

Title	小規模マイクロホンアレーを用いた音声了解度の改善に関する研究
Author(s)	帯辺, 誠之
Citation	
Issue Date	2004-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/1786">http://hdl.handle.net/10119/1786</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修 士 論 文

小規模マイクロホンアレーを用いた  
音声了解度の改善に関する研究

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

帯 辺 誠 之

2004 年 3 月

## 修士論文

# 小規模マイクロホンアレーを用いた 音声了解度の改善に関する研究

指導教官 赤木 正人 教授

審査委員主査 赤木 正人 教授  
審査委員 党 建武 教授  
審査委員 下平 博 助教授

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

110074 帯辺 誠之

提出年月: 2004年2月

## 概要

雑音除去技術の需要はデジタル音響処理技術の発展に伴って益々高まっている。我々を取り巻く雑音環境は数多く存在し、屋内の人混みでの騒音レベルは約 74 dB、電車内や地下鉄構内では約 82 dB にも達する。環境省の報告では、1 m の距離で会話した際に 100 % 明瞭な会話了解度を達成するためには通常、屋内で 45 dB 以下、屋外で 55 dB 以下、会話聴取について特に配慮を要する者（高齢者等）については、これより低いことが望ましいとしている。

このような雑音環境下における了解度改善方法の一つとして雑音を除去することにより目的の音声を強調する手法が数多く提案されている。単一マイクロホンを用いた雑音除去法として、Boll が提案したスペクトルサブトラクションなどが挙げられる。この手法は処理が簡便で実用に適しているが、雑音の定常性を仮定しているために実環境下では利用出来ないことが多い。これに対して、近年、盛んに研究されているマイクロホンアレーを用いた雑音除去法が数多く提案されている。この方法では信号の到来方向という空間情報が利用可能になるため、指向特性を形成することが出来る。しかしながら、遅延和型アレーを用いた受音系において、鋭い指向特性を形成するためには膨大な素子数を必要とする。そのため、日常生活で携帯するような機器としては非常に不向きである。このような問題に対して水町や籠らは減算型アレーを用いた小規模な受音系での雑音除去法を提案した。この手法は時々刻々推定した雑音スペクトルを受音信号のスペクトルから減算することにより、突発的な非定常雑音に対して頑健であった。

本研究では小規模マイクロホンアレーを用いた雑音除去による音声了解度の向上を目的とする。受音系には規模が大きく膨大な素子数を必要とする遅延和マイクロホンアレーを用いず、規模が小さく膨大な素子数を必要としない減算型マイクロホンアレーを用いる。雑音除去法として複数雑音に対して頑健な籠らの提案法を用いる。しかしながら、籠らの提案法は自動音声認識器に適用するためのもので、人間に音を聞かせた場合の考慮がなされていない。そこで本研究では籠らの提案法における問題点である雑音スペクトルの推定精度に注目し、音声了解度を向上させるための改善法について検討を行う。

# 目次

第1章	序論	1
1.1	研究の背景	1
1.2	研究の目的	2
1.3	論文の構成	2
第2章	問題設定と前提条件	4
2.1	問題設定	4
第3章	雑音除去法	5
3.1	マイクロホンアレーを用いた雑音除去法	5
3.1.1	遅延和型と減算型アレー	5
3.1.2	水町らの提案法	7
3.2	籠らの雑音除去法とその定式化	7
3.2.1	周波数帯域分割型雑音除去法	7
3.2.2	雑音の到来方向の推定	8
3.2.3	雑音スペクトルの推定	10
3.2.4	雑音スペクトルの除去	11
第4章	籠らの提案法の問題点とその改善法	13
4.1	ビームフォーマの指向特性	13
4.2	雑音スペクトル推定部における到来方向と周波数帯域毎の分母項	14
4.3	問題点のまとめ	21
4.4	問題点の改善法	21
4.4.1	雑音スペクトル推定部の改善	22
4.4.2	周波数帯域毎の閾値設定による雑音スペクトルの推定精度改善	23
4.5	問題点の改善にあたって目指したこと	25
第5章	計算機シミュレーションを用いた検証	26
5.1	雑音除去アルゴリズムのパラメータ設定	26
5.2	複数雑音に対する雑音除去実験	27
5.2.1	シミュレーション条件	27

5.2.2	雑音除去結果	27
5.3	周波数帯域が重複している雑音に対する雑音除去実験	31
5.3.1	シミュレーション条件	31
5.3.2	雑音除去結果	31
5.4	非定常な突発雑音に対する雑音除去実験	31
5.4.1	シミュレーション条件	31
5.4.2	雑音除去結果	35
5.5	複数の音声が存在する場合の雑音除去実験	35
5.5.1	シミュレーション条件	35
5.5.2	雑音除去結果	35
<b>第6章</b>	<b>単語了解度試験</b>	<b>43</b>
6.1	周波数帯域が重複している雑音に対する単語了解度試験	43
6.2	非定常雑音に対する単語了解度試験	45
6.3	複数の音声が存在する場合に対する単語了解度試験	45
6.4	考察	46
<b>第7章</b>	<b>結論</b>	<b>47</b>
7.1	本論文の要約	47
7.2	今後の課題	47

# 第1章 序論

## 1.1 研究の背景

雑音除去技術の需要はデジタル音響処理技術の発展に伴って益々高まっている。我々を取り巻く雑音環境は数多く存在し、屋内の人混みでの騒音レベルは約 74 dB、電車内や地下鉄構内では約 82 dB にも達する。環境省の報告では、1 m の距離で会話した際に 100 % 明瞭な会話了解度を達成するためには通常、屋内で 45 dB 以下、屋外で 55 dB 以下、会話聴取について特に配慮を要する者（高齢者等）については、これより低いことが望ましいとしている。

では、実際に電車内や地下鉄構内で話し声の聞き取りが全く出来ないかという点を決してそうとは限らない。我々は話者に対して注意を向け、雑音と話し声を聞き分けることが出来る。しかしながら、このような能力の低下している高齢者や難聴者にとって雑音環境下で話し声だけを聞き分けることは容易ではない。また、実用化されている補聴器の多くは雑音環境下で目的とする音声だけを増幅することが出来ず、同時に受信した雑音も同様に増幅するという問題がある。そのため、補聴器の利用者にとって雑音環境下における目的音の聞き分けは非常に困難である。

このような問題を解決する方法の一つとして、雑音を除去することで目的音を強調するという手法が挙げられる。単一マイクロホンを用いた過去の雑音除去研究では Boll が提案したスペクトルサブトラクション (SS) 等が挙げられる。この手法は処理が簡便で実用に適しているが、雑音の定常性を前提としているため上で挙げたような実環境下では利用出来ないことが多い。他にも単一マイクロホンを用いた雑音除去法は数多く提案されている。しかしながら、単一マイクロホンを用いた受信系では鋭い指向特性を形成出来ない上に信号処理も複雑になるという問題があった。

これに対して、近年ではマイクロホンアレー (microphone array : 複数のマイクロホン素子から構成される受信器) を用いた雑音除去法が数多く提案されている。この方法では信号の到来方向という空間情報を利用することが出来るため、指向特性の形成が可能であるため、上で挙げたような雑音環境下での雑音除去による音声の音質・了解度改善が期待出来る。しかしながら、遅延和型アレーを用いた受信系は鋭い指向特性を形成するためには膨大な素子数が必要になるという問題があった。

自動音声認識器に適用される雑音除去法として、水町らは 3 チャンネルの減算型マイクロホンアレーを用いた雑音除去法を提案した。この手法は時々刻々推定した雑音スペクトルを受音信号のスペクトルから減算することにより、突発的な非定常雑音に対して頑健で

あった。しかしながら、目的音と雑音がそれぞれ一つであるという仮定のもとで提案された手法であるために同時に到来する複数雑音には対応していない。これに対して籠らは周波数帯域分割を行い、水町らの提案法を各帯域で繰り返し適用することにより複数雑音に頑健な雑音除去法を提案した。しかしながら、これらの手法は自動音声認識器を対象にした雑音除去であり人間を対象にしたものではない。また、水町らや籠らが用いたスペクトルサブトラクションを用いた雑音除去法では雑音スペクトルの推定誤差のために音声信号のスペクトルに歪みを生じる。雑音スペクトルの推定誤差の原因として、ビームフォーミングの指向特性の形状や、雑音スペクトル推定部における到来方向の誤差等が挙げられるが、水町らや籠らの手法はこの点についての考慮が為されていない。そのために、特に 1 kHz 以下の低い周波数帯域では雑音スペクトルの取り残しや取りすぎが起こる。このような音声信号を人間が受聴した際には雑音感を感じる。

## 1.2 研究の目的

本研究では小規模マイクロホンアレーを用いた雑音除去による音声了解度の向上を目的とする。マイクロホンアレーは膨大な素子数を必要とする遅延和型マイクロホンアレーを使わず減算型マイクロホンアレーを用いて雑音除去を行う。雑音除去アルゴリズムは籠らの提案した雑音除去法を用いる。籠らの提案法において雑音スペクトルの推定誤差が大きくなる周波数帯域・到来方向・ビームフォーミングの組み合わせを調べ、周波数帯域毎に閾値を設けることで雑音スペクトルの推定精度を改善する。また、推定した到来方向と実際に用いる到来方向との間の誤差を小さくする。これによって音声了解度を改善し、補聴器等に応用可能な小規模マイクロホンアレーを用いた雑音除去を実現する。

## 1.3 論文の構成

本論文は全 7 章により構成されている。以下に各章の概要を述べる。

### 第 1 章：序論

補聴器を例にして雑音除去法の問題点を挙げ、過去に提案された改善法について述べる。またマイクロホンアレーを用いた雑音除去法の問題点を明確にし、本研究の目的を示す。

### 第 2 章：問題設定と前提条件

本研究における問題設定と、それにより決定される前提条件について示す。

### 第 3 章：雑音除去法

まず、典型的なマイクロホンアレーを用いた雑音除去法である遅延和型アレーと減算型

アレーについて述べる。次に水町らの提案法の概要を説明し、それを更に発展させた籠らの提案法について説明する。また、籠らの提案法の定式化を行い、雑音の到来方向推定、雑音のスペクトル推定、雑音スペクトルの除去法について説明する。

#### 第4章：籠らの提案法の問題点

第3章での籠らの提案法にどのような問題点があるかを挙げ、改善法について詳細に説明する。本研究における問題設定で籠らの提案法を利用した際のビームフォーマの指向特性の形状、空間分解能、雑音スペクトルの推定精度が低下する周波数帯域を調べ、改善法である周波数帯域毎の閾値設定について説明する。

#### 第5章：計算機シミュレーションを用いた検証

第3章と第4章で記述した雑音除去法を用いて、基礎的な雑音除去性能の検証を行う。複数雑音が存在する場合、周波数帯域が重複した雑音が存在する場合、複数の音声が存在する場合について計算機シミュレーションを行い、結果の考察を行う。

#### 第6章：単語了解度実験

単語了解度実験を行い、籠らの提案法との比較を通して本研究の有効性について検証する。

#### 第7章：結論

本研究で得られた結果を要約し、今後検討すべき課題について述べる。

## 第2章 問題設定と前提条件

### 2.1 問題設定

研究を進める上で扱う問題について考える。人間が会話をする際には大抵、会話する相手と向き合うと考えられる。音源に注意を向けるときは目で音源を探すこともある。そのため目的信号の到来方向を正中面と仮定する。雑音除去法を補聴器などに応用する際にはマイクロホンアレーを眼鏡に装着して利用することが可能である。そのため、一般的な眼鏡の大きさ（約 11 ~ 13 cm 前後）を全長としたマイクロホンアレーとし、素子間隔を 5.5 ~ 6.5 cm 程度とする。本論文では全長 13 cm、素子間隔を 6.5 cm とする。また、携行性を考慮すると規模の大きい AD 変換器を利用することが出来ないためにサンプリング周波数もそれほど高くは出来ない。そのため、前提条件として、

- ・ 目的音は正中面から到来する。
- ・ マイクロホンアレーの全長は 13 cm、素子間隔は 6.5 cm とする。
- ・ サンプリング周波数は 16 kHz とする。

が挙げられる。

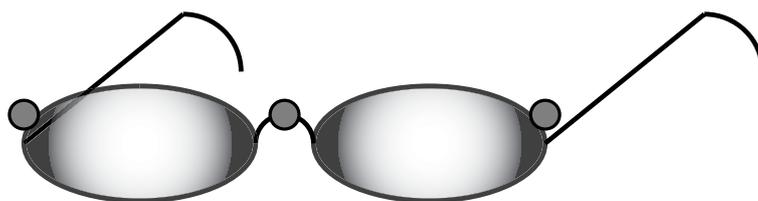


図 2.1: 補聴器眼鏡のイメージ図

## 第3章 雑音除去法

### 3.1 マイクロホンアレーを用いた雑音除去法

本節では、マイクロホンアレーを用いた従来の雑音除去法の概要を説明し、その利点と欠点を明らかにする。序論でも述べた通り、マイクロホンアレーは遅延和型アレーと減算型アレーの大きく二種類に分けられる。

- ・ 遅延和型アレー  
目的信号の到来方向に鋭いビームを形成する。
- ・ 減算型アレー  
雑音の到来方向にビームの死角を形成する。

#### 3.1.1 遅延和型と減算型アレー

遅延和型アレーは図 3.1.1 に示すような形で構成される。各マイクロホンの受信信号に遅延を与え同相化した後、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 3.1.1 では、 $\theta$  方向から目的信号  $s(t)$  が 3 チャンネル遅延和型アレーに到来する場合を考える。音波が平面波で伝搬すると仮定したとき、音速を  $c$ 、隣接マイクロホン間隔を  $d$  とすると、受信信号に与えるべき遅延量  $\tau_1$  は

$$\tau_1 = \frac{d \sin \theta}{c} \quad (3.1)$$

となり、目的信号  $s(t)$  は

$$\hat{s}(t) = \sum_{i=1}^3 w_i x_i(t - (i-1)\tau_1) \quad (3.2)$$

と推定される。ここで

$$\sum_{i=1}^3 w_i = 1 \quad (3.3)$$

より、 $w_i = 1/3$  とする。以上の手順により、目的信号の到来方向に鋭い指向特性を形成する。また、目的信号の到来方向と異なった方向から到来する信号は同相化されないで、信号が相殺されて大きくなるらない。

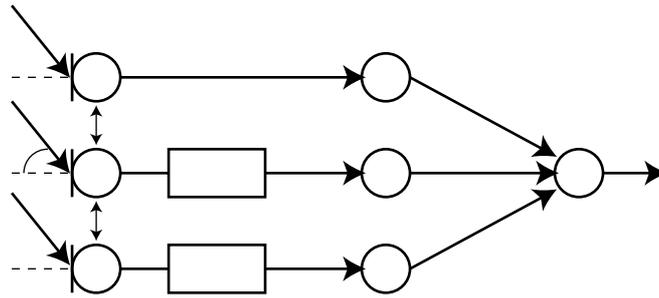


図 3.1: 3チャンネル遅延和アレーのブロックダイアグラム

遅延和型アレーは、雑音の到来方向に無関係にビームフォーミングを行うため、雑音に制約を受けることなく、突発的な非定常雑音や残響のようなあらゆる方向から到来する雑音にも有効である。しかしながら、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、小規模なマイクロホンアレーほど特に低周波数帯域における雑音除去能力が低い。よって素子数やサイズに制約を受けるマイクロホンアレーへの応用は困難である。

減算型アレーは、各マイクロホンの受信信号に遅延を与え同相化した後、減算することにより雑音を除去する。この処理は、周囲から到来する雑音の方向を検出し、その方向に指向特性の死角を形成することにより行われる。周囲の雑音環境に応じた指向特性を自動的に形成することから、このような機能を持つアレーは適応型アレーとも呼ばれている。

以上の操作において、遅延量  $\tau_N$  (雑音の到来方向  $\theta_N$ ) を知っている必要はない。つまり、減算出力のパワー  $P'$  を監視しながら遅延量を変化させ、 $P'$  が最小となったときの遅延量を  $\tau_N$  とすることによって自動的に決定される。遅延を付加することはデジタルフィルタを用いて実現することが出来、雑音成分のパワーが最小となるようにフィルタの係数が制御される(これを適応フィルタと呼ぶ)。また、目的信号が雑音とは異なった方向から到来するとすると、このような遅延操作では目的信号は同相化されないため、減算によって消去されることはない。

しかしながら、適応型アレーは急激な音響的变化に追従することが困難であり、実環境下に多く存在する突発的な非定常雑音に対する除去能力が低い。また、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ると言われており、適応フィルタの係数が不適切に設定されると雑音が除去されないばかりか目的信号に新たな歪みを生じさせてしまう。加えて適応型アレーが形成することが出来る死角数は一般に最大(素子数 - 1)であるため、素子数以上の複数雑音は同時に除去することが出来ない。よって、素子数の少ない減算型アレーを実環境下で用いた場合、非定常雑音及び複数雑音の除去能力低下が大きな問題となる。

### 3.1.2 水町らの提案法

これまで従来の小規模マイクロホンアレーを用いた手法では、実環境下における雑音に対応することが出来ないことを示してきた。この問題に対して水町らは、時々刻々推定した雑音成分を受音信号から減算する、突発的な非定常雑音に対して頑健な3チャンネル解析型アレーによる雑音除去法を提案した。また、籠らは周波数帯域分割を行い、各帯域毎に雑音の到来方向推定と雑音のスペクトル推定を行うことで水町らの提案法を更に発展させた。

水町らの提案法の大まかな手順を以下に示す。

1. 目的信号と最も優勢な雑音の到来方向を推定する。
2. 目的信号を完全に抑圧し、雑音のスペクトルを推定する。
3. スペクトルサブトラクションにより、受音信号のスペクトルから雑音のスペクトルを除去する。

信号の到来方向と雑音スペクトルの推定を行う際に二つのマイクロホンの受音信号を利用し、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形成する。つまり、通常の減算型アレーでは雑音の到来方向に死角を形成するのに対して、水町らの提案法では目的信号の到来方向に死角を形成し、目的音以外の信号を雑音と見なすことで小規模なマイクロホンアレーで精度の高い雑音スペクトルの推定が可能になっている。

しかしながら、水町らの提案法では目的音源と雑音源がそれぞれ一つであるという仮定を設けている。そのために複数の雑音源や残響には対応することが出来ない。

## 3.2 籠らの雑音除去法とその定式化

本節では籠らの提案した雑音除去法の定式化を行う。水町らの手法をどのように周波数帯域分割したのかを説明し、理論的に複数雑音及び突発的な非定常雑音に対応可能であることを示す。

### 3.2.1 周波数帯域分割型雑音除去法

減算型アレーで重要になるのが指向特性の死角と焦点（信号を必ず抽出する方向）である。雑音の到来方向と焦点が一致していない場合、雑音スペクトルを十分に推定することが出来ない。水町らの提案法では全帯域で一度しか雑音の到来方向推定を行わないために推定出来る焦点も一方向のみである。そのため、ビームフォーマの指向特性の死角方向から雑音が到来した場合に雑音スペクトルを十分に推定することが出来ないという問題があった。

そこで籠らは周波数帯域を分割し各帯域毎に雑音の到来方向の推定を行い、帯域毎の雑音スペクトルを推定した。これによって、各周波数帯域で最も優勢な雑音スペクトルの推定が可能になった。上記の改良により、雑音の周波数がそれぞれ異なっていれば、複数雑音を全帯域の一つの雑音スペクトルとして推定することが出来た。

ここで、籠らの提案法の大まかな手順を示す。

1. 白色化相互相関関数を用いて雑音の到来時間差を推定する。
2. 目的信号を完全に抑圧し、雑音のスペクトルを推定する。
3. 1, 2 の処理を周波数帯域毎に行い、全帯域の雑音スペクトルを推定する。
4. スペクトルサブトラクションにより、受信信号のスペクトルから雑音のスペクトルを除去する。

### 3.2.2 雑音の到来方向の推定

本節では、短時間フレーム毎に雑音の到来方向を推定するアルゴリズムについて記述する。籠らの提案法では、受信信号の相互相関を用いて方向推定を行っている。更に方向推定にビームフォーマを利用することで雑音環境下での信号の到来方向推定の精度向上を図っている。

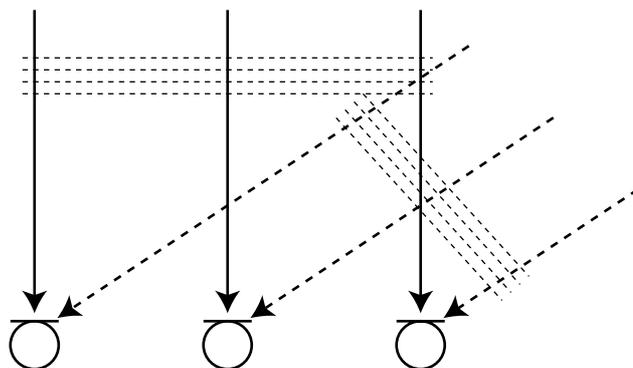


図 3.2: 3チャンネル等間隔直線配列マイクロホンアレー

籠らの提案法で使用されたマイクロホンアレーは図 3.2.2 に示されるような三本の無指向性マイクロホンを等間隔に直線配列したものである。信号が平面波としてマイクロホンアレーに到来すると仮定すると、信号の到来方向推定は二つのマイクロホン間での信号の到来時間差を推定することに相当する。3チャンネルマイクロホンアレーにおいて、左側、

中央、右側のマイクロホンでの受信信号をそれぞれ  $l(t)$ ,  $c(t)$ ,  $r(t)$  とする。目的信号  $s(t)$  以外に雑音  $n_m(t)$  が  $M$  個あるとすると、各受信信号は次式のように表される。

$$\begin{aligned} l(t) &= s(t - \zeta) + \sum_{m=1}^M n_m(t - \delta_m) \\ c(t) &= s(t) + \sum_{m=1}^M n_m(t) \\ l(t) &= s(t + \zeta) + \sum_{m=1}^M n_m(t + \delta_m) \end{aligned} \quad (3.4)$$

ここで  $\zeta$  (本論文では目的信号の到来方向を正中面と仮定しているため、以後  $\zeta = 0$  とする) は目的信号  $s(t)$ 、 $\delta_m$  は雑音  $n_m(t)$  の隣接マイクロホン間における到来時間差である。また、 $l(t)$ ,  $c(t)$ ,  $r(t)$ ,  $s(t)$ ,  $n_m(t)$  の Fourier 変換をそれぞれ  $L(\omega)$ ,  $C(\omega)$ ,  $R(\omega)$ ,  $S(\omega)$ ,  $N(\omega)$  とすれば、上式は以下ようになる。

$$\begin{aligned} L(\omega) &= S(\omega) + \sum_{m=1}^M N_m(\omega) e^{-j\omega\delta_m} \\ C(\omega) &= S(\omega) + \sum_{m=1}^M N_m(\omega) \\ R(\omega) &= S(\omega) + \sum_{m=1}^M N_m(\omega) e^{j\omega\delta_m} \end{aligned} \quad (3.5)$$

受信信号  $l(t)$  と  $r(t)$ 、 $l(t)$  と  $c(t)$ 、 $c(t)$  と  $r(t)$  を組み合わせて目的信号を完全に抑圧する信号、つまり到来時間差が  $\zeta = 0$  である方向に指向特性の死角を形成するような減算型ビームフォーマ  $g_{lr}(t)$ ,  $g_{lc}(t)$ ,  $g_{cr}(t)$  を考える。

$$\begin{aligned} g_{lr}(t) &= \frac{\{l(t + \tau) - l(t - \tau)\} - \{r(t + \tau) - r(t - \tau)\}}{4} \\ g_{lc}(t) &= \frac{\{l(t + \tau) - l(t - \tau)\} - \{c(t + \tau) - c(t - \tau)\}}{4} \\ g_{cr}(t) &= \frac{\{c(t + \tau) - c(t - \tau)\} - \{r(t + \tau) - r(t - \tau)\}}{4} \end{aligned} \quad (3.6)$$

ここで  $\tau$  は 0 以外の任意の定数であり、焦点を制御することが出来る。また、 $g_{lr}(t)$ ,  $g_{lc}(t)$ ,  $g_{cr}(t)$  の Fourier 変換をそれぞれ  $G_{lr}(\omega)$ ,  $G_{lc}(\omega)$ ,  $G_{cr}(\omega)$  とすると上式の Fourier 変換は以下のようになる。

$$\begin{aligned} G_{lr}(\omega) &= \sum_{m=1}^M N_m(\omega) \sin \omega \delta_m \sin \omega \tau \\ G_{lc}(\omega) &= \sum_{m=1}^M N_m e^{-j\omega \frac{\delta_m}{2}}(\omega) \sin \omega \frac{-\delta_m}{2} \sin \omega \tau \\ G_{cr}(\omega) &= \sum_{m=1}^M N_m e^{j\omega \frac{\delta_m}{2}}(\omega) \sin \omega \frac{\delta_m}{2} \sin \omega \tau \end{aligned} \quad (3.7)$$

これらの式に目的信号  $S(\omega)$  に関する項は一切存在せず、目的信号を完全に抑圧していることがわかる。

十分に狭い周波数帯域  $\omega_{k-1} < |\tilde{\omega}| \leq \omega_k$ , ( $k = 1 \sim K$ ) において到来時間差が  $\delta_k$  となる雑音  $N_k(\tilde{\omega})$  を想定する。帯域を十分に狭くした場合、信号を純音として扱うことが出来、純音の和は純音となるので

$$\sum_{m=1}^M N_m(\tilde{\omega}) \sin \tilde{\omega} \delta_m \sin \tilde{\omega} \tau = N_k(\tilde{\omega}) \sin \tilde{\omega} \delta_k \sin \tilde{\omega} \tau \quad (3.8)$$

と書くことが出来る。よって  $G_{lr}(\omega), G_{lc}(\omega), G_{cr}(\omega)$  を周波数帯域の和として表せば

$$\begin{aligned} G_{lr}(\tilde{\omega}) &= \sum_{k=1}^K N_k(\tilde{\omega}) \sin \tilde{\omega} \delta_k \sin \tilde{\omega} \tau \\ G_{lc}(\tilde{\omega}) &= \sum_{k=1}^K N_k e^{-j\tilde{\omega} \frac{\delta_k}{2}}(\tilde{\omega}) \sin \tilde{\omega} \frac{-\delta_k}{2} \sin \tilde{\omega} \tau \\ G_{cr}(\tilde{\omega}) &= \sum_{k=1}^K N_k e^{j\tilde{\omega} \frac{\delta_k}{2}}(\tilde{\omega}) \sin \tilde{\omega} \frac{\delta_k}{2} \sin \tilde{\omega} \tau \end{aligned} \quad (3.9)$$

となる。 $G_{lc}(\tilde{\omega}), G_{cr}(\tilde{\omega})$  より

$$F_{lc}(\tilde{\omega}) = \begin{cases} G_{lc}(\tilde{\omega}) / \sin \tilde{\omega} \tau, & \omega_{k-1} < |\tilde{\omega}| \leq \omega_k \\ 0, & \text{otherwise} \end{cases} \quad (3.10)$$

$$F_{cr}(\tilde{\omega}) = \begin{cases} G_{cr}(\tilde{\omega}) / \sin \tilde{\omega} \tau, & \omega_{k-1} < |\tilde{\omega}| \leq \omega_k \\ 0, & \text{otherwise} \end{cases} \quad (3.11)$$

を計算し、白色化相互相関関数  $\phi_{F_{lc}F_{cr}}(t)$  を用いて到来時間差  $\delta_k$  を推定する。

$$\begin{aligned} \phi_{F_{lc}F_{cr}}(t) &= \text{IFFT} \left[ \frac{F_{lc}(\tilde{\omega}) F_{cr}^*(\tilde{\omega})}{|F_{lc}(\tilde{\omega})| |F_{cr}(\tilde{\omega})|} \right] \\ \delta_k &= \arg \max_t [\phi_{F_{lc}F_{cr}}(t)] \end{aligned} \quad (3.12)$$

このように、クロススペクトルをパワースペクトルで正規化することで、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。

以上の操作により、ある周波数帯域で最も優勢な雑音の到来時間差を目的信号を完全に抑圧した信号から推定することが出来る。

### 3.2.3 雑音スペクトルの推定

前節で推定した到来時間差  $\delta_k$  を利用して、目的信号を完全に抑圧した信号より雑音のスペクトル  $N_k(\tilde{\omega})$ , ( $\omega_{k-1} < |\tilde{\omega}| \leq \omega_k$ ) の推定を行う。

目的信号を完全に抑圧した減算型ビームフォーマ  $G_{lr}(\tilde{\omega}), G_{cr}(\tilde{\omega})$  において焦点を雑音の到来方向に合わせる。マイクロホンアレーの焦点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することが可能である。つまり  $G_{lr}(\tilde{\omega})$  において  $\tau = \delta_k$ 、 $G_{cr}(\tilde{\omega})$  において  $\tau = \frac{\delta_k}{2}$  と代入すれば、周波数帯域  $\omega_{k-1} < |\tilde{\omega}| \leq \omega_k$  において

$$\begin{aligned} G_{lr}(\tilde{\omega}) &= N_k(\tilde{\omega}) \sin^2 \omega \delta_k \\ G_{cr}(\tilde{\omega}) &= N_k(\tilde{\omega}) e^{j\tilde{\omega} \frac{\delta_k}{2}} \sin^2 \tilde{\omega} \frac{\delta_k}{2} \end{aligned} \quad (3.13)$$

となる。しかしながら、 $G_{lr}(\tilde{\omega})$  においては  $\tilde{\omega} = \frac{n\pi}{\delta_k}$ 、( $n$ : 整数)、 $G_{cr}(\tilde{\omega})$  においても  $n$  が偶数となる場合にはビームフォーミングの空間的折り返し現象が生じることで指向特性の死角を形成し、正確に雑音スペクトルを推定することが出来ない。よって、二つの閾値  $\varepsilon_1, \varepsilon_2$  を設けて、次式により雑音スペクトル  $\hat{N}_k(\tilde{\omega})$  を推定する。

$$\hat{N}_k(\tilde{\omega}) = \begin{cases} G_{lr}(\tilde{\omega}) / \sin^2 \tilde{\omega} \delta_k, & \sin^2 \tilde{\omega} \delta_k > \varepsilon_1 \\ G_{cr}(\tilde{\omega}) e^{-j\tilde{\omega} \frac{\delta_k}{2}} / \sin^2 \tilde{\omega} \frac{\delta_k}{2}, & (\sin^2 \tilde{\omega} \delta_k \leq \varepsilon_1) \wedge (\sin^2 \tilde{\omega} \frac{\delta_k}{2} > \varepsilon_2) \\ G_{lr}(\tilde{\omega}) & \sin^2 \tilde{\omega} \frac{\delta_k}{2} \leq \varepsilon_2 \end{cases} \quad (3.14)$$

つまり、分母項である  $\sin^2 \tilde{\omega} \delta_k$  の値が十分大きい場合は、マイクロホン間隔がより大きい  $G_{lr}(\tilde{\omega})$  を用いて雑音スペクトルを推定する。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。分母項が 0 に近づくに従って計算誤差が大きくなるため、そのような場合は  $G_{cr}(\tilde{\omega})$  を用いて雑音スペクトルを推定する。 $G_{cr}(\tilde{\omega})$  を用いても雑音スペクトルを推定することが出来ない場合に限り、雑音スペクトルを近似する。

以上の雑音の到来方向推定及びスペクトル推定を、各周波数帯域で繰り返し行い  $\hat{N}_1(\tilde{\omega}) \sim \hat{N}_K(\tilde{\omega})$ 、( $\omega_{k-1} < |\tilde{\omega}| \leq \omega_k$ ,  $k = 1 \sim K$ ) を求める。最後にそれらを統合することにより、全帯域の雑音スペクトル  $\hat{N}(\tilde{\omega})$  を求める。

### 3.2.4 雑音スペクトルの除去

籠らの提案した雑音除去法では、周波数領域における減算であるスペクトルサブトラクション (SS) によって雑音の除去を行う。

雑音成分の除去は  $m$  マイクロホンアレーの中央のマイクロホンで受信した信号  $c(t)$  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することが出来る。このとき、目的信号の振幅スペクトルの推定値  $|\hat{S}(\omega)|$  を、Boll の SS を改良した非線形 SS

$$|\hat{S}(\omega)| = \begin{cases} |C(\omega)| - \alpha |\hat{N}(\omega)|, & |C(\omega)| \geq \alpha |\hat{N}(\omega)| \\ \beta |C(\omega)|, & \text{otherwise} \end{cases} \quad (3.15)$$

$$\hat{s}(t) = \text{IFFT} \left[ |\hat{S}(\omega)| e^{j \arg C(\omega)} \right] \quad (3.16)$$

により推定する。ここで  $\alpha, \beta$  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のために、受信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ある。その際、目的信号の振幅スペクトルの推定値  $|\hat{S}(\omega)|$  が負の値にならないように、SS に非線形機構が導入されている。

## 第4章 籠らの提案法の問題点とその改善法

本章ではマイクロホンアレーの指向特性、空間分解能と雑音除去法における雑音スペクトルの推定精度を調べ、その観点から籠らの提案法の問題点とその解決法について述べる。

籠らの提案法により雑音除去した音声には1 kHz以下の周波数帯域と5～6 kHz付近に雑音が残る。1 kHz以下の周波数帯域に雑音が残ることで母音の了解度が下がるため、母音に関する聞き間違いが生じる。また、5～6 kHz付近に残る雑音は主に子音の了解度を下げる。

### 4.1 ビームフォーマの指向特性

本節では、減算型アレーのビームフォーマの指向特性について調べる。減算型ビームフォーマは3.2.2節で詳しく説明している。ここでは計算機シミュレーションを用いて各周波数の指向特性について考察を行う。

図4.1から図4.6に素子間隔6.5 cm、全長13 cmのマイクロホンアレーにおける周波数帯域毎のビームフォーマ  $g_{lr}(t)$ ,  $g_{cr}(t)$  の指向特性(死角0°、焦点41°の場合)を示す。図4.3の2 kHzの特性を見ると0°方向に指向特性の谷を形成していることがわかる。また、特性も非常に急峻で全長わずか13 cmのマイクロホンアレーで±20°程度まで死角を絞り込んでいることがわかる。しかしながら、図4.5を見ると、死角の方向以外に指向特性の谷が出来ていることがわかる。この現象は空間的折り返し現象と呼ばれるものである。マイクロホンの素子間隔を  $d$ 、音速を  $c$ 、入力信号の周波数を  $f$  とすると

$$d < \frac{c}{2f} \quad (4.1)$$

を満たすならば、ビームフォーマの指向特性の谷が死角方向以外に生じることはない。つまり、素子間隔が到来する信号の半波長より短ければ谷の折り返しは生じない。マイクロホンアレーの素子間隔を大きくすると、低周波数帯域におけるビームフォーマの指向特性は急峻になるが、周波数が高くなるにつれてこのような空間的折り返し現象が生じることになる。

このような谷が目的音が到来する死角方向以外にあり、谷を形成している方向から雑音が出来ると、雑音のスペクトルを正確に推定することが出来ないために雑音の除去が行われないという問題が起こる。このような問題を回避するために、籠らはビームフォーマ

$g_{lr}(t)$  が谷を形成する場合にはもう一つのビームフォーマ  $g_{cr}(t)$  を用いて雑音スペクトルを推定している。しかしながら、 $\tilde{\omega} = \frac{n\pi}{\delta k}$ , ( $n$ : 偶数) の場合には  $g_{lr}(t), g_{cr}(t)$  共に指向特性の谷を形成するために、この方向から到来する雑音のスペクトルは推定することが出来ない。また、図 4.1 を見ると、ビームフォーマの指向特性の死角が非常に広いことがわかる。このことから、正中面以外から到来する目的信号以外の雑音も抑圧し、1 kHz 以下の低い周波数では殆ど雑音スペクトルを推定することが出来なくなると考えられる。そのため、籠らの提案法では 1 kHz 以下に存在する雑音は除去されることなく、信号に残ることになる。

## 4.2 雑音スペクトル推定部における到来方向と周波数帯域毎の分母項

籠らの提案した雑音除去法で最も重要な要素は雑音スペクトルの推定部である。また、雑音の到来方向推定により求めた到来方向は、雑音スペクトルの推定精度に大きく影響する。まず、推定した到来時間差と各周波数における雑音スペクトルの分母項との関係について調べる。

まず、マイクロホンアレーの空間分解能について説明する。 $d$  だけ離れた二つのマイクロホンにおける受信信号をそれぞれ  $x_1(t), x_2(t)$  とする。このとき、二つのマイクロホンに到来する信号の時間差を  $\delta$ 、音速を  $c$  とすると、

$$x_2(t) = x_1(t - \delta) \quad (4.2)$$

$$\delta = \frac{d \sin \theta}{c} \quad (4.3)$$

が成立する。従って、信号が二つのマイクロホンに到来する時間差を推定することが出来れば

$$\theta = \arcsin\left(\frac{\delta c}{d}\right) \quad (4.4)$$

の式から信号の到来方向  $\theta$  を求めることが出来る。ここで、 $\delta = \frac{\text{Time difference}}{\text{Sampling frequency}}$  とすることでサンプリング周波数と空間分解能の関係を得ることが出来る。表 4.1 に素子間隔 6.5 cm、48 kHz と 16 kHz サンプリングの場合の空間分解能を示す。上記の表からもわか

表 4.1: サンプリング周波数と空間分解能の関係 (素子間隔 6.5 cm)

到来方向 [°]		6	13	19	26	33	41	50	61	79	90
到来時間差 $\delta$	48 kHz	1	2	3	4	5	6	7	8	9	10
	[point]			1			2			3	

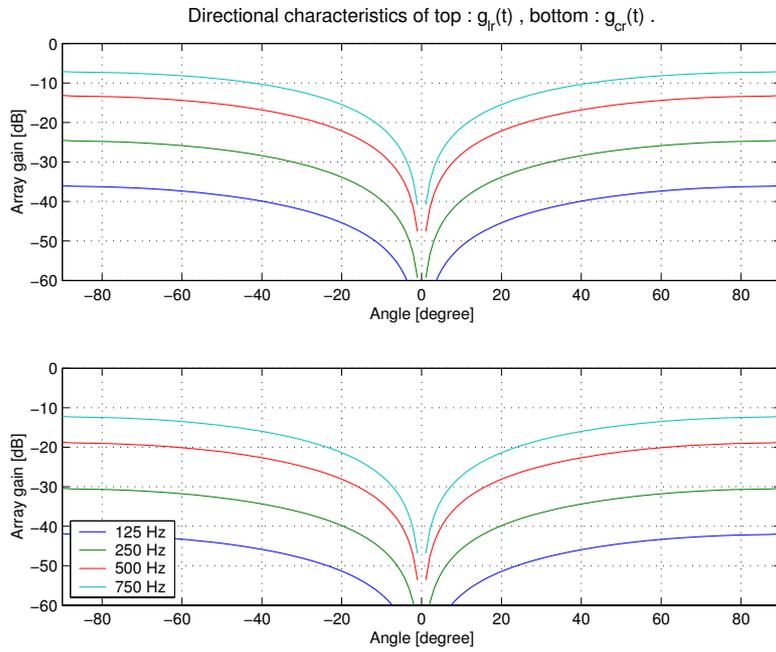


図 4.1: 125 ~ 750 Hz のビームフォーマの指向特性

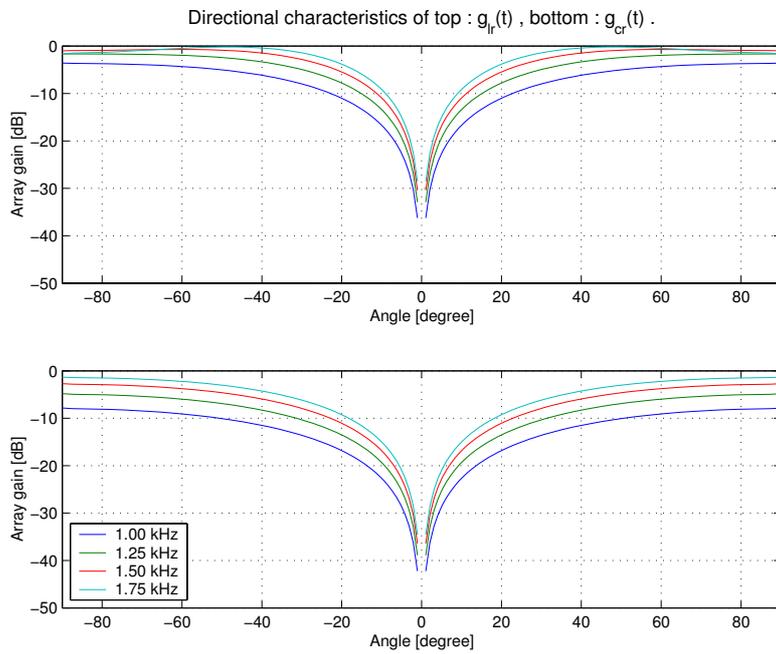


図 4.2: 1 ~ 1.75 kHz のビームフォーマの指向特性

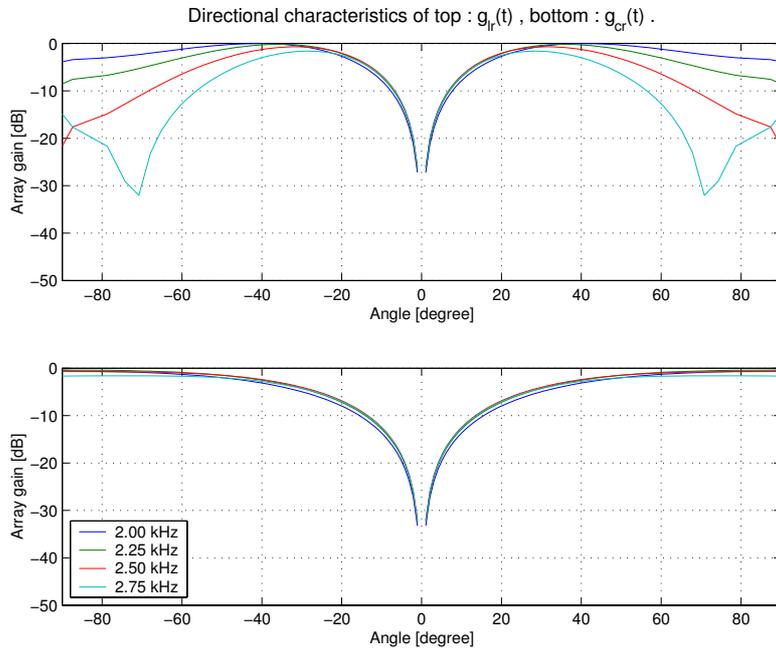


図 4.3: 2 ~ 2.75 kHz のビームフォーマの指向特性

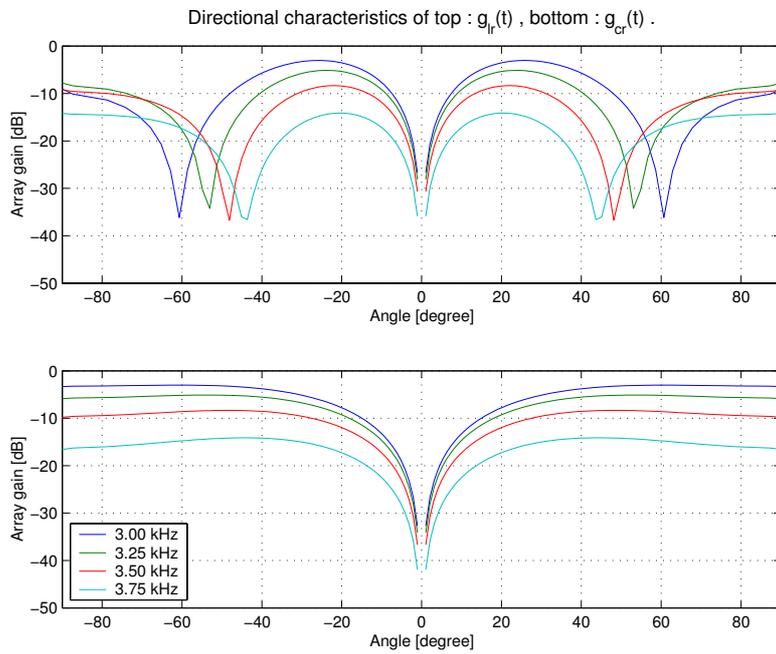


図 4.4: 3 ~ 3.75 kHz のビームフォーマの指向特性

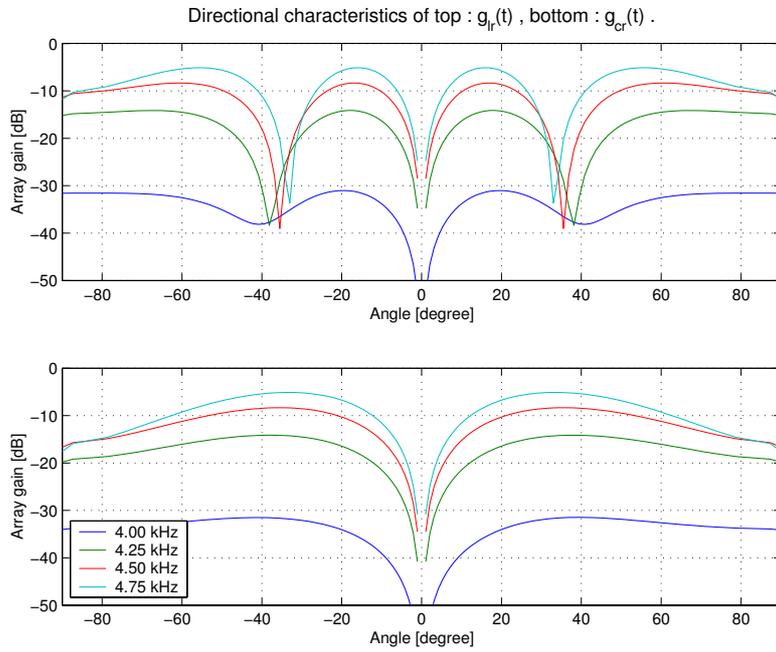


図 4.5: 4 ~ 4.75 Hz のビームフォーマの指向特性

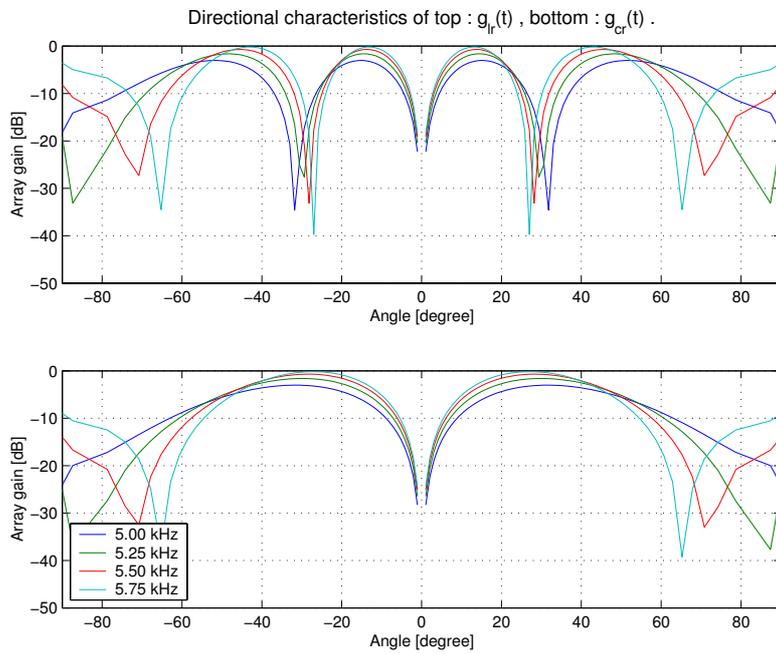


図 4.6: 5 ~ 5.75 kHz のビームフォーマの指向特性

るように、16 kHz サンプルングで信号の到来方向推定を行った場合、 $\pm 19, 41, 79$  のうちから一方向を選ぶことになる。

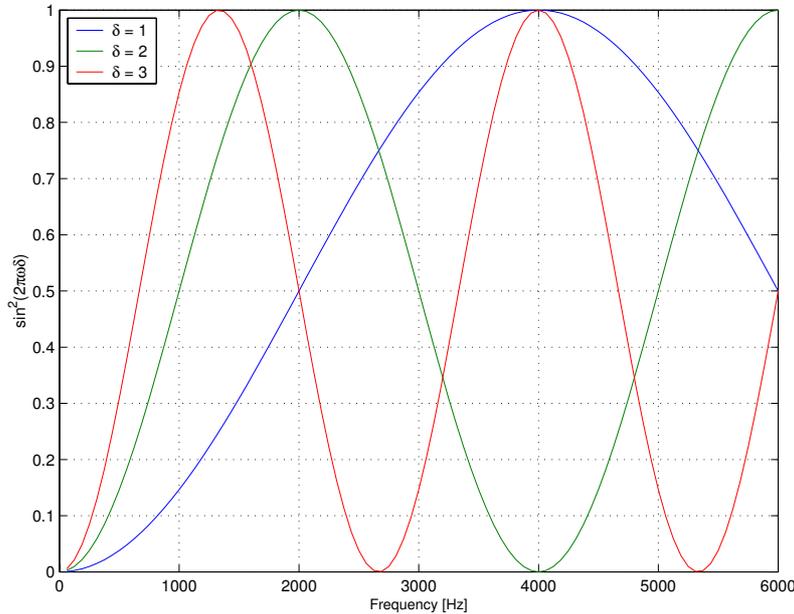


図 4.7: 到来時間差  $\delta$  と周波数と分母項の関係

次に雑音スペクトルの推定精度が低下する到来方向と周波数帯域について述べる。図 4.7 を見ると分母項の値が 0 に近づくのが、1 kHz 以下の周波数に集中していることがわかる。このことから、1 kHz 以下の周波数帯域では雑音スペクトルの推定誤差が大きくなり、雑音の取り残しや音声のスペクトルまでも除去してしまう。図 4.8 を見ると、周波数帯域 5 ~ 6 kHz、到来方向 50 ~ 80 °では二つのビームフォーマが空間的折り返し現象のために、共に死角を形成している。そのため、この方向から到来した雑音を観測することが出来ず、正確に雑音スペクトルを推定することが出来ない。

また、重大な問題点として注目したいのが、到来方向 20 ~ 40 °、5 ~ 6 kHz の周波数帯域でのアルゴリズムの挙動である。ここで図 4.10 に注目すると、5 ~ 6 kHz、到来時間差  $\delta_k = 2$  での分母項の値は 0.5 ~ 1.0 である。籠らの提案法では閾値  $\varepsilon_1 = 0.5, \varepsilon_2 = 0.2$  としているため、ここではビームフォーマ  $g_{lr}(t)$  を利用して雑音スペクトルを推定する(図 4.9 参照)。しかしながら、図 4.10 を見ると  $g_{lr}(t)$  は指向特性の死角を形成していることがわかる。ビームフォーマ  $g_{lr}(t)$  が空間的折り返し現象のために死角を形成した場合、ビームフォーマ  $g_{cr}(t)$  を用いるという動作が働かないために、アルゴリズムが抑圧された信号中から雑音スペクトルを推定しようとするために、殆ど雑音を推定することが出来ない。

また、雑音スペクトルの推定部には大きな問題がある。籠らの研究や本研究で扱う信号処理はデジタル信号処理である。そのため、扱うことが出来る信号は時間的に離散な信号のみである。今、ビームフォーマ  $g_{lr}(t)$  と  $g_{cr}(t)$  に注目すると、遅延操作量  $\tau$  は整数値

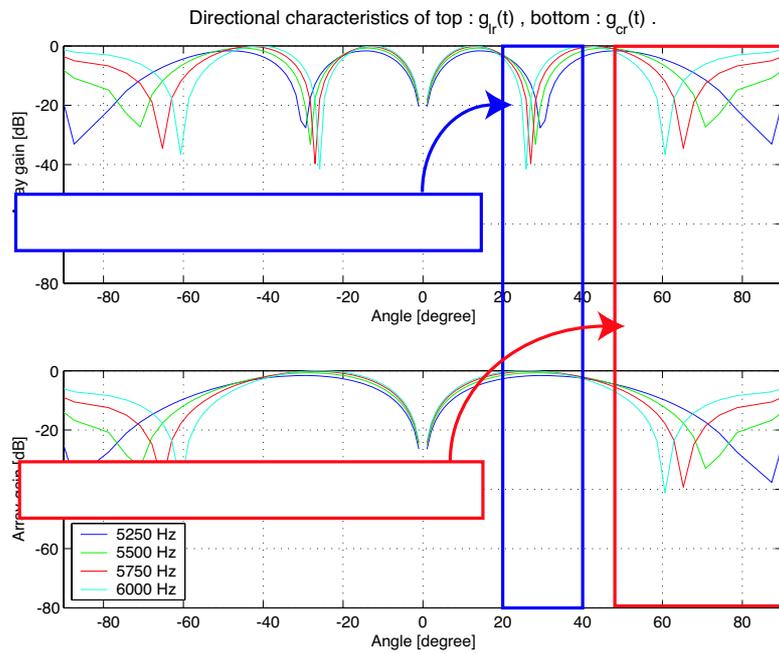


図 4.8: ビームフォーマの補完関係

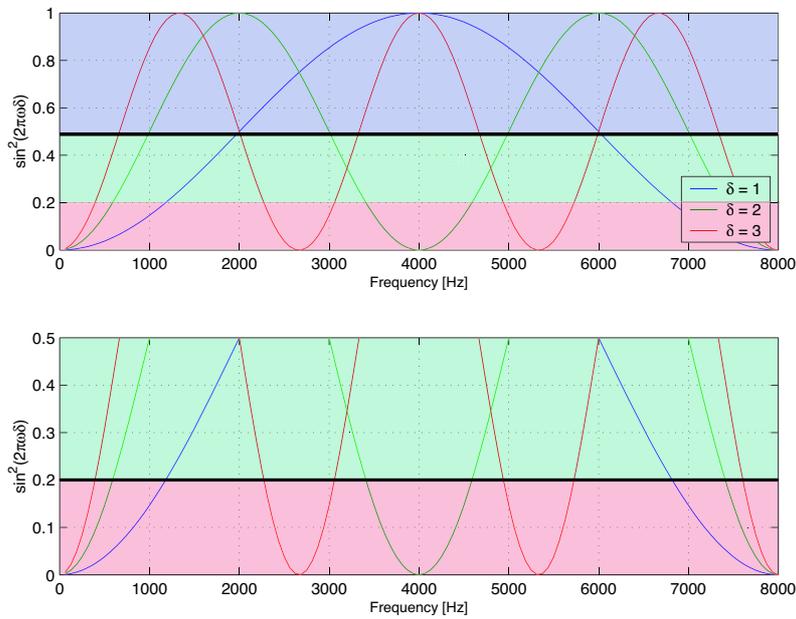


図 4.9: 閾値と雑音スペクトル推定部におけるビームフォーマの関係

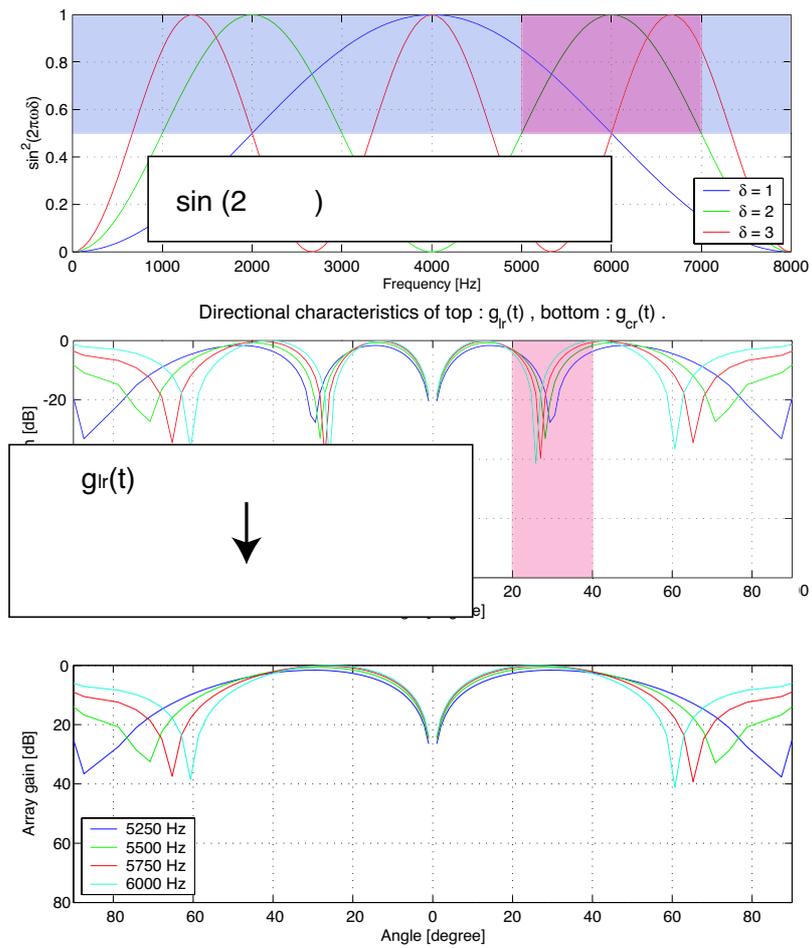


図 4.10: 雑音スペクトル推定部における分母項とビームフォーマの関係

しか取ることが出来ない。しかしながら、仮に到来方向を  $\delta_k = 3$  と推定した場合、ビームフォーマ  $g_{cr}(t)$  で用いられる遅延操作量は  $\tau = \text{round}(\delta_k/2) = 2$  となる。籠らの提案法では、雑音スペクトルの推定にこの丸められた遅延操作量を用いているため、雑音スペクトルの推定精度が低くなっていると考えられる。

### 4.3 問題点のまとめ

前節までに挙げてきた籠らの提案法の問題点として、

1. 1 kHz 以下の低い周波数帯域のビームフォーマの指向特性が非常に緩やかである。このため、殆ど雑音スペクトルを推定することが出来ず、雑音除去した音声に雑音が残る原因になる。
2. 推定した到来方向によっては雑音スペクトルの分母項の値が 1 に近づくにも関わらず、ビームフォーマが死角を形成するために雑音スペクトルの正確な推定が不可能である。ビームフォーマで信号を抑圧しているにも関わらず、抑圧した信号中から雑音スペクトルを推定しようとするために雑音スペクトルを 0 と推定してしまう。
3. 雑音スペクトルの推定部で用いる到来時間差は小数点以下を切り捨てられており、雑音スペクトルの推定に誤差を引き起こす。

というものがあつた。1 はビームフォーマの原理的な問題であり、如何なる方法を用いてもビームフォーマを急峻に形成することは出来ない。次節で 2 と 3 についてどのように解決するかを考える。3 章で説明しているが、雑音スペクトルの推定は二つのビームフォーマを用いて行われる。ここで、ビームフォーマを切り替えるために二つの閾値を設定している。雑音スペクトル推定部の分母項が二つの閾値より大きいかどうかで  $G_{lr}(\tilde{\omega})$  を用いた推定,  $G_{cr}(\tilde{\omega})$  を用いた推定, または  $G_{lr}(\tilde{\omega})$  を用いた近似の三つのうち一つを選ぶ。次節ではこれらの問題の改善法について考える。

### 4.4 問題点の改善法

本節では前章で挙げた問題点の改善法について述べる。まず、雑音スペクトルの推定部で用いる到来時間差は小数点以下を切り捨てられており、雑音スペクトルの推定に誤差を引き起こすという問題を改善する。次に推定した到来方向によっては雑音スペクトルの分母項の値が 1 に近づくにも関わらず、ビームフォーマが死角を形成するために雑音スペクトルの正確な推定が不可能になるという問題の解決法について考える。

#### 4.4.1 雑音スペクトル推定部の改善

籠らの提案法における雑音スペクトルの推定部には大きな問題がある。籠らの研究や本研究で扱う信号処理はデジタル信号処理であるため、扱うことが出来る信号は時間的に離散な信号のみである。ビームフォーマ  $g_{lr}(t)$  と  $g_{cr}(t)$  に注目すると、遅延操作量  $\tau$  は整数値しか取ることが出来ない。しかしながら、仮に到来方向を  $\delta_k = 3$  と推定した場合、ビームフォーマ  $g_{cr}(t)$  で用いられる遅延操作量は  $\tau = \text{round}(\delta_k/2) = 2$  となる。籠らの提案法では、雑音スペクトルの推定にこの丸められた遅延操作量を用いているため、雑音スペクトルの推定精度が低くなっていると考えられる。 $g_{cr}(t)$  を用いて雑音スペクトルを推定すると推定誤差を生じ、雑音スペクトルの取り残しや取り過ぎを引き起こす。そこで、雑音スペクトルの推定精度を改善するために、遅延量の丸め誤差を小さくすることを考える。今、ビームフォーマ  $g_{cr}(t)$  の Fourier 変換  $G_{cr}(\tilde{\omega})$  を考える。

$$G_{cr}(\tilde{\omega}) = N_k(\tilde{\omega})e^{j\tilde{\omega}\frac{\delta_k}{2}} \sin^2 \tilde{\omega} \frac{\delta_k}{2} \quad (4.5)$$

このとき遅延操作量を  $\tau$  (整数) とすると、上式は次のように書くことが出来る。

$$G_{cr}(\tilde{\omega}) = N_k(\tilde{\omega})e^{j\tilde{\omega}\frac{\delta_k}{2}} \sin \tilde{\omega} \frac{\delta_k}{2} \sin \tilde{\omega} \tau \quad (4.6)$$

これを变形して雑音スペクトルの推定部を書き直すと、

$$\hat{N}_k(\tilde{\omega}) = \begin{cases} \frac{G_{lr}(\tilde{\omega})}{(a)}, & (a) > \varepsilon_1 \\ \frac{G_{cr}(\tilde{\omega})e^{-j\tilde{\omega}\frac{\delta_k}{2}}}{(b)}, & ((a) \leq \varepsilon_1) \wedge ((b) > \varepsilon_2) \\ \frac{G_{lr}(\tilde{\omega})}{\varepsilon_1}, & (b) \leq \varepsilon_2 \end{cases} \quad (4.7)$$

となる。ここで (a) と (b) は

$$\begin{aligned} (a) &= \sin^2 \tilde{\omega} \delta_k \\ (b) &= \sin \tilde{\omega} \frac{\delta_k}{2} \sin \tilde{\omega} \tau \end{aligned} \quad (4.8)$$

で、 $\tau = \text{round}(\delta_k/2)$  である。このように遅延量  $\tau$  はそのままに  $\sin \tilde{\omega} \frac{\delta_k}{2}$  を  $\sin \tilde{\omega} \tau$  の代わりに用いることで、雑音スペクトルの推定誤差を小さくすることが出来る。これにより、 $G_{cr}(\tilde{\omega})$  を用いて雑音スペクトルを推定した際にも推定誤差を小さくすることが出来る。また、雑音スペクトルの近似が主に 1 kHz 以下の低い周波数帯域で行われることを考慮して、近似精度を改善するためにビームフォーマ  $G_{lr}(\tilde{\omega})$  を  $\varepsilon_1$  で補正する。ここで、籠らの提案法における雑音スペクトル推定部は

$$\hat{N}_k(\tilde{\omega}) = \begin{cases} G_{lr}(\tilde{\omega}) / \sin^2 \tilde{\omega} \delta_k, & \sin^2 \tilde{\omega} \delta_k > \varepsilon_1 \\ G_{cr}(\tilde{\omega})e^{-j\tilde{\omega}\tau} / \sin^2 \tilde{\omega} \tau, & (\sin^2 \tilde{\omega} \delta_k \leq \varepsilon_1) \wedge (\sin^2 \tilde{\omega} \tau > \varepsilon_2) \\ G_{lr}(\tilde{\omega}) & \sin^2 \tilde{\omega} \tau \leq \varepsilon_2 \end{cases} \quad (4.9)$$

である。二つの方法における主な違いはビームフォーマ  $G_{cr}(\tilde{\omega})$  で雑音スペクトルを推定する際に用いられる到来方向と、ビームフォーマ  $G_{lr}(\tilde{\omega})$  を用いて雑音スペクトルを近似する際の分母項である。これらの改善によって  $G_{cr}(\tilde{\omega})$  を用いた雑音スペクトルの推定での過剰推定を改善することが出来た。また、ビームフォーマ  $G_{lr}(\tilde{\omega})$  を用いて雑音スペクトルを近似する際に近似精度を向上させるため  $\varepsilon_1$  で割っている。これにより低周波数帯域における推定精度の改善に成功した。

#### 4.4.2 周波数帯域毎の閾値設定による雑音スペクトルの推定精度改善

到来方向  $20 \sim 40^\circ$ 、 $5 \sim 6$  kHz の周波数帯域、到来時間差  $\delta_k = 2$  での雑音スペクトルの推定部における分母項の値は  $0.5 \sim 1.0$  である。ここでビームフォーマ  $g_{lr}(t)$  を利用して雑音スペクトルを推定するが、 $g_{lr}(t)$  は指向特性の死角を形成しているために殆ど雑音を推定することが出来ない。そのため、この周波数帯域に存在する雑音は殆ど除去されることなく音声信号に残る。 $5 \sim 6$  kHz 付近に雑音があると子音の了解度を下げる可能性がある。音声の了解度改善を図るためには、この周波数帯域の雑音除去が非常に重要である。

このような問題は閾値の値を変え、ビームフォーマを  $g_{cr}(t)$  に切り替えることで補完機能を働かせることで改善することが出来る。しかしながら、閾値を変えると他の周波数帯域での雑音スペクトルの推定精度に悪影響を及ぼすことが考えられる。そこで、周波数帯域毎に閾値を設定することにより上記のような問題を解決し、最も推定精度が低いと考えられる  $G_{lr}(\tilde{\omega})$  を用いた雑音スペクトルの近似を用いる部分を減らすことを考える。

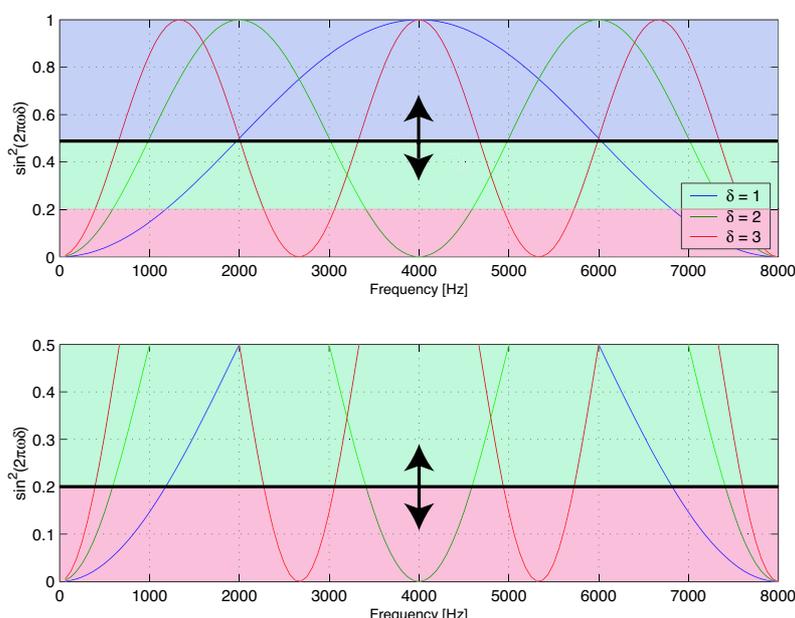


図 4.11: 雑音スペクトル推定部における分母項とビームフォーマの関係

閾値を決定する基準として各周波数帯域において以下の点に注意した。

#### 0 ~ 1 kHz

ビームフォーマの指向特性が緩やかなために雑音スペクトルの推定精度があまりよくなかったが、雑音スペクトルの推定精度改善により、閾値を小さい値にしても雑音スペクトルの推定誤差が小さくなった。750 Hz 以下の推定精度改善と雑音の過剰推定がなるべく少なくなるように閾値を設定した。

#### 1 ~ 3 kHz

ビームフォーマの指向特性が急峻で精度よく雑音スペクトルを推定することが出来るので、更に推定精度が上がるように設定した。

#### 3 ~ 4 kHz

ビームフォーマ  $g_{cr}(t)$  を使う頻度が非常に多い。これは空間的折り返し現象が 3 kHz 付近の帯域から起こり始めるためである。しかしながら、 $g_{cr}(t)$  を用いると雑音スペクトルの過剰推定が大きい傾向があったため、推定誤差が小さいうちは  $g_{lr}(t)$  を使うように設定した。

#### 4 ~ 5 kHz

3 ~ 4 kHz 同様に  $g_{cr}(t)$  を使う頻度を小さくするようにした。しかしながら、空間的折り返し現象が起こるために 3 ~ 5 kHz の帯域では 0 ~ 3 kHz の帯域に比べて  $g_{cr}(t)$  を使う頻度は高い。

#### 5 ~ 6 kHz

最も注意した点は前章節でも挙げているビームフォーマ  $g_{lr}(t)$  が指向特性の谷を形成しているにも関わらず、分母項が大きいために  $g_{cr}(t)$  を用いずに  $g_{lr}(t)$  を用いて雑音スペクトルを推定するという問題が起こらないようにしたことである。この問題が起こると雑音は殆ど除去されないため、子音の了解度が著しく下がる原因になる。 $g_{lr}(t)$  を用いる頻度を少なくすることでビームフォーマの補完が上手く機能するように設定した。

これらの設定により、より正確な雑音スペクトルの推定が可能になった。0 ~ 2 kHz には母音の第一・第二フォルマントが存在しているため、雑音スペクトルの過剰推定が起こらないように注意した。閾値  $\varepsilon_1, \varepsilon_2$  は籠らの提案法で用いられたものに対して、共に小さい値となっている。

## 4.5 問題点の改善にあたって目指したこと

前節まで、籠らの提案法における問題点を挙げ、その解決法について述べた。ここで問題改善にあたって考慮したことについて挙げる。籠らの提案法で処理した音声はその後、6章での了解度実験でも示されるように了解度が低い。聞き間違いを起こす原因として子音の聞き間違いと母音の聞き間違いがある。

男性の日本語母音 /i/, /u/ の第一フォルマントは 100 ~ 400 Hz 付近に存在している。しかしながら、籠らの提案法では 1 kHz 以下の周波数帯域に存在する雑音スペクトルは殆ど推定することが出来ない。そのために母音のフォルマントがマスキングされ、聞き間違いの原因になっていると考えられる。そこで、本手法ではまず低周波数帯域における雑音スペクトルの精度向上について考えた。そして前節で挙げたような雑音スペクトル推定部の改善によって、低周波数帯域における雑音スペクトルの推定精度を向上させることに成功した。

また、子音の聞き間違いについては 5 ~ 6 kHz の周波数帯域の雑音に大きな問題があると考えられる。この周波数帯域は音声の明瞭度にも関係しており、了解度改善を目的とした本研究においても非常に重要な周波数帯域である。籠らの提案法では閾値設定が適切でない上にビームフォーマの補完が機能していない。そのためにこの周波数帯域における雑音除去は殆どなされず、雑音を残すという結果に陥っている。本手法ではビームフォーマの補完機能を手がかりにして、 $g_{lr}(t)$ ,  $g_{cr}(t)$  が誤差の少ない雑音スペクトル推定を行えるように閾値を設定し、各周波数帯域において大幅な雑音スペクトルの推定精度の改善に成功した。

# 第5章 計算機シミュレーションを用いた 検証

本章では、計算機シミュレーションによって雑音除去アルゴリズムの能力を検証する。籠らの提案法を比較対象として本手法の有効性について述べる。

## 5.1 雑音除去アルゴリズムのパラメータ設定

表 5.1: 雑音除去アルゴリズムのパラメータ設定

パラメータ	設定値 (本手法)	籠らの提案法
フレーム長	512 pt	512 pt
フレーム周期	128 pt	128 pt
窓関数	三角窓	三角窓
サブトラクション係数 $\alpha$	1.0	1.0
フロアリング係数 $\beta$	0.001	0.001
閾値 $\varepsilon_1$	表 5.2 参照	0.5
閾値 $\varepsilon_2$	表 5.2 参照	0.2

表 5.2: 帯域毎に設定した閾値  $\varepsilon_1$  と  $\varepsilon_2$

周波数 [kHz]	0 ~ 3	3 ~ 4	4 ~ 5	5 ~ 6
$\varepsilon_1$	0.10	0.35	0.25	0.65
$\varepsilon_2$	0.08	0.08	0.08	0.06

## 5.2 複数雑音に対する雑音除去実験

籠らの提案法は複数雑音に頑健な雑音除去法である。そこで本手法が籠らの提案法をどの程度改善したのかを検証する。籠らの提案法の問題点が実際にどのような結果を引き起こしているのか、本手法がそれをどの程度改善出来ているのかについて計算機シミュレーションを通して検証する。

### 5.2.1 シミュレーション条件

目的音として、ATR 音声データベースに収録されている男性話者 mht のクリーンな音声/aoi/を到来方向  $0^\circ$  から入力する。雑音として計算機で作成した  $0 \sim 1, 3 \sim 4, 5 \sim 6$  kHz の帯域雑音をそれぞれ到来方向  $61, -50, 31^\circ$  から入力した。SNR は目的音に対してそれぞれ 0 dB である。データのサンプリング周波数は入力時が 288 kHz で雑音除去を行う際には 16 kHz にダウンサンプリングした。

### 5.2.2 雑音除去結果

雑音の評価尺度として SD (スペクトル歪み) を用い、以下の式を用いて求めた。

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \{X(\omega) - Y(\omega)\}^2} \text{ [dB]} \quad (5.1)$$

ここで、 $x(t)$  はクリーンな音声、 $y(t)$  は評価対象の音声、 $X(\omega), Y(\omega)$  は対数振幅スペクトルで  $N$  は上限周波数である。

表 5.3: SD による雑音除去評価

	雑音除去前	雑音除去後 (本手法)	雑音除去後 (籠らの提案法)
SD [dB]	1.64	1.29	1.57

籠らの提案法と本手法の推定した雑音スペクトル (図 5.3, 5.4) を比較すると、特に 1 kHz 以下の周波数帯域で大きな違いがあることがわかる。籠らの提案法では殆ど雑音スペクトルを推定することが出来ていないが、本手法では 250 Hz 付近までの雑音スペクトルを推定している。これにより雑音除去後のスペクトルを見ても本手法で雑音除去を行ったものの方がより多くの雑音を除去することが出来ている。SD を見ると 0.3 dB 程度の改善が見られる。

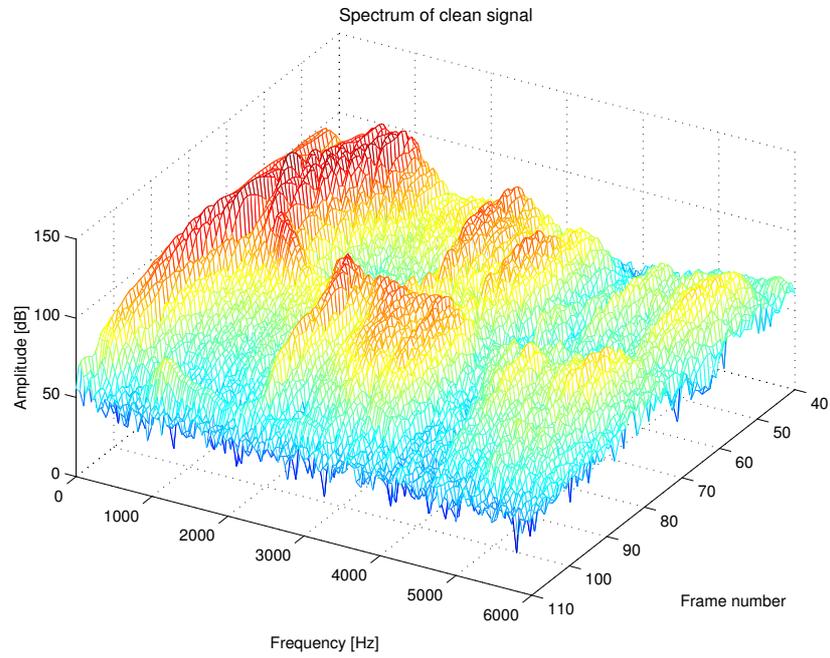


図 5.1: クリーンな音声信号のスペクトル

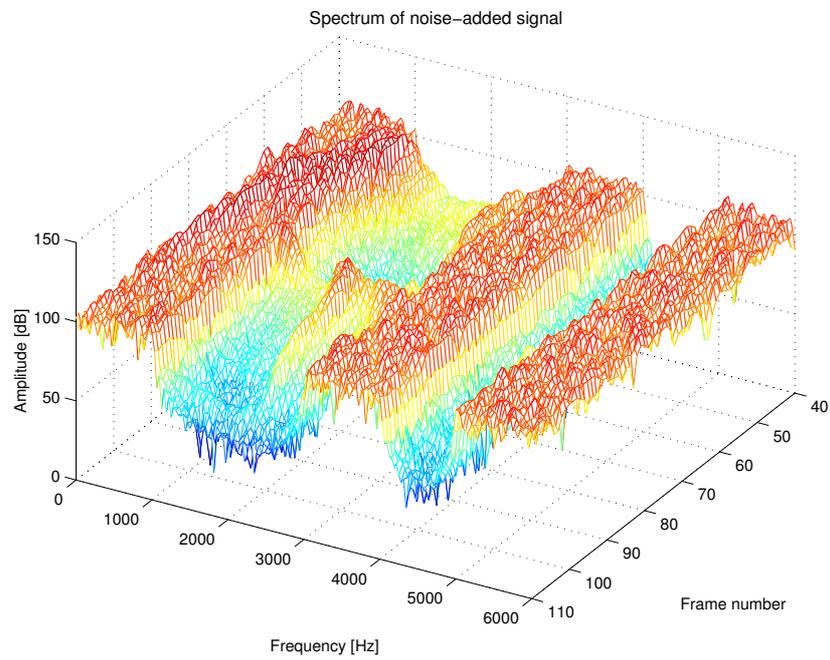


図 5.2: 雑音を付加した音声信号のスペクトル (複数雑音)

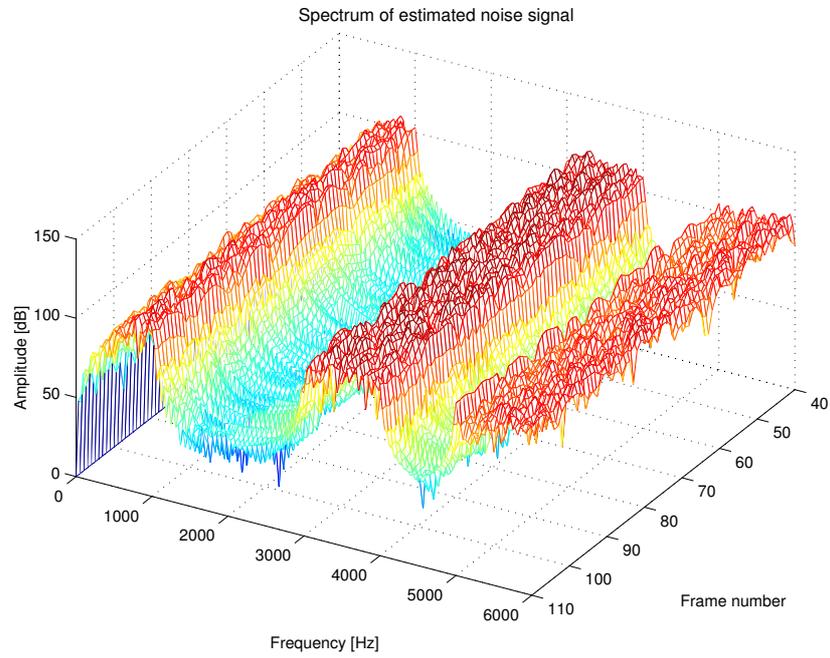


図 5.3: 籠らの提案法を用いて推定した雑音スペクトル (複数雑音)

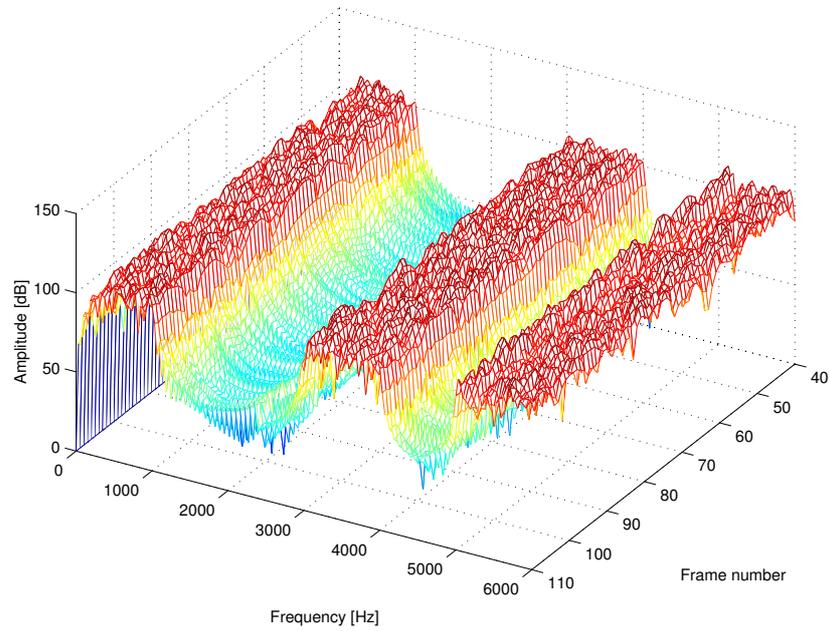


図 5.4: 本手法を用いて推定した雑音スペクトル (複数雑音)

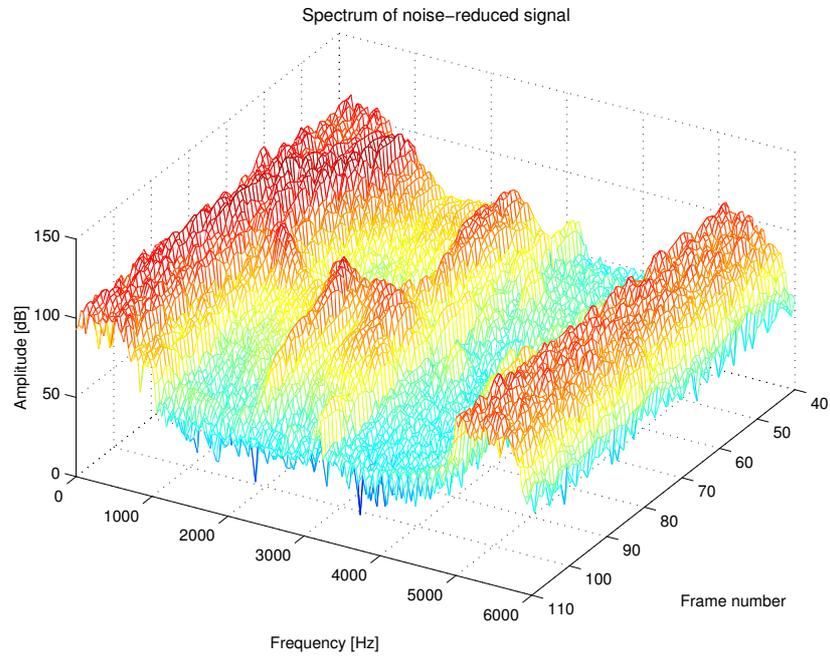


図 5.5: 籠らの提案法を用いて雑音除去した音声信号のスペクトル (複数雑音)

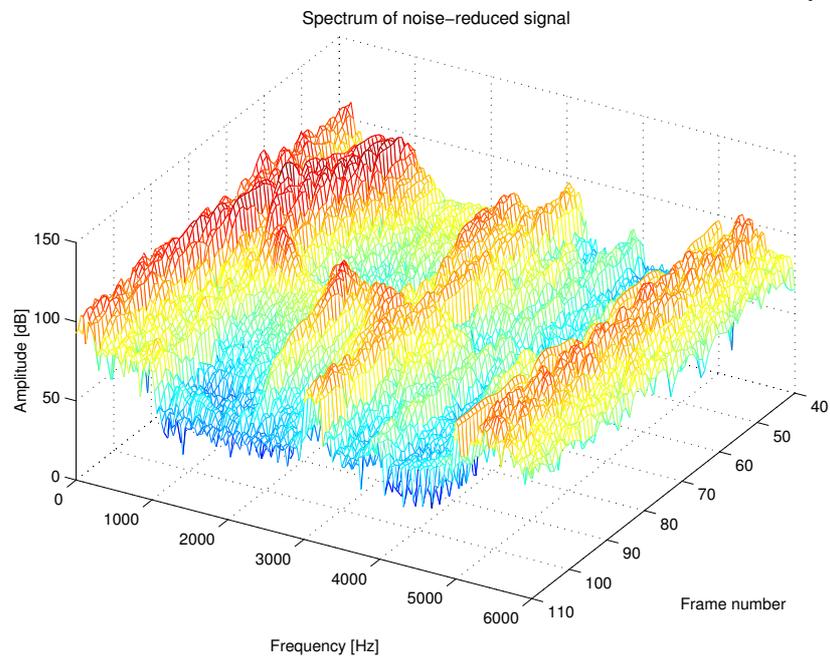


図 5.6: 本手法を用いて雑音除去した音声信号のスペクトル (複数雑音)

## 5.3 周波数帯域が重複している雑音に対する雑音除去実験

複数雑音の周波数帯域が重複している場合に、雑音除去がどの程度行われるのかについて計算機シミュレーションを用いて検証する。

### 5.3.1 シミュレーション条件

目的音として、ATR 音声データベースに収録されている男性話者 mht のクリーンな音声/aoi/を到来方向 0°から入力する。雑音として計算機で作成した 0 ~ 1, 2 ~ 4, 3 ~ 4 kHz の帯域雑音をそれぞれ到来方向 61, -50, 31°から入力した。SNR は目的音に対してそれぞれ 0 dB である。データのサンプリング周波数は入力時が 288 kHz で雑音除去を行う際には 16 kHz にダウンサンプリングした。

### 5.3.2 雑音除去結果

表 5.4: SD による雑音除去評価

	雑音除去前	雑音除去後 (本手法)	雑音除去後 (籠らの提案法)
SD [dB]	1.56	1.28	1.41

籠らの提案法と本手法の推定した雑音スペクトルを比較すると、雑音の重複している帯域ではどちらの方法もよく推定出来ているが、若干籠らの提案法の方がよく推定出来ている。SD の差は主に 1 kHz 以下の周波数帯域の雑音除去による差であると考えられる。

## 5.4 非定常な突発雑音に対する雑音除去実験

実環境において常に存在する非定常な突発雑音に対して本手法の有効性を検証する。

### 5.4.1 シミュレーション条件

目的音として、ATR 音声データベースに収録されている男性話者 mht のクリーンな音声/aoi/を到来方向 0°から入力する。雑音として計算機で作成した 0 ~ 1 (非定常), 3 ~ 4, 5 ~ 6 kHz の帯域雑音をそれぞれ到来方向 61, -50, 31°から入力した。SNR は目的音に対してそれぞれ 0 dB である。データのサンプリング周波数は入力時が 288 kHz で雑音除去を行う際には 16 kHz にダウンサンプリングした。

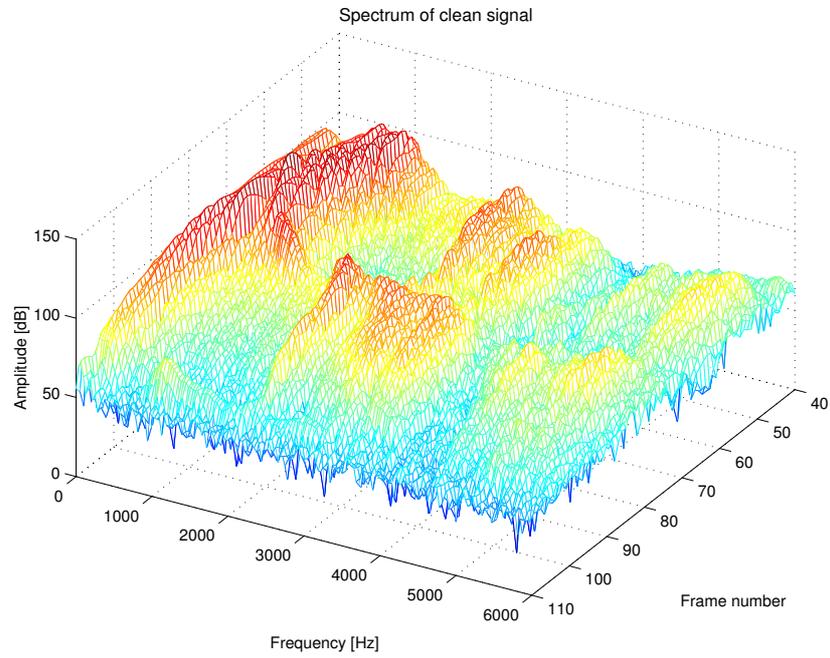


図 5.7: クリーンな音声信号のスペクトル

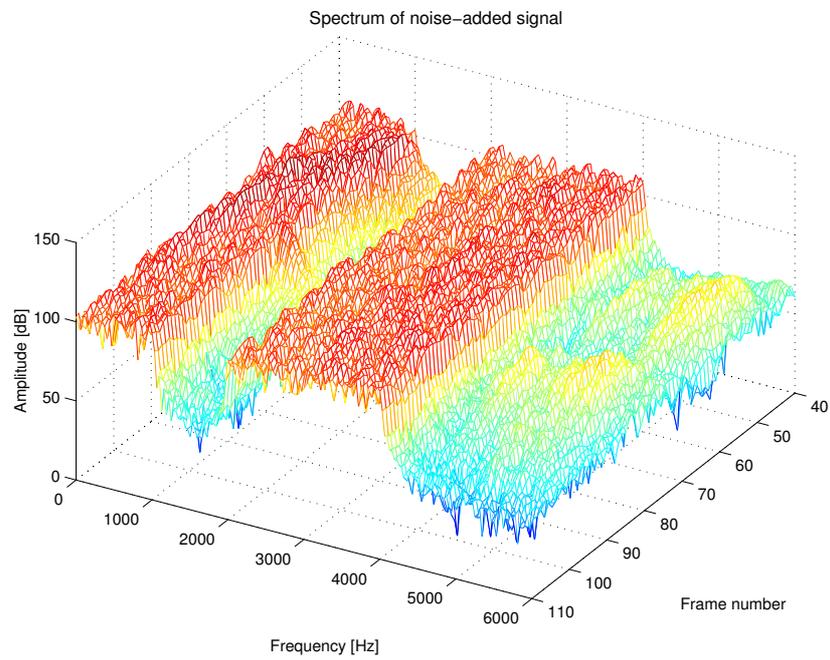


図 5.8: 雑音を付加した音声信号のスペクトル (周波数帯域が重複している雑音)

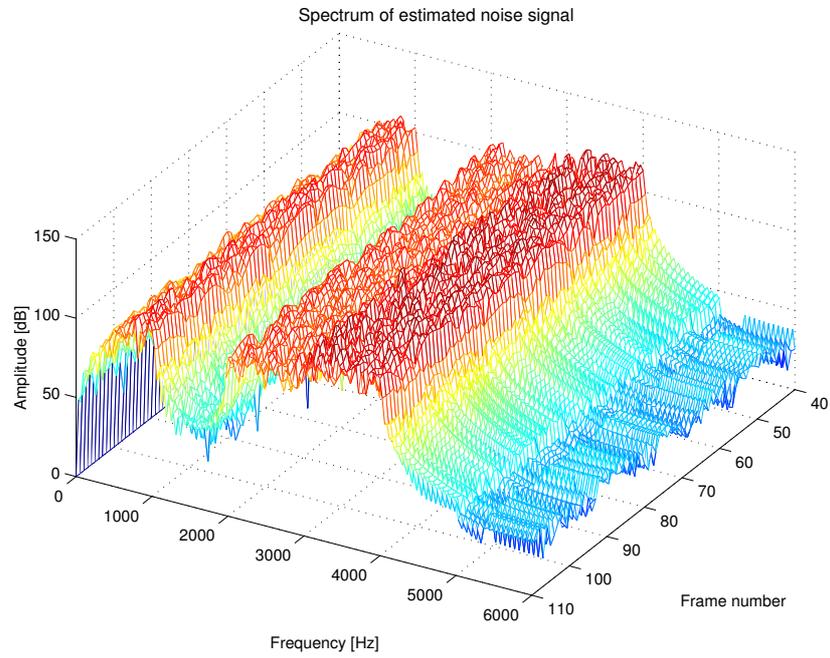


図 5.9: 籠らの提案法を用いて推定した雑音スペクトル (周波数帯域が重複している雑音)

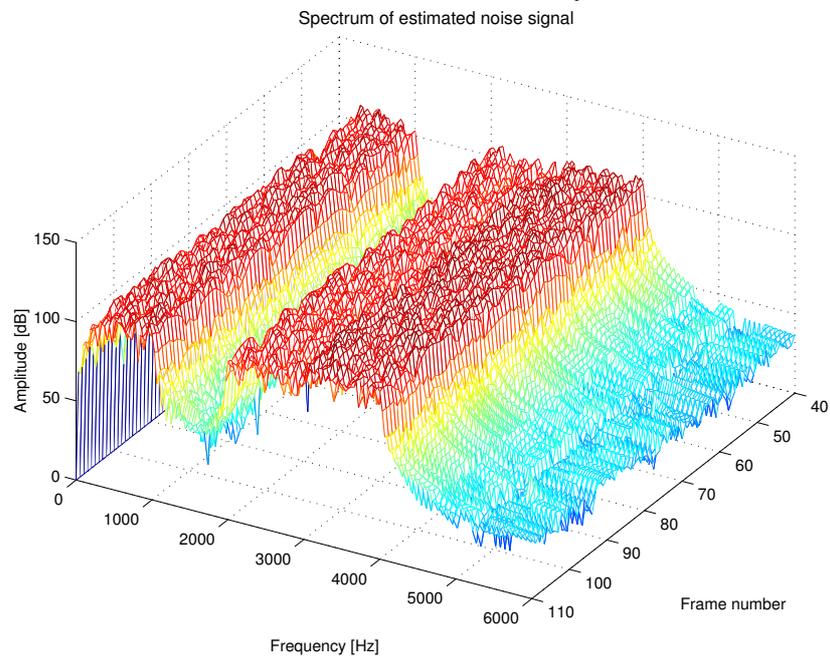


図 5.10: 本手法を用いて推定した雑音スペクトル (周波数帯域が重複している雑音)

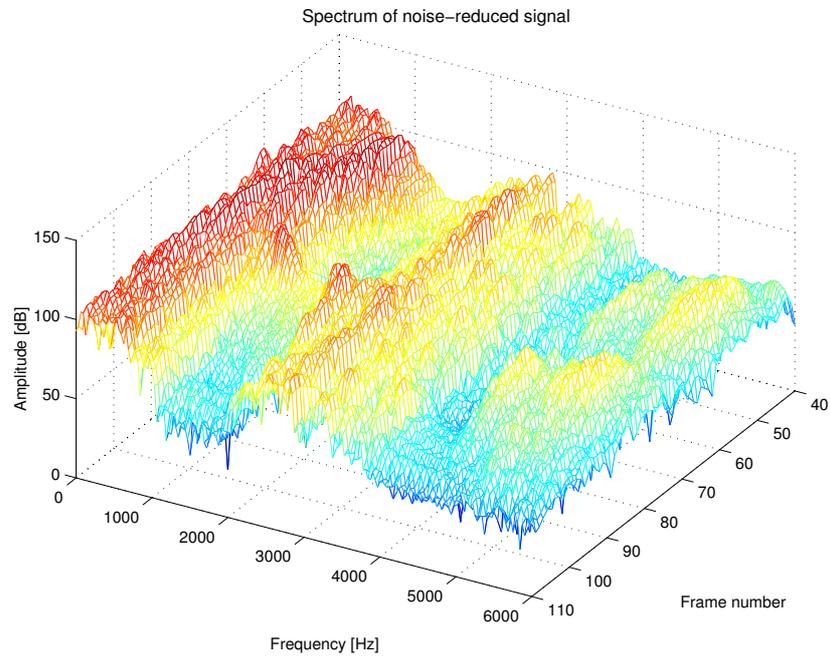


図 5.11: 籠らの提案法を用いて雑音除去した音声信号のスペクトル (周波数帯域が重複している雑音)

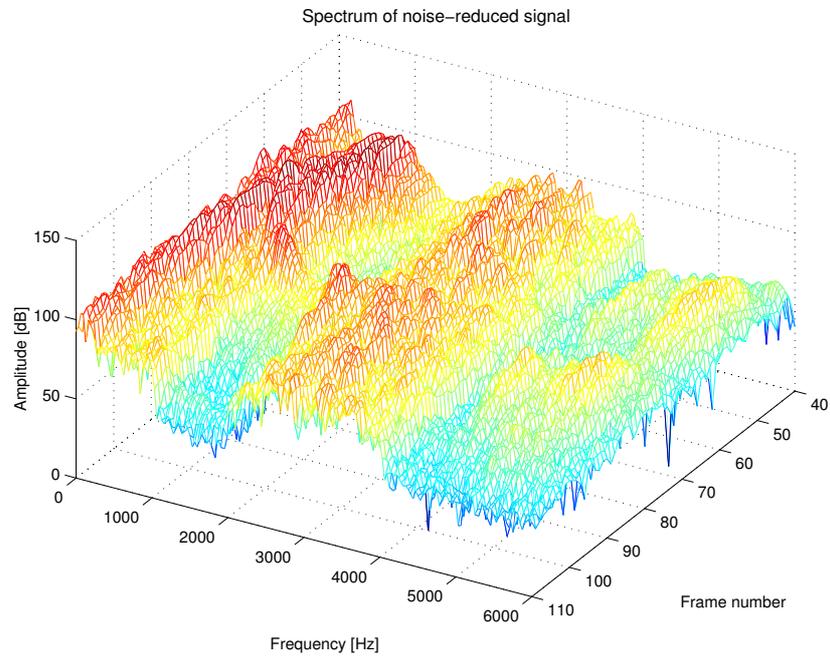


図 5.12: 本手法を用いて雑音除去した音声信号のスペクトル (周波数帯域が重複している雑音)

## 5.4.2 雑音除去結果

表 5.5: SD による雑音除去評価

	雑音除去前	雑音除去後（本手法）	雑音除去後（籠らの提案法）
SD [dB]	1.59	1.27	1.51

推定した雑音スペクトルを比較すると、本手法のものは籠らの提案法で推定したものよりもかなりよく突発雑音を推定出来ていることがわかる。また、5～6 kHzの雑音についても本手法の方がよく推定出来ている。このような推定精度の差がSDによく表れている。籠らの提案法がわずかに0.08 dBしか改善出来ていないのに対して、本手法では0.32 dBほど改善している。

## 5.5 複数の音声が存在する場合の雑音除去実験

人混みの中など、複数の音声が存在する場合に対して本手法の有効性を検証する。

### 5.5.1 シミュレーション条件

目的音として、親密度別単語理解度試験用音声データベースに収録されている男性話者 mya のクリーンな音声「ヒキガネ」を到来方向 0°から入力する。雑音として親密度別単語理解度試験用音声データベースに収録されている女性話者 fhi の「ツキヨミ」、男性話者 mis の「クミイト」、女性話者 fto の「ムラサメ」をそれぞれ到来方向 31, -50, 79°から入力した。SNR は目的音に対してそれぞれ 0 dB である。データのサンプリング周波数は入力時が 288 kHz で雑音除去を行う際には 16 kHz にダウンサンプリングした。

### 5.5.2 雑音除去結果

表 5.6: SD による雑音除去評価

	雑音除去前	雑音除去後（本手法）	雑音除去後（籠らの提案法）
SD [dB]	0.61	1.33	1.39

推定した雑音スペクトルと雑音除去後の音声のスペクトルではわかりづらいが、籠らの提案法では元の音声のスペクトルまで除去しているのに対して、本手法ではある程度元の

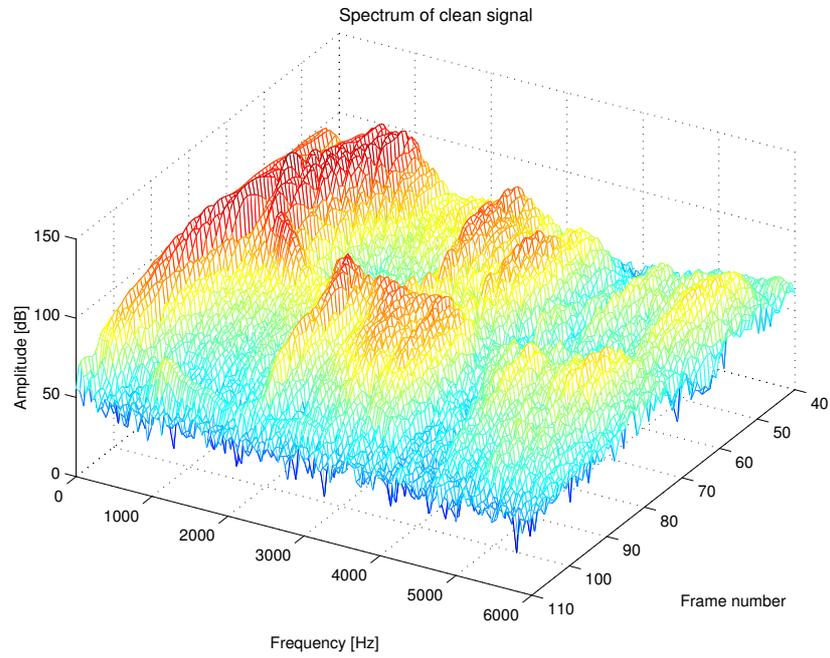


図 5.13: クリーンな音声信号のスペクトル

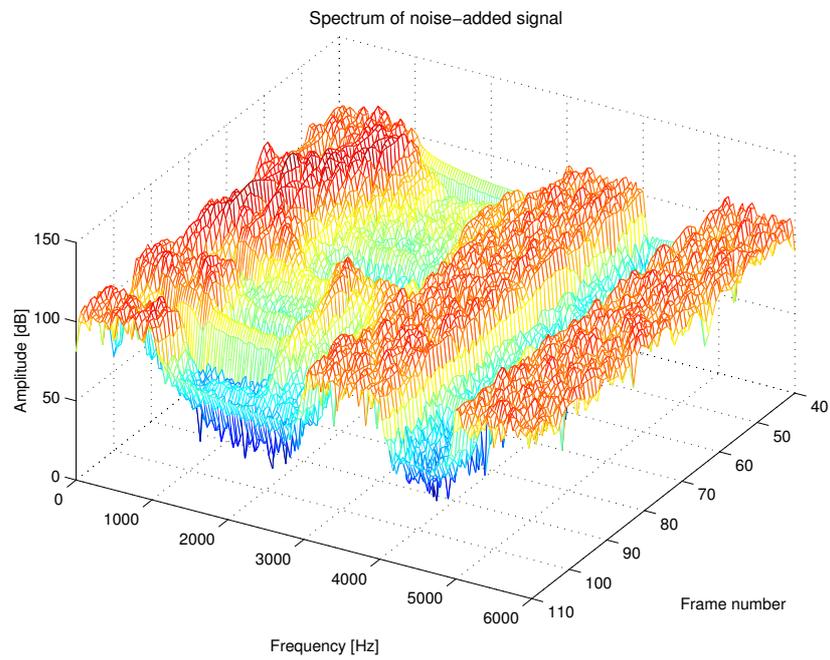


図 5.14: 雑音を付加した音声信号のスペクトル (非定常な突発雑音)

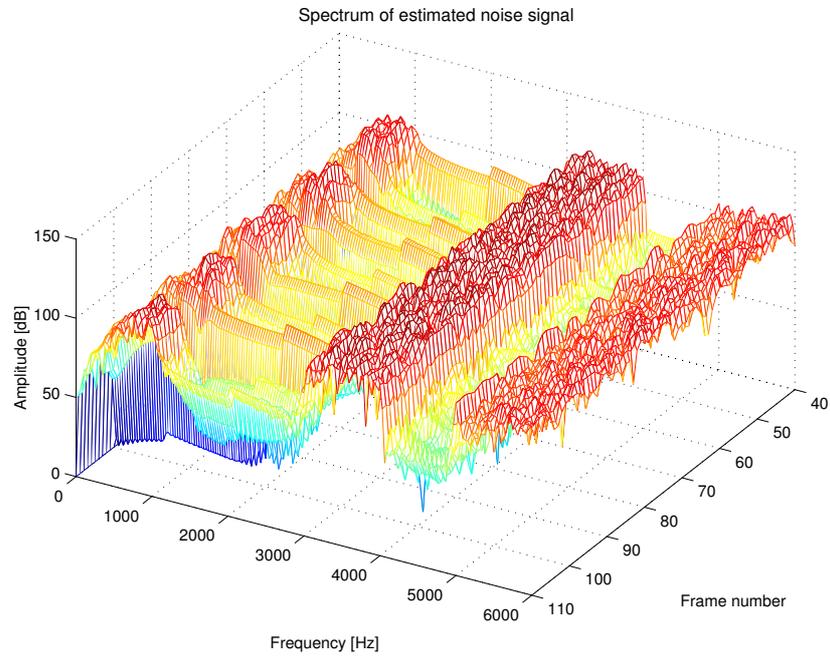


図 5.15: 籠らの提案法を用いて推定した雑音スペクトル (非定常な突発雑音)

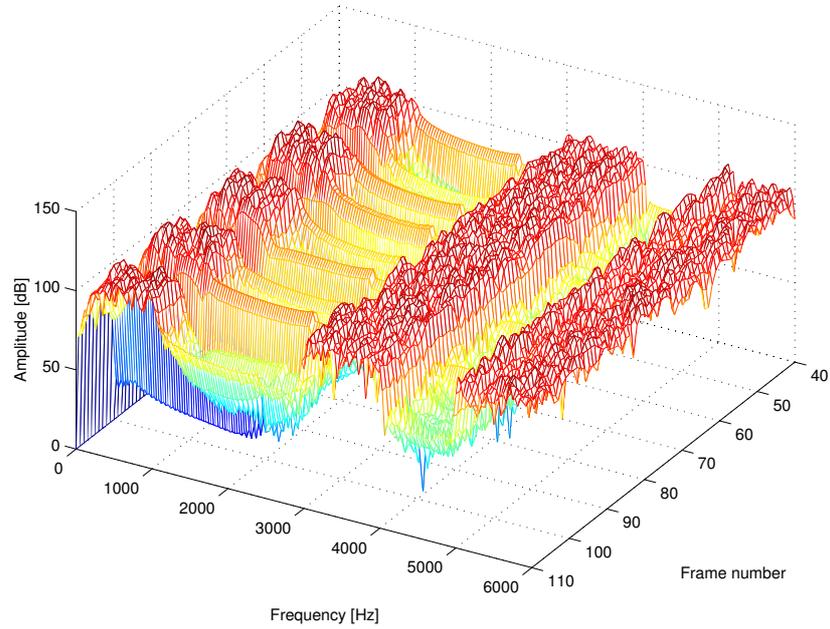


図 5.16: 本手法を用いて推定した雑音スペクトル (非定常な突発雑音)

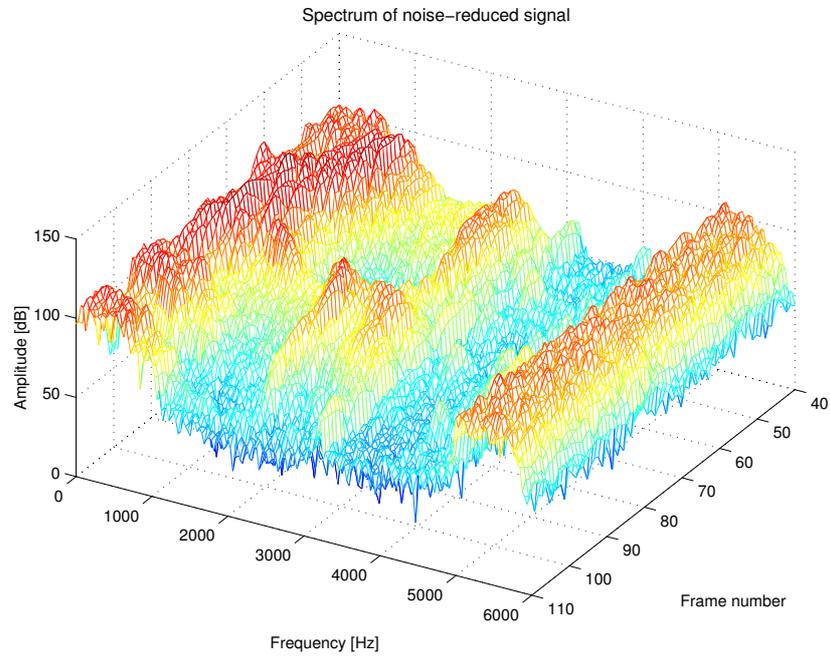


図 5.17: 籠らの提案法を用いて雑音除去した音声信号のスペクトル (非定常な突発雑音)

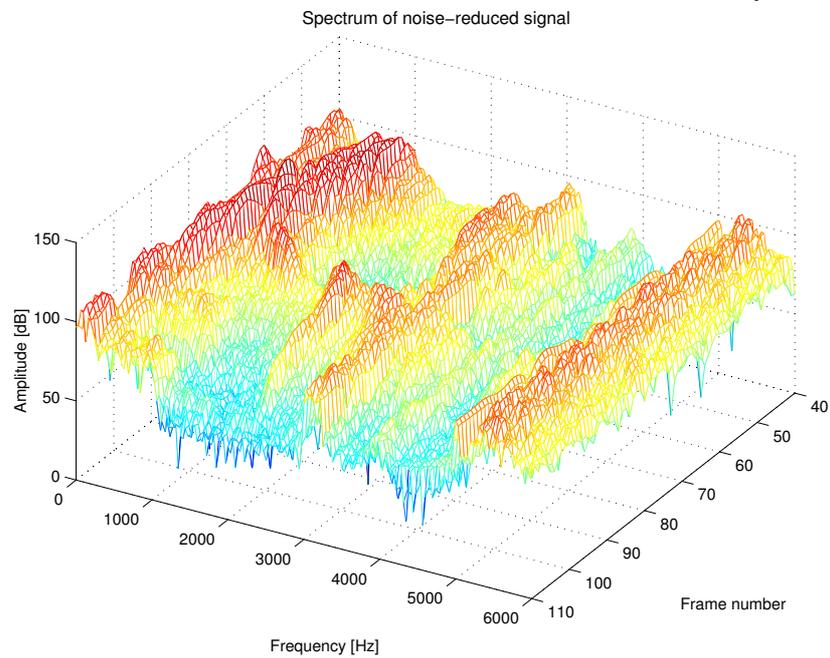


図 5.18: 本手法を用いて雑音除去した音声信号のスペクトル (非定常な突発雑音)

音声のスペクトルを残すことに成功している。しかしながら、籠らの提案法も本手法も雑音除去前よりも SD が増加していることがわかる。

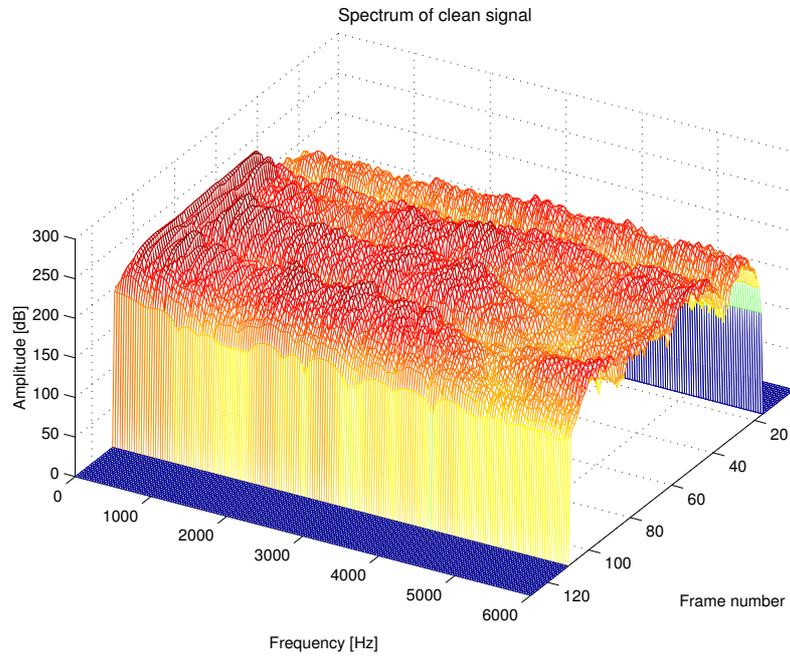


図 5.19: クリーンな音声信号のスペクトル  
Spectrum of noise-added signal

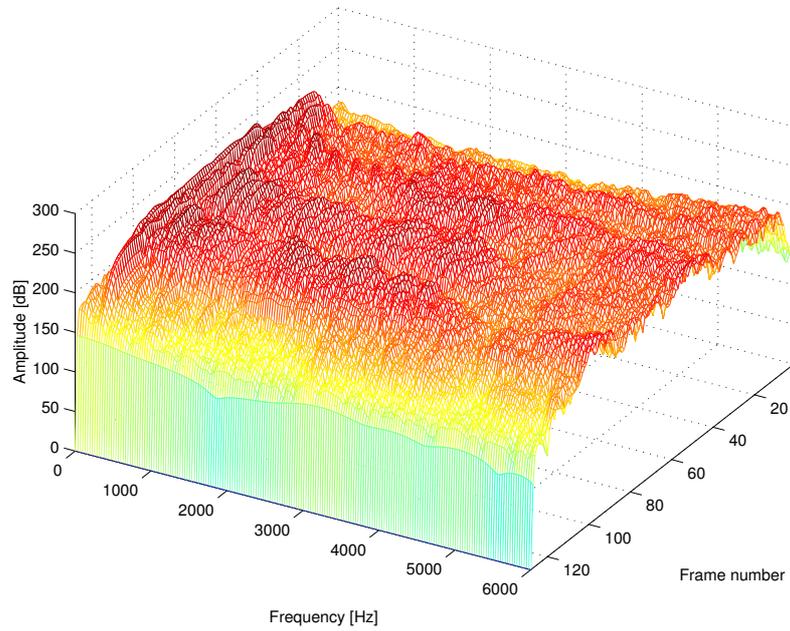


図 5.20: 雑音を付加した音声信号のスペクトル (複数の音声)

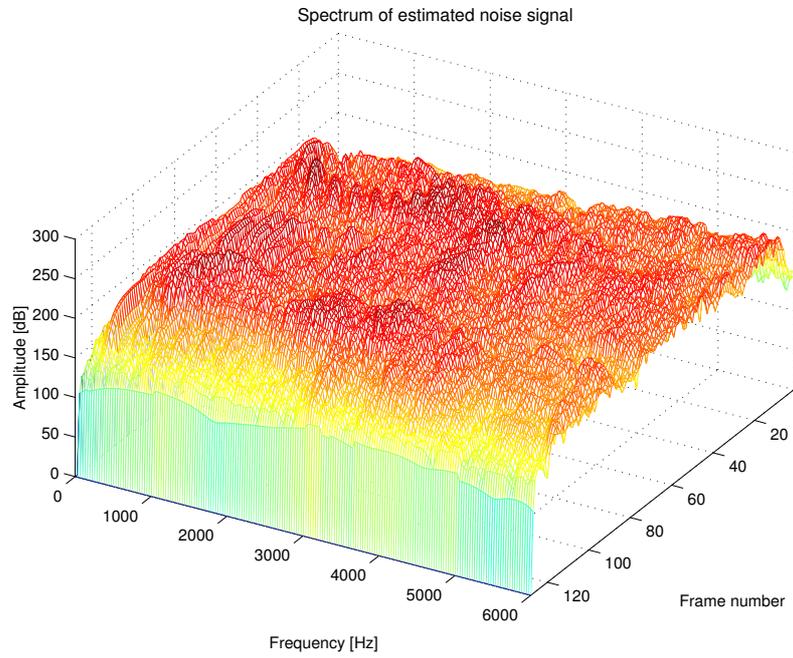


図 5.21: 籠らの提案法を用いて推定した雑音スペクトル (複数の音声)

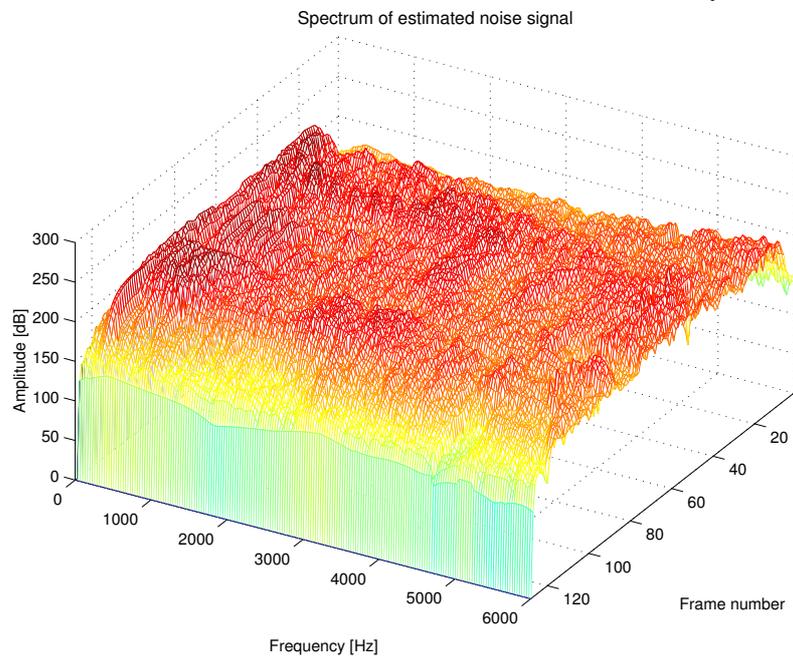


図 5.22: 本手法を用いて推定した雑音スペクトル (複数の音声)

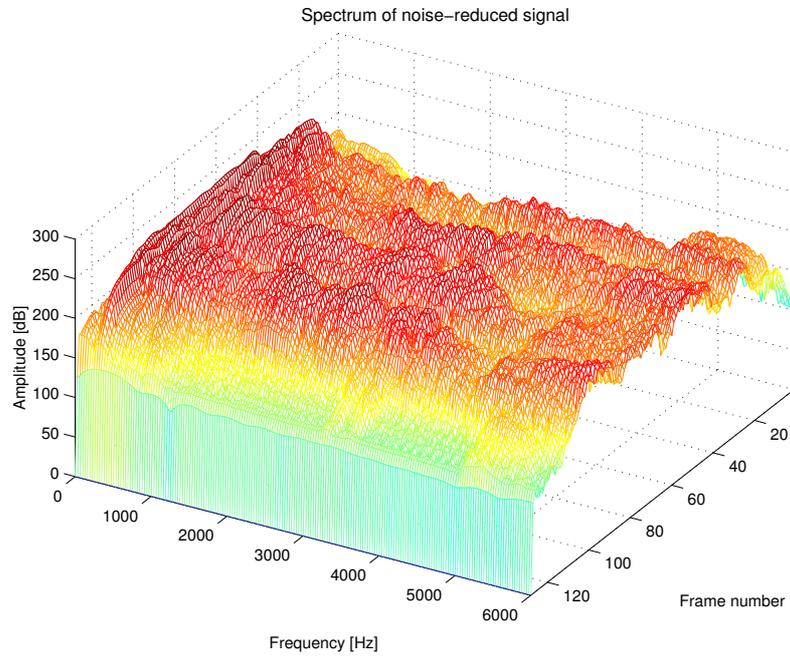


図 5.23: 籠らの提案法を用いて雑音除去した音声信号のスペクトル (複数の音声)

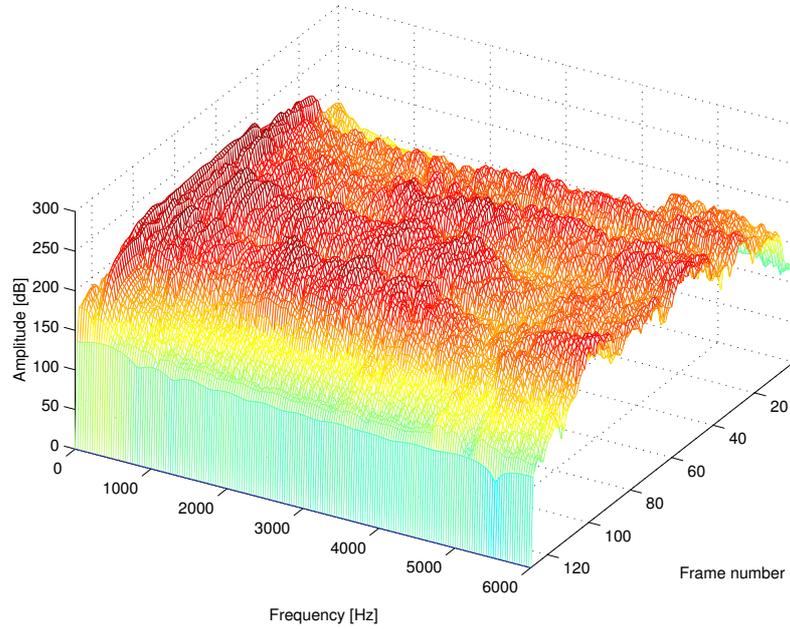


図 5.24: 本手法を用いて雑音除去した音声信号のスペクトル (複数の音声)

## 第6章 単語了解度試験

本章では、籠らの提案法と本論文で挙げている籠らの提案法における問題点を改善した雑音除去法とを単語了解度を尺度として比較、検討を行う。単語了解度試験は防音室内で行った。呈示音は親密度別単語了解度試験用音声データベースの音声に雑音を付加し、籠らの提案法または本論文で挙げている方法を用いて雑音除去を行った音声を両耳受聴により呈示した。回答方法は呈示した音声を音韻列通りにカナで書き取るというもので、得られた回答に対して単語として完全に一致した場合のみを正解、それ以外は不正解として算出した。被験者は正常な聴力を有する大学院生8名である。定常雑音・非定常雑音・複数の音声が付加した三つの場合について単語了解度試験を行った。

表 6.1: 実験に用いた機材

Digital Audio Interface	TOWNSHEND COMPUTER TOOLS	DAT-Link+
A/D D/A Converter	SONY	DTC-57ES
Headphone Amplifier	STAX	SRM-1/MK-2
Headphone	STAX	SR A

### 6.1 周波数帯域が重複している雑音に対する単語了解度試験

親密度別単語了解度試験用音声データベースに収録されている親密度 1.0 ~ 2.5 と 5.5 ~ 7.0 の男性話者 mya の音声データ ( a44, d08, n23 と a34, c42, o36 ) と女性話者 fhi の音声データ ( a20, c27, m43 と b34, g41, k44 ) をそれぞれ籠らの提案法と本手法の目的音とした。雑音として計算機で作成した 0 ~ 1 ( 雑音 1 ) , 2 ~ 4 ( 雑音 2 ) , 3 ~ 5 ( 雑音 3 ) kHz の定常な帯域雑音をそれぞれ 31, -50, 79 ° の到来方向から入力した。SNR は目的音に対してそれぞれ -10 dB である。

親密度 1.0 ~ 2.5 の音声データに対して、籠らの提案法の了解度はわずか 2 % で、本手法でも 7 % 足らずだった。これは籠らの提案法と本手法の両方で子音の聞き間違いが多かったためにこのような結果になったと思われる。親密度 5.5 ~ 7.0 の音声データに対しては籠らの提案法が 38 % なのに対して本手法では 66 % にも達していることがわかる。籠らの提案法では子音以外にも母音の聞き間違いも多かったが、本手法では母音の聞き間違

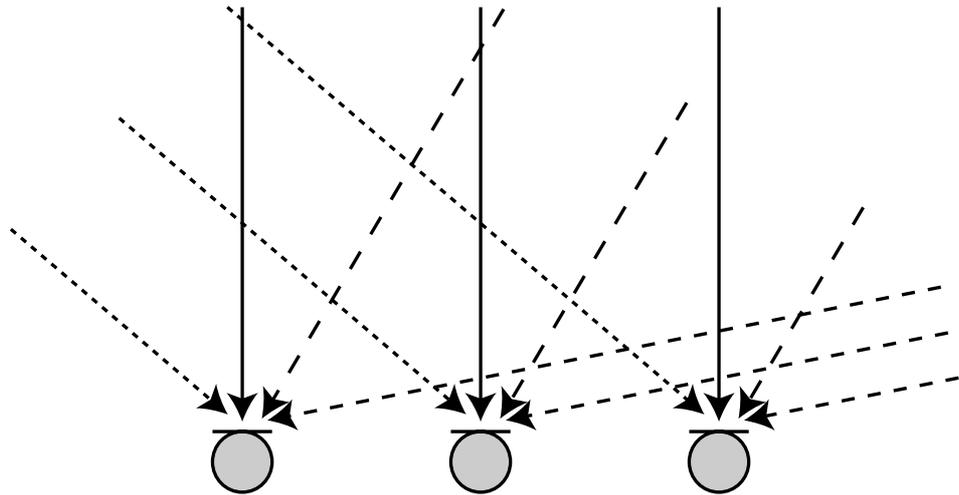


図 6.1: 目的音と雑音の到来方向 (各信号の到来方向は目的音 : 0 °、雑音 1 : 31 °、雑音 2 : -50 °、雑音 3 : 79 °)

表 6.2: 単語了解度試験の正解率 [%] (定常雑音付加)

	親密度 1.0 ~ 2.5	親密度 5.5 ~ 7.0	合計
籠らの提案法	2.38	38.10	20.24
本手法	7.14	66.67	36.90

いはそれほど多くないためにこのような結果になったと思われる。

## 6.2 非定常雑音に対する単語了解度試験

親密度別単語了解度試験用音声データベースに収録されている親密度 1.0 ~ 2.5 と 5.5 ~ 7.0 の男性話者 mya の音声データ (e41, i41, l24 と k23, n40, p24) と女性話者 fhi の音声データ (e09, f31, g49 と e22, g25, j38) をそれぞれ籠らの提案法と本手法の目的音とした。雑音として、計算機で作成した 0 ~ 1 (雑音 1 : 非定常), 2 ~ 3 (雑音 2 : 定常), 5 ~ 6 kHz (雑音 3 : 定常) の帯域雑音をそれぞれ 31, -50, 79 ° の到来方向から入力した。SNR は目的音に対してそれぞれ -10 dB である。

表 6.3: 単語了解度試験の正解率 [%] (非定常雑音付加)

	親密度 1.0 ~ 2.5	親密度 5.5 ~ 7.0	合計
籠らの提案法	14.29	54.76	34.52
本手法	66.67	59.52	63.10

親密度 1.0 ~ 2.5, 5.5 ~ 7.0 の音声データで籠らの提案法は 14, 54 % であるのに対して、本手法は 66, 59 % の了解度を達成している。非定常雑音が 1 kHz 以下の帯域に存在しているために籠らの提案法では殆ど除去することが出来ない。そのため、このような結果になったのではないと思われる。試験結果を見ても籠らの提案法の音声データに対する回答には無回答 (何を言っているのか聞き取れなかった) というものが幾つもあった。

## 6.3 複数の音声が存在する場合に対する単語了解度試験

親密度別単語了解度試験用音声データベースに収録されている親密度 1.0 ~ 2.5 と 5.5 ~ 7.0 の男性話者 mya の音声データ (e41, i41, l24 と k23, n40, p24) と女性話者 fhi の音声データ (e09, f31, g49 と e22, g25, j38) をそれぞれ籠らの提案法と本手法の目的音とした。男性話者 mya に対する雑音として、女性話者 fhi の音声データ (雑音 1 : 2q21), 男性話者 mis の音声データ (雑音 2 : 2n10), 女性話者 fto (雑音 3 : 2h41) をそれぞれ 31, -50, 79 ° の到来方向から入力した。SNR は目的音に対してそれぞれ 0 dB である。

親密度 1.0 ~ 2.5, 5.5 ~ 7.0 の音声データに対して籠らの提案法では了解度が 30 % にも達していない。本手法ではそれぞれ 42, 88 % にも達している。籠らの提案法では母音の第一フォルマントが存在する 1 kHz 以下の周波数帯域の雑音を除去することが出来ないために了解度が低くなっていると思われる。

表 6.4: 単語了解度試験の正解率 [%] (複数の音声付加)

	親密度 1.0 ~ 2.5	親密度 5.5 ~ 7.0	合計
籠らの提案法	28.57	26.19	27.38
本手法	42.86	88.10	65.48

## 6.4 考察

表 6.5: 単語了解度試験の正解率 [%] (合計)

	親密度 1.0 ~ 2.5	親密度 5.5 ~ 7.0	合計
籠らの提案法	15.08	47.73	27.38
本手法	45.29	82.37	55.16

籠らの提案法の試験結果は親密度の低い単語の場合、最も高い了解度が約 28 % 程度である。それに対して本手法では最も高い了解度が約 66 % である。籠らの提案法が最も高い了解度を出した複数の音声信号を付加した場合の試験でも、約 43 % の了解度になっている。総合的な音声了解度を見ても、籠らの提案法が約 27 % であるのに対して本手法では約 55 % にも達している。このことから 4 章で挙げた問題点を改善することで、音声了解度の大幅な改善を達成したと言える。特に 1 kHz 以下の周波数帯域における雑音除去精度改善の効果が大きく、籠らの提案法に比べて母音の聞き間違いを大幅に減らすことが出来た。

試験を行った際に、親密度の低い単語を親密度の高そうな単語に聞き間違えているというものが多々見られた。例えば「ワクデキ」を「バクゲキ」と回答しているものや、「クサモノ」を「ワカモノ」と回答しているものである。これは子音の聞き間違いが原因であると思われる。正解に似た回答(ここでは不正解)の中には「ワ」と「バ」、「カ」を「ハ」と聞き間違えているものが非常に多かった。このことから、音声の了解度改善には子音の聞き取り間違いを減らすような工夫が重要であると考えられる。

## 第7章 結論

### 7.1 本論文の要約

本論文では籠らの提案法を雑音除去に用いることで音声了解度の改善を目指した。しかしながら、籠らの提案法は自動音声認識器に対する雑音除去を目的としており、人間に聞かせることを目的としていない。籠らの提案法の問題点を改善し、音声了解度改善のための雑音除去を達成した。

本研究では低い周波数の雑音除去精度改善を中心に、

6章での結果から、様々な雑音に対して大幅な音声了解度改善を達成した。これにより、人間を対象にした雑音除去、補聴器などへの応用の可能性も示すことが出来た。

### 7.2 今後の課題

#### 子音の了解度改善

6章での音声了解度試験から得られた結果では、不正解の多くは子音の聞き間違いであった。そのため、更なる了解度改善を図るためには子音の了解度を向上させることが不可欠であると言える。高齢者の難聴者の多くは高い周波数の音を聞き取り難いという問題もあるため、子音の了解度向上は非常に重要な問題である。

#### ミュージカルノイズの軽減

ミュージカルノイズは定常雑音を除去した際によく聞こえる。雑音除去を行ったにも関わらず、新たな雑音としてミュージカルノイズが発生するという事は望ましくない。

# 謝辞

本研究を行うにあたり、終始多大なるご指導、ご鞭撻を賜りました北陸先端科学技術大学院大学情報科学研究科 赤木 正人 教授に深く感謝の意を表します。並びに本研究の過程において多くのご協力を賜りました、本学の教官の皆様にご心より感謝いたします。

本研究を進める上で、有益なご議論、ご助言をして戴いた ATR 音声言語通信研究所 水町 光徳 博士、音情報処理学講座の 鷓木 祐史 助手、伊藤 一仁 氏、石本 祐一 氏、西本 博則 氏に厚くお礼申し上げます。また日頃から有益な助言と激励を戴いた、音情報処理学講座の 2004 年度修了生、並びに諸先輩方、単語了解度試験に参加戴いた皆様に深く感謝致します。

最後に大学院における研究生生活を温かく見守って戴きました、両親、兄にご心から感謝致します。

## 参考文献

- [1] 水町 光徳, ”マイクロホン対を用いた雑音除去法に関する研究,” 北陸先端科学技術大学院大学 情報科学研究科 博士論文, 2000.
- [2] 籠 貴司, ”小規模マイクロホンアレーを用いた複数雑音にロバストな雑音除去法に関する研究,” 北陸先端科学技術大学院大学 情報科学研究科 修士論文, 2001.
- [3] 大賀寿郎, 山崎芳男, 金田豊, ”音響システムとデジタル処理,” 電子情報通信学会, 1995.
- [4] 小寺一興, ”補聴器フィッティングの考え方,” 診断と治療社, 1999.
- [5] 古井貞熙, ”音響・音声工学,” 近代科学社, 1998.
- [6] 電子情報通信学会, ”聴覚と音声,” コロナ社, 1994.