

時系列事前分布モデルとスペクトル基底の同時適応を用いたバイノーラル音源分離の実験的評価*

◎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. 本稿では特にバイノーラル形式の音源分離について述べる. バイノーラル信号における音源分離では, 音の臨場感を保つために, 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. このような分離を行うためには, 頭部伝達関数 (head related transfer function: HRTF) などの両耳情報を利用することが考えられる. しかし各ユーザの HRTF 情報は基本的に未知であるため, 本 HRTF の利用は実用上において現実的ではない. そこで著者らは, 事前分布パラメータ推定を用いたゲイン共通型一般化平均二乗誤差最小化短時間振幅スペクトル推定法 (minimum mean-square error shot-time amplitude estimator: MMSE-STSA) によるバイノーラル音楽信号分離手法を提案している [1]. この手法は, 教師あり非負値行列因子分解 (supervised nonnegative matrix factorization: SNMF)[2] による動的な妨害音推定, 高次統計量を用いた事前分布パラメータ推定手法, そして両耳で用いるゲイン関数を共通化するゲイン共通型一般化 MMSE-STSA 推定器の 3 処理から構成されている. この手法では, 両耳観測データから分離対象音の時系列情報である事前分布モデルを左右個別に推定し, 分離に利用する. これにより, HRTF 等の両耳情報を用いることなく, 精度のよい分離を行うことができる.

一方で前述の手法では, 時系列方向の適応に比べ, 周波数方向の適応は十分に行われていなかった. 本手法では, まず分離対象音のスペクトル基底情報を SNMF によって事前に学習し, それを用いて観測音から妨害音成分推定を行う. しかし, 事前学習によって得られた教師成分と実際に観測される目的音信号との差異により, 推定精度が劣化する可能性がある. この問題を解決するために, 観測音信号に合わせて教師基底を変形し分離を行うことが考えられる. このような手法は基底変形型 SNMF[3] として提案されているが, 教師基底の変形と信号の分離を同時に行うため最適化が難しく分離精度が良くないという問題点があった. そこで本研究では, この教師基底を中間的な音源分離音 (MMSE-STSA 推定器出力) に適応させることにより, 時系列事前分布モデルとスペクトル基底を反復的に同時適応する手法を提案する. その後, 評価実験を行うことで提案手法の有用性を確かめる.

2 関連技術

2.1 事前分布パラメータ推定を用いたゲイン共通型一般化 MMSE-STSA 推定器によるバイノーラル音楽信号分離

本章では著者らが提案した, 事前分布パラメータ推定を用いたゲイン共通型一般化 MMSE-STSA 推定法によるバイノーラル音楽音源分離手法について述べる. Figure 1 に従来手法におけるバイノーラル信号モデルを示す. この手法では, 左右の耳に到来する目的音信号振幅スペクトルの確率密度関数 (probability

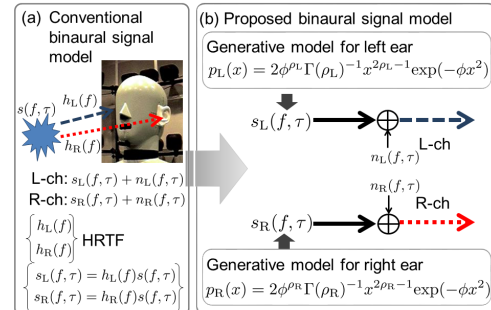


Fig. 1 Conventional and proposed binaural models.

density function: p.d.f.) の差を両耳情報として利用し分離を行うことにより, 品質の良い分離音を得ることができる. 具体的には, 伝達関数が畳み込まれた目的音信号をカイ分布を用いた統計的な信号モデルで表すことで, 従来の決定論的な HRTF の推定問題を統計パラメータの推定問題に変換する. そして観測データから左右チャンネル毎にこの統計パラメータを推定し分離に利用する. この手法では, SNMF による動的な妨害音推定, 高次統計量を用いた閉形式の目的音事前分布パラメータの推定手法, 及び音像定位の改善を目的とした両耳共通スペクトルゲインを用いている. 次節では一般化 MMSE-STSA 推定器, SNMF による妨害音推定について説明する (両耳共通スペクトルゲインの導出, 目的音事前分布パラメータ推定法については文献 [1] を参照されたい).

2.2 信号の混合モデル

本研究では, 左耳と右耳に一個ずつの 2 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音信号 $\mathbf{x}(f, \tau) = [x_L(f, \tau), x_R(f, \tau)]^T$ は, 目的音信号 $s(f, \tau)$ と伝達関数 $\mathbf{h}(f) = [h_L(f), h_R(f)]^T$ 及び妨害音信号 $\mathbf{n}(f, \tau) = [n_L(f, \tau), n_R(f, \tau)]^T$ を用いて以下で表されるとする.

$$\mathbf{x}(f, \tau) = \mathbf{h}(f)s(f, \tau) + \mathbf{n}(f, \tau) \quad (1)$$

ここで, f は周波数ビン, τ は時間フレームを表し, 下付き添え字 $*$ ($*$ = {L, R}) はそれぞれ左耳と右耳での信号を表すとする.

2.3 一般化 MMSE-STSA 推定器 [4]

一般化 MMSE-STSA 推定器は, 目的音を表す事前分布の基で, 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.

$$\tilde{s}_*(f, \tau) = G_*(f, \tau)x_*(f, \tau) \quad (2)$$

$$G_*(f, \tau) = \frac{\sqrt{v_*(f, \tau)}}{\gamma_*(f, \tau)} \cdot \left(\frac{\Gamma(\rho_* + 0.5)}{\Gamma(\rho_*)} \cdot \frac{\Phi(0.5 - \rho_*, 1, -v_*(f, \tau))}{\Phi(1 - \rho_*, 1, -v_*(f, \tau))} \right)^{1/\beta} \quad (3)$$

* "Experimental evaluation of simultaneous adaptation method of time-series a priori statistical model and spectral basis for binaural source separation," by Yuki Murota (NAIST), Daichi Kitamura (SOKENDAI), Shoichi Koyama, Hiroshi Saruwatari (The University of Tokyo), and Satoshi Nakamura (NAIST).

ここで $\Gamma(\cdot)$ はガンマ関数、 $\Phi(a, b; k) = F_1(a, b; k)$ は合流型超幾何関数、 β は振幅圧縮パラメータを表す。また、 $v_*(f, \tau)$ は以下の式で表される

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

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

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

$$\tilde{\gamma}_*(f, \tau) = |x_*(f, \tau)|^2 / P_{\tilde{n}^*}(f) \quad (6)$$

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。

$$p(x) = 2\phi^\rho \Gamma(\rho)^{-1} x^{2\rho-1} \exp(-\phi x^2) \quad (7)$$

$$\phi = \frac{\rho}{E\{|x|^2\}} \quad (8)$$

ここで $p(x)$ は信号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と仮定している。また 1 より小さい ρ では、優ガウス性の分布に従うと仮定している。

一般化 MMSE-STSA 推定器では、 $\tilde{\gamma}_*(f, \tau)$ 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、目的音の事前分布を最もよく表すための最適な形状母数についても推定する必要がある。

2.4 SNMF による妨害音推定

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。

$$A_*(f, \tau) = |x_*(f, \tau)| \approx \sum_k F_*(f, k) V_*(k, \tau) + \sum_n H_*(f, n) U_*(n, \tau) \quad (9)$$

ここで、 $F_*(f, k)$ は事前学習された目的音信号のスペクトルパターンを含む基底行列の要素値である。 $V_*(k, \tau)$ 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 $U_*(n, \tau)$ はそれぞれ、目的音信号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、 $F_*(f, k)$, $V_*(k, \tau)$, $H_*(f, n)$, 及び $U_*(n, \tau)$ はいずれも非負の実数値となる。また、 k は $F_*(f, k)$ の基底のインデックスを表し、 n は $H_*(f, n)$ の基底のインデックスを表す。ここで、事前学習で得られる教師基底 $F_*(f, k)$ は、目的音信号のみが含まれる教師信号を用いて構成される。この分解では $F_*(f, k)$ を固定した状態で $V_*(k, \tau)$, $H_*(f, n)$, 及び $U_*(n, \tau)$ を求めるため、理想的には $\sum_k F_*(f, k) V_*(k, \tau)$ は観測音信号に含まれる目的音信号の成分を、 $\sum_n H_*(f, n) U_*(n, \tau)$ は目的音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。この手法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信号の推定も可能であり、 $(\sum_n H_*(f, n) U_*(n, \tau))^2$ は式 (6) の $P_{\tilde{n}^*}(f)$ のよい推定値であると考えられる。

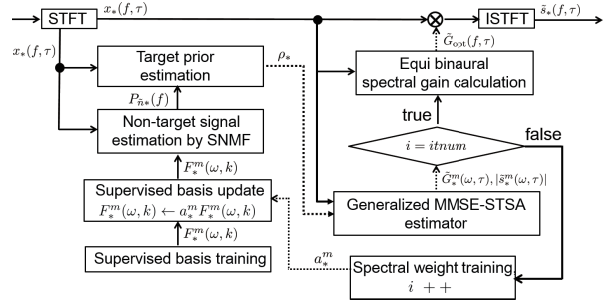


Fig. 2 Block diagram of proposed method.

3 提案手法

3.1 提案手法の概要

バイノーラル信号では、頭部回折などの影響によりスペクトル上での変形が生じ、左右チャンネルでの目的音成分のスペクトル基底は事前学習されたスペクトル基底と異なる。よって、SNMF による妨害音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。これを解決するためには、例えば観測音信号に合わせて教師基底を変形し分離を行う基底変形型 SNMF を導入することが考えられる。

しかし、従来の基底変形型 SNMF では変形成分が各教師基底の各グリッドに対して求まるため、畳み込みで表されるような時不変の変形に対して過度な変形を許容してしまう。また教師基底の変形と信号分離を同時に行うため、最適化が困難である。そこで本研究では、SNMF において一般化 MMSE-STSA 推定器の出力値を用いて教師基底の変形を行う手法を提案する。そして、この基底変形手法を従来手法である事前分布パラメータ推定を備えたゲイン共通型一般化 MMSE-STSA 推定法と組み合わせることを考える。Figure 2 に提案手法の概要を示す。

本手法では、観測音信号に対する周波数方向への適応を SNMF における教師基底の変形により行い、時間方向への適応を目的音信号の事前分布パラメータ推定により行う。具体的にはまず、従来手法により推定目的音信号を得る。次に、その出力値を用いて SNMF に用いる教師基底の変形を行う。そして、変形した教師基底を用いて事前分布パラメータ推定を備えたゲイン共通型一般化 MMSE-STSA 推定器を行う。最後にこれらの処理を任意の反復数 $itnum$ だけ行い、その後得られる推定目的音を最終的な出力値とする。

3.2 一般化 MMSE-STSA 推定器の出力を用いた教師基底の適応

本節では、一般化 MMSE-STSA 推定器の出力を用いた教師基底の変形手法について述べる。提案手法では式 (9) と同じく、目的音信号のみが含まれる教師信号を用いて教師基底 F を学習し、その後、前の反復により得られた推定目的音信号を用いて教師基底の変形成分を学習する。教師基底の変形成分の学習は次式のように行われる。

$$|I_*^m(\omega, \tau) \circ \tilde{s}_*^m(\omega, \tau)| \approx \sum_k (I_*^m(\omega, \tau) \circ \alpha_*^m F_*^m(\omega, k) V_*(k, \tau)) \quad (10)$$

ここで $\tilde{s}_*^m(f, \tau)$ は前の反復において得られる一般化 MMSE-STSA 推定器出力の要素値を表す。 m, ω はそれぞれ要素値が属するサブバンドのインデックス、サブバンド内における周波数インデックスを表す。 $I_*(\omega, \tau)$ は 1 と 0 の値のみを持つバイナリマスクであり、前の反復において式 (2) より計算されるゲイン関数 $G_*^m(\omega, \tau)$ が閾値 θ ($0 \leq \theta \leq 1$) を超えているグリッドは 1 を、それ以外は 0 をとるように定める。ま

た a_*^m は教師基底の変形を行うためのスペクトル重みを表し、 \circ はアダマール積を表す。この学習では、サブバンド毎にスペクトル重み a_*^m を学習し、教師基底に乗ずることで観測音信号中の目的音信号に適応させる。それにより頭部回折や部屋の伝達特性を表すことを期待している。また、推定目的音信号をそのまま用いると、残留している妨害音信号などの影響によりスペクトル重みの学習が適切に実行されない場合がある。そこでバイナリマスク $I_*(\omega, \tau)$ を用いることでスペクトル重みの学習に用いるデータの選抜を行っている。これにより、主に目的音信号のみが存在するグリッドを用いて学習を行うことでスペクトル重みの推定精度を向上させることを期待している。

3.3 目的関数と更新式

本節では、スペクトル重みの学習に用いる NMF の目的関数及び更新式の導出過程を示す。本研究では、一般化 KL ダイバージェンスを NMF の目的関数として用いる。一般化 KL ダイバージェンスは次式で定義される。

$$\mathcal{D}_{\text{KL}}(y||x) = y(\log y - \log x) + x - y \quad (11)$$

これを用いると、スペクトル重みの学習における NMF の目的関数 \mathcal{J} は以下の様に表される。

$$\begin{aligned} \mathcal{J} &= \sum_m \sum_{\omega, \tau} I_*^m(\omega, \tau) \mathcal{D}_{\text{KL}}(|s_*^m(\omega, \tau)| \| \sum_k a_*^m F_*^m(\omega, k) V_*(k, \tau)) \\ &= \sum_m \sum_{\omega, \tau} I_*^m(\omega, \tau) \left[-s_{\omega, \tau}^m \log R_{\omega, \tau}^m + R_{\omega, \tau}^m + C_k \right] \end{aligned} \quad (12)$$

$$R_{\omega, \tau}^m = \sum_k a_*^m f_{\omega, k}^m v_{k, \tau} \quad (13)$$

$$C_k = I_{\omega, \tau}^m (s_{\omega, \tau}^m \log s_{\omega, \tau}^m - s_{\omega, \tau}^m) \quad (14)$$

ここで、 $I_*^m(\omega, \tau)$, $s_*^m(\omega, \tau)$, $f_{\omega, k}^m$, $v_{k, \tau}$ はそれぞれ $I_*^m(\omega, \tau)$, $|s_*^m(\omega, \tau)|$, $F_*^m(\omega, k)$, $V_*(k, \tau)$ に対応する要素値である。式 (12) を最小化する際に更新するものは a_*^m 及び $v_{k, \tau}$ である。しかしこれらを解析的に導出することは困難なため、式 (12) の上限を与える関数を用いた補助関数法より更新式を導出する。式 (12) の対数項に関して補助変数 $\beta_{k, \omega, \tau}^m \geq 0$ と Jensen の不等式を用いると \mathcal{J} の上限関数 \mathcal{J}^+ は以下の様に与えられる。ただし $\sum_k \beta_{k, \omega, \tau}^m = 1$ とする。

$$\begin{aligned} \mathcal{J} &\leq \mathcal{J}^+ \\ &= \sum_m \sum_{\omega, \tau} I_*^m(\omega, \tau) \left[-s_{\omega, \tau}^m \sum_k \beta_{k, \omega, \tau}^m \log(a_*^m f_{\omega, k}^m v_{k, \tau}) \right. \\ &\quad \left. + \sum_k a_*^m f_{\omega, k}^m v_{k, \tau} + C'_k \right] \end{aligned} \quad (15)$$

$$C'_k = C_k + I_{\omega, \tau}^m (s_{\omega, \tau}^m \log \beta_{k, \omega, \tau}^m) \quad (16)$$

ここで、式 (15) の等号成立条件は以下で与えられる。

$$\beta_{k, \omega, \tau}^m = \frac{a_*^m f_{\omega, k}^m v_{k, \tau}}{\sum_k a_*^m f_{\omega, k}^m v_{k, \tau}} \quad (17)$$

\mathcal{J}^+ を各変数で偏微分し 0 とおいた式に、補助関数の等号成立条件を代入することで、各変数の乗法更新式を導出できる。まず a_*^m の更新式を導出する。 $\partial \mathcal{J}^+ / \partial a_*^m = 0$ より以下が成立する。

$$\sum_{\omega, \tau} I_{\omega, \tau}^m \left[-s_{\omega, \tau}^m \sum_k \frac{\beta_{k, \omega, \tau}^m}{a_*^m} + \sum_k f_{\omega, k}^m v_{k, \tau} \right] = 0 \quad (18)$$

式 (18) に補助変数を代入し整理すると、 a_*^m の更新式が導出できる。

$$a_*^m \leftarrow a_*^m \frac{\sum_{\omega, \tau} I_{\omega, \tau}^m s_{\omega, \tau}^m \sum_k (f_{\omega, k}^m v_{k, \tau}) (\sum_k a_*^m f_{\omega, k}^m v_{k, \tau})^{-1}}{\sum_{\omega, \tau} \left(\sum_k I_{\omega, \tau}^m f_{\omega, k}^m v_{k, \tau} \right)} \quad (19)$$

同様に $v_{k, \tau}$ の更新式は以下で与えられる

$$v_{k, \tau} \leftarrow v_{k, \tau} \frac{\sum_{m, \omega} I_{\omega, \tau}^m s_{\omega, \tau}^m (a_*^m f_{\omega, k}^m) (a_*^m \sum_k f_{\omega, k}^m v_{k, \tau})^{-1}}{\sum_{m, \tau} \left(I_{\omega, \tau}^m a_*^m \sum_k f_{\omega, k}^m \right)} \quad (20)$$

本手法では、スペクトル重み a_*^m の学習は他のサブバンドに依存しない。また、この時に更新されるアクティベーション $v_{k, \tau}$ は学習後に破棄する。このようにして学習されたスペクトル重みを教師基底に乗ずる事により、教師基底の変形を行う。その後、変形された教師基底を用いて SNMF を行うことにより妨害音推定を行う。

4 評価実験

4.1 実験条件

提案手法の有用性を確認するために、評価実験を行った。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 (それぞれの楽譜は文献 [5] を参照されたい)。観測音信号に関しては 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、目的音信号、妨害音信号共に -90° から 90° の間で 15° の間隔で等パワーにて配置した。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 [6] を使用し、該当方位の HRTF を畳み込むことで両耳信号を作成した。また教師音信号として MIDI シーケンサで Oboe の人工信号を作成し、観測音信号とは別の正面方位の HRTF を畳み込んだものを使用する。この教師音信号は、観測音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。また教師音信号及び、目的音信号はいずれもサンプリング周波数 44.1 kHz であり、FFT 点数 8192、窓長 4096 点の矩形窓を 512 点でシフトさせてスペクトログラムを作成した。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、振幅圧縮パラメータ β については 1.0 とした。また、客観評価値として BSS EVAL TOOLBOX [7] 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線形ひずみ、非線形ひずみ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。

4.2 実験結果

評価実験では、提案手法と従来手法との比較実験を行う。比較対照法として、反復を行うことなく分離と教師基底の適応を同時に行い妨害音信号を推定する SNMF と、他のゲイン共通型ポストフィルタとを組み合わせた手法を想定する。従来手法に用いるポストフィルタは、ゲイン共通型 WF (Equi-gain WF) [8]、ゲイン共通型 MMSE-STSA 推定器 (Equi-gain MMSE-STSA) [9]、最小ゲインベースのゲイン共通型 MMSE-STSA 推定器 (Gain-min MMSE-STSA) [10]、そして、2 章で述べたゲイン共通型一般化 MMSE-STSA 推定器 (Equi-gain Generalized MMSE-STSA) [1] を用いる。これらの従来手法に用いる SNMF については、付録を参照されたい。また、提案手法における反復数は 4、閾値 θ は 0.8 とし、全てのサブバンドにおいてスペクトル重みの初期値を 1 とした。

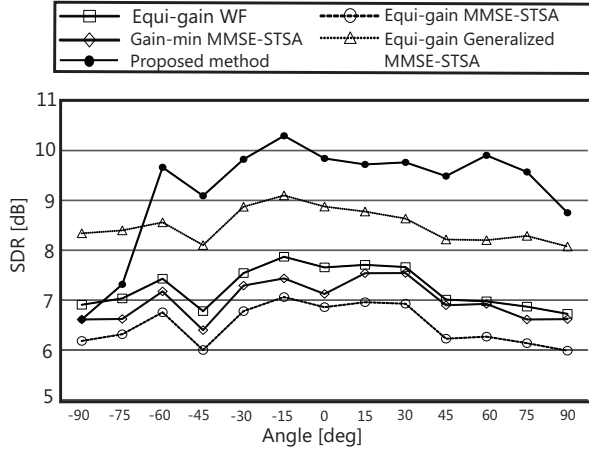


Fig. 3 Average SDRs for each method and each direction.

各手法における音源方位別の評価値を Fig. 3 に示す。本図より、ほぼすべての方位で提案手法が従来手法より良い結果を示していることが分かる。この結果より、教師基底の変形と妨害音の分離を個別に行い、そして学習データの中間的な分離音を用いて教師基底の変形成分を学習することの有効性が確認できた。

5 まとめ

本研究では、SNMF に教師基底の変形成分を反復的に学習する手法を取り入れることにより、従来手法の事前分布パラメータ推定を用いたゲイン共通型一般化 MMSE-STSA 推定法の改良を行った。従来手法の問題点であるスペクトル基底の周波数方向への適応精度が提案手法により向上し、より精度の高い分離ができることを示した。

References

- [1] 室田勇騎, 北村大地, 小山翔一, 猿渡洋, 中村哲, “チャンネル別事前分布推定と両耳共通スペクトルゲインを用いた定位保持型バイノーラル音源分離,” 第 29 回信号処理シンポジウム, pp. 486-491, 2014.
- [2] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, K. Kondo, “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.
- [3] D. Kitamura, H. Saruwatari, K. Shikano, Y. Takahashi, K. Kondo, “Music signal separation by supervised nonnegative matrix factorization with basis deformation,” *Proc. DSP2013*, T3P(C)-1, 2013.
- [4] 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*, pp.4037–4040, 2008.
- [5] Y. Murota, D. Kitamura, S. Nakai, H. Saruwatari, S. Nakamura, K. Shikano, Y. Takahashi, K. Kondo, “Music signal separation based on bayesian spectral amplitude estimator with automatic target prior adaptation,” *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2014)*, pp.7490–7494, 2014.

- [6] S. Shimada, K. Sugiyama, H. Hokari “*Head-Related Transfer Function*,” Cyber Publishing Center Press, Cyber Creative Institute, Japan, 2014 (in Japanese).
- [7] 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.
- [8] K. Reindl, Y. Zheng, W. Kellermann, “Speech enhancement for binaural hearing aids based on blind source separation,” *Proc. 2010 International Symposium on Communications, Control and Signal Processing (ISCCSP2010)*, 2010.
- [9] H. Saruwatari, M. Go, R. Okamoto, K. Shikano, H. Hosoi, “Binaural hearing aid using sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation,” *Proc. 2010 International Workshop on Acoustic Echo and Noise Control (IWAENC2010)*, 2010.
- [10] A. H. Kamkar-Parsi, M. Bouchard, “Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment,” *IEEE Trans. Audio, Speech and Lang. Process.*, vol.17, no.4, pp.521–533, 2009.

A 付録：基底変形型 SNMF

従来手法では、教師基底の変形と分離を同時に行う基底変形型 SNMF を用いて妨害音推定を行う。目的関数 \mathcal{J} は以下の様に表される。

$$\mathcal{J} = \sum_m \sum_{\omega, \tau} \mathcal{D}_{\text{KL}} (A_*^m(\omega, \tau) \parallel \sum_k a_*^m F_*^m(\omega, k) V_*(k, \tau)) + H_*^m(\omega, n) U_*(n, \tau) + \mu \sum_m \sum_{k, n} (\sum_{\omega} F_*^m(\omega, k) H_*^m(\omega, n))^2 \quad (21)$$

第二項は教師基底 $F_*^m(\omega, k)$ とその他の基底 $H_*^m(\omega, n)$ が互いに無相関となる制約を与える直交化罰則項であり、 μ はその重みパラメータである。各変数の更新式は以下の式で与えられる。

$$a_*^m \leftarrow a_*^m \frac{\sum_{\omega, \tau} A_{\omega, \tau}^m Q_{\omega, \tau}^m \sum_k f_{\omega, k}^m v_{k, \tau}^m}{\sum_{\omega, \tau} (\sum_k f_{\omega, k}^m v_{k, \tau}^m)} \quad (22)$$

$$v_{k, \tau}^m \leftarrow v_{k, \tau}^m \frac{\sum_{m, \omega} A_{\omega, \tau}^m Q_{\omega, \tau}^m (a_*^m f_{\omega, k}^m)}{\sum_{m, \tau} (a_*^m \sum_k f_{\omega, k}^m)} \quad (23)$$

$$h_{\omega, n}^m \leftarrow h_{\omega, n}^m \frac{\sum_{\omega, \tau} A_{\omega, \tau}^m Q_{\omega, \tau}^m u_{n, \tau}^m}{\sum_{\tau} u_{n, \tau}^m + \mu \sum_k f_{\omega, k}^m \sum_{\omega'} f_{\omega', k}^m h_{\omega', n}^m} \quad (24)$$

$$u_{n, \tau}^m \leftarrow u_{n, \tau}^m \frac{\sum_{\omega, \tau} A_{\omega, \tau}^m Q_{\omega, \tau}^m h_{\omega, n}^m}{\sum_{m, \tau} h_{\omega, n}^m} \quad (25)$$

$$Q_{\omega, \tau}^m = (a_*^m \sum_k f_{\omega, k}^m v_{k, \tau}^m + \sum_n h_{\omega, n}^m u_{n, \tau}^m)^{-1} \quad (26)$$

ここで、 $A_{\omega, \tau}^m, f_{\omega, k}^m, v_{k, \tau}^m, h_{\omega, n}^m, u_{n, \tau}^m$ はそれぞれ $A_*^m(\omega, \tau), F_*^m(\omega, k), V_*(k, \tau), H_*^m(\omega, n), U_*(n, \tau)$ に対応する要素値である。