MODELS FOR BLIND SPEECH DEREVERBERATION: A SUBBAND ALL-POLE FILTERED BLOCK STATIONARY AUTOREGRESSIVE PROCESS

James R. Hopgood

Institute for Digital Communications,
School of Engineering and Electronics, The University of Edinburgh
http://www.see.ed.ac.uk/~jhopgoo1 James.Hopgood@ed.ac.uk

ABSTRACT

Single channel blind dereverberation of speech acquired in an acoustic environment is approached using parametric modelling and estimation theory to obtain channel estimates; inverse filtering is then applied to derevererate the speech. Models previously used in such an approach deal with relatively simple scenarios such as a gramophone horn modelled with 70 parameters; the weakness of those models, however, is in their attempt to simultaneously model the full channel spectrum by a single all-pole filter. Not only does this lead to a large computational load, it is not parsimonious, nor is it scalable such that the algorithm can be applied to higher dimensional problems. A better approach uses subbands; in this paper, a subband all-pole filter models the channel while the source is still represented by a single-band block stationary AR process. An example is given of blindly dereverberating a signal observed through the aforementioned gramophone horn, demonstrating an equally robust, but more flexible and scalable model.

1. INTRODUCTION

Audio signals acquired in a confined acoustic environment usually exhibit reverberation, due to the physical separation between the source and microphone, which distorts the source signal; this causes problems in two major classes of signal processing applications. The first is in automatic speech recognition, and its variants, where it is more difficult to identify reverberant natural speech than with anechoic or closely coupled speech. This prevents "hands-free" interaction without the undesirable constraint that the user must carry a microphone near their mouth. The second class involves the desire to improve speech quality and intelligibility from devices such as mobile telephones, 'hands-free' tele-conferencing systems and next generation digital hearing aids. In each case, the presence of reverberation should be reduced to adequate levels. In some engineered scenarios, microphone arrays can be used to transform the problem into a multi-channel blind source separation problem. However, it is not always practical to use them due to physical constraints on the size of the array; for example, mobile phones, hearing aids, or a cluttered work environment.

Single channel blind dereverberation is the process of removing reverberation from a single realisation of a reverberant signal modelled as the convolution of an unknown source signal with an unknown acoustic environment. This problem is particularly challenging since audio signals are temporally-correlated and have signal-values that belong to an infinite support (the real line). The scenario has previously been considered in [1], where it is assumed a nonstationary source can be modelled by a block stationary AR (BSAR) process, and the room transfer function (RTF) by an all-pole filter. Robust channel estimates are found using Bayesian parameter estimation theory, and an estimate of the original signal is obtained by inverse filtering the observed reverberant signal. The results [1] show that accurate channel estimates can be obtained for relatively high model orders; examples include source model orders greater than 60, and channel model orders greater than 70.

Although the *all-pole filter* can often parsimoniously approximate rational transfer functions, it is usually applied such that it simultaneously fits the entire frequency range, even though it may well fit some regions in this frequency space better than others. When acoustic impulse responses (AIRs) with reverberation times of $T_{60} \approx 0.5$ seconds are modelled by a linear time-invariant (LTI) all-pole filter, large model orders are typically required, say at least 500 [2]. Thus, attempting to model the entire acoustic spectrum by a single all-pole model often leads to a large computational load; this can be unacceptable in computationally intensive algorithms such as blind dereverberation. As a result, the algorithm in [1] suffers from its lack of scalability; for example, to dereverberate a speech signal measured in a real room, extremely high model-orders are required to approximate the AIR, and the blind deconvolution algorithm becomes numerically impractical.

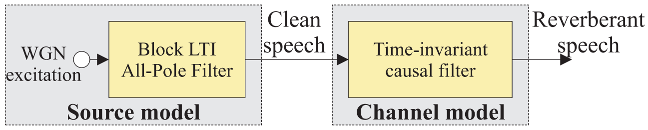

Figure 1: Parametric System Model.

Consequently, it is preferable to simply model a particular frequency band of the filter's spectrum by an all-pole filter, leading to lower model orders. Subband linear prediction was first considered in [3] and developed in [4, 5]. Complete decoupling of the subbands, as discussed in [6] where the subband all-pole filter is applied to room acoustics, leads to discontinuities in the model's spectrum at the subband boundaries, introducing distortion into an equalised channel. To circumvent this problem, continuity across subband boundaries much be ensured by constraining the all-pole parameters such that the end points at one subband boundary is matched to the estimated spectrum in the adjacent subbands, as shown in [7].

In this paper, the subband modelling techniques introduced in [6, 7] are incorporated into the blind dereverberation algorithm introduced in [1], with the aim of extending that algorithm to deal with more practical acoustic impulse responses.

Notation

Except where indicated, the set notation $\mathcal{G} = \{1, \dots, G\} \subset \mathbb{Z}$ is used; *e.g.* $\mathcal{N}_i = \{1, \dots, N_i\}$.

2. SYSTEM MODELS

In single channel blind deconvolution, a degraded observation, $\mathbf{x} = \{x(t), t \in \mathcal{T}\} \in \mathbb{R}^T$, is modelled as a nonlinear filtration of an unknown source signal, $\mathbf{s} = \{s(t), t \in \mathcal{T}\} \in \mathbb{R}^T$; thus, $\mathbf{x} = \mathbf{f}(\mathbf{s})$, where $\mathbf{f}(\cdot)$ denotes the unknown distortion operator. The task is to estimate from the observations, $\mathbf{x}$, either the distortion, $\mathbf{f}$, or a scaled shifted version of the source, $\hat{s}(t) = a\,s(t-\tau)$, where $a, \tau \in \mathbb{R}$. The general form of the parametric system model is shown in Figure 1.

2.1 Nonstationary Source Model

Speech is nonstationary and a particularly appropriate model for a clean speech signal, $s(t)$, is a BSAR process [8–10]; $s(t)$ is partitioned into M contiguous disjoint blocks, with block $i \in \mathcal{M}$ beginning at sample t_i and of length $T_i = t_{i+1} - t_i$. In this block, $s(t)$ is modelled as a stationary autoregressive (AR) model of order Q_i:

$$s(t) = -\sum_{q=1}^{Q_i} b_i(q)\, s(t-q) + e(t), \quad e(t) \sim \mathcal{N}\left(0,\, \sigma_i^2\right) \tag{1}$$

where $e(t)$ is the excitation process whose variance is a positive real number, $\sigma_i \in \mathbb{R}^+$, and where, in this equation, the time index takes on values $t \in \mathscr{T}_i = \{t_i, \ldots, t_{i+1}-1\}$, with $i \in \mathscr{M}$.

2.2 Time-Invariant Channel Model

In previous work [1], the LTI all-pole model is used for the channel:

$$x(t) = -\sum_{p=1}^{P} a(p)x(t-p) + s(t) \tag{2}$$

where $\mathbf{a} = \{a(p),\, p \in \mathscr{P}\}$ is the set of P model parameters. The filter's transfer function is:

$$H(k) = \frac{1}{1 + \sum_{p=1}^{P} a(p)\, e^{-\frac{2j\pi kp}{T}}} \tag{3}$$

However, a more flexible approach models the channel using L subbands, rather than a single "full-band". In each subband, the transfer function, $H(k)$, of $s(t)$ is modelled by an all-pole spectrum in the region $k \in \{k_l, \ldots, k_{l+1}-1\}$, where l denotes the subband index, $K_l = 2(k_{l+1} - k_l) = T_l$ is the number of frequency components in band $l \in \mathscr{L}$, and $\{k_l,\, l \in \mathscr{L}\}$ are the subband boundaries, where $k_0 \triangleq 0$ and $k_L \triangleq T$. Thus, the frequency bin closest to the half sampling frequency is given by $k_{\frac{f_s}{2}} = \lfloor \frac{T}{2} \rfloor$. The transfer function in a particular subband is obtained from equation (3) using the mapping for $kT \leq 2$,

$$k \to \frac{k - k_l}{K_l} = \frac{k - k_l}{2(k_{l+1} - k_l)} \tag{4}$$

such that the transfer function of the subband model is:

$$H(k) = \sum_{l=0}^{L-1} \frac{G_l \frac{K_l}{K}\, \mathbb{I}_{\{k_l, k_{l+1}\}}(k)}{1 + \sum_{p=1}^{P_l} a_l(p)\, e^{-\frac{\pi j(k-k_l)p}{k_{l+1}-k_l}}} \tag{5}$$

where $\mathbf{a}_l = \{a_l(p),\, p \in \mathscr{P}_l\}$ and $G_l \in \mathbb{R}^+$ denote model parameters in subband l, and the indicator function $\mathbb{I}_{\mathscr{A}}(a) = 1$ if $a \in \mathscr{A}$ and zero otherwise. The term $\frac{K_l}{K}$ is required in equation (5) to ensure that the *energy* in the spectrum in conserved through the spectral mapping defined in equation (4). The gain term, G_l, allows a further degree of freedom in the channel model: note that to avoid scaling ambiguities, $G_0 \triangleq 1$. Makhoul [3] suggests a similar model when analysing speech using linear prediction.

A significant problem with this model as presented is its inability to accurately model the phase response of the system transfer function [6]: the phase response of the subband AR model at the subband boundaries is zero, whereas the phase response of the actual transfer function will not be zero at these points. Techniques for dealing with this phase modelling problem are discussed elsewhere [6], but are presently not considered in this paper.

2.3 Prior density functions

Bayesian parameter estimation, as introduced in §3, requires the assignment of a prior probability density function (pdf) on any unknown parameters in the system. For a real, stable, minimum-phase all-pole or AR process, $\mathbf{a}$, should ideally only take on values which lie in the *stability domain*. However, it is usual to place a Gaussian prior on the parameters: $\mathbf{a} \,|\, \sigma^2 \sim \mathcal{N}\left(\mathbf{0}_P,\, \sigma^2 \delta^2 \mathrm{I}_P\right),\, \delta \in \mathbb{R}^+$. A standard prior for scale parameters, such as filter gains and variances, is the inverse-Gamma density: $\sigma^2 \,|\, \alpha, \beta \sim \mathscr{IG}\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)$.[1] As

[1] The inverse-Gamma pdf is $p(x) = \mathscr{IG}(x \,|\, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)}\, e^{-\frac{\beta}{x}}$ for $x \geq 0$, and zero otherwise.

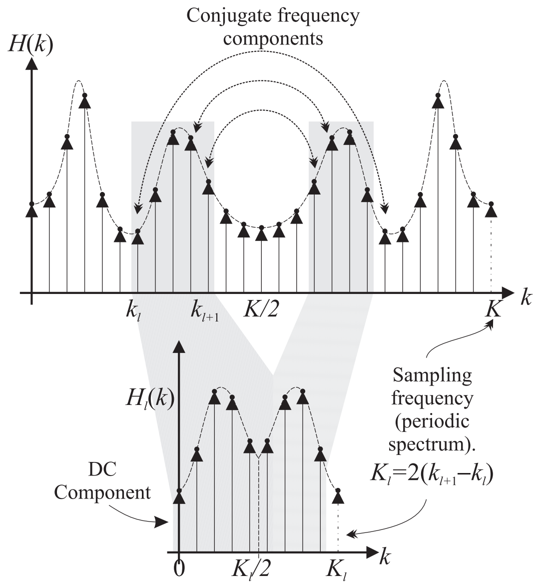

Figure 2: Subband modelling and the indices mapping.

discussed in §4, for simplicity, model-orders, changepoints, and hyperparameters are assumed to be known; of course, in practice, they also need to be estimated.

3. BAYESIAN BLIND DECONVOLUTION

The source, $s(t)$, in block $i \in \mathscr{M}$, as given by (1), may be written as

$$\mathbf{e}_i = \mathbf{s}_i + \mathrm{S}_i\, \mathbf{b}_i$$

where $[\mathbf{e}_i]_{t-t_i+1} = e(t)$, $[\mathbf{s}_i]_{t-t_i+1} = s(t)$, the data matrix is $[\mathrm{S}_i]_{t-t_i+1,q} = s(t-q)$, and $\mathbf{b}_i$ is a vector of parameters $[\mathbf{b}_i]_q = b_i(q)$, and where $t \in \mathscr{T}_i, q \in \mathscr{Q}_i$. Denoting $\sigma = \{\sigma_i^2\}$, $\mathbf{b} = \{\mathbf{b}_i\}$, it is shown in [1] that the likelihood function for the clean speech, $\mathbf{s}$, is

$$p(\mathbf{s} \,|\, \sigma, \mathbf{b}) = \prod_{i=1}^{M} \frac{1}{(\sqrt{2\pi}\sigma_i)^{T_i}} \exp\left\{-\frac{\|\mathbf{s}_i + \mathrm{S}_i \mathbf{b}_i\|^2}{2\sigma_i^2}\right\} \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm. Moreover, due to the causal relationship $x(t) = f(s(t), s(t-1), s(t-2), \ldots)$, the Jacobian, J, in the probability transformation from the random process $s(t)$ to $x(t)$ is a function of the gain parameters, G_l, only, and is given by $J = \prod_l G_l^{K_l}$. Hence, the likelihood function for the observed signal, $\mathbf{x}$, is given by

$$p(\mathbf{x} \,|\, \theta, \sigma, \mathbf{b}, \phi) = \frac{1}{\prod_{l=2}^{L} G_l^{K_l}}\, p(\mathbf{s} \,|\, \sigma, \mathbf{b})$$

where ϕ contains the temporal and spectral changepoints, model orders, and hyperparameters in the priors, and θ is the set of channel parameters including the gain terms. Note that by setting $G_1 = 1$, and observing that the source excitation is constant across all subbands, there is no gain ambiguity between the source signal, and the *relative* subband gains.

Applying Bayes's theorem, the posterior pdf, $p(\theta, \sigma, \mathbf{b} \,|\, \mathbf{x}, \phi)$, for the unknown system parameters given an observation of the state of the system, $\mathbf{x}$, and the assumptions in ϕ, is given by

$$p(\theta, \sigma, \mathbf{b} \,|\, \mathbf{x}, \phi) \propto p(\mathbf{x} \,|\, \theta, \sigma, \mathbf{b}, \phi)\, p(\theta, \sigma, \mathbf{b} \,|\, \phi)$$

where $p(\theta \,|\, \phi)$ represents any prior belief. Assuming $\{b_i, \sigma_i\}$ are independent between blocks, then using the prior densities in §2.3,

Algorithm 1 Inverse filtering of subband channel

Let $X(k) = \mathscr{F}(x(t))$ be the DFT of $\{x(t)\}$. Assume the parameters $a_l(p)$ and G_l are given.
for each subband with boundary frequencies f_l and f_{l+1} **do**
  Determine corresponding frequency bins k_l to $k_{l+1}-1$, and evaluate the baseband AR spectrum:

$$H_l(k) = \frac{K_l}{K} \frac{G_l}{1+\sum_{p=1}^{P_l} a_l(p)\, e^{-\frac{\pi j (k-k_l) p}{k_{l+1}-k_l}}}$$

  for all $k \in \{k_l, \dots, k_{l+1}-1\}$ **do**
    Evaluate $S_{\text{est}}(k) = \frac{X(k)}{H_l(k)}$, and similarly for the corresponding conjugate frequency components (see Figure 2 to see this correspondence).
  end for
end for
Evaluate the inverse DFT $s(t) = \mathscr{F}^{-1}(S_{\text{est}}(k))$.

and marginalising the *nuisance* parameters, $\mathbf{b}$ and σ, it can be shown that the posterior density for the channel parameters is [1]:

$$p(\theta \mid \mathbf{x}, \phi) \propto \frac{1}{\prod_{l=2}^{L} G_l^{K_l}} p(\theta \mid \phi) \times \prod_{i=1}^{M} \frac{\left\{\gamma_i + \mathbf{s}_i^T \mathbf{s}_i - \mathbf{s}_i^T \mathrm{S}_i \left(\mathrm{S}_i^T \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right)^{-1} \mathrm{S}_i^T \mathbf{s}_i\right\}^{-R_i}}{\left|\mathrm{S}_i^T \mathrm{S}_i + \delta_i^{-2} \mathrm{I}_{Q_i}\right|^{\frac{1}{2}}} \tag{7a}$$

where $R_i = \frac{T_i + \nu_i + 1}{2}$, $i \in \mathscr{M}$, and

$$s(t) = f^{-1}(\{x(t)\} \mid \theta) \tag{7b}$$

Equation (7a) is written in terms of $s(t)$ to emphasise that the posterior can be efficiently calculated by 'inverse filtering' the data, $x(t)$; this is achieved as indicated in Algorithm 1, although there are more efficient methods. A maximum marginal *a posteriori* (MMAP) estimate for the parameters θ can by calculated by evaluating:

$$\hat{\theta} = \arg\max_{\theta} p(\theta \mid \mathbf{x}, \phi) \tag{8}$$

In fact, further computational savings can be obtained by noting that since the discrete Fourier transform (DFT) is an orthonormal transformation, the pdf of the DFT of a Gaussian random process is still Gaussian [7]. As a result, equation (7a) can be expressed in an equivalent form where the vectors $\mathbf{s}_i$ and matrices S_i are filled with the DFT signal values rather than the time domain samples. Thus, in Algorithm 1, all the calculations can be performed in the frequency domain. The prior $p(\theta \mid \phi)$ is assumed to be uninformative.

The MMAP estimate for the unknown channel parameters, θ, can by found by solving (8). This optimisation can be performed using deterministic or stochastic optimisation methods; stochastic optimisation methods have a far superior performance when the dimensionality of the parameter space is high, although deterministic methods such as, for example, the Nelder-Mead Simplex method [11] as implemented in MATLAB, are more straightforward to implement. The Gibbs sampler is an example of a stochastic method, and can be used to obtain samples of the channel parameters which can, in turn, be used to to obtain a minimum mean-square error (MMSE) estimate for θ. As discussed in [1], it is reasonable to assume that the MMSE estimate is approximately equal to the MMAP estimate. Details of both of these optimisation are omitted here due to space constraints, although a similar implementation can be found in [1].

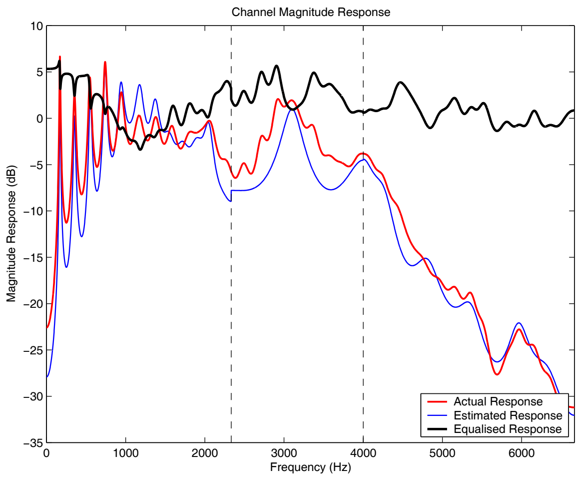

Figure 3: Frequency responses of channel – subband model.

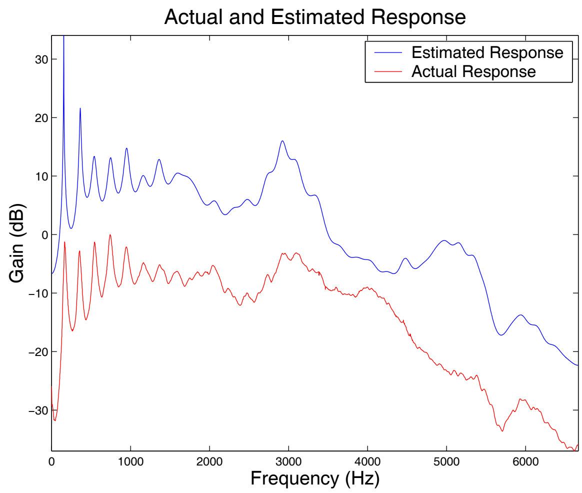

Figure 4: Frequency responses of channel – fullband model.

4. SYNTHETIC EXAMPLE

A synthetic example is presented to demonstrate this approach. It is expected that the results presented here will be extended to a more realistic example in future work. A synthetic BSAR process is used to model the source, whose parameters are generated by modelling discontiguous blocks of a real speech signal as a stationary AR(Q) model and estimating the parameters using the autocorrelation method. This ensures 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 source is modelled as 20th-order, with $L = 20$ blocks, and with each block length $T_l = 500$. The "typical" channel presented in [1, 7], which is that of an acoustic gramophone horn, is also used in these experiments; the magnitude response is shown in Figure 3. In the blind deconvolution problem, the *subband model orders*, *subband boundary positions*, and the *number of subbands* are all unknown and must be chosen. In general, these unknown parameters can be built into a general probabilistic model, and their values estimated by sampling the joint posterior distribution for all these variables using MCMC methods. This is left as further work, and for simplicity, these parameters and all hyperparameters are assumed to be known.

To measure the performance of the algorithm, the ground truth for the channel is known in this experiment. The channel model-orders can be estimated by driving the channel with white noise, and using Akaike's B-Information criterion (BIC), which gives approximately the same solution as if the joint pdf of the AR parameters and model order were maximised. It can be shown that the channel response is accurately modelled over the full frequency band by a 70th-order all-pole model. The parameters for the subband model can be determined using the method described in [7], and if the subband boundaries are chosen to be at angular frequencies $\omega_1 = 0.35\pi$ and $\omega_2 = 0.6\pi$, the required model orders are $P_1 = 24$, $P_2 = 4$ and $P_3 = 13$. Note that these total substantially fewer parameters than for the fullband model.

The synthetic source signal is filtered by the full channel response, and (8) is maximised using the Nelder-Mead Simplex method as implemented in MATLAB.[2]. The estimated channel response is shown in Figure 3, along with the equalised response given by the ratio of the actual and estimated responses. The equalised response is still somewhat *jagged*, but nevertheless flatter than the actual channel response. Note the position of the subband boundaries, as indicated by the vertical dotted lines; there are slight discontinuities at these points. Extensions to the model to prevent such discontinuities are discussed in §5.2. The subband model compares extremely favourably with the results obtained in the fullband case presented in [1], as shown in Figure 4 (an offset between the responses is shown in this figure). In both cases, despite the less than perfect equalised response, listening tests indicate a noticeable improvement.

5. MODEL EXTENSIONS

This section briefly discussions some extensions to the presented model; the performance of these extensions when applied to blind deconvolution are currently being investigated.

5.1 Subband Source Model

The source model in §2.1 is a "full-band" source model. Although speech can be modelled by a relatively low-order BSAR model, a subband BSAR model has the advantage that different frequency bands can have different temporal block lengths over which the signal is modelled as stationary. Hence, there may be some frequency bands over which the speech signal is stationary over a longer period of time: for example, low frequencies, where the pitch of the signal remains approximately constant.

5.2 Subband Boundary Constraints for Channel Model

Since the transfer function is modelled in different subbands independently, discontinuities arise at the subband boundaries. These discontinuities can be avoided by placing a constraint on the subband all-pole parameters to ensure continuity across subbands. This constraint is derived from (5) as follows: note that if $k = k_{l+1}$ in subband $l+1$, then

$$H(k) = \frac{G_{l+1}}{1 + \sum_{p=1}^{P_{l+1}} a_{l+1}(p)} \tag{9a}$$

and $k = k_{l+1}$ in subband l, then:

$$H(k) = \frac{G_l}{1 + \sum_{p=1}^{P_l} (-1)^p a_l(p)} \tag{9b}$$

Equating these leads to the constraint:

$$\sum_{p=1}^{P_{l+1}} a_{l+1}(p) = \left(\frac{G_{l+1}}{G_l} - 1\right) + \frac{G_{l+1}}{G_l} \sum_{p=1}^{P_l} (-1)^p a_l(p) \tag{10}$$

[2]When using the Nelder-Mead Simplex method in a high-dimensional parameter space, the optimisation routine should always be restarted at a point where the algorithm claims to have found a minimum; this is to ensure that the algorithm is not fooled by local minima, or any other anomalies that might influence the convergence criteria [11]

Recall that an all-pole model spectrum has zero gradient at zero frequency, and the half sampling frequency; as a result, this formulation provides continuity in the function value and first derivative of $H(k)$, but not necessarily in the second or higher derivatives.

By using continuity constraints, the unconstrained optimisation exercise in equation (8) becomes constrained. This constraint can be implemented in various ways, either through the prior density $p(\theta \mid \phi)$, or solving (8) subject to the constraint in (10). Unfortunately, this complicates even further an already difficult optimisation problem. An alternative, but suboptimal approach, is to reduce the parameter space of each subband filter by one; this is achieved by rewriting (10) such that, for example, $a_{l+1}(P_{l+1})$ is expressed in terms of the other parameters in that subband, $\{a_{l+1}(1), \ldots, a_{l+1}(P_{l+1} - 1)\}$, G_{l+1}, and the parameters in the adjacent subband $\{a_l(1), \ldots, a_l(P_l)\}$, and G_l. This latter technique has proved successful, and some general details of how this is implemented may be found in [7].

6. CONCLUSIONS

Single channel blind dereverberation has previously been considered in [1], where it is assumed a nonstationary source can be modelled by a BSAR process, and the RTF by an all-pole filter. Unfortunately, this approach suffers from its lack of scalability, and would be unable to deal with, for example, dereverberation of speech measured in a real room where extremely high model-orders are required. A solution is to use a subband model for the RTF. This paper has demonstrated that this approach is feasible and, in the example shown, the performance is equal to the fullband approach. The greatest advantage of the subband model is the ease with which it can be extended to more complicated acoustic impulse responses, by using a greater number of subbands with the smaller model orders in each, such that the overall number of parameters is substantially fewer than in the fullband case.

REFERENCES

[1] J. R. Hopgood and P. J. W. Rayner, "Blind single channel deconvolution using nonstationary signal processing," *IEEE Trans. SAP*, vol. 11, no. 7, pp. 476– 488, Sept. 2003.

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

[3] J. Makhoul, "Spectral analysis of speech by linear prediction," *IEEE Trans. Audio and Electro.*, vol. AU-21, no. 3, pp. 140–148, June 1973.

[4] S. L. Tan and T. R. Fischer, "Linear prediction of subband signals," *IEEE J. Sel. Areas Comms.*, vol. 12, no. 9, pp. 1576–1583, Dec. 1994.

[5] S. Rao and W. A. Pearlman, "Analysis of linear prediction, coding, and spectral estimation from subbands," *IEEE Trans. Information Theory*, vol. 42, no. 4, pp. 1160–1178, July 1996.

[6] J. R. Hopgood and P. J. W. Rayner, "A probabilistic framework for subband autoregressive models applied to room acoustics," in *11th Proc. IEEE SSP*, Aug. 2001, pp. 492 – 494.

[7] J. R. Hopgood and P. J. W. Rayner, "Bayesian formulation of subband autoregressive modelling with boundary continuity constraints," in *Proc. IEEE ICASSP*, Hong Kong, Apr. 2003.

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

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

[10] 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.

[11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 2nd edition, 1992.