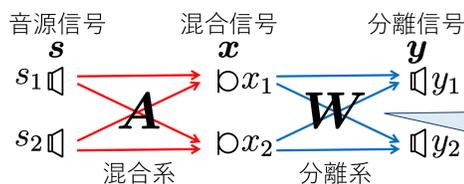


2-1Q-18 メディアン型HPSSを用いた時間周波数マスクに基づくブラインド音源分離

☆大藪宗一郎(香川高専), 北村大地(香川高専), 矢田部浩平(早稲田大)

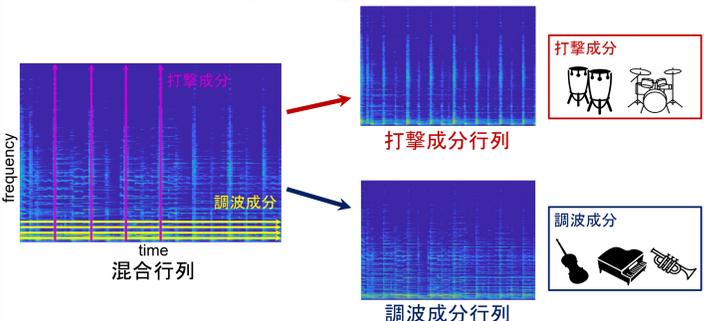
1. 研究背景

- ブラインド音源分離(BSS): 録音マイクや音源の位置が未知での音源分離



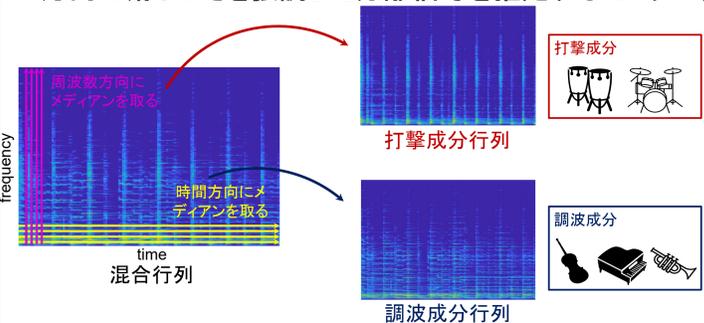
分離音 y を求める為の分離フィルタ W を推定

- 調波打撃音分離(HPSS): 振幅スペクトログラムの時間・周波数方向の滑らかさに基づくモノラル音源分離 [N. Ono et al., 2008]



時間・周波数方向の滑らかさに基づいて分離

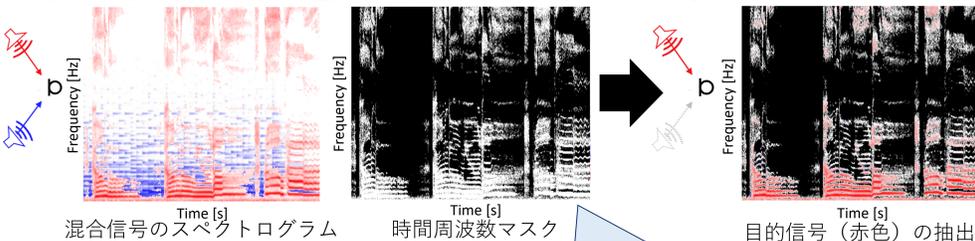
- メディアン型調波打撃音分離(メディアン型HPSS): メディアンをとることで、各軸方向の滑らかさを強調した分離信号を推定するモノラル音源分離 [D. FitzGerald, 2010]



時間・周波数方向にそれぞれメディアンフィルタを適用

これらの分離手法は非線形処理なので、人工歪みが目立つことがある

- 時間周波数マスクに基づく優決定BSS(TFMBSS): 時間周波数マスクで表現される音源モデルに基づいた線形な多チャンネルBSS [K. Yatabe and D. Kitamura, 2019]



赤の音源の部分を1 それ以外を0とするようなマスク

何らかの音源モデルから時間周波数マスクを構築

様々な音源モデルを入れ替えて活用可能

本研究の動機

非線形なモノラル音源分離であるメディアン型HPSSの多チャンネル化規提案のアルゴリズムより高精度かつ高速な音源モデルの探求

本研究の目的

我々はHPSSから生成される時間周波数マスクを用いたTFMBSSを提案済メディアン型HPSSの特徴を活かした、線形な多チャンネル音源分離を実現規提案のアルゴリズムと比較して精度・速度向上の達成 [S. Oyabu et al., 2020]

2. 従来手法

- 従来手法 [S. Oyabu et al., 2020]

モノラル音源分離であるHPSSにTFMBSSを適用
線形な分離とHPSSによる長波打撃音分離の両立を実現

分離性能は他の多チャンネルBSSと比較しても十分に**高性能**
HPSSが反復推定アルゴリズムであるため**低速**

- 排他的マスク処理: 音源分離による音源強調をフィルタとして捉えた処理

調波音成分 Z_P の中に残留する打撃音成分,
打撃音成分 Z_H の中に残留する調波音成分を除去

- スムージング: 反復毎のマスクの変化を小さくし、アルゴリズム安定的に動作
マスク生成の度に、1反復前のマスク M_{old} との要素積を取る

$$M = M \circledast M_{old}^{\beta_{old}}$$

要素ごとの積 (現在のマスク) スムージングパラメータ (反復前のマスク)

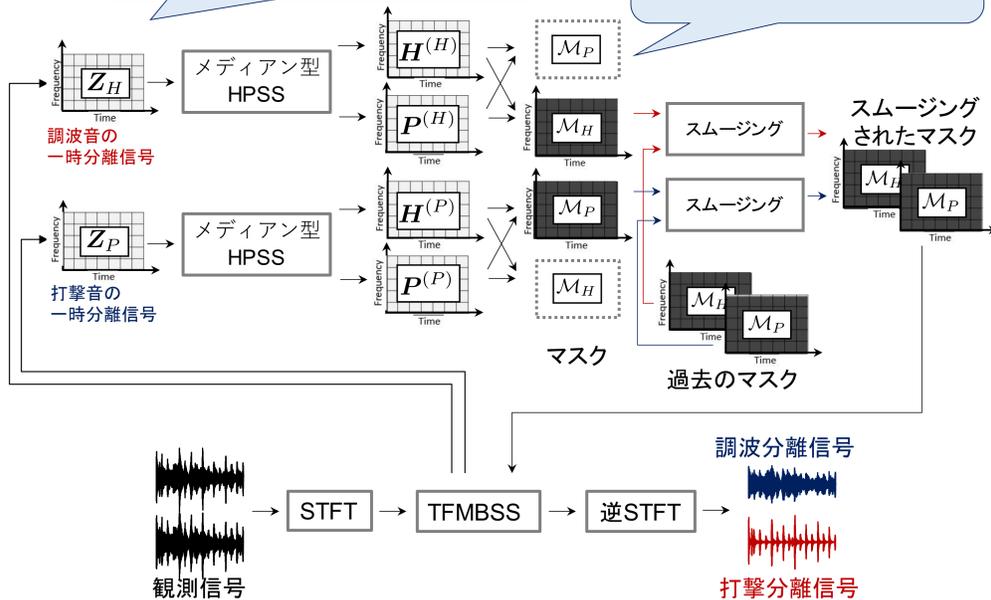
一回の反復更新時のマスクの変動を小さくできる

3. 提案手法

- 提案手法

調波音用と打撃音用HPSSに Z_P 及び Z_H を1/2した値を初期値

残留成分を取り除いた成分のみからマスクを生成



- 分離信号推定の高速化

メディアン型HPSSは比較的**低コスト**で推定可能
さらに、HPSSと**同程度**の性能を確認
メディアン型HPSSのマスクに基づくTFMBSSの実装

従来手法より、**高性能**
かつ**高速な性能**が期待

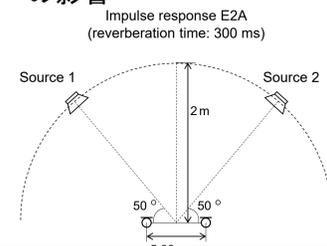
4. 比較実験

- 実験項目

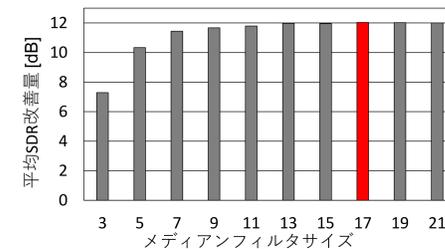
メディアンフィルタサイズによるSDR改善量への影響
スムージングパラメータの変化によるSDR改善量への影響
従来手法とのSDR改善量・分離速度比較

- 実験条件

学習用音声信号	SISECのプロ音楽信号にRWCP収録のマイクアレイパルス応答で畳み込んで作成、2チャンネルで2音源の混合信号
窓関数長	128点 (8 ms)
シフト長	64点 (4 ms, ハン窓)
反復回数	500回
主観評価値	平均SDR改善値

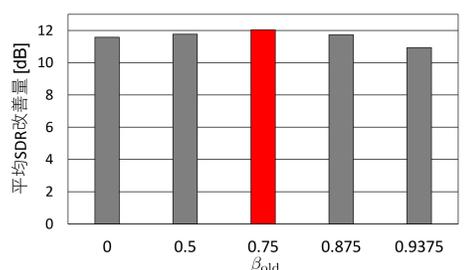
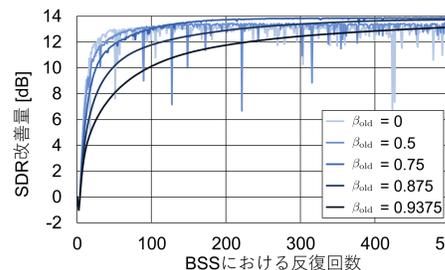


- メディアンフィルタサイズにおける影響検証



提案手法では、HPSSの反復回数が13~21程度であれば平均的に**高性能**
今回の実験条件では、
メディアンフィルタサイズは**17回**が最適

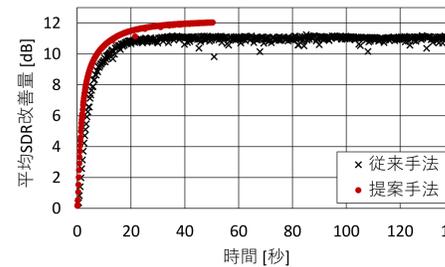
- スムージングパラメータにおける影響検証



スムージングを弱くした場合、SDR改善量の推移が**不安定**
スムージングを強くした場合、推移は**安定**するが収束が遅れる

推移の安定と収束速度・収束値の減衰の**トレードオフ**
これを考慮したパラメータ設定が必要
提案手法では、 $\beta_{old} = 0.75$ 及び $\beta = 0.25$ が最適

- 従来手法との比較結果



提案手法は、従来手法に比べ**高性能**かつ**高速**

調波音と打撃音の多チャンネル音源分離に特化した**高性能**かつ**高速な**
BSSである