修 士 論 文

多数マイクロホンによる音源方向推定に関する研究

指導教官　赤木 正人 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

西田 知之

1999 年 8 月 13 日

目 次

図 目 次

表 目 次

第 1 章

序論

1.1　本研究の背景

音源の定位とは音源の発する音から、音源の位置情報を得ることをさす。我々人間は普段から頻繁にこの音源定位を行なっている。例えば、我々は後ろから呼びかけられた時それに呼応して、自然に振り向く。この時に確かに我々は音の位置情報を判断している。このように、音から位置情報を判断するという動作は、限られた範囲の情報をしか得ることのできない視覚とは違い、周囲の全方向に対する環境認識を行なうことができる。この動作は、人間だけでなく常に周囲の環境を把握し、危険や天敵から身を守るための早期警戒網として重要な役割を担うものとして、地上に住む動物全てが持つ重要な能力である。

このような能力を工学的に実現することは、我々人間などが行なっているのと同様に、ロボットなどの環境認識システムへの応用、音情報と位置情報を同時にとることによって、臨場感のある遠隔会議システムの構築、音源位置に自動的にカメラの照準を合わせる監視カメラシステムなど、様々な分野に応用が考えられている。このため、正確に、そして素早く音源方向の推定を行なうための手法が求められている。

1.2　従来の研究

音源方向推定については現在まで様々な手法が提案され、研究が行なわれている。その中で最も一般的なのが、複数のマイクロホンを配置したアレイを用い、アレイの各々のマイクロホンに入る音の時間差を元にした手法である。

1.2.1 マイクロホンアレイ

アレイの形状についても様々なものが提案され、用いられており、円形に多数のマイクロホンを配置したものや [1] 、格子状にマイクロホンを配置したものなどがある [2]。特に、このように非常に多くのマイクロホンを配置したアレイは、音源分離などと合わせた研究分野において多く見られる傾向がある [2][3]。また、比較的少数のマイクロホンを用いたアレイでは、一般に直線上にマイクロホンを配置したアレイが良く見られる [4]。人間の聴覚では二つの耳が直線上に並んでいるものと見ることができることから、それに即した形状といえ、また構成する部品点数も少なく、入力も少ないことから扱い易いとされる。しかし、人間の聴覚では音源方向推定において様々な処理を行なっているため、このような配置であっても非常に優秀な能力を発揮しているが、マイクロホンをただこの様な配置にしただけでは、前後の判断ができないなど問題点が発生することになる。

1.2.2 時間差測定法

我々人間は様々な手がかりを元にして音源方向の推定を行なっている。主に、両耳間時間差（ITD:interaural time difference）や、両耳間音圧差（ILD:interaural level difference）などがそれにあたる。従来の研究では主に、音圧差に比べ扱い易いという観点から、時間差が用いられている。そして、この時間差を検出するために、相互相関を利用した手法多く用いられている [5]。

相互相関の式は次のように定義される。

$$\phi_{xy} = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)y(t)dt \tag{1.1}$$

ここで、x(t)、y(t) が入力信号。ϕ_{xy}がその相互相関になる。

この手法は、伝達遅れ時間を測定する場合などに用いられる。つまり、一方を入力信号、もう一方を出力信号としてその相互相関をとり、その相関値が大きくなった時間差を遅れ時間と見るわけである。

それは、相互相関を利用した時間差測定の際も同じであり、あるマイクロホン対で、一方の入力を元にし、もう一方の入力の中から同様の信号が入力されるまでの遅れ時間を得ることで、一組のマイクロホン対の中で生じる時間差を測定することができる。しかし、この手法は残響が含まれた環境ではその能力が著しく低下するという問題点がある。

1.3　残響

残響とは、その名の通り音の響きが残る現象のことを指し、古くから知られている音響現象の一つである。

我々人間は、日常残響の存在する環境で生活している。残響は我々の日常生活において重要な役割を果たしており、残響の全く存在しない環境において発声された音声は非常に不自然に感じる。これは、人間の聴覚機構がこのような環境に対する優秀な性能をもっているためである。

現在機械による音情報処理の分野は様々な方面において行なわれている。このようなシステムの実際の応用において、実環境、つまり残響などを含む環境においての処理が重要である。しかし、機械による音処理を行なう場合残響が大きな影響を持っている。残響の含まれた環境による音はいわば情報の混じりあった状態であり、このような情報を分離すること自体がまず困難な問題となる。更に、残響は相関を用いて時間差の測定を行なう上で、その性能に大きな影響を与える特徴を有している。

一般に閉空間において観測される音場は、音源からの直接音と周囲の物体、床、壁面などからの反射音から構成される。この反射音が、残響である。このことを念頭において、直接音、残響の特性を考えてみる。

実際の環境での音の聞こえ方（直接音、反射音を含む）のモデルを図 1.1 に示す。この図を見ると、直接音はその名の通り、直接音源から観測者に到達する音であり、最短距離を通って到達する音である。一方、残響は周囲の壁等に反射して到達するそのため、反射音は直接音よりも長い経路を通過して観測者に届くことになる。

従って音速が一定である室内を考えると、反射音（残響）は以下のような特徴を有することになる。

- 反射音は直接音に比べ遅れて到達する
- 反射音は直接音に比べてそのパワーが小さくなる

つまり、経路長が長くなる分だけ到達が遅れ、また経路長や反射の影響などによりパワーが小さくなるわけである。

ここで、先ほどの相関を用いた時間差検出手法について考えてみる。相関を用いた時間差の検出手法では、あるマイクロホンに入力された信号を元に、他のマイクロホンに入力された信号内を検索し、その中で元になった入力と同様の信号が入力された場所を捜し出すことによって、時間差の検出を行なっている。残響は直接音に対して遅れて到達し、そのパワーが小さいという特徴を持っている。しかし、全体的にパワー成分の減少は生じて

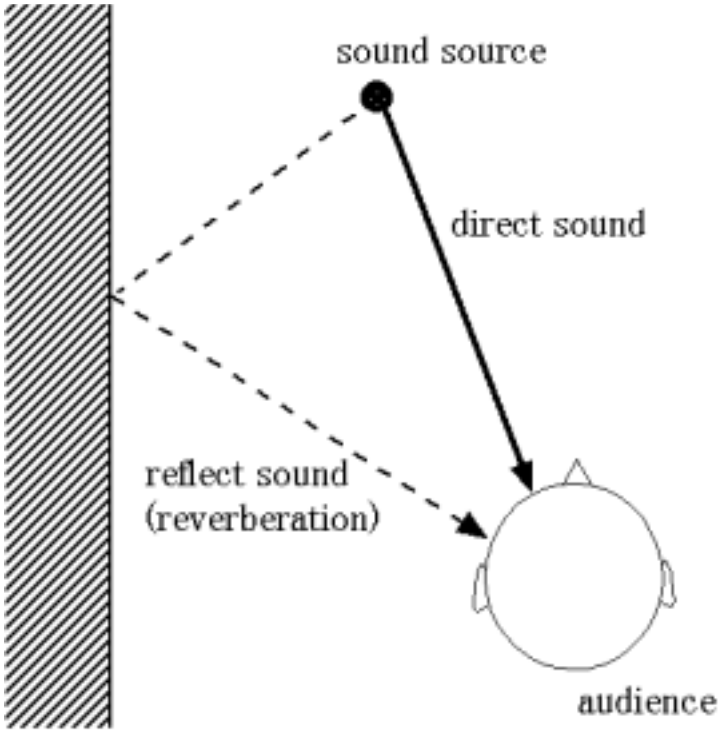

図 1.1: 残響モデル

も、その信号は直接音と同様の波形である。つまり、残響は、相関を用いて時間差を検出する手法で捜し出すべき信号と同様の特徴を有していることになる。このため、相関を用いた時間差検出手法は残響によって、その性能が悪化することになる。

しかし、我々は特殊な場合を除いては、常に残響の存在している環境で生活している。工学的に音源方向推定を実現した場合、それが用いられる環境はやはり同様に残響の存在する環境である。したがって、音源方向推定を行なう上で残響に対する対策が必要不可欠になってくる。

そこで、残響に対する何らかの対策を行なうわけであるが、相関を用いた時間差検出の際に問題となるこの特性を逆に利用することによって可能であると考えられる。事実、我々が音源の方向推定を行なう際には、残響の特徴を利用し、二章で説明する先行音効果と呼ばれるものを用いて、直接音の方向を推定していることが知られている。したがって、計算機上で音源方向の推定を行なう場合も人間の例にのっとり、残響の特徴を利用することで、先行音効果の実現ができる可能性がある。

残響に対する対策も従来の研究の中で様々な手法が提案され、用いられている。従来の、相関を用いた音源方向の推定の研究では、音圧（音の強さ）による方向推定と併用することにより精度を上げているものや、黄らによる研究では立ち上がりを補助的な手がか

りとして用い、残響成分を除去したうえで、時間差を求めるという手法を用いているものなどがある [7]。しかし、これらの手法では音源の方向推定に複数の手がかりを用いているわけであり、処理時間のことなどを含めて考えると好ましくないといえる。

1.4 目的

本研究では、残響の存在する環境で、従来一般的に用いられている相関を用いた手法では、性能の低下が生じるという現状を受けた上で、残響の存在する一般的な環境においての音源方向推定を、計算機を用いた工学的手法によって、実現することを目的とする。

その際に、聴覚で行なわれている音源方向推定の手法に示唆を得、それらの働きを簡易モデルとして工学的に応用することにより、音源方向推定を行なう。

また、三次元空間上の全方向の音源方向推定の前段階として、二次元平面全方位に対する音源方向推定に適したアレイの構築も目標とする。

第 2 章

本研究の特徴

我々人間は、非常に優秀な音源方向推定能力を持っている。本研究で、工学的手法を用いての音源方向推定を行なう上で、それら聴覚の行なっている音源方向推定法に示唆を得ることは有効なアプローチである。

したがって、この章では聴覚が行なっている音源方向推定法と比較を行なう形で、本研究の特徴について述べる。

2.1 聴覚による音源方向推定

聴覚による音源方向推定能力は、二つの耳を用いた知覚能力である。我々人間はこの二つの耳からの情報を元に空間に対する音響的な認識を行なっている。その能力は、非常に優秀であり、先行音効果と言われる反響音のある環境での音源定位や、カクテルパーティー効果と言われる音源分離能力を持つ。

ここで、聴覚において行なわれている音源方向推定について、その概要を述べる。

2.1.1 方向推定の手がかり

いま、ある方向θに音源があるとする。図 2.1 の様に音源方向が正中面以外の方向である場合、左右の耳に到達する音の音源からの距離は異なり、その経路に d という差を生じることになる。音源方向推定の基本的な手がかりとして、この経路差によって生じる物理量の変化を用いている。

まず、この距離の差によって変化する物理量として、時間差がある。例えば、頭の正中面に対して右よりの方向に音源が存在したとすると、音源から観測者までの距離は右耳の方が短く、左耳への距離はそれをりわずかに長い。このような状況では、音声が到達する

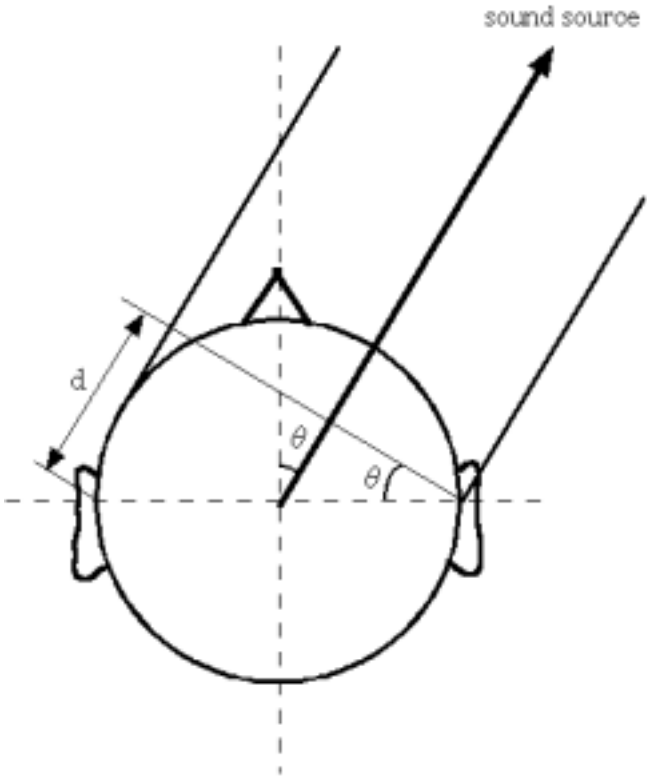

図 2.1: 経路差モデル

時間は左耳の方が遅れる。したがって、左右の耳へ到達する時間差を元にする事により、音源方向の推定ができる。実際に、聴覚における音源方向推定では、この時間差が大きな役割を持っている。しかし、このような手がかりを用いた手法による方向推定は基本的に二次元平面全方位に対してのものである。

方向推定を行なうための、もう一つの手がかりは、スペクトルの変化（音圧差）である。人間の耳は耳介を持ち、頭部の側方に位置している。ここで、正中面から右にずれた位置にある音源を考えると、左耳は頭部の影に位置することになる。しかし、音声は頭部を回り込むことで左耳に到達する。この時に、頭部の形状や耳介の効果によって、スペクトルの変化を生じることになる。聴覚ではこのようなスペクトルの変化も音源方向推定に用いている。

また、このような手がかりを元にした方向推定で低周波の音に対しては時間差が主な手がかりとして働き、高周波の音に対してはスペクトルの変化が主な手がかりとして働くというものがある。

なぜなら、低周波の音の場合を考えてみると。人間が時間差を計測する際には位相差を元にしていると考えられている。ここで、単一周波数成分の純音を考えると、位相差から

時間差を求めることができるのは、両耳間で 180 °の位相差を生じる得る限界周波数以下に限られる。それ以上の周波数では位相差から一義的に時間差を求めることはできなくなる。一方、スペクトル変化を利用した方法では低周波の音では回折が生じ易いために、頭部を回り込んでも減衰が生じ難いので、両耳間にそれほど差は生じなくなる。

次に高周波の音の場合では、時間差による方法では先ほど述べた通り、限界周波数以上の音では方向推定が不可能になるが、一方スペクトル変化を利用した手法では、高周波の音の方が回折が起こり難く、そのために頭部を回り込むと減衰を生じ易いため、両耳間の差が得られることになる。

また、このスペクトルの変化による方向推定のもう一つの大きな効果は前後の判断である。耳には耳介が存在するために、前方向から到達する音と、後ろ方向から到達する音とでは明確なスペクトル変化の違いをもたらすためである。また、これらのスペクトル変化を手がかりとしたものには、高度の知覚も含まれる。

聴覚による音源方向推定はこのように様々な、手がかりを用いることにより３次元空間上の様々な聴覚的なイベントの方向を推定している。

2.1.2 聴覚での時間差検出機構

前節で、音源方向推定の第一の手がかりが音の到達時間差であることを述べた。しかし、音が人間の左右の耳の間で生じる時間差はたかだか数 100 μsec の間しかない。このようなわずかな時間差を検出するための巧妙な神経回路網が存在する。この神経回路は、遅延線と一致検出ニューロンから構成されており、一致検出回路と呼ばれているものである。

その働きを見てみると、左右の耳へ音が到達した時、その音は聴覚末梢系で周波数毎に分解され、一致検出回路の各々のニューロンに到達する。しかし、その刺激は、通過する神経繊維の長さの違いによって、各ニューロンへの到達時間が異なることになる。つまり、右耳からの刺激は右側のニューロンには早く到達し、左側のニューロンへは神経繊維の長さ分遅れ時間を持って到達するわけである。そして、各ニューロンは左右からの刺激が同時に到達した時に強く発火する。したがって、一致検出回路のどのニューロンが発火したかを知ることによって左右の耳に到達した音の時間差を検出しているわけである。そのモデルを図 2.2 に示す。

たとえば、左右の耳に同時に音が到来した場合（正中面に音源のある場合）中央のニューロンが発火し、左耳へ到達する音が遅れを持っていた場合（正中面から右にずれた位置に音源のある場合）では中央から左にずれた位置のニューロンが発火することになる。

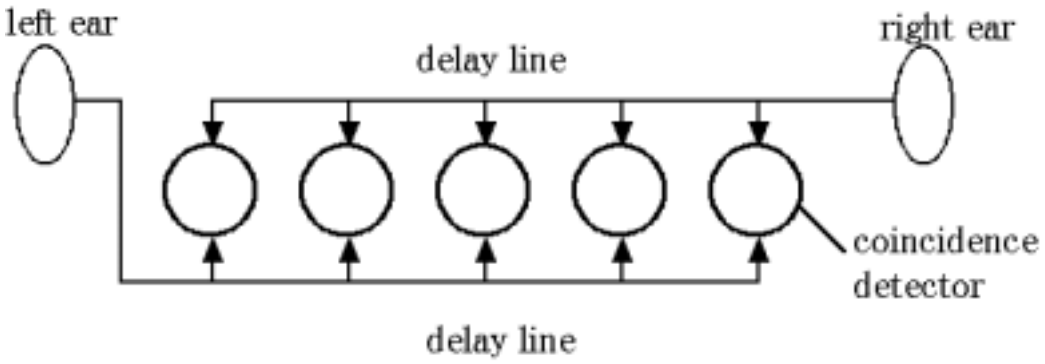

図 2.2: 一致検出回路

この生体での時間差検出機構は両耳間の相互相関に基づくモデルである。また、このモデルは Jeffress モデルとして知られ、簡潔な構造でありながらも、解剖学的研究においても、それと類似する構造が見られるため、時間差検出モデルとして広く支持されている [8]。

2.1.3 先行音効果

人間は普段反響音が多く存在する環境で生活している。しかし、それにも関わらず優れた音源方向推定能力を有している。これは、心理学的な研究によると先行音効果（Hass effect) が働いているためでいわれる。先行音効果とは、人間が音源方向推定を行なう上で、最初に到達した音を用いる働きである。直接音は音源から観測者まで最短距離を通って到達するため、最初に到達する。したがって、先行音効果はこの直接音による情報を積極的に利用することにより、反射音などが存在する環境でも正確に、直接音の音源方向推定を行なうことができる。

2.2 マクロホンアレイを用いた音源方向推定

我々人間は様々な手がかりを元に、音源方向の推定を行なっている。その中で主だったものは、時間差、スペクトルの変化である。マイクロホンアレイを用いて音源方向推定を行なう場合これら全ての手がかりを用いることは困難であるため、それらの中から計算機を用いた処理に適した手がかりを選択する。

ここで、スペクトル変化を利用した手法について考えてみる。人間がスペクトルの変化を用いた方向推定を行なう上で、耳介の存在と、頭部の側方に耳が位置していることを利

用している。そのために、スペクトルの変化を用いた方法ではこれらのものを再現しなければならない。

そこで、本研究では従来の方法と同様に、両耳間の距離差に対しての変化量に線形性が保持されており、最も扱い易い時間差を手がかりとすることにする。

また、人間は二つの耳で、方向推定を行なっている。これは、いわば二つのマイクロホンで構成された直線配置のアレイと見ることができる。しかし、我々はこの二つの耳で３次元上の全方向に対して方向推定を行なうことができる。これは、前節で述べた通り頭部や耳介の存在を利用した上で様々な手がかりを元に音源の位置情報を得ているためである。しかし、マイクロホンアレイを用いた音源方向推定では、ダミーヘッドを用いた手法以外に、受音に影響を与えるような耳介などは存在せず、２つのマイクロホンでは人間と同様な３次元空間上の音源定位は不可能である。

そこで、本研究では３次元空間上の全方位の方向推定の前段階として２次元平面全方位の方向推定が可能なアレイを考える。方向推定の手がかりとして、音声の到達時間差を用いるという前提の上で、最少のマイクロホン数で２次元平面全方位の方向推定が可能なアレイとして、図 2.3 に示すような正三角形の頂点にマイクロホンを配置したアレイを用いることにする。

2.2.1 マイクロホンアレイ形状

一辺の長さが d である正三角形の各頂点に無指向性マイクロホンが配置されているとする。図に示すようにマイクロホン１、２を結ぶ直線の垂直方向で、マイクロホン３と反対方向を基準の方向（０度）とした時、これから右回りにθ 変移した方向に音源があるとする。

この時各マイクロホンに到達する時間差は次式で与えられる。ここで、c は音速とする。

1-2:

$$(t_1 - t_2)\frac{c}{d} = sin(\theta) \tag{2.1}$$

2-3:

$$(t_2 - t_3)\frac{c}{d} = sin(-\frac{\pi}{3} - \theta) \tag{2.2}$$

3-1:

$$(t_3 - t_1)\frac{c}{d} = sin(\frac{\pi}{3} - \theta) \tag{2.3}$$

ここで、各マイクロホンに到達する音の時間差が計測できたとすると、上の３の式より次のようにθが決定できる。

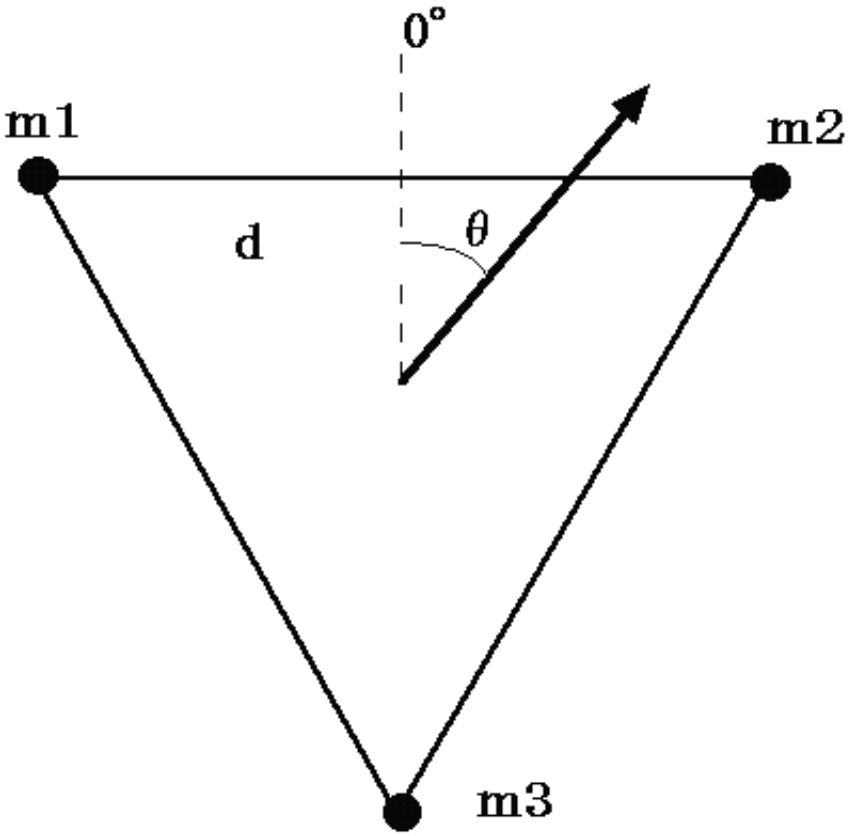

図 2.3: マイクロホンアレイ形状

1-2:

$$\theta = arcsin[k(t_1 - t_2)] \tag{2.4}$$

2-3:

$$\theta = -\frac{\pi}{3} - arcsin[k(t_1 - t_2)] \tag{2.5}$$

3-1:

$$\theta = \frac{\pi}{3} - arcsin[k(t_1 - t_2)] \tag{2.6}$$

ここで、正しく時間差が得られたとして、音源方向を求めてみる。例として次のような結果を示す。

$$(t_1 - t_2)\frac{c}{d} = 0 \Rightarrow \theta = 0, \frac{3\pi}{4}$$

$$(t_2 - t_3)\frac{c}{d} = \frac{\sqrt{3}}{2} \Rightarrow \theta = 0, \frac{\pi}{3}$$

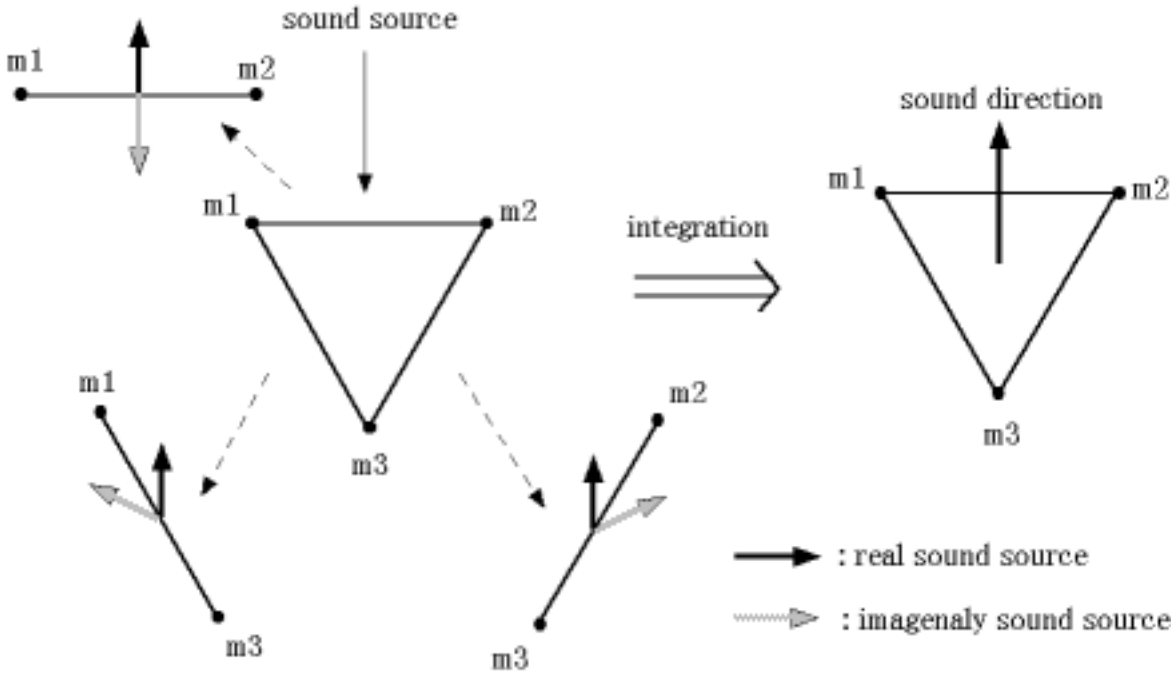

図 2.4: 各々の角度の足し合わせ

$$(t_3 - t_1)\frac{c}{d} = -\frac{\sqrt{3}}{2} \Rightarrow \theta = 0, \frac{-3\pi}{4}$$

つまり、この式によって求められた音源方向は各マイクロホンに対して２組ずつ得られることになる。しかし、各々のマイクロホン対で得られた結果のうち一方は正しい音源方向を示していることは確かである。そこで、各マイクロホン対で得られる方向（３組×２方向＝６方向）の中から正しい方向を選択する必要がある。

各マイクロホン対で正しく方向が得られた場合を考えてみる（図 2.4）。この時、各々のマイクロホン対で実際の音源を示す real sound source と、そのミラーイメージである imagenaly sound source とが生じる。正しい方向は、real sound image であり、実際に各々のマイクロホン対で同じ方向を示している。一方、imagenaly sound source を見てみると、各マイクロホン対で示す方向が異なっているのが分かる。したがって、これらの得られた方向を足し合わせることで、３組のマイクロホン対が同時に示す一方向を正しい音源方向として、得ることができ、従来の直線配置のマイクロホンアレイで生じていた、前後の判断の誤りの問題を解決することができる。

2.2.2 マイクロホンアレイの解像度

一つのマイクロホンアレイを見てみると、音源方向とその解像度にある関係が存在する。

ここで、時間差測定の解像度を考える。いま、一つのマイクロホンアレイを考えると、最大の時間差を生じる方向というのは各々のマイクロホンを結んだ線の垂直二等分線を 0 度して、90 度及び-90 度方向となる。そしてこの時の時間差はマイクロホンの間隔 d によって決定される。

計算機において時間差を扱う場合、信号をある周波数でサンプリングして取り扱うことになる。いま、ある周波数で入力信号をサンプリングしたとすると、各々のマイクロホン対で生じる時間差が大きいほど、多くのサンプリング点がとれることになる。このことを考えると、一つのマイクロホン対で分解できる方向の数を求めることができ、その数はマクロホン対で生じる最大の時間差にサンプリング周波数を掛けた数の 2 倍の数になる（正の時間差と、負の時間差の両方が生じるため）。そして、この数は-90 度から+90 度までの 180 度を分割する数になる。

また、マイクロホンアレイでは得られた時間差によって解像度が異なる。この時間差は、音が各々のマイクロホンに到達する時の経路差によって生じるが、この経路差は音源方向を θとして、$sin\theta$で計算される。そこで、音源からの経路差と音源方向との関係を見てみると、音源方向が 0 度付近では角度差に対する経路差の変化量が大きく、また音源方向が± 90 付近では同じ角度差に対する経路差の変化が小さくなっている。この経路差は直接、時間差に影響を与えるものであり、先ほどのサンプリング周波数の問題を考えてみるとある角度差に対する経路差が大きい（0 度付近）ということは、つまりその角度差の間を細かく分割できるということになる。一方、経路差が小さいところではその角度差の間を荒く分割することになる。

このように、マイクロホンアレイを用いた音源方向推定の解像度は、マイクロホン対で生じる最大の時間差と、得られた時間差に影響される。

このことから、解像度は次のような式で計算される。

$$\eta = \frac{\partial\theta}{\partial(\Delta t)} = \frac{1}{\sqrt{\Delta T^2 - \Delta t^2}} \tag{2.7}$$

ここで、ΔTはそのマイクロホンアレイで起こり得る最大の時間差。Δt は測定した時間差を示す。この式を見てみると、ΔTが大きくなるにつれて（マイクロホンアレイの間隔が大きくなるにつれて）解像度は良くなる。また、Δt が小さくなるにつれて解像度が良くなる。

ここで、一つのマイクロホンアレイの解像度を示すと、図 2.5 のようになる。図に示すように、180 度毎に一回、解像度の最も良い場所が現れるが、逆に解像度の非常に悪い場所も広い範囲に渡って存在している。ここで、本研究で用いる正三角形のマイクロホンアレイを考えてみると、このようなマイクロホンアレイでは各マイクロホンの正中面が 60

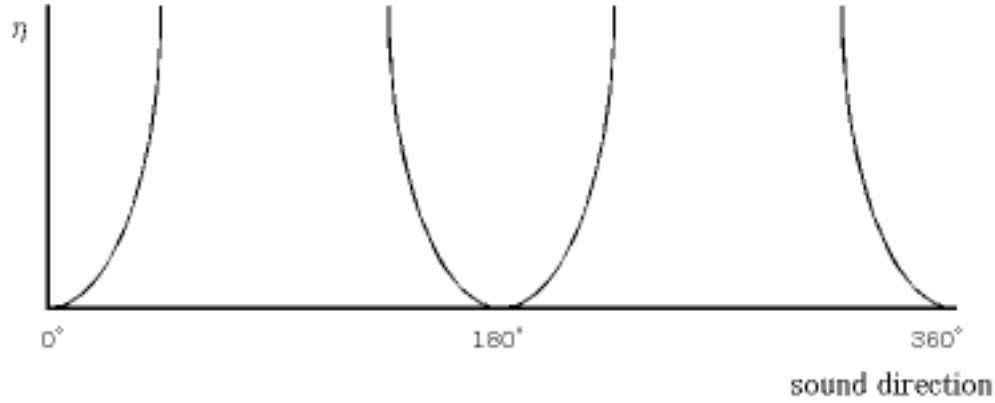

図 2.5: 単独マイクロホンの解像度

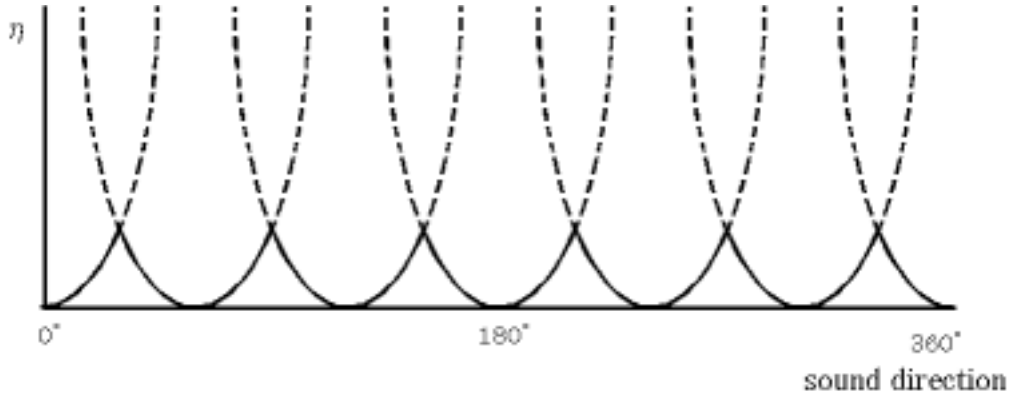

図 2.6: 正三角形配置のマイクロホンの解像度

度毎に存在することが分かる。

このマイクロホンアレイの角度と解像度の関係を示したグラフは図 2.6 のようになる。図中の点線が各々のマイクロホン対での解像度を、また実線が３組のマイクロホンを合わせた解像度を示している。

図のように３つの直線配置のマイクロホンアレイが重なることにより、30 度毎に解像度の良い部分と悪い部分とが繰り返す形となる。つまり、このような形のアレイでは各々のマイクロホン対がお互いの解像度の悪い部分を、各々の解像度の良い部分で補う形となる。もちろんこのような解像度の補完現象は、マイクロホンの数を増やせば増やすほど（マイクロホン対の正中面が増えるにつれ）多く現れ、全体的な解像度は良くなる。しかし、マイクロホンの数を増やすことは、処理時間にそのまま影響を与えることにもなるので、注意が必要である。

2.2.3 時間差検出回路

時間差を検出するにあたり、一致検出回路のモデルを利用することにする。先に紹介した、一致検出回路（Jeffress モデル）は聴覚に類似した構造が見られる上に、非常に簡単な構造で、扱い易いモデルである。そのため、計算機を用いての時間差検出を行なうためのモデルとしても適している。

本研究ではこの一致検出回路の簡易モデルを用いる。本来一致検出回路へは、聴覚末梢系で周波数分割され、位相によって発火された神経パルスが入力されることになる。しかし、本手法においては簡単化のために、時間軸上で存在するイベントを入力に用いる。

このような、モデルを用いることで、容易にそして素早く各マイクロホン間での時間差の検出を行なう。

2.2.4 残響に対する処理

実環境での音源方向推定を行なうには残響に対する対策が必要になり、人間が方向推定をする上で行なっている先行音効果を何らかの方法で実現する必要がある。

先行音効果を実現するためには、いくつかの方法が考えられる。まず、その一つとして、立ち上がりの強調を行なう手法である。これは、その名の通り、音の立ち上がりを強調する処理方法である。立ち上がりを強調することで、最初に到達する情報、つまり直接音を検出しようとするものである。

次に、不応期を用いた手法が考えられる。つまり、閾値処理などを行ない直接音から生じるイベントを検出した後、残響が入力されると予測される期間検出を行なわなくする処理である。

このような手法を用いることで、確かに先行音効果を実現できる。しかし、基本的に入力信号は未知のものであるのが前提である。したがって、これらの手法では何を立ち上がりとしてとらえ強調するのか、そしてどのようなイベントを直接音としてとらえ不応期を設定するのか、という条件設定が困難となる。

そこで、少し別の考え方をしてみる。残響の特性を考慮に入れ、直接音によるイベントを検出した後、一時的に閾値レベルを上昇させる。これにより、直接音よりも遅れて到達し、パワーの小さくなっている残響に対する検出をおさえることができると考える。

この考えを実現するために、本研究では、入力された信号に対して動的に閾値レベルを設定するために、変動閾値という手法を提案する。この手法については三章でアルゴリズムと共に詳しく説明することにする。

第 3 章

音源方向推定法

本研究で構築した音源方向推定法は、図 3.1 に示す様なフローチャートに沿って行なわれる。以下に個々の処理について詳しく述べる。

3.1 ピーク抽出処理

まず、立ち上がり検出処理の前処理として、３つのマイクロホンによって得られた信号に対してピークポイントの抽出を行なう。ここで、ピークポイントは入力信号中の極大、極小点とする。これは、信号のピークポイントを抽出することにより、信号の立ち上がり箇所をより鮮明にするために行なわれる。具体的な手順としては以下の通りである。

- 入力された信号と、その１点前の信号との差分信号を作る
- 差分信号のゼロクロスポイントを検索する
- 入力信号より、差分信号にゼロクロスが生じたポイントの振幅情報を得、ピーク信号を作る。また、ゼロクロスが生じなかった時点のピーク信号の振幅は０とおく

つまり、信号のピークポイントを検出し、その時点の振幅のみを持った信号（ピーク信号）を入力信号より生成する。

例として、入力信号とそれに対するピーク信号の例を図 3.2 に示す。図 3.2(A) が入力された信号、図 3.2(B) が入力信号に対してピーク抽出処理を行なった後のピーク信号である。この、図 3.2(B) を見てみると、入力信号の極大、極小点の振幅のみが残った信号が作られているのが分かる。

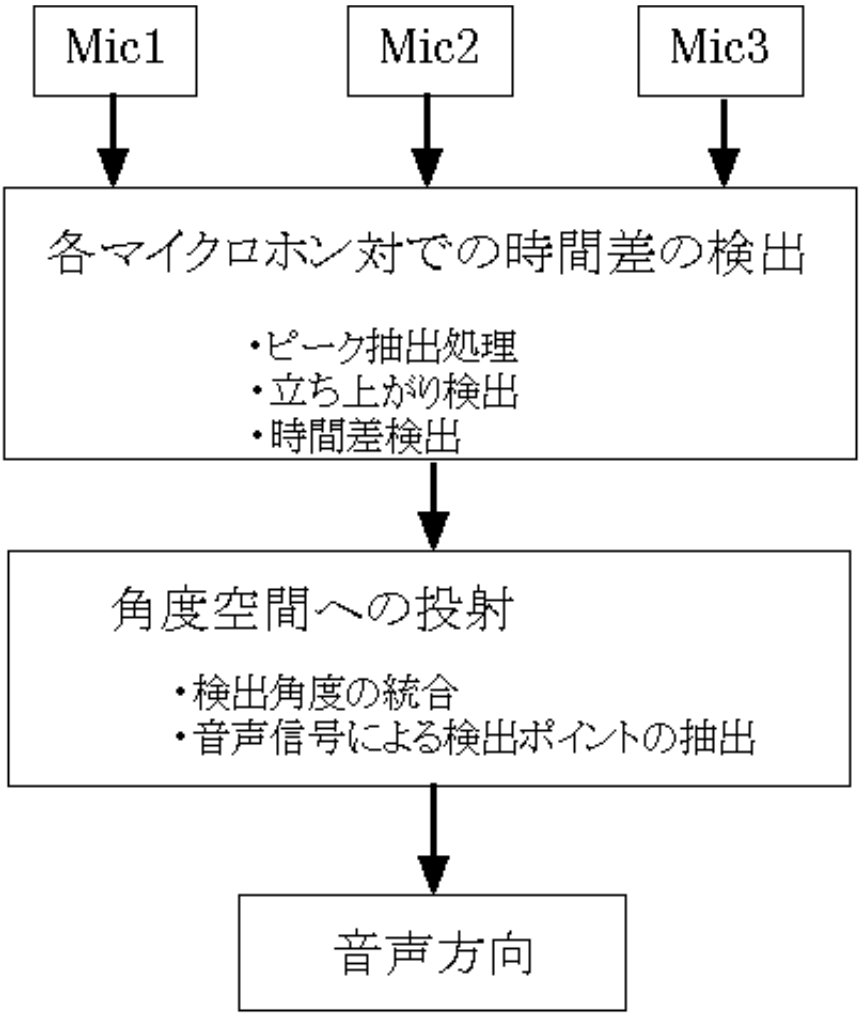

図 3.1: 音源方向推定流れ図

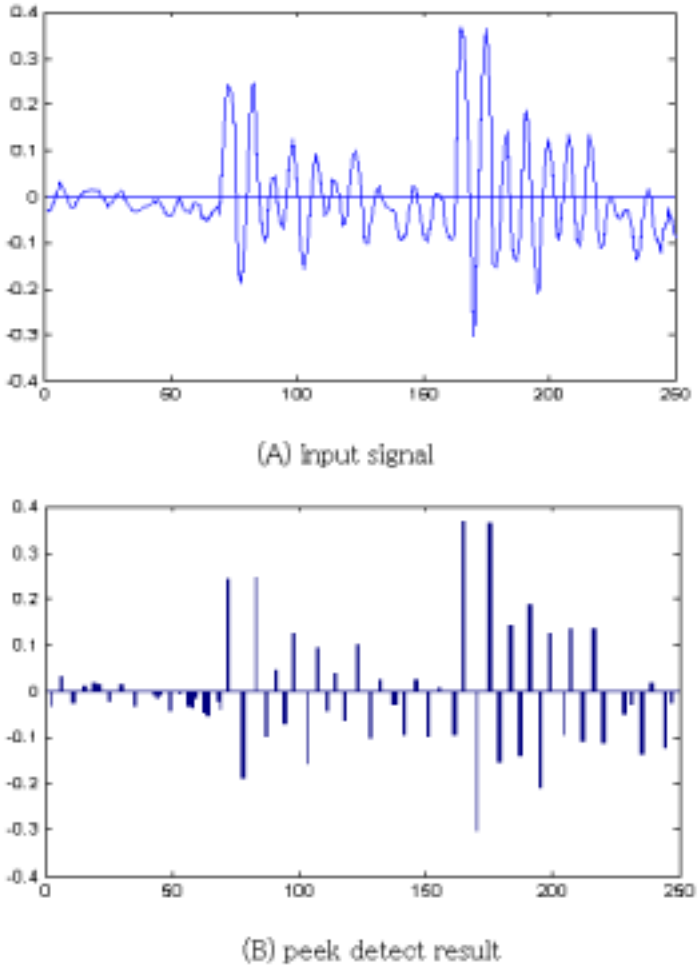

図 3.2: ピーク抽出処理

このようにして生成したピーク信号を元にして、立ち上がりの検出処理を行なうことにより、立ち上がり検出ポイントの絞り込みを行なうことができる。また、正負両方の振幅の情報を利用するために、図 3.2 の信号に全波整流を行なった信号を、この後の処理に用いていく。

3.2 立ち上がり検出処理

時間差を得るために、まず各々のマイクロホンに到達する信号より立ち上がりポイント（振幅が急激に変動するポイント）をとる。ここで、信号に対応して変動する変動閾値を用いることで反射音に対する不応期を実現する。

3.2.1 変動閾値

この変動閾値は、信号の振幅が閾値を越えた時点を急激な振幅の変動が生じた点として捉え、その時点をスタートとして、以下のような振舞いをする。

- 振幅が閾値を越えた時を立ち上がりとして検出する
- また、その時の信号の振幅の値を初期値として保持する
- 閾値は時間と共に指数減少する
- 振幅が閾値を越えるたびに、検出ポイントを得、この動作を繰り返す

このような変動閾値を用いることにより、反射音に対する不応期を設けることができる。つまり、動作を順を追って見ていくと、閾値は一度大きな振幅に反応した直後に最大の値を持つ。そして、そこから徐々に指数減少により閾値のレベルが下がっていくわけである。先に説明した通り、直接音は最も早く到達し、最もそのパワーが大きく、反射音はそれよりも遅れて到達し、そのパワーは小さくなる。したがって、最も早く到達し最もパワーの大きい信号に反応した直後に大きな閾値レベルを設けることにより、その後の遅れて入るパワーの小さな信号に対しての検出が生じなくなる。

また、この変動閾値の手法は正確にいうと音の立ち上がりを検出しているのではなく、急激に大きな振幅が入力されるポイントを検出していることになる。ここで、音の立ち上がりというものについて考えてみる。衝撃音などでは異なるが、音は通常発生時点から緩やかな振幅包絡に沿って徐々にパワーが増大し、定常状態に落ち着くことになる。ここで、この変動閾値の特性と合わせて考えてみると、変動閾値のレベルは入力された信号に対して上下することになる。つまり、振幅の小さな信号が入力された際には低いレベルが設定され、振幅の大きな信号の場合は高いレベルが設定されることになる。

このことをふまえて、徐々にその振幅が増大するような信号を考えると、最初振幅は小さいために閾値のレベルも低く設定される。そして、次にそれよりも大きな振幅を持った信号が入力されると、閾値のレベルは低く設定されているために、反応し易くなる。そして、再び、信号の振幅によって閾値のレベルが再設定され、次の更に振幅の大きくなった信号に対して反応する、といった動作を繰り返すことになる。つまり、変動閾値それ自体は実は大きな振幅を持つ場所を捜し出すための手法ではあるものの、音の立ち上がりの特性をふまえて考えると、その動作はあたかも音の立ち上がり箇所を検出しているように見えるわけである。

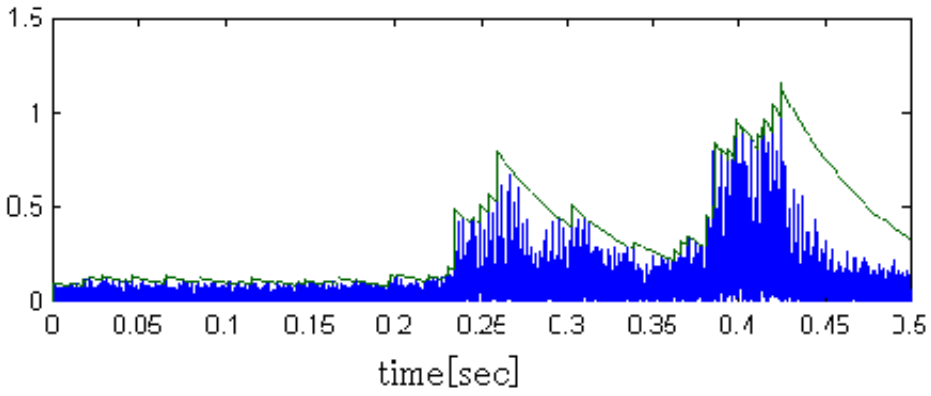

図 3.3: 変動閾値実例

実際の入力信号に対する変動閾値の動きの一例を図 3.3 に示す。これは、入力信号に対しピーク抽出処理を行なった後全波整流を行なった信号と、それに対する変動閾値の動きである。これを見ると分かるように確かに、音の立ち上がり（振幅が徐々に増大する箇所）で、頻繁に反応しそのレベルが再設定されているのが分かる。

3.2.2 入力信号に対するロバスト性

もう一つの変動閾値の効果は信号に対するロバスト性を持たせることができるというものである。通常、マイクロホンによって得られる信号は未知のものであり、しかもマイクロホンの個々の特性などにより、同じ音を受けた場合であっても、異なる振幅の信号を出力するおそれもある。このような信号に対して、ある決められたパラメーターによって固定された閾値を用いた場合では都合が悪い。例えば、入力の最大値が閾値を越えなかった場合、その信号に有用な情報があったとしても検出できなくなってしまう。逆に非常に大きな入力が与えられてしまった場合、ノイズなど不要な情報を検出し続けるかもしれない。

また、変動閾値では閾値レベルの変動という形で不応期を与えていることになるが、この不応期を何らかのパラメータで設定して決定し、不応期の間全く反応検出を行なわないような手法を用いた時を考えてみる。このような手法で、何らかの要因で誤った検出がなされた場合、そこから継続して設定した期間不応期が生じる。ここで、この不応期の期間中に有用な情報が存在しても、その情報を得ることができなくなってしまう。

しかし、この変動閾値はそのレベルを決定するにあたり決まったパラメーターを持たずに、常に信号の入力レベルによって設定されることになり、また閾値レベルの上昇という形で不応期を実現しているために、このような状況を回避することができる。

したがって、変動閾値を用いた手法では、その閾値レベルの決定には人為的な操作は必要とせずに、入力された信号を監視し、それに適した閾値レベルを設定しているため、信号に対するロバスト性が確保できることとなる。

3.3 時間差検出

各マイクロホン対での時間差を測定するために、二章で述べた通り生体での時間差検出機構である一致検出回路のモデルを用いる。

ここで、注目すべきところはこの一致検出回路への入力として、変動閾値の検出結果を用いると言うことである。ここで、先に述べたが一致検出回路は両耳間相互相関に基づいたモデルであることを考えなければならない。したがって、そのまま用いれば、一章で説明した様な残響による性能の低下を生じることになる。しかし、本研究ではこの入力に変動閾値の検出結果、つまり残響の成分を除外し直接音の情報のみを持った信号を入力することにより、これまでの相関を用いた手法とはことなり、直接音によって与えられる時間差のみを検出することが可能となっている。

ここで、注意が必要なことは時間差が検出されるタイミングである。時間差の検出結果はある同一の音に起因するものであっても、必ずしも同時に発生しない。また、１組のマイクロホン対を考えると、検出された時間差ははある時点で出現し、その後すぐに消失する。後の処理で、３組のマイクロホン対から得られる３つの時間差を元に音声方向を推定することになるが、その際この３つの結果を各時点で統合を行なっている。そのために、お互いに対になってるはずの検出結果が同じ時点で発生していないことは都合の悪いことになる。

そこで、あるマイクロホン対で得られた時間差と、対になる時間差が他のマイクロホン対で生じるまでの待ち時間を用意しておく必要がある。この待ち時間は、マイクロホン対で生じる最大の時間差の分だけあれば良く、その間検出値を保持する処理を行なっている。

3.4 音源方向決定手法

一致検出回路を用いた時間差検出により各々のマイクロホン対での時間差を得ることができる。次のステップではこの時間差を角度空間に投射し、その後検出された時間差が音の信号によるものなのか、ノイズなどによるものかの判別を行なう。

3.4.1　時間差の角度空間への投射、統合

各々のマイクロホン対での時間差を元に角度空間への投射を行なう。本手法では、あらかじめ時間差と角度情報を対応付けたテーブルを作成しておき、時間差を元にテーブルを参照するという形で角度空間への投射を行なう。

テーブルの作成については前述の式 (2.4)〜(2.6) によって、時間差と角度の対応づけが得られ、ある時間差が示す角度が分かる。ここで、時間差の検出にはある程度の誤差が含まれると仮定し、角度軸上にある程度の幅を持たせておく。

このようなテーブルを用いて角度空間に投射した結果は、二章で説明したように、一つの時間差につき二つの方向が示されることになる。いわゆる、real sound source と imagenaly sound source が存在しているわけである。しかし、この時点ではどちらの角度が正しいかの判別はできない。そこで、３つの時間差によって得られる角度を統合する。これにより、ある時点で３つの時間差が同時に示す角度が正しい情報として選ばれ、もう一方の imagenaly sound source は意味のない情報として切り捨てられる。

また、このような角度情報の他に、誤った時間差検出結果による検出角度というものが存在する可能性があるが、このような情報も省かれることになる。なぜなら、一組のマイクロホン対で生じた誤った時間差によって示された方向があったとしても、その方向は他の２対のマイクロホン対では必ずしも同時には生じないことになる。したがって、３つの角度情報を統合することにより、それら誤った情報は取り除かれ、正しい音源方向のみが取り出される。

3.4.2　音声信号による検出ポイントの抽出

角度情報を統合することによって極力誤った情報は省かれることになるが、それでもノイズなどによる検出誤りの部分が現れる可能性がある。そこで、音声の信号の特性を考慮した上で、検出された角度が音によるものであるか、そうでないかを判別する。

ここで、入力信号が人間の音声信号である場合を考えてみると、人間の音声は基本周期で規則的なサイクルをもっている。この基本周期はいわば声帯の振動周期によるものである。音声信号の特徴をとらえてみると、声帯が開放される時に瞬間的に大きな振幅が生じている（図 3.4）。言い替えてみると、基本周期毎に一度大きな振幅が現れるということになる。更に、変動閾値の特性を考えると、変動閾値を用いた手法による検出ポイントは前に述べた通り急激な振幅の変動ポイントをとっているわけである。この事を合わせて考えてみると、変動閾値は音声信号の周期の中の大きな振幅、つまりは声帯の開放ポイントをとらえていることになる。

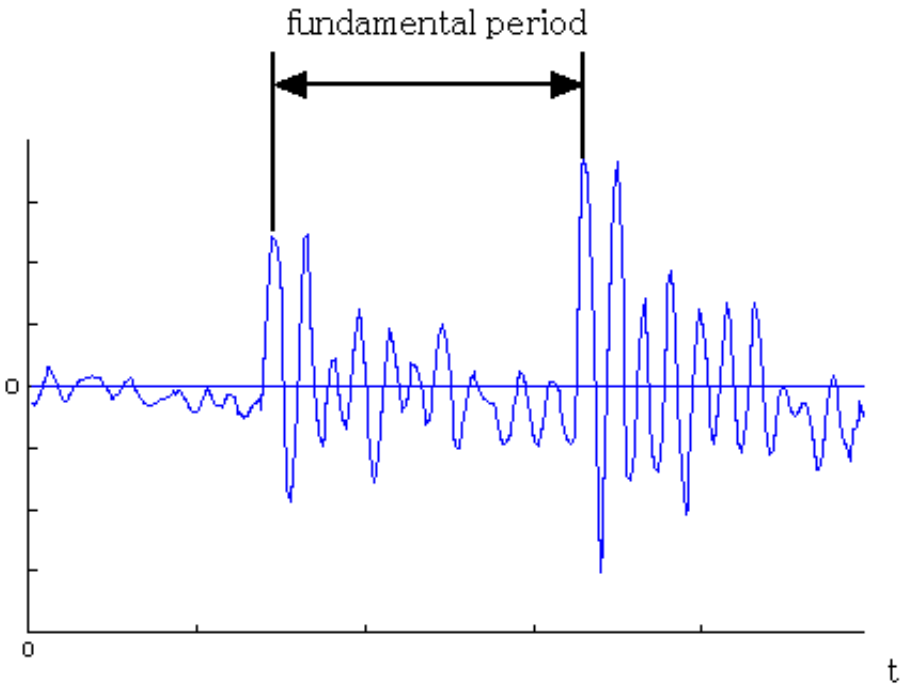

図 3.4: 音声信号例

変動閾値による結果、入力信号に対して最も密に検出ポイントが得られた場合、基本周期毎に得られるわけである。そして、この周期はその後の処理にも影響を与えることになり、時間差の検出及び、角度の統合結果もまた基本的に音声の基本周期毎に検出ポイントが得られることになる。

したがって、検出された角度が、時間軸上で連続的に同じ方向を示している箇所が、音声によって得られた角度であると考えることができる。そのため、音声信号によって得られた角度か、それ以外の誤って検出された角度かを区別するためには、連続的に同じ角度が検出されている場所を探せば良いことになる。

本アルゴリズムでは、時間継続と時間平均操作によって連続的に同じ角度が検出されている場所を捜し出している。その手法は以下の通りである。

まず、これまでの手法によって検出された角度に対して、その値をある程度の時間保持するといった時間継続を行なう。この時間継続幅は、基本周波数を考慮に入れ、一つの基本周期よりも長い期間とする。実際には、必ずしも基本周期毎に検出が得られるという保証はないので、基本周期の数倍の時間継続幅を持たせることが望ましい。

次に、このように時間継続を与えた信号に対して時間平均を行なう。この時、時間平均の窓幅は時間継続幅よりも長く設定しておく。

このように時間平均を行なった結果がどのようになるかを、単独の検出ポイントと、基本周期毎に連続して存在する検出ポイントを比較して考えてみる（図 3.5）。

図 3.5（A）が単独の検出ポイントによる時間平均のモデル図である。図のように、単独の検出ポイントに対して時間継続が行なわれ、それに対して時間平均を行なっている。時間継続を行なった後の信号は図中の点線で示されるようなものになる。このような信号に対して平均処理を行なうわけであるが、平均窓幅は一つの時間継続幅よりも長く設定してある。本手法では検出値は０と１の値で保持するようになっている。そのために、時間継続幅以上の窓幅を持って平均をとった場合、その結果の値は１に満たない値となる。

一方、図 3.5（B）図は連続して検出ポイントが与えられた例である。この例で、まず最初の検出ポイントに対して、時間継続が行なわれる。上の例と異なるところは、その時間継続期間内に他の検出ポイントが存在することである。更に、二つ目の検出ポイントに対しても継続時間が与えられる。そして、二つの継続時間が重なった結果、長時間連続した検出ポイントが得られることになる。その結果として時間平均の窓幅より長い間値を持ち続ける信号が出現する。この信号に対して時間平均を行なった結果の値は１となる。

従って、このように、時間継続と時間平均の操作を行なうことにより、その結果が１であるか、１に満たないかで、連続した検出ポイントが存在する場所、つまり音声信号による検出ポイントとそれ以外の検出ポイントを区別することが可能になる。

しかし、ここで注意しなければならないのは、個人による基本周波数の差である。人間の音声の基本周波数は男性でおよそ 90～130Hz。女性では 250～330Hz と言われている。そして、この周波数の差は時間継続と、平均窓長を決定する上で重要な条件となる。基本的には継続時間を決定する際には、低い基本周波数に合わせる必要がある。またそれに関係して時間平均の窓幅も影響される。なぜなら、高い基本周波数に合わせて継続時間を設定した場合、低い基本周波数では次のサイクルが来る以前に継続時間が終るおそれがあり、結果として基本周波数の低い音声による検出ポイントが得られないことになる。そして、時間平均のステップ幅は高い周波数に合わせて決定される。

しかし、男性、女性を同時に考えると基本周波数は広く分布しているためその大部分の音声に対して適したパラメータを設定する必要がある。

また、あらかじめ何らかの手法により基本周波数を推定した後、適応的に時間継続幅等を決定するといったことも考えられるが、本研究においては処理時間等も考慮に入れた上で、このような手法は用いない。

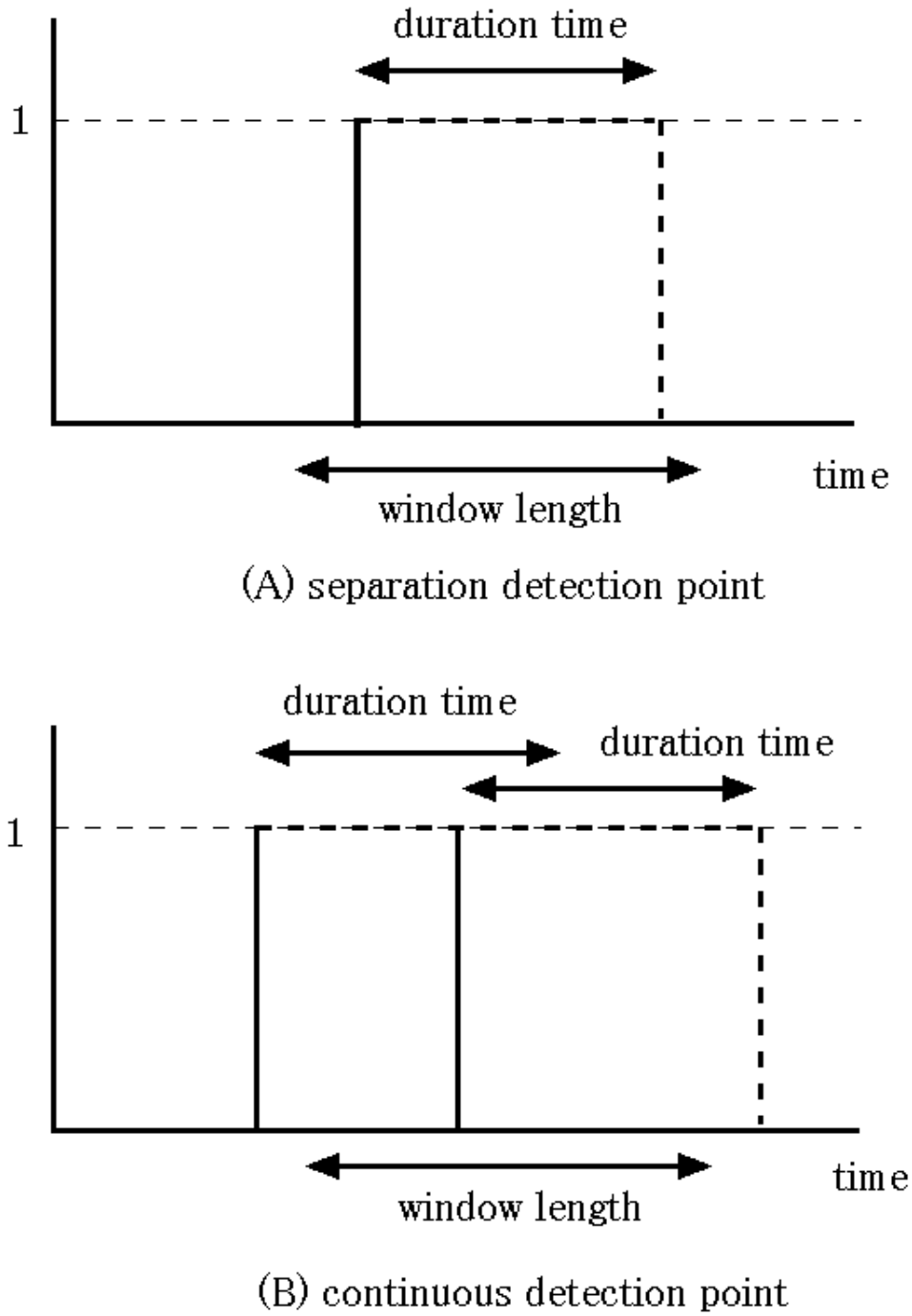
duration time
1
time
window length
(A) separation detection point
duration time
duration time
1
time
window length
(B) continuous detection point

図 3.5: 時間平均モデル

第 4 章

計算機上で作成した信号を用いたシミュレーション

4.1 シミュレーション結果の一例

まず、シミュレーションの結果として個々の処理の結果を示す。ここで、三章で示したフローチャートに沿った形で、その処理毎に結果を示し、各処理で得られる結果の推移及び妥当性を見ていく。

4.1.1 実験条件

例として用いる入力信号の各パラメーターは以下の通りである。

マイクロホン間隔	0.3m
サンプリング周波数	20kHz
使用音声	ATR データベース男性話者単語 /a i ma i/
雑音	白色雑音（SNR 15dB）
音声方向	180 °
第一反射音方向	25 °
直接音対第一反射音パワー比	-4dB
第二反射音方向	120 °
直接音対第二反射音パワー比	-8dB

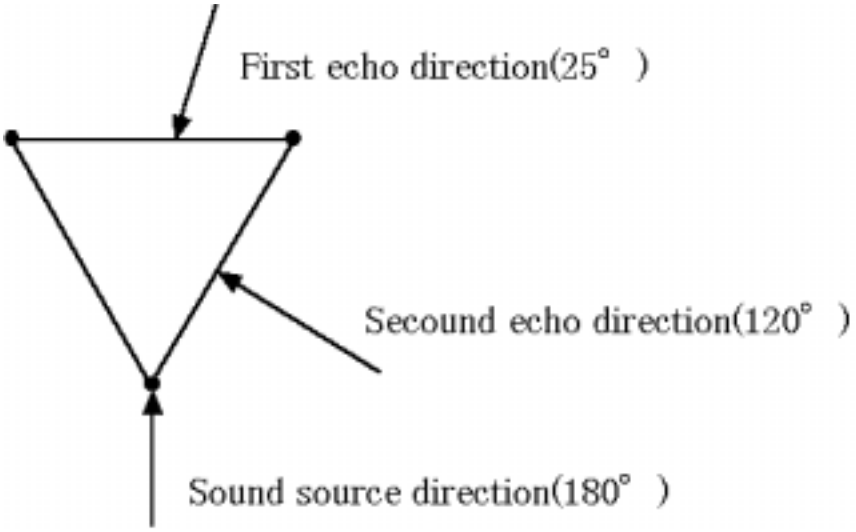

図 4.1: 音声、反射音方向

以上のような条件で、計算機上で入力方向より時間差を求め、各々に時間差を加えた信号を作成し、３つのマイクロホンの入力とする

また、ここでの SNR は雑音と元の信号との比であり、反射音の成分を含まない信号と雑音との比になっている。

入力信号を作成する際には、一度 ATR データベースの信号を 40kHz にアップサンプリングし、そのサンプリングレートで時間差を加えた後、20kHz へダウンサンプリングしている。こうすることで、より厳密に時間差を与えることができる。

4.1.2　立ち上がり検出結果

図 4.2 に入力信号の一例、図 4.3 に立ち上がり検出結果を示す。

ここで、図 4.3(A) はピーク抽出処理を行い、全波整流をかけた波形と、それに対する変動閾値の動き。図 4.3(B) で、パルスの立っている場所が検出点である。これを見てみると、音声の立ち上がり箇所で検出点が得られ、音声の立ち下がり箇所では検出点が得られていない、つまり閾値が反応していないことが分かる。したがって、変動閾値が音声の立ち上がり箇所を見つけている。

更に、この後の処理では図 4.3(B) の信号を用いることになり、入力信号をそのまま用いた場合に比べ、取り扱う情報量が少なくなっているのも注目すべき点である。

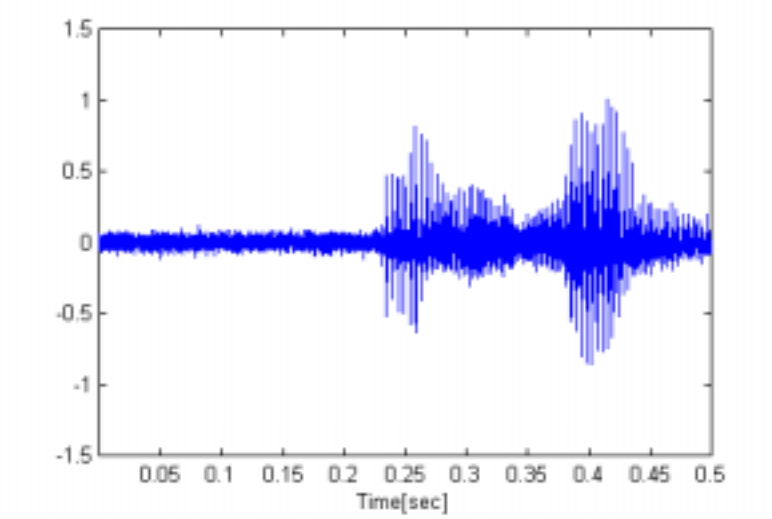

図 4.2: 入力信号例

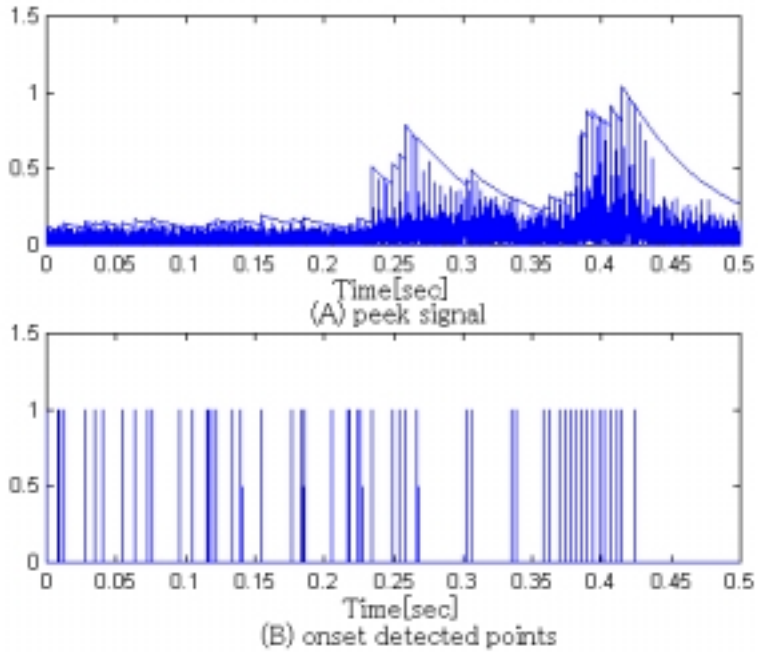

図 4.3: 立ち上がり検出結果

4.1.3 時間差検出

変動閾値を用いた処理により得られた検出点を用い、三章で述べた手法により各マイクロホン対の間での時間差を検出する。その結果を図 4.4 に示す。

上から、各々マイクロホン対 1-2、マイクロホン対 2-3、マイクロホン対 3-1 で得られた時間差である。図のように各々のマイクロホン対で多くの時間差が検出されている。図を注意深く見てみると、同じ検出結果が連続して現れている箇所があるのが分かる（○で囲った部分）。

音声信号に対する処理を行なった場合、変動閾値の特性により基本周期毎に立ち上がりの検出が行なわれ、時間差検出回路への入力もまた基本周期毎に行なわれることになる。したがって、時間差の検出も基本周期毎に得られることになり、音声信号によって得られた時間差はこのように連続して同じ方向を示すような検出点になる。

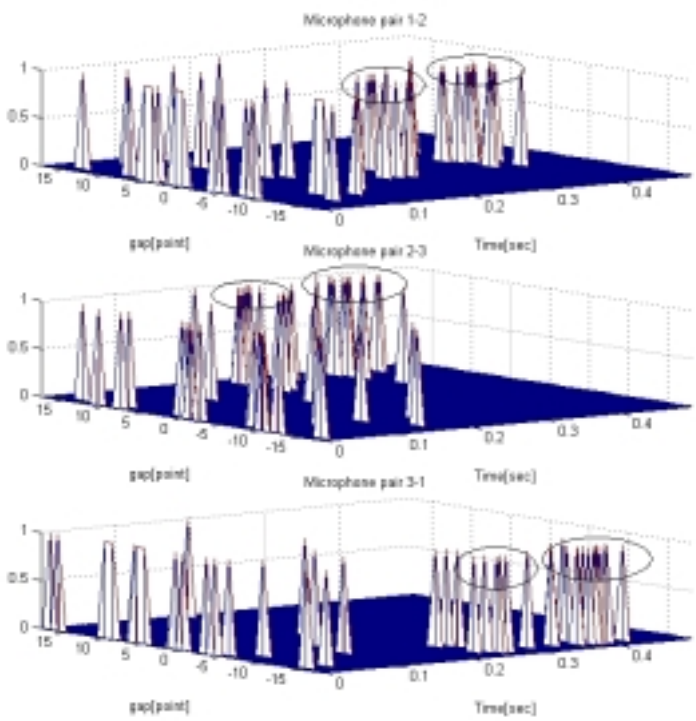

図 4.4: 時間差検出結果

4.1.4 角度空間への投射

時間差検出結果を元にし、式 (2.4)-(2.6) により、あらかじめ用意しておいたテーブルを用いて角度空間への投射を行なう。その結果を図 4.5 に示す。図のように、各々のマイクロホン対の中での検出角度は、時間差検出結果に比べてかなり多く検出されることになる。これは二章で説明したように、一つのマイクロホン対では一つの時間差より二つの方向が示されるからであり、その結果時間差の角度空間への投射を行なうと、これだけ雑多な方向が示されることになる。

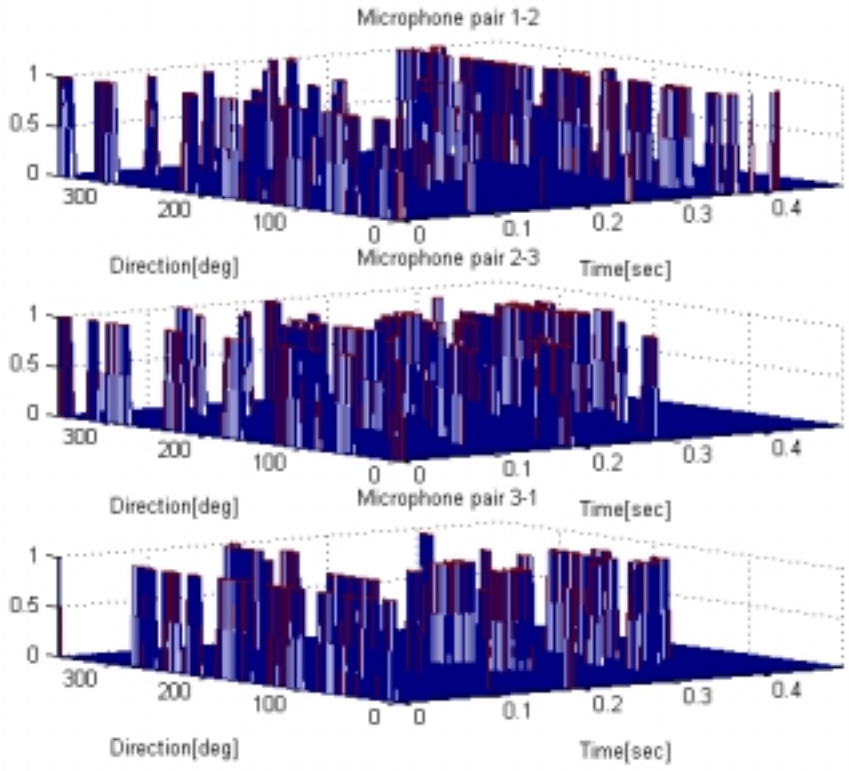

図 4.5: 角度検出結果

4.1.5 検出角度の統合

３組のマイクロホン対で得られた角度を二章で示した手法により統合したものが図 4.6 のようになる。

図のように、一つ一つのマイクロホン対の結果ではあれだけ多く現れていた検出方向が、統合することにより同時に同じ方向を示す結果のみを残して省かれているのが分かる。

4.1.6　音声信号による検出ポイントの抽出

検出角度の統合結果を元に、三章で説明したアルゴリズムを用いて音声信号による検出結果（連続的に同じ角度が検出されている箇所）の抽出を行なう。その結果を図 4.7 に示す。

検出角度の統合結果の時の図 4.6 を見れば分かるが、音声方向の検出角度（180 °）以外の箇所にもいくつか得られている検出角度が存在している。しかし、図 4.6 の結果では、音声方向（180 °）に存在する検出角度のみが残っている。

したがって、時間継続と時間平均の操作によって音声信号によって得られる検出結果の抽出が行なえていることが確認された。

そして、この最終結果を見てみると設定した音声方向である 180 °の方向のみが検出されており、残響として入力している 2 つの反射音の方向には検出点が現れていない。つまり、本手法で直接音の方向のみを検出できることが確認された。

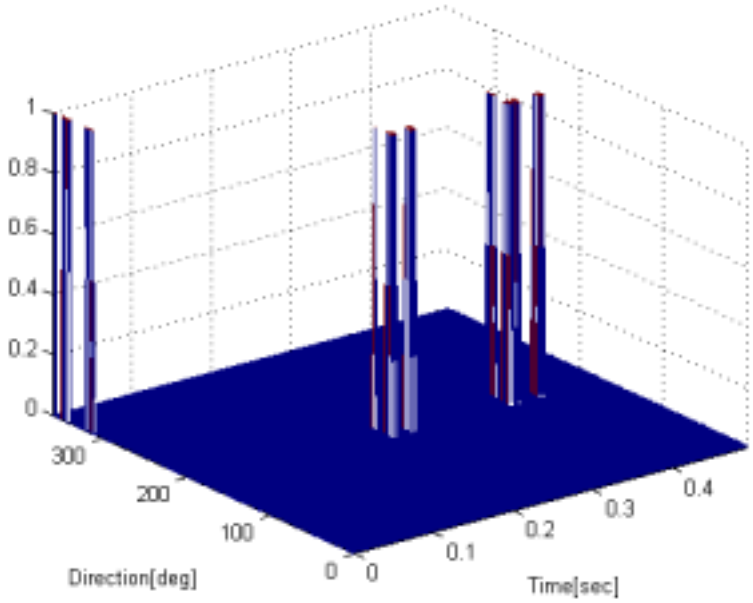

図 4.6: 検出角度統合結果

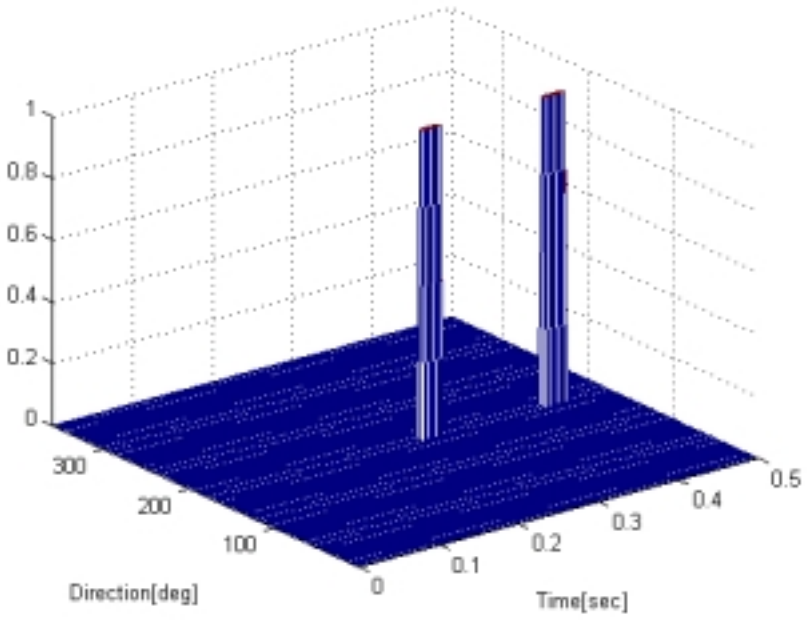

図 4.7: 時間平均処理結果

4.2　シミュレーション結果

本手法の妥当性の調査を行なうために、計算機上で作成した信号を元に単語、及び単音節母音を用いて音源方向推定のシミュレーションを行なった。以下に、その結果について報告する。

4.2.1　単語音声での音源方向推定結果

最初に、単語音声を入力信号として音源方向の検出精度をシミュレーションによって確認する。

シミュレーションの条件はマイクロホン間隔とサンプリング周波数を 4.1 節の時と同じとして、以下の通りである。

使用音声	ATR データベース単語音声
	男性話者３名、女性話者３名
	各の話者について各単語６個
雑音	白色雑音
	SNR を 30dB から-5dB きざみで 10dB まで変化
音声方向	ランダムに設定
第一反射音方向	25 °
直接音対第一反射音パワー比	-4dB
第二反射音方向	ランダムに設定
直接音対第二反射音パワー比	-8dB

以上のような条件で、計算機上で入力方向より時間差を求め、各々に時間差を加えた信号を作成し、３つのマイクロホンに入力する

また、4.1 節の時と同様に作成する。

このシミュレーションによって得られた結果を次の図 4.8、表 4.1 に示す。

図 4.8 は横軸に方向推定結果の精度、つまり推定角度が音声方向に対し±何度以内に入ったかを表し、縦軸に一つの音声信号で得られた検出点の平均数を示しており、最もノイズの大きい SNR10dB の時の結果と、最もノイズの小さい SNR30dB の時の結果を示している。また、表 4.1 にはその中で各々の SNR の条件で± 5 °以内に入った検出点の平均個数を表す。

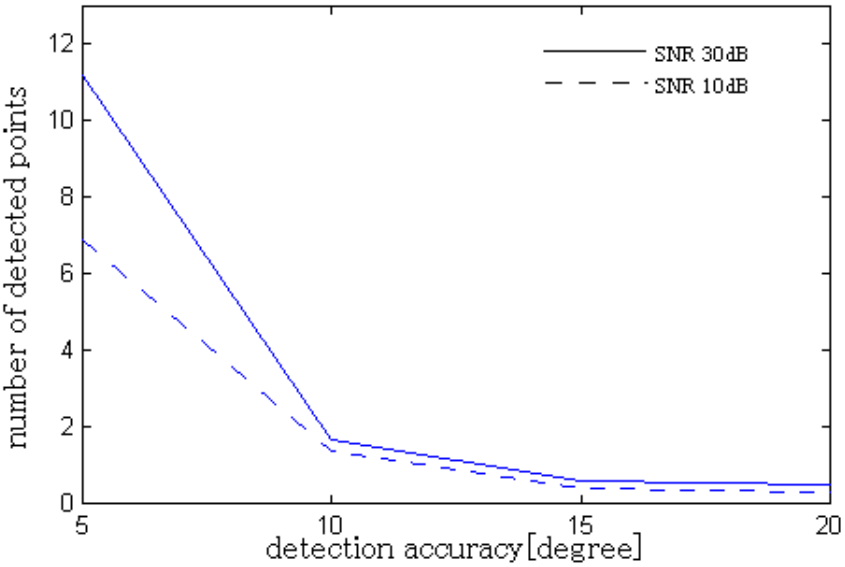

図 4.8: 検出点出現数：単語

図 4.8 の結果を見てみると、± 5 °以内に入った検出点数が最も多く現れ、それから離れた角度では急激に検出点数が減少しているのが分かる。

また、図 4.9 に検出ポイントの分布を示す。ここで横軸は先ほどと同じく精度を表し。縦軸に検出点の分布率を表す。ここで、検出点の分布率とはその角度以内に入った検出点数を正解数として、結果より得られた検出点の総数を有効点数とし、正解数を有効点数で割った値を百分率で表したものである。

これを見てみると、推定された音源方向は± 5 °以内に 70〜90 %存在しているのが分

表 4.1: 検出点分布：単語

		検出点出現位置 [degree]			
		±5	±10	±15	±20
検出点 出現数 [個]	SNR 10dB	6.83	1.37	0.38	0.08
	SNR 15dB	7.05	0.98	0.40	0.22
	SNR 20dB	9.15	1.48	0.48	0.27
	SNR 25dB	9.82	1.48	0.56	0.11
	SNR 30dB	11.17	1.67	0.60	0.55

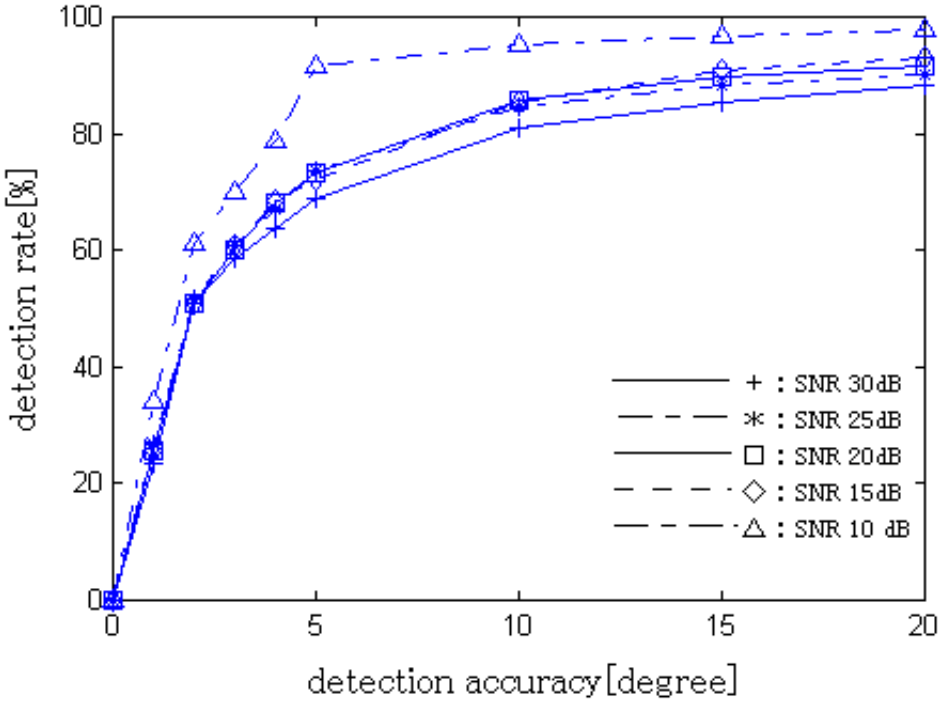

図 4.9: 検出点分布：単語

表 4.2: 音源方向推定結果：単語

	SNR 10dB	SNR 15dB	SNR 20dB	SNR 25dB	SNR 30dB
音源方向推定率 [%]	86.6	67.0	68.4	68.7	63.7

かる。また、± 15 °以内に SNR に関わりなくほぼ 80 %の検出点が入っている。± 0〜3 °の中に入った割合は少なくなっているが、この実験条件では 360 °方向を 200 分割する解像度になるので、実際の音源方向より多少はずれるためである。（マイクロホン間隔を広げる、またはサンプリング周波数を高くすることで解像度は上昇する）

また、± 5 °以内に入ったものを正しく方向推定がされたものと考えて、音源方向推定率を± 5 °内に入った正解数を有効点数で割った値と定義し、SNR と音源方向推定率の関係を表 4.1、図 4.10 に示す。

これを見てみると、ノイズによる音源方向の認識率の変化はほとんど無い様に思われる。むしろ、ノイズが大きい時に高い推定率が出現している場合さえある。

図 4.8 に示した結果を合わせて、この原因を考えてみる。音源方向推定率は± 5 °以内に

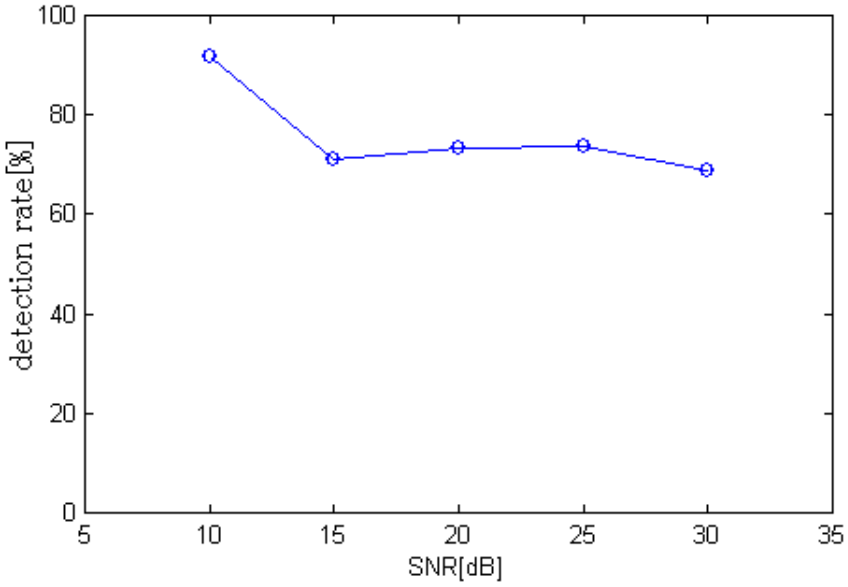

図 4.10: 音源方向推定結果：単語

現れた検出点の数を、入力信号中に現れた検出点の総数（有効点数）で割った値になる。

ここで、SNR に対する検出点の数を比べてみると（図 4.8）、ノイズが大きくなるにつれて、正しい方向を示す検出点の数は減少しているのが分かる。しかし、検出点の総数自体も減っているために、正しい方向を示す割合自体はさほど変化しないような結果が現れることになる。

また、もう一つの傾向としてはノイズが大きくなると正しい方向以外を示す検出点の数が減少するという結果が見られる。したがって、方向推定率を計算する際に分母となる検出点の総数が少なくなるために、方向推定率自体は上昇するという結果が得られている。

4.2.2 母音毎の音源方向推定結果

次に、母音毎の方向推定を行なった結果を示す。

シミュレーション条件は先ほどの単語音声の時と同様であり、使用音声は ATR データベースの男性話者三名、女性話者三名の単音節母音を用いた。

先ほどと同様に、± 5 °以内に検出点が得られた箇所を正しく音源方向が行なわれたと考えて、SNR と音源方向推定率の関係を表 4.3、図 4.11 に示す。

この図を見てみると、SNR に関係なく各母音毎に推定率が異なっているのが分かる。母音 /a/ /o/が、比較的良く方向推定が行なわれており、/i/ /u/ /e/については推定率が低くなっている。

表 4.3: 音源方向推定結果：母音

		SNR 10dB	SNR 15dB	SNR 20dB	SNR 25dB	SNR 30dB
音源方向	母音 /a/	79.2	78.7	58.6	60.8	75.2
推定率 [%]	母音 /i/	61.6	22.8	33.8	33.5	27.1
	母音 /u/	40.5	39.1	44.2	35.1	32.0
	母音 /e/	47.0	48.8	50.3	54.5	43.1
	母音 /o/	78.7	63.1	75.7	70.3	78.0

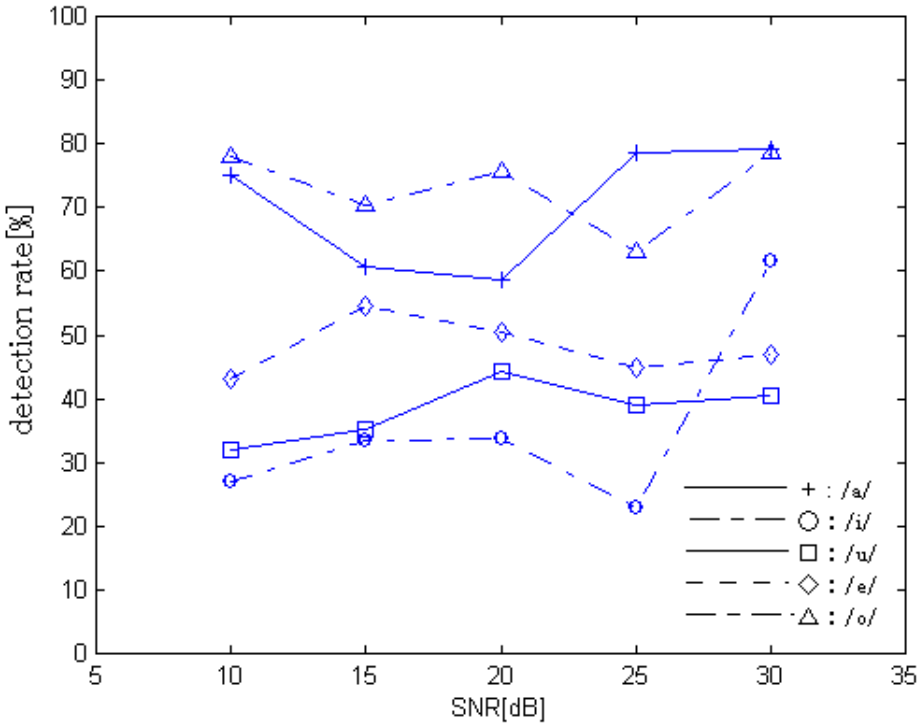

図 4.11: 音源方向推定結果：母音

この原因を各母音の特徴を見た上で考えてみる。単音節母音の信号の一例として、母音/a/と、母音/i/の信号の一例を図 4.12、図 4.13 に示す。

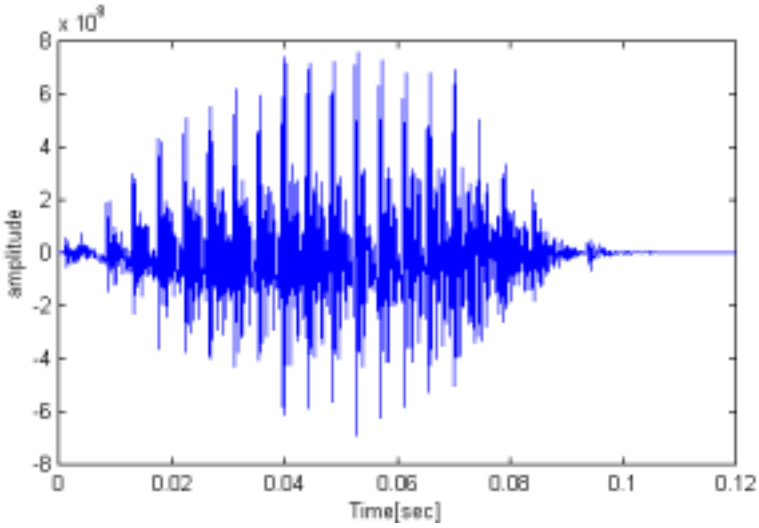

図 4.12: 母音 /a/

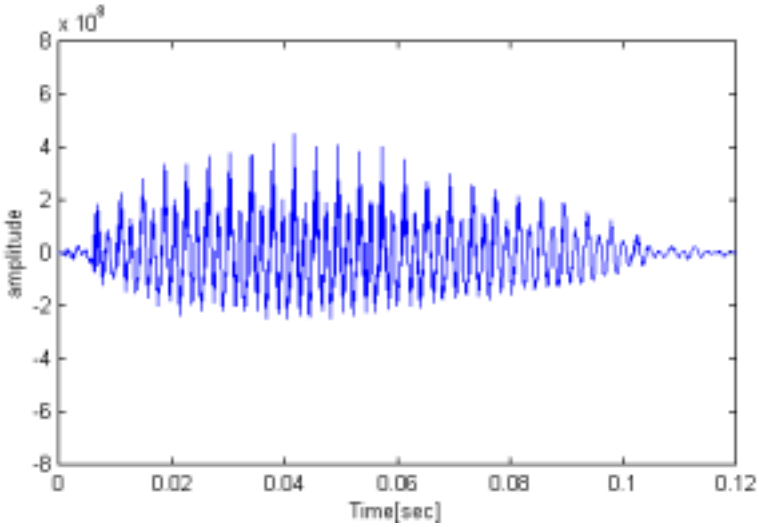

図 4.13: 母音 /i/

ここで両者を比較すると、音声の最大振幅が異なっているのが分かる。母音/a/のほうが、/i/に比べて振幅が大きく。その分ノイズを加えた後も、ノイズに対して peek-to-peek 値が大きくなっている。

変動閾値を用いた立ち上がり検出の際に、手がかりとして振幅が急激に変動する場所を捉える。したがって、音声信号がノイズの振幅よりも大きな振幅を持っている方が、その

手がかりを捉えやすいために、このような差が生じると考えられる。

各母音の中では、/a/と/o/がこのように比較的振幅が大きく、/i/、/u/、/e/ はそれよりも振幅が小さくなっている。各母音に対する方向推定率を比べてみると確かに、振幅の大きいとされる、/a/、/o/は推定率が良く、/i/、/u/、/e/は推定率が低くなっている

また、もう一つの要因として振幅包絡の変化の大きさがあげられる。これは先ほどの最大振幅の大きさにも関係することであるが、図 4.12 の/a/の信号と図 4.13 の/i/の信号とを比較すると、/a/の信号は大きく、はっきりとした山型の振幅包絡を持っている。それに比べ、/i/の信号は最大振幅が小さいことも関係し、小さく、そして一定とも言えるような非常になだらかな振幅包絡を持っている。

ここで、変動閾値の特性を考えてみると、変動閾値は三章で説明した通り、音声の立ち上がり部分、つまり音声の振幅包絡が大きくなっていく場所で反応する特性を持っている。したがって、/i/の様な振幅包絡の変化が小さな信号よりも、/a/の様な振幅包絡の変化が大きい信号に対して良く反応することになる。

以上のような理由から、母音毎の音源方向推定の精度に差が生じることになる。これは単語音声についても関係のあることだが、単語音声の場合は単語を構成する音声の中に様々な母音が含まれ、また母音同士が移り変わる箇所などで振幅包絡が変化する場所が多く存在するために、それら総合的な結果により単音節母音の結果より高い精度で音源方向が行なえる。

4.2.3 相互相関法との比較

次に、残響の含まれた環境における本手法の有効性を確認するために従来の相互相関を用いた手法との比較実験を行なう。

シミュレーション条件は以下の通りである。

使用音声	ATR データベース単語音声
	男性話者３名、女性話者３名
	単語音声 /a i ma i/ /to to no u/
雑音	白色雑音 （SNR 20dB）
音声方向	ランダムに設定
第一反射音方向	25 °
第二反射音方向	ランダムに設定

ここで、音声信号は先の単語音声による音源方向推定シミュレーションによって得られた結果、比較的方向推定率の高かった二つの音声を用いている。ここでは、シミュレーションの条件の通り SNR は固定として、残響のパワーを変更することで、音源方向推定率の変化を見る。

残響として第二反射音まで入力しているが、残響のパワーの変化量として、まず第一反射音として入力される波形は直接音から xdB パワーを減衰させた信号、そして第二反射音は、直接音から 2xdB パワーを減衰させた信号を入力している。

以上のような信号を用いて、相関を用いた手法との比較を行なう。これまでの結果と同様に、音声方向に対し± 5 °に推定された検出点を正しい方向が推定されたものとして、残響のパワーに対する音源方向推定率の変化を表 4.4、図 4.14 に示す。

表 4.4: 音源方向推定結果：相関との比較

		第一反射音パワー減衰量 (x)			
		-∞ dB(no-echo)	-14dB	-8dB	-4dB
音源方向 推定率 [%]	相関法	80.2	67.9	34.5	32.1
	本手法	86.6	89.6	89.5	84.0

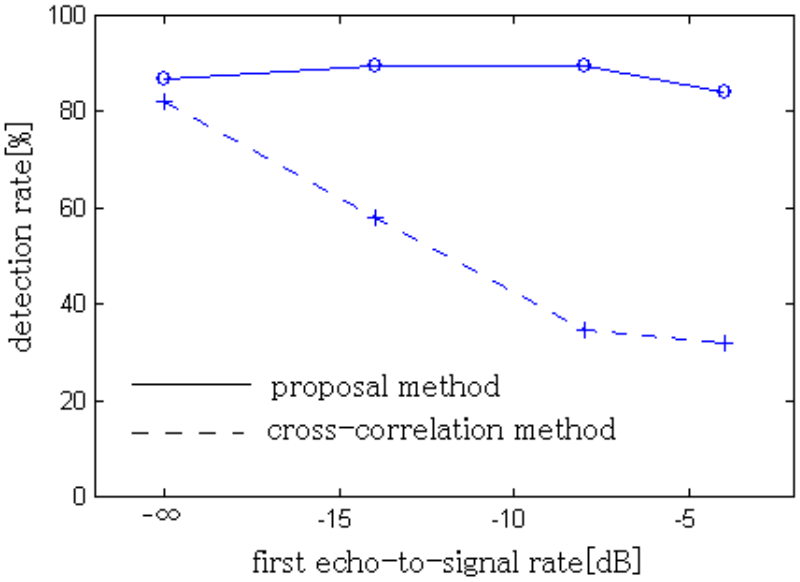

図 4.14: 音源方向推定結果：相関との比較

この結果を見ると、残響が含まれない時（直接音対反射音のパワー比がマイナス無限大の時）は本手法、相互相関を用いた手法の両方ともに高い精度で音源方向推定が行なえているのが分かる。しかし、残響成分のパワーを大きくするにつれて、相互相関を用いた手法では徐々に精度が低下してる。

相互相関を用いた手法では、一章で説明した通り残響によってその精度が悪化するという特徴があり、それはこの結果を見れば明らかである。しかし、本手法では相互相関を用いた手法に見られる残響成分による結果の変化が見られないことが明らかとなった。

4.2.4 シミュレーションまとめ

以上のようなシミュレーション結果より、単語音声、単音節母音など様々な音声信号に対して方向推定が行なえることが確認できた。

また、相関法との比較実験において本手法が残響を含む環境での音源方向推定において優秀な性能を持っていることが確認でき、変動閾値が、残響成分に影響されることなく直接音のみの情報を得ることができることが確認された。

第 5 章

実環境における音源方向推定実験

5.1　実験目的

これまでの結果で、計算機上で作成した信号において音源方向推定アルゴリズムの、ノイズ及び残響の含まれる環境での性能が確認された。

そこで、ここでは本研究において提案した音源方向推定法が実環境に対してどれほど有効であるのかを調査する。

5.2　音声収録

音源方向推定の実験を行なうため、本研究ではクリーンな音声をスピーカーより室内に出力し、設置したマイクロホンアレイで収音することによりデータの収録を行なった。

5.2.1　実験条件、使用機材

その際の実験条件は以下の通りである。

マイクロホン間隔	0.3m
サンプリング周波数	20kHz
	（音声収録時は 24kHz）
使用音声	ATR データベース
	男性話者 3 名、女性話者 3 名
	単語音声 /a i ma i/、/to to no u/
音声方向	180 °

また、音声収録時のサンプリング周波数は 24kHz であるが、音源方向推定には収録後サンプリング周波数 20kHz にダウンサンプリングした波形を用いる。

収録に用いた機材と、収録時のブロック図は以下の通りである。また、本実験で用いたマイクロホンアレイの外見を図 5.2 に紹介する。

表 5.1: 使用機材一覧

使用機材	メーカー	名称
パワー・アンプ	YAMAHA	P2040
スピーカー	JBL	RD135T
マイクロホンアレイ	*laboratory-made*	
マイクロホン	RAMSA	WMC75
マイクロホン・アンプ	TASCAM	MA-8
データ・レコーダー	TEAC	RD-135T

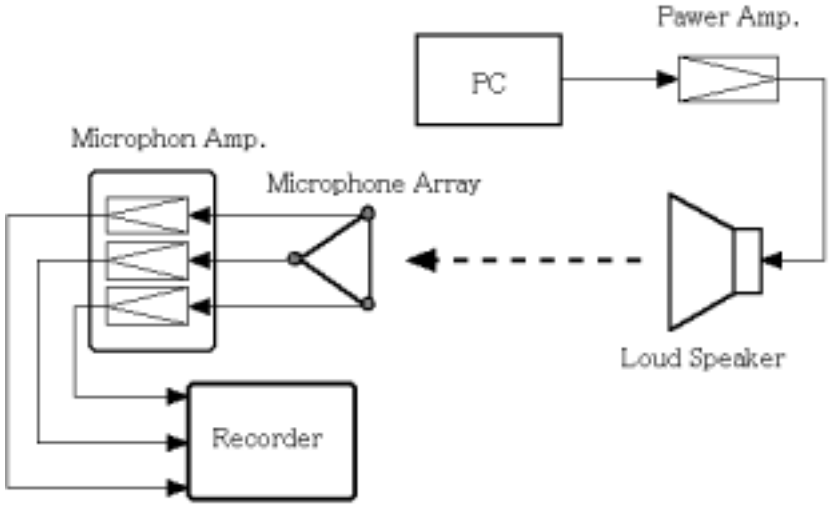

図 5.1: 音声データ収録時のブロック図

5.2.2 環境条件

本研究では、我々が日常生活する環境である残響を含む実環境を考慮して音源方向推定を行なうために、音声の収録の際に四方をコンクリートの壁で囲まれた部屋を利用する。

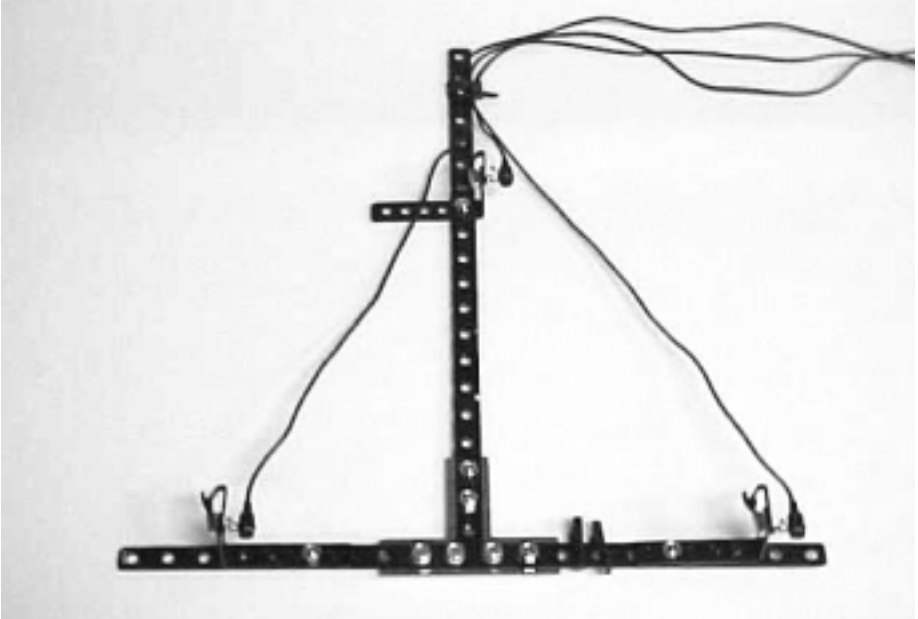

図 5.2: マイクロホンアレイ外見

この部屋での残響時間は、500〜1000Hz の帯域雑音を用いた計測の結果、0.56sec であった。

また、音声収録時の気象条件は、天候曇り、室温 27 度、湿度 38 %であった。

音声収録の際には、スピーカーとマイクロホンアレイの間隔を 3m とした。スピーカー及びマイクロホンアレイは各々部屋の壁から 2m 以上離れた場所に設置し、床から 1m の高さに設置して行なった。

音声収録時における雑音は、より実際に近い形として部屋に存在する環境雑音（各種機材のファンの音、空調システムのファンの音等）を利用する形で行い、その雑音レベルは 49.6dB(A) であった。

この時に、スピーカーから音声を出力する時の音量を調節することによって、SNR を調整した。

5.3 実験結果

実環境で収録したデータを元に音源方向推定を行なった結果を図 5.3 に示す。

これまでと同様に横軸に精度、縦軸に検出点の出現数を示す。

この結果を見ると、シミュレーションの時と同様に± 5 °以内程度にもっとも多くの検出点が出現しているのが分かる。

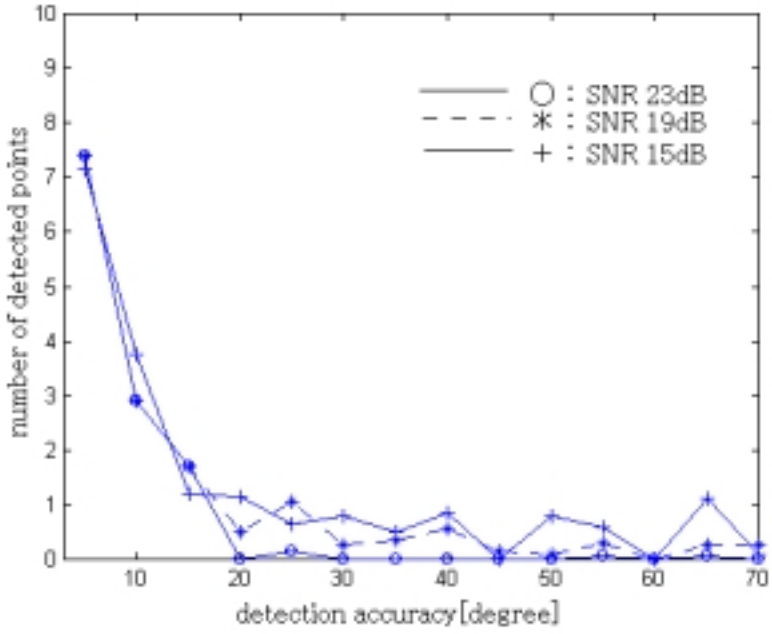

図 5.3: 検出点出現数：実環境

また、シミュレーション実験の時と同様に、ノイズが大きくなるにつれて音源方向付近を示す検出点の数が減少しているのが分かる。しかし、それとは別に音源方向とは離れた箇所に現れる検出点の数は増えている。これは、シミュレーション実験の時には見られなかった現象である。

次に図 5.4 に検出点の分布図を示す。

横軸に精度、縦軸に推定率である。この結果を見ると、± 5 °以内程度に全体の５割から７割程度の検出点が含まれているのが分かる。

また、これまでと同様に± 5 °以内に音源方向が推定された場合を正しく方向推定が行なわれたとして、SNR と推定率の関係を表 5.2、図 5.5 に示す。

この結果を見ると、シミュレーション実験の時とは異なりノイズが大きくなるにつれて推定率が低下していることが分かる。

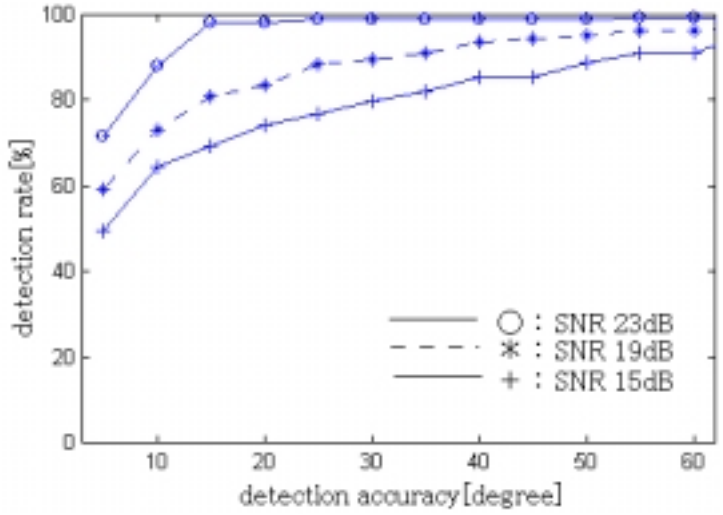

図 5.4: 検出点分布：実環境

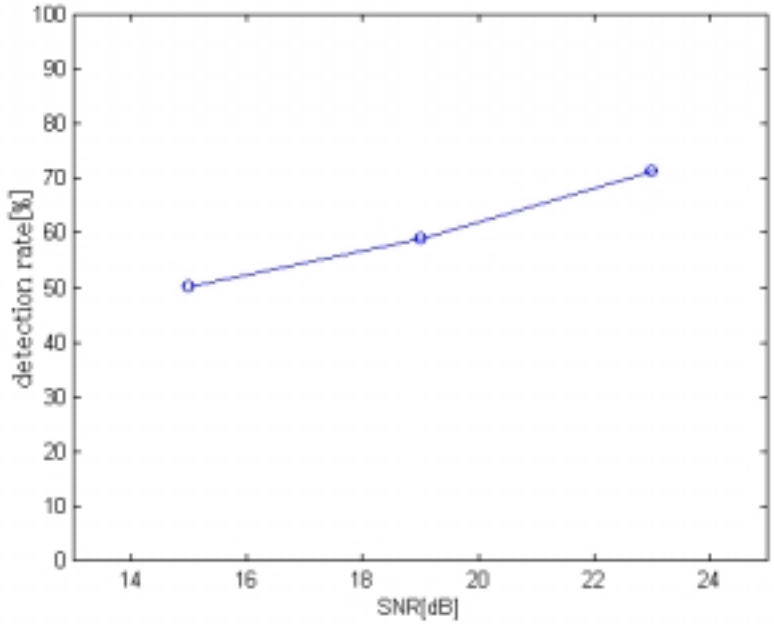

図 5.5: 音源方向推定結果：実環境

表 5.2: 音源方向推定結果：実環境

	SNR 15dB	SNR 19dB	SNR 24dB
音源方向推定率	50.2	59.0	71.3

5.4　考察

音源方向推定実験の結果を元に、本研究で提案手法について考察する。

音源方向推定おいて推定率はシミュレーション実験時よりも低下したものの、実際の音声方向付近に検出点が多く出現し、音源方向に対し±5 °以内に存在する検出点がもっとも多いことが確認された。

また、ノイズの大きさによって、推定率が悪化するという現象が確認された。その原因として以下のようなことが考えられる。

シミュレーション実験の際にノイズは白色雑音を用い、各マイクロホン間で相関性を持たないノイズが入力されていた。つまり、ノイズは方向性を持っていなかった。しかし、実際の環境でのノイズは方向性を持っている可能性があり、その方向性を持ったノイズの方向を検出したのではないかと考えられる。

実際に図 5.3 の検出点の分布を見てみると±65 °の辺りに比較的多くの検出点が存在している。音源方向が 180 °であるので、実際の方向にすると、各々115 °もしくは 245 °の方向になる。ここで、注目した方向は 115 °の方向である。実験環境を設定した時、この方向にちょうどパーソナルコンピュータが存在する配置になっていた。そのために、パーソナルコンピュータのファンなどから発生する音が方向性を持ったノイズとして作用したためであると考えることができる。

音声収録において、ノイズは環境騒音を用いており、そのパワーはほぼ一定に保たれていると見ることができる。今回、音声の音量を変化させることで SNR の調整を行なった。その時に、大きなノイズを得るために音量を小さくすることになる。すなわちこれは、音声信号とノイズの最大振幅が近付くことを意味している。

変動閾値を用いた手法は、三章で述べた通り信号の振幅を監視し、大きな振幅に対し反応するようになっている。したがって、音声信号の最大振幅とノイズの最大振幅が近い値になればなるほど、ノイズによる検出誤りが多く出現する可能性がある。そして、そのノイズが方向性を持っていた場合、そのノイズの方向を音声方向として検出するという現象が起こっていたのではないかと考えることができ、実環境で収録した信号を用いた時の、ノイズレベルの上昇による音源方向推定率の低下の原因と考える。

また、対策としては、音源方向推定アルゴリズムの最終段階で時間継続と時間平均処理によって音声信号による検出ポイントの選別を行なっているが、この時間平均窓長を長くとることによってある程度このような誤った検出点を減少させることができるのではないかと予想される。

第 6 章

結論

6.1　音源方向推定法について

本研究では、実際の環境において含まれる残響に影響されずに２次元平面の全方位に対する音源方向推定法を提案した。その手法は３チャンネルの正三角形配置のマイクロホンアレイを用い、聴覚における方向推定手法に示唆を得、それを簡易モデルとして工学的手法に応用した手法である。また、変動閾値を用いた手法により先行音効果を実現した手法である。

本研究において提案した手法について様々な条件の元でシミュレーション及び実験を行なった。その結果、音声の振幅包絡の上昇するところを音声の立ち上がりと捉えることにより、音源方向推定を行なうことができた。

また、従来の音源方向推定手法で問題点となっていた、残響の影響による精度の悪化を改善するために、本研究で変動閾値を用いた手法を提案した。計算機上で作成した信号によるシミュレーションや、実際に残響の含まれた環境で収録した音声を用いた実験の結果、変動閾値を用いることにより、残響成分を排除し直接音の音源方向のみが推定されることが確認された。

従来多く用いられている相互相関を用いた手法との比較実験の結果、残響を含んだ信号に対して音源方向推定精度の明らかな向上が見られ、本手法で提案した変動閾値を用いた手法が残響に対して従来の手法に比べ非常に優秀な性能を持つことが確認された。

6.2　今後の課題

本手法で、変動閾値を用いた手法によって残響に対して優秀な性能を有する音源方向推定法の構築が行なえた。しかし、実環境等での実験を行なった結果、多少ノイズに対して性能が悪化する傾向がある。これは、基本的な手がかりを信号の振幅に頼っているためであり、ノイズと信号の振幅比が小さくなると、ノイズの情報をイベントとして捉えてしまうためである。したがって、今後ノイズに対する対策を行なう必要がある。

また、本手法は基本的に単一音源に対する方向推定を行なう手法であり、同時に２箇所に音源が存在した場合、パワーの大きな（振幅の大きな）音源の方向しか推定できなくなっている。しかし、方向推定の応用分野によっては同時に２つの音源方向を推定しなければならない状況も考えられるため、このような状況に対して同時に２方向の音声を推定する手法を考える必要がある。

本来音源は３次元空間上に存在しており、その方向を推定することが必要である。本研究では、３次元空間上の全方位に対する音源方向推定の前段階として、２次元平面全方位に対する音源方向推定を前提に研究を行なってきた。その手法として、正三角形配置のマイクロホンアレイを用いたが、将来３次元空間全方位に対する音源方向推定を行なう上では不十分である。そこで、３次元空間上の音源方向を推定するために本手法を拡張する必要がある。具体的な手法として、マイクロホンをもう１つ増やし、正四面体の頂点に配置したマイクロホンアレイを用いることが考えられる。この手法では、本研究の手法と比べると、マイクロホンアレイで存在するマイクロホン対の数が倍になることになる。したがって、取り扱う情報量が倍になることになり、また３次元上の音源を推定するために複雑な幾何学的計算を行なうことが必要になる。

謝辞

本研究を進めるにあたり、終始熱心に御指導下さいました赤木 正人教授に厚く御礼を申し上げます。

また、パターン関連研究室合同セミナーなどで、熱心な議論ならびに多くのアドバイスを下さいました、諸先生方及び学生の方々に厚く御礼申し上げます。

また、日頃から研究および普段の生活にて、多大な御協力をいただきました、赤木研究室の学生そしてＯＢの方々そして友人を始めとする多くの皆様に感謝致します。

参考文献

[1] 永田・安部、'多数センサによる音源波形の推定 '、日本音響学会誌 47 巻 4 号、1991

[2] J.L.Flanagan,'Autodirective Microphone Systems',ACOUSTICA Vol.73,1991

[3] 赤木・水町、'マイクロホン対を用いた雑音除去法 (NORPAM) '、信学技 SP97-34,1997.7

[4] 安部、'多数センサによる音源推定 '、日本音響学会誌 51 巻 5 号、1995

[5] Maurizio Omologo, 'Acoustic Event Localization Using A Crosspower-Spectrum Phase Based Technique',IEEE,1994

[6] Piergiorgi Svaizer,'Acoustic Source Localization In A Three-Demensional Space Using Crosspower Spectrum Phase',IEEE,1997

[7] Jie Huang,'Auditory Spatial Processing in Reverbarant Environment', ITC-CSCC,1997

[8] L.A.Jeffress, 'A Place Theory of Sound Localization', J.Comp.Physiol.Pychol.,1948

[9] 小林・穂刈・島田、'複数マイク自由配置による複数話者位置推定 '、電子情報通信学会論文誌 A Vol.J82-A No.2 pp.193-200、1999

[10] 田中・金田・小島、'音源方向推定法の室内残響下での性能評価 '、音響誌 Vol.47 no.4 pp268-273,1991

[11] 黄・大西・杉江、'生体に示唆を得た音源定位システム 一 反響のある環境での単一音源定位 一 '、電子情報通信学会論文誌 A Vol.J71-A No.10 pp.1780-1789、1988

[12] Jie Huang,'Mobile Robot and Sound Localization',IEEE,1997

[13] 電子情報通信学会、'音声と聴覚 '、コロナ社、1980

[14] J.O.Pickles、’ピクルス聴覚生理学 ’、二瓶社、1995

[15] Brian C.J.Moore,’Hearing’,ACADEMIC PRESS,INC.,1995