

周波数領域 ICA における スペクトログラムに基づくパーミュテーション解法*

小野順貴 (東大院・情報理工)

1 はじめに

周波数領域における独立成分分析 (Independent Component Analysis; ICA) は、音響信号のような時間遅れを有する混合系に対するブラインド信号分離の有効なアプローチの1つであるが、周波数帯域毎に信号分離が行われるため、帯域毎の分離結果を同一音源毎にまとめる、いわゆるパーミュテーション問題を解く必要が生じる。これまでに、音源到来方向 [1]、帯域間の振幅相関 [2]、それらの組み合わせと調波性 [3] などを利用して、様々な手法が提案されている。

我々は、マイクロフォン間隔が大きく、観測信号間の時間同期が保証されていない分散型マイクロフォンアレイへの適用を想定し、音源到来方向を使わないパーミュテーション解法を検討している。近年提案されている、ベクトル型変数を用いた独立成分分析 [4, 5] (以下、IVA と表記する) は、パーミュテーション問題を生じない分離手法として魅力的であるが、残響時間が長い環境では、必ずしもよい分離性能を与えない場合がある。

そこで本研究では、1) IVA に用いられている音源モデルをパーミュテーションの決定に利用し、2) その際の音源モデルとして スペクトルを用いる、という手法について検討し、提案手法と実験結果について報告する。

2 周波数領域パーミュテーション問題

$Y_{t,\omega,k}$, ($1 \leq t \leq N_t$, $1 \leq \omega \leq N_\omega$, $1 \leq k \leq K$) を周波数領域 ICA によって得られた複素スペクトルとする。ただし、 t, ω, k はそれぞれ時間、周波数、分離信号を表すインデックス、 N_t, N_ω は時間フレーム、周波数 bin の総数、 K は音源数を表す。周波数領域 ICA は ω 毎に行われるため、ある k に対し $Y_{t,\omega,k}$ が全ての ω に対して同じ音源信号に対応するとは限らない。パーミュテーション問題とは、 $Y_{t,\omega,\sigma(k)_\omega}$ が ω によらず同一の音源信号に対応するようなパーミュテーション $\sigma(k)_\omega$ を求める問題である。ここで $\sigma(k)_\omega$ はパーミュテーションのインデックスを表し、例えば $K=2$ の場合には、 $(\sigma(1)_\omega, \sigma(2)_\omega) \in \{(1, 2), (2, 1)\}$ である。

3 多変数密度関数による共起の表現

IVA においては、ベクトル型変数: $\mathbf{A} = (A_1 \cdots A_N)^t$ の確率密度分布として、そのノルムである $R = \sqrt{\sum_{i=1}^N |A_i|^2}$ が優ガウス分布に従うというモデルが用いられている。

このモデルが、ベクトル型変数の要素間のどのような関係を表しているのかを確認するため、いま、要素数 2 のベクトル $\mathbf{A} = (A_1 A_2)^t$ を考え、 $R = \|\mathbf{A}\|$ が平均 0、分散 1 のラプラス分布 $(1/2)e^{-R}$ に従うと仮定する。いま、 A_1 の A_2 に対する依存性をみるために、 $A_2 = 0$ 、 $A_2 = 10$ に対する A_1 の条件付確率を求めると、

$$p(A_2 | A_1 = 0) = \frac{1}{2} e^{-|A_1|} \quad (1)$$

$$p(A_2 | A_1 = 10) = C e^{-\sqrt{A_1^2 + 10^2}} \quad (2)$$

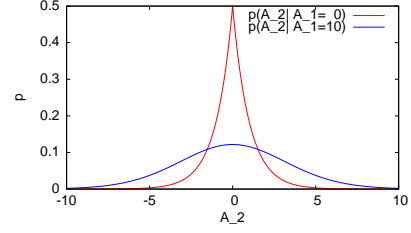


Fig. 1 A_2 の密度分布の A_1 に対する依存性

となる。ここで C は規格化定数で、 $C \approx 2681.14$ である。これらの概形を図 1 に示す。明らかに $p(A_2 | A_1 = 10)$ の方が $p(A_2 | A_1 = 0)$ より裾野が長く、 A_2 が大きな値をとる確率が高いことを表している。すなわちこの確率モデルは大まかにいえば、 A_1 が絶対値の大きな値をとる方が A_2 も絶対値の大きな値をとりやすいという、要素の共起関係を表しているともみることができる。

4 パーミュテーション解法

4.1 スペクトルの利用

IVA では、フレーム毎の複素スペクトルをベクトル型変数として扱い、前述の確率モデルに対して最尤となるように各帯域の分離行列が推定される。しかしながら振幅の同期性よりも、音の立ち上がりのような時間変化部の同期性の方が同一音源をまとめる有効な手がかりに有用なのではないかと、というのが本研究の動機である。以下では共起特徴量として、振幅の他に、その時間差分である スペクトルの利用を検討する。

$$A_{t,\omega,\sigma(k)_\omega}^{\text{amp}} = |Y_{t,\omega,\sigma(k)_\omega}|^\gamma \quad (3)$$

$$A_{t,\omega,\sigma(k)_\omega}^{\text{delta}} = \max(|Y_{t+1,\omega,\sigma(k)_\omega}| - |Y_{t-1,\omega,\sigma(k)_\omega}|, 0)^\gamma \quad (4)$$

ここで γ は、絶対値を圧縮する 0 から 1 の間の定数である。また、スペクトルの振幅を正に限っているのは、前述のモデルが基本的に $A_{t,\omega,\sigma(k)_\omega}$ の符号に依存しない確率モデルであり、このような制限をつけないと音の立ち上がりと立ち下がり同期するような誤った解を生じてしまうためである。

4.2 パーミュテーション決定アルゴリズム

式 (3)、式 (4) で定義されるスペクトログラム特徴量が同一音源で共起しやすいならば、

$$\mathbf{A}_{t,k} = (A_{t,1,\sigma(k)_1} \cdots A_{t,N_\omega,\sigma(k)_{N_\omega}})^t \quad (5)$$

$$R_{t,k} = \sqrt{\sum_{\omega=1}^{N_\omega} |A_{t,\omega,\sigma(k)_\omega}|^2} \quad (6)$$

を定義するとき、 $R_{t,k}$ が一般化正規分布

$$f(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta} \quad (7)$$

*Permutation Correction in Frequency-domain Independent Component Analysis Based on Delta Spectrogram by Nobutaka ONO (The University of Tokyo)

に従うという確率モデルを考えることができる。ここで α は分散を、 β は分布形状を表すパラメータであり、 $0 < \beta < 2$ のときに優ガウス分布となる。一般には分散や分布の性質は音源に依存すると考えられるため、 α と β は音源 k 毎に異なるパラメータと考える。ただし、 β の最尤解は解析的には得られないため、以下では β は音源毎に既知であるものとする。

以上のモデル化より、各帯域におけるパーミュテーション $\sigma(k)_\omega$ と α_k を変数とし、対数尤度関数

$$J = \sum_{t,k} \log \beta_k - \log(2\alpha_k \Gamma(1/\beta_k)) - (R_{t,k}/\alpha_k)^{\beta_k} \quad (8)$$

を反復的に増加させることにより、パーミュテーションを決定するアルゴリズムが得られる。ただしこれは組み合わせ最適化であるため、局所最適性しか保証されない点に注意する必要がある。

- Step 1: $\omega = 1$ とする。
- Step 2: 周波数帯域 ω における全てのパーミュテーション $\sigma(k)_\omega$ に対して J を計算し、 J が最大となる $\sigma(k)_\omega$ を選択し、 $Y_{t,\omega,k} \leftarrow Y_{t,\omega,\sigma(k)_\omega}$ とする。
- Step 3: $\omega \leftarrow \omega + 1$ とする。 $\omega \leq N_\omega$ であれば、Step 2 に戻る。そうでなければ、Step 4 へ。
- Step 4: 全帯域でパーミュテーションの変更がなければ終了。そうでなければ、各 k に対し、

$$\alpha_k^{\beta_k} = \frac{\beta_k}{N_t} \sum_t |R_{t,k}|^{\beta_k} \quad (9)$$

により α_k を更新し、Step 1 へ。

また以下では、式 (3)、式 (4) を共起特徴量として使い、上記のアルゴリズムでパーミュテーションを求めるアルゴリズムをそれぞれ PermAmp、PermDelta と表記する。

5 ブラインド音源分離による評価実験

5.1 実験条件

ブラインド音源分離に適用することにより、提案法の評価を行った。まず評価データはインパルス応答の畳み込みにより生成した。音源信号には、ATR 音声データベースの B セットの音声データをサンプリング周波数を 20kHz から 16kHz にダウンサンプルして用いた。インパルス応答には、RWCP 実環境音声・音響データベースのインパルス応答データの JR1(豊部屋大) というデータセットを用いた。残響時間は 0.60s、マイクロフォン-スピーカ間隔は 2m である。この中から 4 本のマイクロフォン (間隔 2.83cm) により計測されたインパルス応答を選び、等間隔線形アレイによる観測をシミュレートした。観測信号長は 10 秒に揃えた。音源数は 2 とし、1 試行の実験において、音源方向は 50 度から 130 度まで 10 度間隔の 9 方向から 2 方向、音源信号は 90 個の発話データ (話者は 9 人) から 2 個をそれぞれランダムに選んだ。

一方、短時間 Fourier 変換には Hamming 窓関数を用い、フレーム長 4096 点、フレームシフト 2048 点とした。帯域毎に白色化を行った後、補助関数型 ICA[6]20 回の反復を適用し、Projection Back によりスケールを決定した。その後、帯域毎の分離信号に対し、 $\omega = 1$ から昇順に隣接帯域間相関の総和が最大となるパーミュテーションを求め、パーミュテーションの初期値とした。分離性能は [7] に従って求めた SIR により評価を行った。

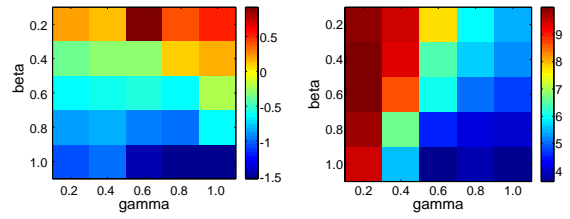


Fig. 2 PermAmp(左), PermDelta(右) の分離性能のパラメータ依存性

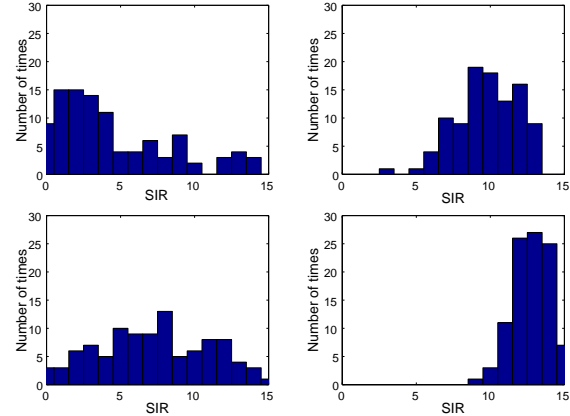


Fig. 3 PermAmp(左上), PermDelta(右上), IVA(左下), 正解パーミュテーション(右下) による分離性能のヒストグラム

5.2 実験結果

まず 20 試行のデータを用い、PermAmp、PermDelta それぞれの手法に対し、 β と γ をそれぞれ 0.2, 0.4, 0.6, 0.8, 1.0 と変化させて提案アルゴリズムを適用し、平均 SIR を求めたものを図 2 に示す。PermAmp の最適パラメータは $\beta = 0.2, \gamma = 0.6$ 、PermDelta の最適パラメータは $\beta = 0.4, \gamma = 0.2$ であった。また平均して、PermDelta の方が分離性能が高いことがわかる。

次にこれらの最適パラメータを用いて、100 試行のデータに対し提案法を適用し、得られた SIR のヒストグラムを図 3 に示す。比較のため、IVA を適用した場合と、音源信号自体を用いて正しいパーミュテーションを与えた場合の結果もあわせて示してある。IVA では、ベクトル型変数のノルムがラプラス分布に従うというモデルを用いた。PermDelta は IVA と比較しても分離性能が安定しており、平均 SIR が高いことが確認された。

今後は他の音源数、残響環境での検証を加えると共に、本研究の知見に基づき、スペクトルの共起を音源モデルとするようなベクトル型の独立成分分析について検討していく予定である。

参考文献

- [1] Kurita *et al.*, Proc. ICASSP, pp. 3140–3143, 2000.
- [2] Murata *et al.*, Neurocomputing, vol. 41, no. 1-4, pp. 1-24, 2001.
- [3] Sawada *et al.*, IEEE Trans. SAP, vol. 12, no. 5, pp. 530–538, 2004.
- [4] A. Hiroe, Proc. ICA, pp. 601–608, 2006.
- [5] Kim *et al.*, IEEE Trans. ASLP, vol. 15, no. 1, pp. 70–79, 2007.
- [6] Ono *et al.*, Proc. LVA/ICA, 2010.
- [7] Vincent *et al.*, IEEE Trans. ASLP, vol. 14, no. 4, pp. 1462–1469, 2006.