Infinite Superimposed Discrete All-pole Modeling for Multipitch Analysis of Wavelet Spectrograms

Kazuyoshi Yoshii[1] Katsutoshi Itoyama[1] Masataka Goto[2]

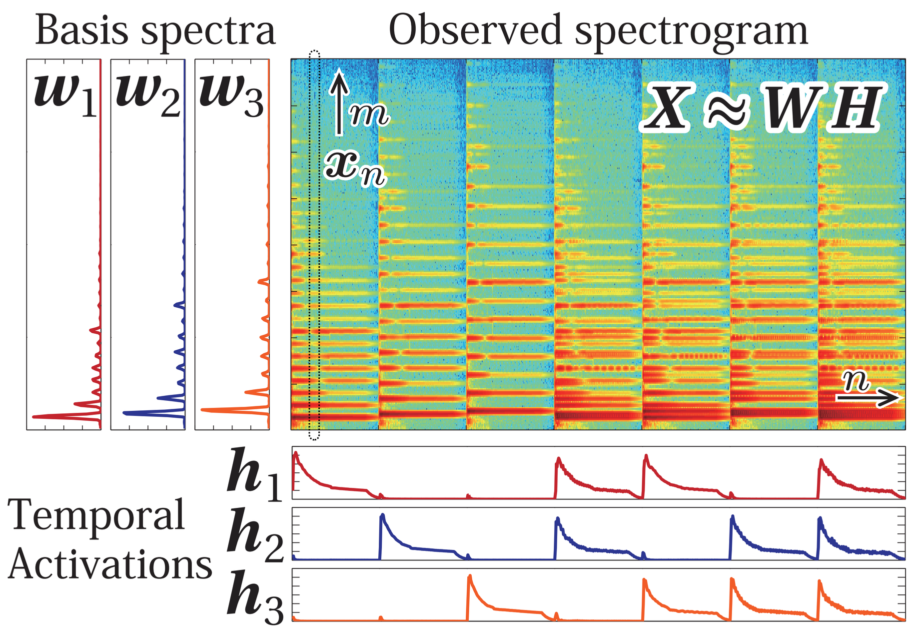
[1]Graduate School of Informatics, Kyoto University, Japan

[2]National Institute of Advanced Industrial Science and Technology (AIST), Japan

Conventional Statistical Modeling of Linear-frequency Spectrograms

Probabilistic models of parts-based representation and spectral envelope estimation have been proposed

Nonnegative Matrix Factorization (NMF)



Each local spectrum is approximated by a weighted sum of basis spectra

$$x_n \approx \sum_{k=1}^{K} w_k h_{kn} \overset{\text{def}}{=} y_n$$

Observation: $X = [x_1, \cdots, x_N] \in \mathbb{R}^{M \times N}$

Basis: $W = [w_1, \cdots, w_K] \in \mathbb{R}^{M \times K}$

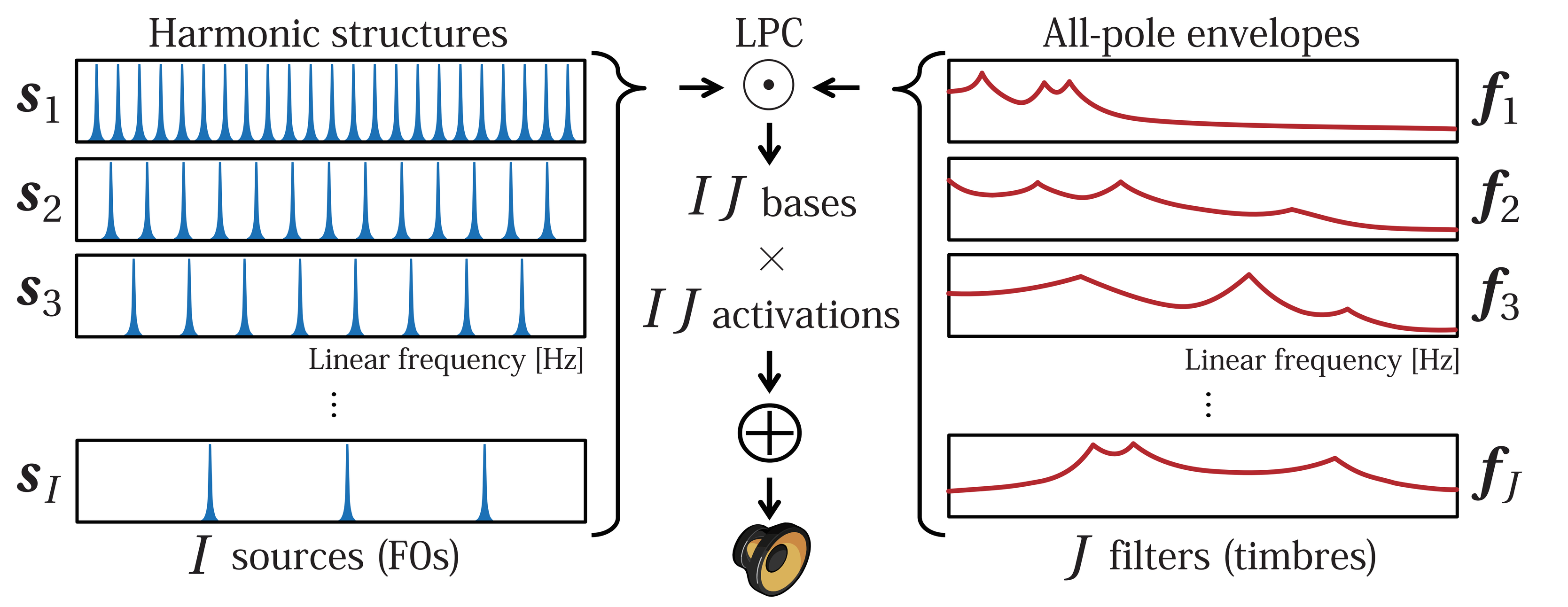
Activation: $H = [h_1, \cdots, h_K]^T \in \mathbb{R}^{K \times N}$

Pros: A probabilisic model can be formulated for maximum likelihood estimation

The number of basis spectra can be automatically adjusted to the observed data (gamma process NMF: GaP-NMF) [Hoffman 2010]

Cons: It is hard to cluter basis spectra into instrument parts $x_{mn} \sim \text{Poisson}(y_{mn})$

Composite Autoregressive Modeling (CAR)



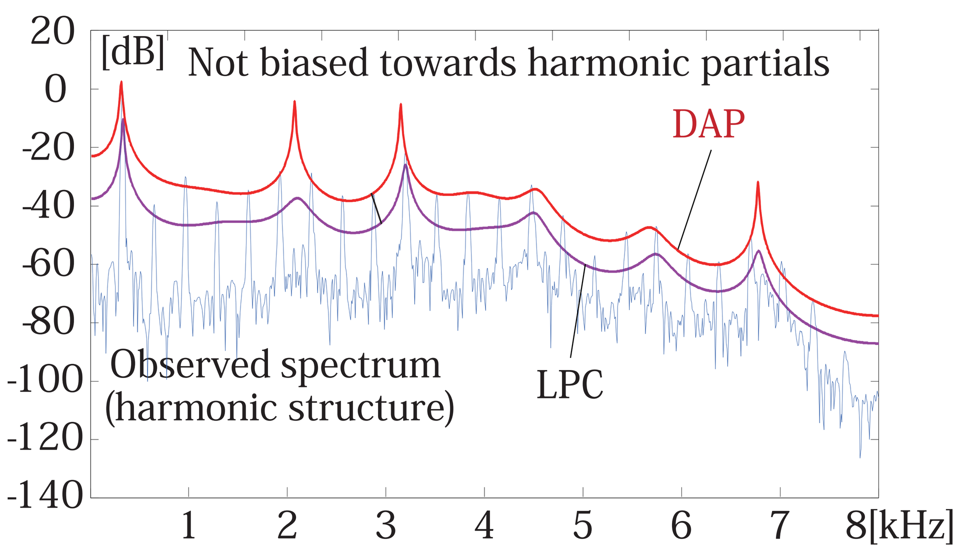
Each local spectram is approximated by combinations of sources and filters

$$x_n \approx \sum_{i=1}^{I} \sum_{j=1}^{J} (s_i \odot f_j) h_{ijn} \overset{\text{def}}{=} y_n$$

(Element-wize product)

Both multipitch estimation and instrtument-part separation can be performed jointly in a unified framework [Yoshii 2012]

Linear Predictive Coding (LPC) and Discrete All-pole Modeling (DAP)



Input: X_m: Observed power spectrum → Output: F_m: Spectral envelope, σ^2: Gain ($\mathbb{E}[X_m] = \mathbb{E}[\sigma^2 F_m]$)

LPC aims to maximize the likelihood function given by

$X_m \sim \text{Exponential}(\sigma^2 F_m)$ $m \in \{1, 2, \cdots, M\}$ (all frequency bins)

DAP aims to maximize the likelihood function given by

$X_m \sim \text{Exponential}(\sigma^2 F_m)$ $m \in \Omega$ (frequency bins of harmonic partials)

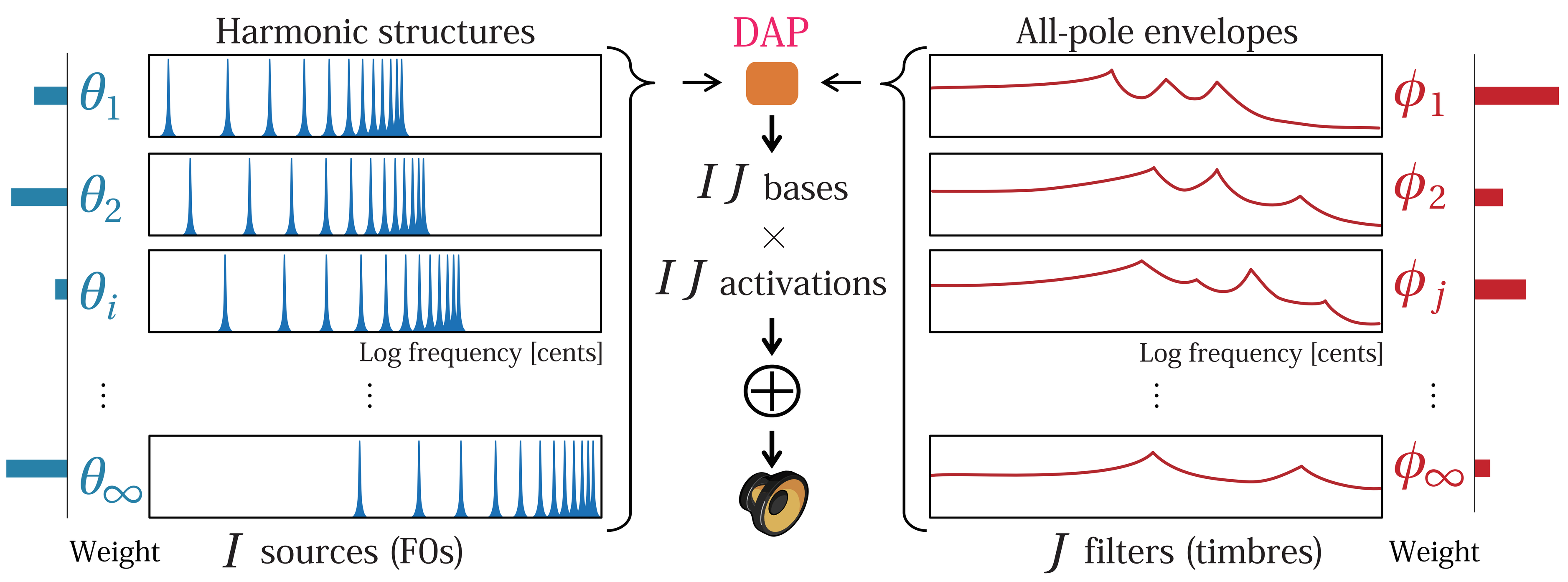
DAP can be used in the linear-frequency domain or in the log-frequency domain

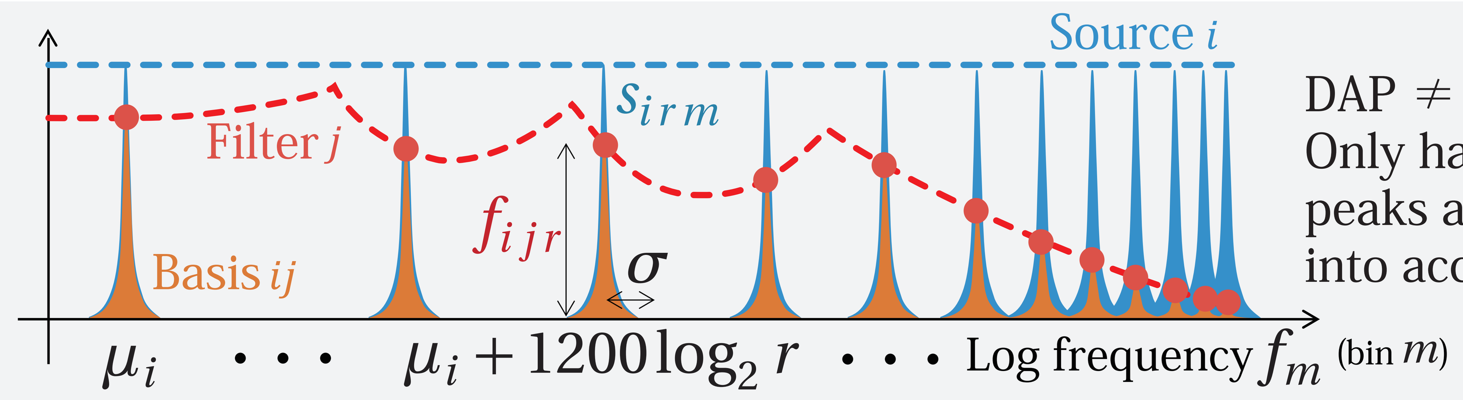
	CAR	DAP
Capable of dealing with superimposed spectra?	✔	
Capable of estimating sources (F0s)?	✔	F0s must be given in advance
Capable of estimating filters (envelopes)?	✔	✔
Can be used in the log-frequency domain?		✔

Source-filter Decomposition of Log-frequency Spectrograms

Suitable to multipitch analysis

We propose a new variant of source-filter NMF by complementing CAR with DAP in the log-frequency domain





DAP ≠ LPC: Only harmonic peaks are taken into account

Likelihood function for the observed spectrogram

$$x_{mn} \sim \text{Poisson}\left(\sum_{i=1}^{I \to \infty} \sum_{j=1}^{J \to \infty} \theta_i \phi_j w_{ijm} h_{ijn} \right)$$

Weighted sum of harmonic partials

$$w_{ijm} = \sum_{r=1}^{R} s_{irm} f_{ijr}$$

Harmonic partial (Gaussian function)

$$s_{irm} = \exp\left(-\frac{1}{2\sigma^2} \left(f_m - (\mu_i + 1200 \log_2 r) \right)^2 \right)$$

Weight (all-pole filter gain)

$$f_{ijr} = \frac{1}{\left| \sum_{p=0}^{P} a_{jp} e^{-p \omega_{ir} \sqrt{-1}} \right|}$$

The posterior distribution of the parameters is approximated by variatinal Bayes (VB)

Gamma process priors are put on the weights of sources and filters (infinite-dimensional nonnegative vectors)

$\theta_i \sim \text{Gamma}\left(\frac{\alpha_\theta}{I}, \alpha_\theta \right)$ $\phi_j \sim \text{Gamma}\left(\frac{\alpha_\phi}{J}, \alpha_\phi \right)$

Only a limited and finite number of elements take non-zero values almost surely

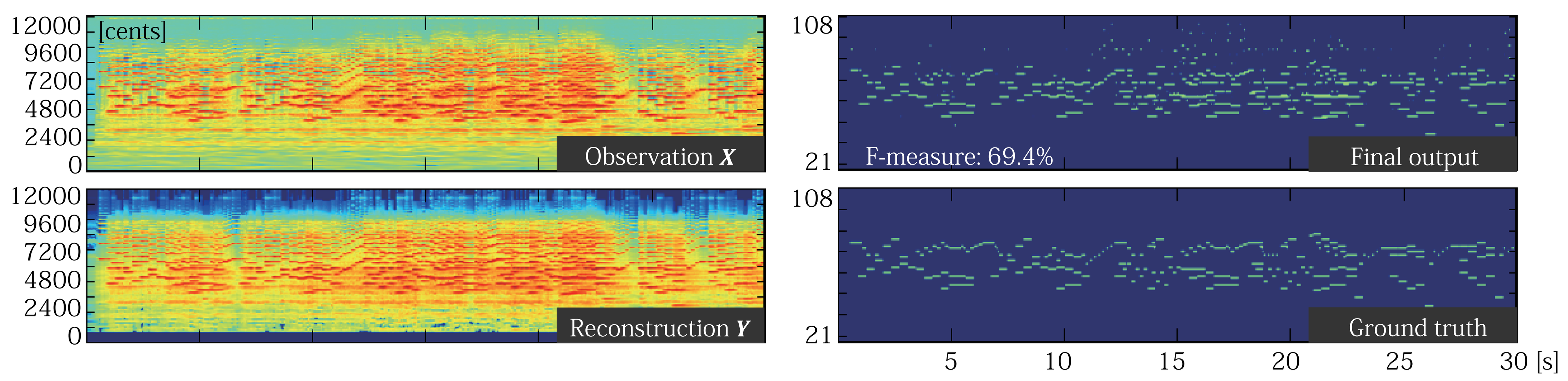
Gamma priors are put on temporal activations

$h_{ijn} \sim \text{Gamma}(a_h, b_h)$

The activation matrix tends to be sparse

Evaluation of Multipitch Estimation on MAPS Database

The proposed model was tested for multipitch analysis of piano recordings (mono-instrument music signals)



Filter learning	HPSS	HMM	$\mathcal{R}$	$\mathcal{P}$	$\mathcal{F}$
Unsupervised			55.3	57.9	56.6
		✓	62.2	60.2	61.2
	✓		62.4	64.3	63.4
	✓	✓	**67.4**	**64.2**	**65.8**
Supervised (open test)	✓		62.4	67.0	64.4
	✓	✓	69.9	64.5	67.3
Supervised (close test)	✓		59.4	69.1	63.9
	✓	✓	67.4	67.8	67.6

A piano timbre (spectral envelope) can be trained in advance

Since the proposed model can deal with only harmonic sounds, HPSS was used as preprocessing

To improve the performance, HMM smoothing was used instead of naive thresholding

The proposed model attained the promising results even if the model was used in a completely unsupervised setting