卒業研究論文

論 文 題 目

マイクロホンアレイを用いた
ドラムセット音源分離のためのデータセット作成

提出年月日	令和8年　2月　19日
学　　科	電 気 情 報 工 学 科
氏　　名	森末 結　　印
指導教員(主査)	北村 大地 准教授　　印
副　　査	雛元 洋一 助教　　印
学　科　長	漆原 史朗 教授　　印

香川高等専門学校

Dataset creation for drum set source separation using microphone array

Yu Morisue

Department of Electrical and Computer Engineering
National Institute of Technology, Kagawa College

Abstract

In drum set recording, multi-track recording is generally employed, where close microphones are placed near each drum part (sound source). The purpose of this method is to record the performance sound of each source individually. However, even if the microphones are closely placed, it is almost impossible to completely prevent the interference of signals observed from other sound sources, which are called bleeding sound. Signals recorded with close microphones are mixed after being processed with equalizers and other effects optimized for each source. bleeding sound hinders the individual processing of each source during the mixing stage and unintentionally degrades sound quality. Therefore, techniques to reduce bleeding sound are required. Bleeding-sound reduction can be formulated as a source separation problem. Among source separation methods, blind source separation (BSS) targets signals observed with multiple microphones. There are several BSS methods have been proposed including independent component analysis (ICA), frequency-domain ICA (FDICA), and independent low-rank matrix analysis (ILRMA). However, BSS assumes the use of a microphone array with microphones arranged at narrow intervals. Consequently, it is difficult to apply BSS to bleeding-sound reduction in multichannel signals obtained by multi-track recording. In this thesis, I record a drum set using a microphone array and release for the purpose of using BSS for drum set bleeding-sound reduction. I also experimentally apply BSS to the recorded drum performance sounds to test the extent to which bleeding sound can be reduced. As a result, when simple ILRMA was applied, separation performance was low, and bleeding-sound reduction was not achieved. However, the effectiveness of bleeding-sound reduction was confirmed with ILRMA with ideal source model and FDICA with ideal permutation solver.

Keywords: Bleeding-sound reduction, drum set, microphone array

（和訳）

ドラムセットの録音では一般的に，各ドラムパーツ（音源）に近接マイクロホンを配置する．この録音方法をマルチトラック録音と呼ぶ．これは，各音源の演奏音を個別に録音することを目的としている．しかしながら，どれだけマイクロホンを近接させても他の音源から観測される信号（被り音）の混入を完全に防ぐことは困難である．近接マイクロホンで録音された信号は各音源に最適化されたイコライザー等で処理されたのちにミキシングされる．被り音はミキシングの過程における各音源への個別の処理を阻害し，意図せず音質を劣化させる．そのため，被り音を抑圧する技術が求められている．被り音抑圧は音源分離の問題としてとらえることができる．音源分離の中でも複数のマイクロホンで観測した信号を対象とした音源分離手法としてブラインド音源分離（blind source separation: BSS）がある．BSS には独立成分分析（independent component analysis: ICA）をはじめ，周波数領域 ICA（frequency-domain ICA: FDICA）や独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）などが提案されている．しかしながら，BSS はマイクロホンを狭い間隔で並べたマイクロホンアレイの使用を前提としている．そのため，マルチトラック録音で得られた多チャネル観測信号の被り音抑圧に BSS を適用することは困難である．そこで本論文では，ドラムセットの被り音抑圧に BSS を用いることを目的としてマイクロホンアレイを用いてドラムセットを録音し公開する．また，録音したドラムセット演奏音に試験的に BSS を適用し，どの程度被り音抑圧を行えるかを実験する．結果として，通常の ILRMA を適用した場合は分離性能が低く被り音抑圧ができるとは言えなかった．しかしながら，理想音源モデル型 ILRMA や理想パーミュテーション解決付き FDICA では，被り音抑圧への有効性が確認できた．

目次

第 1 章

緒言

1.1　ドラムセットの録音

ドラムセットは，Fig. 1.1 のようにキックドラム（kick drum: KD），スネアドラム（snare drum: SD），ハイハット（hi-hat cymbal: HH），クラッシュシンバル（crash cymbal: CC），及びタムタム（ハイタム（hi tom: HT），ロータム（low tom: LT），及びフロアタム（floor tom: FT））など，多くのドラムパーツからなる楽器である．本論文では以降，ドラムセットを構成する個々のドラムパーツを音源と呼ぶ．

ドラムセットの演奏を録音する際には，一般的に Fig. 1.2 のように各音源にマイクロホンを近接するように配置される [1]．ドラムセットに含まれる全ての音源にマイクロホンを近接配置するため，音源数と同じ数以上のマイクロホンを用いた録音となる．この録音方法をマルチトラック録音と呼ぶ．マルチトラック録音はマイクロホンを近接させた音源（目的音源）の音（目的音）のみを録音することを目的としている．しかしながら，Fig. 1.3 に示すように，実際には目的音源以外の音源からの音も混入してしまう．この混入する音を被り音と呼ぶ．

マルチトラック録音により得られた複数の観測信号は各音源に対して個別の音質調整処理を適用し，それらをミキシングすることで最終的なドラムセット全体の演奏音を得る．そのため，近接マイクロホンから得られる観測信号に被り音が含まれる場合，ミキシングの過程で意図せず音質を低下させてしまう可能性がある．以上の理由により，ドラムセットを録音する際には，被り音の混入を防ぐ必要がある．しかしながら，最適な配置に設定したとしてもなお，被り音の混入を完全に防ぐことは不可能である．従って，各マイクロホンで録音された信号に対して，被り音を抑圧する技術が求められている．

1.2　既存の音源分離技術との関連性

前節で述べたように，マルチトラック録音で得られた信号に対して被り音を抑圧する技術が求められている．この被り音抑圧は音源分離に似た問題である．特に，複数のマイクロホンで観測された信号を対象とするブラインド音源分離（blind source separation: BSS）[2, 3, 4, 5] とドラムセットにおけるマルチトラック録音の被り音抑圧は非常に類似した問題といえる．

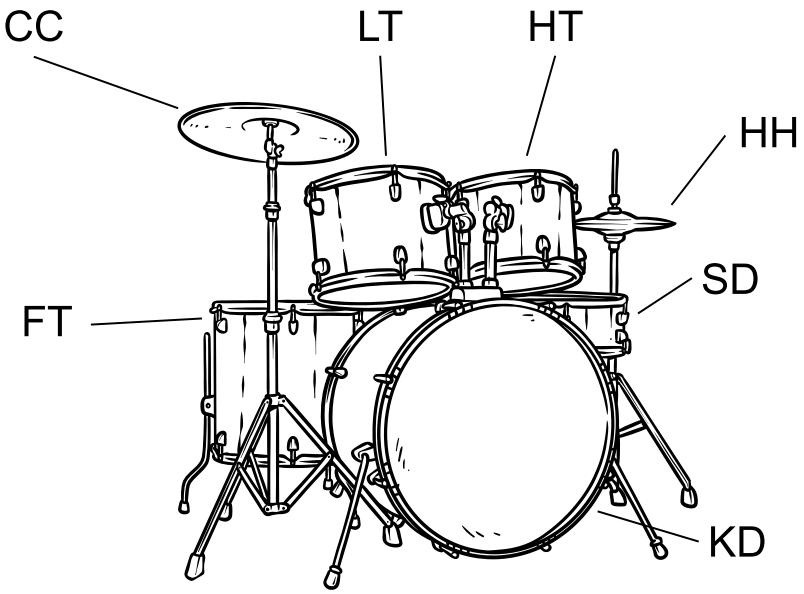


Fig. 1.1: Sound source components of drum set.

(a)

(b)

(c)

(d)

Fig. 1.2: Typical close-miking setup in drum set recording for (a) KD, (b) SD, (c) HH, and (d) CC.

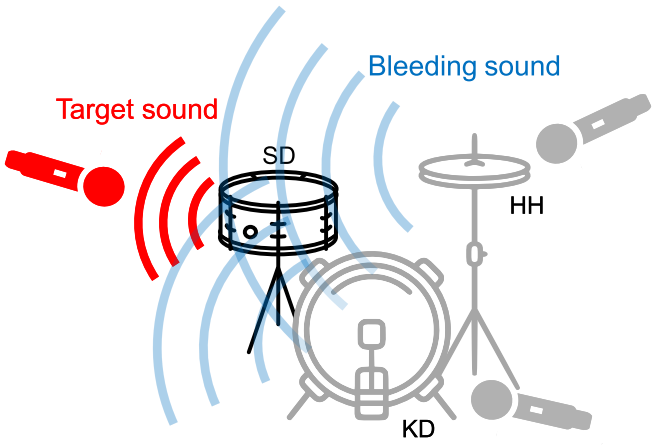


Fig. 1.3: Target sound and bleeding sounds for SD microphone.

BSS とは，マイクロホンの配置や音源の位置が未知の状態で分離系を推定する手法である．BSS には，独立成分分析（independent component analysis: ICA）[6] をはじめとして，周波数領域 ICA（frequency-domain ICA: FDICA）[3]，及び独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[4, 5] などの手法が提案されている．BSS は周波数領域で推定される分離行列を多チャネル観測信号に乗じる形で音源を分離するため線形な音源分離と解釈され，比較的高品質な音質を担保した音源分離を実現できる傾向にある．しかしながら，BSS は一般的に Fig. 1.4 のような，数 mm から数十 mm の狭い間隔で複数のマイクロホンを並べたマイクロホンアレイを用いた観測信号を対象とする．その理由は，マイクロホンの間隔が大きく離れて配置された場合の観測信号には，空間エイリアシングという問題が発生してしまうからである．空間エイリアシングが発生した観測信号に対しては，BSS を適用しても分離系の推定に失敗してしまう．ドラムセットのマルチトラック録音では，各音源に配置された近接マイクロホン間の距離が 1 m 以上離れることがあり，空間エイリアシング問題を回避できない．従って，ドラムセットのマルチトラック録音信号に対して愚直に BSS を適用しても，被り音抑圧は通常失敗してしまう．

一方，近年ではドラムセットを構成する各音源を分離するドラム音源分離（drum source separation: DSS）という問題も提起されるようになり，そのためのデータセットが公開されている [7]．中でも，深層ニューラルネットワーク（deep neural network: DNN）を用いた非線形な DSS が検討されている．しかしながら，高品質で芸術性を損なわない各音源の分離は非常に難しく，現在は十分な性能での DSS は達成されていない．被り音抑圧においても DSS は流用可能と考えられるが，前述の通り，高精度な DSS の実現に DNN を用いることが最善かどうかは不明である．BSS のように線形な音源分離を実現し，ある程度分離音の音質が担保された手法のほうが DSS に適している可能性も十分に考えられる．

Fig. 1.4: Microphone array arranged at millimeter-order intervals.

1.3　本論文の目的

本論文では，マルチトラック録音されたドラムセットの多チャネル観測信号における被り音の抑圧を最終的な目的とし，BSS に基づくアプローチの可能性を開拓するための基礎的な検討を行う．特に，DNN に基づく DSS を応用した被り音抑圧ではなく，マイクロホンアレイと BSS を用いた被り音抑圧に焦点を当てる．これは，線形な音源分離メカニズムを利用する BSS が，DNN に基づく手法より推定信号の音質を担保しやすいことを理由としている．

通常のドラムセットの録音は Fig. 1.2 のように各音源の近接マイクロホンのみを用いるが，本論文ではこの録音形態を拡張し，Fig. 1.4 のようなマイクロホンアレイを追加することと考える．このような録音機材の追加は一般にコストの増加を招くリスクを伴うが，その反面ドラムセットの被り音抑圧という難しい問題を解決できる可能性を高められる利点がある．マイクロホンアレイを用いた観測信号が得られることで，空間エイリアシング問題の発生を回避しながら BSS を適用することが可能となる．

マイクロホンアレイを用いたドラムセットの録音信号は現在データセットとして公開されていないため，本論文ではまず通常の近接マイクロホンとマイクロホンアレイを用いてドラムセットの演奏を録音したデータセットを構築し，今後の技術発展のためにこれを公開する．さらに，この観測信号に対して単純な既存の BSS を適用した場合に，どの程度の被り音抑圧性能が得られるかについて実験的に調査し，今後の更なる手法の拡張に必要な知見を得ることを目的とする．

1.4　本論文の構成

2 章では，ドラムセットの録音や BSS の前提知識及び 4 章の実験で使用する理想的な BSS 手法について説明する．また，被り音抑圧に向けた新しいアプローチについて述べる．3 章ではマイクロホンアレイを用いたドラムセットの録音及び録音したデータセットの詳細について述べる．4 章では，録音したデータセットに対して ILRMA と FDICA を適用した実験結果について述べる．5 章では，本論文の結論と今後の課題について述べる．

第 2 章

ブラインド音源分離とドラムセットの被り音抑圧への応用

2.1　まえがき

本論文では，前章で述べた通り，マイクロホンアレイを用いてドラムセットを録音することを新たに考え，BSS による被り音抑圧の実現可能性について調査する．本章では，前述の内容の遂行に必要な BSS の詳細と，ドラムセットの被り音抑圧問題に BSS を用いることの動機について述べる．2.2 節では，ドラムセットのマルチトラック録音における被り音問題について具体的な例を示しながら説明する．2.3 節では，BSS の理論及び原理について詳しく説明する．2.4 節では，BSS における上限性能を確認するためのアルゴリズムについて説明する．2.5 節では，BSS の性能が劣化する原因となる空間エイリアシング問題について説明する．2.6 節では，本論文における調査内容の動機としてマイクロホンアレイを用いてドラムセットの録音を行うこと及び BSS に基づく被り音抑圧を検討することの動機を述べる．最後に，2.7 節で本章のまとめを述べる．

2.2　マルチトラック録音における被り音問題

1 章で述べたように，ドラムセットのマルチトラック録音は Fig. 1.2 のように，ドラムセットを構成する各音源に近接マイクロホンを配置する録音方法が一般的である．音楽ライブ演奏におけるドラムセットの録音や音楽制作のためのレコーディングにおいても同様の手法が用いられている．各音源に近接マイクロホンを配置する理由は，特定の音源からの音響信号（目的音）のみを録音するためである．しかしながら，プロフェッショナルのサウンドエンジニアがあらゆる条件を考慮してマイクロホンを配置しても，被り音の混入を完全に防ぐことは不可能である．各音源の近接マイクロホンから得られた複数の観測信号は，Fig. 2.1 のようにそれぞれの音源に対して最適化されたヘッドアンプ（head amplifier: HA），コンプレッサー，及びイコライザーを適用させた後にミキシングを行う．被り音を含む観測信号に，目的音源に最適

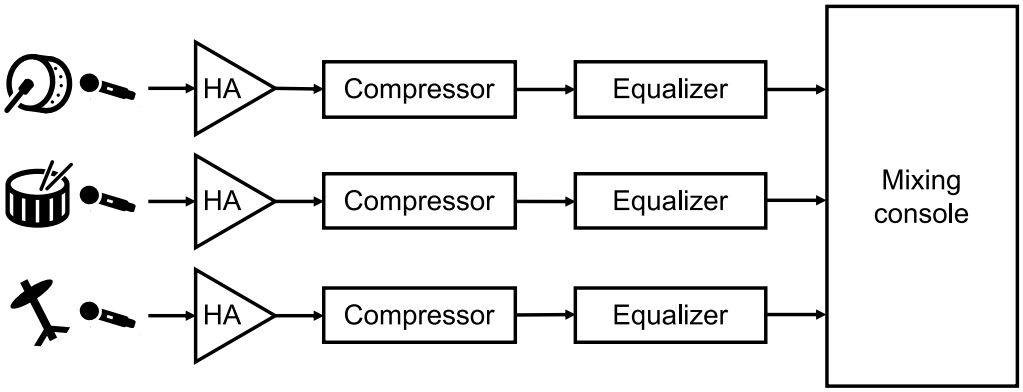


Fig. 2.1: Block diagram of source mixing system.

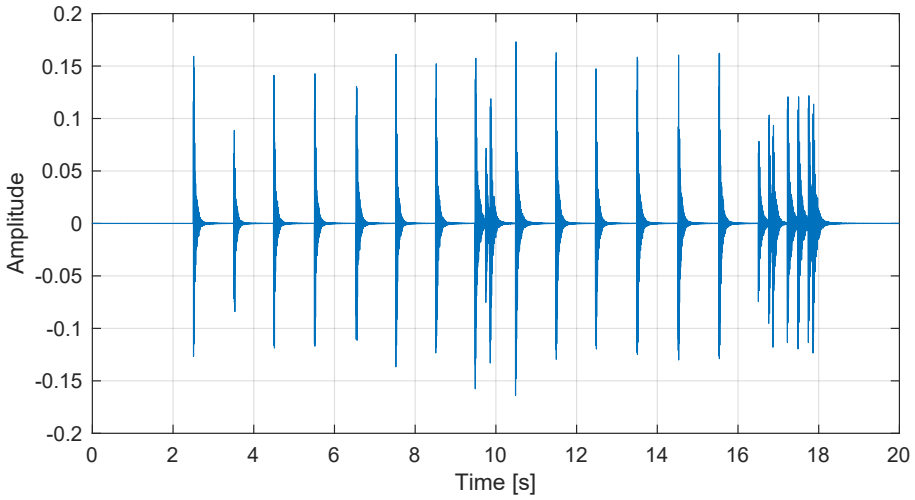
化された音質調整処理を適用すると意図せず音質を低下させる原因となる．従って，マルチトラック録音で得られた観測信号から被り音を抑圧する技術が求められる．

時間的な制約の緩い音楽制作のレコーディング現場では，被り音を完全に回避しながらドラムセットの演奏を録音する手法として，KD のみや SD のみのように音源毎に独立させた演奏及び録音が考えられるが，ドラム奏者にとってはこのような音源毎の演奏は困難であるため非現実的である．現在のレコーディング現場では，ドラムセットのマルチトラック録音における被り音の問題を緩和するために，トリガと呼ばれる技術 [8] が多用されている．これは，近接マイクロホンの観測信号や別のセンサを用いて，ドラム奏者の演奏情報（叩打のタイミングや強弱）をディジタルデータに変換し，これを元に外部のソフトウェアドラム音源（あらかじめ録音されたドラムの各音源のサンプル音）を再生し差し替える又はミックスする技術である．この事実からも，ドラムセットのマルチトラック録音における被り音問題の深刻さや解決の困難さがうかがえる．

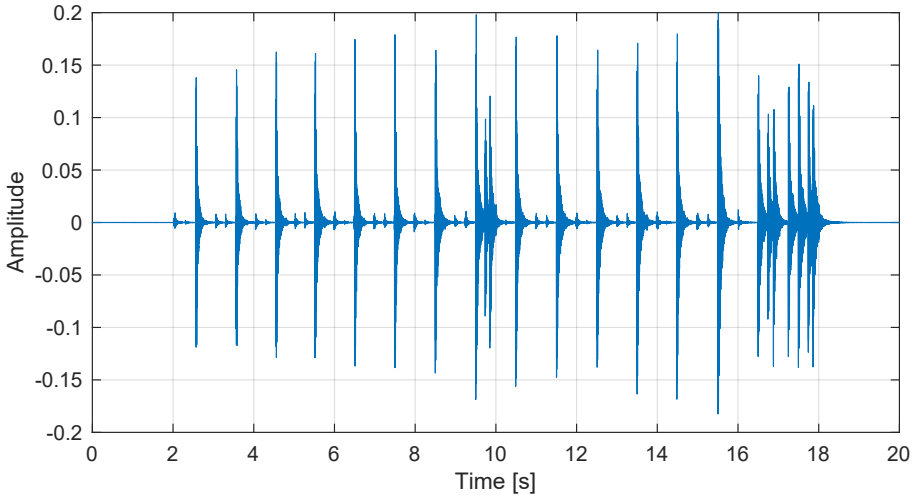
実際に，SD の近接マイクロホンで観測した信号の波形とスペクトログラムを Figs. 2.2 及び 2.3 にそれぞれ示す．ここで，Figs. 2.2 (a) 及び 2.3 (a) は SD のみを演奏した時の観測信号であり，Figs. 2.2 (b) 及び 2.3 (b) は SD，KD，HH，及び CC を同時に演奏した時の観測信号である．従って，Figs. 2.2 (b) 及び 2.3 (b) のみに表れている成分が全て，SD の近接マイクロホンに混入する被り音成分に対応する．例えば，Fig. 2.2 (b) の波形を見れば，SD の叩打の間に被り音成分が確認でき，その周波数成分も Fig. 2.3 (b) に表れている．これらの図から，被り音の影響が信号波形とスペクトログラムに顕著に表れていることが分かる．

2.3 ブラインド音源分離

ドラムセットのマルチトラック録音における被り音抑圧問題は，目的音及び被り音が混合した多チャネル観測信号から目的音のみを分離・推定する BSS の問題として解釈できる．BSS とは，Fig. 2.4 に示すように，混合後の観測信号が与えられ，マイクロホンの空間的な位置，音源の空間的な位置やその種類，及びその学習データなどが一切与えられない（ブラインド



(a)



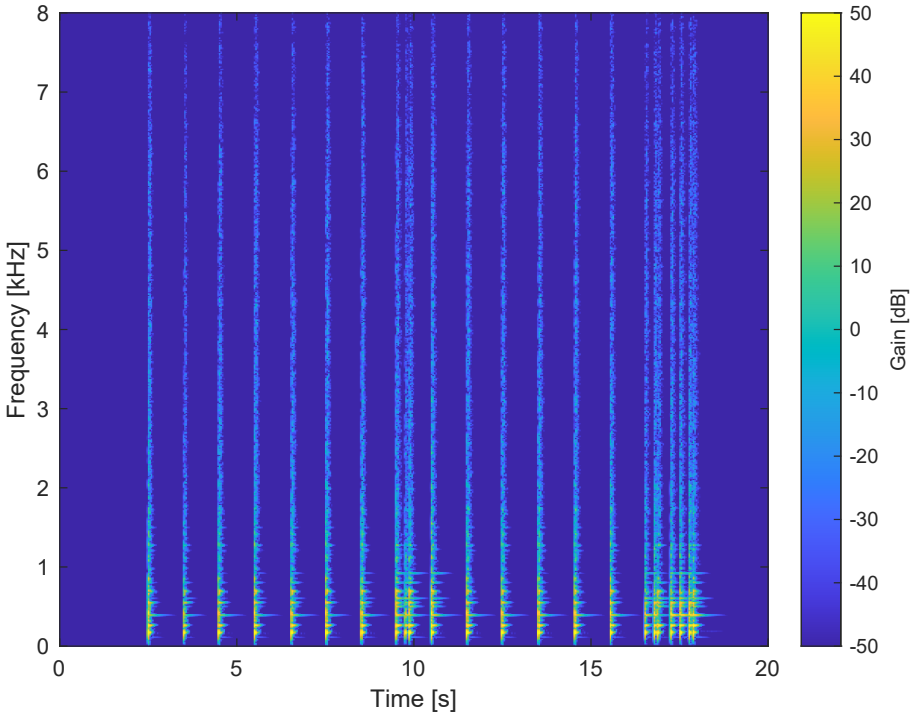
(b)

Fig. 2.2: Waveforms recorded by the SD close microphone (a) when only the SD is played and (b) when all sources are played.

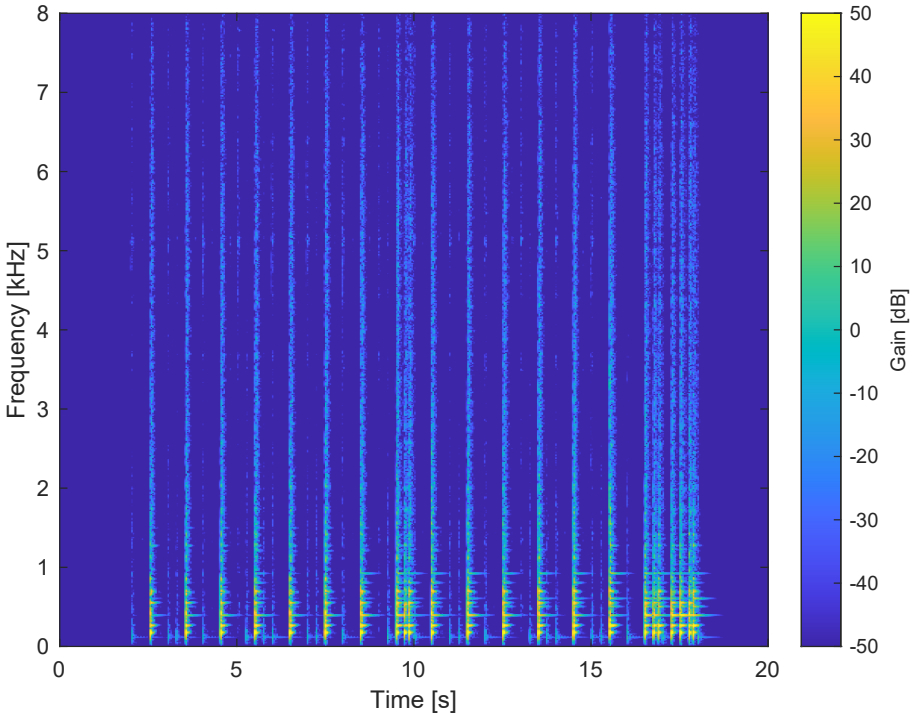
な）条件で，混合前の音源信号を推定する技術である．本節では，BSS の代表的な手法として，ICA，FDICA，及び ILRMA の詳細を示す．

2.3.1　ICA

ICA [6] は，音源信号が時間領域での瞬時混合されること及び音源信号間の統計的性質を仮定した手法である．今，N 個の音源信号の混合を M 個のマイクロホンで観測する状況を考える．このとき，n 番目の音源信号を $\tilde{s}_n(l)$ と表し，m 番目のマイクロホンの観測信号を $\tilde{x}_m(l)$ と表す．ここで，$n = 1, 2, \cdots, N$ 及び $m = 1, 2, \cdots, M$ はそれぞれ音源及びマイクロホン



(a)



(b)

Fig. 2.3: Spectrograms recorded by the SD close microphone (a) when only the SD is played and (b) when all sources are played.

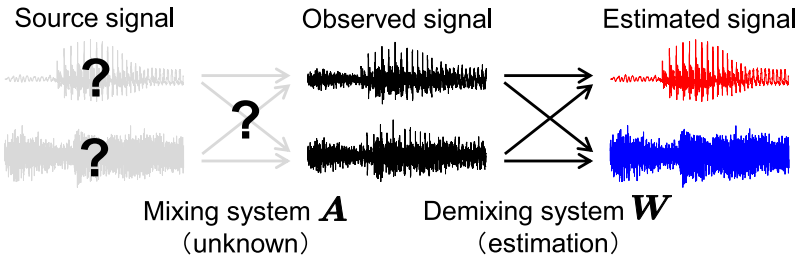


Fig. 2.4: Overview of BSS.

（チャネル）のインデクスである．また，$l = 1, 2, \cdots, L$ は時間サンプルのインデクスを表す．さらに，BSS で推定された n 番目の分離信号を $\tilde{y}_n(l)$ と表す．これらの音源信号，観測信号，及び推定信号をベクトルとしてまとめて，次式で表す．

$$\tilde{\boldsymbol{s}}(l) = [\tilde{s}_1(l), \tilde{s}_2(l), \cdots, \tilde{s}_N(l)]^{\mathrm{T}} \in \mathbb{R}^N \tag{2.1}$$

$$\tilde{\boldsymbol{x}}(l) = [\tilde{x}_1(l), \tilde{x}_2(l), \cdots, \tilde{x}_M(l)]^{\mathrm{T}} \in \mathbb{R}^M \tag{2.2}$$

$$\tilde{\boldsymbol{y}}(l) = [\tilde{y}_1(l), \tilde{y}_2(l), \cdots, \tilde{y}_N(l)]^{\mathrm{T}} \in \mathbb{R}^N \tag{2.3}$$

ここで，$.^{\mathrm{T}}$ はベクトル及び行列の転置を表す．N 個の音源信号が時間領域で瞬時混合する場合，観測信号は次式で表される．

$$\tilde{\boldsymbol{x}}(l) = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}}(l) \tag{2.4}$$

ここで，$\tilde{\boldsymbol{A}} \in \mathbb{R}^{M \times N}$ は時不変な混合行列である．また，観測信号 $\tilde{\boldsymbol{x}}(l)$ から推定信号 $\tilde{\boldsymbol{y}}(l)$ を得る処理は次式となる．

$$\tilde{\boldsymbol{y}}(l) = \tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}(l) \tag{2.5}$$

ここで，$\tilde{\boldsymbol{W}} \in \mathbb{R}^{N \times M}$ は時不変な分離行列である．即ち，式 (2.4) は音源信号の混合過程を表し，式 (2.5) は観測信号の分離過程を表す．もし $\tilde{\boldsymbol{A}}$ が正則ならば，$\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{A}}^{-1}$ を推定することで，観測信号 $\tilde{\boldsymbol{x}}(l)$ から個々の音源に分離された推定信号 $\tilde{\boldsymbol{y}}(l)$（$= \tilde{\boldsymbol{s}}(l)$）を求めることができる．この $\tilde{\boldsymbol{W}}$ を求めることが ICA の目的である．

$\tilde{\boldsymbol{W}}$ を求めるために，ICA では以下の統計的性質を仮定する．

1. 各音源信号 $\tilde{s}_1(l), \tilde{s}_2(l), \cdots, \tilde{s}_N(l)$ は互いに統計的に独立である
2. 各音源信号 $\tilde{s}_1(l), \tilde{s}_2(l), \cdots, \tilde{s}_N(l)$ は非ガウス分布に従う

これらの仮定に基づき，推定信号 $\tilde{y}_1(l), \tilde{y}_2(l), \cdots, \tilde{y}_N(l)$ が互いにできるだけ独立となる分離行列 $\tilde{\boldsymbol{W}}$ を求めるための ICA の最適化問題は，コスト関数 $\mathcal{J}(\tilde{\boldsymbol{W}})$ を用いて次式で表される．

$$\underset{\tilde{\boldsymbol{W}}}{\mathrm{Minimize}} \mathcal{J}(\tilde{\boldsymbol{W}}) \tag{2.6}$$

$$\mathcal{J}(\tilde{\boldsymbol{W}}) = \int_{-\infty}^{\infty} p(\tilde{\boldsymbol{y}}) \log \frac{p(\tilde{\boldsymbol{y}})}{\prod_n p(\tilde{y}_n)} d\boldsymbol{y} \tag{2.7}$$

ここで，$p(\tilde{y}_n)$ は n 番目の分離信号 $\tilde{y}(l)$ の周辺確率密度関数であり，$p(\tilde{\boldsymbol{y}}) = p(\tilde{y}_1(l), \tilde{y}_2(l), \cdots, \tilde{y}_N(l))$ は同時確率密度関数である．最適化問題 (2.6) はコスト関数 $\mathcal{J}(\tilde{\boldsymbol{W}})$ を最小化する $\tilde{\boldsymbol{W}}$ を求めることを意味し，コスト関数を最小化する $\tilde{\boldsymbol{W}}$ は推定信号間の独立性を最大化する系である．このような系は，前述の 2 つの仮定より，観測信号から音源信号を分離する分離系となる．

最適化問題 (2.6) を用いて分離行列 $\tilde{\boldsymbol{W}}$ を推定した場合，推定信号 $\tilde{\boldsymbol{y}}(l)$ はその振幅の大きさや音源の順序を決定できない．つまり，ICA によって推定される分離行列 $\hat{\boldsymbol{W}} \in \mathbb{R}^{N \times M}$ には，以下の任意性が存在する．

1. 分離信号の音源の順序の任意性
2. 分離信号のスケールの任意性

これらの任意性は分離信号に対して Fig. 2.5 のように現れる．これらの任意性を，求めたい分離行列 $\tilde{\boldsymbol{W}}$ 及び推定した分離行列 $\hat{\boldsymbol{W}}$ の関係で表すと次式となる．

$$\hat{\boldsymbol{W}} = \boldsymbol{D}\boldsymbol{P}\tilde{\boldsymbol{W}} \tag{2.8}$$

ここで，$\boldsymbol{D} \in \mathbb{R}^{N \times N}$ は推定信号のスケールを変える対角行列，$\boldsymbol{P} \in \mathbb{R}^{N \times N}$ は推定信号の順序を変える置換行列（パーミュテーション行列）である．前述の音源順序の任意性は $\boldsymbol{P}$ に現れ，スケールの任意性は $\boldsymbol{D}$ に現れる．スケールの任意性に関しては，$\boldsymbol{D}^{-1}$ を解析的に求めるプロジェクションバック（projection back: PB）法 [9] と呼ばれる補正手法が提案されている．推定信号に PB 法を適用すると，n 番目の推定信号 $\hat{y}_n(l)$ は次式で計算される．

$$\hat{y}_n(l) = \hat{\boldsymbol{W}}^{-1}(\boldsymbol{e}_n \odot \tilde{\boldsymbol{y}}(l)) \tag{2.9}$$

ここで，$\boldsymbol{e}_n \in \mathbb{R}^N$ は n 番目の要素が 1，その他の要素が 0 のベクトルであり，$\odot$ はアダマール積（要素毎の積）を示す．式 (2.9) は $\hat{\boldsymbol{W}}$ に含まれる $\boldsymbol{D}$ の影響のみを打ち消す計算に対応する．

2.3.2 FDICA

2.3.1 項で，ICA は音源信号が時間領域で瞬時混合することを仮定すると述べた．しかしながら，現実の音響信号の混合では，音が各マイクロホンに到達するまでの時間差や録音環境の残響により，複数音源の混合は式 (2.4) とは異なり，各マイクロホン及び各音源間の伝達関数（インパルス応答）と音源信号の畳み込み混合となる．残響を含む信号に対する音源分離は，畳み込み混合系の逆系を推定することで達成できる．しかしながら，時間領域における畳み込み混合系の逆系の推定は計算が困難である．そこで，音響信号の BSS では畳み込み混合を積和混合に変換するために，観測信号 $\tilde{\boldsymbol{x}}(l)$ を短時間フーリエ変換（short-time Fourier transform: STFT）して時間周波数領域でモデル化する FDICA [3] が提案された．FDICA では，観測信号 $\tilde{\boldsymbol{x}}(l)$ を STFT して得られる複素スペクトログラムの各周波数において，独立な（複素数の）ICA を適用し，分離信号の複素スペクトログラムを推定する．

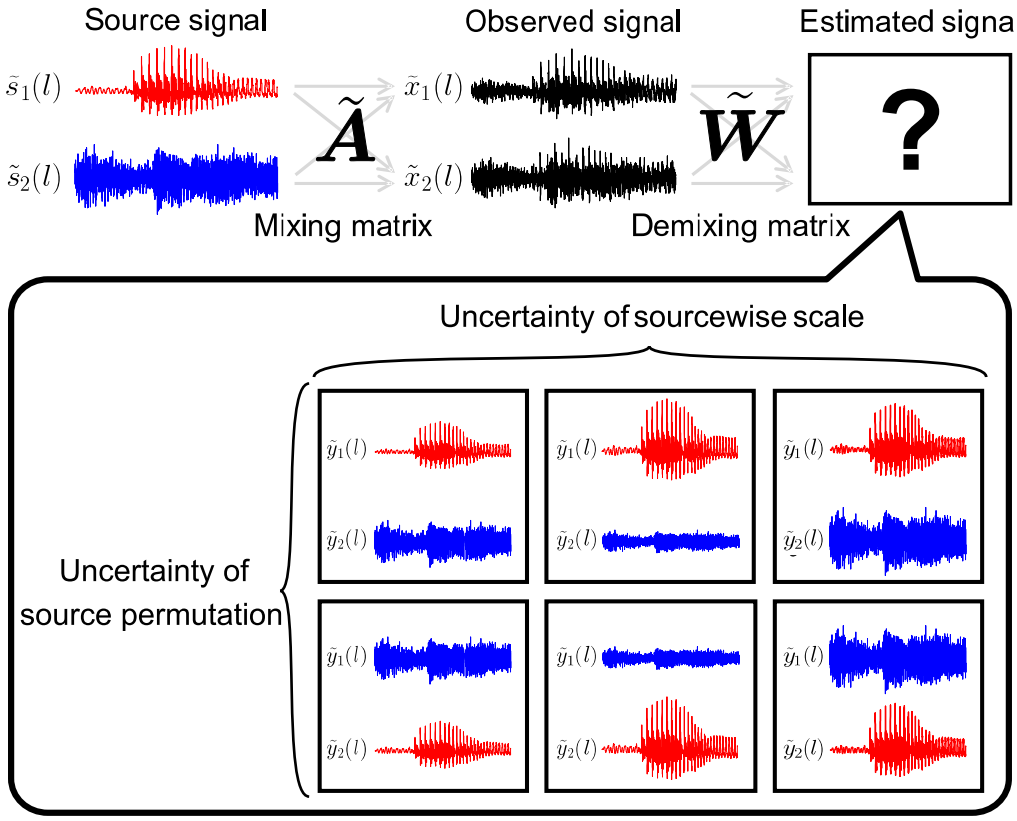


Fig. 2.5: Uncertainty of source permutation and sourcewise scale in ICA.

今, 音源信号 $\tilde{s}_n(l)$, 観測信号 $\tilde{x}_m(l)$, 及び推定信号 $\tilde{y}_n(l)$ に STFT を適用して得られる各信号の複素スペクトログラム行列をそれぞれ $\mathbf{S}_n \in \mathbb{C}^{I\times J}$, $\mathbf{X}_m \in \mathbb{C}^{I\times J}$, 及び $\mathbf{Y}_n \in \mathbb{C}^{I\times J}$ と定義する. また各行列の要素をそれぞれ s_{ijn}, x_{ijm}, 及び y_{ijn} と表す. ここで, $i = 1, 2, \cdots, I$ 及び $j = 1, 2, \cdots, J$ はそれぞれ周波数ビン及び時間フレームのインデクスである. このとき, 音響信号の畳み込み混合系は周波数毎の瞬時混合となり, 分離系と合わせて次式でモデル化できる.

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij} \tag{2.10}$$
$$\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij} \tag{2.11}$$

ここで,

$$\mathbf{s}_{ij} = [\mathrm{s}_{ij1}, \cdots, \mathrm{s}_{ijn}, \cdots, \mathrm{s}_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \tag{2.12}$$
$$\mathbf{x}_{ij} = [\mathrm{x}_{ij1}, \cdots, \mathrm{x}_{ijm}, \cdots, \mathrm{x}_{ijM}]^{\mathrm{T}} \in \mathbb{C}^M \tag{2.13}$$
$$\mathbf{y}_{ij} = [\mathrm{y}_{ij1}, \cdots, \mathrm{y}_{ijn}, \cdots, \mathrm{y}_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \tag{2.14}$$

と定義する. また, $\mathbf{A}_i \in \mathbb{C}^{M\times N}$ は周波数毎の複素混合行列, $\mathbf{W}_i \in \mathbb{C}^{M\times N}$ は周波数毎の複素分離行列である. 式 (2.10) 及び (2.11) の混合系及び分離系のモデルを Fig. 2.6 に示す. Fig. 2.6 (a) は混合系を表しており, 周波数毎の複素混合行列 $\mathbf{A}_i$ が音源スペクトログラムの各時間周波数要素 $\mathbf{s}_{ij}$ に乗算されており, その結果観測信号の各時間周波数要素 $\mathbf{x}_{ij}$ となって

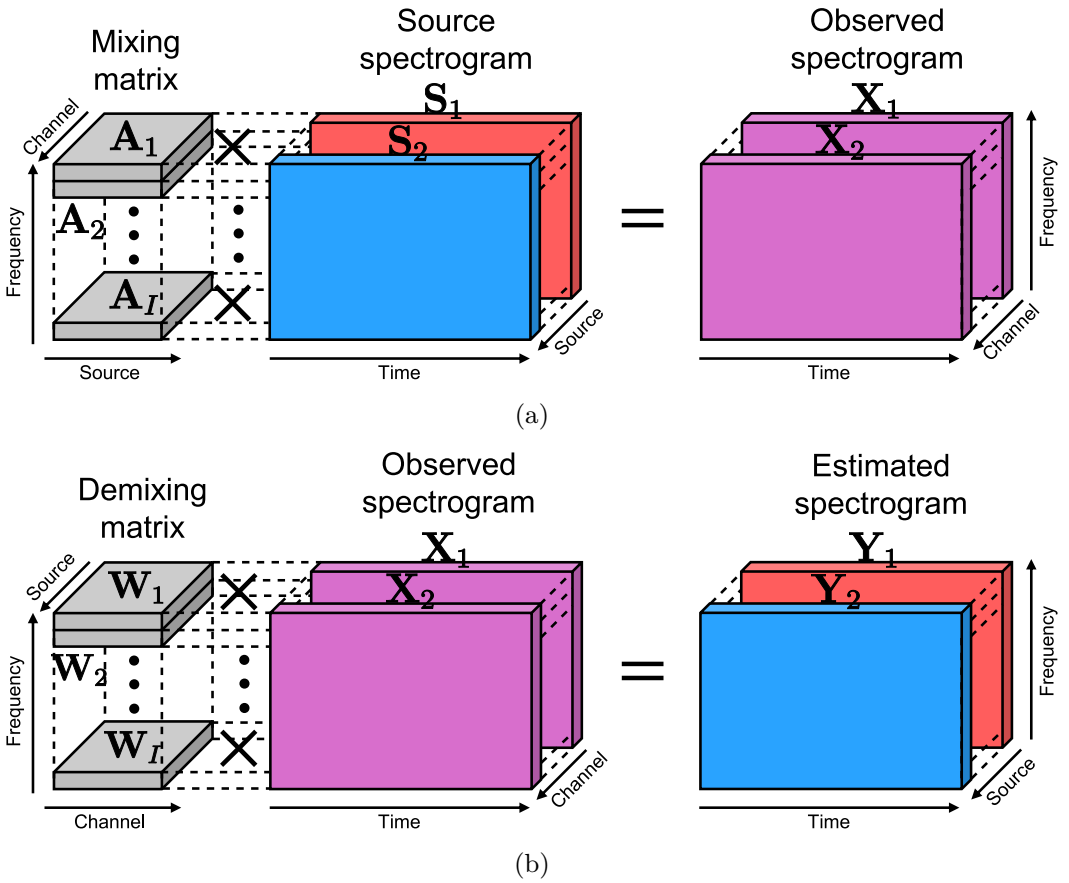


Fig. 2.6: Models of (a) mixing and (b) demixing systems and observed, source, and estimated signals, where $N = M = 2$.

得られている．また，Fig. 2.6 (b) は分離系を表しており，$\mathbf{W}_i = \mathbf{A}_i^{-1}$ なる分離行列を観測信号の各時間周波数要素 $\mathbf{x}_{ij}$ に乗じることで，推定信号の各時間周波数要素 $\mathbf{y}_{ij}$ が得られる．

2.3.1 項で述べたように，ICA では分離信号が非ガウス分布に従うと仮定する．ここで分離信号が従うと仮定する確率分布を音源モデルと呼ぶ．また，ICA の推定信号には順序及びスケールの任意性がある．FDICA では周波数毎に独立な ICA が適用されるため，各周波数ビンで推定した推定信号の複素スペクトログラムの音源順序（パーミュテーション）及びスケールが不統一になるという問題が生じる．周波数毎のスケールの任意性については，式 (2.9) の PB 法 [9] により解析的に復元可能である．一方で，Fig. 2.7 に示すような周波数毎の音源パーミュテーションの順序を適切に並び替えることは困難である．この問題は FDICA におけるパーミュテーション問題と呼ばれている．これを解決することが FDICA における大きな課題であり，現在に至るまで様々な解決方法が提案されている [10, 11, 12, 13, 14].

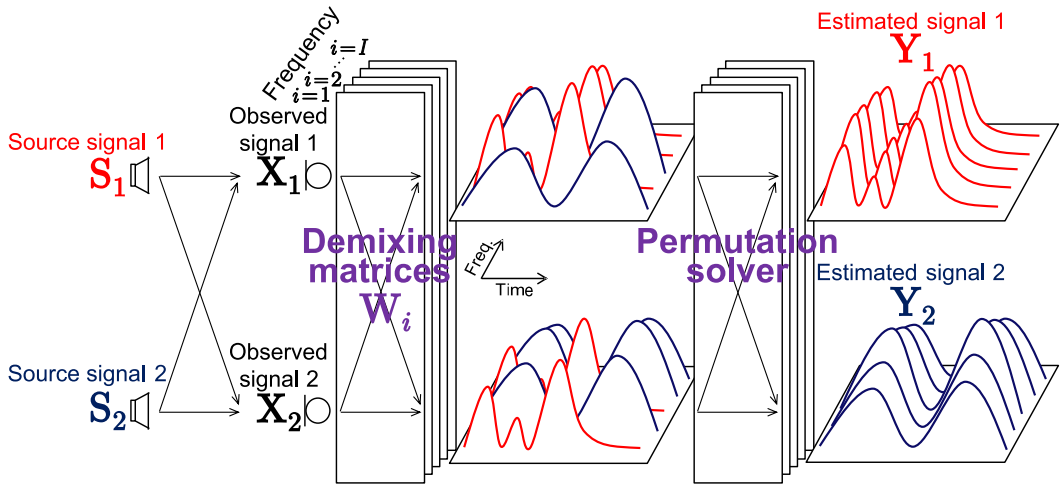


Fig. 2.7: Permutation problem in FDICA.

2.3.3　ILRMA

FDICA におけるパーミュテーション問題をできるだけ回避しながら分離行列 $\mathbf{W}_i$ を推定する技術として，ILRMA [4, 5] が提案されている．ILRMA は，FDICA の音源モデルに非負値行列因子分解（nonnegative matrix factorization: NMF）[15, 16] を導入することで，推定信号のパワースペクトログラムを低ランク近似しながら分離行列を推定するアルゴリズムである．ILRMA による BSS の概要を Fig. 2.8 に示す．図中の $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I\times K}$ 及び $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K\times J}$ は，n 番目の推定音源のパワースペクトログラム $|\mathbf{Y}_n|^{\cdot 2}$ を NMF で低ランク近似した時間周波数モデルを構成する行列であり，$|\mathbf{Y}_n|^{\cdot 2} \approx \boldsymbol{T}_n\boldsymbol{V}_n$ としてモデル化されている．ここで，行列に対する演算子 $|\cdot|^{\cdot 2}$ は，要素毎の絶対値と 2 乗をとった同サイズの行列を表す．$\boldsymbol{T}_n$ を基底行列，$\boldsymbol{V}_n$ をアクティベーション行列と呼ぶ．ILRMA は，音源間の統計的独立性に基づく分離行列 $\mathbf{W}_i$ の反復最適化と NMF の低ランク近似による推定信号のパワースペクトログラム $|\mathbf{Y}_n|^{\cdot 2}$ の低ランクモデル $\boldsymbol{T}_n\boldsymbol{V}_n$（音源モデル）を反復的に交互最適化するアルゴリズムである．これにより，推定信号のパワースペクトログラム $|\mathbf{Y}_n|^{\cdot 2}$ がいずれの音源も低ランクに近づくように分離行列 $\mathbf{W}_i$ が推定される．混合前の音源信号のパワースペクトログラム $|\mathbf{S}_n|^{\cdot 2}$ が本来低ランクである場合，それらが混合した観測信号のパワースペクトログラム $|\mathbf{X}_m|^{\cdot 2}$ はランクが増大するため，低ランクなパワースペクトログラムを持つような分離行列 $\mathbf{W}_i$ を推定することは，パーミュテーション問題の回避とより高精度な BSS を促す作用があり，多くの実験においてこの原理の有効性が確認されている [4, 5].

ILRMA の最適化問題は次のように定義される [4, 5].

$$\underset{\mathbf{W}_1,\cdots,\mathbf{W}_I,\boldsymbol{T}_1,\cdots,\boldsymbol{T}_N,\boldsymbol{V}_1,\cdots,\boldsymbol{V}_N}{\text{Minimize}} \; -2J\sum_i \log|\det \mathbf{W}_i| + \sum_{i,j,n}\left(\frac{|\mathbf{w}_{in}^{\mathrm{H}}\mathbf{x}_{ij}|^2}{\sum_k t_{ikn}v_{kjn}} + \log\sum_k t_{ikn}v_{kjn}\right) \tag{2.15}$$

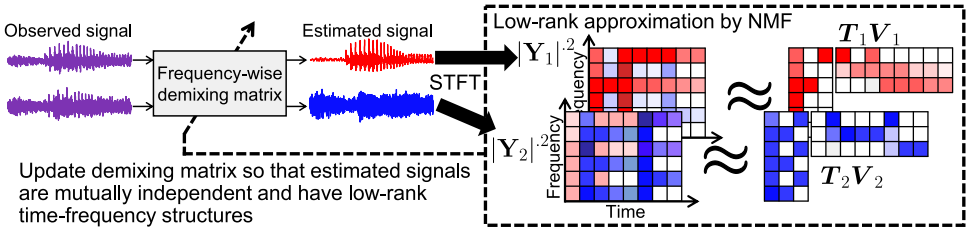


Fig. 2.8: Separation principle of ILRMA.

ここで，t_{ikn} 及び v_{kjn} はそれぞれ $\boldsymbol{T}_n$ 及び $\boldsymbol{V}_n$ の要素，$\mathbf{w}_{in}$ は n 番目の音源を推定するための分離フィルタ（即ち，分離行列 $\mathbf{W}_i$ の n 行目の行ベクトル），$\det\cdot$ は行列式，$\cdot^{\mathrm{H}}$ はエルミート転置を表す．最適化問題 (2.15) を最小化する基底行列及びアクティベーション行列の要素は，次に示す反復最適化アルゴリズムで推定される．

$$t_{ikn} \leftarrow t_{ikn}\sqrt{\frac{\sum_j |\mathbf{w}_{in}^{\mathrm{H}}\mathbf{x}_{ij}|^2 v_{kjn}\left(\sum_{k'} t_{ik'n}v_{k'jn}\right)^{-2}}{\sum_j v_{kjn}\left(\sum_{k'} t_{ik'n}v_{k'jn}\right)^{-1}}} \tag{2.16}$$

$$v_{kjn} \leftarrow v_{kjn}\sqrt{\frac{\sum_i |\mathbf{w}_{in}^{\mathrm{H}}\mathbf{x}_{ij}|^2 t_{ikn}\left(\sum_{k'} t_{ik'n}v_{k'jn}\right)^{-2}}{\sum_i t_{ikn}\left(\sum_{k'} t_{ik'n}v_{k'jn}\right)^{-1}}} \tag{2.17}$$

一方，分離行列 $\mathbf{W}_i$ に関する最適化は，補助関数法 [17] 及び反復射影法（iterative projection: IP）を用いた最適化アルゴリズム [18] によって，高速かつ安定的に解くことができる．その反復更新式は次式となる．

$$\mathbf{U}_{in} = \frac{1}{J}\sum_j \frac{1}{\sum_k t_{ikn}v_{kjn}}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{H}} \tag{2.18}$$

$$\mathbf{w}_{in} \leftarrow (\mathbf{W}_i\mathbf{U}_{in})^{-1}\boldsymbol{e}_n \tag{2.19}$$

$$\mathbf{w}_{in} \leftarrow (\mathbf{w}_{in}^{\mathrm{H}}\mathbf{U}_{in}\mathbf{w}_{in})^{-\frac{1}{2}} \tag{2.20}$$

これらの更新式は，1 回の更新の前後で最適化問題 (2.15) 中の目的関数の値が単調非増加となることが保証されている．

2.4　理想的な BSS

2.3 節で述べた BSS の FDICA や ILRMA では，次の手法を用いることで，上限に近い性能を得ることが可能となる．

- FDICA で得られる分離信号 $\mathbf{Y}_1,\cdots,\mathbf{Y}_N$ に対して，完全に分離された音源信号 $\mathbf{S}_1,\cdots,\mathbf{S}_N$ を用いて正確にパーミュテーション問題を解決する
- ILRMA で仮定する NMF 音源モデルを，完全に分離された音源信号 $\mathbf{S}_1,\cdots,\mathbf{S}_N$ のパワースペクトログラムに置き換えて固定し分離行列を推定する

本論文では，前者の方法を理想パーミュテーション解決付き FDICA，後者の方法を理想音源モデル型 ILRMA と呼ぶ．現実には，完全に分離された音源信号を得ることは不可能なため，これらの手法を実際の問題に適用することは不可能であるが，FDICA や ILRMA に基づく BSS の理想的な性能を調査する目的で実験的に用いることができる．下記では，理想パーミュテーション解決付き FDICA 及び理想音源モデル型 ILRMA の詳細を示す．

2.4.1　理想パーミュテーション解決付き FDICA

理想パーミュテーション解決付き FDICA について述べる．FDICA で得られる N 個の推定信号 $\mathrm{y}_{ij1}, \cdots, \mathrm{y}_{ijN}$ は，音源信号 $\mathrm{s}_{ij1}, \cdots, \mathrm{s}_{ijN}$ の順序と一致していない可能性があり，また順序も周波数ビン i 毎に異なっている状態である．今，$n = 1, \cdots, N$ のあり得るすべての順序を列挙し，これを $r = 1, 2, \cdots, R$ というインデクスで表す．このとき，N 個の異なる順序の総数は $N!$ 個あるため $R = N!$ である．例えば，$N = 2$ であれば $r = 1$ が順序 $(1, 2)$ に対応し，$r = 2$ が順序 $(2, 1)$ に対応する．同様に，$N = 3$ であれば $r = 1, 2, 3, 4, 5, 6$ がそれぞれ順序 $(1, 2, 3)$，$(1, 3, 2)$，$(2, 1, 3)$，$(2, 3, 1)$，$(3, 1, 2)$，及び $(3, 2, 1)$ に対応する．r 番目の音源順序に基づいて $\mathrm{y}_{ij1}, \cdots, \mathrm{y}_{ijN}$ を並び変えた推定信号を $\mathbf{y}_{ij}^{(r)} = [\mathrm{y}_{ij1}^{(r)}, \mathrm{y}_{ij2}^{(r)}, \cdots, \mathrm{y}_{ijN}^{(r)}]^{\mathrm{T}}$ と定義するとき，この順序における音源信号 $\mathbf{s}_{ij}$ との二乗誤差の時間に関する総和を次式で計算できる．

$$E_{ir} = \sum_{j} \|\mathbf{s}_{ij} - \mathbf{y}_{ij}^{(r)}\|_2^2 \tag{2.21}$$

従って，音源信号を用いた理想パーミュテーション解決は，全ての周波数ビン i において，E_{ir} が最も小さくなる順序 r を求めていくことで実現でき，これは次式のように表せる．

$$\hat{r}_i = \arg\min_{r} E_{ir} \tag{2.22}$$

最後に，得られた最適順序 $\hat{r}_i$ に基づいて，推定信号 $\mathrm{y}_{ij1}, \cdots, \mathrm{y}_{ijN}$ の音源順序を入れ替えて上書きする処理をすべての周波数ビンに適用することで，理想パーミュテーション解決付き FDICA の音源分離結果が得られる．

2.4.2　理想音源モデル型 ILRMA

ILRMA は分離行列 $\mathbf{W}_i$ の反復最適化と NMF の低ランク近似による推定信号の音源モデル $\boldsymbol{T}_n\boldsymbol{V}_n$ を反復的に交互最適化するアルゴリズムである．このとき，音源モデルを $\boldsymbol{T}_n\boldsymbol{V}_n$ ではなく，音源信号のパワースペクトログラム $|\mathbf{S_n}|^{\cdot \mathbf{2}}$ に置き換えて固定することで，理想的な音源モデルを考慮した分離行列 $\mathbf{W}_i$ の最適化が可能となる．このような理想音源モデル型 ILRMA の最適化問題は次式となる．

$$\underset{\mathbf{W}_1, \cdots, \mathbf{W}_I}{\text{Minimize}} \; -2J \sum_{i} \log|\det \mathbf{W}_i| + \sum_{i,j,n} \left(\frac{|\mathbf{w}_{in}^{\mathrm{H}} \mathbf{x}_{ij}|^2}{|\mathrm{s}_{ijn}|^2} + \log|\mathrm{s}_{ijn}|^2 \right) \tag{2.23}$$

また，分離行列の反復更新式は次式となる．

$$\mathbf{U}_{in} = \frac{1}{J}\sum_{j}\frac{1}{|\mathrm{s}_{ijn}|^2}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{H}} \tag{2.24}$$

$$\mathbf{w}_{in} \leftarrow (\mathbf{W}_i\mathbf{U}_{in})^{-1}\boldsymbol{e}_n \tag{2.25}$$

$$\mathbf{w}_{in} \leftarrow (\mathbf{w}_{in}^{\mathrm{H}}\mathbf{U}_{in}\mathbf{w}_{in})^{-\frac{1}{2}} \tag{2.26}$$

従って理想音源モデル型 ILRMA は，推定信号間の統計的独立性の最大化基準に加えて，推定信号 $\mathbf{Y}_n$ のパワースペクトログラムが音源信号のパワースペクトログラム $|\mathbf{S}_n|^{\cdot 2}$ に近づくことが考慮された分離行列 $\mathbf{W}_i$ の最適化となる．

2.5 BSS における空間エイリアシング問題

FDICA や ILRMA 等の ICA に基づく BSS は，Fig. 1.4 のように複数のマイクロホンが空間的に密に（狭い間隔で）配置されたマイクロホンアレイの観測信号を想定している．これは，BSS が多チャネル観測信号の周波数毎の位相をそれぞれ適切に回転させて足し合わせることで，他音源を打ち消すという分離原理に基づいていることに起因する．従って，BSS で求まる各音源の分離フィルタ $\mathbf{w}_{in}$ は空間的なディジタルフィルタと解釈でき，FDICA や ILRMA のような時間周波数領域の BSS はビームフォーミング [19] と呼ばれる技術と本質的に等価である [20]．従って，FDICA や ILRMA で推定される周波数毎の分離行列 $\mathbf{W}_i$ は，ビームフォーミングの空間分離フィルタを全音源分まとめたものであり，各音源から各マイクロホンまでの空間伝達系の逆系という物理的性質を持っている．観測信号を得る際のマイクロホンの間隔が離れすぎている場合，各マイクロホンの観測信号間で 1 周期以上の位相差が生じてしまい，正確な位相差が観測できなくなってしまう．この現象は空間エイリアシング問題と呼ばれ，特に高い周波数で顕著に発生する現象である．

位相差の情報の正確な取得に必要なマイクロホン間隔 d の条件は次式となる [21]．

$$d \leq \frac{c}{2f} \tag{2.27}$$

ここで，f 及び c はそれぞれ音響信号に含まれる最高周波数成分及び音速を表す．例えば，$c = 340$ m/s と仮定したとき，マイクロホン間隔 $d = 2$ m の下で空間エイリアシング問題を回避できる周波数成分の最大値は $f = 85$ Hz である．しかしながら，サンプリング周波数として 44,100 Hz が用いられる音響信号の（観測可能な）最高周波数成分は 22,050 Hz である．つまり，ドラムセットの各音源に近接マイクロホンを配置して観測した多チャネル音響信号を BSS の観測信号として扱う場合，ほぼ全ての周波数において空間エイリアシング問題を回避することはできない．以上の理由から，ドラムセットの被り音抑圧問題に対して，近接マイクロホンの観測信号に直接 BSS を適用しても，目的音と被り音の分離は原理的に困難である．

2.6　ドラムセットの被り音抑圧への応用

本論文では，ドラムセットの被り音抑圧問題の解決を目的として，マイクロホンアレイを用いた観測と BSS を用いるアプローチを検討する．すなわち，通常のドラムセットの録音で用いられる近接マイクロホンに加えて，Fig. 1.4 に示すようなマイクロホンアレイを設置して同時に録音する．音楽ライブ演奏や音楽レコーディングの現場において，マイクロホンアレイのような多チャネルの回線を追加することは，機材コストの増大や手間の増加を招くという大きなリスクがある．しかしながら，トリガを用いてドラムセットのソフトウェア音源を駆動するような実際の状況を考慮すると，ドラムセットの被り音の影響は非常に深刻といえる．前述のリスクを加味してもなお，被り音の影響の回避や被り音抑圧を実現する利点は大きいことが予想される．追加するマイクロホンアレイは，ドラムセットの主要な周波数帯域において空間エイリアシングを生じないように設計されたものを用いる．Fig. 1.4 のマイクロホンアレイでは，マイクロホンの間隔がおよそ 0.01 m の直線アレイであり，式 (2.27) を用いると $c = 340$ m/s において空間エイリアシング問題を回避できる周波数成分の最大値は $f = 17000$ Hz であるため，ドラムセットの主要な周波数帯域の位相を正確に観測可能である．

一方で，マイクロホンアレイを用いたドラムセットの演奏の録音データセットは現時点で全く公開されておらず，マイクロホンアレイを用いたドラムセットの音源分離や被り音抑圧の研究を進めるための環境が整備されていないのが現状である．この問題を解決するために，本論文ではまず，マイクロホンアレイを用いたドラムセットの録音実験を実施し，他者が研究目的で無償利用できる実験用データセットとして公開する．また，基礎的な検討として，単純な ILRMA の適用でどの程度の被り音抑圧性能が得られるかを実験的に調査する．さらに，理想的な BSS がもたらす被り音抑圧の性能を確認する目的で，理想パーミュテーション解決付き FDICA 及び理想音源モデル型 ILRMA での被り音抑圧性能についても調査する．

2.7　本章のまとめ

本章では，ドラムセットのマルチトラック録音における被り音問題及び BSS の詳細について述べた．また，ドラムセットのマルチトラック録音で得た観測信号における空間エイリアシング問題と，空間エイリアシング問題を回避しながら被り音抑圧に BSS を用いる動機を述べた．次章では，BSS を用いた被り音抑圧に向けて，近接マイクロホンに加えマイクロホンアレイを用いたドラムセット録音の詳細について述べる．また，録音したドラムセット演奏音のデータセットの詳細について述べる．

第 3 章

マイクロホンアレイを用いたドラムセットの録音

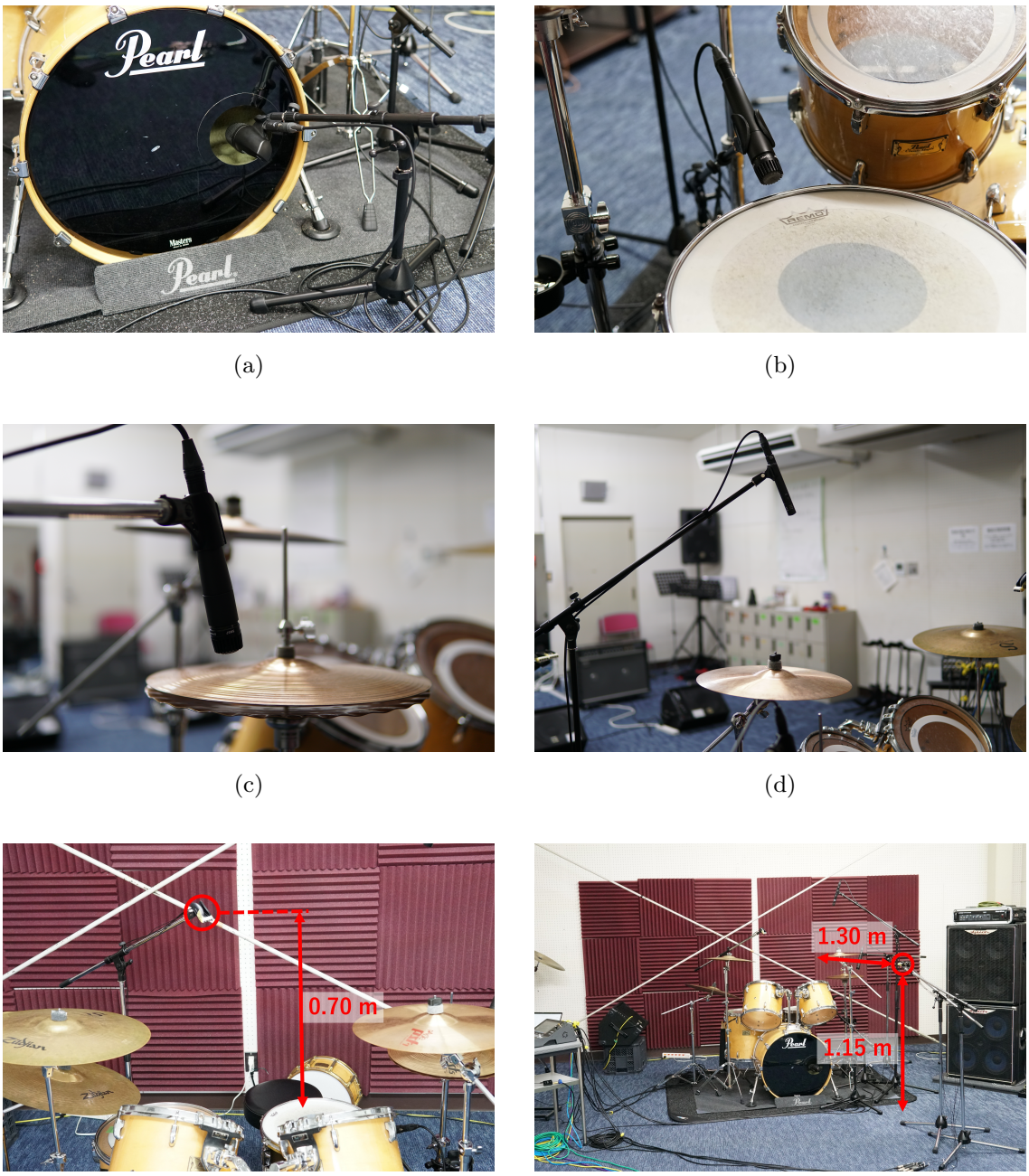
3.1　まえがき

本章では，マイクロホンアレイと BSS を用いたドラムセットの被り音抑圧というアプローチを新たに検討するために，ドラムセットの演奏を近接マイクロホンとマイクロホンアレイの両方を用いて録音したデータセットを作成・公開する．3.2 節では，演奏の録音における録音機器や演奏内容等の種々の条件について述べる．また，3.3 節では実際に作成したデータセットの詳細，ファイル名の命名規則やフォルダ構成について示す．最後に，3.4 節で本章のまとめを述べる．

3.2　録音条件

3.2.1　録音の概要

本項では本研究で作成したデータセットの録音環境，近接マイクロホン及びマイクロホンアレイの空間的な配置，及び録音システムの各種条件を述べる．ドラムセットの録音は，8.5 m×6.7 m 程度の一般的な音楽用リハーサルスタジオ内にて行った．演奏はアマチュアのドラム奏者が，各音源に近接させるマイクロホンの配置はアマチュアのサウンドエンジニア 1 名が，一般的なドラムセットのレコーディング条件を模擬する形で再現した．具体的な画像として，Fig. 3.1 に近接マイクロホン及びマイクロホンアレイの配置を示す．一般的なマルチトラック録音を模擬して，KD，SD，HH，及び CC の 4 音源に対してマイクロホンをそれぞれ近接させて設置した．これらに加えて，ドラムセットの中央上部（演奏者が着座した際の右肩上付近）及び正面左前付近にそれぞれ 4 チャネルのマイクロホンアレイ及び 8 チャネルのマイクロホンアレイを配置した．従って本録音は，近接マイクロホン 4 個，4 チャネルマイクロホンアレイ 1 個，及び 8 チャネルマイクロホンアレイ 1 個の合計 16 チャネルの同期録音している．録音時のサンプリング周波数は 48 kHz，量子化ビット数は 24 bit と設定した．



(a)　(b)

(c)　(d)

(e)　(f)

Fig. 3.1: Microphone setup for (a) KD, (b) SD, (c) HH, (d) CC, (e) four-channel microphone array, and (f) eight-channel microphone array.

3.2.2　録音システム

本データセットの録音に使用した各種機器及び録音時の接続について示す．本録音で使用したドラムセットの各音源の打面やシンバルの型番は Table 3.1 の通りである．これらはい

Table 3.1: Drums equipments used in the dataset recording

equipment	**Model**
KD batter-side head	REMO CS-22B
SD top head	REMO CS-114BA
SD side head	REMO 114SA
HH	PAISTE PST5N Sound Edge Hats 14” (0683114)
CC	PAISTE PST 7 Heavy Crash 16” (1702816)

ずれも一般的なドラムセットのパーツである．また，録音に用いた各マイクロホンの型番を Table 3.2 に示す．マイクロホンの選定についても一般的なドラムセットのレコーディングで広く用いられるマイクロホンの組み合わせとなっている．なお，以降の議論のために，4 チャネルマイクロホンアレイ，8 チャネルマイクロホンアレイ，KD 近接マイクロホン，SD 近接マイクロホン，HH 近接マイクロホン，及び CC 近接マイクロホンの合計 16 チャネルにそれぞれ Table 3.3 のようにチャネル番号を付す．マイクロホンアレイを構成する各マイクロホンのチャネル番号は Fig. 3.2 の通りである．マイクロホンアレイは Fig. 1.4 に示す通り間隔が 9.85 mm の等間隔直線マイクロホンアレイであり，8 チャネルマイクロホンアレイはこの 4 チャネルマイクロホンアレイを横方向に 2 つ並べて構成した．

3.2.3 演奏の条件

本項では，録音したドラムの演奏パターンや演奏における条件を述べる．本データセットで用いたドラムの演奏パターンを楽譜として Fig. 3.3 に示す．この楽譜は一般的なドラム譜面の表記ルールに基づいており，その凡例を Fig. 3.4 に示している．Fig. 3.3 (a) のパターン 1 は KD，SD，及びクローズ HH の 3 音源で構成された一般的な 8 ビートであり，最も単純な演奏パターンである．Fig. 3.3 (b) のパターン 2 は，パターン 1 をベースとして CC やオープン HH も交えたものとなっている．Fig. 3.3 (c) のパターン 3 は比較的遅いテンポで構成されたリズムである．Fig. 3.3 (d) のパターン 4 は，早いテンポの典型的な 16 ビートとなっている．なお，パターン 2，3，及び 4 は KD，SD，HH，及び CC の 4 音源を含む内容となっている．

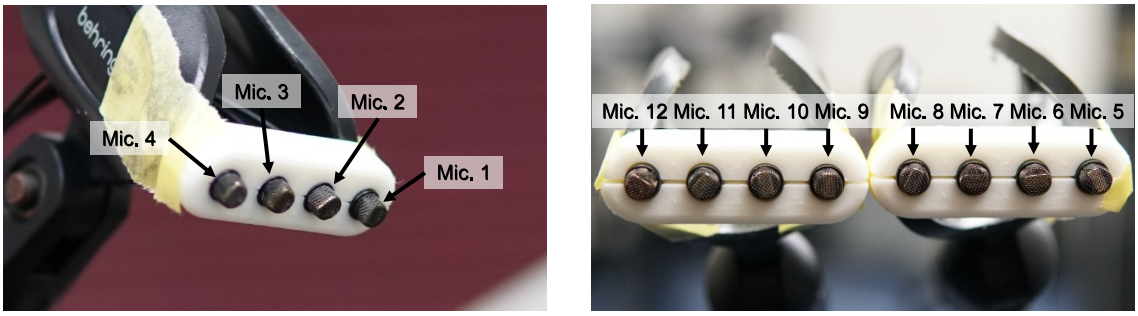
音源分離は通常，性能評価のために客観的な評価尺度を計算・比較する．この客観評価尺度の計算には，音源分離問題における正解となる混合前の各音源の観測信号が必要である．この信号はソースイメージとも呼ばれ，1 つの音源のみを鳴らした際の全マイクロホンの多チャネル観測信号に対応する．従って，音源分離のためのデータセット作成では，ソースイメージを録音し提供しなければならない．ドラムセットの音源分離においては，ソースイメージは「KD のみを演奏した際の全マイクロホンの観測信号」や「SD のみ演奏した際の全マイクロホンの観測信号」等が該当する．このようなドラムセット中の特定の音源を，Fig. 3.3 のようなドラムの演奏パターンに合わせて演奏することは難しいが，本録音では演奏者にテンポを表す

Table 3.2: Model and type of microphones used in the dataset recording

Microphone	Model	Type
Microphone array	JTS CX-500	Condenser microphone
Close microphone for KD	AKG P2	Dynamic microphone
Close microphone for SD	Shure SM57	Dynamic microphone
Close microphone for HH	Shure SM57	Dynamic microphone
Close microphone for CC	AKG P17	Condenser microphone

Table 3.3: Channel indices for each microphone

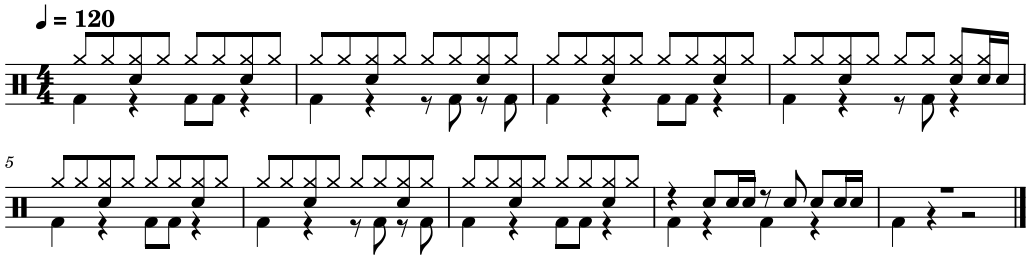
Channel index	Microphone
Mics. 1–4	Four-channel microphone array
Mics. 5–12	Eight-channel microphone array
Mic. 13	KD close microphone
Mic. 14	SD close microphone
Mic. 15	HH close microphone
Mic. 16	CC close microphone



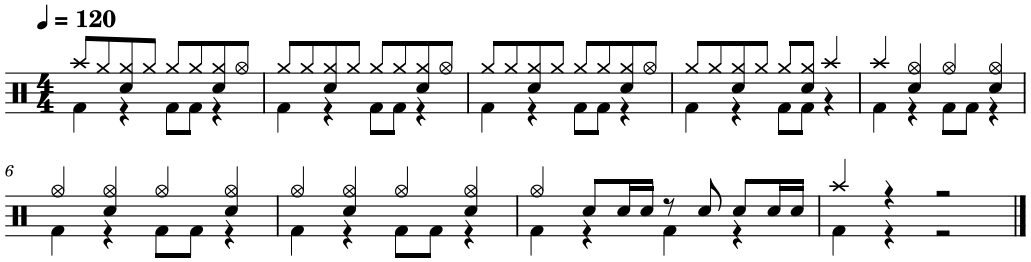
(a)　(b)

Fig. 3.2: Channel indices for each microphone used in (a) four-channel microphone array and (b) eight-channel microphone array.

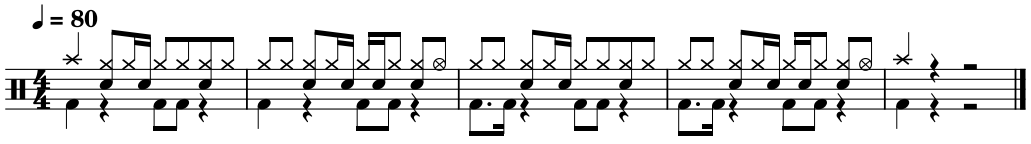
クリック音をインナーイヤー型イヤホンで聞かせながら，可能な限り正確に音源毎の演奏を実施している．録音は Fig. 3.3 の全てのパターンに対して，(a) 全音源を同時に演奏，(b) KD のみ演奏，(c) SD のみ演奏，(e) HH のみ演奏，及び (f) CC のみ演奏の計 5 通りの演奏をそれぞれ 3 回ずつ行った．



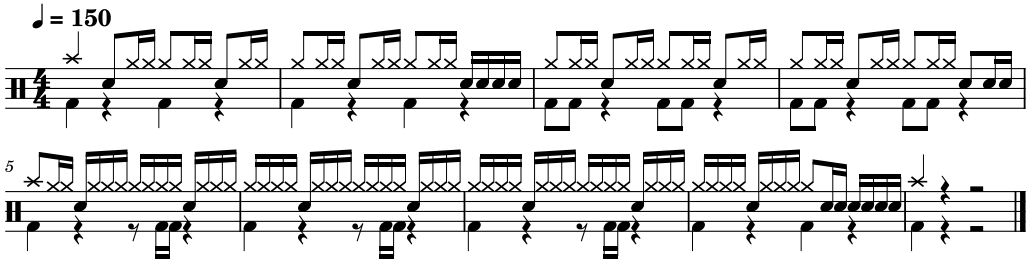
(a)



(b)



(c)



(d)

Fig. 3.3: Drum set scores performed in the dataset recording: (a) Pattern 1, (b) Pattern 2, (c) Pattern 3, and (d) Pattern 4.

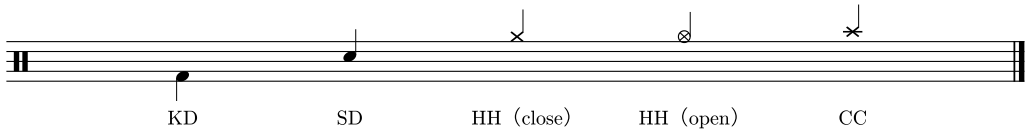


Fig. 3.4: Legend of the sources used in drum set scores.

3.3　データセットの公開

本研究で録音し作成したデータセットは，研究に関する論文・データセット・ソースコード等を無料で公開・保存し，DOI を付与できるオープンリポジトリの Zenodo [22] にて公開した [23]．公開先の QR コードを Fig. 3.5 に示す．データセットのフォルダ構成を Fig. 3.6 に示す．dataset フォルダ内には，Fig. 3.6 に示すような構造で waveform audio file format（.wav 形式）ファイルが格納されている．score フォルダ内には，各演奏パターンの楽譜が portable document format（.pdf）ファイルとして格納されている．README.md は，データセットの詳細を示した Markdown 形式のファイルである．データセットは合計 912 個の WAV ファイルから構成されており，1 つの ZIP ファイルフォーマットに圧縮して提供している．ファイルサイズは 2.8 GB である．WAV ファイルの名前は real_scoreType_takeNum_sourceType_micNum.wav とした．ここで，scoreType，takeNum，sourceType，及び micNum は Table 3.4 に示す命名規則によって適切な文字や数字が当てはまる．

3.4　本章のまとめ

本章では，近接マイクロホンに加えてマイクロホンアレイを用いたドラムセット録音の概要及び録音したドラムセット演奏音のデータセットの詳細を述べた．次章では，録音したデータセットを用いて，愚直に BSS を適用した場合にどの程度の被り音抑圧性能が得られるかを実験的に調査する．また，理想的な BSS における性能についても調査し，本論文で提案するアプローチの有効性や発展可能性について考察する．

Fig. 3.5: QR Code of the recorded dataset in Zenodo.

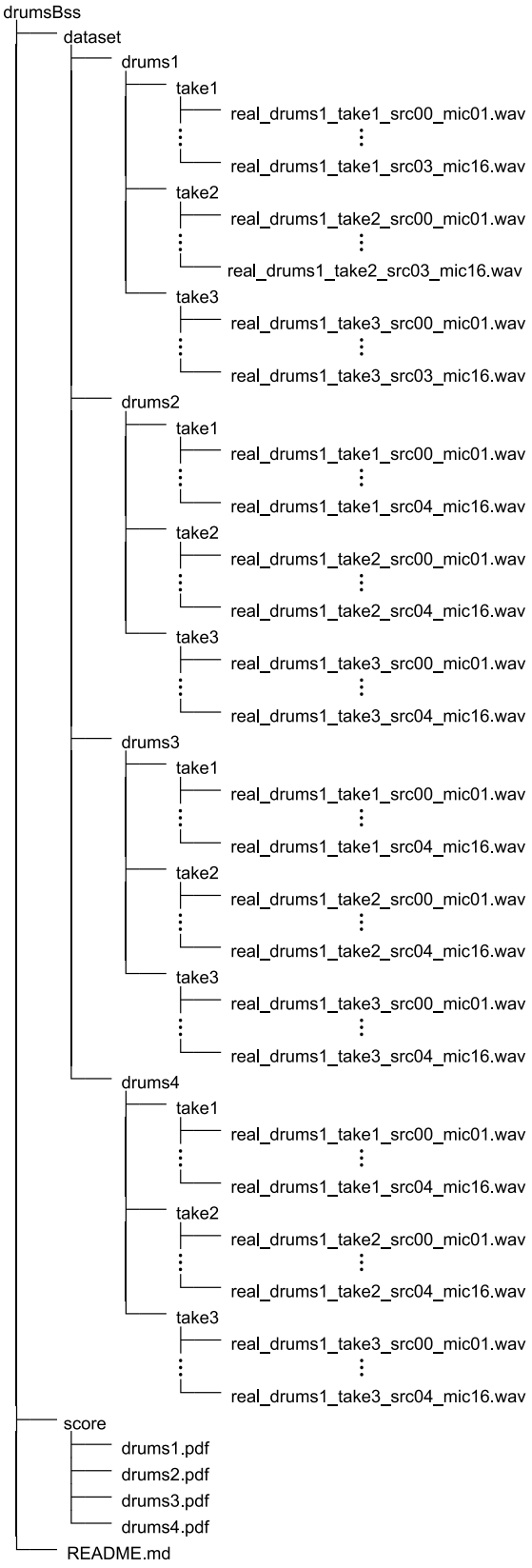


Fig. 3.6: Folder structure of dataset.

Table 3.4: Naming rules for each of recorded audio files

Parameter	**Detail**	**Characters**
`scoreType`	Performance pattern	drums1/drums2/drums3/drums4
`takeNum`	Number of takes	take1/take2/take3
`sourceType`	Active drum sources	src00: All sources src01: KD only src02: SD only src03: HH only src04: CC only
`micNum`	Channel index of microphone	Mics. 01–04: four-channel microphone array Mics. 05–12: eight-channel microphone array Mic. 13: KD close microphone Mic. 14: SD close microphone Mic. 15: HH close microphone Mic. 16: CC close microphone

第 4 章

録音したデータセットでの試験的な BSS

4.1　まえがき

本章では，本論文で対象とするマイクロホンアレイ及び BSS に基づく被り音抑圧が実際にマルチトラック録音されたドラムセットの観測信号に対してどの程度有効かを調査する実験を行う．本実験では，前述のアプローチの基礎的な検討として，3 章で録音・作成したデータセットに対して既存の BSS を適用し，被り音抑圧の客観的な性能を確認する．BSS に対して理想的な条件を与えた場合の性能（すなわち，本アプローチによる被り音抑圧の上限に近い性能）についても確認し，考察する．4.2 節では，本実験で用いる客観評価尺度，観測信号に施す前処理，及び BSS の各種実験条件について述べる．4.3 節では，2 音源，3 音源，及び 4 音源の観測条件における BSS の被り音抑圧の性能をそれぞれ示し，得られた結果について考察する．4.4 節で，本章の実験結果で得られた内容についてまとめる．

4.2　実験条件

本節では，本実験で用いる被り音抑圧性能の客観評価尺度について説明する．また，3 章で作成したドラムセットの観測信号に対して施す前処理をまとめる．さらに，本章で適用する BSS の具体的な実験条件について示す．BSS の性能の評価には信号対歪み比（source-to-distortion ratio: SDR）[24] を使用した．これは，混合音源から目的音源がどれだけ分離できたかを表す分離度合い及び分離音の品質（歪みの少なさ）の両方を加味した音源分離タスクにおける総合的な客観的指標である．SDR の算出方法について説明する．ある音源の混合前の音源 $\tilde{s}(l)$ に対し，これを分離した結果の推定信号を $\tilde{y}(l)$ と表すとき，$\tilde{y}(l)$ は以下の式に示す成分から構成される．

$$\tilde{y}(l) = \tilde{s}_{\mathrm{target}}(l) + \tilde{e}_{\mathrm{interf}}(l) + \tilde{e}_{\mathrm{artif}}(l) \tag{4.1}$$

ここで，$\tilde{s}_{\mathrm{target}}(l)$，$\tilde{e}_{\mathrm{interf}}(l)$，及び $\tilde{e}_{\mathrm{artif}}(l)$ はそれぞれ $\tilde{y}(l)$ 中に含まれる $\tilde{s}(l)$ の成分（目的音源成分），$\tilde{y}(l)$ 中に含まれる $\tilde{s}(l)$ 以外の音源の成分（$\tilde{y}(l)$ に残留した非目的音源成分），及び $\tilde{y}(l)$ 中に含まれるその他の成分（音源分離によって生じた人工的な歪み成分）を表す．SDR は，$\tilde{s}_{\mathrm{target}}(l)$，$\tilde{e}_{\mathrm{interf}}(l)$，及び $\tilde{e}_{\mathrm{artif}}(l)$ を用いて次式で定義される．

$$\mathrm{SDR} = 10\log_{10}\frac{\sum_l |\tilde{s}_{\mathrm{target}}(l)|^2}{\sum_l |\tilde{e}_{\mathrm{interf}}(l) + \tilde{e}_{\mathrm{artif}}(l)|^2}\ [\mathrm{dB}] \tag{4.2}$$

従って，高い SDR 値を達成するには，非目的音源成分 $\tilde{e}_{\mathrm{interf}}(l)$，及び人工的な歪み成分 $\tilde{e}_{\mathrm{artif}}(l)$ が少なく，目的音源成分 $\tilde{s}_{\mathrm{target}}(l)$ が多く抽出されている必要がある．本実験では被り音抑圧の性能評価尺度として，目的音における推定信号の SDR から観測信号の SDR を減算した SDR 改善量を用いる．

本実験では，混合する音源数とマイクロホン数が等しい ($N = M$) 条件にて行う．具体的には，2 音源・2 マイクロホン ($N = M = 2$)，3 音源・3 マイクロホン ($N = M = 3$)，及び 4 音源・4 マイクロホン ($N = M = 4$) の 3 つのケースについて実験を行う．実験に使用するマイクロホンは，同じマイクロホンアレイを構成するマイクロホンから M 個選択し，その他のマイクロホン（マイクロホンアレイを構成する選択されないマイクロホン及び近接マイクロホン）は使用しない．2 音源，3 音源，及び 4 音源で実験を行う際の各マイクロホンの組み合わせ毎のマイクロホンのインデクス $m = 1, \cdots, M$ とデータセットのチャネルインデクス Mics. 1–12 の対応関係を Tables 4.1–4.3 に示す．3.2.3 節で述べたように，音源分離性能の客観的な評価である SDR の算出にはソースイメージが必要である．そのため，本データセットには各音源のソースイメージが含まれている．本実験では，Fig. 4.1 に示すように，各音源のソースイメージを同じチャネルに関して総和することで多チャネル観測信号 $\mathbf{x}_{ij}$ を模擬した．模擬される多チャネル観測信号は以下の式で表される．

$$\mathbf{x}_{ij} = \sum_n \mathbf{s}_{ijn}^{(\mathrm{img})} \tag{4.3}$$

ここで，$\mathbf{s}_{ijn}^{(\mathrm{img})} \in \mathbb{C}^M$ は，n 番目の音源におけるソースイメージの複素スペクトログラムの成分であり，

$$\mathbf{s}_{ijn}^{(\mathrm{img})} = \mathbf{a}_{in} s_{ijn} \tag{4.4}$$

と定義する．また，$\mathbf{a}_{in}$ は混合行列 $\mathbf{A}_i$ の各行に対応するステアリングベクトルである．上記の式は時間周波数領域でソースイメージを足し合わせる場合の計算式であるが，本実験では時間領域でソースイメージを足し合わせて多チャネル観測信号を模擬した．また，SDR の算出において必要な $\tilde{s}_{\mathrm{target}}(l)$，$\tilde{e}_{\mathrm{interf}}(l)$，及び $\tilde{e}_{\mathrm{artif}}(l)$ を計算する際のリファレンス信号は，Tables 4.1–4.3 において $m = 1$ に対応するチャネルのソースイメージを用いた．

本実験では，ILRMA，理想音源モデル型 ILRMA，及び理想パーミュテーション解決付き FDICA を BSS 手法として用いて比較する．この 3 手法の内，ILRMA のみがブラインドな条件であるため，現実的に適用できる手法である．ILRMA の結果は，マイクロホンアレイを

Table 4.1: Correspondence between microphone index m and dataset channel index ($N = M = 2$)

Combination of microphone	$m = 1$	$m = 2$
Mics. 1 & 2	Mic. 1	Mic. 2
Mics. 1 & 3	Mic. 1	Mic. 3
Mics. 1 & 4	Mic. 1	Mic. 4
Mics. 5 & 6	Mic. 5	Mic. 6
Mics. 5 & 7	Mic. 5	Mic. 7
Mics. 5 & 8	Mic. 5	Mic. 8
Mics. 5 & 9	Mic. 5	Mic. 9
Mics. 5 & 10	Mic. 5	Mic. 10
Mics. 5 & 11	Mic. 5	Mic. 11
Mics. 5 & 12	Mic. 5	Mic. 12

Table 4.2: Correspondence between microphone index m and dataset channel index ($N = M = 3$)

Combination of microphone	$m = 1$	$m = 2$	$m = 3$
Mics. 1, 2, & 3	Mic. 1	Mic. 2	Mic. 3
Mics. 5, 6, & 7	Mic. 5	Mic. 6	Mic. 7
Mics. 5, 7, & 9	Mic. 5	Mic. 7	Mic. 9
Mics. 5, 8, & 11	Mic. 5	Mic. 8	Mic. 11

Table 4.3: Correspondence between microphone index m and dataset channel index ($N = M = 4$)

Combination of microphone	$m = 1$	$m = 2$	$m = 3$	$m = 4$
Mics. 1, 2, 3, & 4	Mic. 1	Mic. 2	Mic. 3	Mic. 4
Mics. 5, 6, 7, & 8	Mic. 5	Mic. 6	Mic. 7	Mic. 8
Mics. 5, 7, 9, & 11	Mic. 5	Mic. 7	Mic. 9	Mic. 11

用いたドラムセットの被り音抑圧問題に対して愚直に既存 BSS を適用した場合にどの程度の性能が得られるかを確認することを目的としている．一方で，後者の 2 手法は理想的な条件を与えるために，音源信号（被り音抑圧問題における正解となる信号）を用いている．これは，BSS が最大限性能を発揮した場合にどの程度の性能が得られるかを確認することを目的としている．

Table 4.4 に，これらの手法全てに共通する実験条件を示す．本実験に用いる信号のサンプ

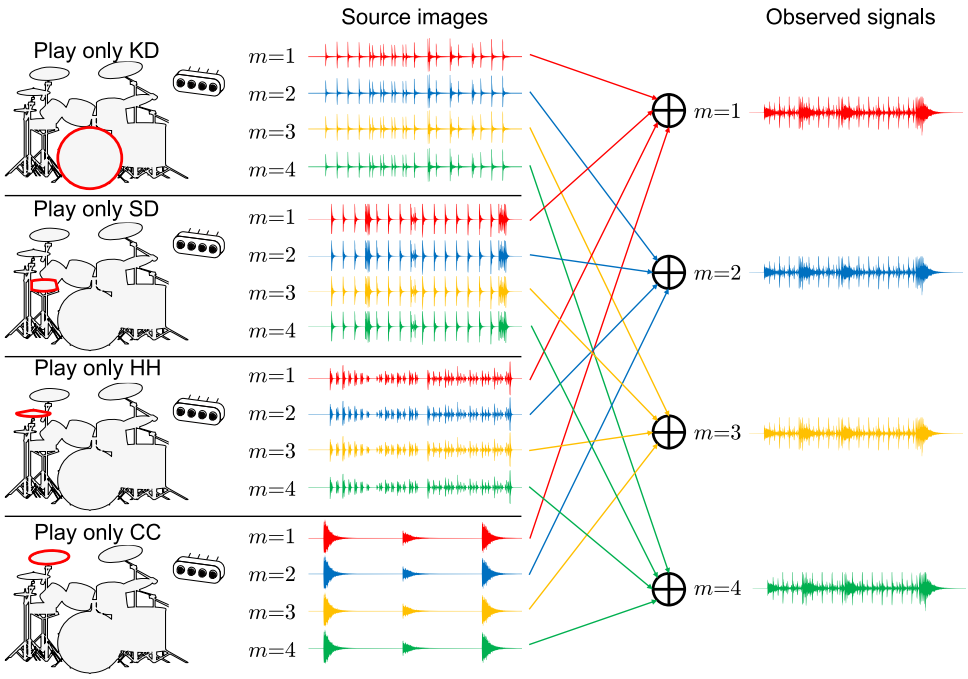


Fig. 4.1: Simulation of multichannel observed signals.

リング周波数は，録音時は 48 kHz であるが，本実験では計算負荷低減のために 44.1 kHz にリサンプリングを行った．STFT の窓長は 16384 点（372 ms），シフト長は 4096 点（93 ms）とした．窓関数には blackman 窓を使用した．BSS の最適化における反復回数は 100 回とした．分離行列 $\mathbf{W}_i$ の初期値は各周波数 i 毎に単位行列とした．また，Table 4.5 に ILRMA 及び理想音源モデル型 ILRMA の実験条件を示す．ILRMA 及び理想音源モデル型 ILRMA における基底数は K=4 とした．ILRMA の基底行列 $\boldsymbol{T}_n$ 及びアクティベーション行列 $\boldsymbol{V}_n$ の初期値はいずれも区間 $(0, 1)$ の一様分布乱数とした．一様分布乱数には 5 種類の擬似乱数シード値を使用し，実験結果を平均した．Table 4.6 に理想パーミュテーション解決付き FDICA の実験条件を示す．理想パーミュテーション解決付き FDICA に置いて仮定する音源モデルを時変分散複素ガウス分布とした．

4.3　実験結果

本節では ILRMA，理想音源モデル型 ILRMA，及び理想パーミュテーション解決付き FDICA による被り音抑圧実験の結果を示す．本実験はマイクロホンアレイを用いたドラムセットの被り音抑圧性能を確認する基礎的な検討であることから，様々なマイクロホンの組み合わせで構成されるマイクロホンアレイでの観測信号を対象として BSS を適用し，どのような配置やマイクロホン間隔に対してどのような性能が得られるかを網羅的に確認する．ただし，本文では代表的なマイクロホンの組み合わせに対する結果を示し，その他の詳細な結果については各手法に対してそれぞれ付録 A に掲載する．本節で抽出して示す結果は，ドラムセッ

Table 4.4: Common experimental conditions

Parameter	Value
Number of sources	2 (KD & SD, KD & HH, or SD & HH) 3 (KD, SD, & HH) 4 (KD, SD, HH, & CC)
Sampling frequency [Hz]	44100
Window size used in STFT	16384 (372 ms)
Window shift size used in STFT	4096 (93 ms)
Window function used in STFT	Blackman window
Number of iterations for parameter update	100
Initial value of demixing matrix	Identity matrix

Table 4.5: Experimental conditions for ILRMA and ILRMA with ideal source model

Parameter	Value
Number of basis vectors	4
Initial value of basis and activation matrices (only for ILRMA)	Uniform random values in the range (0, 1)
Seed for pseudorandom values (only for ILRMA)	1, 2, 3, 4, or 5

Table 4.6: Experimental conditions for FDICA with ideal permutation solver

Parameter	Value
Assumed source model	Complex Gaussian distribution with time-varying variances
Permutation solver	Ideal permutation solver

トの中央上部に配置した 4 チャネルマイクロホンアレイの Mic. 1（Fig. 3.2 (a) 参照）及びドラムセット正面に配置した 8 チャネルマイクロホンアレイの Mic. 5（Fig. 3.2 (b) 参照）を基準となるマイクロホンとし，2 音源・2 マイクロホンの観測信号に対する BSS では基準となるマイクロホンと（同一マイクロホンアレイの中で）マイクロホン間隔が異なるパターンとして Mics. 1 & 2，Mics. 1 & 3，Mics. 1 & 4，Mics. 5 & 6，Mics. 5 & 7，Mics. 5 & 8，Mics. 5 & 9，Mics. 5 & 10，Mics. 5 & 11，及び Mics. 5 & 12 の 10 種類を示す．同様に，3 音源・

3 マイクロホンの観測信号に対する BSS では Mics. 1, 2, & 3，Mics. 5, 6, & 7，Mics. 5, 7, & 9，及び Mics. 5, 8, & 11 の 4 種類，4 音源・4 マイクロホンの観測信号に対する BSS では Mics. 1, 2, 3, & 4，Mics. 5, 6, 7, & 8，及び Mics. 5, 7, 9, & 11 の 3 種類をそれぞれ示す．2.5 節で述べた通り，BSS におけるマイクロホンアレイ中のマイクロホン間隔は空間エイリアシングに直接的に影響するため，これらの観測信号間でどの程度被り音抑圧性能が変化するかについて比較している．

まず，ブラインドな条件で BSS を試みた結果として，ILRMA を観測信号に適用した場合の被り音抑圧性能を Figs. 4.2–4.4 に示す．ここで，Fig. 4.2 は 2 音源・2 マイクロホン，Fig. 4.3 は 3 音源・3 マイクロホン，及び Fig. 4.4 は 4 音源・4 マイクロホンの観測信号に対する結果を示している．Fig. 4.2 (a) の KD 及び SD の混合に対する結果に着目すると，マイクロホンアレイを構成するマイクロホン間隔が長い場合に KD の分離性能が徐々に増加する傾向が確認できる．これはマイクロホンアレイのアレイ長が長いほど低音域に対する位相差の分解能が向上することが原因として推測される．マイクロホン間隔が長くなると空間エイリアシングを回避できる最大周波数が低下するリスクがあるが，KD は低音域に支配的なエネルギーを持つ音源のため，空間エイリアシングの影響は受けなかったと思われる．実際に，Fig. 4.2 (a) において最もマイクロホン間隔が長い例は Mics. 5 & 12 であり，この時のマイクロホン間隔は 0.06895 m 程度となる．式 (2.27) より，このときに空間エイリアシング問題を回避できる周波数成分の最大値はおよそ 2466 Hz 程度となり，これは KD の音源に含まれる主要な周波数成分（500 Hz 以下）を大きく超えていることが分かる．ただし，SD は SDR 改善量が 0 dB を下回っており，その他の音源の組み合わせである Fig. 4.2 (b) 及び (c) に関しては，マイクロホンの組み合わせに応じて性能が一貫性なく変化している．また両音源の SDR 改善量が 0 dB を上回っている例は Fig. 4.2 (c)（SD 及び HH の混合に対する結果）の Mics. 1 & 3 のみであることから，単純な BSS の適用では安定した被り音抑圧を得ることは非常に難しい事実を示している．Figs. 4.3 及び 4.4 は，より困難な 3 音源・3 マイクロホン及び 4 音源・4 マイクロホンの観測信号に対する結果をそれぞれ示している．これらの結果を見ても，やはり全音源の SDR 改善量が 0 dB を上回るような高精度な被り音抑圧結果は得られなかった．以上の結果は，ドラムセットの被り音抑圧に対してマイクロホンアレイ及び BSS を用いるアプローチが非常に困難であることを示唆している．

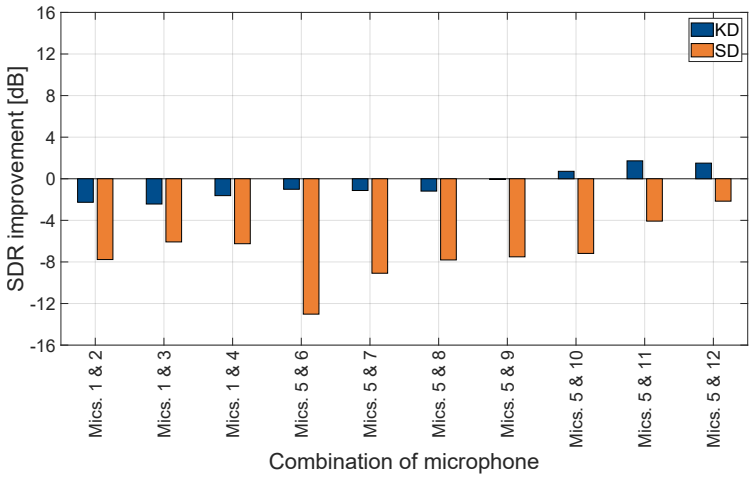
前述の Figs. 4.2–4.4 の結果は BSS による被り音抑圧が困難であることを示している．一方で，これらの低い性能となった結果が，本観測信号に対する被り音抑圧は原理的に困難であることを意味するのか，あるいは単に ILRMA の最適化による分離行列の推定が困難なだけであり，理想的な条件では性能改善が得られるのかについては不明である．そこで以下では，理想的な条件を与える BSS の被り音抑圧性能を確認することで，この観点について実験的に明らかにする．Figs. 4.5–4.7 は理想音源モデル型 ILRMA を用いた場合の 2 音源・2 マイクロホン，3 音源・3 マイクロホン，及び 4 音源・4 マイクロホンの観測信号に対する結果についてそれぞれ示している．同様に，Figs. 4.8–4.10 は理想パーミュテーション解決付き FDICA を用いた場合の各結果をそれぞれ示している．2.4 節で述べた通り，これらの手法はいずれも被

り音抑圧問題の正解となる音源信号そのものを用いていることから，実応用では適用できない手法である．しかしながら，もしこれらの手法が一定の被り音抑圧を達成できる場合は，マイクロホンアレイ及び BSS でドラムセットの被り音抑圧を実現できることが原理的には可能であることを示していることになる．これらの結果をみると，マイクロホンの組み合わせに応じて性能はやや変化しているが，すべての観測信号に対して高精度な被り音抑圧が実現できていることがわかる．本実験において最も難易度が高い 4 音源の混合信号に対しても，すべての音源に対して正の SDR 改善量が得られており，本論文で着目したアプローチが原理的には有効であることを示している．

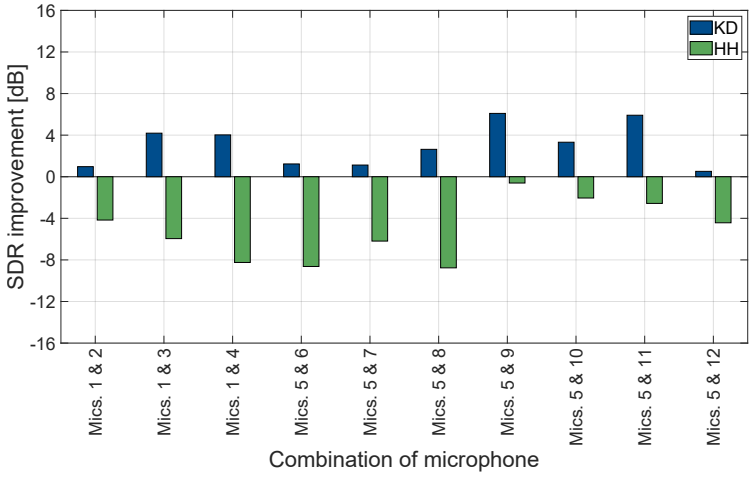
各手法の性能を比較するために，観測信号毎の全音源の SDR 改善量の平均値を Tables 4.4–4.6 に示す．ここで，Table 4.7 は 2 音源・2 マイクロホン，Table 4.8 は 3 音源・3 マイクロホン，Table 4.9 は 4 音源・4 マイクロホンの結果をそれぞれ示している．この結果を見ると，理想音源モデル型 ILRMA 及び理想パーミュテーション解決付き FDICA の分離性能は拮抗しているものの，平均的には理想パーミュテーション解決付き FDICA のほうが分離性能が高いことが分かる．特に，ドラムセットの正面に配置した 8 チャネルマイクロホンアレイから選択したマイクロホンでの実験では，理想パーミュテーション解決付き FDICA の分離性能の優位性が顕著である．

4.4　本章のまとめ

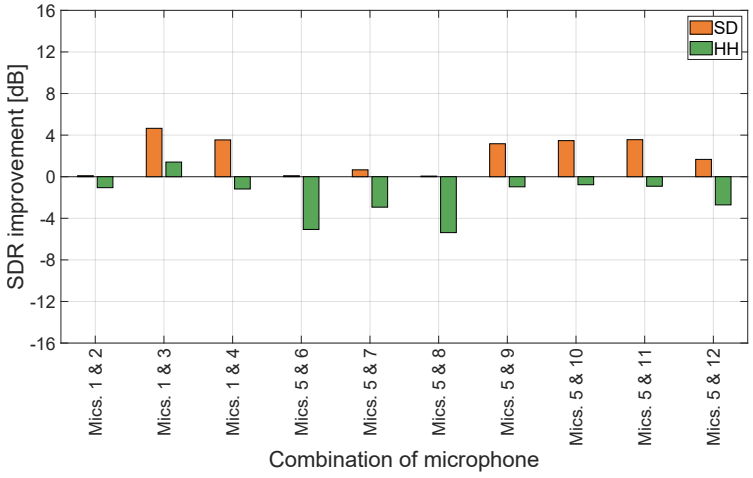
本章では，3 章で録音・作成したデータセットに対して既存の BSS を適用し，マイクロホンアレイを用いたドラムセットの被り音抑圧性能を確認する実験を行った．実験結果から，ブラインドな条件で BSS では被り音抑圧を行うことは困難であることが分かった．一方で，理想的な条件を与えた場合には高精度な被り音抑圧が可能であることが分かり，マイクロホンアレイ及び BSS を用いたドラムセットの被り音抑圧は原理的に有効であることが示された．5 章では，本論文のまとめと今後の課題について述べる．



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. 4.2: SDR improvements of ILRMA for two sources.

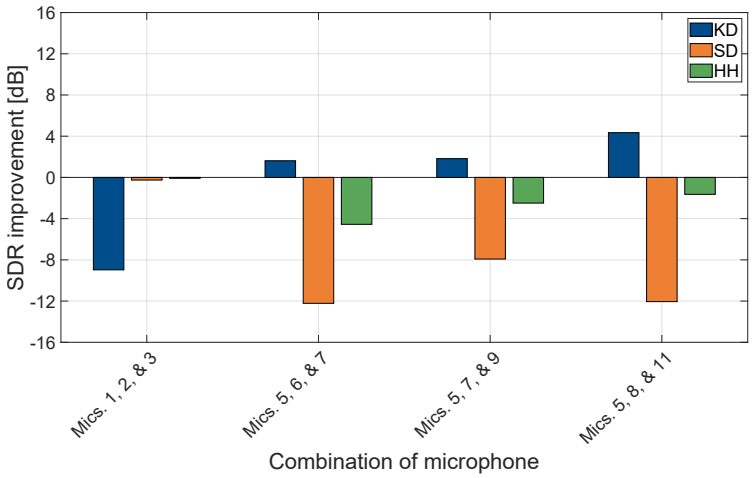


Fig. 4.3: SDR improvements of ILRMA for three sources.

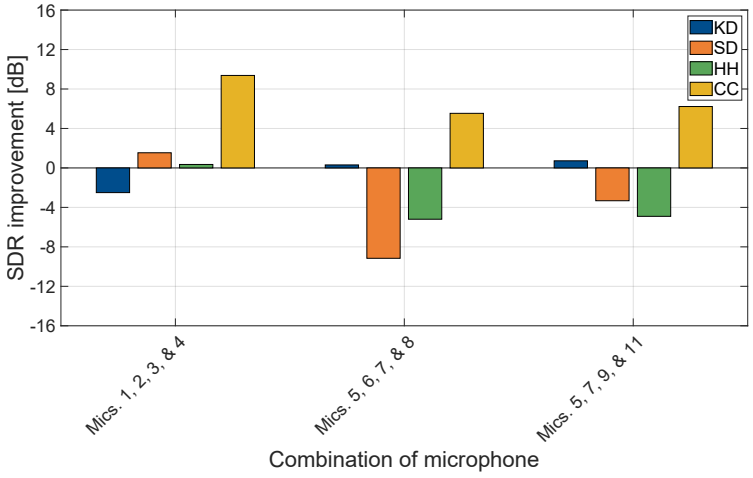
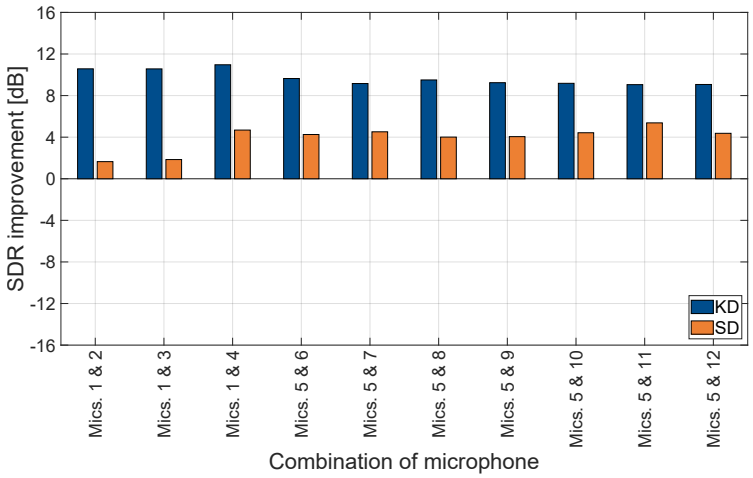
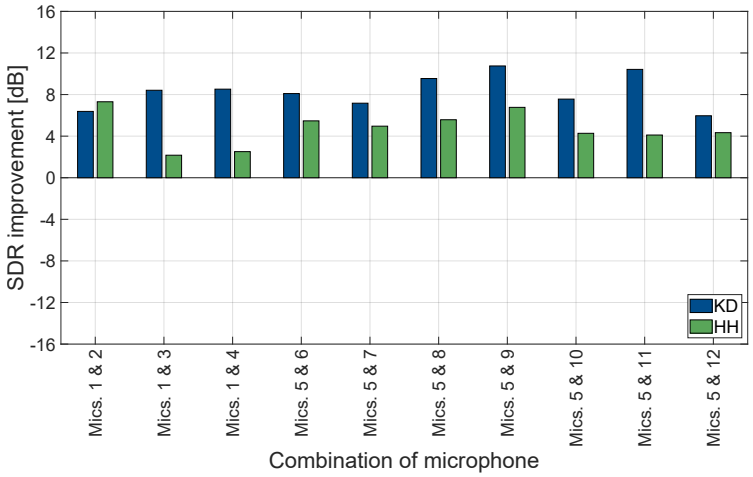


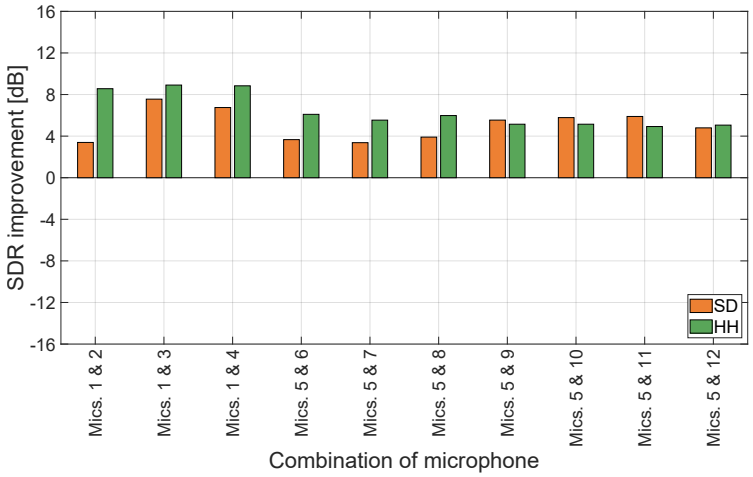
Fig. 4.4: SDR improvements of ILRMA for four sources.



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. 4.5: SDR improvements of ILRMA with ideal source model for two sources.

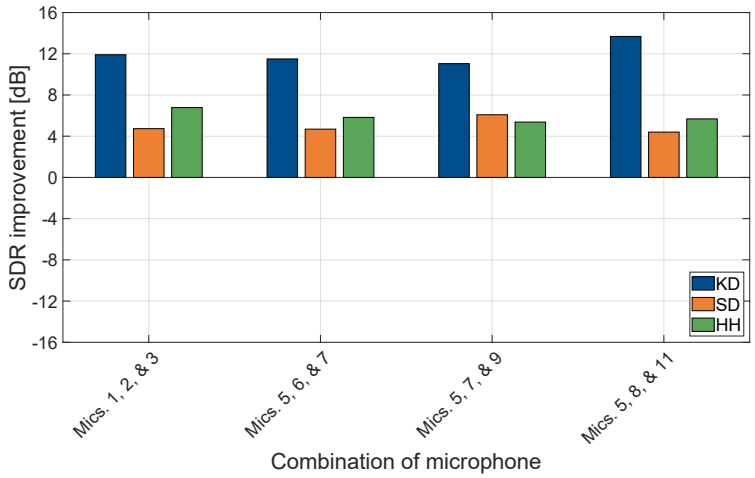


Fig. 4.6: SDR improvements of ILRMA with ideal source model for three sources.

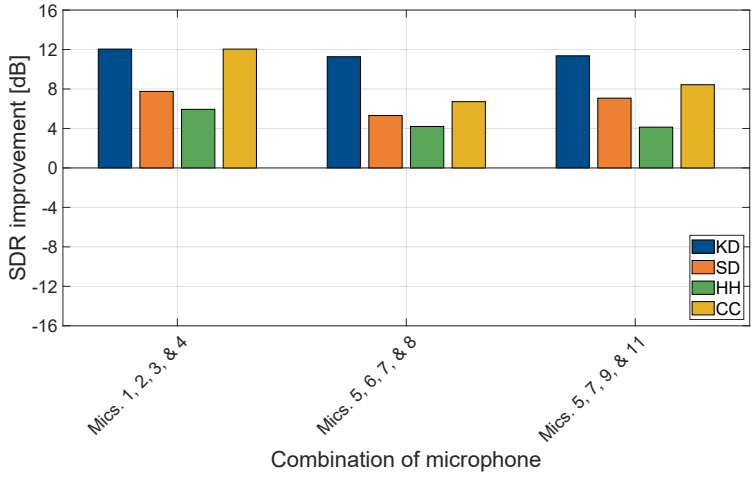
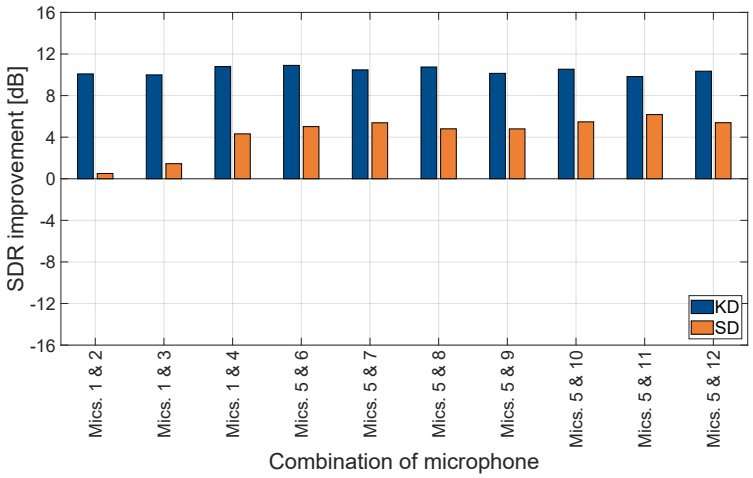
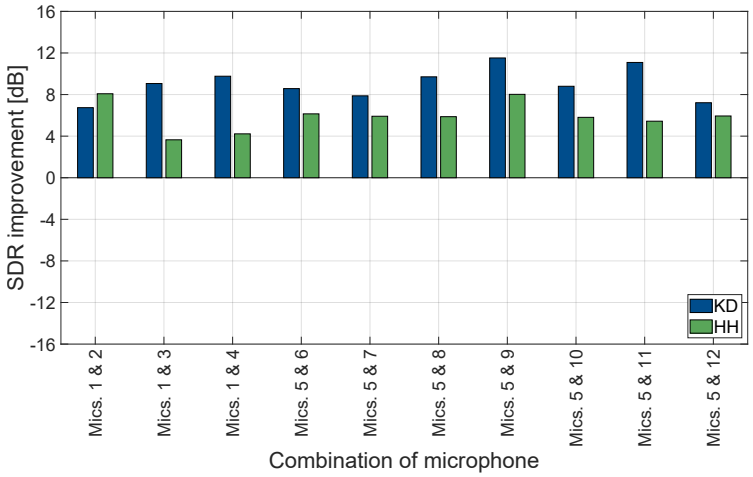


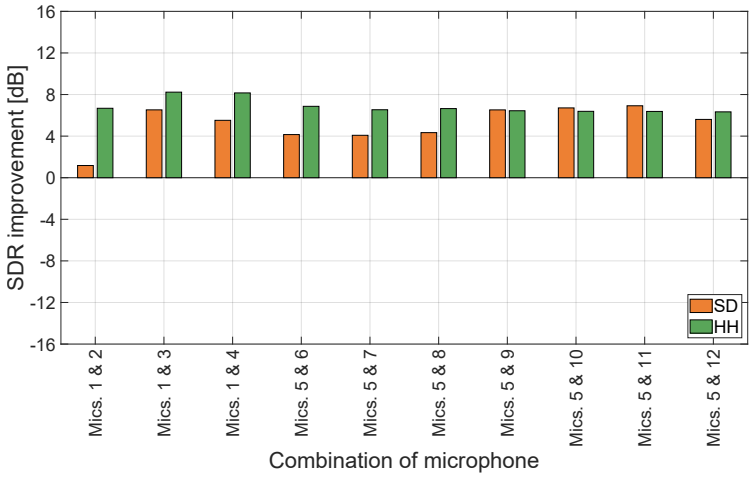
Fig. 4.7: SDR improvements of ILRMA with ideal source model for four sources.



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. 4.8: SDR improvements of FDICA with ideal permutation solver for two sources.

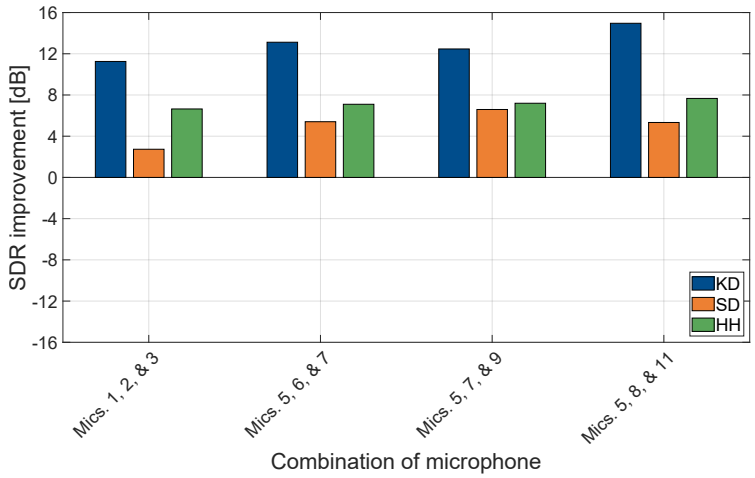


Fig. 4.9: SDR improvements of FDICA with ideal permutation solver for three sources.

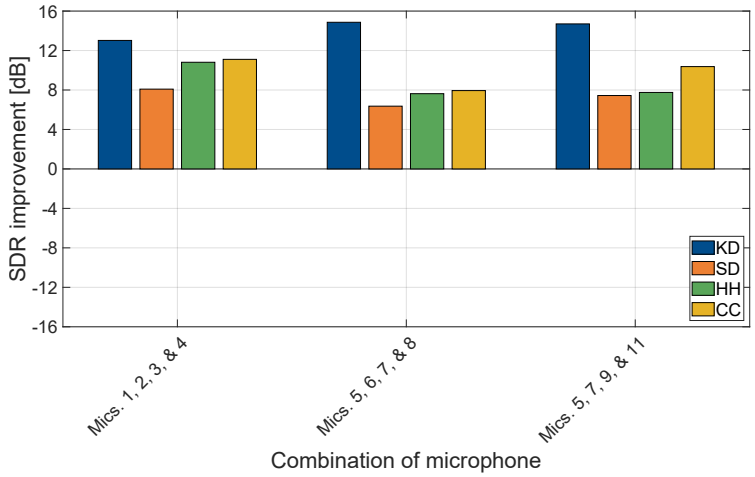


Fig. 4.10: SDR improvements of FDICA with ideal permutation solver for four-sources.

Table 4.7: Comparison of separation performance of various methods for two sources

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1 & 2	-2.36	**6.31**	5.54
Mics. 1 & 3	-0.70	**6.58**	6.48
Mics. 1 & 4	-1.62	7.04	**7.13**
Mics. 5 & 6	-4.40	6.20	**6.94**
Mics. 5 & 7	-2.93	5.79	**6.71**
Mics. 5 & 8	-3.41	6.42	**7.02**
Mics. 5 & 9	0.02	6.92	**7.91**
Mics. 5 & 10	-0.41	6.06	**7.29**
Mics. 5 & 11	0.61	6.63	**7.64**
Mics. 5 & 12	-0.93	5.60	**6.81**

Table 4.8: Comparison of separation performance of various methods for three sources

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1, 2, & 3	-3.10	**7.80**	6.88
Mics. 5, 6, & 7	-5.05	7.34	**8.54**
Mics. 5, 7, & 9	-2.86	7.50	**8.75**
Mics. 5, 8, & 11	-3.12	7.92	**9.32**

Table 4.9: Comparison of separation performance of various methods for four sources

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1, 2, 3, & 4	2.19	9.45	**10.76**
Mics. 5, 6, 7, & 8	-2.13	6.88	**9.20**
Mics. 5, 7, 9, & 11	-0.32	7.75	**10.07**

第 5 章

結言

本論文では，ドラムセットのマルチトラック録音における被り音の混入という問題に対し，従来の近接マイクロホンの設置に加え，マイクロホンアレイを配置するという新たなアプローチを検討した．マイクロホンアレイを配置してドラムセットを録音することで，近接マイクロホンの配置のみでは困難だった，被り音抑圧への BSS の適用を理論的に可能とした．また，今後の技術発展のために，録音したドラムセット演奏音をデータセットとして無償で公開した．実際にマイクロホンアレイで録音したドラムセットの演奏音に BSS を適用した実験では，ブラインドな条件での ILRMA を適用した場合は SDR 改善量が負の値をとる場合が多く，被り音抑圧への有効性が確認できなかった．しかし一方で，理想音源モデル型 ILRMA 及び理想パーミュテーション解決付き FDICA を適用した場合には，SDR 改善量が大幅に向上した．これらの結果から，マイクロホンアレイを用いたドラムセット演奏音に単に BSS を適用するだけでは十分な被り音抑圧は行えないが，理想音源モデル型 ILRMA 及び理想パーミュテーション解決付き FDICA では高精度な音源分離が行え，BSS が被り音抑圧に有効であることが示された．

最後に，今後の課題を述べる．本実験で使用した理想音源モデル型 ILRMA 及び理想パーミュテーション解決付き FDICA は，完全に分離された音源信号（各音源を個別に演奏した信号）を使用することで BSS に理想的な条件を与えた．しかし，実際のドラムセットの録音で各音源を個別に演奏した観測信号を得ることは困難である．そのため，完全分離信号を得ずに理想的な条件を与えた BSS の分離性能に近づける必要がある．これには，本実験では使用しなかった近接マイクロホンでの観測信号を活用する方法が考えられる．近接マイクロホンには目的音が高いエネルギーで観測されている（例えば，KD の近接マイクロホンには KD の演奏音が最も高い音量で観測される）と予想される．この予想より，ILRMA の音源モデルや FDICA におけるパーミュテーション問題の解決に近接マイクロホンの観測信号を活用する方法が有効である可能性があると考えられる．

謝辞

本論文は，香川高等専門学校電気情報工学科北村研究室にて行われた研究に基づくものです．

まず，本研究を進めるにあたり，ご多忙のところ熱心にご指導くださいました指導教員の北村大地准教授に心より感謝申し上げます．北村大地准教授には，毎週のミーティングをはじめ，研究計画の立案から論文の執筆に至るまで，丁寧に指導いただきました．また，研究以外にも，受験勉強に対するアドバイスや学生生活に関する相談等，様々な面でお世話になりました．

本論の副査である雛元洋一助教には，論文の構成や記述に関して大変有益な助言を頂き，大変お世話になりました．ここに厚く御礼申し上げます．

北村研究室の先輩である専攻科生の加藤大輝氏・鈴木慶氏・和気佑弥氏・小川遼氏・谷野宮蒼士氏には，1 年に亘る研究室生活を様々な面で支えていただきました．特に，専攻科 2 年の鈴木慶氏には，研究発表の資料やポスター作成の際にメンターとして多くのアドバイスをいただきました．また，専攻科 1 年の谷野宮蒼士氏には，実験の内容やスクリプト作成において助力をいただきました．北村研究室同期の大喜多景元氏・片山碧人氏とは日常から仲良く交流しながらも，互いに切磋琢磨し，楽しい研究生活を送ることができました．ここに感謝申し上げます．

最後になりますが，現在に至るまで私の学生生活を金銭的・精神的に支え，暖かく見守って下さった両親には感謝の念に堪えません．これまで本当にありがとうございました．

参考文献

[1] O. Gillet and G. Richard, “ENST-Drums: an extensive audio-visual database for drum signals processing,” *in Proc. Int. Conf. Music Inf. Retr.* , pp. 156–159, 2006.

[2] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, “A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF,” *APSIPA Trans. Signal Inf. Process.*, vol. 8, no. e12, pp. 1–14, 2019.

[3] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[4] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.

[5] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation with independent low-rank matrix analysis,” *in Audio Source Separation*, S. Makino, Ed., pp. 125–155. Springer, Cham, 2018.

[6] P. Comon, “Independent component analysis, a new concept?,” *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[7] A. I. Mezza, R. Giampiccolo, A. Bernardini, and A. Sarti, “Toward deep source separation,” *Pattern Recognit. Lett.*, vol. 183, pp. 86–91, 2024.

[8] R. Izhaki, “Drum triggering,” *in Mixing Audio*, R. Izhaki, Ed. London, UK: Taylor & Francis, 2023, ch. 26, pp. 307–314.

[9] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *in Proc. LVA/ICA*, pp. 722–727, 2001.

[10] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.

[11] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, 2004.

[12] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source

separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, 2006.

[13] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS," *in Proc. IEEE Int. Symp. Circuits Syst.*, pp. 3247–3250, 2007.

[14] F. Hasuike, D. Kitamura, and R. Watanabe, "DNN-based frequency-domain permutation solver for multichannel audio source separation," *in Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, pp. 872–877, 2022.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[16] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, 2009.

[17] D. R. Hunter and K. Lange, "A tutorial on MM algorithms." *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[18] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, pp. 189–192, 2011.

[19] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag Berlin Heidelberg, 2001.

[20] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 198923, 2003.

[21] 浅野太, "音のアレイ信号処理: 音源の定位・追跡と分離," コロナ社, 2011.

[22] Zenodo, "Zenodo," https://zenodo.org/, accessed Jan. 30, 2026.

[23] D. Kitamura, "Dataset for drums source separation using microphone arrays," Zenodo, 2025. https://doi.org/10.5281/zenodo.17706651.

[24] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation." *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

発表文献一覧

国内学会

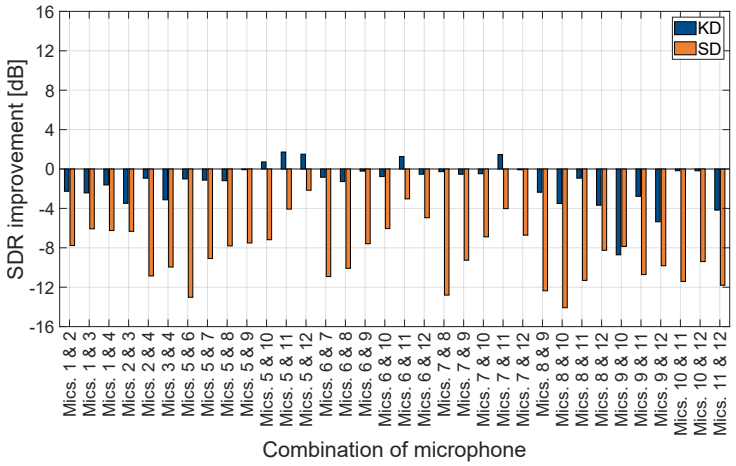
1. 森末結, 北村大地, “マイクロホンアレイを用いたドラムセット音源分離のデータセット収録・公開,” 第 28 回日本音響学会関西支部 若手研究者交流会, p. 18, 京都, 2025 年 12 月（査読無し）.

付録 A

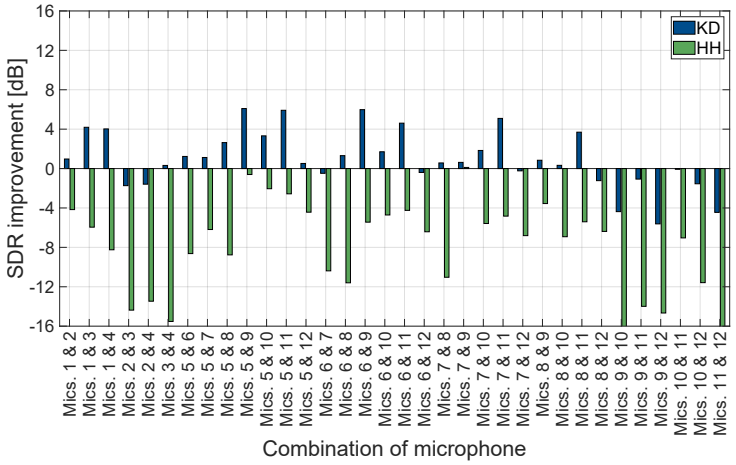
実験結果詳細

A.1　実験結果

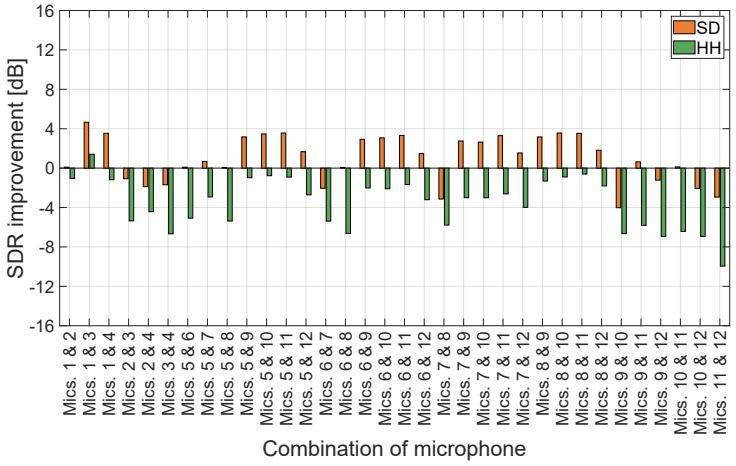
Figs. A.1–A.3 に ILRMA による被り音抑圧実験の結果を示す．Figs. A.4–A.6 に理想音源モデル型 ILRMA による被り音抑圧実験の結果を示す．Figs. A.7–A.9 に理想パーミュテーション解決付き FDICA による被り音抑圧実験の結果を示す．また，Tables A.1–A.3 に各手法の性能を比較するために，観測信号毎の全音源の SDR 改善量の平均値を示す．



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. A.1: SDR improvements of ILRMA for two sources for all observed signals.

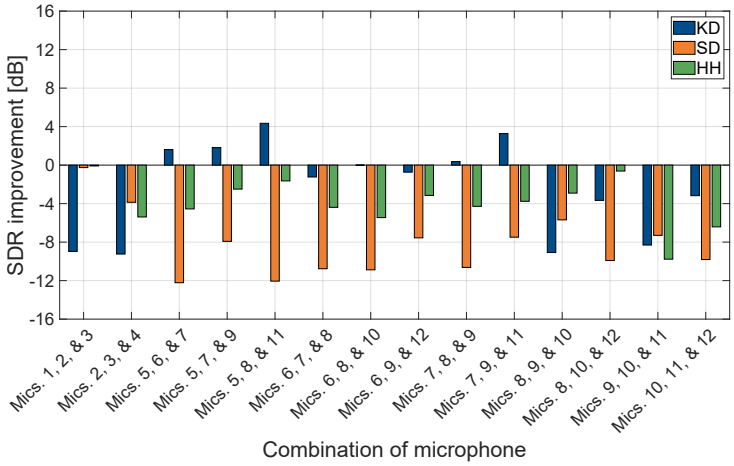


Fig. A.2: SDR improvements of ILRMA for three sources for all observed signals.

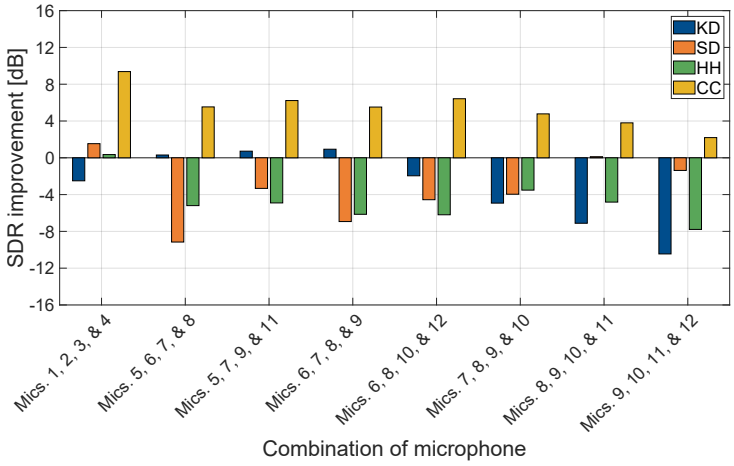
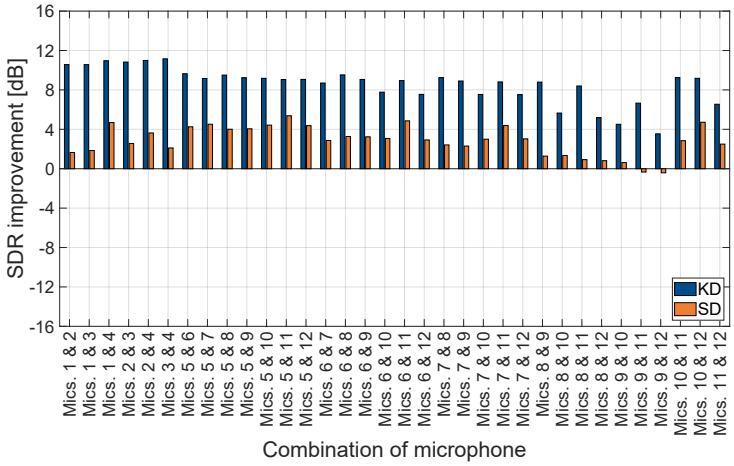
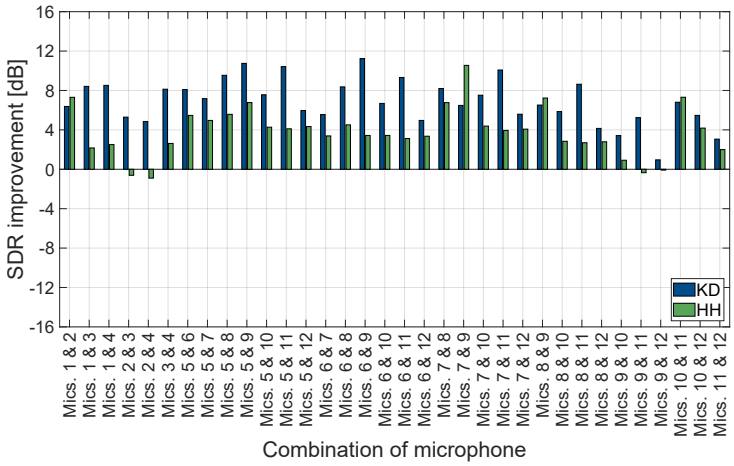


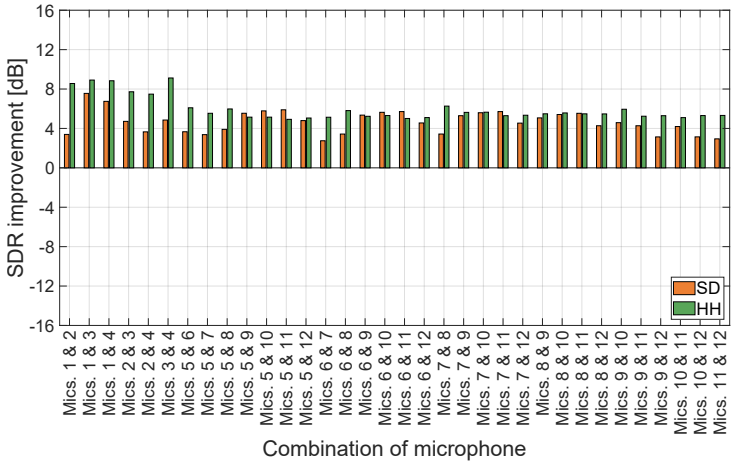
Fig. A.3: SDR improvements of ILRMA for four sources for all observed signals.



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. A.4: SDR improvements of ILRMA with ideal source model for two sources for all observed signals.

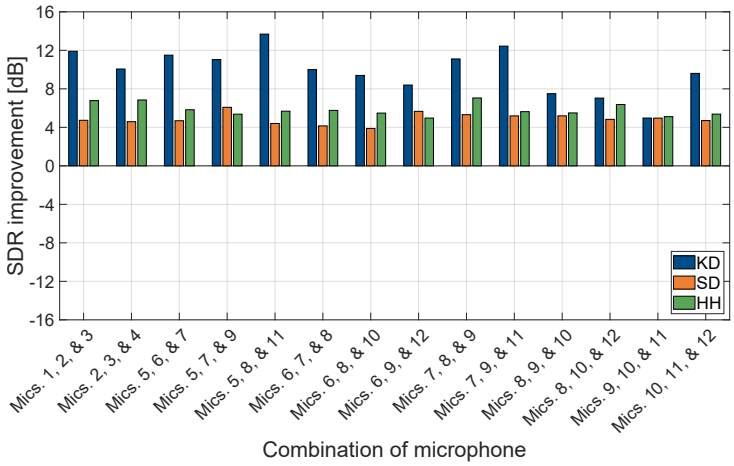


Fig. A.5: SDR improvements of ILRMA with ideal source model for three sources for all observed signals.

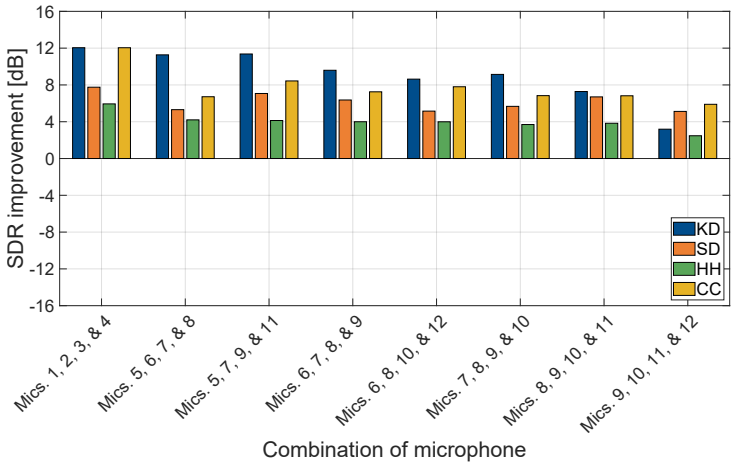
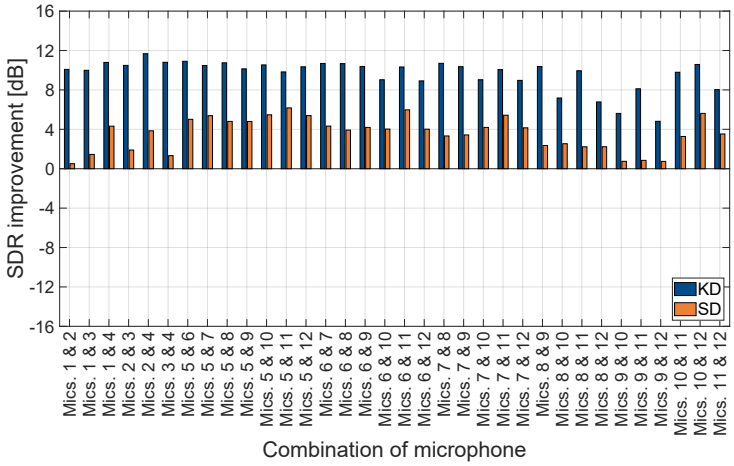
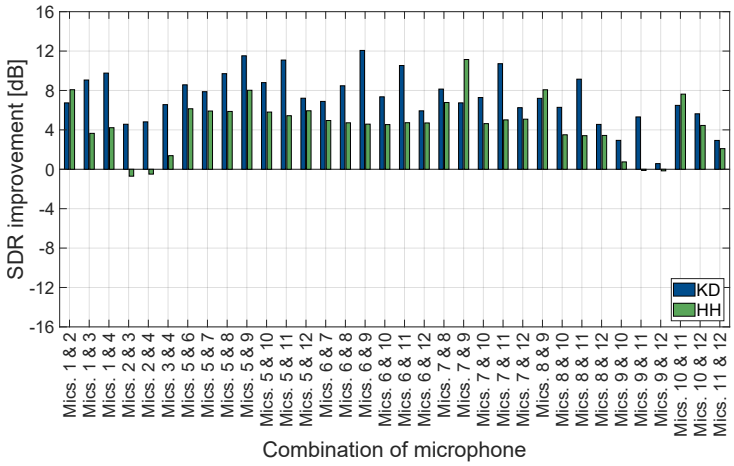


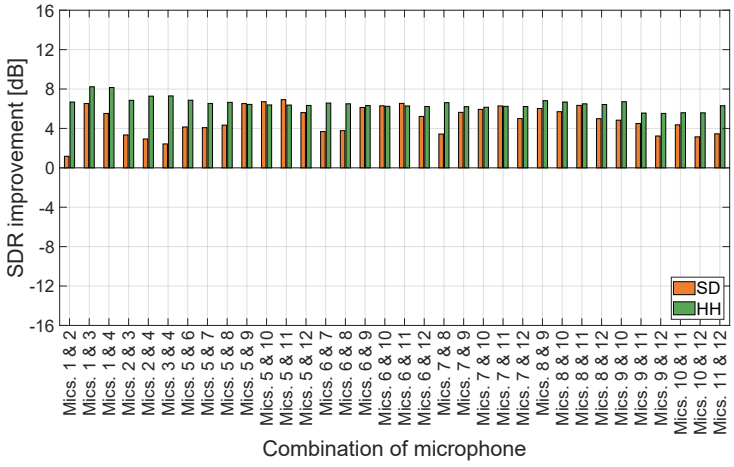
Fig. A.6: SDR improvements of ILRMA with ideal source model for four sources for all observed signals.



(a) KD and SD



(b) KD and HH



(c) SD and HH

Fig. A.7: SDR improvements of FDICA with ideal permutation solver for two sources for all observed signals.

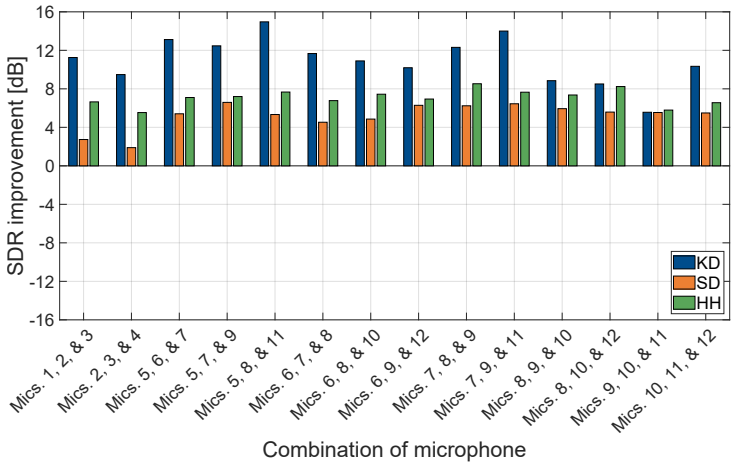


Fig. A.8: SDR improvements of FDICA with ideal permutation solver for three sources for all observed signals.

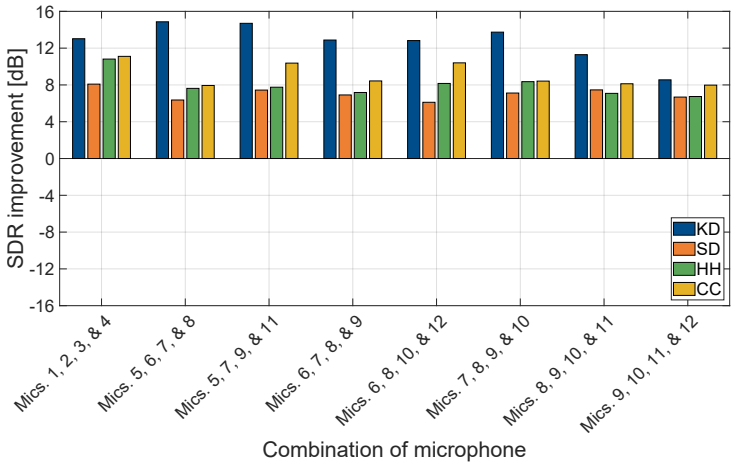


Fig. A.9: SDR improvements of FDICA with ideal permutation solver for four sources for all observed signals.

Table A.1: Comparison of separation performance of various methods for two sources for all observed signals

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1 & 2	-2.36	**6.31**	5.54
Mics. 1 & 3	-0.70	**6.58**	6.48
Mics. 1 & 4	-1.62	7.04	**7.13**
Mics. 2 & 3	-5.39	**5.09**	4.41
Mics. 2 & 4	-5.52	4.95	**5.01**
Mics. 3 & 4	-6.11	**6.33**	4.97
Mics. 5 & 6	-4.40	6.20	**6.94**
Mics. 5 & 7	-2.93	5.79	**6.71**
Mics. 5 & 8	-3.41	6.42	**7.02**
Mics. 5 & 9	0.02	6.92	**7.91**
Mics. 5 & 10	-0.41	6.06	**7.29**
Mics. 5 & 11	0.61	6.63	**7.64**
Mics. 5 & 12	-0.93	5.60	**6.81**
Mics. 6 & 7	-5.01	4.73	**6.19**
Mics. 6 & 8	-4.70	5.82	**6.34**
Mics. 6 & 9	-1.06	6.26	**7.28**
Mics. 6 & 10	-1.48	5.32	**6.25**
Mics. 6 & 11	0.04	6.16	**7.40**
Mics. 6 & 12	-2.34	4.74	**5.84**
Mics. 7 & 8	-5.41	6.05	**6.50**
Mics. 7 & 9	-1.55	6.53	**7.25**
Mics. 7 & 10	-1.91	5.61	**6.20**
Mics. 7 & 11	-0.27	6.37	**7.29**
Mics. 7 & 12	-2.71	5.02	**5.95**
Mics. 8 & 9	-2.60	5.73	**6.81**
Mics. 8 & 10	-3.58	4.45	**5.32**
Mics. 8 & 11	-1.84	5.28	**6.26**
Mics. 8 & 12	-3.26	3.78	**4.74**
Mics. 9 & 10	-8.82	3.33	**3.60**
Mics. 9 & 11	-5.62	3.45	**4.04**
Mics. 9 & 12	-7.27	2.07	**2.46**
Mics. 10 & 11	-4.17	5.92	**6.19**
Mics. 10 & 12	-5.28	5.33	**5.84**
Mics. 11 & 12	-8.93	3.73	**4.39**

Table A.2: Comparison of separation performance of various methods for three sources for all observed signals

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1, 2, & 3	-3.10	**7.80**	6.88
Mics. 2, 3, & 4	-6.16	**7.16**	5.64
Mics. 5, 6, & 7	-5.05	7.34	**8.54**
Mics. 5, 7, & 9	-2.86	7.50	**8.75**
Mics. 5, 8, & 11	-3.12	7.92	**9.32**
Mics. 6, 7, & 8	-5.46	6.64	**7.66**
Mics. 6, 8, & 10	-5.44	6.25	**7.73**
Mics. 6, 9, & 12	-3.81	6.34	**7.81**
Mics. 7, 8, & 9	-4.85	7.82	**9.03**
Mics. 7, 9, & 11	-2.66	7.75	**9.37**
Mics. 8, 9, & 10	-5.89	6.06	**7.38**
Mics. 8, 10, & 12	-4.73	6.08	**7.44**
Mics. 9, 10, & 11	-8.46	5.01	**5.63**
Mics. 10, 11, & 12	-6.47	6.56	**7.46**

Table A.3: Comparison of separation performance of various methods for four sources for all observed signals

Combination of microphones	ILRMA	ILRMA with ideal source model	FDICA with ideal permutation solver
Mics. 1, 2, 3, & 4	2.19	9.45	**10.76**
Mics. 5, 6, 7, & 8	-2.13	6.88	**9.20**
Mics. 5, 7, 9, & 11	-0.32	7.75	**10.07**
Mics. 6, 7, 8, & 9	-1.66	6.80	**8.85**
Mics. 6, 8, 10, & 12	-1.57	6.40	**9.38**
Mics. 7, 8, 9, & 10	-1.91	6.34	**9.41**
Mics. 8, 9, 10, & 11	-2.00	6.16	**8.49**
Mics. 9, 10, 11, & 12	-4.35	4.17	**7.49**