

マイクロホンアレーの位相が観測できない条件での チャンネル間の相関係数の推定

宮部 滋樹[†] 小野 順貴^{††} 牧野 昭二[†]

[†] 筑波大学 生命領域学際研究センター

^{††} 国立情報学研究所 情報学プリンシプル研究系／総合研究大学院大学 複合科学研究科

E-mail: [†]{miyabe,maki}@tara.tsukuba.ac.jp, ^{††}onono@nii.ac.jp

あらまし 本稿では、我々が以前から提案している複素信号の相関係数を位相が観測できない条件で推定する手法を、マイクロホンアレーの音響観測信号に適用し、その有効性を調査する。位相の観測を用いない複素信号の相関分析には確率モデルを利用し、定常ガウスモデルと学生t分布モデルの2つの手法を提案しているが、定常ガウスモデルは非定常な音響信号の分析精度が低く、またt分布モデルは演算量が大きいという問題を抱えていた。提案する時変ガウスモデルは、長時間の分散の推定を短時間区間ごとの推定に置き換えるだけで容易に非定常性をモデル化することができ、演算量の増加はごくわずかである。これらの3つの手法の性質を様々なマイクロホンアレー観測信号について調査し、これらの手法は相関係数の大きさによって性能の関係が変化することと、分析フレーム長が小さい場合に時変ガウスモデルが効率的な分析を達成することを確認した。

キーワード 相関係数, 複素信号, 最尤推定, EM アルゴリズム, 非定常性

1. まえがき

アレー信号処理では2つの複素信号チャンネル間の相関を求める問題が頻繁に現れ、音声処理、無線通信や超音波センシングなど、アレー信号処理の全ての応用分野において重要な役割を果たす。例えばエコーキャンセラ [1] などに利用される適応フィルタは信号の無相関化に基づいているし、無線通信や音声強調に利用されるアレー信号処理 [2] はチャンネル間で相関のある成分を強調、抑圧あるいは検出するのが主な目的である。特に2次統計量に基づく多チャンネル信号処理ではほとんど必ず相関が利用されているといえる。

多くの信号処理は複素振幅の観測が得られたもとの分析になるが、応用によっては信号の位相が正確に観測されない場合がある。たとえば分散型センサアレー [3] のように、センサ同士が独立していて A/D 変換器が共通のクロックを参照できない場合には、チャンネル間で時刻が徐々にずれていくドリフトと呼ばれる現象が起こり、時刻のずれるレートが分析フレーム幅よりも十分小さい場合には振幅は信頼できる情報であるが、チャンネル間の位相差はわずかなドリフトによっても大きく崩れてしまう。このようなチャンネル間の位相差が正確に観測できない問題では、複素信号の相関を求めるのは困難である。

信頼できる位相の観測が得られない場合でも、本来の複素数の信号間の相関係数、つまり相関の絶対値が帯域ごとに得られれば有用な情報となることもある。例えば我々が提案した非同期マイクロホンアレーのためのドリフト補償 [4] は、チャンネル間の有相関性を復元するような時間遅延の推定に基づいており、元々の複素信号の相関の高い帯域が位相を用いず得られれば、

復元に必要な帯域のみを選んで効率的な推定をすることができ。また、位相を用いずに音響信号間の相関がわかれば、共通の場所で録られたウェブ上の動画を音声の分析によって検索するなどの応用が考えられる。

我々はこれまでに、位相の失われた2つの信号の元の複素振幅を最尤法の枠組で推定する手法に取り組んできている。まず、元の複素信号が2変量複素ガウス分布に基づくと仮定し、expectation-maximization (EM) アルゴリズムを用いた最尤推定によって相関係数を推定する手法を提案している [5]。この手法はある程度のモデルミスマッチにも頑健に動作することが確認されたが、しかし音声のような非定常な信号に対しては十分な推定精度が得られなかった。その強い非定常性や無音区間の発生、有声音の調波構造などのために、音声は長時間で定常であると捉えると優ガウス性信号としてふるまうことが知られている。このような優ガウス性信号によく適合するモデル拡張のため、優ガウス性の複素数分布である2変量複素学生t分布を用いた複素振幅モデルによる推定も提案した [6]。この拡張により、音声の狭帯域複素振幅の位相が失われた条件での相関係数推定の精度は向上が得られた。しかし、そのEMアルゴリズムでは特殊関数の1つであるガウスの合流型超幾何関数の計算を何度も行う必要があり、実装の単純さと演算量の面では前者のガウス分布モデルによる推定の方が優れている。また、優ガウ斯的な信号であっても信号の分布形状によってはガウス分布モデルの方が精度が高くなる場合もある。

音声信号は、長時間で定常と捉えた場合には優ガウ斯的な振る舞いを見せるが、5–20 ms 程度の短時間では定常信号のようにふるまうことが知られており、この局所定常状態をガウス分

布としてとらえた分析は、音声の認識や符号化で広く利用されている。そこで本稿では、音声のように分布形状が時間変化しつつも局所的に定常な信号に適したモデルとして、チャンネル間の相関係数が時不変で、各チャンネルの自己分散のみが時間変化する時変ガウス分布を仮定した相関係数推定を検討する。分散の推定を短時間で切り替える以外は定常なガウス分布モデルと同様であり、演算量を増やすことなく非定常な振幅特性に対応できると考えられる。

2. 位相を用いない相関係数推定

偏角が与えられない条件で複素変数の相関係数を推定するための、これまでに提案してきた2種類の手法について述べる。

2.1 問題設定

いま、相関のある2つの零平均複素確率変数 $x_i \in \mathbb{C}$, $i = 1, 2$ があるものとする。また、2つの確率変数はそれぞれ以下のような統計量を持つものとする。

$$E[x_i] = 0 \quad (1)$$

$$E[|x_i|^2] = \sigma_i^2 \quad (2)$$

$$E[x_1 x_2^*] = \sigma_1 \sigma_2 \rho \quad (3)$$

ここで $\{\cdot\}^*$ は複素共役を、 $E[\cdot]$ は確率変数の期待値を表し、 σ_i^2 は分散、 ρ は相互相関で、その絶対値 $0 \leq |\rho| \leq 1$ は相関係数と呼ばれる。また、 x_i の偏角 θ_i が一様分布に従うものとする。

$$p(\theta_i) = \frac{1}{2\pi}, \quad -\pi \leq \theta_i < \pi \quad (4)$$

$$\theta_i = \angle x_i \quad (5)$$

ただし複素変数 x_i , $i = 1, 2$ は観測されず、代わりにその絶対値 $y_i (> 0)$ のみが観測されるものとする。

$$y_i = |x_i| \quad (6)$$

これは偏角 θ_i が失われて観測されることに相当する。

本稿では、以上のような絶対値 y_i , $i = 1, 2$ のみが観測される条件で相関係数 $|\rho|$ を求める問題について議論する。正しい偏角を持った観測 x_1, x_2 が得られる場合には、相関係数の推定は以下のように単純になる。

$$|\rho| \leftarrow \frac{|E[x_1^* x_2]|}{\sqrt{E[|x_1|^2] E[|x_2|^2]}} \quad (7)$$

ただし x_i は確率変数 x_i の N 個の標本を表し、以下では確率変数を大文字で、標本を小文字で表すものとする。また、式(8)のように、 $E[\cdot]$ の引数が標本である場合は標本平均を表すものとする。偏角が与えられない場合には、相互相関の推定 $E[x_1^* x_2]$ を y_i , $i = 1, 2$ によって近似することの難しさのために、相関係数の推定は単純な問題ではなくなる。ここで、言うまでもなく分散 σ_i^2 , $i = 1, 2$ の推定は標本平均で求めることができる。

$$\sigma_i^2 \leftarrow E[y_i^2] \quad (8)$$

分散の推定が単純な標本平均で得られるのに対し、相互相関は一般に絶対値観測の相互相関に一致しない。

$$E[y_1 y_2] \neq \sigma_1 \sigma_2 |\rho| \quad (9)$$

そのため以下のような絶対値の積の平均は相関係数 $|\rho|$ の良い推定を与えない。

$$|\rho| \leftarrow \frac{E[y_1 y_2]}{\sigma_1 \sigma_2} \quad (10)$$

従って、複素変数の相関係数を絶対値のみから推定する問題は自明ではない。

2.2 複素ガウス分布に従う信号を仮定した推定 [5]

本節では我々が以前に提案した、複素確率変数 x_1, x_2 が式(1)–(3)を満たす単純な仮定として、零平均2変量複素ガウス分布に従うと仮定した相関係数推定について述べる。まず、本来の位相を持った複素変数 x_1, x_2 の密度関数を2変量複素ガウス分布で表すと以下ようになる。

$$p(x_1, x_2) = \frac{\exp\left(-\frac{\sigma_2^2 |x_1|^2 + \sigma_1^2 |x_2|^2 - 2\sigma_1 \sigma_2 \operatorname{Re}[\rho^* x_1 x_2^*]}{\sigma_1^2 \sigma_2^2 (1 - |\rho|^2)}\right)}{\pi^2 \sigma_1^2 \sigma_2^2 (1 - |\rho|^2)} \quad (11)$$

これを極座標変換に施すことにより、観測される絶対値 y_i と観測されない偏角 θ_i の同時密度関数が以下のように与えられる。

$$p(y_1, y_2, \theta_1, \theta_2) = \frac{y_1 y_2 \exp\left(-\frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 - 2\sigma_1 \sigma_2 |\rho| y_1 y_2 \cos(\theta_1 - \theta_2 - \angle \rho)}{\sigma_1^2 \sigma_2^2 (1 - |\rho|^2)}\right)}{\pi^2 \sigma_1^2 \sigma_2^2 (1 - |\rho|^2)} \quad (12)$$

この偏角についての周辺化によって絶対値の密度が以下のように得られる。

$$p(y_1, y_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} p(y_1, y_2, \theta_1, \theta_2) d\theta_1 d\theta_2 = \frac{4y_1 y_2 I_0\left(\frac{2|\rho| y_1 y_2}{\sigma_1 \sigma_2 (1 - |\rho|^2)}\right) \exp\left(-\frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2}{\sigma_1^2 \sigma_2^2 (1 - |\rho|^2)}\right)}{\sigma_1^2 \sigma_2^2 (1 - |\rho|^2)} \quad (13)$$

ここで $I_\nu(\cdot)$ は次数 $\nu \in \mathbb{R}$ の第1種変形ベッセル関数である。この密度関数は2変量レイリー分布として知られているものである。この最尤推定は解析的に求められないが、観測されない偏角 θ_1, θ_2 を隠れ変数として扱うことにより、繰り返し最適化で最尤推定を求めるEMアルゴリズムを定式化することができる。隠れ変数の事後密度は、式(13)の完全データの同時密度と式(12)の観測の密度の商から

$$p(\theta_1, \theta_2 | y_1, y_2) = \frac{\exp\left(\frac{2|\rho| y_1 y_2 \cos(\theta_1 - \theta_2 - \angle \rho)}{\sigma_1 \sigma_2 (1 - |\rho|^2)}\right)}{2\pi I_0\left(\frac{2|\rho| y_1 y_2}{\sigma_1 \sigma_2 (1 - |\rho|^2)}\right)} \quad (14)$$

となるため、一時的な相関係数の推定 $|\hat{\rho}|$ を用いて推定の更新 $|\rho|$ を求めるための補助関数 $Q(|\rho|; \bar{|\rho}|)$ は以下のように与えられる。

表 1 ガウス分布を仮定した相関係数推定の EM アルゴリズム

Initialization:

Set maximum iteration number N .

Estimate variances $\sigma_i^2 \leftarrow E[y_i^2]$ for $i = 1, 2$.

Set initial estimates to $|\rho|$.

E-step:

$$\lambda \leftarrow \frac{I_1\left(\frac{2|\rho|y_1y_2}{\sigma_1\sigma_2(1-|\rho|^2)}\right)}{I_0\left(\frac{2|\rho|y_1y_2}{\sigma_1\sigma_2(1-|\rho|^2)}\right)} \quad (= 1 \text{ if } \frac{2|\rho|y_1y_2}{\sigma_1\sigma_2(1-|\rho|^2)} \gg 0)$$

M-step:

$$|\rho| \leftarrow E\left[\frac{\lambda y_1 y_2}{\sigma_1 \sigma_2}\right]$$

Convergence check:

Go back to E-step until the iteration number reaches to the maximum iteratin number N .

$$\begin{aligned} Q(|\rho|; |\bar{\rho}|) &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} p(\theta_1, \theta_2 | y_1, y_2; |\bar{\rho}|) \\ &\quad \cdot \log p(y_1, y_2, \theta_1, \theta_2) d\theta_1 d\theta_2 \\ &= -\frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 - 2\sigma_1 \sigma_2 y_1 y_2 \lambda}{\sigma_1^2 \sigma_2^2 (1 - |\rho|^2)} \\ &\quad - \log(1 - |\rho|^2) + \text{const.} \end{aligned} \quad (15)$$

$$\lambda = \frac{I_1\left(\frac{2|\bar{\rho}|y_1y_2}{\sigma_1\sigma_2(1-|\bar{\rho}|^2)}\right)}{I_0\left(\frac{2|\bar{\rho}|y_1y_2}{\sigma_1\sigma_2(1-|\bar{\rho}|^2)}\right)} \quad (16)$$

ここで密度関数 $p(\cdot)$ のセミコロンの後に表される記号はパラメータを表す。ただし、上の定式化は解析的に得られる分散についての最尤推定 $\sigma_i = E[y_i^2]$ が既に推定されているものとしている。EM アルゴリズムを表 1 に示す。ただし N は EM アルゴリズムの繰り返しの最大値を表す。

2.3 複素 \mathbf{t} 分布に従う優ガウス性信号を仮定した推定 [6]

前節で示した複素信号をガウス分布でモデルする相関係数推定は、信号の分布がガウス分布からかい離している場合にはバイアスが存在してしまうため、音声のような優ガウス性信号の相関係数を推定する場合には精度が劣化してしまう。そこで、形状を表すパラメータを持つ 2 変量複素 \mathbf{t} 分布を用いて同様の信号モデルを立てることにより、音声のような優ガウス性信号とのモデルミスマッチを抑えた手法を提案している [6]。

2 変量複素 \mathbf{t} 分布の確率過程は複素ガウス分布とガンマ分布を用いて定式化される。まず、複素変数 $x_1, x_2, i = 1, 2$ の分散がもう一つの精度確率変数 $b > 0$ を用いて ξ_i^2/b で表され、共分散が $\xi_1 \xi_2 \rho/b$ となり、相関係数が $|\rho|$ で与えられる 2 変量複素ガウス分布を考える。

$$\begin{aligned} p(x_1, x_2 | b) &= \frac{b^2 \exp\left(-b \frac{\xi_2^2 |x_1|^2 + \xi_1^2 |x_2|^2 - 2\xi_1 \xi_2 \text{Re}[\rho^* x_1 x_2^*]}{\xi_1^2 \xi_2^2 (1 - |\rho|^2)}\right)}{\pi^2 \xi_1^2 \xi_2^2 (1 - |\rho|^2)} \end{aligned} \quad (17)$$

更に、 B が形状パラメータが α で期待値 1 のガンマ分布に従うものとする。

$$p(b) = \frac{\alpha^\alpha b^{\alpha-1} \exp(-\alpha b)}{\Gamma(\alpha)} \quad (18)$$

ここで $\Gamma(\cdot)$ はガンマ関数を表す。このとき x_1, x_2 と B の同時密度関数は

$$\begin{aligned} p(x_1, x_2, b) &= \frac{\alpha^\alpha b^{\alpha+1} \exp\left(-b \left(\alpha + \frac{\xi_2^2 |x_1|^2 + \xi_1^2 |x_2|^2 - 2\xi_1 \xi_2 \text{Re}[\rho^* x_1 x_2^*]}{\xi_1^2 \xi_2^2 (1 - |\rho|^2)}\right)\right)}{\pi^2 \xi_1^2 \xi_2^2 (1 - |\rho|^2) \Gamma(\alpha)} \end{aligned} \quad (19)$$

のように表され、この同時分布から b を周辺化すると 2 変量複素 \mathbf{t} 分布が得られる。

$$\begin{aligned} p(x_1, x_2) &= \int_0^\infty p(x_1, x_2 | b) p(b) db \\ &= \frac{(\alpha + 1) \left(1 + \frac{\xi_2 |x_1|^2 + \xi_1^2 |x_2|^2 - 2\xi_1 \xi_2 \text{Re}[\rho^* x_1 x_2^*]}{\alpha \xi_1^2 \xi_2^2 (1 - |\rho|^2)}\right)^{-\alpha-2}}{\pi^2 \alpha \xi_1^2 \xi_2^2 (1 - |\rho|^2)} \end{aligned} \quad (20)$$

ガウス分布モデルの場合と同様に極座標変換と偏角の周辺化の操作を施すことにより、絶対値 y_1, y_2 の密度関数が以下のように定式化できる。

$$\begin{aligned} p(y_1, y_2) &= \frac{4(\alpha + 1)}{\alpha \xi_1^2 \xi_2^2 (1 - |\rho|^2)} y_1 y_2 \\ &\quad \cdot \frac{{}_2F_1\left(\alpha + 2, \frac{1}{2}; 1; \frac{4\xi_1 \xi_2 |\rho| y_1 y_2}{\alpha \xi_1^2 \xi_2^2 (1 - |\rho|^2) + \xi_2^2 y_1^2 + \xi_1^2 y_2^2 + 2\xi_1 \xi_2 |\rho| y_1 y_2}\right)}{\left(1 + \frac{\xi_2^2 y_1^2 + \xi_1^2 y_2^2 + 2\xi_1 \xi_2 |\rho| y_1 y_2}{\alpha \xi_1^2 \xi_2^2 (1 - |\rho|^2)}\right)^{\alpha+2}} \end{aligned} \quad (21)$$

従って、精度 b と偏角 θ_1, θ_2 を隠れ変数とすることにより EM アルゴリズムを定式化することができる。本稿ではの導出は割愛するが、表 2 に EM アルゴリズムを示す。各イタレーションでサンプルごとに 5 種類の異なったパラメータの超幾何関数評価を行う必要があることがわかる。ガウスの合流型超幾何関数は数値評価のための演算コストが高いため、ガウス分布モデルよりも遥かに演算量が大きくなってしまふ。

2.4 時変ガウス分布により非定常性のモデル化した推定

述べたスチューデント t 分布の隠れ変数である、精度 b が表しているのは、サンプルごとの零平均ガウス分布の分散が σ_i^2/b となり、相関が変化しないままで分散が時間変化する時変ガウス分布によって、信号の非定常性をモデル化していると捉えることができる。ここで、音声のような緩やかな特性の時間変化を伴う信号では、分析する信号全体を定常とみなして分散を求めるのではなく、隣接する数フレームは局所的に定常であるとみなし、隣接フレームごとに個別に分散の推定を行えば、時変の分散を小さな誤差で推定することができる。そこで本稿では、時変の分散推定を用いて非定常性をモデル化した相関推定を試みる。

時変分散の推定法にはいくつもの方法があり、フレームの系列を単純にブロックに切り分ける方法、分析対象のフレームを中心に緩やかに両端が減衰する窓で重みづけて推定する方法や、忘却係数を用いる方法等が考えられる。本稿では計算効率を重視し、ブロックごとに分散を推定する方法を採用する。ブロッ

表 2 スチューデント t 分布を仮定した相関係数推定の EM アルゴリズム

Initialization:

Set tuning parameters ϵ ($0 < \epsilon \ll 1$), M and N .

Set initial estimates to $\xi_1^2, \xi_2^2, |\rho|$ and α .

E-step:

$$A \leftarrow \frac{4\xi_1\xi_2|\rho|y_1y_2}{\alpha\xi_1^2\xi_2^2(1-|\rho|^2) + \xi_2^2y_1^2 + \xi_1^2y_2^2 + 2\xi_1\xi_2|\rho|y_1y_2}$$

$$B \leftarrow \frac{\xi_1^2\xi_2^2(1-|\rho|^2)}{2\xi_1\xi_2|\rho|y_1y_2}$$

$$C \leftarrow \frac{\alpha\xi_1^2\xi_2^2(1-|\rho|^2) + \xi_2^2y_1^2 + \xi_1^2y_2^2}{\alpha(\alpha+2)}$$

$$\lambda \leftarrow \frac{2(\alpha+B)(\alpha-B)_2F_1(\alpha+2, \frac{1}{2}; 1; A)}{\cdot(3(C+1)_2F_1(\alpha+3, \frac{1}{2}; 1; A) + ((4\alpha+9)C-3)_2F_1(\alpha+2, \frac{1}{2}; 1; A))}$$

$$\mu \leftarrow \frac{(\alpha+2)_2F_1(\alpha+3, \frac{1}{2}; 1; A)}{(\alpha+B)_2F_1(\alpha+2, \frac{1}{2}; 1; A)}$$

$$\nu \leftarrow \mu - \frac{\alpha+2}{\alpha+B} - \log(\alpha+B) + \Psi(\alpha+2) - \frac{2(\alpha+2)|\rho|y_1y_2}{\xi_1\xi_2(1-|\rho|^2)(\alpha+B)^2} \frac{{}_2F_1(\alpha+3, \frac{3}{2}; 2; A)}{{}_2F_1(\alpha+2, \frac{1}{2}; 2; A)} + \frac{1}{\epsilon} \frac{{}_2F_1(\alpha+\epsilon+2, \frac{1}{2}; 1; A)}{{}_2F_1(\alpha+2, \frac{1}{2}; 1; A)} - \frac{1}{\epsilon}$$

M-step:

$$\xi_1^2 \leftarrow E[\mu y_1^2]$$

$$\xi_2^2 \leftarrow E[\mu y_2^2]$$

$$|\rho| \leftarrow E\left[\frac{\lambda y_1 y_2}{\xi_1 \xi_2}\right]$$

$$\alpha \leftarrow \frac{2(E[\mu] - E[\nu] - 1)}{2(E[\mu] - E[\nu] - 1)}$$

M iterations of generalized Newton update:

$$\alpha \leftarrow \frac{\alpha - \alpha^2 \Psi'(\alpha)}{\log \alpha - \Psi(\alpha) - \alpha \Psi'(\alpha) + E[\nu] - E[\mu] + 2}$$

Convergence check:

Go back to E-step until the iteration number reaches to the maximum iteratin number N .

クごとの分散分析によってサンプルごとの分散 σ_1^2, σ_2^2 が得られたもとでは、これをそのまま表 1 の EM アルゴリズムに代入することにより、ガウスモデルを時変ガウスモデルに拡張することができる。

3. 評価実験

3.1 実験条件

オーディオ信号に対する相関係数推定の挙動を調査するため、マイクロホンアレーで観測した音響信号の狭帯域複素信号の相関係数を、位相が与えられない条件で推定する精度を評価する。観測信号の分布は未知で真の相関は得られないため、複素数の観測が与えられた条件での理想的な相関係数の推定を正解として、振幅のみからの評価の推定誤差を平均二乗誤差平方根 (RMSE) を基準として評価した。評価対象は、式 (10) の振幅相関を用いた推定 (Baseline), 定常性を仮定したガウスモデルに基づく推定 (Gaussian), 新たに提案した非定常性をモデル化する時変ガウスモデルに基づく推定 (Time varying Gaussian/TVG) と、スチューデント t 分布に基づく推定 (Student-t) である。また、これらの推定手法の計算時間

を計測した。

分析する音響信号は、SiSEC データベース [9] の test および test2 データセットの合計 50 種類の全ての混合音を分析した。これらのデータセットは複数の条件の 2 チャンネル音源分離のベンチマークデータとなる録音を集めたものであり、瞬時混合から残響時間数 100 ms までの様々な残響の影響を受けた、音声および音楽演奏の複素音源混合を収録している。観測信号帳はそれぞれ 10 秒で、複素信号の相関係数の推定は短時間フーリエ分析の帯域ごとに評価した。観測信号のサンプリング周波数は 16 kHz, 窓幅は 256 サンプルと 1024 サンプルの 2 種類で、フレームシフトは 1/4 フレームとした。また、時変ガウスモデルの時変分散を推定するためのブロック長は、2 種類のフレーム長の両方で良好な性能が得られた 7 フレームに固定した。また EM アルゴリズムのイタレーションは全て 20 回で打ち切った。

3.2 推定精度の評価

図 1, 2 に、それぞれフレーム長が 1024 の場合と 256 の場合の残響時間の違いでまとめた RMSE と平均的な相関係数の大きさの比較を示す。ここで、平均的な相関係数は、各相関係数の推定の 2 乗平均 (root mean coherence) によって評価している。また図 3, 4 に、それぞれフレーム長が 1024 の場合と 256 の場合の信号減の種類でまとめた RMSE と平均的な相関係数の大きさの比較を示す。更に、それぞれフレーム長が 1024 の場合と 256 の場合の分析結果の例を図 5, 6 に、それぞれの分析例の RMSE を表 3, 4 に示す。そして演算量の比較を図 7 に示す。

全ての図に共通した各手法の推定精度の傾向として、スチューデント t 分布モデル以外の推定法は相関係数を正解よりも高めに推定する傾向があり、結果として相関が高い条件において推定精度が高くなる。反対に、スチューデント t 分布モデルに基づく推定法は相関係数を正解よりも低めに推定する傾向があり、相関が低い条件において推定精度が高くなる。信号の性質は音源の特性や伝達系の特性に影響されるが、多くの場合に単純に相関係数の高さが性能の優劣に影響を及ぼしている。最も多くの条件で高い推定精度を示したのはスチューデント t 分布であった。しかし、図 3, 4, 5, 6 から読み取れるように、残響時間が 0 の瞬時混合の条件では少し違う傾向を示している。図 5, 6 は顕著な例で、全ての推定法が破綻しているのがわかる。これは、瞬時混合は人工的に合成した混合であるため、統計的な分析を破綻させるような何らかの例外を引き起こすことがあるためと考えられる。瞬時混合のいくつかの例において複数の手法の破綻が観察された。

本論文で新たに検討した時変ガウスモデルについて考察する。音声の局所定常性の当てはめは 200 ms 程度までが限界といわれており、フレーム長 1024 サンプルの条件では大きな改善が得られていない。図 4 ではどの音源でも悪化という結果になっているが、これは瞬時混合の条件での分析の破たんの影響を受けたものであり、全体としては図 2 のように瞬時混合の条件を除いて定常ガウスモデルよりも少し改善する傾向にある。しかしフレーム長が 256 サンプルの場合には定常モデルからの顕著な改善がみられ、フレーム長が短い場合において有効な手法で

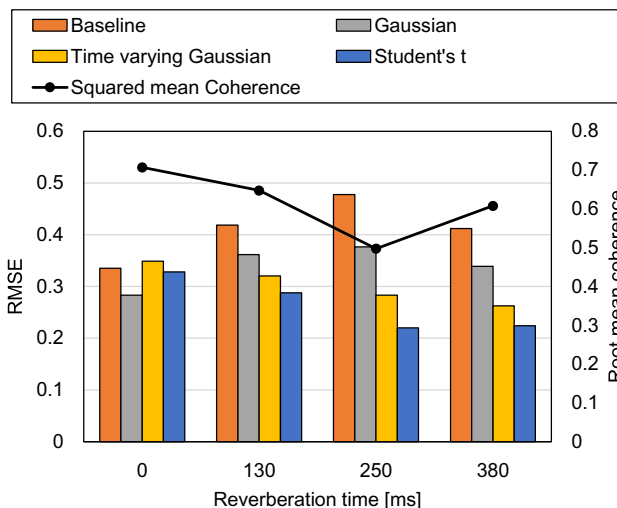


図 1 フレーム長 256 サンプルの分析における残響時間の違いによる相関係数の大きさと推定精度の比較。

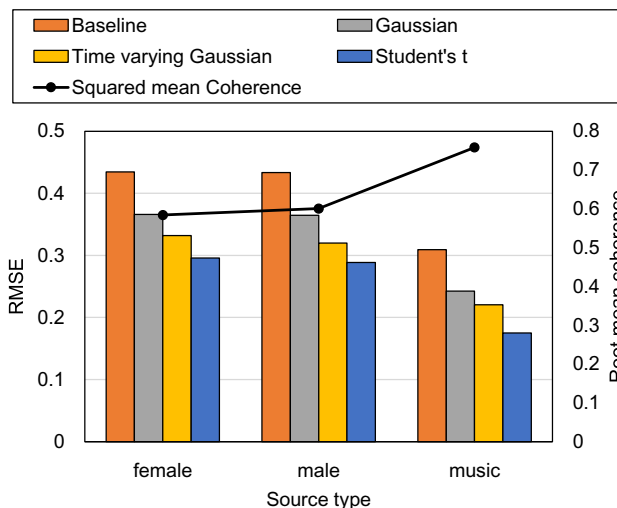


図 3 フレーム長 256 サンプルの分析における音源の種類による相関係数の大きさと推定精度の比較。

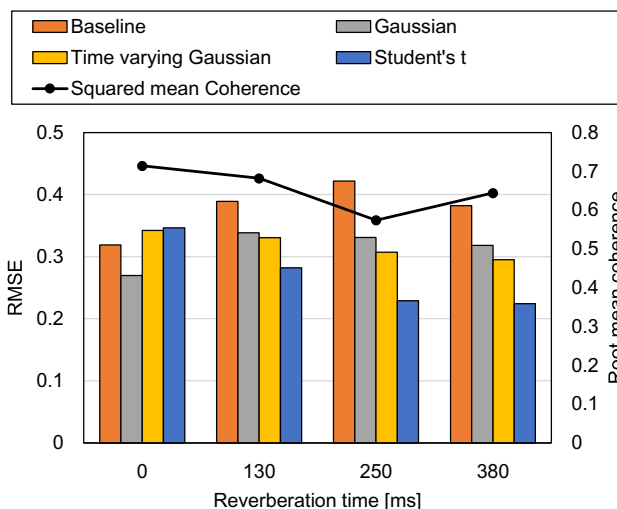


図 2 フレーム長 1024 サンプルの分析における残響時間の違いによる相関係数の大きさと推定精度の比較。

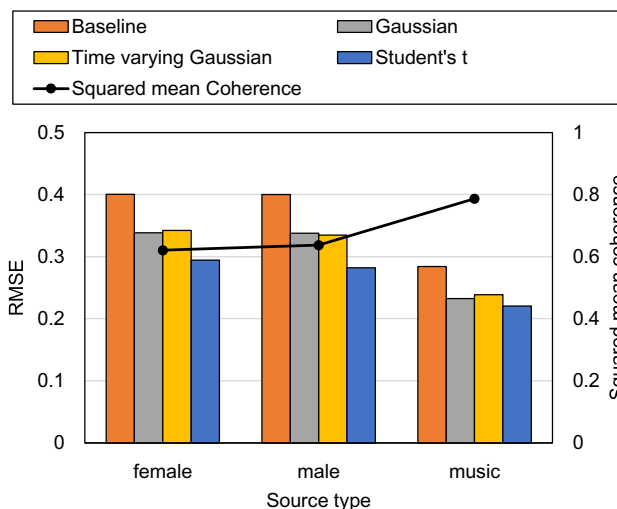


図 4 フレーム長 1024 サンプルの分析における音源の種類による相関係数の大きさと推定精度の比較。

あることがわかる。

最後に計算時間について議論する。スチューデント t 分布モデルの演算量はガウスモデルに比べて演算量のはるかに大きく、ラップトップ PC で観測時間の 10 倍程度の計算時間を要することがわかる。また、標準偏差が平均計算時間と同程度と非常に大きい。これは特殊関数の評価に掛かる計算時間が引数の大きさによって大きく変化するためである。時変ガウスモデルの定常ガウスモデルからの演算量の増加はわずかであり、フレーム数が小さい条件では効果的な改善手法であるといえる。

4. あとがき

本稿では、位相が失われた観測から複素振幅の相関係数を音響信号に適用した場合の性質について調査した。また、時変ガウスモデルによって簡易な計算で信号の非定常性に追従する推定法を新たに提案した。定常ガウスモデル、非定常ガウスモデルとスチューデント t 分布モデルはそれぞれ挙動が少しずつ違

表 3 RMSEs of the estimation example in Fig. 5

Rev. time	Baseline	Gaussian	TVG	Student- t
0 ms	0.54	0.53	0.52	0.36
130 ms	0.29	0.25	0.20	0.21
380 ms	0.37	0.24	0.19	0.17

表 4 RMSEs of the estimation example in Fig. 6

Rev. time	Baseline	Gaussian	TVG	Student- t
Inst.	0.51	0.48	0.52	0.51
130 ms	0.26	0.23	0.21	0.19
380 ms	0.35	0.26	0.24	0.20

い、瞬時混合で作成された信号のような例外を除くと、ほとんどの場合にスチューデント t 分布モデルが最も高い精度を示した。しかし計算量が非常に大きいことを確認した。新たに提案した非定常ガウスモデルは、信号の特性が変化するよりも十分小さいフレームレートの分析では定常モデルよりも高い精度を

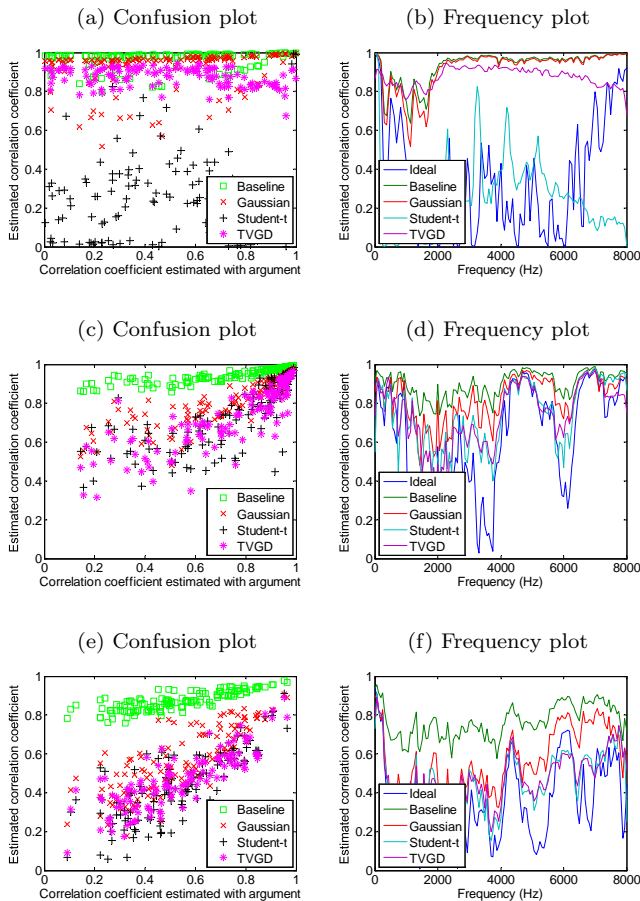


図 5 素子間隔 20 cm のマイクロホンアレーで観測された複数の残響条件におけるドラムを含んだ楽器演奏をフレーム長 256 サンプルの分析例。(a), (b) は瞬時混合, (c), (d) は 130 ms, (e), (f) は 380 ms の残響環境の分析を示す。

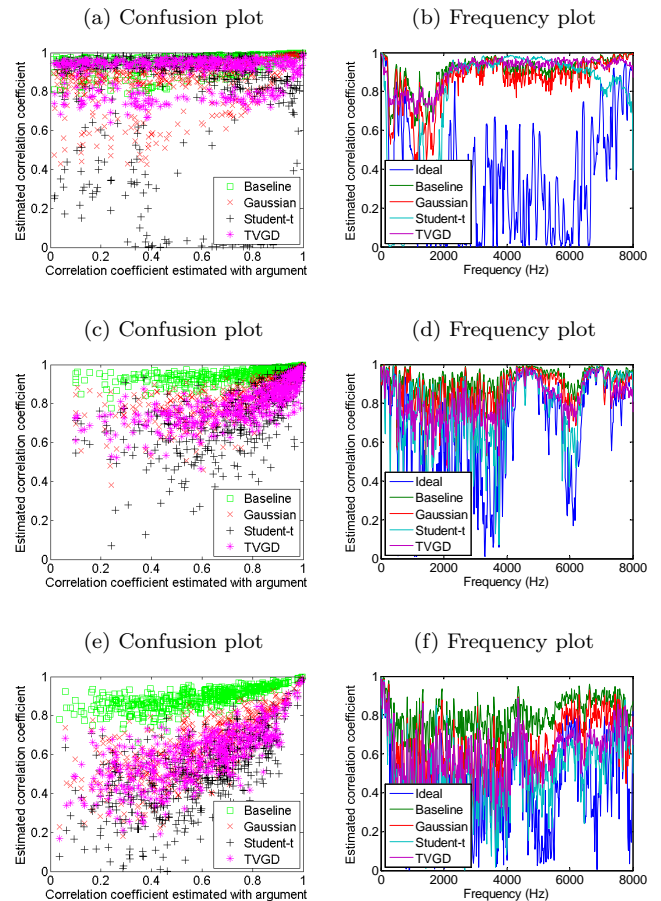


図 6 素子間隔 20 cm のマイクロホンアレーで観測された複数の残響条件におけるドラムを含んだ楽器演奏をフレーム長 1024 サンプルの分析例。(a), (b) は瞬時混合, (c), (d) は 130 ms, (e), (f) は 380 ms の残響環境の分析を示す。

示し、演算量の増加もわずかであるため、場面によっては効果的な手法であることを確認した。

文献

- [1] J. Benesty et al., Advances in network and acoustic echo cancellation, Springer, 2001.
- [2] D. Johnson and D. Dudgeon, Array signal processing: concepts and techniques, Simon & Schuster, 1992.
- [3] A. Bertrand, “Applications and trends in wireless acoustic sensor networks: a signal processing perspective,” Proc. SCVT, 2011.
- [4] S. Miyabe, N. Ono, S. Makino, “Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation,” Signal Processing, vol.107, pp.185–196, 2015.
- [5] 宮部, 小野, 牧野, “絶対値の観測のみを用いた 2 つの複素信号の相関係数推定,” 音講論 (春), pp.735–738, 2014.
- [6] 宮部, 小野, 牧野, “2 つの優ガウス性複素信号の位相観測を用いない相関係数推定,” 信学技法, vol.EA2014-75, no.473, pp.19–24, 2014.
- [7] F. J. Lopez-Martinez et al., “On the bivariate Nakagami cumulative distribution function: closed-form expression and applications,” IEEE Trans. Commun., vol.61, no.4, pp.1404–1414, 2013.

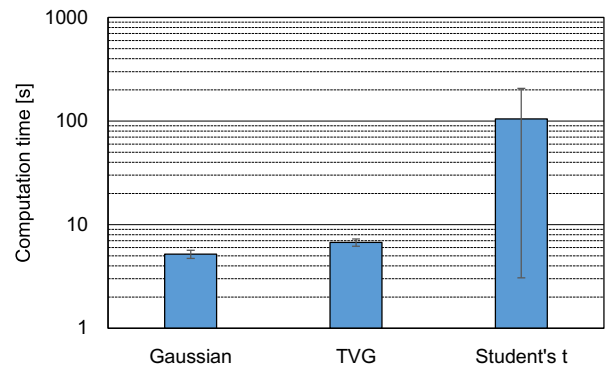


図 7 Windows のラップトップ PC による計算時間の比較。エラーバーは標準偏差を表す。

- [8] M. Novey, T. Adali and A. Roy, “A complex generalized Gaussian distribution—characterization, generation, and estimation,” IEEE Trans. Signal Process., vol.58, no.3, pp.1427–1433, 2010.
- [9] N. Ono et al. “The 2013 Signal Separation Evaluation Campaign,” Proc. MLSP, 6 pages, 2013.