筑波大学大学院博士課程

システム情報工学研究科修士論文

分散型アレー補聴器システムへのブラインド音源分離手法の適用と雑音教師あり手法への拡張に関する研究

宇根 昌和

修士（工学）

（コンピュータサイエンス専攻）

指導教員　牧野 昭二

2021年3月

概要

本論文では，補聴器の両耳に装備されたマイクロホンに，スマートフォンのマイクロホンを加えた分散マイクロホンアレー補聴器システムを新たに提案する．さらに，本論文では既存のブラインド音源抽出 (blind speech extraction: BSE) 手法の雑音情報を用いた半教師あり手法への拡張を行う．BSE とは，音声と雑音との混合信号から，音源及び混合系に関する事前情報を用いずに音声のみを抽出する技術である．これまで多くのブラインド音源分離手法が BSE に適用されており，中でも独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）は各方位に一つの点音源があると仮定し，混合信号を各方位に効率的に分離することを可能にしている．しかし，補聴器利用シーンを含めた実環境では，全方位に雑音源が存在する拡散性雑音下である状況が考えられる．拡散性雑音下では，目的音源方位に雑音源も存在するため，ILRMA では目的音源と背後の雑音は原理的に分離できない．こうした問題を解決するために，ランク制約付き空間共分散行列推定法が提案された．ランク制約付き空間共分散行列推定法は各音源の空間伝達特性を表現する空間相関行列を推定するが，ILRMA で推定された高精度な空間パラメータを用いることでより少ない計算コストで音源抽出を行う手法である．

一般的に，これらの手法は複数のマイクロホンを利用することで多くの空間情報を得ることができ，音源抽出の精度向上に繋がる．しかし，補聴器などのデバイスに多くのマイクロホンを装備することは規模やコストの点で現実的ではない．一方で，近年スマートフォンをはじめとする小型のマイクロホンを搭載した携帯端末が広く普及している．本論文では，スマートフォンに搭載されているマイクロホンを含めた補聴器システムを新たに提案する．スマートフォンに内蔵されたマイクロホンを用いることで，マイクロホンの総数も増え，さらに目的音源に近い位置の空間情報を利用できる．補聴器を装着するユーザの頭や耳の形は人によってそれぞれ異なり，またスマートフォンの位置も特定できない．BSE はこれらの不確定な要素の多い状況に対しても柔軟に処理を行うことができるが，提案する補聴器システムに対する有効性は不明である．そこで，スマートフォンを持った人を模したシミュレータを作成し，データ収録を行い，収録したデータに対する既存手法の有効性を調査する．

一方で，補聴器は基本的に常に周囲の音を収録しており，会話シーンなどでは目的の音，すなわち会話相手の声が発せられる直前に，雑音のみの情報を得られる．得られる雑音情報を利用して音源抽出を行うことでさらなる品質の向上が期待できる．本論文では，ブラインドの枠組みであるランク制約付き空間共分散行列推定法を雑音情報を用いた半教師ありの枠組みへ拡張した新たな手法を提案し，従来の手法と比較して有効であることを示す．

目 次

図目次

第1章　序論

1.1　研究背景

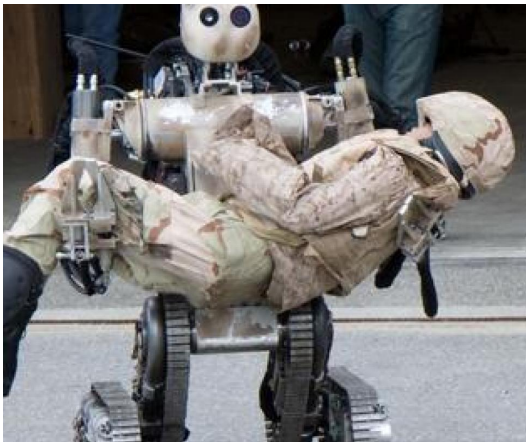
音声は人間にとって最も自然で利用しやすいコミュニケーション手段の一つである．近年では，音声対話ロボットやテレビ会議システム，補聴器など，音声通信に関するシステムが増加しており，音声による情報伝達が多く利用されている．しかし，周囲の雑音の影響で音声の品質が劣化し，音声を用いたアプリケーションの円滑な利用を大きく阻害する．音声を用いたアプリケーションの円滑な利用のためには，雑音下で目的の音を抽出する技術が必要である．このような問題を解決するため，複数の音源からの信号が混合された観測信号から，元の音源信号を推定する音源分離という技術が広く利用されている．

図 1.1 に，音源分離の代表的な応用例を挙げた．一つ目は補聴器システムへの応用である．補聴器システムは聴覚能力が低いユーザの聴覚を補助するための装置である．閑静な場所から人通りが多い街中や駅などまで，様々な音環境の中での使用を想定し，ユーザの聴き取りたい音を自然な形で提示するシステムであることが必要である．音源分離機構をシステムの内部に組み込むことで，例えばユーザが会話するときに背景に存在する様々な雑音を抑圧し会話相手の音声のみを抽出することが可能になる [1]．二つ目はスマートスピーカ，スマートフォン，及びヒューマンオリエンテッドなシステムなどにおける音声認識システムへの応用である．人間との対話型インタフェースに音声を用いるものは音声信号を自然言語へ変換する音声認識を行う必要があるが，雑音が存在する環境では認識精度が悪化してしまう．音源分離により所望の信号を取り出し，分離された信号に対して音声認識を行うことで，様々な状況において音声認識の精度を向上させることができる [2]．三つ目は会議の認識・理解である．会議で議論された内容を録音し，情報としてまとめることにより後から把握することが容易になる．しかし大規模な会議では，録音された音声を書き起こすのにも人手を要する上，しばしば起こり得る発言のオーバーラップは聞き間違いを誘発する．録音音声を一人ひとりの発話に分離し，それぞれに対して音声認識を行うことで適切に議事録の作成が可能になる [3–5]．近年では，図 1.1 (c) のように，デバイスが複数利用されている状況で，それらに内蔵されているマイクロホンを利用して音源の定位や分離を行う分散マイクロホンアレー処理に関する

(a) Hearing-aid device

(b) Smart speaker

(c) Meeting speech recognition and understanding systems



(d) Rescue robot

図 1.1: Applications of speech source separation.

研究も盛んに行われている [6–11]．四つ目は災害用のロボットへの応用である．人間の侵入が困難な場所で，生存者の声を雑音環境下から検知するために音源分離が必要となる．生存者の声を検知する上で，最もクリティカルな問題はロボット内部のモーターなどの駆動音である．ロボット自身が発する雑音（エゴノイズ）が生存者の声の検知を大きく妨げるため，生存者の声とエゴノイズを分離するための試みが行われている [12–14]．

音源分離技術は，マイクロホンの数が 1 つ（単チャネル）か複数（多チャネル）か，及び学習を事前情報無しで行う（ブラインド）か事前情報有りで行うかという 2 つの観点から分

類できる．多チャネルの場合は信号の音響的特徴に加えて空間的な情報を利用することが出来るのに対し，単チャネルの場合は音響的特徴のみしか用いることが出来ないため，その性能は限定的である．ブラインドでない古典的な音源分離手法の代表例として，Wiener フィルタ [15] やビームフォーマ [16–18] がある．Wiener フィルタは音源分離に限らず，広範な種類の時系列データをフィルタリングする手法である．最小平均二乗誤差規範により，目的音源及びその他の雑音源のパワースペクトログラムを用い，目的音源の複素スペクトログラムを推定する．Wiener フィルタが単チャネルの手法である一方で，ビームフォーマは多チャネルの音源分離手法である．マイクロホンアレーの素子同士の位置関係や所望の音源の到来方位などの情報を用い，目的音源を高精度に推定することが可能である．これらの手法の適用には目的音源や雑音源の音響的・空間的情報を要するため，それらの情報が十分な精度で得られない場合には推定精度が劣化してしまう可能性がある．

ブラインド音源分離 (blind source separation: BSS) [19] は，上記の手法ように事前情報を必要とせずに音源の分離を達成することができる．単チャネル及び多チャネルの観測信号から，音響的特徴や空間的特徴をブラインドに推定するため，様々な音響シーンにおいて用いることが可能な技術である．単チャネルの場合のブラインド音源分離手法として，楽器音などの音源のパワースペクトログラムが持つ特徴をモデル化することで分離を行う非負値行列因子分解 (nonnegative matrix factorization: NMF) [20] が提案されている．一方で，多チャネルの場合はさらに空間的特徴に関する手がかりを利用することができるため，より高精度な分離を達成できる．多チャネルのブラインド音源分離は，時間領域における瞬時混合信号を分離する独立成分分析 (independent component analysis: ICA) [21–23] 及び残響が存在する場合の畳み込み混合信号への ICA の適用を可能にした周波数領域独立成分分析 (frequency-domain ICA: FDICA) [24–26] に端を発する．FDICA は各音源が点音源であり空間的混合が線形時不変システムで表されるという仮定に基づき，混合系の逆系である分離系を推定することにより音源分離を行う．さらに FDICA を改良した手法として，周波数間の音源パーミュテーション問題を解決した独立ベクトル分析 (independent vector analysis: IVA) [27–29] や音源のパワースペクトログラムを NMF により表現する独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [30,31] などが提案され，より高精度な音源分離を達成することが可能になった．これらの手法は分離音の歪みを抑えた分離が可能であるが，特に背景雑音として全方位から到来する拡散性の雑音が存在する場合には理論的に拡散性雑音を完全に除去することは不可能であり，分離された目的音には雑音成分が残留してしまう [32]．また，複数音源の混合音から 1 つの独立な成分を抽出する independent vector extraction (IVE) [33] も提案されているが，背景雑音の相互相関は考慮するもののモデル自体は FDICA 等で仮定されている点音源

仮定に基づいているため，背景雑音が拡散性を有する場合の分離性能は限定的である．

上記の問題を解決するため，ビームフォーマや ICA に由来する系統の多チャネル音源分離手法を前段で実行し，その出力音に対し Wiener フィルタやスペクトル減算 [34] など単チャネルのポストフィルタを適用することでさらに目的音と残留雑音の分離を行う手法が数多く提案されている [35–39]．上記手法は残留雑音成分を抑圧するために，目的音とは別の方位に存在する雑音成分の音響的特徴などを用いて後段のポストフィルタを構成する．しかしこれらの推定手法は厳密には統計的枠組みに基づいておらず，その推定精度は限定的であるため，最終的な分離音には大きな歪みが発生してしまう [40]．音声認識システムの前段処理として用いる場合にはこの分離された音声信号の歪みはそれほど大きな悪影響をもたらさないが，人間が受聴した場合には不快感を与えてしまうため，適切な分離が行えているとは言いがたい．多チャネル空間線形フィルタと単チャネルポストフィルタを組み合わせた枠組み全体を統計的にモデル化することで，多チャネル観測信号をより適切に表し，少ない歪みで目的音声を抽出するブラインド音声抽出 (blind speech extraction: BSE) が達成できると考えられるが，そのような手法は今まで提案されてこなかった．

一方で，ICA 系統以外の多チャネルのブラインド音源分離手法として，各音源の空間特徴を表す空間相関行列 (spatial covariance matrix: SCM) [41] を用いるフルランク空間相関行列モデルが提案されている．ICA 系の手法は音源を分離するための線形時不変分離フィルタを推定する一方で，フルランク空間相関行列モデルは各音源の空間的な混合特性を推定する．さらに高精度な推定を達成するために各音源のパワースペクトログラムを NMF を用いて表す多チャネル非負値行列因子分解 (multichannel NMF: MNMF) [42, 43] が提案されており，その計算速度を高速化するため，SCM の同時対角化可能性を仮定する高速多チャネル非負値行列因子分解 (FastMNMF) [44, 45] も提案されている．しかし，これらのフルランクな SCM を推定する手法は分離フィルタを推定する ICA 系統の手法と比べて非常に計算コストが大きい上，初期値に頑健でない，分離音が歪んでしまうなどの欠点を抱えており，実用上課題が残る [30]．

また，近年隆盛を見せている深層学習を用いた音源分離手法も多く提案されている [46–50]．その多くは音響特徴を事前に学習するものであり，教師データとなる特定のクラスの音信号を用意する必要がある．例えば人の音声や特定の楽器音などである．音響特徴は事前学習が可能であるが，空間特徴はその汎化性が極めて低い．すなわち，マイクロホンアレーや部屋の形状に依存する空間特徴はその形状の僅かな変化により大きく変わってしまうため，学習した空間特徴と分離時の空間特徴とのミスマッチが容易に起こり得る．そのため多くの多チャネルの教師有り音源分離手法は，音響特徴は教師有りで学習する一方で，空間特徴はブラインドで推定を行う．特に，補聴器の利用シーンでは想定される空間特徴は多岐にわたるため，

深層学習に基づく分離では十分な効果が望めない．

補聴器などの処理後の信号を人が知覚するアプリケーションなどでは，計算コストが大きいことや，空間特徴の違いによる性能劣化は致命的である．このような問題を解決するため，目的音声と背景に存在する拡散性雑音の空間特性を適切にモデル化し，少ない計算コストかつ高い雑音抑圧性能で目的音声を抽出する手法として，ランク制約付き空間共分散行列推定法が提案されている [51,52]．拡散性雑音中に点音源である音声が存在する場合に ICA 系統の線形時不変分離手法を用いると，目的音声の方位は正確に推定できることが先行研究で指摘されている [53]．ランク制約付き空間共分散行列推定法は，まず前段に ICA 系統の中でも最も高い性能を誇る ILRMA を用い，そこで得られた目的音声と雑音の一部の空間特性から，目的音声方向に存在する雑音成分を推定する．

また，深層学習に基づく音源分離が目的音のデータで学習するのに対し，目的音以外の情報から学習を行い分離を行う半教師ありアプローチの音源分離手法も提案されている [14,54,55]．中でも，基底共有型 ILRMA (basis-shared ILRMA :BS-ILRMA) [14] と呼ばれる手法は，高品質な BSS である ILRMA を半教師ありアプローチへ拡張した手法である．BS-ILRMA は，索状の災害用ロボットのために元々提案された手法であり，ロボット自身が発するエゴノイズから生存者の声を分離する．目的音声の教師信号は得られないが，エゴノイズは前もって収録できるため，半教師ありの枠組みとして利用できる．また，BS-ILRMA は深層学習のように様々なパターンかつ大量のデータは必要なく，前もって収録した雑音信号のみを用いる．さらに，誤差逆伝搬法も必要ないため，計算コストを増加させることなく．ブラインドでの音源分離に比べて高い分離性能を達成することが可能となる．

1.2 本論文の目的

本論文では，補聴器システムに焦点を当てる．図 1.1 (a) に示したように，音源分離の補聴器システムへの応用がなされる一方で，図 1.1 (c) のようにマイクロホンを複数内蔵したデバイスがある状況下を対象にした分散マイクロホンアレー処理に基づく音源分離も広く研究されている [56–65]．分散マイクロホンアレー処理は，その場に存在するマイクロホンを利用してアレー処理を行えるため，得られる音源情報も多くなり，さらに，広い範囲の空間情報も得ることができる．近年はスマートフォンが広く普及しており，会議シーンに限らず様々な場面で分散マイクロホンアレー処理が行える．こうした背景から，本論文では，両耳のマイクロホンだけでなくスマートフォンに内蔵されているマイクロホンも含めた分散マイクロホンアレー補聴器システムを新たに提案する．以下，提案補聴器システムと呼ぶ．補聴器を装着

するユーザの頭や耳の形は人によってそれぞれ異なり，またスマートフォンの位置も特定できない．BSE はこれらの不確定な要素の多い状況に対しても柔軟に処理を行うことができるため，提案補聴器システムには BSE を実装する．ただし，提案補聴器システムに既存の BSE 手法が有効に動作するかは不明である．本論文では，提案補聴器システムに実装する BSE 手法として，実環境にも有用で計算コストの少ないにランク制約付き共分散行列推定法を用いる．そのため，提案補聴器システムを用いてデータ収録を行い，収録したデータに対してランク制約付き共分散行列推定法の音声抽出性能を評価し，ILRMA に比べ有効な手法であることを示す．さらに，提案補聴器システムにおいても，マイク数の増加及び目的音源に近い位置の空間情報が利用できるという 2 点が音源抽出に優位に寄与しているかを ILRMA による分離を行なって調査する．

補聴器を用いて会話するシーンでは，会話が始まる直前まで雑音のみが存在する．この状況を利用して，会話直前の雑音を収録することができる．事前に雑音のサンプルが利用できるため，提案補聴器システムにも BS-ILRMA のような半教師ありアプローチを組み込むことが可能になる．ただし，BS-ILRMA は元来災害用ロボットのために提案された手法である．生存者の声とエゴノイズの分離タスクと比較して，提案する補聴器タスクは雑音の種類が多いなどいくつか分離に不利な要素が考えられる．こうした不利な条件下でも，BS-ILRMA が有効に動作することを確認し，提案補聴器システムに対しても半教師ありアプローチが利用できることを示す．一方で，1.1 節で述べたようにランク制約付き空間共分散行列推定法は，前段に ILRMA を用いた初期化を行い，一部のパラメータを推定している．提案補聴器システムで収録したデータに対する BS-ILRMA の有効性を示したのち，BS-ILRMA をランク制約付き空間共分散行列推定法の初期化方法に用いて，より高品質な音源抽出が達成できることを示す．

ランク制約付き空間共分散行列推定法は，ブラインドで雑音の SCM を推定している．上記に述べたとおり，補聴器利用シーンでは事前に雑音のサンプルが利用できるため，ランク制約付き空間共分散行列推定法の雑音 SCM の推定に対しても雑音のサンプルを用いることが可能になる．従って，本論文では最後に，ブラインドのランク制約付き空間共分散行列推定法を，雑音を用いて半教師ありアプローチへ拡張した手法を提案する．本論文では半教師ありランク制約付き空間共分散行列推定法とする．半教師ありランク制約付き空間共分散行列推定法の提案補聴器システムのデータに対する有効性を示す．

1.3 本論文の構成

本論文の構成は以下の通りである．第 2 章では，本論文で取り扱う既存の音源分離手法について述べる．具体的には，ブラインドの手法として，ILRMA 及びランク制約付き空間共分散行列推定法について，半教師あり音源分離手法として BS-ILRMA について述べる．第 3 章では，新たな補聴器システムとして，両耳のマイクロホンだけでなくスマートフォンのマイクロホンを含めた分散マイクロホンアレー補聴器システムを提案する．第 4 章では，提案補聴器システムに対して，ランク制約付き空間共分散行列推定法が適用可能であるか評価実験により評価する．さらに．提案補聴器システムによりもたらされた，マイク総数の増加及び目的音源に近い位置の空間情報が利用可能という 2 点が優位に働いているかを評価する．第 5 章では半教師あり音源分離手法である BS-ILRMA の提案補聴器システムに対する有効性を示す．その後，BS-ILRMA をランク制約付き空間共分散行列推定法の初期化に用い，さらに高品質な分離を達成することを示す．第 6 章では，元来ブラインドの枠組みであったランク制約付き空間共分散行列推定法を半教師ありアプローチへ拡張した手法を提案する．第 7 章では，半教師ありアプローチへ拡張したランク制約付き制約付き空間共分散行列推定法の，提案補聴器システムのデータに対する有効性を示す．最後に，第 8 章で，本論文の結論を述べる．

第2章　既存手法

2.1　はじめに

本章では，本研究で取り扱う音源分離手法について述べる．まず，2.2 節で基本的な BSS の定式化を行う．次に，ブラインドの枠組みの手法として，2.3 節にて state-of-the-art な手法である ILRMA [30, 31] 及び，2.4 節にて実環境のように拡散性雑音が存在する状況で有効なランク制約付き空間共分散行列推定法 [51, 52] について述べる．最後に，2.5 節では，半教師ありの枠組みの手法として，ILRMA を半教師ありアプローチへ拡張した BS-ILRMA [14] について述べる．

2.2　定式化

N 個の音源信号を M 個のマイクロホンで収録し，観測した信号を分離することを考える．複素時間周波数成分における音源信号 $\boldsymbol{s}_{ij}$，観測信号 $\boldsymbol{x}_{ij}$，及び分離信号 $\boldsymbol{y}_{ij}$ をそれぞれ次のように定義する．

$$\boldsymbol{s}_{ij} = (s_{ij,1}, \ldots, s_{ij,N})^\top \in \mathbb{C}^N \tag{2.1}$$

$$\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,M})^\top \in \mathbb{C}^M \tag{2.2}$$

$$\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,M})^\top \in \mathbb{C}^N \tag{2.3}$$

ここで, $i = 1, \ldots, I$, $j = 1, \ldots, J$, 及び $n = 1, \ldots, N$ はそれぞれ周波数ビン, 時間フレーム, 及び音源信号のインデクスである．$\cdot^\top$ は転置記号を表す．マイク数が音源数以上 $(M \geq N)$ かつ各音源が方向性の点音源であり，短時間フーリエ変換 (short-time Fourier transform: STFT) の窓長が残響時間より十分短い場合，各周波数ビンにおいて混合行列 $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N}) \in \mathbb{C}^{M \times N}$ が存在し，次のように書ける．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{2.4}$$

ただし，$\boldsymbol{a}_{i,n}$ は周波数 i における音源 n のステアリングベクトルである．N 個の音源のソースイメージ（音源から音が発されて空間を伝搬する際の信号）を $\boldsymbol{c}_{ij} = (c_{ij,1}, \ldots, c_{ij,N})^\top \in \mathbb{C}^M$ とすると，式 (2.4) は以下のように書き換えられる．

$$\boldsymbol{x}_{ij} = \sum_n \boldsymbol{c}_{ij,n} \tag{2.5}$$

$$\boldsymbol{c}_{ij,n} = \boldsymbol{a}_{i,n} \boldsymbol{s}_{ij} \tag{2.6}$$

すなわち，$\boldsymbol{A}_i$ の各列ベクトルは各音源からマイクロホンへの伝達特性を表すベクトルである．$M > N$ である場合は主成分分析などの線形次元削減を情報を失わずに施すことができるため，以下では $M = N$ と仮定する．$M = N$ かつ $\boldsymbol{A}_i$ が正則である場合，$\boldsymbol{A}_i$ の逆行列 $\boldsymbol{W}_i \in \mathbb{C}^{N \times M}$ を推定することで，次のように分離信号が得られる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{2.7}$$

2.3 ILRMA

ILRMA では，各時間周波数フレームにおける音源 n の成分が

$$s_{ij,n} \sim \mathcal{N}_c\left(0, r_{ij,n}\right) \tag{2.8}$$

なる単変量複素ガウス分布に従い生起する確率生成モデルを仮定する．$r_{ij,n} > 0$ は時変な分散であり，音源のパワースペクトログラムに対応する．さらに，$r_{ij,n}$ は NMF を用いてモデル化される．

$$r_{ij,n} = \sum_{l=1}^{L} t_{il,n} v_{lj,n} \tag{2.9}$$

ここで，$t_{il,n} \geq 0$，$v_{lj,n} \geq 0$ は NMF 変数であり，$l = 1, \ldots, L$ は NMF 基底のインデクス，L は NMF の基底数である．この時 $\boldsymbol{s}_{ij}$ は多変量複素ガウス分布に従い，式 (2.4)，式 (2.8) 及び式 (2.9) と多変量複素ガウス分布の再生性より，$\boldsymbol{x}_{ij}$ も多変量複素ガウス分布

$$\boldsymbol{x}_{ij} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_n r_{ij,n} \boldsymbol{a}_{i,n} \boldsymbol{a}_{i,n}^{\mathsf{H}}\right) \tag{2.10}$$

に従う．ここで，$r_{ij,n}$ は音源 n の音源モデルに相当し，非負実数である NMF 変数の $t_{il,n}$ と $v_{lj,n}$ を用いて音源パワーのスペクトログラムを低ランク近似したものである．また，$\boldsymbol{a}_{i,n}$ はステアリングベクトル，即ち音源 n における空間基底から構成されるランク 1 空間相関行列であり，音源 n の空間モデルに相当する．$\cdot^{\mathsf{H}}$ はエルミート転置を表す．NMF 変数 $t_{il,n}$，$v_{lj,n}$ 及び分離行列 $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1} = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,N})^{\mathsf{H}}$ は次の負対数尤度関数を反復法に基づき最小化することで推定される．

$$\mathcal{L}(\Theta) = \sum_{i,j,n} \left(\frac{|y_{ij,n}|^2}{\sum_l t_{il,n} v_{lj,n}} + \log \sum_l t_{il,n} v_{lj,n} \right) - 2J \sum_i \log |\det \boldsymbol{W}_i| + \text{const.} \tag{2.11}$$

ここで，$\Theta = \{\boldsymbol{W}_i, t_{il,n}, v_{lj,n}\}$ は目的変数の集合であり，const. は目的変数に依存しない項である．分離フィルタ $\boldsymbol{W}_i$ に関しては，反復射影法 [29, 66] という手法に基づき次のように更新される．

$$\boldsymbol{G}_{i.n} = \frac{1}{J} \sum_j \frac{1}{r_{ij,n}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} \tag{2.12}$$

$$\boldsymbol{w}_{i,n} \leftarrow (\boldsymbol{W}_i \boldsymbol{G}_{i,n})^{-1} \boldsymbol{e}_n \tag{2.13}$$

$$\boldsymbol{w}_{i,n} \leftarrow \boldsymbol{w}_{i,n} (\boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{G}_{i,n} \boldsymbol{w}_{i,n})^{-\frac{1}{2}} \tag{2.14}$$

ここで，$\boldsymbol{e}_n \in \mathbb{R}^N$ は N 次の単位行列 $\boldsymbol{E}_N$ の n 番目の列ベクトルである．NMF 変数に関しては，$|y_{ij,n}|^2$ と $\sum_l t_{il,n} v_{lj,n}$ の間の板倉斎藤ダイバージェンス [67] の最小化に基づき，次の更新式を得る．

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_j \frac{|y_{ij,n}|^2}{\left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^2} v_{lj,n}}{\sum_j \frac{1}{\sum_{l'} t_{il',n} v_{l'j,n}} v_{lj,n}}} \tag{2.15}$$

$$v_{lj,n} \leftarrow v_{lj,n} \sqrt{\frac{\sum_i \frac{|y_{ij,n}|^2}{\left(\sum_{l'} t_{il',n} v_{l'j,n}\right)^2} t_{il,n}}{\sum_i \frac{1}{\sum_{l'} t_{il',n} v_{l'j,n}} t_{il,n}}} \tag{2.16}$$

式 (2.12) – 式 (2.16) に基づく更新において，反復による負対数尤度関数の単調非増加性が成り立つため，収束の保証された最適化を行うことが出来る．

式 (2.6) を用いれば，各音源のソースイメージの分散共分散行列は次のように表せる．

$$\mathbb{E}[\boldsymbol{c}_{ij,n}\boldsymbol{c}_{ij,n}^{\mathsf{H}}] = \mathbb{E}[|s_{ij,n}|^2\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}}] \tag{2.17}$$

$$= r_{ij,n}\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}} \tag{2.18}$$

従って，音源 n の空間共分散行列は $\boldsymbol{R}_{i,n} = \boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}}$ なるランク 1 の行列として表現される．これのような理由から，IVA や ILRMA のモデルはランク 1 空間モデルと言われる．

2.4 ランク制約付き空間共分散行列推定法

2.4.1 動機

ランク制約付き空間共分散モデル推定法は 1 個の方向性目的音源と拡散性雑音が混合している状況を対象とした手法である．ILRMA など空間共分散行列をランク 1 空間モデルとして扱う手法では，各音源が点音音源として仮定している．そのため，拡散性雑音の各音源を点音源として見なすことができないため，原理的に目的音源を抽出することができない [32]．これに対し，拡散性雑音をフルランクの空間共分散行列でモデリングする MNMF や FastNMF を用いることが妥当であると考えられる．しかし．フルランクの空間相関行列を推定することはランク 1 の空間相関行列を推定することと比べ，より大きな計算コストを必要とする．さらに，パラメータ数が多いため性能は ILRMA よりも初期化に関して頑健でない [31]．また，各パラメータを ILRMA による推定値で初期化する手法も提案されているが，そのモデルの複雑さから，推定精度の向上は限定的である．方針として，ILRMA によって得られた各音源の空間基底，及びそれによって構成される M 個のランク 1 空間相関行列を用い，拡散性雑音のフルランク空間相関行列を推定する．これは，拡散性雑音中の目的音源抽出タスクにおいて，ILRMA は目的音源の推定に対する精度は望ましくない一方，拡散性雑音は非常に高い精度で推定できることに由来する [51]．また，この現象は，ILRMA の分離音間の独立性最大化の結果，雑音を推定する分離フィルタが点音源である目的音源を正確に打ち消すヌルビームフォーマを形成するから である [53]．ランク制約付き空間共分散行列推定法はまず，ILRMA を観測信号 $\boldsymbol{x}_{ij}$ に適用し，1 個の目的音と雑音が混ざった信号と $M-1$ 個の雑音のみの信号を得る．次に，得られた信号から空間相関行列を推定するが，ILRMA で得られる拡散性雑音の空間相関行列は目的音源方位の分だけランクが 1 つ不足する．これを補うよう確率的定式化を行い，パラメータを推定する．最後に多チャネル Winer フィルタを構成し，目的音源方向の拡散性雑音を低減する．本手法のアルゴリズムはいくつかの拡張及び高速化がなされてい

る [52] が，本論文では最も基本的な，多変量複素 Gauss 分布を観測信号の生成モデルに用い，expectation-maximization (EM) アルゴリズム [68] によって最適化する方法について述べる．

2.4.2 多変量複素 Gauss 分布を用いた生成モデル

観測信号 $\boldsymbol{x}_{ij}$ を目的音源のソースイメージ $\boldsymbol{h}_{ij} = (h_{ij,1}, \ldots, h_{ij,M})^\top$ と拡散性音源のソースイメージ $\boldsymbol{u}_{ij} = (u_{ij,1}, \ldots, u_{ij,M})^\top$ の和として次のように表す．

$$\boldsymbol{x}_{ij} = \boldsymbol{h}_{ij} + \boldsymbol{u}_{ij} \tag{2.19}$$

目的音源のソースイメージ $\boldsymbol{h}_{ij}$ は，ILRMA によって得られた空間基底 $\boldsymbol{a}_{i,1}, \ldots, \boldsymbol{a}_{i,N}$ のうち目的音源に対応するベクトル $\boldsymbol{a}_i^{(h)} =: \boldsymbol{a}_{i,n_h}$ と，目的音源のドライソース $s_{ij}^{(h)}$ を用いて次のように表す．

$$\boldsymbol{h}_{ij} = \boldsymbol{a}_i^{(h)} s_{ij}^{(h)} \tag{2.20}$$

$$s_{ij}^{(h)} | r_{ij}^{(h)} \sim \mathcal{N}_c\left(0, r_{ij}^{(h)}\right) \tag{2.21}$$

$$p(s_{ij}^{(h)} | r_{ij}^{(h)}) = \frac{1}{\pi r_{ij}^{(h)}} \exp\left(-\frac{|s_{ij}^{(h)}|^2}{r_{ij}^{(h)}}\right) \tag{2.22}$$

ここで，n_h は目的音源に対応する音源インデクス，$r_{ij}^{(h)}$ は目的音源の分散 (パワースペクトログラム) である．目的音源の分散 $r_{ij}^{(h)}$ はスパース性を有するとし，事前分布として逆ガンマ分布を仮定する．

$$p(r_{ij}^{(h)}; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(r_{ij}^{(h)}\right)^{-\alpha-1} \exp\left(-\frac{\beta}{r_{ij}^{(h)}}\right) \tag{2.23}$$

ここで，$\alpha > 0$ は形状母数，$\beta > 0$ は尺度母数，$\Gamma(\cdot)$ はガンマ関数を表す．一方，拡散性音源のソースイメージ $\boldsymbol{u}_{ij}$ は目的音源のソースイメージ $\boldsymbol{h}_{ij}$ とは独立な多変量複素ガウス分布に従うと仮定する．

$$\boldsymbol{u}_{ij} \sim \mathcal{N}_c\left(\boldsymbol{0}, r_{ij}^{(u)} \boldsymbol{R}_i^{(u)}\right) \tag{2.24}$$

$$p(\boldsymbol{u}_{ij}) = \frac{1}{\pi^M (r_{ij}^{(u)})^M \det \boldsymbol{R}_i^{(u)}} \exp\left(-\frac{(r_{ij}^{(u)})^{\mathsf{H}} (\boldsymbol{R}_i^{(u)})^{-1} r_{ij}^{(u)}}{r_{ij}^{(u)}}\right) \tag{2.25}$$

ここで, $r_{ij}^{(u)}$ は拡散性雑音のパワースペクトルに対応する時変な分散パラメータであり, $\boldsymbol{R}_i^{(u)} \in \mathbb{C}^{M\times M}$ は拡散性雑音のフルランクの空間相関行列である．ILRMA によって推定された N 個の分離音 $\hat{\boldsymbol{y}}_{ij,1},\ldots,\hat{\boldsymbol{y}}_{ij,N}$ が得られているため，拡散性音源の空間相関行列 $\boldsymbol{R}_i^{(u)}$ は次のように表現できる．

$$\boldsymbol{R}_i^{(u)} = \boldsymbol{R'}_i^{(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}} \tag{2.26}$$

$$\boldsymbol{R'}_i^{(u)} = \frac{1}{J}\sum_j \hat{\boldsymbol{y}}_{ij}^{(u)} \left(\hat{\boldsymbol{y}}_{ij}^{(u)}\right)^{\mathsf{H}} \tag{2.27}$$

$$\hat{\boldsymbol{y}}_{ij}^{(u)} = \boldsymbol{W}^{-1}(\boldsymbol{w}_{i,1}^{\mathsf{H}}\boldsymbol{x}_{ij},\ldots,\boldsymbol{w}_{i,(n_h-1)}^{\mathsf{H}}\boldsymbol{x}_{ij},0,\boldsymbol{w}_{i,(n_h+1)}^{\mathsf{H}}\boldsymbol{x}_{ij},\ldots,\boldsymbol{w}_{i,M}^{\mathsf{H}}\boldsymbol{x}_{ij})^{\top} \tag{2.28}$$

ここで，$\boldsymbol{R'}_i^{(u)} \in \mathbb{C}^{M\times M}$ は ILRMA によって推定された雑音のランク $M-1$ 空間相関行列である．$\boldsymbol{R'}_i^{(u)}$ は $M-1$ 個の雑音成分から計算されるため，そのランクは $M-1$ である．$\boldsymbol{b}_i \in \mathbb{C}^M$ は $\boldsymbol{R'}_i^{(u)}$ の列ベクトルと $\boldsymbol{b}_i$ が線形独立となるようなベクトルであり，λ_i はスカラー変数である．$\boldsymbol{b}_i$ は例えば $\boldsymbol{a}_{n_h}$ や $\boldsymbol{R'}_i^{(u)}$ の零固有値に対応する単位固有ベクトルとする. $\hat{\boldsymbol{y}}_{ij}^{(u)}$ はプロジェクションバック法 [69] によりスケールが補正された $M-1$ 個の拡散性雑音成分のソースイメージの和である．ここで，$\boldsymbol{R'}_i^{(u)}$ で欠けている空間基底を補完及び復元するため，すから変数 λ_i，目的音源の分散 $r_{ij}^{(h)}$，拡散性雑音の分散 $r_{ij}^{(u)}$ を同時に推定する．ILRMA によって推定された，$\boldsymbol{a}_i^{(h)}$，$\boldsymbol{R'}_i^{(u)}$，及び $\boldsymbol{b}_i$ は固定する．

以上のモデリングと，Gauss 分布の再生性より観測信号は多変量複素 Gauss 分布に従う．

$$\boldsymbol{x}_{ij}|r_{ij}^{(h)} \sim \mathcal{N}_c\left(\boldsymbol{0},\boldsymbol{R}_{ij}^{(x)}\right) \tag{2.29}$$

$$\boldsymbol{R}_{ij}^{(x)} = r_{ij}^{(h)}\boldsymbol{a}_i^{(h)}(\boldsymbol{a}_i^{(h)})^{\mathsf{H}} + r_{ij}^{(u)}\boldsymbol{R}_i^{(u)} \tag{2.30}$$

このモデルにより，$r_{ij}^{(h)}$，$r_{ij}^{(u)}$，及び λ_i は観測信号の負対数尤度関数 $\mathcal{L}(\Theta)$ を最小化することにより，推定される．

$$\mathcal{L}(\Theta) = \sum_{i,j}\left[\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij} + \log\det\boldsymbol{R}_{ij}^{(x)} + (\alpha+1)\log r_{ij}^{(h)} + \frac{\beta}{r_{ij}^{(h)}}\right] + \text{const.} \tag{2.31}$$

ここで，$\Theta = \{r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i\}$ は目的変数の集合である．const. は目的変数に依存しない定数である．この負対数尤度関数 $\mathcal{L}(\Theta)$ を直接最適化することは困難であるため，次節に示す EM アルゴリズムを用いて最適化される [51].

2.4.3 EM アルゴリズムによる最適化

目的音声のドライソース $s_{ij}^{(h)}$ と拡散性雑音のソースイメージ $\boldsymbol{u}_{ij}$ を潜在変数として，事後確率 $p(s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\boldsymbol{x}_{ij}; \tilde{\Theta})$ に関する完全対数尤度の期待値をとることで，Q 関数を次のように計算できる．

$$Q(\Theta; \tilde{\Theta}) = \sum_{i,j} \Bigg[-(\alpha+2)\log r_{ij}^{(h)} - M \log r_{ij}^{(u)} - \frac{\hat{r}_{ij}^{(h)} + \beta}{r_{ij}^{(h)}} - \log\det \boldsymbol{R}_i^{(u)} - \frac{\mathrm{tr}\left(\left(\boldsymbol{R}_i^{(u)}\right)^{-1} \hat{\boldsymbol{R}}_{ij}^{(u)}\right)}{r_{ij}^{(u)}} \Bigg] + \mathrm{const.} \tag{2.32}$$

ここで，$\Theta = \{r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i\}$ は最適化すべき目的変数であり，$\tilde{\Theta} = \{\tilde{r}_{ij}^{(h)}, \tilde{r}_{ij}^{(u)}, \tilde{\lambda}_i\}$ は目的変数の現時点での値である．また，$\hat{r}_{ij}^{(h)}$ 及び $\hat{\boldsymbol{R}}_i^{(u)}$ は E ステップにて次のように計算される事後分布の十分統計量である．

$$\tilde{\boldsymbol{R}}_i^{(u)} = \boldsymbol{R}'^{(u)}_i + \tilde{\lambda}_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}} \tag{2.33}$$

$$\boldsymbol{R}_{ij}^{(x)} = \tilde{r}_{ij}^{(h)} \boldsymbol{a}_i^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} + \tilde{r}_{ij}^{(u)} \tilde{\boldsymbol{R}}_i^{(u)} \tag{2.34}$$

$$\hat{r}_{ij}^{(h)} = \tilde{r}_{ij}^{(h)} - \left(\tilde{r}_{ij}^{(h)}\right)^2 \left(\boldsymbol{a}_i^{(h)}\right)^{\mathsf{H}} \left(\tilde{\boldsymbol{R}}_{ij}^{(x)}\right)^{-1} \boldsymbol{a}_i^{(h)} + \left| \tilde{r}_{ij}^{(h)} \boldsymbol{x}_{ij}^{\mathsf{H}} \left(\tilde{\boldsymbol{R}}_{ij}^{(x)}\right)^{-1} \boldsymbol{a}_i^{(h)} \right|^2 \tag{2.35}$$

$$\hat{\boldsymbol{R}}_{ij}^{(u)} = \tilde{r}_{ij}^{(u)} \tilde{\boldsymbol{R}}_i^{(u)} - \left(\tilde{r}_{ij}^{(u)}\right)^2 \tilde{\boldsymbol{R}}_i^{(u)} \left(\boldsymbol{R}_{ij}^{(x)}\right)^{-1} \tilde{\boldsymbol{R}}_i^{(u)} + \left(\tilde{r}_{ij}^{(u)}\right)^2 \tilde{\boldsymbol{R}}_i^{(u)} \left(\tilde{\boldsymbol{R}}_{ij}^{(x)}\right)^{-1} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} \left(\tilde{\boldsymbol{R}}_{ij}^{(x)}\right)^{-1} \tilde{\boldsymbol{R}}_i^{(u)} \tag{2.36}$$

M ステップでは，Q 関数を各変数に関して座標上昇法を用いて最大化する．

$$r_{ij}^{(h)} \leftarrow \frac{\hat{r}_{ij}^{(h)} + \beta}{\alpha + 2} \tag{2.37}$$

$$\lambda_i \leftarrow \boldsymbol{b}_i^{\mathsf{H}} \left(\frac{1}{J} \sum_j \frac{1}{\tilde{r}_{ij}^{(u)}} \hat{\boldsymbol{R}}_{ij}^{(u)} \right) \boldsymbol{b}_i \tag{2.38}$$

$$\boldsymbol{R}_i^{(u)} \leftarrow \boldsymbol{R}'^{(u)}_i + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}} \tag{2.39}$$

$$r_{ij}^{(u)} \leftarrow \frac{1}{M} \mathrm{tr}\left(\left(\boldsymbol{R}_i^{(u)}\right)^{-1} \hat{\boldsymbol{R}}_{ij}^{(u)}\right) \tag{2.40}$$

2.5 BS-ILRMA

BS-ILRMA は人の進入が困難な災害環境で生存者の声を検知するロボットのための手法として提案された．災害用ロボットの中でも図 2.1 に示すような柔軟索状ロボット [13] は災害環境下において狭く暗い瓦礫の中に進入し，その中の生存者を発見するために開発されている．このロボットは，機体の節に取り付けられた振動モータによって自身を振動させることにより自走することが可能であり，コントロールデバイスやオペレータによる操作なしで移動することができる．ロボットの目的は，ロボットの機体の周りに取り付けられたマイクロホンにより，瓦礫の中に埋もれてしまった生存者の声をとらえることである．ところが，振動モータによる振動が大きな内部雑音 (エゴノイズ) を発生させてしまうため，生存者をロバストに発見するためには観測した生存者の声とエゴノイズを分離しなければならない．BS-ILRMA は前もって収録されたロボットのエゴノイズを利用した半教師あり音源分離である．一方，補聴器を使う状況においても，会話直前の数秒の雑音区間など事前に雑音のサンプルを利用することで半教師あり音源分離を適用することが可能である．

BS-ILRMA は $M = N$ の条件のうち，$N' = M' = N - 1$ 個の雑音源と 1 個の目的音源が存在する状況を仮定している．事前に得られる M' チャネルの雑音のサンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び分離対象である M チャネルの観測信号 $\boldsymbol{x}_{ij'}^{(\mathrm{mix})}$ を以下のように表す．

$$\boldsymbol{x}_{ij'}^{(\mathrm{noise})} = (x_{ij',1}^{(\mathrm{noise})}, \ldots, x_{ij',m'}^{(\mathrm{noise})}, \ldots, x_{ij',M'}^{(\mathrm{noise})}) \tag{2.41}$$

$$\boldsymbol{x}_{ij}^{(\mathrm{mix})} = (x_{ij,1}^{(\mathrm{mix})}, \ldots, x_{ij,M}^{(\mathrm{mix})})^{\top} \tag{2.42}$$

ここで，$j' = 1, \ldots, J'$ 及び $m' = 1, \ldots, M'$ は雑音サンプルの観測信号のフレーム及びインデクスである．

単純に ILRMA を半教師ありにする場合，次のようなステップで半教師あり NMF [54, 55] と ILRMA を組み合わる手法 (semi-supervised ILRMA: SS-ILRMA) が考えられる．

step 1 エゴノイズサンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ に対して ILRMA を用いて分離を行う．

step 2 step 1 の最適化の結果から，学習済みの基底行列 $\boldsymbol{T}_{m'}^{(\mathrm{noise})} \in \mathbb{R}_{\geq 0}^{I \times L}$ が得られる．

step 3 もう一つの ILRMA で観測信号 $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ を分離する．その際に，NMF による音源モデルの学習において，$M - 1$ チャネルを基底行列を学習済みの基底行列 $\boldsymbol{T}_{n'}^{(\mathrm{noise})}$ に固定し，残りの 1 チャネルのみを最適化する．

ここで，$n' = 1, \ldots, N'$ は雑音サンプルの音源のインデクスである．$\boldsymbol{T}_{n'}^{(\mathrm{noise})}$ 以外の他の変数 (教師エゴノイズの音源に対するアクティベーション行列，未知である音声の音源に対する基

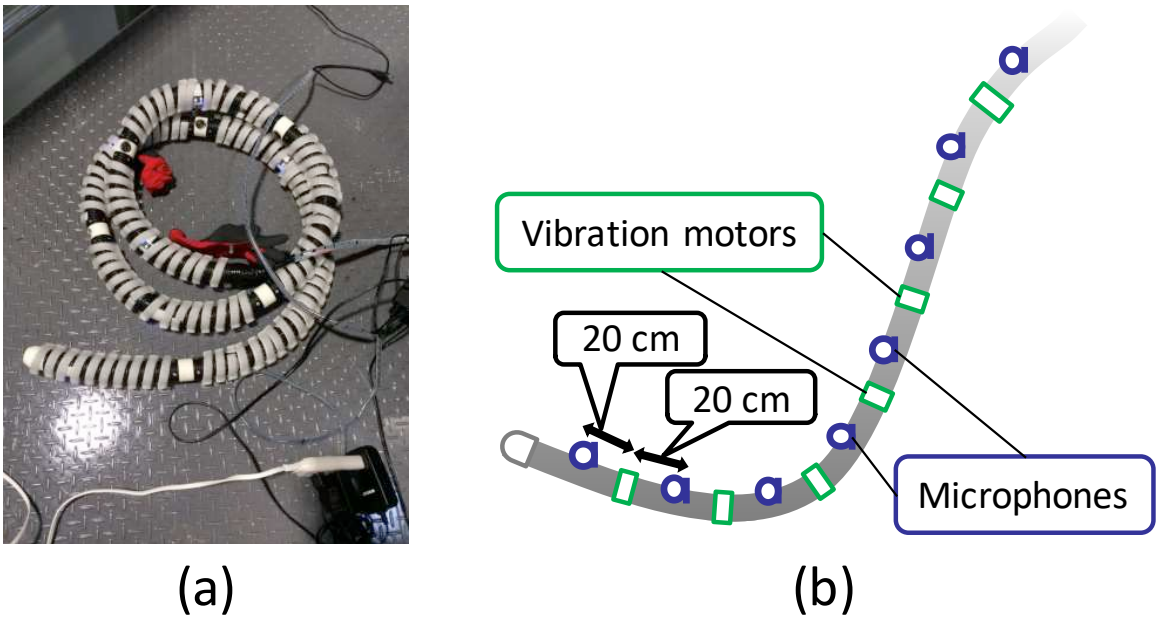


図 2.1: (a) Hose-shaped rescue robot and (b) structure of rescue robot

底行列及びアクティベーション行列，及び分離フィルタ $\boldsymbol{W}_i$) は半教師あり NMF [54,55] と同様に最適化を行う．しかし，単純な半教師ありアプローチでは $\boldsymbol{W}_i$ と基底行列の間のスケールの不定性が雑音の教師基底行列 $\boldsymbol{T}_{n'}^{(\mathrm{noise})}$ のスペクトル構造を崩壊させる可能性がある [14].

この問題に対処するため BS-ILRMA が提案された．BS-ILRMA の概要を Fig. 2.2 に示す．ここで，$\boldsymbol{W}_i^{(\mathrm{noise})} \in \mathbb{C}^{N' \times M'}$ 及び $\boldsymbol{W}_i^{(\mathrm{mix})} \in \mathbb{C}^{N \times M}$ は，雑音サンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び観測信号 $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ に対する分離行列である．$\boldsymbol{X}_{m'}^{(\mathrm{noise})} \in \mathbb{C}^{I \times J'}$ 及び $\boldsymbol{Y}_{n'}^{(\mathrm{noise})} \in \mathbb{C}^{I \times J'}$ はそれぞれ $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び $\boldsymbol{y}_{ij'}^{(\mathrm{noise})} = (y_{ij',1}^{(\mathrm{noise})}, \ldots, y_{ij',N'}^{(\mathrm{noise})})^\top$ の m' 及び n' 番目のスペクトログラムである．$\boldsymbol{X}_m^{(\mathrm{mix})} \in \mathbb{C}^{I \times J}$ 及び $\boldsymbol{Y}_n^{(\mathrm{mix})} \in \mathbb{C}^{I \times J}$ はそれぞれ $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ 及び $\boldsymbol{y}_{ij}^{(\mathrm{mix})} = (y_{ij,1}^{(\mathrm{mix})}, \ldots, y_{ij,N}^{(\mathrm{mix})})^\top$ の m 及び n 番目のスペクトログラムである．また，$|\cdot|^{.2}$ は要素毎の 2 乗を表す．$\boldsymbol{T}_{n'} \in \mathbb{R}_{\geq 0}^{I \times L}$ は雑音サンプルの音源に対する共有基底行列，$\boldsymbol{T}_N \in \mathbb{R}_{\geq 0}^{I \times L}$ は目的音源に対する非共有基底行列，$\boldsymbol{V}_{n'}^{(\mathrm{noise})} \in \mathbb{R}_{\geq 0}^{L \times J'}$ 及び $\boldsymbol{V}_n^{(\mathrm{mix})} \in \mathbb{R}_{\geq 0}^{L \times J}$ は $\boldsymbol{Y}_{n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{Y}_n^{(\mathrm{mix})}$ に対応するアクティベーション行列である．BS-ILRMA は二つの ILRMA を使用する．一方では $\boldsymbol{W}_i^{(\mathrm{noise})}$ 及び $\boldsymbol{y}_{ij'}^{(\mathrm{noise})}$ を推定するため，ILRMA を雑音サンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ に適用する．もう一方では $\boldsymbol{W}_i^{(\mathrm{mix})}$ 及び $\boldsymbol{y}_{ij}^{(\mathrm{mix})}$ を推定するため，ILRMA を観測信号 $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ に適用する．最も重要な点は，雑音サンプルの音源に対する基底行列 $\boldsymbol{T}_{n'}$ は二つの ILRMA 間で共有されており，これらのモデルにおける全ての変数は同時に最適化されていることである．共有基底行列 $\boldsymbol{T}_{n'}$ は $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ の両方で類似したスペクトルを表

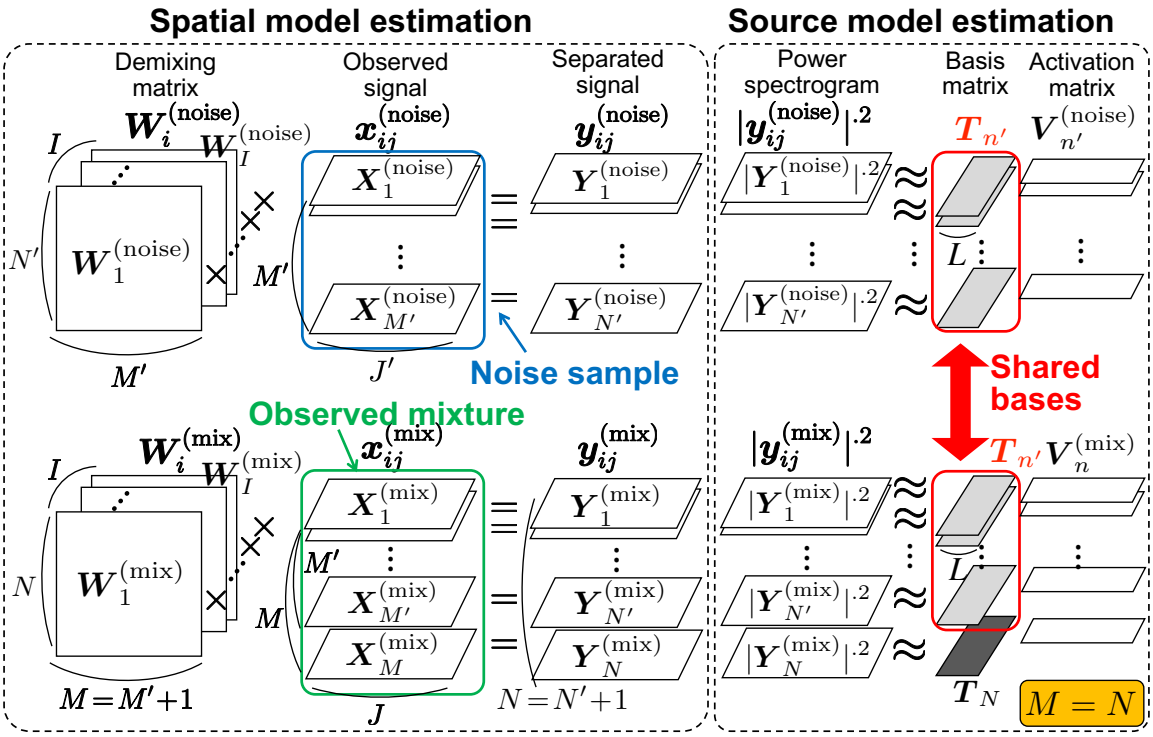


図 2.2: Overview of BS-ILRMA, where upper and lower models are *simultaneously* optimized.

さなければならないため，雑音サンプルのスペクトルパターンは $\boldsymbol{T}_{n'}$ によって捉えられ，基底行列 $\boldsymbol{T}_N$ は結果的に残った目的音源のスペクトルパターンを表現する．

BS-ILRMA のコスト関数は，二つの ILRMA のコストの和として次のように定義される．

$$
\begin{aligned}
\mathcal{J} =& \frac{1}{N'}\left\{\sum_{n'=1}^{N'}\sum_{i,j'}\left[\frac{|y_{i,j',n'}^{(\mathrm{noise})}|^2}{\sum_l t_{il,n'}v_{lj',n'}^{(\mathrm{noise})}} + \log\sum_l t_{il,n'}v_{lj',n'}^{(\mathrm{noise})}\right] - 2J'\sum_i \log|\det \boldsymbol{W}_i^{(\mathrm{noise})}|\right\} \\
&+ \frac{1}{N}\left\{\sum_{n=1}^{N}\sum_{i,j}\left[\frac{|y_{i,j,n}^{(\mathrm{mix})}|^2}{\sum_l t_{il,n}v_{lj,n}^{(\mathrm{mix})}} + \log\sum_l t_{il,n}v_{lj,n}^{(\mathrm{mix})}\right]\right. \\
&+ \sum_{i,j}\left[\frac{|y_{i,j,N}^{(\mathrm{mix})}|^2}{\sum_l t_{il,N}v_{lj,N}^{(\mathrm{mix})}} + \log\sum_l t_{il,N}v_{lj,N}^{(\mathrm{mix})}\right] \\
&\left.- 2J\sum_i \log|\det \boldsymbol{W}_i^{(\mathrm{mix})}|\right\}
\end{aligned}
\tag{2.43}
$$

ここで，$t_{il,n'}$ 及び $t_{il,N}$ はそれぞれ $\boldsymbol{T}_{n'}$ 及び $\boldsymbol{T}_N$ の要素，$\boldsymbol{v}_{lj,n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{v}_{lj,n}^{(\mathrm{mix})}$ はそれぞれ $\boldsymbol{V}_{n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{V}_n^{(\mathrm{mix})}$ の要素を表す．共有されないパラメータの更新式は [30] と同じである．一方，共

有基底 $t_{il,n'}$ に関して式 (2.43) を直接最小化することは困難であるため，補助関数法に基づき補助関数を設計し，これを最小化することで局所最適解を得る [14].

2.6 本章のまとめ

本章では，本論文で取り上げる手法について述べた．まず，基本的な BSS 手法の定式化を行った，次に，ブラインドの枠組みの手法として ILRMA について述べ，実環境下で有効な手法であるランク制約付き空間共分散行列推定法について述べた．さらに，半教師ありの枠組みの手法として，BS-ILRMA について述べた．

第3章　提案補聴器システム

3.1　はじめに

本章では，分散マイクロホンアレーに基づく新たな補聴器システムを提案する．会議シーンでは，PC やスマートフォンなどマイクロホンを内蔵したデバイスが複数ある状況が考えられる．近年では，こうした状況を対象にした分散マイクロホンアレー処理に基づく音源分離も広く研究されている．分散マイクロホンアレー処理は，その場に存在するマイクロホンを利用してアレー処理を行えるため，得られる音源情報も多くなり，さらに，広い範囲の空間情報も得ることができる．近年はスマートフォンが広く普及しているという事実もあり，補聴器ユーザが所持するスマートフォンを用いることで，場面を選ばずに分散マイクロホンアレー処理が行えると考えられる．本論文では，両耳のマイクロホンだけでなくスマートフォンに内蔵されているマイクロホンも含めた分散マイクロホンアレー補聴器システムを新たに提案する．本論文で提案する補聴器システムは，補聴器ユーザ自身が持つスマートフォンを用いてマイクロホンアレーを構成する．将来的には，会話している相手のスマートフォンや，その場にいる多くの人々が持つマイクロホン内蔵のデバイスを使い，さらに高品質な補聴を達成できるなど，提案補聴器システムは高い拡張性を持つ．

想定している提案補聴器システムの利用シーンを図 3.1 に示す．周囲に雑音がある中で，補聴器のユーザが，ユーザ自身の所持するスマートフォンを胸の前方に向けて会話する．まず，提案補聴器システムでデータの収集を行う．図 3.1 に沿って，スマートフォンを持った人間を模したダミーヘッドを用意する．ダミーヘッドの胸部にスマートフォンを取り付け，スマートフォンと両耳を含めた複数のマイクロホンで実環境を想定した収録を行う．ここで，分散マイクロホンアレー処理には．マイクロホンの位置推定とデバイス間のサンプリング同期の問題がある [56]．前者の位置推定の問題については，ブラインドの枠組みによる音源抽出であるためアレーの固定することで致命的な問題とならない後者のサンプリング同期の問題について，マイクロホンを同期するには A-D 変換器を用いて同期させる必要がある．A-D 変換器は高価かつ大規模であり，有線でつなぐ必要があるため，実際の利用シーンで利用することは困難である．本論文では，音源抽出に焦点を当てて A-D 変換器を用いて全てのマイクロ

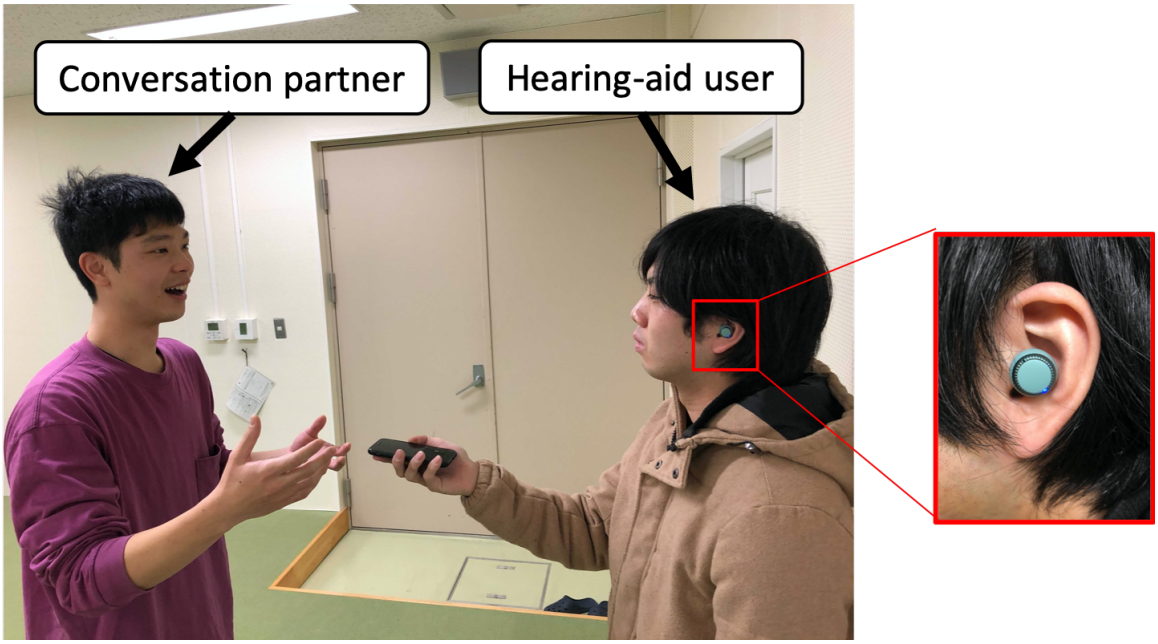


図 3.1: View of conversation using proposed hearing-aid system. Hearing-aid user holds smartphone in front of user's chest.

ホンを予め同期させるが，サンプリングの問題を解決するには，マイクロホンを同期させる手法 [61,70] を適用することで解決できる．提案補聴器システムの音声抽出手法として，実環境下のような拡散性雑音下でも，高速かつ高品質に動作するランク制約付き共分散行列推定法を採用する．

3.2 システムの仕様

本研究では実環境下で対面する人との会話シーンを想定し，8 個のマイクロホンを用いてインパルス応答と拡散性雑音の収録を行う．収録のため，図 3.2 (a) のように，スマートフォンを持った人間を模したダミーヘッドを作成した．ダミーヘッドの両耳には図 3.2 (b) 及び (d) のように，片耳に 3 個ずつ，両耳を合わせて計 6 個の無指向性マイクロホンを取り付けた．スマートフォンは，図 3.2 (c) のようにダミーヘッドの胸部から 20 cm の位置に取り付け，胸部側に先端が向くよう 2 個の無指向性マイクロホンを 4 cm の間隔で取り付けた．合計 8 ch のマイクロホンを装備したダミーヘッドを用いて収録を行う．便宜上，図 3.2 (b)，(c)，(d) のように各マイクロホンに対してナンバリングを行った．ここで，分散マイクロホンアレー処理を行う上で．デバイス間のサンプリング同期の問題がある [56] が，本論文では音源抽出につい

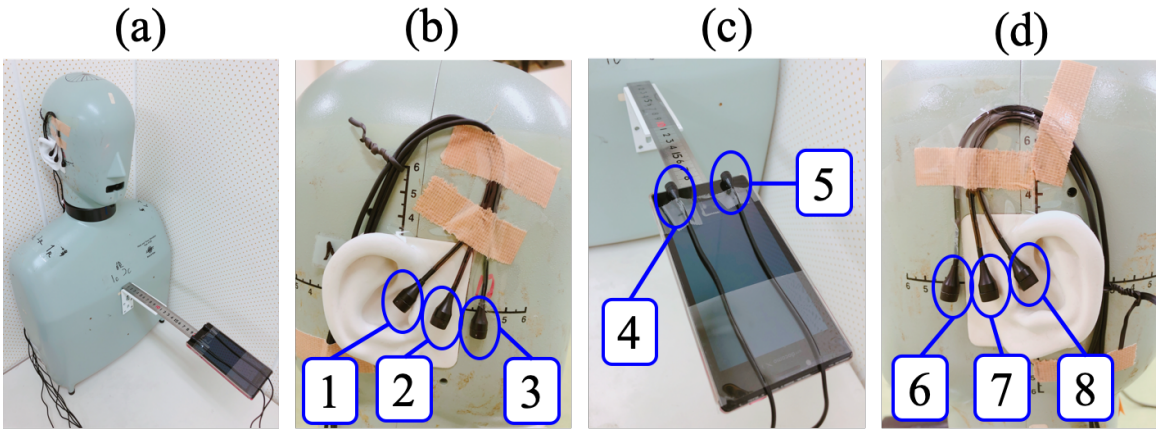


図 3.2: (a) Overall view of head-and-torso dummy, (b) right-ear microphone array, (c) smartphone's microphones, and (d) left-ear microphone array.

て焦点をあて，用いるマイクロホンは全て多チャネル A-D 変換器を用いて予め同期させる．また，分散マイクロホンアレーにおいて同期の問題を解決する手法もいくつか提案されている [61, 70]．ダミーヘッドの身長は 170 cm とし，高さを調節した台に図 3.2 (a) のダミーヘッドを乗せて収録した．また，ダミーヘッドと同身長の人との対話を想定し，床から口元までの高さを測り，スピーカの高さを 152 cm とした．

3.3　インパルス応答と拡散性雑音の収録

インパルス応答の計測方法として, 時間引き伸ばしパルス (time stretched pulse: TSP) 信号 [71] を用いた．収録環境及び TSP 信号の収録条件を表 3.1 に示す．ダミーヘッドからスピーカへの距離を 75 cm, 100 cm, 150 cm に，角度は正面方向 (0°) に加え左右にそれぞれに 20° 変化させ，計 9 パターンのスピーカ位置における TSP 信号を計測した．収録場所は屋内の一室とした．実際にダミーヘッドとその正面方向 (0°) にスピーカを配置した部屋の写真を図 3.3 (a) に示す．収録時のダミーヘッドやスピーカの位置関係，及び計測する 9 パターンのスピーカの位置の概略図を図 3.3 (b) に示す．

雑音データの作成として，数人が自由に移動・会話している状況を想定し収録を行った．約 20 人の協力者を募り，自由に移動・発話してもらった．雑音源は目的話者より外に存在するとしているため，協力者にはダミーヘッド前方半径 150 cm の半円より外側を周回させた．図 3.4 に，拡散性雑音の収録風景を示す．

表 3.1: Recording conditions and devices

Recording location	Studio (see Fig. 3.3)
Reverberation time (T_{60})	300 ms
Microphone	C417 PP (AKG)
Loudspeaker	ADIVA11 (Anthony Gallo)
Audio interface	828x (MOTU)
Data format of TSP signal	48 kHz, 16 bit, WAVE file format
TSP length	65536 samples
Recording sampling freq.	48 kHz
Number of synchronous addition	20 times

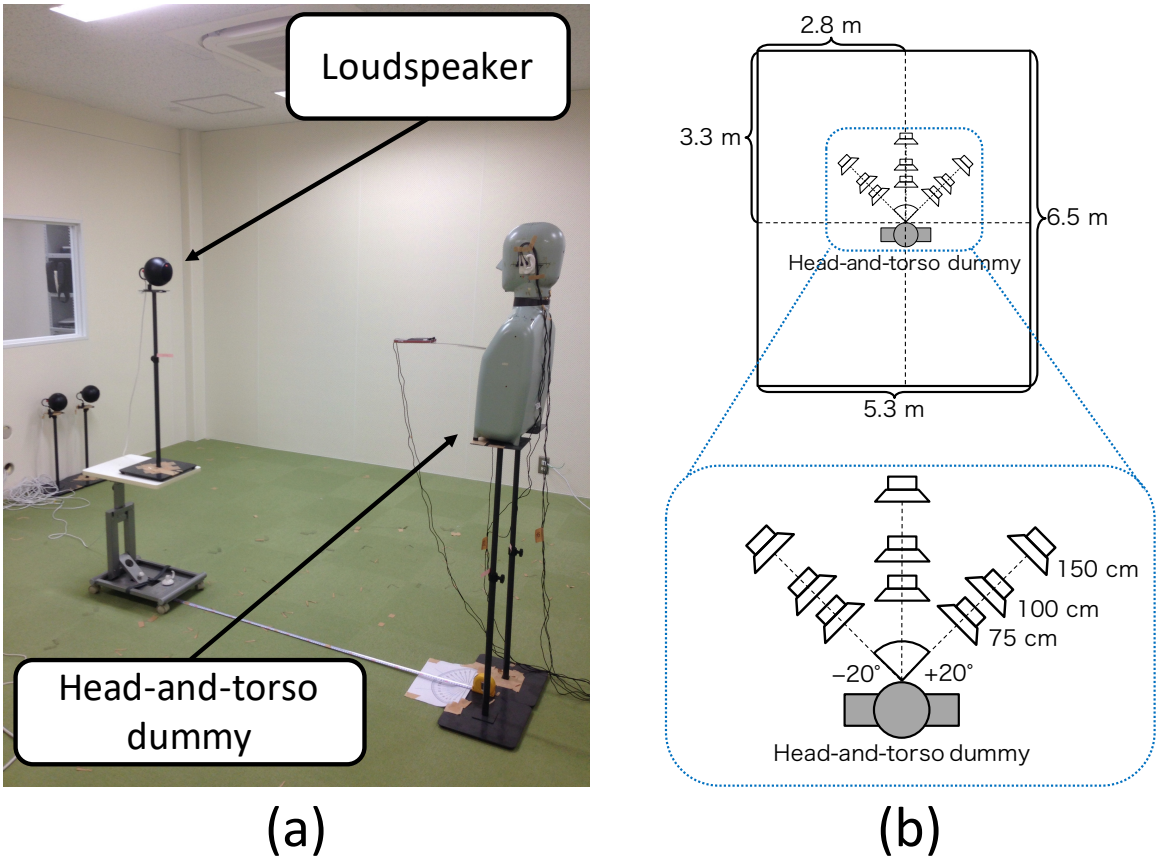


図 3.3: (a) Sets for recording TSP signal in room and (b) room configuration. Position of loudspeaker (mouth of conversation partner) for nine recording cases.

図 3.4: View of noise recording. Approximate 20 people talk and walk around room freely.

3.4 本章のまとめ

本章では，両耳のマイクロホンだけでなくスマートフォンのマイクロホンも含めた新たな分散マイクロホンアレー補聴器システムを提案した．ダミーヘッドを用いて，スマートフォンを持った人間を模した装置を構築し，インパルス応答及び拡散性雑音の収録を行った．

第4章　提案補聴器システムへのBSE手法の利用可能性及び分散マイクロホンアレーによる分離性能改善の評価

4.1　はじめに

本章では，実験提案補聴器システムに対するBSE手法の適用可能性と，提案補聴器システムの有効性を評価するための実験を行う．提案補聴器システムは全く新しいアプローチであるため，既存のびBSE手法が有効に動作することは保証されていない．そのため，まず実環境下で有効なランク制約付き空間共分散行列推定法の3.3節で収録したデータに対する有効性を調査する．ランク制約付き空間共分散行列推定法は式(2.23)にある通り，形状母数αと尺度母数βの2つの内部パラメータを持つ．特に，形状母数αは音源信号のパワースペクトルに対応する分散にスパース性を誘引するパラメータであり，処理後の品質に大きく関係すると考えられる．しかし，αの値と分離性能の関係は明らかにされておらず，この関係を明らかにすることで実際の利用時により高品質な音源抽出が達成できると考えられる．そのため，本章の実験で，提案補聴器システムへの利用可能性と併せて形状母数αの値と分離性能の関係についても調査する．

一方，提案補聴器システムはスマートフォンのマイクロホンを利用することで，利用しない場合と比較して次の2点で有利であると考えられる．

(a) マイクロホンの総数が増え，利用できる音源情報及び空間情報が多くなる．

(b) 目的音源に近い位置の空間情報が得られる．

実際にこれら2点が優位に働いているかどうか，ILRMAによる分離を行い調査する．4.2節では，一連の評価実験の条件を述べる．次に，4.3節では，収録したデータに対してILRMAとランク制約付き空間共分散行列推定法を適用・比較し，ランク制約付き空間共分散行列推定法が提案補聴器システムに利用可能であることを示す．併せて，ランク制約付き空間共分散行列推定法の内部パラメータαと分離性能の関係を調査する．最後に，4.4節で，上で述べ

表 4.1: Experimental conditions

Sampling frequency	16 kHz
FFT length	1024 sample (50% overlap)
Window	Hamming window
Number of bases in low-rank model	10
Number of iterations in ILRMA	50
Initialization of $\boldsymbol{W}_i$ in ILRMA	Identity matrix
Number of iterations in rank-constrained SCM estimation	10

たスマートフォンの利用による 2 つのメリットが，実際に作用しているか実験的に明らかにする．

4.2 実験条件

本評価実験の目的は，両耳とスマートフォンのマイクロホンを用いた補聴器体系において，実環境下での ILRMA とランク制約付き空間共分散モデル推定法を比較し，収録データに対するランク制約付き空間共分散モデル推定法の有効性について調査することである．音声データベース JNAS [72] の女声データ 1 文 に 3.3 節で収録したインパルス応答を畳み込んだものを目的信号とした．拡散性雑音には 3.3 節で収録した雑音を用いた．ただし，使用したコーパスデータのサンプリング周波数が 16 kHz であったため，48 kHz で収録したインパルス応答及び雑音をダウンサンプリングした．実験するにあたり，入力 SNR は -10 dB, -5 dB, 0 dB, ランク制約付き空間共分散モデル推定法における形状母数パラメータ α は 0.5, 1.1, 10, 20 と変化させ，尺度母数パラメータ β は 10^{-16} とした．ILRMA とランク制約付き空間共分散モデル推定法において観測信号を主成分分析を用いて白色化を行い，異なる乱数初期値で 10 回試行した．その他の条件を表 4.1 に示す．以上の条件で，評価尺度として source-to-distortion ratio (SDR) 改善量 [73] を用いて分離性能を比較した．

4.3 収録データに対する既存手法の分離性能

入力 SNR が -10 dB，マイクロホン 1 (右耳外耳道付近, 図 3.2 参照) の，ILRMA 及びランク制約付き空間共分散行列推定法の平均 SDR 改善量を図 4.1 を示す．横軸はランク制約付き空間共分散行列推定法の iteration を表す．ILRMA は反復 50 回目の SDR 改善量を示しており，ランク制約付き空間共分散行列推定法の iteration に依存せず一定となっている．全ての場合

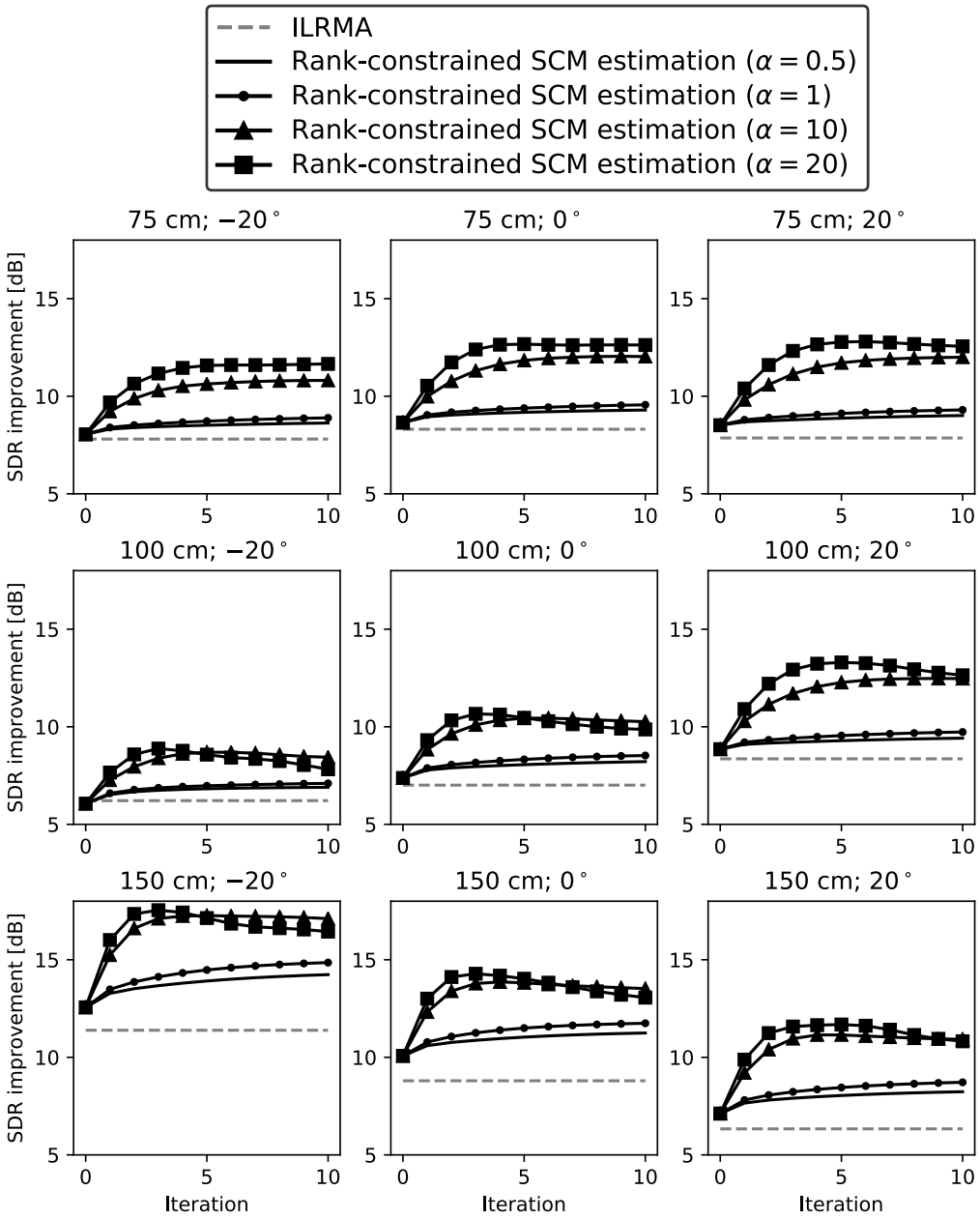


図 4.1: Average SDR improvements for each iteration at microphone 1 under -10 dB input SNR condition. Rows indicate distance from head-and-torso dummy to loudspeaker and columns indicate direction.

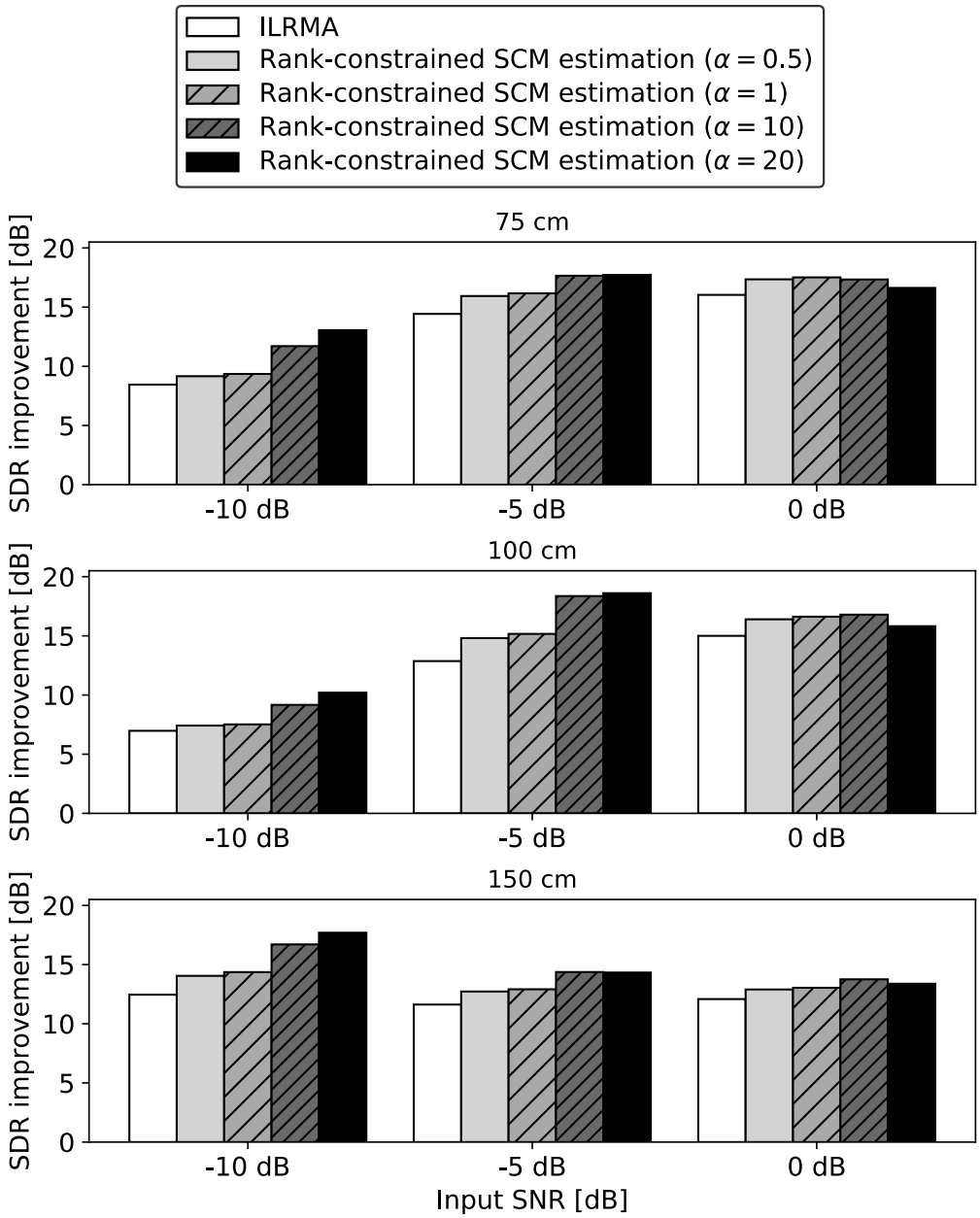


図 4.2: Average SDR improvements of ILRMA and rank-constrained SCM estimation after two iterations at microphone 1 when target source is located at 0°.

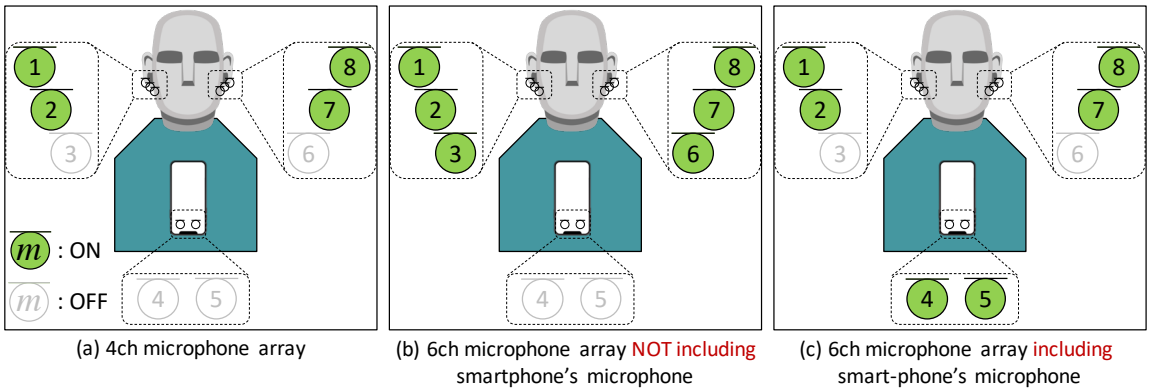


図 4.3: Enabled microphones to evaluate effectiveness of proposed hearing-aid system by including smartphone. Numbers of enabled microphones are (a) four (No. 1, 2, 7, and 8), (b) six (No. 1, 2, 3, 6, 7, and 8) not including smartphone's microphones, and six microphones (No. 1, 2, 4, 5, 7, and 8) including smartphone's microphones.

においてランク制約付き空間共分散モデル推定法が ILRMA を上回っていることがわかる．また，内部パラメータ α の値によって SDR 改善量の変化に大きく違いが現れることが確認できた．内部パラメータ α は，大きいほど少ない反復で高い SDR 改善量を達成し得る．ただし，一定の反復回数を超えると SDR 改善量が減少する傾向にあり，最適な反復回数が不明な場合には，小さな α を設定して安定した分離を達成することも可能である．今回調査した範囲では，2〜5 回程度の少ない反復で高い SDR 改善量を達成することが分かった．これにより，収録データに対しても，ランク制約付き空間共分散モデル推定法は速く収束し，かつ高い SDR 改善量を達成できることが示された．

次に，角度を $-20°$ に限定し，各入力 SNR における SDR 改善量の傾向について，同様に性能を調査する．ただし，ランク制約付き空間共分散モデル推定法は上記の結果に基づき，SDR 改善量が概ね大きい反復 3 回目の結果を用いて比較を行う．マイクロホン 1 での平均 SDR 改善量の結果を図 4.2 に示す．ランク制約付き空間共分散モデル推定法について，入力 SNR が -10 dB，-5 dB の場合の SDR 改善量が ILRMA と比較して大きい．このことから，ランク制約付き空間共分散モデル推定法は低い入力 SNR の場合により高い音声抽出を達成することがわかった．これは，低い入力 SNR でよりクリティカルになった雑音をランク制約付き空間共分散モデル推定法が抑圧できているためと考えられる．

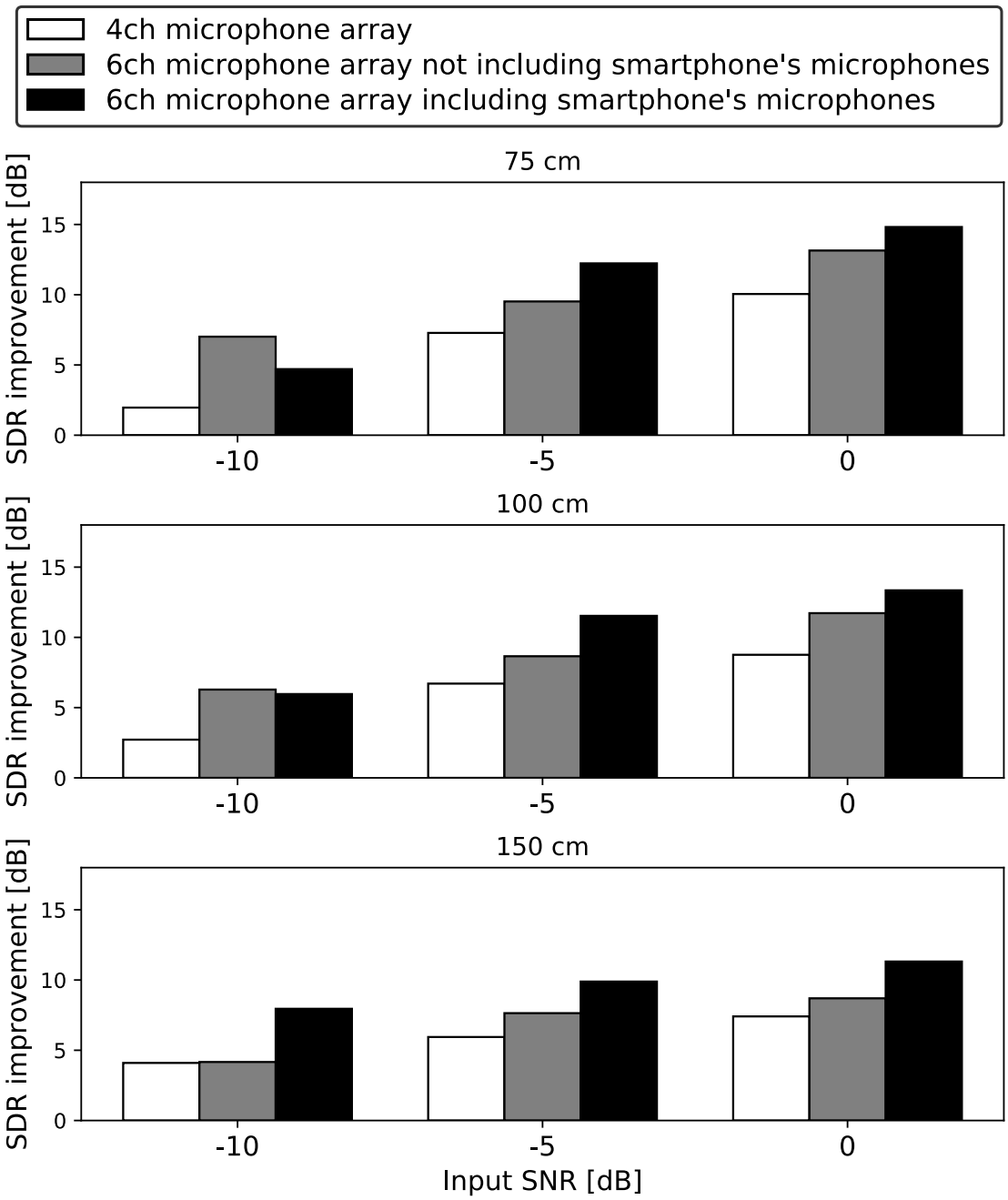


図 4.4: Average SDR improvements of ILRMA for three patterns microphone arrays.

4.4 スマートフォンのマイクロホン利用による分離性能の改善

提案補聴器システムは両耳だけでなくスマートフォンに内蔵されているマイクロホンも用いている．スマートフォンの利用によって，マイクロホンの総数が増え，目的音源に近い位

置の空間情報が得られる．本節の実験で，これら 2 つの利点が実際に分離性能の改善に寄与しているかを調査する．そのために，ダミーヘッドに取り付けた計 8 つのマイクロホンのうち，図 4.3 に示す 3 パターンのマイクロホンに限定して比較を行う．図 4.3 (a) と (b) の場合で分離性能を比較することにより，マイクの総数が増えることによる効果を調査する．また，図 4.3 (b) と (c) の場合で分離性能を比較することにより，目的音源に近い位置での空間情報を得られる効果を調査する．ILRMA を用いて分離を行い，SDR 改善量を用いて分離性能の比較を行った．

結果を図 4.4 に示す．結果から，4ch で分離した場合の SDR 改善量に比べ，6ch で分離した場合の SDR 改善量がほとんどの場合で大きいことがわかる．さらに，スマートフォンのマイクロホンを使った場合の方が，使わなかった場合に比べて SDR 改善量が高いことがわかる．以上から，提案補聴器システムにおいて，スマートフォンのマイクロホンを利用することで，マイクロホンの総数が増え，目的音源に近い位置の空間情報を利用できることによる分離性能の改善効果が確認できた．

4.5 本章のまとめ

本章では，収録した実環境データに対する，ランク制約付き空間共分散行列推定法の有効性を確認した．併せて，明らかにされていなかった内部パラメータと分離性能の関係を明らかにした．さらに，提案補聴器システムにおいてスマートフォンのマイクロホンを利用することによる 2 つの利点が，実際に分離性能の向上に寄与していることを実験的に明らかにした．

第5章　提案補聴器システムへのBS-ILRMAの利用可能性及びランク制約付き空間共分散行列推定法への適用

5.1　はじめに

本章では，半教師あり音源分離の枠組みである BS-ILRMA の提案補聴器システムに対する有効性を示す．さらに，ランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を利用して，さらに高品質な音源抽出が達成できることを示す．補聴器を用いて会話するシーンでは，会話が始まる直前の雑音のみのデータを収録することができる．また，事前に収録する雑音は会話直前の数秒であるため，分離時とアレーの配置を保存することができる．このため，提案補聴器システムにも BS-ILRMA のような半教師ありアプローチを組み込むことが可能になる．ただし，BS-ILRMA は元々災害用ロボットのために提案された手法であり，提案する補聴タスクは生存者の声とエゴノイズの分離タスクと比較して，表 5.1 に示すように 3 つの分離に不利な要素が考えられる．1 つは雑音の種類が多いことである．エゴノイズの分離タスクでは雑音はエゴノイズのみであるが，補聴タスクでは会話相手の声以外の声や，歩行音など様々である．2 つ目は雑音源までの距離である．エゴノイズはロボット自身が発するため雑音源までの距離が近い一方，補聴タスクでは遠方かつ全方位から拡散性の雑音が到来する．3 つ目は事前に得られる雑音情報の長さである．エゴノイズは事前に十分な長さを収録することが可能であるが，補聴タスクでは会話直前の数秒程度である．こうした不利な条件下でも，BS-ILRMA が有効に動作することを確認し，提案補聴器システムに対しても半教師ありアプローチが利用できることを示す．一方で，ランク制約付き空間共分散行列推定法は，初期化方法に ILRMA を用いた初期化を行い，一部のパラメータを推定している．提案補聴器システムのデータに対する BS-ILRMA の有効性を示した後，BS-ILRMA をランク制約付き空間共分散行列推定法の初期化方法に用いて，より高品質な音源抽出が達成できることを示す．

表 5.1: Difficulty of hearing-aid task compared with ego-noise suppression task

	Ego-noise suppression task	Hearing-aid task
Noise type	Ego-noise (only)	Voice or footstep, etc. (plural)
Distance from Microphones to noise source	Close	Far
Length of noise data we can obtain in advance	Any	Few seconds before conversation

5.2 収録データに対する BS-ILRMA の分離性能

まず，提案補聴器システムで収録したデータに対して，通常の ILRMA，雑音サンプルを用いて事前に基底を学習して分離を行う SS-ILRMA，及び基底共有により雑音学習用と分離用のモデルを同時に最適化する BS-ILRMA をそれぞれ適用し，分離性能を調査する．実験条件は 4.2 節と同様である．目的音声が発話される直前の 2 秒間に雑音区間を設け，この 2 秒の雑音区間を学習に用いた．SS-ILRMA については雑音の事前学習として，基底を 50 回更新したものを使用した．異なる乱数初期値で 10 回試行した．以上の条件で，評価尺度として SDR 改善量を用いて，マイクロホン 1 (右耳外耳道付近) での各角度及び異なる乱数初期値での結果を平均して比較した．

結果を図 5.1 に示す．まず，SS-ILRMA は ILRMA に比べて概ね高い SDR 改善量を達成しているが，条件によって ILRMA に劣る場合がある．これは，事前に学習した基底行列と分離時の基底行列のスケールの曖昧さが要因であると考えられる．一方，BS-ILRMA がほぼ全ての場合において通常の ILRMA や SS-ILRMA に比べて高い SDR 改善量を達成していることがわかる．このことから，エゴノイズの分離タスクと比べて不利となる条件があるにも関わらず，提案補聴器システムに対しても BS-ILRMA が利用可能であることがわかった．さらに，BS-ILRMA はランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を用いることで従来に比べ高品質な分離が期待できる．

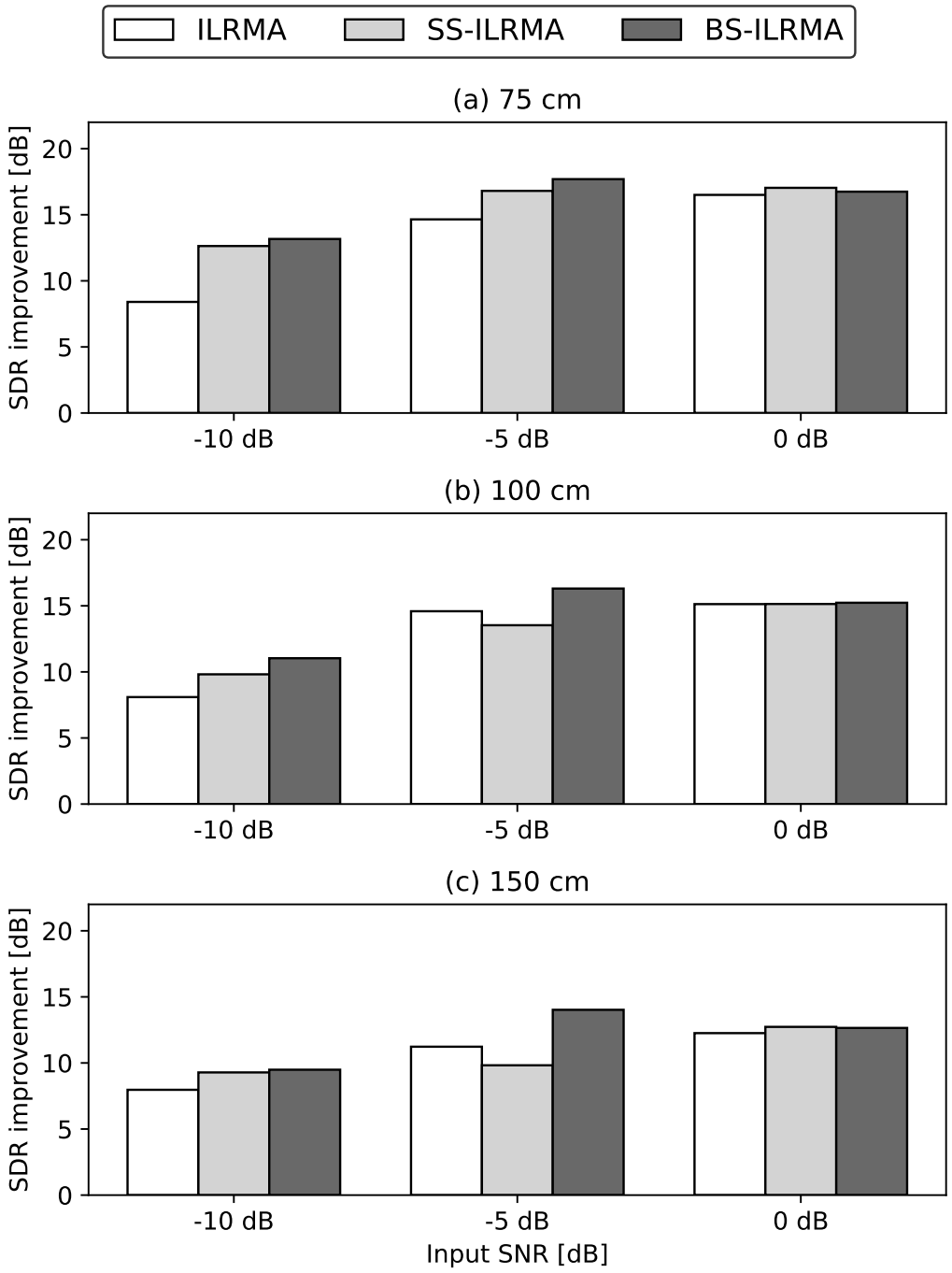


図 5.1: Average SDR improvements of ILRMA, SS-ILRMA, and BS-ILRMA under each input SNR condition. Three figures show results when distance from head-and-torso dummy to target source is set to (a) 75, (b) 100, and (c) 150 cm, respectively.

5.3 BS-ILRMA をランク制約付き空間共分散行列推定法へ適用した場合の性能評価

5.2 節の結果を踏まえ，次に ILRMA，SS-ILRMA，BS-ILRMA のそれぞれを初期化方法としてランク制約付き空間共分散行列推定法を適用した場合の分離性能について調査する．実験データや ILRMA，SS-ILRMA，BS-ILRMA の条件は 5.2 節と同様である．ランク制約付き空間共分散行列推定法における形状母数パラメータ α は 20 とし，尺度母数パラメータ β は 10^{-16} とした．また少ない反復回数で高い分離性能を達成することが分かっているため，ランク制約付き空間共分散行列推定法の反復回数が 2 回目の結果を用いて比較した．

ILRMA, SS-ILRMA, BS-ILRMA と各手法を初期化方法として適用した場合のランク制約付き空間共分散行列推定法の結果を図 5.2 に示す．結果から全ての場合においてランク制約付き空間共分散行列推定法を適用した場合に SDR 改善量が向上している．中でも，BS-ILRMA を初期値とした場合の SDR 改善量が他の場合と比べて高いため，ランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を用いることでより高い分離性能を達成できることが分かる．

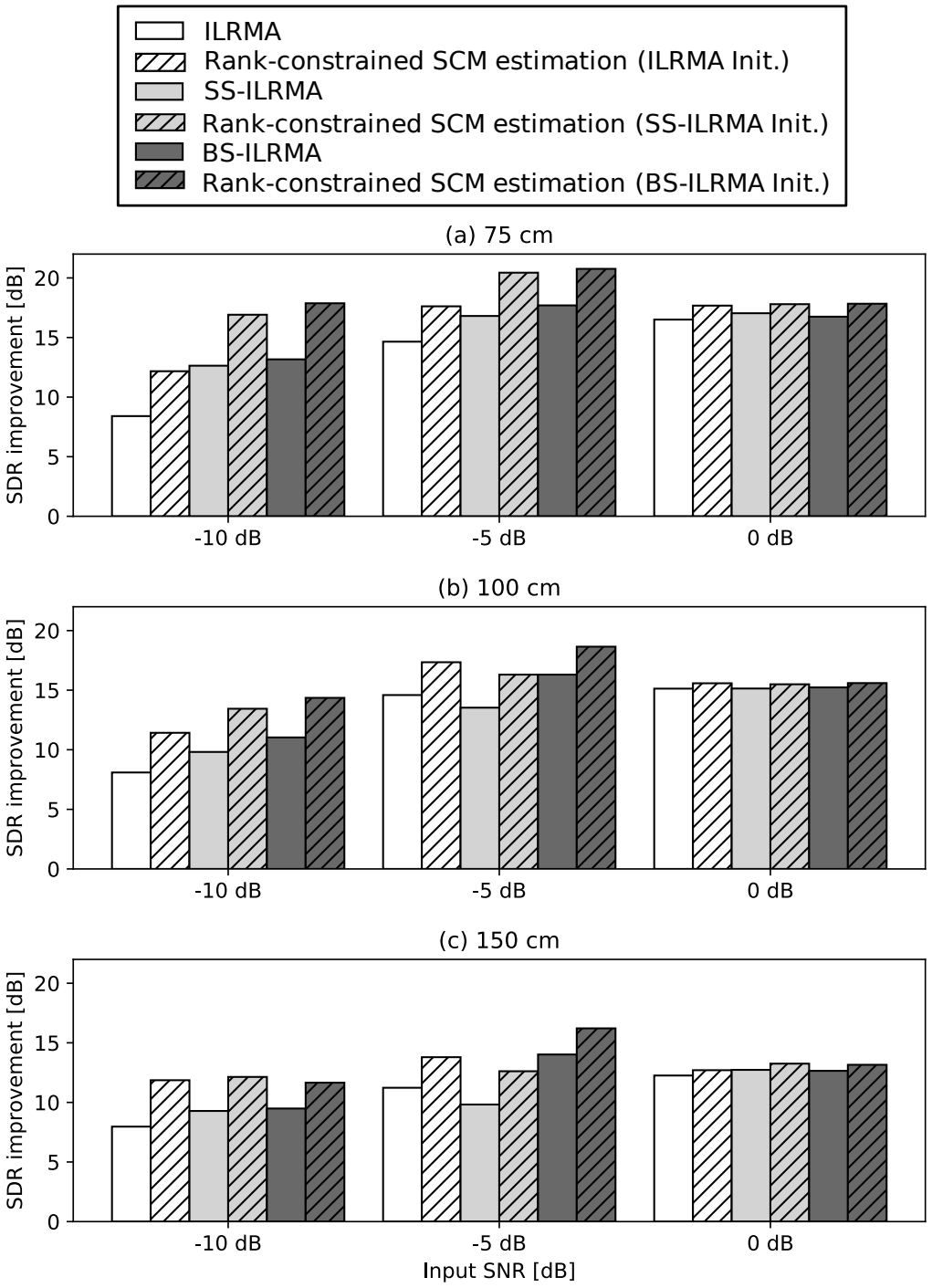


図 5.2: Average SDR improvements of rank-constrained SCM estimation initialized by ILRMA, SS-ILRMA and BS-ILRMA, where number of iterations of rank-constrained SCM estimation was two.

5.4 本章のまとめ

本章では，収録した実環境データに対する，半教師あり音源分離法である BS-ILRMA の有効性について調査した．さらに，ランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を適用した場合の分離性能を調査した．結果として，提案補聴器システムで収録したデータに対して，BS-ILRMA は ILRMA や SS-ILRMA と比較して高い分離性能を達成することを確認した．さらに，ランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を適用した場合も，ILRMA 及び SS-ILRMA で初期化した場合と比較して高い分離性能を達成することを実験的に明らかにした．

第6章　ランク制約付き空間共分散行列推定法の雑音教師ありアプローチへの拡張

6.1　はじめに

本章では，元々ブラインドの枠組みの音源抽出手法であるランク制約付き空間共分散行列推定法を，半教師ありの手法へ拡張する．5 章で述べたように，提案補聴器システムを使うシーンにおいて，雑音情報が事前に得られ，半教師ありアプローチである BS-ILRMA が適用可能である．また，ランク制約付き空間共分散行列推定法の初期化方法に BS-ILRMA を適用し，さらに高品質な音源抽出を達成することを示した．以上から，ランク制約付き空間共分散行列推定法に雑音教師信号を利用した手法に拡張することで，さらに高品質な音源抽出が期待できる．

6.2　半教師ありランク制約付き空間共分散行列推定法

便宜上，教師信号である拡散性雑音のソースイメージを $\breve{\boldsymbol{u}}_{ij'} := \boldsymbol{x}_{ij'}^{(\text{noise})}$ と再定義する．教師信号の雑音の SCM $\breve{\boldsymbol{R}}_i^{(u)}$ は次のように計算できる．

$$\breve{\boldsymbol{R}}_i^{(u)} = \mathbb{E}\left[\breve{\boldsymbol{u}}_{ij'}\breve{\boldsymbol{u}}_{ij'}^{\mathsf{H}}\right] \tag{6.1}$$

$$= \frac{1}{J'}\sum_{j'}\breve{\boldsymbol{u}}_{ij'}\breve{\boldsymbol{u}}_{ij'}^{\mathsf{H}} \tag{6.2}$$

さらに，事前分布として推定する雑音の SCM $\boldsymbol{R}_i^{(u)} \in \mathbb{C}^{M\times M}$ が逆行列ガンマ分布に従うと仮定する．

$$\boldsymbol{R}_i^{(u)} \sim \mathrm{IMG}_M\left(\alpha', \beta', \breve{\boldsymbol{R}}_i^{(u)}\right) \tag{6.3}$$

$$p(\boldsymbol{R}_i^{(u)}) = C(\alpha', \beta', \breve{\boldsymbol{R}}_i^{(u)})|\boldsymbol{R}_i^{(u)}|^{-(\alpha'+M)} \exp\left(-\frac{1}{\beta'}\mathrm{tr}\left((\boldsymbol{R}_i^{(u)})^{-1}\right)\breve{\boldsymbol{R}}_i^{(u)}\right) \tag{6.4}$$

$$C(\alpha', \beta', \breve{\boldsymbol{R}}_i^{(u)}) = \frac{(\breve{\boldsymbol{R}}_i^{(u)})^{\alpha'}}{\beta'^{\alpha' M}\Gamma(\alpha')} \tag{6.5}$$

また，

$$\log p(\boldsymbol{R}_i^{(u)}) = \mathrm{const.} - (\alpha' + M)\log\det \boldsymbol{R}_i^{(u)} - \frac{1}{\beta'}\mathrm{tr}\left((\boldsymbol{R}_i^{(u)})^{-1}\breve{\boldsymbol{R}}_i^{(u)}\right) \tag{6.6}$$

ここで，$\mathrm{IMG}_M(\cdot,\cdot,\cdot)$ は M 次元の逆行列ガンマ分布であり，$\alpha' > (M-1)$ 及び $\beta' > 0$ はそれぞれ逆行列ガンマ分布の形状母数及び尺度母数である．また，const. は $\boldsymbol{R}_i^{(u)}$ に依存しない項である．雑音のソースイメージ $\boldsymbol{u}_{ij}$ は多変量複素ガウス分布であり，$\boldsymbol{u}_{ij}$ から構成される分散共分散行列はウィッシャート分布に従う．さらに，分散共分散行列の共役事前分布に逆ウィッシャート分布が一般的に用いられるため，本論文では逆ウィッシャート分布を一般化した逆行列ガンマ分布を事前分布とした．拡散性雑音の空間共分散行列のモデリングにより完全対数事後分布 $\mathcal{L}(\boldsymbol{x}_{ij}, s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\Theta)$ は次のようにかける．

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x}_{ij}, s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\Theta) &= \log\prod_{i,j} p(\boldsymbol{x}_{ij}, s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\Theta)p(r_{ij}^{(h)})p(\boldsymbol{R}_i^{(u)}) \\
&= \sum_{i,j}\log p(\boldsymbol{x}_{ij}, s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\Theta) + \sum_{i,j}\log p(r_{ij}^{(h)}) + \sum_{i,j}\log p(\boldsymbol{R}_i^{(u)}) \\
&= \sum_{i,j}\Bigg[\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij} + \log\det \boldsymbol{R}_{ij}^{(u)} + (\alpha+1)\log r_{ij}^{(h)} + \frac{\beta}{r_{ij}^{(h)}} \\
&\quad - (\alpha' + M)\log\det \boldsymbol{R}_i^{(u)} - \frac{1}{\beta'}\mathrm{tr}\left((\boldsymbol{R}_i^{(u)})^{-1}\breve{\boldsymbol{R}}_i^{(u)}\right)\Bigg] + \mathrm{const.}
\end{aligned} \tag{6.7}$$

ブラインドのランク制約付き空間共分散行列推定法と同様に EM アルゴリズムに基づき，完

全対数事後分布 $\mathcal{L}(\boldsymbol{x}_{ij}, s_{ij}^{(h)}, \boldsymbol{u}_{ij}|\Theta)$ の期待値をとることで，Q 関数を次のように計算できる．

$$
\begin{aligned}
Q(\Theta;\tilde{\Theta}) = &\sum_{i,j}\left[-(\alpha+2)\log r_{ij}^{(h)} - M\log r_{ij}^{(u)} - \frac{\hat{r}_{ij}^{(h)}+\beta}{r_{ij}^{(h)}}\right] \\
&+ \sum_{i}\left[-(\alpha'+M+J)\log\det \boldsymbol{R}_i^{(u)} - \mathrm{tr}\left\{\left(\boldsymbol{R}_i^{(u)}\right)^{-1}\left(\frac{1}{\beta'}\breve{\boldsymbol{R}}_i^{(u)} + \sum_j \frac{1}{r_{ij}^{(u)}}\hat{\boldsymbol{R}}_{ij}^{(u)}\right)\right\}\right] \\
&+ \mathrm{const.}
\end{aligned} \tag{6.8}
$$

次に，式 (6.8) をそれぞれの変数に関して偏微分し 0 とおいて，Q 関数の最大化を行う．ただし，λ_i に関する最大化は Q 関数に $(\boldsymbol{R}_i)^{-1} = (\boldsymbol{R}_i^{(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}})^{-1}$ や $\log\det \boldsymbol{R}_i^{(u)} = \log\det(\boldsymbol{R}_i^{(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}})$ という項が含まれているため，このままの形では行うことができない．そこで，$(\boldsymbol{R}_i)^{-1}$ を λ_i に関する式として陽に表すための定理 [52] (*claim 1*) を用いることで，偏微分が困難な項を次のように表せる．

$$\log\det \boldsymbol{R}_i^{(u)} = \log\lambda_i + \mathrm{const.} \tag{6.9}$$

$$(\boldsymbol{R}_i^{(u)})^{-1} = (\boldsymbol{R'}_i^{(u)})^{+} + \frac{1}{\lambda_i}\boldsymbol{b}_i\boldsymbol{b}_i^{\mathsf{H}} \tag{6.10}$$

ここで，$\cdot^{+}$ は Moore-Penrose の一般化逆行列を表す．従って，Q 関数の λ_i に関する偏微分は次のように表せる．

$$Q(\lambda_i) = -(\alpha'+M+J)\log\lambda_i - \frac{1}{\lambda_i}\boldsymbol{b}_i^{\mathsf{H}}\left(\frac{1}{\beta'}\breve{\boldsymbol{R}}_i^{(u)} + \sum_j \frac{1}{r_{ij}^{(u)}}\hat{\boldsymbol{R}}_{ij}^{(u)}\right)\boldsymbol{b}_i + \mathrm{const.} \tag{6.11}$$

以上から，式 (6.11) を λ_i に関して偏微分し，0 と置くことで以下の更新式を得る．

$$\lambda_i \leftarrow \frac{1}{\alpha'+M+J}\boldsymbol{b}_i^{\mathsf{H}}\left(\frac{1}{\beta'}\breve{\boldsymbol{R}}_i^{(u)} + \sum_j \frac{1}{r_{ij}^{(u)}}\hat{\boldsymbol{R}}_{ij}^{(u)}\right)\boldsymbol{b}_i \tag{6.12}$$

6.3 本章のまとめ

本章では，元々ブラインドの音源抽出手法であったランク制約付き空間共分散行列推定法を半教師ありの枠組みへ拡張した手法を提案した．事前に収録した雑音情報を利用し，拡散性雑音の空間共分散行列に対して逆行列ガンマ分布を仮定し，新たな更新式を導出した．

第7章　半教師ありランク制約付き空間共分散行列推定法の評価

7.1　はじめに

本章では，半教師ありランク制約付き空間共分散行列推定法の有効性について調査する．また，本手法の内部パラメータである α' 及び β' と分離性能の関係についても調査する．初期化方法に ILRMA，SS-ILRMA 及び BS-ILRMA を使用したブラインドのランク制約付き空間共分散行列推定法と，半教師ありランク制約付き空間共分散行列推定法を比較し，提案手法である後者がより高品質な音声抽出を達成することを示す．

7.2　内部パラメータと音源抽出性能の比較

まず，半教師ありランク制約付き空間共分散行列推定法の内部パラメータ α' 及び β' と分離性能の関係について調査する．ブラインド及び半教師ありランク制約付き空間共分散行列推定法の目的音源に関する内部パラメータ α 及び β はそれぞれ 20 及び 10^{-16} とし，反復回数を 30 回とした．ブラインド及び半教師ありランク制約付き空間共分散行列推定法における目的音源の分散に置いた分布の形状母数パラメータ α は 20 とし，尺度母数パラメータ β は 10^{-16} とした．その他の条件は 4.2 節と同様とした．以上の条件で，α' を 200, 400, 800, β' を 1, 10000 とそれぞれ変化させ，SDR 改善量の関係を調査する．

目的音源の距離が 75 cm での，α' の振る舞いと SDR 改善量の関係に関する結果を図 7.1 に示す．結果から，入力 SNR に対して最適な α' があることがわかる．具体的には入力 SNR が低い場合には小さい α'，入力 SNR が高い場合には大きい α' を設定すると，分離性能が高いことがわかる．また，入力 SNR が低い場合には，iteration に対して頑健な分離性能を達成することがわかる．実環境では，目的音源と拡散性雑音の SNR は不明であることが多いため，α' は 200 程度の大きさに設定することで，比較的安定的な分離性能を達成できる．次に，β' の振る舞いと SDR 改善量の関係に関する結果を図 7.2 に示す．結果から，β' を 1∼10000 の範囲では，SDR 改善量に大きな差は見られないことがわかる．そのため，半教師ありランク

制約付き空間共分散行列推定法を実環境で利用する場合でも，β' の値に関わらず，頑健に音源抽出を行える．

7.3 半教師あり空間共分散行列推定法の音源抽出性能の比較

次に，半教師ありランク制約付き空間共分散行列推定法と従来手法であるブラインドのランク制約付き空間共分散行列推定法との比較を行う．比較手法は ILRMA，SS-ILRMA，及び BS-ILRMA に加え，それぞれをブラインド及び半教師ありランク制約付き空間共分散行列推定法の初期化方法に用いた，計 9 手法の SDR 改善量を比較した．ブラインドのランク制約付き空間共分散行列推定法の iteration は最も SDR 改善量が大きいものを採用した．同様に，半教師ありランク制約付き空間共分散行列推定法のパラメータ α'，β' 及び iteration は，最も SDR 改善量が大きいものを採用して比較を行った．

結果を図 7.3 に示す．結果から，特に入力 SNR が低い場合に半教師ありランク制約付き空間共分散行列推定法の SDR 改善量が最も大きい．これは，教師信号を用いた拡散性雑音のモデリングにより，入力 SNR が大きい中で空間共分散行列の高精度な推定を達成できているためと考えられる．さらに，BS-ILRMA で初期化した半教師ありランク制約付き空間共分散行列推定法の SDR 改善量が最も大きく，雑音情報の利用によってより高精度な音源抽出が可能になることがわかった．

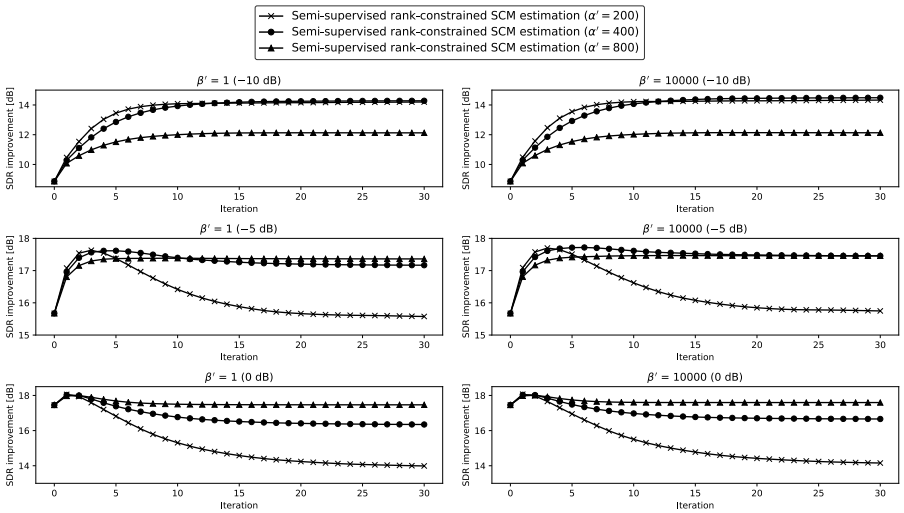


図 7.1: Average SDR improvements of semi-supervised rank-constrained SCM estimation initialized by ILRMA for each update, when distance to target source is set to 75 cm. Three lines ($\alpha' = 200, 400,$ and 800) are plotted in each β' and each input SNR settings.

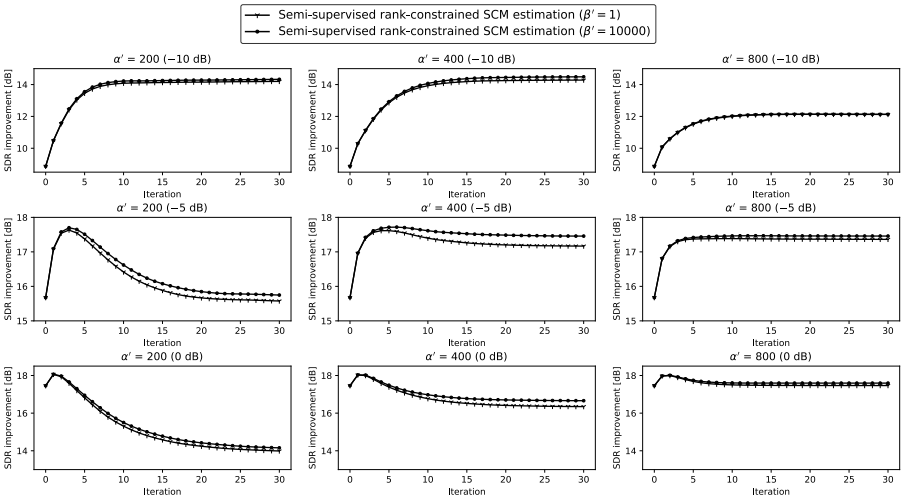


図 7.2: Average SDR improvements of semi-supervised rank-constrained SCM estimation initialized by ILRMA, when distance to target source is set to 75 cm. Two lines ($\beta' = 1$ and 10000) are plotted in each α' and each input SNR settings.

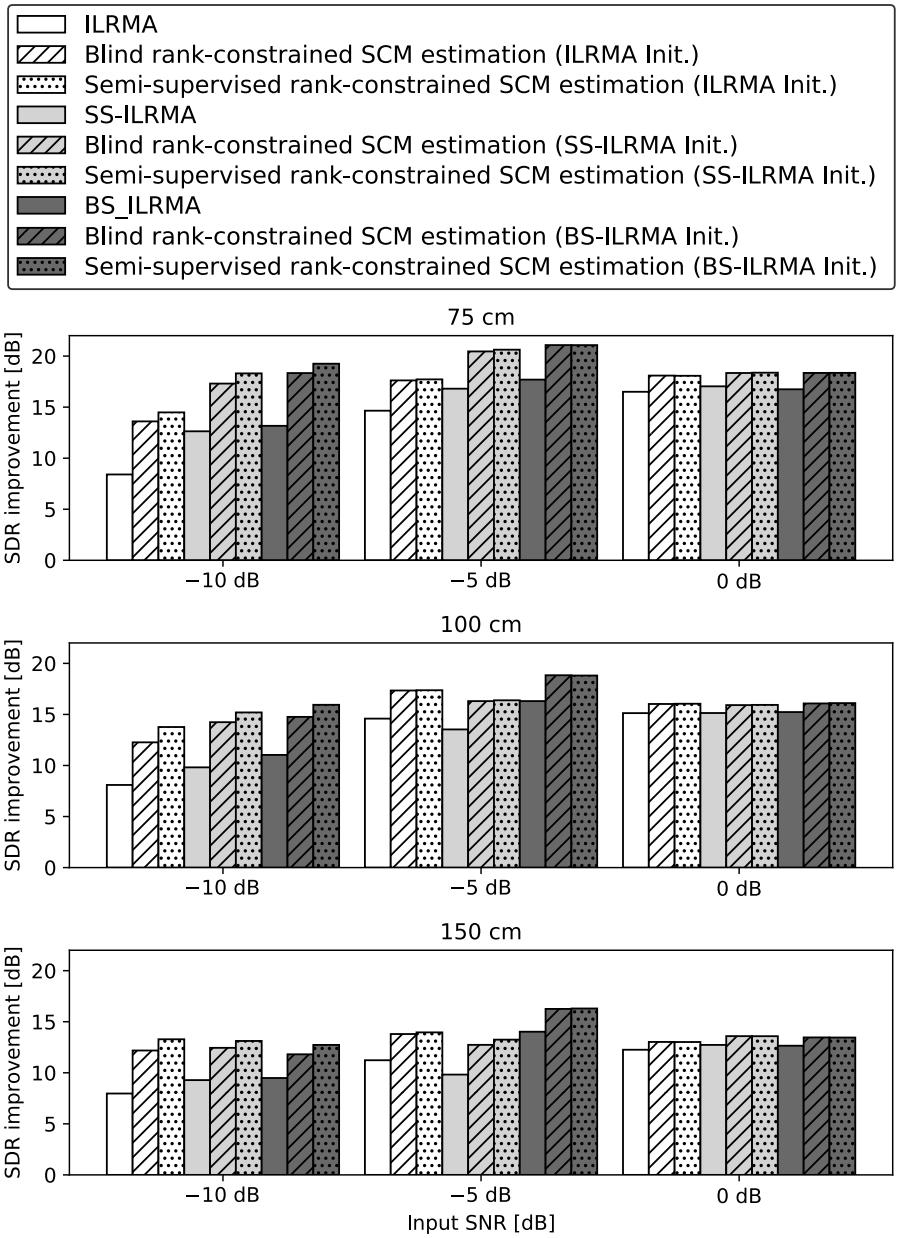


図 7.3: Average SDR improvements of blind/semi-supervised rank-constrained SCM estimation initialized by ILRMA, SS-ILRMA, BS-ILRMA, and each ILRMA. Scores of blind/semi-supervised rank-constrained SCM estimation are the best performance out of 30 iterations.

7.4 本章のまとめ

本章では，半教師ありランク制約付き空間共分散行列推定法について，(1) 内部パラメータと SDR 改善量の関係の調査と，(2) 従来手法との比較実験を行った．(1) の実験結果から，入力 SNR が低い場合には小さい α'，入力 SNR が高い場合には大きい α' を設定すると高い分離性能を達成でき，また，β' は 1〜10000 の広い範囲で設定しても頑健に高い分離性能を達成することがわかった．(2) の実験結果から，半教師ありランク制約付き空間共分散行列推定法は従来手法と比較して高い分離性能を達成し，また初期化方法に BS-ILRMA を採用することでさらに高い分離性能を達成することがわかった．

第8章　結論

本論文では，補聴器システムに焦点を当て，両耳のマイクロホンだけでなくスマートフォンに内蔵されているマイクロホンも含めた分散マイクロホンアレー補聴器システムを新たに提案した．さらに本論文では，元来ブラインドの枠組みであったランク制約付き空間共分散行列推定法を半教師ありアプローチへ拡張した手法を提案した．

第 1 章では，音源分離の必要性及び分散マイクロホンアレーについて述べ，提案補聴器システムを提案する目的を述べた．また，会話直前の数秒の雑音情報を用いて半教師ありアプローチが適用可能であることを述べ，半教師ありランク制約付き空間共分散行列推定法を提案する目的を述べた．

第 2 章では，基本的な BSS の定式化を行い，本論文で取り扱う BSS 手法として ILRMA を，BSE 手法としてランク制約付き空間共分散行列推定法について述べた．また，半教師あり音源分離手法として BS-ILRMA について述べた．

第 3 章では，新たな補聴器システムとして，両耳のマイクロホンだけでなくスマートフォンのマイクロホンを含めた分散マイクロホンアレー補聴器システムを提案し，その動機，目的，及び仕様について述べた．提案補聴器システムに基づいて，データ収録のための装置を構築し，実環境下に基づき，インパルス応答及び拡散性雑音を収録した．

第 4 章では，提案補聴器システムに対して，ランク制約付き空間共分散行列推定法が適用可能であるか評価実験により評価した．併せて，ランク制約付き空間共分散行列推定法の内部パラメータと分離性能の関係を実験的に明らかにした．さらに．提案補聴器システムによりもたらされた，マイク総数の増加及び目的音源に近い位置の空間情報が利用可能という 2 点が優位に働いているかを実験により評価した．

第 5 章では半教師あり音源分離手法である BS-ILRMA の提案補聴器システムに対する有効性を示した．その後，BS-ILRMA をランク制約付き空間共分散行列推定法の初期化方法に用い，ILRMA を初期化に用いた場合と比較して，さらに高品質な分離を達成することを示した．

第 6 章では，ブラインドのランク制約付き空間共分散行列推定法に対して，拡散性雑音の空間共分散行列にモデル新たに導入することで，半教師ありアプローチへ拡張した手法を提案した．

第 7 章では，半教師ありアプローチへ拡張したランク制約付き制約付き空間共分散行列推定法の，内部パラメータと分離性能の関係及びブラインドのランク制約付き制約付き空間共分散行列推定法に対する優位性を調査し，高い分離性能を達成できる内部パラメータと従来手法と比較して高い分離性能を達成できることを示した．

最後に，実用に向けた今後の課題を述べる．まず，リアルタイム処理である．現在，ILRMA やランク制約付き空間共分散行列推定法を用いて，オフラインで音源抽出を行っている．補聴器を利用するシーンにおいては，これらの手法をオンラインで処理する必要があるが，現状 ILRMA やランク制約付き空間共分散行列推定法をオンライン化した手法は知られていない．次に，マイクロホンの同期の問題である．現在は，A-D 変換器を用いて各マイクロホンを同期しているが，実際の利用シーンでは規模やコストの面で持ち運びは困難である．一方で，非同期のマイクロホンに対して同期を行う手法も提案されているが，これらの手法を利用しつつ，既存手法及び本論文で提案した手法が高品質な分離が行えるかを実験により明らかにする必要がある．

謝辞

本論文は，筑波大学大学院システム情報工学研究群情報理工学位プログラムマルチメディア研究室において，著者が修士課程の 2 年間で行った研究に基づくものです．本論文を執筆するにあたり，多くの方々にお世話になりました．この場を借りて，感謝の意を表します．

まずはじめに，システム情報系の牧野昭二教授には，本研究に携わる機会から学会発表できる成果を出せるまで，数多くのご指導やご支援を賜りました．いくつもの国内及び国際発表に参加させていただいたことで，幅広い知識や知見を得ることができました．心より感謝申し上げます．

システム情報系の山田武志准教授には，研究全般に関するご助言を賜りました．ゼミでの様々な方向からのご質問，議論により自身の研究をより良いものへ進めることができました．また，研究発表の練習では，図表の使い方，話し方関して的確なご指摘を頂き，より明瞭な発表を行うことができました．心より御礼申し上げます．

東京大学情報理工学系研究科システム情報系の猿渡洋教授には，他大学の学生にも関わらず，非常に興味深い今回の研究に携わらせていただきました．研究の方針から，論文の添削，発表資料の修正など，猿渡先生のご指導，ご助言がなければ，4 件の対外発表を行えるほどの成果をあげることはできませんでした．深く御礼申し上げます．

香川高等専門学校電気情報工学科の北村大地助教には，独立低ランク行列分析に関して通して大変多くのご助言，ご指導を賜りました．東京大学大学院学術支援専門職員である高宗典玄氏には幅広い数学知識と実装技術に至るまで多くのご指導，ご指摘を頂きました．

久保優騎氏には研究を進めるための様々なご指導，アドバイスを頂けました．無学な私に対しても，優しく分かりやすく説明してくださり，最後までモチベーションを失うことなく研究を続けることができました．ご卒業されてお忙しい中でも，時間を割いて研究以外の様々な相談に乗って頂きました．厚く厚く御礼申し上げます．

研究室 OB である高草木萌氏には，基底共有型独立低ランク行列分析の実装に関する様々なご支援を頂きました．マルチメディア研究室の皆様には，研究活動を通して様々なご協力を頂きました．マルチメディア研究室の松本光雄研究員には，短い期間でしたが，研究全般に関して大変お世話になりました．特に，データ収録のノウハウや機器の接続に関して様々

なご助言を頂きました．心より感謝申し上げます. 秘書の田口朱美氏には，学会の参加や雇用関連の様々な事務手続きで手厚いご支援を頂きました．音響信号処理グループの，博士後期過程の李莉氏，博士前期過程 2 年の井上翔太氏，髙橋理希氏，村島允也氏には，楽しく充実した生活が送れたこと，心から感謝申し上げます. 最後に，家族や友人など多くの方々にご支援を頂きました．研究を支えてくださったすべての方々に，この場を借りて深く感謝申し上げます.

参考文献

[1] F. Mustière, M. Bouchard, H. Najaf-Zadeh, R. Pichevar, L. Thibault, and H. Saruwatari, "Design of multichannel frequency domain statistical-based enhancement systems preserving spatial cues via spectral distances minimization," *Signal Processing*, vol. 93, pp. 321–325, 2013.

[2] 高橋祐, 猿渡洋, and 鹿野清宏, "独立成分分析を導入した空間的サブトラクションアレーによるハンズフリー音声認識システムの開発," 電子情報通信学会論文誌 *D*, vol. 93, no. 3, pp. 312–325, 2010.

[3] A. Waibel, M. Bett, M. Finke, and R. Stiefelhagen, "Meeting browser: tracking and summarizing meetings," *Proceedings of DARPA Broadcast News Transcription and Understanding Workshop*, pp. 281–286, 1998.

[4] S. Renals, T. Hain, and H. Bourlard, "Recognition and understanding of meetings the AMI and AMIDA projects," *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 238–247, 2007.

[5] F. Asano, K. Yamamoto, J. Ogata, M. Yamada, and M. Nakamura, "Detection and separation of speech events in meeting recordings using a microphone array," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, pp. 1–8, 2007.

[6] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," *Proc. Symposium on Communications and Vehicular Technology (SCVT)*, 2011.

[7] R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio processing," *Proc. ICASSP*, pp. 840–843, 2003.

[8] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal of Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.

[9] A. Brutti, M. Omologo, and P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays," *Proc. Interspeech*, pp. 2337–2340, 2005.

[10] Z. Liu, Z. Zhang, L. He, and P. Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," *Proc. ICASSP*, pp. 761–764, 2007.

[11] E. Robledo-Arnuncio, T. S. Wada, and B. H. Juang, "On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation," *Proc. WASPAA*, pp. 34–37, 2007.

[12] "Impulsive Paradigm Change through Distributed Technologies Program (ImPACT)," http://www.jst.go.jp/impact/program07.html.

[13] H. Namari, K. Wakana, M. Ishikura, M. Konyo, and S. Tadokoro, "Tube-type active scope camera with high mobility and practical functionality," *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3679–3686, 2012.

[14] M. Takakusaki, D. Kitamura, N. Ono, T. Yamada, S. Makino, and H. Saruwatari, "Ego-noise reduction for a hose-shaped rescue robot using basis shared semi-supervised independent low-rank matrix analysis," in *Proc. NCSP*, 2018, pp. 351–354.

[15] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series with engineering applications*, Cambridge, MA: MIT Press, 1949.

[16] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[17] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[18] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[19] 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 and Information Processing*, vol. 8, no. e12, pp. 1–14, 2019.

[20] 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.

[21] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[22] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 8, pp. 1129–1159, 1995.

[23] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 757–763, 1996.

[24] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.

[25] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, pp. 365–371, 1999.

[26] 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. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[27] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. of ICA*, 2006, pp. 601–608.

[28] T. Kim, H. T.Attias, S. Y. Lee, and T. W. Lee, "Blind source separation exploiting higher order frequency dependencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.

[29] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.

[30] 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. on ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

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

[32] 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 Journal on Advances in Signal Processing*, vol. 2003, no. 11, pp. 1–10, 2003.

[33] Z. Koldovskỳ and P. Tichavsỳ, "Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence," *IEEE Trans- actions on Signal Processing*, vol. 67, no. 4, pp. 1050–1064, 2019.

[34] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[35] H. Y. Kim, F. Asano, Y. Suzuki, and T. Sone, "Speech enhancement based on short-time spectral amplitude estimation with two-channel beamformer," *IEICE Transactions on Fundamentals*, vol. 79, no. 12, pp. 2151–2158, 1996.

[36] M. Mizumachi and M. Akagi, "Noise reduction by paired-microphones using spectral subtraction," *Proceedings of IEEE International Conference on Acoustics Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 1001–1004, 1998.

[37] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on noise adaptive complementary beamforming," *IEICE Transactions on Fundamentals*, vol. 83, no. 5, pp. 866–876, 2000.

[38] J. Meyer and K. U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 1167–1170, 1997.

[39] I. A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 709–716, 2003.

[40] 猿渡洋, "最近の音声処理に用いられるマイクロホンアレー技術," 日本音響学会誌, vol. 66, no. 10, pp. 521–526, 2010.

[41] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.

[42] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.

[43] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex valued data," *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.

[44] N. Ito and T. Nakatani, "FastMNMF: joint diagonalization based accelerated algorithms for multichannel nonngetive matrix factorization," in *Proc. ICASSP*, 2019, pp. 371–375.

[45] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices," *Proceedings of The European Signal Processing Conference (EUSIPCO)*, p. 5, 2019.

[46] J. R. Hershey, Z. Chen, J. L. Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," *Proceedings of IEEE Inter- national Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 31–35, 2016.

[47] M. Kolbæk, D. Yu, Z.-H. Tan, and J. Jensen, "Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks," *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 1901–1913, 2017.

[48] K. Kinoshita, L. Drude, M. Delcroix, and T. Nakatani, "Listening to each speaker one by one with recurrent selective hearing networks," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 5064–5068, 2018.

[49] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1652–1664, 2016.

[50] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, "Independent deeply learned matrix analysis for deter- mined audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 10, pp. 1601–1615, 2019.

[51] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Efficient full-rank spatial covariance estimation using independent low-rank matrix analysis for blind source separation," in *Proc. EUSIPCO*, 2019.

[52] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Blind speech extraction based on rank-constrained spatial covariance matrix estimation with multivariate generalized gaussian distribution," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1948–1963, 2020,.

[53] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. ASLP*, vol. 17, no. 4, pp. 650–664, 2009.

[54] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *Proc. ICA*, 2007, pp. 414–421.

[55] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, "Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties," in *IEICE Trans. Fundamentals*, 2014, vol. E97-A, pp. 1113–1118.

[56] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009, pp. 161–164.

[57] T. Ono, S. Miyabe, N. Ono, and S. Sagayama, "Blind source separation with distributed microphone pairs using permutation correction by intra-pair TDOA clustering," in *Proc. IWAENC*, 2010.

[58] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," in *Proc. ICASSP*, 2013, pp. 674–678.

[59] S. Miyabe, N. Ono, and S. Makino, "Optimizing frame analysis with non-integer shift for sampling mismatch compensation of long recording," in *Proc. WASPAA*, 2013.

[60] H. Chiba, N. Ono, S. Miyabe, Y. Takahashi, T. Yamada, and S. Makino, "Amplitude-based speech enhancement with nonnegative matrix factorization for asynchronous distributed recording," in *Proc. IWAENC*, 2014, pp. 204–208.

[61] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Elsevier Signal Processing*, vol. 107, pp. 185–196, 2015.

[62] K. Ochi, N. Ono, S. Miyabe, and S. Makino, "Multi-talker speech recognition based on blind source separation with ad hoc microphone array using smartphones and cloud storage," in *Proc. Interspeech*, 2016, pp. 3369–3373.

[63] T.-K. Le and N. Ono, "Closed-form and near closed-form solutions for TOA-based joint source and sensor localization," *IEEE Trans. Signal Processing*, vol. 64, no. 18, pp. 4751–4766, 2016.

[64] T.-K. Le and N. Ono, "Closed-form and near closed-form solutions for TDOA-based joint source and sensor localization," *IEEE Trans. Signal Processing*, vol. 65, no. 5, pp. 1207–1221, 2017.

[65] K. Imoto and N. Ono, "Spatial cepstrum as a spatial feature using distributed microphone array for acoustic scene analysis," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 25, no. 6, pp. 1335–1343, 2017.

[66] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proceedings of International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2010, pp. 165–172.

[67] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proceedings of International Congress on Acoustics (ICA)*, 1968, pp. C–17–C–20.

[68] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," in *Journal of Royal Statistical Society, Series B Statistical Methodology*, 1977, vol. 39, pp. 1–38.

[69] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," in *Neurocomputing*, 2001, vol. 41, pp. 1–24.

[70] R. Sakanashi, N. Ono, and S. Miyabe, "Speech enhancement with ad-hoc microphone array using single source activity," in *Proc. APSIPA2013*, 2015, pp. 1–6.

[71] Y. Suzuki, F. Asano, H. Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *Journal of the Acoustical Society of America*, vol. 65, pp. 1484–1488, 1995.

[72] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *The Journal of Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 199–206, 1999.

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

著者研究発表

査読付国際会議

[C1] M. Une, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, and S. Makino, “Evaluation of multichannel hearing aid system using rank-constrained spatial covariance matrix estimation,” in Proc. *APSIPA*, pp. 1874–1879, November 2019.

[C2] M. Une, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, and S. Makino, “Multichannel hearing-aid system based on basis-shared semi-supervised independent low-rank matrix analysis,” in Proc. *Forum Acusticum*, pp. 763–769, December 2020.

国内会議

[D1] 宇根昌和, 久保優騎, 高宗典玄, 北村大地, 猿渡洋, 牧野昭二, “ランク制約付き空間共分散モデル推定を用いた多チャネル補聴器システム,” 日本音響学会講演論文集, 1-1-3, pp. 161–164, September 2019.

[D2] 宇根昌和, 久保優騎, 高宗典玄, 北村大地, 猿渡洋, 牧野昭二, “基底共有型半教師あり独立低ランク行列分析に基づく多チャネル補聴器システム,” 日本音響学会講演論文集, 1-1-22, pp. 217–220, March 2020.