UNIFYING PROBABILISTIC MODELS FOR TIME-FREQUENCY ANALYSIS

William J. Wilkinson[1]*, *Michael Riis Andersen*[2], *Joshua D. Reiss*[1], *Dan Stowell*[1], *and Arno Solin*[2]†

[1]Centre for Digital Music
Queen Mary University of London
United Kingdom

[2]Department of Computer Science
Aalto University
Finland

ABSTRACT

In audio signal processing, probabilistic time-frequency models have many benefits over their non-probabilistic counterparts. They adapt to the incoming signal, quantify uncertainty, and measure correlation between the signal's amplitude and phase information, making time domain resynthesis straightforward. However, these models are still not widely used since they come at a high computational cost, and because they are formulated in such a way that it can be difficult to interpret all the modelling assumptions. By showing their equivalence to Spectral Mixture Gaussian processes, we illuminate the underlying model assumptions and provide a general framework for constructing more complex models that better approximate real-world signals. Our interpretation makes it intuitive to inspect, compare, and alter the models since all prior knowledge is encoded in the Gaussian process kernel functions. We utilise a state space representation to perform efficient inference via Kalman smoothing, and we demonstrate how our interpretation allows for efficient parameter learning in the frequency domain.

Index Terms— probabilistic time-frequency analysis, Gaussian processes, state space models

1. INTRODUCTION

Time-frequency (TF) analysis is a ubiquitous technique for uncovering the time-varying spectral properties of signals, and it commonly plays the part of a pre-processing module for machine learning and signal processing tasks. However, traditional TF analysis requires various choices to be made regarding windowing functions, transfer functions or wavelets, depending on the representation being used [1]. It is not clear how best to make these choices or what their implications are on tasks such as classification or source separation.

Probabilistic TF analysis [2] promises to remove the need for these difficult decisions by adapting to the incoming signal [3, 4, 5, 6] and by propagating uncertainty information to downstream applications [2, 7]. By specifying a probabilistic model characterised by parameters corresponding to traditional model features, such as centre frequencies and bandwidths of a filter bank, a posterior distribution over the frequency components given the data can be found. Different modelling choices can be compared in a principled manner by evaluating the model likelihood given the parameters, which allows for parameter tuning in order to find the statistically optimal TF representation for a given signal.

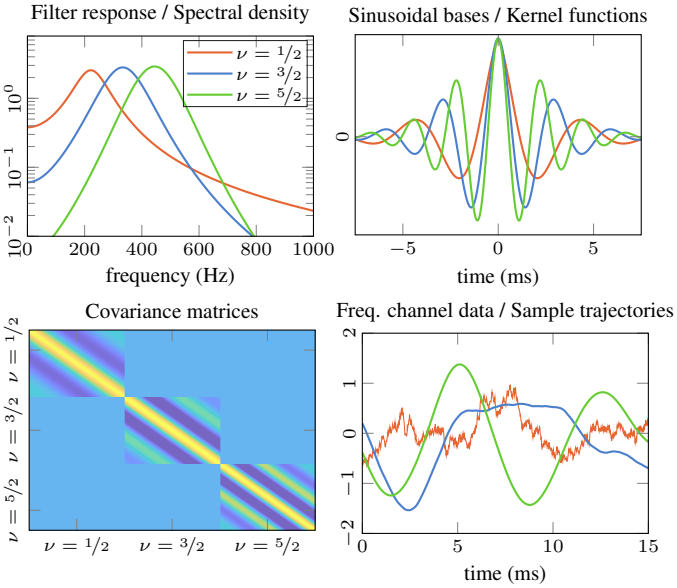

Fig. 1: Four representations of the same Gaussian process-based probabilistic filter bank (with three filters). Each filter / process is a frequency shifted Matérn-ν GP. All three filters have the same lengthscale (bandwidth) parameter, but they exhibit quite different spectral densities (**top-left**). See main text for demonstration of how filter banks can be represented in canonical GP form, such as with kernel functions (**top-right**) and covariance matrices (**bottom-left**). Samples from the prior vary in smoothness (**bottom-right**), suggesting that the choice of ν will affect how the model fits the signal.

Probabilistic models that act directly on the signal waveform implicitly measure correlation between a signal's amplitude and phase information [8], which has the major implication that time-domain synthesis does not require a phase-reconstruction stage. This ability to sample new data from the

*Corresponding author: w.j.wilkinson@qmul.ac.uk.
†AS acknowledges funding from the Academy of Finland (308640).

generative model makes missing data imputation and noise reduction tasks intuitive.

Despite these benefits, existing probabilistic TF models are still not widely used, perhaps due to their higher computational complexity and because they are formulated in such a way that they can be difficult to interpret and understand.

In the field of machine learning, Gaussian processes (GPs) [9] are an increasingly popular non-parametric approach for learning and decision making. In their standard formulation, they are characterised by covariance matrices that capture the hidden structure of data. Covariance matrices are constructed by evaluating *kernel functions* that encode our prior knowledge about the system we are modelling. A major issue with this approach for time-series data is that evaluation of the covariance matrix is impractical for all but the shortest of real-world signals. It is well known that many probabilistic TF methods can be posed as GPs [2], but it is generally unclear how all the modelling assumptions relate to the standard set of GP techniques.

In [10], GPs, along with their neural network counterparts, are presented as *"intelligent agents"* capable of automating the learning and decision making process. It is shown how complex prior knowledge can be encoded in the system by constructing new kernel functions composed of the sum and product of simpler ones. One such class of functions presented in [10] are Spectral Mixture kernels, defined for one-dimensional inputs as

$$\kappa_{\text{sm}}(t,t') = \sum_{d=1}^{D} \sigma_d^2 \cos(\omega_d(t-t')) \exp\!\big(-(t-t')^2/\ell_d^2\big),$$

which comprises a sum of frequency-shifted *radial basis function* kernels [9] and whose spectral density is a mixture of Gaussians. This idea is extended to the entire Matérn kernel class in [11, 12], producing Cauchy-Lorentz densities.

In this work, we show that probabilistic TF analysis and Matérn Spectral Mixture GPs are in fact equivalent. In other words, Spectral Mixture kernels are probabilistic filter banks. By doing so we reinterpret TF modelling assumptions under the GP paradigm. We provide a general procedure for rewriting Spectral Mixture GPs in discrete state space form, such that more complex TF models can be easily constructed, and inference can be performed efficiently via Kalman smoothing, whose computational complexity scales linearly in the number of time steps T and cubicly in state dimensionality M, $\mathcal{O}(M^3T)$. We then show how to utilise the continuous spectral density of the kernel functions to optimise the model parameters in the frequency domain.[1]

After outlining our framework and formalising the equivalence between these two modelling paradigms in Section 2, we go on to illustrate some potential modifications to the standard probabilistic TF approach in Section 3, evaluating the impact of these updates on a missing data synthesis task.

[1] Matlab code for all methods and experiments is available at: `https://github.com/wil-j-wil/unifying-prob-time-freq`

2. STATE SPACE GAUSSIAN PROCESS MODELS FOR TIME-FREQUENCY ANALYSIS

Various models for Bayesian treatment of time-frequency analysis have been proposed, most notably Bayesian spectrum estimation (BSE) [4], and the probabilistic phase vocoder (PPV) [5]. For an overview see [2], where it is also shown that the BSE and PPV models are equivalent up to a shift in frequency. We proceed by considering the PPV model, reformulating it with the aim of unifying these models in a common Gaussian process framework to illuminate the underlying modelling assumptions.

Readers should keep in mind that all the models (1)–(5) *written in this section are exactly equivalent to one another. By presenting them this way, we show multiple perspectives on spectral data analysis.*

2.1. Probabilistic phase vocoder

The standard discrete-time PPV can be written as follows,

$$\begin{aligned} &\text{[Prior]} && \mathrm{z}_{d,k} = \psi_d \mathrm{e}^{\mathrm{i}\,\omega_d} \mathrm{z}_{d,k-1} + \rho_d\, \zeta_{d,k}, \\ &\text{[Likelihood]} && y_k = \sum_{d=1}^{D} \mathrm{Re}[\mathrm{z}_{d,k}] + \sigma_{\mathrm{y}_k} \varepsilon_k, \end{aligned} \tag{1}$$

where k indexes the time step, with complex phasor $z_{d,k} \in \mathbb{C}$ being the (latent) subband signal in frequency channel $d = 1, \ldots, D$. y_k denotes the observed audio signal at t_k and both ε_k and $\zeta_{d,k}$ are i.i.d. Gaussian noise $\mathrm{N}(0,1)$, real-valued and complex-valued, respectively. Note that the noise scale σ_{y_k} can be non-stationary. Parameters ψ_d and ρ_d represent the process and noise variances respectively, whilst ω_d is the instantaneous angular frequency.

Recognising that Eq. (1) is a complex first-order autoregressive process makes it straightforward to write down the model's state space form,

$$\begin{aligned} &\text{[Prior]} && \mathbf{z}_{k+1} = \mathbf{A}\mathbf{z}_k + \mathbf{q}_k, \qquad \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}), \\ &\text{[Likelihood]} && y_k = \mathbf{H}\mathbf{z}_k + \sigma_{\mathrm{y}_k} \varepsilon_k, \end{aligned} \tag{2}$$

for $\mathbf{z}_k = (\mathrm{Re}[\mathrm{z}_{1,k}]\ \mathrm{Im}[\mathrm{z}_{1,k}]\ \ldots\ \mathrm{Re}[\mathrm{z}_{D,k}]\ \mathrm{Im}[\mathrm{z}_{D,k}])^{\mathsf{T}}$ and measurement matrix $\mathbf{H} = (1\ 0\ \ldots\ 1\ 0)$, with transition matrix $\mathbf{A}$ and process noise covariance matrix $\mathbf{Q}$ defined by

$$\mathbf{A} = \begin{pmatrix} \psi_1 \mathbf{R}(\omega_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \psi_D \mathbf{R}(\omega_D) \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \rho_1^2 \mathbf{I} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \rho_D^2 \mathbf{I} \end{pmatrix},$$

for rotation matrix $\mathbf{R}(\omega_d) = \left(\begin{smallmatrix} \cos\omega_d & -\sin\omega_d \\ \sin\omega_d & \cos\omega_d \end{smallmatrix}\right)$.

This linear Gaussian dynamical system is in the precise form required for inference via Kalman filtering and smoothing [13]. The filtering equations provide us with the necessary information required to evaluate the marginal likelihood $p(\mathbf{y}|\theta)$ and hence perform hyperparameter tuning. However, in practice it is much more efficient to tune the hyperparameters in the frequency domain, as discussed in the Section 2.5.

2.2. PPV model in canonical GP form

Gaussian processes are commonly used in Bayesian inference as non-parametric prior distributions on functions [9]. A GP prior, $f \sim \mathrm{GP}\left(\mu(\cdot), \kappa(\cdot,\cdot)\right)$, is completely characterized by a mean function, $\mu(t)$, and a kernel function, $\kappa(t,t')$.

We first write down the PPV's kernel-based GP representation, before going on to show equivalence to Eq. (1),

$$\begin{aligned} &\text{[Prior]} && f(t) \sim \mathrm{GP}(0, \sum_{d=1}^{D} \kappa_{\cos}^{(d)}(t,t')\,\kappa_{\exp}^{(d)}(t,t')), \\ &\text{[Likelihood]} && y_k = f(t_k) + \sigma_{\mathrm{y}_k}\,\varepsilon_k, \end{aligned} \tag{3}$$

where $\kappa_{\cos}^{(d)}(t,t') = \cos(\omega_d\,(t-t'))$ is a deterministic kernel whose function realisations are pure sinusoids, and $\kappa_{\exp}^{(d)}(t,t') = \sigma_d^2 \exp(-|t-t'|/\ell_d)$ is the exponential kernel, otherwise known as the Matérn-$1/2$. The cosine kernel acts as a frequency shift operator, centring the spectral density of the exponential kernel around ω_d.

When written in this form, it becomes apparent that the kernel in Eq. (3) has a close connection with Spectral Mixture kernels. Specifically it belongs to the Matérn Spectral Mixture family used to model harmonic priors over musical audio signals [11, 12]. This explicit link with Spectral Mixture models has not been previously explored, however filter banks composed in this way are reminiscent of the exponential kernels described in [14].

We will now demonstrate the equivalence between the model in Eq. (3) and the PPV model in Eq. (2) by converting it back to discrete state space form. In doing so, we will outline a general procedure for reformulating models in the form of Eq. (3) in this way, such that efficient inference methods can be applied even after the model has been altered.

2.3. The corresponding continuous state space model

The model in Eq. (3) has an equivalent representation in terms of a linear time-invariant (LTI) stochastic differential equation (SDE, see, e.g., [15]). The prior (or dynamics) of the system can be written in terms of a driving Brownian motion (with spectral density $\mathbf{Q}_\mathrm{c}$) in Itô form:

$$\begin{aligned} &\text{[Prior]} && \mathrm{d}\mathbf{f}(t) = \mathbf{F}\mathbf{f}(t)\,\mathrm{d}t + \mathbf{L}\,\mathrm{d}\boldsymbol{\beta}(t), \\ &\text{[Likelihood]} && y_k = \tilde{\mathbf{H}}\mathbf{f}(t_k) + \sigma_{\mathrm{y}_k}\varepsilon_k, \end{aligned} \tag{4}$$

where $\mathbf{f}(t) : \mathbb{R} \to \mathbb{R}^M$ for state space order M, and $\mathbf{F}$, $\mathbf{L}$, and $\tilde{\mathbf{H}}$ are model matrices. Building on previous work [16], we may write the GP prior in Eq. (3) as an LTI SDE of block-Kronecker structure:

$$\mathbf{F} = \mathrm{blkdiag}(\mathbf{F}_{\cos}^{(1)} \oplus \mathbf{F}_{\exp}^{(1)}, \ldots, \mathbf{F}_{\cos}^{(D)} \oplus \mathbf{F}_{\exp}^{(D)}),$$

where '$\oplus$' denotes the Kronecker sum, $\mathbf{F}_{\cos}^{(d)} = \left(\begin{smallmatrix} 0 & -\omega_d \\ \omega_d & 0 \end{smallmatrix}\right)$, and $\mathbf{F}_{\exp}^{(d)} = -1/\ell_d$ (because the exponential covariance function has an exact LTI SDE representation [15]). Similarly, the rest of the model matrices are given in terms of the Kronecker products of the submodel matrices.

2.4. Returning to discrete state space form

LTI SDE models such as Eq. (4) have an exact discrete-time solution, and the corresponding state space model is given by [15]:

$$\begin{aligned} &\text{[Prior]} && \mathbf{f}_{k+1} = \tilde{\mathbf{A}}\mathbf{f}_k + \tilde{\mathbf{q}}_k, \qquad \tilde{\mathbf{q}}_k \sim \mathrm{N}(\mathbf{0}, \tilde{\mathbf{Q}}), \\ &\text{[Likelihood]} && y_k = \tilde{\mathbf{H}}\mathbf{f}_k + \sigma_{\mathrm{y}_k}\varepsilon_k, \end{aligned} \tag{5}$$

where $\mathbf{f}_k$ is the M dimensional state, $\tilde{\mathbf{A}} = \exp(\mathbf{F}\,\Delta t)$ and $\tilde{\mathbf{Q}} = \mathbf{P}_\infty - \tilde{\mathbf{A}}\mathbf{P}_\infty\tilde{\mathbf{A}}^\mathsf{T}$. The stationary state covariance $\mathbf{P}_\infty$ is straightforward to calculate for most common kernel functions [15], and in the exponential kernel case is $\mathbf{P}_\infty = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_D^2)$.

Given that $\exp(\mathbf{F}_{\cos}^{(d)}\Delta t) = \mathbf{R}(\omega_d)$, performing these calculations for our PPV model ($M = 2D$) results in the following parametrisation:

$$\tilde{\mathbf{A}} = \begin{pmatrix} {\scriptstyle \eta_1\mathbf{R}(\omega_1)} & & 0 \\ & \ddots & \\ 0 & & {\scriptstyle \eta_D\mathbf{R}(\omega_D)} \end{pmatrix}, \quad \tilde{\mathbf{Q}} = \begin{pmatrix} {\scriptstyle \alpha_1\mathbf{I}} & & 0 \\ & \ddots & \\ 0 & & {\scriptstyle \alpha_D\mathbf{I}} \end{pmatrix},$$

where $\eta_d = \exp(-\Delta t/\ell_d)$ and $\alpha_d = \sigma_d^2(1-\exp(-2\Delta t/\ell_d))$. Finally it is clear that the models in Eq. (5) and Eq. (2) are of identical form, and hence the probabilistic phase vocoder in Eq. (1) is equivalent to the GP model in Eq. (3) if we select parameters $\psi_d = \exp(-\Delta t/\ell_d)$ and $\rho_d^2 = \sigma_d^2(1 - \exp(-2\Delta t/\ell_d))$.

Whilst we have derived this model for the exponential kernel, the framework outlined above can be followed regardless of the kernel choice, as long as it can be written in state space form. This is possible for most commonly used kernel functions [15, 17]. An intuitive way to proceed now is to alter the kernel in Eq. (3) to fit our requirements for the form of the corresponding filter bank. We explore this idea in Section 3.

2.5. Frequency domain optimisation

The formal connection to probabilistic TF models allows us to utilise Bayesian spectrum analysis [18, 2] for frequency domain hyperparameter tuning in Spectral Mixture GPs. This is significantly faster than time domain optimisation, and we avoid getting stuck in local optima by fitting to a smoothed version of the signal spectrum. By optimising parameters of a stationary GP kernel, rather than coefficients of an autoregressive process, we guarantee stationarity of the filter bank. The Spectral Mixture GP perspective gives us direct access to the model spectrum $\gamma_{y,i}(\theta)$ via the sum of the kernel's spectral density functions, $\gamma_{y,i}(\theta) = \sum_{d=1}^{D} S_{d,i}(\theta) + T\sigma_y^2$, where $S_{d,i}(\theta)$ is the spectral density of the kernel $\kappa^{(d)}$ in Eq. (3) evaluated at frequency bin i.

We fit the parameters via optimisation of the log-likelihood,

$$\log p(y|\theta) = c - \frac{1}{2}\sum_{i=1}^{T}\left(\log(\gamma_{y,i}(\theta)) + \frac{|\tilde{y}_i|^2}{\gamma_{y,i}(\theta)}\right),$$

where $|\tilde{y}_i|^2 = |\sum_{k=1}^{T} \mathrm{FT}_{i,k} y_k|^2$ is the signal spectrum. Note that the cosine kernel in Eq. (3) shifts the spectral density of the exponential kernel such that $S_{d,i} = \frac{1}{2}\left(S^{\exp}_{d,i-\omega_d} + S^{\exp}_{d,i+\omega_d}\right)$.

3. MISSING DATA EXPERIMENT

The methodology outlined in Section 2 allows new TF models to be constructed, with increased freedom over the choice of covariance structure. Here we demonstrate the potential benefits of altering the modelling assumptions via a missing data synthesis task, similar to the one carried out in [2].

The first-order state space form of standard TF models (Eq. (1)) implies that instantaneous frequencies are not correlated through time [8]. Higher-order models encourage slowly-varying instantaneous frequencies, a feature of real-world signals that should be leveraged to aid the highly ill-posed task of inferring a TF representation from data.

Therefore one intuitive example of a way to update the model is to swap the exponential (Matérn-$1/2$, state dimensionality $M = 2D$) kernel with a similar function that admits a higher-order state space representation. This corresponds to a filter bank whose filter transfer functions are no longer first-order autoregressive processes, but take a more complex form. We use the Matérn-$3/2$ ($M = 4D$) and Matérn-$5/2$ ($M = 6D$) kernels, which correspond to second- and third-order filter banks respectively and whose spectral densities have flatter tails and taller peaks (see Fig. 1). Note that the Matérn-$1/2$ model corresponds to the standard PPV.

Each model, with $D = 40$ filters, was trained on 10 short speech excerpts (between 1 and 2 seconds in duration) and then used to filter versions of the recordings in which some data had been removed. Missing data gaps of between 1 ms and 20 ms were studied, with the results shown in Fig. 2. Whilst the differences are subtle (the overall models are similar), the higher-order models' reconstruction achieved an improved signal to noise ratio for all missing data durations averaged across the 10 speakers. We also calculated the PESQ score [19] (a standardised perceptual speech quality metric), which demonstrated some signs of improvement, however all models performed similarly for large gap durations.

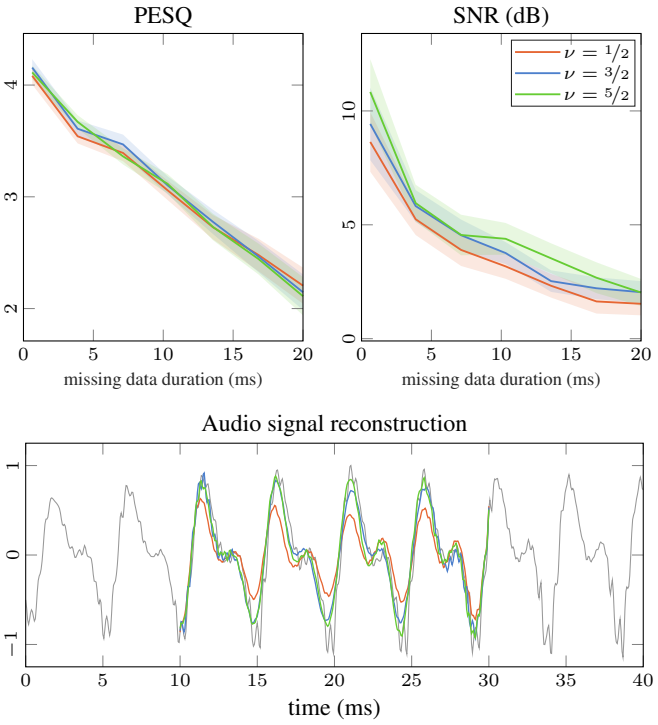

Fig. 2: Missing data synthesis results for three Matérn-ν probabilistic time-frequency models. Segments of data were removed from 10 speech recordings. Performance measured via perceptual quality metric (**top-left**) and signal-to-noise ratio (**top-right**) as a function of gap duration. Median value across speakers shown (shaded area is standard error). A reconstruction example (**bottom**) shows how the higher-order models ($\nu = 3/2, 5/2$) recover the overall shape in clearer detail (ground truth in grey). Matérn-$1/2$ is the standard probabilistic phase vocoder.

4. DISCUSSION

This paper serves to unify the theory surrounding probabilistic time-frequency analysis and explain clearly how it relates to Gaussian process modelling, with the hope of motivating further research at the intersection of these fields. We provide a general framework for converting spectral mixture GP models to a state space form that enables efficient frequency domain optimisation and efficient time domain filtering and prediction. We applied the framework to Matérn Spectral Mixture GPs and demonstrated improved performance over the standard probabilistic phase vocoder on a generative task.

Practical limitations of probabilistic time-frequency models still remain due to the Kalman smoother's cubic computational scaling in the state dimensionality and from the significant memory requirements involved in storing the entire covariance structure for every time step. Future work must focus on these practical issues.

Importantly, the methods presented here assume independence across frequency channels and don't explicitly model time-varying amplitude behaviour. It has been shown previously that a joint model over the TF representation and the amplitudes can result in significant improvement on tasks such as synthesis and noise reduction. Our state space framework provides a foundation on which to construct these more complex models.

5. REFERENCES

[1] Leon Cohen, *Time-frequency Analysis: Theory and Applications*, Prentice Hall, USA, 1995.

[2] Richard E. Turner and Maneesh Sahani, "Time-frequency analysis as probabilistic inference," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6171, 00 2014.

[3] Ervin Sejdić, Igor Djurović, and Jin Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, 2009.

[4] Yuan Qi, Thomas P. Minka, and Rosalind W. Picara, "Bayesian spectrum estimation of unevenly sampled nonstationary data," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2002, vol. 2, pp. II–1473.

[5] Ali Taylan Cemgil and Simon J. Godsill, "Probabilistic phase vocoder and its application to interpolation of missing values in audio signals," in *13th European Signal Processing Conference (EUSIPCO)*, 2005, pp. 1–4.

[6] Jingang Zhong and Yu Huang, "Time-frequency representation based on an adaptive short-time fourier transform," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5118–5128, 2010.

[7] Bradford W. Gillespie and Les E. Atlas, "Optimizing time-frequency kernels for classification," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 485–496, 2001.

[8] Richard E. Turner, *Statistical Models for Natural Sounds*, Ph.D. thesis, UCL, 2010.

[9] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[10] Andrew Wilson and Ryan Adams, "Gaussian process kernels for pattern discovery and extrapolation," in *Proceedings of the 30th International Conference on Machine Learning (ICML)*. 2013, vol. 28 of *Proceedings of Machine Learning Research*, pp. 1067–1075, PMLR.

[11] Pablo A. Alvarado and Dan Stowell, "Efficient learning of harmonic priors for pitch detection in polyphonic music," *arXiv preprint arXiv:1705.07104*, 2017.

[12] Pablo A. Alvarado, Mauricio A. Álvarez, and Dan Stowell, "Sparse gaussian process audio source separation using spectrum priors in the time-domain," *Submitted to International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019.

[13] Simo Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, 2013.

[14] H-I Choi and William J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[15] Simo Särkkä and Arno Solin, *Applied Stochastic Differential Equations*, Cambridge University Press, Cambridge, in press.

[16] Arno Solin and Simo Särkkä, "Explicit link between periodic covariance functions and state space models," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2014, vol. 33 of *Proceedings of Machine Learning Research*, pp. 904–912, PMLR.

[17] Jouni Hartikainen and Simo Särkkä, "Kalman filtering and smoothing solutions to temporal gaussian process regression models," in *International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2010, pp. 379–384.

[18] Larry G. Bretthorst, *Bayesian spectrum analysis and parameter estimation*, vol. 48, Springer Science & Business Media, 2013.

[19] Antony W. Rix, John G. Beerends, Michael P. Hollier, and Andries P. Hekstra, "Perceptual evaluation of speech quality (PESQ)–a new method for speech quality assessment of telephone networks and codecs," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2001, vol. 2, pp. 749–752.