End-to-End Probabilistic Inference for Nonstationary Audio Analysis

(or how to apply Spectral Mixture GPs to audio)

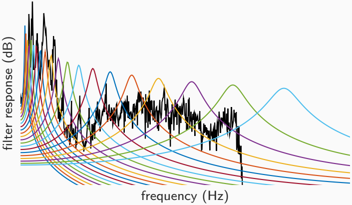
William Wilkinson, Michael Riis Andersen, Josh Reiss, Dan Stowell, Arno Solin

June 12, 2019

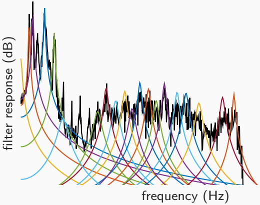
Queen Mary University of London / Aalto University / Technical University of Denmark

Probabilistic time-frequency analysis

We previously showed that a **spectral mixture Gaussian process is equivalent to a probabilistic filter bank**, i.e. a filter bank that adapts to the signal and can make predictions / generate new data.



standard filter bank



probabilistic / adaptive filter bank

Probabilistic time-frequency analysis

We previously showed that a **spectral mixture Gaussian process is equivalent to a probabilistic filter bank**, i.e. a filter bank that adapts to the signal and can make predictions / generate new data.

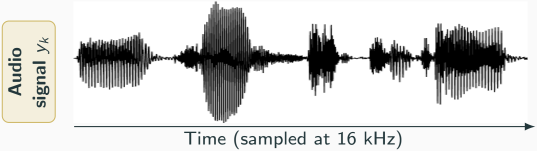
$$\text{[Prior]} \qquad f(t) \sim \mathrm{GP}\left(0, \sum_{d=1}^{D} \sigma_d^2 \exp(-|t - t'|/\ell_d) \cos(\omega_d\,(t - t'))\right),$$

$$\text{[Likelihood]} \qquad y_k = f(t_k) + \sigma_{\mathrm{y}_k}\,\varepsilon_k,$$

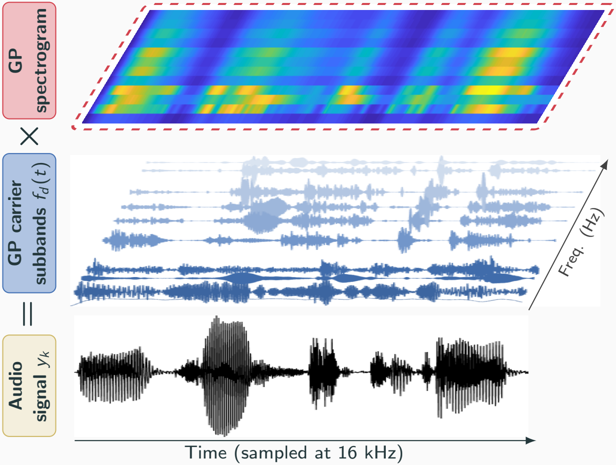
End-to-End probabilistic time-frequency analysis

The next step in the signal processing chain is often to analyse the dependencies in the spectrogram, with e.g. **non-negative matrix factorisation (NMF)**.

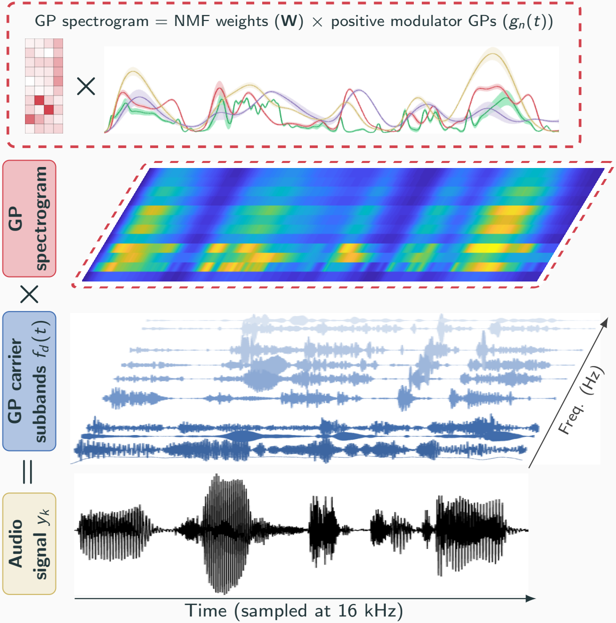
End-to-End probabilistic time-frequency analysis



End-to-End probabilistic time-frequency analysis



End-to-End probabilistic time-frequency analysis



The model

GP prior:

$$f_d(t) \sim \mathrm{GP}\big(0, \sigma_d^2 \exp(-|t - t'|/\ell_d) \cos(\omega_d\,(t - t')\big), \quad d = 1, 2, \ldots, D,$$

$$g_n(t) \sim \mathrm{GP}(0, \kappa_{\mathrm{g}}^{(n)}(t, t')), \quad n = 1, 2, \ldots, N,$$

The model

GP prior:

$$f_d(t) \sim \mathrm{GP}\big(0, \sigma_d^2 \exp(-|t-t'|/\ell_d) \cos(\omega_d\,(t-t')\big), \quad d = 1, 2, \ldots, D,$$
$$g_n(t) \sim \mathrm{GP}(0, \kappa_{\mathrm{g}}^{(n)}(t, t')), \quad n = 1, 2, \ldots, N,$$

Likelihood model:

$$y_k = \sum_d a_d(t_k)\, f_d(t_k) + \sigma_y\, \varepsilon_k,$$

for square amplitudes (the magnitude spectrogram):

$$a_d^2(t_k) = \sum_n W_{d,n}\, \mathrm{softplus}(g_n(t_k)),$$

The model

GP prior:

$$f_d(t) \sim \mathrm{GP}\big(0, \sigma_d^2 \exp(-|t - t'|/\ell_d) \cos(\omega_d\,(t - t')\big), \quad d = 1, 2, \ldots, D,$$
$$g_n(t) \sim \mathrm{GP}(0, \kappa_{\mathrm{g}}^{(n)}(t, t')), \quad n = 1, 2, \ldots, N,$$

Likelihood model:

$$y_k = \sum_d a_d(t_k)\, f_d(t_k) + \sigma_y\, \varepsilon_k,$$

for square amplitudes (the magnitude spectrogram):

$$a_d^2(t_k) = \sum_n W_{d,n}\, \mathrm{softplus}(g_n(t_k)),$$

This is a nonstationary spectral mixture GP

Inference

We show how to write the model as a **stochastic differential equation**:

$$\frac{\mathrm{d}\tilde{\mathbf{f}}(t)}{\mathrm{d}t} = \mathbf{F}\tilde{\mathbf{f}}(t) + \mathbf{L}\mathbf{w}(t),$$
$$y_k = \mathcal{H}(\tilde{\mathbf{f}}(t_k)) + \sigma_y \varepsilon_k,$$

such that inference can proceed via Kalman filtering & smoothing.

Inference

We show how to write the model as a **stochastic differential equation**:

$$\frac{\mathrm{d}\tilde{\mathbf{f}}(t)}{\mathrm{d}t} = \mathbf{F}\tilde{\mathbf{f}}(t) + \mathbf{L}\mathbf{w}(t),$$
$$y_k = \mathcal{H}(\tilde{\mathbf{f}}(t_k)) + \sigma_y \varepsilon_k,$$

such that inference can proceed via Kalman filtering & smoothing.

Usually the nonlinear $\mathcal{H}(\cdot)$ is dealt with via linearisation (EKF), but **we implement full Expectation Propagation (EP) in the Kalman smoother**, and the infinite-horizon solution which scales as:

$$\mathcal{O}(M^2 T)$$

Applications and Results

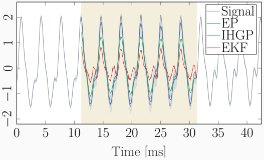
The fully probabilistic model can, **without modification**, be applied to:

Applications and Results

The fully probabilistic model can, **without modification**, be applied to:

Missing Data

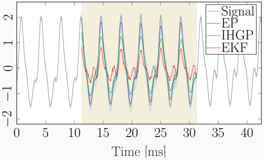
Synthesis



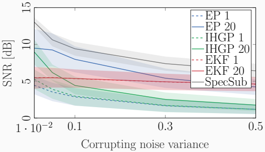
Applications and Results

The fully probabilistic model can, **without modification**, be applied to:

Missing Data Synthesis



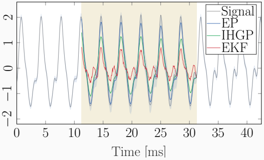
Denoising



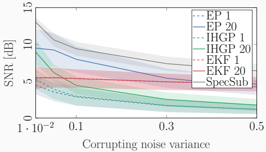
Applications and Results

The fully probabilistic model can, **without modification**, be applied to:

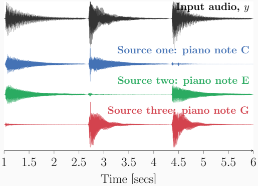
Missing Data Synthesis



Denoising



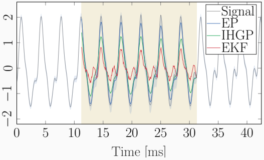
Source Separation



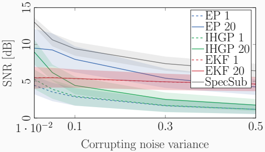
Applications and Results

The fully probabilistic model can, **without modification**, be applied to:

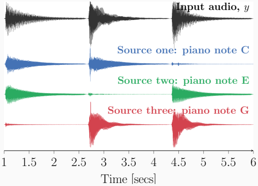
Missing Data Synthesis



Denoising



Source Separation



Thanks for listening! Poster: 6:30pm Weds, **Pacific Ballroom #217**

Contact: william.wilkinson@aalto.fi