NEAR-COMMON ZEROS IN BLIND IDENTIFICATION OF SIMO ACOUSTIC SYSTEMS

Patrick A. Naylor, Xiang (Shawn) Lin and Andy W. H. Khong

Department of Electrical and Electronic Engineering, Imperial College London, UK
E-mail: {p.naylor, shawn.lin04, andy.khong}@imperial.ac.uk

ABSTRACT

The common zeros problem for Blind System Identification (BSI) has been well known to degrade the performance of classic BSI algorithms and therefore limits performance of subsequent speech dereverberation. Recently, we have shown that multichannel systems cannot be well identified if near-common zeros are present. In this work, we further study the near-common zeros problem using channel diversity measure. We then investigate the use of forced spectral diversity (FSD) based on a combination of spectral shaping filters and effective channel undermodelling. Simulation results show the effectiveness of the proposed approach.

Index Terms— blind system identification, near-common zeros, channel identifiability condition, forced spectral diversity

1. INTRODUCTION

Blind System Identification (BSI) has attracted significant interest due to its potential applications in several areas. Since the pioneering work by Tong *et. al.* [1], many second-order statics (SOS)-based methods have been proposed, among which are the subspace [2] and the cross-relation [3] algorithms. Recently, these methods have been adopted from communications into the domain of acoustic signal processing such as for acoustic dereverberation and speech source separation. Both closed-form [4] and adaptive [5] algorithms have been proposed to identify room impulse responses of length up to hundreds of taps for acoustic dereverberation.

Most SOS-based BSI methods, such as [2][3][4][5], rely on the channel identifiability conditions [3] from which one of the conditions is that the channels must be co-prime, i.e., no zeros being common across all channels. This is because when common zeros exist, the BSI algorithms do not have sufficient information to distinguish the common zeros of the channels from ones due to the source signal. Unlike algorithms for communications applications where various pre-processing techniques, such as linear precoding, can be applied to induce specific statistical properties, algorithms for acoustic signal processing in speech dereverberation do not share the similar flexibility. A common way to mitigate the common zeros problem is therefore to increase the number of microphones, which is computationally expensive and practically limited. More recently, the problem of near-common zeros (NCZs) has been addressed in [6][7] where results presented showed that the performance of SOS-based BSI algorithms can degrade as zeros of different channels become close to each other.

In this paper, we first review blind single-input multiple-output (SIMO) system identification in Section 2, where the problem of NCZs is discussed and illustrated. We then show experimentally, in Section 3.1, the relationship between the channel characteristics and the performance of the BSI algorithms such as the normalized multichannel fast least-mean-square (NMCFLMS) algorithm [5]. In

Section 3.2, we introduce the forced spectral diversity (FSD) concept for addressing the detrimental effects of near-common zeros on adaptive BSI algorithms. The FSD concept employs a combination of spectral shaping filters and effective channel undermodelling techniques which in turn improve the performance of the NMCFLMS algorithm. The performance of the proposed approach is illustrated in Section 4.

2. THE NEAR-COMMON ZEROS PROBLEM IN BLIND SIMO SYSTEM IDENTIFICATION

2.1. Review of the NMCFLMS algorithm

Consider a typical acoustic environment with one talker and multiple microphones. This system can be modeled as a linear SIMO system where the relationship between the source signal $s(n)$ and mth output $x_m(n)$ is given by

$$\mathbf{x}_m(n) = \mathbf{H}_m(n)\mathbf{s}(n) + \mathbf{b}_m(n), \qquad m = 1, 2, \ldots, M \quad (1)$$

where M is the number of channels, $\mathbf{x}_m(n) = [x_m(n)\ x_m(n-1)\ \ldots\ x_m(n-L+1)]^T$ and $\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-2L+2)]^T$. The vector $\mathbf{b}_m(n) = [b_m(n)\ b_m(n-1)\ \ldots\ b_m(n-L+1)]^T$ denotes the additive noise and $\mathbf{h}_m(n) = [h_{m,0}(n)\ h_{m,1}(n)\ \ldots\ h_{m,L-1}(n)]^T$ is the vector of channel impulse response of length L, from which the $L \times (2L-1)$ channel matrix is given as

$$\mathbf{H}_m(n) = \begin{bmatrix} h_{m,0}(n) & \cdots & h_{m,L-1}(n) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & h_{m,0}(n) & \cdots & h_{m,L-1}(n) \end{bmatrix} \quad (2)$$

with $[\cdot]^T$ being the transpose operator. A system equation can then be derived by concatenating all M outputs of (1) as follows:

$$\mathbf{x}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{b}(n), \quad (3)$$

where $\mathbf{x}(n) = [\mathbf{x}_1^T(n)\ \mathbf{x}_2^T(n)\ \ldots\ \mathbf{x}_M^T(n)]^T$, $\mathbf{b}(n) = [\mathbf{b}_1^T(n)\ \mathbf{b}_2^T(n)\ \ldots\ \mathbf{b}_M^T(n)]^T$ and $\mathbf{H}(n) = [\mathbf{H}_1^T(n)\ \mathbf{H}_2^T(n)\ \ldots\ \mathbf{H}_M^T(n)]^T$ is the $ML \times (2L-1)$ global channel matrix.

As we are concentrating on the common zeros problem, henceforth we consider only the noise-free case in this paper. In the absence of noise, a multichannel system can be identified using the cross-relation between the ith and jth channel outputs [5], for $i \neq j$, given by $\mathbf{x}_i^T(n)\mathbf{h}_j(n) = \mathbf{x}_j^T(n)\mathbf{h}_i(n)$ for $i, j = 1, 2, \ldots, M$. An error function can then be defined, for $i \neq j$,

$$e_{ij}(n) = \mathbf{x}_i^T(n)\hat{\mathbf{h}}_j(n-1) - \mathbf{x}_j^T(n)\hat{\mathbf{h}}_i(n-1), \quad (4)$$
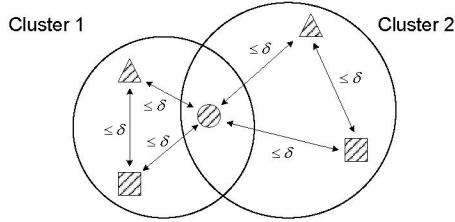
21

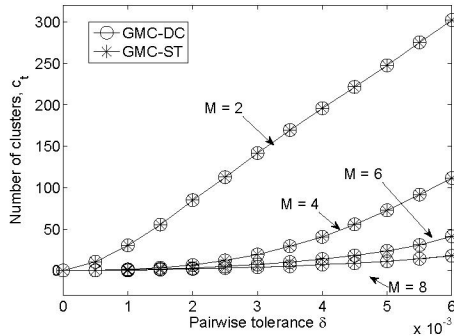**Fig. 1**. An illustrative example of two NCZ clusters for a three-channel SIMO system.

Fig. 2. Number of clusters, c_t, found using the GMC-DC and GMC-ST [7] algorithms against tolerance δ with different number of channels M using simulated impulse responses.

where $\hat{\mathbf{h}}_i(n)$ is the estimated impulse response for the ith channel. Using (4), the NMCFLMS algorithm [5] is derived by minimizing the cost function

$$J(n) = \frac{1}{\| \hat{\mathbf{h}}(n) \|_2^2} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} e_{ij}^2(n), \qquad (5)$$

where $\hat{\mathbf{h}}(n) = [\hat{\mathbf{h}}_1^T(n) \ \hat{\mathbf{h}}_2^T(n) \ \ldots \ \hat{\mathbf{h}}_M^T(n)]^T$ is the channel vector for estimated impulse responses and $\| \cdot \|_2$ denotes l_2-norm. Our objective is to estimate $\hat{\mathbf{h}}_m(n)$ by employing only observations $\mathbf{x}_m(n)$ for $m = 1, 2, \ldots, M$. Note that in this work the channel length L is assumed to be available.

2.2. Effect of near-common zeros on the NMCFLMS algorithm

For an M-channel SIMO system, the channels are said to be co-prime if they do not share the same zeros. However, NCZs are clusters of zeros that satisfy the following conditions [7]: (i) each cluster contains only M zeros with each channel contributing to a zero, and (ii) the Euclidean distance between any pair of zeros in a cluster lies within a pairwise tolerance δ where $\delta \geq 0$. We note that given these two conditions, any zero can be included in more than one cluster. Figure 1 shows an example of two clusters for a three-channel system where the zeros from each channel are represented by triangles, squares and circles.

To illustrate the number of NCZ clusters in SIMO acoustic systems, a set of impulse responses was simulated using the method of images [8] with a linear array of $M = 8$ microphones in a room of dimension $10 \times 10 \times 3$ m. The source is located 1 m in front of the microphone array with uniform spacings of 8 cm. The sampling rate was 16 kHz with each channel impulse response having 512 coefficients and White Gaussian Noise (WGN) is used as source signals. A
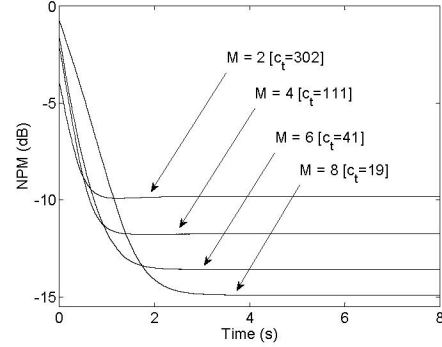

Fig. 3. Performance variation of NMCFLMS [5] with different number of channels using WGN input and simulated impulse responses.

set of 100 SIMO systems are generated by placing this source-sensor configuration in different positions of the room. Finally, results are obtained by averaging across all systems.

Two generalized multichannel clustering (GMC) algorithms, the GMC-DC and GMC-ST proposed in [7] were employed to compute the number of NCZs over the generated impulse responses. Figure 2 shows the average number of NCZs clusters, denoted as c_t, against the tolerance δ with different number of channels M. To show the corresponding performance variation of the BSI algorithm against NCZs, we then employed the NMCFLMS algorithm [5] with a step-size of 0.5. The signal-to-noise ratio (SNR) is set 60 dB to avoid the misconvergence problem [9]. We use the normalized projection misalignment (NPM) [10] to quantify the BSI estimation error. In Fig. 3, performance variation of the NMCFLMS algorithm with different number of channels is shown, where c_t is found for $\delta = 6 \times 10^{-3}$. As can be seen from Fig. 2 and Fig. 3, the number of clusters decreases when M increases, which results in better NPM performance. We also note from Fig. 2 that for all sample systems, there exist no exactly common zeros, i.e., no NCZ clusters were found for $\delta = 0$.

3. CHANNEL AND SPECTRAL DIVERSITY WITH EFFECTIVE UNDERMODELLING

It has been shown in Section 2 that the presence of NCZs results in the performance degradation of the NMCFLMS algorithm. In addition, increasing spatial diversity using more microphones is computationally expensive. In this Section, we show that spectral diversity can be utilized to reduce the number of NCZ clusters. We propose to obtain such extra diversity for a SIMO system with NCZs via the use of spectral shaping filters and effective channel undermodelling, to which we refer as the FSD concept.

3.1. Channel diversity for SIMO systems

Consider a M-channel SIMO system defined in Section 2, if the multiple channels do not share exactly common zeros, the global channel matrix (3) is of full rank [2], i.e., $\mathrm{Rank}(\mathbf{H}(n)) = 2L - 1$. This indicates that the smallest singular value of $\mathbf{H}(n)$, denoted as $\sigma_{2L-1}(\mathbf{H}(n))$, is non-zero. If exactly common zeros exist, $\mathbf{H}(n)$ becomes rank deficient and leads to $\sigma_{2L-1}(\mathbf{H}(n)) = 0$. According to [11], the minimum non-zero singular value of the global channel matrix, i.e., $\lambda_{\min} = \sigma_{2L-1}(\mathbf{H}(n))$ can be used as a measure of *channel diversity* of $\mathbf{H}(n)$.
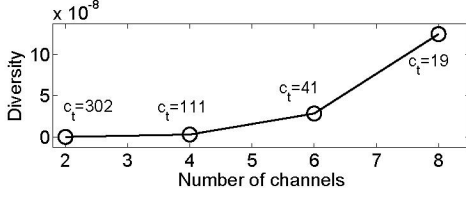
22

Fig. 4. Variation of $\lambda_{\min}$ against the number of channels M for simulated impulse responses used in Fig. 3.

Fig. 5. Schematic for a two-channel FSD SIMO system.

Fig. 6. Frequency responses of the spectral shaping filters.

To show that such measurement can also be used for systems with NCZs, Fig. 4 depicts the variation of $\lambda_{\min}$ against different number of channels over impulse responses used in generating Fig. 3. The number of NCZs found using the GMC-ST algorithm [7] for each data point is also shown for $\delta = 6 \times 10^{-3}$. As can be seen, the channel diversity increases with M. This indicates that as channel diversity increases, the NPM performance of the NMCFLMS algorithm can improve correspondingly since the number of NCZs reduces with M. It is also noted from the figure that for simulated acoustic SIMO systems, the value of $\lambda_{\min}$ is small, which can be expected to result in a large condition number for $\mathbf{H}(n)$ and to further degrade the performance of channel inversion algorithms, such as the MINT algorithm [12], over acoustic SIMO systems.

3.2. Spectral diversity with effective undermodelling

Channel undermodelling was introduced for blind identification of microwave radio impulse responses with small leading and tailing taps in [11], where it was shown that the m-order least-squares (LS) method [3] or subspace method [2] can estimate impulse responses that are close to the m-order "significant" part of the full impulse responses which is obtained by removing small leading and tailing taps. This can only be achieved when the undermodelled system offers sufficient diversity. The attempt of modelling full-length impulse responses, however, can result in much worse estimation performance.

Inspired by results in [11], our proposed FSD concept includes the use of spectral shaping filters and effective channel undermodelling. For clarity of presentation, FSD in this paper is based on a two-channel system case and a schematic for a two-channel SIMO system is shown in Fig. 5. As can be seen, the microphone signal $\mathbf{x}_m(n)$ is pre-processed using a pair of spectral shaping filters, e.g., a pair of high- and lowpass filters (LPF and HPF), the resulting filter outputs $\mathbf{x'}_m(n)$ are then given as

$$\mathbf{x}'_m(n) = \boldsymbol{\mathcal{F}}_m^T \mathbf{H}_m(n)\mathbf{s}(n) \qquad (6)$$

for $m = 1, 2$, where $\boldsymbol{\mathcal{F}}_m$ denotes the mth spectral shaping filter matrix of dimension $L \times (L + L_p - 1)$,

$$\boldsymbol{\mathcal{F}}_m = \begin{bmatrix} f_{m,0} & \cdots & f_{m,L_p-1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & f_{m,0} & \cdots & f_{m,L_p-1} \end{bmatrix}, \qquad (7)$$

and $\mathbf{f}_m = [f_{m,0}, f_{m,1}, \ldots, f_{m,L_p-1}]^T$ is the impulse response of the mth filter of length L_p. According to Fig. 5, $\mathbf{x}'_m(n)$ can be considered as the linear convolution between $\mathbf{s}(n)$ and an equivalent SIMO system of length $L + L_p - 1$, i.e.,

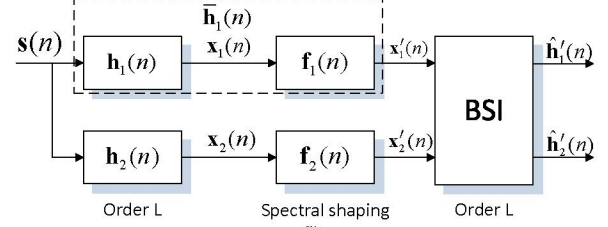$$\bar{\mathbf{h}}_m(n) = \boldsymbol{\mathcal{F}}_m^T \mathbf{h}_m(n), \qquad m = 1, 2. \qquad (8)$$
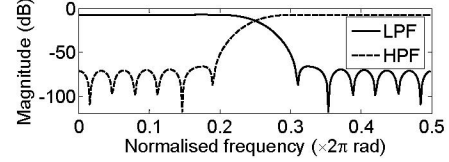
However, the channel length employed within the NMCFLMS algorithm is fixed at L and therefore it is equivalently "blind" to such filtering process. This results in an effective channel undermodelling, i.e., the channel estimates produced by the NMCFLMS algorithm, denoted as $\hat{\mathbf{h}}'_m(n)$, actually correspond to an undermodelled SIMO system of length L.

For acoustic SIMO systems, the room impulse responses have small tailing taps due to the reverberation tail that still exist after being filtered by spectral shaping filters. This characteristic is similar to the microwave radio impulse responses shown in [11]. Therefore, we propose to effectively undermodel $\bar{\mathbf{h}}_m(n)$ by removing the last $L_p - 1$ taps of $\bar{\mathbf{h}}_m(n)$, i.e.,

$$\mathbf{h}'_m(n) = \mathbf{U}\bar{\mathbf{h}}_m(n), \qquad m = 1, 2 \qquad (9)$$

where $\mathbf{U} = [\mathbf{I}_{L \times L}, \ \mathbf{0}_{L \times (L_p-1)}]$ with $\mathbf{I}_{L \times L}$ and $\mathbf{0}_{L \times (L_p-1)}$ being a $L \times L$ identity matrix and a $L \times (L_p - 1)$ null matrix. As will be shown in Section 4, the combination of spectral shaping filters and the effective channel undermodelling can result in increased channel diversity and reduced the number of NCZ clusters.

Finally, we note that the original system can be recovered in principle since the characteristics of the spectral shaping filters are known and could be inverted. In practice, effects of noise amplification in the inversion process will limit accuracy. However, an estimate of $\hat{\mathbf{s}}(n)$ can be obtained without the need to "undo" the effect of the spectral shaping by inverting $\hat{\mathbf{h}}'_m(n)$ using, for example, the MINT algorithm [12]. Similar practical limitations apply but we have found in our tests that such limitations are more than outweighed by the advantages obtained in the BSI due to the increased channel diversity, resulting in an overall improvement in our ability to equalize channels containing NCZs.

4. SIMULATIONS

In this section, we further demonstrate the effect of FSD using the set of two-channel SIMO systems obtained from the impulse response library that has been used throughout this paper. A pair of high- and lowpass shaping filters of length $L_p = 32$ was generated with their magnitude responses shown in Fig. 6.

To demonstrate the effect of FSD on SIMO systems with NCZs, we first measure the channel diversity over the set of two-channel
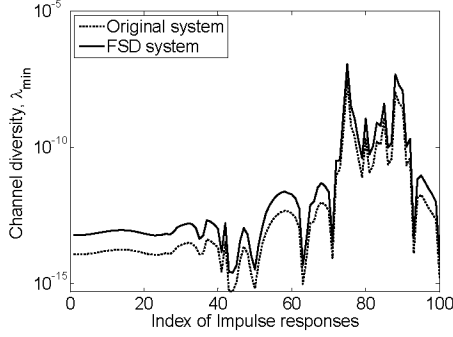
23

Fig. 7. Variation of channel diversity $\lambda_{\min}$ for the original system $\mathbf{h}_m(n)$ and the FSD system $\mathbf{h}'_m(n)$ for different simulated impulse responses.
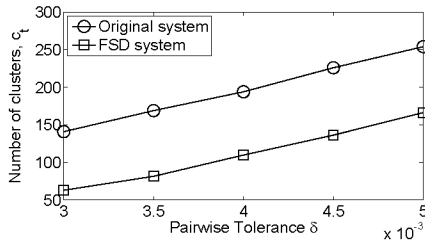

Fig. 8. Comparison of c_t between the original system $\mathbf{h}_m(n)$ and the FSD system $\mathbf{h}'_m(n)$ against various tolerance δ.

SIMO systems chosen from a set of impulse responses described in Section 2.2, where $\mathbf{h}_m(n)$ and $\mathbf{h}'_m(n)$ are obtained using (9) in a similar way as described in [13]. In Fig. 7, the variation of $\lambda_{\min}$ for the original system $\mathbf{h}_m(n)$ and the FSD system $\mathbf{h}'_m(n)$ is plotted against different impulse responses. The FSD system is seen to provides a consistent improvement in terms of channel diversity over the original system, which implies that $\mathbf{h}'_m(n)$ can be better identified. Using the same set of impulse responses, we then compare Fig. 8 the number of NCZ clusters between $\mathbf{h}_m(n)$ and $\mathbf{h}'_m(n)$ against pairwise tolerance δ. As can be seen, the FSD system $\mathbf{h}'_m(n)$ only contains approximately half of the number of NCZs compared to the original system $\mathbf{h}_m(n)$.

Finally, we show in Fig. 9 that the FSD system is more identifiable than the original system in terms of NPM performance. The parameters used are the same as in Section 2 and previous simulations. An approximate 3 dB improvement can be clearly seen which is close to the NPM performance for the case of $M = 6$ in Fig. 3.

5. CONCLUSIONS

We have addressed and illustrated the near-common zeros problem for blind system identification. We showed the link between the channel diversity and the NPM performance of the NMCFLMS algorithm. The FSD concept for blind identification of acoustic SIMO system in the presence of NCZs has been proposed, which combines the use of spectral shaping filters and the effective channel under-modelling method. Simulation results based on two-channel case confirmed the effectiveness of the proposed concept.

6. REFERENCES

[1] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second order statistics: a time domain approach," *IEEE Trans.*
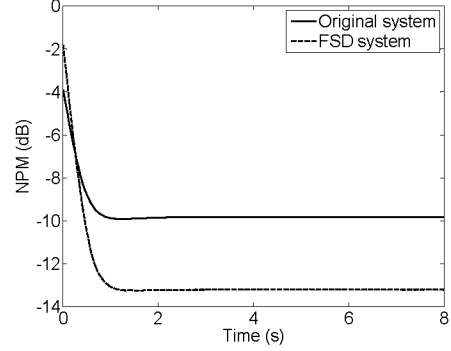
Fig. 9. Performance comparison for the NMCFLMS algorithm between the original system $\mathbf{h}_m(n)$ and the FSD system $\mathbf{h}'_m(n)$.

Inform. Theory, vol. 40, no. 2, pp. 340–349, March 1994.

[2] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 516–525, Feb. 1995.

[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[4] S. Gannot and M. Moonen, "Subspace methods for multi-microphone speech dereverberation," *EURASIP J. Applied Signal Process.*, vol. 2003, no. 11, pp. 1074–1090, Oct. 2003.

[5] 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.

[6] X. S. Lin, N. D. Gaubitch, and P. A. Naylor, "Two-stage blind identification of SIMO systems with common zeros," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2006.

[7] A. W. Khong, X. S. 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 Process. (ICASSP)*, Apr. 2008.

[8] 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, April 1979.

[9] R. Ahmad, A. Khong, and P. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2006.

[10] D. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Letters*, vol. 5, no. 7, pp. 174–176, July 1998.

[11] A. P. Liavas, P. A. Regalia, and J.-P. Delmas, "Robustness of least-squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1636–1645, June 1999.

[12] 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.

[13] M. K. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, "Improving robustness of blind adaptive multichannel identification algorithms using constraints," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sep. 2005.

24