人工知能学会第二種研究会資料
AIチャレンジ研究会
SIG-Challenge-058-05

複数マイクロホンアレイを用いた NMFによる空間音源分離法の残響下での評価

Evaluation of spatial source separation using NMF with multiple microphone arrays under reverberation

鍵本泰宏[1*]　糸山克寿[1]　西田健次[1]　中臺一博[1,2]
Yasuhiro Kagimoto[1]　Katsutoshi Itoyama[1]　Kenji Nishida[1]　Kazuhiro Nakadai[1,2]

[1] 東京工業大学
[1] Tokyo Institute of Technology
[2] （株）ホンダ・リサーチ・インスティチュート・ジャパン
[2] Honda Research Institute Japan, Co., Ltd.

Abstract: 本稿では，複数台のマイクロホンアレイを用いた空間音源分離法について述べる．音源分離は，様々な音源やノイズが混在する中から所望の音源だけを抽出する技術である．音源分離の代表的な手法の一つであるビームフォーミングは，マイクロホンアレイと呼ばれる多チャンネルデバイスで収録した信号から，チャンネル間に生じる位相差に基づいて方向ごとに音源を分離することができる．しかし，ビームフォーミングは方向に基づく手法であり，同方向に複数の音源が存在する場合それらを分離することができないという課題があった．そこで，提案手法は複数台のマイクロホンアレイを用いた，音源の位置に基づく分離を行う．目的音源に対して複数マイクロホンアレイでビームフォーミングし，得られた分離音から非負値行列因子分解（Non-negative Matrix Factorization）によって目的音源だけを抽出することで，同方向に存在する別音源の影響を緩和する．提案手法は，シミュレーションにより遅延和法に比べSDR（Source to Distortion Ratio）がおよそ0.8〜2.3dB向上することがわかった．また，実環境での分離性能を評価し，残響の影響について検証した．

1　はじめに

近年，スマートホンやAIスピーカーの普及に伴い，様々な場面で音響処理技術が利用されるようになってきた．例えば，Apple社のSiriに話しかければ，音声による機器操作や検索，文字起こし等を行える．このようなアプリケーションを使う際に問題となるのが雑音である．実環境では他の音源やノイズが混在しているため，処理精度が低下してしまう．そのため，様々な音源の混合音から目的音源だけを分離する音源分離技術は，音響処理を行う上で重要性を増してきている．音源分離の代表的な手法として，ビームフォーミング（Beamforming，BF）が挙げられる．ビームフォーミングではマイクロホンアレイという多チャンネルデバイスで用い，収録信号のチャンネル間位相差に基づいて音源方向に指向領域を形成する．これにより各音源方向に対してビームフォーミングを適用することで各方向の分離音を得ることができる．しかし，ビームフォーミングは方向に基づく手法であり，同方向に複数の音源が存在する場合，それらを分離することができない．

このような課題に対処するため，本研究では複数のマイクロホンアレイを用いた位置に基づく音源分離手法を提案する．この手法は，音源を方向ではなく，位置に基づいて分離することにより同方向音源の影響を低減する．提案手法の処理は大きく2つのステップで構成されている．まず，図1のように，どのマイクロホンアレイから見ても目的音源S_0方向に別音源が存在するような状況を想定する．一つ目のステップでは，マイクロホンアレイを複数箇所に配置し，それぞれの位置から目的音源方向に対してBFを行う．この時，各マイクロホンアレイの指向領域が重なる部分はスポットと呼ばれ，スポット領域内に存在する音源を抽出することをスポットフォーミング（Spotforming，SF）と呼ぶ．各マイクロホンアレイでBFして得られた目的音源方向の分離音（以下，BF分離音と記す）には，目的音源と目的音源方向に存在する他音源が支配的になって分離される．目的音源は全てのマイクアレイで分離されるのに対し，その他の音源は一部のマイクアレイでしか分離されないと考えられる．そこで，2つ目のス

*連絡先：東京工業大学工学院システム制御系
〒 152-8552 東京都目黒区大岡山 2-12-1 W8W-W310
E-mail: kagimoto@ra.sc.e.titech.ac.jp

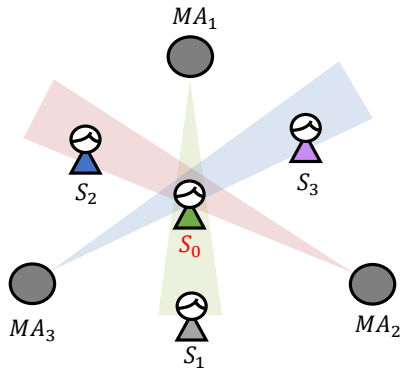


図 1: スポットフォーミングのイメージ．各マイクロホンアレイでビームフォーミングしたときに指向方向が重なる部分をスポット領域とし，スポット内の音源だけを分離する．

テップでは，ブラインド音源分離などで用いられる非負値行列因子分解（Non-negative Matrix Factorization，NMF）[1] を利用して，BF 分離音に共通する音響成分を取り出すことで目的音源を抽出する．ここで，NMF は入力スペクトログラムを構成する代表的な周波数スペクトルを学習し，低ランク近似する機械学習手法の一つである．各処理の詳細については第 3 節で説明する．

本稿では，提案手法について 4 人の音源が同時発話した場合を想定し，提案手法の性能を評価した．シミュレーション環境下で遅延和法に比べ SDR（Source to Distortion Ratio）がおよそ 0.8〜2.3dB 向上することが分かった．また，実環境下での残響の影響についての検証を第 4 節で行い，まとめと今後の課題について第 5 節で述べた．

2　関連研究

前述のとおり，従来のビームフォーミングは同方向に別の音源が存在する場合，それらを分離することが難しいという課題があった．この問題に対処するために，いくつかの手法が提案されている．

関口ら [2] は，複数マイクロホンアレイの配置最適化によるアプローチを提案している．マイクロホンアレイをロボットに搭載し，音源方向が重ならない位置に移動させることで，効率的に音源分離を行うことができる．ただし，このアプローチはロボットが移動するための空間的制約が大きく，使用できる状況が限定的である．

他のアプローチとして，スポットフォーミング法が提案されている．ビームフォーミングは特定方向に存在する音源を強調するのに対し，スポットフォーミングはスポットと呼ばれる領域を音源位置に形成し，スポット内の音源を抽出する手法である．鈴木ら [3] は，二本の超指向性マイクロホンの指向領域が重なる部分をスポットと捉え，収録信号を足し合わせることでスポット内の音源を強調する手法を提案している．指向性マイクロホンは動作が軽量で計算コストが低いという利点がある一方，指向性が一方向に固定されており，スポット位置を変えるためにマイクロホンを動かす必要があるという課題がある．Taseska ら [4] は，超指向性マイクロホンではなく，複数台のマイクロホンアレイを用いた手法を提案している．マイクロホンアレイは信号処理的に指向性を形成するため，任意位置の音を強調することができる．この手法は，分散させたマイクロホンアレイ全体を一つのマイクロホンアレイとみなし，スポット内の音源を強調する空間フィルタの設計を行う．しかし，マイクロホンアレイ間での厳密な同期や，マイクロホンアレイの配置を把握していることが前提となっており，実環境でこのような大規模なシステムを構築する場合配線コストが高くなるという課題がある．

これらを踏まえ，本研究では（1）固定化された複数台のマイクロホンアレイで運用可能である，(2）マイクロホンアレイ間で厳密に同期がされていなくても使える，という二点を重視した手法の構築を目指す．

我々は以前，複数台のマイクロホンアレイでビームフォーミングして得られた分離音のスペクトログラムを，k-means でクラスタリングすることでスポット内音源の分離を行えないか検討した [5]．この手法は，チャンネル間の位相差を用いる信号処理的な工程は各マイクロホンアレイ内で留め，マイクロホンアレイ間の処理は周波数成分のクラスタリングで行っている，これにより，マイクロホンアレイ間の厳密な同期なしに分離を行うことができる．しかし，この手法は空間的に音源を分離できる反面，クラスタリングを行う際に各フレームのスペクトルを一つのデータとして固定化してしまうため，複数音源が同時に発話する場合に分離ができないという欠点がある．そこで我々は，ブラインド音源分離などで用いられる NMF（Non-negative Matrix Factorization）を用いた手法を提案した [6]．NMF は入力される振幅スペクトルを複数音源の混合モデルとして仮定するため，時間的に音源が重なった場合でも分離が可能である．また，本手法ではマイクロホンアレイ間でフレーム単位の緩い同期が必要であるが，位相差などを使った処理は各マイクロホンアレイのビームフォーミング処理に留めているため，サンプル単位の同期を行わなくてもよいというメリットがある．提案手法の詳細については，次の第 3 節で述べる．

3 手法

本節では，提案手法の処理について説明する．全体の処理の流れは図 2 のようになっており，(1) 複数台のマイクロホンアレイによるビームフォーミング，(2) NMF による共通成分抽出，の 2 部で構成されている．以下では，それぞれの工程について説明する．

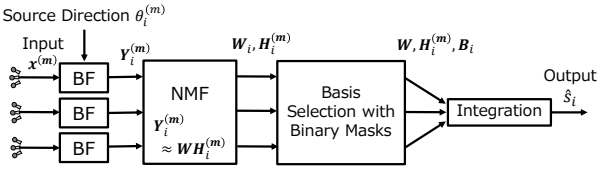


図 2: 提案手法の全体フロー図

3.1 複数台のマイクロホンアレイによる BF

まず，図 1 のような状況を考える．マイクロホンアレイは M 個存在し，各マイクロホンアレイでビームフォーミングを行う．m 番目のマイクロホンアレイで収録された信号を $\boldsymbol{x}^{(m)}(t) = [x_1^{(m)}(t), \ldots, x_N^{(m)}(t)]^\top \in \mathbb{R}^N, t = 1, \ldots, T$ とする．N, T はそれぞれチャンネル数，サンプル数を表している．収録信号に対して短時間フーリエ変換（short-time Fourier transform，STFT）を適用すると，$\boldsymbol{X}^{(m)}(f,\tau) \in \mathbb{C}^N$, $f = 1, \ldots, F$, $\tau = 1, \ldots, T_f$ に変換される．ここで，F および T_f はそれぞれ周波数ビン数とフレーム数である．また，マイクロホンアレイ m からみた目的音源 S_i の方向 $\theta_i^{(m)}$ に対してビームフォーミングすることで得られた分離音（以下 BF 分離音と記述）を $Y_i^{(m)}(f,\tau) \in \mathbb{C}$ とすると，$Y_i^{(m)}(f,\tau)$ は線形フィルタ $\boldsymbol{W}_i^{(m)}(f,\tau) \in \mathbb{C}^{1\times N}$ を用いて次のように得ることができる．

$$Y_i^{(m)}(f,\tau) = \boldsymbol{W}_i^{(m)}(f,\tau)\boldsymbol{X}^{(m)}(f,\tau) \tag{1}$$

ビームフォーミングではこの $\boldsymbol{W}_i^{(m)}(f,\tau)$ を推定することで目的音源方向だけを強調することができる．提案法では，ロボット聴覚用オープンソースソフトウェア HARK（Honda Research Institute Japan Audition for Robots with Kyoto University）[9] に含まれる GHDSS（Geometric High-order Dicorrelation-based Source Separation）[8] を使用する．GHDSS は音源信号間の高次無相関化を行うブラインド音源分離と空間的指向性を形成する BF とのハイブリッド手法である．

BF によって得られた分離音 $Y_i^{(m)}(f,\tau)$ には，目的音源 S_i と目的音源方向に存在するその他の音源が支配的になっていると考えられるため，すべての BF 分離音に共通する音響成分は目的音源の可能性が高い．そこで，すべてのマイクロホンアレイにおける分離音に共通する成分だけを抜き出すことで，スポット内に存在する目的音源を抽出する．次節ではこの考え方に基づき，BF 分離音に共通する音響特徴成分の抽出する方法について説明する．

3.2 NMF による共通成分抽出

本節では，BF 分離音 $Y_i^{(m)}(f,\tau)$ から共通成分を抽出する処理について説明する．NMF は振幅スペクトルがいくつかの基底となるスペクトルの重ね合わせで表現できると考え，非負制約下で低ランク近似を行う．

全 BF 分離音に共通する音響成分を推定するために，各マイクロホンアレイで得られた振幅スペクトログラムを $\boldsymbol{Y}_i^{(m)} \in \mathbb{R}_+^{F\times T_f}$，時間軸方向に結合したものを $\boldsymbol{Y}_i = [\boldsymbol{Y}_i^{(1)}, \ldots, \boldsymbol{Y}_i^{(M)}] \in \mathbb{R}_+^{F\times MT_f}$ とする．この $\boldsymbol{Y}_i$ に対して NMF を適用すると，$\boldsymbol{Y}_i$ を基底行列 $\boldsymbol{W_i}$ とアクティベーション行列 $\boldsymbol{H}_i$ の積に分解できる．

$$\boldsymbol{Y}_i \approx \boldsymbol{V}_i\boldsymbol{H}_i \tag{2}$$

ここで，$\boldsymbol{W}_i \in \mathbb{R}_+^{F\times K}$, $\boldsymbol{H}_i \in \mathbb{R}_+^{K\times MT_f}$ であり，K はあらかじめ決める基底数である．基底行列 $\boldsymbol{V}_i$ は K 種類の基底スペクトルを格納しており，$\boldsymbol{H}_i$ は各フレームにおける基底の強度を表している．NMF は，$\boldsymbol{Y}_i$ と $\boldsymbol{V}_i\boldsymbol{H}_i$ の誤差が小さくなるように基底行列とアクティベーション行列を学習する．学習法としては平均二乗誤差を目的関数とし，乗法更新アルゴリズム [1] を適用する．

続いて，$\boldsymbol{H}_i = [\boldsymbol{H}_i^{(1)}, \ldots, \boldsymbol{H}_i^{(M)}]$ のように $\boldsymbol{H}_i$ を時間フレーム方向に M 個に分割する．この分割により，式 (2) は次のように分解できる．

$$\begin{aligned} \boldsymbol{Y}_i^{(1)} &\approx \boldsymbol{V}_i\boldsymbol{H}_i^{(1)} \\ &\vdots \\ \boldsymbol{Y}_i^{(M)} &\approx \boldsymbol{V}_i\boldsymbol{H}_i^{(M)} \end{aligned} \tag{3}$$

式 (3) の各式は NMF による $\boldsymbol{Y}_i^{(m)}$ の分解になっている．基底行列 $\boldsymbol{V}_i$ はすべてのマイクロホンアレイに共通し，アクティベーション行列 $\boldsymbol{H}_i^{(m)}$ はマイクロホンアレイごとに得られる．これにより全てのマイクロホンアレイのスペクトログラムを通して学習された基底行列 $\boldsymbol{W}_i$ を得ることができる．

NMF では基底と音源の対応づけがされないため，目的音源に対応する基底を各フレームごとに指定する必要がある．本手法では，「各マイクロホンアレイの BF 分離音において，同時刻に出てくる共通の基底（音響成分）は目的音源である」という仮定を置く．類似する周波数構造を持つスペクトルは同じ基底で表される可能性が高くなる．例えば，目的音源はすべてのチャンネルに存在するため，音量の違いがあったとしても同じ基底が同時刻に推定されやすい．この仮定に基づき，基底の出現時刻がどの分離音でも一致している基

底は目的音源に対応すると考える．アクティベーション行列に対するバイナリマスク行列 $\boldsymbol{B}_i$ を次のように作成することで目的音源の基底を抽出する．

$$b_{i,k\tau} = \begin{cases} 1 & (\min(h^{(1)}_{i,k\tau}, h^{(2)}_{i,k\tau}, \cdots, h^{(M)}_{i,k\tau}) > \gamma) \\ 0 & (\text{otherwise}) \end{cases} \quad (4)$$

$h^{(m)}_{i,k\tau}$, $b_{i,k\tau}$ はそれぞれ各基底 k，各時間フレーム τ におけるアクティベーション行列 $\boldsymbol{H}^{(m)}_i$ とバイナリマスク行列 $\boldsymbol{B}_i$ の要素を表している．また，γ は各基底が稼働しているかどうかを決めるための閾値である．式 (4) では，各時刻，各基底ごとに最小値が閾値より大きい時は 1 に，小さい時は 0 にする．この処理はフレームごとに行われるため，フレーム単位での同期が必要である．ただし，ビームフォーミングの出力は同期されているとは限らないため，フレーム単位ではなく数フレーム単位に区切って判定を行うことで時間ずれの影響を緩和できる．また，時間ずれが大きい場合は，自身で時間を合わせる必要がある．この処理を各基底，各フレームごとに行うことでバイナリマスクを得ることができる．

得られたバイナリマスクによって推定信号のスペクトログラム $\hat{\boldsymbol{S}}^{(m)}_i$ は次のように計算できる．

$$\hat{\boldsymbol{S}}^{(m)}_i = \boldsymbol{V}_i(\boldsymbol{H}^{(m)}_i \odot \boldsymbol{B}_i) \quad (5)$$

ここで，$\odot$ は行列の要素積を表す．マイクロホンアレイごとに得られた推定信号に対して逆 STFT した後，時間ずれを相関関数を用いて補正し平均化することで，最終的な出力 $\hat{s}_i$ とした．

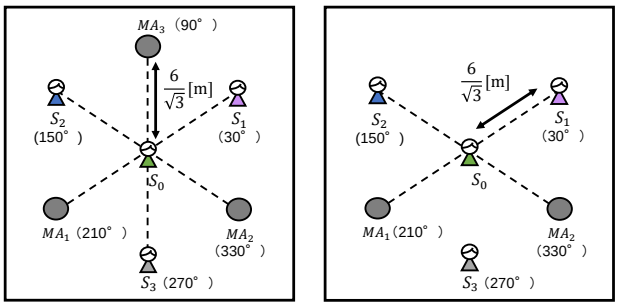
4 実験

本節では，提案手法についてシミュレーションと実環境での実験を行い，分離性能の評価を行う．

4.1 シミュレーション実験

シミュレーションには PyRoomAcoustics[1] という室内音響シミュレーションツールを利用した．図 3 のような 10m×10m の部屋を作成し，音源とマイクロホンアレイを配置した．マイクロホンアレイは円形，8ch，半径 3.65cm とした．また，残響時間 RT60 を 0s，0.3s，0.7s，1.1s に設定して，残響込みのシミュレーションを行った．音源は JNAS の新聞記事読み上げコーパス [7] から 7–10s 程度の 4 つの音源組を 20 パターン用意した．STFT の窓関数は Hamming 窓を用い，窓長 512，シフト幅 256 とし，サンプリング周波数は 16kHz とした．

[1] https://github.com/LCAV/pyroomacoustics



(a) three microphone arrays (b) two microphone arrays

図 3: シミュレーションの設定

表 1: 残響時間（RT60）の変化に伴う音源 S_0 のシミュレーション結果（SDR 改善値 [dB]）

method	アレイ数	0s	0.3s	0.7s	1.0s
NMF–SF	3	**14.3**	**11.3**	**9.7**	**9.1**
Delay–Sum	3	12.0	10.3	8.9	8.3
NMF–SF	2	12.6	9.7	8.2	7.7
Delay–Sum	2	10.0	9.1	7.7	7.1

ビームフォーミングには，ロボット聴覚用オープンソースソフトウェア HARK[9] に含まれる GHDSS [8] ノードを用いて計算した．音源方向は既知とし，GHDSS の入力に方向情報を与えた．NMF の基底数は 100 で統一し，アクティベーション行列の閾値 γ は 1.6×10^{-3} とした．

評価指標として，BssEval [10] に含まれる Source to Distortion Ratio（SDR）を使用する．SDR は推定信号に含まれる目的音源成分と目的音源以外のノイズ成分のパワー比によって定義され，分離音の品質を表す指標である．ソースコードは Bss Eval toolbox [11] を利用し，収録した状態の混合音からどれだけ SDR が改善するかを評価した．

また，比較として BF 分離音を遅延和（Delay–Sum）を適用した．

4.1.1 シミュレーション結果

シミュレーション環境で音源 S_0 に対して分離した結果を表 1 に示す．表から，各残響時間においてマイクロホンアレイが 2 台の場合と 3 台の場合ともに提案法の方が遅延和法よりも SDR 値が改善していることが分かる．また，音源 S_0 のように音源方向に別の音源が存在するような場合，どちらの手法もマイクロホンアレイ数が多い方が SDR 値が大きくなり，精度が向上していることが分かる．

次に，実際に推定された信号のスペクトログラムを図 4 に示す．スペクトログラムをみると，提案法は遅

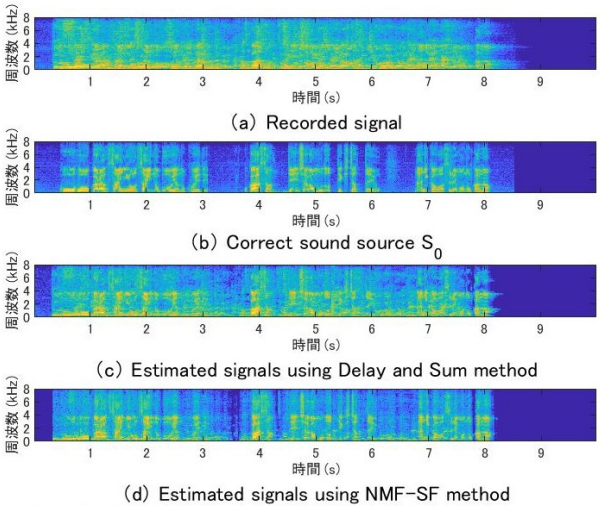


図 4: シミュレーション環境下で残響時間を約 0.7s とした際の分離結果を示す．上から収録信号，正解信号，遅延和による推定信号，提案法による推定信号を表す．横軸が時間縦軸が周波数，色が強度を表している．

延和法に比べノイズが除去され，正解信号に近いスペクトログラムになっている．例えば，3s〜4s の区間を見てみると遅延和法では残っていた雑音が低減されていることがわかる．遅延和法は目的音源を強調する一方で，目的音源以外の音を完全に除去することができない，これに対して，提案法は NMF により推定された基底スペクトルが目的音源かどうかを判定し，目的音源でないものに対してバイナリマスクをかけて除去するため，比較的他音源ノイズの除去に適していると考えられる．

4.2 実環境での実験

次に，実環境での実験を行った．音源はシミュレーションと同様に JNAS の音声コーパス [7] から 7–10s 程度の 4 つの音源組を 20 パターン用意し，GENELEC 8010APM スピーカーから音源を流した．収録装置は図 7 に示すような 16 チャンネルのマイクロホンアレイを使用し，サンプリング周波数 16kHz，量子化ビット数 24bit で収録した．各装置は図 5，6 のように配置し，高さは 1.2m で統一した．伝達関数は，HARKTool5 を利用し，幾何学的計算により作成した．NMF の基底数は 100 とし，アクティベーション行列の閾値 γ は 1.6×10^{-3} とした．比較手法として，遅延和法（Delay–Sum）と第 2 節で説明したクラスタリングベースのスポットフォーミング（LDA-SF）[5] を適用する．

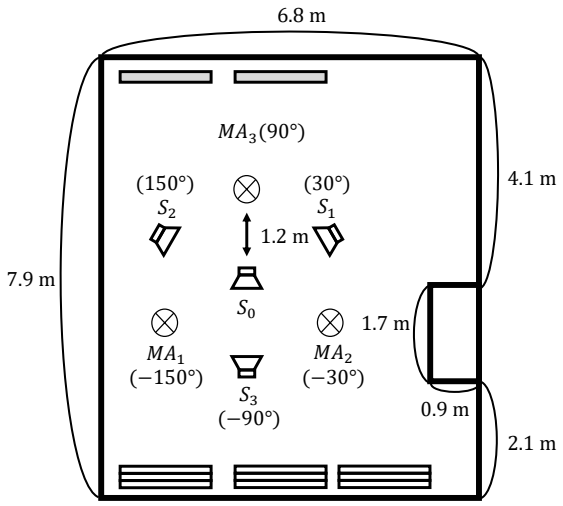


図 5: 実験場の見取り図

図 6: 実験の様子

4.2.1 実験結果と考察

実験結果を表 2 に示す．表は各音源に対する分離結果の SDR 改善値を表している．表を見ると，LDA–SF の性能が低くなっている．LDA–SF は第 2 節で説明したように同時発話音声を分離できないため，分離性能が低くなっていると考えられる．NMF–SF は，遅延和法とほぼ同等の性能になっており，あまり SDR 値が改善していないことがわかる．

この原因として考えられるのが残響音である．残響音が大きい環境では，ビームフォーミングの分離性能が低下してしまう．残響音の影響を見るためにスペクトログラムを図 8 に示す．図は上から，正解音源，各マイクロホンアレイ MA_1，MA_2，MA_3 でビームフォーミングして得られた分離音，提案手法による推定信号が表示されている．各マイクロホンアレイの BF 分離音を見ると，残響音が除去できておらず目的音源方向外の雑音が残っていることがわかる．各 BF 分離音に目的音源方向外の音が混入しているため，第 3.2 節で説明した共通基底選択が機能しなくなってしまってい

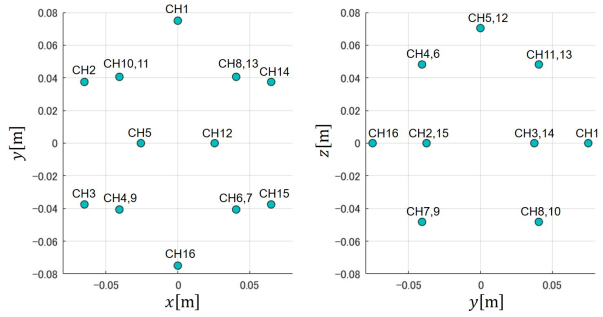


図 7: 使用したマイクロホンアレイの構成

表 2: 実環境での実験結果（SDR 改善値 [dB]）

method	アレイ数	S_0	S_1	S_2	S_3
NMF–SF	3	**3.4**	**5.0**	**11.8**	**12.5**
LDA–SF	3	0.73	1.0	1.3	0.83
Delay–sum	3	3.2	4.8	11.6	11.8
NMF–SF	2	2.6	2.9	2.1	5.3
Delay–sum	2	2.6	3.6	2.1	5.5
LDA–SF	2	0.80	-0.45	-1.2	2.3

ると考えられる．

その他に考えられる原因としては，スピーカーの指向性がある．シミュレーション環境では音源は等方的に広がっていると仮定される一方，実環境では音源に向きがあるため，マイクロホンアレイによって音が入りにくかったり，反射音の方が大きくなったりする可能性がある．現状では NMF の共通成分推定をアクティブ行列の閾値で判定しているため，誤検出が多くなると考えられる．

5　むすび

本稿では，複数台のマイクロホンアレイを用いた NMF による空間音源分離手法の実環境での性能評価を行った．シミュレーションにおいて，提案手法は既存手法よりも SDR がおよそ 0.8dB〜2.3dB 向上することが分かった．しかし，実環境では残響の影響を受けやすく，シミュレーション時に比べて分離精度が向上しにくいということが分かった．今回の実験はケーススタディであるため，様々な会場やシチュエーションで実験評価を進めていき，よりロバストな分離手法を構築することが今後の課題である．

謝辞

本研究は JSPS 科研費 JP19K12017, JP19KK0260 および JP20H00475 の助成を受けた．

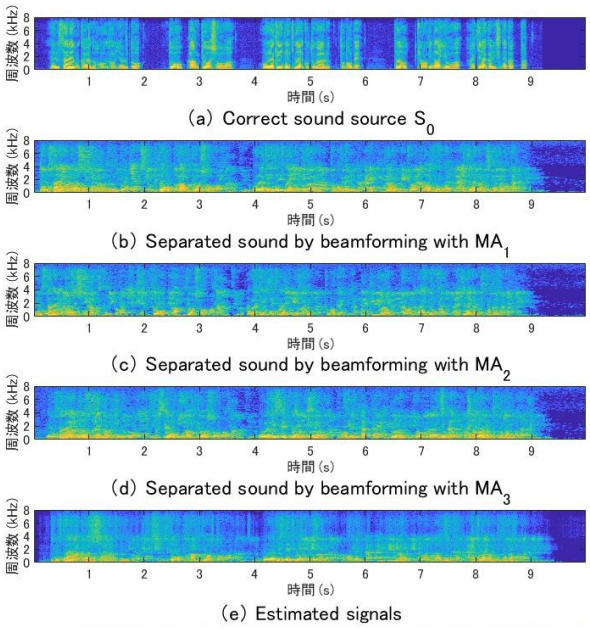


図 8: 分離結果のスペクトログラム．上から正解信号，マイクロホンアレイ MA_1，MA_2，MA_3 でビームフォーミングした分離音，最終的な推定信号を表す．

参考文献

[1] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca and R. J. Plemmons : “Algorithms and applications for approximate nonnegative matrix factorization”, *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[2] K. Sekiguchi, Y. Bando, K. Itoyama and K. Yoshii : “Layout optimization of cooperative distributed microphone arrays based on estimation of source separation performance”, *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 83–93, 2017.

[3] 鈴木基之, 本城剛士 : “2 本の超指向性マイクを用いたスポットフォーミング法の提案”, 研究報告音楽情報科学 (*MUS*), vol. 2014, no. 71, pp. 1–6, 2014.

[4] M. Taseska and E. A. P. Habets : “Spotforming: Spatial filtering with distributed arrays for position-selective sound acquisition”, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 24, no. 7, pp. 1291–1304, 2016.

[5] 鍵本泰宏, 糸山克寿, 西田健次 and 中臺一博: “複数マイクロホンアレイを用いた LDA によるスポッ

トフォーミングの検討”, 第 20 回計測自動制御学会システムインテグレーション部門講演会 (SI2019)) 講演論文集, pp. 1505-1510, 2019.

[6] 鍵本泰宏, 糸山克寿, 西田健次 and 中臺一博: “複数マイクロホンアレイを用いた NMF による空間音源分離法の提案と評価”, 日本ロボット学会誌, vol. 39, no. 7, pp. 669–672, 2021.

[7] 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 the Acoustical Society of Japan (E)* vol. 20, no. 3, pp. 199–206, 1999.

[8] H. Nakajima, K. Nakadai and Y. Hasegawa and H. Tsujino：“Blind source separation with parameter-free adaptive step-size method for robot audition”，*IEEE Transactions on Audio, Speech, and Language Processing*，vol. 18，no. 6，pp. 1476–1485，2009.

[9] K. Nakadai, H. G. Okuno and T. Mizumoto：“Development, deployment and applications of robot audition open source software HARK”，*Journal of Robotics and Mechatronics*，vol. 29，no. 1，pp. 16–25，2017.

[10] E. Vincent, R. Gribonval and C. Févotte：“Performance measurement in blind audio source separation”，*IEEE Transactions on Audio, Speech, and Language Processing*，14(4)，pp. 1462–1469，2006.

[11] F.-R. Stöter, A. Liutkus and N. Ito：“The 2018 signal separation evaluation campaign”，*International Conference on Latent Variable Analysis and Signal Separation*，pp. 293–305，2018.