

音源位置とマイクロフォン位置の同時推定及び受信信号時間原点のブラインドアライメント*

河野仁 (東大工), 小野順貴, 伊藤信貴, 嵯峨山茂樹 (東大情理)

1 はじめに

通常のアレイ信号処理では, 位置が既知のマイクロフォンや, A/D 変換器を介して同期した観測信号が取得されることが前提となっている。これに対し我々は, 多数マイクロフォンの使用を容易にする枠組みとして, PC 内蔵マイクロフォンや IC レコーダなど, 個々に独立した録音機器を, 分散型マイクロフォンアレイとして構成するための研究を進めている。本報告ではこのための基礎技術として, 音源とマイクロフォンの空間位置, 各録音機器の録音時間原点が全て未知の状態から, 観測信号のみを用いてこれらを同時推定するための手法について論じ, 基礎的な実験結果について報告する。

2 定式化

2.1 未知数と観測量の関係

以下では, K 個の音源を L 個のマイクロフォンで観測するものし, 音源, マイクロフォンの位置をそれぞれ $\mathbf{s}_i = (x_i \ y_i \ z_i)^t$ ($1 \leq i \leq K$), $\mathbf{r}_j = (u_n \ v_n \ w_n)^t$ ($1 \leq n \leq L$) で表す。ただし t は転置を表す。また t_n は, ある規準時計で測った n 番目のマイクロフォンの録音開始時刻 (時間原点) とし, 各録音機器での時刻の進み方は等しいことを仮定する。

音源やマイクロフォンの位置推定において最も大きな手がかりの 1 つは観測信号間の時間差である。いま, 音源 i から到来する信号をマイクロフォン m, n で観測し, 信号間の時間差を検出したとすると, n 番目の信号を基準にしたときの m 番目の信号時間遅れは

$$\tau_{imn} = \frac{|\mathbf{s}_i - \mathbf{r}_m| - |\mathbf{s}_i - \mathbf{r}_n|}{c} - (t_m - t_n) \quad (1)$$

と表わされる。式 (1) の第 1 項が真の時間差, 第 2 項が録音開始時刻のずれを表す。

2.2 必要な観測数

各観測信号が同期しており, 式 (1) の第 2 項が 0 の場合には, この時間差の整合性がとれるようにすることで音源位置とマイクロフォン位置を同時推定する手法が提案されている [?]。しかしながら観測信号から求めた時間差に未知の録音開始時刻の差も含まれる場合には一見, 有効な情報は得られないようにも思われる。よってここではまず, 解が求まりうる条件として, 観測数の数と未知数の関係を調べる。

観測量は各観測信号間の時間差 τ_{imn} であり, 1 つの音源に対し $L - 1$ 個の時間差が独立な観測量となる。一方未知数は, 音源とマイクロフォンの 3 次元位置 (x_i, y_i, z_i) , (u_n, v_n, w_n) と録音開始時刻 t_n (ただし $1 \leq i \leq K, 1 \leq n \leq L$) であるが, 時間差に基づく推定の場合にはいずれも相対的にしか決まらず, 基準時計の選び方の 1 自由度と絶対座標系の選び方の並進・回転の 6 自由度は決まらないことに注意すると, 全てのパラメータが未知の場合, 観測量から未知数が決定されるためには少なくとも $K(L - 1) \geq (L - 1) + 3K + 3L - 6$, すなわち, 以下を満たす必要がある。

$$(K - 4)(L - 4) \geq 9 \quad (2)$$

2.3 観測量の取得方法

式 (1) に示したように, 検出される時間差には 1) 音源信号の到来時間差, 2) 録音開始時刻の差, の 2 つが含まれる。1) は音源毎に異なる値をとり, 音源, マイクロフォン間の最大距離を D とすると, 絶対値は高々 D/c を超えない。一方 2) はどんな値でもとり得るため, 各音源に対応する時間差を検出可能にするためにはまず, 2) を大まかに補償する必要がある。よってここでは, 以下のようなアルゴリズムを用いた。

1. ある 1 つの観測信号と, 他の全ての観測信号との相互相関関数を全時間区間を用いて計算し, そのピークから求まる平均的時間差に基づき, 観測信号間の大まかなアライメントをとる。
2. D/c に対して十分大きいフレーム長を選び, 観測信号をフレーム分割する。
3. フレーム毎に観測信号間で正規化した相互相関関数を計算し, ある閾値を超えているピークから, 有意な単一音源の観測信号間時間差を検出する。

3 反復解法の導出

3.1 評価関数の設定

以下では, 理論式である式 (1) の二乗誤差:

$$J(\Theta) = \frac{1}{KL^2} \sum_{i=1}^K \sum_{m=1}^L \sum_{n=1}^L \varepsilon_{imn}^2 \quad (3)$$

$$\varepsilon_{imn} = |\mathbf{s}_i - \mathbf{r}_m| - |\mathbf{s}_i - \mathbf{r}_n| - c(\tau_{imn} + t_m - t_n) \quad (4)$$

* Blind Alignment of Received Signals to Localize Sources and Microphones by Hitoshi KOHNO, Nobutaka ONO, Shigeki SAGAYAMA, Nobutaka ITO (The University of Tokyo)

を最小化することにより、未知パラメータ $\Theta = \{s_i, r_n, t_n \mid 1 \leq i \leq K, 1 \leq n \leq L\}$ を決定することを試みる。

3.2 反復解法式

近年信号処理の分野では、補助関数法と呼ばれる最適化手法が様々な応用されつつある [2, 3, 4]。ここでは紙面の都合上、詳細を述べることができないが、我々は、任意のベクトル x , x_0 , 正の実数 a に対して

$$(|x| - a)^2 \leq \left| x - \frac{x_0}{|x_0|} a \right|^2 \quad (5)$$

が成り立つ (等号条件は $x \parallel x_0$) ことに着目し、式 (3) の適切な補助関数を設計し、以下の更新式を導出した。

$$\mu_{imn}^m \leftarrow |s_i - r_m| - \frac{1}{2} \varepsilon_{imn} \quad (6)$$

$$\mu_{imn}^n \leftarrow |s_i - r_n| + \frac{1}{2} \varepsilon_{imn} \quad (7)$$

$$t_n \leftarrow t_n + \frac{1}{cKL} \sum_{i=1}^K \left(L |s_i - r_n| - \sum_{m=1}^L \mu_{imn}^n \right) \quad (8)$$

$$e_{in} \leftarrow (s_i - r_n) / |s_i - r_n| \quad (9)$$

$$s_i \leftarrow \frac{1}{L^2} \sum_{m=1}^L \left(L r_m + e_{im} \sum_{n=1}^L \mu_{imn}^m \right) \quad (10)$$

$$r_n \leftarrow \frac{1}{KL} \sum_{i=1}^K \left(L s_i - e_{in} \sum_{m=1}^L \mu_{imn}^n \right) \quad (11)$$

パラメータは、1) ε (式 (4) を用いる)、2) 補助変数 μ 、3) t 、4) 補助変数 e 、5) s, r 、の順序で更新を繰り返し、反復推定を行う。

4 評価実験と結果

式 (3) を最小化することで、音源位置、マイクロフォン位置、時間原点の同時推定が可能であるかどうか検証するために行った基礎実験の結果を示す。10 × 10 × 10[m³] の室内を想定し、ほぼ無響環境の球面波伝播を計算機上でシミュレーションした。音源数は 8、マイクロフォン数は 9 とし、位置は乱数で決定した。各音源信号は 1 回の拍手の音を録音したものをを用い、互いに重ならず、単一音源が観測できる条件を仮定した。サンプリング周波数は 44100[Hz]、信号長は 5.0[s] とし、時間原点のずれとして、1.0[s] 以内のランダムな時間差を各観測信号に与えた。シミュレーションで得られた観測信号に対し、2.3 節で論じたように、大まかなアラインメントを求めた後、フレーム長 100[ms] ($> D/c \simeq 50$ [ms]) のフレームに分割し、有意な音響信号が含まれているフレームから時間差を検出し、各パラメータの初期値を乱数で与え

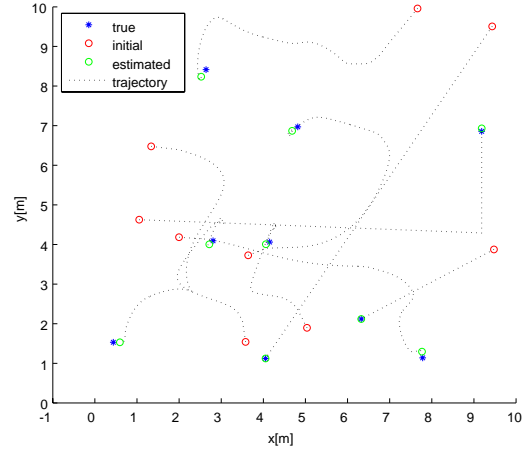


Fig. 1 マイクロフォン位置の推定結果

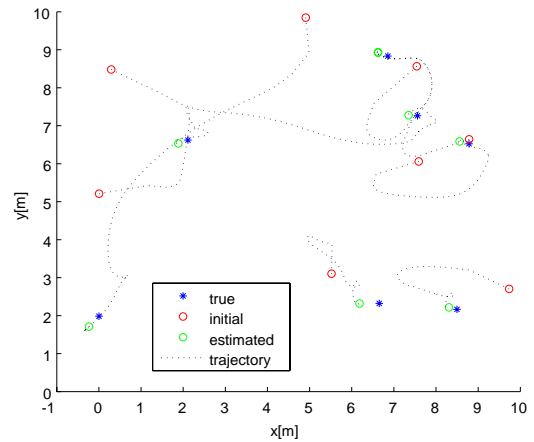


Fig. 2 音源位置の推定結果

て、反復解法による推定を行った。反復回数は 60000 回とした。

マイクロフォン、音源の位置推定に関して、 xy 座標をプロットしたものをそれぞれ Fig. 1, Fig. 2 に示す。時間原点が未知の観測信号から、ほぼ正しくマイクロフォン位置、音源位置が推定されていることがわかる。また、時間原点の推定誤差の標準偏差は 1.0[ms] であり、こちらもほぼ正しく推定されていることを確認した。

今後は、実環境での音響信号を用いた実験、様々なマイクロフォン数と音源数の条件下での検証などを行っていく予定である。

参考文献

- [1] 小林 和則, 古家 賢一, 片山 章俊, 電子情報通信学会論文誌 A, vol. J86-A, No.6, pp. 619 – 627, 2003.
- [2] D. D. Lee and H. S. Seung, Proc. NIPS, pp. 556–562, 2000.
- [3] H. Kameoka, N. Ono, and S. Sagayama, Proc. ICASSP, pp. 29–32, 2008.
- [4] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, Proc. EUSIPCO, 2008.