

# 非同期分散マイクロフォンアレーによる 伝達関数ゲイン基底 NMF を用いた拡散雑音抑圧\*

村瀬慶和, 千葉大将 (筑波大), 小野順貴 (NII/総研大),  
宮部滋樹, 山田武志, 牧野昭二 (筑波大)

## 1 はじめに

近年、音声通話の品質向上や音声認識精度の向上のために利用される雑音抑圧において、携帯電話やボイスレコーダーなどの非同期録音機器を使用したマイクロホンアレー信号処理への期待が高まりつつある。非同期録音機器を使用できる場合、マイクの増設や配置などの構成が柔軟にできるため、高い SN 比での目的音の收音が期待できる。しかし、非同期録音機器を用いたアレー信号処理では、各機器間のサンプリング周波数の微細なずれによって、デジタル信号処理時に各機器間の位相差が時間とともに変化するため、従来のアレー信号処理では性能が劣化する [1, 2]。そのため、従来のマイクロホンアレー信号処理を適用することを目的とした非同期録音機器に対する同期補正 [3, 4] が研究されているが、計算コストが高い、同期時の誤差にアレー信号処理の性能が左右されるなどの問題がある。そこで、同期誤差に頑健なマイクロホンアレー信号処理として位相情報を使用しない振幅ベースの雑音抑圧手法が提案されている。振幅ベースの雑音抑圧手法は、現在までに非目的音と目的音のパワー比を最大化する振幅スペクトルビームフォーマ [5] や、伝達関数のゲインを基底とする非負値行列因子分解 (NMF: Non-negative Matrix Factorization) (伝達関数ゲイン基底 NMF) による時間周波数マスキング [6, 7, 8] が提案されている。振幅スペクトルビームフォーマは各音源の単一音源区間による学習を必要とするが、伝達関数ゲイン基底 NMF はそのような学習が必ずしも必要ではなく、よりブラインドな雑音抑圧手法となり得る。このため、本研究では伝達関数ゲイン基底 NMF による雑音抑圧に焦点をあてる。

伝達関数ゲイン基底 NMF による雑音抑圧手法は、現在までに罰則付き伝達関数ゲイン基底 NMF [7] や、伝達関数ゲインを事前情報によって学習する教師あり伝達関数ゲイン基底 NMF [8] などが提案されている。しかし、拡散雑音のように不特定多数の音源が含まれる場合には、事前情報として必要な音源数の決定が難しく、また、マイク数が音源数よりも多くなるため伝達関数ゲイン基底 NMF をそのまま使用することが難しい。さらに、教師あり伝達関数ゲイン基底 NMF では、事前情報として観測信号に含まれるすべての音源の単一音源区間が必要である。従って、雑音が常に存在するような環境では目的音の単一音源区間を取得することが難しく、実用的ではない。そこで、本研究では拡散雑音環境下でも雑音抑圧が可能な伝達関数ゲイン基底 NMF の拡散雑音の観測モデルを提案する。さらに、雑音が常に存在する場合に雑音のみを学習する半教師あり伝達関数ゲイン基底 NMF を提案する。

## 2 拡散雑音の観測モデル

### 2.1 複素領域における混合モデル

いま、強調を行う目的音源と  $K$  個の雑音源を  $M$  個のマイクロホンで構成された同期マイクロホンアレーで録音する。このとき、目的音と雑音の成分をそれぞれ、上付き文字  $S$  (Signal)、 $N$  (Noise) で表記すると、時間周波数領域の観測信号は目的音  $X^S(\omega)$  と雑音  $X^N(\omega)$  の足し合わせによって、以下のように表される。

$$X(\omega) = X^S(\omega) + X^N(\omega) \quad (1)$$

ここで、 $X(\omega)$ 、 $X^S(\omega)$ 、 $X^N(\omega)$  は  $M \times N$  の行列で、それぞれ  $m$  行  $n$  列に  $X_{mn}(\omega)$ 、 $X_{mn}^S(\omega)$ 、 $X_{mn}^N(\omega)$  の複素数の要素をもつ。また、 $\omega$  は離散時間フーリエ変換の角周波数、 $N$  は時間フレーム数である。さらに、目的音  $X^S(\omega)$  は、

$$X^S(\omega) = A^S(\omega)S^S(\omega) \quad (2)$$

と表わされる。ここで、 $A^S(\omega)$  は  $M \times 1$  の列ベクトルで、その要素  $A_m^S(\omega)$  は目的音から  $m$  番目のマイクロホンまでの伝達関数を表す。さらに、 $S^S(\omega)$  は  $1 \times N$  の行ベクトルで、その要素  $S_n^S(\omega)$  は、目的音の  $n$  番目のフレームにおける時間周波数成分を表す。また、雑音  $X^N(\omega)$  は  $K$  個の雑音源が目的音と同程度の距離にある場合には、雑音源の足し合わせによって、

$$X^N(\omega) = \sum_{k=1}^K A^{N_k}(\omega)S^{N_k}(\omega) \quad (3)$$

と表すことができる。ここで、 $A^{N_k}(\omega)$  は  $M \times 1$  の列ベクトルで、その要素  $A_m^{N_k}(\omega)$  は  $k$  番目の雑音から  $m$  番目のマイクロホンまでの伝達関数を表す。また、 $S^{N_k}(\omega)$  は  $1 \times N$  の行ベクトルで、その要素  $S_n^{N_k}(\omega)$  は  $k$  番目の雑音の  $n$  番目のフレームにおける時間周波数成分を表す。しかし、 $K$  個の雑音源が遠方から到来し、インパルス応答が長い場合には、フレームごとに式 (3) のように表すことができず、結果として拡散雑音のようにふるまう。

以上のモデルは、同期されたマイクロホンアレーでは成立するが、非同期マイクロホンアレーでは録音機器間のサンプリング周波数の微細なずれによって  $X_{mn}(\omega)$  が、

$$Y_{mn}(\omega) \approx X_{mn}(\omega)\exp(-j\omega\epsilon_m n) \quad (4)$$

のように変化する [9]。ここで、 $j$  は虚数 ( $j = \sqrt{-1}$ ) を表し、 $\epsilon_m$  は  $\epsilon_1 = 0$  として  $m = 1$  番目のチャネルからのドリフトの大きさを表している。このように、非同期マイクロホンアレーでは位相ずれが起こるため、処理が難しくなる。従って、次節では位相情報を用いない混合モデルを示す。

\*Suppressing diffuse noise utilizing asynchronous microphone array by transfer-function-gain NMF, by Yoshikazu MURASE, Hironobu CHIBA (University of Tsukuba), Nobutaka ONO (National Institute of Informatics / The Graduate University for Advanced Studies), Shigeki MIYABE, Takeshi YAMADA, Shoji MAKINO (University of Tsukuba)

## 2.2 振幅領域における混合モデル

時間周波数領域における観測信号の振幅の加法性を仮定することによって、位相情報を用いない混合モデルは以下のように表される。

$$\bar{X}(\omega) \approx \bar{X}^S(\omega) + \bar{X}^N(\omega) \quad (5)$$

ここで、オーバーラインは修飾した行列や要素の絶対値を表す。さらに、振幅領域における目的音  $\bar{X}^S(\omega)$  と雑音  $\bar{X}^N(\omega)$  はそれぞれ、

$$\bar{X}^S(\omega) = \bar{A}^S(\omega)\bar{S}^S(\omega) \quad (6)$$

$$\bar{X}^N(\omega) \approx \sum_{k=1}^K \bar{A}^{N_k}(\omega)\bar{S}^{N_k}(\omega) \quad (7)$$

と表わされる。また、 $\bar{A}^S(\omega)$ 、 $\bar{A}^{N_k}(\omega)$  はそれぞれ目的音と雑音の伝達関数ゲインを表し、 $\bar{S}^S(\omega)$ 、 $\bar{S}^{N_k}(\omega)$  はそれぞれ目的音と雑音の絶対値振幅を表す。

ここで、 $K$  個の雑音が遠方から到来し、かつ音源の位置にばらつきを持った拡散雑音だと仮定すると、一般的に拡散雑音の観測モデルは、マイクロホンの感度が同じ場合に観測の平均的なエネルギーが等しいものになる [10, 11]。このため、振幅領域においては平均的に拡散雑音を一つの音源として近似することができ、そのときの観測信号  $\hat{X}^N(\omega)$ 、伝達関数ゲイン  $\hat{A}^N(\omega)$ 、絶対値振幅  $\hat{S}^N(\omega)$  は以下ようになる。

$$\hat{X}^N(\omega) \approx \hat{A}^N\hat{S}^N \quad (8)$$

$$\hat{A}^N(\omega) \approx \bar{A}^{N_1}(\omega) \approx \dots \approx \bar{A}^{N_K}(\omega) \quad (9)$$

$$\hat{S}^N(\omega) \approx \sum_{k=1}^K \bar{S}^{N_k}(\omega) \quad (10)$$

ここで、 $\hat{A}^N(\omega)$  は  $M \times 1$  の列ベクトルで、その要素  $\hat{A}_m^N(\omega)$  は拡散雑音から  $m$  番目のマイクロホンまでの伝達関数を表す。また、 $\hat{S}^N(\omega)$  は  $1 \times N$  の行ベクトルで、その要素  $\hat{S}_n^N(\omega)$  は拡散雑音の  $n$  番目のフレームにおける絶対値振幅を表す。以上により、拡散雑音環境下における振幅領域の観測モデルは以下ようになる。

$$\bar{X}(\omega) \approx \bar{A}(\omega)\bar{S}(\omega) \quad (11)$$

$$\bar{A}(\omega) \approx [\bar{A}^S(\omega) \quad \hat{A}^N(\omega)] \quad (12)$$

$$\bar{S}(\omega) \approx \begin{bmatrix} \bar{S}^S(\omega) \\ \hat{S}^N(\omega) \end{bmatrix} \quad (13)$$

本稿では、このような観測において拡散雑音を抑制することで目的音を強調する。具体的には、非同期録音機器の柔軟な構成により、 $m = 1$  番目のマイクロホンが目的音の最も近くに置かれ、 $A_1^S(\omega)$  は  $A_j^S(\omega), j = 1, \dots, M$  の中で絶対値が最大になるとする。そして、目的音が最も高い SN 比で観測される  $m = 1$  番目のマイクロホンでの観測信号  $X_{1n}(\omega)$  から雑音を抑制することによって目的音を強調する。

以降の議論ではすべての処理を周波数ビンごとに行うため、周波数を表す記号  $\omega$  は省略する。

## 3 伝達関数ゲイン基底 NMF を用いた拡散雑音抑圧

### 3.1 伝達関数ゲイン基底 NMF による雑音抑圧

本節では、2 章で示した振幅領域における拡散雑音の観測モデルにおいて、Fig. 1 に図示するチャネル

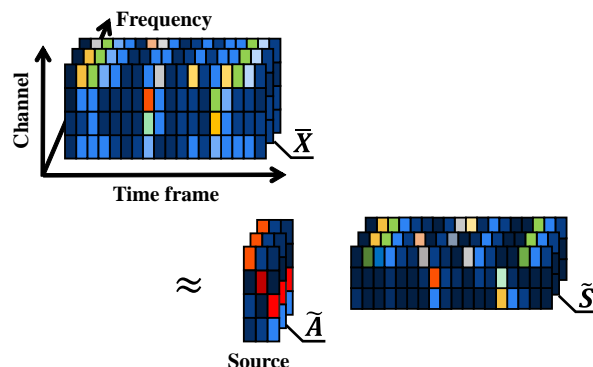


Fig. 1 Channel-time domain representation of observed signals for each frequency bin.

時間領域の NMF を用いて、観測モデルのパラメタを推定することによる雑音抑圧手法について述べる。

NMF は非負行列を 2 つの非負行列の積として

$$\bar{X} \approx \tilde{X} = \tilde{A}\tilde{S} \quad (14)$$

のように低ランク近似する手法である。ここで、NMF で最小化する行列の距離尺度には問題に合わせて様々なものを使用できる。このような低ランク近似は非負の制約により、解がスパースなものに限定される。そのため、式 (9) の観測モデルに対して NMF によるパラメタ推定を行う場合、適切な条件では基底  $\tilde{A}$  が伝達関数ゲイン  $\tilde{A}$  の同定となり、アクティベーション  $\tilde{S}$  が音源の絶対値振幅  $\tilde{S}$  の推定となるような行列分解が得られる。そして、得られた推定解  $\tilde{A}$  と  $\tilde{S}$  を用いて作成したウィナーマスクと、SN 比が最も高い  $m = 1$  番目のマイクロホンで収録した観測信号  $X_{1n}$  によって拡散雑音を抑制した信号  $\tilde{Y}$  は、

$$\tilde{Y} = X_{1n} \frac{(\tilde{A}_1^S \tilde{S}_n^S)^2}{(\tilde{A}_1^S \tilde{S}_n^S)^2 + (\tilde{A}_1^N \tilde{S}_n^N)^2} \quad (15)$$

と表わされる。ここで、ウィナーマスクは観測信号の音源の重ね合わせによるモデル誤差を緩和するために使用する。

また、観測モデルの伝達関数ゲイン行列  $\tilde{A}$  において、目的音の伝達関数ゲインベクトル  $\tilde{A}^S$  は  $A_1^S$  の絶対値が  $A_j^S, j = 1, \dots, M$  の中で最大となる仮定から

$$\tilde{A}_m^S = \begin{cases} 1 - (M-1)\alpha & (m=1) \\ \alpha & (m \neq 1) \end{cases} \quad (16)$$

として、NMF によるパラメタ推定の初期化を行う。ただし、 $\alpha$  は  $0 < \alpha < 1/(M-1)$  となる任意の正の実数である。次に、拡散雑音の伝達関数ゲインベクトル  $\tilde{A}^N$  の初期値はその拡散性から、

$$\tilde{A}_m^N = \frac{1}{M} \quad (m=1, \dots, M) \quad (17)$$

とする。

このように伝達関数ゲイン基底 NMF ではチャネル時間領域において NMF を行うが、音源数とマイク数が近い場合には低ランク近似の拘束力としては弱く、

Table 1 Experimental conditions.

Target source	1
Non-target sources	8
Sampling frequency for synchronous recording	16,000 Hz
Frame length	4096 samples
Frame shift	2048 samples
Signal length for evaluation	10 sec
Signal length for supervised and semi-supervised NMF training	10 sec
Divergence	I-divergence
$\alpha$ (initialization parameter for target signal)	0.25
Number of NMF iterations	200
Reverberation time	0.3 sec

Table 2 Sampling frequencies on each pattern.

	Patt. 1	Patt. 2	Patt. 3
16,000 Hz	Mic 1	Mic 1, 4	Mic 1, 4, 7
16,001 Hz	Mic 2	Mic 2, 5	Mic 2, 5, 8
16,002 Hz	Mic 3	Mic 3, 6	Mic 3, 6, 9

十分な雑音の抑圧性能は期待できない。このため、3.2 ~ 3.4 節では現在提案されている教師あり伝達関数ゲイン基底 NMF と罰則付き伝達関数ゲイン基底 NMF に拡散雑音モデルを適応し、拡散雑音を抑圧する手法と半教師あり伝達関数ゲイン基底 NMF を提案する。

### 3.2 教師あり伝達関数ゲイン基底 NMF

教師あり伝達関数ゲイン基底 NMF は、各音源のみが発話している単一音源区間を用いて伝達関数ゲインベクトルを学習することによって、最適解に近いアクティベーションを推定し、目的音を強調する手法である。まず、本稿では I ダイバージェンス規準の乗法更新式を採用し、

$$\tilde{A}_m^i \leftarrow \tilde{A}_m^i \frac{\sum_n \frac{\tilde{X}_{mn} \tilde{S}_n^i}{\tilde{A}_m^i \tilde{S}_n^i + \tilde{A}_m^N \tilde{S}_n^N}}{\sum_n \tilde{S}_n^i} \quad (i = S, N) \quad (18)$$

$$\tilde{S}_n^i \leftarrow \tilde{S}_n^i \frac{\sum_m \frac{\tilde{X}_{mn} \tilde{A}_m^i}{\tilde{A}_m^i \tilde{S}_n^i + \tilde{A}_m^N \tilde{S}_n^N}}{\sum_m \tilde{A}_m^i} \quad (i = S, N) \quad (19)$$

によって、NMF によるパラメタ推定を行う。そして、目的音と拡散雑音がそれぞれ独立して鳴る単一音源区間の観測信号を用いて、式 (16)、(17) によるパラメタ推定を行うことで  $\tilde{A}^S$  と  $\tilde{A}^N$  を学習する。そして、得られた目的音と拡散雑音の伝達関数ゲインベクトルを結合することによって、伝達関数ゲイン基底  $\tilde{A} = [\tilde{A}^S \ \tilde{A}^N]$  が得られる。その後、学習した基底行列を初期値として、アクティベーション行列のみを式 (19) で更新することによって  $\tilde{S}$  を推定し、式 (15) によって雑音抑圧を行う。

### 3.3 半教師あり伝達関数ゲイン基底 NMF

3.2 節の教師あり伝達関数ゲイン基底 NMF では、目的音と拡散雑音の両方の伝達関数ゲインベクトルに対して学習を行った。しかし、雑音が常に存在する場合には目的音の伝達関数ゲインベクトル  $\tilde{A}^S$  を学習することは難しい。このため、本稿では時間周波数領域において提案されている半教師あり NMF [12] をチャンネル時間領域における NMF に適応した半教師

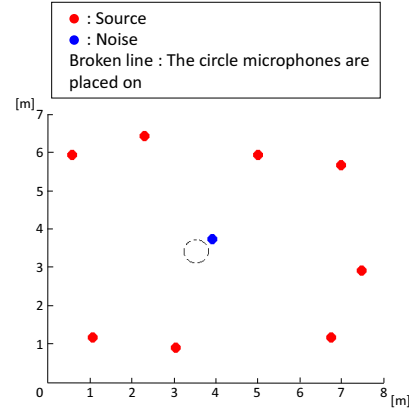


Fig. 2 Arrangement of speakers and microphones.

あり伝達関数ゲイン基底 NMF を提案する。具体的には、拡散雑音の伝達関数ゲインベクトル  $\tilde{A}^N$  のみを単一音源区間によって学習し、式 (18), (19) の更新の際に固定する。

### 3.4 罰則付き伝達関数ゲイン基底 NMF

罰則付き伝達関数ゲイン基底 NMF では、音源のアクティベーションに対してスパースネス制約を導入したもので、その目的関数は、

$$\mathcal{J}(\tilde{X}, \tilde{A}\tilde{S}) = \mathcal{D}_I(\tilde{X}|\tilde{A}\tilde{S}) + \lambda g(\tilde{S}) \quad (20)$$

と表わされる。ここで、 $g(\tilde{S})$  は  $\tilde{S}$  の空間的なスパースネスを測る関数であり、本稿では  $L_{0.5}$  ノルムを用いた。また、 $\lambda$  は非負値の重みを表わし、強調を行う観測信号に合わせて調整を行う。この時の I ダイバージェンス規準における罰則付き乗法型更新式は、

$$\tilde{A}_m^i \leftarrow \tilde{A}_m^i \frac{\frac{\sum_n \tilde{X}_{mn} \tilde{S}_n^i}{\tilde{A}_m^i \tilde{S}_n^i}}{\sum_n \tilde{S}_n^i} \quad (i = S, N) \quad (21)$$

$$\tilde{S}_n^i \leftarrow \tilde{S}_n^i \frac{\sum_m \frac{\tilde{X}_{mn} \tilde{A}_m^i}{\tilde{A}_m^i \tilde{S}_n^i}}{\sum_m \tilde{A}_m^i + \lambda \nabla g(\tilde{S}_n^i)} \quad (i = S, N) \quad (22)$$

となり、更新を行うことによって局所最適解を得る。ここで、 $\nabla g(\tilde{S}_n^i)$  は  $g(\tilde{S}_n^i)$  の勾配を表している。

## 4 評価実験

### 4.1 実験条件

本実験では、提案したモデルを適用した教師あり伝達関数ゲイン基底 NMF と罰則付き伝達関数ゲイン基底 NMF によって、拡散雑音環境下でも雑音抑圧ができるかどうかの確認を行うとともに、半教師あり伝達関数ゲイン基底 NMF によって雑音抑圧が可能かどうかの確認を行う。実験条件を Table 1 に示す。評価に用いる信号は、鏡像法によって生成したインパルス応答を音源信号に畳み込み生成した [13]。また、すべてのマイクロホンは無指向性としてシミュレーションを行い、目的音を強調するための観測信号の目的音と非目的音の SN 比が 0 dB と 5 dB になるような観測信号を作成して実験を行った。また、SN 比の他に、マイクロホンの使用する数を 3, 6, 9 個と変化させて実験を行った。非同期録音データは 16,000 Hz で録音された同期録音データをマイクロホンごとに Table 2 のサンプリング周波数でサンプリングする

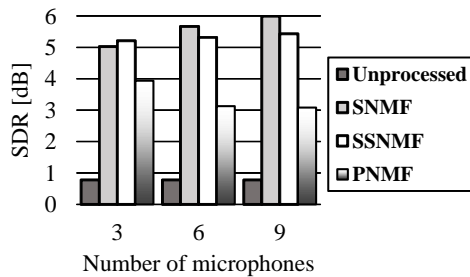


Fig. 3 SDRs of each method on synchronous recording : SNR = 0 dB.

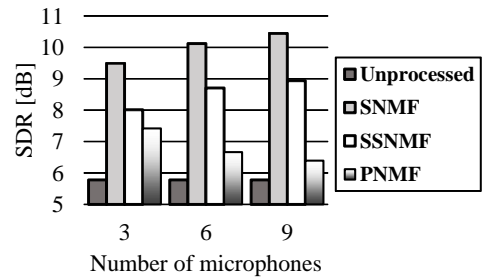


Fig. 5 SDRs of each method on synchronous recording : SNR = 5 dB.

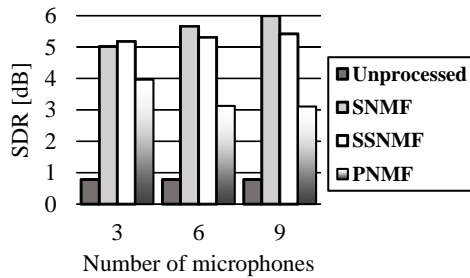


Fig. 4 SDRs of each method on asynchronous recording : SNR = 0 dB.

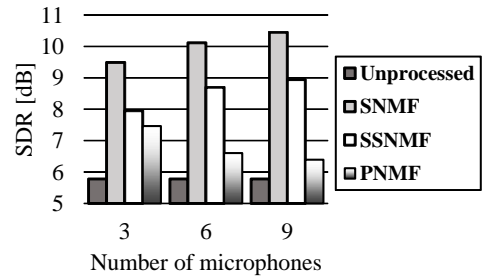


Fig. 6 SDRs of each method on asynchronous recording : SNR = 5 dB.

ことよって作成した。Fig. 2 に音源とマイクロホンの配置を示す。

評価尺度は、Source-to-distortion ratio (SDR) を用いた [14]。これは正答となる元の音源信号と強調信号によって算出され、強調信号の歪みを表し、値が大きいほど雑音抑圧性能が良いことを示す。そして、半教師あり伝達関数ゲイン基底 NMF (SSNMF)、未処理の観測信号 (Unprocessed)、教師あり伝達関数ゲイン基底 NMF (SNMF)、罰則付き伝達関数ゲイン基底 NMF (PNMF) の 4 つの手法に対して評価を行った。ここで、罰則付き伝達関数ゲイン基底 NMF の性能はパラメータを SN 比とマイクロホン数の組み合わせごとに最適値に設定した場合の性能である。

#### 4.2 実験結果

実験結果を Fig. 3, 4, 5, 6 に示す。Fig. 3, 4 は強調するための観測信号の SN 比が 0 dB の同期録音と非同期録音の性能を示し、Fig. 5, 6 は SN 比が 5 dB の性能を示している。これらの結果から、同期録音と非同期録音による雑音抑圧の性能差はほとんど無いことがわかる。また、半教師あり伝達関数ゲイン基底 NMF では、いずれの SN 比やマイク数においても罰則付き伝達関数ゲイン基底 NMF の性能よりも高いことがわかる。さらに、観測信号が 5 dB の場合において、罰則付き伝達関数ゲイン基底 NMF は教師あり伝達関数ゲイン基底 NMF との性能差は大きい、観測信号の SN 比が 0 dB の場合には、教師あり伝達関数ゲイン基底 NMF の性能に近づいていることが分かる。

### 5 おわりに

本研究では拡散雑音環境下でも雑音抑圧が可能な伝達関数ゲイン基底 NMF の観測モデルの提案を行った。また、雑音常在に存在する場合に雑音のみを学習する半教師あり伝達関数ゲイン基底 NMF の提案を行った。結果として、拡散雑音環境下においても教

師あり伝達関数ゲイン基底 NMF、罰則付き伝達関数ゲイン基底 NMF で雑音抑圧を行うことができた。また、半教師あり伝達関数ゲイン基底 NMF では罰則付き伝達関数ゲイン基底 NMF の雑音抑圧性能よりも高い性能を示した。

謝辞 本研究は科学研究費補助金 基盤研究 (B) (25280069) の助成を受けたものである。

#### 参考文献

- [1] Robledo *et al.*, *Proc. WASPAA*, pp. 34-37, 2007.
- [2] Liu *et al.*, *Proc. IWAENC*, 2008.
- [3] Miyabe *et al.*, *Proc. ICASSP*, pp. 674-678, 2013.
- [4] Sakanashi *et al.*, *Proc. APSIPA*, pp. 1-6, 2013.
- [5] Kako *et al.*, *Proc. Acoustic Society of Japan*, pp. 829-830, 2013.
- [6] Lee and Seung, *Proc. NIPS*, vol. 13, pp. 556-562, 2000.
- [7] Togami *et al.*, *Proc. APSIPA*, pp. 522-525, 2010.
- [8] Chiba *et al.*, *Proc. IWAENC*, pp. 204-208, 2014.
- [9] Miyabe *et al.*, *Signal Processing*, vol. 107, pp. 185-196, 2015.
- [10] Itou *et al.*, *Proc. IEEE TASLP*, vol.19, no. 7, pp. 2101-2110, 2011.
- [11] McCowan and Bourlard, *Proc. IEEE TSAP*, vol. 11, no. 6, pp. 709-716, 2003.
- [12] Smaragdakis *et al.*, *Proc. ICA*, pp. 414-421, 2007.
- [13] Habets, Available: [http://home.tiscali.nl/ehabets/rir\\_generator.html](http://home.tiscali.nl/ehabets/rir_generator.html)
- [14] Vincent *et al.*, *Proc. IEEE TASLP*, vol. 14, no. 4, pp. 1462-1469, 2006.