse in (5) is computed using the singular value decomposition of $\mathbf{H}$, i.e.

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{T} \Rightarrow \mathbf{H}^{+} = \mathbf{V}\mathbf{\Sigma}^{+}\mathbf{U}^{T}, \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal $(L_h + L_g - 1) \times (L_h + L_g - 1)$ and $(ML_g) \times (L_h + L_g - 1)$ matrices respectively, and $\mathbf{\Sigma}$ is a diagonal matrix consisting of the singular values $\sigma_i$ of $\mathbf{H}$ in descending order, i.e. $\mathbf{\Sigma} = \text{diag}\{[\sigma_1\ \sigma_2\ \ldots\ \sigma_{L_h+L_g-1}]\}$. When $\mathbf{H}$ contains estimation errors, also the singular values $\sigma_i$ will be perturbed by errors. In order to decrease the sensitivity of the inverse filter to these errors, truncating the singular value decomposition up to a truncation parameter $k$ is proposed, i.e. $\mathbf{\Sigma}_k = \text{diag}\{[\sigma_1\ \sigma_2\ \ldots\ \sigma_k\ 0\ \ldots\ 0]\}$, such that the smallest singular values are disregarded. Similarly to the regularized MINT method, also this approach results in a trade-off between robustness and equalization performance.

In addition, experimental results in [6] show that partial channel equalization techniques such as relaxed multichannel least-squares (RMCLS) and channel shortening (CS) outperform MINT in terms of robustness to channel estimation errors. These methods aim at shortening the equalized impulse response (EIR) defined as

$$c(i) = \sum_{m=1}^{M} h_m(i) * g_m(i), \quad i = 0,\ 1,\ \ldots,\ L_h + L_g - 2. \tag{8}$$

**Relaxed multichannel least-squares.** RMCLS achieves channel shortening by introducing a weighting function $\mathbf{w}$ in (3), i.e.

$$\mathbf{W}\mathbf{H}\mathbf{g} = \mathbf{W}\mathbf{d}, \tag{9}$$

with $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ and

$$\mathbf{w} = [\underbrace{1\ \ldots\ 1}_{\tau}\ \underbrace{1\ 0\ \ldots\ 0}_{L_d}\ 1\ \ldots 1]^{T}_{(L_h+L_g-1)\times(1)}, \tag{10}$$

where $L_d$ denotes the length of the direct path and early reflections (in number of samples), which is typically considered to be between 50 and 80 ms. The inverse filter $\mathbf{g}$ is then computed as

$$\mathbf{g} = (\mathbf{W}\mathbf{H})^{+}\mathbf{W}\mathbf{d}, \tag{11}$$

which ensures that the last taps of the EIR are set to 0, while putting no constraints on the first taps.

**Channel shortening.** The aim of CS is to maximize the energy in the first $L_d$ taps of the equalized impulse response, while minimizing the energy of the remaining taps. This can be expressed in terms of a generalized Rayleigh quotient maximization problem, i.e.

$$\max_{\mathbf{g}} \frac{\mathbf{g}^{T}\mathbf{B}\mathbf{g}}{\mathbf{g}^{T}\mathbf{A}\mathbf{g}}, \tag{12}$$

where

$$\begin{aligned}
\mathbf{B} &= \mathbf{H}^{T}\text{diag}\{\mathbf{w}_d\}^{T}\text{diag}\{\mathbf{w}_d\}\mathbf{H} \\
\mathbf{A} &= \mathbf{H}^{T}\text{diag}\{\mathbf{w}_u\}^{T}\text{diag}\{\mathbf{w}_u\}\mathbf{H} \\
\mathbf{w}_d &= [\underbrace{0\ \ldots\ 0}_{\tau}\ \underbrace{1\ \ldots\ 1}_{L_d}\ 0\ \ldots\ 0]^{T}_{(L_h+L_g-1)\times(1)} \\
\mathbf{w}_u &= \mathbf{1}_{(L_h+L_g-1)\times(1)} - \mathbf{w}_d.
\end{aligned} \tag{13}$$

Maximizing the generalized Rayleigh quotient in (12) is equivalent to the generalized eigenvalue problem $\mathbf{B}\mathbf{g} = \lambda\mathbf{A}\mathbf{g}$, where the optimal filter $\mathbf{g}$ is the generalized eigenvector corresponding to the largest generalized eigenvalue. However, since multiple solutions to (12) exist, a criterion has been proposed in [6] for selecting a perceptually advantageous solution, i.e. the one leading to the minimum $l_2$-norm EIR.

## 3. PARTIAL MULTICHANNEL EQUALIZATION USING MINT (P-MINT)

Although the partial channel equalization techniques presented above ensure that the EIR is shortened, they have no control over the filter coefficients, and hence the frequency response of the shortened equalized impulse response, which may lead to undesired perceptual effects. In order to achieve a more direct control over these filter coefficients, we propose to use the first part of one of the estimated RIRs as the target response in (3) instead of $\mathbf{d}$, i.e.

$$\mathbf{H}\mathbf{g} = \mathbf{h}_m^{d}, \tag{14}$$

where

$$\mathbf{h}_m^{d} = [\underbrace{0\ \ldots\ 0}_{\tau}\ \underbrace{h_m(0)\ \ldots\ h_m(L_d - 1)}_{L_d}\ 0\ \ldots\ 0]^{T}. \tag{15}$$

Assuming that the same conditions as for MINT are satisfied, the solution to (14) is given by

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{h}_m^{d}. \tag{16}$$

Since **P**artial channel equalization is achieved using **MINT**, we refer to this method as P-MINT. Following similar arguments as in [6], it can be shown that the inverse filter calculated by P-MINT satisfies $\mathbf{g}^{T}\mathbf{A}\mathbf{g} = 0$ and $\mathbf{g}^{T}\mathbf{B}\mathbf{g} \neq 0$, therefore also maximizing the generalized Rayleigh quotient in (12). Furthermore, this inverse filter is a linear combination of the multiple generalized eigenvectors obtained in the generalized eigenvalue decomposition $\mathbf{B}\mathbf{g} = \lambda\mathbf{A}\mathbf{g}$.[1] Although it will be shown using the experimental results in Section 4

[1] Due to space constraints, the proof of this statement is omitted here.

that P-MINT is more robust to channel estimation errors than channel shortening, in order to further increase its robustness as channel estimation errors increase, we also propose to extend P-MINT using regularization and truncation similarly as in [4] and [5].

**Regularized P-MINT.** In analogy to (6), the inverse filter is calculated as

$$\mathbf{g} = (\mathbf{H}^T\mathbf{H} + \delta\mathbf{I})^{-1}\mathbf{H}^T\mathbf{h}_m^d, \tag{17}$$

where $\delta$ is a regularization parameter.

**Truncated P-MINT.** Similarly to the TSVD approach discussed in Section 2, we now compute the inverse filter $\mathbf{g}$ as

$$\mathbf{g} = \mathbf{V}\mathbf{\Sigma}_k^+\mathbf{U}^T\mathbf{h}_m^d, \tag{18}$$

where $k$ is a truncation parameter.

## 4. EXPERIMENTAL RESULTS

In order to evaluate the performance and robustness of the proposed approaches when channel estimation errors are present, we have used a 2-channel acoustic system ($T_{60} \approx 300$ ms) from the MARDY database [8] as the true system to be equalized. The sampling frequency is $f_s = 16$ kHz, and the simulation parameters are set to $L_h = 2000$, $L_g = 1999$, and $\tau = 0$. Further, in CS, RMCLS, and P-MINT, we have used $L_d = 0.05 f_s$, corresponding to 50 ms, which is a typical transition time between early and late reflections. Additionally, the desired response in P-MINT is chosen as the direct path of the first estimated channel $\mathbf{h}_1^d$.

The true acoustic system $\mathbf{h}$ is perturbed as in [9] to obtain

$$\hat{h}_m(n) = [1 + e(n)]h_m(n), \tag{19}$$

where $e(n)$ is an uncorrelated Gaussian noise sequence with zero mean and an appropriate variance, such that a desired channel mismatch $C_m$, defined as

$$C_m = 20\log_{10}\frac{\|\mathbf{h} - \hat{\mathbf{h}}\|_2}{\|\mathbf{h}\|_2}, \tag{20}$$

is generated.

We have considered two different channel mismatches, i.e. $C_m = -33$ dB and $C_m = -15$ dB, and the performance of the multichannel equalization algorithms is evaluated by calculating the energy decay curves (EDCs) of the equalized impulse responses, defined as

$$\text{EDC}_{\mathbf{c}}(i) = 10\log_{10}\frac{1}{\|\mathbf{c}\|_2^2}\sum_{j=i}^{L_h+L_g-1} c^2(j),\ i = 0, \ldots, L_h{+}L_g{-}2, \tag{21}$$

where $\text{EDC}_{\mathbf{c}}$ denotes the energy decay curve (EDC) of $\mathbf{c}$. For the sake of readability and to avoid overcrowded plots, the experimental part is structured in three parts.

*Experiment 1.* In the first experiment, the performance of P-MINT is compared to MINT and CS for the considered estimated acoustic systems. Figs. 2 and 3 depict the obtained EDCs for $C_m = -33$ dB and $C_m = -15$ dB, respectively. It can be seen in Fig. 2 that MINT completely fails to equalize the channel, while for CS, the part of the EDC before 50 ms (corresponding to $L_d$) is well above the EDC of $\mathbf{h}_1$. On the other hand, the EDC obtained using P-MINT is significantly below $\text{EDC}_{\mathbf{h}_1}$ and the artificial tail introduced after 125 ms is below audible levels. Fig. 3 shows that as channel estimation errors increase to $-15$ dB, the reverberation suppression is not satisfactory, even though P-MINT is still more robust than CS and MINT.

*Experiment 2.* In an attempt to further increase the robustness, we have investigated the performance of the regularized P-MINT and truncated P-MINT, described in Section 3. Additionally, the performance of the above mentioned procedures is compared to regularized and truncated MINT, described in Section 2. Since all of these procedures require a regularization or truncation parameter to be chosen, a performance measure needs to be defined to select the parameter that yields the optimal EIR. In this paper we have selected the optimal regularization or truncation parameter $p$, as the one that yields the EDC that is closest to the desired EDC in the minimum mean square error sense, i.e.

$$\min_p \frac{1}{L_h + L_g - 1}\sum_{i=0}^{L_h+L_g-2} |\text{EDC}_{\mathbf{c}_p}(i) - \text{EDC}_{\mathbf{h}_1^d}(i)|^2, \tag{22}$$

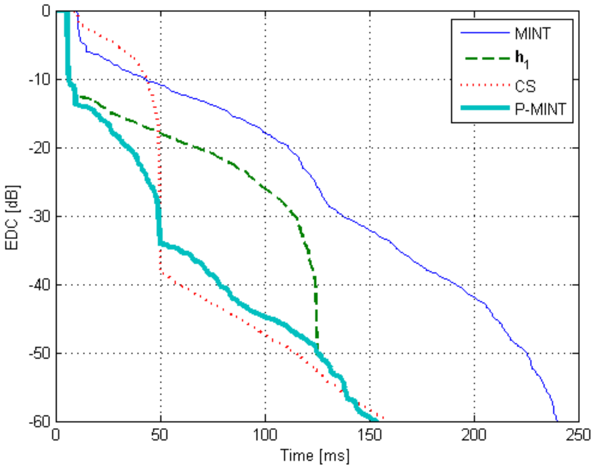

**Fig. 2**: Energy decay curves of MINT, CS, P-MINT, and $\mathbf{h}_1$ ($C_m = -33$ dB)

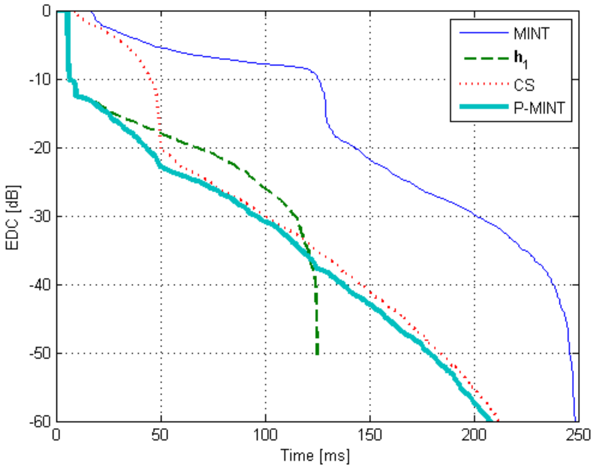

**Fig. 3**: Energy decay curves of MINT, CS, P-MINT, and $\mathbf{h}_1$ ($C_m = -15$ dB)

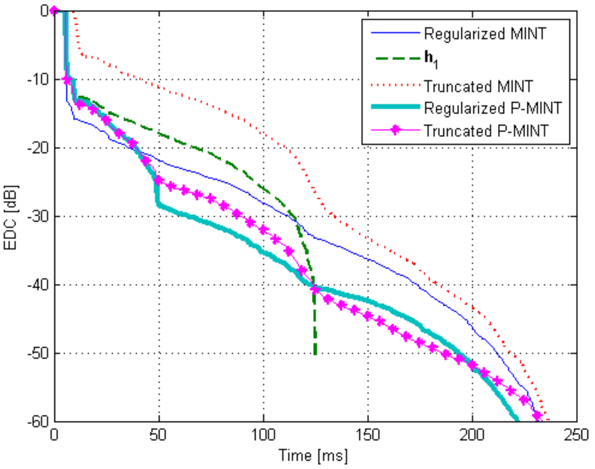

**Fig. 4**: Energy decay curves of regularized MINT, truncated MINT, regularized P-MINT, truncated P-MINT, and $\mathbf{h}_1$ ($C_m = -15$ dB)

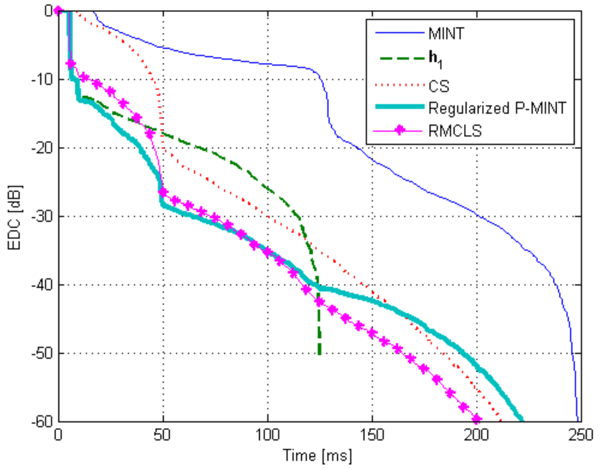

**Fig. 5**: Energy decay curves of MINT, CS, P-MINT, RMCLS, and $\mathbf{h}_1$ ($C_m = -15$ dB)

where $\mathbf{c}_p$ denotes the EIR obtained for a given regularization parameter $\delta$ or truncation value $k$. (Note that to make the implementation feasible, we have set the $-\infty$ taps in $\text{EDC}_{\mathbf{h}_1^d}$ to $-60$ dB). The optimal EDCs determined based on the above mentioned selection procedure are depicted in Fig. 4 for $C_m = -15$ dB. As illustrated in this figure, the truncated MINT solution completely fails to equalize the channel, whereas the regularized MINT fails at suppressing audible reverberation. On the other hand, the regularized and truncated P-MINT methods yield a higher performance, with the regularized P-MINT method outperforming all of the other considered approaches. Based on this experiment, we conclude that the regularized least-squares and the truncated singular value decomposition approach are more effective when applied to P-MINT than MINT. Furthermore, these experimental results show that regularized P-MINT appears to be the most robust method to channel estimation errors.

*Experiment 3.* Finally, we compare the proposed optimal regularized P-MINT solution to the RMCLS solution, which to the best of our knowledge, is the most robust multichannel equalization algorithm that has been proposed [6]. Fig. 5 depicts the obtained EDCs using MINT, CS, regularized P-MINT, and RMCLS for the channel estimation error $C_m = -15$ dB. As illustrated in this figure, also RMCLS fails to equalize the first part of $\mathbf{h}_1$, where the EDC is higher than $\text{EDC}_{\mathbf{h}_1}$. On the other hand, the regularized P-MINT approach yields an EDC that is always lower than $\text{EDC}_{\mathbf{h}_1}$, and even though the tail of its EDC is approximately $5$ dB higher than that of RMCLS, informal listening tests suggest that this reverberation is not audible. Therefore, the proposed regularized P-MINT method outperforms all the considered approaches in terms of robustness to estimation errors.

## 5. CONCLUSION

In this paper, we have presented a novel approach to partial channel equalization for speech dereverberation using the multiple-input/output inverse theorem and the first part of one of the estimated channels as the desired response (P-MINT). Based on experimental results, it has been shown that P-MINT outperforms MINT and CS in terms of robustness to channel estimation errors. In addition, we have investigated two methods to further increase the robustness against channel estimation errors, namely incorporating a regularization parameter in the inverse filter design and a truncated singular value decomposition approach. Experimental results show that the regularized P-MINT method exhibits the highest robustness of all the presented partial channel equalization techniques. Automating the selection of an optimal regularization parameter as well as the comparison of the presented techniques in terms of the perceptual quality of the perceived speech remain topics for future research.

## 6. 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] B. Radlovic, B. Williamson, and R. 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.

[3] M. Hasan and P. Naylor, "Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: robustness issue," in *Proc. EUSIPCO*, Florence, Italy, Sept. 2006.

[4] 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.

[5] Y. Nagata, Y. Tatekura, H. Saruwatari, and K. Shikano, "Iterative inverse filter relaxation algorithm for adaptation to acoustic fluctuation in sound reproduction system," *Electronics and Communications in Japan*, vol. 87, no. 7, July 2004.

[6] W. Zhang, E. Habets, and P. Naylor, "On the Use of Channel Shortening in Multichannel Acoustic System Equalization," in *Proc. IWAENC*, Tel Aviv, Israel, Sept. 2010.

[7] 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.

[8] J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. IWAENC*, Paris, France, Sept. 2006.

[9] 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.