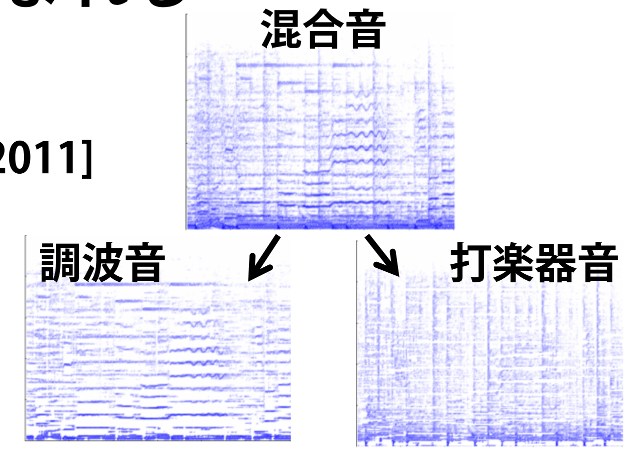
多重音基本周波数解析のための無限重畳離散全極型モデル

吉井 和佳　糸山 克寿 (京大)
後藤 真孝 (産総研)

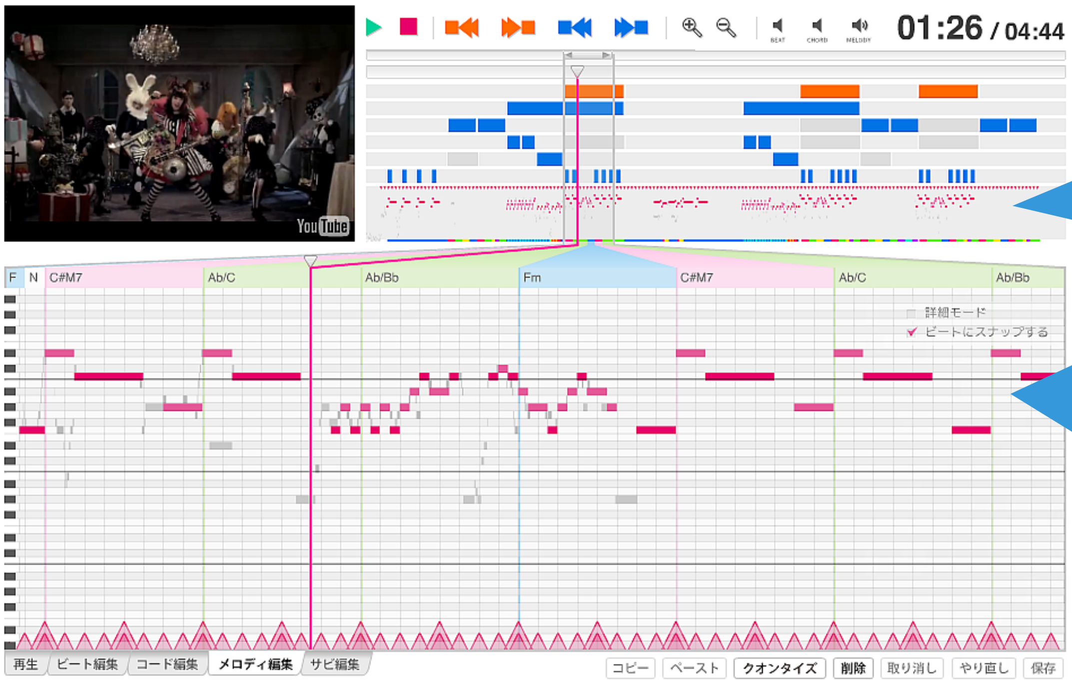
研究の背景

- 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
    - 調波音・打楽器音分離 [宮本2008, Rigaud2011]
    - 楽器音イコライザ [糸山2008, 中村2014]
  - 音楽制作の楽しみ方を広げたい
    - マッシュアップ・二次創作
    - 既存楽曲を構成する「パーツ」の組み合わせ



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

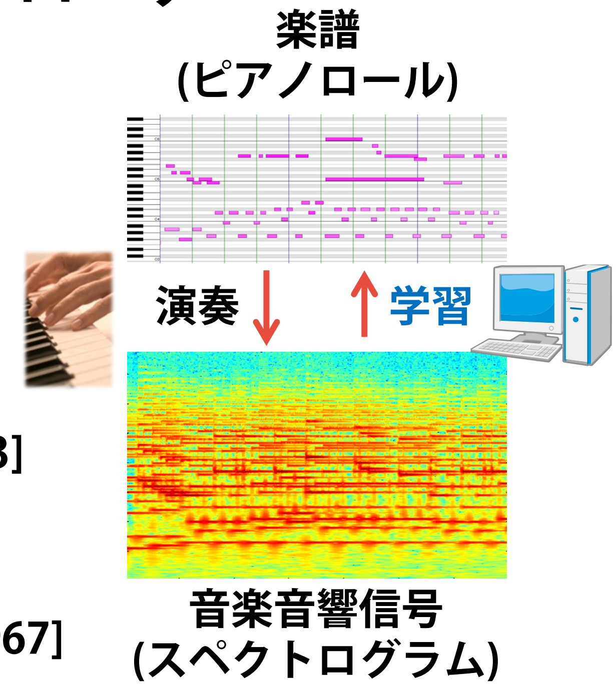


ビート・コード・メロディ・繰り返し構造の自動解析が可能

今はメロディ(歌唱)の音高だけが解析対象
↓
全楽器パートが解析できるように拡張したい

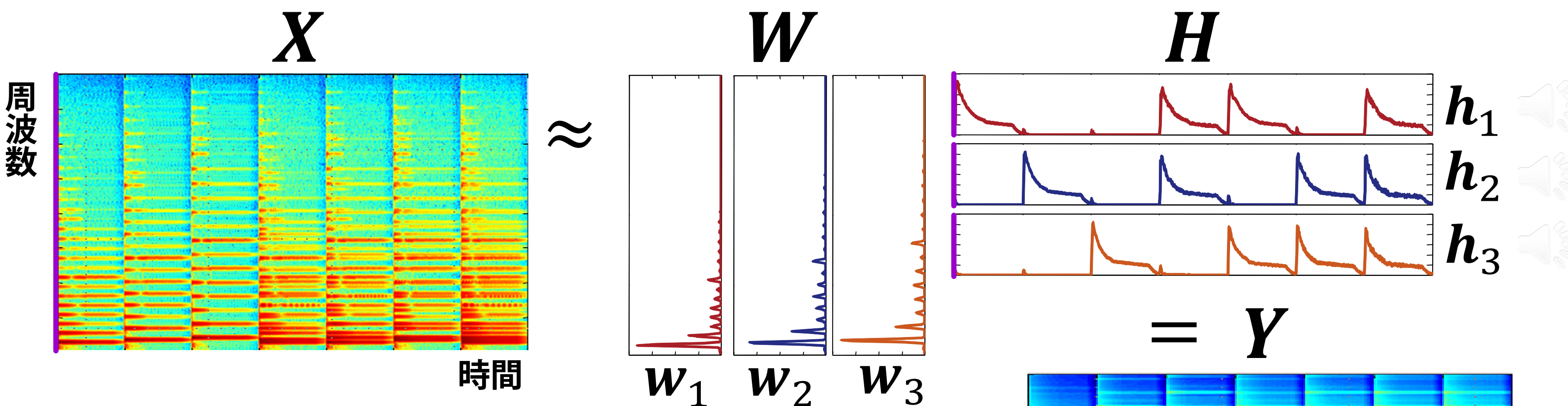
統計的音楽音響信号処理

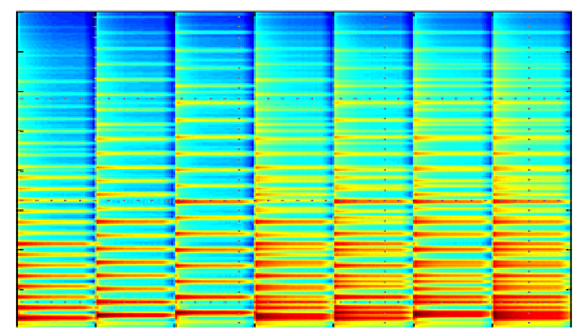
- 音楽音響信号の**生成モデル**に基づくアプローチ
  - 順問題：生成モデル → 音楽音響信号
  - 逆問題：音楽音響信号 → 生成モデル
- 様々な**生成モデル**を確率的に統合可能
  - 非負値行列分解 (NMF)
    - 音楽音響信号のパーツ分解 [Smaragdis 2003]
  - 線形予測分析 (LPC) / 自己回帰過程 (AR)
    - 音声信号のスペクトル包絡推定 [Itakura 1967]
  - 複合自己回帰モデル (CAR)
    - NMFとLPCの組み合わせ [Kameoka 2009, Yoshii 2012]



非負値行列分解 (NMF)

- 音源分離/自動採譜のための強力なアプローチ
  - 振幅スペクトログラムを**基底**と音量とに分解 [Smaragdis 2003]



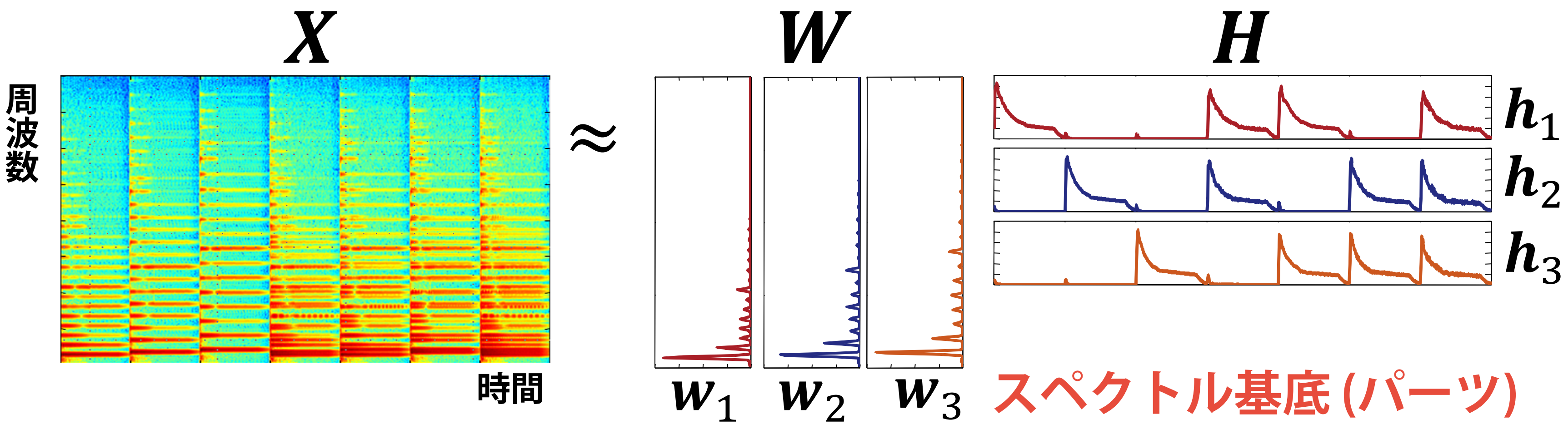


多重音を構成する「パーツ」

観測 X と 再構成 $WH = Y$ との距離を
最小化するように W と H を学習

NMFの問題点

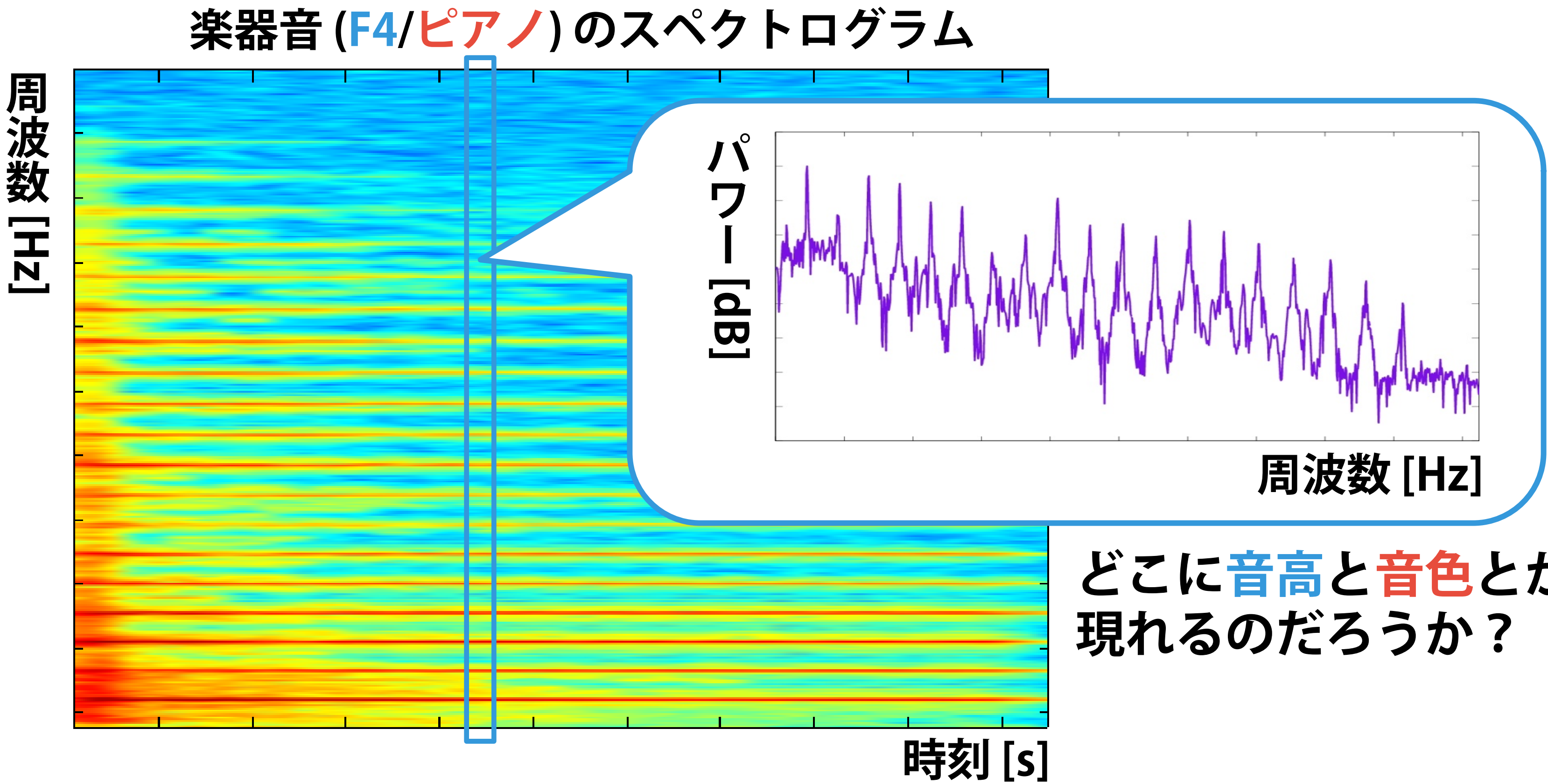
- 観測スペクトログラムが異なる音高ごとに分離される
  - 本当は「音高」ではなく「音色」で分離したい (楽器パート分離)
  - スペクトル基底を音色ごとにクラスタリングするのは難しい



音楽音響信号を構成する「パーツ」とは何かについて再考が必要

楽器音スペクトル

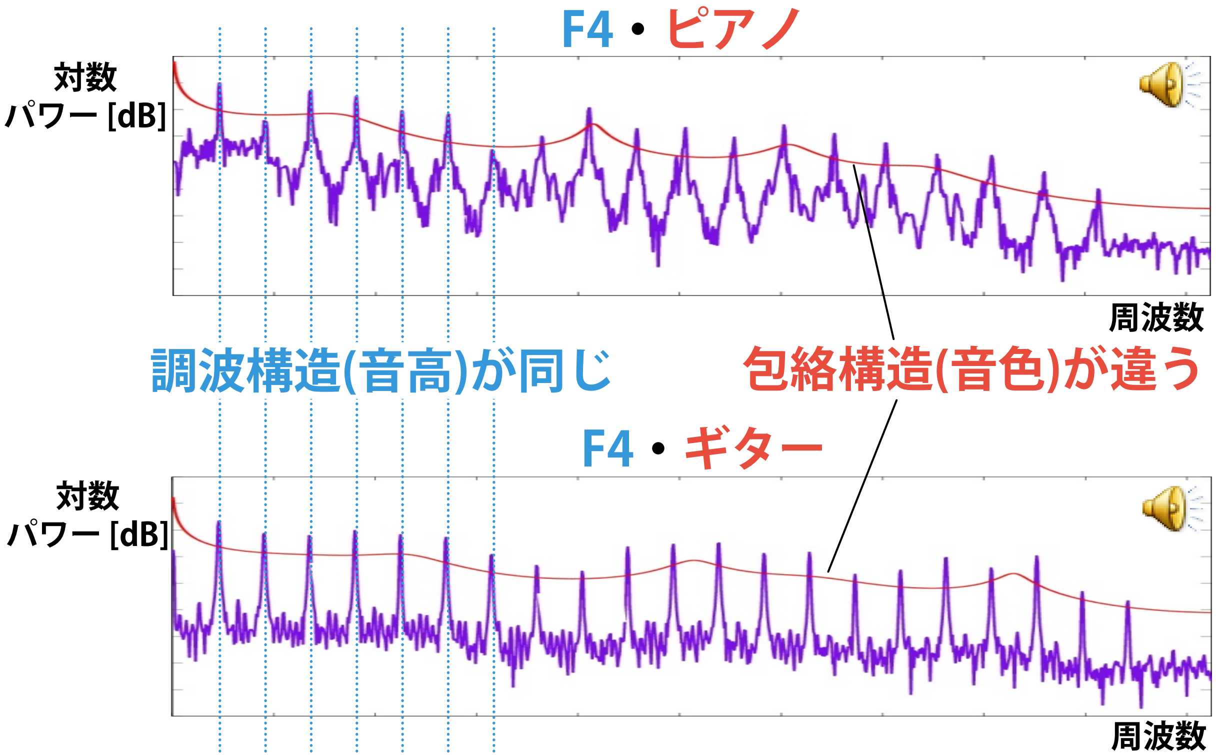
- 短時間フーリエ変換 (STFT) による周波数解析



どこに音高と音色とが現れるのだろうか？

楽器音の生成モデル

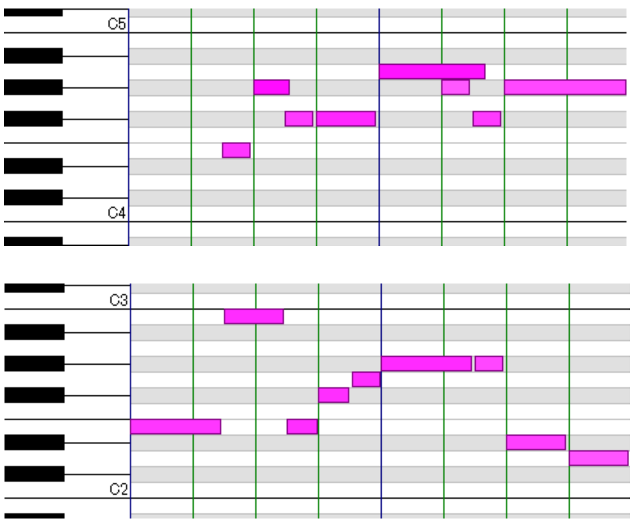
- ソース・フィルタ理論に基づくモデル化
  - 楽器音スペクトル＝調波構造＋包絡構造



多重音の生成モデル

- 音楽音響信号を「パーツ」の組み合わせで表現
  - 楽器音：2種類の要素「音高」と「音色」の組み合わせで構成

音高 (ソース)

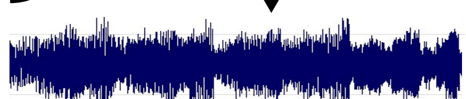


様々な楽器音
パーツを生成

⊗

重ね合わせ

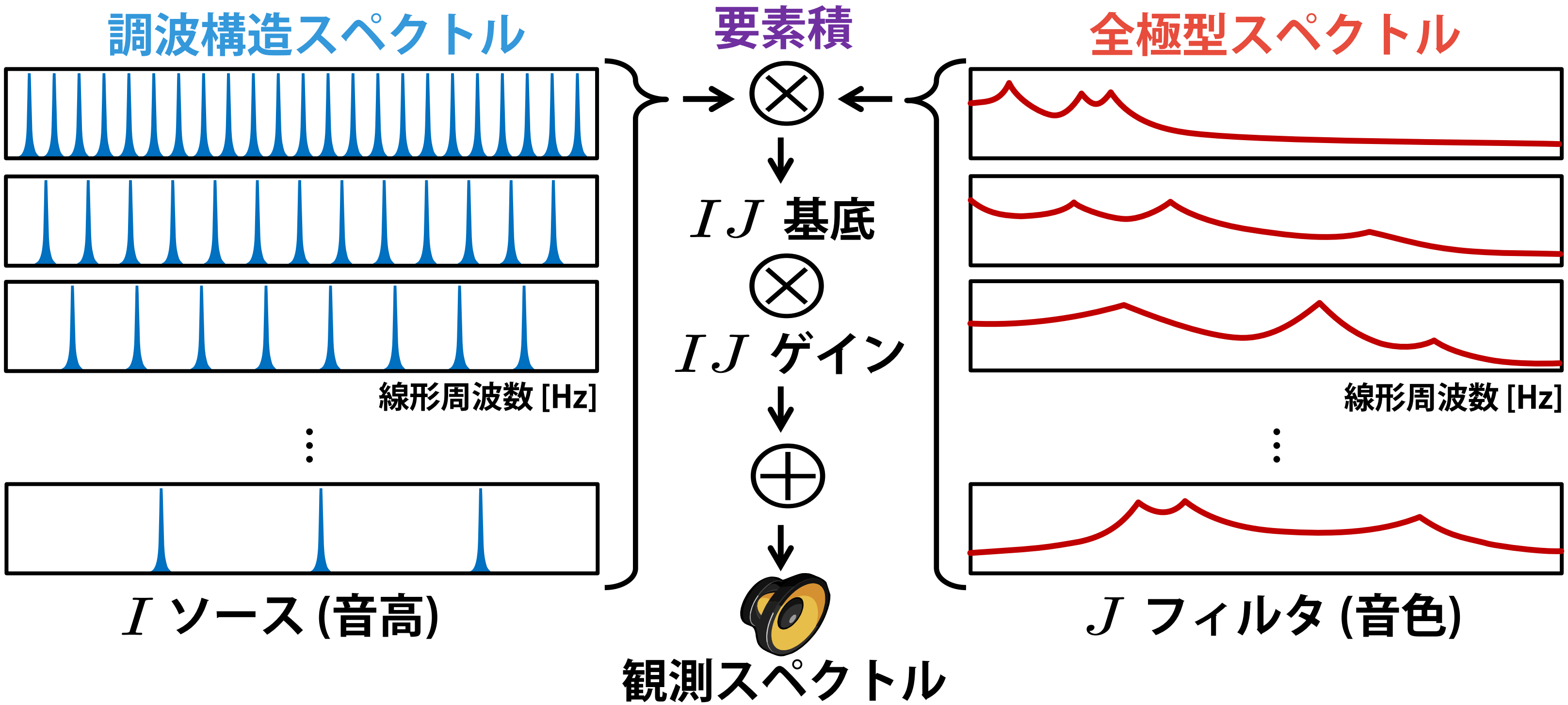
音色 (フィルタ)



混合音

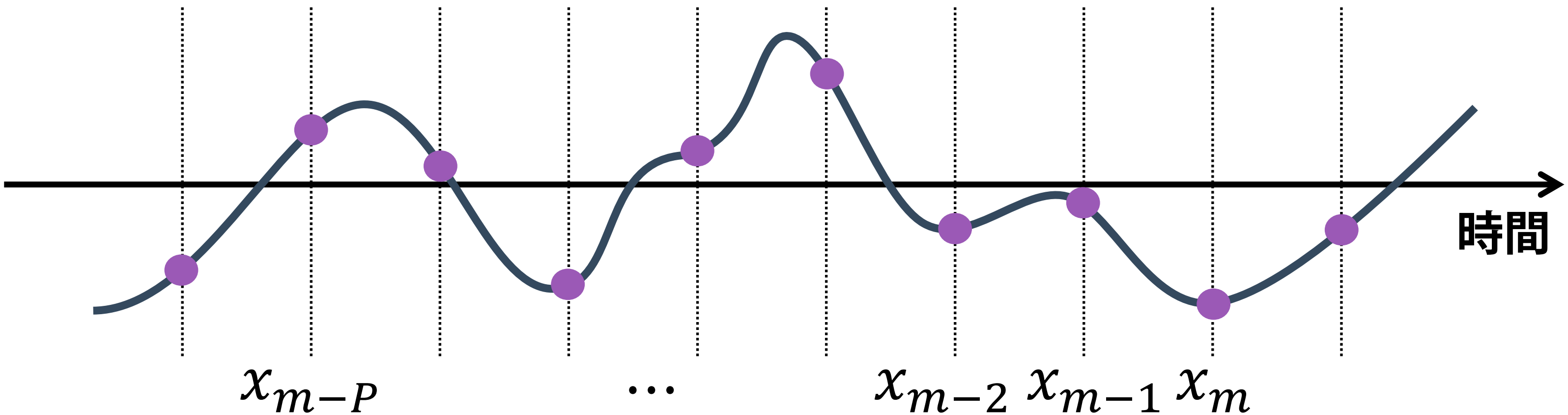
複合自己回帰モデル (CAR)

- 多重音をソースとフィルタの組み合わせで表現 [Kameoka 2009]
  - 無限個のソースとフィルタを仮定する拡張も可能 [Yoshii 2012]



楽器音の生成モデル

- 楽器信号が自己回帰 (AR) 過程に従うと仮定
  - **線形予測分析**：過去のサンプルの線形結合で次のサンプルを予測



$$x_m = \sum_{p=1}^{P} \hat{a}_p x_{m-p} + s_m$$

観測信号　予測係数　予測誤差

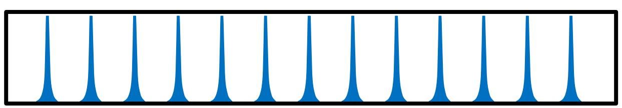
ソース・フィルタ理論の観点では音源信号に対応

楽器音の生成モデル

- **楽器音スペクトル生成モデルのフーリエ領域表現**
  - 時間領域での畳み込み → フーリエ領域での要素積

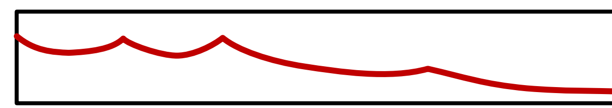
$$x_m = \sum_{p=1}^{P} \hat{a}_p x_{m-p} + s_m \rightarrow s_m = \sum_{p=0}^{P} a_p x_{m-p}$$

観測信号　予測係数　音源信号

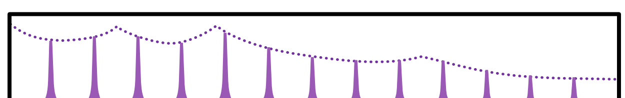


$\{S_m\}_{m=1}^{M}$ 周波数m

要素積



$\{F_m\}_{m=1}^{M}$ 周波数m



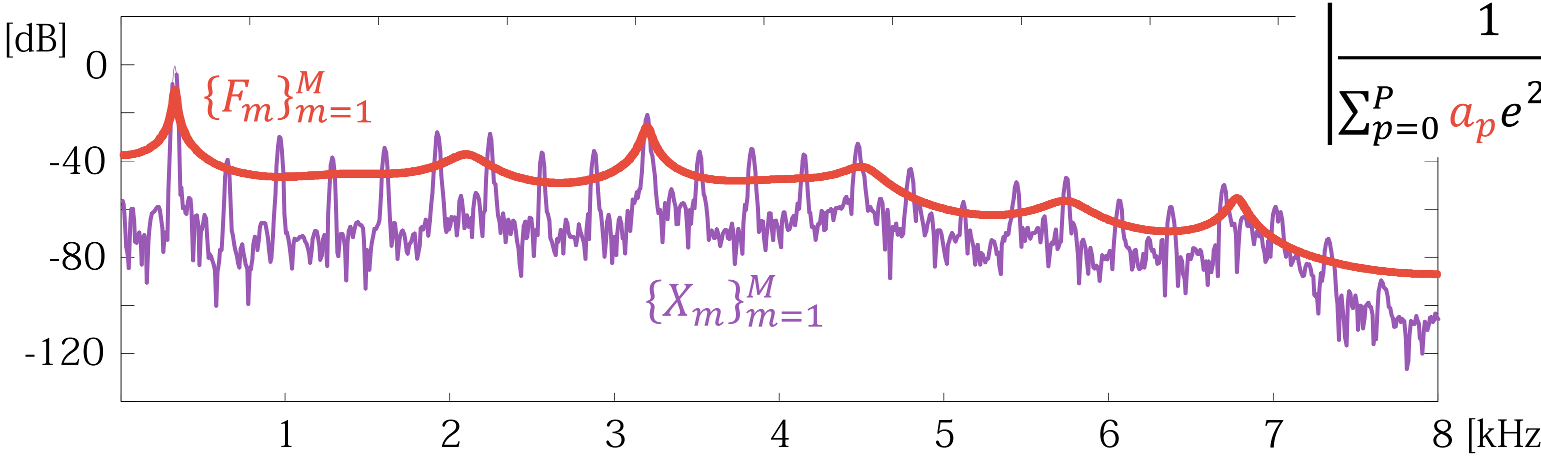
$\{X_m\}_{m=1}^{M}$ 周波数m

$$= \left| \frac{1}{\sum_{p=0}^{P} a_p e^{2\pi i \frac{m}{M} p}} \right|$$

線形予測分析 (LPC)

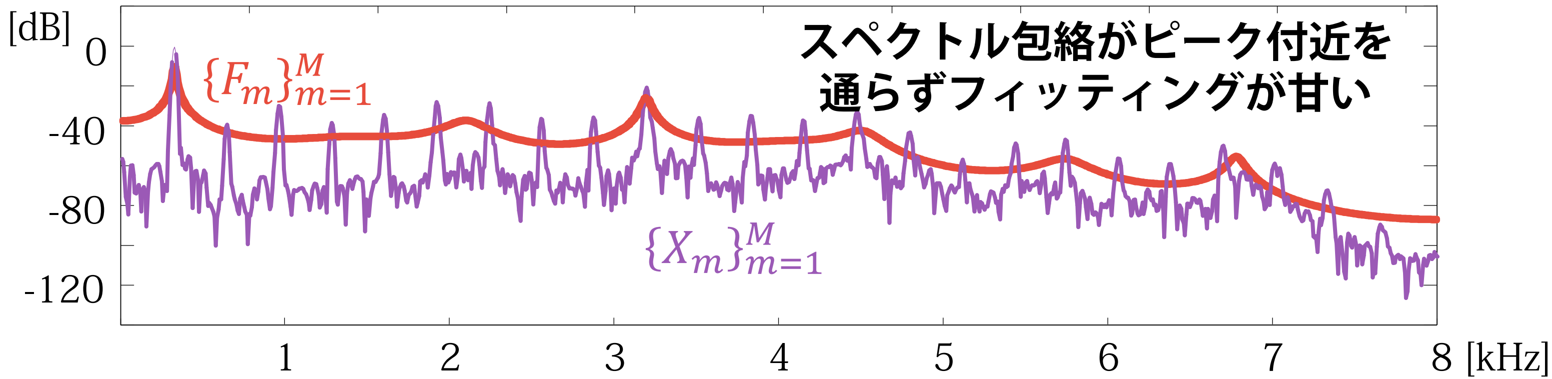
- 観測スペクトル$\{X_m\}_{m=1}^{M}$からスペクトル包絡$\{F_m\}_{m=1}^{M}$を推定
  - 音源信号にガウス性白色雑音を仮定 (パワースペクトルは指数分布)
  - 生成モデルの最尤推定としての解釈 [Itakura 1967]

$$X_m = S_m F_m$$
$$S_m \sim \mathrm{Exponential}\,(\sigma^2)$$
$$\longrightarrow X_m \sim \mathrm{Exponential}\,(\sigma^2 F_m)$$
$$F_m = \left| \frac{1}{\sum_{p=0}^{P} a_p e^{2\pi i \frac{m}{M} p}} \right|^2$$



LPCの問題点

- 音源信号にガウス性白色雑音を仮定　$X_m = S_m F_m$
  - 音源スペクトル$\{S_m\}_{m=1}^{M}$自体を緻密にモデル化したい
- 全周波数帯域の利用
  - 調波構造のピークだけでスペクトル包絡を推定したい
- 線形周波数領域での定式化
  - 人間の聴覚特性に合致した対数周波数領域で定式化したい

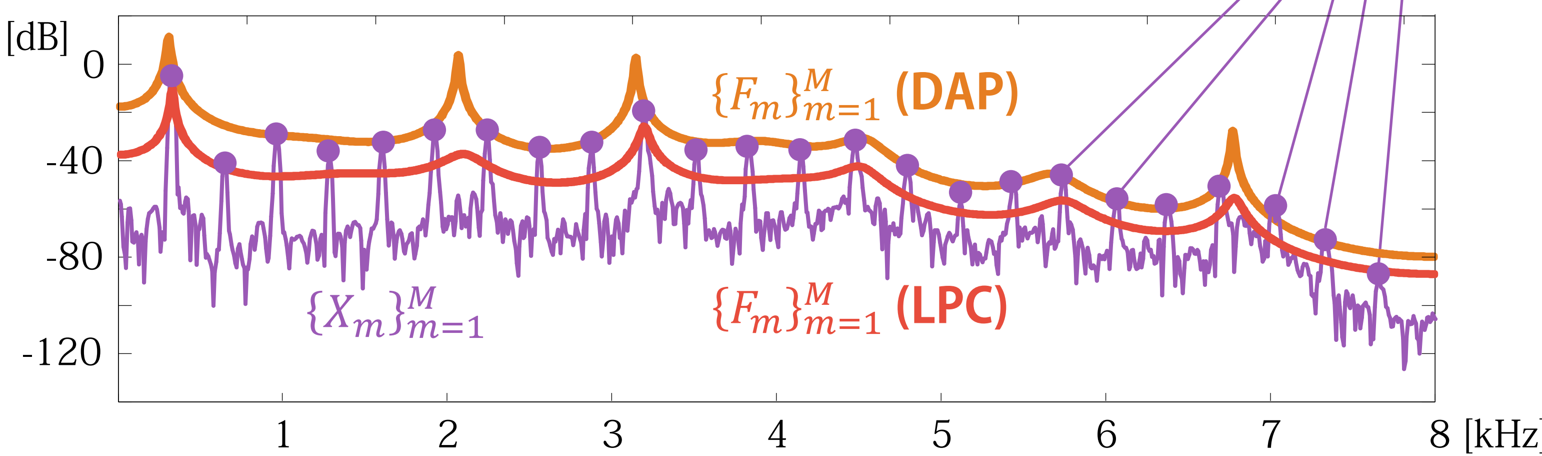


離散全極型モデル (DAP)

- **LPCを改良したスペクトル包絡推定法** [El-Jaroudi 1991]
  - 利点：調波構造ピークのみを考慮 / 対数周波数領域に適用可能
  - 欠点：<u>音源信号</u>に対する基本周波数推定が必要

$$X_m \sim \mathrm{Exponential}\left(\sigma^2 F_m\right) \quad m \in \Omega$$

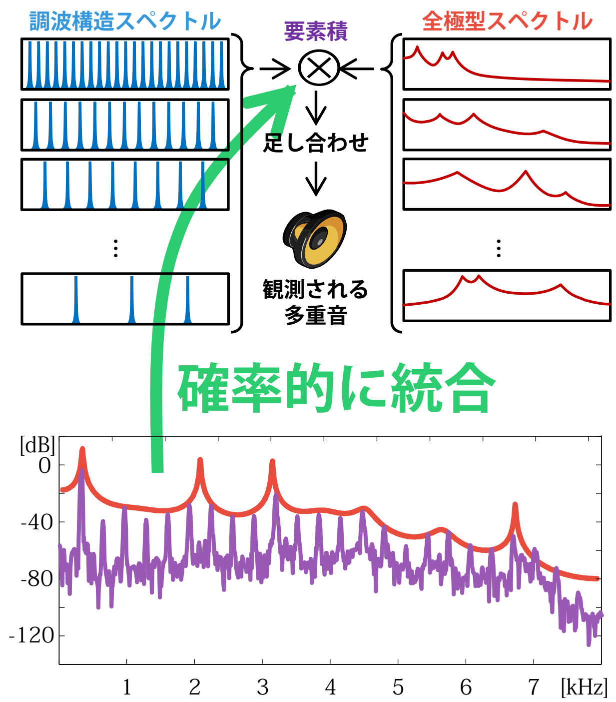
考慮すべき
周波数ビンの集合



提案するアプローチ

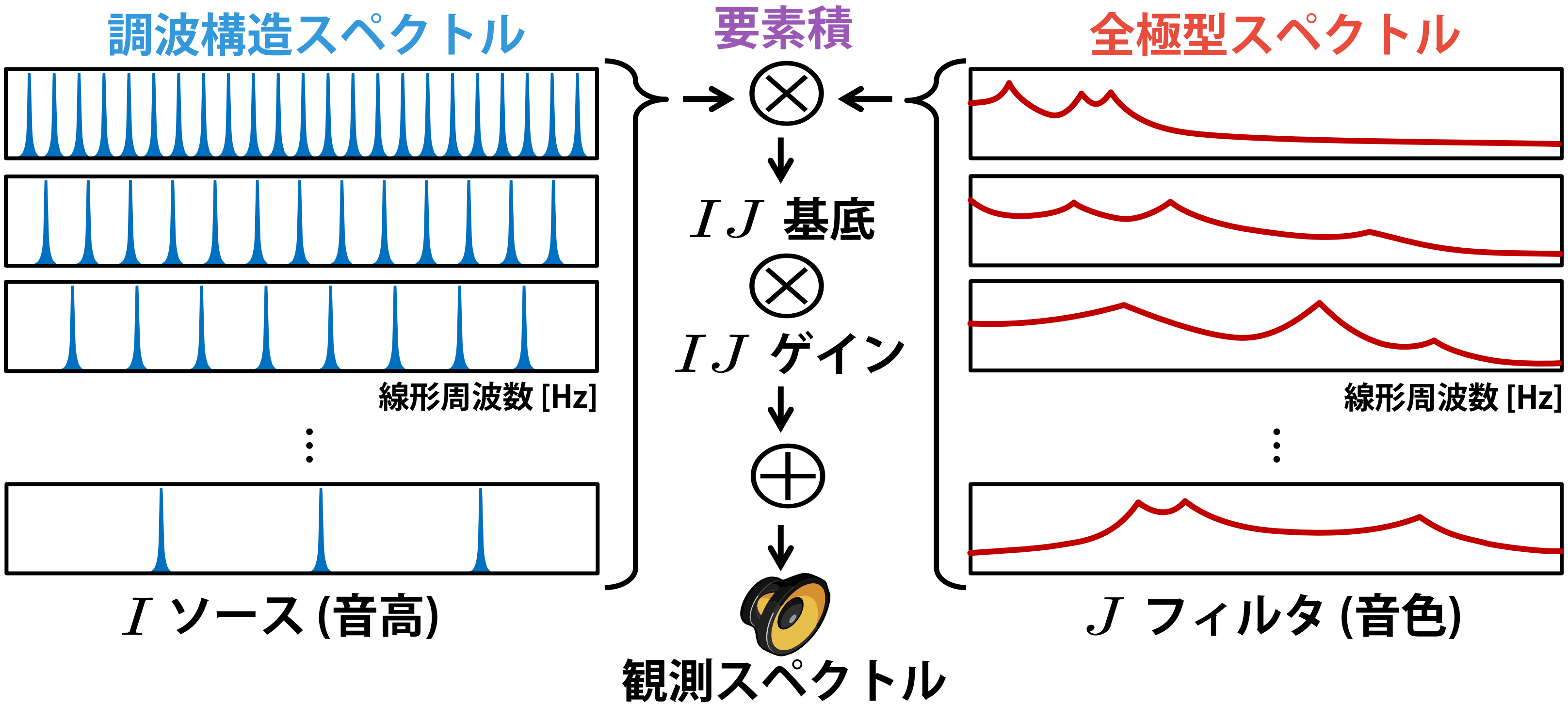
- **複合自己回帰モデル + 離散全極型モデル**
  - 確率的な枠組みのもとで統一的な確率モデルを定式化

	複合自己回帰モデル (CAR)	離散全極型モデル (DAP)
多重音の取り扱い	複数のソース・フィルタに分解可能	×
音源信号の取り扱い	音源スペクトル自体を推定可能 (基本周波数も推定可能)	×
対数周波数の取り扱い	×	周波数軸によらずに定式化可能



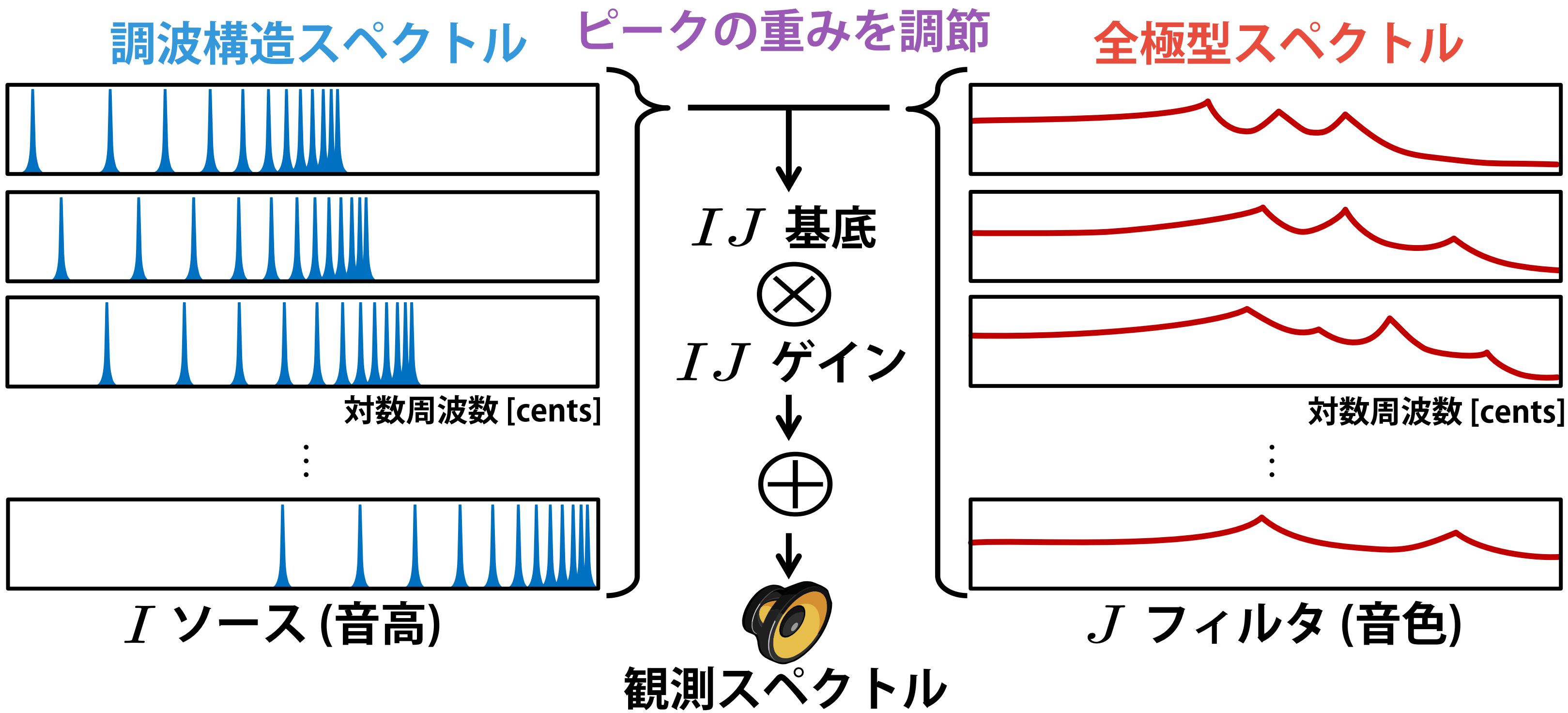
複合自己回帰モデル (CAR)

- 線形周波数領域で定式化されたソース・フィルタNMF
  - 無限個のソースとフィルタを仮定する拡張も可能 (iCAR)



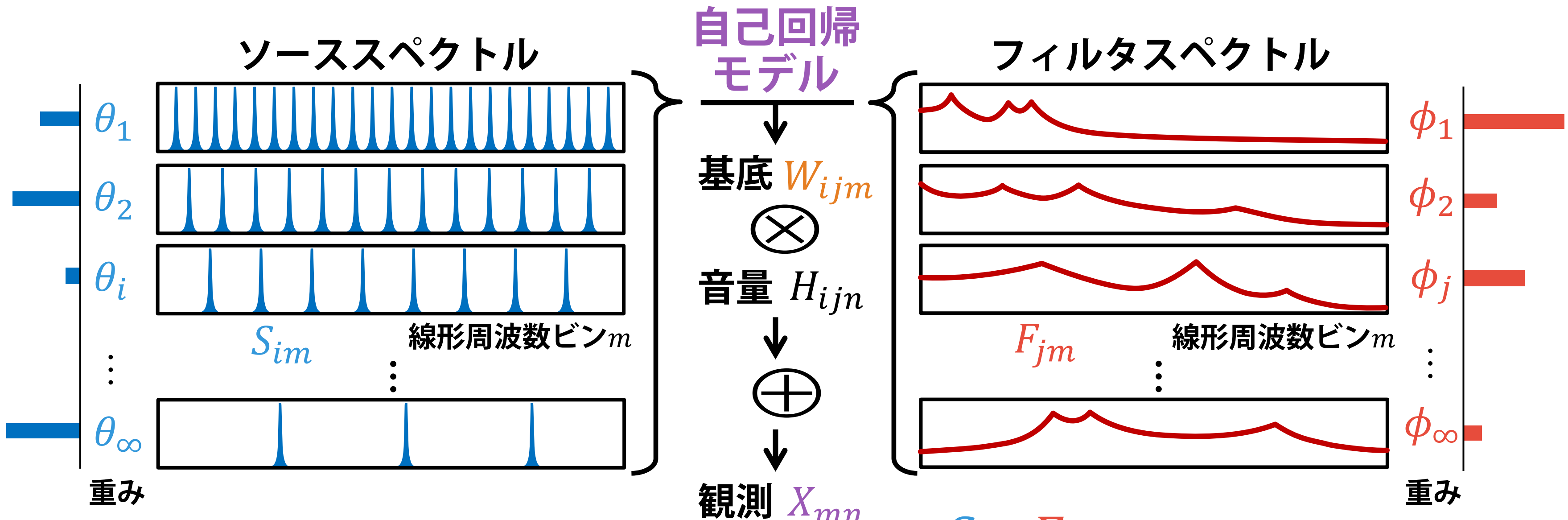
重畳離散全極型モデル (SDAP)

- 対数周波数領域で定式化されたソース・フィルタNMF
  - 無限個のソースとフィルタを仮定する拡張も可能 (iSDAP)



iCARの確率モデル

- NMFの拡張：ソース・フィルタモデル + 無限モデル

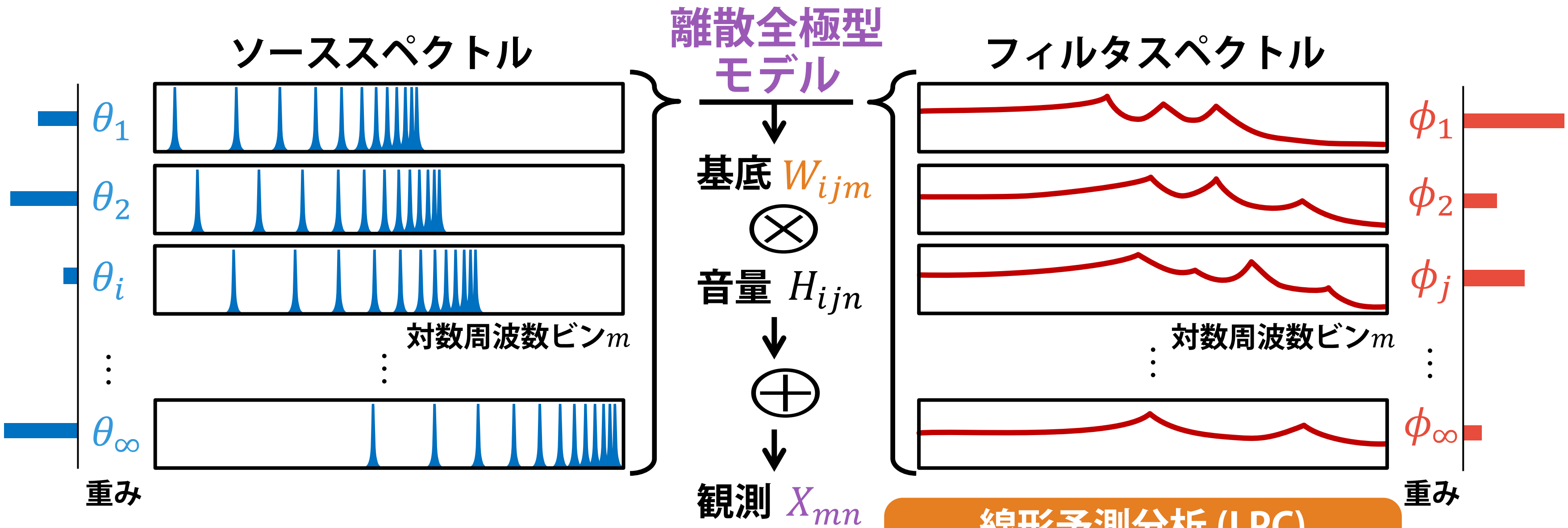


$S_{im}F_{jm}$

$$X_{mn} \sim \mathrm{Poisson}\left(\sum_{i=1}^{I\to\infty}\sum_{j=1}^{J\to\infty}\theta_i\phi_j W_{ijm}H_{ijn}\right)$$

iSDAPの確率モデル

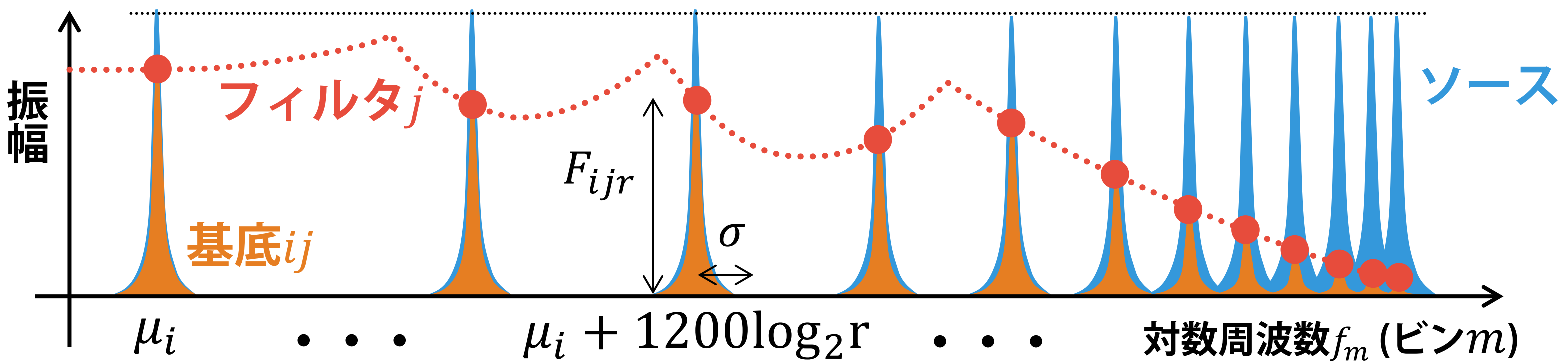
- iCARと同じ形式：ソース・フィルタモデルの改良



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

スペクトル基底のモデル

- スペクトル基底に対する<u>離散</u>全極型モデル (DAP)
  - 各倍音成分の重みをスペクトル包絡で決定



$$W_{ijm} = \sum_{r=1}^{R} S_{irm} F_{ijr} \begin{cases} S_{irm} = e^{-\frac{1}{2\sigma^2}(f_m - (\mu_i + 1200\log_2 r))^2} & \text{ガウス関数 (倍音成分)} \\ F_{ijr} = \dfrac{1}{\left|\sum_{p=0}^{P} a_{jp} e^{-\omega_{ir} p i}\right|} & \text{全極型関数 (倍音の重み)} \end{cases}$$

ベイズ的取り扱い

- ノンパラメトリックベイズモデルの定式化

ポアソン尤度

$$X_{mn} \sim \mathrm{Poisson}\left(\sum_{i=1}^{I\to\infty}\sum_{j=1}^{J\to\infty}\theta_i\phi_j W_{ijm}H_{ijn}\right)$$

モデルの学習

変分ベイズ法 (VB)
＋
乗法更新 (MU)

ガウス関数 (倍音スペクトル)

$$S_{irm} = e^{-\frac{1}{2\sigma^2}(f_m-(\mu_i+1200\log_2 r))^2}$$

全極型関数 (倍音の重み)

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

ガンマ過程事前分布

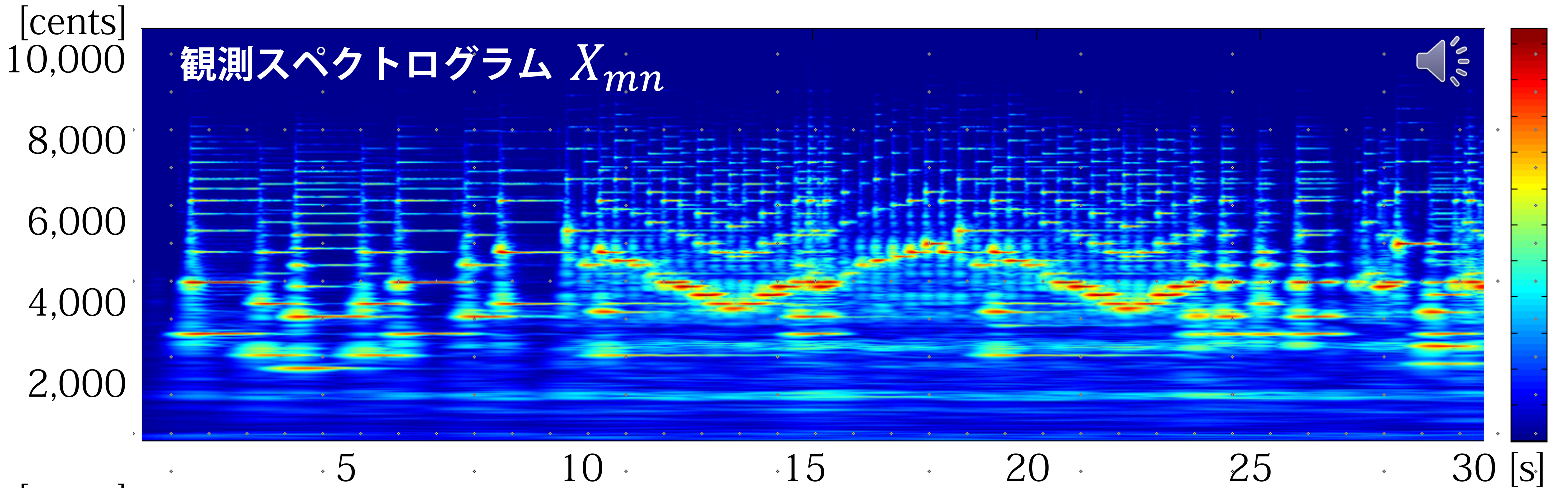
$$\theta_i \sim \mathrm{GaP}(a_\theta, \mathbf{1})$$

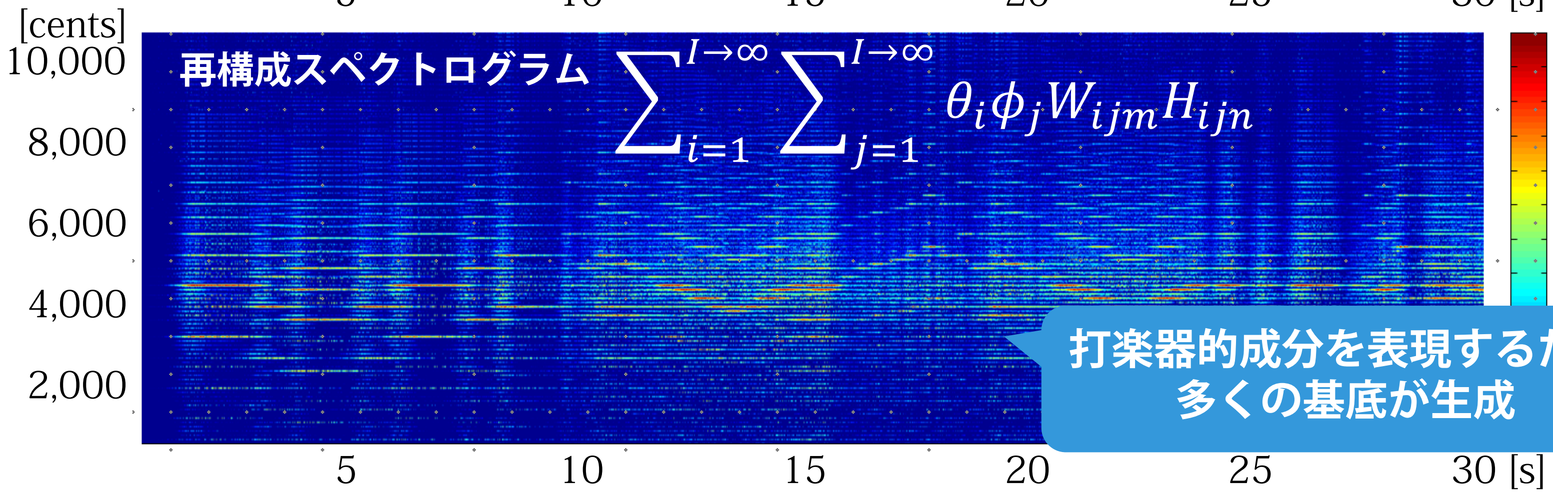
$$\phi_j \sim \mathrm{GaP}(a_\phi, \mathbf{1})$$

ガンマ事前分布

$$H_{nij} \sim \mathrm{Gamma}(a_H, b_H)$$

実験結果

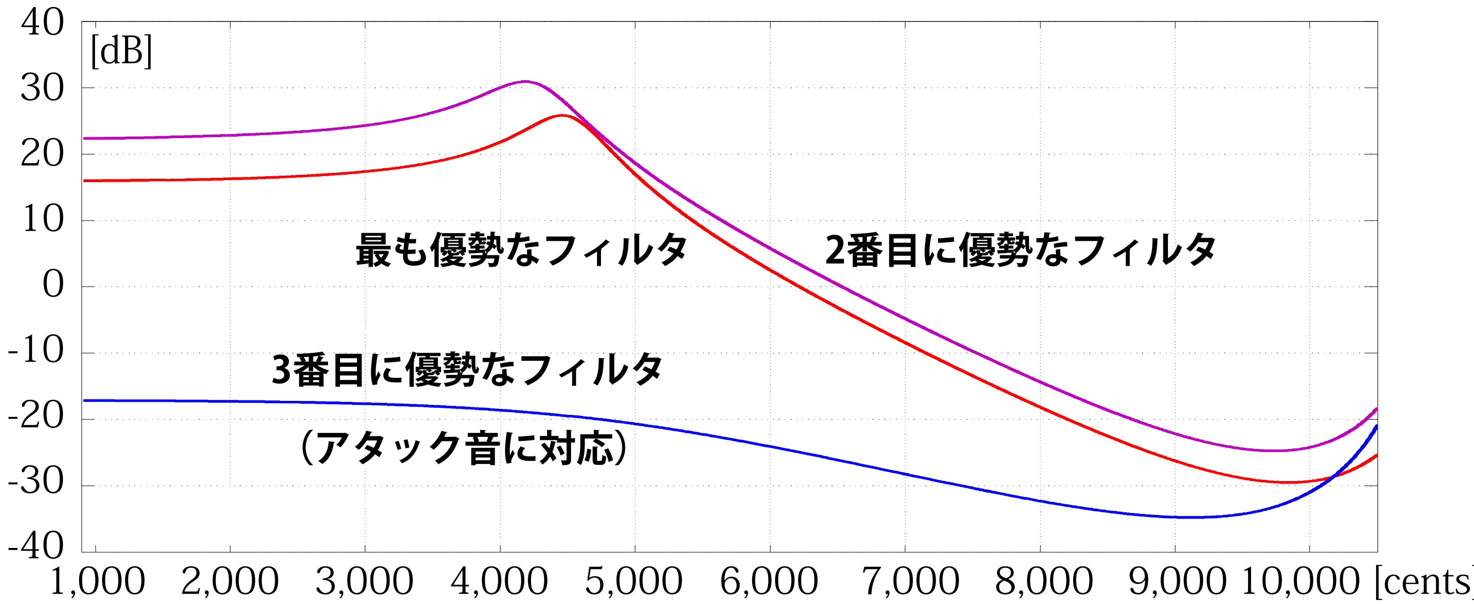




打楽器的成分を表現するため
多くの基底が生成

フィルタの推定結果

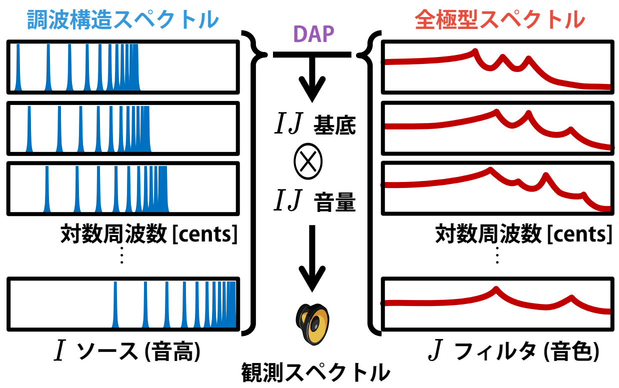
- 調波構造に対してもなめらかなスペクトル包絡が推定可能
  - データから倍音の減衰比率が学習可能
  - 従来法(PreFEst, HTC)では倍音の相対強度に対する事前分布が必要



まとめと今後の課題

- 無限重畳離散全極型モデル (iSDAP) を提案
  - 確率的な枠組みのもとで２つの確率モデルを統合

	複合自己回帰モデル (CAR)	離散全極型モデル (DAP)
多重音の取り扱い	複数のソース・フィルタに分解可能	×
音源信号の取り扱い	音源スペクトル自体を推定可能 (基本周波数も推定可能)	×
対数周波数の取り扱い	×	周波数軸によらずに定式化可能



今後の課題
- 打楽器音の取り扱い
- 楽譜情報に相当する事前分布を導入