

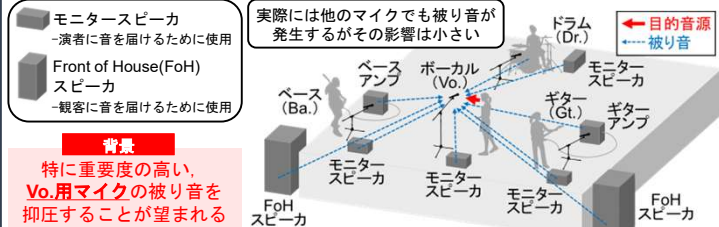
本発表の概要

- 音楽ライブ演奏の収録では、各マイクに意図していない音源信号(被り音)が混入
- 特にボーカル(Vo.)用マイクには許容できない大きさの被り音が混入し、悪影響を及ぼす
- ➡ **本研究では、Vo.用マイクにフォーカスした被り音抑圧手法を提案**
- 他の音源であるギター(Gt.), ベース(Ba.), ドラム(Dr.)用マイクの被り音は小さい
- ➡ **Vo.以外の音源信号は既知であるというセミブラインドなモデルを仮定し、独立低ランク行列分析(independent low-rank matrix analysis: ILRMA)に導入**
- 実験では、実際に音楽ライブ演奏環境を構築しステージ上のインパルス応答を測定
- 測定結果を使用し、提案手法が効果的なVo.用マイク被り音抑圧を達成することを確認

1. 研究背景

被り音(bleeding sound)

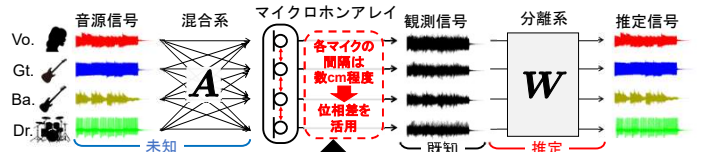
- 音楽ライブ演奏で近接マイクでの収録を行う際、**Vo.音源**の明瞭な収録が重要となる
- しかし、ステージ上に存在する他の音源が**被り音**としてVo.用マイクに混入してしまう



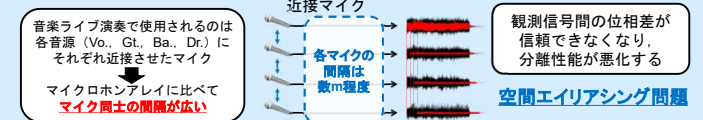
2. 従来手法

ブラインド音源分離(blind source separation: BSS)

- 音源やマイクの空間的配置(混合系A)が不明な状況下で分離系Wを推定する手法
- 観測信号同士の**位相差**を利用するためマイクロホンレイでの収録を前提としている

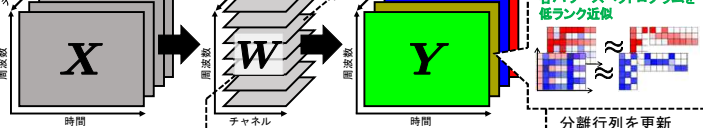


BSSによる被り音抑圧の問題点



独立低ランク行列分析(independent low-rank matrix analysis: ILRMA) [Klitamura+, 2016]

- 時間周波数領域の推定信号に対して、「各信号間における統計的独立性の最大化」及び「パワースペクトログラムの低ランクモデリング」を同時に達成するBSS



ILRMAによる被り音抑圧の問題点



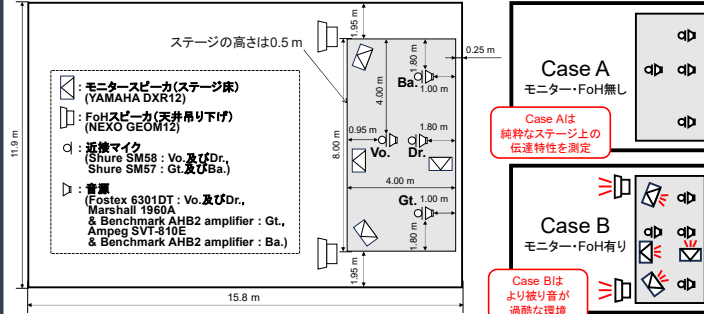
本研究の目的

ILRMAに改良を施し、空間エイリアシングに対して頑健に動作するVo.用マイクの被り音抑圧手法を提案する

3. インパルス応答測定

音楽ライブ演奏環境のインパルス応答測定

- 被り音の特性を調査するため、実際にステージを構築しインパルス応答を測定
- モニター及びFoHスピーカの影響を調査するため、2つのCaseで測定を実施



測定結果(被り音相対エネルギー解析)

- 各音源から各近接マイクへのインパルス応答のエネルギーを算出
- 近接マイク毎に(表中では行方向)目的音源のエネルギーで正規化(単位はdB)

Case A(モニター・FoH無し)

近接マイク基準の相対パワー [dB]	Vo.音源	Gt.音源	Ba.音源	Dr.音源
Vo.用マイク	0.0	-17.3	-12.9	-1.2
Gt.用マイク	-56.3	0.0	-25.3	-29.4
Ba.用マイク	-59.7	-34.8	0.0	-31.2
Dr.用マイク	-72.7	-44.6	-40.6	0.0

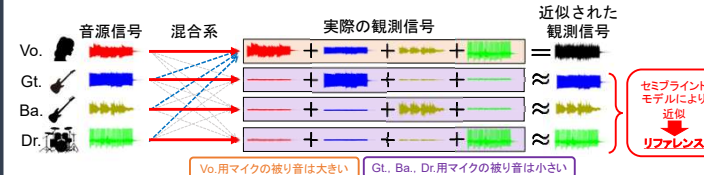
Case B(モニター・FoH有り)

近接マイク基準の相対パワー [dB]	Vo.音源	Gt.音源	Ba.音源	Dr.音源
Vo.用マイク	0.0	-9.8	-6.9	-1.3
Gt.用マイク	-31.5	0.0	-16.9	-19.8
Ba.用マイク	-32.5	-21.5	0.0	-23.1
Dr.用マイク	-43.9	-34.1	-31.4	0.0

4. 提案手法

セミブラインドモデル

- 被り音の小さいマイク観測信号は、音源信号そのものとして扱えるというモデル
- 音楽ライブ演奏の場合は、Vo.以外(Gt., Ba., Dr.)の観測信号が**リファレンス**となる



Semi-blind ILRMA

- ILRMAに先述のセミブラインドモデルを導入
- 信号のリファレンス化により、通常のBSSモデルから分離行列の形状が変化

BSSの分離行列

$$W_f = \begin{bmatrix} w_{f11} & w_{f12} & w_{f13} & w_{f14} \\ w_{f21} & w_{f22} & w_{f23} & w_{f24} \\ w_{f31} & w_{f32} & w_{f33} & w_{f34} \\ w_{f41} & w_{f42} & w_{f43} & w_{f44} \end{bmatrix} \in \mathbb{C}^{4 \times 4}$$

リファレンス信号は存在せず、全てのパラメータを推定する必要がある

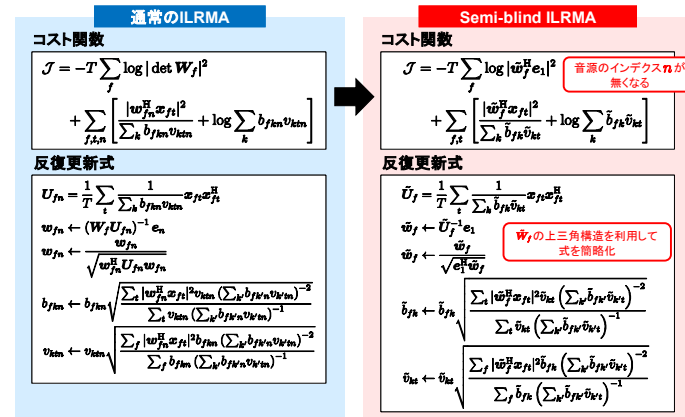
セミブラインドモデルの分離行列

$$\tilde{W}_f = \begin{bmatrix} \tilde{w}_{f11} & \tilde{w}_{f12} & \tilde{w}_{f13} & \tilde{w}_{f14} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{C}^{4 \times 4}$$

2, 3, 4行目(Gt., Ba., Dr.に対応)が単位行列の形に固定される

推定すべきパラメータが大幅に削減されることで音源分離の難易度が低下し、空間エイリアシングが発生している状況下でも効果的な分離行列の推定が可能

- パラメータ削減された分離行列をILRMAのコスト関数に代入することで導出
- パラメータ削減後の分離行列は上三角構造を有しており、一部の簡略化が可能



5. 被り音抑圧実験

被り音抑圧性能の比較実験

- 以下の3つの音源分離手法を用いて、それぞれVo.用マイクの被り音抑圧を実施
- TCNMF [Mizobuchi+, 2025] ILRMA Semi-blind ILRMA (提案手法)

- 被り音抑圧性能をSDR改善量で評価し、各手法の性能を比較

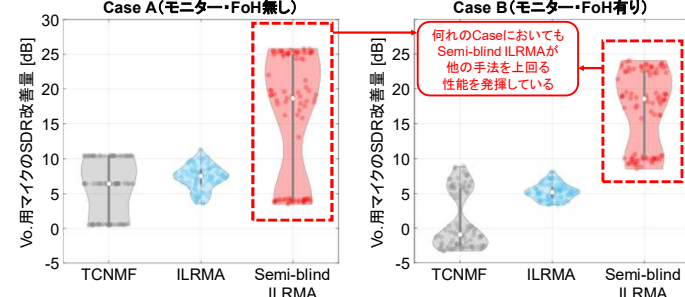
実験条件

- 観測信号は、測定で得たインパルス応答に3種のドライソースを重畳することで得る
- Semi-blind ILRMAではGt., Ba., 及びDr.観測信号をリファレンスとする
- Case A及びCase Bの2種類の観測信号に対して被り音抑圧を行う
- 窓長は分離性能に大きく影響するため、事前に探索によって最適な値を得る
- TCNMFの3つのハイパーパラメータに関しても同様に、事前に探索を行う
- 30個の異なる乱数シードで施行

その他のパラメータ	条件
窓関数	Blackman窓
シフト長	窓長の1/8
反復回数	100回
ILRMA及びSemi-blind ILRMAの基底数	10本

実験結果

- Vo.用マイクのSDR改善量をバイオリンプロットで表示(手法毎に3曲×30シード)



提案手法であるSemi-blind ILRMAがVo.用マイクの被り音を従来手法よりも効果的に抑圧することを、実際の音楽ライブ環境の測定データを用いて確認