

# 多くの背景音からの主要音源のブラインド抽出

Blind Extraction of Dominant Target Sources from Many Background Interference Sources

澤田 宏  
Hiroshi Sawada

荒木 章子  
Shoko Araki

向井 良  
Ryo Mukai

牧野 昭二  
Shoji Makino

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories, NTT Corporation

## 1 まえがき

ブラインド音源分離 (BSS: Blind Source Separation) は、音源や混合系に関する情報を用いずに、観測信号のみから音を分離する技術である。我々は、周波数領域で独立成分分析 (ICA: Independent Component Analysis) を適用する際に発生する permutation の問題を効率的に解決する手法を提案し、残響のある実環境で多数の音源を分離することに成功した [1]。本稿では、より現実的な状況として、全体の音源数がマイクロホン数よりも多い、あるいは音源数が無数で数えられないような状況を考える。ただし、これら無数の音源をすべて分離することは難しいため、マイクロホンに比較的近いいくつかの主要音源のみを分離・抽出する。文献 [2] では、主要音源の数を 1 個と仮定しているが、本稿では、主要音源が複数の場合を考慮する。

## 2 提案手法

図 1 に提案手法の処理の流れを示す。まず、観測信号を短時間フーリエ変換 (STFT: Short-Time Fourier Transform) し、周波数毎の時間系列に変換する。そして、周波数毎に正方行列による ICA を適用して、観測信号をマイクロホン数と同じ数の信号に分解する。ICA の結果には permutation の任意性があるため、これを基底ベクトルのクラスタリングにより解決する。その詳細を示す [2] では、分散が最小であるクラスタを主要音源のクラスタとして選んでいる。本稿では、複数の主要音源に対処するため、分散がある閾値以下のクラスタを主要音源のクラスタとして選ぶ。次に、時間周波数マスクング [2] により、分離信号内に存在する背景音の残留成分をさらに抑圧する。最後に、ISTFT: Inverse STFT により時間領域での分離信号を得る。

## 3 実験結果

図 2 に示す条件で実験を行った。4 個のマイクロホン、それらに比較的近い 3 個の主要音源、背景音として 6 個の比較的遠い音源を用意した。合計 9 個のスピーカからは、それぞれ異なる英語や日本語の 6 秒間の音声を再生した。音声の組合せを変え、10 回試行した平均の分離結果を表 1 に示す。全体としての音源数がマイクロホン数よりはるかに多い状況にも関わらず、SIR: Signal-to-Interference Ratio の改善が十分に行われている。また、ICA だけでなく時間周波数マスクングを併用することで、SIR はさらに改善された。音のサンプルは [3] で聞くことができる。

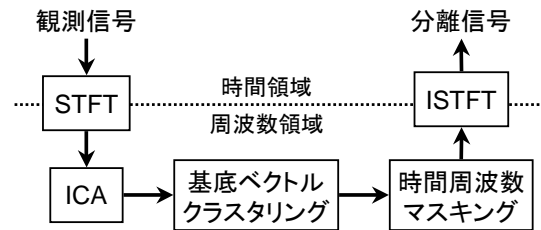


図 1 処理の流れ

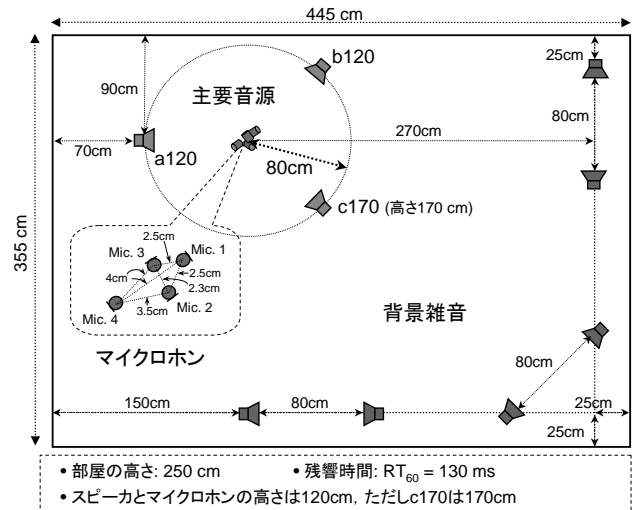


図 2 実験条件

表 1 SIR 改善量 (dB), 10 回の平均

主要音源の位置	a120	b120	c170
入力 SIR	-3.9	-3.6	-5.9
ICA のみ	12.5	13.6	14.5
ICA と時間周波数マスクング	15.1	16.5	17.6

## 4 まとめ

背景音の数が無数であるような現実的な状況でも有効に働くブラインド音源分離手法を提案した。なお、本手法のリアルタイム処理システムも実現されている。

### 参考文献

- [1] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Frequency-domain blind source separation," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Springer, Mar. 2005, pp. 299–327.
- [2] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of a dominant source signal from mixtures of many sources," in *Proc. ICASSP 2005*, vol. III, Mar. 2005, pp. 61–64.
- [3] <http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/dominant/>