



非負値行列因子分解に基づく音響的変化の可視化手法の基礎検討

大村 美結¹, 大島風雅², 中山 仁史²

¹広島市立大学 情報科学部, ²広島市立大学 情報科学研究科

1. はじめに

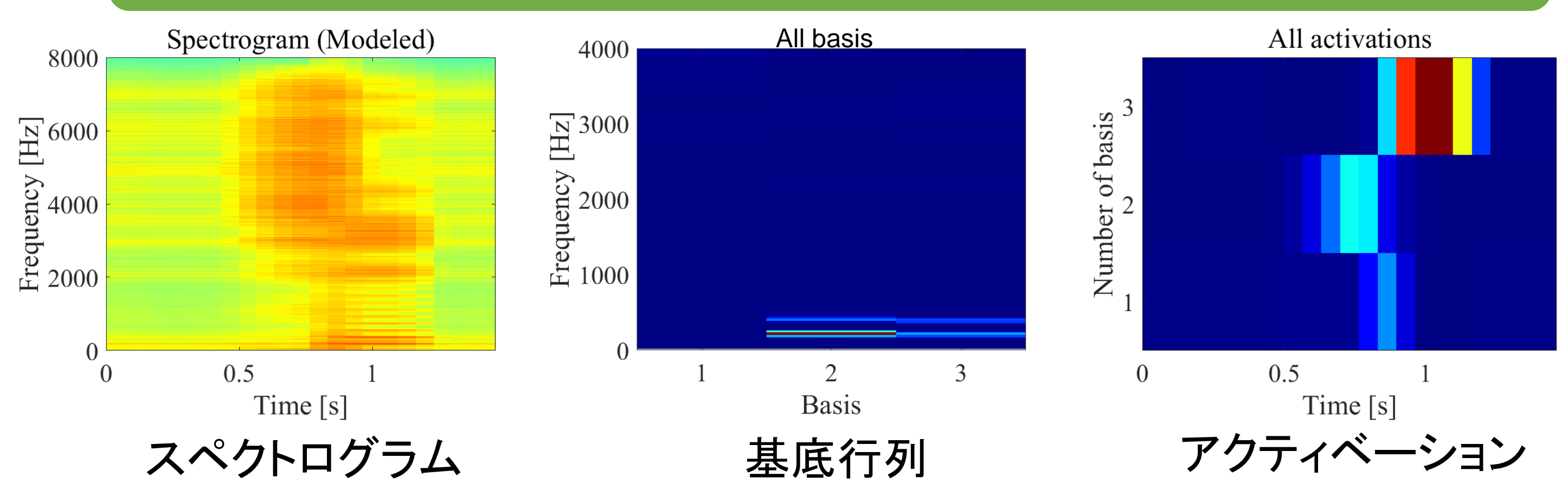
<研究背景>

- 非負値行列因子分解 (nonnegative matrix factorization: NMF)
 - 非負値行列を対象とした行列分解アルゴリズム
 - 時間周波数領域の信号に適用
 - 周波数スペクトルとその発音パワーに行列分解することができる
- ブラインド音声分離
 - 文脈依存の発声を音素・音節ごとに分解
 - 独立音源として音響的特徴を明らかにしたい

<研究目的>

- 各基底行列におけるパワー(振幅)の任意性
 - 従来のNMF
 - 基底行列とアクティベーション行列のどちらにパワーが表現されるか不定
 - 基底行列を正規化し, アクティベーションを補正
- アクティベーションによるパワーの視覚化による比較
 - 正規化された基底とパワー
 - それぞれの割合を明らかにして可視化
 - 音響的特徴を明らかにする

2. Nonnegative Matrix Factorization



スペクトログラム 基底行列 アクティベーション

スペクトログラムとしては保証されるが, 基底行列・アクティベーションのどちら振り分けられるか不明⇒各アクティベーションで比較できない

⇒視覚化による比較のためには, 基底行列とアクティベーションの正規化が必要

3. 基底関数・アクティベーションの正規化

<基底の正規化>

- 基底ベクトル T についてユークリッドノルムを計算
- T をユークリッドノルムを正規化
 - $\|x\| = \sqrt{\sum_{n=1}^N x_n^2}$

<アクティベーションの正規化>

- T のユークリッドノルムを係数ベクトルにかける
 - アクティベーションを補正
- $$\|Z\| = TV = \frac{T}{\|T\|} \|T\|V$$

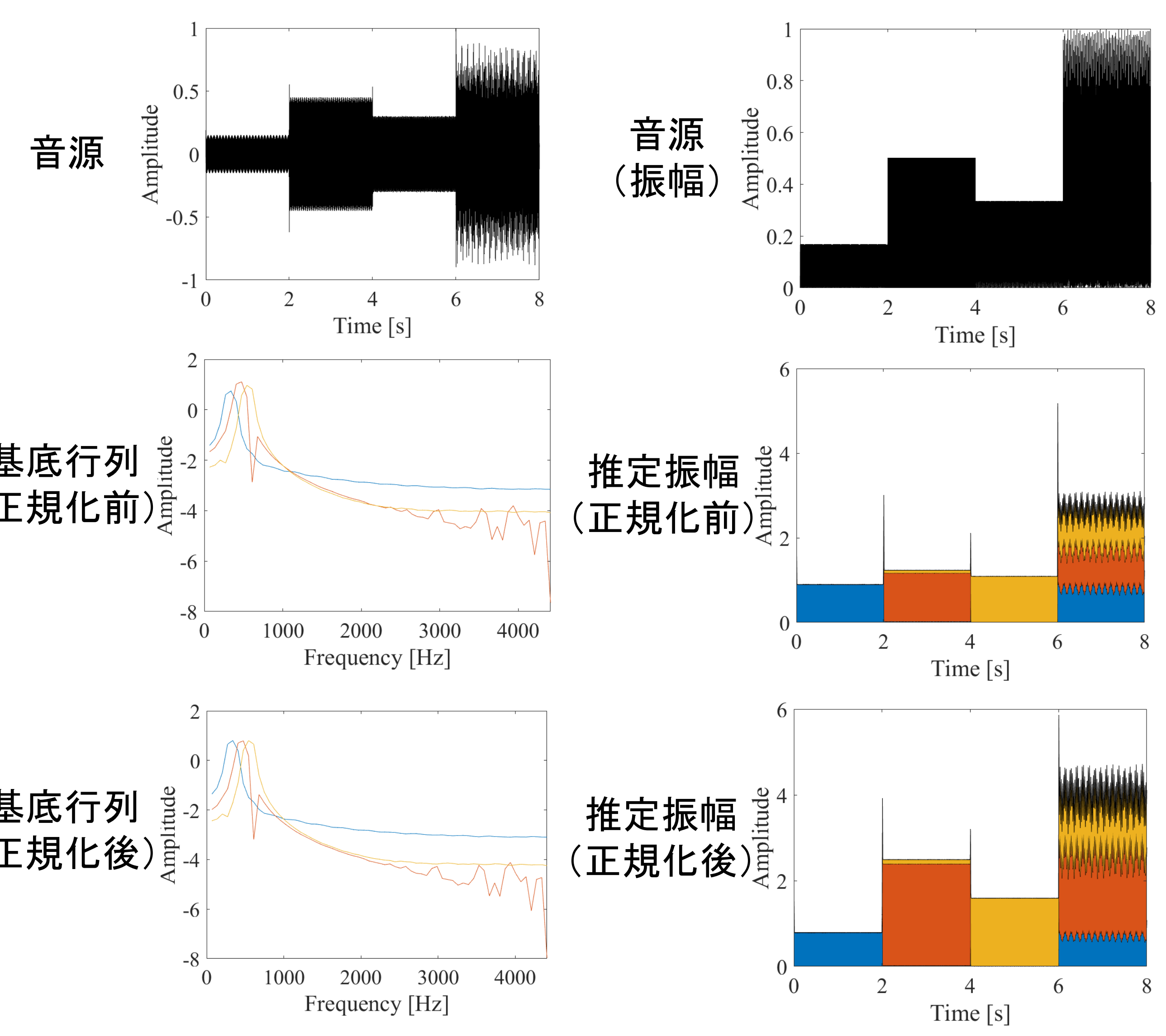
4. 音響的可視化のための音声分離実験

<実験1: 正規化手法による性能確認>

あらかじめ設定した振幅(パワー)が分かっているデータに対して, どの程度の推定性能が得られているかを確認

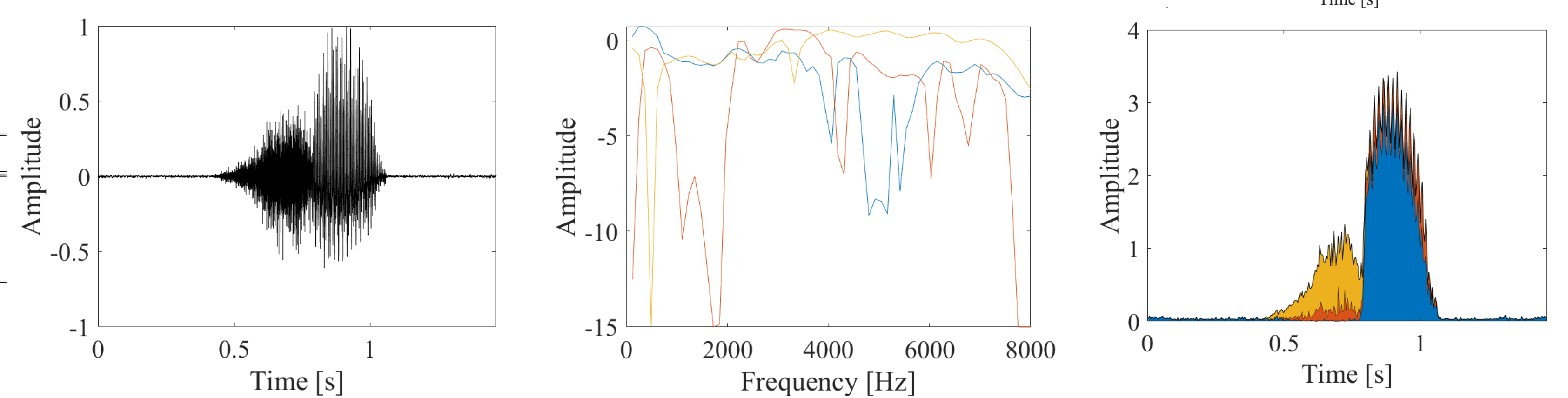
対象音源:
C4, G4及びE4の正弦波合成音
(261.63Hz, 392Hz, 329.63Hz)

Pitch	Amp.	Obs. nonreg. amp.	Obs. reg. amp.
C4	1.00	1.00	1.00
G4	3.00	0.97	3.03
E4	2.00	1.43	2.03



<実験2: 単音節を対象とした音声分離>

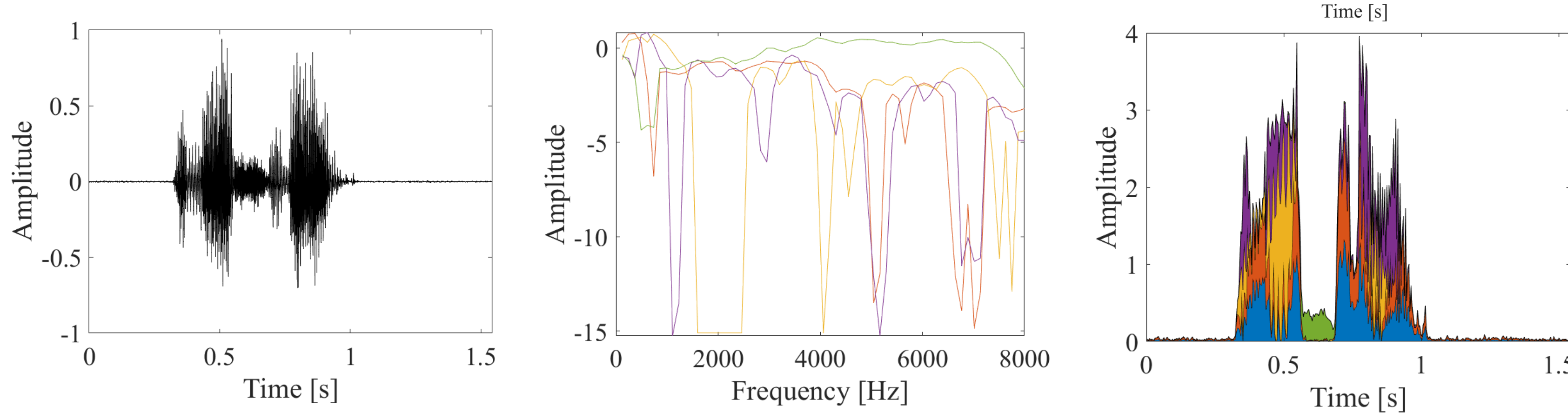
- 対象音源: 音声データ① M6SBY, /shi/



⇒子音から母音にかけての音響的変化を可視化できた

<実験3: 単語を対象とした音声分離>

- 対象音源: 音声データ② M6ATR, 「omoshiroi」



⇒単語でも分解できたが, 調音結合があるため境界が判断しづらい

5. まとめと今後の予定

<まとめ>

- 提案法により, 音素ごとの大きさの変化について比較が可能となった
- 母音の重複のような, 特徴の似た基底が複数あるデータに関しては, 基底の分離が混同するため適用が難しい

<今後の予定>

- 対象音源を増加させて正規化による適用範囲を確認すること
- 文脈依存音素における発生メカニズムや音韻論などの解明に資する音声分析ツールの構築

参考文献

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, Vol. 401, No. 6755, pp. 788-791, 1999.

[2] 板橋秀一, 重点領域研究「音声言語」・試験研究「音声 DB」連続音声データベース(PASL-DSR), 国立情報学研究所音声資源コンソーシアム. (データセット), 2006. <https://doi.org/10.32130/src.PASL-DSR>