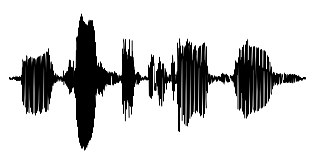
Gaussian Processes for Audio Feature Extraction

Dr. Richard E. Turner (ret26@cam.ac.uk)

Computational and Biological Learning Lab
Department of Engineering
University of Cambridge

Machine hearing pipeline

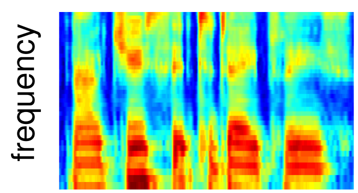
y signal



T samples

Machine hearing pipeline

$\mathcal{S}$ time-frequency (TF) analysis

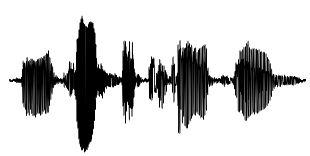


T'$\times$D > T samples

short time Fourier transform
spectrogram
wavelet
filter bank

(non-linear)

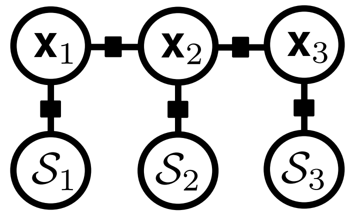
y signal



T samples

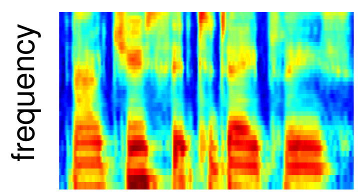
Machine hearing pipeline

probabilistic
model



HMM (speech recognition)
NMF (source separation, denoising)
ICA (source separation, denoising)

$\mathcal{S}$ time-frequency
(TF) analysis

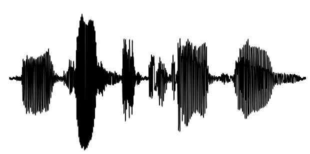


T'$\times$D > T samples

short time Fourier transform
spectrogram
wavelet
filter bank

(non-linear)

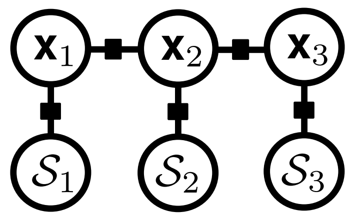
y signal



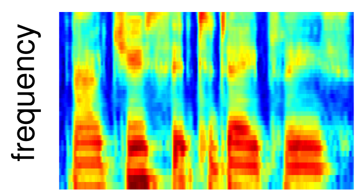
T samples

Problems with conventional pipeline

probabilistic model



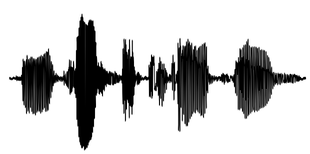
$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

(non-linear)

y signal



T samples

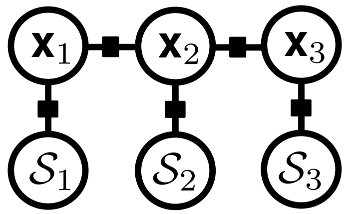
noise (source mixtures) hard to model in TF domain

$$\mathcal{S}(\mathrm{y}_1 + \mathrm{y}_2) \neq \mathcal{S}(\mathrm{y}_1) + \mathcal{S}(\mathrm{y}_2)$$

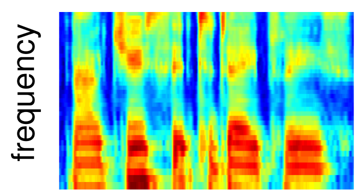
(hard to propagate uncertainty noise/missing data - from signal to TF domain)

Problems with conventional pipeline

probabilistic model



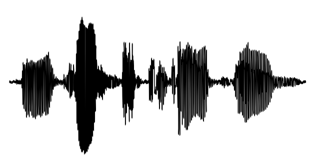
$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

(non-linear)

y signal



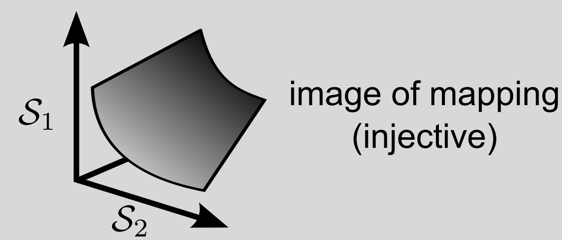
T samples

noise (source mixtures) hard to model in TF domain

$$\mathcal{S}(\mathsf{y}_1 + \mathsf{y}_2) \neq \mathcal{S}(\mathsf{y}_1) + \mathcal{S}(\mathsf{y}_2)$$

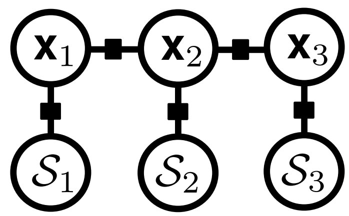
(hard to propagate uncertainty noise/missing data - from signal to TF domain)

hard to enforce/learn dependencies intrinsic to the FT analysis

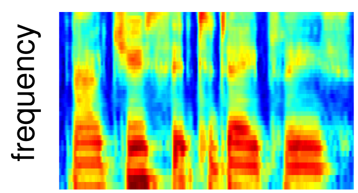


Problems with conventional pipeline

probabilistic model



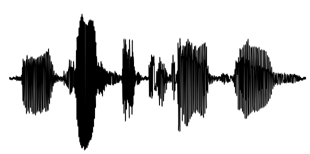
$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

(non-linear)

y signal



T samples

noise (source mixtures) hard to model in TF domain

$$\mathcal{S}(\mathrm{y}_1 + \mathrm{y}_2) \neq \mathcal{S}(\mathrm{y}_1) + \mathcal{S}(\mathrm{y}_2)$$

(hard to propagate uncertainty noise/missing data - from signal to TF domain)

hard to enforce/learn dependencies intrinsic to the FT analysis

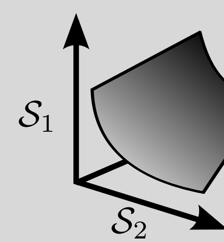


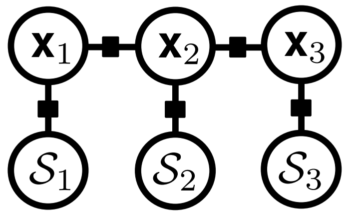
image of mapping (injective)

learning based on time-frequency representation ignores Jacobian

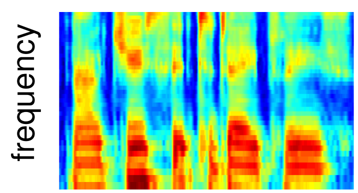
$$p(\mathcal{S}|\theta)$$

Problems with conventional pipeline

probabilistic model



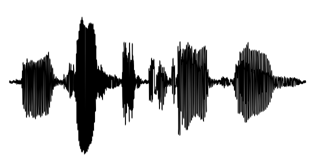
$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

(non-linear)

y signal



T samples

noise (source mixtures) hard to model in TF domain

$$\mathcal{S}(\mathsf{y}_1 + \mathsf{y}_2) \neq \mathcal{S}(\mathsf{y}_1) + \mathcal{S}(\mathsf{y}_2)$$

(hard to propagate uncertainty noise/missing data - from signal to TF domain)

hard to enforce/learn dependencies intrinsic to the FT analysis

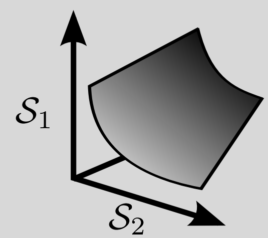


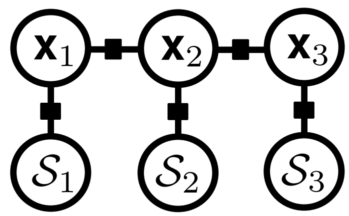
image of mapping (injective)

learning based on time-frequency representation ignores Jacobian

$$p(\mathsf{y}|\theta) = p(\mathcal{S}|\theta)\left|\frac{\mathrm{d}\mathcal{S}}{\mathrm{dy}}\right|$$

Problems with conventional pipeline

probabilistic model

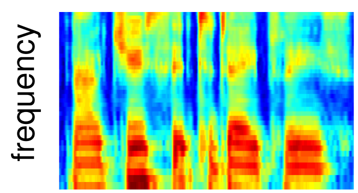


noise (source mixtures) hard to model in TF domain

$$\mathcal{S}(\mathrm{y}_1 + \mathrm{y}_2) \neq \mathcal{S}(\mathrm{y}_1) + \mathcal{S}(\mathrm{y}_2)$$

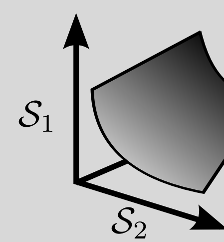
(hard to propagate uncertainty noise/missing data - from signal to TF domain)

$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

hard to enforce/learn dependencies intrinsic to the FT analysis

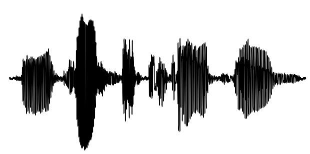


(non-linear)

learning based on time-frequency representation ignores Jacobian

$$p(\mathrm{y}|\theta) = p(\mathcal{S}|\theta)\left|\frac{\mathrm{d}\mathcal{S}}{\mathrm{dy}}\right|$$

y signal

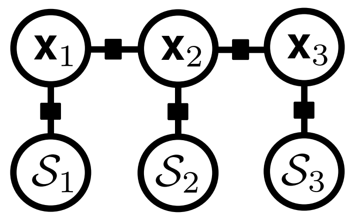


T samples

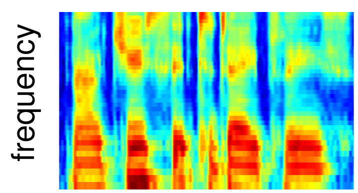
hard to adapt both top and bottom layers

Goal of this talk

probabilistic model



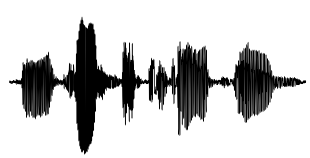
$\mathcal{S}$ time-frequency (TF) analysis



T'$\times$D > T samples

(non-linear)

y signal



T samples

probabilise time-frequency analysis
(construct generative model in which inference corresponds to classical time-frequency analysis)

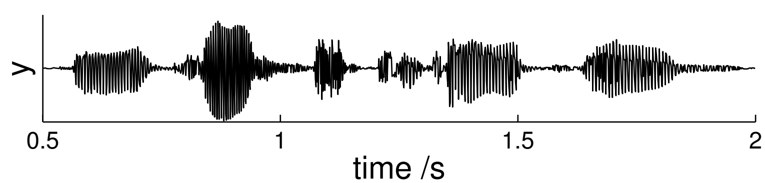
build a hierachical model that incorporates downstream processing module

classical signal processing

machine learning

A typical audio pipeline

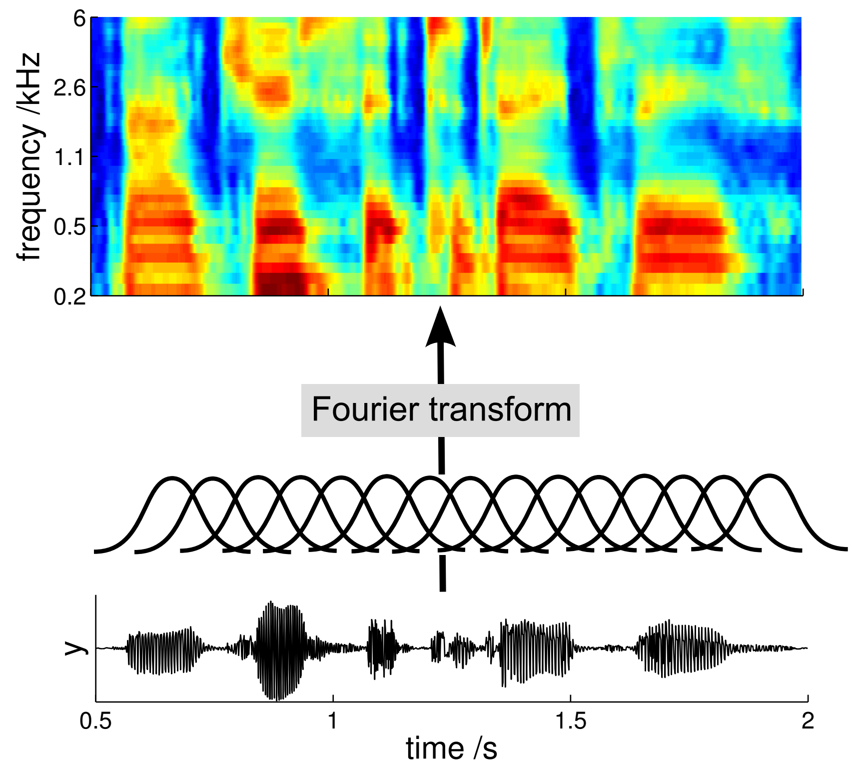
signal



A typical audio pipeline

spectrogram

signal

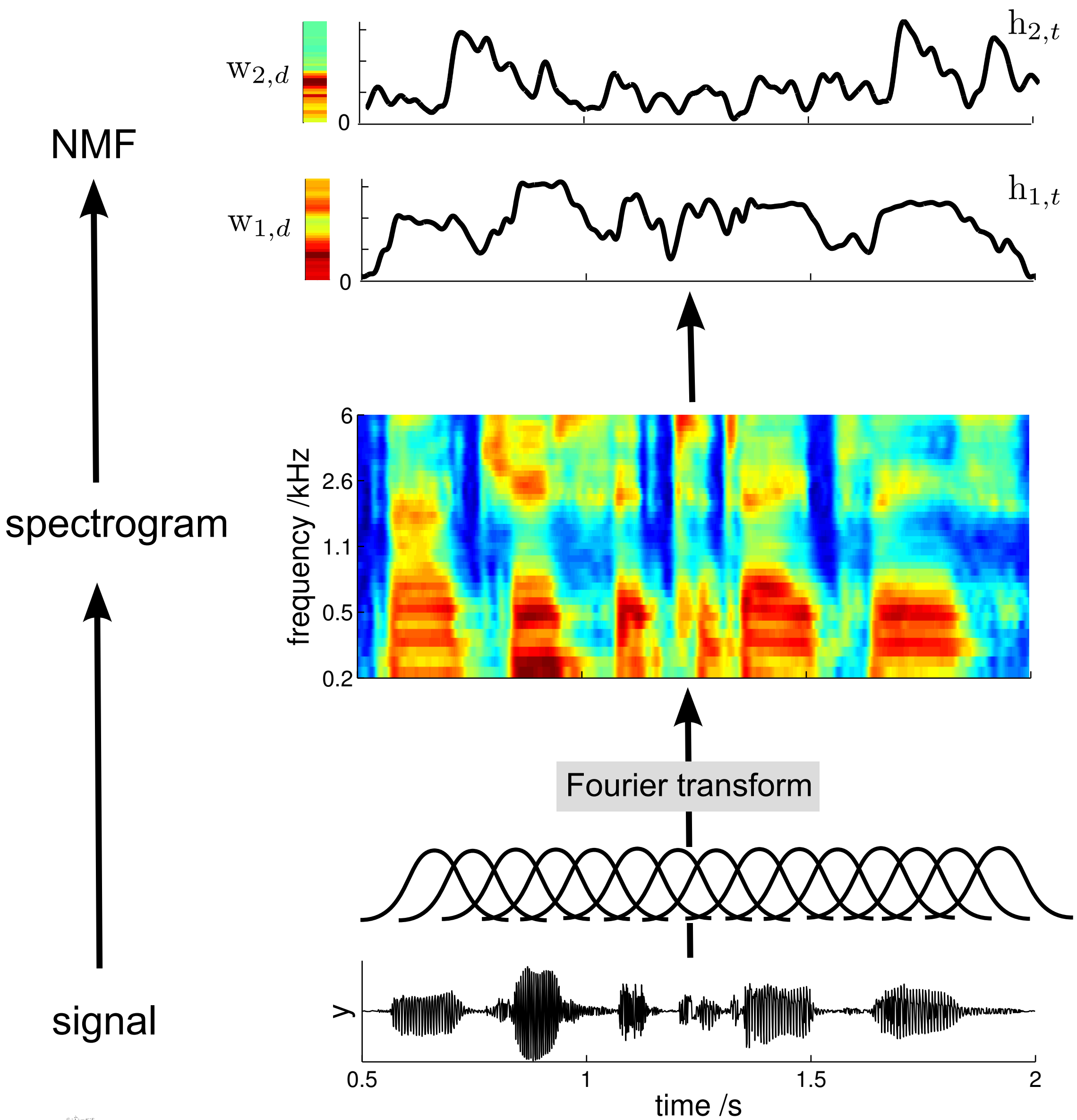


magnitude
$$\mathcal{S}_{d,t} = |\mathrm{s}^{\text{STFT}}_{d,t}|^2$$

short time
Fourier transform
$$\mathrm{s}^{\text{STFT}}_{d,t} = \sum_{t'} e^{-i\omega_d t'} \mathrm{W}_{t-t'} \mathbf{y}_{t'}$$

window
$$\mathrm{W}_{t-t'} \mathbf{y}_{t'}$$

A typical audio pipeline



$$\mathcal{S}_{d,t} = \sum_{l=1}^{L} h_{l,t} w_{l,d}$$

$$h_{l,t} > 0, w_{l,d} > 0$$

magnitude

$$\mathcal{S}_{d,t} = |s_{d,t}^{\text{STFT}}|^2$$

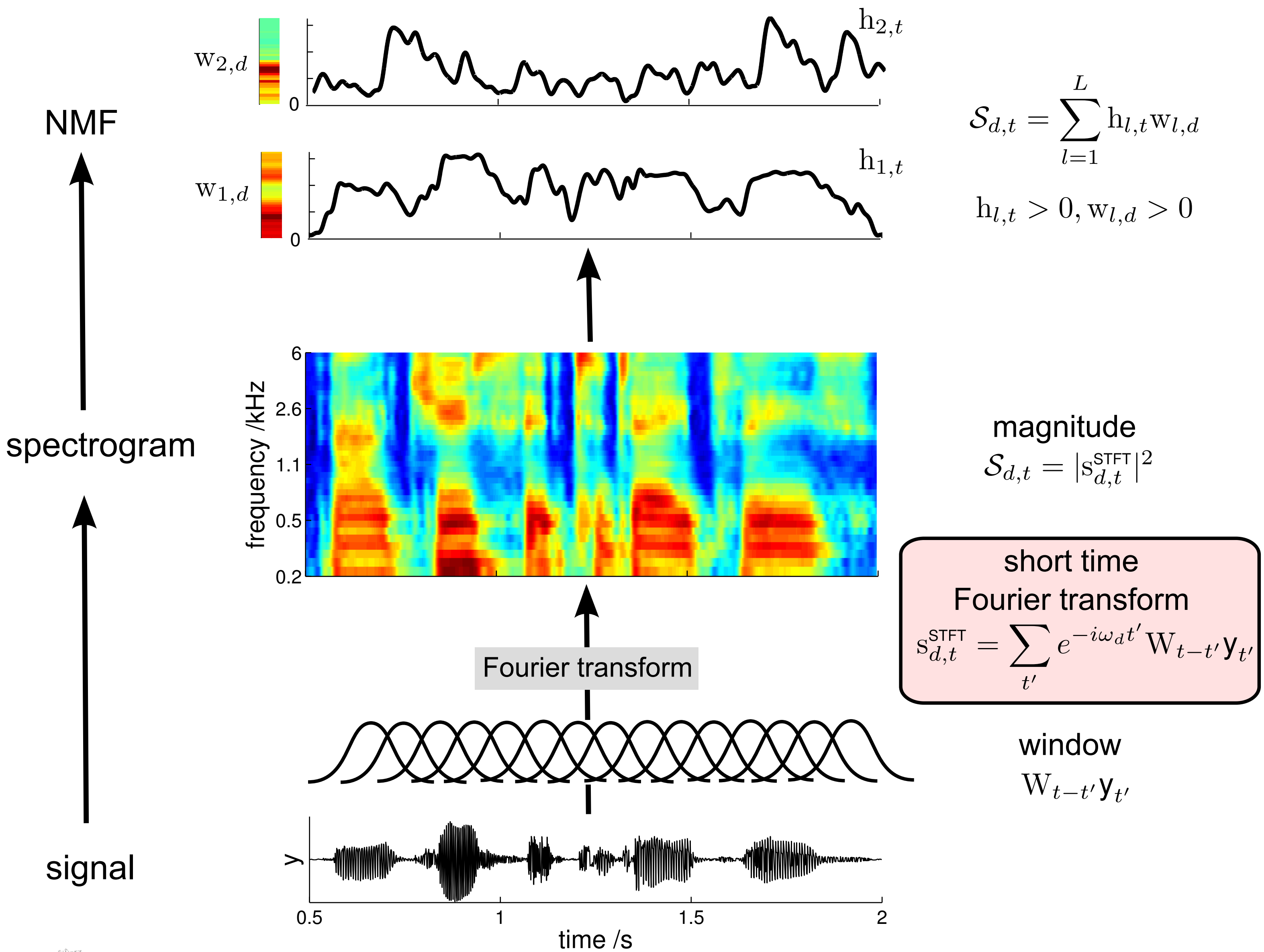
short time
Fourier transform

$$s_{d,t}^{\text{STFT}} = \sum_{t'} e^{-i\omega_d t'} W_{t-t'} y_{t'}$$

window

$$W_{t-t'} y_{t'}$$

A typical audio pipeline



What form of generative model corresponds to the STFT?

desire: expected value of latent time-frequency coefficients $\mathrm{s}_{d,1:T}$ = STFT

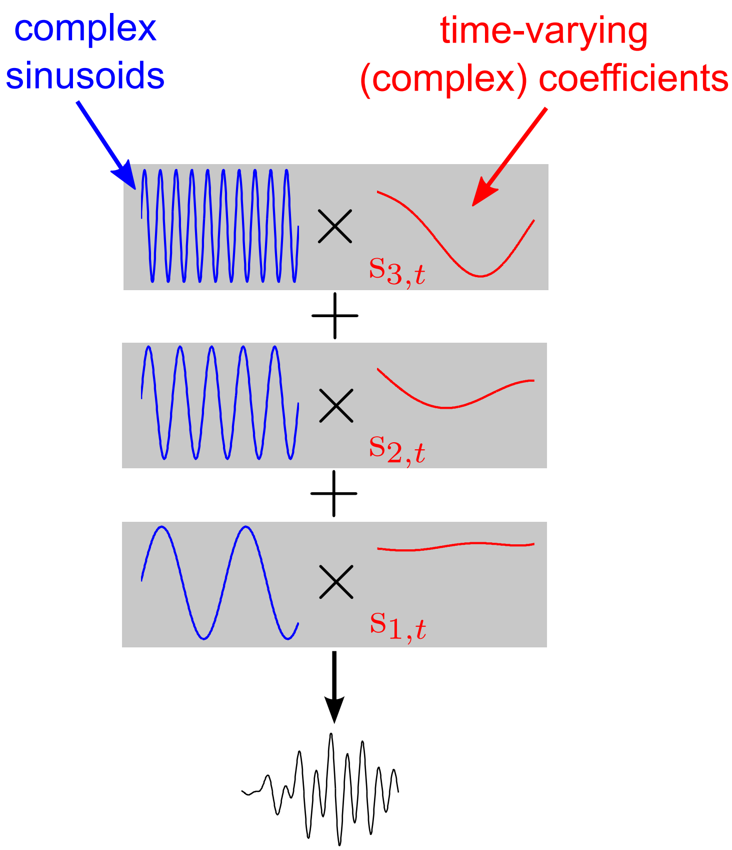
- assume y formed by (weighted) superposition of band-limited signals $\mathrm{s}_{d,1:T}$
- linearity of inference can be assured by setting the distributions of each $\mathrm{s}_{d,1:T}$ and the noise to be Gaussian
- time-invariance $\implies$ generative model statistically stationary

$\implies$ GP prior over STFT coefficients, $p(\mathrm{s}_{d,1:T}) = \mathcal{G}(\mathrm{s}_{d,1:T}; 0, \Gamma)$, stationary

$$\Gamma_{t,t'} \approx \sum_{k=1}^{T} \mathrm{FT}^{-1}_{t,k} \gamma_k \mathrm{FT}_{k,t'} \text{ where } \mathrm{FT}_{k,t} = e^{-2\pi i (k-1)(t-1)/T}$$

Time-frequency analysis as inference

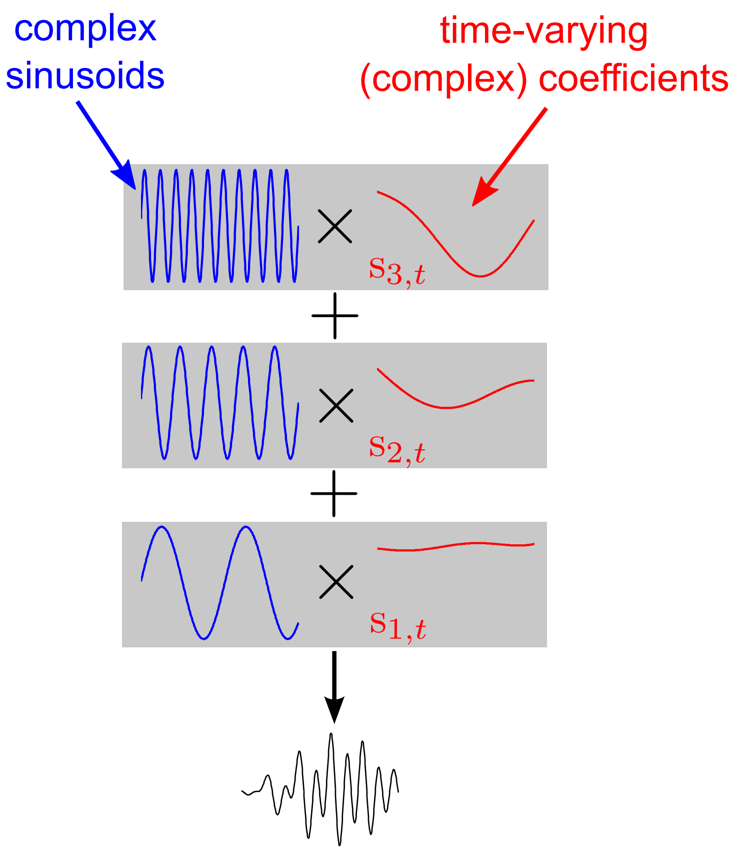
generation



$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

Time-frequency analysis as inference

generation



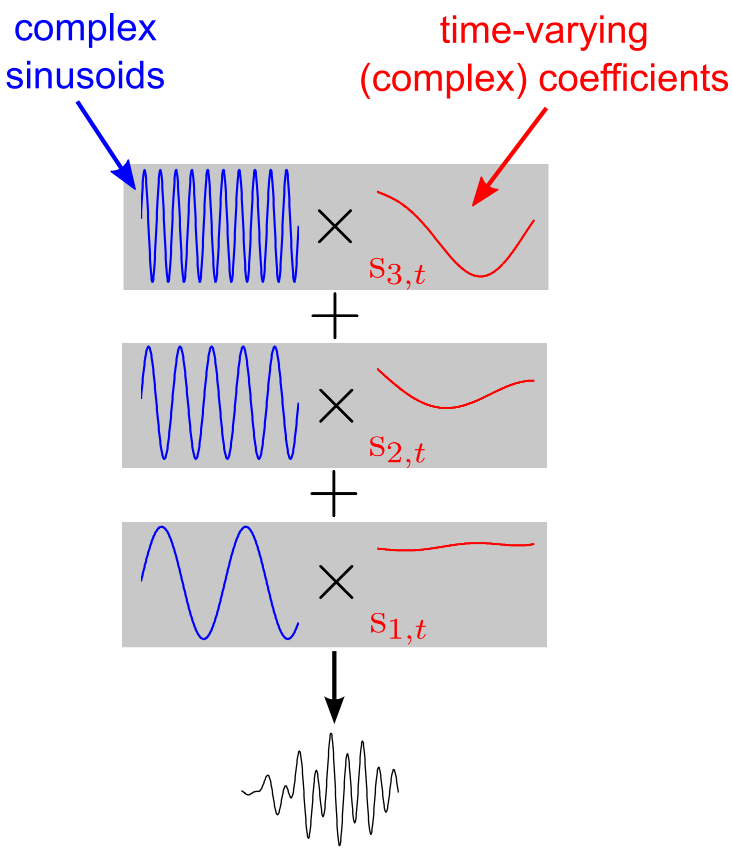
$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

Time-frequency analysis as inference

generation

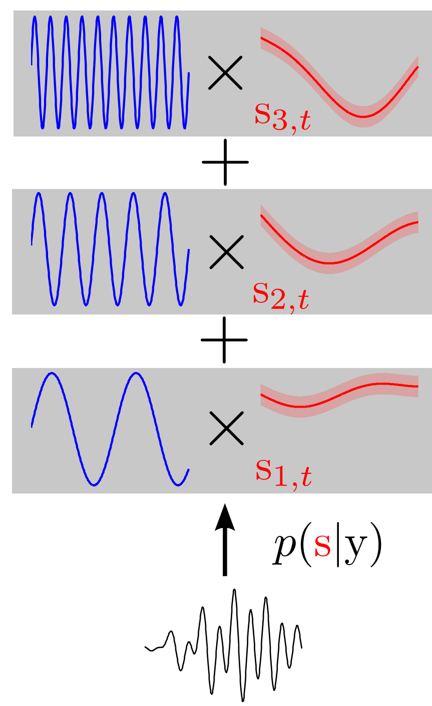


$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

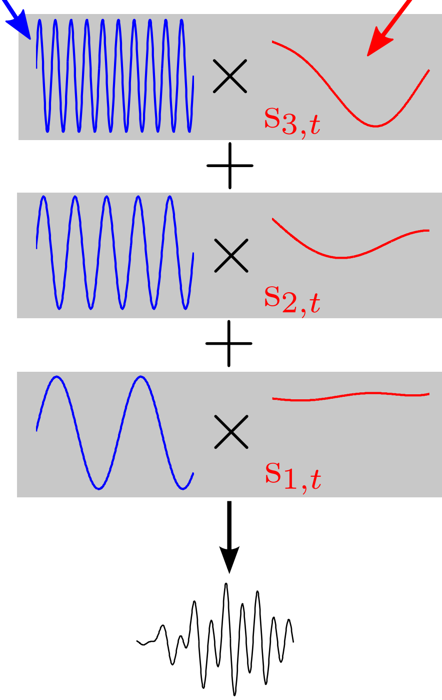


Time-frequency analysis as inference

generation

complex sinusoids

time-varying (complex) coefficients



$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

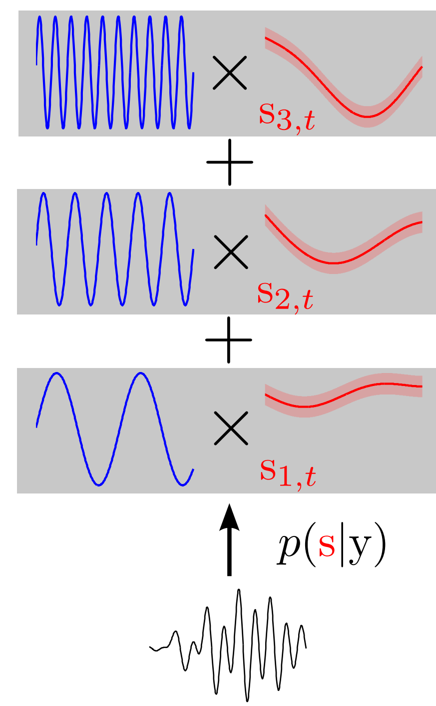
$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

most probable coefficients given the signal is the STFT

$$\text{STFT} \equiv \arg\max_s p(s|y)$$



STFT window = prior covariance

$\otimes$

frequency shifted inverse signal covariance

Time-frequency analysis as inference

generation

inference

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

Time-frequency analysis as inference

generation

inference

$$\boxed{\mathrm{x}_{d,t} = \mathrm{e}^{i\omega_d t}\mathrm{s}_{d,t}}$$

$$\mathrm{y}_t = \sum_d \Re\left(\mathrm{e}^{i\omega_d t}\mathrm{s}_{d,t}\right) + \sigma_{\mathrm{y}}\eta_t$$

$$\Re\left(\mathrm{s}_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(\mathrm{s}_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d(t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d(t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d (t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

depends on $\mathbf{y}$

independent of $\mathbf{y}$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d (t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

depends on $\mathbf{y}$ independent of $\mathbf{y}$

$$y_t = \underbrace{x_{d,t}}_{\text{signal}} + \underbrace{\sum_{d' \neq d} \Re\left(x_{d',t}\right) + \sigma_y \eta_t}_{\text{noise}}$$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d (t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

depends on $\mathbf{y}$ independent of $\mathbf{y}$

$$y_t = \underbrace{x_{d,t}}_{\text{signal}} + \underbrace{\sum_{d' \neq d} \Re\left(x_{d',t}\right) + \sigma_y \eta_t}_{\text{noise}}$$

$$\mu_{x_{d,t}|\mathbf{y}} = \sum_{t'} V_{d,t-t'} y_{t'}$$

Wiener filter

$$\tilde{V}_{d,\omega} = \frac{\gamma_{d,\omega}}{\gamma_{\mathbf{y},\omega}}$$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \Gamma_{t,t'} \cos(\omega_d(t - t'))$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

depends on $\mathbf{y}$ independent of $\mathbf{y}$

$$y_t = \underbrace{x_{d,t}}_{\text{signal}} + \sum_{d' \neq d} \underbrace{\Re\left(x_{d',t}\right) + \sigma_y \eta_t}_{\text{noise}}$$

$$\mu_{x_{d,t}|\mathbf{y}} = \sum_{t'} V_{d,t-t'} y_{t'}$$

Wiener filter

$$\tilde{V}_{d,\omega} = \frac{\gamma_{d,\omega}}{\gamma_{\mathbf{y},\omega}}$$

$$\boxed{\mu_{s_{d,t}|\mathbf{y}} = e^{-i\omega_d t} \mu_{x_{d,t}|\mathbf{y}}}$$

Time-frequency analysis as inference

generation

$$y_t = \sum_d \Re\left(x_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(x_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \boxed{\Gamma_{t,t'} \cos(\omega_d (t - t'))}$$

$$\boxed{x_{d,t} = e^{i\omega_d t} s_{d,t}}$$

$$y_t = \sum_d \Re\left(e^{i\omega_d t} s_{d,t}\right) + \sigma_y \eta_t$$

$$\Re\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(s_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(X|\mathbf{y}) = \mathcal{G}(X; \mu_{X|\mathbf{y}}, \Sigma_{X|\mathbf{y}})$$

depends on $\mathbf{y}$ independent of $\mathbf{y}$

$$y_t = \underbrace{x_{d,t}}_{\text{signal}} + \sum_{d' \neq d} \underbrace{\Re\left(x_{d',t}\right) + \sigma_y \eta_t}_{\text{noise}}$$

$$\mu_{x_{d,t}|\mathbf{y}} = \sum_{t'} V_{d,t-t'} y_{t'}$$

Wiener filter

$$\tilde{V}_{d,\omega} = \frac{\boxed{\gamma_{d,\omega}}}{\gamma_{\mathbf{y},\omega}}$$

$$\boxed{\mu_{s_{d,t}|\mathbf{y}} = e^{-i\omega_d t} \mu_{x_{d,t}|\mathbf{y}}}$$

$$\mu_{s_{d,t}|\mathbf{y}} = \sum_{t'} e^{-i\omega_d t'} W_{t-t'} y_{t'}$$

STFT window = prior covariance $\otimes$ frequency shifted inverse signal covariance

Time-frequency analysis as inference

generation

probabilistic filter bank

$$\mathrm{y}_t = \sum_d \Re\left(\mathrm{x}_{d,t}\right) + \sigma_\mathrm{y}\eta_t$$

$$\Re\left(\mathrm{x}_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Im\left(\mathrm{x}_{d,t}\right) \sim \mathcal{GP}(0, \Sigma_d)$$

$$\Sigma_{d,t,t'} = \boxed{\Gamma_{t,t'} \cos(\omega_d(t-t'))}$$

$$\mathrm{x}_{d,t} = \mathrm{e}^{i\omega_d t}\mathrm{s}_{d,t}$$

probabilistic STFT

$$\mathrm{y}_t = \sum_d \Re\left(\mathrm{e}^{i\omega_d t}\mathrm{s}_{d,t}\right) + \sigma_\mathrm{y}\eta_t$$

$$\Re\left(\mathrm{s}_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

$$\Im\left(\mathrm{s}_{d,t}\right) \sim \mathcal{GP}(0, \Gamma)$$

inference

$$p(\mathrm{X}|\mathbf{y}) = \mathcal{G}(\mathrm{X}; \mu_{\mathrm{X}|\mathbf{y}}, \Sigma_{\mathrm{X}|\mathbf{y}})$$

depends on $\mathbf{y}$ independent of $\mathbf{y}$

$$\mathrm{y}_t = \underbrace{\mathrm{x}_{d,t}}_{\text{signal}} + \underbrace{\sum_{d' \neq d} \Re\left(\mathrm{x}_{d',t}\right) + \sigma_\mathrm{y}\eta_t}_{\text{noise}}$$

$$\mu_{\mathrm{x}_{d,t}|\mathbf{y}} = \sum_{t'} \mathrm{V}_{d,t-t'}\mathrm{y}_{t'}$$

Wiener filter

$$\tilde{\mathrm{V}}_{d,\omega} = \frac{\boxed{\gamma_{d,\omega}}}{\gamma_{\mathbf{y},\omega}}$$

$$\mu_{\mathrm{s}_{d,t}|\mathbf{y}} = \mathrm{e}^{-i\omega_d t}\mu_{\mathrm{x}_{d,t}|\mathbf{y}}$$

$$\mu_{\mathrm{s}_{d,t}|\mathbf{y}} = \sum_{t'} e^{-i\omega_d t'}\mathrm{W}_{t-t'}\mathrm{y}_{t'}$$

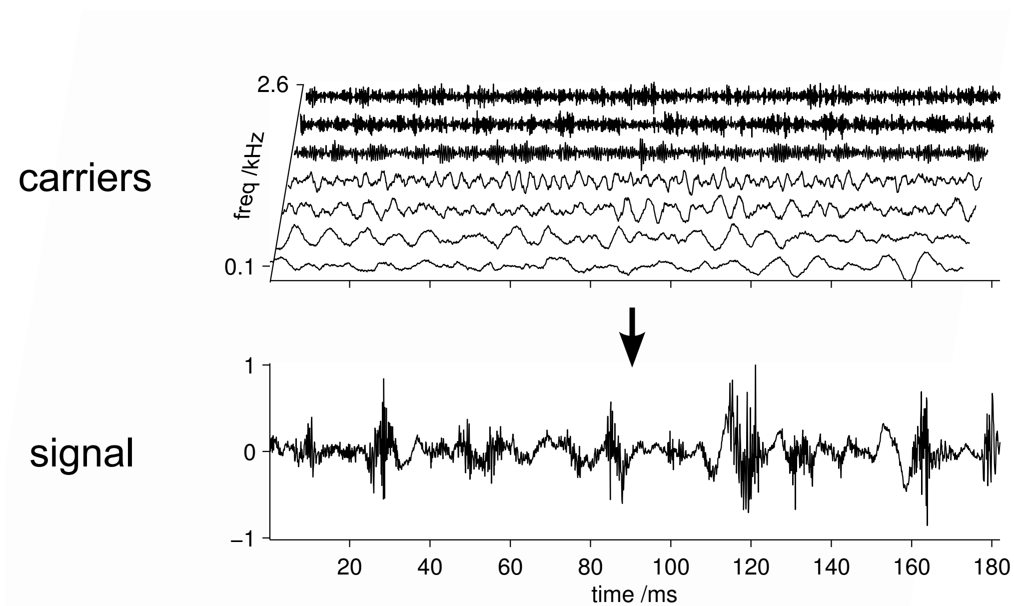
STFT window = prior covariance $\otimes$
frequency shifted inverse signal covariance

Time-frequency analysis as inference

- probabilistic models in which **inference recovers STFT, filter bank, wavelet analysis**
  - **unifes a number of existing probabilistic time-series models** & connects to traditional sig. proc.
  - can learn window of STFT and frequencies (equivalently filter properties)
  - frequency shift relationship mimics classical relationship between these time-frequency relationships

- **hops/down-sampling and finite window used correspond to FITC** (uniformly spaced pseudo-points) and sparse-covariance approximations
  - rediscover Nyquist in the context of approximation GPs

Probabilistic audio processing pipeline



mean spectrum

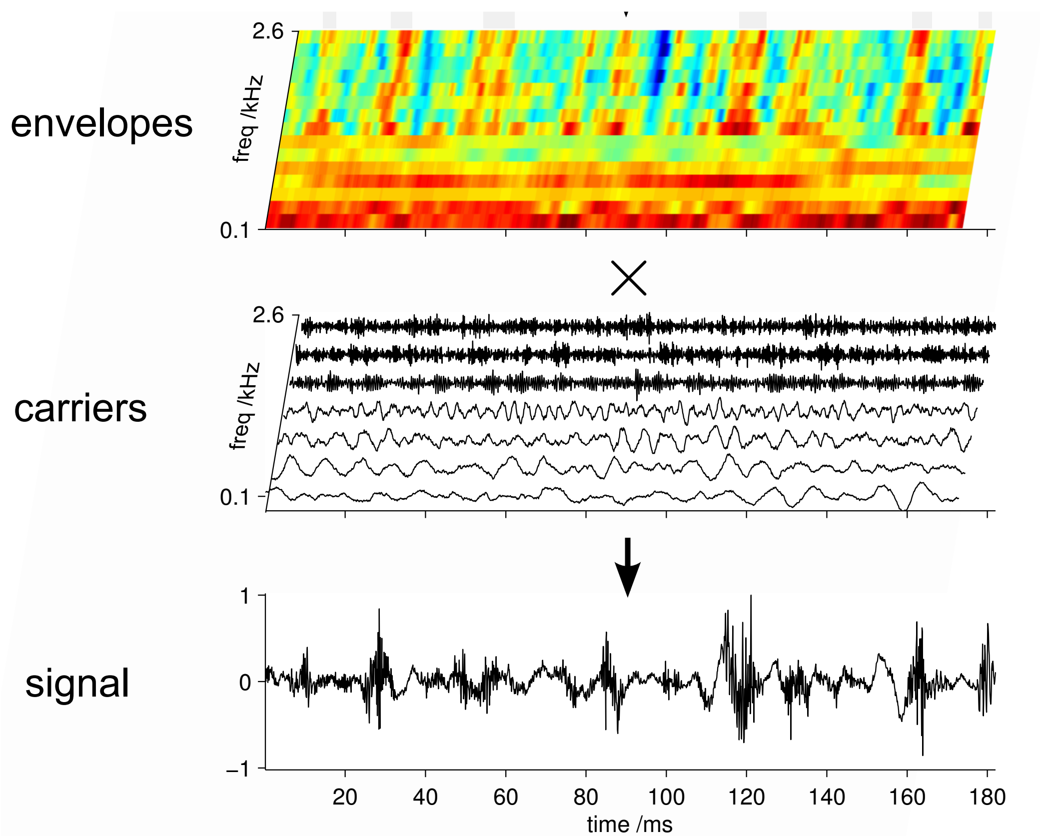
$$\mathsf{c}_d(t) \sim \mathrm{GP}\left(0, \ \cdot \ \right)$$

0 f

= bandpass Gaussian noise

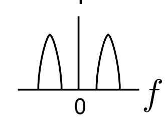
$$\mathsf{y}(t) = \sum_{d=1}^{D} \Re\left(\mathsf{x}_d(t)\right)$$

Probabilistic audio processing pipeline



$a_{d,t}$

mean spectrum

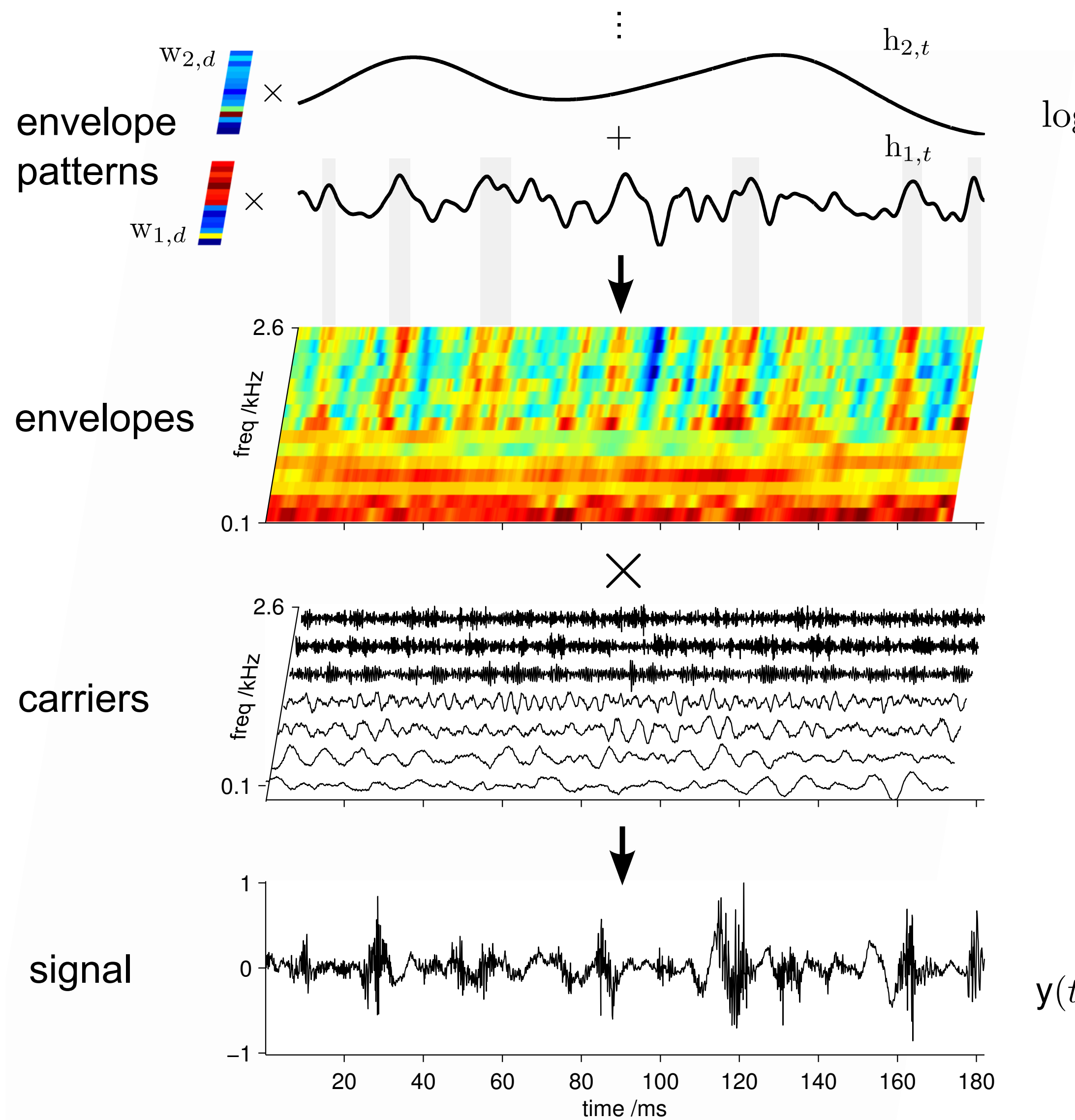
$c_d(t) \sim \mathrm{GP}\Big(0,$ 

$\Big)$

= bandpass Gaussian noise

$$y(t) = \sum_{d=1}^{D} \Re\left(x_d(t)\right) a_d(t)$$

Probabilistic audio processing pipeline



mean spectrum

$$\log h_{l,t} \sim \text{GP}\left(\mu_k, \; 0 \; f\right)$$

= slow Gaussian process

$$a_{d,t} = \sum_{l=1}^{L} h_{l,t} w_{l,d}$$

mean spectrum

$$c_d(t) \sim \text{GP}\left(0, \; 0 \; f\right)$$

= bandpass Gaussian noise

$$y(t) = \sum_{d=1}^{D} \Re\left(x_d(t)\right) a_d(t)$$

Inference and Learning

- **Key Observation** – fix envelopes:
  - posterior over carriers is Gaussian
  - posterior mean given by an (adaptive) filter

- Leads to MAP estimation of the envelopes (or HMCMC), let $\mathrm{z}_{lt} = \log \mathrm{h}_{lt}$

$$\mathrm{Z}^{\mathsf{MAP}} = \arg\max_{\mathrm{Z}} p(\mathrm{Z}|\mathrm{Y})$$

$$p(\mathrm{Z}|\mathrm{Y}) = \frac{1}{Z} p(\mathrm{Z},\mathrm{Y}) = \frac{1}{Z} \int d\mathrm{X} p(\mathrm{Z},\mathrm{Y},\mathrm{X}) = \frac{1}{Z} p(\mathrm{Z}) \int d\mathrm{X} p(\mathrm{Y}|\mathrm{A},\mathrm{X}) p(\mathrm{X})$$

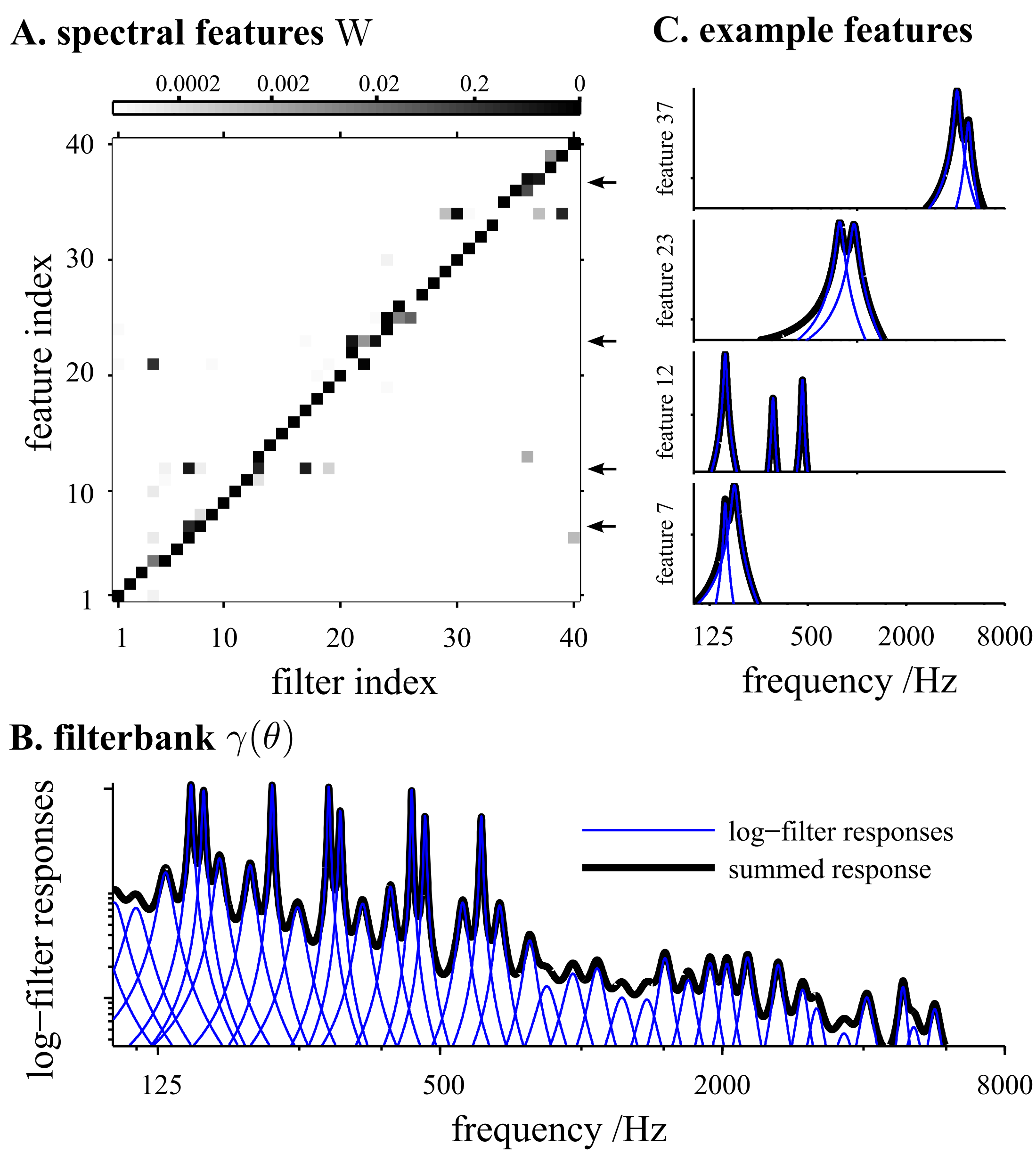
- Compute integral efficiently using **chain stuctured approximation** and **Kalman Smoothing**

- Leads to gradient based optimisation for transformed amplitudes

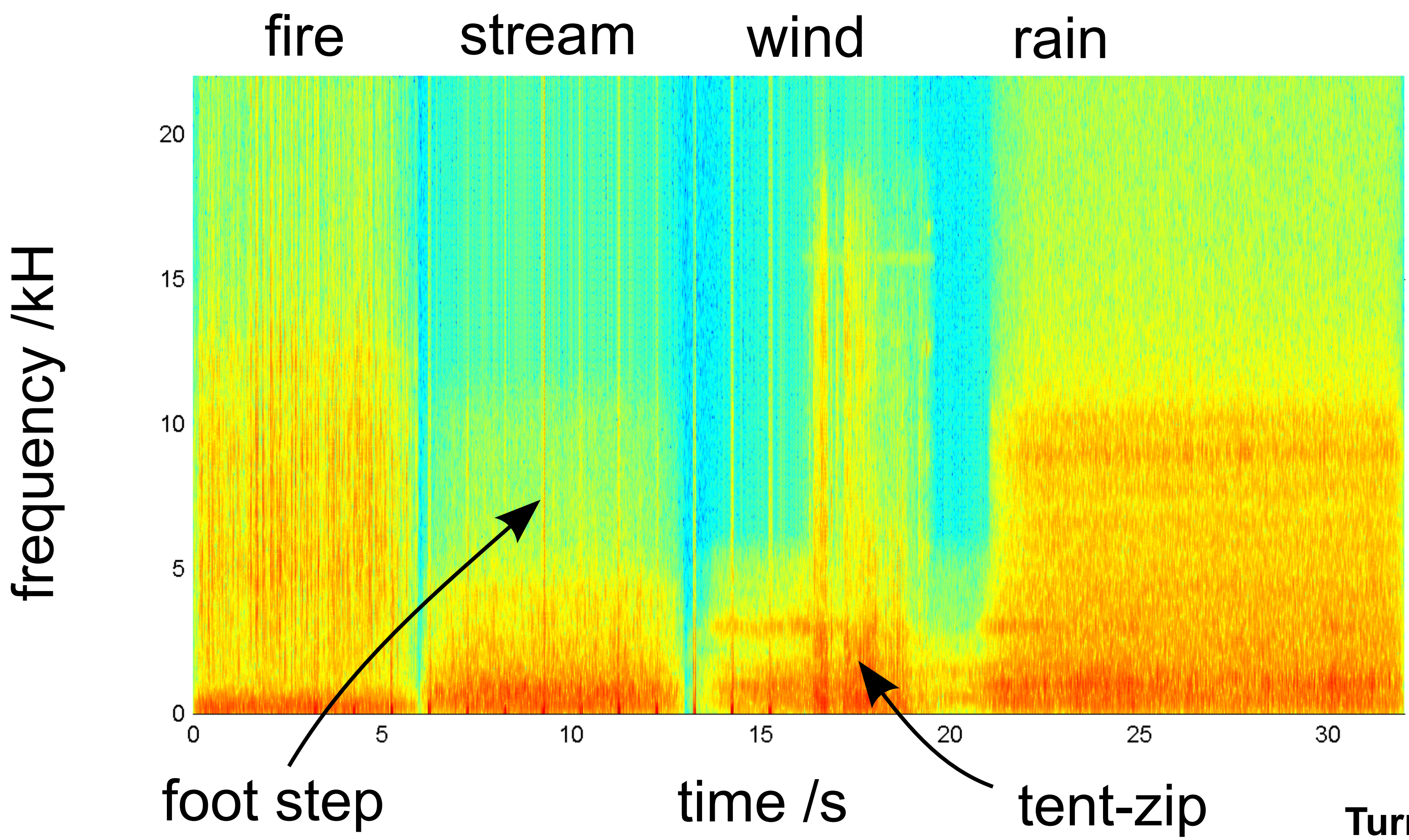
- Learning: approximate Maximum Likelihood $\theta = \arg\max_{\theta} p(\mathrm{Y}|\theta)$

- **NMF**: zero-temperature EM, one E-Step, initialise constant envelopes

Audio modelling

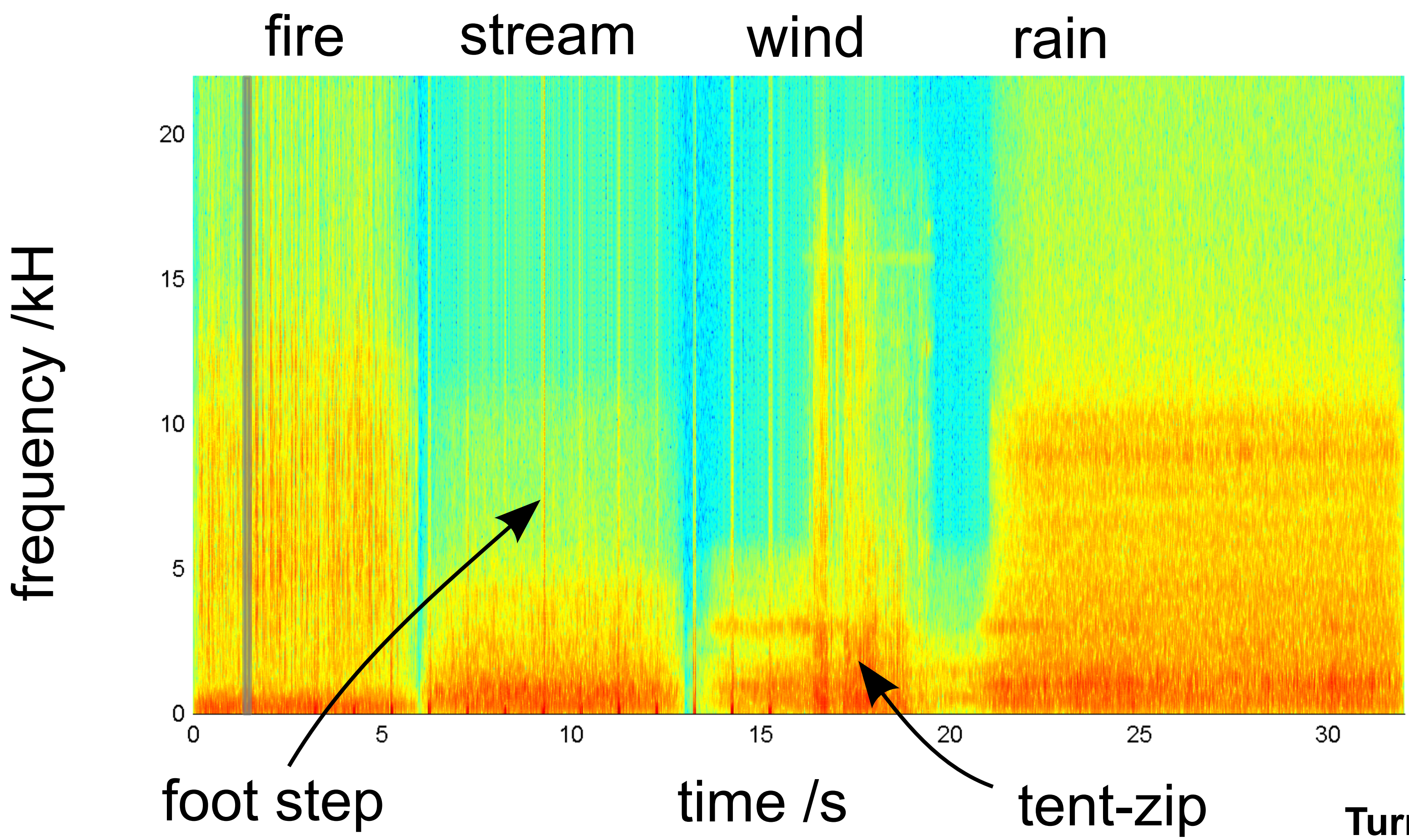


Audio modelling



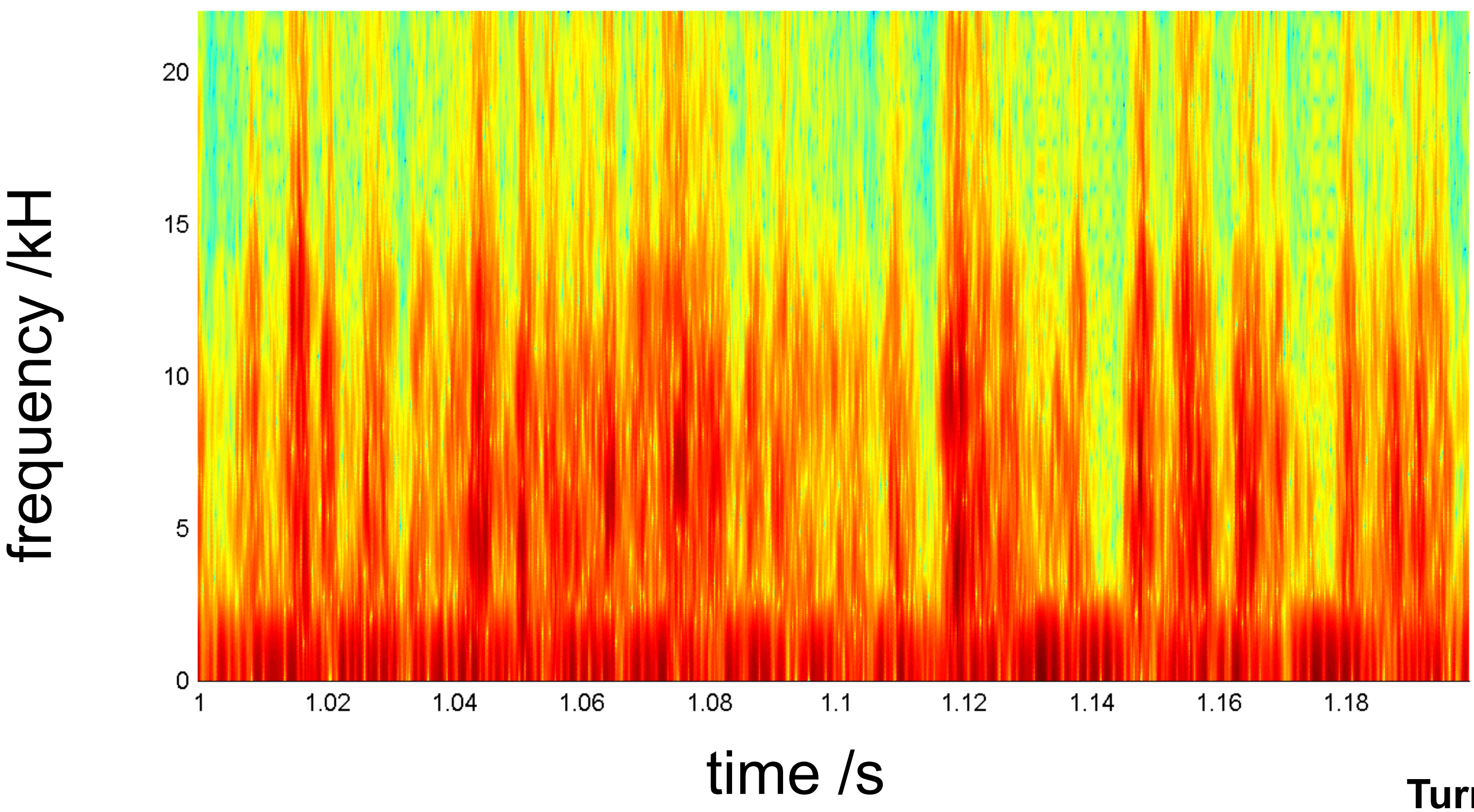
Turner, 2010

Audio modelling



Turner, 2010

Audio modelling

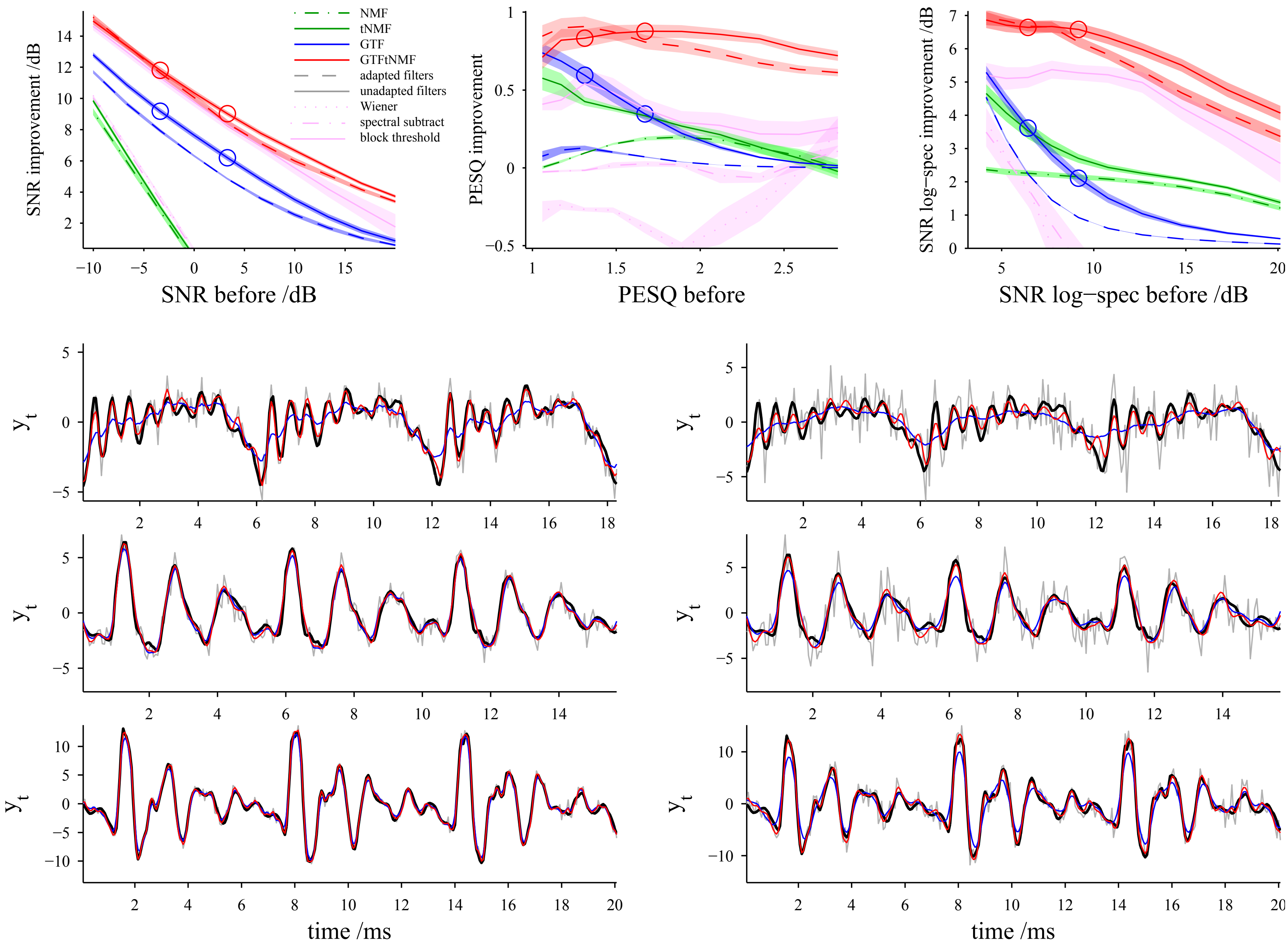


Turner, 2010

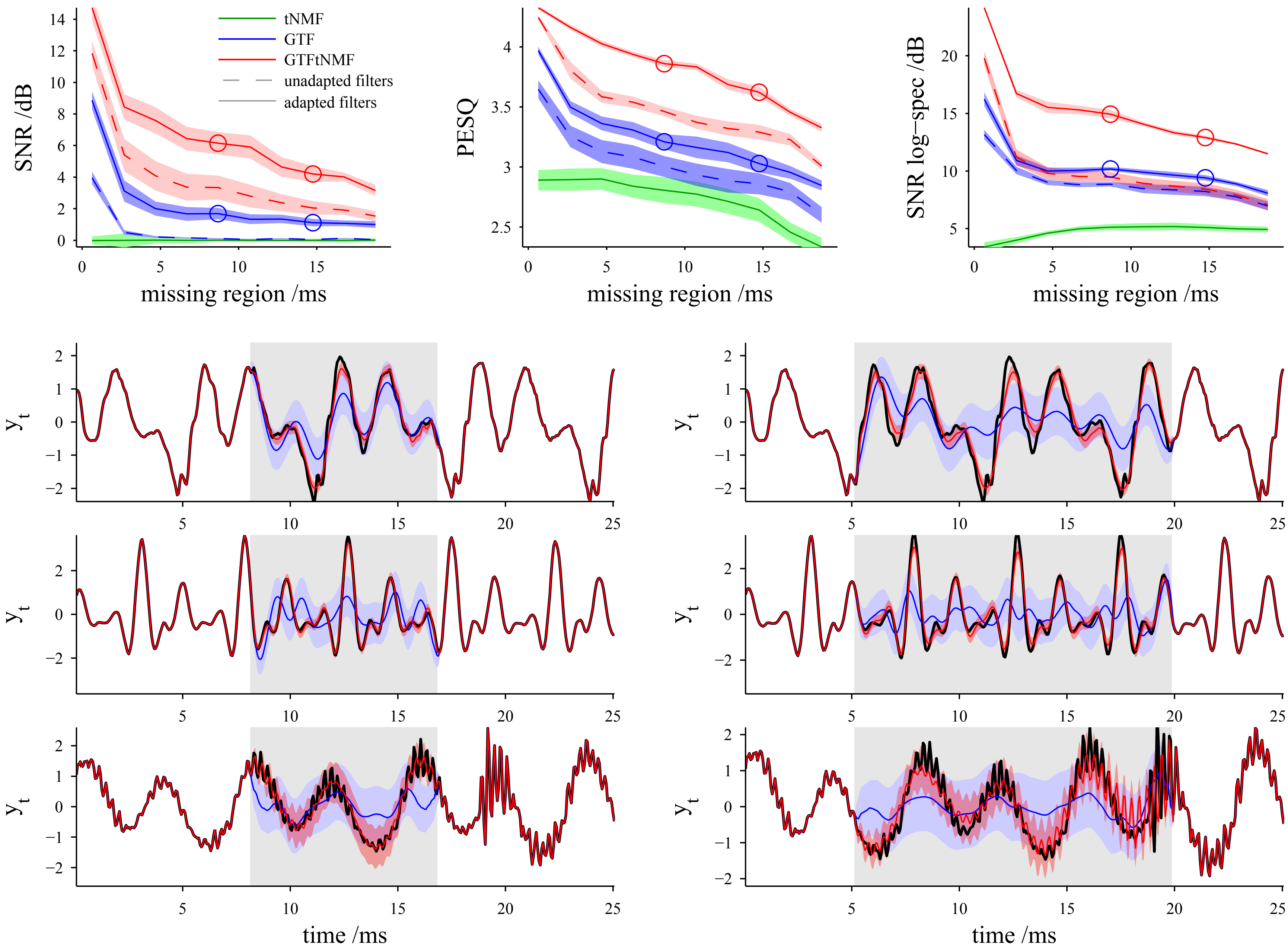
Statistical texture synthesis

- Old approach: build **detailed physical models** (e.g. rain drops)
- New approach
  - **train model** on your favourite texture
  - **sample** from the prior, and then from the likelihood.
- Waveform unique, but statistically matched to original
- Often perceptually indistinguishable

Audio denoising



Audio missing data imputation



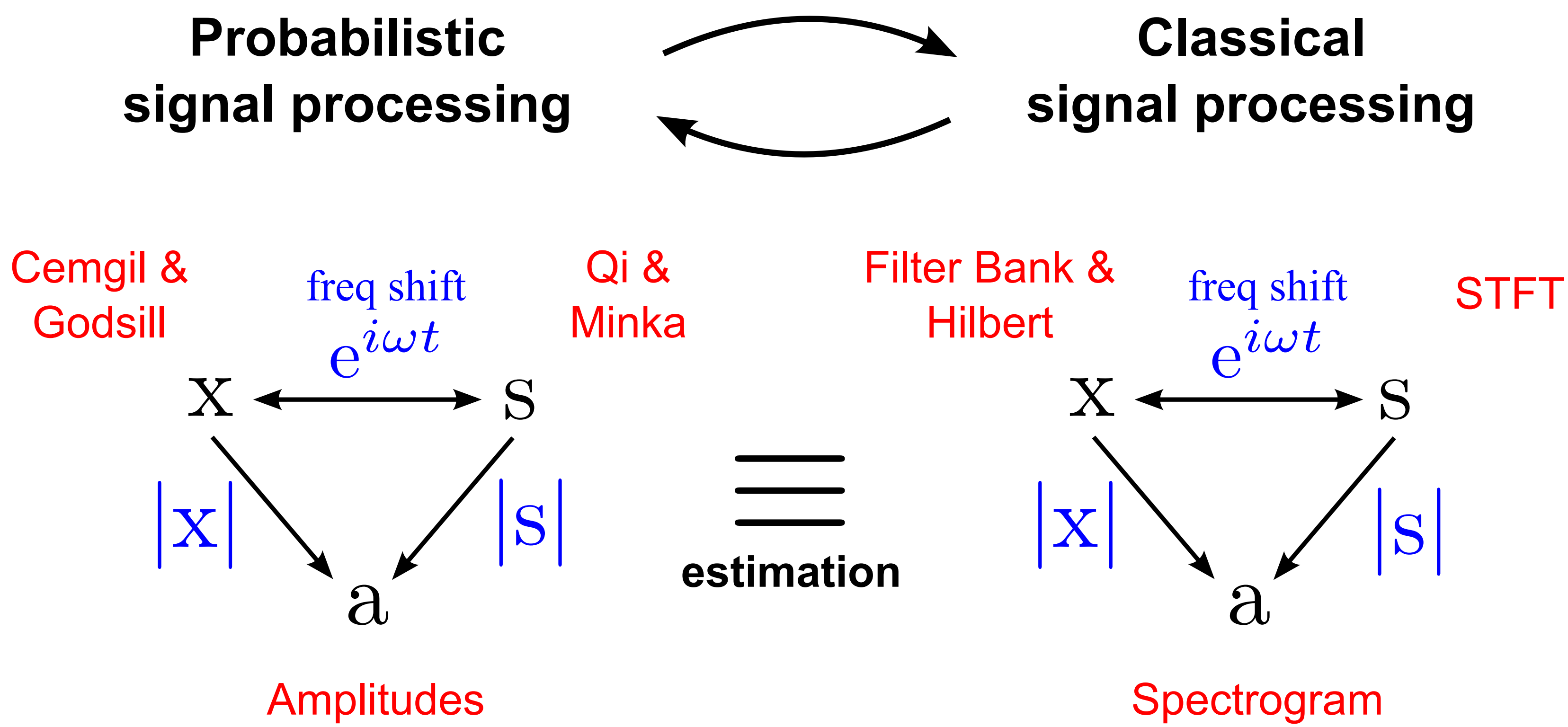
Unifying classical and probabilistic audio signal processing

robustness
adaptation

Probabilistic signal processing

Classical signal processing

fast methods
important variables


Probabilistic signal processing
Classical signal processing
Cemgil & Godsill
freq shift
$e^{i\omega t}$
Qi & Minka
x
s
|x|
|s|
a
Amplitudes
estimation
Filter Bank & Hilbert
freq shift
$e^{i\omega t}$
STFT
x
s
|x|
|s|
a
Spectrogram

Additional slides