

空間勾配スペクトルのベイズモデリングによる複数音源の瞬時定位*

©植野夏樹（東大院情報理工，NTT CS研），亀岡弘和（NTT CS研）

1 はじめに

波源定位は，レーダやソナーといった幅広い応用を有している。特に，小さいアレイで，移動する波源を瞬時に定位し追跡できるようにすることは重要課題である。波源定位問題に対する従来法としては，Multiple Signal Classification (MUSIC) 法 [1]，Generalized Cross-Correlation methods with Phase Transform (GCC-PHAT) 法 [2]，波源拘束偏微分方程式に基づく手法 [3, 4, 5] などがある。

MUSIC 法や GCC-PHAT 法は，音源に対し平面波を仮定し各音源のセンサ間での到来時間差を定位の手がかりとするため，一般にアレイサイズは大きい方が有利となる。また，いずれもセンサアレイの受信信号間の自己相関関数や相互相関関数といった，統計量に基づく手法であるため，音源を高い精度で定位するためには観測時間幅を十分長く取る必要がある。このため，これらの手法は小さいアレイサイズと瞬時的な観測のみによる波源定位には必ずしも向いていない。一方，波源拘束偏微分方程式に基づく手法は，各時刻ごとに成立する音響信号の時空間偏微分方程式を元に音源定位を行うもので，理論的には瞬時の小領域観測のみで波源定位を行うことが可能である。ただし，この手法は単一波源に対して成立する方程式をベースとしているため，雑音や複数の点音源が存在する場合などのように観測音響信号が偏微分方程式から逸脱する場合に脆弱であるという欠点を有している。

そこで本研究では，本枠組においてランダム雑音や複数の点音源が存在する場合への拡張を可能にするため，波源拘束偏微分方程式を基にしたアレイ観測信号の確率モデルを導き，ベイズ推論に基づく波源定位アルゴリズムを導出する。また，数値シミュレーションによって提案手法の有効性を検証する。

2 音源定位問題の定式化

2.1 波源拘束偏微分方程式

観測点の基準となる位置ベクトルを $\mathbf{r} = (x, y, z)^T \in \mathbb{R}^3$ とし，単一波源の位置ベクトルを $\mathbf{r}_0 = (x_0, y_0, z_0)^T \in \mathbb{R}^3$ とする。波源の信号を $g(t)$ ，音速を c とし，単一点波源からの球面波伝播を仮定すると観測点における観測値は

$$f(\mathbf{r}, t) = \frac{1}{\|\mathbf{r} - \mathbf{r}_0\|} g\left(t - \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}\right) \quad (1)$$

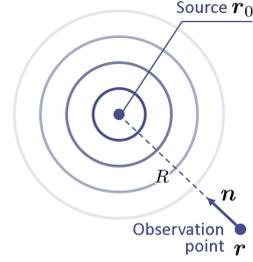
と表される。ただし，

$$\|\mathbf{r} - \mathbf{r}_0\| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \quad (2)$$

である。観測点から波源方向へ向かう単位ベクトルを \mathbf{n} とすると， $f(\mathbf{r}, t)$ の空間微分は，

$$\nabla f(\mathbf{r}, t) = \frac{1}{\|\mathbf{r} - \mathbf{r}_0\|^2} g\left(t - \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}\right) \mathbf{n} \quad (3)$$

$$+ \frac{1}{c\|\mathbf{r} - \mathbf{r}_0\|} \dot{g}\left(t - \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}\right) \mathbf{n} \quad (4)$$

Fig. 1 点音源 \mathbf{r}_0 から観測点 \mathbf{r} へ到来する球面波。

となる。また， $f(\mathbf{r}, t)$ の時間微分は

$$\dot{f}(\mathbf{r}, t) = \frac{1}{\|\mathbf{r} - \mathbf{r}_0\|} \dot{g}\left(t - \frac{\|\mathbf{r} - \mathbf{r}_0\|}{c}\right) \quad (5)$$

となるので， g を消去して，

$$\nabla f(\mathbf{r}, t) = \left\{ \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right\} \mathbf{n} \quad (6)$$

のように，観測信号とその時間・空間微分のみを含む方程式を立てることができる。ただし， $R = \|\mathbf{r} - \mathbf{r}_0\|$ は観測点から波源までの距離である。この式を音源拘束式と呼ぶ [3, 4, 5]。以上のように音源拘束式は，任意の音源信号波形で成り立つ，音源の位置と空間の場の一意な関係を記述する偏微分方程式である。

2.2 波源拘束式に基づく音響信号の確率モデル化

観測信号の空間微分を空間差分で近似するため，以下では図2に示すマイクロホンアレーを仮定する。ただし，マイクロホンの配置は観測信号の空間微分を空間差分で近似できるものであれば良く，以下の理論は図2の配置に限らない。図2のマイクロホンアレーの場合，7本のマイクロホンを用いて各時刻 t_j で，基準点における信号 $f_{0,j}$ およびその各方向の空間差分 $f_{x,j} = \frac{f_{1,j} - f_{2,j}}{2D}$ ， $f_{y,j} = \frac{f_{3,j} - f_{4,j}}{2D}$ ， $f_{z,j} = \frac{f_{5,j} - f_{6,j}}{2D}$ を得ることができる。ただし， j は標本時刻のインデックスを表す。このとき式 (6) は

$$f_{i,j} = \left(\frac{1}{R} f_{0,j} + \frac{1}{c} \dot{f}_{0,j} \right) n_i \quad (7)$$

と表せる。ただし， $i = x, y, z$ で， n_x, n_y, n_z はそれぞれ \mathbf{n} の x, y, z 方向の成分である。式 (7) の左辺を右辺に移項すると

$$-f_{i,j} + \left(\frac{1}{R} f_{0,j} + \frac{1}{c} \dot{f}_{0,j} \right) n_i = 0 \quad (8)$$

が得られる。ここで， $f_{0,j}, f_{x,j}, f_{y,j}, f_{z,j}$ を窓関数で窓掛けして取得された信号とする。切り出し区間の両端点の影響を無視できるものとする，式 (8) は周波数領域で

$$-F_{i,m} + \left(\frac{1}{R} + j \frac{\omega_m}{c} \right) n_i F_{0,m} = 0 \quad (9)$$

*Multiple source localization with Bayesian modeling of spatial gradient spectra. by UENO, Natsuki (University of Tokyo, NTT), KAMEOKA, Hirokazu (NTT)

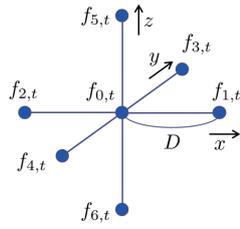


Fig. 2 観測信号 f の空間微分を取得するためのアレイ幾何の例。例えば x 方向の f の空間微分は $\frac{f_{1,j}-f_{2,j}}{2D}$ で近似できる。

と表される。ただし、 $F_{0,m}, F_{x,m}, F_{y,m}, F_{z,m}$ は $f_{0,j}, f_{x,j}, f_{y,j}, f_{z,j}$ の離散 Fourier 変換であり、 m は離散周波数インデックスである。

式 (9) の右辺は差分近似に伴う誤差により実際には必ずしも厳密に 0 にはならない。そこで、式 (8) の右辺を

$$-F_{i,m} + \left(\frac{1}{R} + j \frac{\omega_m}{c} \right) n_i F_{0,m} = \epsilon_{i,m} \quad (10)$$

のように誤差変数 $\epsilon_{x,m}, \epsilon_{y,m}, \epsilon_{z,m}$ に置き換え、これらを平均が 0 で互いに独立な正規確率変数（複素正規分布に従う確率変数）

$$\epsilon_{i,m} \sim \mathcal{N}_{\mathbb{C}}(\epsilon_{i,m}; 0, \sigma_m^2) \quad (11)$$

と仮定する。また、観測点における観測信号の各周波数成分を平均が 0、分散が $\sigma_{0,m}^2$ の正規確率変数とする。これは、

$$F_{0,m} = \epsilon_{0,m} \quad (12)$$

$$\epsilon_{0,m} \sim \mathcal{N}_{\mathbb{C}}(\epsilon_{0,m}; 0, \sigma_{0,m}^2) \quad (13)$$

と仮定することに相当する。

ここで、 $F_{x,m}, F_{y,m}, F_{z,m}, F_{0,m}$ を並べたベクトルと $\epsilon_{x,m}, \epsilon_{y,m}, \epsilon_{z,m}, \epsilon_{0,m}$ を並べたベクトルを

$$\mathbf{f}_m = (F_{x,m}, F_{y,m}, F_{z,m}, F_{0,m})^T \in \mathbb{C}^4 \quad (14)$$

$$\boldsymbol{\epsilon}_m = (\epsilon_{x,m}, \epsilon_{y,m}, \epsilon_{z,m}, \epsilon_{0,m})^T \in \mathbb{C}^4 \quad (15)$$

とする。式 (10) は

$$\mathbf{C}_m(\boldsymbol{\theta}) \mathbf{f}_m = \boldsymbol{\epsilon}_m \quad (16)$$

の形で書ける。ただし $\boldsymbol{\theta} = \{R, \mathbf{n}\}$ であり、 $\mathbf{C}_m(\boldsymbol{\theta})$ は

$$\mathbf{C}_m(\boldsymbol{\theta}) = \begin{bmatrix} -1 & & & \left(\frac{1}{R} + j \frac{\omega_m}{c}\right) n_x \\ & -1 & & \left(\frac{1}{R} + j \frac{\omega_m}{c}\right) n_y \\ & & -1 & \left(\frac{1}{R} + j \frac{\omega_m}{c}\right) n_z \\ & & & 1 \end{bmatrix} \quad (17)$$

で与えられる。式 (11), (13) より、 $\boldsymbol{\epsilon}_m$ は平均が $\mathbf{0}$ 、分散共分散行列 $\boldsymbol{\Sigma}_m$ が

$$\boldsymbol{\Sigma}_m = \text{diag}(\sigma_m^2, \sigma_m^2, \sigma_m^2, \sigma_{0,m}^2) \quad (18)$$

の複素正規分布

$$\boldsymbol{\epsilon}_m \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\epsilon}_m; \mathbf{0}, \boldsymbol{\Sigma}_m) \quad (19)$$

に従う。 $\det \mathbf{C}_m(\boldsymbol{\theta}) = -1 \neq 0$ ($\mathbf{C}_m(\boldsymbol{\theta})$ は正則) であるので、 \mathbf{f}_m は

$$\mathbf{f}_m = \mathbf{C}_m(\boldsymbol{\theta})^{-1} \boldsymbol{\epsilon}_m \quad (20)$$

と表される。よって式 (19) より、 \mathbf{f}_m の確率密度関数

$$p(\mathbf{f}_m | \boldsymbol{\theta}) = \mathcal{N}_{\mathbb{C}}(\mathbf{f}_m; \mathbf{0}, \mathbf{C}_m(\boldsymbol{\theta})^{-1} \boldsymbol{\Sigma}_m \mathbf{C}_m(\boldsymbol{\theta})^{-H}) \quad (21)$$

を得る。以上より、観測スペクトルおよびその空間差分 $\mathbf{f} = \{\mathbf{f}_m\}_m$ が与えられた下で、単一音源を定位する問題は、

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \log p(\mathbf{f} | \boldsymbol{\theta}) \quad (22)$$

$$p(\mathbf{f} | \boldsymbol{\theta}) = \prod_m p(\mathbf{f}_m | \boldsymbol{\theta}) \quad (23)$$

を解く最尤推定問題に帰着する。

2.3 混合音信号の確率モデル化

音声信号や楽音など実世界の音響信号の多くは時間周波数成分がスパースである。従って、複数の音源が同時に混在する場合でも、各時間周波数点では高々一つの音源のみが支配的であると仮定できる場合が多い。この音源の時間周波数成分のスパース性の仮定と以上の \mathbf{f} の確率モデル化に基づき、音源が複数個存在する場合、および雑音が存在する場合の観測信号の確率分布を導くことができる。

信号の切り出しフレームの時刻のインデックスを l 、音源インデックスを $k = 0, \dots, K$ とし、 $k = 0$ は雑音、 $k \neq 0$ は点音源に対応するものとする。また、点音源 k の位置を $\boldsymbol{\theta}^{(k)} = \{R^{(k)}, \mathbf{n}^{(k)}\}$ とする。ここで、雑音を含む全音源の時間周波数成分のスパース性を仮定し、周波数 m 、時刻 l において $z_{m,l}$ 番目の音源のみが非零のパワーをもち、それ以外の音源のパワーを 0 とする。このとき、所与の $z_{m,l}$ の下での観測信号の時間周波数成分とその空間差分 $\mathbf{y}_{m,l} \in \mathbb{C}^4$ (以後、観測信号) の条件付き確率密度関数は

$$p(\mathbf{y}_{m,l} | z_{m,l} = k, \boldsymbol{\vartheta}) = \mathcal{N}_{\mathbb{C}}(\mathbf{y}_{m,l}; \mathbf{0}, \boldsymbol{\Lambda}_{m,l}^{(k)}) \quad (24)$$

で与えられる。ここで、

$$\boldsymbol{\Lambda}_{m,l}^{(k)} = \begin{cases} \mathbf{C}_m(\boldsymbol{\theta}^{(k)})^{-1} \boldsymbol{\Sigma}_{m,l}^{(k)} \mathbf{C}_m(\boldsymbol{\theta}^{(k)})^{-H} & (k \neq 0) \\ \boldsymbol{\Gamma}_{m,l} & (k = 0) \end{cases} \quad (25)$$

$$\boldsymbol{\Sigma}_{m,l}^{(k)} = \text{diag}(\sigma_m^2, \sigma_m^2, \sigma_m^2, \sigma_{0,m,l}^{(k)2}) \quad (26)$$

である。また、 $\boldsymbol{\Gamma}_{m,l}$ は雑音の時間周波数成分の分散共分散行列で、周波数にのみ依存する正規化分散共分散行列モデル \mathbf{W}_m と時刻にも依存する雑音のパワー $\nu_{m,l}^2$ の積

$$\boldsymbol{\Gamma}_{m,l} = \nu_{m,l}^2 \mathbf{W}_m \quad (27)$$

で表されるものとする。拡散雑音を仮定する場合の \mathbf{W}_m の設定方法については [6] を参照されたい。 $\boldsymbol{\vartheta}$ はすべての未知パラメータ $\{\boldsymbol{\theta}^{(k)}\}_k, \{\sigma_m^2\}_m, \{\sigma_{0,m,l}^{(k)}\}_{k,m,l}, \{\nu_{m,l}^2\}_{m,l}$ を表す。 $z_{m,l}$ の事前確率を $p(z_{m,l} = k) = \alpha_{m,l}^{(k)}$ とすると、観測信号 $\mathbf{y} = \{\mathbf{y}_{m,l}\}_{m,l}$ の確率密度

関数 (ϑ の尤度関数) は

$$\begin{aligned} p(\mathbf{y}|\vartheta) &= \prod_{m,l} \sum_{k=0}^K p(\mathbf{y}_{m,l}|z_{m,l}=k, \vartheta) p(z_{m,l}=k) \\ &= \prod_{m,l} \sum_{k=0}^K \alpha_{m,l}^{(k)} \mathcal{N}_{\mathbb{C}}(\mathbf{y}_{m,l}; \mathbf{0}, \mathbf{\Lambda}_{m,l}^{(k)}) \end{aligned} \quad (28)$$

と書ける。以上より、複数の音源と雑音が存在する場合の各音源の位置 $\hat{\theta} = \{\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)}\}$ を推定する問題は、観測信号 \mathbf{y} が与えられた下で

$$\hat{\vartheta} = \underset{\vartheta}{\operatorname{argmax}} \log p(\mathbf{y}|\vartheta) \quad (29)$$

を解く最尤推定問題に帰着する。この最適化問題の大域解は解析的に解くことはできないが、 $Z = \{z_{m,l}\}_{m,l}$ を潜在変数とした Expectation-Maximization (EM) アルゴリズムにより停留点を探索することができる。

2.4 ノンパラメトリックベイズモデリング

実環境では音源数が未知の場合が多い。上記の定式化では音源数 K を既知と仮定したが、音源数を仮定せずともモデルの複雑度を適応させながら実際の音源数に合わせて音源定位を行えるようにすることが望ましい。そこで、本章では上記の生成モデルを [7] と同様に Dirichlet 過程混合モデルに拡張する。 $k=0$ を雑音、 $k=1, \dots, \infty$ を点音源のインデックスとし、 $z_{m,l}$ を加算無限次元の離散分布 $\pi_0, \pi_1, \dots, \pi_\infty$ に従って生成される確率変数とする。

$$z_{m,l} \sim \text{Categorical}(\pi_0, \pi_1, \dots, \pi_\infty) \quad (30)$$

ここで、点 (m,l) において雑音が支配的である ($z_{m,l}=0$ となる) 確率 π_0 を、超パラメータ α_b, β_b のベータ分布に従って生成される変数

$$\pi_0 = v_0 \quad (31)$$

$$v_0 \sim \text{Beta}(\alpha_b, \beta_b) \quad (32)$$

とし、点音源 $k \neq 0$ が支配的である ($z_{m,l}=k$ となる) 確率 π_k を、棒折り過程 [8] に従って決まる変数

$$\pi_k = v_k \prod_{j=0}^{k-1} (1 - v_j) \quad (33)$$

$$v_k \sim \text{Beta}(1, \beta_s) \quad (34)$$

とする。以上のプロセスで生成される π_1, π_2, \dots の期待値は、 k が大きいほど指数的に小さくなる傾向を持つため、大きい k に対応した音源ほどアクティブになる確率が低くなる。よって、観測信号からパラメータを推論する際、必要最小限の音源インデックス数の混合モデルで観測信号を説明しようとする効果がもたらされる。以上の生成モデルにおいて、全未知変数 Θ は以下となる。

$$\Theta = \{N, \rho, Z, V, \lambda, \zeta, \gamma\} \quad (35)$$

$$N = \{\mathbf{n}^{(k)}\}_k \quad (36)$$

$$\rho = \{\rho^{(k)}\}_k \quad (\rho^{(k)} = 1/R^{(k)}) \quad (37)$$

$$Z = \{z_{m,l}\}_{m,l} \quad (38)$$

$$V = \{v_k\}_k \quad (39)$$

$$\lambda = \{\lambda_m\} \quad (\lambda_m = 1/\sigma_m^2) \quad (40)$$

$$\zeta = \{\zeta_{m,l}^{(k)}\} \quad (\zeta_{m,l}^{(k)} = 1/\sigma_{0,m,l}^{(k)2}) \quad (41)$$

$$\gamma = \{\gamma_{m,l}\} \quad (\gamma_{m,l} = 1/\nu_{m,l}^2) \quad (42)$$

2.5 変分推論アルゴリズム

以上の生成モデル化により、

$$\begin{aligned} p(Y|N, \rho, Z, \lambda, \zeta, \gamma) \\ = \prod_{m,l} \prod_{k=0}^{\infty} \mathcal{N}_{\mathbb{C}}(\mathbf{y}_{m,l}; \mathbf{0}, \mathbf{\Lambda}_{m,l}^{(k)}) \mathbf{1}^{[z_{m,l}=k]} \end{aligned} \quad (43)$$

$$p(Z|V) = \prod_{m,l} \prod_{k=0}^{\infty} \pi_k \mathbf{1}^{[z_{m,l}=k]} \quad (44)$$

$$p(V) = \text{Beta}(v_0; \alpha_b, \beta_b) \prod_{k=1}^{\infty} \text{Beta}(v_k; 1, \beta_s) \quad (45)$$

と書ける。また、その他の事前分布 $p(V)$, $p(N)$, $p(\rho)$, $p(\lambda)$, $p(\zeta)$ 及び $p(\gamma)$ を

$$p(N) = \prod_{k=1}^{\infty} \text{vMF}(\mathbf{n}^{(k)}; \boldsymbol{\xi}^{(k)}, \kappa^{(k)}) \quad (46)$$

$$p(\rho) = \prod_{k=1}^{\infty} \mathcal{N}(\rho^{(k)}; \mu^{(k)}, \eta^{(k)}) \quad (47)$$

$$p(\lambda) = \prod_{m,l} \text{Gamma}(\lambda_m; \alpha_{\lambda,m}, \beta_{\lambda,m}) \quad (48)$$

$$p(\zeta) = \prod_{m,l} \prod_{k=1}^{\infty} \text{Gamma}(\zeta_{m,l}^{(k)}; \alpha_{\zeta,m,l}^{(k)}, \beta_{\zeta,m,l}^{(k)}) \quad (49)$$

$$p(\gamma) = \prod_{m,l} \text{Gamma}(\gamma_{m,l}; \alpha_{\gamma,m,l}, \beta_{\gamma,m,l}) \quad (50)$$

と仮定すると、観測信号 $Y = \{\mathbf{y}_{m,l}\}_{m,l}$ と未知変数 Θ の同時分布

$$\begin{aligned} p(\Theta, Y) &= p(Y|N, \rho, Z, \lambda, \zeta, \gamma) p(Z|V) \\ &\quad p(V) p(N) p(\rho) p(\lambda) p(\zeta) p(\gamma) \end{aligned} \quad (51)$$

を具体的に記述することができる。ただし、 $\text{vMF}(\cdot)$ は von Mises-Fisher 分布、 $\mathcal{N}(\cdot)$ は実正規分布、 $\text{Gamma}(\cdot)$ はガンマ分布を表す。

Θ の事後分布 $p(\Theta|Y) = p(\Theta, Y)/p(Y)$ を解析的に得ることは難しいが、変分推論法に基づき近似分布を反復計算により得ることができる。変分推論は、 $\int q(\Theta) d\Theta = 1$ を満たす非負の変関数 $q(\Theta)$ を、事後分布 $p(\Theta|Y)$ との間の Kullback-Leibler (KL) ダイバージェンス

$$\text{KL}(q(\Theta)||p(\Theta|Y)) = \int q(\Theta) \log \frac{q(\Theta)}{p(\Theta|Y)} d\Theta \quad (52)$$

が小さくなるように求める方法である。 $q(\Theta)$ を

$$q(\Theta) = q(N) q(\rho) q(Z) q(V) q(\lambda) q(\zeta) q(\gamma) \quad (53)$$

のように近似できると仮定すると、 $q(N)$, $q(\rho)$, $q(Z)$, $q(V)$, $q(\lambda)$, $q(\zeta)$, $q(\gamma)$ について反復的に式 (52) を最小化することで $p(\Theta|Y)$ の近似分布を得ることができる。また、 $q(Z)$ に関して、以下の打ち切り近似

$$q(z_{m,l} = K^* + 1) = \dots = q(z_{m,l} = \infty) = 0 \quad (54)$$

を行う。この近似はモデルの複雑度 (音源数) を固定したことを意味するのではなく q の関数空間をある範囲に限定したことを意味する。従って、 $p(\Theta|Y)$ をできるだけ良く近似したければ K^* を大きくして $q(\Theta)$ が取りうる範囲を広くすれば良い。

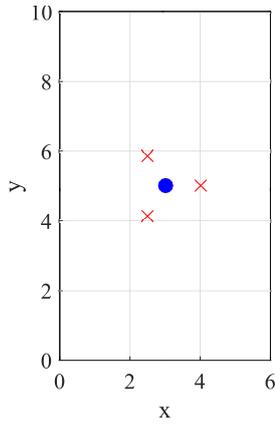


Fig. 3 音源位置とマイクロホン位置。赤バツ印が音源位置。青マル印がマイクロホンアレーの中心位置

導出は省略するが、式 (52) を各 q について最小化する更新式は以下の形となる。

$$q(N) = \prod_{k=1}^{K^*} \text{vMF}(\mathbf{n}^{(k)}; \bar{\boldsymbol{\xi}}^{(k)}, \bar{\boldsymbol{\kappa}}^{(k)}) \quad (55)$$

$$q(\rho) = \prod_{k=1}^{K^*} \mathcal{N}(\rho^{(k)}; \bar{\mu}^{(k)}, \bar{\eta}^{(k)}) \quad (56)$$

$$q(Z) = \prod_{m,l} \bar{\varphi}_{m,l}^{(z_{m,l})} \quad (57)$$

$$q(V) = \prod_{k=0}^{K^*} \text{Beta}(v_k; \bar{\beta}_0^{(k)}, \bar{\beta}_1^{(k)}) \quad (58)$$

$$q(\lambda) = \prod_{m,l} \text{Gamma}(\lambda_{m,l}; \bar{\alpha}_{\lambda,m,l}, \bar{\beta}_{\lambda,m,l}) \quad (59)$$

$$q(\zeta) = \prod_{m,l} \prod_{k=1}^{K^*} \text{Gamma}(\zeta_{m,l}^{(k)}; \bar{\alpha}_{\zeta,m,l}^{(k)}, \bar{\beta}_{\zeta,m,l}^{(k)}) \quad (60)$$

$$q(\gamma) = \prod_{m,l} \text{Gamma}(\gamma_{m,l}; \bar{\alpha}_{\gamma,m,l}, \bar{\beta}_{\gamma,m,l}) \quad (61)$$

3 数値実験

提案手法の性能を検証するため、残響環境下の複数音源定位の数値シミュレーションを行った。今回、 x 方向、 y 方向の2次元モデルを使用した。部屋サイズは $6 \text{ m} \times 10 \text{ m} \times 4 \text{ m}$ とし、中心に 0.03 m 間隔で7つのマイクロホンを図2のように配置した。壁面の反射係数は 0.7308 及び 0.4566 (Sabineの残響公式による残響時間がそれぞれ 0.5 s 及び 0.2 s に相当) とした。マイクロホンのサンプリング周波数は 32 kHz 、短時間 Fourier 変換のフレーム長は 64 ms (オーバーラップは 32 ms)、観測信号全体の長さは 2779 ms 及び 1665 ms とした。音源数は3個、音源の位置は部屋の中からそれぞれ $(1, 0, 0) \text{ m}$ 、 $(-0.5, 0.87, 0) \text{ m}$ 、 $(-0.5, -0.87, 0) \text{ m}$ とした。音源信号は話速バリエーション型音声データベース (SRV-DB) を利用した。提案法は変分推論に基づく手法 (VBEM)、正しい音源数 (3個) を仮定した EM アルゴリズムによる手法 (EM1)、誤った音源数 (6個) を仮定した EM アルゴリズムによる手法 (EM2) を評価し、従来法である MUSIC 法との比較を行った。EM アルゴリズムにおいては定常雑音を仮定し $\nu_{m,l}^2 = \hat{\nu}_m^2$ とした。どの手法においても、検出閾値をもとに音源の有無及び方向を推定し、そのうち真の音源方向との角度の誤差が $\pm\tau$ 以内のものがあれば正解、なければ脱落誤り、検出された音源方向のうちどの真の音源方向にも属さ

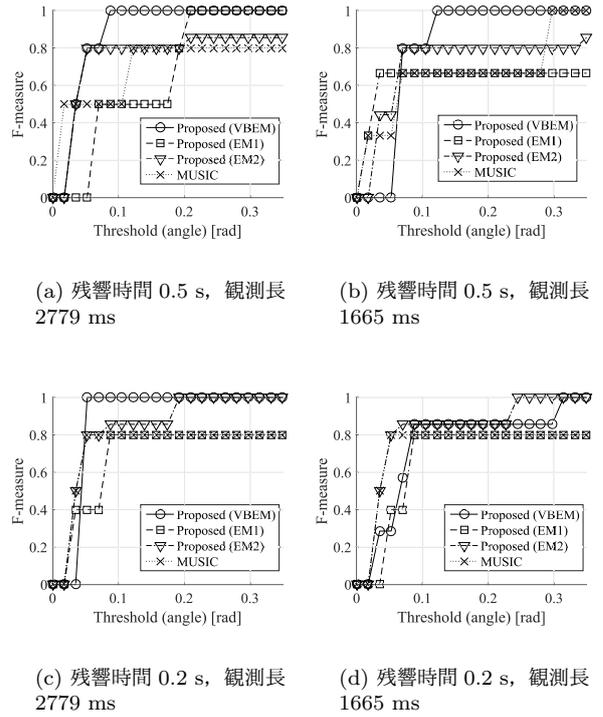


Fig. 4 実験結果

ないものを誤挿入として、F 尺度を算出した。各 τ について、検出閾値を変化させたときの最も高い F 尺度のプロットを図4に示す。多くの場合提案法、特に変分推論に基づく手法が他の手法より高精度な定位を達成していることが確認できる。

4 結論

本研究は、波源拘束変微分方程式を基に音響信号の空間勾配スペクトルをベイズモデリングすることで小領域・瞬時観測による複数音源の定位を可能にする手法を提案し、数値実験で有用性を示した。

謝辞 本研究は JSPS 科研費 26730100 の助成を受けて行われた。

参考文献

- [1] R. O. Schmidt, IEEE Trans. AP, 34 (3), 276–280, 1986.
- [2] C. H. Knapp et al., IEEE Trans. ASSP, 24(4), 320–327, 1976.
- [3] 藤田他, 音講論 (秋), 3-1-3, 483–484, 2006.
- [4] S. Ando et al., Proc. ICSV14, 2007.
- [5] 小山他, 音講論 (秋), 2-8-20, 679–682, 2008.
- [6] 鈴木&亀岡, 音講論 (春), 3-3-4, 615–618, 2016.
- [7] H. Kameoka et al., IEICE Trans. FE, E96-A(10), 1928–1937, 2013.
- [8] J. Sethuraman, Statistica Sinica, 4, 639–650, 1994.