

高次モーメント分析に基づく非線形GCCによる高精度到来時間差推定*

☆タク暁東, 杉本侑哉, 宮部滋樹, 山田武志, 牧野昭二 (筑波大)

1 はじめに

到来時間差 (TDOA: Time Difference of Arrival) 推定技術は, マイクロホンアレーで観測した信号の分析により, 音源信号の各チャンネルに対する到来時間差を推定する技術である. 音源信号の到来時間差は音源分離や, 移動ロボットの環境認識システム, 監視システム, 騒音の分析など, 多くの応用分野で事前情報として必要であり, 目的や環境に合わせて, 様々な研究が行われている.

代表的な TDOA 推定手法の1つとして, チャンネル間の相関分析に基づき, 雑音・残響などの環境に合わせた周波数重み付けによって推定精度を向上させた GCC 法 [1] が知られている. GCC 法は, ステレオ録音など少ないチャンネル数の信号しか利用できない条件でも高精度な TDOA 推定が可能であるという利点を持つ. 現実の多くの録音データはステレオ環境のものであることから実用性が高く, 大規模なマイクロホンアレーによる多数音源の位置同定問題や, マイクロホン, 音源とともに位置情報は未知として信号処理を行う分散型マイクロホンアレーなどの, ある1つの音源が各マイクペアに与える到来時間差を正確に求める必要のあるアプリケーションにおいても欠かすことができない.

本稿では, GCC 法の推定性能を更に向上させるために, 非線形関数 [2, 3] により観測信号を高次元空間へ写像し, 写像空間において GCC 法と同様の分析を行う写像 GCC 法を提案する. 提案手法の TDOA 推定精度を評価するため, 実環境実験とシミュレーション実験を行い, GCC 法や MUSIC [4] などの従来手法よりも雑音と残響に対して頑健であることを確認する.

2 問題設定

本研究で扱う TDOA 推定問題を定式化するために, 2チャンネル, 1音源を仮定した時間領域の観測信号モデルを次式に示す.

$$x_1(t) \approx (\alpha_1 \text{sinc}(t - \tau_1) + h_1(t)) * s(t) + n_1(t) \quad (1)$$

$$x_2(t) \approx (\alpha_2 \text{sinc}(t - \tau_1 + \Delta\tau) + h_2(t)) * s(t) + n_2(t) \quad (2)$$

$$\Delta\tau \triangleq \tau_1 - \tau_2 \quad (3)$$

$$\text{sinc}(t) = \frac{\sin\pi t}{\pi t} \quad (4)$$

ここで, * は畳み込み, t は時間, α_i は直接波が第 i チャンネルに伝搬するときの減衰係数, τ_i は直接波が第 i チャンネルに到達するまでの時間, x_i は第 i チャンネルの観測信号, h_i は音源と第 i チャンネル間の直接波成分を除いたインパルス応答, s は音源信号, n は観測雑音を表す. 音源信号と観測雑音は 0 平均の定常過程で互いに統計的独立, インパルス応答は時不変とする.

TDOA 推定問題で目的とするのは音源から各チャンネルへの直接波が形成する TDOA $\Delta\tau$ であるため, 本稿の問題設定を観測信号 $x_1(t)$, $x_2(t)$ の分析によって, 目的の音源信号が与える真の TDOA $\Delta\tau$ を推定することとする.

3 従来手法

3.1 GCC 法

TDOA 推定の最も基本的なアプローチである相互相関法の原理と, GCC 法によるその拡張について述べる. 2章の信号モデルに基づく, チャンネル間の相互相関関数 $C_{x_1x_2}(\tau)$ は以下のように展開できる.

$$\begin{aligned} C_{x_1x_2}(\tau) &= E[x_1(t)x_2(t-\tau)] \\ &\approx \alpha_1\alpha_2 E[s(t-\tau_1)s(t-\tau_1+\Delta\tau-\tau)] \\ &\quad + \alpha_1 E[s(t-\tau_1)(h_2(t-\tau) * s(t-\tau))] \\ &\quad + \alpha_2 E[s(t-\tau_1+\Delta\tau-\tau)(h_1(t) * s(t))] \\ &\quad + E[(h_1(t) * s(t))(h_2(t-\tau) * s(t-\tau))] \end{aligned} \quad (5)$$

このとき, チャンネル間で直接波が形成する真の TDOA $\Delta\tau$ の推定が目的であるため, 式 (5) において, 直接波どうしの相互相関である第 1 項のみを評価できればよい. したがって, 直接波は干渉音よりも十分大きなパワーを持っていると仮定し, $C_{x_1x_2}(\tau)$ 全体の評価を第 1 項の評価の近似と見なして, τ による最大化を行うのが相互相関法である.

$$\hat{\tau} = \arg \max_{\tau} C_{x_1x_2}(\tau) \quad (6)$$

ここで, $\hat{\tau}$ は相互相関関数の最大化により推定された TDOA を表す. このとき, 相互相関関数の近似は以下のような手順により行う. 直接波のパワーが大きい場合, 式 (5) に現れる時間方向のデータ系列を, 各

* Time Difference of Arrival estimation by the non-linear GCC based on higher-order moment analysis. by Xiaodong ZHAI, Yuya SUGIMOTO, Shigeki MIYABE, Takeshi YAMADA, and Shoji MAKINO (University of Tsukuba)

チャンネルで観測する直接波は $\mathbf{d}_1, \mathbf{d}_2$ 、直接波以外の干渉音は $\mathbf{r}_1, \mathbf{r}_2$ とベクトル表現すると、これらのノルムの間には式 (9) の関係が成り立つ。

$$\mathbf{d}_i = [\dots, \alpha_i s(t), \dots] \quad (7)$$

$$\mathbf{r}_i = [\dots, h_i(t) * s(t), \dots] \quad (i = 1, 2) \quad (8)$$

$$\|\mathbf{d}_1\| \approx \|\mathbf{d}_2\| \gg \|\mathbf{r}_1\| \approx \|\mathbf{r}_2\| \quad (9)$$

式 (7)–(9) より、式 (5) の第 1 項以外は無視できるものとする。

$$C_{x_1 x_2}(\tau) \approx \alpha_1 \alpha_2 E [s(t - \tau_1) s(t - \tau_1 + \Delta\tau - \tau)] \quad (10)$$

しかし、現実の多くの環境では上の仮定は成り立たず、相互相関法によって十分な精度の TDOA 推定が行えるとは言い難い。そこで、雑音や反射波に対する頑健性を確保するために、相互相関に周波数フィルタリングを行う試みが成されており、GCC 法と呼ばれている。GCC 法によるフィルタリングには、チャンネル間のクロススペクトル白色化によって相互相関関数のピークを鋭くする、干渉音の影響を低減するなどの効果があり、環境に合わせてフィルタを適切に設計することで、雑音・残響下の TDOA 推定精度を大きく向上させる。次節で、周波数領域における GCC 法の処理を述べる。

3.2 GCC 法の行列表現による計算

一般化相関行列とステアリングベクトルを用いた GCC 法の TDOA 推定アルゴリズムについて述べる。なお、時間周波数領域の信号は $\mathbf{x}(\omega, t)$ と表し (ただし t は時間フレームとする)、 i, j ($i, j = 1, 2$) はマイクのチャンネルインデックスとする。

周波数領域における GCC 法の計算は、マイクペア $\{x_1, x_2\}$ に対して定義される相互相関関数 $C_{x_1 x_2}$ とクロススペクトル $G_{x_1 x_2}$ がフーリエ変換対であることを利用している。クロススペクトルは周波数領域の観測ベクトルが与える共分散行列の非対角要素に相当するため、観測信号の共分散行列 $\mathbf{R}(\omega)$ を求め、さらに GCC 法の周波数フィルタを表現するフィルタ行列 $\Psi(\omega)$ とのアダマール積をとることで一般化相関行列 $\mathbf{R}_g(\omega)$ を構成する。

$$\mathbf{R}_g(\omega) = \Psi(\omega) \oplus \mathbf{R}(\omega) \quad (11)$$

$$\Psi(\omega) = [\psi_{ij}(\omega)]_{ij} \quad (12)$$

$$\begin{aligned} \mathbf{R}(\omega) &= E [\mathbf{x}(\omega, t) \mathbf{x}^H(\omega, t)] \\ &= [G_{x_i x_j}(\omega)]_{ij} \end{aligned} \quad (13)$$

$$\mathbf{x}(\omega, t) = [x_1(\omega, t), x_2(\omega, t)]^T \quad (14)$$

$$G_{x_i x_j}(\omega) = E [x_i(\omega, t) x_j^*(\omega, t)] \quad (15)$$

ここで、 $[\cdot]^H$ は複素共役転置、 \oplus はアダマール積演算、 $\psi_{ij}(\omega)$ はマイクペア $\{x_i, x_j\}$ に対する周波数フィ

ルタ、 $[\cdot]_{ij}$ は (i, j) 要素が引数により与えられる行列を表す。

3.1 節では時間領域において、相互相関関数を最大化する時間差 τ の探索によって TDOA を推定したが、これを周波数領域における等価な計算に置き換えると、クロススペクトルに位相差 $\exp(j\omega\tau)$ を付与して、周波数方向の積分を最大化するような τ を探索する操作となる。クロススペクトルは一般化相関行列 $\mathbf{R}_g(\omega)$ の非対角要素に相当することから、この操作は以下のような狭帯域の 2 次形式 TDOA スコア $f_g(\omega; \tau)$ を周波数平均した $\bar{f}_g(\omega; \tau)$ の最大化問題として表現できる。

$$f_g(\omega; \tau) \triangleq \mathbf{b}^H(\omega; \tau) \mathbf{R}_g(\omega) \mathbf{b}(\omega; \tau) \quad (16)$$

$$\begin{aligned} \mathbf{b}(\omega; \tau) &\triangleq \frac{1}{\sqrt{2}} [1, \exp(-j\omega\tau)] \\ (\|\mathbf{b}(\omega; \tau)\|_2 &= 1) \end{aligned} \quad (17)$$

$$\bar{f}_g(\omega; \tau) \triangleq \frac{1}{|\Omega|} \sum_{\omega \in \Omega} f_g(\omega; \tau) \quad (18)$$

$$\hat{\tau} = \arg \max_{\tau} \bar{f}_g(\omega; \tau) \quad (19)$$

ここで、 $\mathbf{b}(\omega; \tau)$ はチャンネル間に位相差を付与するステアリングベクトル、 Ω は平均する周波数ビンの集合を表す。なお、 $\mathbf{b}(\omega; \tau)$ はノルムが 1 となるように正規化されているが、2 次形式の非対角要素に対応する項を最大化するには位相情報のみが寄与し、伝搬における振幅の減衰を考慮する必要がないためである。

GCC 法では、周波数フィルタ $\psi_{ij}(\omega)$ を観測信号の性質に合わせて適切に設計することで、誤差が大きいと推定される周波数帯域の抑制、白色化による推定ピークの先鋭化などの効果が得られる。以下では、具体的なフィルタの設計例として GCC-SCOT と GCC-PHAT を示し、簡潔に特徴を述べる。

• GCC-SCOT

$$\psi_{ij}(\omega) = \frac{1}{\sqrt{G_{x_i x_i}(\omega) G_{x_j x_j}(\omega)}} \quad (20)$$

コヒーレンス関数に基づくフィルタ。雑音の自己スペクトル $G_{x_i x_i}, G_{x_j x_j}$ の影響をよく低減するため耐雑音性が高いとされる。

• GCC-PHAT

$$\psi_{ij}(\omega) = \frac{1}{|G_{x_i x_j}(\omega)|} \quad (21)$$

振幅特性を打ち消して、信号の白色化を行うフィルタである。残響の影響に強く、反射波耐性が高いとされる。

4 提案手法

4.1 写像 GCC による TDOA 推定のアルゴリズム

GCC 法を高次元写像により拡張し、TDOA の推定精度を向上させた写像 GCC 法を提案する。以降、マイク数を 2 チャンネルから M チャンネルに一般化して議論する ($i, j = 1, \dots, M$)。写像 GCC 法は、観測信号 $\mathbf{x}(\omega, t)$ を非線形関数 $\phi: \mathbb{C}^M \rightarrow \mathbb{C}^{M'}$ ($M' \geq M$) により高次元写像し、 M' 次元ユークリッド空間でベクトル $\phi(\mathbf{x}(\omega, t))$ を用いて GCC 法と同様の分析を行う。以下では、写像 GCC 法の TDOA 推定アルゴリズムについて述べる。写像 GCC 法が正しく方向推定を行うためには、写像の前で各素子間の相関情報が保持される必要があり、この条件は式 (21)–(23) により表される。

1. ノルムの大小関係を維持する

$$\|\mathbf{x}\|_2 \geq \|\mathbf{y}\|_2 \rightarrow \|\phi(\mathbf{x})\|_2 \geq \|\phi(\mathbf{y})\|_2 \quad (22)$$

2. 原点を移動させない

$$\mathbf{x} = \mathbf{0} \rightarrow \phi(\mathbf{x}) = \mathbf{0} \quad (23)$$

3. 信号の直交性を維持する

$$\mathbf{x}^H \mathbf{y} = 0 \rightarrow \phi^H(\mathbf{x}) \phi(\mathbf{y}) = 0 \quad (24)$$

条件 (21)–(23) を満たす写像 ϕ により観測信号 $\mathbf{x}(\omega, t)$ を写像し、 $\phi(\mathbf{x}(\omega, t))$ の一般化相関行列 $\mathbf{R}^{G'}(\omega)$ を得る。

$$\mathbf{R}'_g(\omega) = \Psi'(\omega) \oplus \mathbf{R}'(\omega) \quad (25)$$

$$\mathbf{R}'(\omega) = E[\phi(\mathbf{x}(\omega, t)) \phi^H(\mathbf{x}(\omega, t))] \quad (26)$$

$$\Psi'(\omega) = [\psi'_{ij}(\omega)]_{ij} \quad (i, j = 1, \dots, M') \quad (27)$$

ここで、 $\Psi'(\omega)$ は写像 GCC 法のフィルタ行列を表す。ステアリングベクトル $\mathbf{b}(\omega; \tau)$ を写像し、 $\phi(\mathbf{b}(\omega; \tau))$ と $\mathbf{R}^{G'}(\omega)$ によって、TDOA スコア $f'_g(\omega; \tau)$ を得る。

$$f'_g(\omega; \tau) \triangleq \phi^H(\mathbf{b}(\omega; \tau)) \mathbf{R}'_g(\omega) \phi(\mathbf{b}(\omega; \tau)) \quad (28)$$

$f'_g(\omega; \tau)$ を周波数平均した $\bar{f}'_g(\omega; \tau)$ の最大化により推定 TDOA $\hat{\tau}$ を得る。

$$\bar{f}'_g(\omega; \tau) \triangleq \frac{1}{|\Omega|} \sum_{\omega \in \Omega} f'_g(\omega; \tau) \quad (29)$$

$$\hat{\tau} = \arg \max_{\tau} \bar{f}'_g(\omega; \tau) \quad (30)$$

以下では、高次元写像 $\phi: \mathbb{C}^M \rightarrow \mathbb{C}^{M'}$ ($M' \geq M$) が GCC 法の推定性能に与える影響について議論する。GCC 法と写像 GCC 法の TDOA 推定は、それぞれ式 (16), (28) に基づいて行われるが、これは真の

TDOA を表現したベクトルとより線形従属性の高いステアリングベクトルを探索することと等価である。このとき、写像によりベクトルの次元数が增大すると、線形従属性の評価は厳しくなる。真の到来時間差と類似の結果を与える到来時間差の幅が狭まることから、高分解能化が達成され、推定精度が向上する。幾何的にはより簡単な解釈が与えられ、高次元写像 ϕ によるベクトルの次元数の増大が、バーチャルなマイク数の増加に相当するものだと考えられる。

4.2 偶数次モーメント分析のための写像 ϕ_d

[3] において、任意の $2d$ 次クロスモーメント分析のための写像 $\phi_d: \mathbb{C}^M \rightarrow \mathbb{C}^{M^d}$ が提案されている。

$$\phi_d(\mathbf{x}(\omega, t)) \triangleq \left[\prod_{l=1}^d \mathbf{x}_{c_{1l}}^{l \otimes}(\omega, t), \dots, \prod_{l=1}^d \mathbf{x}_{c_{M^d l}}^{l \otimes}(\omega, t) \right]^T \quad (31)$$

$$\mathbf{x}^{l \otimes} \triangleq \begin{cases} \mathbf{x} & (\text{if } l \text{ is odd}) \\ \mathbf{x}^* & (\text{if } l \text{ is even}) \end{cases} \quad (32)$$

ただし、 $[\cdot]^*$ は複素共役を表す。ここで、 c_{kl} は ϕ_d の計算に用いる \mathbf{x} の要素番号を指定するインデックスであり、 $\{1, \dots, M\}$ から d 要素選択する全 M^d 通りの重複順列を、任意の順番で行ベクトルとして並べた $M^d \times d$ インデックス行列 \mathbf{C} の (k, l) 成分として与えられる。

$$\mathbf{C} \triangleq [c_{kl}]_{kl} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 2 \\ & & \vdots & \\ M & M & \dots & M \end{bmatrix} \quad (33)$$

4.3 写像 GCC 法のフィルタ設計

ここでは、4.2 節の写像 ϕ_d を用いた写像 GCC 法のフィルタ設計について述べる。これらは、GCC 法のフィルタに準拠しており、性質も類似のため、式の紹介に留める。

- 写像 GCC-SCOT

$$\psi'_{ij}(\omega) = \frac{1}{\prod_{l=1}^d \sqrt{E[|\mathbf{x}_{il}|^{2d}] E[|\mathbf{x}_{jl}|^{2d}]}} \quad (34)$$

- 写像 GCC-PHAT

$$\psi'_{ij}(\omega) = \frac{1}{\left| E \left[\left(\prod_{l=1}^d \mathbf{x}_{il}^{\oplus} \right) \left(\prod_{l=1}^d \mathbf{x}_{jl}^{\oplus} \right)^* \right] \right|} \quad (35)$$

ただし、 $(i, j = 1, \dots, M)$ とする。

5 実験

5.1 実験条件

提案する写像 GCC 法の有効性を確認するために、TDOA 推定精度の評価実験を行った。比較のための

Table 1 Experimental condition.

素子数	2
音源数	1
部屋のサイズ	5.5 × 9 × 2.6 m
素子間隔	0.2 m
残響時間 (T_{60})	Exp. A: 0.8 sec Exp. B: 0.3, 0.5, 0.8 sec
雑音種類	拡散性雑音
SNR	5, 0, -5 dB
サンプリング周波数	16 kHz
フレーム長	1024 samples
フレームシフト長	256 samples
窓関数	ハミング窓
周波数帯域	0.2–4 kHz
音の種類	男女各 2 名
マイクとスピーカの距離	1, 1.5 m
音声長	1 s
音源到来方向	(正面:0°, 時計回り) -90°, -60°, ..., 90°

Exp. A: 実環境実験

Exp. B: シミュレーション実験

従来手法として、3章に示した GCC-SCOT, GCC-PHAT に加えて、MUSIC, 2q-MUSIC($q=3$) を用いた。また、GCC 法の周波数フィルタとして、2素子単一音源の条件では GCC-SCOT, GCC-PHAT より優れた推定精度を示す GCC-ML が知られているが、このフィルタは高雑音、高残響下では大きく性能が劣化するため、条件の該当する本実験には含めない。

提案手法の実利用での有効性と、性質の定量的な評価のため、実環境 (Exp. A) とシミュレーション (Exp. B) によって実験を行った。Exp. A はスタジオで収録したインパルス応答をクリーン音声に畳み込んだ信号を用いた。残響時間に対する頑健性を評価するために、Exp. B では Exp. A の部屋を模擬し、鏡像法により求めたインパルス応答を用いて、0.3, 0.5, 0.8 s の3つの残響時間で音声混合した。耐雑音性の評価は、観測信号に拡散性雑音 [5] を 5, 0, -5 dB の3つの SNR で重畳して行った。実験の各条件を Table 1 に示す。TDOA 推定は 2素子単一音源の条件で行い、音源は 2素子の midpoint を基準点として $\{-90^\circ, -60^\circ, \dots, 90^\circ\}$ の 30° 間隔 7 方向に配置した。推定精度の評価には、真の到来時間差と推定到来時間差の間の二乗平均平方根誤差 (RMSE) を用いた。

5.2 実験結果

提案手法と従来手法の、各条件における TDOA 推定の結果を Fig. 1-2 に示す。提案手法は耐雑音性、耐残響性ともに従来手法より優れ、高精度での TDOA 推定を達成している。写像 GCC-SCOT と写像 GCC-PHAT を比較すると、雑音が強い場合は写像 GCC-SCOT、残響が強い場合は写像 GCC-PHAT が優れた結果となる傾向があり、それぞれ GCC-SCOT, GCC-

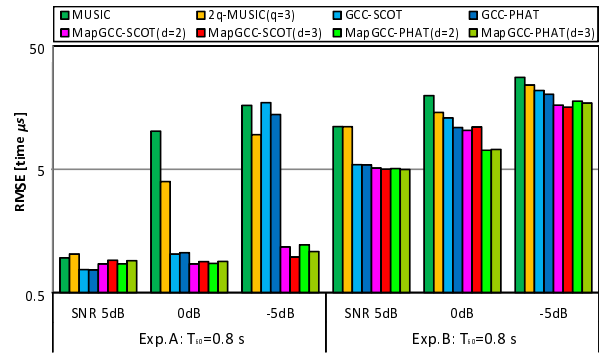


Fig. 1 The results of Exp. A and Exp. B.

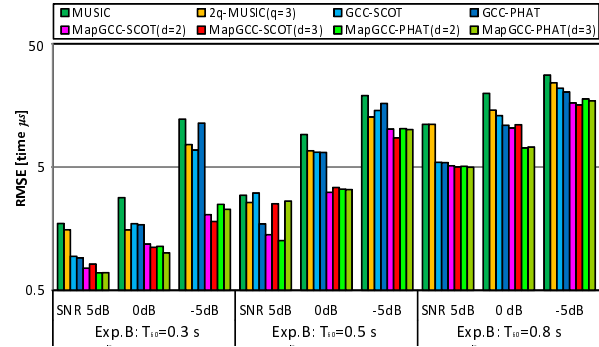


Fig. 2 The results of Exp. B.

PHAT と同様の性質を示している。

6 おわりに

本稿では、雑音・残響環境下において高精度に音源の TDOA 推定を行うため、非線形関数による観測信号の高次元写像を用いて分析を行う写像 GCC 法を提案した。実環境実験とシミュレーション実験で推定精度を評価した結果、雑音・残響下で、提案手法は MUSIC などの従来手法より高精度な TDOA 推定が可能であることを確認した。特に、雑音下では写像 GCC-SCOT、残響下では写像 GCC-PHAT により優れた推定精度が得られた。

参考文献

- [1] C. H. Knapp *et al.*, *IEEE Trans. Acoust. Speech Signal Process.*, vol.24, pp.320–327, 1976.
- [2] S. Miyabe *et al.*, *Proc. ICASSP*, pp.1641–1644, 2009.
- [3] Y. Sugimoto *et al.*, *Proc. IWAENC*, 2012.
- [4] R. O. Schmidt, *IEEE Trans. Antennas Propag.*, vol.AP-34, pp.276–280, 1986.
- [5] I. A. McCowan *et al.*, *IEEE Trans. Speech Audio Process.*, vol.11, pp.709–716, 2003.