Multichannel Online Blind Speech Dereverberation with Marginalization of Static Observation Parameters in a Rao-Blackwellized Particle Filter

Noname manuscript No.
(will be inserted by the editor)

Multichannel online blind speech dereverberation with marginalization of static observation parameters in a Rao-Blackwellized particle filter

Christine Evers · **James R. Hopgood**

Received: September 4, 2013

Abstract Room reverberation leads to reduced intelligibility of audio signals and spectral coloration of audio signals. Enhancement of acoustic signals is thus crucial for high-quality audio and scene analysis applications. Multiple sensors can be used to exploit statistical evidence from multiple observations of the same event to improve enhancement. Whilst traditional beamforming techniques suffer from interfering reverberant reflections with the beam path, other approaches to dereverberation often require at least partial knowledge of the room impulse response which is not available in practice, or rely on inverse filtering of a channel estimate to obtain a clean speech estimate, resulting in difficulties with non-minimum phase acoustic impulse responses. This paper proposes a multi-sensor approach to blind dereverberation in which both the source signal and acoustic channel are directly estimated from the distorted observations using their optimal estimators. The remaining model parameters are sampled from hypothesis distributions using a particle filter, thus facilitating real-time dereverberation. This approach was previously successfully applied to single-sensor blind dereverberation. In this paper, the single-channel approach is extended to multiple sensors. Performance improvements due to the use of multiple sensors are demonstrated on synthetic and real speech examples.

Keywords Blind dereverberation · Multi-sensor processing · Speech enhancement · Kalman filter · Particle filter · Rao-Blackwellization · Bayesian estimation

1 Introduction

Audio signals in confined spaces exhibit reverberation due to reflections off surrounding obstacles. In addition to the direct path signal, time-shifted reflections are received, leading to reduced intelligibility. Due to the length of the different paths of propagation and different

C. Evers acknowledges the support from the Scottish Funding Council of her position within the Joint Research Institute in Signal & Image Processing with the University of Edinburgh which is a part of the Edinburgh Research Partnership in Engineering & Mathematics (ERPem)

C. Evers · J. R. Hopgood
Institute for Digital Communications, Joint Research Institute for Signal & Image Processing, School of Engineering and Electronics, The University of Edinburgh, King's Buildings, Edinburgh, EH9 3JL, UK
E-mail: c.evers@ieee.org

amounts of energy being absorbed by reflecting obstacles, each wavefront arrives with a different phase and amplitude at the receiver. Reverberation causes echoey effects, spectral coloration, and leads to articulation loss of consonants [1] and masking effects of phonemes [2] due to distortion of sound onsets and decays [3]. The distorting effects of reverberation are especially pertinent for high quality audio applications, e.g., automatic speech recognition, hearing aids, or scene analysis applications (source localization, tracking, or identification).

Thus, an important engineering problem is the blind enhancement of the reverberant signal from the observed signal in order to obtain a clean speech estimate. Signal enhancement in reverberant environments can be considered as a blind deconvolution problem and can be performed using a single microphone [4–8]. Estimates can be improved using multiple sensors in order to exploit spatial diversity and statistical evidence from multiple observations of the same event. Blind dereverberation could be further improved with accurate channel modeling which depends on knowledge of the target position. However, many passive target tracking methods suffer from the presence of reverberation leading to substantial tracking errors. Therefore, enhancement could be improved through a joint enhancement and tracking algorithm. As single microphones cannot exploit spatial diversity of the received signal required for tracking, multiple microphones would be necessary in such a framework. This paper thus proposes a parametric multi-sensor blind dereverberation approach for stationary speakers that will be extended in future research to incorporate joint tracking for moving speakers.

The proposed blind dereverberation algorithm utilizes the optimal estimators for the source signal and room impulse response (RIR) by exploiting sub-structures of the underlying state-space. The source signal and acoustic channel can thus be estimated directly from the observations using modified versions of the Kalman filter (KF). Thus, issues due to channel inversion often used to reconstruct the source signal are avoided, and real-time speech enhancement is facilitated due to the recursive nature of the KF. The Kalman recursions rely on knowledge of any underlying model parameters. However, as the source signal and channel need to be estimated blindly from the distorted measurements, the model parameters are unknown in practice. Instead, an *ensemble* of KFs is evaluated for stochastically selected parameters sampled from hypothesis distributions using a particle filter (PF) framework.

As analytically tractable substructures of the system are obtained using their optimal estimator, the variance of the estimates is decreased [9]. Rao-Blackwellized PFs marginalizing the source signal by means of KF estimation from the remaining unknowns are well known in the literature (see, e.g., [10, 11]). The novelty of the approach proposed in this paper thus lies in the marginalization of the *channel*. As a consequence, particle impoverishment is avoided that arises when sampling static parameters using PFs due to their implicit enforcement of a dynamic upon the estimated variables. The proposed algorithm will be referred to as the **Ma**rginalized **R**ao-**Bl**ackw**e**llized (MARBLE) PF. [8] shows how this approach can be extended to a multitude of other signal processing applications.

Sect. §2 gives a brief overview of multi-sensor dereverberation approaches in the literature and how the MARBLE PF differs and improves upon these methods. Sect. §3 presents the models used for the speech source and acoustical channel. Sects. §4 and §5 discuss the methodology and derive the necessary marginalizations. Results for synthetic and speech data are demonstrated in sect. §6, a multirate extension for reduced model orders is discussed in sect. §7, and conclusions are drawn in sect. §8.

2 Multi-sensor blind dereverberation

Incoming plane waves from spatially distinct sources arrive at a sensor array with slight time delays. By introducing suitable delays to each channel, beamformers [12] enforce that the source signal arrives coherently at each sensor. By adding all sensor outputs, the source signal is thus amplified whilst incoherent interference is attenuated. However, reverberant reflections arrive from multiple different positions in the room and are likely to interfere with the beam path. Hence, traditional beamforming approaches only remove reverberation to some degree. Affes and Grenier [13] use a matched filter beamformer that adaptively estimates the channel response and convolve the received signals with the inverse of the resulting RIR. Flanagan *et al.* [14], use three dimensional microphone arrays to steer beams in the direction of strong initial reflections. However, both [13] and [14] require at least partial knowledge of the RIR.

Instead of utilizing beamformers, Allen [15] estimates the parameters of an infinite impulse response (IIR) filter approximating the vocal tract using linear prediction analysis to synthesize a signal that approximates the clean speech signal. This approach can be extended to explicitly use specific speech models and discriminate between reverberation and speech based on impulse clustering of the reverberant signal [16–18]. However, although impulses due to reverberation are reduced, natural speech components are neglected due to signal *synthesis*. Nakatani *et al.* [19–21] synthesize an enhanced speech signal in order to iteratively estimate a filter approximating the reverberant channel based on the reverberant and synthetic signal. The clean speech signal is estimated by inverse filtering the reverberant signal with the channel estimate. However, RIRs are often non-minimum phase [22], leading to difficulties with channel inversion. Furthermore, a high number of channel parameters is necessary in order to accurately reflect the RIR. Habets [23, 24] reconstructs the speech signal by spectral subtraction of the estimated power spectral density (PSD) of the RIR from the received signal. However, again, prior information about the RIR is required.

The MARBLE PF circumvents the issues encountered in channel inversion, spectral subtraction, and linear predictive coding (LPC) analysis approaches [25] by i) *direct source signal estimation*, i.e., neither speech synthesis nor channel inversion are necessary to reconstruct the clean speech signal, thus circumventing non-minimum phase problems or scaling of errors; ii) *Sequential processing* facilitating real-time speech enhancement; iii) *Blind channel estimation*, i.e., no prior knowledge of the RIR is necessary, and iv) *marginalization of the static channel* considering uncertainty introduced through channel estimation. Furthermore, as described in sect. §7, *multi-rate filtering* can be straightforwardly applied such that fewer channel parameters are necessary in each sub-band to approximate the RIR.

3 System model

Data recorded in realistic environments is often hard to analyze and interpret, especially if the functional relationship of factors influencing the environment is non-linear and/or non-Gaussian. Mathematical models can be used to represent essential aspects of a system in usable form. Although models are, by definition, never entirely accurate, statistical properties of the underlying model are exploited to provide a better understanding of the data.

Parametric models associate each entity of the system with a model characterized by a finite set of parameters that control its properties. The model is fitted to the data by estimating the parameters according to some criterion [26]. A signal estimate can be obtained by applying the parameters and model to a random excitation. Given the parameters of the

models, the characteristics of the system can be described and differences between the measurements and the model can often be detected.

3.1 Source model

Models of speech production systems describe how *unstructured* airflow pressed out of the lungs is *structured* by a system producing speech sounds by means of the vocal tract. Parametric speech models are thus based on modeling the human vocal tract and production of human sound.

Kelly and Lochbaum [27] proposed that the human vocal tract can be modeled as slowly time-varying circular one-dimensional acoustic tube. An extended model [28] assumes that the vocal tract can be represented by a concatenation of lossless acoustic tubes, where the constant cross-sectional areas of the individual tubes approximate the vocal tract.

Autoregressive (AR) processes can be interpreted as simplified models of lossless acoustics tubes [28, 29]. The acoustic tube is excited by either periodic glottal pulse waveforms to produce voiced speech, or by turbulent noise to produce unvoiced speech (Fig. 1). This paper uses a model for unvoiced speech, thus focusing on the excitation by turbulent noise. Local correlations in the signal are exploited by linearly combining past samples, i.e.,

$$x_t = \sum_{q \in \mathscr{Q}} a_q \, x_{t-q} + \sigma_v \, v_t, \qquad v_t \sim \mathscr{N}(0,1) \tag{1}$$

where Q is the model order, $\mathbf{x}_t = \begin{bmatrix} x_0 & \cdots & x_t \end{bmatrix}^T$ are the signal samples up to time t, the AR coefficients are given by $\mathbf{a} = \{a_q\}_{q \in \mathscr{Q}}$, and σ_v^2 is the covariance of the excitation, v_t. AR processes in contrast to autoregressive moving average processes cannot capture anti-resonances – or time-delays – represented by zeros. Hence, certain sounds such as French nasals [30] cannot be modeled accurately. Nonetheless, the inclusion of zeros in the model requires the solution of a non-linear set of equations, whereas AR processes allow for linear analysis. The popularity of AR processes thus mainly stems from their simplicity and analytical tractability [31].

Many estimation methods impose time-invariance on the signal model primarily to constructively exploit ergodicity. However, the vocal tract is continually changing with time and thus the limitation of stationarity results in poor modeling for speech signals. Hence, the variation of speech parameter should be modeled as a *non-stationary* process.

The local time-variation of speech signals can be captured using time-varying AR (TVAR) processes. The signal can be represented by a Q^{th} order TVAR process as

$$x_t = \sum_{q \in \mathscr{Q}} a_{q,t} \, x_{t-q} + \sigma_{v_t} \, v_t, \qquad v_t \sim \mathscr{N}(0,1) \tag{2}$$

where $\mathbf{a}_t = \left\{ a_{q,t} \right\}_{q \in \mathscr{Q}}$ are the TVAR coefficients. The time-varying characteristics of $\mathbf{x}_t$ are thus described by the parameters. The time-varying source parameters and excitation sequence can be modeled as a stochastic process specified by the first-order Markov chain [10],

$$\left.\begin{aligned} a_{q,t} &= a_{q,t-1} + \sigma_{a_{q,t}} \, r_{a_t} \\ \phi_{v_t} &= \phi_{v_{t-1}} + \sigma_{\phi_{v_t}} \, r_{\phi_{v_t}} \end{aligned}\right\} \left\{ r_{a_t}, r_{\phi_{v_t}} \right\} \sim \mathscr{N}(0,1) \tag{3}$$

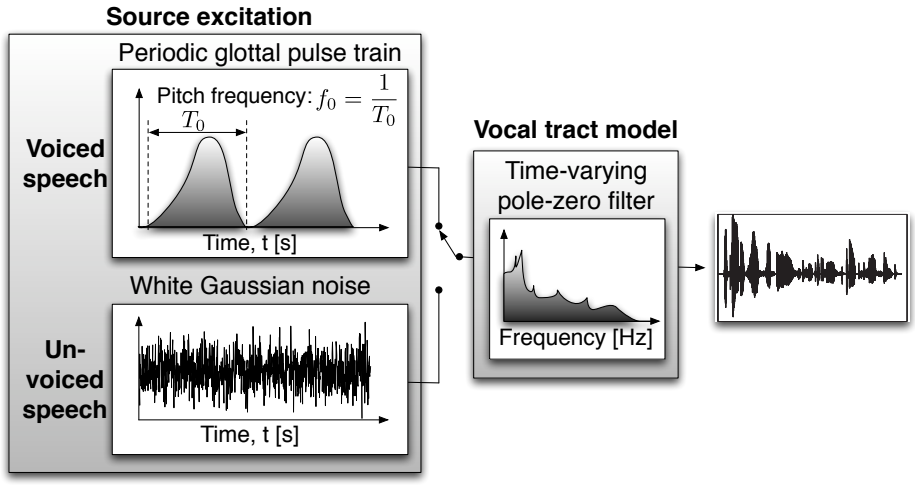


Fig. 1: Speech production model

where $\phi_{v_t} \triangleq \ln \sigma^2_{v_t}$,[1] and $\sigma^2_{a_{q,t}}$ and $\sigma^2_{\phi_{v_t}}$ are the variance terms on the random walks on the source parameters and logarithm of the source excitation. Since the source parameters are of stochastic nature, sequential estimation frameworks are particularly apt at tracking the unknown random variables, x_t, $a_{q,t}$, and ϕ_{v_t}. Stability constraints can be enforced by applying the indicator function, $\mathbb{I}_{\mathscr{A}_Q}(\mathbf{a}_t)$, over the region of support, $\mathscr{A}_Q$, of the source parameters. The indicator function accepts $\mathbf{a}_t$ if the roots, $\mathbf{p}_t$, of $\mathbf{a}_t$ lie within the unit circle. Otherwise, the roots are reflected back into the unit circle by setting $\mathbf{p}_t \rightarrow {}^1/_{\mathbf{p}_t}$ [32]. The probability density functions (pdfs) of eqn. (3) are thus given by

$$p\left(\mathbf{a}_t \,|\, \mathbf{a}_{t-1}\right) \propto \mathscr{N}\left(\mathbf{a}_t \,\middle|\, \mathbf{a}_{t-1}, \boldsymbol{\Sigma}_{\mathbf{a}_t}\right) \mathbb{I}_{\mathscr{A}_Q}(\mathbf{a}_t) \tag{4a}$$

$$p\left(\phi_{v_t} \,|\, \phi_{v_{t-1}}\right) = \mathscr{N}\left(\phi_{v_t} \,\middle|\, \phi_{v_{t-1}}, \sigma^2_{\phi_{v_t}}\right) \tag{4b}$$

where the covariances $\boldsymbol{\Sigma}_{\mathbf{a}_t}$, $\sigma^2_{\phi_{v_t}}$ are assumed known.

3.2 Channel model

Many different techniques for modeling an room impulse response (RIR) exist. In general, each model applies to a different frequency range of the audible spectrum. A complete characterization of the acoustic impulse response (AIR) can be obtained by a parametric model through the solution of the acoustic wave equation in terms of a linear combination of damped harmonics,

$$h_t = \begin{cases} 0 & \text{for } t < 0 \\ \sum_n A_n e^{-\delta_n t} \cos\left(\omega_n t + \theta_n\right) & \text{for } t \geq 0 \end{cases} \tag{5}$$

[1] By definition, variance terms are bound between $0 \leq \sigma^2 \leq \infty$. Sampling from $\ln \sigma^2_{v_t}$ reinforces this constraint.

where the coefficients A_n implicitly contain the location of the source and observer, δ_n, ω_n, and θ_n are the damping constant, undamped natural frequency, and phase terms respectively. However, this model is intractable for many estimation problems and does not lead to an analytical expression in the Bayesian framework for blind dereverberation.

The solution of the acoustic wave equation does however indicate that a room transfer function can be expressed by a rational expression, and therefore can be modelled by a conventional pole-zero model [33]. From a physical point of view, poles represent resonances, and zeros represent time delays and anti-resonances. Another commonly used model is the all-zero model. There are several main limitations of finite impulse response (FIR) filters imposed by the nature of room acoustics [33, 34]. First, AIRs are, in general, very long and an all-zero filter typically requires $n_s = T_{60}\, f_s$ coefficients where f_s is the sampling frequency. For example, if $T_{60} = 0.5$ seconds and $f_s = 10$ kHz, the all-zero filter requires $n_s = 5000$ coefficients. Secondly, the resulting FIR filter may be effective only for a very limited spatial combination of source and receiver positions, as all-zero models lead to large variations in the room transfer function (RTF) for small changes in source–observer positions [33, 34].

As an alternative, all-pole models for approximating rational transfer functions are widely used. Typical all-pole model orders required for approximating RTFs are in the range $50 < P < 500$ [33] depending on the frequency range of the acoustic spectrum considered. A significant advantage of the all-pole model over the all-zero model is its lower sensitivity to changes in source and observer positions. In many signal processing applications dealing with room acoustics, it is thus sufficient and more efficient to manipulate all-pole models rather than high-order all-zero models.

A source signal distorted by white Gaussian noise with variance $\sigma^2_{w_t}$ and filtered through an all-pole channel of order P is observed at the m^{th} sensor, $m \in \mathscr{M}$, as

$$y_{m,t} = \sum_{p \in \mathscr{P}} b_{m,p} y_{t-p,m} + x_t + \sigma_{w_{m,t}}\, w_{m,t}, \tag{6}$$

where $w_{m,t} \sim \mathscr{N}(0,\, 1)$ and the channel coefficients are $\mathbf{b}_m = \left\{ b_{m,p} \right\}_{p \in \mathscr{P}}$. Similar to eqn. (3), the logarithm of the measurement noise variance, $\phi_{w_{m,t}} \triangleq \ln \sigma^2_{w_{m,t}}$, is assumed to vary according to a first-order Markov chain, i.e.,

$$\phi_{w_{m,t}} = \phi_{w_{m,t-1}} + \sigma_{\phi_{w_{m,t}}}\, r_{\phi_{w_{m,t}}} \tag{7a}$$

$$p\left(\phi_{w_{m,t}} \,\middle|\, \phi_{w_{m,t-1}} \right) = \mathscr{N}\left(\phi_{w_{m,t}} \,\middle|\, \phi_{w_{m,t-1}},\, \sigma^2_{\phi_{w_{m,t}}} \right), \tag{7b}$$

where $\sigma^2_{\phi_{w_{m,t}}}$ is assumed known and constant.

3.3 CGSS model

The source model in eqn. (2) and the measurement model in eqn. (6) can be easily rewritten in state space form. Conditionally on the model parameters the signal model is linear, leading to the conditionally Gaussian state-space (CGSS) representation [10],

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{\Sigma}_{\mathbf{v}_t} \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathscr{N}\left(\mathbf{0}_{Q\times 1}, \mathbf{I}_Q \right), \tag{8a}$$

$$\mathbf{y}_t = \mathbf{Y}_{t-1}\mathbf{b} + \mathbf{C}^T \mathbf{x}_t + \mathbf{\Sigma}_{\mathbf{w}_t} \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathscr{N}\left(\mathbf{0}_{M\times 1}, \mathbf{I}_M \right) \tag{8b}$$

where $\mathbf{y}_t \triangleq \begin{bmatrix} y_{t,1} & \dots & y_{t,M} \end{bmatrix}^T$ and $\mathbf{b} \triangleq \begin{bmatrix} \mathbf{b}_1^T & \dots & \mathbf{b}_M^T \end{bmatrix}^T$. Furthermore, $\mathbf{C}^T \triangleq \mathbf{1}_{M\times 1} \begin{bmatrix} 1 & \mathbf{0}_{1\times Q-1} \end{bmatrix}$, and $\boldsymbol{\Sigma}_{\mathbf{v}_t}^T \triangleq \begin{bmatrix} \sigma_{v_t} & \mathbf{0}_{1\times Q-1} \end{bmatrix}$, and

$$\mathbf{A}_t \triangleq \begin{bmatrix} & \mathbf{a}_t^T \\ \mathbf{I}_{Q-1} & \mathbf{0}_{Q-1\times 1} \end{bmatrix} \qquad \mathbf{Y}_{t-1} \triangleq \begin{bmatrix} \widehat{\mathbf{y}}_{t-1,1}^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \widehat{\mathbf{y}}_{t-1,M}^T \end{bmatrix}$$

where $\widehat{\mathbf{y}}_{t-1,i} \triangleq \begin{bmatrix} y_{t-1,i} & \dots & y_{t-P,i} \end{bmatrix}^T$ and $\mathbf{a}_t \triangleq \{a_{m,t}\}_{m\in\mathscr{M}}$.

4 Marginalizing channel parameters from source signal

Given the stochastic model in eqn. (8), an optimal estimator is sought of the clean speech signal, $\mathbf{x}_{0:t}$. If all system variables are considered as stochastic entities, an estimate of the source signal can be obtained by maximizing its posterior pdf, $p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t})$, where $\boldsymbol{\theta}_{0:t} \triangleq \left\{ \mathbf{a}_{0:t}, \boldsymbol{\phi}_{v_{0:t}}, \boldsymbol{\phi}_{w_{0:t}} \right\}$ are defined as the time-varying model parameters and assumed known in this section. However, according to eqn. (8), the posterior pdf of the source signal is dependent upon the parameters of the RIR, $\mathbf{b}$, which are unknown in practice. In order to estimate $\mathbf{x}_{0:t}$, an estimate of $\mathbf{b}$ is thus required, i.e.,

$$p(\mathbf{z}_{0:t} \mid \boldsymbol{\psi}_{0:t}) = p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t}, \mathbf{b})\, p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t}) \tag{9}$$

where $\mathbf{z}_{0:t} \triangleq \{\mathbf{x}_{0:t}, \mathbf{b}\}$ and $\boldsymbol{\psi}_{0:t} \triangleq \{\mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}\}$. The mean squared error (MSE) between $\mathbf{z}_{0:t}$ and its estimate, $\widehat{\mathbf{z}}_{0:t}$, can thus be expressed as

$$\mathrm{MSE}_{\widehat{\mathbf{z}}_{0:t}} = \int \|\widehat{\mathbf{z}}_{0:t} - \mathbf{z}_{0:t}\|^2\, p(\mathbf{z}_{0:t} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{z}_{0:t}. \tag{10}$$

Differentiating eqn. (10) with respect to $\widehat{\mathbf{z}}_{0:t}$ and setting to zero, the minimum mean-square error (MMSE) estimate is

$$\widehat{\mathbf{z}}_{0:t} = \int \mathbf{z}_{0:t}\, p(\mathbf{z}_{0:t} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{z}_{0:t} \tag{11a}$$

which, by inserting eqn. (9), is equivalent to

$$\begin{aligned} &= \iint \begin{bmatrix} \mathbf{x}_{0:t} \\ \mathbf{b} \end{bmatrix} p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t}, \mathbf{b})\, p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{b}\, d\mathbf{x}_{0:t} \\ &= \begin{bmatrix} \int \mathbf{x}_{0:t} \int p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t}, \mathbf{b})\, p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{b} d\mathbf{x}_{0:t} \\ \int \mathbf{b}\, p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t}) \int p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t}, \mathbf{b})\, d\mathbf{x}_{0:t}\, d\mathbf{b} \end{bmatrix} \\ &= \begin{bmatrix} \int \mathbf{x}_{0:t}\, p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{x}_{0:t} \\ \int \mathbf{b}\, p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t})\, d\mathbf{b} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{x}}_{0:t} \\ \widehat{\mathbf{b}} \end{bmatrix} \end{aligned} \tag{11b}$$

Thus, the optimal estimator of the source signal maximizes the *marginalized* posterior pdf, $p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t},)$. The posterior pdf, $p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t}, \mathbf{b})$, can be expressed in terms of the KF equations due to the linear, Gaussian substructure in eqn. (8) as discussed in sect. §4.1. Likewise, as derived in sect. §4.2, $p(\mathbf{b} \mid \boldsymbol{\psi}_{0:t})$ can be updated using the KF. Finally, sect. §4.3 shows that $\mathbf{x}_{0:t}$ is *linearly dependent* in $\mathbf{b}$ such that $p(\mathbf{x}_{0:t} \mid \boldsymbol{\psi}_{0:t})$ can be derived straightforwardly from the results in sects. §4.1 and §4.2.

4.1 Estimation of source signal

The KF is the optimal estimator of the source signal for known model parameters in CGSS systems (see eqn. (8)).

4.1.1 Marginalization from posterior pdf

KFs sequentially predict $\mathbf{x}_{0:t}$ based on the model parameters and correct the prediction using the most recent measurement. Thus, the posterior,

$$p(\mathbf{x}_{0:t} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) \propto \prod_{k=1}^{t} p(\mathbf{x}_k \mid \mathbf{y}_{1:k}, \mathbf{x}_{0:k-1}, \boldsymbol{\theta}_{0:k}, \mathbf{b}) \tag{12}$$

is to be estimated. Assuming that the posterior pdf at $t-1$, $p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b})$ is available, the source signal sample at t can be predicted using $\mathbf{y}_{1:t-1}$ by marginalizing the previous states, $\mathbf{x}_{t-1}$, i.e.,

$$\begin{aligned} p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, d\mathbf{x}_{t-1} \\ &= \int p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{x}_{t-1}, \mathbf{b})\, p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, d\mathbf{x}_{t-1} \end{aligned}$$

Recalling eqn. (8a), given $\mathbf{x}_{t-1}$, $\mathbf{x}_t$ only depends on $\mathbf{x}_{t-1}$ and $\boldsymbol{\theta}_{0:t}$, then

$$= \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \boldsymbol{\theta}_{0:t})\, p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b})\, d\mathbf{x}_{t-1}. \tag{13}$$

When the measurement at t becomes available, the prediction $p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$ can be updated using the current observations via Bayes's theorem to obtain the corrected pdf, $p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$, i.e.,

$$\begin{aligned} &p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) \\ &\quad = \frac{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \mathbf{x}_t, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})}{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})}. \end{aligned} \tag{14}$$

where $p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \mathbf{x}_t, \boldsymbol{\theta}_{0:t}, \mathbf{b})$, is given by probability transformation of eqn. (8), i.e.,

$$p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \mathbf{x}_t, \boldsymbol{\theta}_{0:t}, \mathbf{b}) = \mathcal{N}\left(\mathbf{y}_t \,\middle|\, \mathbf{Y}_{t-1}\mathbf{b} + \mathbf{C}^T\mathbf{x}_t,\, \boldsymbol{\Sigma}_{\mathbf{w}_t}\right) \tag{15}$$

The evidence term, $p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$, is independent of $\mathbf{x}_t$, thus acting as a scaling constant only.

For the linear Gaussian state space in eqn. (8), the recursive propagation in time of the posterior pdf, as described by eqns. (13) and (14), can be completely characterized by a Kalman filter, such that [10, 35],

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) = \mathcal{N}\left(\mathbf{x}_t \,\middle|\, \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right), \tag{16a}$$

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) = \mathcal{N}\left(\mathbf{x}_t \,\middle|\, \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}\right). \tag{16b}$$

The mean and covariance are obtained using the KF equations (see, e.g, [35]),

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{A}_t \boldsymbol{\mu}_{t-1|t-1}, \tag{17a}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{\Sigma}_{\mathbf{v}_t}\boldsymbol{\Sigma}_{\mathbf{v}_t}^T + \mathbf{A}_t\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}_t^T \tag{17b}$$

$$\boldsymbol{\mu}_{t|t} = \left(\mathbf{I}_Q - \mathbf{K}_t\mathbf{C}^T\right)\boldsymbol{\mu}_{t|t-1} - \mathbf{K}_t\left(\mathbf{Y}_{t-1}\mathbf{b} - \mathbf{y}_t\right) \tag{17c}$$

$$\boldsymbol{\Sigma}_{t|t} = \left(\mathbf{I}_Q - \mathbf{K}_t\mathbf{C}^T\right)\boldsymbol{\Sigma}_{t|t-1}. \tag{17d}$$

with residual covariance, $\boldsymbol{\Sigma}_{\mathbf{z}_t}$, and the Kalman gain, $\mathbf{K}_t$,

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tag{18}$$

$$\boldsymbol{\Sigma}_{\mathbf{z}_t} = \boldsymbol{\Sigma}_{w_t} + \mathbf{C}^T\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}. \tag{19}$$

As $p\left(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right)$, is Gaussian, $\boldsymbol{\mu}_{t|t}$ corresponds to both the expected value and maximum of the posterior pdf of the source signal and is thus both the maximum *a posteriori* (MAP) and MMSE estimate of $\mathbf{x}_t$. Therefore, using eqns. (13) and (14), $\mathbf{x}_{0:t}$ can be recursively estimated by i) predicting the states by marginalization of the trajectory of past states, $\mathbf{x}_{0:t-1}$, and ii) updating the estimate using y_t by applying Bayes's theorem .

Furthermore, the source signal is marginalized from the likelihood in eqn. (15) via

$$\begin{aligned} &p\left(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right) \\ &= \int p\left(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \mathbf{x}_{0:t}, \mathbf{b}\right) p\left(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right) d\mathbf{x}_t \\ &= \mathcal{N}\left(\mathbf{y}_t \,\middle|\, \mathbf{Y}_{t-1}\mathbf{b} + \mathbf{C}^T\boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{\mathbf{z}_t}\right). \end{aligned} \tag{20}$$

4.1.2 Linearity of KF in channel parameters

The posterior pdf of the source signal in eqn. (16b) is dependent on the unknown channel parameters, $\mathbf{b}$, via $\boldsymbol{\mu}_{t|t}$ in eqn. (17a). In fact, it can be shown by induction that $\boldsymbol{\mu}_{t|t}$ is *linearly dependent* in $\mathbf{b}$, such that eqn. (17a) at $t-1$ is

$$\boldsymbol{\mu}_{t-1|t-1} = \boldsymbol{\alpha}_{t-1|t-1} + \boldsymbol{\Gamma}_{t-1|t-1}\mathbf{b}. \tag{21}$$

Inserting eqn. (21) into the prediction, $\boldsymbol{\mu}_{t|t-1}$, in eqn. (17a) at t, the predicted states become

$$\boldsymbol{\mu}_{t|t-1} = \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\Gamma}_{t|t-1}\mathbf{b} \tag{22}$$

where

$$\boldsymbol{\alpha}_{t|t-1} = \mathbf{A}_t\boldsymbol{\alpha}_{t-1|t-1} \tag{23a}$$

$$\boldsymbol{\Gamma}_{t|t-1} = \mathbf{A}_t\boldsymbol{\Gamma}_{t-1|t-1} \tag{23b}$$

Inserting into eqn. (17c), the corrected states thus are *implicitly dependent* on the channel parameters via

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\alpha}_{t|t} + \boldsymbol{\Gamma}_{t|t}\mathbf{b} \tag{24}$$

where

$$\boldsymbol{\alpha}_{t|t} = \left(\mathbf{I}_Q - \mathbf{K}_t\mathbf{C}^T\right)\boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t\mathbf{y}_t \tag{25a}$$

$$\boldsymbol{\Gamma}_{t|t} = \left(\mathbf{I}_Q - \mathbf{K}_t\mathbf{C}^T\right)\boldsymbol{\Gamma}_{t|t-1} - \mathbf{K}_t\mathbf{Y}_{t-1} \tag{25b}$$

This linear dependency of $\boldsymbol{\mu}_{t|t}$ in $\mathbf{b}$ facilitates the marginalization of $\mathbf{b}$ from the posterior pdf $p\left(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right)$ as derived in sect. §4.3 in eqn. (11b).

4.2 Estimation of channel parameters

The static channel, $\mathbf{b}$, does not exhibit a dynamic over time. Predicting future values would thus be futile. Nonetheless, *belief* in the static parameters can be updated as new data becomes available. Using Bayes's theorem, this belief can be expressed as (see Appen. A)

$$\begin{aligned} &p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) \\ &\quad = \frac{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}. \end{aligned} \tag{26}$$

Assuming that $p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})$ is Gaussian with mean $\boldsymbol{\mu}_{\mathbf{b},t-1}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{b},t}$, such that

$$p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}) = \mathcal{N}\left(\mathbf{b} \,\middle|\, \boldsymbol{\mu}_{\mathbf{b},t-1}, \boldsymbol{\Sigma}_{\mathbf{b},t-1}\right), \tag{27}$$

then inserting $p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})$ and $p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$ (eqn. (20)) into eqn. (26), the posterior pdf of the channel is Gaussian itself as shown in Appen. A, i.e.,

$$p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) = \mathcal{N}\left(\mathbf{b} \,\middle|\, \boldsymbol{\mu}_{\mathbf{b},t}, \boldsymbol{\Sigma}_{\mathbf{b},t}\right) \tag{28}$$

where the mean, $\boldsymbol{\mu}_{\mathbf{b},t}$, and covariance, $\boldsymbol{\Sigma}_{\mathbf{b},t}$, are

$$\boldsymbol{\mu}_{\mathbf{b},t} = \left(\mathbf{I}_P - \mathbf{K}_{\mathbf{b},t}\tilde{\mathbf{Y}}_{t-1}^T\right)\boldsymbol{\mu}_{\mathbf{b},t-1} + \mathbf{K}_{\mathbf{b},t}\tilde{\mathbf{y}}_t \tag{29a}$$

$$\boldsymbol{\Sigma}_{\mathbf{b},t} = \left(\mathbf{I}_P - \mathbf{K}_{\mathbf{b},t}\tilde{\mathbf{Y}}_{t-1}^T\right)\boldsymbol{\Sigma}_{\mathbf{b},t-1}, \tag{29b}$$

where the Kalman gain, $\mathbf{K}_{\mathbf{b},t}$, and the residual covariance, $\boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}}$, are defined as

$$\mathbf{K}_{\mathbf{b},t} = \boldsymbol{\Sigma}_{\mathbf{b},t-1}\tilde{\mathbf{Y}}_{t-1}^T\boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}}^{-1} \tag{30}$$

$$\boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}} = \boldsymbol{\Sigma}_{\mathbf{z}_t} + \tilde{\mathbf{Y}}_{t-1}^T\boldsymbol{\Sigma}_{\mathbf{b},t-1}\tilde{\mathbf{Y}}_{t-1}. \tag{31}$$

and where $\tilde{\mathbf{y}}_t \triangleq \mathbf{y}_t - \mathbf{C}^T\boldsymbol{\alpha}_{t|t-1}$ and $\tilde{\mathbf{Y}}_{t-1}^T \triangleq \mathbf{Y}_{t-1} + \mathbf{C}^T\boldsymbol{\Gamma}_{t|t-1}$. The channel estimation is thus of the form of the update Kalman equations. As more knowledge about the observations becomes available, the *belief* in the static channel is updated (as opposed to predicting a dynamic into the future and correcting using measurements as in eqn. (17)).

4.3 Marginalization of channel parameters

Recall the marginalization of $\mathbf{b}$ in eqn. (11b),

$$p(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) = \int p(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t})\, d\mathbf{b}$$

The posterior pdf of the source signal, $p(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$, is given in sect. §4.1 by eqn. (16b) and the corresponding KF equations in eqs. (22), (24), (17b), and (17d). The posterior pdf of the channel, $p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t})$, is given in sect. §4.2 by eqn. (28) and the KF equations in eqn. (29).

Inserting eqns. (16b) and (28) into eqn. (11b) and solving the integral using the standard Gaussian identity (see Appen. B), the marginal posterior pdf of $\mathbf{x}_t$ becomes

$$p(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) = \mathcal{N}\left(\mathbf{x}_t \,\middle|\, \hat{\boldsymbol{\mu}}_{t|t}, \hat{\boldsymbol{\Sigma}}_{t|t}\right) \tag{32}$$

KF prediction of $\boldsymbol{\mu}_{t|t-1}^{(i)}$, $\boldsymbol{\Sigma}_{t|t-1}^{(i)}$ (eqs.(17b), (17a)).;
Evaluate $\boldsymbol{\alpha}_{t|t-1}^{(i)}$, $\boldsymbol{\Gamma}_{t|t-1}^{(i)}$, $\boldsymbol{\alpha}_{t|t}^{(i)}$, $\boldsymbol{\Gamma}_{t|t}^{(i)}$ (eqs. (23), (25)).;
KF estimation of $\boldsymbol{\mu}_{\mathbf{b},t}^{(i)}$ and $\boldsymbol{\Sigma}_{\mathbf{b},t}^{(i)}$ (eqn. (29));
KF correction of $\boldsymbol{\mu}_{t|t}^{(i)}$, $\boldsymbol{\Sigma}_{t|t}^{(i)}$ (eqn. (33)).

Algorithm 1: MARBLE KF for source signal estimation at time t given estimates at $t-1$

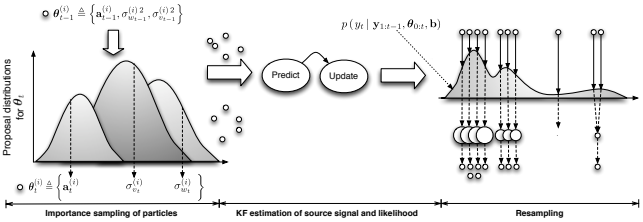


Fig. 2: Rao-Blackwellized particle filter

with marginal covariance, $\hat{\boldsymbol{\Sigma}}_{t|t}$, and marginal mean, $\hat{\boldsymbol{\mu}}_{t|t}$,

$$\hat{\boldsymbol{\mu}}_{t|t} = \boldsymbol{\alpha}_{t|t} + \boldsymbol{\Gamma}_{t|t}\boldsymbol{\mu}_{\mathbf{b},t} \tag{33a}$$

$$\hat{\boldsymbol{\Sigma}}_{t|t} = \left(\mathbf{I}_Q - \mathbf{k}_t\mathbf{c}^T\right)\boldsymbol{\Sigma}_{t|t-1} + \boldsymbol{\Gamma}_{t|t}\boldsymbol{\Sigma}_{\mathbf{b},t}\boldsymbol{\Gamma}_{t|t}^T. \tag{33b}$$

Thus, the MMSE estimator of $\mathbf{x}_t$ is given by $\hat{\boldsymbol{\mu}}_{t|t}$ with error covariance $\hat{\boldsymbol{\Sigma}}_{t|t}$. Comparing eqn. (33a) to the decomposed posterior pdf of the source signal in eqn. (24), the MMSE estimate of $\mathbf{x}_{0:t}$ is equivalent to inserting a MAP estimate of $\mathbf{b}$ in the KF update equation. The error covariance of the estimate takes into account uncertainty introduced by channel estimation through introduction of a weighted version $\boldsymbol{\Sigma}_{\mathbf{b},t}$ by $\boldsymbol{\Gamma}_{t|t}$ in eqn. (33b).

Furthermore, the channel is marginalized from the likelihood eqn. (20) via

$$\begin{aligned} &p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}\right) \\ &= \int p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right) p\left(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}\right) d\mathbf{b} \\ &= \mathcal{N}\left(\mathbf{y}_t \,\middle|\, \boldsymbol{\mu}_{\mathbf{y}_t}, \boldsymbol{\Sigma}_{\mathbf{z}_t,\mathbf{b}}\right) \end{aligned} \tag{34}$$

where $\boldsymbol{\mu}_{\mathbf{y}_t} \triangleq \mathbf{Y}_{t-1}\boldsymbol{\mu}_{\mathbf{b},t-1} + \mathbf{C}^T\left(\boldsymbol{\alpha}_{t|t-1} + \boldsymbol{\Gamma}_{t|t-1}\boldsymbol{\mu}_{\mathbf{b},t-1}\right)$ is defined as the mean (see Appen. C.2). The algorithm for source signal and channel estimation is summarized in Algorithm 1.

5 MARBLE PF for blind speech dereverberation

In sect. §4 the time-varying model parameters, $\boldsymbol{\theta}_{0:t}$ were assumed known. However, in practice, the source signal has to be estimated blindly, i.e., only the measurements are available. Thus, the KF cannot be applied directly. Instead, an optimal estimate of the source signal can be obtained by estimating an ensemble of KFs for all possible parameter choices. As this requires the search of an infinite parameter space, an ensemble of KFs can be evaluated for *stochastically* selected parameters for reduced dimensionality. The parameters are

sampled in a particle filter framework. PFs have been widely discussed in the literature, see, e.g., [36, 37]. For completeness, a brief explanation of PFs is given in sects. §5.1 and §5.2. Sect. §5.3 derives the necessary pdfs required specifically for the MARBLE PF.

5.1 Rao-Blackwellized estimation of $\mathbf{z}_{0:t}$ and $\boldsymbol{\theta}_{0:t}$

Similarly to sect. §4, denoting $\mathbf{f}_{0:t} \triangleq \{\mathbf{z}_{0:t}, \boldsymbol{\theta}_{0:t}\}$, the posterior pdf of the all unknown parameters is given as

$$p(\mathbf{f}_{0:t} \mid \mathbf{y}_{1:t}) = p(\mathbf{z}_{0:t} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t}) \tag{35}$$

If $\widehat{\mathbf{f}}_{0:t}$ is an estimate of $\mathbf{f}_{0:t}$, the MSE between $\mathbf{f}_{0:t}$ and $\widehat{\mathbf{f}}_{0:t}$ is

$$\mathrm{MSE}_{\widehat{\mathbf{f}}_{0:t}} = \int \|\widehat{\mathbf{f}}_{0:t} - \mathbf{f}_{0:t}\|^2\, p(\mathbf{f}_{0:t} \mid \mathbf{y}_{1:t})\, d\mathbf{f}_{0:t}$$

The MMSE estimate of $\mathbf{f}_{0:t}$ is found by differentiating with respect to $\widehat{\mathbf{f}}_{0:t}$ and setting to zero. The optimal estimator can be derived similarly to eqn. (11), such that

$$\widehat{\mathbf{f}}_{0:t} = \begin{bmatrix} \int \mathbf{z}_{0:t}\, p(\mathbf{z}_{0:t} \mid \mathbf{y}_{1:t})\, d\mathbf{z}_{0:t} \\ \int \boldsymbol{\theta}_{0:t}\, p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})\, d\boldsymbol{\theta}_{0:t} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{z}}_{0:t} \\ \widehat{\boldsymbol{\theta}}_{0:t} \end{bmatrix} \tag{36a}$$

Thus, the unknown model parameters can be obtained by marginalizing $\mathbf{z}_{0:t}$ from the joint posterior pdf, such that $\boldsymbol{\theta}_{0:t}$ and $\mathbf{z}_{0:t}$ can be estimated separately. As both the source signal and the channel are obtained using their optimal estimator as described in sect. §4, the estimators of $\mathbf{z}_{0:t}$ and $\boldsymbol{\theta}_{0:t}$ is Rao-Blackwellized [9] by marginalization of the analytically tractable sub-structure from the joint posterior pdf, $p(\mathbf{z}_{0:t}, \boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$. As subsets of smaller dimension are evaluated, the variance of the estimates is decreased compared to direct evaluation of the joint pdf.

By applying Bayes's theorem to $p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$ and reordering slightly, eqn. (36a) can be written as

$$\widehat{\boldsymbol{\theta}}_{0:t} = \frac{\int \boldsymbol{\theta}_{0:t}\, p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_{0:t})\, d\boldsymbol{\theta}_{0:t}}{\int p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_{0:t})\, d\boldsymbol{\theta}_{0:t}}$$

Assuming a proposal distribution, $\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$, is available that is easy to sample from and approximates $p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$ with the same support, the MMSE estimate becomes

$$= \frac{\int \boldsymbol{\theta}_{0:t} \frac{p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}_{0:t}) p(\boldsymbol{\theta}_{0:t})}{\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})}\, \pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})\, d\boldsymbol{\theta}_{0:t}}{\int \frac{p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}_{0:t}) p(\boldsymbol{\theta}_{0:t})}{\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})}\, \pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})\, d\boldsymbol{\theta}_{0:t}}. \tag{37}$$

Defining so-called importance weights, w_t,

$$\begin{aligned} w_t &\triangleq \frac{p(\mathbf{y}_{1:t} \mid \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_{0:t})}{\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})} \\ &= w_{t-1} \times \frac{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{0:t-1})}{\pi(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t-1})} \end{aligned} \tag{38}$$

and recalling $\mathbb{E}_{p(\boldsymbol{\theta}_{0:t} \mid \cdot)}[\mathbf{h}_{0:t}] \triangleq \int \mathbf{h}_{0:t}\, p(\boldsymbol{\theta}_{0:t} \mid \cdot)\, d\boldsymbol{\theta}_{0:t}$ for any function $\mathbf{h}_{0:t}$ of $\boldsymbol{\theta}_{0:t}$, the MMSE estimate in eqn. (37) is

$$\widehat{\boldsymbol{\theta}}_{0:t} = \frac{\mathbb{E}_{\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})}[\boldsymbol{\theta}_{0:t}\, w_t]}{\mathbb{E}_{\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})}[w_t]} \tag{39}$$

5.2 Sequential Monte Carlo sampling

In order to solve eqn. (39), the solution of several high-dimensional, non-linear integrals is required. As an exercise in *stochastic* integration, numerical Monte Carlo methods can be used to approximate these integrals by drawing N independent and identically distributed (i. i. d.) samples, $\boldsymbol{\theta}_{0:t}^{(i)}$, $i \in \mathcal{N}$ from the proposal pdf, leading to the approximation of the MMSE estimate, $\widehat{\boldsymbol{\theta}}_{0:t}$, as

$$\widehat{\boldsymbol{\theta}}_{0:t} = \frac{1}{N} \sum_{i \in \mathcal{N}} \boldsymbol{\theta}_{0:t} \left(\boldsymbol{\theta}_{0:t}^{(i)} \right) \tilde{w}_t^{(i)}$$

where the weight of the i^{th} particle, $w_t^{(i)}$, is normalized via

$$\tilde{w}_t^{(i)} \triangleq w_t^{(i)} \Big/ \sum_{j \in \mathcal{N}} w_t^{(j)}. \tag{40}$$

The desired posterior pdf, $p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$, can thus be approximated by a point-mass distribution [10], i.e.,

$$\widehat{p}_N(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \sum_{i \in \mathcal{N}} \tilde{w}_{0:t}^{(i)} \, \delta \left(\boldsymbol{\theta}_{0:t} - \boldsymbol{\theta}_{0:t}^{(i)} \right). \tag{41}$$

As the proposal distribution, $\pi(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$, only approximates the parameter posterior, $p(\boldsymbol{\theta}_{0:t} \mid \mathbf{y}_{1:t})$, the discrepancy between the proposal and the posterior pdf increases stochastically with time. After few iterations all but one importance weight are close to zero and computational effort is dissipated to tracking particle trajectories not contributing to the final estimate. Resampling ensures that only statistically relevant samples are retained by eliminating degenerate trajectories with small importance weights [38]. A measure of degeneracy is given by the effective sample size,

$$\hat{N}_{eff} = 1 \Big/ \sum_{i \in \mathcal{N}} \left(\tilde{w}_t^{(i)} \right)^2 \tag{42}$$

If $\hat{N}_{eff}$ is below a defined threshold, the particles are resampled. An overview of resampling schemes can be found in [39]. In this paper, systematic resampling is utilized. The principle of the proposed algorithm is illustrated in Fig. 2: N samples of the time-varying model parameters are chosen from a hypothesis distribution, reflecting *belief* in the parameter production system. An ensemble of the KFs for source signal and channel parameters is evaluated for all N choices of model parameters. Using the resulting likelihood, weights are assigned to each resulting set of estimates and parameters. Only statistically relevant samples are retained by resampling all sets corresponding to their weights.

5.3 Choice of importance sampling function

The performance of particle filters is highly dependent on the choice of the proposal distribution that $\boldsymbol{\theta}_{0:t}^{(i)}$ are drawn from. The optimal importance function minimizing the variance upon $\boldsymbol{\theta}_{0:t}^{(i)}$ and the $\mathbf{y}_{1:t}$ is given by $p\left(\boldsymbol{\theta}_t^{(i)} \,\middle|\, \mathbf{y}_t, \boldsymbol{\theta}_{t-1}^{(i)} \right)$ with $w_t^{(i)} \propto p\left(\mathbf{y}_t \mid \boldsymbol{\theta}_{t-1}^{(i)} \right)$ [40]. However, as $\boldsymbol{\theta}_t^{(i)}$ are non-linear in the likelihood due to the form of eqn. (8), the optimal weights, $w_t^{(i)}$ are not analytically tractable.

for $t = 1, \ldots,$ *number of samples* **do**
  for $i = 1, \ldots,$ *number of particles* **do**
    Importance sampling of $\boldsymbol{\theta}_{0:t}^{(i)}$ (eqs. (3), (7b));
    Evaluate MARBLE KF as described in Algorithm 1;
    Evaluation of weights $w_t^{(i)}$ (eqs. (38), (34));
  end
  Normalization of importance weights (eqn. (40));
  Resampling;
  Computation of particle average:

$$\widehat{\mathbf{x}}_t = \sum_{i \in \mathcal{N}} \hat{\boldsymbol{\mu}}_{t|t}^{(i)} \qquad \widehat{\boldsymbol{\theta}}_t = \sum_{i \in \mathcal{N}} \boldsymbol{\theta}_{0:t}^{(i)}, \qquad \widehat{\mathbf{b}} = \sum_{i \in \mathcal{N}} \boldsymbol{\mu}_{\mathbf{b},t}^{(i)}.$$

end

Algorithm 2: MARBLE PF

Sampling from the prior, $p\left(\boldsymbol{\theta}_t^{(i)} \,\middle|\, \boldsymbol{\theta}_{t-1}^{(i)}\right)$, is often used instead of optimal importance sampling. Inserting the prior pdf into eqn. (38), the weights reduce to

$$w_t^{(i)} = w_{t-1}^{(i)}\, p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}^{(i)}\right) \propto p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}^{(i)}\right) \tag{43}$$

where the proportionality is due to the uniform distribution of weights after resampling. $p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}^{(i)}\right)$ can be obtained by marginalizing $\mathbf{z}_{0:t}$ from the joint likelihood, i.e.,

$$p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}^{(i)}\right) = \int p\left(\mathbf{y}_t, \mathbf{z}_{0:t}^{(i)} \,\middle|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}^{(i)}\right) d\mathbf{z}_{0:t}^{(i)} \tag{44}$$

Thus, similar to the posterior pdf of the model parameters, the weights of the particles are obtained by marginalization of source signal and channel from the likelihood. Firstly, the predicted source signal pdf in eqn. (16a) is marginalized from the measurement likelihood (eqn. (15)) in eqn. (20). The likelihood required in eqn. (44) is obtained by marginalizing the channel posterior pdf in eqn. (28) from eqn. (20) as derived in eqn. (34). The MARBLE PF is summarized in Algorithm 2.

6 Experimental results

Examples are presented in the following to demonstrate the proposed approach. To measure the performance of the proposed approach, synthetic examples are used in order to compare the results to an underlying ground truth which is not available for speech data. Speech data is subsequently used to present efficiency for blind speech dereverberation.

In order to test the general performance of the algorithm against an underlying truth, a TVAR signal of order $Q = 4$ is synthetically generated according to eqns. (2) and (3). The source signal is distorted by an acoustic gramophone horn with optimal model order $P = 72$ (corresponding parameters are extracted in, e.g., [41]), white Gaussian noise (WGN) with 25dB signal-to-noise ratio (SNR), and using a single sensor. The MARBLE PF is executed for $N = 200$ particles. The corresponding source signal and channel parameter estimates are shown in Fig.s 3 and 4a. After approximately 1000 samples, the estimates converge towards an accurate approximation of the source signal. However, as particle filters are sequential estimators, the same estimation performance should be evident for the whole cycle of samples and so-called 'burn-in' periods, observed in Markov chain Monte Carlo approaches, should not occur. However, this 'burn-in' period of the source signal estimate is in fact due to the

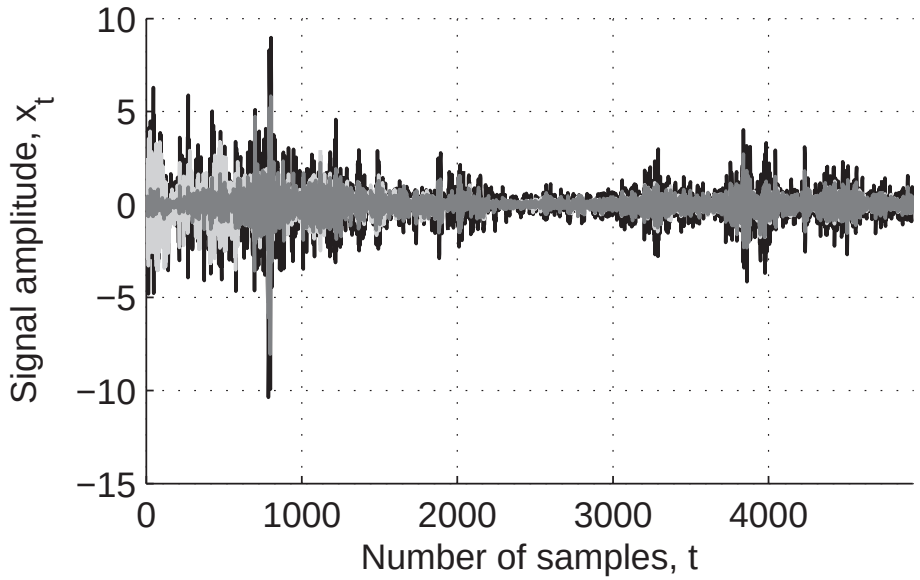


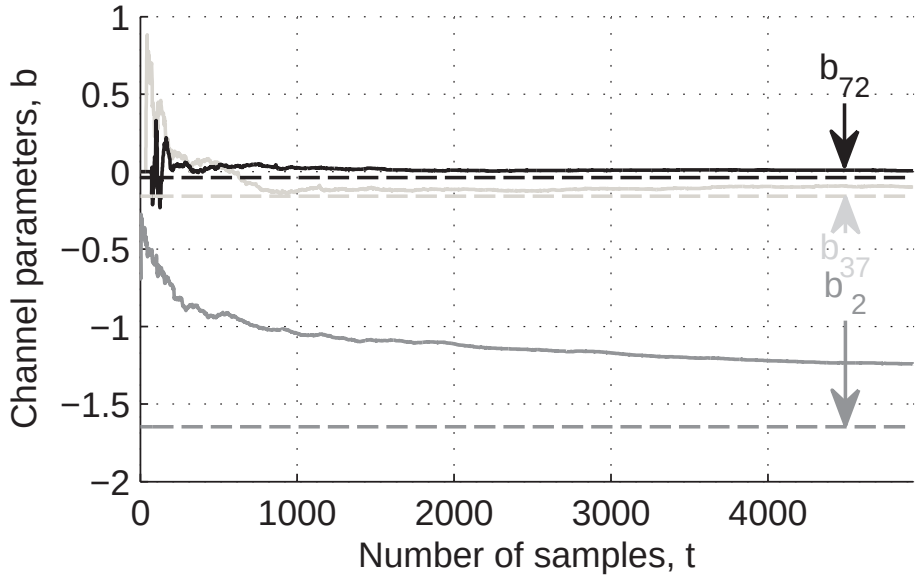
Fig. 3: Accuracy of estimated signal (—) with source signal (—) and observed signal (—) for 200 particles using an acoustic horn channel and TVAR data.

Bayesian update procedure on the static channel parameters: As can be seen from the trajectory of channel parameter estimates in Fig. 4, the channel parameters require approximately 1000 samples to converge towards the actual channel parameters. Note that whilst most parameter estimates converge towards the actual parameters, some (such as b_2) converge to a false constant. Regardless, the channel pole estimate (i.e., the roots of the parameters) accurately approximate the actual channel poles. This is due to the close spacing between poles such that few pole estimates approximate neighboring pole positions.

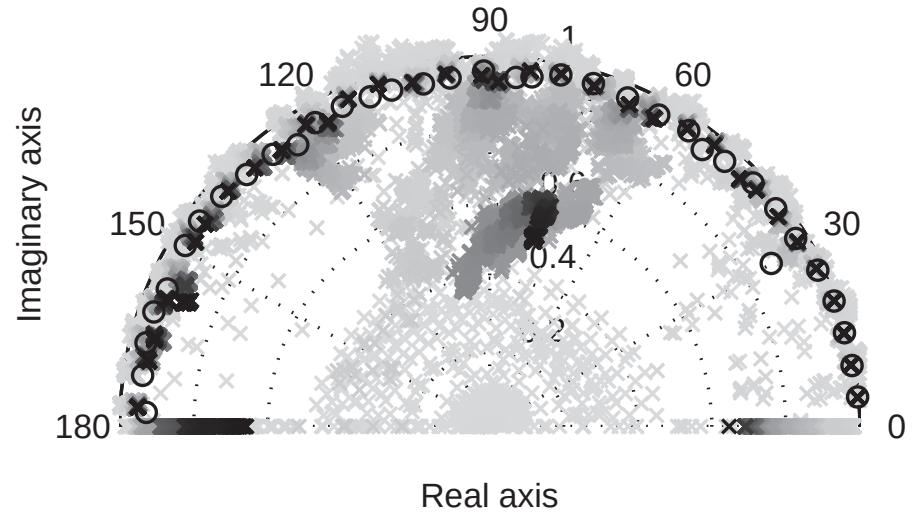
To confirm that the 'burn-in' period of the source signal estimates is in fact an artefact of the channel estimator, the experiment is repeated assuming the channel parameters are *known* in the particle filter framework. The resulting source signal estimate in Fig. 5 accurately estimates the source signal from $t = 1$, thus confirming that the 'burn-in' period apparent in Fig. 3 is, in fact, due to the number of samples required to obtain convergence on the channel parameter estimates.

Using the image-source method [42], the impulse response of a $4.68 \times 2.78 \times 3.2$m room is simulated at $f_s = 500$ Hz. The source is placed 1.54m away from the West wall, the sensors are 10cm apart and 1.415m away from the East wall. The RIR is modelled as an AR process[2] used to filter a 4.2s speech signal at $f_s = 500$ Hz distorted by WGN at SNR of 25dB. The MARBLE PF is applied assuming 10 source parameters using 50 particles and 20 Monte Carlo iterations. Performance of the results is evaluated using the signal-to-

[2] Note that any non-minimum phase components are excluded.



(a) Convergence of estimated channel parameters (solid) towards the actual channel parameters (dashed) after 1000 samples for 200 particles.



(b) Pole trajectory with time from early estimates at $t = P$ (light grey) to estimate at $t = 5000$ (black) vs. actual poles (red circles)

Fig. 4: Convergence of estimated channel parameters and poles of the acoustic horn channel with time towards the actual static channel using 200 particles for a TVAR signal.

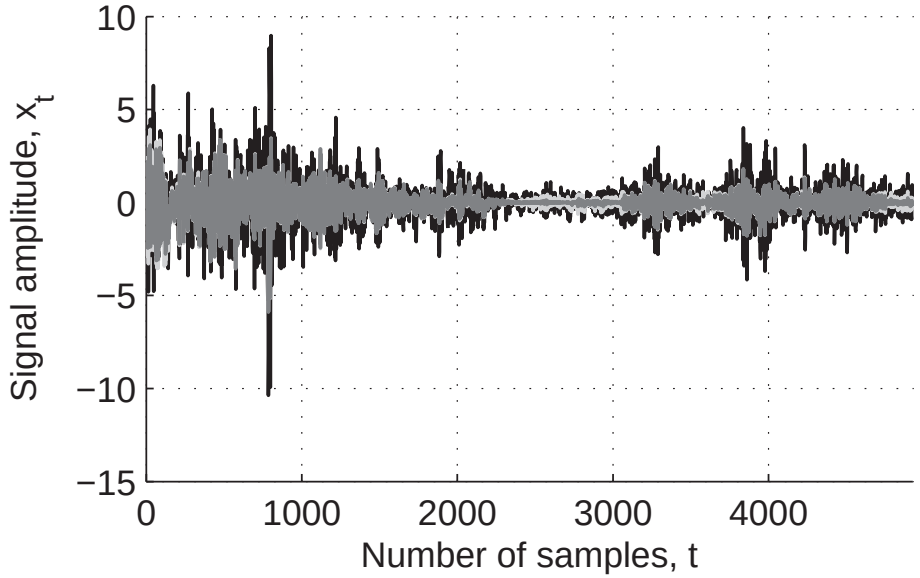


Fig. 5: Clean source vs. estimated vs. observed signal for $P = 72$ acoustic horn and TVAR data assuming *known* channel parameters.

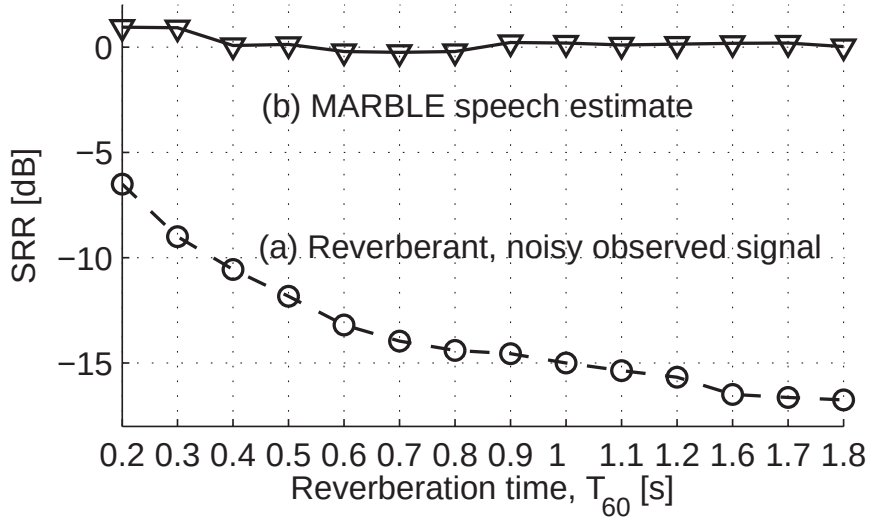


Fig. 6: SRR degradation with increasing reverberation time.

reverberant component ratio (SRR) that can be found as

$$\mathrm{SRR}_{\mathrm{lin}} \triangleq \mathbb{E}\left[\frac{\mathbf{e}_{0:t}^T\mathbf{e}_{0:t}}{\mathbf{x}_{0:t}^T\mathbf{x}_{0:t}}\right] = \frac{\sum_{k=0}^{t} \mathrm{tr}\left[\bar{\boldsymbol{\Sigma}}_{k|k}\right] + \|\bar{\boldsymbol{\mu}}_{k|k} - \mathbf{x}_k\|^2}{\sum_{i=0}^{t} \mathbf{x}_i^T\mathbf{x}_i},$$

where $\mathbf{e}_{0:t} \triangleq \widehat{\mathbf{x}}_{0:t} - \mathbf{x}_{0:t}$ is the error between the estimate, $\widehat{\mathbf{x}}_{0:t}$, and the actual source signal, $\mathbf{x}_{0:t}$; $\mathrm{tr}[\cdot]$ denotes the trace of a matrix; $\bar{\boldsymbol{\mu}}_{t|t}$ and $\bar{\boldsymbol{\Sigma}}_{t|t}$ are the particle average of the corrected Kalman filter mean and covariance respectively in eqn. (33); and the SRR in dB is $\mathrm{SRR}_{\mathrm{dB}} = 10\log 10\{\mathrm{SRR}_{\mathrm{lin}}\}$. Typical reverberation times, T_{60}, for rooms where dereverberation is of interest lie between 0.2 and 2s. The PF is hence applied for impulse responses generated using these T_{60} times to measure the approach's performance with increasing reverberation effects. The corresponding SRR of the observed signal approximately follows an exponential decay, decreasing by 10dB with increasing T_{60} (Fig. 6). In contrast, the SRR of the MARBLE PF remains almost constant and in total decreases by only 1dB, thus leading to a robust performance improvement of 17dB as compared to the reverberant observations at $T_{60} = 1.8$s. The MARBLE PF is thus comparatively robust against the reverberation time of a room. Using $T_{60} = 0.45$s, the experiment is repeated for $M = 1$ to 10 sensors (Fig. 7). Whilst the MARBLE PF achieves an SRR improvement of 12dB using a single sensor, the SRR is increased by a further 4dB when using 10 sensors. Multi-sensor processing thus leads to significant SRR improvement. Fig. 8 illustrates that further performance improvement can be achieved by increasing the distance between sensors, thus exploiting spatial diversity of sensors.

Fig. 9 compares the time-series of the speech, observed, and estimated signals at 250Hz distorted by the simulated image-source method (ISM) data and WGN for a single and multiple sensors. Whilst the estimate accurately approximates the source signal within approx-

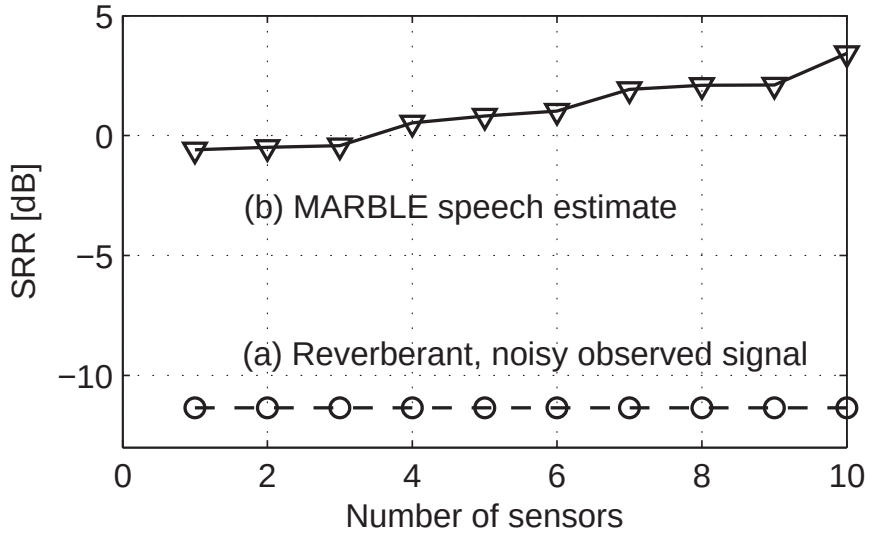


Fig. 7: SRR improvement with increasing number of sensors.

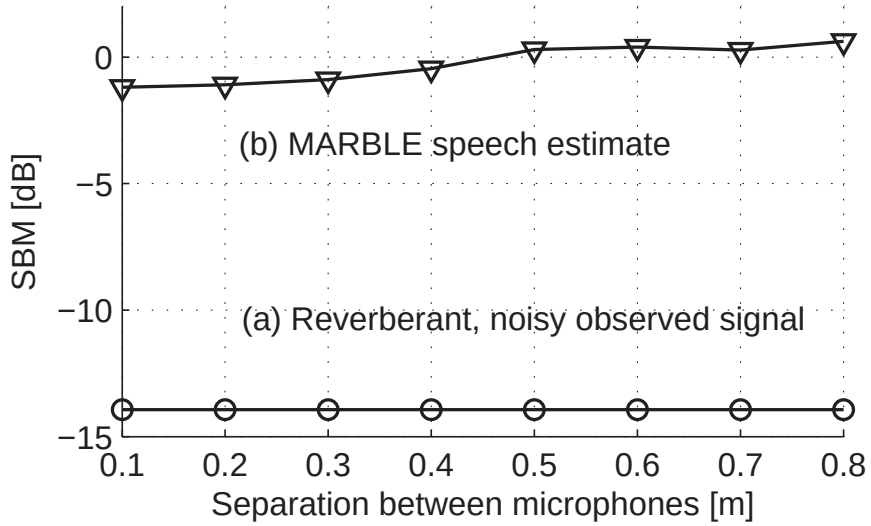


Fig. 8: SRR improvement with increasing separation between microphones

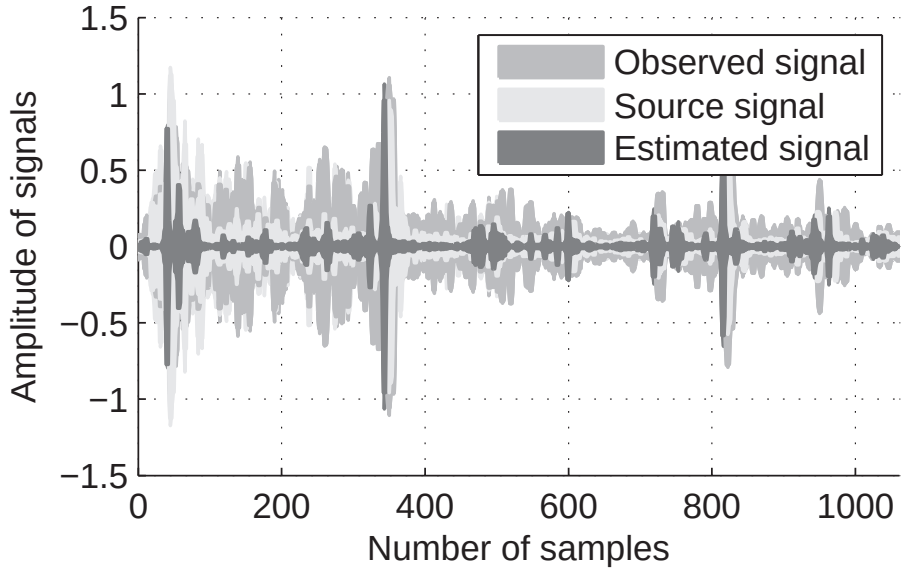
imately 200 samples when using multiple sensors, a 'burn-in' period of about 600 samples is required for the single sensor case to obtain accurate estimation.

7 Model extension: Multirate filtering

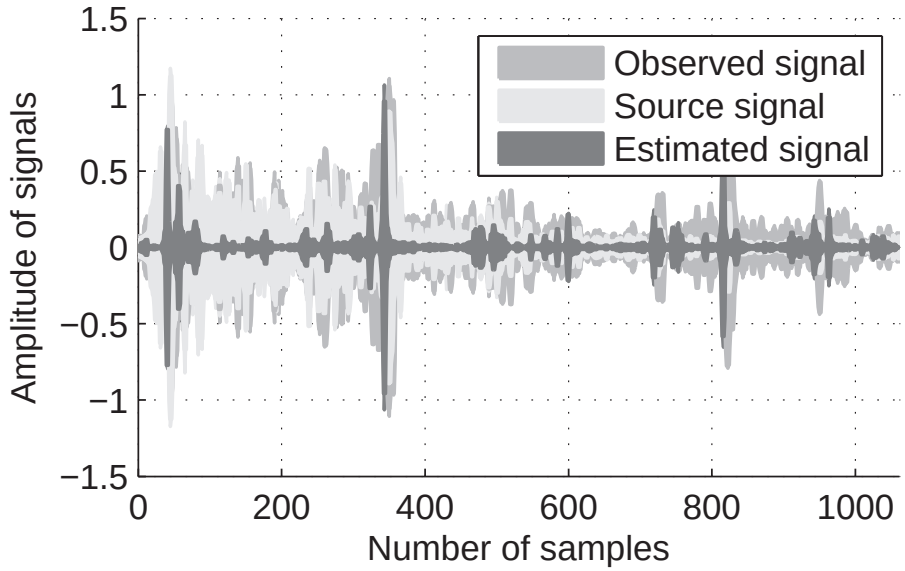
In sect. §6, results were demonstrated for subband speech signals sampled at 500 and 250Hz. However, fundamental frequencies of speech phonemes start at approximately 125Hz for male speakers (e.g., /u/ has a fundamental frequency of $f_0 = 141$Hz), whilst formant frequencies for male speech and even fundamental frequencies in female speech lie well above 230Hz (e.g., /u/ corresponds to $f_0 = 231$Hz in female speech, whilst the first formant frequency lie at 300Hz in male speech and and 370Hz in female speech; second and higher formant frequencies start from 870Hz) [43]. Therefore, harmonic components generated by the speech formants are disregarded when decimating speech signals to a sampling frequency of 500Hz and below.

Nevertheless, the model order of RIRs increases with increasing sampling rate (see sect. §3.2). Thus, several hundred channel parameters are necessary in order to accurately capture the RIR when using fullband speech data (at least 8kHz sampling rate). Thus, evaluation of the KF equations in eqs. (29), (33) and (34) involves processing of extremely large matrices for speech signals. Therefore, to avoid computational overhead, the sampling rate in this paper was limited to 250-500Hz. However, the presented results can be extended to fullband signals combining several subband signals in a multirate filtering approach.

Processing can be carried out more efficiently by splitting the fullband signal into several sub-band signals. An analysis filter bank consisting of K filters, $h_{k,t}$, $k \in \mathscr{K}$, channelizes the observed signal, $\mathbf{y}_{1:t}$, into K sub-band signals, $\mathbf{y}_{k,t}$, and decimates the resulting signals by a factor of K (denoted as $\downarrow K$). Because of the reduced sampling frequency, fewer model parameters and samples are required, leading to more efficient and faster processing. After



(a) Accuracy of estimated signal approximating source signal using multiple sensors.



(b) Estimated signal approximating source signal after an initial 'burn-in' period using a single sensor.

Fig. 9: Estimated signal, source, and observed signal for multi- and single-sensor MARBLE PF for speech filtered by a RIR generated using the ISM method with $T_{60} = 0.45$s.

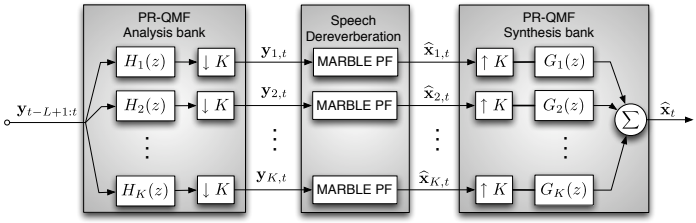


Fig. 10: Sequential multirate filtering of fullband reverberant speech, $\mathbf{y}_{1:t}$, into K sub-band signals, $y_{k,t}, k \in \mathscr{K}$ using GDFT analysis and synthesis banks with filter length L of the dereverberated speech estimated, $\hat{\mathbf{x}}_t$

processing, the estimated subband signals, $\widehat{\mathbf{x}}_{k,t}$ are recombined by interpolating by a factor of K (denoted as $\uparrow K$) and applying K synthesis filters, $g_{k,t}$. The fullband signal is the sum over the output of the synthesis filters. This multirate processing approach is shown in Fig. 10 and is commonly referred to as a K-channel filterbank [44]. For near-perfect reconstruction at critical sampling, the generalised discrete Fourier transform (GDFT) filterbank [45] can be used. Multirate filtering using the GDFT was applied successfully in blind dereverberation problems in, e.g., [46, 47]. Even though the design results in complex subband signals, only $K/2$ subbands need to be processed if K is even. The remaining $K/2$ subbands are given as the complex conjugates of the processed subbands. The analysis filters, $h_{k,i}$, can be computed from a single prototype filter, $h_{pr,i}$, of length L_{pr} and bandwidth $2\pi/K$,

$$h_{k,i} = h_{pr,i} \exp\left\{ j\frac{2\pi}{K}(k+k_0)(i+i_0) \right\} \qquad i = 0,1,\ldots,L_{pr}-1$$

where $i_0 = 0$ and $k_0 = 1/2$ [45]. The synthesis filter satisfying near-perfect reconstruction for the reconstruction of the fullband signal is given by the time-reversed conjugate of the analysis filter, i.e., $v_{k,i} = h^{\star}_{k,L_{pr}-i-1}$.

8 Conclusion

This paper presents a Bayesian approach to multi-sensor blind dereverberation of speech from a stationary speaker. The source signal and reverberant channel are directly estimated from the distorted observations by means of their optimal estimators, the KF. As knowledge of model parameters is required but unavailable, an ensemble of KFs is evaluated for stochastically selected parameters. The parameters are obtained by importance sampling in a PF framework. Due to Rao-Blackwellization, the variance of the estimates is decreased. Multi-sensor blind dereverberation is compared to utilization of a single sensor by means of examples. Results show that whilst the single-sensor case offers accurate estimation of the source signal after an initial burn-in period, multiple sensors allow for instant accuracy using less particles and increased SRR performance.

Future research will expand the framework to fullband signals via multirate filtering, extend the model to facilitate blind dereverberation of speech from moving speakers, and incorporate joint enhancement and target tracking of the source.

Acknowledgement

The author's would like to thank the reviewers for their helpful comments.

A Estimation of channel parameters

This section derives the results discussed in sect. §4.2. Using Bayes's theorem, the channel posterior pdf is expressed as

$$\begin{aligned} &p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) \\ &= \frac{p(\mathbf{y}_t, \boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b})\, p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}{p(\mathbf{y}_t, \boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})} \end{aligned} \tag{45}$$

By applying the probability chain rule,

$$\begin{aligned} &\frac{p(\mathbf{y}_t, \boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b})}{p(\mathbf{y}_t, \boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})} \\ &\quad = \frac{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b})}{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t})\, p(\boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}. \end{aligned} \tag{46}$$

Since $\boldsymbol{\theta}_t$ depends only on $\boldsymbol{\theta}_{t-1}$ and $\mathbf{y}_t$, the pdfs of $\boldsymbol{\theta}_t$ reduce to the prior pdf, $p(\boldsymbol{\theta}_t \,|\, \boldsymbol{\theta}_{t-1}, \mathbf{y}_{1:t-1}, \cdot) = p(\boldsymbol{\theta}_t \,|\, \boldsymbol{\theta}_{t-1})$,

$$\begin{aligned} &p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) \\ &\quad = \frac{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})\, p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})}. \end{aligned} \tag{26}$$

Inserting $p(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})$ from eqn. (27) and the likelihood, $p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b})$, in eqn. (20) into eqn. (26) and defining $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{C}^T \boldsymbol{\alpha}_{t|t-1}$ and $\tilde{\mathbf{Y}}_{t-1}^T = \mathbf{Y}_{t-1} + \mathbf{C}^T \boldsymbol{\Gamma}_{t|t-1}$, then after a little rearrangement:

$$\begin{aligned} p(\mathbf{b} \,|\, \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) &= \frac{(2\pi)^{-(P+1)}}{|\boldsymbol{\Sigma}_{\mathbf{b},t-1}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{\mathbf{z}_t}|^{\frac{1}{2}}} \frac{1}{p(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1})} \\ &\quad \times \exp\left\{ -\frac{1}{2} \left[\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \mathbf{b} - 2\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \boldsymbol{\mu}_{\mathbf{b},t} + \beta \right] \right\} \\ &= \mathcal{N}\left(\mathbf{b} \,|\, \boldsymbol{\mu}_{\mathbf{b},t}, \boldsymbol{\Sigma}_{\mathbf{b},t}\right) \end{aligned} \tag{47}$$

where $\beta \triangleq \boldsymbol{\mu}_{\mathbf{b},t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1}^{-1} \boldsymbol{\mu}_{\mathbf{b},t-1} + \tilde{\mathbf{y}}_t^T \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t$ and

$$\boldsymbol{\Sigma}_{\mathbf{b},t} = \left(\boldsymbol{\Sigma}_{\mathbf{b},t-1}^{-1} + \tilde{\mathbf{Y}}_{t-1} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{Y}}_{t-1}^T \right)^{-1} \tag{48a}$$

$$\boldsymbol{\mu}_{\mathbf{b},t} = \boldsymbol{\Sigma}_{\mathbf{b},t} \left(\boldsymbol{\Sigma}_{\mathbf{b},t-1}^{-1} \boldsymbol{\mu}_{\mathbf{b},t-1} + \tilde{\mathbf{Y}}_{t-1} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t \right) \tag{48b}$$

Now, using the Woodbury matrix identity,

$$(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U} \left(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U} \right)^{-1} \mathbf{V}\mathbf{A}^{-1} \tag{49}$$

eqn. (48a) can be rewritten as

$$\boldsymbol{\Sigma}_{\mathbf{b},t} = \boldsymbol{\Sigma}_{\mathbf{b},t-1} - \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \left(\boldsymbol{\Sigma}_{\mathbf{z}_t} + \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \right)^{-1}$$

$$\times \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1}$$
$$= \left\{ \mathbf{I}_P - \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \left(\boldsymbol{\Sigma}_{\mathbf{z}_t} + \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \right)^{-1} \tilde{\mathbf{Y}}_{t-1}^T \right\} \boldsymbol{\Sigma}_{\mathbf{b},t-1}$$

which simplifies to eqn. (29) in sect. §4.2. Eqn. (48b) can thus be written as

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{b},t} &= \left\{ \mathbf{I}_P - \mathbf{K}_{\mathbf{b},t} \tilde{\mathbf{Y}}_{t-1}^T \right\} \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t \\
&\quad + \left\{ \mathbf{I}_P - \mathbf{K}_{\mathbf{b},t} \tilde{\mathbf{Y}}_{t-1}^T \right\} \boldsymbol{\mu}_{\mathbf{b},t-1} \\
&= \boldsymbol{\mu}_{\mathbf{b},t-1} - \mathbf{K}_{\mathbf{b},t} \left(\tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\mu}_{\mathbf{b},t-1} + \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t \right) \\
&\quad + \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t
\end{aligned}$$

Recalling $\mathbf{K}_{\mathbf{b},t} = \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}}^{-1}$ (eqn. (30)) and reordering

$$\begin{aligned}
&= \boldsymbol{\mu}_{\mathbf{b},t-1} - \mathbf{K}_{\mathbf{b},t} \Bigg(\tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\mu}_{\mathbf{b},t-1} \\
&\quad + \left\{ \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1} + \boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}} \right\} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t \Bigg)
\end{aligned}$$

By rearranging eqn. (31) to $\boldsymbol{\Sigma}_{\mathbf{z}_{t,\mathbf{b}}} - \boldsymbol{\Sigma}_{\mathbf{z}_t} = \tilde{\mathbf{Y}}_{t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1} \tilde{\mathbf{Y}}_{t-1}$ and inserting, the mean can be simplified to eqn. (29).

B Marginalization of channel

This section derives the results discussed in sect. §4.3. Inserting eqns. (16b) and (29) into eqn. (11b) and regrouping according to terms gives

$$\begin{aligned}
&p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) \\
&= \frac{(2\pi)^{-\frac{P+Q}{2}}}{|\boldsymbol{\Sigma}_{t|t}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{\mathbf{b},t}|^{\frac{1}{2}}} \int \exp\Bigg\{ -\frac{1}{2} \Big[\mathbf{b}^T \left(\boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} + \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\Gamma}_{t|t} \right) \mathbf{b} \\
&- 2\mathbf{b}^T \left(\boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \boldsymbol{\mu}_{\mathbf{b},t} + \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{x}_t - \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\alpha}_{t|t} \right) + \alpha \Big] \Bigg\} d\mathbf{b}
\end{aligned}$$

where

$$\alpha \triangleq \mathbf{x}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{x}_t - 2\mathbf{x}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\alpha}_{t|t} + \boldsymbol{\alpha}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\alpha}_{t|t} + \boldsymbol{\mu}_{\mathbf{b},t}^T \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \boldsymbol{\mu}_{\mathbf{b},t}$$

contains all terms independent of $\mathbf{b}$. Defining $-\boldsymbol{\Gamma} \triangleq \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} + \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\Gamma}_{t|t}$ and $\mathbf{e} \triangleq \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\alpha}_{t|t} - \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \boldsymbol{\mu}_{\mathbf{b},t}$ and applying the Gaussian identity,

$$\begin{aligned}
&p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}) \\
&= \frac{(2\pi)^{-\frac{P+Q}{2}}}{|\boldsymbol{\Sigma}_{t|t}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{\mathbf{b},t}|^{\frac{1}{2}}} \frac{(2\pi)^{\frac{P}{2}}}{|\boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} + \boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\Gamma}_{t|t}|^{\frac{1}{2}}} \exp\Bigg\{ -\frac{1}{2} \Big[\alpha \\
&\quad - \left(\boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{x}_t + \mathbf{e} \right)^T \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{\Gamma}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{x}_t + \mathbf{e} \right) \Big] \Bigg\} \\
&= \mathcal{N}\left(\mathbf{x}_t \,\middle|\, \hat{\boldsymbol{\mu}}_{t|t}, \hat{\boldsymbol{\Sigma}}_{t|t} \right)
\end{aligned} \tag{32}$$

where the covariance, $\hat{\boldsymbol{\Sigma}}_{t|t}$, and mean, $\hat{\boldsymbol{\mu}}_{t|t}$, are given as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{t|t} &= \left(\boldsymbol{\Sigma}_{t|t}^{-1} - \boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\right)^{-1} \\ &= \left(\boldsymbol{\Sigma}_{t|t}^{-1} - \hat{\mathbf{C}}\boldsymbol{\Gamma}^{-1}\hat{\mathbf{C}}^{T}\right)^{-1} \qquad (50a) \\ \hat{\boldsymbol{\mu}}_{t|t} &= \hat{\boldsymbol{\Sigma}}_{t|t}\left(\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\alpha}_{t|t} + \boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t}\boldsymbol{\Gamma}^{-1}\mathbf{e}\right) \\ &= \hat{\boldsymbol{\Sigma}}_{t|t}\left(\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\alpha}_{t|t} + \hat{\mathbf{C}}\boldsymbol{\Gamma}^{-1}\mathbf{e}\right) \qquad (50b)\end{aligned}$$

where $\hat{\mathbf{C}} \triangleq \boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t}$. Thus, eqn. (50) is identical in form with eqn. (48), and can hence be rewritten as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{t|t} &= \left(\mathbf{I}_Q - \hat{\mathbf{K}}_t\hat{\mathbf{C}}^{T}\right)\boldsymbol{\Sigma}_{t|t} \\ \hat{\boldsymbol{\mu}}_{t|t} &= \left(\mathbf{I}_Q - \hat{\mathbf{K}}_t\hat{\mathbf{C}}^{T}\right)\boldsymbol{\alpha}_{t|t} + \hat{\mathbf{K}}_t\mathbf{e}\end{aligned}$$

where

$$\begin{aligned}\hat{\mathbf{K}}_t &\triangleq \boldsymbol{\Sigma}_{t|t}\hat{\mathbf{C}}\left(\boldsymbol{\Gamma} + \hat{\mathbf{C}}^{T}\boldsymbol{\Sigma}_{t|t}\hat{\mathbf{C}}\right) \\ &= \boldsymbol{\Sigma}_{t|t}\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t}\left(-\boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} - \boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t} + \boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\Gamma}_{t|t}\right)^{-1} \\ &= -\boldsymbol{\Gamma}_{t|t}\boldsymbol{\Sigma}_{\mathbf{b},t}\end{aligned}$$

such that but inserting back into $\hat{\boldsymbol{\Sigma}}_{t|t}$,

$$\hat{\boldsymbol{\Sigma}}_{t|t} = \left(\mathbf{I}_Q + \boldsymbol{\Gamma}_{t|t}\boldsymbol{\Sigma}_{\mathbf{b},t}^{-1}\boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\right)\boldsymbol{\Sigma}_{t|t}$$

which can be rearranged to eqn. (33b). Furthermore, by inserting $\hat{\mathbf{K}}$, $\hat{\mathbf{C}}$, and $\mathbf{e}$ in $\hat{\boldsymbol{\mu}}_{t|t}$,

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{t|t} = &\left(\mathbf{I}_Q + \boldsymbol{\Gamma}_{t|t}\boldsymbol{\Sigma}_{\mathbf{b},t}\boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\right)\boldsymbol{\alpha}_{t|t} \\ &- \boldsymbol{\Gamma}_{t|t}\boldsymbol{\Sigma}_{\mathbf{b},t}\left(\boldsymbol{\Gamma}_{t|t}^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\alpha}_{t|t} - \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1}\boldsymbol{\mu}_{\mathbf{b},t}\right)\end{aligned}$$

which, by eliminating terms, can be rearranged to eqn. (33a).

C Marginalization from likelihood

This section derives the results discussed in sect. §5.3.

C.1 Marginalization of source signal

Inserting eqns. (15) and (17b) into eqn. (14),

$$\begin{aligned}&p\left(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right) \\ &\quad= \frac{1}{(2\pi)^{\frac{Q+1}{2}}\,\sigma_{w_t}^{2}\,|\boldsymbol{\Sigma}_{t|t-1}|^{\frac{1}{2}}}\,\frac{1}{p\left(y_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}\right)} \\ &\qquad\times \exp\Bigg\{-\frac{1}{2}\Big[\mathbf{x}_t^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\mathbf{x}_t - 2\mathbf{x}_t^{T}\boldsymbol{\Sigma}_{t|t}^{-1}\boldsymbol{\mu}_{t|t} \\ &\qquad\quad + \hat{\mathbf{y}}_{t,\mathbf{b}}^{T}\boldsymbol{\Sigma}_{w_t}^{-1}\hat{\mathbf{y}}_{t,\mathbf{b}} + \boldsymbol{\mu}_{t|t-1}^{T}\boldsymbol{\Sigma}_{t|t-1}^{-1}\boldsymbol{\mu}_{t|t-1}\Big]\Bigg\}\end{aligned} \qquad (51)$$

where $\hat{\mathbf{y}}_{t,\mathbf{b}} = \mathbf{y}_t - \mathbf{Y}_{t-1}\mathbf{b}$. Integrating both sides of eqn. (51) with respect to $\mathbf{x}_t$ (since the left hand side must integrate to unity):

$$\begin{aligned} p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) &= \frac{1}{(2\pi)^{\frac{Q+P}{2}} |\boldsymbol{\Sigma}_{w_t}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{t|t-1}|^{\frac{1}{2}}} \\ &\times \int \exp\left\{ -\frac{1}{2}\left[\mathbf{x}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{x}_t - 2\mathbf{x}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\mu}_{t|t} \right.\right. \\ &\left.\left. + \hat{\mathbf{y}}_{t,\mathbf{b}}^T \boldsymbol{\Sigma}_{w_t}^{-1} \hat{\mathbf{y}}_{t,\mathbf{b}} + \boldsymbol{\mu}_{t|t-1}^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} \right] \right\} d\mathbf{x}_t \end{aligned} \tag{52}$$

By applying the Gaussian identity,

$$\begin{aligned} &\int_{\mathbb{R}^N} \exp\left\{ -\frac{1}{2}\left[\alpha + 2\boldsymbol{\beta}^T \mathbf{y} + \mathbf{y}\boldsymbol{\Gamma}\mathbf{y} \right] \right\} d\mathbf{y} \\ &\qquad = \frac{(2\pi)^{\frac{N}{2}}}{|\boldsymbol{\Gamma}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\left[\alpha - \boldsymbol{\beta}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\beta} \right] \right\} \end{aligned} \tag{53}$$

eqn. (52) can be rewritten as

$$\begin{aligned} p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b}) &= \frac{|\boldsymbol{\Sigma}_{t|t}|^{\frac{1}{2}}}{(2\pi)^{\frac{M}{2}} |\boldsymbol{\Sigma}_{w_t}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{t|t-1}|^{\frac{1}{2}}} \\ \times \exp\left\{ -\frac{1}{2}\left[\hat{\mathbf{y}}_{t,\mathbf{b}}^T \boldsymbol{\Sigma}_{w_t}^{-1} \hat{\mathbf{y}}_{t,\mathbf{b}} + \boldsymbol{\mu}_{t|t-1}^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} - \boldsymbol{\mu}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\mu}_{t|t} \right] \right\} \end{aligned} \tag{54}$$

Defining $\hat{\mathbf{K}}_t = \mathbf{I}_Q - \mathbf{K}_t \mathbf{C}^T$, such that from eqn. (17a), $\boldsymbol{\mu}_{t|t} = \hat{\mathbf{K}}_t \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \hat{\mathbf{y}}_{t,\mathbf{b}}$, then the term

$$\begin{aligned} &\boldsymbol{\mu}_{t|t-1}^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} - \boldsymbol{\mu}_{t|t}^T \boldsymbol{\Sigma}_{t|t}^{-1} \boldsymbol{\mu}_{t|t} - \eta \\ &= \boldsymbol{\mu}_{t|t-1}^T \left(\boldsymbol{\Sigma}_{t|t-1}^{-1} - \hat{\mathbf{K}}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \hat{\mathbf{K}}_t \right) \boldsymbol{\mu}_{t|t-1} - 2\hat{\mathbf{y}}_{t,\mathbf{b}} \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \hat{\mathbf{K}}_t \boldsymbol{\mu}_{t|t-1} \end{aligned}$$

where $\eta \triangleq \hat{\mathbf{y}}_{t,\mathbf{b}}^T \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{K}_t \hat{\mathbf{y}}_{t,\mathbf{b}}$ is independent of $\boldsymbol{\mu}_{t|t-1}$. Now, as $\boldsymbol{\Sigma}_{t|t} = \hat{\mathbf{K}}_t \boldsymbol{\Sigma}_{t|t-1}$ using eqn. (17d),

$$\begin{aligned} &= \boldsymbol{\mu}_{t|t-1}^T \left(\mathbf{I}_Q - \hat{\mathbf{K}}_t^T \right) \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} - 2\hat{\mathbf{y}}_{t,\mathbf{b}} \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} \\ &= \boldsymbol{\mu}_{t|t-1}^T \mathbf{C} \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} - 2\hat{\mathbf{y}}_{t,\mathbf{b}} \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t-1}^{-1} \boldsymbol{\mu}_{t|t-1} \end{aligned}$$

Using eqn. (18), $\mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t-1}^{-1} = \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \mathbf{C}^T$:

$$= \boldsymbol{\mu}_{t|t-1}^T \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \mathbf{C}^T \boldsymbol{\mu}_{t|t-1} - 2\hat{\mathbf{y}}_{t,\mathbf{b}} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \mathbf{C}^T \boldsymbol{\mu}_{t|t-1}$$

Inserting into eqn. (54), note that the terms independent of $\boldsymbol{\mu}_{t|t-1}$ can be rewritten as

$$\begin{aligned} &\hat{\mathbf{y}}_{t,\mathbf{b}}^T \boldsymbol{\Sigma}_{w_t}^{-1} \hat{\mathbf{y}}_{t,\mathbf{b}} - \hat{\mathbf{y}}_{t,\mathbf{b}}^T \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{K}_t \hat{\mathbf{y}}_{t,\mathbf{b}} \\ &\quad = \hat{\mathbf{y}}_{t,\mathbf{b}}^T \left(\boldsymbol{\Sigma}_{w_t}^{-1} - \mathbf{K}_t^T \boldsymbol{\Sigma}_{t|t}^{-1} \mathbf{K}_t \right) \hat{\mathbf{y}}_{t,\mathbf{b}} \\ &\quad = \hat{\mathbf{y}}_{t,\mathbf{b}}^T \left(\boldsymbol{\Sigma}_{w_t}^{-1} - \mathbf{K}_t^T \left(\mathbf{C} \boldsymbol{\Sigma}_{w_t}^{-1} \mathbf{C}^T + \boldsymbol{\Sigma}_{t|t-1}^{-1} \right) \mathbf{K}_t \right) \hat{\mathbf{y}}_{t,\mathbf{b}} \end{aligned}$$

$$= \hat{\mathbf{y}}_{t,\mathbf{b}}^T \left(\boldsymbol{\Sigma}_{w_t}^{-1} - \mathbf{K}_t^T \mathbf{C} \boldsymbol{\Sigma}_{w_t}^{-1} \mathbf{C}^T \mathbf{K}_t - \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \mathbf{C}^T \mathbf{K}_t \right) \hat{\mathbf{y}}_{t,\mathbf{b}} \tag{55}$$

Inserting eqn. (18) into eqn. (19),

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{z}_t} = \boldsymbol{\Sigma}_{w_t} + \mathbf{C}^T \boldsymbol{\Sigma}_{t|t-1} \mathbf{C} = \boldsymbol{\Sigma}_{w_t} + \mathbf{C}^T \mathbf{K}_t \boldsymbol{\Sigma}_{\mathbf{z}_t} \\ \Rightarrow \mathbf{C}^T \mathbf{K}_t = \mathbf{I}_P - \boldsymbol{\Sigma}_{w_t} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \end{aligned} \tag{56}$$

such that eqn. (55) can be written as

$$\begin{aligned} &\hat{\mathbf{y}}_{t,\mathbf{b}}^T \left(\boldsymbol{\Sigma}_{w_t}^{-1} - \mathbf{K}_t^T \mathbf{C} \boldsymbol{\Sigma}_{w_t}^{-1} \left(\mathbf{I}_P - \boldsymbol{\Sigma}_{w_t} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \right) - \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \mathbf{C}^T \mathbf{K}_t \right) \hat{\mathbf{y}}_{t,\mathbf{b}} \\ &= \hat{\mathbf{y}}_{t,\mathbf{b}}^T \left(\mathbf{I}_P - \mathbf{K}_t^T \mathbf{C} \right) \boldsymbol{\Sigma}_{w_t}^{-1} \hat{\mathbf{y}}_{t,\mathbf{b}} = \hat{\mathbf{y}}_{t,\mathbf{b}}^T \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \hat{\mathbf{y}}_{t,\mathbf{b}} \end{aligned}$$

Finally, note that

$$\frac{\det(\boldsymbol{\Sigma}_{t|t})}{\det(\boldsymbol{\Sigma}_{t|t-1})} = \det\left(\mathbf{I}_Q - \mathbf{K}_t \mathbf{C}^T \right) = 1 - \mathbf{C}^T \mathbf{K}_t = \boldsymbol{\Sigma}_{w_t} \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1}$$

by using eqn. (56) and the identity $\det(\mathbf{I}_Q + \mathbf{u}\mathbf{v}^T) = 1 + \mathbf{v}^T \mathbf{y}$. Hence, eqn. (52) can be written in the simplified form:

$$p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b} \right) = \mathcal{N}\left(\mathbf{y}_t \,|\, \mathbf{Y}_{t-1}\mathbf{b} + \mathbf{C}^T \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{\mathbf{z}_t} \right) \tag{20}$$

C.2 Marginalization of channel

$$\begin{aligned} p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t} \right) &= \int p\left(\mathbf{y}_t, \mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t} \right) d\mathbf{b} \\ &= \int p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b} \right) p\left(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{1:t-1}, \boldsymbol{\theta}_t \right) d\mathbf{b} \\ &= \int p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b} \right) \\ &\quad \times \frac{p\left(\boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \mathbf{b} \right) p\left(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1} \right)}{p\left(\boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1} \right)} d\mathbf{b} \end{aligned}$$

Applying first-order Markov properties of $\boldsymbol{\theta}_t$, and independence of $\mathbf{y}_{1:t-1}$ and channel, $p\left(\boldsymbol{\theta}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1}, \cdot \right)$ reduces to $p\left(\boldsymbol{\theta}_t \,|\, \boldsymbol{\theta}_{t-1} \right)$. As both terms cancel,

$$= \int p\left(\mathbf{y}_t \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t}, \mathbf{b} \right) p\left(\mathbf{b} \,|\, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}_{0:t-1} \right) d\mathbf{b}$$

by inserting eqns. (44) and (29), then after rearrangement,

$$\begin{aligned} = \frac{(2\pi)^{-P}}{|\boldsymbol{\Sigma}_{\mathbf{z}_t}|^{\frac{1}{2}} |\boldsymbol{\Sigma}_{\mathbf{b},t}|^{\frac{1}{2}}} \int \exp\Bigg\{ -\frac{1}{2} \Big[\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \mathbf{b} - 2\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{b},t}^{-1} \boldsymbol{\mu}_{\mathbf{b},t} \\ + \tilde{\mathbf{y}}_t^T \boldsymbol{\Sigma}_{\mathbf{z}_t}^{-1} \tilde{\mathbf{y}}_t + \boldsymbol{\mu}_{\mathbf{b},t-1}^T \boldsymbol{\Sigma}_{\mathbf{b},t-1}^{-1} \boldsymbol{\mu}_{\mathbf{b},t-1} \Big] \Bigg\} d\mathbf{b}. \end{aligned} \tag{57}$$

Eqn. (57) is identical to eqn. (52) and reduces to eqn. (34).

References

1. V. M. A. Peutz, "Articulation loss of consonants as a criterion for speech transmission in a room," *J. Audio Eng. Soc.*, vol. 19, no. 11, pp. 915–919, Dec. 1971.
2. A. K. Nábelek, T. R. Letowski, and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, vol. 86, no. 4, pp. 1259–1265, 1989.
3. R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," *J. Royal Stat. Soc.*, vol. 21, pp. 577–580, 1949.
4. J. R. Hopgood and P. J. W. Rayner, "Blind single channel deconvolution using nonstationary signal processing," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 476–488, Sep. 2003.
5. J. R. Hopgood, C. Evers, and S. Fortune, "Bayesian single channel blind dereverberation of speech from a moving speaker," in *Speech dereverberation*, P. A. Naylor and N. Gaubitch, Eds. Springer, 2009.
6. C. Evers, J. R. Hopgood, and J. Bell, "Acoustic models for online blind source dereverberation using sequential monte carlo methods," in *Proc. IEEE Conf. ICASSP*, Las Vegas, NV, 24 Mar. - 4 Apr. 2008.
7. ——, "Blind speech dereverberation using batch and sequential monte carlo methods," in *Proc. IEEE Conf. ISCAS*, Seattle, WA, 18-21 May 2008.
8. C. Evers and J. R. Hopgood, "Marginalization of static observation parameters in a Rao-Blackwellized particle filter with application to sequential blind speech dereverberation," in *Proc. EUSIPCO*, Glasgow, UK, Aug. 2009.
9. G. Casella and C. P. Robert, "Rao-Blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.
10. 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.
11. T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, Jul. 2005.
12. B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, April 1988.
13. S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, 1997.
14. J. L. Flanagan, A. C. Surendran, and E. E. Jan, "Spatially selective sound capture for speech and audio processing," *Speech Commun.*, vol. 13, no. 1-2, pp. 207–222, Oct. 1993.
15. J. B. Allen, "Synthesis of pure speech from a reverberant signal," U.S. Patent No. 3786188, Jan. 1974.
16. M. S. Brandstein, "On the use of explicit speech modeling in microphone array applications," in *Proc. IEEE Conf. ICASSP*, vol. 6, Seattle, WA, May 1998, pp. 3613–3616.
17. S. Griebel and M. Brandstein, "Wavelet transform extrema clustering for multi-channel speech dereverberation," in *Proc. IEEE Conf. WASPAA*, 1999, pp. 27–30.
18. B. Yegnanarayana, S. R. M. Prasanna, and K. S. Rao, "Speech enhancement using excitation source information," in *Proc. IEEE Conf. ICASSP*, vol. 1, Orlando, FL, May 2002, pp. 541–544.
19. T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," in *Proc. IEEE Conf. ICASSP*, vol. 1, 2003, pp. 92–95.
20. T. Nakatani, M. Miyoshi, and K. Kinoshita, "Single-microphone blind dereverberation," in *Speech enhancement*, ser. Signals and Commnication Technology, J. Benesty, S. Makino, and J. Chen, Eds. Springer, 2005, pp. 247–270.
21. T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity-based blind dereverberation for single-channel speech signals," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 1, pp. 80–95, Jan. 2007.
22. B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 728–737, Nov. 2000.
23. E. A. P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proc. IEEE Conf. ICASSP*, vol. 4, Mar. 2005, pp. 173–176.
24. E. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, Eindhoven, Netherlands, Jun. 2007.
25. P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Proc. IEEE Conf. IWAENC*, Eindhoven, Netherlands, 2005.
26. J. V. Candy, *Model-based signal processing*. New Jersey, NJ: John Wiley & Sons, 2006.
27. J. L. Kelly and C. C. Lochbaum, "Speech synthesis," in *Proc. Int. Congress Acoustics*, Copenhagen, Denmark, 1962, pp. 1–4.
28. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
29. J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-time processing of speech signals*. Englewood Cliffs, NJ: Macmillan Publishing Company, 1993.

30. E. Bognar and H. Fujisaki, "Analysis, synthesis and perception of the French nasal vowels," in *Proc. IEEE Conf. ICASSP*, vol. 11, Apr. 1986, pp. 1601–1604.
31. J. Vermaak, M. Niranjan, and S. J. Godsill, "An improved speech production model for voiced speech utilising a seasonal AR-AR model and Markov Chain Monte Carlo simulation," Cambridge University, UK, CUED/F-INFENG/TR.325, June 1998.
32. C. W. Therrien, *Discrete random signals and statistical signal processing*, ser. Signal processing series, A. V. Oppenheim, Ed. Englewood Cliffs, NJ: Prentice Hall, 1992.
33. 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.
34. J. N. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound Vibr.*, vol. 102, no. 2, pp. 217–228, September 1985.
35. B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter - Particle Filters for Tracking Applications*. Artech House, 2004.
36. A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo methods in practice*. New York: Springer, 2000.
37. A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering." *Statist. Comp.*, vol. 10, pp. 197–208, 2000.
38. A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Stat. Assoc.*, vol. 89, pp. 278–288, 1994.
39. R. Douc, O. Cappe, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Proc. IEEE Conf. ISPA*, 2005, pp. 64–69.
40. V. S. Zaritskii, V. B. Svetnik, and L. I. Shimelevich, "Monte Carlo technique in problems of optimal data processing," *Automation and Remote Control*, vol. 12, pp. 95–103, 1975.
41. P. S. Spencer, "System identification with application to the restoration of archived gramophone recordings," PhD Thesis, University of Cambridge, UK, Jun. 1990.
42. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
43. G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, Mar. 1952.
44. P. P. Vaidyanathan, *Multirate systems and filter banks*. New Jersey, NJ: Prentice Hall, 1993.
45. J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2730–2743, Nov. 2002.
46. M. J. Daly and J. R. Reilly, "Blind deconvolution using Bayesian methods with application to the dereverberation of speech," in *Proc. IEEE Conf. ICASSP*, vol. 2, May 2004, pp. 1009–1012.
47. N. Gaubitch, X. S. Lin, and P. A. Naylor, "Scale factor ambiguity correction for subband multichannel identification," in *Proc. IEEE Conf. IWAENC*, Seattle, WA, Sep. 2008.