

# 複数移動ロボットによる協調音源分離のための 分離精度予測を用いた配置最適化

## Layout Optimization of Multiple Mobile Robots for Cooperative Sound Source Separation

by Predicting Source Separation Performance

関口航平, 坂東昭宜, 糸山 克寿, 吉井 和佳

Kouhei Sekiguchi, Yoshiaki Bando, Katsutoshi Itoyama, Kazuyoshi Yoshii

京都大学 大学院情報学研究科

Graduate School of Informatics, Kyoto University

sekiguch@kuis.kyoto-u.ac.jp

### Abstract

本稿では、複数音源が存在する状況において、注目したい音源を高精度に分離することを目的として、マイクロホンアレイを搭載した複数移動ロボットの配置を最適化する手法について述べる。音源分離はマイクロホンアレイを搭載した全てのロボットを1つのマイクロホンアレイとみなして行う。音源分離の精度はロボットと音源の位置関係に依存するため、ロボットを最適な配置に移動させることで音源分離性能の向上を行うことができる。しかし、音源分離に最適な複数台のロボットの配置は自明ではない。本研究では、ロボット配置の分離精度を事前予測して、複数台のロボットの配置を最適化する。分離精度の予測値は、瞬時混合モデル上での混合行列と分離行列から計算する。実験では、提案法によりランダムな場合に比べてSDRが最大8.6 dB向上することを確認した。さらに、各ロボットで独立に分離音を生成してから統合する場合よりも提案法での分離精度が高くなることを確認した。

## 1 はじめに

近年の通信技術の発達に伴い遠隔地とのコミュニケーションを行う様々な手段が開発されている。その一つがテレプレゼンスロボットである。テレプレゼンスロボットとは、移動機構にカメラやマイクロホンを搭載したロボットで、遠隔地にいる操縦者がまるで現地にいるかのようにコミュニケーションを行うことを可能にする。例えば、在宅勤務者が自宅から社内の人とコミュニケーションをとるなどの目的でテレプレゼンスロボットが使用されている [Ng 15, Berri 14, Yan 13].



図 1: 複数ロボットの配置最適化の一例

テレプレゼンスロボットを用いて遠隔地とのコミュニケーションを円滑に行うためには、目的音以外の雑音への対策が不可欠となる。一般に、実際の環境では他人の話し声や音楽、空調機などの様々な雑音が存在しており、操縦者が実際に聞く音は複数の音を含む混合音となり、目的音の認識が困難になる。このような状況に対処するため、マイクロホンアレイ処理を用いて混合音を各音源信号に分離する研究が行われている [Makino 05, Lee 07, Nakajima 10]. 水本らは音源分離を用いて、操縦者が指定した方向の音だけを聞くテレプレゼンスロボットを開発した [Mizumoto 11].

音源分離の精度はマイクロホンと音源の位置関係に依存し、音源とロボットの位置関係によっては分離が困難となる問題がある [Nakadai 02]. 例えば、ロボットから見て複数の音源が同一方向に存在する場合や、音源間の距離差が大きい場合などである。注目音源が一つならば音源に近づくことで聞きやすくなるが、目的音源が複数存在する場合には、最適なロボットの配置は自明ではない。

本研究では、複数の子機ロボットを用いた音源分離支援システムの開発を行う (図 1). 操縦者が聞きたい音源を指定することで、その音源の配置に応じてマイクロホンアレイを搭載した子機ロボットが適切な位置に移動し、分離精度を向上させる。このとき、複数のロボットに搭載されたマイクロホンアレイ全体を一つの大きなマイクロホンアレイとみなし、すべてのマイクロホンでの観測音を

用いて音源分離を行う。ロボットの最適配置は音源分離精度を予測して決定する。各ロボット配置での実際の音源分離精度は、もとの音源信号が未知であるため計算することができない。そこで、瞬時混合モデルを仮定し、音源とロボットの位置関係を用いて音源分離をシミュレーションすることで音源分離精度の予測を行う。実験では、ランダムな配置と提案法による最適配置での分離精度の比較と、各ロボットで独立に分離音を生成して統合した場合と、複数のロボット全体を一つのマイクロホンアレイとみなして分離音を生成した場合の比較を行った。

## 2 音源分離に最適なロボット配置の探索

音源が複数存在する環境において、複数ロボットの配置を最適化することで、目的音源を高精度に分離する手法について述べる。本研究の課題は、音源分離に最適な複数ロボットの配置が自明ではないことである。ある配置で実際に音を録音して音源分離を行っても、元信号がないため分離音から分離精度を計算できず、最適配置を探索することができない。したがって、複数ロボットの最適配置探索には、実際に音源分離を行わず各ロボット配置での音源分離精度を予測することが必要となる。本研究では、音源分離に GICA や GHDSS [Nakajima 10] などの幾何制約付きブラインド音源分離手法を使用する。この手法は分離性能や環境適応性が高く計算量も少ないため、実時間での動作が望まれるロボット聴覚に適した手法である。一方、この手法では音源分離精度の予測が困難であるという問題がある。そこで、GICA や GHDSS と分離精度について相関のある遅延和ビームフォーミング (DSBF [Johnson 92]) の利得 (図 2) を用いて音源分離精度の推定を行う。利得とは分離音中に含まれる目的音と雑音の比率であり、分離音と音源信号の関係から計算することができる。この関係は音の混合過程と分離過程を推定することによって求めることができる。注目音源についての利得を用いた評価関数により遺伝的アルゴリズムで最適配置を決定する。

本稿で扱う配置最適化問題を以下のように定める。

- 
- 入力  $\mathbf{X}_t = [\mathbf{x}_{t1}, \dots, \mathbf{x}_{tM}]^T \in \mathbb{C}^{M \times F}$   
 $N$  個の音源が混合した  $M$  チャンネル観測音
- 出力 (1)  $\mathbf{Y}_t = [\mathbf{y}_{t1}, \dots, \mathbf{y}_{tN'}]^T \in \mathbb{C}^{N' \times F}$   
 注目している  $N'$  個の音源の分離音
- (2)  $\mathbf{B}^* = [\mathbf{b}_1^*, \dots, \mathbf{b}_R^*] \in \mathbb{R}^{R \times 2}$   
 $R$  台のロボットの最適配置の座標
- 仮定 (1) 各マイクロホンはすべて同期済み
- (2)  $N$  個の音源座標  $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_N] \in \mathbb{R}^{N \times 2}$   
 は音源定位と三角測量により既知 [Sasaki 06]
- 

ここで、音源の総数を  $N$  とし、そのうち注目する音源の数を  $N'$  と定める。 $\mathbf{X}_t, \mathbf{Y}_t$  はそれぞれ、録音した音響信号、分離音の  $t$  フレーム目を短時間フーリエ変換して得る。 $F$  は周波数ビンの数を表し、 $\mathbf{x}_{tm} = [x_{tm1}, \dots, x_{tmF}]$ ,

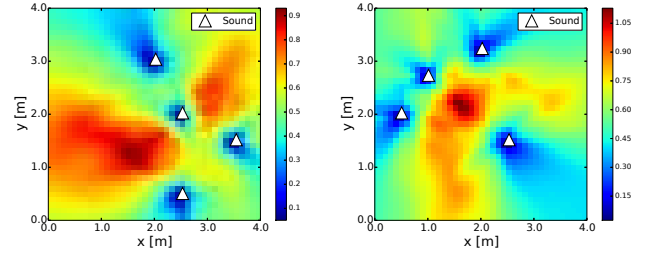


図 2: 1 台のロボットを部屋の各点に配置した場合の利得の一例。三角は音源位置を表し、値が大きい位置ほど分離精度が高くなると予想される。

$\mathbf{y}_{tm} = [y_{tm1}, \dots, y_{tmF}]$  である。

マイクロホンアレイの配置最適化の関連研究には Martinson らの手法 [Martinson 11] と佐々木らの手法 [Sasaki 11] がある。前者は 1 チャンネルマイクロホンを搭載した複数のロボットを用いる。音源の配置から音源定位に最適なロボット配置を幾何的に決定する。後者はマイクロホンアレイを搭載した 1 台のロボットを用いる。DSBF の利得を用いて、すべての方向に対して高い分離精度をもつ、音源配置によらないマイクロホンアレイの最適配置を探索する。

### 2.1 音の混合過程

音源信号  $\mathbf{S}_t = [s_{t1}, \dots, s_{tN}] \in \mathbb{C}^{N \times F}$  とマイクロホンによる観測音  $\mathbf{X}_t$  の関係について述べる。ここで、 $s_{ti}$  は音源  $i$  の  $t$  フレーム目の音源信号の短時間フーリエ変換を表す。音の伝搬を線形時不変システムと仮定すると、音源信号と観測音の関係は以下のように表される。

$$\mathbf{x}_{t,f} = \mathbf{H}_f \mathbf{s}_{t,f} \quad (1)$$

ここで、 $\mathbf{x}_{t,f} = [x_{t1f}, \dots, x_{tMf}]^T \in \mathbb{C}^M$ ,  $\mathbf{s}_{t,f} = [s_{t1f}, \dots, s_{tNf}]^T \in \mathbb{C}^N$  であり、 $\mathbf{H}_f \in \mathbb{C}^{M \times N}$  は混合行列である。雑音と残響を考慮せず、音の距離減衰と到達時間差のみを考慮した場合、 $x_{tmf}$  と  $s_{tnf}$  の関係は次のように表される。

$$x_{tmf} = \sum_{n=1}^N \frac{1}{d_{nm}} s_{tnf} \exp(-j2\pi f \tau_{nm}) \quad (2)$$

ここで、 $d_{nm}$  は音源  $n$  とマイクロホン  $m$  の間の距離を表し、 $\tau_{nm}$  は音源  $n$  のマイクロホン  $m$  への到達時間を表し、 $\tau_{nm} = d_{nm}/c$  ( $c$  は音速) で計算される。音の振幅は距離に反比例するため、 $1/d_{nm}$  の項は距離減衰を表す。式 (1) と式 (2) を比較すると、混合行列  $\mathbf{H}_f$  の  $(m, n)$  成分  $h_{mnf}$  は以下のように表される。

$$h_{mnf} = \frac{1}{d_{nm}} \exp(-j2\pi f \tau_{nm}) \quad (3)$$

### 2.2 音源分離

マイクロホンでの観測音  $\mathbf{x}(t)$  と分離音  $\mathbf{y}(t)$  の関係について述べる。音の混合過程と同様に、音源分離が線形時不変

システムで表されると仮定すると、観測音と分離音の関係は以下の式で表される。

$$\mathbf{y}_{t,f} = \mathbf{W}_f \mathbf{x}_{t,f} \quad (4)$$

ここで、 $\mathbf{y}_{t,f} = [y_{t1f}, \dots, y_{tNf}]^T \in \mathbb{C}^N$  であり、 $\mathbf{W}_f \in \mathbb{C}^{N \times M}$  は分離行列を表す。式 (1) と式 (4) から、 $\mathbf{W}_f = \mathbf{H}_f^{-1}$  のとき、 $\mathbf{y}_{t,f} = \mathbf{W}_f \mathbf{x}_{t,f} = \mathbf{W}_f \mathbf{H}_f \mathbf{s}_{t,f} = \mathbf{s}_{t,f}$  となり、分離音は音源信号と等しくなる。

GICA は ICA を基にした手法であり、音源信号の独立性を過程して、分離音が独立となるような分離行列  $\mathbf{W}$  を推定する。分離行列  $\mathbf{W}$  を推定するために、以下の二つのコスト関数を用いる。

$$\mathbf{J}_{\text{ICA}}(\mathbf{W}) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_k p(y_k)} \quad (5)$$

$$\mathbf{J}_{\text{GC}}(\mathbf{W}) = \|\mathbf{W}\mathbf{H} - \mathbf{I}\|^2 \quad (6)$$

ただし、 $p(\mathbf{y}) = p(y_1, \dots, y_N)$  である。 $\mathbf{J}_{\text{ICA}}(\mathbf{W})$  は  $p(\mathbf{y})$  と  $\prod_k p(y_k)$  の KL-divergence であり、独立性の尺度となっている。 $\mathbf{J}_{\text{GC}}$  は幾何制約を表す。実際の環境での混合行列  $\mathbf{H}$  は未知であるため、ここで与える  $\mathbf{H}$  はあらかじめ録音したインパルス応答や幾何的に計算したインパルス応答から作成する。本研究ではリアルタイムで音源分離を行うために以下の更新式を用いて、逐次的に分離行列  $\mathbf{W}$  を推定する。

$$\mathbf{W}_{t+1} = \mathbf{W}_t - \alpha \mathbf{J}'_{\text{ICA}} - \beta \mathbf{J}'_{\text{GC}} \quad (7)$$

ただし、 $\alpha, \beta$  はステップサイズパラメータであり、 $\mathbf{J}'_{\text{ICA}} = \nabla_{\mathbf{W}} \mathbf{J}_{\text{ICA}}$ 、 $\mathbf{J}'_{\text{GC}} = \nabla_{\mathbf{W}} \mathbf{J}_{\text{GC}}$  である。 $\{\}^*$  は複素共役を、 $\nabla$  は微分作用素を表す。GHDSS は GICA と類似した手法であり、分離行列  $\mathbf{W}$  を推定するために、 $\mathbf{J}_{\text{ICA}}$  の代わりに以下で定める  $\mathbf{J}_{\text{HDSS}}$  を用いる。

$$\mathbf{J}_{\text{HDSS}} = \|E[\mathbf{E}_\phi]\|^2 \quad (8)$$

ただし、

$$\mathbf{E}_\phi = \phi(\mathbf{y})\mathbf{y}^H - \text{diag}[\phi(\mathbf{y})\mathbf{y}^H] \quad (9)$$

$$\phi(y_i) = -\frac{\partial}{\partial y_i} \log p(y_i) \quad (10)$$

GICA や GHDSS では観測音などに応じて分離行列  $\mathbf{W}$  が異なるため、事前に分離行列を推定することが困難である。したがって、これらの手法を用いた場合、利得を計算することができないという問題がある。

本研究では遅延とビームフォーミング (DSBF) という手法に注目する。DSBF と GICA や GHDSS の分離精度には高い正の相関があるため、本研究では、この相関を利用して DSBF の利得を使って GICA や GHDSS での分離精度の予測を行う。音源分離手法として DSBF を用いた場合、分離行列  $\mathbf{W}_f$  の要素はマイクロホンと音源の位置

関係のみから決定される。DSBF とは注目音源の座標から各マイクロホンへの到達時間差を推定し、観測信号を到達時間差だけ時間シフトして足し合わせるにより注目音を強調する音源分離手法である。本研究では、各マイクロホンと音源の距離を考慮し、音源に近いマイクロホンの観測音の比率を高く、音源と遠いマイクロホンの観測音の比率を低くして足し合わせる。したがって、分離音と観測音の関係は周波数領域では次のように表される。

$$y_{tnf} = \sum_m \frac{1}{d_{nm}} x_{tmf} \exp(j2\pi f \tau_{nm}) \quad (11)$$

式 (4) と式 (11) から、分離行列  $\mathbf{W}_f$  の  $(n, m)$  成分  $w_{nmf}$  は以下の式で表される。

$$w_{nmf} = \frac{1}{d_{nm}} \exp(j2\pi f \tau_{nm}) \quad (12)$$

### 2.3 目的関数

複数ロボットの配置最適化における目的関数を DSBF の利得の調和平均として定める。ロボット配置  $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_R]$  における目的関数の値を  $f(\mathbf{B})$  とすると、

$$f(\mathbf{B}) = \frac{N'}{\sum_{n \in D} \frac{1}{g_n(\mathbf{B})}} \quad (13)$$

ここで、 $D$  は注目音源の集合を表す。 $g_n(\mathbf{B})$  は音源  $n$  の利得を表し、音源  $n$  の分離音中の音源  $n$  と雑音の比率として定める。利得の調和平均を目的関数としたのは、本研究ではすべての注目音源の高精度な分離を目的とするためである。もし一つでも分離精度が悪い音源が存在する場合、目的関数の値は大きく低下する。

利得は分離音と音源信号の関係式から計算することが可能である。式 (1) と式 (4) から、分離音と音源信号の関係は周波数領域で以下のように表される。

$$\mathbf{y}_{t,f} = \mathbf{A}_f \mathbf{s}_{t,f} \quad (14)$$

ここで、 $\mathbf{A}_f \in \mathbb{C}^{N \times N}$  は利得行列であり、 $\mathbf{A}_f = \mathbf{W}_f \mathbf{H}_f$  として定める。利得行列  $\mathbf{A}_f$  の対角成分は分離音中に含まれる目的音源の比率を、非対角成分は雑音の比率を表している。したがって、音源  $n$  の利得  $g_n(\mathbf{B})$  は以下のようになる。

$$g_n(\mathbf{B}) = \frac{\sum_f a_{nnf}}{\sum_{n \neq k} \sum_f a_{nkf}} \quad (15)$$

ここで、 $a_{nkf}$  は利得行列  $\mathbf{A}_f$  の  $(n, k)$  成分であり、音源  $n$  の分離音に含まれる音源  $k$  の割合を示す。

DSBF の利得を用いた場合には、式 (11) と式 (2) から  $a_{nkf}$  は以下のようになる。

$$a_{nkf} = \left| \sum_{m=1}^M \frac{1}{d_{nm} d_{km}} \exp(j2\pi f (\tau_{nm} - \tau_{km})) \right| \quad (16)$$

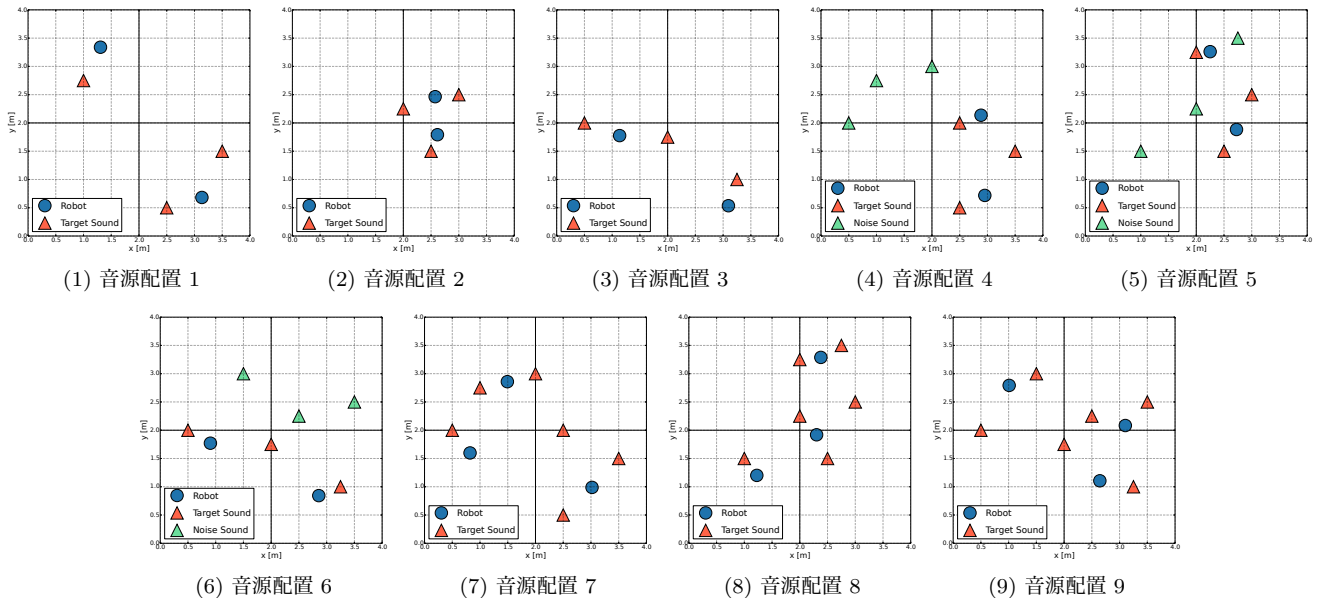


図 3: 音源配置と提案法で求めた複数ロボットの最適配置。青丸がロボット、赤三角が注目の音源、緑三角が雑音を表す。

## 2.4 最適配置探索

本研究では遺伝的アルゴリズムを用いて最適配置探索を行う。これは、グリッドサーチによる全探索を用いた場合、ロボットの台数に指数的に比例して計算量が増加してしまい、また、勾配法などを用いた場合には局所最適解に陥ってしまう可能性があるためである。複数ロボットの座標と方向の組を1つの個体とみなし、個体の組み替えは現在位置の近傍へ移動することで行い、突然変異によりランダムに移動することで局所最適解に陥ることを防ぐ。ロボットの向きはランダムに与える。ただし、ロボット間の距離が離れすぎた場合、1つの時間フレーム内に含まれる音源信号の区間がロボット間で大きく異なってしまうため、分離精度が低下してしまう問題がある。そのため、ロボット間の距離が一定距離以内に収まるように制約を設ける。選択はエリート選択とルーレット選択を併用する。目的関数は複数台のロボットに搭載した全てのマイクを1つのマイクロホンアレイとみなして計算を行う。世代交代を一定回数行った後、評価関数の値が最大となる個体を最適配置とする。

## 3 評価実験

ランダムな配置と提案法による最適配置の比較と、複数ロボット全体を一つのマイクロホンアレイとみなして音源分離する場合と各ロボットで独立に分離音を生成して統合した場合を比較するために、シミュレーション混合を用いた評価実験を行った。

### 3.1 実験条件

一辺 4 m の正方形の部屋に音源  $N$  個、ロボット  $R$  台がある場合を想定する。  $N$  個の音源の中で、  $N'$  個の音源を注目したい音源、  $N - N'$  個の音源を雑音とみなす。以下

の 3 つの条件それぞれについて 3 種類の音源配置を用意し実験を行った (図 3)。

1. 注目の音源 3 つ、ロボット 2 台 ( $N = N' = 3, R = 2$ )
2. 注目の音源 3 つ、雑音 3 つ、ロボット 2 台 ( $N = 6, N' = 3, R = 2$ )
3. 注目の音源 6 つ、ロボット 3 台 ( $N = N' = 6, R = 3$ )

各ロボットは 8 チャンネルマイクロホンアレイを搭載し、ロボットの配置を提案法を用いて最適化する。最適配置での観測音は幾何学的に計算したインパルス応答を使ったシミュレーション混合を用いて生成する。音源信号は JNAS 音素バランス文 (量子化 24 bit, サンプリング周波数 16 kHz)[Sagisaka 92] を用いた。

ランダムな配置と、提案法での最適配置で、複数のロボット全体を一つのマイクロホンアレイとみなして音源分離を行った場合と、最適配置で各ロボットで独立に分離音を生成して統合した場合の比較を行った。音源分離手法には DSBF, GHDSS, GICA の 3 つを用いた。GICA と GHDSS を用いて各ロボットで独立に分離音を生成した場合、各分離音の位相がわからないという問題がある。そのため、分離音をそのまま足しただけでは、位相が合わず分離精度が悪くなってしまう可能性がある。この問題を考慮して、分離音の統合方法として以下の 3 つの手法を用いた。

- (a) Single: 各音源に最も近いロボットでの観測音のみを用いて分離音を生成する。
- (b) Average: ロボット  $R$  台の場合、  $R - 1$  個の分離音をそれぞれ  $0.5n$  サンプルずらして加算し、分離精度を計算する ( $n = 0, 1, \dots, 20$ )。  $n^{R-1}$  の組み合わせの中で最も分離精度が高くなったものを、分離音として出力する。

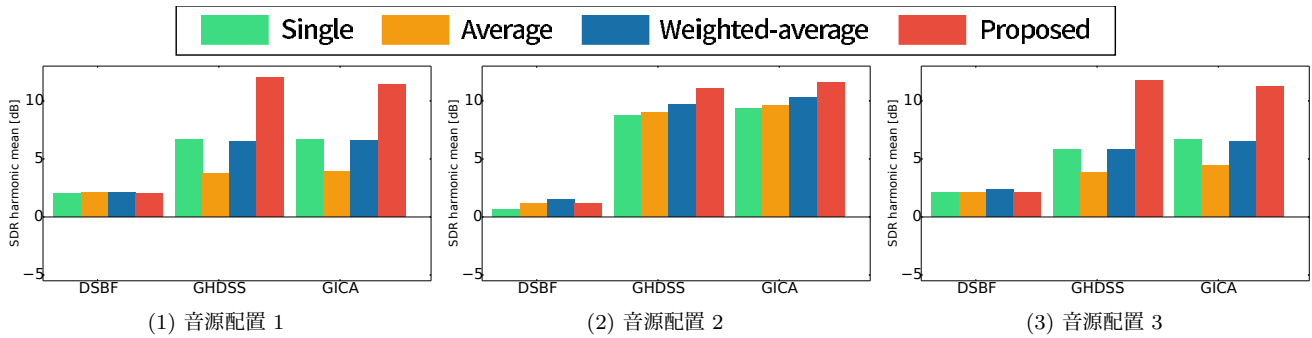


図 4: ロボット数：2，注目音源数：3，雑音数：0 の場合の SDR の調和平均

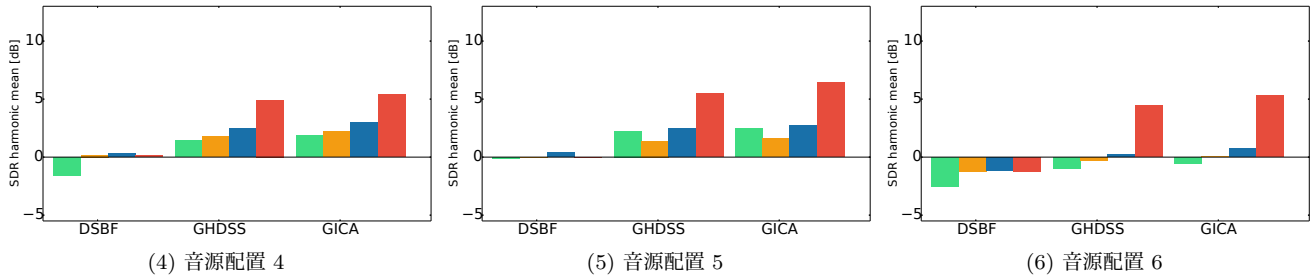


図 5: ロボット数：2，注目音源数：3，雑音数：3 の場合の SDR の調和平均

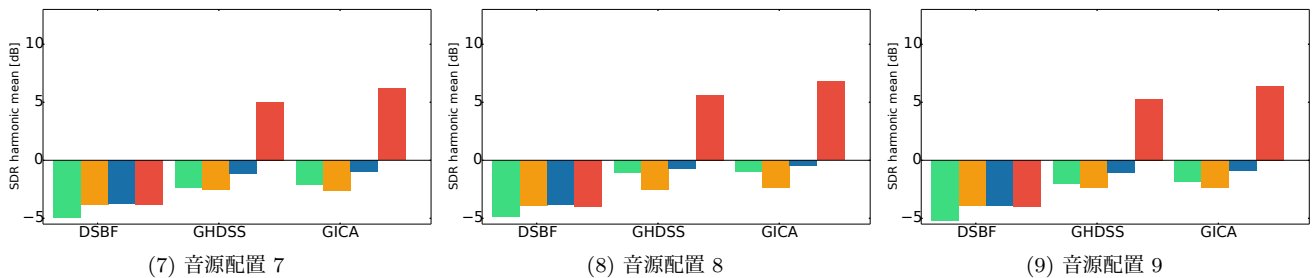


図 6: ロボット数：3，注目音源数：6，雑音数：0 の場合の SDR の調和平均

(c) Weighted-average：音源とロボットの距離で重み付けて (b) を行う。

音源分離精度の指標には sound-to-distortion ratio(SDR)[Vincent 06, Raffenl 14] の注目音源についての調和平均を用いた。調和平均を使用したのは、目的関数で利得の調和平均を用いたのと同様、本研究では全ての注目音源を高精度に分離することを目的としており、調和平均は一つでも分離精度が低い音源が存在すると値が大きく下がるためである。最適配置探索は遺伝的アルゴリズムを用いることによりランダム性を持つため、各音源配置について提案法による最適化を 30 回行い、各ロボット配置について注目音源の SDR の調和平均を求め、その平均を計算した。

### 3.2 実験結果

ランダムな配置と提案法による最適配置での分離精度を表 1 に示す。すべての場合において、提案法による最適配置がランダムな配置を上回っている。提案法による最適配置で、複数ロボットを 1 つのマイクロホンアレイと見なす場合と、各ロボットで独立に分離を行う場合の実験結果を図 4, 図 5, 図 6 に示す。全ての条件で、DSBF を用いた場合よりも GHDSS や GICA を用いた場合の分離精度が上

分離	配置	1	2	3	4	5	6	7	8	9
DSBF	ランダム	-2.7	-2.5	-2.5	-6.3	-6.6	-6.3	-8.7	-8.8	-9.5
	最適位置	2.1	1.2	2.1	0.2	-0.1	-1.3	-3.8	-4.0	-4.0
GHDSS	ランダム	4.5	2.4	4.4	-3.0	-3.8	-3.5	-3.5	-5.5	-4.8
	最適位置	12.0	11.0	11.8	4.9	5.5	4.4	5.0	5.6	5.2
GICA	ランダム	5.0	4.4	5.5	-1.8	-2.5	-2.4	-1.2	-2.6	-1.8
	最適位置	11.4	11.5	11.2	5.4	6.4	5.3	6.2	6.8	6.4

表 1: ランダムな配置と提案法による最適配置での分離精度 (SDR [dB])

回っている。これは、DSBF では雑音によらず注目音源の位相に合わせて信号をずらして足し合わせるだけであるが、GHDSS や GICA では、雑音方向に Null ビームを形成することにより、雑音を消すことができるためである。

全てのマイクロホンと同時に用いる提案法と、各ロボットごとに分離音を生成する Single, Average, Weighted-average を比較する。DSBF を用いた場合には、提案法と Average が同一となり、提案法と比べて Single は平均で 0.78 dB 低く、Weighted-average は平均で 0.16 dB 高くなり、ほとんど差がなかった。提案法と Average が同一となったのは、Average が行っているロボットごとに平均を取り、さらにロボット間で平均を取る操作は、全てのロボットで一度に平均を取る操作と同じためである。一方、GHDSS や GICA では、提案法が他の手法を平均で 5.2 dB

以上上回る結果となった。これは、GICA や GHDSS では、目的音源をほとんど含まない観測音も Null ビームを形成するために使うことができるためである。各ロボットで独立に分離音を生成した場合、目的音源をほとんど含まない観測音を有効に使うことができず、また、この観測音から生成される分離音は分離精度が悪くなるため、その音を足すことで全体の分離精度も下がってしまう場合があると考えられる。

今回試した手法以外にも複数の音源を統合する手法が考えられるが、GHDSS と GICA では平均で 5.2 dB 以上他の手法を上回り、また、実際には各ロボットで分離音を生成する場合、統合の際に位相のずれの問題があるため分離精度が更に低下することが予想される。これらのことから、提案法の有効性が確認できたと言える。

### 3.3 まとめ

本稿では、複数の音源が存在する状況において、注目したい音源に応じて複数ロボットの配置を最適化することで、音源分離精度の向上を行う手法を開発した。音源分離は複数のロボット全体を一つの大きなマイクロホンアレイとみなして行った。複数ロボットの最適配置はロボットと音源の位置関係から分離精度を予測することで決定した。実験では、提案法によりランダムな場合に比べて SDR が最大 8.6 dB 向上することを確認した。さらに、各ロボットで独立に分離音を生成してから統合する場合よりも提案法での分離精度が高くなることを確認した。今後は、実環境で複数ロボットを用いて音源分離を行う際に問題となるロボット間の同期に取り組む予定である。

謝辞 本研究の一部は、科研費 24220006、および ImPACT「タフ・ロボティクス・チャレンジ」の支援を受けた。

### 参考文献

- [Berri 14] Berri, R., et al.: Telepresence Robot with Image-based Face Tracking and 3D Perception with Human Gesture Interface using Kinect Sensor, in *JCRIS*, pp. 205–210 (2014)
- [Johnson 92] Johnson, D. H. and Dudgeon, D. E.: *Array Signal Processing: Concepts and Techniques*, Prentice Hall (1992)
- [Lee 07] Lee, I., et al.: Fast Fixed-point Independent Vector Analysis Algorithms for Convolutional Blind Source Separation, *J. Signal Processing*, Vol. 87, No. 8, pp. 1859–1871 (2007)
- [Makino 05] Makino, S., et al.: Blind Source Separation of Convolutional Mixtures of Speech in Frequency Domain, *IEEE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, Vol. 88, pp. 1640–1655 (2005)
- [Martinson 11] Martinson, E., et al.: Optimizing a Reconfigurable Robotic Microphone Array, in *IEEE/RSJ IROS*, pp. 125–130 (2011)
- [Mizumoto 11] Mizumoto, T., et al.: Design and Implementation of Selectable Sound Source Separation on the Texai Telepresence System using HARK, in *IEEE ICRA*, pp. 2130–2137 (2011)
- [Nakadai 02] Nakadai, K., et al.: Real-Time Sound Source Localization and Separation for Robot Audition, in *ICSLP*, pp. 193–196 (2002)
- [Nakajima 10] Nakajima, H., et al.: Blind Source Separation With Parameter-Free Adaptive Step-Size Method for Robot Audition, *IEEE Trans. Audio, Speech and Language Processing*, Vol. 18, No. 6, pp. 1476–1485 (2010)
- [Ng 15] Ng, M. K., et al.: A cloud robotics system for telepresence enabling mobility impaired people to enjoy the whole museum experience, in *IEEE DTIS*, pp. 1–6 (2015)
- [Raffel 14] Raffel, C., et al.: mir\_eval: A Transparent Implementation of Common MIR Metrics, in *ISMIR*, pp. 367–372 (2014)
- [Sagisaka 92] Sagisaka, Y. and Uratani, N.: ATR Spoken Language Database, *J. The Acoustic Society of Japan*, Vol. 48, No. 12, pp. 878–882 (1992)
- [Sasaki 06] Sasaki, Y., et al.: Multiple Sound Source Mapping for a Mobile Robot by Self-motion Triangulation, in *IEEE/RSJ IROS*, pp. 380–385 (2006)
- [Sasaki 11] Sasaki, Y., et al.: 32-Channel Omnidirectional Microphone Array Design and Implementation, *J. Robotics and Mechatronics*, Vol. 23, No. 3, pp. 378–385 (2011)
- [Vincent 06] Vincent, E., et al.: Performance Measurement in Blind Audio Source Separation, *IEEE Trans. Audio, Speech and Language Processing*, Vol. 14, No. 4, pp. 1462–1469 (2006)
- [Yan 13] Yan, R., et al.: An Attention-Directed Robot for Social Telepresence, in *HAI*, pp. III–1–2 (2013)