社団法人　人工知能学会
Japanese Society for
Artificial Intelligence

人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B002-8

ICA に基づく音声対話ロボット雑音抑圧における確率統計モデルを用いたパーミュテーション解決法

Permutation solver using probability statistics model for ICA-based noise reduction in spoken dialogue robot

† 平田将久，† 脇坂龍，†† 八田俊之，† 猿渡洋，† 鹿野清宏，††† 高谷智哉

†Nobuhisa Hirata, †Ryo Wakisaka, ††Toshiyuki Hatta, †Hiroshi Saruwatari
†Kiyohiro Shikano and †††Tomoya Takatani

† 奈良先端科学技術大学院大学　†Nara Institute of Science and Technology
†† 大阪府立工業高等専門学校　††Osaka Prefectural College of Technology
††† トヨタ自動車株式会社　†††TOYOTA MOTOR CORPORATION

nobuhisa-h@is.naist.jp

Abstract

In this paper, first, a new permutation solving method using probability statistics model is proposed for realizing high performance ICA-based noise reduction used in a spoken dialogue robot. In this method, a shape difference between probability density functions of sources can cope with the permutation problem in realistic sound mixtures consisting of point-source speech and diffuse noise. Next, to achieve high recognition accuracy for the early utterance of the target speaker, we introduce a new rapid ICA initialization method combining image information and a pre-stored initial separation filter bank. The experimental results show that the proposed approaches can remarkably improve the word recognition accuracy in the real-time ICA-based noise reduction developed in the robot dialogue system.

1　はじめに

近年，人と音声コミュニケーションができる音声対話ロボットの研究が盛んに行われている．しかし，実環境下においてロボットから離れた位置から対話ができるハンズフリー音声対話システムを実現する際，環境雑音によって音声認識率が低下するという問題点がある．従来の雑音抑圧技術として独立成分分析 (independent component analysis: ICA) [1]があるが，ICA は音声と環境雑音が混合した信号から環境雑音を推定する能力が高いことがわかっている [2]．そこで，Takahashi らはブラインド空間的サブトラクションアレー (blind spatial subtraction array: BSSA) [3]という雑音抑圧手法を提案している．BSSA は，環境雑音を含んだ観測信号から，ICA によって推定した環境雑音をスペクトル減算 (spectral subtraction: SS) [4]することで目的音抽出を行う手法であり，リアルタイム化も行われている [5]．しかしリアルタイム BSSA では，ユーザ方位情報が未知であるため，いかなる方位のユーザに対しても，ICA における分離フィルタ初期値として正面方位の死角ビームフォーマ (null beamformer: NBF) [6]を使用せざるを得ない．更に ICA にて精度良く雑音推定するには，ある程度分離フィルタの学習時間が必要である．また，ICA は信号間の独立性のみを用いて分離を行うため，分離信号における順序の不定性の問題（パーミュテーション問題）が生じる．従って，異なる周波数毎に ICA を行う周波数領域 ICA (FDICA) では，この問題が生じ，分離信号を大きく歪ませてしまう可能性がある．従って，フィルタの学習が収束するまでに入力される信号に対しては雑音抑圧性能が低く，ロボット音声対話におけるユーザの第一発話目の音声認識率が極めて低い．

上記を解決するため本研究では，ロボットにはカメラが搭載されていて，そのカメラの画像情報からユーザ方位情報を瞬時に推定できると仮定し，予め過去に学習した ICA フィルタを得られたユーザ方位情報にタグ付けをして保存することでフィルタバンクを作成し，そのフィルタバンクに存在する話者方位の ICA フィルタを初期値として使うことで，ロボット音声対話におけるユーザの第一発話目の音声認識率の向上を目指す．また，ICA におけるパーミュテーション問題解決として，音声と拡散性雑音の分離問題に対応させるため，ガンマ分布に分離信号をフィッティングさせる方法を提案する．

2　ICA を用いた目的音声抽出

2.1　ICA による雑音推定

本稿では，点音源で近似される一つの目的信号と，点音源で近似されない雑音信号がある環境を想定する．このような環境の場合，ICA は目的信号を推定するよりも，雑音信号を推定する精度のほうが高いということが明らかになっている [2]．マイクロホン数を J とすると，時間

周波数領域における観測信号は以下のように表現できる．

$$\boldsymbol{x}(f,\tau) = \boldsymbol{h}(f,\theta)s(f,\tau,\theta) + \boldsymbol{n}(f,\tau) \tag{1}$$

ここで，$\boldsymbol{x}(f,\tau) = [x_1(f,\tau),\ldots,x_J(f,\tau)]^{\mathrm{T}}$ は観測信号ベクトル，$\boldsymbol{h}(f,\theta) = [h_1(f,\theta),\ldots,h_J(f,\theta)]^{\mathrm{T}}$ は，目的音源から各マイクロホンへの伝達関数ベクトル，$s(f,\tau,\theta)$ は目的信号，$\boldsymbol{n}(f,\tau) = [n_1(f,\tau),\ldots,n_J(f,\tau)]^{\mathrm{T}}$ は加法性の雑音信号ベクトルを示す．ただし，f は周波数領域番号，τ は分析フレーム番号，θ は画像情報に基づいて推定された目的信号方位を表す．FDICA では，観測信号を以下の式に基づいて分離を行う．

$$\boldsymbol{o}(f,\tau,\theta) = \boldsymbol{W}_{\mathrm{ICA}}(f,\theta)\boldsymbol{x}(f,\tau) \tag{2}$$

$$\boldsymbol{o}(f,\tau,\theta) = [o_1(f,\tau,\theta),\ldots,o_K(f,\tau,\theta)]^{\mathrm{T}} \tag{3}$$

ここで $\boldsymbol{o}(f,\tau,\theta)$ は分離信号ベクトル，K は出力音源数，$\boldsymbol{W}_{\mathrm{ICA}}(f,\theta)$ は θ 方位の信号をキャンセルするための分離行列である．分離行列は以下の更新式に基づいて反復的に求められる．

$$\boldsymbol{W}_{\mathrm{ICA}}^{[p+1]}(f,\theta) = \mu[\boldsymbol{I} - \langle\boldsymbol{\varphi}(\boldsymbol{o}(f,\tau,\theta))\boldsymbol{o}^{\mathrm{H}}(f,\tau,\theta)\rangle_\tau]\boldsymbol{W}_{\mathrm{ICA}}^{[p]}(f,\theta) + \boldsymbol{W}_{\mathrm{ICA}}^{[p]}(f,\theta) \tag{4}$$

ここで p は反復回数，μ はステップサイズ，$\boldsymbol{M}^{\mathrm{H}}$ は行列 $\boldsymbol{M}$ の複素共役転置，$\langle\rangle_\tau$ は時間平均，$\phi(\cdot)$ は非線形関数ベクトルを表す．雑音推定を行うため，分離信号ベクトルから，目的音推定信号 $o_U(f,\tau,\theta)$ を以下のように取り除いた信号ベクトル $\boldsymbol{q}(f,\tau,\theta)$ を得る．

$$\boldsymbol{q}(f,\tau,\theta) = [o_1(f,\tau,\theta),\ldots,o_{U-1}(f,\tau,\theta),0, o_{U+1}(f,\tau,\theta),\ldots,o_K(f,\tau,\theta)]^{\mathrm{T}} \tag{5}$$

次に射影法によって，利得の正規化を行う．この処理は以下の式によって与えられる．

$$\hat{\boldsymbol{q}}(f,\tau,\theta) = [\hat{q}_1(f,\tau,\theta),\ldots,\hat{q}_J(f,\tau,\theta)]^{\mathrm{T}} \tag{6}$$

$$= \boldsymbol{W}_{\mathrm{ICA}}^{+}(f,\theta)\boldsymbol{q}(f,\tau,\theta) \tag{7}$$

ここで，$\boldsymbol{M}^{+}$ は行列 $\boldsymbol{M}$ の Moore-Penrose 型一般逆行列を表す．ICA では，信号間の独立性のみを用いて分離を行うため，分離信号における順序の不定性の問題（パーミュテーション問題）が生じる．従って，異なる周波数毎に ICA を行う FDICA では，この問題が生じ，分離信号を大きく歪ませてしまう可能性がある．

2.2 目的音声抽出

目的音声抽出におけるポスト処理として，本研究では Wiener filter (WF) [10]を使用する．ICA によって推定した雑音信号を用いて，以下のように各チャネル毎に WF のゲイン係数を得る．

$$g_j(f,\tau,\theta) = \frac{|x_j(f,\tau,\theta)|^2}{|x_j(f,\tau,\theta)|^2 + \beta|\hat{q}_j(f,\tau,\theta)|^2} \tag{8}$$

ここで $g_j(f,\tau,\theta)$ は j チャネルにおけるゲイン係数，β は雑音抑圧の処理強度パラメータを表す．最終的に，各チャネル毎にゲイン係数 $g_j(f,\tau,\theta)$ をマイクロホンの観測信号に適用することで，以下のように推定目的信号を得る．

$$s_j^{(\mathrm{WF})}(f,\tau,\theta) = \sqrt{g_j(f,\tau,\theta)|x_j(f,\tau,\theta)|^2}\,\frac{x_j(f,\tau,\theta)}{|x_j(f,\tau,\theta)|} \tag{9}$$

ここで，$s_j^{(\mathrm{WF})}(f,\tau,\theta)$ は j チャネルにおける推定目的音声信号を表す．最後に，WF によって得られた各チャネル毎の推定目的音声信号に対して，遅延話法 (delay and sum: DS) により目的音声強調を行い，最終出力音声信号を得る．

$$s_{\mathrm{DS}}(f,\tau) = \boldsymbol{w}_{\mathrm{DS}}(f,\theta)^{\mathrm{T}}[s_1^{(\mathrm{WF})}(f,\tau,\theta),\ldots,s_J^{(\mathrm{WF})}(f,\tau,\theta)]^{\mathrm{T}} \tag{10}$$

$$\boldsymbol{w}_{\mathrm{DS}}(f,\theta) = [w_1^{(\mathrm{DS})}(f,\theta),\ldots,w_J^{(\mathrm{DS})}(f,\theta)]^{\mathrm{T}} \tag{11}$$

$$w_j^{(\mathrm{DS})}(f,\theta) = \frac{1}{J}\exp\left(-i2\pi(f/N)f_{\mathrm{s}}d_j\sin\theta/c\right) \tag{12}$$

ここで $s_{\mathrm{DS}}(f,\tau)$ は最終出力音声信号，$\boldsymbol{w}_{\mathrm{DS}}(f,\theta)$ は DS のフィルタ係数ベクトル，θ は DS の目的音声方位を表し，ロボットのカメラの画像情報から得られるユーザ方位である．ここで f_{s} はサンプリング周波数，d_j $(j = 1,\ldots,J)$ はマイクロホン位置，N は DFT 長，c は音速を表す．

3 提案法 1: ガンマ分布に基づくパーミュテーション解決

パーミュテーション問題の解決法として様々な提案がなされている [6] [7] [8]．これらは点音源である音声と音声の分離問題においては有効な解決法であるが，音声と拡散性の環境雑音の分離問題においては，うまく機能しない．そこで本研究では，ICA におけるパーミュテーション問題を，ガンマ分布に信号をモデリングすることで解決する方法を提案する．本手法は分離信号の確率統計量を用いるので音声と拡散性雑音の分離問題にも有効であると考えられる．ガンマ分布は，一般に，パワースペクトル領域の音声信号や実環境の雑音信号を表現可能であると言われている [9]．また，ガンマ関数に基づく分布であるので，数学的に有用な性質が多く，高次統計量を表現する目的にも利用しやすい特徴を持つ．ガンマ分布の確率密度関数 (PDF) は，

$$P(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\theta}} \tag{13}$$

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx \tag{14}$$

と表せる．ここで，$x \geq 0$ は信号のパワースペクトル系列であり，$\alpha > 0$ かつ $\theta > 0$ である．また，α は形状母

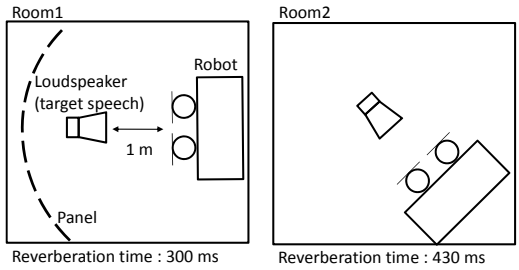

Figure 1: Layout of two reverberant rooms used in our simulation.

数，θ は尺度母数，$\Gamma(\alpha)$ はガンマ関数である．$\alpha = 1$ の場合，式 (13) は指数分布と一致することが知られており，これはガウス性信号のパワースペクトルに対応する．また，$0 < \alpha < 1$ の場合は，優ガウス性信号であることを示す．ガンマ分布の平均値は，以下の式で表現できる．

$$\mathrm{E}[P(x)] = \alpha\theta \tag{15}$$

ここで $\mathrm{E}[\cdot]$ は期待値演算子である．ガンマ分布によるモデリングは，観測される生のデータサンプルから，形状母数 α と尺度母数 θ を推定することで行われる．これらの母数は，以下のように最尤推定法に基づき推定される．

$$\hat{\alpha} = \frac{3 - \gamma + \sqrt{(\gamma - 3)^2 + 24\gamma}}{12\gamma} \tag{16}$$

$$\hat{\theta} = \frac{\mathrm{E}[x]}{\hat{\alpha}} \tag{17}$$

ここで $\gamma = \log(\mathrm{E}[x]) - \mathrm{E}[\log x]$ である．推定された $\hat{\alpha}$ の値が小さい程，優ガウス性が高い分布形状となり，$\hat{\alpha} = 1$ のとき，ガウス性の分布形状となる．一般に音声の PDF は優ガウス性の分布形状であり，拡散性雑音の PDF はガウス性の分布形状であると言われている．よって，それぞれの分離信号を用いて $\hat{\alpha}$ の値を求め，その大小を比較することによってパーミュテーション問題を解決する．

以上の手法の有効性を確認するために，予備実験として ICA による音源分離実験を行った．Fig.1 の Room2 で収録したインパルス応答をクリーン音声データベースに畳み込み，入力 SNR が 10 dB となるように拡散性雑音を付加した．拡散性雑音は，実収録による人ごみ雑音を用いた．マイクロホンアレーには 2 素子を用い，音声とマイクロホンアレーの距離は 1.0 m とした．比較手法として以下の 4 つを評価した．

- パーミュテーション問題未解決 (unprocessed).
- 方位特性に基づく解決法 (DOA-based) [6].
- ガンマ分布に基づく解決法 (proposed).
- 真の分離信号を用いる理想的な解決 (ideal).

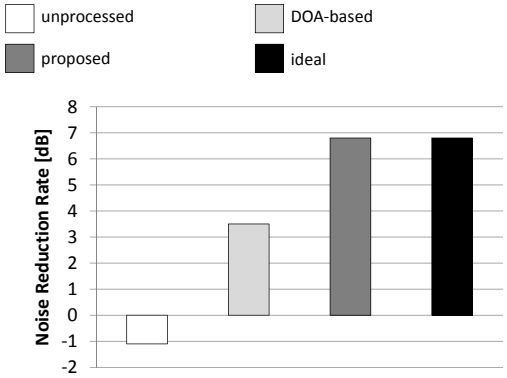

Figure 2: result of preliminary experiment 1.

また，分離性能の評価には雑音抑圧量 (noise reduction rate: NRR) を用いた．NRR は値が高いほど良い性能を示す．実験結果を Fig 2 に示す．この結果より，本稿のような音声と拡散性雑音の分離問題におけるパーミュテーション問題の解決法としては，提案法が有効であることが確認できる．

4 提案法 2: 画像情報に基づく第一発話処理

4.1 概要

本研究では，音声対話ロボットにはカメラが搭載されており，カメラから得られる画像情報からユーザ方位を瞬時に推定できると仮定する．予め過去に学習した ICA フィルタをユーザ方位にタグ付けをして保存することでフィルタバンクを作成し，そのフィルタバンクに存在する話者方位の ICA フィルタを初期値として使うことで雑音を推定する．さらに推定した雑音を用いて，マイクロホンアレーで観測した信号に WF を適用することで目的音抽出を行い，最後に DS によって目的音声強調をする手法を提案する．提案法における ICA のパーミュテーション解決法としては，ガンマ分布に基づく解決法を用いる．これらの処理によって得られた目的音声を Julius [11]によって機械音声認識することで，提案法の有効性を示す．アルゴリズムの詳細を以下で述べる．

4.2 予備実験

ここで，ある部屋で予め ϕ 方位のユーザに対して学習を行い $\boldsymbol{W}_{\mathrm{ICA}}(f, \phi)$ を保存し，それとは別の部屋で ϕ 方位のユーザに対して保存した $\boldsymbol{W}_{\mathrm{ICA}}(f, \phi)$ を用いて目的音声抽出を行う場合を考える．ICA は信号間の独立性のみを用いて分離学習を行うが，結果的には部屋の残響特性を含めたユーザ方位の音を抑圧するフィルタを学習している．よって，部屋の残響特性が変化すると過去に学習した ICA フィルタ $\boldsymbol{W}_{\mathrm{ICA}}(f, \phi)$ を用いても精度良く雑音推定できるとは限らない．ロボット音声対話を想定した場合，ユーザ方位から到来する音波は，順に直接波，ロボット本体によ

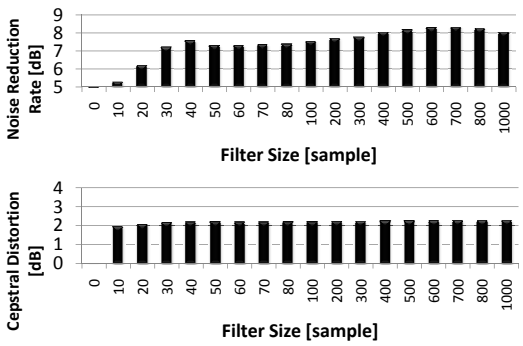

Figure 3: result of preliminary experiment 2.

る回折波，初期反射波，部屋の伝達特性による残響波であるが，部屋が変わることによって伝達特性に変化があるのは，初期反射波以降の部屋の伝達特性による波である．よって，この成分を抑圧する ICA フィルタの一部分は再度 ICA で学習させなければならないが，保存したフィルタをそのまま使って学習させるほうが良いか，初期反射以降の部屋の伝達特性を抑圧する部分を無くすために ICA フィルタを短くして学習させるほうが良いか，またその場合フィルタの長さはどれくらいが良いかを確認しなければならない．そこで最適な ICA フィルタの長さを確認するために予備実験を行った．まず，0 °方位の話者に対して ICA で学習を行い，そのフィルタを長さをそれぞれ操作したものを保存する．そしてそれとは別の残響特性のインパルス応答を用いて観測信号にそれぞれの長さの ICA フィルタを学習させずに適用させ，WF によって目的音声抽出を行う．ICA フィルタ長は最大で 1024 サンプルである．実験条件としては前節の予備実験と同じで，評価値には NRR 及びケプストラム歪み (cepstral distortion: CD) を用いた．CD は値が小さいほど良い性能を示す．この予備実験の結果を Fig. 3 に示す．一定のフィルタ長までは性能が上がっていくが，それ以降は変化がないことが分かる．この結果より，部屋の残響特性の変化による雑音抑圧性能は ICA フィルタの長さによらないということが言える．よって以降，本稿では ICA フィルタを保存する際，フィルタ長を操作しない方法を用いる．

4.3 リアルタイム処理のフロー

本手法では，以下のステップでリアルタイム音声強調を行う．処理フローを Fig. 4 に示す．

Step 1 事前フィルタバンクの構成

本手法では，画像情報から得られたユーザ方位を，正面方位を 0 度とし，-90 度から 90 度まで 15 度間隔で区切った 13 方位のうち，最も近い方位を θ とし，処理に使用する．13 方位全てにおいて，過去に十分学習を行ったユーザ方位 θ に関する ICA の分離行列フィルタ $\boldsymbol{W}_{\mathrm{ICA}}(f,\theta)$ を，3 節で述べたパーミュテーション解決法でパーミュテーション解決を行い，フィルタバンクに保存する．

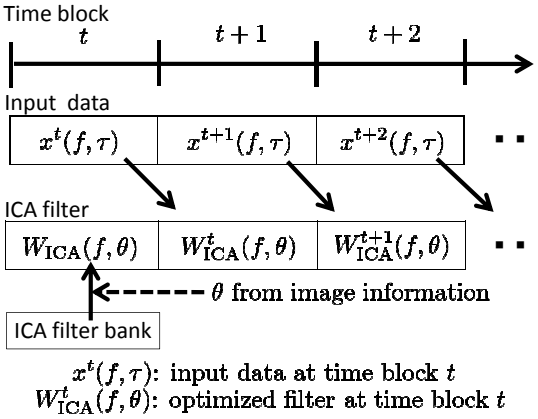

Figure 4: Signal flow of updating ICA filter in real-time simulation.

Step 2 画像による方位推定

ユーザがカメラに映ったときに，ユーザ方位推定を行い θ を取得する．

Step 3 第一発話目に対する処理

Step 2 で取得した θ に基づき，フィルタバンクから $\boldsymbol{W}_{\mathrm{ICA}}(f,\theta)$ を読み込む．それを用いて，現在の入力データブロックに対して雑音推定を行い，推定雑音を用いて入力データに対して各チャネル毎に WF を適用させて雑音抑圧処理を行い，強調音声を出力する．最後に DS をすることで目的音声強調を行い，最終出力音声信号を得る．

Step 4 第二発話目以降に対する処理

第二発話目以降については，Fig. 4 のように入力信号を時間ブロックに分け，各ブロックで ICA のフィルタを学習させ，更新していく．このときのパーミュテーション問題についても 3 節で述べた手法でパーミュテーション解決を行う．雑音抑圧処理は Step 3 と同様に行う．

Step 5 ユーザ発話終了後の処理

ユーザの発話が終了したときは，$\boldsymbol{W}_{\mathrm{ICA}}(f,\theta)$ を未来のユーザに対する第一発話目用のフィルタとして，フィルタバンクに上書きをする．ここでも 3 節で述べた手法でパーミュテーション解決を行う．その後，Step 1 へ戻る．

5 音声認識実験

5.1 実験条件

提案法の有効性を確認するため，Julius を用いて機械音声認識を行った．実験は以下に示す 3 つの手法を用いて行った．

- 正面方位の NBF を初期値として雑音推定を行った従来法 (Conventional).
- 画像により取得されたユーザ方位の NBF を初期値として雑音推定を行った手法 (Supervised).
- 提案法 (Proposed).

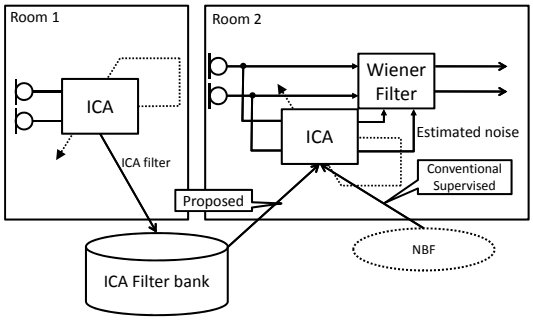

Figure 5: Block diagram of speech recognition experiment.

Table 1: 音声認識実験の条件

テストデータ	JNAS テストセット (男女話者による 200 文)
音声認識タスク	新聞記事読み上げ (語彙数: 20 k)
音響モデル	音素内タイドミクスチャーモデル (phonetic-tied mixture model: PTM) に基づく 25 dB オフィス雑音重畳モデル
音響モデルの学習データ	JNAS 260 話者 (1 話者あたり 150 文)
認識デコーダ	Julius ver. 3.4.2

提案法の処理ブロック図を Fig. 5 に示す．提案法における ICA フィルタのデータベース作成は Fig. 1 の Room1 で行い，実験は Room2 で行う．Room1 と Room2 では，パネルを用いて残響時間を変え，ロボットの位置も変えた．Room1，Room2 のいずれにおいても，各音響環境で収録したインパルス応答を JNAS のクリーン音声に畳み込んだ信号を目的音声信号とした．この信号に対して SNR が 10 dB となるように，実収録の駅環境雑音を付加した．マイクロホンアレーの素子数は 2 個で，SHURE 製の指向性マイクロホン MX-184 を使用した．実験における WF の処理強度パラメータは，いずれの手法においても音声認識精度を基にして最適な値を選んだ．ICA フィルタ学習時の時間ブロック長は 3 s とし，学習回数は 100 回とした．音声認識に実験の条件を表 1 に示す．なお，Conventional 及び Supervised におけるパーミュテーション解決法としては，方位特性に基づく解決法を用いた [6].

5.2 実験結果

Fig. 6 に音声認識結果を示す．この結果より，いずれの話者方位でも過去に学習した ICA フィルタを初期値にし，パーミュテーション解決法としてガンマ分布に基づく解決法を用いた提案法のほうが音声認識率が改善されていることがわかる．Conventional と Supervised の認識率がそれほど変わらない原因は，マイクロホンアレーに指向性マイクロホンを使用したこと及び目的信号のロボット本体による回折波成分が大きいため，NBF の死角が正しく形成されていないことが原因であると考えられる．従って，提案法はロボット音声対話におけるユーザの第一発話目の認識率向上に有効であると言える．

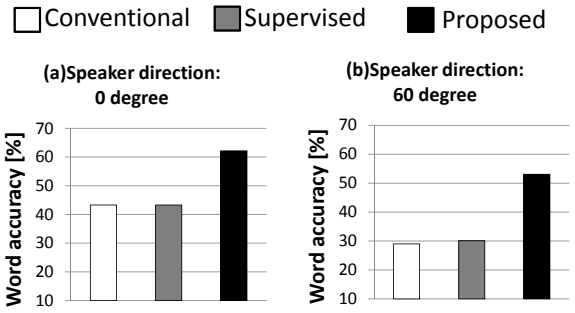

Figure 6: Word accuracy of different speaker directions; (a) 0 degrees and (b) 60 degrees.

6 まとめ

本稿では，音声対話ロボットにおいてユーザ方位が瞬時に推定できたときの，リアルタイムを想定した雑音抑圧におけるパーミュテーション解決法を提案した．音声認識実験により提案法の有効性を確認した．今後はさらなる音声認識率の向上を目指す．

謝辞 本研究の一部は総務省・戦略的情報通信研究開発推進制度 (SCOPE) の支援を受けた．

参考文献

[1] P. Comon, "Independent component analysis, a new concept", *Signal processing,* vol. 36, pp. 287–314, 1994.

[2] Y. Takahashi, et al., "Blind source extraction for hands-free speech recognition based on wiener filtering and ICA-based noise estimation," *Proc. HSCMA*, 2008.

[3] Y. Takahashi, et al., "Blind spatial subtraction array for noisy environment," *IEEE Trans. Audio, Speech, and Language Processing*, vol.17, no.4, pp.650–664, 2009.

[4] S. F. Boll, *IEEE Trans. Acoustics, Speech, Signal Proc,* vol. ASSP-27, no. 2, pp. 113–120, 1979.

[5] 高橋祐，他 "独立成分分析を導入した空間的サブトラクションアレーによるハンズフリー音声認識システムの開発," 電子情報通信学会論文誌 D, vol.J93-D, no.3, pp.312–325, 2010.

[6] H. Saruwatari, S. Kiumura, K Takeda, F. Itakura, and T. Nishikawa. "Blind source separation combining independent component analysis and beamform-

ing." *EURASIP Journal on Applied Signal Processing,* vol. 2003, pp. 1135–1146, 2003.

[7] N. Murata, S. Ikeda, and A. Ziehe. "An approach to blind source separation based on temporal structure of speech signal." *Neurocomputing,* vol. 41, no. 1-4, pp. 1–24, 2001.

[8] H. Sawada, R. Mukai, S. Araki, and S. Mikano. "A robust and precise method for solving the permutation problem of frequency-domain blind source separation." *IEEE Transactions on Speech and Audio Processing,* vol.12, no. 5, pp. 530–538, 2004.

[9] T. Inoue, Y. Takahashi, H. Saruwatari, K. Shikano and K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction: why should not use power/amplitude subtraction?," *Proc. EUSIPCO European Signal Processing Conference*, pp. 994–998, 2010.

[10] P. Loizou, *Speech Enhancement: Theory and Practice,* CRC Press, 2007.

[11] A. Lee, et al., "Julius An open source realtime large vocabulary recognition engine," *Proc. Eur. Conf. Speech Commun. Technol.*, pp.1691–1694, 2001.