解説　両耳処理と独立成分解析*

池田思朗 (九州工業大学 & 科技団)**

1. はじめに

なぜ耳は2つあるのか，とはなかなか難しい質問である．今回ここに投稿させて頂く機会を得たが，独立成分解析 (Independent Component Analysis: 以下 ICA) の立場から何か言えないか，という委員からの御意見である．

ICA とは，多次元信号を独立な成分に分解するため統計的手法である．簡単に想像のつく通り，このような理論的な統計的手法から，冒頭の哲学的とも取れる質問に答えることは難しい．

本稿では，ICA の立場から両耳処理にあたる処理としてどのようなことが可能であるか，そしてそれが耳で行なっている処理とどのような点で類似，あるいは異なっているかを述べたい．特に複数の音声を同時に観測した場合に，それを分離する手法，いわゆるカクテルパーティー効果について考える．

2. ICA の問題

2.1. 基本 ICA

本節では ICA の問題について説明する．ICA の基本的なアイディアは 1980 年代の終りに Jutten と Herault によって提案された [11]．その後理論的にセミパラメトリック推定法の枠組から整備され [1, 5]，幾つかの具体的なアルゴリズムが提案された [3, 4, 8]．応用においても様々な分野で用いられ始めている．脳計測データの解析では，データ解析のための有効な結果が得られている [10, 12]．画像処理においても，視覚野における細胞の持つ特徴と ICA の結果との類似性が指摘されている [7, 15]．音響の分野においても，ICA による分離は 1 つの注目されているトピックである [6, 14, 17]．

本稿では音の分離，特に音声の分離を中心に扱うが，まず基本となる ICA の問題を示す．

信号源が次のベクトルで与えられるとする．

$$\boldsymbol{s}(t) = (s_1(t), \ldots, s_n(t))^T \quad t = 0, 1, 2, \ldots$$

$\boldsymbol{s}(t)$ の各成分の平均は 0，各成分は互いに独立であるとする．また，信号は正規分布ではないある確率分布から発生しているとする．^{T} は転置を表わす．観測は，

$$\boldsymbol{x}(t) = (x_1(t), \ldots, x_m(t))^T \quad t = 0, 1, 2, \ldots$$

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だと考えればよい．センサーの数 m と信号源の数 n は必ずしも一致しない．ここで $\boldsymbol{s}(t)$ と $\boldsymbol{x}(t)$ との間に，

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t), \tag{1}$$

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．BSS の問題は $\boldsymbol{s}(t)$ の確率分布の形と A に関する具体的な知識を持たずに $\boldsymbol{x}(t)$ を n 個の独立な信号成分に分離することである．

$n \leq m$ ならば線形な解が存在する．すなわち，ある $n \times m$ の実数行列 W が存在し，

$$\boldsymbol{y}(t) = W\boldsymbol{x}(t), \tag{2}$$

によって互いに独立な $\boldsymbol{y}(t)$ を再構成できる．$WA = I$ (I は $n \times n$ の単位行列) となれば $\boldsymbol{y}(t)$ と $\boldsymbol{s}(t)$ は一致する．しかし，$\boldsymbol{y}(t)$ の成分の順番を入れ替えても独立性は保たれ，各成分の大きさも独立性には影響しないことから amplitude と permutation の 2 つの任意性は許容した上で独立な信号成分に分離できればそれを解とする．

2.2. 畳み込み混合の問題

一方実空間で音を複数のマイクロフォンで録音した，あるいは両耳で聞いた場合の観測信号は

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

* Binaural processing and independent component analysis

** Shiro Ikeda (Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 808–0196) *shiro@brain.kyutech.ac.jp*

のように畳み込みを用いて定義できる．

この問題に対しても ICA により独立性に基づく分離法は有効であることが確かめられている．この問題では，$\boldsymbol{y}(t)$ の各成分が独立となるようなフィルター $W(t)$ を求めることで信号を分離できる．

$$\boldsymbol{y}(t) = W(t) * \boldsymbol{x}(t) \tag{4}$$

この問題を特に Blind Source Decomposision と呼ぶ場合もある．解は線形フィルタの設計を行なうことと等しく，推定するパラメータの数が増えることから，一般に問題が難しくなる．

3. ICA の解法

3.1. 確率分布の独立性に基づく解法

本節では ICA の基本問題に対する 2 つの解法について述べる．ICA 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．今，$\boldsymbol{y}(t)$ の同時分布密度関数を，$p(\boldsymbol{y}) = p(y_1, \cdots, y_n)$ と置く．行列 W によって観測信号 $\boldsymbol{x}$ が正しく分離できたならば $\boldsymbol{y}$ の各成分 y_i は独立になる．このとき $p(y_i)$ を y_i についての周辺分布密度関数とすれば $p(\boldsymbol{y})$ はこの周辺分布密度関数の積として $p(\boldsymbol{y}) = \prod_{i=1}^{n} p(y_i)$ とかける．そこで，$p(\boldsymbol{y})$ と $\prod_{i=1}^{n} p(y_i)$ とが一致するように W を求めるのがこの手法である．多くの手法では評価関数として $p(\boldsymbol{y})$ 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 の定義は次の通りである，

$$\begin{aligned} KL(W) &= \int p(\boldsymbol{y}) \log \frac{p(\boldsymbol{y})}{\prod_{i=1}^{n} p(y_i)} d\boldsymbol{y} \\ &= -H(\boldsymbol{Y}; W) + \textstyle\sum_{i=1}^{n} H(Y_i; W). \end{aligned}$$

$H(\boldsymbol{Y}; W)$ 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．$p(\boldsymbol{y})d\boldsymbol{y} = p(\boldsymbol{x})d\boldsymbol{x}$，$p(\boldsymbol{y}) = p(\boldsymbol{x})/|W|$ ($|W|$ は W の行列式) に注意すると，$H(\boldsymbol{Y}; W)$ と $H(Y_i; W)$ は $p(\boldsymbol{x})$ と W によって書き直せる．

$$\begin{aligned} H(\boldsymbol{Y}; W) &= H(\boldsymbol{X}) + \log |W|, \\ H(Y_i; W) &= -\int p(\boldsymbol{x}) \log p(y_i) d\boldsymbol{x}. \end{aligned}$$

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，最急降下法を行なえば良い．

$$\begin{aligned} \Delta W &\propto -\frac{\partial KL(W)}{\partial W} \\ &= \left(I - E_{\boldsymbol{x}}[\varphi(\boldsymbol{y})\boldsymbol{y}^T]\right)(W^T)^{-1} \end{aligned} \tag{5}$$

$$\varphi(\boldsymbol{y}) = -\left(\frac{\partial \log p(y_1)}{\partial y_1}, \ldots, \frac{\partial \log p(y_n)}{\partial y_n}\right)^T \tag{6}$$

と更新していくことで W を求められる．計算上 (5) 式中の逆行列 $(W^T)^{-1}$ が問題となる．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，

$$\Delta W \propto \left(I - E_{\boldsymbol{x}}[\varphi(\boldsymbol{y})\boldsymbol{y}^T]\right) W \tag{7}$$

を新たな学習則とする．この方が計算量も少なく，収束も速い．信号に強定常性の仮定が置ける場合，アンサンブル平均を時間平均に置きかえ，η を正の定数とし，データが観測される毎に (8) 式にしたがってパラメタを更新すれば収束点として W が得られる．

$$W_{t+1} = W_t + \eta \left(I - \varphi(\boldsymbol{y}(t))\boldsymbol{y}(t)^T\right) W_t. \tag{8}$$

この更新則は $\varphi(\boldsymbol{y})$ を含んでおり，厳密には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 W は正しく求まる [1)]．近似の具体的な方法によって，様々な手法が提案されている．一般に，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とされている．

3.2. 時間構造に基づく分離法

音声信号のように強定常ではない信号を考える場合には，確率分布の独立性に基づく場合と異なり，2 次の統計量のみで信号の分離が可能である．ただし，時間構造を用いるなど，他の情報を併用する．時間構造に基づく手法にも複数の手法があるが，ここでは Molgedey と Schuster の自己相関関数に基づく手法 [13)] を紹介する．

信号にエルゴード性を仮定し，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，

$$\left\langle \boldsymbol{x}(t)\boldsymbol{x}(t+\tau)^T \right\rangle = A \left\langle \boldsymbol{s}(t)\boldsymbol{s}(t+\tau)^T \right\rangle A^T$$
$$= A \begin{pmatrix} R_{s_1}(\tau) & & 0 \\ & \ddots & \\ 0 & & R_{s_n}(\tau) \end{pmatrix} A^T, \qquad (9)$$

とかける．$\langle\cdot\rangle$ は $\boldsymbol{x}(t)$ 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 を求めたとすると $\boldsymbol{y}(t)$ の相関関数は，

$$\left\langle \boldsymbol{y}(t)\boldsymbol{y}(t+\tau)^T \right\rangle = \left\langle (WA\boldsymbol{s}(t))\,(WA\boldsymbol{s}(t+\tau))^T \right\rangle$$
$$= \begin{pmatrix} \lambda_{1'}^2 R_{s_{1'}}(\tau) & & 0 \\ & \ddots & \\ 0 & & \lambda_{n'}^2 R_{s_{n'}}(\tau) \end{pmatrix}, \qquad (10)$$

となる．$1', 2', \ldots, n'$ は $1, 2, \ldots, n$ の置換を表し，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 $\boldsymbol{y}(t)$ の相関関数を対角行列とする ((10) 式)．したがって $\boldsymbol{x}(t)$ 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．

$$W\langle \boldsymbol{x}(t)\boldsymbol{x}(t+\tau_i)^T\rangle W^T = \Lambda_i, \quad i = 1, \ldots, r,$$

Λ_i は対角行列である．$r = 2$ の場合は行列の固有値問題に帰着され，代数的に一意に W が求まる [13]．ノイズのある場合には 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．ただし，完全には対角化できないので，適当な評価関数を定義するのが一般的である [18]．

4. 畳み込み混合信号の分離

時間遅れのある混合に対しては，前節の解法では逆フィルタの推定はできない．この問題に対する 1 つの解法は，分離行列を線型な FIR フィルターとして拡張し，BSS の問題を解くときに用いた評価関数を基に逆フィルタの係数を求める方法である [6]．

しかし，この方法からは両耳の処理とは類似点を見ることができない．本稿では別のアプローチについて説明する．すなわち，信号を時間周波数方向に展開し，各周波数で ICA を行なう方法である [14, 17]．(4) 式の $\boldsymbol{x}(t)$, $\boldsymbol{s}(t)$, $A(t)$ を Fourier 変換したものを ω を周波数としてそれぞれ $\hat{\boldsymbol{x}}(\omega)$, $\hat{\boldsymbol{s}}(\omega)$, $\hat{A}(\omega)$ と置くと，

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

という関係が成りたつ．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く，$A(t)$ の時間応答も長くないとし，近似的に，

$$\hat{\boldsymbol{x}}(\omega, t_s) = \hat{A}(\omega)\hat{\boldsymbol{s}}(\omega, t_s), \qquad (11)$$

という関係が成り立つとする．$\hat{\boldsymbol{x}}(\omega, t_s)$ と $\hat{\boldsymbol{s}}(\omega, t_s)$ は $\boldsymbol{x}(t)$, $\boldsymbol{s}(t)$ 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 の問題を解き，その結果をまとめれば分離できるように見える．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．amplitude と permutation の任意性があることから，単に周波数毎に並べても大きさがまちまちであり，順番の入れ違いが起ってしまう．

この問題を解決する方法として，我々が 1 つの提案をしている [9, 14]．

Amplitude の任意性

ICA では音源についての知識が不足していることから，各音源の大きさは推定できない．しかし，観測点における，各音源の大きさを推定することはできる．周波数 ω で ICA を行ない，分離のための行列 $\hat{W}(\omega)$ を求めたとしよう．

$$\hat{\boldsymbol{y}}(\omega, t_s) = \hat{W}(\omega)\hat{\boldsymbol{x}}(\omega, t_s)$$

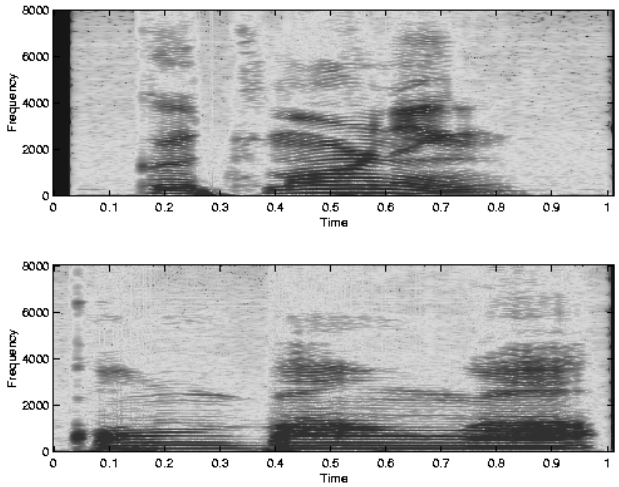
この後で，$\hat{\boldsymbol{y}}(\omega, t_s)$ の各成分を $\hat{W}(\omega)^{-1}$ によって “戻す” ことを考える．仮に $\hat{\boldsymbol{y}}(\omega, t_s)$ が 2 次元のベクトルであったとすれば，

$$\hat{\boldsymbol{u}}_1(\omega, t_s) = \hat{W}(\omega)^{-1} \begin{pmatrix} \hat{y}_1(\omega, t_s) \\ 0 \end{pmatrix}$$
$$\hat{\boldsymbol{u}}_2(\omega, t_s) = \hat{W}(\omega)^{-1} \begin{pmatrix} 0 \\ \hat{y}_2(\omega, t_s) \end{pmatrix}$$

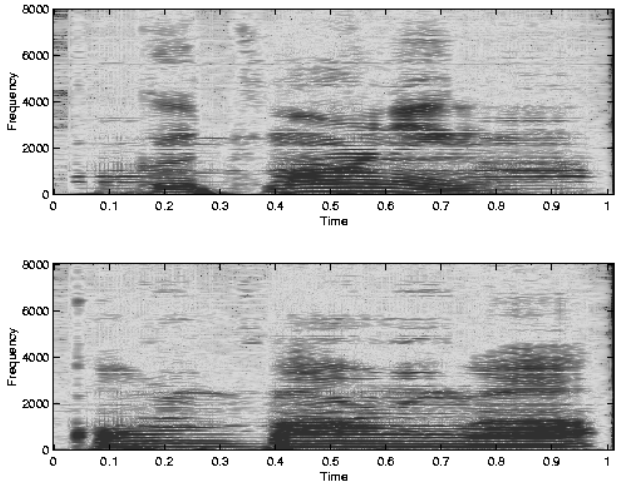
とするわけである．明らかに $\hat{\boldsymbol{x}}(\omega, t_s) = \hat{\boldsymbol{u}}_1(\omega, t_s) + \hat{\boldsymbol{u}}_2(\omega, t_s)$ である．したがって，観測における各独立成分の大きさを推定したという点で，$\hat{\boldsymbol{u}}_i(\omega, t_s)$ には大きさの任意性は無いと言える．

Permutation の任意性

Amplitude の任意性を取り除いただけでは未だ不十分である．例を示そう． 図–1 と図–2 とを比

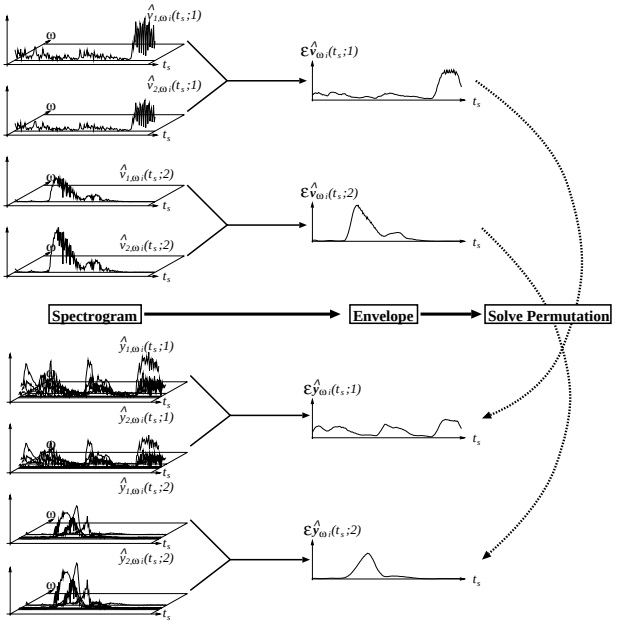


図–1 2 つの音源のスペクトログラム



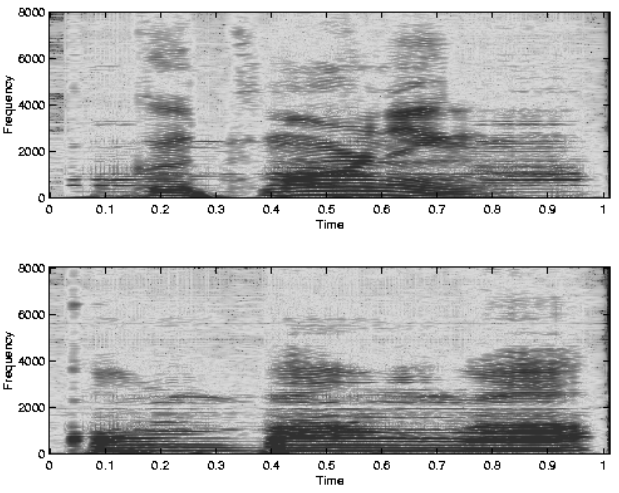
図–2 時間周波数における ICA の基づき信号を分離し，amplitude の任意性を取り除いた 2 つの信号

べると，いくつかの周波数において，分離がうまくいっていないようにみえる．この問題を解くために，角周波数における信号の強度を求め，そのエンベロープをグルーピングすることで，permutation の任意性を取り除く手法を提案した．図–3 に図を示す．この図のように，各周波数での時間的な変動を，他の周波数における時間変動と比べ，ベクトルの相関を取ることで近さを定義し，グルーピングを行なった．この手法によって求まった信号



図–3 Permutation の任意性を解く

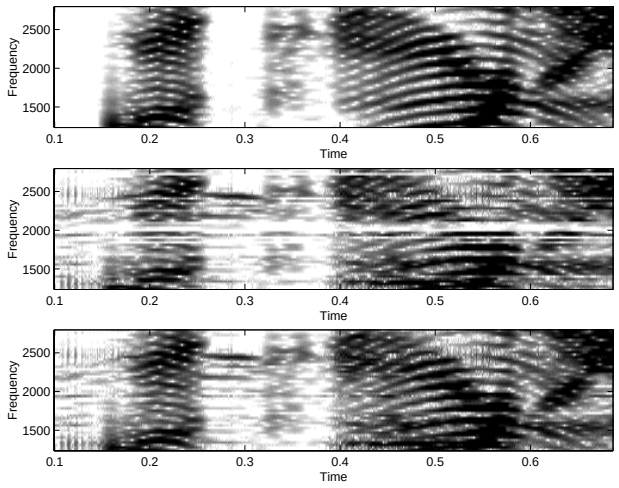
を図–4 に示す．



図–4 Amplitude と permutation の任意性を取り除き，復元された 2 つの音源のスペクトル

この手法はかなら簡単な方法に基づいているが，音の分離では，十分うまくいっている．各信号の対応する部分を拡大したものを図 5 に示す．

我々の行なった方法は，信号の情報をできるだけ用いず，permutation の任意性を取り除こうというものである．もちろん，他にも信号の 3 次元的な位置を用い，分離行列 $\hat{W}(\omega)$ の類似性から permutation の任意性を除くなどのアプローチは



図–5 図–1 (上) と図–2(中), 図– 4(下) の対応する部分を拡大したもの

考えられる．いずれにせよ，時間周波数における ICA では，amplitude と permutation の任意性を正しく扱わなければいけない，という点が重要である．

5. 両耳処理との相違

両耳での処理において，実際に音を上の結果のように分離しているかは意見が別れるだろうが，両耳で行なっている処理との相異点を考えるのは興味深い．両耳処理では，複数の信号源からの信号がそれぞれの耳に届く時間差，強度差を用いてなんらかの処理をしているはずである．

厳密な議論ではないが，蝸牛における処理は信号をバンドパスフィルタを通すことに対応している．したがって，両耳では時間信号を周波数軸上に展開し，脳においては時間周波数領域での処理を行なっていると考えられる．時間差は周波数軸では位相の差として表われる．したがって，この位相差と強度差とを用いることで，両耳ではカクテルパーティー効果と呼ばれる処理が可能になっているはずである．

反射などが無ければ，位相差と強度差とは実は 3 次元的な音源の位置の関数として書くことができるはずである．したがって，3 次元的な位置と両耳での処理とを直接結びつけるのも 1 つの手法である．これは伝統的なマイクロフォンアレイ処理の考え方とも一致する．

一方，ICA による音源分離の手法は 3 次元的な位置の情報は用いていない．単に独立な成分への分離を行なっているに過ぎない．理想的な場合であれば，求まった分離フィルタはビームフォーミングによって求まるフィルタと一致するだろうが，出発点が異なっている．

例えば，位相差のみ，あるいは強度差のみを作り込んで混合した信号を観測したとする．このとき，真に 3 次元的な位置情報と逆フィルタとを結びつけているとすると，信号の分離は行なえない．しかし，ICA であれば，どちらの場合でも問題なく分離することができる．これはある意味では両耳で行なっている信号処理に近い．

以上のことは，工学的には ICA とマイクロフォンアレイ的な処理を組み合わせることによって，音声分離の手法が向上できる可能性があることを示唆している．この方向については，幾つかのグループで積極的に研究が行なわれている [2, 16]．

一方，信号分離の 1 つの考え方に，Audio Scene Analysis のように，時間周波数軸上で，信号になんらかの意味を見出そうというアプローチもある．ICA は周波数間の関係の利用の仕方が Audio Scene Analysis とは異なるし，複数の観測を用いる点で，このような手法とも多少異なる．しかし，Audio Scene Analysis のような周波数間の関係に基づく知識を ICA の処理と組み合わせれば，より両耳の処理 “らしい” 処理が可能となるだろう．

また，人間は 2 つ以上の音を聞き分けられるが，ICA ではそれが可能かという疑問を持つ人もいるだろう．確かに，ICA は観測の数以上の信号源に分離することは基本的に不可能である．しかし，時間周波数領域で考えると，各周波数においてはそんなに多くの信号が重なっていないと考えるほうが妥当だろう．このスパース性を用いれば，各周波数では 2 つ程度の信号を分離することで，全体としてはそれ以上の信号を分離することも可能となる．

6. 今後の展望

最後に，ICA のような計算が脳において実現可能かという疑問が残る．基本 ICA についてはそれほど難しい計算を行なってはいないので，神経回路網による分離も可能だと考えられる．しかし，各周波数で基本 ICA を解き，さらに amplitude, permutation の任意性を，我々が提案したような手法で神経回路網が解いていると考えるのは無理

がある．他の予備的な知識を総合的に用いて，高次の処理をしていると考える方が妥当だろう．

各周波数での ICA 的な処理はボトムアップ的な処理ではないかと考える．一方，amplitude, permutation の任意性を取り除くための処理はトップダウン的な処理によって可能となるのではないかと考える．この 2 つをどのように結びつけるか，それがマイクロフォンアレイ的な処理なのか，Audio Scene Analysis 的な処理なのかは分らないが，

各方面の手法と ICA の考え方を組み合わせることが，これからの研究の方向として重要となると感じる．そのとき，重要な方向性を与えてくれるのは，両耳における処理ではないだろう．

文　献

1) S. Amari and J.-F. Cardoso. “Blind source separation – semiparametric statistical approach”, *IEEE Trans. Signal Processing*, 45(11):2692(1997).
2) F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki. “Blind source separation in reflective sound fields”, In *Proceedings of ICASSP2001*, MULT-P2.1(2001).
3) A. J. Bell and T. J. Sejnowski. “An information maximization approach to blind separation and blind deconvolution”, *Neural Computation*, 7(6):1129(1995).
4) J.-F. Cardoso. “Higher-order contrasts for independent component analysis”, *Neural Computation*, 11(1):157(1999).
5) P. Comon. “Independent component analysis, a new concept?”, *Signal Processing*, 36(3):287(1994).
6) S. C. Douglas and A. Cichocki. “Neural networks for blind decorrelation of signals”, *IEEE Trans. on Signal Processing*, 45(11):2829(1997).
7) A. Hyvärinen and P. Hoyer. “Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces”, *Neural Computation*, 12(7):1705(2000).
8) A. Hyvärinen and E. Oja. “A fast fixed-point algorithm for independent component analysis”, *Neural Computation*, 9(7):1483(1997).
9) S. Ikeda and N. Murata. “A method of blind separation based on temporal structure of signals”, In *Proceedings of ICONIP’98*, 737(1998).
10) S. Ikeda and K. Toyama. “Independent component analysis for noisy data –MEG data analysis”, *Neural Networks*, 13(10):1063(2000).
11) C. Jutten and J. Herault. “Separation of sources, part i”, *Signal Processing*, 24(1):1(1991).
12) S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. “Blind separation of auditory event-related brain responses into independent components”, *Proc. Natl. Acad. Sci. USA*, (94):10979(1997).
13) L. Molgedey and H. G. Schuster. “Separation of a mixture of independent signals using time delayed correlations”, *Phys. Rev. Lett.*, 72(23):3634(1994).
14) N. Murata, S. Ikeda, and A. Ziehe. “An approach to blind source separation based on temporal structure of speech signals”, *Neurocomputing*, 41(1-4):1(2001).
15) B. A. Olshausen and D. J. Field. “Emergence of simple-cell receptive field properties by learning a sparce code for natural images”, *Nature*, 381:607(1996).
16) 猿渡, 栗田, 武田, 板倉, 鹿野. 帯域分割型 ICA とアルゴリズムダイバーシチに基づくブラインドビームフォーマ. 日本音響学会 2000 年秋季研究発表会 講演論文集, 443(2000).
17) P. Smaragdis. “Blind separation of convolved mixtures in the frequency domain”, *Neurocomputing*, 22(1-3):21(1998).
18) A. Ziehe and K.-R. Müller. “TDSEP – an efficient algorithm for blind separation using time structure”, In *Proceedings of ICANN’98*, 675(1998).