社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

変分ベイズ多チャネルロバストNMFに基づく マイクロホンの移動・被覆を許容する音声強調

坂東　宜昭† 糸山　克寿† 昆陽　雅司†† 田所　諭†† 中臺　一博†††
吉井　和佳† 河原　達也† 奥乃　博††††

† 京都大学 大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
†† 東北大学 大学院情報科学研究科　〒 980–8579 宮城県仙台市青葉区荒巻字青葉 6–6–01
††† 東京工業大学 大学院情報理工学研究科 / ホンダ RI ジャパン　〒 152–8552 東京都目黒区大岡山
†††† 早稲田大学 理工学術院　〒 169–0072 東京都新宿区大久保 2–4–12
E-mail: †{yoshiaki, itoyama, yoshii}@kuis.kyoto-u.ac.jp, kawahara@i.kyoto-u.ac.jp,
††{konyo, tadokoro}@rm.is.tohoku.ac.jp, †††nakadai@jp.honda-ri.com, ††††okuno@nue.org

あらまし　本稿では，マイクロホンの移動・被覆を許容する多チャネル音声強調法について述べる．瓦礫に埋もれた被災者を捜索するために，細長い形状が特徴の柔軟索状レスキューロボットが開発されている．視界が制限される瓦礫環境下では被災者の声を手がかりにした捜索が重要であるが，本ロボットでは自身の非定常な走行雑音によって音声が聞き取りづらくなる問題があった．さらに本ロボットでは，1) 駆動に伴ってマイクロホンアレイの配置が動的に変化するため (移動問題)，2) 一部のマイクロホンが障害物に遮蔽され音声を十分大きく収録できなくなるため (遮蔽問題)，従来のブラインド音源分離法の使用が困難であった．本研究では，マイクロホンアレイの配置に強く依存する位相情報を用いず振幅スペクトログラム領域での音声強調を行う．さらに，各マイクロホンでの目的音声の音量を同時推定し，遮蔽問題に対処する．これらの機能は，入力音響信号である多チャネル振幅スペクトログラムを低ランク成分 (雑音) とスパース成分 (目的音声) に分離することで，事前情報を用いずに行われる．8 チャネル・マイクロホンアレイを搭載する 3m の柔軟索状レスキューロボットを用いた評価実験で，信号対雑音比が従来のブラインド音源分離法に比べて 2.7 dB 向上することを確認した．

キーワード　ブラインド音声強調, レスキューロボット, 分散マイクロホンアレイ, 低ランク・スパース分解

1. はじめに

レスキューロボットは人や動物が侵入できない環境を探索するために開発されており [1]，その中でも柔軟索状レスキューロボット [2] 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できる．例えば，繊毛の振動で駆動する Active Scope Camera (ASC) [2] が報告されている．図 1 に示すように本ロボットは，マイクロホンアレイと先端に取付けられた小型カメラを用いて被災者を捜索する．本ロボット上のマイクロホンアレイは，全てのマイクロホンが同時に障害物に覆われないように，ロボット全体に配置されている [3]．

被災者の音声を手がかりにした捜索は，障害物が多く視野が狭くなる瓦礫下で有効であるが，柔軟索状レスキューロボットでは，自身の走行雑音により音声が聞き取りづらくなる問題があった．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り非効率的であった．ロボットの走行雑音などの自己生成音を事前に学習し抑圧する音声強調法が開発されているが [4–7]，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，これらの従来法は適用が困難である．

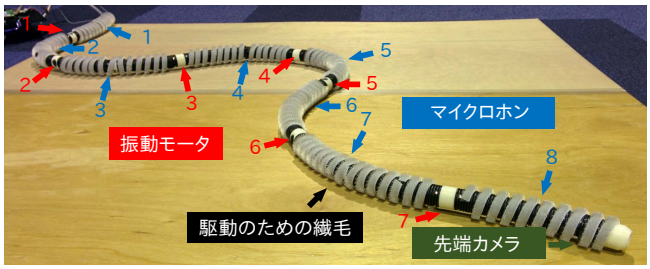


図 1: 8 チャネル・マイクロホンアレイを搭載した柔軟索状レスキューロボット

マイクロホンアレイ信号処理の観点から，柔軟索状レスキューロボットの音声強調では，マイクロホンの移動問題と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，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．マイクロホンの移動問題は，ロボットの姿勢 (形状) を推定し対処する方

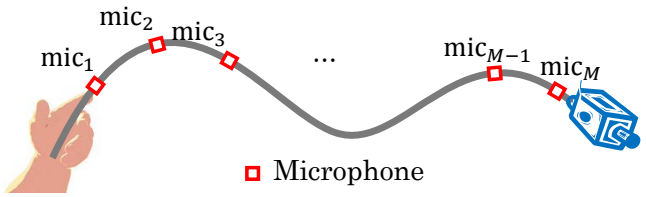


図 2: ロボット上のマイクロホンの配置

法 [8] が考えられるが，従来の姿勢推定法の精度は，マイクロホンの遮蔽問題に対処できる位相情報に基づく音源分離法には不十分である [9,10].

本稿では，マイクロホンの移動と遮蔽に頑健な多チャネル音声強調法について述べる．提案法は，マイクロホンアレイの配置に強く依存する位相情報を用いずに，振幅スペクトログラム領域で音声強調を行う．走行雑音のスペクトログラムが低ランクであり，音声のスペクトログラムがスパースであることに注目し，多チャネル振幅スペクトログラムから事前情報を用いずに走行雑音と音声を分離する [11–14]．また，各マイクロホンにおける目的音声の音量を推定し，マイクロホンの遮蔽問題に対処する．提案法の有効性は，マイクロホンアレイを搭載した 3m の柔軟索状ロボットを用いた評価実験で確認する．

2. 関 連 研 究

本章では，音声強調を行うための，雑音と音声を分離する従来の音源分離法を，位相情報に基づく手法と振幅情報に基づく手法に分類し議論する．

2.1 位相情報に基づく音源分離

マイクロホン間の位相差に基づくブラインド音源分離法は，音源やマイクロホンに関する事前知識を用いない音源分離法として広く研究されている [9,15–18]. 例えば，多チャネル非負値行列因子分解 (Multi-channel nonnegative matrix factorization: MNMF) [16–18] は，混合音を観測した多チャネル複素スペクトログラムを，各音源信号を表す低ランク性のスペクトログラムとそのマイクロホンへの伝達関数に分解する．Kounades-Bastin ら [18] は，従来音源に静止仮定を置いていた MNMF を拡張し，移動音源に対応した．本手法では，時々刻々と変化する伝達関数にマルコフ性のみを仮定し，音源分離を行う．これらの手法は，全てのマイクロホンで全ての音源信号が観測されることを仮定するため，柔軟索状ロボットの走行雑音のように異なる雑音が各マイクロホンに含まれる場合，その分離性能は大きく劣化する．さらに，軟索状ロボットでは振動モータにより伝達関数が小刻みに変化するため，伝達関数にマルコフ性を仮定し正確に推定することは難しい．

2.2 振幅情報に基づく音源分離

時間変化する音源とマイクロホン間の伝達関数の推定を避ける方法として，振幅スペクトログラム領域での多チャネル音源分離法が考えられる．例えば，千葉ら [19] は，非同期分散マイクロホンアレイのための振幅スペクトログラム領域で動作する音源分離法を開発した．非同期分散マイクロホンアレイでは，マイクロホン間の位相差が A/D 変換器のクロックのずれに起因して時間経過と共に少しずつ変動する．このような状況では位相情報が信頼できないため，NMF を多チャネル振幅スペクトログラムに対して適用し，音源分離を行った．しかしこの手法では，各マイクロホン間の音量差情報が既知である必要があり，障害物が存在する瓦礫内環境では，本情報を取得することが困難である．

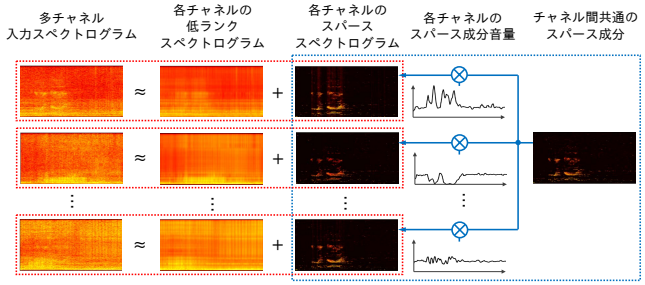


図 3: 変分ベイズ多チャネルロバスト NMF の概要

事前情報を用いずに非定常な雑音を抑圧し，音声を強調する手法として，低ランク・スパース分解に基づく音源分離がある [11–14, 20]. 例えば，ロバスト PCA (Robust Principal Component Analysis: RPCA) は，入力である単チャネル振幅スペクトログラムに含まれる雑音と音声を，それぞれ低ランク振幅スペクトログラムとスパース振幅スペクトログラムに分解できる [11, 12]. さらに，RPCA は低ランク成分とスパース成分の不確かさを考慮できるようにベイズ的な拡張がなされている [21, 22]. 例えば，ビデオ映像の前景と背景を分離するために，Ding ら [21] はマルコフ制約をスパース成分 (前景) に仮定し，分離成分に現れるごま塩ノイズの軽減を行った．また，Babacan ら [22] は，変分ベイズ (Variational Bayesian: VB) 推論法に基づくベイジアン RPCA (VB-RPCA) を開発し，計算量を削減した．しかし，RPCA は入力，低ランク成分，スパース成分に負の値を許してしまうため，RPCA の音響信号 (振幅スペクトログラム) への応用は物理モデルとしての妥当性が欠如していた．

音響信号や映像の解析のために，入力の非負値行列を非負の低ランク成分とスパース成分に分解する，ロバスト NMF (RNMF) が提案されている [14, 20]. Sun ら [20] は，Kullback-Leibler (KL) 距離に基づく RNMF を報告している．KL 距離は，音響信号の分離において広く用いられている距離尺度である．ベイジアン RPCA のように，RNMF をベイズ的に定式化することで，多チャネル音響信号の分離など更なる拡張性が期待できる．

3. 変分ベイズ多チャネルロバスト NMF

本章では，提案法である変分ベイズ多チャネルロバスト NMF (VB multi-channel RNMF: VB-MRNMF) について述べる．VB-MRNMF で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) に分解する (図 3)．また，VB-MRNMF は同時に各チャネルのスパース成分の音量を推定する．以降では，振幅スペクトログラムを扱いやすくするために非負制約

を置いた VB-RPCA である変分ベイズロバスト NMF (VB-RNMF) を定式化し，その後 VB-RNMF を多チャネル拡張した VB-MRNMF を定式化する．

3.1 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 1 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 とする (図 2)．また，F および T をそれぞれ周波数ビン数，時間フレーム数とし，f と t をそれぞれのインデックスとする．本稿で扱う音声強調の問題設定を以下に示す:

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

ここで，$\mathbb{R}_+$ は，非負実数値の集合を示す．振幅スペクトログラムは，時間領域信号を短時間フーリエ変換 (Short Time Fourier Transform: STFT) し，絶対値を取ることで得られる．

3.2 単チャネル音響信号のための VB-RNMF

本稿ではまず，変分ベイズロバスト NMF (VB-RNMF) を単チャネル振幅スペクトログラム $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T] \in \mathbb{R}_+^{F\times T}$ に対し定式化する．VB-RNMF では，入力振幅スペクトログラム $\mathbf{Y}$ を，低ランク振幅スペクトログラム $\mathbf{L} = [\boldsymbol{l}_1, \ldots, \boldsymbol{l}_T] \in \mathbb{R}_+^{F\times T}$ と，スパース振幅スペクトログラム $\mathbf{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_T] \in \mathbb{R}_+^{F\times T}$ の和に分解する:

$$\boldsymbol{y}_t \approx \boldsymbol{l}_t + \boldsymbol{s}_t \tag{1}$$

低ランク振幅スペクトログラムは，VB-RPCA [22] と同様に，K 個の基底スペクトル $\mathbf{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K] \in \mathbb{R}_+^{F\times K}$ とそれらのアクティベーションベクトル $\mathbf{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_T] \in \mathbb{R}_+^{K\times T}$ の積として表現される:

$$\boldsymbol{y}_t \approx \mathbf{W}\boldsymbol{h}_t + \boldsymbol{s}_t \tag{2}$$

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

3.2.1 尤度関数

VB-RNMF では，入力振幅スペクトログラムの近似誤差を Kullback-Leibler (KL) 距離に基づいて最小化する．KL 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定義する．

$$p(\mathbf{Y}|\mathbf{W},\mathbf{H},\mathbf{S}) = \prod_{f,t} \mathcal{P}\left(y_{ft} \middle| \sum_k w_{fk}h_{kt} + s_{ft}\right) \tag{3}$$

ここで，$\mathcal{P}$ は Poisson 分布を表す．

3.2.2 低ランク成分に対する事前分布

低ランク成分は Bayesian NMF [23] 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 $\mathbf{W}$ およびアクティベーション行列 $\mathbf{H}$ に対し，Poisson 尤度関数の共役事前分布である gamma 分布を置く．

$$p(\mathbf{W}|\alpha^{wh},\beta^{wh}) = \prod_{f,k} \mathcal{G}(w_{fk}|\alpha^{wh},\beta^{wh}) \tag{4}$$

$$p(\mathbf{H}|\alpha^{wh},\beta^{wh}) = \prod_{k,t} \mathcal{G}(h_{kt}|\alpha^{wh},\beta^{wh}) \tag{5}$$

ここで，$\mathcal{G}$ は gamma 分布を表し，$\alpha^{wh} \in \mathbb{R}_+$ および $\beta^{wh} \in \mathbb{R}_+$ は gamma 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 [23]．これによって，低ランク成分 $\mathbf{L}$ を低ランクに誘導する．

3.2.3 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 分布とその分散パラメータに Jeffreys 超事前分布を置き同時推定することで，スパース性を表現していた [22]．スパース成分を非負値に制限するため，VB-RNMF では，gamma 分布の rate パラメータ $\boldsymbol{\beta}^s$ に Jeffreys 超事前分布を置くことでスパース性を表現する:

$$p(\mathbf{S}|\alpha^s,\boldsymbol{\beta}^s) = \prod_{f,t} \mathcal{G}(s_{ft}|\alpha^s,\beta^s_{ft}) \tag{6}$$

$$p(\beta^s_{ft}) \propto (\beta^s_{ft})^{-1} \tag{7}$$

ここで，$\alpha^s \in \mathbb{R}_+$ は gamma 分布の超パラメータを表す．VB-RNMF では，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．

3.3 多チャネル音響信号のための VB-MRNMF

本節では，前節で定義した VB-RNMF に基づいて，その多チャネル拡張である変分ベイズ多チャネルロバスト NMF (VB-MRNMF) を定式化する．VB-MRNMF では，目的音声 $\boldsymbol{s}_t \in \mathbb{R}_+^F$ とその各マイクロホン m での観測 $\boldsymbol{y}'_{mt} \in \mathbb{R}_+^F$ との関係を，以下に示す周波数非依存時変線形システムと仮定する．

$$\boldsymbol{y}'_{mt} \approx g_{mt}\boldsymbol{s}_t \tag{8}$$

ここで，$g_{mt} \in \mathbb{R}_+$ は各マイクロホン m および時刻 t での目的音声の音量を表す．本仮定に基づき，VB-MRNMF では各入力チャネルの振幅スペクトログラム $\mathbf{Y}_m = [\boldsymbol{y}_{m1}, \ldots, \boldsymbol{y}_{mT}]$ を，各チャネル独立の低ランク振幅スペクトログラム (走行雑音) とチャネル間共通のスパース振幅スペクトログラム (目的音声) $\mathbf{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_T] \in \mathbb{R}_+^{F\times T}$ に分解する．

$$\boldsymbol{y}_{mt} \approx \mathbf{W}_m\boldsymbol{h}_{mt} + g_{mt}\boldsymbol{s}_t, \tag{9}$$

ここで，$\mathbf{W}_m \in \mathbb{R}_+^{F\times K}$ および $\mathbf{H}_m = [\mathbf{h}_{m1}, \ldots, \mathbf{h}_{mT}] \in \mathbb{R}_+^{K\times T}$ は，各チャネルの低ランク成分を表現する基底およびアクティベーション行列を表す．

3.3.1 尤度関数および事前分布

尤度関数および事前分布は，音量パラメータ g_{mt} を除き，VB-RNMF と同様に定義する (式 3–7)．音量パラメータ 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，音量パラメータ g_{mt} のマイクロホン間でのばらつき度合いを表す超パラメータである．

3.4 変分ベイズ法に基づく推論

本節で述べる VB-MRNMF の推論の目的は，未知パラメータの真の事後分布 $p(\mathbf{W}_{1:m},\mathbf{H}_{1:m},\boldsymbol{g}_{1:m},\mathbf{S},\boldsymbol{\beta}^s|\mathbf{Y}_{1:m})$ 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 (Variational Bayes: VB) に基づいた近似推論を行う [23]．以降，$\boldsymbol{\Theta}$ を全てのパラメータの集合とし，$q(x)$ を変分事後分布とする．真の事後分布は以下のように各パラメータの変分事後分布で近似される: $p(\boldsymbol{\Theta}|\mathbf{Y}_{1:M}) \approx \left\{\prod_m q(\mathbf{W}_m)q(\mathbf{H}_m)q(\boldsymbol{g}_m)\right\} q(\mathbf{S})q(\boldsymbol{\beta}^s)$．本仮定に基づき，各変分事後分布は真の事後分布と変分事後分布の間の KL 距離を最小化し推定する．

VB-MRNMF で用いる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 [23]．以降，$\langle x \rangle$ を x の事後分布の期待値すると，各変分事後分布は，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．

$$
\begin{aligned}
q(w_{mfk}) &= \mathcal{G}(\alpha^{wh} + \sum_t y_{mft}\lambda^{wh}_{mftk}, \beta^{wh} + \sum_t \langle h_{mtk}\rangle),\\
q(h_{mtk}) &= \mathcal{G}(\alpha^{wh} + \sum_f y_{mft}\lambda^{wh}_{mftk}, \beta^{wh} + \sum_f \langle w_{mfk}\rangle),\\
q(g_{mt}) &= \mathcal{G}(\alpha^{g} + \sum_f y_{mft}\lambda^{gs}_{mft}, \alpha^{g} + \sum_f \langle s_{ft}\rangle),\\
q(s_{ft}) &= \mathcal{G}(\alpha^{s} + \sum_m y_{mft}\lambda^{gs}_{mft}, \langle\beta^s_{ft}\rangle + \sum_m \langle g_{mt}\rangle),\\
q(\beta^s_{ft}) &= \mathcal{G}(\alpha^{s}, \langle s_{ft}\rangle),\\
\lambda^{wh}_{mftk} &= \frac{\mathbb{G}[w_{mfk}]\mathbb{G}[h_{mtk}]}{\sum_k \mathbb{G}[w_{mfk}]\mathbb{G}[w_{mtk}] + \mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]},\\
\lambda^{gs}_{mft} &= \frac{\mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]}{\sum_k \mathbb{G}[h_{mfk}]\mathbb{G}[h_{mtk}] + \mathbb{G}[g_{mt}]\mathbb{G}[s_{ft}]},
\end{aligned}
$$

ここで，$\mathbb{G}[x]$ は x の幾何平均を表し，λ^{wh}_{mftk} と λ^{gs}_{mft} は Jensen の不等式による補助変数を表す．

4. 評価実験

本章では，実際に柔軟索状レスキュー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．

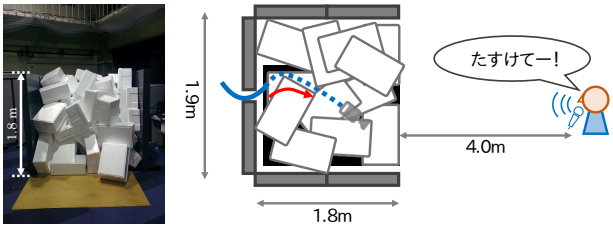
4.1 実　　装

図 1 に，マイクロホンアレイを搭載した柔軟索状ロボットの写真を示す．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．M=8 のマイクロホンをロボット表面に 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．本ロボットは，Fukuda らの Tube-type Active Scope Camera [2] と同様，繊毛と振動モータを用いた駆動で前進する．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．

VB-MRNMF の各パラメータは以下を使用した．STFT のシフト長と窓長はそれぞれ，160 サンプル，1024 サンプルとした．超パラメータ α^{wh}, β^{wh}, α^{g}, および α^{s} は，それぞれ，1.0, 1.0, 5.0, 0.7 とした．基底数 K は 20 とした．これらのパラメータは実験的に決定した．VB-MRNMF の推論結果は 200 回反復更新することで得た．

4.2 実験 1: 実録音を用いた評価

本節では，模擬瓦礫内にロボットを侵入させて収録した実録音の音響信号による評価を報告する．



(a) 模擬瓦礫　　(b) 模擬瓦礫と被験者の配置

図 4: 実験 1 における模擬瓦礫と被験者の実験条件．

4.2.1 実 験 設 定

音声の伝搬を阻害する瓦礫を模擬するために，図 4-(a) のように発泡スチレン箱を積み上げた模擬瓦礫を制作した．図 4-(b) に示すように，男性の被験者にこの瓦礫から 4 m 離れた地点に座ってもらい，「たすけてー」，「おーい」，「だれかー」などと助けを求める声を発してもらった．ロボットは，瓦礫の背後から挿入され，走行雑音と目的音声の混合音である 8 チャネル音響信号を 60 秒間録音した．また，被験者の口元のマイクロホンを用いてリファレンス録音を個別に収録した．

ロボット上のマイクロホンで観測される目的音声の正解データを得ることは不可能なので，本実験では以下で計算される信号対雑音比 (Signal-to-Noise Ratio: SNR) を用いて評価した:

$$
\mathrm{SNR}(\hat{\mathbf{S}}, \mathbf{S}, \alpha) = 10\log_{10}\frac{\sum_{f,t}\alpha^2 s^2_{ft}}{\sum_{f,t}(\hat{s}_{ft} - \alpha s_{ft})^2}, \qquad (11)
$$

ここで，$\mathbf{S} \in \mathbb{R}_+^{F\times T}$ と $\hat{\mathbf{S}} \in \mathbb{R}_+^{F\times T}$ はそれぞれ，リファレンス音声と音声強調結果の振幅スペクトログラムである．また，α は，$\mathbf{S}$ と $\hat{\mathbf{S}}$ の音量差を表すパラメータで，$\alpha\mathbf{S}$ と $\hat{\mathbf{S}}$ の間での最小 2 乗誤差推定 (Minimum Mean-Square Error: MMSE) で得た．観測音響信号の SNR は −14.7 dB であった．

VB-MRNMF と VB-RNMF は，RPCA [11]，多チャネル音源分離の従来法である MNMF [17] および IVA (Independent Vector Analysis) [15]，スペクトル減算法の一種である HRLE (Histogram-based Recursive Level Estimation) [24] と比較した．MNMF および IVA に指定する音源数は 8 個とした．MNMF と IVA は目的音源と走行雑音を区別できないので，これらの SNR は 8 つの分離音のうち最も高い SNR の結果を用いた．VB-RNMF と RPCA, HRLE の結果は先端のマイクロホンを用いて得た．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) [3] とも比較を行った．

4.2.2 実 験 結 果

表 1 に示すように，SNR 向上量において VB-MRNMF がもっとも性能が良い．2 番目に性能がよい Med-RPCA と比較すると，SNR が 2.7 dB 向上している．VB-MRNMF と VB-RNMF を比較すると，提案法の多チャネル拡張によって SNR が 3.8 dB 向上した．また，図 5 に観測信号 (先端マイクロホン) と VB-MRNMF による強調音声の振幅スペクトログラムを示す．本結果は，提案法が時間変化する走行雑音を抑圧でき

表 1: 実験 1 における SNR 向上量 (dB)

VB-MRNMF (3.3 節)	VB-RNMF (3.2 節)	Med-RPCA [3]	RPCA [11]	MNMF [17]	IVA [15]	HRLE [24]
4.29	0.49	1.62	-0.57	-0.18	0.02	-0.89

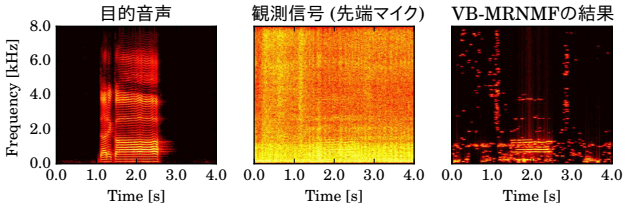


図 5: 実験 1 における音声強調結果例.

ていることを示している.

4.3 実験 2: シミュレーション混合音を用いた評価

本節では，提案法を詳細に評価するために行った，シミュレーション混合音を用いた実験を報告する.

4.3.1 実 験 設 定

本実験では走行雑音と目的音声をそれぞれ独立に収録し，SNR を −20 dB から +5 dB まで 5 dB 刻みで変動させながら混合して評価を行った．図 6 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 種類で評価した.

1) Open-Front: ロボットを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 であった.

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 であった.

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．目的音声はそれぞれ 1 分間の 2 種の男声と 2 種の女声からなる，計 4 分間の録音信号を用いた．本実験では，目的音声の収録時にはロボットは静止していたので，目的音源は本実験では静止している．評価尺度には信号対歪比 (Signal-to-distortion ratio: SDR) [25, 26] を用いた．SDR は総合的な分離精度を表し，計算には Python tool-kit の MIR-EVAL [26] を用いた.

4.3.2 実 験 結 果

図 7 に示すように，Open-Front および Open-Right 条件では，VB-MRNMF が最も高い SDR となった．一部のマイクロホンが隠れている Door-4ch および Door-2ch の条件では，従来の多チャネル音源分離法 (MNMF, IVA，および Med-RPCA) は単チャネルの強調法より性能が劣化している．VB-MRNMF もこれらの条件では，SDR が低下しているが，Door-4ch 条件では SNR が −20 dB のときを除き，単チャネル強調法である VB-RNMF と同程度の性能となった.

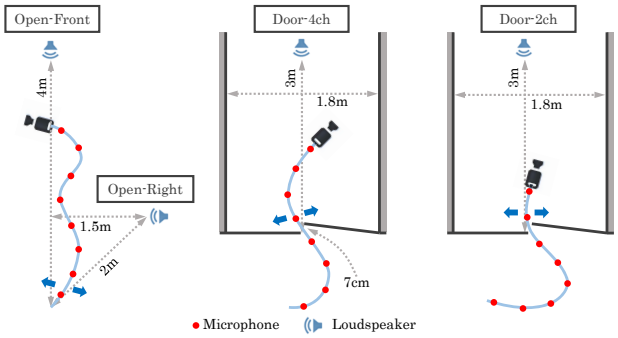


図 6: 実験 2 で用いたスピーカとロボットの配置条件

図 8 に，VB-MRNMF により推定されたスパース成分の各チャネルでの (時変) 音量を示す．Door-2ch および Door-4ch の条件では，ドアで隠れているマイクロホン (Door-2ch では 1 番目から 4 番目，Door-4ch では 1 番目から 6 番目のマイクロホン) において，明らかに音量が小さい．この結果より，スパース成分の音量推定結果が各マイクロホンの信頼度推定に利用できると考えられる.

VB-MRNMF は，信頼度に基づくマイクロホンの選択機構を導入することで，更なる精度向上が期待できる．2 つのマイクロホンのみ音声を大きく収録できていた Door-2ch 条件では，VB-MRNMF の SDR より，先端のマイクのみを用いる VB-RNMF の方が高かった．本結果より，有効なマイクロホン数が少ない時は，マイクロホンの選択機構によって性能が向上できると考えられる．本拡張は，β 過程 NMF [27] の導入により実現できる.

5. お わ り に

本稿では，多チャネルのブラインド音声強調法である変分ベイズ多チャネルロバスト NMF (VB-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，多チャネル音響信号からスパース成分 (目的音声) と低ランク成分 (走行雑音) を分離するベイズモデルに基づく音声強調法を開発した．本手法では，振幅スペクトログラム上で処理を行うため移動問題の影響を軽減でき，各マイクロホンの目的音源の音量を推定するため遮蔽問題に対処できる．8 チャネルのマイクロホンアレイを搭載した 3m の柔軟索状レスキューロボットを用いた評価実験で，信号対雑音比が従来のブラインド音源分離法に比べて 2.7 dB 向上することを確認した．提案法は，雑音の低ランク性と音声のスパース性のみ仮定するため，例えばドローンの飛行雑音抑圧など，他のロボットにも事前学習無しで容易に適用できる．今後は，実用に不可欠なリアルタイム処理のために，オンライン更新則の導出を行う.

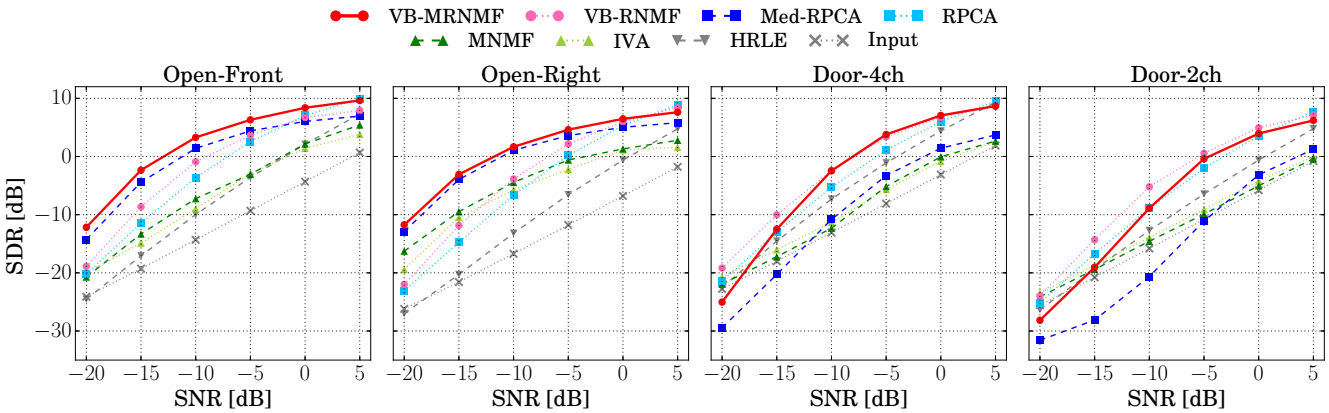


図 7: 実験 2 における音声強調結果の SDR

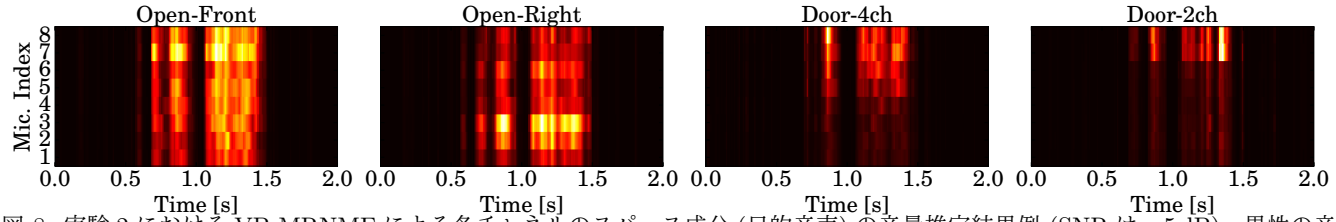


図 8: 実験 2 における VB-MRNMF による各チャネルのスパース成分 (目的音声) の音量推定結果例 (SNR は −5 dB)．男性の音声が 0.5 秒から 1.5 秒の区間で再生されている．

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．

文　献

[1] Robin R. Murphy. *Disaster Robotics*. MIT Press, 2014.

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

[3] 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.

[4] A. Deleforge et al. Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures. In *IEEE ICASSP*, pages 355–359, 2015.

[5] B. Cauchi et al. Reduction of non-stationary noise for a robotic living assistant using sparse non-negative matrix factorization. In *SMIAE*, pages 28–33, 2012.

[6] K. Furukawa et al. Noise correlation matrix estimation for improving sound source localization by multirotor UAV. In *IEEE/RSJ IROS*, pages 3943–3948, 2013.

[7] G. Ince et al. Assessment of general applicability of ego noise estimation. In *IEEE ICRA*, pages 3517–3522, 2011.

[8] Yoshiaki Bando et al. Microphone-accelerometer based 3D posture estimation for a hose-shaped rescue robot. In *IEEE/RSJ IROS*, pages 5580–5586, 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] Yosuke Tatekura et al. Sound source separation with shaded microphone array. *JARP*, 3(2), 2013.

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

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

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

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

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

[16] Alexey Ozerov et al. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE TASLP*, 18(3):550–563, 2010.

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

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

[19] H. Chiba et al. Amplitude-based speech enhancement with nonnegative matrix factorization for asynchronous distributed recording. In *IWAENC*, pages 203–207, 2014.

[20] Meng Sun et al. Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback–Leibler divergence. *IEEE/ACM TASLP*, 23(7):1233–1242, 2015.

[21] X. Ding et al. Bayesian robust principal component analysis. *IEEE TIP*, 20(12):3419–3430, 2011.

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

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

[24] H. Nakajima et al. An easily-configurable robot audition system using histogram-based recursive level estimation. In *IEEE/RSJ IROS*, pages 958–963, 2010.

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

[26] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. PW Ellis. mir eval: a transparent implementation of common MIR metrics. In *ISMIR*, pages 367–372, 2014.

[27] Dawen Liang et al. Beta process non-negative matrix factorization with stochastic structured mean-field variational inference. *arXiv preprint arXiv:1411.1804*, 2014.