

1. 概要

- 非負値行列因子分解(nonnegative matrix factorization: NMF)による音源分離技術:スペクトログラムを非負の基底行列とアクティベーション行列の積に分解する
- 教師ありNMF (SNMF): 分離する楽器の教師音を用いる音源分離手法
- SNMFの欠点
→教師音と目的音の音色が異なる場合に分離性能が低下

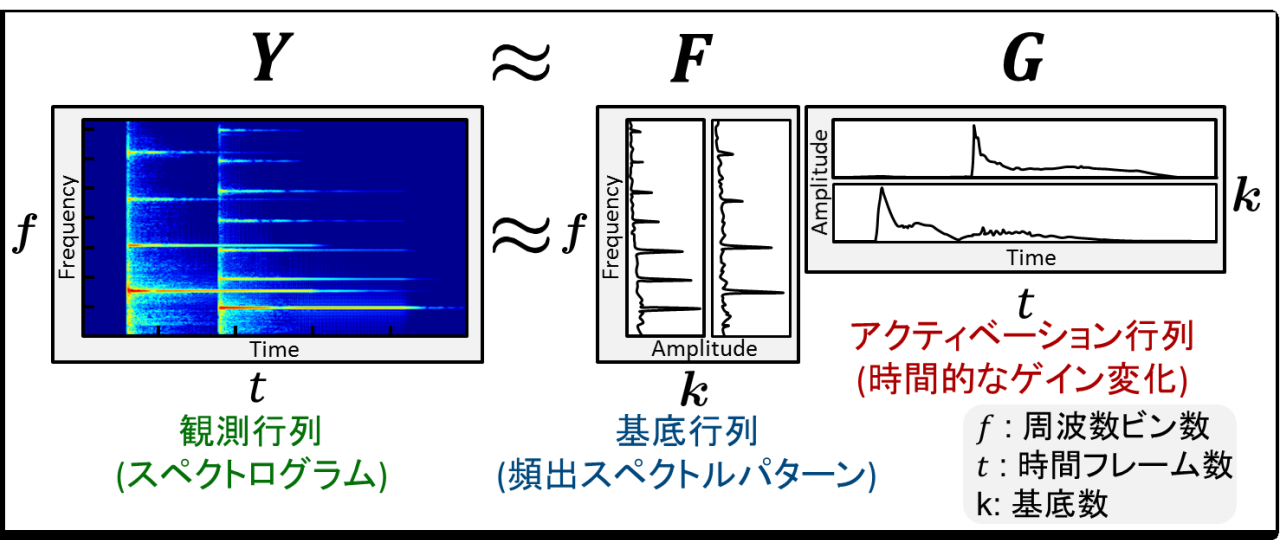
提案手法

- 線形時不変な全極モデルに基づく教師基底の変形における高い自由度での挙動の観察
- 時変な全極モデルに基づく教師基底の変形の提案

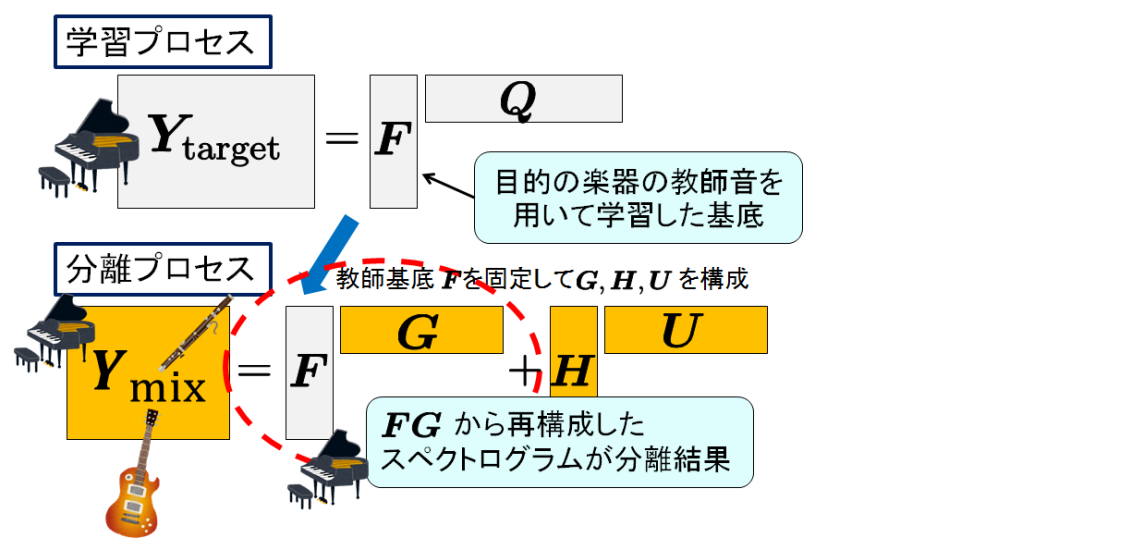
2. 従来手法

NMFとSNMF [1]

☆NMF
スパース分解表現による特徴量抽出手法



☆SNMF
教師音から基底行列Fを学習し、混合音をY ≅ FG + HUと近似して目的音をFGと抽出する



加法型基底変形SNMF [2]

教師音より得られるスペクトル基底Fを加法項Dで変形し、分離目的音に適応させる分解手法

分解表現モデル:

$$Y_{\text{mix}} \approx (F + D)G + HU$$

- 直交化罰則が多くなり、重みパラメータの調整が困難
- 初期値依存性が強い

統計的音声スペクトル推定器を用いたポストフィルタ (一般化MMSE-STSA推定器) [3]

初期値としての推定妨害音と目的音の事前分布を与えることで生成できるポストフィルタ(目的音抽出フィルタ)。

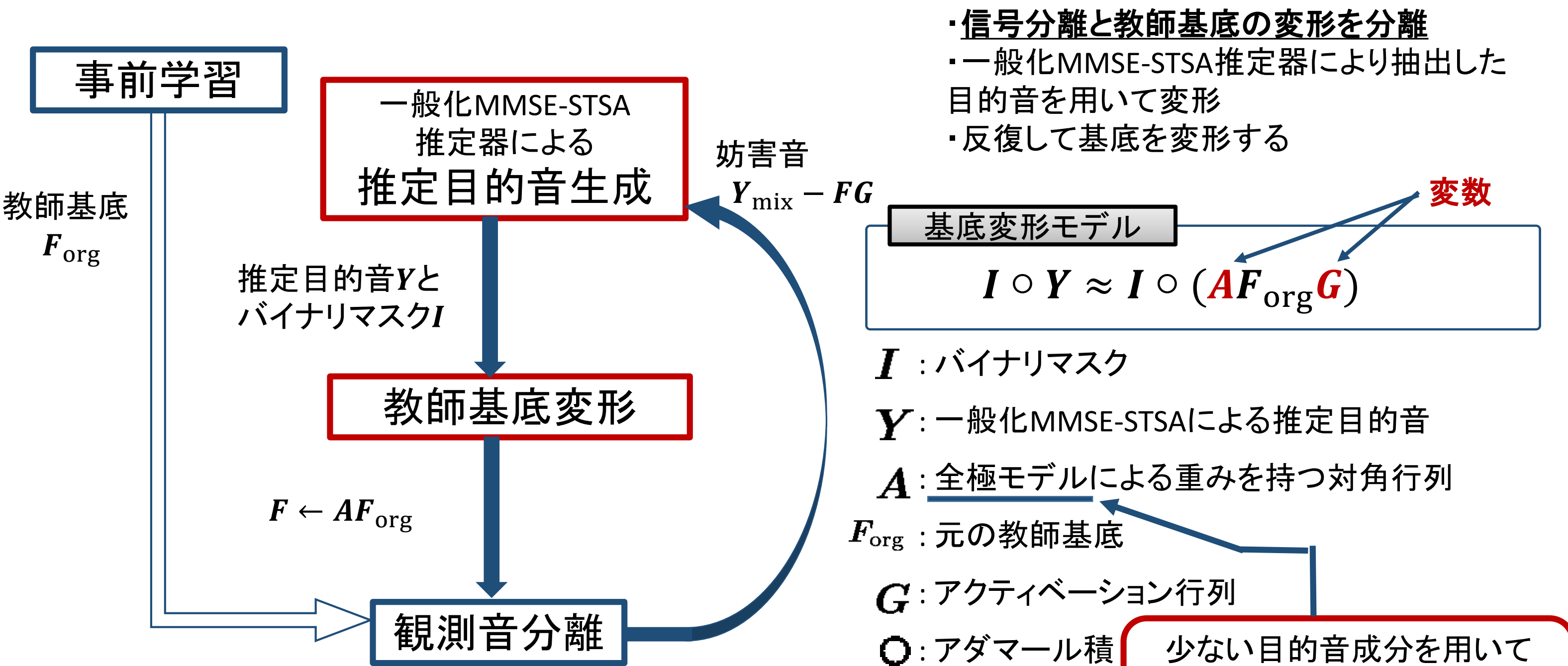
$$Y = J \circ Y_{\text{mix}}$$

○ はアダマール積

目的音振幅 Y_{target} をカイ分布と仮定し、妨害音を与えて $\min E[(Y_{\text{target}} - Y)^2]$ を解くことで得られるフィルタ J

3. 提案手法

システムの流れ



バイナリマスクによる推定目的音の選別

一般化MMSE-STSA推定器の推定目的音は精度が悪い

→精度の高い推定目的音成分のみ変形に使用

妨害音との重複が少ない成分

ポストフィルタ J の値がある程度大きい成分

ポストフィルタ J を閾値処理して推定目的音から精度の高い成分を取り出すバイナリマスク I を生成する

3.2 目的関数と更新式

一般化MMSE-STSA推定器の出力 Y を用いてスペクトル重み A を学習

$$\mathcal{J} = \sum_{\omega,t} i_{\omega,t} \left\{ -y_{\omega,t} + \frac{\sum_k f_{\omega,k} g_{k,t}}{|A_{\omega}|} + y_{\omega,t} \log \frac{y_{\omega,t}}{\sum_k f_{\omega,k} g_{k,t} / |A_{\omega}|} \right\}$$

$i_{\omega,t}$: I の要素値 $y_{\omega,t}$: Y の要素値

$f_{\omega,k}$: F_{org} の要素値 $g_{k,t}$: G の要素値

$$A_{\omega} = 1 - \sum_{k=1}^p \alpha_k \exp(-\pi j k \frac{\omega}{\Omega})$$

Ω : ナイキスト周波数

変数の更新式

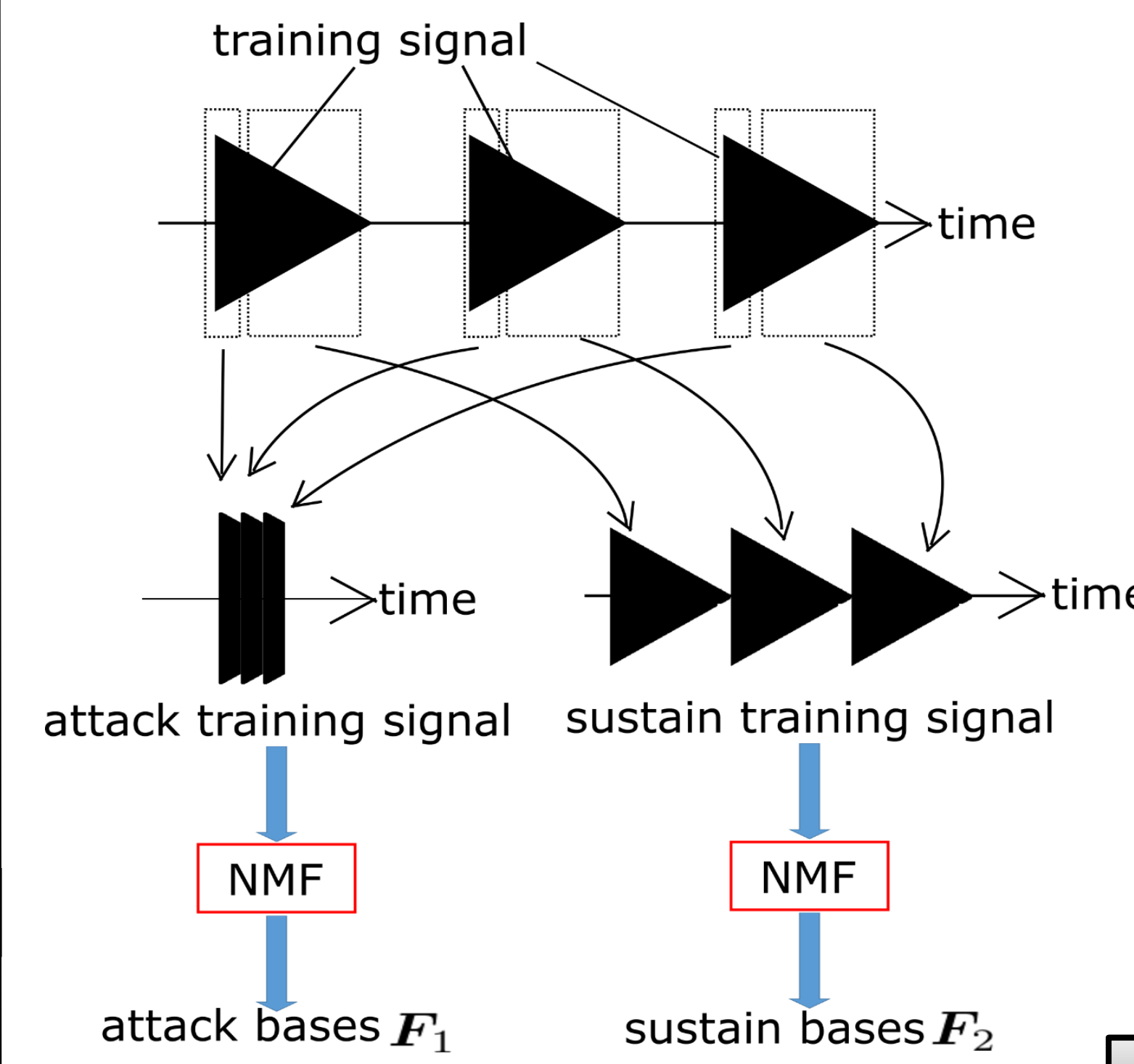
$$\alpha \leftarrow R^{-1}r \quad (\alpha \text{は全極モデルの係数ベクトル})$$

$$R_{k,q} = \sum_{\omega,t} \left[i_{\omega,t} \left(\sum_k f_{\omega,k} g_{k,t} \frac{1}{|A_{\omega}|^3} + y_{\omega,t} \frac{1}{2|A_{\omega}|^2} \right) \left(\exp(-2\pi j \frac{\omega}{\Omega} (k-q)) + \exp(2\pi j \frac{\omega}{\Omega} (k-q)) \right) \right]$$

$$r_q = \sum_{\omega,t} i_{\omega,t} \left[\left(\sum_k f_{\omega,k} g_{k,t} \frac{1}{|A_{\omega}|^3} + y_{\omega,t} \frac{1}{2|A_{\omega}|^2} \right) \left(\exp(-2\pi j \frac{\omega}{\Omega} q) + \exp(2\pi j \frac{\omega}{\Omega} q) \right) - \frac{3}{|A_{\omega}|^2} \sum_k f_{\omega,k} g_{k,t} \text{Re} \left[\frac{A_{\omega}^*}{|A_{\omega}|} \exp(-2\pi j \frac{\omega}{\Omega} q) \right] \right]$$

$$g_{k,t} \leftarrow g_{k,t} \frac{\sum_{\omega} i_{\omega,t} y_{\omega,t} f_{\omega,k} / (\sum_k f_{\omega,k} g_{k,t})}{\sum_{\omega} i_{\omega,t} f_{\omega,k} / |A_{\omega}|}$$

4. 時変な教師基底変形



コスト関数

$$\mathcal{J} = \sum_{\omega,t} i_{\omega,t} \left\{ -y_{\omega,t} + \frac{\sum_k f_{\omega,k,1} g_{k,t,1}}{|A_{\omega}|} + \frac{\sum_l f_{\omega,l,2} g_{l,t,2}}{|B_{\omega}|} + y_{\omega,t} \log \frac{y_{\omega,t}}{\sum_k f_{\omega,k,1} g_{k,t,1} / |A_{\omega}| + \sum_l f_{\omega,l,2} g_{l,t,2} / |B_{\omega}|} \right\}$$

$i_{\omega,t}$: I の要素値

$y_{\omega,t}$: Y の要素値

$f_{\omega,t,1}$: F_1 の要素値

$g_{\omega,t,1}$: G_1 の要素値

$f_{\omega,t,2}$: F_2 の要素値

$g_{\omega,t,2}$: G_2 の要素値

$$A_{\omega} = 1 - \sum_{k=1}^p \alpha_k \exp(-\pi j k \frac{\omega}{\Omega})$$

$$B_{\omega} = 1 - \sum_{k=1}^p \beta_k \exp(-\pi j k \frac{\omega}{\Omega})$$

Ω : ナイキスト周波数

- 楽器音の時変構造に応じた基底の変形
- 教師音を、信号の鳴り始めと鳴り終りに分割し、別々に基底を学習
- それぞれの基底に別々の変形を施す

基底変形モデル

$$I \circ Y \approx I \circ (A F_1 G_1 + B F_2 G_2)$$

I : バイナリマスク

Y : 一般化MMSE-STSAによる推定目的音

A : F_1 を変形する全極モデルによる重み行列

B : F_2 を変形する全極モデルによる重み行列

F_1 : 鳴り始め部分のみによって学習した教師基底

F_2 : 鳴り終わり部分のみによって学習した教師基底

G : アクティベーション行列

$\circ$: アダマール積

変数の更新式

$$\alpha = W^{-1}w \quad \beta = V^{-1}v \quad (\alpha, \beta \text{は全極モデル} A, B \text{の係数ベクトル})$$

$$W_{m,q} = \sum_{\omega,t} \left[i_{\omega,t} \left(\sum_k f_{\omega,k,1} g_{k,t,1} \frac{1}{|A_{\omega}|^3} + y_{\omega,t} \frac{1}{2|A_{\omega}|^2} \sum_k f_{\omega,k,1} g_{k,t,1} + |A_{\omega}| \sum_k f_{\omega,k,1} g_{k,t,1} / |A_{\omega,t}| \right) \left(\exp(-\pi j \frac{\omega}{\Omega} (m-q)) + \exp(\pi j \frac{\omega}{\Omega} (m-q)) \right) \right]$$

$$w_q = \sum_{\omega,t} i_{\omega,t} \left[\left(\sum_k f_{\omega,k,1} g_{k,t,1} \frac{1}{|A_{\omega}|^3} + y_{\omega,t} \frac{1}{2|A_{\omega}|^2} \sum_k f_{\omega,k,1} g_{k,t,1} + |A_{\omega}| \sum_k f_{\omega,k,1} g_{k,t,1} / |A_{\omega,t}| \right) \left(\exp(-\pi j \frac{\omega}{\Omega} q) + \exp(\pi j \frac{\omega}{\Omega} q) \right) - \frac{3}{|A_{\omega}|^2} \sum_k f_{\omega,k,1} g_{k,t,1} \text{Re} \left[\frac{A_{\omega}^*}{|A_{\omega}|} \exp(-\pi j \frac{\omega}{\Omega} q) \right] \right]$$

$$V_{m,q} = \sum_{\omega,t} i_{\omega,t} \left[\left(\sum_l f_{\omega,l,2} g_{l,t,2} \frac{1}{|B_{\omega}|^3} + y_{\omega,t} \frac{1}{2|B_{\omega}|^2} \sum_l f_{\omega,l,2} g_{l,t,2} + |B_{\omega}| \sum_l f_{\omega,l,2} g_{l,t,2} / |B_{\omega,t}| \right) \left(\exp(-\pi j \frac{\omega}{\Omega} (m-q)) + \exp(\pi j \frac{\omega}{\Omega} (m-q)) \right) \right]$$

$$v_q = \sum_{\omega,t} i_{\omega,t} \left[\left(\sum_l f_{\omega,l,2} g_{l,t,2} \frac{1}{|B_{\omega}|^3} + y_{\omega,t} \frac{1}{2|B_{\omega}|^2} \sum_l f_{\omega,l,2} g_{l,t,2} + |B_{\omega}| \sum_l f_{\omega,l,2} g_{l,t,2} / |B_{\omega,t}| \right) \left(\exp(-\pi j \frac{\omega}{\Omega} q) + \exp(\pi j \frac{\omega}{\Omega} q) \right) - \frac{3}{|B_{\omega}|^2} \sum_l f_{\omega,l,2} g_{l,t,2} \text{Re} \left[\frac{B_{\omega}^*}{|B_{\omega}|} \exp(-\pi j \frac{\omega}{\Omega} q) \right] \right]$$

$$g_{k,t,1} = g_{k,t,1} \frac{\sum_{\omega} i_{\omega,t} f_{\omega,k,1} / (\sum_k f_{\omega,k,1} g_{k,t,1})}{\sum_{\omega} i_{\omega,t} f_{\omega,k,1} / |A_{\omega}|}$$
$$g_{l,t,2} = g_{l,t,2} \frac{\sum_{\omega} i_{\omega,t} f_{\omega,l,2} / (\sum_l f_{\omega,l,2} g_{l,t,2})}{\sum_{\omega} i_{\omega,t} f_{\omega,l,2} / |B_{\omega}|}$$

5. 実験

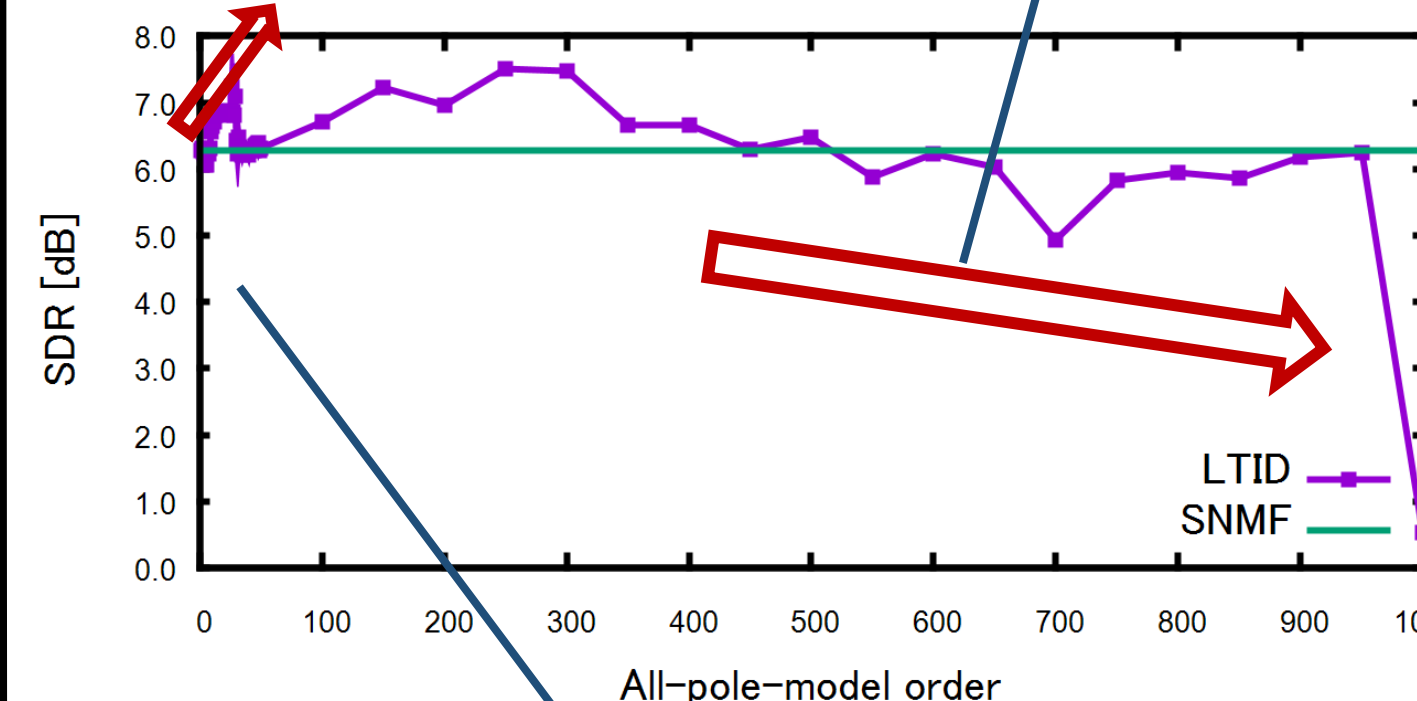
実験条件

楽器	Oboe (Ob.), Piano (Pf.), Trombone (Tb.)
教師音 (MIDI)	目的音の半音階で2オクターブ上昇する24音 Garritan Professional Orchestra
目的音 (MIDI)	Microsoft GS Wavetable SW Synth
サンプリング周波数	44100 Hz
FFT長	4096点(約100 ms)
シフト長	512点(約15 ms)
基底数	目的音: 100, 妨害音: 30
反復回数	学習時: 500, 分離時: 400
評価値	SDR: 分離度合いと人工歪みの少なさを含む総合的な分離音源の品質
比較手法	SNMF, 時不変な全極モデルによる教師基底変形(LTID), 時変な教師基底変形(LTVD)

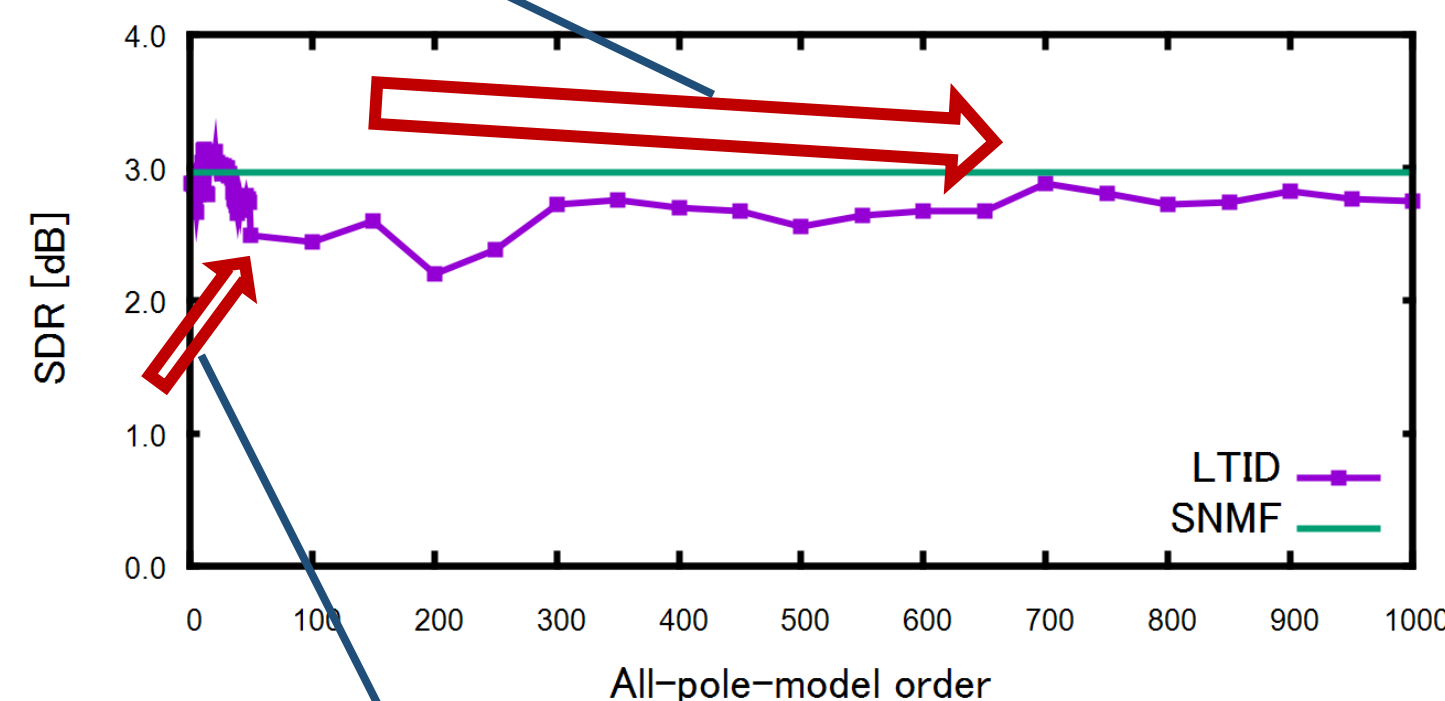
時不変な全極モデルによる教師基底変形の
高次数における挙動

②最大値をとる次数以上の次数ではむしろSDRは低下する傾向

Ob.とPf.の混合音からOb.を分離抽出する場合におけるSDR



Ob.とTb.の混合音からOb.を分離抽出する場合におけるSDR

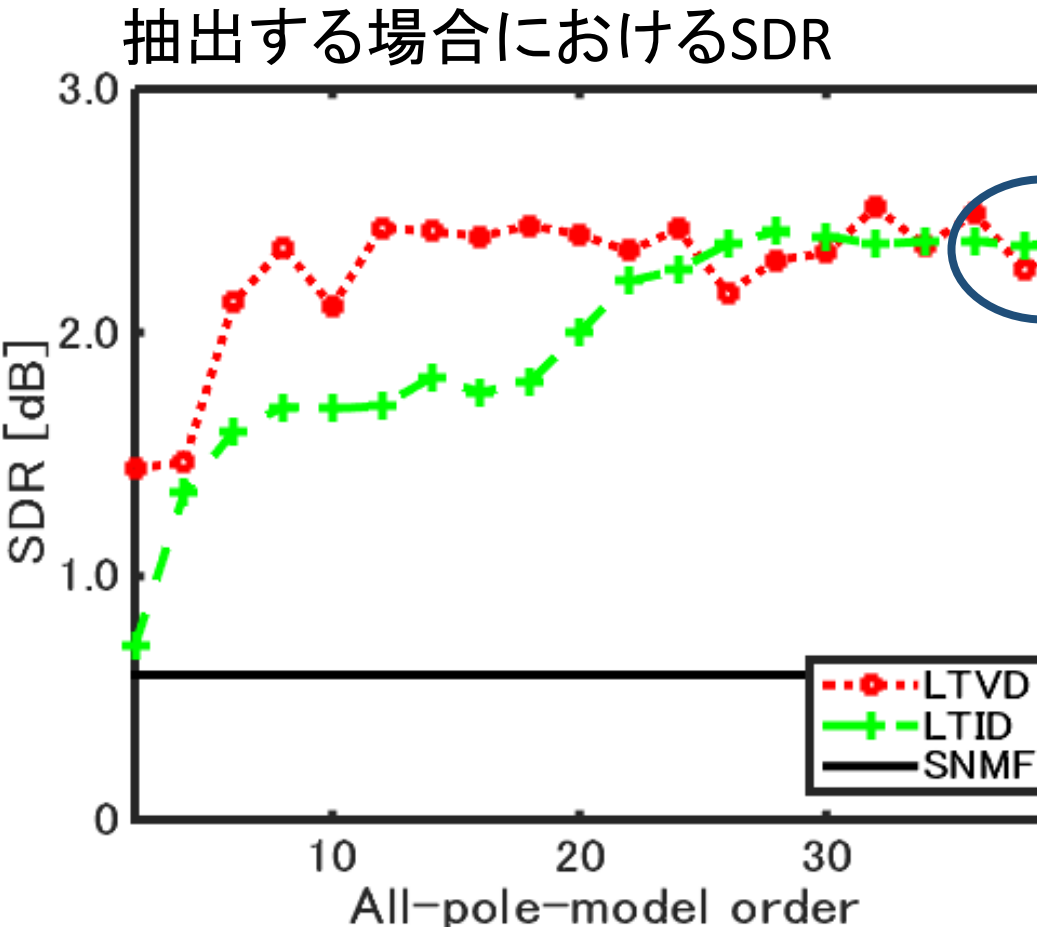


①次数を上昇させるとSDRが上昇

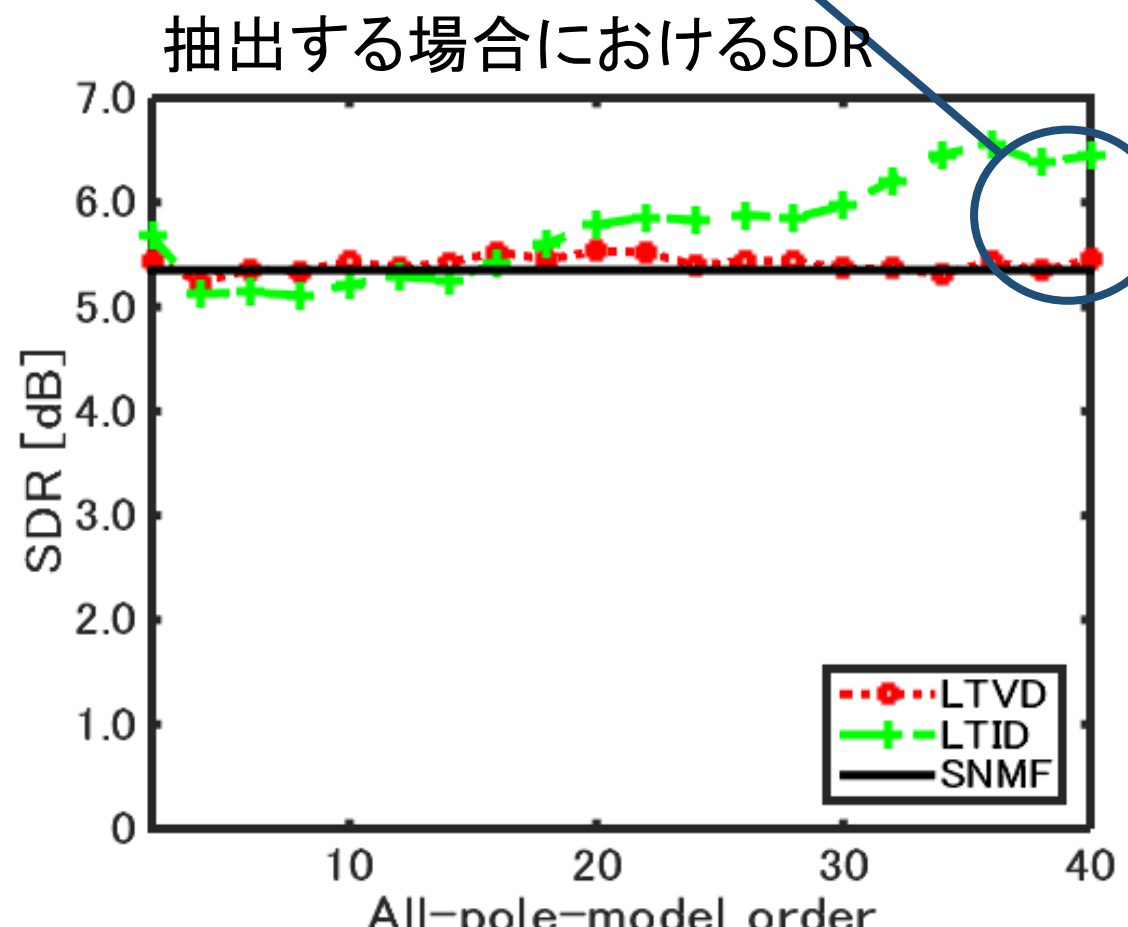
時変な教師基底変形の挙動

教師基底の変形の自由度を向上させることは必ずしも分離性能の改善につながらない

Tb.とOb.の混合音からTb.を分離抽出する場合におけるSDR



Pf.とOb.の混合音からTb.を分離抽出する場合におけるSDR



考察

過度な自由度の上昇はSDRを低下させる＝適切な自由度設定の必要性

理由

- 一般化MMSE-STSA推定器の推定目的音の不確かさ
- バイナリマスクによって変形に利用できる成分が減少した目的音に過適応しても分離はうまく出来ない