EUSIPCO 2013 1569742013

ROBUST SPEECH DEREVERBERATION USING SUBBAND MULTICHANNEL LEAST SQUARES WITH VARIABLE RELAXATION

Felicia Lim and Patrick A. Naylor

Dept. of Electrical and Electronic Engineering, Imperial College London, UK
{felicia.lim06, p.naylor}@imperial.ac.uk

ABSTRACT

Multichannel equalization algorithms which are robust to system identification errors (SIEs) are important for practical speech dereverberation. We present an equalizer employing variable relaxation within the framework of the relaxed multichannel least squares (RMCLS) algorithm in frequency subbands. We show that varying the relaxation of constraints in RMCLS leads to a trade-off between robustness to SIEs and improved suppression of the early reflections that can be useful to achieve improved overall perceived speech quality after dereverberation processing. We then develop a method of controlling the amount of relaxation based on the expected level of SIE in each subband. Additionally, our algorithm guarantees robustness even in the presence of very high SIEs by backing off dereverberation in the relevant subbands.

Index Terms— Dereverberation, equalization, robustness, system identification errors, subband

1. INTRODUCTION

Speech signals recorded using hands-free communications devices typically suffer from reverberation due to the multipath propagation of the source signal to the microphones through acoustic channels. This degrades perceived speech quality and reduces the performance of other speech processing algorithms such as speech recognizers [1].

A promising approach to dereverberation is acoustic multichannel equalization, where the acoustic impulse responses (AIRs) between the source and microphones are estimated using blind system identification (BSI) algorithms [2, 3] and an inverse filter is subsequently designed to counteract the effect of room acoustics.

The problem is formulated as follows for a speech signal $s(n)$ propagating through an M-channel ($M \geq 2$) acoustic system modeled as M finite impulse responses, $\mathbf{h}_m = [h_m(0)\ h_m(1)\ \dots\ h_m(L-1)]^T$ for $m = 1, 2, \dots, M$. The reverberant signal at the m-th microphone is given by

$$\mathbf{x}_m(n) = \mathbf{H}_m^T \mathbf{s}(n) + \mathbf{v}(n), \tag{1}$$

where $\mathbf{x}_m(n) = [x_m(n)\ x_m(n-1)\ \dots\ x_m(n-L_i+1)]^T$, $\mathbf{s}(n) = [s(n)\ s(n-1)\ \dots\ s(n-L-L_i+2)]^T$ and $\mathbf{v}(n) = [v(n)\ v(n-1)\ \dots\ v(n-L_i+1)]^T$ are segments of the microphone, speech and noise signals respectively, $\mathbf{H}_m$ is the $(L+L_i-1) \times L_i$ convolution matrix of $\mathbf{h}_m$ and L_i is the equalizing filter length.

A set of equalizing filters $\mathbf{g}_m = [g_m(0)\ g_m(1)\ \dots\ g_m(L_i-1)]^T$ can be designed to counteract the effect of $\mathbf{h}_m$ to give an equalized impulse response (EIR) we denote as $\mathbf{d}$ such that

$$\mathbf{H}\mathbf{g} = \mathbf{d}, \tag{2}$$

where $\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2\ \dots\ \mathbf{H}_M]$, $\mathbf{g} = [\mathbf{g}_1^T\ \mathbf{g}_2^T\ \dots\ \mathbf{g}_M^T]^T$,

$$\mathbf{d} = [\underbrace{0\ 0\ \dots\ 0}_{\tau}\ 1\ 0\ \dots\ 0]^T_{[(L+L_i-1)\times 1]}, \tag{3}$$

and τ is a delay.

In practical applications, the AIRs are estimated as $\hat{\mathbf{h}}$ with system identification errors (SIEs). The problem is therefore to find a set of filters $\mathbf{g}$, given $\hat{\mathbf{h}}$, which will equalize $\mathbf{h}$ in a robust way to reduce the degradations due to reverberation in the speech signal.

For the ideal case with no SIEs, the multiple-input/output inverse theorem (MINT) [4] provides exact multichannel inverse filters $\mathbf{g} = \mathbf{H}^+\mathbf{d}$, where $\{\cdot\}^+$ denotes the Moore-Penrose pseudo-inverse, subject to [4, 5]:

C-1 $H_m(z)$, the z-transforms of $\mathbf{h}_m$, do not share any common zero.

C-2 $L_i \geq \lceil \frac{L-1}{M-1} \rceil$, where $\lceil \cdot \rceil$ is the ceiling operator.

However, in the presence of SIEs, the exact inverse filters of $\hat{\mathbf{h}}$ provided by MINT do not equalize $\mathbf{h}$ and reverberation is added to the EIR rather than suppressed as desired.

The relaxed multichannel least squares (RMCLS) algorithm [6] detailed in Section 2 improves robustness over MINT in moderate levels of SIEs in a manner desirable for speech dereverberation. However, its performance remains limited in severe levels of SIEs as additional reverberation can be introduced in the EIR. To improve the robustness of RMCLS, [7] incorporated regularization in the matrix inversion to reduce the energy of the inverse filters and consequently, the distortion introduced. An alternative approach in [8], gated subband RMCLS (G-RMCLS), implemented

RMCLS in subbands to reduce numerical errors from inverting large and poorly conditioned matrices, and introduced gated dereverberation in each subband to place an upper limit on the degradation introduced in the EIR.

In this work we extend [8] and investigate a way to control the amount of relaxation applied in subband RMCLS based on the level of SIEs to obtain a better solution than simple gating. As in previous work [6] the focus of this work is on equalizing the response of the acoustic channel and not estimating $s(n)$ in the presence of noise. The effect of the additive noise is to introduce SIEs into the estimation of $\mathbf{H}_m$ and hence our interest is to design an equalizer that is robust to these SIEs.

2. RELAXED MULTICHANNEL LEAST SQUARES

RMCLS [6] relaxes constraints placed by MINT to improve robustness to SIEs. Its motivation stems from the psychoacoustics principle that the late reverberant tail coefficients beyond approximately the first 0.05 s of an AIR are most damaging to perceived speech quality while the early coefficients do not impair speech intelligibility significantly [1]. RMCLS therefore aims to design a set of equalizing filters $\mathbf{g}$ to give an EIR where the early coefficients are unconstrained but the late coefficients should tend towards zero to suppress the late reverberation. To achieve the above, the following cost function is minimized

$$J = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2, \tag{4}$$

where $\hat{\mathbf{H}}$ is an estimate of $\mathbf{H}$ with SIEs, $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ and

$$\mathbf{w} = [\underbrace{1 \ \dots \ 1}_{\tau} \ \underbrace{1 \, 0 \ \dots \ 0}_{L_w} \ 1 \ \dots \ 1]^T_{[(L+L_i-1)\times 1]}. \tag{5}$$

The term L_w defines an interval referred to as the 'relaxation window', and typically corresponds to the region of the unconstrained early coefficients in the EIR. The first weight in the relaxation window is set to unity to avoid the trivial solution. The minimum ℓ_2-norm solution is then given by

$$\mathbf{g} = (\mathbf{W}\hat{\mathbf{H}})^{+}\mathbf{W}\mathbf{d}, \tag{6}$$

subject to conditions C-1 and C-2 in Section 1 being satisfied.

RMCLS is more robust than MINT given moderate SIEs. However, its performance deteriorates in severe levels of SIEs and can result in additional reverberation being introduced in the EIR when the robustness limits of RMCLS are exceeded.

3. GATED SUBBAND RMCLS

The G-RMCLS algorithm [8] extends the robustness of RMCLS and limits additional reverberation introduced in the EIR in high levels of SIEs. It implements RMCLS in subbands and gating equalization in each subband is employed such that equalization is applied only if the expected level of SIEs in that subband exceeds a predetermined threshold. The SIEs are quantified using normalized projection misalignment (NPM) [1] and the threshold, NPM_t, is selected as the case where the energy in the reverberant tail of the EIR is greater than the energy in the reverberant tail of $\mathbf{h}$. The energy in the reverberant tail is quantified with energy decay curves (EDCs) [1] and the region of reverberant tail is defined as $n > 0.05 f_s$, where n is the sample index of the EDC.

In practical applications, NPM in each subband can be estimated from the signal-to-noise ratio (SNR) for a given system identification algorithm such as shown in [3] using techniques for example based on non-intrusive SNR estimation [9, 10]. In this work, the oracle case where knowledge of the exact NPM in each subband is assumed to avoid introducing NPM estimation errors.

The design of subband equalizing filters $\mathbf{g}'$ requires subband estimated AIRs $\hat{\mathbf{h}}'_{km}$ to be found from $\hat{\mathbf{h}}_m$ such that the total transfer function of the subband filters is equivalent to the full-band filter up to an arbitrary scale factor and delay. A K-subband system with decimation factor N is first constructed based on the generalized discrete Fourier transform (GDFT) filter-bank [11]. Analysis filters $u_k(n)$ for $k = 0, \dots, K$ are obtained by modulating a prototype filter $p(n)$ of length L_{pr} as

$$u_k(n) = p(n) \cdot e^{j\frac{2\pi}{K}(k+k_0)(n+n_0)}, \tag{7}$$

where k_0 and n_0 are frequency and time offsets, set to $k_0 = 1/2$ and $n_0 = 0$ [12, 13]. Synthesis filters are given by the time-reversed and conjugated analysis filters [12], $v_k(n) = u_k^*(L_{\text{pr}} - n - 1)$. The parameters $K = 32$, $N = 24$ and $L_{\text{pr}} = 512$-taps were chosen for good trade-offs between aliasing suppression and sufficiently short subband equalization filters [13]. Complex subband decomposition [12, 13] is next employed to find $\hat{\mathbf{h}}'_{km}$ given by

$$\hat{\mathbf{h}}'_{km} = \mathbf{U}^{+}_{N,k}\mathbf{c}_{N,km}, \tag{8}$$

where

$$\mathbf{U}_{N,k} = \begin{bmatrix} u_k(0) & 0 & \cdots & 0 \\ u_k(N) & u_k(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ u_k(L_{\text{pr}}-1) & \cdots & \vdots & 0 \\ 0 & u_k(L_{\text{pr}}-1) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & u_k(L_{\text{pr}}-1) \end{bmatrix}$$

and $\mathbf{c}_{N,km} = [c_{km}(0), c_{km}(N), \dots, c_{km}\,(N(L-1))]^T$ is an $\lceil (L + L_{\text{pr}} - 1)/N \rceil \times 1$ vector with $c_{km}(n) = \hat{h}_m(n) * u_k(n)$. The length of $\hat{\mathbf{h}}'_{km}$ is $L' = \left\lceil \frac{L+L_{\text{pr}}-1}{N} \right\rceil - \left\lceil \frac{L_{\text{pr}}}{N} \right\rceil + 1$.

With this filter design, the first $K/2$ subbands are complex conjugates of the remaining subbands [12]. Therefore, processing of only the first $K/2$ subbands is required.

Given $\hat{\mathbf{h}}'_{km}$, $\hat{\mathbf{H}}'_k$ can be found in a similar way to $\hat{\mathbf{H}}$ and the subband RMCLS solution is given by modifying (6) as

$$\mathbf{g}'_k = (\mathbf{W}'_{r,k}\hat{\mathbf{H}}'_k)^+\mathbf{W}'_{r,k}\mathbf{d}' \quad \text{for } k = 0, 1, \ldots, K/2-1, \tag{9}$$

where $\mathbf{W}'_{r,k} = \text{diag}\{\mathbf{w}'_{r,k}\}$ with

$$\mathbf{w}'_{r,k} = [\underbrace{1 \ \ldots \ 1}_{\tau'} \ \underbrace{1 \, 0 \ \ldots \ 0}_{L'_{w,k}} \ 1 \ \ldots \ 1]^T_{(L'+L'_i-1)\times 1}, \tag{10}$$

where $\tau' = \lceil \tau/N \rceil$ and $L'_{w,k} = \lceil L_w/N \rceil$.

The gated approach to dereverberation in G-RMCLS ensures robustness to severe levels of SIEs, but does not otherwise control the amount of dereverberation applied. It is therefore desirable to achieve better control of the performance of the equalizer.

4. VARIABLE RELAXATION RMCLS

The length of L_w in (5) applied in RMCLS has a trade-off between suppression of the early coefficients and late coefficients, as will be demonstrated with simulation results in Section 5.1. The aim of variable relaxation RMCLS (VR-RMCLS) is to exploit this trade-off by varying L_w independently in each subband according to the corresponding level of NPM (known or estimated as described in Section 3). This enables potentially better control of robustness over simple gating in G-RMCLS. In this manner, subbands with worse SIEs can employ longer L_w to increase robustness at the expense of lower dereverberation performance. In subbands with small SIEs, less robustness is required and shorter L_w can be used to improve dereverberation by suppressing more of the EIR coefficients. In the lower limit where there are no SIEs, $L_w = 0$ is used, giving the MINT solution [4]. In subbands with exceptionally high SIEs, gating can be applied in the same way as G-RMCLS to avoid adding reverberation in the EIR, thereby exploiting a merge of the advantageous properties of both G-RMCLS and VR-RMCLS. The remainder of this section discusses practical considerations and the method of determining L_w in each subband, $L'_{w_o,k}$.

The value of $L'_{w_o,k}$ is chosen in this work as a function of NPM as follows. Given an NPM and its corresponding $\hat{\mathbf{h}}'_{km}$, subband RMCLS is first performed for a range of $L'_{w,k}$ and the subband EIRs found as

$$\text{EIR}_k = \mathbf{H}'_k\mathbf{g}'_k. \tag{11}$$

The known initial delays caused by subband filtering are removed and the EDCs are calculated. For a given NPM, $L'_{w_o,k}$ is selected to meet two criteria. The first criteria is to suppress the reverberant tail of the EIR to an acceptable level, the choice of which is discussed below. The second criteria is to avoid introducing additional degradation in the reverberant tail of the EIR over the AIRs. Two threshold EDCs are defined for each of the criteria above. The first threshold, EDC_{te}, defines the maximum acceptable level of reverberant tail suppression in this work as the maximum EDC across all subbands for $L'_{w,k} = \lceil 0.05 f_s/N \rceil$ at sample index $n_r = \lceil 0.05 f_s/N \rceil + 1$. The second threshold, EDC_{th}, defines the EDC of $\mathbf{h}'_{k,m}$ at n_r, where $\mathbf{h}'_{k,m}$ is found in a similar manner as $\hat{\mathbf{h}}'_{k,m}$ using $\mathbf{h}$. The value of $L'_{w_o,k}$ for the NPM under consideration can now be selected as the minimum $L'_{w,k}$ in each subband where its corresponding EDC value at n_r is below the minimum value of EDC_{th} and EDC_{te}. In the case where no $L'_{w,k}$ satisfies the above, the NPM is considered to be too large and the gating dereverberation method in G-RMCLS is applied. The $L'_{w_o,k}$ values can be pre-trained for a given NPM such that for practical application, only NPM estimates are required.

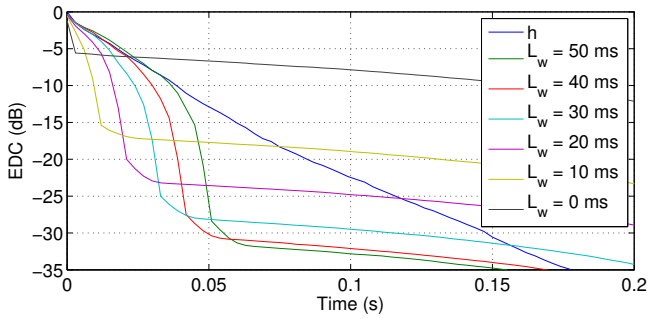
5. SIMULATIONS AND RESULTS

Two simulations were performed. Simulation 1 illustrates the basic concept of VR-RMCLS by varying L_w in single subbands to show the effect on the robustness of the equalizer. Simulation 2 evaluates the performance of VR-RMCLS against RMCLS and G-RMCLS with SIEs in all subbands.

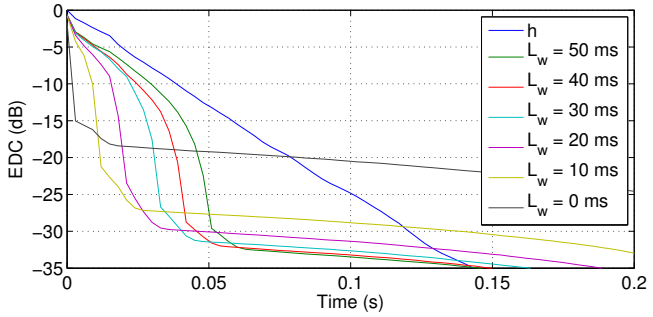
5.1. Simulation 1

A 2-channel system was simulated using the image method [14, 15] for a room size of 6.4 x 5 x 3.6 m with a distance of 2 m between the source and centre of the microphone array, an inter-microphone distance of 0.1 m and reverberation time $T_{60} = 0.3$ s. The fractional delay before the direct path in the AIRs was removed such that $\tau = 0$ and the channels truncated to $L = 2000$. Input speech signals were taken from the TIMIT database [16] and resampled to $f_s = 8\,\text{kHz}$. The AIRs were subsequently filtered into $K = 32$ subbands and SIEs artificially introduced by addition of subband filtered white Gaussian noise to achieve a desired level of NPM [17]. In this work, NPM $= -30, -27, \ldots, -6$ dB were simulated. Complex subband decomposition was applied as (8) and subband RMCLS equalizers were designed using $L'_{w,k}$ derived from $L_w = \{0, 0.01, \ldots, 0.05\} f_s$. EIRs and their corresponding EDCs were calculated in subbands as (11). Each simulation was repeated 50 times with randomly varying locations of the source and microphone array while maintaining constant source-sensor distances to give spatially averaged results.

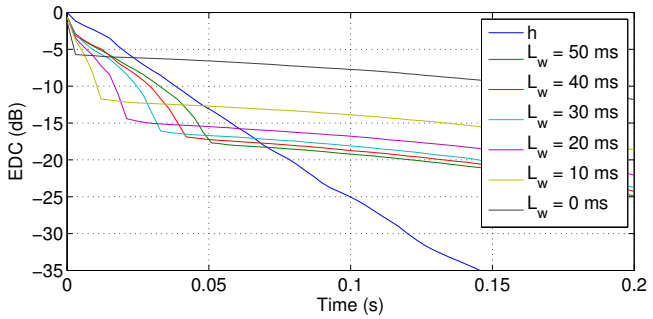
Illustrative examples of some subband EDC results are given in Figure 1, where it can be seen that the choice of $L'_{w,k}$ involves a trade-off between the suppression of early coefficients and late reverberant tail of the EIRs. Furthermore, the levels of suppression achieved can be seen to vary between subbands for the same NPM and L_w values, and it is this characteristic which is exploited to select the desirable subband $L'_{w,k}$ values. From these results, $L'_{w_o,k}$ values are found according to the method described in Section 4.



(a) NPM = -30 dB, subband k = 1



(b) NPM = -30 dB, subband k = 7



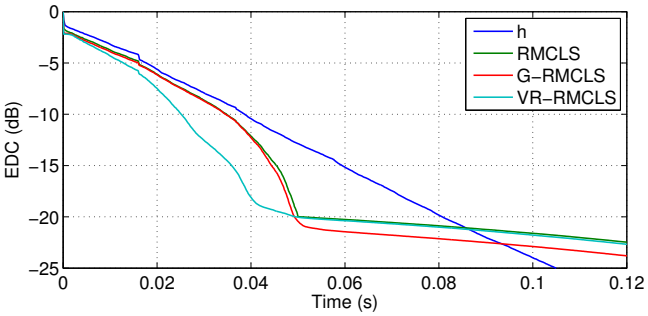
(c) NPM = -15 dB, subband k = 7

Fig. 1. Averaged EDCs in different subbands for different NPM values.

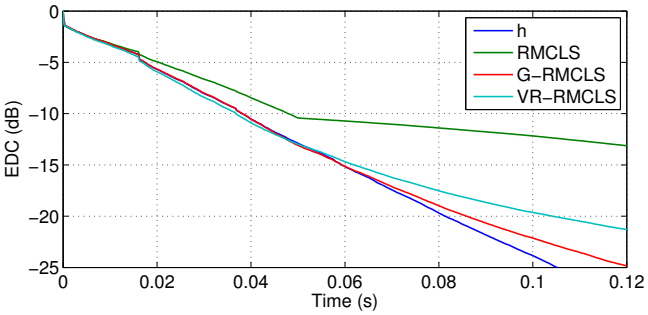
5.2. Simulation 2

Two simulations were run with SIEs of varying NPM levels in all subbands. The performance of VR-RMCLS was evaluated against RMCLS and G-RMCLS based on the full-band EIR and equalized speech signal. In addition to EDC, evaluation of perceived speech quality was carried out using ITU-T P.862 (PESQ) scores, which provides an estimate using a predicted mean opinion score (PMOS) ranging from $1-4.5$ [18]. To facilitate a comparison between the microphone and equalized speech signals, the difference in their PESQ scores was calculated as ΔP. It is desirable to simply observe $\Delta P > 0$ since PESQ is known to not be a reliable measure of reverberation, and instead was selected specifically to provide some assurance that there was no measurable degradation in overall speech quality.

The same 2-channel acoustic system from Section 5.1 was simulated. SIEs were artificially added in subbands with the $K/2$ subband NPM levels pseudorandomly drawn from a uniform distribution on two intervals with moderate SIEs, $(-25, -15)$ dB, and severe SIEs, $(-15, -5)$ dB. Subband equalization for VR-RMCLS was performed using the $L'_{w_o,k}$ values found in Section 5.1. Each simulation was repeated 50 times with randomly varying locations of the source and microphone array while maintaining constant source-sensor distances to give spatially averaged results.



(a) SIEs with NPM drawn from the interval $(-25, -15)$ dB.



(b) SIEs with NPM drawn from the interval $(-15, -5)$ dB.

Fig. 2. Averaged EDCs (evaluated over the full bandwidth).

NPM (dB)	RMCLS	G-RMCLS	VR-RMCLS
$(-25, -15)$	0	0.1	0.2
$(-15, -5)$	-0.3	0	0

Table 1. Averaged ΔP showing that the G- and VR-RMCLS methods do not degrade PESQ even with severe SIEs.

The EDCs based on the reconstructed full-band EIRs are shown in Fig. 2 and the ΔP results are shown in Table 1. In the presence of moderate SIEs, VR-RMCLS improved early coefficients suppression up to -5.7 dB over G-RMCLS and full-band RMCLS for $t \leq 0.05$ s. In the presence of severe SIEs, both VR-RMCLS and G-RMCLS successfully avoid introducing additional distortion over the true AIRs, except where the EIR is already suppressed by at least the EDC value of the AIRs at $t = 0.05$ s. The ΔP scores indicated

that VR-RMCLS did not degrade the perceived quality of the equalized speech signal compared to the microphone signal, which is desirable.

6. CONCLUSION

We have presented a novel equalizer for dereverberation of speech employing variable relaxation of RMCLS in frequency subbands. We demonstrated through experimental results that 1) the robustness of RMCLS can be varied as desired from subband to subband to exploit as well as possible the available accuracy of the BSI, and 2) the amount of relaxation applied for RMCLS in subbands involves a trade-off between the robustness to SIEs in terms of the reverberation tail suppression, and the suppression of early coefficients of the EIR. The VR-RMCLS algorithm was proposed, exploiting this trade-off and further employs gated dereverberation from G-RMCLS to guarantee robustness even in the presence of severe SIEs. Experimental results demonstrate that improved suppression of the early coefficients was achieved without significantly adversely affecting the robustness of the reverberant tail suppression.

7. REFERENCES

[1] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[2] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[3] M.A. Haque and M.K. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 305–308, 2008.

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

[5] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. Signal Process.*, vol. 46, pp. 215–218, 1998.

[6] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[7] I. Kodrasi and S. Doclo, "Robust partial multichannel equalization techniques for speech dereverberation," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Apr. 2012.

[8] F. Lim and P. A. Naylor, "Robust low-complexity multichannel equalization for dereverberation," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

[9] R.C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 4266–4269.

[10] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383 –1393, May 2012.

[11] S. Weiss and R. W. Stewart, *On adaptive filtering in oversampled subbands*, Shaker Verlag, 1998.

[12] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2730–2743, Nov. 2002.

[13] N. D. Gaubitch and P. A. Naylor, "Equalization of multichannel acoustic systems in oversampled subbands," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1061–1070, Aug. 2009.

[14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[15] E. A. P. Habets, "Room impulse response (RIR) generator," http://home.tiscali.nl/ehabets/rirgenerator.html, May 2008.

[16] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," Technical report, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, Dec. 1988.

[17] W. Zhang and P. A. Naylor, "An algorithm to generate representations of system identification errors," *Research Letters in Signal Processing*, vol. 2008, pp. 13:1–13:4, Jan. 2008.

[18] 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," Recommendation P.862, International Telecommunications Union (ITU-T), Feb. 2001.