AN EXPERIMENTAL STUDY OF THE ROBUSTNESS OF MULTICHANNEL INVERSE FILTERING SYSTEMS TO NEAR-COMMON ZEROS

Wancheng Zhang, and Patrick A. Naylor

Department of EEE, Imperial College London, London, SW7 2AZ
{wancheng.zhang07, p.naylor}@imperial.ac.uk

ABSTRACT

Reverberation can significantly reduce the intelligibility and degrade naturalness of speech signals in hands-free communications. One method to achieve dereverberation is to perform identification and inverse filtering of multichannel acoustic systems. In this approach, we will show that the 'energy' of the components of a multichannel inverse filtering system play a key role in its robustness to channel noise and system identification errors, where the 'energy' of an MA filter is defined as the sum of the squared coefficients of it. The energy of the components of inverse filtering systems obtained from multiple-input/output inverse theorem (MINT) is usually high, resulting in non-robust inverse filtering systems. In this paper, the characteristics of acoustic channels which give rise to high energy in the components of inverse filtering systems are investigated. Specifically, the features of channel zeros that cause the components of inverse filtering systems to be of high energy are shown.

1. INTRODUCTION

In hands-free communications, speech signals 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 multichannel acoustic systems. The methodology is illustrated in Fig. 1. Consider a clean speech signal

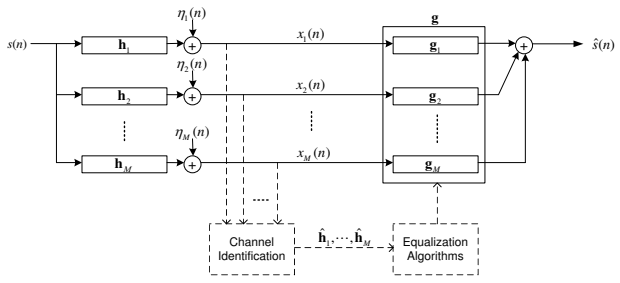

Figure 1: Illustration of identification and inverse filtering of acoustic systems.

$s(n)$ propagating through an M-channel acoustic system, the channels of 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 $\{\cdot\}^T$ denotes the transpose operation. Using the noisy reverberant speech signals

$$x_m(n) = s(n) * h_m(n) + \eta_m(n), \tag{1}$$

estimates of the room impulse responses (RIRs) $\mathbf{h}_m$ can be obtained with blind system identification techniques such as in [1], [2], [3] and [4], where $*$ denotes linear convolution, and $\eta_m(n)$ is the channel noise of the mth channel. Then, with the estimates $\hat{\mathbf{h}}_m = [\hat{h}_m(0)\ \hat{h}_m(1)\ \cdots\ \hat{h}_m(L-1)]^T$, 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 components $\mathbf{g}_m = [g_m(0)\ g_m(1)\ \ldots\ g_m(L_i-1)]^T$, can be designed with some chosen equalization algorithm. Then, by summing up the outputs of $\mathbf{g}_m$ with input $x_m(n)$, we expect a good estimate, $\hat{s}(n)$, of $s(n)$ can be obtained.

In practice, system identification usually induces some errors into the estimates due to the existence of channel noise. The estimate of the mth channel can be written as

$$\hat{\mathbf{h}}_m = \mathbf{h}_m + \mathbf{e}_m, \tag{2}$$

where $\mathbf{e}_m = [e_m(0)\ e_m(1)\ \cdots\ e_m(L-1)]^T$ is the system identification error vector of the mth channel.

Given the multichannel estimates $\hat{\mathbf{h}}_m$, where $M \geq 2$, an inverse filtering system $\mathbf{g}$ can be obtained using the multiple-input/output inverse theorem (MINT) [5] in the case that $\hat{\mathbf{h}}_m$ do not share any common zeros [5]. This means a $\mathbf{g}$ can be obtained so that

$$\sum_{m=1}^{M} \hat{h}_m(n) * g_m(n) = d(n-\tau), \tag{3}$$

where $d(n)$ is the delta function and τ is a delay.

Using MINT, the inverse filtering system $\mathbf{g}$ can be obtained by

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{d}, \tag{4}$$

where $\mathbf{H} = [\mathbf{H}_1\ \cdots\ \mathbf{H}_M]$ is defined as the system matrix formed by the convolution matrices $\mathbf{H}_m$, $\{\cdot\}^+$ denotes pseudo inverse, and

$$\mathbf{d} = [\underbrace{0\ \ldots\ 0}_{\tau} 1\ 0\ \ldots\ 0]^T \tag{5}$$

is an $(L+L_i-1)\times 1$ vector. $\mathbf{H}_m$ is an $(L+L_i-1)\times L_i$ convolution matrix of $\mathbf{h}_m$

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

The minimum length of the components $\mathbf{g}_m$ should be used to ensure such an inverse filtering system can be obtained is $L_i = L_c$, where $L_c = \lceil \frac{L-1}{M-1} \rceil$ is the critical length.

Using such an inverse filtering system $\mathbf{g}$ to filter the observations $x_m(n)$, an estimate $\hat{s}(n)$ can be obtained

$$\begin{aligned}
\hat{s}(n) &= \sum_{m=1}^{M} x_m(n) * g_m(n) \\
&= \sum_{m=1}^{M} (s(n) * h_m(n) + \eta_m(n)) * g_m(n) \\
&= \sum_{m=1}^{M} (s(n) * (\hat{h}_m(n) - e_m(n)) + \eta_m(n)) * g_m(n) \\
&= s(n-\tau) - \sum_{m=1}^{M} (s(n) * e_m(n) - \eta_m(n)) * g_m(n). \quad (6)
\end{aligned}$$

In (6), the first term is the desired (delayed) original speech signal, and the second term forms distortion to the speech signal.

It has been shown in [6], [7], [8] and [9] that the distortion caused by system identification errors or channel noise is usually very strong, making the estimated speech signal $\hat{s}(n)$ unsatisfactory. It can be seen that, in such situations, this inverse filtering system $\mathbf{g}$ obtained from MINT is not robust to channel noise and system identification errors. In [9], it is assumed that the RIRs are exactly known and the robustness of $\mathbf{g}$ to channel noise is considered. It is believed in [9] that reducing the energy of $\mathbf{g}_m$ is the key to make $\mathbf{g}$ more robust to channel noise. The energy of $\mathbf{g}_m$ is defined as

$$E_{\mathbf{g}_m} = \sum_{n=0}^{L_i-1} g_m^2(n), \quad (7)$$

and the energy of $\mathbf{g}$ is defined as

$$E_{\mathbf{g}} = \sum_{m=1}^{M} E_{\mathbf{g}_m}. \quad (8)$$

In fact, we can see in the second term of (6) that no matter if only channel noise exists, or both channel noise and system identification errors exist, $E_{\mathbf{g}_m}$ play a key role in the robustness of the inverse filtering system.

In this paper, we will investigate the characteristics of the RIRs that make the inverse filtering system obtained from MINT be of high energy. More specifically, the features of the zeros of the RIRs that cause the inverse filtering systems to be of high energy will be described. It has been shown in [9] that lengthening the components $\mathbf{g}_m$ and introducing delay can help to reduce the energy, so the effect of L_i and τ in reducing the high energy caused by the different features of the zeros will also be studied.

The remainder of this paper is arranged as follows. In Section 2, the features of zeros of acoustic channels will be summarized. In Section 3, the relationship between the features of the channel zeros and the energy of the inverse filtering system will be investigated. In Section 4, we will draw some conclusions.

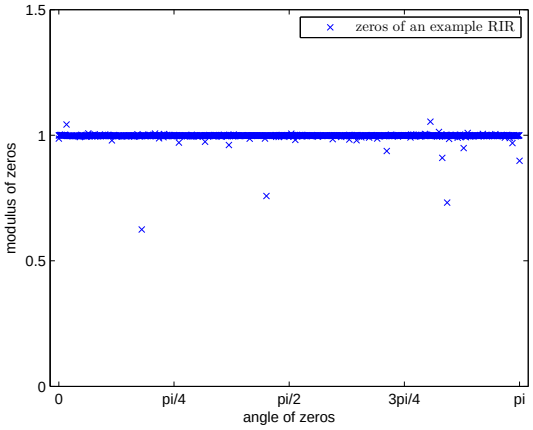

Figure 2: An example showing the distribution of the zeros of a typical RIR.

2. SUMMARY OF FEATURES OF CHANNEL ZEROS

In this Section, we summarize the features of zeros of the RIRs. For a typical RIR of thousands taps, the zeros of it have the following features:

1. The angles θ of the zeros are approximately uniformly distributed on $(-\pi, \pi]$ and the modulus r of most of the zeros is close to 1 [10]. Our studies have shown that the modulus r of most of the zeros lies in (0.995, 1.002).
2. The modulus of a few zeros is evidently greater than 1. The modulus of such zeros can be $r > 1.03$.
3. Among multiple channels, near-common zeros usually exist.

Here we use the definition of [11] that a cluster of near-common zeros is defined when M zeros from M different RIRs are located in the same vicinity in the z-plane, the vicinity being characterized by a small 'tolerance' δ.

In Figure 2 we give an example showing the distribution of the zeros of a RIR from the MARDY database [12]. The length of this RIR is $L = 2000$. In Fig. 2, the zeros, the angles of which are in $[0, \pi]$, are shown; complex conjugates of these zeros are not shown for explicitness. It can be seen in this figure that there are two zeros of modulus greater than 1.05.

In Fig. 3, we give an example analysis of number of clusters of near-common zeros against the tolerance δ between two MARDY channels. The clusters are identified with the clustering algorithms described in [11]. It can be seen that since most of the zeros are close to the unit circle, near-common zeros that are within the vicinity of $\delta > 10^{-5}$ commonly exist.

In the next Section, the effect of the inter-channel commonality δ and the modulus r on the energy of the inverse filtering systems will be studied.

3. INTER-CHANNEL COMMONALITY OF ZEROS AND ENERGY OF INVERSE FILTERING SYSTEMS

In this Section, the relationship between the inter-channel commonality of the zeros of RIRs and the energy

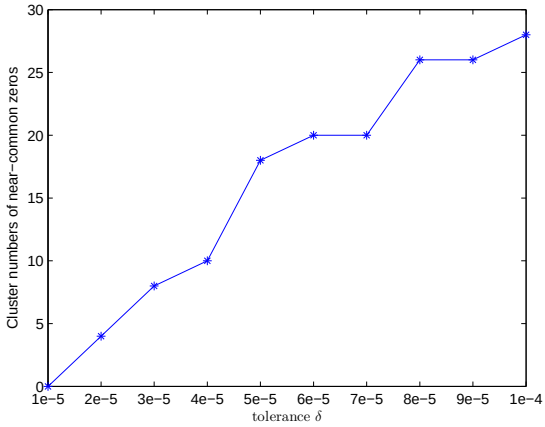

Figure 3: An example showing the number of clusters of near-common zeros against the tolerance δ.

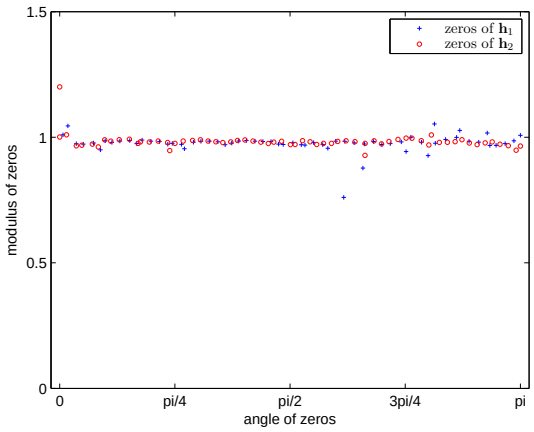

Figure 4: Zeros of the RIRs used to obtain the experimental results of Fig. 5.

of the inverse filtering systems will be investigated experimentally. The RIRs used in the experiments are all from MARDY database and are truncated to $L = 128$. For the sake of comparison of the energy of the inverse filtering systems in different cases, all the RIRs used are normalized to be of unit energy.

3.1 Clusters of near-common zeros close to the unit circle

The zeros of the two RIRs $\mathbf{h}_1$ and $\mathbf{h}_2$ used in this experiment are shown in Fig. 4. Since these two RIRs are truncated to $L = 128$, there is no cluster of zeros within vicinity of δ of order 10^{-5} between these two channels. The closest pair of zeros between these two channels is of $\delta = 3.54 \times 10^{-3}$ (Case 1). To study the effect of the near-common zeros, we experiment by moving a zero of $\mathbf{h}_2$ at the angle of $\theta = 0.6812\pi$ towards a zero of $\mathbf{h}_1$ at $\theta = 0.6823\pi$ and make these two zeros within the vicinity of $\delta = 3 \times 10^{-5}$ (Case 2). The modulus of this pair of near-common zeros is $r = 0.983$. The frequency responses of filters $\mathbf{g}_1$ and $\mathbf{g}_2$ obtained from (4) with

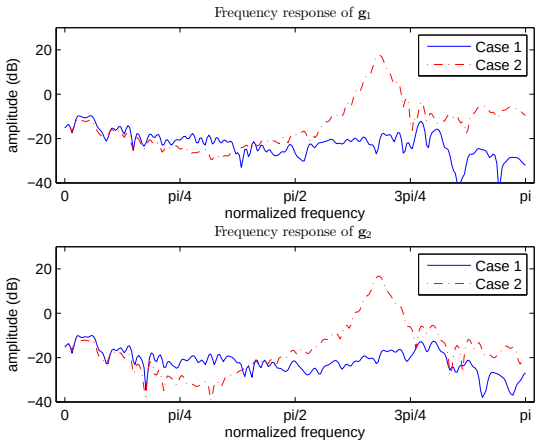

Figure 5: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ for Case 1 and Case 2.

$L_i = L_c$ and $\tau = 0$ are shown in Fig. 5. It can be seen in Fig. 5 that, compared with Case 1, the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ have a strong peak at the normalized frequencies around 0.68π in Case 2. In addition, it is found that $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ in Case 2 are much higher than that obtained with RIRs of Case 1. In Case 1, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ are 6.60 and 6.70 respectively; with near-common zeros of $\delta = 3 \times 10^{-5}$ in Case 2, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ are as high as 814.53 and 698.54.

We also studied the effect of increasing L_i and introducing τ in reducing the high energy caused by near-common zeros. The delay used is the best delay, i.e., the delay with which (4) gives a $\mathbf{g}$ of the lowest $E_{\mathbf{g}}$. All the experiment data obtained in this paper is listed in Table 1. We can see in Case 2 that with delay, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ have not been reduced, but with increasing the length to be $L_i = 3L_c$, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ have been greatly reduced to 5.34 and 5.81.

3.2 Zeros of large modulus

Now we study the effect of zeros of $r > 1.03$. The zeros of the RIRs are shown in Fig. 6. The closest pair of zeros between these two channels is of $\delta = 3.27 \times 10^{-3}$. It can be seen from Fig. 6 that at angles around 0.02π and 0.81π, both of the two channels have zeros of $r > 1.03$ (Case 3). The two zeros at angle 0.02π are in the vicinity of $\delta = 6.86 \times 10^{-3}$; the modulus of them is $r = 1.045$ and $r = 1.048$. The zeros at angle 0.81π are of $\delta = 2.14 \times 10^{-2}$, the modulus of them is $r = 1.053$ and $r = 1.032$. Compared with the near-common zeros of $\delta = 3 \times 10^{-5}$ in Case 2, the tolerance δ of these two clusters is much larger. However, these clusters of zeros also cause $\mathbf{g}$ to be of high energy. With $L_i = L_c$ and $\tau = 0$, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ are 899.40 and 1197.85. Experimenting with $L_i = 3L_c$, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ are still as high as 477.86 and 839.26. It can be seen therefore that increasing L_i does not help significantly with respect to the zeros of large modulus. On the other hand, controlling the delay has very significant value. With $L_i = 3L_c$ and best delay, $E_{\mathbf{g}_1}$ and $E_{\mathbf{g}_2}$ are reduced to 2.16 and 2.79, which are substantially reduced.

The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ with $L_i = 3L_c$

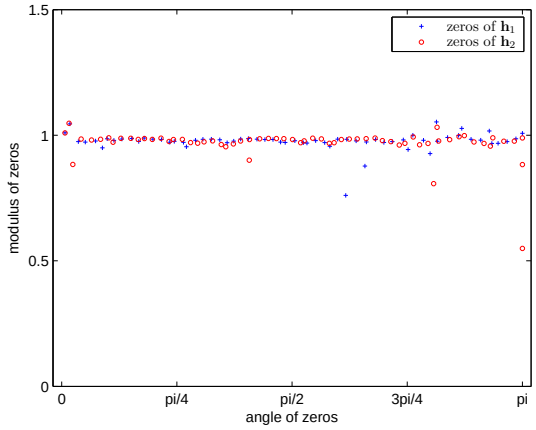

Figure 6: Zeros of the RIRs used to obtain the experimental results of Fig. 7.

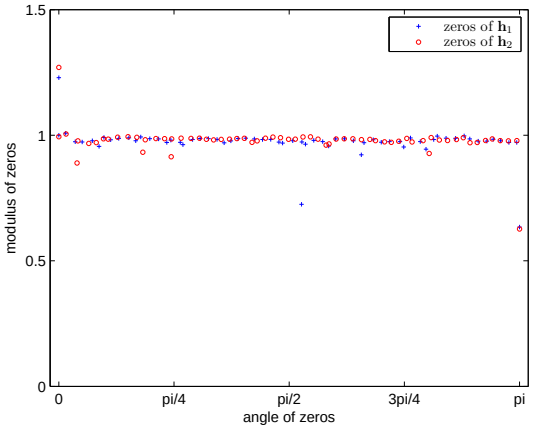

Figure 8: Zeros of the RIRs used to obtain the experimental results of Fig. 9.

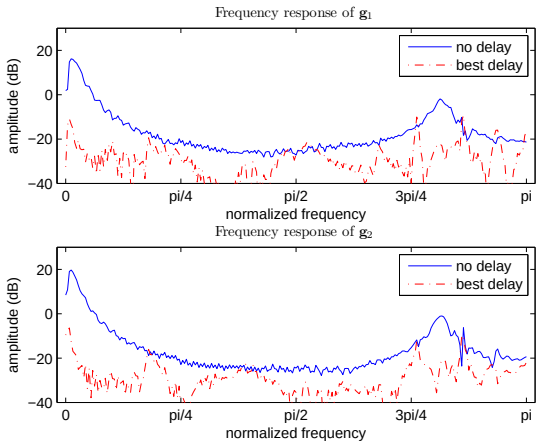

Figure 7: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ for Case 3.

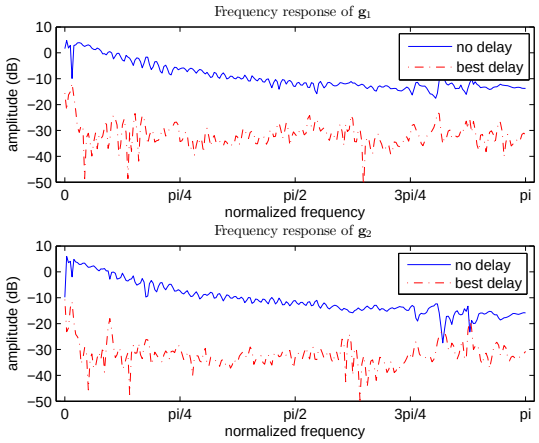

Figure 9: The frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ for Case 4.

are shown in Fig. 7. It can be seen in Fig. 7 that there are two peaks. One peak is at low frequencies corresponding to the cluster of zeros at 0.02π and the other is at frequencies around 0.81π corresponding to the cluster at 0.81π. Controlling the delay is very helpful in suppressing these peaks.

Compared with $\delta = 3 \times 10^{-5}$ and $r = 0.983$ in Case 2, we can see that with relatively larger $\delta = 6.86 \times 10^{-3}$, cluster of zeros of large modulus $r = 1.045$ can also cause strong peaks in the frequency responses of $\mathbf{g}_m$.

In the next experiment, another example will be given showing the effect of clusters of zeros of large modulus. The frequency responses of $\mathbf{g}_m$ corresponding to the RIRs of the zeros shown in Fig. 8 are shown in Fig. 9. The length of $\mathbf{g}_m$ used to obtain the experimental results of Fig. 9 is $L_i = 3L_c$. It can be seen in Fig. 8 that at the angle of 0, $\mathbf{h}_1$ has a zero of $r = 1.23$ and $\mathbf{h}_2$ has a zero of $r = 1.27$ (Case 4). The modulus r of these two zeros is much larger than that of the clusters in Case 3. The tolerance of these two zeros is $\delta = 4 \times 10^{-2}$. However, in the frequency responses in Fig. 9, we can see that the peak at low frequencies corresponding to this cluster of zeros is weaker than the peak at low frequencies in Case 3. This shows that, though the modulus of these zeros is much larger than that in Case 3, since δ is larger in this example, the resulting $E_{\mathbf{g}_m}$ is smaller. This indicates that in the cases of zeros of large modulus, compared with r, the inter-channel commonality δ is more important in influencing the energy of inverse filtering systems. This observation also agrees with the fact that the peak around 0.81π in the frequency responses of $\mathbf{g}_1$ and $\mathbf{g}_2$ is weaker than the peak at low frequencies in Case 3.

We can see from Table 1 that although increasing the length L_i and introducing delay τ can help to reduce the energy caused by these special zeros, they cannot eliminate the effect caused by them.

4. CONCLUSIONS

In this paper, the relationship between the inter-channel commonality of zeros of multichannel acoustic systems and the robustness of their inverse filtering systems to channel noise and system identification errors is investi-

	L_i	$E_{\mathbf{g}_1}$		$E_{\mathbf{g}_2}$	
		$\tau = 0$	best τ	$\tau = 0$	best τ
Case 1	L_c	6.60	3.10	6.70	2.04
	$3L_c$	5.20	0.56	5.71	0.53
Case 2	L_c	814.53	814.53	698.54	698.54
	$3L_c$	5.34	1.32	5.81	1.42
Case 3	L_c	899.40	39.06	1197.85	51.12
	$3L_c$	477.86	2.16	839.26	2.79
Case 4	L_c	216.38	3.84	201.36	5.24
	$3L_c$	147.62	0.77	142.09	0.85

Table 1: Energy of the components of the inverse filtering systems.

gated. Clusters of near-common zeros close to the unit circle cause strong peaks in the frequency responses of the components of the inverse filtering systems and make the components be of high energy. The normalized frequencies of these peaks correspond to the angles of these clusters. Clusters of zeros of large modulus also cause strong peaks in the frequency responses of the components of the inverse filtering systems at corresponding frequencies. To cause peaks of equivalent amplitude, the inter-channel commonality of the zeros in the clusters of larger modulus can be lower than the inter-channel commonality of the clusters of smaller modulus.

REFERENCES

[1] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, 1998.

[2] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, pp. 1127–1138, 2002.

[3] ——, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Processing*, vol. 51, pp. 11–24, 2003.

[4] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

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

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

[7] N. D. Gaubitch and P. A. Naylor, "Equalization of multichannel acoustic systems in oversampled subbands," *to appear in IEEE Trans. Audio, Speech, Language Processing*, 2009.

[8] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2006.

[9] ——, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[10] X. Lin, N. D. Gaubitch, and P. A. Naylor, "Two-stage blind identification of SIMO systems with common zeros," in *Proc. European Signal Processing Conf.*, 2006.

[11] A. W. H. Khong, X. Lin, and P. A. Naylor, "Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2008.

[12] 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. Int. Workshop Acoust. Echo Noise Control*, 2006.