

平成 19 年度
学士学位論文

音の到着時間差を用いた音源空間推定

Estimation of Sound Source Zone using The Arrival
Time Interval

1080351 金井 宏一郎

指導教員 福本 昌弘

2008 年 3 月 7 日

高知工科大学 情報システム工学科

要 旨

音の到着時間差を用いた音源空間推定

金井 宏一郎

マイクロホンアレイを用いた受音では，指向性の制御や音の到来方向の推定を行うことが可能であり，目的とする音と不要な音を到来方向から分離することで，受信信号の S/N を向上させることができる．しかしながら，受信対象が移動する場合には，常に移動に合わせて指向性の制御をしなければならないといった欠点がある．マイクロホンアレイの指向性制御を自動化するためには，刻々と変化する音源位置の情報が必要となってくる．そこで，本論文ではマイクロホン 2 本を用いた音源空間の推定法を提案している．2 本のマイクロホンの受信信号について相関をとり，マイクロホン間の音の到着時間差を求める．音の到着時間差を求めることができれば，2 本のマイクロホンのうち先に音が到着したマイクロホンを特定することが可能となる．提案手法により，音源が複数存在する環境においても推定可能であることを計算機シミュレーションにより確認している．

キーワード 相互相関，無指向性マイクロホン，指向性マイクロホン，音源

Abstract

Estimation of Sound Source Zone using The Arrival Time Interval

Koichiro Kanai

The microphone array can control the directivity and presume the direction where the sound comes. Therefore, it is possible to separate a desired sound and an unnecessary sound from the direction of coming. Hence, the microphone array can improve S/N of the receiving sound. However, when the target of receiving sound moves, the microphone array always should match the direction of the microphone according to the movement. Information of the changing sound source position becomes necessary in order to automate directivity control of the microphone array. In this paper, estimation of sound source zone using two microphones is proposed. The proposed system is used for the correlation of two input signals. If the arrival time interval is estimated, the microphone which is near to the sound source is decided. The proposed system can estimate when two sound source exist at the same time.

key words cross-correlation, omni directional microphone, directional microphone, sound source

目次

第 1 章	序論	1
1.1	背景と目的	1
1.2	構成	2
第 2 章	音源から受音点までの音の伝わり方	3
2.1	まえがき	3
2.2	音の到着時間差	3
2.3	マイクロホンの種類	4
2.4	無指向性マイクロホン	4
2.5	単一指向性マイクロホン	4
2.6	単一指向性の性能	5
第 3 章	音源が存在する空間の推定	7
3.1	まえがき	7
3.2	音源空間推定	7
3.3	相互相関	8
3.4	実験環境	10
3.5	室内伝達特性の計測	10
3.6	シミュレーション方法	11
3.7	評価方法	13
3.8	計算機シミュレーションの結果	15
3.9	マイクロホンと音源の配置を変更した場合の変化	15
第 4 章	音源が複数存在する場合の性能評価	20
4.1	まえがき	20

目次

4.2	音源が複数存在する場合の音源空間推定法	20
4.3	シミュレーション方法	21
4.4	評価方法	23
4.5	計算機シミュレーションの結果	24
第 5 章	結論	30
5.1	本研究のまとめ	30
5.2	今後の課題	30
謝辞		31
参考文献		32
付録 A	適応信号処理	33
A.1	パラメータ推定問題	33
A.2	適応アルゴリズム	33
A.3	学習同定法	34

目次

2.1	音の到着時間差について	4
2.2	ソース音源	5
2.3	単一指向性マイクロホンの性能	6
3.1	音源空間推定システム	8
3.2	実験環境での伝達環境測定モデル	11
3.3	実験環境の説明	12
3.4	実験環境の写真	13
3.5	シミュレーションに用いた信号	13
3.6	受音対象空間と推定結果が不一致	14
3.7	受音対象空間と推定結果が一致	14
3.8	間隔 100mm での推定結果	17
3.9	間隔 200mm での推定結果	17
3.10	間隔 300mm での推定結果	18
3.11	間隔 400mm での推定結果	18
3.12	音源とマイクの配置	19
3.13	間隔 200mm45deg での推定結果	19
4.1	複数音源が存在する場合の音源空間推定法	21
4.2	シミュレーションに用いた信号 (1)	22
4.3	シミュレーションに用いた信号 (2)	22
4.4	マイク A の観測信号	23
4.5	マイク B の観測信号	23
4.6	間隔 100mm45deg での推定結果 (相関対象 4,410 サンプル)	26

図目次

4.7	間隔 100mm45deg での相関値の時間的变化 (相関対象 4,410 サンプル) . . .	27
4.8	間隔 100mm 一直線上での推定結果 (相関対象 4,410 サンプル)	27
4.9	間隔 100mm 一直線上での相関値の時間的变化 (相関対象 4,410 サンプル) .	28
4.10	間隔 200mm45deg での推定結果 (相関対象 4,410 サンプル)	28
4.11	間隔 200mm45deg での相関値の時間的变化 (相関対象 4,410 サンプル) . . .	29
4.12	間隔 200mm 一直線上での推定結果 (相関対象 4,410 サンプル)	29

表目次

3.1	相関係数	10
3.2	マイクロホン間 200mm での性能比較	16
3.3	マイクロホン間 300mm での性能比較	16
3.4	マイクロホン間 400mm での性能比較	16
4.1	マイク間 100mm45deg での性能比較	25
4.2	マイク間 100mm 一直線上での性能比較	25
4.3	マイク間 200mm45deg での性能比較	26

第 1 章

序論

1.1 背景と目的

テレビ会議等で発言者の声を録音したい場合，音源が空間内に複数存在し，発言者以外のノイズが集音される可能性がある．このような状況下でのマイクロホンによる受音方法として，マイクロホンアレイの利用が考えられる．マイクロホンアレイを用いた受音では，指向性の制御や音の到来方向の推定を行うことが可能であり，目的とする音と不要な音を到来方向から分離することで，受音信号の S/N を向上させることができる．例えば，テレビ会議においてマイクロホンアレイを用い発言中の人を特定し，その方向のマイクロホンのみの感度を高くすれば，話者の音圧レベルのみを上げることができ，結果として周囲の騒音を低減することができる．しかしながら，受音対象が移動する場合には常に移動に合わせマイクロホンの方向も合わせなければならないといった欠点がある．これまでに，ビームフォーマ法 [1]，各マイクロホン出力の DFT スペクトル間の位相差とパワー比から目的音検出を行う方法 [2] が提案されている．しかしながら，これらは指向性マイクロホンを用いるため，利用するにあたって指向性の方向調整が必要であり，設置する際にユーザに負担をかけさせてしまう．本研究では，無指向性マイクロホン 2 本のみを用い，音源である空間の推定を行う．無指向性マイクロホンを用いるため，音源の向きを考慮すること無くマイクロホンを配置することができる．よって，ユーザへの負担は軽くなると考えられる．音源の推定には，相互相関を用いて 2 本のマイクロホン間の音の到着時間差により求める．この推定結果を用いることで，常に音源である方向に指向性を向けるためのパラメータを提供することを可能とする．

1.2 構成

本研究では，前説で述べたマイクロホンアレイの問題を解決するために，マイクロホン 2 本を用いた音源である空間の推定を行うシステムを提案する．

本論文は 5 章より構成されている．以下に各章の概要を述べる．第 2 章では，本研究の着眼点である音の到着時間差について説明する．また，使用するマイクロホンによる受信信号の変化について述べる．

第 3 章では，音源である空間の推定法についての説明と音源が単一である場合のシミュレーション結果を示す．

第 4 章では，音源が複数存在する場合のシミュレーション方法とその結果を示す．

第 5 章では，本研究で得られた結果を元に結論を述べる．

第 2 章

音源から受音点までの音の伝わり方

2.1 まえがき

音源定位とは、生物が音源の方向や距離を知覚する能力のことである。人間は、両耳時間差と両耳強度差を手がかりとして、単一音源ならば数 10ms の時間で 1 度以下という優れた定位能力をもつ。ある空間内で、マイクロホンを用いて目的とする音を録音したい場合、音源の方向を知らなければ音源に対して指向性を向けることができない。音源の方向を推定するには、最低 2 本のマイクロホンが必要となる。

この章では、音源空間推定を行うにおいて必要となる音の到着時間差、またマイクロホンの性質について述べる。

2.2 音の到着時間差

一定の間隔をあけ設置したマイクロホンに対し音を出すと、それぞれのマイクロホンは同時に同一の信号を受音するのではなく、信号源に近いマイクロホンに対し遠方のマイクロホンは遅延や減衰を受けた信号が観測される。

音は、秒速 340m/s で進むため、マイクロホン間の距離を 200mm 離し、録音のサンプリング周波数を 44.1KHz に設定した場合、

$$\frac{0.2(m)}{340(m/s)} \times 44100(Hz) = 25.94 \quad (2.1)$$

およそ 26 サンプルの遅延が生まれる。この遅延サンプル数を求めることが可能であれば、どちらのマイクロホンがより音源に近いのかを推定することが可能である。

2.3 マイクロホンの種類

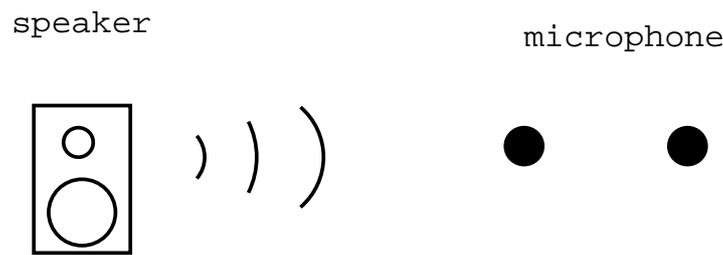


図 2.1 音の到着時間差について

2.3 マイクロホンの種類

マイクロホンには、無指向性、双指向性、単一指向性、鋭指向性、超指向性といったものがある。本研究では、無指向性と単一指向性のマイクロホンに注目する。

2.4 無指向性マイクロホン

無指向性マイクロホンは、置かれた場所に集まった音の全てが振動板に届いて電気出力となる。よって、マイクロホン本体の向きや角度に関係なく、音の大きさだけに反応する性格を持つ。動板の前方だけが音場に対して拡げられている構造で、カセットテープレコーダーなどの内蔵用、インタビュー用や議事収録用コンパクトマイクロホンなどに主に使われる。効果音の生録でも力を発揮する。場所を離して2本立てれば、本来は味わう事のできない独特なステレオ感を創り出す事も可能である。

2.5 単一指向性マイクロホン

単一指向性マイクロホンは、特定の方向をとらえやすい性質を持っている。振動板の後側にも音の通り道として穴や溝が設けられている点が、無指向性との違いである。後方で鳴った音は、まずこの穴や溝から入って振動板の裏側に届く。同じ音は回り込み、少し遅れて振動板の裏側にも届く。そこで、穴や溝から振動板の裏側までに障害物などを置いて間接音の速度を遅らせて直接音と同時に到達するようにすると、この音は振動板の表と裏で同時に生

2.6 単一指向性の性能

じた同量のエネルギーとして相殺され、電気出力にならない。前方で鳴った音は、まず先に振動板の表側に伝わり、その後の裏側への回り込みは、障害物によって到達が更に遅くなる。この時間差によってエネルギーは相殺されずに電気出力される。よって、前方への単一指向性を持つ事になる。

2.6 単一指向性の性能

図 2.2 は、人間の声をサンプリング周波数 44.1kHz で 20 秒間録音したものである。最初の 10 秒間、つまり 441,000 サンプルまでは、指向性マイクロホンを音源側に向けておき、10 秒後に音源を 180 度反対側に移動させることにより、指向性マイクロホンの音源到来方向による入力感度差について調べた。その結果を図 2.3 に示す。10 秒後の入力信号を見ると、指向性を向けた方向以外からの音に対しては、入力が抑圧されていることが確認できる。しかし、指向性を向けた方向以外の音を完璧に消去出来ているわけではない。また、受音目的の音源が移動した場合は、マイクロホンの向きを変える必要がある。

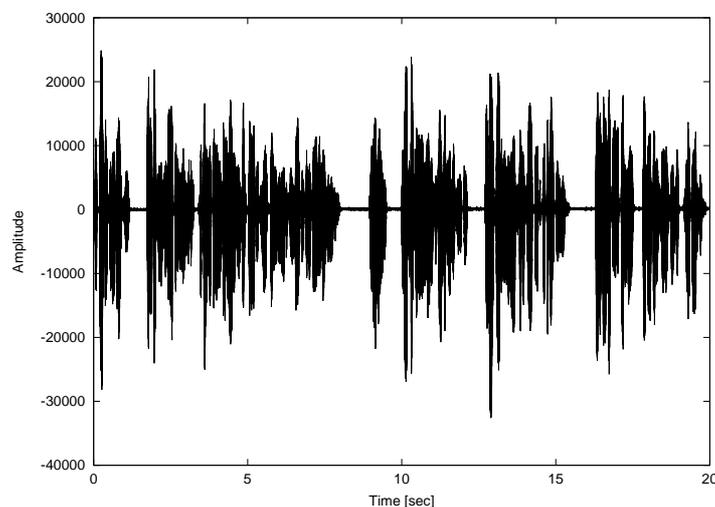


図 2.2 ソース音源

2.6 単一指向性の性能

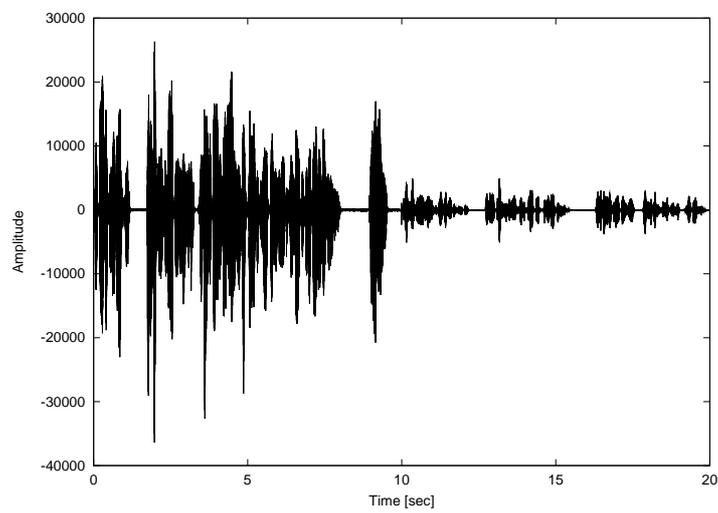


図 2.3 単一指向性マイクロホンの性能

第 3 章

音源が存在する空間の推定

3.1 まえがき

この章では、音源空間推定を行う際に用いる相互相関について説明する。また、音源を単一にした場合の音源空間推定結果ならびに、推定結果を用いた不要音除去性能を指向性マイクロホンの性能と比較する。さらに、音源とマイクロホンの配置を変更した場合の性能変化について検証する。

3.2 音源空間推定

2本のマイクロホンを用い、音源の方向とマイクロホンの位置関係により、それぞれのマイクロホンへ音が到着する時間差が生じる。この性質を利用した音源空間推定について説明する。

音源空間推定システムを、図 3.1 に示す。提案システムでは、相互相関を用いて音源空間推定を行う。一方のマイクロホンへの入力信号を 1 サンプルずつ順にずらし、その都度、相互相関係数を求めていく。相関を取る際、対象信号の平均値を求め、その値を差し引いておくことにより、変化に着目した相関を求めることとする。その後、最も相関係数が大きくなった地点を探索し、その際のサンプルシフト数を得る。このサンプルシフト数が音の到着時間差により発生した時間ずれであると推定できる。

本システムでは、サンプルシフト数が 0 より小さくなった場合は音源がマイクロホン B 側に、サンプルシフト数が 0 より大きくなった場合はマイクロホン A 側に音源があると推

3.3 相互相関

定されていることとなる。また，サンプルシフト数が 0 となった場合は，音源から 2 本それぞれのマイクロホンまでの距離が等しい，または音源が存在しないと推定されていることとなる。

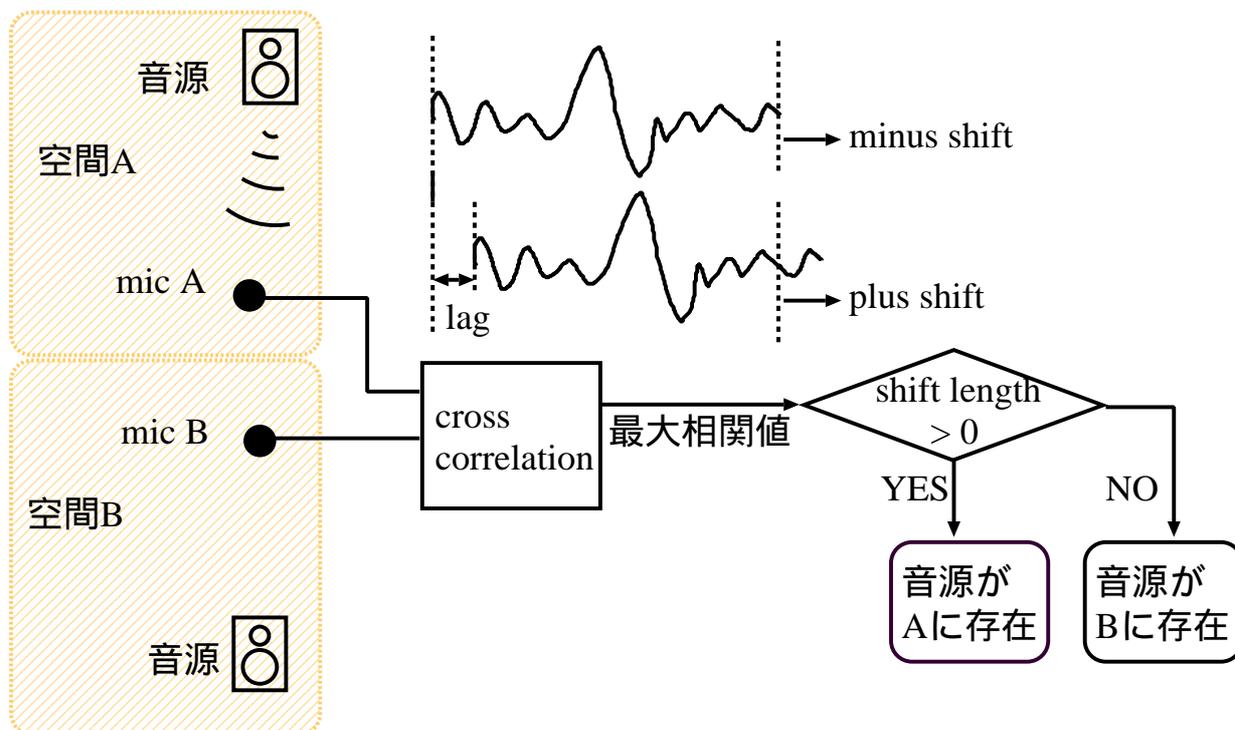


図 3.1 音源空間推定システム

3.3 相互相関

2 つの N サンプルの信号， $f = \{f_0, f_1, \dots, f_{N-1}\}$ ， $g = \{g_0, g_1, \dots, g_{N-1}\}$ の相関値を $R(fg)$ と表せば，

$$R(fg) = \frac{\langle f, g \rangle}{\|f\| \|g\|} \quad (3.1)$$

で定義される。式 3.1 で計算される相関値は感覚的に，2 つの信号の類似度を示すものと考えることができる。例えば，2 つの信号が同一である場合には，自分自身の信号の中に同じような信号パターンの存在を与えることになる。すなわち，ある周期性を示すパラメータを検出することに相当する。

3.3 相互相関

2つの波形の類似性を時間のずれに関係無く評価したいとか、あるいは2つの信号が時間的にどれくらいずれているかを知りたい時に、利用される評価尺度が相互相関係数である。評価したい2つの信号を $f = \{f_0, f_1, \dots, f_{N-1}\}$, $g = \{g_0, g_1, \dots, g_{N-1}\}$ とするとき、その相互相関係数 $\tilde{R}_n^{(fg)}$ は、

$$\tilde{R}_n^{(fg)} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} f_i g_{i+n}}{\sqrt{\frac{1}{N} \sum_{i=0}^{N-1} f_i^2} \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} g_{i+n}^2}} \quad (3.2)$$

と定義される。この式は、信号 f の時間軸はそのままにして、信号 g のみを n サンプルだけ左へずらして得られる信号、すなわち、

$$g^n = \{g_{0+n}, g_{1+n}, \dots, g_{N-1+n}\} \quad (3.3)$$

との内積を計算している。内積は2つの信号間の類似性を評価する尺度であるから、この類似性を時間ずれ n サンプルを変数にもつ係数として表しているわけである。なお、式は次のように変形される。

$$\tilde{R}_n^{(fg)} = \frac{\sum_{i=0}^{N-1} f_i g_{i+n}}{\sqrt{\sum_{i=0}^{N-1} f_i^2} \sqrt{\sum_{i=0}^{N-1} g_i^2}} \quad (3.4)$$

この式で与えられる相互相関係数は、 -1 から $+1$ までの値をもち、正規化された相関値となる。また、あらかじめ信号の平均値を差し引くことにより、直流分を取り除くことができ、変化分に着目した相関を調べることが可能となる。このときの相互相関係数を $\hat{R}_n^{(fg)}$ と表すと、

$$\hat{R}_n^{(fg)} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} \hat{f}_i \hat{g}_{i+n}}{\sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \hat{f}_i^2} \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \hat{g}_i^2}} = \frac{\sum_{i=0}^{N-1} \hat{f}_i \hat{g}_{i+n}}{\sqrt{\sum_{i=0}^{N-1} \hat{f}_i^2} \sqrt{\sum_{i=0}^{N-1} \hat{g}_i^2}} \quad (3.5)$$

で与えられることになる。

3.4 実験環境

相関係数は，2変数の規則的な関係を表している．図 3.1 に，その関係を示す．相関係数の値が大きいほど，強い相関があることを示す．逆に，相関係数の値が 0 に近ければ，ほとんど相関が無いことを示す．

表 3.1 相関係数

$0.0 \leq \ r\ \leq 0.2$	ほとんど相関がない
$0.2 < \ r\ \leq 0.4$	弱い相関がある
$0.4 < \ r\ \leq 0.7$	中程度の相関がある
$0.7 < \ r\ \leq 1.0$	強い相関がある

3.4 実験環境

実験環境を，図 3.2 に示す．マイクロホン間の距離は，100mm，200mm，300mm，400mm の 4 種類に設定する．音源と音源に近い側のマイクロホンとの間隔は 1,200mm に設定する．音源とマイクロホン 2 本を一直線上に並ばせた状態で実験を行う．

ここからは，図 3.3 のように，上側に設置したマイクロホン側の空間を空間 A，下側に設置したマイクロホン側の空間を空間 B と呼ぶこととする．

3.5 室内伝達特性の計測

上記で説明した実験環境において，室内伝達特性の計測を行った．計測は，研究室 (A260) の輪講スペースで行った．ラウドスピーカと無指向性マイクロホンの位置関係は，図 3.2 のように 5 箇所的位置を設定し，インパルス応答の測定を行った．

3.6 シミュレーション方法

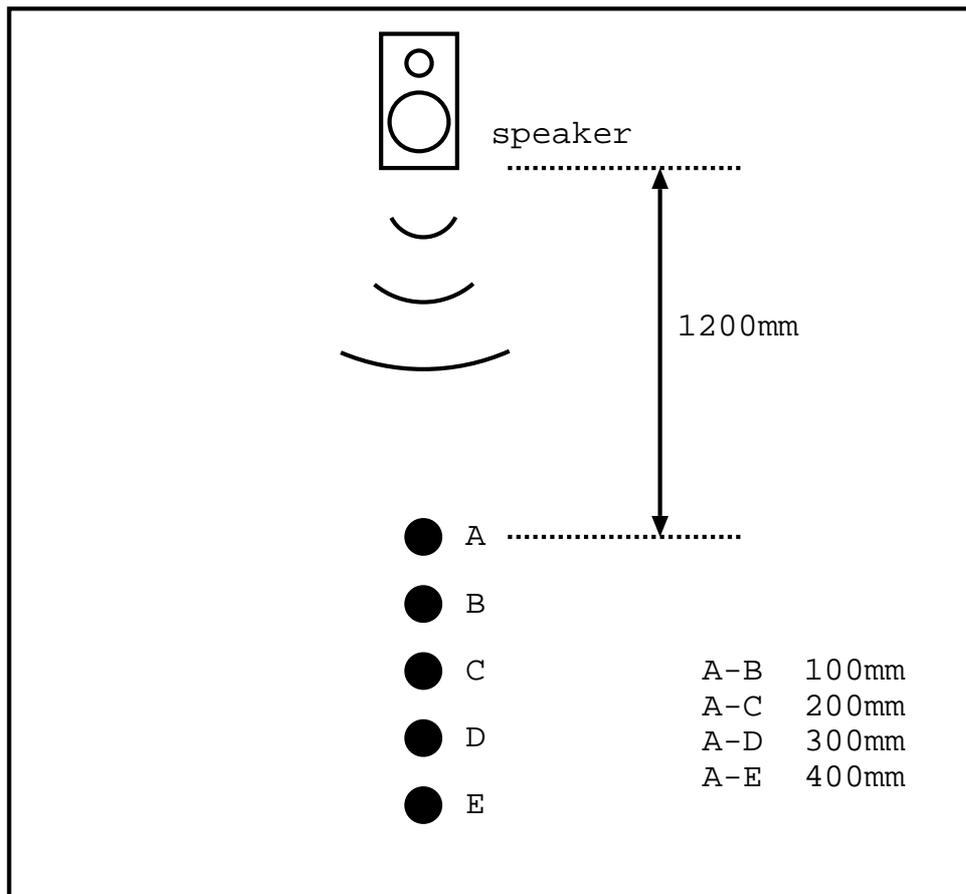


図 3.2 実験環境での伝達環境測定モデル

3.6 シミュレーション方法

入力信号には、図 3.5 に示すサンプリング周波数 44.1kHz の音声信号を用いる。測定した室内伝達特性を用いて、シミュレーション用の信号を作成する。

相関対象信号数を、4,410、8,820、13,230、17,640 サンプルの 4 種類に設定しそれぞれシミュレーションを行う。2 入力信号の相互相関を取る際、相関対象サンプル数の信号を方形窓により切り出した。

$$h(n) = \begin{cases} 1 & (0 \leq n < N - 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (3.6)$$

また、前後の相関対象信号を滑らかにつなぐため、オーバーラップは相関対象サンプル数の 1/2 に設定した。

3.6 シミュレーション方法



図 3.3 実験環境の説明

最初は、音源を空間 A 側に設定し、10 秒後に空間 B 側に音源を移動させるようにした。その際、同時に 2 つの空間に音源が存在しないようにする。これは、向かい合った 2 人が交互に話をしている環境を想定したものである。

これらの音源空間推定結果を用いて、目的空間音以外の音を適応信号処理により抑圧する。適応アルゴリズムには、ステップゲイン 1.0、フィルタ次数 4,098 の学習同定法を用いた。適応フィルタへの入力信号は、音源空間側のマイクロホン入力信号とし、もう一方のマイクロホン入力信号を未知系の出力とした。よって、2 本のマイクロホン間のパラメータ推定が行われる。受音対象空間が推定結果と一致した場合は、適応フィルタの推定出力を出力とし、受音対象空間と推定結果が一致しなかった場合には、推定誤差を出力とする。

シミュレーションでは、受音対象空間を空間 A に設定する。よって、空間 B が音源である音に対しては抑圧処理が行われるべきである。

今回は、音源推定を行うマイクロホンを用いた抑圧処理を行うが、本来はマイクロホンアレイを用いた目的音抽出を行う。本システムの音源推定結果を行い、その音源情報のパラ

3.7 評価方法



図 3.4 実験環境の写真

メータをマイクロホンアレイに提供し，指向性制御に利用する．

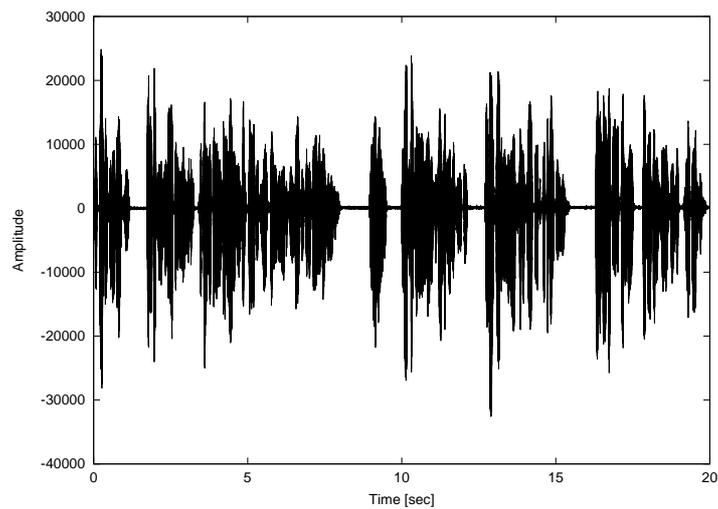


図 3.5 シミュレーションに用いた信号

3.7 評価方法

不要音除去効果は，式 3.7 で求められるシステム入力信号 $x(t)$ とシステム出力信号 $y(t)$ を比較した入出力比を用いる．

入出力比 S は，負の方向に大きいほど不要音の抑圧性能が高いことを示す．シミュレー

3.7 評価方法

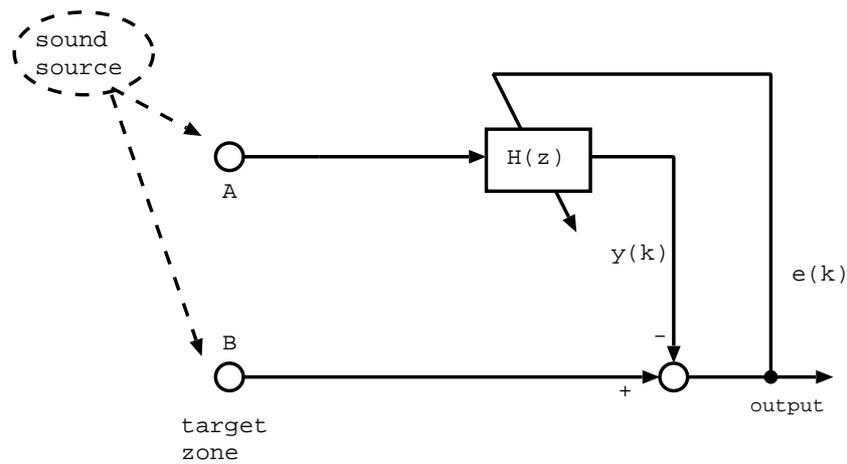


図 3.6 受音対象空間と推定結果が不一致

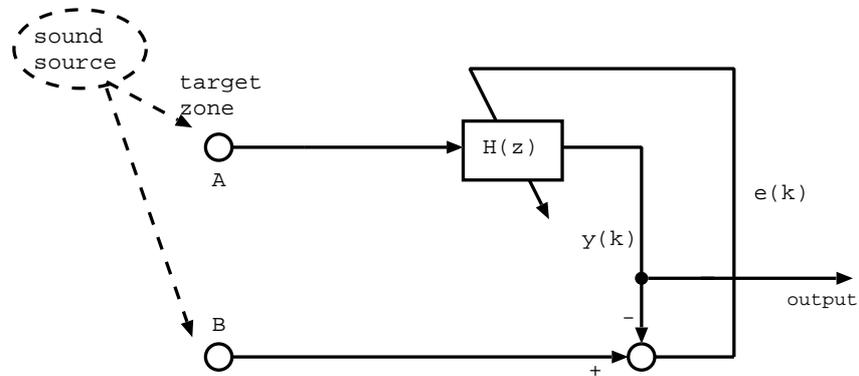


図 3.7 受音対象空間と推定結果が一致

ションでは、空間 A を受音対象空間としているため、開始 10 秒後の音源が空間 B に変化し
てからの抑圧性能について比較を行う。

$$S = 10 \log_{10} \frac{\sum y(t)}{\sum x(t)} [\text{dB}] \quad (3.7)$$

適応信号処理に関しては、音源空間推定結果の性能を示す際の評価を出すための一つの手段として行っているものである。よって、適応信号処理に用いるアルゴリズムやパラメータの変更により評価結果は変化する。また、入出力比が 0dB に近い結果が出た場合は、音源空間の推定誤りによるもの、もしくはシステム同定が追いついていない場合によるものである。

3.8 計算機シミュレーションの結果

相関対象信号を変化させて求められた音の到着サンプル差を，図 3.8 から 3.11 に示す． y 軸はずらしたサンプル数であり，プラス側は音源がマイクロホン A 側，マイナス側はマイクロホン B 側に推定されていることとなる．図 3.8 は，マイクロホン間隔を 100mm に設定した際のシミュレーション結果であるが，正しく推定結果が出ていないことが確認できる．これは，マイクロホンの間隔が近すぎたため，前方のマイクロホンによる影響を受けた信号を後方のマイクロホンが受けたことによるものであると考えられる．

図 3.9 は，マイクロホン間隔を 200mm に設定したものであるが，100mm でのシミュレーションと比べると，明確に強いピークが出ていることが確認できる．相関対象信号数は，多いほど正しい音源空間推定が可能となっている．2 本のマイクロホンの間隔を離す毎に，強いピークが無くなっていき，推定結果は悪化していくことを確認した．

推定結果を用いて不要音除去を行った結果を，表 3.2 から 3.4 に示す．マイクロホン間隔を 100mm に設定した際の不要音除去については，音源空間推定に失敗しているため，行わないこととする．

表中の指向性は指向性マイクロホンにおける抑圧効果を示しており，これよりも値が小さい場合に，提案手法の抑圧効果の方が大きいといえる．相関対象信号数が 4,410 サンプルの場合は，音源空間推定に誤判定が多いため，入出力比 S の値は，0dB に近づいてしまう結果となっているが，それ以外の相関対象信号数での場合は，良好な結果が出ている．

3.9 マイクロホンと音源の配置を変更した場合の変化

音源とマイクロホンの角度を 45 度に設定し，同様に音源空間推定を行う．これにより，音源とマイクロホンが一直線上に並んでいない状態であっても，同様の推定結果を得ることができるか検証する．音源とマイクロホンの配置については，図 3.12 のようにする．マイクロホン間距離は 200mm に設定した．

シミュレーション結果を図 3.13 に示す．音源とマイクロホン 2 本が一直線上に並んでい

3.9 マイクロホンと音源の配置を変更した場合の変化

表 3.2 マイクロホン間 200mm での性能比較

Sample Number	指向性	4,410	8,820	13,230	17,640
445,410	-19.66	-20.51	-18.91	-19.36	-18.16
533,610	-19.62	-17.91	-19.90	-19.81	-19.16
621,810	-19.08	-16.27	-17.11	-16.11	-17.13
710,010	-18.26	-25.26	-24.70	-25.00	-25.60
798,210	-20.75	1.17	-24.50	-24.86	-25.04

表 3.3 マイクロホン間 300mm での性能比較

Sample Number	指向性	4,410	8,820	13,230	17,640
445,410	-19.66	-18.30	-19.49	-16.96	-18.61
533,610	-19.62	-20.79	-20.69	-18.52	-19.00
621,810	-19.08	-17.46	-17.11	-16.98	-16.82
710,010	-18.26	-24.23	-24.63	-24.98	-24.39
798,210	-20.75	4.51	-26.68	-26.22	-26.59

表 3.4 マイクロホン間 400mm での性能比較

Sample Number	指向性	4,410	8,820	13,230	17,640
445,410	-19.66	-16.19	-15.39	-18.71	-16.05
533,610	-19.62	-17.80	-16.83	-16.63	-16.43
621,810	-19.08	-16.43	-16.27	-15.47	-15.62
710,010	-18.26	-31.47	-31.09	-29.82	-30.50
798,210	-20.75	-27.15	-24.96	-26.61	-27.05

3.9 マイクロホンと音源の配置を変更した場合の変化

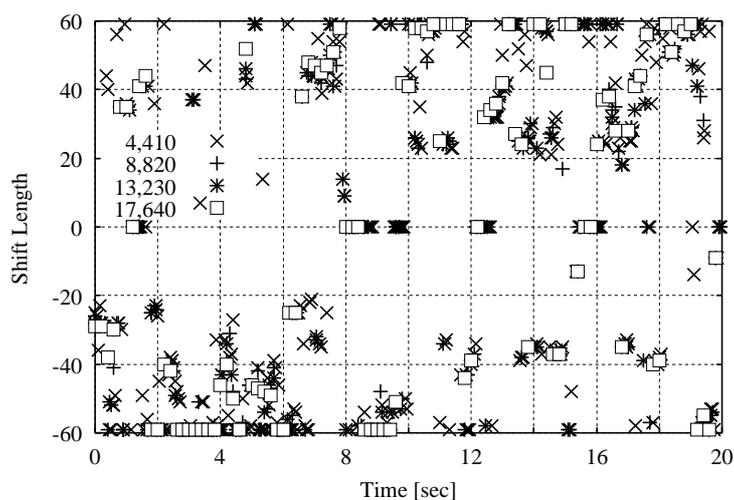


図 3.8 間隔 100mm での推定結果

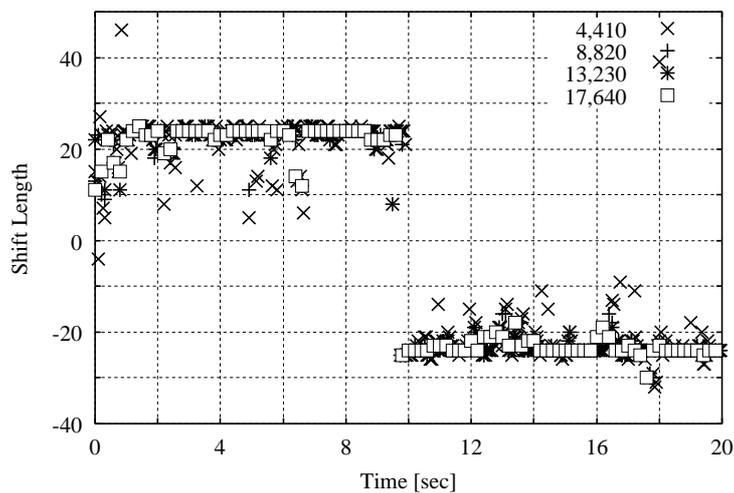


図 3.9 間隔 200mm での推定結果

ない場合においては、マイクロホン間隔が 100, 200mm の場合に推定が可能であることを確認した。しかし、間隔が 300mm 以上となると、推定が不可能であった。これらの結果から、マイクロホンの配置方法を変更した場合においても、音源空間の推定が可能であることを確認した。

3.9 マイクロホンと音源の配置を変更した場合の変化

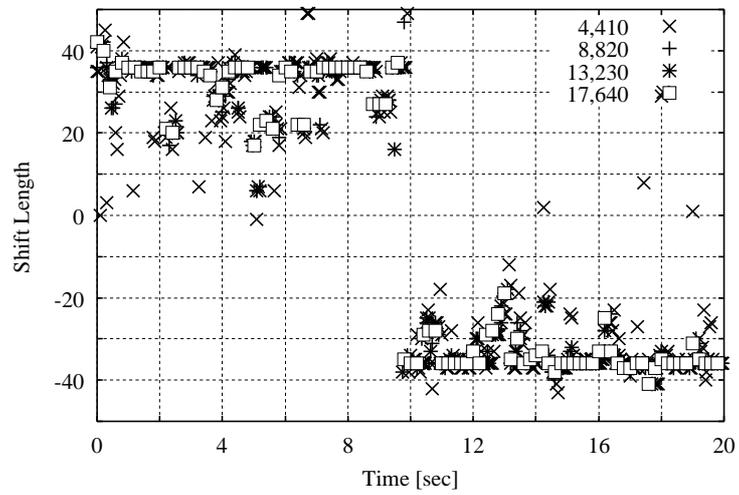


図 3.10 間隔 300mm での推定結果

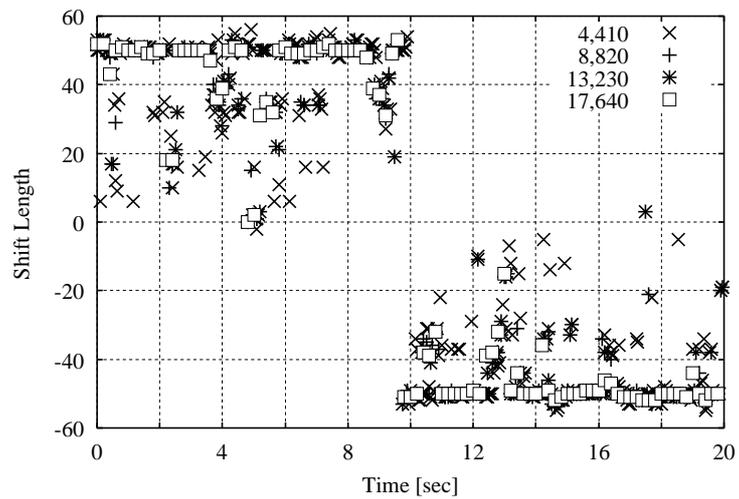


図 3.11 間隔 400mm での推定結果

3.9 マイクロホンと音源の配置を変更した場合の変化

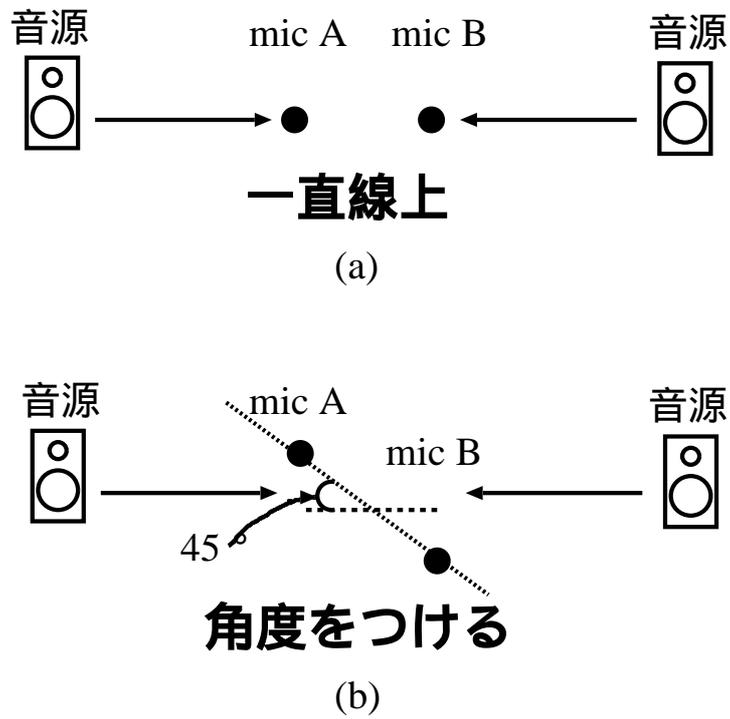


図 3.12 音源とマイクの配置

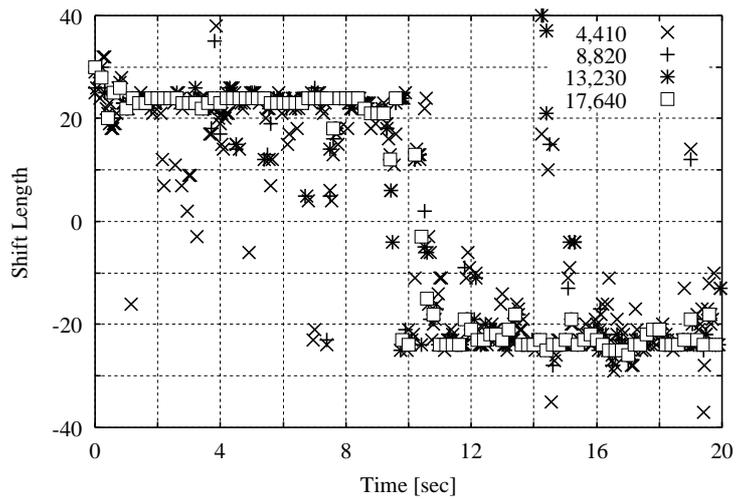


図 3.13 間隔 200mm45deg での推定結果

第 4 章

音源が複数存在する場合の性能評価

4.1 まえがき

前章で示した音源が単一である場合の音源空間推定法について示した。しかし、実際にシステムとして動作させる場合は、さまざまな方面からの雑音がマイクロホンで集音されることとなる。本章では、音源を同時に複数存在させるという設定条件においての音源空間推定法とその評価について示す。

4.2 音源が複数存在する場合の音源空間推定法

図 4.1 に、システム構成を示す。まず、マイクロホン B の観測信号をシフトさせ、相関値の最大値となるサンプルシフト数を求める。この動作を、プラス方向へのシフトと呼ぶこととする。次に、マイクロホン A の観測信号をシフトさせ、同様に相関値が最大値となるサンプルシフト数を求める。この動作を、マイナス方向へのシフトと呼ぶこととする。求めたそれぞれのサンプルシフト数は、プラスシフトが空間 A、マイナスシフトが空間 B の音源空間推定結果となる。空間に音源がないと推定された場合は、推定結果が 0 となり、音源があると判断された場合は、それ以外の値となる。よって、別々の空間にいる 2 人が同時に声を出した場合でも、それぞれの空間に音源が存在していることを推定することが可能となる。

4.3 シミュレーション方法

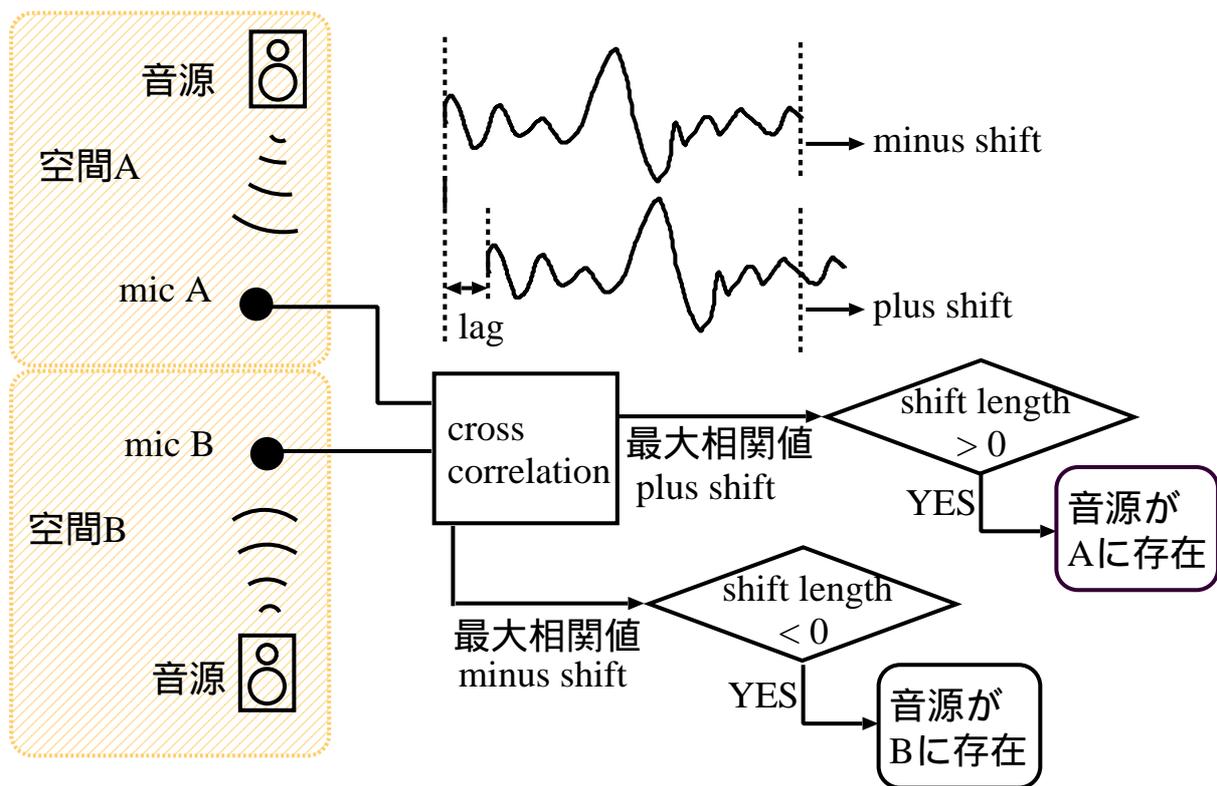


図 4.1 複数音源が存在する場合の音源空間推定法

4.3 シミュレーション方法

入力信号には、サンプリング周波数 44.1kHz の音声信号を用いる。実験環境は A503 とし、音源を単一とした際よりも反響音の多い環境での検証を行った。これは、実際に本システムが利用される環境を想定して、有効性を確認するためのものである。

図 4.4 と 4.5 に、それぞれのマイクの観測信号について示す。

相関対象信号数を 13,230 サンプルに設定し、マイクロホン間隔は単一音源音源でのシミュレーションと同等の条件でシミュレーションを行う。2 入力信号の相互相関を取る際、相関対象サンプル数を方形窓により切り出し、4,410 サンプル毎に推定結果を算出するように設定した。相関対象区間のシフトは、プラス方向、マイナス方向それぞれ 70 サンプルまでとする。

複数の音源が同時に存在する場合のシミュレーション方法について説明する。最初の 10

4.3 シミュレーション方法

秒間は、空間 A で 1 人が声を出し、10 秒後に空間 B から 1 人が声を出し始める状況を設定した。よって、10 秒後からは同時に 2 人が向かい合って話している環境を想定したものとなっている。このような設定環境において、A と B それぞれの空間から同時に音が出された場合でも音源空間の推定が行えることを確認する。本シミュレーションでは、最初の 10 秒間は空間 A 側、10 秒以降は両空間に音源判別結果が現れることが期待される。

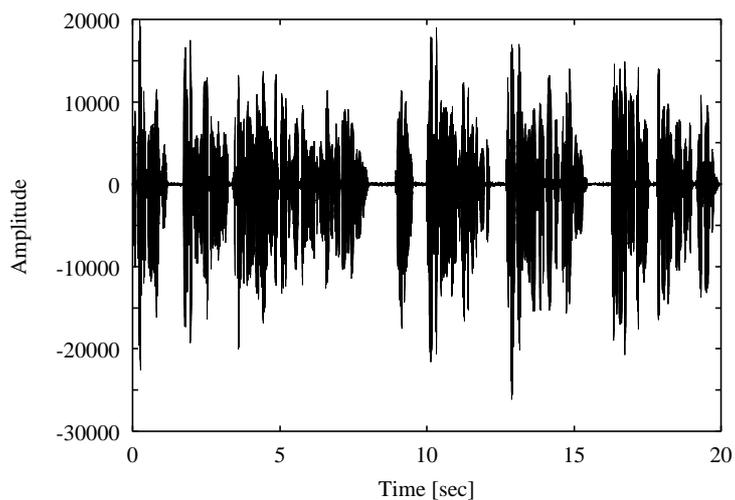


図 4.2 シミュレーションに用いた信号 (1)

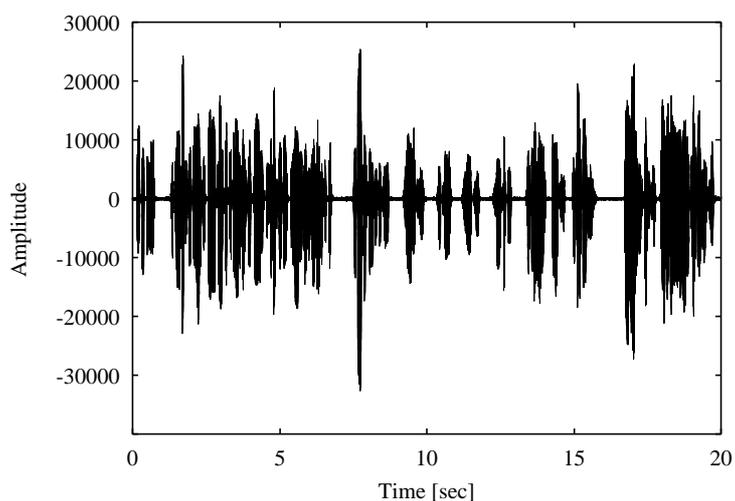


図 4.3 シミュレーションに用いた信号 (2)

4.4 評価方法

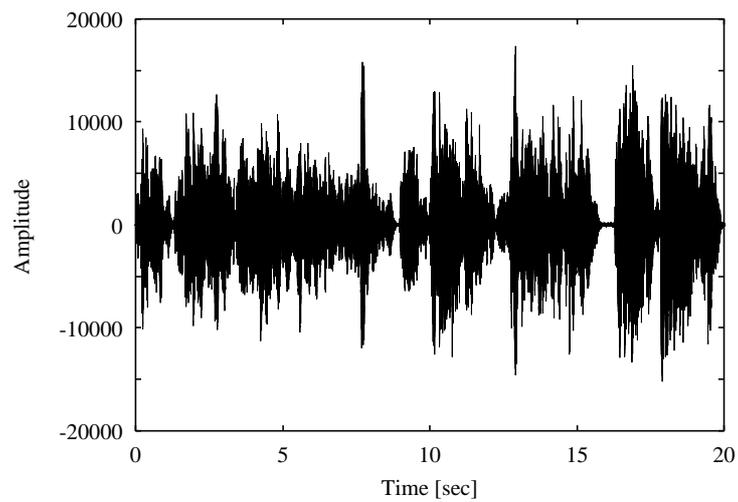


図 4.4 マイク A の観測信号

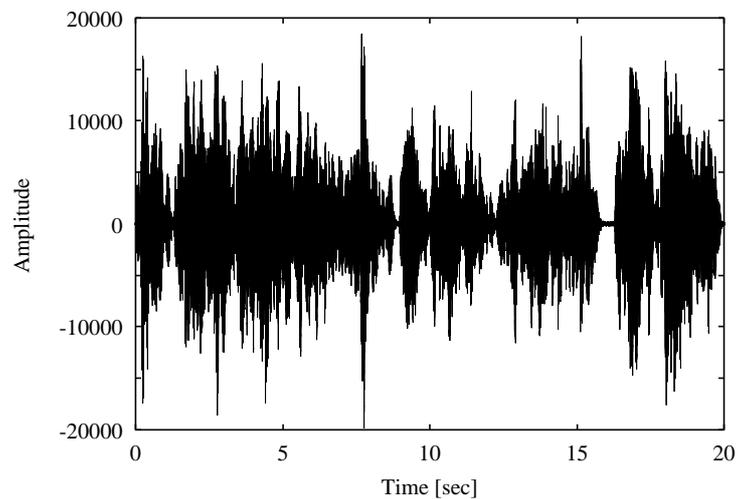


図 4.5 マイク B の観測信号

4.4 評価方法

マイクロホン間の距離，音源とマイクロホンの配置方法，また相関対象サンプル数を変化させることにより，推定性能の比較を行う．推定性能については，開始 10 秒以降の推定正解率により比較する．推定正解率は，正しい音源判別結果数を全判別数で割ることにより求められる．

4.5 計算機シミュレーションの結果

表 4.1 から 4.3 に、マイクロホン間距離と配置方法、相関対象区間を変化させた際の推定正解率について示す。図 4.6 から 4.12 に、相関対象区間を 4,410 サンプルに設定した際のシミュレーション結果と、推定するにあたって用いた相関値分布を示す。この相関値は、シフトした際に最も大きい相関値となった部分について表している。

図中の“shift B”がマイク B の入力信号をシフトさせた際に、最大の相関値を得た音源推定結果である。“shift A”は、マイク A の入力信号をシフトさせた際に、最大の相関値を得た音源推定結果である。最初の 10 秒間までは、プラス側に判定され、マイナス側にはほとんど判定結果が現れていない。10 秒以降は、空間 A, B それぞれに推定結果が現れていることが確認できる。推定結果が 0 となっている部分は、音源がその空間に無く、もう一方の空間に存在するため、サンプルシフトさせない時に最も相関値が高くなるのが原因によるものである。よって、推定結果が 0 である場合は、その空間に音源が存在しないと推定されていることとなる。また、一直線上にマイクロホンを設置した場合においては、200, 300, 400mm, 45 度の角度をつけて設置した場合においては 300, 400mm での推定が正常に行えなかった。図 4.12 は、マイクロホン間隔 200mm で一直線上に配置した場合の音源判別結果を示したものであるが、ほとんどの判別結果がシフト幅のプラス方向、マイナス方向それぞれの最大値である 70 に集まっていることが確認できる。これは、本来とは違うシフト幅において相関値の最大値が現れたことが原因であり、正しく音源判別ができていないといえない。

これらの結果は、音源を単一とした場合のシミュレーション時と比べ、マイクロホン間隔を広く取ることができなかった。これは、本章の最初に述べたように、実験環境の変化によるものである。マイクロホン間の距離は、大きく取り過ぎると相関が取れなくなることから、できる限り間隔を狭めることで、あらゆる環境に対応できるといえる。

これらの音源推定結果を、正常に推定が行えた条件においてのみ、推定正解率により性能比較する。相関対象区間を短くした方が両空間の音源判別の正解率が高くなるといった結

4.5 計算機シミュレーションの結果

果が出た。また、相関対象区間を広く取る毎に、一方の空間の推定正解率が高くなり、もう一方の空間の正解率が低くなった。これは、一方の空間から出す音において、音量が小さい部分が多い場合には、もう一方の空間から出ている音の影響を受けやすくなる。よって、相関を取った際にシフトさせない場合が最も相関値が高いといった結果が出てしまうためである。相関対象区間を広く取った場合は、音の小さい区間が短期間に連続して発生しない音、例えば早口で話している話者側に音源が判別されやすくなる。しかし、相関対象区間を小さく取った場合には、その影響を抑えることができる。このため、相関対象区間は短い方が正解率が高くなったといえる。

また、両空間それぞれの声の大きさによっても判別結果が異なってくる。同時に両空間の話者が発言した場合、声が大きい話者がいる空間側に相関結果の最大値が現れやすい傾向があるといえる。

表 4.1 マイク間 100mm45deg での性能比較

相関対象区間	空間 A	空間 B
4,410 (0.10sec)	0.54	0.61
8,820 (0.20sec)	0.45	0.66
13,230 (0.30sec)	0.42	0.68
17,640 (0.40sec)	0.42	0.67

表 4.2 マイク間 100mm 一直線上での性能比較

相関対象区間	空間 A	空間 B
4,410 (0.10sec)	0.56	0.58
8,820 (0.20sec)	0.45	0.64
13,230 (0.30sec)	0.41	0.67
17,640 (0.40sec)	0.42	0.67

4.5 計算機シミュレーションの結果

表 4.3 マイク間 200mm45deg での性能比較

相関対象区間	空間 A	空間 B
4,410 (0.10sec)	0.65	0.69
8,820 (0.20sec)	0.59	0.74
13,230 (0.30sec)	0.54	0.77
17,640 (0.40sec)	0.52	0.75

相関値の時間的変化を見ると、最初の 10 秒間はそれぞれの方向にシフトした際の相関値に大きな差が表れている。よって、音源判別において一方の空間のみに判別結果が表れているといえる。しかし、10 秒後からは 2 値に大きな差が表れず、また全体的に相関値が小さくなっていることが確認できる。これは、同時に 2 箇所音源が存在することによる影響である。そのため、判別結果においても両空間に判別結果が表れているといえる。

また、マイクロホン間隔の大きい方が、プラス方向、マイナス方向それぞれにシフトさせた際の最大相関値の差が広がっていくことも確認できる。この差が大きい方が音源判定結果に良い結果をもたらすことが確認できた。

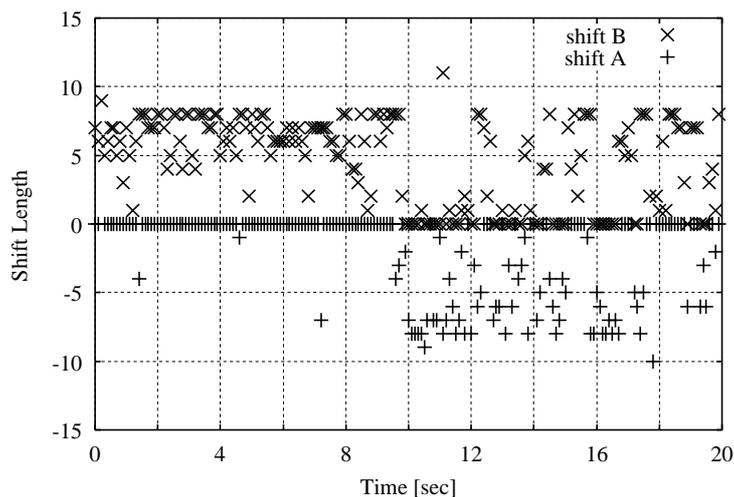


図 4.6 間隔 100mm45deg での推定結果 (相関対象 4,410 サンプル)

4.5 計算機シミュレーションの結果

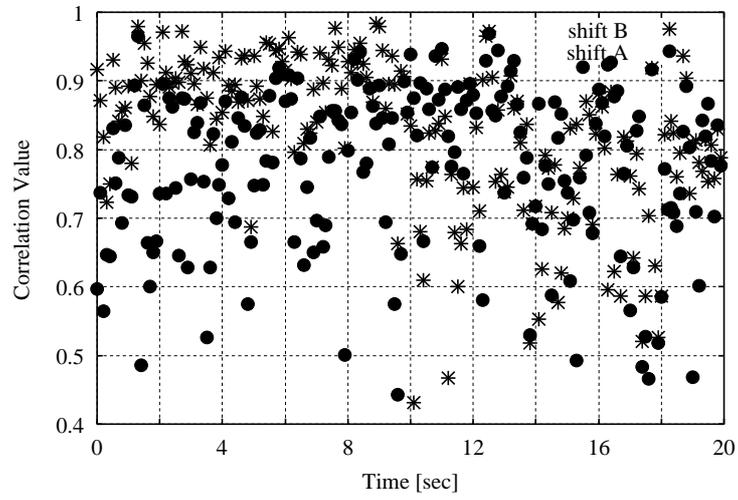


図 4.7 間隔 100mm45deg での相関値の時間的变化 (相関対象 4,410 サンプル)

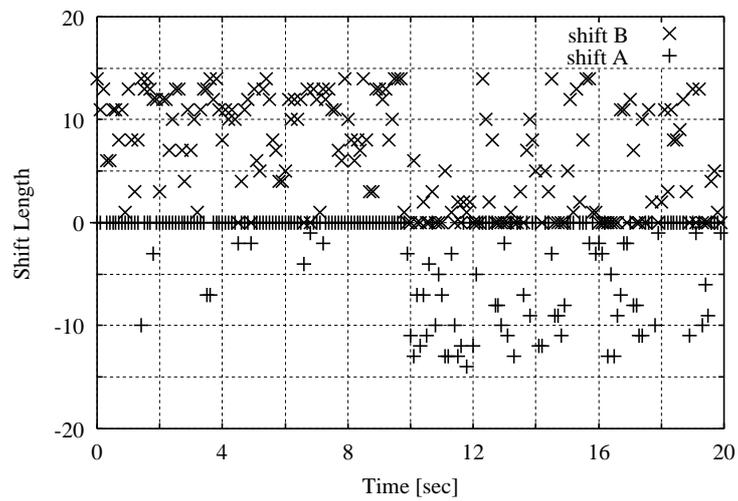


図 4.8 間隔 100mm 一直線上での推定結果 (相関対象 4,410 サンプル)

4.5 計算機シミュレーションの結果

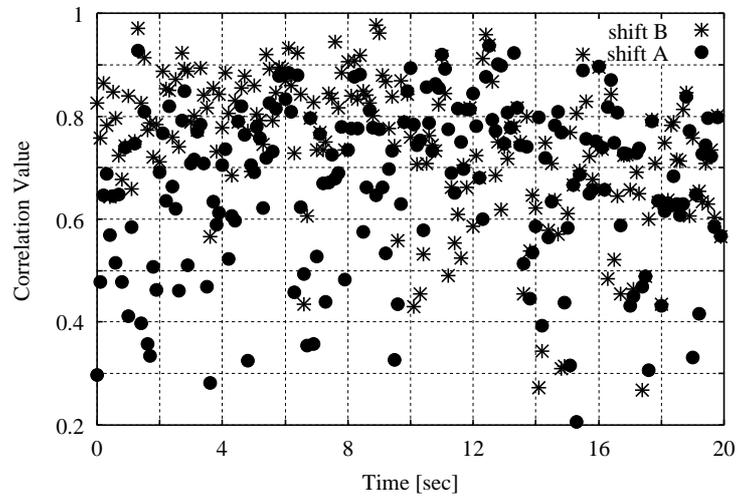


図 4.9 間隔 100mm 一直線上での相関値の時間的变化 (相関対象 4,410 サンプル)

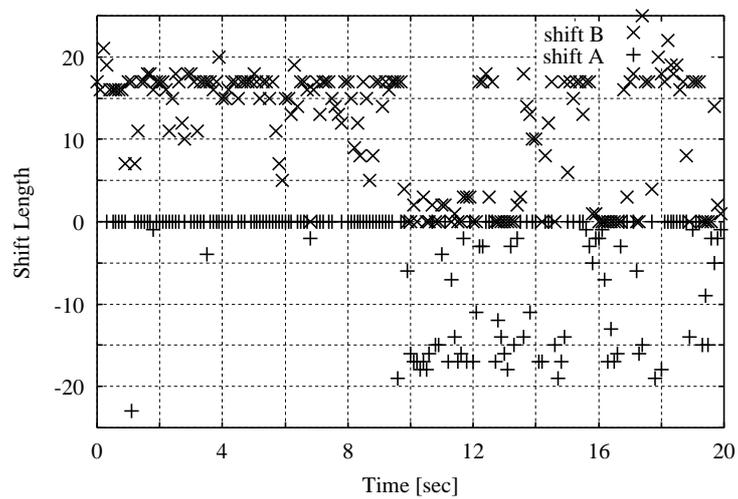


図 4.10 間隔 200mm45deg での推定結果 (相関対象 4,410 サンプル)

4.5 計算機シミュレーションの結果

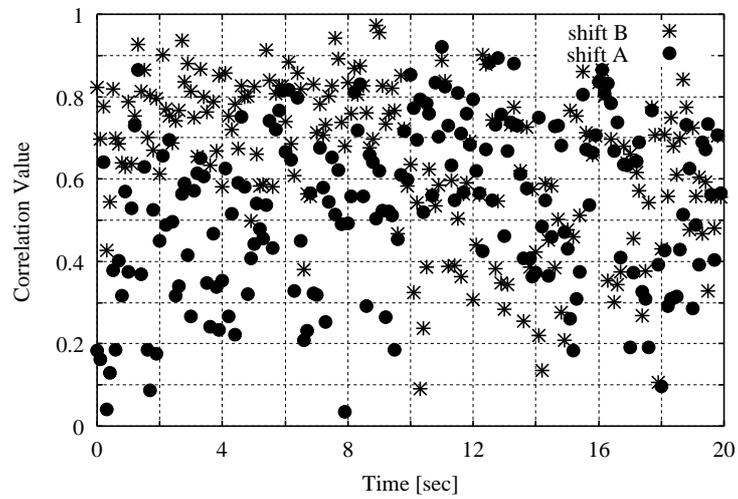


図 4.11 間隔 200mm45deg での相関値の時間的变化 (相関対象 4,410 サンプル)

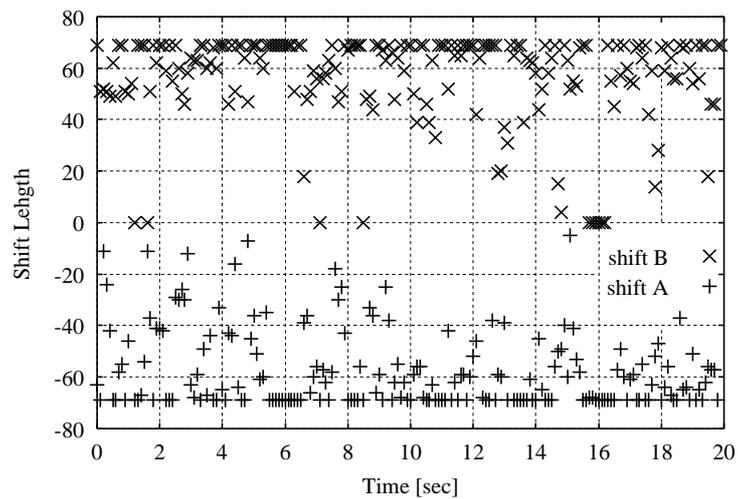


図 4.12 間隔 200mm 一直線上での推定結果 (相関対象 4,410 サンプル)

第 5 章

結論

5.1 本研究のまとめ

本研究により、マイクロホン 2 本を用いた音源空間推定が提案システムにより可能であることを示した。音源が単一である場合については、受音対象空間でない場合に適応信号処理による不要音除去性能が指向性マイクロホンの性能よりも高いことを示すことにより、音源空間推定結果が指向性の制御に利用可能であることを確認した。今回は、音源推定を行ったマイクロホンによる不要音除去を行ったが、マイクロホンアレイの指向性制御に本システムの音源推定結果を利用することにより、より性能を向上させることが可能である。

また、音源が複数である場合については、同時に音源が複数存在しても音源空間推定を行うことが可能であることを示した。このシミュレーションでは、音の反響しやすい空間を想定したものであったため、実際の会議等でも利用可能であると考えられる。

提案手法により求めた音源空間推定結果は、物理的な操作を伴わない指向性制御が可能であるマイクロホンアレイへ与えるパラメータとしての利用が考えられる。

5.2 今後の課題

今回は対象とするゾーンの分割を最も簡単である 2 つとしているが、使用するマイクの本数は増えるものの、分割数を増やすことでより精度の高い音源推定を録音対象が移動するような環境下においても簡単に実現できるものと考えられる。

謝辞

本研究を行うにあたって、夜遅くまで御指導して頂いた福本昌弘准教授に心から感謝します。

また、本研究の審査をして頂いた坂本明雄教授、吉田真一講師にも感謝致します。

佐伯幸郎氏、福富英次氏には、プレゼンのアドバイスなど様々な面で御世話になり、とても感謝しております。特に佐伯氏には、研究活動以外の面でも御世話になりました。巨人ファンである私を広島カープの試合観戦に2回も連れて行っていただきました。おかげさまで、少し広島カープに詳しくなっていました。次こそ、巨人に絡んだ試合に連れて行ってほしいです。

同4年生の山田君、小林君、野村君には、散々ハバネロパウダーの餌食になってもらい、日頃溜ったストレスの発散をさせていただきました。この中でも一番の被害者である山田君には、どれだけ御世話になったかわかりません。このパウダーの提供者である小林君には、とても感謝しています。また、野村君は運悪く私の隣の席を選んだため、毎日のようにしょうもないちょっかいが入ったと思いますが、飽きずによく相手をしてもらいました。このメンバー無くして、卒論を完成させることはできなかったと思っています。

最後に、私の大学生活に関わって頂いた全ての方々に感謝致します。

参考文献

- [1] 永田仁史, 阿倍正人, “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 (A), vol.J82-A, no.6, pp.860-866, Jun, 1999
- [2] 永田仁史, 藤岡豊太, 阿倍正人, “二つの指向性マイクロホンを用いた目的音検出に関する検討”, 信学論 (A), vol.J83-A, no.12, pp.1445-1454, Dec, 2000
- [3] 辻井重男, “適応信号処理”, 昭晃堂, 1995
- [4] 三谷政昭, “やり直しのための信号数学”, CQ 出版社, 2005

付録 A

適応信号処理

A.1 パラメータ推定問題

入出力データからその未知システムの構造とパラメータを推定することをシステム同定と呼び、適応信号処理はその大半がパラメータ（インパルス応答）推定問題として取り扱われる。

A.2 適応アルゴリズム

LMS アルゴリズムは、2乗平均誤差を最急降下に基づいて最小にする方式で、安定感があり、演算量が少ないという特徴から代表的な適応アルゴリズムとなっている。一方、1967年に野田と南雲が、学習同定法を開発した。これは、先に述べた LMS アルゴリズムに比べると複雑ではあるが、収束速度が入力信号の大きさに依存しないという特徴を持っており、実用的にも優れた適応アルゴリズムといえる。これらのアルゴリズムは、推定すべきパラメータの変化にある程度追従できる特徴がある。しかし、入力信号が有色である場合、収束速度が著しく劣化するといった欠点もある。また、1960年、Kalman により離散時間カルマンフィルタが提案された。カルマンフィルタにおいて、状態変数を推定すべき未知パラメータとし、このパラメータが時間的に変動しないと仮定すると、このカルマンフィルタはよく知られた逐次最小 2 乗アルゴリズム (RLS) アルゴリズムと一致する。RLS アルゴリズムは、推定すべきパラメータの個数を N とすると、1 サンプルあたり N^2 に比例する回数の乗算を必要とする。LMS アルゴリズムや学習同定法の乗算回数が N に比例するのと比較

A.3 学習同定法

すれば，RLS アルゴリズムは演算量からリアルタイム処理は困難といえる．

A.3 学習同定法

学習同定法は，別名 NLMS アルゴリズムと呼ばれ LMS アルゴリズムのパラメータ修正項をフィルタの状態ベクトルで正規化されたものとみなすことができる．時刻 t における適応フィルタの出力 $y(t)$ が未知システムの出力 $d(t)$ に等しいとすると，

$$d(t) = h_N^T x_N(t) \quad (\text{A.1})$$

と表すことができる．

しかし， $h_N = w_N$ を満たすためには，すべての入力信号 $x(t)$ に対して式 A.1 が成り立たなければならない．

そこで，式 A.1 を満たす解集合の代表ベクトルを $h_N(t)$ とする．この解集合は式 A.1 より，入力ベクトル $x_N(t)$ に直交しているといえる．更に， w_N はこの解集合に含まれているので， $h_N(t)$ はある点から $x_N(t)$ 方向にパラメータ修正したとき，最も w_N に近い点といえる．

したがって， $h_N(t)$ を $w_N(t)$ に更に近づけるためには，適当に定めたある点よりも w_N により近い $h_N(t+1)$ を次の修正パラメータの初期値とすれば良い．以上のことより，

$$\begin{aligned} h_N(t+1) &= h_N(t) + \{h_N(t+1) - h_N(t)\} \\ &= h_N(t) + \frac{\{w_N - h_N(t)\}^T \{h_N(t+1) - h_N(t)\}}{\|h_N(t+1) - h_N(t)\|} \\ &\quad \times \frac{h_N(t+1) - h_N(t)}{\|h_N(t+1) - h_N(t)\|} \end{aligned} \quad (\text{A.2})$$

となる．ただし， $\|\cdot\|$ はベクトルのユークリッドノルムを表し，要素の 2 乗和の平方根と定義する．ここで，

$$\frac{h_N(t+1) - h_N(t)}{\|h_N(t+1) - h_N(t)\|} = \frac{x_N(t)}{\|x_N(t)\|} \quad (\text{A.3})$$

$$\begin{aligned} \{w_N - h_N(t)\}^T x_N(t) &= d(t) - y(t) \\ &= e(t) \end{aligned} \quad (\text{A.4})$$

A.3 学習同定法

が成立するので，式 A.2 は，

$$h_N(t+1) = h_N(t) + \frac{x_N(t)}{\|x_N(t)\|^2} e(t) \quad (\text{A.5})$$

のように変形できる．学習同定法は，式 A.5 の修正ベクトルにステップゲインを掛け

$$h_N(t+1) = h_N(t) + \alpha \frac{x_N(t)}{\|x_N(t)\|^2} e(t) \quad (\text{A.6})$$

で与えられる．