

音楽音響信号解析のための ガンマ過程に基づく無限重畳離散全極モデル

吉井 和佳
京都大学

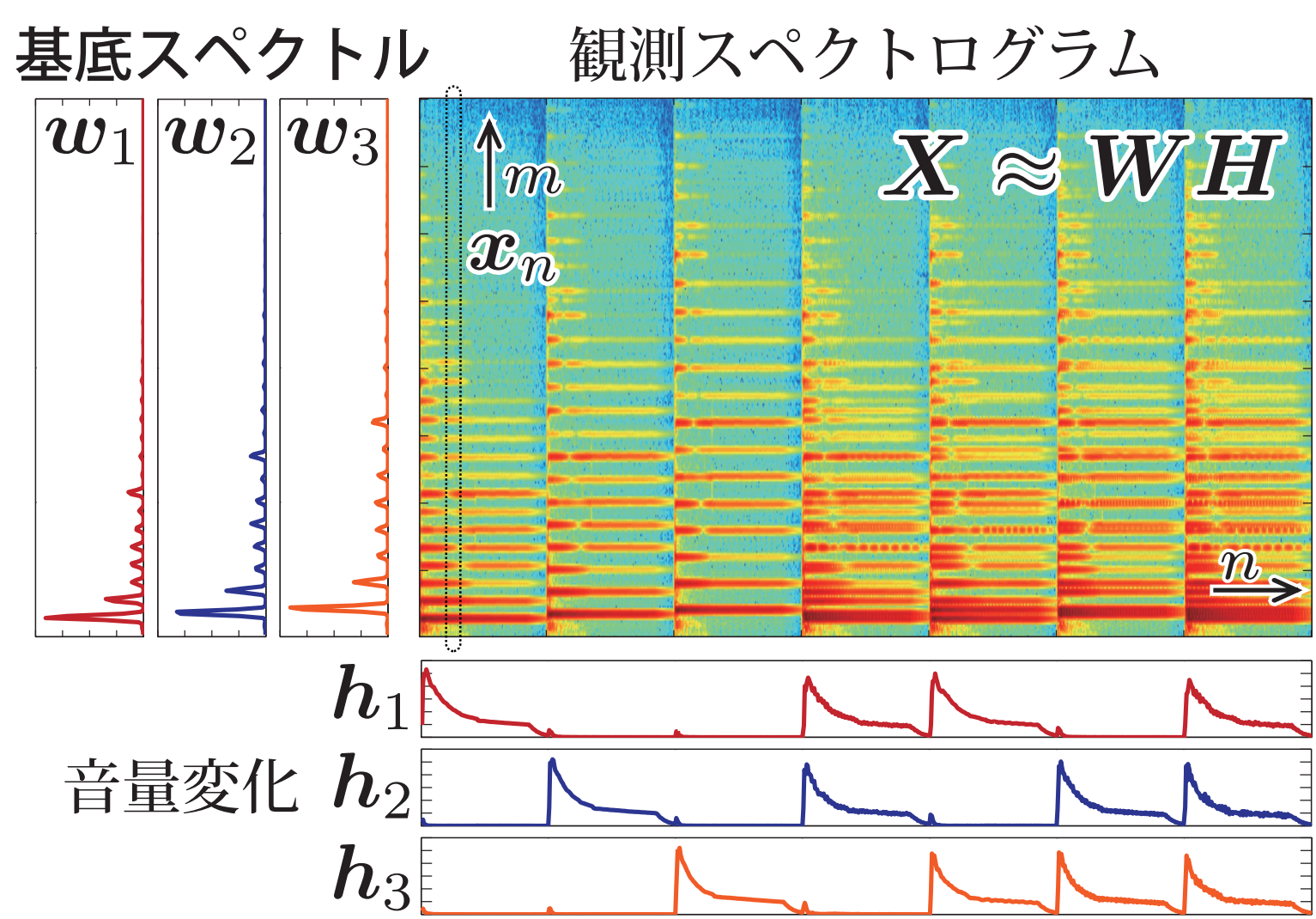
糸山 克寿
京都大学

後藤 真孝
産業技術総合研究所

音楽音響信号の高々有限個のパーツへの分解

何を部品とみなすかによって異なる確率モデルが定式化 → 部品の個数が未知であるのでノンパラメトリックベイズモデル

非負値行列分解 (Nonnegative Matrix Factorization: NMF)

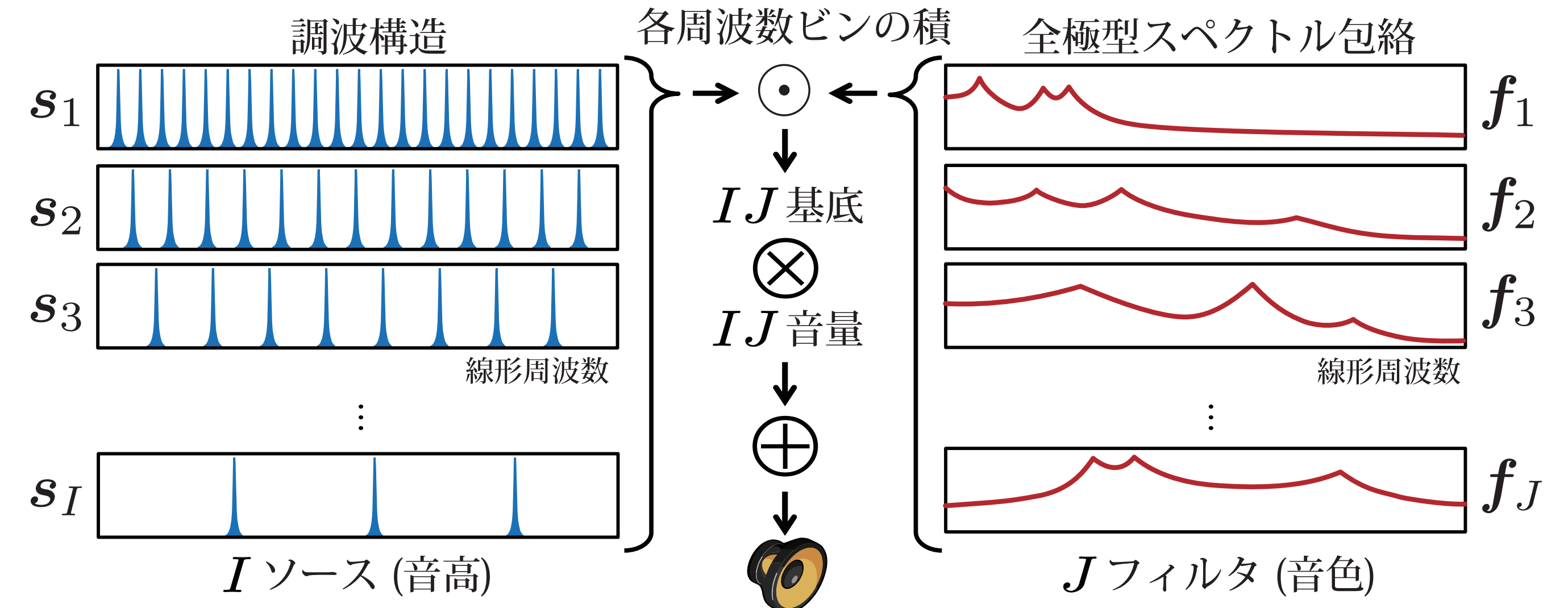


非負ベクトルを非負ベクトルの凸結合で近似

$$\mathbf{x}_n \approx \sum_{k=1}^K \mathbf{w}_k h_{kn} \stackrel{\text{def}}{=} \mathbf{y}_n$$

観測行列 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$
 基底行列 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K] \in \mathbb{R}^{M \times K}$
 音量行列 $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^T \in \mathbb{R}^{K \times N}$
 全て非負ベクトル

無限複合自己回帰モデル (Infinite Composite Autoregressive Model: iCAR)



音楽音響信号は無限個の音高・音色の組み合わせから生成されていると仮定

$$\mathbf{x}_n \approx \sum_{i=1}^I \sum_{j=1}^J (\mathbf{s}_i \odot \mathbf{f}_j) h_{ijn} \stackrel{\text{def}}{=} \mathbf{y}_n$$

音高で分離：多重音F0推定
 音色で分解：楽器パート分離
 [Kameoka 2009, Yoshii 2012]

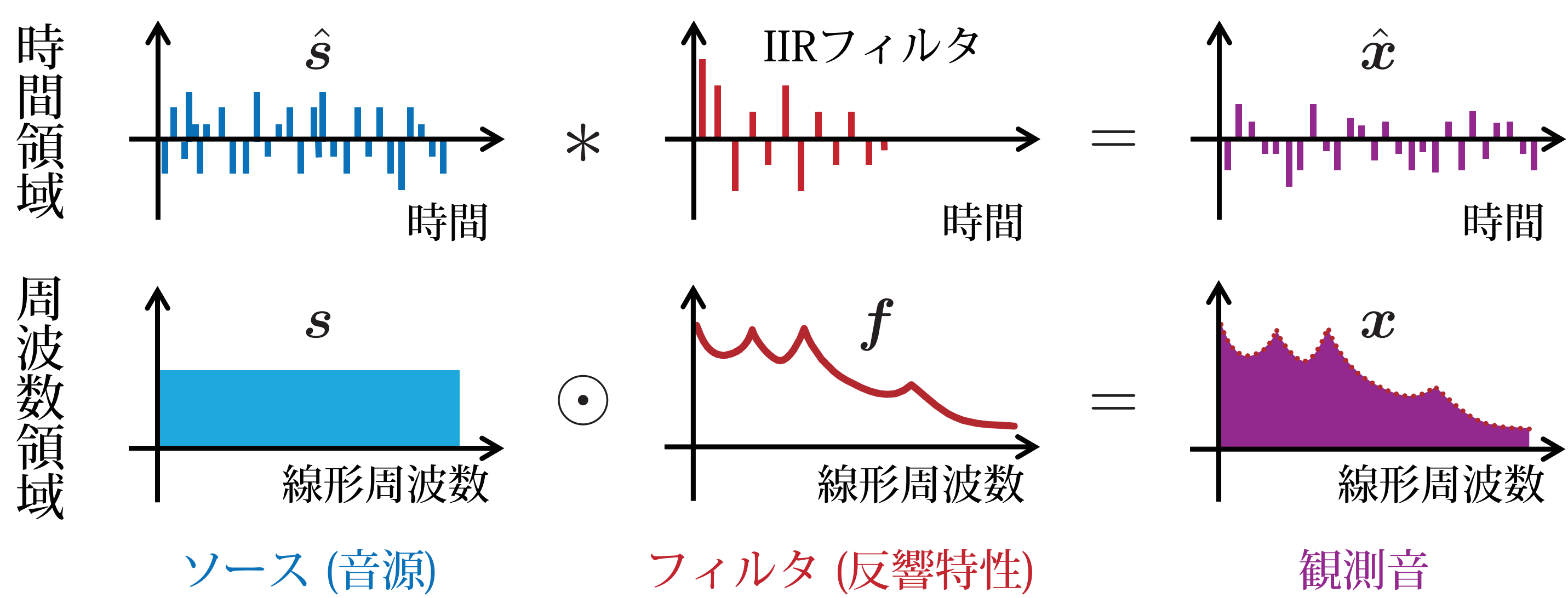
Bregmanダイバージェンスの最小化 = ある特定の確率モデルの最尤推定

$$\mathcal{D}_\phi(\mathbf{x}_n | \mathbf{y}_n) = \phi(\mathbf{x}_n) - \phi(\mathbf{y}_n) - \phi'(\mathbf{y}_n)^T (\mathbf{x}_n - \mathbf{y}_n) \quad \phi: \text{凸関数}$$

Pros: ガンマ過程事前分布を導入することで基底数の無限が可能 [Hoffman 2010]

Cons: 音色単位の分解が不可能・周波数成分が全て独立であるという強い仮定

線形予測分析 (Linear Predictive Coding: LPC)



観測信号の各サンプルを過去のPサンプルの線形結合で近似 (自己回帰過程)

$$\hat{x}_m = \sum_{p=1}^P a_p \hat{x}_{m-p} + \hat{s}_m \Rightarrow \hat{x}_m - \sum_{p=1}^P a_p \hat{x}_{m-p} = \hat{s}_m \quad \text{畳み込み演算}$$

$$Z \text{ 領域 } X(z) = S(z)F(z) \quad \text{ただし } F(z) = \frac{1}{1 - \sum_{p=1}^P a_p z^{-p}} \quad \text{全極型伝達関数}$$

フーリエ領域へ変換 $z = e^{2\pi i \frac{m}{M}}$

$$\mathbf{x} \approx \mathbf{s} \odot \mathbf{f} \stackrel{\text{def}}{=} \mathbf{y} \quad \text{楽器音の生成過程に対するソース・フィルタモデル}$$

s および f の同時推定は不良設定問題 → 音源信号がガウス性白色雑音であると仮定

問題 1：周期性をもつ音源信号 (音高をもつ観測音) には対応できない

推定されるスペクトル包絡には、倍音周波数で不要に急峻なフォルマント (極) が存在

問題 2：人間の聴覚特性に合致する対数周波数領域では定式化できない

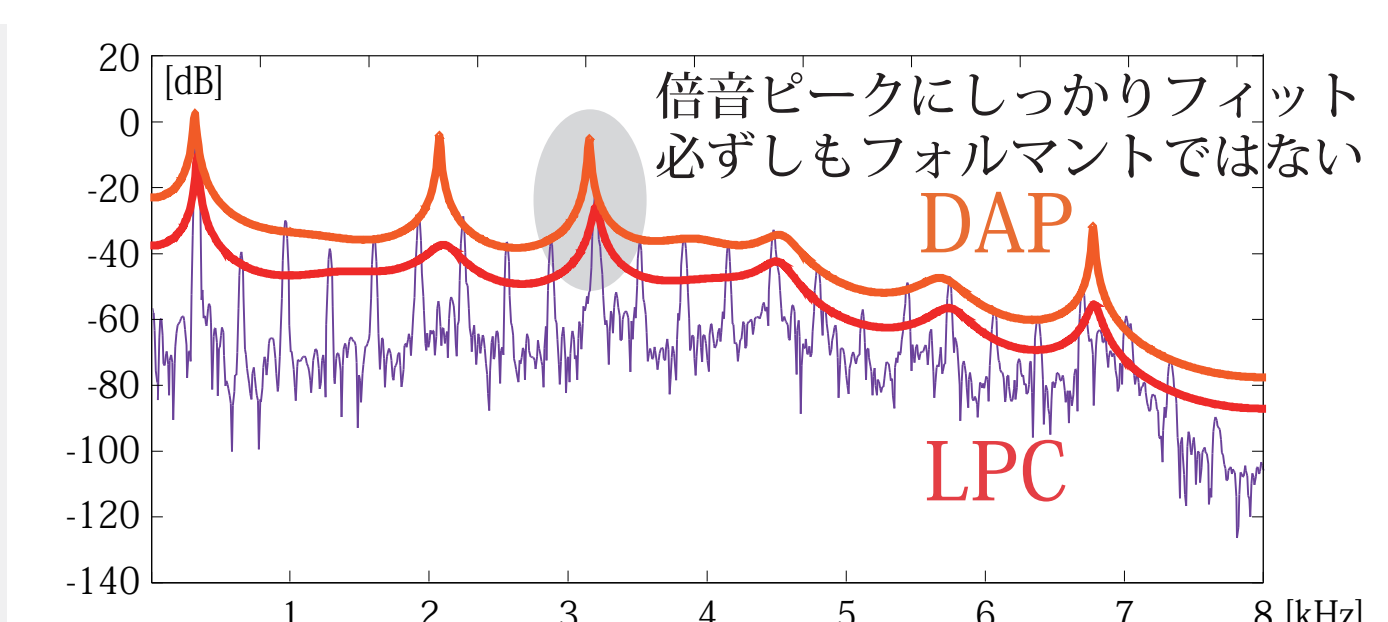
多重音解析と親和性が高い対数周波数スペクトログラムを取り扱うことができない

無限複合自己回帰モデル

音源スペクトルも同時に推定

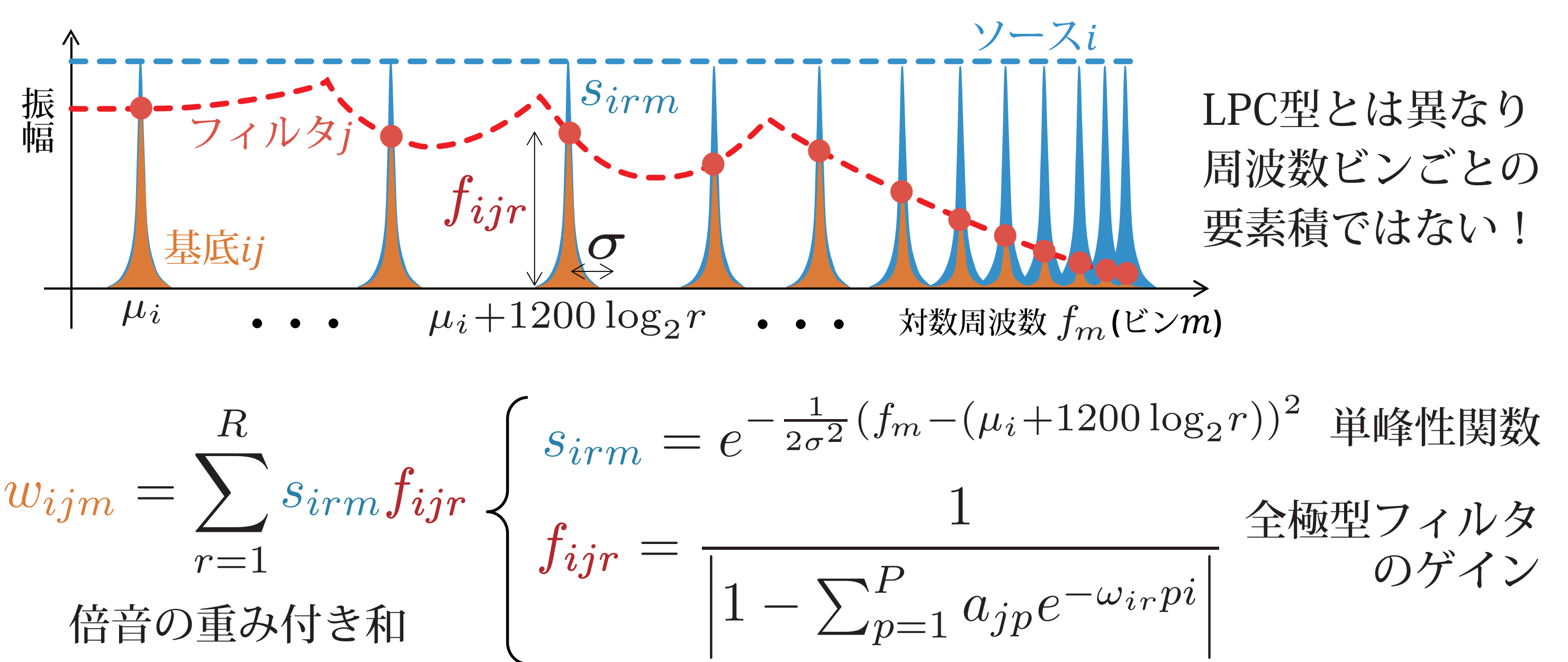
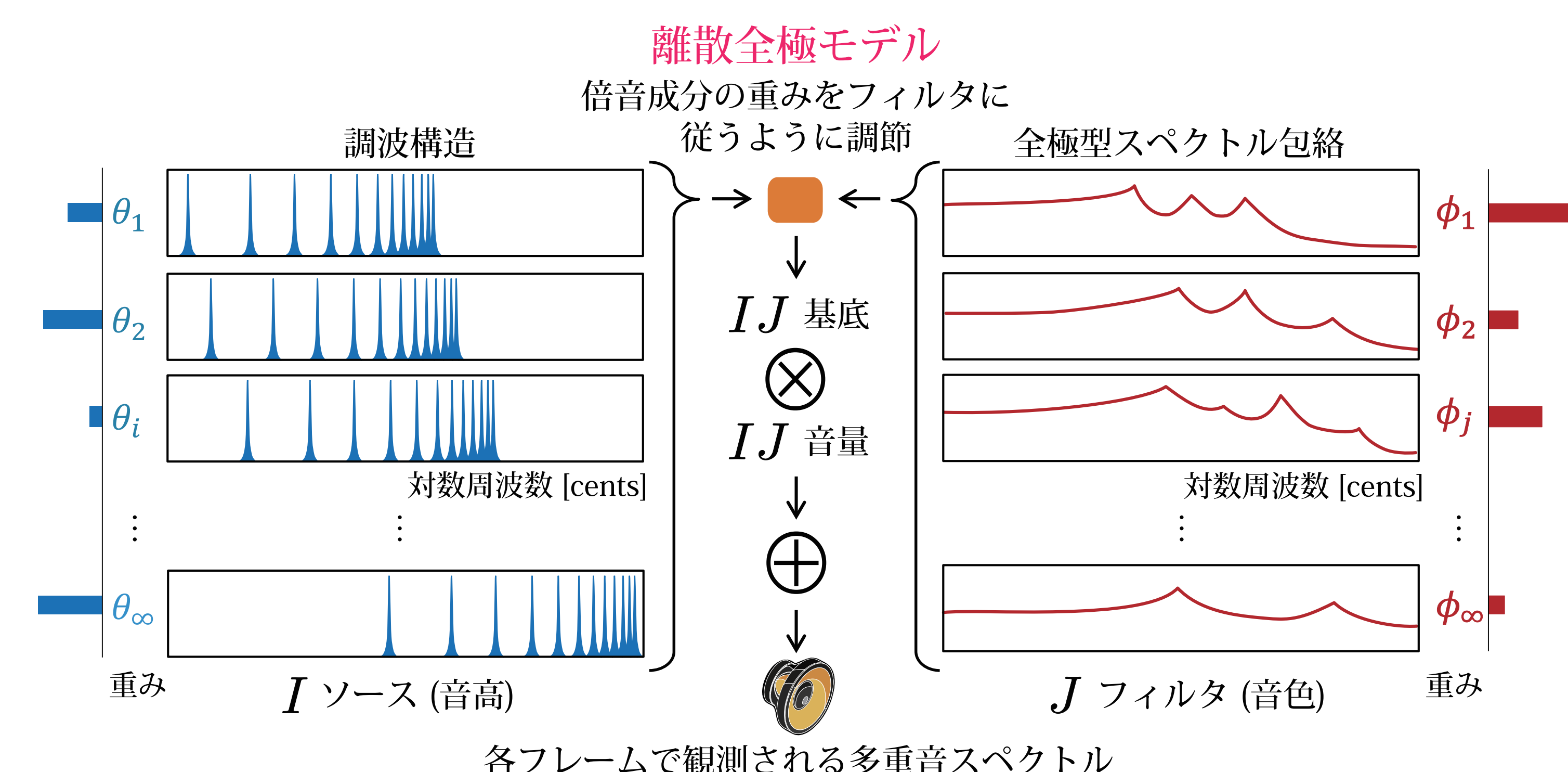
離散全極モデル (LPCの改良)

倍音ピークのみから包絡を推定



対数周波数領域における複数の音高と複数の音色への分解

無限重畳離散全極モデル：無限複合自己回帰モデルと離散全極モデルとの統合に基づく混合音のソース・フィルタ分解



観測データに対する確率モデルの尤度

$$X_{mn} \sim \text{Poisson} \left(\sum_{i=1}^{I \rightarrow \infty} \sum_{j=1}^{J \rightarrow \infty} \theta_i \phi_j w_{ijm} h_{ijn} \right) \quad h_{ijn} \sim \text{Gamma}(a_H, b_H)$$

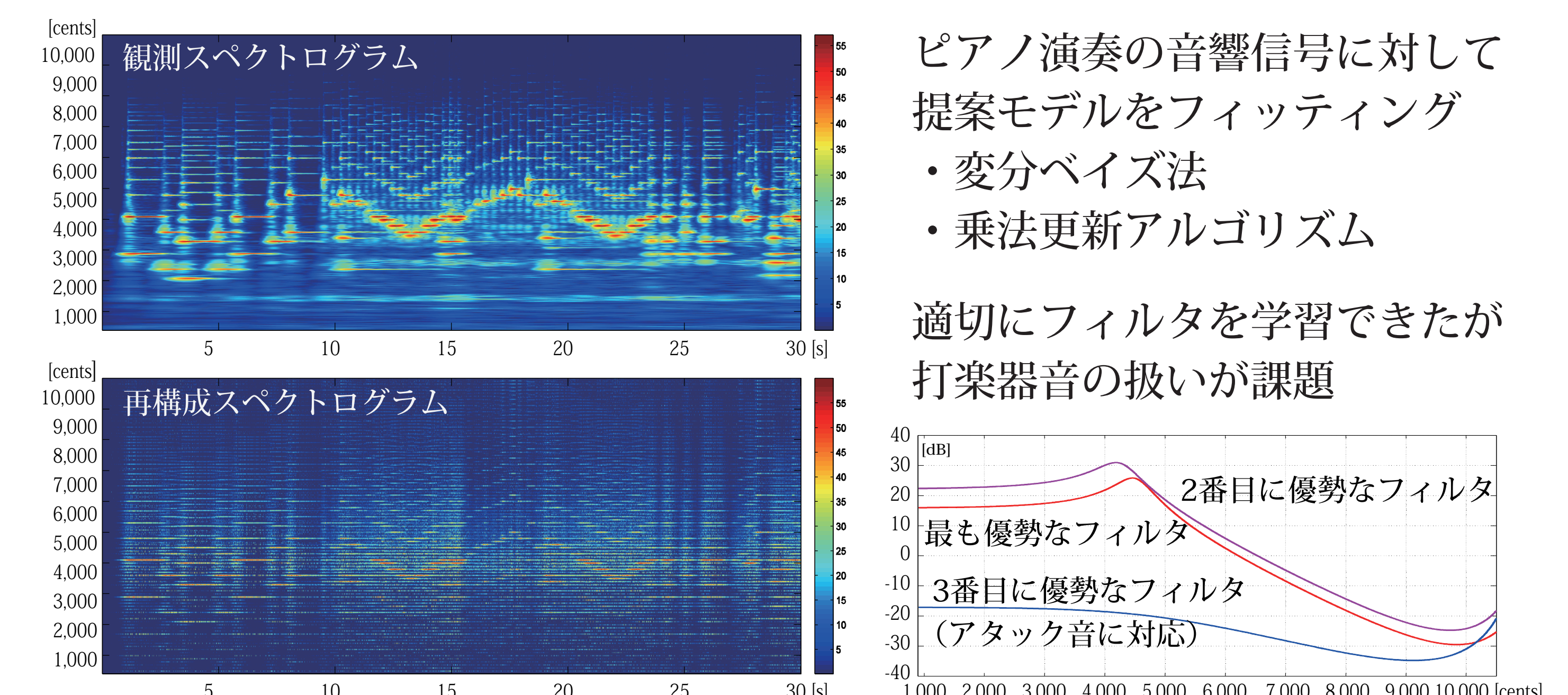
音量に対するガンマ事前分布 (時間的連続性も考慮可能)

スパースな解に誘導する効果

ソース・フィルタの重みに対するガンマ過程 (無限次元の非負ベクトル上の事前分布)

$$\theta_i \sim \text{GaP}(\alpha_\theta, \mathbf{1}) \quad \text{ほとんどすべての要素が限りなくゼロに近く、}$$

$$\phi_j \sim \text{GaP}(\alpha_\phi, \mathbf{1}) \quad \text{高々有限個の要素のみ値をもつように誘導できる}$$



ピアノ演奏の音響信号に対して提案モデルをフィッティング
 ・変分ベイズ法
 ・乗法更新アルゴリズム

適切にフィルタを学習できたが打楽器音の扱いが課題