

音源方位ヒストグラム法を用いた遅れ時間推定による音源分離

Estimation of propagation delays by orientation histogram for blind source separation

龍田 成揮† 山下 純司‡ 平井 有三十
Shigeki Tatsuta Junji Yamashita Yuzo Hirai

1. はじめに

ブラインド信号源分離問題は、信号処理やニューラルネットワークの分野で盛んに研究が行われている[1][2][3]。ICAに基づいているアルゴリズムが多く、その他時間領域で2次統計量を扱うもの[4]や、周波数領域で扱うもの[5]などがあり、畳み込み混合のようなより現実的な混合にも適用されている。我々は、実環境においてリアルタイムな音源分離を行うためのアプローチとして、音源方位ヒストグラム法[6][7]を提案している。

本手法は、ICAで生じる出力位置や強度の任意性の問題がなく、学習を含まないため繰り返し演算が不要である。本論文では、この音源方位ヒストグラム法を発展させ、バンドパスフィルタと組み合わせることで、音の伝播遅れ時間を推定し、分離を行う手法を提案する。また、仮想空間における音声と音楽の混合音や、実環境で録音された混合音[8]の分離を効果的に実行できることを示す。

2. 音源方位ヒストグラム法

2.1 問題の定式化

$x(t)$ を N 個の観測信号による観測信号ベクトル、 $A(t)$ を混合行列、 $s(t)$ を M 個の独立した信号源から発生する源信号ベクトルとすると、源信号の混合過程は

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

となる。以下、本論文では $M = N = 2$ を取り扱う。

2.2 音源方位の検出と分離

2つの観測信号の同時分布は図1のようになる。また、各点と原点を通る直線が x 軸(水平軸)となす角の角度ヒストグラムは図2のようになる。ヒストグラムの2つのピークが2つの信号源方位として推定される。

検出された信号源方位の単位ベクトルを u_1, u_2 とすると、推定される混合行列は $U = (u_1, u_2)$ となり、分離信号 $y(t)$ を以下のように求めることができる。

$$y(t) = U^{-1}x(t) \quad (2)$$

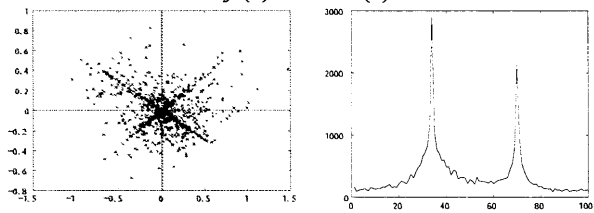


図1. 観測信号の同時分布

図2. 角度のヒストグラム

3. バンドパスフィルタを用いた分離

音源方位ヒストグラムは、音声 - 音楽などスーパーガウシアン分布とサブガウシアン分布の信号の混合などに対し

† 筑波大学大学院 システム情報工学研究科
〒305-8573 茨城県つくば市天王台 1-1-1
‡ Hewlett Packard

ては、図3のように2つのピークが現れず、うまく分離を行うことができない。そこで、観測信号に対して狭帯域のバンドパスフィルタをずらしながらかけ、通過帯域ごとに音源方位ヒストグラムを作成し、平均することで、図4のように2つのピークが出現するようになり、この方位をデータに射影することで分離信号を得ることができる[6][7]。

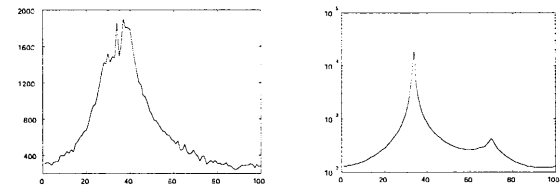


図3. フィルタなし 図4. フィルタあり (Y軸 log-scale)

4. 反射のない畳み込み混合の分離

4.1 仮想空間による混合

時間遅れのある畳み込み混合をシミュレートするために、仮想空間混合モデルを考える。マイク、信号源はすべて同一平面上にあるとする。

S_j, M_i をそれぞれ j 番目の信号源、 i 番目のマイクとし、信号源とマイク座標をそれぞれ $(X_{S_j}, Y_{S_j}), (X_{M_i}, Y_{M_i})$ とすると、距離 δ_{ji} 、遅延サンプル数 Δ_{ji} は以下の式で表せる。

$$\delta_{ji} = \sqrt{(X_{S_j} - X_{M_i})^2 + (Y_{S_j} - Y_{M_i})^2} \quad (3)$$

$$\Delta_{ji} = \left\lfloor \frac{\delta_{ji}}{v t_s} + 0.5 \right\rfloor \quad (4)$$

ここで、 v は音速、 t_s は標本化間隔である。よって、 i 番目のマイクでの観測信号 x_i は、以下ようになる。

$$x_i(t) = \sum_j \frac{1}{\delta_{ji}^2} s_j(t - \Delta_{ji}) \quad (5)$$

4.2 分離アルゴリズム

このアルゴリズムは、時間遅れによって対応関係が失われていた観測信号から、観測信号間をずらしながら同時分布を求め、音源方位を求める。このとき、最も高い高密度方位が現れるずらしサンプル数 τ_j が、実際の遅延時間に対応するものと考えられる。そこで、その高密度方位に射影を行い、分離を行う。

実際に前節で示した仮想空間における観測信号を1サンプルずつずらして同時分布を取った結果を図5に示す。±12 サンプルのところ強いピークが現れていることわかる。このようにして推定された遅延サンプル数 τ_j をずらして、同時分布を取ったときの音源方位 θ_1, θ_2 とし、その単位ベクトルを u_1, u_2 、分離行列を

$$U = (u_1, u_2) \quad (6)$$

とすると、以下のように射影信号を得ることができる。

$$w_i(t) = \begin{pmatrix} w_{i1}(t) \\ w_{i2}(t) \end{pmatrix} = U^{-1} \begin{pmatrix} x_1(t) \\ x_2(t - \tau_i) \end{pmatrix} \quad (7)$$

ここで、一つの遅延サンプル数に対応する源信号は一つであるから、式(7)より以下の分離信号を得ることができる。

$$y(t) = \begin{pmatrix} w_{11}(t) \\ w_{22}(t) \end{pmatrix} \text{ or } \begin{pmatrix} w_{21}(t) \\ w_{12}(t) \end{pmatrix} \quad (8)$$

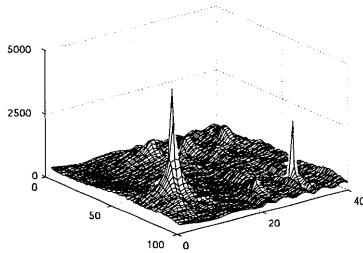


図5. ずらしサンプル数と方位ヒストグラム
X軸: ずらしサンプル数、Y軸: 方位、Z軸: 投票数

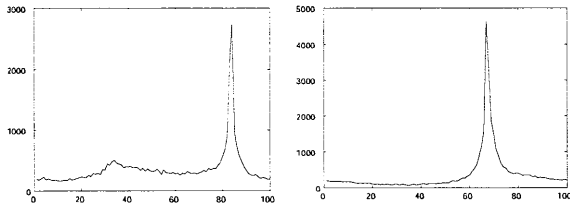


図6. τ_1 でのヒストグラム 図7. τ_2 でのヒストグラム

4.3 実験結果

仮想空間によりシミュレートされた信号(約3.5秒)を観測信号とした。サンプリングレートは8kHz、サンプル数は37170である。実験結果を図8に示す。実験の結果、式(9)で表される源信号対分離信号比(Source-to-Separated signal Ratio: 以下SSR、単位[dB])は、それぞれ11.0dB, 9.7dB、計算時間は約22秒であった。

$$SSR = 10 \log_{10} \frac{\sum_i s(t)^2}{\sum_i (y(t) - s(t))^2} \quad (9)$$

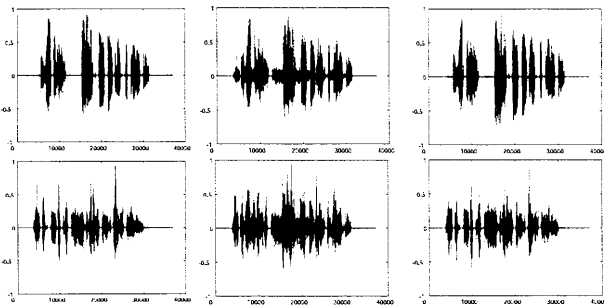


図8 源信号(左) 観測信号(中) 分離信号(右)

5. 実環境音での分離

提案したアルゴリズムを用いて、実環境音[8]の分離を行った。結果を図9に示す。ある程度二つの音声分離できている。以下のURLで実際の音声を聞くことができる。
<http://www.viplab.is.tsukuba.ac.jp/>

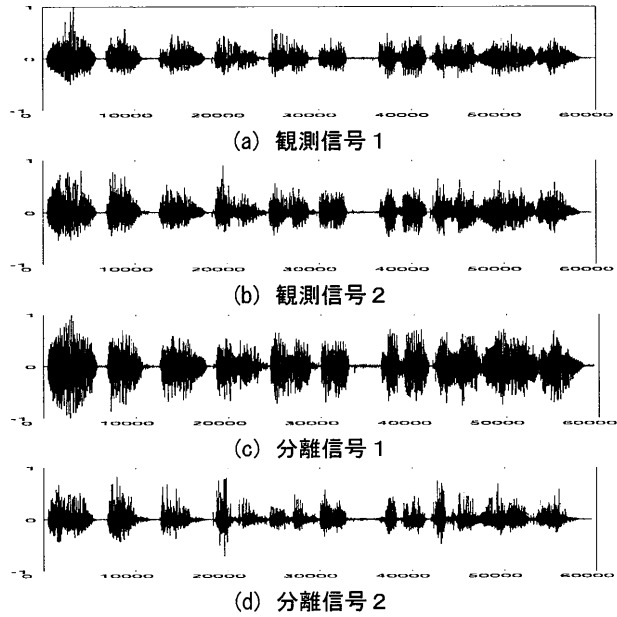


図9 実環境音[8]の観測信号と分離信号

6. 結論

本研究では、音源方位ヒストグラム法をもとにして、遅延の差を推定し、補正することで分離を行うアルゴリズムを提案した。この手法では遅延のみを考慮した混合信号に対しては良い分離性能を示した。また、実環境音の分離を行った結果、完全とは言えないが分離を行うことができた。今後は反響などを考慮したアルゴリズムの開発を進めていく予定である。

7. 参考文献

- [1] Anthony J. Bell and Terrence J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129-1159, 1995.
- [2] Shun-ichi Amari, Andrzej Cichocki, and Howard H. Yang. *A new learning algorithm for blind signal separation*, volume 8, pages 757-763. San Mateo. Morgan Kaufmann, 1996.
- [3] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.
- [4] L.Parra and C.Spence. Convolutional blind separation of non-stationary sources. In *IEEE Transactions on Speech and Audio Processing*. vol.8 pp. 320-327, 2000.
- [5] N.Mutara, S.Ikeda, and Z.Andreas. An Approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, pp. 1-24, 2001.
- [6] H.Shindo and Y.Hirai, Blind source separation by a geometric method. In *IEEE World Congress on Computational intelligence Workshop*, pp1109-1114 2002.
- [7] J.Yamashita, S.Tatsuta and Y.Hirai, Estimation of Propagation Delays using Orientation Histograms for Anechoic Blind Source Separation. *International Joint Conference on Neural Networks*, in Budapest 2004.
- [8] T.-W. Lee, A.J.Bell and R.Orglmeister, Blind source separation of real world signals. In *Proceedings of IEEE International Conference on Neural Network*, pp.2129-2135 1997