

ロボットのためのマイクアレイ音響信号処理

吉井 和佳^{1,2,a)}

概要：本講演では、マイクアレイを搭載したロボットのためのマルチチャネル音響信号処理技術について解説する。周囲の環境に応じて自律的に行動することができるロボットを開発するためには、カメラや深度センサから得られる画像情報だけではなく、マイクから得られる音響情報を適切に解析・理解する技術が必要となる。そのような音環境理解 (computational auditory scene analysis, CASA) 技術の中でも、マイクアレイを用いた音源定位・分離は、最も基本的な研究課題の一つである [1]。その基本原理は、もし、マイクアレイから見て、ある方向に音源が一つ存在すると、音源信号の到達時刻にはマイク間で差があることから、逆に、到達時間差に着目すれば方向が推定できることにある。時間領域における到達時間差は、周波数領域における位相差に相当し、マイク間の複素スペクトルの相関は空間相関行列で与えられる。空間相関行列は、マイク配置から幾何的に計算するか、無響室あるいは実際の環境下で、音源の方向ごとにインパルス応答を測定することで求められるが、理論値からの乖離や環境の変化に弱いといった問題がある。そのため、音源定位・分離を行うと同時に、空間相関行列を同時に推定するブラインド音源分離の研究が盛んにおこなわれている [2-5]。近年は、マルチチャネル信号処理ならではの空間情報に着目するだけではなく、シングルチャネル信号処理でしばしば手掛かりとなる音源情報を取り込むことで、音源定位・分離の精度向上が試みられている。例えば、音源信号（それ自身が音楽のような混合音の可能性があるので注意）のスペクトログラムが低ランク構造を持つと仮定すると、非負値行列因子分解 (nonnegative matrix factorization, NMF) を用いて音源信号のスペクトログラムを基底行列とアクティベーション行列との積で近似することができる [6]。このとき、近似誤差の最小化に基づく NMF は、ある確率モデルの最尤推定に対応していることが知られており [7]、その確率モデルはシングルチャネルの音源スペクトログラムの生成モデルとして利用できる。音源信号が空間を伝播してマイクアレイに到達すると、空間相関行列に基づくマルチチャネルの音源スペクトログラム (イメージと呼ばれる) の生成モデルが定式化できる。逆に、これらイメージが重畳したスペクトログラムが観測された時に、音源を構成する基底やアクティベーション、空間相関行列を一挙に同時推定する音源分離技法が提案されている。このとき、事前に測定した空間相関行列を事前分布としてベイズ推定を行うことで、実際の環境に適応しつつ、音源分離と定位を同時に行うことも可能になっている。

参考文献

- [1] 浅野太：音のアレイ信号処理（音響テクノロジーシリーズ），コロナ社（2011）。
- [2] Ozerov, A. and Févotte, C.: Multichannel Nonnegative Matrix Factorization in Convolutional Mixtures for Audio Source Separation, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 3, pp. 550-563 (2010).
- [3] Sawada, H., Kameoka, H., Araki, S. and Ueda, N.: Multichannel Extensions of Non-Negative Matrix Factorization With Complex-Valued Data, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 5, pp. 971-982 (2013).
- [4] Itakura, K., Bando, Y., Nakamura, E., Itoyama, K., Yoshii, K. and Kawahara, T.: Bayesian Multichannel Audio Source Separation Based on Nested Source and Spatial Models, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 550-563 (2017). under review.
- [5] 板倉光佑, 坂東宜昭, 中村栄太, 糸山克寿, 吉井和佳, 河原達也：マルチチャネル音源分離のための低ランク音源モデルとスパース重畳過程に基づくネスト型ベイズ混合・因子モデル，電子情報通信学会第 19 回情報論的学習理論ワークショップ，Vol. 116, No. 300, pp. 353-359 (2016).
- [6] Smaragdakis, P. and Brown, J. C.: Non-negative Matrix Factorization for Polyphonic Music Transcription, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 177-180 (2003).
- [7] Févotte, C., Bertin, N. and Durrieu, J.-L.: Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis, *Neural Computation*, Vol. 21, No. 3, pp. 793-830 (2009).

¹ 京都大学 大学院情報学研究所 知能情報学専攻

Yoshida-honmachi, Sakyo, Kyoto, Kyoto 606-8501, Japan

² 理化学研究所 革新知能統合研究センター (AIP) 音響情景理解チーム

15F, 1-4-1 Nihonbashi, Chuo, Tokyo 103-0027, Japan

a) yoshii(at)kuis.kyoto-u.ac.jp