論　　文 43.60.+d

非同期マイクロホンアレーにおける伝達関数ゲイン基底非負値行列因子分解を用いた遠方音源抑圧*

村瀬慶和[*1]　小野順貴[*2,*3]　宮部滋樹[*1]
山田武志[*1]　牧野昭二[*1]

［要旨］ ビームフォーミングなどの従来のアレー信号処理による雑音抑圧手法は，位相情報を活用した指向性制御に基づいており，特定方向から到来する雑音に対しては指向性の零点を向けることで高い効果が得られる。しかし，到来方向が特定できないような，いわゆる背景雑音の抑圧は，一般に難しかった。本論文では，伝達関数ゲイン基底 NMF により，遠方から到来する雑音を複数マイクを用いて効果的に抑圧する手法を提案する。提案手法では，背景雑音が遠方から到来することを仮定し，時間周波数領域における振幅情報のみに着目することで，様々な方向から到来する遠方音源を一つの混合音源としてモデル化する。次にこの振幅の混合モデルを従来提案されている制約付き伝達関数ゲイン基底 NMF に適用し，遠方音源の抑圧を行う。更に，半教師あり伝達関数ゲイン基底 NMF を適用し，遠方音源の抑圧を行う。本手法は振幅情報のみを用いているため，非同期録音機器を用いることができる利点も有する。シミュレーション実験により，様々な方向から到来する複数の遠方音源や移動する遠方音源が，提案手法により効果的に抑圧できることを確認した。

キーワード　遠方音源抑圧，非同期録音機器，非負値行列因子分解
Far-noise suppression, Asynchronous recording, Non-negative matrix factorization

1. はじめに

会話音声や会議録音における音声品質の向上を目的とした雑音抑圧において，ビームフォーミングなどの位相情報を活用した指向性制御に基づく手法が従来提案されてきた [1–4]。このようなアレー信号処理による雑音抑圧手法は，特定方向から到来する雑音に対しては指向性の零点を向けることで高い効果が得られる。しかし，到来方向が特定できないような，いわゆる背景雑音の抑圧は，一般に難しかった。そこで，本論文では，背景雑音が遠方から到来することを仮定し，時間周波数領域における振幅情報のみに着目することで，様々な方向から到来する複数の遠方音源を一つの混合音源としてモデル化する。

振幅領域の雑音抑圧手法は，現在までに，目的音と非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，非負値行列因子分解（NMF：Non-negative Matrix Factorization）[6] を用いて音源の伝達関数ゲインと絶対値振幅を推定する伝達関数ゲイン基底 NMF [7–9] が提案されている。これらの手法は，非同期録音機器と呼ばれる携帯電話や IC レコーダ，ノートパソコンなどに内蔵されている複数のマイクを使用して雑音抑圧を行う枠組み [10–18] において提案されている。非同期録音機器を用いる場合には，低コスト化や，持ち運びが容易などのメリットがある。これらのメリットは，会話音声や会議音声などの使用場所を特定できない場合には重要な要素であることから，本論文では非同期録音機器を用いた振幅領域における背景雑音抑圧に焦点を当てる。

なお，同期アレーを用いた背景雑音除去 [19, 20] が提案されているが，これらは非同期アレーに適用することはできない。一方，非同期アレーを同期させる技術 [17, 18] も提案されているが，[17] では単一音源区間の情報が必要であるし，[18] では音場が定常であるという仮定が成り立つ必要がある。ここでは，非同期アレーに適用できる遠方音源抑圧法を提案する。また，モノラル雑音除去技術 [21, 22] も非同期アレーによる背景雑音除去に適用できるが，ここではマルチチャネル録音を前提としているため，性能向上が期待できるマルチチャネル信号処理について検討する。

---

* Far-noise suppression by transfer-function-gain non-negative matrix factorization in ad hoc microphone array,
by Yoshikazu Murase, Nobutaka Ono, Shigeki Miyabe, Takeshi Yamada and Shoji Makino.

*1 筑波大学
*2 国立情報学研究所
*3 総合研究大学院大学
（問合先：牧野昭二　e-mail: maki@tara.tsukuba.ac.jp）

（2016 年 4 月 27 日受付，2017 年 1 月 11 日採録決定）

振幅スペクトルビームフォーマは，観測に含まれるすべての音源の単一音源区間による学習が必要であるのに対して，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い。そのため，本論文では学習を必要としないブラインドな雑音抑圧手法となり得る伝達関数ゲイン基底 NMF に焦点をあてる。そして，非同期録音機器を用いた伝達関数ゲイン基底 NMF による雑音抑圧に，上述した新たな混合モデルを適用することによって，背景雑音の抑圧を行うことを提案する。更に，背景雑音は常時存在することが多いことに対応するために，雑音の伝達関数ゲインを事前に学習したもので固定し，目的音の伝達関数ゲインとアクティベーションを NMF によりパラメタ推定する半教師あり伝達関数ゲイン基底 NMF を提案する。

本論文の構成は以下のとおりである。第 2 章では，遠方音源の混合モデルについて説明する。第 3 章では，伝達関数ゲイン基底 NMF を用いた遠方音源抑圧について述べる。第 4 章では，評価実験について述べる。第 5 章では，本論文の結論を述べる。

2. 遠方音源の混合モデル

2.1 従来表現による遠方音源の混合モデル

いま，目的音源と背景雑音を構成する K 個の遠方音源を 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，目的音と遠方音源の成分をそれぞれ，上付き文字 $\mathbf{S}$，$\mathbf{N}$ で表記すると，時間周波数領域の観測信号は目的音 $\boldsymbol{X}^{\mathbf{S}}(\omega)$ と遠方音源 $\boldsymbol{X}^{\mathbf{N}}(\omega)$ の足し合わせによって，以下のように表される。

$$\boldsymbol{X}(\omega) = \boldsymbol{X}^{\mathbf{S}}(\omega) + \boldsymbol{X}^{\mathbf{N}}(\omega) \tag{1}$$

ここで，$\boldsymbol{X}(\omega)$，$\boldsymbol{X}^{\mathbf{S}}(\omega)$，$\boldsymbol{X}^{\mathbf{N}}(\omega)$ は $M \times N$ の行列で，それぞれ m 行 n 列に $x_{mn}(\omega)$，$x^{\mathbf{S}}_{mn}(\omega)$，$x^{\mathbf{N}}_{mn}(\omega)$ の複素数の要素を持つ。また，ω 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。更に，目的音 $\boldsymbol{X}^{\mathbf{S}}(\omega)$ は，

$$\boldsymbol{X}^{\mathbf{S}}(\omega) = \boldsymbol{a}^{\mathbf{S}}(\omega)\boldsymbol{s}^{\mathbf{S}}(\omega) \tag{2}$$

と表される。ここで，$\boldsymbol{a}^{\mathbf{S}}(\omega)$ は $M \times 1$ の列ベクトルで，その要素 $a^{\mathbf{S}}_m(\omega)$ 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。更に，$\boldsymbol{s}^{\mathbf{S}}(\omega)$ は $1 \times N$ の行ベクトルで，その要素 $s^{\mathbf{S}}_n(\omega)$ 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，遠方音源 $\boldsymbol{X}^{\mathbf{N}}(\omega)$ は K 個の遠方音源の足し合わせによって，

$$\boldsymbol{X}^{\mathbf{N}}(\omega) = \sum_{k=1}^{K} \boldsymbol{a}^{\mathbf{N}_k}(\omega)\boldsymbol{s}^{\mathbf{N}_k}(\omega) \tag{3}$$

と表すことができる。ここで，$\boldsymbol{a}^{\mathbf{N}_k}(\omega)$ は $M \times 1$ の列ベクトルで，その要素 $a^{\mathbf{N}_k}_m(\omega)$ は k 番目の遠方音源から m 番目のマイクロホンまでの伝達関数を表す。また，$\boldsymbol{s}^{\mathbf{N}_k}(\omega)$ は $1 \times N$ の行ベクトルで，その要素 $s^{\mathbf{N}_k}_n(\omega)$ は k 番目の遠方音源の n 番目のフレームにおける時間周波数成分を表す。

以上のような複素領域の観測において，位相差を利用したビームフォーミングは，目的音源の位置が変動する場合や，遠方音源と同じ方向にある場合には性能が低下するため，背景雑音環境下では性能が期待できない。また，非同期録音機器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，

$$y_{mn}(\omega) \approx x_{mn}(\omega)\exp(-\jmath\omega\epsilon_m n) \tag{4}$$

のように変化する [18]。ここで，$\jmath$ は虚数 ($\jmath = \sqrt{-1}$) を表し，ϵ_m は $\epsilon_1 = 0$ として 1 番目（$m = 1$）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，非同期録音機器では位相ずれが起こるため，処理が難しくなる。以上のことから，非同期録音機器を用いた背景雑音抑圧では振幅領域における混合モデルを用いる。

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。

$$|\boldsymbol{X}(\omega)| \approx |\boldsymbol{X}^{\mathbf{S}}(\omega)| + |\boldsymbol{X}^{\mathbf{N}}(\omega)| \tag{5}$$

更に，振幅領域における目的音源 $|\boldsymbol{X}^{\mathbf{S}}(\omega)|$ と遠方音源 $|\boldsymbol{X}^{\mathbf{N}}(\omega)|$ はそれぞれ，

$$|\boldsymbol{X}^{\mathbf{S}}(\omega)| = |\boldsymbol{a}^{\mathbf{S}}(\omega)||\boldsymbol{s}^{\mathbf{S}}(\omega)| \tag{6}$$

$$|\boldsymbol{X}^{\mathbf{N}}(\omega)| \approx \sum_{k=1}^{K} |\boldsymbol{a}^{\mathbf{N}_k}(\omega)||\boldsymbol{s}^{\mathbf{N}_k}(\omega)| \tag{7}$$

と表される。ここで，$|\boldsymbol{a}^{\mathbf{S}}(\omega)|$ と $|\boldsymbol{a}^{\mathbf{N}_k}(\omega)|$ はそれぞれ目的音源と遠方音源の伝達関数ゲインを表し，$|\boldsymbol{s}^{\mathbf{S}}(\omega)|$ と $|\boldsymbol{s}^{\mathbf{N}_k}(\omega)|$ はそれぞれ，目的音源と遠方音源の絶対値振幅を表している。

このような振幅領域における混合モデルを従来の伝達関数ゲイン基底 NMF [7–9] に適用する場合には，NMF の初期行列の設定の際に音源数 K が既知である必要があるが，無数の遠方音源からなる背景雑音の場合には遠方音源の数 K は未知となり，雑音抑圧を行うことができない。このため，次節では無数の遠方音源からなる背景雑音を抑圧するための混合モデルを提案する。

2.2 無数の遠方音源抑圧のための混合モデル

いま，2.1 節のような観測において点音源 k における $m = 1$ 番目のマイクと各マイクの音圧レベル差は，

$$L_1 - L_m = 10\log\frac{P_1^2}{P_0^2} - 10\log\frac{P_m^2}{P_0^2}\ (m=2,\ldots,M) \quad (8)$$

と表すことができる。ここで，P_m，L_m は点音源 k における各マイクの音圧と音圧レベルを表している。また，P_0 は人間の最小可聴音圧である基準値 $P_0 = 20\times10^{-6}\,\mathrm{Pa}$ を表している。このとき，m 番目のマイクの音圧を $m=1$ 番目のマイクの音圧 P_1 を用いて表現すると，

$$P_m^2 = \left(\frac{\beta_m}{\beta_1}\right)\times\left(\frac{r_m^2}{r_1^2}\right)^{-1}\times P_1^2\ (m=2,\ldots,M) \quad (9)$$

と表すことができる。ここで，β_m，r_m はそれぞれ m 番目のマイクのマイク感度と点音源 k までの距離を表している。式 (8)，(9) より，マイク間の音圧レベル差は，

$$L_1 - L_m = -10\log\frac{\beta_m}{\beta_1} - 20\log\frac{r_1}{r_m}\ (m=2,\ldots,M) \quad (10)$$

と表すことができる。ここで，点音源 k がマイク間隔に対して十分遠方から到来すると仮定すると，$r_1 \approx r_m$ となり，マイク間のゲイン差は音源の音量や方向に依存せず，マイクの感度にのみ依存していることが分かる。つまり，すべての遠方音源で伝達関数ゲイン $|\boldsymbol{a}^{\mathbf{N}_k}|$ が類似する。

そこで，背景雑音を構成するすべての遠方音源の伝達関数ゲインを，代表とする伝達関数ゲインで近似し，遠方から到来するすべての音源を一つの遠方音源として近似する。このとき，振幅領域における遠方音源の観測信号 $|\hat{\boldsymbol{X}}^{\mathbf{N}}(\omega)|$，伝達関数ゲイン $|\hat{\boldsymbol{a}}^{\mathbf{N}}(\omega)|$，絶対値振幅 $|\hat{\boldsymbol{s}}^{\mathbf{N}}(\omega)|$ は以下のように表される。

$$|\hat{\boldsymbol{X}}^{\mathbf{N}}(\omega)| \approx |\hat{\boldsymbol{a}}^{\mathbf{N}}(\omega)||\hat{\boldsymbol{s}}^{\mathbf{N}}(\omega)| \quad (11)$$

$$|\hat{\boldsymbol{a}}^{\mathbf{N}}(\omega)| \approx |\boldsymbol{a}^{\mathbf{N}_1}(\omega)| \approx \ldots \approx |\boldsymbol{a}^{\mathbf{N}_K}(\omega)| \quad (12)$$

$$|\hat{\boldsymbol{s}}^{\mathbf{N}}(\omega)| \approx \sum_{k=1}^{K} |\boldsymbol{s}^{\mathbf{N}_k}(\omega)| \quad (13)$$

ここで，$|\hat{\boldsymbol{a}}^{\mathbf{N}}(\omega)|$ は $M\times1$ の列ベクトルで，その要素 $|\hat{a}_m^{\mathbf{N}}(\omega)|$ は遠方音源から m 番目のマイクロホンまでの伝達関数ゲインを表す。また，$|\hat{\boldsymbol{s}}^{\mathbf{N}}(\omega)|$ は $1\times N$ の行ベクトルで，その要素 $|\hat{s}_n^{\mathbf{N}}(\omega)|$ は n 番目のフレームにおける遠方音源の絶対値振幅の総和を表している。以上により，遠方音源が存在する環境下における振幅領域の混合モデルは以下のようになる。

$$|\boldsymbol{X}(\omega)| \approx |\boldsymbol{A}(\omega)||\boldsymbol{S}(\omega)| \quad (14)$$

$$|\boldsymbol{A}(\omega)| \approx \left[|\boldsymbol{a}^{\mathbf{S}}(\omega)| \quad |\hat{\boldsymbol{a}}^{\mathbf{N}}(\omega)|\right] \quad (15)$$

$$|\boldsymbol{S}(\omega)| \approx \begin{bmatrix}|\boldsymbol{s}^{\mathbf{S}}(\omega)|\\ |\hat{\boldsymbol{s}}^{\mathbf{N}}(\omega)|\end{bmatrix} \quad (16)$$

以上のモデルによって，NMF に必要な音源数の決定を可能にすると共に，NMF による低ランク近似が可能になる。本論文では，このような混合モデルにおいて遠方音源を抑圧する。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，$a_1^{\mathbf{S}}(\omega)$ は $a_j^{\mathbf{S}}(\omega), j=1,\ldots,M$ 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 $x_{1n}(\omega)$ に対して時間周波数マスクを作成することによって遠方音源を抑圧する。

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 ω は省略する。

3. 伝達関数ゲイン基底 NMF を用いた遠方音源抑圧

3.1 NMF による遠方音源抑圧

本節では，モデルのパラメタを NMF を用いて推定することによって雑音抑圧を行う手法 [7–9] を 2 章で提案した振幅領域における遠方音源の混合モデルに適用することによって遠方音源を抑圧する手順について述べる。ここで，**Fig. 1** は NMF を用いて推定する混合モデルを示しており，音声・音響信号処理で一般的なスペクトルパターンとアクティベーションへの分解 [23–25] とは異なり，周波数ビンごとのチャネル時間スペクトルを伝達関数ゲインと絶対値振幅に分解する。

NMF は非負値行列を二つの非負値行列の積として

$$|\boldsymbol{X}| \approx \tilde{\boldsymbol{X}} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{S}} \quad (17)$$

のように低ランク近似する手法である [6]。ここで，$\tilde{\boldsymbol{X}}$，$\tilde{\boldsymbol{A}}$，$\tilde{\boldsymbol{S}}$ は観測に対する推定解を表している。このよう

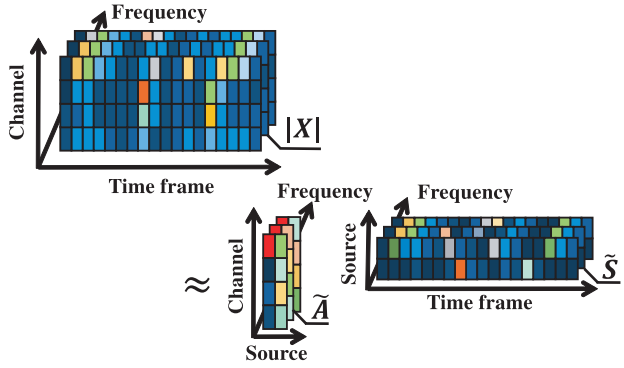


Fig. 1 Channel-time domain representation of observed signals for each frequency bin.

な低ランク近似を行うためには，元の行列と推定した行列の距離を最小化すれば良いが，直接最小化することは困難である。このため，NMF では補助関数法によって得られた更新式を用いることによって低ランク近似を行う。また，NMF で最小化する行列の距離尺度には様々なものを用いることができ，問題にあわせて適切なものを選択する。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，適切な条件のもとでは基底 $\tilde{\boldsymbol{A}}$ が伝達関数ゲイン $|\boldsymbol{A}|$ の同定となり，アクティベーション $\tilde{\boldsymbol{S}}$ が音源の絶対値振幅 $|\boldsymbol{S}|$ の推定となるような行列分解が得られる。

本論文では，距離尺度として I ダイバージェンス規準を採用する。このとき，更新式は以下のように表される。

$$\tilde{a}_m^i \leftarrow \tilde{a}_m^i \frac{\sum_n \frac{|x_{mn}|\tilde{s}_n^i}{\tilde{a}_m^{\mathbf{S}}\tilde{s}_n^{\mathbf{S}}+\tilde{a}_m^{\mathbf{N}}\tilde{s}_n^{\mathbf{N}}}}{\sum_n \tilde{s}_n^i} \quad (i=\mathbf{S},\mathbf{N}) \tag{18}$$

$$\tilde{s}_n^i \leftarrow \tilde{s}_n^i \frac{\sum_m \frac{|x_{mn}|\tilde{a}_m^i}{\tilde{a}_m^{\mathbf{S}}\tilde{s}_n^{\mathbf{S}}+\tilde{a}_m^{\mathbf{N}}\tilde{s}_n^{\mathbf{N}}}}{\sum_m \tilde{a}_m^i} \quad (i=\mathbf{S},\mathbf{N}) \tag{19}$$

また，式 (18)，(19) の更新ごとに，

$$\tilde{a}_m^i \leftarrow \frac{\tilde{a}_m^i}{\tilde{a}_m^{\mathbf{S}}+\tilde{a}_m^{\mathbf{N}}} \quad (i=\mathbf{S},\mathbf{N}) \tag{20}$$

$$\tilde{s}_n^i \leftarrow \left(\tilde{a}_m^{\mathbf{S}}+\tilde{a}_m^{\mathbf{N}}\right)\tilde{s}_n^i \quad (i=\mathbf{S},\mathbf{N}) \tag{21}$$

として正規化を行う。

更に，NMF では周波数領域 ICA などの従来の周波数独立の音源分離手法 [26, 27] と同様に，それぞれの周波数ビンにおいて分離信号の各周波数成分が異なる順番で現れるというパーミュテーション問題が発生する。そこで，$a_1^{\mathbf{S}}$ の絶対値が $a_j^{\mathbf{S}}, j=1,\ldots,M$ の中で最大となる仮定から基底の初期値を

$$\tilde{a}_m^{\mathbf{S}} = \begin{cases} 1-(1-M)\alpha & (m=1) \\ \alpha & (m\neq 1) \end{cases} \tag{22}$$

とすることでパーミュテーション問題を解決する。ここで，α は非目的音に対する初期値であり，$0<\alpha<1/(M-1)$ となる任意の正の実数である。また，遠方音源の基底 $\tilde{\boldsymbol{a}}^{\mathbf{N}}$ は，

$$\tilde{a}_m^{\mathbf{N}} = \frac{1}{M}\ (m=1,\ldots,M) \tag{23}$$

として，初期化を行う。

目的音源を強調した信号 y_n は SN 比が最も高い $m=1$ 番目のマイクで収音した観測信号 x_{1n} と目的音源以外を抑圧するウィーナマスクとの積として，

$$y_n = x_{1n}\frac{\left(\tilde{a}_1^{\mathbf{S}}\tilde{s}_n^{\mathbf{S}}\right)^2}{(\tilde{a}_1^{\mathbf{S}}\tilde{s}_n^{\mathbf{S}})^2+(\tilde{a}_1^{\mathbf{N}}\tilde{s}_n^{\mathbf{N}})^2} \tag{24}$$

と表される。ここで，ウィーナ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。最後に，y_n を離散時間フーリエ逆変換することによって時間領域における目的音源の強調信号を求める。

3.2 制約付き伝達関数ゲイン基底 NMF と半教師あり伝達関数ゲイン基底 NMF

従来の伝達関数ゲイン基底 NMF において，3.1 節のような単純な NMF によるパラメタ推定では，音源数とマイク数が近い値の場合に低ランク近似の拘束力としては弱く，十分な雑音の抑圧性能は期待できないことが確認されている [28]。更に，観測信号のスパース性が十分に保たれない場合には，NMF の最適解が任意性を持ってしまい，雑音抑圧に有用な基底とアクティベーションが得られないということが，多重音解析の分野で議論されている [29]。その解決法として，アクティベーション行列をスパースにするための罰則項を導入した罰則付き伝達関数ゲイン基底 NMF [7] や，基底行列を各音源の単一音源区間によって学習した教師あり伝達関数ゲイン基底 NMF [8, 9] などの制約をつけた手法が提案されている。本節では，それらの制約付き伝達関数ゲイン基底 NMF の特徴について述べる。また，半教師あり伝達関数ゲイン基底 NMF を提案する。

罰則付き伝達関数ゲイン基底 NMF は，事前情報を必要としないブラインドな雑音抑圧手法 [7] である。具体的には，以下のようにアクティベーション行列にスパースネス制約を導入した雑音抑圧手法である。

$$\mathcal{J}(|\boldsymbol{X}|,\tilde{\boldsymbol{A}}\tilde{\boldsymbol{S}}) = \mathcal{D}(|\boldsymbol{X}||\tilde{\boldsymbol{A}}\tilde{\boldsymbol{S}}) + \lambda g(\tilde{\boldsymbol{S}}) \tag{25}$$

ここで，$g(\tilde{\boldsymbol{S}})$ は時間フレームごとに音源のアクティベーション $\tilde{\boldsymbol{S}}$ に対するスパースネス制約を評価する関数であり，λ はその重みを表している。また，距離尺度として I ダイバージェンス，スパースネス制約の評価関数として $L_{0.5}$ ノルムを採用した場合の更新式は以下のように表される。

$$\tilde{A}_m^i \leftarrow \tilde{A}_m^i \frac{\frac{\Sigma_n|X_{mn}|\tilde{S}_n^i}{\tilde{A}_m^i\tilde{S}_n^i}}{\Sigma_n\tilde{S}_n^i} \quad (i=\mathbf{S},\mathbf{N}) \tag{26}$$

$$\tilde{S}_n^i \leftarrow \tilde{S}_n^i \frac{\Sigma_m\frac{|X_{mn}|\tilde{A}_k^i}{\tilde{A}_m^i\tilde{S}_n^i}}{\Sigma_m\tilde{A}_m^i+\lambda\nabla g(\tilde{S}_n^i)} \quad (i=\mathbf{S},\mathbf{N}) \tag{27}$$

ここで，$\nabla g(\tilde{S}_n^i)$ は式 (25) における $g(\tilde{S}_n^i)$ の勾配を表している。罰則付き伝達関数ゲイン基底 NMF はこの更新式を用いて目的音源の伝達関数ゲインと絶対値振幅を推定し，3.1 節と同様にウィーナマスクの作成を行うことによって雑音抑圧を行う。

教師あり伝達関数ゲイン基底 NMF は伝達関数のゲ

インを事前に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する手法である [8, 9]。以下にその手順を説明する。まず，目的音源と遠方音源の単一音源区間を録音した学習信号 $\boldsymbol{X}^{\mathbf{S}}$，$\boldsymbol{X}^{\mathbf{N}}$ を取得する。次に，その学習信号に対して NMF による推定を行うことによって目的音源と遠方音源の伝達関数ゲインベクトル $\boldsymbol{a}^{\mathbf{S}}$，$\boldsymbol{a}^{\mathbf{N}}$ を得る。そして，雑音抑圧を行う観測信号に対して式 (18)，(19) の更新を行う際に，$\tilde{\boldsymbol{A}} = [\boldsymbol{a}^{\mathbf{S}}\boldsymbol{a}^{\mathbf{N}}]$ として基底行列を固定し，式 (19) のみを更新することによって，最適解に近い目的音源の絶対値振幅を推定する。推定後は 3.1 節と同様に，ウィーナマスクの作成を行うことによって雑音抑圧を行う。

上記のように教師あり伝達関数ゲイン基底 NMF では，目的音源と遠方音源の伝達関数ゲインベクトルに対して学習を行う。しかし，遠方音源が常時存在する場合には，目的音源の学習信号 $\boldsymbol{X}^{\mathbf{S}}$ を取得することは難しい。一方で，遠方音源の学習信号 $\boldsymbol{X}^{\mathbf{N}}$ を取得することは比較的容易である。このため，本論文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 [30] を時間チャネル領域における NMF に適応した半教師あり伝達関数ゲイン基底 NMF を提案する。まず，遠方音源の単一音源区間を録音した学習信号 $\boldsymbol{X}^{\mathbf{N}}$ を取得する。その後，学習信号に対して NMF による推定を行うことによって遠方音源の伝達関数ゲインベクトル $\boldsymbol{a}^{\mathbf{N}}$ を得る。そして，遠方音源の伝達関数ゲインベクトルのみを式 (18), (19) の更新の際に固定することによって，真値に近い目的音の伝達関数ゲインベクトルと絶対値振幅を推定する。

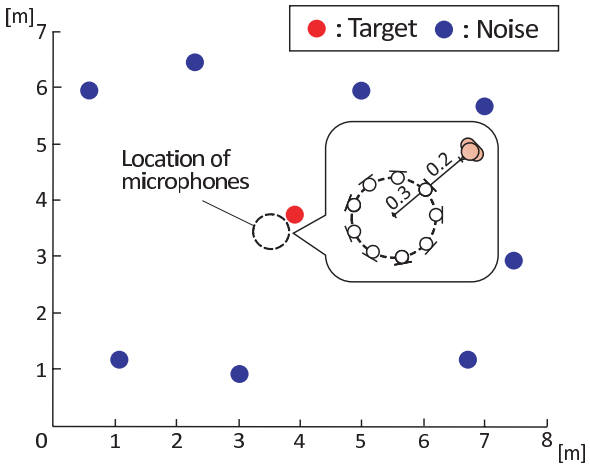


Fig. 2 Experiment 1: Arrangement of a target speaker, multiple far noises, and microphones.

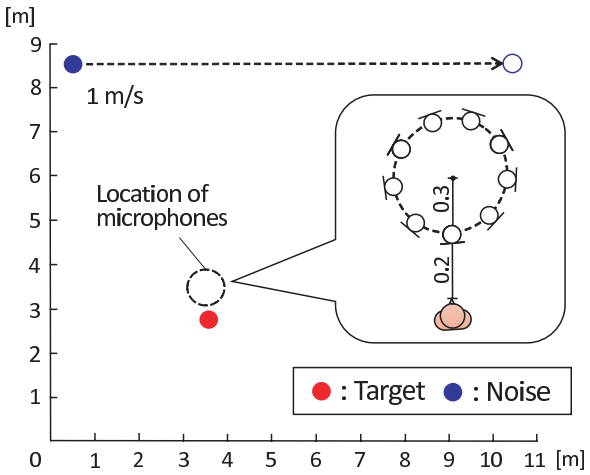


Fig. 3 Experiment 2: Arrangement of a target speaker, a moving noise, and microphones.

4. 評 価 実 験

4.1 実 験 条 件

提案した遠方音源の混合モデルと半教師あり伝達関数ゲイン基底 NMF によって，遠方音源が存在する環境下で雑音抑圧ができるかどうかの確認を行う。

遠方音源を二つの種類に分けて実験を行った。マイクと音源の配置を **Fig. 2**，**Fig. 3** に示す。実験 1 では Fig. 2 のように複数の静止した遠方音源を抑圧できるかどうか確認を行う。また，実験 2 では Fig. 3 のように一つの移動する遠方音源を抑圧できるかどうかの確認を行う。主な実験条件を **Table 1**，**Table 2** 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 [31, 32] を音源信号に畳み込み生成した。ここで，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リサンプリングすることによって作成した。マイクは図の破線上に等間隔に並ぶ。このとき，Mic 1 はウィーナマスクをかける観測信号を録音するためのマイクであり，目的音の前に固定され，すべてのパターンで共通の位置に配置される。なお，伝達関数ゲイン基底 NMF によるパラメタ推定において，同期録音データと非同期録音データの性能差がとても小さいことが確認されているため [8, 9]，本実験では非同期録音データのみを対象とした。

評価尺度は，Signal-to-distortion ratio（SDR）と Source-to-interference ratio（SIR）を用いた [33]。これら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る。SDR は強調信号のひずみ，SIR は非目的信号の抑圧率を表し，ともに値が大きいほど抑圧性能が良いことを示す。評価値は，提案した混合モデルを適用した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），半教師あり伝達関数ゲイン基底 NMF（SSNMF）に対して算出する。更に，比較のために，未処理の観測信号（Unprocessed），SN 比最大化ビームフォーマ（mSNRbf），目的音の学習に雑音を含んだ観測信号を使用した SN

Table 1 Experimental conditions.

Sampling frequency for synchronousrecording	16,000 Hz
Dry-sources of target and noise	PASL-DSR
Frame length	4,096 samples
Frame shift	2,048 samples
Signal length for evaluation	10 s
Signal length for pre-training	10 s
Divergence	I-divergence
α (initialization parameter)	0.1
Number of NMF iterations	200
Reverberation time	0.3 s
Number of microphones	3, 6, 9 ch
Signal to far noise ratio	0, 5 dB

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

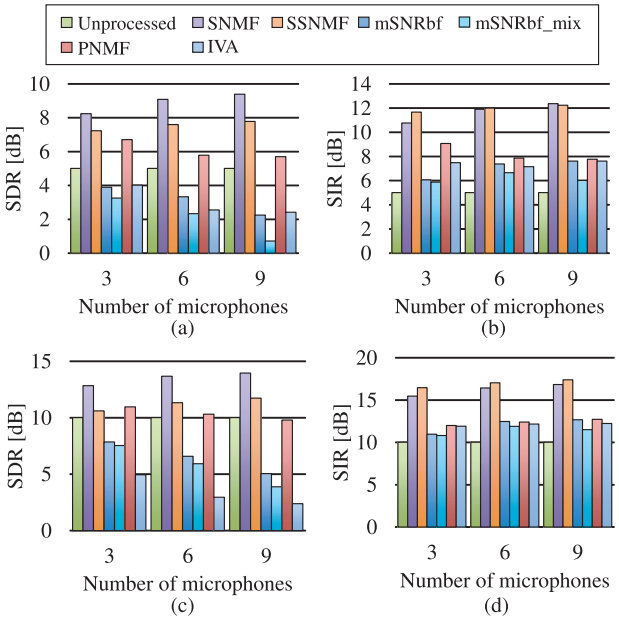


Fig. 4 Experiment 1: SDRs in (a) SNR = 5 dB and (b) SNR = 10 dB, SIRs in (c) SNR = 5 dB and (d) SNR = 10 dB with asynchronous recording.

比最大化ビームフォーマ（mSNRbf_mix），独立ベクトル分析（IVA）に対しても評価値を算出した。ここで，SNMF は，遠方音源の単一音源区間に加えて，目的音の単一音源区間も取得できたと仮定した理想性能を示している。SSNMF は，遠方音源の伝達関数ゲインのみを遠方音源の単一音源区間によって学習している。また，mSNRbf_mix は SSNMF との比較のために用いられ，目的音の学習信号には遠方音源を含んだ混合信号を，遠方音源の学習信号には，遠方音源の単一音源区間を使用している。PNMF における罰則の重み調整は，文献 [34] を参考に 0.01 から 0.15 の間で 0.01 刻みで SDR が最も高くなる値を採用している。また，ブラインドな手法である IVA は PNMF との比較のために用いられる。また，観測における抑圧区間と学習区間を変更した計 6 回の平均を算出している。

4.2 実験結果

実験結果を **Fig. 4**, **Fig. 5** に示す。Fig. 4 に，実験 1 における SN 比が 5 dB と 10 dB の場合の複数の遠方音源を抑圧した結果を示す。まず, SSNMF ではいずれの条件でもマイク数が増加するに従って SDR 値と SIR 値が向上している。一方で, mSNRbf と mSNRbf_mix では，SDR 値がいずれの条件でも未処理時の性能よりも低下し，マイク数の増加に伴って SDR 値が低下することを確認した。また，SIR 値はマイク数の増加に伴って向上が見られたが，SSNMF の SIR 値よりも低いことを確認した。次に，PNMF はマイク数の増加に伴って SDR 値が低下することを確認した。IVA は，いずれの条件においても SDR 値，SIR 値ともに

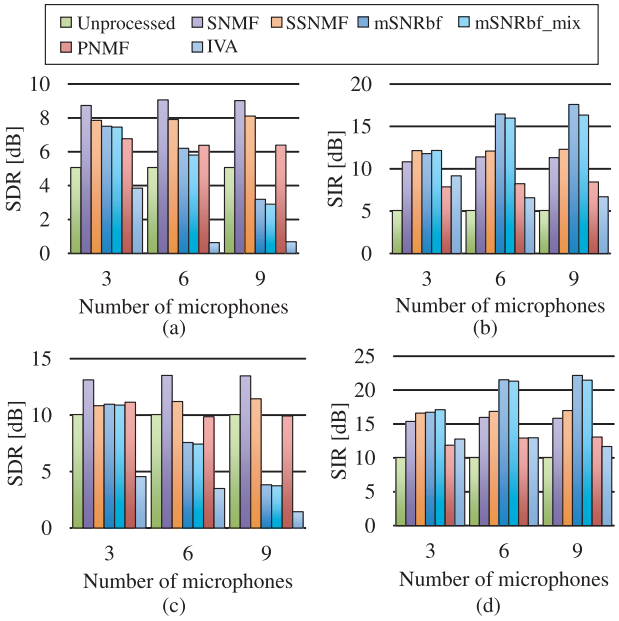


Fig. 5 Experiment 2: SDRs in (a) SNR = 5 dB and (b) SNR = 10 dB, SIRs in (c) SNR = 5 dB and (d) SNR = 10 dB with asynchronous recording.

PNMF よりも低い値を示し，マイク数の増加に伴って SDR 値が低下することを確認した。

Fig. 5 に，実験 2 における SN 比が 5 dB と 10 dB の場合の遠方から到来する一つの移動音源を抑圧した結果を示す。まず, SSNMF は実験 1 と同様に，いずれの条件でもマイク数が増加するに従って SDR 値と SIR 値が向上している。一方で, mSNRbf と mSNRbf_mix では，SDR 値がマイク数の増加に伴って低下し，SIR 値は向上していることを確認した。次に，PNMF はマ

イク数の増加に伴ってSDR値が低下したが，SIR値は向上することを確認した。IVAは，いずれの条件においてもSDR値，SIR値ともにPNMFよりも低い値を示し，マイク数の増加に伴ってSDR値が低下することを確認した。

この結果から，位相情報を用いるSN比最大化ビームフォーマや，独立ベクトル分析は，非同期録音機器の場合には，マイク数の増加に伴ってSDR値が低下することを確認した。これは，マイク数が増えるに従って，観測の位相ずれが大きくなるためだと考えられる。一方で，伝達関数ゲイン基底NMFではマイク数の増加に伴って，SDRとSIRが向上することを確認した。これは，マイク数の増加に伴って，アクティベーションの任意性が制限されるため，NMFによるパラメタ推定が最適解に近い基底とアクティベーションを推定したためだと考えられる。

5. おわりに

本論文では，背景雑音を抑圧するために，振幅領域における音源の遠方性を仮定した混合モデルの提案を行った。そして，非同期録音機器を使用可能な雑音抑圧手法である伝達関数ゲイン基底NMFに適用することによって，遠方音源を抑圧できるかどうかの検討を行った。その結果，遠方音源の混合モデルを適用した制約付き伝達関数ゲイン基底NMFで遠方音源を抑圧することができた。また，半教師あり伝達関数ゲイン基底NMFでは，遠方音源のみを事前に学習することによって，実用の範囲内で高い雑音抑圧性能を確認した。

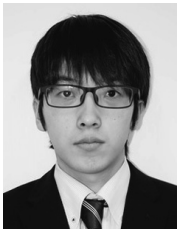
謝　辞

本論文は，科学研究費補助金基盤研究(B)(25280069)，基盤研究（A)(16H01735)，及びセコム科学技術振興財団の支援を受けた。

文　献

[1] J. L. Flanagan, D. A. Berkley, G. W. Elko and W. M. M. Sondhi, “Autodirective microphone systems,” *Acustica*, **73**, 58–71 (1991).
[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*: *Concepts and Techniques* (Prentice Hall, Englewood Cliffs, 1993).
[3] 金田　豊, “騒音下音声認識のためのマイクロホンアレー技術,” 音響学会誌, **53**, 872–876 (1997).
[4] 宝珠山治, 杉山昭彦, “音響ビームフォーミングと多次元信号処理,” 信学技報, CAS97-98 (1998).
[5] T. Kako, K. Kobayashi and H. Ohmuro, “A proposal of amplitude-spectrum beamformer for an asynchronous distributed microphone array,” *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 829–830 (2013) (in Japanese).
[6] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. NIPS*, pp. 556–562 (2000).
[7] M. Togami, Y. Kawaguchi, H. Kokubo and Y. Obuchi, “Acoustic echo suppressor with multichannel semi-blind non-negative matrix factorization,” *Proc. APSIPA*, pp. 522–525 (2010).
[8] H. Chiba, N. Ono, S. Miyabe, Y. Takahashi, T. Yamada and S. Makino, “Amplitude-based speech enhancement with nonnegative matrix factorization for asynchronous distributed recording,” *Proc. IWAENC*, pp. 204–208 (2014).
[9] 千葉大将, 小野順貴, 宮部滋樹, 高橋　祐, 山田武志, 牧野昭二, “アドホックマイクロホンアレーにおける時間チャネル領域での非負値行列因子分解を用いた振幅ベースの音声強調,” 音響学会誌, **72**, 462–470 (2016).
[10] 小野順貴, 宮部滋樹, 牧野昭二, “非同期分散マイクロホンアレイに基づく音響信号処理,” 音響学会誌, **70**, 391–396 (2014).
[11] 小野順貴, T.-K. Le, 宮部滋樹, 牧野昭二, “アドホックマイクロホンアレー―複数のモバイル録音機器で行う音響信号処理―,” 信学会 *Fundam. Rev.*, **7**, 336–347 (2014).
[12] M. Souden, K. Kinoshita, M. Delcroix and T. Nakatani, “Distributed microphone array processing for speech source separation with classifier fusion,” *Proc. MLSP*, pp. 1–6 (2012).
[13] A. Bertrand, “Applications and trends in wireless acoustic sensor networks: A signal processing perspective,” *Proc. SCVT*, pp. 1–6 (2011).
[14] E. Robledo-Arnuncio, T. S. Wada and B.-H. Juang, “On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation,” *Proc. WASPAA*, pp. 34–37 (2007).
[15] Z. Liu, “Sound source separation with distributed microphone arrays in the presence of clock synchronization errors,” *Proc. IWAENC*, pp. 1–4 (2008).
[16] S. Miyabe, N. Ono and S. Makino, “Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain,” *Proc. ICASSP*, pp. 674–678 (2013).
[17] R. Sakanashi, N. Ono, S. Miyabe, T. Yamada and S. Makino, “Speech enhancement with ad-hoc microphone array using single source activity,” *Proc. APSIPA*, pp. 1–6 (2013).
[18] S. Miyabe, N. Ono and S. Makino, “Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation,” *Signal Process.*, **107**, 185–196 (2015).
[19] H. Sawada, S. Araki, R. Mukai and S. Makino, “Blind extraction of a dominant source signal from mixtures of many sources,” *Proc. ICASSP*, pp. 61–64 (2005).
[20] T. Yoshioka, N. Ito, M. Delcroix, A. Ogawa, K. Kinoshita, M. Fujimoto, C. Yu, W. J. Fabian, M. Espi, T. Higuchi, S. Araki and T. Nakatani, “The NTT CHiME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices,” *Proc. ASRU*, pp. 436–443 (2015).
[21] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. Audio Speech Signal Process.*, **ASSP**-32, 1109–1121 (1984).
[22] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano and K. Kondo, “Musical-noise-free speech enhancement based on optimized iterative spectral subtraction,” *IEEE Trans. Audio Speech Lang. Process.*, **20**, 2080–2094 (2012).
[23] P. Smaragdis and J. C. Brown, “Non-negative matrix factorization for music transcription,” *Proc. WASPAA*, pp. 177–180 (2003).
[24] T. Virtanen, “Monaural sound source separation

by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio Speech Lang. Process.*, **15**, 1066–1074 (2007).
[25] C. Févotte, N. Bertin and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Comput.*, **21**, 793–830 (2009).
[26] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," *Proc ICA*, pp. 365–371 (1999).
[27] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis* (John Wiley & Sons, New York, 2001).
[28] Y. Murase, H. Chiba, N. Ono, S. Miyabe, T. Yamada and S. Makino, "On microphone arrangement for multichannel speech enhancement based on nonnegative matrix factorization in time-channel domain," *Proc. APSIPA*, FA1-1-3, pp. 1–5 (2014).
[29] F. Rigaud, A. Falaize, B. David and L. Daudet, "Does inharmonicity improve an NMF-based piano transcription model?," *Proc. WASPAA*, pp. 1–5 (2013).
[30] P. Smaragdis, B. Raj and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," *Proc. ICA*, pp. 414–421 (2007).
[31] E. A. P. Habets, "Room impulse response (RIR) generator," Available at: http://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator（参照 2016-04-27）.
[32] E. Vincent and D. R. Campbell "Roomsimove," Available at: http://www.irisa.fr/metiss/members/evincent/software/Roomsimove.zip（参照 2016-04-27）.
[33] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, **14**, 1462–1469 (2006).
[34] 千葉大将, 小野順貴, 宮部滋樹, 山田武志, 牧野昭二, "教師なし伝達関数ゲイン基底 NMF による目的音強調における罰則項の特性評価," 音講論集, 1-1-15, pp. 527–530 (2014).



村瀬　慶和

2014 筑波大・情報・情報科卒。2016 同大学大学院・シス情・CS 博士前期課程了。修士（工学）。同年株式会社リコーに入社。在学中，アレー信号処理に関する研究に従事。日本音響学会会員。

小野　順貴

2001 東大博士後期課程修了。同年 同大学助手。2005 同大学講師。2011 国立情報学研究所 准教授。アレー信号処理，音源定位，音源分離などの音響信号処理の研究に従事。博士（工学）。IEEE Senior member，日本音響学会，電子情報通信学会，情報処理学会，計測自動制御学会，各会員。

宮部　滋樹

2007 奈良先端大博士後期課程了。2008 米ジョージア工科大学客員研究員。2009 東大特任研究員。2010 同大助教。2011 筑波大助教。音響信号処理の研究に従事。博士（工学）。日本音響学会，IEEE，電子情報通信学会，各会員。

山田　武志

1999 奈良先端大博士後期課程了。同年，筑波大学講師。現在，同准教授。音声認識，音環境理解，多チャネル信号処理，メディア品質評価，e ラーニングの研究に従事。博士（工学）。IEEE，電子情報通信学会，情報処理学会，日本音響学会，日本言語テスト学会，各会員。

牧野　昭二

1981 東北大大学院修士課程了。同年日本電信電話公社入社。以来，NTT 研究所において，電気音響変換器，音響エコーキャンセラ，ブラインド音源分離などの音響信号処理の研究に従事。工博。現在，筑波大学生命領域学際研究センター教授。文部科学大臣表彰（科学技術賞 研究部門），ICA Unsupervised Learning Pioneer Award，IEEE Signal Processing Society Best Paper Award 受賞。IEEE Distingushed Lecturer。IEEE Fellow。電子情報通信学会 Fellow。