

高次統計量の不動点に着目した統計推定及びその音響センシングへの応用

代表研究者 猿渡 洋
代表分担者 小山翔一

東京大学・大学院情報理工学系研究科 教授
東京大学・大学院情報理工学系研究科 助教

1 研究調査の要旨

本研究では、高次統計量の不動点に着目した新しい反復型統計推定理論及びその音響センシングへの応用に関して検討を行う。特に、弱いベイズ推定を反復適用する手法を確立し、高精度な音響・音声信号強調システムの構築を行う。一般的に、高次統計量は「聴覚品質(ミュージカルノイズの程度)」を表し、先行研究により高次統計量に不動点が存在する場合があることが明らかになっている。本不動点に対応する推定を「弱い推定器」と見なし、それを繰り返すことで最終的に推定誤差が小さくかつミュージカルノイズが少ない推定を実現することができる。現在までに、最尤推定に関してはその不動点の存在が確認されているが、MAP推定やベイズ推定においては未だにその存在が明らかでなく、更なる検討を行う必要があった。本研究では第一に、通常の複素ガウス分布(振幅スペクトルがレイリー分布に従うと仮定)を目的音事前モデルとするベイズ推定に関しては高次統計量の不動点は存在しないため、振幅スペクトルをカイ分布にてモデリングする一般化された事前モデルを導入し、その場合における高次統計量の不動点を解析した。ここでは、高次統計量の不動点を生じさせるパラメータとして、事前 SN 比の推定ブロックにおいて若干の劣化(人工バイアス)を加えることとした。このバイアス値(> 0)を増加させ、出力のカートシス比の値が 1 となる点を探索した。第二に、上記解析を容易にするため、高次キュムラント領域における加法性を經由して高次モーメントの計算を行う手法を新たに導出した。最後に、提案する反復型バイアス付きベイズ推定器を用いて、従来の統計的音声強調法と音質評価実験を行った。本結果より、提案推定器の品質が最も良いことが分かった。

2 研究背景と目的

近年、ビッグデータに代表される超多次元データ情報処理が盛んに研究されている。これまではテキスト系の構造化データが中心であったが、これからは非テキスト・非構造化データが爆発的に増えると予想される(IDC レポートによると今後 9 割以上が非構造化データになる見込み)。このような状況を鑑みると、実計測信号からの統計的信号推定は、データクレンジング処理としてますますその重要性を増している。本著者らも、人間の対話インタラクションや周りの音響状況を多チャンネル音響センサにて収録し、その中から重要なイベントを分離解析し、「場の雰囲気・盛り上がり」を定量的に測る技術の研究開発を行っている[1](図 1 参照)。本研究も、この音響データの分離解析技術に関するものである。

このような統計的信号処理は、従来のマンマシンインターフェイス応用と異なり、リアルタイム性が要求されないという利点を持つ。つまり、データ取得後に大規模な計算コストと時間をかけてクレンジング・解析することが出来る。本研究では、この特徴を生かした新しい統計的信号解析(雑音除去、目的音推定等)理論を構築することを主目的とする。特に、本申請者が昨年度までに提案した「弱い統計推定器を繰り返し適用する」聴感印象不動理論に着目し、その一般化を図ることによって目的を達成する。

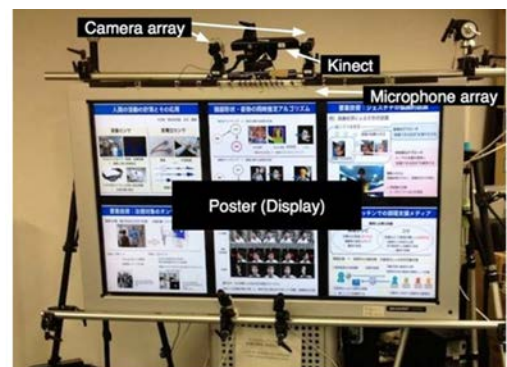


図1. 場の雰囲気を察知するセンサシステム

3 高次統計量の不動点に着目した音響信号統計推定法の提案

2-1 提案法の着想及び動機

本提案テーマ発想の動機となったのは、本著者によって提唱された高次統計量追跡による信号処理システム解析である[2, 3]。ここでは、4 次の統計量に基づくカートシスの値の変化が、人間の聴覚的印象と強い相関があることを明らかにした。端的に言うと、4 次統計量を大きく変化させるような信号処理は、概して瞬

間的に音のトーン性（音程感）を付与することに等しく、また過度にトーン性を強調するような信号処理はしばしば自然界の聞きなれた音を人工的な音に変えてしまい、結果として人が聞いた時に不快感や了解性の低下を引き起こす。そこで我々は、「さまざまな統計推定を行った際に信号の高次統計量がどのように変形を受けたか」を追跡すれば、世の中全ての音響信号に対する処理結果を予測・評価できるという理論を構築した。因みに、高次統計量が「質」を表すのに対し、従来の低次（1次）統計量の変化は統計推定誤差量（つまり「量」）を表すことに相当し、一般に両者はトレードオフの関係にあるとされてきた。

上記に鑑み、図1に示した音響センサシステムの聴感品質を制御するため、高次統計量の不動点に着目した統計推定の研究を行う必要性が生じた。本不動点の存在は聴覚印象を全く変化させずに特定音源を抽出できることを示しており、本著者によって世界で初めて見出されたものである[4, 5]（図2参照）。本不動点に対応する推定を「弱い推定器」と見なし、それを繰り返すことで最終的に推定誤差が小さくかつ聴覚印象劣化が少ない推定を実現することができる。現在までに、最尤推定に関してはその存在が確認されているが、MAP推定やベイズ推定においては未だにその存在が明らかでなく、更なる検討を行う必要がある。

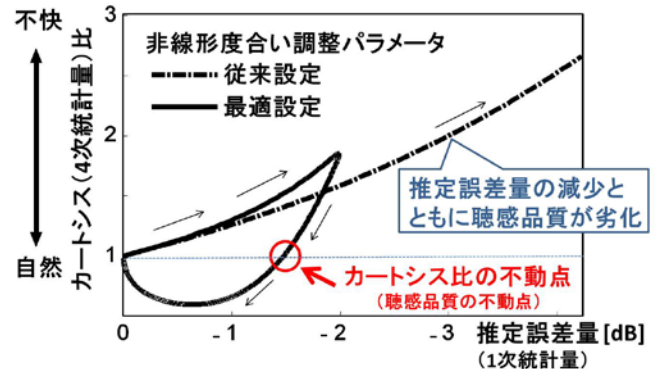


図2. 最尤推定における高次統計量不動点の存在

2-2 提案システム概要

本研究にて想定される「高次統計量不動点を利用した反復型の統計推定」システムの概要について説明する。ここでは、以下2つの条件を満たす(単体では推定誤差の多い「弱い」)統計推定器を見つけることが課題となる。

【(A) カートシス比[6]不動条件】 推定誤差(残留不要信号)や抽出対象信号の高次統計量が推定の前後で変化しない。つまり人間の聴感印象が変わらないという条件を課す。ここで「カートシス」とは、一般に確率信号の4次モーメントと二乗2次モーメントとの比で与えられる無次元数である。

【(B) 推定誤差減少条件】 条件(A)を満たす無意味な解として、「何もしない」という状態が存在する。これを避けるため、最低でも推定誤差の二乗平均値が処理前後で減少するという条件を課す。

ある統計推定器において上記(A)及び(B)を満たす不動点パラメータが存在するならば、その推定を繰り返し行うことで、高次統計量を変化させずに低次統計量(推定誤差平均値)のみを所望の値まで下げることが可能である。本研究では、本理論の一般化を目指し、目的抽出信号の事前仮説モデルを導入したベイズ推定[7, 8]に関して、その不動点の存在および反復型推定の可能性について検討を行った。

2-3 高次統計量不動点を持つベイズ推定器の設計

本著者らの過去の研究において、通常の複素ガウス分布(振幅スペクトルがレイリー分布に従うと仮定)を目的音事前モデルとするベイズ推定(MMSE-STSA estimator [7])に関しては高次統計量の不動点は存在しないことが明らかになっている[9]。そこで、本申請研究では、振幅スペクトルをカイ分布にてモデリングする一般化された事前モデルを導入し、その場合における高次統計量の不動点を解析した。ここでは、高次統計量の不動点を生じさせるパラメータとして、事前SN比の推定ブロックにおいて若干の劣化(人工バイアス)を加えることとした。このバイアス値(> 0)を増加させ、出力のカートシス比の値が1となる点を探索した。

2-4 高次統計量不動点に対応するパラメータの高速計算法

ベイズ推定器に関する高次統計量の解析においては、再帰型非線形システムの出力における高次モーメントを計算する必要がある。しかし、最尤推定の場合と異なり、高次モーメントを推定器のパラメータに関する解析的な関数で表すことは不可能であることより、何らかの工夫が必要である。そこで、Breithauptらによって導入された非再帰型(FIR型)非線形システム近似[8]を導入し、高次キウムラント領域における加法

性を経由して高次モーメントの計算を行う手法を導出した（導出の詳細は<外部発表資料>を参照のこと）。

以下に計算の概略を示す。まず、本過程において、入力雑音のパワースペクトルは以下のガンマ分布に従うものと仮定する。ここで η は雑音の性質を表す形状母数であり、小さいほど優ガウス性の雑音であったことに相当する。また θ はスケール母数であるが、本研究では任意の値を定めることが出来る。

$$p_{\Gamma}(x) = x^{\eta-1} \exp(-x/\theta) \theta^{-\eta} \Gamma(\eta)^{-1}$$

次に、目的とする音声の振幅スペクトルの事前分布を以下のカイ分布で表す。ここで、 ρ は音声のスパース性を表す形状母数であり、小さいほどスパースな分布に対応する。 ϕ はスケール母数である。本研究では、この ρ をハイパーパラメータとして用意し、ベイズ型統計推定器の出力の性質を求める。

$$p_{\chi}(x) = 2\phi^{\rho} \Gamma(\rho)^{-1} x^{2\rho-1} \exp(-\phi x^2)$$

この事前分布の元で、観測信号を以下で表す。

$$X(f, \tau) = S(f, \tau) + N(f, \tau)$$

ここで S は目的音声信号の時間周波数領域表現（ f は周波数・ τ は時間インデックスを表す）、 N は雑音を表す。

ここで、事前仮説を導入したベイズ推定器の出力 Y は以下で与えられる。

$$Y(f, \tau) = G(f, \tau) X(f, \tau)$$

$$G(f, \tau) = \frac{\sqrt{\nu(f, \tau)}}{\gamma(f, \tau)} \left[\frac{\Gamma(\rho + \beta/2)}{\Gamma(\rho)} \cdot \frac{\Phi(1 - \rho - \beta/2, 1, -\nu(f, \tau))}{\Phi(1 - \rho, 1, -\nu(f, \tau))} \right]^{\frac{1}{\beta}}$$

$$\nu(\tau) = \xi(f, \tau) (1 + \xi(f, \tau))^{-1} \gamma(f, \tau)$$

上記の G はゲイン関数であり、 Φ は合流型超幾何関数である。これは事前 SN 比 ξ と事後 SN 比 γ の関数となっている。また Γ は完全ガンマ関数、 β は信号推定を行うドメインの圧縮パラメータである。

本研究では、前記の事前 SN 比 ξ に関して、以下のようなバイアス ε を付加した形式で逐次推定を行う。これは意図的に推定信号の低次統計量の変動を抑えると同時に、高次統計量の変化も抑える狙いがある。

$$\begin{aligned} \hat{\xi}_{\text{bias}}(f, \tau) &= \alpha \hat{\gamma}(f, \tau - 1) G^2(f, \tau - 1) \\ &\quad + (1 - \alpha) \max[\hat{\gamma}(f, \tau) - 1, \varepsilon] \end{aligned}$$

ここで α は逐次推定における忘却係数である。本来、このバイアス ε に関する推定出力 Y の高次統計量を計算する必要があるが、 G の挙動を解析的に書くことが困難であることより、本研究では Breithaupt らによって導入された非再帰型（FIR 型）非線形システム近似を行う。本近似を用いて推定出力のパワー（二乗）系列を求めると以下ようになる。

$$|Y_{\text{bias}}(f, \tau)|^2 \simeq h_{\xi}(\tau) * \hat{\gamma}_{\text{bias}}(f, \tau) \cdot R(\rho, \beta) \cdot P_{\hat{N}}(f) \quad \dots (1)$$

ここでは

$$\hat{\gamma}_{\text{bias}}(f, \tau) = \max[\hat{\gamma}(f, \tau) - 1, \varepsilon],$$

$$h_{\xi}(f, \tau) = (1 - \alpha) \exp(-\lambda_{\xi} \tau),$$

$$\lambda_{\xi} = \ln(2\alpha^{-1} R(\rho, \beta)^{-1}).$$

$$R(\rho, \beta) = \begin{cases} \rho^{-1} \left[\Gamma(\rho + \beta/2) / \Gamma(\rho) \right]^{2/\beta}, & \text{if } \beta \neq 0, \\ \rho^{-1} \exp(-c), & \text{if } \beta = 0 \text{ and } \rho = 1 \end{cases}$$

であり、 $*$ は畳み込み演算を表す。

上記式 (1) で与えられた近似推定出力パワー系列 $|Y_{\text{bias}}|^2$ の高次モーメントを解析的かつ閉形式で求める。まず、式 (1) が畳み込み演算を含む形式であることより、加法確率信号の高次モーメントを求めねばならないが、これは非常に困難であることが知られている。そこで、本研究では高次キュムラントを経由することで、本問題を解決する。一般に、 m 次キュムラントは以下の加法性（厳密には m 次斉次性）を有する。

$$\kappa_m(\sum_i a_i x_i) = \sum_i a_i^m \kappa_m(x_i) \quad (\kappa_m(\cdot) \text{ is the } m\text{-th-order cumulant})$$

この関係を用いることにより、式(1)の m 次キュムラントは以下で与えられる。

$$\kappa_m(|Y_{\text{bias}}(f, \tau)|^2) = \frac{\exp(-m\lambda_\xi)}{1 - \exp(-m\lambda_\xi)} \{(1 - \alpha)R(\rho, \beta)\eta\theta\}^m \cdot \kappa_m(\hat{\gamma}_{\text{bias}}(f, \tau)) \quad \dots (2)$$

ここで、上式右辺の最終項に現れるバイアス付き事後 SN 比 γ_{bias} の m 次キュムラントを導出する。まず、入力雑音パワースペクトルがガンマ分布に従うという仮定より、バイアス付き事後 SN 比 γ_{bias} は以下の確率密度分布を有する。

$$p_{[\hat{\gamma}_{\text{bias}}]}(x) = \begin{cases} 0 & (x < \varepsilon) \\ \int_{-1}^{\varepsilon} p_{[\hat{\gamma}-1]}(x)dx & (x = \varepsilon) \\ \frac{(x+1)^{\eta-1} \exp(-\eta(x+1))}{(1/\eta)^\eta \Gamma(\eta)} & (x > \varepsilon) \end{cases}$$

よって、上式に基づいてその m 次モーメントを計算することができ、その値は以下で与えられる。

$$\begin{aligned} \mu_m(\hat{\gamma}_{\text{bias}}(f, \tau)) &= \int_{-\infty}^{\infty} x^m p_{[\hat{\gamma}_{\text{bias}}]}(x) dx \\ &= \int_{\varepsilon}^{\infty} x^m \frac{(x+1)^{\eta-1} \exp(-\eta(x+1))}{(1/\eta)^\eta \Gamma(\eta)} dx \\ &\quad + \varepsilon^m \frac{1}{\Gamma(\eta)} \Gamma_L(\eta, \eta(\varepsilon+1)), \end{aligned}$$

この式の右辺第一項は積分を含む形なのでこのままでは高速に計算することが出来ないが、以下のような展開形式で書くことができる。

$$\begin{aligned} &\int_{\varepsilon}^{\infty} x^m \frac{(x+1)^{\eta-1} \exp(-\eta(x+1))}{(1/\eta)^\eta \Gamma(\eta)} dx \\ &= \int_{\eta(\varepsilon+1)}^{\infty} \frac{1}{\eta^m \Gamma(\eta)} (t - \eta)^m t^{\eta-1} \exp(-t) dt \\ &= \frac{1}{\eta^m \Gamma(\eta)} \sum_{l=0}^m (-\eta)^l \frac{\Gamma(m+1)}{\Gamma(l+1) \Gamma(m-l+1)} \\ &\quad \cdot \Gamma_U(\eta+m-l, \eta(\varepsilon+1)), \end{aligned}$$

ここで、 Γ_U は第一種不完全ガンマ関数である。以上より、バイアス付き事後 SN 比 γ_{bias} のモーメントを閉形式で算出することが出来た。このモーメントに基づき、それらをキュムラントへ変換すると、以下のような高次キュムラントを得ることが出来る。

$$\left\{ \begin{aligned} \kappa_1(\hat{\gamma}_{\text{bias}}(f, \tau)) &= \mu_1(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ \kappa_2(\hat{\gamma}_{\text{bias}}(f, \tau)) &= \mu_2(\hat{\gamma}_{\text{bias}}(f, \tau)) - \mu_1^2(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ \kappa_3(\hat{\gamma}_{\text{bias}}(f, \tau)) &= \mu_3(\hat{\gamma}_{\text{bias}}(f, \tau)) + 2\mu_1^3(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ &\quad - 3\mu_2(\hat{\gamma}_{\text{bias}}(f, \tau))\mu_1(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ \kappa_4(\hat{\gamma}_{\text{bias}}(f, \tau)) &= \mu_4(\hat{\gamma}_{\text{bias}}(f, \tau)) - 3\mu_2^2(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ &\quad - 4\mu_3(\hat{\gamma}_{\text{bias}}(f, \tau))\mu_1(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ &\quad + 12\mu_2(\hat{\gamma}_{\text{bias}}(f, \tau))\mu_1^2(\hat{\gamma}_{\text{bias}}(f, \tau)) \\ &\quad - 6\mu_1^4(\hat{\gamma}_{\text{bias}}(f, \tau)) \end{aligned} \right.$$

これらの値を式(2)の右辺最終項に代入することにより、近似推定出力パワー系列 $|Y_{\text{bias}}|^2$ の高次キュムラントが計算出来る。更に、その結果をモーメントへ変換することにより、以下のような値で近似推定出力パワー系列 $|Y_{\text{bias}}|^2$ の高次モーメントを得ることが出来る。これらの値は、全て積分を含まない形で導出されているため、非常に高速に計算することが出来る。

$$\begin{cases} \mu_1(|Y_{\text{bias}}(f, \tau)|^2) = \kappa_1(|Y_{\text{bias}}(f, \tau)|^2) \\ \mu_2(|Y_{\text{bias}}(f, \tau)|^2) = \kappa_2(|Y_{\text{bias}}(f, \tau)|^2) + \kappa_1^2(|Y_{\text{bias}}(f, \tau)|^2) \\ \mu_4(|Y_{\text{bias}}(f, \tau)|^2) = \kappa_4(|Y_{\text{bias}}(f, \tau)|^2) + 3\kappa_2^2(|Y_{\text{bias}}(f, \tau)|^2) \\ \quad + 4\kappa_3(|Y_{\text{bias}}(f, \tau)|^2)\kappa_1(|Y_{\text{bias}}(f, \tau)|^2) \\ \quad + 6\kappa_2(|Y_{\text{bias}}(f, \tau)|^2)\kappa_1^2(|Y_{\text{bias}}(f, \tau)|^2) \\ \quad + \kappa_1^4(|Y_{\text{bias}}(f, \tau)|^2) \end{cases} \quad \dots (3)$$

最後に 2-2 節にて述べたカートシス比と推定誤差減少率を定義し、上記の高次モーメントを用いて算出する。まず、カートシス比とは、入力雑音パワースペクトルのカートシスと推定器の出力に現れる雑音成分のカートシスとの比で与えられる。一般に、入力雑音分布（ガンマ分布）の m 次モーメントは

$$\mu_m(x) = \int_0^\infty x^m p_\Gamma(x) dx = \theta^m \eta(\eta+1) \cdots (\eta+m-1)$$

で与えられるため、式(3)との比をとると以下のカートシス比を得ることが出来る。

$$\text{Kurtosis ratio}(\alpha, \beta, \varepsilon, \eta, \rho) = \frac{\mu_4(|Y_{\text{bias}}(f, \tau)|^2)}{\mu_2^2(|Y_{\text{bias}}(f, \tau)|^2)} \cdot \frac{\eta(\eta+1)}{(\eta+2)(\eta+3)} \quad \dots (4)$$

また同様に、推定誤差減少率の計算には 1 次モーメントを用い、以下のような Noise Reduction Rate (NRR) を定義する。

$$\text{NRR}(\alpha, \beta, \varepsilon, \eta, \rho) = 10 \log_{10} \frac{\theta \eta}{\mu_1(|Y_{\text{bias}}(f, \tau)|^2)} \quad \dots (5)$$

式(4)及び(5)とも、事前 SN 比逐次推定の忘却係数 α 、圧縮パラメータ β 、バイアス ε 、雑音の形状母数 η 、目的音声の事前分布形状母数 ρ の関数であり、これらの値を変化させることにより、カートシス比の不動点が存在するか否かを予測することが出来る。

4 実験的評価

上記の手法によって算出されたカートシス比の数値計算例と、実際の音声強調を行った場合におけるカートシス比の実測値とを比較した。ここでは、加法雑音をガウス信号 ($\eta=1$) とし、事前 SN 比逐次推定の忘却係数を 0.98 に固定した。図 3 に、目的音事前モデルをガウス性とした場合におけるカートシス比と雑音抑圧量 (NRR) の挙動を示す。各線種は圧縮パラメータ β に対応し、またそれぞれの曲線は右上が $\varepsilon=0$ (バイアス無し; 通常の統計推定)、左下が ε を大きくした場合に相当する。数値計算値と実測値とに NRR の乖離はあるものの、どちらもカートシス比が 1 となる交差点が存在し、またその曲線の挙動も似ていることが分かる。これより、高次統計量不動点の存在が明らかになったとともに、提案する高速計算手法の有効性が確認できる。

次に、目的音事前モデルを優ガウス性とした場合におけるカートシス比と NRR の挙動を図 4 に示す。本条件下では、提案する数値計算例において高次統計量不動点が存在するものの、実測値においてほとんど高次統計量不動点は生じにくくなっている。これは、目的音事前モデルがよりカートシスの高いものになるため、残留雑音のカートシスも比較的高い値に留まるためだと思われる。本傾向は、目的音事前モデルをより優ガウス性にするほど顕著であった。以上より、一般に事前分布の優ガウス化は音声モデルにより近づくため音声強調には有効と言われているが、一方で残留雑音におけるカートシス不動点（聴覚印象の不動点）を得ることは困難になるということが分かった。つまり、両者にはトレードオフの関係があり、適切な事前モデルを設定することが重要である。

最後に、提案するバイアス付きベイズ推定器を用いて、従来の統計的音声強調法と音質評価実験を行った。比較対照法は、MMSE-STSA estimator [7]、MMSE-log spectral amplitude (LSA) estimator [10]、musical-noise-free spectral subtraction (SS) [4]、musical-noise-free Wiener filter (WF) [11]、及び musical-noise-free MMSE-STSA estimator [12] (提案法において $\rho=1$ に固定したもの) である。提案法の ρ と β はそれぞれ 0.5 及び 0.001 に設定した (つまり目的音事前分布は優ガウスに設定されている)。実験

結果を図5に示す。本図においては、全ての手法が $\text{NRR}=16\text{ dB}$ となるように調整されており、その同一条件下でのケプストラム歪み（音声品質を表す代表的スコア）の値を示している。本ケプストラム歪みの値が小さいほど、良い音声強調法であると言える。因みに、MMSE-STSA estimator 及び MMSE-LSA estimator のカートシス比は 2.7 であり、その他のミュージカルノイズフリー手法は 1.0 であった。本結果より、提案するバイアス付きベイズ推定器はミュージカルノイズの発生も無く、かつ音声の品質が最も良いことが分かった。

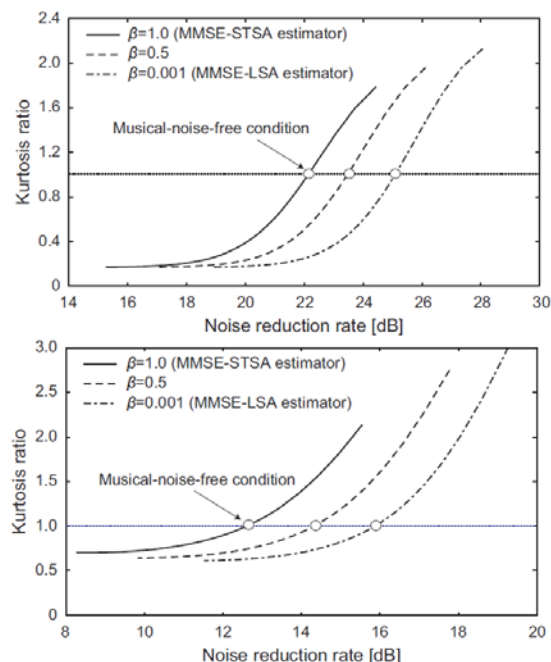


図3: ガウス性事前モデルにおける挙動。
(上)理論値, (下)実測値。

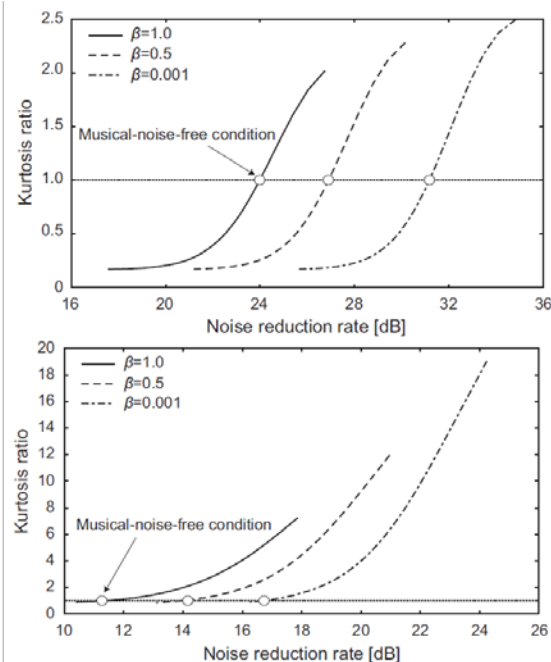


図4: 優ガウス性事前モデルにおける挙動。
(上)理論値, (下)実測値。

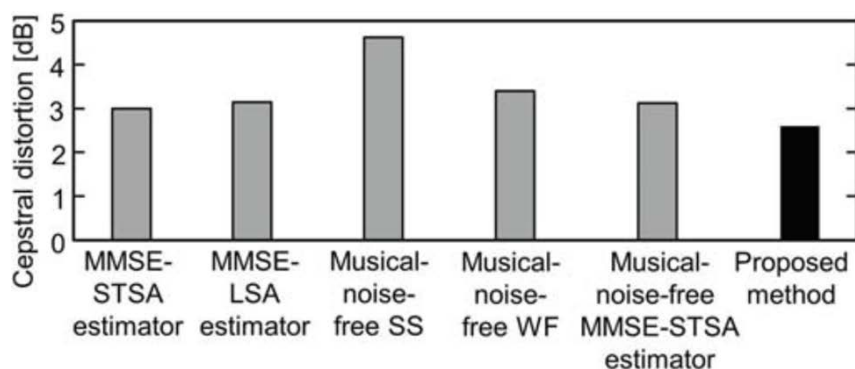


図5: 従来の音声強調手法と提案法におけるケプストラム歪みの比較

5 まとめ

本研究では、高次統計量の不動点に着目した新しい反復型統計推定理論及びその音響センシングへの応用に関して検討を行った。特に、弱いベイズ推定を反復適用する手法を確立し、高精度な音響・音声信号強調システムを構築した。本研究では第一に、通常の複素ガウス分布を目的音事前モデルとするベイズ推定に関しては高次統計量の不動点は存在しないため、振幅スペクトルをカイ分布にてモデリングする一般化された事前モデルを導入し、その場合における高次統計量の不動点を解析した。ここでは、高次統計量の不動点を生じさせるパラメータとして、事前 SN 比の推定ブロックにおいて若干の劣化（人工バイアス）を加えることとした。このバイアス値を増加させ、出力のカートシス比の値が 1 となる点を探索した。第二に、上記解析

を容易にするため、高次キュムラント領域における加法性を経由して高次モーメントの計算を行う手法を新たに導出した。最後に、提案する反復型バイアス付きベイズ推定器を用いて、従来の統計的音声強調法と音質評価実験を行った。本結果より、提案推定器の品質が最も良いことが分かった。

本研究における最大の知見は、目的音事前モデルを優ガウス性とした場合におけるカートシス比と NRR の挙動が明らかになったことである。本条件下では、実測値においてほとんど高次統計量不動点が生じにくくなった。これは、目的音事前モデルがよりカートシスの高いものになるため、残留雑音のカートシスも比較的高い値に留まるためだと思われる。以上より、一般に事前分布の優ガウス化は音声モデルにより近づくため音声強調には有効と言われている[13]が、一方で残留雑音におけるカートシス不動点を得ることは困難になるということが分かった。つまり、両者にはトレードオフの関係があり、適切な事前モデルの設定が重要であることを示唆している。今後は、上記の適切なモデルパラメータをいかにして自動獲得するかについて検討を進める予定である。

【参考文献】

- [1] 中井駿介, 宮崎亮一, 猿渡洋, 中村哲, 井上昂治, 若林佑幸, 河原達也, “スマートポスターボードにおける実環境を想定した複数話者分離,” 日本音響学会春季研究発表会講演論文集, no.2-Q4-8, 2014.
- [2] Takayuki Inoue, Hiroshi Saruwatari, Yu Takahashi, Kiyohiro Shikano, Kazunobu Kondo, “Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics,” IEEE Transactions on Audio, Speech & Language Processing, vol. 19, no. 6, pp. 1770-1779, 2011.
- [3] Hiroshi Saruwatari, Yohei Ishikawa, Yu Takahashi, Takayuki Inoue, Kiyohiro Shikano, Kazunobu Kondo, “Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher order statistics,” IEEE Transactions on Audio, Speech & Language Processing, vol. 19, no. 6, pp. 1457-1466, 2011.
- [4] Ryoichi Miyazaki, Hiroshi Saruwatari, Takayuki Inoue, Yu Takahashi, Kiyohiro Shikano, Kazunobu Kondo, “Musical-noise-free speech enhancement based on optimized iterative spectral subtraction,” IEEE Transactions on Audio, Speech & Language Processing, vol. 20, no. 7, pp. 2080-2094, 2012.
- [5] Ryoichi Miyazaki, Hiroshi Saruwatari, Satoshi Nakamura, Kiyohiro Shikano, Kazunobu Kondo, Jonathan Blanchette, Martin Bouchard, “Musical-noise-free blind speech extraction integrating microphone array and iterative spectral subtraction,” Signal Processing (Elsevier), vol. 102, pp. 226-239, 2014.
- [6] Yoshihisa Uemura, Yu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano, Kazunobu Kondo, “Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics,” Proc. International Workshop on Acoustic Echo and Noise Control, 2008.
- [7] Y. Ephraim, D. Malah, “Speech enhancement using a minimum mean square error short-time spectral amplitude estimator,” IEEE Trans. Acoustics, Speech and Signal Processing, vol. 32, no. 6, pp. 1109-1121, 1984.
- [8] C. Breithaupt, R. Martin, “Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions,” IEEE Trans. Audio, Speech, and Language Processing, vol. 19, no. 2, pp. 277-289, 2011.
- [9] Hiroshi Saruwatari, Suzumi Kanehara, Ryoichi Miyazaki, Kiyohiro Shikano, Kazunobu Kondo, “Musical noise analysis for Bayesian minimum mean-square error speech amplitude estimators based on higher-order statistics,” Proc. Interspeech2013, pp. 441-445, 2013.
- [10] Y. Ephraim, D. Malah, “Speech enhancement using a minimum mean square error log-spectral amplitude estimator,” IEEE Trans. Acoustics, Speech and Signal Processing, vol. 33, no. 2, pp. 443-445, 1985.
- [11] Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, Kazunobu Kondo, “Musical-noise-free speech enhancement based on iterative Wiener filtering,” Proc. IEEE International Symposium on Signal Processing and Information Technology, 2012.

- [12] Syunsuke Nakai, Hiroshi Saruwatari, Roichi Miyazaki, Satoshi Nakamura, Kazunobu Kondo, "Theoretical analysis of biased MMSE short-time spectral amplitude estimator and its extension to musical-noise-free speech enhancement," Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA), 2014.
- [13] Ryo Wakisaka, Hiroshi Saruwatari, Kiyohiro Shikano, Tomoya Takatani, "Speech prior estimation for generalized minimum mean-square error short-time spectral amplitude estimator," IEICE Transactions, vol. 95-A, no. 2, pp. 591-595, 2012.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Statistical-model-based speech enhancement with musical-noise-free properties	Proc. 2015 IEEE International Conference on Digital Signal Processing (DSP2015)	2015 年 7 月 (招待講演)