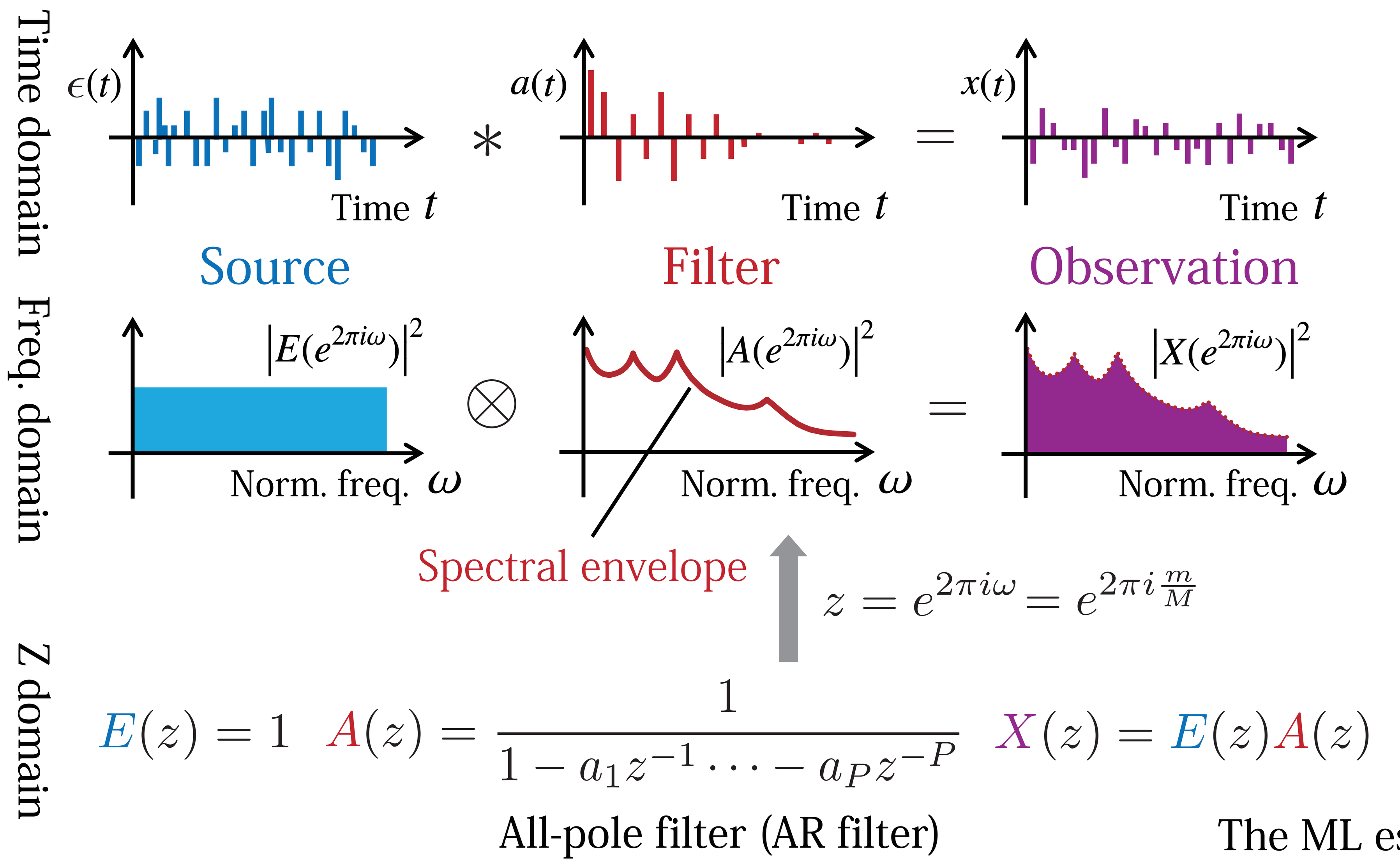
Infinite Kernel Linear Prediction for Joint Estimation of Spectral Envelope and Fundamental Frequency

Kazuyoshi Yoshii Masataka Goto (AIST, Japan)

Objective: Estimate the correct spectral envelope of an observed audio signal

Linear Prediction (LP): A probabilistic model that assumes the observed signal to follow an autoregressive (AR) process



$$x_m = \sum_{p=1}^{P} a_p x_{m-p} + \epsilon_m \Rightarrow \boldsymbol{\epsilon} = \boldsymbol{\Psi}\boldsymbol{x} \Rightarrow \boldsymbol{x} = \boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon}$$

$\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_M)^T$ Source signal $\boldsymbol{a} = (a_1, \cdots, a_P)^T$ Filter coefficients $\boldsymbol{x} = (x_1, \cdots, x_M)^T$ Observed signal

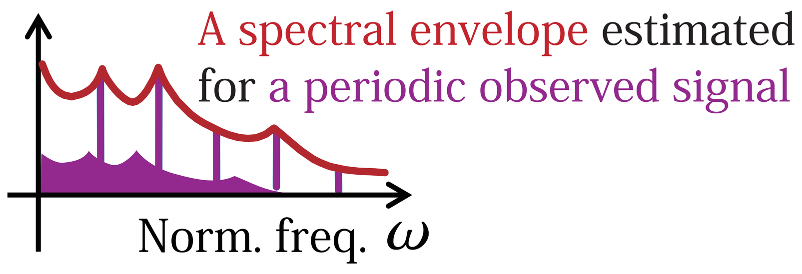
$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & & & & \mathbf{0} \\ -a_1 & \ddots & & & \\ \vdots & \ddots & \ddots & & \\ -a_P & & \ddots & \ddots & \\ \mathbf{0} & \ddots & -a_P & \cdots -a_1 & 1 \end{bmatrix}_{M\times M} \quad \boldsymbol{X} = \begin{bmatrix} 0 & \cdots & 0 \\ x_1 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ \vdots & & x_1 \\ \vdots & & \vdots \\ x_{M-1} & \cdots & x_{M-P} \end{bmatrix}_{M\times P}$$

If the source signal is white Gaussian noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \nu\boldsymbol{I})$
the observed signal follows a Gaussian $\boldsymbol{x} \sim \mathcal{N}(0, \nu\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}^{-T})$

The ML estimate of $\boldsymbol{a}$ is given by solving a normal equation $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{x}$

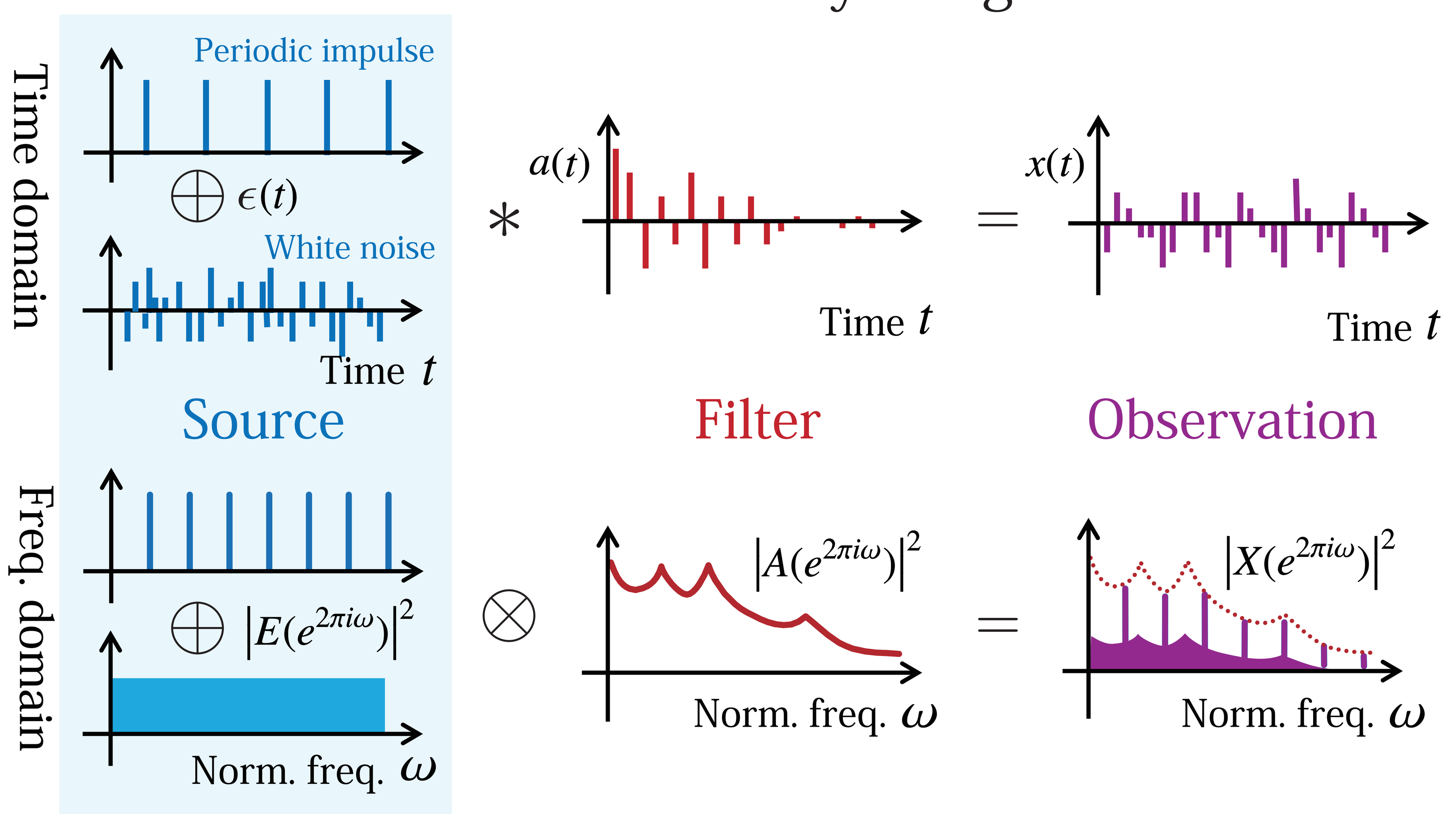
Problem: The white-Gaussian assumption is violated if the observed signal is periodic

The estimated spectral envelope has unnecessary sharp peaks at harmonic partials



Approach: Jointly estimate a fundamental frequency and a spectral envelope

Infinite Kernel LP (IKLP): A probabilistic model that represents the periodicity of a source signal by using a convex combination of infinitely many kernels



Conventional: Multiple Kernel LP (MKLP) [Kameoka2010]

The source signal is precisely modeled by using a Gaussian process (GP)

$$\epsilon(t) = \sum_{j=1}^{J} w_j \phi_j(t) + \eta(t) = \boldsymbol{\phi}(t)^T\boldsymbol{w} + \eta(t) \Rightarrow \boldsymbol{\epsilon} = \underset{M\times J}{\boldsymbol{\Phi}}\underset{J\times 1}{\boldsymbol{w}} + \boldsymbol{\eta}$$

Weighted sum of basis functions White noise M observed points

If we assume $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \nu_w\boldsymbol{I})$ and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \nu_e\boldsymbol{I})$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \nu_w\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \nu_e\boldsymbol{I}) \xrightarrow{\text{Kernelize}} \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \nu_w\boldsymbol{K} + \nu_e\boldsymbol{I})$$

Linear regression model GP regression model

The periodicity parameter of $\boldsymbol{K}$ is unknown → Multiple Kernel Learning

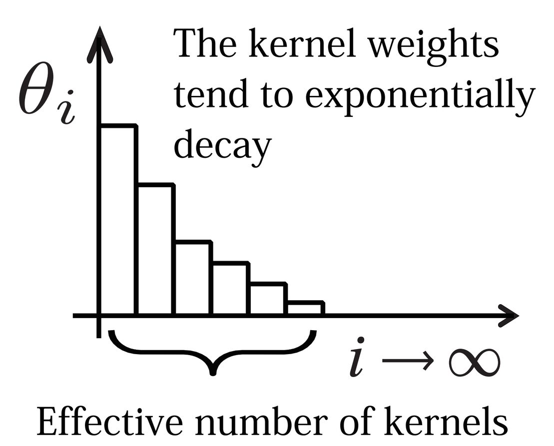
$$\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^{-1}(\nu_w\boldsymbol{K} + \nu_e\boldsymbol{I})\boldsymbol{\Psi}^{-T}) \quad \boldsymbol{K} = \sum_{i=1}^{I}\theta_i\boldsymbol{K}_i$$

Periodic compo. White noise
Includes a conventional LP as a special case
A convex combination of many kernels corresponding to different F0s

$$\boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}^{-1}\left(\nu_w\sum_{i=1}^{I\to\infty}\theta_i\boldsymbol{K}_i + \nu_e\boldsymbol{I}\right)\boldsymbol{\Psi}^{-T}\right)$$ Likelihood by IKLP!

Proposed: Nonparametric Bayesian Kernel Learning

A gamma process prior is put on infinitely many kernel weights



$\boldsymbol{\theta} \sim \mathrm{GaP}(\alpha, \mathrm{Uniform})$

⬇ Truncate at a sufficiently large level

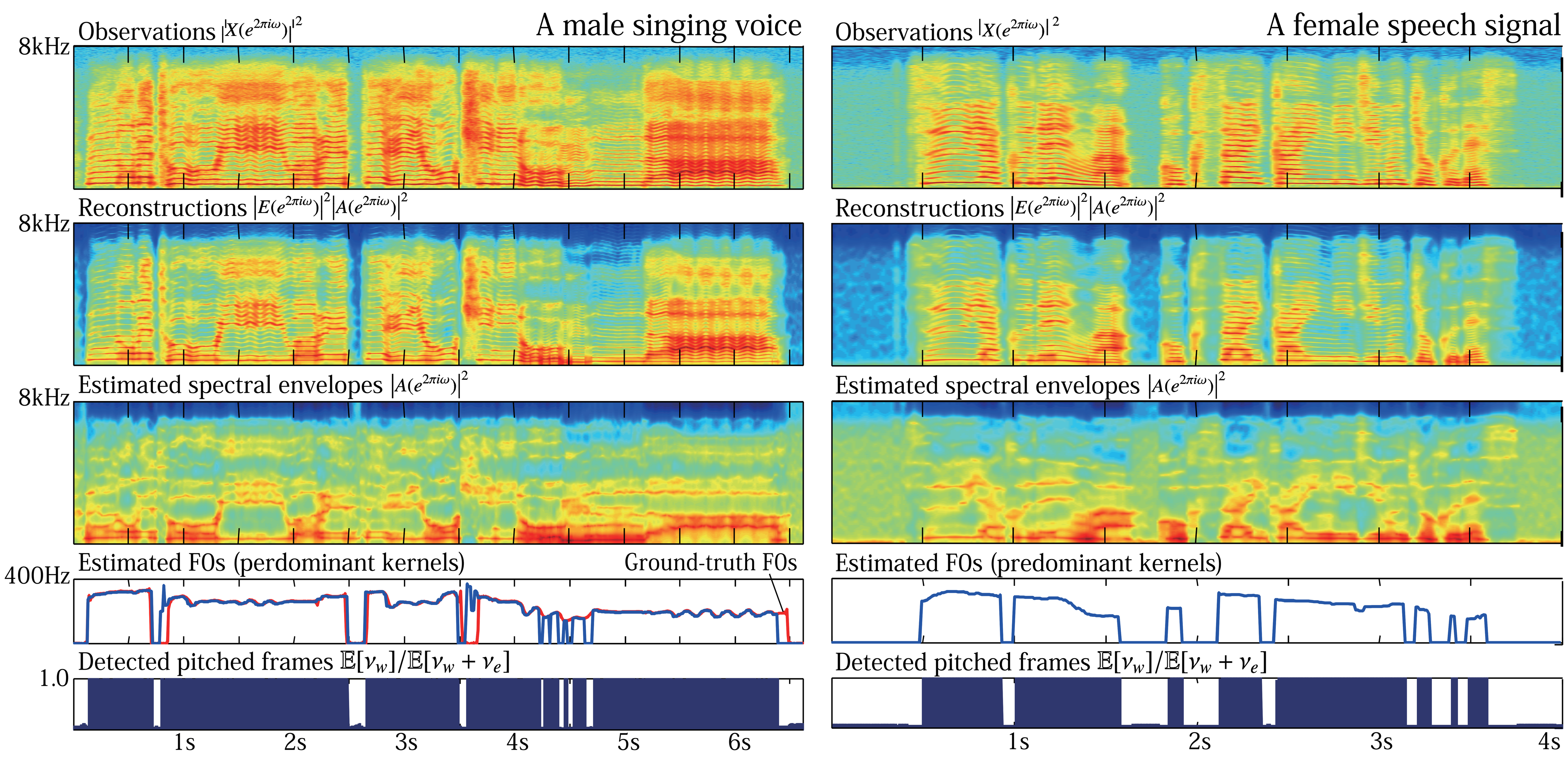
$\theta_i \sim \mathrm{Gamma}(\alpha/I, \alpha)$

α : concentration parameter controlling the tail heaviness

We also put priors on other unknown variables as follows:

$\nu_w \sim \mathrm{Gamma}(a_w, b_w)$ $\nu_e \sim \mathrm{Gamma}(a_e, b_e)$ $\boldsymbol{a} \sim \mathcal{N}(\mathbf{0}, \lambda\boldsymbol{I})$

We derived a variational Bayesian (VB) algorithm for closed-form parameter optimization
This algorithm can be viewed as a new efficient solution of multiple kernel learning



Conclusion

We proposed a nonparametric Bayesian model that represents the periodicity of a source signal by using infinitely many kernels

The joint estimation of a F0 and a spectral envelope was achieved in a principled way

Future Work

We plan to extend the model such that it can deal with not only infinitely many sources but also have infinitely many filters for timbre-based separation of music signals
→ Extension of infinite composite autoregressive models [Yoshii2012]