

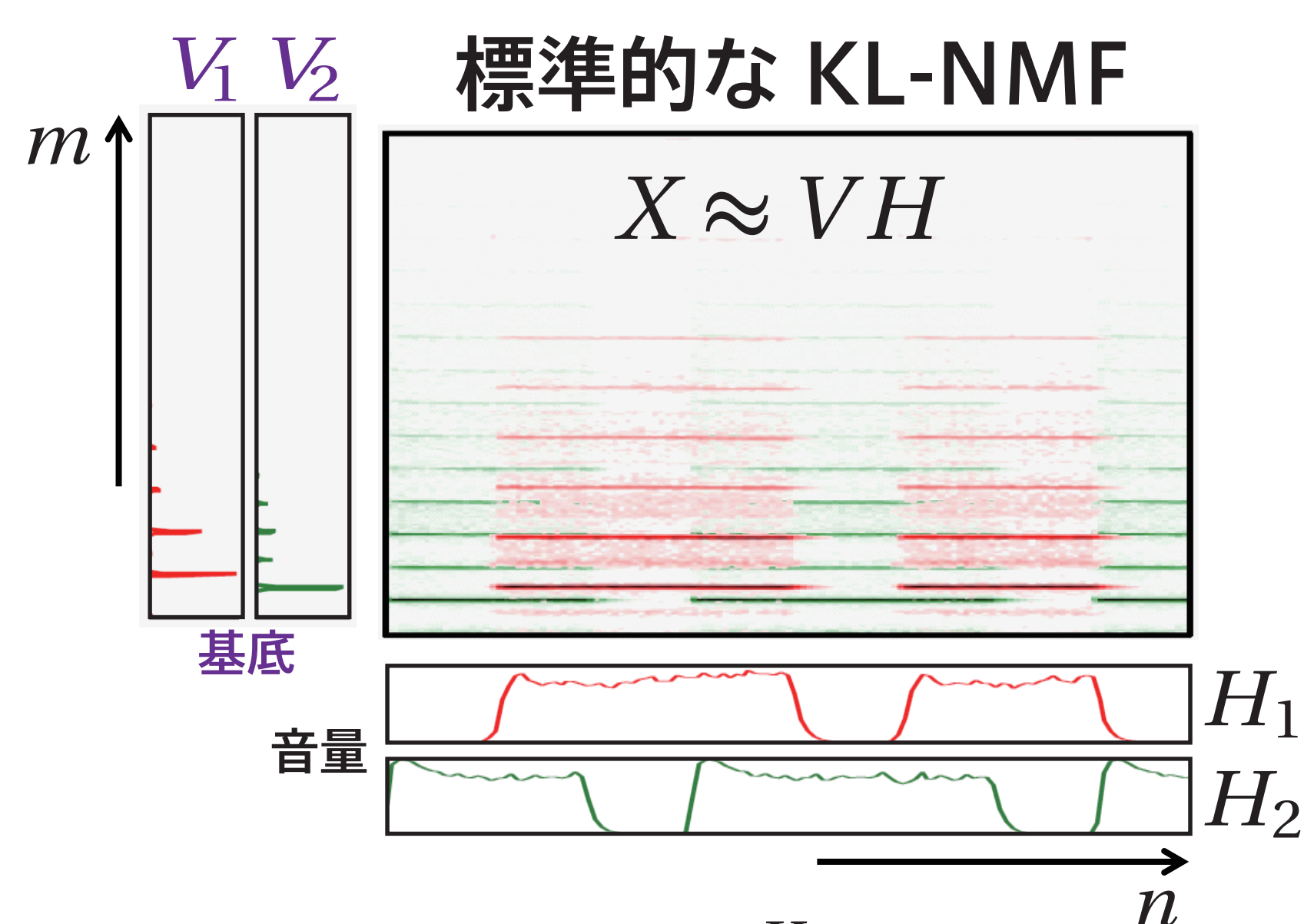


音楽音響信号解析のための ガンマ過程に基づく無限複合自己回帰モデル

吉井 和佳 後藤 真孝
産業技術総合研究所 情報技術研究部門

ノンパラベイズはこわくない!

目的：音楽音響信号に対する複数基本周波数解析・音色に基づく音源分離
標準的な非負値行列因子分解 (NMF) における3つの問題点を解決したい



- (1) モデルの自由度を増やしすぎると過学習しやすくなるにもかかわらず、多様な楽器音の音色をカバーするには多数の独立したスペクトル基底が必要
アイデア：スペクトル基底をソースとフィルタの積に因子分解
- (2) 得られたスペクトル基底それぞれに対して、調波構造の存在を判定し、その音高（基本周波数：F0）を推定する後処理が必要
アイデア：F0をパラメータにもつ関数としてソースを表現
- (3) スペクトル基底の個数を適切に設定することが必要
アイデア：ノンパラメトリックベイズモデルにより自動調節

亀岡 ASJ 2009
安良岡 MUS 2012
Hennequin DAFX 2010
Hoffman ICML 2010

これら3つのアイデアを統一的なベイズモデルとして定式化・学習則を導出

音高・音色・音高の個数・音色の個数を一挙に教師なし学習する枠組みを提案

手法：ソース・フィルタNMFに対するノンパラメトリックベイズモデリング
音楽音響信号の生成過程を考慮した事前分布・関数形を各因子に対して仮定する

(1) $|X_{mn}| \approx \sum_{k=1}^K V_{km} H_{kn}$

(2) $|X_{mn}| \sim \text{Poisson}(Y_{mn})$
 $|X_{mn}|^2 \sim \text{Exponential}(Y_{mn})$ equiv. to $\begin{cases} \min \text{KL}(|X_{mn}| | Y_{mn}) \\ \min \text{IS}(|X_{mn}|^2 | Y_{mn}) \end{cases}$ (ソースが白色雑音の場合に限って理論的に正しい)

(3) ソースの個数 I ・フィルタの個数 J を無限化する!

$|X_{mn}|$ or $|X_{mn}|^2 \approx \sum_{i,j=1}^{I,J \rightarrow \infty} \theta_i \phi_j W_{im} A_{jm} H_{ijn} = Y_{mn}$

ソース i のグローバルな重み θ_i / フィルタ j のグローバルな重み ϕ_j

生成過程: I ソース (くし型関数) × J フィルタ (自己回帰系) = IJ 基底 $W_i \odot A_j$ × IJ 音量 H_{ijn} = 多重音スペクトル

事前分布:
 $\theta_i \sim \text{Gamma}(\alpha/I, \alpha)$
 $\phi_j \sim \text{Gamma}(\gamma/J, \gamma)$
 $G_{ijn} \sim \text{Gamma}(\beta, \beta H_{ijn-1})$
 $H_{ijn} \sim \text{Gamma}(\beta, \beta G_{ijn})$

関数形:
 $W_{im} = \sum_{h=1}^H \exp\left(-\frac{(m-h\mu_i)^2}{2\sigma^2}\right)$
 $A_{jm}^{\text{KL}} = \sqrt{A_{jm}^{\text{IS}}}$
 $A_{jm}^{\text{IS}} = \frac{1}{\sum_{p=0}^P a_p^j e^{-2\pi \frac{m}{2M} p i}}^2$

エビデンス最大化のため変分ベイズ法 (VB) と乗法更新則 (MU) との組み合わせた効率的な反復学習アルゴリズムを導出

実験：ピアノ楽曲・ポピュラー楽曲に対するソース・フィルタ分解

複数基本周波数推定：MAPS ピアノデータベース (ソースに着目)

30曲に対する平均推定精度 (F値)
48.4% (KL-iCARM) 35.1% (IS-iCARM)

KL-iCARMは、ソースの個数・フィルタの個数を教師なしで適切に推定することができた
ソースに着目することで調波構成成分と白色雑音成分との音源分離が可能
IS-iCARMはソースの個数・フィルタの個数ともに過剰に見積もり傾向があった
モデルの再構成値が観測値を下回ると大きなペナルティがかかるので、多数の基底が必要になった

音色に基づく音源分離：RWC 音楽データベース (フィルタに着目)

同じモデルで音高推定も音色推定も同時にできる!

多くの楽曲で、最も優勢なフィルタはベースギターに、次に優勢なフィルタはボーカルに対応していた
打楽器 (ベースドラム・スネアドラム・ハイハットなど) はいくつかのフィルタに分かれていた

今後の展開

管楽器などにおけるフィルタの音高依存性を考慮する
人間の聴覚を参考に、対数周波数領域で再定式化を行う
自動採譜のために、言語モデルと音響モデルを統合する