多チャネル低ランク・スパース分解に基づく柔軟索状レスキューロボットのためのリアルタイム音声強調

Real-Time Human-Voice Enhancement for a Hose-Shaped Rescue Robot Based on Multi-Channel Low-Rank Sparse Decomposition

○ 学 坂東 宜昭 (京大)　安部 祐一 (東北大)　糸山克寿 (京大)　正 昆陽雅司
正 田所諭 (東北大)　中臺一博 (東工大/HRI)　吉井和佳 (京大)　奥乃博 (早大)

Yoshiaki bando[1], Yuichi Ambe[2], Katsutoshi Itoyama[1], Masashi Konyo[2], Satoshi Tadokoro[2], Kazuhiro Nakadai[3], Kazuyoshi Yoshii[1], and Hiroshi G. Okuno[4]
[1]Kyoto University, [2]Tohoku University,
[3]Tokyo Institute of Technology / Honda Research Institute Japan, and [4]Waseda University

This paper presents a real-time human-voice enhancement method for a hose-shaped rescue robot based on multi-channel low-rank sparse decomposition. Although microphone arrays equipped on hose-shaped robots are crucial for finding victims under collapsed buildings, human voices captured by the microphone array are contaminated by environment-dependent and non-stationary ego-noise. Our method decomposes multi-channel amplitude spectrograms into sparse and low-rank components (human voice and noise) without any prior training. This decomposition is conducted with a state-space model representing the dynamics of these components in a mini-batch manner. Experimental results show that the performance difference between our method and its offline version is less than 3 dB in signal-to-distortion ratio.

Key Words: Hose-shaped rescue robot, Human-voice enhancement, Bayesian signal processing

1　はじめに

柔軟索状レスキューロボット [1] は細長い形状が特徴のロボットで，瓦礫の隙間に挿入し被災者を捜索するために開発されている．例えば，繊毛の振動で駆動する Active Scope Camera (ASC) [1] がある．本ロボットでは，瓦礫に埋もれた被災者を音声で捜索するために，自身の走行雑音の抑圧が不可欠である．より広い範囲を限られた時間で探索するためにロボットは駆動し続ける必要があるが，従来は声を聞くために定期的にアクチュエータを静止させる必要があった．さらに，本ロボットの走行雑音には摩擦音が含まれ，接地面の材質・形状に依存して変化するため，雑音を事前学習する従来法の適用が困難だった．

本ロボットの走行雑音の抑圧には，低ランク・スパース分解による音声強調が有効である [2–5]．低ランク・スパース分解法の一つである RPCA (robust principal component analysis) [2–4] は，そのスペクトログラムが低ランク性である走行雑音と，スパース性である目的音声に事前学習せず分離できる．低ランク・スパース分解は単チャネル音響信号の音源分離法として開発されているが，本手法を多チャネル音響信号の音声強調法として拡張すれば，より高い性能が期待できる．

柔軟索状レスキューロボット上のマイクロホンアレイを用いた音声強調では，1) マイクロホンの移動問題と 2) 瓦礫によるマイクロホンの遮蔽問題に対処する必要がある．本ロボット上のマイクロホンアレイは，ロボットが柔軟であるためにその位置関係がロボットの運動に伴って変動する．また，瓦礫環境では，ロボット上の一部のマイクロホンが瓦礫に隠れ目的音声を全てのマイクロホンで収録できないことがある．従来の多チャネル音声強調法や音源分離法では，これら 2 つの問題のうちどちらかしか対応できなかった [6–9]．さらに，実際の捜索活動ではリアルタイムの音声強調が不可欠となる．

本稿では，入力の多チャネル振幅スペクトログラムを逐次的に低ランク・スパース分解する SVB-MRNMF (streaming variational Bayesian multi-channel robust non-negative matrix factorization) について述べる．本手法は，振幅スペクトログラム上で動作するためマイクロホンの移動に頑健で，目的音声の各マイクロホンでの音量を推定するため遮蔽問題に対処できる．これらの機能は，バッチ処理の従来法である VB-MRNMF を状態空間モデルに拡張し，逐次推論することで実現される．

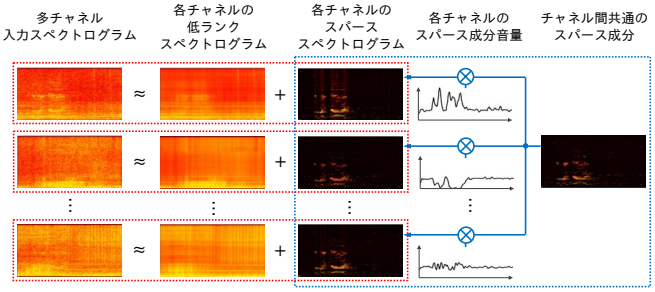

Fig.1 SVB-MRNMF の概要.

2　SVB-MRNMF に基づくミニバッチ音声強調

SVB-MRNMF は，入力である多チャネル音響信号をチャネル毎の低ランク成分 (走行雑音) と，チャネル間共通のスパース成分 (目的音声) に分解し，同時に各チャネルのスパース成分の音量を推定する (図 1).

2.1　問題設定

本稿で扱うマイクロホンアレイを搭載した柔軟索状レスキューロボットを図 2 に示す．ロボット上のマイクロホンは根本側を 1，先端側を $M = 8$ とする．SVB-MRNMF は長さ T に分割されたミニバッチ振幅スペクトログラムを逐次的に音声強調する．本稿で扱う音声強調の問題設定を以下に示す:

入力: M チャネル振幅スペクトログラム $\mathbf{Y}_m^{(n)} \in \mathbb{R}_+^{F\times T}$
出力: 音声強調された振幅スペクトログラム $\mathbf{S}^{(n)} \in \mathbb{R}_+^{F\times T}$

ここで，$\mathbb{R}_+$ は，非負実数値の集合を表し，n はミニバッチのイ

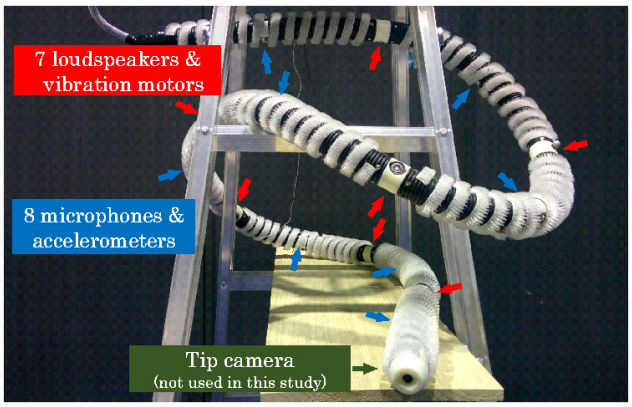

Fig.2 8 チャネル・マイクロホンアレイを搭載した柔軟索状レスキューロボット

ンデックスである ($n = 1, 2, 3, \ldots$). 以降では，F を周波数ビン数とし，f と t を周波数ビン，時間フレームインデックスとする．

2.2 VB-MRNMF

バッチ処理で動作する従来法である VB-MRNMF [10] を概説する．VB-MRNMF は，入力の多チャネル振幅スペクトログラム $\mathbf{Y}_m = [\boldsymbol{y}_{m1}, \ldots, \boldsymbol{y}_{mT}] \in \mathbb{R}_+^{F\times T}$ を，チャネル毎の低ランクスペクトログラム $\mathbf{L}_m = [\boldsymbol{l}_{m1}, \ldots, \boldsymbol{l}_{mT}] \in \mathbb{R}_+^{F\times T}$ と，スパーススペクトログラム $\mathbf{S}_m = [\boldsymbol{s}_{m1}, \ldots, \boldsymbol{s}_{mT}] \in \mathbb{R}_+^{F\times T}$ の和に分解する:

$$\boldsymbol{y}_{mt} \approx \boldsymbol{l}_{mt} + \boldsymbol{s}_{mt}. \tag{1}$$

低ランク振幅スペクトログラムは，VB-RPCA [11] と同様に，K 個の基底スペクトル $\mathbf{W}_m = [\boldsymbol{w}_{m1}, \ldots, \boldsymbol{w}_{mK}] \in \mathbb{R}_+^{F\times K}$ とそれらのアクティベーションベクトル $\mathbf{H}_m = [\boldsymbol{h}_{m1}, \ldots, \boldsymbol{h}_{mT}] \in \mathbb{R}_+^{K\times T}$ の積として表現される:

$$\boldsymbol{y}_{mt} \approx \mathbf{W}\boldsymbol{h}_{mt} + \boldsymbol{s}_{mt}. \tag{2}$$

さらに，目的音声 $\boldsymbol{s}_t \in \mathbb{R}_+^F$ とその各マイクロホンでの観測 $\boldsymbol{s}_{mt} \in \mathbb{R}_+^F$ との関係を周波数非依存時変線形システムと仮定する:

$$\boldsymbol{s}_{mt} \approx g_{mt}\boldsymbol{s}_t, \tag{3}$$

ここで，$g_{mt} \in \mathbb{R}_+$ は各マイクロホン m および時刻 t での目的音源の音量を表す．以上より，観測 $\boldsymbol{y}_{mt}$ は次のように分解される:

$$\boldsymbol{y}_{mt} \approx \mathbf{W}\boldsymbol{h}_{mt} + g_{mt}\boldsymbol{s}_t. \tag{4}$$

各スペクトログラムの低ランク性とスパース性は，以降で述べるようにベイズ的に定式化される．

2.2.1 尤度関数

VB-MRNMF では，入力振幅スペクトログラムの近似誤差を Kullback-Leibler (KL) 儀距離に基づいて最小化する．KL 儀距離の最小化は Poisson 分布の最尤推定に相当するため，尤度関数は以下のように各マイクロホンごとに定義する．

$$p(\mathbf{Y}_m|\mathbf{W}_m, \mathbf{H}_m, \mathbf{S}_m) = \prod_{f,t} \mathcal{P}\left(y_{mft} \left| \sum_k w_{mfk}h_{mkt} + g_{mt}s_{ft}\right.\right) \tag{5}$$

ここで，$\mathcal{P}$ は Poisson 分布を表す．

2.2.2 低ランク成分に対する事前分布

低ランク成分は Bayesian NMF [12] を参考に定式化する．低ランク成分の潜在変数である基底行列 $\mathbf{W}_m$ およびアクティベーション行列 $\mathbf{H}_m$ に対し，Poisson 尤度関数の共役事前分布であるガンマ分布を置く．

$$p(\mathbf{W}_m|\alpha^{wh}, \beta^{wh}) = \prod_{f,k} \mathcal{G}(w_{mfk}|\alpha^{wh}, \beta^{wh}) \tag{6}$$

$$p(\mathbf{H}_m|\alpha^{wh}, \beta^{wh}) = \prod_{k,t} \mathcal{G}(h_{mkt}|\alpha^{wh}, \beta^{wh}) \tag{7}$$

ここで，$\mathcal{G}$ はガンマ分布を表し，$\alpha^{wh} \in \mathbb{R}_+$ および $\beta^{wh} \in \mathbb{R}_+$ はガンマ分布の shape および rate パラメータを表す．Shape パラメータ α^{wh} を 1 以下にすることで，基底およびアクティベーション行列をスパースに誘導できることが知られている [12]．これによって，低ランク成分 $\mathbf{L}$ を低ランクに誘導する．

2.2.3 スパース成分に対する事前分布

VB-RPCA では，スパース成分に Gauss 分布と Jeffreys 超事前分布を置き，スパース性を表現していた [11]．スパース成分を非負値に制限するため，VB-MRNMF では，ガンマ分布の rate パラメータに Jeffreys 超事前分布を置き，スパース性を表現する:

$$p(\mathbf{S}|\alpha^s, \boldsymbol{\beta}^s) = \prod_{f,t} \mathcal{G}(s_{ft}|\alpha^s, \beta^s_{ft}) \tag{8}$$

$$p(\beta^s_{ft}) \propto (\beta^s_{ft})^{-1} \tag{9}$$

ここで，$\alpha^s \in \mathbb{R}_+$ はガンマ分布の超パラメータを表す．この shape パラメータ α^s によって $\mathbf{S}$ のスパース性を調整する．また，各マイクロホンの音量 g_{mt} には以下のようにガンマ事前分布を置く:

$$p(g_{mt}|\alpha^g) = \mathcal{G}(g_{mt}|\alpha^g, \alpha^g), \tag{10}$$

ここで，$\alpha^g \in \mathbb{R}_+$ は，マイクロホン間のばらつき度合いを表す超パラメータである．

2.3 SVB-MRNMF

VB-MRNMF をミニバッチ推論できるように拡張するため，低ランク成分とスパース成分を時変な潜在変数とする状態空間モデルとして定式化する．本モデルは，各ミニバッチの振幅スペクトログラム $y^{(n)}_{mft}$ を VB-MRNMF と同様に低ランク成分とスパース成分に分解する:

$$\boldsymbol{y}^{(n)}_{mt} \approx \mathbf{W}^{(n)}_m \boldsymbol{h}^{(n)}_{mt} + g^{(n)}_{mt}\boldsymbol{s}^{(n)}_t. \tag{11}$$

ここで，$\mathbf{W}^{(n)}_m \in \mathbb{R}_+^{F\times K}$, $\boldsymbol{h}^{(n)}_{mt} \in \mathbb{R}_+^K$, $g^{(n)}_{mt} \in \mathbb{R}_+$ および $\boldsymbol{s}^{(n)}_t \in \mathbb{R}_+^F$ は，n 番目のミニバッチでの VB-MRNMF における $\mathbf{W}_m$, $\boldsymbol{h}_{mt}$, g_{mt} および $\boldsymbol{s}_t$ を表す．n 番目のミニバッチにおける潜在変数の集合 $\{\mathbf{W}^{(n)}_{1:M}, \mathbf{H}^{(n)}_{1:M}, \boldsymbol{g}^{(n)}_{1:M}, \mathbf{S}^{(n)}, \boldsymbol{\beta}^{s(n)}\}$ を $\boldsymbol{\Theta}^{(n)}$ とする．以降，潜在変数の時間変動を表す状態空間モデルとして，観測モデル $p(\mathbf{Y}^{(n)}_{1:M}|\boldsymbol{\Theta}^{(n)})$ と状態更新モデル $p(\boldsymbol{\Theta}^{(n)}|\boldsymbol{\Theta}^{(n-1)})$ を定式化する．

2.3.1 観測モデル

SVB-MRNMF の観測モデル $p(\mathbf{Y}^{(n)}_{1:M}|\boldsymbol{\theta}^{(n)})$ は，VB-MRNMF と同様に Poisson 分布で表現する:

$$p\left(\mathbf{Y}^{(n)}_{1:m}\middle|\boldsymbol{\theta}^{(n)}\right) = \prod_{mft} \mathcal{P}\left(y^{(n)}_{mft}\left|\sum_k w^{(n)}_{mfk}h^{(n)}_{mkt} + g^{(n)}_{mt}s^{(n)}_{mft}\right.\right). \tag{12}$$

2.3.2 状態更新モデル

スパース成分の潜在変数 $\boldsymbol{g}^{(n)}_{1:M}$, $\mathbf{S}^{(n)}$ と $\boldsymbol{\beta}^{s(n)}$ および，低ランク成分のアクティベーション行列 $\boldsymbol{H}^{(n)}_{1:M}$ は各時間フレームごとに独立なので，本状態空間モデルでは，基底行列 $\mathbf{W}^{(n)}_m$ のみ過去の状態 $\mathbf{W}^{(n-1)}_m$ に依存する:

$$p(\boldsymbol{\theta}^{(n)}|\boldsymbol{\theta}^{(n-1)}) = p(\mathbf{W}^{(n)}_m|\mathbf{W}^{(n-1)}_m)p(\mathbf{H}^{(n)}_m)\times p(\boldsymbol{g}^{(n)}_m)p(\mathbf{S}^{(n)})p(\boldsymbol{\beta}^{s(n)}) \tag{13}$$

過去の状態に依存しない $\mathbf{H}^{(n)}_m$, $\boldsymbol{g}^{(n)}_m$, $\mathbf{S}^{(n)}$ と $\boldsymbol{\beta}^{s(n)}$ の事前分布は，VB-MRNMF と同様に置く．

基底行列 $\mathbf{W}^{(n)}_m$ の状態更新モデルは，$\mathbf{W}^{(n)}_m$ の各要素ごとに独立に定義する:

$$p(\mathbf{W}^{(n)}_m|\mathbf{W}^{(n-1)}_m) = \prod_{m,f,k} p(w^{(n)}_{mfk}|w^{(n-1)}_{mfk}) \tag{14}$$

この要素ごとの状態更新モデル $p(w^{(n)}_{mfk}|w^{(n-1)}_{mfk})$ は，過去の状態との関係を表す $p_1(w^{(n)}_{mfk}|w^{(n-1)}_{mfk})$ と，時間に普遍な事前分布

$p_2(w_{mfk}^{(n)})$ からなる．現在の状態 $w_{mfk}^{(n)}$ は，1 ミニバッチ前の状態 $w_{mfk}^{(n-1)}$ に乗法性ノイズ $\phi_{mfk}^{(n)} \in \mathbb{R}_+$ を乗じて更新される:

$$w_{mfk}^{(n)} = \phi_{mfk}^{(n)} w_{mfk}^{(n-1)}. \tag{15}$$

ここで，このプロセスノイズは $w_{mfk}^{(n-1)}$ の平均を保存し分散のみ増大する特性が望ましい．言い換えると，1 ミニバッチ前の事後分布 $p(w_{mfk}^{(n-1)}|\mathbf{Y}^{(1:n-1)})$ を $\mathcal{G}(\hat{\alpha}_{mfk}^{(n-1)}, \hat{\beta}_{mfk}^{(n-1)})$ とすると，現在のミニバッチの予測分布 $p_1(w_{mfk}^{(n)}|\mathbf{Y}^{(1:n-1)})$ は以下となることが望ましい:

$$p_1(w_{mfk}^{(n)}|\mathbf{Y}^{(1:n-1)}) = \mathcal{G}(\gamma\hat{\alpha}_{mfk}^{(n-1)}, \gamma\hat{\beta}_{mfk}^{(n-1)}). \tag{16}$$

ここで，$\gamma \in \mathbb{R}_+$ は，$w_{mfk}^{(n)}$ の分散を調節するスケールパラメータであり，本予測分布の平均は $p(w_{mfk}^{(n-1)}|\mathbf{Y}^{(1:n-1)})$ の平均と一致し，分散は γ^{-1} 倍されている．文献 [13] で報告されているように，本性質を満たすプロセスノイズはベータ分布を用いて以下のように表現される:

$$p_1(w_{mfk}^{(n)}|w_{mfk}^{(n-1)}) = \mathcal{B}\left(\gamma\frac{w_{mfk}^{(n)}}{w_{mfk}^{(n-1)}}\middle|\gamma\hat{\alpha}_{mfk}^{(n-1)}, (1-\gamma)\hat{\alpha}_{mfk}^{(n-1)}\right). \tag{17}$$

ただし，$\mathcal{B}(\alpha, \beta)$ は形状母数 α と β を持つベータ分布を表す．一方，時間に普遍な事前分布 $p_2(w_{mfk}^{(n)})$ は VB-MRNMF と同様に以下のように定式化する:

$$p_2(w_{mfk}^{(n)}) = \mathcal{G}(w_{mfk}^{(n)}|\alpha^{wh}, \beta^{wh}). \tag{18}$$

これら 2 つの分布は，状態更新モデル $p(w_{mfk}^{(n)}|w_{mfk}^{(n-1)})$ として，Product of Experts [14] の考え方により，以下のように統合する:

$$p(w_{mfk}^{(n)}|w_{mfk}^{(n-1)}) \propto p_1(w_{mfk}^{(n)})^{\kappa} p_2(w_{mfk}^{(n)}|w_{mfk}^{(n-1)})^{1-\kappa} \tag{19}$$

ここで，$\kappa \in \mathbb{R}_+$ $(0.0 < \kappa < 1.0)$ は，$p1$ と $p2$ それぞれの影響を制御する重みパラメータである．

2.4 変分ベイズ法に基づく推論

本推論の目的は，未知パラメータの真の事後分布を求めることである．真の事後分布は解析的に導出困難なので，本稿では，変分ベイズ法に基づいた近似推論を行う [12].

2.4.1 VB-MRNMF

VB-MRNMF で用いた確率分布は全て共役指数分布族上で定義されているので，各変分事後分布は Jensen の不等式と Lagrange の未定乗数法を用いることで計算できる [12]. 以降，$\langle x \rangle$ を x の事後分布の期待値すると，各辺分事後分布は，以下に従い順に各変数を他の変数を固定しながら更新することで反復推定できる．

$$q(w_{mfk}) = \mathcal{G}(\alpha^{wh} + \sum_t y_{mft}\lambda_{mftk}^{wh}, \beta^{wh} + \sum_t \langle h_{mtk}\rangle), \tag{20}$$

$$q(h_{mtk}) = \mathcal{G}(\alpha^{wh} + \sum_f y_{mft}\lambda_{mftk}^{wh}, \beta^{wh} + \sum_f \langle w_{mfk}\rangle), \tag{21}$$

$$q(g_{mt}) = \mathcal{G}(\alpha^{g} + \sum_f y_{mft}\lambda_{mft}^{gs}, \alpha^{g} + \sum_f \langle s_{ft}\rangle), \tag{22}$$

$$q(s_{ft}) = \mathcal{G}(\alpha^{s} + \sum_m y_{mft}\lambda_{mft}^{gs}, \langle\beta_{ft}^{s}\rangle + \sum_m \langle g_{mt}\rangle), \tag{23}$$

$$q(\beta_{ft}^{s}) = \mathcal{G}(\alpha^{s}, \langle s_{ft}\rangle), \tag{24}$$

$$\lambda_{mftk}^{wh} = \frac{\mathbb{G}[w_{mfk}]\mathbb{G}[h_{mtk}]}{\sum_k \mathbb{G}[w_{mfk}]\mathbb{G}[w_{mtk}] + \mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]}, \tag{25}$$

$$\lambda_{mft}^{gs} = \frac{\mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]}{\sum_k \mathbb{G}[h_{mfk}]\mathbb{G}[h_{mtk}] + \mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]}. \tag{26}$$

ここで，$\mathbb{G}[x]$ は x の幾何平均を表し，λ_{mftk}^{wh} と λ_{mft}^{gs} は補助変数を表す．

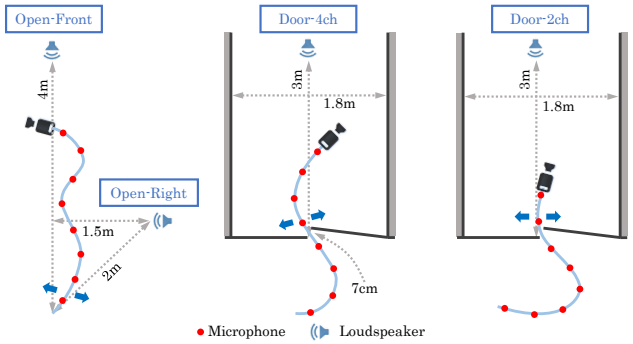

Fig.3 評価実験で用いたスピーカとロボットの配置条件

2.4.2 SVB-MRNMF

SVB-MRNMF の各ミニバッチ入力に対する事後分布は，予測ステップと修正ステップを逐次的に計算することで推定する．予測ステップでは，予測分布 $p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n-1)})$ を直前の事後分布 $p(\mathbf{\Theta}^{(n-1)}|\mathbf{Y}^{(1:n-1)})$ から計算する:

$$p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n-1)}) = \int p(\mathbf{\Theta}^{(n)}|\mathbf{\Theta}^{(n-1)}) p(\mathbf{\Theta}^{(n-1)}|\mathbf{Y}^{(1:n-1)}) d\mathbf{\Theta}^{(n-1)}. \tag{27}$$

修正ステップでは，現在のミニバッチ入力 $\mathbf{Y}^{(n)}$ に対する事後分布 $p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n)})$ を以下のように計算する:

$$p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n)}) \propto p(\mathbf{Y}^{(n)}|\mathbf{\Theta}^{(n)}) p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n-1)}) \tag{28}$$

式 13 および 14 より，予測分布は以下のように計算できる:

$$p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n-1)}) = \prod_{m,f,k} p(w_{mfk}^{(n)}|\mathbf{Y}^{(1:n-1)}) \times p(\mathbf{H}^{(n)}) p(\boldsymbol{g}_m^{(n)}) p(\mathbf{S}^{(n)}) p(\boldsymbol{\beta}^{s(n)}). \tag{29}$$

このうち，基底成分の予測分布 $p(w_{mft}^{(n)}|\mathbf{Y}^{(1:n-1)})$ は以下となる:

$$p\left(w_{mft}^{(n)}\middle|\mathbf{Y}^{(1:n-1)}\right) = \mathrm{Gamma}\left(\hat{\alpha}_{ftk}^{(n)}, \hat{\beta}_{ftk}^{(n)}\right), \tag{30}$$

$$\hat{\alpha}_{ftk}^{(n)} = \kappa\alpha^{wh} + (1-\kappa)\gamma\hat{\alpha}^{(n-1)}, \tag{31}$$

$$\hat{\beta}_{ftk}^{(n)} = \kappa\beta^{wh} + (1-\kappa)\gamma\hat{\beta}^{(n-1)}. \tag{32}$$

本予測分布を VB-MRNMF の事前分布とし式 20–26 を計算すれば，現在のミニバッチの事後分布 $p(\mathbf{\Theta}^{(n)}|\mathbf{Y}^{(1:n)})$ を得る．

3 評価実験

本章では，実ロボットを用いて収録した走行雑音を用いた評価実験を報告する．

3.1 実験設定

図 2 に示すように柔軟索状ロボットの本体は，直径 38 mm のコルゲートチューブからなり，全長 3 m である．8 つのマイクロホンアレイ (M=8) をロボット表面に 40 cm 間隔で 90 度ずつ回転して装着した．両端のマイクロホン間の距離は 2.8 m である．マイクロホンは手元から順番にインデックス m で区別する $(m = 1, \cdots, M)$．本ロボットは，Fukuda らの Tube-type ASC [1] と同様，繊毛と振動モータを用いて前進する．振動モータはロボット内に 40 cm 間隔で 7 つ直列に装着されている．本実験ではマイクロホンアレイを 16 kHz，24 ビットで同期録音した．

本稿で用いる走行雑音と目的音声はそれぞれ独立に収録し，信号対雑音比 (signal-to-noise ratio: SNR) を変動させながら混合して評価に用いた．SNR は，−20 dB から +5 dB まで 5 dB 刻みで変動させた．図 3 に示すように，本実験では目的音声を再生するスピーカとロボットの配置を以下の 4 種で評価した．

1. **Open-Front**: ロボットは障害物のない実験室内に配置し，スピーカはロボットの正面に配置した．本実験室の残響時間 (RT_{60}) は 750 ms だった．

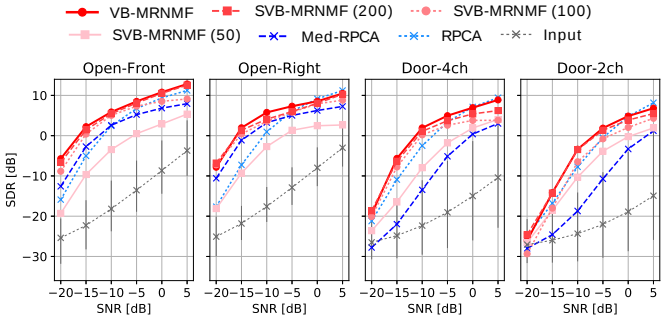

Fig.4 評価実験での音声強調結果の SDR．入力 SDR のエラーバーは各マイクロホンの最大値と最小値を表す．

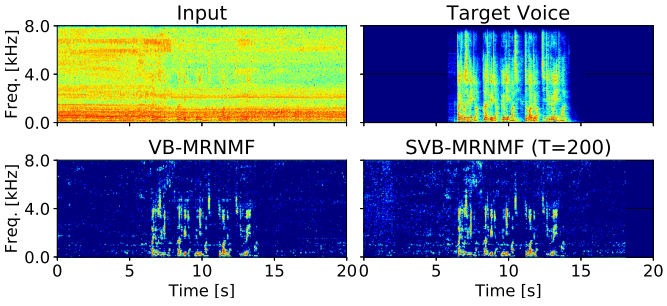

Fig.5 VB-MRNMF と SVB-MRNMF による音声強調の結果例 (SNR は-5.0 dB)．

2. **Open-Right**: 音源がロボットの右側に配置されていることを除いて，Open-Front と同様に配置した．
3. **Door-4ch**: ロボットはドアに挟まれており，スピーカはロボットの正面に配置されている．ドアにより後方 4 つのマイクロホンがスピーカから隠れている．残響時間は 990 ms だった．
4. **Door-2ch**: 後方 6 つのマイクロホンがドアで隠されていることを除いて，Door-4ch と同様に配置した．

走行雑音は各条件においてロボットを駆動させ，手を用いて左右にロボット振りながら，60 秒の動作雑音を録音した．目的音声は ASJ JNAS 音素バランス文読み上げ音声に含まれる，男女それぞれ 3 名ずつ計 24 発話を用いた．これらの音声と 20 秒間に切り出した収録雑音を混合して入力信号を作成した．

評価尺度には Signal-to-distortion ratio (SDR) [15] を用いた．SVB-MRNMF のミニバッチサイズ T を 50, 100, 200 フレームと変動させて性能を評価した．従来法として，VB-MRNMF，先端のマイクロホンに RPCA を適用した場合 [2] と比較した．また，各マイクロホンの RPCA の結果を中央値選択で統合する従来法 (Med-RPCA) とも比較を行った [16].

3.2 実験結果

図 4 に示すように，バッチ法である VB-MRNMF が最も高い SDR を示しいている．ミニバッチ法である SVB-MRNMF もミニバッチサイズ T が 200 フレームのとき，SNR が-10 dB 以下で VB-MRNMF に次いで高い SDR であった．また，VB-MRNMF との差は 3 dB 以下である．各マイクロホンの RPCA 結果を中央値選択で統合した Med-RPCA は，一部のマイクロホンが遮蔽される Door-4ch と-2ch の条件で大きく SDR が低下している．SVB-MRNMF と VB-MRNMF も低下しているが，最も音源に近いマイクロホンに適用した RPCA と同程度以上の SDR となった．以上より，SVB-MRNMF はマイクロホンの異動問題と遮蔽問題を解決しながらミニバッチでの音声強調を実現した．一方，ミニバッチサイズが 50 フレームの場合大きく性能が低下しており，ミニバッチサイズは 100 フレーム以上必要であることが分かる．図 5 に，SVB-MRNMF と VB-MRNMF による音声強調結果例を示す．観測の音響信号には，時々刻々と変化する走行雑音が混入しているが，両手法ともこの走行雑音を抑圧できてる．

屋外で利用可能なロボットシステムを実現するために，組み込み GPGPU ボードの一つである Jetson TX1 上に SVB-MRNMF を実装した．実装には C++11 と CUDA 8.0 を使用した．20.0 秒間の入力信号をミニバッチサイズ 200 フレームで解析するために要した時間は 19.8 秒であった．この時間は入力信号の長さより短く，継続的に入力信号を音声強調する処理時間を達成した．

4 おわりに

本稿では，事前学習不要でリアルタイムに動作する音声強調法である SVB-MRNMF について述べた．柔軟索状レスキューロボットの音声強調には，マイクロホンの移動問題と遮蔽問題の 2 つの課題があった．これらの問題に対処するため，多チャネル振幅スペクトログラムからスパース成分 (目的音声) と低ランク成分 (走行雑音) を分離するミニバッチで分離する SVB-MRNMF を開発した．本手法では，各マイクロホンの目的音源の音量を推定するため，障害物で遮蔽されたマイクロホンの影響を低減できる．実験では，バッチ処理の従来法である VB-MRNMF と比較し，SDR が最大 3dB 程度の差であり，リアルタイムに高精度な音声強調が行えることを確認した．今後は，模擬評価フィールドでのユーザビリティ評価や，ビームフォーマを用いたポストフィルタによる音質改善を行う．

謝辞 本研究は，科研費基盤 (S) No.24220006，特別研究員奨励費 No. 15J08765，および ImPACT「タフ・ロボティクス・チャレンジ」の支援を受けた．

参考文献

[1] J. Fukuda et al. Remote vertical exploration by active scope camera into collapsed buildings. In *IEEE/RSJ IROS*, pages 1882–1888, 2014.

[2] C. Sun et al. Noise reduction based on robust principal component analysis. *JCIS*, 10(10):4403–4410, 2014.

[3] E. J. Candès et al. Robust principal component analysis? *JACM*, 58(3):11, 2011.

[4] Z. Chen et al. Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition. In *IEEE WASPAA*, pages 1–4, 2013.

[5] N. Dobigeon et al. Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images. In *WHISPERS*, pages 1–4, 2013.

[6] N. Ono. Stable and fast update rules for independent vector analysis based on auxiliary function technique. In *IEEE WASPAA*, pages 189–192, 2011.

[7] D. Kitamura et al. Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model. In *IEEE ICASSP*, pages 276–280, 2015.

[8] D. Kounades-Bastian et al. A variational EM algorithm for the separation of moving sound sources. In *IEEE WASPAA*, pages 1–5, 2015.

[9] J. Nikunen et al. Direction of arrival based spatial covariance model for blind sound source separation. *IEEE/ACM TASLP*, 22(3):727–739, 2014.

[10] Y. Bando et al. Variational bayesian multi-channel robust nmf for human-voice enhancement with a deformable and partially-occluded microphone array. In *EUSIPCO*, pages 1018–1022, 2016.

[11] S. D. Babacan et al. Sparse Bayesian methods for low-rank matrix estimation. *IEEE TSP*, 60(8):3964–3977, 2012.

[12] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *CIN*, 2009(785152):1–17, 2009.

[13] D. Gamerman et al. A non-Gaussian family of state-space models with exact marginal likelihood. *JTSA*, 34(6):625–645, 2013.

[14] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

[15] E. Vincent et al. Performance measurement in blind audio source separation. *IEEE TASLP*, 14(4):1462–1469, 2006.

[16] Y. Bando et al. Human-voice enhancement based on online RPCA for a hose-shaped rescue robot with a microphone array. In *IEEE SSRR*, pages 1–6, 2015.