

2010年度（平成22年度）

博士論文

音響測距とマイクロホンアレーを用いた  
雑音下における遠隔発話音声受音に関する研究

中山 雅人



# 音響測距とマイクロホンアレーを用いた 雑音下における遠隔発話音声受音に関する研究

中山 雅人

## 内容梗概

人間がコミュニケーションをする際にストレスなく最も効率のよい手段として音声によるハンズフリーインタフェースの実現が強く望まれている。ストレスなく利用できるハンズフリー音声インタフェースとして、遠隔設置型のマイクロホンが検討されているが、口元から離れた位置のマイクロホンで音声を受音するため、環境雑音や残響の影響により音声が歪みを受けて劣化する問題がある。この問題を解決し、遠隔発話の音声を高品質に受音・認識するために、複数のマイクロホン素子を空間的に配置したマイクロホンアレーと呼ばれる受音器を利用する技術が検討されている。本論文では、このマイクロホンアレーを用いて遠隔発話音声を高品質に受音する手法を検討する。

マイクロホンアレー技術による高品質遠隔発話音声受音では、話者の方位もしくは位置を推定し、推定した話者位置を利用して話者の音声のみを受音する指向特性の制御（ビームフォーミング）を実現する必要がある。

そこで、本論文では、(I) 雑音下においても頑健に動作する位相干渉に基づく音響測距法の検討、(II) 位相干渉に基づく音響測距法をマイクロホンアレーに拡張した話者位置推定の検討、(III) 母音/子音特徴量を利用した適応型アレーの検討、以上3つの検討を行うことによって遠隔発話音声を高品質に受音する手法の実現を試みる。

(I) では、位相干渉に基づく音響測距法を雑音環境下でも頑健に動作するように理論の拡張を行い、この測距法に適した雑音抑圧手法を提案する。この音響測距法は、対人・対物センサとして利用可能である。従来の雑音抑圧手法である同期加算とスペクトルサブトラクションの考え方を位相干渉に基づく音響測距法に導入すること

で低 SNR ( Signal-to-Noise Ratio ) 環境でも測距可能な手法を検討する。また，実環境において雑音環境を考慮した評価実験の結果，提案する音響測距法が低 SNR の雑音環境においても有効であることを確認した。

(II) では，音響測距法で 1 つであったマイクロホンをマイクロホンアレーに拡張することで，話者 ( 対象物 ) までの距離だけでなくその位置まで推定する手法を提案する。加えて，話者の発話/非発話によって，話者の発話を利用する CSP ( Cross-power Spectrum Phase analysis ) 法による音源方位 ( Direction Of Arrival; DOA ) 推定と提案手法を切り替えて適用する手法も検討する。また，実環境における評価実験の結果，提案手法の有効性を確認した。

最後に，(III) では，(I), (II) で検討してきた話者位置推定によって得られた話者位置を利用して，母音/子音 ( Vowel/Consonant; V/C ) 特徴量に特化した適応型アレーを目的信号の劣化を許容する拘束条件下で構築することにより雑音下音声認識性能の改善した手法を提案する。提案手法は雑音下観測音声の V/C 時系列に対して正確な V/C ビームフォーミングの適用が必要不可欠であるため，音素認識による V/C セグメンテーションの検討も行う。また，実環境における評価実験の結果，低 SNR の雑音環境において提案手法の有効性を確認した。

## キーワード

マイクロホンアレー, 高品質遠隔発話音声受信, 音響測距法, 話者位置推定, 適応型アレー, 母音/子音特徴量

# Distant-talking Speech Acquisition Using Acoustic Distance Measurement and Microphone-array in Noisy Environments

Masato Nakayama

## Abstract

A hands-free speech interface has become essential as a stress-free, efficient controller of information equipment by general users. And it requires distant-talking speech acquisition. However, distant-talking speech is distorted and degraded by environmental noises and reverberations. Microphone-array technology is an ideal candidate for overcoming this problem. In high-quality distant-talking speech acquisition using a microphone-array, two technologies are required: 1) direction (or position) estimation of the talker and 2) beamforming by steering high-sensitive directivity for the estimated direction (or position) of the talker.

Accordingly, in this thesis, the author discusses a noise-robust distant-talking speech acquisition method consisting of three components: (I) robust acoustic distance measurement based on interference in noisy environments, (II) talker localization based on interference between transmitted and reflected audible sounds, (III) noise-robust distant-talking speech acquisition using an adaptive microphone-array based on vowel/consonant features.

As component (I), the acoustic distance measurement method based on the interference between the transmitted wave and the reflected waves is expanded to a noise-robust method in a noisy environment, which is necessary for a distance sensor. Distance sensors are very important in numerous engineering fields. In order

to measure distance in a low-signal-to-noise ratio (low-SNR) environment, the author introduces a concept from synchronous addition and spectral subtraction to the proposed method. Evaluation experiments in a real environment reveal that the proposed method was valid and effective because the proposed method could measure distance in a low-SNR environment.

As component (II), a talker localization method is proposed by expanding the acoustic distance measurement method to a microphone-array because the acoustic distance measurement method can measure distance with component (I) in a noisy environment. In addition, the talker's speech can be used for estimating the direction of arrival (DOA) of the talker. Therefore, the author proposes the combination of the proposed method and the cross-power spectrum phase (CSP) analysis method, which is a DOA estimation method. As a result of evaluation experiments in a real environment, the effectiveness of the proposed method was confirmed.

Finally, as component (III), using a talker position estimated by the talker localization described above, the author proposes an adaptive microphone-array based on vowel/consonant features designed under the condition of a constraint that tolerates distortion of the desired signal. In addition, vowel/consonant segmentation is required in order to realize the proposed method. Therefore, the author uses phoneme recognition with the hidden Markov model (HMM) for vowel/consonant segmentation. The results of the evaluation experiments revealed that the proposed method was effective in a low-SNR environment.

**Keywords:**

Microphone-array, High-quality distant-talking speech acquisition, Acoustic distance measurement, Talker localization, Adaptive microphone-array, Vowel/consonant features

# 目次

第1章 序論	1
1.1. 研究の背景と目的	1
1.2. 本論文の構成	4
第2章 マイクロホンアレーを用いた遠隔発話音声受音の基礎	5
2.1. はじめに	5
2.2. 話者方位推定	5
2.2.1 2素子マイクロホンアレーによる受音	7
2.2.2 CSP法による音源の到来方位推定	8
2.3. ビームフォーミング技術	9
2.3.1 遅延和アレーの原理	10
2.3.2 適応型アレーの原理	11
2.3.3 AMNOR (Adaptive Microphone-array for NOise Reduction)	14
2.3.4 S-AMNOR (average Speech spectrum-based AMNOR)	16
2.4. まとめ	18
第3章 可聴音の送信波と反射波の位相干渉に基づく距離推定	19
3.1. はじめに	19
3.2. 実環境における可聴音の送信波と反射波の位相干渉に基づく距離推定の原理	20
3.3. 雑音環境に頑健な可聴音の送信波と反射波の位相干渉に基づく測距法	25
3.3.1 従来の前処理による雑音抑圧手法	25
3.3.2 位相干渉による音響測距法に適した雑音抑圧手法	25

3.4.	実環境における評価実験	26
3.4.1	実験条件	27
3.4.2	実験結果	31
3.5.	まとめ	41
<b>第4章</b>	<b>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</b>	<b>42</b>
4.1.	はじめに	42
4.2.	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	43
4.3.	実環境における評価実験	45
4.3.1	実験条件	45
4.3.2	実験結果	48
4.4.	まとめ	56
<b>第5章</b>	<b>適応型アレーを用いた母音/子音特徴量に基づく音声強調</b>	<b>58</b>
5.1.	はじめに	58
5.2.	VC-AMNOR ( V/C 特徴量を利用した適応型アレー ) の提案	60
5.2.1	VC-AMNOR のアルゴリズム	61
5.2.2	音素の特徴量	62
5.2.3	音素認識を用いた V/C セグメンテーション	62
5.2.4	V/C に対して歪みを軽減するビームフォーミング用フィルタ の設計	64
5.3.	実環境評価実験	67
5.3.1	実験条件	67
5.3.2	実験結果	72
5.4.	まとめ	80
<b>第6章</b>	<b>結論</b>	<b>81</b>
6.1.	本論文のまとめ	81
6.2.	今後の課題	82
	謝辞	84



参考文献	86
研究業績	97

# 目 次

1.1	単一マイクロホンを用いた遠隔発話音声の受音 . . . . .	2
1.2	7素子等間隔直線型マイクロホンアレー . . . . .	3
2.1	マイクロホンアレーを用いた遠隔発話音声の高品質受音 . . . . .	6
2.2	2素子マイクロホンアレーで受音される信号 . . . . .	7
2.3	$M$ 素子マイクロホンアレーを用いた遅延和アレー . . . . .	10
2.4	2素子マイクロホンアレーによる減算型アレー . . . . .	12
2.5	アレーの目的信号に対する流れ図 . . . . .	13
2.6	$M$ 素子マイクロホンアレーにおける AMNOR 方式 . . . . .	15
2.7	白色雑音のスペクトル . . . . .	17
2.8	音声平均スペクトルを重み付けした信号のスペクトル . . . . .	17
3.1	対象物が一つの場合の $\Delta p(f, 0)$ . . . . .	24
3.2	対象物が一つの場合の距離スペクトル . . . . .	24
3.3	雑音環境下の音響測距実験における実験環境 . . . . .	27
3.4	雑音環境下音響測距実験における雑音信号 . . . . .	28
3.5	雑音環境下音響測距実験における雑音信号のパワースペクトル . . . . .	29
3.6	雑音環境下音響測距実験における $\Delta p(f, 0)$ (雑音: 白色雑音) . . . . .	32
3.7	雑音環境下音響測距実験における距離スペクトル (雑音: 白色雑音) . . . . .	33
3.8	測距可能であった最低 SNR (雑音: 白色雑音) . . . . .	34
3.9	NRR の実験結果 (雑音: 白色雑音) . . . . .	34
3.10	RSR の実験結果 (雑音: 白色雑音) . . . . .	35
3.11	PSFR の実験結果 (雑音: 白色雑音) . . . . .	35

3.12	雑音環境下音響測距実験における $\Delta p(f, 0)$ (雑音: ファンノイズ, SNR = 0 dB)	36
3.13	雑音環境下音響測距実験における距離スペクトル (雑音: ファンノイズ, SNR = 0 dB)	37
3.14	測距可能であった最低 SNR (雑音: ファンノイズ)	39
3.15	NRN の実験結果 (雑音: ファンノイズ)	39
3.16	RSR の実験結果 (雑音: ファンノイズ)	40
3.17	PSFR の実験結果 (雑音: ファンノイズ)	40
4.1	話者位置推定を行う提案手法における話者, スピーカとマイクロホンの関係	43
4.2	話者位置推定実験における実験状況 (上方図)	47
4.3	話者位置推定実験における実験状況 (正面図)	47
4.4	話者位置推定実験における提案手法の距離スペクトル (Room A, 0.5 m, 60 度)	49
4.5	話者位置推定実験における提案手法の距離スペクトル (Room A, 1.0 m, 60 度)	49
4.6	話者位置推定実験における提案手法の距離スペクトル (Room B, 1.0 m, 60 度)	49
4.7	話者位置推定実験における提案手法による距離推定結果 (Room A, 0.5 m)	50
4.8	話者位置推定実験における提案手法による距離推定結果 (Room A, 1.0 m)	50
4.9	話者位置推定実験における提案手法による距離推定結果 (Room B, 1.0 m)	51
4.10	話者位置推定実験における提案手法による DOT 推定結果 (Room A, 0.5 m)	52
4.11	話者位置推定実験における提案手法による DOT 推定結果 (Room A, 1.0 m)	52

4.12	話者位置推定実験における提案手法による DOT 推定結果 ( Room B, 1.0 m )	53
4.13	CSP 係数 ( Room A, 0.5 m, 60 度 )	54
4.14	CSP 係数 ( Room A, 1.0 m, 60 度 )	54
4.15	CSP 係数 ( Room B, 1.0 m, 60 度 )	54
4.16	CSP 法による DOA 推定結果 ( Room A, 0.5 m )	55
4.17	CSP 法による DOA 推定結果 ( Room A, 1.0 m )	55
4.18	CSP 法による DOA 推定結果 ( Room B, 1.0 m )	56
5.1	VC-AMNOR の概略図	61
5.2	母音のスペクトルを重み付けした信号のスペクトル	66
5.3	子音のスペクトルを重み付けした信号のスペクトル	66
5.4	VC-AMNOR の評価実験における音源とマイクロホンアレーの配置 ( Room A )	68
5.5	VC-AMNOR の評価実験における音源とマイクロホンアレーの配置 ( Room B )	68
5.6	VC-AMNOR の評価実験における劣化許容量 0.2 , SNR -5 dB の条件における波形	76
5.7	VC-AMNOR の評価実験における劣化許容量 0.2 , SNR -5 dB の条件におけるスペクトログラム	77

# 表 目 次

3.1	雑音環境下の音響測距実験で使用した機材 . . . . .	30
4.1	話者位置推定における実験で使用した機材 . . . . .	46
5.1	日本語の音素表記 . . . . .	63
5.2	V/C 平均振幅スペクトルの作成条件 . . . . .	65
5.3	VC-AMNOR の評価実験における収録条件 . . . . .	69
5.4	VC-AMNOR の評価実験における収録機材 . . . . .	69
5.5	VC-AMNOR の評価実験における実験条件 . . . . .	70
5.6	音響実験室 ( Room A ) の V/C 識別率 . . . . .	72
5.7	研究室内 ( Room B ) の V/C 識別率 . . . . .	72
5.8	VC-AMNOR 評価実験の劣化許容量 0.01 における SNR の改善量 . . . . .	73
5.9	VC-AMNOR 評価実験の劣化許容量 0.2 における SNR の改善量 . . . . .	73
5.10	VC-AMNOR 評価実験の劣化許容量 0.01 における SD . . . . .	74
5.11	VC-AMNOR 評価実験の劣化許容量 0.2 における SD . . . . .	74
5.12	VC-AMNOR の評価実験における音声認識結果 . . . . .	75

# 第1章 序論

## 1.1. 研究の背景と目的

近年の情報通信技術( Information and Communication Technology; ICT )の急速な発展に伴い、様々なものが情報化されて、情報化されたコンテンツやサービスを「いつでも、どこでも、何でも、誰でも」使えるユビキタスなネットワーク社会の実現が望まれている。しかし、情報格差( デジタル・ディバイド )のないICT社会の実現には、これまでのPCなどの情報端末を積極的に利用していなかった人々にもストレスなく利用可能なインタフェースが必要不可欠である。そこで、人間がコミュニケーションをする際に最も効率のよいインタフェースとして音声インタフェースの実現が強く望まれている。音声インタフェースの実現によって、これまで情報化が困難であった分野においてもより一層の情報化による利便性の向上が期待でき、ICTの恩恵・利便を実感できる社会の基盤を担う革新的な技術として非常に期待されている。

しかし、現在の音声インタフェースは携帯電話やヘッドセットなどの接話型マイクロホンが主流である。これは、音声インタフェースを実現するためには、音声を高品質に受音・認識する必要があるためである。口元近くの接話型マイクロホンにより音声を受音すると、容易に高品質の受音が可能であるが、マイクロホンを身につける煩わしさは利用者への大きな負担に繋がり、音声インタフェースの普及を阻害する大きな要因とされている。そこで、図 1.1 のように遠隔設置型のマイクロホンの利用が検討されているが、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、環境雑音や残響の影響により音声が劣化する問題がある。この問題を解決し、利用者がマイクロホンを気にせず、どこで発話された音声でも高品質に受音可能なハンズフリー音声インタフェースの実現は急務である。

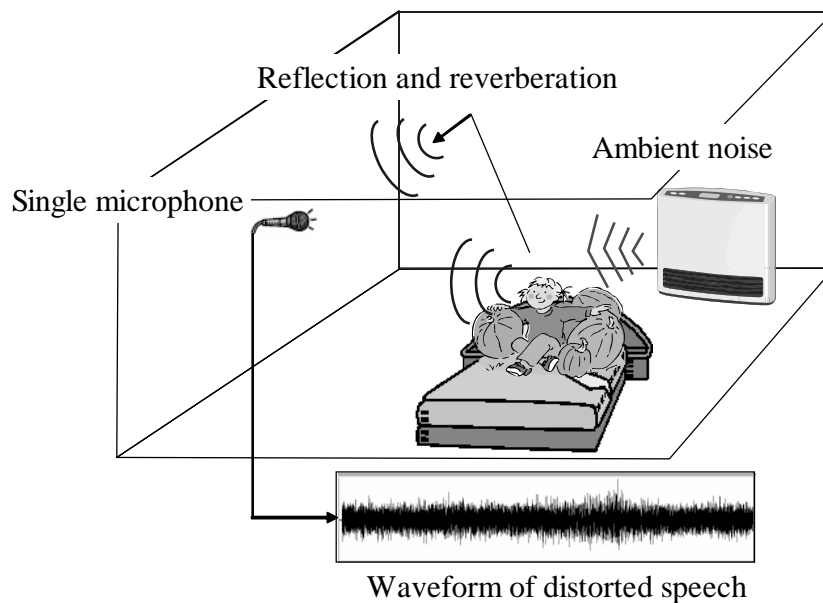


図 1.1 単一マイクロホンを用いた遠隔発話音声の受信

この問題を解決し，遠隔発話の音声を高品質に受信するために，図 1.2 のような複数のマイクロホン素子を空間的に配置したマイクロホンアレーと呼ばれる受信器を利用する技術が検討されている．このマイクロホンアレー技術を用いると，空間的な情報を利用して，口元（音声の音源）方位を高感度に雑音の音源方位を低感度に作り変えるビームフォーミング（指向性の制御）が可能であり，音声のみを高品質に受信可能となる [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]．そのため，マイクロホンアレー技術は，音声認識 [11, 12, 13, 14]，対話型ロボット [15, 16, 17] や音声通信 [18, 19, 20] など広く応用されている．

マイクロホンアレー技術による遠隔発話音声の高品質受信では，前述のビームフォーミングによる音声受信を行うために，事前に発話者の位置情報を取得しておくことが非常に重要となる．これは，音声の到来方位（空間情報）が既知でなければ，ビームフォーミングができないためである．また，発話者の位置が既知となれば，その発話者のみに音情報を伝えることも可能となる．音源の方位推定は，マイクロホンアレー技術を利用することで可能である．ただし，話者の発話を利用した

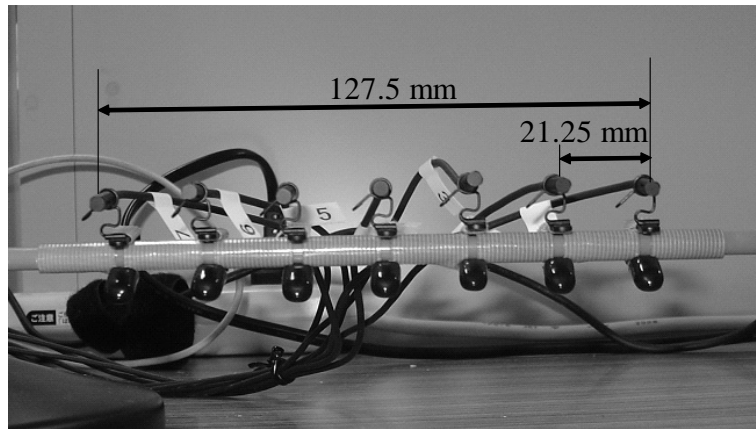


図 1.2 7 素子等間隔直線型マイクロホンアレー

方位推定であるため、話者が発話していない区間では方位を推定できず、話者が移動する環境では推定が難しい問題がある [21]。また、アレー長が短い直線型マイクロホンアレーでは、話者位置（話者の方位と距離）を推定することは困難である [22]。複数のマイクロホンを空間的に立体的に配置することで、話者位置（音源位置）まで推定・高音質受音する手法も提案されているが、多数のマイクロホン素子を持つ大規模マイクロホンアレーが必要となり、非常に高コストとなる問題がある [23, 24, 25]。そこで、本論文では、低コストで構築可能な素子数の少ない等間隔直線型マイクロホンアレー（以後、単にマイクロホンアレーと表記する）の利用を前提条件とする。

一方、話者（対象物）までの距離を可聴音の送信波と反射波の位相干渉に基づいて距離推定する技術も提案されている [27, 28, 29, 30, 31, 32, 33, 34, 35, 36]。その原理は、送信波と反射波の干渉によって生じる定在波のパワースペクトルが周期関数であり、その周期がマイクロホン - 対象物間の距離に逆比例することである。この技術では、スピーカとその近傍に設置したマイクロホンを利用して、特定方位に存在する対象物までの距離を推定する。この手法をマイクロホンアレーに拡張することで、話者位置の推定が可能となる。ただし、この手法は、反射音の計測を行うという観点から言えば遠隔マイクロホンを利用する手法であるため、雑音の影響によって測距性能が低下する問題がある。



そこで本論文では、まず可聴音の位相干渉による測距法を雑音環境下でも頑健に動作させるための検討を行い、次に可聴音の位相干渉による測距法をマイクロホンアレーに拡張した話者位置推定手法を用いて話者位置推定を行い、最後にその話者位置を利用してビームフォーミングによる雑音下遠隔発話音声より目的音声のみを抽出することで、音声を高品質に受音する手法について検討する。

## 1.2. 本論文の構成

本論文の構成を以下に述べる。本論文は6章から構成される。2章では、マイクロホンアレーを用いた高品質音声受音では必要不可欠な技術である、話者方位推定、ビームフォーミングの原理と既存手法の問題点について述べる。3章では、可聴音の送信波と反射波の位相干渉に基づく距離推定の原理と雑音環境に頑健な音響測距法について述べる。4章では、可聴音の送信波と反射波の位相干渉に基づく距離推定をマイクロホンアレーに拡張した話者位置推定手法について述べる。5章では、音声を母音/子音に分類して母音/子音特徴量に適したビームフォーミングを音声の劣化を許容する拘束条件下で行う、雑音下遠隔発話音声の高品質音声受音手法について述べる。最後に、6章で本論文の結論と今後の課題について述べる。

## 第2章 マイクロホンアレーを用いた 遠隔発話音声受音の基礎

### 2.1. はじめに

マイクロホンアレー技術を用いることによって話者方位の推定が可能であり，その話者方位から図 2.1 のように話者方位に対しては高感度の指向特性，雑音方位に対しては死角を形成するビームフォーミングを行うことで，遠隔発話音声を高品質に受音できる．本章では，マイクロホンアレーを用いた遠隔発話音声受音の要素技術である話者方位推定とビームフォーミングの原理と既存の手法について述べる．2.2 節では，既存の話者方位推定で利用される音声の到来方位 ( Direction Of Arrival; DOA ) 推定手法についての概要を述べ，DOA 推定の原理と特に白色化相互相関 ( Cross-power Spectrum Phase analysis; CSP ) 法の詳細について述べる．2.3 節では，マイクロホンアレーのビームフォーミング技術である遅延和アレーと適応型アレーの原理，特に本論文において提案手法の基礎となる適応型アレーである AMNOR ( Adaptive Microphone-array for NOise Reduction ) ，音声受音用途に改良された AMNOR ( average Speech spectrum-based AMNOR; S-AMNOR ) の詳細について述べる．最後に，2.4 節で本章のまとめを述べる．

### 2.2. 話者方位推定

マイクロホンアレー技術によるハンズフリー音声インタフェースの実現には，話者方位 ( もしくは話者位置 ) が既知であるか，逐次推定する必要がある．利用者が音声インタフェースを利用するために発話する際には，当然ながら発話の音源は口元となるため，マイクロホンアレーに到来する発話 ( 音声 ) の到来方位 ( Direction Of

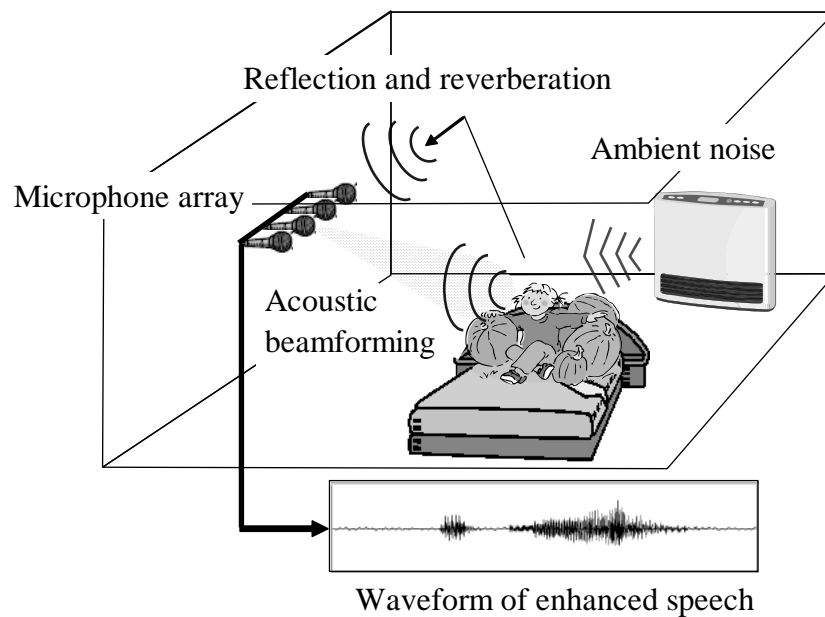


図 2.1 マイクロホンアレーを用いた遠隔発話音声の高品質受音

Arrival; DOA) を推定することで話者方位を推定することが可能となる。DOA 推定は話者位置推定以外にも様々な応用分野が存在するため、多種多様なマイクロホンアレーを用いた DOA 推定手法が提案されている [1, 37, 38, 39, 40, 41, 42, 43, 44, 45]。

遅延和 (Delay-and-Sum; DS) 法は、各方位にのみ高感度の指向特性を形成することで得られる信号の分散が最大となる方位を音源方位とする手法である [1, 37]。しかし、DS 法には周波数依存性があり、高い空間分解能を得るためには多素子でアレー長の長い大規模マイクロホンアレーを必要とする問題がある [23, 24, 25]。最小分散 (Minimum Variance; MV) 法は、各方位にのみ指向特性の死角を形成することで得られる信号の分散が最小になる方位を音源方位とする手法である [38, 39]。線形予測法は、受音信号の線形予測符号 (Linear Predictive Coding; LPC) [46] に基づく手法である [40, 41]。MUSIC (Multiple Signal Classification) 法は、信号の到来方位ベクトル (信号部分空間) と雑音部分空間の直交性を利用する手法である [42, 43]。

MV 法、線形予測法と MUSIC 法において、 $M$  素子等間隔直線型マイクロホンア

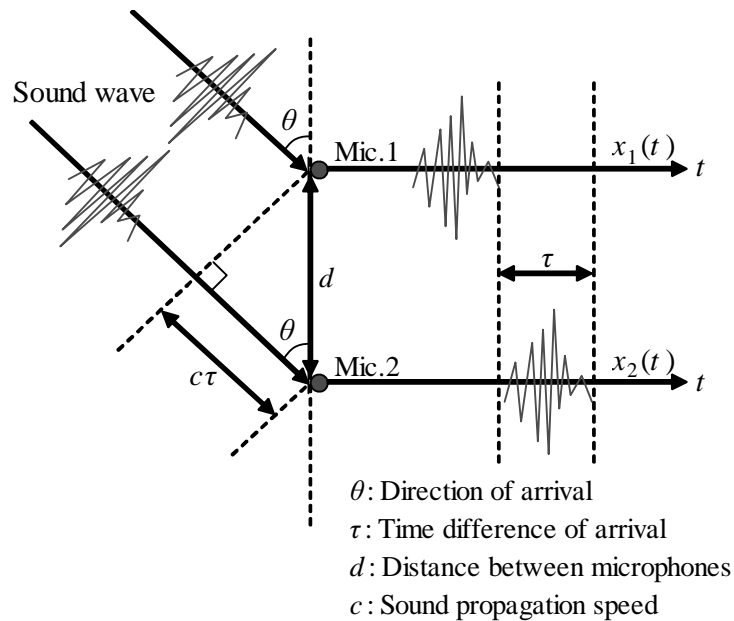


図 2.2 2 素子マイクロホンアレーで受信される信号

レーを用いた場合に推定できる DOA の最大数は  $M - 1$  個となる [24] . そのため , 音声だけでなく雑音と反射音を受音する雑音・残響環境下において高精度に話者音声の到来方位を推定するためには , DS 法と同様に大規模なマイクロホンアレーが必要となる . 一方 , 受信信号間の時間差情報に基づく白色化相互相関 ( Cross-power Spectrum Phase; CSP ) 法は , 雑音・残響環境下では性能が劣化するものの 2 素子マイクロホンアレーのみで複数の DOA を推定可能である [44, 45] . そこで , 本節では 2 素子マイクロホンアレーによる受信と CSP 法による音源の到来方位推定について述べる .

### 2.2.1 2 素子マイクロホンアレーによる受信

図 2.2 は , 自由空間においてマイクロホン間隔  $d$  で配置されたマイクロホン Mic.1, Mic.2 に角度  $\theta$  から音速  $c$  で平面波として到来している音響信号  $x(t)$  を受信している様子である . 図 2.2 において , Mic.1,2 の受信信号をそれぞれ  $x_1(t), x_2(t)$  , 音源から

Mic.1,2 に音響信号が受信されるまでの時間をそれぞれ  $\tau_1, \tau_2$  とすると、音響信号の到来時間差 ( Time Difference Of Arrival; TDOA ) と DOA は以下の式により表せる。

$$x_1(t) = x(t - \tau_1), \quad (2.1)$$

$$x_2(t) = x(t - \tau_2), \quad (2.2)$$

$$\begin{aligned} \tau &= \tau_2 - \tau_1 \\ &= \frac{d \cos \theta}{c}, \end{aligned} \quad (2.3)$$

$$\theta = \cos^{-1} \left( \frac{c\tau}{d} \right), \quad (2.4)$$

ここで、 $t$  は時間、 $\tau$  と  $\theta$  はそれぞれ音響信号の TDOA、DOA を表す。このように、マイクロホンアレーで受信された音響信号の TDOA を推定することで、音響信号の DOA、すなわち音響信号を音声と考えれば話者の方位を推定できる。

### 2.2.2 CSP 法による音源の到来方位推定

CSP 法は、2 素子マイクロホンアレーの受信信号を利用して音源の到来方位推定を推定する手法である [44, 45]。

まず、式 (2.5) のように、2 素子マイクロホンアレーの受信信号をフーリエ変換して、受信信号それぞれの振幅スペクトルの正規化により白色化した受信信号間のクロススペクトルを求めた後、逆フーリエ変換することで CSP 係数 ( 白色化されたアレー信号間の相互相関係数 ) を得る。

$$\text{CSP}(t) = \text{IFT} \left[ \frac{\text{FT}[x_1(t)]\text{FT}[x_2(t)]^*}{|\text{FT}[x_1(t)]||\text{FT}[x_2(t)]|} \right], \quad (2.5)$$

ここで、 $\text{CSP}(t)$  は CSP 係数、 $\text{FT}[\cdot]$  はフーリエ変換 ( Fourier Transform; FT )、 $\text{IFT}[\cdot]$  は逆フーリエ変換 ( Inverse Fourier Transform; IFT )、 $*$  は複素共役を表す。

次に、CSP 係数のピークが存在する時間差 ( 受信信号間の相関が強くなる時間差 ) を TDOA とする。音源を 1 つ ( 話者が 1 名 ) とした場合では、次式により TDOA を推定し、TDOA から DOA すなわち話者方位を推定できる。

$$\tau = \underset{t}{\text{argmax}}(\text{CSP}(t)), \quad (2.6)$$

$$\theta = \cos^{-1} \left( \frac{c\tau}{d} \right). \quad (2.7)$$

話者が複数の場合は，式 (2.6) により得られた  $\tau$  における CSP 係数のピークである  $\text{CSP}(\tau)$  を削除して，式 (2.6),(2.7) を繰り返す操作を話者の数だけ行えばよい．

CSP 法では，式 (2.6) のように，白色化したクロススペクトルを求めることで，受信信号間の位相差のみを利用しているため，受信信号の振幅特性に依存しない DOA 推定を実現できる [47]．しかし，雑音・残響環境下，特に指向性雑音環境下において DOA 推定性能が劣化する問題がある．また，2 素子マイクロホンアレーを利用する場合には，DOA は推定できるが話者の位置（話者までの距離）は推定できない [22]．

## 2.3. ビームフォーミング技術

ビームフォーミング技術では，話者方位を高感度に，雑音の音源方位を低感度（感度の死角）に作り変えるビームフォーミングによって，残響・雑音環境下で受信した劣化した音声から目的音声のみを抽出することで高品質な受信を実現する．これまでに，マイクロホンアレーを用いたビームフォーミングを行う様々な手法が提案されている [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]．ビームフォーミングは，大きく分けて二種類に分類される．一つは，話者方位に高感度な指向特性を形成する遅延和（Delay-and-Sum; DS）アレーによるビームフォーミングである [1]．もう一方は，雑音の音源方位に感度の死角を形成する適応型アレーによるビームフォーミングである．さらに適応型アレーにはその拘束条件によって様々な手法が提案されている [2, 3, 4, 5, 6, 7, 8, 9, 10]．

Griffith-Jim 型アレーは，話者の方位にメインローブ（高感度の指向特性）を形成し，それ以外のサイドローブ（メインローブを除く高感度な指向特性）を打ち消す手法である [3]．最小分散ビームフォーミング（Minimum Variance Beamforming; MVB）型アレーは，話者の方位を除く雑音の到来方位に感度の死角を形成する手法である [6]．微分ビームフォーミング（Differential Beamforming; DB）型アレーは，指向特性の微分を行うことで死角を形成する手法である [7]．固有値ビームフォーミング（Eigen Beamforming; EB）型アレーは，信号の到来方位ベクトル（信号部分空間）と雑音部分空間の直交性を利用する手法である [8]．AMNOR 方式は，話者の音声の劣化を許容することで，MVB 型アレー以上に雑音抑圧が可能な死角形成手法である [4, 5]．S-AMNOR 方式は，音声の劣化を音声の特徴量を考慮して緩和した

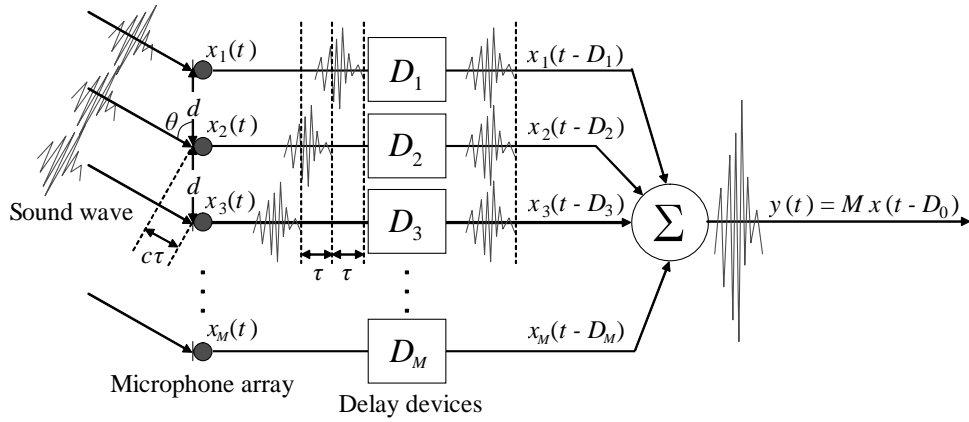


図 2.3  $M$  素子マイクロホンアレーを用いた遅延和アレー

AMNOR 方式である [9, 10] .

本節では，遅延和アレーと適応型アレーの原理について述べ，特に本論文において提案手法の基礎となる適応型アレーである AMNOR，S-AMNOR の詳細について述べる .

### 2.3.1 遅延和アレーの原理

遅延和アレーは，各マイクロホンにおける受信信号に対して話者方位の TDOA を同相化する遅延を付加した後，総和をとることにより，話者方位を強調するビームフォーミングを行う [1, 37] . 図 2.3 に，角度  $\theta$  から音速  $c$  で到来する音響信号  $x(t)$  を強調するマイクロホン間隔  $d$  で配置された  $M$  素子マイクロホンアレーを用いた遅延和アレーを示す .

図 2.3 に示した角度  $\theta$  から音速  $c$  で到来する音響信号  $x(t)$  は DOA に対応した TDOA を持つ信号として次式のように各マイクロホンで受信される .

$$x_i(t) = x(t - (i - 1)\tau), \quad (2.8)$$

$$\tau = \frac{d \cos \theta}{c}, \quad (2.9)$$

ここで， $i$  は  $i = 1, 2, \dots, M$  であり， $c$  は音速である .

式 (2.8) の TDOA を同相化する遅延を各受信信号に付与した後に総和をとることで、音響信号の DOA を強調するビームフォーミングを行う。同相化するための遅延と総和は次式で実現できる。

$$y(t) = \sum_{i=1}^M x_i(t - D_i), \quad (2.10)$$

$$D_i = D_0 - (i - 1)\tau, \quad (2.11)$$

ここで、 $D_0$  は遅延が非因果的になることを防ぐために付加する固定遅延量である。

式 (2.10),(2.11) より、音響信号の DOA に対応する TDOA は同相化されるため音響信号は  $Mx(t)$  に増幅されるが、それ以外の DOA では同相化されないため、ゲインの増幅効果が低い。結果として、音響信号の DOA のみが強調されるビームフォーミングとなる。

遅延和アレーは、マイクロホン数  $M$  とマイクロホン間隔  $d$  を大きく選び、アレー長  $Md$  を大きくすることで、音響信号の DOA に鋭い指向特性を形成できる。ただし、マイクロホン間隔  $d$  は音響信号の周波数を  $f$  とすると、 $d < c/2f$  を満たさない場合、指向特性に折り返し現象（空間エイリアシング）が発生する。また、遅延和アレーの指向特性は、周波数依存性があり、高い周波数ほど指向特性が鋭くなるが、特に素子数とアレー長が小規模なマイクロホンアレーでは低い周波数において鋭い指向特性を形成できないため [24]、高雑音環境下では十分な性能を発揮できない問題がある。

### 2.3.2 適応型アレーの原理

適応型アレー（Adaptive Array）は、各マイクロホンにおける受信信号に対して雑音の TDOA を同位相化する遅延を付加した後、減算を行うことにより、周囲から到来する雑音の DOA に対して死角を形成するビームフォーミングを行う [2, 3, 4, 5, 6, 7, 8, 9, 10]。図 2.4 に適応型アレーの中で最も単純な 2 素子マイクロホンアレーによる減算型アレーを示す。

図 2.4 において、雑音は未知の方向  $\theta$  から到来し、2 つのマイクロホン Mic.1, 2 で受信される。このとき、Mic.1 の受信信号  $x_1(t)$  は Mic.2 の受信信号  $x_2(t)$  より  $\tau$  だ



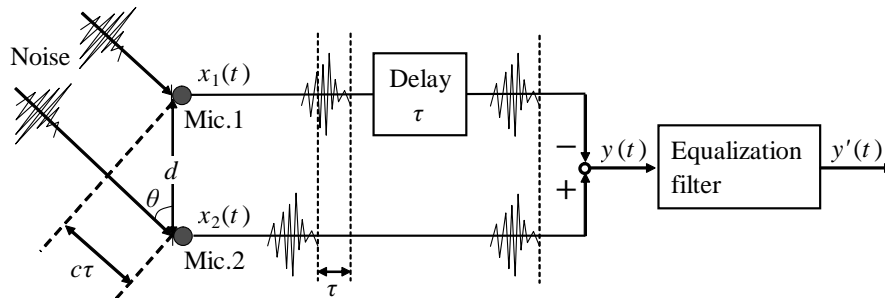


図 2.4 2 素子マイクロホンアレーによる減算型アレー

け遅れた信号になっている．この TDOA と  $x_1(t), x_2(t)$  の関係は次式のように表される．

$$x_2(t) = x_1(t - \tau), \quad (2.12)$$

$$\tau = \frac{d \cos \theta}{c}, \quad (2.13)$$

ここで， $c$  は音速である．

すなわち，受信信号  $x_1(t)$  に  $\tau$  の遅延を付加すればこれらの受信信号を同相化され，これらを次式のように減算すれば雑音は消去される．

$$\begin{aligned} y(t) &= x_2(t) - x_1(t - \tau) \\ &= 0. \end{aligned} \quad (2.14)$$

以上の操作では，雑音の TDOA，すなわち，雑音の DOA が既知である必要はない．これは，雑音のみが到来している状況で，遅延量を変化させ，アレー出力のパワーを最小化することで自動的に遅延量が  $\tau$  となり，アレー出力のパワーが 0 となることから雑音が消去され，結果として雑音の DOA に感度の死角を形成するビームフォーミングが行われるからである [2]．

一方，雑音とは異なった方向から到来してきた目的信号は，TDOA が異なるため，減算によって消去されることはないが，歪みが生じてしまう問題がある．そこで，図 2.4 のように，減算の後段に補正フィルタを接続して周波数特性を補正する手法が提案されている [3, 4, 5, 6, 9, 10, 26] ．

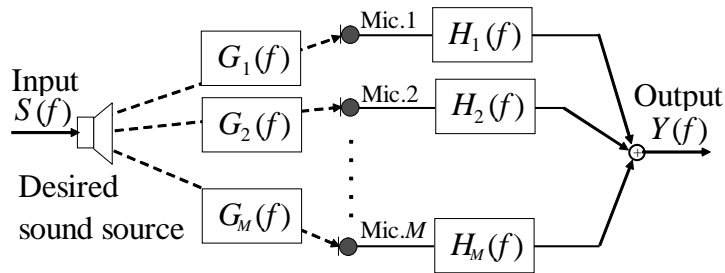


図 2.5 アレーの目的信号に対する流れ図

この補正フィルタと遅延量を一つのフィルタとみなして，適応型アレーのフィルタ特性と考えることができる．周波数  $f$ ，適応型アレーのフィルタ特性  $H_i(f)$ ，目的音源から各マイクロホンまでの伝達特性  $G_i(f)$  を用いて目的信号とアレー出力の関係を図 2.5 のように表すことができる．

図 2.5 の目的音源から各マイクロホンまでの伝達特性  $G_i(f)$ ，適応型アレーのフィルタ特性  $H_i(f)$  を用いて，このアレーの目的信号に対する周波数特性  $J(f)$  は，

$$J(f) = \sum_{i=1}^M G_i(f)H_i(f), \quad (2.15)$$

と表すことができる．

式 (2.15) において，単純に雑音を最小化するとすべての適応型アレーのフィルタ係数が 0 になるため，拘束条件付きの雑音最小化を行う必要がある．雑音パワーの最小化に対する代表的な拘束条件として， $J(f)$  に関する次式の 2 つの拘束条件がある [24]．

$$\text{I} \quad J(f) = 1, \quad (2.16)$$

$$\text{II} \quad E = \int |1 - J(f)|^2 df \leq \hat{E}, \quad (2.17)$$

ここで， $E$  は目的音の劣化量であり， $\hat{E}$  はあらかじめ決められた許容劣化量である．

I の拘束条件では，目的信号に対する周波数特性  $J(f)$  を平坦に保つ拘束条件であるため，目的信号に対する歪みは発生しない．Griffith-Jim 型アレー [3] や MVB 型アレー [6] はこの拘束条件 I を用いた適応型アレーである．

Ⅱの拘束条件は，目的信号に対する劣化を許容する拘束条件である．Ⅱの拘束条件では目的信号に対して歪みが発生するが，Ⅰの拘束条件以上に雑音を抑圧可能であるため，特に低 SNR 環境において高い性能を発揮する．AMNOR 方式や S-AMNOR 方式は，この拘束条件Ⅱを用いた適応型アレーである [4, 5]．

適応型アレーの指向特性において，死角を形成できる DOA の数は  $M$  素子マイクロホンアレーの場合  $M - 1$  個である [24]．そのため，雑音の到来方向が形成できる死角の数よりも少ない場合には，遅延和アレーと比べて大きな雑音低減効果が得られるが，高残響下ではその性能が著しく低下してしまう問題がある．

### 2.3.3 AMNOR (Adaptive Microphone-array for NOise Reduction)

AMNOR は金田らによって提案された適応型アレーの一つで，式 (2.17) の雑音パワーを最小化する拘束条件を持つ適応型アレーである [4, 5]．AMNOR は目的音の劣化量  $E$  と出力雑音パワーとのトレードオフ関係を利用する．すなわち，目的音の劣化許容量  $\hat{E}$  を大きくすれば，出力雑音パワーを小さくすることができ，逆に  $\hat{E}$  を小さくすれば出力雑音パワーは大きくなる．この結果，多少目的音の周波数特性は劣化するが，雑音を大幅に抑圧することができる． $M$  素子マイクロホンアレーにおける AMNOR 方式を図 2.6 に示す．

図 2.6 の VF1, VF2, AF は  $M$  入力 1 出力の FIR フィルタである．AF は適応フィルタであり，VF1, VF2 は AF と同じフィルタ係数を持つ可変フィルタである．AMNOR はフィルタ設計の際，擬似目的信号  $s'(k)$  が必要になる．これは，AMNOR が環境の周辺雑音と擬似目的信号を用いることで，雑音抑圧の最大化を行うためである．擬似目的信号  $s'(k)$  は，増幅器のゲイン  $A$  と時間遅れ  $\tau_i$  (以後， $i$  は  $i = 1, \dots, M$ ) から， $As'_i(k - \tau_i)$  で表す．ただし，時間遅れ  $\tau_i$  は，目的信号の到来方向を既知として計算する．次に，マイクロホンアレーを用いて目的音が含まれていない雑音のみの信号  $n_i(k)$  を受信する．この雑音のみの信号  $n_i(k)$  と擬似目的信号  $As'_i(k - \tau_i)$  を加算した信号が適応フィルタ AF の入力になる．適応フィルタ AF は式 (2.18) の  $e(k)$

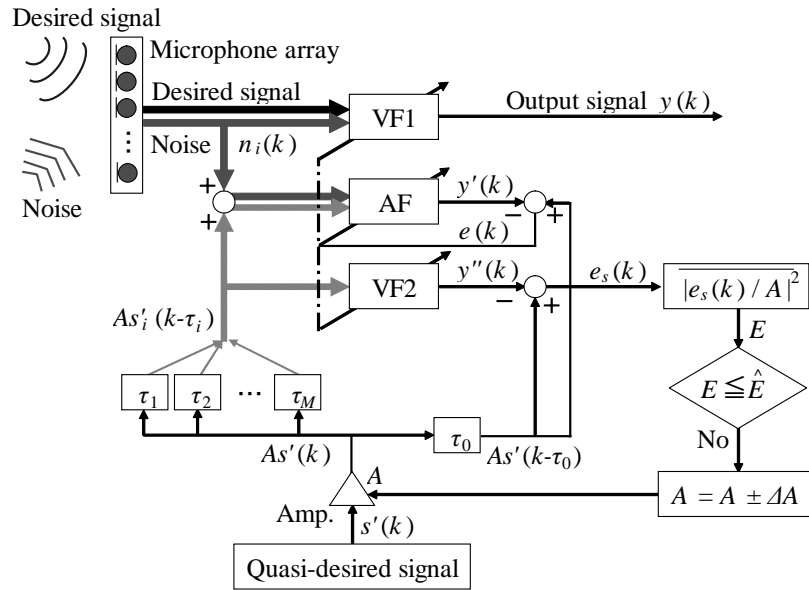


図 2.6  $M$  素子マイクロホンアレーにおける AMNOR 方式

により，フィルタ係数を制御する．

$$e(k) = As'(k - \tau_0) - y'(k), \quad (2.18)$$

ここで， $\tau_0$  は因果律を満たすための固定遅延量である．

$e_s(k)$  は適応フィルタ AF でフィルタ係数が設計した後，FIR フィルタ VF2 の出力  $y''(k)$  を用いて次式のように計算する．

$$e_s(k) = As'(k - \tau_0) - y''(k). \quad (2.19)$$

現在の劣化量  $E$  は式 (2.20) より得られる．

$$E = \overline{\left| \frac{e_s(k)}{A} \right|^2}, \quad (2.20)$$

ここで， $\overline{|\cdot|^2}$  は，二乗平均の操作を表す．

現在の劣化量  $E$  と劣化許容量  $\hat{E}$  を比較することで，増幅器のゲイン  $A$  は， $E \leq \hat{E}$  を満たすまでに更新される．上記のようなアルゴリズムにより，AMNOR は実環境において，高い雑音抑圧性能を達成する．

### 2.3.4 S-AMNOR( average Speech spectrum-based AMNOR )

S-AMNOR 方式は西浦らにより提案された音声受音用の AMNOR 方式である [9, 10]．従来の AMNOR では，擬似目的信号にすべての周波数成分のパワーが均一に含まれている白色雑音を使用する．しかし，音声受音用途に利用を限定する場合，音声の特徴を考慮することでより適切なビームフォーミングが可能である．そこで S-AMNOR は，AMNOR のアルゴリズムで用いる擬似目的信号のスペクトル特性に音声の長時間平均スペクトルを重み付けすることで，音声に対する歪みを軽減するビームフォーミングを行う．

まず，音声の平均スペクトル重みは式 (2.21) のように計算する．

$$W_{sp}(\omega) = \sum_{l=1}^L \frac{1}{N_l} \sum_{n=1}^{N_l} SP_l(\omega; n), \quad (2.21)$$

ここで， $L$  は音声（単語）の数， $N_l$  は  $l$  番目の音声のフレーム数， $SP_l(\omega; n)$  は音声信号のフーリエ変換， $W_{sp}(\omega)$  は音声の平均スペクトルを表す．音声の平均スペクトルに基づく擬似目的信号は，白色雑音に音声の平均スペクトル  $W_{sp}(\omega)$  を重み付けすることで作成する．

従来の AMNOR で用いる擬似目的信号である白色雑音のスペクトルを図 2.7 に，S-AMNOR で用いる擬似目的信号である音声の平均スペクトルを重み付けした信号のスペクトルを図 2.8 に示す．

S-AMNOR は特に音声の重要な特徴成分である低次フォルマントに対する歪みを軽減可能である．しかし，音声の短時間スペクトルは音素によって特徴が大きく異なるため，音声すべての長時間平均スペクトルを考慮したビームフォーミングでは十分に歪みを軽減できない問題点がある．

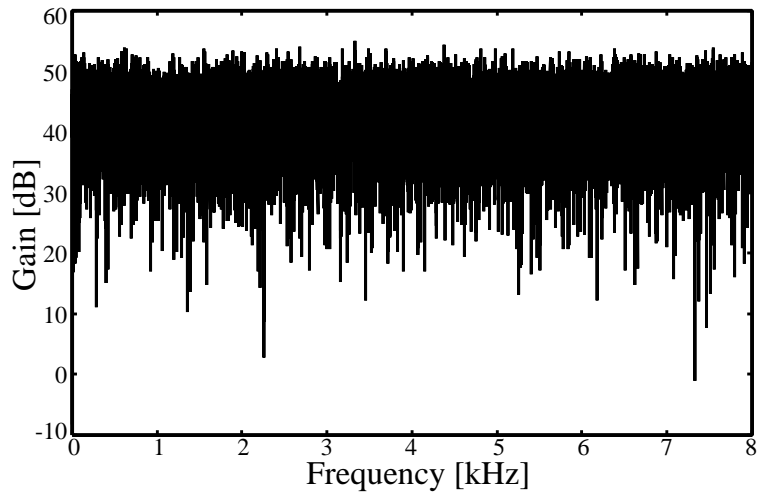


図 2.7 白色雑音のスペクトル

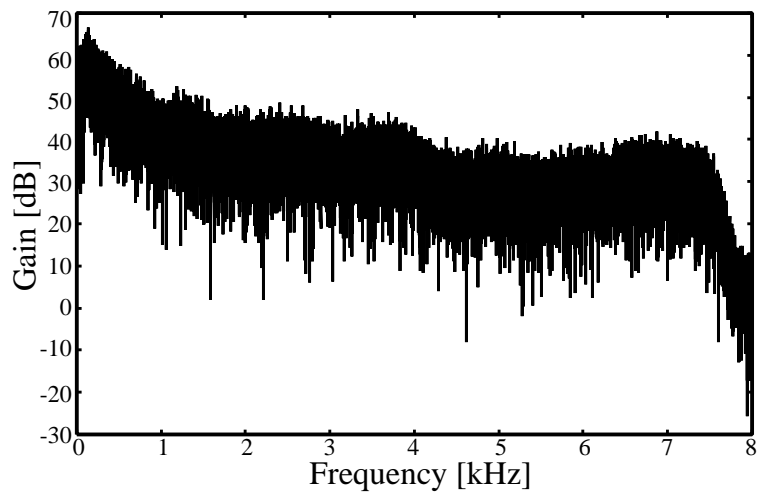


図 2.8 音声平均スペクトルを重み付けした信号のスペクトル

## 2.4. まとめ

本章では、マイクロホンアレーを用いた遠隔発話音声受音の要素技術である話者方位推定とビームフォーミングの原理と既存の手法について述べた。2.2節では、既存の話者方位推定で利用される音声の DOA 推定の原理と CSP 法の詳細について述べ、CSP 法は雑音・残響環境下、特に指向性雑音環境下において DOA 推定性能が劣化し、2素子マイクロホンアレーでは、DOA は推定できるが話者の位置（話者までの距離）は推定できないことを示した。2.3節では、遅延和アレーと適応型アレーの原理、特に AMNOR と S-AMNOR の詳細について述べ、低 SNR 環境において高い性能を発揮するが音声に歪みが生じることを示した。

# 第3章 可聴音の送信波と反射波の 位相干渉に基づく距離推定

## 3.1. はじめに

多くの工学分野において，対象物までの距離は基本的かつ非常に重要な情報である．一般的な距離推定法では，送信波と対象物からの反射波との時間差が利用されている [48, 49, 50, 51]．しかし，この手法では距離が短く送信波が十分減衰しない間に反射波が返ってくるような場合には，反射波が送信波に埋もれてしまい，対象物までの距離を推定することが困難になる [52]．

マイクロ波レーダの分野において，送信波と反射波の干渉を利用した距離推定法が知られている [53, 54, 55]．これを可聴音域に拡張した距離推定法（以降，単に音響測距法もしくは提案手法と表記する）について検討されている [27, 28, 29, 30, 31, 32, 33, 34, 35, 36]．この手法は，反射音の計測を行うという観点から言えば遠隔マイクロホンを利用する手法であるため，雑音の影響によって測距性能が低下することが予想される．しかし，これまでは，実環境における評価実験は行っていたが，雑音環境下における詳細な評価実験や雑音対策などは行われていなかった．

そこで，本論文では提案手法の加法性雑音（以後，単に雑音と表記する）が存在する環境における動作について議論する．そのため，まず位相干渉に基づく音響測距法を雑音環境下における実環境の測定系を考慮した理論に拡張する．また，雑音を抑圧するためにこれまで多くの手法が提案されている [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 56, 57, 58, 59, 60, 61, 62, 63, 64]．雑音抑圧では，雑音の時間的な情報，もしくは空間的な情報を利用して観測信号に含まれる雑音を推定することが重要である．提案手法では，マイクロホンが1つであるため，複数のマイクロホンにより空間的な情報を取得して雑音を抑圧するマイクロホンアレー技術 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] は



適用できない。また，参照マイクロホンを用いて雑音を推定・減算するアダプティブノイズキャンセラ ( Adaptive Noise Canceller; ANC ) [56] も利用できない。

単一マイクロホンを利用した雑音下音声受信においても様々な雑音抑圧手法が提案されている [57, 58, 59, 60, 61]。ウィナーフィルタやカルマンフィルタによるフィルタリングは雑音をモデル化することで雑音を抑圧する手法である [57, 58, 59, 60]。スペクトルサブトラクション ( Spectral Subtraction; SS ) は，雑音の平均スペクトルを推定して観測信号のスペクトルから減算することで定常雑音を抑圧する手法である [61]。特に，SS は非常に簡易かつ効果的に雑音抑圧が可能であるため，様々な受信システムの前処理として利用されている [67, 68, 69]。

一方，単一マイクロホンを利用した雑音下におけるインパルス応答計測においても様々な雑音抑圧手法が提案されている [62, 63, 64]。同期加算法 ( Synchronous Addition; SA ) は，同一の計測信号を複数回計測・加算平均することで突発性雑音や非周期性雑音などを抑圧する手法である [24, 64, 65, 66]。

本論文では，一般的によく用いられる手法である，SA と SS の考え方を位相干渉に基づく音響測距法に導入することで，この音響測距法に適した雑音抑圧手法を検討する。

本章の構成を以下に示す。3.2 節では，実環境における可聴音の直接波と反射波の位相干渉に基づく距離推定の原理の詳細について述べる。3.3 節では，提案手法である雑音環境に頑健に動作する可聴音の直接波と反射波の位相干渉に基づく音響測距法について述べる。3.4 節では，実環境において行った雑音環境を考慮した評価実験とその結果について述べる。最後に，3.5 節で本章のまとめを述べる。

## 3.2. 実環境における可聴音の送信波と反射波の位相干渉に基づく距離推定の原理

$t$  を時刻 [s]， $x$  を信号源と対象物を結ぶ軸上の位置 [m] ( 原点は信号源 ) として，信号源より対象物へ向けて放射される送信波  $v_T(t, x)$  を次式とする。

$$v_T(t, x) = \sum_{i=1}^N A_i e^{j(2\pi f_i t - \frac{2\pi f_i x}{c} + \theta_i)}, \quad (3.1)$$

ここで、 $f_i$  は第  $i$  周波数 [Hz] ( $f_1, f_N$  はそれぞれ送信波の最低と最高の周波数 [Hz]) であり、 $A_i, \theta_i$  はそれぞれ第  $i$  周波数成分の振幅および初期位相、 $c$  は音速 [m/s] である。

この送信波が  $n$  番目の対象物によって反射された反射波  $v_{R_n}(t, x)$  は次式で表される。

$$v_{R_n}(t, x) = \sum_{i=1}^N A_i \gamma_{i,n} e^{j\phi_{i,n}} e^{j(2\pi f_i t - \frac{2\pi f_i}{c}(2d_n - x) + \theta_i)}, \quad (3.2)$$

ここで、 $d_n$  は  $n$  番目の対象物までの距離であり、 $\gamma_{i,n} e^{j\phi_{i,n}}$  は  $n$  番目の対象物の反射係数である。

対象物が  $m$  個の場合の実環境において観測される合成波  $v_C(t, x)$  は次式となる。

$$v_C(t, x) = b(t) * (v_T(t, x) + \sum_{n=1}^m v_{R_n}(t, x)) + w(t), \quad (3.3)$$

ここで、 $*$  は畳み込み演算子、 $b(t)$  は実環境の計測系のインパルス応答 (乗法性雑音) であり、 $w(t)$  は環境雑音 (加法性雑音) である。

マイクロホン位置  $x = 0$  で観測された合成波  $v_C(t, 0)$  を次式のようにフーリエ変換する。

$$V_C(f, 0) = \int_{-\infty}^{\infty} v_C(t, 0) e^{-j2\pi ft} dt, \quad (3.4)$$

ここで、 $f_i$  は変数  $f$  に置き換えている。

$v_C(t, 0)$  のパワースペクトル  $p(f, 0)$  は次式となる。

$$p(f, 0) = |V_C(f, 0)|^2. \quad (3.5)$$

反射係数  $\gamma_{i,n} e^{j\phi_{i,n}}$  が周波数に依らず一定 ( $\gamma_n e^{j\phi_n}$ ) であり、反射係数の大きさ  $\gamma_n$  が十分小さい ( $\gamma_1, \dots, \gamma_n \ll 1$ ) と仮定すると、式 (3.1) ~ (3.5) より、 $p(f, 0)$  は次式で近似できる。

$$p(f, 0) \approx |A(f)|^2 |B(f)|^2 \{1 + C(f)\} + |W(f)|^2, \quad (3.6)$$

$$C(f) = 2 \sum_{n=1}^m \gamma_n \cos\left(\frac{4\pi f d_n}{c} - \phi_n\right),$$

ここで、 $A(f)$  は送信波のスペクトル、 $B(f)$  は計測系の伝達関数、 $W(f)$  は環境雑音のスペクトル、 $|A(f)|^2|B(f)|^2$  は計測系の伝達関数によって歪んだ送信波のパワーであり、 $C(f)$  は定在波によるパワースペクトルの揺らぎ成分（位相干渉項）である。

式 (3.6) において、 $|A(f)|^2|B(f)|^2$  の項を減算することで、位相干渉項  $C(f)$  を抽出可能である。式 (3.3), (3.6) より、 $|A(f)|^2|B(f)|^2$  の項は、対象物の存在しない環境における計測によって次式のように得られる。

$$\hat{v}_C(t, x) = b(t) * v_T(t, x) + \hat{w}(t), \quad (3.7)$$

$$\hat{p}(f, 0) \approx |A(f)|^2|B(f)|^2 + |\hat{W}(f)|^2, \quad (3.8)$$

ここで、 $\hat{w}(t)$  ( $\neq w(t)$ ) は環境雑音、 $\hat{W}(f)$  は  $\hat{w}(t)$  のスペクトルであり、 $v_T$  は式 (3.1) と同様の送信波である。

式 (3.6), (3.8) より、 $\Delta p(f, 0)$  は  $p(f, 0)$  から  $\hat{p}(f, 0)$  を減算することで次式のように得られる。

$$\Delta p(f, 0) \approx |A(f)|^2|B(f)|^2 C(f) + |W(f)|^2 - |\hat{W}(f)|^2. \quad (3.9)$$

この処理をバックグラウンド要素消去処理 (Back-Ground Components Cancellation Processing; BGCCP) という。もう一つの BGCCP として白色化に基づく BGCCP [36] も提案されているが、白色化に基づく BGCCP は白色化フィルタの設計が雑音環境下では困難という問題がある。

$SD[\cdot]$  を標準偏差演算子として、次式のように雑音の振幅が干渉成分に比べて十分に小さいと仮定すると

$$SD[|W(f)|^2 - |\hat{W}(f)|^2] \ll SD[|A(f)|^2|B(f)|^2 C(f)], \quad (3.10)$$

式 (3.9) より、 $\Delta p(f, 0)$  は周波数  $f$  に関して周期関数となり、その周期はマイクロホン-対象物間の距離  $d_n$  に反比例する。したがって、 $\Delta p(f, 0)$  を  $f$  の関数としてフーリエ変換することで、距離を推定できる。すなわち、フーリエ変換の公式：

$$F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} dt, \quad (3.11)$$

において、 $f$  を  $2x/c$ 、 $t$  を  $f$ 、 $f(t)$  を  $\Delta p(f, 0)$  と置き換えることで、 $P(x)$  は次式のように定義できる。

$$P(x) = \int_{f_1}^{f_N} \Delta p(f, 0) e^{-j2\pi \frac{2x}{c} f} df. \quad (3.12)$$

この  $P(x)$  の振幅スペクトル  $|P(x)|$  を距離スペクトルと呼び、距離スペクトルのピーク位置がマイクロホン-対象物間の距離  $d_n$  の推定値となる。ただし、 $x_{AB}$  を  $|\text{FT}[|A(f)|^2|B(f)|^2]|$  のピーク位置 (FT[·]: フーリエ変換)、 $x_C$  を  $|\text{FT}[C(f)]|$  のピーク位置とすると、式 (3.9) より、 $2x_{AB} < x_C$  が減算の BGCCP に基づく音響測距法の十分条件となる。式 (3.9) より、 $2x_{AB} < x_C$  が満たされた場合、 $C(f)$  が振幅変調された信号となる。そのため、減算の BGCCP に基づく音響測距法は、出来る限り計測系が校正されている必要があり、基本周波数を含む信号を送信波として利用することが難しい。さらに、環境雑音の振幅が大きい ( $\text{SD}[|A(f)|^2|B(f)|^2C(f)] < \text{SD}[|W(f)|^2 - |\hat{W}(f)|^2]$ ) とき、測距性能が劣化する。

ここで、図 3.1 に、雑音が存在せず、 $|A(f)|^2|B(f)|^2$  が均一であり、距離  $d$  の位置に対象物が一つ存在する場合に観測される合成波より得られた  $\Delta p(f, 0)$  を、図 3.2 にそのときの距離スペクトル  $|P(x)|$  を示す。

音響測距法の距離推定精度は音速  $c$  と  $\Delta p(f, 0)$  の周波数分解能に依存する。測距可能な最小距離  $d_{\min}$  (距離分解能に相当) は図 3.1 の帯域幅  $f_w$  によって決まり、式 (3.9) の周期は帯域幅よりも小さくならないことから、

$$d_{\min} > \frac{c}{2f_w}, \quad (3.13)$$

によって求められる。

一方、測距可能な最大距離  $d_{\max}$  は周波数間隔  $\Delta f = f_{i+1} - f_i$  によって決まり、標本化定理により、 $\Delta p(f, 0)$  の周期は  $2\Delta f$  よりも大きくなければならないことから、

$$d_{\max} < \frac{c}{4\Delta f}, \quad (3.14)$$

によって求められる。

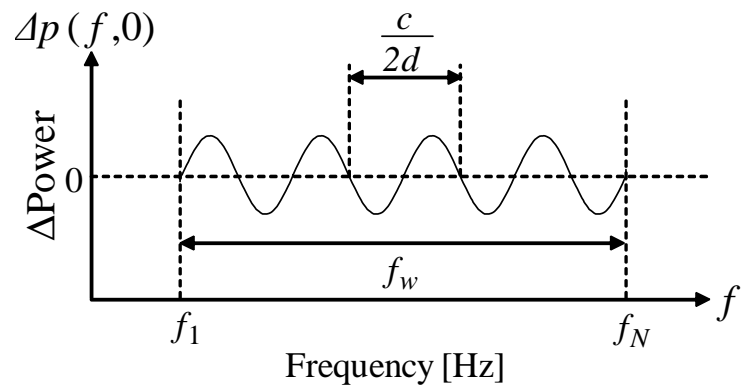


図 3.1 対象物が一つの場合の  $\Delta p(f,0)$

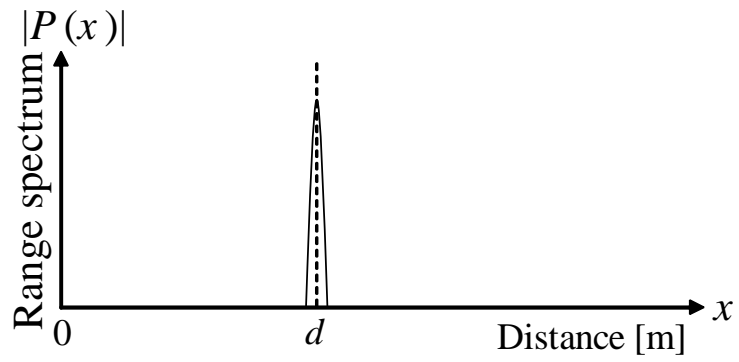


図 3.2 対象物が一つの場合の距離スペクトル

### 3.3. 雑音環境に頑健な可聴音の送信波と反射波の位相干渉に基づく測距法

#### 3.3.1 従来の前処理による雑音抑圧手法

本節では、従来の前処理による雑音抑圧手法について説明する。同期加算 (Synchronous Addition; SA) [64] とスペクトルサブトラクション (Spectral Subtraction; SS) [61] は前処理における雑音抑圧技術の従来手法である。SA は、インパルス応答の計測時に発生する加法性の突発性・非周期性雑音に対して効果的な雑音抑圧手法である。SA は次式のように定義される。

$$\hat{x}(t) = \overline{x(t)} + \frac{1}{L} \sum_{i=1}^L w_i(t), \quad (3.15)$$

ここで、 $\hat{x}(t)$  は強調されたインパルス応答、 $\overline{x(t)}$  は平均化されたインパルス応答、 $w_i(t)$  は環境雑音であり、 $L$  は同期加算回数である。式 (3.15) において、 $\overline{x(t)}$  はインパルス応答  $x(t)$  と等しくなり、 $\frac{1}{L} \sum_{i=1}^L w_i(t)$  は雑音がエルゴード性を有している場合は長時間平均  $\overline{w(t)}$  と等しくなる。 $w_i(t)$  が非周期信号であるならば、SA によって SNR (Signal-to-Noise Ratio) を改善できる。

SS は加法性の定常雑音に対して効果的な従来手法である。SS は雑音の長時間平均スペクトルを観測信号のスペクトルより減算することで、定常な雑音を抑圧することができる。SS は次式のように定義される。

$$|\hat{X}(f)| = |Y(f)| - \alpha \overline{|W(f)|}, \quad (3.16)$$

ここで、 $|\hat{X}(f)|$  は雑音を抑圧された観測信号のスペクトル、 $|Y(f)|$  は観測信号のスペクトル、 $\overline{|W(f)|}$  は雑音の長時間平均スペクトルであり、 $\alpha$  はサブトラクション係数である。

#### 3.3.2 位相干渉による音響測距法に適した雑音抑圧手法

本節では、位相干渉による音響測距法に適した雑音抑圧手法について説明する。本論文では、前節で説明した SA と SS の考え方を位相干渉による音響測距法に導入

することで雑音を抑圧する手法を検討する．まず，同一の送信波を複数回測定することで，SAを行う．式(3.3),(3.7),(3.15)より， $v_C(t, x)$ と $\hat{v}_C(t, x)$ は次式のように平均化される．

$$\overline{v_C(t, x)} = \overline{b(t) * (v_T(t, x) + \sum_{n=1}^m v_{Rn}(t, x)) + w_i(t)}, \quad (3.17)$$

$$\overline{\hat{v}_C(t, x)} = \overline{b(t) * v_T(t, x) + \hat{w}_i(t)}. \quad (3.18)$$

SAの条件下では， $\overline{b(t) * (v_T(t, x) + \sum_{n=1}^m v_{Rn}(t, x))}$ と $b(t) * (v_T(t, x) + \sum_{n=1}^m v_{Rn}(t, x))$ ， $\overline{b(t) * v_T(t, x)}$ と $b(t) * v_T(t, x)$ は等しくなるため，式(3.6),(3.8),(3.17),(3.18)より， $p(f, 0)$ と $\hat{p}(f, 0)$ は次式のように近似される．

$$\overline{p(f, 0)} \approx |A(f)|^2 |B(f)|^2 \{1 + C(f)\} + \overline{|W(f)|^2}, \quad (3.19)$$

$$\overline{\hat{p}(f, 0)} \approx |A(f)|^2 |B(f)|^2 + \overline{|\hat{W}(f)|^2}, \quad (3.20)$$

もし， $\overline{|W(f)|}$ が $\overline{|\hat{W}(f)|}$ と等しくない場合は，式(3.19),(3.20)より， $\Delta p'(f, 0)$ は $\overline{p(f, 0)}$ から $\overline{\hat{p}(f, 0)}$ を減算することで次式のように得られる．

$$\Delta p'(f, 0) \approx |A(f)|^2 |B(f)|^2 C(f) + \overline{|W(f)|^2} - \overline{|\hat{W}(f)|^2}. \quad (3.21)$$

もし， $\overline{|W(f)|}$ が $\overline{|\hat{W}(f)|}$ と等しい場合は，式(3.19),(3.20)より， $\Delta p'(f, 0)$ は $\overline{p(f, 0)}$ から $\overline{\hat{p}(f, 0)}$ を減算することで次式のように得られる．

$$\Delta p'(f, 0) \approx |A(f)|^2 |B(f)|^2 C(f). \quad (3.22)$$

式(3.21),(3.22)は突発性・非周期性雑音であればSAによって雑音が抑圧され，定常雑音であればSSによって雑音が抑圧されることを表している．この雑音抑圧された $\Delta p'(f, 0)$ を式(3.12)の $\Delta p(f, 0)$ に代入することで，距離スペクトルが算出できる．

### 3.4. 実環境における評価実験

提案手法の雑音環境における頑健性を確認するため，様々なSNRの雑音環境で提案手法による測距実験を行った．

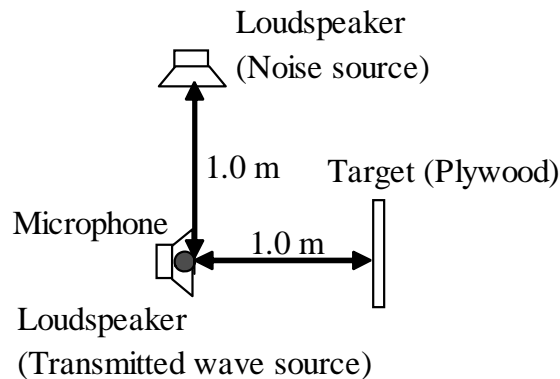
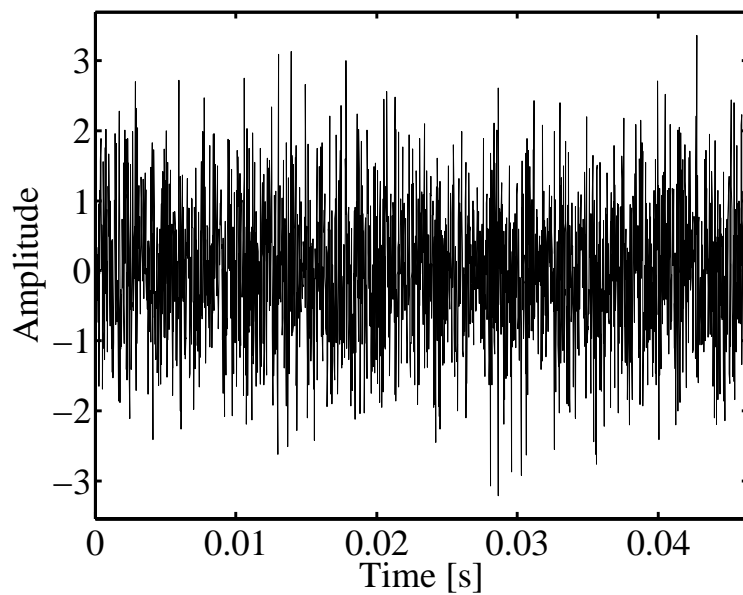


図 3.3 雑音環境下の音響測距実験における実験環境

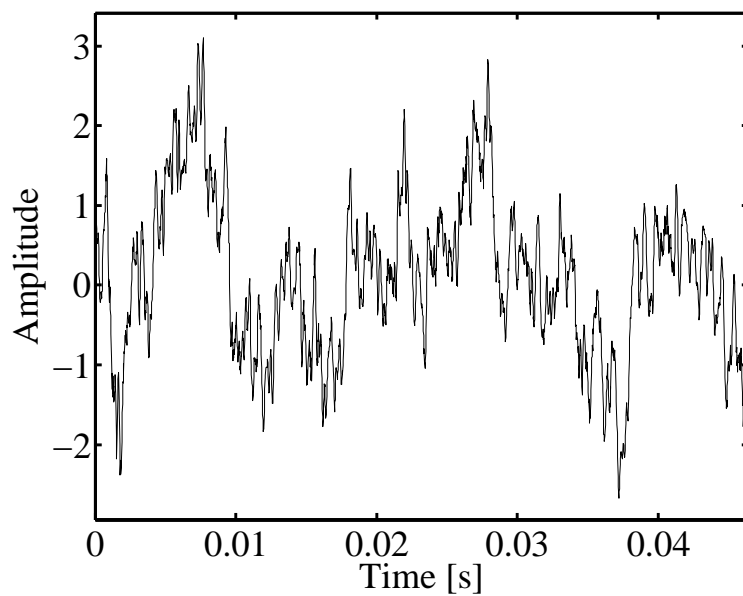
### 3.4.1 実験条件

図 3.3 に対象物，スピーカとマイクロホンの関係を，表 3.1 に実験で使用した機材を示す．マイクロホンはスピーカの上に設置（スピーカ・マイクロホン間距離: 0.06 m）し，対象物はベニヤ板とした．本実験では，送信波として白色雑音，雑音として送信波とは異なる白色雑音と PC のファンノイズ（有色性雑音）を利用した．図 3.4, 3.5 は，それぞれ白色雑音とファンノイズの波形とスペクトルである．雑音は，図 3.3 のように，マイクロホンから 1.0 m 離れた位置に設置したスピーカ（BOSE, 101MM）より放射した．室内の残響時間は  $T_{[60]} = 0.6$  s であり，背景騒音レベルは 32 dBA であった．実験では，TSP 法 [63] で計測したインパルス応答を利用した．サンプリング周波数は 44.1 kHz，計測データ長は 2,048 点，音速は 340 m/s とした．測距に利用した周波数帯域は十分に反射波の発生する帯域である 2.1 kHz ( $= f_1$ ) から 7.6 kHz ( $= f_N$ ) までの帯域幅 5.5 kHz とした．この帯域幅における距離解像度は式 (3.13) より約 0.03 m となる．合成波（観測信号）は白色雑音とインパルス応答を畳み込むことによって作成し， $-5$  dB から 25 dB まで 1 dB 刻みと， $\infty$  dB の SNR で雑音である白色雑音もしくはファンノイズを観測信号である合成波に加算した．SNR =  $\infty$  dB はクリーン環境を意味する．本実験条件では，定常性雑音である白色信号と，定常性雑音ではあるが突発性雑音を含むファンノイズを雑音とした場合の耐雑音性評価となる．



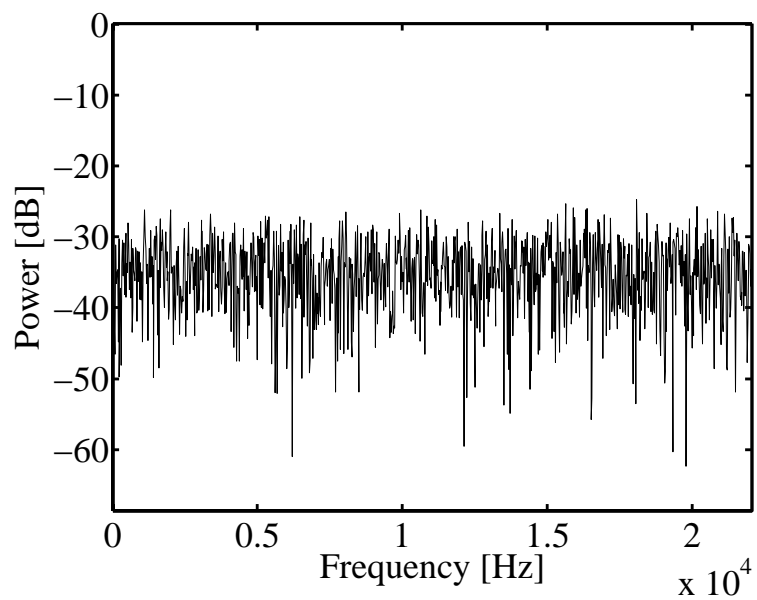


(a) 白色雑音

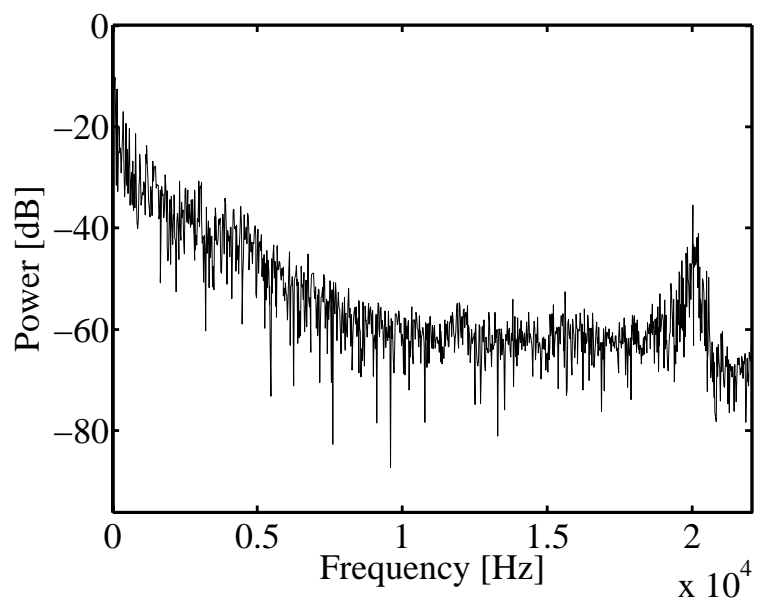


(b) ファンノイズ

図 3.4 雑音環境下音響測距実験における雑音信号



(a) 白色雑音



(b) ファンノイズ

図 3.5 雑音環境下音響測距実験における雑音信号のパワースペクトル

表 3.1 雑音環境下の音響測距実験で使用した機材

Microphone	AUDIO-TECHNICA, AT-805F
Loudspeaker	FOSTEX, FE87E
Speaker box	DAITO-VOICE, SV-70
Power amplifier	YAMAHA, A100a
Microphone amplifier	AUDIO-TECHNICA, AT-MA2
Audio interface	M-AUDIO, MOBILE PRE USB
Target	Plywood square (H 0.3 m × W 0.225 m × D 0.05 m)

本実験条件では，スピーカとマイクロホンが非常に近接して設置されているため，送信波は雑音と比べて非常に大きくなる．そのため，本実験における低 SNR の条件では非常に騒がしい環境となる．

本実験において，測距可能であった最低 SNR，雑音抑圧量( Noise Reduction Ratio; NRR )，距離スペクトルにおける SNR ( Range Spectrum Ratio; RSR )，パワースペクトルの揺らぎにおける SNR ( Power Spectral Fluctuation Ratio; PSFR ) を評価した．

NRR は次式で定義する．

$$\text{NRR} = 10 \log_{10} \frac{\sigma_i^2}{\sigma_o^2}, \quad (3.23)$$

ここで， $\sigma_i^2$  は入力雑音の分散であり， $\sigma_o^2$  は雑音抑圧された出力雑音の分散である．NRR は時間領域における雑音抑圧性能を示す．

RSR は次式で定義する．

$$\text{RSR} = 10 \log_{10} \frac{\sum_{i=0}^H |P_c(x_i)|^2}{\sum_{i=0}^H |P_n(x_i)|^2}, \quad (3.24)$$

ここで， $H$  は距離スペクトルのデータ数， $|P_c(x_i)|$  はクリーン環境( 雑音なし )における距離スペクトルであり， $|P_n(x_i)|$  は各 SNR の雑音環境における距離スペクトルである．RSR は距離スペクトルのピーク検出の難しさを示す．

PSFR は次式で定義する .

$$\text{PSFR} = 10 \log_{10} \frac{\sum_{i=1}^G |\Delta p_c(f_i, 0)|^2}{\sum_{i=1}^G |\Delta p_n(f_i, 0)|^2}, \quad (3.25)$$

ここで ,  $G$  は分析スペクトル帯域のデータ数 ,  $\Delta p_c(f_i, 0)$  はクリーン環境における BGCCP 後のパワースペクトルであり ,  $\Delta p_n(f_i, 0)$  は各 SNR の雑音環境における BGCCP 後のパワースペクトルである . PSFR はパワースペクトルの揺らぎにおける位相干渉成分と雑音の比を示す .

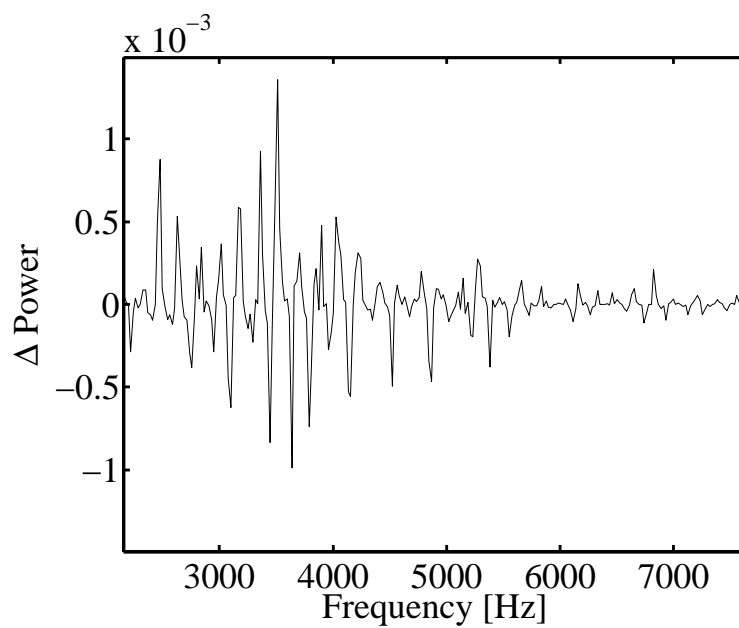
### 3.4.2 実験結果

図 3.6 にクリーン環境と SNR = 0 dB の白色雑音が加算されたときの  $\Delta p(f, 0)$  を , 図 3.7 に同条件における距離スペクトルをそれぞれ示す . 図 3.6, 3.7 より , 低 SNR 環境では雑音に距離に相当する干渉成分が埋もれてしまっていることがわかる .

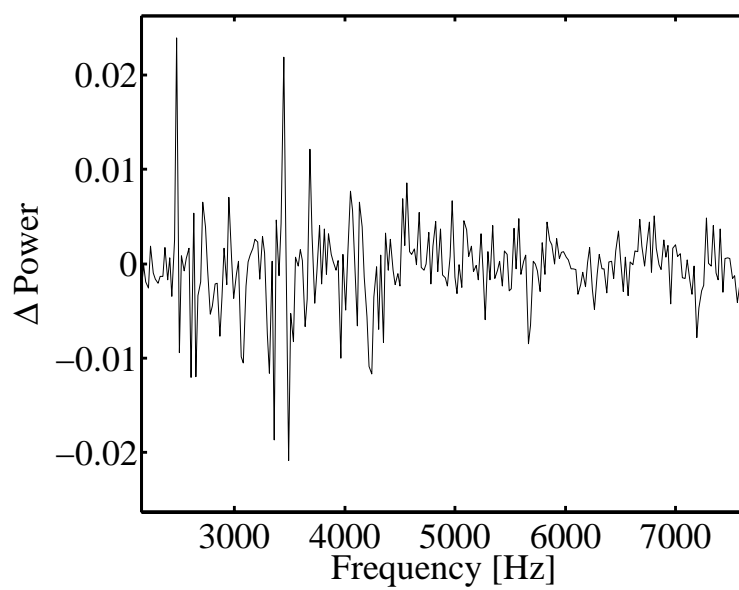
図 3.8, 3.9, 3.10, 3.11 に雑音を白色雑音とした場合の測距可能であった最低 SNR , NRR , RSR , PSFR をそれぞれ示す . 図 3.8, 3.9, 3.10, 3.11 より , 同期加算回数を増やすごとにそれぞれの性能が改善していることがわかる .

図 3.8 の測距性能において , 同期加算数 32 回の方が 16 回よりも性能が劣化しているのは , 同期加算数 16 回における減算後の雑音による距離スペクトルが対象物による距離スペクトルと重なっているためである . また , 図 3.10, 3.11 の RSR, PSFR の結果より , RSR と PSFR の間には非常に高い相関関係があり , RSR, PSFR を -10 dB 以上に改善することで測距が可能となる傾向があった .

図 3.12 に SNR = 0 dB のファンノイズが加算されたときに同期加算を行わない従来手法 ( Number of synchronous addition: L=1 ) と 32 回の同期加算を行った提案手法 ( Number of synchronous addition: L=32 ) の  $\Delta p(f, 0)$  を , 図 3.13 に同条件における距離スペクトルをそれぞれ示す . クリーン環境の結果は , 図 3.6(a), 図 3.7(a) と同様である . 図 3.12, 3.13 より , 白色雑音の結果である図 3.6, 3.7 と同様に低 SNR 環境では従来の音響測距法では雑音に距離に相当する干渉成分が埋もれてしまっているが , 提案手法を適用することによって低 SNR 環境でも距離スペクトルのピークが検出できることがわかる .

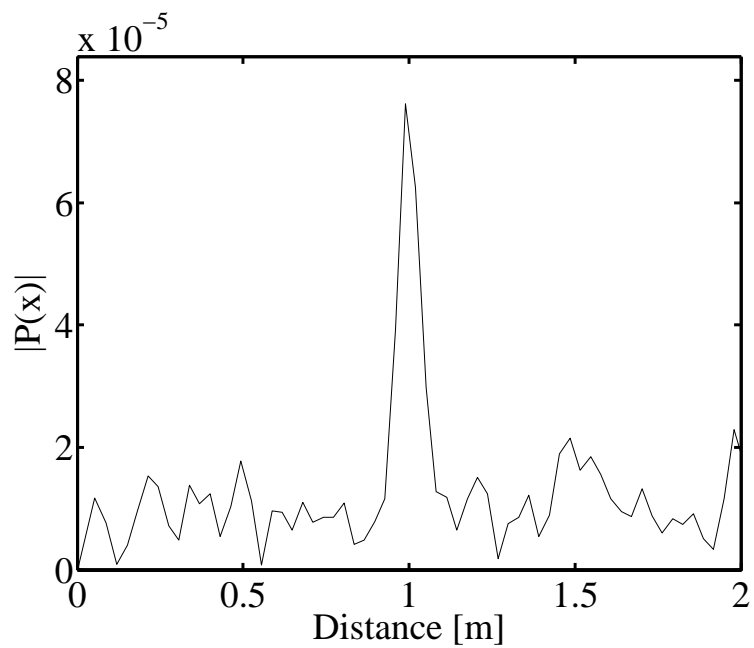


(a) クリーン環境

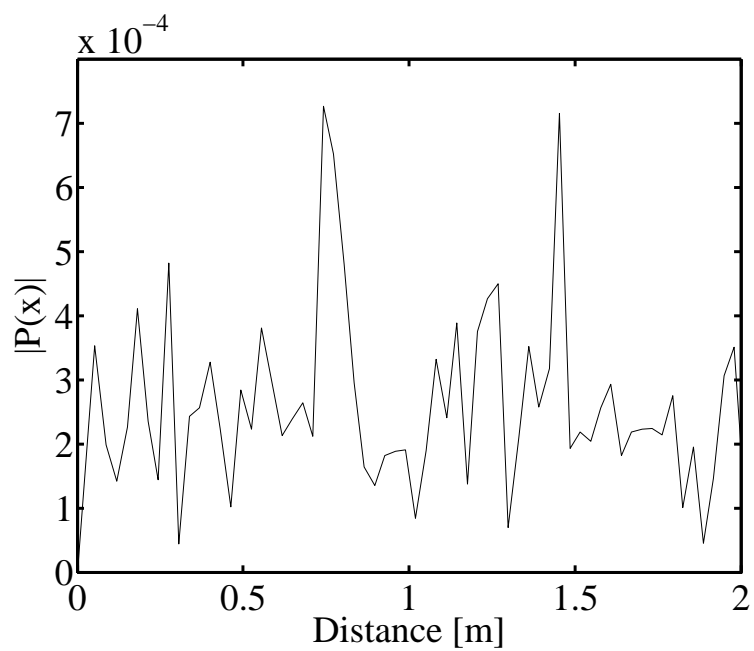


(b) SNR = 0 dB 環境

図 3.6 雑音環境下音響測距実験における  $\Delta p(f, 0)$  (雑音: 白色雑音)



(a) クリーン環境



(b) SNR = 0 dB 環境

図 3.7 雑音環境下音響測距実験における距離スペクトル (雑音: 白色雑音)

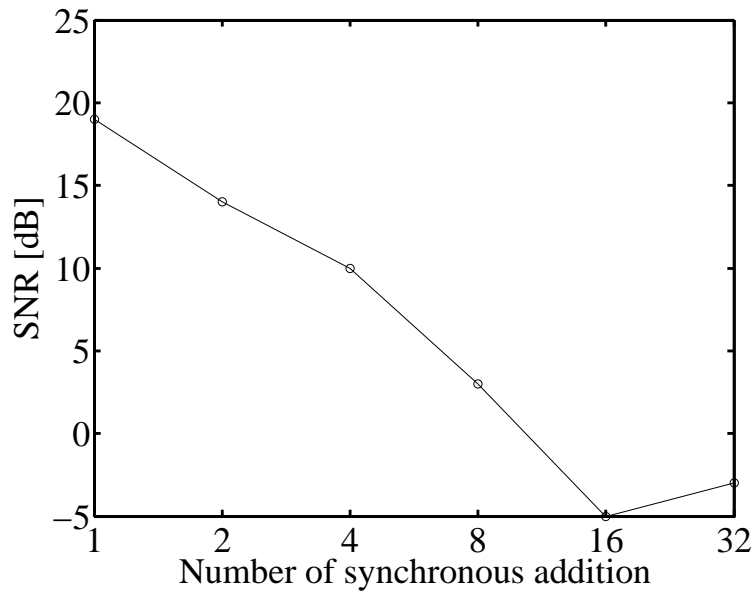


図 3.8 測距可能であった最低 SNR (雑音: 白色雑音)

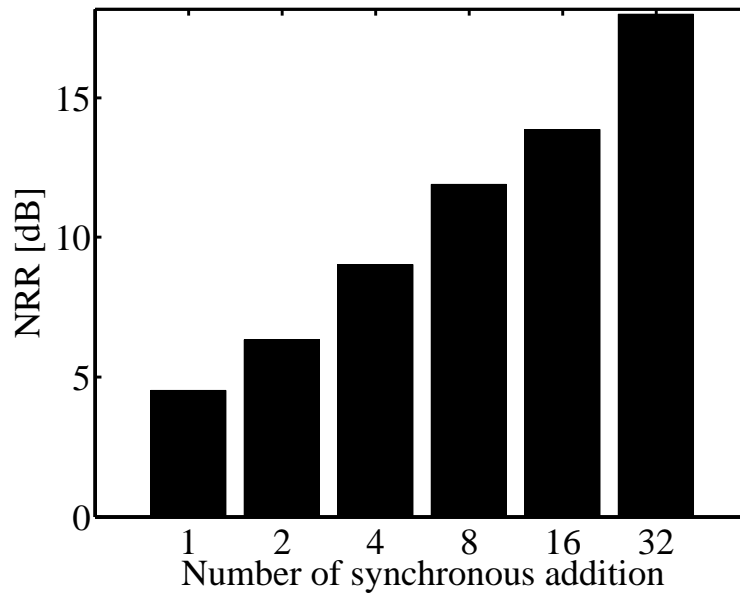


図 3.9 NRR の実験結果 (雑音: 白色雑音)

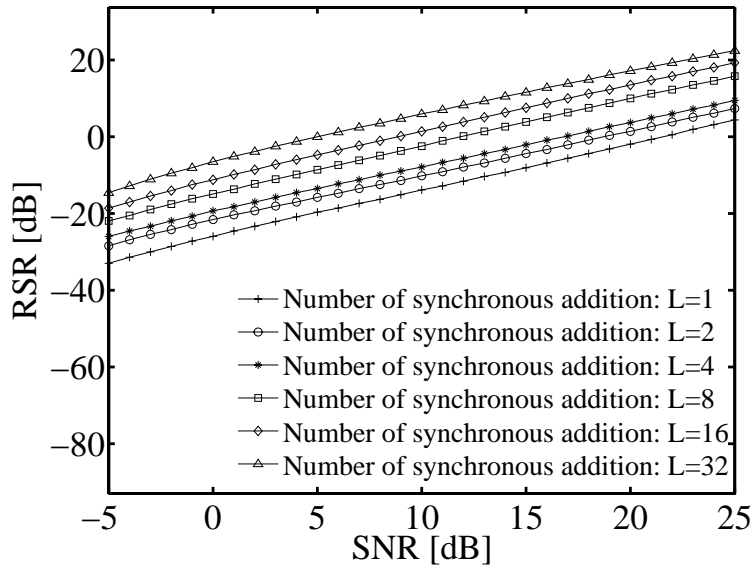


図 3.10 RSRの実験結果 (雑音: 白色雑音)

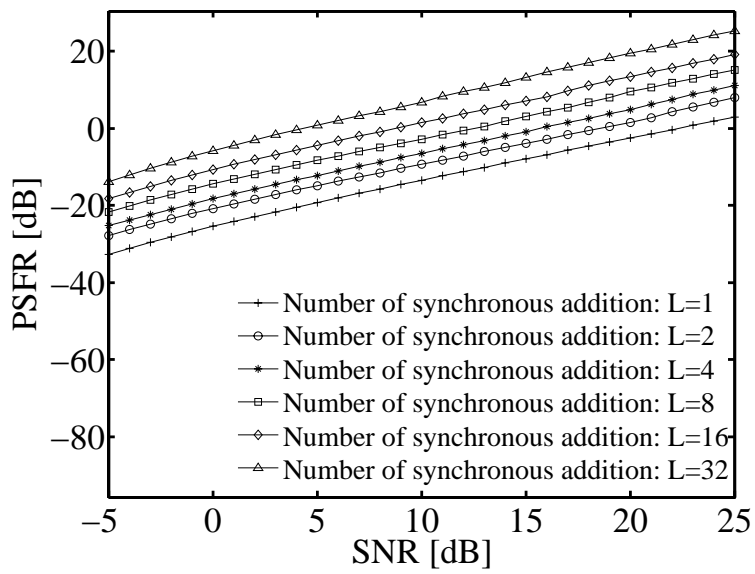
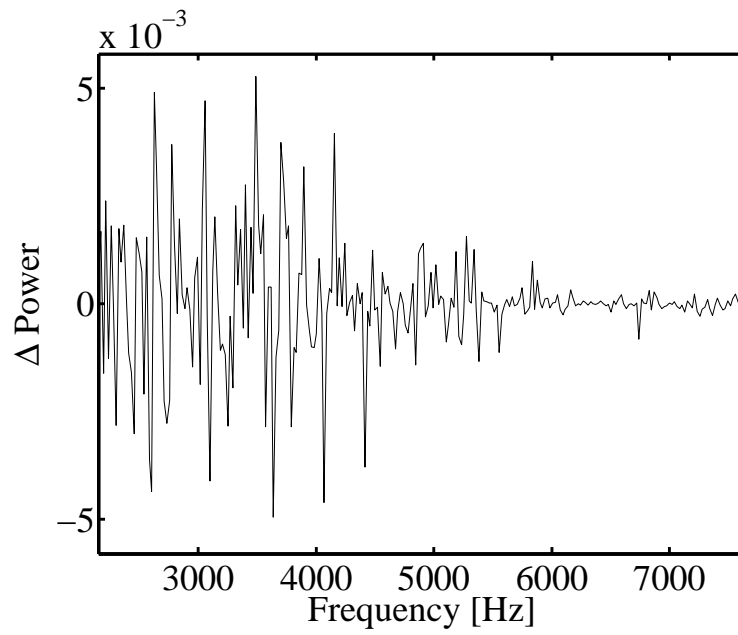
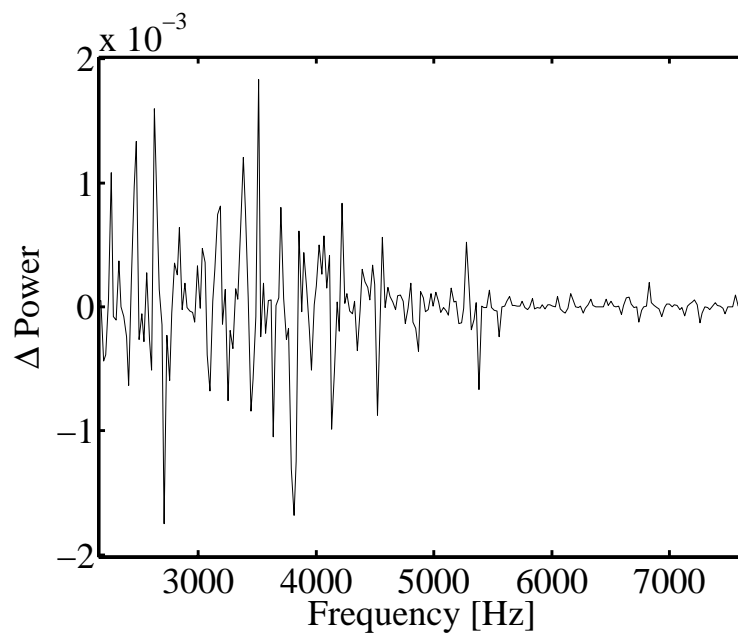


図 3.11 PSFRの実験結果 (雑音: 白色雑音)



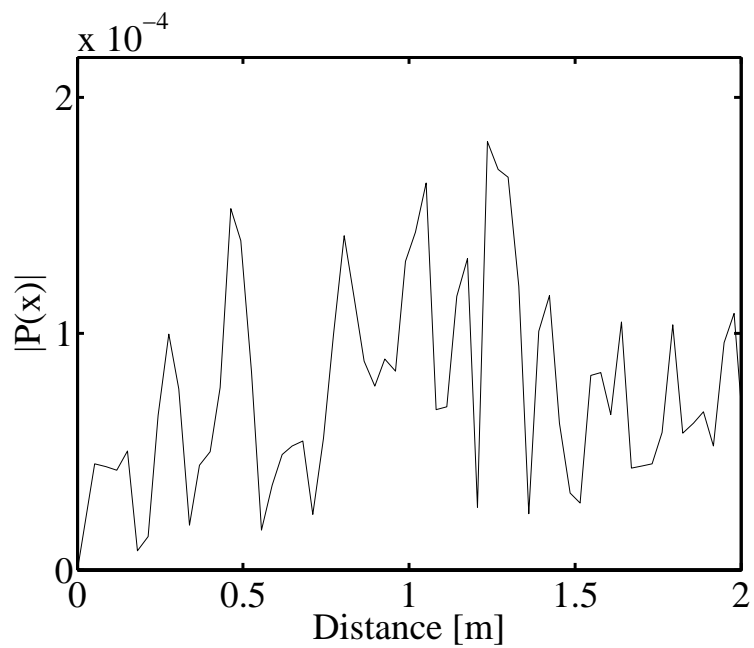


(a) 同期加算回数:  $L=1$

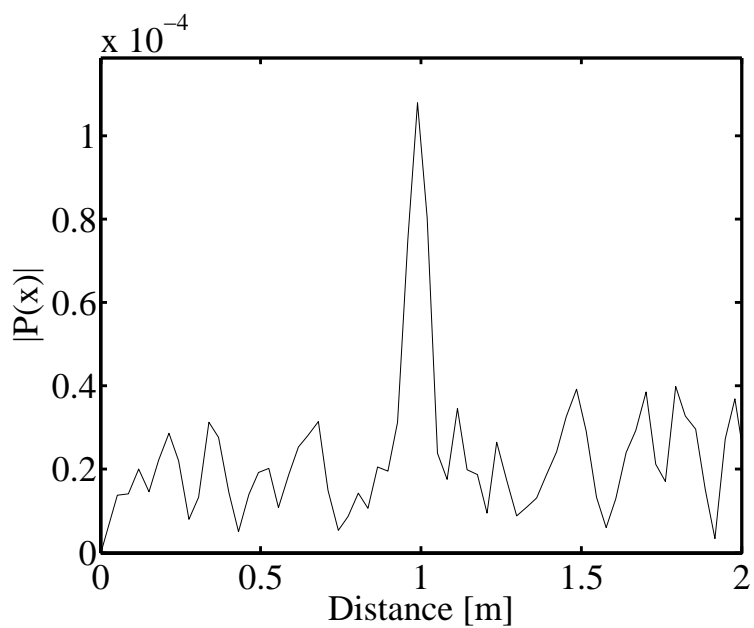


(b) 同期加算回数:  $L=32$

図 3.12 雑音環境下音響測距実験における  $\Delta p(f, 0)$  (雑音: ファンノイズ,  $\text{SNR} = 0 \text{ dB}$ )



(a) 同期加算回数:  $L=1$



(b) 同期加算回数:  $L=32$

図 3.13 雑音環境下音響測距実験における距離スペクトル(雑音: ファンノイズ,  $\text{SNR} = 0 \text{ dB}$ )

図 3.14, 3.15, 3.16, 3.17 に雑音をファンノイズとした場合の測距可能であった最低 SNR, NRR, RSR, PSFR をそれぞれ示す。図 3.14, 3.15, 3.16, 3.17 より、白色雑音の場合と同様に同期加算回数を増やすごとにそれぞれの性能が改善していることがわかる。

図 3.14 の測距性能では、図 3.8 の白色雑音の結果と比較して、低 SNR でも高い測距性能となっている。これは、ファンノイズが白色雑音よりも低い周波数帯域に大きな周波数成分を持っているため、分析周波数帯域の SNR が設定 SNR よりも高い条件になっていること、ファンノイズが定常な雑音の成分だけではなく、突発性・非周期性の雑音成分も含んでいることが原因であると考えられる。

図 3.15 の NRR は、図 3.9 の白色雑音での NRR と比べて少ない同期加算回数で高い NRR となっており、十分な加算回数ではさらに高い NRR となっている。これは、ファンノイズは突発性・非周期性の雑音成分も含むため、SA による雑音抑圧効果によって少ない同期加算回数で高い NRR となり、また十分な平均化が可能な加算回数では SS の効果も加わり、白色雑音の場合に比べて高い NRR となったと考えられる。

図 3.16, 3.17 の RSR, PSFR の結果は、図 3.10, 3.11 の白色雑音における RSR, PSFR と比べて 10 dB 前後高い値となっている。このことから、高い測距性能となっていることがわかる。また、図 3.10, 3.11 の白色雑音の結果と同様に、RSR, PSFR を  $-10$  dB 以上に改善することで測距が可能となる傾向があった。

以上の結果より、送信波の周波数帯域における SNR が測距性能を決定する重要な要素であり、提案手法によって実環境の低 SNR 条件下においても測距性能を十分に改善できることが明らかとなった。

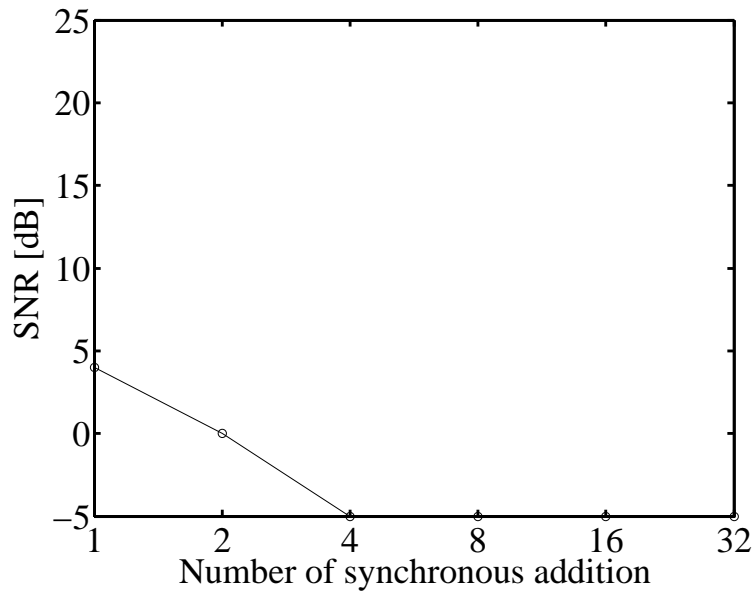


図 3.14 測距可能であった最低 SNR (雑音: ファンノイズ)

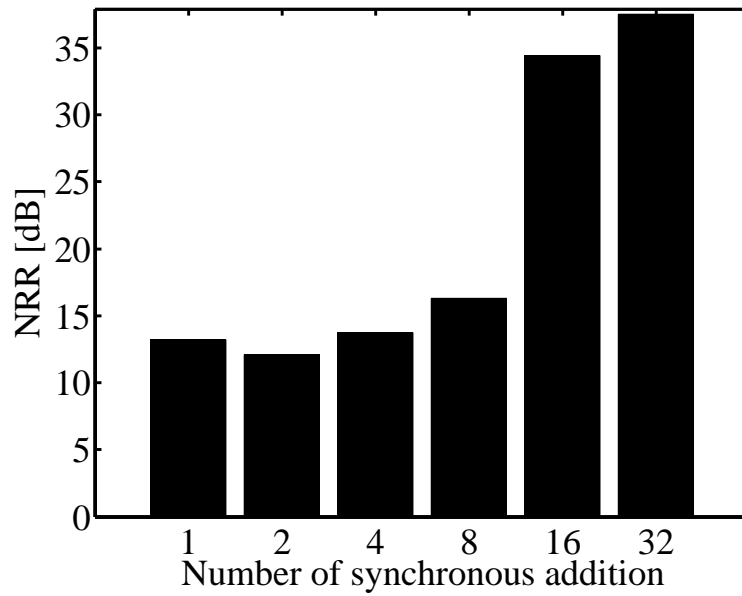


図 3.15 NRR の実験結果 (雑音: ファンノイズ)

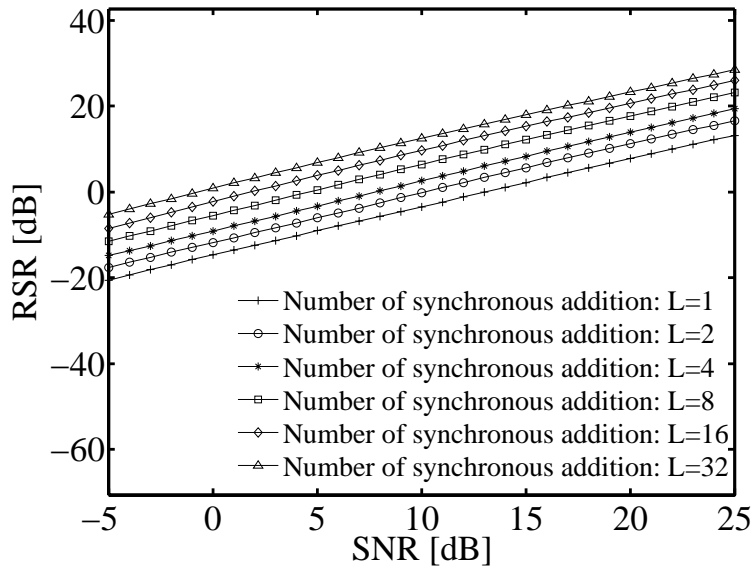


図 3.16 RSRの実験結果 (雑音: ファンノイズ)

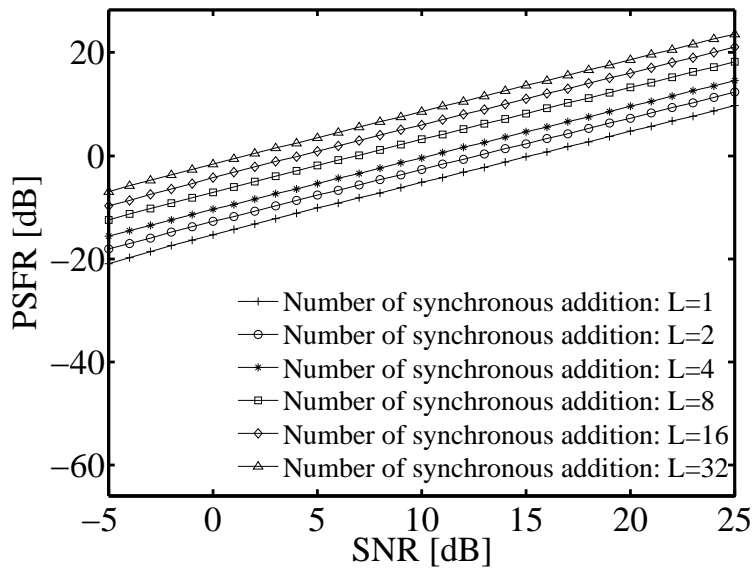


図 3.17 PSFRの実験結果 (雑音: ファンノイズ)

### 3.5. まとめ

本章では，対人・対物センサとして利用可能ある可聴音の直接波と反射波の位相干渉に基づく距離推定の原理とその実現法について述べた．3.2節では可聴音の直接波と反射波の位相干渉に基づく距離推定の原理の詳細について述べ，3.2節では実環境において問題となる送信波のスペクトル特性や計測系の歪みを緩和する可聴音の直接波と反射波の位相干渉に基づく距離推定の実現手法の一つについて述べ，実環境において対象物までの距離が可聴音を利用して測定可能であることを述べた．

本論文では，送信波と反射波の干渉を利用した音響測距法を雑音環境下における実環境の測定系を考慮した理論に拡張し，従来の雑音抑圧手法であるSAとSSの考え方を提案手法に導入することで，提案手法に適した雑音抑圧手法を提案した．

実環境において雑音環境を考慮した評価実験を行った結果，提案手法が低SNRの雑音環境においても有効であることが明らかとなった．

今後の課題として，独立成分分析 ( Independent Component Analysis; ICA ) [70, 71] などを利用した位相干渉成分の抽出手法やカルマンフィルタ [58, 59, 60]・パーティクルフィルタ [72, 73] など他の雑音抑圧手法の検討が挙げられる．また，本論文では，雑音として白色雑音とファンノイズを用いたが，他にも音声などが雑音となった場合の評価実験も行いたい．

## 第4章 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

### 4.1. はじめに

話者（物体）もしくは音源の位置検出は，多くの工学分野において非常に重要な情報である．物体の位置検出では，レーダの分野において様々な研究が行われている [48, 49, 50, 51]．特に近距離の距離測定においては，送信波と反射波の干渉を利用した距離推定法が知られている [53, 54, 55]．3章では，この距離推定技術を可聴音域に拡張し，雑音環境下においても動作できるように検討した．

また，音源の位置検出手法として，マイクロホンアレー技術が注目されている．マイクロホンアレー技術では，複数マイクロホンで受信した信号（発話音声）の時間差情報を利用して音源位置を推定する [1, 37, 38, 39, 40, 41, 42, 43, 44, 45]．しかし，話者の発話による話者位置推定では，非発話状態での話者位置の推定が困難であった．従来手法として全方位カメラによる話者の画像情報と音源位置情報を組み合わせることで非発話状態での話者位置の推定を行う手法 [21] も提案されているが，マイクロホン以外の観測機材を組み合わせる必要があった．

本論文では，可聴音の送信波と反射波の位相干渉に基づく距離推定法を複数マイクロホンに拡張することで，対象物までの距離と方位より話者（対象物）位置を推定する手法を提案する．提案手法は，2つのマイクロホンと1つのスピーカの構成となっている．一般的なハンズフリー音声対話型インタフェースでは，音声入力用のマイクロホンと音声出力用のスピーカがよく用いられるため，特別な観測機材を必要としないという利点がある．また，提案手法では，話者位置を推定するために測定音を出力する必要があるが，音声対話システムの場合にはアナウンス音声や効果音などが測定音として利用可能である．

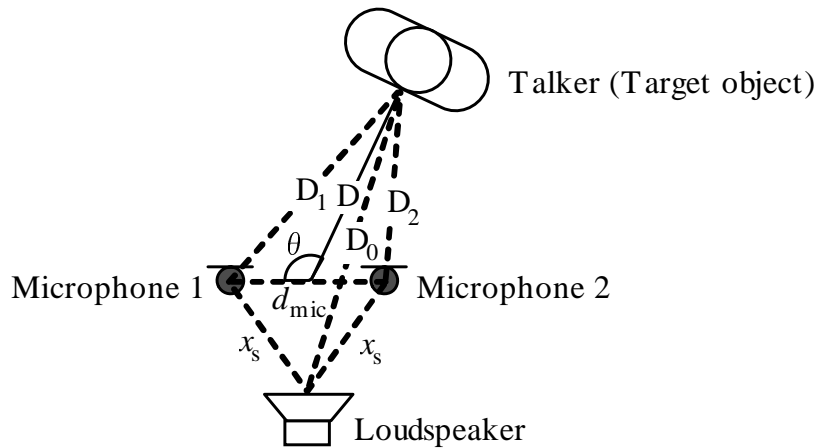


図 4.1 話者位置推定を行う提案手法における話者，スピーカとマイクロホンの関係

また，話者の発話区間では DOA 推定手法である CSP 法 [44, 45] を利用し，非発話区間では提案する話者位置推定手法を適用する手法も提案する．

本章の構成を以下に示す．4.2 節では，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法について述べる．4.3 節では，実環境において行った話者位置推定実験とその結果について述べる．最後に，4.4 節で本章のまとめを述べる．

## 4.2. 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

本節では，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手法を提案する．

提案手法は，従来の可聴音の送信波と反射波の位相干渉に基づく距離推定法では 1 つであった観測用のマイクロホンをマイクロホンアレー (2ch マイクロホン) に拡張する．提案手法における話者 (対象物)，スピーカとマイクロホンの関係を図 4.1 に示す．図 4.1 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．ただし，



従来手法である式 (3.12) では、スピーカ・対象物軸上にマイクロホンを設置する条件のため、往復距離を 2 分の 1 (距離軸を 0.5 倍にスケールリング) にすることで片道距離に相当する距離スペクトルへ射影していた。しかし、提案手法ではマイクロホンアレーを利用するため、スピーカ・対象物軸上にマイクロホンを設置できない。そこで、提案手法の距離スペクトル  $|P_1(x)|$ ,  $|P_2(x)|$  は式 (3.12) ではなく式 (4.1) より算出する。

$$P_k(x) = \sum_{i=1}^N p_k(f_i, x_s) e^{-j2\pi\frac{x}{c}f_i}, \quad (4.1)$$

ここで、 $k$  は、1 もしくは 2 であり、式 (3.12) と式 (4.1) の違いは、式 (3.12) では  $2x$  となっている距離軸への変換が式 (4.1) では  $x$  とすることで距離軸のスケールリングを行わない点である。このため、従来の距離推定で式 (3.12) ではなく式 (4.1) を用いると、対象物-マイクロホン間を往復した距離からスピーカ-マイクロホン間の距離を減算した距離  $2d_n - x_s$  が算出される。

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  $x_1$ ,  $x_2$  は式 (4.1) より式 (4.2), (4.3) で与えられる。

$$x_1 = \operatorname{argmax}_x (|P_1(x)|), \quad (4.2)$$

$$x_2 = \operatorname{argmax}_x (|P_2(x)|). \quad (4.3)$$

この距離  $x_1$ ,  $x_2$  は図 4.1 の  $x_s$ ,  $D_0$ ,  $D_1$ ,  $D_2$  を用いると式 (4.4), (4.5) となる。

$$x_1 = D_1 + D_0 - x_s, \quad (4.4)$$

$$x_2 = D_2 + D_0 - x_s, \quad (4.5)$$

ここで、スピーカとマイクロホン間の距離  $x_s$  は既知である。

図 4.1 では、理解しやすく記載するためスピーカの前面にマイクロホンを設置している図となっているが、このような設置状況ではマイクロホンの固定器具とスピーカ筐体の間に多重反射が発生して近距離に誤った距離スペクトルのピークが発生してしまう問題がある [75]。

そのため提案手法では、実際にはスピーカの上部にマイクロホンアレーを設置する。スピーカの上部にマイクロホンを設置すると多重反射の影響は軽減されるが、

図 4.1 の対象物，スピーカとマイクロホンアレーが同一平面上に存在しないため，マイクロホンアレーと対象物間の距離  $D$  を正確に求めることが困難となる．

そこで，マイクロホンアレーと対象物間の距離  $D$  を式 (4.6) 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．

$$D \approx \frac{x_1 + x_2}{4}. \quad (4.6)$$

ここで，マイクロホン間隔  $d_{\text{mic}}$  が既知， $\delta = x_2 - x_1$ ， $D_2 = D_1 + \delta$  とすると， $D_1$  は余弦定理より式 (4.7) となる．

$$D_1 = \sqrt{\frac{d_{\text{mic}}^2}{4} + D^2 - \frac{3\delta^2}{8}} - \frac{\delta}{2}. \quad (4.7)$$

式 (4.7) の  $D_1$  を利用して，図 4.1 のマイクロホンアレーから対象物までの方位  $\theta$  は式 (4.8) より推定できる．

$$\theta = \cos^{-1} \left( \frac{d_{\text{mic}}^2/4 + D^2 - D_1^2}{D \cdot d_{\text{mic}}} \right). \quad (4.8)$$

また，話者が発話している状況では，話者の発話音声は提案手法では雑音となるため，提案手法による話者位置の推定が困難となる．そこで，話者の発話区間では，CSP 法による話者音声の DOA 推定を行い，非発話区間では提案手法を利用する組み合わせ手法を検討する．

### 4.3. 実環境における評価実験

提案手法の有効性を示すために，実環境において評価実験を行った．評価実験では，提案手法との組み合わせ手法を検討するために，CSP 法による音源方位の推定実験もあわせて行った．

#### 4.3.1 実験条件

実験で使用した機材を表 4.1 に示す．話者（対象物），スピーカとマイクロホンの関係を図 4.2, 4.3 に示す．スピーカは床から 1.1 m に設置し，マイクロホンは図 4.2,

表 4.1 話者位置推定における実験で使用した機材

PCM recorder	SONY, PCM-D50
Microphone	SONY, ECM-88B
Microphone amplifier	AUDIO-TECHNICA, AT-MA2
Loudspeaker	BOSE, 101MM
Power amplifier	BOSE, 1705II
Audio interface	ROLAND, UA-25EX

4.3 のようにスピーカの上部に設置した。また，話者はマイクロホンアレーの中心から 0～180 度まで 10 度刻みの位置で直立不動の状態ですべて計測を行った。マイクロホンアレーと話者間の距離は，0.5 m と 1.0 m の 2 条件で計測を行った。これは，一般的な遠隔発話音声を受音するハンズフリー音声対話型インタフェースでは，話者とマイクロホンまでの距離を 0.5～1.0 m に設定していることが多いためである。実験は，残響時間 0.6 秒の演習室 ( Room A, 部屋の大きさ: 7.0 m × 9.0 m ) 中央付近で行った。また，残響時間 0.5 秒の室内 ( Room B, 部屋の大きさ: 5.0 m × 8.0 m ) の壁際においても，同様の実験を行った。この実験では，マイクロホンアレーと話者間の距離は，1.0 m で計測を行った。

実験では，TSP 法で計測したインパルス応答を利用した。提案手法による話者位置推定では，計測したインパルス応答を観測信号とした。そのため，送信音はインパルス信号であり，加算雑音がほとんど存在しない条件となる。CSP 法による DOA 推定では，話者位置にスピーカを設置して計測したインパルス応答とドライソースである男性話者音声を畳み込んだ信号を観測信号とした。

サンプリング周波数を 44.1 kHz，距離推定に利用するデータ点数を 4,096 点，音速を 342.48 m/s，距離推定に使用する周波数帯域は可聴音域であり，十分な反射波が発生する 2.1 kHz から 7.6 kHz まで ( 周波数帯域幅 5.5 kHz，512 点 ) とした。また，距離スペクトルへの変換の際にデータ点数が 512 点の周波数軸上のデータを 4,096 点となるように 0 詰めを行った。この操作により，距離スペクトルの解像度 ( 刻み幅 ) が 8 倍となる。

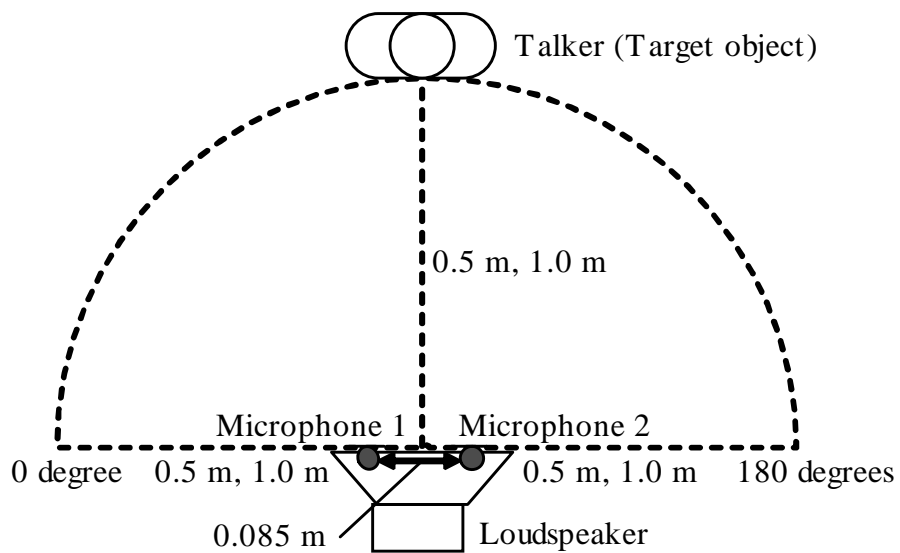


図 4.2 話者位置推定実験における実験状況 (上方図)

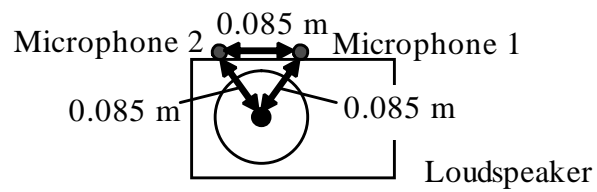


図 4.3 話者位置推定実験における実験状況 (正面図)

### 4.3.2 実験結果

図 4.4, 4.5, 4.6 は話者の方位が 60 度の条件での提案手法の距離スペクトルであり, 図 4.4 はマイクロホンアレーと話者間の距離 0.5 m ( Room A ), 図 4.5 はマイクロホンアレーと話者間の距離 1.0 m ( Room A ), 図 4.6 はマイクロホンアレーと話者間の距離 1.0 m ( Room B ) の条件における実験結果である .

図 4.4 の距離スペクトルのピークは, Microphone 1 は 0.991 m, Microphone 2 は 1.022 m にあり, そのときの推定距離は 0.503 m, 方位推定角は 62 度であった . 図 4.5 の距離スペクトルのピークは, Microphone 1 は 1.912 m, Microphone 2 は 1.952 m にあり, そのときの推定距離は 0.985 m, 方位推定角は 57 度であった . 図 4.6 の距離スペクトルのピークは, Microphone 1 は 1.936 m, Microphone 2 は 1.999 m にあり, そのときの推定距離は 0.984 m, 方位推定角は 57 度であった .

図 4.4, 4.5, 4.6 の距離スペクトルのピークが設定距離の約 2 倍となっているのは, 提案手法の距離スペクトルの変換式である式 (4.1) で算出しているためである .

図 4.7, 4.8, 4.9 は話者の方位が 0 ~ 180 度の条件での提案手法による距離推定の結果であり, 図 4.7 はマイクロホンアレーと話者間の距離 0.5 m ( Room A ), 図 4.8 はマイクロホンアレーと話者間の距離 1.0 m ( Room A ), 図 4.9 はマイクロホンアレーと話者間の距離 1.0 m ( Room B ) の条件における実験結果である .

図 4.7, 4.8, 4.9 の提案手法による距離推定の実験結果より, マイクロホンアレーと話者間の距離 0.5 m ( Room A ) における推定距離の平均誤差は 0.018 m, マイクロホンアレーと話者間の距離 1.0 m における推定距離の平均誤差は 0.020 m ( Room A ), 0.020 m ( Room B ) であった . このことから, 約 0.02 m の誤差で, 距離推定が可能であることがわかった .

図 4.10, 4.11, 4.12 は話者の方位が 0 ~ 180 度の条件での提案手法による話者 ( 対象物 ) の方位 ( Direction Of Target; DOT ) 推定の結果であり, 図 4.10 はマイクロホンアレーと話者間の距離 0.5 m ( Room A ), 図 4.11 はマイクロホンアレーと話者間の距離 1.0 m ( Room A ), 図 4.12 はマイクロホンアレーと話者間の距離 1.0 m ( Room B ) の条件における実験結果である .

図 4.10, 4.11, 4.12 の話者の方位推定の実験結果より, マイクロホンアレーと話者間の距離 0.5 m ( Room A ) における方位推定角の平均誤差は 7 度, マイクロホン

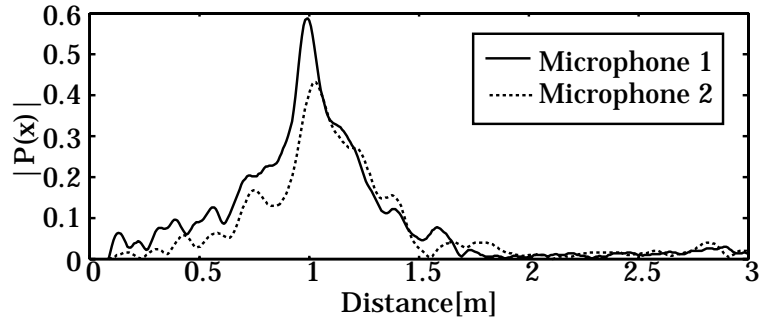


図 4.4 話者位置推定実験における提案手法の距離スペクトル( Room A, 0.5 m, 60 度 )

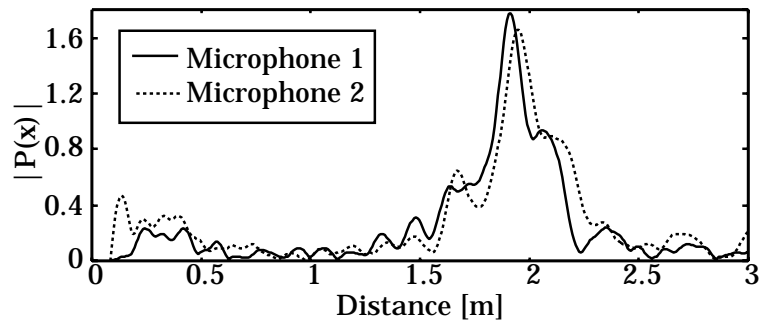


図 4.5 話者位置推定実験における提案手法の距離スペクトル( Room A, 1.0 m, 60 度 )

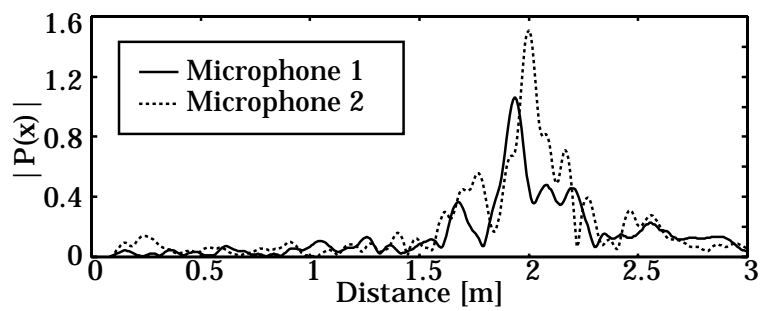


図 4.6 話者位置推定実験における提案手法の距離スペクトル( Room B, 1.0 m, 60 度 )

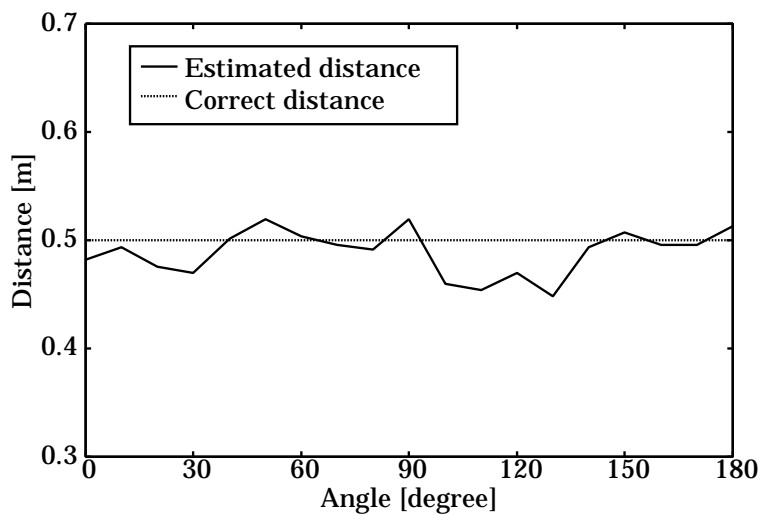


図 4.7 話者位置推定実験における提案手法による距離推定結果 ( Room A, 0.5 m )

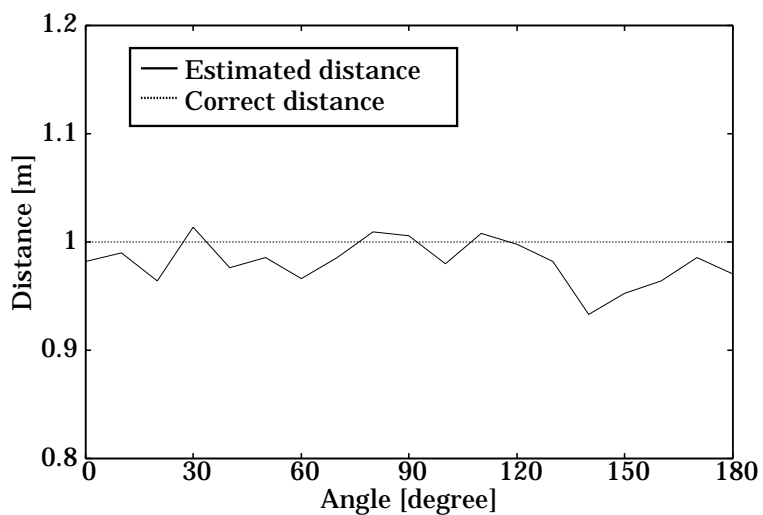


図 4.8 話者位置推定実験における提案手法による距離推定結果 ( Room A, 1.0 m )

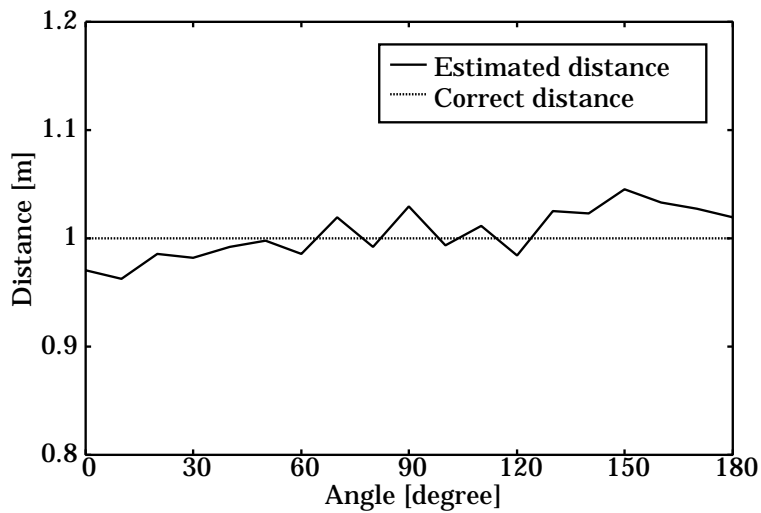


図 4.9 話者位置推定実験における提案手法による距離推定結果 ( Room B, 1.0 m )

アレーと話者間の距離 1.0 m における方位推定角の平均誤差は 7.5 度 ( Room A ) , 8.5 度 ( Room B ) であった . また , 特に高精度に推定できた 30 ~ 150 度の範囲では , マイクロホンアレーと話者間の距離 0.5 m ( Room A ) における方位推定角の平均誤差は 4.3 度 , マイクロホンアレーと話者間の距離 1.0 m における方位推定角の平均誤差は 4.8 度 ( Room A ) , 7.9 度 ( Room B ) であった . このことから , 30 ~ 150 度の範囲では , 約 5 度の誤差で話者の方位を推定可能であることがわかった .

図 4.13, 4.14, 4.15 は話者の方位が 60 度の条件での CSP 法による CSP 係数であり , 図 4.13 はマイクロホンアレーと話者間の距離 0.5 m ( Room A ) , 図 4.14 はマイクロホンアレーと話者間の距離 1.0 m ( Room A ) , 図 4.15 はマイクロホンアレーと話者間の距離 1.0 m ( Room B ) の条件における実験結果である .

図 4.13 の CSP 係数のピーク , 図 4.14 の CSP 係数のピークは共に 62 度であった . また , 図 4.15 の CSP 係数のピークは 59 度であった .

図 4.16, 4.17, 4.18 は話者の方位が 0 ~ 180 度の条件での CSP 法による話者音声の DOA 推定の結果であり , 図 4.16 はマイクロホンアレー-話者間の距離 0.5 m ( Room A ) , 図 4.17 はマイクロホンアレーと話者間の距離 1.0 m ( Room A ) , 図 4.18 はマイクロホンアレーと話者間の距離 1.0 m ( Room B ) の条件における実験結果である .



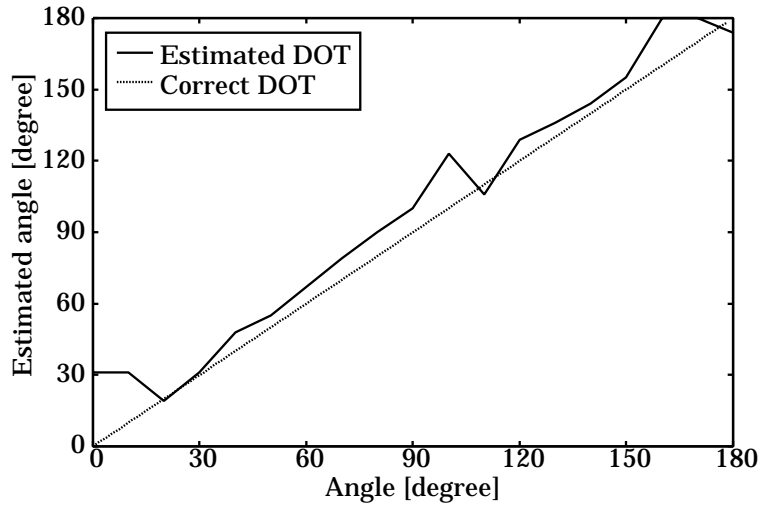


図 4.10 話者位置推定実験における提案手法による DOT 推定結果 ( Room A, 0.5 m )

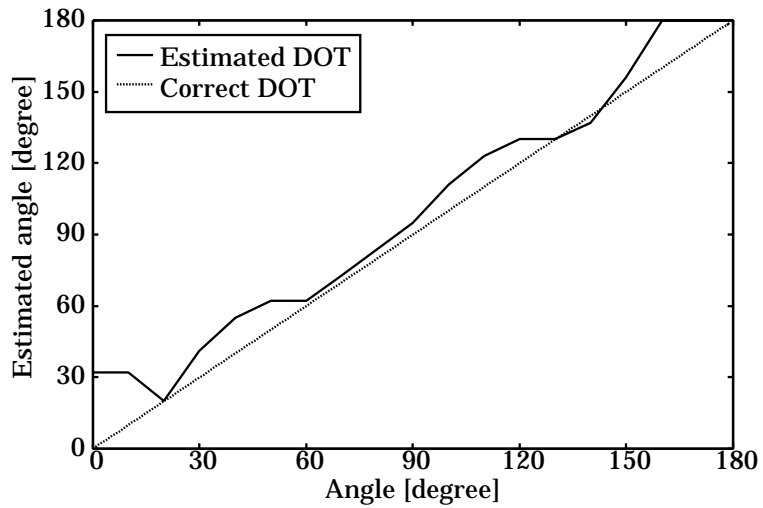


図 4.11 話者位置推定実験における提案手法による DOT 推定結果 ( Room A, 1.0 m )

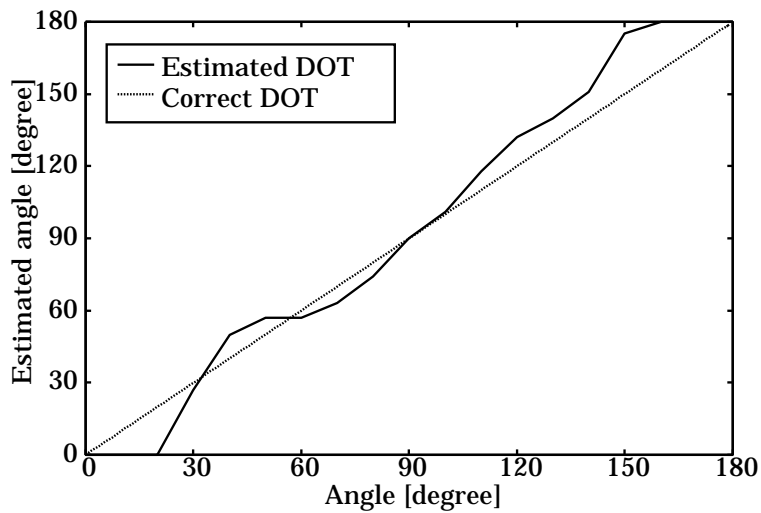


図 4.12 話者位置推定実験における提案手法による DOT 推定結果 ( Room B, 1.0 m )

図 4.16, 4.17, 4.18 の CSP 法による DOA 推定の実験結果より，マイクロホンアレーと話者間の距離 0.5 m ( Room A ) おける方位推定角の平均誤差は 18.8 度，マイクロホンアレーと話者間の距離 1.0 m における方位推定角の平均誤差は 15.0 度 ( Room A )，17.4 度 ( Room B ) であった．また，特に高精度に推定できた 30～150 度の範囲では，マイクロホンアレーと話者間の距離 0.5 m ( Room A ) における方位推定角の平均誤差は 3.0 度，マイクロホンアレーと話者間の距離 1.0 m における方位推定角の平均誤差は 3.5 度 ( Room A )，4.4 度 ( Room B ) であった．このことから，30～150 度の範囲では，約 4 度の誤差で話者の方位を推定可能であることがわかった．

図 4.10, 4.11, 4.12, 4.16, 4.17, 4.18 の方位推定の実験結果では，0～20 度，160～180 度で推定精度が劣化している．この原因として， $\cos^{-1}$  による角度射影の関係上，角度の解像度が 0～20 度，160～180 度の範囲では低いことに加えて，マイクロホンアレーをスピーカの上に設置していることによって発生するスピーカ筐体からの反射音の影響が考えられる．この反射音の影響は，スピーカ筐体に吸音材を設置することで軽減できる可能性がある．

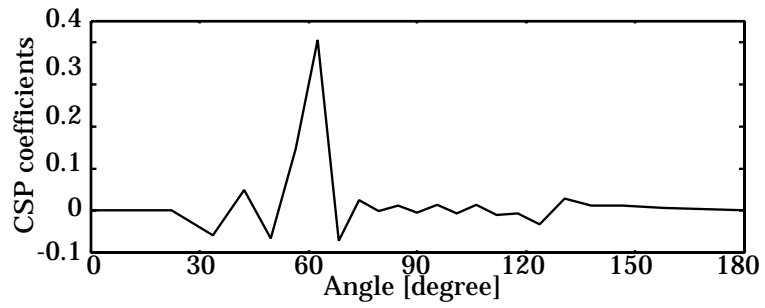


图 4.13 CSP 系数 ( Room A, 0.5 m, 60 度 )

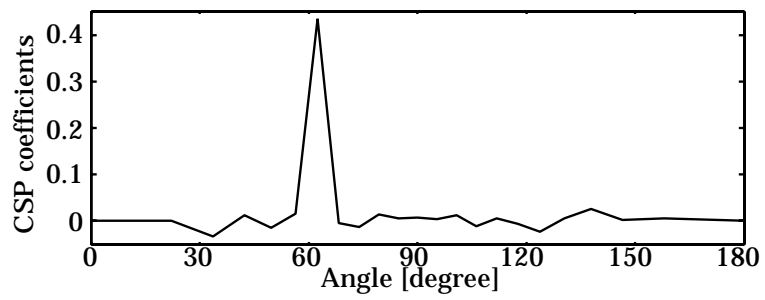


图 4.14 CSP 系数 ( Room A, 1.0 m, 60 度 )

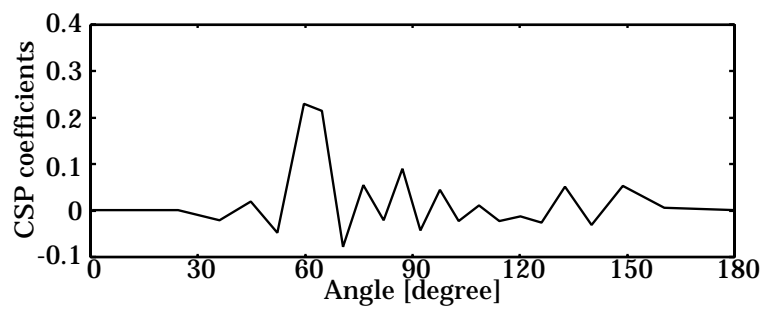


图 4.15 CSP 系数 ( Room B, 1.0 m, 60 度 )

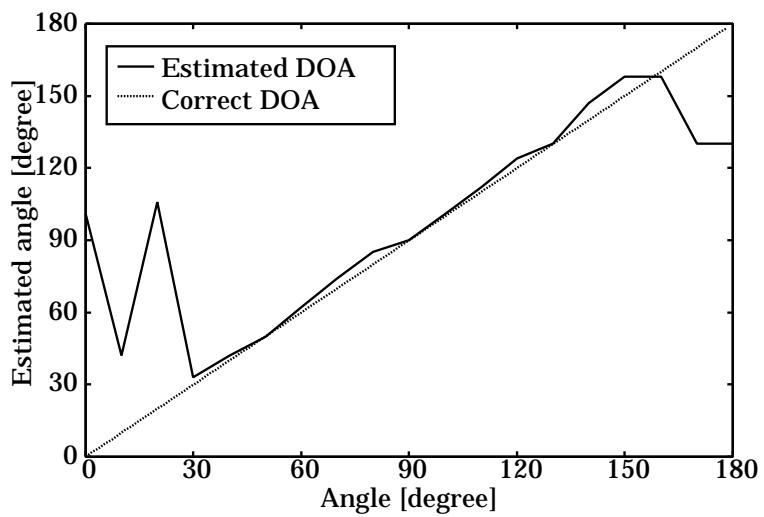


図 4.16 CSP 法による DOA 推定結果 ( Room A, 0.5 m )

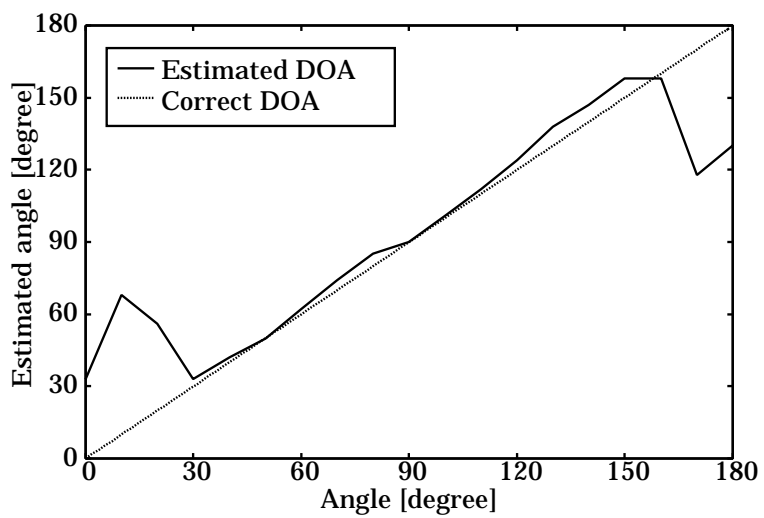


図 4.17 CSP 法による DOA 推定結果 ( Room A, 1.0 m )

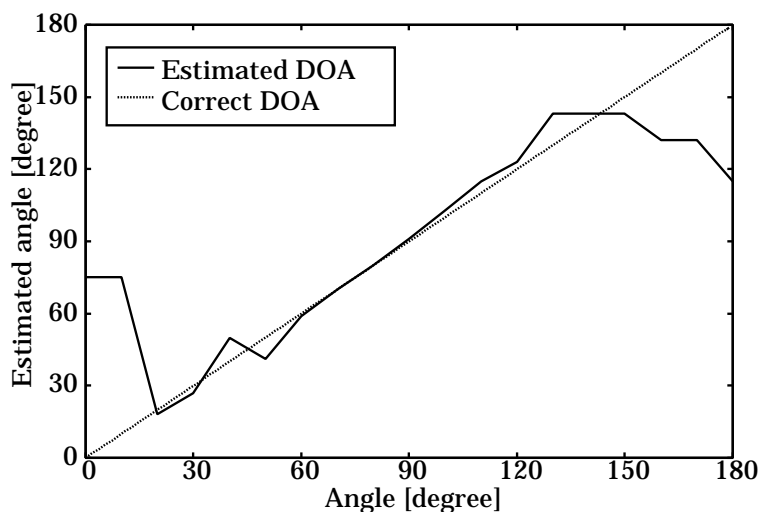


図 4.18 CSP 法による DOA 推定結果 ( Room B, 1.0 m )

#### 4.4. まとめ

マイクロ波レーダの分野における定在波を利用した近距離計測法を可聴音域に拡張し、近距離の測定対象物に対する距離推定法が提案されている。本章では、可聴音による定在波を利用した近距離計測法の受音用マイクロホンをマイクロホンアレーに拡張することで、話者（対象物）位置を推定する手法を提案した。また、話者の発話/非発話によって、話者の発話を利用する CSP 法による音源方位推定と提案手法を切り替えて適用する手法を検討した。

提案手法を利用すれば、ロボットやハンズフリー音声対話型システムにおける物体の位置推定や話者検出に役立つと考えられる。

実環境における実験の結果、30～150度の範囲では話者の発話区間では CSP 法によって約 3 度の誤差、非発話区間では提案手法によって約 5 度の誤差で方位推定が可能であり、非発話区間では提案手法によって約 0.02 m の誤差で距離推定が可能であることがわかった。

今後は、送信波にインパルス波ではなく音声対話システムで利用されるアナウンス音声や音楽を用いた場合の評価実験、非可聴域に近い高い周波数の送信波を用い

て発話区間でも提案手法による話者位置推定を行えるように拡張したい。また，提案手法を多チャンネルに拡張することで，音響イメージングが可能である。さらに，位相干渉を利用した多チャンネル音響測距法と合成開口 [74] による音響イメージングも検討したい。

# 第5章 適応型アレーを用いた 母音/子音特徴量に基づく 音声強調

## 5.1. はじめに

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）の受音が可能のためである．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ため，背景雑音や残響の影響により音声が悪化する問題がある．この問題に対して，音声認識の分野では雑音下音声を音響モデルに適応学習する手法 [76]，音声強調の分野では雑音下音声から推定した雑音を減算する手法 [57, 58, 59, 61] などが提案されている．しかし，モノラル録音された雑音下音声に対する耐雑音では雑音の統計情報に大きく依存するため，雑音の統計情報が変動する環境においては十分な性能を発揮できない．一方，雑音と音声の空間情報を利用した遠隔発話音声の受音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（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．

2章において述べたように，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, 3, 4, 5, 6, 7, 8, 9, 10] がある。また，適応型アレーは目的信号に対する歪みを許容しない拘束条件と許容する拘束条件があり，目的信号に対して歪みを許容する拘束条件では大幅に雑音を抑圧可能である。

本論文では，適応型アレーの AMNOR のアルゴリズムを利用する。AMNOR は目的信号の周波数特性の劣化を許すことで大幅に雑音を抑圧できる適応型アレーであり，擬似的に作成した目的信号と背景雑音区間を利用して背景雑音に死角を形成する。AMNOR は，様々な目的信号を対応するため，擬似目的信号に全周波数帯域に均一な周波数特性をもつ白色信号を仮定している。そのため，音声認識において重要な特徴成分である低次フォルマントに対しても大きな歪みが発生して，音声認識性能を劣化させる原因となる。この問題に対して，目的信号を音声のみに限定することで，目的信号である音声に対する歪みを緩和した S-AMNOR [9, 10] が西浦らにより提案されている。

S-AMNOR は，音声に対する歪みを軽減するため音声の長時間平均スペクトルを利用している。しかし，音声の短時間スペクトルは音素によって特徴が大きく異なることが明らかとなっている [78, 79, 80]。音素の特徴量を考慮した従来の単一マイクによる雑音下音声強調手法として音声の短時間スペクトルにおける基本周波数，ピッチ構造を利用した手法 [81] などが提案されている。この音声の特徴量を考慮した考え方を拡張して適応型アレーの拘束条件として取り入れることで，さらに音声に対する歪みを軽減できる可能性がある。また，この拡張は収録環境に依存しないため，従来のマイクロホンアレーと同様に雑音の統計情報が変動する環境においても頑健な耐雑音が可能であると考えられる。

そこで本章では，音声を母音/子音(以下，V/C)の2つに分類して雑音下観測音声の V/C 区間に対応して V/C に対する歪みを緩和したビームフォーミングを AMNOR の拘束条件下で行うことにより，音声認識性能の改善を試みる。提案手法では，受信した観測音声の V/C 区間に正しく V/C に対する歪みを緩和したビームフォーミングを適用するために，観測音声の V/C セグメンテーションが必要不可欠となる。そこで，提案手法では観測音声の V/C セグメンテーションに従来の AMNOR と音素認識の利用を検討する。

本章の構成を以下に示す。5.2 節では，提案手法の詳細について述べる。5.3 節で



は，実環境において行った評価実験とその結果について述べる．最後に，5.4節で本章のまとめを述べる．

## 5.2. VC-AMNOR ( V/C 特徴量を利用した 適応型アレー ) の提案

従来手法である S-AMNOR は，音声の長時間平均スペクトルを考慮したビームフォーミングを行っているため，音声の短時間スペクトルに対して十分に歪み軽減ができない問題点があった．しかし，従来の単一マイクによる雑音下音声強調手法には，音声の短時間スペクトルにおける基本周波数，ピッチ構造を利用した手法 [81] などが提案されている．この考え方を拡張して適応型アレーの拘束条件に取り入れることで，さらに音声に対する歪みを軽減できる可能性がある．

本論文では，音声の短時間スペクトルの中でも母音/子音 ( V/C ) に注目する．観測された雑音下音声 ( 短時間フレーム ) の V/C の種別を把握することができれば，音声の V/C に対してさらに歪みを軽減できる可能性がある．そこで本章にて，V/C の特徴量を利用した適応型アレー ( VC-AMNOR: average Vowel/Consonant spectrum-based AMNOR ) を提案する．

提案する VC-AMNOR は，計 2 回のビームフォーミングを必要とする．まず，1 回目に従来の AMNOR による大雑把なビームフォーミングと音素認識で観測された雑音下音声 ( 短時間フレーム ) の背景雑音 ( 非発話 ) 区間検出を含む V/C セグメンテーションを行い，2 回目に V/C セグメンテーション結果を利用して V/C の特徴量を考慮した詳細なビームフォーミングを行うことで，音声の短時間スペクトルである V/C に対する歪みを軽減する構造とした．これは，V/C のセグメンテーション程度であれば，大雑把なビームフォーミングと音素認識でもある程度の V/C 正解精度を保てる可能性が高いことを利用している．また，V/C セグメンテーションで背景雑音と判定された区間では雑音を抑圧するため出力ゲインを小さくしている．

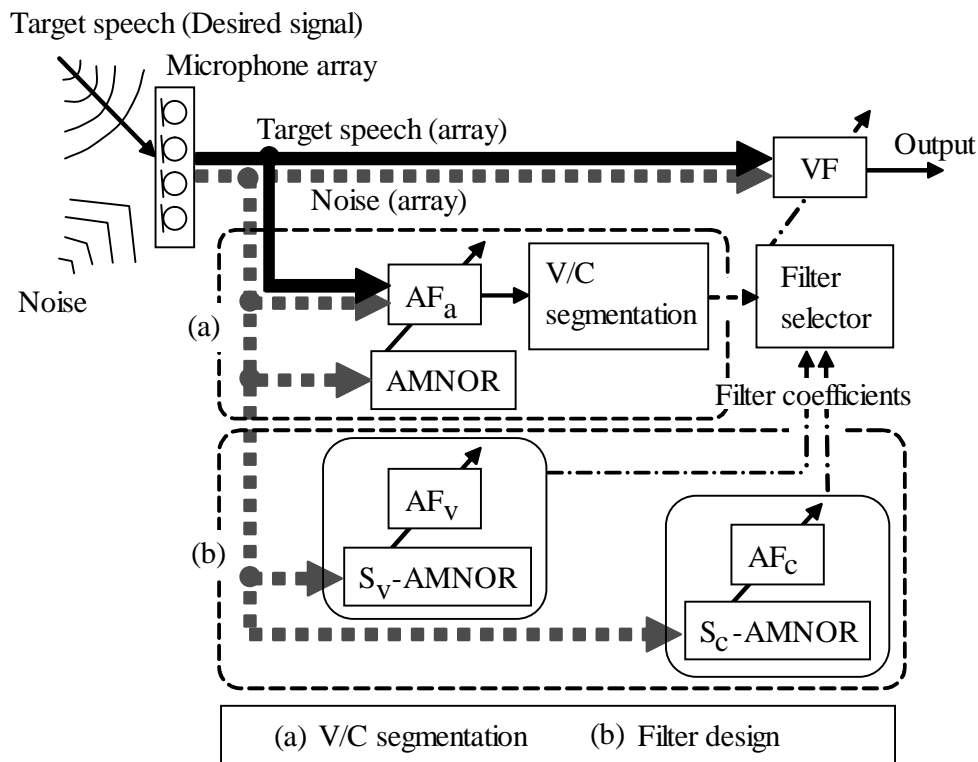


図 5.1 VC-AMNOR の概略図

### 5.2.1 VC-AMNOR のアルゴリズム

本節では、提案する VC-AMNOR の動作アルゴリズムについて説明する。図 5.1 に提案手法である VC-AMNOR の概略図を示す。図 5.1 において、 $AF_a, AF_v, AF_c$  は  $M$  入力 1 出力の FIR フィルタを示す。 $AF_a$  は従来の AMNOR で設計したビームフォーミング用 FIR フィルタである。 $AF_v$  は母音に対して歪みを軽減するように AMNOR 拘束条件下 ( $S_v$ -AMNOR) で設計したビームフォーミング用 FIR フィルタ、 $AF_c$  は子音に対して歪みを軽減するように AMNOR 拘束条件下 ( $S_c$ -AMNOR) で設計したビームフォーミング用 FIR フィルタである。 $VF$  は  $AF_v, AF_c, AF_a$  のいずれかと同じフィルタ係数をもつビームフォーミング用 FIR フィルタである。

提案手法である VC-AMNOR は、以下の (i) ~ (vii) の処理手順に従って動作する。

- (i) マイクロホンアレーによる背景雑音の受音
- (ii) (i) で受音した背景雑音を利用して AMNOR ,  $S_v$ -AMNOR ,  $S_c$ -AMNOR でビームフォーミング用 FIR フィルタ  $AF_a$  ,  $AF_v$  ,  $AF_c$  を設計 ( 図 5.1 中 (a),(b) )
- (iii) マイクロホンアレーによる雑音下音声の受音
- (iv) (ii) で設計した  $AF_a$  により , (iii) で受音した雑音下音声に対して 1 段階目のビームフォーミング ( 図 5.1 中 (a) )
- (v) (iv) で雑音抑圧処理された音声に対して V/C セグメンテーション ( 図 5.1 中 (a) )
- (vi) (v) の V/C セグメンテーション結果より , (iii) で受音した雑音下音声に V/C 区間に対応したフィルタ  $AF_v$  ,  $AF_c$  による 2 段階目のビームフォーミングした信号を出力 . (iii) の処理手順に移行
- (vii) (v) の V/C セグメンテーション結果 , 非発話区間と判定された場合 , フィルタ  $AF_a$  を (iii) で受音した雑音下音声に適用後 , ゲイン減衰係数 ( 本論文では 0.1 ) をかけた信号を出力 . (i) の処理手順に移行

## 5.2.2 音素の特徴量

音声を V/C に分類するためには , 日本語の音素を定義する必要がある . 日本語は 43 音素に分類可能である [82] . 表 5.1 に 43 音素と V/C の分類表を示す . 提案手法では , V/C の分類表のように音声を母音と子音の 2 つに分類してそれぞれに対して歪みを緩和するビームフォーミングを AMNOR の拘束条件下で行う .

## 5.2.3 音素認識を用いた V/C セグメンテーション

V/C に対して歪みを軽減したビームフォーミング用 FIR フィルタを雑音下観測音声に適用するためには , 図 5.1 中 (a) のように雑音下観測音声の非発話区間の検出を含む V/C セグメンテーションが必要不可欠である . そこで , 本節では音素認識を利用した雑音下観測音声の V/C セグメンテーションを検討する .

雑音環境下において観測音声をそのまま音素認識すると , 雑音環境に依存して V/C セグメンテーション性能が大きく劣化してしまう . そこで , 式 (5.1) のように雑音下

表 5.1 日本語の音素表記

43 phonemes	
	a i u e o a: i: u: e: o: N w y j my
	ky by gy ny hy ry py p t k ts ch
	b d dy g z m n s sh h f r q
Vowel/consonant	
Vowel	a i u e o a: i: u: e: o:
Consonant	N w y j my ky by gy ny hy ry py p t
	k ts ch b d dy g z m n s sh h f r q

観測音声に対して AMNOR によりビームフォーミングを行う。

$$S(\omega; k) = \sum_{i=1}^M AF_i(\omega) O_i(\omega; k), \quad (5.1)$$

ここで  $S(\omega; k)$  は  $k$  フレームにおける AMNOR によって強調された音声のスペクトル,  $|AF_i(\omega)|$  は  $i$  チャンネルに対応する AMNOR で設計したビームフォーミング用 FIR フィルタのスペクトル,  $O_i(\omega; k)$  は  $k$  フレームにおける  $i$  チャンネルの雑音下観測音声のスペクトル,  $M$  は最大チャンネル数である。式 (5.1) の出力信号  $S(\omega; k)$  を Mel-Frequency Cepstrum Coefficients ( MFCC ) に変換したデータ系列に対して式 (5.2) を用いた HMM に基づく音素認識 [83] による V/C セグメンテーションを行う。

$$\hat{\lambda}_k = \operatorname{argmax}_i P(C_k^{(S)} | \lambda^{(i)}), \quad (5.2)$$

ここで  $C_k^{(S)}$  は  $S(\omega; k)$  から変換した  $k$  フレームにおける MFCC,  $\hat{\lambda}_k$  は  $k$  フレームにおいて最大となった統計モデル,  $i$  は  $sp, v, c$  のいずれかであり,  $\lambda^{(sp)}$  は背景雑音 ( 非発話区間 ) の統計モデル,  $\lambda^{(v)}$  は母音の統計モデル,  $\lambda^{(c)}$  は子音の統計モデルである。

## 5.2.4 V/C に対して歪みを軽減するビームフォーミング用フィルタ の設計

図 5.1 中 (b) のように，提案手法では V/C に対して歪みを軽減するビームフォーミング用 FIR フィルタをそれぞれ AMNOR 拘束条件下で設計することが必要不可欠である．そこで，本節では V/C に対して歪みを軽減するビームフォーミング用フィルタの設計方法について説明する．

手動ラベリングなどにより既に V/C セグメンテーションが行われている音声データベースを利用して母音および子音の平均振幅スペクトル（スペクトル重み係数）を式 (5.3), (5.4) のように計算する．

$$w_v(\omega) = h(k) * \left( \frac{1}{N_v} \sum_{n=1}^{N_v} |SP_v(\omega; n)| \right) + g_v(\omega), \quad (5.3)$$

$$w_c(\omega) = h(k) * \left( \frac{1}{N_c} \sum_{n=1}^{N_c} |SP_c(\omega; n)| \right) + g_c(\omega), \quad (5.4)$$

ここで， $w_v(\omega)$  は母音のスペクトル重み係数， $w_c(\omega)$  は子音のスペクトル重み係数， $\omega$  は周波数インデックス， $h(k)$  はローパスフィルタ， $N_v$  は音声データベースに含まれる母音のフレーム数， $N_c$  は音声データベースに含まれる子音のフレーム数， $|SP_v(\omega; n)|$  は母音フレーム  $sp_v(t; n)$  の振幅スペクトル， $|SP_c(\omega; n)|$  は子音フレーム  $sp_c(t; n)$  の振幅スペクトル， $g_v(\omega)$  は母音に対する振幅スペクトル補正パラメータ， $g_c(\omega)$  は子音に対する振幅スペクトル補正パラメータを表す．振幅スペクトル補正パラメータ  $g_v(\omega)$ ,  $g_c(\omega)$  は平均振幅スペクトルを対数軸上で一定範囲内に補正する振幅補正係数である．また，ローパスフィルタ  $h(k)$  は振幅スペクトルの包絡を得るための平滑処理である（移動平均で代用可）．振幅補正と平滑処理を行うことで，V/C に対するビームフォーミングの出力信号がある特定の周波数帯域のみ支配的となることを防ぐことが出来る．

本論文では，音声データベースとして手動により音素単位でセグメンテーションが行われている JNAS データベース [84] の日本語文を利用して，V/C のスペクトル重み係数を設計した．また，振幅スペクトル補正パラメータは全周波数帯域で同値とし，重み係数の最大と最小のスペクトルの差が 20dB となるように適応的に算出

表 5.2 V/C 平均振幅スペクトルの作成条件

Speech database	JNAS-DB [84]
Phoneme addition	30000 frames
Frame length	32 msec (Hamming window)
Frame interval	8 msec
Sampling frequency	16 kHz
Quantization	16 bits

した．V/C のスペクトル重み係数を重み付けした白色信号を AMNOR の擬似目的信号として利用して式 (2.17) の AMNOR 拘束条件下で V/C に対して歪みを軽減するビームフォーミング用 FIR フィルタを設計する．

白色雑音の周波数特性を  $H(\omega)$  で表したとき，V/C の擬似目的信号  $H_v(\omega), H_c(\omega)$  は式 (5.5), (5.6) のようになる．

$$H_v(\omega) = w_v(\omega)H(\omega), \quad (5.5)$$

$$H_c(\omega) = w_c(\omega)H(\omega), \quad (5.6)$$

ここで， $w_v(\omega), w_c(\omega)$  は V/C の重み係数である．

V/C に対して歪みを軽減するビームフォーミング用 FIR フィルタ設計の一例として，白色雑音に V/C のスペクトル重み係数を重み付けした信号を図 5.2, 5.3 に示す．V/C の平均振幅スペクトルは，表 5.2 の実験条件にて作成した．この擬似目的信号を利用して AMNOR 拘束条件下で V/C に対して歪みを軽減するビームフォーミング用 FIR フィルタを設計する．提案する VC-AMNOR は，本節で作成した V/C 用ビームフォーミング用 FIR フィルタを，観測音声の V/C 区間に対応して切り替えて適用することで，音声に対する歪みの軽減を試みる．

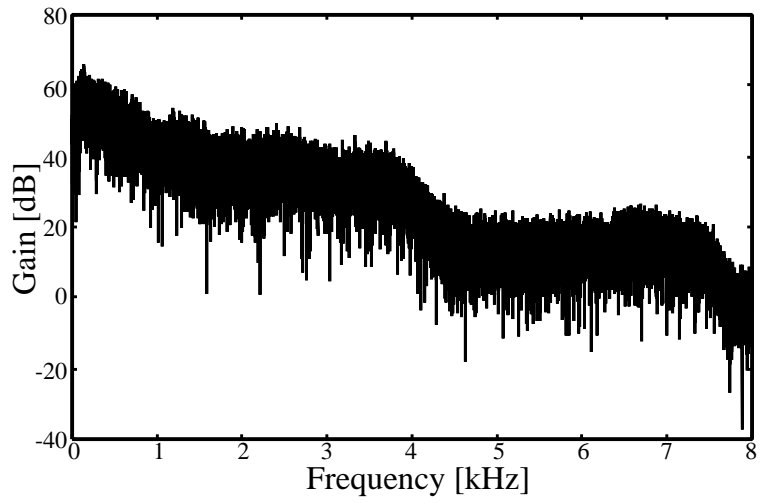


図 5.2 母音のスペクトルを重み付けした信号のスペクトル

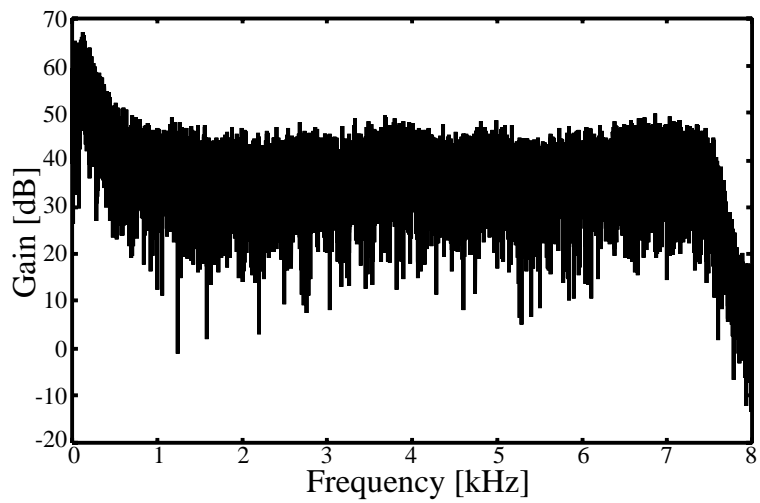


図 5.3 子音のスペクトルを重み付けした信号のスペクトル

## 5.3. 実環境評価実験

提案手法の性能と有効性を検証するために，残響時間の短い音響実験室において音声対雑音比 (Signal-to-Noise Ratio; SNR) の改善量，ビームフォーミング後の音声に対する歪み量，音声認識率 (孤立単語認識) に対する評価を行った．また，オフィスと同程度の残響時間と環境騒音である研究室内において音声認識率 (孤立単語認識) に対する評価を行った．

### 5.3.1 実験条件

音響実験室 (Room A) と研究室内 (Room B) の目的信号源，雑音源，マイクロホンアレーの関係を図 5.4, 5.5 に示す．また，実環境における収録条件を表 5.3 に，収録機材を表 5.4 に，実験条件を表 5.5 に示す．

目的信号と雑音は別々に収録を行い，SNR ごとに加算処理を行うことで観測信号を作成した．残響による音声明瞭度の劣化に関する検討を行うため，残響時間が非常に短い音響実験室 (Room A) とオフィス環境に近い実環境である研究室内 (Room B) において実験を行った．V/C 平均振幅スペクトルは男女不特定話者モデルである．

提案手法の有効性を示すため，実験では以下 (i) ~ (iv) の項目に対する評価を行った．

- (i) V/C セグメンテーションによる V/C 識別性能の評価実験 (Room A, Room B)
- (ii) SNR の改善量と信号の歪み量の評価 (Room A)
- (iii) 実験結果の波形とスペクトログラムの評価 (Room A)
- (iv) 音声認識性能の評価 (Room A, Room B)

(i) の V/C 識別率は，手動により作成した V/C・非発話区間ラベルを正解として，全フレーム数に対する V/C セグメンテーションの正解フレーム数の割合により V/C と非発話状態の 3 状態に対する識別率を算出した．また，V/C セグメンテーションは HTK[86] を利用した音素認識により行った．

(ii) の SNR の改善量は，ビームフォーミング前の SNR とビームフォーミング後の SNR の比により評価した．また，音声の歪み量は式 (5.7) の Spectral Distortion



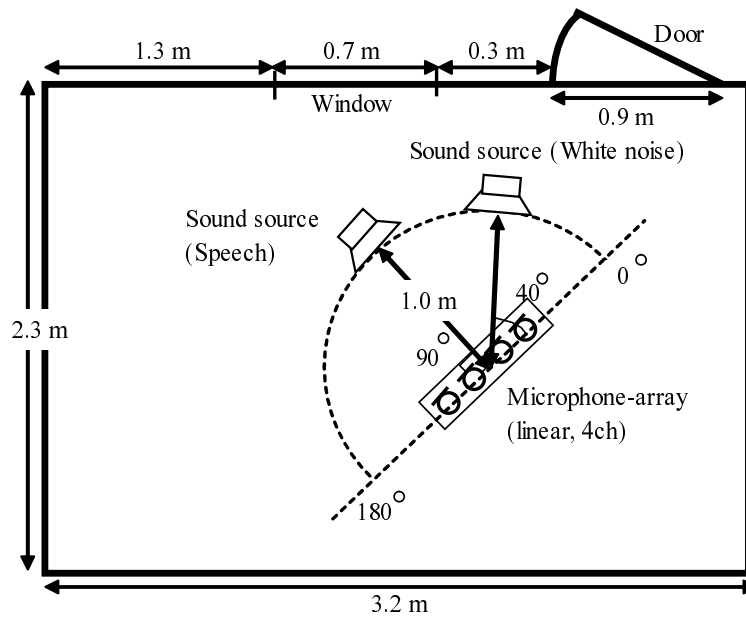


図 5.4 VC-AMNOR の評価実験における音源とマイクロホンアレーの配置 ( Room A )

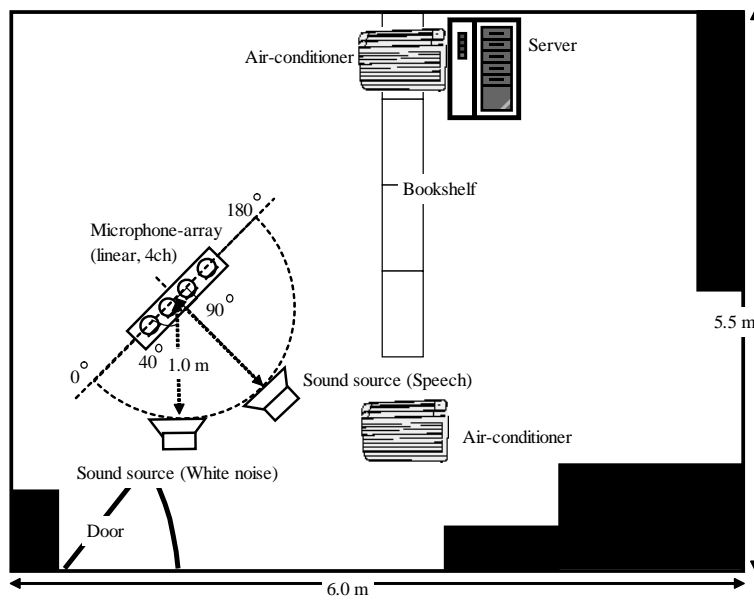


図 5.5 VC-AMNOR の評価実験における音源とマイクロホンアレーの配置 ( Room B )

表 5.3 VC-AMNOR の評価実験における収録条件

<b>Room A</b>	
Room size	2.3 m × 3.2 m
Reverberation time	$T_{[60]} = 0.12$ sec
Ambient noise level	17 dBA
<b>Room B</b>	
Room size	5.5 m × 6.0 m
Reverberation time	$T_{[60]} = 0.48$ sec
Ambient noise level	56 dBA
<b>Recording conditions</b>	
Temperature	21.5 °C
Humidity	46 %
Sampling frequency	16 kHz
Quantization	16 bits

表 5.4 VC-AMNOR の評価実験における収録機材

Microphone-array	Linear type 4 transducers
Microphone interval	2.125 cm
Microphone	HOSHIDEN, KUC-1333
Microphone amplifier	PAVEC, MA-2016B
Loudspeaker	MITSUBISHI, DIATONE DS-7
Speaker amplifier	YAMAHA, P4050
AD/DA converter	PAVEC, DF-2021/22

表 5.5 VC-AMNOR の評価実験における実験条件

<b>Parameter settings</b>	
Filter tap	61
Distortion	0, 0.01, 0.05, 0.1, 0.2, 0.3
Frame length	32 msec (Hanning window)
Frame interval	16 msec
<b>Test data (open)</b>	
Desired sound source	Phoneme balanced 216 words in ATR-DB [85] (2 females and 2 males)
Noise sound source	White Gaussian noise
SNR	-5, 0, 5, 10 dB
Acoustic model (HMM)	IPA monophone model [83] (Clean training and gender model)

measure ( SD ) [78] で評価した .

$$SD = \sqrt{\frac{10^2}{N} \sum_{\omega=1}^N (\log_{10} |S(\omega)| - \log_{10} |S'(\omega)|)^2}, \quad (5.7)$$

ここで,  $N$  は最大サンプル数,  $SD$  は信号の歪み量,  $S(\omega)$  はクリーンな音声のスペクトル,  $S'(\omega)$  はビームフォーミングされた音声のスペクトルである . また, 式 (2.17) の劣化許容量を 0.01, 0.2 ( 入力に対する出力の歪みを 1 %, 20 % の劣化まで許容 ) とした場合の SNR の改善量と  $SD$  を評価した .

(iii) の実験結果の波形とスペクトログラムは, 式 (2.17) の劣化許容量を 0.2 ( 入力に対する出力の歪みを 20 % の劣化まで許容 ), SNR を  $-5$  dB とした条件において評価した .

(iv) の音声認識性能を評価するために音声認識エンジン Julius-3.4 [77] を使用した . クリーンな音声に対して音声認識を行った場合の音声認識率は 96 % であった . 音声認識率は, 劣化許容量を変化させた場合の最高値を実験結果とした . 音声認識性能は言語モデルを利用せず, クリーン音声で学習した音響モデルと単語辞書を利用した孤立単語認識により評価した . この音声認識評価では雑音と音声の歪みにのみ依存して音声認識率が劣化するため, SNR と  $SD$  の両方の影響を音声認識率として評価可能である . さらに, 従来手法である音響モデルの適応学習との比較を行うため, AMNOR による音声への歪みを音響モデルに MLLR 適応 [76] した場合の音声認識率も評価した . MLLR 適応には HTK [86] を利用し, JNAS-DB の 20 話者の 20 文による教師なしの適応学習を行った . この MLLR 適応は, 上記のクリーン音声の音響モデルに対してのみ行ない, テスト環境下での適応は行っていない .

また, 実験では従来手法との比較のため, 以下の手法について評価を行った .

1. AMNOR : 基礎となる AMNOR
2. S-AMNOR : 従来手法である S-AMNOR
3. VC-AMNOR : 提案手法である VC-AMNOR
4. VC-AMNOR (with clean label) : 提案手法である VC-AMNOR において雑音下観測音声の V/C 区間が既知

表 5.6 音響実験室 ( Room A ) の V/C 識別率

SNR	-5 dB	0 dB	5 dB	10 dB
Correct V/C rate	75.2 %	79.4 %	81.9 %	83.9 %

表 5.7 研究室内 ( Room B ) の V/C 識別率

SNR	-5 dB	0 dB	5 dB	10 dB
Correct V/C rate	70.1 %	73.6 %	75.9 %	76.5 %

### 5.3.2 実験結果

#### V/C セグメンテーションによる V/C 識別性能評価の実験結果

表 5.6 に残響時間の短い音響実験室 ( Room A ) における V/C 識別率を示す . 表 5.6 にオフィス環境に近い実環境である研究室内 ( Room B ) における V/C 識別率を示す . 表 5.6 より , 残響時間の短い環境において V/C セグメンテーション性能は低 SNR において劣化する傾向にあるが , V/C 識別程度であれば -5 dB の環境下でも 75 % 程度の識別が可能であることがわかった . 表 5.7 より , オフィス環境に近い実環境においても V/C セグメンテーション性能は表 5.6 の場合と同様に低 SNR において劣化する傾向にあるが , V/C 識別程度であれば -5 dB の環境下でも 70 % 程度の識別が可能であることがわかった .

以上の結果より , 提案手法の前提条件である大雑把なビームフォーミングと音素認識でもある程度の V/C 正解精度を保てることが示唆される . また , 残響時間が長い環境における V/C 識別性能の劣化は , 残響により雑音に対する死角形成が困難となり , かつ目的音声の母音のエネルギーが非発話区間に後続したことが原因として考えられる . また , VC-AMNOR (with clean label) は V/C 識別率が 100 % である状態を示す . そのため , 後述する音声認識結果における VC-AMNOR と VC-AMNOR (with clean label) の性能差は V/C 識別精度の劣化が原因であると考えられる .

表 5.8 VC-AMNOR 評価実験の劣化許容量 0.01 における SNR の改善量

SNR	-5 dB	0 dB	5 dB	10 dB
AMNOR	20.2 dB	20.3 dB	20.3 dB	20.4 dB
S-AMNOR	16.1 dB	16.1 dB	16.1 dB	16.2 dB
VC-AMNOR	18.0 dB	18.0 dB	18.1 dB	18.1 dB
VC-AMNOR (with clean label)	19.0 dB	18.9 dB	19.0 dB	18.6 dB

表 5.9 VC-AMNOR 評価実験の劣化許容量 0.2 における SNR の改善量

SNR	-5 dB	0 dB	5 dB	10 dB
AMNOR	25.5 dB	25.6 dB	25.8 dB	26.1 dB
S-AMNOR	22.5 dB	22.6 dB	22.6 dB	22.7 dB
VC-AMNOR	24.9 dB	24.9 dB	25.0 dB	25.1 dB
VC-AMNOR (with clean label)	25.6 dB	25.6 dB	25.6 dB	25.8 dB

### SNR の改善量と SD の実験結果

表 5.8, 5.9 に SNR の改善量を示す．表 5.10, 5.11 にビームフォーミングされた音声の歪み量である SD を示す．表 5.8, 5.10 は式 (2.17) の劣化許容量を 0.01, 表 5.9, 5.11 は式 (2.17) の劣化許容量を 0.2 とした場合の SNR 改善量と SD である．

表 5.8, 5.10 より, 劣化許容量を小さく設定した場合には, SNR は基礎となる AMNOR に比べて改善していないが SD は改善する傾向があることがわかった．表 5.9, 5.11 より, 劣化許容量を大きく設定した場合には, SNR は基礎となる AMNOR と同程度であるが, SD は AMNOR に比べて大きく改善する傾向があることがわかった．また, 従来手法の S-AMNOR は基礎となる AMNOR に比べて SNR が劣化する傾向にあるが, 提案手法である VC-AMNOR は AMNOR と同程度の SNR を保持している．これらの結果より, 従来の S-AMNOR は基礎となる AMNOR に比べて SNR は劣化するが音声に対する歪みである SD を改善可能であり, VC-AMNOR は

表 5.10 VC-AMNOR 評価実験の劣化許容量 0.01 における SD

SNR	-5 dB	0 dB	5 dB	10 dB
AMNOR	3.0 dB	2.4 dB	1.9 dB	1.7 dB
S-AMNOR	2.7 dB	2.0 dB	1.7 dB	1.4 dB
VC-AMNOR	2.3 dB	1.8 dB	1.5 dB	1.3 dB
VC-AMNOR (with clean label)	2.3 dB	1.8 dB	1.5 dB	1.3 dB

表 5.11 VC-AMNOR 評価実験の劣化許容量 0.2 における SD

SNR	-5 dB	0 dB	5 dB	10 dB
AMNOR	8.1 dB	8.1 dB	8.2 dB	8.2 dB
S-AMNOR	2.5 dB	2.4 dB	2.3 dB	2.2 dB
VC-AMNOR	2.4 dB	2.3 dB	2.3 dB	2.3 dB
VC-AMNOR (with clean label)	2.4 dB	2.4 dB	2.4 dB	2.4 dB

表 5.12 VC-AMNOR の評価実験における音声認識結果

Clean training(%WRR)											
	Room A (0.12 sec)					Room B (0.48 sec)					Overall
	-5 dB	0 dB	5 dB	10 dB	Average	-5 dB	0 dB	5 dB	10 dB	Average	Average
AMNOR	81.3	87.3	88.9	90.0	86.9	57.7	72.4	80.4	85.2	73.9	80.4
S-AMNOR	84.3	88.1	89.4	90.1	88.0	60.5	74.9	82.4	85.9	75.9	82.0
VC-AMNOR	85.0	88.4	90.1	90.4	88.5	63.6	76.7	83.8	87.8	78.0	83.2
VC-AMNOR with clean label	86.7	89.2	90.5	90.6	89.2	70.2	79.6	84.7	88.7	80.8	85.0
Clean training and MLLR adaptation to AMNOR distortion (%WRR)											
	Room A (0.12 sec)					Room B (0.48 sec)					Overall
	-5 dB	0 dB	5 dB	10 dB	Average	-5 dB	0 dB	5 dB	10 dB	Average	Average
AMNOR	83.7	90.1	91.7	91.3	89.2	64.6	78.8	86.8	92.7	80.7	85.0
S-AMNOR	86.2	88.6	90.8	89.8	88.9	63.2	80.9	87.0	88.3	79.9	84.4
VC-AMNOR	86.8	91.2	91.0	91.2	90.1	65.4	77.6	85.6	90.6	79.8	84.9
VC-AMNOR with clean label	88.0	91.5	92.3	91.3	90.8	74.8	83.3	91.1	94.3	85.9	88.3

S-AMNOR に比べてさらに SD と SNR を改善していることがわかる。

#### 実験結果の波形とスペクトログラム

図 5.6 に劣化許容量を 0.2, SNR を  $-5$  dB とした場合の波形を, 図 5.7 に同条件下のスペクトログラムを示す。

図 5.6, 5.7 より, 劣化許容量を大きく設定した場合, 基礎となる AMNOR では低周波数帯域の信号が歪みにより大きく欠落しているのに対して, S-AMNOR, VC-AMNOR では低周波数帯域の信号(フォルマントなど)が保存されていることがわかった。図 5.6(e), 5.7(e)において一部の音声区間が大きく抑圧されている原因は, 音声(V/C)区間を背景雑音区間と誤認識したためである。

#### 音声認識性能評価の実験結果

表 5.12 に残響時間の短い音響実験室(Room A)とオフィス環境に近い実環境である研究室内(Room B)における音声認識率を示す。

表 5.12 の Room A の音声認識率より, 残響時間の短い環境において提案手法は従来手法よりも音声認識性能の改善が可能であり, 特に低 SNR において高い効果を発揮することがわかった。表 5.12 の Room B の音声認識率より, Room A の認識結果に比べてオフィス環境に近い実環境では残響の影響により全体的に音声認識率が低



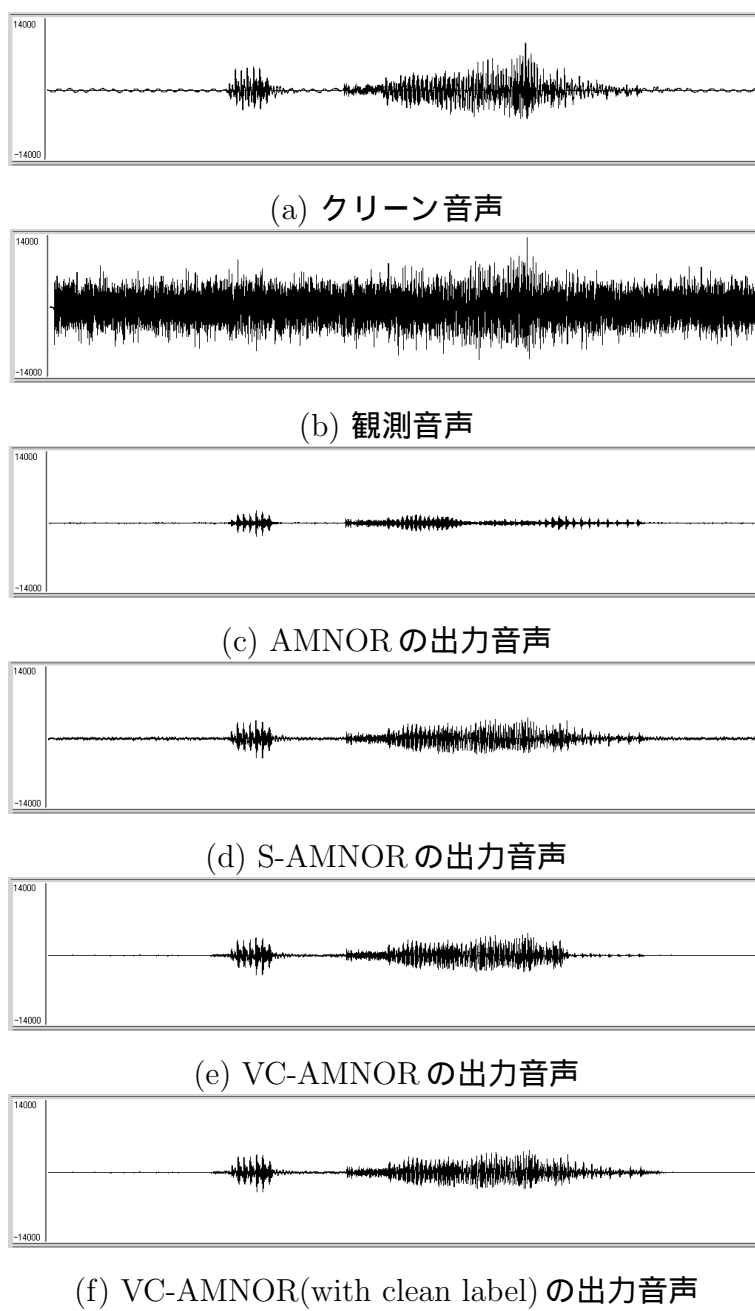
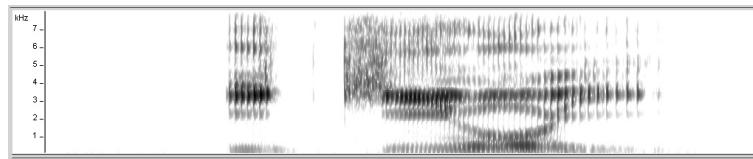
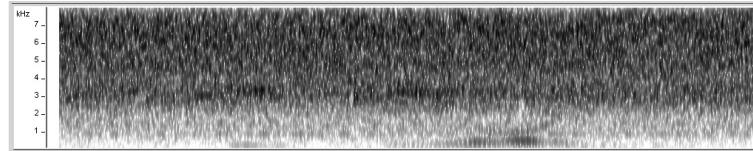


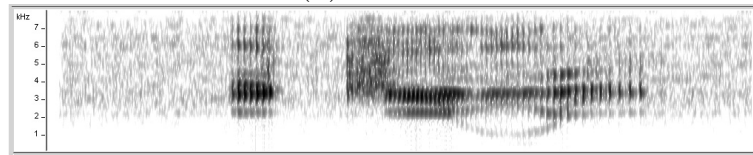
図 5.6 VC-AMNOR の評価実験における劣化許容量 0.2 , SNR  $-5$  dB の条件における波形



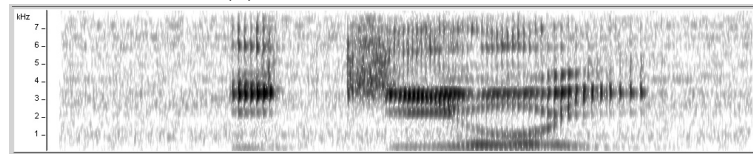
(a) クリーン音声



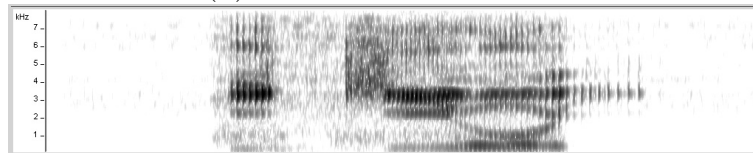
(b) 観測音声



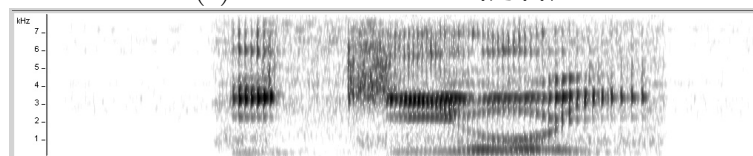
(c) AMNORの出力音声



(d) S-AMNORの出力音声



(e) VC-AMNORの出力音声



(f) VC-AMNOR(with clean label)の出力音声

図 5.7 VC-AMNOR の評価実験における劣化許容量 0.2 , SNR  $-5$  dB の条件におけるスペクトログラム

下しているが，提案手法は従来手法よりも音声認識性能の改善が可能であり，特に低 SNR において高い効果を発揮することがわかった．

以上の結果より，提案手法は特に高残響低 SNR の環境において有効であり，従来手法に比べて残響・雑音に対して頑健な手法であることが示唆される．VC-AMNOR (with clean label) は正確な V/C に対応したビームフォーミングと残響により非発話区間に後続した母音を非発話区間として抑圧できるため，高残響低 SNR 環境でも音声認識率の改善が可能であったと考えられる．VC-AMNOR は，VC-AMNOR (with clean label) と比べて音声認識性能は劣化しているが，表 5.6, 5.7 の V/C 識別結果のように高残響低 SNR 環境でも 7 割程度の V/C 識別精度があるため，従来手法と比べて音声認識率を改善できたと考えられる．

表 5.12 の AMNOR の歪みに対する MLLR 適応を行った音響モデルを利用した場合の音声認識結果では，VC-AMNOR (with clean label) 以外の手法は同程度の音声認識率となった．これは，AMNOR によるスペクトル歪みを MLLR 適応でも同様に軽減できることを意味している．また，AMNOR の歪みに対する MLLR 適応であるため，AMNOR の音声認識率が 5 %程度改善しているのに比べて，他の手法では 2~3%の改善となっている．しかし，VC-AMNOR (with clean label) は，AMNOR の歪みを MLLR 適応した音響モデルを利用した条件下においても，従来手法と比べて音声認識率を改善することができた．また，AMNOR の歪みに対する MLLR 適応と同様に S-AMNOR や VC-AMNOR による歪みに対して MLLR 適応を行った音響モデルを使用することで，それぞれの手法の音声認識率をさらに改善できる可能性がある．ただし，AMNOR, S-AMNOR, VC-AMNOR のいずれにおいても，雑音環境と話者の方位が変わることによってビームフォーミング後の音声に対する歪みの特性が変化するため，MLLR 適応は雑音環境や話者の方位が変わることに行う必要がある．そのため，話者の方位や雑音環境が大きく変動しない条件以外では MLLR 適応と組み合わせる手法は実現が難しいと考えられる．

以上の評価実験の結果，提案手法は SNR の改善量においては従来手法と同程度であるが，音声に対する歪みを軽減することで音声認識率を改善していることが確認できた．この傾向は従来手法である S-AMNOR において顕著であり，S-AMNOR では基礎となる AMNOR に比べて SNR が劣化しているにもかかわらず SD と音声認

識率は改善する傾向があった。提案手法は、基礎となる AMNOR と同程度の SNR であり、かつ SD をさらに改善したため、従来手法である S-AMNOR 以上に音声認識率を改善できたと考えられる。

## 5.4. まとめ

雑音下音声認識の性能改善のために，本論文では適応型アレーの1つとして提案されている AMNOR を利用し，V/C に特化した適応型アレーを構築することにより雑音下音声認識性能の改善した VC-AMNOR を提案した．提案手法は雑音下観測音声の V/C 時系列に対して正確な V/C ビームフォーミングの適用が必要不可欠であるため，音素認識による V/C セグメンテーションの検討を行った．実環境における評価実験の結果，提案手法は従来手法よりも音声認識性能を改善できたことから，その有効性を確認できた．

提案手法では，V/C セグメンテーションにクリーン音響モデルを利用しているが，音響モデルに対する適応学習 [88, 89] や雑音を考慮した音響モデル [90, 91] の利用により，さらに V/C 区間のセグメンテーション性能を向上することができれば，提案手法の上限値である VC-AMNOR (with clean label) の性能に近づける可能性が高い．そこで今後は，V/C セグメンテーションの性能向上による提案手法の改善を検討したい．また，提案手法の V/C セグメンテーションの前処理として，AMNOR を利用しているが，前処理の音声にはスペクトル歪みが発生しても問題にはならないため，スペクトルサブトラクション [61] などを利用することを検討したい．

## 第6章 結論

### 6.1. 本論文のまとめ

これまでのPCなどの情報端末を積極的に利用していなかった人々がストレスなく情報機器を利用するには、音声入力インタフェースが適しており、利用者が音声入力時にマイクロホンを気にせず、どこで発話された音声でも高品質に受音・認識可能なハンズフリー音声インタフェースの実現は急務である。しかし、実環境では、環境雑音や室内残響の影響で、遠隔発話音声は歪みを受けて劣化してしまう。そこで、本論文ではマイクロホンアレーを用いて遠隔発話音声を高品質に受音する手法を検討した。

マイクロホンアレーを用いて遠隔発話音声を高品質に受音するためには、まず、話者位置を推定する必要がある。その話者の位置に対してのみ高感度の指向特性として、それ以外の雑音の方向には感度の死角を形成するビームフォーミングを行うことで、遠隔発話音声を高品質に受音できる。まず、2章では、マイクロホンアレーを用いた遠隔発話音声受音の基礎として、マイクロホンアレー受音の原理と話者位置推定に利用されるDOA推定手法の一つであるCSP法の詳細を述べ、その問題点を示した。また、ビームフォーミングの従来手法である遅延和アレーと適応型アレーの原理と提案手法の基となるAMNOR, S-AMNORの詳細について述べ、その問題点を示した。これらの問題を解決するために、本論文では以下の手法から構成されるマイクロホンアレーを用いた高品質音声受音手法を提案した。

3章では、可聴音の直接波と反射波の位相干渉に基づく距離推定法(以後、単に音響測距法と表記する)を雑音環境下においても頑健に動作するように理論拡張し、この測距法に適した雑音抑圧手法を提案した。この音響測距法は、対人・対物センサとして必要不可欠である。提案手法は、1つのスピーカと1つのマイクロホンを

利用して、測定音の計測から対象物、すなわち話者までの距離を測定する技術である。ただし、距離の測定には、対象物が存在ある/ない場合の測定が必要である。そこで、このバックグラウンド要素除去処理( Back-Ground Components Cancellation Processing; BGCCP )に同期加算とスペクトルサブトラクションの考え方を導入することで低 SNR 環境でも測距可能な手法を検討した。実環境において雑音環境を考慮した評価実験を行った結果、提案する音響測距法が低 SNR の雑音環境においても有効であることが明らかとなった。

4章では、3章の研究成果である雑音に頑健な音響測距法の受音用マイクロホンマイクロホンアレーに拡張することで、話者(対象物)までの距離だけでなくその位置まで推定する手法を提案した。また、話者の発話/非発話によって、話者の発話を利用する CSP 法による音源方位推定と提案手法を切り替えて適用する手法を検討した。実環境における評価実験の結果、30~150度の範囲では話者の発話区間では CSP 法によって約3度の誤差、非発話区間では提案手法によって約5度の誤差で方位推定が可能であり、非発話区間では提案手法によって約0.02 mの誤差で距離推定が可能であることがわかった。

最後に5章では、4, 3章で提案した話者位置推定の結果を利用して、雑音下音声認識の性能改善のために、適応型アレーの1つとして提案されている AMNOR を利用し、母音/子音( Vowel/Consonant; V/C )に特化した適応型アレーを構築することにより雑音下音声認識性能の改善した VC-AMNOR を提案した。提案手法は雑音下観測音声の V/C 時系列に対して正確な V/C ビームフォーミングの適用が必要不可欠であるため、音素認識による V/C セグメンテーションの検討を行った。実環境における評価実験の結果、提案手法は従来手法よりも音声認識性能を改善できたことから、その有効性を確認できた。

## 6.2. 今後の課題

本論文で提案した手法によって遠隔発話音声を高品質に受音できることを確認した。しかし、実環境の高雑音・残響下で頑健に遠隔発話音声受音を行い、ハンズフリー音声インタフェースとして利用するためには、以下の検討すべき点が存在する。

1. 音響測距法に適用する雑音抑圧手法の検討
2. 発話・非発話を問わない話者位置推定の検討
3. 高雑音環境下におけるさらなる音声受音品質向上の検討

1. の検討課題に対しては，独立成分分析( Independent Component Analysis; ICA ) [70, 71] などを利用した位相干渉成分の抽出手法など他の雑音抑圧手法の検討を今後行いたい．

2. の検討課題に対しては，非可聴域に近い高い周波数の送信波を用いて発話区間でも提案手法による話者位置推定を行えるように拡張したい．

3. の検討課題に対しては，V/C セグメンテーション性能の改善によって，提案手法の上限値である VC-AMNOR (with clean label) の性能に近づけたい．また，音素を母音/子音に分類するのではなく，音素のスペクトル包絡を考慮した分類を行い，ビームフォーミングする手法を検討したい．

以上の検討を行い，提案手法の改善を行うことで，より高品質な音声受音を実現，音声認識の性能を向上により，テレビ会議システムや音声入力インタフェースなどに実用化されることを期待している．また，本論文で提案した手法は，ハンズフリー音声インタフェース以外にも様々な分野に応用可能である．音響測距法を応用することで，ロボットセンサにおける物体の位置推定や話者検出にも役立つと考えられる．

これらの研究によって，誰でも ICT の利便性を実感できる社会実現の一助となり，僅かなりとも社会貢献できれば幸いである．



## 謝辞

本博士論文は、著者が立命館大学大学院理工学研究科博士後期課程在籍時において行った研究の成果と近畿大学生物理工学部において現在行っている研究成果の一部をまとめたものです。本研究を遂行するにあたり、様々な方々にお世話になり、また現在もお世話になっております。ここに感謝の意を示したいと思います。

立命館大学情報理工学部山下洋一教授には、著者が本大学院在籍中、指導教員として終始懇切なる御指導、御助言を頂きました。山下先生の熱心な御指導と的確な御助言があったからこそ、著者は大学院在籍時から満了後を経て現在に至るまで研究活動を継続し、本博士論文の執筆にまで辿り着けたと実感しております。ここに深甚なる感謝の意を表します。

同学部樋口宜男教授には、本博士論文の執筆において御指導とご助力を頂きました。樋口先生から頂いた有益な御指摘のおかげで本論文の完成度を高めることができました。心より深く御礼申し上げます。

同学部西浦敬信准教授には、著者の本大学院在籍中だけではなく、博士前期課程の大学院生として和歌山大学大学院に在籍していた頃より、研究活動への取り組み方から実験機材の使用方法に至るまで様々な院生として必要不可欠な知識と姿勢を真摯に御指導いただきました。また、本博士論文の執筆にあたっては、多くの御指導、御鞭撻を頂きました。ここに心から深く感謝の意を表します。

和歌山大学システム工学部河原英紀教授、同学部入野俊夫教授には、著者が和歌山大学大学院システム工学研究科在籍時に指導教員として熱心な御指導を頂きました。河原先生と入野先生の御指導と御助言があればこそ、研究者としての現在の自分があると言っても過言ではありません。ここに厚く御礼申し上げます。また、同学部坂野秀樹助手（現在、名城大学准教授）、同大学高橋徹研究員（現在、京都大

学助教)には、著者が和歌山大学大学院在籍時に、様々な御指導、御討論を頂きました。心より御礼申し上げます。著者の在籍時には直接御指導いただけていませんが、著者の修了後に同学部に着任された西村竜一助教には、研究会などで、的確な御指摘と御討論を頂きました。深く感謝いたします。

立命館大学情報理工学部趙國助手には、音声認識に関する多くの御助言を頂きました。厚く御礼申し上げます。

和歌山大学在籍時から立命館大学在籍時まで、同大学研究室に在籍していた傳田遊亀博士(現在、村田機械株式会社)には、研究遂行の上で様々な御討論を頂きました。深く感謝いたします。

本研究の遂行にあたり、旧情報処理学会雑音下音声認識評価ワーキンググループの皆様には数々の有益な御助言を頂きました。心より御礼申し上げます。

近畿大学生物理工学部中迫昇教授には、著者の同大学学部在籍時から著者が同大学非常勤講師として現在に至るまで、熱心かつ厳格な御指導と御助言を頂きました。ここに心から深く感謝の意を表します。

ここでは個々には御名前は省略させていただきますが、著者の研究活動にご助力いただいた立命館大学情報理工学部音情報処理研究室、和歌山大学システム工学部聴覚メディア研究室、近畿大学生物理工学部音響信号処理研究室の多くの先輩、同期、後輩、秘書の皆様は心より御礼申し上げます。

最後に、本博士論文の執筆のみならず、常日頃より著者を支えてくれる家族に心から感謝いたします。

## 参考文献

- [1] J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, “Computer-steered Microphone Arrays for Sound Transduction in Large Rooms,” *Journal of the American Statistical Association.*, **JASA**, vol.78, no.5, pp.1508–1518, Nov. 1985.
- [2] O.L. Frost, “An Algorithm for Linearly Constrained Adaptive Array Processing,” *Proc. IEEE*, vol.60, no.8, pp.926–935, Aug. 1972.
- [3] L.J. Griffiths, C.W. Jim, “An Alternative Approach to Linearly Constrained Adaptive Beamforming,” *IEEE. Trans. Antennas Propag.*, vol.AP-30, no.1, pp.27–34, Jan. 1982.
- [4] Y. Kaneda and J. Ohga, “Adaptive Microphone-array system for NOise Reduction,” *IEEE Trans. Acoust. Speech Signal Proc.*, vol.ASSP-34, no.6, pp.1391–1400, Dec. 1986.
- [5] 金田 豊, “適応形雑音抑圧マイクロホンアレー (AMNOR) の指向特性,” *日本音響学会誌*, vol.44, no.1, pp.23–30, Jan. 1988.
- [6] D. Johnson and D. Dudgeon, *Array signal processing: Concepts and Techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [7] G.W. Elko, “A Steerable and Variable First-order Differential Microphone Array,” *Proc. 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1997)*, vol.1, pp.223–226, Apr. 1997.

- [8] J.S. Hammerschmidt, C. Brunner, and C. Drewes, “Eigenbeamforming - a Novel Concept in Array Signal Processing,” *Proc. European Wireless Conference 2000*, Sept. 2000.
- [9] T. Nishiura, S. Nakamura, Y. Okada, T. Yamada, and K. Shikano, “Suitable Design of Adaptive Beamformer Based on Average Speech Spectrum for Noisy Speech Recognition,” *Proc. 7th International Conference on Spoken Language Processing (ICSLP 2002)*, pp.1789–1792, Sept. 2002.
- [10] Y. Okada, T. Nishiura, S. Nakamura, T. Yamada, and K. Shikano, “A Design of Adaptive Beamformer Based on Average Speech Spectrum for Noisy Speech Recognition,” *Acoustical Science and Technology*, vol.23, no.6, pp.323–327, Nov. 2002.
- [11] M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, “Microphone Array Based Speech Recognition With Different Talker-array Positions,” *Proc. 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1997)*, vol.1, pp.227–230, Apr. 1997.
- [12] E. Lleida, J. Fernandez, and E. Masgrau, “Robust Continuous Speech Recognition System Based on a Microphone Array,” *Proc. 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1998)*, vol.1, pp.241–244, May 1998.
- [13] T. Yamada, S. Nakamura, and K. Shikano, “Hands-free Speech Recognition Based on 3-D Viterbi Search Using a Microphone Array,” *Proc. 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1998)*, vol.1, pp.245–248, May 1998.
- [14] 高橋 祐, 猿渡 洋, 鹿野 清宏, “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, Mar. 2010.

- [15] 松井 俊浩, 麻生 英樹, John Fry, 浅野 太, 本村 陽一, 原 功, 栗田 多喜夫, 速水 悟, 山崎 信行, “オフィス移動ロボット Jijo-2 の音声対話システム,” 日本ロボット学会誌, vol.18, no.2, pp.300–307, Mar. 2000.
- [16] Y. Tamai, S. Kagami, Y. Amemiya, Y. Sasaki, H. Mizoguchi, and T. Takano, “Circular Microphone Array for Robot’s Audition,” *Proc. IEEE 2004 Sensors*, vol.2, pp.565–570, Oct. 2004.
- [17] S. Yamamoto, J.M. Valin, K. Nakadai, J. Rouat, F. Michaud, T. Ogata, H.G. Okuno, “Enhanced Robot Speech Recognition Based on Microphone Array Source Separation and Missing Feature Theory,” *Proc. 2005 IEEE International Conference on Robotics and Automation (ICRA 2005)*, pp.1477–1482, Apr. 2005.
- [18] G.W. Elko, “Microphone Array Systems for Hands-free Speech Telecommunication,” *Speech Communication*, vol.20, pp.229–240, Dec. 1996.
- [19] 小林 和則, 古家 賢一, 羽田 陽一, 片岡 章俊, “方向別自動音量調整マイクロホンアレー,” 電子情報通信学会論文誌 A, vol.J87-A, no.12, pp.1491–1501, Dec. 2004.
- [20] W. Herbordt, T. Horiuchi, M. Fujimoto, T. Jitsuhiro, and S. Nakamura, “Hands-free Speech Recognition and Communication on PDAs Using Microphone Array Technology,” *Proc. 2005 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2005)*, pp.302–307, Nov. 2005.
- [21] Y. Denda, T. Nishiura, and Y. Yamashita, “Omnidirectional Audio-Visual Talker Localization Based on Dynamic Fusion of Audio-Visual Features Using Validity and Reliability Criteria,” *IEICE Trans on Information and Systems*, vol.E91-D, no.3, pp.598–606, Mar. 2008.
- [22] 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,” 電子情報通信学会論文誌 D-II, vol.83-D-II, no.8, pp.1713–1721, Aug. 2000.

- [23] J. Flanagan, D. Berkley, G.W. Elko, J. West, M. Sondhi, “Autodirective microphone systems,” *Acustica*, vol.73, no.2, pp.58–71, 1991.
- [24] 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, 東京, 1995.
- [25] 菊池 慶子, 醍醐 徹, 中島 弘史, 中臺 一博, 長谷川 雄二, 金田 豊, “大規模マイクロホンアレイによる発話方向推定の検討,” 電子情報通信学会技術研究報告 [応用音響] EA2008-39, vol.108, no.143, pp.13–18, Jul. 2008.
- [26] H. Saruwatari, S. Kajita, K. Takeda, F. Itakura, “Speech Enhancement Using Nonlinear Microphone Array Based on Noise Adaptive Complementary Beamforming,” *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol.E83-A, no.5, pp.866–876, May 2000.
- [27] N. Ohmata, T. Uebo, N. Nakasako, and A. Mori, “Distance measurement by audible sound (Utilization of band-limited signal with random phase),” *Proc. of The China-Japan Joint Conference of Acoustics 2007*, pp.1–6, Jun. 2007.
- [28] N. Nakasako, T. Uebo, N. Ohmata, and A. Mori, “On Distance Measurement Using Band-limited Noise,” *Proc. of SICE Annual Conference 2007*, pp.151–156, Sept. 2007.
- [29] N. Nakasako, T. Uebo, A. Mori, and N. Ohmata, “Fundamental Consideration on Distance Estimation Using Acoustical Standing Wave,” *IEICE Trans. Fundamentals*, vol.E91-A, no.4, pp.1218–1221, Apr. 2008.
- [30] 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,” 電気学会論文誌 C, vol.128, no.7, pp.1117–1122, Jul. 2008.
- [31] 土田 悠太, 中迫 昇, 上保 徹志, 篠原 寿広, “定在波に基づく距離推定法とその実装に関する一考察: Chirp 音信号の利用,” 電子情報通信学会技術研究報告 [応用音響] EA2008-121, vol.108, no.411, pp.29–34, Jan. 2009.

- [32] T. Uebo, N. Nakasako, N. Ohmata, and A. Mori, “Distance Measurement Based on Standing Wave for Band-limited Audible Sound With Random Phase,” *Acoustical Science and Technology*, vol.30, no.1, pp.18–24, Jan. 2009.
- [33] 大亦 紀光, 上保 徹志, 中迫 昇, 篠原 寿広, “可聴音域の定在波に基づく距離推定法の実装に関する一試み,” *電気学会論文誌 C*, vol.129, no.2, pp.314–319, Feb. 2009.
- [34] N. Nakasako, S. Hanabusa, T. Shinohara, and T. Uebo, “Distance Estimation and Compensation Method Using Band-limited Transmitted Wave with Uniform Amplitude and Random Phase Characteristics,” *Proc. 2009 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2009)*, pp.580–583, Jul. 2009.
- [35] S. Hanabusa, N. Nakasako, T. Uebo, and T. Shinohara, “Correction Method of Estimated Distance Using Geometrical Information on Sound Source and Microphone,” *Proc. 2009 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009)*, pp. 237–240, Oct. 2009.
- [36] 中迫 昇, 英 慎平, 篠原 寿広, 上保 徹志, “音声による対象物までの距離推定の試み (定在波理論の利用),” *日本音響学会 2010 年春季研究発表会講演論文集*, pp.785–786, Mar. 2010.
- [37] 盛田 章, “超指向性マイクロホン,” *日本音響学会誌*, vol.51, no.5, pp.395–399, May 1995.
- [38] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol.57, no.8, pp.1408–1418, Aug. 1969.
- [39] 安倍 正人, “多数センサによる音源推定,” *日本音響学会誌*, vol.51, no.5, pp.384–389, May 1995.

- [40] H.P. Bucker, “High-resolution Cross-sensor Beamforming for a Uniform Line Array,” *Journal of the Acoustical Society of America*, vol.63, no.2, pp.420–424, Aug. 1978.
- [41] D.P. Skinner, “Maximum Entropy Array Processing,” *Journal of the Acoustical Society of America*, vol.66, no.2, pp.488–493, Aug. 1978.
- [42] R.O. Schmidt, “Multiple Emitter Location and Signal Parameter Estimation,” *IEEE Transactions on Antennas and Propagation*, vol.34, no.3, pp.276–280, Mar. 1986.
- [43] S.U. Pillai, “Array Signal Processing,” Springer-Verlag, New York, 1989.
- [44] C. Knapp and G. Carter, “The Generalized Correlation Method for Estimation of Time Delay,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.24, no.4, pp.320–327, Aug. 1976.
- [45] M. Omologo and P. Svaizer, “Use of The Crosspower-spectrum Phase in Acoustic Event Location,” *IEEE Transactions on Speech and Audio Processing*, vol.5, no.3, pp.288–292, May 1997.
- [46] J. Markel and A. Gray, “On Autocorrelation Equations as Applied to Speech Analysis,” *IEEE Transactions on Audio and Electroacoustics*, vol.21, no.2, pp.69–79, Apr. 1973.
- [47] I. Yamada and N. Hayashi, “Improvement of The Performance of Cross Correlation Method for Identifying Aircraft Noise With Pre-whitening of Signals,” *Journal of the Acoustical Society of Japan (E)*, vol.13, no.4, pp.241–252, Jul. 1992.
- [48] M. Parrilla, J.J. Anaya, and C. Fritsch, “Digital Signal Processing Techniques for High Accuracy Ultrasonic Range Measurements,” *IEEE Trans. Instrum. Means.*, vol.40, no.4, pp.759–763, Aug. 1991.



- [49] D. Marioli, C. Narduzzi, C. Offelli, D. Petri, E. Sardini, and A. Taroni, “Digital Time-of-Flight Measurement for Ultrasonic Sensors,” *IEEE Trans. Instrum. Means.*, vol.41, no.1, pp.93–97, Feb. 1992.
- [50] M. Yang, S.L. Hill, B. Bury, and J.O. Gray, “A Multifrequency AM-Based Ultrasonic System for Accuracy Distance Measurement,” *IEEE Trans. Instrum. Means.*, vol.43, no.6, pp.861–866, Dec. 1994.
- [51] 奥雲 正樹, 木村 昭彦, 大木 誠, 大北 正昭, “高機能超音波センサシステムの開発研究,” *電気学会論文誌 C*, vol.128. no.1, pp.55–61, Jan, 2008.
- [52] 伊藤 健一, “超音波のはなし,” 日刊工業新聞社, 1992.
- [53] 上保 徹志, “0 mから測距可能な定在波レーダ,” *電気学会論文誌 C*, vol.125, no.11, pp.1646–1651, Nov. 2005.
- [54] T. Uebo, Y. Okubo, and T. Iritani, “Standing Wave Radar Capable of Measuring Distances down to Zero Meters,” *IEICE Trans. Commun.*, vol.E88-B, no.6, pp.2609–2615, Jun. 2005.
- [55] T. Uebo, T. Kitagawa, and T. Iritani, “Short Range Radar Utilizing Standing Wave of Microwave or Millimeter Wave,” *Proc. IEEE Intelligent Vehicles Symposium 2001*, pp.95–99, May 2001.
- [56] S.F. Boll and D.C. Pulsipher, “Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation,” *IEEE Trans. Acoust. Speech Signal Process.*, vol.28, no.6, pp.751–753, Dec. 1980.
- [57] J.S. Lim and A.V. Oppenheim, “All-pole Modeling of Degraded Speech,” *IEEE Trans. Acoust. Speech Signal Process.*, vol.26, no.3, pp.197–210, Jun. 1978.
- [58] E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Transactions of the American Society of Mechanical Engineers, Journal of Basic Engineering*, vol.82, no.3, pp.35–45, Mar. 1977.

- [59] 有本 卓, “カルマン・フィルター,” 産業図書, 1977.
- [60] 田邊 造, 小野塚 康浩, 田畑 雅崇, 古川 利博, 辻井 重男, “有色性駆動源を持つカルマンフィルタを用いた高速な雑音抑圧法,” 電子情報通信学会技術研究報告 [応用音響] EA2008-137, vol.108, no.491, pp.7–11, 2009.
- [61] S.F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” *IEEE Trans. Acoust. Speech Signal Process.*, vol.27, no.2, pp.133–120, Apr. 1979.
- [62] 金田 豊, “M 系列を用いたインパルス応答測定における誤差の実験的検討,” 日本音響学会誌, vol.52, no.10, pp.752–759, Oct. 1996.
- [63] Y. Suzuki, F. Asano, H.Y. Kim, and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, vol.97, no.2, pp.1119–1123, 1995.
- [64] Y. Kuroiwa and T. Shimamura, “An Improvement of LPC Based on Noise Reduction Using Pitch Synchronous Addition,” *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems (ISCAS’99)*, vol.3, pp.122–125, Jun. 1999.
- [65] D. Hochmann, M. Sadok, “Theory of synchronous averaging,” *Proc. 2004 IEEE Aerospace Conference*, vol.6, pp.3636–3653, Mar. 2004.
- [66] S. Zhang, Y. Li, X. Li, Z. Yang, and L. Zhao, “System of Synchronous Averaging De-noising Based on Labview,” *Proc. 2008 International Conference on MultiMedia and Information Technology (MMIT’08)*, pp.446–448, Dec. 2008.
- [67] 旭 健作, 三好 史泰, 小川 明, “音響による接近車両検出における自車エンジンの低減手法について,” 電子情報通信学会論文誌 A, vol.J91-A, no.1, pp.68–77, Jan. 2008.

- [68] H.M. Goodarzi and S. Seyedtabaii, “Application of Spectral Subtraction for Reducing Industrial Noises,” *Proc. 6th International Symposium on Image and Signal Processing and Analysis (ISPA 2009)*, pp.79–83, Sept. 2009.
- [69] L. Gao, Y. Guo, S. Li, and F. Chen, “Speech Enhancement Algorithm Based on Improved Spectral Subtraction,” *Proc. 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2009)*, vol.3, pp.140–143, Nov. 2009.
- [70] 中迫 昇, 小倉 久直, “独立成分分析の基礎と音響信号処理,” システム制御情報学会誌, vol.46, no.7, pp.400–408, 2002.
- [71] H. Park, S. Oh, and S. Lee, “Adaptive Noise Cancelling Based on Independent Component Analysis,” *IEEE Electronics Letters*, vol.38, no.15, pp.832–833, Jul. 2002.
- [72] 樋口 知之, “粒子フィルタ,” 電子情報通信学会誌, vol.88, no.12, pp.980–994, Dec. 2005.
- [73] 實廣 貴敏, 鳥山 朋二, 小暮 潔, “実環境下音声認識のためのパーティクルフィルタを統合した複数モデル雑音抑圧手法,” 電子情報通信学会論文誌 D, vol.J91-D, no.10, pp.2519–2528, Oct. 2008.
- [74] 伊藤 俊夫, 杉本 雅則, 橋爪 宏達, “最適化したマルチキャリア信号と合成送信開口による高画質音響イメージング,” 電子情報通信学会論文誌 A, vol.J93-A, no.5, pp.341–352, May 2010.
- [75] 中山 雅人, 定野 一也, 篠原 寿広, 中迫 昇, “定在波を利用した距離推定とマイクロホンアレーに基づく話者位置推定の検討,” 日本音響学会 2009 年春季研究発表会, pp.763–764, Mar. 2009.
- [76] C.J. Leggetter and P.C. Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models,” *Computer Speech and Language*, pp.171–185, Apr. 1995.

- [77] A. Lee, T. Kawahara, and K. Shikano, “JULIUS – An Open Source Real-Time Large Vocabulary Recognition Engine,” *Proc. 2001 European Conference on Speech Technology (Eurospeech 2001)*, pp.1691–1694, Sept. 2001.
- [78] 古井 貞熙, “電子情報通信工学シリーズ 音声信号処理,” 森北出版株式会社, 東京, 1998.
- [79] 鈴木 久喜, “音声のデジタル信号処理(上),” コロナ社, 東京, 1983.
- [80] 古井 貞熙, “音響・音声工学,” 近代科学社, 東京, 1992.
- [81] J.S. Lim and A.V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” *Proceedings of the IEEE*, vol.67, no.12, pp.1586–604, 1979.
- [82] International Phonetic Association, “Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet,” Cambridge Univ Pr (Sd), Aug. 1999.
- [83] 鹿野 清宏, 伊藤 克亘, 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.
- [84] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, and K. Shikano, “JNAS:Japanese Speechcorpus for Large Vocabulary Continuous Speech Recognition Research,” *Journal of Acoustic Society of Japan*, vol.20, no.3, pp.199–206, 1999.
- [85] K. Takeda, Y. Sagisaka, and S. Katagiri, “Acoustic-Phonetic Labels in a Japanese Speech Database,” *Proc. 1987 European Conference on Speech Technology (Eurospeech 1987)*, vol.2, pp.13–16, Oct. 1987.
- [86] <http://htk.eng.cam.ac.uk/>
- [87] 中川 聖一, “確率・統計的手法による音声認識,” *日本音響学会誌*, vol.50, no.2, pp.126–132, 1994.

- [88] 松浪 加奈子, 芳澤 伸一, 馬場 朗, 李 晃伸, 猿渡 洋, 鹿野 清宏, “十分統計量を用いた教師なし話者適応および環境適応,” 情報処理学会論文誌, vol.43, no.7, pp.2038–2045, Jul. 2002.
- [89] 山出 慎吾, 馬場 朗, 芳澤 伸一, 李 晃伸, 猿渡 洋, 鹿野 清宏, “実環境における頑健な音声認識のための音韻モデルの教師なし話者適応,” 電子情報通信学会論文誌 D-II, vol.J87-D-II, no.4, pp.933–941, Apr. 2004.
- [90] 藤村 浩司, 宮島 千代美, 伊藤 克巨, 武田 一哉, “実環境における SNR 別音響モデルの評価,” 電子情報通信学会技術研究報告 [音声] SP2004-152, vol.104, no.631, pp.43–48, Jan. 2005.
- [91] 鈴木 基之, 加藤 裕介, 伊藤 彰則, 牧野 正三, “SN 比に頑健なマルチミクスチャー HMM の性能評価,” 電子情報通信学会技術研究報告 [音声] SP2005-31, vol.105, no.133, pp.25–30, Jun. 2005.

# 研究業績

## 学術論文

1. 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” 電気学会論文誌 C, vol.130-C, no.11, pp.1994–2000, Nov. 2010.
2. 中山 雅人, 西浦 敬信, 山下 洋一, “母音/子音特徴量に基づく適応形マイクロホンアレーを用いた雑音下音声認識,” 電子情報通信学会論文誌 D, vol.J92-D, no.9, pp.1568–1578, Sept. 2009.
3. Norihide Kitaoka, Takeshi Yamada, Satoru Tsuge, Chiyomi Miyajima, Kazumasa Yamamoto, Takanobu Nishiura, **Masato Nakayama**, Yuki Denda, Masakiyo Fujimoto, Tetsuya Takiguchi, Satoshi Tamura, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, “CENSREC-1-C: An evaluation framework for voice activity detection under noisy environments,” *Acoustical Science and Technology*, vol.30, no.5, pp.363–371, Sept. 2009.
4. 傳田 遊亀, 田中 貴雅, 溝口 遊, 中山 雅人, 西浦 敬信, 山下 洋一, “話者方位推定を利用した動的な時間領域処理に基づく遠隔発話区間検出,” 電子情報通信学会論文誌 D, vol.J92-D, no.1, pp.112–122, Jan. 2009.

## 国際会議

5. **Masato Nakayama**, Shimpei Hanabusa, Noboru Nakasako, and Tetsuji Uebo, “Robust Acoustic Distance Measurement Method Based on Interference in

- Noisy Environments,” *Proc. 2010 10th International Symposium on Communications and Information Technologies 2010 (ISCIT 2010)*, Oct. 2010 (accepted).
6. Satoshi Tamura, Chiyomi Miyajima, Norihide Kitaoka, Takeshi Yamada, Satoru Tsuge, Tetsuya Takiguchi, Kazumasa Yamamoto, Takanobu Nishiura, **Masato Nakayama**, Yuki Denda, Masakiyo Fujimoto, Shigeki Matsuda, Tetsuji Ogawa, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, “CENSREC-1-AV: An Audio-visual Corpus for Noisy Bimodal Speech Recognition,” *Proc. 2010 International Conference on Auditory-Visual Speech Processing (AVSP 2010)*, Oct. 2010 (accepted).
  7. **Masato Nakayama**, Takanobu Nishiura, Yuki Denda, Norihide Kitaoka, Kazumasa Yamamoto, Takeshi Yamada, Satoru Tsuge, Chiyomi Miyajima, Masakiyo Fujimoto, Tetsuya Takiguchi, Satoshi Tamura, Tetsuji Ogawa, Shigeki Matsuda, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, “CENSREC-4: Development of Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments,” *Proc. INTERSPEECH 2008*, pp.968–971, Sep. 2008.
  8. Takanobu Nishiura, **Masato Nakayama**, Yuki Denda, Norihide Kitaoka, Kazumasa Yamamoto, Takeshi Yamada, Satoru Tsuge, Chiyomi Miyajima, Masakiyo Fujimoto, Tetsuya Takiguchi, Satoshi Tamura, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, “Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments — Newest Part of the CENSREC Series —,” *Proc. LREC 2008*, CD-ROM proceedings, May 2008.
  9. Norihide Kitaoka, Kazumasa Yamamoto, Tomohiro Kusamizu, Seiichi Nakagawa, Takeshi Yamada, Satoru Tsuge, Chiyomi Miyajima, Takanobu Nishiura, **Masato Nakayama**, Yuki Denda, Masakiyo Fujimoto, Tetsuya Takiguchi, Satoshi Tamura, Shingo Kuroiwa, Kazuya Takeda, and Satoshi Nakamura, “Development of VAD Evaluation Framework CENSREC-1-C and Investigation of Relationship Between VAD and Speech Recognition Performance,” *Proc. 2007 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*

- 2007), pp.607–612, Dec. 2007.
10. Hiroaki Ishii, **Masato Nakayama**, Takanobu Nishiura, and Shinichi Nakagawa, “A Suggestion of Noise Reduction Wall With Acoustic Fresnel Lens and Active Noise Control System,” *Proc. 2007 International Commission for Acoustics (ICA 2007)*, Sep. 2007.
  11. Takanobu Nishiura, Yoshiki Hirano, Yuki Denda, and **Masato Nakayama**, “A Study of Suitable Reverberation Criteria for Distant-talking Speech Recognition With Room Acoustic Parameters,” *Proc. 2007 International Commission for Acoustics (ICA 2007)*, Sep. 2007.
  12. Takanobu Nishiura, Yoshiki Hirano, Yuki Denda, and **Masato Nakayama**, “Investigations into Early and Late Reflections on Distant-talking Speech Recognition Toward Suitable Reverberation Criteria,” *Proc. INTERSPEECH 2007*, pp.1052–1055, Aug. 2007.
  13. Yuki Denda, Takamasa Tanaka, **Masato Nakayama**, Takanobu Nishiura, and Yoichi Yamashita, “Noise-Robust Hands-free Voice Activity Detection With Adaptive Zero Crossing Detection using Talker Direction Estimation,” *Proc. INTERSPEECH 2007*, pp.222–225, Aug. 2007.
  14. Takamasa Tanaka, Yuki Denda, **Masato Nakayama**, and Takanobu Nishiura, “An Evaluation of Hands-free Voice Activity Detection Algorithm based on Weighted CSP Analysis and Zero Crossing Detector,” *Proc. 4th Joint Meeting of the ASA and the ASJ*, p.3040, Nov. 2006.
  15. Yoshiki Hirano, Yuki Denda, **Masato Nakayama**, and Takanobu Nishiura, “An investigation of early reflection signal based on impulse response on automatic speech recognition,” *Proc. 4th Joint Meeting of the ASA and the ASJ*, p.3042, Nov. 2006.
  16. Keiichiro Yamada, Yuki Denda, **Masato Nakayama**, and Takanobu Nishiura, “A design of virtual sound field simulator based on acoustic ray tracing and wall impedance,” *Proc. 4th Joint Meeting of the ASA and the ASJ*, p.3213, Nov. 2006.



17. Keitaro Kuroda, Yuki Denda, **Masato Nakayama**, and Takanobu Nishiura, "Acoustic sound field simulator based on head-related transfer function and 3-dimensional mirror-image," *Proc. 4th Joint Meeting of the ASA and the ASJ*, p.3214, Nov. 2006.
18. **Masato Nakayama**, Takanobu Nishiura, and Yoichi Yamashita, "A Design of Fast Steering Filters Based on the Adaptive Fusion of Predesigned Finite Impulse Response Filters for Microphone Array," *Proc. 4th Joint Meeting of the ASA and the ASJ*, p.3210, Nov. 2006.
19. Takamasa Tanaka, Yuki Denda, **Masato Nakayama**, and Takanobu Nishiura, "A Study of Hands-free Voice Activity Detection based on Weighted CSP Analysis and Zero Crossing Detector," *Proc. 9th Western Pacific Acoustics Conference (WESPAC IX 2006)*, PaperID:469, Jun. 2006.
20. **Masato Nakayama**, Takanobu Nishiura, and Yoichi Yamashita, "Hands-Free Speech Recognition With Average Phoneme-based AMNOR," *Proc. 9th Western Pacific Acoustics Conference (WESPAC IX 2006)*, PaperID:343, Jun. 2006.
21. **Masato Nakayama**, Yuki Denda, Takanobu Nishiura, Hideki Kawahara, and Toshio Irino, "An Evaluation of In-CAR Speech Enhancement Techniques With Microphone Array Steering," *Proc. 2004 International Conference on Acoustics (ICA2004)*, vol.4, pp.3041–3044, Apr. 2004.
22. **Masato Nakayama**, Takanobu Nishiura, and Hideki Kawahara, "Adaptive Beamformer Based on Average Vowel / Consonant Spectrum with Phoneme Identification," *Proc. 2003 International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*, pp.243–246, Sept. 2003.
23. Takanobu Nishiura, **Masato Nakayama**, and Satoshi Nakamura, "An Evaluation of Adaptive Beamformer Based on Average Speech Spectrum for Noisy Speech Recognition," *Proc. 4th. IEEE International Conference on Multimedia and Expo (ICME2003)*, no.3, pp.209–212, Jul. 2003.
24. Takanobu Nishiura, **Masato Nakayama**, and Satoshi Nakamura, "An Evalu-

ation of Adaptive Beamformer Based on Average Speech Spectrum for Noisy Speech Recognition,” *Proc. 28th. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003)*, no.1, pp.668–671, Apr. 2003.

25. Masato Nakayama, Takanobu Nishiura, and Hideki Kawahara, “Adaptive Beamformer Based on Average Vowels/Consonant Spectrum Weights for Noisy Speech Recognition,” *Proc. First Pan-American/Iberian Meeting on Acoustics (Joint Meeting ASA/FIA/IMA)*, pp.2323, Cancun, Mexico, Dec. 2002.

## 研究会

26. 田村 哲嗣, 宮島 千代美, 北岡 教英, 武田 一哉, 山田 武志, 滝口 哲也, 柘植 覚, 山本 一公, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 松田 繁樹, 小川 哲司, 黒岩 眞吾, 中村 哲, “雑音下マルチモーダル音声認識評価基盤 CENSREC-1-AV の構築,” 情報処理学会研究報告 [音声言語情報処理], vol.2010-SLP-82, no.7, Jul. 2010.
27. 中山 雅人, 中迫 昇, “左右の近距離スピーカを利用した距離推定に基づく音場再現の検討,” 電子情報通信学会技術研究報告 [応用音響] EA2010-13, vol.110, no.54, pp.73–78, May 2010.
28. 北岡 教英, 山田 武志, 滝口 哲也, 柘植 覚, 山本 一公, 宮島 千代美, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 田村 哲嗣, 松田 繁樹, 小川 哲司, 黒岩 眞吾, 武田 一哉, 中村 哲, “雑音下音声認識評価ワーキンググループ活動報告：認識に影響する要因の個別評価環境 (3) (SIG-SLP 内組織の活動報告),” 情報処理学会研究報告 [音声言語情報処理], vol.2008, no.102, pp.41–46, Oct. 2008.
29. 篠原 寿広, 長谷川 敦史, 中山 雅人, 中迫 昇, “脳血管の構造分析のための頭部 3 次元 CT 画像を用いた血管追跡による血管位置情報の推定,” センシングフォーラム資料 26, pp.97–100, Sept. 2008.
30. 傳田 遊亀, 田中 貴雅, 中山 雅人, 西浦 敬信, 山下 洋一, “話者方位情報とゼロ交差情報に基づくハンズフリー発話間検出の評価,” 第 65 回 音声言語情報処理研究会 (SIG-SLP-65), pp.109–114, Feb. 2007.

31. 北岡 教英, 山田 武志, 滝口 哲也, 柘植 覚, 山本 一公, 宮島 千代美, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 田村 哲嗣, 松田 繁樹, 小川 哲司, 黒岩 眞吾, 武田 一哉, 中村 哲, “雑音下音声認識評価ワーキンググループ活動報告: 認識に影響する要因の個別評価環境 (2) (雑音・VAD, 第9回音声言語シンポジウム),” 情報処理学会研究報告 [音声言語情報処理], vol.2007, no.129, pp.1-6, Dec. 2007.
32. 北岡 教英, 山田 武志, 滝口 哲也, 柘植 覚, 山本 一公, 宮島 千代美, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 田村 哲嗣, 黒岩 眞吾, 武田 一哉, 中村 哲, “雑音下音声認識評価ワーキンググループ活動報告: 認識に影響する要因の個別評価環境,” 第64回音声言語情報処理研究会 (SIG-SLP-64), pp.1-6, Dec. 2006.
33. 北岡 教英, 山田 武志, 柘植 覚, 宮島 千代美, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 山本 一公, 滝口 哲也, 黒岩 眞吾, 武田 一哉, 中村 哲, “CENSREC-1-C: 雑音下音声区間検出評価基盤の構築,” 情報処理学会研究報告 [音声言語情報処理], vol.2006, no.107, pp.1-6, Oct. 2006.
34. 傳田 遊亀, 田中 貴雅, 中山 雅人, 西浦 敬信, 山下 洋一, “マイクロホンアレーを用いた時間/空間情報に基づくハンズフリー発話区間検出の検討,” 第62回音声言語情報処理研究会 (SIG-SLP-62), pp.7-12, Jul. 2006.
35. 中山 雅人, 西浦 敬信, 河原 英紀, “母音/子音区間検出を備えた適応形ビームフォーマを用いた母音/子音平均スペクトルに基づく雑音下音声認識,” 電子情報通信学会技術研究報告 [応用音響] EA2003-56, vol.103, no.251, pp.39-44, Aug. 2003.

## 大会発表

36. 英 慎平, 中山 雅人, 中迫 昇, 篠原 寿広, 上保 徹志, “音響測距法を利用した話者位置推定システム構築に向けた検討,” 第13回日本音響学会関西支部若手研究者交流研究発表会, Dec. 2010.
37. 太田 将史, 中山 雅人, 中迫 昇, 篠原 寿広, 上保 徹志, “マフラー音の位相干渉を利用したリアソナーシステム開発に向けた基礎的検討,” 第13回日本音響学会関西支部若手研究者交流研究発表会, Dec. 2010.

38. 西前 達矢, 中山 雅人, 中迫 昇, 篠原 寿広, 上保 徹志, “アナウンス音声の位相干渉と音素認識に基づく音響測距法の検討,” 第 13 回日本音響学会関西支部若手研究者交流研究発表会, Dec. 2010.
39. 廣畑 和紀, 中山 雅人, 中迫 昇, 篠原 寿広, 上保 徹志, “近距離スピーカを利用した音響測距に基づく音場再現手法の音質に関する検討,” 第 13 回日本音響学会関西支部若手研究者交流研究発表会, Dec. 2010.
40. 河西 慶治, 中迫 昇, 篠原 寿広, 中山 雅人, 上保 徹志, “位相干渉に基づく音響測距法における数値微分を利用した DC 成分除去の検討,” 第 13 回日本音響学会関西支部若手研究者交流研究発表会, Dec. 2010.
41. 中山 雅人, 太田 将史, 篠原 寿広, 中迫 昇, 上保 徹志, “2ch 音響測距法と CSP 法を統合した話者位置推定法の検討,” 日本音響学会 2010 年秋季研究発表会, pp.593–594, Sept. 2010.
42. 中山 雅人, 廣畑 和紀, 中迫 昇, “近距離スピーカを利用した音響測距法に基づく音場再現手法 -再現音質の検証とその補正法の検討-,” 日本音響学会 2010 年秋季研究発表会, pp.757–758, Sept. 2010.
43. 中迫 昇, 英 慎平, 中山 雅人, 篠原 寿広, 上保 徹志, “既知信号の減算を利用した位相干渉に基づく音響測距法の検討,” 日本音響学会 2010 年秋季研究発表会, pp.599–600, Sept. 2010.
44. 英 慎平, 中山 雅人, 篠原 寿広, 中迫 昇, 上保 徹志, “移動物体に対する時間幅を持つ送信波の位相干渉に基づく音響測距法の基礎的検討,” 日本音響学会 2010 年秋季研究発表会, pp.687–688, Sept. 2010.
45. 中山 雅人, 英 慎平, 篠原 寿広, 中迫 昇, 上保 徹志, “移動物体に対する可聴音の干渉を利用した近距離推定手法の検討,” 日本音響学会 2010 年春季研究発表会, pp.825–826, Mar. 2010.
46. 田村 哲嗣, 宮島 千代美, 北岡 教英, 武田 一哉, 山田 武志, 滝口 哲也, 柘植 覚, 山本 一公, 西浦 敬信, 中山 雅人, 傳田 遊亀, 藤本 雅清, 松田 繁樹, 小川 哲司, 黒岩 眞吾, 中村 哲, “CENSREC-1-AV: マルチモーダル音声認識コーパスの構築,” 日本音響学会 2010 年春季研究発表会, pp.219–220, Mar. 2010.
47. 中山 雅人, 定野 一也, 篠原 寿広, 中迫 昇, “定在波を利用した距離推定とマイク

- ロホンアレーに基づく話者位置推定の検討,” 日本音響学会 2009 年春季研究発表会, pp.763–764, Mar. 2009.
48. 川畑 尚也, 中山 雅人, 西浦 敬信, 飯田 雅巨, “高雑音下音声受音用ヘッドセットマイクロホンアレーの開発,” 日本音響学会 2008 年秋季研究発表会, pp.797–798, Sept. 2008.
49. 西浦 敬信, 傳田 遊亀, 中山 雅人, “マイクロホンアレーを用いた遠隔発話音声処理,” 第 52 回システム制御情報学会研究発表講演会, pp.353–354, May 2008.
50. 西浦 敬信, 中山 雅人, 傳田 遊亀, 北岡 教英, 山本 一公, 山田 武志, 藤本 雅清, 柘植 覚, 宮島 千代美, 滝口 哲也, 田村 哲嗣, 小川 哲司, 松田 繁樹, 黒岩 眞吾, 武田 一哉, 中村 哲, “残響下音声認識評価基盤 ( CENSREC-4 ) の構築,” 日本音響学会 2008 年春季研究発表会, pp.175–178, Mar. 2008.
51. 平野 良季, 傳田 遊亀, 中山 雅人, 西浦 敬信, “室内音響指標を用いた残響下音声認識性能の評価,” 日本音響学会 2008 年春季研究発表会, pp.183–184, Mar. 2008.
52. 石井 宏明, 中山 雅人, 西浦 敬信, 仲川 真一, “音響レンズと ANC を併用した環境雑音抑圧の実験的検討 ( 2 ) - 低域騒音抑圧に最適な音響レンズの形状に関する検討 - ,” 日本音響学会 2008 年春季研究発表会, pp.613–614, Mar. 2008.
53. 西川 浩太郎, 傳田 遊亀, 中山 雅人, 西浦 敬信, 南條 浩輝, “音素時系列を用いたスペクトルサブトラクション法の検討,” 日本音響学会 2008 年春季研究発表会, pp.185–186, Mar. 2008 .
54. 南條 浩輝, 国松 卓, 川野 弘, 中山 雅人, 西浦 敬信, “音響防犯システムのための叫び声の基礎的検討,” 日本音響学会 2008 年春季研究発表会, pp.195–198, Mar. 2008.
55. 石井 宏明, 中山 雅人, 西浦 敬信, 仲川 真一, “音響フレネルレンズと ANC を併用した環境雑音抑圧の実験的検討,” 日本音響学会 2007 年度秋季研究発表会, pp.785–789, Sept. 2007.
56. 平野 良季, 傳田 遊亀, 中山 雅人, 西浦 敬信, “室内音響指標を用いた残響下音声認識性能の分析と推定,” 日本音響学会 2007 年度秋季研究発表会, pp.205–206, Sep. 2007.
57. 傳田 遊亀, 田中 貴雅, 中山 雅人, 西浦 敬信, 山下 洋一, “話者方位推定と適応ゼ

- 口交差検出を用いた発話区間形成法の評価,” 日本音響学会 2007 年度秋季研究発表会, pp.203–204, Sep. 2007.
58. 傳田 遊亀, 田中 貴雅, 中山 雅人, 西浦 敬信, 山下 洋一, “話者方位推定に基づく適応型ゼロ交差検出を用いたハンズフリー発話区間検出法の評価,” 日本音響学会 2007 年度春季研究発表会, pp.103–104, Mar. 2007.
59. 傳田 遊亀, 田中 貴雅, 中山 雅人, 西浦 敬信, 山下 洋一, “話者方位推定と適応型ゼロ交差検出に基づく時間・空間領域発話区間検出法の検討,” 第 9 回日本音響学会関西支部若手研究者交流研究発表会, pp.11, Dec. 2006.
60. 北岡 教英, 西浦 敬信, 中山 雅人, 藤本 雅清, 山田 武志, 滝口 哲也, 山本 一公, 宮島 千代美, 柘植 覚, 中村 哲, 武田 一哉, 黒岩 眞吾, “雑音下音声区間検出手法評価基盤の構築,” 日本音響学会 2006 年度秋季研究発表会, pp.103–104, Sep. 2006.
61. 中山 雅人, 西浦 敬信, 山下 洋一, “フィルタ合成型 AMNOR における最適フィルタ係数の決定法に関する検討,” 日本音響学会 2006 年度秋季研究発表会, pp.489–490, Sep. 2006.
62. 田中 貴雅, 傳田 遊亀, 中山 雅人, 西浦 敬信, “マイクロホンアレーを用いた時間/空間情報に基づくハンズフリー発話区間検出法の評価,” 日本音響学会 2006 年度秋季研究発表会, pp.25–26, Sep. 2006.
63. 中山 雅人, 西浦 敬信, 山下 洋一, “音素特徴量に基づく適応形アレーの検討,” 日本音響学会 2006 年度春季全国大会講演論文集, pp.645–646, Mar. 2006.
64. 田中 貴雅, 傳田 遊亀, 中山 雅人, 西浦 敬信, “Weighted CSP 法と音声特徴量に基づくハンズフリー発話区間検出の検討,” 日本音響学会 2006 年度春季全国大会講演論文集, pp.149–150, Mar. 2006.
65. 中山 雅人, 西浦 敬信, 山下 洋一, “マイクロホンアレーを用いた温度変化に頑健な音声受音に関する検討,” 日本音響学会関西支部 若手研究者交流研究発表会, pp.13, Dec. 2005.
66. 中山 雅人, 西浦 敬信, 山下 洋一, “適応アルゴリズムを利用したフィルタ係数合成のためのマイクロホン素子校正法の検討,” 日本音響学会 2005 年度秋季全国大会講演論文集, pp.697–698, Sep. 2005.

67. 中山 雅人, 西浦 敬信, 山下 洋一, “フィルタ加算に基づく適応形ビームフォーミングの指向性制御に関する検討,” 日本音響学会 2005 年度春季全国大会講演論文集, pp 523–524, Mar. 2005.
68. 中山 雅人, 西浦 敬信, 河原 英紀, 入野 俊夫, “GMM による母音/子音区間検出を用いた母音/子音平均スペクトルに基づく適応形ビームフォーミングの検討,” 日本音響学会 2004 年度春季全国大会講演論文集, pp.647–648, Mar. 2004.
69. 中山 雅人, 西浦 敬信, 河原 英紀, “GMM による母音/子音区間の自動検出と適応形ビームフォーミングを用いた母音/子音平均スペクトルに基づく雑音下音声認識の検討,” 日本音響学会 2003 年度秋季全国大会講演論文集, pp.507–508, Sep. 2003.
70. 中山 雅人, 西浦 敬信, 河原 英紀, “適応型ビームフォーミングを用いた母音/子音平均スペクトルに基づく音声強調,” 日本音響学会 2003 年度春季全国大会講演論文集, pp.757–758, Mar. 2003.

## 特許

71. 特願 2007-178552, 特公 2009-13710 “防音装置,” Sep. 2007.