1-5-22

事前学習を用いる周波数領域Pearson-ICAの高速化*

○ 加藤比呂子(NTT 研究所), 永原裕一(明大), 荒木章子(NTT 研究所),
澤田宏(NTT 研究所), 牧野昭二(NTT 研究所)

1 はじめに

独立成分分析(Independent Component Analysis 以下 ICA)を用いた周波数領域ブラインド音源分離手法に対し， [1]では，各周波数で，ピアソン(Pearson) 分布により ICA のスコア関数をモデル化する手法を提案した。これにより，全周波数領域のスコア関数に対して同じ非線形関数を用いた従来法よりも分離精度が向上することを示した。この手法を，周波数領域 (Frequency-domain，以下 FD) Pearson-ICA ということにする。

[1]の FD-Pearson-ICA では，1) スコア関数の係数を各周波数でグリッドサーチによって求める方法と，2) スコア関数のモデル化に際し各周波数の系列から直接モーメントを求めピアソン分布型を決める方法について実験した。これらの手法から分離精度は上がったが，反面，全周波数で同じ非線形関数を適用した従来手法と比べて計算時間がかかるという点が問題となった。そこで 1)で各周波数に分解された系列のピアソン分布を調べると，幾つかの音声の組み合わせに対して，中高域の分布形状の傾向が似ていることがわかった。このことから，この帯域の分布パラメータをあらかじめ学習すれば，その帯域部分のスコア関数推定に関する計算量を節約できると考えた。

そこで 本稿では，ある周波数より低域には 2)を適用し，中高域には，1)を適用するという混合手法を提案する。1)においては，混合音声の何組かに対して，あらかじめグリッドサーチによりパラメータを推定し (この過程を事前学習とする)，混合音声の組全体で算出された平均値を用いた。これにより、分離精度の向上を維持しながら，計算時間も改善された。

2 音源分離システム

周波数領域に変換された観測信号を $\mathbf{X}(f,m)$ とすると，ICA による分離により各周波数の分離信号は

$$\mathbf{Y}(f,m)=\mathbf{W}(f)\mathbf{X}(f,m) \tag{1}$$

で得られる。分離行列 $\mathbf{W}(f)$ の推定は ICA を用いて更新式

$$\mathbf{W}_{i+1}(f)=\mathbf{W}_i(f)+\eta[\mathbf{I}-\langle\Phi(\mathbf{Y}(f,m))\mathbf{Y}^H(f,m)\rangle]\cdot\mathbf{W}_i(f)$$

$\langle x(f,m)\rangle$: m についての平均操作、 (2)

η : 学習ステップサイズ、i : 更新回数

によりおこなう[1]。ここで $\Phi(\cdot)$ はスコア関数で，周波数領域に分解された系列の分布 p が既知であれば，

$$\Phi(Y)=-\frac{p'(|Y|)}{p(|Y|)}\exp(j\angle Y)$$

$p'=dp(x)/dx$ (3)

Y : 複素数， $|\cdot|$: 絶対値， $\angle$: 偏角

となる。実際には， p は既知ではないため，従来は，全周波数領域に対して同じ非線形関数 *tanh* 関数や generalized Gaussian distribution(GGD)等を用いてモデル化した。

3 ピアソン分布の適用

3.1 スコア関数の定義

ピアソン分布系は微分可能な分布で、データの分布 $p(\cdot)$ の微分方程式は、確率変数の多項式で定義されている。多項式の係数は，データのサンプルモーメントから推定される。本稿で扱う分離システムでは，周波数領域に変換された個々の複素系列の分布を対象とするため，[1]により微分方程式は次のように示される:

$$-\frac{p'(Y)}{p(Y)}=\frac{b_0(f)+b_1(f)|Y|}{c_0(f)+c_1(f)|Y|+c_2(f)|Y|^2}\exp(j\angle Y)$$

Y: 複素数 (4)

$b_0(f),b_1(f),c_0(f),c_1(f),c_2(f)$:

周波数fにおける多項式の係数

この式は，(3)式のスコア関数に対応し，多項式の係数は各周波数の系列から推定される。これにより，各周波数において分布形状に見合ったスコア関数を推定することが可能となる。

3.2 スコア関数の推定方法

旧手法[1]は，2 つのスコア関数推定方法を提案した：1) (2)式中 $\mathbf{I}-\langle\Phi(\mathbf{Y})\mathbf{Y}^H\rangle$ の非対角要素を最小

---

* Efficient frequency-domain Pearson-ICA with learning parameters by KATO, Hiroko (NTT), NAGAHARA, Yuichi (Meiji Univ.), ARAKI, Shoko(NTT), SAWADA, Hiroshi (NTT) and MAKINO, Shoji (NTT).

にする係数をグリッドサーチで決めた（**手法1**）。 2) 系列のサンプルモーメントから各周波数のピアソン分布の型を判定し，その分布型に対応する分布パラメータを個々に算出した（**手法2**）。

本稿では，どの音声の組み合わせに対しても，低域ではピアソン分布の型や，同一の型でもその形状にばらつきがあり，また，中高域の周波数帯では同一の分布型で，各周波数における形状のばらつきは小さいことに着目し，以下のように手法1と2を組み合わせた方法を提案する：

手法3 手法1をあらかじめあるデータの組み合わせに適用し(4)式の係数を事前学習させ，平均値 $\{\bar{b}_0(f),\bar{b}(f)_1,\bar{c}_0(f),\bar{c}_1(f),\bar{c}_2(f)\}$ を算出する。ICAによる分離時には，ある周波数 f_0 に対し， $f \le f_0$ には系列それぞれに対して手法2で分布パラメータを推定， $f > f_0$ には事前学習による平均値を適用する。

手法4 FD-Pearson-ICA の性能の上限を調べるために，既知の室内インパルス応答，原音声データを用い(unblind) SIR(後述)を最良にする(4)式の係数をグリッドサーチする手法を用いる。手法3と同様に事前学習し，係数の各平均値を算出し， $f \le f_0$ には手法2を、 $f > f_0$ には算出された平均値を適用する。

4　実験と結果

4.1 実験条件

2マイク2音源の場合について実験をおこなった。マイクロホンアレイに対する話者の位置は、右を0度、正面を90度としたときにそれぞれ30度と150度とした。残響時間 130ms の部屋で実測したインパルス応答を音声信号に畳み込み、加算を行い、マイクの観測信号を模擬した。サンプリング周波数は8kHz、STFTのフレームサイズは512である。

分離精度の良さに対する評価には Signal-to-interference-ratio (以下 SIR)を用いた。

f_0 は周波数ビン60から100の間に定めた。

4.2 実験結果

分離精度

事前学習に用いた音声データの組み合わせは2組で，その他2組を加え全4組に対し各手法を適用した。比較する従来法としては，全周波数領域に同じ *tanh* 関数を用いる場合，各周波数でGGDのスケーリングパラメータを推定した関数を用いる場合を考えた。4組のデータから得られたSIR値を平均した結果を Fig.1 に示す。手法4の結果は unblind なため最も分離精度が高く提案手法の性能の上限値とみなせる。従来法と比較すると，提案手法は1-2.5dB以上の向上を示した。また，手法1と3を比較すると，事前学習による平均値を用いても分離精度を落とさないことが確認できた。さらに，事前学習に使われていないデータに対しても、分離精度の向上を確認できた。手法4で示した unblind な条件であらかじめ何組かの混合音から代表的な平均値を事前学習し，他の組み合わせに適用することで，分離精度が向上する可能性を示すことができた。

計算時間

MATLAB® の *profile* コマンドにより，計算時間を計測した。Table1 に，従来の *tanh* 関数を用いた場合と，手法1-3の計算時間をまとめた。これによると，*tanh* 関数を用いる場合は圧倒的に速く，手法1はグリッドサーチによりかなり時間がかかっていたが，事前学習を用いた提案手法3では，FD-Pearson-ICA としての計算時間が大幅に改善され，全周波数領域で同じ *tanh* 関数を用いる従来法と同程度まで短縮されることが明らかになった。

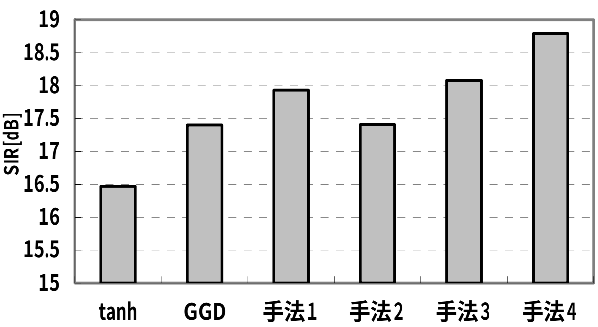

Fig. 1 実験結果

Table 1　計算時間

適用した手法	時間（秒）
非線形関数	10.38
手法1	6888.66
手法2	56.11
手法3	13.07

5　おわりに

FD-Pearson-ICA による音源分離手法で，[1]で提案された2つの手法に，事前学習を導入する手法を提案した。従来法よりも高い分離精度を維持したまま処理計算も短縮された。

参考文献

[1]　加藤他，音講論（秋）, 593-594, 2005