JAIST Repository
https://dspace.jaist.ac.jp/

Title	マイクロホン対を用いた雑音除去法に関する研究
Author(s)	水町，光徳
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/902
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

博 士 論 文

マイクロホン対を用いた雑音除去法に関する研究

指導教官　赤木 正人 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

水町 光徳

2000 年 1 月 14 日

Copyright © 2000 by Mitsunori MIZUMACHI

目 次

1 序論 **1**

1.1 本研究の背景 2

1.1.1 雑音除去の必要性 2

1.1.2 雑音除去技術の現状及び問題点 2

1.2 本論文の目的 7

1.3 本論文の構成と各章の概要 9

2 各雑音除去法の非定常雑音に対する有効性 **13**

2.1 はじめに 14

2.2 スペクトルサブトラクション 14

2.2.1 単点受音雑音適応型 SS 15

2.2.2 2ch 雑音推定 SS 17

2.3 遅延和アレー 19

2.4 減算型アレー 22

2.5 まとめ 27

3 雑音除去アルゴリズムの定式化 **28**

3.1 はじめに 29

3.2 定式化 30

3.2.1 目的信号および雑音の到来方向推定 31

3.2.2 雑音スペクトルの推定 38
3.2.3 雑音スペクトルの除去 40
3.3 まとめ 41

4 雑音除去アルゴリズムの基礎的検証 42
4.1 はじめに 43
4.2 雑音除去アルゴリズムのパラメータ設定 43
4.3 方向推定アルゴリズムの精度検証 46
4.4 最適マイクロホン対選択機構の有効性検証 51
4.5 突発性雑音の除去実験 53
4.6 まとめ 57

5 音声認識のフロントエンドとしての有効性検証 58
5.1 はじめに 59
5.2 音声認識実験による有効性検証 59
5.3 ASR 雑音除去フロントエンドのための客観的評価尺度 62
5.3.1 客観的歪み音声評価尺度概説 62
5.3.2 客観的評価尺度 LPC-SED の定義 62
5.3.3 客観的評価尺度 LPC-SED の妥当性検証 64
5.4 客観的評価尺度による雑音除去アルゴリズムの評価 66
5.4.1 客観的評価尺度 LPC-SED による有効性検証 66
5.4.2 客観評価尺度 NDSR による有効性検証 72
5.5 まとめ 75

6 聴感上の歪み感低減能力の検証 76
6.1 はじめに 77
6.2 主観評価実験による提案法の有効性検証 77
6.3 聴感上の印象を考慮した客観的評価尺度 80

6.3.1 聴感上の歪み印象に対応した客観的評価尺度概説 80
6.3.2 客観的歪み評価尺度 ASD 82
6.4 客観的評価尺度による雑音除去アルゴリズムの評価 83
6.5 まとめ 88

7 実環境における有効性検証 **89**
7.1 はじめに 90
7.2 実環境における雑音除去実験 90
7.3 まとめ 98

8 結論 **99**
8.1 本論文の要約 100
8.2 今後の課題 103

付録 **105**
A 1. 聴覚特性を考慮した加法性雑音に対する歪み評価尺度の構築 106
A 2. 聴感上の印象を考慮した評価尺度 ASD の妥当性検証 113

謝辞 **119**

参考文献 **120**

本研究に関する発表論文 **128**

第 1 章

序論

1.1 本研究の背景

1.1.1 雑音除去の必要性

雑音除去技術の需要は、ディジタル音響処理技術の発展につれて益々高まっている。音声情報処理に関する研究では、基礎研究から実用化段階へ進むにつれ、雑音による性能低下が表面化する。その顕著な例が、自動音声認識器 (Automatic Speech Recognizer; ASR) である。計算機シミュレーションという理想的な環境ではほぼ 100 % の認識率を誇る ASR であっても、実環境においては認識率は大幅に低下することが報告されている [1]。この要因の一つが、実環境には必ず存在する雑音や残響の影響である [2][3][4]。

また、難聴者や高齢者は、健聴者と比較し、雑音や残響の影響によって音声聴取が困難になる傾向がより顕著に現われるという報告がある [5]。そのため、補聴器への雑音除去技術の導入が盛んに試みられており、その有効性が報告されている [6]。高齢化が進む現代社会において、ASR や補聴器などの社会福祉関連機器への雑音除去の需要は、今後更に高まることが予想される。雑音除去技術は、現在の音声情報処理にとって既に欠くことのできない要素となっており [7]、今後その重要性は益々高まるであろう。

1.1.2 雑音除去技術の現状及び問題点

雑音除去に関する研究は非常に長い歴史を持っており、特に 1970 年代後半からは音声情報処理研究の重要な一部門として注目を浴び続けている。これまでに様々な雑音除去法が提案されてきたが、そのほとんどが定常的な雑音を対象としており、非定常雑音への対応については決して十分とは言えない。従来の雑音除去法には、信号の統計的性質に制約を設け、目的信号と雑音とが無相関であることを前提にした手法が多い。例えば、雑音除去を行なう短時間フレーム内では、目的信号と雑音がともに定常であると仮定し、目的信号と雑音との混合信号から目的信号を推定するウィーナーフィルタ [8] を用いた雑音除去法 [9]、ウィーナーフィルタを拡張することにより非定常信号も扱うことのできるカルマンフィルタ [8] を用いた雑音除去法 [10] などが提案されている。しかし、モデル化した信号と

実環境に存在する信号との統計的性質の差が大きすぎるため、これらは実環境において有効な手法とは言えない。今日の雑音除去技術には、ドアの開閉音や電話のベル、あるいは複数の話者が同時に喋っている場合など、我々の日常生活に存在するような非定常で、しかも目的信号と相関の高い雑音の除去能力こそが求められている。

現在、単点受音の雑音除去法としては、Boll の スペクトルサブトラクション (Spectral Subtraction; SS)[11] およびその改良法がよく用いられている。また、近年は、マイクロホンアレーを用いた雑音除去法も数多く提案されている。そこで、以下に SS による雑音除去法、マイクロホンアレーによる雑音除去法、そして ASR や補聴器など各用途に応じた雑音除去法、最後に我々の聴覚系における雑音除去処理とそれを利用した雑音除去法について概説する。

a. スペクトルサブトラクションによる雑音除去法

1979 年に Boll が提案した SS は、ヘリコプタのコックピットに存在するような定常的な雑音のみを除去対象としており非定常雑音へは対応できないが、ノンパラメトリックで、処理が非常に簡潔であることから実用化に適した手法である。SS は、Boll の提案以来現在に至るまで、定常雑音の除去精度を向上させようという試み、非定常雑音を除去しようという試みなど、非常にたくさんの改良法が提案されてきた。

SS 改良法は、Boll の SS と同じく単点受音、雑音スペクトルの推定精度を向上させることを目的とした多点受音に分類できる。単点受音の SS 改良法としては、音声／非音声の判定を行って非音声区間で雑音スペクトルを推定するのではなく、雑音スペクトルを受音信号をもとに随時変形していく手法 [12] などが提案された。この手法は、雑音の音響的性質の緩やかな変化には追従できるが、実環境に存在するような非定常雑音に対応することはできない。雑音の大きさの変動に対しては、SS の後処理としての周波数平滑化 [13]、自己相関関数による雑音レベルの逐次推定 [14]、適応的なゲイン制御 [15] なども有効である。一方、受音点を複数とすることにより、信号の到来方向という空間情報を利用して雑音スペクトルの推定を行なおうとするアプローチもある。2 点受音の差信号により雑音スペク

トルを推定する手法 [16] が提案されたが、雑音スペクトルの推定精度に問題があり、雑音除去処理によって雑音により音声に生じた歪みを増大させる可能性もある。

また、SS 改良法としては、雑音スペクトルの推定精度を向上させることが目的ではなく、雑音除去音声の品質向上を目指した手法も多い。例えば、ケプストラム領域において雑音スペクトルの平滑化を行なう手法 [17]、相関領域での SS [18]、Wavelet 変換と SS とによる雑音除去法 [19]、聴覚末梢系の周波数分析機構を模擬したガンマチャープフィルタバンク [20] と SS による雑音除去法 [21] などが提案されている。

b. マイクロホンアレーによる雑音除去法

マイクロホンアレーを用いることにより、単点受音では得ることができない信号の到来方向という空間情報を得ることができる。我々の聴覚系が 2 つの受音器を備えていることからも、空間情報が雑音除去処理に重要な役割を果たすであろうことが予測できる。マイクロホンアレーを用いた雑音除去法の中には、空間情報に加え、音声の統計的性質、調音機構、ピッチ、ホルマント構造など、単点受音の雑音除去法で用いられている音声特有の情報を積極的に利用する手法 [22] なども提案されている。

マイクロホンアレーに関する研究は、遅延和アレーと減算型アレーとに分類されることが多い。遅延和アレーとは、目的信号を同相化、すなわち、目的信号が各マイクロホンへ到来する際に生じる時間差をすべて揃えるために各受音信号に適当な遅延を与え、それらを加算することにより、特定の方向から到来する信号のみを強調する雑音除去法である。遅延和アレーの雑音除去精度の向上を図るためには、目的信号方向に非常に鋭い指向特性を構成する必要がある。鋭い指向特性を実現するためには、莫大な素子数が必要となることから、遅延和アレーの用途は限られる。例えば、Flanagan らは講堂に設置することを目的とした 400 素子の 2 次元アレー [23]、金森らは 30 素子の 1 次元アレー [24] を提案している。

減算型アレーとは、より少ない素子で高精度の雑音除去を実現するため、ビームパターン (マイクロホンアレーの指向特性) の死角に着目し、それを積極的に利用するビームフォー

ミング手法である。例えば、2 つのマイクロホンでの受音信号に対して雑音を同相化した後、その 2 つの信号を減算することにより雑音は完全に抑圧される。これが、雑音の到来方向にビームパターンの死角を形成すること、つまり雑音方向に鋭いディップを持つ空間フィルタを設計することに相当する。現在の減算型アレーの多くは、適応フィルタを用いてビームフォーミングを行なっている。適応型アレーとしては、線形拘束条件を適用した手法 [25][26]、非線形拘束条件を適用した手法 [27][28] など様々提案されており、各拘束条件のもとで雑音のパワーが最小になるように適応フィルタの係数を決定する。適応型アレーの代表として挙げられることの多い Griffiths-Jim 型アレー [26] は、目的信号に対する周波数特性を平坦に保つという厳しい線形拘束条件を採用している。AMNOR 方式 [28] は、若干の偏差を許容した緩やかな非線形拘束条件を用いることにより、演算量は増えるものの雑音除去精度が向上する。近年、ハードウェア技術の飛躍的な発展に伴い、演算量に関する制限が緩和され、適応型アレーを用いた雑音除去法に関する研究は、益々盛んに行なわれている。浅野らは、適応ビームフォーマとサブスペース法 [29] とを組み合わせた雑音除去法を提案している [30]。猿渡らは、より少ない素子数で、より多くの死角を形成するための相補的ビームフォーミング手法 [31] を利用した非線形アレー信号処理に基づく雑音除去法を提案している [32]。また、音源の方向情報（音源移動に関する情報）を積極的に利用する手法も提案されており、永田らは 2ch 適応型アレーを用いた音源追尾機構を備えた雑音抑圧法 [33][34] 、中村らは音源方向探索と適応型アレーとを組み合わせた雑音除去法 [35][36] を提案している。しかしながら、これらの適応フィルタを用いた雑音除去法は、目的信号と雑音の無相関性を前提としているため、残響のある環境においては適応フィルタがなかなか収束しないという問題を抱えている [37]。

c. 各用途に応じた雑音除去法

雑音や残響などの周囲環境の影響は、ASR にとって脅威的存在であり [2] [3] [4]、それゆえ数多くの対策方法が提案されている [1][38]。これらは、分析部における耐雑音対策、照合部における耐雑音対策の 2 つに大別できる [39]。前者に関しては、単点受音の場合には

SS が採用されることが多いが、マイクロホンアレーを用いる場合には多種多様な手法が提案されている [38]。後者に関しては、音声の統計的モデル化手法である隠れマルコフモデル (Hidden Markov Model; HMM)[39] を用いた ASR においては、クリーン音声 HMM と雑音 HMM を準備しておき、それらを合成することにより雑音に適応した HMM を作成するモデル分解合成法 [40][41] などがある。

また、マイクロホンアレーによる雑音除去技術は、ディジタル補聴器にも導入されている [42][43][44][45] [46]。この背景には、補聴器には必ずマイクロホンが使用されており、通常は両耳に装着することから補聴器は 2ch マイクロホンアレーの受音系と見做すことができ、また補聴器のディジタル化が進むにつれ高度な雑音除去処理の導入が可能になったことがある。本論文で提案する雑音除去法は、3ch 直線配列マイクロホンアレーを利用するため、直接補聴器へ応用することは難しい。しかし、例えば 2 つのマイクロホンを補聴器へ、残りの 1 つをメガネに装着するなどマイクロホン配置を工夫することにより、提案法は補聴器のための雑音除去法としても利用可能である。

d. 聴覚系における雑音除去処理

我々の聴覚系には優れた雑音除去能力が備わっているが、残念ながらその機構はまだ解明されていない。例えば、複数の音源が存在する環境において、我々はある特定の音のみを分離抽出して聴くことができる。これは、カクテルパーティ効果と呼ばれている。カクテルパーティ効果を解明するためには、情報の選択と統合という高次の情報処理機構を知る必要があり [47]、それが如何に困難な課題であるかは容易に想像できる。カクテルパーティ効果も含めて聴覚心理学で得られた数多くの知見を説明するため、ゲシュタルト心理学的発想から聴覚情景解析 (Auditory Scene Analysis; ASA) [48] という概念が提案され、これを数理工学的に解釈して利用した雑音除去法も提案されている [49][50]。

また、両耳聴を考慮し、両耳間時間差に基づいた二重母音の分離モデル [51]、信号の両耳間位相差に起因する両耳間マスキングレベル差 [52] によって特定方向から到来する信号を強調あるいは抑圧する EC 理論 [53] などが提唱されている。

1.2 本論文の目的

音声情報処理の高度化に伴い、雑音除去技術の需要は高まり、それに応えて第 1.1 節に示したように様々な雑音除去法が提案されてきた。従来の雑音除去に関する研究は、定常雑音を対象とした手法がその大半を占めており、定常雑音の除去に関しては満足できるレベルに達している。近年、非定常雑音の除去を目指した手法も提案されており、それらは音響的特性が滑らかに変化するような準定常的な雑音は除去できるが、音声区間中で突発的に発生するような雑音などを除去することは困難である。例えば適応フィルタを利用した手法は、雑音の音響的性質は激しく変化せず、目的信号と雑音とは無相関であるという非現実的な仮定を必要とする。

本論文では、目的信号に対して加法的関係 [39] にあるすべての信号を除去対象の雑音と考える。つまり、従来の雑音除去法では扱うことが困難で対象外とされてきた非定常雑音、例えば目的信号中で突発的に発生し、目的信号との相関が高い雑音などを除去することを本論文の目的とする。そして、雑音除去技術を早急に必要としている ASR のフロントエンド、あるいは補聴器などの聴感上の歪み感が問題となる機器への応用を目指す。

本論文では、時々刻々変化するような非定常雑音の除去を実現するため、マイクロホン対を基礎とした雑音除去アルゴリズムを構築する。マイクロホンアレーを用いた雑音除去に関する研究では、一般に素子数を増やすほど雑音除去精度は向上する。この典型的な例が遅延和アレーである。しかし、例えば可搬型 ASR や補聴器のための雑音除去を考えた場合、受音系のサイズあるいは設置場所には制約があり、素子数はできる限り少ない方が望ましい。本研究では 3 素子のマイクロホンアレーを使用し、3 素子のうち各処理において最適な 2 つの素子を選択してマイクロホン対を構成する。

本雑音除去法は、特定の方向から到来する信号を抽出する空間フィルタリング手法と、周波数領域における雑音除去法とを融合させた雑音除去アルゴリズムである。空間フィルタリングに関しては、先に目的信号および雑音の到来方向を推定し、その情報をもとに解析的に空間フィルタを設計する。現在、空間フィルタリングには一般的に適応処理が利用されている。これは雑音環境下での信号の到来方向推定が非常に困難であることに起因し

ており、本手法では方向推定精度を大幅に向上させ、適応処理を用いず解析的に空間フィルタを設計することを試みる。そして、空間フィルタリングによって時々刻々推定した雑音成分を受音信号から減算することにより、非定常雑音の除去を実現する。本手法は、従来の雑音除去法のように雑音の音響的性質に対して制約を設ける必要がないため、ドアの開閉音のように突発的に発生する雑音、目的信号と雑音がともに音声であるような現実的な環境においても有効であることが期待できる。

本論文では、提案法および従来法に対し、非定常雑音に対する雑音除去精度を検証し、提案法の雑音除去能力、従来法に対する優位性についても詳細な検討を加える。具体的な応用例を想定し、ASR のフロントエンドとしての有効性、補聴器などで問題となる聴感上の歪み感低減能力についても検証する。前者に関しては音声認識実験による評価を行なうとともに、ASR のフロントエンドとしての有効性を評価するための客観的評価尺度を定義し、後者に関しても主観評価実験に加え、聴覚特性を考慮した聴感上の歪み印象と対応のよい客観的評価尺度を構築し、各目的に応じて効率的な評価を行なう。また、残響や暗騒音のある環境において雑音除去実験を行ない、提案法の実環境における有効性についても調査検討する。

1.3　本論文の構成と各章の概要

本論文は、全 8 章により構成されている。各章の概要を以下に述べ、各章の相互関係を図 1.1 に示す。

第 1 章：　序論

雑音除去研究の現状および問題点を明確にし、本研究の目的を述べる。

第 2 章：　各雑音除去法の非定常雑音に対する有効性

計算機シミュレーションにより、従来の雑音除去法が、非定常雑音をどの程度除去可能であるのかを詳細に調査する。

本章では、序論で概説した従来の雑音除去法のうち、周波数領域におけるフィルタリング手法として非定常雑音の除去を目的とした SS 改良法、空間フィルタリング手法として遅延和アレー、減算型アレーの代表としての Griffiths-Jim 適応型アレーに対してより詳細な検討を加える。特に、時々刻々変化する非定常雑音に対する各手法の雑音除去能力について調査する。

第 3 章：　雑音除去アルゴリズムの定式化

本章では、2 本のマイクロホンにより構成されるマイクロホン対という概念に基づき、非定常雑音にも対応可能な雑音除去アルゴリズムの定式化を行なう。

まず、雑音環境下での目的信号および雑音の到来方向を推定する。ここでは、受音信号の相互相関を利用して最も優勢な信号の到来方向を推定し、その方向から到来する信号を完全に抑圧するような空間フィルタを設計する。その空間フィルタの出力に対してもう一つの信号到来方向を推定し、目的信号および雑音の到来方向を決定する。次に、雑音スペクトルの推定に関しては、信号の到来方向という空間情報を利用した空間フィルタリング

手法を用いる。ここでは、マイクロホン対の大きさによって空間的零点が生じる方向が異なることを積極的に利用し、周波数帯域ごとに最適なマイクロホン対を選択し、雑音スペクトルを推定する。最後に、スペクトルサブトラクションにより、受音信号から雑音成分を除去する。以上により、時々刻々変化する非定常雑音の除去が実現できる。

第 4 章： 雑音除去アルゴリズムの基礎的検証

第 3 章で定式化した雑音除去アルゴリズムの基礎的な雑音除去能力を検証する。

提案法の特徴である雑音環境下での目的信号および雑音の到来方向推定、雑音スペクトル推定における周波数ごとの最適マイクロホン対選択機構の有効性について調査する。そして、本研究の主目的である突発性雑音の除去に関して、計算機シミュレーションによる雑音除去実験を行ない、従来法に対する提案法の優位性について検討する。

第 5 章： 音声認識のフロントエンドとしての有効性検証

提案法の応用例として音声認識のフロントエンドを想定し、提案法の有効性を検証する。

音声認識実験を行ない、提案法の音声認識のフロントエンドとしての有効性を検証する。また、音声認識率をある程度反映可能な客観的歪み評価尺度を定義し、その妥当性を検証する。そして、客観評価尺度を用いて、提案法と従来法の音声認識のフロントエンドとしての有効性について調査検討する。提案法と従来法との性能比較に関しては、突発性雑音の除去能力差について検討するとともに、これまで同種の実験でよく用いられてきた定常雑音に対し、広く知られている遅延和アレーを比較対象アルゴリズムとすることにより、提案法の音声認識のフロントエンドとしての有効性を示す。

第 6 章： 聴感上の歪み感低減能力の検証

補聴器など聴感上の歪み感が問題となる機器への雑音除去技術の導入を想定し、提案法の聴感上の歪み感低減能力について検証する。

本章においても第 5 章と同様に、聴感上の歪み印象と相性のよい客観的的歪み評価尺度を作成し、効率的に雑音除去アルゴリズムの評価を行なう。そして、客観評価尺度あるいは聴取実験により、聴感上の歪み感低減能力に関して雑音除去アルゴリズムの評価を行なう。本章においても音声区間中に発生する非定常雑音に対し、提案法と従来法との性能比較を行なう。また、適応フィルタを利用した手法が苦手とするような目的信号と雑音がともに音声であるような状況において、提案法の雑音除去能力を検証する。

第 7 章： 実環境における有効性検証

本章では、残響や暗騒音が存在する実環境における本雑音除去アルゴリズムの有効性について検証する。

雑音除去実験は、残響時間が短く暗騒音は存在しない環境、一般的なオフィス環境、残響時間が長く暗騒音も存在する劣悪な環境において行なう。本章においても非定常雑音の除去能力に着目し、提案法と従来法との性能比較を行なう。

第 8 章： 結論

本論文で得られた結果を要約するとともに、今後検討すべき課題について述べる。

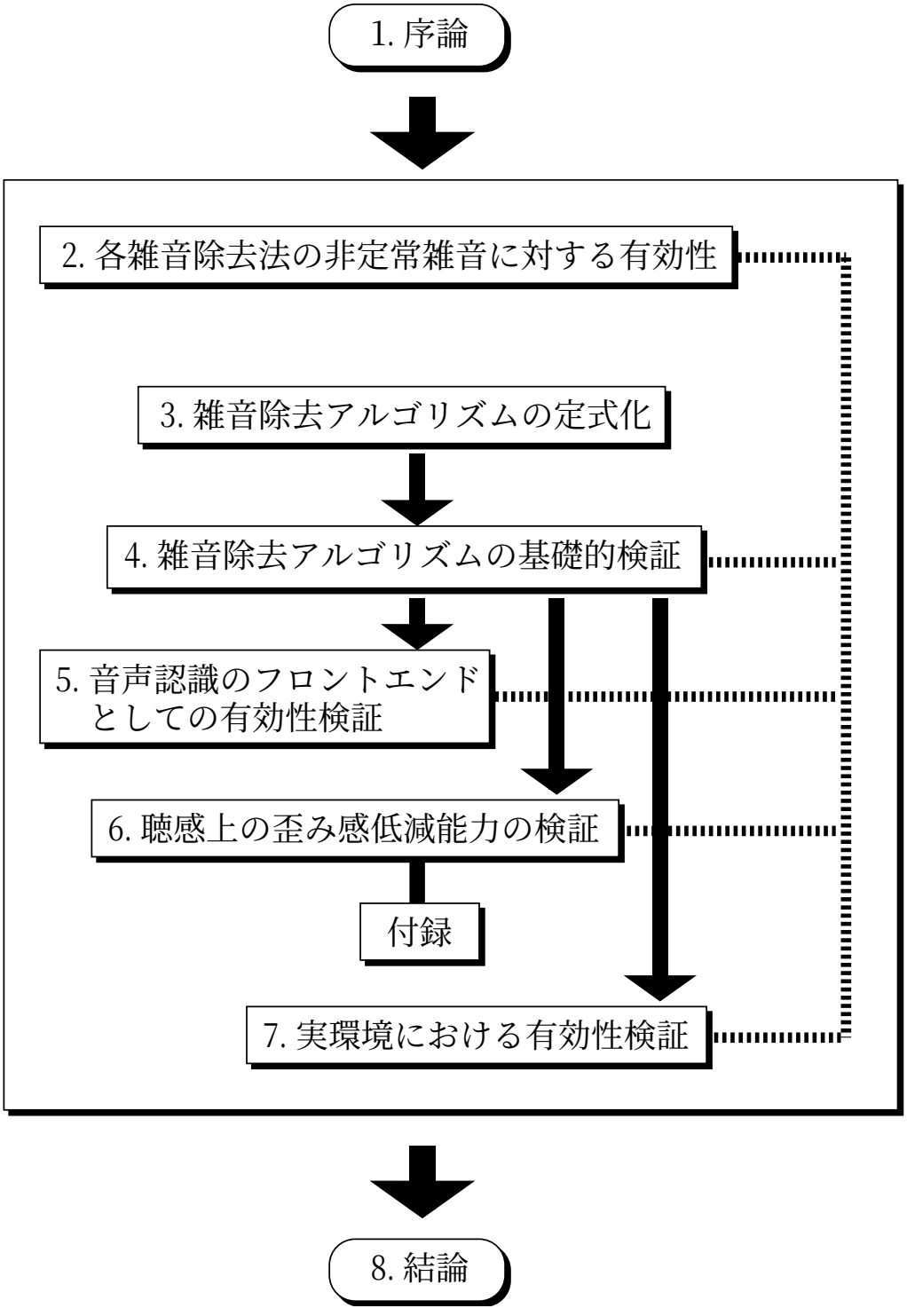
1. 序論
2. 各雑音除去法の非定常雑音に対する有効性
3. 雑音除去アルゴリズムの定式化
4. 雑音除去アルゴリズムの基礎的検証
5. 音声認識のフロントエンド
としての有効性検証
6. 聴感上の歪み感低減能力の検証
付録
7. 実環境における有効性検証
8. 結論

図 1.1: 各章の相互関係

第 2 章

各雑音除去法の非定常雑音に対する有効性

2.1 はじめに

本章では、これまでに提案された雑音除去法のうち本研究と関連のある手法についてより詳細に検討し、それらの利点・欠点を明らかにする。特に、非定常雑音をどの程度除去できるのかに着目し、従来の雑音除去研究に対する本研究の位置付けを明確にする。

従来の雑音除去法は、いずれも定常雑音に対する有効性は示しているが、実環境に存在するような非定常雑に対応できているとは言い難い。非定常雑音への対応を試みた雑音除去法も提案されているが、それらのほとんどは非音声区間に存在していた雑音の特性は滑らかに変化すると仮定している。我々の生活環境には、空調機のノイズのような定常的な雑音も存在するが、一方、ドアの開閉音や電話の呼出音のような突発的な非定常雑音も数多く存在する。目的音が音声である場合、第三者の音声は非定常で目的音と非常に相関の高い雑音となる。従って、従来の雑音除去法は、このような非定常雑音を除去できないだけでなく、雑音により生じた歪みを増大させてしまうことも考えられる。本章では、計算機シミュレーションにより、このような従来法の問題点を明確にする。

本研究では、3 ch 等間隔直線配列マイクロホンアレーにより雑音スペクトルを推定し、SS による雑音スペクトルの除去を行なう。そこで、本章では、単点受音の雑音除去法として SS とその改良法、多点受音の雑音除去法として 3ch 遅延和アレーと 3ch 適応 (減算型) アレーの非定常雑音の除去能力について検討する。

2.2 スペクトルサブトラクション

Boll が提案した SS [11] は、雑音は定常であると仮定し、音声と雑音とが加算された受音信号の振幅スペクトルから、非音声区間で推定した雑音の振幅スペクトル成分を引き去る手法である。つまり、定常的な雑音に対しては有効であるが、雑音スペクトルが時間とともに変化する場合には全く対応できない。また、雑音環境下での音声／非音声区間の判定も必要となる。そこで、音声／非音声区間の判定機構を排除し、非定常雑音への対応を目指した SS 改良法が提案されている。

2.2.1 単点受音雑音適応型 SS

谷口らは、非音声区間において推定した雑音の振幅スペクトルの初期値 $N_{i,k}$ に対して

$$N_{i,k+1} = N_{i,k} + \mu X_{i,k} E_{i,k} \tag{2.1}$$

により、雑音の振幅スペクトルを入力環境雑音に追従するように更新する手法を提案した [12]。ここで、k は分析フレームを表す添字、i は周波数を表す添字、μ は雑音推定の追従速度を決める定数、$X_{i,k}$ は入力信号の振幅スペクトル、$E_{i,k}$ は雑音除去後の振幅スペクトルである。この SS 改良法は、音声／非音声区間の判定を行なう必要がないという利点はあるが、雑音の準定常性を前提としている。谷口らは、自動車内や計算機室のようなある程度定常的な雑音には効果があるが、幹線道路や人混みの雑音では効果は少ないと報告している [12]。

ここで、式 (2.1) により雑音推定を行なう単点受音雑音適応型 SS の有効性を検証する。ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/ に、雑音として線形対数 sweep 音を計算機上で付加した音声を作成する。雑音として用いる線形対数 sweep 音は、信号長は音声の信号長と等しく 200 msec とし、始端周波数 1000 Hz、終端周波数 1100 Hz の変化の緩やかな場合と、始端周波数 1000 Hz、終端周波数 2000 Hz の比較的速く変化する場合の 2 通りとする。これらの雑音付加音声に対し、単点受音雑音適応型 SS により雑音除去を行なった結果を図 2.1 に示す。雑音推定の追従速度を決める式 (2.1) のパラメータ μ は、文献 [12] に従い 10^{-7} とした。

図 2.1 より、単点受音雑音適応型 SS は、変化の緩やかな線形対数 sweep 音は除去できることがわかる。しかし、変化の激しい線形対数 sweep 音に対しては、単点受音雑音適応型 SS は雑音スペクトルを正確に推定できず、雑音を除去できないばかりでなく、音声の振幅スペクトルに新たな歪みを生じさせている。以上より、式 (2.1) により雑音推定を行なう単点受音雑音適応型 SS は、音響的性質が急速に変化するような雑音や、音声区間中において発生する突発性雑音などには対応できないことがわかる。

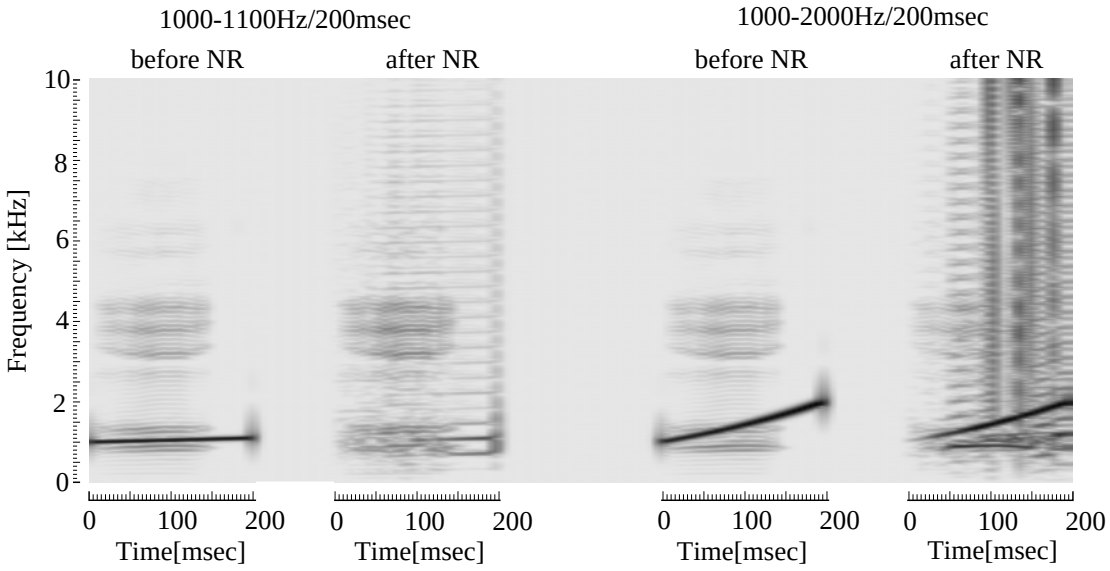

図 2.1: 単点受音雑音適応型 SS による雑音除去結果 (左から、変化の緩やかな線形対数 sweep 音に対する雑音除去前、雑音除去後、変化の激しい線形対数 sweep 音に対する雑音除去前、雑音除去後信号のスペクトログラム)

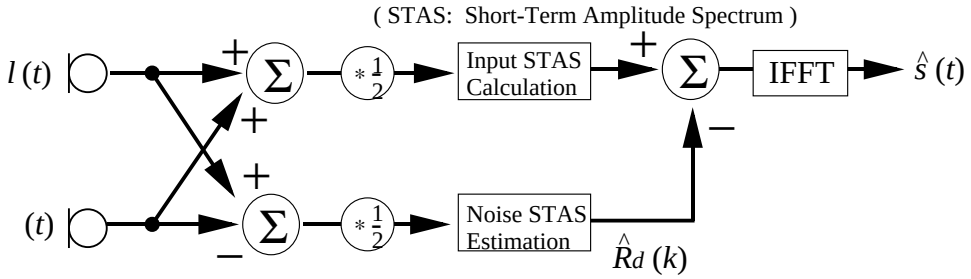

図 2.2: 2ch 雑音推定 SS のブロックダイアグラム

2.2.2 2ch 雑音推定 SS

非定常雑音の除去精度は、雑音推定能力に大きく依存する。単点受音の場合と比較し、多点受音の場合には信号到来方向という空間的な情報も得ることができる。そこで、多点受音信号を利用して雑音スペクトルを推定し、SS により雑音除去を行なう手法も提案されている。

金らは、短時間振幅スペクトル推定法とビームフォーマの利点を組み合わせて雑音の振幅スペクトルを推定する SS 改良法を提案した [16]。この 2ch 雑音推定 SS は、図 2.2 に示すように、2 つの受音信号の差信号から雑音スペクトルを推定する。

雑音の短時間振幅スペクトル $|R_d(\omega)|$ の推定では、まず非音声区間の受音信号を $|\hat{R}_d(\omega)|$ とする。そして、その後の音声区間においては、雑音源から受音点までの伝達関数 $H(\omega)$ を用いて

$$|\hat{R}_d(\omega)| = \sqrt{|R(\omega) - L(\omega)|^2/|H(\omega)|^2} \tag{2.2}$$

により $|\hat{R}_d(\omega)|$ を推定し、目的信号の短時間パワースペクトルを

$$|\hat{S}(\omega)|^2 = |R(\omega) + L(\omega)|^2 - |\hat{R}_d(\omega)|^2 \tag{2.3}$$

により求める。但し、目的信号は正中面方向から到来し、雑音源は移動しないと仮定する。この仮定により、雑音源から受音点までの伝達関数 $H(\omega)$ は線形時不変となり、雑音の短時間振幅スペクトル $|\hat{R}_d(\omega)|$ の推定が可能となる。

ここで、式 (2.2) により雑音推定を行なう 2ch 雑音推定 SS の有効性を検証する。目的

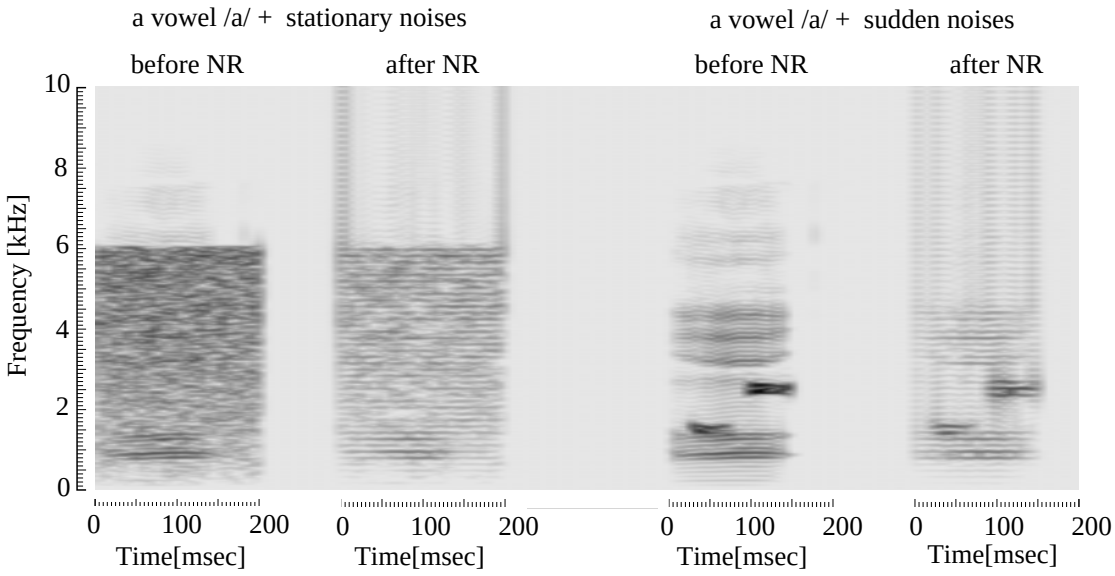

図 2.3: 2ch 雑音推定 SS による雑音除去結果 (左から、ランダム帯域雑音に対する雑音除去前、雑音除去後、突発性雑音に対する雑音除去前、雑音除去後信号のスペクトログラム)

信号は、正中面方向から到来する ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/ とする。雑音は、右 30 °方向から到来する定常的な 125Hz - 6kHz のランダム帯域雑音、あるいは音声区間中で発生する突発的な狭帯域雑音とする。これらを計算機上で付加した雑音付加音声に対し、2ch 雑音推定 SS により雑音除去を行なった結果を図 2.3 に示す。

図 2.3 より、2ch 雑音推定 SS は、非音声区間と音声区間とに等しく存在する定常的なランダム帯域雑音はある程度除去できている。しかしながら、雑音除去を行なった結果、目的信号である音声のホルマント構造も不明瞭になっている。突発性雑音の除去に関しては、図 2.3 より雑音のパワーは確かに小さくなっていることが確認できる。しかしながら、定常雑音の場合と同様に音声のパワーも小さくなっており、また雑音を正確に推定できていないために雑音の周波数帯域が広がっている。以上より、2ch 雑音推定 SS は、音源の移動に対して厳しい制約を設けており、雑音推定精度が低いため、実環境において有効な手法とは言い難い。

これまでに提案されてきた非定常雑音にも対応可能な SS 改良法は、雑音の準定常性を必要とし、典型的な非定常雑音である突発性雑音などへは対応できていない。ただし、雑音適応あるいは雑音推定を行なうことにより、非常に困難な問題である雑音環境中における音声／非音声区間判定を不必要としたことは評価に値する。

2.3 遅延和アレー

遅延和アレーは、図 2.4 に示すように、各受音信号に遅延を与え、それらに係数を乗じて加算することにより目的信号を強調する。図 2.4 において、目的信号が θ 方向から 3ch 遅延和アレーに到来する場合を考える。音速を c、隣接マイクロホン間距離を d とすると、受音信号に与えるべき遅延量 τ_1 は

$$\tau_1 = \frac{d \sin\theta}{c} \tag{2.4}$$

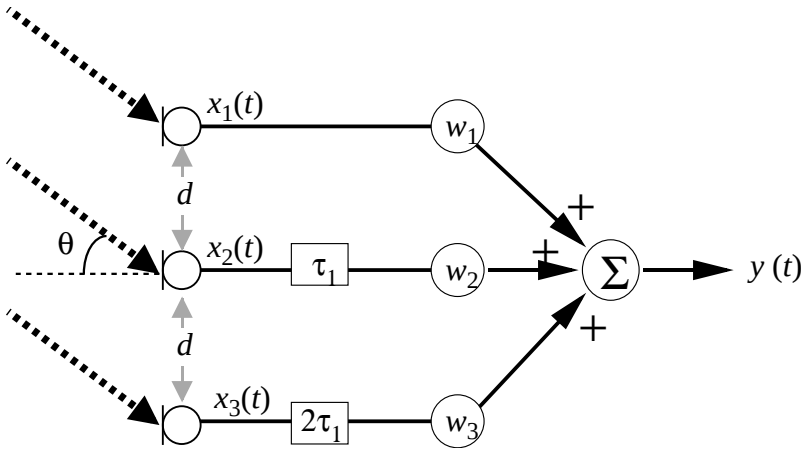

図 2.4: 3ch 遅延和アレーのブロックダイアグラム

となり、

$$\hat{y}(t) = \sum_{i=1}^{3} w_i x_i(t - (i-1)\tau_1) \tag{2.5}$$

により、雑音成分を抑圧し目的信号の推定値 $\hat{y}(t)$ が得られる。ここで、

$$\sum_{i=1}^{3} w_i = 1 \tag{2.6}$$

より $w_i = 1/3$ とする。

目的信号が正中面方向 $(\theta = 0)$ から到来する場合、3ch 遅延和アレーは図 2.5 に示すような指向特性を形成する。500 Hz 以下の低周波数帯域では、主ローブ幅が非常に広いため、雑音成分の除去、すなわち目的信号の強調は非常に困難である。つまり、音声のパワーが集中している低周波数帯域における雑音除去精度は低い。一般に、遅延和アレーにおいて鋭い指向特性を得るためには、素子数と素子間隔の積を大きくすることが必要である [37]。また、高周波数帯域では、目的信号の到来方向の主ローブ以外にもサイドローブが生じる。これは、空間エイリアジングと呼ばれる現象であり、空間的標本化定理

$$d < \frac{c}{2f} \tag{2.7}$$

を満足しない周波数 f において生じる [37]。

遅延和アレーは、雑音の到来方向に無関係にビームフォーミングを行なうため、雑音の音響的性質に対して制約を設けることなく、残響のようにあらゆる方向から到来する雑音

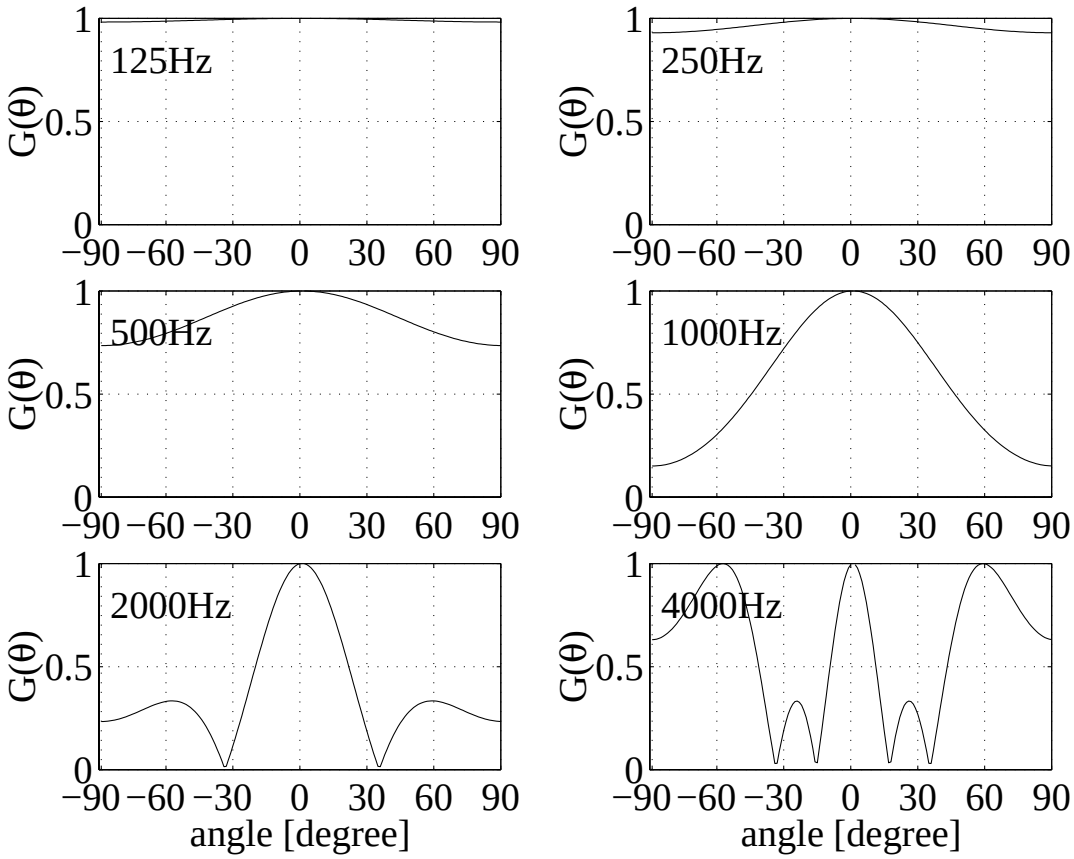

図 2.5: 3ch 遅延和アレーの指向特性 (左上から順に、125 Hz、250 Hz、500 Hz、1000 Hz、2000 Hz、4000 Hz の指向特性)

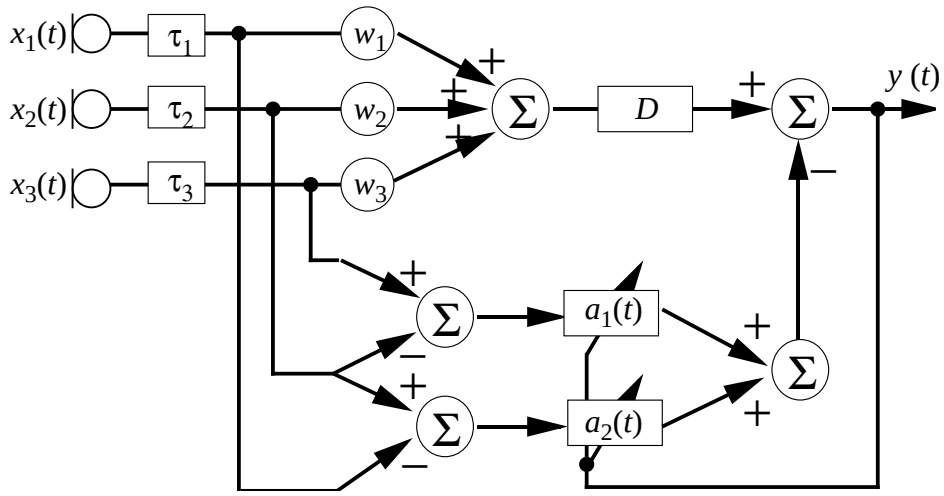

図 2.6: 3ch Griffiths-Jim 適応型アレーのブロックダイアグラム

にも有効な手法である。つまり、遅延和アレーは、高精度の雑音除去を実現するためには莫大な素子数を必要とし、他の手法と比較すると特に低周波数帯域における雑音除去能力は低いが、非定常雑音や残響のある環境下において有効な雑音除去法である。また、遅延和アレーは線形演算によるビームフォーマであるため、雑音除去音声に不快な歪みを生じないことも大きな利点である。

2.4 減算型アレー

適応減算型アレーの代表として、3ch Griffiths-Jim 適応型アレーのブロックダイアグラムを図 2.6 に示す。各受音信号は、遅延和アレーと同様に適当な遅延を与えられた後、適応フィルタを通して統合される。Griffiths-Jim 型アレーは、目的信号の信号源からアレー出力までの周波数特性を平坦に保ちながら、雑音のパワーが最小になるように適応フィルタのフィルタ係数を定める。本節における雑音除去の計算機シミュレーションでは、文献 [26] に忠実に実装したアルゴリズムを用いる。

適応型アレーは、一般に急激な環境の変化に追従することが困難であり、目的信号と雑音との相関が高い場合には適応フィルタの収束が困難になり雑音除去性能が低下する。そ

こで、突発性雑音をどの程度除去できるのか、また目的信号と雑音がともに音声信号である場合の雑音除去能力について調査する。目的信号は、正中面方向から到来する ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/ とする。雑音は、右 30 °方向から到来する突発性雑音、あるいは同一話者発話の単母音 /i/ の 2 種類とする。

図 2.7 に、雑音除去前後の信号の対数振幅スペクトル包絡を示す。図 2.7 より、3ch Griffiths-Jim 適応型アレーは、全般的に雑音成分を除去することは可能であるが、突発性雑音の立ち上がり部を除去することは困難であると言える。これは、適応型アレーが急激な環境の変化には追従できないという問題点を顕著に示す例である。参考のため、図 2.1 に示した音声に変化速度の異なる 2 種類の線形対数 sweep 音を付加したデータに対し、Griffiths-Jim 型アレーで雑音除去を行なった結果を図 2.8 に示す。雑音スペクトルの変化が速い場合の雑音除去結果 (図 2.8 右端) を見ると、適応型アレーは雑音の音響的性質の変化に追従できないため、雑音除去処理により新たな歪みを生じている。

図 2.9 に、単母音 /a/ と単母音 /i/ の混合信号である二重母音から、単母音 /i/ を雑音とみなして単母音 /a/ の分離抽出を行なった結果を示す。Griffiths-Jim 適応型アレーにより分離抽出された信号 (図 2.9 下図) を見ると、単母音 /i/ の第 2 ホルマント (F2) は除去できているが、第 1 ホルマント (F1) が全く除去できていない。このため、雑音除去後の音声を単母音 /a/ と聴き取ることはできなかった。この要因として、/a/ の第 1 ホルマント周波数と /i/ の第 1 ホルマント周波数が比較的近い、つまり目的信号と雑音の相関が高いことが挙げられる。

以上より、適応型の減算ビームフォーマは、目的信号と雑音が無相関であり、雑音の音響的特性が緩やかに変化する状況においては有効であるが、これらの制約条件が満たされない実環境においては有効である保障はない。また、適応フィルタの係数が一旦不適切な値に設定されると、後々までその影響が及ぶことから、適応型アレーは雑音を除去できないばかりか目的信号に新たな歪みまでも生じさせてしまうこともわかった。

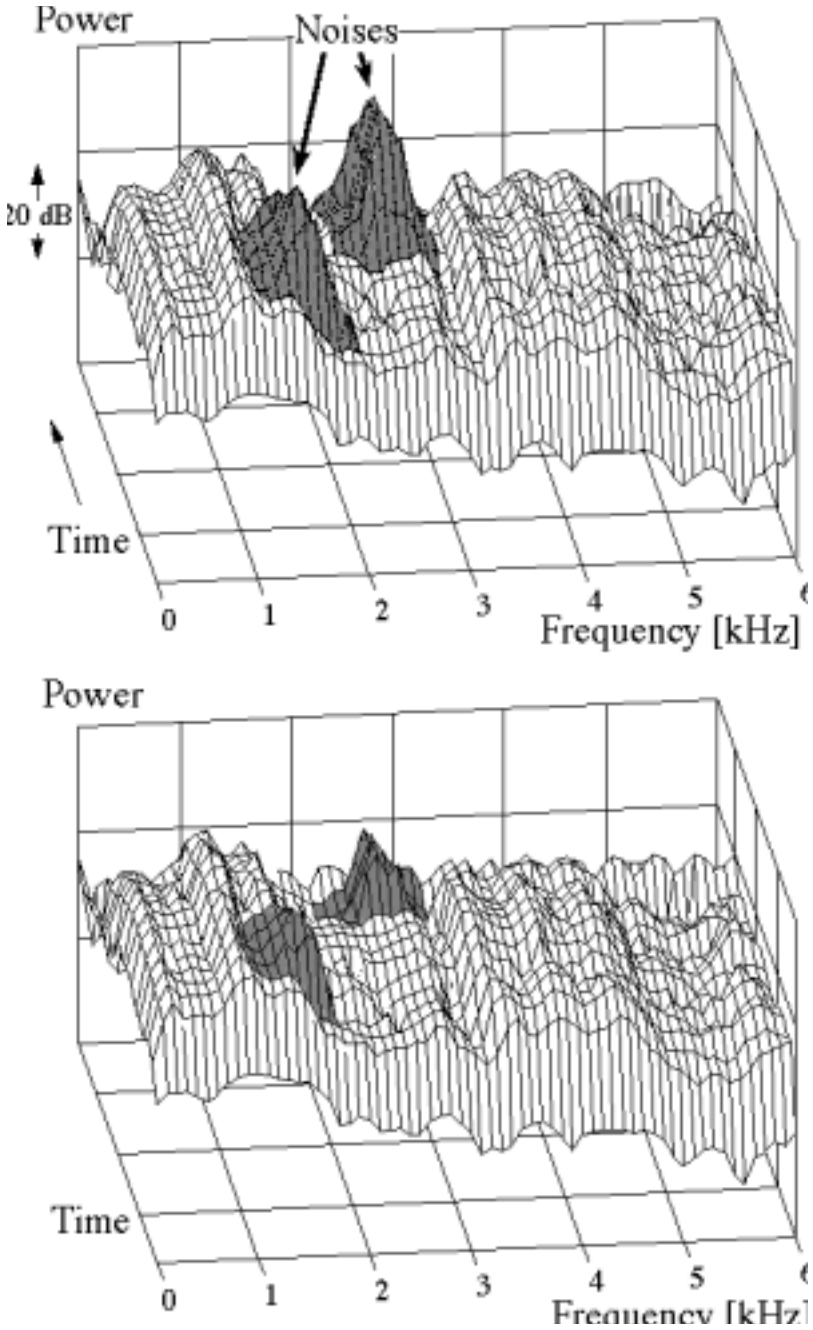

図 2.7: Griffiths-Jim 適応型アレーによる非定常雑音の除去結果 (上図は突発性雑音を付加した単母音 /a/、下図は雑音除去後の単母音/a/)

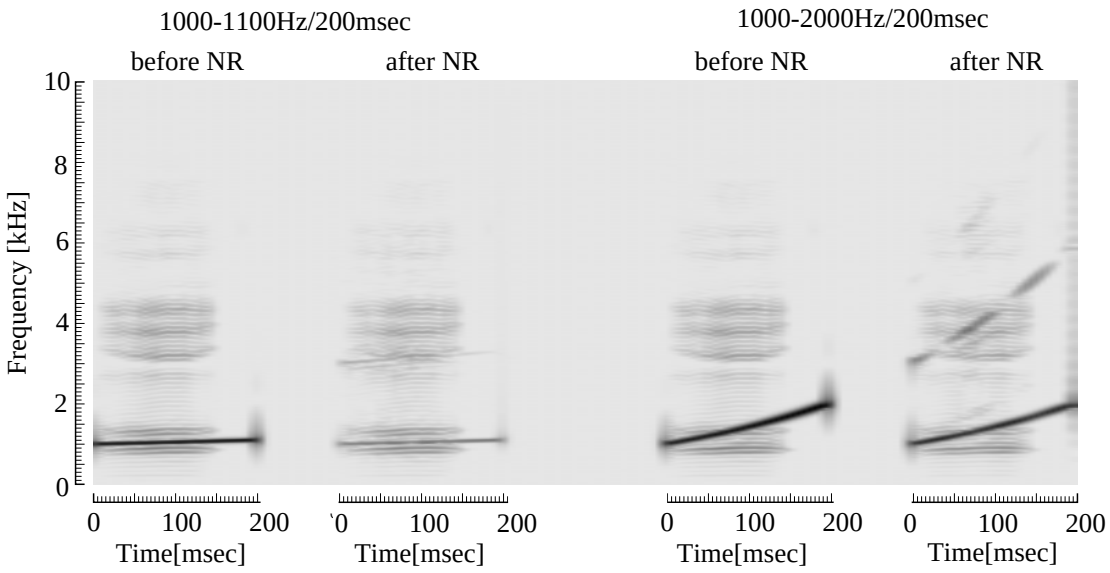

図 2.8: Griffiths-Jim 型アレーによる雑音除去結果 (左から、変化の緩やかな線形対数 sweep 音に対する雑音除去前、雑音除去後、変化の激しい線形対数 sweep 音に対する雑音除去前、雑音除去後信号のスペクトログラム)

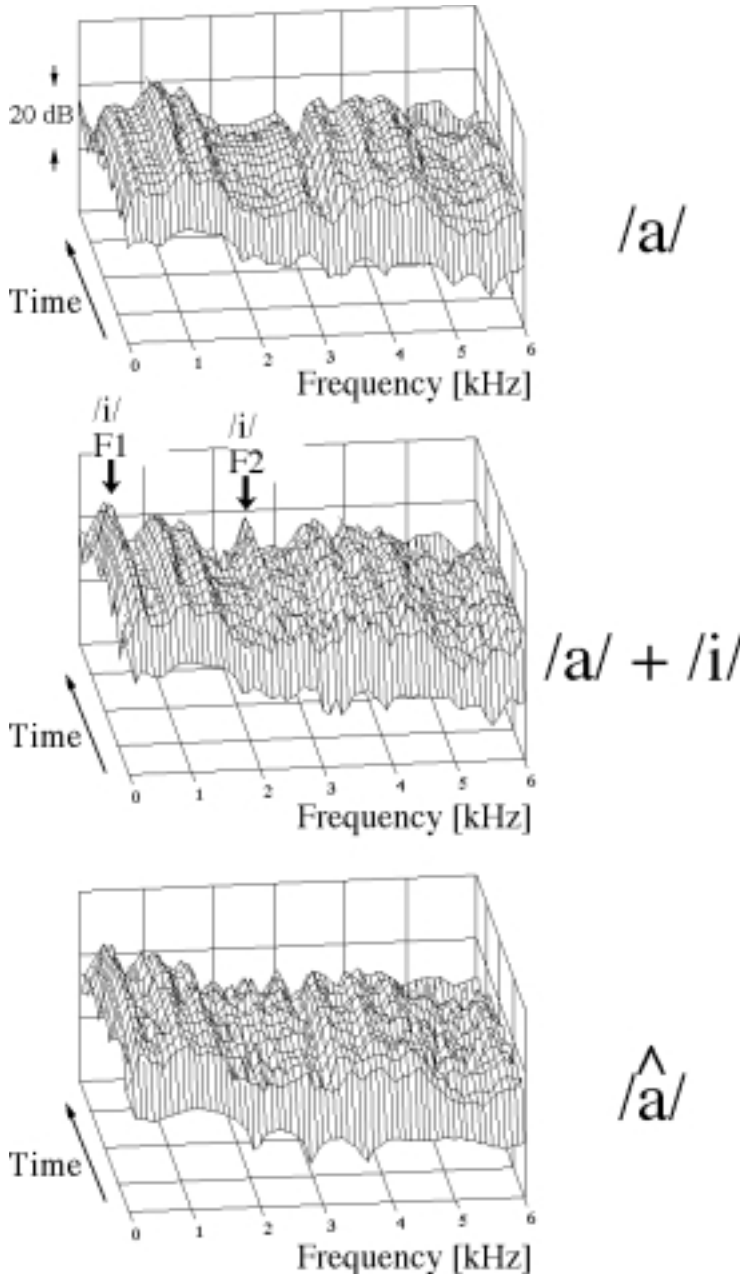

図 2.9: Griffiths-Jim 適応型アレーによる二重母音の分離抽出 (上から順に、クリーンな単母音 /a/、単母音 /a/ と単母音 /i/ の混合信号、分離抽出した単母音 /a/)

表 2.1: 従来法の雑音適応表

雑音除去アルゴリズム (素子数)	SS (1)	適応 SS (1)	2ch SS (2)	遅延和アレー (多数)	適応型アレー (少数 (3))
定常雑音	◎	◎	○	○	◎
音源が移動する定常雑音	○	○	×	○	○
緩変化の非定常雑音	×	○	△	○	○
急激に変化する非定常雑音	×	×	×	○	×
音声区間中で発生する雑音	×	×	×	○	△
音源が移動する非定常雑音	×	△	×	○	○

2.5 まとめ

本章では、計算機シミュレーションにより、従来の雑音除去法が非定常雑音をどの程度除去できるのかを検証した。各雑音除去アルゴリズムの雑音適応性を表 2.1 に示す。

遅延和アレーは、雑音に関する制約を全く設けないために非定常雑音にも対応できるが、素子数が少ない場合には低周波数帯域における雑音除去能力は非常に低い。表 2.1 における遅延和アレーは、信号の到来方向を既知とし、十分な素子数を持つものと仮定する。表 2.1 は、この最適化された遅延和アレーを基準とし、他の手法の雑音適応性を表す。

遅延和アレーと比較すると、SS 改良法あるいは適応型アレーは、雑音に対する制約を設けることにより、素子数は少なくとも高精度の雑音除去を目指した手法と言える。しかし、制約条件が実環境と大きく掛け離れる状況が多々あり、またこれらの手法は多かれ少なかれ過去に推定した雑音スペクトルを用いるため、雑音を除去できないばかりでなく、歪みを増大、あるいは新たな歪みを生じさせることもあり得ることがわかった。

本研究では、本章で明らかになった従来法の問題点を踏まえ、突発性雑音などの時々刻々変化する非定常雑音の除去について検討する。第 3 章において、できる限り少ない素子数で、適応フィルタに依らず解析的に設計する減算型ビームフォーマと SS とを組み合わせた雑音除去アルゴリズムを構築する。

第 3 章

雑音除去アルゴリズムの定式化

3.1 はじめに

第 2 章において、従来の雑音除去法は、定常雑音あるいは変化の緩やかな非定常雑音は除去できるが、音響的特性が急激に変化する雑音や突発性雑音の除去は困難であることを確認した。本章では、突発性雑音のような非定常雑音の除去を実現するため、マイクロホン対による雑音除去アルゴリズムを構築する。

非定常雑音を除去するためには、時々刻々変化する雑音スペクトルを推定する必要がある。本研究では、雑音スペクトルの急激な変化への対応が困難な適応フィルタは利用せず、解析的に減算型ビームフォーマを設計する。このためには、雑音環境下で信号到来方向を推定する必要がある。減算型アレーにおいて、適応フィルタを用いたビームフォーミングが主流となった背景には、雑音環境における信号の到来方向推定が非常に困難であり、これまではこの問題を克服できなかったことがある。

本雑音除去アルゴリズムは、まずマイクロホン対により得られる空間情報を利用し、目的音および雑音の到来方向を推定する。ここで積極的に減算型ビームフォーマを利用することにより、雑音環境下での信号到来方向の推定精度を飛躍的に向上させることが期待できる。信号到来方向がわかれば、解析的にビームパターン (空間フィルタ) を設計することが可能になり、ビームパターンの死角を適切に制御することによって目的信号を完全に抑圧し、雑音成分を時々刻々推定することが可能になる。本雑音除去アルゴリズムは、周波数ごとに最適なマイクロホン対を選択し、より少ない素子数で高精度の雑音スペクトル推定を行なう。このため、少数のマイクロホンしか使用しなくとも高精度の雑音スペクトル推定が実現できる。そして、周波数領域におけるフィルタリング手法である SS を利用し、受音信号から時々刻々推定した雑音成分を除去することにより、非定常雑音の除去も可能となる。

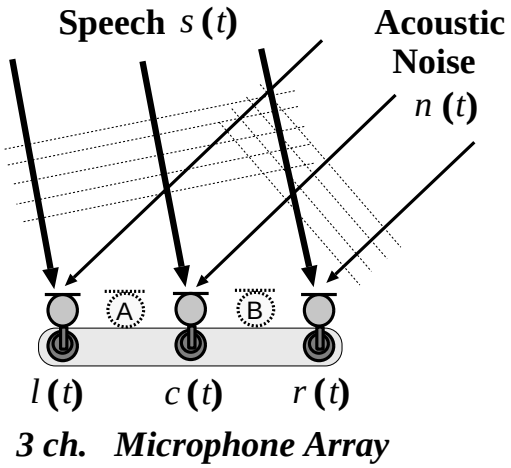

図 3.1: マイクロホンアレーと音声信号、雑音信号の相互関係

3.2 定式化

本雑音除去アルゴリズムでは、図 3.1 に示すような 3 本の無指向性マイクロホンを等間隔に直線配列したマイクロホンアレーを用いる。ここで扱う信号は、すべて平面波でマイクロホンアレーへ到来する、すなわち音源とマイクロホンアレーは十分離れていると仮定する。本仮定が成り立つための音源とマイクロホンアレー間の具体的な距離は、音速、サンプリング周波数 (マイクロホンアレーの空間分解能に相当) に依存して決定される。

ここで、目的信号を $s(t)$、除去対象とする加法性雑音を $n(t)$ とし、マイクロホンアレーの左端、中央、右端のマイクロホンで受音された信号をそれぞれ $l(t)$、$c(t)$、$r(t)$ とする。目的信号 $s(t)$ が、マイクロホンアレーの両端のマイクロホンから成るマイクロホン対（主対）へ到来する際の時間差を 2ζ とし、雑音 $n(t)$ の主対への到来時間差を 2δ とする。

以上より、受音信号 $l(t)$、$c(t)$、$r(t)$ と音声 $s(t)$、雑音 $n(t)$ との間には

$$l(t) = s(t-\zeta) + n(t-\delta) \tag{3.1}$$

$$c(t) = s(t) + n(t) \tag{3.2}$$

$$r(t) = s(t+\zeta) + n(t+\delta) \tag{3.3}$$

の関係が成り立つ。

受音信号 $l(t)$、$c(t)$、$r(t)$、および音声 $s(t)$、雑音 $n(t)$ の Fourier 変換をそれぞれ $L(\omega)$、$C(\omega)$、$R(\omega)$、$S(\omega)$、$N(\omega)$ とすると、式 (3.1)、(3.2)、(3.3) はそれぞれ

$$L(\omega) = S(\omega)e^{-j\omega\zeta} + N(\omega)e^{-j\omega\delta} \tag{3.4}$$

$$C(\omega) = S(\omega) + N(\omega) \tag{3.5}$$

$$R(\omega) = S(\omega)e^{j\omega\zeta} + N(\omega)e^{j\omega\delta} \tag{3.6}$$

となる。式 (3.4)、(3.5)、(3.6) の関係をもとに、以下の手順で雑音除去を行なう。

1. 3 本のマイクロホンの受音信号を用いて、目的信号および雑音の到来方向を推定する。
2. 3 本のうちの 2 本のマイクロホン (マイクロホン対) で受音した信号を用いて、中央のマイクロホン位置における雑音スペクトルを推定する。
3. 中央のマイクロホンでの受音信号から、先に推定した雑音成分を引き去ることにより雑音除去を行なう。

本章では、目的信号および雑音の到来方向推定、雑音スペクトルの推定、雑音除去の順に定式化を行なう。

3.2.1 目的信号および雑音の到来方向推定

本節では、短時間フレームごとに目的信号および雑音の 2 方向を推定するアルゴリズムを構築する。本アルゴリズムは、受音信号の相互相関に基づき方向推定を行なうが、方向推定にも減算型ビームフォーマを導入することにより、雑音環境下での 2 つの信号到来方向の高精度推定を実現する。

相互相関を利用しない方向推定法として、例えば我々の聴覚系を考慮した信号の立ち上がり部分に着目した方向推定法 [55] などがある。これは、残響のある環境において、音源が複数存在する場合であってもそれらの方向を推定できる魅力的な手法である。しかし、方向推定の手がかりは、あくまで音の立ち上がり部分であり、レベル変動の少ない定常的な信号の到来方向を推定／追尾することは困難である。

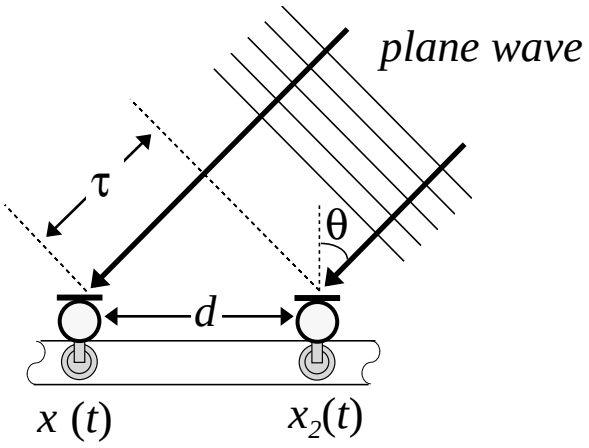

図 3.2: 平面波に対する到来時間差推定

信号が平面波としてマイクロホンアレーへ到来する場合、信号の到来方向の推定は、信号が 2 つのマイクロホン (マイクロホン対) へ到来する際の時間差を推定することに相当する。図 3.2 に示すように、信号はマイクロホンアレーの正中面に対して θ 方向から到来し、d だけ離れた 2 つのマイクロホンでの受音信号をそれぞれ $x_1(t)$、$x_2(t)$ とする。このとき、2 つのマイクロホンへの信号の到来時間差を τ、音速を c とすると

$$x_2(t) = x_1(t - \tau) \tag{3.7}$$

$$\tau = \frac{d \sin\theta}{c} \tag{3.8}$$

の関係が成り立つ。従って、信号のマイクロホン対への到来時間差 τ が推定できれば

$$\theta = \arcsin(\frac{\tau c}{d}) \tag{3.9}$$

により、信号の到来方向 θ を求めることができる。

信号の到来時間差の推定に関しては、2 つの受音信号の相互相関関数を計算し、その最大値を与えるタイムラグを信号の到来時間差とする [56]。本時間差推定法は、受音信号の周波数帯域が広いほど、また受音信号のデータ長が長いほど推定精度が高くなる [37]。受音信号 $x_1(t)$、$x_2(t)$ の Fourier 変換をそれぞれ $X_1(\omega)$、$X_2(\omega)$ とすると、

$$\phi_{x_1 x_2}(t) = \mathrm{IFFT}\left[\frac{X_1(\omega)X_2^*(\omega)}{|X_1(\omega)||X_2(\omega)|}\right] \tag{3.10}$$

により白色化相互相関関数 $\phi_{x_1x_2}(t)$ が計算でき

$$\tau = \underset{t}{\operatorname{argmax}} [\, \phi_{x_1x_2}(t) \,] \tag{3.11}$$

によって信号到来時間差 τ が得られる。式 (3.10) のように、クロススペクトルを受音信号のパワーで正規化することにより、信号の振幅スペクトル形状に依存せず、より高精度の時間差推定が可能になる [57]。

本時間差推定法は、短時間フレームごとに、目的信号と最も優勢な雑音 (パワー最大の雑音) の到来時間差を以下の手順により推定する。

STEP 1: [雑音除去 (方向推定) の必要性判定]

雑音除去を行なうか否かの判定には、マイクロホンアレーの中央マイクロホンの受音信号を用いる。受音信号の平均パワーが閾値を越えた場合に限り、そのフレームにおいて信号の到来時間差推定、雑音除去処理を行なう。この閾値は、計算機シミュレーションや実環境における実験において使用するデータを考慮し、適宜適切な値に設定した。

STEP 2: [一方の信号の到来時間差推定]

受音信号に含まれる様々な信号の中から、最もパワーの大きな信号の隣接マイクロホン間への到来時間差を推定する。これは、目的信号が最もパワーが大きい場合には目的信号の到来時間差となり、そうでない場合には除去対象となる雑音の到来時間差となる。

一般に、マイクロホン間隔が大きいほど高精度の到来時間差推定が可能であるため、ここでは主対での受音信号 $L(\omega)$、$R(\omega)$ を用い、白色化相関法により最もパワーの大きな信号の到来時間差を推定する。但し、式 (3.12) において使用する信号は、100 Hz から 6 kHz までの音声帯域に限定する。相互相関関数 $\phi_{lr}(t)$ を用いて式 (3.13) より、隣接マイクロホン間への到来時間差 d_1 を推定する。

$$\phi_{lr}(t) = \mathrm{IFFT}\left[\frac{L(\omega)R^*(\omega)}{|L(\omega)||R(\omega)|}\right] \tag{3.12}$$

表 3.1: 信号の到来時間差と音源方向との関係 (サンプリング周波数 48 kHz の場合)

時間差 [pt]	方向 [°]	時間差 [pt]	方向 [°]
1	4	8	35
2	8	9	40
3	12	10	45
4	16	11	51
5	21	12	58
6	25	13	67
7	30	14	83

$$2d_1 = \underset{t \in C_1}{\operatorname{argmax}} [\ \phi_{lr}(t)\ ] \tag{3.13}$$

方向推定に用いたマイクロホンの間隔は、隣接マイクロホン間隔の 2 倍であるため、隣接マイクロホン間への到来時間差 d_1 は式 (3.13) の右辺の 1/2 となる。

式 (3.13) における探索範囲 C_1 は、前フレームで推定した 2 つの信号の到来時間差の近傍、具体的には到来時間差の変化を 2pt (1pt はサンプリング周期) 以内とする。これは、音源は高速では移動しないという仮定に基づく制約条件である。方向推定を行なう際のフレーム長を 21.3 msec (サンプリング周波数が 48 kHz の場合は 1024 pt)、フレーム周期を 10.6msec (512 pt) とすると、本制約条件のもとでは正中面方向で 0.52 rad/sec、つまり、マイクロホンアレーから 1m 離れた地点を 48 km/h で通過する物体まで方向推定／追尾可能である。よって、雑音除去を必要とする現実的な環境においては、本制約条件が障害になるとは考え難い。サンプリング周波数を 48 kHz とした場合、信号の隣接マイクロホンへの到来時間差と音源方向との間には、表 3.1 に示すような関係が成り立つ。なお、前フレームにおいて時間差推定を行わなかった場合、本フレームにおいて新たに音源が生じたと考える。従って、式 (3.13) では t の探索範囲 C_1 には制約を設けず、音源方向の探索範囲は全方位とする。

STEP 3: [もう一つの信号の到来時間差推定]

受音信号に含まれる信号のうち、2 番目にパワーの大きな信号の隣接マイクロホンへの到来時間差を推定する。式 (3.13) で推定した到来時間差 d_1 が目的信号の到来時間差であった場合、ここでは除去対象となる雑音の到来時間差を推定することとなる。そうでなければ、目的信号の到来時間差を推定する。

相互相関による方向推定法は、信号がただ一つ存在する状況であれば、その到来方向を正確に推定することが可能である。しかし、同程度のパワーを持つ複数の信号が共存する環境においては方向推定精度は大幅に低下する。この問題を解決するために、本方向推定アルゴリズムでは、**STEP 2** において推定した隣接マイクロホンに時間差 d_1 で到来する信号を完全に抑圧し、それ以外の信号成分に対して到来時間差を推定する。

まず、副対での受音信号 $l(t)$、$c(t)$ を用いて、信号の到来時間差が d_1 である方向にマイクロホンアレーの指向性の死角 (零点) を形成するようなビームフォーマ $g_{lc}(t)$：

$$g_{lc}(t) = \frac{\{l(t+d_1+\tau) - l(t+d_1-\tau)\} - \{c(t+\tau) - c(t-\tau)\}}{4} \tag{3.14}$$

を考える。ここで、τ は 0 以外の任意の定数とする。ビームフォーマ $g_{lc}(t)$ は、d_1 によりマイクロホンアレーの死角、τ により焦点を制御することが可能である。マイクロホンアレーの焦点とは、ビームフォーマ $g_{lc}(t)$ により抽出する信号の到来方向を意味する。ここで、死角を正中面方向 ($d_1 = 0$)、焦点を 30 °方向 ($\tau = 14$) と設定した場合のビームフォーマ $g_{lc}(t)$ の指向特性を図 3.3 に示す。図 3.3 より、マイクロホンアレーの焦点方向では、各周波数において感度が等しく、雑音スペクトルを正確に推定できることがわかる。

前述のとおり、**STEP 2** の式 (3.13) で推定した到来時間差 d_1 は、目的信号の到来時間差 ζ か、雑音の到来時間差 δ のいずれかである。d_1 が目的信号の到来時間差である場合、すなわち $d_1 = \zeta$ の場合、式 (3.14) の Fourier 変換 $G_{lc}(\omega)$ は、式 (3.1)、(3.2)、(3.4)、(3.5) より

$$G_{lc}(\omega) = N(\omega)\, e^{j\omega\frac{\zeta-\delta}{2}} \sin\omega\tau\ \sin\omega\frac{\delta-\zeta}{2} \tag{3.15}$$

となり、目的信号 $S(\omega)$ に関する項は一切存在しない。また、$d_1 = \delta$ の場合、式 (3.14) の

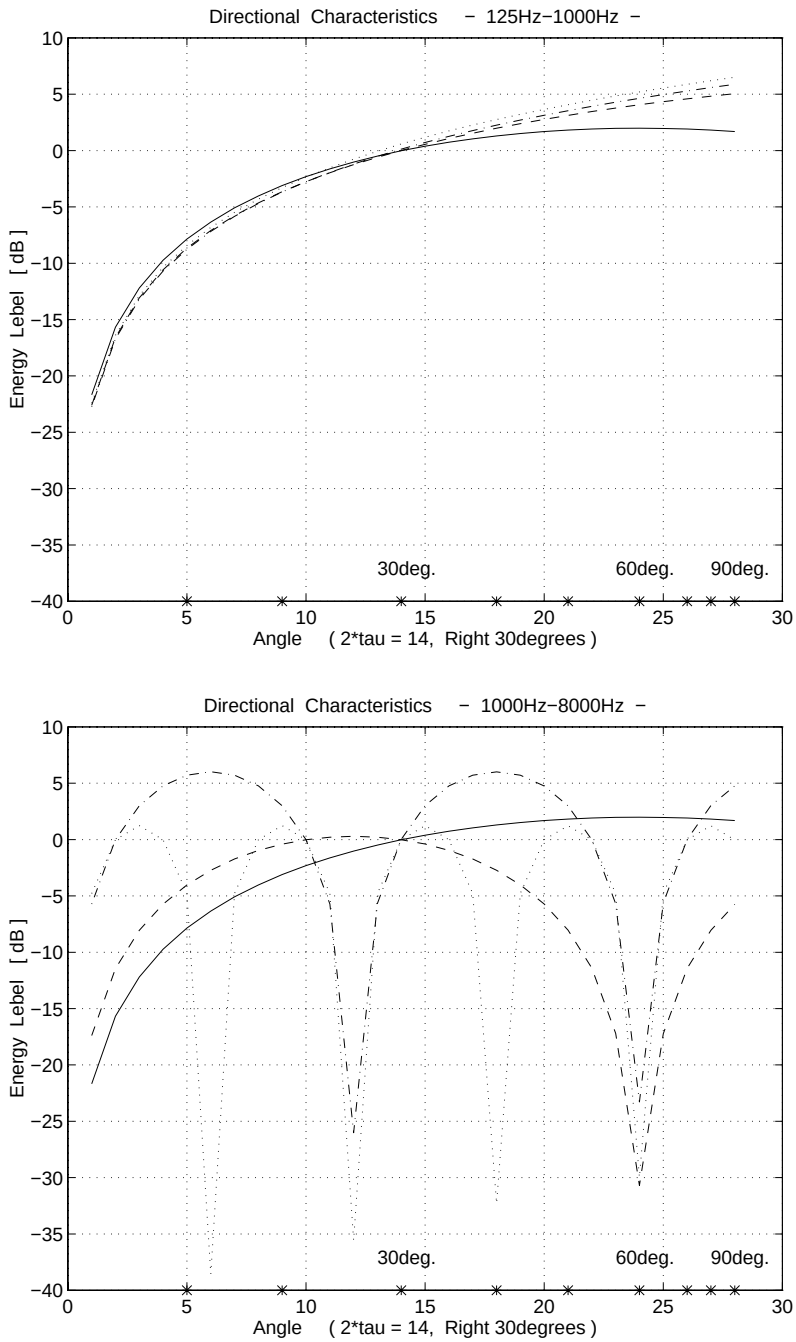

図 3.3: マイクロホン対による減算型ビームフォーマの指向特性 (死角 0 °、焦点 30 °の場合) [上図は低周波数特性を表し、実線は 1000 Hz、破線は 500 Hz、鎖線は 250 Hz、点線は 125 Hz の指向特性。下図は高周波数特性を表し、実線は 1000 Hz、破線は 2000 Hz、鎖線は 4000 Hz、点線は 8000 Hz の指向特性。]

Fourier 変換 $G_{lc}(\omega)$ は

$$G_{lc}(\omega) = S(\omega)\, e^{-j\omega\frac{\zeta-\delta}{2}} \sin\omega\tau\ \sin\omega\frac{\zeta-\delta}{2} \tag{3.16}$$

となり、雑音 $N(\omega)$ に関する項は一切存在しない。つまり、式 (3.14) は、図 3.1 の仮想マイクロホン **A** の受音信号のうち、時間差 d_1 で到来する信号を完全に抑圧した信号に相当する。同様に、副対での受音信号 $c(t)$、$r(t)$ を用いて、時間差 d_1 で到来する信号を完全に抑圧するような指向特性を持つ仮想マイクロホン **B** の受音信号 $g_{cr}(t)$ を求める。

次に、$g_{lc}(t)$ と $g_{cr}(t)$ の音声帯域の成分に対して白色化相互相関関数 $\phi_{g_{lc}g_{cr}}(t)$ を計算し、

$$d_2 = \underset{t\in C_2}{\operatorname{argmax}}\,[\ \phi_{g_{lc}g_{cr}}(t)\ ] \tag{3.17}$$

により受音信号のうち 2 番目にパワーの大きな信号の到来時間差 d_2 を推定する。探索範囲 C_2 は、パワーが最大の信号の到来時間差 d_1 の近傍 (具体的には 2pt 以内) を除くと定める。以上より、本方向推定アルゴリズムは、隣接マイクロホンに時間差 d_1 で到来する信号を完全に抑圧し、次にパワーの大きな信号の到来時間差 d_2 を推定することが可能になる。

STEP 4: [目的信号および雑音の到来時間差判定]

最後に、推定した 2 つの到来時間差 d_1、d_2 に対し、それぞれ目的信号の到来時間差 ζ であるか、雑音の到来時間差 δ であるかの判定を行なう必要がある。

ここでは、音源は滑らかに移動すると仮定し、d_1、d_2 は、それぞれ前フレームにおいて推定した目的信号および雑音の到来時間差の推定結果に近い方を選択する。雑音除去処理の開始時点では、目的信号あるいは雑音の少なくとも一方の到来時間差がわかっていなければならない。本アルゴリズムでは、処理の開始時点において、マイクロホンアレーの正中面により近い方向から到来する信号を目的信号と仮定する。目的信号の初期到来方向も含めて信号到来方向の自動推定を行なうことが望ましいが、そのためには目的信号あるいは雑音に関する何らかの知識を記憶しておく必要がある。

3.2.2 雑音スペクトルの推定

本節では、前節において推定した目的信号および雑音の隣接マイクロホンへの到来時間差を利用し、目的信号を完全に抑圧して雑音スペクトルを推定する。本研究では、減算型アレーの主流となっている適応フィルタによるビームフォーミングではなく、到来時間差を利用する解析的ビームフォーミング手法を提案する。

主対の受音信号 $l(t)$、$r(t)$ に、それぞれ $\zeta \pm \tau_1$、$-\zeta \pm \tau_1$ の遅延を与えた信号を加減算することにより、目的信号 $s(t)$ を完全に抑圧するようなビームフォーマ $g_{lr}(t)$：

$$g_{lr}(t) = \frac{\{l(t+\zeta+\tau_1) - l(t+\zeta-\tau_1)\} - \{r(t-\zeta+\tau_1) - r(t-\zeta-\tau_1)\}}{4} \tag{3.18}$$

を設計する。ここで、τ_1 は 0 以外の任意の定数である。目的信号 $s(t)$ を抑圧するだけであれば、時間のシフトに関しては $+\tau$ あるいは $-\tau$ のみで十分である。しかし、マイクロホンアレーの中央のマイクロホン位置における雑音のスペクトルを正確に推定するためには $\pm\tau$ の時間シフトを考慮する必要がある。これにより、振幅スペクトルのみならず、位相スペクトルも含め、厳密な雑音スペクトルの推定が可能になる。

次に、ビームフォーマ $g_{lr}(t)$、すなわち目的信号を完全に抑圧した信号から、雑音のスペクトルを推定する。式 (3.18) のビームフォーマ $g_{lr}(t)$ の Fourier 変換 $G_{lr}(\omega)$ は

$$G_{lr}(\omega) = N(\omega)\ \sin\omega(\delta-\zeta)\ \sin\omega\tau_1 \tag{3.19}$$

となる。式 (3.18) の τ_1 は任意定数であるため、先に推定した目的信号および雑音の隣接マイクロホンへの到来時間差 ζ、δ を利用して

$$\tau_1 = \delta - \zeta \tag{3.20}$$

と定める。このとき、式 (3.19) は

$$G(\omega) = N(\omega)\ \sin^2\omega(\delta-\zeta) \tag{3.21}$$

となり、雑音のスペクトル $N(\omega)$ は

$$N(\omega) = G(\omega)/\sin^2\omega(\delta-\zeta), \qquad \omega(\delta-\zeta) \neq n\pi \quad (n\text{: 整数}) \tag{3.22}$$

と推定できる。しかし、これでは雑音スペクトルのうち

$$\omega = \frac{n\pi}{\delta - \zeta} \qquad (n\text{: 整数}) \tag{3.23}$$

となる周波数成分は計算できず、その近傍では推定誤差が生じる。例えば、目的信号が正中面方向 ($\zeta = 0$)、雑音が正中面に対して 30 °方向 ($\delta = 146\mu$sec) から到来する場合、約 $3.4 \cdot n$ kHz (n: 整数) 付近が計算できない帯域に相当する。

そこで、式 (3.18) で定義した主対ビームフォーマで雑音を推定できない周波数成分に対しては、マイクロホンアレーの中央と右端のマイクロホンから副対での受音信号 $c(t)$、$r(t)$ を利用して推定する。副対ビームフォーマ $g_{cr}(t)$ は、式 (3.18) で設計した主対ビームフォーマと同様に

$$g_{cr}(t) = \frac{\{c(t+\tau_2) - c(t-\tau_2)\} - \{r(t-\zeta+\tau_2) - r(t-\zeta-\tau_2)\}}{4} \tag{3.24}$$

と設計する。ここで、τ_2 は任意定数であり、$\tau_2 = (\delta - \zeta)/2$ とすると、$g_{cr}(t)$ の Fourier 変換 $G_{cr}(\omega)$ は

$$G_{cr}(\omega) = N(\omega)\, e^{j\omega\frac{\delta-\zeta}{2}} \sin^2 \omega \frac{\delta - \zeta}{2} \tag{3.25}$$

となる。副対ビームフォーマ $g_{cr}(t)$ において、雑音を正確に推定できない周波数帯域は

$$\omega \approx \frac{2m\pi}{\delta - \zeta} \qquad (m\text{: 整数}) \tag{3.26}$$

となる。従って、副対ビームフォーマ $g_{cr}(t)$ を使用することにより、主対ビームフォーマ $g_{lr}(t)$ では雑音スペクトルを正確に推定できない周波数帯域のうち、n が奇数の場合に限っては雑音スペクトルの推定が可能になる。なお、n が偶数の場合、副対ビームフォーマ $g_{cr}(t)$ を使用しても雑音スペクトルを正確には推定できない。

このとき、雑音スペクトル $\hat{N}(\omega)$ は、各周波数帯域においてより高精度の推定を行なうため、2 つの閾値 ε_1、ε_2 を設けて式 (3.27) により推定する。

$$\hat{N}(\omega) = \begin{cases} G_{lr}(\omega) / \sin^2 \omega(\delta - \zeta), & \sin^2 \omega(\delta - \zeta) > \varepsilon_1 & \text{(a)} \\ G_{cr}(\omega)\, e^{-j\omega\frac{\delta-\zeta}{2}} / \sin^2 \omega\frac{\delta-\zeta}{2}, & \sin^2 \omega(\delta - \zeta) \le \varepsilon_1 \ and \ \sin^2 \omega\frac{\delta-\zeta}{2} > \varepsilon_2 & \text{(b)} \\ G_{lr}(\omega)/\varepsilon_2^2, & \sin^2 \omega\frac{\delta-\zeta}{2} \le \varepsilon_2 & \text{(c)} \end{cases} \tag{3.27}$$

つまり、$\sin\omega(\delta-\zeta)$ の値が十分大きい周波数帯域 (a) においては、マイクロホン間隔がより大きい主対ビームフォーマを用いて雑音スペクトルの推定を行なう。これは、マイクロホン間隔が大きいほど、より高精度の雑音スペクトルの推定が可能になるためである。$\sin\omega(\delta-\zeta)$ が 0 近づくに従って計算誤差が大きくなるため、そのような周波数帯域 (b) においては副対ビームフォーマを用いて雑音スペクトルを推定する。副対ビームフォーマを使用してもなお雑音スペクトルを推定できない周波数帯域 (c) に限って、雑音スペクトルを近似的に推定する。広帯域のランダム雑音を用いて雑音スペクトルの推定実験を行なった結果、式 (3.27) (c) に相当する周波数帯域は、雑音の到来方向に依存するが、音声帯域のうち高々数パーセントにすぎないことを確認した。

3.2.3 雑音スペクトルの除去

本雑音除去アルゴリズムは、周波数領域における減算である SS[11] を利用して雑音成分の除去を行なう。

Boll が提案した SS は、音声／非音声区間の判定を必要とし、非音声区間で推定した定常雑音の振幅スペクトルを音声区間の受音信号の振幅スペクトルから引き去る手法である。しかし、雑音のある環境で収音された音声に対して非音声区間を検出することは決して容易ではなく [59]、実環境にはたくさんの非定常雑音が存在する。この問題を解決すべく、数多くの SS 改良法が提案されてきたが、非定常雑音の除去に関してはまだ満足できる手法はない。本雑音除去法は、短時間フレームごとに高精度で雑音スペクトルを推定できることから、非定常雑音への対応が期待できる。

雑音成分の除去は、マイクロホンアレーの中央マイクロホンで受音した信号 $c(t)$ の Fourier 変換 $C(\omega)$ から求めた振幅スペクトル $|C(\omega)|$ から、先に推定した雑音の振幅スペクトル $|\hat{N}(\omega)|$ を引き去ることにより実現できる。このとき、音声の振幅スペクトルの推定値 $|\hat{S}(\omega)|$ は、Boll の SS を改良した非線形 SS:

$$|\hat{S}(\omega)| = \begin{cases} |C(\omega)| - \alpha\cdot|\hat{N}(\omega)|, & |C(\omega)|\geq\alpha\cdot|\hat{N}(\omega)| \\ \beta\cdot|C(\omega)|, & \text{otherwise} \end{cases} \tag{3.28}$$

により推定する。式 (3.28) における α、β は定数であり、それぞれサブトラクション係数、フロアリング係数と呼ばれる。雑音スペクトルの推定誤差により、受音信号の振幅スペクトル $|C(\omega)|$ よりも雑音の振幅スペクトル $|\hat{N}(\omega)|$ の方が大きくなる可能性も考えられる。その際、音声の振幅スペクトルの推定値 $|\hat{S}(\omega)|$ が非負となるように非線形 SS を用いる。

3.3 まとめ

本章では、3ch 小規模マイクロホンアレーを利用し、3 つの受音点から 2 つの受音点を選択して構成されるマイクロホン対による雑音除去アルゴリズムを構築した。

本雑音除去アルゴリズムは、目的信号と雑音の到来方向を推定し、信号到来方向を利用して目的信号を完全に抑圧するような減算型ビームフォーマを解析的に設計し、スペクトル領域で受音信号から雑音成分を引き去る。信号の到来方向推定にも減算型ビームフォーマを利用することにより、雑音環境下での信号到来方向推定が可能になり、信号到来方向を利用することにより解析的ビームフォーミングが実現できる。また、雑音スペクトルの推定に関しては、各周波数において最適マイクロホン対選択機構の導入により、大幅な精度向上が期待できる。従って、本雑音除去法は、従来の雑音除去法では扱うことが困難であった時間・周波数領域において局在する非定常雑音を除去することが可能である。

第 4 章

雑音除去アルゴリズムの基礎的検証

4.1　はじめに

本章では、計算機シミュレーションにより、第 3 章で構築した雑音除去アルゴリズムの基礎的能力を検証する。

本雑音除去アルゴリズムの特徴は、時々刻々変化する非定常雑音の除去を実現するために

- 雑音環境下での音声および雑音の到来方向推定機構
- 各周波数における最適マイクロホン対の選択機構

を備えていることである。従って、本章では、雑音環境下での音声および雑音の到来方向推定精度、雑音スペクトルの推定精度 (最適マイクロホン対選択機構の有効性) について検証する。そして、典型的な非定常雑音として、目的信号と時間・周波数領域で局在する突発性雑音の除去精度についても検証する。

雑音環境下における信号の到来方向推定は非常に困難な問題であり、これまではこの課題が解決できなかった。これが、現在、適応フィルタによるビームフォーミングが減算型アレーの主流となっている要因の一つである。本研究では、方向推定にも積極的に減算型ビームフォーマを導入し、2 つの信号到来方向を推定することを試みた。本方向推定アルゴリズムが有効であれば、適応フィルタを用いる必要がなくなり、提案法では適応ビームフォーマが苦手としていた音響的特徴が急激に変化するような雑音の除去も可能になる。後者も本研究の独創的成果の一つであり、特に音声のような広帯域信号に対しては、本機構が高精度の雑音除去の実現に必要不可欠であると予想される。

4.2　雑音除去アルゴリズムのパラメータ設定

本雑音除去アルゴリズムの受音系は、隣接マイクロホン間隔 10 cm、両端のマイクロホンの間隔が 20 cm の 3ch 等間隔直線配列マイクロホンアレーとする。雑音除去アルゴリズムの各種パラメータは、以後特に断りのない限り、表 4.1 に示す値に設定する。

表 4.1: 雑音除去アルゴリズムのパラメータ設定値

パラメータ	設定値
フレーム長	5.3 msec
フレーム周期	2.7 msec
窓関数	hamming 窓
閾値 ε_1	0.5
閾値 ε_2	0.2
サブトラクション係数 α	1
フロアリング係数 β	0.001

フレーム長とフレーム周期は、一般的な音声情報処理と比較すると非常に短いが、これは時々刻々変化する非定常雑音にも対応できるように短く設定した。本雑音除去アルゴリズムでは、信号到来時間差の推定精度を向上させるため、到来時間差推定と雑音推定・除去処理とでは異なる分析条件を採用する。4 フレーム分の信号を統合した信号に対して方向推定を行ない、その各フレームにおいて雑音除去処理を行なう。具体的には、第 $(4n+1)$ フレーム（$n = 0, 1, 2, \cdots$）において、第 $(4n+4)$ フレームまでの 4 フレーム分の信号を一旦統合して到来時間差推定を行なう。第 $(4n+2)$、第 $(4n+3)$、第 $(4n+4)$ フレームに関しては、第 $(4n+1)$ フレームで推定した信号の到来時間差を利用し、雑音スペクトルの推定・除去のみを短時間フレームごとに行なう。閾値 ε_1、ε_2 は、広帯域のランダム雑音を用いた雑音スペクトル推定実験の結果に基づき、経験的に定めた値である。

ここで、目的信号が正中面方向、雑音が右 35 °から到来する場合の提案法の指向特性を図 4.1 に示す。提案法の指向特性と、同一マイクロホン配置の遅延和アレーの指向特性 (図 2.5) とを比較すると、高周波数帯域では両者とも空間的エイリアジングのため、目的信号の到来方向以外に受音感度の高いサイドローブが生じている。一般に、音声のパワー分布

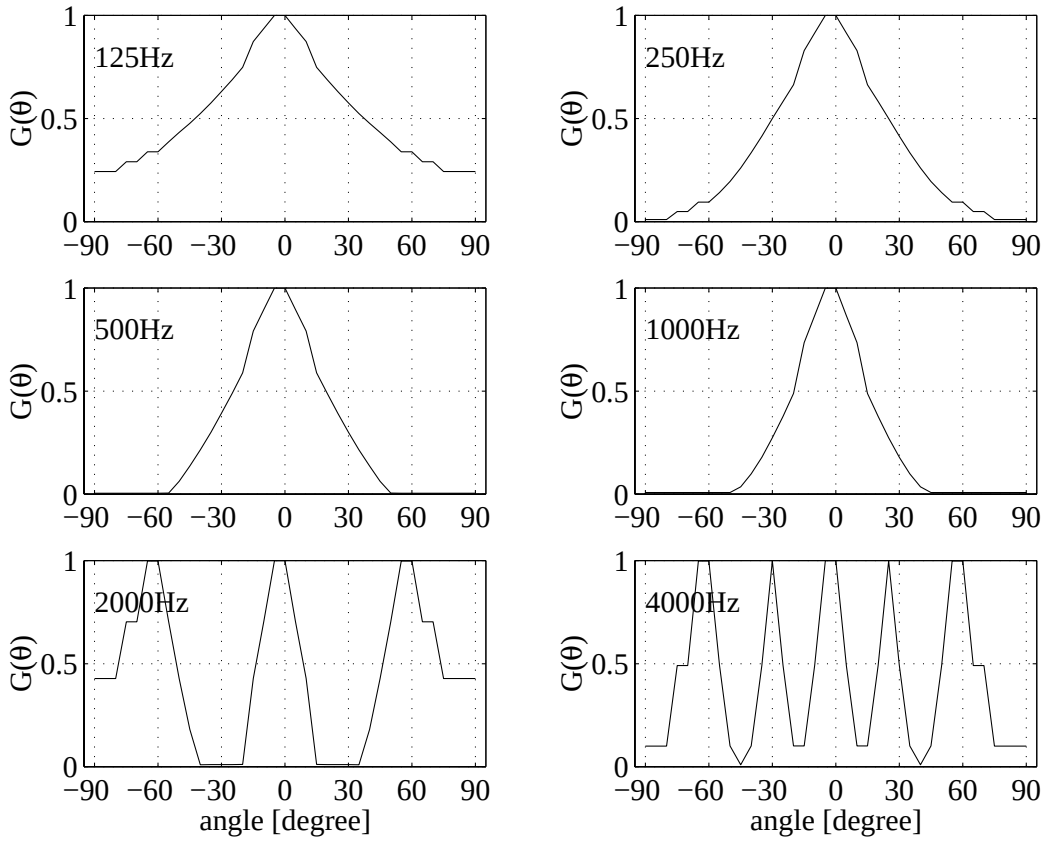

図 4.1: 提案法の指向特性 (左上から順に、125 Hz、250 Hz、500 Hz、1000 Hz、2000 Hz、4000 Hz の指向特性)

を考慮すると、高周波数帯域よりも低周波数帯域の雑音除去能力が重要となる [60]。提案法は、低周波数帯域において主ローブ幅が狭く、遅延和アレーよりも圧倒的に雑音除去能力が高いことがわかる。

4.3 方向推定アルゴリズムの精度検証

目的：

目的信号と雑音が各々一つずつ存在するような雑音環境において、目的信号及び雑音の到来方向推定精度を検証する。ここでは、目的信号源、雑音源がともに移動する場合について検討する。

シミュレーション条件：

方向推定の計算機シミュレーションには、目的信号として ATR 音声データベース [54] に収録されている男性話者 mht 発話によるクリーンな音声 /aoi/ と、雑音として中心周波数 1.5kHz、帯域幅 1 kHz のランダム帯域雑音を用いる。音声信号 /aoi/ は、3 つの単母音 /a/、/o/、/i/ を計算機上で結合したものである。音声信号および雑音信号の時間波形を図 4.2 に示す。信号到来方向の推定には、サンプリング周波数が 48 kHz で、16 bit で量子化された信号を用い、3 つの受音信号 $l(t)$、$l(t)$、$c(t)$ は、音声信号 $s(t)$、雑音信号 $n(t)$ にそれぞれ適切な時間シフトを与え、それらを計算機上で加算して作成する。

音声信号源 (話者) はアレーから見て右前方に存在して 35 °から正中面方向に向かって 25 °まで、雑音源は左前方に存在して 45 °から正中面方向に向かって 20 °まで移動するものとする。図 4.3 に、横軸を時間の経過を表すフレーム番号、縦軸を信号到来時間差を表す隣接マイクロホンへの到来時間差として、各フレームにおける音声および雑音の到来時間差の設定値を示す。方向推定を行なう際のフレーム長は 21.3 msec (1024 pt)、フレーム周期は 10.6msec (512 pt) とした。

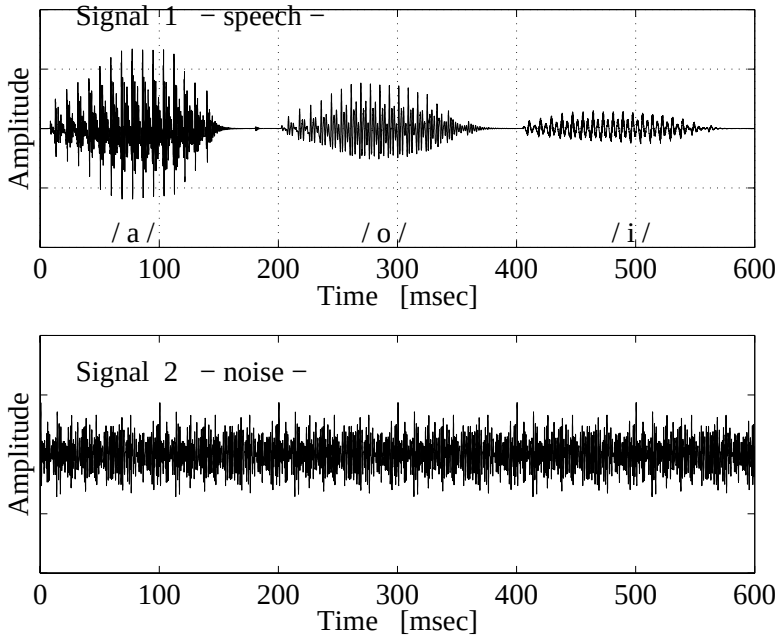

図 4.2: 信号到来方向の推定実験に用いる信号 (上図： 音声信号、下図： 雑音信号)

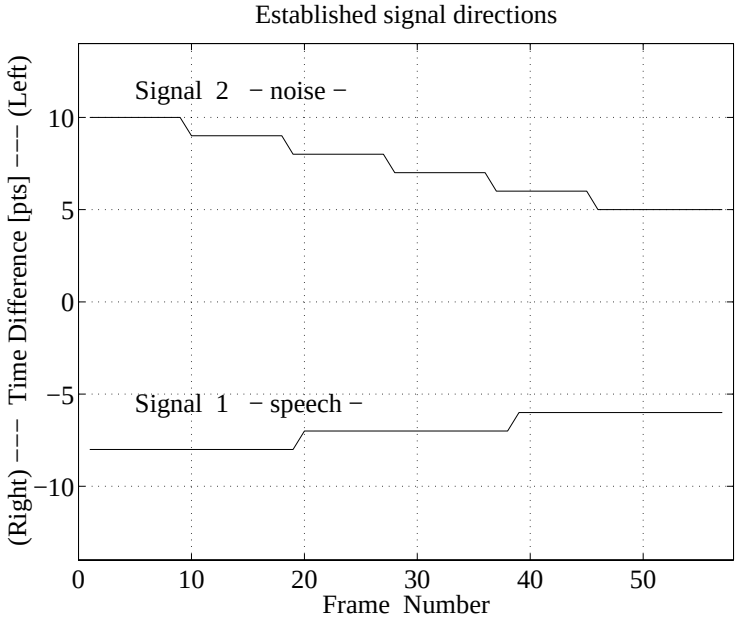

図 4.3: 音声および雑音の到来時間差の設定値

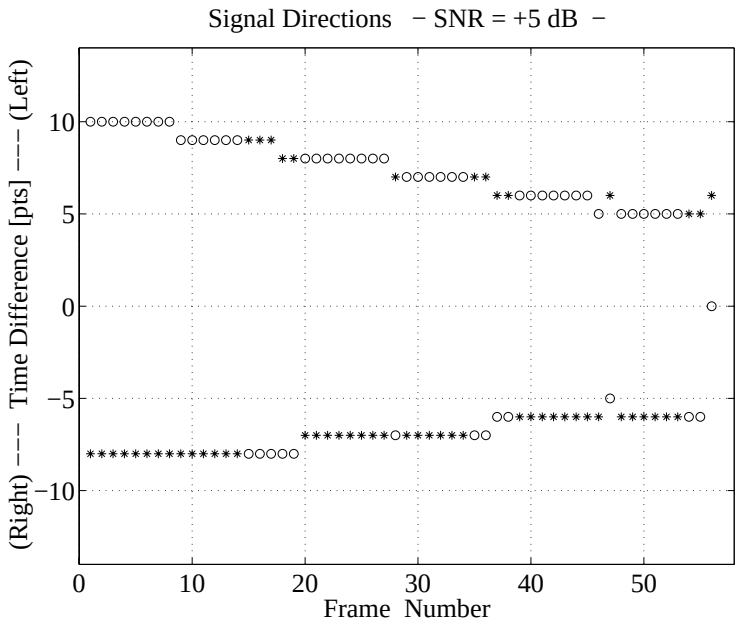

図 4.4: 音声および雑音の到来時間差推定結果 (SNR 5 dB の場合)

シミュレーション結果：

図 4.4 に、音声の SNR が 5 dB である場合の音声および雑音の到来時間差推定結果を示す。本方向推定アルゴリズムでは、各フレームごとに 2 つの信号の到来時間差を推定するため、最初に推定された到来時間差 (d_1) を＊印、2 番目に推定された到来時間差 (d_2) を○印で表す。これらに対し、**3.2.1** の **STEP 4** で述べた手法により、それぞれの方向が、音声あるいは雑音のどちらの到来方向であるかを判定する。

雑音に対する音声の SNR を +20 dB から 5 dB 刻みで −20 dB まで変化させ、音声および雑音の到来時間差推定を行なった結果を図 4.5 に示す。図 4.5 の横軸は音声の SNR であり、縦軸は到来時間差推定結果として、正確に推定できた場合、到来時間差 ±1 pt の誤差で推定できた場合、到来時間差の推定誤差が ±2 pt 以上生じた場合の割合を表す。参考のため、到来時間差推定に用いた信号のうち、SNR が ±20 dB、±10 dB、0 dB の場合の時間波形を図 4.6 に示す。

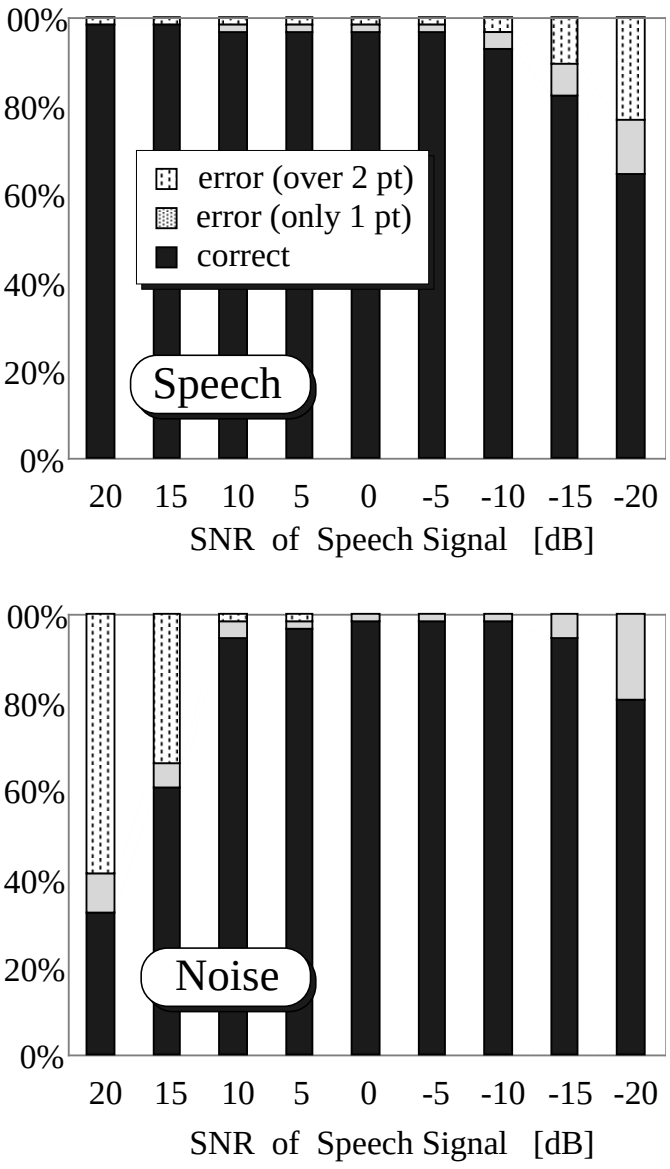

図 4.5: 音声および雑音の到来時間差推定結果 (SNR: –20 dB〜+20 dB)

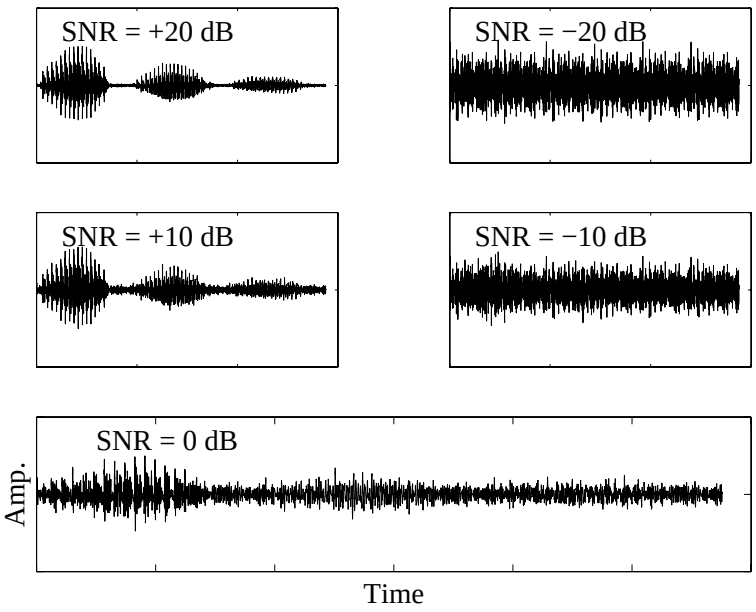

図 4.6: 音声および雑音の到来時間差推定に用いた信号 (SNR: ±20 dB、±10 dB、0 dB)

考察：

方向推定結果に関しては、図 4.4 の第 56 フレームにおいて大きな推定誤差が生じているが、これが雑音除去結果に及ぼす影響は非常に小さい。この理由は、図 4.2 に示した音声波形を見ると、誤差が生じているフレームでは音声信号自体のパワーが非常に小さいためである。図 4.5 より、音声の SNR が ±10 dB 以内であれば、本手法は音声と雑音の両到来時間差を精度よく推定できている。音声の到来時間差の推定に着目すれば、SNR が -5 dB 以上であれば 90% 以上の高精度で到来方向を正確に推定可能である。

以上より、2 つの信号が局在する環境において、本方向推定アルゴリズムが両信号の到来時間差を非常に正確に推定できることを確認できた。

4.4 最適マイクロホン対選択機構の有効性検証

目的：

本雑音除去アルゴリズムは、雑音スペクトルの推定過程において、主対ビームフォーマで推定誤差が大きくなる周波数帯域においては副対ビームフォーマを利用する。雑音スペクトル推定時に誤差が生じるのは、理論上は特定周波数のみであるが、実際にはその近傍の周波数帯域においても推定誤差が生じることが予想される。

本節では、計算機シミュレーションにより、スペクトルの推定誤差が生じる周波数帯域幅の調査、副対を利用することの有効性 (重要性) について検討する。

シミュレーション条件：

計算機シミュレーションでは、目的信号が正中面方向、雑音が正中面に対して 35 °方向から到来する状況を想定する。ここでは雑音スペクトル推定精度を調査することが目的であるため、音源の移動については考えない。

目的信号は、ATR 音声データベース [54] に収録されている男性話者 mht 発話によるクリーンな音声 /a/ の定常部分をループさせて作成した信号長 1 sec の定常的な信号とする。雑音は、始端周波数 1 kHz、終端周波数 6 kHz の線形対数 sweep 音で、信号長は 1 sec とする。雑音の変化速度は、図 2.1 あるいは図 2.8 の雑音の変化が激しい場合と等しい。

シミュレーション結果：

図 4.7 上段に、クリーンな音声に線形対数 sweep 音を付加した歪み音声のスペクトログラムを示す。歪み音声の SNR は -10 dB である。図 4.7 上段に示した低 SNR の音声を入力信号とし、雑音除去を行なった結果を図 4.7 の中段、下段に示す。図 4.7 中段図は雑音スペクトルの推定を行なう際に主対ビームフォーマのみを用いた場合、図 4.7 下段図は主対ビームフォーマと副対ビームフォーマとを併用した場合の雑音除去結果である。なお、主対ビームフォーマにおいて雑音スペクトルを計算できない周波数は、式 (3.23) より $3.4 \cdot n$ kHz (n: 整数) となる。

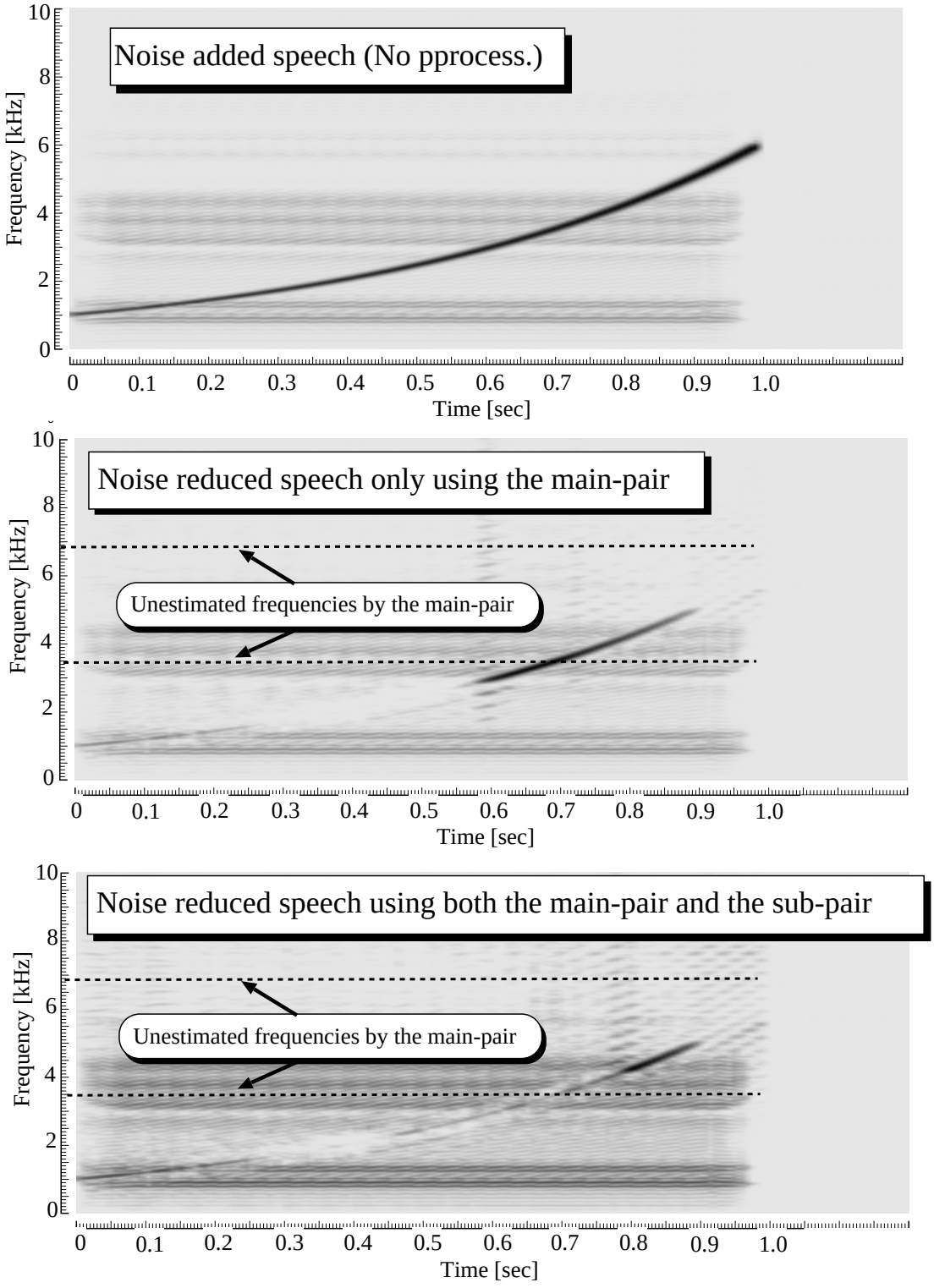

図 4.7: 線形対数 sweep 音を付加した音声と雑音除去音声のスペクトログラム (上段： 雑音付加音声、中段： 主対ビームフォーマのみによる雑音除去音声、下段： 主対ビームフォーマと副対ビームフォーマとを併用して雑音除去を行なった音声)

考察：

主対ビームフォーマのみで雑音推定を行なった場合、雑音除去後信号のスペクトルグラム (図 4.7 中段図) において、3.4 kHz とその近傍の周波数帯域において雑音を除去できていない。雑音を除去できない周波数帯域は、約 3kHz から 5kHz までと非常に広い。一方、副対ビームフォーマを併用して雑音スペクトルの推定・除去を行なうことにより、この問題が解決できることが図 4.7 下段図よりわかる。以上より、雑音スペクトル推定過程において、周波数ごとの最適マイクロホン対選択機構が非常に重要であることを確認できる。

提案法の雑音除去精度 (図 4.7 下段図) は、変化速度が等しい線形対数 sweep 音に対する従来法の雑音除去結果と比較すると、1ch 雑音適応型 SS (図 2.1 右端図)、Griffiths-Jim 適応型アレー (図 2.8 右端図) よりも圧倒的によいことがわかる。

4.5 突発性雑音の除去実験

目的：

Boll が提案した SS は非音声区間で雑音を推定し、非定常雑音への対応を目指した SS 改良法でも多かれ少なかれ非音声区間で推定した定常雑音を利用している。つまり、これまでに提案された SS 改良法では、非音声区間には存在せずに音声区間中で発生するような雑音、例えば会話中のドア開閉音などを除去することはできない。一方、提案法は、時々刻々雑音スペクトルを推定するため、非定常雑音を除去することも可能である。

ここでは、音声区間中で発生し、発生後直ちに消滅するような雑音で、しかも雑音源が移動する状況における提案法の雑音除去能力を検証する。

シミュレーション条件：

目的信号は、ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/ とし、正中面方向から到来すると設定した。一方、雑音は、継続時間が 50 msec で帯域幅が 200 Hz、中心周波数が 1.5 kHz と 2.5 kHz の 2 種類の狭帯域雑音が、音声区間中で突発的に発生するよう作成した。これらの雑音は、前者が右 35 °、後者が右 45 °から到

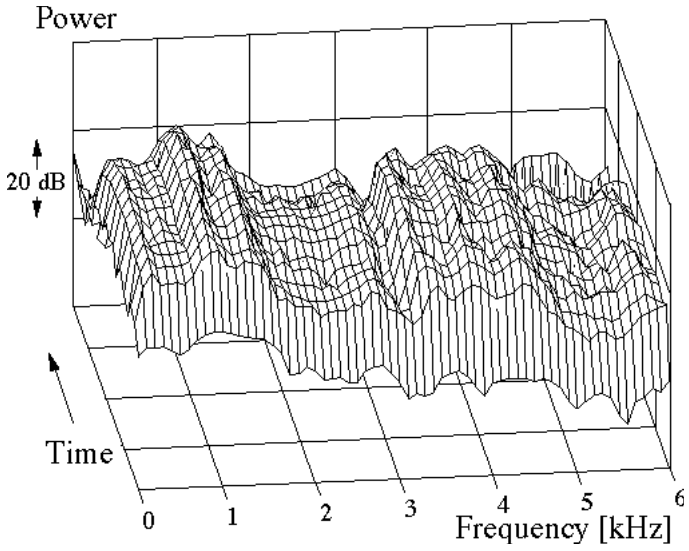

図 4.8: クリーンな音声の対数スペクトル包絡

来するように設定した。音声と雑音との時間的関係については、音声が約 10 msec から約 150 msec まで存在し、中心周波数 1.5 kHz の雑音が 30 msec から 80 msec まで、中心周波数 2.5 kHz の雑音が 90 msec から 140 msec まで存在する。クリーンな音声 /a/ の対数スペクトル包絡を図 4.8 に示す。同様に、SNR が −10 dB となるように突発性雑音を付加した音声を図 4.9 に示す。

シミュレーション結果：

雑音除去を行なった信号の対数スペクトル包絡を図 4.10 に、音声および雑音の到来時間差の推定結果を図 4.11 に示す。図 4.11 の横軸は時間、縦軸は隣接マイクロホンへの到来時間差を表す。図中の実線は音声および雑音の到来時間差の設定値、＊印、○印はそれぞれ最初に推定された信号の到来時間差、2 番目に推定された信号の到来時間差である。

考察：

雑音除去音声の対数スペクトル包絡 (図 4.10) を見ると、提案法は、従来の SS 改良法では除去が困難であった突発性雑音を高精度で除去できている。図 4.11 の音声および雑音

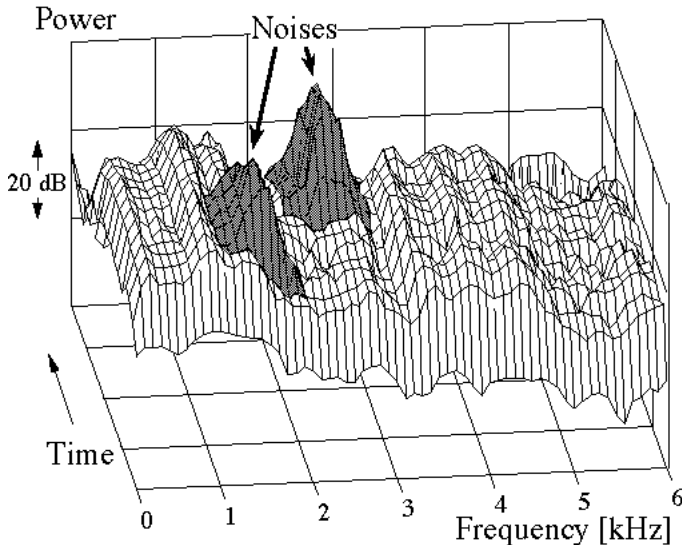

図 4.9: 雑音を付加した音声の対数スペクトル包絡

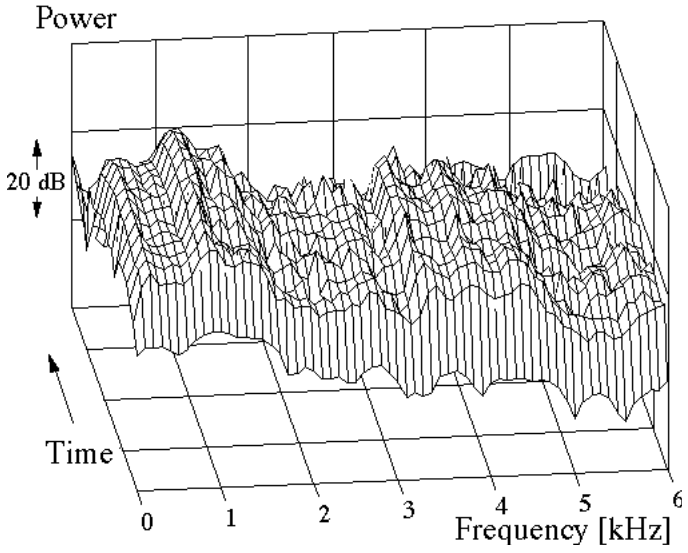

図 4.10: 雑音除去音声の対数スペクトル包絡

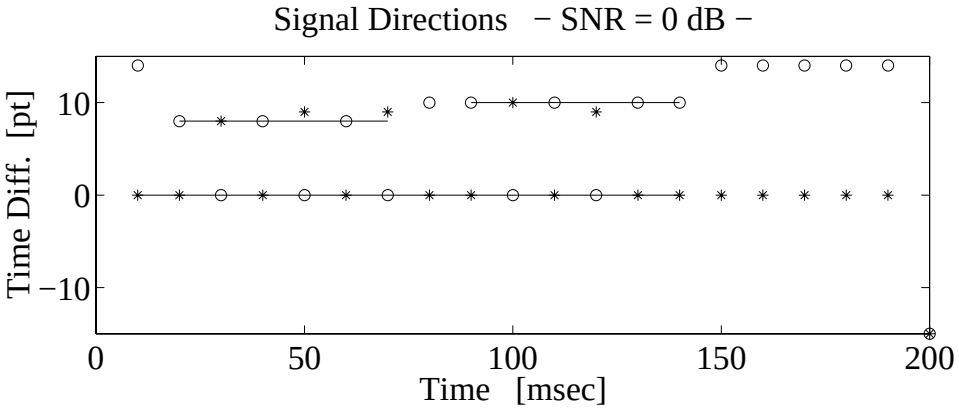

図 4.11: 音声および雑音の到来時間差の設定値 (実線) とその推定結果（＊印： 最初に推定された信号の到来時間差、○印： 2 番目に推定された信号の到来時間差）

の到来時間差の推定結果を見ると、雑音が存在する区間において若干の誤差が生じている。しかし、今回のシミュレーション条件においては、到来時間差 1 pt 程度の誤差であれば、それが雑音除去結果へ及ぼす影響は非常に小さいこともわかる。

雑音除去音声のスペクトル包絡には、雑音が存在していた帯域を中心に若干の歪みが生じている。これは、musical noise と呼ばれる歪みであり、SS の本質的な問題点である。例えば Zelinski は、この musical noise を低減させるためにウィナーフィルタを用いたポストフィルタリング手法を提案し、その有効性を報告している [58]。提案法においても、musical noise の除去・低減については今後の検討課題である。

ここで、同一のデータに対して、提案法で雑音除去を行なった結果 (図 4.10) と、Griffiths-Jim 適応型アレーで雑音除去を行なった結果 (図 2.7 下図) との比較を行なう。但し、提案法の場合には 2 つの突発雑音の到来方向は異なる (雑音源が移動する) が、Griffiths-Jim 適応型アレーの場合には 2 つの突発雑音は同一方向から到来すると仮定している。提案法は、雑音源が移動しているにも関わらず、Griffiths-Jim 適応型アレーよりも雑音除去能力が高いことが一目瞭然である。特に、雑音の立ち上がり部において、その差は顕著に現われている。我々の聴覚は、音の立ち上がり部に非常に敏感であるため、聴感上の印象を考慮する場合には、提案法は非常に優位であると考えられる。

表 4.2: 非定常雑音に対する提案法および従来法の適応表

雑音除去アルゴリズム (素子数)	SS (1)	適応 SS (1)	2ch SS (2)	遅延和アレー (多数)	適応アレー (3)	提案法 (3)
急激に変化する非定常雑音	×	×	×	○	×	◎
音声区間中で発生する雑音	×	×	×	○	△	◎
音源が移動する非定常雑音	×	△	×	○	○	◎

4.6 まとめ

本章では、第 3 章で構築した雑音除去アルゴリズムの基礎的能力を計算機シミュレーションにより確認した。

音声と雑音が時間・周波数領域で局在するような環境における 2 つの信号到来方向の推定に関しては、提案法が音声および雑音の到来方向をともに高精度で推定できることを確認した。また、雑音スペクトルの推定・除去精度に関しては、周波数ごとの最適マイクロホン対選択機構の有効性、そして重要性を確認した。また、提案法は、従来の SS 改良法や適応型アレーでは除去することが困難であった音声区間中の突発雑音も良好に除去できることもわかった。本章で明らかになった提案法の各種非定常雑音に対する適応性は、表 4.2 に示すように、従来法を大きく凌ぐものである。

本雑音除去アルゴリズムの各目的に応じた有効性、実環境における有効性に関しては、次章以降において順次調査検討する。

第 5 章

音声認識のフロントエンドとしての有効性検証

5.1 はじめに

本章では、雑音除去の応用例として自動音声認識器 (Automatic Speech Recognizer; ASR) を想定し、提案法の ASR のフロントエンドとしての有効性を検証する。

検証実験は、ASR に雑音除去アルゴリズムを実装し、フィールドテストによる性能評価を行なうことも可能である。しかし、これは評価を行なう度に莫大なコストを要し、評価結果が認識アルゴリズム、実験に用いるデータへ大きく依存する危険もある。より効率的かつ定量的に評価を行なうためには、客観的な歪み評価尺度の使用が望ましい。本章では、まず簡単な音声認識実験を行ない、次に雑音除去アルゴリズムの ASR のフロントエンドとしての有効性を評価するために相応しい客観的歪み評価尺度を定義する。客観評価尺度による評価結果と、先に行なった認識実験の結果とを比較し、客観評価尺度の妥当性を検証する。

そして、客観的歪み評価尺度を用い、提案法の ASR のフロントエンドとしての有効性を調査する。本章では提案法の評価として、非定常雑音に対する従来法との性能比較、定常雑音に対する遅延和アレーと提案法との性能比較を行ない、提案法の優位性について多角的に検討する。

5.2 音声認識実験による有効性検証

目的：

雑音を付加あるいは除去することによる音声認識率の変化を調査する。雑音除去アルゴリズムとして提案法に加え、遅延和アレーも用い、提案法の優位性についても検討する。

実験条件・実験手法：

音声認識における雑音の影響を調査するため、認識器として HTK[61] を用い、メルケプストラム (MFCC) 12 次 を特徴量とした音韻 HMM を構築し、話者の個人性の問題が混入しないように特定話者音韻認識実験を行なう。

HMM の学習に用いたデータは、ATR 音声データベース [54] に収録されている男性話者 mht 発話の重要語 5240 単語から抜粋した 1048 単語のクリーンな音声とする。テストデータは、同データベースに収録されている同話者発話の音韻バランス 216 単語に、計算機上で 125 Hz–6 kHz の定常的なランダム帯域雑音を付加した音声、あるいはそれに対して雑音除去を行なった音声とする。

雑音除去アルゴリズムとしては、提案法と、目的信号である音声の到来方向を既知と最適化した 3 ch 遅延和アレーとを用いる。

実験結果：

各 SNR の雑音付加音声、それに対して遅延和アレーあるいは提案法により雑音除去を行なった音声のそれぞれに対する音韻認識率を図 5.1、誤認識率を図 5.2 に示す。なお、クリーンな音声に対する認識率は 84.6 % であった。

考察：

提案法は、音声の歪みの程度 (SNR 値) に関わらず、雑音の影響により低下した認識率を大幅に向上させることが可能であることがわかる。同一マイクロホン配置の遅延和アレーとの比較においては、提案法の方が圧倒的に雑音除去能力が高く、その差は非常に大きい。誤認識率を示した図 5.2 より、雑音除去前と比較し、遅延和アレーでは誤認識率を高々 5% 程度しか低減できていないが、提案法では誤認識率を最大で約 20% も低減できている。これは、目的音源と雑音源とが移動せず、遅延和アレーは信号到来方向を既知と最適化した場合の実験結果である。遅延和アレーで方向を自動推定する場合、特に音源が移動する状況においては、提案法と遅延和アレーとの雑音除去能力の差はより大きくなると考えられる。

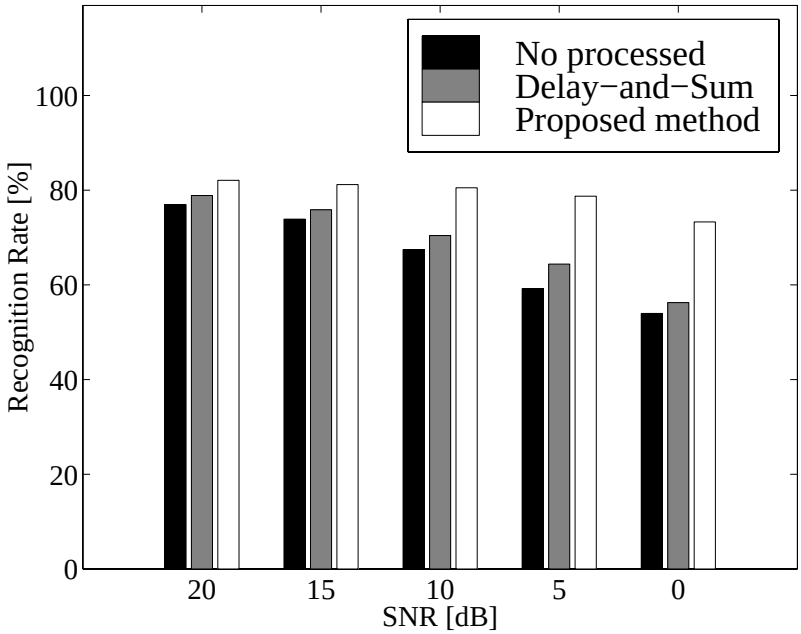

図 5.1: 音韻認識実験結果 (雑音除去付加音声、遅延和アレーによる雑音除去音声、提案法による雑音除去音声)

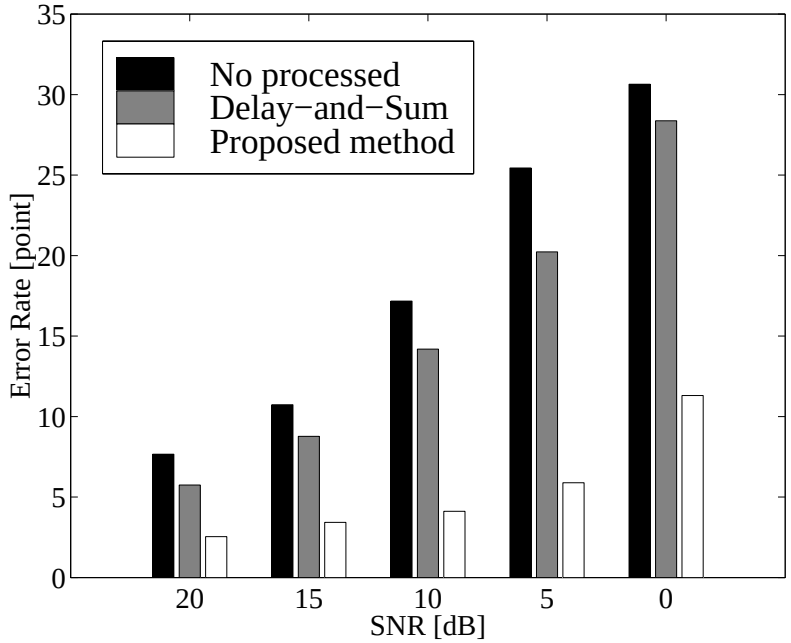

図 5.2: 雑音の影響による誤認識率 (雑音除去付加音声、遅延和アレーによる雑音除去音声、提案法による雑音除去音声)

5.3 ASR 雑音除去フロントエンドのための客観的評価尺度

5.3.1 客観的歪み音声評価尺度概説

雑音による音声の歪みの程度を客観的に評価するための尺度として、式 (5.1) で定義される信号対雑音比 (Signal to Noise Ratio; SNR) や、式 (5.2) で定義されるスペクトル歪 (Spectral Distortion; SD) がよく用いられている。

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{i=1}^{M} x^2(i)}{\sum_{i=1}^{M}\{x(i)-y(i)\}^2}\ [\ \mathrm{dB}\ ] \tag{5.1}$$

$$\mathrm{SD} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\{X(j)-Y(j)\}^2}\ [\ \mathrm{dB}\ ] \tag{5.2}$$

ここで、$x(t)$、$y(t)$ はそれぞれ評価の基準となるクリーンな音声、評価対象の音声とし、それらの対数振幅スペクトルを $X(\omega)$、$Y(\omega)$ とする。また、M は評価に用いる信号の信号長、N は評価の対象周波数の上限を表す。

5.3.2 客観的評価尺度 LPC-SED の定義

ASR 雑音除去フロントエンドの性能評価を行なうためには、 ASR の照合部で利用されているスペクトル距離尺度を考慮すべきである。現在の ASR の多くが LPC 分析に基づくスペクトル距離尺度を利用していることから、本節では LPC ケプストラムの低次係数から求めた対数スペクトル包絡に生じた歪みを評価可能な尺度について考える。これを LPC 対数スペクトル包絡歪 (LPC log Spectral Envelope Distortion; LPC-SED) と呼ぶ。突発音などの非定常雑音により生じる歪みを正確に評価するためには、短時間ごとに歪み量を求める必要がある。そこで、評価の基準となる信号 (クリーンな音声) と評価対象信号 (雑音除去前、雑音除去後の信号) のそれぞれに対し、短時間フレームごとに LPC 対数スペクトル包絡を表 5.1 に示す条件で求め、それらを $S_x(\omega)$、$S_y(\omega)$ とする。このとき、LPC-SED は、各フレームごとに

$$\text{LPC-SED} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{S_x(i)-S_y(i)\}^2}\ [\ \mathrm{dB}\ ] \tag{5.3}$$

表 5.1: LPC-SED 値算出のための LPC 分析条件

分析パラメータ	設定値
サンプリング周波数	12 kHz
フレーム長	21.3 msec
フレーム周期	5.3 msec
窓関数	hamming 窓
LPC 次数	16
プリエンファシス	0.98

により計算される歪み量の全音声区間での平均値と定める。ここで、評価の上限周波数 N は、対象信号が音声であることから、音声の明瞭性のみならず個人性、自然性まで考慮して 6 kHz とする [59]。

LPC-SED 値を計算するために、提案法は、雑音除去処理後に音声の振幅スペクトルと受音信号の位相スペクトルを用い、逆 Fourier 変換により一旦時間領域へ変換した信号を作成する。これは、ASR で通常用いられている分析条件を考慮し、雑音除去処理の分析条件 (表 4.1) とは異なる分析条件 (表 5.1) で評価を行うためである。また、LPC-SED にユークリッド距離を用いた理由は、ユークリッド距離が距離尺度の対称性、正値性を満たすことができ、音声認識の距離尺度として利用価値が高いためである [59]。

なお、LPC-SED は、Parseval の定理より理論上は LPC ケプストラム距離 (Cepstrum Distance; CD) と等価である [62]。CD は符号化電話音声の品質評価に用いられているが [59]、LPC-SED は雑音除去アルゴリズムの ASR のフロントエンドとしての有効性を調査することを目的とする。

客観的評価尺度 SNR、SD、LPC-SED の関係について考える。SNR は、時間波形に対して評価を行なうために位相も考慮した評価尺度である。SD は、対数振幅スペクトルに対して評価を行なうため、位相に関しては全く考慮していない。本節で定義した LPC-SED は、LPC 対数スペクトル包絡を用いて歪み量を計算するため、SD では考慮する振幅スペ

クトルの微細変動成分を無視したものである。また、LPC 対数スペクトル包絡は、スペクトルの山形部分を重視して推定されるため [59]、SD の計算に用いる FFT により計算されたスペクトルに含まれるスペクトル包絡情報とは若干異なるものである。

5.3.3 客観的評価尺度 LPC-SED の妥当性検証

本節では、音声認識実験により、前節で定義した LPC-SED による評価値が認識率をある程度反映しているのかを調査する。雑音による歪み量を表す LPC-SED 値が大きくなると認識率が低下し、LPC-SED 値が小さくなると認識率が向上するという傾向を確認できれば、客観的評価尺度 LPC-SED は音声認識率をある程度は反映できると言える。音声認識率は、実験条件や実験データに大きく依存するため、客観評価値により認識率を厳密に予測することは不可能である。仮に、ある条件下で認識率を正確に予測できるような評価尺度を作成しても、実験条件や実験データが異なると再度評価尺度を作成する必要がある。従って、LPC-SED 値が、音声認識率を厳密に予測できるか否かではなく、ある程度反映しているか否かについて検討する。つまり、評価尺度 LPC-SED は、評価値の大小に対しての順序は保障されている必要があるが、評価値の等間隔性まで保障する必要はない序数尺度 [63] の条件を満たしていれば十分である。

目的：

本実験の目的は、ASR 雑音除去フロントエンドを評価するための客観的歪み評価尺度 LPC-SED の序数尺度としての妥当性を検証することである。

実験手法：

第 5.2 節の音声認識実験に用いた雑音を付加あるいは除去した音声に対して LPC-SED 値を計算し、認識率と LPC-SED 値との関係について検討する。

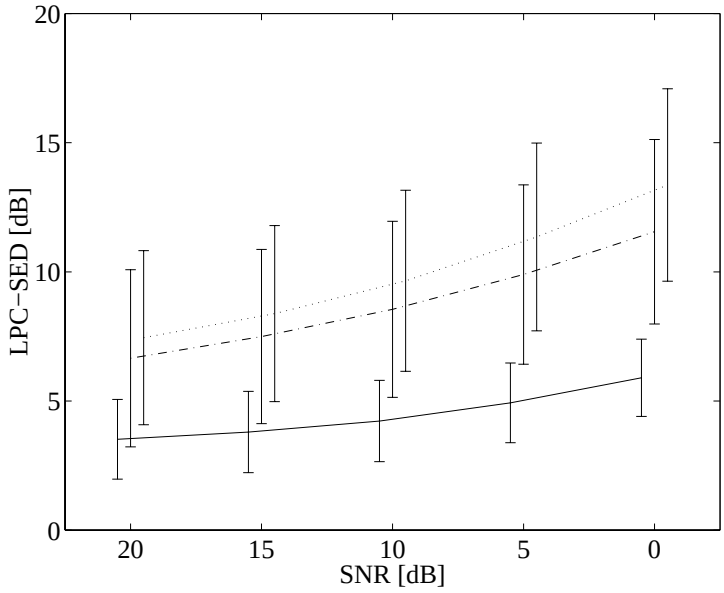

図 5.3: SNR と LPC-SED 値との関係 (点線は雑音付加音声、鎖線は遅延和アレーによる雑音除去音声、実線は提案法による雑音除去音声)

実験結果：

各 SNR の雑音付加音声、遅延和アレーによる雑音除去音声、提案法による雑音除去音声に対して計算した LPC-SED 値を図 5.3 に示す。図 5.3 において、点線は雑音付加音声、鎖線は遅延和アレーによる雑音除去音声、実線は提案法による雑音除去音声の各 216 単語に対して計算した LPC-SED 値の平均値、エラーバーはそれぞれの標準偏差を表す。

考察：

SNR 値と音韻認識率との関係 (図 5.1) と、SNR 値と LPC-SED 値との関係 (図 5.3) より、図 5.4 に示すような LPC-SED 値と音韻認識率との関係が得られる。図 5.4 における点線は、各データに対し、最小自乗誤差法 [64] によりフィッティングした 2 次曲線である。図 5.4 より、客観的歪み評価尺度 LPC-SED は、大局的に見ると序数尺度の条件を満たしており、雑音除去アルゴリズムの ASR のフロントエンドとしての有効性を評価するための尺度として妥当であると言える。

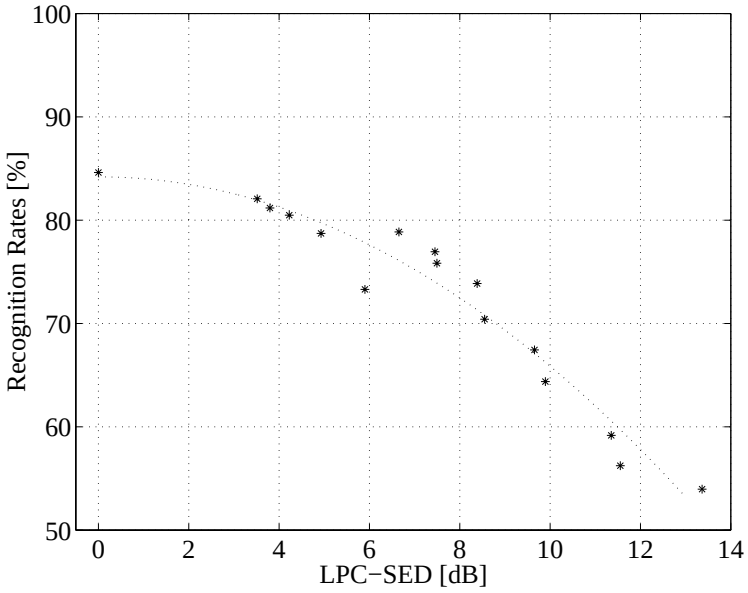

図 5.4: LPC-SED 値と音韻認識率との関係

5.4 客観的評価尺度による雑音除去アルゴリズムの評価

本節では、客観的評価尺度 LPC-SED による雑音除去アルゴリズムの評価を行なう。さらに、音声認識率と相性のよい客観的音声歪み評価尺度として提案されている Normalize Deviation to Signal Ratio (NDSR) [65][66] も使用し、多角的に本雑音除去アルゴリズムの評価を行なう。

5.4.1 客観的評価尺度 LPC-SED による有効性検証

目的：

雑音除去による LPC-SED 値の変化を調査し、本雑音除アルゴリズムの ASR のフロントエンドとしての有効性を検証する。また、LPC-SED を評価基準とした場合のサブトラクション係数 α の最適値について調査する。

シミュレーション条件：

まず、ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/ (図 4.8) に、継続時間が 50 msec で帯域幅が 200 Hz、中心周波数が 1.5 kHz と 2.5 kHz の 2 つの突発性雑音を付加する。音声は正中面方向から、雑音はそれぞれ右 35 °と右 45 °から到来するものと仮定し、雑音付加音声の SNR が -10 dB から 5 dB 刻みで 10 dB までの信号を作成する。第 2 章で解説した 1ch 雑音適応型 SS、2ch 雑音推定型 SS、3ch 遅延和アレー、3ch Griffiths-Jim 適応型アレーによる雑音除去も行ない、雑音除去性能の差異について検討する。

次に、目的信号がより複雑な母音と子音が混在する単語音声に対する提案法の雑音除去能力について検討する。実験に使用する信号は、ATR 音声データベース [54] に収録されている男性話者 mht 発話の音韻バランス単語 216 語とし、正中面方向から到来ように設定する。雑音は、125 Hz から 6 kHz までのランダム帯域雑音とし、右 35 °から到来するように設定する。音声の SNR が 10 dB となるようにクリーンな音声に雑音を付加し、雑音除去を行なう。

最後に、LPC-SED を評価基準とし、サブトラクション係数 α の最適化実験を行なう。実験には、ATR 音声データベース [54] に収録されている重要語データセット中の男性話者 mht 氏発声の連続母音 /ao/ のわたり部に、2kHz–3kHz のランダム帯域雑音を付加した音声を用いる。実験手法は、式 (3.28) のサブトラクション係数 α を 0 から 0.1 刻みで 2.0 まで変化させて雑音除去を行ない、各雑音除去音声の LPC-SED 値を求め、LPC-SED 値が最小となる場合の α を最適値とする。

シミュレーション結果：

まず、単母音に 2 つの突発性雑音を付加した音声 (SNR 0 dB) に対する雑音除去結果として、雑音除去による LPC-SED 値の変化を図 5.5 に示す。図 5.5 の横軸はすべて時間を表し、上段から順に、クリーンな音声、雑音付加音声の時間波形、評価フレームごとの雑音除去前音声 の LPC-SED 値、雑音除去後音声の LPC-SED 値、雑音除去による LPC-SED

値の改善量を表す。図 5.5 の 2 段目に雑音付加音声とともに示した斜線部は、音声と雑音が共存する区間を表し、この区間における LPC-SED 値の平均改善量は 4.85 dB であった。

雑音付加音声の SNR が異なる場合の雑音除去性能については、音声と雑音が共存する区間における LPC-SED 値の平均改善量を図 5.6 に示す。図 5.6 の横軸は雑音除去前の音声の SNR、縦軸は雑音除去による LPC-SED の平均改善量を示す。図中の＊印は音声および雑音の到来時間差を自動推定して雑音除去を行なった場合、実線はそれらを既知として雑音除去を行なった場合の LPC-SED 値の平均改善量を表す。

SNR 0 dB の雑音付加音声に対し、第 2 章で解説した 1ch 雑音適応型 SS、2ch 雑音推定型 SS、3ch 遅延和アレー、3ch Griffiths-Jim 適応型アレーにより雑音除去を行なった結果を図 5.7 に示す。

次に、単語音声を目的音とした場合の雑音除去性能については、図 5.8 に示すように、雑音除去前の単語音声の LPC-SED 値は全 216 単語の音声区間での平均で 11.15 dB (標準偏差は 1.54 dB)、雑音除去音声では 5.15 dB (標準偏差は 0.68 dB) であった。

LPC-SED を評価基準とした場合、すなわち ASR のフロントエンドとしてのサブトラクション係数 α の最適化実験の結果を図 5.9 に示す。図 5.9 は、サブトラクション係数 α と LPC-SED 値との関係を表しており、式 (3.28) からもわかるように $\alpha = 0$ は雑音除去を行なわないことを意味する。図 5.9 より、LPC-SED を評価基準としてサブトラクション係数 α の最適化を行なった場合、本実験における最適値は明確に 1.0 である。様々なデータに対し、LPC-SED を基準としたサブトラクション係数 α の最適化実験も行なった結果、α の最適値はいずれの場合も図 5.9 と同様に明確に 1.0 であった。

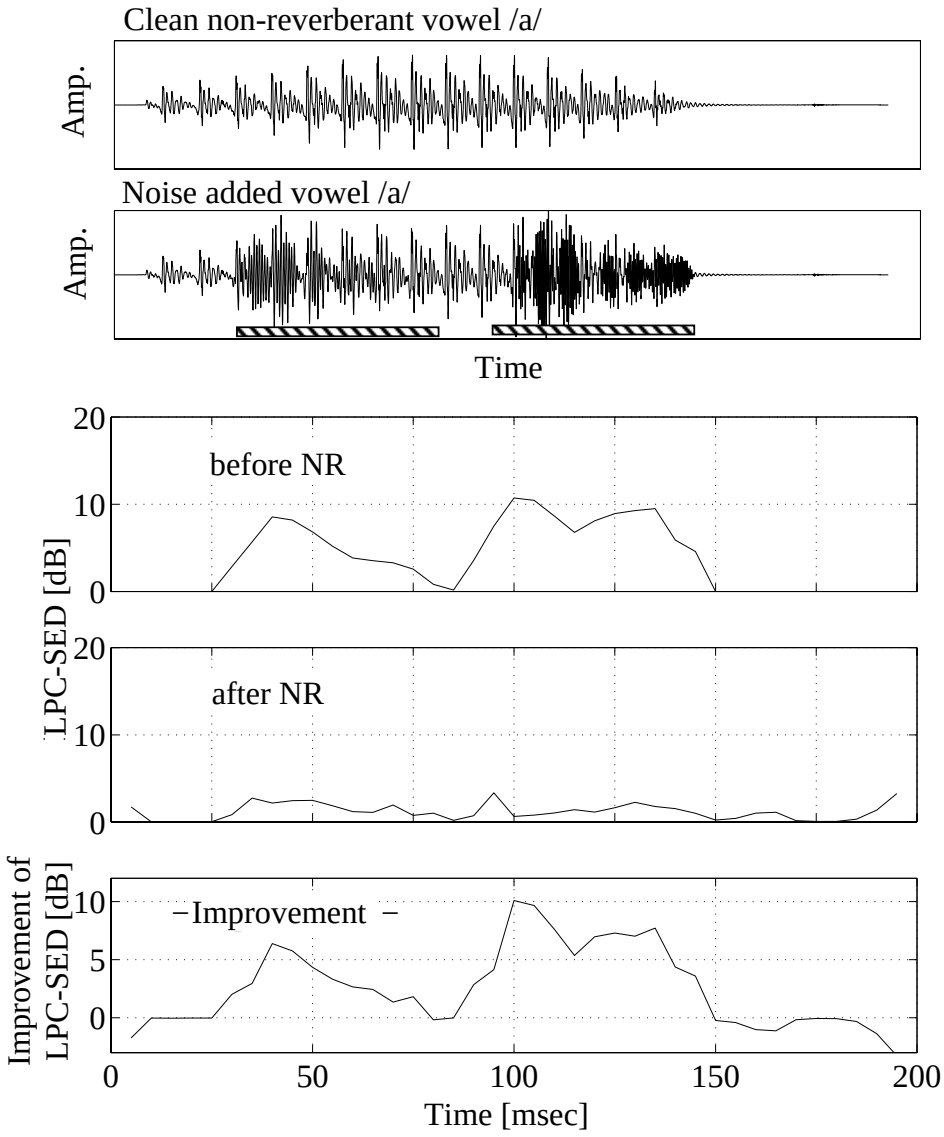

図 5.5: 雑音除去による LPC-SED 値の変化 (上段から順に、クリーンな音声、雑音付加音声 (斜線部に雑音付加)、雑音除去前の LPC-SED 値、雑音除去後の LPC-SED 値、雑音除去による LPC-SED 値の改善量)

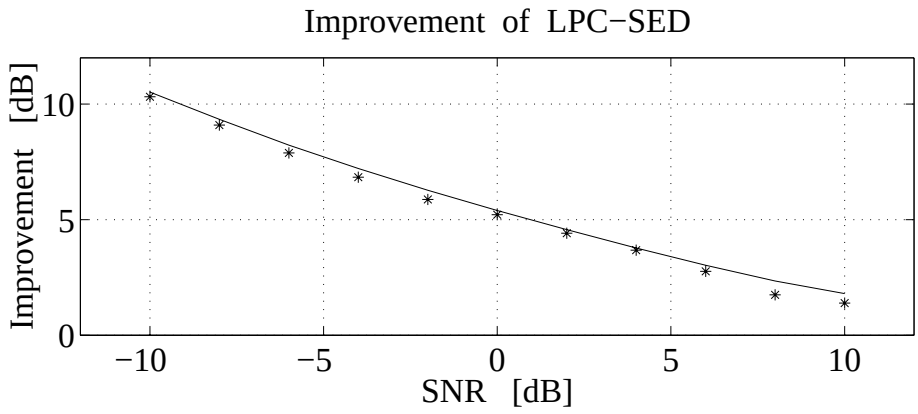

図 5.6: 各 SNR に対する LPC-SED 値の平均改善量 (∗： 音声および雑音の到来時間差を自動推定した場合、実線： 到来時間差を既知とした場合)

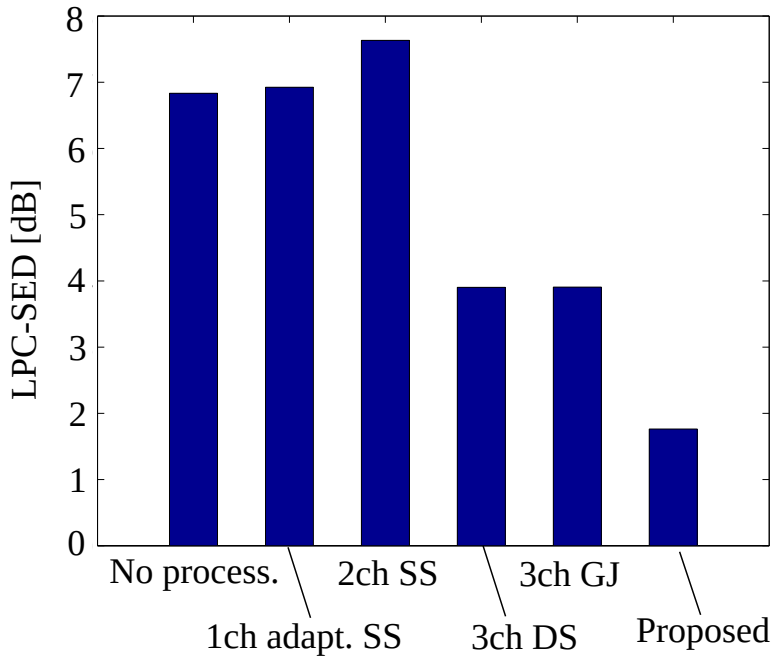

図 5.7: 提案法と従来法との雑音除去能力の比較 (左から順に、雑音付加音声の LPC-SED 値 (No process.)、1ch 雑音適応型 SS(1ch adapt. SS)、2ch 雑音推定型 SS (2ch SS)、3ch 遅延和アレー (3ch DS)、3ch Griffiths-Jim 適応型アレー (3ch GJ)、提案法 (Proposed) による雑音除去音声の LPC-SED 値)

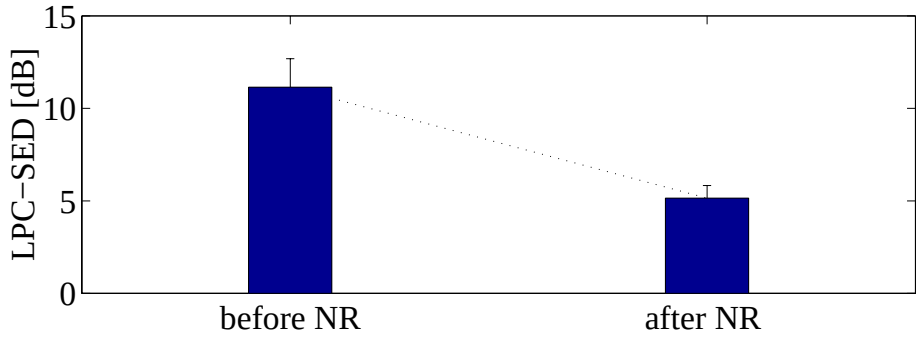

図 5.8: 単語音声に対する雑音除去による LPC-SED 値の変化 (216 単語に対して計算した LPC-SED 値の平均値と標準偏差)

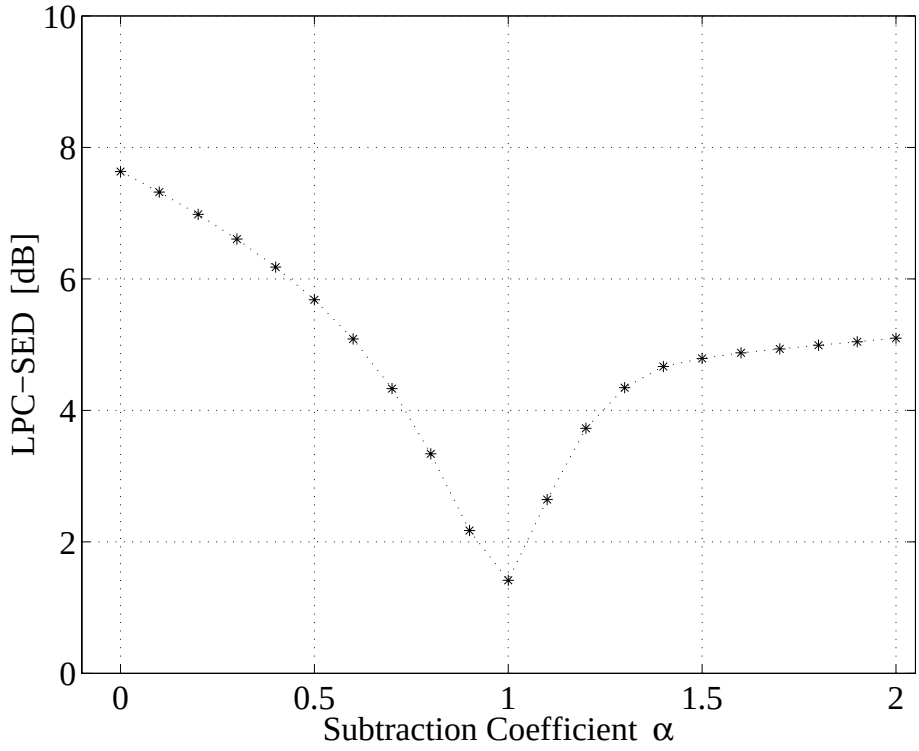

図 5.9: サブトラクション係数 α と LPC-SED 値との関係

考察：

LPC-SED による評価 (図 5.5) において、提案法が音声区間中で発生した突発性雑音を除去できることを確認できる。図 5.6 において、信号の到来方向を自動推定した場合 (＊印) と信号の到来方向を与えた場合 (実線) とを比較しても雑音除去精度には有意な差はない。なお、方向推定の誤差の大きさに関しては、推定誤差のほとんどが到来時間差 1pt であった。つまり、提案法は、方向推定で誤りが生じたとしても、それが小さければ雑音除去精度が大きく低下することはない。

様々な雑音除去法により突発性雑音の雑音除去を行なった結果、図 5.7 より、提案法が最も雑音除去精度が高いことがわかる。3ch の遅延和アレー、Griffiths-Jim 適応型アレーは、突発性雑音に対してははぼ同等の雑音除去能力を有する。これは、前者は突発性雑音を全般的に抑圧できるが抑圧量は少なく、後者は雑音の立ち上り部では雑音を除去できていないためである。

また、母音のみならず子音も含み、SNR が 10 dB と比較的歪みの少ない単語音声に対しても本雑音除去アルゴリズムは有効であることもわかった。

サブトラクション係数 α の最適化実験においては、最適値は 1.0 であることがわかった。これは、推定した雑音成分を受音信号からそのまま引き去ることを意味し、本雑音除去アルゴリズムの雑音スペクトルの推定精度の高さを裏付ける結果でもある。

5.4.2 客観評価尺度 NDSR による有効性検証

NDSR の定義：

NDSR は、マイクロホンアレーによる各種雑音除去法の評価 [65]、音声認識に適したマイクロホンアレーの設計 [66] を行なうことを目的に作成された客観的な音声歪み評価尺度である。目的信号のパワースペクトルを $S(\omega)$、雑音抑圧後の信号のパワースペクトルを $\hat{S}(\omega)$ とすると、NDSR は式 (5.4) で与えられる [65]。

$$\mathrm{NDSR} = -10\log\left(\sum_{\omega}^{N}\frac{|\hat{S}(\omega)-S(\omega)|}{S(\omega)}+1\right)\ [\ \mathrm{dB}\ ] \tag{5.4}$$

表 5.2: LPC-SED と NDSR による雑音除去前後の歪み音声評価結果

	LPC-SED [dB]	NDSR [dB]
before NR	11.15 (1.54)	-6.57 (0.63)
DS	8.22 (1.27)	-3.33 (0.53)
proposed	5.15 (0.68)	-2.30 (0.26)

井上らは、マイクロホン配置の異なる様々な遅延和アレーによる雑音除去結果に対し、NDSR が音声認識率を反映することのできる評価尺度であることを確認している [65][66]。

目的：

提案法と遅延和アレーの 2 種類の雑音除去アルゴリズムを用いて雑音除去を行ない、LPC-SED と NDSR の両評価尺度による評価値を求め、本雑音除去アルゴリズムの有効性を検証するとともに、LPC-SED と NDSR との関係を調査する。

実験条件・実験手法：

実験に使用する音声は、ATR 音声データベース [54] に収録されている男性話者 mht 発話の音韻バランス単語 216 語とし、正中面方向から到来するよう設定した。雑音は、125 Hz–6 kHz のランダム帯域雑音が右 35 °から到来するよう設定し、SNR が 10 dB となるように計算機上で音声に付加した。但し、遅延和アレーは、マイクロホン配置は提案法と同様とし、目的信号の到来方向は既知と最適化する。

実験結果：

雑音付加音声、遅延和アレーおよび提案法による雑音除去音声に対する LPC-SED 値、NDSR 値の平均値と標準偏差を表 5.2 に示す。LPC-SED、NDSR は、両者とも評価値が 0 dB に近いほど評価音声の歪みが小さいことを意味する。

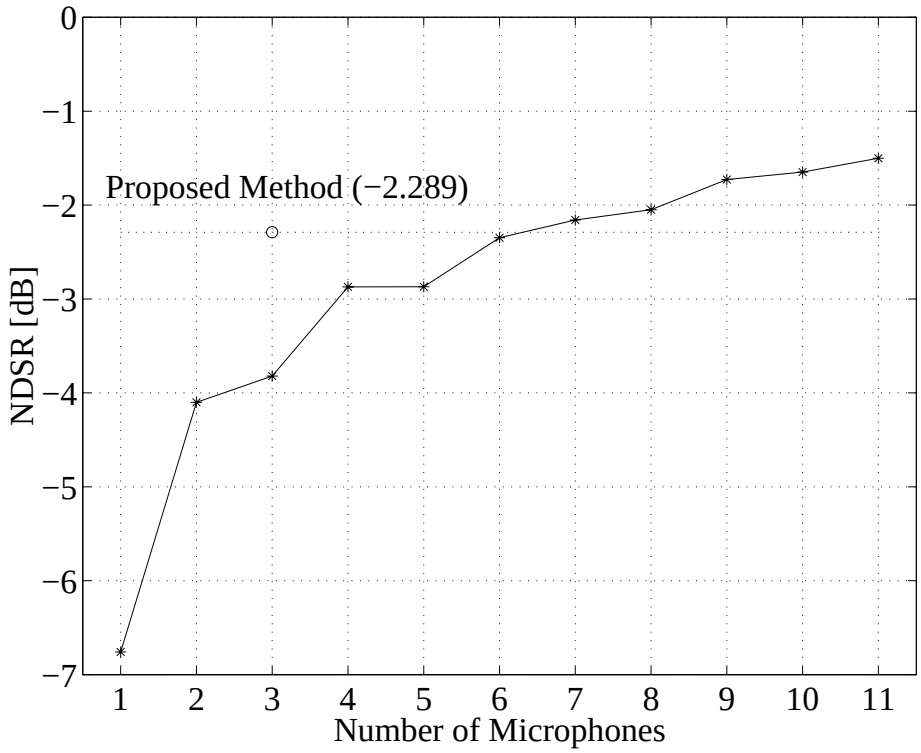

図 5.10: NDSR による提案法と遅延和アレーとの性能比較

考察

表 5.2 より、LPC-SED、NDSR の両評価尺度において、雑音除去前よりは遅延和アレーで雑音除去を行なった方が、また遅延和アレーで雑音除去を行なった場合よりも提案法で雑音除去を行なった方が、歪みが小さくなっている。

ここで、遅延和アレーと提案法との性能をより詳細に比較するため、遅延和アレーの素子数 (マイクロホン数) と NDSR 値との関係を調査する。マイクロホン間隔を 6 cm とし、遅延和アレーの素子数を変化させて雑音除去を行ない、雑音除去音声を NDSR により評価した結果を図 5.10 に示す。図 5.10 より、ASR のフロントエンドとして、遅延和アレーで提案法と同等の雑音除去性能を得るためには、少なくとも 6 本のマイクロホンが必要であることがわかる。

5.5 まとめ

本章では、ASR のフロントエンドとしての有効性を調査するため、まず音声認識実験を行ない、提案法の有効性ならびに遅延和アレーに対する優位性を確認した。

また、客観的歪み評価尺度 LPC-SED を定義し、序数尺度としての妥当性を検証した。音声認識率をある程度反映できる客観的評価尺度 LPC-SED、NDSR の両者において、提案法が ASR のフロントエンドとして有効であることがわかった。

提案法と従来法との性能比較においては、ASR のフロントエンドとしての提案法の雑音除去能力は、同一マイクロホン配置の遅延和アレー、Griffiths-Jim 適応型アレーよりも優れていることがわかった。特に、提案法と遅延和アレーとの比較においては、素子数 3 の提案法が、素子数 6 の遅延和アレーと同等の性能を持つこともわかった。

第 6 章

聴感上の歪み感低減能力の検証

6.1 はじめに

本章では、補聴器などで問題となる聴感上の歪み感の低減能力に着目し、雑音除去アルゴリズムの評価を行なう。主観評価実験による雑音除去アルゴリズムの評価を行なうとともに、第 5 章と同様に聴感上の歪み感と相性のよい客観的評価尺度を定義し、これを用いた客観評価も行なう。本評価尺度は、聴覚特性として継時マスキング特性も考慮しているため、時々刻々変化する非定常雑音にも対応できる。

また、聴感上の歪み感低減能力に関して提案法と従来法との性能比較を行ない、提案法の優位性についても検討する。ここでは、音声区間中に発生するような突発性雑音の除去能力、目的信号と雑音がともに同一話者発話の音声である高相関雑音の除去能力について検討する。

6.2 主観評価実験による提案法の有効性検証

目的：

本節では、非定常雑音により生じた聴感上の歪み感低減能力に関して、主観評価実験により提案法の有効性を検証する。

実験条件：

LPC-SED に対するサブトラクション係数 α の最適化実験と同様に、サブトラクション係数 α を 0 から 0.1 刻みで 2.0 まで変化させて雑音除去を行ない、雑音除去音声に対する主観評価値 MOS を求める。

実験には、音声区間中に突発的に発生し、目的信号との相関が高い非定常雑音を用いる。目的信号と雑音は、ともに ATR 音声データベース [54] に収録されている男性話者 mht 氏発話の文章データの一部とする。目的信号は「通訳電話国際会議事務局です」、雑音は「題目の締切を教えて下さい」の各文章とし、これらを SNR が 0 dB になるように加算したデータに対して雑音除去を行なう。これは、雑音除去問題の中では最も困難な課題の一つ

であるカクテルパーティ効果 [47] に相当する。

主観評価実験は、正常聴力を有する大学院生 8 名を被験者とし、防音室内で各雑音除去音声を 4 回ずつランダムにヘッドホン呈示する。被験者には、雑音による音声の歪み感を 5 段階の MOS により回答させた。目的信号に生じる歪みには、妨害音声の存在による目的信号の聞き取り難さ (サブトラクション係数 α が 0 に近い場合)、目的信号本来のスペクトル構造が崩壊することによる不快感 (α が 2.0 に近い場合) の 2 種類がある。被験者へは、実験を行なう前にこれらの歪みに関する説明を行ない、実験に使用する刺激音を用いて十分に練習を行なわせた。

実験結果：

実験結果として、主観評価実験により得られたサブトラクション係数 α の各値に対する被験者個別の平均 MOS を図 6.1 に示す。図 6.1 を見ると、被験者 S01、S02、S05、S06 の 4 名は、$1.1 \leq \alpha \leq 1.3$ で MOS が最大となっている。被験者 S03、S04、S07、S08 は、α が 1.7 あるいは 2.0 で MOS が最大となっているが、いずれも $1.1 \leq \alpha \leq 1.5$ に MOS の極大値が存在する。また、被験者全員の平均値を図 6.2 に示す。全被験者の平均 MOS に対するサブトラクション係数 α の最適値は 1.2 である。

考察：

提案法は、目的信号と雑音がともに音声である場合であっても良好に雑音を除去できることがわかった。また、聴感上の印象を最大限向上させるためには、サブトラクション係数 α は 1.2 程度に設定すべきであり、ASR のフロントエンドとして (LPC-SED を評価基準とした場合) の最適値 1.0 とは異なることもわかった。

主観評価実験の結果に関して、サブトラクション係数 α が 0.1、0.3 と 0 に近い場合、被験者ごとの主観評価値 MOS のバラツキが大きくなっている。これは、サブトラクション係数 α が 0 に近い場合は、雑音として用いた妨害音声がはっきりと聴こえしまい、被験者がこれを完全に分離して聴いたためであろうと推測できる。

更に、この結果が実験データと被験者に依存した傾向ではないことを確認するため、ATR

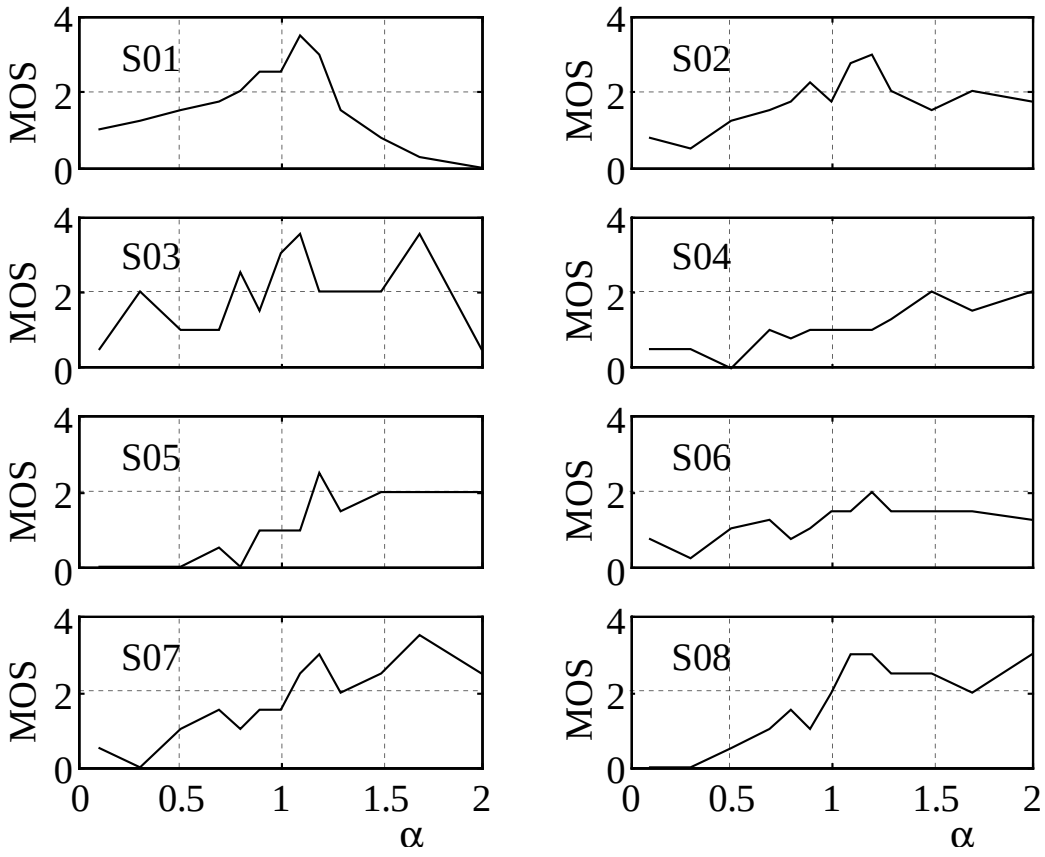

図 6.1: 被験者個別のサブトラクション係数 α と MOS との関係

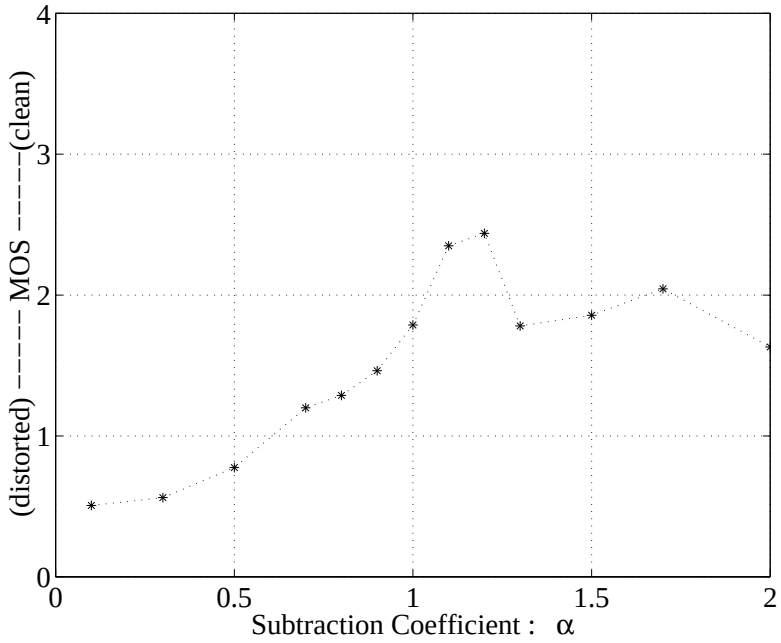

図 6.2: サブトラクション係数 α と 全被験者の平均 MOS との関係

英語音声データベース [67] に収録されている 2 つの英語会話音声を用いて、これまで聴取実験の被験者の経験がない大学院生 5 名に対して同様の実験を行なった。この結果、これまでに聴取実験の経験のない被験者であったこと、また彼らの英語に対する習熟度に起因し、主観評価実験により得られた MOS に若干のバラツキは見られたが、サブトラクション係数 α の最適値は 1.1 であった。この結果は、本節の実験結果と矛盾するものではない。

6.3 聴感上の印象を考慮した客観的評価尺度

6.3.1 聴感上の歪み印象に対応した客観的評価尺度概説

我々の聴覚は非常に優れた評価器であり、これに勝る評価器は存在しない。しかし、聴取実験を行なうためには莫大なコストが必要となり、被験者間、あるいは同一被験者内で

あっても評価実験ごとに主観評価値が異なる可能性がある。そこで、評価の効率、評価結果の安定性を考慮すると、客観的評価尺度による音声の品質評価が望ましい。

音声の客観的品質評価には、信号対雑音比 (Signal to Noise Ratio; SNR) がよく用いられている。しかし、SNR が主観評価に対応した評価尺度であるという保障は全くない。例えば、符号化音声の評価に関しては、主観的な品質は SNR 値による評価結果とは異なるという報告もあり [68]、SNR よりもスペクトル歪 (Spectral Distortion; SD) が利用されることが多い [59]。現在、あらゆる雑音に対して主観評価値と対応のよい客観的評価尺度は存在せず、それぞれの目的、雑音の種類に応じた評価尺度が提案されている。

主観評価に代わる代表的な客観的評価尺度として、International Telecommunication Union (ITU) が、300 Hz から 3400 Hz までの電話帯域に限定し、マスキング現象などの聴知覚特性を考慮した音声歪み評価尺度 Perceptual Speech Quality Measure (PSQM) を提唱している [69]。PSQM が主観評価値と非常に相性がよいことは、様々なデータベースを用いた大規模実験により確認された [70]。客観評価に要する演算量を削減することを目的に、PSQM の簡易モデルも提案されている [71]。川口らは、符号誤りのある音声をより精度よく評価するために、有声フレームと無声フレームにおける歪みの総量の違いを考慮し、PSQM を改良した W-PSQM を提案している [72]。また、小坂らは、電話伝送系において通話品質を劣化させる様々な要因を考慮し、MOS を予測するための客観的評価モデルを提案している [73] [74]。特定の電話回線に限定し、できる限り高精度の評価尺度を作成しようとする試みもある [75]。聴覚特性の客観的評価尺度への導入という観点では、ラウドネスに基づいた簡単なマスキングモデルを導入し、我々の聴覚系を考慮して線形周波数軸ではなく、Bark 軸 [76] 上でのスペクトル歪を計算する Modified Bark Spectral Distortion (MBSD) [77] などもある。更に、より厳密に聴覚末梢系における信号処理機構を考慮するため、Meddis が提案した内有毛細胞モデル [78] を利用した符号化電話音声の客観評価モデル [79] なども提案されている。このように、聴覚特性を導入した客観的評価モデルは数多く提案され、各々の使用条件における有効性も検証されている。但し、これらは、すべて符号化電話音声を評価対象としており、主観評価値 MOS を予測することを目的とした比率尺度である。比率尺度は、絶対零点を有し、尺度上での等間隔、等比率が保障される必要がある [63]。

6.3.2 客観的歪み評価尺度 ASD

本節では、様々な加法性雑音により音声に生じた歪み量を定量的に算出し、それに基づき雑音除去アルゴリズムの評価を行なう。この目的においては、主観評価値 MOS の絶対的な値を厳密に予測する必要はなく、雑音除去処理により歪み量をどの程度低減できるのかを定量的に知ることができれば十分である。つまり、本節で用いる尺度は、比率尺度である必要はなく、絶対的な零点は持たないが尺度上での等間隔が保障される間隔尺度 [63] であれば十分である。

本節では、加法性雑音により音声に生じた歪みの程度を客観的に評価することを目的に、同時マスキング現象、継時マスキング現象を考慮した歪み評価尺度 Auditory-oriented Spectral Distortion (ASD) を構築する。継時マスキング現象を考慮することにより、非定常雑音の評価も可能になる。我々の聴覚におけるマスキング特性は、信号の音圧レベルに依存することがわかっているが [80]、本節で作成する評価尺度ではマスキング特性は音圧に依存しないと簡略化する。これは、客観的評価尺度の作成にあたり、評価に要する演算量をできる限り少なくすることも重要であるためである [81]。

ASD 算出の流れを概説する。まず、フレーム長 21.3 msec、フレーム周期 5.3 msec の短時間フレームごとに、評価対象の信号に対し、同時マスキングと継時マスキングにおけるマスキング閾値を計算し、それらを統合して相対可聴閾値を計算する。クリーンな音声 /a/ の定常部を短時間フレームで切り出した信号に対し、ASD の算出過程を図 6.3 に示す。図中の細実線は音声の対数振幅スペクトル、＊印は同時マスキングのマスカー成分として検出されたスペクトル成分、点線は各マスカーに対するマスキング領域、破線は継時マスキングを考慮した過去の短時間フレームで計算された相対可聴閾値、太実線が本フレームにおける相対可聴閾値を表す。

相対可聴閾値以下のスペクトル成分は、例え歪みが生じたとしても、我々には知覚できないと考えられる。そこで、各短時間フレームごとに、相対可聴閾値を越える成分のみに対して

$$\mathrm{ASD} = \sqrt{\underset{i}{\mathrm{MEAN}}\{S_{target}(i) - S_{clean}(i)\}^2} \ [\ \mathrm{dB}\ ] \tag{6.1}$$

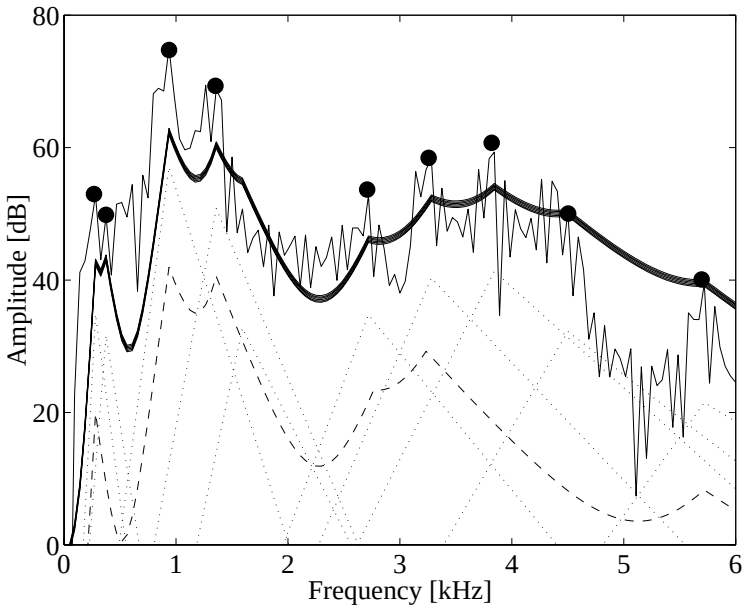

図 6.3: ASD 算出過程の解説図

により歪み量を計算する。ここで、$S_{target}(\omega)$、$S_{clean}(\omega)$ は、評価対象音声、クリーンな音声の振幅スペクトルとする。また、周波数 i は、評価対象周波数の 100 Hz から 6kHz のうち、評価対象音声の振幅スペクトル $X_{target}(\omega)$ が、マスキング閾値を越える周波数である。従って、評価に用いられる周波数は、各短時間フレームごとに異なる。評価対象音声の客観的な歪み量 ASD 値は、各フレームごとに式 (6.1) で求められる歪み量の音声区間における平均値と定義する。

相対可聴閾値算出の詳細は付録 A 1 に記しており、ASD の間隔尺度としての妥当性については付録 A 2 で検証している。

6.4 客観的評価尺度による雑音除去アルゴリズムの評価

目的：

本節では、客観的評価尺度 ASD により、提案法と従来法との非定常雑音により生じた聴感上の歪み感低減能力について調査検討する。

実験条件：

客観的評価尺度 ASD を基準とした雑音除去実験に用いた音声は、ATR 音声データベース [54] に収録されている男性話者 mht 氏発話の日本語連続母音 /ao/ である。このクリーンな音声 /ao/ の /a/ から /o/ へのわたり部に、2 kHz-3 kHz のランダム帯域雑音を突発的に、SNR が 10 dB になるように付加する。このデータは、客観的評価尺度 ASD の妥当性検証 (付録 A 2) に用いたものと同様である。

実験は、サブトラクション係数 α を 0 から 0.1 刻みで 2.0 まで変化させ、客観評価尺度 ASD を基準とした場合のサブトラクション係数 α の最適値を求める。これが、主観評価値 MOS を評価基準とした場合と同様の傾向を示せば、本実験からも客観評価尺度 ASD の妥当性が証明できる。

また、聴感上の歪み感低減能力に関する提案法と従来法との性能比較として、第 2.4 節で用いた 2 つの突発性雑音を付加した単母音 /a/ に対する雑音除去音声を ASD により評価する。

実験結果：

聴感上の印象を考慮した客観的歪み評価尺度 ASD を評価基準とし、サブトラクション係数 α を変化させ、雑音除去を行なった結果を図 6.4 に示す。この場合、サブトラクション係数 α の最適値は 1.2 である。

また、SNR 0 dB の雑音付加音声に対し、第 2 章で解説した 1ch 雑音適応型 SS、2ch 雑音推定型 SS、3ch 遅延和アレー、3ch Griffiths-Jim 適応型アレーにより雑音除去を行なった音声に対する ASD 値を図 6.5 に示す。

考察：

本節の実験において、ASD を評価基準とした場合のサブトラクション係数 α の最適値は、主観評価値 MOS を評価基準にした場合 ($1.1 \leq \alpha \leq 1.3$ に最適値は存在するが明確ではなく、$1.3 < \alpha \leq 2.0$ でも歪みは比較的小さい) と類似しており、ASR のフロントエンドとしての最適値 (明確に 1.0 が最適値) とは異なることがわかった。つまり、聴感上の歪

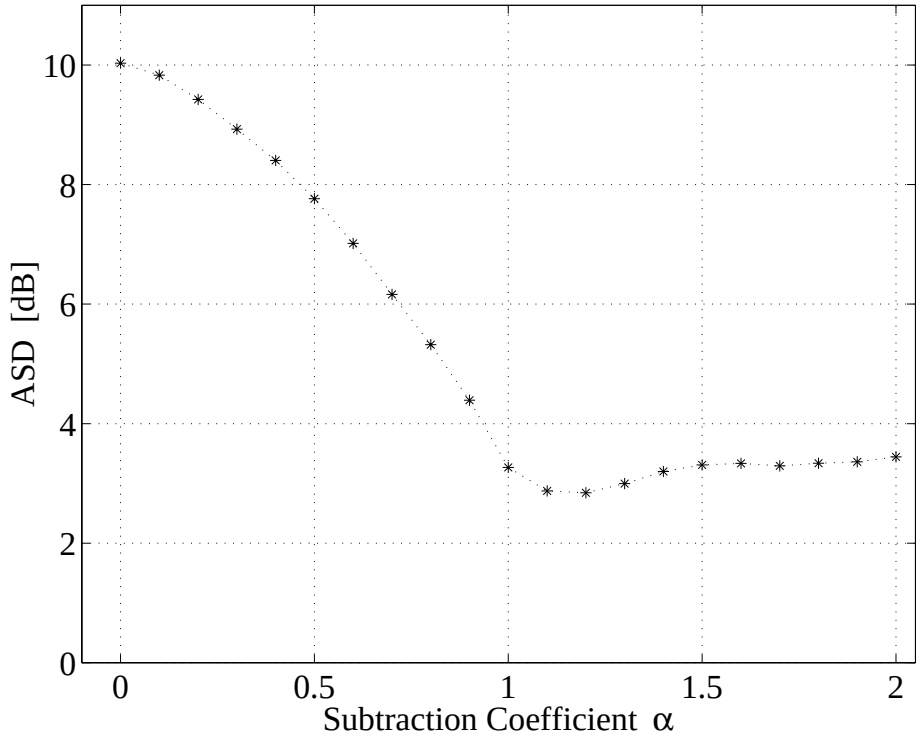

図 6.4: サブトラクション係数 α と ASD 値との関係

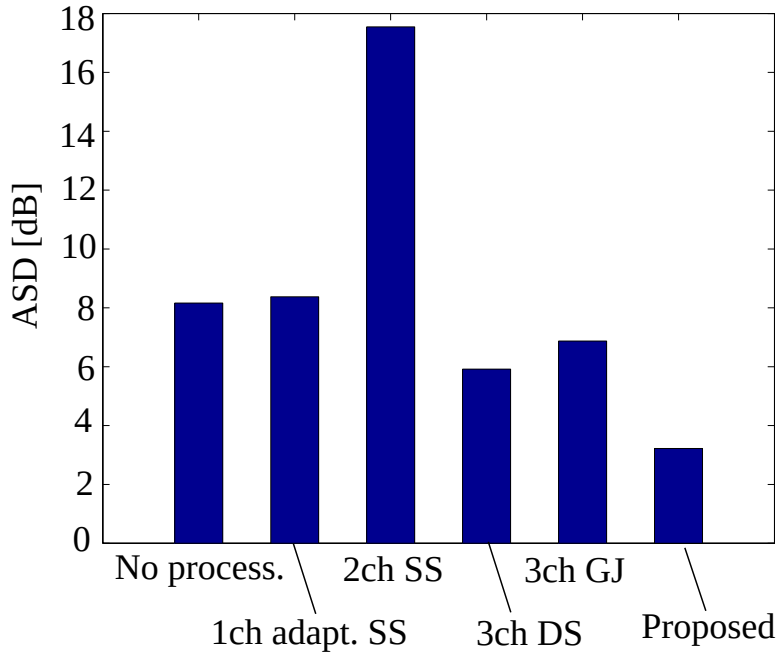

図 6.5: 客観的評価尺度 ASD に基づいた提案法と従来法との雑音除去能力の比較 (左から順に、雑音付加音声の ASD 値 (No process.)、1ch 雑音適応型 SS(1ch adapt. SS)、2ch 雑音推定型 SS (2ch SS)、3ch 遅延和アレー (3ch DS)、3ch Griffiths-Jim 適応型アレー (3ch GJ)、提案法 (Proposed) による雑音除去音声の ASD 値)

み感を評価するためには、SNR や第 5 章で定義した LPC-SED ではなく、本章で作成した ASD を用いることが望ましい。評価尺度 ASD の妥当性に関しては、付録 A 2 において、聴覚特性を考慮しない客観的歪み評価尺度 SD との比較により、ASD が聴感上の歪み感をより忠実に評価できることも検証している。以上より、加法性雑音により生じた聴感上の歪み感の評価において、客観的評価尺度 ASD が適していることがわかる。また同時に、ASD による評価において、提案法が非定常雑音を除去できることを確認した。

提案法と従来法の聴感上の歪み感低減能力の比較に関しては、ASD 値による各雑音除去アルゴリズムの評価結果 (図 6.5) より、提案法が最も優れていることがわかる。図 6.5 において、2ch 雑音推定型 SS により雑音除去を行なった結果、歪み量が 2 倍以上に増大している。これは 2ch 雑音推定 SS が、音声区間中で発生する突発的な雑音を推定できなかっただけでなく、目的信号である音声に新たに歪みを生じさせたためである。また、LPC-SED 値による評価 (図 5.7) では、遅延和アレーと Griffiths-Jim 適応型アレーはほぼ同等の雑音除去能力を示していたが、聴感上の歪み印象に対応した評価尺度 ASD により評価を行なうと、遅延和アレーの方が雑音除去能力が優れていることがわかる。これは、適応型アレーが、我々が敏感に反応する雑音の立ち上がり部を除去できないためである。

客観的評価尺度 ASD の妥当性は、以下の考察により証明することも可能である。目的信号と一つの雑音のみが存在する状況において、サブトラクション係数 α が 1.0 よりも大きい場合、雑音除去音声の振幅スペクトルは本来の大きさよりも小さくなる。これより、我々の聴覚系は、雑音による極わずかな歪みにも敏感に反応する一方、振幅スペクトル形状の若干の崩壊には鈍感であろうことが予想される。雑音スペクトルを本来の大きさよりも大きく見積もって引き去ることは、スペクトルのピーク部、ディップ部に無関係に、全周波数帯域に等しく影響を及ぼす。しかし、スペクトルのディップ部あるいはスペクトルレベルが小さな周波数帯域に関しては、マスキングの影響により我々がそれらを知覚することは困難である。以上より、マスキングの影響を受けない振幅スペクトルのピーク部に生じた歪みのみを評価する ASD は、我々の聴覚系に近い評価尺度であると言える。

6.5 まとめ

本章では、主観評価実験、聴感上の歪み印象に対応した歪み評価尺度 ASD による客観的評価により、提案法が非定常雑音により生じた聴感上の歪み感を低減できることを確認した。提案法と従来法との聴感上の歪み感低減能力の差異に関しては、特に突発性雑音の除去においては提案法の優位性を確認することができ、適応フィルタを利用する Griffiths-Jim 型アレーは、突発性雑音の立ち上がり部を除去できないため、聴感上の印象を向上させることが困難であることを確認した。

サブトラクション係数 α の最適値は、評価基準として LPC-SED 値を用いた場合 ($\alpha = 1.0$) と、MOS あるいは ASD 値を用いた場合 ($1.1 \leq \alpha \leq 1.3$) とでは異なることがわかった。提案法は、一つのパラメータ値 α の変更により、ASR の雑音除去フロントエンド、聴感上の歪み感低減の両目的に応用できるため、汎用性が高い雑音除去アルゴリズムである。

第 7 章

実環境における有効性検証

7.1 はじめに

第 3 章において雑音除去アルゴリズムの定式化を行ない、それ以後の章では提案法の基礎的な性能評価、ASR のフロントエンドとしての有効性、聴感上の歪み感低減能力について検証してきた。その結果、あらゆる評価実験において、提案法は高精度の雑音除去が可能であり、特に突発雑音のような従来法が苦手としてきた非定常雑音の除去に関しては、提案法は従来法よりも雑音除去能力が高いことを確認した。但し、前章までの評価実験は、暗騒音や残響が存在しない環境を想定した計算機上での雑音除去実験の結果である。

本章では、提案法の実環境適応性を調査するために、実環境における雑音除去実験を行なう。雑音除去実験に用いた実環境は、暗騒音はほとんど存在せず残響時間が非常に短い環境、我々が日常生活で使用するような一般的なオフィス、暗騒音が多く残響時間も非常に長い劣悪な環境の 3 種類である。

7.2 実環境における雑音除去実験

目的：

本節では、実環境における雑音除去実験を行ない、提案法の実環境適応性について検討する。また、提案法と同一マイクロホン配置の遅延和アレーとの残響のある環境下での雑音除去能力の差異についても比較検討する。

実験条件：

雑音除去実験を行なった環境は、暗騒音が存在せず残響も少ない防音室と、適度な残響のある一般的なオフィス、非常に残響がある工作室の 3 種類である。

防音室は、4.0 m × 4.5 m × 2.4 m の大きさで、遮音処理、防音処理を施されているが、反射音の影響を抑えるために、更に室側面に吸音材としてグラスウールを設置した。TSP 信号 [82] を建築音響特性を測定するために最適化した OATSP 信号 [83] を用いて室インパルスレスポンスを測定し、インパルス自乗積分法 [84] により各周波数ごとに求めた残響時

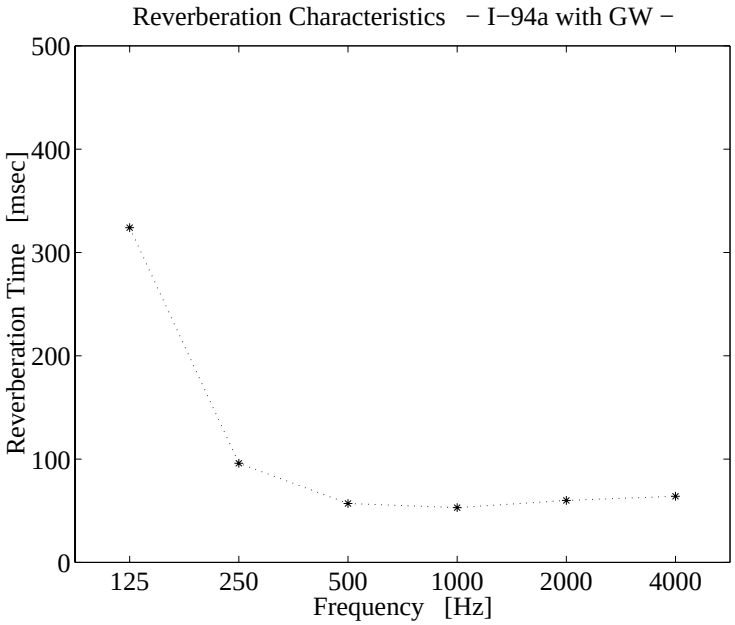

図 7.1: グラスウールを設置した防音室の残響特性

間を図 7.1 に示す。防音室の残響特性は、500 Hz 以上の帯域においては残響時間が約 50 msec と非常に短いが、低周波数帯域における残響時間はかなり長い。これより、本防音室は、音声の基本周波数帯付近に反射音が存在することがわかる。 実験に使用したオフィスは、4.0 m × 5.4 m × 2.7 m と防音室よりやや広く、机や本棚などが設置され、床にはカーペットが敷かれており、暗騒音レベルは約 33 dB(A) である。図 7.2 に、このオフィスの残響特性を示す。一方、工作室は、5.1 m × 10.5 m × 2.8 m と防音室の約 3.5 倍の室容積を持ち、工作機器などが設置され、遮音・防音処理は一切施されておらず、暗騒音レベルは約 34 dB(A) である。工作室の残響特性を図 7.3 に示す。工作室は、500 Hz 以上の周波数帯域における残響時間は 500 msec 程度であるが、防音室と同様に低周波数帯域での残響時間が長いため、聴感上は残響時間が非常に長く感じられる。

実環境における雑音除去実験では、クリーンな音声と雑音をそれぞれ別のスピーカから実音場に放射し、それをサンプリング周波数 48 kHz、16 bit 量子化でデータレコーダに一旦録音したものを計算機に取り込み、雑音除去を行なった。実環境における残響特性測定

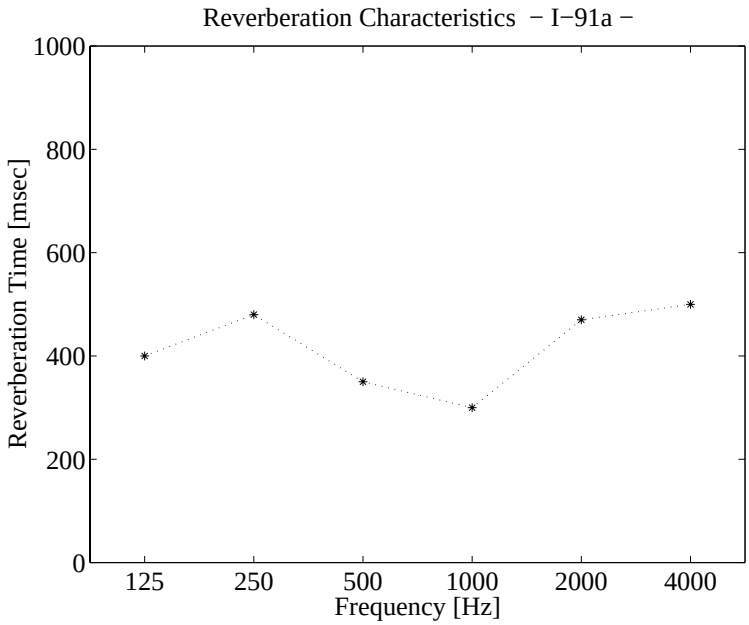

図 7.2: オフィスの残響特性

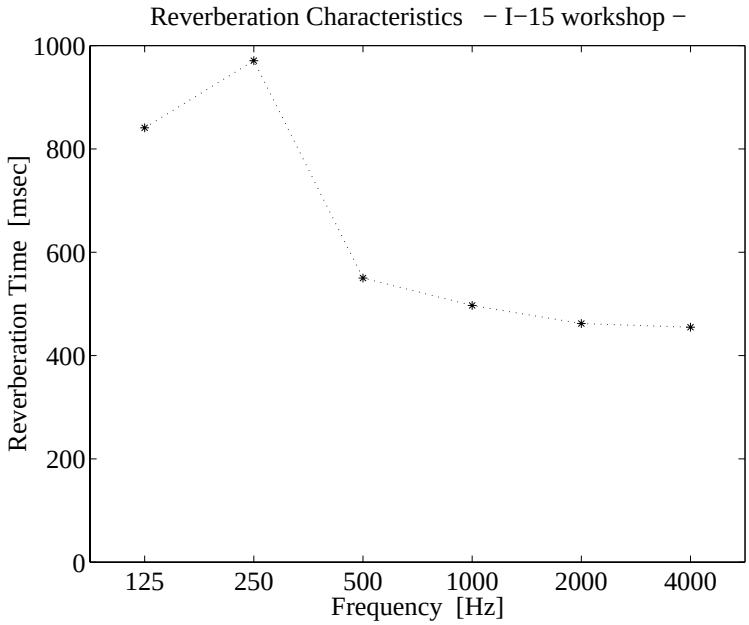

図 7.3: 工作室の残響特性

表 7.1: 残響特性測定および実環境実験に使用した機器一覧

機器	メーカ	型番
DAT Deck	SONY	DTC-57ES
Power Amp.	SONY	TA-F333ESJ
Loud Speaker	AURATONE	5PSC
Speaker Stand	ULTIMATE	TS-40
Microphone	SONY	ECM-44B
Mic. Stand	SONY	B-305B
Mic. Amp.	TASCAM	MA-8
DAT DATA Recorder	TEAC	RD-135T
Sound Level Meter	RION	NL-04
DAT	SONY	DT-60RA

およびデータ作成には、表 7.1 に示す機材、図 7.4 に示すような素子間隔 10 cm の 3 ch 直線配列マイクロホンアレーを使用した。

音声信号を放射するスピーカはマイクロホンアレーに対して正中面方向、雑音信号を放射するスピーカは右 35 °方向に、マイクロホンアレーとの距離がそれぞれ 2.0 m となるように設置した。音声は ATR 音声データベース [54] に収録されている男性話者 mht 発話の単母音 /a/、/i/、/u/、/e/、/o/ の 5 種類とし、雑音は第 2.4 節で用いた継続時間 50 msec、帯域幅 200 Hz、中心周波数 1.5 kHz または 2.5 kHz の連続した 2 つの狭帯域雑音とする。音声と雑音の音圧レベルに関しては、音声の SNR が約 0 dB になるように調整した。なお、実環境において収音した雑音付加音声の SNR は、録音再生系において生じる乗法性歪み [39] の影響をできる限り少なくするため、クリーンな音声のみを音場に放射し、収音した音声を基準信号として計算する。

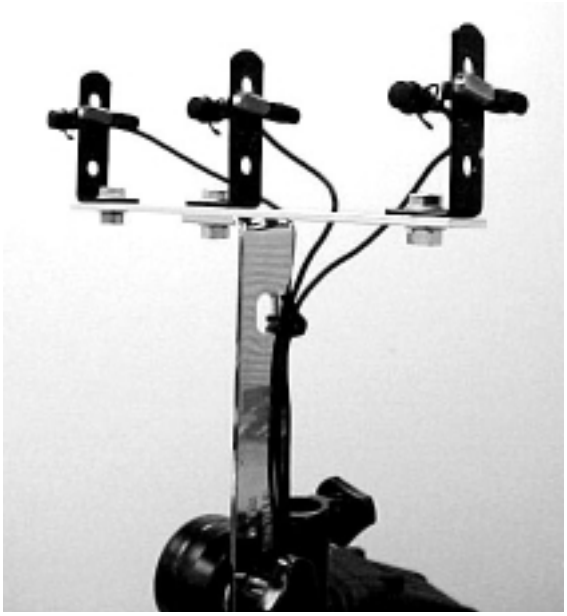

図 7.4: 実環境実験に使用した 3 ch 直線配列マイクロホンアレー

実験結果：

まず、防音室にて収音した単母音 /a/ に対し、雑音除去前後の LPC-SED 値と、雑音除去によるその改善量を図 7.5 に示す。クリーンな音声のみをスピーカで実環境に放射し、収音した信号に対する雑音付加音声の SNR は −0.6 dB であり、音声と雑音が共存する区間 (図 7.5 2 段目の雑音付加波形の斜線部) における LPC-SED 値の平均改善量は 3.14 dB であった。これは、同一環境で収音した 10 個の信号に対する雑音除去結果の平均値であり、標準偏差は 0.10 dB であった。

実環境における雑音除去実験の結果として、日本語 5 母音に対する LPC-SED の平均改善量を図 7.6 に示す。図 7.6 には、理想的な環境としての計算機シミュレーション、防音室、残響の少ないオフィス、残響の多い工作室における雑音除去実験の結果を示す。

ここで、実環境で収録した雑音付加音声に対し、信号到来方向を既知として最適化した遅延和アレー、方向も自動推定する提案法により雑音除去を行なった結果より、提案法と遅延和アレーとの雑音除去能力の差異を図 7.7 に示す。図 7.7 は、日本語 5 母音に対し、提案法あるいは遅延和アレーにより雑音除去を行なった音声に対して LPC-SED 値の平均改

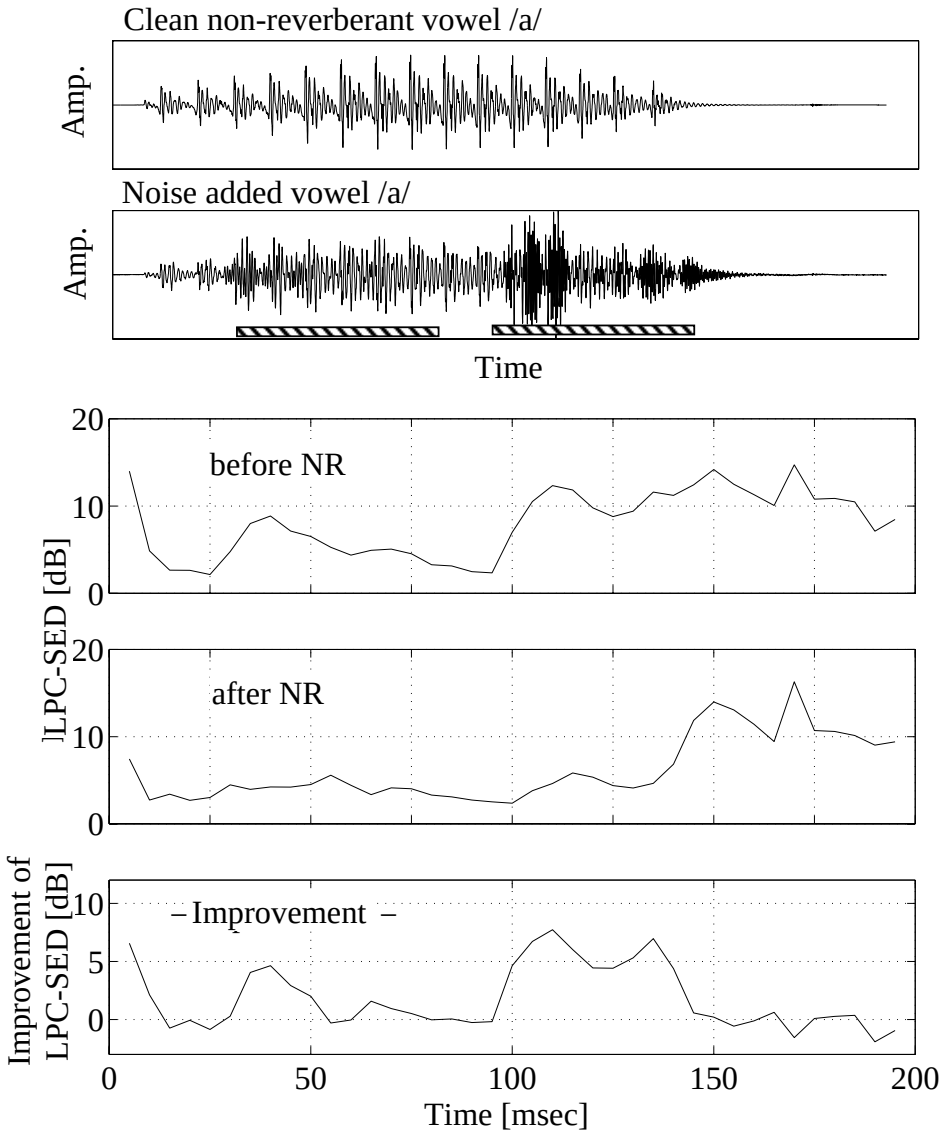

図 7.5: 防音室における雑音除去実験の結果 (上から順に、クリーンな音声 /a/、実環境において雑音を付加した音声 /a/、受音信号の LPC-SED 値、雑音除去後の LPC-SED 値、雑音除去による LPC-SED 値の改善量)

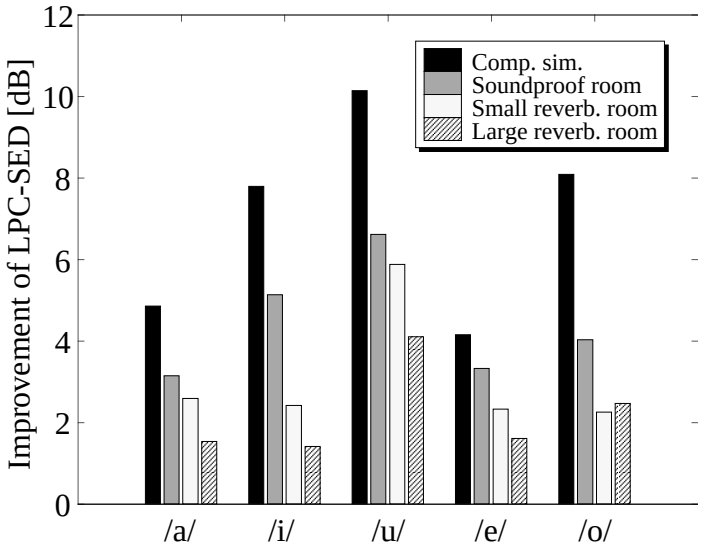

図 7.6: 雑音除去による LPC-SED 値の平均改善量。左から、計算機シミュレーション、防音室、残響の少ないオフィス、残響の多い工作室における実験結果を表す。

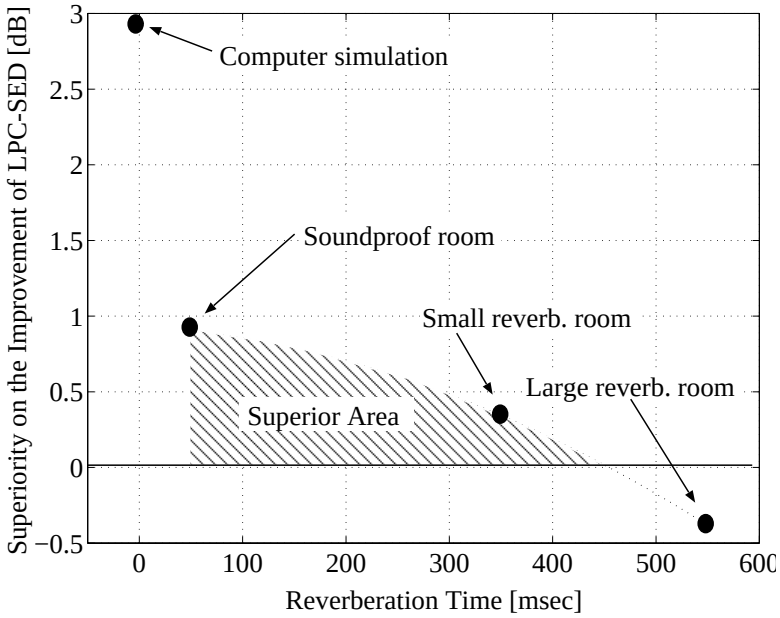

図 7.7: 実環境における提案法の遅延和アレーに対する優位性

善量を計算し、提案法と遅延和アレーとの差、すなわち提案法の優位性を表す。図 7.7 の横軸は 500 Hz における残響時間、縦軸は提案法の優位量として LPC-SED 値の平均改善量の差 (提案法による平均改善量から最適化遅延和アレーによる平均改善量を差し引いた値)、図中の斜線部は提案法が遅延和アレーよりも優位である領域を表す。

考察：

実環境における雑音除去実験の結果より、提案法は、残響時間が長くなるに従い雑音除去精度の低下は生じるが、残響のある環境においても有効であることがわかった。実環境における提案法と遅延和アレーとの雑音除去能力に関しては、雑音除去による LPC-SED 値の平均改善量の差を示した図 7.7 より、残響時間が 450 msec 程度までの適度な残響環境においては、提案法の方が雑音除去能力が高いであろうと予測できる。

残響のある環境における提案法の雑音除去精度をより向上させるためには、残響にロバストな方向推定アルゴリズムを構築する必要がある。なお、方向推定に誤りが生じても雑音除去の効果があった理由は、提案法が特定の雑音方向以外から到来する信号成分もある程度は除去できるためである。これは、提案法が雑音スペクトルを推定する際に、特定の一方向から到来する信号を正確に推定するだけでなく、若干の誤差は生じるものの目的信号方向以外から到来するあらゆる雑音成分を推定できるためである。

表 7.2: 残響環境下での提案法および従来法の雑音適応表

雑音除去アルゴリズム (素子数)	SS (1)	適応 SS (1)	2ch SS (2)	遅延和アレー (多数)	適応アレー (3)	提案法 (3)
残響環境下の定常雑音	△	△	×	○	○	○
残響環境下の非定常雑音	×	×	×	○	○	○

7.3 まとめ

本章では、残響や暗騒音のある実環境における提案法の有効性について検証した。実環境における雑音除去実験の結果、提案法は、残響時間に比例して精度は低下するものの残響環境においては非定常雑音を除去可能であることを確認した。

残響環境下での提案法および従来法の雑音除去能力については、本章の実験結果から、残響時間が 450 msec までの適度な残響環境においては、提案法は、一般に残響の影響を受けにくいと言われている遅延和アレー (音声の到来方向は既知) よりも優れていることがわかった。しかし、残響時間が長い環境においては、提案法は、最適化した遅延和アレーの雑音除去能力には及ばない。以上より、実環境適応性に関しては、表 7.2 に示すように、提案法は、遅延和アレーや適応型アレーに対して雑音除去能力に有意な差はないが、有効であると考えられる。SS および雑音適応 SS は、残響時間が短く、雑音が定常であればある程度は有効であるが、雑音が非定常であれば除去することはできない。2ch SS は、残響が存在すると、音源と受音点間の系が線形時不変の仮定を外れるため、雑音スペクトルの推定において大きな推定誤差が生じることになる。

一般に、小規模マイクロホンアレーによる雑音除去法は、残響のある環境では雑音除去性能が低下することがわかっている。従って、提案法の残響耐性を向上させるためには、方向推定アルゴリズム残響耐性を向上させるとともに、素子数を増加させることについても検討する必要がある。

第 8 章

結論

8.1 本論文の要約

本論文では、従来の雑音除去法では扱うことが困難であった時々刻々変化する非定常雑音の除去を目標に、マイクロホン対による雑音除去アルゴリズムを構築した。雑音除去は以下の手順で行なう。

1. 目的信号と最も優勢な雑音の 2 つの信号到来方向の推定

2. 目的信号を完全に抑圧し、雑音スペクトルの推定

3. スペクトル領域において、受音信号から雑音成分の除去

信号到来方向の推定および雑音スペクトルの推定を行なう際、2 つのマイクロホン（マイクロホン対）の受音信号を利用し、特定方向から到来する信号を完全に抑圧するような減算型ビームフォーマを解析的に設計することが本研究の特徴である。減算型ビームフォーマは、遅延和ビームフォーマと比較し、より少ないマイクロホン数でより高い雑音除去性能が得られるため、マイクロホンアレーによる空間フィルタリング手法の主流となっている。現在提案されている減算型ビームフォーマの多くが、適応フィルタによるビームフォーミング手法を採用しているが、それらは時間的に変化の激しい非定常雑音や目的信号と相関の高い雑音の除去は困難であり、これらの問題は本論文においても確認した。本手法は、定常雑音、非定常雑音という枠を越え、また目的信号と雑音がともに音声であるような非常に困難な状況においても雑音を除去できる。また、受音信号から雑音成分を除去する際には、Boll によって提案され、現在でも ASR のフロントエンドなどに用いられている周波数領域でのフィルタリング手法である SS を利用する。従来の SS は、単点受音信号に対して非音声区間で雑音の推定を行なうため、音声／非音声区間の判定が必要であり、定常雑音にしか対応できなかった。SS 改良法は数多く提案され、音声／非音声区間の判定機構は不必要になったが、非定常雑音を除去することは困難であった。しかし、本雑音除去法は、信号の到来方向という空間情報を利用した空間フィルタリングにより、時々刻々雑音スペクトルを推定することが可能であり非定常雑音にも対応できる。

表 8.1: 各雑音除去法の雑音適応表

雑音除去アルゴリズム (素子数)	SS (1)	適応 SS (1)	2ch SS (2)	遅延和アレー (多数)	適応アレー (3)	提案法 (3)
定常雑音	◎	◎	○	○	◎	◎
音源が移動する定常雑音	○	○	×	○	◎	◎
残響環境下の定常雑音	△	△	×	○	○	○
緩変化の非定常雑音	×	○	△	○	○	◎
急激に変化する非定常雑音	×	×	×	○	×	◎
音声区間中で発生する雑音	×	×	×	○	△	◎
音源が移動する非定常雑音	×	△	×	○	○	◎
残響環境下の非定常雑音	×	×	×	○	○	○

最後に、提案法と従来法との雑音除去能力の差異を明確にするため、本論文で明らかになった各雑音除去アルゴリズムの様々な雑音への適応性を表 8.1 にまとめて示す。提案法は、様々な実験を通し、従来の雑音除去法では対象外とされてきた非定常雑音を除去できることがわかった。定常雑音であれば如何なる雑音除去法も対応できるが、非定常的な雑音、より現実的な雑音には対応できない手法が多いのが現状である。雑音の音響的性質が急激に変化する場合や音声区間中で突発的に発生する雑音に対し、提案法の雑音除去能力は他を大きく凌いでいる。このような雑音は、決して稀な存在ではなく、ドアの開閉音、足音、あるいは話し声など一般的な生活環境にはたくさん存在する。提案法のような少数素子のマイクロホンアレーを用いた雑音除去法では適応処理の利用が常識となっているが、適応処理では信号の急激な変化に追従することが難しく、適応処理を採用した手法では雑音の音響的特性が大きく変化する区間では雑音を除去できない。一方、我々の聴覚は、動物的本能により、定常的な信号よりも時間・周波数方向に動きのある非定常な信号ほど敏感に知覚する。よって、聴感上の印象に基づいた評価において、適応処理を利用した手法

は、我々の知覚に強い影響力を持つ非定常雑音を除去できないため、同素子数の古典的な遅延和アレーよりも評価が低い。なお、遅延和アレーは、線形処理のため雑音除去音声は高音質であるが、高精度の雑音除去を実現するためには莫大な素子数を必要とするため使用環境が限定されてしまう。例えば、雑音除去アルゴリズムを可搬型 ASR や補聴器へ導入する場合、大規模マイクロホンアレーを用いることは困難であり、遅延和アレーでは十分な雑音除去精度を期待できない。そこで、少数素子の減算型アレーの導入が考えられるが、雑音除去が必要とされているより現実的な環境においては、表 8.1 より現在の主流である適応型アレーよりも提案法が適していることがわかる。

提案法は、具体的な応用例として ASR のフロントエンド、補聴器などを想定した評価において、雑音により生じた音声認識率の低下、あるいは音声の歪み感を良好に回復できることも確認した。音声認識の雑音耐性向上に関しては、提案法は遅延和アレーよりも圧倒的に優れており、3 素子の提案法が 6 素子の遅延和アレーと同等の性能を有することもわかった。以上より、本雑音除去アルゴリズムは、本論文で目的として掲げた非定常雑音の除去を達成できており、ASR のフロントエンド、あるいは補聴器などの聴感上の歪み低減を必要とする機器への利用が期待できる。

8.2 今後の課題

本雑音除去アルゴリズムの実用性をより高めるためにも、今後、以下に挙げる課題を解決することが望ましい。これにより、提案法の工学的有用性は益々向上し、提案法が ASR のフロントエンド、補聴器のように現段階で早急に雑音除去技術を必要としている分野のみならず、現在研究が進められている様々な音響処理技術の将来的な実用化にも大きく貢献できるであろうと期待できる。

近接音源への対応

本雑音除去アルゴリズムは、目的信号と雑音はともに平面波としてマイクロホンアレーに到来するという仮定のもと構築した。これより、音源はマイクロホンアレーから離れていなければならないという制約が生じる。

音源とマイクロホンアレーとの距離が近い場合には、マイクロホン間隔を狭めることで平面波として扱うことは可能であるが、空間分解能が低下するという問題が生じる。従って、音源とマイクロホンアレーとの距離を推定し、平面波として扱うことができないほど音源がマイクロホンアレーに近い場合は、信号は球面波で到来するという前提で雑音除去アルゴリズムの再定式化を行なうことが望ましい。なお、音源の方向推定のみならず距離推定を高精度で行なうためには、マイクロホンアレー形状 (マイクロホンの配置) についても再検討する必要があろう。

複数の雑音・残響のある環境への対応

本論文では、基本的に目的音源と雑音源とが一つずつ存在する状況を想定していた。しかし、一般に雑音除去が必要とされる環境には、複数の雑音源や残響が存在する。

本雑音除去アルゴリズムは、最も優勢な雑音を完全に抑圧するとともに、残響も含め、それ以外の雑音もある程度は除去できることが減算型ビームフォーマの指向特性から類推できる。本雑音除去アルゴリズムの複数雑音源あるいは残響への有効性を明確にするため

には、雑音源の数および配置に対し、雑音除去能力を定量的に評価することが期待される。

また、実環境における雑音除去実験の結果より、計算機シミュレーションにおける理想的環境と実環境との雑音除去精度の差は、方向推定の誤差に起因することがわかった。推定誤差が小さい場合には、雑音除去精度が大幅に低下することはないことは確認した。しかし、実環境には無数の反射音 (残響) が存在し、本来推定されるべき直接音方向とは全く異なる反射音方向を誤って推定すると誤差量が大きくなり、雑音除去精度が低下する。従って、複数の雑音源や残響のある環境において、本雑音除去アルゴリズムをより有効に働かせるためには、雑音や残響によりロバストな方向推定アルゴリズムを構築することが望ましい。本雑音除去アルゴリズムはモジュール化されており、それぞれが独立しているため、方向推定アルゴリズムを入れ換えた場合でも、雑音スペクトルの推定、雑音スペクトルの除去の各モジュールはそのまま利用できる。

リアルタイム雑音除去システムの実現に向けて

本雑音除去アルゴリズムは、解析的にビームフォーミングを行なうため、適応処理を利用した減算型ビームフォーミング手法よりも演算量は圧倒的に少ないと予想される。リアルタイム処理を前提に実装を行なう場合には、演算量と雑音除去精度の関係をより詳細に調査し、雑音除去の精度に大きな影響を与えないような処理は簡略化することにより、更なるコストパフォーマンスの向上が期待できる。

例えば、マイクロホンアレーのマイクロホン数 (素子数) は 3 本でよいのかを再検討することも必要かも知れない。マイクロホン数を増やしてより多くの空間情報を用いることが、全体的な演算コスト削減に繋がる可能性もある。ディジタル信号処理の DSP による実装を考えた場合、一般的な 2ch 用 DSP と特殊用途である 3ch 用 DSP とでは要するコストの差は非常に大きい。本雑音除去アルゴリズムの基礎概念は、マイクロホン対であり、3 本目のマイクロホンは方向推定も含めて雑音除去の精度向上を実現するために導入した。例えば、本雑音除去アルゴリズムを回頭可能なロボットの受音系に導入するような場合、2ch マイクロホンアレーでも十分な雑音除去精度を実現できる可能性もある。

付録： 聴感上の印象を考慮した評価尺度 ASD

A 1. 聴覚特性を考慮した加法性雑音に対する歪み評価尺度の構築

A 1.1 客観的歪み評価尺度 ASD の概要

本研究で提案する Auditory-oriented Spectral Distortion (ASD) は、加法性雑音により音声に生じた歪みの程度を客観的に評価することを目的に、同時マスキング現象、継時マスキング現象を考慮した聴感上の歪み印象に対応した客観的評価尺度である。我々の聴覚におけるマスキング特性は、信号の音圧レベルに依存することがわかっているが [80]、本節で作成する評価尺度ではマスキング特性は音圧に依存しないと簡略化する。これは、客観的評価尺度の作成にあたり、評価に要する演算量をできる限り少なくすることも重要であるためである [81]。

ASD では、以下の手順で音声に生じた歪み量を計算する。

1. 短時間スペクトルの算出 (図 A.1 の細実線)
2. 同時マスキングのマスカー検出 [A 1.2.a] (図 A.1 の＊印)
3. 2. の各マスカーに対するマスキング領域の算出 [A 1.2.b] (図 A.1 の点線)
4. 継時マスキングのマスキング閾値の算出 [A 1.3] (図 A.1 の破線)
5. マスキング閾値の統合 [A 1.4] (図 A.1 の太実線)
6. ASD 値の算出 [A 1.5]

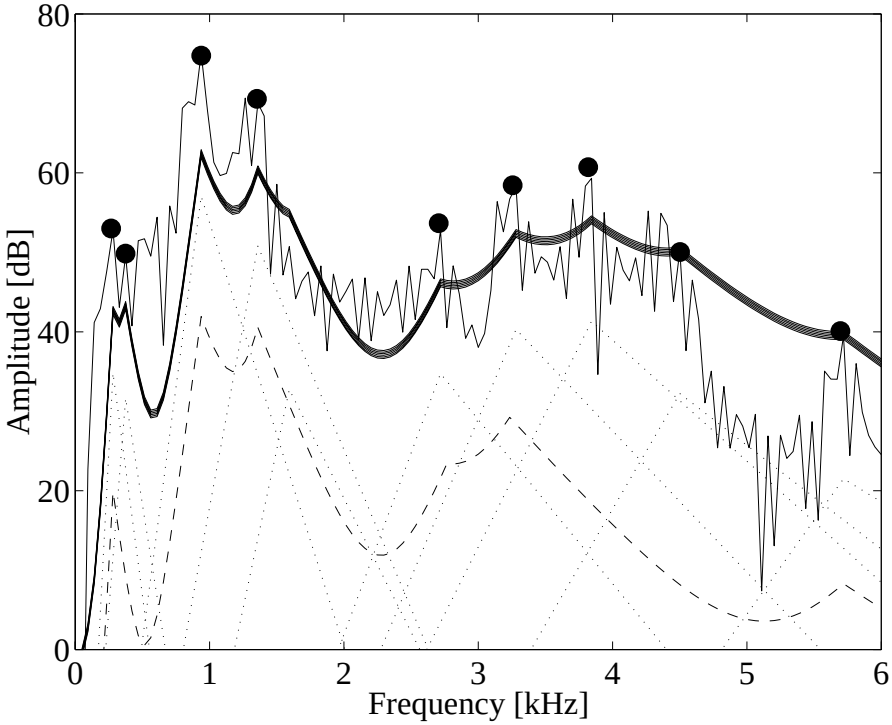

図 A.1: ASD 算出過程の解説図

A 1.2. 同時マスキング特性の実装

同時マスキング特性の実装に関しては、各短時間フレーム内でマスカー成分を検出し、各々のマスカー成分に対してマスキング領域を計算する。

a. マスカー成分の検出

我々の音の大きさ (ラウドネス) に対する知覚特性は、周波数ごとに異なっており、最小可聴閾値 (絶対可聴閾値) として測定されている [85]。図 A.2 は、騒音計の周波数補正特性にも採用されている絶対可聴閾値を 2 次関数で近似した A 特性曲線である [84]。A 特性フィルタに通した評価対象音声より短時間振幅スペクトル (図 A.1 の細実線) を計算し、これに対して同時マスキングにおいてマスカーとなり得るスペクトル成分 (図 A.1 の＊印) を検出する。

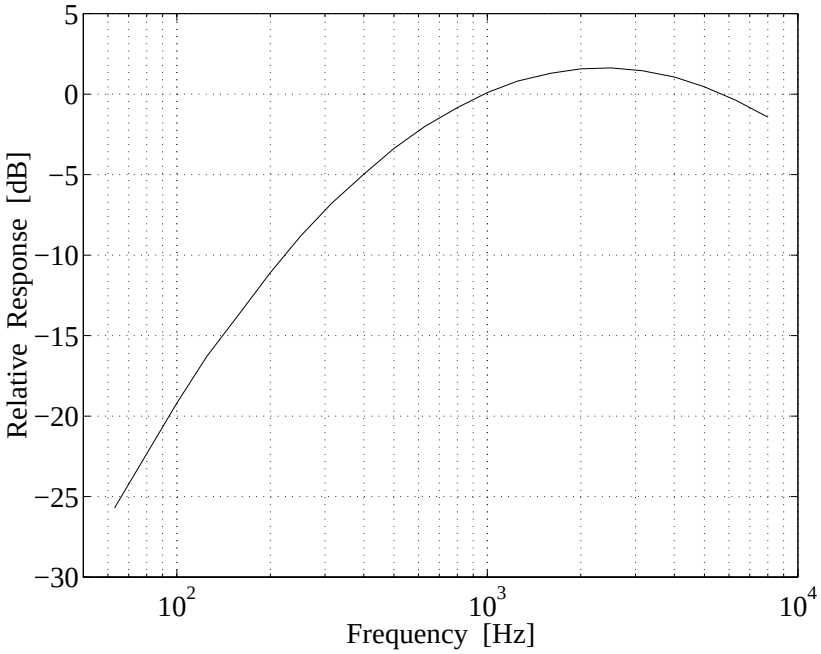

図 A.2: 絶対可聴閾値を近似した A 特性曲線

A 特性フィルタ通過後の評価対象信号の振幅スペクトル、すなわち我々が知覚しているであろう振幅スペクトル $X(\omega)$ に対し、式 (A.1) を満たすスペクトル成分 $X(k)$ をマスカー成分の候補として検出する。

$$\begin{cases} X(k) > X(k-1) \\ X(k) \geq X(k+1) \\ X(k) - X(k+j) > 3 \text{ [dB]}, \ j = 1, 2, \cdots, J \end{cases} \tag{A.1}$$

ここで、探索範囲 J は、k を中心周波数と考えた場合の聴覚フィルタ幅 ERB(k) [86]：

$$\text{ERB}(k) = 24.7(4.37 \cdot k/1000 + 1) \qquad \text{[Hz]}, \tag{A.2}$$

により決定するが、高域では ERB(k) の値が非常に大きくなるため、天井値を設けて

$$J = \min\left\{\text{ERB}(k), \ \text{ERB}(1.5[kHz])\right\} \tag{A.3}$$

とする。

更に、聴覚フィルタ内に含まれる情報はそれらが独立に処理されることはないという報告 [87] に基づき、同一聴覚フィルタ内に存在可能なマスカーの数は高々 1 つと制限する。ある聴覚フィルタ内に複数のマスカー候補が存在する場合、それらのうちで最もパワーの大きなマスカー候補をマスカー成分と定める。

以上のマスカー検出手法は、International Standardization Organization (ISO) が作成した MPEG 1 オーディオ符号化アルゴリズム [88] を参考にしたものである。MPEG アルゴリズムでは、人の聴覚特性を考慮した非線形周波数写像のスケールとして Bark を採用しているが、Bark よりも ERB の方が望ましいことから [76]、本評価尺度では ERB を採用する。また、聴覚フィルタに関する概念は、MPEG アルゴリズムには存在せず本評価尺度において新たに導入したものである。

b. 各マスカー成分に対するマスキング領域の計算

各々のマスカー成分に対するマスキング領域は、Egan らにより測定された狭帯域刺激に対するマスキングパターン [89] に基づいて算出する。但し、本評価尺度では計算を簡略化するため、マスカーの音圧レベルが 60 dB の場合の測定結果を直線近似したものを常時利用する。具体的には、マスカーの周波数を k [Hz]、その音圧レベルを $X(k)$ [dB] とし、次の 3 点 A、B、C:

$$\begin{cases} A: & (k,\ X(k) - 18) \\ B: & (k - 2 \cdot \mathrm{ERB}(k),\ X(k) - 48) \\ C: & (k + 4.5 \cdot \mathrm{ERB}(k),\ X(k) - 48) \end{cases} \tag{A.4}$$

を計算し、図 A.3 の灰色部分に示すようなマスキング領域を定める (図 A.1 の点線)。

A 1.3. 継時マスキング特性の実装

継時マスキングに関しては、あるフレームにおいて検出された同時マスキングにおけるマスカー成分が、それ以降のフレームへ影響を及ぼすと考える (図 A.1 の破線)。つまり、順向性マスキングの特性を実装する。継時マスキングには、順向性マスキングの他に逆向

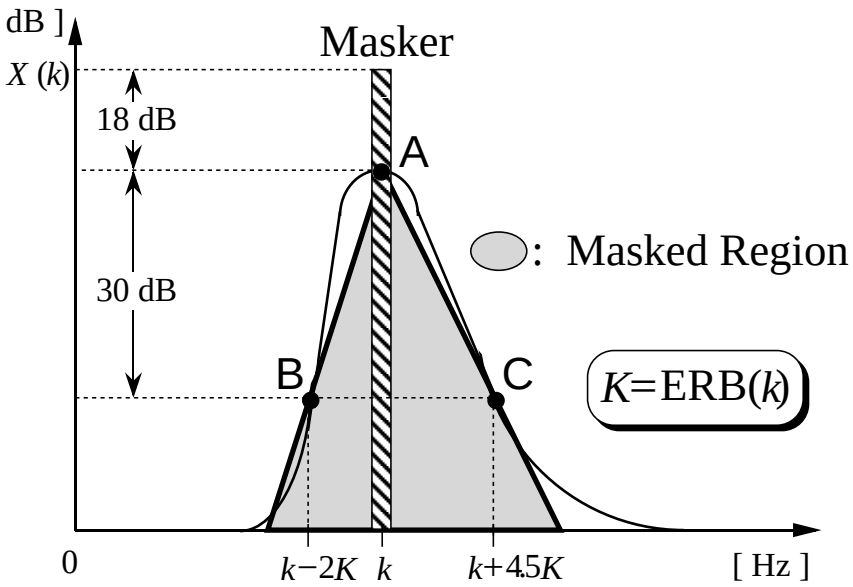

図 A.3: 各マスカーに対するマスキング領域

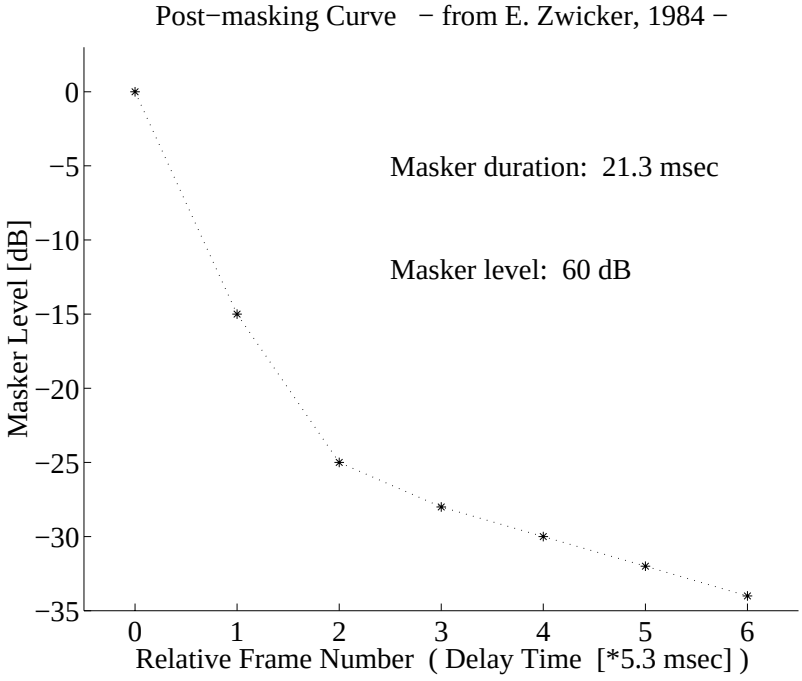

図 A.4: マスカーの継時レベル減衰量

性マスキングがあるが、逆向性マスキングは順向性マスキングと比較するとその影響は非常に小さく、定量的に測定することが困難である [80]。従って、本評価尺度では逆向性マスキングの実装は行なわない。

マスカーの継時レベル減衰量は、Zwicker の実験結果 [90] を参考に、図 A.4 に示す値に設定した。図 A.4 の横軸は、マスカーが存在するフレーム以降の相対的なフレーム数 (時間軸)、縦軸は各フレームにおけるマスカーのレベル、図中の＊印はマスカー長をフレーム長とした場合のマスカーの継時レベル減衰量を表す。本評価尺度においては、継時マスキングの影響は 6 フレーム後 (約 30 msec) まで及ぼすものとする。

A 1.4. 相対可聴閾値の計算

各フレームにおける相対可聴閾値 $X_{total}(\omega)$ (図 A.1 の太実線) は、同時マスキング、継時マスキングのそれぞれにおいて求めたマスキング閾値を統合して得られる。本評価尺度では、マスキング閾値の統合に Lutfi が提案したベキ乗則変換 [91] を採用する。ベキ乗則変換モデルは、心理物理実験により得られた数多くのマスキング現象に関する知見を説明づけるために提案され、一パラメータのみにより様々な条件におけるマスキング現象を表現可能なモデルである。相対可聴閾値 $X_{total}(\omega)$ は、同時マスキングにおける周波数 ω_i の単一マスカー成分に対するマスキング閾値 $X_{\omega_i}(\omega)$、継時マスキングにおける過去の第 t フレームで計算されたマスキング閾値を $X_t(\omega)$ とすると、

$$X_{total}(\omega) = \mathrm{inv}F\left[\sum_{\omega_i} F\left[X_{\omega_i}(\omega)\right] + \sum_{t=1}^{6} F\left[X_t(\omega)\right]\right],\ F(z) = z^p, \qquad (A.5)$$

により求められる。ここで、p は定数である。

Lutfi は、同時マスキングのみを考える場合、マスカーの数が 4 つしか存在しない状況においては $p = 0.33$ が心理物理実験の結果と最も整合することを報告している [91]。また、より多くのマスカーが存在する場合、あるいは継時マスキングを考慮する場合には、$p = 0.33$ が最適値ではないことも報告している [92] [93]。本評価尺度は、対象とする信号は数多くのマスカーが存在する音声信号であり、増すキング現象として同時マスキングと継

時マスキングとの両者を考慮しているため、$p = 0.33$ が最適であるとは考えられない。そこで、本評価尺度による客観評価値と主観評価値との間の相関が最も高くなるように、パラメータ p の最適化を行なった。これに関しては、付録 A 2 にて検討する。

A 1.5. 聴覚特性を考慮した客観的歪量の算出

相対可聴閾値以下のスペクトル成分に関しては、例え歪みが生じていても、我々には知覚できないと考えられる。そこで、評価の各フレームごとに、相対可聴閾値を越える成分のみに対して

$$\mathrm{ASD} = \sqrt{\underset{i}{\mathrm{MEAN}}\{S_{target}(i) - S_{clean}(i)\}^2}\ [\ \mathrm{dB}\ ], \tag{A.6}$$

により歪み量を計算する。ここで、$S_{target}(\omega)$、$S_{clean}(\omega)$ は、評価対象音声、クリーンな音声の振幅スペクトルとする。周波数 i は、評価対象周波数帯域を 100 Hz から 6kHz の音声帯域とし、その中で評価対象音声の振幅スペクトル $X_{target}(\omega)$ が、マスキング閾値を越える周波数とする。

評価対象音声の客観的な歪み量である ASD 値は、各フレームごとに式 (A.6) で計算される歪み量の全音声区間での平均値と定義する。

A 2. 聴感上の印象を考慮した評価尺度 ASD の妥当性検証

実験の目的：

客観的評価尺度の評価基準は、評価尺度による客観評価値と、主観評価値との間に線形的関係が保たれているほど望ましいとする。もしそれらの間に線形的な関係が成り立てば、客観評価値より主観的な歪みの程度を容易に予測することが可能になるためである。そこで、SD あるいは ASD の各評価尺度により得られる客観評価値と、聴取実験により得られる主観評価値としての 5 段階の平均オピニオン値 Mean Opinion Score (MOS) との関係を調査することにより、SD に対する ASD の優位性を確認する。また、ASD の一パラメータである式 (A.5) のベキ乗則パラメータ p の最適化も行なう。

実験条件・実験手法：

主観評価実験により、クリーンな音声に雑音を付加した音声の 5 段階 MOS を求める。クリーンな音声は、ATR 音声データベース [54] に収録されている重要語データセット中の男性話者 mht 氏発声の連続母音 /ao/ を用いた。雑音は、2kHz–3kHz のランダム帯域雑音とし、それを計算機上でクリーンな音声/ao/ のわたり部に付加する。まず、任意の大きさの雑音を付加した音声を作成し、予備実験にて被験者に主観評価させた結果、被験者全員がこの音声を MOS が 0(歪んでいる) と評価した。この音声の SNR は -2.3 dB であり、これを雑音最大の音声、つまり最も歪んだ音声とする。クリーンな音声と雑音最大の音声、そして雑音最大音声の SD および ASD(パラメータ p は 0.60 とする) による評価結果を図 A.5 に示す。雑音最大音声に対する SD 値は 16.7 dB、ASD($p = 0.60$) 値は 21.5 dB であった。主観評価実験に用いる信号は、雑音最大音声に対する SD 値 16.7 dB、ASD($p = 0.60$) 値 21.5 dB のそれぞれ 0.25 倍、0.5 倍、0.75 倍となるように雑音の音圧レベルを調整した雑音付加音声とする。具体的には、図 A.6 に示すように、SD 値が 4.2 dB、8.4 dB、12.6 dB、ASD 値が 5.4 dB、10.8 dB、16.1 dB である歪み音声を作成する。

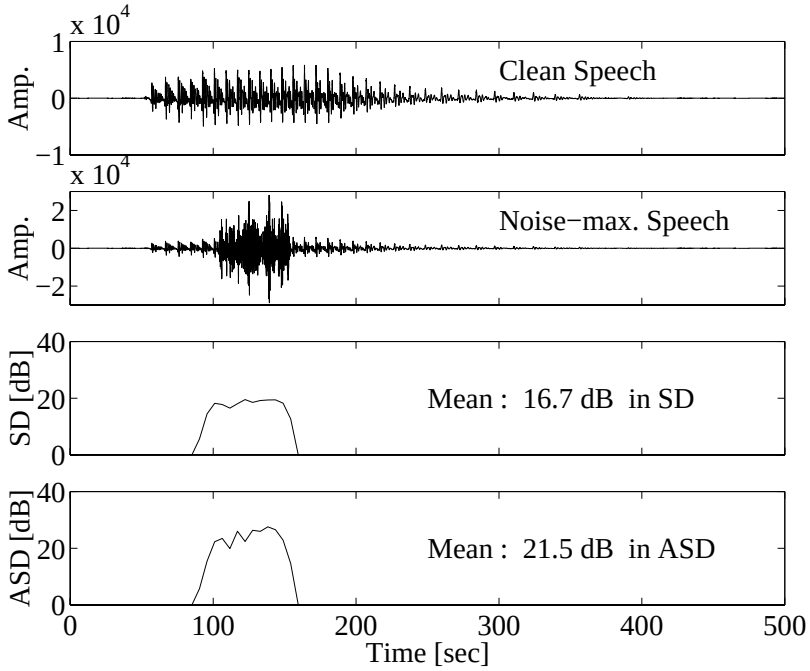

図 A.5: クリーン音声、雑音最大音声、SD 値および ASD($p = 0.60$) 値

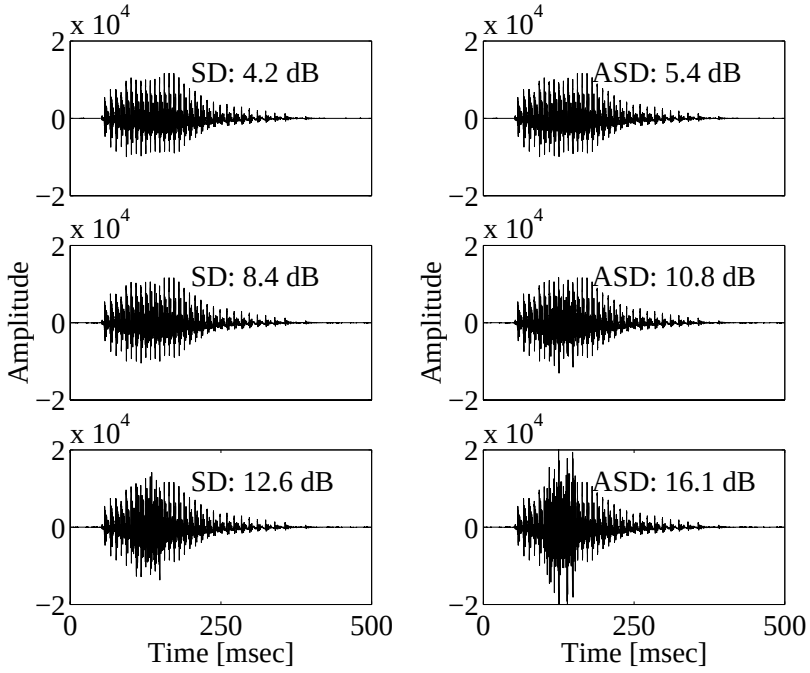

図 A.6: 主観評価実験に用いる雑音付加音声

表 A.1: 主観評価実験に使用した機器一覧

機器	メーカ	型番
DSP	T. C. T.	DAT-Link+
DAC	STAX	DAC-TALENT BD
安定化電源	SINANO	HSP-510
ヘッドホン	STAX	Lambda Nova Signature
ヘッドホンアンプ	STAX	SRM-1/MK-2 P.P.
スピーカ	AURATONE	5PSC
パワーアンプ	SANSUI	AU- α 907 MR
騒音計	B&K	2231
マイクロホン	B&K	4134
人工耳	B&K	4153

被験者は、これまでに何らかの聴取実験において被験者としての経験を持ち、正常聴力を有する大学院生 8 名である。実験は、表 A.1 に示す機器を用い、防音室内でヘッドホン受聴あるいはスピーカ受聴により行なった。被験者には、クリーンな音声、雑音最大音声、評価対象音声の順、あるいは雑音最大音声、クリーンな音声、評価対象音声の順に呈示し、3 番目の音声の主観的な歪み量 (雑音の大きさ) を、最初の 2 つの音声を参考に、4〜0 で 5 段階評価させた。それぞれの刺激は、ランダムに合計 6 回評価させるが、最初の 2 回の評価結果はトレーニングセットとして破棄する。つまり、各刺激に対して、各被験者が 4 回ずつ評価を行なうことになる。被験者への呈示音圧レベルは、クリーンな音声で約 66 dB(A)、雑音レベルが最大の音声で約 75 dB(A) であった。

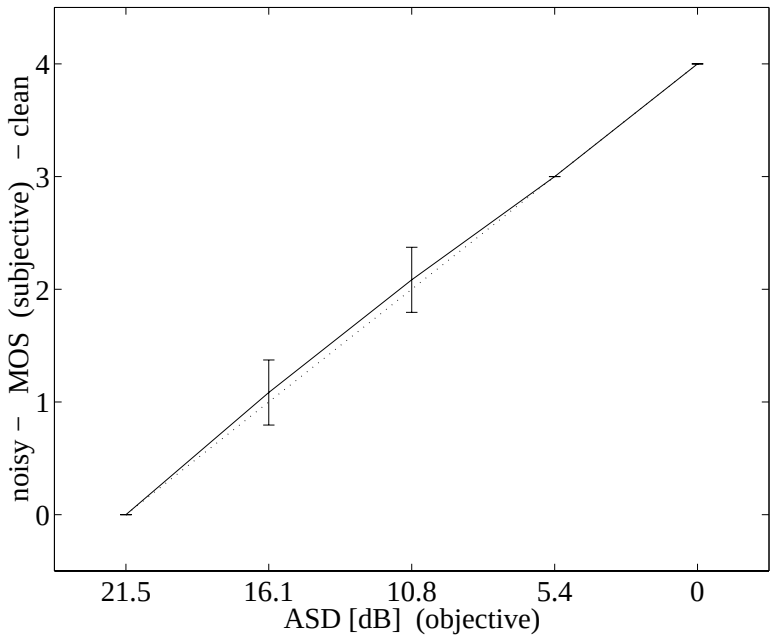

図 A.7: ASD($p = 0.60$) に対する主観-客観対応曲線

実験結果：

主観評価実験により求めた MOS と、客観評価値 ASD 値との関係を図 A.7 に、SD 値との関係を図 A.8 に示す。図 A.7 および図 A.8 は、横軸が ASD($p = 0.60$) および SD による評価値、縦軸は MOS を表す。図中の実線は、被験者全員の MOS の平均値を表し、以後これを主観-客観対応曲線と呼ぶ。図中の点線は、主観評価値と客観評価値とが線形的な関係にある場合の理想直線、エラーバーは MOS の標準偏差を表す。

次に、式 (A.5) におけるパラメータ p の最適化を行なうため、p を 0.3 から 0.1 刻みで 0.9 まで変化させ、客観評価値 ASD 値と主観評価値 MOS との関係を調査する。p の各値に対して、図 A.7 と同様に ASD に対する主観-客観対応曲線を描き、それらと理想直線との相関係数を計算する。図 A.9 は、横軸にパラメータ p の値、縦軸に相関係数をとり、パラメータ p と相関係数との関係を示す。なお、SD に対する主観-客観対応曲線と、理想直線との相関係数は 0.951 であった。参考のため、相関係数が SD の場合とほぼ同じ値となった、ASD($p = 0.30$) に対する主観-客観対応曲線を図 A.10 に示す。

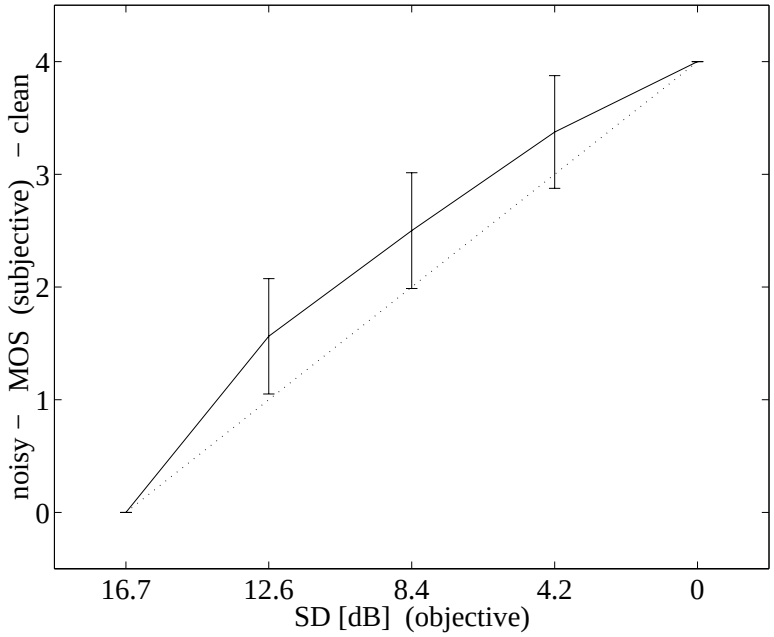

図 A.8: SD に対する主観-客観対応曲線

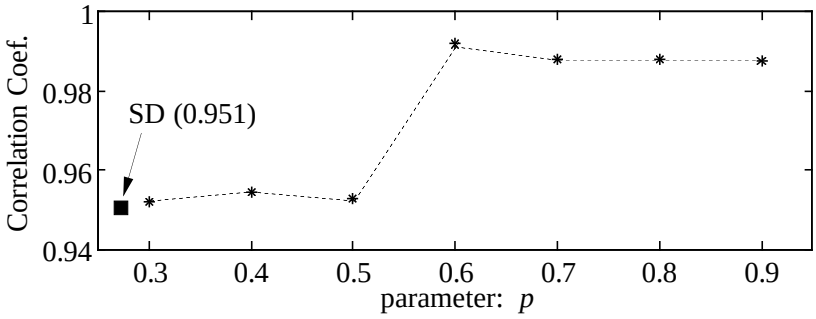

図 A.9: ベキ乗則変換のパラメータ p と相関係数との関係

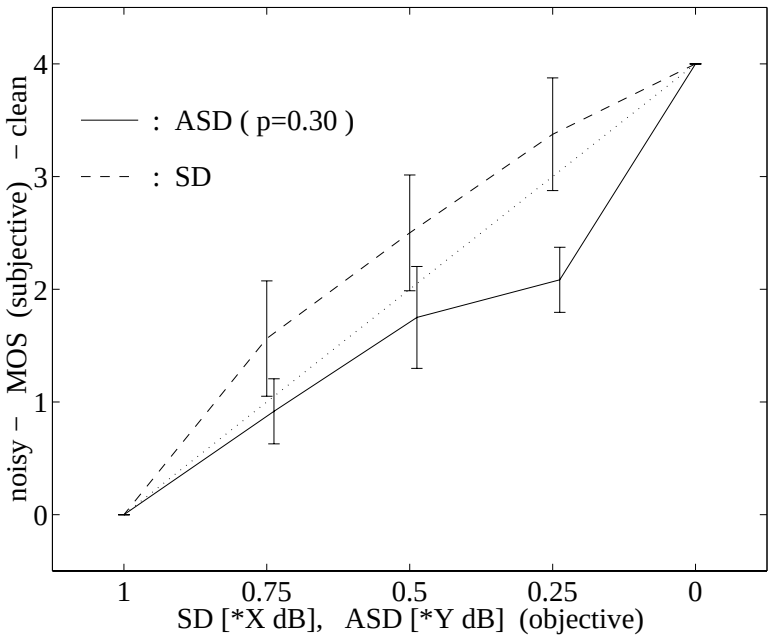

図 A.10: ASD($p = 0.30$) に対する主観-客観対応曲線 (破線は SD の主観-客観対応曲線)

考察：

評価尺度 ASD のベキ乗則パラメータ p の最適値は、図 A.9 より 0.60 であることがわかる。ASD($p = 0.60$) に対する主観-客観対応曲線 (図 A.7) は、SD に対する主観-客観対応曲線 (図 A.8) と比較し、理想直線に非常に近く、主観評価値 MOS と客観評価値 ASD との間に線形的な関係が成り立っている。つまり、客観的評価尺度 ASD は、間隔尺度の条件を満たしており、SD よりも聴感上の歪み印象と対応がよいことがわかる。

本実験に関しては、被験者間のバラツキ、またヘッドホン呈示とスピーカ呈示との呈示条件による相違は明確には現れなかった。実験結果のバラツキの程度を客観的に評価するため、実験結果に対して F-検定 [94] を行なった結果、有意水準 5% では有意な差は見られなかった。これより、母分散は一定と仮定して t-検定 [94] を行なった結果、SD に対する主観-客観対応曲線は、理想直線に対して有意水準 1% で有意差が見られたが、ASD($p = 0.60$) の場合には有意差は見られなかった。従って、ASD($p = 0.60$) は、聴感上の歪み感の定量的評価に適した客観的評価尺度と言える。

謝辞

本研究を行なうに当たり、終始御指導を賜わった北陸先端科学技術大学院大学 情報科学研究科教授 赤木正人 博士に深謝致します。

本論文をまとめるにあたり、草稿の段階から貴重な御助言と御指導を賜わりました北陸先端科学技術大学院大学 情報科学研究科教授 嵯峨山茂樹 博士、情報科学研究科助教授 小谷一孔 博士、情報科学研究科助教授 下平博 博士、奈良先端科学技術大学院大学 情報科学研究科助教授 中村哲 博士に心より感謝致します。

また、日頃から有益な御助言をいただき、御指導いただいた北陸先端科学技術大学院大学 情報科学研究科教授 宮原誠 博士をはじめとする諸先生方、北村達也 博士をはじめとする赤木研究室の諸先輩方に厚く御礼申し上げます。

本研究を行なうにあたり、多面に渡り御協力いただいた赤木研究室の皆様に感謝します。

本研究の一部は、科学技術振興財団による戦略的基礎研究推進事業 (CREST)、並びに文部省科学研究費補助金 (No. 10680374) による援助を受けて行なわれたものであります。ここに感謝の意を表します。

最後に、大学院在学中に自由な研究の場を与えていただき、暖かく見守ってくれた両親、祖母、弟に心から感謝致します。

参考文献

[1] J. C. Junqua and J. P. Haton, *Robustness in automatic speech recognition*, Kluwer Academic Publishers, Boston (1996).

[2] 飛田 瑞広, 管村 昇, 中津 良平" 単語音声認識における残響の影響," 音講論 (春), Vol. I, pp. 13-14 (1990).

[3] 飛田 瑞広, " 異なるマイク特性による騒音下での音声認識性能," 音講論 (秋), Vol. I, pp. 55-56 (1991).

[4] 飛田 瑞広, 管村 昇" 音声認識における周囲環境の影響," 音響誌, Vol. 51, No. 4, pp. 331-335 (1995).

[5] A. J. Duquesnoy and R. Plomp, "Effect of reverberation and noise on the intelligibility of sentences in case of presbycusis," J. Acoust. Soc. Am., Vol. 68, pp. 537-544 (1980).

[6] 水島 昌英, 伊藤憲三, "自動利得制御と雑音抑圧処理が難聴者の音声知覚に及ぼす影響," 信学技報, SP 96-35, pp. 17-24 (1996).

[7] 小畑 秀文, "雑音下での音声処理技術," 音講論 (秋), Vol. I, pp. 233-236 (1998).

[8] A. Papoulis, "Probability, random variables, and stochastic process (Third edition)," McGraw-Hill, New York (1995).

[9] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. ASSP, Vol. 26, pp. 197-210 (1978).

[10] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," Proc. ICASSP'87, Vol. 1, pp. 177-180 (1987).

[11] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. ASSP, Vol. 27, No. 2, pp. 113-120 (1979).

[12] 谷口 賢一，津村 尚志，福留 公利, “スペクトルサブトラクション法における雑音推定方式,” 音講論 (秋), Vol. I, pp. 175-176 (1994).

[13] 二村 善則, 松本 弘, “スペクトルサブトラクション法の改良と非定常雑音による評価,” 音講論 (春), Vol. I, pp. 5-6 (1994).

[14] 國枝 伸行, “雑音レベルの変動を考慮したスペクトルサブトラクション法,” 音講論 (秋), Vol. I, pp. 245-246 (1998).

[15] H. Gustafsson, S. Nordholm, I. Claesson, “Spectral Subtraction with Adaptive Averaging of the Gain Function,” EUROSPEECH’99, Vol. 6, pp. 2599-1602 (1999).

[16] 金 学胤，浅野 太，鈴木 陽一，曽根 敏男, ” 短時間振幅スペクトル推定を用いた 2 チャネル音声強調法における振幅スペクトル推定について,” 音講論 (秋), Vol. I, pp. 533-534 (1995).

[17] 中島 栄俊, 宇佐川 毅, 江端 正直, “スペクトルサブトラクション法とケプストラム処理による音声強調,” 音講論 (秋), Vol. I, pp. 499-500 (1999).

[18] 國枝 伸行, 島村 徹也, 鈴木 誠史, “LPC 分析のための雑音除去の検討 – 相関領域でのスペクトルサブトラクション–,” 音講論 (秋), Vol. I, pp. 259-260 (1996).

[19] 西村 竜一, 浅野 太, 鈴木 陽一, 曽根 敏夫, “Wavelet 変換を用いた Spectral Subtraction による音声強調,” 信学論, Vol. J79-A, No. 12, pp. 1986-1993 (1996).

[20] T. Irino, R. D. Patterson, “A time-domain, level-dependent auditory filter: the gammachirp,” J. Acoust. Soc. Am., Vol. 101, No. 1, pp. 412-419 (1997).

[21] T. Irino, “Noise suppression using a time-varying, analysis/synthesis gammachirp filterbank, ” Proc. ICASSP’99, Vol. I, pp. 97-100 (1999).

[22] M. S. Brandstein, “On the use of explicit modeling in microphone array applications,” Proc. ICASSP’98, Vol. VI, pp. 3613-3616 (1998).

[23] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West and M. M. Sondhi, “Autodirective microphone systems,” ACUSTICA, Vol. 73, pp. 58-71 (1991).

[24] 金森 丈郎, 茨木 悟, 古川 博基, 直野 博之, 齋藤 浩, 西川 清, "2 次元ディジタルフィルタを用いた超指向性マイクロホン," 信学技報, Vol. EA91-84 (1992).

[25] O. L. Flost, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE, Vol. 60, No. 8, pp. 926-935 (1972).

[26] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. AP, Vol. 30, No. 1, pp. 27-34 (1982).

[27] M. M. Sondhi and G. W. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint," Proc. ICASSP'86, Vol. 2, pp. 981-984 (1986).

[28] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Trans. ASSP, Vol. 34, No. 6, pp. 1391-1400 (1986).

[29] 浅野 太, 速水 悟, 山田 武志, 中村 哲, " サブスペース法を用いた音声強調法とその音声認識への応用," 信学技報, Vol. EA97-17, pp. 17-24 (1997).

[30] 浅野 太, 速水 悟, 山田 武志, 中村 哲, " サブスペース法を用いた音声強調法とその音声認識への応用（第 2 報）," 信学技報, Vol. EA98-51, pp. 37-44 (1998).

[31] 猿渡 洋, 古村 光夫, "非線形サイドローブキャンセラを用いた合成開口空中ソーナ," 信学論 (A), Vol. J81-A, No. 5, pp. 815-826 (1998).

[32] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on complementary beamforming," IEICE Trans. Fundamentals, Vol. E82-A, No. 8, pp. 1501-1510 (1999).

[33] Y. Nagata and H. Tsuboi, "A two-channel adaptive microphone array with target tracking," Proc. EUROSPEECH'97, Vol. 1, pp. 343-346 (1997).

[34] 永田 仁史，安倍 正人, "話者追尾 2 チャンネルマイクロホンアレーに関する検討," 信学論, Vol. J82-A, No. 6, pp. 860-866 (1999).

[35] T. Yamada, S. Nakamura and K. Shikano, "An effect of adaptive beamforming on hands-free speech recognition based on 3-D Viterbi search," Proc. ICSLP'98, pp. 381-284 (1998).

[36] S. Nakamura, T. Yamada, P. Heracleous and K. Shikano, “Recognition of distant talking speech based on 3-D trellis search using a microphone array and adaptive beamforming,” Proc. IEEE Workshop on Robust Methods for Speech Recognition in Adverse Conditions, pp. 219-222 (1999).

[37] 大賀 寿郎，山崎 芳男，金田 豊, “音響システムとディジタル処理,” 電子情報通信学会 (1995).

[38] 中村 哲, “音声認識系へのマイクロホンアレーの応用,” 音講論 (春), Vol. I, pp. 515-518 (1998).

[39] 鹿野 清宏, 中村 哲，伊勢 史郎, “音声・音情報のディジタル信号処理,” 昭晃堂 (1997).

[40] M. J. Gales and S. Young, ”An improved approach to the Hidden Markov Models decomposition of speech and noise,” Proc. ICASSP’92, pp. 233-236 (1992).

[41] F. Martin, K. Shikano, and Y. Okabe, ”Recognition of noisy speech by composition of Hidden Markov Models,” 信学技報, Vol. SP92-96, pp. 9-16 (1992).

[42] 浅野 太, 鈴木 陽一, 曽根 敏夫, “適応マイクロホンアレイを用いた補聴器のための雑音抑圧手法に関する一考察,” 信学技報, Vol. EA90-68, pp. 45-52 (1990).

[43] F. Asano, S. Hayamizu, Y. Suzuki, S. Tsukui and T. Sone, “Array signal processing applicable to hearing aids,” Proc. ASA and ASJ third joint meeting, pp. 261-266 (1996).

[44] J. E. Greenberg and P. M. Zurek, “Evaluation of an adaptive beamforming method for hearing aids,” J. Acoust. Soc. Am., Vol. 91, No. 3, pp. 1662-1676 (1992).

[45] J. G. Desloge, W. M. Rabinowitz and P. M. Zurek, “Microphone-array hearing aids with binaural output – Part. I: Fixed-processing systems,” IEEE Trans. Speech and Audio Processing, Vol. 5, No. 6, pp. 529-542 (1997).

[46] D,. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, “Microphone-array hearing aids with binaural output – Part. II: A two-microphone adaptive system,” IEEE Trans. Speech and Audio Processing, Vol. 5, No. 6, pp. 543-551 (1997).

[47] 柏野 牧夫, “雑音と音声の知覚,” 音講論 (秋), Vol. I, pp. 229-232 (1998).

[48] A. S. Bregman, *Auditory scene analysis*, MIT Press, Massachusetts (1989).

[49] G. J. Brown and M. Cooke, "Computational auditory scene analysis," Computer Speech and Language, Vol. 8, pp. 297-336 (1994).

[50] M. Unoki and M. Akagi, "A method of signal extraction from noisy signal based on auditory scene analysis," Speech Communication, Vol. 27, No. 3, pp. 261-279 (1999).

[51] J. F. Culling and Q. Summerfield, "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am., Vol. 98, No. 2, pp. 785-797 (1995).

[52] S. A. Gelfand, *Hearing (Third edition)*, Marcel Dekker, Inc., New York (1998).

[53] N. I. Durlach, "Equalization and cancellation theory of binaural masking level differencec," J. Acoust. Soc. Am., Vol. 35, pp. 1206-1218 (1963).

[54] 武田 一哉, 匂坂 芳典, 片桐 滋, 阿部 匡伸, 桑原 尚夫, "研究用日本語音声データベース利用解説書," ATR Tech. Rep. TR-I-0028 (1988).

[55] J. Huang, N. Ohbishi, and n. Sugie, "A biomimetic system for localization and separation of multi sound sources," IEEE Trans. Instrumentation and Measurement, Vol. 44, No. 3, pp. 733-738 (1995).

[56] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. ASSP, Vol. 24, No. 4, pp. 320-327 (1976).

[57] I. Yamada and N. Hayashi, "Improvement of the performance of cross correlation method for identifying aircraft noise with pre-whitening of signals," J. Acoust. Soc. Jpn., (E), Vol. 13, No. 4, pp. 241-252 (1992).

[58] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," Proc. ICASSP'88, Vol. 5, pp. 2578-2581 (1988).

[59] 古井 貞熙, "ディジタル音声処理," 東海大学出版会 (1985).

[60] 中田 和男, "音声," コロナ社 (1977).

[61] S. Young, J. Odell, D. Ollason, V. Valtchev and P. Woodland, *The HTK BOOK*, Cambridge University (1995).

[62] B. S. Atal, “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification,” J. Acoust. Soc. Am., Vol. 55, pp. 1304-1312 (1974).

[63] 難波 精一郎, 桑野 園子, “音の評価のための心理学的測定法,” コロナ社 (1998).

[64] The MathWorks, Inc., *Using MATLAB* (1996).

[65] 井上 雅晶，山田 武志，中村 哲，鹿野 清宏, “種々のマイクロホンアレーによる音声認識の検討,” 信学技報, SP 96-89, pp. 13-20 (1997).

[66] M. Inoue, S. Nakamura, T. Yamada, K. Shikano, “Microphone array design measures for hands-free speech recognition,” Proc. EUROSPEECH’97, Vol. 1, pp. 13-20 (1997).

[67] N. Campbell, “The ATR British English speech database,” ATR Tech. Rep. TR-I-0363 (1993).

[68] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall, New Jersey (1978).

[69] ITUT Recommendation P. 861, *Objective quality measurement of telephone-band(300-3400 Hz) speech codecs* (1996).

[70] M. Kahrs and K. Brandenburg, *Applications of digital signal processing to audio and acoustics*, Kluwer Academic Publishers, Massachusetts (1998).

[71] S. Voran, “A simplified version of the ITU algorithm for objective measurement od speech codec quality,” Proc. ICASSP’98, Vol. I, pp. 537-540 (1998).

[72] 川口 恵子, 清水 美都子, 山田 武志, 北脇 信彦, 浅野 太, “符合誤り音声に対する客観品質評価尺度の検討,” 音講論 (秋), Vol. I, pp. 369-370 (1995).

[73] 小坂 直敏, 筧 一彦, “基本的支配要因を対象とした通話品質客観評価モデル,” 信学論, Vol. J68-A, No. 1, pp. 70-77 (1985).

[74] 小坂 直敏, 筧 一彦, “通話品質のオピニオン評価に影響を与える心理的要因の検討,” 信学論, Vol. J69-A, No. 5, pp. 652-662 (1986).

[75] M. Szarvas, T. Fegyó, P. Tatai, G. Gordos, "Qualiphone-A: A perceptual speech quality evaluation system for analog mobile networks," EUROSPEECH'99, Vol. 6, pp. 2563-2566 (1999).

[76] 赤木 正人, "聴覚フィルタとそのモデル," 信学誌, Vol. J77-A, No. 9, pp. 948-956 (1994).

[77] W. Yang, M. Benbouchta, and R. Yantorno, "Performance of the modified Bark spectral distortion as an objective speech quality measure," Proc. ICASSP'98, Vol. I, pp. 541-544 (1998).

[78] R. Meddis, "Simulation of mechanical to neural transduction in the auditory recepter," J. Acoust. Soc. Am., Vol. 79, No. 3, pp. 702-711 (1986).

[79] M. Hauenstein, "Application of Meddis' inner hair-cell model to the prediction of subjective speech-quality," Proc. ICASSP'98, Vol. I, pp. 545-548 (1998).

[80] E. Zwicker, *Psychoakustik*, Springer-Verlag, Berlins (1982) [Japanese edition: Y. Yamada, 西村書店 (1992)].

[81] 伊藤 憲三, 北脇 信彦, 筧 一彦, "音声のディジタル波形符合化方式の客観的品質評価尺度の検討," 信学論, Vol. J66-A, No. 3, pp. 274-281 (1983).

[82] S. Aoshima, "Computer-generated pulse signal applied for sound measurement," J. Acoust. Soc. Am., Vol. 69, No. 5, pp. 1484-1488 (1981).

[83] 鈴木 陽一, 浅野 太, 金 学胤, 曽根 敏夫, "時間引き延ばしパルスの設計法に関する考察," 信学技報, Vol. EA 92-86, pp. 17-24 (1992).

[84] 永田 穂, 飯田 一嘉, 古宇田 潔, 橘 秀樹, 古川 宣一, 安岡 正人, 山本 照二, "建築音響," コロナ社 (1998).

[85] 境 久雄, 中山 剛, "聴覚と音響心理," コロナ社 (1978).

[86] B. R. Glasberg and B. J. C. Moore, "Derivation of auditory filter shapes from notched-noise data," Hearing Research, Vol. 47, pp. 103-138 (1990).

[87] R. D. Patterson, "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am., Vol. 82, No. 5, pp. 1560-1586 (1987).

[88] ISO/IEC 11172, *Coding of moving pictures and associated audio for degital strage media at up to about 1.5 Mbit/s* (1993).

[89] J. P. Egan and H. W. Hake, "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am., Vol. 22, pp. 622-630 (1950).

[90] E. Zwicker, "Dependence of post-masking on masker duration and its relation to temporal effects in loudness," J. Acoust. Soc. Am., Vol. 75, No. 1, pp. 219-223 (1984).

[91] R. A. Lutfi, "Additivity of simultaneous masking," J. Acoust. Soc. Am., Vol. 73, No. 1, pp. 262-267 (1983).

[92] R. A. Lutfi, "A power-law transformation predicting masking by sounds with complex spectra," J. Acoust. Soc. Am., Vol. 77, No. 6, pp. 2128-2136 (1985).

[93] L. E. Humes and W. Jesteadt, "Models of the additivity of masking," J. Acoust. Soc. Am., Vol. 77, No. 3, pp. 1285-1294 (1989).

[94] 押川 元重, 阪口 紘治, "基礎 統計学," 培風館 (1989).

本研究に関する発表論文

論文

[1] 水町 光徳, 赤木 正人," マイクロホン対を用いたスペクトルサブトラクションによる雑音除去法," 信学論 (A) vol. J82-A, No. 4, pp. 503-512 (1999).

[2] M. Mizumachi and M. Akagi,"The auditory-oriented spectral distortion for evaluating speech signals distorted by additive noises," J. Acoust. Soc. Jpn. (E) (条件付採録)

国際会議

[1] M. Mizumachi and M. Akagi,"Noise reduction by paired-microphones using spectral subtraction," Proc. ICASSP'98, Seattle, USA, Vol. II, pp. 1001-1004 (1998).

[2] M. Mizumachi and M. Akagi,"Noise reduction method that is equipped for a robust direction finder in adverse environments," Proc. IEEE Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, FINLAND, pp. 179-182 (1999).

[3] M. Mizumachi and M. Akagi,"An objective distortion estimator for hearing aids and its application to noise reduction," Proc. EUROSPEECH'99, Budapest, HUNGARY, Vol. 6, pp. 2619-2622 (1999).

口頭発表

[1] 水町 光徳, 赤木 正人," マイクロホン対を用いた残響除去法（DERPAM）," 音講論(春), Vol. I, pp. 621-622 (1997).

[2] 水町 光徳, 赤木 正人," 複数マイクロホンを用いたスペクトルサブトラクションによる雑音除去法," 信学技報, SP 97-35, pp. 9-16 (1997).

[3] 水町 光徳, 赤木 正人," 複数マイクロホンを用いたスペクトルサブトラクションによる雑音除去法," 音講論 (秋), Vol. I, pp. 541-542 (1997).

[4] 水町 光徳, 赤木 正人," 雑音除去を前提とした音声及び雑音の到来方向推定法," 音講論 (春), Vol. I, pp. 521-522 (1998).

[5] 水町 光徳, 赤木 正人," 聴覚特性を考慮した低品質音声に対する歪み評価尺度の提案," 音講論 (秋), Vol. I, pp. 369-370 (1998).

[6] 水町 光徳, 赤木 正人," マスキング特性を考慮した低品質音声に対する歪み評価尺度の提案," 音講論 (春), Vol. I, pp. 401-402 (1999).

[7] 水町 光徳, 赤木 正人," スペクトルサブトラクションの最適化に関する検討," 音講論(秋), Vol. I, pp. 373-374 (1999).

[8] 水町 光徳, 赤木 正人," 加法性雑音に対する客観的歪み音声評価尺度," 聴覚研究会資料, H-2000-4 (2000).

[9] 水町 光徳, 赤木 正人," 小規模マイクロホンアレーによる音声認識の雑音耐性向上," 音講論 (春), 1-8-5 (2000).

第一著者以外の発表論文

[1] M. Akagi and M. Mizumachi,"Noise Reduction by Paired Microphones," Proc. EUROSPEECH'97, Rhodes, GREECE, Vol. 1, pp. 335-338 (1997).

[2] 赤木 正人，水町 光徳," マイクロホン対を用いた雑音除去法（NORPAM）," 音講論(春), Vol. I, pp. 619-620 (1997).

[3] 赤木 正人，水町 光徳," マイクロホン対を用いた雑音除去法（NORPAM）," 信学技報, SP 97-34, pp. 1-8 (1997).