

ロボットのためのハンズフリー音声認識システム

細谷 耕佑[†] 小川 哲司^{††} 藤江 真也^{††} 渡辺 大地[†] 市川 悠飛[†]

谷山 輝[†] 小林 哲則[†]

[†] 早稲田大学 基幹理工学部 情報理工学科

〒 169-8555 東京都新宿区大久保 3-4-1

^{††} 早稲田大学 高等研究所

〒 169-8050 東京都新宿区西早稲田 1-6-1

あらまし 自律移動型ロボットに適した音源分離方式を開発し、対話ロボットにおけるハンズフリー音声認識システムの前処理として適用した。提案する音源分離方式は、対話相手の発話音声に加えてマイクロホンに混入する指向性雑音、拡散性雑音、ロボット動作音、ロボットの自発話といった多様な雑音を同時に除去可能な枠組である。本研究では、この枠組を、ロボット頭部に設置した超小型無指向性マイクロホンをを用い、低計算コストのアルゴリズムで実現する。ロボットとの対話において、対話相手はロボットの正面にすることが多いため、提案手法はロボットの正面方向からの音声のみを抽出する枠組となっている。このような、ロボットとの対話という場面において妥当な制約を設けることで、低計算コストの雑音抑圧処理を実現することが可能となる。また、対話相手がロボットの正面方向から移動した場合にも対処するため、ロボットの目に設置したカメラから得られる画像情報を用いて、対話相手の顔を検出、追跡する枠組を構築し、音源定位として利用する。このように、画像情報を併用できるというロボットの利点を活かすことで、多様な雑音を実時間で高精度に抑圧することが可能となること、および得られた雑音抑圧後の音声を認識に用いることで、実環境においてハンズフリー音声対話システムが良好に動作することを確認した。

キーワード 雑音除去, ハンズフリー音声認識, 顔検出, 音声対話ロボット

Hands-free speech recognition system for robot

Kosuke HOSOYA[†], Tetsuji OGAWA^{††}, Shinya FUJIE^{††}, Daichi WATANABE[†], Yuhi

ICHIKAWA[†], Hikaru TANIYAMA[†], and Tetsunori KOBAYASHI[†]

[†] Dept. of Computer Science, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, JAPAN 169-8555

^{††} Waseda Institute for Advanced Study

1-6-1 Nishi-waseda, Shinjuku-ku, Tokyo, JAPAN 169-8050

Abstract A new type of noise reduction method suitable for autonomous mobile robots is proposed and applied to pre-processing of a hands-free spoken dialogue system. The proposed method can reduce various kinds of noise such as directional noise, diffuse noise, moving noise of the robot, and speech utterance from the robot, which are mixed with the target speech for the case in which people talk with the robot, by using small and light-weighted devices and low-computational-cost algorithms. Here, we assume that the people talking with the robot is in the front of the robot, and thus the proposed method aims at extracting speech signals coming from the frontal direction of the robot. In addition, for the case in which the people moves from the front of the robot, the sound source can be localized by face detection and tracking using facial images obtained from a camera mounted on eyes of the robot. By taking advantage of the robot, which can combine speech information with image information, real-time reduction of the various noise can be achieved, and thus the hands-free spoken dialogue system can work well in real environments.

Key words noise reduction, hands-free speech recognition, face detection, spoken dialogue system for robot

1. はじめに

対話相手の発話音声のみを抽出する音源分離処理を、小型・軽量のデバイスと低計算コストのアルゴリズムを用いて実現し、音声対話ロボットの必須機能であるハンズフリー音声認識の前処理として適用することを試みる。

実環境においてロボットと人が会話を行う場合、様々な雑音が入り混じった音声の中から、目的とする対話相手の音声のみを分離・抽出し、認識する機能を実現する必要がある。特に、自律移動型ロボットにおいては、積載可能な装置の重量・サイズの制限から、マイクロホンや信号処理デバイスなどは、物理的な小型化、軽量化が必要とされる。また、そのようなデバイスの小型化に伴い、必然的に処理アルゴリズムの低コスト化も求められる。

ロボットに搭載したマイクロホンによる雑音抑圧手法としては、厳密な頭部伝達関数を用いる手法がある [1]。しかし、この手法では、ロボットおよびマイクロホン配置ごとに膨大な数の頭部伝達関数の測定が必要となる。このような頭部伝達関数の測定を避けるため、頭部を球状と仮定することで、厳密な頭部伝達関数の測定を必要としない手法が提案されている [2]。しかし、頭部を球状と仮定できない場合も多い。

それに対し我々は、ロボット頭部側面に配置した 4 系統の指向性マイクロホンを用いた、頭部伝達関数の測定を必要としない指向性雑音抑圧方式を提案してきた [3]。しかし、この方式は指向性マイクロホンが必要なことから、micro electro mechanical systems (MEMS) 技術の導入が難しく、単純には小型化できないという問題があった。また、ロボットとの音声対話においては、対話相手以外の話者の発話音声などの指向性雑音のみならず、エアコンや計算機などを音源とする拡散性の雑音、ロボット自身が動作する際に発生する雑音、ロボット自身の発話音声といった多様な雑音がマイクロホンに混入するため、それらを同時に除去する必要がある。

そこで本稿では、頭部伝達関数の影響の少ない頭頂部に設置した 4 系統の無指向性アナログ MEMS マイクロホンを用いて、多様な雑音が入り混じった環境下において、ロボットの正面方向から到来する音声のみを強調する音源分離方式 [4] [5] を提案し、対話ロボットによるハンズフリー音声対話システムに適用することを試みる。提案手法では、対話相手がロボットの正面にいてと仮定しているが、対話相手がロボットの正面から移動した場合にも対処する必要がある。本研究では、ロボットの目に設置されたカメラから得られる画像情報を用いて対話相手の顔を検出・追跡することで、音源である対話相手を常に正面に捉えることにより、この問題を解決する。このように、ロボットによる対話システムでは画像情報を併用することで、音声のみで音源の定位と分離を行うブラインド音源分離に比べ、低計算コストの音源分離手法を容易に構築することが可能となる。

本論文の構成は下記の通りである。2. では、ハンズフリー音声対話システムに適用する雑音抑圧システムについて延べ、指向性雑音と拡散性雑音が混在する雑音環境下における基本性能を単語正解精度により評価する。続いて、3. では、画像情報を

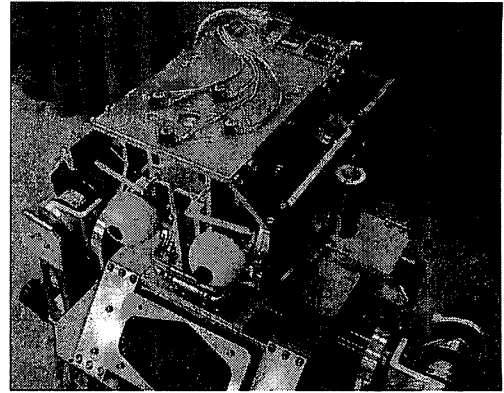


図 1 The robot and the microphones mounted on the robot head.

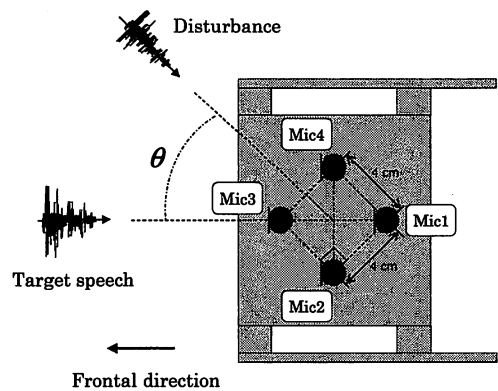


図 2 A microphone arrangement. This figure shows the top view of the robot.

用いて対話相手の顔を検出・追跡することで、音源を定位する手法について述べる。最後に、4. でまとめを述べる。

2. 雑音抑圧システム

対話ロボットのための雑音抑圧システムについて述べる。

2.1 マイクロホンシステム

2.1.1 MEMS マイクロホン

本研究では、4 系統の無指向性アナログ MEMS マイクロホンを使用する。MEMS マイクロホンは半導体集積技術を用いて作られたマイクロホンであり、非常に小型、軽量のデバイスである。ここでは、Knowles 社製のアナログ MEMS マイクロホン SPM0208HD5 を用いた。MEMS マイクロホンに加えアンプ等の周辺回路を搭載した 1.5 cm 角の基板を作成し、ロボットに設置した。

2.1.2 マイクロホン配置

マイクロホンはロボット頭頂部に図 1 のように正方形型に配置する。マイクロホンをロボット頭頂部に設置するのは、ロボット自身によって生じる反射や回折の影響を軽減するためである。マイクロホン間隔は隣接するマイクロホン間で 4 cm とした。このとき、対角位置にあるマイクロホンの間隔は 5.66

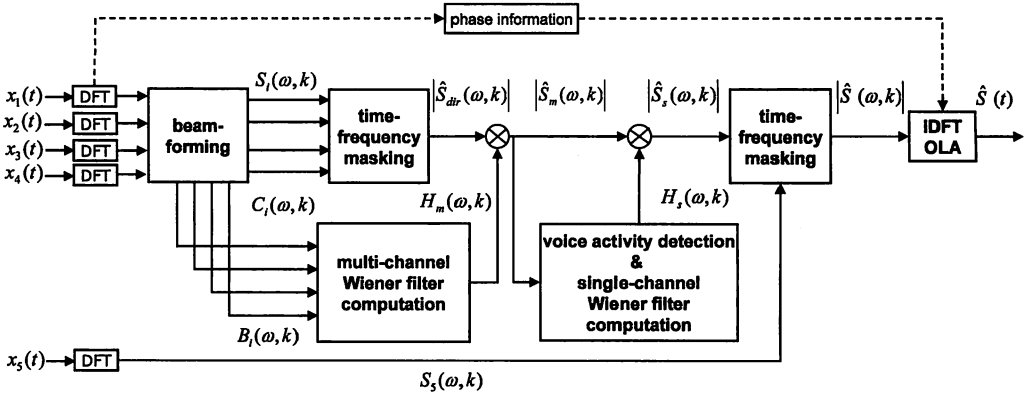


図3 Block diagram of the proposed noise reduction system.

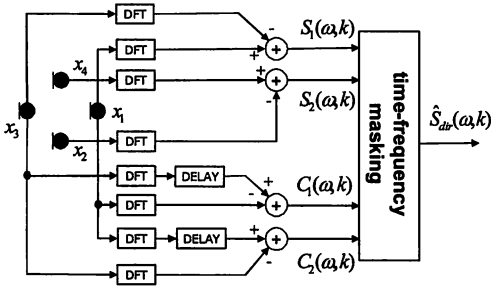


図4 Directional noise reduction method.

cm となる。強調すべき音声 (目的音声) の到来方向をロボットの正面方向とし、マイクロホンの番号を図2のように定める。

2.2 雑音抑圧方式

ロボット頭頂部に設置した4系統のマイクロホンの入力音声から指向性雑音、拡散性雑音、ロボット動作音、ロボットの自発話音声を同時に抑圧し、対話相手 (目的話者) の音声のみを強調する手法について述べる。

提案する雑音抑圧システムの処理の流れを図3に示す。提案システムは、指向性雑音抑圧部、拡散性雑音抑圧部、残留雑音抑圧部、ロボット自発話音声抑圧部から構成される。指向性雑音抑圧部では、主にロボット正面以外の方向から到来する対話相手以外の話者の発話音声や、ロボットの動作音を抑圧する。拡散性雑音抑圧部では、エアコンなどを音源とする拡散性雑音を抑圧する。残留雑音抑圧部では、両方で除去しきれずに残留した定常的な雑音を抑圧する。ロボット自発話音声抑圧部では、ロボット自身の発話音声を抑圧する。

以下では、その各々の処理について述べる。

2.2.1 指向性雑音抑圧

指向性雑音抑圧の処理の流れを図4に示す。時刻 t においてチャンネル i のマイクロホンで観測した信号を $x_i(t)$ 、短時間フーリエ変換されたスペクトルを $S_i(\omega, k)$ と表す。ここで、 k は離散フレーム、 ω は離散周波数を表す。また、 $x_5(t)$ は、ロ

ボットの発話音声の信号とする。

まず、目的音源方向に対し縦にならんだ Mic1, Mic3 の入力信号を用いて遅延処理と減算処理による死角制御型ビームフォーマを形成し、その出力信号 c_1, c_2 を得る。このとき付加する遅延は、マイクロホン間隔を音声が発達するのに要する時間に相当する。対角位置にあるマイクロホン間の遅延時間を τ_{cross} としたとき、 c_1 と c_2 は次式のように得られる。

$$c_1(t) = x_3(t - \tau_{cross}) - x_1(t) \quad (1)$$

$$c_2(t) = x_1(t - \tau_{cross}) - x_3(t) \quad (2)$$

また、減算処理による減算型ビームフォーマの出力信号として次式の s_1 を得る。

$$s_1(t) = x_1(t) - x_3(t) \quad (3)$$

さらに、目的音源方向に対し横に並んだ Mic2, Mic4 の入力信号を用いて形成される減算型ビームフォーマの出力信号として、次式の s_2 を得る。

$$s_2(t) = x_4(t) - x_2(t) \quad (4)$$

このとき、 c_1 は 0° 方向に死角を持ち、 c_2 は 180° 方向に死角を持つ指向特性を形成する。また、 s_1 は $0^\circ, 180^\circ$ 方向に最大のゲインを持ち $\pm 90^\circ$ 方向に死角を持つ指向特性を形成し、 s_2 は $\pm 90^\circ$ 方向に最大のゲインを持ち、 $0^\circ, 180^\circ$ 方向に死角を持つ指向特性を形成する。 c_1, c_2 のゲインパターンを図5に、 s_1, s_2 のゲインパターンを図6に示す。この4つの信号を用いて次式の帯域選択を行うことで、正面方向から到来する信号 \hat{S}_{dir} を推定する [6]。

$$\hat{S}_{dir}(\omega, k) = \begin{cases} S_1(\omega, k), & \text{if } |S_1(\omega, k)| > |S_2(\omega, k)| \\ & \text{and } |C_1(\omega, k)| < |C_2(\omega, k)| \\ \beta, & \text{otherwise} \end{cases} \quad (5)$$

ここで、 β はフロアリング定数を表す。

この帯域選択は、 $S_1(\omega, k)$ と $S_2(\omega, k)$ を比較し、正面、後ろ方向にゲインを持つ信号 $S_1(\omega, k)$ の方が大きい時間・周波数成

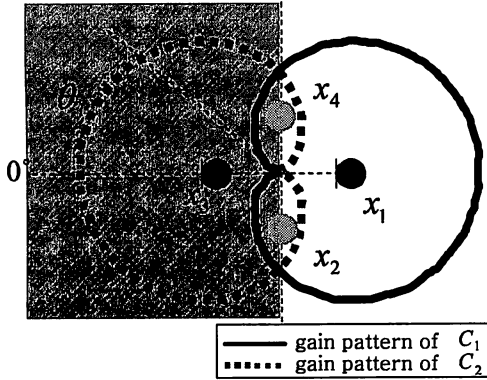


図 5 Directivity patterns of null beamformer. Time frequency masking removes spectral components of the signals coming from the backward of the robot (unshaded area).

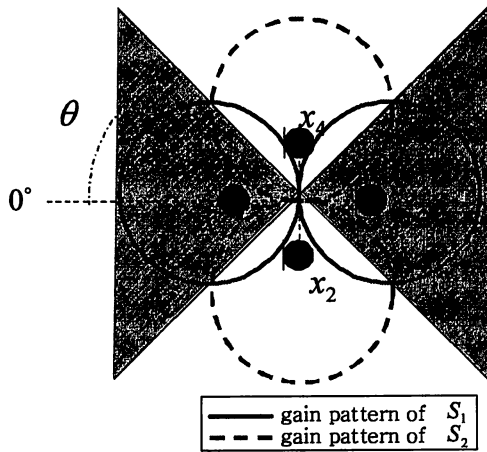


図 6 Directivity patterns of subtractive beamformer. Time frequency masking removes spectral components of the signals coming from the side of the robot (unshaded area).

分のみを残すことで、横方向から到来する指向性雑音を抑圧する(図 6)。さらに、 $C_2(\omega, k)$ と $C_1(\omega, k)$ を比較し、正面方向にゲインを持つ信号の方が大きい時間・周波数成分の $S_1(\omega, k)$ のみを残すことで、先の選択で選ばれた時間・周波数成分のうち、後方から到来する指向性雑音を抑圧する(図 5)。最終的に図 7 に示す範囲の音声を抽出する。

このように、マイクロホンと同一平面上にあるロボット正面方向の音源に対して鋭い指向性を形成することで、ロボットと対面する人の声を抽出することができる。また、この処理により、ロボット頭頂部に設置されたマイクロホンよりも下部に音源があるロボット自体の動作音やモータ音も同時に抑圧することが可能である。

2.2.2 拡散性雑音抑圧

指向性雑音が抑圧された $\hat{S}_{dir}(\omega, k)$ 中に存在する拡散性の雑音を抑圧するため、マルチチャネルウィナーフィルタを用いる。まず、死角制御型ビームフォーマの出力信号として、下記の 4

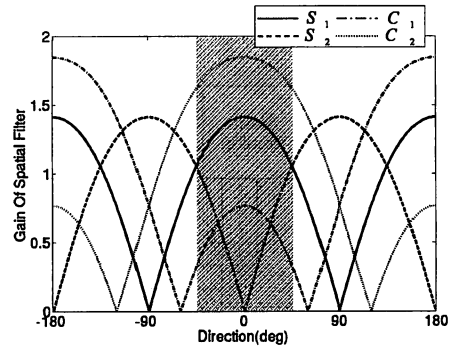


図 7 Estimates of the directional noise reduction system. Spectral components of the signal coming from the direction of the shaded area.

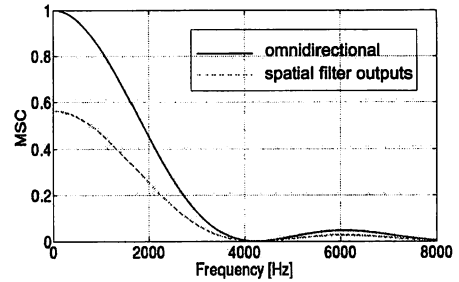


図 8 Theoretical magnitude-squared coherences as a function of frequencies for the case in which microphones spacing is 4 cm.

つの信号を得る。

$$b_1(t) = x_2(t - \tau_{next}) - x_1(t) \quad (6)$$

$$b_2(t) = x_3(t - \tau_{next}) - x_2(t) \quad (7)$$

$$b_3(t) = x_3(t) - x_4(t - \tau_{next}) \quad (8)$$

$$b_4(t) = x_4(t) - x_1(t - \tau_{next}) \quad (9)$$

ここで、 τ_{next} は、隣接するマイクロホン間隔に相当する遅延時間である。このとき、拡散性雑音抑圧フィルタ $H_m(\omega, k)$ を、上記ビームフォーマの出力 $b_i(t)$ ($i = 1, 2, 3, 4$) のスペクトル $B_i(\omega, k)$ を用いて、次式のように算出する。

$$H_m(\omega, k) = \frac{\frac{1}{2} \sum [\text{abs}\{B_l(\omega, k)B_m^*(\omega, k)\}]}{\frac{1}{4} \sum_{i=1}^4 [B_i(\omega, k)B_i^*(\omega, k)]} \quad (10)$$

ここで、 l, m は $\{(l, m)\} = \{(1, 2), (3, 4)\}$ なる組合せであり、目的音方向に対して線対称となるゲインパターンを持つ死角制御型ビームフォーマを選択して用いた。本手法は、Zelinski によるマルチチャネルウィナーフィルタ [7] と同様の定式化であるが、マイクロホンの入力をそのまま用いるのではなく、ビームフォーマの出力を用いている点が大きく異なる。マイクロホンの入力をそのまま用いる従来のマルチチャネルウィナーフィルタでは、本研究のようにマイクロホン間隔が小さい場合、図 8 に示すように低周波数帯域において MSC が低減せず、十分

な拡散性雑音抑圧効果が得られない。一方、提案手法では、互いに対称な指向性をもつビームフォーマの出力を用いることで、マイクロホン間隔が小さい場合でも、高い拡散性雑音抑圧効果が得られる [4]。

拡散性雑音が抑圧された信号の振幅スペクトルは、式 (10) で表されるマルチチャンネルウィナーフィルタを用いて、次式のように求まる。

$$|\hat{S}_m(\omega, k)| = H_m(\omega, k) \cdot |\hat{S}_{dir}(\omega, k)| \quad (11)$$

2.2.3 残留雑音抑圧

2.2.1 で述べた指向性雑音抑圧と 2.2.2 で述べた拡散性雑音抑圧を施した信号に残留している定常的な雑音を、シングルチャンネルウィナーフィルタにより抑圧する。

ここでは、拡散性雑音抑圧部において算出するコヒーレンス値とパワーを用いて推定した非発話区間から残留雑音を推定し、ウィナーフィルタ $H_s(\omega, k)$ を算出する。これをマルチチャンネルウィナーフィルタを適用した信号 $|\hat{S}_m(\omega, k)|$ に施すことで、残留雑音が抑圧された目的音の推定値 $|\hat{S}_s(\omega, k)|$ を得る。

$$|\hat{S}_s(\omega, k)| = H_s(\omega, k) \cdot |\hat{S}_m(\omega, k)| \quad (12)$$

2.2.4 ロボットの自発話音声抑圧

2.2.3 までに述べた雑音抑圧を行うことで、指向性雑音、拡散性雑音、ロボット動作音が抑圧された、ロボットの正面方向から到来する信号が得られる。しかし、ロボットの胴体に設置されたスピーカから出力されるロボット自身の発話音声、目的話者の音声とともにマイクロホンに混入し、音声認識、発話区間検出に悪影響を及ぼす。

そこで、2.2.3 までの処理で得られた $|\hat{S}_s(\omega, k)|$ と、ロボットの発話として出力する合成音声信号 $x_s(t)$ のスペクトル $S_s(\omega, k)$ を用いて以下の帯域選択を行うことで、ロボットの自発話音声を除去する。

$$|\hat{S}(\omega, k)| = \begin{cases} |\hat{S}_s(\omega, k)|, & \text{if } |S_s(\omega, k)| < \text{threshold} \\ \beta, & \text{otherwise} \end{cases} \quad (13)$$

ここで、 β はフロアリング定数を表す。また、閾値は実験的に求めた。最終的に得られた振幅スペクトル $|\hat{S}(\omega, k)|$ に入力信号の位相を付与し、逆フーリエ変換を施すことで、雑音が抑圧された音声信号を復元する。

2.3 評価実験

提案する雑音抑圧手法の有効性を調査するために、指向性雑音と拡散性雑音が混在する環境下において、音源分離実験を行った。指向性雑音としては妨害話者の発話音声を、拡散性雑音としては大型空調器の雑音を用いた。本評価実験においては、ロボットの動作音と自発話音声は対象としない。評価は単語正解精度を用いた。従来手法として、遅延和アレイの後段に Zelinski のマルチチャンネルウィナーフィルタを施した方式 [7] (4ch-DS+MWF), generalized sidelobe canceler (GSC) [8] (2ch-GSC), マイクロホン間の位相差に基づく時間・周波数マスクング [9] (2ch-TFmasking) の雑音抑圧性能を評価した。2ch-GSC, 2ch-TFmasking については、Mic2 と Mic4 で観測された 2 チャンネルの音声信号を用いて評価を行った。

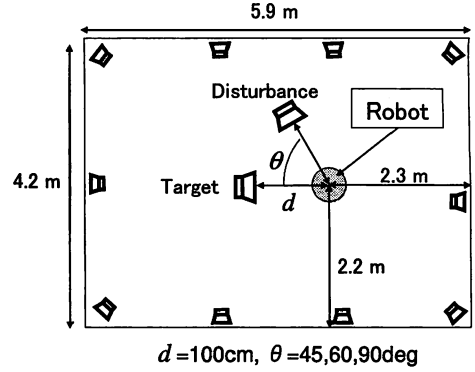


図 9 Recording environment.

表 1 Setup for analysis

sampling frequency	16kHz
frame length	32ms
frame shift	8ms
analysis window	Hamming window
analysis range of frequencies	300 - 5600Hz

表 2 Setup for acoustic feature extraction

sampling frequency	16kHz
frame length	25ms
frame shift	10ms
analysis window	Hamming window
pre-emphasis	$1-0.97z^{-1}$
feature parameters	12 MFCCs, 12ΔMFCCs, Δlog energy

2.3.1 音声収録

音声の収録環境を図 9 に示す。マイクロホンは、対話ロボット ROBISUKE [10] の頭部に設置した。目的音声、妨害音声の音源位置はロボットから 1 m とし、ロボットの正面方向を 0° としたとき、目的音声を 0° に、妨害音声を 45°, 60°, 90° に配置した。

認識対象音声(目的音)は ASJ-JNAS [11] に含まれる男性話者 23 人が発話した計 100 発話を使用した。妨害音は、ASJ-JNAS のうち、目的音に含まれない話者の発話 100 発話を使用した。このとき、妨害音の発話長、発話エネルギーが目的音と同じになるように調整した。つまり、指向性雑音と目的音の SN 比はほぼ 0dB である。

拡散性雑音は電子協騒音データベースより空調機(大型)の騒音を、部屋の周囲に配置した 10 個のスピーカから再生することで模擬した。この拡散性雑音を目的音との SN 比が 15dB となるように重畳した。

2.3.2 実験条件

雑音抑圧を行う際の分析条件を表 1 に、音声認識に用いる音響特徴量の抽出条件を表 2 に示す。音響モデルは、接話マイクロホンで収録された男性 133 名による新聞記事読み上げ音声 20414 文から学習を行った状態共有トライフォン(状態数 2000,

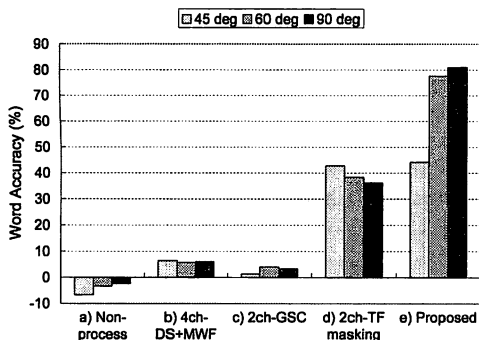


図 10 Word accuracy for evaluation items. Input signal consists of a target speech utterance, directional noise coming from the direction of 45°, 60° or 90°, and diffuse noise at an SNR of 15dB.

混合数 16) である。言語モデルには、新聞記事から学習した語彙数 20000 のトライグラムを用いた。また、認識器には本研究室で開発した SKOOD [12] を用いた。

2.3.3 実験結果

単語正解精度による評価結果を図 10 に示す。

提案手法 (e) では、指向性雑音の到来方向が目的音声から 60° 以上離れている場合、75% 以上の単語正解精度を達成した。雑音抑圧処理を行わない場合 (a), 単語正解精度は 0% を下回っており、認識器は機能していないことがわかる。また、遅延和アレイの後段にマルチチャネルウィナーフィルタを施した手法 (b) や、GSC (c) についても単語正解精度の向上は見られない。マイクロホン間の位相差に基づく時間・周波数マスクング (d) は、上述の方法と比較して高い性能を与えており、特に、妨害音の方向が 45° のとき提案法と同等の性能を与えている。しかし、妨害音の方向が 60° 以上離れているとき提案手法に比べて大きく性能が劣化しており、提案手法の有効性が示された。

3. 画像情報に基づく音源の検出・追跡

2. で述べた音源分離処理システムは、対話相手が正面にいることを仮定し、ロボットの正面方向から到来する音声を抽出する。そこで、対話相手が正面方向から移動してしまった場合に対処するため、ロボットの目に搭載されたカメラから得られる画像情報を用いて顔検出・追跡を行うことで、目的音源である対話相手を正面に捉える枠組を構築した。

まず、ロボットに搭載されたカメラの画像内に人がいない場合、キャプチャーされた画像からオプティカルフローを計算し、動く物体を検出する。ここで、検出される対象物は人の体や手振りとは仮定しているため、ロボットは、検出された方向に視線を向ける。次に、画像内に人の顔が検出された場合、その方向に視線を向ける。こうすることで、対話相手を常にロボットの正面に捉え続けることが可能となる。

なお、オプティカルフローの計算には、Lucas・Kanade が提案したアルゴリズム [13] を採用し、顔検出の機能は、Haar-Like 特徴量を用いた Viola・Jones のアルゴリズム [14] を採用した。

また、音声処理、画像処理、ロボット制御各モジュールの協調動作のために、ロボット用通信アーキテクチャである MONEA [15] を用いた。

4. まとめ

ロボット頭部に設置した小型の無指向性マイクロホンを用い、多様な雑音が混在する環境において、ロボットの正面にいる対話相手の発話音声のみを強調する手法を提案し、対話ロボットにおけるハンズフリー音声認識システムの前処理として適用した。また、ロボットに搭載されたカメラから得られる画像情報を用いて、対話相手の顔を検出・追跡することで、音源である対話相手を常に正面に捉える枠組を構築した。

謝辞 本研究の一部は、NEDO 戦略的先端ロボット要素技術プロジェクト・高齢者コミュニケーション RT システムの支援によるものである。

文献

- [1] F. Asano *et al.*, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Process.*, vol.SAP-8, no.5, pp.497-507, Sept. 2000.
- [2] K. Nakadai *et al.*, "Applying scattering theory to robot audition system," *Proc. IROS2003*, pp.1147-1152, Oct. 2003.
- [3] N. Mochiki *et al.*, "Ears of the robot: Three simultaneous speech segregation and recognition using robot-mounted microphones," *IEICE Letter on Information and Systems*, vol.E90-D, no.9, Sept. 2007.
- [4] S. Takada *et al.*, "Speech enhancement using square microphone array for mobile devices," *Proc. ICASSP2008*, pp.313-316, March 2008.
- [5] T. Ogawa *et al.*, "Ears of the robot: Noise reduction using four-line ultra-micro omni-directional microphones mounted on a robot head," *Proc. EUSIPCO2008*, Aug. 2008.
- [6] 赤桐健三, 他, "2 個のマイクによる空間フィルタを用いた音源分離方式," *音誌論集*, pp.443-444, March 2005.
- [7] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *Proc. ICASSP1988*, vol.5, pp.2578-2581, April 1988.
- [8] L. J. Griffiths *et al.*, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propagation*, vol. 30, no. 1, pp. 27-34, 1982.
- [9] O. Yilmaz *et al.*, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol.52, no.7, July 2004.
- [10] 小林哲則, 他, "マルチモーダル会話ロボット: ロボットが会話において行う「聴く」行為に付いて," *計測と制御*, vol.46, no.6, pp.466-471, June 2007.
- [11] K. Itou *et al.*, "The design of the newspaper-based Japanese large vocabulary continuous speech recognition corpus," *Proc. ICSLP1998*, pp.3261-3264, Nov. 1998.
- [12] 柴田大輔, 他, "ワンパストライグラムデコーダにおける単語履歴の束ね処理に関する検討," *音誌論集*, pp.151-152, Sept. 2002.
- [13] B.Lucas *et al.*, "An Iterative Image Registration Technique with an Application to Stereo Visio" *Proc. of 7th IJCAI*, pp. 674-679, 1981.
- [14] P. Viola *et al.*, "Robust real-time object detection," *Intl. J. Computer Vision*, vol.57, no.2, pp.137-154, 2004.
- [15] 中野健兵, 他, "MONEA:効率的な多機能ロボット開発環境を実現するメッセージ指向ネットワークロボットアーキテクチャ," *日本ロボット学会誌*, vol.24, no.4, pp.115-125, April 2005.