修 士 論 文

パーティクルフィルタを用いた雑音に頑健な音声スペクトル上の複数ローカルピーク推定に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

友池 誠二

2008 年 3 月

修 士 論 文

パーティクルフィルタを用いた雑音に頑健な音声スペクトル上の複数ローカルピーク推定に関する研究

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　鵜木 祐史 准教授
審査委員　徳田 功 准教授

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

0610061 友池 誠二

提出年月: 2008 年 2 月

概 要

本論文では，パーティクルフィルタを用いた雑音に頑健な音声スペクトル上の複数ローカルピーク推定法を提案する．本研究が対象とするローカルピークは，音声スペクトル上のあらゆる高調波に対するローカルピークである．音声スペクトル上のローカルピーク推定に関して，従来法では過去のフレームで推定されるピークに関する知識を利用していない．つまり従来法では，現在のフレームの情報のみを使ってローカルピークの推定を行うため，従来法の推定精度は現在のフレームに存在する雑音に大きく影響される．このため，従来法は雑音に対する頑健性が低いという問題点がある．そこで，過去のフレームで推定されるローカルピークに関する知識を学習ようなローカルピーク推定を考える．

本論文では，パーティクルフィルタを用いた提案法によって過去のフレームで推定されるローカルピークを学習し，現在のフレームのローカルピークを推定する．パーティクルフィルタはシステムの状態推定に際して，事後分布を正確に近似することで，システムの状態を推定する方法である．提案法では，ローカルピーク推定に際して，音声スペクトル上の複数のローカルピークを同時に推定する尤度を導入する．この尤度を用いることで複数のローカルピークの存在確率を同時に推定可能であるため，提案法は個数が未知である複数のローカルピーク推定に適用可能である．さらに，状態推定に尤度と事後分布のみを用いているため，ローカルピークの動きのモデル化を必要としない．入力される音声によって，ピークの個数は異なるため，ピークの個数を仮定することなく過去のピーク情報から複数のローカルピーク推定が可能であるのは大きな特長である．

提案法は大きく分けて 2 つの手順で構成される．第 1 の手順は，ケプストラムから得られるスペクトル包絡を尤度とする，ピークの存在確率推定である．高調波は，高域成分ほど基本周波数の整数倍にはなっていない．提案法では，個数が未知である各倍音がわずかなゆらぎを持つような独立した動きを持つ場合であってもパーティクルフィルタによってピークの存在確率を細かく表現することによって，ピークの存在確率を推定可能とする．第 2 の手順は，ピークの存在確率から得られる，ローカルピークとなりうる候補からローカルピークを抽出する手法である．

提案法が雑音環境下で精度良くローカルピークの推定が行われているかを検証するために実験を行った．条件 1 として，基本周波数の変化に直線的または正弦波的変化を適用し，それを基音とする倍音によって合成波形を生成する．合成波形の振幅は一定とする．この合成波形に，白色雑音，ピンク雑音，狭帯域雑音を付加したものを入力とし，雑音に対する頑健性の評価を行った．条件 2 では，条件 1 で使用する合成波形の振幅変化に実音声の持つ値を適用したものを使用する．実験の結果，推定ピーク数と正解ピーク数との差，推定ピークの正解との距離の両方の尺度から，提案法は定常雑音と非定常雑音の双方で従来法よりも推定精度が高いことが明らかになった．このことから，提案法は過去のフレームのローカルピークの学習によって，従来法の問題点であった雑音に対する頑健性を高めることがわかった．

目 次

図 目 次

表 目 次

第1章　序論

1.1　本研究の背景

1.1.1　ローカルピーク推定の必要性

人間の音声の主な特徴量として高調波がある．高調波は音声認識，基本周波数推定，音声強調といった音声情報処理における重要な役割を担っている．そして，高調波は周波数領域における音声スペクトルのローカルピークと密接な関係がある．有声音は声帯の振動によって発生する準周期的なパルスで声道が励振されて生じる．図 1.1 に男性話者による発話/a/の有声区間における第 1 フレームの振幅スペクトルを示す．この図から，高調波が基本周波数とそのほぼ整数倍の成分から構成されていることがわかる．各高調波成分は振幅と周波数によって表現される．高調波と音声の関係を示した応用例として，McAulay と Quatieri[1] は正弦波の足し合わせに基づいた音声の分析合成システムを提案している．つまり，McAulay と Quatieri は，有声音声は瞬時振幅や周波数が正弦波の和として表すことができ，正弦波の数，各正弦波の瞬時振幅，瞬時周波数，初期位相の各パラメータが与えられれば，そこから有声音声を得られることを示した．しかし，高調波の倍音の数は未知であり，高域成分はゆらぎを持つため，完全な基本周波数の整数倍にはなっていない．そのため，倍音の数を仮定せず，高調波の各倍音を独立に推定する手法が求められる．

本研究では，高調波を与える 1 つのパラメータとして，高調波の各成分の周波数と対応する音声スペクトルのローカルピーク推定を行う．本研究が対象とするローカルピークは，音声スペクトル上のあらゆる高調波に対するローカルピークである．音声スペクトルからローカルピークが推定できれば，高調波の周波数領域における雑音抑圧が可能となる．高精度な音声の高調波周波数は，様々な音声信号処理が実現可能となる．例えば，高調波成分のスペクトル強調に基づいた音声認識 [2] が可能である．この手法は，高調波のローカルピーク近辺のスペクトルを強調することで，高調波成分の周波数を音声認識のパラメータとして利用する．ここで，ローカルピークが精度良く推定できれば，さらなる認識率向上が望める．また，高調波の周波数を規範とした櫛型フィルタを形成することで基本周波数の推定 [3] が可能である．この手法では，櫛形フィルタによって雑音を抑圧し，雑音抑圧された音声に対して基本周波数推定を行う．ここで，ローカルピークが精度良く推定できれば，雑音らしい部分を除去し，より高精度な基本周波数推定が望める．さらに，高調波の倍音の周波数が得られれば，倍音の各周波数ビンの SNR を推定することで振幅方向の雑音除去が可能となり，音声強調を実現できる [4]．このように音声の特徴

を利用した様々な音声情報処理において，高調波の情報が必要となる．そのため，高調波を与えるパラメータとしてローカルピーク推定が重要な課題であるといえる．

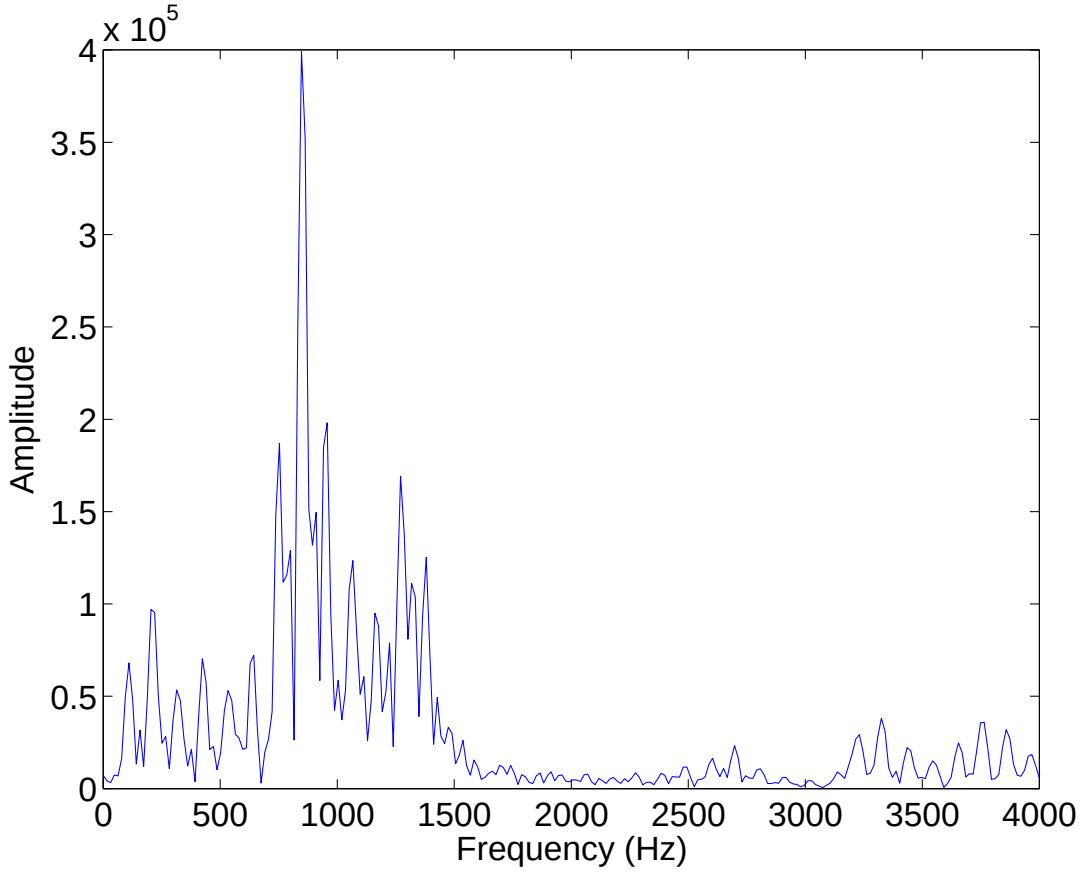


図 1.1: 男性話者発話の母音/a/の振幅スペクトル (平均基本周波数 110.3846Hz)

1.1.2 ローカルピーク推定の問題点

1.1.1 節で述べたように，高調波の倍音の数は未知であり，高域成分はゆらぎを持つため，完全な基本周波数の整数倍にはなっていないという問題がある．高調波の個数が未知であり，各高調波の動きが異なるものに対してそれぞれ動きを定式化するのは困難である．そのため，一般的に雑音環境下でローカルピークを推定することは非常に困難であると考えられている．ローカルピーク推定に関する研究は古くから行われている．しかしながら，入力信号の種類に関係なくあらゆるローカルピークを推定する手法では，高調波によるローカルピークであるのか，それとも雑音によるローカルピークであるのかを区別する判断材料がなく，パラメータの調整が困難である．そのため，音声スペクトルのローカルピーク推定という目的においては，高調波の情報をローカルピークの選択基準にすると

いったローカルピーク推定法が重要となる．また，従来法の多くは，音声はクリーンな状況で観測されるものと仮定されてきたため，あらかじめ雑音抑圧などの前処理を必要とする．しかし，実環境においては様々な雑音が存在し，音声信号処理システムの性能は前処理の精度に大きく依存する．そして深刻な問題の１つとして，従来法では，現在のフレームの情報のみを使ってローカルピークの推定を行うため，従来法の推定精度は現在のフレームに存在する雑音に大きく影響されるという点が挙げられる．よって，雑音環境下におけるローカルピーク推定の精度に関しては十分とはいえない．

従来法の問題に関して，過去のフレームで推定されたローカルピークの学習が行われていれば，音声スペクトルの変化が緩やかな場合において，過去に推定されたローカルピークの位置の学習によってピークの存在確率が得らる．ピークの存在確率が得られれば，現在のフレームにおけるピークの存在の確からしさが得られるため，現在のフレームにおけるローカルピークの周波数位置の推定が可能である．つまり，ローカルピークの周波数位置の推定によって，多数得られたローカルピークの候補の中から真の音声のローカルピークらしい周波数位置が得られる．

雑音を含む観測値からシステムの状態を過去のフレームを用いて推定する手法として，カルマンフィルタ [5] が広く用いられている．システムの状態をローカルピークとし，カルマンフィルタによって単一のローカルピークの動きをトラッキングすることが可能である．しかし，ローカルピークの予測に関して，音声スペクトルの変化をモデル化するのは困難である．また，トラッキングするローカルピークの個数が未知であり，しかも複数のローカルピークを同時に推定するような場合，複雑な形状を持つ確率密度分布の表現が困難である．よって，複数のローカルピーク推定にカルマンフィルタを用いるのは困難である．

パーティクルフィルタ [6] は，雑音を含む観測値からシステムの状態を過去のフレームを用いて推定する新たな手法として近年注目されている．パーティクルフィルタの特徴は，離散的な値と重みを持つ多数のパーティクルの集合によって事後確率分布を表現し，推定する点である．パーティクルの密度によって事後確率の大きさを表現する．ここで，複数のローカルピークの存在確率を事後確率分布で表現すれば，観測値と事後確率分布により，次のフレームにおけるピークの存在確率が予測できる．そのため，状態遷移に関するモデル化を用いることなく，ピークの存在確率を動的に更新できる．

1.2　本研究の目的

人間の音声の特徴量である高調波と密接な関係にある音声スペクトルのローカルピークの推定は，様々な音声情報処理に応用できる重要なパラメータであり，これまでに様々なローカルピーク推定法が提案されてきた．従来のローカルピーク推定に関する研究では，高 SNR での定常雑音に対してはローカルピークの推定が可能であるものの，低 SNR で音声区間中に突発的に発生するような雑音などに対しての推定精度は非常に悪い．実環境では，目的音と雑音が周波数成分・時間成分で予測不可能な重なりを持つような非定常雑音

が付加された音声が観測される場合が多い．そのため，突発的な雑音のような非定常雑音に頑健なローカルピーク推定法が求められる．

本論文では，パーティクルフィルタを用いた提案法によって過去のフレームで推定されるローカルピークを学習し，現在のフレームのローカルピークを推定する．パーティクルフィルタはシステムの状態推定に際して，事後分布を正確に近似することで，システムの状態を推定する方法である．提案法では，ローカルピーク推定に際して，音声スペクトル上の複数のローカルピークを同時に推定する尤度を導入する．この尤度を用いることで複数のローカルピークの存在確率を同時に推定可能である．ローカルピークの存在確率が高い部分は近接フレームにおいてもピークが存在する確率が高い．そのため，提案法は個数が未知である複数のローカルピーク推定に適用可能である．さらに，状態推定に尤度と事後分布のみを用いているため，ローカルピークの動きのモデル化を必要としない．入力される音声によって，ローカルピークの個数は異なるそのため，ローカルピークの個数を仮定することなく過去のローカルピーク情報から複数のローカルピーク推定が可能であるのは大きな特徴である．

提案法は大きく分けて 2 つの手順で構成される．第 1 の手順は，ケプストラムから得られるスペクトル包絡を尤度とする，ピークの存在確率推定である．この方法で得られるピークの存在確率は尤度を用いて動的に更新される．そのため，提案法は音声スペクトルのローカルピークの変化に関するモデルを用いることなく，ピークの存在確率を推定可能である点が特徴である．高調波は，高域成分ほど基本周波数の整数倍にはなっていない．提案法では，個数が未知である各倍音がわずかなゆらぎを持つような独立した動きを持つ場合であってもパーティクルフィルタによってピークの存在確率を細かく表現することによって，ピークの存在確率を推定可能とする．第 2 の手順は，ピークの存在確率から得られる，ローカルピークとなりうる候補からローカルピークを抽出する手法である．よって，従来法の問題点である，過去のフレームで推定されたローカルピークの学習が生かされていないという点が解決した手法が実現できる．

提案法が，過去のフレームのローカルピークの学習によって，雑音環境下で精度良くローカルピークの推定が行われているかを検証するために実験を行う．この目的のためには，まず音声スペクトルのローカルピーク推定に関する従来法が定常雑音環境下と非定常雑音環境下での推定精度を評価する必要がある．その後，この提案法の定常雑音環境下と非定常雑音環境下でのローカルピークの推定精度を評価することで有効性を示す．1.1.2 節に示した考え方から，本論文では，高調波の学習が有効に動作する制約として

- 音声スペクトルの変化は緩やかである
- パーティクルフィルタの初期値として，初期フレームに高 SNR 部を仮定する

という条件を設け，提案法によってこれらの条件下で精度良く複数ローカルピークが推定できることを示す．

1.3　本論文の構成

本論文は，全5章により構成されている．各章の見出しと概要を以下に述べる．

第1章：序論

ローカルピーク推定法の現状および問題点を明確にし，本研究の目的を述べる．

第2章：従来のローカルピーク推定法の有効性

計算機シミュレーションにより，従来のローカルピーク推定法が，雑音環境下でどの程度ローカルピークを推定可能であるのかを詳細に調査する．

本章では，従来のローカルピーク推定法のうち，あらゆるローカルピークを抽出することを目的とした二階微分を用いた手法，山登り法を用いた手法と，高調波の追跡を目的とした瞬時周波数を用いた手法に対してより詳細な検討を加える．

第3章：パーティクルフィルタを用いた複数ローカルピーク推定法の提案

本章では，非定常雑音に対応可能な複数ローカルピーク推定アルゴリズムの定式化を行う．まず，ピークの存在確率を求める．次に，ピークの存在確率から，ローカルピークの候補を決定し，その中からローカルピークを抽出する．以上により，非定常雑音にも対応可能な複数のローカルピークの推定を実現する．

第4章：複数ローカルピーク推定法の有効性

第3章で定式化した複数ローカルピーク推定アルゴリズムのローカルピーク推定能力を検証する．提案法の特徴である雑音環境下での複数ローカルピーク推定に関して，計算機シミュレーションによる複数ローカルピーク推定実験を行う．そして，従来法に対する提案法の優位性について検討し，検証結果およびに本研究全体を考察する．

第5章：結論

本論文で得られた結果を要約するとともに，今後検討すべき課題について述べる．

第2章　従来のローカルピーク推定法の有効性

2.1　従来のローカルピーク推定法

従来のローカルピーク推定法について述べる．従来のローカルピーク推定は大別すると，

1. 存在するあらゆるローカルピークを抽出する手法

2. 音声の特徴を利用してローカルピークを抽出する手法

に分類できる．入力されるデータ列から，どのようにローカルピークを取捨選択するかによってローカルピーク推定の精度と適した推定対象が決定できるため，どのようなローカルピークを選択するかを決めることが重要である．あらゆるローカルピークを抽出し，ピークの鋭さやピークとディップの差の大小で判断する方法が，1. の存在するあらゆるローカルピークを抽出する手法である．この考えで取捨選択を行う場合，雑音によって発生するピークがローカルピークの選択基準となるしきい値を超えるような低SNR の場合に，本来のローカルピークであるのか雑音によるローカルピークであるのかを区別できないという問題がある．また，しきい値を調整することでピークの個数を制御できるが，それが目的に合ったものであるかどうかの判断ができない．よって高 SNR の場合の定常雑音環境下での使用に適しているといえる．

ローカルピークの取捨選択に関して，フォルマントやスペクトル包絡などを用いてより用途を限定して効果的にローカルピークを推定するのが，2. の音声の特徴を利用してピークを抽出する手法である．この考え方では，あらかじめどのローカルピークを選択するかを判断する選択基準を用意する必要があるが，目的に合うローカルピークを重み付けして集中的に推定を行う．そのため，比較的低 SNR であっても推定が可能であるという特徴がある．

2.1.1　存在するあらゆるローカルピークを抽出する手法

二階微分を用いたピークピッキング

音声スペクトル全体に対して，一階微分と二階微分をとり，一階微分値が 0 かつ二階微分値が大の条件を満たす点をローカルピークの候補とする．各ローカルピーク候補のス

ロープをしきい値としてスロープの大小によってピークを選択する．スロープを選択基準としているため，最終的に得られるローカルピークとして，各ローカルピーク候補のスロープが鋭いものが抽出される．そのため雑音環境下では，定常雑音で高 SNR の入力に対して有効なローカルピーク抽出法である．

山登り法を用いたピークピッキング [7]

音声スペクトル全体を順に走査していき，ピークとディップの差をしきい値とし，しきい値より小さいような極小のピークを無視することで雑音によって発生したと思われるピークを除いたローカルピークが得られる．スロープが鋭いピークはピークとディップの差が大きく抽出されやすくなるため，二階微分を用いた手法で選択されるようなローカルピークが得られる．また，山登り法では二階微分を用いた手法と異なり，幅が広いようなローカルピークや，微分不可能なフラットなローカルピークも得られるのが特徴である．

2.1.2 音声の特徴を利用してピークを抽出する手法

瞬時周波数を用いた高調波推定法 [8]

音声が持つ高調波の特徴を用いて，入力データが音声の場合に高調波を精度良く抽出する手法である．音声信号を倍音成分に分解するために帯域通過フィルタ群を用いている．各帯域通過フィルタの中心周波数は出力の瞬時周波数を追跡することで層的な更新を行う．この手法では，倍音の数が不明である場合のパラメータの初期値設定が困難な点や，追跡対象の増減に対応できない問題点がある．そのため，入力の音声が男であるか女であるかを仮定する，学習のためのクリーンな音声区間を用意するといった制約を設けることで高調波の推定ができる．

2.2 雑音環境に対する頑健性の評価

本節では，従来のローカルピーク推定法の対雑音特性の評価を行う．そこで本論文では，ローカルピークの選択基準として，ピークの鋭さ，ピークとディップの振幅差を用いた多用途向けの手法と，高調波の推定に特化した手法という性質の異なる選択基準を用いたローカルピーク推定法を取り上げ，対雑音特性の評価を行う．対象とする手法は次の 3 つの手法である．

- 二階微分を用いたローカルピーク推定法
- 山登り法を用いたローカルピーク推定法
- 瞬時周波数を用いたローカルピーク推定法

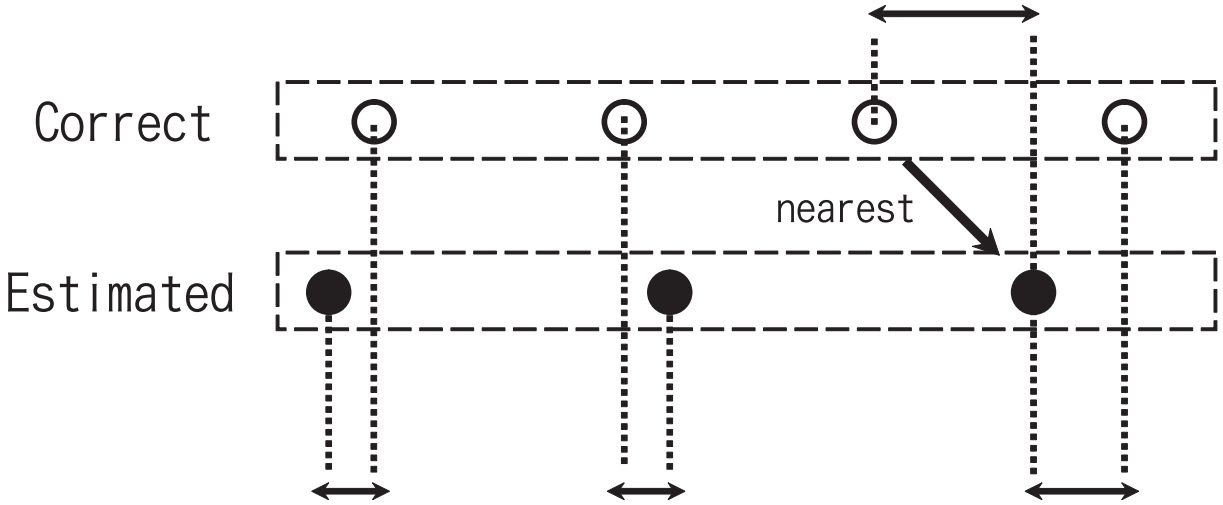


図 2.1: 距離の定義

2.2.1 評価方法

評価は，推定ピーク数から正解ピーク数を減じたピークの過大推定数と，推定ピーク位置と正解ピーク位置の距離によって比較を行う．ここで，ピーク間の距離を次のように定義する．正解ピークに対応する推定ピークとの周波数の差をとる，つまり正解ピーク位置に最も近い推定ピーク位置との差を取る．(図 2.1)

$$df = \overline{|f_{cor} - \hat{f}_{est}|}\ , \tag{2.1}$$

ここで，df は周波数距離，f_{cor} は正解ピークの周波数，f_{est} は推定されたピークの周波数である．1 フレーム内に存在するピーク全体の差の平均を距離とする．ただし，差が基本周波数の値を超えた場合は対応するピークがないと見なし，基本周波数の値に相当する距離を加算する．推定ピーク数が正解ピーク数より大きい場合は過大推定数だけ基本周波数の値を加算する．正解ピークの個数を比較することで，ローカルピークの個数が正確に推定できているかどうかを確認できる．また，個数だけ正確でローカルピークの周波数位置が正解ピークの位置から大きくずれている場合も考えられるので，ピーク間の距離を比較することで正解ピークから大きくずれていないかどうかを確認できる．よって，この 2 つの評価尺度はともに重要である．

実験は入力 SNR を $\infty, 20, 10, 0, -10$dB とした雑音付加音声に対して，二階微分を用いた手法，山登りを用いた手法，瞬時周波数を用いた手法をそれぞれ適用し，推定結果の比較を行う．定量的な評価を行うために，あらかじめ推定結果に対する正解となるピークを用意する．100Hz から 200Hz まで基本周波数が直線的に変化するデータのそれぞれの整数倍を正解ピークとし，時間長は 485ms，倍音は 40 個とする．正解ピーク位置に対して正弦波の足し合わせによって合成 [1] した波形を用いる．図 2.2 は合成波形生成時に用いる正解ピークである．図 2.2 の各時刻において，縦に並んでいる点の集合は，フレーム毎

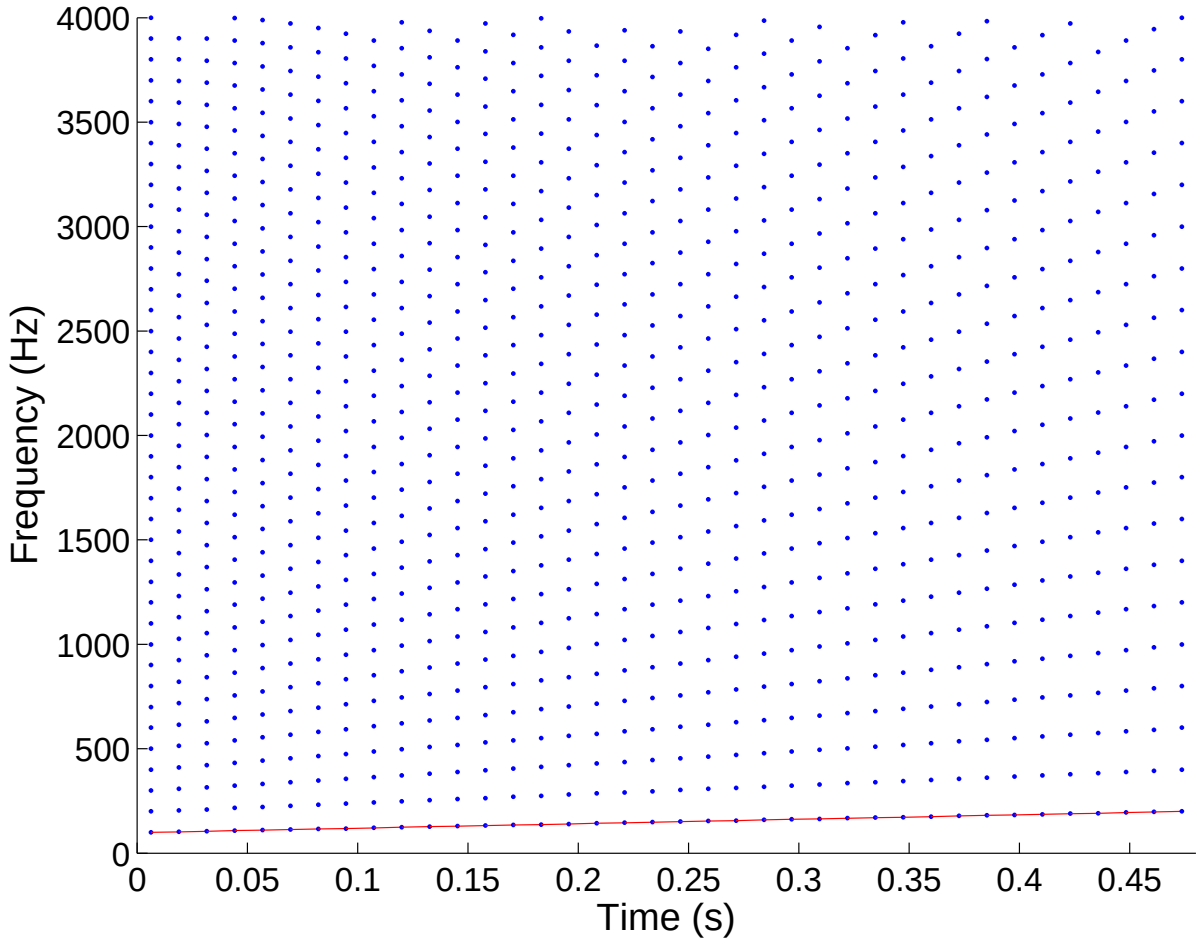


図 2.2: 時間長 485ms, 基本周波数直線的変化時の正解ピーク

の正解ピークの周波数位置を示している．振幅成分の変化は実音声の振幅成分と同じとする．振幅として使用する音声は，ATR 研究用音声データベースの A セットに収録されている男性話者 MHT 発話の短母音/aoi/である．正解と各手法によって推定されたピークとの距離と個数を比較する．

雑音にはピンク雑音，白色雑音，狭帯域雑音を用いる．狭帯域雑音の中心周波数を 1000Hz，帯域幅を 1000Hz，時間幅を 1 から 4 フレームとする．ピンク雑音，白色雑音は，音声区間全体に付加し，狭帯域雑音は 5 フレーム目を始点として付加する．評価区間については，ピンク雑音，白色雑音は音声区間全体，狭帯域雑音は雑音付加区間に加え，各手法の局所最適を回避するために直前 3 フレームも評価区間に加える．サンプリング周波数は 8kHz，フレームの切り出しには Hamming 窓を用い，窓長は 32ms，スライド幅は窓長の 0.4，FFT フレーム長は 512 とする．ピーク同士のピークを分離できる周波数分解能を得るために，窓によって切り出された信号の両端にゼロ値を挿入し，FFT フレーム長と同じにすることで，スペクトルをオーバーサンプリングすることと等価な処理を行う．

2.2.2 パラメータ設定

二階微分を用いたローカルピーク推定法 (SOD)

2.1 節で述べたように，ローカルピークの選択基準はスロープ値である．入力の違いによってローカルピークの個数は異なる．実験では，各ローカルピーク候補のスロープ値の平均を基準に係数 $slope\ rate$ で除した値をしきい値として設定する．

山登り法を用いたローカルピーク推定法 (HC)

2.1 節で述べたように，ローカルピークの選択基準はピークとディップの振幅差である．入力の違いによってローカルピークの個数は異なるため，相対的なパラメータを設定するのが望ましい．実験では，入力の最大値と最小値の差による振幅幅に係数 $height$ を乗じた値をしきい値として設定する．

瞬時周波数を用いたローカルピーク推定法 (IF)

2.1 節で述べたように，ローカルピークの選択基準は帯域通過フィルタの中心周波数である．帯域通過フィルタは最初，等間隔に配置する．中心周波数の間隔が高調波の基本周波数よりも大きいと正しく追跡が行えないため，フィルタ間隔 $interval$ は基本周波数以下とする．入力の音声が男性のものであるか女性のものであるかは未知であるので，男性に合わせるのが望ましい．また，帯域通過フィルタの通過域 $bandwidth$ はフィルタ間隔のおよそ 1.5 倍以上とする．tc は時定数である．

2.2.3 パラメータ設定方針

パラメータは，クリーン音声で精度良くローカルピーク推定が可能な値の範囲から複数回の試行で得られた最適値を手動で設定する．各手法に適用したパラメータを表 2.1 に示す.

表 2.1: 従来法のパラメータ（左：二階微分法，中央：山登り法，右：瞬時振幅法）

slope rate	*height*	*bandwidth*	*interval*	*tc*
10	0.01	190	100	0.05

2.2.4 評価結果

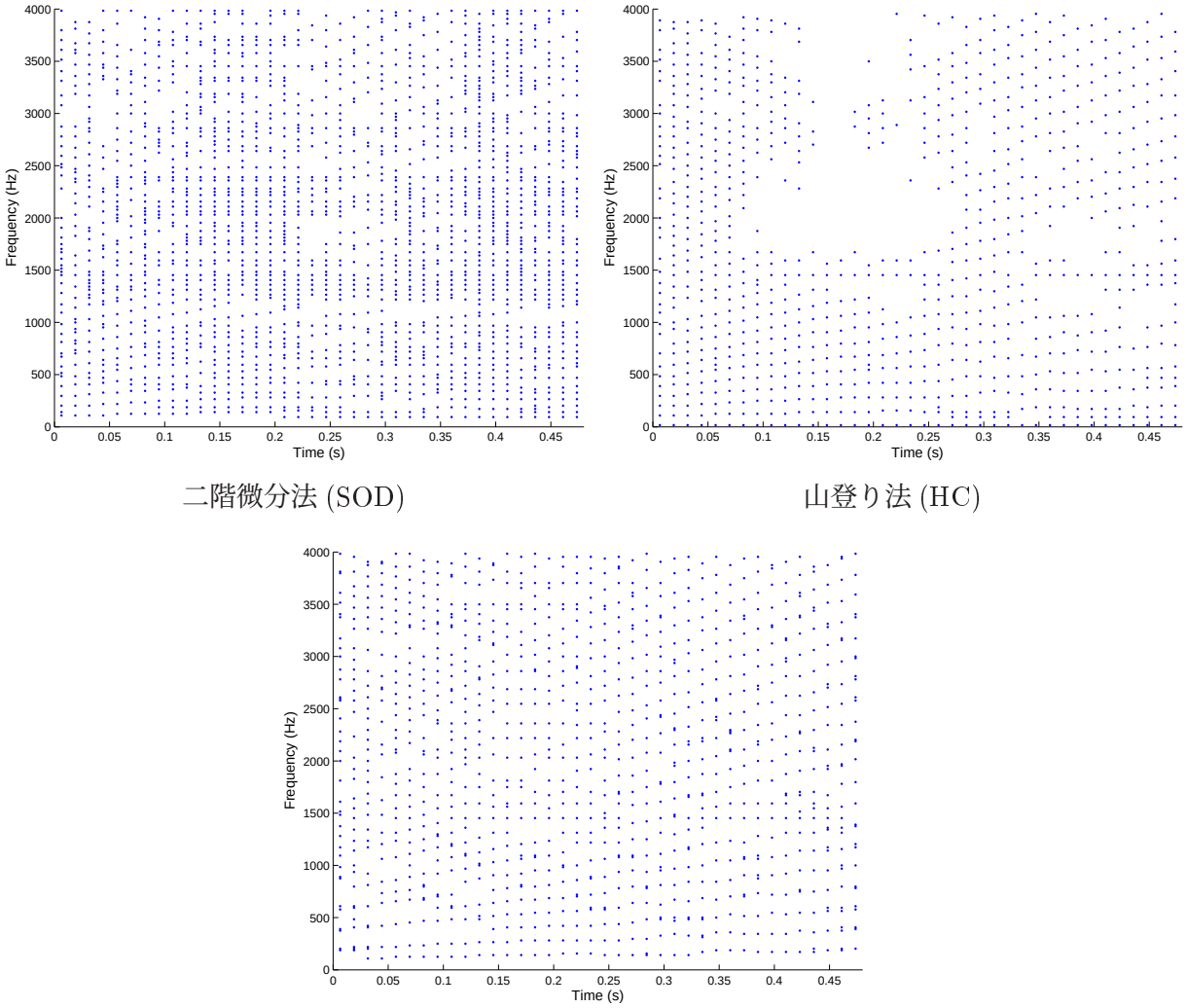
実験の結果を図 2.3 と，図 2.4 から図 2.9 に示す．図 2.3 は，入力 SNR0dB, 狭帯域雑音幅 4 フレームにおける推定ピークの時間-周波数領域の関係を表しており，各時刻において，縦に並んでいる点の集合は，フレーム毎の正解ピークの周波数位置を示している．正解ピークである図 2.2 と比較すると，定常雑音で比較的高 SNR の条件であってもローカルピーク推定精度に大きな違いが出ていることが見てとれる．これらの実験から，ローカルピーク推定に対して，過去の情報を効果的に利用することが重要であることがわかる．

図 2.4 はピンク雑音を付加した条件での結果，図 2.5 は白色雑音を付加した条件での結果，図 2.6 から図 2.9 はそれぞれ狭帯域雑音の長さを 1 フレームから 4 フレームに変えた時の結果である．横軸は入力 SNR を示しており，$-10, 0, 10, 20, \infty$dB の順に並んでいる．図はいずれも縦軸が 0 に近いほど正確にローカルピークの推定がなされたといえる．

実験の結果から，SOD と HC は，どの雑音を用いた場合でも，入力 SNR が低くなるにつれ距離と過大推定ピーク数が増加傾向にあることがわかった．特に負の入力 SNR における距離の増加度合いが大きい．SOD と HC にピンク雑音と白色雑音を用いた場合，ピークの推定個数に関して過大であったり，過小であったりと結果が安定していないことがわかった．図 2.5 を見ると，過小推定と過大推定の遷移が見て取れ，さらに個数の入力 SNR が ∞dB の場合の距離と，入力 SNR が -10dB の場合の距離は 5 倍近くの差が出た．SOD と HC は，クリーン音声用に設定したパラメータを用いたため，雑音区間で雑音の影響を大きく受けたためと考えられる．SOD と HC に狭帯域雑音を用いた場合には，ピークの推定数はどの入力 SNR においても SOD は過大推定，HC は過小推定となった．よって，SOD と HC は定常雑音か非定常雑音のどちらかに最適化したパラメータが設定可能といえる．

一方，IF はピンク雑音，白色雑音においては距離と推定ピーク数の双方で SOD と HC の精度を上回った．狭帯域雑音を用いた場合，IF はピークを過小推定しているものの入力 SNR によらず距離，推定ピーク数双方で安定した推定精度が得られている．IF は，学習こそ行っていないものの，直前のフレームで推定されたローカルピークの周波数位置を

元に現在のフレームの推定ピーク位置を推定する．IF の更新の度合いは時定数によって制御可能であるので，適切なパラメータを設定できれば精度良くローカルピークの追跡が可能となる．



二階微分法 (SOD)　　山登り法 (HC)

瞬時周波数法 (IF)

図 2.3: 入力 SNR20dB, ピンク雑音での各従来法の推定ピークのプロット

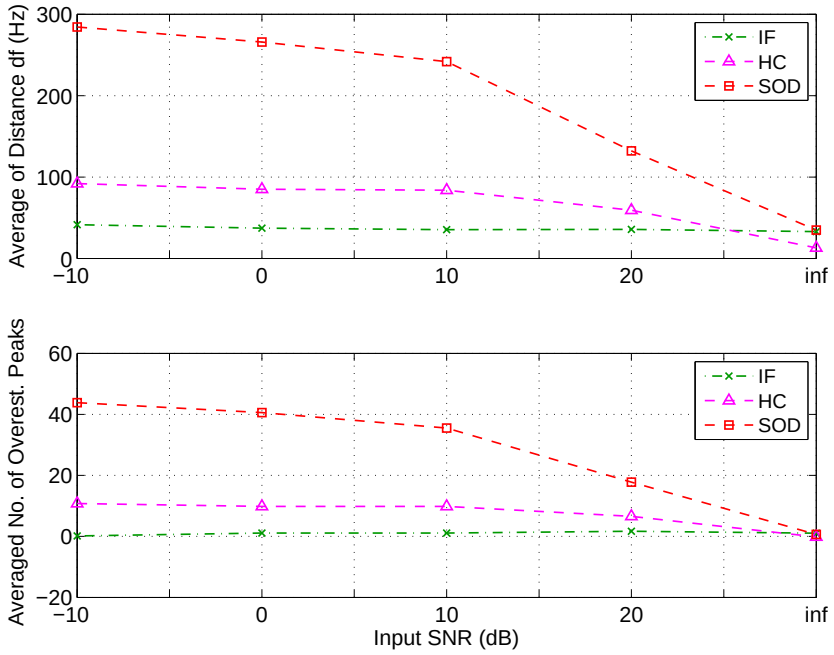


図 2.4: ピンク雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

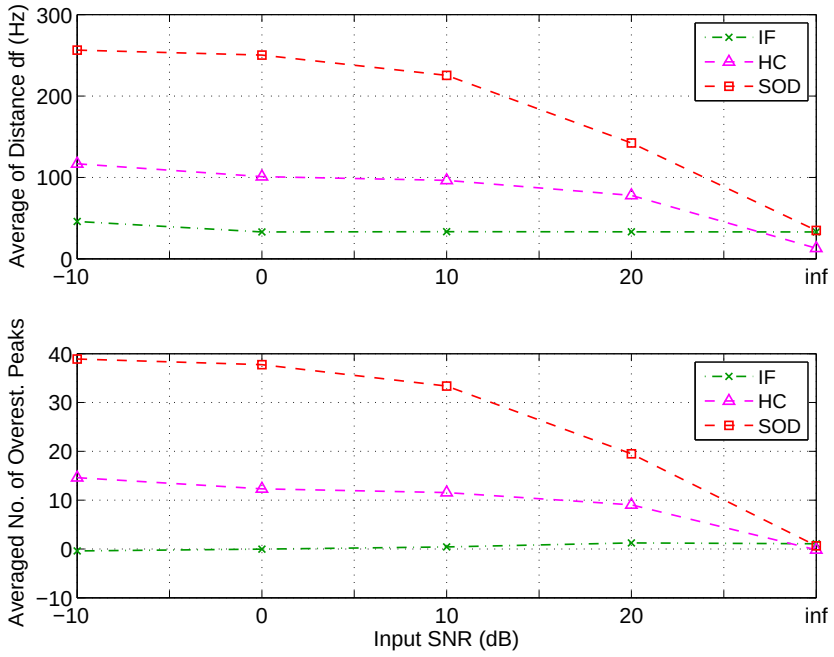


図 2.5: 白色雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

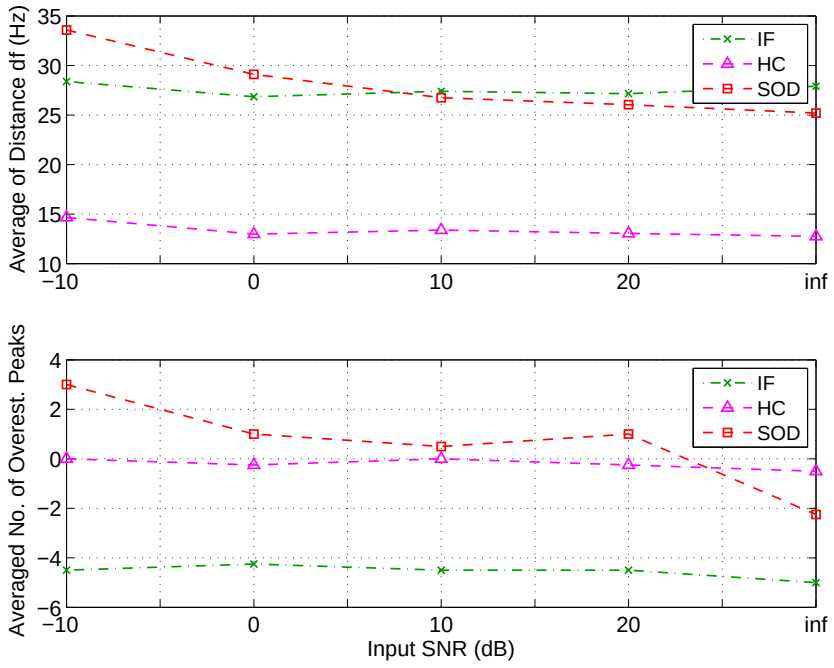


図 2.6: 時間幅 1 フレームの狭帯域雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

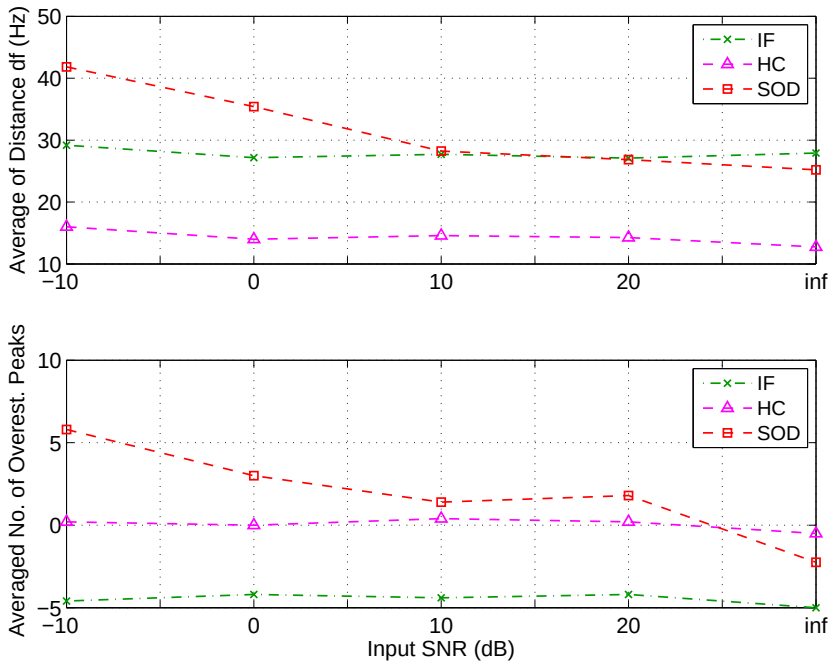


図 2.7: 時間幅 2 フレームの狭帯域雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

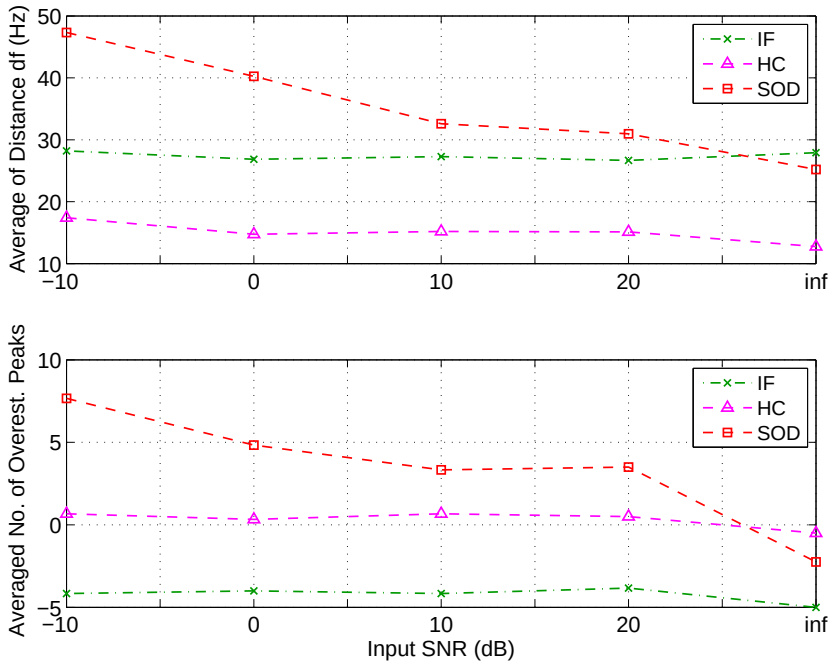


図 2.8: 時間幅 3 フレームの狭帯域雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

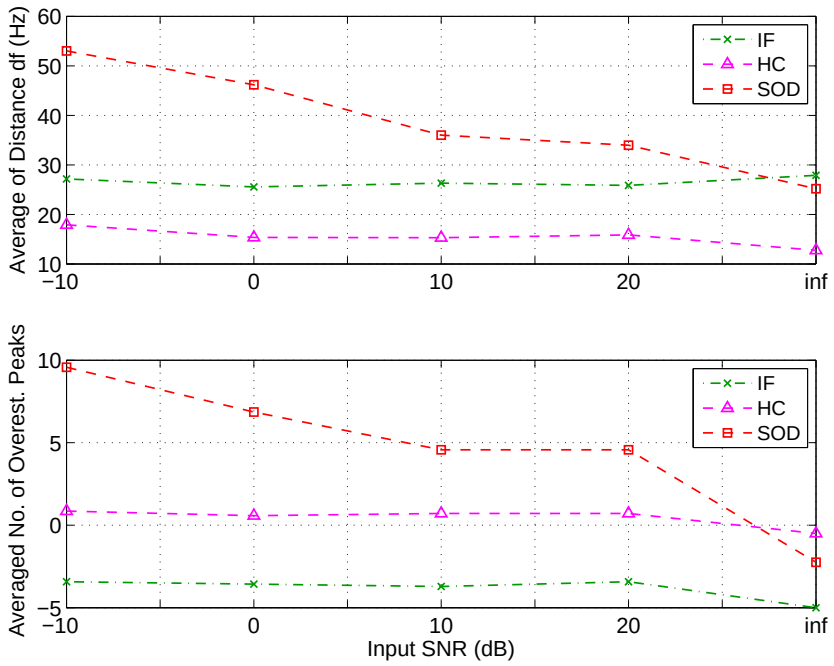


図 2.9: 時間幅 4 フレームの狭帯域雑音を用いた実験 (上段：平均距離，下段：平均過大推定ピーク数)

第3章　パーティクルフィルタを用いた複数ローカルピーク推定法の提案

3.1　問題の定式化

時刻 t における観測信号 $y(t)$ は音声信号 $s(t)$ と観測雑音信号 $w(t)$ の線形和で表現できると仮定し，

$$y(t) = s(t) + w(t) \tag{3.1}$$

と表す．このとき $s(t)$ と $w(t)$ が独立であれば，フーリエ変換 $\mathcal{F}[\cdot]$ により式 (3.1) は，

$$Y_t(f) = S_t(f) + W_t(f) \tag{3.2}$$

となる．ただし，$Y_t(f) = \mathcal{F}[y(t)]$，$S_t(f) = \mathcal{F}[s(t)]$，$W_t(f) = \mathcal{F}[w(t)]$ である．問題を簡単化するために，残響による信号の歪みは考慮しない．

観測信号 $y(t)$ を表現するのに必要な変数のベクトルを $x(t)$ とする．この $x(t)$ は状態ベクトルと呼ばれ，直接観測できない隠れ変数である．過去のフレームで推定されたピークの学習によって確率的に状態ベクトルを更新するために，提案法では状態ベクトルでピークの存在確率を表現する．このとき，過去のピーク位置を近辺の遷移確率を高くする．状態ベクトルでピークそのものを推定せずに，ピークの存在確率を推定することにより，ピークの動きが未知の状況でもピークの存在確率の学習で過去の推定ピーク位置付近のピークが推定可能となる．目的信号のフーリエ変換 $S_t(f)$ と状態ベクトルのフーリエ変換 $X_t(f)$ の関係は式 (3.3) のようにする．

$$S_t(f) = A_t(X_t(f)) \tag{3.3}$$

ここで，$X_t(f)$ は状態ベクトルのフーリエ変換である．状態ベクトル $X_t(f)$ は，高調波のローカルピークを与えるピークの存在確率を表現する．しかし，実際にはピークの存在確率 $X_t(f)$ から音声スペクトル $S_t(f)$ は完全に得られないため，A_t は $X_t(f)$ から $S_t(f)$ を近似する関係を表現する関数とする．ピークの存在確率は，過去の状態から現在の状態へと非線形な状態遷移関数 $Q_t(\cdot)$ によって状態遷移する．よって，式 (3.4) のような状態遷移のモデルが得られる．

$$X_t(f) = Q_t(X_{t-1}(f)) + V_t(f) \tag{3.4}$$

ここで，$V_t(f)$ はシステム雑音である．式 (3.4) に関して，提案法では，パーティクルフィルタによって動的に更新される状態ベクトルの推定値 $X_t(f)$ と尤度を用いて推定を行う自己組織型の状態空間モデルを扱う．カルマンフィルタでは，状態遷移を陽に記述する必要があり，1 次近似による状態更新が行われる．また，状態遷移に関して AR モデルや声道情報といった情報を用いるとフォルマントに特化された推定がなされ，一部ローカルピークが得られなくなる．そのため，推定の途中で状態の次元数が増減した場合に，状態の予測・更新ステップでのローカルピークの対応が取れず，正しく予測・更新が行われない．提案法では，パーティクルフィルタの確率密度分布の自由度を利用し，状態ベクトルの遷移を，$X_t(f)$ の推定と統一的に行う．そのため，提案法では状態関数 Q_t をモデル化する必要がない．式 (3.2) と式 (3.4) をあわせて状態空間モデルと呼ぶ．状態空間モデルは，パラメータの一部が時間とともに変化する非定常な現象の表現に適している [9]．

3.2 パーティクルフィルタのアルゴリズム

システムモデルと観測モデルに基づいて，雑音を含む観測値 $Y_t(f)$ から状態 $X_t(f)$ を推定する．時刻 t までの観測値の集合を $Y_{1:t}(f)$ とすると，問題は状態ベクトルの推定値 $\hat{X}_t(f)$ や，事後確率 $P(X_t(f)|Y_{1:t}(f))$ を推定する問題に帰着する．状態 $X_t(f)$ は，事後確率 $P(X_t(f)|Y_{1:t}(f))$ を最大化するように推定される．

パーティクルフィルタは尤度 $P(Y_t(f)|X_t(f))$ と状態遷移確率 $P(X_t(f)|X_{t-1}(f))$ を用いたベイズ理論によって推定を行う．事後確率の近似は離散値と重みを持つ多数のパーティクルを用いたモンテカルロ近似によって行われる．(式 3.5)

$$P(X_t(f)|Y_{1:t}(f)) \propto \prod_{m=1}^{M} P(Y_t(f)|X_t(f))P(X_t(f)|X_{t-1}(f)) \ , \tag{3.5}$$

ここで，M は，未知である推定対象の数である．中心極限定理により，パーティクルの数が ∞ に近づくように増えるほど，パーティクルによって近似される分布は真の事後確率に近づく [10]．

事後確率の近似における多数のパーティクルを用いたモンテカルロ近似は式 (3.6) のように表される．

$$\begin{aligned} p(X_t(f)|Y_{1:t-1}(f)) &\simeq \frac{1}{N}\sum_{i=1}^{N}\delta(X_t(f) - X_{t|t-1}^{(i)}) \\ p(X_t(f)|Y_{1:t}(f)) &\simeq \frac{1}{N}\sum_{i=1}^{N}\delta(X_t(f) - X_{t|t}^{(i)}) \end{aligned} \tag{3.6}$$

ここで，N はパーティクル数，$\delta(\cdot)$ はデルタ関数である．$X_{t|t-1}^{(i)}$ をパーティクルと呼び，状態ベクトルを表現する値とその重みで構成される．重みはパーティクルの存在する状態空間における重要度を表している．

3.3 複数ローカルピークの推定

3.3.1 ピークの存在確率の推定

複数ローカルピーク推定の第一段階として，ケプストラムのスペクトル包絡を規範とする尤度を用いたピークの存在確率推定を行う．ピークの存在確率は次節で説明するピーク抽出や，状態遷移関数の更新のために用いる．高調波の複数ローカルピークを推定するためには多次元尤度が必要となる．ここで多次元尤度は，観測の各周波数成分においてピークが存在するかどうかを判断する基準を与える．尤度 $P(Y_t(f)|X_t(f))$ はケプストラムで平滑化されたスペクトル包絡を用いて式 (3.7) のように動的に更新される．

$$P(Y_t(f)|X_t(f))=\begin{cases} 1 & ,Y_t(f)/C_x \geq 1 \\ Y_t(f)/C_x & ,Y_t(f)/C_x < 1 \end{cases} \tag{3.7}$$

C_x はケプストラムで平滑化されたスペクトル包絡である．スペクトル包絡は，ケフレンシー $x_c(k)$ にリフター $lif(\cdot)$ を適用することによって得られる低ケフレンシー部のフーリエ変換より，式 (3.8) から式 (3.10) のように得られる．

$$C_x = \mathcal{F}[lif(x_c(k))] \tag{3.8}$$

$$x_c(k) = \mathcal{F}^{-1}[log|Y_t(f)|] \tag{3.9}$$

$$lif(q) = \begin{cases} 1 & ,q \leq q_{lif} \\ 0 & ,q > q_{lif} \end{cases} \tag{3.10}$$

ただし，$q_{lif} = 2.5\text{ms}$ とし，$\mathcal{F}[\cdot]$，$\mathcal{F}^{-1}[\cdot]$ はそれぞれフーリエ変換，逆フーリエ変換である．

図 3.1 と図 3.2 は，式 (3.7) を説明した図である．図 3.1 の細線は観測値のスペクトル，太線はスペクトル包絡を表している．スペクトル包絡は，観測スペクトルのピーク部のみを得るような形状であるため，ピークの存在確率の尤もらしさを表現する尤度として用いられる．スペクトル包絡より大きい振幅を持つそれぞれ周波数帯域は，ローカルピークが存在する確率が同程度に高いため，振幅を揃える．スペクトル包絡より小さい振幅を持つ周波数帯域は，ローカルピークが存在する確率が低い部分であるが，音声スペクトルの時間変化によってローカルピークがわずかに変化することを考慮し，存在確率を完全には 0 にはしない．これを 0 から 1 の範囲で正規化することで，ピークの存在確率の推定値が得られる．推定されるピークの存在確率は図 3.2 のような形状となる．ピークの存在確率は高確率部分が平坦な形をしており，音声のスペクトル変化が緩やかであれば，ローカルピークは次の時刻に，平坦部それぞれの周波数帯域の中にピークが存在する可能性が高い．そのため，推定される状態 $P(\hat{X}_t(f))$ を式 (3.11) のように更新する．

$$\hat{X}_t(f) = P(Y_t(f)|X_t(f))P(X_t(f)|X_{t-1}(f)) \tag{3.11}$$

状態遷移確率は，推定されるローカルピークが次の時刻にピークの存在確率の高確率となる平坦部に遷移している可能性が高いことを利用して，ピークの存在確率を事後確率分布

として，式 (3.12) のように更新する．

$$P(X_t(f)|X_{t-1}(f))= \frac{\hat{X}_{t-1}(f)}{\sum_f \hat{X}_{t-1}(f)},\ f=1,2,\cdots \tag{3.12}$$

3.3.2 再サンプリング

パーティクルフィルタでは，過去のフレームからの状態の過学習によって状態の漸近的な変化に追随できない状況や，推定値の収束によって状態空間の縮退が起こる状況が発生する可能性がある．再サンプリングは，状態空間におけるパーティクルの縮退を防いだり，収束を促進させることで，パーティクルを適切な数に分散させる処理である．基本的な処理は重みの小さいパーティクルを，より重みの大きいパーティクルに併合することである．ここで，併合したパーティクルの再分布方法によって，状態空間の収束度合を調整することになる．再サンプリングの頻度を大きくすれば，状態の変化への追随精度が良くなる一方，過去のフレームからの学習で得られたパーティクルの重みが考慮されにくくなる．再サンプリングの頻度を小さくすれば，過去のフレームからの学習で得られたパーティクルの重みにより，突発的な状態の変化に影響されなくなる一方，状態の変化への追随精度が悪くなる．提案法では，収束に関して重みの小さいパーティクルを近傍パーティクルとして平滑化を行うランダムサンプリングに基づくアルゴリズム [11] を用いる．この再サンプリング手法に加え，高調波の変化が大きい場合にでも精度良くローカルピークの推定が行えるように，各状態空間に最低 1 個のパーティクルの分布を保証するような拡散手法を導入する．この方法により，パーティクルフィルタにおける各状態空間でのピークの再発生を保証でき，複数対象のどうじつ遺跡が可能となる．再サンプリングの手順は次の通りである．

- すべてのフレーム t について
  1. 一様乱数 $u_t^{(i)} \in U[0,1]$ を生成
  2. $\frac{1}{C}\sum_{l=1}^{k-1}\pi_t^{(l)} < u_t^{(i)} < \frac{1}{C}\sum_{l=1}^{k}\pi_t^{(l)}$ を満たす k を探索する．
  3. k 番目のパーティクルを新たなパーティクル列に加える

ただし，i はパーティクル番号，$\pi_t^{(i)}$ は i 番目のパーティクルの重み，N はパーティクル数，$C=\sum_{l=1}^{N}\pi_t^{(l)}$ である．

3.3.3 ローカルピーク抽出

複数の推定値を得る必要がある場合，単純に重みと状態遷移確率からは計算できない．状態遷移確率は，推定されるローカルピークが次の時刻にピークの存在確率の高確率とな

る平坦部に遷移している可能性が高いため，その範囲の 1 点を選択する事でローカルピークの抽出を行う．提案法では，ピークの存在確率の各平坦部から中央値をローカルピークとして抽出し，パラメータ化を行う．尤度によってピークの存在確率が正しく求められているのであれば，連続する各ピーク存在範囲にはただ 1 つのピークが含まれているはずなので，ピークの動きうる範囲の中心をピークとして抽出することは，ピーク位置の平均から考えると妥当な選択である．

3.3.4 アルゴリズム

提案法のアルゴリズムを以下に示す．また，提案法のアルゴリズムの図示を図 3.3 に示す．

1. 初期化
   $P(X_t(f)|X_{t-1}(f))= U(f)$
   ただし，$U(f)$ は一様分布である．

2. すべてのフレーム t について

   - すべてのパーティクル i について

     (a) ピークの存在確率を計算する．[式 (3.11)]

     (b) 状態遷移確率を更新する． [式 (3.12)]

   - 再サンプリング
     ランダムサンプリング [11] に基づくアルゴリズムを用いる．各状態空間に最低 1 個のパーティクルの分布させる．

3. ローカルピークの抽出
   ピークの存在確率の各平坦部から中央値をローカルピークとして抽出する．

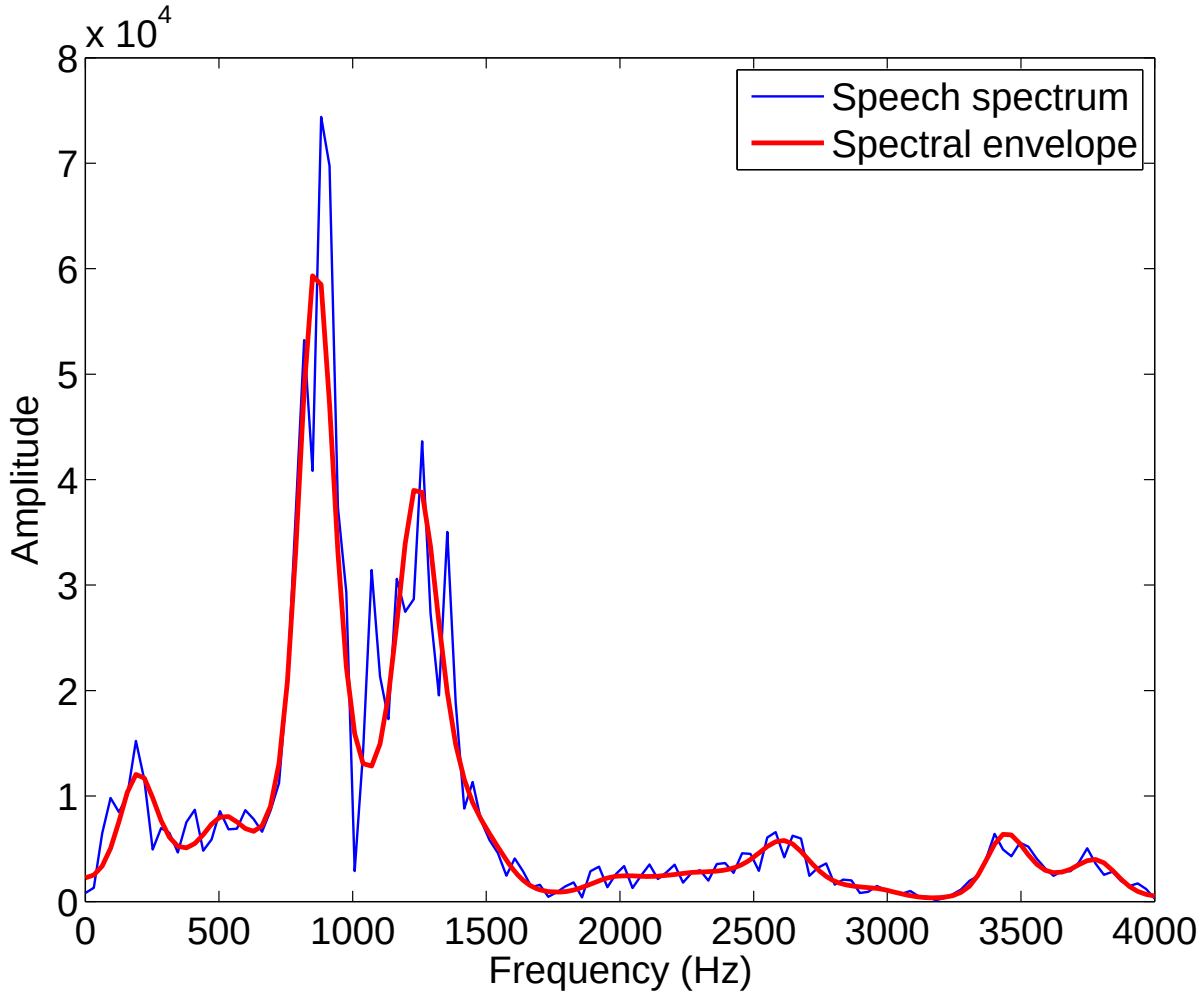


図 3.1: 観測値とスペクトル包絡を規範とした尤度の図示

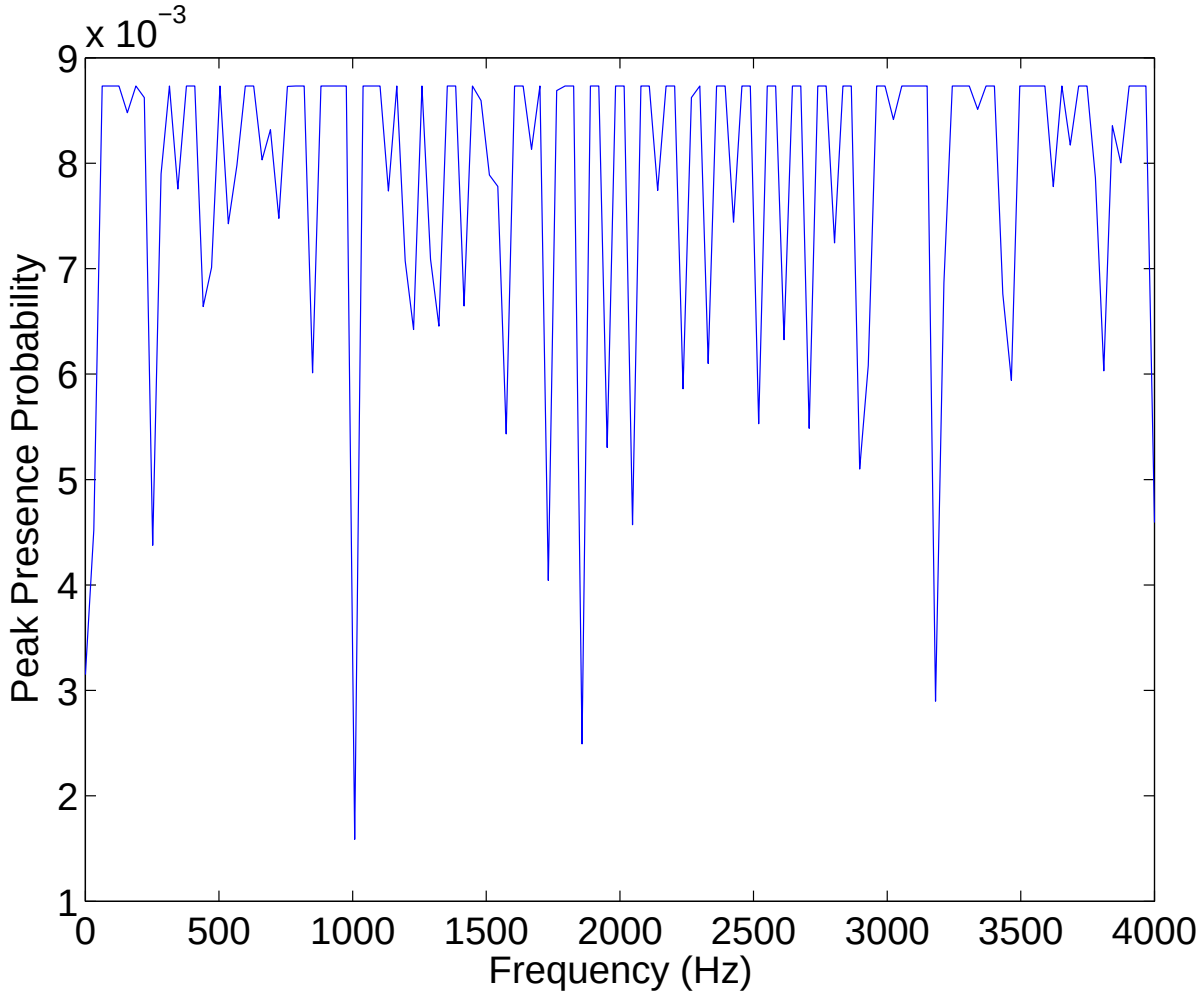


図 3.2: ピークの存在確率の推定値

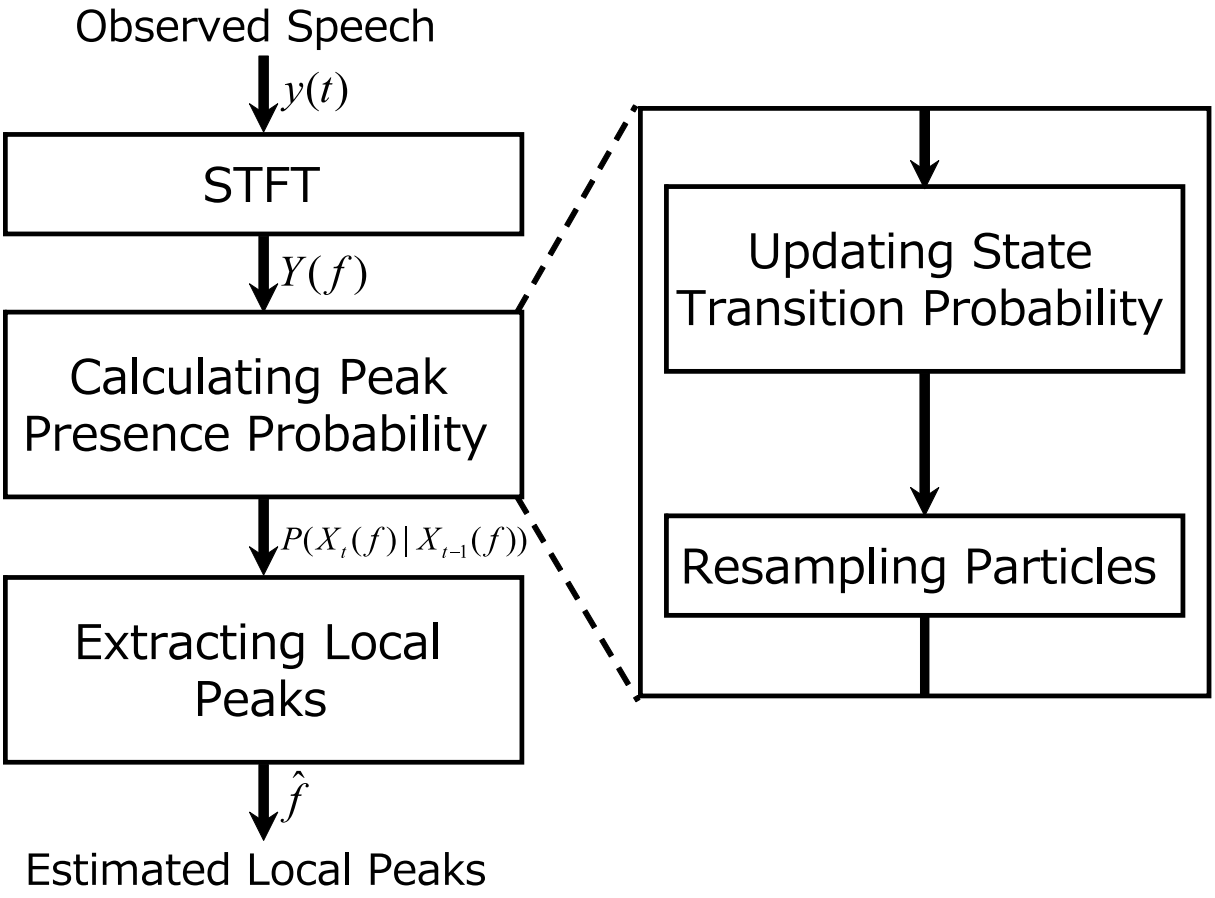


図 3.3: 提案法のアルゴリズム

第4章　複数ローカルピーク推定法の有効性

4.1　評価方法

2.2.1 節で述べた距離と過大推定ピーク数の 2 つの尺度によって評価を行う．実験は入力 SNR を $\infty, 20, 10, 0, -10$dB とした雑音付加音声に対して，提案法と二階微分を用いた手法，山登り法，瞬時周波数を用いた手法をそれぞれ適用し，推定結果の比較を行う．ここで，実験条件を 2 つ設定した．

条件 1 では，基本周波数が直線的に上昇する変化と，正弦波状に上昇下降する変化の 2 通りのデータをそれぞれ用い，振幅はすべての周波数成分で一定とする．基本周波数の直線的変化では，基本周波数が 100Hz から 200Hz まで直線的に変化し，倍音成分を 40 個持つ正解ピークデータを用いる (Cond.1-1)．基本周波数の正弦波的変化では，基本周波数が 100Hz を中心に幅 50Hz で基本周波数が正弦波状に変化し，倍音成分を 30 個持つ正解ピークデータを用いる (Cond.1-2)．それぞれの正解ピークに対する正弦波の足し合わせによって合成 [1] した波形を各手法の入力とする．信号の時間長は 485ms とする．正解と各手法によって推定されるピークとの距離と個数を比較する．

条件 2 では，基本周波数が直線的に上昇する変化を用い，振幅は実音声の振幅を用いる．基本周波数の変化は，100Hz から 200Hz まで基本周波数が直線的に変化する正解ピークデータを用いる．振幅設定に使用する音声は，ATR 研究用音声データベースの A セットに収録されている男性話者 MHT，MAU と女性話者 FYN，FKN 発話の短母音/a/と連続母音/aoi/である．それぞれの正解ピークに対する正弦波の足し合わせによって合成 [1] した波形を各手法の入力とする．正解と各手法によって推定されるピークとの距離と個数を比較する．入力信号の時間長は，/a/を用いる場合は 185ms，/aoi/を用いる場合には 485ms とした．振幅情報，基本周波数変化，高調波の倍音成分の個数の各条件が異なる 8 種類の試行を行う (Cond.2-1〜Cond.2-8)．基本周波数の直線的変化に対する正解ピークを図 4.1 に，正弦波的変化に対する正解ピークを図 4.2 に示す．実験 1 と実験 2 の各実験条件を表 4.1 に示す．

雑音にはピンク雑音，白色雑音，狭帯域雑音を用いる．狭帯域雑音の中心周波数を 1000Hz，帯域幅を 1000Hz，時間幅を 1 から 4 フレームとする．倍音がナイキスト周波数を超えた場合には，ナイキスト周波数以上の周波数を持つピークは無効とする．基本周波数が直線的な変化をする正解ピークを図 4.1 に，正弦波状に変化をする正解ピークを図 4.2 に示す．サンプリング周波数は 8kHz，フレームの切り出しには Hamming 窓を用い，

表 4.1: 実験条件

	speaker	*content*	*F0 variation*	*number of harmonics*
Cond.1-1	-	-	Straight	40
Cond.1-2	-	-	Sinusoid	30
Cond.2-1	MHT	/a/	Straight	40
Cond.2-2	FYN	/a/	Straight	30
Cond.2-3	MAU	/a/	Straight	40
Cond.2-4	FKN	/a/	Straight	30
Cond.2-5	MHT	/aoi/	Straight	40
Cond.2-6	FYN	/aoi/	Straight	30
Cond.2-7	MAU	/aoi/	Straight	40
Cond.2-8	FKN	/aoi/	Straight	30

窓長は 32ms，スライド幅は窓長の 0.4，FFT フレーム長は 512 とする．ピーク同士のピークを分離できる周波数分解能を得るために，窓によって切り出された信号の両端にゼロ値を挿入し，FFT フレーム長と同じにすることで，スペクトルをオーバーサンプリングすることと等価な処理を行う．また，パーティクル数は 40000 とする．フィルタ特性はあらかじめ除去し，音源特性を推定対象とする．

評価は，推定ピーク数から正解ピーク数を減じたピークの過大推定数と，推定ピーク位置と正解ピーク位置の距離によって比較を行う．ここで，ピーク間の距離を次のように定義する．正解ピークに対応する推定ピークとの周波数領域の差をとる，つまり正解ピーク位置に最も近い推定ピーク位置との差を取る．1 フレーム内に存在するピーク全体の差の平均を距離とする．ただし，差が基本周波数の値を超えた場合は対応するピークがないと見なし，基本周波数の値に相当する距離を加算する．推定ピーク数が正解ピーク数より大きい場合は過大推定数だけ基本周波数の値を加算する．

4.1.1 パラメータ設定

二階微分を用いたローカルピーク推定法 (SOD)

2.2.2 節と同様である．

山登り法を用いたローカルピーク推定法 (HC)

2.2.2 節と同様である．

瞬時周波数を用いたローカルピーク推定法 (IF)

2.2.2 節と同様である．

提案法 (PF)

提案法では，尤度の振幅方向での補正度合，パーティクルフィルタにおける再サンプリングの頻度をパラメータとする．尤度の確率方向での補正度合は，音声スペクトルからピークであるとみなす振幅範囲を，振幅値の平均を基準に調整する．再サンプリングの頻度が大きければ，学習による過去の情報が利用されにくくなり，頻度が小さければ過去の情報が利用されやすくなる．過去の情報を利用すればするほど過去のピーク位置近辺のピークの存在確率が大きくなり雑音に対する頑健性が高くなり，ローカルピークへの追随性能が低下する．

4.1.2 パラメータ設定方針

ピンク雑音，白色雑音に関して，各 SNR における各手法の最適なパラメータを設定したものを用いる．

狭帯域雑音に関して，雑音区間以外はクリーンであるため，クリーン音声に対する複数の最適パラメータを狭帯域雑音環境で適用した際の最適値を手動で設定する．

各手法に適用したパラメータを表 4.2 に示す．

表 4.2: 提案法と従来法のパラメータ（左から：提案法，二階微分法，山登り法，瞬時振幅法）

likelihood rate	*resample rate*	*slope rate*	*height*	*bandwidth*	*interval*	*tc*
1.02	0.5	10	0.01	190	100	0.05

4.2 評価結果

Cond.1-1 における入力 SNR 0dB, 狭帯域雑音幅 4 フレームにおける推定ピークの時間-周波数領域の関係を図 4.3 に示す．正解ピークの図 4.1 と比較すると，狭帯域雑音付加部分である 5 フレーム目から 8 フレーム目で顕著な違いが見てとれる．PF と IF は比較的正解ピークの概形を保持しているが，SOD と HC は特に雑音付加区間での雑音による影響を受けていることがわかる．定量的な評価を行うために，あらかじめ推定結果に対する正解となるピークを用意し，推定ピークと正解ピークの個数および推定ピークと正解ピークの距離の 2 つの評価尺度によって比較を行った．

実験データを比較した結果,「男女の差異」以外の,「正解ピークの基本周波数の変化の差異」,「話者の差異」,「/a/と/aoi/の差異」による実験結果の違いはほとんど見られなかった．そこで，差異のない結果については平均を行う．実験 1 の結果を平均したもの，つまり Cond.1-1 と Cond.1-2 の結果の平均をとったものを図 4.4 から図 4.9 に示す．図 4.4 はピンク雑音を付加した条件での結果，図 4.5 は白色雑音を付加した条件での結果，図 4.6 から図 4.9 はそれぞれ狭帯域雑音の長さを 1 フレームから 4 フレームに変えた時の結果である．同様に，男女の差異のみに注目して，Cond.2-1, Cond.2-3, Cond.2-5, Cond.2-7 の結果の平均をとったものを図 4.10 から図 4.15 に，Cond.2-2, Cond.2-4, Cond.2-6, Cond.2-8 の結果の平均をとったものを図 4.16 から図 4.21 に示す．図 4.10 はピンク雑音を付加した条件での結果，図 4.11 は白色雑音を付加した条件での結果，図 4.12 から図 4.15 はそれぞれ狭帯域雑音の長さを 1 フレームから 4 フレームに変えた時の結果である．図 4.16 はピンク雑音を付加した条件での結果，図 4.17 は白色雑音を付加した条件での結果，図 4.18 から図 4.21 はそれぞれ狭帯域雑音の長さを 1 フレームから 4 フレームに変えた時の結果である．横軸は入力 SNR を示しており，$-10, 0, 10, 20, \infty$dB の順に並んでいる．図はいずれも縦軸が 0 に近いほど正確にローカルピークの推定がなされたといえる．

4.3 考察

4.3.1 条件 1

各手法における過大推定ピーク数の違いについて，SOD は入力 SNR が ∞dB を除くすべての入力 SNR で推定精度が悪かった．SOD は狭帯域雑音の場合でも同様の傾向を示している．HC は，狭帯域雑音を用いた場合には提案法に匹敵する推定精度が得られた．しかし，ピンク雑音と白色雑音を付加した条件でのローカルピーク推定精度は悪かった．SOD および HC は，クリーン音声時に精度良く推定できるようなパラメータを用いたため，雑音区間で雑音の影響を大きく受けたためと考えられる．特に実験では入力が未知である状況を想定しているため，この条件に最適化したパラメータを用いると別の条件で精度が悪くなる可能性がある．IF はピンク雑音や白色雑音を付加した条件で，推定ピーク数で SOD と HC の精度を上回った．そして狭帯域雑音を用いた場合，IF はピークを全体的に過小推定する傾向にある．

各手法における距離の違いについて，SOD は入力 SNR が ∞dB を除くすべての入力 SNR で推定精度が悪かった．特に入力 SNR が -10dB のときは，∞dB 時の数倍もの距離になった．SOD は狭帯域雑音の場合でも同様の傾向を示している．HC は，推定ピーク数と同様の傾向で，狭帯域雑音を用いた場合には提案法に匹敵する推定精度が得られた．そして，ピンク雑音と白色雑音を付加した条件でのローカルピーク推定精度は悪かった．SOD および HC は，推定ピーク数と同様の理由で，雑音区間で雑音の影響を大きく受けたため低 SNR における推定精度が低かったと考えられる．IF はピンク雑音や白色雑音を付加した条件では，距離で SOD と HC の精度を上回った．狭帯域雑音を用いた場合，IF

は入力 SNR によらず距離で安定した推定精度が得られた．IF は，直前のフレームで推定されたローカルピークの周波数位置を元に現在のフレームの推定ピーク位置を推定するために定常雑音の影響を少なくできたと考えられる．IF の更新の度合いは時定数によって更新の度合いが制御可能であるので，適切なパラメータを設定できれば精度良くローカルピークの追跡が可能となる．

一方，PF はピンク雑音や白色雑音を付加した条件では，入力 SNR が正である場合に SOD,HC,IF よりも距離は短く，過大推定ピーク数は少ないという良い結果が得られた．また，狭帯域雑音を付加した条件では，PF は負の入力 SNR であっても SOD,HC,IF より距離は短く，過大推定ピーク数は少ないという良い結果が得られた．PF では，雑音に埋もれなかった僅かなピークの突出であっても尤度によってピークらしいと判断されており，さらにピークの学習によって振幅値の小さなピークであっても高い確率でピークが推定できたために精度が良かったと考えられる．また，PF は狭帯域雑音環境下では負の SNR であっても SOD,HC,IF よりも精度良くローカルピーク推定が可能であった．狭帯域雑音を用いた条件では，初期フレームは雑音を付加していないクリーンな音声であるという仮定は一般的ではないが，前に述べたように PF はピンク雑音と白色雑音に対する正の SNR に対して頑健性を持つため，この初期フレームへのクリーン音声の対応付けが有効である．ピンク雑音や白色雑音を付加した環境下でも正の SNR でなら精度良くローカルピーク推定が行えるといえる．

実験 1 の平均値の結果から，振幅が一定の合成波形を用いた理想的な環境において提案法は従来法を上回る推定精度が得られることがわかった．これは提案法において，雑音のレベルがローカルピークを完全に上回らないような場合であれば，わずかなローカルピークの突出を手がかりに過去のフレームからのピーク位置の学習を推定できているためと考えられる．逆に，入力 SNR が負である場合には雑音によって生み出された本来ローカルピークとなり得ない部分を誤学習しているため従来法より結果が悪くなったと考えられる．また，非定常雑音においても入力 SNR が正である場合には精度良くローカルピークの推定が行えることがわかる．

4.3.2 条件 2

実験 2 の男性平均値の結果，SOD, HC, IF については実験 1 と同様の傾向が見られた．PF はピンク雑音や白色雑音を付加した条件では，実音声の振幅を持つ合成波形においても実験 1 と同様の傾向が見られた．狭帯域雑音を付加した条件では，雑音長が 1 フレームでは負の入力 SNR であっても，SOD, HC, IF を上回る結果であったが，$2, 3, 4$ フレーム長では正の SNR で SOD,HC,IF を上回る結果となった．実音声の振幅を用いたことによって雑音に埋もれた極小ローカルピークを推定するために重みのしきい値を上げる必要があるが，しきい値を上げることにより，過剰にローカルピークを推定したり，ピークの存在確率で高確率部が狭まるためにローカルピークの時間変化に追随しにくくなるのが精度低下の原因と考えられる．

実験 2 の女性平均値の結果，ピンク雑音，白色雑音の双方で PF, SOD, HC, IF すべてにおいて，実験 1 と同様の傾向が見られた．しかし距離に関しては，女性の距離は男性のものに比べおよそ 2 倍の距離を持っている．一般に女性話者の基本周波数は男性話者のものより基本周波数がおよそ 2 倍高いため，推定誤差が男性話者の場合よりも大きくなったためと考えられる実験 2 の女性平均値と実験 2 の男性平均値を比較すると，実験結果からも PF の距離がおよそ 2 倍になっていることからも見て取れる．

狭帯域雑音の帯域幅について，本実験で設定した条件で最適化した場合における頑健性は数フレーム程度であったが，再サンプリングで過去の情報をできるだけ多く利用するようにパラメータ $resample\ rate$ を調整することで，さらに長い区間での雑音に対する頑健性が得られる．しかし，過去のフレームからの学習が多ければ多いほど，現在のフレームからの情報がピークの存在確率の更新に反映されにくくなるため，ローカルピークの追随が困難になると考えられる．

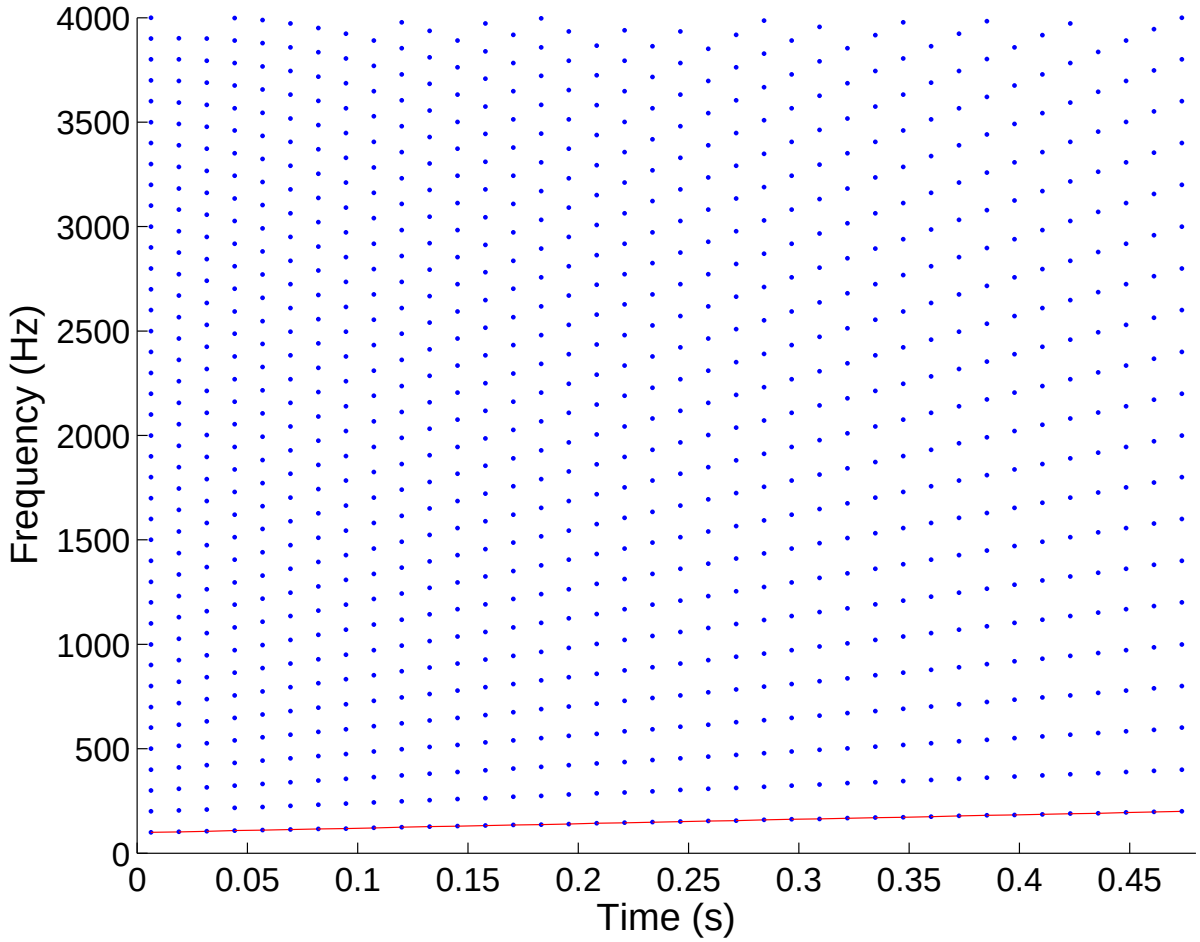


図 4.1: 時間長 485ms, 基本周波数直線的変化時の正解ピーク

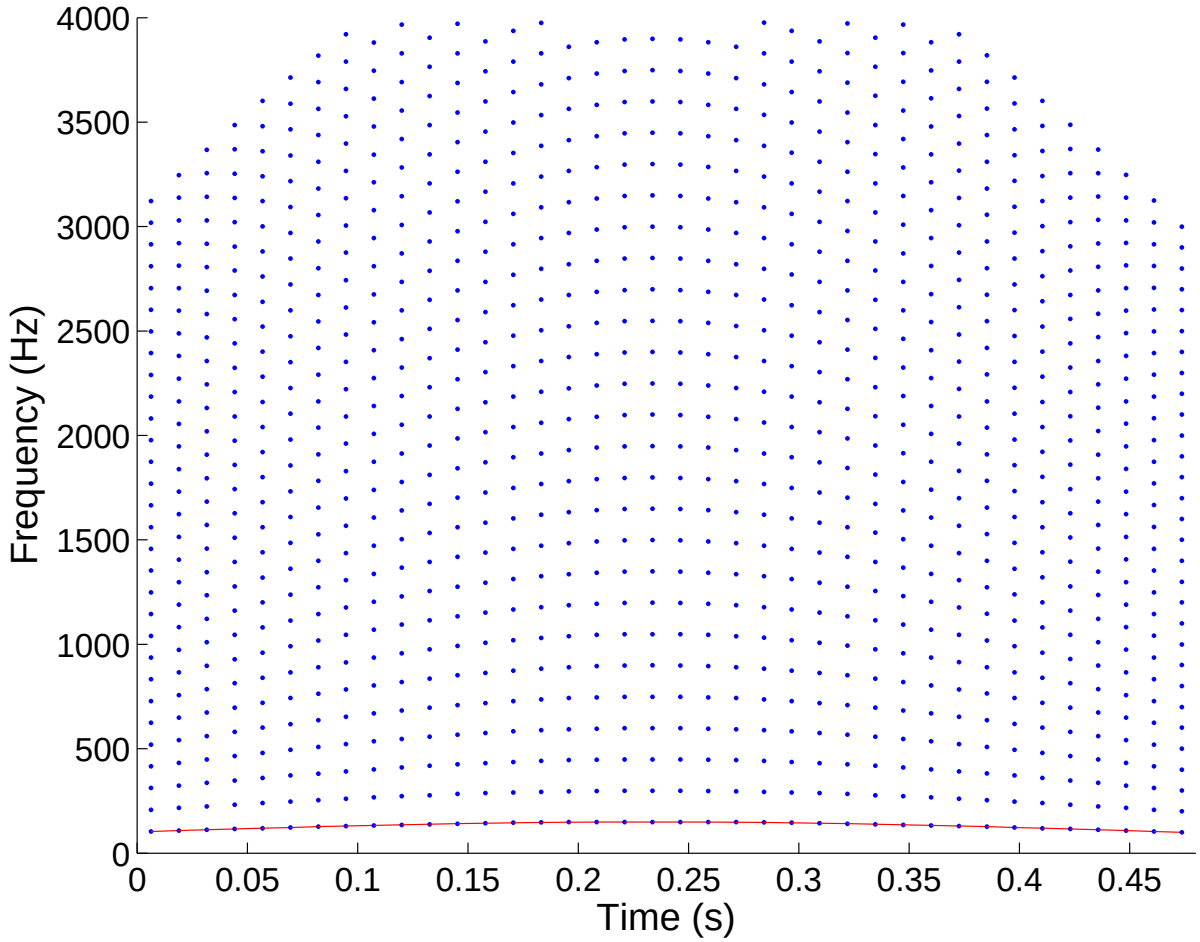
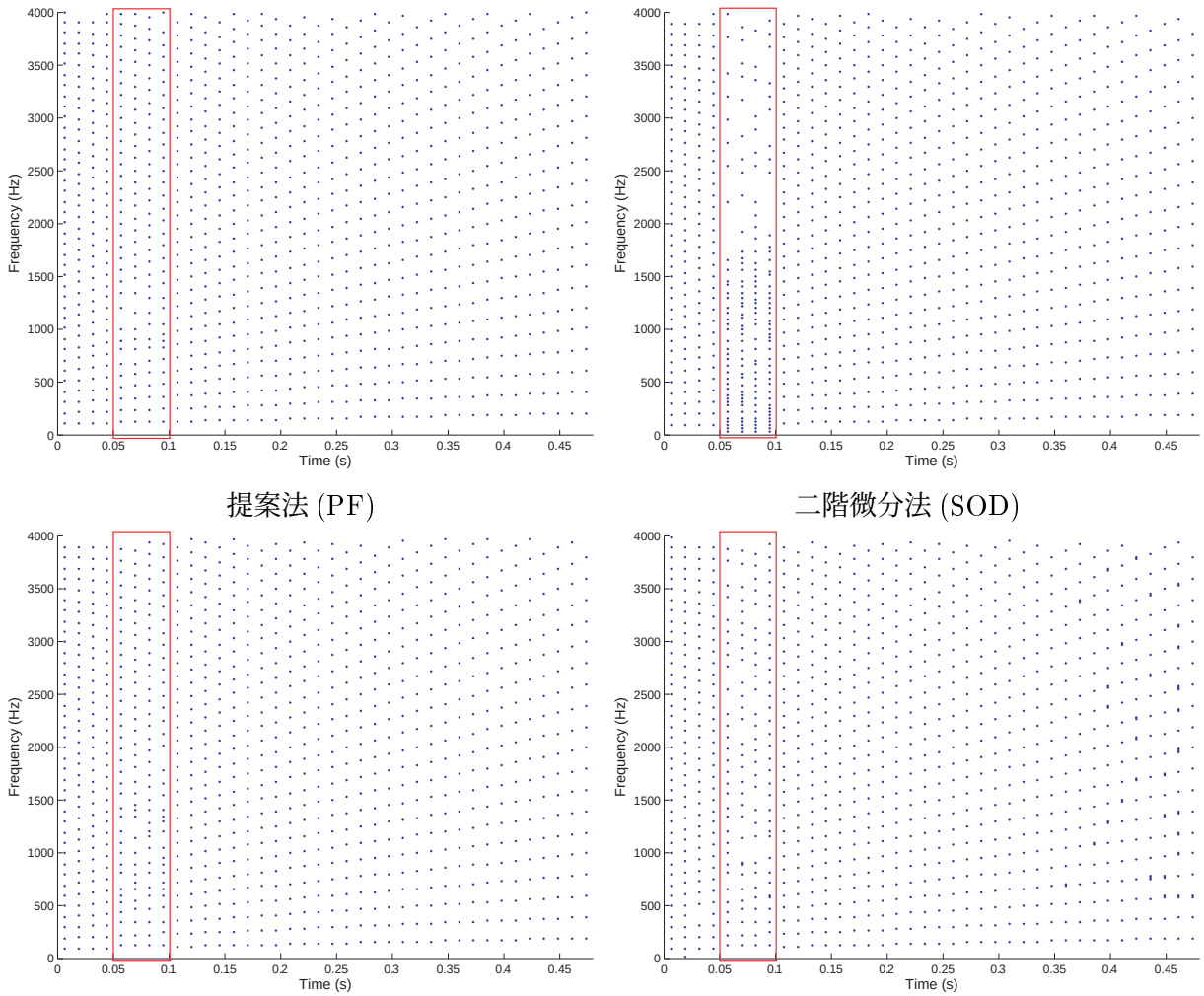


図 4.2: 時間長 485ms, 基本周波数正弦波的変化時の正解ピーク



提案法 (PF)　　二階微分法 (SOD)

山登り法 (HC)　　瞬時周波数法 (IF)

図 4.3: Cond.1-1, 入力 SNR0dB, 狭帯域雑音幅 4 フレームでの推定ピークのプロット

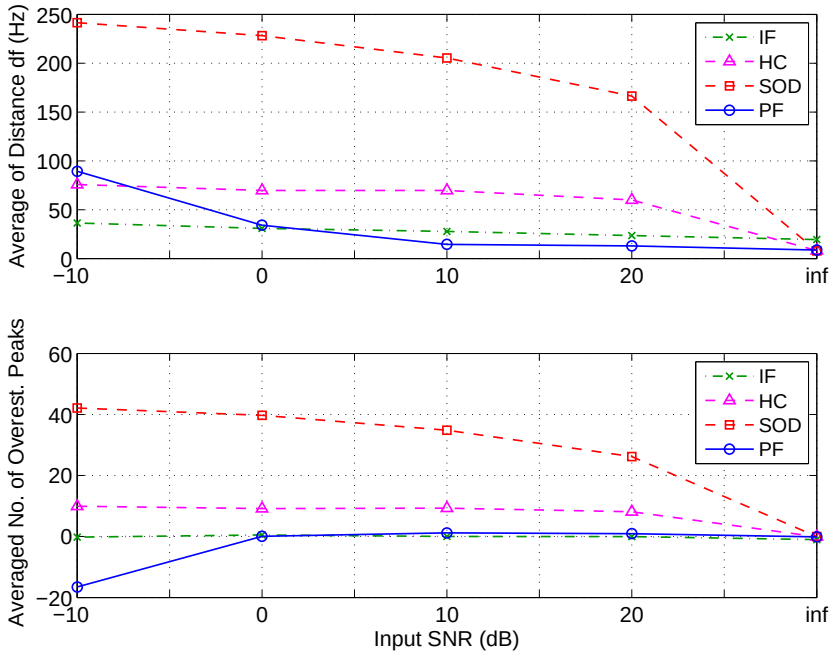


図 4.4: Cond.1-1,Cond.1-2 の平均値，ピンク雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

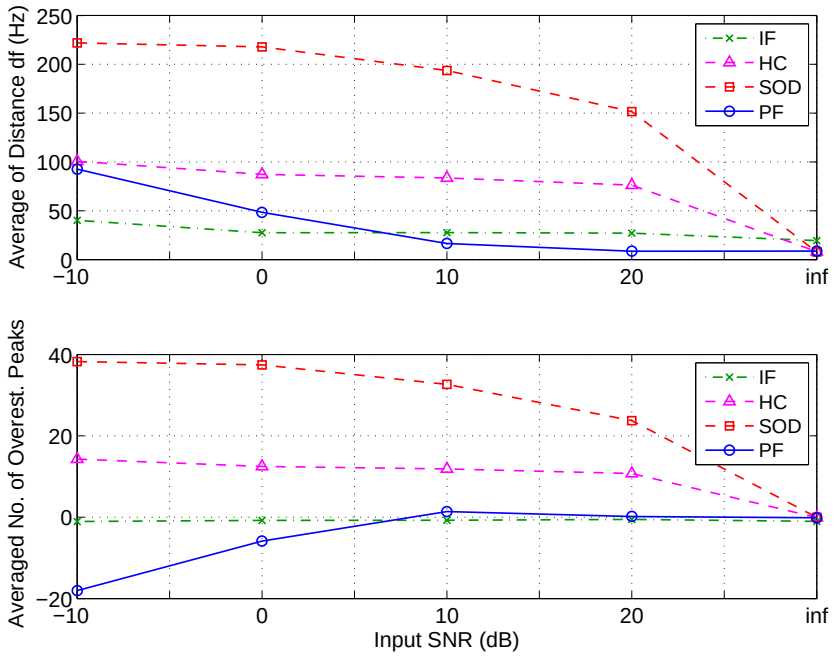


図 4.5: Cond.1-1,Cond.1-2 の平均値，白色雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

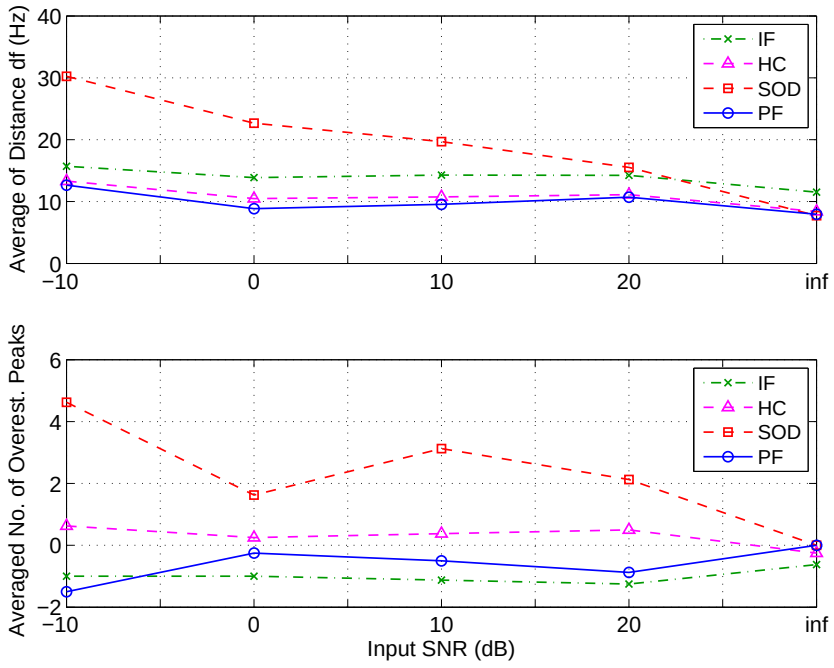


図 4.6: Cond.1-1,Cond.1-2 の平均値，時間幅 1 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

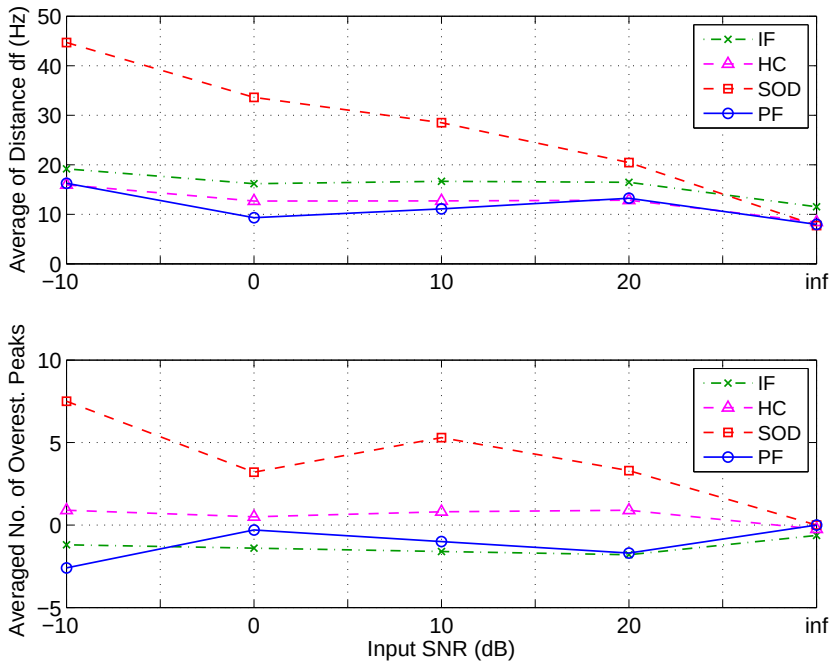


図 4.7: Cond.1-1,Cond.1-2 の平均値，時間幅 2 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

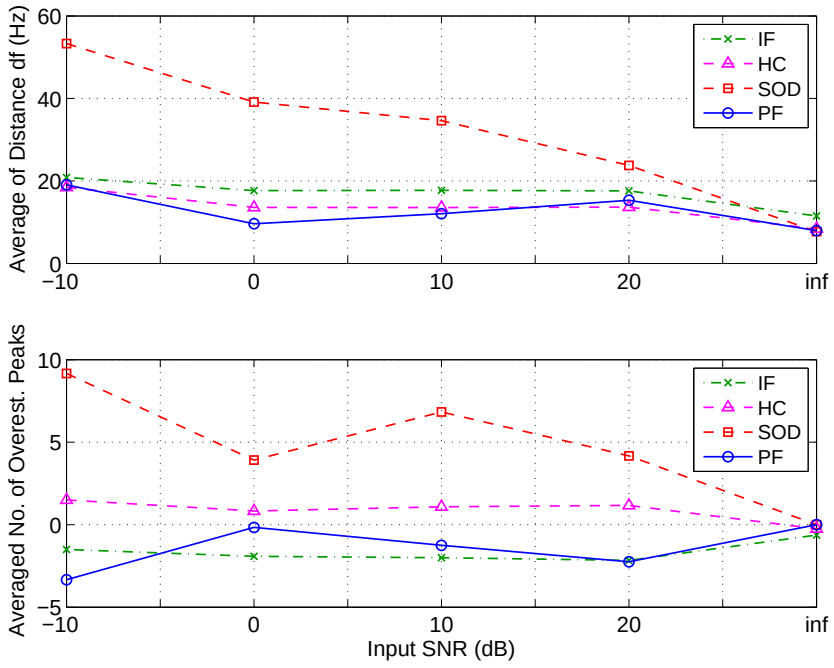


図 4.8: Cond.1-1,Cond.1-2 の平均値，時間幅 3 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

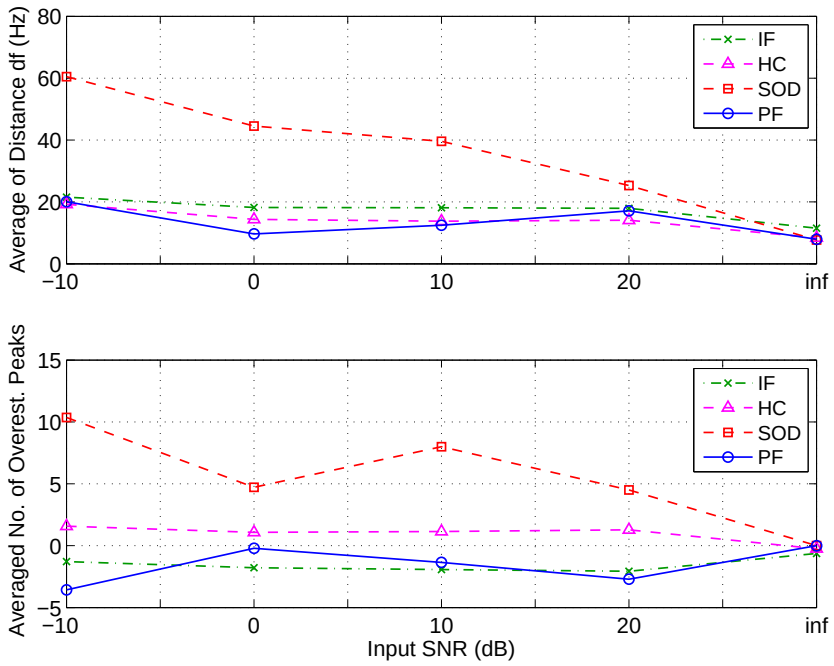


図 4.9: Cond.1-1,Cond.1-2 の平均値，時間幅 4 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

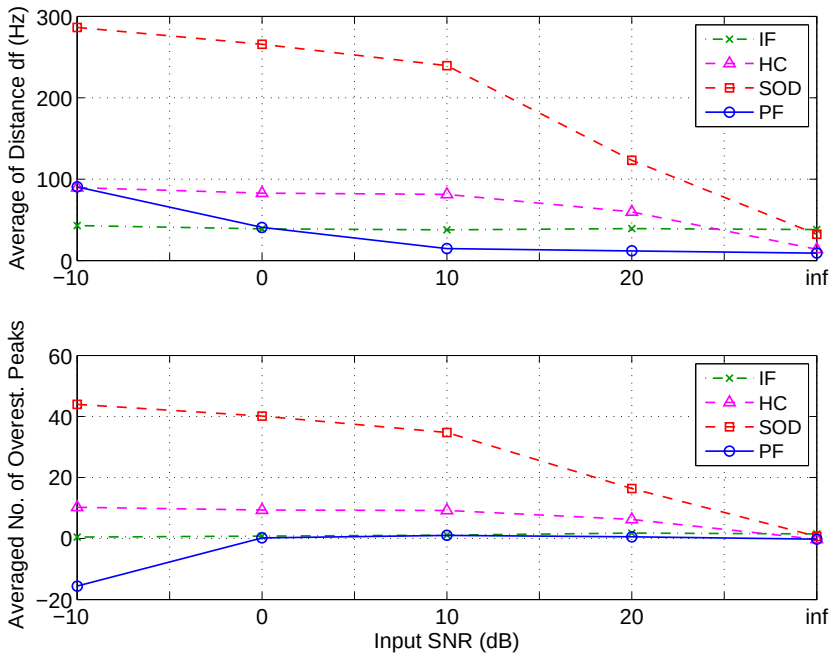


図 4.10: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，ピンク雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

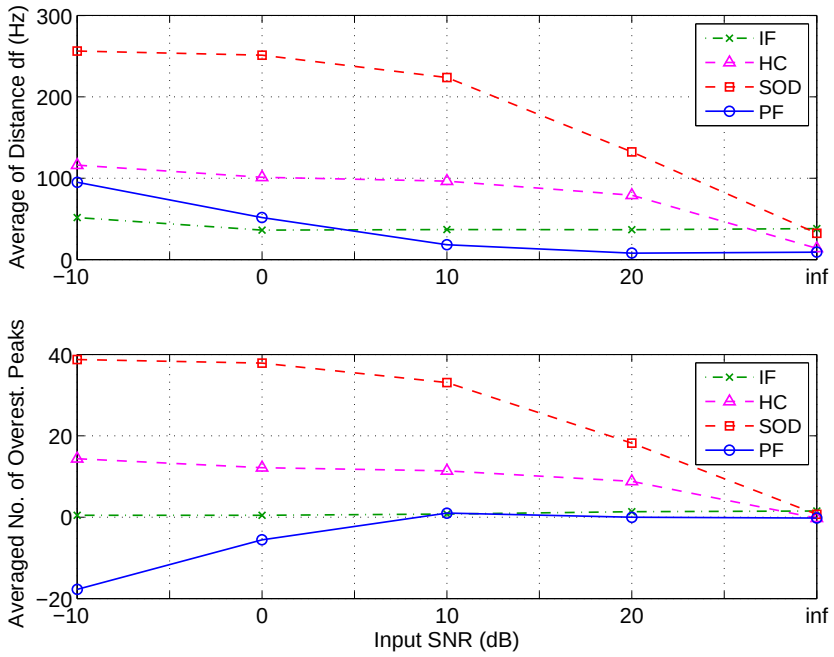


図 4.11: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，白色雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

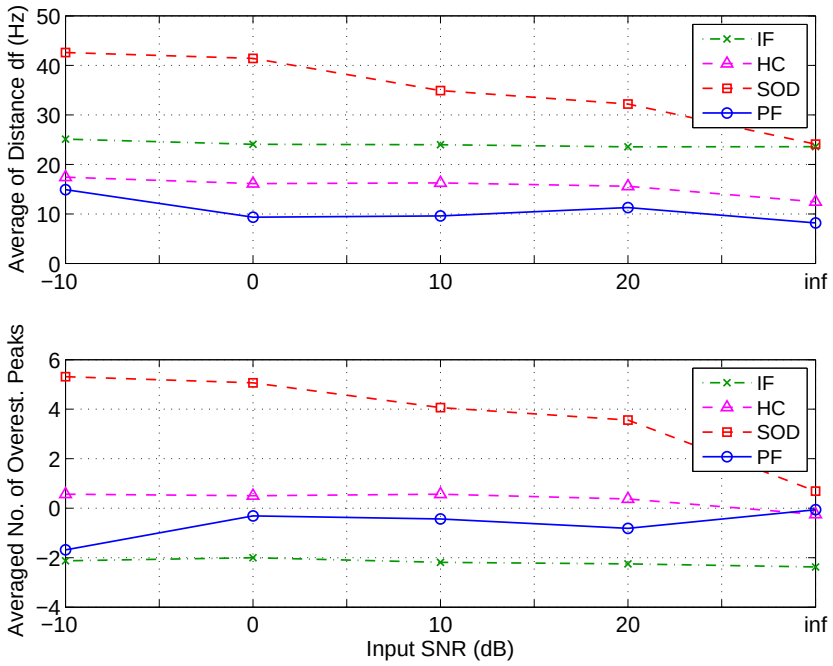


図 4.12: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，時間幅 1 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

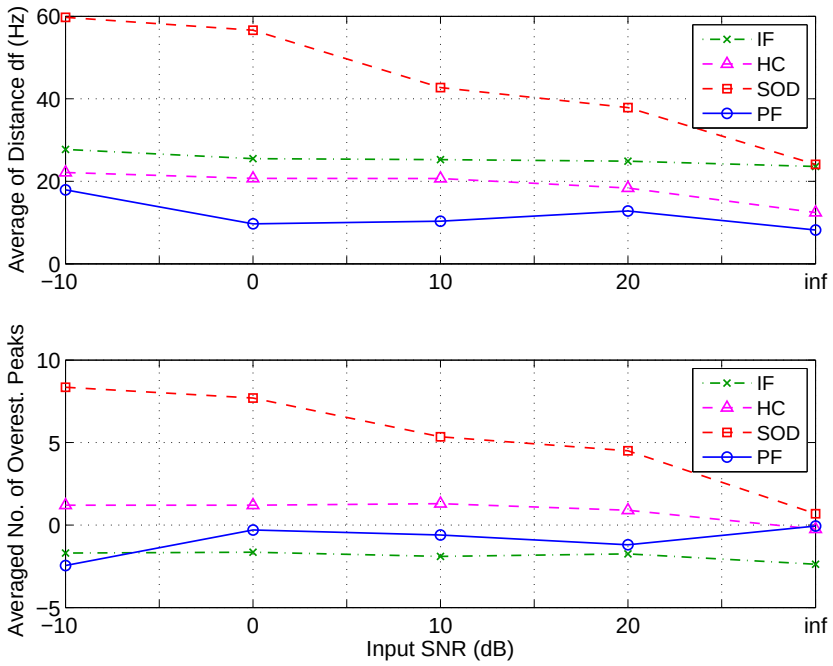


図 4.13: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，時間幅 2 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

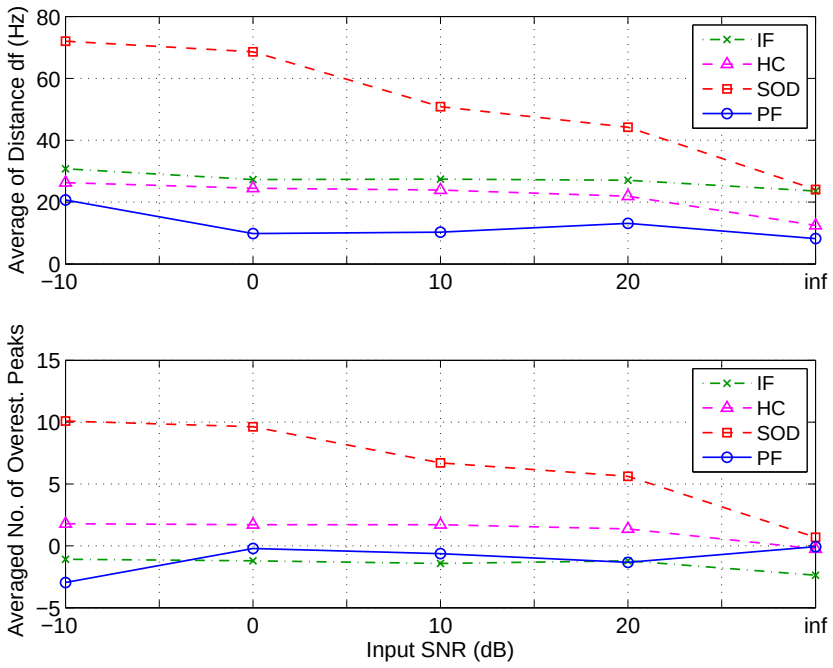


図 4.14: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，時間幅 3 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

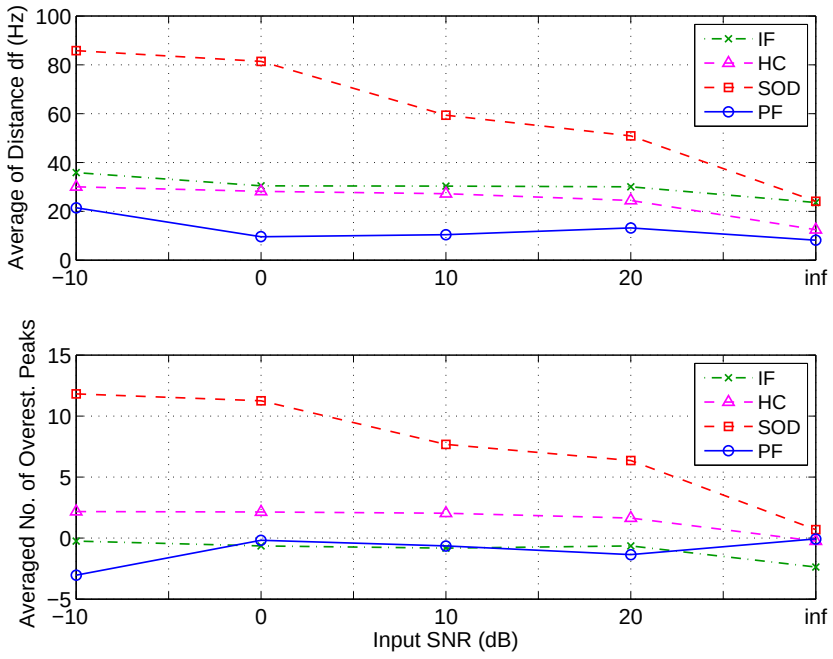


図 4.15: Cond.2-1,Cond.2-3,Cond.2-5,Cond.2-7 の平均値，時間幅 4 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

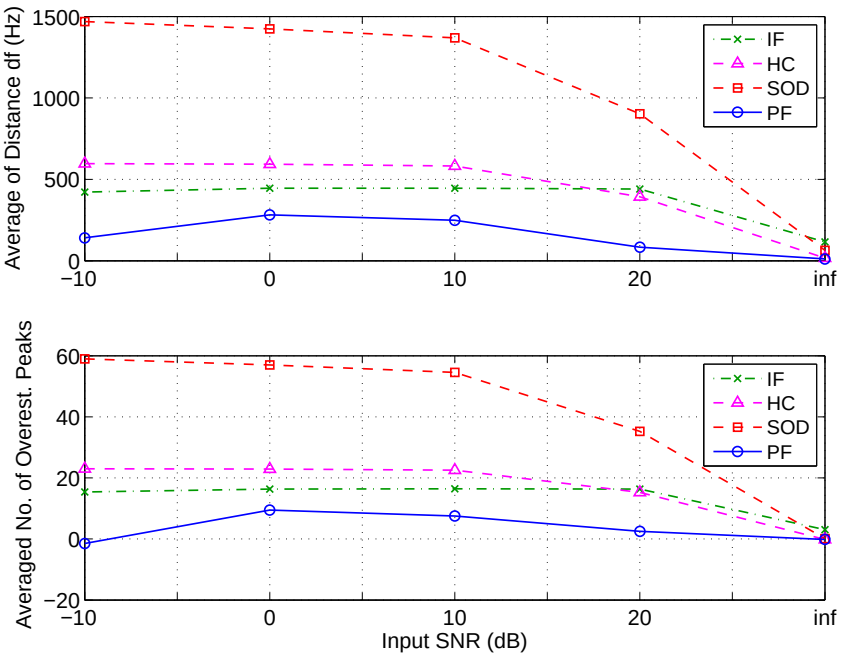


図 4.16: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，ピンク雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

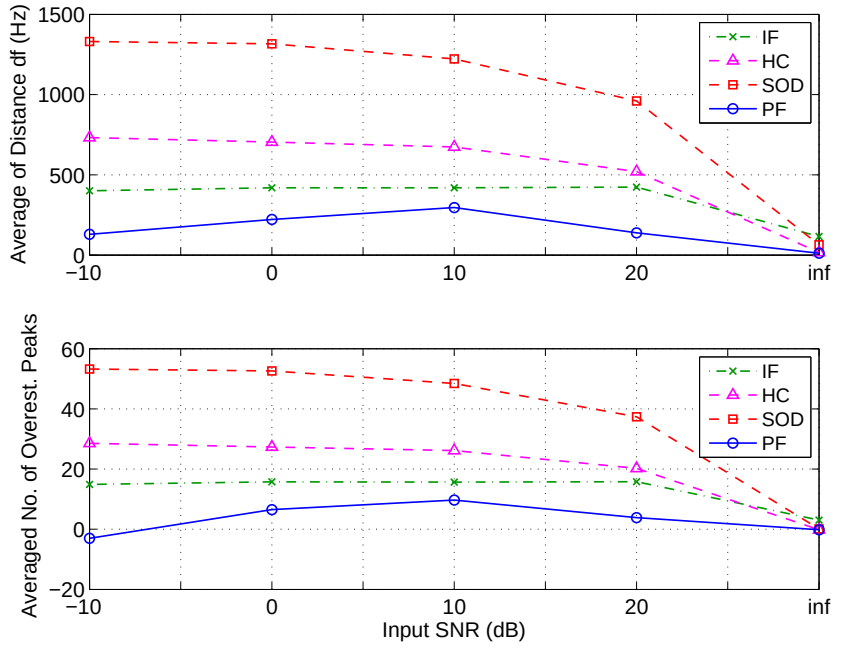


図 4.17: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，白色雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

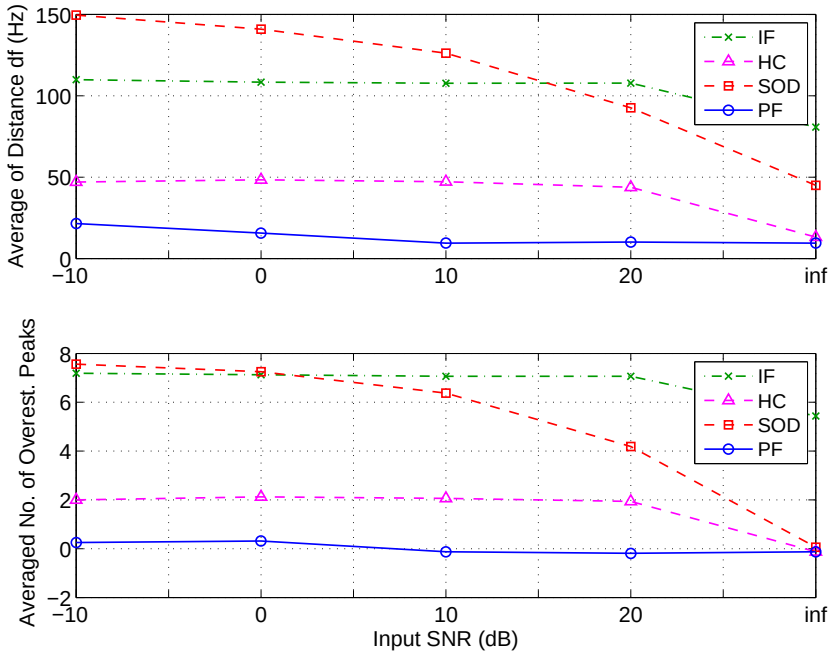


図 4.18: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，時間幅 1 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

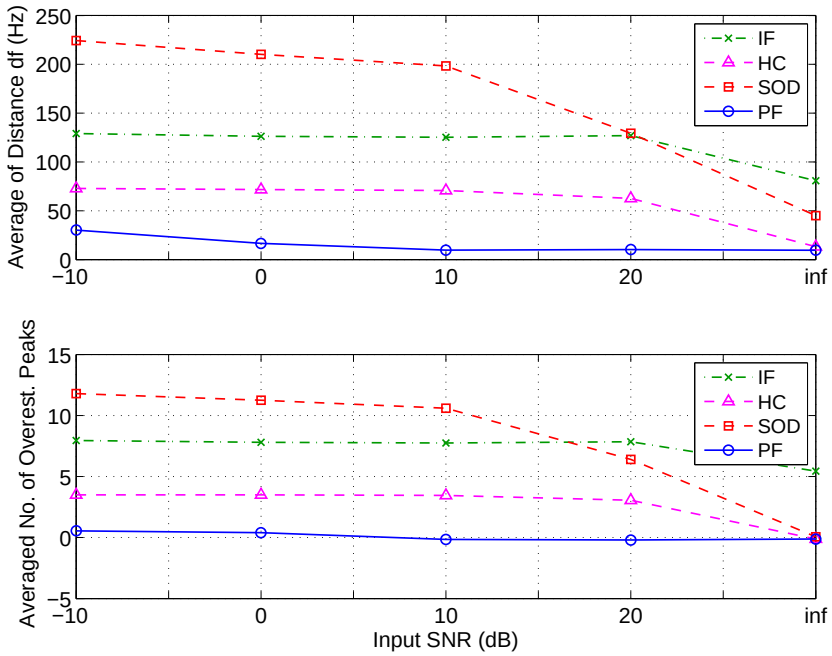


図 4.19: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，時間幅 2 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

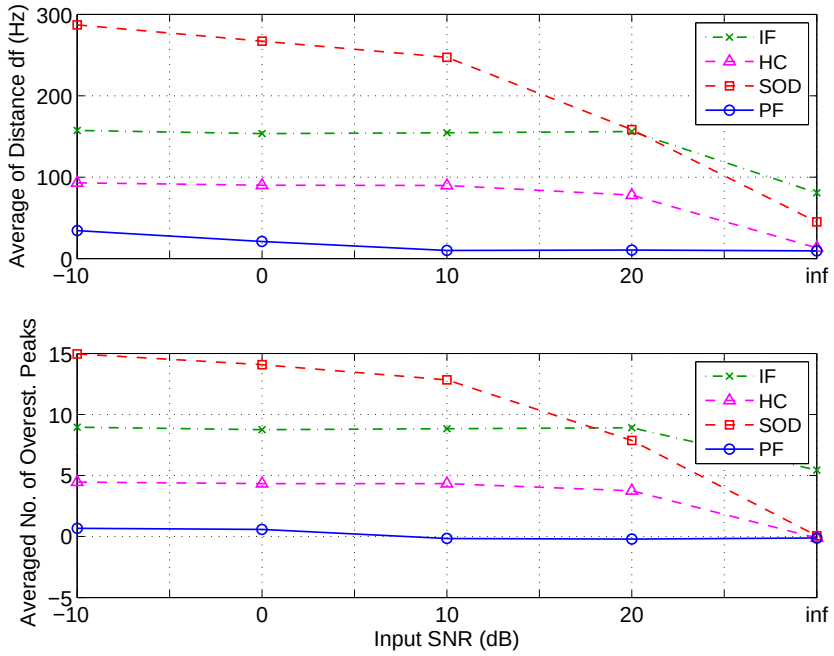


図 4.20: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，時間幅 3 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

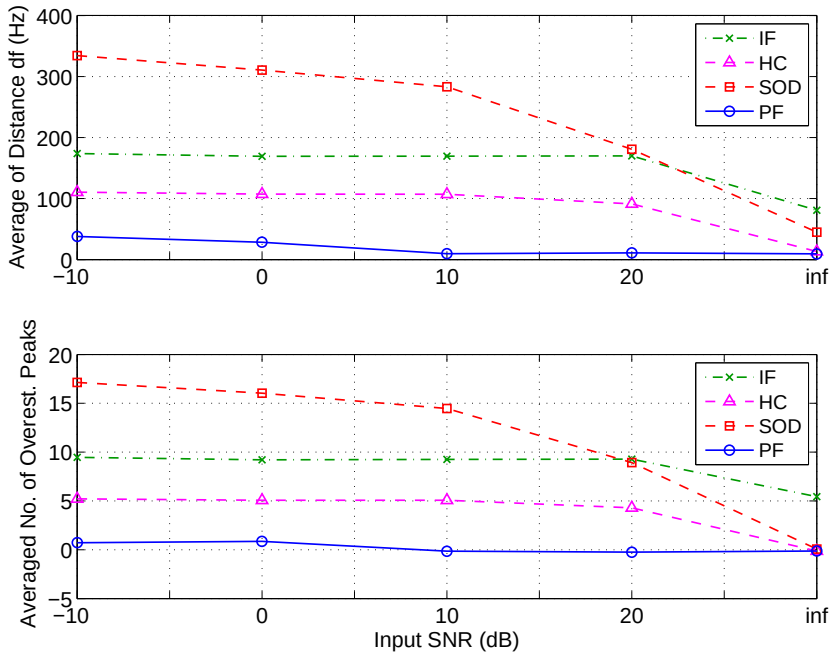


図 4.21: Cond.2-2,Cond.2-4,Cond.2-6,Cond.2-8 の平均値，時間幅 4 フレームの狭帯域雑音を用いた結果 (上段：平均距離，下段：平均過大推定ピーク数)

第5章　結論

5.1　本論文で明らかにしたこと

高調波は様々な音声信号処理における重要な役割を担っている．音声スペクトルのローカルピークは，高調波成分の周波数を与える基本的な特徴量である．そのため，様々なローカルピーク推定法が提案されてきたが，非定常雑音に頑健なローカルピーク推定法は報告されておらず，また非常に困難であると考えられている．本論文の目的は，雑音環境下での複数ローカルピーク推定法を提案することである．

そのために，パーティクルフィルタを用いた提案法によって過去のフレームで推定されるローカルピークの学習によって，現在のフレームのローカルピークを推定する．パーティクルフィルタはシステムの状態推定に際して，事後分布を正確に近似することで，システムの状態を推定する方法である．提案法では，ローカルピーク推定に際して，音声スペクトル上の複数のローカルピークを同時に推定する尤度を導入する．この尤度を用いることで複数のローカルピークの存在確率を同時に推定可能であるため，提案法は個数が未知である複数のローカルピーク推定に適用可能である．さらに，状態推定に尤度と事後分布のみを用いているため，ローカルピークの動きをモデル化を必要としない．入力される音声によって，ローカルピークの個数は異なるそのため，ローカルピークの個数を仮定することなく過去のローカルピーク情報から複数のローカルピーク推定が可能であるのは大きな特徴である．

雑音環境下での音声のローカルピークを精度良く推定できれば，実環境での様々な音声信号処理に応用が可能である．まず，従来のローカルピーク推定法が雑音環境下で機能するかどうかを見極めるために，従来のローカルピーク推定法を概説し，その中で3つの手法について耐雑音性の評価を行った．評価は，ローカルピークの推定個数の差，正解ピークと推定ピーク間の距離の差の2つの尺度で行った．その結果，従来の手法はいずれも雑音の影響を大きく受け，特に非定常雑音に対する推定精度は高いとは言えなかった．この評価結果から，従来法は雑音の混入が時間的・周波数的に予測困難であるような非定常雑音に対しての頑健性が低いことが分かった．従来法では，現在のフレームの情報のみを使ってローカルピークの推定を行うため，従来法の推定精度は現在のフレームに存在する雑音に大きく影響されるのが原因であると考えられる．

そこで，音声スペクトルの変化は緩やかであるとし，過去のフレームで推定されたローカルピークの周波数位置を学習することを考える．提案法では，パーティクルフィルタを用いた非定常雑音に頑健な複数ローカルピークの推定を行う．提案法は大きく分けて2つ

の手順で構成される．第 1 の手順は，ケプストラムから得られるスペクトル包絡を尤度とする，ピークの存在確率推定である．この方法で得られるピークの存在確率は尤度を用いて動的に更新される．提案法は音声スペクトルのローカルピークの変化に関するモデルを用いることなく，ピークの存在確率を推定可能である点が特徴である．また，個数が未知である各倍音がわずかなゆらぎを持つような独立した動きを持つ場合であってもパーティクルフィルタによってピークの存在確率を細かく表現することによって，ピークの存在確率を推定可能とする．第 2 の手順は，ピークの存在確率から得られる，ローカルピークとなりうる候補からローカルピークを抽出する手法である．よって，従来法の問題点である，過去のフレームで推定されたローカルピークの学習が生かされていないという点が解決した手法が実現できる．

提案法が雑音環境下で精度良くローカルピークの推定が行われているかを検証するために実験を行った．提案法は，非定常雑音環境下においても従来法を上回る推定精度であった．よって，提案法は非定常雑音に対して頑健な特徴であることがわかった．ただし，従来法に比べ，非定常雑音に対して頑健ではあったが，大幅な改善には至らなかった．

5.2 今後の課題

音声スペクトルのローカルピークが明瞭で，その動きが緩やかな変化である場合，提案法によって精度良くローカルピークが推定できることがわかった．会話中の音声においては，多くの子音を含むため，上記の条件を満たす音声ばかりで構成されるとは限らない．そこで，音声スペクトルのローカルピークが不明瞭であり，またその動きが急激な変化であるような音声に対するスペクトルのローカルピーク推定を考慮する必要がある．また，実環境での適用を考えると，雑音ばかりでなく残響の影響も推定精度に関係すると考えられるため，残響環境に対しても適応させていく必要がある．

謝辞

本研究を遂行するにあたり，終始多大なる御指導並びに御鞭撻賜りました，北陸先端科学技術大学院大学 情報科学研究科 赤木正人 教授に深く感謝致します．

本研究を遂行するにあたり，日頃から有益な御助言をいただき，御指導いただいた，北陸先端科学技術大学院大学 情報科学研究科 鵜木祐史 准教授，李軍鋒 助教に心より感謝致します．

研究発表にあたり，熱心な議論並びに貴重な御助言賜りました，北陸先端科学技術大学院大学 情報科学研究科 党建武 教授と徳田功 准教授に心より感謝致します．

日頃の研究生活において，多くの御助言ならびに激励を賜りました，北陸先端科学技術大学院大学 情報科学研究科 博士後期課程 羽二生篤さん，村上泰樹さん，並びに，音情報処理学講座の皆様，知能情報処理学講座の皆様，及び諸先輩方に厚く御礼申し上げます．

最後に，石川県での 2 年間の研究生活をいつも暖かく見守ってくれ，心の支えとなってくれた両親，兄姉，並びに多くの友人たち，諸先輩方，後輩たち，そして多くの知人の方々に心より感謝致します．

これまでの人生の中で出会った多くの方々の支えがあって，本論文を執筆することができました．音の研究は実に奥が深く，科学としての新たな知見を得る楽しさ，工学としての応用する楽しさなどを通して人間の素晴らしい聴覚に迫っていくという素晴らしい研究対象です．本論文が世界中の研究者にとって研究の足がかりになればと願っています．本当にありがとうございました．

参考文献

[1] R. J. McAulay, T. F. Quatieri, “Low-rate speech coding based on the sinusoidal model,” Advances in Speech Signal Processing, Chapter 6, pp.165-208. Marcel Dekker, 1991.

[2] O. Ichikawa, T. Fukuda, M. Nishimura, “Local Peak Enhancement for Automatic Speech Recognition,” Proc. of The Acoustical Society of Japan, 1-P-24, pp.185-186, 2007 Sep.

[3] 石本祐一, “時間情報と周波数情報を用いた雑音環境における音声の基本周波数推定に関する研究,” JAIST 博士論文, 2004.

[4] Y. Ephraim, D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,” IEEE Trans. ASSP, Vol.32, No.6,pp1109-1121, 1984.

[5] R. E. Kalman, “A new approach to linear filtering and prediction problems,” IEEE Trans. ASME - Journal of Basic Engineering, 82-D, pp.35-45, 1960.

[6] 北川 源四郎, “モンテカルロ・フィルタおよび平滑化について,” 統計数理 第 44 巻 第 1 号 pp31-48, 1996.

[7] `http://www.tulane.edu/~park/`

[8] T. Abe, T. Kobayashi, S. Imai, “Harmonics Tracking and Pitch Extraction Based on Instantaneous Frequency,” Proc. IEEE ICASSP, pp.756-759, 1995.

[9] 北川 源四郎, “時系列解析における数値的方法 –計算統計学的な方法–,” 計算機統計学 第 15 巻 第 2 号 pp159-170, 2002.

[10] 伊庭 幸人, 種村 正美, 大森 裕浩, 和合 肇, 佐藤 整尚, 高橋 明彦, “計算統計 II マルコフ連鎖モンテカルロ法とその周辺,” 岩波書店, 2005.

[11] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, “A Tutorial on Particle Filter for Online Nonlinear/Non-Gaussian Bayesian Tracking,” IEEE Trans. Signal Processing, Vol.50, No.2, pp174-188, 2002.

本研究に関する研究業績

口頭発表

Seiji Tomoike and Masato Akagi, “Estimation of local peaks based on particle filter in adverse environments,” *Proc. NCSP'08*, March 2008 (to be appear).