JOINT BAYESIAN REMOVAL OF IMPULSE AND BACKGROUND NOISE

James Murphy* and Simon Godsill

Cambridge University
Department of Engineering
Cambridge, CB2 1PZ, U.K.
{jm362,sjg}@eng.cam.ac.uk

ABSTRACT

We present a method for the removal of noise including non-Gaussian impulses from a signal. Impulse noise is removed jointly a homogenous Gaussian noise floor using a Gabor regression model [1]. The problem is formulated in a joint Bayesian framework and we use a Gibbs MCMC sampler to estimate parameters. We show how to deal with variable magnitude impulses using a shifted inverse gamma distribution for their variance. Our results show improved signal to noise ratios and perceived audio quality by explicitly modelling impulses with a discrete switching process and a new heavy-tailed amplitude model.

Index Terms— Impulse, Noise removal, Gabor, MCMC

1. INTRODUCTION

Noise reduction is an important component of audio restoration and enhancement which aims to improve the perceived quality of corrupted audio signals. Early work in noise reduction can be found in [2], but the area continues to be active, e.g. [3, 4]. An overview of a range of methods can be found in [5] and the references therein, but alternative psychoacoustically-based approaches such as [6] have also been popular. A technique common to several methods is the representation of the signal as a weighted sum of basis functions. Common choices of basis function are modified discrete cosines functions [7], which provide an orthogonal basis, and Gabor functions [1, 8], which do not. Such decompositions are useful for a range of audio processing tasks including noise reduction [7] and missing data interpolation [8].

Since the composition of audio signals varies with time, decomposition is performed in blocks on short sub-samples of the whole signal and, in order to reduce blocking effects, these sub-samples generally overlap. This leads to multiple possible decompositions of the signal into the (local) basis functions (overcompleteness). Of these, sparse representations are frequently preferred as they give a parsimonious representation of the original signal and numerous methods of finding the best sparse signal respresentation exist [9, 10, 1]. Here we follow the work of [1] and [8] and use a Bayesian problem formulation combined with a Markov chain Monte Carlo (MCMC) sampler.

In [1] and [8], Gabor decomposition is used to remove homogeneous background noise. However, an important form of noise in audio signals is impulse noise, which takes the form of large but brief deviations between the observed value and the true signal. Impulses can be caused by, amongst other things, wear, dirt or scratches

on vinyl records and are perceived as audible pops and clicks in a recording. Because they can derive from a number of sources, their size can vary across multiple orders of magnitude. We aim to extend the Bayesian Gabor framework of [1] and [8] to allow the removal of impulsive noise. Much previous impulse removal work has been carried out using autoregressive methods, described in [5], for example.

1.1. Gabor Noise Reduction

In [1] the authors use *Gabor analysis* to decompose a signal into a weighted sum of time- and frequency-shifted versions of a single Gabor window function $\tilde{g}$. This is typically a smooth function with compact support. The shifted versions of it are known as *Gabor synthesis atoms* $\tilde{g}_{m,n}$ and can be thought of as lying at discrete points (m, n) on a time-frequency plane. Markov chain Monte Carlo (MCMC) is used in [1] to determine the weighting coefficient $c_{m,n}$ for each atom in order to represent a given input signal $x(t)$:

$$x(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{m,n} c_{m,n} \tilde{g}_{m,n}(t). \tag{1}$$

Typically the atoms used in decomposing audio signals overlap each other in both the time and frequency domains, so coefficient determination is complicated by the fact that there is not a unique representation of the signal. A parsimonious, sparse representation is favoured, however, through the use of indicator variables $\gamma_{m,n} \in \{0, 1\}$ in conjunction with the weighting coefficients that determine whether or not a particular atom is included in the final representation. An advantage of this formulation is that time and frequency dependence can be incorporated through the use of appropriate priors on these indicators (for example a Markov random field structure or a Markov chain preferring time or frequency persistence, depending on the signal and application, see [1] and [8]).

Such analysis can be useful for practical purposes including noise removal, where we assume that the noisy signal observations y_t are the true signal x at time t corrupted by additive noise:

$$y_t = x(t) + v_t, \qquad v_t \sim \mathcal{N}(0, \sigma_v^2). \tag{2}$$

By using this model as the basis for the analysis, the corrupting noise can be removed; see [1]. In [8] a similar approach is used as the basis of a technique to interpolate missing data in audio signals.

Our work extends that of [1] and [8] to deal with cases where impulse noises are present in the signal. We introduce a new latent process z_t which has the same homogeneous noise structure as the processes treated in [1], allowing the removal of impulse noises

---

within a similar Bayesian framework as that for standard Gabor analysis. This allows removal of both impulsive and additive background noise be removed jointly from the signal. We also introduce a shifted inverse gamma distribution as a model for the impulse noise variance. This leads to a variance-mixture model for the impulse noise and permits the removal of impulses with a wide range of magnitudes.

The basic method with fixed impulse scale is presented in section 2; section 3 extends this to variable scale impulses; section 4 gives results comparing the algorithms and section 5 draws conclusions.

2. NON-GAUSSIAN NOISE MODEL

The model in [1] can be extended to accommodate impulse noise by allowing the variance of the noise term at each time $\sigma_{v_t}^2$ to take one of two values, 'normal' or 'impulsive', depending on an indicator i_t, so that

$$v_t \sim \mathcal{N}(0, \sigma_{v_t}^2), \tag{3}$$

with

$$\sigma_{v_t}^2 = (1 + i_t \lambda)\sigma^2, \ \ i_t \in \{0, 1\}, \tag{4}$$

and where $\lambda > 0$ is a scale factor relating the magnitude of impulse noise to that of background noise (typically $\lambda \gg 0$). Thus the noise variance is σ^2 when no impulse is present and $(1 + \lambda)\sigma^2$ when it is.

Impulse noise can be removed using an MCMC scheme to estimate the noise variance, impulse scale factor, the impulse indicator variables and ultimately the signal value x_t. Such schemes draw samples from the posterior distribution of the variables of interest and thus can be used to estimate them; see e.g. [11] for an overview. We use Gibbs sampling, a type of MCMC in which each variable (or block of variables) is updated one at a time (or in blocks), drawing the new value from the distribution of that variable conditional on all other variables in the state space.

The Gabor analysis described in [1] also uses a Gibbs sampling scheme and so is easy to integrate with our method. However, it is designed to work on processes with homogeneous Gaussian noise. In order to use it with our impulse noise reduction, we introduce a second, artificial, latent process z_t with the required noise distribution, to which we can apply the Gabor analysis in an efficient manner:

$$z_t = x_t + w_t \tag{5}$$

with $w_t \sim \mathcal{N}(0, \sigma^2)$. The observed process y_t is then given by

$$y_t = z_t + i_t u_t, \tag{6}$$

with

$$u_t \sim \mathcal{N}(0, \lambda \sigma^2). \tag{7}$$

This structure has the property that

$$
\begin{aligned}
p(x \mid y, z, i) &\propto p(y \mid z, i, x)p(z, i \mid x)p(x) \\
&= p(y \mid z, i)p(i)p(z \mid x)p(x) \\
&\propto p(x \mid z),
\end{aligned} \tag{8}
$$

where x denotes the collection of x_t variables at all times $\{x_t; t \in 0, .., T\}$, and similarly for other variables. This means that the posterior distribution $p(x \mid z)$ is the same as that in [1], since the z process is simply the x process perturbed by homogeneous Gaussian noise. The sampling of the decomposition of x and its parameters, conditional on z, can thus be performed in exactly the same way as in [1].

A sampling iteration, therefore, will consist of sampling the z process as described below, followed by sampling the x process through its Gabor decomposition, as described in [1].

In order to sample the z and i processes jointly in a block-based Gibbs sampler, we require the distribution of their components at a particular time t conditional on everything else (x, y and the values of i and z at all times other than t, denoted i_{-t} and z_{-t}). This is given by

$$p(z_t, i_t \mid x, y, i_{-t}, z_{-t}) = p(z_t \mid i_t, x, i_{-t}, z_{-t})p(i_t \mid x, i_{-t}, z_{-t}). \tag{9}$$

To sample from this distribution, we may draw a sample from $p(i_t \mid x, y, i_{-t}, z_{-t})$ followed by a draw from $p(z_t \mid i_t, x, y, i_{-t}, z_{-t})$. We can come up with expressions for both of these distributions from which we can easily draw samples. For i_t,

$$
\begin{aligned}
p(i_t \mid x, i_{-t}, z_{-t}) &\propto p(i_t \mid i_{-t})p((y - x)_t \mid i_t) \\
&= p(i_t \mid i_{-t})\mathcal{N}\left(0, (1 + i_t \lambda)\sigma^2\right). \tag{10}
\end{aligned}
$$

In order to allow for temporal dependence in the impulses (since an impulse is likely to last for a number of digital samples) we model the impulse indicator as a two-state Markov chain. Its distribution $p(i_t \mid i_{-t})$ is then given by the transition probabilities for the Markov chain, which could themselves be sampled as parameters of the process:

$$p(i_t \mid i_{-t}) \propto p(i_{t+1} \mid i_t)p(i_t \mid i_{t-1}). \tag{11}$$

Since the impulse indicator can only take one of two values we can sample directly from the distribution in equation (10) by evaluating the expression for both $i_t = 0$ and $i_t = 1$ and normalizing to give the probabilities for a sample from a Bernoulli distribution.

Once i_t has been sampled, we can sample z_t from its conditional distribution

$$
\begin{aligned}
p(z_t \mid i_t, x, y, i_{-t}, z_{-t}) &\propto p(y_t \mid z_t, i_t)p(z_t \mid x_t) \\
&= \mathcal{N}(y_t \mid z_t, i_t \lambda \sigma^2)\mathcal{N}(z_t \mid x_t, \sigma^2) \\
&\propto \mathcal{N}\left(z_t \mid \frac{y_t + i_t \lambda x_t}{1 + i_t \lambda}, \frac{i_t \lambda \sigma^2}{1 + i_t \lambda}\right) \tag{12}
\end{aligned}
$$

Note that if $i_t = 0$ then $z_t = y_t$.

These distributions allow us to sample i_t and z_t directly using Gibbs sampling for each time t.

3. SHIFTED INVERSE GAMMA NOISE VARIANCE MODEL

Impulsive noise can originate from a number of different physical sources and so in some cases a single scale factor λ might not lead to a noise distribution sufficiently heavy-tailed to capture all impulses. We can extend the above model to allow the scale factor to vary with time, replacing the single scale parameter λ with a scale λ_t sampled at each sample time. The process noise variance given in equation (4) is replaced with

$$\sigma_{v_t}^2 = (1 + i_t \lambda_t)\sigma^2, \ \ i_t \in \{0, 1\}. \tag{13}$$

Although in principle many prior structures $p(\lambda_t)$ are possible for λ_t, a convenient one, as used in [4] in a different context, is a shifted inverse gamma model. This is a truncated and shifted version

262

of the inverted gamma distribution (note the offset of +1 in the λ_t arguments) and takes the following form:

$$p(\lambda_t) = \frac{\beta_\lambda^{\alpha_\lambda}(1+\lambda_t)^{-(\alpha_\lambda+1)}\exp(-\beta_\lambda/(1+\lambda_t))}{\gamma(\alpha_\lambda,\beta_\lambda)}, \quad \lambda \geq 0,$$
$$\propto \mathcal{IG}(1+\lambda_t;\alpha_\lambda,\beta_\lambda) \tag{14}$$

where $\mathcal{IG}(1+\lambda_t;\alpha_\lambda,\beta_\lambda)$ is the inverse gamma pdf with parameters α_λ and β_λ, evaluated at $1+\lambda_t$ and $\gamma(\alpha_\lambda,\beta_\lambda)$ is the lower incomplete gamma function defined as

$$\gamma(\alpha_\lambda,\beta_\lambda) = \int_0^\beta t^{\alpha-1}e^{-t}dt. \tag{15}$$

We need to sample i, λ and z at each sample point so, in order to use Gibbs sampling, we require their distribution at time t, given all other state variables

$$p(i_t, z_t, \lambda_t \mid x, y, i_{-t}, z_{-t}, \lambda_{-t}, \sigma^2) =$$
$$p(z_t \mid i_t, \lambda_t, x_t, y_t, \sigma^2)p(\lambda_t \mid i_t, x_t, y_t, \sigma^2)p(i_t \mid i_{-t}, x_t, y_t, \sigma^2). \tag{16}$$

As before, we can jointly sample i_t, λ_t and z_t by sampling sequentially (in that order) from the distributions on the right of equation (16).

The distribution from which to sample i_t is given by

$$p(i_t \mid i_{-t}, x_t, y_t, \sigma^2) \propto p(i_t \mid i_{-t})p(y_t \mid x_t, i_t, \sigma^2), \tag{17}$$

with

$$p(y_t \mid x_t, i_t, \sigma^2) = \begin{cases} \mathcal{N}(y_t \mid x_t, \sigma^2), & i_t = 0 \\ p(y_t \mid x_t, i_t = 1, \sigma^2), & i_t = 1. \end{cases} \tag{18}$$

Thanks to the special form of the prior $p(\lambda_t)$ it is possible to find $p(y_t \mid x_t, i_t = 1, \sigma^2)$ in closed form, as described in [4]:

$$p(y_t \mid x_t, i_t = 1, \sigma^2) = \int_0^\infty p(y_t \mid \lambda_t, x_t, i_t = 1, \sigma^2)p(\lambda_t)d\lambda_t$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}}\frac{\gamma(\alpha_p,\beta_p)}{\gamma(\alpha_\lambda,\beta_\lambda)}\frac{\beta_\lambda^{\alpha_\lambda}}{\beta_p^{\alpha_p}}, \tag{19}$$

where $\alpha_p = \alpha_\lambda + 1/2$ and $\beta_p = \beta_\lambda + \frac{(y_t - x_t)^2}{2\sigma^2}$.

Again, we can sample from this distribution by evaluating the expressions for the cases $i_t = 0$ and $i_t = 1$ in equation (17) and normalizing to get the probability for a Bernoulli sample.

We then sample λ_t from its conditional distribution

$$p(\lambda_t \mid i_t, x_t, y_t, \sigma^2) \propto p(y_t \mid x_t, \lambda_t, i_t, \sigma^2)p(\lambda_t)$$
$$= \mathcal{N}(y_t \mid x_t, (1+i_t\lambda_t)\sigma^2)p(\lambda_t)$$
$$\propto \begin{cases} p(\lambda_t), & i_t = 0 \\ \mathcal{IG}(1+\lambda_t;\alpha_p,\beta_p), & i_t = 1. \end{cases} \tag{20}$$

Both distributions in the last line of equation (20) are shifted inverse gamma distributions and can be sampled using a rejection sampling trick. First, we define a variable $l_t = 1 + \lambda_t$, then sample this from the inverse gamma distribution with the appropriate parameters. If it is less than 1, we reject it and resample, otherwise we accept it and subtract 1 to give a sample for λ_t.

Finally, since we condition on λ_t the derivation of the conditional distribution $p(z_t \mid i_t, \lambda_t, x_t, y_t, \sigma^2)$ is the same as that above, replacing λ with λ_t in equation (12).

Since all the full conditional distributions can be sampled from directly it is straightforward to update the state using a Gibbs sampling scheme.

3.1. Algorithm summary

The following gives a brief summary of the steps involved in removing impulse noise using the shifted inverse gamma model.

1. For each sample time t :

   (a) Sample i_t from $p(i_t \mid i_{-t}, x_t, y_t, \sigma^2)$ as given in equations (18) and (19)

   (b) Sample λ_t from $p(\lambda_t \mid i_t, x_t, y_t, \sigma^2)$, given in equation (20). Both distributions are shifted inverse Gamma distributions so can be sampled using the rejection sampling trick described above.

   (c) Sample z_t from $p(z_t \mid i_t, \lambda_t, x_t, y_t, \sigma^2)$, a normal distribution given in equation (12) with λ replaced with λ_t.

2. Perform Gabor analysis as in [1], using the sampled z_t variables (which should be the true signal perturbed by homogenous Gaussian noise) in place of the observed signal y_t.

3. Iterate to obtain a set of samples representing an estimate for the Gabor coefficients and the true signal x_t.

4. RESULTS

In order to evaluate the methods we compared them to each other and to the algorithm of [1], which they extend. To do so, we took clean audio signals and corrupted them with additive noise at a signal to noise ratio (SNR) of around 7dB. The tests used a mixture of artificially generated and real impulse noise. For the artifical impulsese, we first added homogenous Gaussian noise to a SNR of 15dB and then added impulse noise with either a constant ('Fixed λ') or variable ('Variable λ') impulse variance. The former used a scale factor of $\lambda = 100$, while the latter had a scale factor with an inverse gamma distribution with parameters $\alpha = 1$ and $\beta = 20$, giving roughly the same SNR. For the 'Real' impulse noise, we added a suitable multiple of noise from the run-in track from an old vinyl recording, high-pass filtered to remove low-frequency distortions.

Algorithm	Impulse Type	Noisy SNR	Final SNR
Original	Fixed λ	6.97	12.32
Original	Variable λ	6.96	9.65
Original	Real	6.61	8.83
Fixed λ	Fixed λ	6.97	13.83
Fixed λ	Variable λ	6.96	13.36
Fixed $\lambda = 15$	Real	6.61	11.54
Fixed $\lambda = 100$	Real	6.61	12.79
Variable λ	Fixed λ	6.97	13.36
Variable λ	Variable λ	6.96	13.47
Variable λ	Real	6.61	12.81

Table 1. SNRs (dB) before and after application of each algorithm

After running each noise reduction algorithm we looked at the SNRs of the restored signals, taken as the mean signal over the last 100 MCMC samples. All algorithms were run for 200 samples with the first 100 samples being considered burn-in (i.e. time for the chain to reach its stable distribution) and discarded. Parameter and impulse estimates were generally observed to converge within the 100 iteration burn-in time (see figure 2). The noisy signal reconstruction z_t was initialized to the observed signal y_t and the final signal reconstruction x_t was initialized to zero. When using the fixed variance
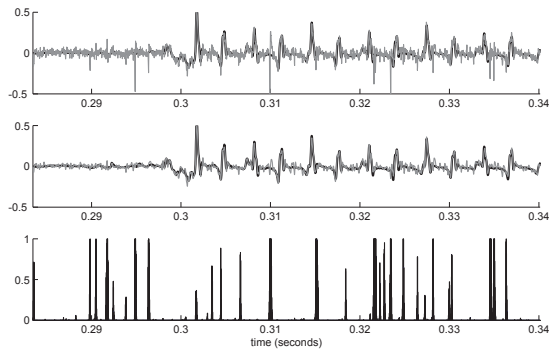
Fig. 1. A short excerpt showing the removal of impulses from a track using the variable impulse size algorithm. The top chart shows the noisy waveform (grey) superimposed on the clean signal (heavy black). The second chart shows the reconstructed waveform superimposed on the clean signal. The final chart shows the estimated posterior probability of an impulse being present at each location.

algorithm with variable variance impulses, the impulse variance parameter λ was set to the mean impulse variance. Figure 1 shows the impulse detection and removal results for a small section of audio using the variable impulse size algorithm.

The results in table 1 show that the original algorithm performs considerably worse than the new algorithms when impulses are present. As might be expected, the variable impulse method performs slightly worse for fixed impulses as the fixed impulse algorithm is correctly tuned to the impulse size. For variable size and real impulses the variable impulse algorithm performs best with no need for tuning (as opposed to the fixed impulse algorithm, where choosing different values of λ has a sizeable impact on performance).

It is difficult to make a full assessment of the results based only on the SNR and a more complete evaluation would include a psychoacoustical metric. Some distortion due to the filtering can be evident in badly corrupted examples and this can sometimes be ameliorated by increasing the amount of noise in the final reconstruction. This idea could be incorporated into these algorithms by introducing a further latent process 'between' x and z, which would be a reconstruction with some background noise deliberately left in to improve the perception of the reconstruction.

The examples here and others can be found at `http://www-sigproc.eng.cam.ac.uk/~jm362`.

5. CONCLUSIONS

In this work we have outlined a way to remove impulse noise from audio signals within a Bayesian context. We show how impulses with a wide range of magnitudes can be removed using an inverse gamma structure for their variance distribution. Results showed that this structure was successful in removing impulse noise from a range of samples, including real-life examples.

Possible extensions to the work include the the use of the shifted inverse gamma prior to the Gabor coefficients $c_{m,n}$, allowing the more efficient modelling of fatter tails for these and the implementation of a sequential version of the algorithm, allowing the removal of impulses to be performed on-line.
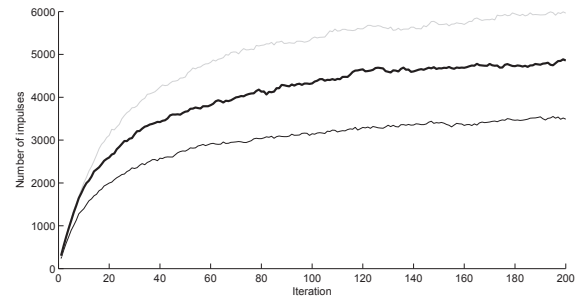
Fig. 2. Number of detected impulses with iteration number for real impulse example (87723 samples) with the fixed impulse size ($\lambda = 15$, grey; $\lambda = 100$, thin black) and variable impulse size (thick black) algorithms.

6. REFERENCES

[1] P. J. Wolfe, S. J. Godsill, and W.J. Ng, "Bayesian variable selection and regularisation for time-frequency surface estimation," *Journal of the Royal Statistical Society, Series B*, vol. 66, no. 3, pp. 575–589, 2004, Read paper (with discussion).

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[3] J.S. Erkelens and R. Heusdens, "Tracking of nonstationary noise based on data-driven recursive noise power estimation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 6, pp. 1112–1123, 2008.

[4] S.J. Godsill, "The shifted inverse-gamma model for noise floor estimation in archived audio recordings," *Applied Signal Processing*, 2010, Special Issue on Preservation of Ethological Recordings.

[5] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model-Based Approach*, Berlin: Springer, ISBN 3 540 76222 1, Sept. 1998.

[6] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 5, pp. 245–256, 2002.

[7] I.Y. Soon, S.N. Koh, and C.K. Yeo, "Noisy speech enhancement using discrete cosine transform," *Speech communication*, vol. 24, no. 3, pp. 249–257, 1998.

[8] P.J. Wolfe and S. J. Godsill, "Interpolation of missing data values for audio signal restoration using a Gabor regression model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2005.

[9] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[10] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[11] W.R. Gilks, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov chain Monte Carlo in practice*, Chapman & Hall/CRC, 1996.