PROBABILISTIC PHASE VOCODER AND ITS APPLICATION TO INTERPOLATION OF MISSING VALUES IN AUDIO SIGNALS

Ali Taylan Cemgil and Simon J. Godsill

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ, UK
{atc27, sjg}@cam.ac.uk

ABSTRACT

We formulate the phase vocoder – an audio synthesis method very closely related to inverse short time Fourier Transform synthesis – as a Gaussian state space model and demonstrate simulation results on interpolation of missing values. The audio signal is modelled as a superposition of quasi-sinusoidal signals generated by a linear dynamical system. The advantage of our "generative" perspective is that it allows a full Bayesian treatment of the problem; e.g. one can perform the analysis while arbitrary chunks of sample values are missing or model parameters are unknown. To perform audio restoration, we derive an expectation-maximisation (EM) algorithm that infers the expectations of missing samples and maximum a-posteriori model parameters. We demonstrate the validity of our approach on a set of challenging real audio examples and compare to existing methods.

1. INTRODUCTION

Entire blocks of audio data can get lost during transmission over a noisy channel or during storage on physically degrading media such as magnetic tapes. In contrast to denoising applications, in this scenario the actual observed value of a corrupted sample is assumed to be completely independent from the original value, hence we assume that the corrupted samples are missing. The missing value interpolation problem is restoration of such distorted audio material when the indices of corrupted samples are known [1, 2, 3]. Let us denote the "clean" samples by $x_{\boldsymbol{\kappa}}$ and missing samples by $x_{\neg\boldsymbol{\kappa}}$. The missing value interpolation problem can be stated as the following generic Bayesian inference problem:

$$p(x_{\neg\boldsymbol{\kappa}}|x_{\boldsymbol{\kappa}}) \quad \propto \quad \int d\mathcal{H} p(x_{\neg\boldsymbol{\kappa}}|\mathcal{H})p(x_{\boldsymbol{\kappa}}|\mathcal{H})p(\mathcal{H}) \qquad (1)$$

Here, $\mathcal{H} = (\Theta, \mathcal{S})$ denotes the collection of unknown model parameters Θ and other unobserved latent state variables $\mathcal{S}$ that describe the sound generation mechanism $p(x_{0:K-1}, \mathcal{H})$ where $x_{0:K-1} = x_{\boldsymbol{\kappa}} \cup x_{\neg\boldsymbol{\kappa}}$. The model structure in (1) is somewhat restricted since it assumes that both $x_{\neg\boldsymbol{\kappa}}$ and $x_{\boldsymbol{\kappa}}$ share the same hidden cause and are independent otherwise. On the other hand, for most audio signals such as music or speech, it is realistic to assume that the same dynamical physical mechanism governs generation of both missing and observed samples.

In this paper, we will first describe a generic probabilistic generative model $p(x|\mathcal{H})p(\mathcal{H})$ for audio signals. Our model is closely related to the phase vocoder of [4]. Subsequently, we will describe inference methods and present results on real audio examples.

2. THE PHASE VOCODER

The phase vocoder (PVOC) is a well known tool for time scaling and pitch shifting of speech and music via modification of original short-time Fourier transform (STFT) coefficients. Interestingly, the first formulation of PVOC [4] was aimed primarily at a different application: low bit rate speech coding – hence the name "voice coder". The algorithm exploits the fact that most natural audio signals contain resonances that can be described by simple sinusoidal oscillations. The time varying amplitudes and phase shifts of these resonances are estimated via the STFT. For coding, the amplitudes and phases are quantised and sent over a channel to the decoder. Alternatively, to synthesise a time stretched version of the original [5], new STFT coefficients are created that have the same amplitudes as the original transform but with adjusted phases such that individual oscillations in each frequency band "last for more cycles". Several improvements and extensions were proposed over last decades to cope with some perceptual artifacts, e.g. see [6].

In our view, one conceptual problem with the original formulation of PVOC and later extensions is that they describe merely an encoding or time stretching algorithm and do not explicitly refer to the underlying signal model. It is fairly clear that the performance of the PVOC algorithm hinges critically upon the fact that the characteristics of the analysed signal are well matched by a sinusoidal model. In our view, explicitly stating the signal model can help in fixing the problems in a principled way, and, as we wish to demonstrate in this paper, to extend the basic model to scenarios where it is not at all obvious how to apply the original algorithm.

In the next section, we will describe the inverse discrete Fourier transform (DFT) as a dynamic generative process. This starting point will be useful later to describe the probabilistic phase vocoder model: an adaptive model directly applicable to spectral analysis and restoration of nonstationary signals.

2.1 (Inverse) Discrete Fourier Transform as a [illegible]

Consider a sequence $\mathbf{x} \equiv (x_0, x_1, \dots, x_k, \dots x_{K-1})^T$ with time index $k = 0, \dots, K-1$. Here, x_k are complex numbers and T denotes non-conjugate transpose. The Fourier transform $\mathbf{s} \equiv (s^0, s^1, \dots, s^\nu, \dots s^{W-1})^T$ with the frequency index $\nu = 0, \dots, W-1$ is given by

$$\mathbf{s} \quad = \quad F\mathbf{x}$$

where $F = \{F_\nu^k\}$ is the DFT matrix defined by entries[1]

$$F_\nu^k \quad = \quad e^{-2\pi j \nu k/K}$$

This mapping between "time domain" and "frequency domain" is invertible when the transform matrix is square, i.e.

A. T. Cemgil acknowledges European Commission funded Research Training Networks MUSCLE for support of this work.

[1] We omit irrelevant scaling factors.

$W = K$. In this case, one can compute the inverse DFT to reconstruct the original signal

$$\mathbf{x} \quad = \quad F^H \mathbf{s} \tag{2}$$

Here, H denotes the Hermitian transpose, $*$ denotes complex conjugation and the entries of the inverse DFT matrix $F^H = \{F_k^{*\nu}\}$ are given as

$$F_k^{*\nu} \quad = \quad e^{2\pi j \nu k / W}$$

Due to the special structure of the Fourier basis, the matrix entries can be defined recursively, i.e. k'th row of F^H can be "generated" from $k-1$'th row by a linear transformation:

$$F_k^{*\nu} \quad = \quad e^{j\omega\nu} F_{k-1}^{*\nu} \tag{3}$$

where $\omega \equiv 2\pi/W$ and $F_0^{*\nu} = 1$. The k'th sample of $\mathbf{x}$ defined in (2) is

$$x_k \quad = \quad \sum_\nu s_k^\nu \tag{4}$$

where $s_k^\nu \equiv F_k^{*\nu} s^\nu$. Substituting this definition to (3), by representing a complex number as a 2-D vector and multiplication by a complex number with unit magnitude as a rotation, we get (by a slight abuse of notation)

$$s_k^\nu \quad = \quad B(\omega\nu) s_{k-1}^\nu \tag{5}$$

where $s_k^\nu \equiv \begin{pmatrix} s_{\Re,k}^\nu & s_{\Im,k}^\nu \end{pmatrix}^T$ with $s_\Re \equiv \mathrm{Re}\{s\}$, $s_\Im \equiv \mathrm{Im}\{s\}$. Here, $B(\theta)$ is a Givens rotation matrix defined as

$$B(\theta) \quad = \quad \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

In this vector notation, we can get rid of the frequency index ν in (3) by defining a $2W \times 1$ state vector

$$\mathbf{s}_k \quad = \quad \left(s_{\Re,k}^0, s_{\Im,k}^0, \dots, s_{\Re,k}^\nu, s_{\Im,k}^\nu, \dots, s_{\Re,k}^{W-1}, s_{\Im,k}^{W-1} \right)^T$$

and rewriting (4) as

$$x_k \quad = \quad \begin{pmatrix} x_{\Re,k} \\ x_{\Im,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \end{pmatrix} \mathbf{s}_k$$

When $\mathbf{x}$ is real, which is the case for single channel audio signals, we have redundancy in transform coefficients due to the conjugacy relationship $s^\nu = (s^{W-\nu})^*$. In this case (assuming W is even) we redefine a $W \times 1$ state vector as

$$\mathbf{s}_k = \left(s_{\Re,k}^0, s_{\Re,k}^1, s_{\Im,k}^1, \dots, s_{\Re,k}^\nu, s_{\Im,k}^\nu, \dots, s_{\Re,k}^{W/2-1}, s_{\Im,k}^{W/2-1}, s_{\Re,k}^{W/2} \right)^T$$

Hence we can rewrite (5) and (4) respectively as

$$\mathbf{s}_k \quad = \quad A\mathbf{s}_{k-1} \tag{6}$$

$$x_k \quad = \quad x_{\Re,k} = C\mathbf{s}_k \tag{7}$$

$$A \quad \equiv \quad \mathrm{blkdiag}\{B(0), B(\omega), \dots, B(\nu\omega), \dots, B(\frac{W}{2}\omega)\} \tag{8}$$

$$C \quad \equiv \quad \begin{pmatrix} 1 & 2 & 0 & 2 & 0 & \cdots & 2 & 0 & 1 \end{pmatrix} \tag{9}$$

Here, $\mathrm{blkdiag}(B_1, B_2, \dots)$ denotes a block diagonal matrix with blocks $B_1, B_2, \dots$. [2]

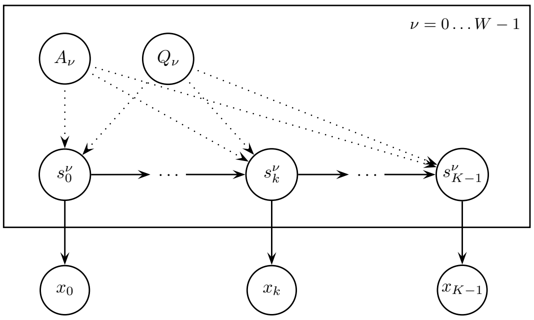

Figure 1: The probabilistic phase vocoder as a graphical model. The rectangle denotes a plate, W copies of the nodes inside. The directed links between the time slices are parametrised by the transition model given in (10). The observation model is given by (11) and the first time slice is drawn according to (12). The links from parameters A and Q are drawn dotted to maintain clarity. For fixed parameters, this model topology is a factorial Kalman filter model. The inverse DFT corresponds to the special case when $W = K$, $R = 0$, $Q = 0$ and A is defined as in (8) ,*i.e.*, all links are deterministic. The initial conditions are given by the Fourier transform coefficients $s_0^\nu = s^\nu$. The observation model is given by (4) and the links between the time slices correspond to (5).

2. 2 Probabilistic Phase Vocoder

The probabilistic phase vocoder (PPVOC) model is a stochastic version of the deterministic transition and observation equations in (6) and (7) respectively

$$\mathbf{s}_k | \mathbf{s}_{k-1} \quad \sim \quad \mathcal{N}(\mathbf{s}_k; A\mathbf{s}_{k-1}, Q) \tag{10}$$

$$x_k | \mathbf{s}_k \quad \sim \quad \mathcal{N}(x_k, C\mathbf{s}_k, R) \tag{11}$$

$$\mathbf{s}_0 \quad \sim \quad \mathcal{N}(\mathbf{s}_0; 0, P) \tag{12}$$

here, $\mathcal{N}(x; \mu, \Sigma)$ denotes a Gaussian distribution on index set x with mean μ and covariance matrix Σ. Similar formulations are used previously to model pitched music instruments[7] or in the econometrics literature to model seasonal fluctuations [8]. To treat the parameter estimation in a Bayesian framework, we assume the following conjugate prior distributions on the parameters

$$\mathrm{vec}\, A \quad \sim \quad \mathcal{N}(\mathrm{vec}\, A; \mathrm{vec}\, \Omega_A, \Sigma_A) \tag{13}$$

$$\mathrm{diag}\, Q \quad \sim \quad \prod_i \mathcal{IG}(q_i; a_{Qi}, b_{Qi}) \tag{14}$$

Here, $\mathcal{IG}(x; a, b)$ denotes a inverse Gamma distribution with location parameter a and shape parameter b. The operator vec "reshapes" a matrix as a column vector by concatenating its columns and $\mathrm{diag}\, X$ is a column vector equal to the diagonal of the square matrix X. To maintain interpretability, we assume that C is known and is fixed as in (9). Similarly, we fix the hyper-parameter Ω_A ("expected transition matrix") to (8). In this paper, we further assume, that the covariance matrices P and R are known; it is straightforward to include them in the estimation procedure by using priors analogous to (14). The graphical model is shown in figure 1.

3. INFERENCE

To solve the problem in (1) exactly, we need to first infer the posterior distribution

$$p(\mathcal{S}, \Theta | x_{\boldsymbol{\kappa}}) = \frac{1}{Z_x} p(x_{\boldsymbol{\kappa}} | \mathcal{S}, \Theta) p(\mathcal{S} | \Theta) p(\Theta) \equiv \frac{1}{Z_x} \phi(\mathcal{S}, \Theta) \equiv \mathcal{P} \tag{15}$$

[2] Notice that, we have $B(0) = I$ and $B(\frac{W}{2}\omega) = B(\pi) = -I$ respectively – the rotation matrices of the DC and highest frequency band. Hence $s_\Im^0$ and $s_\Im^{W/2}$ are unobservable; they can be removed from the model by redefining $B(0) \equiv 1$ and $B(\pi) \equiv -1$.

where $\mathcal{S} = s_{0:K-1}^{0:W-1}$, $\Theta = (A, Q)$ and $Z_x = p(x_{\boldsymbol{\kappa}})$ is a normalising constant (also known as the evidence or data likelihood). Subsequently, we need to compute the predictive distribution

$$p(x_{\neg\boldsymbol{\kappa}}|x_{\boldsymbol{\kappa}}) \quad = \quad \int d\mathcal{S}d\Theta p(x_{\neg\boldsymbol{\kappa}}|\mathcal{S},\Theta)p(\mathcal{S},\Theta|x_{\boldsymbol{\kappa}}) \quad (16)$$

Exact evaluation of the posterior distribution in (15) is intractable due to couplings between Θ and $\mathcal{S}$, so we will resort to approximations.

3 . 1 Structured Mean Field

One possible approximation method, that leads to a practical optimisation procedure is the *mean field* approach, also known as *variational Bayes* [9, 10, 11]. In the particular case of (1), mean field boils down to approximating the exact posterior $\mathcal{P}$ in (15) with a simple distribution $\mathcal{Q}$ in such a way that the integral expression (16) becomes tractable. An intuitive interpretation of mean field is minimising the KL divergence with respect to (the parameters of) $\mathcal{Q}$ where

$$KL(\mathcal{Q}||\mathcal{P}) \quad = \quad \langle \log \mathcal{Q} \rangle_{\mathcal{Q}} - \left\langle \log \frac{1}{Z_x}\phi(\mathcal{S},\Theta) \right\rangle_{\mathcal{Q}} \quad (17)$$

Here, $\langle f(x) \rangle_{p(x)} \equiv \int dxp(x)f(x)$ denotes the expectation of f w.r.t. p. Using non-negativity of KL [12] we obtain a lower bound on the evidence

$$\log Z_x \quad \geq \quad \langle \log \phi(\mathcal{S},\Theta) \rangle_Q - \langle \log \mathcal{Q} \rangle_{\mathcal{Q}} \quad (18)$$

It is clear that maximising this lower bound is equivalent to finding the "nearest" $\mathcal{Q}$ to $\mathcal{P}$ in terms of KL. For the PPVOC model, we choose the approximating distribution $\mathcal{Q}$ of form

$$\mathcal{Q} \quad \equiv \quad \prod_{\alpha \in \mathcal{C}} q(s_{0:K-1}^{\alpha})q(\Theta_\alpha) \equiv \prod_{\alpha \in \mathcal{C}} \mathcal{Q}_\alpha(\mathcal{S}_\alpha)\mathcal{Q}_\alpha(\Theta_\alpha)$$

where $\mathcal{C} = \{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N\}$ is a set of disjoint clusters of frequency bands such that $\boldsymbol{\nu}_i \cap \boldsymbol{\nu}_j = \emptyset$ for $i \neq j$ and $\bigcup_i \boldsymbol{\nu}_i = \{0, \dots, W-1\}$. The parameter $\Theta_\alpha = (A_\alpha, Q_\alpha)$ denotes the diagonal blocks of A and Q matrices, that correspond to the frequency bands $\nu \in \alpha$. In contrast to naive mean field where $\mathcal{Q}$ is fully factorized, it is natural to adapt a *structured* mean field method [13, 14] by choosing a Gauss-Markov chain for $q(s_{0:K-1}^{\alpha})$ terms:

$$\begin{aligned} q(s_{0:K-1}^{\alpha}) \quad &\equiv \quad q(s_0^{\alpha}) \prod_{k=1}^{K} q(s_k^{\alpha}|s_{k-1}^{\alpha}) \\ q(\Theta_\alpha) \quad &\equiv \quad q(A_\alpha)q(Q_\alpha) \end{aligned}$$

The parameter distributions can chosen to have the same functional form as (13) and (14). Although a closed form solution for $\mathcal{Q}$ still can not be found, it can be easily shown, e.g. see [15], that each factor potential $\mathcal{Q}_\alpha$ of the optimal approximating distribution should satisfy the following fixed point equation

$$\mathcal{Q}_\alpha \quad \propto \quad \exp\left(\langle \log \phi(\mathcal{S},\Theta) \rangle_{\mathcal{Q}_{\neg\alpha}} \right) \quad (19)$$

where $\mathcal{Q}_{\neg\boldsymbol{\alpha}} \equiv \mathcal{Q}/\mathcal{Q}_\alpha$, i.e. product of all factors excluding $\mathcal{Q}_\alpha$. Hence, the mean field approach leads to a set of fixed point equations that need to be iterated.

3 . 2 Connection to EM

There is a direct link between the mean field approximation and the EM algorithm for parameter estimation [16]. One way to see this is to constrain the parameter distribution to have the form $\hat{q}(\Theta) = \delta(\Theta - \Theta^*)$, where δ is a Dirac pulse. Because of the additional constraint, we need to find the "closest" degenerate distribution to the actual mean field solution $q(\Theta)$ given in (19). Hence, we minimize a second KL as

$$\begin{aligned} \Theta^* \quad &= \quad \arg\min_{\theta} KL(\delta(\Theta - \theta)||q(\Theta)) \\ &= \quad \arg\max_{\theta} \langle \log \phi(\mathcal{S},\theta) \rangle_{q(\mathcal{S})} = \arg\max_{\theta} q(\theta) \quad (20) \end{aligned}$$

This resulting algorithm is equivalent to EM, where E and M steps correspond to computing the expectation w.r.t. $q(\mathcal{S})$ and subsequently finding the best parameter Θ^*. Due to the degenerate form of $\hat{q}(\theta)$, the update step for $q(\mathcal{S})$ is trivialized as:

$$\log q(\mathcal{S}) \quad = \quad \langle \log \phi(\mathcal{S},\Theta) \rangle_{\delta(\Theta-\Theta^*)} = \log \phi(\mathcal{S},\Theta^*) \quad (21)$$

The algorithm proceeds iterating (20) and (21) as in regular EM. When there is only one cluster with $\mathcal{C} = \{\{0 \dots W-1\}\}$, this algorithm is equivalent to EM for the linear dynamical system [17]. In this paper, we will provide results for the EM case only.

4 . RESULTS

The simulations were performed on the data set used in [3]: 16-bit 44.1 kHz signals are downsampled to 11.025 Hz and corrupted by a series of gaps in the range of 2ms-4ms. The two examples (piano, trumpet) are destorted such that 36.5% and 37.2% of the samples are missing. The data set and reconstruction results will be made available at `http://www-sigproc.eng.cam.ac.uk/~atc27/em-restore/`.

In the first experiment, our aim is to test the effect of the clustering choice on to the SNR (Signal-to-noise ratio) of the reconsturction, where the reconstruction is calculated as $\langle y \rangle_{\mathcal{Q}}$. We run our test with the number of frequency bands $W = 64$ and on the first 600 samples of the corrupted trumpet signal. Adjacent frequency bands are put into clusters, e.g., when $CS = 8$, $\mathcal{C} = \{\{0, \dots, 7\}, \{8, \dots, 15\}, \dots\}$. The results in table 1 suggest that adapting the transition matrix improves the SNR, significantly so when the clusters are small.

CS	adapt A	fixed A
1	4.8	2.3
2	5.5	3.1
4	9.7	5.3
8	9.6	5.1
16	10.9	8.8
32	7.6	8.9

Table 1: Comparison of SNR (in dB) of the reconstruction with and without adapting the transition matrix A versus cluster size CS.

In Table 2, we give a short table a typical results when $W = CS = 40$.

Our results and informal subjective listening tests suggest that we can get reconstruction performance comparable to existing methods. One advantage of our variational approximation technique, when compared to MCMC, is that it tends to converge quite fast to a solution. This suggests that, whilst the exact problem is intractable, the parameter regime useful for restoration of audio admits a simple deterministic approximation.

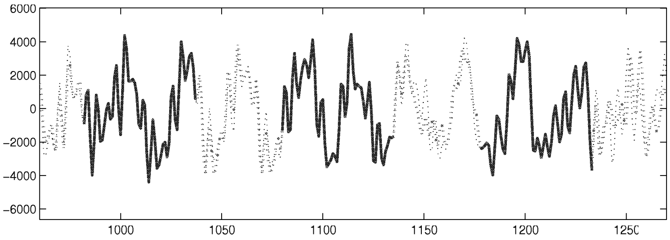

Figure 2: Illustration of the reconstruction result on segments from trumpet signal. In this example, $W = 32$ frequency bands are divided into two clusters with $CS = 16$. Solid line denote the observed samples. Small dotted line is the original signal, big dotted line is the reconstruction.

	sin-AR [1]	W&G [3]	PPVOC-EM
piano	1.46	10.17	7.68
trumpet	-	5.94	7.10

Table 2: Comparison of improvement in SNR (in dB). Missing samples are assumed to be zero when computing the SNR. The model achieves comperable performance to W&G [3].

5 . CONCLUSIONS

The PPVOC model described in this article is an instance of the linear dynamical system, also known as the Kalman filter model, e.g. see [18]. In our case, the interpretation of the particular parametrization is more important: The transition model at each chain $s^{\nu}_{0:K-1}$ generates an independent "basis function". The observation model adds up each basis function to generate the observation $x_{0:K-1}$. In contrast to the DFT, where there is a strictly sinusoidal basis, in PPVOC, the individual basis functions are "generated" by a stochastic process. By fine tuning the parameters A and Q of this process, the model has more flexibility to adapt itself to the statistics of the observed audio signal. For example, adjusting a diagonal block of A is effectively equivalent to tuning the center frequency of the corresponding frequency band. This flexibility also allows a Bayesian treatment of the problem in (1). Moreover, the dynamic system formulation circumvents problems associated with using a fixed analysis window length or finding an optimal basis set.

Although our initial results are promising, two questions are still open and require further research for an answer: (1) Does the PPVOC model provide a practical alternative for audio restoration? (2) Is the EM approach sufficient as an approximation method ? At this stage, it is still early to answer the first question, and more simulation studies have to be carried out. To answer the second question, we are currently testing full variational approximation where the parameters are integrated out. This direction is attractive, since for related models good results were reported, e.g. see [10]. Alternatively, sampling methods should also be compared.

Acknowledgments A. T. Cemgil would like to thank to Christopher Raphael for helpful discussions about the phase vocoder.

REFERENCES

[1] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1124–1135, 1996.

[2] P. A. A. Esquef, "Interpolation of long gaps in audio signals using line spectrum pair polynomials," Helsinki University of Technology, Tech. Rep. 72, 2004.

[3] P. J. Wolfe and S. Godsill, "Interpolation of missing values using a gabor regression model," in *Submitted to ICASSP 2005*.

[4] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, pp. 1493–1509, November 1966.

[5] M. Dolson, "The phase vocoder: A tutorial," *Computer Music Journal*, vol. 10, no. 4, pp. 14–27, 1986.

[6] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323–332, May 1999.

[7] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *Accepted to IEEE Transactions on Speech and Audio Processing*, 2004.

[8] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge, U.K.: Cambridge Univ. Press, 1989.

[9] W. Wiegerinck, "Variational approximations between mean field theory and the junction tree algorithm," in *UAI-2000 (16-th conference)*, pp. 626–633.

[10] Z. Ghahramani and M. Beal, "Propagation algorithms for variational Bayesian learning," in *Neural Information Processing Systems 13*, 2000.

[11] M. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," Department of Statistics, UC Berkeley, Tech. Rep. 649, September 2003.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, Inc., 1991.

[13] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, no. 29, pp. 245–273, 1997.

[14] D. Barber and W. Wiegerinck, "Tractable variational structures for approximating graphical models," in *Advances in Neural Information Processing Systems (NIPS)*, M. Kearns, S. Solla, , and D. Cohn, Eds., 1999, pp. 183–189.

[15] J. Winn and C. Bishop, "Variational message passing," *Submitted to Journal of Machine Learning Research*, 2004.

[16] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in graphical models.* MIT Press, 1999, pp. 355–368.

[17] R. Shumway and D. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *J. Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.

[18] K. P. Murphy, "Dynamic Bayesian networks: Representation, inference and learning," Ph.D. dissertation, University of California, Berkeley, 2002.