4T-07

分散型マイクロホンアレイを用いた音源分離のための複数移動ロボットの配置最適化

関口 航平[1]　坂東 宜昭[2]　糸山 克寿[2]　吉井 和佳[2]
[1]京都大学 工学部 情報学科　[2]京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

ロボットが自律して行動するためには周辺環境の情報を得ることが必要であり，ロボットはカメラやマイクなどのセンサーを用いてこれを行う．マイクを用いて周囲の音環境を把握するためには音源の定位や認識を行う必要がある．また，複数の音源が存在する環境ではそれぞれの音を認識するために混合音の分離が必要となる．そのため，これまでに音源定位や分離，認識といったロボット聴覚の研究がなされてきた．例えば，1台のロボットによる3話者同時認識 [1] や，1台の移動ロボットでのアクティブオーディション [2]，複数の移動ロボットでのアクティブオーディション [3] などの研究が行われてきた．[3] では，音源数は1つを想定し，音源とロボットの幾何制約によって音源定位に最適な複数ロボットの配置を決定する．しかし，実環境におけるロボット聴覚では複数の音源への対応が不可欠である．

本研究では音源が複数ある場合を想定し，複数台のマイクロホンアレイを搭載した移動ロボットを分散型マイクロホンアレイとみなし，音源分離に最適なロボット配置の探索を行う．音源とマイクの位置関係が音源分離の性能に大きく影響するという知見 [4] に基づき，音源分離性能を予測することでロボットを最適な配置に移動させ音源分離性能を改善する．

2. 複数ロボットの音源分離に最適な配置探索

本稿では音源が複数存在する環境において，各ロボットに搭載された M チャンネルのマイクロホンアレイで録音した混合音から，全ての音を精度良く分離することを目的とする．ロボットの台数より音源数が少ない場合は，各音源にロボットが接近することで目的を達成できるため，本稿ではロボットの台数より音源数が多い場合を想定する．複数ロボットの最適配置探索の課題は，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推定することである．本研究では，音源分離に遅延和ビームフォーミング (DSBF) を使用し，DSBF の利得を用いて音源分離精度の推定を行う．利得とは混合音の分離音に含まれる雑音成分の比率であり，各マイクへの音の到達時間差から計算できる．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	M チャンネルマイクロホンアレイで録音した観測音 $x_1(t), \cdots, x_M(t) \in \mathbb{R}$
出力	N 個の音源の分離音 $y_1(t), \cdots, y_N(t) \in \mathbb{R}$ ロボットの最適配置 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_R \in \mathbb{R}^2$
仮定	(1) 各マイクは同期済み (2) 音源の座標 $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_N \in \mathbb{R}^2$ は既知

本稿ではロボット，音源の位置は二次元平面上で表され，ロボット r，音源 n の各座標を $\boldsymbol{a}_r, \boldsymbol{b}_n$ とする．

Optimizing the Positions of Microphone Arrays for Sound Source Separation: *Kouhei Sekiguchi, Yoshiaki Bando, Katsutoshi Itoyama, Kazuyoshi Yoshii (Kyoto Univ.)*

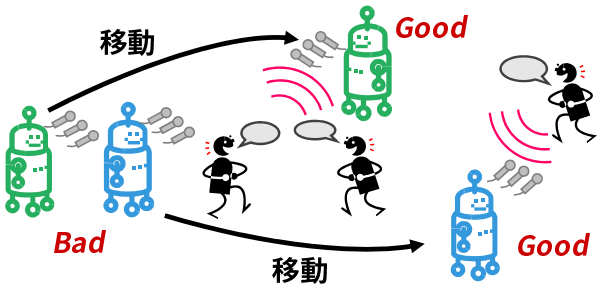

図 1: 音源分離性能向上のためのロボット配置最適化

DSBF の利得を用いたマイクロホンアレイの配置最適化の関連研究には佐々木らの手法 [5] がある．この手法では，メインローブ幅が広く，サイドローブでの感度が低くなることを目的としており，音源方向に依らない最適配置を決定する．本研究では音源の座標を用いて，音源の配置に応じた最適なロボット配置を探索する．

2.1 音源分離

本研究では，複数台のロボットに搭載したマイクロホンアレイを1つのマイクロホンアレイとみなして，DSBF により音源分離を行う．DSBF とは注目音源の座標から各マイク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ことにより注目音を強調する音源分離手法である．このとき，各マイクと音源の距離を考慮し，音源に近いマイクの観測音の比率を高く，音源と遠いマイクの観測音の比率を低くして足し合わせる．つまり，時刻 t でのマイク m の観測音を $x_m(t)$，音源 n とマイク m 間の距離を d_{nm}，音速を c として，音源 n の分離音 $y_n(t)$ を以下のように計算する．

$$y_n(t) = \sum_{m=1}^{M} \frac{1}{d_{nm}} x_m\left(t + \frac{d_{nm}}{c}\right)$$

2.2 遅延和ビームフォーミングの利得

各周波数 f について各音源が正弦波であると仮定し，音源 n に対する利得 $g(n)$ を以下のように定める:

$$g(n) = \frac{\sum_{k \neq n}^{N} \alpha_{nk}}{\alpha_{nn}}$$

ただし，

$$\alpha_{nk} = \left| \sum_{m=1}^{M} \frac{1}{{d_{km}}^2} e^{i\omega\theta_{km}} \right|, \quad \theta_{km} = \frac{d_{km} - d_{nm}}{c}$$

ここで，ω は角周波数を表し，$\omega = 2\pi f$ の関係が成り立つ．α_{nk} は音源 n に注目した際の音源 k の振幅を示す．$(1/d_{km})^2$ の項は，音の振幅が距離に反比例することと，音源分離時に音源から遠いマイクの観測音の比率を低くして足し合わせることを反映している．α_{nn} が注目音源の振幅を表し，$\sum_{k \neq n}^{N} \alpha_{nk}$ が雑音の振幅を表し，利得は注目音源の振幅に対する雑音の振幅の比である．

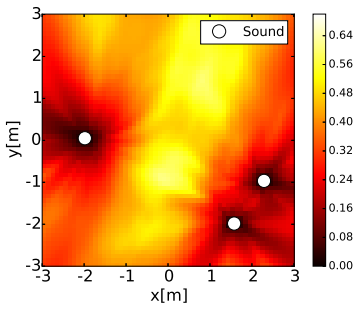

図 2: ロボット 1 台，音源 3 つの場合の評価値の一例

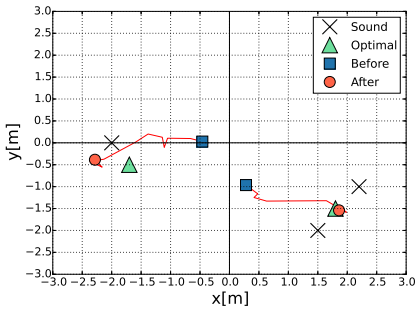

図 3: 評価関数の値が最大となる個体の軌跡の一例

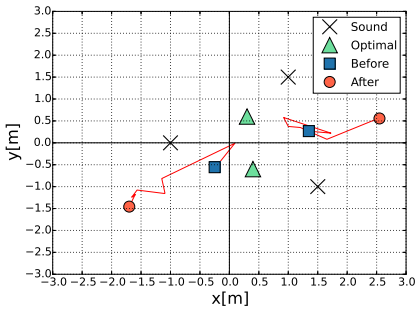

図 4: 正解配置と提案法による最適解が異なる場合の一例

2.3 評価関数

R 台のロボットの配置が $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_R$ の場合における評価関数 $f(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_R)$ を利得の調和平均を用いて以下のように定める:

$$f(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_R) = \frac{1}{\sum_{n=1}^{N} \frac{1}{g(n)}}$$

利得の調和平均をとることで，全音源の分離精度が良い場合に評価値が高くなるため，目的とするロボット配置を求めることができる．図 1 に音源 3 つ，ロボット 1 台の場合の，各点における評価関数の値の一例を示す．

2.4 最適配置探索

本研究では遺伝的アルゴリズム (GA) を用いて最適配置探索を行う．複数ロボットの配置を個体とみなし，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．突然変異率は最初は大きく，徐々に小さくなるように設定する．選択はエリート選択とルーレット選択を併用する．このとき，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評価関数の計算を行う．世代交代を一定回数行った後，評価関数の値が最大の個体を最適配置とする．

3. 評価実験

実験条件 複数ロボットをランダムに配置した場合, 提案法による最適配置, 正解配置での分離精度を比較し, 提案法の有効性を確認するためシミュレーションによる評価実験を行った．各配置での分離精度は，シミュレーションにより作成した混合音を DSBF を用いて音源分離を行い，各分離音の Signal to Distortion Ratio (SDR) [6] の調和平均とした．正解配置は 0.2[m] 間隔のグリッドサーチで各配置について実際に分離精度を計算し，分離精度が最大となった配置とした．1 辺 6[m] の正方形の部屋に音源が 3 つ，ロボットが 2 台ある場合を想定し, 以下の操作を 8 パターンの音源配置に対し, それぞれ 30 回行った．まず 2 台のロボットをランダムに配置し，その配置で分離精度を計算する．次に提案法により 2 台のロボットの最適配置を探索し，この配置で分離精度を計算する．音源配置 (1,2) は 3 つの音源が一直線に並ぶ場合，(3,4,5) は 2 つの音源が近く，1 つの音源が離れている場合，(6,7,8) は 3 つの音源全てが離れている場合である．音源は ATR 音素バランス文からランダムに選んだ．

実験結果 図 4 に，各音源配置にてランダムにロボットを配置した場合の分離精度の平均値，提案法による最適配置での分離精度の平均値，正解配置での分離精度を示す．全ての場合において提案法はランダムに配置した場合よりも分離精度が向上している．図 2 は音源配置 (3) の場合の評価値最大の個体の軌跡の一例であり，四角の印がロボットの初期配置を，丸印が提案法による最適配置を，三角の印が正解配置である．この音源配置では，提案法による最適配置と正解配置が近く，正解データと同等の分離精度を実現した．

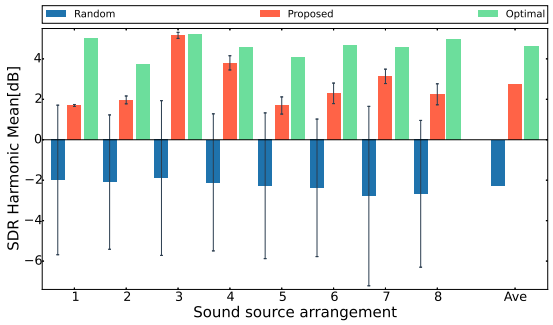

図 5: 各音源配置での分離精度

一方，図 3 は音源配置 (8) の軌跡の一例であり，途中で正解配置に近づいたが最終的には正解配置と離れた配置となっており，提案法による最適配置と正解配置が異なっていることがわかる．ランダムに配置した場合よりは分離精度が向上しているが，正解配置での分離精度は大きく下回っている．これは，提案法の評価関数では周波数ビンごとに重みをつけていないが，実際には音が含まれる周波数ビンが重要となることに起因すると考えられる．録音した音から周波数ビンごとの重みを推定し，評価関数に与えることで正解に近づくことが期待できる．

4. まとめ

本稿では音源分離に最適な複数ロボットの最適配置を探索する手法を開発した．シミュレーション混合音を用いた評価実験で，提案法により音源分離精度が平均 5.0db 向上することを確認した．今後は実環境での録音を用いた評価実験や，強化学習の枠組みを用いた音源・ロボット位置と最適配置のオンライン同時推定を行う予定である．

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

[1] K. Nakadai *et al.* Robot recognizes three simultaneous speech by active audition. *ICRA–2003*, 398–405.
[2] T. Yoshida *et al.* Active audio-visual integration for voice activity detection based on a causal bayesian network. *Humanoids–2012*, 370–375.
[3] E. Martinson *et al.* Optimizing a reconfigurable robotic microphone array. *IROS–2011*, 125–130.
[4] K. Nakadai *et al.* Real-time sound source localization and separation for robot audition. *ICSLP–2002*, 193–196.
[5] 佐々木 洋子 *et al.* 32ch 低サイドローブ設計マイクロホンアレイの開発と屋外での音源定位評価. *ROBOMECH–2010*, 2A2–B02(1)–(4).
[6] C. Raffel *et al.* mir-eval: A transparent implementation of common MIR metrics. *ISMIR–2014*, 367–372.