

ICA の音響処理への応用

○池田 思朗 (さきがけ研究 21, 科技団) *

1 はじめに

ICA (Independent Component Analysis, 独立成分解析) と呼ばれる手法が, 近年, 注目されるようになった. この手法は良く知られた PCA (Principal Component Analysis, 主成分分析) と同様, 多次元信号の解析手法である. PCA は信号の 2 次の相関を用い, 信号を無相関にする変換を求めが, ICA は高次の統計量, あるいは時間的な相関に基づく独立性により信号を分離する. 正規分布以外の確率分布では一般に無相関と独立性とは一致しない. したがって PCA と ICA は異なる結果を与える.

この ICA が近年注目されるようになったのは, 問題の単純さとその応用の可能性からである. ICA では信号源が独立であるという仮定のみを置き, そのパラメータの更新則は比較的簡単である. 一方, 応用範囲は広く, 脳などの生理計測データの解析, または音響信号の分離, 携帯電話などの電波信号の分離などにおいて興味深い結果を与えている.

ここでは, ICA の応用のなかから, 音響信号の分離に関し, 我々の行なっているアプローチについて説明したい.

2 ICA の問題

本節では ICA の問題について説明する. 信号源が $s(t) = (s_1(t), \dots, s_n(t))^T$, ($t = 0, 1, 2, \dots$) であるとすると, $s(t)$ の各成分の平均は 0, 各成分は互いに独立であるとする.^T は転置を表わす. 一方観測は, $x(t) = (x_1(t), \dots, x_m(t))^T$ ($t = 0, 1, 2, \dots$) で表すものとする. これは m 個のセンサーで観測された信号である. センサーの数 m と信号源の数 n は必ずしも一致しない. $s(t)$ と $x(t)$ との間に,

$$x(t) = As(t), \quad (1)$$

*"ICA for sound signal separation", Shiro Ikeda (PRESTO, JST)

という線型の関係を仮定する. A は $m \times n$ の実数行列である.

ICA の目的は $s(t)$ の確率分布の形と A に関する知識を持たずに $x(t)$ を n 個の独立な信号成分に分離することである. $n \leq m$ ならば解は存在する. すなわち $n \times m$ の実数行列 W が存在し,

$$y(t) = Wx(t), \quad (2)$$

によって互いに独立な $y(t)$ を再構成できる. $WA = I_n$ (I_n は n 次の単位行列) ならば $y(t)$ と $s(t)$ は一致するが, $y(t)$ の成分の順番の入れ替えと大きさは独立性には影響しないことから, この 2 つの任意性は許容する.

一方, 音響信号の分離の問題は,

$$\begin{aligned} x(t) &= A(t) * s(t) \\ x_i(t) &= \sum_k a_{ik}(t) * s_k(t), \quad (3) \\ a_{ik}(t) * s_k(t) &= \sum_{\tau=0}^{\infty} a_{ik}(\tau) s_k(t - \tau), \end{aligned}$$

のように畳み込みを用いて定義する. この問題では $y(t)$ の各成分が独立となるようなフィルター $W(t)$ を求めることが目的となる.

$$y(t) = W(t) * x(t). \quad (4)$$

3 ICA の解法

3.1 確率分布の独立性に基づく分離法

本節では ICA の解法について, 概略を述べる [9]. 代表的な手法の一つは確率分布の独立性に基づく手法である. この手法では, 信号源の各成分 $s_i(t)$ が正規分布ではない確率分布にしたがっていると仮定する. ただし確率分布は未知である. $y(t)$ の同時分布密度関数を, $p(y) = p(y_1, \dots, y_n)$ と置く. 行列 W によって観測信号 $x(t)$ が正しく分離できたならば y の各成分 y_i は独立にな

り, $p(\mathbf{y}) = \prod_{i=1}^n p(y_i)$ とかける. そこで, $p(\mathbf{y})$ と $\prod_{i=1}^n p(y_i)$ とが一致するように W を求める. 評価関数としては $p(\mathbf{y})$ と $\prod_{i=1}^n p(y_i)$ との間の Kullback-Leibler divergence を小さくするように W を求める. K-L divergence の定義は,

$$KL(W) = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^n p(y_i)} d\mathbf{y} \\ = -H(\mathbf{Y}; W) + \sum_{i=1}^n H(Y_i; W). \quad (5)$$

$H(\mathbf{Y}; W)$ は同時分布のエントロピー, $H(Y_i; W)$ は周辺分布のエントロピーである. (5) 式の $H(\mathbf{Y}; W)$ と $H(Y_i; W)$ は $p(x)$ と W によって以下のように書き直せる.

$$H(\mathbf{Y}; W) = H(\mathbf{X}) + \log |W|, \quad (6)$$

$$H(Y_i; W) = - \int p(x) \log p(y_i) dx. \quad (7)$$

信号源が正規分布でないという仮定から $KL(W)$ は $p(y_i)$ が互いに独立な場合に限り 0 となる. W を求めるには $KL(W)$ の W に関する勾配を求め, 最急降下法を行えば良い.

$$\Delta W \propto - \frac{\partial KL(W)}{\partial W} = (W^T)^{-1} - E_x[\varphi(\mathbf{y})\mathbf{x}^T] \\ = (I - E_x[\varphi(\mathbf{y})\mathbf{y}^T]) (W^T)^{-1} \quad (8)$$

$$\varphi(\mathbf{y}) = - \left(\frac{\partial \log p(y_1)}{\partial y_1}, \dots, \frac{\partial \log p(y_n)}{\partial y_n} \right)^T \quad (9)$$

と更新していくことで W を求められる. 収束性に関しては正定値行列を掛けても構わないことから $W^T W$ を掛けて [3],

$$\Delta W \propto (I - E_x[\varphi(\mathbf{y})\mathbf{y}^T]) W \quad (10)$$

とした方が, 計算量も少なく, 収束も速い [2].

この更新則は $\varphi(\mathbf{y})$ を含んでおり, 密度関数の形が分らなければ計算できない. しかし, 適当なパラメトリックな非線型関数や統計的な展開法によって近似しても正しく W が求まる [1]. 多項式を用いてこれを近似する手法 [6, 10] や, sigmoid 関数を用いているもの [5]. 漸近展開を用いるもの [3, 7] などが有名である. 経験的には, 正規分布より裾が“重い”(sub-Gaussian) 場合は多項式などで近似し, 正規分布より裾が“軽い”場合 (super-Gaussian) Sigmoid 関数などで近似するのがよい.

3.2 時間構造に基づく分離法

音声信号では, 時間構造に基づく分離法を用いることができる. Molgedey と Schuster は, 自己相関関数に基づく手法 [12] を提案している. 信号がエルゴード的であるとし, 各信号源のスペクトル密度が異なるとする. 観測データの相関関数は信号源の独立性より,

$$\langle \mathbf{x}(t)\mathbf{x}(t+\tau)^T \rangle = A \langle \mathbf{s}(t)\mathbf{s}(t+\tau)^T \rangle A^T \\ = A \begin{pmatrix} R_{s_1}(\tau) & & 0 \\ & \ddots & \\ 0 & & R_{s_n}(\tau) \end{pmatrix} A^T, \quad (11)$$

とかける. $\langle \cdot \rangle$ は $\mathbf{x}(t)$ の確率分布での平均を表わし, $R_{s_i}(\tau)$ は信号源 $s_i(t)$ の自己相関関数である. 正しく W が求まるれば $\mathbf{y}(t)$ の相関関数は,

$$\langle \mathbf{y}(t)\mathbf{y}(t+\tau)^T \rangle = \langle W A \mathbf{s}(t) (W A \mathbf{s}(t+\tau))^T \rangle \\ = \begin{pmatrix} \lambda_1^2 R_{s_{1'}}(\tau) & & 0 \\ & \ddots & \\ 0 & & \lambda_n^2 R_{s_{n'}}(\tau) \end{pmatrix}, \quad (12)$$

となる. $1', 2', \dots, n'$ は置換の任意性を, λ_i は大きさの任意性を考慮したものである. ノイズがなく各信号が完全に独立であるならば, 最適な W は全ての τ に対し $\mathbf{y}(t)$ の相関関数を対角行列とする ((12) 式). したがって $\mathbf{x}(t)$ の相関関数を複数の時間差 τ_i に対して求め, 同時に対角化する行列として W を求めれば良い.

$$W \langle \mathbf{x}(t)\mathbf{x}(t+\tau_i)^T \rangle W^T = \Lambda_i, i = 1, \dots, r.$$

Λ_i は対角行列である. 行列の数が 2 つの場合はこれは行列の固有値問題として一意に W が求まる [12]. ノイズのある場合は 2 つ以上の τ_i を選び, 相関行列を同時対角化する解を求めたほうがロバストな解が得られるが, 完全には対角化ができないので, 何か別の評価関数を考える必要がある [14, 16].

3.3 音響信号の分離

時間遅れのある混合に対しては, ICA の解法を拡張する必要がある. 1 つの拡張は, 分離行列を

線型な FIR フィルターとし、ICA の問題を解くときに用いた評価関数を基に FIR フィルターの係数を求める方法である [8, 11] .

もう 1 つの拡張としては、信号を時間周波数方向に展開し、各周波数軸で ICA を行なう方法がある [14, 15] . (4) 式の $x(t)$, $s(t)$, $A(t)$ を Fourier 変換したものを ω を周波数としてそれぞれ $\hat{x}(\omega)$, $\hat{s}(\omega)$, $\hat{A}(\omega)$ と置くと、

$$\hat{x}(\omega) = \hat{A}(\omega)\hat{s}(\omega),$$

という関係が成り立つ . 特に音声信号は数 10msec では定常と見なせるが、それ以上長い時間では定常では無いと考えられるので、信号が短い時間ではある種の定常性があり、長い時間では非定常性が強いと仮定し、近似的に、

$$\hat{x}(\omega, t_s) = \hat{A}(\omega)\hat{s}(\omega, t_s), \quad (13)$$

という関係が成り立つとする . $\hat{x}(\omega, t_s)$ と $\hat{s}(\omega, t_s)$ は $x(t)$, $s(t)$ を windowed Fourier 変換したものである . 周波数を固定するとこの式は ICA の問題と等しい . すなわち各周波数で独立に ICA を行ない、その結果をまとめれば分離できる .

しかし、このままでは分離した結果を周波数でまとめ、時間信号に復元するときに問題が生じる . 大きさと順番の任意性があることから、単に周波数毎に並べても周波数毎に大きさと順番の入れ違いが起ってしまうからだ . この 2 つの問題を解決すれば、以上の方法で時間遅れを含む ICA を行なえる . 我々はこの 2 つの問題を解く手法を提案している [14] . まず、何らかの方法で、各周波数での分離が完了したとする . 我々は 3.1 節の方法に基づいた方法 [13] と 3.2 節の時間相関を用いたもの [14] とを提案している . いずれにせよ、各周波数での分離行列が $B(\omega)$ として求まったとすると、

$$\hat{u}_\omega(t_s) = B(\omega)\hat{x}_\omega(t_s).$$

として各周波数での分離された信号が得られる . まず、大きさと順番を除くため、 $B(\omega)^{-1}$ を用い、 $i = 1, \dots, n$ に対し、 $\hat{v}_\omega(t_s; i)$ を定義する .

$$\hat{v}_\omega(t_s; i) = B(\omega)^{-1}(0 \dots 0, \hat{u}_{i,\omega}(t_s), 0 \dots 0)^T.$$

すると、 $\hat{v}_\omega(t_s; i)$ は、 $\hat{x}_\omega(t_s) = \sum \hat{v}_\omega(t_s; i)$ であることから、信号 i が、元のセンサー上でどの程度の大きさであったかを示すことが分る . したがって、 $\hat{v}_\omega(t_s; i)$ には大きさの任意性は無くなっている .

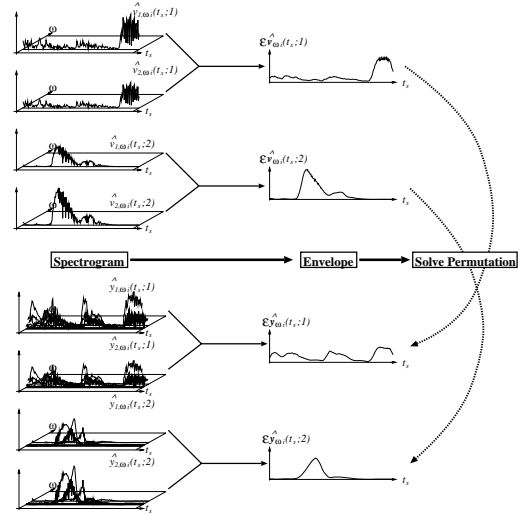


図 1: 順番の任意性を除く

残る問題は順序の入れ換えである . この問題を解決するためには、何らかの情報を用いる必要がある . 我々は異なる周波数であってもそのエンベロープはある程度類似しているだろうという仮定を置いた . エンベロープは次のように定義した .

$$\mathcal{E}\hat{v}_\omega(t_s; i) = \frac{1}{2M} \sum_{t'_s=t_s-M}^{t_s+M} \sum_{k=1}^n |\hat{v}_{k,\omega}(t'_s; i)|,$$

ここで M は正の定数である . $\hat{v}_{k,\omega}(t_s; i)$ は $\hat{v}_\omega(t_s; i)$ の k 番目の成分である . 互いがどのくらい似ているかは、次の内積を用いて、

$$\begin{aligned} \mathcal{E}\hat{v}_\omega(i) \cdot \mathcal{E}\hat{v}_{\omega'}(k) &= \sum_{t_s} \mathcal{E}\hat{v}_\omega(t_s; i) \mathcal{E}\hat{v}_{\omega'}(t_s; k), \\ \|\mathcal{E}v_\omega(i)\| &= \sqrt{\mathcal{E}\hat{v}_\omega(i) \cdot \mathcal{E}\hat{v}_\omega(i)}, \end{aligned}$$

次の式で定義する .

$$\text{sim}(\omega) = \sum_{i \neq k} \frac{\mathcal{E}\hat{v}_\omega(i) \cdot \mathcal{E}\hat{v}_\omega(k)}{\|\mathcal{E}\hat{v}_\omega(i)\| \|\mathcal{E}\hat{v}_\omega(k)\|}. \quad (14)$$

この内積を用いて、図 1) のように順番の任意性を解くことにする [14] .

これらの手法を用い、フーリエ逆変換によって、時間信号として、音響信号を分離することができる .

4 考察

本稿では、我々が提案している音響信号の分離に対する ICA の応用の概略について述べた。

ICA を時間軸のみで考え、FIR の係数を求める手法に比べ、時間周波数に展開し、その中で各周波数毎に分離を考える手法は、最近特に注目されるようになってきた。その理由の一つは計算量が少ないことが挙げられるだろう。我々も、実時間信号分離への応用も実現を行なっている [4]。

この手法の問題点の一つにはパラメータの値の選択である。サンプリング周波数、FFT の点数、窓関数の移動幅、エンベロープの計算のための Moving Average を取る時間幅などである。これらの選択は、試行錯誤の中から得た知見から求めている。このパラメータの最適化は今後の課題の一つである。

また、音響信号の性質によっては、分離ができない場合もある。片方の信号がある周波数においてパワーを持たなければ、その周波数では分離行列を求めることができない。また、窓関数を用いて時間周波数空間に信号を展開しているが、環境の持つ反射などが極端に長い時間残る場合には、(13) 式が成り立たず、窓関数の長さ以上の FIR の係数は推定できない。

課題は残るが、ICA を音声分離に応用することは大きな可能性を持っていると考えている。参考までに、我々の行なった実験に関しては、<http://www.mns.brain.riken.go.jp/~shiro/> から聞くことができる。

参考文献

- [1] S. Amari and J.-F. Cardoso. Blind source separation – semiparametric statistical approach. *IEEE Trans. SP*, 45(11):2692–2700, 1997.
- [2] S. Amari, T. Chen, and A. Cichocki. Stability analysis of learning algorithms for blind source separation. *Neural Networks*, 10(8):1345–1351, 1997.
- [3] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, Vol. 8, pp. 757–763. The MIT Press, Cambridge MA, 1996.
- [4] F. Asano and S. Ikeda. Evaluation and real-time implementation of blind source separation system using time-delayed decorrelation. In *Proceedings of ICA2000*, pp. 411–415, 2000.
- [5] A. J. Bell and T. J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.
- [6] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEEE-Proceedings-F*, 140(6):362–370, 1993.
- [7] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.
- [8] S. C. Douglas and A. Cichocki. Neural networks for blind decorrelation of signals. *IEEE Trans. SP*, 45(11):2829–2842, 1997.
- [9] 池田. 独立成分解析の信号処理への応用. 計測自動制御学会「計測と制御」, 38(7):461–467, 1999.
- [10] C. Jutten and J. Herault. Separation of sources, part i. *Signal Processing*, 24(1):1–10, 1991.
- [11] M. Kawamoto, K. Matsuoka, and N. Ohnishi. A method of blind separation for convolved non-stationary signals. *Neurocomputing*, 22(1-3):157–171, 1998.
- [12] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72(23):3634–3637, 1994.
- [13] N. Murata and S. Ikeda. A on-line algorithm for blind source separation on speech signals. In *Proceedings of NOLTA'98*, pp. 923–926, 1998.
- [14] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. to appear in *Neurocomputing*.
- [15] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22(1-3):21–34, 1998.
- [16] A. Ziehe and K.-R. Müller. TDSEP – an efficient algorithm for blind separation using time structure. In *Proceedings of ICANN'98*, pp. 675–680, 1998.