

話者と音源の位置推定に基づく複数死角制御型 ビームフォーマの基礎的検討

中山 雅人[†] 西浦 敬信^{††} 山下 洋一^{††} 中迫 昇[†]

[†] 近畿大学 生物理工学部 〒 649-6493 和歌山県紀の川市西三谷 930

^{††} 立命館大学情報理工学部 〒 525-8577 草津市野路東 1-1-1

E-mail: †nakayama@info.waka.kindai.ac.jp

あらまし 近年、雑音環境下における遠隔発話音声認識のための前処理としてマイクロホンアレーを用いたビームフォーミングが注目されている。適応型ビームフォーマは雑音の方位に感度の死角を形成することで、少ないマイクロホン素子でビームフォーミングを実現できるが、話者もしくは雑音源が移動するような環境では、死角制御フィルタ（適応フィルタ）の環境雑音に対する学習が環境の変化に追従できない問題があった。しかし、事前に話者と雑音源の方位が推定可能であれば、環境雑音に対する適応学習を逐次行わなくても雑音の方位に感度の死角を形成することは可能である。そこで、本研究では、話者と雑音源の位置推定を行い、その推定結果と移動予測に基づき事前に設計した複数の死角制御フィルタを用いる複数死角制御型ビームフォーマを提案する。最後に、シミュレーションを行い、提案手法の有用性を確認した。

キーワード 遠隔発話音声認識, マイクロホンアレー, ビームフォーミング, 話者と雑音源の位置推定, 複数死角制御型ビームフォーマ

A fundamental study of multiple-nulls-steering beamformer based on both talkers and noises localization

Masato NAKAYAMA[†], Takanobu NISHIURA^{††}, Yoichi YAMASHITA^{††}, and Noboru NAKASAKO[†]

[†] Factory of Bio.-Ori. Sci. & Tech., Kinki Univ., Nishi-mitani 930, Kinokawa, Wakayama, 649-6493 Japan

^{††} College of Info. Sci. & Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577 Japan

E-mail: †nakayama@info.waka.kindai.ac.jp

Abstract Beamforming with microphone array is ideal candidate for distant-talking speech recognition. Adaptive beamformer can realize beamforming with small size microphone-array, but those had difficulty in extracting distant-moving speeches and in reducing moving noise, because those must rapidly train multiple-channel adaptive filters with long filter length by using observed noises with microphone-array. However, if positions of talkers and noise can be estimated, it may be not necessary that adaptive filters are trained in real noisy environments. Therefore, we propose the multiple-nulls-steering beamformer based on talkers and noises localization, which doesn't require training with observed noises.

Key words Beamforming, Microphone array, Distant-talking speech recognition, Talkers and noises localization, multiple-nulls-steering beamformer

1. ま え が き

近年、テレビ会議システムや音声入力で機器を制御するシステムとしてハンズフリーによる音声入力インターフェースの需要が高まっている。現在のハンズフリーシステムはヘッドセット

やピンマイクなどの身につけるタイプ（接話型）のマイクロホンが主流である。これは、口元近くの接話型マイクロホンにより音声を受音するため容易に高 SNR (Signal-to-Noise Ratio) の受音が可能なためである。しかし、接話型マイクロホンを身につける煩わしさは、ユーザへの大きな負担になる。一方、遠

隔設置型のマイクロホンは、ユーザへの負担にならないハンズフリーインタフェースであるが、口元から離れた位置のマイクロホンで音声を受音する必要があるため、背景雑音や残響の影響により音声が劣化する問題がある。遠隔発話音声を高 SNR に受音するための技術として、マイクロホンアレーが注目されている。マイクロホンアレーによるビームフォーミング（ステアリングフィルタの設計による指向性の制御）を行うことで、マイクロホンから離れた話者の音声を高品質に受音できる。

代表的なビームフォーミング技術として、遅延和アレー [1] と適応型アレー [2]~[4] がある。遅延和アレーは目的音の到来方位のみでステアリングフィルタが決まるため、適応学習を必要としない。一方、適応型アレーは雑音の到来方位が未知であっても雑音の最小化により適応的に死角を形成するが、多チャンネルのステアリングフィルタ（死角制御フィルタ）の適応学習が必要不可欠であるため、ユーザが移動する環境や雑音が急激に変動する環境に適応学習が追従できない問題がある [5]。また、適応学習を行わない適応型アレーとして、2ch 減算型アレー [5] があるが、形成できる死角が非常に狭いため、実環境では十分な性能を発揮できない問題がある。

しかし、事前に話者と雑音源の方位が推定可能であれば、環境雑音に対する適応学習を逐次行わなくても雑音の方位に感度の死角を形成することは可能である。そこで、本研究では、話者と雑音源の位置推定を行い、その推定結果と移動予測に基づき事前に設計した複数の死角制御フィルタを用いる複数死角制御型ビームフォーマを提案する。提案手法において、指向性雑音の到来方位は CSP 法 [6] を用いて推定を行い、話者の方位は、位相干渉に基づく音響測距法 [7] を用いて話者位置推定を行う。

最後に、シミュレーションを行い、提案手法の有用性を確認した。

2. 従来手法

2.1 CSP 法による音源の到来方位推定 [6]

マイクロホン i, j で信号 $s_i(t), s_j(t)$ を受音すると、CSP 法により 2ch の受音信号の CSP 係数及び到来時間差を推定できる。CSP 係数及び到来時間差の推定式を式 (1), (2) に示す。

$$\text{CSP}_{i,j}(k) = \text{DFT}^{-1} \left[\frac{\text{DFT}[s_i(t)]\text{DFT}[s_j(t)]^*}{|\text{DFT}[s_i(t)]||\text{DFT}[s_j(t)]|} \right] \quad (1)$$

$$\tau = \underset{k}{\text{argmax}}(\text{CSP}_{i,j}(k)) \quad (2)$$

ここで、 t は離散時刻 [s]、 $\text{CSP}_{i,j}(k)$ は CSP 係数、 τ は到来時間差、 $\text{DFT}[s_j(t)]^*$ は $\text{DFT}[s_j(t)]$ の複素共役を表す。

式 (1) では、2ch の受音信号をフーリエ変換して、振幅で正規化を行う。そして位相差を求めて逆フーリエ変換を行い CSP 係数を求めている。さらに式 (2) により CSP 係数の大きくなる時間差（相関の強い時間差）を求めることにより到来時間差の推定を行う。

音源が一つしか存在しない環境において音源の DOA を推定するときは、式 (1), (2) により τ を求めて式 (3) により音源方位を推定すればよい。

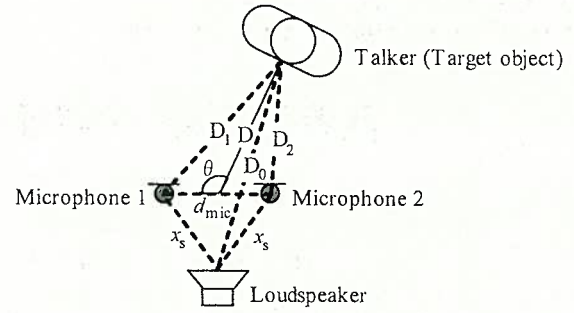


図 1 話者位置推定手法の概略図

Fig. 1 Overview of talker localization method.

$$\theta = \cos^{-1} \left(\frac{c \cdot \tau / F_s}{d_{\text{mic}}} \right) \quad (3)$$

ここで、 c は音速 [m/s]、 d_{mic} はマイクロホン間隔 [m]、 F_s はサンプリング周波数 [Hz] である。

2.2 位相干渉に基づく音響測距法による話者位置推定 [8]

本手法は、可聴音の送信波と反射波の位相干渉に基づく音響測距法 (Acoustic Distance Measurement method; ADM method) [7] を話者位置推定に応用した手法である。本手法では、図 1 のような 2ch マイクロホンとスピーカのシステムを利用して、話者位置を推定する。図 1 の Microphone 1 で観測された合成波による距離スペクトルを $|P_1(x)|$ 、Microphone 2 で観測された合成波による距離スペクトルを $|P_2(x)|$ とする。

対象物である話者が 1 人である場合、提案手法の距離スペクトル $|P_1(x)|, |P_2(x)|$ のピークが存在する距離 x_1, x_2 は図 1 の x_s, D_0, D_1, D_2 を用いると式 (4) で与えられる。

$$\begin{aligned} x_i &= \underset{x}{\text{argmax}}(|P_i(x)|) \\ &= D_i + D_0 - x_s \end{aligned} \quad (4)$$

ここで、 i は 1 もしくは 2 であり、スピーカとマイクロホン間の距離 x_s は既知である。

次に、マイクロホン 1, 2 の中点と対象物間の距離 D を式 (5) のようにマイクロホン 1, 2 までの距離の平均値によって近似する。

$$D = \frac{x_1 + x_2}{4} \quad (5)$$

ここで、マイクロホン間隔 d_{mic} が既知、 $\delta = x_2 - x_1$ 、 $D_2 = D_1 + \delta$ とすると、図 1 のマイクロホンアレーから対象物までの方位 θ は式 (6) より推定できる。

$$\theta = \cos^{-1} \left(\frac{d_{\text{mic}}^2/4 + D^2 - D_1^2}{D \cdot d_{\text{mic}}} \right) \quad (6)$$

$$D_1 = \sqrt{\frac{d_{\text{mic}}^2}{4} + D^2 - \frac{\delta^2}{4} - \frac{\delta}{2}} \quad (7)$$

2.3 AMNOR (Adaptive Microphone-array for NOise Reduction) [3], [4]

AMNOR [4] は適応型アレーの 1 つで、式 (8) の拘束条件を持つ適応型アレーである。

$$D_1 = \int |1 - F(\omega)|^2 d\omega \leq \hat{D}_1, \quad (8)$$

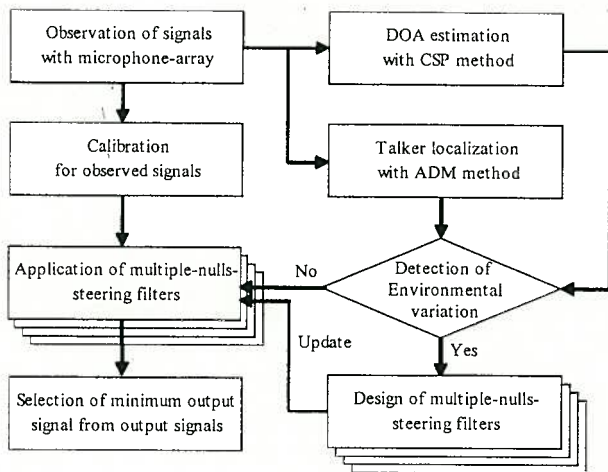


図2 提案手法のフローチャート
Fig. 2 Flowchart of proposed method.

ここで、 D_1 は目的信号の劣化量であり、 $F(\omega)$ はアレーの目的信号に対する周波数特性、 \hat{D}_1 は事前に設定した劣化許容量である。この拘束条件は、目的信号の周波数特性 $F(\omega)$ の劣化を許容する条件である。この拘束条件により AMNOR は目的信号の劣化量 D_1 と出力雑音パワーとの間にトレードオフの関係をもつ。すなわち、目的信号の劣化許容量 \hat{D}_1 を大きく許容すれば、出力雑音パワーは小さくなり、逆に \hat{D}_1 を小さく許容すれば出力雑音パワーは大きくなる。この結果、劣化許容量 \hat{D}_1 を適切に決定できれば、目的信号の周波数特性の劣化を最小に留めながら、雑音を大幅に抑圧できる。

AMNOR の適応フィルタを設計する際には、擬似目的信号が必要になる。これは、AMNOR が擬似目的信号の出力に対する劣化量を拘束条件として環境雑音の最小化を行うためである。

3. 話者と音源の位置推定に基づく複数死角制御型ビームフォーマの提案

従来手法である適応型アレーは、少ない素子数でも比較的高い性能を発揮できるが、雑音環境の急激な変動や常に雑音源の位置が変動するような環境では適応フィルタの学習が困難であるため、十分な性能を発揮できない問題があった。そこで本章にて、話者と音源の位置推定に基づく複数死角制御型ビームフォーマを提案する。

3.1 提案手法のアルゴリズム

本節では、提案手法の動作アルゴリズムについて説明する。図2に提案手法のフローチャートを示す。

図2において、到来方位推定 (DOA estimation) は CSP 法 [6] により、話者位置推定 (Talker localization) は ADM 法 [7] を利用することで取得可能である。本論文では到来方位、話者位置はそれぞれ従来手法により取得可能であることを前提とする。計測系の校正 (Calibration) はマイクロホンアレーの各マイクロホンのばらつきを次式で校正している。

$$\hat{S}_i(f) = S_i(f)\sigma_{g_i}/\sigma_{g_1}, \quad (9)$$

ここで、 i はチャンネル番号 ($i = 1, \dots, M$)、 f は周波数、 $\hat{S}_i(f)$ は校正された i チャンネルの観測信号のスペクトル、 $S_i(f)$ は校

正前の i チャンネルの観測信号のスペクトル、 σ_{g_i} は、マイクロホンアレー近傍に設置したスピーカで計測したインパルス応答の第一波面より算出した i チャンネルのゲインである。環境の変化の判定 (Detection of environmental variation) は、話者位置もしくは、音源の到来方位が変化したかを判定している。この判定により環境の変動が検出された場合、音源の到来方位と話者位置の情報、またそれらの移動予測 (現在は、左右に 5 度移動する状況を考慮) を用いてシミュレートし、複数の死角制御フィルタを設計する。この死角制御フィルタは、AMNOR の一括演算 [4] により算出している。この死角制御フィルタの設計は、雑音の到来方位と話者音声の到来方位の情報だけで可能であり、また死角制御フィルタの容量は小さいため、通信機能を利用してフィルタの設計を高速な外部サーバで行うことも、死角制御フィルタの大規模データベースを利用することも可能である。

このように設計した複数の死角制御フィルタを校正されたアレー信号に適用することで、複数の出力信号が次式のように得られる。

$$Y_j(f) = \sum_{i=1}^M \hat{S}_i(f)H_{ij}(f), \quad (10)$$

ここで、 j は死角制御フィルタの番号 ($j = 1, \dots, N$)、 $Y_j(f)$ は j 番目の死角制御フィルタの出力信号、 $H_{ij}(f)$ は j 番目の死角制御フィルタの i チャンネルのフィルタである。提案手法では、次式のように、これら N 個の出力信号の中から最小の信号を出力信号として利用する。

$$Y(f) = \min_{j=1}^N Y_j(f), \quad (11)$$

上記の一連の流れは、フレーム処理によって行われる。各フレームごとに、到来方位推定と話者位置推定を行うため、一見すると非常に計算コストが高いように考えられるが、実際には、CSP 法と ADM 法は Fourier 変換 2 回程度の計算コストで実現可能であり、従来のビームフォーマにおいても、話者の到来方位の推定が必要であるため、計算コストの増加は最小限となっている。また、観測信号のフーリエ変換を行っているため、複数の死角制御フィルタの適用も容易であるが、それぞれの死角制御フィルタを適用した信号に対する逆 Fourier 変換の計算コストは必要となる。

3.2 死角制御フィルタの構築

提案手法は、死角制御を行うためのステアリングフィルタを複数設計する必要がある。提案手法において、ステアリングフィルタの設計には AMNOR のアルゴリズムを利用する。ただし、擬似目的信号、雑音はすべてシミュレーションで構築するため、AMNOR の拘束条件下で雑音を最小化するステアリングフィルタは適応的に学習する必要はない。そこで、AMNOR の一括演算アルゴリズム [4] を利用して、式 (12)、(13) によりシミュレーションで作成した擬似目的信号と環境雑音によりステアリングフィルタを設計する。

$$S(\omega) = A^{(s)}D^{(s)}(\omega)X(\omega), \quad (12)$$

$$\mathbf{N}(\omega) = \frac{1}{M} \sum_{j=1}^M A_j^{(n)} \mathbf{D}_j^{(n)}(\omega) \mathbf{Y}_j(\omega), \quad (13)$$

ここで、 $\mathbf{S}(\omega)$ は擬似目的信号のスペクトル、 $\mathbf{N}(\omega)$ は環境雑音のスペクトル、 M は想定する環境雑音の最大数、 s は擬似目的信号に対応する添字、 n は環境雑音に対応する添字、 $\mathbf{D}^{(s)}(\omega)$ は擬似目的信号の到来方位に対応する遅延フィルタのスペクトル、 $\mathbf{D}_j^{(n)}(\omega)$ は環境雑音の到来方位に対応する遅延フィルタのスペクトル、 $A_j^{(s)}$ は、擬似目的信号の到来方位に対応するゲイン係数、 $A_j^{(n)}$ は、環境雑音の到来方位に対応するゲイン係数、 $\mathbf{X}(\omega)$ 、 $\mathbf{Y}_j(\omega)$ は互いに無相関な白色信号のスペクトルである。

また、式(13)の $\mathbf{N}(\omega)$ には、擬似目的信号の方位を除く全指向性雑音を付加することで、目的信号の方位以外の出力ゲインが大きくなる問題を解決できる。さらに、話者が複数いる場合には、それぞれの話者の方位を擬似目的信号の方位として、他方を雑音の方位と考えて、フィルタを設計すればよい。移動予測などを考慮すると、シミュレートする雑音環境の組み合わせをすべて網羅するためには、大きな計算コストが必要となる。

しかし、事前に様々な環境を考慮して、大規模データベースを構築するか、通信機能を利用して、外部の高速なクラスタマシンにてフィルタ設計を行うことでこれらの問題は解決できる可能性が高い。これは、観測したアレー信号そのものは非常に多量のデータとなるため、現在の高速な通信環境でもリアルタイムで転送することが非常に難しいが、到来方位や話者位置の情報、フィルタ係数などはデータ容量が非常に小さいため、リアルタイムでデータの送受信が可能であるためである。

ただし、提案手法は複数のステアリングフィルタを利用するため、従来手法に比べてメモリの使用量が大きくなる傾向がある。0~90度方位に目的信号が、90~180度方位に雑音が存在する環境で5度刻みでステアリングフィルタを設計した場合を想定すると、チャンネル数4、フィルタ長100tap、目的信号18方位、雑音18方位、浮動小数点精度8bytes換算で1MB程度のメモリ容量が一時記憶として必要となる。

4. シミュレーション

提案手法の性能と有用性を検証するために、残響のないシミュレーションにおいて雑音抑圧性能、音声認識性能(孤立単語認識)に関する評価を行った。

4.1 シミュレーション条件

目的信号源、雑音源、マイクロホンアレーの関係を図3に、シミュレーション条件を表1に示す。

提案手法の評価は雑音の変動する環境におけるシミュレーションや実験が妥当であるが、雑音の移動する条件はシミュレーションや実験の再現が難しい。そこで、本シミュレーションでは、雑音方位を2つとし、0.2 secごとに雑音の方位が切り替わる条件とした。また、話者数も2名として、目的音声と妨害音声は常に0 dBで混合されている条件とした。雑音は白色雑音を用い、雑音は-5~15 dBで付加した。このシミュレーション条件におけるSNR=25 dBとは、目的音声に対して、妨害音声0 dBと白色雑音25 dBが付加されている条件を表して

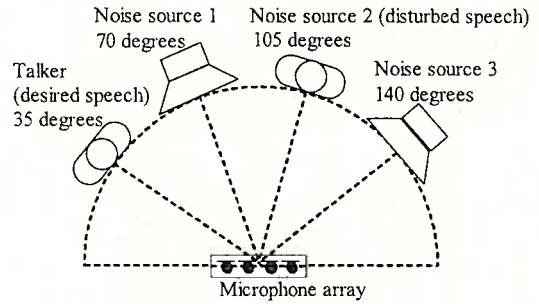


図3 音源とマイクロホンアレーの配置

Fig. 3 Experimental environment.

いる。マイクロホンアレーは4ch直線型、マイクロホン間隔は2.125 cm、目的音声は男性話者音声(35度方位より到来)、妨害音声は女性話者音声(105度方位より到来)とした。音声としてATR音素バランス216単語[9]を用い、妨害音声は逆順で付加した。

従来手法として、適応フィルタの学習区間を用いず受音した雑音下音声のみを利用して適応学習を行ったGriffiths-Jim型アレー(GJ array)によるシミュレーションも行った。

また、提案手法では、複数のステアリングフィルタの出力を各フレームごとに最小化することで出力信号を得るが、それぞれのステアリングフィルタの出力信号に対しても雑音抑圧性能と音声認識性能の評価を行った。各ステアリングフィルタは、雑音を設定した方位にて表記する。本シミュレーションにおいて、AMNORの劣化許容量は0.0%とした。

音声認識には、音声認識エンジンJulius[11]を用いた。雑音を付加しない場合の条件における音声認識率は96%であった。

4.2 シミュレーション結果

4.2.1 死角制御フィルタの指向性パターン

図4に設計した死角制御フィルタの指向性パターンを示す。図4(a)~(g)はそれぞれ、70度へのみ、105度へのみ、140度へのみ、70, 105度へのみ、70, 140度へのみ、105, 140度へのみ、70, 105, 140度に雑音を設定したステアリングフィルタの指向性パターンである。以降、NSFはNull-steering-filterの略語表記として用い、NSF for 70 deg.は70度方位に死角を形成した死角制御フィルタであることを示す。図4の指向性パターンより、多くの死角を形成する条件であるほど、急峻な死角を形成することが難しくなる傾向があることがわかる。

4.2.2 アレー処理後の出力信号波形と雑音抑圧比の評価

図5にSNR = -5 dBの条件における観測信号とアレー処理後の出力信号の波形を示す。図5(a)~(g)はそれぞれ70度へのみ、105度へのみ、140度へのみ、70, 105度へのみ、70, 140度へのみ、105, 140度へのみ、70, 105, 140度に雑音を設定したステアリングフィルタによる出力信号である。図5(h)は提案手法による出力信号、図5(i)はGJ arrayによる出力信号、図5(j)は観測信号の波形である。図5(a)~(g)のシミュレーション結果より、死角を形成していない方位から雑音到来する条件では、十分に雑音を抑圧できないことがわかる。しかし、すべての雑音の方位に対して死角を形成した図5(h)の場合、十分に雑音が抑

表 1 シミュレーション条件
Table 1 Simulation conditions.

Parameter settings	
Sampling Frequency	16 kHz (16 bit)
Frame length	32 msec (Hanning window)
Frame interval	16 msec
Filter design conditions	
Filter tap	101
Distortion	0.0
DOA of desired sound source	35 degrees
DOAs of noise sound source	70 degrees, 105 degrees, 140 degrees, 70 and 105 degrees, 70 and 140 degrees, 105 and 140 degrees, 70, 105 and 140 degrees
Ambient noise	30 dB
Test data(open)	
Desired sound source	Phoneme balanced 216 words in ATR-DB [9] (1 male)
Disturbed sound source	Phoneme balanced 216 words in ATR-DB (1 female)
Noise sound source	White noise
SNR	-5, 0, 5, 10, 15 dB
Acoustic model (HMM)	IPA monophone model (Clean training and gender model) [10]

表 2 SNR の評価

Table 2 The evaluation of SNR (Signal-to-Noise Ratio).

SNR [dB]	-5	0	5	10	15
NSF for 70 deg.	4.7	5.9	6.4	6.5	6.6
NSF for 105 deg.	4.0	8.9	13.6	17.9	21.1
NSF for 140 deg.	4.6	8.9	12.3	14.3	15.2
NSF for 70 and 105 deg.	8.3	12.2	14.8	16.2	16.7
NSF for 70 and 140 deg.	7.7	8.6	8.9	9.0	9.0
NSF for 105 and 140 deg.	4.4	9.4	14.2	18.7	22.4
NSF for 70, 105 and 140 deg.	11.3	14.4	16.0	16.7	16.9
Proposed method	13.2	16.3	18.8	22.1	24.7
GJ array	1.7	1.2	-0.0	-0.3	1.2

圧されている。さらに、提案手法では、図 5(a)~(h) までの出力信号の最小化を行っているため、さらに SNR が改善していることがわかる。また、図 5(j) の従来手法である GJ array のシミュレーション結果より、本シミュレーション条件では、適応学習が困難であり、十分な性能を発揮できないことがわかる。

表 2 にアレー出力後の SNR の結果を示す。表 2 のシミュレーション結果より、本シミュレーション条件では、従来手法である GJ array では SNR の改善が難しいことがわかった。また、高 SNR 環境、すなわち、雑音方位によって SNR が異なる環境では、NSF for 70, 105 and 140 deg. のようにすべての雑音方位に死角を形成するよりも、NSF for 105 and 140 deg. のように雑音が優位な方位に深い死角を形勢する方が SNR の改善量が高いことがわかった。これは、マイクロホン数が少ない

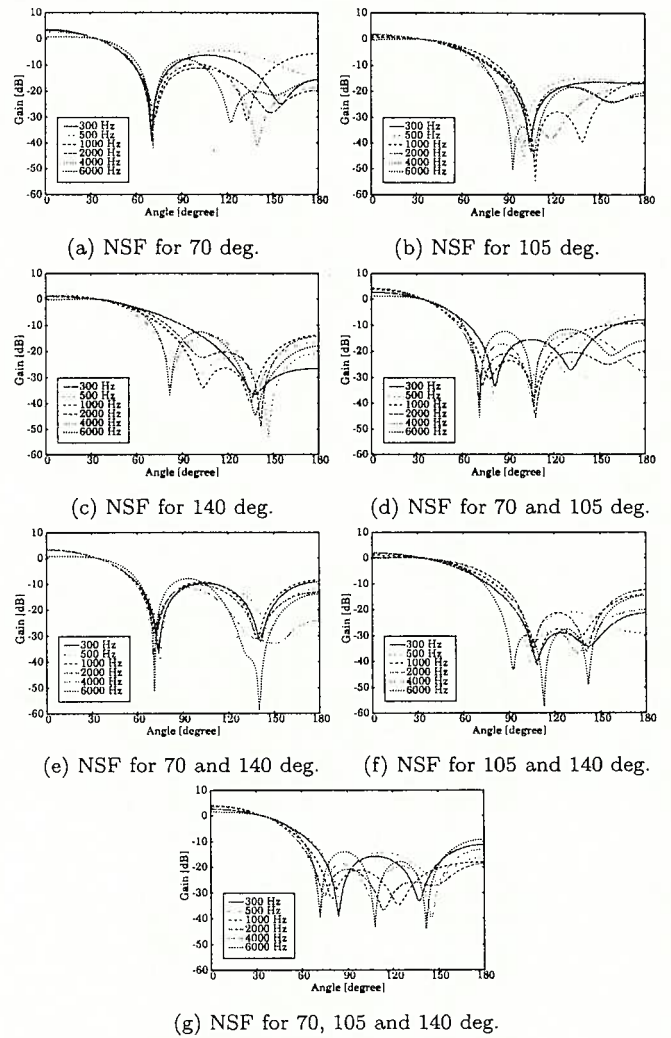


図 4 死角制御フィルタの指向性パターン

Fig. 4 Directionality pattern of null-steering filter.

条件であること、死角制御フィルタには雑音方位によって SNR が異なる環境を考慮していないことが原因として考えられる。このように、単一の死角制御フィルタのみを利用した場合、各雑音の SNR が変化することで性能が劣化する傾向があったが、複数の死角制御フィルタを併用する提案手法は低 SNR から高 SNR 環境まで高い SNR の改善が可能であることがわかった。

4.2.3 音声認識結果

表 3 に各環境における各手法の音声認識率を示す。表 3 の音声認識結果より、本シミュレーション条件では従来手法である GJ array では音声認識を行うことが難しいことがわかった。また、表 2 の SNR の改善量と同様に、高 SNR 環境、すなわち、雑音方位によって SNR が異なる環境では、NSF for 70, 105 and 140 deg. のようにすべての雑音方位に死角を形成するよりも、NSF for 105 and 140 deg. のように雑音が優位な方位に深い死角を形勢する方が音声認識率が高いことがわかった。ただし、低 SNR 環境においては、死角を形成できていない方位が存在するため、NSF for 105 and 140 deg. の音声認識率は低下する傾向があった。また、提案手法は、低 SNR ~ 高 SNR 環境において高い SNR、音声認識率となっていることから提案手法の有効性を確認できた。

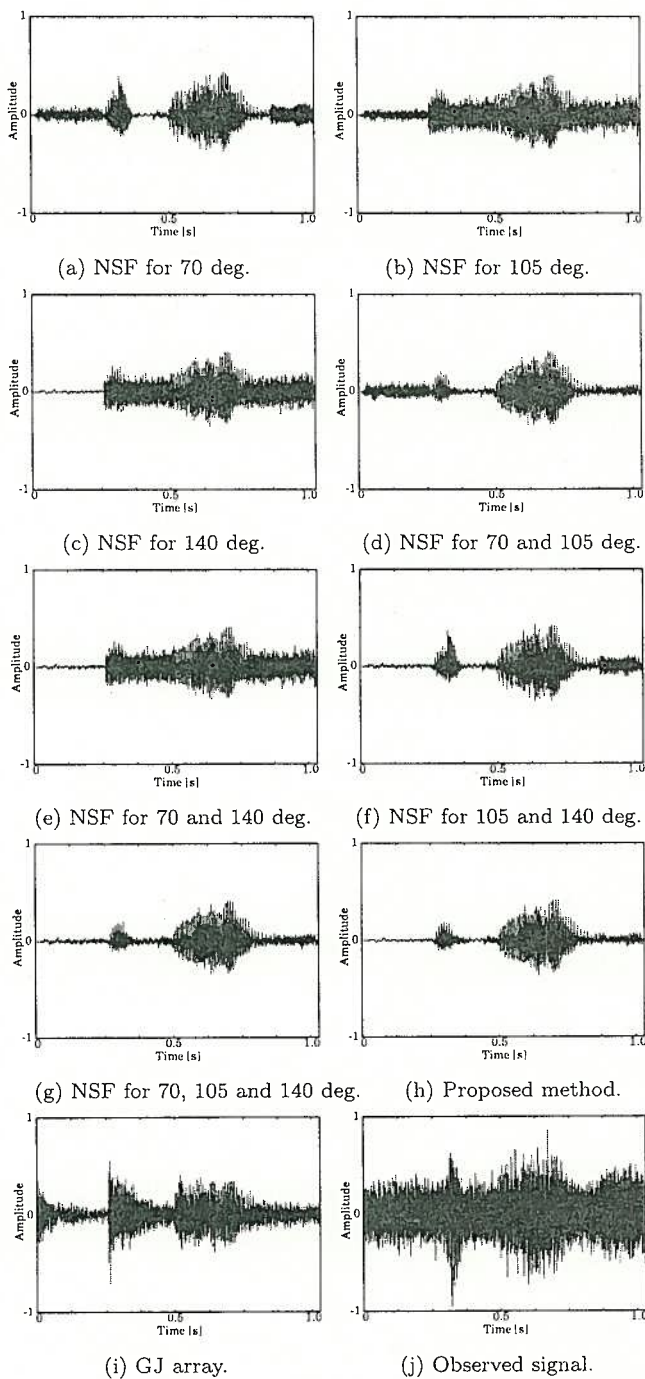


図 5 出力信号と観測信号の波形

Fig. 5 Waveform of output signals and observed signal.

5. むすび

雑音の変動する環境においても頑健なビームフォーミングを実現するために、本稿では話者と音源の位置推定に基づいて複数の死角制御を行うステアリングフィルタを利用することで、雑音環境が変動する環境においても頑健にビームフォーミング可能な死角制御型ビームフォーマを提案した。シミュレーションにおける評価の結果、提案手法である GJ array よりも SNR、音声認識率が改善可能であったことより、雑音環境の変動する条件における有用性を確認できた。今後は、実環境における実験を行い提案手法の有効性を検証したい。

表 3 音声認識率 [%] の評価

Table 3 The evaluation of ASR (Automatic Speech Recognition) [%].

SNR [dB]	-5	0	5	10	15
NSF for 70 deg.	29.6	31.0	31.5	28.7	31.9
NSF for 105 deg.	26.7	50.5	67.6	80.1	83.8
NSF for 140 deg.	21.3	34.7	52.3	63.9	66.7
NSF for 70 and 105 deg.	57.9	65.3	66.7	69.4	71.3
NSF for 70 and 140 deg.	34.7	34.7	37.5	36.6	36.6
NSF for 105 and 140 deg.	22.7	46.3	65.7	85.6	89.8
NSF for 70, 105 and 140 deg.	64.8	69.0	69.0	69.4	72.7
Proposed method	72.7	74.1	78.7	86.1	87.0
GJ array	2.3	2.3	2.8	1.9	2.8

また、さらに計算コストに余裕がある場合、並列に従来の適応型アレーを動作させれば、環境の変動が小さい環境では、従来の逐次学習を行った適応型アレーの出力信号を用いることで、従来の適応型アレーと同程度の性能を実現することも可能である。そこで、従来の適応型アレーと組み合わせる検討も行いたい。

謝辞 本研究の一部は、科学研究費補助金 No. 20700169, 21560454 による研究助成を受けた。

文 献

- [1] J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, vol.78, no.5, pp.1508-1518, Nov. 1985.
- [2] L.J. Griffiths, C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol.AP-23, no.1, pp.27-34, Jan. 1982.
- [3] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-34, pp.1391-1400, 1986.
- [4] 一ノ瀬 裕, 片岡 章俊, "適応型雑音抑圧マイクロホンアレー," *NTT R&D*, Vol. 38, No. 10, pp. 1153-1162, 1986.
- [5] 大賀 寿郎, 山崎 芳男, 金田 豊, "音響システムとデジタル処理," コロナ社, 1995.
- [6] 西浦 敬信, 山田 武志, 中村 哲, 鹿野 清宏, "マイクロホンアレーを用いた CSP 法による複数音源位置推定," *信学論*, Vol. J83-D-II, No. 8, pp. 1713-1721, 2000.
- [7] 上保徹志, 中迫昇, 大亦紀光, 板垣英恵, "帯域雑音信号による複数対象物の距離推定," *電学論 C*, Vol.128, No.7, pp.1117-1122, 2008.
- [8] 中山雅人, 中迫昇, 篠原寿広, 上保徹志, "可聴音の送信波と反射波の位相干渉に基づく話者位置推定," *電学論 C*, Vol.130-C, No.11, pp.1994-2000, 2010.
- [9] K. Takeda, Y. Sagisaka, and S. Katagiri, "Acoustic-Phonetic Labels in a Japanese Speech Database," *Proc. European Conference on Speech Technology*, vol.2, pp.13-16, Oct. 1987.
- [10] 鹿野清宏, 伊藤克巨, 河原達也, 武田一哉, 山本幹雄, "音声認識システム," オーム社出版局, 2001.
- [11] A. Lee, T. Kawahara, and K. Shikano, "JULIUS - An Open Source Real-Time Large Vocabulary Recognition Engine," *EUROSPEECH2001*, pp.1691-1694, Sept. 2001.