

Title	多点受音信号を用いた残響除去法に関する基礎的研究
Author(s)	水町, 光徳
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1039
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

修 士 論 文

多点受音信号を用いた残響除去法に関する基礎的研究

指導教官 赤木 正人 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

水町 光徳

1997 年 2 月 14 日

～ 要 旨 ～

現在、音声認識システムの実用化には大きな関心が寄せられている。しかし、現実には、雑音や残響が存在する実使用環境においてはその性能は著しく低下するという問題に直面している。そこで、本研究では、実環境における音声認識装置の実用化を支援できるような音声認識のプリプロセッサとしての残響除去法を提案する。

本研究で提案する残響除去法では、3本の無指向性マイクロホンを等間隔直線配置したマイクロホンアレーを使用する。そして、その特徴は、まず初めに3本のマイクロホンのうち両端の2本のマイクロホンで受音した信号に含まれる、到来方向に関する情報に基づいて中央のマイクロホン位置における残響成分を解析的に推定し、次に、3本のマイクロホンのうち残りの1本（中央）のマイクロホンで受音した信号、すなわち、直接音と残響成分との合成信号から、先に求めた残響成分のみを減算することにより残響除去を行なうことである。

そして、残響除去実験を行なって本残響除去法の有用性についても検討した。その結果、本方法は母音区間では有効であると言えるが、単語などの子音区間では、残響除去を行なうとスペクトルの歪みが増すという問題が存在することも分かった。

よって、この問題点については、今後、更なる検討が必要であるが、残響除去に関する他の研究と比較すると、本手法は音声認識の残響除去プリプロセッサとしての素質を十分に備えている。

目次

1	はじめに	1
1.1	本研究の背景	2
1.1.1	残響とは	2
1.1.2	残響除去の必要性	2
1.1.3	残響除去に関する研究	3
1.2	本研究の特徴	5
2	残響除去アルゴリズムの概要	6
2.1	残響除去を行なうために	6
2.2	残響除去アルゴリズムの概要	7
3	残響成分の抽出	8
3.1	残響成分抽出アルゴリズム	8
3.1.1	残響成分抽出アルゴリズムの概説	8
3.1.2	2 波形分離アルゴリズムとは	8
3.1.3	信号の到来方向推定	9
3.1.4	残響成分抽出のためのアルゴリズム	11
3.2	計算機シミュレーションによる信号抽出アルゴリズムの評価	18
3.2.1	計算機シミュレーションの目的	18
3.2.2	シミュレーションに使用するデータ	18
3.2.3	シミュレーション結果	18
3.2.4	考察	21
3.3	考察	21

4	残響成分の除去	22
4.1	残響成分除去アルゴリズム	22
4.1.1	残響成分除去アルゴリズムの概説	22
4.1.2	スペクトルサブトラクションとは	22
4.1.3	スペクトルサブトラクションの応用	23
4.2	考察	24
5	残響除去実験	25
5.1	実験の目的	25
5.2	残響のある音声データの作成	25
5.2.1	データ作成方法	25
5.2.2	環境条件	31
5.3	評価尺度の設定	32
5.4	実験結果	33
5.5	考察	38
6	おわりに	40
6.1	本残響除去法について	40
6.2	今後の課題	41
	謝辞	42
	参考文献	43
	学会発表リスト	45

目 次

2.1	残響のある環境とマイクロホンアレー配置	6
3.1	信号抽出のための解説図	14
32	マイクロホンアレーの指向特性(実線: 1 kHz、破線: 500 Hz、鎖線: 20 Hz、点線: 15 Hz)	16
33	マイクロホンアレーの指向特性(実線: 1 kHz、破線: 2 kHz、鎖線: 4 kHz、点線: 8 kHz)	16
34	帯域制限したピンクノイズの時間波形とパワースペクトル	19
35	焦点を信号到来方向(右 30°)に一致させた場合の信号抽出結果	19
36	焦点を右 60°と設定した場合の信号抽出結果	20
37	焦点を右 90°と設定した場合の信号抽出結果	20
51	実験データ作成時のブロックダイアグラム	26
52	クリーンな信号(/a/)	27
53	残響付加を行なった信号(/a/、中央のマイクロホンでの受音信号)	27
54	残響付加を行なった信号(/a/、左端のマイクロホンでの受音信号)	28
55	残響付加を行なった信号(/a/、右端のマイクロホンでの受音信号)	28
56	クリーンな信号(/bunri/)	29
57	残響付加を行なった信号(/bunri/、中央のマイクロホンでの受音信号)	29
58	残響付加を行なった信号(/bunri/、左端のマイクロホンでの受音信号)	30
59	残響付加を行なった信号(/bunri/、右端のマイクロホンでの受音信号)	30
510	データ作成に使用した室の平面図	31
511	データ作成に使用した室の残響特性	32
512	/a/ のクリーンな信号(上段)と残響除去の結果(下段)	35

5.13 /bunri/ のクリーンな信号（上段）と残響除去の結果（下段）	35
5.14 /a/ のクリーンな音声（上段） 残響のある音声（中段） 各フレームにお ける信号の到来方向推定結果（下段）	36
5.15 /a/ の信号到来方向の推定結果より作成したヒストグラム	36
5.16 /bunri/ のクリーンな音声（上段） 残響のある音声（中段） 各フレーム における信号の到来方向推定結果（下段）	37
5.17 /bunri/ の信号到来方向の推定結果より作成したヒストグラム	37

表 目 次

5.1 使用機材一覧	26
5.2 音声帯域セグメンタル SD の音声区間平均改善量	34
5.3 残響成分の到来方向の検出結果	34

第 1 章

はじめに

近年のデジタル信号処理技術の急速な発展に伴い、注目を集めているものの 1 つに音響情報処理がある。

そもそも、「音」は人間にとって欠くことのできない情報であり、その「音」に関する研究は、古くは紀元前の Pythagoras も行なっていたほど長い歴史を持っている [1]。そこで、「音」に関するこれまでの膨大な研究成果と高度なデジタル信号処理技術とを融合させることが可能になった現在、音響情報処理に関する研究に非常に高い期待が寄せられている訳である。

現在、音響情報処理のなかでも特に音声に関しては、音声認識、音声合成、音声の高能率符号化などについて盛んに研究が行なわれ、その研究成果は、徐々に我々の日常生活においても利用され始めている。ここで、音声認識に着目すると、音声認識の実用化により、機器（コンピュータ）と人間とのコミュニケーションを考えた場合の人に優しいインターフェイスが構築できるのではないかと期待できる。

また、音声認識装置の実用化は一部、実現されている。しかし、これまでの研究は、雑音や残響が全く存在しない理想的な環境において使用することを前提にした認識アルゴリズムに関する研究がほとんどであった。そのため、雑音や残響が存在する実使用環境においては、音声認識装置の性能（認識率）が大幅に低下するという報告がある [8]。そこで、音声認識アルゴリズムに関する研究と同様に、雑音や残響の周囲環境の影響を少なくするための研究も重要となってきた。

よって、本研究では、実環境における音声認識装置の実用化を支援できるような残響除去法を提案し、その有用性について検討している。

1.1 本研究の背景

1.1.1 残響とは

残響 (reverberation) とは、その名のとおり音の響きが残る現象である。

この残響という現象は古くから知られており、紀元前に建設されたギリシャの野外劇場には数多くの壺を埋めることにより、残響を調節していたことが記録に残っている [2]。そして、日本の能楽堂の床下に埋められている瓶や北欧の中世の教会の壁に埋められている瓶や壺は、現在もそのままの形で残っており、これらは主に残響特性を調整する目的で設置されたと考えられている [5]。

また、残響現象を定量的に研究したのは W. C. Sabine (1868 ~ 1919) が最初であり、彼が確立した残響理論 (1895) は現在の室内音響学の礎となっている。この残響理論の特徴は、室内の響きを音源停止後、室内の音エネルギー密度が定常状態の 100 万分の 1 (音圧レベルで -60 dB) に減衰するまでの時間を残響時間という簡単に計測可能な物理量で表したことである。

この残響時間は、室内音響特性を表す代表的な客観的評価値であり、室の主観的な音響効果とよく対応することから、現在でも重要な尺度として用いられている。

1.1.2 残響除去の必要性

残響は、我々の日常生活において非常に重要な役割を果たしており、残響が全く存在しない自由音場 (例えば無響室) において発声された音声は非常に不自然に感じる。

しかし、機械による音声認識という観点からは、残響のある環境において発声された音声は、残響による歪みの影響で認識性能が大幅に低下することが報告されている [8]。これは、現行の音声認識システムが、入力音声と標準パターンとのパターンマッチングを行なうことにより認識を行っており、その標準パターンの作成にクリーンな音声を用いているためである。

そこで、標準パターンを残響のある音声を用いて作成するという研究も行なわれている [10]。その手法では、音声認識システムと話者間の室内伝達特性が変化しないことを前提にしているため、ある特定の音場のみでしか有効ではない。また、同一の音場であっても気温や湿度が変化することにより室内伝達特性は変化するため [11]、その手法はあまり望ましくない。従って、本論文では他の残響除去法について検討する。

なお、残響時間と音声認識性能との関係を調べる場合には、残響時間と話者-マイクロホン間距離、そして、マイクロホンの種類（指向特性）の3つを変数として調査する必要がある。文献[9]によると、単語音声認識実験の結果では、残響がない無響室で收音した場合であれば、マイクロホンの種類と話者-マイクロホン間距離に関係なくほぼ100%の認識性能が得られるのに対し、残響のある環境で收音すると、残響時間が長く、收音距離が大きく、マイクロホンの指向性が無指向性に近くなるほど認識性能が大きく低下することが報告されている。例えば、残響時間が1.7 secの室内で無指向性マイクロホンを使用し、話者-マイクロホン間距離を1 mとして收音すると、認識率は約60%にまで低下する。

1.1.3 残響除去に関する研究

一般に、多点受信信号を用いた残響除去法に関する研究の代表例として、2点受信信号の相関値を利用して残響除去を行なう振幅圧縮法、また、室内伝達特性の逆フィルタを計算し、それを受音信号に畳み込むことにより残響除去を行なう逆フィルタ法がある。そこで、それらの特徴について以下で概説する。

a. 振幅圧縮法

多点受信信号を用いた残響除去に関する研究の第一歩は、1970年代にAllenらが提案した、2つのマイクロホンで受信した信号間の相関に着目した残響除去法[12]である。

この手法の特徴は、マイクロホンで受信される信号を直接音と残響成分とに大別し、直接音に関しては2つのマイクロホンで受信される信号は若干、位相はずれるものの波形としてはほぼ同じであるためそれらの相関は高く、2つのマイクロホンで受信した信号に含まれる残響成分に関してはそれらの波形はかなり異なるため、相関が低いという仮定を設けたことである。

そこで、2つのマイクロホンで受信した信号をそれぞれ複数の周波数帯域に分割し、その各帯域ごとに相関値を計算し、相関値に比例したゲインを付加する。すなわち、相関値の低い帯域には仮定より残響成分が存在すると考えられ、その帯域の信号（残響成分）に小さいゲインを付加することにより、残響成分のパワーを縮小させることが可能になる。

なお、この手法では、残響成分は良好に低減されるが、演算が非線形であるため残留雑音が耳障りなものとなり聴感上の印象はそれほど向上しない。

b. 逆フィルタ法

逆フィルタを用いた残響除去法に関する研究は、振幅圧縮法が発表された数年後に始められた。なお、逆フィルタ法の研究には、振幅圧縮法を提案した Allen も従事していた。

しかし、室内のインパルスレスポンスは一般に非最小位相であり、そのままでは室内伝達関数の逆フィルタを作成することは不可能であるため、Neyley と Allen は、それを最小位相化することにより逆フィルタを作成して残響除去を行なったが良好な結果は得られなかった [13]。その後、三好らによって、複数のマイクロホンを使用し、より高度な信号処理技術を用いることにより、室内伝達関数が非最小位相であっても原理的にはその逆フィルタが作成できることが示された [14]。

そこで、Wang ら [15]、山田ら [16] は、複数のマイクロホンと多チャンネルサブバンド信号処理を利用して室内伝達関数の逆フィルタを作成し、それを用いて残響除去を行なう手法を提案し、残響除去実験を行なってその有用性を示した。これは、マイクロホンの位置が異なれば非最小位相の零点を含んだ周波数帯域も異なることに着目し、各帯域において複数の伝達関数に対して 2 乗最小誤差で原信号を回復するフィルタを計算し、最も高精度のフィルタを利用することにより残響除去を行なう手法である。

この手法を用いると、高精度の残響除去が可能であるが、周りの環境が変化する度に逆フィルタを作成する必要がある。ここで、実環境の伝達特性は時々刻々変化する [11] ことを考慮すると、逆フィルタ法の実環境における有効性については疑問が残る。

1.2 本研究の特徴

本研究では、残響のある環境においては、現行の音声認識システムは大幅に性能が低下するという現状を鑑み、残響のある一般的な環境においても音声認識システムを広く使用できるようにするため、残響による音声の歪みを回復させることを目標とする。

なお、本研究では、音声が入力される前に残響を除去することを考え、音声認識アルゴリズムに依存しない音声認識のプリプロセッサとしての残響除去アルゴリズムを構築することが目的である。

そこで、3本のマイクロホンを等間隔直線配置したマイクロホンアレーを用いて、信号の到来方向に関する情報、すなわち、各マイクロホンへの信号の到来する時間差情報を利用した残響除去法を提案する。

そして、本研究の特徴は、残響除去という処理を残響成分を推定する過程と残響成分を除去する過程に大別し、その前者ではマイクロホンアレーの両端のマイクロホンで受信した信号を用いて幾何音響学的に、後者ではマイクロホンアレーの残りの1本(中央)のマイクロホンで受信した信号を用いて統計音響学的に考えることにより複雑な残響除去問題を簡略化して解いている点である。

第 2 章

残響除去アルゴリズムの概要

2.1 残響除去を行なうために

本研究では、図 2.1 に示すような 3 本の無指向性マイクロホン等を等間隔直線配置したマイクロホンアレーを使用する。

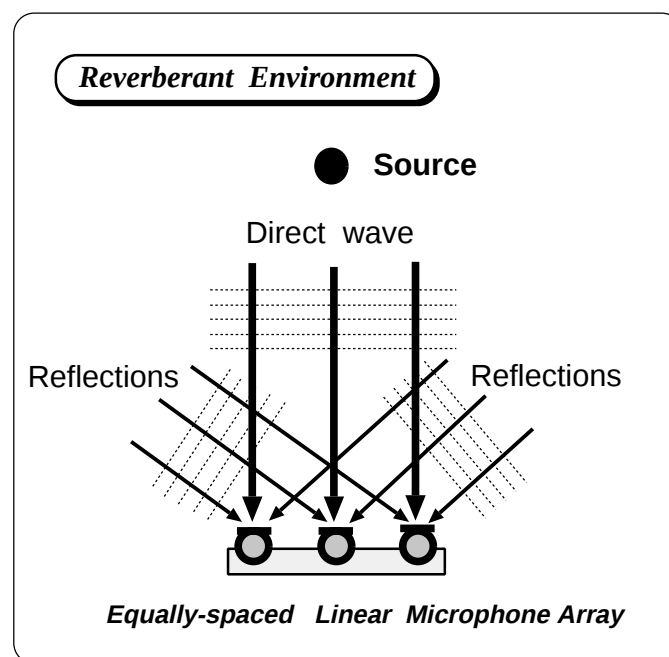


図 2. 1: 残響のある環境とマイクロホンアレー配置

ここで、一般に残響除去問題は非常に複雑であるため、問題の簡略化を行なうために以下のような制約を設ける。また、本残響除去法が音声認識の残響除去プリプロセッサという位置づけをしているのには、問題を簡略化するという目的も含まれている。

本残響除去法における制約：

- 音波は平面波として伝搬すると仮定する
- マイクロホンの音響中心を含む水平面方向の残響除去しか考えない
- 直接音と同一方向、あるいは、それに対して 180° 方向から到来する残響成分については考慮しない
- 振幅スペクトルについてのみ歪みを回復することを考え、位相スペクトルの歪みの回復については考慮しない（音声認識では振幅スペクトルのみで十分である）

2.2 残響除去アルゴリズムの概要

ここで、本研究で提案した残響除去アルゴリズムについて概説する。

まず初めに、3本のマイクロホンのうち両端の2本のマイクロホン（マイクロホン対）で受音した信号を利用し、中央のマイクロホン位置における残響成分を解析的に推定する。

次に、3本のマイクロホンのうち残りの1本（中央）のマイクロホンで受音した信号、すなわち、直接音と残響成分との合成信号から、先に求めた残響成分のみを減算することにより残響除去を行なう。

これらの詳細については、残響成分の推定過程については第3章、残響成分の除去過程については第4章で詳しく解説する。

第 3 章

残響成分の抽出

3.1 残響成分抽出アルゴリズム

3.1.1 残響成分抽出アルゴリズムの概説

本研究では、残響成分の抽出に、今田らによる 2 波形分離アルゴリズム [6][7] を応用する。なお、2 波形分離アルゴリズムの詳細については次節で紹介するが、その特徴は、正中面方向とそれ以外の 1 方向から到来する 2 つの信号のみが存在するような状況において、2 本のマイクロホンで構成されるマイクロホン対で受音した信号より、正中面方向から到来する信号を消去し、もう一方の信号のみを抽出することである。ただし、抽出したい信号の到来方向は予め与える必要があり、本論文では、その方向をマイクロホンアレーの焦点と呼ぶ。

そこで、本研究では、今田らの 2 波形分離アルゴリズムに改良を加え、正中面方向から直接音が到来すると仮定した場合に、それ以外の全方向から到来する残響成分を抽出するための残響成分抽出アルゴリズムを作成した。

3.1.2 2 波形分離アルゴリズムとは

今田らは、2 つの音響信号が正中面方向とそれ以外の方向から到来するような場合に、それぞれの信号が 2 つのマイクロホンへ到来する時間差を利用し、正中面方向以外から到来する信号を抽出する 2 波形分離アルゴリズムを提案した [6][7]。

そのアルゴリズムでは、左右のマイクロホンで受音した信号をそれぞれ $l(t)$ 、 $r(t)$ とすると、それらを Wavelet 分析系を用いて多数の周波数帯域に分割し、その各周波数帯域ごとに目的信号を抽出する。このとき、中心角周波数が ω_i である帯域における目的信号の推定値 $\hat{s}_i(t)$ は

$$\hat{s}_i(t) = \frac{\{l_i(t) - l_i(t - 2\tau)\} - \{r_i(t) - r_i(t - 2\tau)\}}{2(1 - \cos 2\omega_i\tau)} \quad (3.1)$$

により求められる。

ここで、上式の 2τ は、抽出したい目的信号が左右のマイクロホンへ到来する際の時間差であり、 $l_i(t)$ 、 $r_i(t)$ はそれぞれ $l(t)$ 、 $r(t)$ の中心角周波数が ω_i である帯域における成分である。なお、この場合の $\hat{s}_i(t)$ は左のマイクロホン位置における目的信号の推定値となっている。

そして、この手法を用いることにより、式 (3.1) の分母が 0 に近くなり $\hat{s}_i(t)$ が計算できない (発散する) ような周波数成分を除いては、かなりの高精度で目的信号を抽出できることが報告されている。

3.1.3 信号の到来方向推定

信号抽出を行なう際には、抽出したい目的信号の到来方向を推定しなければならない。

なお、信号の到来方向推定に関しては、本研究では既存の手法の中から、2つのマイクロホンで受音した信号に含まれる各マイクロホンへの到来時間差に関する情報を利用する方向推定法を応用する。これを選択した理由は、信号抽出を行なう際にも信号の到来時間差を利用するためである。

ここで、信号の到来時間差に基づく方向推定法とは、ある音波が平面波として2つのマイクロホンに到来した場合に、それらの受音信号を $x_1(t)$ 、 $x_2(t)$ とすると

$$x_2(t) = x_1(t - \tau_s) \quad (3.2)$$

の関係が成り立ち、その時間差 τ_s を $x_1(t)$ と $x_2(t)$ との相互相関関数に基づき導出する手法である。ここで、実際の計算方法について簡単に説明する。まず、 $\phi_{11}(\tau)$ を $x_1(t)$ の自己相関関数、 $\phi_{12}(\tau)$ を $x_1(t)$ と $x_2(t)$ との相互相関関数とすると

$$\begin{aligned}
\phi_{12}(t) &= E[x_1(t) \cdot x_2(t + \tau)] \\
&= E[x_1(t) \cdot x_1(t + \tau - \tau_s)] \\
&= \phi_{11}(\tau - \tau_s)
\end{aligned} \tag{3.3}$$

の関係が成り立つ。但し、 $E[\cdot]$ は期待値を表す。

次に、自己相関関数 $\phi_{11}(\tau)$ は $\tau = 0$ で最大値をとるため、式 (3.3) より $\phi_{12}(\tau)$ が $\tau = \tau_s$ で最大値をとることが分かる。従って、 $x_1(t)$ と $x_2(t)$ との相互相関関数を計算し、その最大値を与える τ を求めることが τ_s を求めることに相当する。

よって、音波が2つのマイクロホンに到来する時間差 τ_s さえ分かれば、音速とマイクロホン間距離に関する情報を用いて信号の到来方向を知ることができる。

本研究では、残響成分抽出アルゴリズムで構成されるマイクロホンアレーの焦点を定める際に、上記の方向推定アルゴリズムを利用する。

その焦点の決定方法とは、まず始めに、マイクロホンアレーの両端のマイクロホンで受信した信号をそれぞれ短時間フレームに分割し、その各フレーム内のデータを用いて方向推定を行なう。この場合、各フレームにおける方向推定の結果としては、そのフレーム内において最大のパワーを有する信号の到来方向が得られるため、直接音と残響成分が共存するようなフレームにおいては、直接音の到来方向（本論文では正中面方向）となることが容易に想像できる。

そこで、入力データの全フレームにおいて方向推定を行ない、その全フレームの方向推定結果を用いてヒストグラムを作成し、正中面方向付近（正確には左右 5° 程度）を除いて最大の頻度をとる方向、すなわち、残響成分の中で最もパワーの大きい成分の到来方向を残響成分の抽出過程におけるマイクロホンアレーの焦点とする。

この残響成分の到来方向推定の精度については、第5章において、残響除去実験の実験結果として残響成分の到来方向の推定結果も併せて表示しているため、この手法の妥当性に関する考察はそちらに譲る。

3.1.4 残響成分抽出のためのアルゴリズム

まず、今田らによる2波形分離アルゴリズムを改良し、より高精度の信号抽出(残響成分の抽出)を目指す。本研究では、以下の4点について改良を加えた。

1. 時間領域の信号を周波数領域に変換する際に、DWT(離散 Wavelet 変換)ではなく DFT(離散 Fourier 変換)を利用し、計算精度の向上と計算時間の短縮を図る
2. フレーム処理を導入することにより、時間分解能を向上させる
3. 位相の補正を行なう(マイクロホン対の中央の位置における信号を抽出する)
4. 計算不可能な周波数における計算方法についての再考

ここで、上記の改良点についてのより詳細な説明を加える。

1. については、今田らは 12ch. の Wavelet 分析系により、入力信号を 12 個の狭帯域信号に分割し、それぞれの周波数帯域において2波形分離を行ない、その結果を Wavelet 合成系により合成することにより目的信号の抽出を行っていた。

しかし、12ch. の Wavelet 分析・合成系の計算には莫大な時間が必要となる。そして、Wavelet 合成系を用いて複数の周波数帯域に分割することの利点は、特定周波数帯域における振幅情報と位相情報を正確に得ることができる点であるが、今田らの2波形分離アルゴリズムでは、入力信号の位相情報は重要であるが、その特定周波数帯域における位相情報は不要であるため、DWT ではなく DFT で十分であることが分かった。

また、この改良により、計算量の劇的な減少は容易に想像できるが、その恩恵を受け、DFT の計算を行なうためのポイント数(帯域分割する際の帯域の数)を 12 点よりも大幅に増やすことが可能になり、周波数分解能もそれだけ向上するという利点が生じる。

2. については、今田らは受信信号を時系列上で分割せずにまとめて目的信号の抽出を行なうことを考えていた。しかし、それでは音源が移動する場合には全く対応できない。

よって、音源の移動に対してもそれを追従し、目的信号を正確に抽出できるように短時間フレームごとに信号抽出を行なうように改良した。ちなみに、本研究では、フレーム長を 512pt (サンプリング周波数 48kHz で 10.7ms) フレーム周期を 256pt (同様に 5.3ms) フレームごとにデータを切り出す際に使用する窓関数は Hamming 窓を採用した。これらのパラメータに関しては、抽出する目的信号としてピンクノイズと音声(単

母音)を用いて、各データに対してフレーム長を 128 pts、256 pts、512 pts、1024 pts、2048 pts の 5 通り、フレーム周期を $1/2$ フレーム、 $1/4$ フレーム、 $1/8$ フレームの 3 通り、窓関数を blackman 窓、bartlett窓、hamming 窓、hanning 窓、三角窓、矩形窓 の 6 通りの全ての組合せ (180 通り) に対して信号抽出実験を行ない、音声を用いた場合の信号抽出精度が最も高かった組合せである。なお、ピンクノイズのような定常的な信号の抽出を行なう場合には、フレーム長を 1024 pts に設定することが望ましいことも分かった。

但し、本研究では、現段階では音源の移動までは考慮しておらず、音源の移動に追従し、その直接音のみを抽出できるような残響除去アルゴリズムの構築は、将来的な課題であることを付記しておく。

3. については、今田らは式 (3.1) に示したように、受信信号そのものとそれに遅延 (2τ) を与えた信号とを用いて目的信号の抽出を行っていた。その結果、抽出された信号は、左右どちらかのマイクロホン位置における推定値となる。

そこで、本研究では、目的信号の抽出を行なう際に、受信信号そのものは使用せず、受信信号にプラス方向とマイナス方向の遅延 (τ) を与えたものを使用するように改良することにより、2つのマイクロホンの中央の位置における推定値を求めることを考える。これは、残響成分の推定を行なった後、それを除去する際に2つのマイクロホンの中央の位置における推定値を求めることが必要になるためである。従って、これは信号抽出過程において、位相の補正を行なうための改良である。

4. については、本研究では、受信信号から抽出した残響成分を受信信号から減算した結果が目的信号となる。ここで、残響成分を抽出する過程において、式 (3.1) の分母が小さくなると、ある周波数成分のレベルが極端に大きくなる。その結果、減算を行なうことが不可能になるか、たとえ減算を行なうことは可能であったとしても、その周波数成分のレベルが極端に小さくなる (情報が欠落する) ことが考えられる。

そこで、この問題に対処するために以下の3種の対策を試みた。

- a. 信号抽出を行わない (左右のマイクロホンで受信した信号の加算平均をとる)
- b. 分母を定数で置換する
- c. 信号の抽出結果を 0 置換する

最初に検討したのが、a. の左右のマイクロホンで受信した信号の加算平均をとる方法であるが、時間領域において2つの信号を加算すると、受信信号に含まれる特定の信号がその位相差により完全に相殺される可能性がある。また、実際にピンクノイズを目的信号として信号抽出の計算機シミュレーションを行なった結果、他の手法に比較して良好な結果が得られなかったことも分かったので本研究ではこの手法は採用しなかった。

次に、b. の分母を定数で置換する方法と c. の信号の抽出結果を0置換する方法に関しては、計算機シミュレーションを行なった結果、ほぼ同程度の抽出精度が得られたので、今田らも考察していた c. の信号の抽出結果を0置換する方法を採用した。

そこで、残響成分の抽出過程において、式(3.1)の分母が0¹よりも小さくなるような周波数では、式(3.1)に従って計算を行わず、目的信号の該当周波数帯域における成分を0として置換する。実際に計算を行なった結果、0置換を行なう周波数帯域の数は、音声帯域(15 Hz ~ 6 kHz)の1、2%程度であることを確認した。

なお、この0置換を行なうという手法は、先にも述べた通り、今田らが既に提案していた。しかし、今田らは信号の抽出のみを考えていたので最終結果の特定成分に欠落が生じていた。これに対し、本研究では非目的信号(残響成分)の抽出結果の特定成分を欠落させ、最終結果では特定成分に関しては、受信信号をそのまま残すという違いがある。

従って、残響除去という観点からは、直接音と残響成分とはその性質が非常に良く類似していることを考慮すると、この改良は有効であると考えられる。

以上の改良点を踏まえ、残響成分抽出アルゴリズムを以下のように定式化する。

まず、図31に示すように、正中面方向から直接音 $x(t)$ 、正中面方向に対して角度 θ 方向から残響成分の一部 $y(t)$ が到来するとした場合、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信される信号をそれぞれ $l(t)$ 、 $c(t)$ 、 $r(t)$ とすると

$$\begin{cases} l(t) = x(t) + y(t - \delta) \\ c(t) = x(t) + y(t) \\ r(t) = x(t) + y(t + \delta) \end{cases} \quad (3.4)$$

¹この閾値は、これを0.0001~1までの様々な値に変化させて後述の残響除去実験を行ない、最も望ましい実験結果が得られた場合の値として経験的に定めたものである。

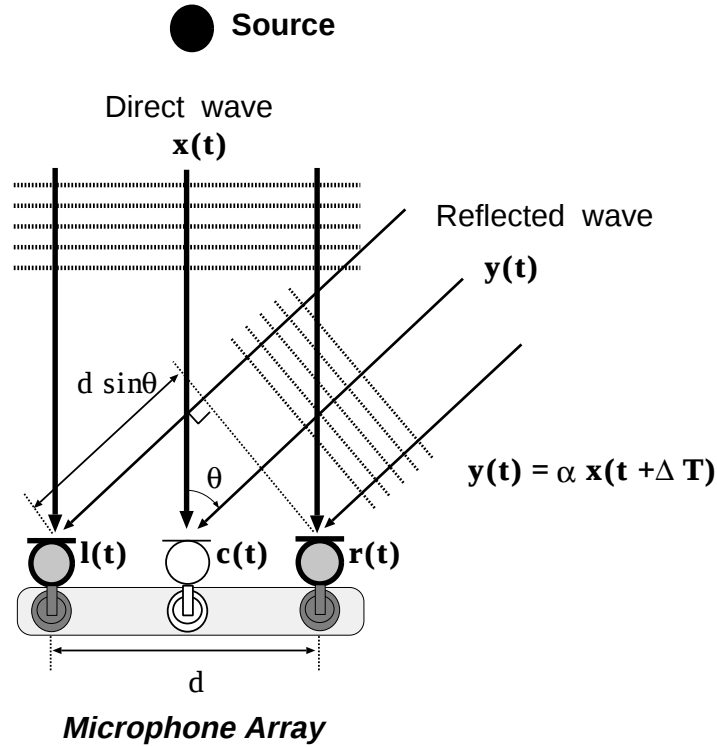


図 3.1: 信号抽出のための解説図

となる。ここで、上式の δ は抽出したい目的信号 $y(t)$ が左右両端のマイクロホンに到来する際の時間差の $1/2$ であり、 $y(t)$ の到来方向 θ と δ との間には、左右両端のマイクロホン間距離を d m、音速を v m/sec とすると

$$2\delta = \frac{d \sin \theta}{v} \quad [\text{sec}] \quad (3.5)$$

の関係が成り立つ。

このとき、 $l(t)$ 、 $r(t)$ 、 $x(t)$ 、 $y(t)$ をフーリエ変換したものをそれぞれ $L(\omega)$ 、 $R(\omega)$ 、 $X(\omega)$ 、 $Y(\omega)$ とすると、式 (3.4) より

$$\begin{cases} L(\omega) = X(\omega) + Y(\omega) e^{-j\omega\delta} \\ R(\omega) = X(\omega) + Y(\omega) e^{j\omega\delta} \end{cases} \quad (3.6)$$

となる。

ここで、抽出したい目的信号 $y(t)$ と左右両端のマイクロホンで受信信号 $l(t)$ 、 $r(t)$ との関係を、正中面方向から到来する非目的信号 $x(t)$ を用いずに表現することを考える。

そこで、 $l(t)$ 、 $r(t)$ をそれぞれ $\pm\tau$ だけずらした信号を利用し

$$g(t) = \{l(t + \tau) - l(t - \tau)\} - \{r(t + \tau) - r(t - \tau)\} \quad (3.7)$$

と新たに $g(t)$ を定めると、この $g(t)$ をフーリエ変換した $G(\omega)$ は

$$G(\omega) = \{L(\omega)e^{j\omega\tau} - L(\omega)e^{-j\omega\tau}\} - \{R(\omega)e^{j\omega\tau} - R(\omega)e^{-j\omega\tau}\} \quad (3.8)$$

となる。そこで、上式に式 (3.6) の $L(\omega)$ 、 $R(\omega)$ を代入してまとめると

$$G(\omega) = 4Y(\omega) \sin \omega\delta \sin \omega\tau \quad (3.9)$$

となり、目的信号の実際の到来方向 (時間差 2δ 方向) が分かれば、マイクロホンアレーの焦点 (目的信号が左右両端のマイクロホンに到来する際の時間差 2τ 方向) をその方向に一致させる、つまり、 $2\tau = 2\delta$ とすることにより、目的信号 $y(t)$ の中央のマイクロホン位置 (左右両端のマイクロホンの中央の位置) における推定値 $\hat{y}(t)$ は

$$\hat{y}(t) = \begin{cases} \text{IFFT} \left[\frac{G(\omega)}{4 \sin^2 \omega\tau} \right], & 4 \sin^2 \omega\tau > 0.02 \\ 0, & \text{other wise} \end{cases} \quad (3.10)$$

により計算できる。

ここで、式 (3.10) を計算するためには、マイクロホンアレーの焦点 (2τ) を定めなければならないし、更に式 (3.10) で計算できるのは、マイクロホンアレーの焦点方向から到来する残響成分のみであるにも関わらず、何故、式 (3.10) により残響成分が抽出できるのかという疑問が生じる。そこで、これらに対する解答として以下のような検討を加えた。

まず、上記の信号抽出アルゴリズムにより構成されるマイクロホンアレーを用いて信号抽出を行なった場合に、マイクロホンアレーの焦点方向とそれ以外の方向でどのくらいの感度の差があるのか、すなわち、マイクロホンアレーの指向特性についての調査を行なった。図 3.2 と図 3.3 に、右 30° 方向に焦点を定めた場合のマイクロホンアレーの指向特性を示しているが、横軸は信号が 2 つのマイクロホンへ音波が到来する際の時間差 (正中間方向からの角度)、縦軸はマイクロホンアレーの感度を表す。

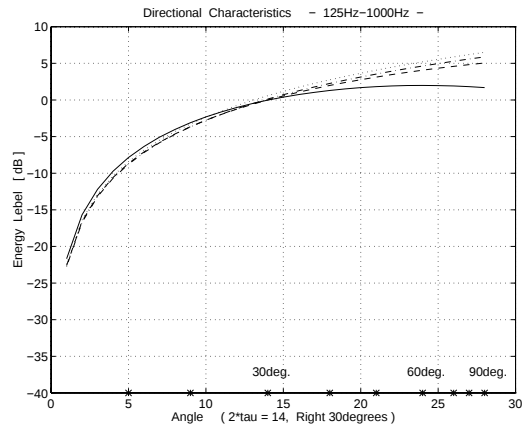


図 3.2: マイクロホンアレーの指向特性 (実線 : 1 kHz、破線 : 500 Hz、鎖線 : 250 Hz、点線 : 125 Hz)

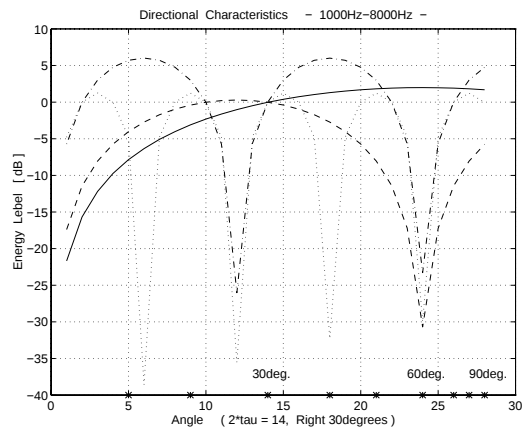


図 3.3: マイクロホンアレーの指向特性 (実線 : 1 kHz、破線 : 2 kHz、鎖線 : 4 kHz、点線 : 8 kHz)

ここで、マイクロホンアレーの焦点方向 (30° 方向、 $2\tau = 14$ pts) では、全周波数帯域においてマイクロホンアレーの感度が一致しており、その方向から到来する信号の正確な抽出が可能であることが分かる。

なお、マイクロホンアレーの指向特性に関しては、式 (3. 10) により目的信号を計算する際に、その到来方向 (時間差 2δ 方向) に対してマイクロホンアレーの焦点を時間差 $2\tau'$

方向 ($2\tau' \neq 2\delta$) と誤って設定した場合に生じる誤差 $e(\omega)$ として

$$e(\omega) = \frac{\sin^2 \omega \tau'}{\sin^2 \omega \tau} \quad (3.11)$$

より計算できる。図 3.2 と図 3.3 に示しているマイクロホンアレーの指向特性は、式 (3.11) の 2τ を 14 pts (信号が 30° 方向から到来する場合に相当) とし、 $2\tau'$ を 1 ~ 28 pts ($2^\circ \sim 90^\circ$ 方向) に変化させた場合の各周波数成分に対する抽出誤差を表したものである。

そして、図 3.2 と図 3.3 から分かるように、焦点方向以外では、低周波数成分に関しては死角方向 (正中面方向) 付近以外をほぼ等感度で抽出できるが、高周波数成分に関しては高周波数帯域になるほど指向特性に凸凹が生じることが分かる。つまり、高周波数成分に関しては、到来方向によってはかなり大きな抽出誤差が生じることが予想される。

また、残響成分は、直接音の到来方向以外から到来する全ての信号の総和と考えてもよいので、死角以外の方向に関しては全て等感度で信号を抽出することができれば、正確に残響成分を抽出できることになる。しかし、実際にはそれが不可能であるため、本残響除去法では、前述の通り残響成分の中で最もパワーの大きい成分が到来する方向を自動推定し、その方向にマイクロホンアレーの焦点を定めることにより、パワーが最大となる残響成分を忠実に抽出することを目指す。

従って、本研究では、マイクロホンアレーの焦点を残響成分の中で最もパワーの大きい成分が到来する方向と定め、式 (3.10) により抽出した信号を中央のマイクロホン位置における残響成分と考える。

3.2 計算機シミュレーションによる信号抽出アルゴリズムの評価

3.2.1 計算機シミュレーションの目的

信号抽出の計算機シミュレーションを行なう目的は、信号抽出アルゴリズムの基本性能を調査することである。そこで、信号は1方向のみから到来するものとして計算機シミュレーションを行なう。

なお、マイクロホンアレーの焦点を信号の到来方向と一致させた場合だけでなく、敢えて、マイクロホンアレーの焦点を信号の到来方向とずらした場合についても信号抽出の計算機シミュレーションを行ない、焦点がずれた場合の抽出結果に生じる歪みについても検討する。

3.2.2 シミュレーションに使用するデータ

信号抽出の計算機シミュレーションには、音声帯域(125 Hz ~ 6 kHz)に帯域制限したピンクノイズを使用する。これは、本手法では音声信号の残響成分さえ抽出できれば良いためであり、音声の周波数特性に類似したピンクノイズ(信号長: 100msec)を、信号通過帯域 125Hz ~ 6 kHz /3 dB のバンドパスフィルタに通して作成した。この信号の時間波形とパワースペクトルとを図 3.4 に示している。

そして、シミュレーションでは、2つのマイクロホン間距離を 20cm、帯域制限したピンクノイズを発する音源は、正中面方向に対して右 30°方向で2つのマイクロホンの中央点から 2.5m の距離に存在し、音速は 340m/sec であると想定する。

3.2.3 シミュレーション結果

まず、マイクロホンアレーの焦点を信号の到来方向(右 30°)と一致させた場合の信号抽出結果を図 3.5 に、次に、マイクロホンアレーの焦点を信号の到来方向とずらした場合については、焦点を右 60°、右 90°と設定した場合の信号抽出結果をそれぞれ図 3.6 図 3.7 に示している。

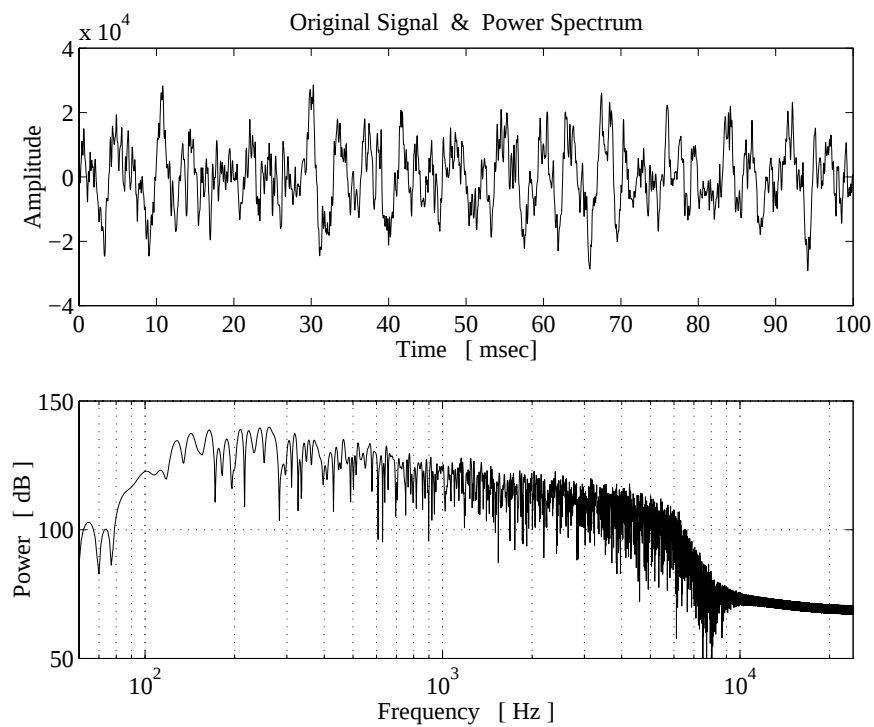


図 3.4: 帯域制限したピンクノイズの時間波形とパワースペクトル

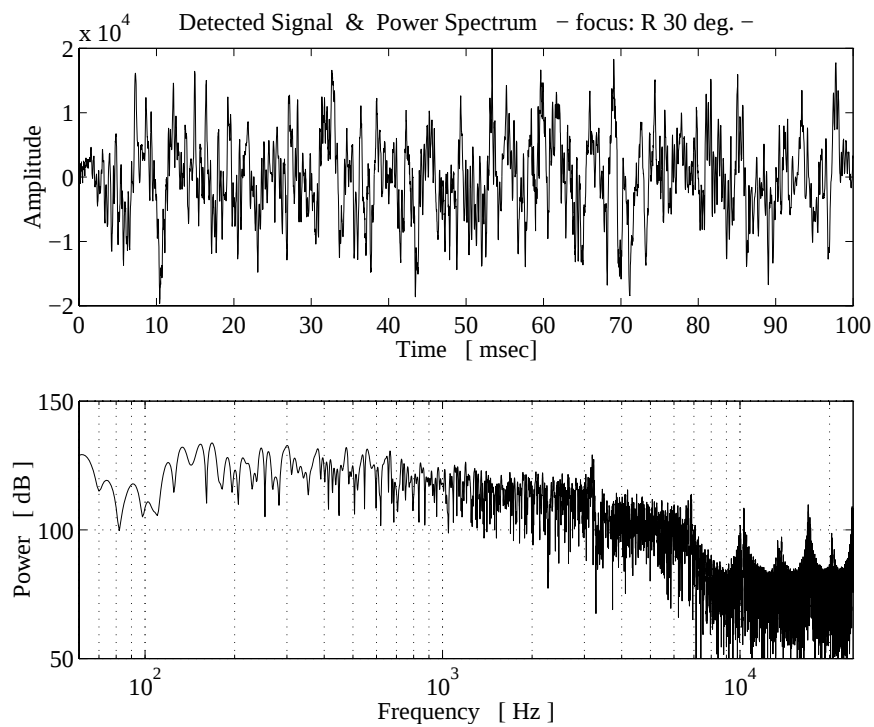


図 3.5: 焦点を信号到来方向 (右 30°) に一致させた場合の信号抽出結果

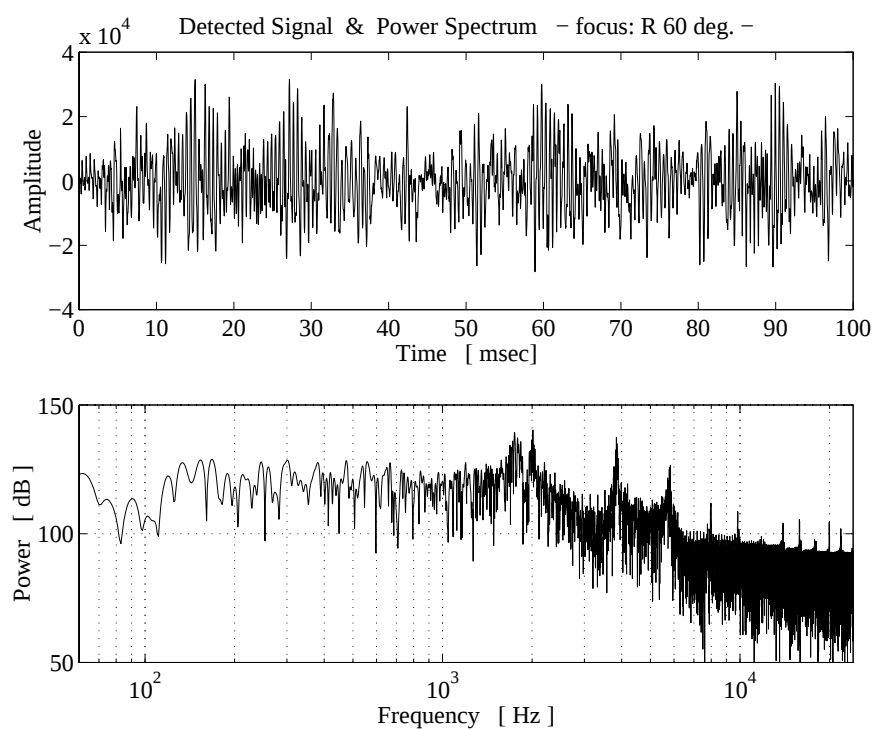


図 3.6: 焦点を右 60 °と設定した場合の信号抽出結果

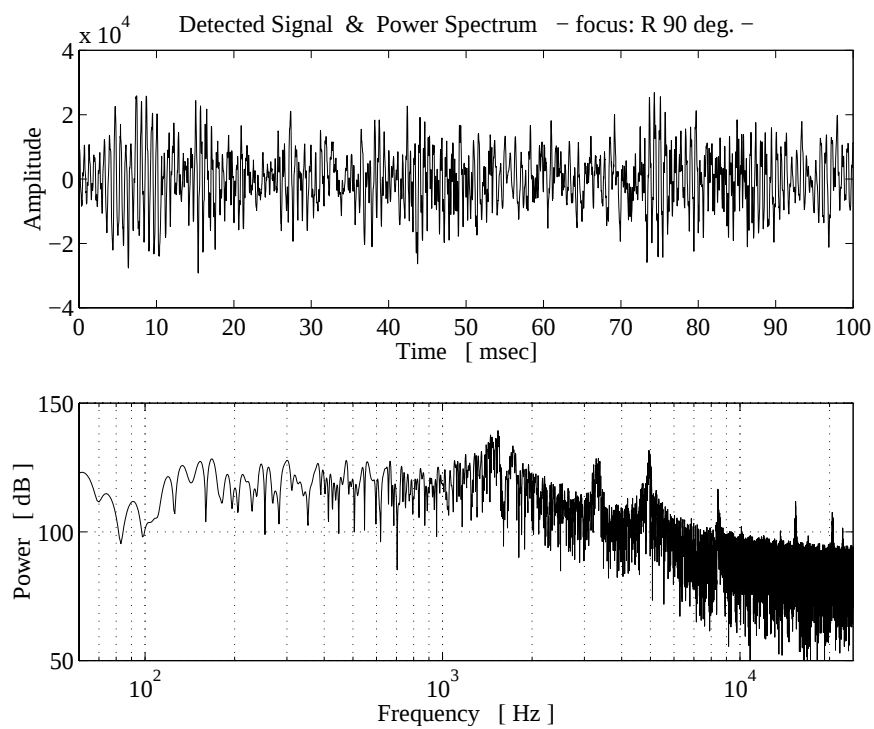


図 3.7: 焦点を右 90 °と設定した場合の信号抽出結果

3.2.4 考察

シミュレーション結果に関しては、抽出したい目的信号とシミュレーションにより得られた信号のパワースペクトルに着目し、信号抽出アルゴリズムの性能を主観評価する。

まず、マイクロホンアレーの焦点が信号の到来方向と一致した場合には、式(3.10)の分母が0に近くなる周波数(約1.7kHz、約3.4kHz、約5.1kHz...)であっても誤差は少なく、かなり高精度の信号抽出が可能であることが分かる。

次に、焦点が信号の到来方向とずれた場合は、300 Hz ~ 1 kHz 程度までの周波数帯域においては、焦点が到来方向に一致した場合の抽出結果とほぼ同じであるが、低・高周波数領域においてかなりの抽出誤差が生じていることが分かる。この原因は、低周波数領域における誤差は、フレーム内のデータ数が少ないため計算誤差ではないかと考えられ、高周波数領域における誤差は、マイクロホンアレーの指向特性に生じる凸凹の影響である。

3.3 考察

本章では、残響除去アルゴリズムのうち、残響成分の抽出過程について解説した。その残響成分抽出アルゴリズムでは、マイクロホンアレーの両端で受信した信号を用いて、音波は平面波として伝搬すると仮定し、幾何音響学的な概念に基づき解析的に中央のマイクロホン位置における残響成分を推定する。

本手法は、ピンクノイズを用いた計算機シミュレーション結果から分かるように、1方向からのみ信号が到来する場合は、かなりの高精度で信号を抽出できることが分かった。しかし、本研究では、この手法を直接音以外の全ての残響成分、すなわち直接音の到来方向(死角方向)以外から到来する残響成分を等感度で抽出するために使用する。

そこで、マイクロホンアレーの指向特性を考慮すると、全周波数帯域において残響成分を正確に抽出することが不可能であるが、マイクロホンアレーの指向性を見る限り、1 kHz 程度までの低中周波数帯域においては、焦点がずれていても正中面方向付近から到来する信号以外をほぼ等感度で抽出できることが想像でき、実際に、計算機シミュレーションを行なうことによりそれを確認した。

従って、低中周波数帯域にパワーが集中しているという音声の特徴を考慮すると[3]、音声認識の残響除去プリプロセッサを目指す本残響除去法で、今田らの2波形分離アルゴリズムを残響成分の抽出に応用したことは、完璧ではないがあながち誤りでもないだろう。

第 4 章

残響成分の除去

4.1 残響成分除去アルゴリズム

4.1.1 残響成分除去アルゴリズムの概説

本残響除去法が、残響成分の抽出過程と残響成分の除去過程の 2 つに大別できることは前述の通りである。そして、残響成分の抽出に関しては第 3 章で既に解説したので、次段階として、直接音と残響成分の合成信号である受音信号から、推定した残響成分の減算を考えなければならない。

減算処理については、まず時間領域において、波形どうしを単純に減算することが考えられるが、この演算を行なうためには、位相特性も十分に考慮しなければならない。しかし、本残響除去法は音声認識システムのプリプロセッサを目指しており、現行の音声認識アルゴリズムのほとんどが位相情報を必要としていないという現状では、時間領域において減算を行なうことは非合理的である。

そこで、次に考えられるのが周波数領域における減算である。そこで、本研究ではその演算にスペクトルサブトラクションを利用する。

4.1.2 スペクトルサブトラクションとは

スペクトルサブトラクションとは、周波数領域においてスペクトルの減算を行なうことにより、雑音に埋もれた音響信号を抽出する雑音除去の一手法であり、1979 年に Boll によって提案された [18]。

Boll は、ヘリコプタのコックピット内で収録した、雑音と音声の混合された音響信号から、スペクトルサブトラクションを使用して音声のみを抽出する実験を行ない、その手法が定常雑音の除去には非常に有効であることを示した。

また、時間領域における波形どうしの減算を考えた場合、信号の位相までも厳密に考慮しないと正確な演算は行なえない。それに対し、周波数領域における減算であれば、位相に関してはそれほど厳密に考えなくても良いことから、スペクトルサブトラクションは画期的な手法であり、Boll が発表した後に様々な改良が加えられ、現在でも広く利用されている。

4.1.3 スペクトルサブトラクションの応用

本研究では、直接音と残響成分が合成された信号から、残響成分のみを減算するためにスペクトルサブトラクションを利用するが、スペクトルサブトラクションを利用する前提条件として、残響成分は定常性を有している必要があるため [18]、短時間フレームごとに計算を行なうこととする。

また、前述の通り、残響成分を正確に計算することは不可能である。これに関しては、マイクロホンアレーの指向特性のため、高周波数成分に関しては実際のレベルよりも大きく抽出されることを計算機シミュレーションにより確認した。そこで、減算を行なう際に残響成分のレベルを小さくするため、残響成分に 1 よりも小さい定数を乗じることを考える。以後、この定数をサブトラクション係数と呼ぶ。この定数を乗じるという処理は、スペクトルサブトラクションを利用する場合には良く用いられている概念である [19]。

ここで、あるフレームにおける残響成分は、その 1 つ前のフレームにおける残響成分を除去した結果、すなわち、直接音の推定値と高い相関があること¹が予想される。よって、それらの相関値を計算し、その絶対値に 0.01 を乗じた値をサブトラクション係数と定める。この背景には、複数のデータを用いて残響除去実験を行なった結果、サブトラクション係数を 0.001 ~ 0.01 に設定することが望ましいと分かったため、0 ~ 1 の値をとる相関値の絶対値に 0.01 を乗じることにより、その最適値に近付けるという理由がある。

本研究では、フレームごとに残響成分の抽出、サブトラクション係数の計算、残響成分の除去を行ない、より高精度の残響成分除去を目指す。

¹ フレーム長を 10.7 msec、フレーム周期を 5.31msec としたため、あるフレームに存在する信号の反射音が次のフレームに含まれている可能性が高いと考える

4.2 考察

本章では、本研究で提案した残響除去法のうち残響成分の除去過程について解説した。この残響成分除去アルゴリズムでは、統計音響学的に受音信号と残響成分が各フレーム内においては定常であると仮定し、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を利用した。

また、スペクトルサブトラクションを行なう際に、マイクロホンアレーの指向特性を考慮してサブトラクション係数を導入した。そこで、サブトラクション係数という概念は広く用いられているが、本研究ではそれを定数とせず、フレームごとに逐次計算することにより、時変性の信号にも対応できる残響成分除去アルゴリズムが構築できた。

なお、本研究で対象としている信号は音声であるため、残響除去アルゴリズムにおける各種のパラメータ²は音声などの時変性の信号を前提として設定した。そして、次章においてピンクノイズのような定常的な信号ではなく、より現実に近い、残響のある音声を用いた残響除去実験を行なっているので、本残響除去法の有用性に関する考察はそちらに譲る。

²フレーム長 (512 pts、サンプリング周波数 48 kHz で 10.7 msec)、フレーム周期 (256 pts、同様に 5.3 msec)、窓関数 (hamming 窓)、サブトラクション係数

第 5 章

残響除去実験

5.1 実験の目的

残響除去実験を行なう目的は、本研究で提案した残響除去法が、実環境においてどれほど有効であるかを調査することである。

そこで、本章では、残響実験を行なうために必要となる、残響のある音声データの作成方法、ならびに、残響除去を行なった結果を評価するために本研究で提案した客観的評価尺度について解説する。

5.2 残響のある音声データの作成

5.2.1 データ作成方法

残響除去実験を行なうためには、まず、残響のある音声データを作成しなければならない。そこで、本研究では、クリーンな音声をスピーカを用いて室内に放射し、それを 3 チャンネルマイクロホンアレーで收音することにより残響のある音声データを作成した。なお、スピーカから実音場に放射するクリーンな音声としては、ATR 音声データベース [22] に収録されている音声データから、男性話者により発話された単母音 /a/ (mhtsy003.ad) と単語 /bunri/ (nh t 14348. ad) を使用した。

その際に、各マイクロホンを隣接するマイクロホンとの距離が 10 cm となるように直線等間隔配置し、マイクロホンの音響中心が床上 1 m の水平面上に存在するように設置

した。また、スピーカとマイクロホンアレーとの距離は 2.5 m とし、スピーカから放射する音声データの音圧レベルは、マイクロホンアレーの中央のマイクロホンの振動板位置において約 90 dB(A) となるように調整し、サンプリング周波数 48 kHz、量子化ビット数 16 bit で収録した。これは、人の発声時の音圧レベルに比較すると大きいが、暗騒音レベルが約 33 dB(A) であったことを考慮し、高 SNR で収録をするための措置である。

ここで、残響のある音声データを作成するために使用した機材を表 5. 1 に、データ作成時のブロックダイアグラムを図 5. 1 に示している。

表 5. 1: 使用機材一覧

使用機材	メーカ	名称	備考
DAT	SONY	PCM-2700A	
パワー・アンプ	SANSUI	AU α 907 MR	パワーアンプ部にダイレクト入力
スピーカ	<i>laboratory-made</i>		使用ユニット: FOSTEX FE127
マイクロホンアレー	<i>laboratory-made</i>		使用マイクロホン: RAMSA WM C75 (3本)
マイクロホン・アンプ	TASCAM	MA 8	内蔵パワーサプライも使用
レコーダ	TEAC	RD 135T	8 ch. デジタル MTR
騒音計	HCN	NL-04	音圧レベルの測定にのみ使用
DAT テープ	SONY	DF-60	録音用、再生用に各 1 本使用

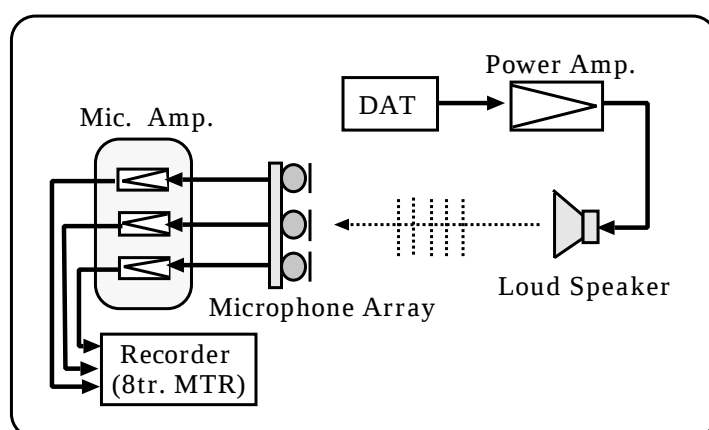


図 5. 1: 実験データ作成時のブロックダイアグラム

ここで、クリーンな信号と残響付加を行なった信号について、それぞれ時間波形と平均パワースペクトルを以下に示すが、図 5.2～図 5.5 が単母音 /a/ の場合で、図 5.6～図 5.9 が単語 /bunri/ の場合である。

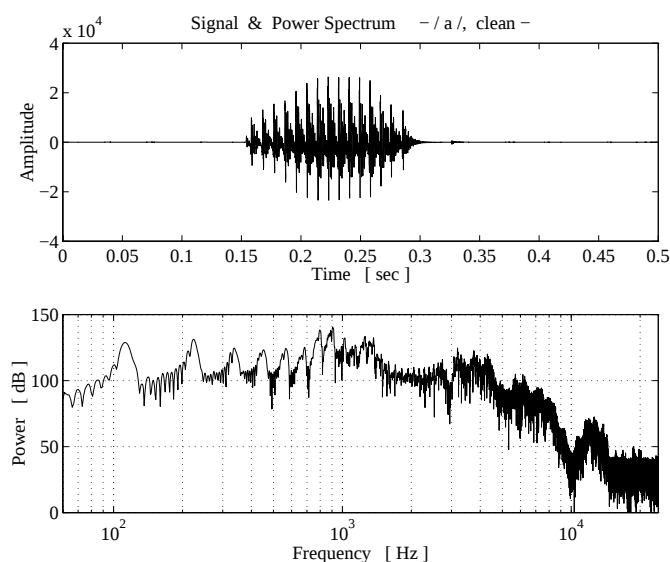


図 5. 2: クリーンな信号 (/a/)

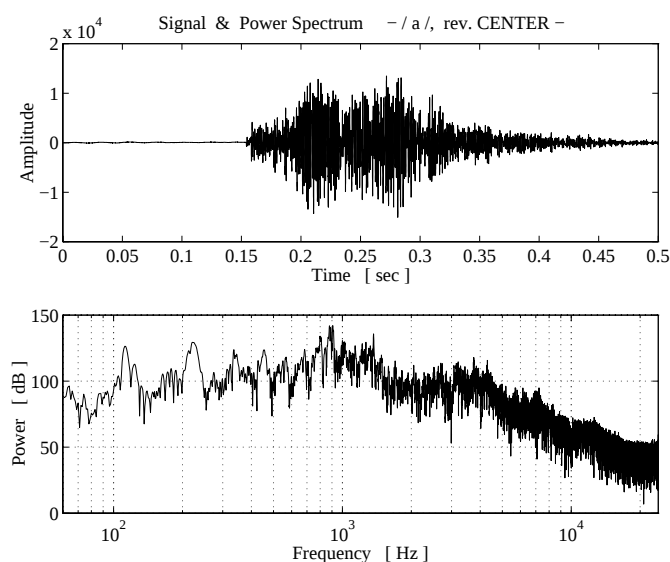


図 5. 3: 残響付加を行なった信号 (/a/、中央のマイクロホンでの受信信号)

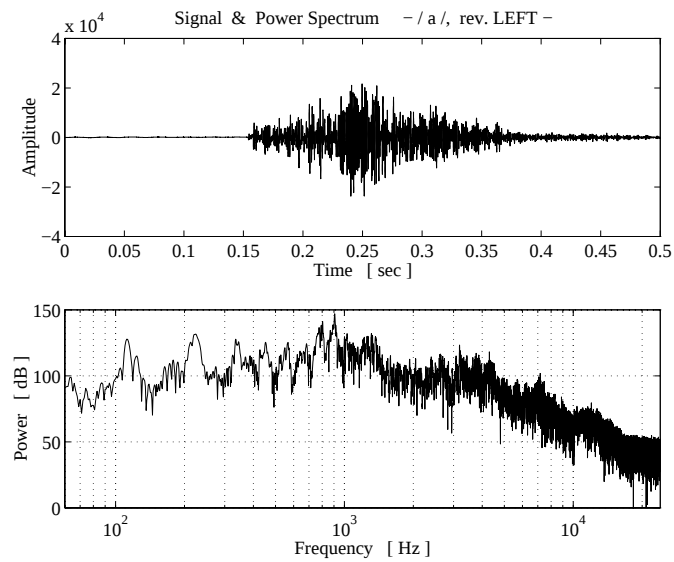


図 5.4: 残響付加を行なった信号 (/a/, 左端のマイクロホンでの受信信号)

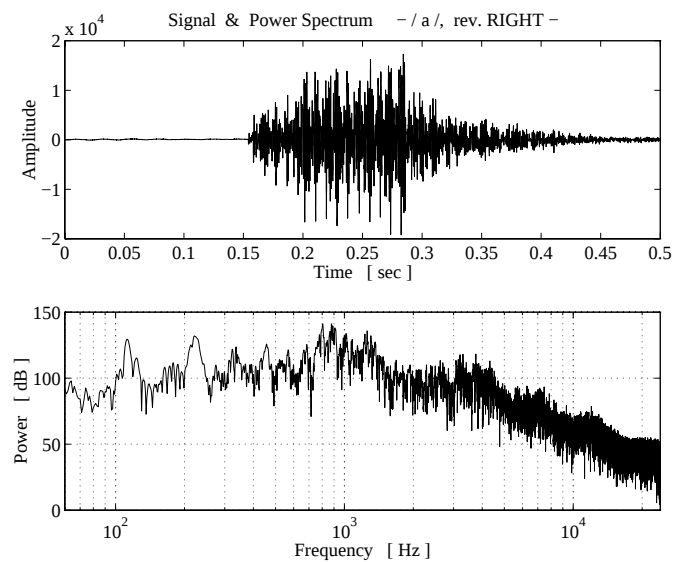


図 5.5: 残響付加を行なった信号 (/a/, 右端のマイクロホンでの受信信号)

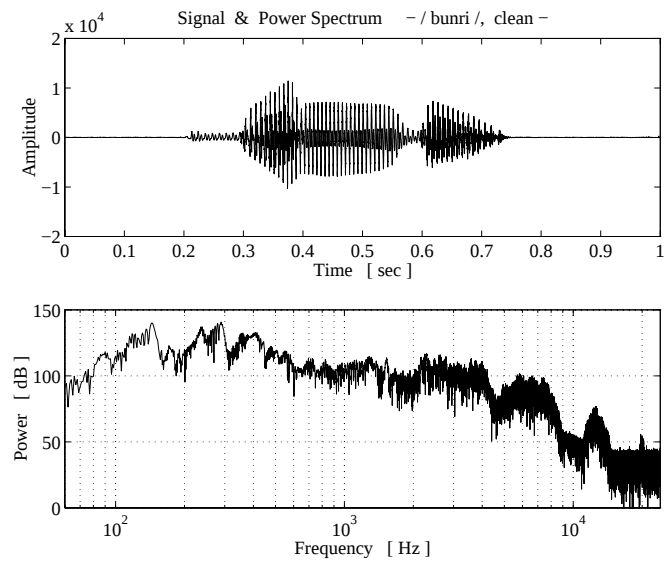


図 5.6: クリーンな信号 (/bunri/)

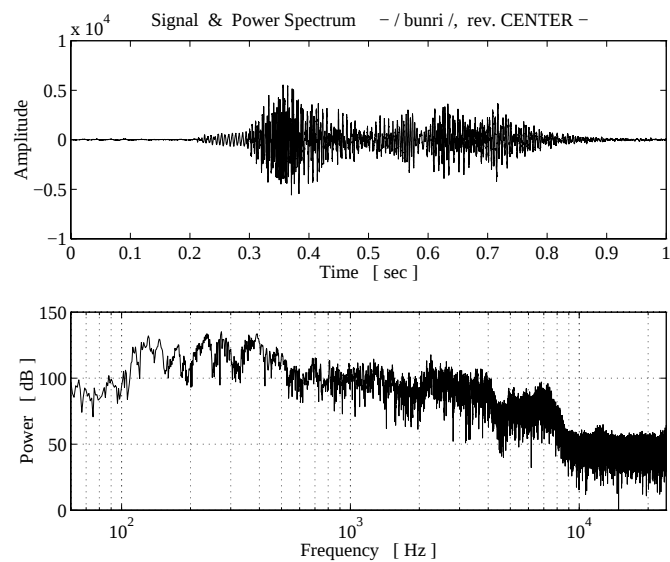


図 5.7: 残響付加を行なった信号 (/bunri /、中央のマイクロホンでの受信信号)

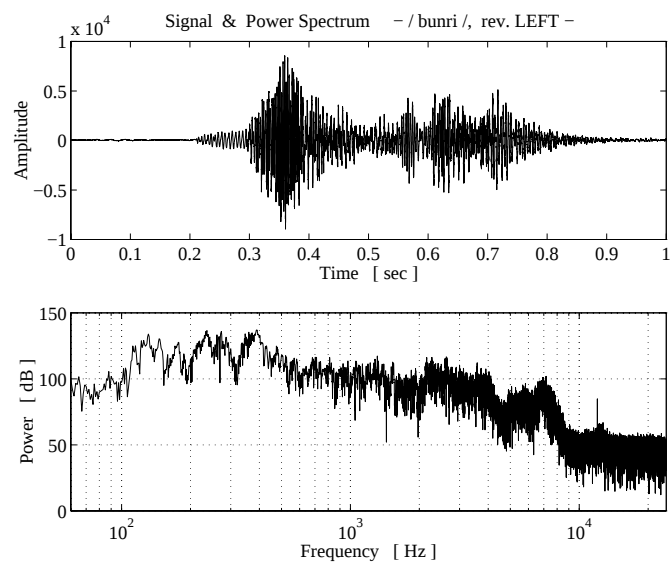


図 5.8: 残響付加を行なった信号 (/bunri/、左端のマイクロホンでの受信信号)

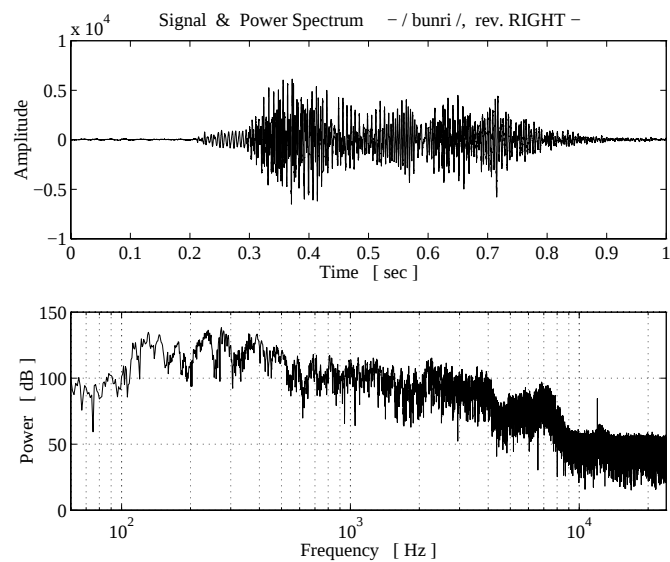


図 5.9: 残響付加を行なった信号 (/bunri/、右端のマイクロホンでの受信信号)

5.2.2 環境条件

本研究では、我々の日常生活で使用するような環境を想定し、室容積が約 50 m^3 で、 500 Hz における残響時間が 0.52 sec の室をデータ作成に使用した。この室の平面図を図 5.10 に示しており、床面からの反射音を抑えるために床にはグラスウール¹を設置した。また、データ作成時の気象条件は、天候は曇、室温 17.1 、湿度 22.3% であった。

ここで、データ作成に使用した室の残響特性を図 5.11 に示しているが、その計測方法について説明する。まず、TSP (Time Stretched Pulse) 信号を用いて室のインパルスレスポンスを測定する [20]。このインパルスレスポンスには測定システム系のインパルスレスポンスも含まれているが、本研究においては、その影響は極僅かであると考えて無視することとする。次に、そのインパルスレスポンスをバンドパスフィルタに通過させることにより、インパルスレスポンスの特定周波数帯域成分を抽出する。そして、各周波数帯域ごとに 2 乗積分法により残響減衰曲線を求め [4]、それより残響時間を読みとった。

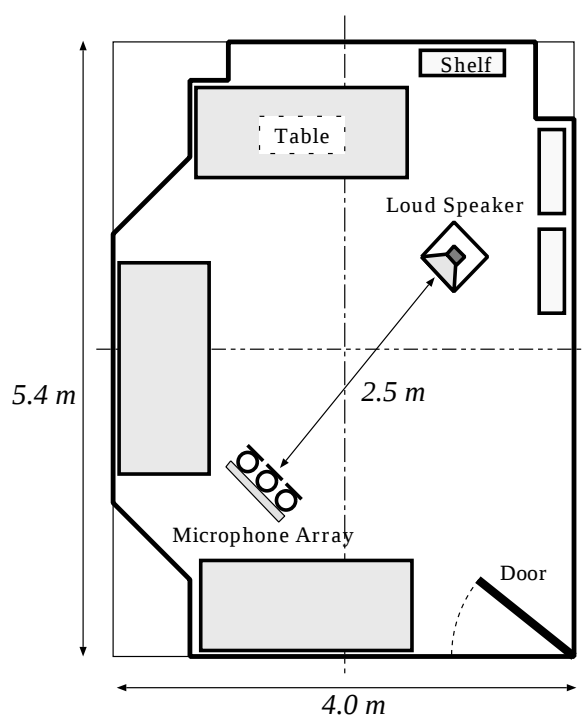


図 5.10: データ作成に使用した室の平面図

¹JIS 規格 A9505、A6306 に遵守した、密度 32 kg/m^3 、厚さ 50 mm のグラスウールにカラスクロスを貼ったものである。吸音特性については文献 [21] を参照のこと。

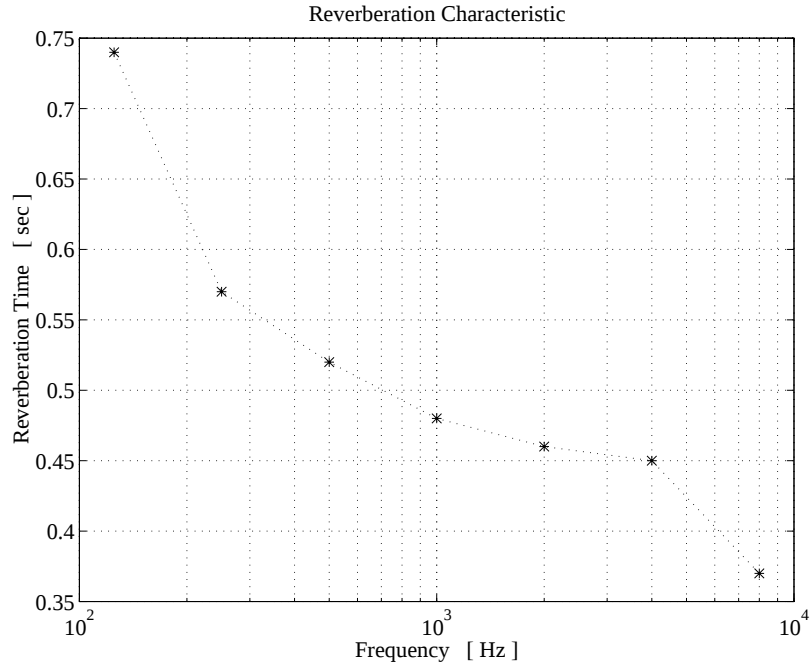


図 5.11: データ作成に使用した室の残響特性

5.3 評価尺度の設定

本研究では、残響除去実験の結果を客観的に評価するために、従来の評価尺度に基づき、それに若干の改良を加えた新たな評価尺度を提案する。

現在、残響あるいは雑音による音響信号の歪みを評価するために、一般的に使用されている客観的尺度としては、以下のように時間領域で定義されている信号対雑音比 (SNR: Signal to Noise Ratio) と、周波数領域で定義されているスペクトル歪 (SD: Spectral Distortion) がある。

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N x(n)^2}{\sum_{n=1}^N \{x(n) - y(n)\}^2} \quad [dB] \quad (51)$$

$$SD = \left[\frac{1}{W} \sum_{\omega=1}^W \{S_x(\omega) - S_y(\omega)\}^2 \right]^{\frac{1}{2}} \quad [dB] \quad (52)$$

ここで、式 (51) の $x(n)$ 、 $y(n)$ は信号のサンプル値、 N は信号測定区間のサンプル数であり、式 (52) の $S_x(\omega)$ 、 $S_y(\omega)$ は信号の対数スペクトル、 W は対数スペクトルを計算する際に使用する FFT のポイント数を表している。

また、本研究で提案した残響除去法は、音声認識の残響除去プリプロセッサを前提にしているため、信号の位相情報にも依存する SNR ではなく、振幅スペクトルのみに依存し、位相情報を必要としない SD を用いる方が望ましい。

しかし、式(5.2)で定義される SD そのものでは、本残響除去法の性能を十分に評価することはできない。従って、SD を計算する際の周波数帯域を音声帯域(125 Hz ~ 6 kHz)のみと制限し、さらに本研究で提案した残響除去法がフレーム処理を行なっているため、SD の評価も短時間フレームごとに計算を行なうことにする。以上の改良を加えた SD を音声帯域セグメンタル SD と名付け、フレームごとに音声帯域 SD (SD_{speech}) を式(5.3)のように計算する。

$$SD_{speech} = \left[\frac{1}{W_{\max} - W_{\min} + 1} \sum_{\omega=W_{\min}}^{W_{\max}} \{S_x(\omega) - S_y(\omega)\}^2 \right]^{\frac{1}{2}} \quad [\text{dB}] \quad (5.3)$$

ここで、 $S_x(\omega)$ 、 $S_y(\omega)$ は信号の対数スペクトル、 $W_{\min}$ は音声帯域の最低周波数(約 125 Hz)に対応した周波数領域におけるサンプル番号、 $W_{\max}$ は音声帯域の最高周波数(約 6 kHz)に対応した周波数領域におけるサンプル番号を表している。また、音声帯域セグメンタル SD の計算を行なう際には、フレーム長 1024 pts (サンプリング周波数 48 kHz で 21.3 msec)、フレーム周期 512 pts (同様に 10.7 msec)とし²、データの切り出しには窓関数として hamming 窓を利用する。

5.4 実験結果

残響除去実験の結果については、本手法では位相情報を全く考慮していないため、時間波形について議論することはできない。そこで、残響除去を行なった結果、前節で提案した音声帯域セグメンタル SD がどれくらい改善されたかについてのみ考察する。但し、本研究における音声帯域セグメンタル SD には、それを計算する際の基準信号としてクリーンな音声を用いているため、スピーカなどの録音再生システムの伝達関数による歪みも含まれていることに注意しなければならない。

そこで、データとして /a/、/bunri/ を用いた残響除去実験の結果をそれぞれ図 5.12 図 5.13 に、クリーンな信号(目的信号)と共に残響除去による音声帯域セグメンタル SD

²本来は、残響除去アルゴリズムで設定した値と一致させることが望ましいが、それでは周波数分解能が低過ぎるため、評価にはフレーム長とフレーム周期をそれぞれ 2 倍にした。

表 5.2: 音声帯域セグメンタル SD の音声区間平均改善量

データ	音声区間	平均改善量
/ a /	14 - 27 [frame no.]	3.18 [dB]
/ bunri /	18 - 65 [frame no.]	1.29 [dB]

の改善量を表示している。そして、音声区間における音声帯域セグメンタル SD の平均改善量を表 5.2 に示している。

次に、残響成分の到来方向の推定結果について、方向推定を行なったフレームごとの信号の到来方向推定結果と、それを基に各信号ごとに作成した方向推定結果のヒストグラムをデータ /a/、/bunri/ の各場合についてそれぞれ図 5.14~ 図 5.17 に示している。

また、本研究では、信号到来方向の推定結果より作成したヒストグラムの正中面方向（直接音の到来方向）付近を除いたピークが存在する方向を残響成分の到来方向と定めた。そこで、/a/、/bunri/ の各場合についての残響成分の到来方向の検出結果、すなわち、残響成分の抽出を行なう際のマイクロホンアレーの焦点を表 5.3 に示している。なお、残響成分の左右両端のマイクロホンへの到来時間差については、正中面方向に対して左側が + 方向である。

表 5.3: 残響成分の到来方向の検出結果

データ	残響成分の到来方向の検出結果	
	左右両端のマイクロホンへの到来時間差	到来方向
/ a /	+ 6 [pts] (+ 0.125 [ms])	左 12 °
/ bunri /	+ 6 [pts] (+ 0.125 [ms])	左 12 °

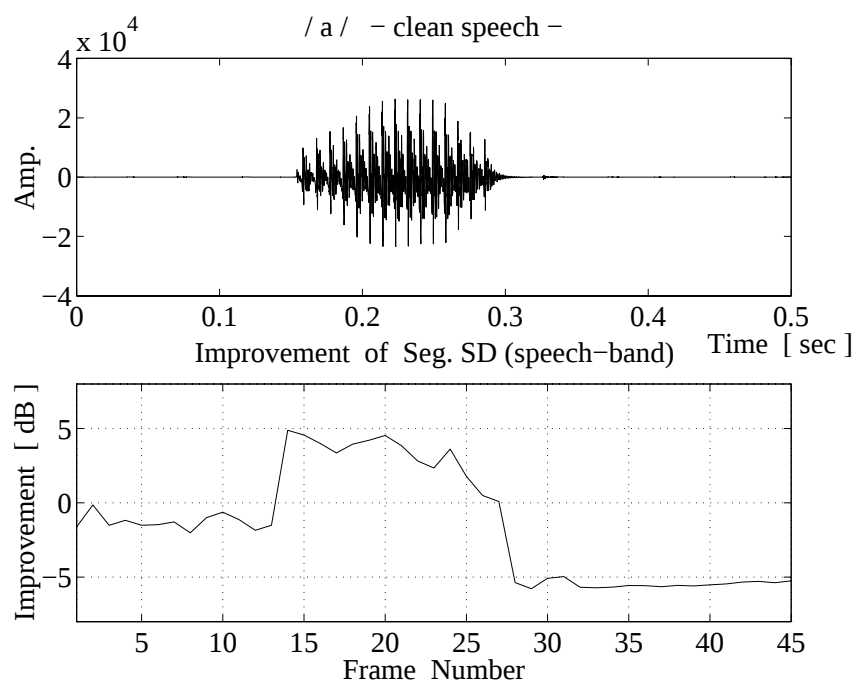


図 5.12: /a/ のクリーンな信号（上段）と残響除去の結果（下段）

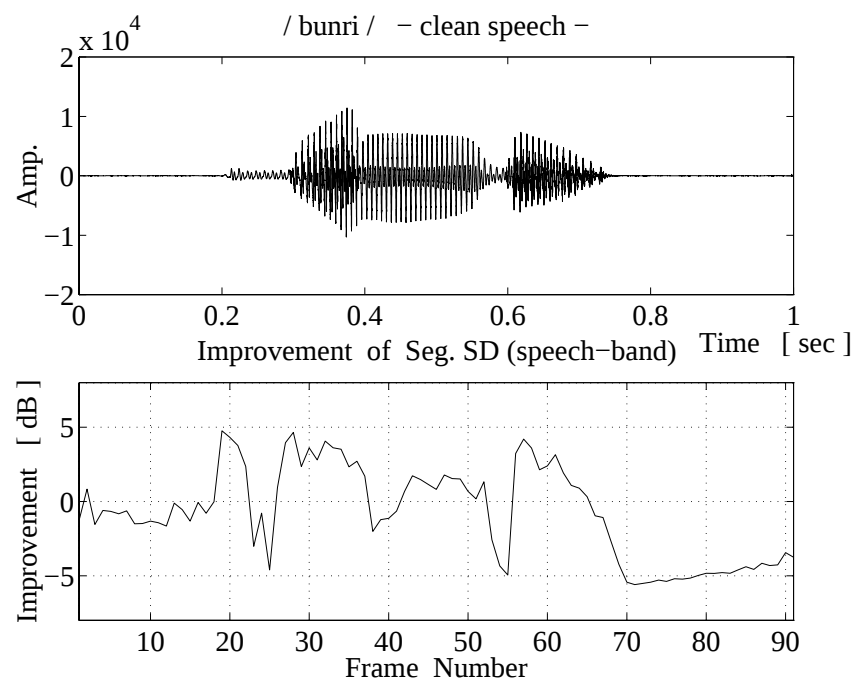


図 5.13: /bunri/ のクリーンな信号（上段）と残響除去の結果（下段）

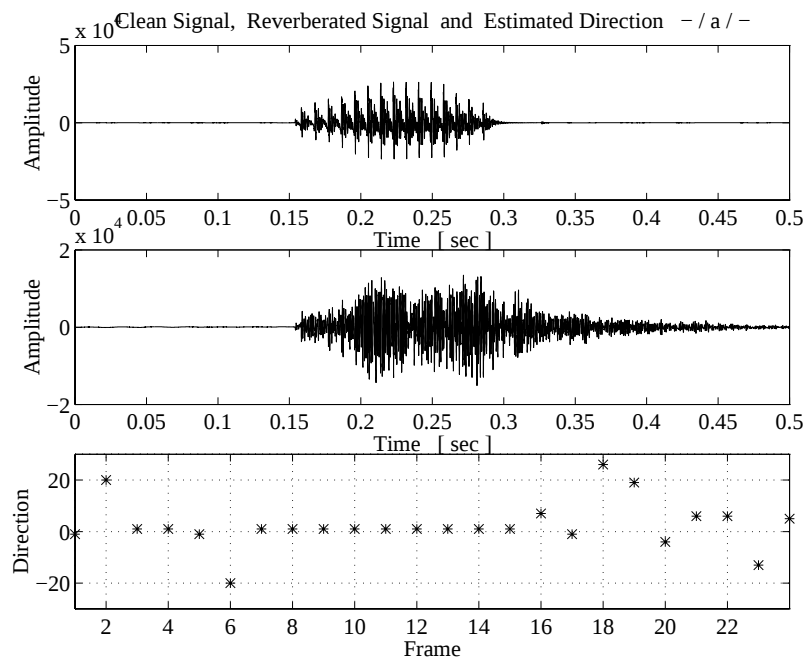


図 5.14: /a/ のクリーンな音声（上段）、残響のある音声（中段）、各フレームにおける信号の到来方向推定結果（下段）

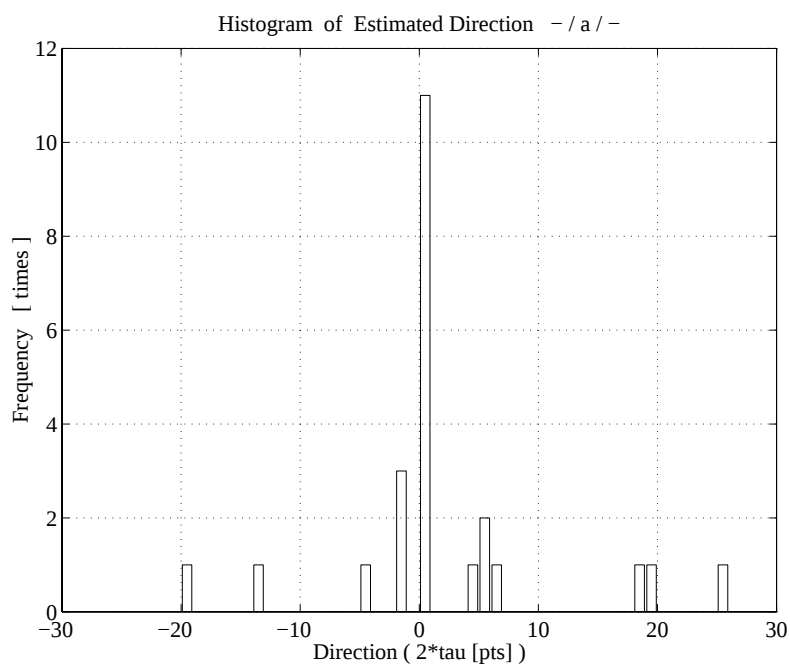


図 5.15: /a/ の信号到来方向の推定結果より作成したヒストグラム

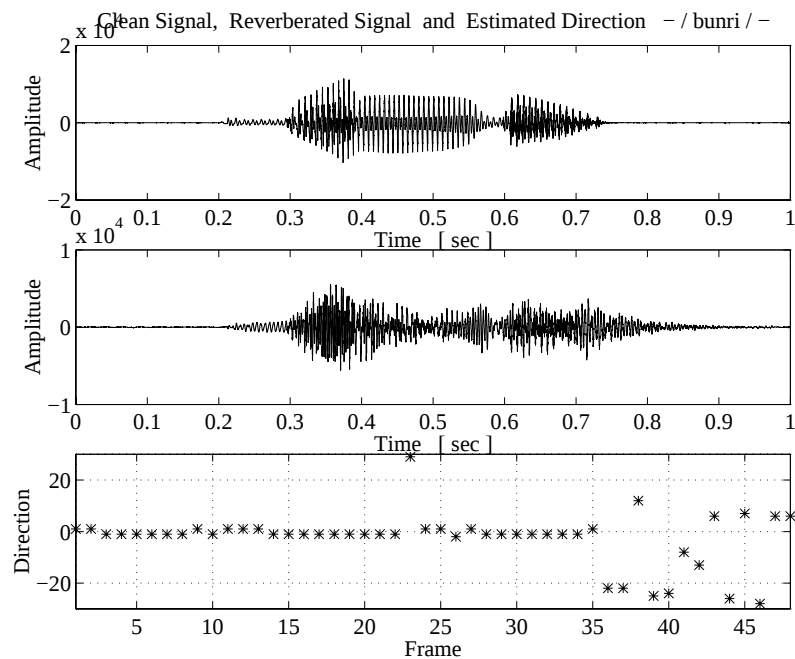


図 5.16: /bunri/ のクリーンな音声（上段）、残響のある音声（中段）、各フレームにおける信号の到来方向推定結果（下段）

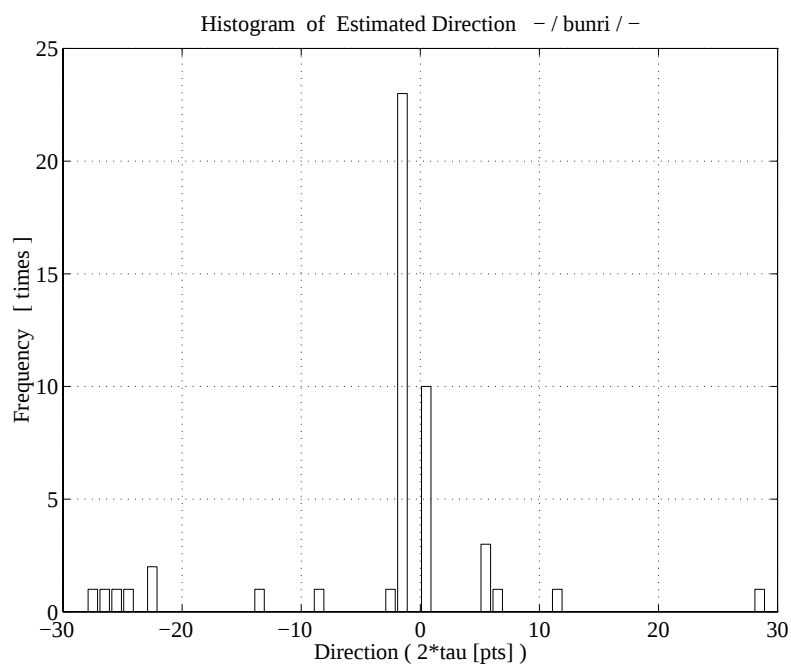


図 5.17: /bunri/ の信号到来方向の推定結果より作成したヒストグラム

5.5 考察

残響除去実験の結果に基づき、本研究で提案した残響除去法について考察する。

まず、残響成分の到来方向推定に関しては、同じ環境条件で収録した2つのデータ /a/、/bunri/ に対して全く同一の方向が検出されたので、その精度は問題ないと考えられる。

次に、残響除去の精度に関しては、単母音 /a/ の場合は図 5.12 に示しているように、上段の目的信号の音声区間に対応するフレームに関しては、そのすべてフレームにおいて音声帯域セグメンタル SD が改善されている。それに対して、図 5.13 の単語 /bunri/ の場合には、音声帯域セグメンタル SD の改善量がマイナス、すなわち、残響除去を行なった結果、音声帯域セグメンタル SD が更に大きくなったフレームも存在することが分かった。これを同図の上段に示した目的信号と見比べると、音声の子音区間であることが確認できる。

よって、本研究で提案した残響除去アルゴリズムを用いて残響除去を行なうと、音声の母音区間においては非常に有効であるが、子音区間においては逆に歪みが大きくなるという問題が生じることが分かった。

ここで、その原因について考える。まず、その最大の原因は、母音と子音とのパワー差に起因しているのではないかと考えている。つまり、時系列上で子音の前に母音がある場合には、パワーの小さな子音の直接音にパワーの大きな母音の残響成分が付加されるために残響除去が非常に困難になったのではないかと考えている。これに関しては、コンサートホールなどの残響時間の長い室内においては、子音が聴き取りにくくなるという我々の日常的な経験からも納得できるが、それゆえに子音の残響除去は重要な問題と考えている。また、スペクトルサブトラクションを行なう際に、その処理を行なうフレーム内では信号は定常であると仮定したが、音声の子音は非定常であるためにその仮定を満たすことができないにも関わらず強引に残響除去を行なったことも、子音区間において、音声の振幅スペクトルの歪みが大きくなった一因ではないかと考えられる。この子音区間において歪みが大きくなる原因を解明するためには、理想的な環境を想定し、残響除去の計算機シミュレーションを行なうことも有効な手段であろう。

なお、各フレームごとに、音声のパワーを用いて母音と子音の区別を行ない、その母音区間にのみ本手法を適用すれば、子音区間においてスペクトルの歪みが増すという結果にはならないがこれは本質的な解決法ではない。そこで、残響除去の精度向上のためには、

サブトラクション係数をマイクロホンアレーの指向特性を考慮して周波数帯域ごとに変化させることや、それを適応的に計算することなども考えられる。

最後に、評価尺度についても再検討が必要であると考えている。まず、原段階では録音再生システム系の伝達関数特性を補償³していないため、音声帯域セグメンタル SD には残響による歪みだけではなく、システム系の歪みも含まれている。よって、録音再生システム系の伝達関数特性を補償するか、もしくは残響による歪みとシステム系の歪みとの性質の違いについて調査することにより、残響除去の結果をより正確に知ることが可能になるであろう。そして、本研究の目的を考慮すると、本残響除去法を音声認識装置の残響除去プリプロセッサとして用いた場合に、音声認識性能がどの程度向上するかを調査することが望ましい。但し、音声認識に必要な周波数帯域さえ分かれば、SD の計算を行なう際に重み付けを行なうことが可能になり、実際に音声認識実験を行なわなくても、それと同程度に信頼できる評価を簡単に実現できるのではないかと考えている。

³ 収音システム系の伝達関数を測定し、その逆フィルタを受音信号に畳み込めばよい。

第 6 章

おわりに

6.1 本残響除去法について

本研究では、音声認識システムの実用化を支援するための残響除去法を提案した。それは、3チャンネルの等間隔直線配置マイクロホンアレーを用い、その両端のマイクロホン対で受音した信号に含まれる、到来方向に関する情報を利用した残響除去法である。

本残響除去法の特徴は、まず初めに、幾何音響学的概念に基づいて解析的に残響成分を抽出し、次に、統計音響学的手法を用いて残響成分の除去を行なうという2つの概念を融合させている点である。これにより、非常に複雑な残響除去問題の簡潔化が実現できた。

まず、残響成分の抽出に関しては、任意の方向にマイクロホンアレーの焦点を設定することにより、約 1 kHz までの低中周波数成分に対しては、正中面方向を除いてあらゆる方向から到来する信号、すなわち、残響成分をほぼ等感度で抽出できることが分かった。そのため、本手法を用いることにより、その低中周波数帯域にパワーが集中している音声に対して残響成分の抽出が実現できた。

次に、残響成分の除去に関しては、スペクトルサブトラクションを用いたが、それを音声のような時変性の信号にも対応できるように、短時間フレーム処理、そして、フレームごとに定めるサブトラクション係数を導入した。

そして、残響除去実験を行なった結果、本残響除去法を利用することにより、音声の母音区間においては残響により生じた歪みを回復させることができ、本手法の有効性を確認

できた。しかし、単語などの子音区間においては、本手法を用いて残響除去を行なうとスペクトルの歪みが増すという問題点が存在することも明らかになった。

最後に、本残響除去法の音声認識のための残響除去プリプロセッサとしての有用性について考えてみる。一般に、残響除去は非常に困難な問題であり、王らは残響除去を行なった結果が SD で約 1 dB しか改善されなくても残響除去の効果はあると述べているため [17]、本残響除去法は、現段階ではまだ問題点を抱えているため完璧な手法とは言えないが、音声認識の残響除去プリプロセッサとしての素質は十分に備えている。そして、その目標を達成できる日はそう遠くないものと信じている。

6.2 今後の課題

本研究では、残響除去の第一歩として、音声認識の残響除去プリプロセッサとしての残響除去法を提案したが、その精度についてはまだ満足できるレベルには達していない。

そこで、今後の課題としては、本残響除去法が抱えている問題点を解決するための一方として、受音点の数を増やすことが考えられる。これに関しては、本研究で提案したマイクロホンアレーは、1つのマイクロホン対とその中央に1本のマイクロホンを使用したものであり、残響成分の抽出にはマイクロホン対のみを用いている。そこで、このマイクロホン対を増やし、それらを円形配置することにより、残響成分の抽出精度を向上させること可能になるのではないかと考えている。

また、将来的には、本残響除去法が目標としている音声認識のプリプロセッサという枠組を超えた、より一般的な残響除去法にまで拡張したい。但し、そのためには位相情報までも考慮しなければならず、本研究で提案したスペクトルサブトラクションを利用した残響成分除去アルゴリズムの再検討が必要になるであろう。

謝辞

本論文の作成にあたり、一年間にわたり熱心に御指導して頂きました赤木 正人 助教授をはじめ同講座の飯島 泰蔵 教授、岩城 護 助手に厚く御礼を申し上げます。そして、本研究を行なうに過程において、熱心な御議論ならびに様々なアドバイスを下さいました学内外の諸先生方にも、この場を借りて御礼申し上げます。

また、北村 達也 氏をはじめとする赤木研究室の先輩方には、本研究を進めるうえで、あらゆる局面において多大なるアドバイスを頂きましたことに感謝致します。

最後に、本研究に対する議論とアドバイスを頂いたすべての皆様に謝意を表します。

参考文献

- [1] 小橋 豊: “音と音波,” 裳華房 (1969)
- [2] 永田、飯田、古宇田、橘、古川、安岡、山本: “建築音響,” コロナ社 (1988)
- [3] 古井 貞熙: “デジタル音声処理,” 東海大学出版会 (1985)
- [4] 大賀、山崎、金田: “音響システムとデジタル処理,” 電子情報通信学会 (1995)
- [5] 東倉、赤木、阪上、鈴木、中村、山田: “音のなんでも小辞典,” 講談社 (1996)
- [6] 今田 稔一郎: “複数受音波形による音源分離に関する研究,” JAIST 修士論文 (1996)
- [7] 今田、赤木: “2点受音波形からの音源分離,” 音講論集, 3-3-14 (1996.3)
- [8] 飛田、菅村: “音声認識における周囲環境の影響,” 日本音響学会誌, vol. 51, no. 4, pp. 331-335 (1995)
- [9] 飛田、中津、菅村: “単語音声認識における残響の影響,” 音講論集, 137 (1996.3)
- [10] 飛田 瑞広: “単語音声認識における残響の影響とその改善法,” 音講論集, 188 (1999)
- [11] 引地、王、板倉: “室内伝達関数の時間変動特性,” 信学技報, EA 98-12 (1998)
- [12] J. B. Allen, D. A. Berkley and J. Baer: “Multimicrophone signal-processing technique to remove room reverberation from speech signal,” J. Acoust. Soc. Am. vol. 62, no. 4, pp. 912-915 (1977)
- [13] S. T. Neely and J. B. Allen: “Invertibility of a room impulse response,” J. Acoust. Soc. Am. vol. 66, no. 1, pp. 165-169 (1979)

- [14] M. Miyoshi and Y. Kaneda: "Inverse Filtering of Room Acoustic Echoes," *IEEE Trans. ASSP*-36(2) pp. 145-151 (1988)
- [15] Wang, F. and Takusagawa: "An Approach to the Problem of Multi-microphone Subband Equalization," *IEEE Trans. ASSP* 39(1) pp. 953-960 (1991)
- [16] 山田、王、板倉: "帯域分割 MINTEによる残響音声の回復," *信学技報*, E A 9 0- 1 9 (1 9 9 0)
- [17] 王、板倉: "多数マイクロホン帯域包絡推定による残響音声の回復 — マイクロホンの配置及び雑音の影響 —," *音講論集*, 1 - 3 (1 9 8 9) . 9
- [18] F. Boll: "Suppression of Acoustic Noise in Speech by Spectral Subtraction," *IEEE Trans. ASSP* - (227) (1979)
- [19] 庄境、中村、鹿野: "音声認識における音声強調手法及びモデル適応化の検討," *信学技報*, E A 9 6 - 1 (1 9 9 6) 6
- [20] 鈴木、浅野、金、曾根: "時間引き延ばしパルスの設計法に関する考察," *信学技報*, E A 9 2 - 2 (1 9 9 2) 2
- [21] 日本音響材料協会: "騒音・振動対策ハンドブック," *技報堂* (1 9 8 8) 2
- [22] 武田、匂坂、片桐、阿部、桑原: "研究用日本語音声データベース利用解説書," *ATR Techn. Report* R- 1 - 0 (1 9 8 8) 8

学会発表リスト

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),” 日本音響学会 平成 9 年春季研究発表会 2-P- 24(1997)