Supervised Identification and Removal of Common Filter Components in Adaptive Blind SIMO System Identification

Mark R. P. Thomas, Nikolay D. Gaubitch, Emanuël A. P. Habets and Patrick A. Naylor
Communications and Signal Processing Research Group
Imperial College London
London SW7 2AZ
Email: {mrt102, ndg, e.habets, p.naylor}@imperial.ac.uk

***Abstract*—Adaptive blind system identification with LMS-type algorithms is prone to misconvergence in the presence of noise. In this paper we consider the hypothesis that such misconvergence is due to the introduction of a common filter to the estimated impulse respones. A technique is presented for identifying and removing the common filter using prior knowledge of the true channels. Experimental results with this approach show an improved rate of convergence and reduced system error. Furthermore, misconvergent behaviour is no longer observed, offering a plausible explanation as to the source of misconvergence in adaptive blind system identification.**

I. Introduction

Blind system identification (BSI) is a common problem encountered with the analysis of signals captured in a reverberant environment. A number of multichannel least-squares and subspace methods have been proposed that are able to identify channels from multichannel observations providing the channel order is known and identifiability conditions are satisfied [1], [2]. With these constraints satisfied, perfect identification and equalization of the observed signals is possible in the absence of noise. However, the robustness of adaptive algorithms in the presence of additive noise is a problem that has received much attention in the literature. It is known that with additive noise such adaptive algorithms converge towards the correct solution before catastrophic misconvergence [3]. To overcome this, various methods employ constraints based upon *a priori* knowledge of the channel to improve noise robustness [2], [4].

Many studies have been conducted into the behaviour of least mean square (LMS)-type BSI algorithms in noise. In this paper we present a preliminary study into the effect of uncorrelated additive noise and show that it causes a common filter to be present in all estimated channels. Such a common component has been shown to be the result of overmodelling a system [5], though experimental evidence in this paper shows that a common component may occur even where the channel order is modelled exactly. It is further shown that the estimation and removal of a common component using knowledge of the true impulse responses increases the rate of convergence, with an error related to the noise floor, and prevents misconvergence of the identified system. This result suggests that a misconvergence can be viewed as the addition of a common filter to the estimated channels and not convergence to an entirely incorrect solution.

The remainder of this paper is organized as follows. In Section II the problem is formulated formally. In Section III, the NMCFLMS algorithm is derived. Common filtering in the estimated channels is described in Section IV. Experimental results are presented in Section V. Conclusions are given in Section VI.

II. Problem Formulation

Consider a speech signal recorded in a noisy environment with an array of microphones. The observed signals at channel $i \in \{1, 2, \dots, M\}$ are given by

$$\mathbf{x}_i(n) = \mathbf{y}_i(n) + \boldsymbol{\nu}_i(n) \tag{1}$$

$$\mathbf{y}_i(n) = \mathbf{H}_i \mathbf{s}(n), \tag{2}$$

where $\mathbf{s}_i(n) = [s(n)\ s(n-1)\ \dots\ s(n-2L+1)]^T$, $\mathbf{x}_i(n) = [x_i(n)\ x_i(n-1)\ \dots\ x_i(n-L+1)]^T$, $\mathbf{y}_i(n) = [y_i(n)\ y_i(n-1) \dots\ y_i(n-L+1)]^T$, $\boldsymbol{\nu}_i(n) = [\nu_i(n)\ \nu_i(n-1)\ \dots\ \nu_i(n-L+1)]^T$ are segments of the speech signal, noisy observation, clean observation, noise and starting at sample n respectively and $\mathbf{H}_i$ denotes the filtering matrix. The length of each segment is L samples. The filtering matrix is defined by

$$\mathbf{H}_i = \begin{bmatrix} h_{i,0} & \dots & h_{i,L-1} & \dots & \dots & 0 \\ 0 & h_{i,0} & \dots & h_{i,L-1} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & h_{i,0} & \dots & h_{i,L-1} \end{bmatrix}. \tag{3}$$

The noise signals are assumed to be uncorrelated and white such that $E\{\boldsymbol{\nu}_i(n)\boldsymbol{\nu}_j(n)\} = 0\ \forall\ i \neq j$ and $E\{\boldsymbol{\nu}_i(n)\boldsymbol{\nu}_i(n-n')\} = 0\ \forall\ i, n' \neq 0$. The source and noise are also uncorrelated as $E\{\boldsymbol{\nu}_i(n)\mathbf{s}(n)\} = 0\ \forall\ i$.

Blind system identification is the process of estimating the filters $\mathbf{h}_i$ (the first row of the filtering matrix) from the observations $\mathbf{x}_i(n)$ alone. The identifiability conditions [1] state that the autocorrelation matrix of $\mathbf{s}(n)$ be full rank and that there are no common zeros shared between channels;

in the noiseless case the filters $\mathbf{h}_i(n)$ can then be identified exactly. In the presence of noise it is known that adaptive LMS-type algorithms converge towards the correct solution before misconvergence [2]. The nature of this misconvergence and its relationship to common filtering between the identified channels is investigated in this paper.

III. Review of the NMCFLMS Algorithm

The normalized multichannel frequency-domain LMS (NMCFLMS) algorithm [6] is a computationally-efficient adaptive BSI technique which is briefly summarized in this section. From (2) the following relationship can be deduced [1] in the absence of noise

$$x_i * h_j = s * h_i * h_j = x_j * h_i, \tag{4}$$

where $*$ denotes linear convolution, from which an error function can be derived

$$e_{ij}(n) = \mathbf{x}_i^T(n)\hat{\mathbf{h}}_j - \mathbf{x}_j^T(n)\hat{\mathbf{h}}_i, \; i,j = 1,2,\dots,M, \; i \neq j. \tag{5}$$

In order to avoid trivial estimates of all zero elements, a unit-norm constraint is imposed on $\mathbf{h}$ and the normalized error signal becomes

$$\epsilon_{ij}(n) = \frac{e_{ij}(n)}{\|\hat{\mathbf{h}}\|} \; i,j = 1,2,\dots,M, \; i \neq j. \tag{6}$$

This result may be used to calculate the instantaneous square error in the frequency domain at the mth processing block which is minimized by the NMCFLMS algorithm

$$J_f(m) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \underline{\boldsymbol{e}}_{ij}^H(m)\underline{\boldsymbol{e}}_{ij}(m), \tag{7}$$

where $\underline{\boldsymbol{e}}_{ij}(m)$ is the frequency-domain block error signal between channels i and j. The algorithm is summarized as [6]

$$\begin{aligned} \hat{\underline{\boldsymbol{h}}}_k^{10}(m+1) =& \hat{\underline{\boldsymbol{h}}}_k^{10}(m) - \rho(\boldsymbol{\mathcal{P}}_k(m) + \delta\mathbf{I}_{2L\times 2L})^{-1} \\ & \times \sum_{i=1}^{M} \boldsymbol{\mathcal{D}}_{x_i}^*(m)\underline{\boldsymbol{e}}_{ik}^{01}(m), \; k = 1,2,\dots,M, \end{aligned} \tag{8}$$

where $\hat{\underline{\boldsymbol{h}}}^{10}$ and $\underline{\boldsymbol{e}}_{ik}^{01}(m)$ are the length $2L$ DFTs of h_k and e_{ik} at block m respectively as defined in [6], $\mathbf{I}$ is an identity matrix and

$$\boldsymbol{\mathcal{P}}_k(m) = \lambda\boldsymbol{\mathcal{P}}_k(m-1) + (1-\lambda) \tag{9}$$

$$\times \sum_{i=1, i\neq k}^{M} \boldsymbol{\mathcal{D}}_{x_i}^*(m)\boldsymbol{\mathcal{D}}_{x_i}(m), \; k = 1,2,\dots,M, \tag{10}$$

where $\boldsymbol{\mathcal{D}}_{x_i}(m)$ is a diagonal matrix of DFT coefficients for a block m of observation x_i. Symbols ρ, δ and λ denote control parameters. The estimate of the kth channel coefficient vector is $\hat{\mathbf{h}}_k(m) = [\hat{h}_{k,0}(m) \; \hat{h}_{k,1}(m) \; \dots \; \hat{h}_{k,L-1}(m)]^T$.

IV. Identification and Removal of Common Filter

The hypothesis investigated in this paper is that the effect of additive noise is to cause the LMS-type BSI adaptive filtering to 'see' a common filter h^{c} as a component in each channel, and whose estimation and removal will provide an improved estimate of the true channels. We now investigate this phenomenon further. The equality in (4), and hence the cross-relation error in (5), are blind to common filtering in the channel estimates,

$$x_i * \hat{h}_j = x_j * \hat{h}_i \tag{11}$$

$$x_i * \hat{h}_j * h^{\mathrm{c}} = x_j * \hat{h}_i * h^{\mathrm{c}}, \tag{12}$$

where h^{c} is a common filtering component. Consider then a decomposition of the estimated impulse responses,

$$\hat{h}_k = h^{\mathrm{c}} * h'_k, \tag{13}$$

where h'_k is a channel-dependent component at channel k. A supervised estimate of the common component may be found by the following summation, assuming estimation errors are uncorrelated between channels,

$$\hat{h}^{\mathrm{c}} = \frac{1}{M}\sum_{k=1}^{M} g_k * \hat{h}_k, \tag{14}$$

where g_k is the L-sample inverse of the true channel response h_k. The filter g_k is obtained by

$$g_k = \arg\min_{g_k} \|g_k * h_k - \delta(n)\|_2^2 \tag{15}$$

where $\delta(n)$ is a unit impulse function. Note that g_k is a noncausal function[1]. With the common component estimated in (14), the channel-dependent component h'_k can be estimated by the convolution

$$\hat{h}'_k = \hat{g}^{\mathrm{c}} * \hat{h}_k \tag{16}$$

where $\hat{g}^{\mathrm{c}}$ the L-sample is a least-squares inverse of $\hat{h}^{\mathrm{c}}$, found in a similar way to (15) by the minimization

$$\hat{g}^{\mathrm{c}} = \arg\min_{\hat{g}^{\mathrm{c}}} \|\hat{g}^{\mathrm{c}} * \hat{h}^{\mathrm{c}} - \delta(n)\|_2^2 \tag{17}$$

and is also a noncausal function. The decomposed channels may be represented in vector form as,

$$\hat{\mathbf{h}}_k = [h_{k,0} \; h_{k,1} \; \dots \; h_{k,L-1}] \tag{18}$$

$$\hat{\mathbf{h}}'_k = [h'_{k,0} \; h'_{k,1} \; \dots \; h'_{k,L-1}] \tag{19}$$

$$\hat{\mathbf{h}}^{\mathrm{c}} = [\hat{h}^{\mathrm{c}}_{-L/2} \; \hat{h}^{\mathrm{c}}_{-L/2+1} \; \dots \; \hat{h}^{\mathrm{c}}_{3L/2-1}]. \tag{20}$$

In the final case the anticausality is limited to $L/2$ samples and the total filter is $2L$ samples long; empirical results in Sec. V justify this choice. The appended argument $\hat{\mathbf{h}}_k(n)$ denotes the estimate at time n and concatenated channels are denoted by the matrices $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T \; \hat{\mathbf{h}}_2^T \; \dots \; \hat{\mathbf{h}}_M^T]^T$ and $\hat{\mathbf{h}}' = [\hat{\mathbf{h}}_1'^T \; \hat{\mathbf{h}}_2'^T \; \dots \; \hat{\mathbf{h}}_M'^T]^T$.

[1] The filters h_k and h^{c} are not generally minimum phase so an exact inverse does not exist.

V. Experimental Results

Three types of channel were simulated: room impulse responses (RIRs) for a room measuring $5 \times 6 \times 3$ m of length 128 taps and $T_{60} = 0.3$ s with an array of 5 microphones spaced at 0.2 m intervals using the method of images [7], five random channels of length 16 taps, and five exponentially-decaying channels of length 16 taps. A sampling rate of $f_s = 8000$ Hz was employed. Care was taken to ensure the absence of common zeros in the simulated channels. The channels were excited by 10 seconds of white Gaussian noise and noise was added to the observed signals to obtain a signal-to-noise ratio (SNR) of $\{10, 30, 60\}$ dB at the received signals $\mathbf{x}_i(n)$. A total of 20 Monte Carlo realizations were simulated and averaged.

The normalized projection misalignment (NPM) between the true impulse response $\mathbf{h}$ and estimated impulse responses $\hat{\mathbf{h}}(n)$ is defined as [8]

$$\text{NPM}[\mathbf{h}(n), \hat{\mathbf{h}}(n)] = 20 \log_{10} \left(\frac{\|\mathbf{h} - \kappa(n)\hat{\mathbf{h}}(n)\|_2}{\|\mathbf{h}\|_2} \right), \quad (21)$$

where $\kappa(n)$ is defined as

$$\kappa(n) = \frac{\mathbf{h}^T \hat{\mathbf{h}}(n)}{\hat{\mathbf{h}}^T(n)\hat{\mathbf{h}}(n)}. \quad (22)$$

The results Figs. 1, 2 and 3 show (a) the NPM between true and estimated systems, (b) the cost function in (7) and (c) the NPM between true and channel-dependent component of the estimated system as a function of time for the RIRs, random and exponentially-decaying random channels respectively. In all cases (a) converges more slowly than (b), in particular in the 10 dB SNR case where misconvergence occurs in (a). Furthermore, the NPM in (c) is consistently lower than (a) and does not misconverge. The implication of these results is that a common component is present in the estimated channels both before and after misconvergence, and that the system error in (c) converges to a level that is related to the SNR. The curves in (c) also follow a similar convergence characteristic to (b) suggesting that, provided a common component is removed, the cost function is a good measure of system error.

The curves in Fig. 4 show, for the RIR case, (a) the ℓ_2-norm between the common component and a unit impulse

$$\|\hat{\mathbf{h}}^{\text{c}} - \boldsymbol{\delta}\|_2, \quad (23)$$

where $\boldsymbol{\delta}$ is a vector of M concatenated delayed unit impulses

$$\boldsymbol{\delta} = [\underbrace{\mathbf{0}_{1\times L/2} \; 1 \; \mathbf{0}_{1\times 3L/2-1}}_{M \text{ times}} \; \ldots]^T, \quad (24)$$

and (b) the NPM between neighbouring estimates of $\hat{\mathbf{h}}^{\text{c}}$. These results show that the common filter's amplitude varies as a function of time but that its shape varies relatively little. In the case of 10 dB SNR, the common component amplitude varies in a similar fashion to the misconverging curve in Fig. 1, showing that the misconvergence is due largely to an increased common component and not convergence to an entirely incorrect solution. The ℓ_2-norm of the common component is seen in all cases to peak shortly after the filter begins to adapt before reducing in subsequent iterations; this can also be seen around 0.25 s in Fig. 2 (c). Only if the estimate begins to misconverge does the ℓ_2-norm of the common component later increase. Similar results have been observed with longer channels of up to 200 taps and with time-domain LMS BSI algorithms.

The plot in Fig. 5 shows (a) the common component $\hat{\mathbf{h}}^{\text{c}}$ and (b) its energy decay curve (obtained by backwards integration) [9] at the last iteration of the misconverged estimate of the room impulse responses in the 10 dB SNR case. The common component decays to -60 dB of its peak value in approximately $3L/2$ samples, suggesting that the length of this filter is long compared with the true channels.

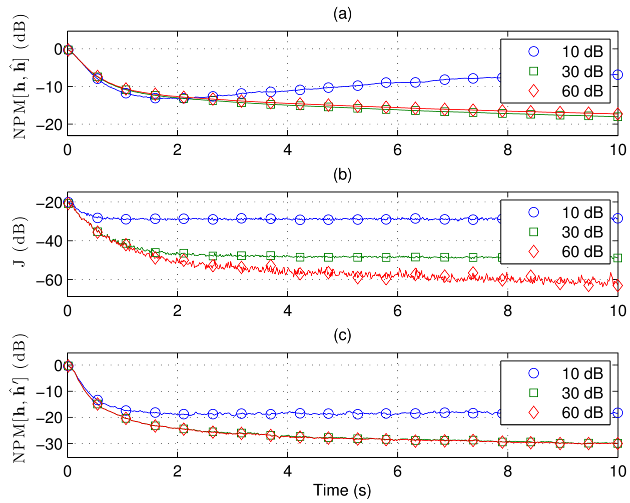

Fig. 1. (a) NPM between true and estimated systems, (b) Cost function, (c) NPM between true and channel-dependent component of estimated system for room impulse responses with varying SNR.

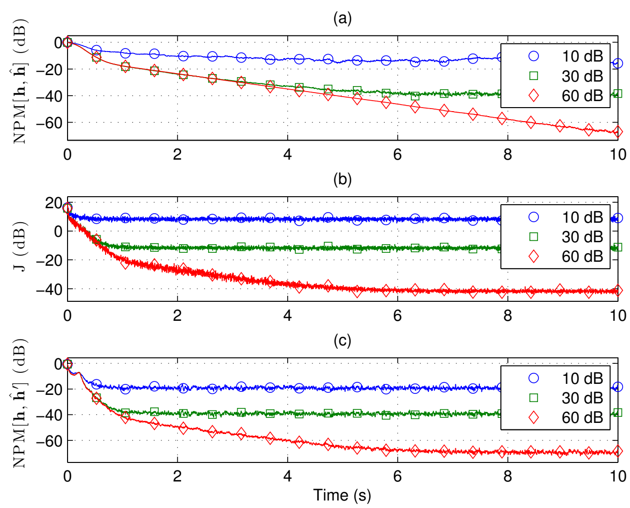

Fig. 2. (a) NPM between true and estimated systems, (b) Cost function, (c) NPM between true and channel-dependent component of estimated system for random channels with varying SNR.

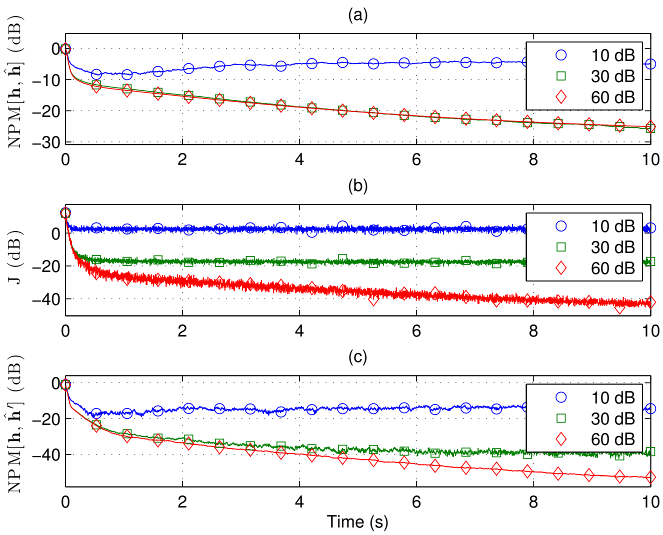

Fig. 3. (a) NPM between true and estimated systems, (b) Cost function, (c) NPM between true and channel-dependent component of estimated system for exponentially-decaying random channels with varying SNR.

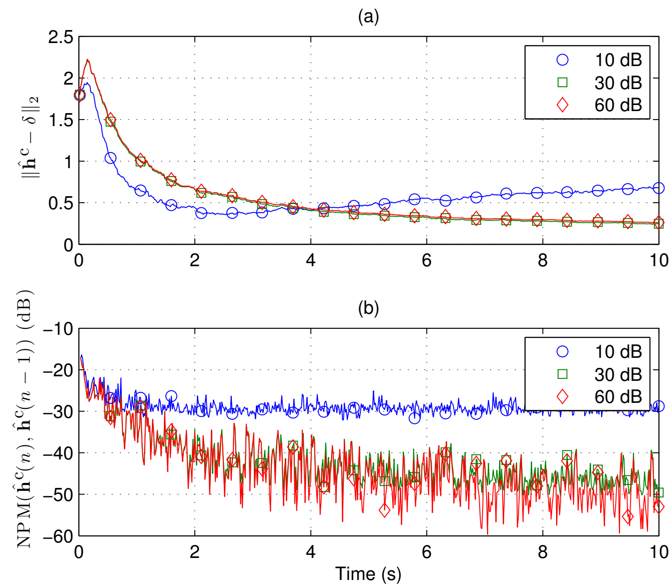

Fig. 4. (a) ℓ_2-norm between common component and delayed delta, (b) NPM between neighbouring common components for the acoustic transfer function.

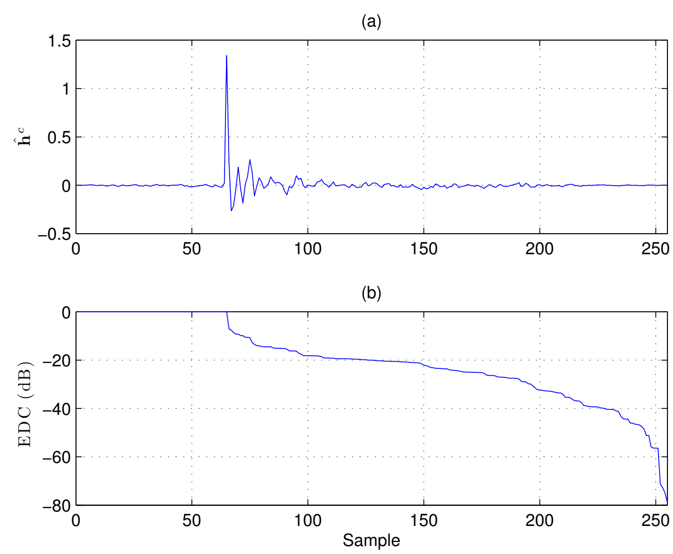

Fig. 5. (a) Common component for a misconverged system, (b) Energy decay curve of the common component.

VI. Conclusion

A study has been conducted into common filtering in impulse responses estimated with noisy adaptive blind system identification. A supervised technique was proposed for studying the effects of common filtering that requires prior knowledge of the true system. Experimental evidence with room impulse responses and random channels has shown that common filters can be identified and removed from the estimated channels, improving the rate of convergence and system error. It is also observed that these channel estimates do not misconverge, offering the explanation that misconvergence is due to a common filtering and not convergence to an entirely incorrect solution. The cause of common filtering in adaptive BSI remains an open question, however these findings suggest that future algorithms exploiting knowledge of common filter behaviour will exhibit improved noise robustness.

VII. Acknowledgement

The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 226007 SCENIC

References

[1] 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.

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

[3] M. K. Hasan and P. A. Naylor, "Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: robustness issue," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sep. 2006.

[4] 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 Processing Conf. (EUSIPCO)*, 2005.

[5] A. Liavas and P. Regalia, "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, Jun. 1999.

[6] 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.

[7] 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.

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

[9] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.