

1. はじめに

研究背景

非負値行列因子分解(nonnegative matrix factorization: NMF)とは、非負値行列を対象とした行列分解アルゴリズムである。これを短時間フーリエ変換(short-time Fourier transform: STFT)した時間周波数領域の信号に適用することで、周波数スペクトルとその発音エネルギーに分解することができる。NMFによる信号解析を行う上で適切な基底数を設定することが重要である。ここで、適切な基底数とは音声为例に挙げると音素・音節などの意味のある音響的単位を示す。これを自動推定することで高品質な音響分析や音源分離などが期待できる。

研究目的

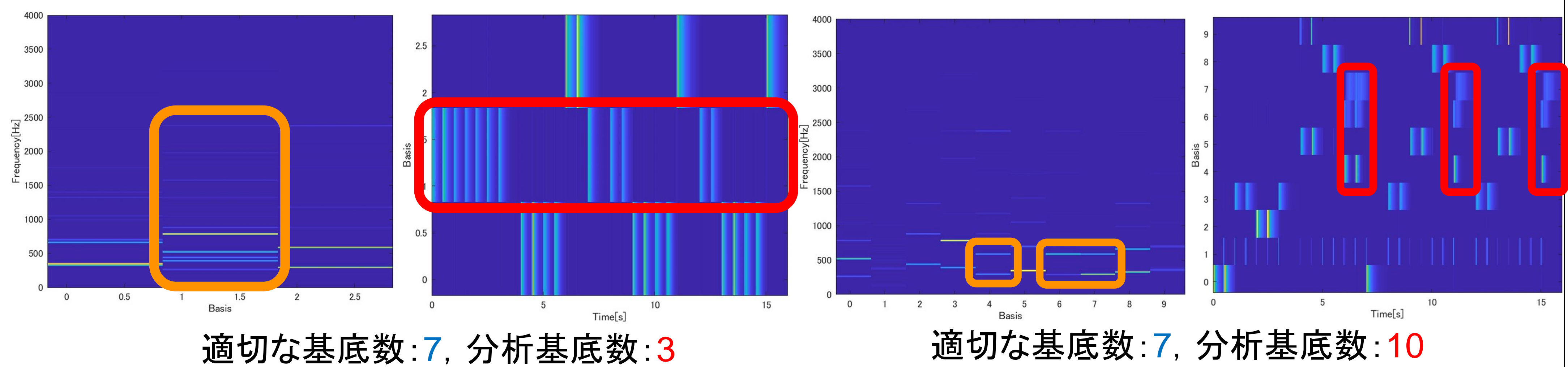
既存のNMFでは、信号中の潜在的パターンを示す基底数はユーザが設定する必要がある。一方で、任意音源に対して適切な基底数を設定することは困難である上に、出力結果に大きな影響を与える。そこで、我々は音声や楽音を対象に基底数を自動推定した上で非負値行列因子分解(NMF)を行う手法を提案する。

音素を音響的単位とした基底数推定を実現するための基礎検討として、単調なスペクトルパターンを有する楽音を対象とした基底数推定を行う。そして、音素や音階などが複数回同時に発音していたとしても、潜在する最適な基底数を自動推定できるような新たなNMFのフレームワークの提案を目指す。

2. NMFにおける基底数の関係

適切な基底数が設定されていないときのNMFの分解結果

- ・基底数が小さすぎる場合(左図)
⇒1つの基底が**複数の音階を表現**
- ・基底数が大きすぎる場合(右図)
⇒複数の基底が**単一の音階を表現**



3. 基底数推定を用いたNMFによる音源分離

提案手法

以下の手順により基底数を自動推定し、音階順にソートされた基底行列とアクティベーション行列を求める

Phase 1: 最大基底数の推定

- ・基底数1のときNMFを行い、このアクティベーションから音源の発音数を推定
⇒**最大基底数を求める**

Phase 2: 基底行列のクラスタリング

- ・各基底の類似度を評価し、類似度の高いアクティベーション行列のベクトルを結合
⇒**同じ音色の基底をマージして基底数を削減**

Phase 3: 基底ベクトルのソート

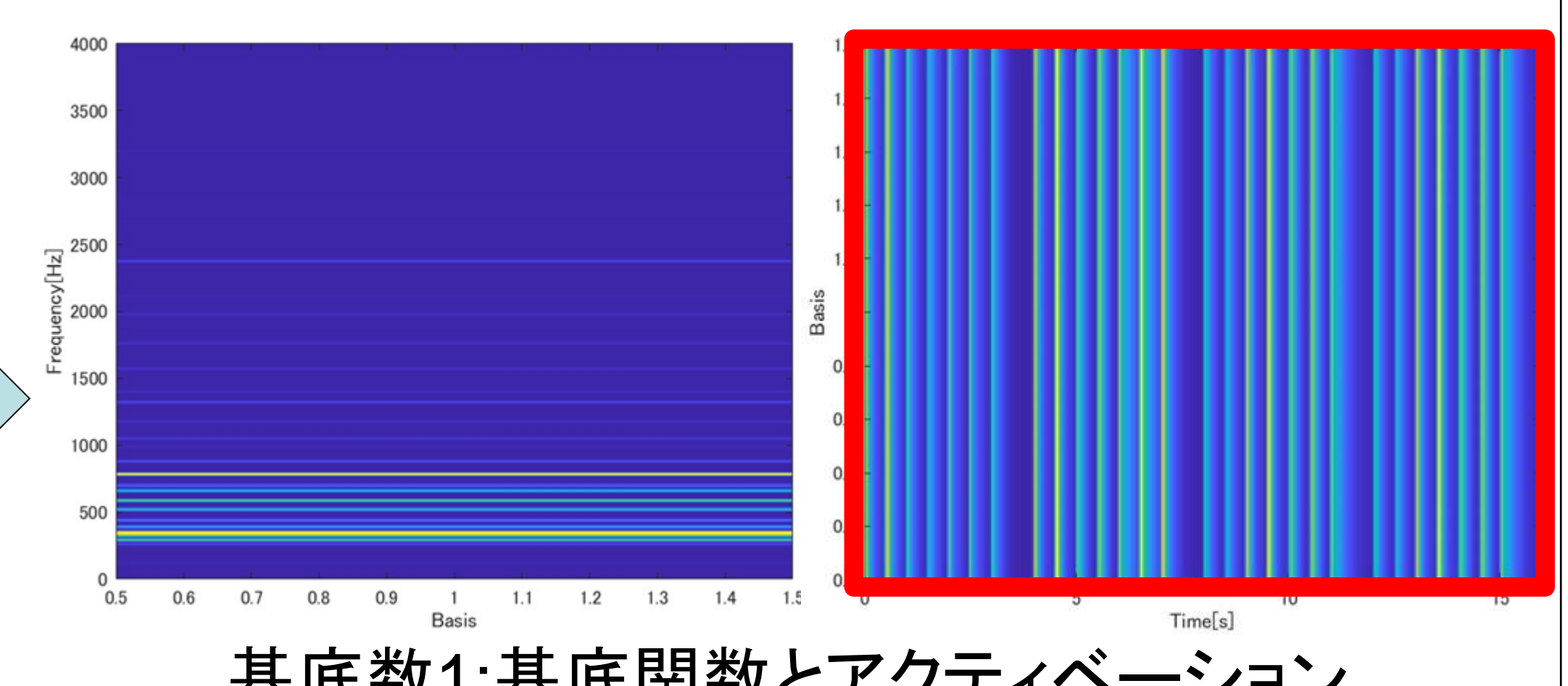
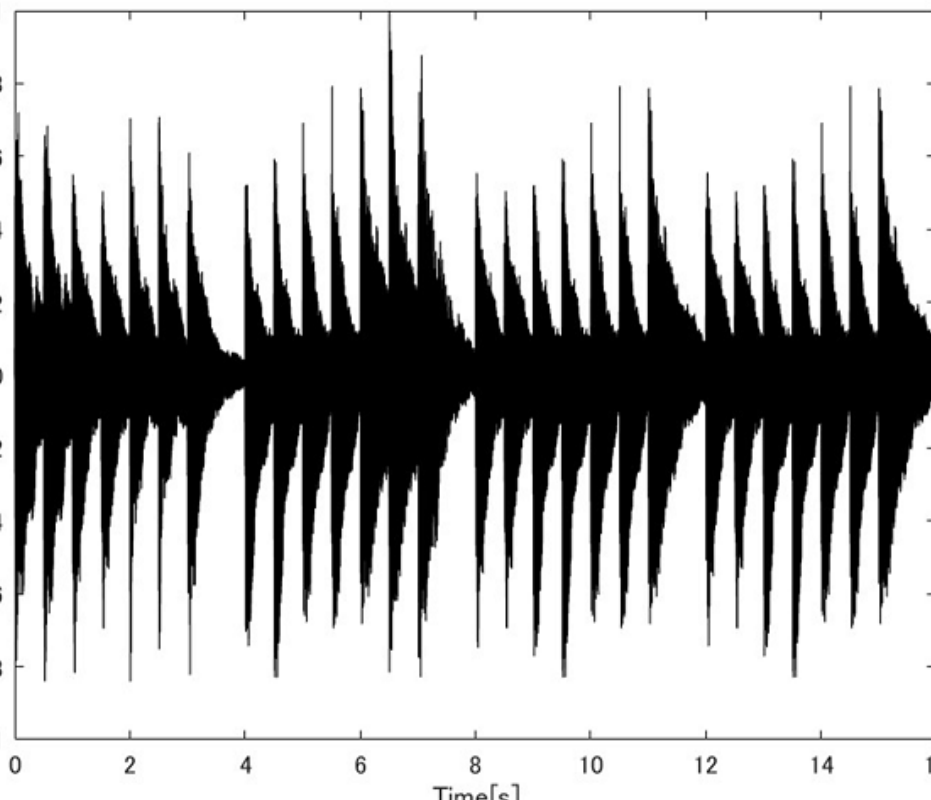
- ・推定された各基底の基本周波数でソート
※ただし、ノイズを表す基底は基本周波数にかかわらず最下段へ配置

実験条件

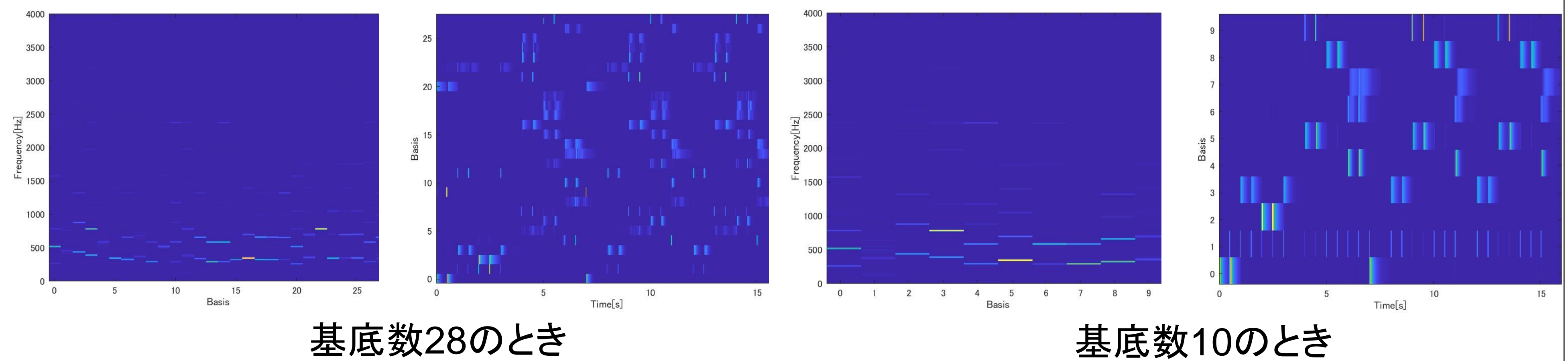
- 音源:
 - "Twinkle, twinkle, little star" 打ち込み音源
 - GarageBand / Grand Piano
- サンプリング周波数: 44,100Hz, BPM: 120, STFTの条件:
 - 窓長: 4,096, シフト長: 2,048, 窓関数: ハニング窓

実験結果

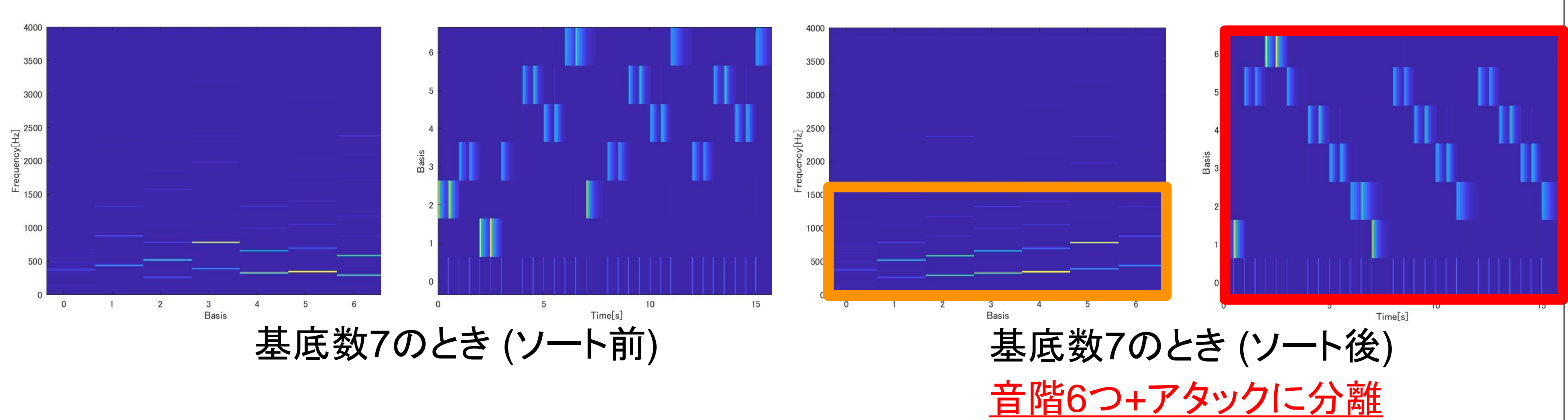
Phase 1: 最大基底数の推定



Phase 2: 基底行列のクラスタリング



Phase 3: 基底ベクトルのソート



4. おわりに

まとめ

定常的なスペクトルパターンを有する楽音を対象とした基底数推定を行った。提案手法では、自動で適切な基底数を推定できることを実験から確認した。

今後の予定

複雑なスペクトルパターンを有する楽音や自由発話の音声などを対象とした提案手法の性能を確認する。また、提案手法の自動音響分析や音源分離といった応用手法への発展も同時に検討する。