ACOUSTIC MODELS FOR ONLINE BLIND SOURCE DEREVERBERATION USING SEQUENTIAL MONTE CARLO METHODS

*Christine Evers** and *James R. Hopgood*

Institute for Digital Communications,
School of Engineering and Electronics, University of Edinburgh, UK
{c.evers,james.hopgood}@ed.ac.uk

Judith Bell

School of Engineering and Physical Sciences,
Heriot-Watt University, UK
j.bell@hw.ac.uk

ABSTRACT

Reverberation and noise cause significant deterioration of audio quality and intelligibility to signals recorded in acoustic environments. Noise is usually modeled as a common signal observed in the room and independent of room acoustics. However, this simplistic model cannot necessarily capture the effects of separate noise sources at different locations in the room. This paper proposes a noise model that considers distinct noise sources whose individual acoustic impulse responses are separated into source-sensor specific and common acoustical resonances. Further to noise, the signal is distorted by reverberation. Using parametric models of the system, recursive expressions of the filtering distribution can be derived. Based on these results, a sequential Monte Carlo approach for online dereverberation and enhancement is proposed. Simulation results for speech are presented to verify the effectiveness of the model and method.

***Index Terms*—** Acoustic signal processing, speech enhancement, speech dereverberation, sequential estimation, Monte Carlo

1. INTRODUCTION

Audio signals in confined spaces exhibit reverberation due to reflections off surrounding obstacles. In addition to a direct path signal, time-shifted echoes are received, leading to spectral coloration and reduced intelligibility. Moreover, the signal is distorted by noise, which is usually considered to be a common signal observed within a room. In this paper, we propose a model that considers spatially distinct noise sources with individual acoustic impulse responses (AIRs) exhibiting reverberation (Fig. 1). The individual channels of the audio and noise sources are separated into source-sensor specific and common acoustical resonances.

Multiple sensor blind dereverberation techniques exploit spatial diversity of acoustic channels. Where sensor arrays are impractical, single-channel blind dereverberation proves effective. However, spatial diversity cannot be utilized in this inherently underdetermined problem and prior knowledge must be incorporated.

In a model-based approach, parametric models are assumed for both the source and the channel. Based on the state-space representation of the speech and channel model, recursive expressions can be derived for the filtering distribution. Estimates of the source signal and the model parameters can be obtained from the recursive representation of the filtering distribution in a sequential manner. Sequential estimation facilitates online processing of the signal, which is of particular interest for applications such as security surveillance systems where results should become available as soon as a signal sample is measured, i.e., where batch methods are impractical. Particle filters (or sequential Monte Carlo (SMC) methods) represent a target distribution by a large number of random variates from a hypothesis distribution. Incorporation of knowledge about the current and past measured samples allows for correction and evolution of the particles in time. Particle filters were shown to effectively enhance systems distorted by white Gaussian noise (WGN) [1] and for reverberant all-zero channels [2]. This paper extends this work to reverberant all-pole channels and spatially distinct noise sources.

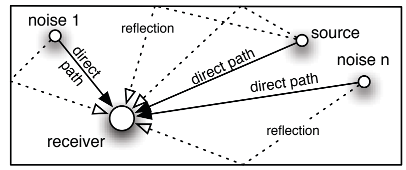


Figure 1: Distant noise source filtered through separate channels

The system model is presented in sect. §2. Sect. §3 discusses blind signal and parameter estimation using particle filters. Experimental results are presented and conclusions drawn in §4 and §5.

2. SYSTEM MODEL

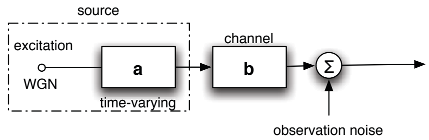
2.1. Reverberation and background noise

Noise is usually modeled as an additive common signal close to the microphone and unaffected by the room acoustics. Therefore, it can be added at the output of the system (Fig. 2a). In a more realistic model, spatially distinct noise sources are each observed after they have propagated through the acoustic system, and therefore have corresponding but distinct AIRs. Hence, a combination of signals filtered by separate channels is observed at the receiver (Fig. 2b).

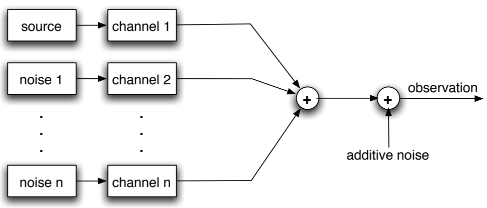
While the model in Fig. 2b is idealistic, it is also overly complicated, making it difficult to estimate all the relevant system parameters. Moreover, it can be simplified using the notion of common acoustical poles. In [3], Haneda *et al.* decompose individual channels into a combination of two components: one that is dependent on the source-sensor geometry, and one that is acoustically common to all source-sensor arrangements. Motivated by the presence of common resonances, we propose to apply this separation to the model in Fig. 2b to obtain a more realistic room acoustical model (Fig. 3a).

Although the general model in Fig. 3a is of great interest, the presence of general room transfer functions (RTFs) dependent on source-sensor geometries leads to difficulties in uniquely identifying the source signals in the blind deconvolution problem. Identifiability results are required before this model can be used with confidence. Therefore, this paper uses a simplified model (Fig. 3b) to investigate if the signal enhancement methodology proposed in the following is appropriate for the model in Fig. 3a.

* This work was funded by the Edinburgh Research Partnership in Engineering and Mathematics, and partially sponsored through the German Academic Exchange Service (DAAD) PhD scholarship



(a) Additive noise model



(b) Remote noise sources filtered through separate channels

Figure 2: Conventional noise model assuming noise at sensor vs. proposed noise model assuming remote noise sources

2.2. Source Model

Autoregressive (AR) processes are a popular approach for modeling the vocal tract of a speaker due to their accurate modeling of the short-term spectrum of speech. However, stationary autoregressive (AR) processes result in poor modeling of speech signals due to the continually changing nature of the vocal tract. To reconcile this time-variance, the parameter variation of the signal is modeled as a non-stationary process. A Q^{th} order time-varying AR (TVAR) process can be expressed by

$$x_t = \sum_{q \in \mathcal{Q}} a_{q,t} x_{t-q} + \sigma_{e_t} e_t, \tag{1}$$

where $e_t \sim \mathcal{N}(0, 1)$, $\mathbf{x} = \begin{bmatrix} x_0 & \ldots & x_t \end{bmatrix}^T$ is the speech sequence for $t \in \mathcal{T}$ samples,[1] $\mathbf{a}_t = \begin{bmatrix} a_{1,t} & \ldots & a_{Q,t} \end{bmatrix}^T$ is the set of time-varying coefficients, and $\sigma^2_{e_t}$ is the variance of the innovation.

Often, the signal is modeled as a linear time-varying process by transforming the time-varying parameters through a set of known basis functions to a space where they can be analyzed as a linear time-invariant process (see e.g. [4] and references therein). However, the accuracy of the estimate in this model is highly dependent on the choice of basis functions [4]. An alternative approach that facilitates the use of sequential Bayesian methods is to model the time-varying parameters and excitation sequence as a random walk [1] specified by the first-order Markov chain,

$$p\left(\mathbf{a}_t \mid \mathbf{a}_{t-1}\right) \propto \mathcal{N}\left(\mathbf{a}_t \,\middle|\, \mathbf{a}_{t-1}, \boldsymbol{\Delta}_{\mathbf{a}}\right) \mathbb{I}_{\mathcal{A}_Q}(\mathbf{a}_t) \tag{2a}$$

$$p\left(\phi_{e_t} \mid \phi_{e_{t-1}}\right) = \mathcal{N}\left(\phi_{e_t} \,\middle|\, \phi_{e_{t-1}}, \delta^2_e\right) \tag{2b}$$

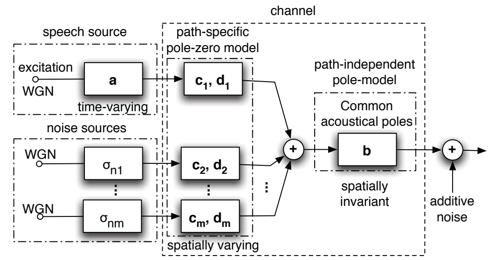
where[2] $\phi_{e_t} = \ln \sigma^2_{e_t}$, $\mathbb{I}_{\mathcal{A}_Q}(\mathbf{a}_t)$ denotes the indicator function for the region of support, $\mathcal{A}_Q$, of the source parameters, $\mathbf{a}_t \in \mathcal{A}$. The initial states are given by $p(\mathbf{a}_0) \propto \mathcal{N}\left(\mathbf{a}_0 \,\middle|\, \mathbf{0}_{Q\times 1}, \boldsymbol{\Delta}_{\mathbf{a}_0}\right) \mathbb{I}_{\mathcal{A}_Q}(\mathbf{a}_0)$ and $p(\phi_{e_0}) \triangleq \mathcal{N}\left(\phi_{e_0} \,\middle|\, 0, \delta^2_{e_0}\right)$. The set of Markov parameters $\left\{\boldsymbol{\Delta}_{\mathbf{a}}, \boldsymbol{\Delta}_{\mathbf{a}_0}, \delta^2_e, \delta^2_{e_0}\right\}$ is assumed known and constant.

2.3. Channel Model

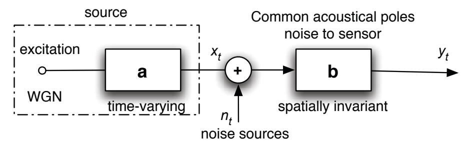
The acoustic wave equation indicates that room transfer functions can be expressed as conventional pole-zero models. While these models capture both resonances and time-delays, the inclusion of zeros in the model requires the solution of a set of nonlinear equations.

[1] The set notation $\mathcal{U} = \{1, \ldots, U\}$, $U \in \mathbf{N}$, is used for simplicity.

[2] By definition variances are positive, enforced by sampling from $\ln \sigma^2_{e_t}$.



(a) Proposed model with spatially separated noise sources and separation of common acoustic poles and zeros



(b) Simplification of Fig. 3a to audio and noise sources filtered through the same channel

Figure 3: Proposed model and its simplification used in this paper

All-pole models allow for linear modeling. Even though they cannot capture time-delays or anti-resonances, all-pole models are less sensitive to changes in the source / observer positions and are particularly useful for wave acoustics and high-sound frequencies [5].

A source signal, x_t, distorted by WGN with variance $\sigma^2_{n_t}$ and filtered by a P^{th} order all-pole channel (Fig. 3b) results in

$$y_t = \sum_{p \in \mathcal{P}} b_{p,t} y_{t-p} + x_t + \sigma_{n_t} n_t, \tag{3}$$

where $n_t \sim \mathcal{N}(0, 1)$, $\mathbf{y}_t = \begin{bmatrix} y_0 & \ldots & y_t \end{bmatrix}^T$ is the vector of observations, and the channel coefficients are $\mathbf{b}_t = \begin{bmatrix} b_{1,t} & \cdots & b_{P,t} \end{bmatrix}^T$. Generally, a scaling channel gain term should be included in the system after the channel in Fig. 3b. However, due to scaling ambiguities, the channel gain can be omitted from the estimation process.

As the channel parameters are assumed spatially invariant for a fixed source-sensor geometry, the channel posterior and noise prior can be expressed by the Markov chains

$$p\left(\mathbf{b} \,\middle|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}^{(-\mathbf{b})}_{0:t-1}\right) \triangleq \mathcal{N}\left(\mathbf{b} \,\middle|\, \boldsymbol{\mu}_{\mathbf{b},t-1}, \mathbf{P}_{\mathbf{b},t-1}\right), \tag{4a}$$

$$p\left(\phi_{n_t} \mid \phi_{n_{t-1}}\right) \triangleq \mathcal{N}\left(\phi_{n_t} \,\middle|\, \phi_{n_{t-1}}, \delta^2_n\right) \tag{4b}$$

where $\phi_{n_t} = \ln \sigma^2_{n_t}$, $\boldsymbol{\theta}_{0:t-1} \triangleq \left\{\mathbf{a}_{0:t-1}, \mathbf{b}_{0:t-1}, \phi_{e_{0:t-1}}, \phi_{n_{0:t-1}}\right\}$, $\boldsymbol{\theta}^{(-\mathbf{b})}_{0:t-1} \triangleq \left\{\mathbf{a}_{0:t-1}, \phi_{e_{0:t-1}}, \phi_{n_{0:t-1}}\right\}$, and initial states $p(\phi_{n_0}) \triangleq \mathcal{N}\left(\phi_{n_0} \,\middle|\, 0, \delta^2_{n_0}\right)$. $\boldsymbol{\mu}_{\mathbf{b},0}$ and $\mathbf{P}_{\mathbf{b},0}$ as well as $\left\{\delta^2_n, \delta^2_{n_0}\right\}$ are assumed constant and known. The time-varying mean, $\boldsymbol{\mu}_{\mathbf{b},t-1}$, and covariance, $\mathbf{P}_{\mathbf{b},t-1}$, of the spatially invariant channel evolve through and facilitate a recursive update rule on the channel parameters (sect. §3.3).

2.4. Conditionally Gaussian State Space

As the process and measurement noise are WGN, (1) and (3) can be expressed in conditionally Gaussian state-space (CGSS) form,

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{D}_t e_t, \tag{5a}$$

$$y_t = \mathbf{B}_t \mathbf{y}_{t-1:t-P} + \mathbf{C}_t \mathbf{x}_t + \sigma_{n_t} n_t, \tag{5b}$$

for $t > 0$ samples, where $\{e_t, n_t\} \sim \mathcal{N}(0, 1)$ and where

$$\mathbf{A}_t \triangleq \begin{bmatrix} & \mathbf{a}_t^T \\ \mathbf{I}_{Q-1} & \mathbf{0}_{Q-1\times 1} \end{bmatrix} \text{ and } \mathbf{D}_t \triangleq \begin{bmatrix} \sigma_{e_t} \\ \mathbf{0}_{Q-1\times 1} \end{bmatrix}$$

are the source transition and process noise model respectively, $\mathbf{B}_t \triangleq \mathbf{b}_t^T$ contains the channel parameters, $\mathbf{C}_t \triangleq \begin{bmatrix} 1 & \mathbf{0}_{1\times Q-1} \end{bmatrix}$ denotes the observation model, and $\mathbf{y}_{t-1:t-P} = \begin{bmatrix} \mathbf{y}_{t-1} & \cdots & \mathbf{y}_{t-P} \end{bmatrix}^T$ contains the P previous observations.

3. METHODOLOGY

The aim is to reconstruct the source signal, $\mathbf{x}_{0:t}$, and the set of parameters, $\boldsymbol{\theta}_{0:t}$, given only the distorted signal, $\mathbf{y}_{1:t}$. This can be achieved by sampling from the posterior distribution of the source signal and unknown parameters. Since the source signal is dependent on the parameters, whereas the parameters are independent of the signal, the joint posterior can be written as

$$p\left(\mathbf{x}_{0:t},\, \boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_{0:t} \mid \boldsymbol{\theta}_{0:t},\, \mathbf{y}_{1:t}\right) p\left(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t}\right).$$

As the joint posterior often cannot be sampled from directly, estimates of the source signal and model parameters can thus rather be obtained by drawing samples from their posterior distributions separately. As the system described in sect. §2.4 is a CGSS, the likelihood of the clean signal, $p\left(\mathbf{x}_{0:t} \mid \boldsymbol{\theta}_{0:t},\, \mathbf{y}_{1:t}\right)$, can be estimated by the Kalman filter assuming known parameters [1,2] (see sect. §3.1).

3.1. Direct source signal estimation through Kalman filtering

The derivations of the Kalman filter equations for the system in sect. §2.4 are similar to the standard Kalman filter (KF) for systems distorted by WGN, see e.g. Ristic *et al.* [6]. Following this approach, the Kalman states (or source signal estimate) and their error covariance after prediction and correction are

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{A}_t \boldsymbol{\mu}_{t-1|t-1} \qquad \text{(state prediction)} \tag{6a}$$

$$\mathbf{P}_{t|t-1} = \mathbf{D}_t \mathbf{D}_t^T + \mathbf{A}_t \mathbf{P}_{t-1|t-1} \mathbf{A}_t^T \tag{6b}$$

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(y_t - y_{t|t-1}\right) \qquad \text{(state correction)} \tag{6c}$$

$$\mathbf{P}_{t|t} = \left(\mathbf{I}_Q - \mathbf{K}_t \mathbf{C}_t\right) \mathbf{P}_{t|t-1}. \tag{6d}$$

The optimal Kalman gain, $\mathbf{K}_t$, and the measurement residual covariance, $\sigma_{z_t}^2$, are

$$\mathbf{K}_t = \frac{1}{\sigma_{z_t}^2} \mathbf{P}_{t|t-1} \mathbf{C}_t^T, \text{ with } \sigma_{z_t}^2 = \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t^T + \sigma_{n_t}^2, \tag{7}$$

with $y_{t|t-1} = \mathbf{B}_t \mathbf{y}_{t-1:t-P} + \mathbf{C}_t \boldsymbol{\mu}_{t|t-1}$, such that

$$p\left(y_t \mid \mathbf{y}_{1:t-1},\, \boldsymbol{\theta}_{0:t}\right) = \mathcal{N}\left(y_t \,\middle|\, y_{t|t-1},\, \sigma_{z_t}^2\right). \tag{8}$$

Thus, the standard KF [6] and that in eqns. (6) differ solely in the corrected Kalman states, $\boldsymbol{\mu}_{t|t}$, and eqn. (8).

By nature of the blind deconvolution problem, the set of parameters is unknown and direct application of the KF leads to poor estimates. Instead, the KF is incorporated within a sequential framework where at each time step, eqns. (6) are evaluated using the current parameter estimates. The parameter estimates are corrected using the results of eqns. (6).

3.2. Signal and parameter estimation using SMC methods

This sequence of steps for direct source signal and parameter estimation can be realized in a SMC framework. As $p\left(\mathbf{x}_{0:t} \mid \boldsymbol{\theta}_{0:t},\, \mathbf{y}_{1:t}\right)$ can be estimated using the KF, estimation of the joint posterior reduces to the estimation of $p\left(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t}\right)$. For sequential importance resampling (SIR) particle filters, the parameters are estimated by sampling a large number of random variates (or particles) from a hypothesis distribution approximating $p\left(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t}\right)$. The KF is bombarded with these particles. The source signal estimate hence corresponds to the mean of the state estimates, $\boldsymbol{\mu}_{t|t}$, over all particles. The particles are corrected by resampling according to a function of the observation likelihood. Resampling also ensures that only statistically significant particles are retained. The parameter estimates correspond to the mean of the particle swarm per parameter.

In this paper, prior importance sampling is utilized, i.e., the particles are drawn from the priors in eqns. (2a), (2b) and (4b). Thus, the importance weights reduce to eqn. (8) [1].

3.3. Channel estimation using Bayesian channel updates

Particle filters assume time-varying parameters and implicitly impose a dynamic on time-invariant coefficients. Thus, particle filters perform poorly for importance sampling of the spatially invariant channel parameters. Instead, these are estimated using Bayesian updates. Using Bayes's theorem, the channel posterior is,

$$p\left(\mathbf{b} \mid \mathbf{y}_{1:t},\, \boldsymbol{\theta}_{0:t}^{(-\mathbf{b})}\right) \propto p\left(y_t \mid \mathbf{y}_{1:t-1},\, \boldsymbol{\theta}_{0:t}\right) p\left(\mathbf{b} \mid \mathbf{y}_{1:t-1},\, \boldsymbol{\theta}_{0:t-1}^{(-\mathbf{b})}\right)$$

where eqn. (4a) is a prior in this context. The evidence term in the denominator is independent of the channel parameters, and thus an omittable scaling factor. Using eqn. (8), the channel posterior is

$$p\left(\mathbf{b} \mid \mathbf{y}_{1:t},\, \boldsymbol{\theta}_{0:t}^{(-\mathbf{b})}\right) \propto \mathcal{N}\left(\mathbf{b} \,\middle|\, \boldsymbol{\mu}_{\mathbf{b},t},\, \mathbf{P}_{\mathbf{b},t}\right), \tag{9}$$

with covariance and mean

$$\mathbf{P}_{\mathbf{b},t} = \left(\mathbf{P}_{\mathbf{b},t-1}^{-1} + \frac{1}{\sigma_{z_t}^2} \mathbf{y}_{t-1:t-P} \mathbf{y}_{t-1:t-P}^T\right)^{-1} \tag{10a}$$

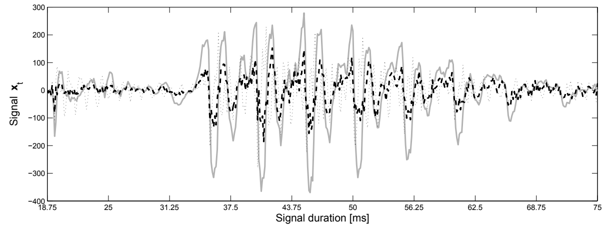
$$\boldsymbol{\mu}_{\mathbf{b},t} = \mathbf{P}_{\mathbf{b},t} \Big(y_t \frac{1}{\sigma_{z_t}^2} \mathbf{y}_{t-1:t-P} + \mathbf{P}_{\mathbf{b},t-1}^{-1} \boldsymbol{\mu}_{\mathbf{b},t-1} - \frac{1}{\sigma_{z_t}^2} \mathbf{y}_{t-1:t-P} \mathbf{C}_t \boldsymbol{\mu}_{t|t-1} \Big). \tag{10b}$$

Given the posterior, the maximum *a posteriori* (MAP) estimate of the channel is evaluated to be used for the KF correction (eqn. (6c)) and evaluation of the weights (eqn. (8)). Since the channel posterior is Gaussian, its maximum is located at the mean. Thus, the MAP estimate of the channel parameters is $\mathbf{b}_{\text{MAP}} = \boldsymbol{\mu}_{\mathbf{b},t}$.

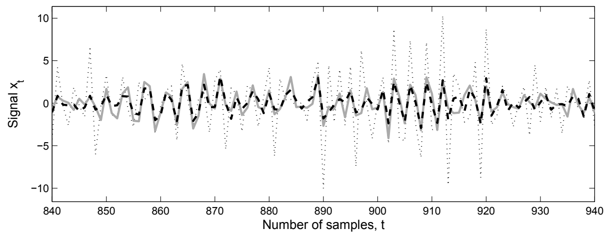
At $t+1$, the previous posterior, $p\left(\mathbf{b} \mid \mathbf{y}_{1:t},\, \boldsymbol{\theta}_{0:t}^{(-\mathbf{b})}\right)$, is used as the prior to compute the current posterior, leading to a sequential Bayesian update procedure of the channel parameters. The complete particle filter is summarized in Algorithm 1.

```
for t = 1, ..., number of samples do
    for i = 1, ..., number of particles do
        Sample a proposal of θ_t^(-b) from (2a), (2b), (4b);
        Prediction step of KF: (6a), (6b);
        Evaluation of K_t, σ²_{z_t}: (7);
        Bayesian update of channel parameters: (10);
        MAP estimation of channel: b_MAP = μ_{b,t};
        Evaluation of importance weights with b_MAP;
        Correction step of KF: (6c), (6d) ;
    end
    Normalization of importance weights;
    Resampling step (see, e.g., [7]);
end
```

Algorithm 1: SIR particle filter for reverberant system



(a) Speech signal segment



(b) Synthetic data

Figure 4: Accuracy of source signal estimate (- - -) vs. the actual source signal (—) and observed signal (·········)

$\delta^2_{e_0}$	$\delta^2_{n_0}$	δ^2_e	δ^2_n	$\mathbf{\Delta}_{a_0}$	$\mathbf{\Delta}_a$
0.5	0.5	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$0.5\mathbf{I}_Q$	$5 \cdot 10^{-4}\mathbf{I}_Q$

Table 1: Markov parameters for synthesis and estimation

4. EXPERIMENTAL RESULTS

A 4^{th} order synthetic source signal is filtered through an 8^{th} order channel according to Fig. 3b. The signal-based measure (SBM)[3] of the distorted signal is -6.15dB. As the results are insensitive to the specific Markov parameters [1], these are set to the values used for data synthesis (Table 1). These values correspond to the Markov parameters chosen in [1], selected heuristically to ensure sufficiently broad sampling of the sampling space. The particle filter is executed for 1000 samples and 800 particles and $\boldsymbol{\mu}_{\mathbf{b},0} = 0.5 \times \mathbf{1}_{P\times 1}$, $\mathbf{P}_{\mathbf{b},0} = \mathbf{0}_Q$. Even though the source parameter estimates appear inaccurate (Fig. 6b), the SBM of the enhanced signal is 4.42dB, an improvement of 10.57dB. The accuracy of the estimated signal compared to the clean source signal and the observed signal is shown in Fig. 4b. The evolution of the poles with time of the MAP estimates of the stationary channel parameters are shown in Fig. 5. After few iterations, the estimates converge towards the actual channel poles. Likewise, the channel parameters converge after ca. 200 samples to the actual coefficients (Fig. 6a).

The particle filter is then run for 1000 particles on an utterance of the words "the farmer's" by a female American speaker for 9372 samples with sampling frequency 8kHz distorted by the same 8^{th} order channel as above. 15 source parameters are estimated. The SBM of the observed signal is -1.93dB, that of the estimated source signal is 1.68dB, an improvement of 3.61dB. The accuracy of the estimated signal[4] is shown in Fig. 4a. Although additional musical noise is introduced, the intelligibility of the estimated source signal is clearly improved over the distorted signal. The effect of musical noise could be reduced and intelligibility further improved by a fixed-lag smoothing Markov chain Monte Carlo (MCMC) step [1].

[3] $\mathrm{SBM_{dB}} = 10\log_{10}\left(\frac{\|\mathbf{x}_{0:t-1}\|^2}{\|\tilde{\mathbf{x}}_{0:t-1} - \mathbf{x}_{0:t-1}\|^2}\right)$, where $\tilde{\mathbf{x}}_{0:t-1}$ is either the estimated or the distorted signal sequence.

[4] Corresponding audio files online at `http://www.see.ed.ac.uk/~s0565868/Conferences/2008-04-ICASSP/`.

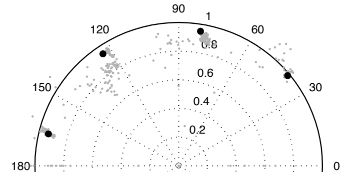


Figure 5: Convergence of estimated (●) to actual channel poles (●)

5. CONCLUSIONS

This paper proposed a model where noise as well as the audio source are subject to reverberation by means of spatially distinct AIRs separated into source-sensor specific and common acoustical resonances. For enhancement, a particle filter with a Bayesian update procedure of the channel parameters was proposed. A simplified version of the model was applied to speech and synthetic data. Results verified the effectiveness of the investigated model and method for speech.

Future research will investigate the applicability of the complete proposed model in Fig. 3a for high-order channels. Further, it will be determined if the discrepancy of the source parameter estimates can be alleviated or if this is due to identifiability issues of the model.

6. REFERENCES

[1] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Particle methods for Bayesian modeling and enhancement of speech signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 3, pp. 173–185, Mar. 2002.

[2] M. J. Daly, J. P. Reilly, and J. H. Manton, "A Bayesian approach to blind source recovery," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, vol. 1, Nov. 2004, pp. 989–993.

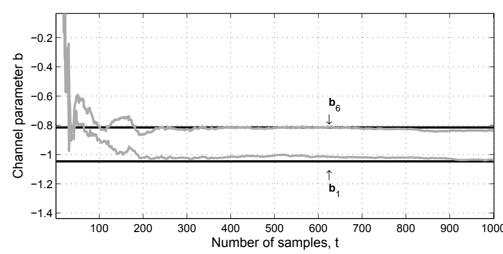
[3] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, Apr. 1994.

[4] J. J. Rajan, P. J. W. Rayner, and S. J. Godsill, "Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *IEE Proc.-Vis. Image Signal Process.*, vol. 144, no. 4, pp. 249–256, August 1997.

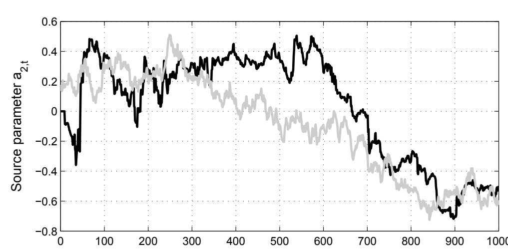
[5] J. N. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vibr.*, vol. 146, no. 2, pp. 281–302, April 1991.

[6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter - Particle Filters for Tracking Applications*. Artech House, 2004.

[7] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo methods in practice*. New York: Springer Verlag, 2000.



(a) Convergence of channel estimate to actual parameters $\mathbf{b}_{\{1,6\}}$



(b) Discrepancy between estimate and actual source parameter $\mathbf{a}_{2,t}$

Figure 6: Estimated (—) and actual parameters (—)