6P-1

音源数同定とブラインド音源分離を同時に行う infinite ICA

柳楽 浩平 † 高橋 徹 ‡ 尾形 哲也 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 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

近年, 音声対話ロボット等を目的として, 実環境下での音声認識の研究が進められている. その場合, 接話マイクを使用する場合と異なり, マイクには目的音声とともに残響や雑音が入力される. よって, 音声認識を行う際にはマイクの入力のみから目的話者の音声を復元する必要がある. このような状況での音源分離をブラインド音源分離 (BSS) と呼ぶ. さらに, 実環境では潜在的に無数の音源が存在するため, 音源数を正確に仮定できない. よって, 実環境下での音源分離には音源数未知の状況下で BSS を実現することが要求される.

そのような BSS を行う手法としてノンパラメトリックベイズを適用した infinite Independent Component Analysis (iICA) や infinite sparse Factor Analysis (isFA) が提案されている [1]. これらの手法は, 潜在的に無数の音源が存在することを許容した BSS であり, モンテカルロ法によりモデルパラメータを推定する. しかし, 従来の iICA/isFA では実数信号の瞬時混合問題のみを扱っており, 残響をモデル化した畳み込み問題を解くことができない.

本稿では実数 iICA/isFA を用いた, 畳み込み問題に適用可能な手法を提案する. 本手法の従来法に対する位置づけは表 1 の通りである. まず, 短時間フーリエ変換 (STFT) を適用し, 周波数領域での複素信号に変換する. これにより, 畳み込み問題が複素領域での瞬時混合問題に帰着される. さらに, 複素数を等価な実数ベクトルとして表現することで, 実数 iICA/isFA で複素信号分離を実現する.

2. 実数領域での iICA/isFA とその課題

2.1 実数信号に対する iICA/isFA [1] の概要

音源数を K, マイク数を D, 信号の長さを N とする. iICA/isFA では以下の瞬時混合モデルを考える.

$$\mathbf{x}_t = \mathbf{A}(\mathbf{z}_t \odot \mathbf{s}_t) + \boldsymbol{\varepsilon}_t \tag{1}$$

ここで, $\mathbf{x}_t = [x_{1t}, x_{2t}, \cdots, x_{Dt}]^{\mathrm{T}}$ は時刻 t での混合音ベクトル, $\mathbf{s}_t = [s_{1t}, s_{2t}, \cdots, s_{Kt}]^{\mathrm{T}}$ は時刻 t での音源ベクトル, $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{Dt}]^{\mathrm{T}}$ は時刻 t でのガウス性雑音ベクトル, $\mathbf{A}$ は $D \times K$ 次元の混合行列である. $\mathbf{z}_t = [z_{1t}, z_{2t}, \cdots, z_{Kt}]^{\mathrm{T}}$ であり, z_{kt} はバイナリ変数で, 時刻 t で音源 k が ON なら 1 を, OFF なら 0 を表現する. ここでは $\mathbf{z}_t$ を ON/OFF データ列と呼ぶ. また, 演算子 $\odot$ はベクトルの要素ごと積を表す.

Knowles ら [1] の手法では, パラメータ $\mathbf{A}, \mathbf{z}_t, \mathbf{s}_t (t = 1 \cdots N)$ を, ベイズ理論に基づきマルコフ連鎖モンテカルロ法を用いて反復推定する. この時, 混合行列の事前分布は $\mathcal{N}(0,1)$ で, 音源信号の事前分布は isFA の場合 $\mathcal{N}(0,1)$, iICA の場合 $\mathcal{L}(1)$ とする. ただし, $\mathcal{N}(\mu, \sigma^2)$ は平均 μ で分散 σ^2 の正規分布を, $\mathcal{L}(\phi)$ は尺度パラメータ ϕ のラプラス分布を表す. また, 無限の音源数を仮定しているため, $\mathbf{z}_t$ はインディアンビュッフェ過程 (IBP) [2] で実現する. これらの事前分布と尤度関数から事後分布を計算し, 事後分布からサンプリングを行う. (尤度関数や事後分布の詳細は [1] を参照)

表 1: 本手法の位置付け

	瞬時混合問題	畳み込み問題
音源数未知	実数 iICA/isFA [1]	**本手法**
音源数既知	時間領域 ICA	周波数領域 ICA

2.2 周波数領域での適用における課題

実数領域での iICA/isFA は瞬時混合問題にのみ適用できるが, 残響を含んだ信号を扱えない. なぜなら残響を含んだ信号の混合モデルは畳み込みモデルで表されるからである. 畳み込みモデルについては文献 [3] でも議論されている. 本手法では, 畳み込みモデルに対応するために, iICA/isFA を周波数領域に適用する. この時, 以下の 3 点が課題となる.

1. 実数 iICA/isFA での複素信号の扱い
2. パーミュテーション問題
3. スケーリング問題

以下では, これらの課題とその解決法について考察する.

3. 周波数領域における iICA/isFA

3.1 実数 iICA/isFA での複素信号の扱い

音声信号をフーリエ変換した時, その実部と虚部に相関はないので, 実部と虚部を独立な確率変数として扱える. つまり, 各複素信号を実部と虚部に分けて独立な音源と見なす. これは複素数の積 $r = pq$ は以下の行列で表せることによる.

$$\begin{pmatrix} \mathrm{Re}\, r \\ \mathrm{Im}\, r \end{pmatrix} = \begin{pmatrix} \mathrm{Re}\, p & -\mathrm{Im}\, p \\ \mathrm{Im}\, p & \mathrm{Re}\, p \end{pmatrix} \begin{pmatrix} \mathrm{Re}\, q \\ \mathrm{Im}\, q \end{pmatrix} \tag{2}$$

これにより, 複素数の積を実数の行列積に置き換えられるので, 実数領域の iICA/isFA が適用できる. そこで得られた分離結果の信号は, 各音源信号の実部と虚部を表す. この時, 混合行列の推定において, 本来推定すべきパラメータは混合行列の各成分の実部と虚部の計 $2DK$ 個であるのに対して, この場合は計 $4DK$ 個のパラメータを推定しなければならない.

3.2 パーミュテーションとスケーリング

次に, パーミュテーション問題について考える. iICA/isFA では, 各音源の実部と虚部が順不同に出力されるので, 音源ごとにうまく並び替える必要がある. この問題をパーミュテーション問題と呼ぶ. この並び替えには実数領域の iICA/isFA の出力の一つである ON/OFF データ列を用いる. 同一音源の実部と虚部では, その音源の ON/OFF データ列の推定結果は類似すると予想されるので, 各音源の ON/OFF データ列の類似性を調べて, どの 2 つの信号が同一の信号の実部と虚部になるのかを判断する.

Simultaneous blind source separation and estimation of the number of sources using infinite ICA: *Kohei Nagira, Toru Takahashi, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

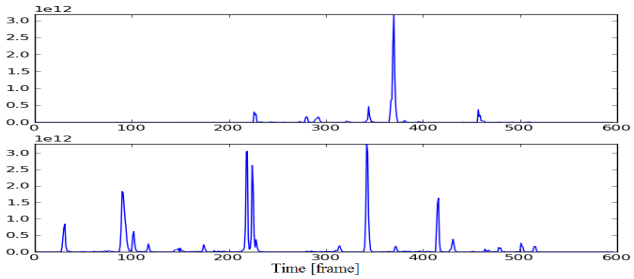

図 1: 元信号のパワー

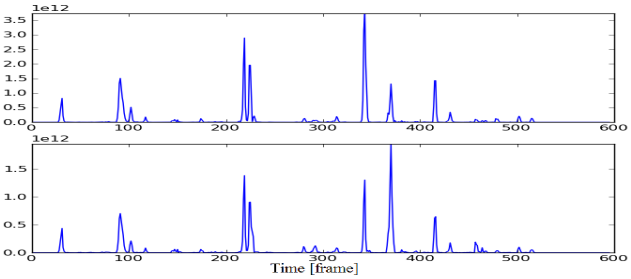

図 2: 混合後の信号のパワー

表 2: 評価実験の条件

音源数 K	2
マイク数 D	2
音声の長さ	12 秒
サンプリング周波数	16 kHz
インパルス応答	無響室で測定
STFT の窓長	1024
STFT のシフト長	320
用いたモデル	isFA
反復回数	1000

音源同士の ON/OFF データ列の類似性を表す尺度は次のようにして算出する. 調べる音源の ON/OFF データ列を $\mathbf{z}_{s_1} = [z_{s_1,1}, \cdots, z_{s_1,N}], \mathbf{z}_{s_2} = [z_{s_2,1}, \cdots, z_{s_2,N}]$ とする. ここで, $\mathbf{z}_{xor} = \mathbf{z}_{s_1} \otimes \mathbf{z}_{s_2}$ で表される $\mathbf{z}_{xor}$ を考える. 演算子 $\otimes$ は排他的論理和 (XOR) を表す. この $\mathbf{z}_{xor}$ の総和が小さいほど $\mathbf{z}_{s_1}$ と $\mathbf{z}_{s_2}$ は類似している. 各信号の組に対してこの値を算出し, 値が最も低くなる 2 つの音源同士がある音源の実部と虚部に対応する. しかし, 同一音源と判定された 2 つの信号のどちらが実部でどちらが虚部なのかは判定できない.

最後に, 残された実部・虚部の判定とスケーリング問題について考える. ここでは projection back [4] を応用する. 推定された k 番目の音源 $\widehat{\mathbf{s}}_k$ は元音源 $\mathbf{s}_k$ を用いて $\widehat{\mathbf{s}}_k = \mathbf{P}_k \Lambda_k \mathbf{s}_k$ と表される. ただし, 実部・虚部の入れ替わりを表す行列を $\mathbf{P}_k$, スケーリング行列を $\Lambda_k = \mathrm{diag}([\lambda_{k_1}, \lambda_{k_2}])$ とする. しかし, 混合モデルを考えると $\widehat{\mathbf{A}}_{dk}\widehat{\mathbf{s}}_k = \mathbf{A}_{dk}\mathbf{s}_k$ となる. ここで, $\mathbf{A}_{dk}, \widehat{\mathbf{A}}_{dk}$ は, それぞれ混合行列の音源 k とマイク d に対応する部分の真値と推定値を表す. $\mathbf{A}_{dk}\mathbf{s}_k$ はマイク d で観測される音源 k からの音声を表す. これにより実部・虚部のパーミュテーション問題及びスケーリング問題が解決される.

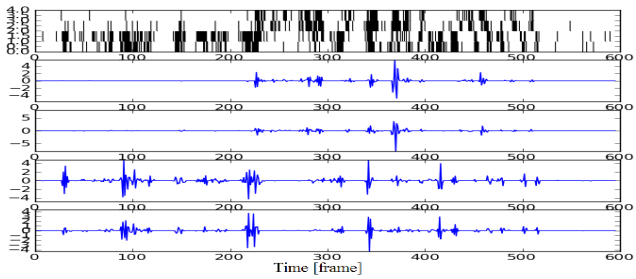

図 3: 4 つの分離信号の ON/OFF データ列と信号波形

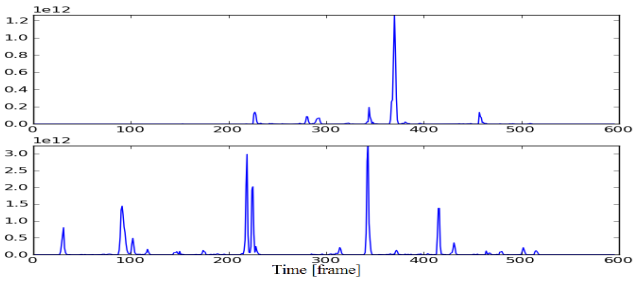

図 4: スケーリング修正後の信号のパワー

4. 評価実験

本手法の分離性能を確認するために音声信号にインパルス応答を畳み込んだデータを混合し, そのうちの一つの周波数ビンについて分離実験を行った. 実験条件を表 2 に示す.

混合前の信号のパワーを図 1 に, 混合後の信号のパワーを図 2 に, この混合した信号を実部と虚部に分け実数領域 isFA を行い, 各音源ごとにクラス分けした信号波形の推定結果を図 3 に示す. 図 3 の一番上のグラフは ON/OFF データ列の推定結果であり, 黒が ON, 白が OFF を表す. その後, スケーリングを修正した結果の信号のパワーを図 4 に示す. 図 1, 図 4 を比較すると, 混合前の信号が復元されている事がわかる. また SN 比による評価では, 分離前が 0.9dB であったのに対し分離後には 15.4dB となり, 14.5 ポイントの改善が見られた.

5. おわりに

本稿では, 実環境において音源数同定とブラインド音源分離を同時に行うシステムについて報告した. 畳み込みモデルに対応するために, 実数 iICA/isFA を周波数領域で適用し, 実音声信号の 1 つの周波数ビンでの分離実験で本手法の有効性を確認した. 今後は, 全周波数帯域を用いた分離実験, Semi-Blind 音源分離への応用などが考えられる.

謝辞 本研究の一部は科研費基盤研究 (S), 特定領域, GCOE の支援を受けた.

参考文献

[1] David Knowles *et. al*: "Infinite Sparse Factor Analysis and Infinite Indpendent Components Analysis", *Proc. of ICA*, 2007.
[2] Tom Griffiths *et. al*: "Infinite latent feature models and the Indian buffet process", *Proc. of NIPS*, 2005.
[3] Aapo Hyvärinen *et. al*: "Independent Component Analysis", Wiley-Interscience, 2001.
[4] Noboru Murata *et. al*: "An approach to blind source separation based on temporal structure of speech signals", Neurocomputing, vol.41, pp. 1-24, Oct. 2001.