

バイノーラル信号音源分離における両耳事前分布モデルの考察*

☆室田 勇騎 (奈良先端大), 北村 大地 (総研大),
 小山 翔一, 猿渡 洋 (東京大学),
 中村 哲 (奈良先端大)

1 はじめに

近年, 信号処理の分野において音源分離技術が盛んに研究されている. この技術は3D オーディオシステムの実現や自動採譜への応用など様々な可能性を秘めている. 音源分離, 特にシングルチャネルの音声の分離において一般的に用いられるのはウィーナフィルタ (Wiener filter: WF) や平均二乗誤差最小化短時間振幅スペクトル推定器 (minimum mean-square error short-time spectral amplitude estimator: MMSE-STSA estimator) [1] などの非線形フィルタリングである. 特に MMSE-STSA 推定器については, 目的信号の事前分布モデルを用いたベイズ型推定器であり高い分離精度を誇る. しかしこれらの手法は, 非定常な妨害信号に対応することが難しく, また事前分布モデルも分離性能に影響を及ぼすが, 抽出音源にとって最適な分布を事前を知ることは困難である.

音源分離に用いられている別の技術として, 非負値行列因子分解 (nonnegative matrix factorization: NMF) があり, 近年盛んに研究されている [2]. NMF は入力された非負値行列を二つの非負値行列に分解する手法であり, これを用いることで振幅 (パワー) スペクトルドメインでの分解を行うことができる. また, NMF に事前学習を取り入れた教師あり NMF (supervised NMF: SNMF) [3] では, 妨害信号が非定常なものであっても精度よく分離を行うことができる. しかし, SNMF では振幅スペクトルの加法性などを仮定していることから, 近似分解となっており, 厳密には正しい分解にはなっていない.

そこで著者らはこれら 2 つの手法の利点を組み合わせることにより, 隠れた目的音の統計モデルパラメータを自動推定し, 目的音に適応した最適なベイズ推定を行うことが可能となる統計モデルパラメータの自動推定を備えた一般化 MMSE-STSA 推定法による音源分離手法¹を提案している [4].

本稿では上述の手法を拡張し, バイノーラル信号に対応させることを考える. バイノーラル信号における音源分離において, 両耳情報の利用が可能ならば分離性能の向上が期待されるが, このような情報は基本的に未知であるため, 実用面においてこれらを使用することは現実的ではない. そこでブラインドに推定可能な両耳情報として, 左右での目的音信号の事前分布の違いに着目し, 前述の提案手法を個別に導入することにより両耳での目的信号の事前分布を推定する. これを両耳情報として用いることで, 分離性能及び再減温の空間的品質の向上につながることを期待される. 以上を踏まえ, 本稿では基礎実験として, バイノーラル信号において適切な事前分布を推定することによる効果を確認することを目的とし, バイノーラル信号に対応した統計モデルパラメータの自動推定を備えた一般化 MMSE-STSA 推定法による音源分離手法を提案する. その後, 評価実験を行うことで, 提案手法の有効性を確かめる.

2 関連技術

2.1 一般化 MMSE-STSA 推定器 [5]

一般化 MMSE-STSA 推定器における混合モデルは以下のように, 時間周波数領域での足し合わせとし

て表現される

$$Y_R(f, \tau) + iY_I(f, \tau) = (S_R(f, \tau) + iS_I(f, \tau)) + (N_R(f, \tau) + iN_I(f, \tau)) \quad (1)$$

ここで, $Y_*(f, \tau)$ は観測音, $S_*(f, \tau)$ は目的音, $N_*(f, \tau)$ は妨害音を表し, $*$ = {R, I} は信号の実部及び虚部を表すインデックス, また f は周波数ビン, τ は時間フレームを表している. MMSE-STSA 推定器 [1] は最小平均二乗誤差基準に基づく目的音の推定法であり, 目的音の振幅スペクトルがレイリー分布に従うと仮定している. 一方, 一般化 MMSE-STSA 推定器 [5] では, 目的音の振幅スペクトルがカイ分布に従うと仮定しており, その母数を変化させることで多様な事前分布を表現することが可能である.

2.2 一般化 MMSE-STSA 推定器による音源分離

一般化 MMSE-STSA 推定器は, 目的音にとって最適な事前分布の下で, 真の目的音振幅スペクトルとその推定値との平均二乗誤差を最小化するスペクトルゲインを求める. このスペクトルゲインを観測信号スペクトルに乗算することで, 推定目的音が得られ, 音源分離が達成される. 一般化 MMSE-STSA 推定器における推定目的音を $\tilde{S}_*(f, \tau)$ とすると, これは以下の式で与えられる.

$$\tilde{S}_*(f, \tau) = G(f, \tau) Y_*(f, \tau) \quad (2)$$

$$G(f, \tau) = \frac{\sqrt{\nu(f, \tau)}}{\gamma(f, \tau)} \cdot \frac{\Gamma(\rho + 0.5)}{\Gamma(\rho)} \cdot \frac{\Phi(0.5 - \rho, 1, -\nu(f, \tau))}{\Phi(1 - \rho, 1, -\nu(f, \tau))} \quad (3)$$

ここで $\Gamma(\cdot)$ はガンマ関数, ρ は目的音の事前分布であるカイ分布の形状母数, $\Phi(a, b; k) = F_1(a, b; k)$ は合流型超幾何関数であり, $\nu(f, \tau)$ は以下の式で表される

$$\nu(f, \tau) = \tilde{\gamma}(f, \tau) \tilde{\xi}(f, \tau) (1 + \tilde{\xi}(f, \tau))^{-1} \quad (4)$$

$\tilde{\xi}(f, \tau)$, $\tilde{\gamma}(f, \tau)$ はそれぞれ事前 SNR, 事後 SNR であり, 以下の式で表される.

$$\tilde{\xi}(f, \tau) = \alpha \tilde{\gamma}(f, \tau - 1) G^2(f, \tau) + (1 - \alpha) \max[\gamma(f, \tau) - 1, 0] \quad (5)$$

$$\tilde{\gamma}(f, \tau) = (Y_R^2 + Y_I^2) / P_{\tilde{N}}(f) \quad (6)$$

ここで $P_{\tilde{N}}(f)$ は推定妨害音のパワースペクトル, α は忘却係数を表す.

一般化 MMSE-STSA 推定器における目的音の振幅スペクトルの事前分布は, 次のカイ分布で表される.

$$p(x) = \frac{2}{\Gamma(\rho)} \left(\frac{\rho}{E[x^2]} \right)^\rho x^{2\rho-1} \exp\left(-\frac{\rho}{E[x^2]} x^2 \right) \quad (7)$$

ここで $p(x)$ は信号 x の振幅スペクトルの p.d.f. であり ρ は形状母数である. カイ分布において, $\rho = 1$ であるとき, これはレイリー分布と一致し, 時間領域では信号が複素ガウス分布に従うことを仮定している.

*"Study on a priori statistical model of target signal for binaural signal source separation," by Yuki Murota (NAIST), Daichi Kitamura (SOKENDAI), Shoichi Koyama, Hiroshi Saruwatari, (The University of Tokyo) and Satoru Nakamura (NAIST)

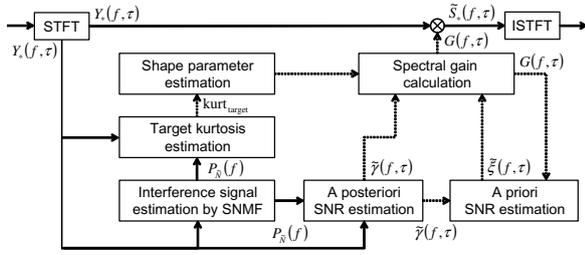


Fig. 1 Block diagram of proposed method.

また1より小さい ρ では優ガウス性の分布に従うことを仮定している。

一般化MMSE-STSA推定器において $\tilde{\gamma}(f, \tau)$ の計算に妨害音のパワースペクトルが必要となるが、妨害音が非定常である場合、動的な推定が必要となる。また、目的音の事前分布の形は形状母数 ρ によって決定されるが、形状母数 ρ は目的音の種類に応じて最適な値が異なり、事前にこれを知ることが出来ない。しかしながら、妨害音成分が推定できた場合、高次統計量を用いて解析的に形状母数 ρ を推定することができる。

Figure 1に事前分布パラメータ推定を備えた一般化MMSE-STSA推定法のブロック線図を示す。本手法では、まず動的な妨害音の推定が可能な手法を用いて推定妨害音振幅スペクトルを求め、そこから目的音の適切な事前分布(形状母数 ρ)を算出する。次に、推定された事前分布を用いた一般化MMSE-STSA推定器により音源分離を行う。次節では、推定妨害音が与えられた場合における形状母数の推定手法について述べる

2.3 目的音の事前分布推定

式(7)のカイ分布 $p(x)$ において、形状母数 ρ は以下のように表される。

$$\rho = (\mu_4/\mu_2^2 - 1)^{-1} \quad (8)$$

ここで、 μ_m は目的信号の振幅スペクトル分布の m 次モーメントであり、特に2次モーメントと4次モーメントにより算出される値 μ_4/μ_2^2 はカートシスと呼ばれる。 m 次モーメントは以下のように定義される。

$$\mu_m = \int_0^\infty x^m p(x) dx \quad (9)$$

式(8)から、観測信号より目的信号の振幅スペクトル分布のカートシスを推定できれば、観測される目的信号毎に最適な形状母数の推定が可能となり、それを用いて最適なスペクトルゲインの計算を行うことができる。しかし、実環境下では、目的信号は常に妨害信号の影響を受けるため、観測信号から目的信号のカートシスを直接推定することは困難である。そこで、高次統計量を利用した目的信号の振幅スペクトル分布の形状母数推定法について、次節で述べる。

2.3.1 目的信号の振幅スペクトルカートシスの推定法

本節では簡略化のために、各信号における f 及び τ の表記を省略する。まず、式(1)のような複素数領域(時間周波数領域)での混合モデルを考え、観測信号及び推定妨害信号から以下の m 次モーメントが得られたとする。

$$\mu_m(Y_R) = E[Y_R^m] \quad (10)$$

$$\mu_m(Y_I) = E[Y_I^m] \quad (11)$$

$$\mu_m(N_R) = E[N_R^m] \quad (12)$$

$$\mu_m(N_I) = E[N_I^m] \quad (13)$$

これらを用いると、複素数領域における目的信号の振

幅スペクトルのカートシスは以下のように表される。

$$\begin{aligned} \text{kurt}_{\text{target}} &= \frac{\mu_4((S_R^2 + S_I^2)^{\frac{1}{2}})}{\mu_2^2((S_R^2 + S_I^2)^{\frac{1}{2}})} \\ &= \frac{\mathcal{N}(\mu_m(Y_R), \mu_m(Y_I), \mu_m(N_R), \mu_m(N_I))}{\mathcal{D}(\mu_m(Y_R), \mu_m(Y_I), \mu_m(N_R), \mu_m(N_I))} \quad (14) \end{aligned}$$

ここで、各分子・分母は以下で与えられる。

$$\begin{aligned} \mathcal{N}(\mu_m(Y_R), \mu_m(Y_I), \mu_m(N_R), \mu_m(N_I)) &= \mu_4(Y_R) + \mu_4(Y_I) - \mu_4(N_R) - \mu_4(N_I) \\ &\quad + 6\mu_2^2(N_R) + 6\mu_2^2(N_I) + 2\mu_2(Y_R)\mu_2(Y_I) + 2\mu_2(N_R)\mu_2(N_I) \\ &\quad - 6\mu_2(Y_R)\mu_2(N_R) - 6\mu_2(Y_I)\mu_2(N_I) \\ &\quad - 2\mu_2(Y_R)\mu_2(N_I) - 2\mu_2(Y_I)\mu_2(N_R) \quad (15) \end{aligned}$$

$$\begin{aligned} \mathcal{D}(\mu_m(Y_R), \mu_m(Y_I), \mu_m(N_R), \mu_m(N_I)) &= \mu_2^2(Y_R) + \mu_2^2(Y_I) + \mu_2^2(N_R) + \mu_2^2(N_I) + 2\mu_2(Y_R)\mu_2(Y_I) \\ &\quad - 2\mu_2(Y_R)\mu_2(N_R) - 2\mu_2(Y_R)\mu_2(N_I) - 2\mu_2(Y_I)\mu_2(N_R) \\ &\quad - 2\mu_2(Y_I)\mu_2(N_I) + 2\mu_2(N_R)\mu_2(N_I) \quad (16) \end{aligned}$$

次に、上記の複素数領域で表されたカートシスの推定式を振幅スペクトルを用いて表すことを考える。これは妨害音が非定常な場合において、SNMF[3]を用いることで動的な妨害音の振幅スペクトルを推定できることを想定している[4]。この場合、妨害信号は振幅スペクトルのみ得られるため、実部 N_R 及び虚部 N_I を得られず、式(14)-(16)を直接計算することは出来ない。そこで、信号の実部、虚部で独立同一分布を仮定することにより、目的信号のカートシス推定式は振幅スペクトルを用いて以下のように表すことができる。

$$\text{kurt}_{\text{target}} = \frac{\mu_4(A) - \mu_4(I) + 4\mu_2^2(I) - 4\mu_2(A)\mu_2(I)}{\mu_2^2(A) + \mu_2^2(I) - 2\mu_2(A)\mu_2(I)} \quad (17)$$

ここで A は観測信号の振幅スペクトル $(Y_R^2 + Y_I^2)^{1/2}$ 、 I は妨害信号の振幅スペクトル $(N_R^2 + N_I^2)^{1/2}$ である。式(14)-(17)の導出については文献[4]を参照されたい。以上から、式(8)と式(17)を用いることにより、観測信号とSNMFの出力値のみから解析的に目的音の振幅スペクトルのカートシスを推定することが出来る。

3 事前分布パラメータ推定を用いた一般化MMSE-STSA推定法によるバイノーラル信号音源分離

3.1 概要

音源分離技術において、ヘッドホンのように両耳で音を聴取するシステムを考えた場合、音の臨場感などを保つために、分離音の定位や残響などは保持されていることが望ましい。このような分離を実現するためには、頭部伝達関数(head related transfer function: HRTF)などの両耳情報の利用が考えられるが、各ユーザのHRTF情報は基本的に未知であるため、実用面においてこれを利用することは現実的ではない。そこで、ブラインドに推定可能な両耳情報として、左右での目的音信号の事前分布の違いに着目する。具体的には、第三章で述べた事前分布パラメータ推定を用いた一般化MMSE-STSA推定法を利用し、左右それぞれのチャンネルで最適な目的音の事前分布を適用することで性能の向上が期待でき、また左右間の事前分布の差を調節することで分離後の目的音の知覚を変化させることが出来ると考える。Figure 2に事前分布を用いたバイノーラル信号のモデルを示す。

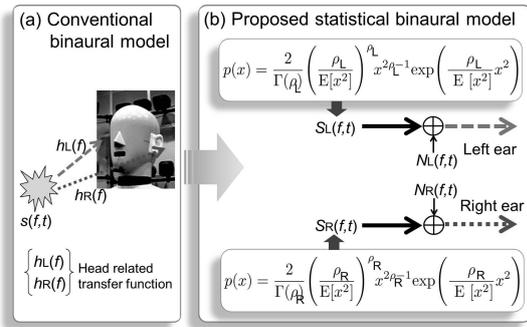


Fig. 2 Overview of proposed binaural model.

以上を踏まえ、本章以降では基礎実験として、バイノーラル信号において適切な事前分布を推定することによる効果を確認し、左右の事前分布を変化させたときの音の知覚影響の有無を確認する。次節以後において、事前分布パラメータ推定を用いた一般化 MMSE-STSA 推定法によるバイノーラル音源分離手法について述べる。まずバイノーラル信号における混合モデルを示す。その後、動的な雑音推定として SNMF を導入し、最後に評価実験を行い、得られた結果について考察する。

3.2 バイノーラル信号における混合モデル

左耳と右耳に 1 個ずつの計 2 個のマイクロホンで観測されたバイノーラル信号を考える。このとき時間周波数領域における観測信号 $\mathbf{x}(f, \tau) = [x^l(f, \tau), x^r(f, \tau)]^T$ は、目的信号 $s(f, \tau)$ と伝達関数 $\mathbf{h}(f) = [h^l(f), h^r(f)]^T$ 及び妨害信号 $\mathbf{n}(f, \tau) = [n^l(f, \tau), n^r(f, \tau)]^T$ を用いて以下のように表される（以下上付き文字 l は左耳、 r は右耳での信号を表すものとする）。

$$\mathbf{x}(f, \tau) = \mathbf{h}(f)s(f, \tau) + \mathbf{n}(f, \tau) \quad (18)$$

上記で示される左右それぞれのチャンネルに対して、第三章で述べた事前分布パラメータ推定を用いた一般化 MMSE-STSA 推定法を個別に適用することで音源分離を行う。

3.3 SNMF を用いた妨害音推定

SNMF は、事前学習によって得られた教師基底 $F(f, k)$ を用いて観測信号のスペクトログラムを近似分解する手法である。SNMF による観測スペクトログラムの分解は以下の式で表される。

$$A^m(f, \tau) \approx \sum_k F^m(f, k)V^m(k, \tau) + \sum_n H^m(f, n)U^m(n, \tau) \quad (19)$$

ここで m はマイクロホンのチャンネル ($m=l$ or r) を表す。 $F(f, k)$ は事前学習された目的音のスペクトルパターンを含む基底行列の要素値である。 $V(f, k)$ は $F(f, k)$ に対応するアクティベーション行列の要素値であり、各スペクトルパターンの時間強度変化を示す。 $H(f, n)$ 及び $U(n, t)$ はそれぞれ、目的音以外の成分を表すための基底行列及びアクティベーション行列の要素値であり、 $F(f, k)$, $V(k, t)$, $H(f, n)$, 及び $U(n, t)$ はいずれも非負の実数値となる。

また、 k は $F(f, k)$ の基底数を表し、 n は $H(f, n)$ の基底数を表す。ここで、事前学習で得られる教師基底 $F^m(f, k)$ は、目的音のみが含まれる教師信号を用いて構成される。SNMF では、この教師基底 $F^m(f, k)$ を固定した状態で、 $V^m(k, t)$, $H^m(f, n)$, 及び $U^m(n, t)$ を求めるため、理想的には $\sum_k F(f, k)V(k, \tau)$ は観測信号に含まれる目的信号の成分を、 $\sum_n H(f, n)U(n, \tau)$ は目的信号以外の成分をそれぞれ表している。

これより得られた $\sum_n H(f, n)U(n, \tau)$ (又は、 $A(f, \tau) - \sum_k F(f, k)V(k, \tau)$) を推定妨害音の振幅スペクトルとして一般化 MMSE-STSA 推定器に用いる。SNMF では教師基底を用いるので、音楽信号のように非定常で動

的な妨害音の推定も可能であり、 $(\sum_n H(f, n)U(n, \tau))^2$ は式 (6) の $P_N(f)$ のよい推定値になっていると考えられる。

3.4 バイノーラル信号における妨害音推定手法

バイノーラル信号では、頭部回折や部屋の残響特性などの影響により、同じ時間周波数グリッドでも左右チャンネルで目的音成分が異なる。よって、先ほど述べた SNMF による妨害信号の推定をそのまま用いると、事前学習によって得られた教師成分と実際の両耳の目的音成分との差異により、推定精度が劣化する可能性がある。これを考慮し、以下に示す三種類の手法によって上記の影響を補正した妨害信号を推定する。

基底変形型 SNMF による妨害音推定 [6]

基底変形型 SNMF は観測信号中の目的音に合わせて、事前学習によって得られた教師基底を変形させ、それを用いて観測信号を分解する。観測信号は次式のように分解される

$$A^m(f, \tau) \approx \sum_k (F^m(f, k) + D^m(f, k))V^m(k, \tau) + \sum_n H^m(f, n)U^m(n, \tau) \quad (20)$$

ここで、 $D^m(f, k)$ は教師基底 $F^m(f, k)$ に対する変形成分を示し、 $\eta F^m(f, k) + D^m(f, k)$ が非負となる範囲で、 $D^m(f, k)$ は正負の実数値をもつ。但し、 η は教師基底 $F^m(f, k)$ に対する負の変形許容範囲を表すパラメータである。この分解では、教師スペクトル $F^m(f, k)$ が実際の目的音スペクトルとわずかに異なっていた場合に、その差分を $D^m(f, k)$ で表現することで、教師基底を目的音に適応的に変形させることを期待している。これにより、頭部回折や部屋の伝達特性を含んだ妨害音を推定する。

伝達関数を畳み込んだ教師基底を用いた SNMF

実験的に、観測系の伝達関数 (HRTF 等) が既知であるという仮定をおき、その伝達関数を畳み込んだサンプルデータを用いて教師基底 $F^m(f, k)$ を作成する。従って、得られる教師基底は両耳の特性を含んだものとなるため、頭部回折や部屋の伝達特性を含んだより精度の高い妨害音推定が可能となる。

真の妨害音を用いる

比較参考値として、観測信号に混合されている真の妨害音を推定妨害音として用いる。

4 評価実験

4.1 実験条件

事前分布パラメータ推定を用いた一般化 MMSE-STSA 推定法において、バイノーラル信号音源分離への効果と左右の事前分布の違いによる影響を考察するために、HRTF を畳み込んだバイノーラル音源を作成して音源分離実験を行った。まず、実験用データとして MIDI 音で作成した Oboe, Clarinet, Piano, Cello の 4 種類を用意し、Oboe を抽出対象音として選んだ。それぞれのスコア (楽譜) を Fig.3 に示す。観測信号は Oboe 以外の 3 種類の音源データの中から 1 つを妨害音として選び、目的音、妨害音共に左側 90° の方向に配置し等パワーで混合したものを使用する。各音源の作成にあたっては、ダミーヘッドにより収録された HRTF[7] を使用し、該当方向の HRTF 畳み込むことで作成した。また教師音として MIDI シーケンサで Oboe の人工信号を作成した。この教師音は、観測信号中の各楽器音の音域を含む範囲で、2 オクターブを半音階ずつ上昇する 24 音で構成されている。教師信号及び、目的信号はいずれもサンプリング周波数 44.1kHz であり、FFT 点数 8192、窓長 4096 の矩形窓を 512 でシフトさせてスペクトログラムを作成した。

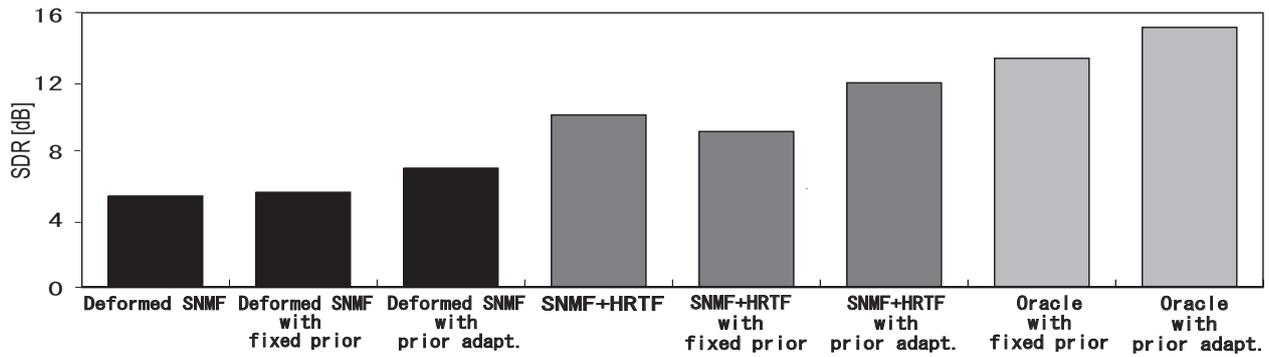


Fig. 4 Average SDRs of each method.



Fig. 3 Scores of each part.

また事前学習における学習基底数は 100, 信号分離における妨害音信号の基底数は 50 とした. 忘却係数 α については 0.97 とした.

比較手法として, バイノーラル信号における目的音の事前分布推定の効果を確認するために, それぞれ左右のチャンネルで MMSE-STSA 推定法を行った場合 (“with fixed prior”) と両耳で事前分布パラメータ推定を用いた一般化 MMSE-STSA 推定法を行った場合 (“with prior adapt.”) とを比較する. また妨害音の推定方法も 4.3 節で示した 3 種類の推定手法を行い, それぞれの場合において評価値への影響と事前分布推定の効果があるのかを考察する. 次節では基底変形型 SNMF による妨害音推定を Deformed SNMF, 伝達関数を畳み込んだ教師基底を用いた SNMF による妨害音推定を SNMF+HRTF, 真の妨害音を用いた場合を Oracle とする.

4.2 実験結果及び考察

本実験では客観評価値として BSS EVAL TOOLBOX[8] による source to distortion ratio (SDR) を用いる. SDR は目的音と妨害音の分離度合いと, 一連の信号処理によって生じた目的音のひずみの両方を考慮した, 推定目的音の品質を表すものであり, 値が大きくなるほど精度が良いことを示す.

Figure 4 に, 各比較手法で分離を行った場合の左チャンネルの平均 SDR 値を示す. 各棒グラフ名について前者はポスト処理の違いを, 後者は妨害音推定手法の違いを示す. 例えば Deformed SNMF with fixed prior の場合, 基底変形型 SNMF で推定された推定妨害音を用いて MMSE-STSA 推定法を行ったものとなっている. また Deformed SNMF, SNMF+HRTF のように, 妨害音推定手法だけが示されているものはポスト処理を行わずに, それぞれ推定手法で得られた目的音を出力したものの評価値となっている. Figure 4 より事前分布パラメータ推定を用いた一般化 MMSE-STSA 推定器を用いることで SDR 値がより改善していることがわかる. また事前分布推定によって得られた形状母数 ρ は, いつも 1 以下の値を示していた. これはバイノーラル音楽信号の事前分布が, 伝達関数などの影響を受けてもより急なピークを持つ分布であることを示している. よってあらかじめ固定された事前分布を仮定した MMSE-STSA 推定法よりも, 適切な

事前分布を推定しそれに合わせた分離を行う方が良いと考えられる. 以上から, バイノーラル信号における事前分布の推定の効果を確認することができた.

5 まとめ

本稿では, バイノーラル音楽信号分離において両耳間での目的信号の分布の違いによる分離精度や音の知覚による影響を調べるため, 事前分布推定を備えた一般化 MMSE-STSA 推定器をバイノーラル信号に拡張し実験を行った. 実験結果より, バイノーラル信号においても目的音の事前分布の推定を行うことで, 分離精度が向上する確認を確認した.

References

- [1] 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.
- [2] D. D. Lee, H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. Advances in Neural Information Processing Systems*, vol.13, pp.556–562, 2001. 2001.
- [3] D. Kitamura *et al.*, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol.E97-A, no.5, pp.1113–1118, 2014.
- [4] Y. Murota *et al.*, “Music Signal Separation Based on Bayesian Spectral Amplitude Estimator with Automatic Target Prior Adaptation,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.7490–7494, 2014
- [5] C. Breithaupt, M. Krawczyk, R. Martin, “Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.4037–4040, 2008.
- [6] D. Kitamura *et al.*, “Music signal separation by supervised nonnegative matrix factorization with basis deformation,” *Proc. DSP2013*, T3P(C)-1, 2013.
- [7] 島田 正治, 他, “頭部伝達関数 (HRTF),” サイバー出版センター, 東京, 2014.
- [8] E. Vincent, R. Gribonval, C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech and Language Processing*, vol.14, no.4, pp.1462–1469, 2006.