ADAPTIVE INVERSE FILTERING OF ROOM ACOUSTICS

[†]*Wancheng Zhang,* [‡]*Andy W. H. Khong, and* [†]*Patrick A. Naylor*

[†]Imperial College London, Department of Electrical and Electronic Engineering
[‡]Nanyang Technological University, School of Electrical and Electronic Engineering

ABSTRACT

Equalization techniques for high order, multichannel, FIR systems are important for dereverberation of speech observed in reverberation using multiple microphones. In this case the multichannel system represents the room impulse responses (RIRs). The existence of near-common zeros in multichannel RIRs can slow down the convergence rate of adaptive inverse filtering algorithms. In this paper, the effect of common and near-common zeros on both the closed-form and the adaptive inverse filtering algorithms is studied. An adaptive shortening algorithm of room acoustics is presented based on this study.

1. INTRODUCTION

In hands-free communications, the speech signal can be distorted by room reverberation, resulting in reduced intelligibility to listeners. One method to achieve dereverberation is to perform identification and inverse filtering of the room impulse responses (RIRs). The methodology is illustrated in Fig. 1. Consider a
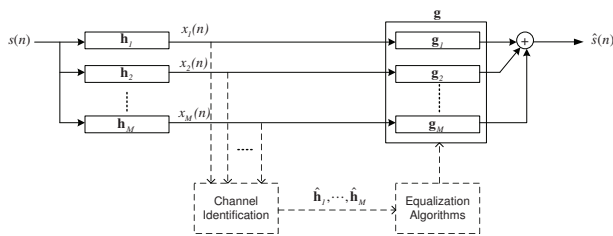

Fig. 1. Illustration of identification and inverse filtering of acoustic systems.

clean speech signal $s(n)$ propagating through M acoustic channels, which are characterized by their impulse responses $\mathbf{h}_m = [h_m(0)\ h_m(1)\ \cdots\ h_m(L-1)]^T$, $m = 1, \cdots, M$, where L is the length of the RIRs, and $\{\cdot\}^T$ denotes the transpose operation. Using the reverberant speech signals $x_m(n)$, $m = 1, \cdots, M$, estimates of the RIRs $\mathbf{h}_m$, $m = 1, \cdots, M$ can be obtained with blind system identification techniques, such as in [1]. Then, with the estimates $\hat{\mathbf{h}}_m$, $m = 1, \cdots, M$, an inverse filtering system $\mathbf{g} = [\mathbf{g}_1^T\ \mathbf{g}_2^T\ \ldots\ \mathbf{g}_M^T]^T$, which is formed by stacking column vectors of the filters $\mathbf{g}_m = [g_m(0)\ g_m(1)\ \ldots\ g_m(L_i - 1)]^T$ of each channel, can be designed with some equalization algorithm. Then, by filtering $x_m(n)$ using the inverse filtering system $\mathbf{g}$, we expect a good estimate $\hat{s}(n)$ of $s(n)$ can be obtained. In this paper, we do not consider the possible errors induced by the system identification, so we assume $\hat{\mathbf{h}}_m = \mathbf{h}_m$, $m = 1, \cdots, M$.

Traditionally, inverse systems can be obtained, for the single channel case, by using the method of least squares (LS), or employ-

ing the multiple-input/output inverse theorem (MINT) when multiple microphones are deployed [2]. Although LS inverse filters can be used to approximately invert the RIRs, which are usually of non-minimum phase, such techniques necessitate the use of very long inverse filters as well as significantly long delay [3]. From a theoretical perspective, reverberation can be completely removed by using multiple microphones and techniques based on MINT in the case that the multichannel room transfer functions (RTFs) do not share any common zeros [2]. In practice, MINT is computationally expensive [4], and this motivates the use of the subband algorithm [4]. On the other hand, multichannel adaptive systems have been used for acoustic system equalization [5][6], and it is shown that an identical inverse filtering system to MINT can be obtained [7]. However, the existence of common and *near-common* zeros [8] causes problems in both closed-form and adaptive inverse filtering algorithms. MINT has been generalized to a multichannel least squares (MCLS) method [4], and it can overcome the problems due to the existence of common zeros. In adaptive inverse filtering, we will show that near-common zeros slow down the convergence rate of the adaptive algorithms. Therefore, after any finite period of adaptation, the tail of the equalized impulse response will not be completely suppressed.

In this paper, we address the performance degradation in adaptive inverse filtering algorithms due to the presence of near-common zeros. We achieve this by leaving the parts of the RIRs with common and near-common zeros without equalization, which leads to a process known as channel shortening. Channel shortening has been extensively developed in the context of digital communications to mitigate the inter-symbol and inter-carrier interference. The techniques are firstly developed for the single-input/single-output (SISO) cases. Both closed form [9] and, iterative and adaptive [10][11] methods have been well studied. These techniques have been extended to the multiple-input/multiple-output (MIMO) case in [12][13]. A common frame work and an overview of the design techniques for channel shortening can be found in [14]. The motivation behind employing such techniques for our acoustic system equalization application is based on the fact that the early reflections in room acoustics can, in certain cases, enhance the speech intelligibility [15]. Therefore, it can be argued that it is not necessary to use the delta function as the target impulse response (TIR) in RIRs equalization for the purpose of dereverberation. Shortening the RIRs may therefore be indeed satisfactory for enhancing the quality and intelligibility of reverberant speech. By relaxing the TIR to be less constrained than the delta function, we expect that the common and near-common parts of the RIRs can be manifest in the early part of the equalized impulse response and the equalization tail correspondingly suppressed.

First, we will study the LS and MCLS algorithms. It will be shown that when common zeros exist, the MCLS is able to invert those parts of the channels with factors which are not common in the multichannel RIRs and to perform the LS inversion on the parts with common zeros. Then, the performance of an adaptive inverse fil-

tering algorithm when common or near-common zeros exist will be studied. It will be shown that the near-common zeros can slow down the convergence rate of the adaptive algorithm. After this, some improvements to the adaptive algorithm based on this study will be made, which will lead to an adaptive channel shortening algorithm for the RIRs.

2. FORMULATION OF INVERSE FILTERING

Inverse filtering of room acoustics aims to use an inverse system of the RIRs to compensate for the distortion to the original signal caused by the RIRs. It usually aims to force the equalized impulse response

$$\mathbf{y} = \mathbf{h}_1 * \mathbf{g}_1 + \mathbf{h}_2 * \mathbf{g}_2 + \cdots + \mathbf{h}_M * \mathbf{g}_M = \sum_{m=1}^{M} \mathbf{h}_m * \mathbf{g}_m \quad (1)$$

to be a target impulse response (TIR) of the delta function

$$\mathbf{d} = \underbrace{[1 \ 0 \ \dots \ 0]}_{L+L_i-1}{}^T, \quad (2)$$

where $*$ denotes linear convolution. The aim is to minimize the cost function

$$J = \|\mathbf{d} - \mathbf{y}\|^2, \quad (3)$$

where $\| \cdot \|$ denotes the Euclidean norm.

The inverse system $\mathbf{g}$ can be obtained by

$$\mathbf{g} = \mathbf{H}^+\mathbf{d}, \quad (4)$$

where $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \cdots \ \mathbf{H}_M]$ is the system matrix, and $\{\cdot\}^+$ denotes pseudo inverse. $\mathbf{H}_m$ is an $(L + L_i - 1) \times L_i$ convolution matrix of $\mathbf{h}_m$

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \cdots & 0 \\ h_m(1) & h_m(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_m(L-1) & \cdots & \vdots & \vdots \\ 0 & h_m(L-1) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & h_m(L-1) \end{bmatrix}.$$

If $M = 1$, (4) gives a single channel LS optimal inverse system [2]. If $M \geq 2$, the multichannel RTFs do not share any common zeros, and $L_i \geq L_c$, where $L_c = \lceil \frac{L-1}{M-1} \rceil$ is defined as the critical filter length of the inverse system, (4) gives an exact inverse system, with which the TIR (2) can be perfectly achieved [2]. For the case that multichannel RTFs have common zeros, or $L_i < L_c$, (4) gives an MCLS inverse system [4].

An inverse system $\mathbf{g}$ minimizing (3) can also be obtained adaptively, which will be introduced later in this paper.

In this paper, we will focus on the $M = 2$ channel systems to study the effect of common and near-common zeros on the inverse filtering algorithms. Suppose we have an $M = 2$ channel system, and the design length of the filters of the inverse system is $L_i = L_c$, which leads to a square system matrix $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$. The degree of rank deficiency of $\mathbf{H}$ is equivalent to the number of common zeros of the transfer functions of $\mathbf{h}_1$ and $\mathbf{h}_2$ [16]. If these two channels do not have any common zeros, i.e. $\mathbf{H}$ is of full rank, then $\mathbf{H}^+ = \mathbf{H}^{-1}$. If these two channels are identical, i.e. all zeros of these two channels
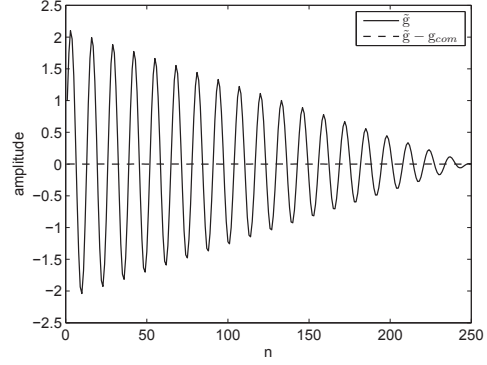

Fig. 2. $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{g}} - \mathbf{g}_{com}$.

are correspondingly common, then calculating $\mathbf{H}^+$ is identical to calculating the single channel LS inverse system.

To study the effect of common and near-common zeros by experiments, we will use some synthetic impulse responses, the zeros of which are manually located on the z-plane.

3. FROM LS TO MINT

Studies of RIRs measured in rooms indicate that common zeros are normally present. Therefore, the inverse system $\mathbf{g}$ in (4) is usually given by the MCLS. In this Section, we will show that the MCLS works to fully invert the non-common parts in the impulse responses, and performs LS inversion on the common parts.

Two synthetic impulse responses $\mathbf{h}_1$ and $\mathbf{h}_2$, the transfer functions of which have two common zeros are used to show this. These two zeros are a pair of conjugate zeros. The length of $\mathbf{h}_1$ and $\mathbf{h}_2$ is $L = 127$. $\mathbf{h}_1$ and $\mathbf{h}_2$ can be written as

$$\mathbf{h}_1 = \tilde{\mathbf{h}}_1 * \mathbf{h}_{com}, \quad (5)$$
$$\mathbf{h}_2 = \tilde{\mathbf{h}}_2 * \mathbf{h}_{com}, \quad (6)$$

where $\tilde{\mathbf{h}}_1$ and $\tilde{\mathbf{h}}_2$ are the non-common parts, and $\mathbf{h}_{com}$, which is of 3 taps in this example, is the common part.

Consider an inverse system $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T]^T$, where $L_i = L_c$, is obtained by using (4). By applying $\mathbf{g}$ to $\mathbf{h} = [\mathbf{h}_1 \ \mathbf{h}_2]$, an equalized impulse response can be obtained,

$$\begin{aligned} \mathbf{y} &= \mathbf{g}_1 * \mathbf{h}_1 + \mathbf{g}_2 * \mathbf{h}_2 \\ &= (\mathbf{g}_1 * \tilde{\mathbf{h}}_1 + \mathbf{g}_2 * \tilde{\mathbf{h}}_2) * \mathbf{h}_{com} \\ &= \tilde{\mathbf{g}} * \mathbf{h}_{com}, \quad (7) \end{aligned}$$

where $\tilde{\mathbf{g}}$ is of $L + L_i - 3$ taps. On the other hand, an LS optimal inverse filter, $\mathbf{g}_{com}$, of $\mathbf{h}_{com}$, with a design length of $L + L_i - 3$ can be obtained by the LS algorithm.

Experiment 1
In this experiment, we study the relationship between $\tilde{\mathbf{g}}$ and $\mathbf{g}_{com}$. Figure 2 shows that $\tilde{\mathbf{g}}$ and $\mathbf{g}_{com}$ are identical, which means that the MCLS performs an inversion on the non-common parts in full, and performs LS inversion on the common parts. In practice, to avoid the problems caused by rounding errors in computing $\mathbf{H}^+$, singular values and their singular vectors corresponding to the near-common zeros of very small δ, found to be of order, for example, 10^{-16} in our MATLAB simulations which use the IEEE floating-point
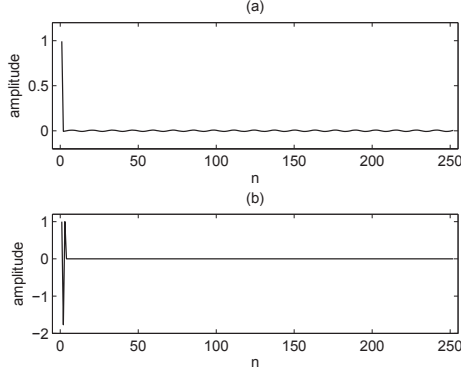
789

Fig. 3. Equalization results for (a) TIR from (2) and (b) TIR from (8).

Fig. 4. Convergence of J.

double-precision computation, are also truncated. This can be understood in another way that the near-common zeros of very small δ are processed as common zeros in MCLS, the order of which is subject to different numerical computation systems.

Experiment 2
In this experiment, we will use

$$\mathbf{d} = [\mathbf{h}_{com}^T \ 0 \ \cdots \ 0]^T \tag{8}$$

as the TIR to calculate $\mathbf{g}$ from $\mathbf{h} = [\mathbf{h}_1 \ \mathbf{h}_2]$. The equalization result with (2) as TIR is shown in Fig. 3(a), and Fig. 3(b) shows the result obtained with (8). It can be seen that in Fig. 3(a) that the equalization result is equal to a LS inversion of $\mathbf{h}_{com}$ with the characteristic non-zero tail exhibiting ripple. In Fig. 3(b), since no attempt is made to equalize the common part $\mathbf{h}_{com}$, the equalization tail is completely suppressed with no evidence of ripple.

4. ADAPTIVE INVERSE FILTERING

In this Section, an adaptive algorithm aiming to minimize the cost function (3) using the steepest descent (SD) method [17] will be proposed.

The gradient is given by

$$\nabla J = -2\mathbf{H}^T\mathbf{d} + 2\mathbf{H}^T\mathbf{H}\mathbf{g} \tag{9}$$

and the inverse system $\mathbf{g}$ can be obtained by

$$\mathbf{g}(k+1) = \mathbf{g}(k) - \mu\nabla J, \tag{10}$$

where k denotes the index of iteration, and μ is the step-size. The algorithm is given in Algorithm 1.

Algorithm 1 Proposed adaptive inverse filtering.

$\quad \mathbf{g}(0) = \mathbf{0}_{M(L+L_i-1)}$
$\quad \mathbf{b} = \mathbf{H}^T\mathbf{d}, \ \mathbf{A} = \mathbf{H}^T\mathbf{H}$
$\quad \textbf{for } k = 0, 1, 2, \ldots \textbf{ do}$
$\quad\quad \nabla J = -2\mathbf{b} + 2\mathbf{A}\mathbf{g}(k)$
$\quad\quad \mathbf{g}(k+1) = \mathbf{g}(k) - \mu\nabla J$
$\quad \textbf{end for}$
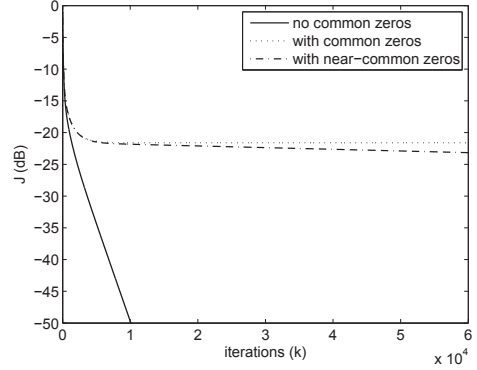

In the following, the effect of common or near-common zeros on the performance of Algorithm 1 will be studied.

Experiment 3
In this experiment, Algorithm 1 will be studied for the case in which no common or near-common zeros exist. The first impulse response used is $\mathbf{h}_1$ in (5). The second impulse response will include $\tilde{\mathbf{h}}_2$ in (6), but the other part is different from $\mathbf{h}_{com}$, to ensure that the two channels do not share common zeros. The convergence of J in (3) is shown in Fig. 4.

Experiment 4
In this experiment, Algorithm 1 will be studied for the case that common zeros exist. The impulse responses used are $\mathbf{h} = [\mathbf{h}_1 \ \mathbf{h}_2]$ in (5) and (6). The convergence of J is shown in Fig. 4.

Experiment 5
Algorithm 1 will next be studied for the case that near-common zeros exist. The first impulse response will be $\mathbf{h}_1$ in (5), and the second is obtained by replacing $\mathbf{h}_{com}$ with some impulse response with zeros separated by $\delta = 1 \times 10^{-4}$ from the corresponding zeros of the first channel. The convergence of J is shown in Fig. 4.

We can see in Fig. 4 that without common zeros, J converges quickly. When common zeros exist, Fig. 4 shows that J converges to an asymptotic performance level of about -21 dB, which corresponds to the LS inverse filtering of $\mathbf{h}_{com}$ shown in Fig. 3(a). When near-common zeros exist, J converges quickly to the asymptotic level of the case when common zeros exist, and then continues to converge but very slowly. Therefore, we can conclude that near-common zeros will slow down the convergence rate of Algorithm 1 after the adaptive filter has first been able to equalize the parts without common or near-common zeros.

5. CHANNEL SHORTENING

In this Section, we will study the performance of Algorithm 1 with some different TIRs, for the cases that common or near-common zeros exist.

Experiment 6
As in Exp. 2, we use (8) as the TIR when common zeros exist. We can see in Fig. 5(a) that J converges quickly. Figure 5(b) shows that the equalized impulse response is identical to the one given in Fig.
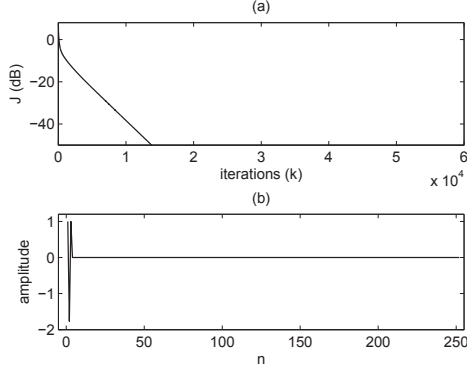
790

Fig. 5. Adaptive equalization with TIR of (8).

Fig. 6. Adaptive channel shortening with $L_r = 8$.

3(b) which is obtained using the closed-form MCLS.

However, in practice, we do not know the zeros of the true RTFs, and we do not know how many common or near-common zeros exist. Therefore, we cannot use an exactly known TIR such as (8) which we have used in Exp. 6. In room acoustics, the early reflections can enhance the speech intelligibility in certain circumstances [15]. Therefore, we relax the early part of TIR, to make it less constrained than the delta function (2). We propose to achieve this by using a weighting function in the cost function

$$J = \|\mathbf{w} \circ (\mathbf{d} - \mathbf{y})\|^2, \tag{11}$$

where

$$\mathbf{w} = [\underbrace{1\,0\,\ldots\,0}_{L_r}\,1\,\ldots\,1]^T \tag{12}$$

is the weighting function and $\circ$ denotes the Hadamard product. Here L_r is the length of the 'relaxing' window. We use $w(1) = 1$ to avoid the trivial solution.

With the weighting function used, the gradient of J defined in (11) can be written as

$$\nabla J = -2(\mathbf{WH})^T \mathbf{d} + 2(\mathbf{WH})^T (\mathbf{WH})\mathbf{g}, \tag{13}$$

where $\mathbf{W} = \mathrm{diag}\{\mathbf{w}\}$. The corresponding channel shortening algorithm to compute the shortening system $\mathbf{g}$ is given in Algorithm 2.

Algorithm 2 Proposed adaptive channel shortening.

$\mathbf{g}(0) = \mathbf{0}_{M(L+Li-1)}$
$\mathbf{b} = (\mathbf{WH})^T \mathbf{d}$, $\mathbf{A} = (\mathbf{WH})^T (\mathbf{WH})$
for $k = 0, 1, 2, \ldots$ **do**
   $\nabla J = -2\mathbf{b} + 2\mathbf{Ag}(k)$
   $\mathbf{g}(k+1) = \mathbf{g}(k) - \mu \nabla J$
end for

By using the 'relaxing' window, the equalization tail may still not be able to fully removed, for example, when L_r is less than the length of $\mathbf{h}_{com}$. However, any L_r greater than one can reduce the effect of the common and near-common zeros on the adaptive inverse filtering algorithm. Tests show that the equalization tail does not need to be fully suppressed in speech dereverberation. It is satisfactory to suppress it to some given level.
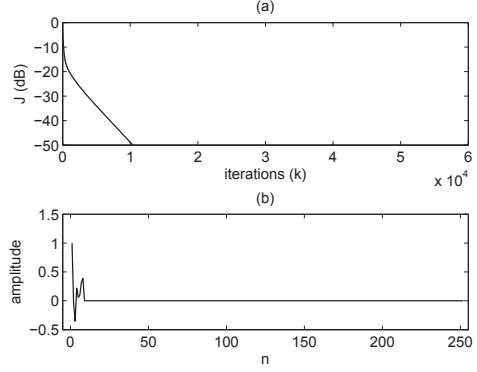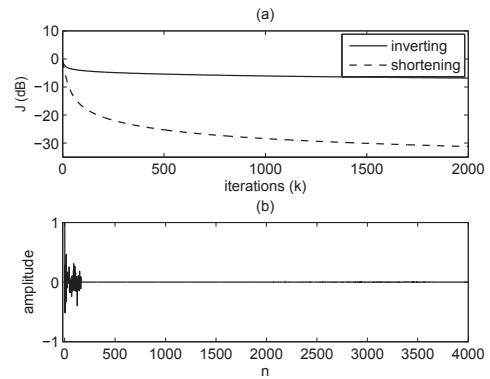

Fig. 7. Adaptive inverse filtering and shortening of RIRs.

Experiment 7
In this experiment, we will use $L_r = 8$ to test Algorithm 2. The impulse responses used in Exp. 5 (near-common zeros case) will be used. It can be seen in Fig. 6(a) that with the 'relaxing' window employed, J converges more quickly. The equalized impulse response is given in Fig. 6(b). We can see that after the 8th tap, the late part is fully suppressed.

6. ADAPTIVE CHANNEL SHORTENING USED IN TRUE ROOM ENVIRONMENTS

In this Section, the adaptive channel shortening algorithm will be tested with true RIRs.

Experiment 8
In this experiment, a $M = 2$ channel acoustic system will be used and the RIRs are taken from the MARDY database [18]. The length of the RIRs is $L = 2000$ with a sampling frequency of 8 kHz. The filter length of the shortening system L_i is used as $L_i = L_c = 1999$.

Since reflections arriving within 20 ms (160 taps) of the direct sound cause little or no disturbance in hearing even when the amplitude of the reflections is greater than the direct sound [15], we will use $L_r = 160$ in (12) in this experiment.

The comparison of convergence of J of inverse filtering ($L_r = 1$) and shortening ($L_r = 160$) is shown in Fig. 7(a). The shortening result at iteration 2000 is shown in Fig. 7(b). We can see that for

791

shortening, J converges more quickly than inverse filtering.

7. CONCLUSIONS

In this paper, we analyzed the performance of LS and MINT algorithms for the inverse filtering of RIRs. An adaptive approach for inverse filtering of room acoustics has been introduced and studied. Also, an adaptive channel shortening algorithm has been developed. Experiments show that when common zeros among multichannel RTFs exist, the MCLS inverts the non-common parts of the RIRs and performs LS inversion on the common parts. For adaptive inverse filtering, the existence of near-common zeros slows its convergence rate. Adaptive channel shortening can speed up the convergence rate and effectively suppress the late part of the RIRs.

8. REFERENCES

[1] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 10, Oct. 1998.

[2] M. Miyoshi and K. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 145–152, 1988.

[3] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Int. Workshop Acoust. Echo Noise Control*, Eindhoven, Sept 2005.

[4] N. D. Gaubitch and P. A. Naylor, "Equalization of multichannel acoustic systems in oversampled subbands," *accepted for publication in IEEE Trans. Audio, Speech, Language Processing*.

[5] S. J. Elliott and P. A. Nelson, "Algorithm for multichannel LMS adaptive filtering," *Electronics Letters*, vol. 21, no. 21, pp. 979–981, Oct. 1985.

[6] S. J. Elliott, I. M. Stothers, and P. A. Nelson, "A multiple error LMS algorithm and its application to the active control of sound and vibration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 10, pp. 1423–1434, Oct. 1987.

[7] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 3, pp. 185–192, May 1995.

[8] A. W.H. Khong, X. S. Lin, and P. A. Naylor, "Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization," in *ICASSP*, 2008, pp. 229–232.

[9] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, Dec. 1996.

[10] M. Nafie and A. Gatherer, "Time-domain equalizer training for ADSL," in *IEEE Int. Conf. Commun*, June 1997, vol. 2, pp. 1085–1089.

[11] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Process. Lett.*, vol. 9, pp. 341–343, Nov. 2002.

[12] N. Al-Dhahir, "FIR channel shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[13] R. K. Martin, J. M. Walsh, and C. R. Johnson Jr., "Low-complexity MIMO blind, adaptive channel shortening," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1324–1334, Apr. 2005.

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

[15] H. Kuttruff, *Room Acoustics, 4th edition*, Taylor & Frances, 2000.

[16] S. Barnett, "Degrees of greatest common divisors of invariant factors of two regular polynomial matrices," *Proc. Camb. Phil. Soc.*, vol. 66, pp. 241–245, 1970.

[17] S. S. Haykin, *Adaptive filter theory, 4th edition*, Prentice Hall, 2002.

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