拡散音場モデルに基づく残響環境下での信頼度付時間差検出*

◎小野順貴, 和泉洋介, 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

音声のスパース性を利用したブラインド音源分離 [1, 2, 3, 4] は観測信号数より音源信号数が多くても適用可能であり，マイクロフォン数が少ない場合に特に有効な分離手法である。スパース性に基づく音源分離は通常，1) 時間差・強度比のクラスタリングによる音源定位と，2) 定位結果に基づく時間周波数マスキングや逆フィルタによる分離，により行なわれるが，残響環境下においては周囲の壁面等からの反射音が複雑に直接音に重畳し，少ないマイクロフォン数では音源定位自体が困難な状況となる。

残響は大きさも到来方向も非定常であり，また観測雑音としては，マイクロフォン間で周波数と残響到来方向に依存した相関をもつ点がその扱いを難しくしている。本稿ではこれに対し，1) 残響環境を拡散音場とみなして残響の空間的相関の周波数依存性を確率的にモデル化し，2) ある連続したフレームの観測区間毎に，音源方向 (時間差) と共に非定常な観測雑音の大きさや各観測区間の信頼度を推定し，信頼度の高い結果を統合することで残響環境下での音源定位 (時間差検出) を改善する手法を提案する。

2 信頼度付時間差検出の理論

2.1 観測モデル

以下ではマイクロフォン 2 個の場合に議論をしぼり，2 信号 $m_L(t)$，$m_R(t)$ を時間周波数分解し，ベクトル表示したものを

$$\boldsymbol{M}_i(\omega) = \begin{pmatrix} M_{Li}(\omega) \\ M_{Ri}(\omega) \end{pmatrix} \tag{1}$$

のように表す。ただし，i はフレーム番号，ω は角周波数，L，R の添え字はそれぞれ，左右のマイクロフォンで取得された信号であることを表す。

ある連続したフレームの観測区間において active な音源が 1 つであると仮定すると，観測信号は，

$$\boldsymbol{M}_i(\omega) = S_i(\omega)\boldsymbol{b}(\omega) + \boldsymbol{N}_i(\omega) \tag{2}$$

のように表される。ここで $S_i(\omega)$ は音源信号，$N_i(\omega)$ は残響などに起因する観測誤差，$\boldsymbol{b}(\omega)$ は音源位置に対応するベクトルを表し，また $|\boldsymbol{b}(\omega)| = 1$ のように規格化されているものとする。$|\boldsymbol{b}(\omega)|$ の 2 成分

$$\boldsymbol{b}(\omega) = \begin{pmatrix} b_L(\omega) \\ b_R(\omega) \end{pmatrix} \tag{3}$$

に対して，

$$a = \mathrm{Re}\left[\mathrm{Log}\frac{b_R(\omega)}{b_L(\omega)}\right],\ \tau = \mathrm{Im}\left[\mathrm{Log}\frac{b_R(\omega)}{b_L(\omega)}\right]/\omega \tag{4}$$

を定義すると，a，τ は，2 つのマイクロフォンで観測される観測信号の振幅比の対数と時間差に対応し，理想的には角周波数 ω に依らず，音源位置で決まる定数となる。以下では式の煩雑さを避けるため，引数である ω などは省略して表記する。

2.2 残響成分の拡散音場モデル

拡散音場とは，あらゆる方向からエネルギーが等しく無相関な平面波が等確率で到来する音場であり，3 次元中で D 離れた 2 点間の音圧の相互相関係数は，

$$\eta = \sin\frac{\omega D}{c}\Big/\left(\frac{\omega D}{c}\right) \tag{5}$$

のように表される [5]。式 (2) 中の観測誤差がこの拡散音場モデルに従うとすると，共分散行列は

$$E[\boldsymbol{N}\boldsymbol{N}^h] = \sigma^2 V,\quad V = \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix} \tag{6}$$

とかける。ただし，^{h} はエルミート転置，σ^2 は誤差分散の大きさを表す。密度分布としてガウス分布を仮定すれば，観測量 $\boldsymbol{M}_i(i = 1, \cdots, k)$ に対する未知数 S_i，$\boldsymbol{b}$，σ の対数尤度は

$$\begin{aligned} L(\boldsymbol{S}, \boldsymbol{b}, \sigma) &= -k\log 2\pi\sigma - \frac{k}{2}\log(1-\eta^2) \\ &\quad -\frac{1}{2\sigma^2}\sum_{i=1}^{k}(\boldsymbol{M}_i - S_i\boldsymbol{b})^h V^{-1}(\boldsymbol{M}_i - S_i\boldsymbol{b}) \end{aligned} \tag{7}$$

のように表される。ただし $\boldsymbol{S} = (S_1\ \cdots\ S_k)$ である。

2.3 最尤解の導出

紙面の制約のため導出は省略するが，$|\boldsymbol{b}| = 1$ の条件の下で式 (7) を最大にする最尤解は

$$\hat{\sigma}^2 = \frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{M}_i - \hat{S}_i\hat{\boldsymbol{b}})^h V^{-1}(\boldsymbol{M}_i - \hat{S}_i\hat{\boldsymbol{b}}) \tag{8}$$

$$\hat{S}_i = \frac{\hat{\boldsymbol{b}}^h V^{-1}\boldsymbol{M}_i}{\hat{\boldsymbol{b}}^h V^{-1}\hat{\boldsymbol{b}}},\quad \hat{\boldsymbol{b}} = \frac{UD^{1/2}\tilde{\boldsymbol{b}}_{max}}{|UD^{1/2}\tilde{\boldsymbol{b}}_{max}|} \tag{9}$$

のように得られる。ただし，

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},\ D = \begin{pmatrix} 1-\eta & 0 \\ 0 & 1+\eta \end{pmatrix} \tag{10}$$

であり，$\tilde{\boldsymbol{b}}_{max}$ は相関行列

$$\Phi = \sum_{i=1}^{k}\boldsymbol{M}V^{-1}\boldsymbol{M}^h \tag{11}$$

の最大固有値 λ_{max} に対応する固有ベクトルである。このとき最大化された対数尤度は，

$$L(\hat{\boldsymbol{S}}, \hat{\boldsymbol{b}}, \hat{\sigma}) = C - k\log\lambda_{min} \tag{12}$$

と表される。ただし C は定数，λ_{min} は Φ の最小固有値である。

---

*Time Delay Detection with Reliability Measure Based on Diffuse Sound Field Model under Reverbration Environment by ONO, Nobutaka, IZUMI, Yosuke, and SAGAYAMA, Shigeki (The University of Tokyo)

2.4 信頼度の指標

信頼度の指標としては対数尤度 (式 (12)) が考えられるが，本稿での定式化では観測区間における誤差分散を標本分散から求めているため，本来信頼できないはずの無音区間で誤差分散が小さくなり，対数尤度が大きくなってしまう問題がある。別の指標として考えられるのは，最尤方向 $\hat{\boldsymbol{b}}$ 近傍での対数尤度の 2 階微分の大きさである。この値が大きいほど，対数尤度は $\hat{\boldsymbol{b}}$ 近傍で急峻なピークをもつため，信頼性が高いと考えられる。$|\boldsymbol{b}| = 1$ に注意すると，$\boldsymbol{b}$ は複素定数倍の任意性を除き

$$\boldsymbol{b} = \boldsymbol{b}_{max} \cos\psi + e^{jp} \boldsymbol{b}_{min} \sin\psi \qquad (13)$$

と角度 ψ をパラメータとして表せるので，$\psi \ll 1$(最尤方向である $\boldsymbol{b}_{max}$ 近傍) の場合，対数尤度は

$$\begin{aligned} & L(\hat{\boldsymbol{S}}, \boldsymbol{b}, \hat{\sigma}) - C \\ = \; & -k \log(\lambda_{min} + \lambda_{max} - \lambda_{max} \cos^2\psi - \lambda_{min} \sin^2\psi) \\ \simeq \; & -k \log \lambda_{min} - k \left(\frac{\lambda_{max}}{\lambda_{min}} - 1 \right) \psi^2 \qquad (14) \end{aligned}$$

のように表される。すなわち，最尤方向近傍での 2 階微分の大きさは，$\lambda_{max}/\lambda_{min} - 1$ により得られる。

3 時間差検出実験と結果

3.1 実験条件

鏡像法 [6] により残響環境下での音場をシミュレートし，本手法を適用して時間差検出の実験を行なった。シミュレーションで用いた音源やマイクロフォンの 2 次元的な配置は Fig. 1 に示す通りであり，高さ方向の位置は全て床から 1.5[m] とした。壁面の反射率は 0.9 とし，470[ms] の残響時間の環境をシミュレートした。音源信号には ATR の音声データベース B セットの連続音声をサンプリング周波数 16kHz に変換したものを用いた。分析条件はフレーム長 1024 点，またフレームシフトは 16 点と短くとり，中心フレームの前後 15 フレーム計 31 フレームを観測区間として，15 フレーム毎に推定を行なった。

3.2 実験結果

Fig. 2 に時間差検出の結果を示す。(a)，(b)，(c) ともに横軸はサンプル単位で表した時間差であり，全フレームから求めた結果を散布図として表している。また，青い縦線は真の時間差を表す。(a) は 1 フレームの観測信号 $M_{Li}(\omega)$，$M_{Ri}(\omega)$ の位相差から時間差を求めた従来法の結果であり，各フレームのパワーを信頼度と考え，縦軸にプロットしたものである。残響のために検出結果は真の値の周りに広く分散し，音源ごとのピークは形成されない。これに対し (b) は本提案手法であり，縦軸には信頼度として $\lambda_{max}/\lambda_{min} - 1$ をプロットした。信頼度が高い検出結果は真の音源位置の近くに分布し，分布のピークが真の音源位置に対応している様子が確認できる。一方 (c) は，式 (6) において $\eta = 0$ とし，拡散音場モデルではなく独立なガウス雑音を仮定して，提案法と同様の処理で求めた結果である。相関のある残響を独立な雑音として扱ったために，ピーク位置が中心方向にバイアスされているようにみえる。(b) と (c) の比較から拡散音場モデルが，残響環境下での時間差検出に有効に働いていることが確認できる。本手法を用いたブラインド音源分離への適用は [7] で報告する。

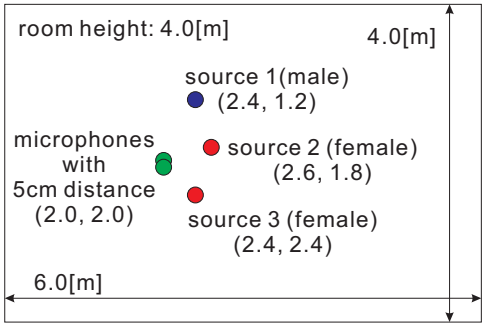

Fig. 1 音源とマイクロフォンの 2 次元配置。() 内は [m] 単位の座標値を表す。

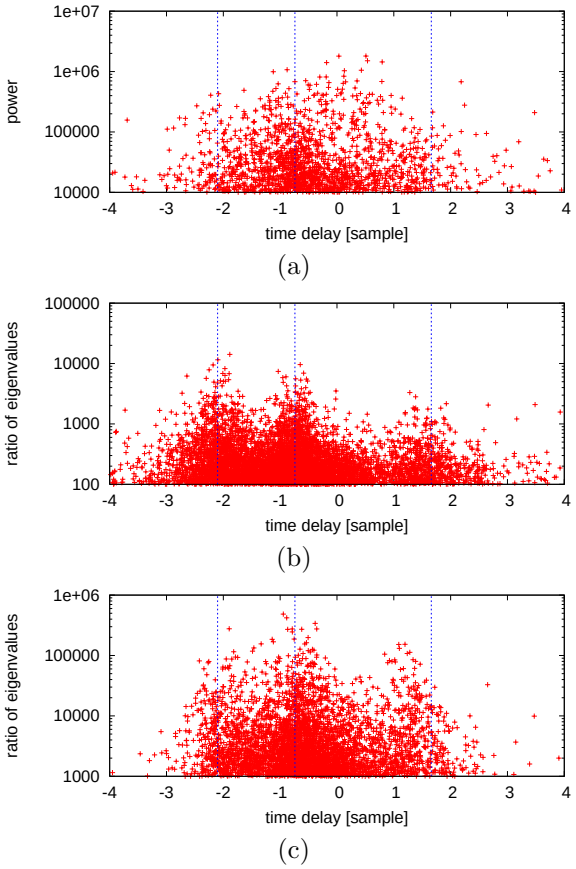

Fig. 2 時間差検出結果。(a)：1 フレームの左右信号位相差から求める従来法、(b)：提案法、(c)：拡散音場モデルを使わなかった場合の提案法 ($\eta = 0$)

謝辞 本研究の一部は科学研究費補助金・若手研究 (B)(課題番号 18760303) の補助を受けて行なわれたので，ここに謝意を表する。

参考文献

[1] M. Baeck et al., Proc. DAFx-03, Sep., 2003.
[2] Ö. Yilmaz et al., IEEE Trans. on SP, vol. 52, no. 7, pp. 1830–1847, 2004.
[3] A. Blin, et al., IEICE Trans. Fundamentals, vol. E88-A, no. 7, pp. 1693-1700, July 2005.
[4] 荒木他, 音講論 (秋), pp. 591-592, 9 月, 2005.
[5] R. K. Cook et al., JASA, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.
[6] J. B. Allen et al., JASA, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[7] 和泉他, 音講論 (秋), 1-1-11 in CD-ROM, 9 月, 2006.