

# STATISTICAL MODELLING OF MULTICHANNEL BLIND SYSTEM IDENTIFICATION ERRORS

*Felicia Lim, Patrick A. Naylor*

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

## ABSTRACT

It is well known that blind system identification (BSI) algorithms misconverge in the presence of noise and that applications relying on such channel estimates must be designed to be robust to these blind system identification errors (BSIEs). However, there is currently no generalized model of BSIEs in the literature and instead, white Gaussian noise (WGN) is commonly assumed. This paper investigates the statistics of BSIEs based on a robust state-of-the-art BSI algorithm using both simulated and real impulse responses. A BSIE model is proposed based on Gaussian mixture models (GMMs) and a method for generating artificial BSIEs based on this model for simulations is given. Comparisons against alternative assumptions used in the literature are given and it is shown through experimental results that the proposed model gives BSIEs that are most statistically similar to the ground truth.

**Index Terms**— blind system identification, modelling acoustic channel errors

## 1. INTRODUCTION

Channel estimates provided by blind system identification (BSI) algorithms are important for a range of applications [1], including speech dereverberation [2] and automatic camera steering for teleconferencing [3]. In the presence of noise, the accuracy of the BSI algorithms suffers and therefore downstream algorithms must be designed to be robust to the resulting blind system identification errors (BSIEs) [4]. An accurate model of the BSIEs would therefore be useful for development of such algorithms, and further enable realistic simulations as a convenient means of performance evaluation.

To the best knowledge of the authors, there is currently no established systematic model for BSIEs. In the literature, a commonly used model is white Gaussian noise (WGN), for example in [4, 5, 6, 7]. In [8], the BSIEs are modelled as WGN with an exponential damping factor governed by the room's reverberation time.

This paper investigates the statistics of BSIEs associated with a robust BSI algorithm from the literature for both simulated and real acoustic impulse responses (AIRs). It will be shown that the BSIEs are related to the AIR coefficients'

amplitudes and a model based on a trained Gaussian mixture model (GMM) is proposed. A method to subsequently generate artificial BSIEs based on the proposed model will be described and shown to be statistically more similar to the true BSIEs when compared against alternative models.

## 2. OVERVIEW OF BSI ALGORITHMS

Consider an  $M \geq 2$  channel acoustic single-input multiple-output (SIMO) finite impulse response (FIR) system, where the  $m$ -th channel AIR is denoted  $h_m$ . The noisy and reverberant signal captured at the  $m$ -th microphone is given as

$$x_m(n) = h_m * s(n) + v_m(n), \quad (1)$$

where  $s(n)$  is the source signal,  $v_m(n)$  is an additive noise and  $*$  denotes linear convolution. The additive noise in the different channels are assumed to be WGN that are uncorrelated with each other and with the source signal.

BSI algorithms aim to blindly estimate the AIRs from the noisy and reverberant microphone signals and can be classified into two categories, higher-order statistics (HOS) and second-order statistics (SOS) methods [9]. In particular, the normalized multichannel frequency domain least mean squares (NMCFLMS) algorithm proposed in [10] has attracted significant interest as an adaptive algorithm that is computationally efficient and converges relatively quickly to the true filter coefficients. It is based on the cross-relation property

$$x_m * h_{m'} = s * h_m * h_{m'} = x_{m'} * h_m \quad (2)$$

that can be deduced from (1) for the noiseless case. In the presence of noise, an error signal can be derived as

$$e_{mm'}(n) = \mathbf{x}_m^T(n) \hat{\mathbf{h}}_{m'} - \mathbf{x}_{m'}^T(n) \hat{\mathbf{h}}_m, \quad m \neq m' \quad (3)$$

for  $m, m' = 1, 2, \dots, M$ , where  $\mathbf{x}_m(n) = [x_m(n) \ x_m(n-1) \ \dots \ x_m(n-L+1)]^T$  and  $\mathbf{h}_m = [h_{m,0} \ \dots \ h_{m,L-1}]^T$ . NMCFLMS aims to minimize a normalized error signal derived from (3) in the frequency domain. Unfortunately, in the presence of noise it suffers from misconvergence as the adaptation process progresses. In [11], the robust normalized multichannel frequency-domain least mean squares

(RNMCFMLS) algorithm was proposed and addresses the misconvergence issue by introducing a flat spectral constraint on the filter coefficients. The RNMCFMLS cost function is then defined for the  $b$ -th processing block as [11]

$$J(b) = J_f(b) - \beta_R(b) J_p(b), \quad (4)$$

where  $\beta_R(b)$  is the Lagrange multiplier addressed later,

$$J_f(b) = \sum_{m=1}^{M-1} \sum_{m'=m+1}^M \mathcal{E}_{mm'}^H(b) \mathcal{E}_{mm'}(b), \quad (5)$$

where  $\mathcal{E}_{mm'}(b)$  is the frequency-domain error signal between channels  $m$  and  $m'$ ,  $\{\cdot\}^H$  is the Hermitian transpose,

$$J_p(b) = \sum_{i=1}^{ML} \ln \left( |\hat{h}(b, i)|^2 \right) \quad (6)$$

is the spectral constraint where  $\hat{h}(b, i)$  is the  $i$ -th element of  $\hat{\underline{h}}(b)$ ,  $\hat{\underline{h}}(b) = [\hat{h}_1^T(b) \dots \hat{h}_M^T(b)]^T$  and  $\hat{h}_m(b)$  is the discrete Fourier transform (DFT) of  $\hat{h}_m$  at block  $b$ . The gradient of (5) is  $\nabla J_f(b) = [\nabla J_{f,1}^T(b) \dots \nabla J_{f,M}^T(b)]^T$ , where

$$\nabla J_{f,m}(b) = \mathcal{P}_m^{-1}(b) \times \sum_{m'=1}^M \mathcal{D}_{m'}^*(b) \mathcal{E}_{m'm}^{01}(b), \quad (7)$$

in which  $*$  is the complex conjugate,  $\mathcal{E}_{m'm}^{01}(b)$  is the length  $2L$  DFTs of  $e_{m'm}$  at block  $b$ ,  $\mathcal{D}_m(b)$  is the diagonal matrix of DFT coefficients for block  $b$  of  $x_m$  and

$$\mathcal{P}_m(b) = \gamma \mathcal{P}_m(b-1) + (1-\gamma) \sum_{\substack{m'=1 \\ m' \neq m}}^M \mathcal{D}_{m'}^*(b) \mathcal{D}_{m'}(b), \quad (8)$$

where  $\gamma = [1 - 1/(3L)]^L$ . The gradient of (6) is obtained as  $\nabla J_p(b) = [\nabla J_{p,1}^T(b) \dots \nabla J_{p,M}^T(b)]^T$ , where

$$\nabla J_{p,m}(b) = \mathcal{Q}_m(b) \hat{\underline{h}}_m(b), \quad (9)$$

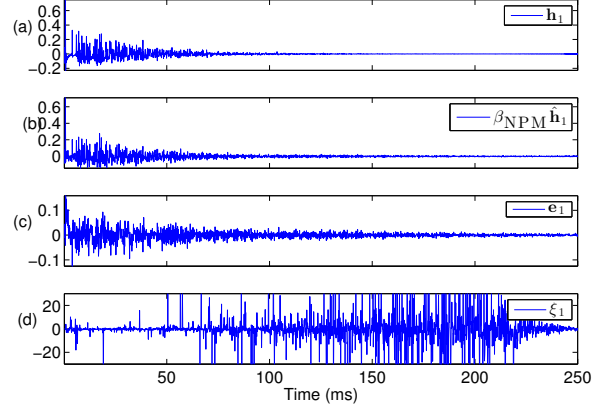
and  $\mathcal{Q}_m(b)$  is the diagonal matrix of  $2/|\hat{h}_m(b, i)|^2$ ,  $i = 1, \dots, L$ . The Lagrange multiplier can now be given as

$$\beta_R(b) = \left| \frac{\nabla J_p^H(b) \nabla J_f(b)}{\|\nabla J_p(b)\|^2} \right| \quad (10)$$

and the RNMCFMLS update equation summarized as

$$\hat{\underline{h}}_m^{10}(b+1) = \hat{\underline{h}}_m^{10}(b) - \rho \nabla J_{f,m}(b) + \rho \beta_R(b) \nabla J_{p,m}(b), \quad (11)$$

where  $\hat{\underline{h}}_m^{10}(b)$  is the length  $2L$  DFT of  $\hat{h}_m$  at block  $b$  and  $0 < \rho < 2$  is the step-size. In this work, the BSIEs resulting from RNMCFMLS will be investigated and modelled.



**Fig. 1:** Example of coefficients in the first channel of (a) a true AIR, (b) the normalized AIR estimated using RNMCFMLS, (c) the resultant BSIEs and (d) the corresponding BSIE ratios.

### 3. A STUDY OF THE BSIE COEFFICIENTS

Estimated AIRs obtained using cross-correlation based BSI algorithms contain an inherent scale factor, which can be ignored, deduced or assumed in some applications such as speech dereverberation [12]. Therefore it is desirable to investigate  $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T \dots \hat{\mathbf{h}}_M^T]^T$  independently of the scaling factor. A widely used measure of BSIE levels is the normalized projection misalignment (NPM) measure [13], where  $\hat{\mathbf{h}}$  is normalized by applying a gain factor  $\beta_{\text{NPM}} = (\mathbf{h}^T \hat{\mathbf{h}}) / (\hat{\mathbf{h}}^T \hat{\mathbf{h}})$ , with  $\mathbf{h} = [\mathbf{h}_1^T \dots \mathbf{h}_M^T]^T$ . Following this method, the normalized BSIEs are calculated as

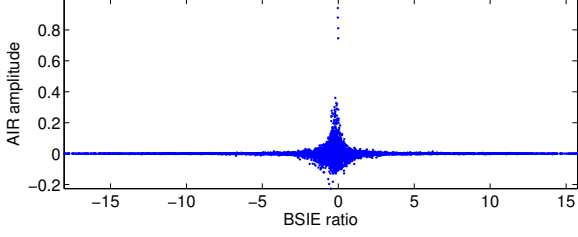
$$\mathbf{e}_m = \beta_{\text{NPM}} \hat{\mathbf{h}}_m - \mathbf{h}_m. \quad (12)$$

An example of  $\mathbf{e}_m$  obtained using RNMCFMLS is given in Fig 1. While it may appear that the BSIEs can be modelled by WGN with an exponential decay as proposed in [8], the behaviour exhibited by BSIE coefficients in the region of early reflections suggests that a better model might be found by investigating the relationship between the BSIEs and their corresponding  $\mathbf{h}$  amplitudes. Let the BSIE ratios be defined as  $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T \dots \boldsymbol{\xi}_M^T]^T$ , where

$$\boldsymbol{\xi}_m = [e_{m,0} \dots e_{m,L-1}] \odot \left[ \frac{1}{h_{m,0}} \dots \frac{1}{h_{m,L-1}} \right], \quad (13)$$

and  $\odot$  denotes the Hadamard (element-wise) multiplication.

Fig. 1 shows, for an arbitrarily chosen example, the relationship between  $\boldsymbol{\xi}_m$  and the amplitudes of  $\mathbf{h}_m$ , with (d) in Fig. 1 showing the BSIE ratios (13). Let their joint observations be defined as  $\mathbf{O} = [\mathbf{o}_1^T \dots \mathbf{o}_M^T]^T$ , where  $\mathbf{o}_m = [[h_{m,0}, \xi_{m,0}]^T \dots [h_{m,L-1}, \xi_{m,L-1}]^T]^T$ , with an example plot given in Fig. 2. It can be observed that  $\mathbf{h}$  coefficients with larger magnitudes (e.g. in the direct path and strong early reflections) are associated with  $\boldsymbol{\xi}$  that have



**Fig. 2:** An example of the joint distribution between the AIR coefficient amplitudes and their BSIE ratios.

smaller variances, while  $\mathbf{h}$  coefficients with smaller magnitudes, such as those found in the later part of the AIR, are associated with  $\xi$  that have larger variances. The mean of  $\xi$  is approximately  $\mathbf{0}$  for all  $\mathbf{h}$  amplitudes. These characteristics suggest that RNMCFLMS performs better in estimating AIR coefficients with magnitudes in the higher part of the amplitude probability density function (PDF) than for the smaller amplitudes.

#### 4. PROPOSED BSIE MODEL

It is desirable to find a parameterizable model for  $\mathbf{O}$  from which statistically realistic BSIEs can be generated. A GMM is used to model the distribution [14]. For  $K$  number of Gaussians, the joint probability distribution of  $\mathbf{O}$  can be approximated as

$$p(\mathbf{O}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{O} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (14)$$

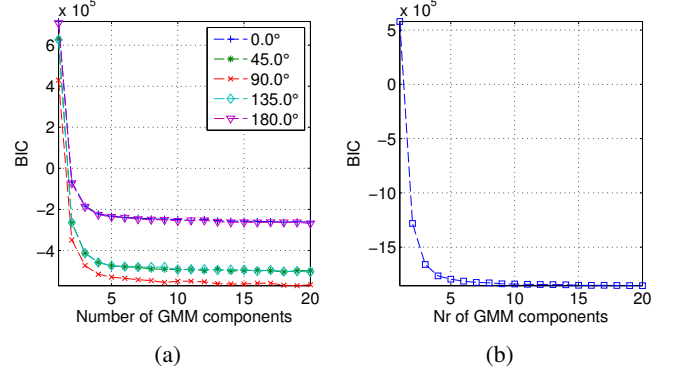
where each  $\mathcal{N}(\mathbf{O} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is a component of the GMM with its own mean  $\boldsymbol{\mu}_k$  and covariance  $\boldsymbol{\Sigma}_k$ , and  $\pi_k$  is the mixing coefficient, with  $\sum_{k=1}^K \pi_k = 1$ ,  $0 \leq \pi_k \leq 1$ . A best-fit GMM for a given  $\mathbf{O}$  can be found using the method of expectation-maximization (EM) [14]. A more complex GMM with larger  $K$  will almost always result in a better fit to the data; however this runs the risk of overfitting to the specific data considered. In this work, the trade-off between model complexity and goodness-of-fit is evaluated using Bayesian information criterion (BIC) [15], given as

$$\text{BIC} = K \ln(N_{\text{obs}}) - 2 \ln \mathcal{L}(\boldsymbol{\theta}_{\text{GMM}}), \quad (15)$$

where  $N_{\text{obs}}$  is the number of observations in  $\mathbf{O}$  and  $\mathcal{L}(\boldsymbol{\theta}_{\text{GMM}})$  is the maximized likelihood function of the best-fit GMM with parameters  $\{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{\forall k \in \{1 \dots K\}}$ . The optimum  $K$ ,  $K_{\text{opt}}$  can then be found as the  $K$  that gives the minimum BIC.

In the remainder of this section, the GMM parameters and their BICs are computed to describe BSIE models for various acoustic scenarios using both simulated and real AIRs.

The simulated AIRs were obtained using the image method [16, 17] with the following acoustic setup. A 5-channel endfire uniform linear array (ULA) with an inter-microphone spacing of 4 cm was placed in the middle of a



**Fig. 3:** BICs over a range of  $K$  GMM components for (a) image method AIRs, and (b) MARDY AIRs

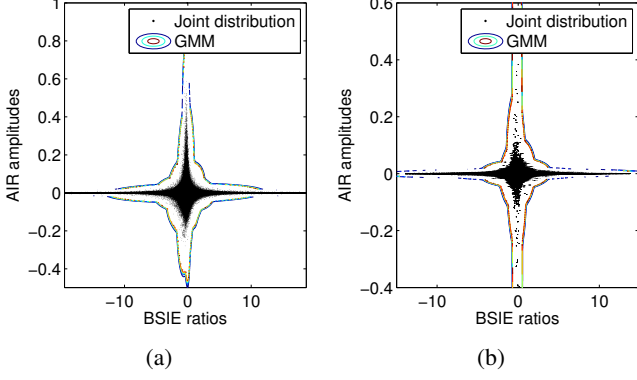
room. A source was stepped around the centre of the ULA at a 1 m radius for  $P = 5$  angles from  $0^\circ$  to  $180^\circ$  on the horizontal plane. At each angle, the AIRs were simulated with sampling frequency  $f_s = 8$  kHz, reverberation time  $T_{60} = 250$  ms and length  $L = T_{60}f_s = 2000$  coefficients. A total of 30 AIRs were obtained for each angle, each corresponding to a randomly generated room between  $3 \times 3 \times 2$  m and  $4 \times 5 \times 3$  m.

Additionally, real AIRs were taken from the MARDY database [18], which comprises 8-channel impulse responses for 3 source-to-microphone distances at  $\{1, 2, 3\}$  m and 3 different source locations, giving a total of 9 different acoustic setups. To maintain consistency with the image method experiments, the number of channels were limited to  $M = 5$  by generating 4 subsets from each 8-channel acoustic setup, where the  $i$ -th subset is formed as  $\{m_i \dots m_{i+4}\}$ . Therefore, in total, 36 experiments were formed from 4 subsets  $\times$  9 acoustic setups.

The microphone signals were then generated from the AIRs as (1) using WGN for  $v_m(n)$  with a signal-to-noise ratio (SNR) of 15 dB and RNMCFLMS was used to estimate  $\hat{\mathbf{h}}$ . BSI algorithms are known to misconverge in the presence of noise and in this work, the  $\hat{\mathbf{h}}$  that gives the minimum NPM over 300 s of  $\mathbf{x}_m$  was used. In practice, this could potentially be achieved using algorithm control techniques [19].

For each set of the  $P$  image method AIRs and the collective MARDY AIRs,  $\xi$  was computed using (13) and aggregated over all AIRs within the set. Outliers were avoided by removing the top and bottom 5-th percentiles in amplitude, along with their corresponding  $\mathbf{h}$  coefficients. The best-fit GMMs were then found for the remaining joint observations  $\mathbf{O}$  for  $K \in \{1, 2, \dots, 20\}$  and their BICs were computed.

The BIC results for the simulated AIRs are given in Fig. 3a with an example of the best-fit GMM with  $K_{\text{opt}} = 20$  in Fig. 4a for  $p = 5$  ( $180^\circ$ ). The BIC results for the real AIRs are given in Fig. 3b and Fig. 4b shows the best-fit GMM with  $K_{\text{opt}} = 20$ .



**Fig. 4:** Best-fit GMMs obtained with  $K_{\text{opt}}$  for (a) image method AIRs with  $p = 5$  ( $180^\circ$ ), and (b) MARDY AIRs

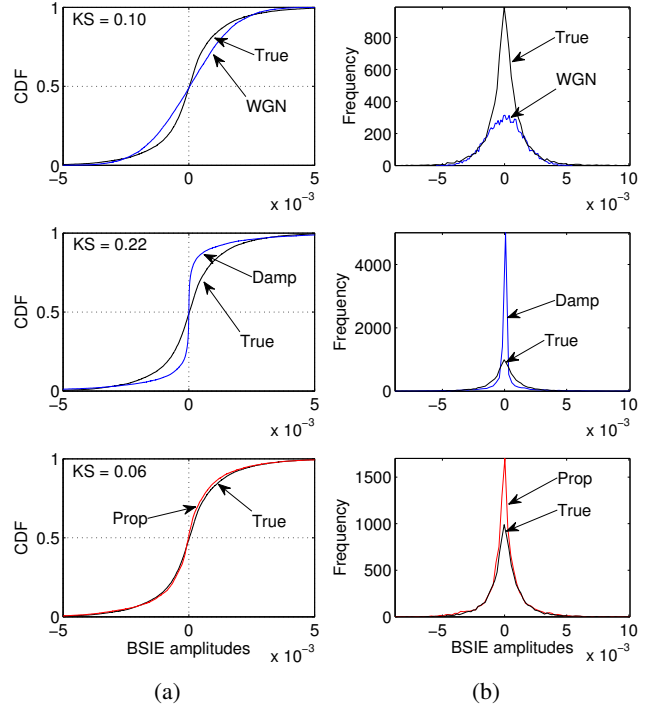
## 5. VALIDATION OF PROPOSED MODEL

Given knowledge of  $\mathbf{h}$  and the trained GMM parameters, artificial BSIEs can be generated from the proposed model as follows. For each  $\mathbf{h}$  coefficient,  $h_{m,i}$ , the conditional probability distribution of the BSIE ratio,  $P(\xi_{m,i}|h_{m,i}, \theta_{\text{GMM}})$  is computed [20] and a corresponding artificial BSIE ratio,  $\xi_{m,i}^A$ , is generated by drawing randomly from  $P(\xi_{m,i}|h_{m,i}, \theta_{\text{GMM}})$ . The artificial BSIE coefficients are then obtained as  $\mathbf{e}_m^A = \mathbf{h}_m \odot \boldsymbol{\xi}_m^A$ , where  $\boldsymbol{\xi}_m^A = [\xi_{m,0}^A \dots \xi_{m,L-1}^A]^T$ , and the artificial estimated AIRs given as  $\hat{\mathbf{h}}_m^A = \mathbf{h}_m + \mathbf{e}_m^A$ . In applications of artificial BSIE models, it may be desirable to subsequently adjust the NPM, which can be achieved as described in [21].

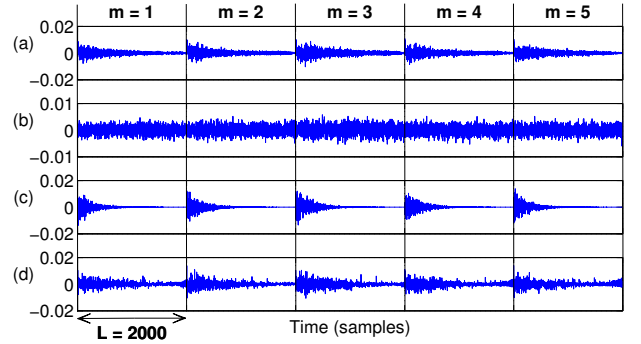
The statistics of the proposed proportional BSIE model can now be compared against the true BSIEs and two alternative models, 1) WGN and 2) Damp [8]. In all cases, the BSIEs were first scaled to  $\text{NPM} = -30$  dB and the cumulative distribution functions (CDFs) of  $\mathbf{e}^A$  then computed and compared against the CDFs of  $\mathbf{e}$ . The differences can be quantitatively evaluated with the Kolmogorov-Smirnov (K-S) statistic,  $\kappa$ , which is a distance measure between two CDFs [22], where a smaller  $\kappa$  indicates that the sample is more similar to the desired distribution. The comparison CDFs plots and their corresponding  $\kappa$  values are given in Fig. 5a where it can be seen that  $\kappa$  for the proposed BSIE model is the minimum amongst all models considered. The BSIE histograms and coefficients are given in Fig. 5b and Fig. 6 for qualitative comparison, where it can be seen that the proposed model is most similar to the true BSIEs. Results given here are for an example set of true BSIEs obtained with simulated AIRs. Similar trends were observed for additional experiments that we have performed with real AIRs.

## 6. CONCLUSIONS

The statistics of BSIEs was investigated for a robust BSI algorithm, RNMCFMLS. A model was subsequently developed



**Fig. 5:** (a) CDF comparison plots with their associated K-S statistics. (b) Histograms of the true and artificial BSIEs.



**Fig. 6:** The multichannel concatenated BSIE amplitudes for the different models: (a) True (b) WGN (c) Damp (d) Prop.

based on GMMs that can be used to generate realistic BSIEs for simulation purposes when investigating related applications, such as channel equalization. It was shown both quantitatively (using the K-S statistic) and qualitatively (through the histograms and time-domain plots) that BSIEs artificially generated from the proposed model are statistically more similar to the true BSIEs than alternative models. This model is therefore highly suitable for use in simulation experiments to evaluate the performance of, for example, dereverberation algorithms based on inverse of estimated AIRs with known levels of estimation error.

## 7. REFERENCES

- [1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
- [2] 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.
- [3] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 509–519, Sep. 2004.
- [4] 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.
- [5] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 718–724, Nov. 1999.
- [6] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.
- [7] R. Rashobh and A. W. H. Khong, "A multichannel time-domain subspace approach exploiting multiple time-delays for acoustic channel equalization," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.
- [8] W. Zhang, E. A. P. Habets, and P. A. Naylor, "A system-identification-error-robust method for equalization of multichannel acoustic systems," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, Mar. 2010.
- [9] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.
- [10] 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.
- [11] 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.
- [12] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.
- [13] 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.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [15] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [16] 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.
- [17] E. A. P. Habets. (2008, May) Room impulse response (RIR) generator. [Online]. Available: <http://home.tiscali.nl/ehabets/rirgenerator.html>
- [18] J. Wen, N. D. Gaubitch, E. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Paris, France, Sep. 2006.
- [19] R. Ahmad, N. D. Gaubitch, and P. A. Naylor, "A noise-robust dual filter approach to multichannel blind system identification," in *Proc. European Signal Processing Conf.*, 2007.
- [20] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, 2006. [Online]. Available: <http://www.gaussianprocess.org/gpml/>
- [21] 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.
- [22] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.