Blind Single Channel Deconvolution using Nonstationary Signal Processing

Blind Single Channel Deconvolution using Nonstationary Signal Processing

James R. Hopgood*, *Member, IEEE*, and Peter J. W. Rayner

*Corresponding author. e-mail: jrh1008@eng.cam.ac.uk
This work was supported by the Schiff Foundation, University of Cambridge.
The authors are with the Signal Processing Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge. CB2 1PZ. UK. Tel: +44 1223 335562. Fax: +44 1223 332662. e-mail: {jrh1008,pjwr}@eng.cam.ac.uk
EDICS: 1-ENHA

Abstract

Blind deconvolution is fundamental in signal processing applications and, in particular, the single channel case remains a challenging and formidable problem. This paper considers single channel blind deconvolution in the case where the degraded observed signal may be modelled as the convolution of a nonstationary source signal with a stationary distortion operator. The important feature that the source is nonstationary while the channel is stationary facilitates the unambiguous identification of either the source or channel, and deconvolution is possible, whereas if the source and channel are both stationary, identification is ambiguous. The parameters for the channel are estimated by modelling the source as a time-varying AR process and the distortion by an all-pole filter, and using the Bayesian framework for parameter estimation. This estimate can then be used to deconvolve the observed signal. In contrast to the classical histogram approach for estimating the channel poles, where the technique merely relies on the fact that the channel is actually stationary rather than modelling is as so, the proposed Bayesian method does take account for the channel's stationarity in the model and, consequently, is more robust. The properties of this model is investigated, and the advantage of utilising the nonstationarity of a system rather than considering it as a curse is discussed.

Keywords

nonstationary processes, single channel blind deconvolution, speech dereverberation

I. Introduction

Blind deconvolution is fundamental in signal processing applications and, in particular, the single channel case remains a challenging and formidable problem. In single channel blind deconvolution, a degraded observed signal, $\mathbf{x} = \{x(t),\, t \in \mathcal{T}\} \in \mathbb{R}^T$, $\mathcal{T} = \{1, \ldots, T\} \subset \mathbb{Z}$, is modelled as the convolution of an unknown source signal, $\mathbf{s} = \{s(t),\, t \in \mathcal{T}\} \subset \mathbb{R}^T$, with an unknown distortion operator, $\mathcal{A}$. The distortion operator could, for example, represent the acoustical properties of a room, the effect of *multipath propagation* in the reception of radio, or a non-impulsive excitation in seismic applications. Acoustic reverberation [1] is a cause of significant degradation in speech intelligibility for users of hearing aids [2] and is a problem in 'hands-free' telephony and speech recognition. The process of removing the effects of reverberation is known as *blind dereverberation*, and belongs to the general problem of blind deconvolution. The task is to estimate the distortion, $\mathcal{A}$, *or* a scaled shifted version of the original signal, $\hat{s}(t) = a\, s(t - \tau)$, where $a,\, \tau \in \mathbb{R}$, given only the observed signal, $\mathbf{x}$. Throughout this paper, except where indicated, the set notation $\mathcal{G} = \{1, \ldots, G\}$ is used; *e.g.* $\mathcal{U}_i = \{1, \ldots, U_i\}$. Further, if $\gamma \triangleq \{\alpha,\, \beta\}$, then $\gamma_{-\beta} \triangleq \{\alpha\}$.

A. Nonstationary Signal Processing

Signal processing techniques over the past three decades have been dominated by the constraint of *stationarity* – an appealing notion as many processes are endowed with the property of ergodicity which allows quantities defined as ensemble averages, for example, autocorrelation functions, to be estimated from a single realisation of the process by calculating time domain averages. However, since the statistical properties of many processes, including speech, are nonstationary, many existing techniques which assume stationarity produce poor results when applied to audio restoration and other problems [3]. Since the estimation of the autocorrelation function of a nonstationary processes is difficult, often requiring multiple data records which may not always be available, nonstationarity is usually regarded as an undesirable feature. Nevertheless, it has recently been recognised that nonstationarity can actually be a useful feature, and can be utilised to produce superior results both in existing problems attempted using the stationarity assumption, and previously intractable problems [4–7]. The advantage of utilising the nonstationarity of a system rather than considering it as a curse is demonstrated in this paper, where solutions are presented for the single channel blind deconvolution problem which could not be obtained if nonstationarity was not taken into account.

B. Characteristics of Blind Deconvolution

Many blind deconvolution techniques assume the source signal is contained within a finite support [8–10] and, or, are independent and identically distributed (i. i. d.) [9, 11]. However, when the source signal is highly correlated and belongs to a set of infinite support, these techniques cannot be directly applied. Furthermore, many techniques assume *quasi-stationarity* of the system, and do not take *global nonstationarity* into account. Utilising the global nonstationarity of a system allows the identification of system characteristics which may otherwise be unattainable.

The problem is under-constrained and *can only* be solved by incorporating varying degrees of prior knowledge regarding $s(t)$ and $\mathcal{A}$. A characteristic of blind deconvolution is that the source signal and impulse response of the distortion operator must be *irreducible* for unambiguous deconvolution [12]. An (real) irreducible signal is one that cannot be expressed as the convolution of two or more signal components, on the understanding that the delta function is not a signal component. Suppose the distortion operator, $\mathcal{A}$, is linear time-invariant (LTI), then the observed signal may be expressed as $x(t) = h(t) \star s(t)$, where $\star$ denotes convolution. If either $h(t)$ or $s(t)$

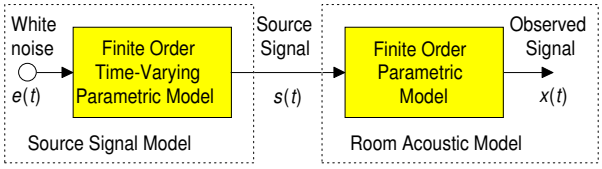


Fig. 1. General parametric model for blind deconvolution.

are *reducible* such that $h(t) = h_1(t) \star h_2(t)$, and $s(t) = s_1(t) \star s_2(t)$, then $x(t) = h_1(t) \star h_2(t) \star s_1(t) \star s_2(t)$, and it is impossible to decide which component belongs to the source signal or to the distortion operator without additional knowledge. Consequently, many linear systems become reducible when they are considered stationary, and *blind deconvolution* is impossible. However, if, in fact, $s(t)$ and $\mathcal{A}$ are both quasi-stationary signals and *locally reducible*, but possess different *rates* of *global* nonstationarity, then $s(t)$ and $\mathcal{A}$ are no longer *globally reducible* and, therefore, in this case blind deconvolution is possible.

There are two distinct approaches to this problem: 1] estimate $s(t)$ as a 'missing data' problem by treating the parameters of $\mathcal{A}$ as *nuisance parameters*, or, 2] estimate the parameters of $\mathcal{A}$ by treating $s(t)$ as a nuisance parameter, and then deconvolve $x(t)$ with $\mathcal{A}$ to recover $s(t)$. The second approach is considered using parametric models for $s(t)$ and $\mathcal{A}$. The resulting model is shown in Figure 1.

II. System Models

A. Nonstationary Models for Source Signal

For many time-series, the limitation of assuming a signal is stationary often results in poor modelling and, in such cases, stationary models can prove ineffective for some applications. The most common approach to modelling nonstationary processes is to represent the signal in the form of a stationary model, commonly the autoregressive moving average (ARMA) model [13–15], and to represent the time-varying *parameters* either as a linear combination of deterministic time-varying basis functions, or as an unobserved random process. The choice of basis functions, of which many have been proposed [14, 16, 17], is dependent on any prior belief of the variation of the parameters [18, 19] and, without this knowledge, there exists no general rule for choosing these functions. However, while a suitable nonstationary model must reflect the statistics of the source signal, it is proposed that the sensitivity of a chosen model on the nature of the underlying signal structure decreases when the additional information inherent in the nonstationarity of the

process is utilised.

The form for a time-varying ARMA (TVARMA) model of a time-series $s(t)$ is given in terms of a zero-mean *nonstationary* white noise process, $e(t) \sim \mathcal{N}\left(0,\, \sigma_{e,t}^2\right)$, as:

$$s(t) = -\sum_{q=1}^{Q_t} b(t,q)\, s(t-q) + \sum_{r=0}^{R_t} c(t,r)\, e(t-r) \tag{1}$$

where $\{b(t,q),\, q \in \mathcal{Q}_t,\, t \in \mathcal{T}\}$ and $\{c(t,r),\, r \in \mathcal{R}_t,\, t \in \mathcal{T}\}$ are the time-varying model parameters; note, to avoid scaling ambiguities, $\{c(t,0) \triangleq 1; t \in \mathcal{T}\}$. The most general case of this model is where the parameters are completely uncorrelated at each sample and, therefore, each sample of $s(t)$ would be represented by $Q_t + R_t + 1$ unknown coefficients. Since the parameter space is increased at each sample by $Q_t + R_t$, numerical problems result as there isn't enough data, at least from a single realisation of a process, to facilitate parameter estimation. Ergo, a practical case is one where the parameters are highly correlated, and one such case is for stochastic processes that are *globally* nonstationary, yet approximately *locally* stationary; these can be represented by a *quasi-stationary model*.

A time-varying AR (TVAR) process is proposed for modelling a wide range of input signals; this model is particularly appropriate for speech signals [14, 17, 20, 21]. Specifically, $s(t)$, is modelled by a block stationary AR (BSAR) process. Here, $s(t)$ is partitioned into M contiguous disjoint blocks, block $i \in \mathcal{M}$ beginning at sample t_i with length $T_i = t_{i+1} - t_i$ and, in this block, is given by a stationary AR model of order Q_i. Using (1), this is equivalent to setting $Q_t = Q_i$, $R_t = 0$, $b(t,q) = b_i(q)$, $q \in \mathcal{Q}_i$, $\sigma_t = \sigma_i$, $\forall t \in \mathcal{T}_i = \{t_i, \ldots, t_{i+1} - 1\} \subset \mathbb{Z}^{T_i}$ and $i \in \mathcal{M}$, yielding:

$$s(t) = -\sum_{q=1}^{Q_i} b_i(q)\, s(t-q) + e(t) \tag{2}$$

where $e(t) \sim \mathcal{N}\left(0,\, \sigma_i^2\right)$, $\sigma_i \in \mathbb{R}^+$, $t \in \mathcal{T}_i$, $i \in \mathcal{M}$.

B. Time-Invariant Channel Model

In this paper, the channel model is restricted to ones which are linear time-invariant. The use of a linear time-varying (LTV) model will be considered in future work. The all-pole model is used for the channel since, not only is it mathematically convenient, but it is widely used in many fields for approximating rational transfer functions. The distortion operator, $\mathcal{A}$, is modelled by

a LTI all-pole filter of order P, such that:

$$x(t) = -\sum_{p=1}^{P} a(p)\, x(t-p) + s(t), \quad t \in \mathbb{Z} \tag{3}$$

where $\mathbf{a} = \{a(p),\, p \in \mathcal{P}\}$ is the set of P model parameters. The appropriateness of this model for the particular application of modelling room acoustics is discussed in §VI-A.

III. Bayesian Parameter Estimation

The *posterior probability*, $p\left(\boldsymbol{\theta} \,|\, \mathbf{x}, \mathcal{I}\right)$, of a particular set of system parameters, $\boldsymbol{\theta}$, given the state of the system, $\mathbf{x}$, and an underlying model, $\mathcal{I}$, is given by Bayes's theorem:

$$p\left(\boldsymbol{\theta} \,|\, \mathbf{x}, \mathcal{I}\right) = \frac{p\left(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathcal{I}\right) p\left(\boldsymbol{\theta} \,|\, \mathcal{I}\right)}{p\left(\mathbf{x} \,|\, \mathcal{I}\right)} \tag{4}$$

where $p\left(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathcal{I}\right)$ is the *likelihood* function and $p\left(\boldsymbol{\theta} \,|\, \mathcal{I}\right)$ represents any prior belief. The term $p\left(\mathbf{x} \,|\, \mathcal{I}\right)$ is the *evidence* and, although usually regarded as a normalising constant, is of interest for model selection. The choice of priors for the system model introduced in §II are described below.

A. Prior distribution on AR coefficients

In the case when a process is modelled by a real, stable, minimum-phase AR process of order P with parameters $\mathbf{a}$ and excitation variance σ^2, the parameter vector, $\mathbf{a}$, should ideally only take on values which lie in the *stability domain*. However, the terms in the likelihood function for the AR parameters are usually in the form of a Gaussian distribution (see [22]) and, in order to obtain analytically tractable results, a Gaussian prior is placed on the parameter values: $\mathbf{a} \,|\, \sigma^2 \sim \mathcal{N}\left(\mathbf{0}_P,\, \sigma^2\, \delta^2\, \mathbf{I}_P\right),\, \delta \in \mathbb{R}^+$, where the prior on $\mathbf{a}$ becomes uninformative as the hyperparameter, $\delta \to \infty$. $\mathbf{I}_P \in \mathbb{R}^{P \times P}$ is the identity matrix.

B. Prior distribution for the Excitation Variance

A standard prior for application to scale parameters, such as variances, is the inverse-Gamma density with the form $\sigma^2 \,|\, \nu, \gamma \sim \mathcal{IG}\left(\sigma^2,\, \nu\right) \gamma$, where:

$$\mathcal{IG}\left(\sigma^2 \,|\, \nu,\, \gamma\right) = \frac{\gamma^\nu}{\Gamma(\nu)} (\sigma^2)^{-(\nu+1)} \exp\left[-\frac{\gamma}{\sigma^2}\right] \mathbb{I}_{(0,+\infty]}\left(\sigma^2\right) \tag{5}$$

where ν, γ are the hyperparameters, and $\mathbb{I}_{\mathcal{A}}\left(a\right) = 1$ if $a \in \mathcal{A}$ and zero otherwise.

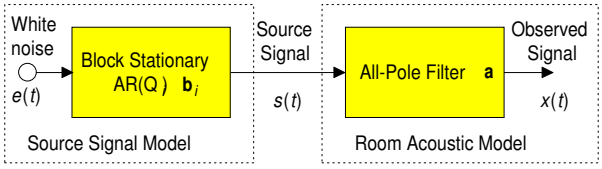


Fig. 2. Complete system model; the output, $x(t)$, is a BSAR process, $s(t)$, filtered by all-pole model with parameters $\mathbf{a}$.

IV. Histogram Technique

In this section, the histogram technique introduced in [23, 24], and recently implemented in real time [25], is used to estimate the parameters for the distortion filter, $\mathcal{A}$. This technique relies on the fact that the channel is stationary, while the source is nonstationary. However, since the channel is not *explicitly* modelled as stationary, a technique which explicitly models the stationarity will perform better, and such an approach is outlined in §V.

A. Principle Underpinning Histogram Technique

Consider, in block i, the output, $x(t)$, of a BSAR process, $s(t)$, filtered by an all-pole model, shown in Figure 2, as a windowed version of the infinite stationary sequence, $x_i(t)$:

$$\left.\begin{aligned} x_i(t) &= -\sum_{p=1}^{P} a(p)\, x_i(t-p) + s_i(t) \\ s_i(t) &= -\sum_{q=1}^{Q_i} b_i(q)\, s_i(t-q) + e_i(t) \end{aligned}\right\} \; t \in \mathcal{T} \subset \mathbb{Z} \tag{6}$$

where $e_i(t) \sim \mathcal{N}\left(0,\, \sigma_i^2\right)$, such that $x(t) = x_i(t)$, $\forall t \in \mathcal{T}_i \subset \mathcal{T}$. The power spectral density (PSD), $\mathcal{P}_i(e^{j\omega})$, of $x_i(t)$ is:

$$\mathcal{P}_i(e^{j\omega}) \equiv \frac{1}{\prod_{p=1}^{P} \left|1 - r_{\mathbf{a}}(p)\, e^{-j\omega}\right|^2} \frac{\sigma_i^2}{\prod_{q=1}^{Q_i} \left|1 - r_{\mathbf{b}_i}(q)\, e^{-j\omega}\right|^2}$$

where $\mathbf{r_a} \equiv [\, r_{\mathbf{a}}(1) \;\; \ldots \;\; r_{\mathbf{a}}(P)\,]$ are the roots of the all-pole filter, $\mathcal{A}$, and $\mathbf{r}_{\mathbf{b}_i} \equiv [\, r_{\mathbf{b}_i}(1) \;\; \ldots \;\; r_{\mathbf{b}_i}(Q_i)\,]$ are the roots of the TVAR source model. This can be written as:

$$\mathcal{P}_i(e^{j\omega}) = \frac{\sigma_i^2}{\left|1 + \sum_{k=1}^{R_i} c_i(k)\, e^{-j\omega k}\right|^2} \equiv \frac{\sigma_i^2}{\prod_{k=1}^{R_i} \left|1 - r_i(k)\, e^{-j\omega}\right|^2}$$

where $R_i = P + Q_i$, $c_i(k) = [a(p) \star b_i(q)]_k$, and $\mathbf{r}_i \equiv [\, r_i(1) \;\; \ldots \;\; r_i(R_i)\,] = \{\mathbf{r_a}, \mathbf{r_{b_i}}\}$ are the roots of the combined AR model. In each block, suppose there is an estimate of $\mathbf{r}_i$ denoted by $\hat{\mathbf{r}}_i = \{\hat{\mathbf{r}}_{\mathbf{a}_i}, \hat{\mathbf{r}}_{\mathbf{b}_i}\}$, where $\hat{\mathbf{r}}_{\mathbf{a}_i}$ is an estimate of $\mathbf{r_a}$ using the data in block i and, correspondingly, $\hat{\mathbf{r}}_{\mathbf{b}_i}$ is an estimate of $\mathbf{r}_{\mathbf{b}_i}$. Initially, it may appear impossible, without considerable prior knowledge, to partition each pole set, $\hat{\mathbf{r}}_i$, into subsets separately containing $\hat{\mathbf{r}}_{\mathbf{a}_i}$ and $\hat{\mathbf{r}}_{\mathbf{b}_i}$. However, consider constructing the set of all poles obtained by concatenating the poles estimated from each block, $i \in \mathcal{M}$; *i.e.* consider the set $\mathcal{R} \triangleq \{\hat{\mathbf{r}}_i, i \in \mathcal{M}\}$. If all the poles in the set $\mathcal{R}$ are compared simultaneously, a number of poles will be contained in a number of *small local regions of support*, with each of these subsets containing estimates that are just statistical variations of the same pole. Furthermore, there will be many subsets which contain just a few pole estimates; the number of estimates within each subset obviously depends on the size of the local region of support. It can be concluded that where the number of estimates within a particular fixed region is large, the corresponding subset represents a *stationary* pole, and where the number is small, the subset represents a *nonstationary* pole. If $s(t)$ is modelled as a TVAR process and is known to be comprised only of *nonstationary* poles, while $\mathcal{A}$ is known to be stationary, the channel can be estimated from some estimate of the *stationary* poles based on the subsets which contain a large number of pole estimates.

B. Probablistic Framework

To formalise the histogram method in a probablistic framework, consider a process given by (6), defined for $t \in \mathcal{T}_i = \{t_i, \ldots, t_i + T_i\}$, where T_i is the block length, and $\mathcal{T}_i \cap \mathcal{T}_j \neq \{\emptyset\}$ so that the blocks are overlapping with spacing $L_i = T_{i+1} - T_i$. The excitation, $e(t)$, in block i is related to the observed signal, $x(t)$, by $\mathbf{e}_i = \mathbf{x}_i + \mathbf{X}_i\,\mathbf{c}_i$, where $\mathbf{e}_i$ is a vector of samples $[\mathbf{e}_i]_{t-t_i+1} = e(t)$, $t \in \mathcal{T}_i$, and, similarly, $[\mathbf{x}_i]_{t-t_i+1} = x(t)$ and $[\mathbf{X}_i]_{t-t_i+1,q} = x(t-q)$, $t \in \mathcal{T}_i, q \in \mathcal{Q}_i$ and $[\mathbf{c}_i]_k = c_i(k)$, $k \in \mathcal{R}_i$ are the parameters of the combined AR model. The Jacobian $J(\mathbf{x}_i, \mathbf{e}_i)$ is unity and, thus, the likelihood function in block i is given by:

$$p\left(\mathbf{x}_i \,|\, \mathbf{c}_i,\, \sigma_i^2,\, \mathbf{x}_{i-1}\right) = \mathcal{N}\left(\mathbf{e}_i \,\middle|\, 0,\, \sigma_i^2\,\mathbf{I}_{T_i}\right) \tag{7}$$

Assuming again the prior distributions in §III:

$$\begin{aligned} \mathbf{c}_i \,|\, \sigma_i^2 &\sim \mathcal{N}\left(\mathbf{0}_{R_i},\, \delta_i^2\,\sigma_i^2 \mathbf{I}_{R_i}\right),\, \delta_i \in \mathbb{R}^+, \\ \sigma_i^2 &\sim \mathcal{IG}\left(\frac{\nu_i}{2},\, \frac{\gamma_i}{2}\right) \end{aligned} \tag{8}$$

with δ_i, ν_i and γ_i are hyperparameters then, using Bayes's rule in (4), it follows:

$$p\left(\mathbf{c}_i,\, \sigma_i^2 \,\middle|\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right) \propto \frac{1}{\sigma_i^{\hat{R}_i}} \exp\left\{-\frac{\gamma_i + \mathbf{e}_i^T\,\mathbf{e}_i + \delta_i^{-2}\,\mathbf{c}_i^T\,\mathbf{c}_i}{2\sigma_i^2}\right\} \tag{9}$$

where $\hat{R}_i = T_i + R_i + \nu_i + 2$. The excitation variance can be marginalised using the standard Gamma integral [26] to give the conditional density $p\left(\mathbf{c}_i \,|\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right)$. Thus, by sampling and histogramming the variates $\{\mathbf{c}_i,\, i \in \mathcal{M}\}$ from this distribution, estimates of the parameter $\mathbf{a}$ can be obtained as described in §IV-A.

However, it is difficult to sample the parameters $\mathbf{c}_i$ from the resulting distribution and, therefore, the *Gibbs sampler* is employed. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) technique that allows samples to be drawn from complicated probability density functions (pdfs) by drawing samples from simpler conditional densities. This sampling method is described in detail in, for example, [27]. In order to implement the Gibbs sampler, the conditional probabilities $p\left(\mathbf{c}_i \,|\, \sigma_i^2,\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right)$ and $p\left(\sigma_i^2 \,\middle|\, \mathbf{c}_i,\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right)$ must be easily sampled. Using Bayes's rule and the given priors, it is easy to show that $p\left(\mathbf{c}_i \,|\, \sigma_i^2,\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right)$, $i \in \mathcal{M}$, is given by a multivariate Gaussian with inverse covariance:

$$\mathbf{C}_i^{-1} = \frac{\mathbf{X}_i^T\,\mathbf{X}_i + \delta_i^{-2}\mathbf{I}_{R_i}}{\sigma_i^2}$$

and mode:

$$\hat{\mathbf{c}}_i = -\left(\mathbf{X}_i^T\,\mathbf{X}_i + \delta_i^{-2}\mathbf{I}_{R_i}\right)^{-1}\mathbf{X}_i^T\,\mathbf{x}_i$$

while,

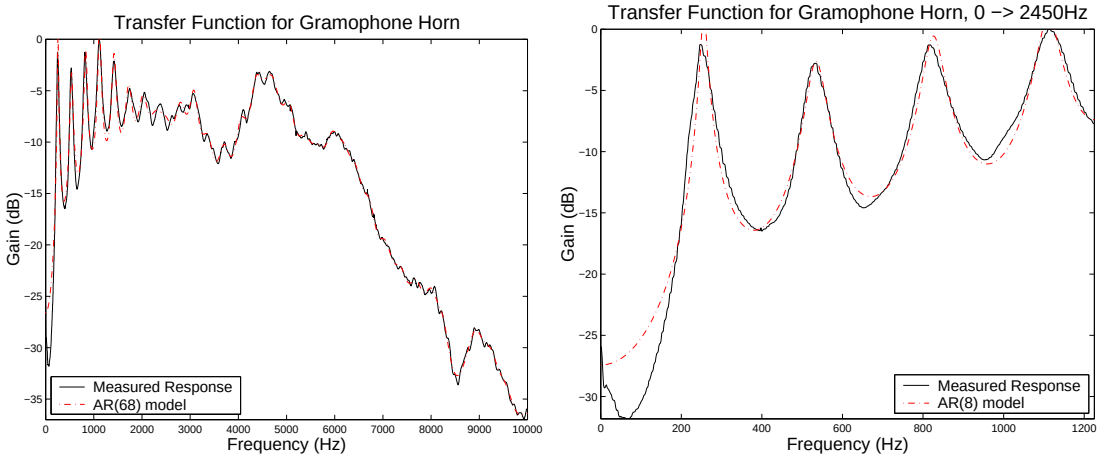
$$p\left(\sigma_i^2 \,\middle|\, \mathbf{c}_i,\, \mathbf{x}_i,\, \mathbf{x}_{i-1}\right) = \mathcal{IG}\left(\sigma_i^2 \,\middle|\, \frac{T_i + \nu_i}{2},\, \frac{\mathbf{e}_i^T\,\mathbf{e}_i + \gamma_i}{2}\right)$$

Details for sampling these distribution is discussed in [27].

C. Examples

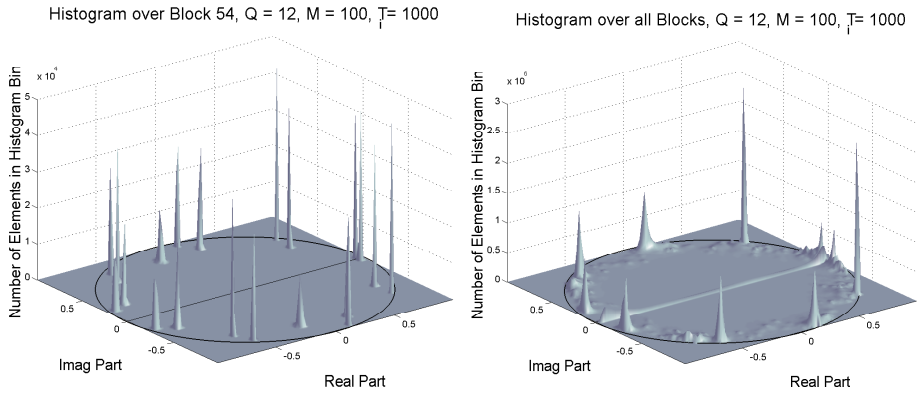
As a typical channel for a system, consider the frequency response of an acoustic gramophone horn as measured in [24]. The magnitude response is shown in Figure 3(a), and using a maximum-likelihood (ML) estimator for model selection, Spencer [24] showed that this response can be accurately modelled by a 68th-order all-pole model. Moreover, the response upto 2.45 kHz can be accurately modelled by an 8th-order model, as shown in Figure 3(b).

A synthetic BSAR process is used to model the source signal. The parameters for this process are estimated by modelling discontiguous blocks of a real speech signal as a stationary AR(Q) model and estimating the parameters using the autocorrelation method [20, 28]. This ensures



(a)Response, $0 \rightarrow 10.00$ kHz (b)Response, $0 \rightarrow 2.45$ kHz

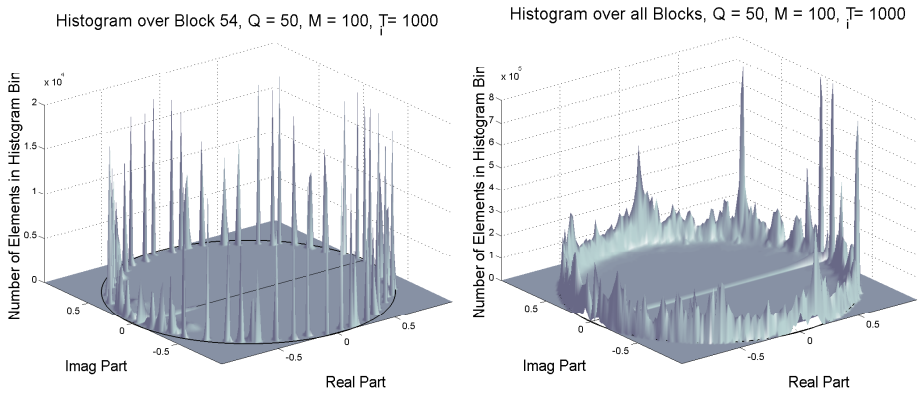
Fig. 3. Power Spectral Density of Signal Recorded at Neck of Horn, and the corresponding AR model.



(a)Histogram of samples in *one* block (b)Histogram of samples over *all* block

Fig. 4. Histogram for case when source model order $Q = 12$, number of blocks $N = 100$, and length of each block $T_i = 1000$.

that the synthetic data, generated in contiguous blocks using these parameters, is stable, non-stationary, and partially reflects the statistical properties of a real signal. The model orders of both the filter and source are assumed to be known. Samples are drawn from the posterior distribution given by (9) as discussed in §IV-B. Figure 4 shows the histogram of all the samples on a grid covering the unit circle for the case when the source model order $Q = 12$, number of blocks $M = 100$, and length of each block $T_i = 1000$. Figure 4(a) shows the histogram of samples drawn from a single block, where it is seen that the peaks are located at the positions of the BSAR poles, as well as the positions of the poles due to the filter; clearly, it is impossible



(a)Histogram of samples in *one* block (b)Histogram of samples over *all* block

Fig. 5. Histogram for case when source model order $Q = 50$, number of blocks $N = 100$, and length of each block $T_i = 1000$.

to distinguish the channel from the source, when a stationary model is used for the complete system. However, Figure 4(b) shows the histogram of the samples from all blocks, and it is seen that the peaks are now located at the position of the stationary poles which must belong to the filter and to any poles representing the stationary components of the source signal. In this case, there are no dominant peaks due to the source signal, and the channel has been identified.

If the source model order is increased to $Q = 50$, as in Figure 5, the pole-plane becomes highly crowded, as emphasised in Figure 5(a). Thus, as shown in Figure 5(b), the stationary peaks with high variance become swamped by the considerably larger number of source poles. Hence, in the case of a large difference between the number of source and channel parameters, the histogram model fails.

V. Bayesian Blind Deconvolution

The histogram technique *does not* constrain $\mathcal{A}$ to be stationary across block boundaries. It is by virtue of the fact that the filter *is actually* stationary that the technique can detect the filter parameters by considering the system poles. A more robust method of parameter estimation is to account for the filter's stationarity, and such an approach is demonstrated in this section. This probabilistic approach was first presented in [6] and [7], and it will be shown that channels can be estimated with model orders considerably higher than can be dealt with by the histogram technique.

A. Posterior Distribution for Channel Parameters

The source signal, $s(t)$, is given by (2) and, therefore, the excitation samples in block $i \in \mathcal{M}$ may be written as $\mathbf{e}_i = \mathbf{s}_i + \mathrm{S}_i\,\mathbf{b}_i$, where $[\mathbf{e}_i]_{t-t_i+1} = e(t)$, $t \in \mathcal{T}_i$, and, similarly, $[\mathbf{s}_i]_{t-t_i+1} = s(t)$, $t \in \mathcal{T}_i$, $\mathbf{b}_i$ is a vector of parameters $[\mathbf{b}_i]_q = b_i(q)$, $q \in \mathcal{Q}_i$, and the data matrix $[\mathrm{S}_i]_{t-t_i+1,q} = s(t-q)$, $t \in \mathcal{T}_i, q \in \mathcal{Q}_i$. The probability distribution for the excitation in block i is given by $\mathbf{e}_i \sim \mathcal{N}\left(\mathbf{0}_{T_i},\, \sigma_i^2\,\mathbf{I}_{T_i}\right)$. Since $p\left(\mathbf{s} \mid \mathcal{I}\right) = p\left(\mathbf{s}_i,\, i \in \mathcal{M} \mid \mathcal{I}\right)$, and the BSAR process depends only on the previous Q_i outputs, such that, iff $Q_i \leq T_i$, $p\left(\mathbf{s}_i \mid \mathbf{s}_{i-1},\, \ldots,\, \mathbf{s}_1,\, \mathcal{I}\right) = p\left(\mathbf{s}_i \mid \mathbf{s}_{i-1},\, \mathcal{I}\right)$, then the probability *chain rule* identity can be written as:

$$p\left(\mathbf{s} \mid \mathcal{I}\right) = p\left(\mathbf{s}_1 \mid \mathcal{I}\right) \prod_{i=2}^{M} p\left(\mathbf{s}_i \mid \mathbf{s}_{i-1},\, \mathcal{I}\right) \tag{10}$$

Denoting $\boldsymbol{\sigma} = \{\sigma_i^2,\, i \in \mathcal{M}\}$, $\mathbf{b} = \{\mathbf{b}_i,\, i \in \mathcal{M}\}$, $\boldsymbol{\theta} = \{\mathbf{a},\, \boldsymbol{\sigma}, \mathbf{b}\}$, the likelihood function for the source signal, $\mathbf{s}_i$, in block $i \in \mathcal{M}_{\{-1\}}$, where $\mathcal{M}_{\{-1\}}$ denotes the set $\mathcal{M}$ not including the element 1, is:

$$p\left(\mathbf{s}_i \mid \mathbf{s}_{i-1}, \boldsymbol{\theta}_{-\mathbf{a}},\, \mathcal{I}\right) = \frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{-\frac{\|\mathbf{s}_i + \mathrm{S}_i\mathbf{b}_i\|^2}{2\sigma_i^2}\right\} \tag{11}$$

where $i \in \mathcal{M}_{\{-1\}}$, and $\|\cdot\|$ denotes the Euclidean norm. Note $p\left(\mathbf{s}_1 \mid \mathcal{I}\right)$ is dependent on the initial values, $\mathbf{s}_0$, of $\mathbf{s}$ and, therefore, has a different distribution to (11), as discussed in [21, 22]. However, if $Q_1 \ll T_1$, which is often the case with audio signals, it is common practice to approximate $p\left(\mathbf{s}_1 \mid \mathcal{I}\right)$ by (11). Using (10), the likelihood function for the observed signal, $\mathbf{x}$, is:

$$p\left(\mathbf{x} \mid \boldsymbol{\theta},\, \boldsymbol{\phi},\, \mathcal{I}\right) = \prod_{i=1}^{M} p\left(\mathbf{s}_i \mid \mathbf{s}_{i-1}, \boldsymbol{\theta}_{-\mathbf{a}},\, \mathcal{I}\right) \tag{12}$$

where $s(t) \equiv s(t, \mathbf{a},\, \mathbf{x})$ is given by the relationship $\mathbf{x} = f(\mathbf{s}, \mathbf{a})$, and $\boldsymbol{\phi} = \{\boldsymbol{\tau},\, \boldsymbol{\Xi},\, \boldsymbol{\delta},\, \boldsymbol{\nu},\, \boldsymbol{\gamma}\}$ contains the vector of changepoints, $\boldsymbol{\tau} = \{t_i,\, i \in \mathcal{M}\}$, the vector of model orders, $\boldsymbol{\Xi} = \{Q_i,\, i \in \mathcal{M}\}$, and the vectors of hyperparameters, $\boldsymbol{\delta} = \{\delta_i,\, i \in \mathcal{M}\}$, $\boldsymbol{\nu} = \{\nu_i,\, i \in \mathcal{M}\}$, and $\boldsymbol{\gamma} = \{\gamma_i,\, i \in \mathcal{M}\}$ as defined in the assigned priors below. Applying Bayes's rule, the posterior pdf for the unknown parameters $\boldsymbol{\theta}$ becomes:

$$p\left(\boldsymbol{\theta} \mid \mathbf{x},\, \boldsymbol{\phi},\, \mathcal{I}\right) \propto\; p\left(\mathbf{x} \mid \boldsymbol{\theta},\, \boldsymbol{\phi},\, \mathcal{I}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{\phi},\, \mathcal{I}\right)$$

assuming $\boldsymbol{\phi}$ is known. Assuming $\{b_i,\, \sigma_i\}$ are independent between blocks, the assigned priors are $\mathbf{b}_i \,|\, \sigma_i^2 \sim \mathcal{N}\left(\mathbf{0}_{Q_i},\, \sigma_i^2\,\delta_i^2\,\mathbf{I}_{Q_i}\right)$, $\delta_i \in \mathbb{R}^+$, and $\sigma_i^2 \sim \mathcal{IG}\left(\frac{\nu_i}{2},\, \frac{\gamma_i}{2}\right)$, $i \in \mathcal{M}$. Hence:

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{\phi},\, \mathcal{I}\right) = p\left(\mathbf{a} \mid \boldsymbol{\phi},\, \mathcal{I}\right)\, p\left(\mathbf{b} \mid \boldsymbol{\sigma},\, \boldsymbol{\phi},\, \mathcal{I}\right)\, p\left(\boldsymbol{\sigma} \mid \boldsymbol{\phi},\, \mathcal{I}\right)$$

Since it is only of interest to estimate the channel parameters, $\mathbf{a}$, the *nuisance* parameters, $\mathbf{b}$ and $\boldsymbol{\sigma}$, can be *marginalised* from equation (12) by integrating over $\boldsymbol{\theta}_{-\mathbf{a}}$, as shown in the Appendix, yielding the posterior density:

$$\begin{aligned} p\left(\mathbf{a} \mid \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) &\propto p\left(\mathbf{a} \mid \boldsymbol{\phi}, \mathcal{I}\right) \\ &\times \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T \mathrm{S}_i \left(\mathrm{S}_i^T \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right)^{-1} \mathrm{S}_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|\mathrm{S}_i^T \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right|^{\frac{1}{2}}} \end{aligned} \tag{13}$$

where $R_i = \frac{T_i + \nu_i + 1}{2}$, $i \in \mathcal{M}$. Equation (13) is written in terms of $\mathbf{s}(t)$ to emphasise that the posterior can be efficiently calculated by 'inverse filtering' the data, $\mathbf{x}(t)$, before performing matrix products. A maximum marginal *a posteriori* (MMAP) estimate for the parameters $\mathbf{a}$ can by calculated by evaluating:

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a}} p\left(\mathbf{a} \mid \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) \tag{14}$$

B. Principle

The principle of the histogram technique of §IV can be extended to the constrained channel model discussed above by writing (13) as:

$$\ln p\left(\mathbf{a} \mid \mathbf{x}, \boldsymbol{\phi}, \mathcal{I}\right) = \ln p\left(\mathbf{a} \mid \mathcal{I}\right) + \sum_{i=1}^{M} p_i\left(\mathbf{a} \mid \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, \mathcal{I}\right)$$

where the log-pdf of the parameters, $\mathbf{a}$, given only the data in the i-th block and the initial conditions, $\mathbf{x}_{i-1}$, is given by $p_i\left(\mathbf{a} \mid \mathbf{x}_i, \mathbf{x}_{i-1}, \boldsymbol{\phi}, \mathcal{I}\right)$: the log of the term in the RHS of (13). Using the transformation between the AR parameters and the AR poles, denoted by $\mathbf{r}_{\mathbf{a}} = \mathrm{roots}\left(\mathbf{a}\right)$, it follows:

$$\ln \hat{p}\left(\mathbf{r}_{\mathbf{a}} \mid \mathbf{x}, \cdot\right) = \ln \hat{p}\left(\mathbf{r}_{\mathbf{a}} \mid \cdot\right) + \sum_{i=1}^{M} \hat{p}_i\left(\mathbf{r}_{\mathbf{a}} \mid \mathbf{x}_i, \cdot\right) \tag{15}$$

where $\hat{p}\left(\mathbf{r}_{\mathbf{a}} \mid \mathbf{x}, \cdot\right)$ is the probability, $p\left(\mathbf{a} \mid \mathbf{x}, \cdot\right)$, of the set of channel parameters, $\mathbf{a}$, with corresponding poles $\mathbf{r}_{\mathbf{a}}$, when plotted in the pole plane. The '$\cdot$' is used for brevity to denote all known parameters. Consequently, $\hat{p}_i\left(\mathbf{r}_{\mathbf{a}} \mid \mathbf{x}_i, \cdot\right)$ can be considered as a probabilistic version of the histogram over a block as discussed in §IV. The important distinguishing feature is that, in contrast to the histogram approach where the peaks corresponding to the channel are not *constrained* to be in the same position between blocks, $\mathbf{r}_{\mathbf{a}}$ is constrained to be stationary over each block, and thus the peaks in the density $\hat{p}_i\left(\mathbf{r}_{\mathbf{a}} \mid \mathbf{x}_i, \cdot\right)$ corresponding to the channel are in very similar positions for each block. Hence, the sum of the densities $\hat{p}_i\left(\mathbf{r}_{\mathbf{a}} \mid \cdot\right)$, $i \in \mathcal{M}$ of

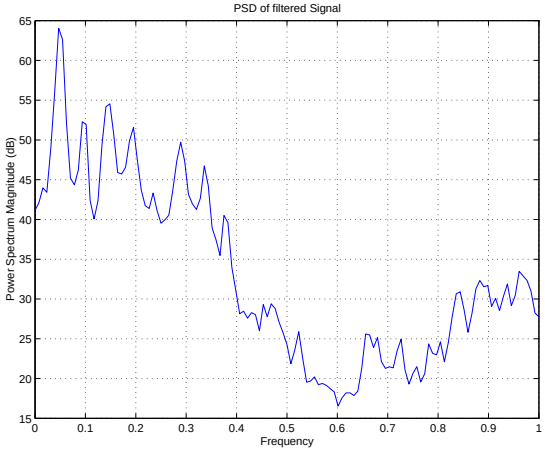


Fig. 6. Power spectral density of the BSAR(2) synthetic source discussed in §V-B.

equation (15) is large where a number of poles coincide – *i.e.* at the location the stationary poles – while the sum is small near the locations of the source poles which do not coincide in the parameter or pole space.

To demonstrate this, consider filtering a BSAR(2) synthetic source signal by a second-order all-pole filter, with $N = 20$ and $T_i = 1000$, $\forall i \in \mathcal{M}$. The phase and magnitude of the pole locations for this BSAR(2) process change linearly with block number. Thus, $\mathbf{b}_i = [\,-2r_i \cos\theta_i,\ r_i^2\,]^T$, which corresponds to a single complex pole-pair at $\mathbf{r}_{\mathbf{b}_i} = r_i\, e^{\pm j\theta_i}$, $i \in \mathcal{M}$. The rate of time-variation is slow enough that the frozen-state approximation of system poles is appropriate [29]. For a typical data sequence generated from this process, Figure 7 shows the contour plots of p_i in the complex domain for $i = 5$ and 15, and Figure 8 shows a plot of $\exp(p_i)$ for $i = 15$ to highlight the 'sharpness' of the peaks in the densities. Figure 9 shows a contour plot of $\ln \hat{p}_i$. Figure 6 shows the PSD of the source signal, which is equivalent to averaging the PSDs of the source signal in each block; clearly, the channel, which has a resonant at $\omega = 0.3\pi$, cannot be detected from this plot.

The stationary poles are estimated by maximising $\ln p\,(\mathbf{a}\,|\,\mathbf{x},\, \boldsymbol{\phi},\, \mathcal{I})$. In the figures, the actual location of the stationary pole is denoted by a cross, the channel pole by a diamond, and the MMAP estimate of the stationary pole is denoted by a square. Note that the estimate of the channel using the data in one block matches the position of the source pole, not the channel pole as desired. Further, the large circular dots denote the estimate of the channel when the entire system is modelled as stationary, rather than accounting for the nonstationarity. This estimate

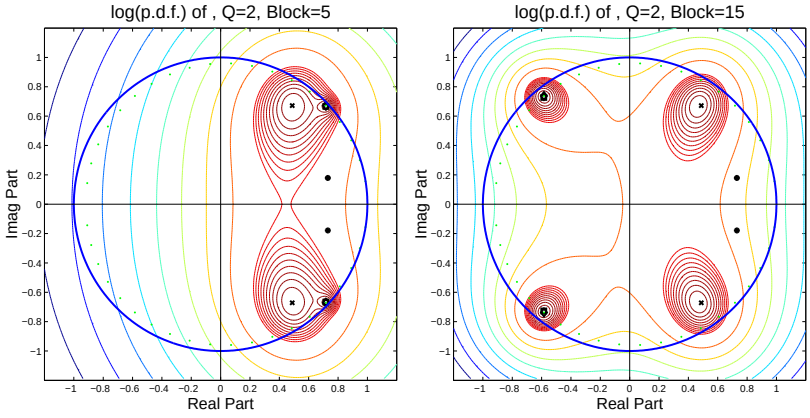


Fig. 7. Plot of $\ln \hat{p}_i\left(\mathbf{r_a} \mid \mathbf{x}_i, \cdot\right)$ for $i = 5$ and 15. The actual location of the stationary pole is denoted by a cross, the channel by a diamond, and the MMAP estimate of the stationary pole by a square. Additionally, the small dots denotes the position of the full trajectory of the time-varying channel pole. Contours are plotted at $\{10\%, 20\%, \cdots, 90\%, 92\%, \cdots, 98\%\}$. The unit circle is also plotted.

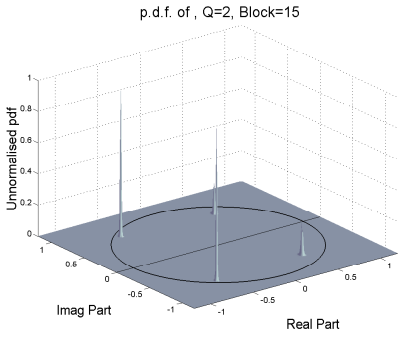


Fig. 8. Plot of $\hat{p}_{15}\left(\mathbf{r_a} \mid \mathbf{x}_{15}, \cdot\right)$.

is completely wrong, while the estimate using the pdf of the channels over all blocks is very accurate. Hence the benefits of explicitly utilising nonstationarity are clear.

C. Exploration of Parameter Space using Gibbs Sampler

In principle, a MMAP estimate for the unknown channel parameters, $\mathbf{a}$, can by found by solving (14). This optimisation can be performed using deterministic or stochastic optimisation methods. It is not the intention of this paper to investigate these various techniques, as their *pros* and *cons* are discussed elsewhere, *e.g.*, in [27]. However, since sampling from the distribution in (13) is difficult, estimates of the channel, $\mathbf{a}$, are obtained using the Gibbs sampler by drawing variates, $\boldsymbol{\theta}$, from the distribution $p\left(\boldsymbol{\theta} \mid \mathbf{x}\right)$. The Monte Carlo method can then be used to marginalise the

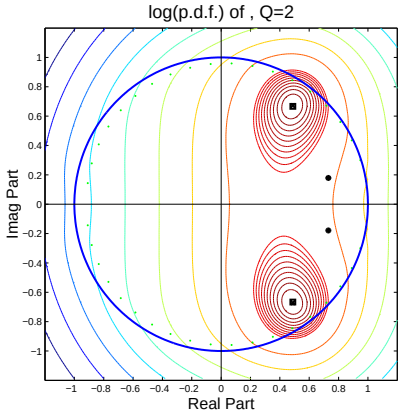


Fig. 9. Plot of $\ln \hat{p}\,(\mathbf{r_a} \mid \mathbf{x}, \cdot)$, showing how coninciding peaks in individual blocks sum to give a reliable estimate of the channel peak. The same legends are used as in Figure 7.

nuisance parameters $\mathbf{b}$ and $\boldsymbol{\sigma}$, and a minimum mean-square error (MMSE) estimate for $\mathbf{a}$ can be calculated by finding the expected value of the samples. Since, as described in the previous section, a dominant peak is expected in the distribution, it is reasonable to assume that the MMSE estimate is approximately equal to the MMAP estimate.

To ensure the distributions are not dependent on $\boldsymbol{\delta}$, the Bayesian model is extended so that $\boldsymbol{\delta}$ is also considered as a parameter. The excitation in block i may be written as:

$$\mathbf{e}_i = \mathbf{s}_i + \mathbf{S}_i\,\mathbf{b}_i \quad \text{or} \quad \mathbf{e}_i = \mathbf{y}_i + \mathbf{Y}_i\,\mathbf{a} \tag{16}$$

where $[\mathbf{y}]_{t-t_i+1} = y(t)$ and $[\mathbf{Y}]_{t-t_i+1,p} = y(t-p)$, $t \in \mathcal{T}_i, p \in \mathcal{P}$, and $y(t) \equiv y(t, \mathbf{b}_i, \mathbf{x})$ is a function of the data, $\mathbf{x}$, and the AR parameters in block i, $\mathbf{b}_i$:

$$y(t) = x(t) + \sum_{q=1}^{Q_i} b_i(q)\,x(t-q) \tag{17}$$

which is $x(t)$ filtered by $\mathbf{b}_i$. The likelihood function is:

$$p\,(\mathbf{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{M} \mathcal{N}\left(\mathbf{e}_i \,\middle|\, 0,\, \sigma_i^2\,\mathbf{I}_{T_i}\right) \tag{18}$$

Bayes's rule gives: $p\,(\mathbf{b}_i \mid \boldsymbol{\theta}_{-\mathbf{b}_i}, \mathbf{x}) \propto p\,(\mathbf{x} \mid \boldsymbol{\theta})\,p\,(\mathbf{b}_i)$. Using the priors discussed in §V-A, it follows that $p\,(\mathbf{b}_i \mid \boldsymbol{\theta}_{-\mathbf{b}_i}, \mathbf{x})$ is multivariate Gaussian with:

covariance: $$\mathrm{C}_i = \left\{ \frac{\mathbf{S}_i^T\,\mathbf{S}_i + \delta_i^{-2}\mathbf{I}_{Q_i}}{\sigma_i^2} \right\}^{-1}$$

and mode: $$\hat{\mathbf{b}}_i = -\left(\mathbf{S}_i^T\,\mathbf{S}_i + \delta_i^{-2}\mathbf{I}_{Q_i}\right)^{-1}\mathbf{S}_i^T\,\mathbf{s}_i$$

and $$p\left(\sigma_i^2 \,\middle|\, \boldsymbol{\theta}_{-\sigma_i}\,\mathbf{x}\right) = \mathcal{IG}\left(\sigma_i^2 \,\middle|\, \frac{T_i+\alpha_i}{2},\ \frac{\mathbf{e}_i^T\,\mathbf{e}_i+\beta_i}{2}\right)$$

Assuming $\mathbf{a}\,|\,\sigma_a^2 \sim \mathcal{N}\left(\mathbf{0}_P,\, \sigma_a^2\,\mathbf{I}_P\right)$, where $\sigma_a \in \mathbb{R}^+$ is a hyperparameter, it may be shown that $p\left(\mathbf{a}\,|\,\boldsymbol{\theta}_{-\mathbf{a}},\,\mathbf{x}\right)$ is a multivariate Gaussian with:

covariance: $$\mathrm{C}_a = \left\{\frac{1}{\sigma_a} + \sum_{i=1}^{M}\frac{\mathbf{Y}_i^T\,\mathbf{Y}_i}{\sigma_i^2}\right\}^{-1}$$

and mode: $$\hat{\mathbf{a}} = -\mathrm{C}_a\left(\sum_{i=1}^{M}\frac{\mathbf{Y}_i^T\,\mathbf{y}_i}{\sigma_i^2}\right)$$

The Gibbs sampler is subsequently used in the experiments to obtain a MMSE estimate of $\mathbf{a}$, as discussed above.

D. Effect of Model Order

Thus far, it has been assumed the correct form of model for the problem is known, and that the models for the source signal and channel accurately represent the system under consideration. The validity of this assumption is briefly discussed in §II-B. Given this assumption, however, it is necessary to find the most appropriate *model order* for the data. Both of these questions form the problem of *model selection*. It is impossible to do an exhaustive investigation on the effect of model order and so, as such, a number of experiments are selected to give sufficient evidence from which conclusions can be drawn.

The results are summarised by considering an experiment in which the channel filter, $\mathcal{A}$, is *known* to be second-order, and the source signal consists of extracts of recorded speech. This is simplest case when a surface plot of $p\left(\mathbf{a}\,|\,\mathbf{x},\,\cdot\right)$ can be made, and the effects of varying $\boldsymbol{\Xi} \triangleq \{Q_i,\, i \in \mathcal{M}\}$ investigated by visualising the changes in these plots. This is feasible only if $Q_i = Q,\, i \in \mathcal{M}$. Figure 10 shows the following, in columns, from top to bottom, for different beliefs of *source signal* model orders:

Contour plot of $\ln \hat{p}_i\left(\mathbf{r}_{\mathbf{a}}\,|\,\mathbf{x}_i,\,\cdot\right)$

In a particular data block, the source, which is actually a Q_A-th order AR process, where $Q_A = 12$, is modelled as a Q_M-th order AR process, with the parameters estimated using the covariance method [20]: the poles corresponding to this source model are denoted by a $\diamond$ in the

contour plot of $\ln \hat{p}_i(\cdot)$ in Figure 10. The position of the true source pole can be identified from the plot in the second column, when the proposed source model order is equal to the actual model order. Hence, this plot again indicates how the source poles influence the pdf of the channel parameters when estimated using just a single block.

Contour and surface plots of $\ln \hat{p}(\mathbf{r_a} \mid \mathbf{x}, \cdot)$

The surface plots of $\ln \hat{p}(\mathbf{r_a} \mid \mathbf{x}, \cdot)$ indicate the level of multimodality in the log distribution, which is not apparent in the plot of the actual distributions. The large dots in the contour plots denote the locations of the resulting channel estimates *if* the entire system is modelled as stationary. The MMAP estimate obtained using the proposed deconvolution method is shown as a ($\square$), and the actual location of the filter parameters are denoted by a ($\times$).

Surface plots of $\hat{p}(\mathbf{r_a} \mid \mathbf{x}, \cdot)$

These indicate how pronounced the dominant modes are.

Discussion

As the hypothesised model order increases, $\ln \hat{p}(\mathbf{r_a} \mid \mathbf{x}, \cdot)$ (and consequently $\ln p(\mathbf{a} \mid \mathbf{x}, \cdot)$) flattens out considerably. However, whilst considerable over-modelling makes it difficult for the peaks corresponding to the resonances of the filter to remain prominent in a particular data block, *over-modelling* by a factor of 2 to 3 relative to the true model order has little impact on the pdf of the channel parameters, given the entire data set. However, *under-modelling* the source signal *sometimes* leads to unsatisfactory results. When an AR process is *under-modelled*, the estimated spectrum often results in being relatively flat since the estimator is trying to fit the entire spectrum simultaneously, and not just a particular subband containing one of the resonant peaks. Therefore, when the source signal is *under-modelled*, the *estimated* source spectrum remains flat, and the pole locations due to the source may appear stationary. As the pdf flattens out, anomalous peaks emerge from these 'false' stationary pole locations. Additional experiments indicate that if the *channel* is *over-modelled*, but not *under-modelled*, the estimated spectrum is reasonably accurate and independent of hypothesised source model order. These results suggest that utilising nonstationarity reduces the requirement of belief regarding the models and accurate model order estimates may not be needed.

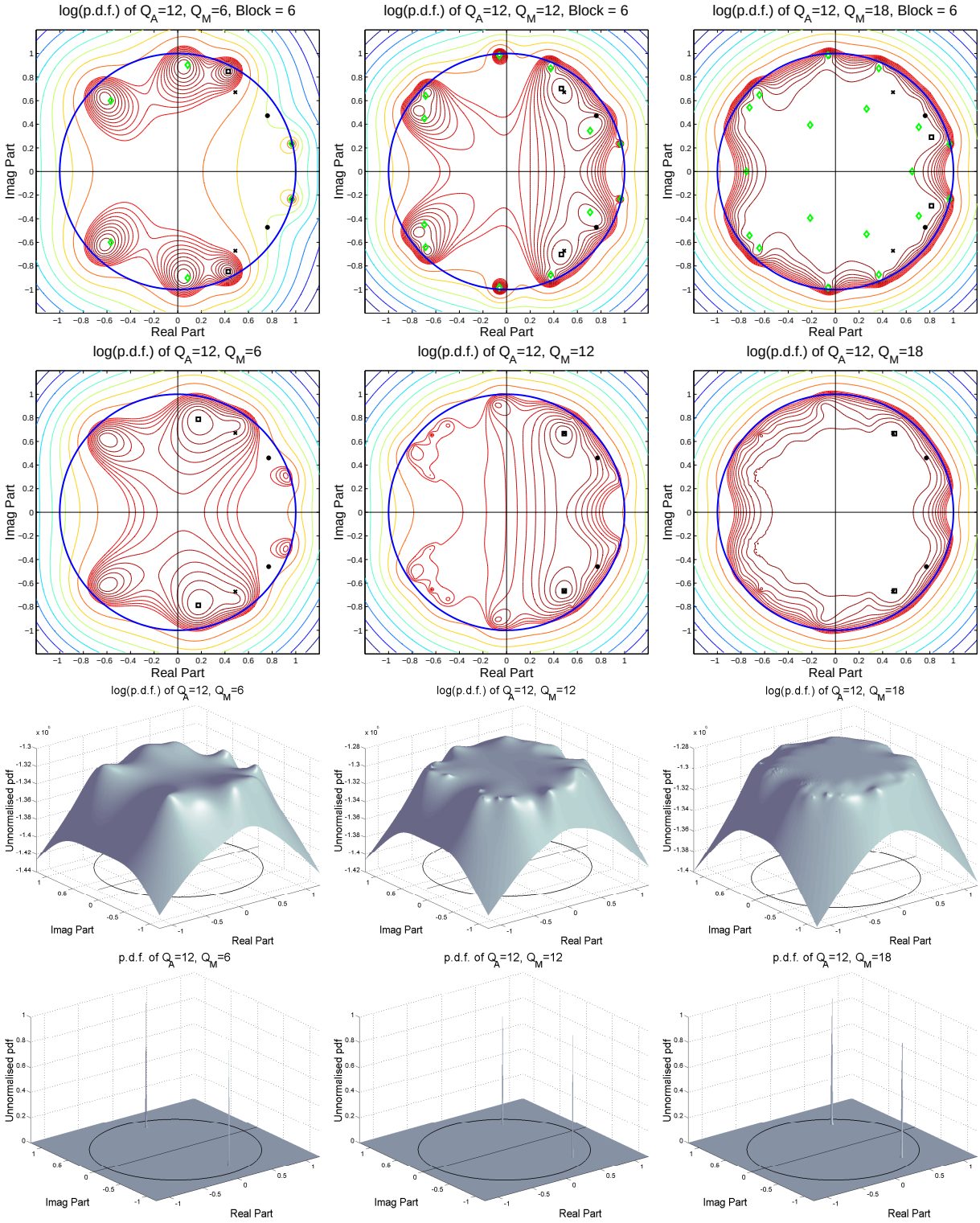


Fig. 10. These plots show the effect of model order on parameter estimation and the posterior density for an AR(2) system. From left to right, the TVAR model orders are $Q = \{6, 12, 18\}$. A synthetic 12-th order AR process is passed through the channel. The channel is estimated assuming there are $N = 50$ blocks, and $T_i = 1000$. The symbols on the contour plots are described in the text.

E. Effect of Observation Noise

The model shown in Figure 2 does not account for observation noise. In a more realistic model, the observed signal, $y(t) = x(t) + w(t)$, is given as the sum of the filtered speech, $x(t)$, and the observation noise, $w(t)$, which has variance σ_w^2, *i.e.* $w(t) \sim \mathcal{N}\left(0,\, \sigma_w^2\right)$. It is necessary to defined an "average signal-to-noise ratio (SNR)" as the average over all blocks of the SNRs for each block:

$$\hat{SNR} = \frac{1}{M}\sum_{i=1}^{M} 10\log\left\{\frac{\sum_{t=t_i}^{t_{i+1}-1} x^2(t)}{\sum_{t=t_i}^{t_{i+1}-1} w^2(t)}\right\} \tag{19}$$

If, for a nonstationary signal, the SNR is defined in the usual sense as the ratio of the total signal power over all the data to the total noise power, then in some data blocks, the "local SNR" might be very good, whilst in other blocks, the SNR may be very bad. There are three relevant measures used for the accuracy of the parameter estimates: the spectral distortion measure, the pole error function (PoEF) and the parameter error function (PaEF).

Spectral Distortion Measure

If $H_{\mathbf{a}}(e^{j\omega})$ is the frequency response of the AR process with parameters $\mathbf{a}$, the spectral distortion measure gives an indication of the similarity of the estimated response of the channel, $H_{\hat{\mathbf{a}}}(e^{j\omega})$, and actual response, $H_{\mathbf{a}}(e^{j\omega})$:

$$J_{H(e^{j\omega})} = \mathbb{E}\left\{\sum_{\omega}\left|20\log\left|\frac{H_{\hat{\mathbf{a}}}(e^{j\omega})}{H_{\mathbf{a}}(e^{j\omega})}\right|\right|\right\} \tag{20a}$$

Pole Error Function

The PoEF is a measure of the fit of pole estimates to the true pole locations and, using the notation in §IV-A:

$$J_{\mathbf{r_a}} = \mathbb{E}\Big\{\arg\min_{\mathcal{Q}_{\text{perm}}=\{q(p),\, p\in\mathcal{P}\}}\sum_{p\in\mathcal{P}}\left\|\mathbf{r_a}(p) - \hat{\mathbf{r}}_{\mathbf{a}}(q(p))\right\|^2\Big\}$$

where the set $\mathcal{Q}_{\text{perm}} = \text{perm}\,\{1,\ldots,P\}$ is a permutation of the elements in $\mathcal{P}$: *i.e.* $J_{\mathbf{r_a}}$ uniquely associates each $\hat{\mathbf{r}}_{\mathbf{a}}$ with an actual pole so as to minimise the total distance between the estimated and actual poles.

Parameter Error Function

The PaEF is the expected "distance" between the actual parameters and their estimates:

$$J_{\mathbf{a}} = \mathbb{E}\left\{\|\hat{\mathbf{a}} - \mathbf{a}\|\right\} \tag{21}$$

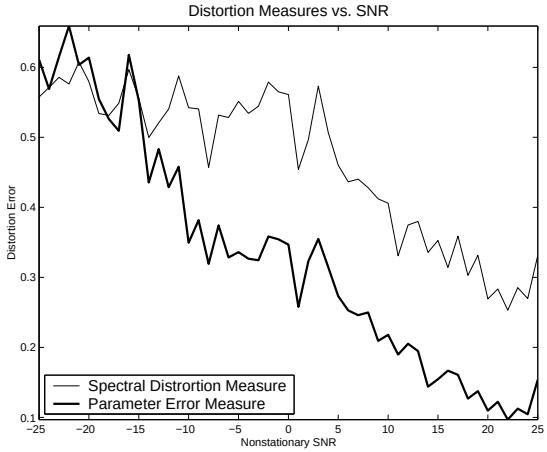


Fig. 11. Effect of observation noise on parameter estimates.

Parameter Accuracy vs. SNR

The influence of SNR on the accuracy of the parameter estimates can be investigated by driving a channel with real speech, adding noise with variance σ_w^2, and calculating the accuracy of the estimates using the three measures discussed above. The noise variance is a function of the average SNR, and is given by:

$$\sigma_w^2 = \frac{10^{\frac{1}{10}(S\hat{N}R - \hat{P})}}{T},\ \hat{P} = \frac{10}{M}\sum_{i=1}^{M} \log\left\{\sum_{t=t_i}^{t_{i+1}-1} x^2(t)\right\} \tag{22}$$

which has been derived from (19) by assuming the noise variance is constant across all blocks, and $T = T_i$, $i \in \mathcal{M}$. The channel estimates are calculated using the Gibbs sampler as discussed in §V-C. The simulation is run a number of times, with different noise realisations. This ensures that an *average* measure can be determined which is independent of a particular noise realisation.

In the simulation results given here, the channel used is shown in Figure 3(b), the speech is modelled as a 20-th order AR process, 40 realisations of the noise sequence are generated, and the Gibbs sampler is run for 2000 iterations, each run using a different initial condition. The channel is estimated for average SNRs between -25 dB and 25 dB. Figure 11 shows the average spectral distortion, and the PaEF, where the former has been scaled to fit on the same plot. Note that the PaEF falls off approximately linearly across a wide range of SNRs. Interestingly, the PoEF, not shown for clarity, falls off in the same way.

The performance of the proposed algorithm clearly falls with SNR and is susceptible to noise, although this is as expected. This is since the implicit filtering of (3) used to obtain an estimate

of the clean signal, $s(t)$, for use in the calculation of $p(\mathbf{a} \mid \mathbf{x})$ in (13), has a noise gain greater than one. Although the observation noise can be incorporated into the marginal posterior $p(\mathbf{a} \mid \mathbf{y})$, it would appear that to do so leads to an intractable distribution where numerical methods are required. The Kalman filter is particularly appropriate in such a case, and will be discussed elsewhere.

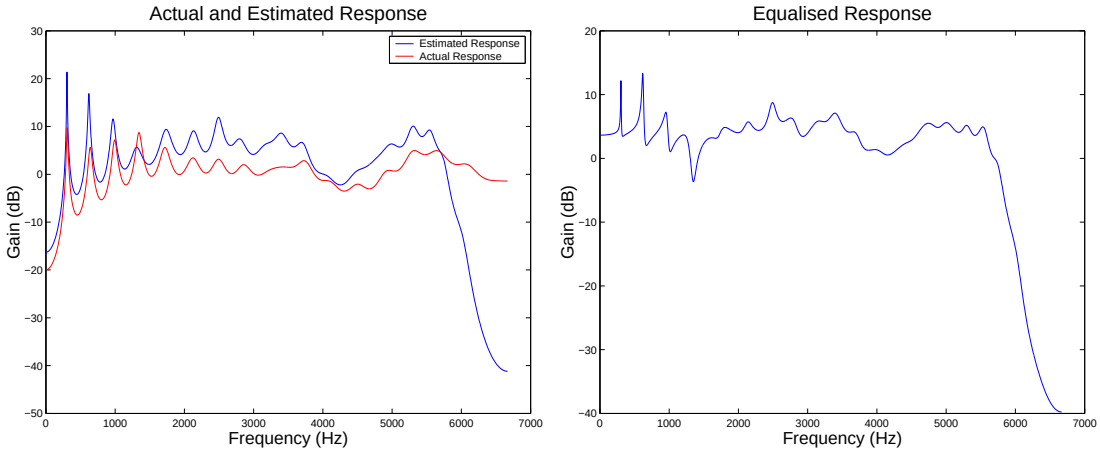
However, a far simpler, and just as effective, approach to reduce the effect of observation noise is to use the Wiener-Hopf filter (WHF) to obtain a maximum-likelihood estimate (MLE) of the clean signal, $s(t)$. The likelihood in equation (12) is then calculated using the MLE, $\hat{s}(t)$. This may straightforwardly be interpreted as replacing the 'direct inverse' relationship between $x(t)$ and $s(t)$ in (3) by:

$$\hat{s}(t) = \sum h(t,q)\, y(t-q) \tag{23}$$

where the Wiener-Hopf filter, $h(t,q)$, is chosen to minimise the mean squared error (MSE) between the estimate of the desired signal of (23), $\hat{s}(t)$, and $s(t)$. The WHF can be expressed in terms of the unknown channel parameters and the correlation function of the observed signal, $y(t)$. Hence, *given* a proposed set of channel parameters, the WHF can be calculated, and the pdf of the channel parameters calculated as before. The marginal posterior for these channel parameters can then be calculated by evaluating (13), or by modifying the Gibbs sampler as appropriate.

F. Effect of Length and Number of Blocks

There is an inherent problem in the modelling of nonstationary processes by a block stationary process: if the block length is large, the variance of the parameter estimate is small; however, in that block, the actual parameter values may change significantly, such that the model no longer accurately reflects the time-varying nature of the underlying signal. On the other hand, if the block length is small, the variance of the estimate is large, although the block stationary model will better represent the time-varying nature of the signal. This raises the question of whether an optimum block length exists. Although the changepoints in τ (subsumed in $\boldsymbol{\phi}$) could be blindly estimated, it is important to have an understanding of what the optimum is, and why. However, experimentation with block length, particularly in the examples discussed in §VI suggests that only an approximate estimate of the 'optimal' block length is required when taking nonstationarity into account.



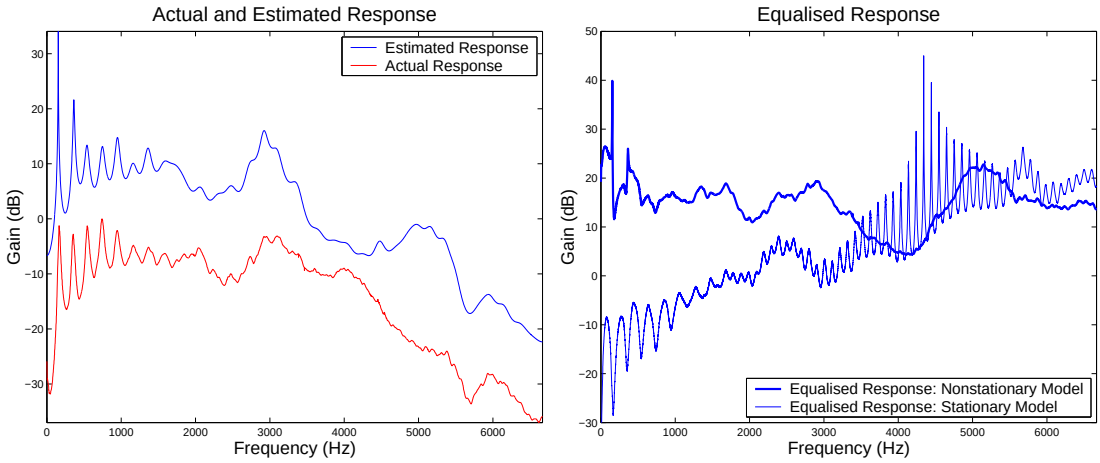
(a)Actual and Estimated Response (b)Equalised Magnitude Response

Fig. 12. Actual and estimated frequency responses of the acoustic gramophone horn in the first example.

VI. Blind deconvolution in a simple acoustic environment: Gramophone horn

As an example of the approach used in this paper, consider again speech signal which is recorded through an acoustic gramophone horn, as discussed in §IV-C. The frequency response of the horn is shown in Figure 3(a). In the first of two examples, a synthetic BSAR process is used to model the source signal, and generated as in §IV-C. This synthetic signal, with model order $Q = 80$, is filtered by the horn. Choosing $N = 80$ blocks of length $T_i = 1000$, the channel upto a bandwidth of 6667 Hz is estimated using the proposed Bayesian algorithm. The Gibbs sampler is used to generate samples from which a MMSE estimate of the channel is made. Although the hyper-hyperparameters were fixed at $\nu_i = \gamma_i = 0$, the hyperparameters δ_i are estimated from the data, using the prior distribution discussed in the Appendix. Figure 12 shows the actual and estimated responses and it can be seen that although some resonances remain in the equalised response, the magnitude of these resonances are far smaller than in the unequalised response. Note that the accuracy of the channel estimate falls off at high-frequencies where there is no signal energy. In this example, the model orders for the AR processes are assumed to be known, and the block lengths are chosen heuristically. Reversible-jump MCMC techniques [30] could be used to tackle the case when the AR models are unknown [31]. However, the investigations presented in §V-D question whether accurate estimation of these unknowns is *really* necessary, since the utilisation of nonstationarity reduces the dependence on such parameters.

In the second example, real speech is filtered by the full measured response of the acoustic



(a)Actual and Estimated Response (b)Equalised Magnitude Response

Fig. 13. Actual and estimated frequency responses of the acoustic gramophone horn in the second example.

horn. This test is as close as possible to actually measuring speech that has been "played" through the horn. The results are shown in Figure 13, where an offset between the actual and estimated responses has been included to make the graph clearer. The channel is modelled as 72-nd order, and the speech as 80-th order, with $N = 80$, $T_i = 1000$. Although the equalised response is not particularly flat, acoustic listening tests indicate that the restored version is more pleasing to the ear than the speech heard directly from the horn. This example highlights the problem that although the frequency response of the channel estimate is close to the actual response, any slight errors can introduce additional resonances in the equalised response; this is a well-known problem with equalisation of resonant responses. However, the algorithm finds a good estimate of the channel which can be used for restoration. Additionally, the equalised response obtained when the system is modelled as stationary across all the data is extremely poor. In both examples, the Gibbs sampler was started from a variety of positions and, in all cases, converged towards the same solution.

A. All-pole model for Room Acoustics

The transfer function due to the acoustics of a room generally do not change considerably with time, but do vary with the spatial locations of the sound source and observer. If the observer is assumed to be spatially stationary, a LTI model for the room transfer function (RTF) is appropriate. Typical all-pole model orders required for approximating RTFs are in the range

$50 \leq P \leq 500$ [32]. Mourjopoulos and Paraskevas [32] state that all-pole model orders are typically a factor of 40 lower than all-zero model orders, while several studies [33, 34] state that the gain achieved using pole-zero over all-zero modelling of reverberant environments is not as high as generally thought throughout the literature. A significant advantage of the all-pole model over other LTI models is its lower sensitivity to changes in source and observer positions.

For a rectangular room the number of poles, P_{theo}, upto a frequency, $f_s \gg 500$ Hz, increases $\propto f_s^3$ [35]. As $f_s \to \infty$, P_{theo} applies to arbitrary shaped rooms [1]. Since this estimate does not take into account correlations between modes, it is a high upper bound and, for typical rooms, is much higher than required for all-pole modelling. If the all-pole order, $P = P_{\text{theo}}$, the model corresponds well with the actual response; if $P \ll P_{\text{theo}}$, as is typically the case, the estimated poles correspond to the major resonances which have high Q factors [35]. Mourjopoulos [32] concludes that in many applications dealing with room acoustics, it may be both sufficient and more efficient to manipulate all-pole rather than high-order all-zero models, and is thus a reasonable model for a range of acoustic environments.

VII. Conclusions

Single channel blind deconvolution is tackled by modelling the source signal as a BSAR process, and the distortion operator as an all-pole filter. The Bayesian paradigm is used as a means of parameter estimation, and the posterior density for the distortion filter parameters conditional on the observed data is derived. The issues of selecting model order and block length have been investigated. By utilising the nonstationarity of the system, less specific belief regarding the model of the source signal is required. As long as the model of the source is nonstationary, the stationary component of the system can be estimated. While there exists a plethora of nonstationary, linear models, each of which is appropriate for different nonstationary systems, the purpose of this paper is not to investigate their properties but, rather, it shows how *nonstationarity* can provide additional degrees of freedom that allow strong requirements on prior belief to be relaxed, and the model investigated here has shown that. Several examples of blind deconvolution of reasonably high-order channels have been investigated, and the results are extremely encouraging, and far superior to the estimates obtained from a histogram approach.

Appendix

This appendix details some of the steps in deriving the pdf in (13). The prior densities for the BSAR coefficients and excitation variance are introduced in §III and §V, and rely on the hyperparameters $\{\boldsymbol{\delta},\, \boldsymbol{\nu},\, \boldsymbol{\gamma}\}$. The values for $\{\boldsymbol{\delta},\, \boldsymbol{\nu},\, \boldsymbol{\gamma}\}$ are unknown and often have a prior assigned to them; these 'hyper-priors' also depend on hyper-hyperparameters. The form of the posterior is less susceptible to changes in the hyper-hyperparameters, than to changes in the hyperparameters. A complete Bayesian hierarchical model for a BSAR process may be found in [31]. Here, a slightly less general form of Bayesian hierarchical model is chosen to ensure that the underlying principle of utilising nonstationarity isn't obscured. As such, $\{\boldsymbol{\nu},\, \boldsymbol{\gamma}\}$ are assumed to be known, and a hyperprior is placed on $\boldsymbol{\delta}$, such that the influence on the posterior of this hyper-hyperparameter is minimal. A vague conjugate prior density is ascribed to δ^2 using an inverse-Gamma density: $\delta^2 \sim \mathcal{IG}\left(\alpha_{\delta^2},\, \beta_{\delta^2}\right)$. Assigning these priors to each block, and modifying Bayes's rule in §V-A, the joint density is:

$$
\begin{aligned}
p\left(\boldsymbol{\theta},\, \boldsymbol{\delta} \mid \mathbf{x}, \boldsymbol{\phi}_{-\boldsymbol{\delta}}, \mathcal{I}\right) &= \frac{p\left(\mathbf{a} \mid \mathcal{I}\right)}{p\left(\mathbf{x} \mid \mathcal{I}\right) \mathcal{J}(\mathbf{x},\, \mathbf{s})} \\
&\times \prod_{i=1}^{M} \left[\frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{ -\frac{(\mathbf{s}_i + \mathrm{S}_i\mathbf{b}_i)^T(\mathbf{s}_i + \mathrm{S}_i\mathbf{b}_i)}{2\sigma_i^2} \right\} \right. \\
&\times \frac{1}{(\sqrt{2\pi}\delta_i\sigma_i)^{Q_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}} (\sigma_i^2)^{-(\frac{\nu_i}{2}+1)}}{\Gamma(\frac{\nu_i}{2})} \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}} \left(\delta_i^2\right)^{-(\alpha_{\delta_i^2}+1)}}{\Gamma(\alpha_{\delta_i^2})} \\
&\left. \times \exp\left\{ -\frac{\gamma_i}{2\sigma_i^2} - \frac{\mathbf{b}_i^T\,\mathbf{b}_i}{2\delta_i^2\sigma_i^2} - \frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \right]
\end{aligned}
$$

Since the transformation from $\mathbf{s}$ to $\mathbf{x}$ is linear, $\mathcal{J}(\mathbf{x},\, \mathbf{s}) = 1$. The AR parameters, $\mathbf{b}$, are marginalised using the form:

$$
p\left(\boldsymbol{\theta}_{-\mathbf{b}} \mid \mathbf{x}\right) = \int_{\mathbb{R}^{Q_1}} \cdots \int_{\mathbb{R}^{Q_M}} p\left(\boldsymbol{\theta}_{-\mathbf{b}},\, \mathbf{b} \mid \mathbf{x}\right)\, d\mathbf{b}_M \,\ldots\, d\mathbf{b}_1
$$

using the standard Gaussian integral [26] to give:

$$
\begin{aligned}
& p\left(\boldsymbol{\theta}_{-\mathbf{b}},\, \boldsymbol{\delta} \mid \mathbf{x},\, \boldsymbol{\phi}_{-\boldsymbol{\delta}},\, \mathcal{I}\right) \\
& = \frac{p\left(\mathbf{a} \mid \mathcal{I}\right)}{p\left(\mathbf{x} \mid \mathcal{I}\right)} \prod_{i=1}^{M} \left[\frac{1}{(\sqrt{2\pi})^{T_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} \frac{1}{\left|\mathrm{S}_i^T\, \mathrm{S}_i + \delta_i^{-2}\, \mathrm{I}_{Q_i}\right|^{\frac{1}{2}}} \right. \\
& \times \frac{1}{\sigma_i^{T_i+\nu_i+2}} \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}}}{\Gamma(\alpha_{\delta_i^2})} (\delta_i^2)^{-(\frac{Q_i}{2}+\alpha_{\delta_i^2}+1)} \exp\left\{ -\frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \\
& \left. \times \exp\left\{ -\frac{\gamma_i + \mathbf{s}_i^T\, \mathbf{s}_i - \mathbf{s}_i^T\, \mathrm{S}_i \left(\mathrm{S}_i^T\, \mathrm{S}_i + \delta_i^{-2}\, \mathbf{I}_{Q_i}\right)^{-1} \mathrm{S}_i^T\, \mathbf{s}_i}{2\sigma_i^2} \right\} \right]
\end{aligned}
$$

Finally, marginalising σ_i^2:

$$
p\left(\mathbf{a} \mid \mathbf{x}\right) = \int_0^\infty \cdots \int_0^\infty p\left(\mathbf{a},\, \boldsymbol{\sigma} \mid \mathbf{x}\right)\, d\sigma_M^2 \ldots\, d\sigma_1^2 \tag{25}
$$

using the standard Gamma integral [26] gives:

$$
\begin{aligned}
& p\left(\mathbf{a},\, \boldsymbol{\delta} \mid \mathbf{x},\, \boldsymbol{\phi}_{-\boldsymbol{\delta}},\, \mathcal{I}\right) \\
& = \frac{p\left(\mathbf{a} \mid \mathcal{I}\right)}{p\left(\mathbf{x} \mid \mathcal{I}\right)} \prod_{i=1}^{M} \left[\frac{1}{(\sqrt{2\pi})^{T_i}} \frac{\left(\frac{\gamma_i}{2}\right)^{\frac{\nu_i}{2}}}{\Gamma(\frac{\nu_i}{2})} \frac{1}{\left|\mathrm{S}_i^T\, \mathrm{S}_i + \delta_i^{-2}\, \mathrm{I}_{Q_i}\right|^{\frac{1}{2}}} \right. \\
& \times \frac{\beta_{\delta_i^2}^{\alpha_{\delta_i^2}}}{\Gamma(\alpha_{\delta_i^2})} (\delta_i^2)^{-(\frac{Q_i}{2}+\alpha_{\delta_i^2}+1)} \exp\left\{ -\frac{\beta_{\delta_i^2}}{\delta_i^2} \right\} \\
& \left. \times \frac{\Gamma(R_i)}{\left\{ \gamma_i + \mathbf{s}_i^T\, \mathbf{s}_i - \mathbf{s}_i^T\, \mathrm{S}_i \left(\mathrm{S}_i^T\, \mathrm{S}_i + \delta_i^{-2}\, \mathbf{I}_{Q_i}\right)^{-1} \mathrm{S}_i^T\, \mathbf{s}_i \right\}^{R_i}} \right]
\end{aligned}
$$

where $R_i = \frac{T_i+\nu_i+1}{2}$. If the prior on the hyper-parameter $\boldsymbol{\delta}$ is omitted, this expression reduces to (13).

References

[1] H. Kuttruff, *Room Acoustics*, Spon Press, London, England, fourth edition, 2000.

[2] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, no. 1, pp. 117–128, Jan. - Feb. 2000.

[3] S. V. Vaseghi and P. J. W. Rayner, "The effects of non-stationary signal characteristics on the performance of adaptive audio restoration systems," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, UK, May 1989, vol. 1, pp. 377–380.

[4] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 2055–2063, Aug. 1996.

[5] A. Ahmed, P. J. W. Rayner, and S. J. Godsill, "Considering non-stationarity for blind signal separation," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New York, Oct. 1999, pp. 111–114.

[6] J. R. Hopgood and P. J. W. Rayner, "Bayesian single channel blind deconvolution using parametric signal and channel models," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New York, Oct. 1999, pp. 151 – 154.

[7] J. R. Hopgood, *Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments*, Ph. D. Thesis, University of Cambridge, UK, Nov. 2000.

[8] T. H. Li, "Estimation and blind deconvolution of autoregressive systems with nonstationary binary inputs," *Journal of Time Series Analysis*, vol. 14, no. 6, pp. 575–588, 1993.

[9] R. Chen and T. H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampling," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2410–2413, Oct. 1995.

[10] T. H. Li and K. Mbarek, "A blind equalizer for nonstationary discrete-valued signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 247–254, Jan. 1997.

[11] O. Cappé, A. Doucet, M. Lavielle, and E. Moulines, "Simulation-based methods for blind maximum-likelihood filter identification," *Signal Processing*, vol. 73, no. 1-2, pp. 3–25, Feb. 1999.

[12] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, May 1996.

[13] L. A. Liporace, "Linear estimation of nonstationary signals," *Journal of the Acoustical Society of America*, vol. 58, no. 6, pp. 1288–1295, Dec. 1975.

[14] Y. Grenier, "Time-dependent ARMA modelling of nonstationary signals," *IEEE Transactions on Speech, and Signal Processing*, vol. 31, no. 4, pp. 899–911, Aug. 1983.

[15] R. Charbonnier, M. Barlaud, G. Alengrin, and J. Menez, "Results on AR-modelling of nonstationary signals," *Signal Processing*, vol. 12, no. 2, pp. 143–151, Mar. 1987.

[16] T. Subba Rao, "The fitting of nonstationary time-series models with time-dependent parameters," *Journal of the Royal Statistical Society – Series B*, vol. 32, no. 2, pp. 312–322, 1979.

[17] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Processing*, vol. 5, no. 3, pp. 267–285, May 1983.

[18] J. J. Rajan and P. J. W. Rayner, "Generalised feature extraction for time-varying autoregressive models," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2498–2507, Oct. 1996.

[19] J. J. Rajan, P. J. W. Rayner, and S. J. Godsill, "Bayesian approach to parametric estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 144, no. 4, pp. 249–256, Aug. 1997.

[20] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[21] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag, Germany, 1998.

[22] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Holden-Day, third edition, 1994.

[23] P. S. Spencer and P. J. W. Rayner, "Separation of stationary and time-varying systems and its application to the restoration of gramophone recordings," in *Proc. IEEE International Symposium on Circuits and Systems*, Portland, OR, USA, May 1989, vol. 1, pp. 292–295.

[24] P. S. Spencer, *System Identification with Application to the Restoration of Archived Gramophone Recordings*, Ph. D. Thesis, University of Cambridge, UK, June 1990.

[25] B. Theobald, S. Cox, G. Cawley, and B. Milner, "Fast method of channel equalisation for speech signals and its implementation on a DSP," *Electronic Letters*, vol. 35, no. 16, pp. 1309–1311, Aug. 1999.

[26] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Inc., fifth edition, 1994, Alan Jeffrey, editor.

[27] J. J. K. O'Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer-Verlag, New York, 1996.

[28] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, Inc., 1992.

[29] L. A. Zadeh, "Time-varying networks, I," *Proc. IRE*, vol. 49, pp. 1488–1503, Oct. 1961.

[30] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

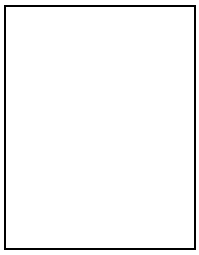
[31] O. Punska, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Bayesian segmentation of piecewise constant autoregressive processes using MCMC methods," Tech. Rep. CUED/F-INFENG/TR. 344, Department of Engineering, University of Cambridge, UK, 1999.

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

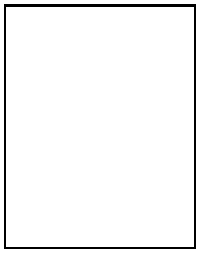
[33] S. Gudvangen and S. J. Flockton, "Comparison of pole-zero and all-zero modelling of acoustic transfer functions," *Electronic Letters*, vol. 28, no. 21, pp. 1976–1978, Oct. 1992.

[34] S. Gudvangen and S. J. Flockton, "Modelling of acoustic transfer functions for echo cancellers," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 142, no. 1, pp. 47–51, Feb. 1995.

[35] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modelling of room transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 320–328, Apr. 1994.



James R. Hopgood received the M. A., M.Eng. degree in Electrical and Information Sciences from the Univeristy of Cambridge in 1997, and a Ph.D. from Cambridge in July 2001. He is currently a Research Fellow at Queens' College, Cambridge, and is with the Signal Processing and Research Group at the Department of Engineering at Cambridge University working on nonstationary signal processing with application to reverberation cancellation in acoustic environments.



Peter J. W. Rayner received the M.A. degree from the University of Cambridge, UK, in 1968 and the Ph.D. degree from Aston University in 1969. Since 1968 he has been with the Department of Engineering at Cambridge University and is head of the Signal Processing Research Laboratory. In 1999 he was appointed an ad-hominem professorship in Signal Processing. He teaches courses in random signal theory, digital signal processing, image processing and communication systems. His current research interests include: optimal nonlinear filters for signal detection and estimation, nonlinear system identification and equalisation, blind identification of acoustic systems, signal separation, document and image sequence restoration, speech enhancement in low signal-to-noise ratio conditions, restoration of signals in nonstationary non-Gaussian noise, and Bayesian techniques in signal reconstruction and classification.