

観測信号ベクトルのクラスタリングに基づく スパース信号の到来方向推定*

○荒木 章子, 澤田 宏, 向井 良, 牧野 昭二 (NTT 研究所)

1 はじめに

本稿では、信号のスパース性の利用により、センサ数より信号数が多い場合にも対応できる信号の到来方向推定方法を提案する。信号のスパース性を用いた手法はこれまでいくつか提案されているが、従来法は2センサ [1] あるいは3センサ [2] の場合に特化された手法であったため、3次元的な方向推定ができなかった。これに対し本手法は、4センサ以上への適用も容易であり、到来方向を3次元ベクトルとして推定できる。提案法により、残響のある環境においても正確な到来方向推定ができたので報告する。

2 問題設定

2.1 観測信号のモデル

N 個の信号が M 個のセンサにて観測されるとする。センサ j による観測信号 x_j は、 $x_j(n) = \sum_{i=1}^N \sum_{t=1}^{\infty} h_{ji}(t)s_i(n-t+1)$ ($j = 1, \dots, M$) とモデル化される。ここで s_i は信号 i 、 h_{ji} は信号源 i とセンサ j 間のインパルス応答である。本稿の目的は、原信号 s_i を用いることなく、観測信号 x_j およびセンサ位置情報を用いて、全ての信号の到来方向推定をすることである。

本稿では、時間領域での観測信号 $x_j(n)$ ($j = 1, \dots, M$) に短時間フーリエ変換 (STFT) を適用し、時間周波数領域にて信号を扱う。時間周波数領域における観測信号 $x_j(f, \tau)$ は近似的に $x_j(f, \tau) \approx \sum_{i=1}^N h_{ji}(f)s_i(f, \tau)$ と書け、ベクトル表記では

$$\mathbf{x}(f, \tau) \approx \sum_{i=1}^N \mathbf{h}_i(f)s_i(f, \tau) \quad (1)$$

と書ける。ここで $s_i(f, \tau)$ ($i = 1, \dots, N$) は原信号の STFT 結果、 $h_{ji}(f)$ は信号源 i からセンサ j への周波数応答、 $\mathbf{x} = [x_1, \dots, x_M]^T$ は観測信号ベクトル、 $\mathbf{h}_i = [h_{1i}, \dots, h_{Mi}]^T$ である。

また本稿では、スパース信号を仮定する。これより、各時間周波数 (f, τ) において原信号のうちの1つ s_k のみが支配的であり、式 (1) が

$$\mathbf{x}(f, \tau) \approx \mathbf{h}_k(f)s_k(f, \tau), \quad k \in \{1, \dots, N\} \quad (2)$$

と書けることを仮定できる。これは音声信号などで確認される。

2.2 到来方向の定義

ここでは到来方向の定義を行う。 \mathbf{q}_i を信号源 s_i の方向を示す大きさ1の3次元ベクトルとする [Fig. 1(a)]。また、各センサ j の3次元位置ベクトル \mathbf{d}_j が与えられるとする [Fig. 1(a)]。Fig. 1(b) に示すように、方位角 θ_i と仰角 ϕ_i を用いれば到来方向ベクトル \mathbf{q}_i は

$$\mathbf{q}_i = [\cos \theta_i \cos \phi_i, \sin \theta_i \cos \phi_i, \sin \phi_i]^T \quad (3)$$

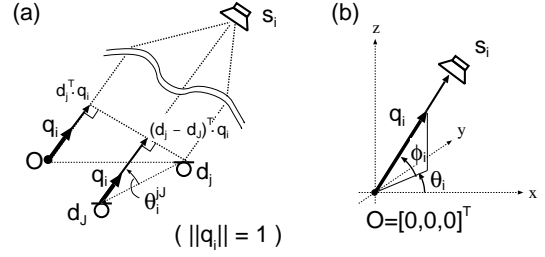


Fig. 1 (a) 遠距離場モデル (b) 到来方向ベクトル

と書ける。

ここでは信号の到来方向推定のため、周波数応答 $h_{ji}(f)$ を Fig. 1(a) に示す遠距離場モデルで近似する:

$$h_{ji}(f) \approx \exp[-j2\pi f c^{-1} \mathbf{d}_j^T \mathbf{q}_i] \quad (4)$$

(c は伝播速度)。すなわち周波数応答 $h_{ji}(f)$ は信号 i から原点 O までと、信号 i からセンサ j までの行路差 $\mathbf{d}_j^T \mathbf{q}_i$ による遅延により決まると仮定する [Fig. 1(a)]。

本稿では、観測信号 $\mathbf{x}(f, \tau)$ およびセンサ位置情報 \mathbf{d}_j ($j = 1, \dots, M$) を用いて、到来方向ベクトル \mathbf{q}_i ($i = 1, \dots, N$) を求めることが目的である。

3 提案手法

1. 観測信号ベクトルの正規化: まず観測信号ベクトル $\mathbf{x} = [x_1, \dots, x_M]^T$ の各要素について、下記の正規化を行う (詳細は [3, 4] 参照)。

$$\bar{x}_j(f, \tau) \leftarrow |x_j(f, \tau)| \exp \left[j \frac{\arg[x_j(f, \tau)/x_J(f, \tau)]}{4\pi f c^{-1} d_{\max}} \right] \quad (5)$$

$$\bar{\mathbf{x}}(f, \tau) \leftarrow \bar{\mathbf{x}}(f, \tau) / \|\bar{\mathbf{x}}(f, \tau)\| \quad (6)$$

ここで d_{\max} はあるセンサ J と $\forall j \in \{1, \dots, M\}$ の距離の最大値であることが望ましい。これにより、正規化観測信号ベクトル $\bar{\mathbf{x}} = [\bar{x}_1, \dots, \bar{x}_M]^T$ を得る。

2. クラスタリング: 次に、正規化観測信号ベクトル $\bar{\mathbf{x}}(f, \tau)$ を、全ての (f, τ) をまとめて、観測信号ベクトルと同じ次元 (M 次元複素空間) にクラスタリングし、 N 個のクラスタ C_1, \dots, C_N を求める。このクラスタリングは、適切なセントロイドの初期値 \mathbf{c}_k ($k = 1, \dots, N$) を設定した後、セントロイドと各正規化観測信号ベクトルとの距離の最小化によるクラスタメンバの決定 (7) と、セントロイドの再設定 (8) との反復学習により実現される。

$$C_k = \{\bar{\mathbf{x}}(f, \tau) \mid k = \operatorname{argmin}_i \|\bar{\mathbf{x}}(f, \tau) - \mathbf{c}_i\|^2\} \quad (7)$$

$$\mathbf{c}_k \leftarrow E[\bar{\mathbf{x}}(f, \tau)]_{\bar{\mathbf{x}} \in C_k}, \quad \mathbf{c}_k \leftarrow \mathbf{c}_k / \|\mathbf{c}_k\|, \quad (8)$$

ここで $E[\cdot]_{\bar{\mathbf{x}} \in C_k}$ はクラスタ C_k のメンバに対する平均操作を表す。この学習は k-means 法 [5] などにより効果的に実現できる。

*Observation vector clustering based DOA estimation for multiple sparse sources. by ARAKI, Shoko, SAWADA, Hiroshi, MUKAI, Ryo, MAKINO, Shoji (NTT Communication Science Laboratories, NTT Corporation.)

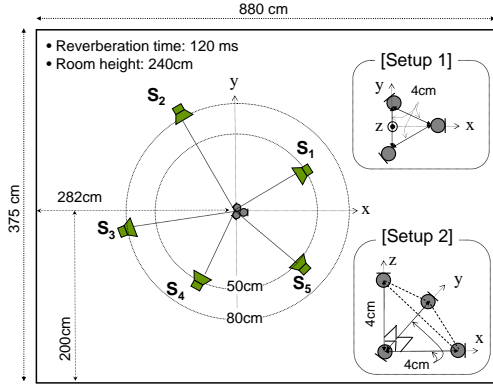


Fig. 2 実験条件。Setup 1 の時: 音源は全てマイクホンと同じ高さとした。Setup 2 の時: 音源の高さは様々とした。

Table 1 $N = 4, M = 3$ (Setup 1) の場合の結果

音源	s_1	s_2	s_3	s_4
正解	24°	117°	217°	311°
提案法 (5 秒データ)	25°	114°	214°	313°
提案法 (0.5 秒データ)	23°	109°	213°	328°

このように求めたセントロイド \mathbf{c}_k (8) は、(1) に見られる周波数応答ベクトル \mathbf{h}_k の推定値となる。実際、(2), (5), (6), (8) より

$$\mathbf{c}_k = E[\bar{\mathbf{x}}(f, \tau)]_{\bar{\mathbf{x}} \in C_k} = E[\bar{\mathbf{h}}_k(f)]_f \quad (9)$$

を示すことができる。ここで $\bar{\mathbf{h}}_k(f)$ は (5)(6) と同様の操作で正規化された周波数応答ベクトル、 $E[\cdot]_f$ は周波数に関する平均操作である。求めたい到来方向の情報は、(4) のモデル化により周波数応答ベクトル $\mathbf{h}_k(f)$ に含まれる。よってセントロイドを用いることで到来方向推定が可能となる。

3. 到来方向推定: 次に遠距離場モデル (4) と (2), (5), (6), (9) より、 k 番目のセントロイド \mathbf{c}_k の j 番目の要素 $\{c_k\}_j$ は

$$\begin{aligned} \{c_k\}_j &= E[\bar{x}_j(f, \tau)]_{\bar{\mathbf{x}} \in C_k} = E[\bar{h}_{jk}(f)]_f \\ &= \frac{1}{\sqrt{M}} \exp \left[j \frac{\pi}{2d_{\max}} (\mathbf{d}_j - \mathbf{d}_J)^T \mathbf{q}_k \right] \end{aligned} \quad (10)$$

と表すことができる。これはセンサペア jJ に関する情報のみ含んでいる。センサ J とその他のセンサによる全てのセンサペアの情報を用いると (10) は

$$\mathbf{r}_k = \frac{\pi}{2d_{\max}} \mathbf{D} \mathbf{q}_k \quad (11)$$

と書ける。ここで

$$\begin{aligned} \mathbf{r}_k &= [\arg\{\{c_k\}_1\}, \dots, \arg\{\{c_k\}_M\}]^T \\ \mathbf{D} &= [\mathbf{d}_1 - \mathbf{d}_J, \dots, \mathbf{d}_M - \mathbf{d}_J]^T \end{aligned}$$

である。これを \mathbf{q}_k について解けば到来方向ベクトル \mathbf{q}_k が求まるが、普通 (11) は厳密解を持たないので、擬似逆行列 \mathbf{D}^+ を用いて近似的に解く。これより最終的な到来方向ベクトル \mathbf{q}_k は

$$\mathbf{q}_k = \frac{2d_{\max}}{\pi} \mathbf{D}^+ \mathbf{r}_k. \quad (12)$$

にて与えられる。 $\text{rank}(\mathbf{D}) \geq 3$ であれば、到来方向を 3 次元的に求めることが可能である。

Table 2 $N = 5, M = 3$ (Setup 1) の場合の結果

音源	s_1	s_2	s_3	s_4	s_5
正解	24°	117°	176°	217°	311°
提案法 (5 秒データ)	23°	112°	175°	218°	314°

Table 3 $N = 5, M = 4$ (Setup 2) の場合の結果

音源		s_1	s_2	s_3	s_4	s_5
正解	θ	31°	85°	133°	222°	302°
	ϕ	-26°	6°	30°	39°	-8°
提案法 (5 秒データ)	θ	30°	79°	132°	221°	298°
	ϕ	-22°	7°	28°	35°	-9°

4 実験と結果

Fig. 2 に示す環境で実験を行った。混合信号は、原則 5 秒間の英語音声に、Fig. 2 の環境で計測したインパルス応答を畳み込んで得た。サンプリング周波数は 8kHz、STFT フレーム長 L は 512、フレームシフトは $L/2$ とした。到来方向の正解は、各信号を 1 つずつ鳴らした時に MUSIC 法にて推定した到来方向とした。推定結果は 8 組の音声組合せの平均値である。

結果を Tables 1~3 に示す。Tables 1,2 では 2 次元の方向推定である (仰角 $\phi = 0$ である) ため方位角 θ のみを、Table 3 では 3 次元の方向推定なので方位角 θ と仰角 ϕ の双方を推定した [(3) 参照]。音源数 > センサ数であるにもかかわらず、高い精度で到来方向推定ができています。また Table 1 では、データ長が 0.5 秒の場合の推定結果も示している。短いデータでも良い推定ができていたことから、本手法は移動音源の追跡にも応用可能である。

また実験で用いた音声および残響時間というスパース性の程度の場合、角度差が 20 度の 2 つの到来方向を推定できた [6]。一方 MUSIC 法では 2 つの到来方向を分離して求めることはできなかった [6]。これはスパース性を用いる提案法が高い角度分解能を示すことを示唆すると考えている。

参考文献

- [1] S. Rickard and F. Dietrich, "DOA estimation of many W-disjoint orthogonal sources from two mixtures using DUET," in *Proc. SSAP2000*, Aug. 2000, pp. 311-314.
- [2] M. Matsuo, Y. Hioka, and N. Hamada, "Estimating DOA of multiple speech signals by improved histogram mapping method," in *Proc. IWAENC2005*, Sept. 2005, pp. 129-132.
- [3] 荒木章子, 澤田宏, 向井良, 牧野昭二, "観測信号ベクトル正規化とクラスタリングによる音源分離手法とその評価," 音講論 (秋), pp. 591-592, 2005.
- [4] S. Araki, H. Sawada, R. Mukai, and S. Makino, "A novel blind source separation method with observation vector clustering," in *Proc. IWAENC2005*, Sept. 2005, pp. 117-120.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley Interscience, 2nd ed., 2000.
- [6] S. Araki, H. Sawada, S. Makino, and R. Mukai, "DOA estimation for multiple sparse sources with normalized observation vector clustering," ICASSP2006 (accepted).