BAYESIAN INTERPOLATION IN A DYNAMIC SINUSOIDAL MODEL WITH APPLICATION TO PACKET-LOSS CONCEALMENT

Jesper Kjær Nielsen[†], *Mads Græsbøll Christensen*[†], *Ali Taylan Cemgil*[††], *Simon J. Godsill*[†††], *and Søren Holdt Jensen*[†]

[†]Aalborg University
Department of Electronic Systems
Niels Jernes Vej 12, DK-9220 Aalborg
{jkn,mgc,shj,tl}@es.aau.dk

[††]Boğaziçi University
Department of Computer Engineering
34342 Bebek Istanbul, TR
taylan.cemgil@boun.edu.tr

[†††]University of Cambridge
Department of Engineering
Trumpington St., Cambridge, CB2 1PZ, UK
sjg@eng.cam.ac.uk

ABSTRACT

In this paper, we consider Bayesian interpolation and parameter estimation in a dynamic sinusoidal model. This model is more flexible than the static sinusoidal model since it enables the amplitudes and phases of the sinusoids to be time-varying. For the dynamic sinusoidal model, we derive a Bayesian inference scheme for the missing observations, hidden states and model parameters of the dynamic model. The inference scheme is based on a Markov chain Monte Carlo method known as Gibbs sampler. We illustrate the performance of the inference scheme to the application of packet-loss concealment of lost audio and speech packets.

1. INTRODUCTION

Interpolation of missing, corrupted and future samples in signal waveforms is an important task in several applications. For example, speech and audio signals are often transmitted over packet-based networks in which packets may be lost, delayed or corrupted. If the contents of neighbouring packets are correlated, the erroneous packets can be approximately reconstructed by using suitable interpolation techniques. The simplest interpolation techniques employ signal repetition [1] and signal stretching [2], whereas more advanced interpolation techniques are based on filter bank methods such as GAPES and MAPES [3], and signal modelling such as autoregressive models [4, 5], hidden Markov models [6], and sinusoidal models [7]. An integral part of the techniques based on signal modelling is the estimation of the signal parameters. Given estimates of these parameters, the interpolation task is simply a question of simulating data from the model. In this paper, we develop an interpolation and parameter estimation scheme by assuming a dynamic sinusoidal model for an observed signal segment. This model can be written as a linear Gaussian time-invariant state space model given by

$$\begin{aligned} y_n &= \boldsymbol{b}^T \boldsymbol{s}_n + w_n && \text{(observation equation)} \\ \boldsymbol{s}_{n+1} &= \boldsymbol{A}\boldsymbol{s}_n + \boldsymbol{v}_n && \text{(state equation)} \end{aligned} \quad (1)$$

where $n = 1, \ldots, N$ label the uniform sampled data in time, and

$$\boldsymbol{b} = [1 \quad 0 \quad \cdots \quad 1 \quad 0]^T \quad (2)$$

$$\boldsymbol{A} = \text{diag}(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_l, \cdots \boldsymbol{A}_L) \quad (3)$$

$$\boldsymbol{A}_l = \exp(-\gamma_l) \begin{bmatrix} \cos\omega_l & \sin\omega_l \\ -\sin\omega_l & \cos\omega_l \end{bmatrix} , \quad (4)$$

with $\omega_l \in [0, \pi]$ and $\gamma_l > 0$ denoting the (angular) frequency, and the log-damping coefficient of the l'th sinusoid, respectively. Further, $\boldsymbol{s}_n$ is the state vector, and $\boldsymbol{v}_n$ and w_n are white Gaussian state and observation noise sequences with covariance matrix $\boldsymbol{Q}$ and variance σ_w^2, respectively. We also assume a Gaussian prior for the initial state vector $\boldsymbol{s}_1$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{P}$. For a non-zero state covariance matrix, the dynamic sinusoidal model in (1) is able to model non-stationary tonal signals such as a wide range of speech and audio signal segments. We are here concerned with the problem of performing interpolation and parameter estimation in the model in (1) from a Bayesian perspective which offer some conceptual advantages to classical statistics (see, e.g., [8]). For example, the Bayesian approach offers a standardised way of dealing with nuisance parameters and signal interpolation [4]. The downside of using the Bayesian methods is that they struggle with practical problems such as evaluation of high-dimensional and intractable integrals. Although various developments in Markov chain Monte Carlo (MCMC) methods (see, e.g., [9]) in recent years have overcome these problems to a great extend, the methods still remain very computational intensive.

Within the field of econometrics, the dynamic sinusoidal model in (1) is well-known and referred to as the stochastic cyclical model [10]. Two slightly different stochastic cyclical models were given a fully Bayesian treatment using MCMC inference techniques in [11] and [12]. Neither of these, however, considered the case where some observations are missing. In the audio and speech processing field, the dynamic sinusoidal model has also been considered by Cemgil et al. in [13, 14, 15]. However, they considered the frequency parameter as a discrete random variable and based their inference on approximate variational Bayesian methods.

In this paper, we extend the above work by developing an inference scheme for the dynamic sinusoidal model based on MCMC inference techniques. We consider the frequency parameter as a continuous random variable and allow some of the observations to be missing. To achieve this, we develop a Gibbs sampling scheme. The output of this sampler can be used for forming histograms of the unknown parameters of interest. These histograms have the desirable property that they converge to the probability distribution of these unknown parameters when the number of generated samples is increased, and they therefore enable us to extract statistical features for the model parameters as well as for performing the interpolation of the missing observations. It should be noted that although this inference scheme can be used for estimating parameters of signals with no missing observations, the primary focus of this paper is on the application of reconstructing missing observations from signal segments which are assumed to have been generated by a dynamic sinusoidal model.

The paper is organised as follows. In Sec. 2, we formalise the problem by setting up the Bayesian framework. This enables us in Sec. 3 to develop the interpolation and inference scheme. In Sec. 4, we illustrate the performance of the inter-

The work of J.K. Nielsen was supported by the Oticon Foundation's Scholarship.

polating scheme by use of simulations, and Sec. 5 concludes this paper.

2. PROBLEM FORMULATION

In the Bayesian approach, all variables of the model in (1) are random variables, and we partition them as

$$\begin{aligned} &\text{Observations:} && \boldsymbol{y} = [y_1, y_2, \cdots, y_N]^T \\ &\text{Latent variables:} && \boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_N] \\ &\text{Model parameters:} && \boldsymbol{\theta} = \{\boldsymbol{\omega}, \boldsymbol{\gamma}, \boldsymbol{q}, \sigma_w^2\} \end{aligned}$$

where $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$ and $\boldsymbol{q}$ are L-dimensional vectors consisting of the L frequencies, the L log-damping parameters and the L state noise variances, respectively. The nth state vector $\boldsymbol{s}_n = \left[\boldsymbol{s}_{n,1}^T, \cdots, \boldsymbol{s}_{n,L}^T\right]^T$ consists of L two-dimensional state vectors pertaining to the L sinusoids. Conditioned on the previous state vector, each of these L two-dimensional state vectors has isotropic covariance matrix $q_l \boldsymbol{I}_2$, where $\boldsymbol{I}_2$ is the 2×2 identity matrix, so that $\boldsymbol{Q} = \text{diag}(\boldsymbol{q}) \otimes \boldsymbol{I}_2$ where $\otimes$ is the Kronecker product. We also assume that R of the elements in $\boldsymbol{y}$ are missing or corrupted, and that we know their indices $\mathcal{I} \subset \{1, \ldots, N\}$. Using this set of indices, we define the vectors $\boldsymbol{y}_\text{m} \triangleq \boldsymbol{y}_\mathcal{I}$ and $\boldsymbol{y}_\text{o} \triangleq \boldsymbol{y}_{\setminus\mathcal{I}}$ containing the R missing or corrupted observations and the $N - R$ valid observations, respectively. The notation $(\cdot)_{\setminus *}$ denotes 'without element $*$'.

The primary objective of this paper is to recover $\boldsymbol{y}_\text{m}$ from $\boldsymbol{y}_\text{o}$. This can be achieved in various ways, e.g., by using MAP/MMSE estimate w.r.t. the posterior distribution $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$ or by drawing a sample from $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$. The MAP-based interpolation produces the most probable interpolants. For audio and speech signals, however, MAP/MMSE-based interpolation tends to produce over-smoothed interpolants in the sense that they do not agree with the stochastic part of the valid observations [16]. A more typical interpolant can be obtained by drawing a single sample from $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$ [4]. The posterior distribution for the missing samples given the valid samples is given by

$$p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o}) = \int p(\boldsymbol{y}_\text{m}|\boldsymbol{S}_\mathcal{I}, \sigma_w^2) p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o}) d\boldsymbol{S} d\boldsymbol{\theta} \ . \tag{5}$$

We are not able to draw a sample directly from $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$ since we are not able to integrate the nuisance parameters $\boldsymbol{S}$ and $\boldsymbol{\theta}$ out analytically. However, we can obtain a sample from $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$ by taking a single sample from the joint posterior distribution $p(\boldsymbol{y}_\text{m}, \boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o})$ and simply ignore the generated values for $\boldsymbol{S}$ and $\boldsymbol{\theta}$. From the observation equation of (1), we know the distribution of $p(\boldsymbol{y}_\text{m}|\boldsymbol{S}_\mathcal{I}, \sigma_w^2)$, so the only problem left is computing $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o})$. This distribution is by Bayes' theorem given by

$$p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o}) = \frac{p(\boldsymbol{y}_\text{o}, \boldsymbol{S}_{\setminus 1}|\boldsymbol{s}_1, \boldsymbol{\theta}) p(\boldsymbol{s}_1, \boldsymbol{\theta})}{p(\boldsymbol{y}_\text{o})} \tag{6}$$

where $p(\boldsymbol{y}_\text{o}, \boldsymbol{S}_{\setminus 1}|\boldsymbol{s}_1, \boldsymbol{\theta})$, $p(\boldsymbol{s}_1, \boldsymbol{\theta})$ and $p(\boldsymbol{y}_\text{o})$ are referred to as the likelihood, the prior and the model evidence, respectively. Under the above assumption, the likelihood can be factored as

$$\begin{aligned} p(\boldsymbol{y}_\text{o}, \boldsymbol{S}_{\setminus 1}|\boldsymbol{s}_1, \boldsymbol{\theta}) =& p(\boldsymbol{y}_\text{o}|\boldsymbol{S}_{\setminus\mathcal{I}}, \sigma_w^2) \\ &\times \prod_{n=1}^{N-1} \prod_{l=1}^{L} p(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, q_l, \omega_l, \gamma_l) \end{aligned} \tag{7}$$

which from (1) is seen to be a product of normal distributions. For the prior distribution, we assume the factorisation

$$\begin{aligned} p(\boldsymbol{s}_1, \boldsymbol{\theta}) &= p(\boldsymbol{s}_1) p(\boldsymbol{\omega}) p(\boldsymbol{\gamma}) p(\boldsymbol{q}) p(\sigma_w^2) \\ &= p(\boldsymbol{s}_1) \left[\prod_{l=1}^{L} p(\omega_l) p(\gamma_l) p(q_l)\right] p(\sigma_w^2) \end{aligned} \tag{8}$$

where $p(\boldsymbol{s}_1)$ has a normal distribution $\mathcal{N}(\boldsymbol{s}_1; \boldsymbol{\mu}, \boldsymbol{P})$, $p(\omega_l)$ has a uniform distribution $\mathcal{U}(\omega_l; 0, \pi)$, $p(\gamma_l)$ has an exponential distribution $\text{Exp}(\gamma_l; \lambda_l)$, and $p(\sigma_w^2)$ and $p(q_l)$ have inverse gamma distributions $\mathcal{IG}(\sigma_w^2; \alpha_w, \beta_w)$ and $\mathcal{IG}(q_l; \alpha_{v,l}, \beta_{v,l})$. The model evidence $p(\boldsymbol{y}_\text{o})$ is independent of $\boldsymbol{S}$ and $\boldsymbol{\theta}$ and is therefore a mere scale factor which can be ignored in the inference stage.

3. INFERENCE SCHEME

In the Bayesian framework, all statistical inference is based on the posterior distribution over the unknown variables or a marginal posterior distribution over some of these. As derived in the previous section, we have to generate samples from $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o})$ in order to be able to do this. Unfortunately, this distribution has a very complicated form, and we are therefore not able to sample directly from it. We therefore have to resort to other sampling techniques in order to enable statistical inference based on this distribution. One of the simplest and most popular numerical sampling techniques is the Gibbs sampler [17] which is an MCMC-based algorithm and suitable for this task. The Gibbs sampler draws samples from a multivariate distribution, say $p(\boldsymbol{x}) = p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)$, by breaking it into a number of conditional distributions $p(\boldsymbol{x}_k|\boldsymbol{x}_{\setminus k})$ of smaller dimensionality from which samples are obtained in an alternating pattern. Specifically, for the τ's iteration, we sample for $k = 1, \ldots, K$ from

$$\boldsymbol{x}_k^{[\tau+1]} \sim p(\boldsymbol{x}_k|\boldsymbol{x}_1^{[\tau+1]}, \ldots, \boldsymbol{x}_{k-1}^{[\tau+1]}, \boldsymbol{x}_{k+1}^{[\tau]}, \ldots, \boldsymbol{x}_K^{[\tau]}) \ . \tag{9}$$

After an initial burn-in time during which the sampling scheme converges, the samples obtained from sampling these lower dimensional conditional distributions can be regarded as samples from the joint posterior distribution. In this paper, the posterior distribution $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\text{o})$ is broken into the two conditional distributions given by

$$\text{States:} \qquad p(\boldsymbol{S}|\boldsymbol{\theta}, \boldsymbol{y}_\text{o}) \tag{10}$$

$$\text{Model parameters:} \qquad p(\boldsymbol{\theta}|\boldsymbol{S}, \boldsymbol{y}_\text{o}) \tag{11}$$

The selected grouping of variables in (10) and (11) leads to a set of conditional distributions which are fairly easy to sample from. In the next sections, we derive the particular form of these conditional distributions.

3.1 States

The conditional state distribution in (10) can be shown to be a multivariate Gaussian distribution. However, the dimension of this distribution is $2LN \times 1$ which would render direct sampling from it infeasible for most applications. Instead, we use the simulation smoother [18], which is an efficient sampling scheme using standard Kalman smoothing, for drawing samples from (10). Since some of the observations are missing, we have to modify the simulation smoother slightly. This is easily done by skipping the update step of the build-in Kalman filter for these observations.

3.2 Model Parameters

Since the model parameter of the observation equation, σ_w^2, and the L sets of model parameters of the state equation, $(\omega_l, \gamma_l, q_l)$, are mutually independent conditioned on the states $\boldsymbol{S}$, we can factor (11) as

$$p(\boldsymbol{\theta}|\boldsymbol{S},\boldsymbol{y}_\text{o}) = \left[\prod_{l=1}^{L} p(\omega_l,\gamma_l,q_l|\boldsymbol{S})\right] p(\sigma_w^2|\boldsymbol{S},\boldsymbol{y}_\text{o}) \ . \qquad (12)$$

Thus, sampling from the conditional distribution in (11) can be done by sampling the $L+1$ conditional distributions on the right side of (12) independently.

3.2.1 Frequency, Log-damping and State Noise Variance

To our knowledge, it is not possible to sample directly from the conditional distribution $p(\omega_l,\gamma_l,q_l|\boldsymbol{S})$. A Gibbs sampling scheme is also not straight-forward since it suffers from poor mixing and since the l'th log-damping coefficient conditioned on the l'th frequency parameter and state noise variance has a non-standard distribution. In order to improve mixing of the parameters and lower the overall computational complexity, we therefore propose sampling from $p(\omega_l,\gamma_l,q_l|\boldsymbol{S})$ by use of a Metropolis-Hastings (MH) sampler [19]. In the MH sampler, samples generated from the desired posterior distribution, say $p(\boldsymbol{x})$, which we know up to some normalising constant Z with $p(\boldsymbol{x}) = \tilde{p}(\boldsymbol{x})/Z$, are generated by use of a user-defined proposal distribution $q(\boldsymbol{x}|\boldsymbol{x}^{[\tau]})$, where $\boldsymbol{x}^{[\tau]}$ is the τth generated sample. In general, $p(\boldsymbol{x}) \neq q(\boldsymbol{x}|\boldsymbol{x}^{[\tau]})$ so a proposed sample $\boldsymbol{x}' \sim q(\boldsymbol{x}|\boldsymbol{x}^{[\tau]})$ is only accepted as a sample from $p(\boldsymbol{x})$ with probability

$$\alpha(\boldsymbol{x}^{[\tau]},\boldsymbol{x}') = \min\left[1, \frac{\tilde{p}(\boldsymbol{x}')q(\boldsymbol{x}^{[\tau]}|\boldsymbol{x}')}{\tilde{p}(\boldsymbol{x}^{[\tau]})q(\boldsymbol{x}'|\boldsymbol{x}^{[\tau]})}\right] \ . \qquad (13)$$

Otherwise, the previous accepted sample is retained, i.e., $\boldsymbol{x}^{[\tau+1]} = \boldsymbol{x}^{[\tau]}$.

For $p(\omega_l,\gamma_l,q_l|\boldsymbol{S})$, the proposal samples $(\omega_l',\gamma_l',q_l')$ are generated in two simple steps: First, we generate a sample for the mean and variance of a bivariate normal-scaled inverse gamma distribution with isotropic covariance matrix. This is done by sampling from

$$q_l' \sim \mathcal{IG}(\alpha_{q_l},\beta_{q_l}) \qquad (14)$$

$$\tau_l' \sim \mathcal{IG}(\alpha_{q_l},1/2) \qquad (15)$$

$$\boldsymbol{a}_l' = \begin{bmatrix} a_{1,l}' & a_{2,l}' \end{bmatrix}^T \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{a},l}, 2\beta_{q_l}\tau_l'\sigma_{\boldsymbol{a},l}^2\boldsymbol{I}_2) \qquad (16)$$

where we have defined

$$\boldsymbol{\varphi}_l \triangleq \begin{bmatrix} \boldsymbol{s}_{2,l}^T & \boldsymbol{s}_{3,l}^T & \cdots & \boldsymbol{s}_{N,l}^T \end{bmatrix}^T \qquad (17)$$

$$\boldsymbol{\phi}_l \triangleq \begin{bmatrix} \boldsymbol{s}_{1,l}^T & \boldsymbol{s}_{2,l}^T & \cdots & \boldsymbol{s}_{N-1,l}^T \end{bmatrix}^T \qquad (18)$$

$$\tilde{\boldsymbol{\phi}}_l \triangleq \begin{bmatrix} (\boldsymbol{s}_{1,l}^\perp)^T & (\boldsymbol{s}_{2,l}^\perp)^T & \cdots & (\boldsymbol{s}_{N-1,l}^\perp)^T \end{bmatrix}^T \qquad (19)$$

$$\boldsymbol{\Phi}_l \triangleq \begin{bmatrix} \boldsymbol{\phi}_l & \tilde{\boldsymbol{\phi}}_l \end{bmatrix} \qquad (20)$$

$$\sigma_{\boldsymbol{a},l}^2 \triangleq (\boldsymbol{\phi}_l^T\boldsymbol{\phi}_l)^{-1} \qquad (21)$$

$$\boldsymbol{\mu}_{\boldsymbol{a},l} \triangleq \sigma_{\boldsymbol{a},l}^2\boldsymbol{\Phi}_l^T\boldsymbol{\varphi}_l \qquad (22)$$

$$\alpha_{q_l} \triangleq \alpha_{v,l} + N - 1 \qquad (23)$$

$$\beta_{q_l} \triangleq \beta_{v,l} + (\boldsymbol{\varphi}_l^T\boldsymbol{\varphi}_l - \sigma_{\boldsymbol{a},l}^{-2}\boldsymbol{\mu}_{\boldsymbol{a},l}^T\boldsymbol{\mu}_{\boldsymbol{a},l})/2 \ , \qquad (24)$$

and $\boldsymbol{s}_{n,l}^\perp$ is obtained by a 90° clockwise rotation of $\boldsymbol{s}_{n,l}$. Second, we transform $\boldsymbol{a}_l'$ into (ω_l',γ_l') by the relations

$$\omega_l' = \arctan(a_{2,l}'/a_{1,l}') \qquad (25)$$

$$\gamma_l' = -\ln\left(\boldsymbol{a}_l'^T\boldsymbol{a}_l'\right)/2 \ . \qquad (26)$$

1. Select hyperparameters and initialise the Gibbs sampler.
2. Repeat for $k = 0,1,2,\ldots,K$
   - (a) $\boldsymbol{S}^{[k+1]} \sim p(\boldsymbol{S}|\boldsymbol{\theta}^{[k]},\boldsymbol{y}_\text{o})$ (simulation smoother)
   - (b) Repeat for $l = 1,2,\ldots,L$
     - i. $q_l' \sim \mathcal{IG}(\alpha_{q_l}^{[\tau]},\beta_{q_l}^{[\tau]})$
     - ii. $\tau_l' \sim \mathcal{IG}(\alpha_{q_l}^{[\tau]},1/2)$
     - iii. $\boldsymbol{a}_l' \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{a},l}^{[\tau]}, 2\beta_{q_l}^{[\tau]}\tau_l'\sigma_{\boldsymbol{a},l}^{2\ [\tau]}\boldsymbol{I}_2)$
     - iv. $\omega_l' = \arctan(a_{2,l}'/a_{1,l}')$
     - v. $\gamma_l' = -\ln\left(\boldsymbol{a}_l'^T\boldsymbol{a}_l'\right)/2$
     - vi. $u_l = \mathcal{U}(0,1)$
     - vii. if $u_l \leq \alpha\big((\omega_l^{[\tau]},\gamma_l^{[\tau]},q_l^{[\tau]}),(\omega_l',\gamma_l',q_l')\big)$
       - $(\omega_l^{[k+1]},\gamma_l^{[k+1]},q_l^{[k+1]}) = (\omega_l',\gamma_l',q_l')$

       else
       - $(\omega_l^{[k+1]},\gamma_l^{[k+1]},q_l^{[k+1]}) = (\omega_l^{[\tau]},\gamma_l^{[\tau]},q_l^{[\tau]})$
   - (c) $\sigma_w^{2\ [k+1]} \sim \mathcal{IG}(\alpha_{\sigma_w^2}^{[\tau]},\beta_{\sigma_w^2}^{[\tau]})$

Table 1: Summary of proposed Gibbs sampler for generating samples from $p(\boldsymbol{S},\boldsymbol{\theta}|\boldsymbol{y}_\text{o})$.

Then, if $a_{2,l}' \geq 0$, the proposal samples $(\omega_l',\gamma_l',q_l')$ are accepted as samples from $p(\omega_l,\gamma_l,q_l|\boldsymbol{S})$ with probability

$$\begin{aligned}\alpha\big((\omega_l^{[\tau]},&\gamma_l^{[\tau]},q_l^{[\tau]}),(\omega_l',\gamma_l',q_l')\big)\\ &= \min\left[1,\exp\left\{(\lambda_l-2)(\gamma_l^{[\tau]}-\gamma_l')\right\}\right] \ . \end{aligned} \qquad (27)$$

Otherwise, the previous values $(\omega_l^{[\tau]},\gamma_l^{[\tau]},q_l^{[\tau]})$ are retained. Notice that if the rate parameter λ_l of the prior for γ_l is equal to two, $\alpha = 1$ for any γ_l'. The details of the derivation of this sampling scheme can be found in [20].

3.2.2 Observation Noise Variance

By Bayes' theorem, we can write $p(\sigma_w^2|\boldsymbol{S},\boldsymbol{y}_\text{o})$ as

$$p(\sigma_w^2|\boldsymbol{S},\boldsymbol{y}_\text{o}) \propto p(\boldsymbol{y}_\text{o}|\boldsymbol{S}_{\setminus\mathcal{I}},\sigma_w^2)p(\sigma_w^2) \qquad (28)$$

where $p(\boldsymbol{y}_\text{o}|\boldsymbol{S}_{\setminus\mathcal{I}},\sigma_w^2)$ is the likelihood of the observation equation in (1) and $p(\sigma_w^2)$ is the prior distribution for σ_w^2. Since $p(\boldsymbol{y}_\text{o}|\boldsymbol{S}_{\setminus\mathcal{I}},\sigma_w^2) = \mathcal{N}(\boldsymbol{S}_{\setminus\mathcal{I}}^T\boldsymbol{b},\sigma_w^2\boldsymbol{I}_{N-R})$ and $p(\sigma_w^2) = \mathcal{IG}(\alpha_w,\beta_w)$, the posterior distribution $p(\sigma_w^2|\boldsymbol{S}_{\setminus\mathcal{I}},\boldsymbol{y}_\text{o})$ is an inverse gamma distribution, $\mathcal{IG}(\sigma_w^2;\alpha_{\sigma_w^2},\beta_{\sigma_w^2})$, with parameters

$$\alpha_{\sigma_w^2} = \alpha_w + N/2 \qquad (29)$$

$$\beta_{\sigma_w^2} = \beta_w + \frac{1}{2}(\boldsymbol{y}_\text{o} - \boldsymbol{S}_{\setminus\mathcal{I}}^T\boldsymbol{b})^T(\boldsymbol{y}_\text{o} - \boldsymbol{S}_{\setminus\mathcal{I}}^T\boldsymbol{b}) \ . \qquad (30)$$

3.3 Summary of Inference Scheme

Table 1 summarises our proposed Gibbs sampler for generating samples from $p(\boldsymbol{S},\boldsymbol{\theta}|\boldsymbol{y}_\text{o})$. The computational complexity of the algorithm is fairly high primarily due to the generation of the states by the simulation smoother. In our implementation with $N = 600$ observations and $L = 6$ sinusoids, it takes approximately 40 ms for generating a state sample $\boldsymbol{S}^{[\tau]}$. This corresponds to nearly 97 % of the time consumption of one iteration of the Gibbs sampler. For the application of interpolation, we only need a single sample for the states and model parameters from the invariant distribution of the underlying Markov chain of the sampler. Once these have been generated, we may perform the interpolation by simulating from the observation equation of (1). Therefore, the computational complexity of the algorithm heavily depends on proper initialisation and the convergence speed of the chain.

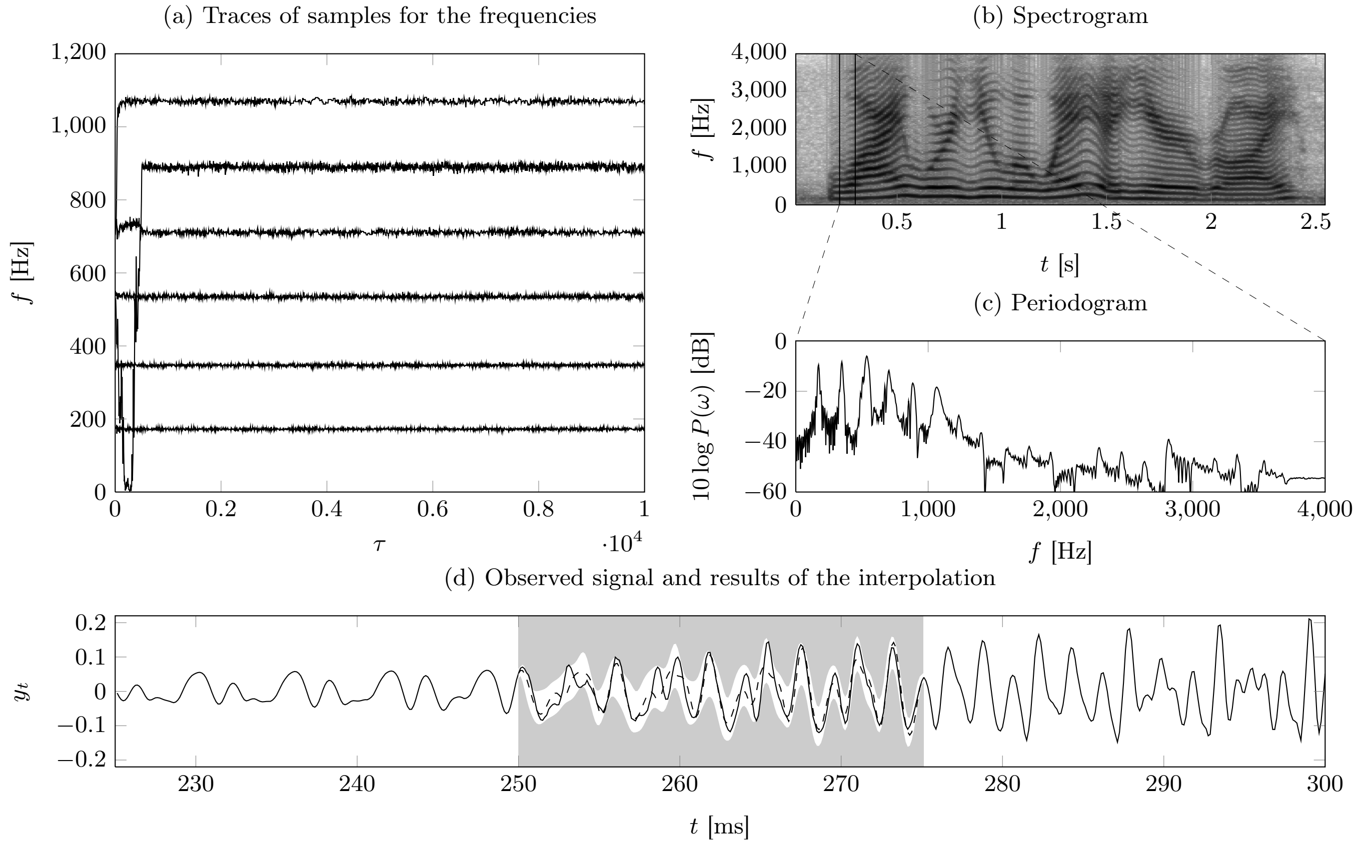

Figure 1: Plot (a) shows the six traces for the frequencies each consisting of 10,000 samples. Plot (b) shows the spectrogram for the complete speech signal whereas plot (c) shows the periodogram for the section indicated in plot (b). The time series corresponding to this section is shown in plot (d) with the middle section of 25 ms audio missing. The plot also shows the result of the interpolation in terms of the 95 % credible interval, a sample from the marginal posterior distribution $p(\boldsymbol{y}_\mathrm{m}|\boldsymbol{y}_\mathrm{o})$ and the true missing observations (dashed).

4. SIMULATIONS

We consider the problem of reconstructing missing or corrupted packets on a packet-based network. First, we illustrate the reconstruction process and, second, we present the results of a small-scale listening test.

4.1 Speech Signal Reconstruction

We used a snapshot from a speech signal (see Fig. 1.d) consisting of $N = 600$ samples corresponding to 75 ms of speech at a sampling frequency of 8000 kHz. The speech signal is generated by a female voice uttering, *"Why were you away a year, Roy?"* and its spectrogram in shown in Fig. 1.b. The periodogram of the 75 ms speech signal segment is shown in Fig. 1.c. Prior to running the Gibbs sampler, we removed the middle section thus emulating a lost audio packet of 25 ms. For the setup of the Gibbs sampler, we assumed $L = 6$ sinusoidal components, and we selected the hyperparameters such that the prior distributions were diffuse. The initial values for the frequency and the observation noise variance were computed by using a matching pursuit algorithm. The initial values for the log-damping coefficients and the state noise covariances were somewhat heuristically set to 0 and $\sigma_w^{2\,[0]}/10$, respectively. Fig. 1 shows the main results of the simulation. Fig. 1.a shows the six traces of samples obtained for the frequency parameters. After a burn-in length of approximately 1000 samples the underlying Markov chain seems to have converged to the true posterior distribution for the frequencies. Inference for the frequency parameters can thus be based on histograms formed by the the last approximately 9000 samples. In a similar way, histograms for the remaining model parameters can be formed. Fig. 1.d shows a typical sample obtained for the missing observations compared to the true signal. Notice, that unlike maximum likelihood- and EM-restoration techniques, the noise is also modelled when performing the interpolation in the Bayesian framework. Fig. 1.d also shows an estimate of the 95 % credible interval for the missing observations.

4.2 Listening Test

We conducted a small-scale MUSHRA listening test [21, 22] to evaluate the performance of the interpolation scheme. In addition to the speech signal, we also used an excerpt from a trumpet signal. Both of these signals were partitioned in 25 ms packets and transmitted through four artificial channels where packets were lost independently with probabilities of 5 %, 10 %, 20 % and 30 %, respectively. On the receiver side, we applied our proposed interpolation scheme to the missing packets. For every gap of one or more consecutive missing packets, we used the valid packet before and after the gap as in Fig. 1. We compared the interpolant (A) from $p(\boldsymbol{y}_\mathrm{m}|\boldsymbol{y}_\mathrm{o})$ against the MMSE interpolant (B) $E\{\boldsymbol{y}_\mathrm{m}|\boldsymbol{y}_\mathrm{o}\}$ and the interpolant (C) from $p(\boldsymbol{y}_\mathrm{m}|\boldsymbol{y}_\mathrm{o}, \boldsymbol{\theta}^\mathrm{MAP})$. The MMSE and MAP estimates were computed from the last 9000 generated samples from the Gibbs sampler. For the anchor signal, we used zeros for the interpolation. Fig. 2 shows the results ob-

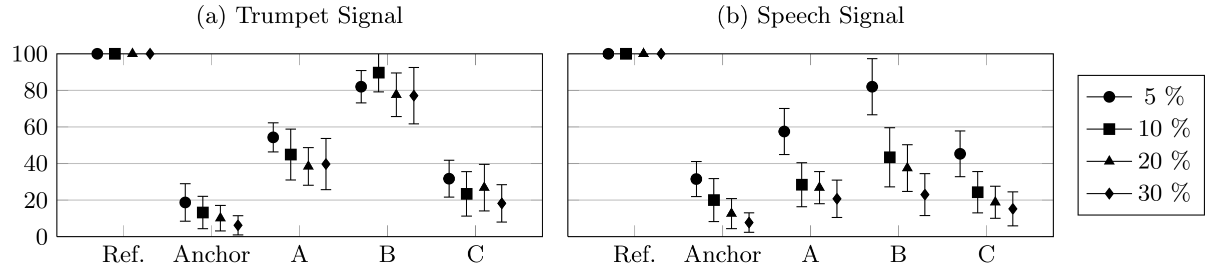

Figure 2: Mean and 95 % confidence intervals for the MUSHRA listening test. The reference signal was transmitted through four artificial channels with independent packet-loss probabilities of 5 %, 10 %, 20 % and 30 %, respectively. For the anchor signal, the missing packets were interpolated with zeros while the interpolation for A, B and C were based on $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o})$, $E\{\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o}\}$ and $p(\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o}, \boldsymbol{\theta}^\text{MAP})$, respectively.

tained by applying the statistical analysis suggested in [21] to the scores given by ten listeners. The listening test clearly revealed that reconstructing missing packets of the highly tonal and fairly stationary trumpet signal was much more successful than for the speech signal. The results also revealed that the interpolation based on $E\{\boldsymbol{y}_\text{m}|\boldsymbol{y}_\text{o}\}$ performed better than the other methods.

5. CONCLUSION

Based on a Gibbs sampler, we have presented a Bayesian inference scheme for the missing observations, the states and the model parameters of a dynamic sinusoidal model. This model is able to model some non-stationary signal segments which are often encountered in music or speech signal processing. In the simulations, we demonstrated that the algorithm can be used for interpolation of audio and speech signals. This is an integral part of many signal processing applications such as packet-loss concealment, pitch- and time-scale modification. Additionally, the inference scheme can also be used for making inference about the unknown model parameters of the dynamic sinusoidal model.

REFERENCES

[1] D. Goodman, G. Lockhart, O. Wasem, and Wai-Choong Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 6, pp. 1440–1448, Dec. 1986.

[2] Y.J. Liang, N. Farber, and B. Girod, "Adaptive playout scheduling and loss concealment for voice communication over IP networks," *IEEE Trans. Multimedia*, vol. 5, no. 4, pp. 532–543, Dec. 2003.

[3] H. Ofir and D. Malah, "Packet loss concealment for audio streaming based on the GAPES and MAPES algorithms," in *IEEE Conv. Electrical and Electronics Engineers in Isreal*, Nov. 2006, pp. 280–284.

[4] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*, Springer-Verlag, London, 1998.

[5] Mathieu Lagrange, Sylvain Marchand, and Jean-Bernard Rault, "Enhancing the tracking of partials for the sinusoidal modeling of polyphonic sounds," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1625–1634, Jul. 2007.

[6] C. A. Rødbro, M. N. Murthi, S. V. Andersen, and S. H. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1609–1623, Sep. 2006.

[7] J. Lindblom and P. Hedelin, "Packet loss concealment based on sinusoidal modeling," in *IEEE Proc. Workshop on Speech Coding*, Oct. 2002, pp. 65–67.

[8] J. M. Bernardo and Adrian Smith, *Bayesian Theory*, John Wiley and Sons Ltd, 1. edition, 1994.

[9] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 edition, 2005.

[10] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1989.

[11] A.C. Harvey, T.M. Trimbur, and H.K. van Dijk, "Cyclical components in economic time series," Econometric Institute Report EI 2002-20, Erasmus University Rotterdam, Econometric Institute, Nov. 2002.

[12] R. Kleijn and H. K. van Dijk, "Bayes model averaging of cyclical decompositions in economic time series," *Journal of Applied Econometrics*, vol. 21, no. 2, pp. 191–212, Mar. 2006.

[13] A. T. Cemgil and S. J. Godsill, "Probabilistic phase vocoder and its application to interpolation of missing values in audio signals," in *Proc. European Signal Processing Conf.*, 2005.

[14] A. T. Cemgil and S. J. Godsill, "Efficient variational inference for the dynamic harmonic model," in *Proc. IEEE Workshop on Appl. of Signal Process. to Aud. and Acoust.*, Oct. 2005, pp. 271–274.

[15] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 679–694, 2006.

[16] J.J.K. O Ruanaidh and W.J. Fitzgerald, "Interpolation of missing samples for audio restoration," *Electronics Letters*, vol. 30, no. 8, pp. 622–623, Apr. 1994.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, November 1984.

[18] J. Durbin and S. J. Koopman, "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, vol. 89, no. 3, pp. 603–615, 2002.

[19] Siddhartha Chib and Edward Greenberg, "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[20] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, S. J. Godsill, and S. H. Jensen, "Bayesian interpolation and parameter estimation in a dynamic sinusoidal model," *IEEE Trans. Audio, Speech, Lang. Process.*, 2010, Submitted in February 2010.

[21] "Method for the subjective assessment of intermediate quality levels of coding systems," 2003, ITU BS.1534-1.

[22] E. Vincent, *MUSHRAM: A MATLAB interface for MUSHRA listening tests*, 2005.