

内部雑音抑圧型ロボット音声対話システムにおけるマイクロホンアレー配置の検討

A Study on Microphone Array Layout for Spoken-Oriented Robot Dialogue System with
Internal-Noise Suppressor

†澤田紘志, †Jani Even, †猿渡洋, †鹿野清宏, †高谷智哉

†Hiroshi Sawada, †Jani Even, †Hiroshi Saruwatari, †Kiyohiro Shikano, and †Tomoya Takatani

†奈良先端科学技術大学院大学

‡トヨタ自動車株式会社

†Nara Institute of Science and Technology

‡TOYOTA MOTOR CORPORATION

hiroshi-s@is.naist.jp

Abstract

本論文では、内部雑音抑圧型ロボット音声対話システムのために有効なマイクロホンアレー配置について提案する。我々は既に、ハンズフリー音声認識のための内部雑音抑圧手法として、独立成分分析 (independent component analysis: ICA) に基づくセミブラインド音源分離 (semi-blind source separation: SBSS) と wiener filter (WF) を統合した手法を提案している。本稿でははじめに、従来のマイクロホンアレー配置では SBSS の雑音推定性能が劣化し、音声認識性能があまり高くないという問題を示す。次に、内部雑音の音源到来方位 (direction of arrival: DOA) を解析し、内部雑音が 0° に定位することを示す。この解析結果に基づき、マイクロホンアレーをブロードサイドアレーからエンドファイアアレーに回転させることを提案する。最後に、提案するマイクロホンアレー配置の有効性を確かめるためにシミュレーション実験を行った。従来に比べて高い音声認識性能を有することを示す。

1 はじめに

人と音声コミュニケーション可能なパートナーロボットでは、ユーザから離れた位置にマイクロホンを設置して音声認識を行うハンズフリー音声認識が必要不可欠である。しかし、実環境下においては、周囲に存在する環境雑音や残響、さらにはファンノイズやモータ音などのロボット自身が発する内部雑音によって、音声認識性能が低下する問題がある。そこで、著者らの一人である Even らによって、内部雑音測定用センサの観測信号を教師情報に用いる内部雑音抑圧手法として、セミブラインド音源分離 (semi-blind source separation: SBSS) [1] と wiener filter (WF) [2] を統合した手法が提案されている [3]。SBSS と WF を統合した手法を Fig. 1 に示す。この手法では、はじめに独立成分

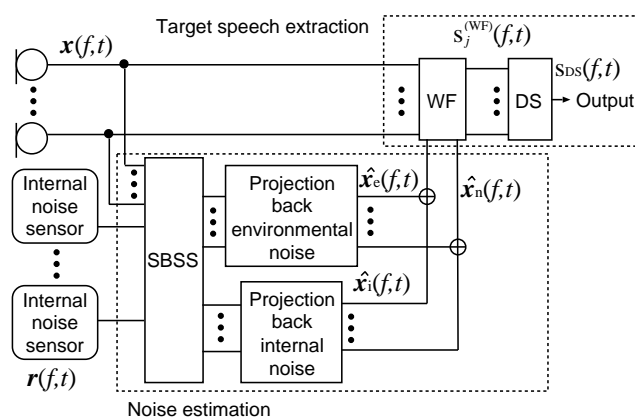


Figure 1: Block diagram of speech extraction method.

分析 (independent component analysis: ICA) [4] に基づく SBSS によって環境雑音と内部雑音を動的に推定する。次に、推定した環境雑音と内部雑音を基に、マイクロホンの観測信号に WF を適用することによって、各チャンネル毎に目的音声抽出を実現する。最後に、各チャンネル毎の目的音声抽出結果を遅延和アレー (delay-and-sum: DS) [5] によって目的音声強調し、最終出力音声信号を得る。しかし、ユーザがロボットの正面に立ち、かつロボットに設置されているマイクロホンアレーがブロードサイドアレーのとき、SBSS の雑音推定性能が劣化し音声認識性能があまり高くないという問題があった。

そこで本論文では、まず、従来のマイクロホンアレー配置における SBSS の雑音推定性能があまり高くないことを示す。次に、この原因を検証するために、内部雑音の音源到来方位 (direction of arrival: DOA) について解析を行い、内部雑音が 0° に定位していることを示す。この結果から、ユーザがロボット正面に立ちロボットに設置されているマイクロホンアレーがブロードサイドアレーのとき、目的音声と内部雑音の DOA が 0° で重なることにより十分な SBSS の雑音推定性能が得られていなかったことを確かめる。この知見に基づき、本稿では、SBSS の

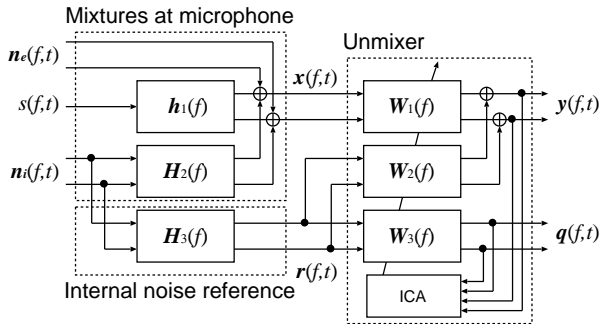


Figure 2: Block diagram of mixing and unmixing at f -th frequency bin.

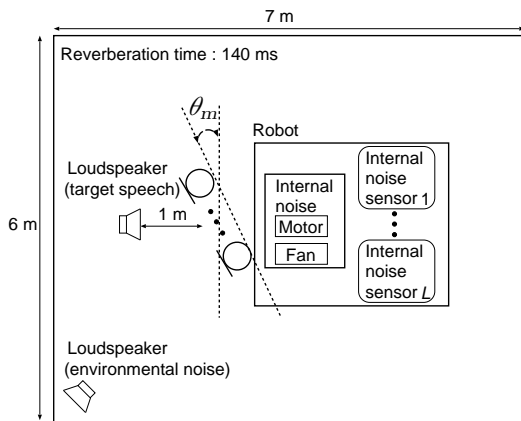


Figure 3: Layout of reverberant room used in our simulation.

雑音推定性能を高めるためにロボットに搭載する直線上マイクロホンアレーをブロードサイドアレーからエンドファイアアレーに回転させ、目的音声と内部雑音のDOAを空間的に異なる方位にすることを提案する。最後に、提案するマイクロホンアレー配置の有効性を確かめるために、シミュレーション実験を行う。提案法よって得られた最終出力音声信号をJulius [6]によって機械音声認識を行い、客観評価から提案法の有効性を示す。

2 セミブラインド音源分離を用いた目的音声抽出

2.1 混合過程

本稿では、点音源で近似できる目的音声信号、点音源で近似されない(非点音源)環境雑音信号、ロボットが発する内部雑音信号が同時に存在する環境を想定する。ここで内部雑音信号とは、ロボットの筐体振動であると推測される。ロボット内部に搭載されているサーボモータやCPUファン等がロボットの筐体を振動させ、筐体に設置されているマイクロホンアレーが振動し、その振動が雑音として観測される。そこで、内部雑音測定用センサをロボットの筐体に設置し、筐体振動である内部雑音信号を測定する。想定する環境の混合過程と分離過程をFig. 2に

示す。以下では混合過程について説明する。

マイクロホン数を J 、内部雑音測定用センサ数を L としたとき、マイクロホンと内部雑音測定用センサにおけるそれぞれの観測信号は以下のように表現できる。

$$x(f, t) = \mathbf{h}_1(f)s(f, t) + \mathbf{n}_e(f, t) + \mathbf{H}_2(f)\mathbf{n}_i(f, t) \quad (1)$$

$$\mathbf{r}(f, t) = \mathbf{H}_3(f)\mathbf{n}_i(f, t) \quad (2)$$

ここで、 f は周波数帯域番号、 t は時間フレーム番号を表す。また、 $\mathbf{x}(f, t) = [x_1(f, t), \dots, x_J(f, t)]^T$ はマイクロホンにおける観測信号ベクトル、 $\mathbf{n}_e(f, t) = [n_1^{(e)}(f, t), \dots, n_J^{(e)}(f, t)]^T$ は環境雑音ベクトル、 $\mathbf{n}_i(f, t) = [n_1^{(i)}(f, t), \dots, n_K^{(i)}(f, t)]^T$ は内部雑音ベクトル (K は内部雑音の音源数)、 $\mathbf{r}(f, t) = [r_1(f, t), \dots, r_L(f, t)]^T$ 内部雑音測定用センサにおける観測信号ベクトル、 $s(f, t)$ は目的音声信号を表す。さらに、 $\mathbf{h}_1(f) = [h_1^{(1)}(f), \dots, h_J^{(1)}(f)]^T$ は目的音声声源から各マイクロホンへの伝達関数ベクトル、 $\mathbf{H}_2(f) = [h_1^{(2)}(f), \dots, h_K^{(2)}(f)]$ は各内部雑音源から各マイクロホンへの伝達関数行列、 $\mathbf{H}_3(f) = [h_1^{(3)}(f), \dots, h_K^{(3)}(f)]$ は各内部雑音源から各内部雑音測定用センサへの伝達関数行列を表す。 $\mathbf{h}_k^{(2)}(f) = [h_{1k}^{(2)}(f), \dots, h_{Jk}^{(2)}(f)]^T$ ($k = 1, \dots, K$) は k 番目の内部雑音源から各マイクロホンへの伝達関数ベクトル、 $\mathbf{h}_k^{(3)}(f) = [h_{1k}^{(3)}(f), \dots, h_{Lk}^{(3)}(f)]^T$ ($k = 1, \dots, K$) は k 番目の内部雑音源から各内部雑音測定用センサへの伝達関数ベクトルを表す。また内部雑音測定用センサは筐体振動のみを観測することができ、空気振動である目的音声信号や環境雑音信号をほとんど観測しないという特徴を持つ。そのため混合過程モデルでは、内部雑音測定用センサは空気振動の成分を全く観測しないものとして定式化される。

2.2 セミブラインド音源分離に基づく雑音推定

ICAに基づくSBSSでは分離行列 $\mathbf{W}_i(f)$ ($i = 1, 2, 3$) を用いて、目的音声と環境雑音の分離信号ベクトル $\mathbf{y}(f, t) = [y_s(f, t), y_n(f, t)]^T$ と内部雑音の分離信号ベクトル $\mathbf{q}(f, t) = [q_1(f, t), \dots, q_K(f, t)]^T$ を、各周波数毎に以下の式で求めることができる [1]。

$$\mathbf{y}(f, t) = \mathbf{W}_1(f)\mathbf{x}(f, t) + \mathbf{W}_2(f)\mathbf{r}(f, t) \quad (3)$$

$$\mathbf{q}(f, t) = \mathbf{W}_3(f)\mathbf{r}(f, t) \quad (4)$$

ここで、 $y_s(f, t)$ は目的音声推定信号、 $y_n(f, t)$ は環境雑音推定信号を示す。観測信号を分離する最適な分離行列 $\mathbf{W}_i(f)$ ($i = 1, 2, 3$) は以下の反復学習式によって求めることができる。

ネルにおける推定目的音声信号 $s_j^{(\text{WF})}(f, t)$ を得る .

$$\mathbf{W}_i^{[k+1]}(f) = \mathbf{W}_i^{[k]}(f) - \mu \Delta \mathbf{W}_i^{[k]}(f) \quad (5)$$

$$\Delta \mathbf{W}_1^{[k+1]}(f) = (\mathbf{I} - \langle \phi(\mathbf{y}(f, t)^{[k]}) \mathbf{y}^{\text{H}}(f, t)^{[k]} \rangle_t) \mathbf{W}_1^{[k]}(f) \quad (6)$$

$$\Delta \mathbf{W}_2^{[k+1]}(f) = (\mathbf{I} - \langle \phi(\mathbf{y}(f, t)^{[k]}) \mathbf{y}^{\text{H}}(f, t)^{[k]} \rangle_t) \mathbf{W}_2^{[k]}(f) - (\langle \phi(\mathbf{y}(f, t)^{[k]}) \mathbf{q}^{\text{H}}(f, t)^{[k]} \rangle_t) \mathbf{W}_3^{[k]}(f) \quad (7)$$

$$\Delta \mathbf{W}_3^{[k+1]}(f) = (\mathbf{I} - \langle \phi(\mathbf{q}(f, t)^{[k]}) \mathbf{q}^{\text{H}}(f, t)^{[k]} \rangle_t) \mathbf{W}_3^{[k]}(f) \quad (8)$$

$\phi(\cdot)$ は非線形関数ベクトルを表し, 分離信号 $\mathbf{y}(f, t)$, $\mathbf{q}(f, t)$ のサンプルデータから, カーネルに基づくスコア関数の推定により求める [7]. また, \mathbf{I} は単位行列, μ は更新係数, $\langle \cdot \rangle_t$ は時間平均演算子, M^{H} は M の複素共役転置, $[\cdot]^{[k]}$ は k 回目の反復学習であることを表す. また, ICA における permutation 問題は, 確率密度分布推定と DOA 推定を組み合わせた手法を用いて解決している [8].

ここで, 非点音源雑音環境下において, ICA の目的音声信号推定精度はあまり優れていないのに対して, 雑音信号推定精度は非常に高いことが知られている [9]. このような事実に基づき, ICA を雑音推定器として用いる. よって, 最適化された分離フィルタを基に以下の式を用いて, 環境雑音信号とマイクロホンで観測される内部雑音信号の推定を行う.

$$\hat{\mathbf{x}}_e(f, t) = \mathbf{W}_1^+(f) [0, y_n(f, t)]^{\text{T}} \quad (9)$$

$$\hat{\mathbf{x}}_i(f, t) = -\mathbf{W}_1^+(f) \mathbf{W}_2(f) \mathbf{W}_3^+(f) \mathbf{q}(f, t) \quad (10)$$

ここで $\hat{\mathbf{x}}_e(f, t) = [\hat{x}_1^{(e)}(f, t), \dots, \hat{x}_J^{(e)}(f, t)]^{\text{T}}$ は推定環境雑音ベクトル, $\hat{\mathbf{x}}_i(f, t) = [\hat{x}_1^{(i)}(f, t), \dots, \hat{x}_J^{(i)}(f, t)]^{\text{T}}$ はマイクロホン地点での推定内部雑音ベクトルを表す.

2.3 目的音声抽出

SBSS によって求めた推定環境雑音ベクトル, 推定内部雑音ベクトル, マイクロホンの観測信号ベクトルを用いて, 各チャンネル毎に WF のゲイン係数設計と適用を行い, 目的音声抽出を実現する. 以下でその処理の詳細を述べる.

抑圧すべき全ての雑音信号は推定環境雑音信号と推定内部雑音信号の和で表され, これを用いて各チャンネル毎に WF のゲイン係数を設計する.

$$\hat{\mathbf{x}}_n(f, t) = \hat{\mathbf{x}}_e(f, t) + \hat{\mathbf{x}}_i(f, t) \quad (11)$$

$$g_j(f, t) = \frac{|x_j(f, t)|^2}{|x_j(f, t)|^2 + \beta |\hat{x}_j^{(n)}(f, t)|^2} \quad (12)$$

ここで, $\hat{\mathbf{x}}_n(f, t) = [\hat{x}_1^{(n)}(f, t), \dots, \hat{x}_J^{(n)}(f, t)]^{\text{T}}$ は全ての推定雑音ベクトル, $g_j(f, t)$ は j チャン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, 各チャンネル毎にゲイン係数 $g_j(f, t)$ 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 j チャ

$$s_j^{(\text{WF})}(f, t) = \sqrt{g_j(f, t) |x_j(f, t)|^2} \frac{x_j(f, t)}{|x_j(f, t)|} \quad (13)$$

最後に, WF によって得られた各チャンネル毎の推定目的音声信号に対して 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.

$$s_{\text{DS}}(f, t) = \mathbf{w}_{\text{DS}}(f, \theta_{\text{U}})^{\text{T}} [s_1^{(\text{WF})}(f, t), \dots, s_J^{(\text{WF})}(f, t)]^{\text{T}} \quad (14)$$

$$\mathbf{w}_{\text{DS}}(f, \theta) = [w_1^{(\text{DS})}(f, \theta), \dots, w_J^{(\text{DS})}(f, \theta)]^{\text{T}} \quad (15)$$

$$w_j^{(\text{DS})}(f, \theta) = \frac{1}{J} \exp(-i2\pi(f/N) f_s d_j \sin \theta / c) \quad (16)$$

ここで $s_{\text{DS}}(f, t)$ は最終出力音声信号, $\mathbf{w}_{\text{DS}}(f, \theta)$ は DS のフィルタ係数ベクトル, θ_{U} は DS の目的音声方位を表し, ICA によって学習した分離行列から推定することが可能である [10]. ここで f_s はサンプリング周波数, d_j ($j = 1, \dots, J$) はマイクロホン位置, N は DFT 長, c は音速を表す.

2.4 セミブラインド音源分離の問題点

従来のマイクロホンアレー配置では, SBSS の雑音推定性能があまり高くないという問題があった. このことを確かめるために, 予備実験を行った. Figure 3 に示す音響環境で収録したインパルス応答を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, 実収録内部雑音信号における静止区間 (CPU ファン雑音のみ) と目的音声信号の信号対雑音電力比 (signal-to-noise ratio: SNR) が 20 dB になるように内部雑音信号を付加した. 目的音声信号と動作区間を含んだ内部雑音信号全体における入力 SNR は 16.6 dB である. さらに, 目的音声信号と環境雑音信号の入力 SNR が 10 dB になるように実収録の環境雑音信号を付加した. マイクロホンの素子数は 4, 内部雑音測定用センサの素子数は 3, $\theta_{\text{m}} = 0^\circ$ とした.

Figure 4 に目的音声信号 10 文において, SBSS を適用したときの環境雑音信号と内部雑音信号による全ての雑音信号におけるスペクトル歪み (spectral distortion: SD) $e(f)$ の平均値を示す. $e(f)$ は以下の式で与えられる.

$$e(f) = 10 \log_{10} \left(\frac{1}{J} \sum_j \sum_t |x_j^{(n)}(f, t) - \hat{x}_j^{(n)}(f, t)|^2 \right) \quad (17)$$

ここで, $x_j^{(n)}$ は j 番目のマイクロホン地点における真の雑音信号ベクトルを表す. SD はスペクトルドメインにおける歪みを表しており, 値が小さいほど歪みが小さく雑音推定性能が良いことを表す指標である. Figure 4 より, 低域の SD が大きく, SBSS の雑音推定性能が高くないことが分かる.

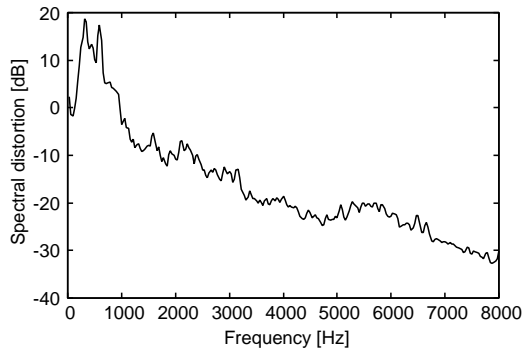


Figure 4: Spectral distortion between components of true noise and estimated noise (averaged on channels and utterances).

3 提案法

3.1 概要

本章では、SBSS の雑音推定性能が高くない原因を検証し、その改善策を提案する。まず、内部雑音信号の DOA を最小分散法 (minimum variance method: MV) によって解析する。この解析より、内部雑音信号の DOA が 0° に定位していることを示す。この結果から、従来のマイクロホンアレー配置では正面の目的音声信号と内部雑音信号の DOA が 0° で重なり、十分な SBSS の雑音推定性能が得られていなかったことを示す。この知見に基づき、SBSS の雑音推定性能を改善するために有効なマイクロホンアレー配置を提案する。

3.2 DOA に基づく内部雑音の解析

はじめに、内部雑音信号の DOA を解析する予備実験を行った。実験における音響環境と条件は 2.4 節と同じである。ただしロボットの内部雑音に関しては、機械音やモータ音を含む異なる動作によって発生した 4 種類について検討を行う。

目的音声信号と内部雑音信号の DOA を解析するために MV を用いる。はじめに、マイクロホンの受信信号に対して MV を適用したときの出力パワー $P(f, \theta)$ を以下のように求める。

$$P(f, \theta) = \frac{1}{\mathbf{a}^H(f, \theta) \mathbf{R}^{-1}(f) \mathbf{a}(f, \theta)} \quad (18)$$

$$\mathbf{R}(f) = \mathbb{E}[\mathbf{z}(f, t) \mathbf{z}^H(f, t)] \quad (19)$$

$$\mathbf{a}(f, \theta) = [a_1(f, \theta), \dots, a_J(f, \theta)]^T \quad (20)$$

$$a_j(f, \theta) = \exp(i2\pi(f/N)f_s d_j \sin \theta/c) \quad (21)$$

ここで、 $\mathbf{a}(f, \theta)$ はステアリングベクトル、 $\mathbf{R}(f)$ は共分散行列、 $\mathbf{z}(f, t) = [z_1(f, t), \dots, z_J(f, t)]^T$ は受信信号ベクトルを表し、予備実験においては目的音声信号、もしくは内部雑音信号を表す。さらに、 $\mathbb{E}[\cdot]$ は平均値操作を表し、ここでは時間フレームに対して平均値操作を行う。また、

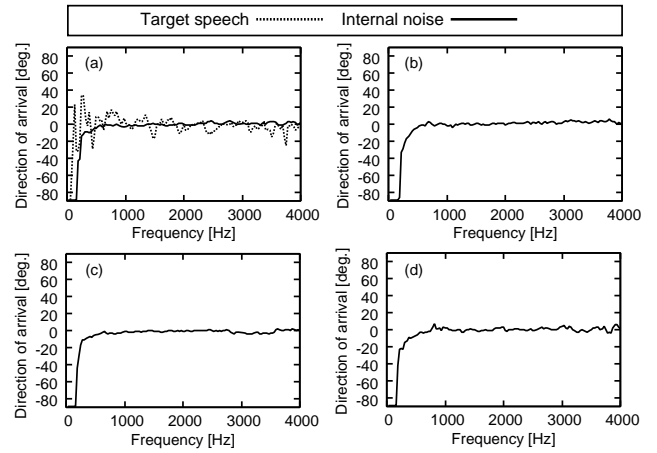


Figure 5: DOAs of internal noise: (a) type 1 target speech, (b) type 2, (c) type 3 and (d) type 4.

各周波数毎に θ を -90° から 90° まで 1° ずつ変化させ、 $|P(f, \theta)|$ が最も大きくなったときの θ を受信信号の DOA とする。

4 種類の内部雑音信号における DOA の結果を Figs. 5(a), (b), (c), (d) に示す。また、Fig. 5(a) には目的音声信号の DOA の結果も示す。Figure 5 より、内部雑音タイプによらず内部雑音信号の DOA が 0° に定位していることが確かめられる。これは、ロボットに取り付けられたマイクロホンアレーで観測される内部雑音信号が、固体中を伝わる筐体振動であることが原因だと考えられる。一般的に、固体中の音速は空気中の音速よりも 10 倍以上速い。そのため、各マイクロホンに入力する内部雑音信号の時間差は非常に小さくなる。その結果、内部雑音信号の DOA は 0° になると考えられる。よって、従来のマイクロホンアレー配置では Fig. 5(a) に示すように目的音声信号と内部雑音信号の DOA が同じ方位になってしまう。

3.3 マイクロホンアレー配置の提案

ロボット音声対話システムでは、ロボットの正面にユーザが立っている状況を想定するのが一般的である。よって、ロボットに搭載するマイクロホンアレーがブロードサイドアレーの場合、Fig. 6(a) に示すように目的音声信号と内部雑音信号の DOA が 0° で同じ方位になる。Figure 6(a) の場合、ICA が目的音声信号に対して死角を形成すると、同時に内部雑音信号も抑圧することになり、適切に内部雑音信号を推定することができない。そこで、マイクロホンアレー配置をブロードサイドアレーからエンドファイアアレーに回転させることを提案する。これによって、マイクロホンアレーに同位相で入力する内部雑音信号の DOA は、Fig. 6(b) のように、目的音声信号とは異なる方位にマッピングされる。その結果、ロボット音声対話システムにおけるユーザとロボットの対面関係を保ったまま、目的音声信号と内部雑音信号の DOA を空間的に異なる方位にすることができ、SBSS の雑音推定性能が改善すると考え

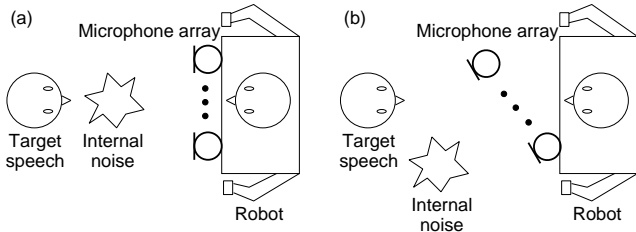


Figure 6: (a) Conventional and (a) proposed microphone array structures.

られる．本実験で用いている内部雑音抑圧型ロボットにおける最適なマイクロホンアレー配置 (θ_m) を以下の実験により評価する．

4 実験

4.1 実験条件

マイクロホンアレー配置の有効性を検証するために，シミュレーション実験を行った．実験における音響環境は，2.4 節で示した環境と同じである．目的音声信号には，Fig. 3 で収録したインパルス応答を畳み込んだ日本語新聞記事読み上げ音声コーパス (Japanese newspaper article sentences: JNAS) のテストセット 100 文 (女性発声) を用い，内部雑音信号は前述の 4 種類を用いた．目的音声信号と内部雑音信号の入力 SNR は，内部雑音タイプ 1 が 16.6 dB，タイプ 2 が 4.54 dB，タイプ 3 が 0.39 dB，タイプ 4 が 5.37 dB である．さらに，目的音声信号と環境雑音信号の入力 SNR が 10 dB になるように実収録の環境雑音信号を付加した．音声認識実験の条件を表 1 に示す．WF の処理強度パラメータ β は内部雑音タイプによらず 5 で固定した．以上の実験条件において，Fig. 3 に示す θ_m を -90° から 90° の範囲で計 11 パターン変化させて実験を行った．

4.2 実験結果

マイクロホンアレー配置を回転させたときの単語認識精度，SD，雑音抑圧量 (noise reduction rate: NRR)，ケプストラム歪み (cepstral distortion: CD) の比較を行った．NRR は出力 SNR [dB] - 入力 SNR [dB] で定義され，値が大きくなるほど雑音抑圧性能が高いことを表す指標である．また，CD はスペクトル包絡の歪み具合を表す尺度で，値が小さいほど処理による歪みが小さいことを表す指標である [11]．

Figure 7 に音声認識実験の結果を示す．Figure 7 より，全ての内部雑音タイプにおいて $\theta_m = 60^\circ$ のときのほうが $\theta_m = 0^\circ$ よりも単語認識精度が優れていることが分かる． $\theta_m = 60^\circ$ のとき $\theta_m = 0^\circ$ に比べて，単語認識精度が内部雑音タイプ 1 では 14%，タイプ 2 では 11%，タイプ 3 では 7%，タイプ 4 では 13% 改善された．また，SBSS 地点における全ての雑音信号における SD を Figs. 8(a), (b), (c), (d)，最終出力音声信号における NRR と CD の

Table 1: 音声認識実験の条件

テストデータ	JNAS テストセット (女性話者 23 名 100 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) [12] 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. 3.5.1

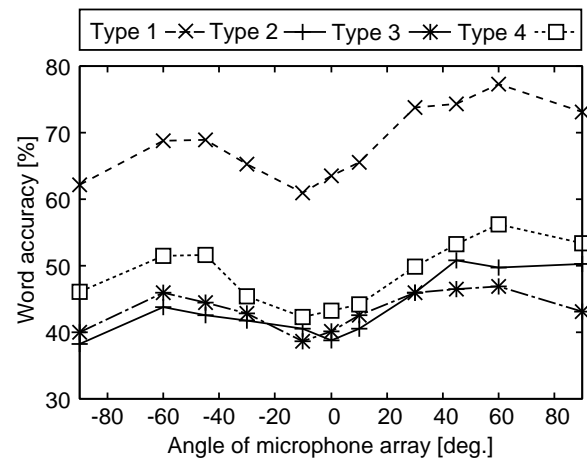


Figure 7: Word accuracy for all internal noise types with different angles of microphone array.

結果を Figs. 9(a), (b) に示す．Figure 8 より， $\theta_m = 60^\circ$ のときのほうが SD が小さくなっており，SBSS の雑音推定性能が改善されていることが分かる．また Fig. 9(a) より，NRR はタイプ 3 以外の内部雑音タイプではほとんど性能が変わらないことが示されている．さらに Fig. 9(b) より，CD は全ての内部雑音タイプにおいて小さくなっていることが分かる．これらの結果から，従来のマイクロホンアレー配置の場合，雑音推定性能が高くないことにより，WF において雑音信号を多く見積もり雑音抑圧されたため，目的音声信号の CD が大きかったと考えられる．しかし提案するマイクロホンアレー配置では，雑音推定性能が改善されたことにより適切に雑音抑圧を実現することができ，雑音抑圧性能を低下させずに CD 値を改善することができたと考えられる．

5 まとめ

本論文では，内部雑音の DOA が固体を伝わる振動であることによって 0° に定位することを示した．この知見に基づき，マイクロホンアレーの配置を回転させることを提案し，音声認識実験によりその有効性を確認した．この提案はロボットにおける内部雑音に限らず，同じように固

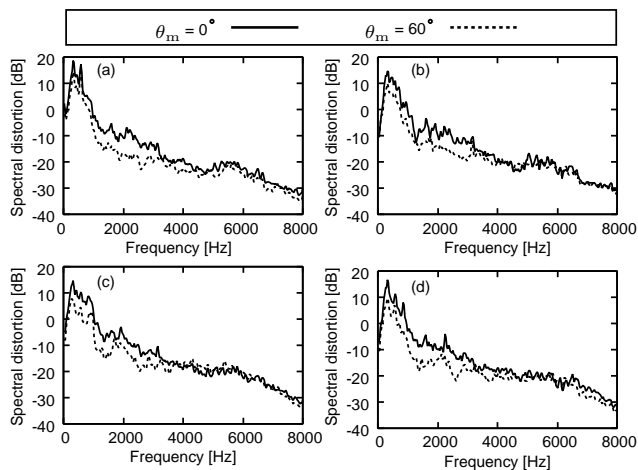


Figure 8: Experimental results of spectral distortion of internal noise: (a) type 1, (b) type 2, (c) type 3 and (d) type 4 for cases of $\theta_m = 0^\circ$ and $\theta_m = 60^\circ$.

体中を振動して伝わる車のロードノイズにも応用することができると考えられる [13].

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] J. Even, et al., "Frequency domain semi-blind signal separation: application to the rejection of internal noises," *Proc. International Conference on Acoustic Speech and Signal Processing*, pp. 157–160, 2008.
- [2] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.
- [3] J. Even, et al., "Semi-blind suppression of internal noise for hands-free robot spoken dialog system," *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 658–663, 2009.
- [4] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, pp. 287–314, 1994.
- [5] M. Brandstein, et al., *Microphone Arrays Signal Processing Techniques and Applications*, Springer-Verlag, 2001.
- [6] A. Lee, et al., "Julius? An open source real-time large vocabulary recognition engine," *Proc. Eur. Conf. Speech Commun. Technol.*, pp. 1691–1694, 2001.

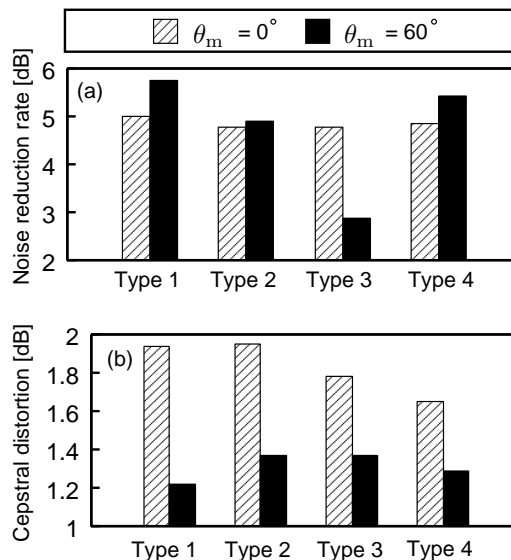


Figure 9: Experimental results of (a) noise reduction rate (b) cepstral distortion for cases of $\theta_m = 0^\circ$ and $\theta_m = 60^\circ$.

- [7] N. Vlassis, et al., "Efficient source adaptivity in independent analysis," *IEEE Trans. Neural Networks*, vol.12, no.3, pp. 559–566, 2001.
- [8] J. Even, et al., "An improved permutation solver for blind signal separation based front-ends in robot audition," *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2172–2177, 2008.
- [9] 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.
- [10] H. Saruwatari, et al., "Blind source separation combining independent component analysis and beamforming," *EURASIP J. Applied Signal Proc.*, vol.2003, no.11, pp. 1135–1146, 2003.
- [11] L. Rabiner, et al., *Fundamentals of speech recognition*, Upper Saddle River, NJ: Prentice Hall PTR, 1993.
- [12] A. Lee, et al., "A new phonetic-tied mixture model for efficient decoding," *In Proceedings of ICASSP.*, pp. 1269–1272, 2000.
- [13] H. Saruwatari, et al., "Speech enhancement in car environment using blind source separation," *Proc. International Conference on Spoken Language Processing*, pp. 1781–1784, 2002.