Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

ROBUST TREATMENT OF IMPULSIVE NOISE IN SPEECH AND AUDIO SIGNALS*

BY SIMON J. GODSILL AND PETER J.W. RAYNER
Signal Processing and Communications Group
University of Cambridge

Markov chain Monte Carlo methods are presented for treatment of localized, impulsive noise (outliers) in digitized waveforms, within a Bayesian hierarchical framework. Outliers in audio signals occur as 'clicks' and 'crackles' in degraded sound recordings and impulsive noise in communications channels. Sampling-based methods for detection and correction of such artefacts are presented, in which individual noise sources are modelled as Gaussian with unknown scale, allowing for robustness to heavy-tailed noise distributions. Results are presented for speech and audio signals obtained from digitized sound recordings.

1. Introduction. This paper is concerned with the reconstruction of acoustically recorded signals, such as speech and music, which are degraded by impulsive noise sources ('outliers'). In the case of gramophone recordings there are several mechanisms for such defects in the physical storage medium, including natural irregularities, scratches, cracks and dust particles. These all give rise to localized noise artefacts in the recorded sound which are perceived as the 'click' and 'crackle' noise associated with old recordings. In analogue communications channels impulsive noise occurs as a result of electromagnetic interference, switching noise and atmospheric noise, all of which exhibit impulsive properties.

Godsill, Rayner and Cappé (1996) give a thorough review coverage of methods currently available for correction of impulsive and other types of degradation found in audio material. One approach which has been very successful involves modelling the audio signal as an autoregressive (AR) process (Vaseghi and Rayner, 1990). Identification (detection) of outliers is achieved by thresholding the estimated AR innovation sequence, while reconstruction is performed by least-squares interpolation of the corrupted data values. Disadvantages of the method include the inability to detect small impulses in the presence of much larger disturbances as well as the introduction of distortion in the presence of certain signal characteristics. Godsill and Rayner (1992, 1995) have developed recursive Bayesian methods which improve performance by allowing very accurate detection of audio outliers occurring in arbitrary configurations and bursts.

The rapid increases in available computational power which have occurred over the last few years have led to a revival of interest in Markov

*Work Supported by British Library's Film and Audio Restoration Project

chain Monte Carlo (MCMC) simulation methods (Hastings, 1970; Geman and Geman, 1984; Gelfand and Smith, 1990) amongst Bayesian statisticians. The Gibbs Sampler is perhaps the most popular form of MCMC currently in use for the exploration of posterior distributions. This method requires full specification of the conditional posterior densities of any unknown parameters. The sampler is initialized with arbitrary starting values for all the parameters. New values are generated iteratively by sampling in turn from the individual parameter conditional densities with the remaining parameters fixed at their most recent values. Convergence to the joint posterior density is then guaranteed in the limit under mild conditions.

Our work can be related to the non-linear state-space work of Carlin, Polson and Stoffer (1992), in which the Gibbs Sampler is employed for the solution of non-linear state-space systems in non-Gaussian noise. The work of Carter and Kohn (1994) is also relevant, in which observation noise and errors are modelled as Gaussian mixtures, allowing simultaneous generation of the whole state vector using the standard Kalman filter. This strategy is shown to give much improved convergence speed for several examples compared with the more general but univariate sampling schemes of Carlin *et al.* . In the audio processing field the Gibbs Sampler has been applied to the interpolation of missing samples for data that can be modelled as an AR process (Ó Ruanaidh and Fitzgerald, 1994).

The signal and noise models used here are an adaptation of those used by McCulloch and Tsay (1994) for analysis of autoregressive (AR) time series. The main differences are in the switched noise model, which is based on a continuous mixture of Gaussians in order to give greater robustness to heavy-tailed noise sources, and in the discrete Markov chain prior which models prior dependence between outlier indicator variables. The Gibbs Sampler is applied to the problem with several modifications aimed at improving its convergence. In particular, the binary outlier indicator variables and reconstructed signal elements are sampled *jointly* in subgroups in order to overcome a perceived limitation in the McCulloch and Tsay approach to outlier rejection. It is noted that Kalman filter-type operations can be used to generate the reconstructed data samples (in a similar fashion to Carter and Kohn (1994)), but that computational savings will be small when the number of outliers is a small proportion of the total data length.

2. Model and Prior Specifications.

2.1 Noise Specification. The types of degradation we are concerned with here can be regarded as additive and localized in time, which may be represented within the classical additive outlier (AO) framework as

$$y_t = x_t + i_t v_t \tag{1}$$

where y_t is the observed (corrupted) waveform, x_t is the underlying audio signal and $i_t\, v_t$ is an outlier process. i_t is a binary (0/1) variable which indicates outlier positions and v_t is a continuous noise process.

We model v_t as Gaussian with time-varying scale parameter, i.e. $v_t \sim \mathrm{N}(0, \sigma_{v_t}^2)$ where $\sigma_{v_t}^2$ is dependent upon t. Such a model allows for the fact that defects in a single extract of recorded audio material are typically present at widely varying scales. In a gramophone recording, for example, these range from microscopic surface defects in the pressing medium up to large particles of dust adhering to the groove walls. Note that for an application such as this it is not reasonable to assume that noise variances are scaled relative to signal power, since the two physical generation mechanisms are essentially independent.

An independent inverse-gamma prior $\sigma_{v_t}^2 \sim \mathrm{IG}(\alpha_v, \beta_v) = cx^{-(\alpha_v+1)}e^{-\beta_v/x}$ is chosen to express uncertainty about the scale of the noise defects. Some vague prior information about the scale variation will be available from physical considerations of the problem (see section 4) and this can be incorporated into the IG prior by specifying mean and variance values. Such a prior/model combination can be expected to lead to robustness in the presence of heavy-tailed noise distributions, and indeed it is well known that the normal-IG scale mixture is equivalent to the Student-t distribution. West (1984) has proposed the use of scale mixtures of normals for modelling heavy-tailed noise sources from a wide range of distribution families (Andrews and Mallows, 1974), including the Student-t. We restrict attention here to the IG family owing to its convenience as the natural conjugate density for the Gaussian and its simple specification in terms of mean and variance parameters (when $\alpha > 2$). However, we note that the sampling methods described here can easily be extended to other classes of distribution and a sensitivity study would be a useful exercise.

The preceding robustness considerations will be of particular importance in our situation, where a high-fidelity reconstruction of x_t is required. Specifically, the reconstruction procedure must be capable of extracting useful audio information where the scale of defect is small, whilst effectively ignoring data values which are subject to corruption at the largest scales. These objectives are not attainable if a common-variance Gaussian is assumed for the noise distribution.

Finally, a prior must be specified for the indicator vector $\mathbf{i}$. We will refer to this as the *noise generator prior*, $p(\mathbf{i})$. The choice of this prior is not limited by computational considerations since any prior which can easily be evaluated for all $\mathbf{i}$ will fit readily into the sampling framework of the next section. The Bernoulli prior $p(\mathbf{i}) = \alpha^l(1-\alpha)^{N-l}$ (where l is the total number of outliers indicated by $\mathbf{i}$) has typically been used in outlier problems (Box and Tiao, 1968; Abraham and Box, 1979). Such a distribution implies prior independence of the indicators i_t. This is not a reasonable assumption for the defects typical of audio data, where outliers occur in short 'bursts' of adjacent data elements, corresponding to a small area of physical damage in the recorded medium. A correlated prior such as a discrete Markov chain prior $p(\mathbf{i}) = \prod_t P_{i_{t-1} \to i_t}$ may be more appropriate, in which the clustering of outliers in time is modelled by the transition probabilities of the Markov chain, $P_{i \to j}$. The transition probabilities can in principle be treated as unknown and sampled with the rest of the variables in the Gibbs Sampler. However, in this case it is probably more effective to use values which are known to be reasonable from past experience with similar data-sets. We assume the discrete Markov chain prior with fixed transition probabilities for the remainder of this paper.

2.2 Signal Specification. Speech and music signals are known to have a strong local autocorrelation structure in the time domain. This structure can be used for distinguishing between an uncorrupted audio waveform and unwanted noise artefacts. We choose to model the autocorrelation structure in the uncorrupted data sequence x_t as an autoregressive (AR) process whose coefficients a_i are constant in the short term:

$$x_t = \sum_{i=1}^{P} x_{t-i}\, a_i \;+\; e_t \tag{2}$$

where $e_t \sim \mathrm{N}(0, \sigma_e^2)$ is the i.i.d. innovations sequence. In matrix-vector form we have, for N data samples:

$$\mathbf{e} = \mathbf{A}\mathbf{x} = \mathbf{x} - \mathbf{X}\mathbf{a}$$

where the rows of $\mathbf{A}$ and $\mathbf{X}$ are constructed in such a way as to form (2) for successive values of t.

Assuming that the AR parameter vector $\mathbf{a}$ is *a priori* independent of σ_e^2, we assign the simple improper prior $p(\mathbf{a}, \sigma_e^2) \propto \mathrm{IG}(\alpha_e, \beta_e)$. Note that $\mathbf{a}$ and σ_e^2 will generally be well supported by the data since $\mathbf{x}$ will contain many hundreds of data values. A very vague prior on σ_e^2 will then suffice.

It is of course an approximation to assume that the AR parameters and excitation variance remain fixed in any given block of data. A time-varying system might be a more realistic representation of the signal and should certainly be considered in future adaptations to this work. However, we have found this assumption to be much less critical in practice than choice of an appropriate impulsive noise model, since the signal parameters for typical speech and audio signals typically vary slowly and smoothly with time.

3. Gibbs Sampler. The Gibbs Sampler may be used to sample from the joint posterior distribution for all the unknowns. The conditional densities for model parameters $\boldsymbol{\theta} = \{\mathbf{a}, \sigma_e^2, \sigma_{v_t}^2\ \forall t\}$ are obtained straightforwardly as

$$(3) \qquad (\mathbf{a}|\mathbf{i}, \mathbf{x}, \boldsymbol{\theta}_{-(\mathbf{a})}, \mathbf{y}) \sim \mathrm{N}_P(\, \mathbf{a}^{\mathrm{MAP}},\, \sigma_e^2\,(\mathbf{X}^T\mathbf{X})^{-1})$$

$$(4) \qquad (\sigma_e^2|\mathbf{i}, \mathbf{x}, \boldsymbol{\theta}_{-(\sigma_e^2)}, \mathbf{y}) \sim \mathrm{IG}(\, \alpha_e + (N-P)/2,\, \beta_e + \sum_t {e_t}^2/2\,)$$

$$(5) \qquad (\sigma_{v_t}^2|\mathbf{i}, \mathbf{x}, \boldsymbol{\theta}_{-(\sigma_{v_t}^2)}, \mathbf{y}) \sim \begin{cases} \mathrm{IG}(\, \alpha_v + 1/2,\, \beta_v + {v_t}^2/2) & \forall\, t : i_t = 1 \\ \mathrm{IG}(\, \alpha_v,\, \beta_v\,) & \textit{otherwise} \end{cases}$$

where $(b|c)$ denotes a conditional probability distribution, $\boldsymbol{\theta}_{-(\mathbf{a})}$ denotes all members of $\boldsymbol{\theta}$ except $\mathbf{a}$ and $\mathbf{a}^{\mathrm{MAP}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}$ is the conditional MAP estimate for the AR parameters. Note that this estimate uses the approximate likelihood in which the likelihood for the first P data samples is neglected (Box and Jenkins, 1970, Appendix A7.5), but that the full likelihood can be incorporated by a simple Hastings-Metropolis (Hastings, 1970) modification to (3). The terms e_t and v_t, required in (4) and (5) can be obtained for a particular $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{a}$ from (1) and (2). Note also that when $i_t = 0$, (5) requires that we sample from the prior on $\sigma_{v_t}^2$, since there is then no information available from the data or other parameters about this hyperparameter.

Now consider the remaining unknowns, $\mathbf{x}$ and $\mathbf{i}$. The 'detection' of outliers involves estimation of discrete vector $\mathbf{i}$, whilst reconstruction requires estimation of $\mathbf{x}$. McCulloch and Tsay (1994) propose a Gibbs sampling solution to a closely related outlier problem, which is similar in principle to George and McCulloch's (1993) variable selection method. Their approach performs univariate sampling from the conditionals for all the elements i_t and v_t (from which the sampled reconstruction can be obtained as $x_t = y_t - i_t v_t$). This scheme has two drawbacks which will affect convergence of the sampler. Firstly, sampling is univariate even though there are likely to be strong posterior correlations between successive elements of i_t and v_t. Secondly, the scheme involves sampling from the prior for v_t when $i_t = 0$, since v_t is conditionally independent of the input data in this situation. i_t will then only stand a reasonable chance of detecting a true outlier when v_t happens to take a sampled value close to its 'true' value, with the result that the evolution of i_t with iteration number is rather slow. Note that in equation (5) of our scheme it is necessary to sample from the prior on $\sigma_{v_t}^2$ when $i_t = 0$. However, at this level of the hierarchy the convergence of i_t is less likely to be affected adversely. The drawbacks of univariate sampling can be overcome by sampling from the joint multivariate conditional for $\mathbf{x}$ and $\mathbf{i}$. Note that this is distinct from the approach of Carter and Kohn (1994), in which indicator variables ($\mathbf{i}$) are sampled jointly *conditional* upon state variables

($\mathbf{x}$), and *vice versa*. The joint conditional expression is specified by

$$(\mathbf{i}|\boldsymbol{\theta},\mathbf{y}) \sim c\,p(\mathbf{i})\,\frac{(\sigma_e^2)^{l/2}\exp(-S^2|_{\mathbf{x}_{(\mathrm{i})}=\mathbf{x}_{(\mathrm{i})}^{\mathrm{MAP}}})}{(2\pi\sigma_e^2)^{(N-P)/2}\,|\mathbf{R}_{\mathbf{v}_{(\mathrm{i})}}|^{1/2}|\,|\boldsymbol{\Phi}|^{1/2}} \tag{6}$$

$$(\mathbf{x}_{(\mathrm{i})}|\mathbf{i},\boldsymbol{\theta},\mathbf{y}) \sim \mathrm{N}_l(\,\mathbf{x}_{(\mathrm{i})}^{\mathrm{MAP}},\,\sigma_e^2\,\boldsymbol{\Phi}^{-1}\,) \tag{7}$$

where

$$\begin{aligned} S^2 &= \sum_t {e_t}^2/2\sigma_e^2 \;+\; \sum_{\{t:\,i_t=1\}} {v_t}^2/2\sigma_{v_t}^2 \\ \boldsymbol{\Phi} &= \mathbf{A}_{(\mathrm{i})}^T\mathbf{A}_{(\mathrm{i})} \;+\; \sigma_e^2\mathbf{R}_{\mathbf{v}_{(\mathrm{i})}}^{-1} \\ \mathbf{x}_{(\mathrm{i})}^{\mathrm{MAP}} &= -\boldsymbol{\Phi}^{-1}\left(\mathbf{A}_{(\mathrm{i})}^T\mathbf{A}_{-(\mathrm{i})}\,\mathbf{y}_{-(\mathrm{i})} \;-\; \sigma_e^2\,\mathbf{R}_{\mathbf{v}_{(\mathrm{i})}}^{-1}\mathbf{y}_{(\mathrm{i})}\right) \end{aligned} \tag{8}$$

Here we have introduced a notation for vector/matrix partitioning such that subscript '$_{(\mathrm{i})}$' denotes elements/columns corresponding to corrupted data (i.e. $i_t = 1$), while '$_{-(\mathrm{i})}$' denotes the remaining elements/columns. $\mathbf{R}_{\mathbf{v}_{(\mathrm{i})}}$ is the covariance matrix for the l corrupting outliers indicated by $\mathbf{i}$, and is diagonal with elements $\sigma_{v_t}^2$ in this case. l is obtained directly from a particular $\mathbf{i}$ as the number of non-zero elements in $\mathbf{i}$. Note that S^2 in (6) is evaluated using (8) with the conditional MAP reconstruction $\mathbf{x}_{(\mathrm{i})}^{\mathrm{MAP}}$ substituted into $\mathbf{x}$ for $\mathbf{x}_{(\mathrm{i})}$. The marginal conditional for $\mathbf{i}$, given in (6), is derived straightforwardly from the joint conditional using multivariate normal identities. Full details can be found in Godsill and Rayner (1995). As discussed earlier the noise generator prior $p(\mathbf{i})$ can take any form which is easily evaluated for any given $\mathbf{i}$. In our results, however, we assume the 2-state discrete Markov chain prior with fixed transition probabilities.

As in (3) the results of (6) and (7) are based on the approximate likelihood expression, for the sake of simplicity. Incorporation of the full likelihood is achieved by a minor modification which maintains the Gaussian form of (7) (Box and Jenkins, 1970, Appendix A7.5) and hence makes no fundamental change to the sampling algorithm. This modification is required only for accurate detection of outliers which occur in the first P samples of data.

Joint sampling of $\mathbf{i}$ and $\mathbf{x}$ now involves sampling in turn from (6) and (7). (6) is a multivariate discrete distribution defined for all 2^N permutations of $\mathbf{i}$. Direct sampling from this distribution will thus require 2^N evaluations of (6), so some other scheme must be adopted. A Metropolis-Hastings step (Hastings (1970)) is certainly one possibility for this challenging task, although we can expect low acceptance rates due to the high dimensionality of $\mathbf{i}$. Here we adopt a Gibbs sampling approach which performs the sampling given by (6) and (7) in sub-blocks of size q. A compromise can thus be achieved between computational load and convergence performance. Sampling in sub-blocks is achieved simply by fixing elements of i_t and y_t to 0 and the current sample of x_t, respectively, for all elements which lie outside the current sub-block.

Compared with the $O(2^N)$ sampling operation for $\mathbf{i}$ in (6) there is much less computational difficulty in the full multidimensional sampling operation of $\mathbf{x}_{(\mathbf{i})}$ (7) (an operation which is at most $O(l^3)$, depending on the spacing between 'bursts' of outliers), especially when the percentage of outliers is small. Thus this operation can be performed on an occasional basis for the whole data block or for large sub-blocks, in addition to the joint sampling of $\mathbf{i}$ and $\mathbf{x}_{(\mathbf{i})}$ in small sub-blocks of size q. This should alleviate any convergence problems which may arise as a result of posterior correlation between the x_t's for a given $\mathbf{i}$.

4. Implementational Issues and Results. The method was tested using data digitally sampled at 44.1kHz and 16-bit integer resolution from degraded gramophone recordings containing speech and music material. It is typically realistic to assume that the autoregressive process parameters are fixed for time intervals up to 25ms in speech and music signals. Hence data block lengths N of around 1100 samples will be appropriate. Processing for listening purposes can then proceed sequentially through the data, block by block, re-initializing the sampler for each new block and storing the reconstructed output data sequentially in a new data file. When processing for a particular extract is completed, the output data file may be listened to for appraisal purposes using a suitable digital-to-analogue conversion system.

The hyperparameters for the prior on the noise variances, α_v and β_v, are chosen initially from the physical constraints of the measurement (digitization) system and the assumption that the transfer engineer has adjusted the overall digitized signal levels to be centered (on a dB scale) within the 16-bit integer range. After the first blocks of data have been processed, the prior parameters may be informally updated, based on the noise variance distributions from earlier blocks. This relies on the assumption that noise characteristics remain roughly unchanged throughout a particular extract, which is empirically observed to be the case for most recordings. A very vague proper IG prior is chosen for the innovations variance σ_e^2, since there is likely to be considerable support for this parameter from the data.

In figure 1 we can see the results obtained from running one instance of the Gibbs Sampler for a single block of classical music, digitized from a typical noisy 78rpm recording (figure 1(a) shows this data block). An AR model order $P = 30$ was chosen, which is adequate for representation of moderately complex classical music extracts. Each iteration started with steps (3)-(5). This was followed by steps (6) and (7), using a sub-block size $q = 3$. Finally step (7) was performed for the entire block, as discussed. Hyperparameters on the noise variance priors were fixed at $\alpha_e = \beta_e = 10^{-10}$ (see above) and

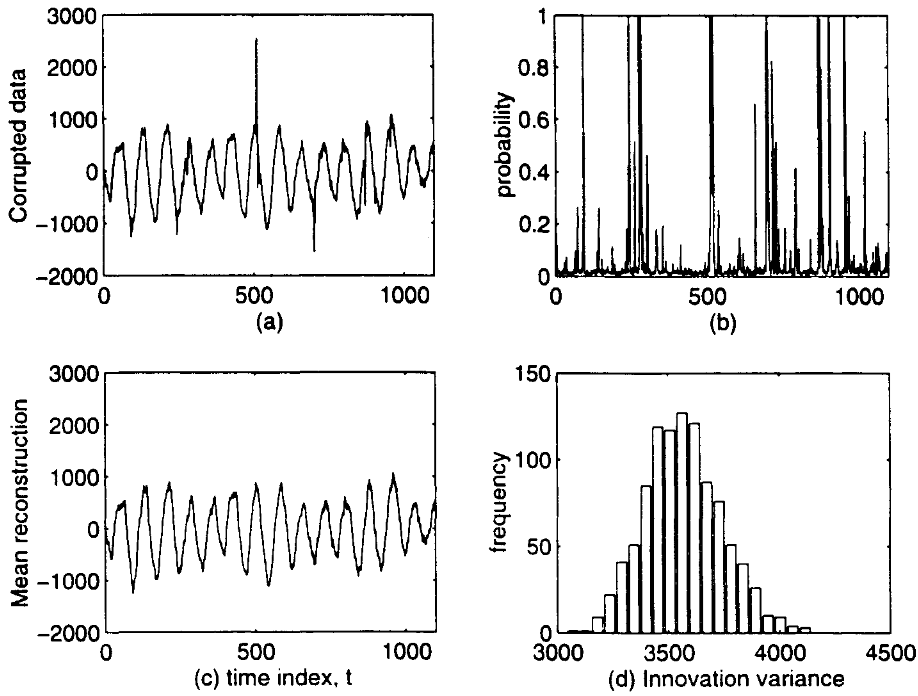

Figure 1: *Reconstruction and detection estimates*

$\alpha_v = 0.8, \beta_v = 10000$ (chosen approximately on the basis of earlier processed data blocks). In fact the reconstructed output has been found to be rather insensitive to the precise value chosen for these parameters. In this case, for example, values of α_v ranging between 0.3 and 10 and β_v from 2000 to 500000 led to reconstructions which were largely indistinguishable to the eye.

The sampler was run for 1000 iterations, following a 'burn-in' period of 100 iterations. Figure 1(b) shows the histogram of sampled indicator vectors $\mathbf{i}$, which gives an estimate of the marginal detection probabilities $(i_t|\mathbf{y})$. While most samples appear to have a non-zero posterior probability of being an outlier, time indices corresponding to high probabilities can be identified clearly with 'spikes' in the input waveform, figure 1(a). Figure 1(c) shows the estimated posterior expectation of the reconstructed data $\mathbf{x}$ obtained by taking the mean of the reconstructions sampled after the burn-in period, while Figure 1(d) shows the histogram of sampled σ_e^2 values following burn-in. The distribution is quite well-determined around its mode.

Initialization of the sampler can be achieved in many ways but, if the method is to be of practical value, we should choose starting points which lead to very rapid and reliable convergence. The AR model parameters and excitation variance are initialized to maximum likelihood estimates obtained as if the data were uncorrupted. This should be a fairly robust starting point when only a small proportion of the data are corrupted. Noise variances are initially sampled from their inverse Gamma prior. Probably the most critical initialization is for the detection vector **i**. Two schemes are considered: assign all zeros to **i** or assign a robust estimate obtained from some other simple detection procedure. The first scheme is used for the graphical results presented here. From visual examination of the reconstructed output samples, the algorithm appears to converge within 10 iterations, and this is supported empirically by examination of the evolution with time of the individual unknowns. Even faster convergence is observed when the second initialization scheme is used for **i**. A rough and ready initial value for **i** is obtained by thresholding the estimated AR innovations sequence corresponding to the corrupted data (Godsill *et al.* 1996), with a threshold set low enough to detect all sizeable outliers. This gives the algorithm a good starting point from which the detection vector **i** will converge very rapidly.

The first 20 iterations of σ_e^2 are displayed in figure 2(a) and (b) under various sub-block sampling schemes and initializing **i** to be all zeros. Under our proposed scheme elements of i_t and x_t are sampled *jointly* in sub-blocks of size q. This is contrasted with the independent sampling scheme used by McCulloch and Tsay (1994), applied here in sub-blocks. Figure 2(a) shows the comparison for the minimal sub-block size $q = 1$. The joint sampling scheme is seen to converge significantly faster than the independent scheme which did not in fact reach the 'true' value of σ_e^2 for many hundreds of iterations. In figure 2(b) the sub-block size is 4. Once again convergence is significantly faster for the joint sampling method, although the independent method does this time converge successfully after approximately 10 iterations. Note that the initial rate of change in σ_e^2 does not depend strongly on the block length chosen. This is as a result of the additional reconstruction operation (7) which is performed each iteration for the whole data block and helps to reduce the dependency on q. Thus we recommend that a small value of q is used in a practical situation. The convergence time of the independent sampling scheme was also found to be far less reliable than the joint scheme since it can take many iterations before very large outliers are first detected under the independent scheme. The differences in convergence demonstrated here will be a significant factor in an application such as this where processing speed is of the essence.

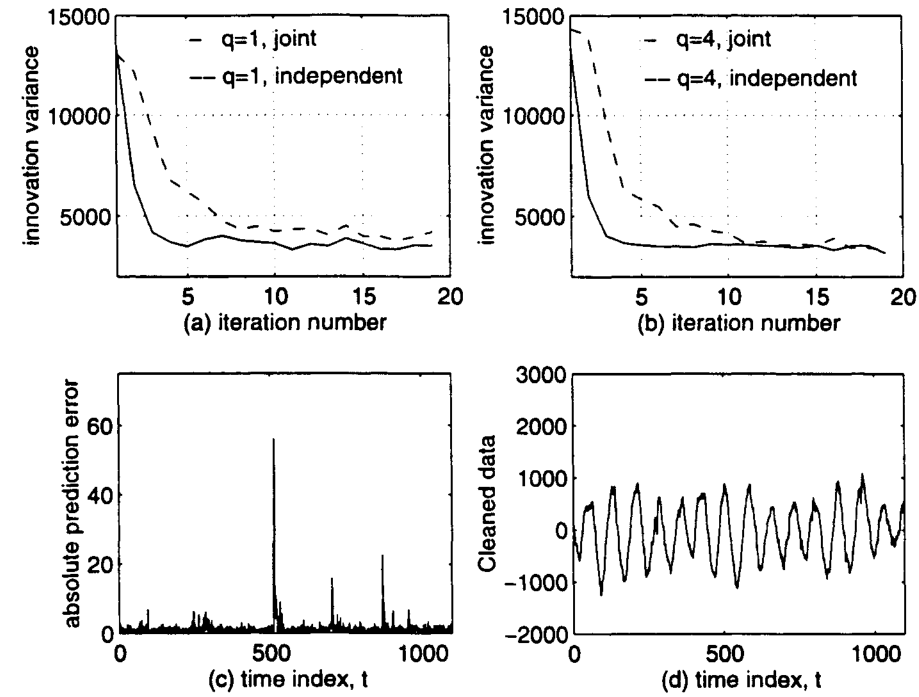

Figure 2: *(a) and (b): Convergence of innovations variance under different sampling strategies. (c) Prediction error from* $\mathbf{y}$. *(d) 'Cleaned' data using thresholded prediction error for detection*

From a computational point of view it will not be possible to run the sampler for very many iterations if the scheme is to be any practical use in audio restoration. Longer recorded extracts processed using the sampler were thus limited to 15 iterations per data block. Restorations were taken as either the mean of the last 5 reconstructions or the last sampled reconstruction from the chain. This latter approach is appropriate for listening purposes since it can be regarded as a *typical* realization of the restored data. Informal listening tests show that either scheme leads perceptually to very high quality restorations. The processed material is rendered entirely free from audible clicks and crackles in one single procedure, with minimal distortion of the underlying audio signal quality. The same degree of noise reduction is not achievable by any other single procedure currently available for audio restoration, and certainly not without much greater distortion of the sound quality.

For visual comparison purposes we compare results with those obtained from the detection and interpolation procedures commonly used for restoration of AR-modelled audio signals, as described in Vaseghi and Rayner (1990) and Godsill *et al.* (1996). AR coefficients $\mathbf{a}$ and innovations variance σ_e^2 are estimated by maximum likelihood from the corrupted data sequence and the corresponding prediction error sequence is calculated by inverse filtering. Outliers are detected at those samples whose prediction error magnitude lies above a chosen factor times the prediction error's standard deviation. Corrupted samples are then restored by maximum likelihood missing-data interpolation, conditional upon the AR coefficients (Janssen, Veldhuis and Vries, (1986)). The whole procedure can be iterated to improve robustness. The results shown in figure 2(c) and (d) used AR coefficients estimated from the Gibbs Sampler output in order to see a comparison with the 'best case' of the standard method. A detection threshold of $3\sigma_e$ was applied to the prediction error sequence, figure 2(c), leading to the 'cleaned' data sequence of figure 2(d). Comparison with the Gibbs Sampler output (figure 1 (c)) shows undetected outliers around sample numbers 250 and 950 which are likely to be audible in listening tests. Lower detection thresholds can of course lead to the removal of all outliers but at the expense of many 'false alarm' detections and a corresponding reduction in sound quality as compared with the Gibbs Sampler.

REFERENCES

ABRAHAM, B. AND BOX, G. E. P. (1979) Bayesian analysis of some outlier problems in time series. *Biometrika* **66**(2) 229–36.

ANDREWS, D. F. AND MALLOWS, C. L. (1974) Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36** 99–102.

BOX, G. E. P. AND JENKINS, G. M. (1970) *Time Series Analysis, Forecasting and Control*. Holden-Day.

BOX, G. E. P. AND TIAO, G. C. (1968) A Bayesian approach to some outlier problems. *Biometrika* **55** 119–129.

CARLIN, B. P., POLSON, N. G. AND STOFFER, D. S. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of American Statistical Association* **87**(418) 493–500.

CARTER, C. K. AND KOHN, R. (1994) On Gibbs Sampling for state space models. *Biometrika* **81**(3) 541–553.

GELFAND, A. E. AND SMITH, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85** 398–409.

GEMAN, S. AND GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721–741.

GEORGE, E. I. AND MCCULLOCH, R. E. (1993) Variable selection via Gibbs sampling. *Journal of American Statistical Association* **88**(423) 881–889.

GODSILL, S. J. AND RAYNER, P. J. W. (1992) A Bayesian approach to the detection and correction of bursts of errors in audio signals. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* **II** 261–264. San Francisco.

GODSILL, S. J. AND RAYNER, P. J. W. (1995) A Bayesian approach to the restoration of degraded audio signals. *IEEE Trans. on Speech and Audio Processing* **3**(4) 267–278.

GODSILL, S. J., RAYNER, P. J. W. AND CAPPÉ, O. (1996) *Digital Audio Restoration. In* Applications of Digital Signal Processing to Audio and Acoustics, K. Brandenburg and M. Kahrs, eds. , Kluwer Academic Publishers (*to appear*).

HASTINGS, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

JANSSEN, A. J. E. M., VELDHUIS, R. AND VRIES, L. B. (1986) Adaptive interpolation of discrete-time signals that can be modeled as AR processes. *IEEE Trans. Acoustics, Speech and Signal Processing* **ASSP-34**(2) 317–330.

MCCULLOCH, R. E. AND TSAY, R. S. (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* **15**(2) 235–250.

Ó RUANAIDH, J. J. K. AND FITZGERALD, W. J. (1994) Interpolation of missing samples for audio restoration. *Electronic Letters* **30**(8).

VASEGHI, S. V. AND RAYNER, P. J. W. (1990) Detection and suppression of impulsive noise in speech communication systems. *IEE Proceedings, Part 1* **137**(1) 38–46.

WEST, M. (1984) Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society, Series B* **46**(3) 431–439.

SIGNAL PROCESSING AND COMMUNICATIONS GROUP
CAMBRIDGE UNIVERSITY ENGINEERING DEPARTMENT
CAMBRIDGE, CB2 1PZ
ENGLAND