

多重音基本周波数解析のための 無限複合自己回帰モデル

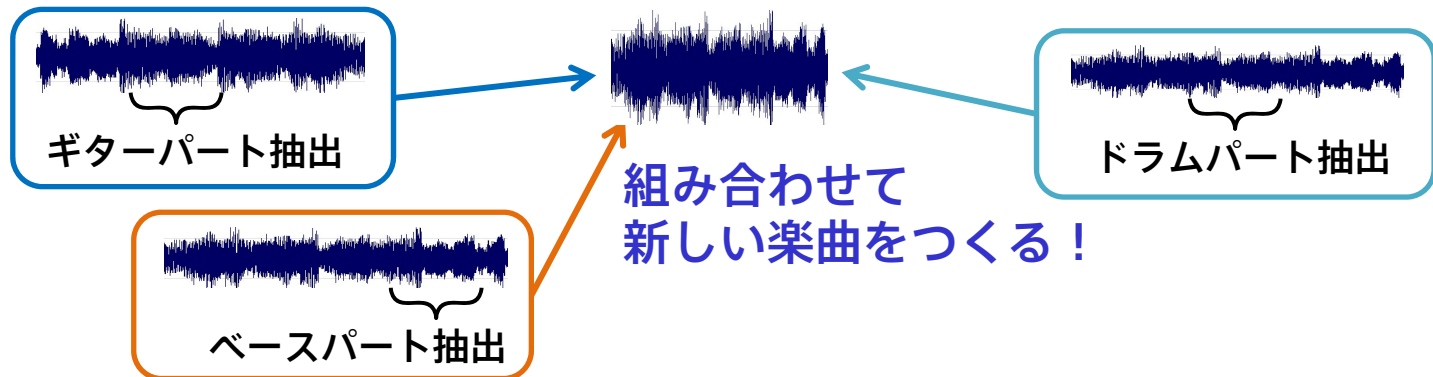
産業技術総合研究所

吉井 和佳 後藤 真孝

{k.yoshii, m.goto}@aist.go.jp

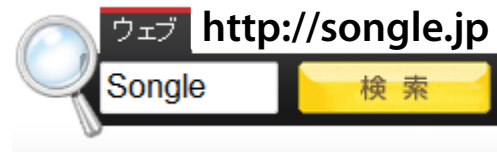
研究の背景

- 音楽音響信号の分解・再構成技術が望まれる
 - 音楽鑑賞の楽しみ方を広げたい
 - 楽器パートごとの音量調整・編集
 - 調波音・打楽器音分離 [宮本2007]
 - 楽器音イコライザ [糸山2008]
 - Drumix: ドラムパートのリアルタイム編集 [吉井2006]
 - 音楽制作の楽しみ方を広げたい
 - 次世代のマッシュアップ・二次創作
 - 既存楽曲に含まれる「パーツ」を組み合わせた



想定する応用

- Songle: 能動的音楽鑑賞用WEBサービス
 - 音楽音響信号の自動解析結果を表示
 - 誤り訂正も可能 → システムが賢くなる

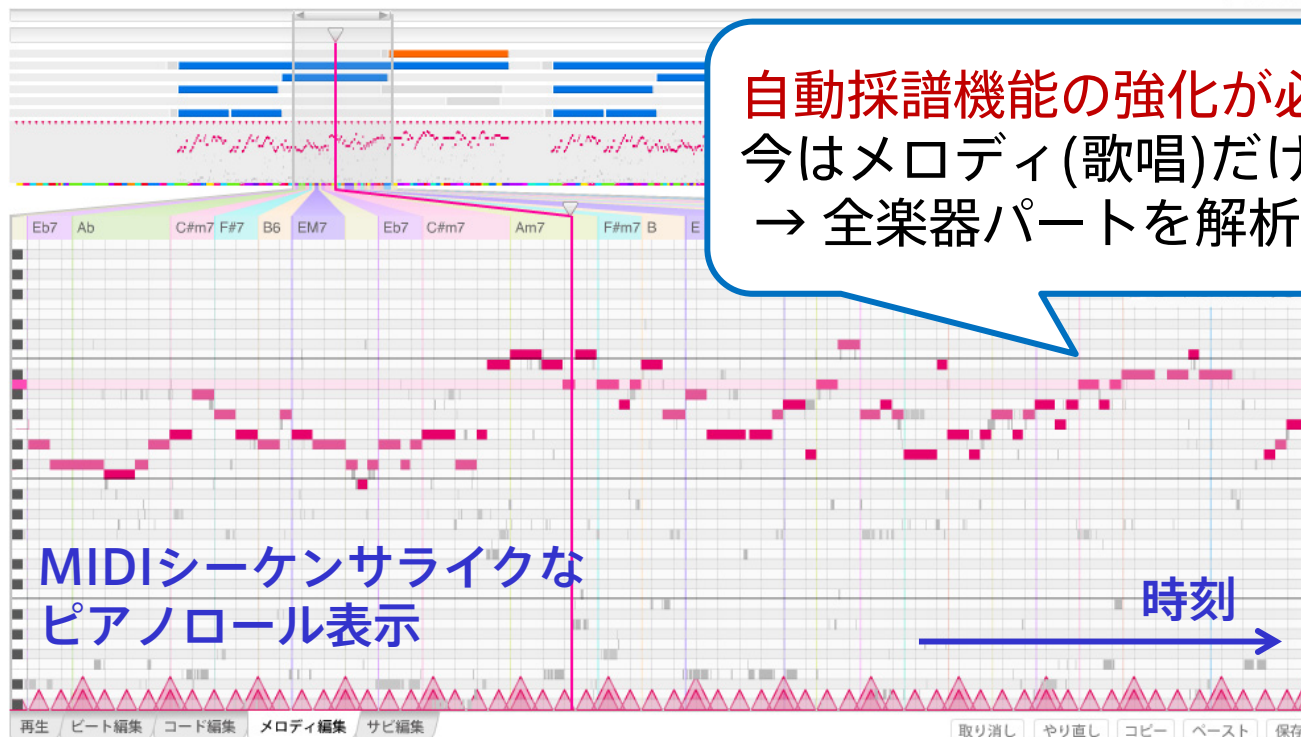


♪ マージナル by OSTER project

[リンク](#) / [好き](#) / [嫌い](#) / [次の曲へ](#) / [編集履歴](#) / [キー操作一覧](#)

1426 Plays 34581 Edits 9 Lists

01:02 / 04:11



自動採譜機能の強化が必要
今はメロディ(歌唱)だけが対象
→ 全楽器パートを解析したい

音高

時刻

音楽の成り立ち

- 音楽音響信号を構成する「パーツ」とは？
 - 楽器音：2種類の要素の組み合わせ

音色 (楽器種)



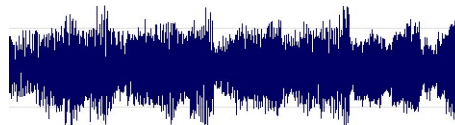
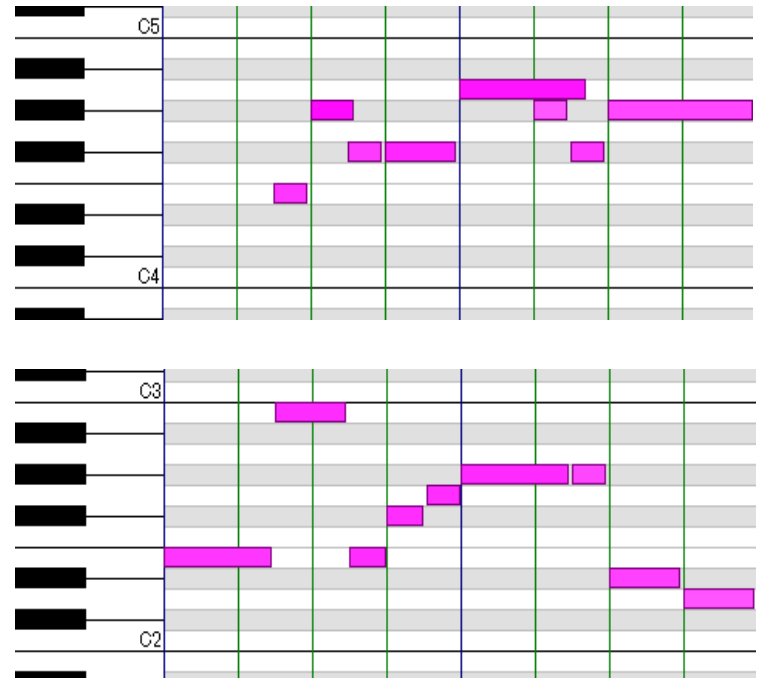
様々な
楽器音を
生成



重ね合わせ




音高 (基本周波数)

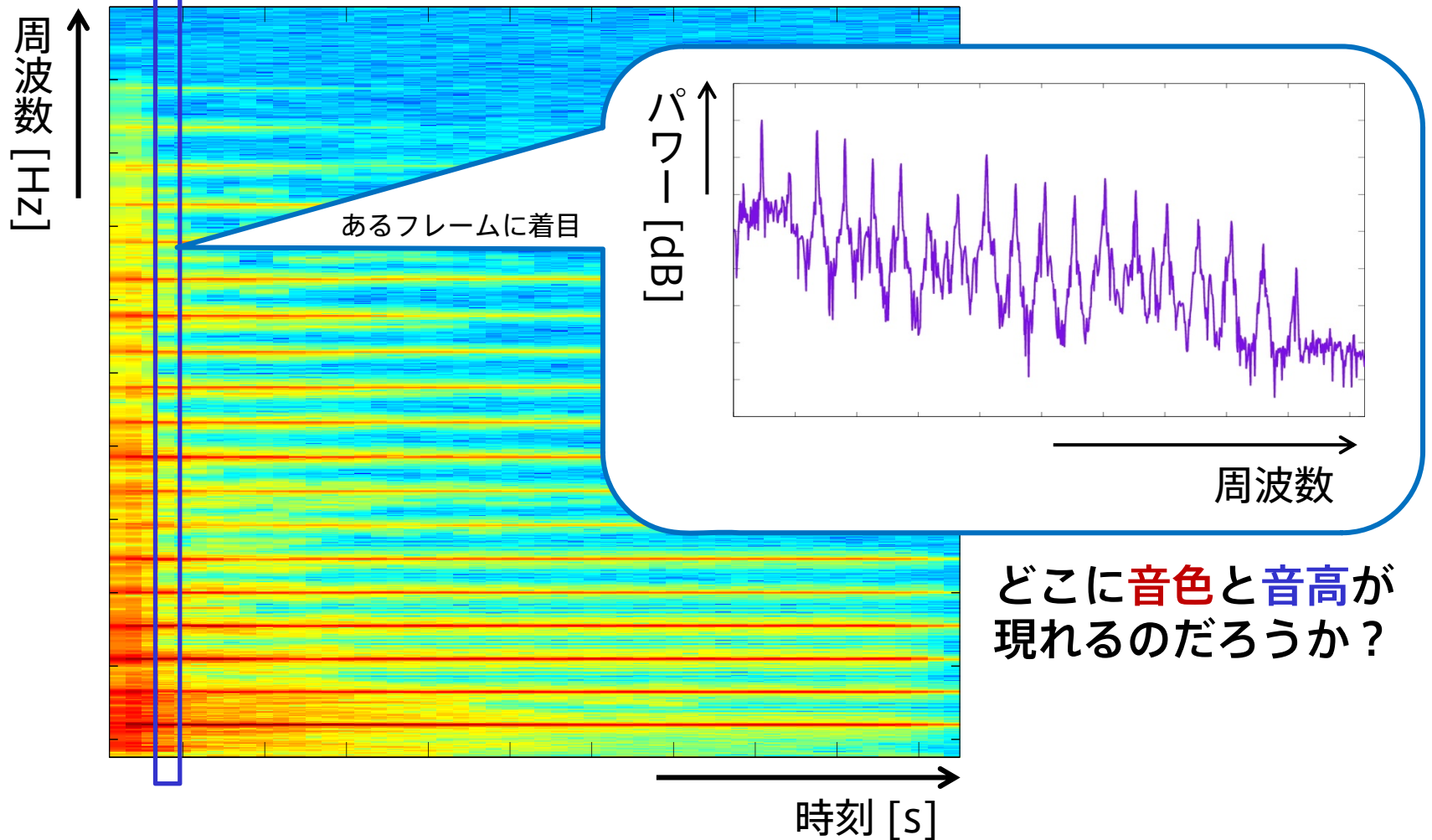


音楽音響信号

楽器音スペクトル

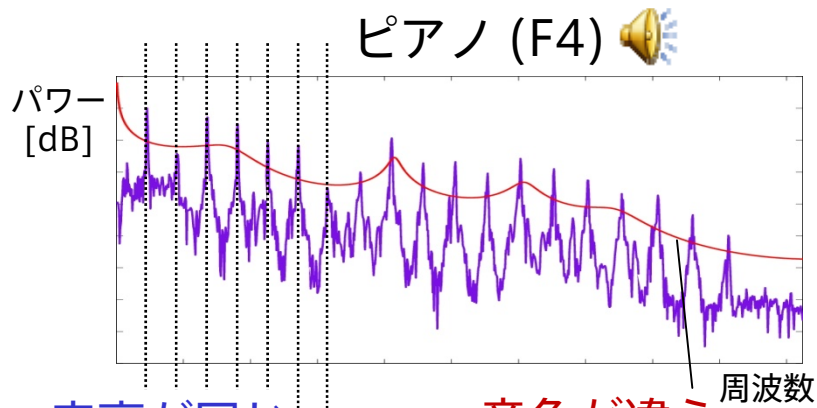
- 短時間フーリエ変換 (STFT) による周波数解析

楽器音 (ピアノ・F4) のスペクトログラム 



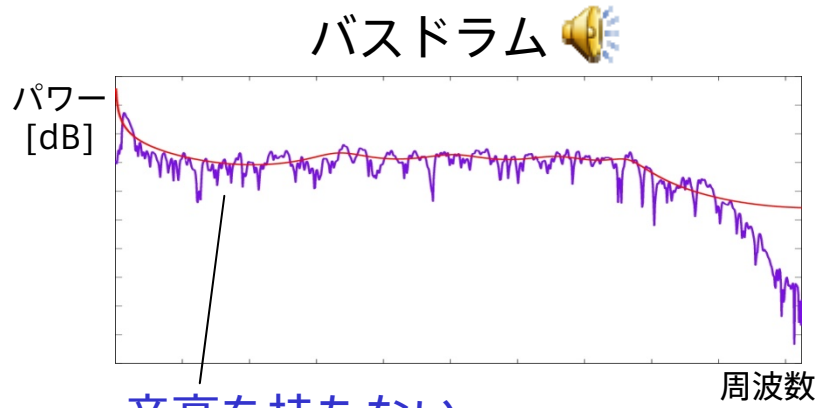
楽器音スペクトル

- 「包絡構造」と「調波構造」とに分離
 - 「音色」と「音高」とに対応
 - 打楽器音も対象

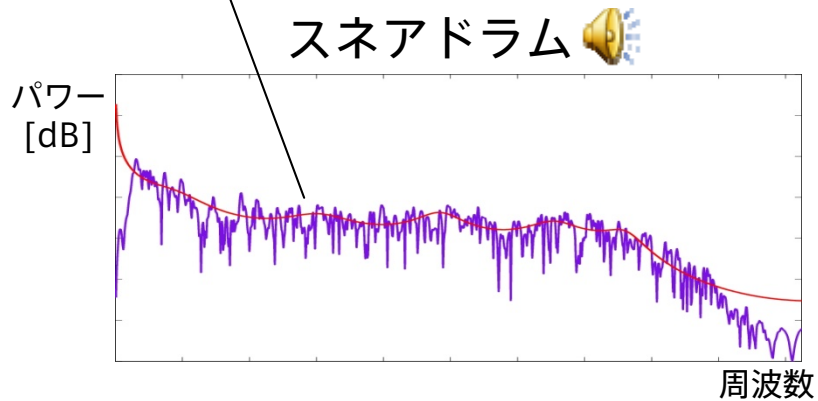
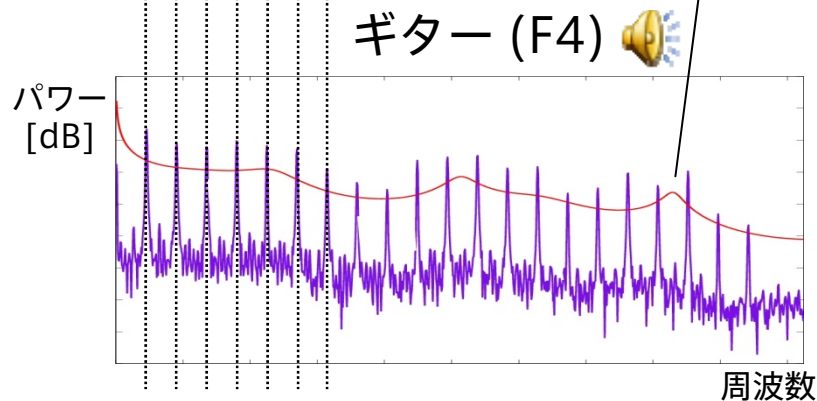


音高が同じ
→ 調波構造が同じ

音色が違う
→ 包絡線が異なる

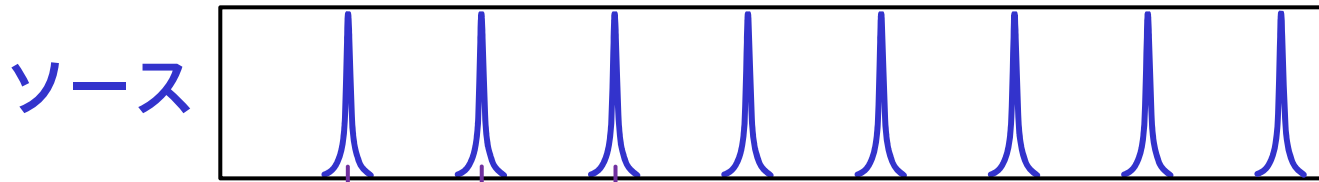


音高を持たない
→ ならかな分布 (雑音)



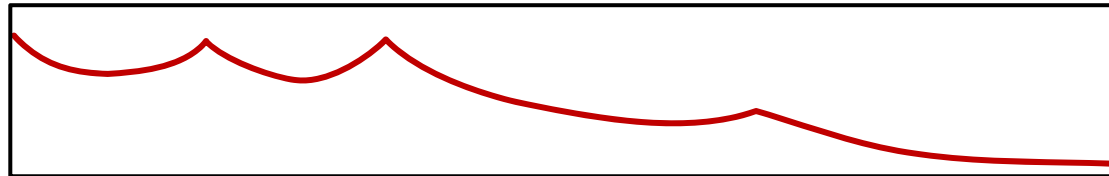
調波音の生成モデル

- 調波構造をもつ楽器音の生成過程
 - ソース：等間隔に急峻なピークを持つ関数 (くし型)
 - フィルタ：周波数応答を表す関数 (なだらかな)

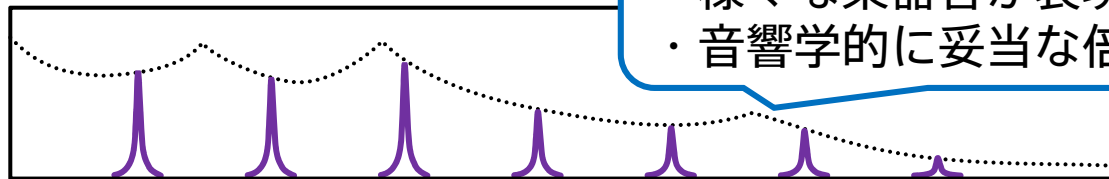


基本周波数 μ 2μ 3μ \otimes 各周波数ビンごとに掛け算

フィルタ



調波音
スペクトル



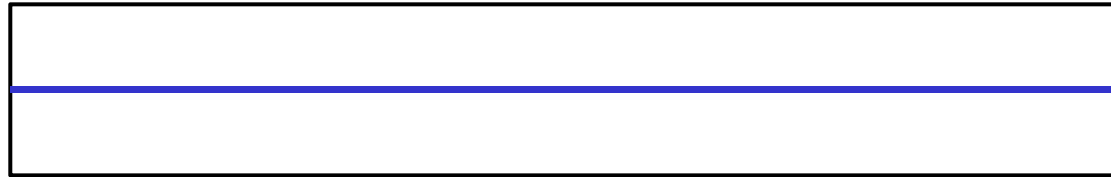
- ソースやフィルタを変えれば
様々な楽器音が表現可能！
- 音響学的に妥当な倍音比

※縦軸は
リニア

非調波音の生成モデル

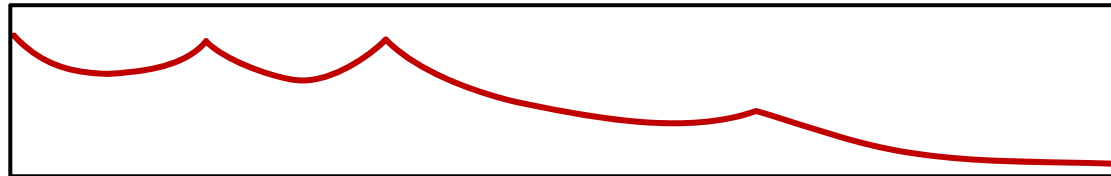
- 打楽器音や有色雑音の生成過程
 - ソース：白色雑音に対応するスペクトル (平坦)
 - フィルタ：周波数応答を表す関数 (なだらかな)

ソース



⊗ 各周波数ビンごとに掛け算

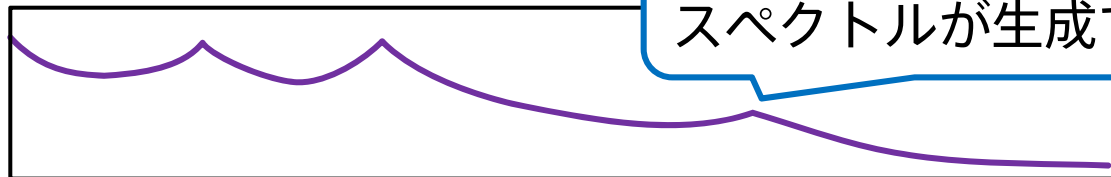
フィルタ



||

周波数方向になだらかな
スペクトルが生成できる

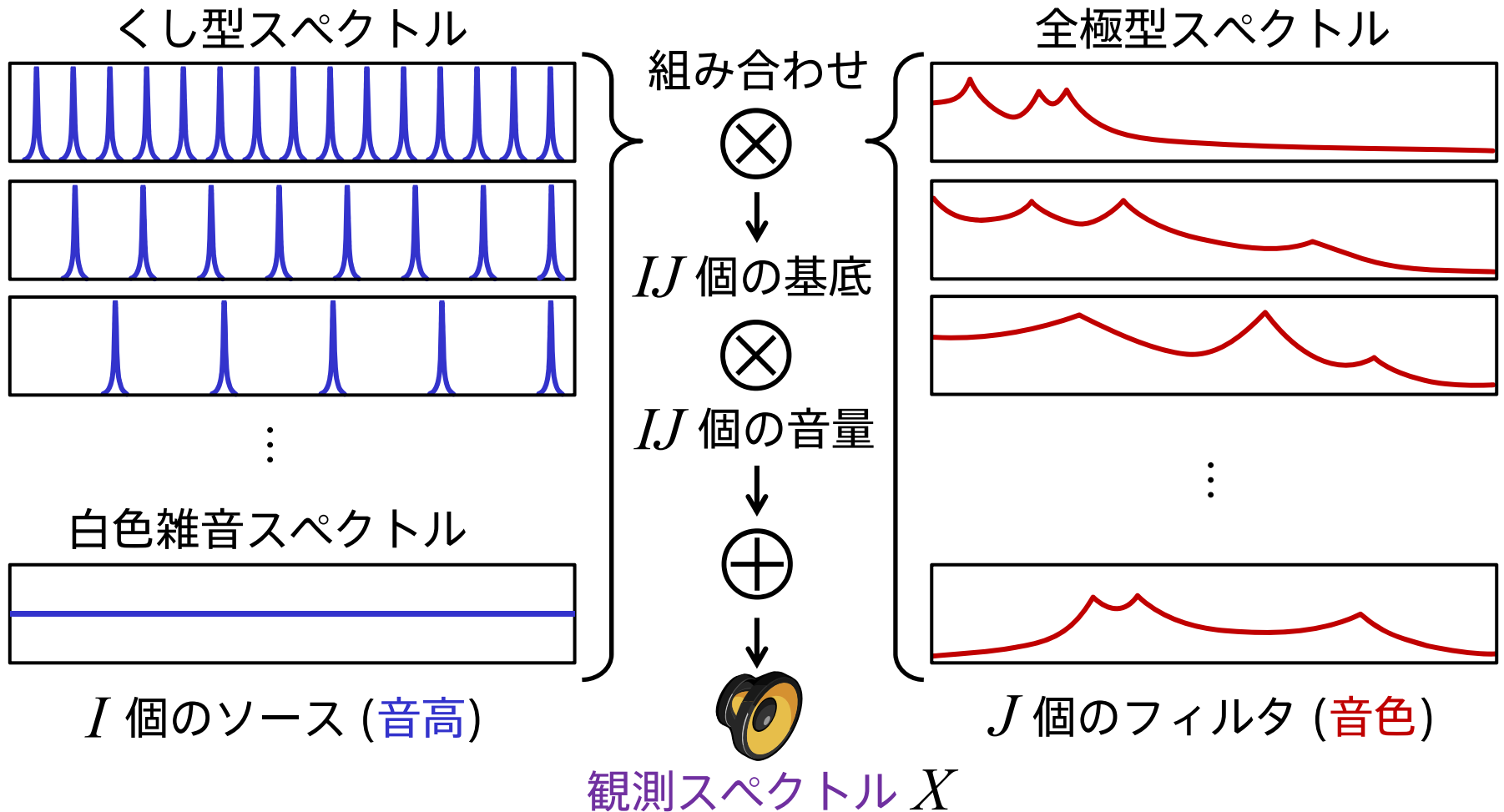
非調波音
スペクトル



※縦軸は
リニア

混合音の生成モデル

- 混合音スペクトルのソース・フィルタへの分解
 - 複合自己回帰モデル [亀岡2007, 安良岡2012]



従来の問題点

- モデルの複雑さを事前に指定する必要あり
 - ソースの個数は未知
 - フィルタの個数は未知
- モデル選択が必要

ソースの個数

フィルタの
個数

| | 10 | 20 | 30 | 40 | ... |
|-----|---------|--------|--------|--------|-----|
| 1 | -20,000 | -9,000 | -8,000 | -8,000 | |
| 2 | -10,000 | -8,500 | -7,000 | -7,500 | |
| 3 | -9,000 | -8,300 | -7,500 | -7,900 | |
| 4 | -8,800 | -8,400 | -8,000 | -8,500 | |
| ... | | | | | |

それぞれの設定で
独立に分解を行う

$\log p(X)$ が最大となるモデルを網羅的に探索

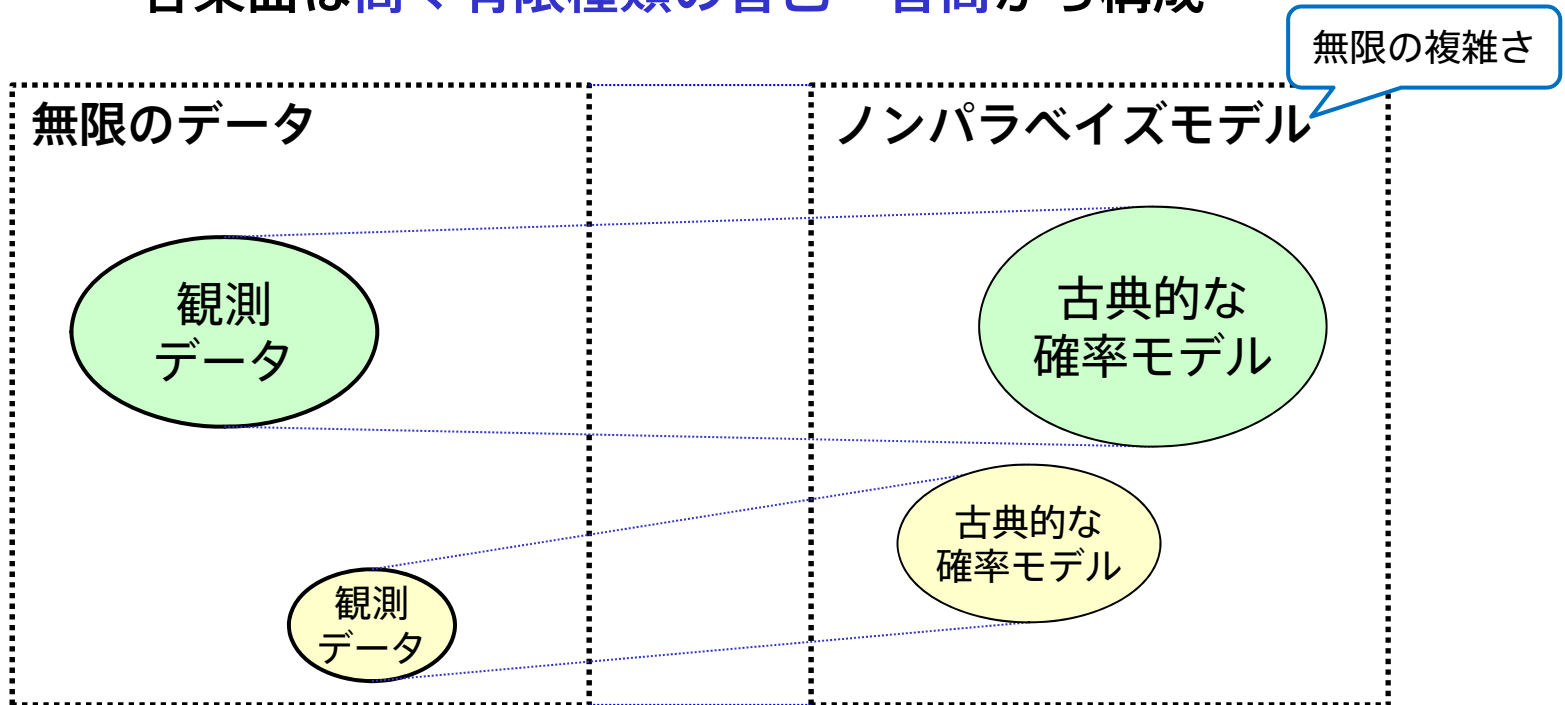
→ 計算量が組み合わせ爆発！

本研究のアプローチ

- ノンパラメトリックベイズ理論の活用

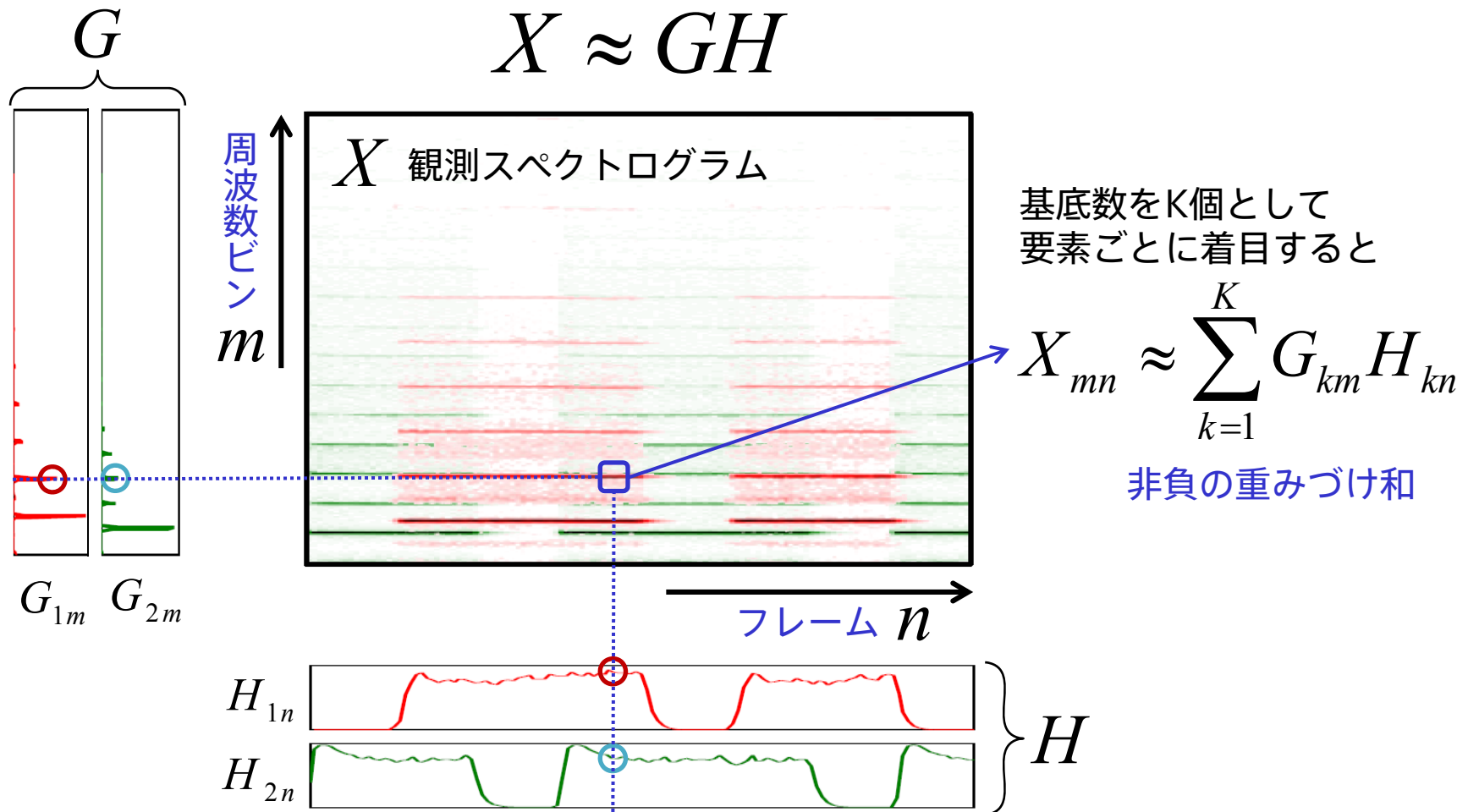
- **無限複合自己回帰モデル**

- 無限個のソース・フィルタを考慮可能
 - ソース・フィルタの個数を事前に指定する必要なし
- 観測データを説明するのに必要な分だけ実体化
 - 各楽曲は高々有限種類の音色・音高から構成



非負値行列因子分解

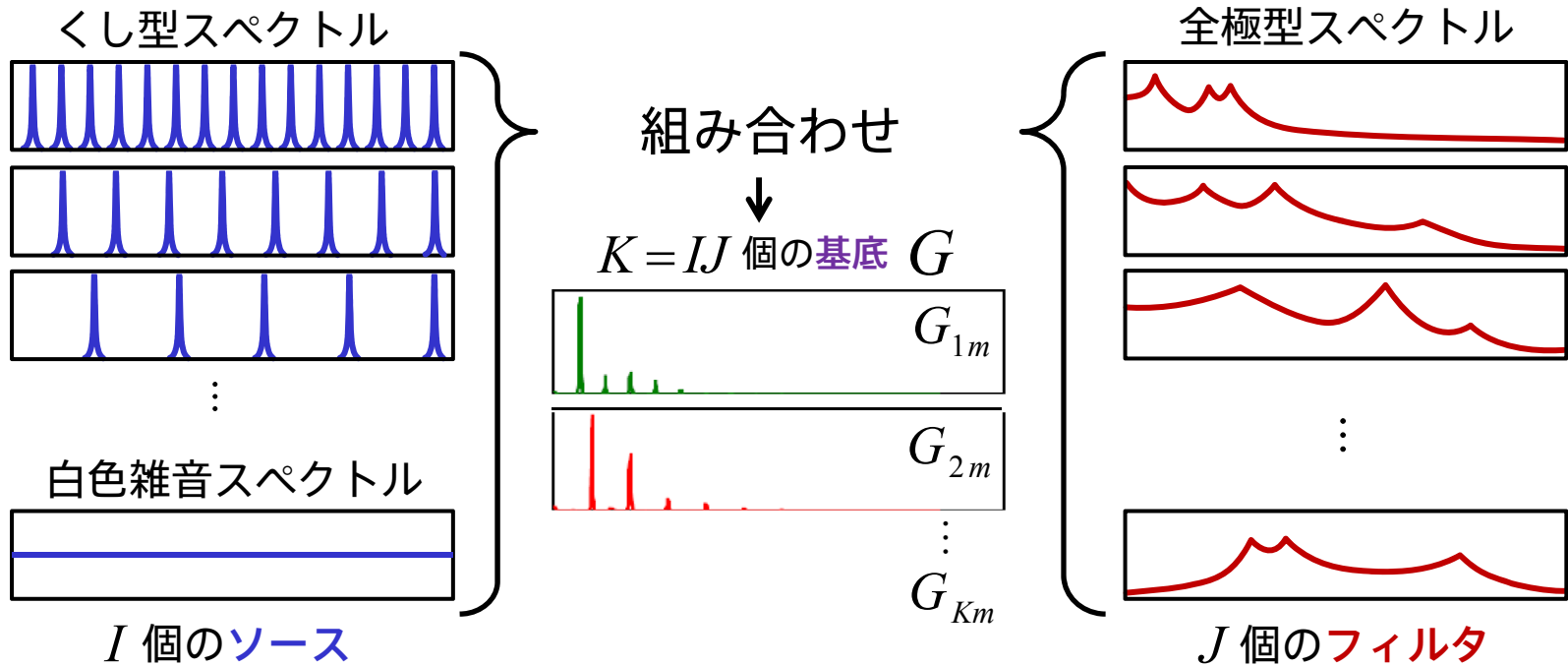
- 観測スペクトログラムを非負の行列積で近似
 - 少数の「基底」と「音量」との積に分解



複合自己回帰モデル

- 非負値行列因子分解 (NMF) の拡張

- 「基底」を「ソース」と「フィルタ」の積に分解



$$X_{mn} \approx \sum_{k=1}^K G_{km} H_{kn} \quad \Rightarrow \quad X_{mn} \approx \sum_{i,j}^{I,J} W_{im} A_{jm} H_{ijn}$$

基底 音量 ソース フィルタ 音量

無限複合自己回帰モデル

- ノンパラメトリックベイズモデル
 - 無限個のソース・フィルタをあらかじめ想定
 - 観測データを説明するのに必要な分だけ有効化
 - 各楽曲には「高々有限種類」しか含まれない

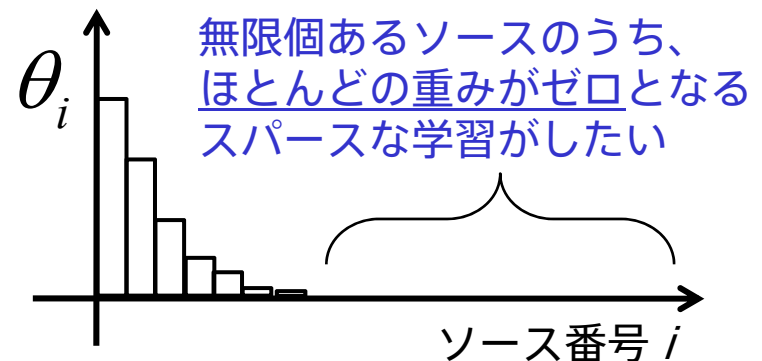
$$X_{mn} \approx \sum_{i,j}^{\infty} W_{im} A_{jm} H_{ijn}$$

無限種類の組み合わせの和

ソース*i* フィルタ*j* このペアの音量

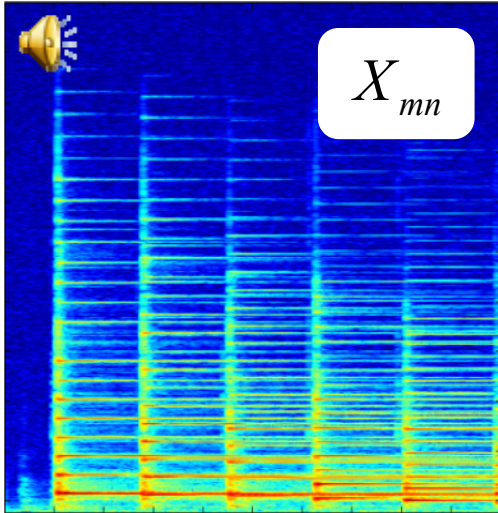
ソース*i*の重み

フィルタ*j*の重み

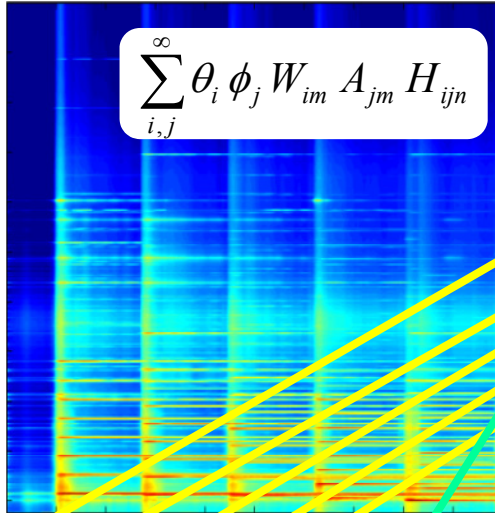


モデルの学習例

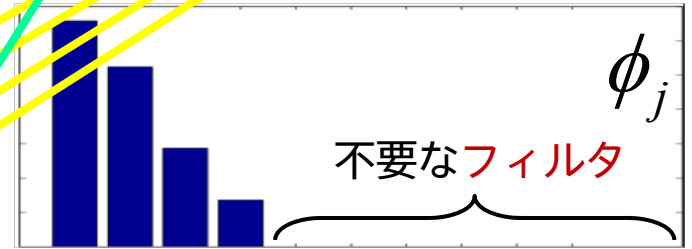
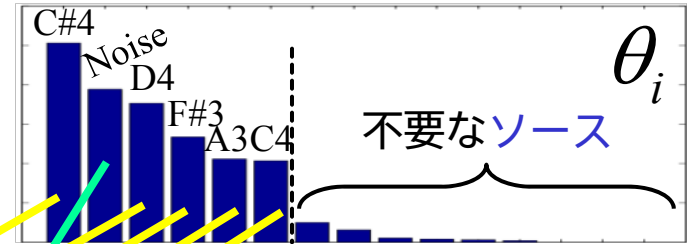
観測スペクトログラム



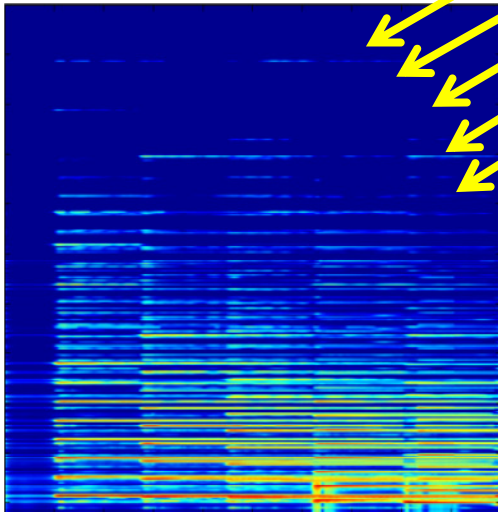
再構成スペクトログラム



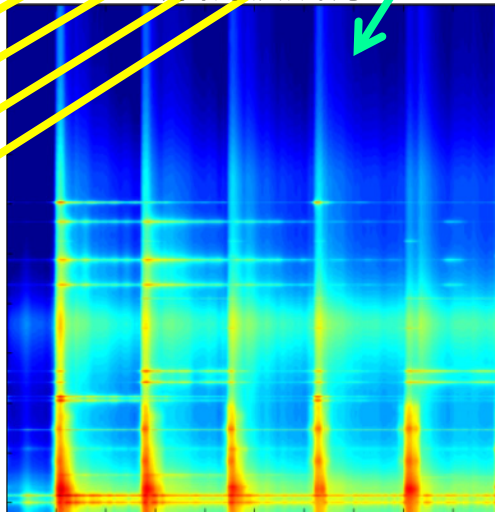
グローバルな重み



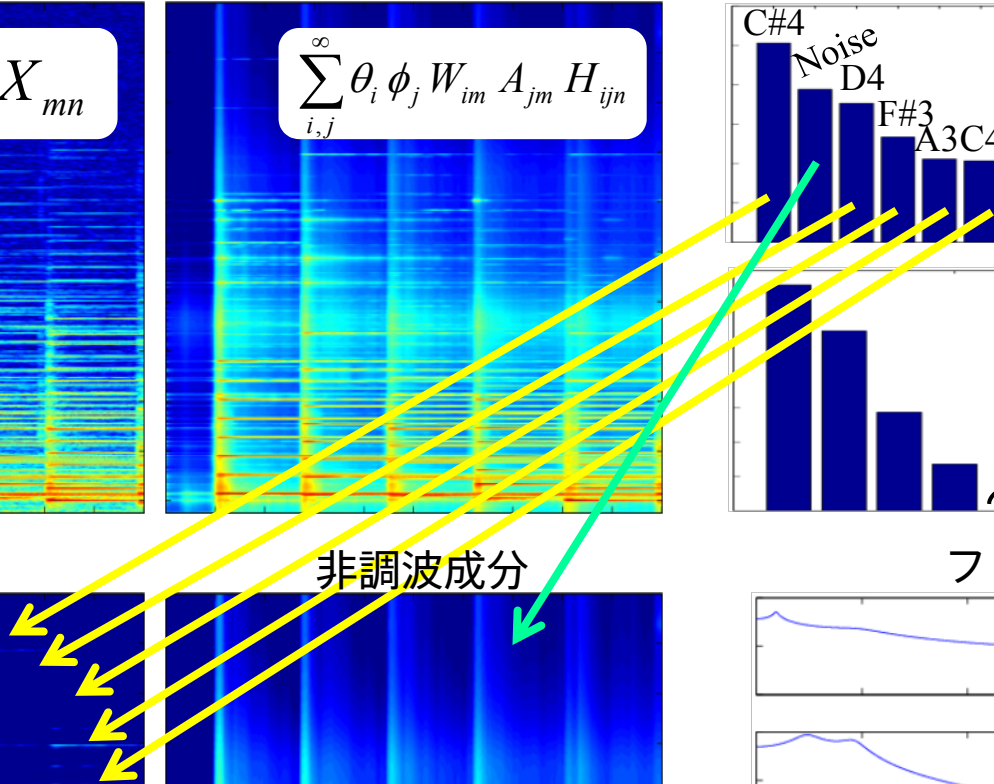
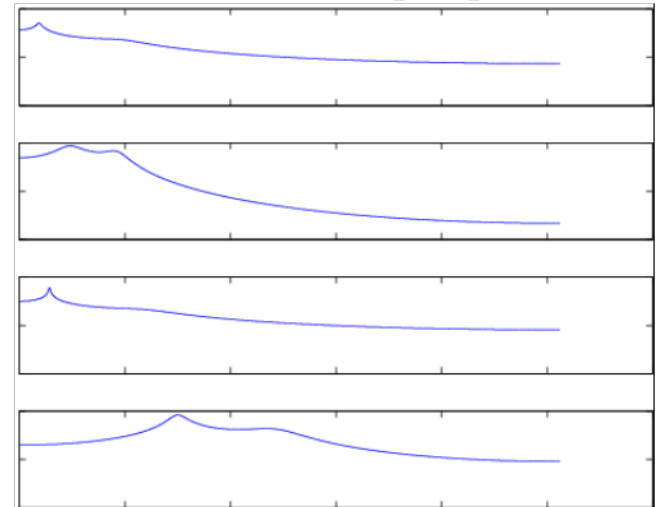
調波成分



非調波成分



フィルタ [dB]



ベイズモデルの構築

- 無限複合自己回帰モデルの数学的定式化

$$X_{mn} \approx \sum_{i,j}^{\infty} \theta_i \phi_j W_{im} A_{jm} H_{ijn}$$

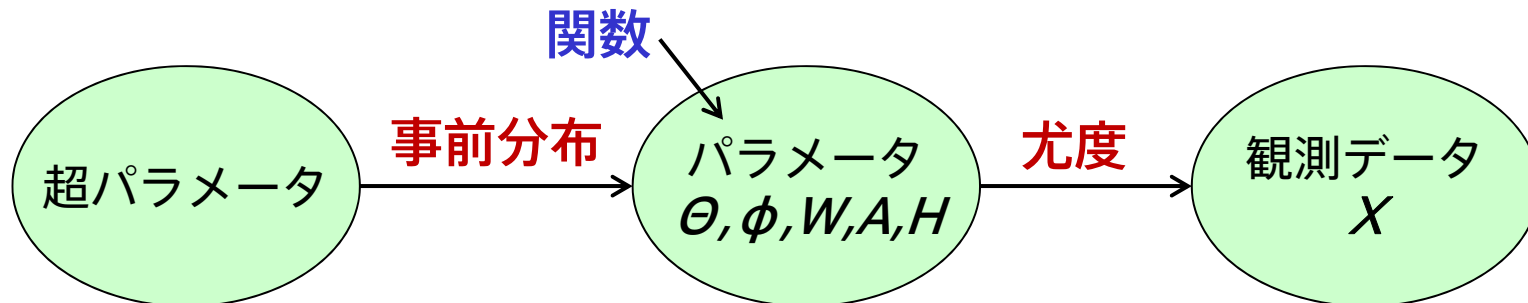
ソース*i* フィルタ*j* ペアの音量

– 尤度 (距離尺度) の設計

- ポアソン分布：KLダイバージェンス基準NMFに対応
- 指数分布：ISダイバージェンス基準NMFに対応

– 事前分布の設計

- グローバルな重み θ, ϕ ・音量 H に対する事前分布
- ソース W ・ フィルタ A の形状を表す関数



事前分布の設計

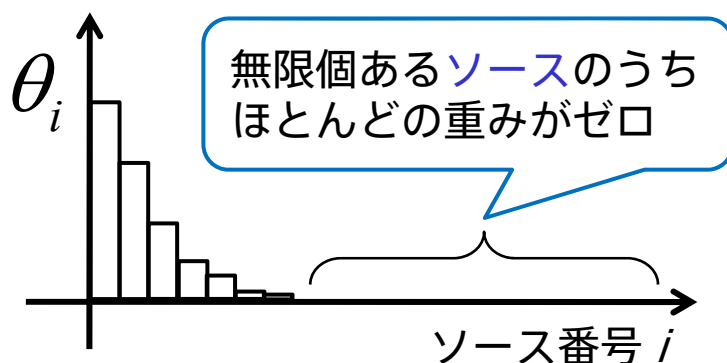
- グローバルな重み θ, ϕ に対する事前分布

- ガンマ過程事前分布 [Hoffman 2010]

- 無限次元の非負値ベクトルを生成
 - 要素を大きいもの順に並べると指数減衰

$$\theta \sim \text{GammaProcess}(\alpha)$$
$$\phi \sim \text{GammaProcess}(\gamma)$$

超パラメータ：
スパースさを制御



モデルの実効的な複雑さが自動決定される！

事前分布の設計

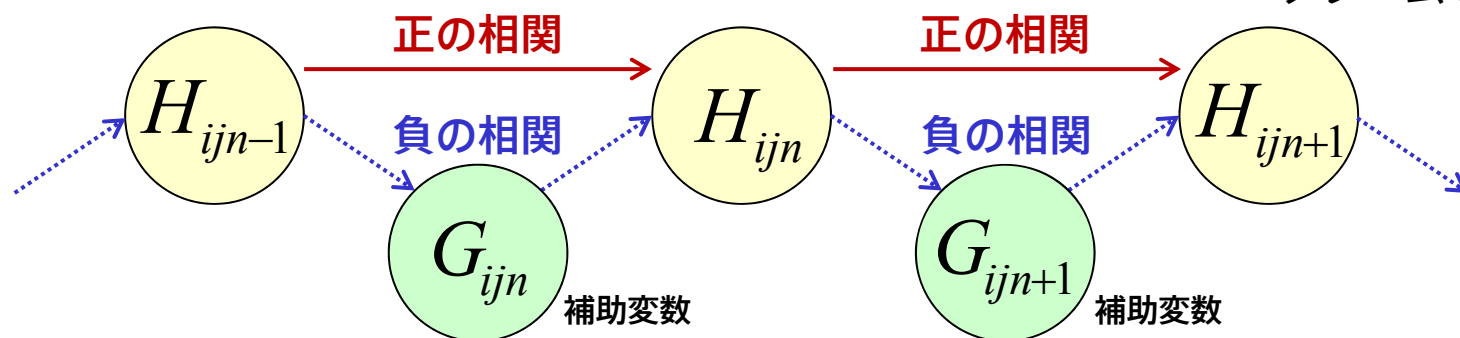
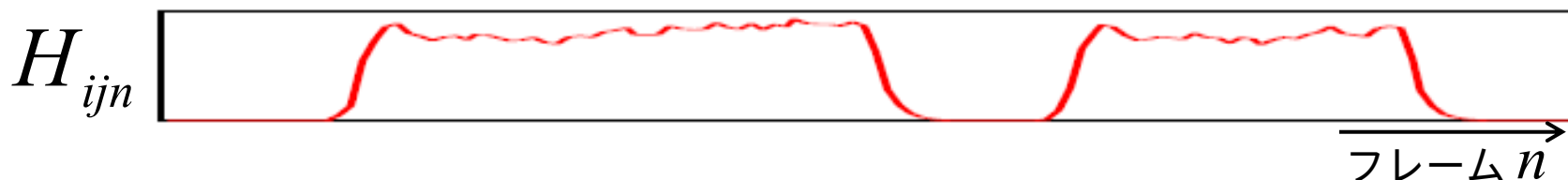
- 音量 H に対する事前分布

- ガンマチェイン事前分布 [Cemgil 2007]

- 音量を時間方向に滑らかに変化させる
 - 隣り合う音量が正の相関を持つマルコフ連鎖

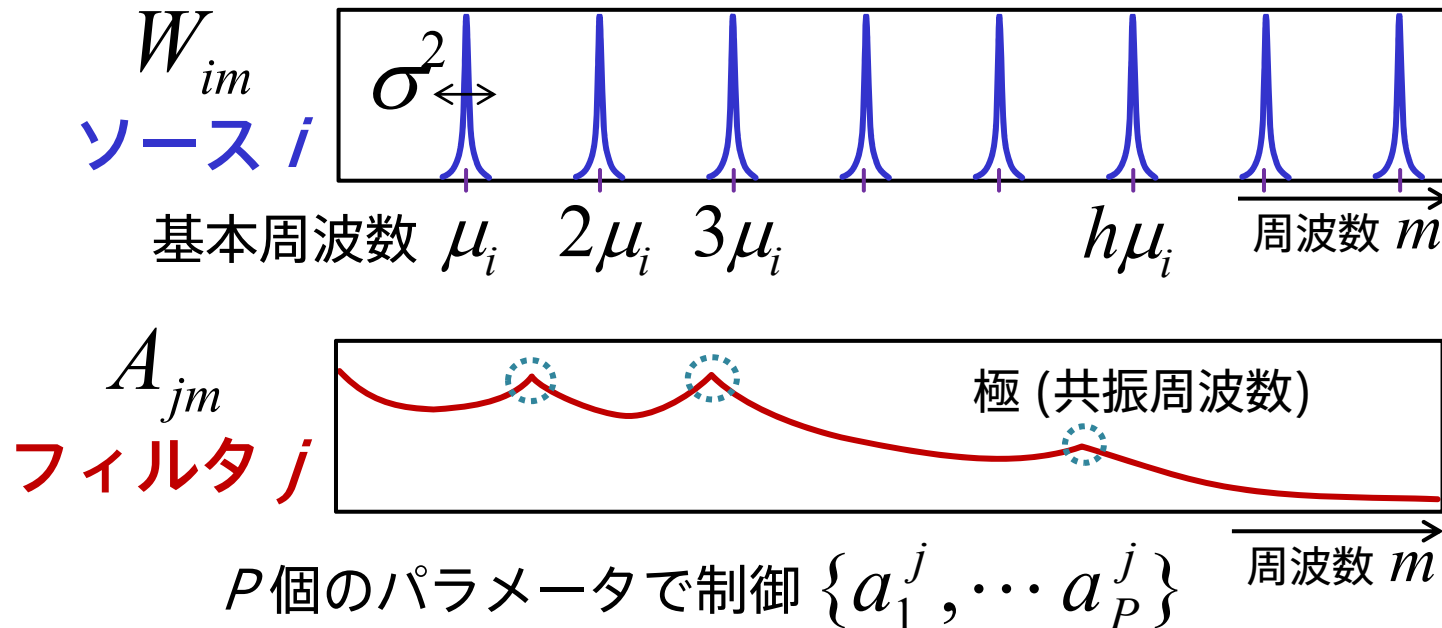
$$G_{ijn} \sim \text{Gamma}(\beta, \beta H_{ijn-1}) \quad E[G_{ijn-1}] = H_{ijn-1}^{-1}$$

$$H_{ijn} \sim \text{Gamma}(\beta, \beta G_{ijn}) \quad E[H_{ijn}] = G_{ijn}^{-1}$$



関数形的设计

- ソース W ・ フィルタ A に対する関数形
 - **くし型関数** [後藤2001, 亀岡2007, 安良岡2012など]
 - 急峻なピークを持つガウス関数を等間隔に配置
 - **全極型伝達関数** [音声分野で多くの文献]
 - なららかな周波数応答をもち、複数のピークが存在



モデルの学習

- 統一的なコスト関数のもとでの反復最適化

$$X_{mn} \approx \sum_{i,j}^{\infty} \theta_i \phi_j W_{im} A_{jm} H_{ijn}$$

ソース*i* フィルタ*j* ペアの音量

- $p(X)$ が単調増加するようにモデルを更新

- 変分ベイズ法 (VB)

- グローバルな重み θ, ϕ の事後分布
- 音量 H の事後分布

- 乗法更新則 (MU)

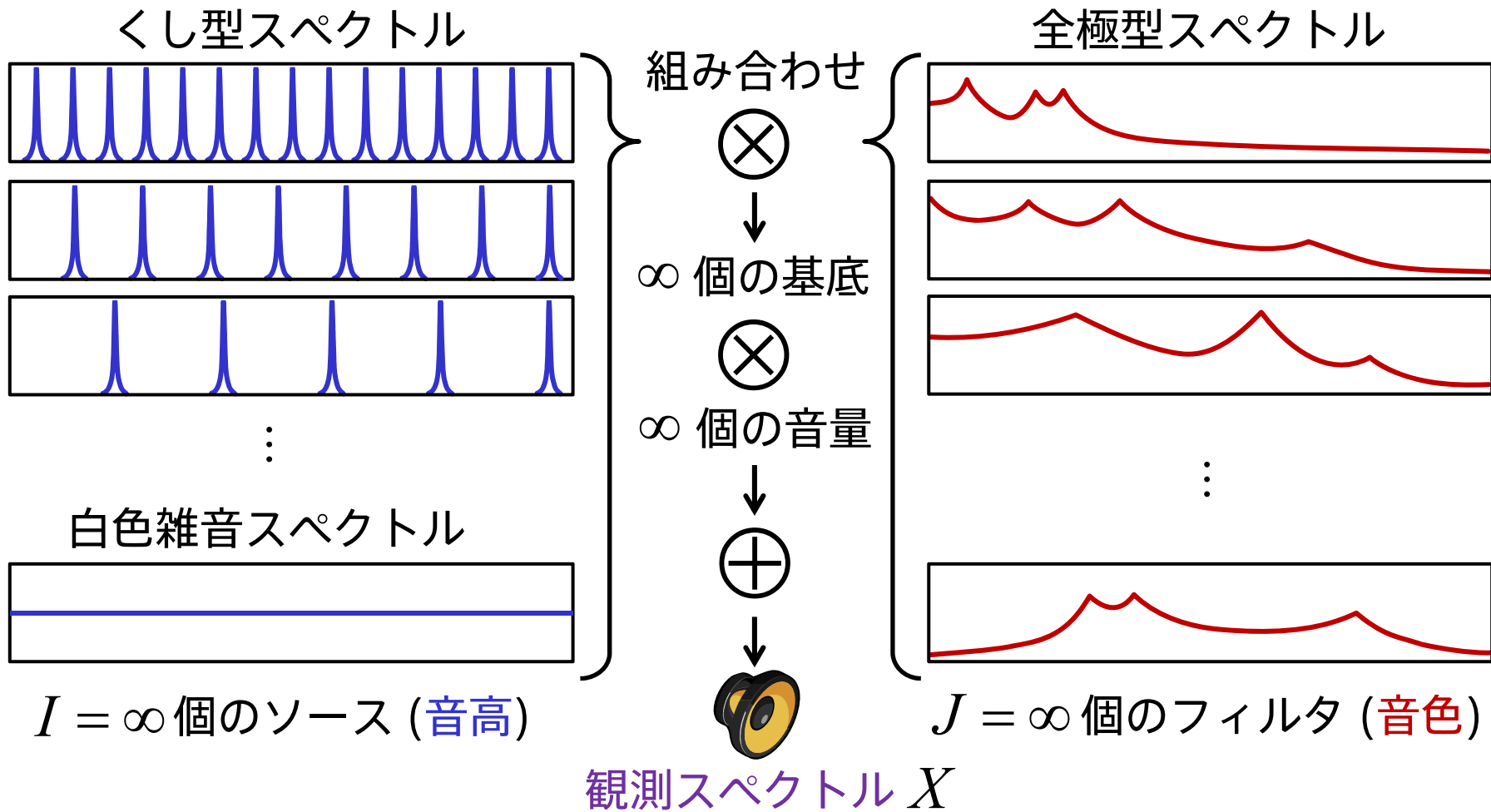
- ソース W のパラメータ
 - » 基本周波数・分散
- フィルタ A のパラメータ
 - » 線形予測係数

VBとMUの統合

本研究が初めて

無限複合自己回帰モデル

- 混合音スペクトルのソース・フィルタへの分解



評価実験

- ピアノ曲の基本周波数推定実験

- 実験条件


- MAPSピアノデータベース
 - 30秒 x 30曲を利用
 - 16 [kHz]・モノラルに変換
 - 窓幅2048点・シフト長160点で短時間フーリエ変換
 - 1024 x 3000 の観測スペクトログラム

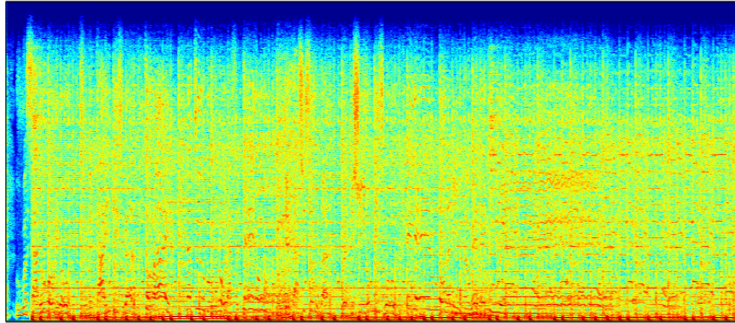
- 推定精度

- KL-Divergence 無限複合自己回帰モデル : 48.4%
 - IS-Divergence 無限複合自己回帰モデル : 35.1%

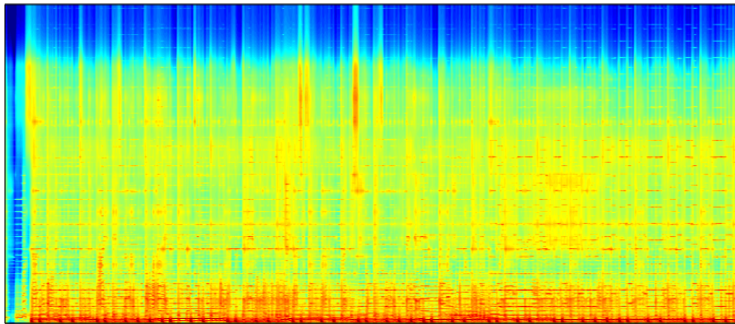
他文献で報告されている60~70%程度には及ばないが、今後のモデルの改良により十分達成可能と考えられる

音源分離 : RWC-MDB-P-2001 No. 1

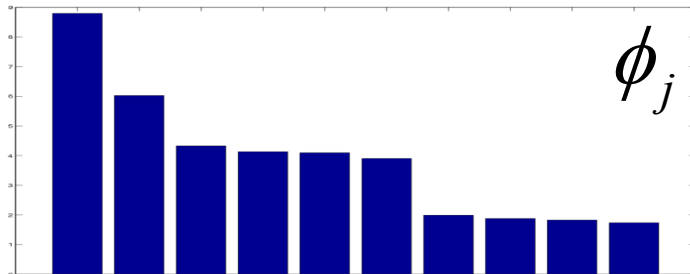
観測スペクトログラム 



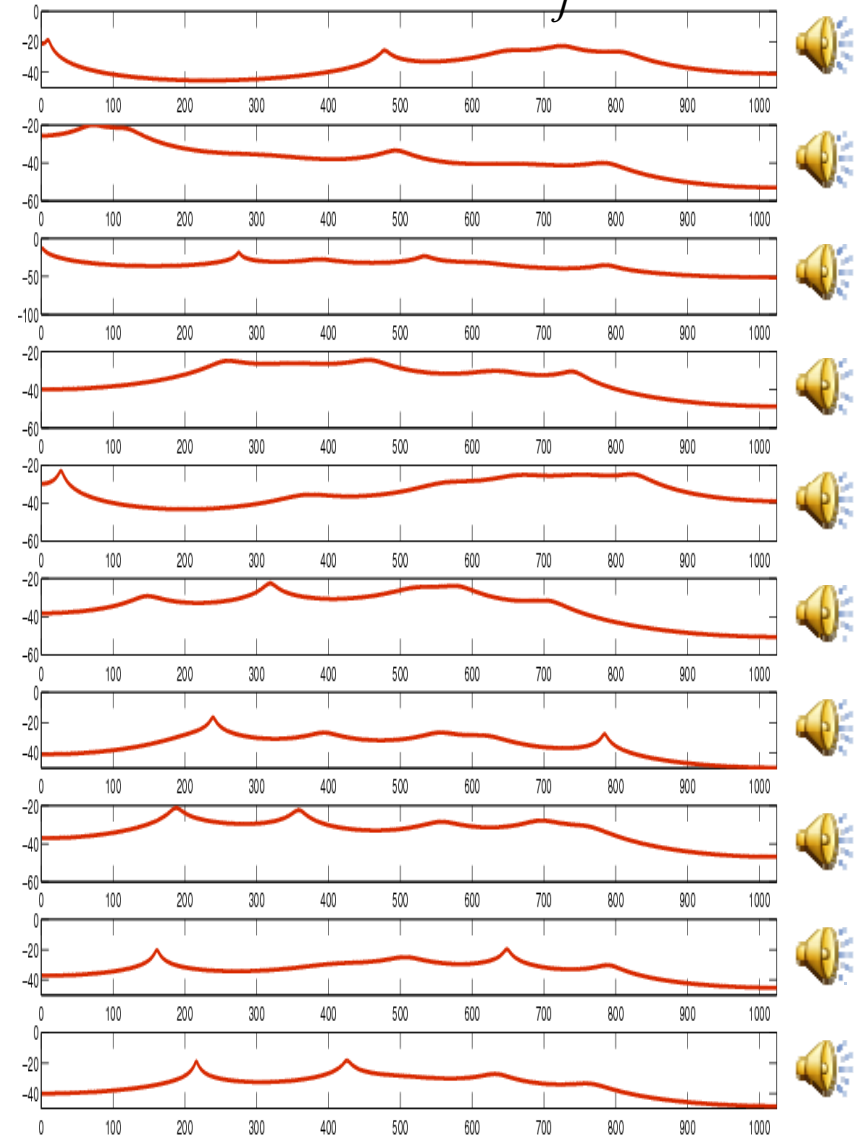
再構成スペクトログラム




フィルタの重み

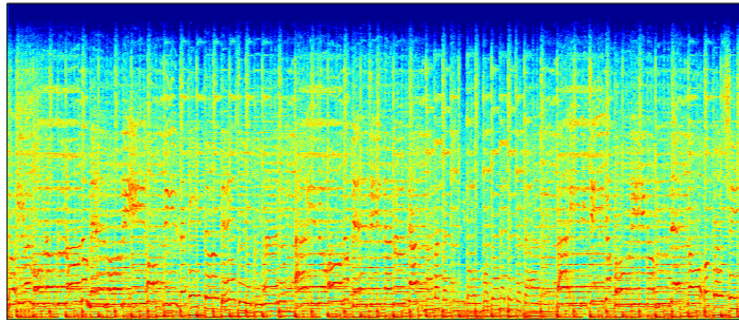


フィルタ A_j

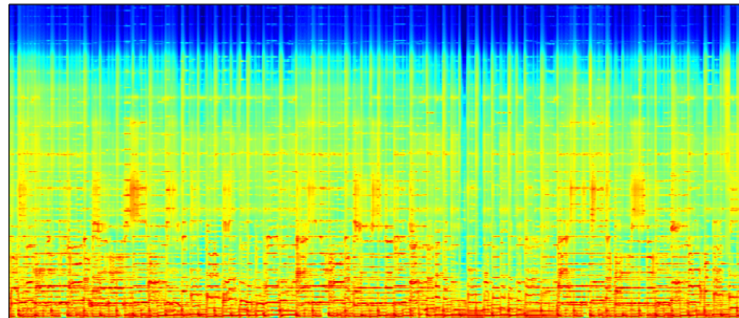


音源分離 : RWC-MDB-P-2001 No. 5

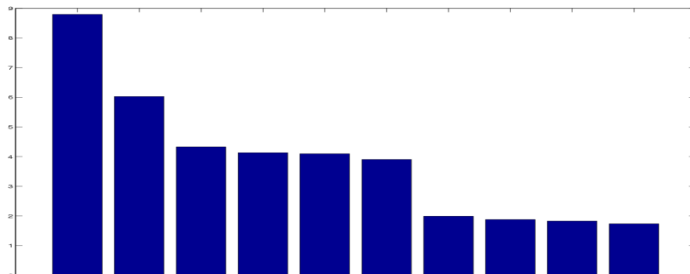
観測スペクトログラム 



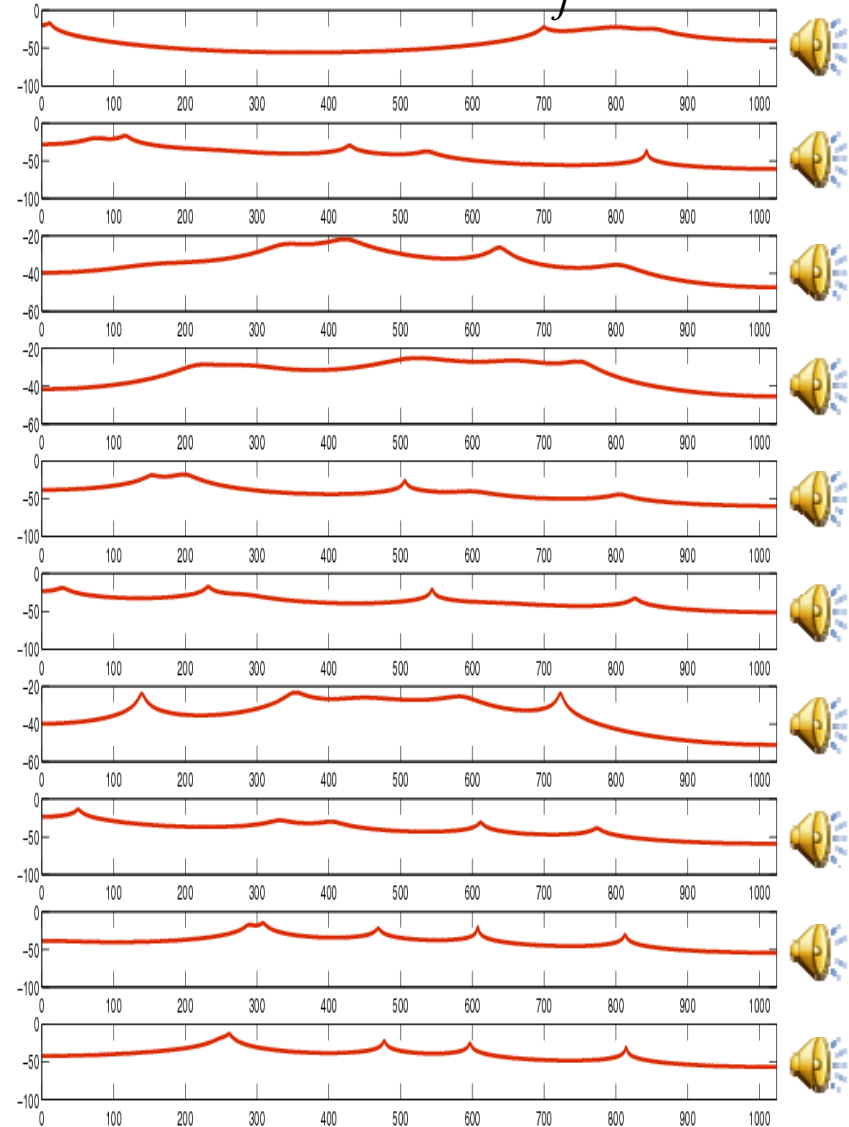
再構成スペクトログラム




フィルタの重み

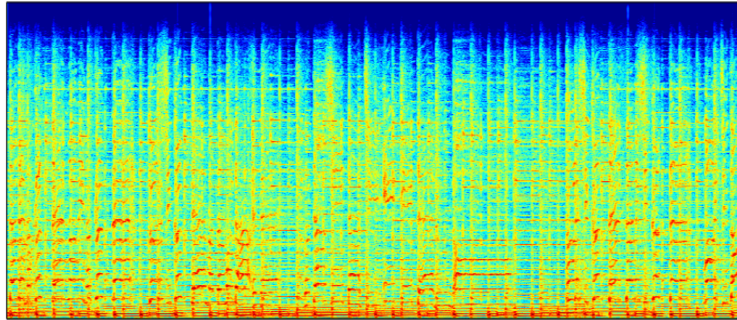


フィルタ A_j

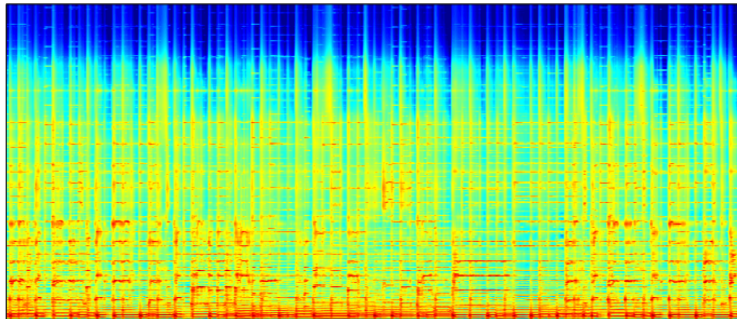


音源分離 : RWC-MDB-P-2001 No. 8

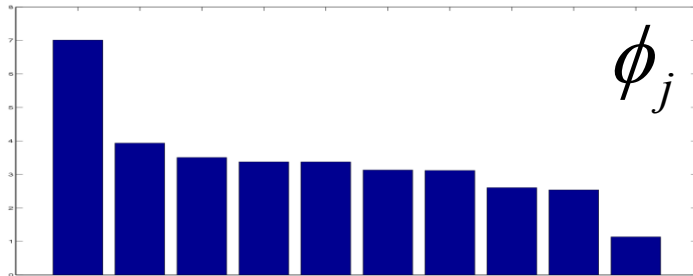
観測スペクトログラム 



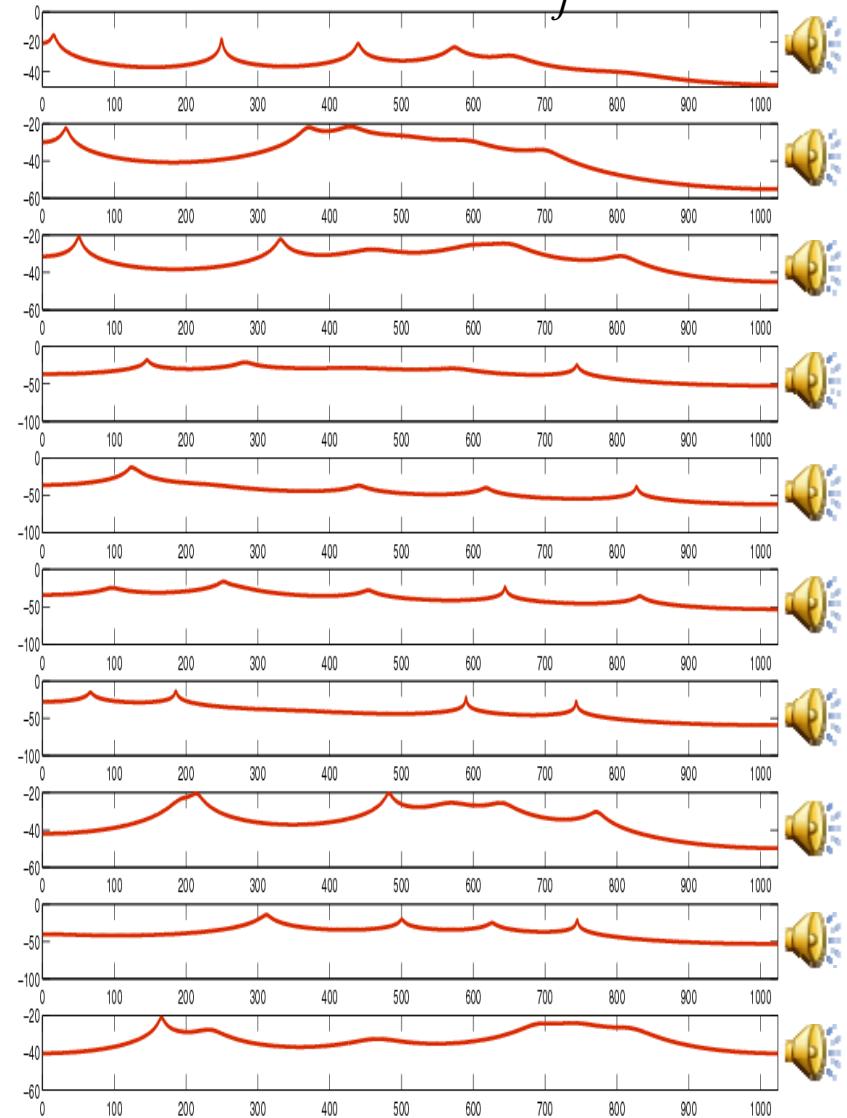
再構成スペクトログラム



フィルタの重み



フィルタ A_j



今後の課題

- 自動採譜精度の高精度化

- 確率モデルの再定式化

- 人間の知覚スケールに合わせる
 - 対数周波数領域
 - 対数パワー領域

- フィルタの音高依存性を考慮

- ピアノ・ギター・打楽器
 - 反響板の特性で音色が決まる → フィルタは一定
 - 管楽器
 - 音響管の長さが変化 → フィルタは可変

- 音源分離への取り組み

- 複雑な音響信号の楽器パート分離

- 例：ポピュラー音楽 → ボーカル＋ギター＋ベース＋打楽器