

Title	雑音抑圧処理を付加した音源方向推定に関する研究
Author(s)	小林, 利秋
Citation	
Issue Date	2001-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1472
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修士論文

雑音抑圧処理を付加した音源方向推定に関する研究

指導教官 赤木 正人 教授

審査委員主査 赤木 正人 教授

審査委員 小谷 一孔 助教授

審査委員 下平 博 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

910048 小林 利秋

2001年2月

目次

1	序論	1
1.1	はじめに	1
1.2	研究の背景	2
1.2.1	複数のマイクロホンを用いた音源方向推定	2
1.2.2	従来の研究	2
1.3	本研究の目的	4
2	雑音抑圧処理	6
2.1	音源方向推定の前処理としての雑音抑圧	6
2.1.1	目的音の基本周期に基づく時間 - 周波数フィルタ	6
2.2	雑音、反射音を含む音声の基本周波数推定	11
2.2.1	反射音を含む音声の基本周波数推定	11
2.2.2	反射音を含む音声のスペクトル	11
2.2.3	基本周波数推定の雑音に対する強化	12
3	音源方向推定	18
3.1	マイクロホンアレイ	18
3.1.1	マイクロホンアレイの形状	18
3.1.2	マイクロホン対の検出角度の統合	19
3.1.3	マイクロホンアレイにおける角度量子化幅	20
3.2	音源方向推定のアルゴリズム	24
3.2.1	雑音抑圧処理	24
3.2.2	立ち上がり検出	24
3.2.3	時間差検出と角度空間への投射	27
3.2.4	検出点の絞り込み	28

4	方向推定シミュレーション	30
4.1	単音節母音に対する音源方向推定結果	30
4.1.1	雑音に方向性がない場合のシミュレーション結果	30
4.1.2	方向性雑音を付加した場合のシミュレーション結果	32
4.2	単語音声に対する音源方向推定結果	37
4.2.1	雑音に方向性がない場合のシミュレーション結果	37
4.2.2	方向性雑音を付加した場合のシミュレーション結果	41
4.3	音場シミュレーションを用いた音源方向推定実験	44
4.3.1	イメージ法	44
4.3.2	シミュレーションに使用した室内モデル	44
4.3.3	シミュレーション条件	45
4.3.4	残響環境下で性能が低下した原因	50
5	結論	53
5.1	本研究についてのまとめ	53
5.2	今後の課題	54

目次

1.1	2つのマイクロホンで受音される音波	3
2.1	雑音抑圧ブロック図	7
2.2	時間 - 周波数フィルタ ブロック図	7
2.3	基本周期 5ms に合わせた時間 - 周波数フィルタの振幅特性 (上) = 0.2 (中) = 0.5 (下) = 0.8	9
2.4	雑音抑圧処理例 (a) マイクロホン入力波形 ($x(t)$) (b) 基本周波数推定結果 ($1/T$) (c) 雑音抑圧波形 ($\hat{s}(t)$)	10
2.5	基本周波数推定結果例 (a) S N R 10dB の白色雑音と反射音が付加された faf /o/ (b) ケプストラム法による基本周波数推定結果	13
2.6	前置フィルタの特性 ($f_0 = 200$ の場合)	14
2.7	前置フィルタ処理を適用した基本周波数 (F_0) 推定のフローチャート	15
2.8	フレーム保護を含んだ音声区間判定の状態遷移図	16
2.9	基本周波数推定結果例 (a) S N R 10dB の白色雑音と反射音が付加された faf /o/ (b) ケプストラム法による基本周波数推定結果 (c) 前置フィルタによる強化後の基本周波数推定結果	17
3.1	マイクロホンアレイ形状 ($d = 0.3m$)	19
3.2	マイクロホン対 3 組の検出角度の統合	20
3.3	マイクロホン対 m1-m2、音源方向	21
3.4	角度量子化幅 $\Delta\theta$ の角度依存性 (マイクロホン対 : m1-m2)	22
3.5	正三角形配置のマイクロホン対の角度量子化幅 $\Delta\theta$ (上) 各マイクロホン対の $\Delta\theta$ (下) マイクロホン 3 対の統合結果	23
3.6	音源方向推定流れ図	25
3.7	雑音抑圧処理 (a) マイクロホン入力波形 (b) 基本周波数推定結果 ($1/T$) (c) 雑音抑圧波形	26

3.8	変動閾値による立ち上がり検出 (a) ピーク抽出波形と変動閾値 (b) 立ち上がり検出結果	27
3.9	(a) マイクロホン入力波形 (b) 音源方向推定結果	29
4.1	母音に対する音源方向推定結果	32
4.2	母音に対する音源方向推定結果：方向性雑音付加	33
4.3	方向性雑音に関する方向推定結果例 (母音) (a) 入力波形 mau/a/ + 雑音 + 反射音 (b) 本研究の提案法による推定結果 (c) 西田らの推定法による結果	34
4.4	母音毎の音源方向推定結果：方向性雑音	36
4.5	単語音声に対する音源方向推定結果：雑音方向性無、反射音なし	38
4.6	正しい音声方向が得られたサンプル数：雑音方向性無、反射音なし	38
4.7	単語音声に対する音源方向推定結果：雑音方向性無、反射音あり	40
4.8	正しい音声方向が得られたサンプル数：雑音方向性無、反射音あり	40
4.9	単語音声に対する音源方向推定結果：方向性雑音、反射音あり	42
4.10	正しい音声方向が得られたサンプル数：方向性雑音、反射音あり	42
4.11	方向性雑音に関する方向推定結果例 (単語) (a) マイクロホン入力波形 (b) 本研究の提案法 (c) 西田らの推定法 (d) 白色化相互相関法 (CPSP) に基づく推定法	43
4.12	音の反射と仮想音源	46
4.13	シミュレーションで用いた室内のモデル	47
4.14	残響環境の音源方向推定結果：SNR 10dB	48
4.15	正しい音声方向が得られたサンプル数：SNR 10dB	48
4.16	残響環境の音源方向推定結果：SNR 20dB	49
4.17	正しい音声方向が得られたサンプル数：SNR 20dB	49
4.18	残響環境の音源方向推定結果：SNR 30dB	50
4.19	正しい音声方向が得られたサンプル数：SNR 30dB	50
4.20	mau /a i ma i/ の元波形と残響が付加された波形	52

表 目 次

3.1 動作例に用いたシミュレーション条件	24
4.1 母音に対する音源方向推定結果	31
4.2 母音に対する音源方向推定結果：方向性雑音付加	33

第 1 章

序論

1.1 はじめに

我々は実生活の場面で、複数の音源から到来する混合音を聴いている。そして我々はそのような環境下においても、通常特定の音源の音だけを抽出して聴くことを行っており、このような現象は「カクテルパーティ現象」と呼ばれている [31]。このカクテルパーティ現象にはいまだ未知の部分が多いが、それをもたらす要因の一つとして、音源の方向を検知する能力が関わっているとされている。

このことを工学的に応用して、音の到来方向の情報を利用した、目的音の抽出、または目的音以外の雑音の抑圧、さらにそれらを前処理として用いた音声認識等の研究が、マイクロホンアレイを用いた音情報処理の分野において、現在までに極めて数多く行われている [12][24][4] [9]。そしてこれらの研究を支える前提条件のひとつは、目的音の到来方向を正しく検出することにある。

一方、パーティ会場のように、たくさんの人の話声、音楽の音、食器が触れ合う音等、多くの音が存在する環境で、ふいに後ろから自分を呼ぶ声に反応し、振り返って声の主を確認することも、よくある光景と思われる。この光景から連想し、著者は音源方向推定を研究テーマに設定した際、カクテルパーティ現象を変形した以下の問を自問した。

「我々は複数の音源から到来する音の混合音から、いかに目的音の到来方向を知るのか？」

その問に対し著者自らの体験から、音源の種類を区分することにより目的音に着目して、その到来方向を検知するのではないだろうかと考えた。冒頭に述べたカクテルパーティ現象の件では、音の到来方向の情報が音源の区分に役立つと記したが、これとは逆に、音源を区分することが、音の到来方向の検知を補助するということも、また言えるのではないだろうか。

以上は個人の思いつきであるが、このことを音源方向推定システム上で工学的に応用できないかと考えたことが、本研究の雑音抑圧を付加した音源方向推定に着手した発端である。

1.2 研究の背景

音源方向推定は、前節で述べたような音源方向の情報を用いた雑音抑圧、音声認識の前処理の他に、ロボットなどの環境認識システムへの応用 [10]、ビデオ会議システム等で話者方向に自動的にカメラを合わせる自動カメラ制御 [22] 等、広範囲な分野でその応用がもとめられ、さまざまな研究が行われている。

1.2.1 複数のマイクロホンを用いた音源方向推定

複数のマイクロホンを用いた音源方向推定は、一般に各マイクロホンに到達した音の時間差をもとに推定される。図 1.1 を用いてこのことを説明する。前提として、ここでは自由空間において、音波が平面波として到達すると仮定する。そして、この音波を間隔 d 離れた 2 つのマイクロホン (m_1, m_2) で受信することを考える。

図 1.1 の 方向から到来した音波は、まずマイクロホン m_2 において受信される。次に、音波は図に示す通り $d \sin \theta$ だけ進行してマイクロホン m_1 に到達する。従って、マイクロホン m_1 での受信信号は、マイクロホン m_2 での受信信号と比べて、音波が距離 $d \sin \theta$ だけ進むのに要する時間 τ だけ遅れた信号となっている。すなわちマイクロホン m_1 での受信信号を $x_1(t)$ 、マイクロホン m_2 での受信信号を $x_2(t)$ とすると

$$x_1(t) = x_2(t - \tau) \quad (1.1)$$

そして音速を c とすれば、

$$c\tau = d \sin \theta \quad \theta = \arcsin(c\tau/d) \quad (1.2)$$

従って、音速 c とマイクロホン間隔 d は既知であるため、 $x_1(t)$ と $x_2(t)$ の時間差 τ を決定できれば、式 (1.2) より音波の到来方向 θ を推定することができる。

1.2.2 従来の研究

式 (1.1) の時間差 τ は、通常 $x_1(t)$ と $x_2(t)$ の相互相関を最大とする時間差として決定する方法が最も一般的である。しかし雑音、反射音を含む実際の環境では、各マイクロホ

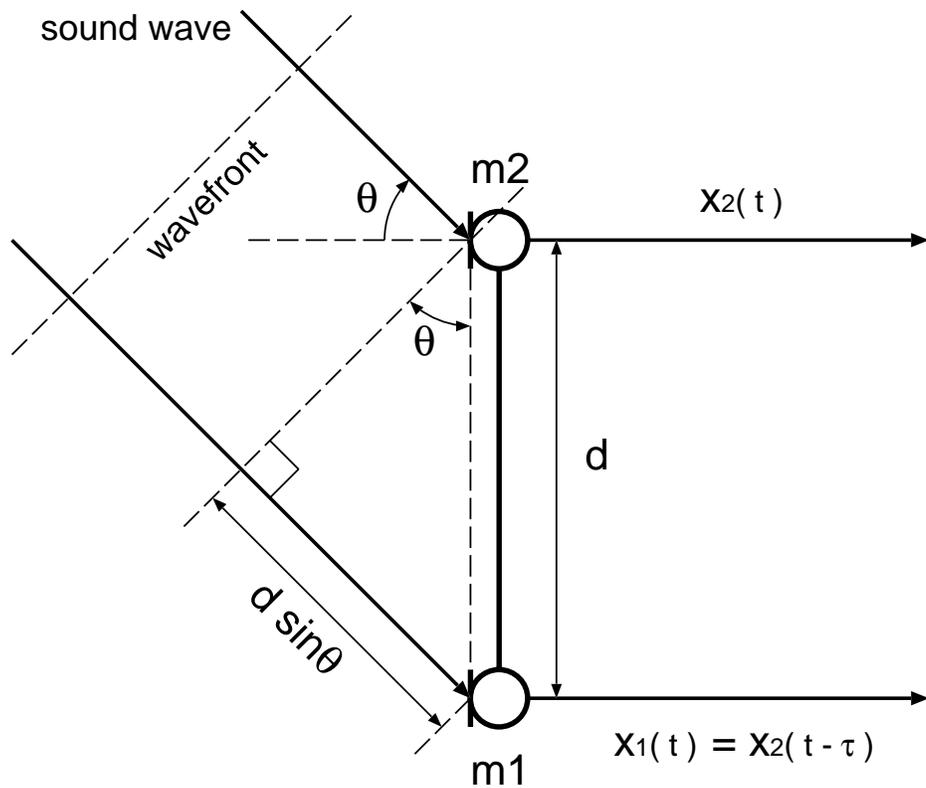


図 1.1: 2つのマイクロホンで受信される音波

ンの受信波形に不一致が生じるため、その場合この方法では、 $x_1(t)$ と $x_2(t)$ の時間差に対応した相関のピークを得ることが困難となり、時間差検出の精度が低下してしまう。

そこでこの問題点を改善するために、相互相関により得られた時間差検出結果に対し、選別処理を付加したさまざまな手法が考えられている。一例として Rabinkin ら [1] は、相互相関関数の最大値を与えるピークの形状に着目した。ピークが緩やかな相関係数から得られた時間差検出結果には誤りが含まれている可能性が高いため、その結果を破棄し、ピークが急峻であるものから得られた結果のみを正しい結果として採用する手法を提案している。また別の研究として Brandstein らの [8] では、相互相関により得られた時間差の時系列データの分散をもとめ、その分散が大きい区間の時間差検出結果を破棄し、分散が小さい区間の結果のみを採用する方法を提案している。これらの研究は、バックグラウンドノイズとして方向性を持たない雑音に対しては有効性がみとめられるが、しかし反面、目的音の他に、相互相関関数に鋭いピークを生じるような特定方向から到来する雑音が、定常的に存在する環境下では、選別を誤って雑音到来方向に対応した時間差を検出してしまふ可能性が大きい。実環境においては、雑音源があって、そこから雑音が到来する場合も多いことを考慮すると、方向性の雑音に対応していないこれらの方法には問題があると

言える。

一方、相関を用いない全く異なるアプローチで、反射音、残響に強い音源方向推定としては、西田らが提案する各々のマイクロホンに到達した音の立ち上がり点の時間差に基づく推定法がある [25]。この推定法では、反射音の一般的な性質である

- 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

ことに着目し、検出した音の立ち上がり点を利用して反射音に対する不応期を設定している。そしてその不応期制御により反射音による性能低下を回避している。しかし西田らの推定法では、各マイクロホンの受信波形の時間振幅情報のみを用いているため、定常雑音等、目的音と無関連な雑音が存在する環境下では、雑音の立ち上がりを検出してしまい、方向推定を誤るといった問題点があった。

これらの問題点を改善するには、目的音が音声である場合であれば、目的音声の有声音に見られる周期性、および調波構造に基づいて、マイクロホンで受信した信号中の音声と他の音を区別する等の処理を、音源方向推定に組み合わせることが有効と考えられる。そして実際にいくつかの研究でもそのことは指摘されている [10][9][22]。しかしその具体的な方法を提示し、かつ方向推定における効果を確認したという報告はごく少数しか見られない。

その数少ない報告例の中で、Brandstein の [6],[7] では、相互相関関数に以下に述べる特別な重みづけ関数を乗じることで、雑音と音声の区分をはかりつつ、到来時間差を推定する方法を提案している。そこではまず相互相関処理のフレームに合わせて、入力信号を声帯と声道からなるソースフィルタモデルで近似する。そしてそのモデルの近似誤差の逆数を、上述の重みづけ関数として用いることで、音声と異なる性質をもつ雑音に対して、その影響の軽減をはかっている。しかしこの手法ではこの重みづけ関数の処理において、目的音声と類似の性質をもつ反射音を区分することが困難であり、かつ相互相関による方法自体が反射音に対して弱いこととあいまって、残響環境下では性能が劣化してしまう。

1.3 本研究の目的

本研究では、音源方向推定の適用分野の多くで主たる目的音を音声としていることに着目し、目的音として音声を対象とした音源方向推定を扱う。

そして前節で記したように、雑音と反射音が混在する環境下において、現行の音源方向推定法は性能が低下するという現状に鑑み、本研究では、雑音と反射音の各々に対応

し、雑音と反射音が混在する環境下において有効な音源方向推定を提案することを目的とする。

その中で雑音に対応するために、音源方向推定システムに目的音声の有声音の基本周期を手掛かりとした時間 - 周波数フィルタによる雑音抑圧処理を組み合わせることを提案し、その雑音抑圧処理の方向推定における有効性を確認する。またその際にバックグラウンドノイズとして従来の研究でよく性能評価に使用される、方向性を持たない雑音だけでなく、方向性の雑音に対してもその性能を確認する。

さらにその雑音抑圧波形に対して、反射音に対する不応期制御を含んだ西田らの提案法を適用することで反射音に対応することにより、システム全体として、雑音と反射音に対し頑健な音源方向推定のアルゴリズム構築を目指す。

第 2 章

雑音抑圧処理

本研究では、音源方向推定の応用分野の多くで目的音として人間の音声を対象とすることに着目し、目的音として音声を対象とした音源方向推定を扱う。

その枠組の中で雑音に強いシステムを実現するために、方向推定の前処理に、目的音の音声の特徴を利用した雑音抑圧を設けたことを特色とする。

本章では、本研究で使用する雑音抑圧の詳細と、その雑音抑圧を実現するために要求される基本周波数推定について述べる

2.1 音源方向推定の前処理としての雑音抑圧

時間差に基づく音源方向推定の前処理としての雑音抑圧では、雑音抑圧波形が元の目的音のもつ時間情報を保存している必要がある。

そこで本研究では、そのような雑音抑圧手段として、推定した雑音成分を混合音より時間領域で引き去るキャンセレーションの時間 - 周波数フィルタにより雑音抑圧を行う。

この雑音抑圧のブロック図を図 2.1 に示す。 $s(t)$ を調波複合音である目的音、 $n(t)$ を雑音、 $x(t)$ を目的音と雑音の混合音とする。まずはじめに基本周期推定部で目的音の基本周期 T を推定する。そしてそこで得られた基本周期 T をもとに、時間 - 周波数フィルタ部にて、混合音中の雑音成分を抑圧し、雑音抑圧波形としての目的音の推定値 $\hat{s}(t)$ を得る。以下に、この時間 - 周波数フィルタ部の処理について説明する。

2.1.1 目的音の基本周期に基づく時間 - 周波数フィルタ

目的音の基本周期 T に基づく時間 - 周波数フィルタには、石本らのキャンセレーションによる帯域可変楕円フィルタ [13] に、若干の変更を加えたものを使用した。そのフィルタ

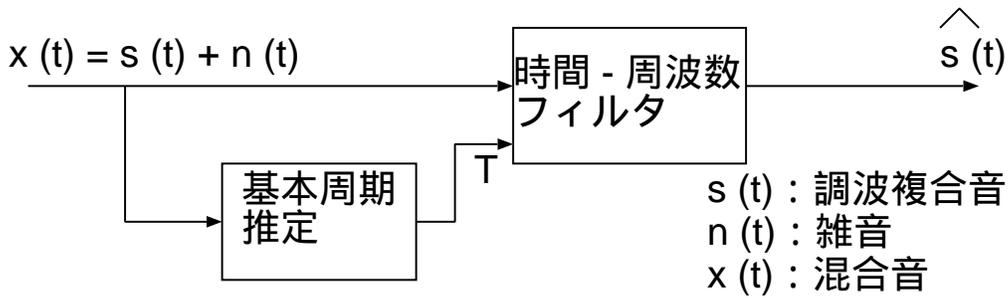


図 2.1: 雑音抑圧ブロック図

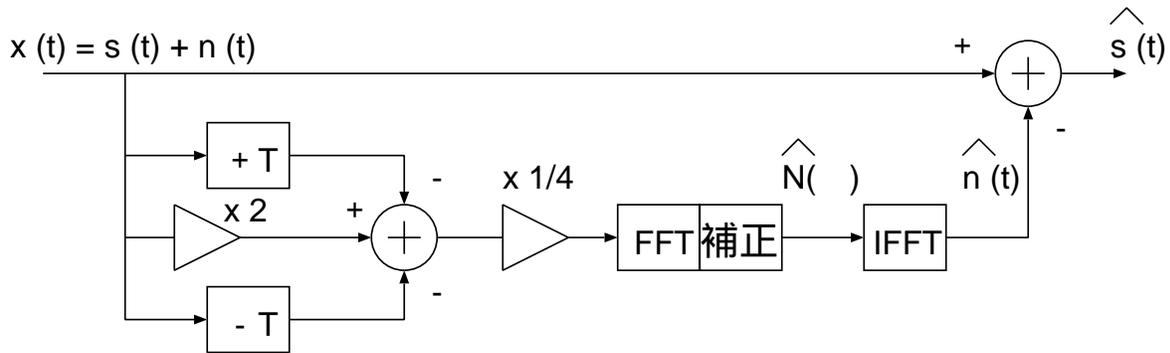


図 2.2: 時間 - 周波数フィルタ ブロック図

の処理について以下に述べる。

目的音 $s(t)$ の音声が、基本周期 T (基本周波数 ω_0) の調波複合音で代表されると仮定すると、マイクロホンに入力する混合音は、次のように表せる。

$$x(t) = s(t) + n(t) \quad (2.1)$$

$$= \sum_n a_n e^{j(n\omega_0 t + \theta_n)} + \sum_k b_k e^{j(n\omega_k t + \theta_k)} \quad (2.2)$$

キャンセレーション法では、まず混合音 $x(t)$ より目的音 $s(t)$ をキャンセルし、雑音成分 $n(t)$ を推定する。そのためにまず下記の如く $g(t)$ を定義する。

$$g(t) = \frac{2x(t) - x(t - T) - x(t + T)}{4} \quad (2.3)$$

$$= \sum_k b_k e^{j(n\omega_k t + \theta_k)} \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (2.4)$$

上式には目的音の成分が、振幅、位相項ともに全く含まれていないことに留意する必要があり、かつそれが重要なポイントとなっている。よってこの式を以下のように変換して得られる雑音成分には、理論的に目的音の成分は全く含まれないことになる。

次に (2.4) 式をフーリエ変換する。 $n(t)$ のフーリエ変換を $N(\omega_k)$ 、 $g(t)$ のフーリエ変換を $G(\omega_k)$ とすれば、

$$G(\omega_k) = N(\omega_k) \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (2.5)$$

よって雑音スペクトルは

$$N(\omega_k) = G(\omega_k) / \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (2.6)$$

となる。

上式では $\omega_k/\omega_0 =$ 整数の時、分母がゼロとなり、計算処理に不都合が生じる。よって上式を、微小量 ε を設定し、以下のように場合分けする。

$$N(\omega_k) = \begin{cases} G(\omega_k) / \sin^2 \frac{\omega_k}{\omega_0} \pi & \text{for } |\sin \frac{\omega_k}{\omega_0} \pi| \geq \varepsilon \\ 0 & \text{for } |\sin \frac{\omega_k}{\omega_0} \pi| < \varepsilon \end{cases} \quad (2.7)$$

この $N(\omega_k)$ を逆フーリエ変換した雑音 $n(t)$ を、もとの混合音から引き去ることにより目的音 $s(t)$ を推定としての雑音抑圧信号を得る。参考として、以上の処理を表したブロック図を図 2.1 に、本処理で実現される時間 - 周波数フィルタの振幅特性の一例を図 2.3 に示す。図 2.3 は、目的音の基本周期を 5ms(基本周波数 200Hz) とした場合であり、これをみると目的音の調波構造に合わせた楕円フィルタの特性を有し、かつ (2.7) 式の ε の値により、その通過帯域幅が変化していることがわかる。

雑音抑圧結果の例として、ATR データベースの男性話者の単語音声に SNR 10dB 相当の白色雑音と、直接音より 4dB 減衰した反射音を付加した入力に対する結果を図 3.7 に示す。本図では、本節で述べた処理に加え、基本周期 (基本周波数) が得られない区間に対しては目的音が存在しないものとみなし、その区間の雑音抑圧波形の振幅を 0 する処理を施している。

以上の雑音抑圧処理は、上述したように目的音の成分を完全に含まない雑音成分を混合音より引き去る処理であるため、得られる目的音の推定値は、元の目的音がもつ時間情報を理論的に保存しており、時間差検出の前処理に用いる雑音抑圧として適しているといえる。

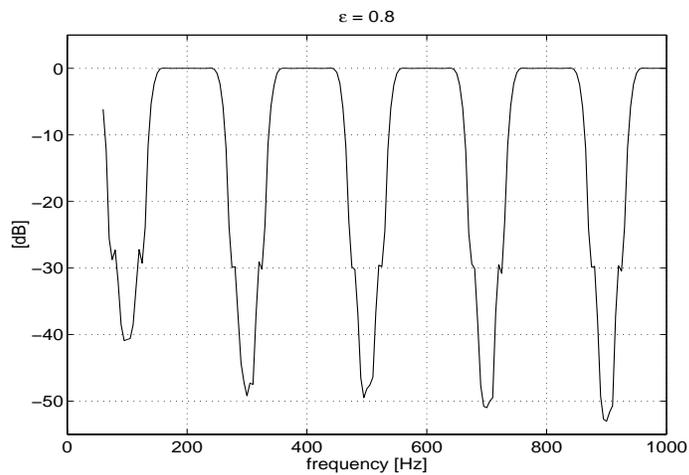
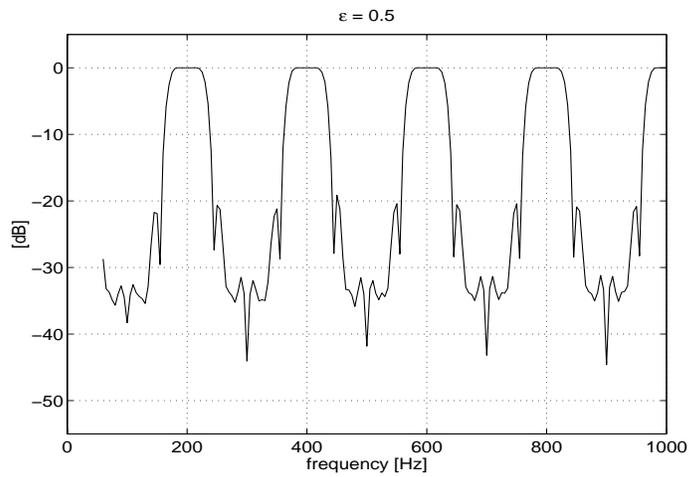
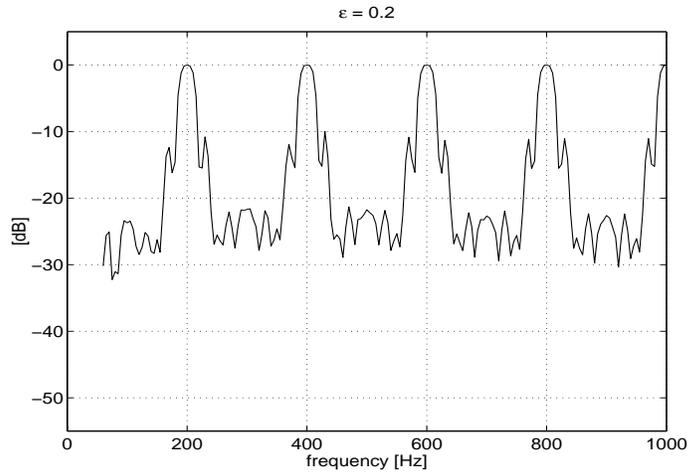


図 2.3: 基本周期 5ms に合わせた時間 - 周波数フィルタの振幅特性 (上) $\epsilon = 0.2$ (中)
 $\epsilon = 0.5$ (下) $\epsilon = 0.8$

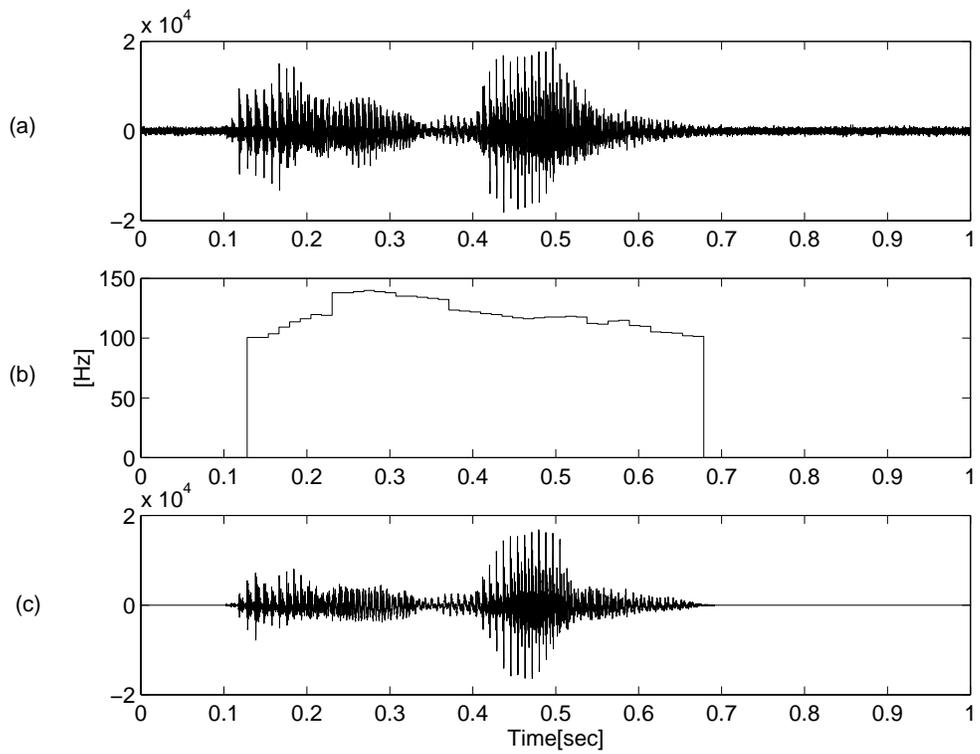


図 2.4: 雑音抑圧処理例 (a) マイクロホン入力波形 ($x(t)$) (b) 基本周波数推定結果 ($1/T$) (c) 雑音抑圧波形 ($\hat{s}(t)$)

2.2 雑音、反射音を含む音声の基本周波数推定

本研究の音源方向推定は、雑音と反射音が混在する環境での適用を想定しているため、本章の雑音抑圧を成立させる前提条件として、この環境で目的音の基本周期 (基本周波数) が推定できなければならない。

本節では、雑音と反射音が付加された音声の基本周波数推定について検討する。その際に音源方向推定の適用分野を想定し、

- SNR 10dB までの雑音環境
- 直接音からの遅れが数 ms の早い反射音

に対応することを念頭におくこととする。

音声の基本周波数推定法にはさまざまな方法があるが、主に自己相関、AMDF等の相関処理によるものと、ケプストラム等のスペクトル処理によるものに大別される。

それらに対し、石本らの研究 [13] を含め、雑音環境の基本周波数推定に関する研究報告は数多く存在するが、先の条件で記したような反射音に対応した基本周波数推定に関する知見は見られない。そこでまず、反射音を含む音声の基本周波数推定について検討する。

2.2.1 反射音を含む音声の基本周波数推定

反射音を含む音声に対して、相関処理による基本周波数推定法を用いると仮定した場合、反射音の到来遅延時間に対応した時間差で相関のピークが生じ、音声の基本周期に対応する相関のピークを不確かにする懸念がある。特に先にあげた条件をみたす反射音は、音声の基本周期より短い遅延時間で到来するため、この場合、推定がいつそう困難となることが予想される。

よって本節の検討は、以下に記すように、スペクトル処理によるものを主に行った。

2.2.2 反射音を含む音声のスペクトル

反射音を含む音声信号がどのようなものになるかを机上検討してみる。そしてまず問題を簡略化し、音声の第一調波 (すなわち基本周波数) 成分について考える。

音声の基本周波数を $\omega_0/2\pi$ とすると、その第一調波は時間波形で $a \sin(\omega_0 t + \phi_a)$ と表せる。そして上述したような短い遅れ時間で到来する反射音が付加された場合について検討する。

音声の基本周波数は時間的に緩やかに変動することが知られている。そのことに着目すれば、そのような反射音の基本周波数は直接音とほぼ等しくなり、その反射音の第一調波成分は $b \sin(\omega_0 t + \phi_b)$ と近似的に表せる (但し $b < a$)。

従って反射音と直接音の混合音 $y(t)$ は

$$y(t) = a \sin(\omega_0 t + \phi_a) + b \sin(\omega_0 t + \phi_b) \quad (2.8)$$

$$= \sqrt{a^2 + b^2} \sin(\omega_0 t + \phi_c) \quad (2.9)$$

すなわち混合音 $y(t)$ は直接音と、振幅、位相が変化するが、周波数は変わらないことが分かる。

次に直接音より大きく遅れて到来する反射音について検討する。この場合は音声の基本周波数の時間変動を考慮し、反射音の基本周波数が、直接音のそれから離れることを想定し、その反射音を、 $b \sin(\omega'_0 t + \phi_b)$ と表すこととする ($\omega_0 \rightarrow \omega'_0$)。しかし反射音は、遅れて到来するものほど減衰することを考慮すれば、この反射音と直接音の混合音 $y(t)$ は

$$y(t) = a \sin(\omega_0 t + \phi_a) + b \sin(\omega'_0 t + \phi_b) \quad (2.10)$$

$$a \sin(\omega_0 t + \phi_a) \quad (a > b) \quad (2.11)$$

従ってこの場合も、混合音の周波数は近似的に直接音の周波数となる。

これらの解析結果が各次数の調波成分で成り立つならば、結論として、直接音とその反射音の混合音では、その基本周波数は保存され、さらに式 (2.9) における混合音の振幅 $\sqrt{a^2 + b^2}$ の方が、直接音の振幅 a より大きいことに着目すると、特に早い初期反射音の重畳は、スペクトル領域において目的音声の調波成分を強調しているとみることができる。

以上は簡略化した解析ではあるが、この結論は、Y. Pan と A. Waibel が、彼らの論文 [15] 中”残響 対 ピッチ”の節で、実験データを附して“理論的に残響のマルチパス効果は、楕形フィルタを通した特性の如し”と論じていることと、その意味するところは合致している。

以上を踏まえると、反射音を含む音声の基本周波数推定にはスペクトル処理によるものが有利と考えられる。

そこで本研究では基本周波数推定に関し、スペクトル処理によるものの一つとしてよく知られる、ノルのケプストラム法 [11] をベースに検討した。

2.2.3 基本周波数推定の雑音に対する強化

ケプストラム法による基本周波数推定は、前節の検討より反射音に対しては有力視されるが、一方白色雑音等の帯域の広い雑音が付加されるとその性能が低下する問題点がある

[13]。

実際に、SNR 10dB 相当の白色雑音と、反射音を付加した単音節母音の性能評価実験において、いくつかのデータで、音声区間中に、ケプストラム法の基本周波数推定結果が得られない現象が確認された。

図 2.5にその一例として、ATR データベースの女性話者 *faf* の /o/ に、上述の雑音、反射音を付加した信号に対する基本周波数推定結果を示す。

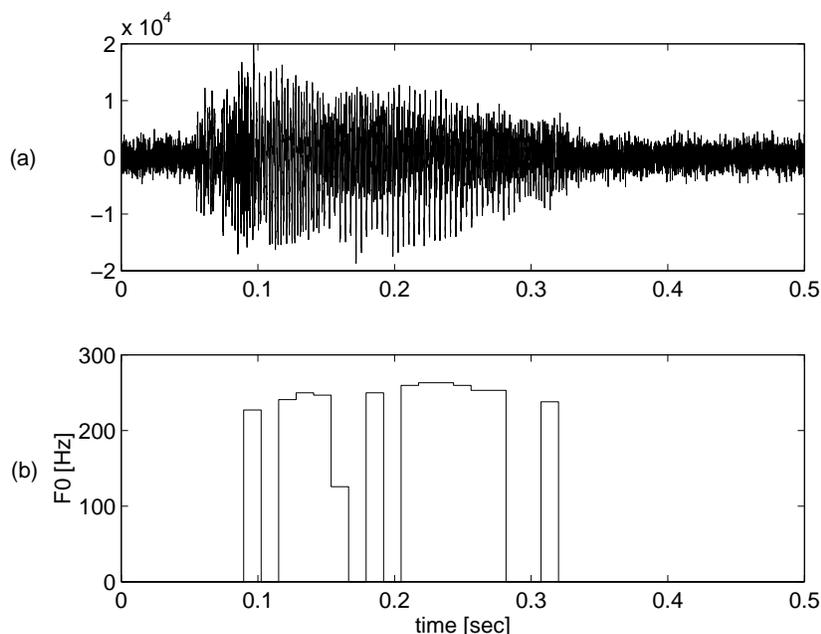


図 2.5: 基本周波数推定結果例 (a) SNR 10dB の白色雑音と反射音が付加された *faf* /o/ (b) ケプストラム法による基本周波数推定結果

図 2.5(a) の入力に対し、ケプストラム法による基本周波数推定は、図 2.5(b) に示すように推定結果に頻繁な欠けが見られる。

このような基本周波数推定結果に欠けが生じると、この基本周波数 (周期) を用いた雑音抑圧処理で、目的音声除去してしまい、方向推定に支障が生じる。

そこでこのような雑音に対応するため、ケプストラム法の基本周波数推定結果が得られない区間に対して、ケプストラム法の基本周波数推定に、雑音を抑圧するための前置フィルタを設けることで、基本周波数推定処理の強化をはかった。

この前置フィルタには、後段のケプストラム法を補助するために、目的音声の調波にあった楕円フィルタが望まれる。その楕円フィルタの実現には、フィルタの中心周波数 f_0 を決定する必要がある。

そこで本研究では、音声の基本周波数がゆるやかに変動することに期待し、その f_0 として前フレームで推定された基本周波数を使用することとした。

この前置フィルタは、入力を $x(t)$ 、出力を $y(t)$ として、以下の式で表されるものを使用した。

$$y(t) = \frac{x(t) + x(t + 1/f_0)}{2} \quad (2.12)$$

以上の式により、中心周波数 f_0 の楕形フィルタを構成できる。一例として $f_0 = 200\text{Hz}$ の場合のフィルタの振幅特性を図 2.6 に示す。

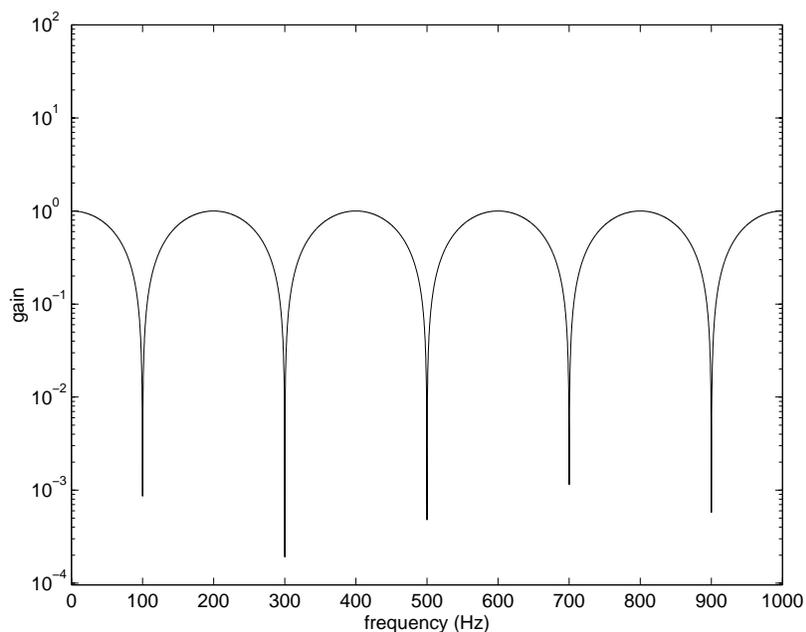


図 2.6: 前置フィルタの特性 ($f_0 = 200$ の場合)

実際のこの前置フィルタによる処理は以下の如くとし、その適用は音声区間に限定した。

基本周波数推定における前置フィルタ適用のアルゴリズム

- ケプストラム法により、基本周波数を推定 (1次推定)。推定ができれば、その推定値をそのまま採用。
- 1次推定値が得られなかった場合で、かつ処理フレームが音声区間ならば、直前フレームの基本周波数推定値を用いて前置フィルタを設定。その前置フィルタを通した波形について、再度ケプストラム法により、基本周波数を推定 (2次推定)。

- 以上で推定出来なかった場合は、その処理フレームは目的音声が存在しない区間と判定し、推定値なしとする。

以上の処理のフローチャートを図 2.7に示す。

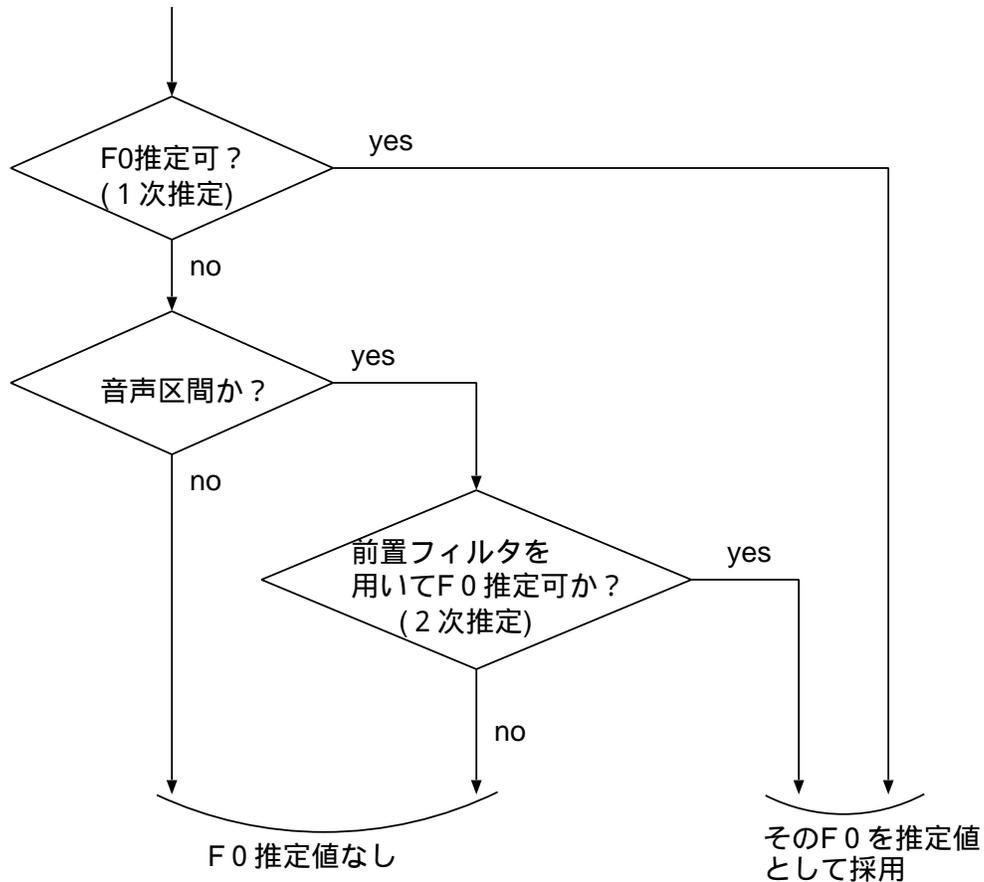


図 2.7: 前置フィルタ処理を適用した基本周波数 (F0) 推定のフローチャート

ここで音声区間の判定は、連続した処理フレームで基本周波数が得られる区間を音声区間とみなし、下記の前方3段、後方4段のフレーム保護を設けて行った。フレーム保護の概要は以下のとおりである。

前方3段

- 非音声区間において、3フレームで連続して1次推定値が得られたならば、そのフレームより音声区間とする。

後方4段

- 音声区間において、1次推定値が得られない区間が4フレーム続いていたなら、そのフレームで音声区間終了とみなす。

以上のフレーム保護に関する状態遷移図を図 2.8に示す。

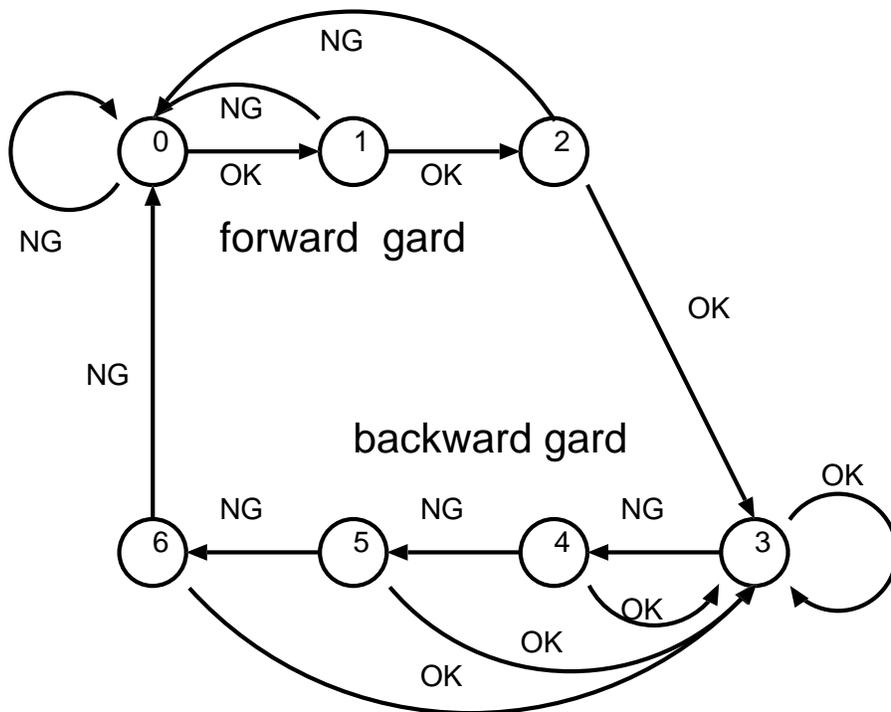


図 2.8: フレーム保護を含んだ音声区間判定の状態遷移図

図 2.8の状態 0 が、非音声区間かつ前方保護の始点である。基本周波数の推定値が得られたならば、OK の矢印へ状態遷移する。基本周波数の推定値が得られないなら、NG の矢印をたどる。状態 0 から、3 フレームで連続して基本周波数が推定できると状態 3 へ至り、この状態 3 で音声区間となる (前方保護)。その状態 3 から 4、5、6、0 への遷移が、後方保護を表している。先の説明と照らすと、現在の処理フレームがこの状態遷移図中の状態 3 ~ 6 に位置付けられるとき、そのフレームを音声区間と判定する。

以上の前置フィルタによる処理を加えた基本周波数推定結果の一例を図 2.9(c) に示す。入力波形は、先に示した図 2.5と同様である。

これをみると、前置フィルタ処理を行わない図 2.9(b) において基本周波数が得られなかった区間でも、基本周波数の推定値が得られており、本手法の有効性が認められる。

以上述べた、ケプストラム法を音声区間に合わせて前置フィルタ処理により強化した基本周波数推定と、それに基づく時間 - 周波数フィルタによる雑音抑圧処理を用いて、次章

に記す音源方向推定アルゴリズムを構築した。

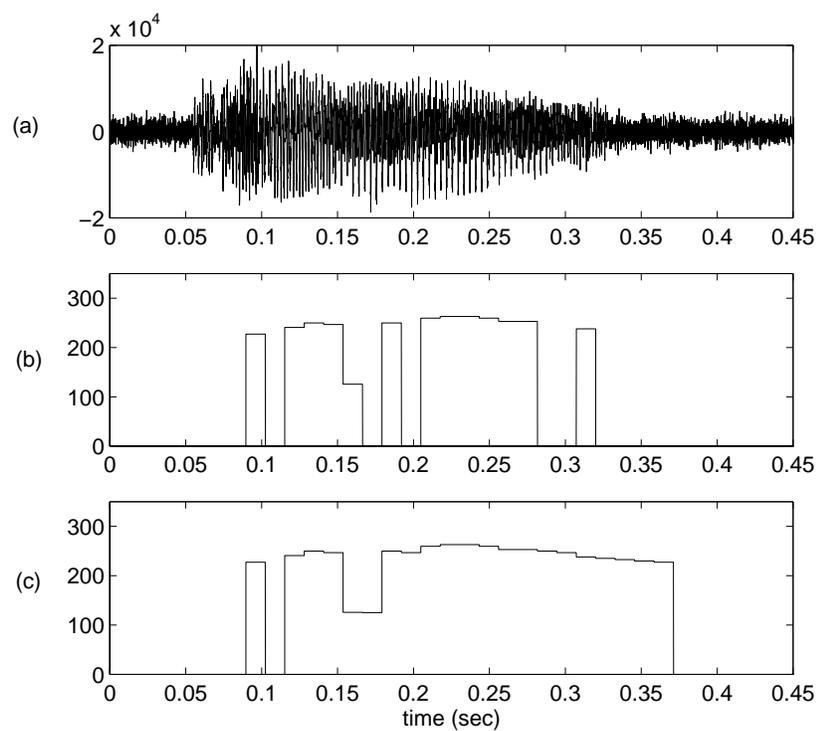


図 2.9: 基本周波数推定結果例 (a) S N R 10dB の白色雑音と反射音が付加された *faf /o/*
(b) ケプストラム法による基本周波数推定結果 (c) 前置フィルタによる強化後の基本周波数推定結果

第 3 章

音源方向推定

マイクロホンレイを用いた音源方向推定は、使用するマイクロホンアレイの形状、寸法により、そしてさらに閉じられた空間内で使用する場合は、その空間のどこにアレイを配置するかにより、その性能は大きく変化する [17]。そこでまずはじめに、本研究で使用するマイクロホンアレイについて説明する。

その次に前章で述べた雑音抑圧処理と西田らの提案法を含んだ音源方向推定のアルゴリズムについて述べる。

3.1 マイクロホンアレイ

3.1.1 マイクロホンアレイの形状

本研究の音源方向推定では、水平面上の全方位 (0° から 360°) に対して方向推定が行うこと前提とする。そのために必要な最少数のマイクロホンを用いたマイクロホンアレイとして、西田らの方向推定法 [25] と同様に、図 3.1 に示すような、一辺の長さ d ($0.3m$) の正三角形の各頂点にマイクロホンを配置したアレイを使用することとした。

このマイクロホンアレイは、3つのマイクロホン対が正三角形の各辺に配置しているととらえることができる。

今、図 3.1 に示すように、マイクロホン m_1 、 m_2 を結ぶ直線の垂直方向で、マイクロホン m_3 と反対方向を基準の方向 (0°) 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。

この時各々のマイクロホン対に対して、第 1 章で記した時間差と到来方向の関係を適用することにより、以下の 3 式が得られる。

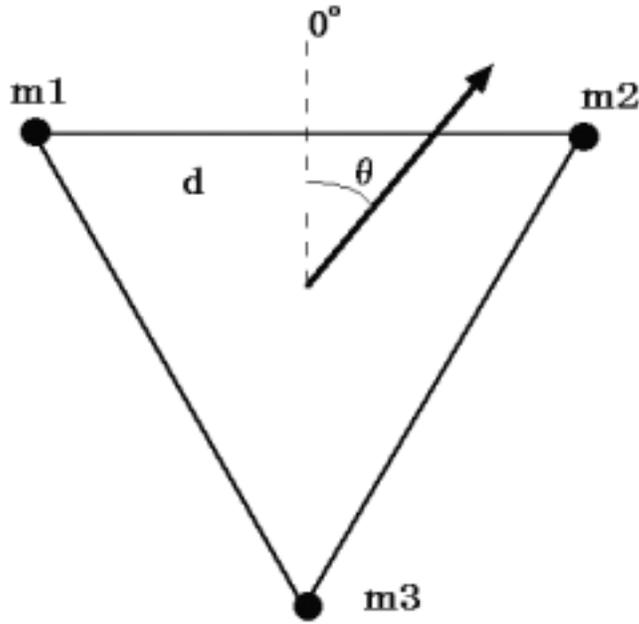


図 3.1: マイクロホンアレイ形状 ($d=0.3m$)

マイクロホン対 m1-m2:

$$\theta = \arcsin(c\tau_{12}/d) \quad (3.1)$$

マイクロホン対 m2-m3:

$$\theta = -\frac{\pi}{3} - \arcsin(c\tau_{23}/d) \quad (3.2)$$

マイクロホン対 m3-m1:

$$\theta = \frac{\pi}{3} - \arcsin(c\tau_{31}/d) \quad (3.3)$$

ここで τ_{12} は m1-m2、 τ_{23} は m2-m3、 τ_{31} は m3-m1 の到来時間差である。音源方向はこれらの時間差をもとに推定される。

3.1.2 マイクロホン対の検出角度の統合

前節の3つの式で留意すべきは、各式はいずれも、一つの時間差に対して、 0° から 360° の範囲で2通りの θ の値が得られる点である。この2通りのうち、どちらか一方が正しい音源方向を代表している。

この正しい音源方向の決定は、各式より得られた各マイクロホン対の θ の値を、統合することにより達せられる。その様子を図 3.2 に示す。図中、real sound source が正しい音源を代表しており、imaginary sound source がそのミラーイメージである。この図を見ると、各マイクロホン対が水平面上の同一の方向を指示している θ が、real sound source

の方向と一致していることがわかる。最終的にこの角度統合処理により、水平面上全方位の方向推定を可能となる。

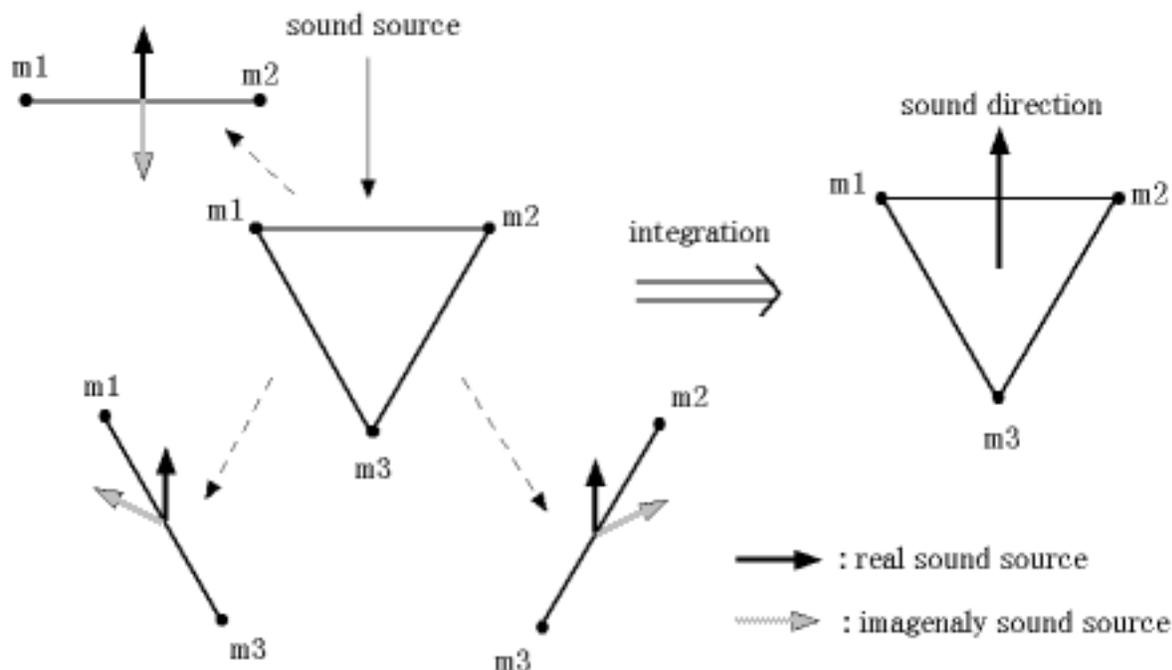


図 3.2: マイクロホン対 3 組の検出角度の統合

またこの統合処理は、任意のマイクロホン対にて、雑音等により誤って検出された結果が生じて、それを除去する付随効果が期待でき、さらに次節で述べる角度量子化幅の偏差を抑える優れた効果を有している。

3.1.3 マイクロホンアレイにおける角度量子化幅

マイクロホンアレイを構成している 1 対のマイクロホン対、例えば m1-m2 に着目すると、各マイクロホン受信信号を周期 T_s でサンプリングして扱う場合、図 3.3 の経路差 $d \sin \theta$ の決定に際し周期 cT_s の空間的なサンプリング効果が生じる (但し c は音速)。

今、経路差を l とおくと

$$l = d \sin \theta \quad (3.4)$$

辺々微分して

$$dl = d \cos \theta d\theta \quad d\theta = \frac{dl}{d \cos \theta} \quad (3.5)$$

ここで dl を空間サンプリング周期 cT_s とし、その時の $d\theta$ を $\Delta\theta$ とすれば、

$$\Delta\theta = \frac{cT_s}{d \cos \theta} \quad (3.6)$$

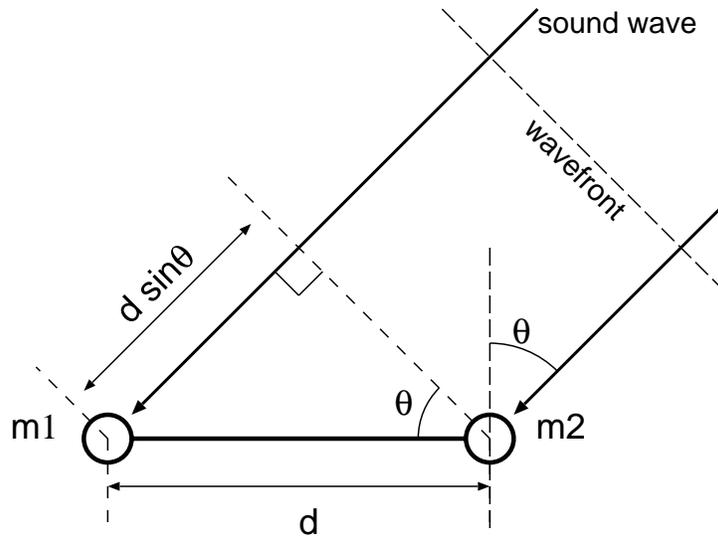


図 3.3: マイクロホン対 m1-m2、音源方向

この $\Delta\theta$ は、空間サンプリングより派生する角度 θ に関する量子化幅とみることができ、図 3.3 のマイクロホン対における方向推定精度の方式上の限界を与えている。さらに式 3.6 より、その推定精度の限界値は、角度 θ 自身の値に依っていることがわかる。

実際に本研究の音源方向推定システムで使用するパラメタを、上式に代入して得られる角度量子化幅 $\Delta\theta$ の角度依存性を図 3.4 に示す ($d = 0.3\text{m}$, $c = 340\text{m/s}$, $T_s = 1/20\text{k} = 50\mu\text{s}$)。

$\Delta\theta$ は $\theta = 0^\circ, 180^\circ$ で最小となり (このとき $\Delta\theta = cT_s/d$)、 $\theta = 90^\circ, 270^\circ$ では $\Delta\theta$ は無限大となる。すなわち $\theta = 90^\circ$ または 270° に近づくにつれ、角度 θ に対するサンプリングが、極端に荒くなることを示している。

さらにマイクロホンアレイを構成するすべてのマイクロホン対で、同様の解析を行った結果を図 3.5(上) に示す。この図を見ると、各マイクロホン対の特性が、互いに 60° づつシフトしているものであることがわかる。

ここで前節で述べた角度統合処理を施すことにより、マイクロホンアレイ全体で図 3.5(下) に示す特性が得られる。図 3.5(下) を見ると、全方位を通して角度量子化誤差は 4° 以内であり、かつその偏差は 1° 以内となることがわかる。

本マイクロホンアレイは、水平面上の全方位 (0° から 360°) に対して方向推定が行うために必要な最少数のマイクロホンを用いて構成されておりながら、その中の各マイクロホン対は、互いに角度の量子化が荒くなる部分を極めて効果的に補償しあい、アレイ全体としてその偏差を 1° 以内に抑えていることは注目に値する。

以上の検討により、本マイクロホンアレイを使用した方向推定結果には、全方位に対して原理的に $\pm 4^\circ$ 以下の角度量子化誤差を含むことが明らかとなった。

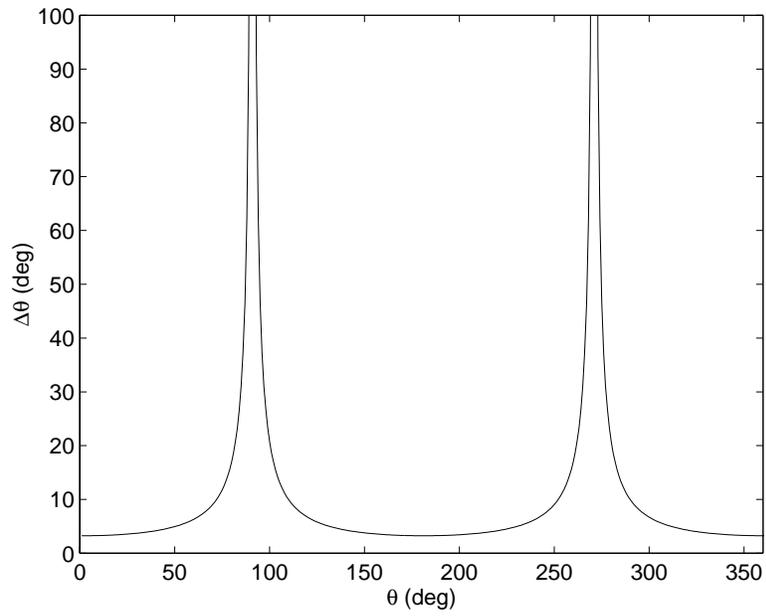


図 3.4: 角度量子化幅 $\Delta\theta$ の角度依存性 (マイクロホン対 : m1-m2)

そこで後の章の方向推定シミュレーションにおいては、本検討結果に 1° のマージンを付加し、方向推定結果の正常、異常の判別閾値として、 $\pm 5^\circ$ を採用することとする。

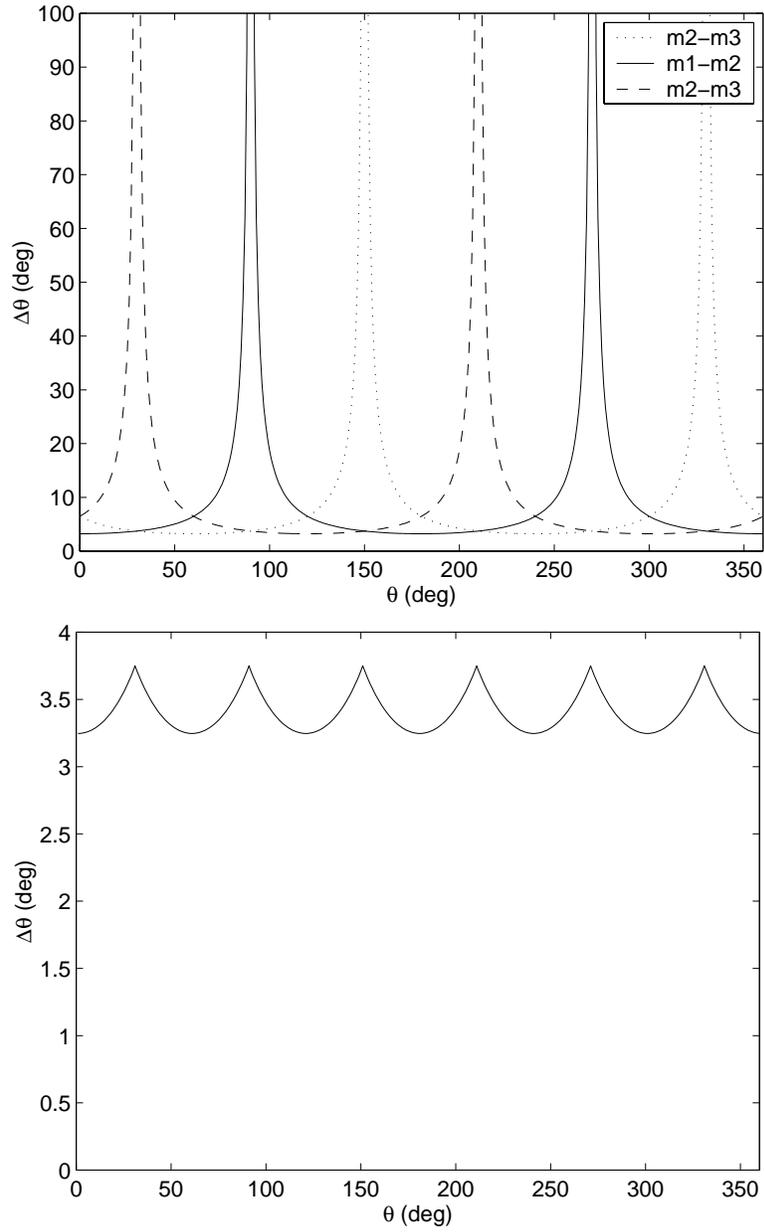


図 3.5: 正三角形配置のマイクロホン対の角度量子化幅 $\Delta\theta$ (上) 各マイクロホン対の $\Delta\theta$ (下) マイクロホン 3 対の統合結果

3.2 音源方向推定のアルゴリズム

本節では、本研究で構築した音源方向推定のアルゴリズムについて、表 3.1 の条件で行った計算機シミュレーションの動作例を示しながら、図 3.6 の流れ図に沿って説明する。基本的に図 3.6 の立ち上がり検出以降の処理は、その反射音に対する不応期制御に期待し、西田らの推定法を流用している。

表 3.1: 動作例に用いたシミュレーション条件

話者方向	180 °
使用音声	ATR 音声データベース 男性話者単語 /a i ma i/
雑音	白色雑音 (S N R 15dB) 到来方向 : 60 °
第一反射音	直接音とのパワー比 : -4dB 到来方向 : 300 ° 直接音からの遅れ時間 : 2msec
第二反射音	直接音とのパワー比 : -8dB 到来方向 : 270 ° 直接音からの遅れ時間 : 15msec
サンプリング周波数	20kHz

3.2.1 雑音抑圧処理

まず雑音に対応するため、各マイクロホンで受信した信号に対し、前章で記した目的音声の基本周期 T に基づく時間 - 周波数フィルタによる雑音抑圧を行う。この雑音抑圧処理の結果の一例を図 3.7 に示す。

3.2.2 立ち上がり検出

立ち上がり検出では、反射音に対応するため、人間の音源定位に有効な先行音効果を工学的にモデル化し以下のように実装される。

ここでは先ず雑音抑圧波形を全波整流した後、ピーク抽出を行う (図 3.8(a))。この処理により、なだらかに変化する波形のピークのみを残すことで、時間差検出に不必要な情報を取り除く。

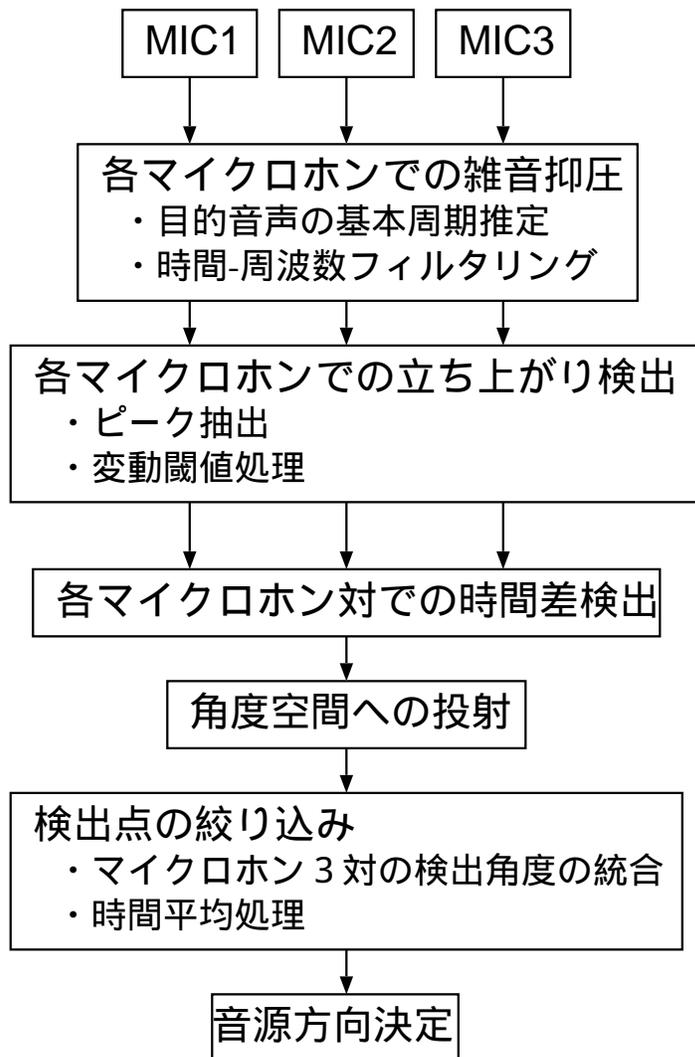


図 3.6: 音源方向推定流れ図

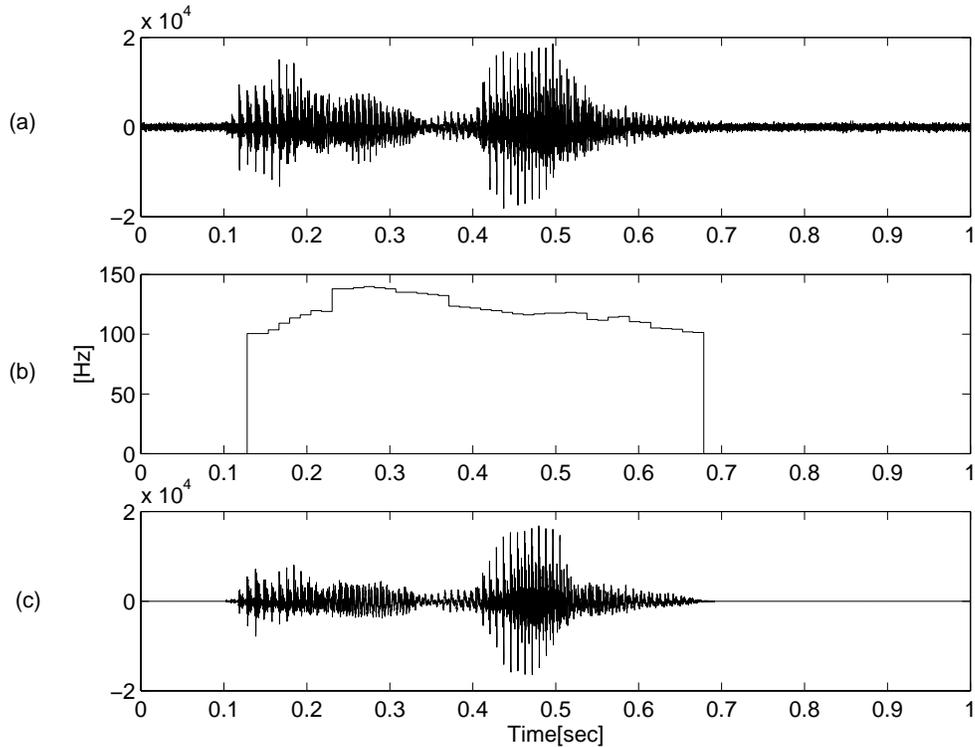


図 3.7: 雑音抑圧処理 (a) マイクロホン入力波形 (b) 基本周波数推定結果 ($1/T$) (c) 雑音抑圧波形

次にピーク抽出波形に対して、波形に適応する形で変動閾値を設定し、振幅が急激に変化するポイントの検出を行う。この変動閾値は、信号の振幅が閾値を越えた時点を音の立ち上がり点（急激な振幅の変動が生じた点）として捉え、その時点を始点として、以下のように動作する。

- 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

実際の入力信号に対する変動閾値の例を図 3.8(a)、その変動閾値による立ち上がり検出点を図 3.8(b) に示す。これらを見ると、変動閾値は、立ち上がり検出点毎にそのレベル

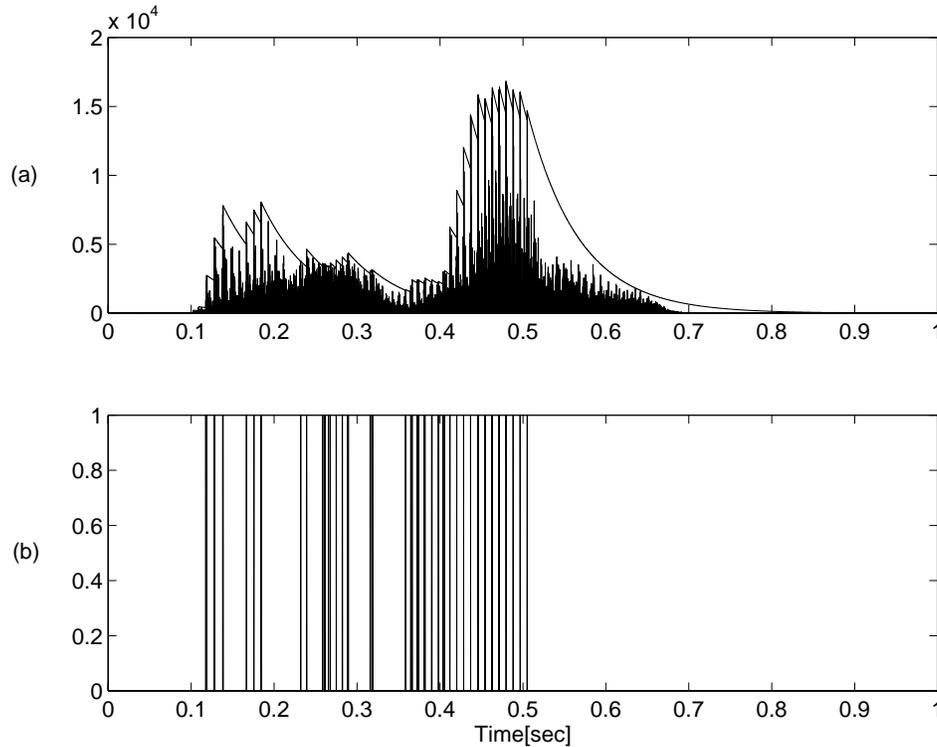


図 3.8: 変動閾値による立ち上がり検出 (a) ピーク抽出波形と変動閾値 (b) 立ち上がり検出結果

が再設定され、そして次の立ち上がり検出点まで、閾値が指数的に低下していく様子が分かる。

一般に反射音は直接音より遅れて到来し、かつその振幅は直接音より小さいという反射音の性質を考慮すれば、このピーク抽出波形に対する変動閾値処理により、反射音に対する不応期を実現しているとみることができる。

3.2.3 時間差検出と角度空間への投射

各マイクロホンの立ち上がり検出結果から、各々のマイクロホン対で時間差を検出する。さらにその結果を、事前に作成したテーブルをもとに角度空間へ投射する。その結果として各マイクロホン対に対応した 3 組の角度検出結果を得る。

3.2.4 検出点の絞り込み

検出角度の統合

以上のようにして求められた各マイクロホン対の角度検出結果には、本章の前半で述べたように、1つの検出時間差に対し、2つの角度検出点が生じている。また雑音による誤った検出点が含まれている可能性がある。

そこで各マイクロホン対の角度検出結果に対して、前に述べた角度統合処理を行い、正しい音源方向に対応する検出点を抽出し、誤った検出点を取り除く。

時間平均処理

さらに方向推定精度の向上を目的とし、目的音の音声の基本周期毎に振幅のピークを持つことに着目し、上述の角度統合結果に対して、一定方向で時間的に連続して生じている検出点を抽出することにより、最終的な目的音声の到来方向を決定する。

以上の処理を施した最終的な音源方向推定結果を図 3.9 に示す。図 3.9 を見ると、表 3.1 で示した雑音と反射音が付加された音声に対して、その立ち上がり部分に同期して正しい音源方向 (180°) のみが検出されており、本手法が有効であることが分かる。

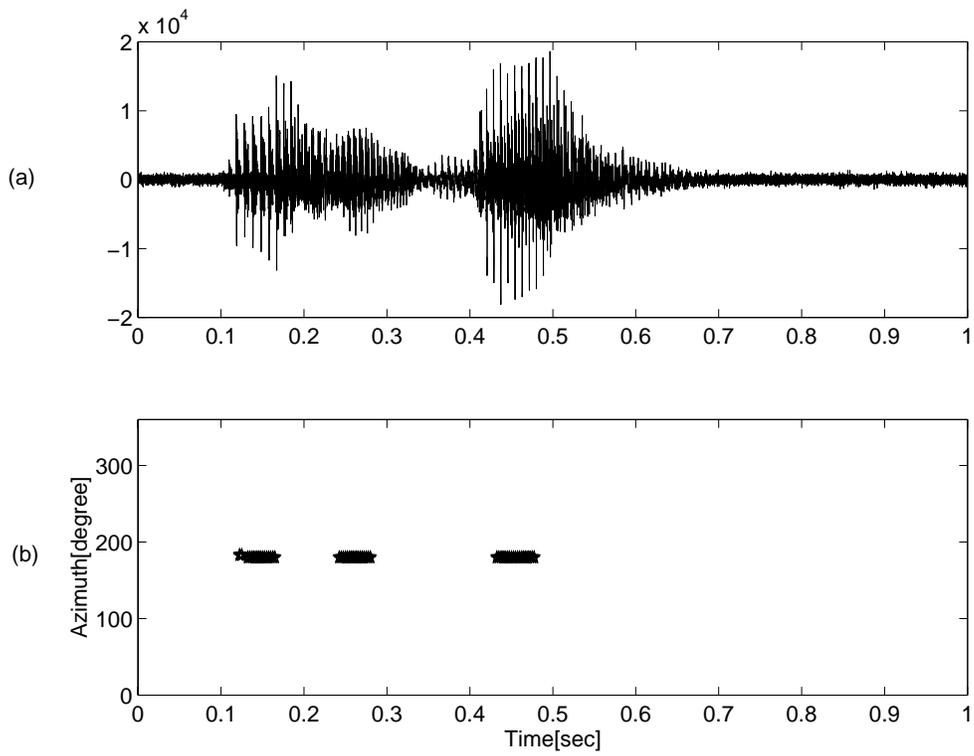


図 3.9: (a) マイクロホン入力波形 (b) 音源方向推定結果

第 4 章

方向推定シミュレーション

本研究で構築した音源方向推定法の性能を調査するため、計算機上で作成した信号を元に単音節母音及び単語音声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。

またその中で、西田らの推定法、及び現行で最も一般的な方向推定法と考えられる白色化相互相関 (別名 Cross-Power Spectrum Phase: CPSP) による推定法 [16] [17][18][5][1][24] との比較実験を行った結果についてもあわせて報告する。

4.1 単音節母音に対する音源方向推定結果

最初に本研究の推定法と西田らの推定法に関し、単音節母音を入力信号として音源方向の推定シミュレーションを行い、その性能を比較した。

4.1.1 雑音に方向性がない場合のシミュレーション結果

シミュレーションの条件は、西田らの研究のシミュレーション条件にあわせて、以下の通りとした。

サンプリング周波数	20kHz
使用音声	ATR データベース 単音節母音 (男性話者 3 名、女性話者 4 名) 到来方向：ランダムに設定
雑音	白色雑音 SNR を 30dB から -5dB きざみで 10dB まで変化
第一反射音	直接音とのパワー比：-4dB 到来方向：ランダムに設定 直接音からの遅れ時間：2msec
第二反射音	直接音とのパワー比：-8dB 到来方向：ランダムに設定 直接音からの遅れ時間：10 ~ 15msec

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。

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。なお入力信号を作成する際には、より厳密に時間差を与えるために、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 ヘダウンサンプリングを行っている。

シミュレーションによって得られた方向推定結果について、西田らの研究と同様、正しい音源方向から $\pm 5^\circ$ 以内に入った検出点を正常、それより外れた検出点を誤りと判定することにする。

方向推定結果で得られた全検出点数に対する、誤り検出点数の比の百分率を誤り率と定義し、それを SNR 毎に算出した結果を図 4.1 に、全入力データに対して算出した誤り率を表 4.1 に示す。

表 4.1: 母音に対する音源方向推定結果

	本研究の推定法	西田らの推定法
音源方向推定誤り率 [%]	28.2	39.9

2つの推定法の性能を比較してみると、図 4.1 では、SNR 15dB 以外では、本研究の推定法の方が誤り率が低く、また表 4.1 みると、本研究の推定法の方が西田らの推定法より、10% 程度、誤り率が低いことがわかる。

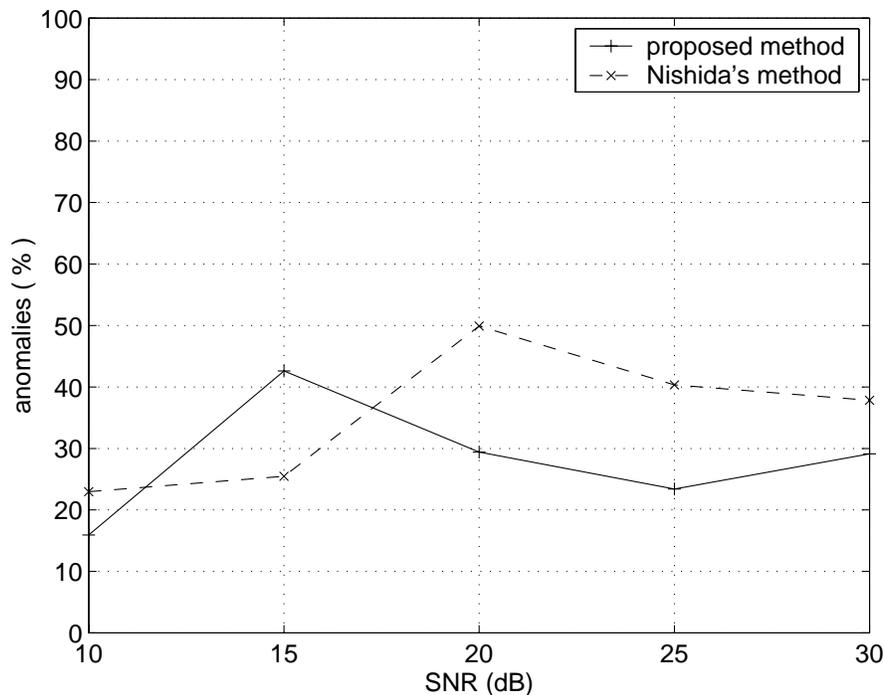


図 4.1: 母音に対する音源方向推定結果

さらに図 4.1をみると、本研究の推定法、西田ら推定法のいずれも SNR が大きいほど誤り率が低減するとは、必ずしもいえない結果となっている。このことは、西田らの研究のシミュレーションにも現れており、原因としては、西田らの推定法では、SNR が小さくなるにつれて、正常な検出点の数が減少するが、推定結果の全検出点数自体も減少するため、SNR が小さくなることで、必ずしも誤り率を大きくするとは限らない結果が得られる。西田らの推定法を流用している本研究の推定法についても、それと同様の傾向が見られる。

4.1.2 方向性雑音を付加した場合のシミュレーション結果

以上のシミュレーションでは、西田らの行ったシミュレーション条件に合わせて、雑音は方向性をもたないものと仮定した。次に上述のシミュレーション条件で、白色雑音を、60°の方向より到来すると仮定して行ったシミュレーション結果について示す。

図 4.2に その SNR 毎の誤り率を、表 4.2に全入力データに対して算出した誤り率を示す。

これらを見ると、方向性のない雑音に対する結果と比べて、本研究の推定法では、誤り率が低減しているのに対し、西田らの推定法では大幅に誤り率が增大していることがわかる。

表 4.2: 母音に対する音源方向推定結果：方向性雑音付加

	本研究の推定法	西田らの推定法
音源方向推定誤り率 [%]	15.4	72.3

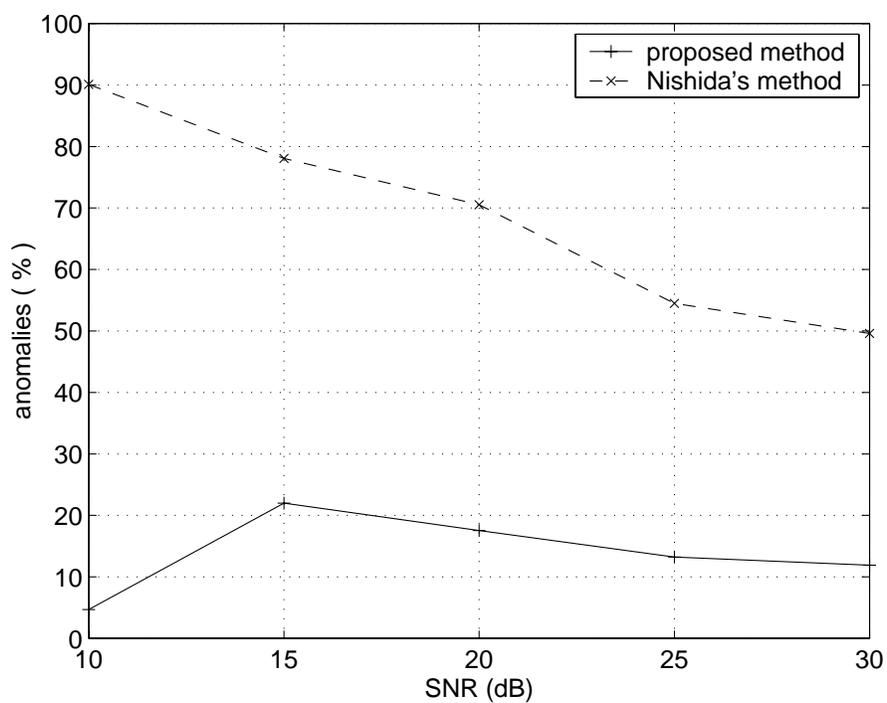


図 4.2: 母音に対する音源方向推定結果：方向性雑音付加

この原因を個別のシミュレーション結果を見た上で考えてみる。一例として、本シミュレーションの中で、男性話者 mau /a/、目的音方向 359 °、雑音レベル SNR 10dB に対する方向推定結果を図 4.3に示す。

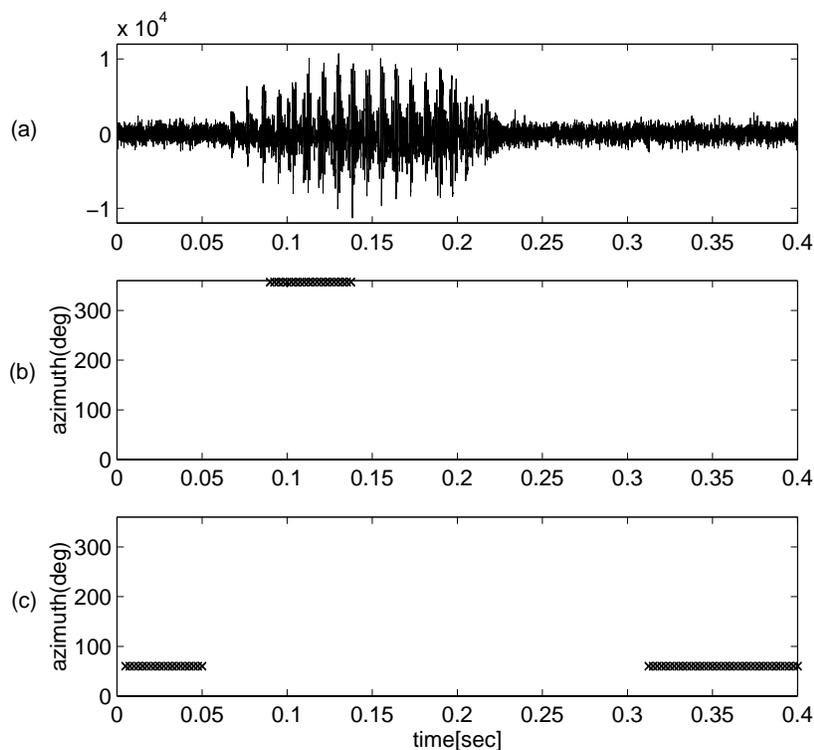


図 4.3: 方向性雑音に関する方向推定結果例 (母音) (a) 入力波形 mau/a/ + 雑音 + 反射音 (b) 本研究の提案法による推定結果 (c) 西田らの推定法による結果

図 4.3を見ると、本研究の提案法では目的音の立ち上がり部分で正しく目的音方向のみが得られているのに対し、西田らの推定法では、目的音到来方向の検出点が全く得られていない。さらに西田らの推定法では目的音が存在しない区間で雑音到来方向 (60 °) の検出点が顕著に現れていることがわかる。この結果は次のように説明できる。

西田らの推定法では、時間差検出までは、時間波形処理で雑音と音声を区分するような処理を行っていないため、雑音による検出点が含まれることが避けられない。その雑音による誤った時間差検出点を、主に角度統合処理により取り除くことで、方向推定システムとして成立しているが、本シミュレーションで付加したような方向性の雑音に対しては、角度統合処理による誤り検出点の除去は全く無効となる。そのため目的音の存在しない区間で雑音方向を検出してしまう。またその後の時間平均処理も本シミュレーション条件では効果が無く、最終方向推定結果に雑音方向が瀕出している。さらに本シミュレーション

においては、目的音の存在する区間でも、目的音の立ち上がり点の情報が雑音により乱されて、目的音の到来方向が得られなかったものと推測される。

一方本研究の提案法では、雑音抑圧処理が有効に作用し、西田らの方法でみられた問題点を回避したため、性能の低下が見られなかった。

次に本シミュレーション結果に対し、母音毎に誤り率を算出した結果を図 4.4に示す。

これをみると、どの母音に対しても、西田らの推定法に比べ本研究の提案法の方が誤り率が低く、図 4.2と同様、SNR が大きいほど本研究の優位性が大きくなる傾向が見られる。

図 4.4を、本研究の提案法の結果にのみ着目してみると、一般に信号振幅が大きいとされる /a/、/o/について誤り率が低く、振幅が小さいとされる /i/、/u/で誤り率が比較的高くなり、このことは、西田らが行った彼らの推定法の実験結果と同様の傾向が見られた。

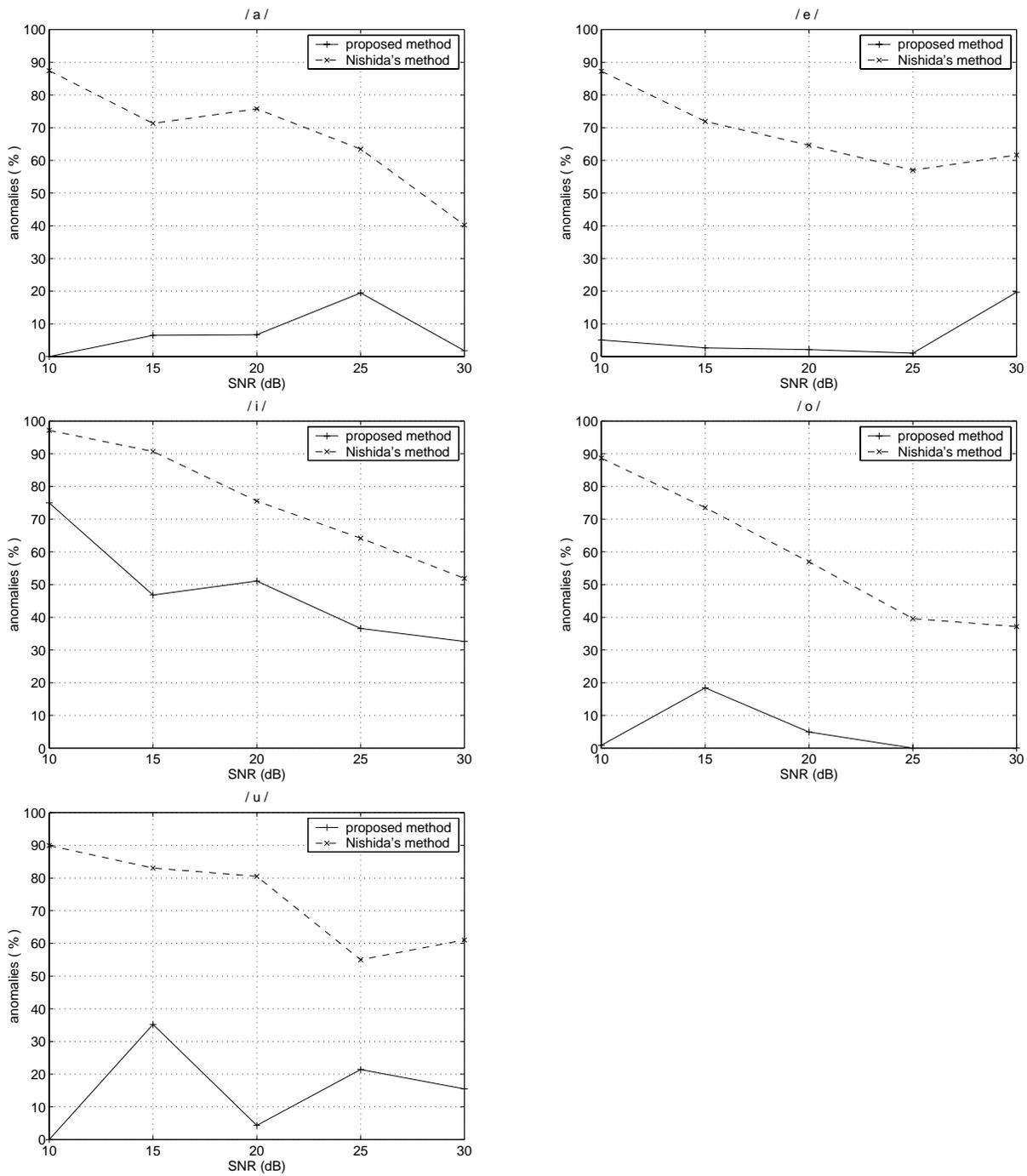


図 4.4: 母音毎の音源方向推定結果：方向性雑音

4.2 単語音声に対する音源方向推定結果

つぎに単語音声に対して行った結果を示す。本節のシミュレーションでは、本研究の提案法、西田らの推定法に加え、白色化相互相関法 (以後 CPSP とする) による推定法について、性能比較を行った。

なお CPSP については、条件をあわせるため、本研究と同様のマイクロホンアレイを使用することを仮定し、かつそれで得られた時間差検出結果に対して、処理フレーム単位で角度統合を行った結果について示している。

4.2.1 雑音に方向性がない場合のシミュレーション結果

はじめに方向性のない雑音のみを付加した単語音声のシミュレーション結果を示す。シミュレーション条件は下記の通りである。

サンプリング周波数	20kHz
使用音声	ATR データベース 単語音声、8 単語、計 18 サンプル (男性話者 4 名、女性話者 3 名) 到来方向：ランダムに設定
雑音	白色雑音 (方向性なし) SNR を 30dB から -5dB きざみで 10dB まで変化

それぞれの推定法について、SNR と誤り率の関係を図 4.5 に示す。これをみると、本研究の推定法では、SNR 10dB で誤り率が 23% であり、SNR が大きくなるに従い、誤り率が低減していく傾向が見られる。一方 CPSP の結果に着目すると、SNR の値によらず、極めて低い誤り率を維持している。この誤り率に関する限り、本条件では CPSP の性能が最も優れる。

さらに本シミュレーション結果に対し、各推定法ごとに、全 18 サンプル中、正しい音声方向が得られたサンプル数を調査した。その結果を図 4.6 に示す。図 4.6 では、1 ポイントもしくは 1 フレームでも正常な検出点を得られれば、そのサンプルをカウントした。

これを見ると本研究の提案法では、全サンプルに対して、全 SNR で正常検出点を得られていることがわかる。これに対し CPSP では、SNR が低くなるにつれ、正常検出点を得られたサンプル数が極端に低下している。特に SNR 10dB では、半数以下の 7 サンプルでしか正しい推定結果が得られていない。それにもかかわらず、SNR が低い条件でも CPSP の誤り率が低いことは、正しい音源方向が多く得られているというより、むしろ先に記した角度統合処理の補助により、誤り検出点が除かれたことによる受動的な要因が大

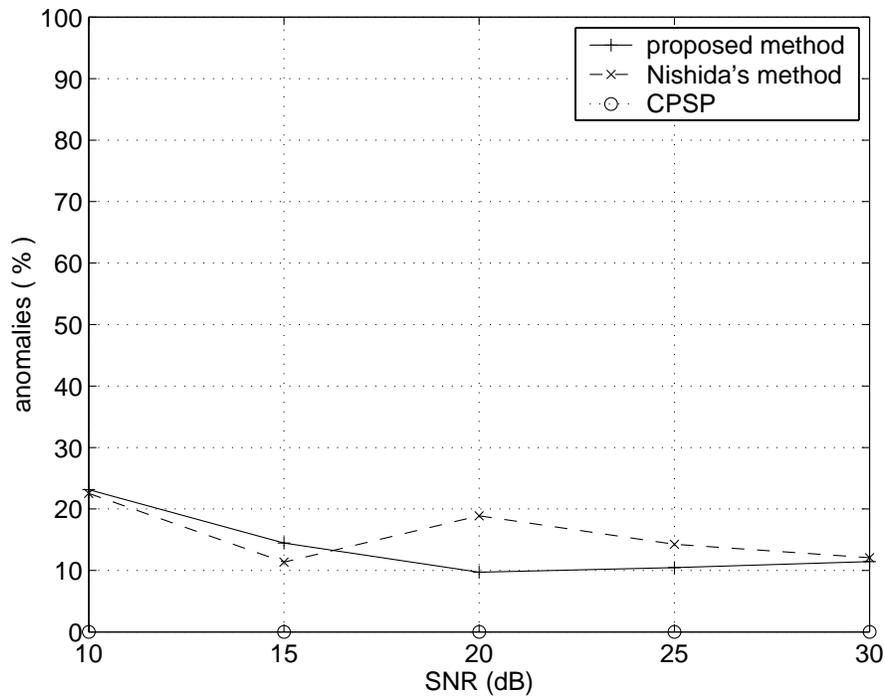


図 4.5: 単語音声に対する音源方向推定結果：雑音方向性無、反射音なし

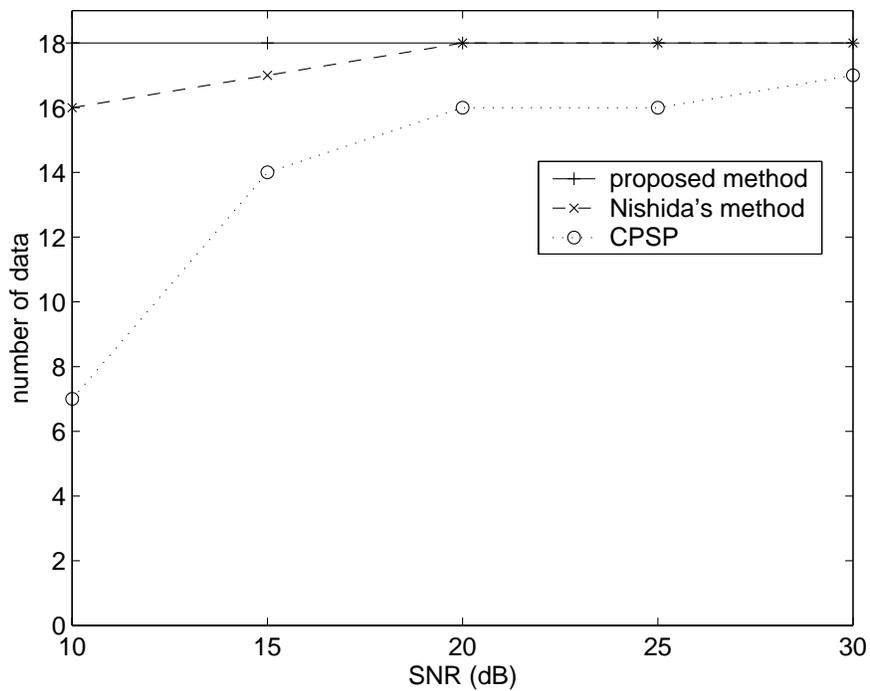


図 4.6: 正しい音声方向が得られたサンプル数：雑音方向性無、反射音なし

きいものと推測される。

次に本条件に反射音を加えた方向推定結果について示す。シミュレーション条件は、以下の通りとした。

サンプリング周波数	20kHz
使用音声	ATR データベース 単語音声、8 単語、計 17 サンプル (男性話者 4 名、女性話者 3 名) 到来方向：ランダムに設定
雑音	白色雑音 (方向性なし) SNR を 30dB から -5dB きざみで 10dB まで変化
第一反射音	直接音とのパワー比：-4dB 到来方向：ランダムに設定 直接音からの遅れ時間：2msec
第二反射音	直接音とのパワー比：-8dB 到来方向：ランダムに設定 直接音からの遅れ時間：10 ~ 15msec

以上の条件で行ったシミュレーション結果に関し、図 4.7に誤り率、図 4.8に正常検出点
が得られたサンプル数を示す。

この結果を先のシミュレーション結果と比較してみると、誤り率は、本研究の提案法と
西田らの推定法で、低い SNR において増加しているが、CPSP ではほとんど変化がみら
れない。正常検出点を得られたサンプル数については、各提案法ともに、反射音の影響を
受け減少している。

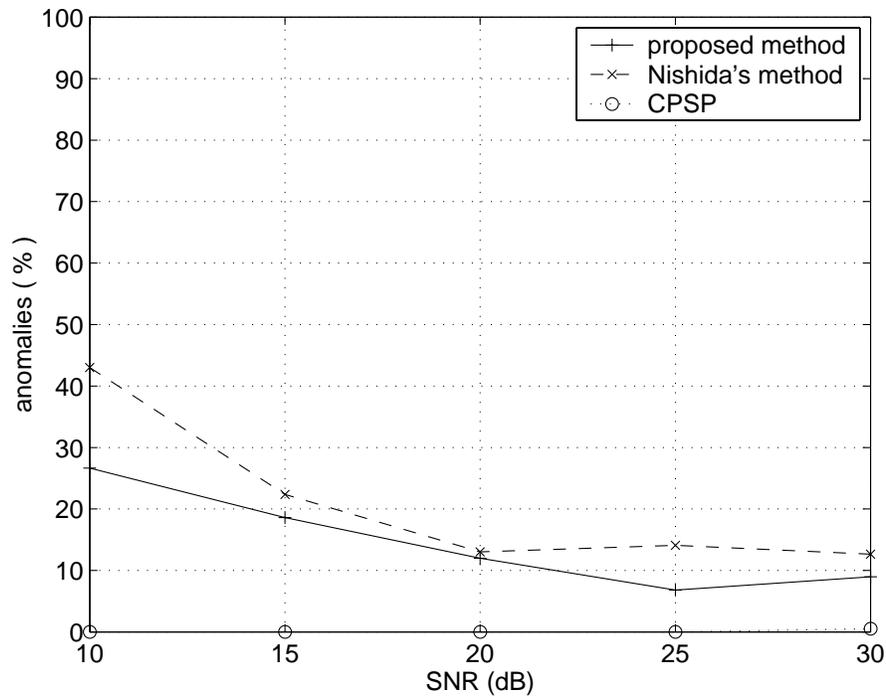


図 4.7: 単語音声に対する音源方向推定結果：雑音方向性無、反射音あり

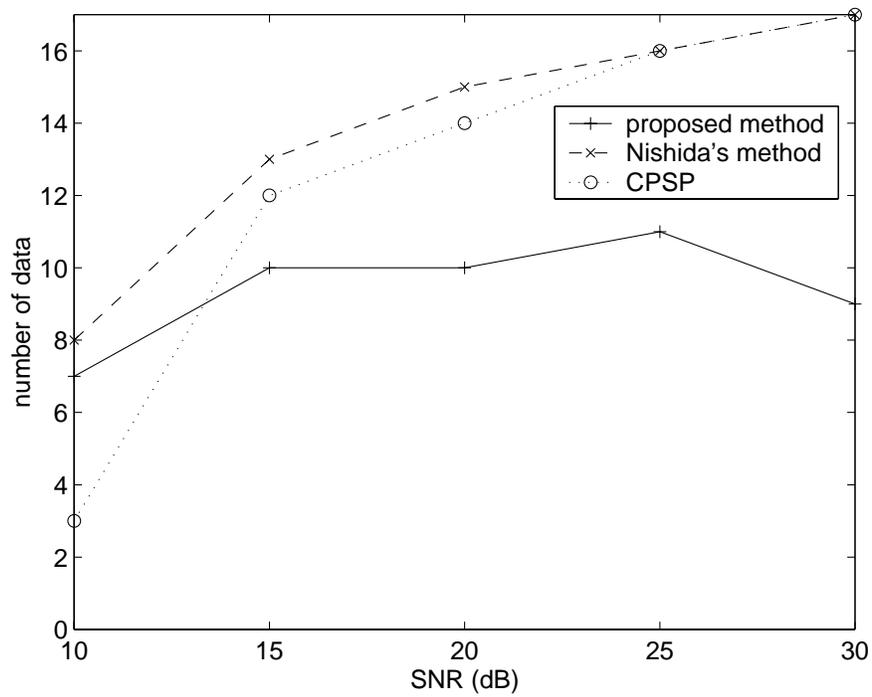


図 4.8: 正しい音声方向が得られたサンプル数：雑音方向性無、反射音あり

4.2.2 方向性雑音を付加した場合のシミュレーション結果

さらにこのシミュレーション条件で、雑音を、 60° の方向より到来する白色雑音に変えた場合のシミュレーション結果を、図 4.9、図 4.10 に示す。雑音条件以外は、先のシミュレーション条件と全く同様である。図 4.9の誤り率の結果をみると、西田らの提案法と CPSP の特性が大きく劣化しており、方向性雑音の付加に対しては、本研究の提案法がもっとも優位であることが示されている。

この誤り率に関する方向推定法間の差異について、個別サンプルのシミュレーション結果を調査することで考察を試みる。一例として、本シミュレーションの中の、女性話者 fks /hougaku/、目的音方向 238° 、雑音レベル SNR 10dB に対する方向推定結果を図 4.11に示す。

これをみると、本研究の提案法では正しく目的音方向のみが得られているのに対し、西田らの推定法では、前節の方向性雑音による結果と同様に、目的音が存在しない区間で雑音到来方向 (60°) の検出点が顕著に現れている。CPSP についてしてみると、目的音の存在しない区間のみならず、目的音の区間においても雑音方向が推定されている。本図を見る限り、CPSP がもっとも方向性雑音の影響を被っており、そして実際に図 4.9において、SNR が 20dB 以下では、CPSP の誤り率ももっとも大きい。

振り返って、図 4.6、図 4.8を見ると、先に触れたように、SNR が 10dB において正常検出点が得られるサンプル数は、他の推定法と比較して CPSP は極端に少なく、本シミュレーションではその点を方向性雑音につかれた形で、図 4.11のような結果が生じたものと考えられる。

この図 4.11と、図 4.9の誤り率の結果は、本研究の特色である、目的音声の調波に合わせた雑音抑圧処理の有効性を、よく表していると言える。

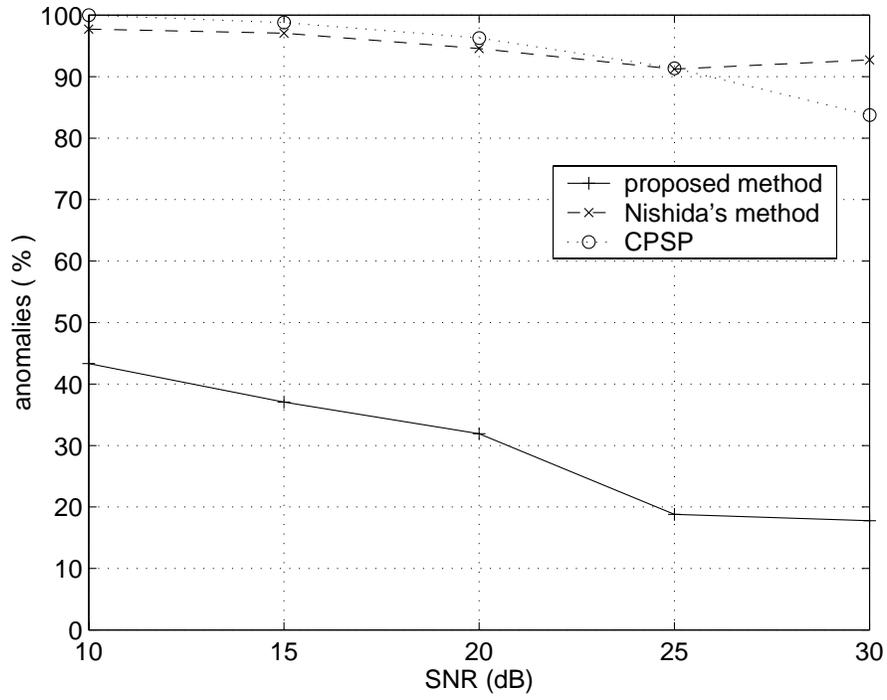


図 4.9: 単語音声に対する音源方向推定結果：方向性雑音、反射音あり

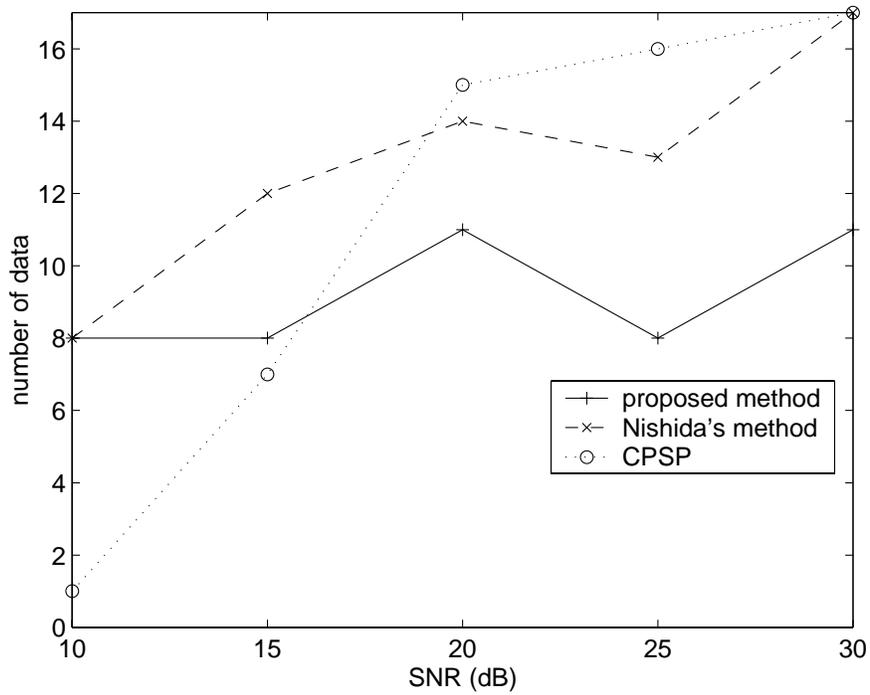


図 4.10: 正しい音声方向が得られたサンプル数：方向性雑音、反射音あり

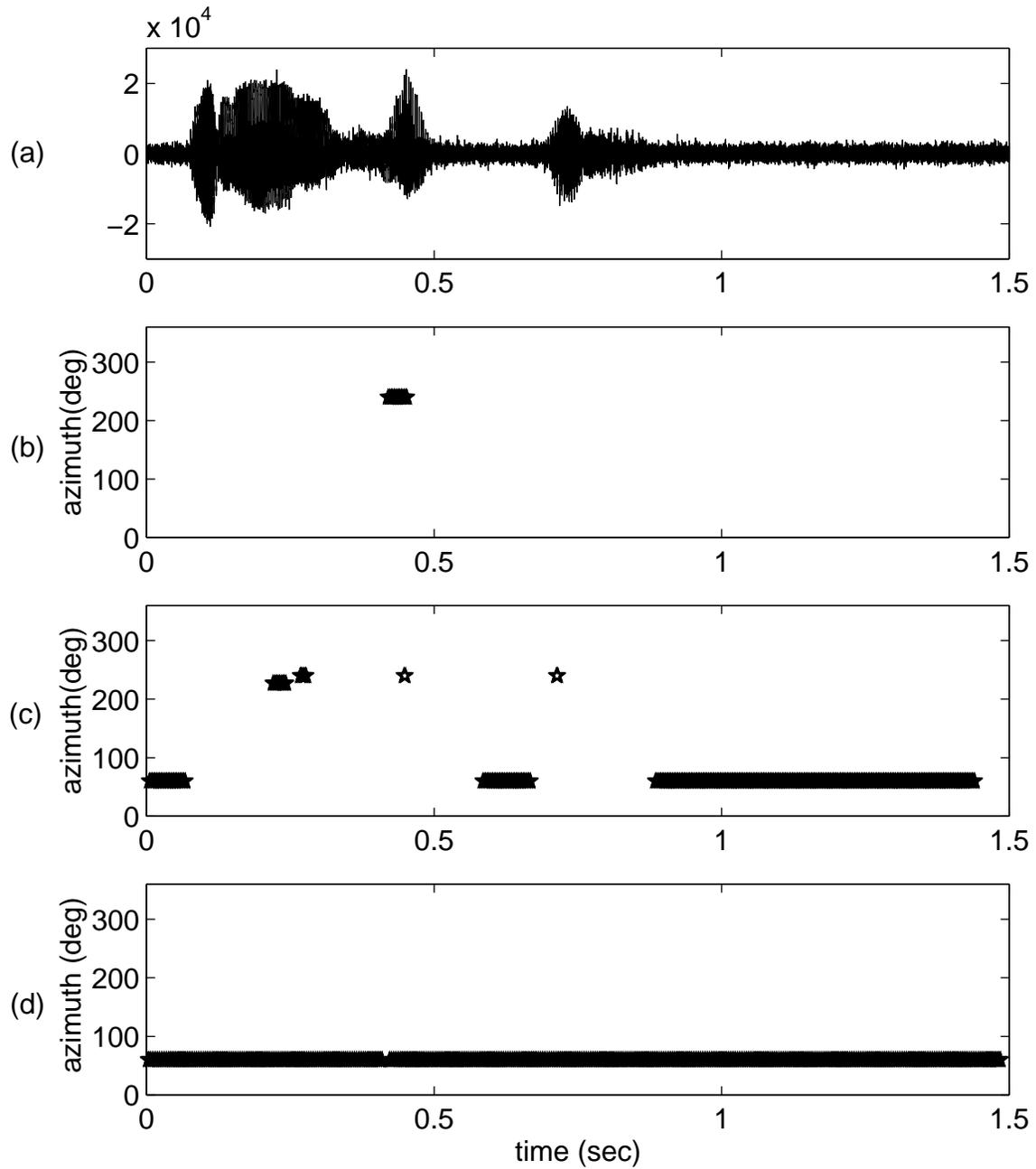


図 4.11: 方向性雑音に関する方向推定結果例 (単語) (a) マイクロホン入力波形 (b) 本研究の提案法 (c) 西田らの推定法 (d) 白色化相互相関法 (CPSP) に基づく推定法

4.3 音場シミュレーションを用いた音源方向推定実験

本節では、室内音場シミュレーションの手法を用いた残響環境を計算機上に構築し、残響時間をパラメタとした性能評価シミュレーションを行った結果について報告する。

4.3.1 イメージ法

本シミュレーションでは、ある一つの室内空間を仮定する。この空間全体の反射音の構造は、以下のような仮想音源分布で記述するイメージ法を用いることで、理解できる。例として図 4.12 に直方体の部屋の 2 つの壁面に対する反射の伝搬の様子を示す。これらの反射波は、壁を取り除いた状態での仮想音源からの直接波として置き換えられることを示している。これにより、反射音の到来方向を知ることができ、さらに仮想音源からの直接波に対して、壁の反射係数を、透過係数に置き換えることで、マイクロホンで受信する反射音の音圧を見積もることができる。またさらに、このような仮想音源が存在する鏡像の部屋を、実体の部屋の四方八方に幾重にも配置することにより、反射音が多数存在する残響環境を仮定することが可能となる。

4.3.2 シミュレーションに使用した室内モデル

以上のイメージ法を用いて、例えば先の直方体の部屋のインパルス応答をシミュレートすることが出来、計算機上でこのインパルス応答を音声と畳み込むことにより、残響を被った音声信号を作成することができる。本実験では、図 4.13 に示すような、寸法 $11 \times 6 \times 2.75\text{m}$ の直方体の室内を仮定し、イメージ法に基づく J.B.Allen らの提案法 [14] を用いて、そのインパルス応答を作成した。音源とマイクロホンの配置は、この部屋の左下隅を原点とし、そして長辺方向を x 軸、短辺方向を y 軸、高さ方向を z 軸にとり、以下の通りとした (単位は m)。

	x 座標	y 座標	z 座標
音源	10.2	0.96	1.50
マイクロホン m1	2.35	3.12	1.50
マイクロホン m2	2.65	3.12	1.50
マイクロホン m3	2.50	2.86	1.50

以上の配置により、マイクロホンアレイに対する音源方向は 105° となる。そしてこの部屋の 6 面の壁の反射係数を変化させることで、残響時間 50ms から 700ms までの範囲

に対応したインパルス応答を、残響時間毎、各マイクロホン毎に複数作成する。この際、仮想室内の残響時間と壁面の反射係数の関連付けに、Eyring の定式 [35] を使用した。

4.3.3 シミュレーション条件

本研究の提案法、西田らの推定法、CPSP に関して、上述により作成した室内インパルス応答と、音源の音声波形を畳み込むことにより、残響を被った音声波形をマイクロホン毎に作成し、かつ以下の条件で方向シミュレーションを行った結果について示す。

サンプリング周波数	20kHz
使用音声	ATR データベース 単語音声 計 18 サンプル (男性話者 4 名、女性話者 3 名) 到来方向 : 105 °
雑音	白色雑音 SNR 10dB, 20dB, 30dB
残響時間	50ms から 700ms まで変化

本シミュレーション結果に関し、雑音レベルの SNR 毎に、残響時間に対する誤り率と、正しい音声方向が得られたサンプル数の算出した。その結果を

SNR 10dB に関し、図 4.14、図 4.15 に、

SNR 20dB に関し、図 4.16、図 4.17 に、

SNR 30dB に関し、図 4.18、図 4.19 にそれぞれ示す。

これらの図をみると、雑音レベルが SNR 10dB、残響時間が 70ms 以下で、提案法の誤り率が、他の手法より低いことが確認できる。

CPSP は通常残響時間が 200 ~ 300ms 以上で、その性能が劣化することが知られており、本シミュレーション結果の誤り率をみても、そのことが確認できる。本研究の提案法は、耐残響性能において CPSP を上回ることを期待したが、結果として、どの SNR に対しても、200ms 以上の残響時間ではその有効性が見出せない。

提案法と西田らの推定法の誤り率が、数 100ms の残響時間に対し安定せずに大きく振れていることは、この範囲の残響時間で正しい方向が得られるサンプル数が大きく落ち込んでいることを考慮すると、おそらく誤り率を計算する検出点数が少なくなるため、1ポイントの検出点増減が誤り率を大きく変化させてしまう事が原因と推測される。このことを考えると、本提案法が、数 100ms の残響時間で部分的に誤り率が低い箇所があったとしても、提案法が有効であるとは必ずしも言えないだろう。

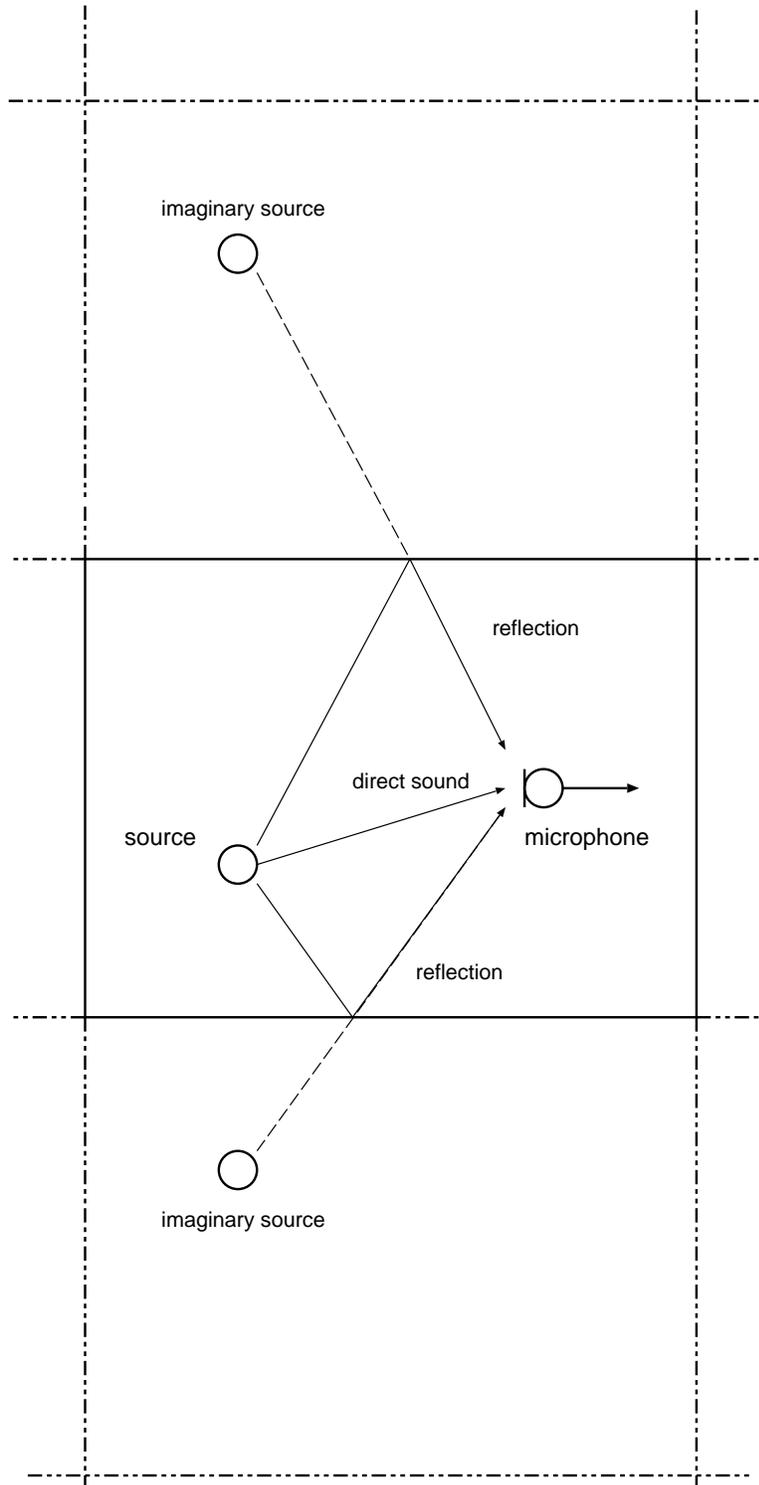


図 4.12: 音の反射と仮想音源

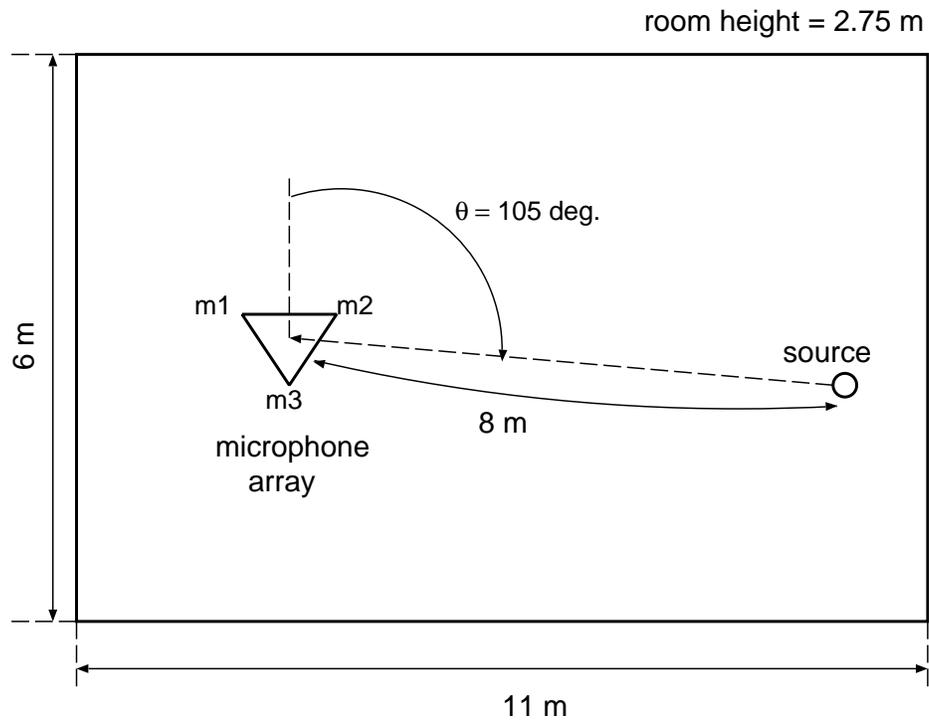


図 4.13: シミュレーションで用いた室内のモデル

正しい音声方向が得られたサンプル数を比較すると、数 100ms の残響時間では、全ての SNR において、CPSP が他の 2 つに比べてかなり多く、この点でも本研究の提案法の優位性は認められなかった。

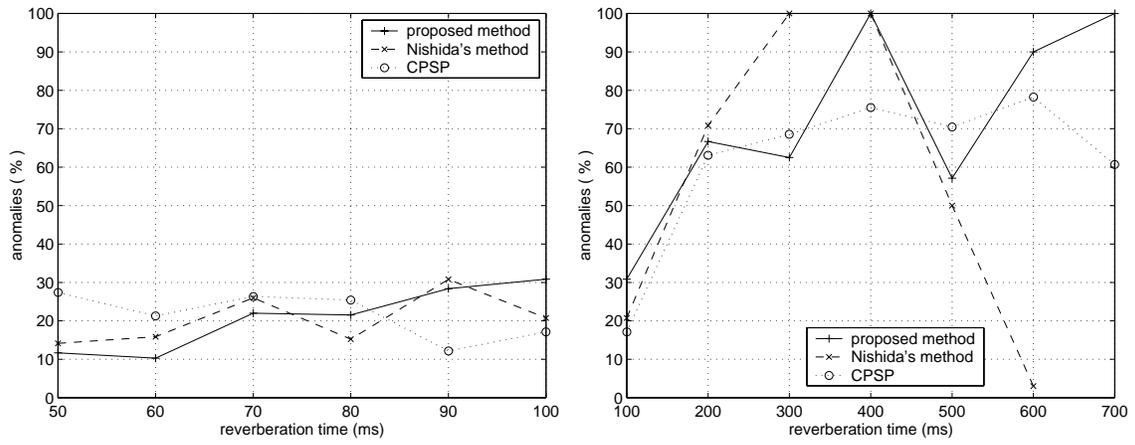


図 4.14: 残響環境の音源方向推定結果 : SNR 10dB

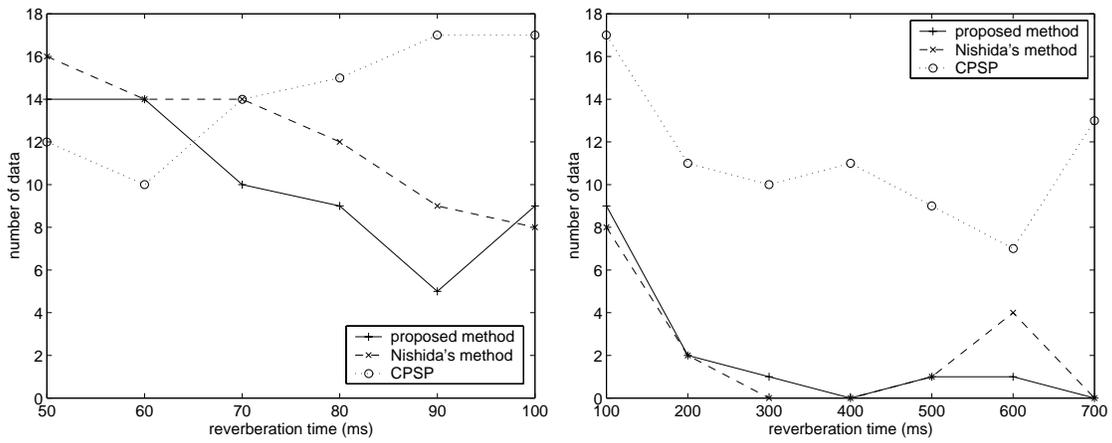


図 4.15: 正しい音声方向が得られたサンプル数 : SNR 10dB

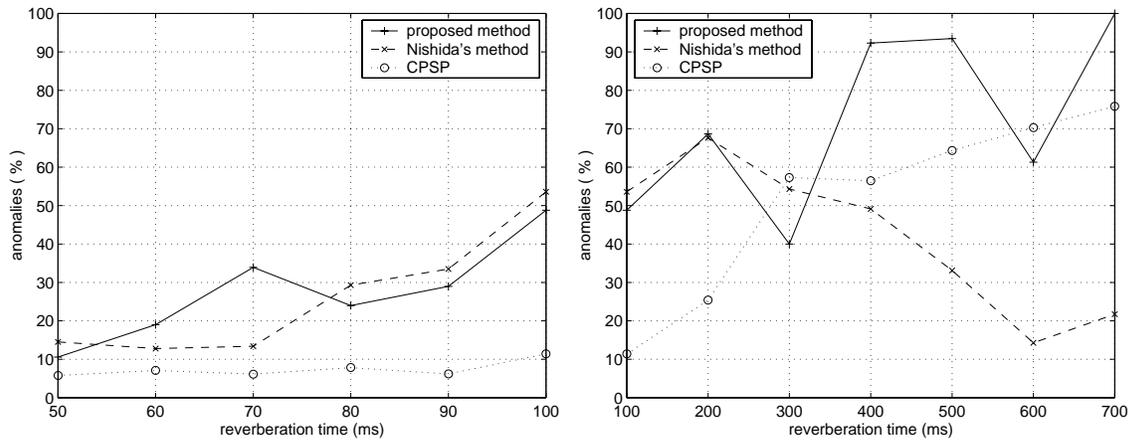


図 4.16: 残響環境の音源方向推定結果 : SNR 20dB

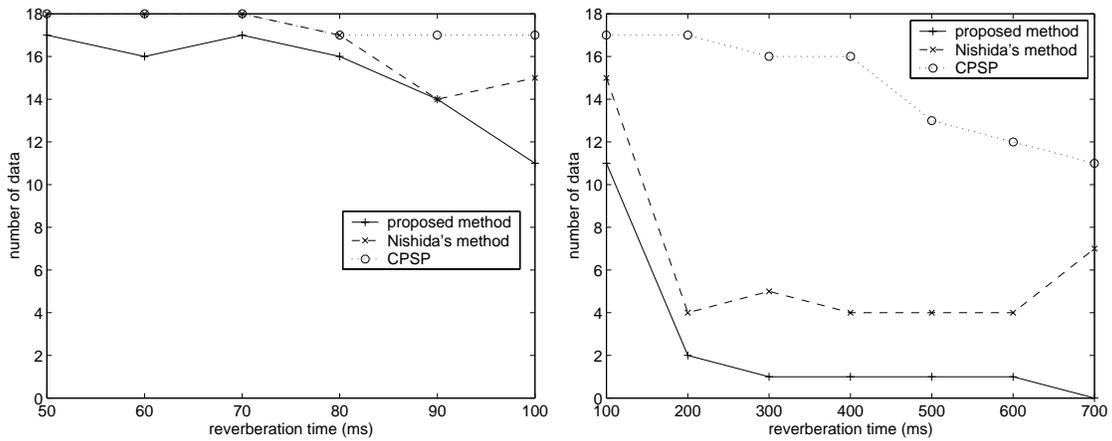


図 4.17: 正しい音声方向が得られたサンプル数 : SNR 20dB

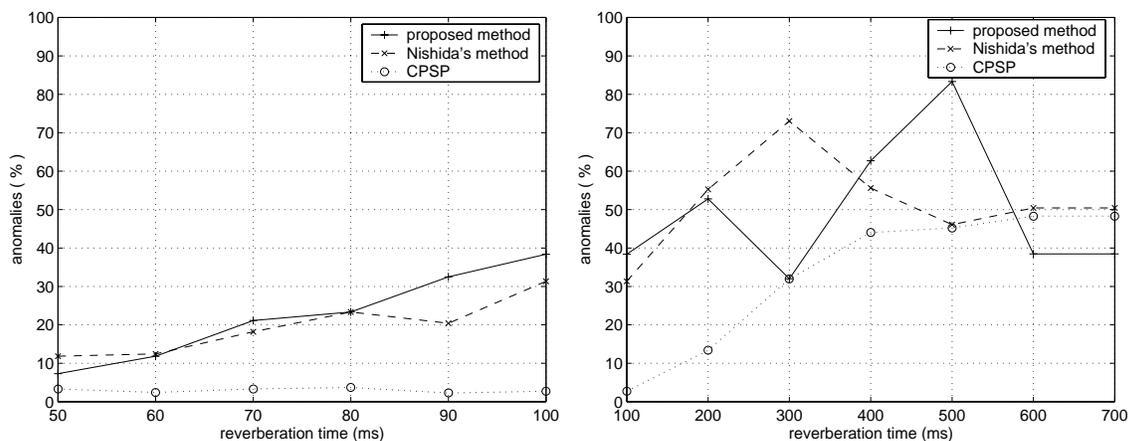


図 4.18: 残響環境の音源方向推定結果 : SNR 30dB

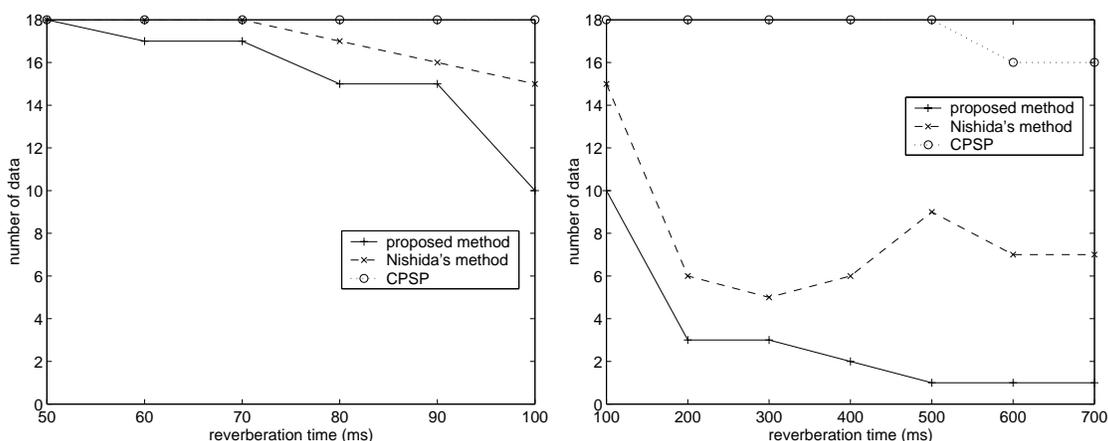


図 4.19: 正しい音声方向が得られたサンプル数 : SNR 30dB

4.3.4 残響環境下で性能が低下した原因

本シミュレーション結果において、数 100ms の残響時間で、CPSP と同様、もしくはそれ以上に、本研究の推定法と西田らの推定法の性能が低下した原因について考察する。本研究の反射音、残響対策は、西田らの手法を流用した反射音に対する不応期制御にあった。この制御の有効性は、第 1 章で述べたように以下の仮定に基づいている。

- 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

この不応期制御は時間波形に関する処理であるため、実際に本シミュレーションで使用したインパルス応答を用いて、残響波形を確認してみた。図 4.20 は、男性話者 mau /aimai/ に関し、残響を含まないものから、残響時間 600ms までの残響波形を示している。左上が残響が付加されていない元信号である。

これを見ると明らかに、残響時間が長くなるにつれ、付加される残響の量が多くなり、信号振幅が大きくなっていくことがわかる。残響時間 600ms の残響が付加された波形においては、その振幅がピーク・ツー・ピークで、元信号の 4 倍から 5 倍となっている。この事実は著者の予想を越えていた。そしてこのような波形に対して、本研究もしくは西田らの方法で、正確な直接音の立ち上がりを数多く検出できるとは到底思われぬ。

上述の不応期制御を支える仮定のうち、“反射音は直接音に比べてそのパワーが小さくなる”ことは、一つ一つの反射音には当てはまるが、多数の反射音が重畳する残響については成り立たず、そのことが本シミュレーションの数 100ms の残響条件で、西田らの提案法と、それを流用した本研究の提案法の反射音に対する不応期制御、立ち上がり検出を困難とし、性能低下に至ったものと考えられる。

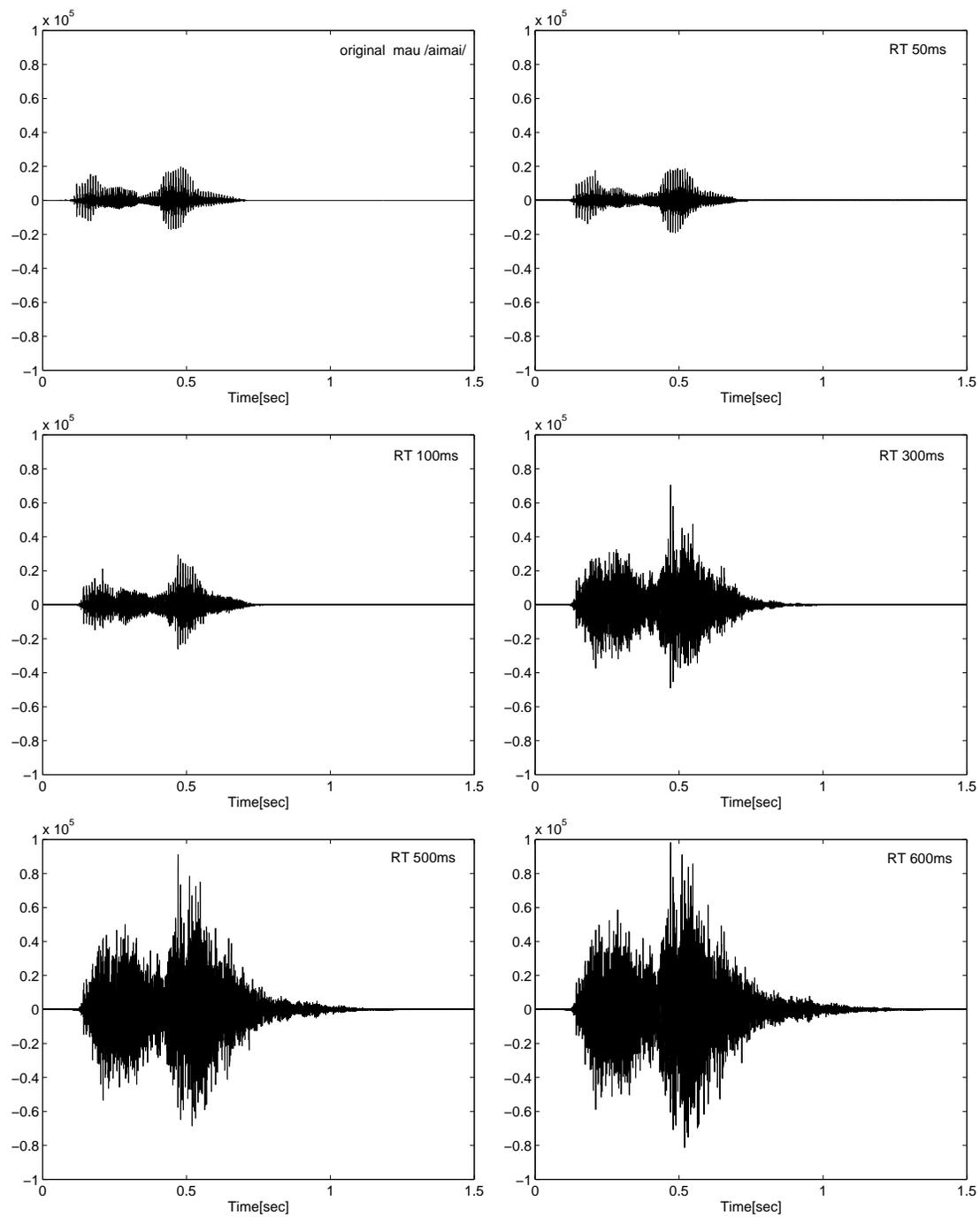


図 4.20: mau /a i ma i/ の元波形と残響が付加された波形

第 5 章

結論

5.1 本研究についてのまとめ

本研究では音声を目的音とする音源方向推定に関し、音声の特徴に利用した雑音抑圧処理を付加した音源方向推定法を提案した。その雑音抑圧に関し、方向推定に欠かせない元の目的音声の持つ時間差情報を保存しつつ実現することに留意し、音声の基本周期を手掛かりとした時間 - 周波数フィルタの適用をはかった。

その中でこの時間 - 周波数フィルタを実現するために必須となる、雑音と反射音が混在する環境下における目的音声の基本周波数推定について検討し、ケプストラム法をベースにし、音声区間の判定を含んだ基本周波数推定法を新たに考案した。実際本研究中、この基本周波数推定に関わる検討が最も大きなウェイトを占め、提案法で確認された方向推定結果の優位性は、この基本周波数推定法の適用に依るところが極めて大きいと考える。

さらにこの雑音抑圧処理と、反射音に対する不応期制御を含む西田らの音源方向推定を組み合わせることで、雑音と反射音に対応した音源方向システムの構築を行った。

シミュレーションによる性能評価では、特に方向性雑音に対し、白色化相互相関 (CPSP) による推定法と西田らの提案法に対して明らかに優位であることが認められ、本研究の雑音抑圧処理の有効性が確認された。

しかし一方、イメージ法を用いた音場シミュレーションによる残響に対する評価において、数 100ms の残響時間に対し性能低下がみられ、そのレベルの残響環境において CPSP に対する優位性は認められなかった。この原因としては、提案法が反射音に対して備えた上述の不応期制御が、反射音と直接音のレベル差に期待した処理であるため、反射音が集合した残響のレベルが直接音のレベルを凌駕する環境では正常に機能せず、方向推定が行えなかったためと考えられる。

5.2 今後の課題

本研究では、雑音と反射音の双方に対応した方向推定アルゴリズムを目指したが、前節で述べたように、複数の反射音の集合である残響に対し、従来法を上回る性能が得られなかった。残響への対応は、西田らの不応期制御のような、時間波形に対するレベル処理だけでは、従来法 (CPSP) の性能を越えることは難しいと思われる。物理的に初期反射と残響は異なる性質をもつ [34] ということであれば、残響に対応するためには、反射音に対するものとは異なる、残響の特徴を利用した残響抑圧処理を検討する必要があると考える。

他の課題として、例えば雑音抑圧の、変動閾値の減衰の時定数等、提案法で使用する各種パラメタの最適化があげられる。本研究では雑音抑圧処理を付加したことによる方向推定結果における有効性を、出来るだけ様々な環境条件を想定して検証することに優先して取り組んだ為、これらパラメタのチューニングは十分に行えなかった。今後、本論文で示したシミュレーション条件で、各種パラメタの設定を変えてその性能を調査して、各パラメタの最適化を図る必要がある。さらにより上位のレベルで言えば、雑音抑圧処理に関し、提案法で用いたものとは異なる他の雑音抑圧方法の適用を検討する余地も十分に残されている。

さらに他の課題として方向推定区間の拡張があげられる。本研究の提案法では音の立ち上がり点のみに着目しているため、目的音区間の一部分においてしか、方向推定結果が得られない。このことは突発雑音に対応した雑音抑圧 [24]、移動音源追尾等への適用に問題があり、本研究の提案法の適用分野を狭くしているため、より広い区間を方向推定の対象とすることが好ましい。但し提案法が音の立ち上がり点に着目しているのは、上述の反射音に対する不応期を設定するためであり、先に述べたような残響抑圧処理がこの不応期制御に変わるものであれば、この制約から解放される可能性がある。

この他に本研究の水平面の方向推定から 3次元空間の方向推定への拡張、平面波の仮定が成り立たないような近接音源への対応等があげられる。

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。

最後に、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] D. Rabinkin, R. Renomeron, J. French and J. Flanagan, "Estimation of Wavefront Arrival Delay Using the Cross-Power Spectrum Phase Technique" , J. Acous. Soc. Am., Vol. 100, No. 4 Pt. 2, pp. 2697, October 1996
- [2] E. E. Jan and J. L. Flanagan, "Sound Source Localization in Reverberant Environments using an Outlier Elimination Algorithm", Proceedings of ICASSP , Atlanta, 1996.
- [3] 中村、猿渡、鹿野, "移動マイクロホンアレーを用いた残響下における音源位置推定," 音論講集 (秋), Vol.1, 3-P-24, pp.249-496, 2000.
- [4] 山田、中村、鹿野, "マイクロホンアレーを用いた話者方向検出による音声認識," 音論講集 (秋), Vol.1, 1-2-4, pp.7-8, 1995.
- [5] 西浦、山田、中村、鹿野, "マイクロホンアレーを用いた CSP 法に基づく複数音源位置推定," 信学論, Vol. J83-D-II, No.8, pp.1713-1721, 2000.
- [6] M. Brandstein, "Time-Delay Estimation of Reverberated Speech Exploiting Harmonic Structure," J. Acoust. Soc. Am., 105(5), pp.2914-2919, 1999.
- [7] M. Brandstein, "A Pitch-Based Approach to Time-Delay Estimation of Reverberant Speech," In 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, October 19-22 1997.
- [8] M. Brandstein, J. Adcock, and H. Silverman, "A Practical Time-Delay Estimator for Localizing Speech Sources with a Microphone Array," Computer, Speech, and Language, 9, pp.153-169, April 1995.
- [9] 浅野、速水、松井, "話者方向同定と雑音抑制による音声認識性能の改善," 日本音響学会誌 , 第 53 卷 11 号, pp.889-894, 1997.

- [10] 王、竹内、大西、杉江, "能動的に音源を定位し識別する自動移動ロボット," 日本ロボット学会誌, vol15, pp.223-229, 1997.
- [11] A.M. Noll, "Cepstrum pitch determination," Journal of the Acoustical Society of America, 41, pp. 293-309, Feb 1967.
- [12] M.Akagi et.al., "Speech Enhancement and Segregation based on Human Auditory Mechanisms," Proc, Int. Conf. Information Society in the 21st Century(IS2000), pp.102-126, Aizu-Wakamatsu, November 2000.
- [13] 石本 祐一, "雑音環境における基本周波数の推定とこれを用いた雑音抑圧に関する研究," JAIST 修士論文, 2000
- [14] Jont B. Allen, David A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," J.Acoust.Soc.Am. 65(4), pp. 954-950, April 1979.
- [15] Yue Pan, Alex Waibel, "THE EFFECTS OF THE ROOM ACOUSTICS ON M-FCC SPEECH PARAMETER," International Conference on Speech and Language Processing, Beijing, China, October. 2000.
- [16] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," Proceedings of ICASSP 94, vol.2, pp. 273-276, Adelaide, 1994.
- [17] M. Omologo and P. Svaizer, "Acoustic source location in noisy and reverberant environment using CSP analysis," Proceedings of ICASSP 96, pp. 921-924, Atlanta, 1996.
- [18] M. Omologo, P. Svaizer, "Use of the Crosspower-Spectrum Phase in Acoustic Event Location," IEEE Transactions on Speech and Audio Processing, pp.288-292, May 1997.
- [19] B. Champagne et.al., "Performance of Time-Delay Estimation in the Presence of Room Reverberation," IEEE Transactions on Speech and Audio Processing, pp.148-152, March 1997.
- [20] 吉本 功一, "音声認識におけるHMMの残響適応に関する研究," JAIST 修士論文, 2000.

- [21] 王 紅, "残響音響信号処理," 信学技報 SP92-78, pp.35-42, 1992.
- [22] H. Wang, P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing", Proc. of ICASSP 97, Munich, Germany, April 1997.
- [23] 水町 光徳, 多点受音信号を用いた残響除去法に関する基礎的研究, JAIST 修士論文, 1997.
- [24] M.Mizumachi, M.Akagi, "Noise reduction method that is equipped for a robust direction finder in adverse environments," Proc. IEEE Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, FINLAND, pp.179-182, 1999.
- [25] 西田 知之, "多数マイクロホンによる音源推定方向に関する研究," JAIST 修士論文, 1999.
- [26] 黄・大西・杉江, "生体に示唆を得た音源定位システム — 反響のある環境での単一音源定位 —", 電子情報通信学会論文誌 A Vol.J71-A No.10 pp.1780-1789, 1988.
- [27] Jie Huang, 'Auditory Spatial Processing in Reverberant Environment', ITC-CSCC, 1997
- [28] Jie Huang, 'Mobile Robot and Sound Localization', IEEE, 1997
- [29] 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 pp.268-273, 1991
- [30] 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
- [31] 日本音響学会編, 音のなんでも小辞典, 講談社ブルーバックス, 1996.
- [32] 古井 貞熙, デジタル音声処理, 東海大学出版会, 1985.
- [33] 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.
- [34] B.Gold, N.Morgan, Speech and audio signal processing: Processing and Perception of Speech and Music, chap.13, Wiley, New York, 1999.
- [35] H.Kuttruff, Room Acoustics, 3rd ed., Elsevier, London, 1991.

学会発表リスト

1. 小林 利秋, 西田 知之, 赤木 正人, "雑音と反射音に対してロバストな話者方向推定法," 日本音響学会 2001 年春季研究発表会, 2-7-9, 2001.