振幅調節と変調を施した振幅スペクトルを用いた雑音混入音声の基本周波数推定

静岡大学　博士論文

振幅調節と変調を施した振幅スペクトルを用いた雑音混入音声の基本周波数推定

2008年7月

静岡大学大学院電子科学研究科
電子応用工学専攻

小川　啓太

要旨

本論文では，雑音混入音声の基本周波数（F0）推定誤りを減らすことを目的とし，そのための新しいF0推定法の提案と実験についての研究を取りまとめたものである．

様々な環境で音声機器を使用する機会が多くなり，それに伴って実環境下音声からの高い精度のF0推定が必要となっている．音声情報処理で用いられる重要な特徴量であるため，これまでに多くのF0推定法が提案されている．しかし，雑音混入音声に対応した決定的な手法は確立されていない．雑音混入音声のスペクトルは雑音の影響で調波構造が明瞭でない帯域が多くなり，F0の推定誤りが多くなっている．そこで，スペクトルの調波構造の明瞭な帯域を増やすために，振幅スペクトルの変調と調波構造の特徴を利用した雑音低減，そして自己相関関数（ACF）の変調を用いている．音声の振幅スペクトルのピークが雑音の影響を受け難い場合に，そのスペクトルの情報を用いてF0推定の誤りを低減できるところに特徴がある．ここでは，シングルチャネル入力の雑音混入音声を観測信号として，雑音の情報を事前に与えない観測信号から処理フレーム内でF0を推定する．

まず，基本波成分を明瞭にさせるため，観測信号スペクトルとこの方法の後半で得られるACFに変調処理を適用する．ここで，変調周波数は観測信号スペクトルのピークの周波数から求めている．雑音の影響を低減するために，観測信号スペクトルの変調後に得られる変調スペクトルから雑音スペクトルを推定して，そのACFを変調スペクトルのACFから引く処理（ACS）を行っている．雑音スペクトルの推定は，スペクトルの調波構造を利用した変調スペクトルからの粗い雑音推定と，この粗い雑音スペクトルの精度を高めるための周波数領域での2入力ブラインド信号分離（BSS）技術による精度の高い雑音スペクトル推定の2つの処理からなる．このBSS出力の精度の高い雑音スペクトルのACFがACSに用いられる．これらの処理を組み込んだF0推定法（ACS-CM）の有効性を比較実験によって検討している．合成雑音混入音声の実験結果は，ACS-CMによってF0推定誤りが減ることを示した．基準F0の± 5 %以内の推定にならない推定誤りをGross F0 errorとした場合に，白色雑音混入音声で信号対雑音比（SNR）が-5 dBのF0推定では，自己相関を用いたF0推定法（AUTOC）のGross F0 errorと比べて11 %程度の誤り低減を実現した．工場（板金）雑音混入音声でSNRが0 dBのF0推定では，AUTOCのGross F0 errorと比べて3 %程度の誤り低減を

実現した．しかし，走行自動車内の雑音が混入した音声のF0推定で誤りを低減できない結果となった．走行自動車内雑音のような場合では，混入した雑音が音声のある帯域に偏在し，大きなパワーを持ち，その帯域の調波構造を大きく乱している．このような場合には，そのままの振幅スペクトルを用いると振幅の大きい雑音を多く含む帯域をそのまま移動することになり，雑音の影響が大きく，また変調の効果も小さい．

そこで，雑音がある帯域に偏在し，大きなパワーを持つ場合にも対応するため，大きなパワーを持つ雑音の影響を抑圧して，変調の効果がより大きく出るように，前処理として振幅調節を導入した改良をACS-CMに加える．そして，変調において複数の変調周波数を用いて行い，さらに反復の効果を検討することで効果のある反復回数を組み込んだF0推定を行う．

白色雑音混入音声に対しては振幅調節を行うことで雑音を強調することになり，効果がないため，まず振幅スペクトルの振幅の調節が必要であるかどうかの判断をする．この判断には，振幅スペクトルの全帯域のスペクトルの振幅分布の分散を振幅スペクトルの2乗平均で正規化した振幅分散を用いる．振幅調節は2段階で行う．最初は線形予測分析を応用して，バンド幅拡大を施した線形予測係数で構成される逆フィルタを通す．雑音混入音声スペクトルの大まかな傾きを含め，ホルマントや雑音によって偏在する大きなパワーを持つスペクトルの起伏を緩やかにすることで雑音の影響を抑圧する．次にF0探索範囲を考慮した帯域幅600 Hz程度の振幅スペクトルの平均を使った各周波数成分の振幅調節を行う．その後，複数の変調周波数を用いた振幅スペクトルの反復変調を組み込むことで，低域を含めて調波構造の明瞭な帯域を増やす．反復変調後のスペクトルにおいて，スペクトルの調波構造を利用した雑音推定を用いてスペクトル減算をする．この操作によってさらに調波構造は明瞭になる．そして，このスペクトルに対するACFを求める．振幅スペクトルを変調した同じ複数の変調周波数の情報を用いてACFの反復変調を行った後，このACFからF0の推定を行っている．この実験結果は，ACS-CMで問題となった特徴を持つ雑音においても有効性を示している．ACS-CMのGross F0 errorと比べて走行自動車内雑音混入音声のSNRが0 dBの場合，改良法は30 %程度の誤り低減を実現した．

実環境の雑音混入音声でF0推定の誤りを低減できる本研究の成果は，音声情報処理の幅広い分野において役立てることができる．

目 次

略語リスト

信号対雑音比	(	SNR	:	signal-to-noise ratio)
デシベル	(	dB	:	decibel)
離散フーリエ変換	(	DFT	:	discrete fourier transform)
低域通過フィルタ	(	LPF	:	low-pass filter)
自己相関関数	(	ACF	:	autocorrelation function)
ACS-CM	(	ACS-CM	:	autocorrelation subtraction and cosine modulation)
ブラインド信号分離	(	BSS	:	blind signal separation)
自己相関減算	(	ACS	:	autocorrelation subtraction)
スペクトルサブトラクション	(	SS	:	spectrum subtraction)

第1章　序論

1.1　音声の基本周波数

音声は，声帯振動の有無によって有声音と無声音に大きく分けることができる．有声音と無声音の決定は，周期性と非周期性の特徴と同一視して行われる．有声音は，声帯の振動によって断続された励振波で，短時間で考えた場合にほぼ相似的な周期波がみられる．このときの声帯の振動周期を基本周期と呼び，この逆数が基本周波数（F0）と呼ばれる．つまり，周期が長くなるとF0は低くなる関係にある．この声帯音源によって母音が生成され，子音は乱流雑音源や破裂音源なども音源とする [1]．声帯の振動は，声門の開閉，繰り返し回数，呼吸の送出量，喉頭の緊張度を調節することで，発話者が意識的に制御することができる．このことで，F0のほかに持続時間や圧力に対応する波形の強さが変化する．そして，音声波形は図1.1のように音源の生成，声道の形による調音，唇または鼻孔からの放射によって生成される [2]．そのため，声帯から口腔を経て唇に至る声道のフィルタによって特徴づけられ，女声や男声の個人の音色の違いとして現れる [3]．一般的な平均では，女声のF0が男声のF0の2倍程度で，F0変動の標準偏差も女声は男声の2倍程度であると知られている．F0は声帯長と関係があり，日本人の平均F0は，成人女性で250 Hz，成人男性で125 Hzとされている [4]．人間の発声のF0は広い周波数範囲を持つことができるが，その僅かな周波数部分だけが会話の発話で用いられる．声帯振動の周波数が高さ，振動の継続や休止の時間が長さ，振動振幅の違いが声の大きさを表す時間的変化パターンなどの要素によって言語が構成される．この時間的変化が，単語のアクセント，文のイントネーション，ストレス，リズム，意味の強調などを表現する．また，発話の自然さ，話者の個人性，性別，年齢，感情などの情報も持っている [5]．そのため，話者のくせや方言 [6] などの特徴として現れ，個人の音色の違い [7, 8] となる．そして，聴覚

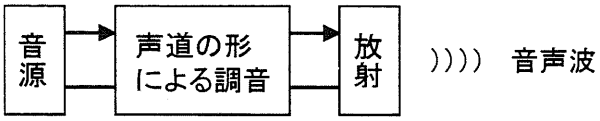


図 1.1: 音声生成の基本形 [2]

上では音の高さに関係している．同時に発話された音声や雑音を含む音声から目的の話者の発話を聞き分けることや音源の違いを特徴づけるひとつに，人間はF0を利用していると考えられている．

F0は，言語学や工学での話者照合 [9]，音声認識 [10]，音声の分析合成など利用されている．また，音声信号を利用するアプリケーションの多くで，F0のパラメータが用いられる．F0を使って得られたアクセントやイントネーションから，話者の認識や情緒性を推定する情報のひとつとして用いることもできる．そして，音声の分析，合成，符号化などの自然性に影響する．そのため，F0は音声信号の重要な特徴パラメータであり，信頼度の高い推定が必要とされている．

音声信号の周期性を時間領域で図1.2に示す．この音声信号は基本周期3.8 msで，4.4 msと8.2 msの2つの縦線付近に注目した場合，ほぼ相似した波形を確認できる．また，有声音スペクトルの調波性を周波数領域で図1.3に示す．F0の整数倍に相当する周波数の上下にプラス印を示す．有声音の周波数スペクトルは，F0とその整数倍成分となる周波数の調波成分によって，調波構造になっている．実際に発話された音声の低域周波数のスペクトルは，F0の整数倍付近で振幅が大きくなっていることを図1.3から確認できる．しかし，高域周波数のスペクトルでは，振幅が小さく調波構造が明瞭でないことを確認できる．

有声音を周期性と基本波の表現で見た場合，次のような特徴がある．周期性では，時間領域で時間的な繰り返しの構造，周波数領域で調波構造，相関領域で基本周期のピークとなって現れる．基本波では，時間領域で最も特徴的な部分の波形，周波数領域でF0のピークとなって現れる．

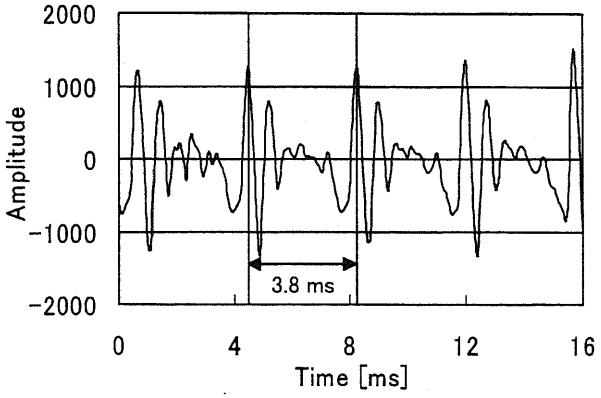


図 1.2: 女声/a/の時間領域の波形

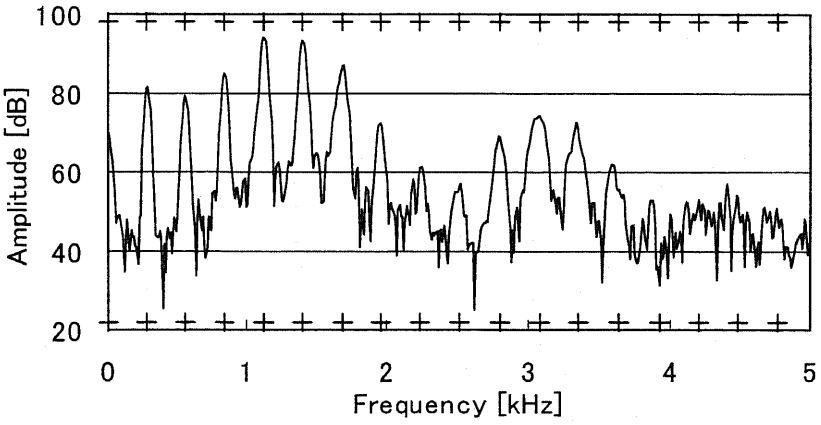


図 1.3: 女声/a/のスペクトル

1.2 基本周波数推定の問題

音声の特徴でF0に着目する．精度の高いF0推定は，以下の理由から困難になることが知られている．

音声は，時間によって徐々に変化する準周期的な波であり，完全な周期波ではない．そして，会話音声には音声区間と音声休止区間が混在している．そこで，音声波形のどの時間区間を周期として抽出するか自明ではない．これは，図1.2の全体的な振幅の違いや振幅の小さな部分の波形が完全な一致ではないことからも確認できる．特に，音声休止区間との変化部分の語頭や語尾，無声音と有声音

の変化部分などでは声帯振動が完全な周期性を持たないことで困難になる．これに伴い，実音声の有声／無声区間推定 [11, 12] における判断も重要な課題である．

また，声帯振動に伴う生理的現象を電気信号で観測した場合などでなければ，口から放射された音波を処理する必要がある．そのため，声道特性が影響することで，声帯音源を直接観測信号として扱えない．また，F0 の変化幅が広帯域であることや雑音の混入についても問題である．同一話者でも短時間のうちに 1 オクターブ以上変化する場合もある [13]．そして，実環境で収録された音声は，目的となる音声以外に雑音を含む場合が多くある．そこで，雑音が周期性や調波性に悪影響を及ぼす．このため，F0 の 2 倍の周波数を推定する誤りや F0 の半分の周波数を推定する誤り，さらに F0 付近を推定できるが正確な F0 ではない誤りなどが起こる．また，先ほどの声帯振動が完全な周期性を持たない部分に雑音が影響することで，ますます F0 推定は困難になる．

以上の問題点から精度の高い F0 推定が困難であり，特に雑音混入音声は F0 推定精度が大きく低下するために問題である．そのため雑音混入音声の F0 推定について，研究が必要であると考えられる．

1.3 従来の音声の基本周波数推定法

音声波形は，振幅と位相が時間的に緩やかに変化する正弦波の和で構成されていると考えることができる．そこで，低域通過フィルタ（LPF : low-pass filter）に通すことで基本波成分のみを得ることが可能な場合，F0 は求められる．しかし，F0 推定の問題点があるため困難である．そのため，F0 推定は以前から多くの研究で報告されている [14, 15, 16, 17, 18]．準周期信号の周期性を安定して抽出する方法，周期性の乱れによる F0 推定誤差の補正を行う方法，ホルマントの影響を取り除く方法などが考えられた．

これまでに提案されている F0 推定法の多くが．主に次のように分けられる [19, 20]．

- 波形処理

  周期性の特徴を利用した手法で波形の時間間隔を利用する．手法の実装が

比較的容易で計算量の少ない場合が多いが，声道特性によって大きく変化する．

手法には，波形のピークを多種類から推定された基本周期の多数決で求める並列処理法 [21] や波形データから基本波の候補以外の情報を除いていくデータ減少法 [22] や波形の零交差数に関する繰り返しパターンに着目した零交差計数法 [23] や減衰振動の包絡を強調することで波高のパルスを作成する方法 [24] などがある．

- 相関処理
  波形の位相歪みに強いとされ，比較的ランダム性雑音に頑健である．

  手法には，音声波形の自己相関から求める自己相関法 [25] や LPC 分析の残差信号の自己相関から求める変形相関法や平均振幅差によって周期性を検出する AMDF 法 [26] などがある．

- スペクトル処理
  声道の影響を取り除き，励振波を求めることに適している．

  手法には，対数のパワースペクトルの逆フーリエ変換によりスペクトルの包絡と微細構造を分離するケプストラム法 [27] やスペクトル上の F0 の高調波成分のヒストグラムから求めた高調波の公約数によって決定するピリオドヒストグラム法 [28] などがある．

それぞれの処理における利点がある．そのため，各処理を組み合わせた手法も多く提案されている．

1.4 背景と目的

種々の環境で音声機器を使用する機会が多くなり，それに伴って音声処理システムの多くで F0 が必要になる．この音声を用いたシステムの多くでは，周期性を利用するため，正確な F0 抽出を必要とする．しかし，声帯振動が完全な周期性でないことや声道特性などによる，推定誤りが指摘されている．そこで，正確な音声の F0 推定の研究が必要となっている．さらに，システムが実際利用され

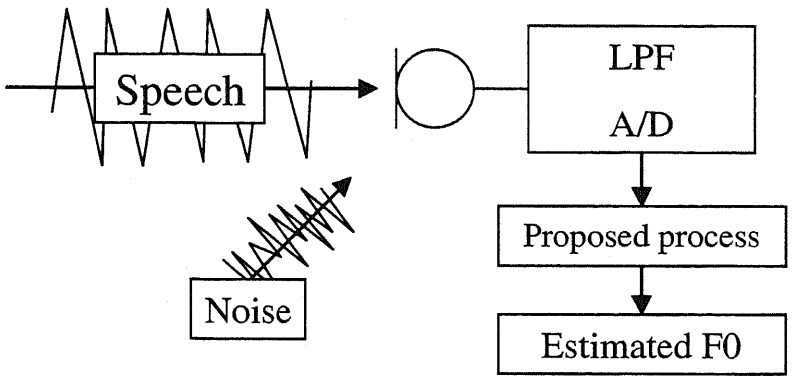


図 1.4: 雑音混入音声の F0 推定の概観

る場合，周囲には常に雑音の存在が考えられる．実環境で収録された音声は，目的となる音声以外に雑音を含む場合が多くある．そのため，雑音が調波成分に悪影響を及ぼし，F0 の推定率を大きく低下させる．その結果，F0 推定の精度低下に比例して，各システムでの精度低下が考えられる [9, 10]．このために雑音混入音声についての研究がさらに必要となっている．

音声をコンピュータを用いて処理するため，アナログ波形の連続した音声をデジタル音声に変換した離散的なデータとして扱った．雑音混入音声の F0 推定を図 1.4 に示す．ここでの対象信号の取得には，シングルマイクを使ったシステムを想定している．大型な装置などについては，周りが雑音混入の少ない環境である場合や装置内に雑音が入り難い環境になっているものが多くある．しかし，場所を選ばない装置や携帯が可能なものなどは，大型な装置と違い雑音の混入が少なくなるように周りの環境を配慮することが困難であり，観測される場所や場面によって雑音の種類も変化する．そこで，雑音混入音声の F0 推定が特に重要となるのは，構造が単純で装置規模が小さい汎用性のものが多いと考え，シングルチャネル音声信号の F0 推定を行う．

また，利用される実環境も様々な場所が考えられ，その違いによって雑音の傾向が変わってくる．目的とする音声以外の各音源から発生した信号は，それぞれが異なった周波数スペクトルの特性となる．走行自動車内の雑音では，車内にエンジン音や走行ノイズなど様々な音が存在することになる．このように，白色雑音と比べて特定の周波数帯域にエネルギーが偏在する特性を持つ場合もある．時

間によって雑音のスペクトルは変化する．そこで，どの周波数帯域にエネルギーが偏在しているのかを時間とともに求める必要がある．

F0の精度が高ければ，F0を利用するシステムの精度も向上することが期待できる．そこで，この雑音混入音声からF0推定の精度向上を目的にする．しかし，理想的なF0推定は困難で，雑音の影響で大幅な推定率の低下になる．正しいF0とはまったく異なったF0が推定されることや正しいF0付近の周波数を推定するが，F0の数パーセント程度違った周波数を推定される問題がある．そこで，これらの改善を目標にしたF0推定法を提案する．F0推定に影響する雑音の成分を低減することや明瞭な情報を用いて必要な成分を強調する必要がある．F0推定率を上げるためにスペクトルの調波構造の明瞭な帯域を増やす．そのために，振幅スペクトルの変調と調波構造の特徴を利用した雑音低減，そして自己相関関数の変調を用いる．

雑音混入音声では，有声／無声区間推定[29, 30]も重要な課題である．ただし，ここでは事前の雑音情報を与えず，有声／無声判定は考慮しないとする．

1.5 本論文の構成

本論文は，全6章で構成される．各章の概要を以下に述べる．

- 第1章
  音声のF0の概説と研究の関係について述べた．

- 第2章
  雑音混入音声のF0について述べ，自己相関関数と自己相関法について説明する．

- 第3章
  雑音混入音声のF0推定法となるACS-CMの処理全体の構成と流れを述べ，各処理の詳細を説明する．

- 第4章
  ACS-CMの有効性を検討するために雑音混入音声を用いた実験について述

べる．実験で必要となる条件や評価方法などと，実験で用いた音声サンプルについて説明する．実験結果は，合成雑音混入の場合と実雑音混入の場合について述べる．実験では既存のF0推定法と比較を行い，同様に結果を示す．

- 第5章
  偏在する雑音のための改良について述べる．ACS-CMに改良を加えた改良法の処理全体の構成と流れを説明する．改良法の有効性を検討するために行った実験についてと，その結果を合成雑音混入の場合と実雑音混入の場合について述べる．

- 第6章
  本論文で得られた結果を要約し，今後の課題を述べる．

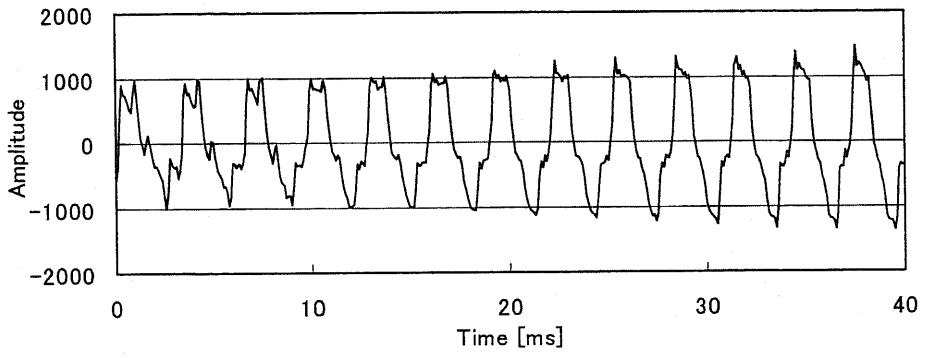
第2章　雑音混入音声の基本周波数推定

2.1　はじめに

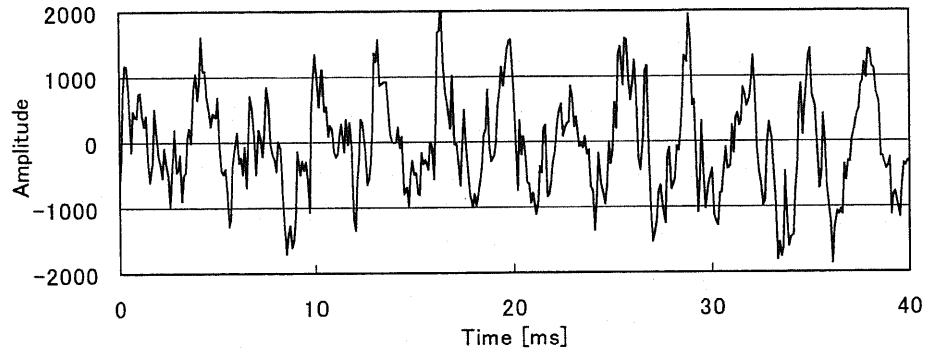
実際の環境では，周りに雑音が存在するため，観測されるほとんどの音声に雑音が混入する．観測した信号では，目的音声以外を雑音であると考えることができる．雑音は，発生原因により周波数でそれぞれ異なった特性のスペクトルとなり，目的音声は影響した雑音の特性により歪み方が変わる．時間領域で音声波形が大きく変化し，波形のピーク位置の間隔が歪み周期性を乱す．環境によって雑音が異なるため，あらかじめ雑音の特性を予測することは困難である．そこで，各観測信号ごとに対応できる雑音混入音声のF0推定が必要である．

音声はある程度の間隔で考えた場合に非定常で，そのスペクトルは時間とともに変化する．音声と比べて背景雑音などの雑音は，定常かゆっくりとした変化であると仮定できる．ここで，本研究での処理はshort-timeを基にしているため，短い間隔で考えた場合に雑音信号のスペクトルは，定常であると考えることができる．

本研究では，信号対雑音比（SNR : signal-to-noise ratio）がそれぞれのデシベル（dB : decibel）となる実雑音混入音声のシミュレートをするために，コンピュータを用いてプログラムによりクリーン音声へ実雑音を加えた．そこから得られる信号を，実環境で収録された雑音混入音声であると仮定した音声サンプルを用いて，実験を行っている．そこで，雑音の混入について以下で述べる．



(a) クリーンな音声



(b) 白色雑音混入音声（SNR −5 dB）

図 2.1: 女声話者が発話した/soozoo…/の音声波形の一部

2.2 雑音混入音声信号

雑音の混入した音声を観測信号とした場合，時間 t での観測信号 $v'(t)$ は下式によって与えられる．

$$v'(t) = s(t) + z(t) \tag{2.1}$$

ここで，$s(t)$ はクリーンな音声，$z(t)$ は雑音である．

クリーンな音声と雑音混入音声について，同じ音声部分の波形を図 2.1 に示す．図 2.1(a) のクリーンな波形では，徐々に変化があるが周期性を確認できる．そのため，基本周期を求めることができる．しかし，式 (2.1) のように雑音が混入した図 2.1(b) 場合は，雑音の影響で類似した波形を見つけることが困難である．そのため，(a) と比べて周期性を確認できない．

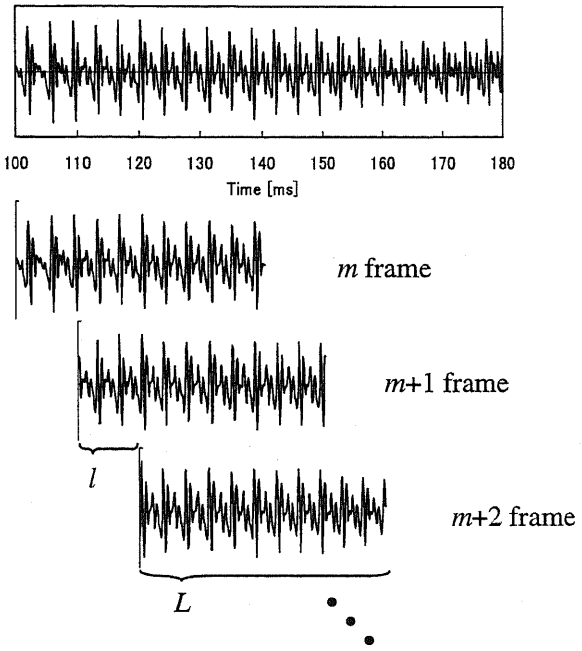


図 2.2: フレーム化の概略

観測信号を重複するフレームに分けることで short-time の信号サンプルを得て，処理はフレームごとに行う．このフレーム化した観測信号 $v_m(n)$ は，次式で求められる．

$$v_m(n) = v'(n + ml) \tag{2.2}$$

ここで，n は離散時間のサンプル番号 $(n = 0, 1, ..., L-1)$，m は時間フレームの番号 $(m = 0, 1, ...)$，L はフレーム長，l はフレームのシフト幅である．フレーム化の概略を図 2.2 に示す．このフレーム化した観測信号についてフレームごとに F0 推定を行い，それぞれのフレームでの F0 が求まる．

多くの雑音低減アルゴリズムで，音声に混入した雑音を取り除き易くするために short-time 離散フーリエ変換（DFT : discrete fourier transform）が用いられる．時間領域の信号は，DFT で周波数領域に変換することができる．これは，信号成分と雑音成分の混合した信号を周波数領域で扱うためである．周波数領域の雑音混入音声は DFT によって，次式で表すことができる．

$$\sum_{n=0}^{N-1} v_m(n) \exp\left(-j\frac{2\pi}{N}hn\right)$$
$$= \sum_{n=0}^{N-1} s_m(n) \exp\left(-j\frac{2\pi}{N}hn\right) + \sum_{n=0}^{N-1} z_m(n) \exp\left(-j\frac{2\pi}{N}hn\right)$$
$$V_m(f_h) = S_m(f_h) + Z_m(f_h) \tag{2.3}$$

ここで，$V_m(f_h)$，$S_m(f_h)$，$Z_m(f_h)$ は，それぞれ周波数領域の観測信号，クリーンな信号，雑音混入信号で，f_h は周波数点である．

雑音混入音声の F0 推定では，このように雑音の影響した信号を用いて推定する必要がある．そこで，雑音混入音声に着目した既存の F0 推定法の一部を以下で述べる．

2.3 既存の雑音混入音声の基本周波数推定法

F0 推定の研究では，実環境における雑音の存在に対応した雑音混入音声の F0 推定法も検討されている．雑音の影響による問題があるため，様々な手法が提案されてきた [31, 32, 33]．

雑音混入音声について提案されてきた F0 推定法は，これまでに次のような手法が挙げられる．複数の推定から総合的に求める手法には，スペクトル包絡パターンの時間的連続性を利用して複数個の基本周期候補から決定する手法 [34] や複数の異なる幅の分析窓を用いた自己相関関数から得られた窓の数分の基本周期候補から重み付けにより選択される手法 [13] などがある．ケプストラムを有効に利用する手法には，対数スペクトルのうち特に雑音の影響を受けやすい高周波数成分とスペクトルの谷の部分を除去し音声信号の調波構造を明瞭にした上でケプストラムを求める手法 [35] やケプストラム法にハフ変換を適用し様々な雑音に対して頑健な手法 [36] などがある．また，調波性の特徴を利用した手法には，コムフィルタの中心周波数を求めて調波の存在する周波数を決定する振幅スペクトルのコムフィルタリングを利用する手法 [37] や対数スペクトルの自己相関関数を利用する手法 [38] などがある．そして，瞬時周波数に着目し [39, 40]，瞬時振幅に表れる音声の周期性と調波性を利用した手法 [41] や音声の周期性と調波性に対してエントロピーによる重み付けを利用した手法 [42] などがある．雑音混入音声の基本周波数推定法について，主な従来法を付録 A にまとめる．

有用な手法の一つとして，自己相関法が挙げられる．雑音に頑強であり，特に白色雑音においては効果的である．その中で，スペクトルサブトラクションを用いる手法 [43] があるが，適切な減算が困難な問題を持っている．そして，ブラインド信号分離などを適宜利用する方法 [44] の研究もある．これらで必要となる雑音の推定スペクトルについて，より正確に求めることも課題である．

既存の雑音混入音声のF0推定に関する手法の中で，白色雑音混入に頑健とされる手法と走行自動車内雑音に頑健とされる手法に着目する．比較的新しく提案された手法の中で特に，F0の存在する帯域内で振幅スペクトルのべき乗を施すF0抽出法（BPPAS）と瞬時振幅の周期性と調波性を考慮したF0抽出法（EWPH）に着目し，次に説明する．

BPPAS

島村らは，白色雑音で劣化した音声に着目し，帯域制限をかけた振幅スペクトルのべき乗に基づく基本周波数抽出法（BPPAS）[45] を提案している．この流れ図を図2.3に示す．まず，信号は周波数領域に変換され，そのスペクトルをべき乗する．そして，F0 推定は処理後のスペクトルをIDFTで求めた値で行われる．このスペクトルのパワーに用いられるべき乗の指数は，入力のSNRによって設定される．この基本関数は，帯域制限された振幅スペクトルのべき乗を逆フーリエ変換することで得られる．雑音推定部では，あらかじめ既知の音声休止区間の分析フレームを用いる．

この方法の特徴は，白色雑音混入音声に効果的で，対象雑音の定常性が必要になる．

EWPH

石本らは，走行自動車内雑音のような一部の帯域にパワーが偏在する雑音に着目し，瞬時振幅の周期性と調波性を考慮したエントロピーで重み付けした周期性・調波性特徴を用いる方法（EWPH）[46] を提案している．この流れ図を図2.4に示す．

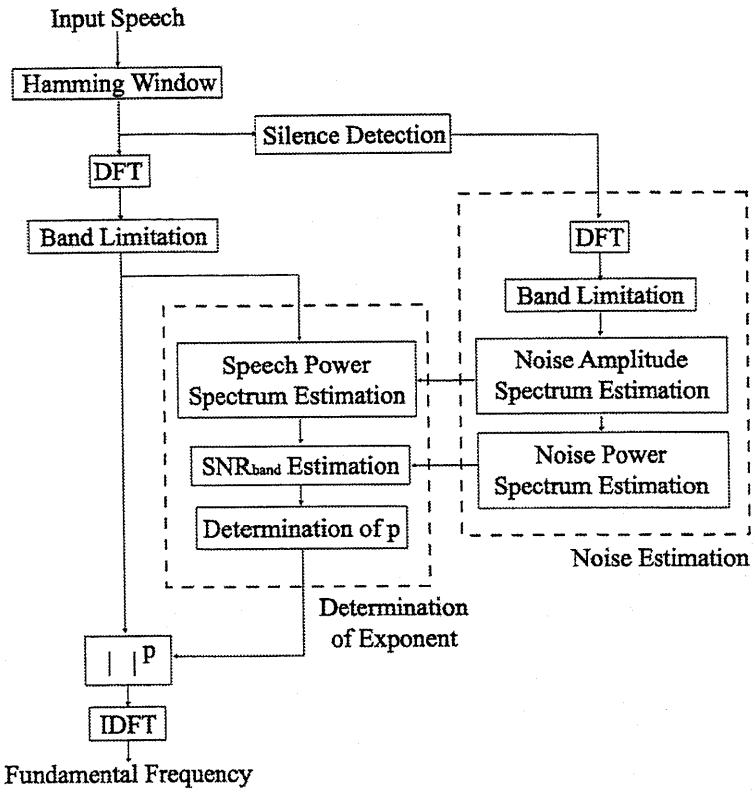


図 2.3: BPPAS の流れ図（島村ら [45]）

瞬時振幅の周期性 (時間情報) と調波性 (周波数情報) を基にした雑音に頑健な基本周波数推定と帯域幅可変櫛形フィルタによる雑音抑圧，瞬時周波数を用いた高精度な F0 推定の STRAIGHT-TEMPO[47] を組み合わせた手法である．

この手法の特徴は，雑音のエネルギーが小さな周波数帯域がある場合となる特定の周波数帯域にパワーが偏在する雑音混入音声で高い耐雑音性能を示す．

2.4 基本周波数推定法の比較

これまでに，雑音を含む音声の F0 推定方が複数提案されている．その中で，基本的であると考えられる自己相関，ケプストラムを用いたそれぞれの F0 推定法で耐雑音性の傾向について比較を行う．

それぞれの処理を図 2.5 に示す．ここで，図 2.5 の自己相関，ケプストラムをそれぞれ AUTOC，CEPST と呼ぶことにする．AUTOC は，2.5 節で詳細を述べ

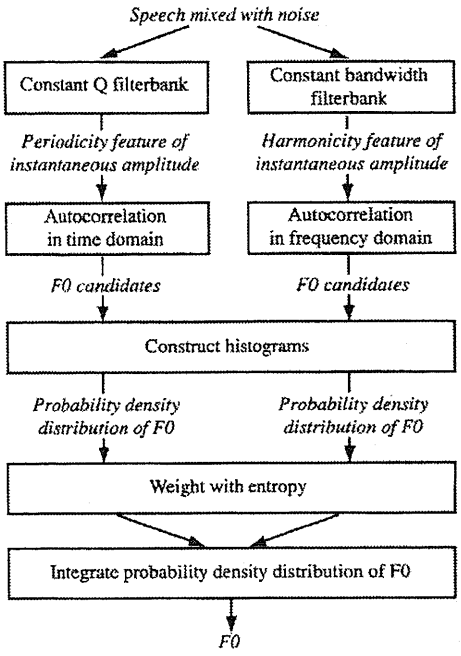


図 2.4: EWPH の流れ図（石本ら [46]）

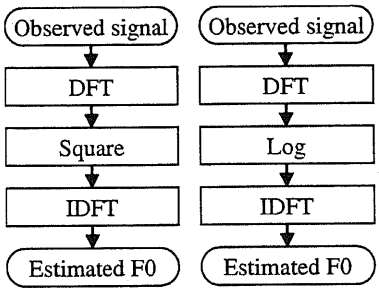


図 2.5: F0 推定の基となる手法の概略

る．CEPST は，ケプストラム分析を行い，その高ケフレンシー部分のピークを探索範囲内から推定し，ケフレンシーの逆数を求める．

ここでは，傾向を求めて検討するために，あらかじめ設定した F0 となる合成音を用いて実験を行う．処理の比較実験に用いた合成音とその結果を次に示す．

表 2.1: 合成音のホルマント

/a/	周波数 [Hz]	1160	1570	3090	4200
	帯域幅 [Hz]	60	70	130	200
/i/	周波数 [Hz]	340	2630	3480	4200
	帯域幅 [Hz]	50	110	150	200
/u/	周波数 [Hz]	340	1270	2750	4200
	帯域幅 [Hz]	50	60	110	200
/e/	周波数 [Hz]	500	2260	3130	4200
	帯域幅 [Hz]	50	90	130	200
/o/	周波数 [Hz]	580	910	3240	4200
	帯域幅 [Hz]	50	60	140	200

合成音サンプル

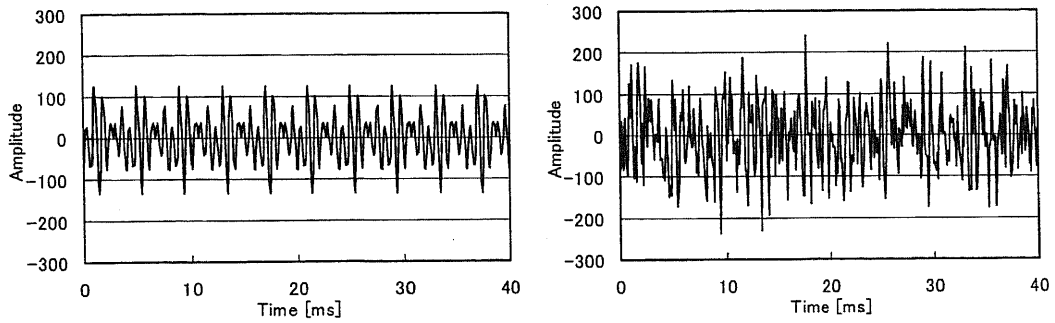
コンピュータを用いて合成する．声門開口比 0.7 のローゼンベルグ波を入力として用い，サンプリング周波数 10 kHz で合成した．また，放射特性は，差分特性 $(1 - z^{-1})$ を用いた．女声の典型的な 4 ms と男声の典型的な 8 ms となるような，基本周期を想定した．ホルマントについては，表 2.1 のように設定した．また，図 2.6(a) に生成したものを示す．さらに，合成音に白色雑音を付加した波形を図 2.6(b) に示す．時間領域において周期性の識別が，極端に困難なったことを確認できる．

この合成音の単母音は，コンピュータで合成した白色雑音を付加することで劣化させた．独立に 2000 回合成した雑音をそれぞれに混入させた，各フレームについて 2000 データを用いた．

実験結果

F0 推定の評価では，250 Hz または 125 Hz のそれぞれ ±20 %以内の周波数が推定できた場合を正解とした．

クリーンな合成母音では，AUTOC と CEPST ともに正しい推定ができる．



(a) クリーンな波形　　(b) 雑音を付加した波形

図 2.6: 基本周期を 4 ms の設定で合成した/a/の波形

雑音混入の合成母音それぞれを各 F0 推定法で求めた．表 2.2 に F0 が 250 Hz となるように合成した音声の F0 推定誤り率を示す．表 2.3 では，同様に F0 が 125 Hz の場合を示す．

表 2.2 の AUTOC は，CEPST と比べ比較的に誤りが少ないことを確認できる．また，表 2.3 の AUTOC も同様の傾向を確認できる．そのため，基礎的な F0 推定において AUTOC と CEPST を比べた場合，AUTOC の耐雑音性が白色雑音について比較的高いと考えられる．

考察

雑音混入音声の F0 推定をする場合，耐雑音性のある手法を用いることが望まれる．今回の比較では，相関処理をする AUTOC が有効であると考えられる．よって，以後の実験では AUTOC を基とした音声の F0 推定を提案することとする．ただし，白色雑音が混入した場合から判断している．

2.5 自己相関関数を用いた基本周波数推定

波形の周期性を求めるために自己相関を用いる．自己相関を用いた処理は雑音に比較的頑健で，特に白色雑音混入で有効な処理のひとつとされている．そこで，2.4 節の比較からも雑音混入音声の F0 推定に AUTOC が有効であると考えられる．以下に，自己相関関数（ACF）と AUTOC について述べる．

表 2.2: F0 が 250 Hz の雑音混入合成音の F0 推定誤り率 [%]

	F0 推定の誤り率 [%]			
	SNR 5 dB		SNR 0 dB	
母音	AUTOC	CEPST	AUTOC	CEPST
/a/	0.00	19.19	5.61	39.15
/i/	0.00	20.50	3.84	33.74
/u/	0.00	23.92	1.41	37.04
/e/	0.23	30.97	13.87	48.20
/o/	0.00	17.27	2.33	31.22
平均	0.05	22.37	5.41	37.87

表 2.3: F0 が 125 Hz の雑音混入合成音の F0 推定誤り率 [%]

	F0 推定の誤り率 [%]			
	SNR 5 dB		SNR 0 dB	
母音	AUTOC	CEPST	AUTOC	CEPST
/a/	0.00	0.64	0.18	18.22
/i/	0.00	2.10	0.00	18.45
/u/	0.00	5.72	0.00	20.55
/e/	0.03	4.84	8.10	28.64
/o/	0.00	0.41	0.00	8.55
平均	0.01	2.74	1.66	18.88

自己相関関数

周期性のある波形では，以前の波形とその波形自体の相関が高いかを調べることで，周期性を持つ区間を導くことができる．そこで，ACF を用いる．時間軸をシフトした波形と元の時間における波形を用いることで，最も類似した波形を別の時間帯から求め，基本周期を推定することができる．

この自己相関の推定量計算には，標本データの値から直接計算する方法かDFTを用いることで求めることができる．

まず，直接計算する場合について述べる．データ数をL，遅延点をi_{d}とするときACF $R'(i_{\mathrm{d}})$は，

$$R'(i_{\mathrm{d}}) = \frac{1}{L - i_{\mathrm{d}}} \sum_{n=0}^{L-i_{\mathrm{d}}-1} x(n)x(n+i_{\mathrm{d}}) \tag{2.4}$$

となる．ここで，$i_{\mathrm{d}} = 0, 1, \cdots, n_{\mathrm{L}}$である．さらに，$L \gg n_{\mathrm{L}}$の場合，

$$R'(i_{\mathrm{d}}) = \frac{1}{L} \sum_{n=0}^{L-i_{\mathrm{d}}-1} x(n)x(n+i_{\mathrm{d}}) \tag{2.5}$$

と示せる．これは，ACFの偏りを持った推定量となる．ここで，遅延がない状態$(i_{\mathrm{d}} = 0)$は，

$$R'(0) = \frac{1}{L} \sum_{n=0}^{L-1} x^2(n) \tag{2.6}$$

となり，最大のACFとなる．そのため，この値を用いACFの正規化を行うことで，正規化したACFを求める．

$$\begin{aligned} R(i_{\mathrm{d}}) &= R'(i_{\mathrm{d}})/R'(0) \\ &= \left(\sum_{n=0}^{L-i_{\mathrm{d}}-1} x(n)x(n+i_{\mathrm{d}}) \right) \Big/ \left(\sum_{n=0}^{L-1} x^2(n) \right) \end{aligned} \tag{2.7}$$

すなわち，

$$-1 \leq \frac{R'(i_{\mathrm{d}})}{R'(0)} \leq 1 \tag{2.8}$$

となる．

ウィナー・ヒンチンの定理より，ACFとパワースペクトル密度は，互いにフーリエ変換の関係がある．これにより，正規化前のACFを導くことにする．

$$R'(i_{\mathrm{d}}) = \mathrm{IDFT}\left(|X_1(f_h)|^2 \right) \tag{2.9}$$

$$\tag{2.10}$$

さらに，$R'(0)$で正規化を行い正規化したACFを求める．ここで，$x(n)$の周波数表現を$X_1(f_h)$，IDFTを逆フーリエ変換とする．

$$R(i_{\mathrm{d}}) = \frac{R'(i_{\mathrm{d}})}{R'(0)} \tag{2.11}$$

このフーリエ変換を用いた求め方は，全体の演算を直接求めるよりも効率よくすることができる．

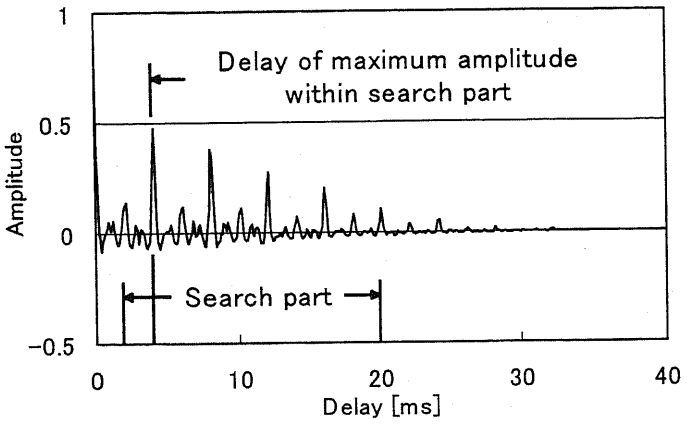


図 2.7: 探索範囲内の最大振幅の抽出

自己相関法

本研究で基礎となるAUTOCについて説明する．時間領域の観測信号はDFTを行うことで周波数領域表現$V_m(f_h)$に変換され，式(2.9)から観測信号のACF $R_m(i_\mathrm{d})$はパワースペクトル$|V_m(f_h)|^2$の逆DFT（IDFT）で得られる．$R_m(i_\mathrm{d})$は次式で求められる．

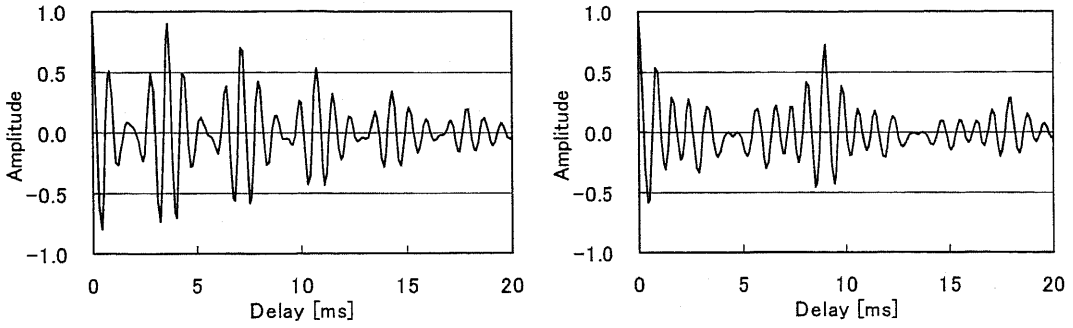
$$R'_m(i_\mathrm{d}) = \frac{1}{N}\sum_{h=0}^{N-1}\left|V_m(f_h)\right|^2 \exp\left(j\frac{2\pi}{N}hi_\mathrm{d}\right) \tag{2.12}$$

ここで，i_dは遅延時間のサンプル番号である．正規化したACFは次式で求められる．

$$R_m(i_\mathrm{d}) = \frac{R'_m(i_\mathrm{d})}{R'_m(0)}. \tag{2.13}$$

ACFのピーク点が遅延のない位置（$i_\mathrm{d}=0$）となる．探索範囲内でのACFのピーク点は，図2.7に相当するサンプル点のように抽出される．この遅延時間が基本周期として推定され，振幅が信号の周期性の値として求められる．図2.7では，フレーム長が40 msの場合に遅延時間は40 msまで求められる．想定されるF0の範囲に相当する基本周期によって決めた探索範囲の中から最大の振幅を求め，基本周期を抽出する．サンプリング周波数がS_Fで，遅延のサンプル点（$i_\mathrm{d}=d_\mathrm{p}$）で振幅$R_m(d_\mathrm{p})$が探索範囲の最大のとき，基本周期$T_\mathrm{p}$は

$$T_\mathrm{p} = d_\mathrm{p}/S_\mathrm{F} \tag{2.14}$$



(a) 女声　　(b) 男声

図 2.8: クリーンな/a/の ACF

となり，逆数から F0

$$F_0 = 1/T_{\mathrm{p}} \tag{2.15}$$

が求められる．

クリーンな女声と男声の単母音/a/の ACF 例をそれぞれ図 2.8(a) と図 2.8(b) に示す．図 2.8 の (a) では，基本周期付近 4 ms に相当する遅延時間付近に大きな振幅が見られる．(b) では，基本周期付近 8 ms のため (a) で見られた大きな振幅は見られず，次の大きな振幅が探索範囲内の最大振幅となる．/a/では，基本周期付近以外の遅延点で顕著な振幅の起伏が見られる．同様に/a/以外の単母音について女声の ACF を示す．図 2.9 の/i/，/u/では，基本周期付近に大きな振幅が見られる．/e/では，基本周期付近の振幅以外にも大きな振幅が確認できる．そのため，基本周期の 1/2 倍あたりや 2 倍あたりに見られる振幅を探索範囲内の最大に誤る可能性が高くなる．

比較的 ACF に規則性が見られた単母音/i/に雑音の混入した場合の ACF を図 2.10 に示す．雑音の影響によって周期性が乱れたため，遅延のない位置（0 ms）以外で相関が少なくなり，全体的に振幅が小さくなったことを確認できる．特に遅延が 1 サンプル点（$i_{\mathrm{d}} = 1$）の位置と基本周期に相当する遅延点の位置で，振幅の変化に注目できる．そして，探索範囲内での最大振幅は本来の基本周期と比べて 2 倍の位置になっている．これは，F0 の 1/2 倍の周波数を誤って推定することになる．

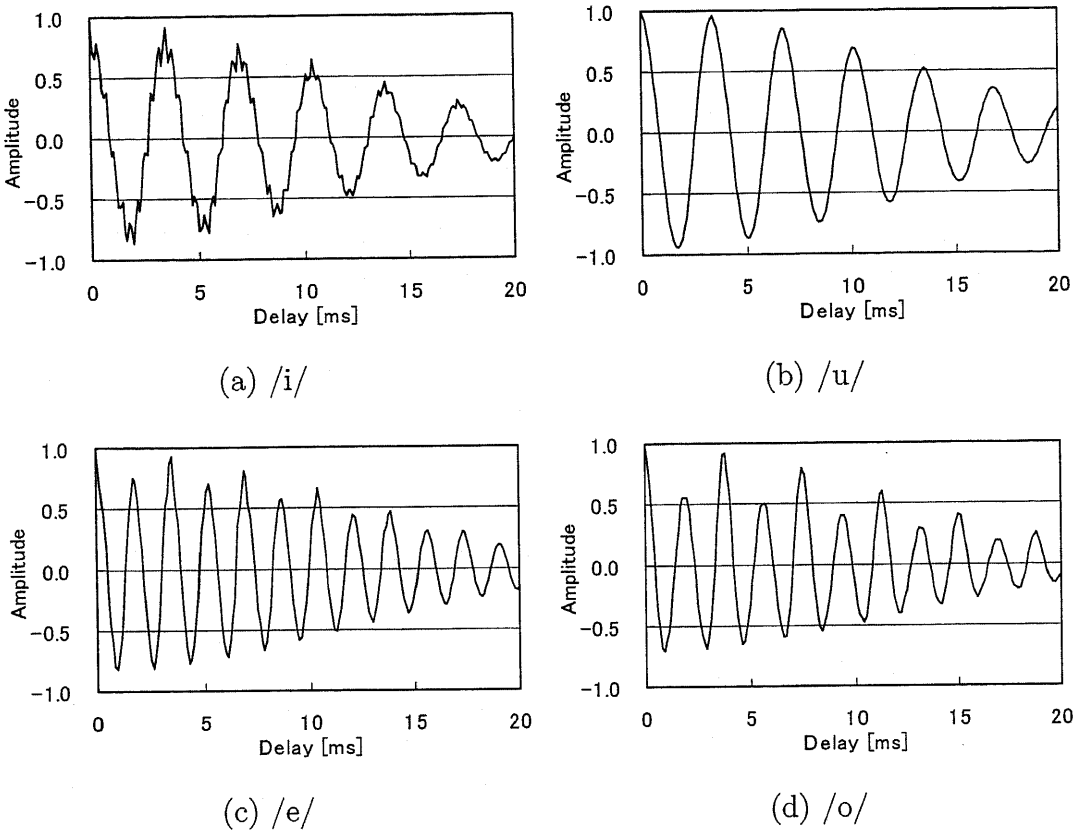


図 2.9: 女声のクリーンな単母音の ACF

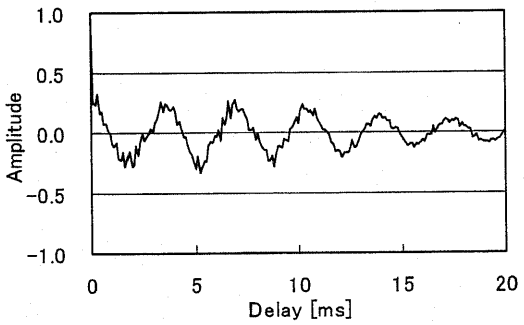


図 2.10: 女声の雑音混入音声/i/の ACF

2.6 雑音の低減

雑音混入音声の処理として，雑音の影響を減算や抑圧することが考えられる．

混入した雑音が既知の場合や雑音のみの入力信号を得られる場合は，雑音の影響を除去することが比較的容易になる．しかし，実際には困難であり，システムが限られる．

そこで，混入した雑音が未知の観測信号から雑音の情報を得る必要がある．このためには，混入した雑音を推定する必要がある．主な雑音推定には，音源の到来方向の空間情報を得ることで推定するマイクロホンアレーによる雑音推定や定常雑音の除去を対象とするスペクトルサブトラクションによる雑音推定などがある．F0 推定での影響は少ないと考えられるが，推定された雑音の精度によっては，雑音低減の処理でミュージカルノイズなどの問題がある．また，ブラインド信号分離によって音声信号と雑音成分を分離する [48] ことも考えられる．

混入した雑音が完全に推定できた場合は，観測信号から減算することで雑音の混入していないクリーンな音声を処理でき，精度の良い F0 推定ができる．そこで，雑音が推定できる場合は，雑音混入音声の F0 推定で雑音の減算を用いることが有効であると考えられる．

第3章　自己相関減算とコサイン変調を用いた基本周波数推定

3.1　はじめに

雑音混入音声のスペクトルは雑音の影響で調波構造が明瞭でない帯域が多くなり，F0の推定を困難にしている．そこで，スペクトルの調波構造の明瞭な帯域を増やすために，振幅スペクトルの変調と調波構造の特徴を利用した雑音低減，そしてACFの変調を用いる．

本研究で，観測信号の変調，Rough雑音推定，ブラインド信号分離（BSS : blind signal separation），自己相関減算（ACS : autocorrelation subtraction），コサイン変調の組み合わせたF0推定法をACS-CM（ACS-CM : autocorrelation subtraction and cosine modulation ）と呼ぶことにする．この章では，それぞれの処理について説明する．

図3.1はACS-CMの流れ図を示している．観測信号をフレーム化して，窓掛け処理を行う．フーリエ変換から観測信号のスペクトルを求め，帯域制限を施す．帯域制限後のスペクトルを変調して，Rough雑音を推定する．BSSを用いて，Rough雑音からFine雑音を推定する．帯域制限後に変調したスペクトルとFine雑音のスペクトルをそれぞれ二乗のIDFTからACFを求める．そのACFを用いたACSから雑音の影響を低減したACFを求め，コサイン変調を施す．そこから得たACFの探索範囲内の最大振幅となる遅延に相当する周期点でサンプリング周波数を割り，F0を推定する．

3.2　観測信号の変調

変調の原理を以下で述べる．

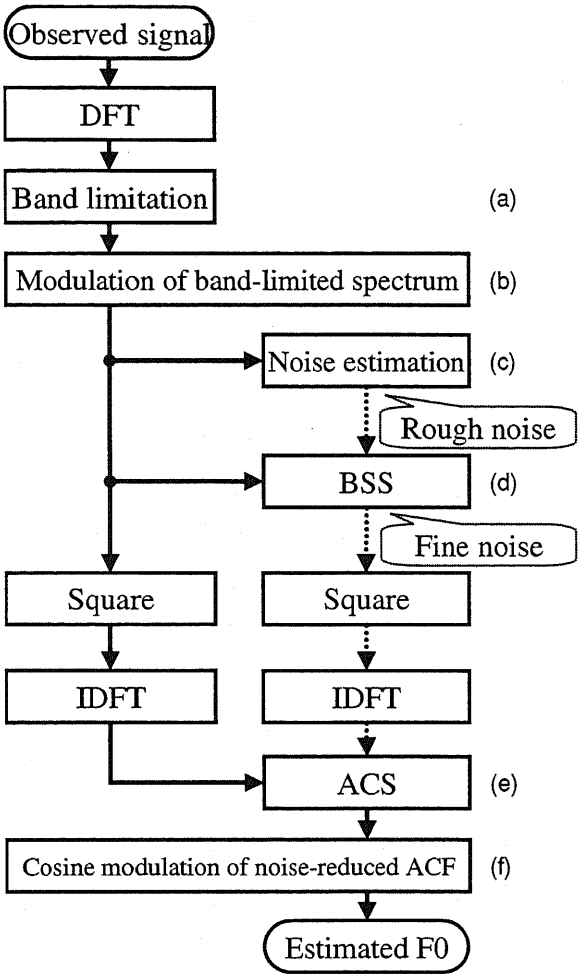


図 3.1: ACS-CM

F0 を f_0，$\omega_0 = 2\pi f_0$ として，周期信号 $s(t)$ を次式で表す．

$$s(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t + \theta_n) \tag{3.1}$$

$s(t)$ の ACF $R(\tau)$ は次式で表せる．

$$R(\tau) = \frac{A_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty} A_n^2 \cos(n\omega_0 \tau) \tag{3.2}$$

式 (3.2) に，$\tau = T_p = 1/f_0$ を代入すると，

$$R(T_p) = \frac{A_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty} A_n^2 \tag{3.3}$$

式 (3.3) は F0 推定にはすべての調波成分が関係することを示す．実際の処理では信号に窓掛けを行うが，基本周期 T_p の整数倍に窓の長さを設定することはできない．F0 は未知であるため，多くの場合 T_p は離散時間領域で整数倍のサンプリング周期ではない．また，多くの場合に信号は雑音が混入している．したがって，式 (3.2) の振幅 A_1 が大きくなれば F0 推定の誤りが少なくなる．この理由の説明を次に述べる．

基本周期 T_p がサンプリング周期の T の整数倍でない次の場合を考える．ここで n と m を整数とする．

$$T_p = nT + \frac{T}{m} \tag{3.4}$$

式 (3.4) から，T_p に近いサンプリング周期の T の整数（n）倍は次式で表せる．

$$nT = T_p - \frac{T}{m} \tag{3.5}$$

T_p の整数（m）倍は次式で表せるように，サンプリング周期 T の整数倍（$mn+1$）になる．

$$mT_p = (mn+1)T \tag{3.6}$$

簡単のため，式 (3.2) を次式で表す．

$$R(\tau) = \frac{1}{2}A_1^2 \cos\omega_0\tau + \frac{1}{2}A_m^2 \cos m\omega_0\tau + R_{dm}(\tau) \tag{3.7}$$

ここで，窓の影響を $W_{af}(\tau)$ とする．

$$\begin{aligned} W_{af}(nT)R(nT) = &\frac{1}{2}A_1^2 W_{af}(nT)\cos\omega_0(T_p - \frac{T}{m}) \\ &+\frac{1}{2}A_m^2 W_{af}(nT)\cos\omega_0(T_p - \frac{T}{m}) \\ &+ W_{af}(nT)R_{dm}(nT) \end{aligned} \tag{3.8}$$

$$\begin{aligned} W_{af}(mT_p)R(mT_p) = &\frac{1}{2}A_1^2 W_{af}(mT_p) \\ &+\frac{1}{2}A_m^2 W_{af}(mT_p) + W_{af}(mT_p)R_{dm}(mT_p) \end{aligned} \tag{3.9}$$

一般に $W_{af}(nT) > W_{af}(mT_p)$ である．式 (3.8) と式 (3.9) において，右辺の第一項の関係が

$$\frac{1}{2}A_1^2 W_{af}(nT)\cos\omega_0(T_p - \frac{T}{m}) > \frac{1}{2}A_1^2 W_{af}(mT_p) \tag{3.10}$$

で，右辺の第二項の関係が

$$\frac{1}{2}A_m^2 W_{af}(nT)\cos m\omega_0(T_p - \frac{T}{m}) < \frac{1}{2}A_m^2 W_{af}(mT_p) \tag{3.11}$$

のときもあり得る．(3.10) と (3.11) の条件のもとで以下の関係を満たす状態を考える．

$$W_{af}(nT)R(nT) > W_{af}(mT_p)R(mT_p) \tag{3.12}$$

であるためには右辺の第一項が第二項より大きければよいから，式 (3.7) の A_1 が大きい程，上式を満たす可能性が高くなる．

$\tau = \frac{T_p}{m}$ でそれが次の整数倍サンプリング周期の場合を考える．式 (3.8) と (3.9) と同様である．

$$W_{af}(T_p)R(T_p) = \frac{1}{2}A_1^2 W_{af}(T_p) + \frac{1}{2}A_m^2 W_{af}(T_p) + W_{af}(T_p)R_{dm}(T_p) \tag{3.13}$$

$$W_{af}(\frac{T_p}{m})R(\frac{T_p}{m}) = \frac{1}{2}A_1^2 W_{af}(\frac{T_p}{m})\cos(\frac{2\pi}{m}) + \frac{1}{2}A_m^2 W_{af}(\frac{T_p}{m}) + W_{af}(\frac{T_p}{m})R_{dm}(\frac{T_p}{m}) \tag{3.14}$$

通常のように $W_{af}(\frac{T_p}{m}) > W_{af}(T_p)$，次の 2 つの関係の成立が可能である．式 (3.13) と (3.14) の右側の最初の部分で，

$$\frac{1}{2}A_1^2 W_{af}(T_p) > \frac{1}{2}A_1^2 W_{af}(\frac{T_p}{m})\cos(\frac{2\pi}{m}) \tag{3.15}$$

となる．そして，それらの次の部分で，

$$\frac{1}{2}A_m^2 W_{af}(T_p) < \frac{1}{2}A_m^2 W_{af}(\frac{T_p}{m}) \tag{3.16}$$

となる．条件式 (3.15) と (3.16) で次の関係を満たす状態を考える．

$$W_{af}(T_p)R(T_p) > W_{af}(\frac{T_p}{m})R(\frac{T_p}{m}) \tag{3.17}$$

したがって，式 (3.7) の A_1 がより大きければ，上の関係を満たす可能性はより高くなる．

最大の振幅 A_m をもつ角周波数を $m\omega_0$ として，$s_m(t)$ のコサイン変調（振幅変調）は次式で表せる．

$$\begin{aligned} s_m(t) &= s(t)\cos m\omega_0 t \\ &= \frac{A_0}{2}\cos m\omega_0 t \\ &\quad +\frac{1}{2}\sum_{n=1}^{\infty} A_n \cos\{(n-m)\omega_0 t+\theta_n\} \\ &\quad +\frac{1}{2}\sum_{n=1}^{\infty} A_n \cos\{(n+m)\omega_0 t+\theta_n\} \end{aligned} \tag{3.18}$$

変調によって式 (3.1) においての，$m\omega_0$ の両隣接調波成分 $(m-1)\omega_0$ と $(m+1)\omega_0$ は式 (3.18) から分かるように ω_0 の成分になる．式 (3.18) から，ω_0 成分の振幅 $A_{m,1}$ は次式で表せる．

$$A_{m,1} = \frac{1}{2}\sqrt{A_{m-1}^2 + A_{m+1}^2 + 2A_{m-1}A_{m+1}\cos(\theta_{m-1}+\theta_{m+1})} \tag{3.19}$$

音声においては，ホルマントの影響で，一般に最大振幅をもつ調波の隣接調波の振幅は相対的に大きい．したがって，$m \neq 1$ のときは次式を満たす場合が多い．

$$A_{m,1} > A_1 \tag{3.20}$$

式 (3.20) を満たさない場合，$m = 1$ のとき，および振幅スペクトルのピークの周波数が調波でない場合を考慮し，F0 推定には次式で表せる信号を用いる．

$$s_m(t) + s(t) \tag{3.21}$$

式 (3.18) の最後の式の第 3 項は，$A_{m-1}/2$ と $A_m/2$ と $A_{m+1}/2$ の振幅をもつ成分がそれぞれ $(2m-1)\omega_0$，$2m\omega_0$，$(2m+1)\omega_0$ であることを示す．雑音を含む場合，振幅スペクトル上での振幅の小さい部分は雑音に埋もれて，調波構造がはっきりしなくなるが，振幅スペクトルのピークの周波数が調波成分であれば，式 (3.21) で表せる信号は調波構造のはっきりした部分が増える．調波構造のはっきりする部分が増えると，式 (3.3) から F0 推定誤りを少なくすることが期待できる．また，調波構造のはっきりした部分を多く含むことは，後で述べる F0 推定に悪影響を及ぼす雑音の推定の精度を高めることが期待できる．これらの効果については実験で示す．

ここではこの変調を周波数領域で行う．また，雑音混入により調波構造がはっきりしない部分が多い高域による悪影響を少なくするために，帯域制限後，変調を施して行う．利用帯域は4.4節の予備実験で決定した．離散時間表現においては帯域制限と式(3.21)の周波数領域表現はそれぞれ以下の処理で用いる．雑音の影響によって$s(t)$が劣化する場合，スペクトルの比較的小さい振幅は雑音に埋もれて，調波構造の明瞭な成分は減少する．そして，振幅スペクトルのピークの周波数が調波成分$m\omega_0$であるなら，$s_m(t)+s(t)$の調波構造の明瞭な成分が増加する．明瞭な部分が増加した場合，F0推定の誤りは減少し，F0推定で悪い影響を与える雑音の推定精度が向上すると考えられる．

離散の周波数領域で，明瞭な調波構造を持っていない高周波部分から悪影響を低減するため，この変調を帯域制限後に行う．

3.2.1 スペクトルの帯域制限

図3.1の処理(a)を説明する．

帯域制限は次式のように表すことができる．

$$S_{\mathrm{B}}(f_h) = S(f_h)B(f_h) \tag{3.22}$$

ここで，$S(f_h)$は雑音の混入によって劣化した$s(t)$に対応する観測信号のスペクトル，$B(f_h)$はローパスフィルタ，$S_{\mathrm{B}}(f_h)$は帯域制限後の観測信号のスペクトルで帯域制限スペクトルと呼ぶことにする．$B(f_h)$は次式のように表される．

$$B(f_h) = \begin{cases} 1, & 0 \le h \le K_{\mathrm{B}}, N-K_{\mathrm{B}} \le h < N \\ 0, & K_{\mathrm{B}} < h < N-K_{\mathrm{B}} \end{cases} \tag{3.23}$$

ここで，K_{B}は次式によって得られる．

$$K_{\mathrm{B}} \quad = \min\{K_{\mathrm{x}}\} \tag{3.24}$$

ここで，$\min\{\cdot\}$は次式のK_{x}の最小値である．

$$\sum_{h=0}^{K_{\mathrm{X}}} \left|S(f_h)\right| \ge b_d \sum_{h=0}^{\frac{N}{2}-1} \left|S(f_h)\right| \tag{3.25}$$

ここで，b_dは予備実験で求めた0.8に設定する．

3.2.2 帯域制限スペクトルの変調

図3.1の処理(b)を説明する.

スペクトルの調波構造の明瞭な帯域を増やすため変調処理を行う. 図3.2に帯域制限スペクトルの変調の概観を示す. $S_{\mathrm{B}}(f_h)$ のコサイン変調後の $s_m(t)$ に対応するスペクトル $S_{\mathrm{M}}(f_h)$ は次式のように表される.

$$S_{\mathrm{M}}(f_h) = \begin{cases} \dfrac{S_{\mathrm{B}}(f_{h+N-K_{\mathrm{M}}}) + S_{\mathrm{B}}(f_{h+K_{\mathrm{M}}})}{2} & , 0 \le h < K_{\mathrm{M}} \\ \dfrac{S_{\mathrm{B}}(f_{h-K_{\mathrm{M}}}) + S_{\mathrm{B}}(f_{h+K_{\mathrm{M}}})}{2} & , K_{\mathrm{M}} \le h < N - K_{\mathrm{M}} \\ \dfrac{S_{\mathrm{B}}(f_{h-K_{\mathrm{M}}}) + S_{\mathrm{B}}(f_{h-N+K_{\mathrm{M}}})}{2} & , N - K_{\mathrm{M}} \le h < N \end{cases} \tag{3.26}$$

ここで, K_{M} は振幅スペクトル $|S_{\mathrm{B}}(f_h)|$ のピークに相当する周波数点を示す.

$s_m(t) + s(t)$ に対応する $Y_1(f_h)$ は, 次式のように表される.

$$Y_1(f_h) = S_{\mathrm{B}}(f_h) + S_{\mathrm{M}}(f_h) \tag{3.27}$$

ここで, $Y_1(f_h)$ を変調スペクトルと呼ぶことにする.

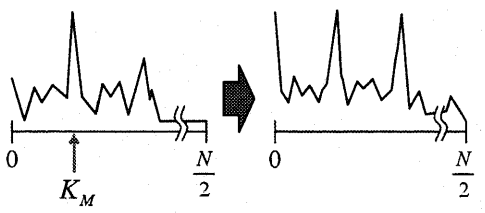


図 3.2: 帯域制限スペクトルの変調の概観

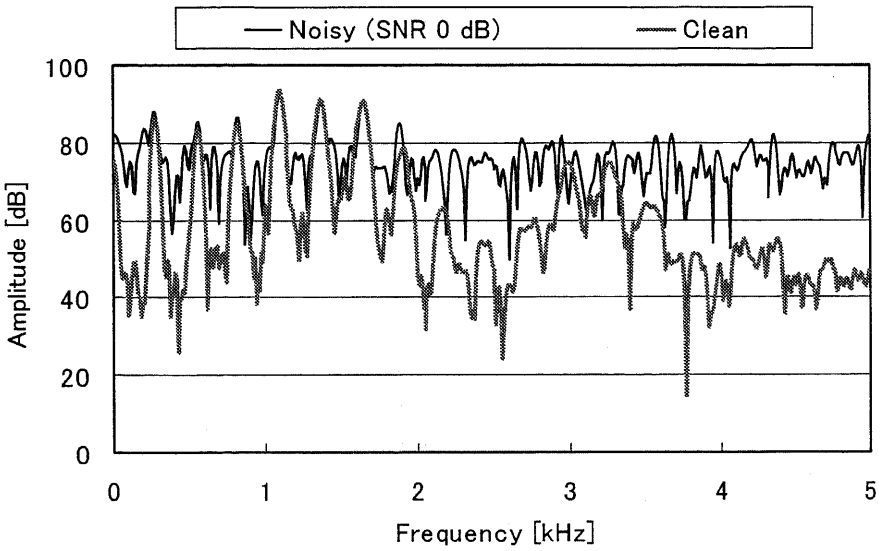


図 3.3: 雑音混入音声とクリーン音声のスペクトル

3.3 粗い雑音推定

図 3.1 の処理 (c) を説明する.

雑音成分の影響を低減するために雑音の情報が必要となる. 事前の雑音情報を用いないため, 観測信号からの推定が必要である. そこで, 以下の想定によって変調後の観測スペクトルから粗い (Rough) 雑音成分を推定する.

有声音のスペクトルには, 調波構造を持つことが知られている. 雑音が混入したスペクトルでは, 隣接した調波間成分の多くが雑音であると想定できる [51, 52]. 図 3.3 は, この仮定の例を示す. 雑音混入音声 (Noisy) のスペクトルとクリーンな音声 (Clean) のスペクトルを比べた場合, 調波成分の大きな山部分では同じような振幅である. しかし, 隣接調波間となるスペクトルの谷部分では, 雑音混入音声スペクトルの帯域の多くがクリーンな音声スペクトルよりも振幅が大きなことを確認できる. そのため, 調波成分が明瞭な低域部分において, 特に隣接調波間成分は雑音による影響であると考えられる. この仮定を基に, 変調スペクトル $Y_1(f_h)$ から Rough 雑音を推定する.

変調後の Rough 雑音スペクトル $Y_2(f_h)$ は，root mean square（RMS）を

$$Y_{\mathrm{rms}} = \sqrt{\frac{1}{N}\sum_{h=0}^{N-1} Y_1^2(f_h)} \tag{3.28}$$

とする 4 ステップ $|Y_1(f_h)|$ の谷部分から推定される．

1. $|Y_1(f_h)|$ の極小点を検出する．
2. 各隣接極小点間を線形補間して，3 点の移動平均を取ることで平滑化した RMS が Z_{av} の $Z_{\mathrm{S}}(f_h)$ を得る．
3. 事前に求めた関数からフレームごとに導かれる雑音パワーと雑音の混入した観測信号パワーの割合を表す，3.5 章で説明される雑音の程度 D_m が用いられる．$Z_{\mathrm{w}} = \sqrt{D_m} \times Y_{\mathrm{rms}}$ による調節は以下のようになる．

$$Z(f_h) = Z_{\mathrm{S}}(f_h) - Z_{\mathrm{av}} + Z_{\mathrm{w}} \tag{3.29}$$

4. 変調観測信号と推定したスペクトルを比較して，次式のように調節する．

$$Y_2(f_h) = \begin{cases} Y_1(f_h), & Z(f_h) > |Y_1(f_h)| \\ Z(f_h)e^{j\theta}, & \text{otherwise} \end{cases} \tag{3.30}$$

ここで，θ は $Y_1(f_h)$ の位相と同様に $|Y_1(f_h)|e^{j\theta}$ と表される．

この方法を用いて推定した Rough 雑音と変調観測信号のスペクトルを図 3.4 に示す．図より，調波成分の部分を大まかに除いて推定していることが確認できる．

3.4 ブラインド信号分離を用いた精度の高い雑音推定

次の ACS では推定された精度の高い雑音から ACF を求める必要がある．そこで，変調スペクトルから推定した Rough 雑音について，BSS を用いることで精度の高い（Fine）雑音を推定する．

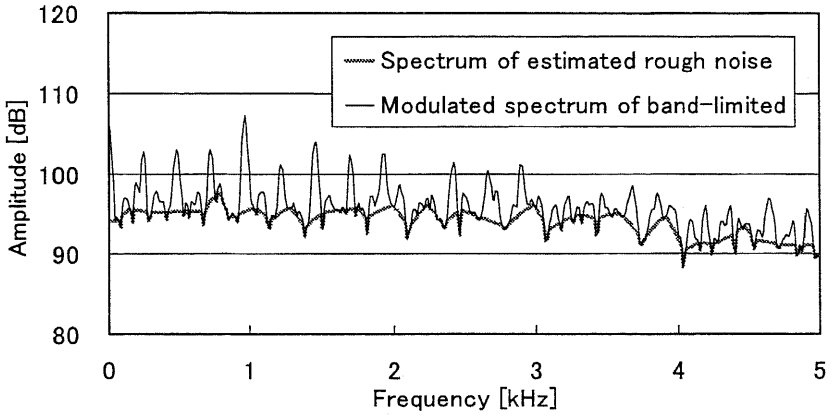


図 3.4: 推定した Rough 雑音と観測信号のスペクトル

3.4.1 ブラインド信号分離の原理

複数の信号源が存在する場合に，そこからの混合した信号を用いて，分離や復元することにより混合する前の信号を推定する場合に BSS の手法がある．これは，音源からマイクロホンまでの伝達特性から分離や復元の推定を行う．しかし，伝達特性が分からない場合が多くある．この混合プロセスが分からない場合における分離や復元が必要になる．一般的な混合プロセスが未知の場合には，観測信号以外に，源信号に関する何らかの事前情報に基づいて逆プロセスを推定することが必要となる．そのため，源信号の持つある性質が分かっていることが条件となる．

ここで，複数の信号源が統計的に独立であると仮定することで行う．

混合過程

2 つの信号源 $S_1(z)$，$S_2(z)$ があり，これを同数のマイクロホンで観測したと想定する．また，既知の混合フィルタを H_{12}，H_{21} とする．このとき，2 つの信号が混合した信号を $X_1(z)$，$X_2(z)$ とする場合，混合プロセスは

$$\begin{aligned} X_1(z) &= S_1(z) + H_{12}(z)S_2(z) \\ X_2(z) &= S_2(z) + H_{21}(z)S_1(z) \end{aligned} \tag{3.31}$$

となる．この原理図を図 3.5 に示す．

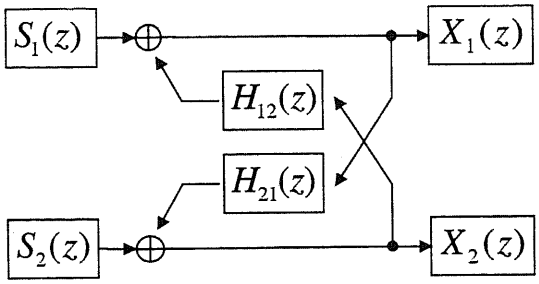


図 3.5: 信号の混合過程

分離過程

混合過程についての逆過程を構築することで，分離を行う．分離フィルタを $W_{12}(z)$，$W_{21}(z)$ とすると，分離プロセスは

$$\begin{aligned} S_1(z) &= X_1(z) - W_{12}(z)X_2(z) \\ S_2(z) &= X_2(z) - W_{21}(z)X_1(z) \end{aligned} \tag{3.32}$$

となる．この原理図を図 3.6 に示す．

ここで，混合過程が既知であれば，元の信号を復元できる．

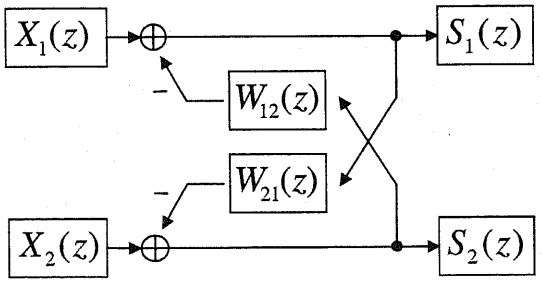


図 3.6: 信号の分離過程

3.4.2 精度の高い雑音推定

図 3.1 の (d) を説明する．

BSS を用いることで，Rough 雑音の推定精度の改善を行う．周波数領域での BSS[48] をこのために用いる．ここでは，変調スペクトル $Y_1(f_h)$ と Rough 雑音

$Y_2(f_h)$ の 2 入力を使った方法である．そのため，シングルチャネルの信号で BSS を用いることができる．この手法によって，より正確に混入した雑音の状態を推定することが可能となる．

雑音の精度を高めるために，先に求めた 1 次雑音と観測信号に対する，2 つの分離フィルタを構成し，BSS 技術を用いて，2 次雑音を求める．2 つの分離フィルタを次に示す．

$$\begin{aligned} \hat{x}_1(n) &= y_1(n) - \sum_k w_{12}(k)\hat{x}_2(n-k) \\ \hat{x}_2(n) &= y_2(n) - \sum_k w_{21}(k)\hat{x}_1(n-k) \end{aligned} \tag{3.33}$$

分離基準として無相関化基準 J を用い，

$$\begin{aligned} J &= \sum_k [\sum_n \hat{x}_i(n)\hat{x}_j(n-k)]^2 \\ &= \sum_k [\sum_n y_i(n) - \sum_k w_{ij}(k)\hat{x}_j(n-k)\hat{x}_j(n-k)]^2 \end{aligned} \tag{3.34}$$

とする．最小二乗平均 (LMS) アルゴリズムを用いると，フィルタの更新式は

$$w_{ij}(m+1,k) = w_{ij}(m,k) + \frac{\mu_i \sum_n \hat{x}_i(n)\hat{x}_j(n-k)}{0.5\sum_n \left\{\hat{x_i}^2(n) + \hat{x_j}^2(n)\right\}} \tag{3.35}$$
$$i,j \in \{1,2\} \quad (i \neq j)$$

となる．ここで，k はフィルタ係数の番号を表す．また，μ_i をここではステップサイズパラメータと呼ぶことにする．G を繰り返し回数とした場合，式 (3.33) と式 (3.35) の周波数領域表現は次式で表せる．

周波数領域 BSS システムの出力は，

$$\begin{aligned} \hat{X}_1(f_h) &= Y_1(f_h) - W_{12}^{(g)}(m,f_h)\hat{X}_2(f_h) \\ \hat{X}_2(f_h) &= Y_2(f_h) - W_{21}^{(g)}(m,f_h)\hat{X}_1(f_h) \quad , \\ & \qquad h = 0,1,...,\frac{N}{2}-1 \end{aligned} \tag{3.36}$$

となる．ここで，h は周波数点，m はフレーム番号，g は各フレームでの反復回数 $g = 0,1,...,G-1$ である．

ここで，重み付けフィルタの更新にLMSアルゴリズムを用いる．次の反復番号について同じフレーム内の重み付けフィルタの更新は，

$$W_{ij}^{(g+1)}(m, f_h) = W_{ij}^{(g)}(m, f_h) + \frac{\mu_i \sum_{n=-h_i}^{h_i} \hat{X}_i(f_{h+n})\hat{X}_j^*(f_{h+n})}{0.5 \sum_{n=-h_i}^{h_i} \left\{|\hat{X}_i(f_{h+n})|^2 + |\hat{X}_j(f_{h+n})|^2\right\}} \quad (3.37)$$
$$i, j = 1, 2 \quad (i \neq j)$$

となる．ここで，μ_i はステップサイズパラメータ，$*$ は共役をそれぞれ表し，更新は

$$W_{ij}^{(0)}(m+1, f_h) = W_{ij}^{(G)}(m, f_h) \quad (3.38)$$

となる．この式 (3.37) は無相関化基準をによって成り立っている．G，μ_1，μ_2，h_1，h_2 は，予備実験によって 2，0.03，0.0003，250，15 にそれぞれ設定される．式 (3.36) の $\hat{X}_2(h)$ は，推定されたFine雑音のスペクトルを示す．このBSSの効果は，予備実験で表される．

3.5 自己相関減算

雑音が影響した信号 $Y_1(h)$ に対応する $y_1(t)$ を次式で表す．

$$y_1(t) = y_c(t) + z(t) \quad (3.39)$$

ここで，$y_c(t)$ はクリーン音声と変調後のクリーン音声からなる音声を示し，$z(t)$ は雑音と変調後の雑音からなる信号を示す．$y_1(t)$ のACF $R_{obs}(\tau)$ は次式で表せる．

$$\begin{aligned} R_{obs}(\tau) &= E[y_1(t)y_1(t+\tau)] \\ &= E[y_c(t)y_c(t+\tau)] + E[y_c(t)z(t+\tau)] \\ &\quad + E[z(t)y_c(t+\tau)] + E[z(t)z(t+\tau)] \end{aligned} \quad (3.40)$$

ここで，$E[\cdot]$ は期待値を示す．信号と雑音が無相関であれば，式 (3.40) からクリーン音声と雑音が互いに無相関であると仮定しているが，変調したクリーン音声と

変調した雑音は，変調は時不変な働きでないので互いに無相関ではないのは明白である．しかしながら，$y_1(t)$ における $z(t)$ の影響を低減させるために，あえて $y_c(t)$ と $z(t)$ が互いに無相関であると仮定する．この仮定に基づく処理の効果は，予備実験で示される．$y_c(t)$ と $z(t)$ が互いに無相関であると仮定した場合，式 (3.40) は，

$$\begin{aligned} R_{\rm obs}(\tau) &= E[y_c(t)y_c(t+\tau)] + E[z(t)z(t+\tau)] \\ &= R_{\rm S}(\tau) + R_{\rm N}(\tau) \end{aligned} \tag{3.41}$$

のように表される．ここで，$R_{\rm S}(\tau) = E[y_c(t)y_c(t+\tau)]$，$R_{\rm N}(\tau) = E[z(t)z(t+\tau)]$ である．式 (3.41) から，対象信号 $y_c(t)$ の ACF $R_{\rm S}(\tau)$ は次式で表せる．

$$R_{\rm S}(\tau) = R_{\rm obs}(\tau) - R_{\rm N}(\tau) \tag{3.42}$$

ここで，式 (3.42) の処理を自己相関関数減算（ACS : autocorrelation subtraction）と呼ぶことにする．式 (3.42) は，$R_{\rm obs}(\tau)$ と雑音の ACF $R_{\rm N}(\tau)$ が分かれば $R_{\rm S}(\tau)$ を求めることができることを示す．$R_{\rm N}(\tau)$ の推定は有声音のスペクトル構造を利用した大まかな雑音推定と後述する雑音の程度を利用したその振幅の調整からなる Rough 雑音スペクトル推定，その Rough 雑音スペクトルを入力とし，もう一方の入力を観測信号スペクトルとする無相関基準に基づく 2 入力ブラインド分離を応用した Fine 雑音スペクトルの推定，そのスペクトルに対する ACF の導出と雑音の程度による ACF の振幅調整からなる．これらの処理の離散時間表現を次に示す．

本研究では，離散時間領域で式 (3.42) の処理を実行する．これは，Rough 雑音スペクトルの推定と BSS による Fine 雑音スペクトルの推定と Fine 雑音スペクトルを ACF へ変換し，ACF の振幅を調節する 4 つの処理で構成された $R_{\rm N}(\tau)$ の推定によって行う．

推定雑音を用いた ACS

図 3.1 の (e) を説明する．

変調スペクトル $Y_1(h)$ を ACF $R_{\rm obs}(i_{\rm d})$ に変換し，推定した Fine 雑音スペクトル $\hat{X}_2(h)$ を ACF $R_{\rm Ns}(i_{\rm d})$ に変換する．音声信号は雑音と無相関であると仮定した

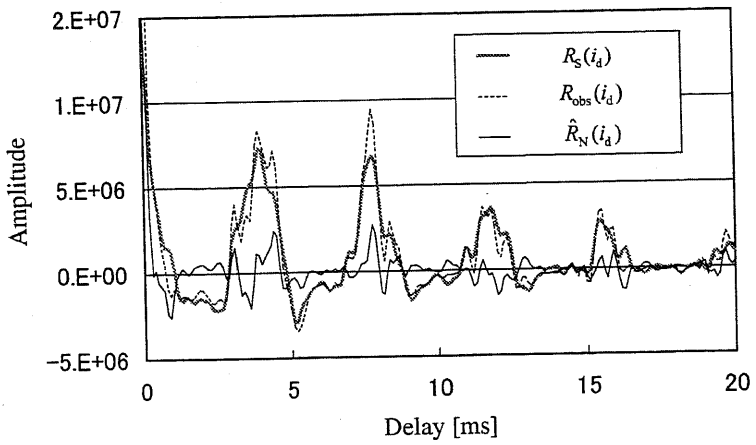


図 3.7: 雑音低減後の ACF $R_S(i_d)$ と変調スペクトルからの ACF $R_{obs}(i_d)$ と Fine 雑音スペクトルからの ACF $\hat{R}_N(i_d)$ （有色雑音混入文章，SNR −5 dB）

場合，$R_S(\tau)$ に対応する雑音低減した ACF $R_S(i_d)$ は，次式で得られる．推定雑音を用いた ACS によって得られる ACF は，

$$R_S(i_d) = R_{obs}(i_d) - \hat{R}_N(i_d) \tag{3.43}$$

で求められる．ここで，

$$\hat{R}_N(i_d) = D_m \frac{R_{obs}(0)}{R_{Ns}(0)} R_{Ns}(i_d) \tag{3.44}$$

となる．式 (3.44) は Fine 雑音の ACF が雑音の程度 D_m によって調節されることを表している．図 3.7 で雑音低減した ACF の例を示す．比較のため同様に変調スペクトルからの ACF と雑音低減した ACF に対応する Fine 雑音スペクトルからの ACF を示す．このフレームの基本周期は 4 ms 付近である．しかし，雑音の混入によって雑音の ACF が 8 ms 付近の遅延位置に大きな振幅を持つと考えられる．そのため，$R_{obs}(i_d)$ は 8 ms 付近の遅延位置に探索範囲内での最大振幅を持っている．推定によって求められた $\hat{R}_n(i_d)$ を用いた ACS で得られた $R_S(i_d)$ は，探索範囲内での最大振幅が基本周期付近に改善できている．このため，$\hat{R}_n(i_d)$ の推定も精度が高いと考えられる．

雑音の程度

ここでは，観測信号から音声に混入している雑音の程度を求めることを考える．そのために用いる手法 [49] を説明する．あらかじめ用意した種々の SNR の雑音混入母音を用いて，雑音パワーと観測可能な雑音を含む信号のパワーの関係を求めておく．この関係（関数）から雑音の程度を推定し，雑音スペクトルの振幅を調整するために用いる．

定常な周期信号で，それが無限に続く場合は，その相関関数 $R(i_{\mathrm{d}}T) = R(i_{\mathrm{d}})$ はその最初の一周期と全く同じ振幅で繰り返す．ここで i_{d} は遅れ数，T はサンプリング周期である．周期を pT，j_n を任意な整数とすると，

$$R(i_{\mathrm{d}}) = R(i_{\mathrm{d}} + j_{\mathrm{n}} \times p) \tag{3.45}$$

と表すことができる．無相関な雑音が含まれるとその ACF は遅延（ラグ）0 の振幅 $R_{\mathrm{obs}}(0)$ が大きくなる．他は変化がない．振幅が大きくなった分は雑音のパワー P_{N} によるものである．

$$P_{\mathrm{N}} = R_{\mathrm{obs}}(0) - R(0) = R_{\mathrm{obs}}(0) - R(j_{\mathrm{n}} \times p) \tag{3.46}$$

この性質を用いて雑音の程度を推定する実験式（関数）を導き出す．$R_{\mathrm{obs}}(0)$ は雑音を含む観測信号のパワー P_{obs} に等しいから，P_{obs} で正規化した式

$$A_m = 1 - \frac{R(0)}{P_{\mathrm{obs}}} = 1 - \frac{R(j_{\mathrm{n}} \times p)}{P_{\mathrm{obs}}} \tag{3.47}$$

は式 (3.46) が成り立つ場合は雑音の程度を表す式になる．実際の処理の場合は信号も有限長で，しかも窓掛け等も行うのが一般的であるので，式 (3.46) は成り立たない．これは真の雑音の程度を表さない．したがって，式 (3.47) を用いた窓掛け等の影響を考慮した真の雑音の程度を表す実験式を導き出す必要がある．信号と雑音が無相関であるとすると，

$$P_{\mathrm{obs}} = P_{\mathrm{S}} + P_{\mathrm{N}} \tag{3.48}$$

と表すことができる．ここで，P_{S} は信号のパワーである．真の雑音の程度を D_{mt} として，

$$D_{mt} = \frac{P_{\mathrm{N}}}{P_{\mathrm{obs}}} = \frac{P_{\mathrm{N}}}{P_{\mathrm{S}} + P_{\mathrm{N}}} \tag{3.49}$$

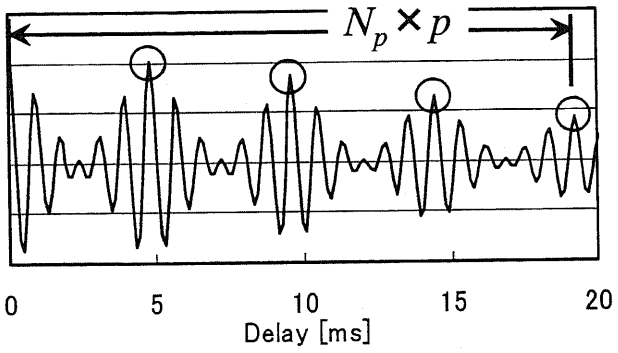


図 3.8: 想定する最大の基本周期 $N_p \times p$ 内の周期間隔ごとの振幅

となる．式 (3.47) の第 2 項はフレーム化された信号でしかも窓掛けされている場合は $j_n \times p$ が大きくなるにしたがって小さくなる．できるだけ信号ごとに異なる周期長の影響が入らないように，想定する最大の基本周期 $T_{p\max}$ ($> N_p \times p$) 内の周期間隔ごとの振幅値の平均を次式により求め利用した．観測信号の ACF を $\hat{R}_{\rm obs}$ としたとき

$$\bar{A_m} = \frac{1}{N_p} \sum_{j_{\rm n}=1}^{N_p} \frac{\hat{R}_{\rm obs}(j_{\rm n} \times p)}{P_{\rm obs}} \tag{3.50}$$

$$B_m = 1 - \bar{A}_m \tag{3.51}$$

とおき，推定した雑音の程度

$$D_m = aB_m - b \tag{3.52}$$

の関係式を女性 3 名，男性 3 名の各 5 母音を利用して，最小二乗法により a と b を求めた．ここで B_m を雑音の程度のパラメータと呼ぶ．B_m は観測信号が得られる度に計算でき，推定した雑音の程度 D_m は式 (3.52) にその B_m を代入することで求めることができる．N_p は想定の最大基本周期までで得られる値の数で，これは図 3.8 の丸印の数で見ることができる．

以上の内容から雑音パワーと観測可能な雑音を含む信号のパワーの関係を求めた．雑音の程度の関係を図 3.9 に示す．

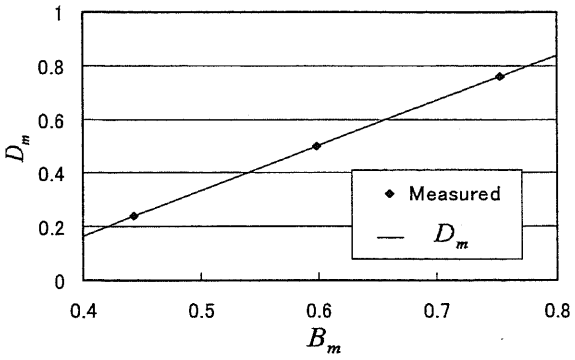


図 3.9: 雑音の程度の関係

3.6 雑音を低減した自己相関関数のコサイン変調

雑音低減したACFのコサイン変調は，基本周期T_pの付近に残っている雑音の影響を低減するために適応される．この変調は，雑音低減したACFのω_0の成分を強調することで，T_pの付近に残っている雑音の影響を低減する．式(3.1)で表せる周期信号が雑音，ここでは簡略化のため，白色雑音$z(t)$に一部埋もれる観測信号$s_{\rm obs}(t)$を近似的に次式で表す．

$$\begin{aligned} s_{\rm obs}(t) = \quad & A_m \cos(m\omega_0 t + \theta_m) \\ & + A_{m+1} \cos\{(m+1)\omega_0 t + \theta_{m+1}\} + z(t) \end{aligned} \tag{3.53}$$

ここで，$\omega_0 = 2\pi f_0$で，f_0はF0を示す．雑音$z(t)$は，必ず白色雑音であるというわけではない．雑音混入信号で，A_mとA_{m+1}がスペクトル$z(t)$のピークより大きな必要がある．式(3.53)は$m \neq 1$のとき，少なくてもはっきりした調波が2つ必要であるという，ここで扱う雑音混入音声のF0推定の一つの条件を含む式を表す．$s_{\rm obs}(t)$のACFは次式で表せる．

$$\begin{aligned} R_{\rm S}(\tau) = \quad & \frac{1}{2} A_m^2 \cos m\omega_0 \tau \\ & + \frac{1}{2} A_{m+1}^2 \cos(m+1)\omega_0 \tau + R_{rsd}(\tau) \end{aligned} \tag{3.54}$$

ここで，$R_{rsd}(\tau)$は$z(t)$のACFを含む残りを示す．

$R_{\rm S}(\tau)$のコサイン変調を次式で表す．

$$R_{\rm M}(\tau) \quad = \quad R_{\rm S}(\tau) \cos m\omega_0 \tau$$

$$
\begin{aligned}
&= \frac{1}{4}A_m^2 + \frac{1}{4}A_{m+1}^2\cos\omega_0\tau + \frac{1}{4}A_m^2\cos 2m\omega_o\tau \\
&\quad + \frac{1}{4}A_m^2\cos(2m+1)\omega_0\tau + R_{rsd}(\tau)\cos m\omega_0\tau
\end{aligned} \tag{3.55}
$$

$m=1$ のとき，振幅スペクトルの第1ピークと第2ピークの周波数が調波成分ではない場合，以下の ACF を用いて F0 推定を行う．

$$
\begin{aligned}
R_{\mathrm{CM}}(\tau) &= R_{\mathrm{S}}(\tau) + R_{\mathrm{S}}(\tau)\cos m\omega_0\tau = R_{\mathrm{S}}(\tau)(1+\cos m\omega_0\tau) \\
&= \frac{1}{4}A_m^2 + \frac{1}{4}A_{m+1}^2\cos\omega_0\tau \\
&\quad + \frac{1}{2}A_m^2\cos m\omega_0\tau + \frac{1}{2}A_{m+1}^2\cos(m+1)\omega_0\tau \\
&\quad + \frac{1}{4}A_m^2\cos 2m\omega_o\tau + \frac{1}{4}A_m^2\cos(2m+1)\omega_0\tau \\
&\quad + R_{rsd}(\tau)(1+\cos m\omega_0\tau)
\end{aligned} \tag{3.56}
$$

式 (3.53) において，右辺の第1と第2項の調波成分がはっきりしていると，式 (3.56) から分かるように，ω_0 にはっきりした成分が現れ，F0 推定の誤りを少なくできる可能性がある．また，$\cos(m+1)\omega_0\tau$ で変調しても同様なことがいえる．$\cos\omega_0\tau$ と $\cos(m+1)\omega_0\tau$ での変調後，これらの調波の差，差が ω_0 でなくても，差の周波数でのコサイン変調は低域の調波成分をよりはっきりさせる．

離散時間におけるこれらの処理を次に述べる．

雑音低減した ACF のコサイン変調の適用

図 3.1 の (f) を説明する．

コサイン変調は，式 (3.54) の $R_{\mathrm{S}}(\tau)$ に対応する雑音低減した ACF に適用される．$m\omega_0$ と $(m+1)\omega_0$ とその間隔に対応する変調のための周波数は，上記のスペクトル $|S_{\mathrm{B}}(h)|$ の第1ピークと第2ピークとその間隔からそれぞれ求められる．これらは，それぞれ f_1，f_2，$f_3 = |f_1 - f_2|$ と示される．f_1 と f_2 が隣接調波に対応するように，次の条件を用いる．

$$
f_3 < f_{\mathrm{c}} \tag{3.57}
$$

ここで，f_{c} は F0 探索範囲内での周波数の上限を示す．本研究では，F0 の探索を

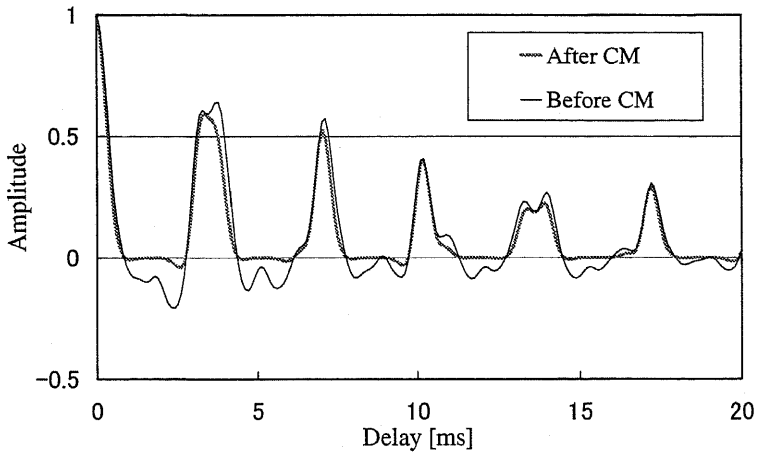


図 3.10: コサイン変調の処理前と処理後の雑音低減 ACF（有色雑音混入文章, SNR −5 dB）

50 Hz から 500 Hz としている．次式で表される変調をコサイン変調とする．

$$R_{\mathrm{CM}}(i_{\mathrm{d}}) = \begin{cases} R_{\mathrm{S}}(i_{\mathrm{d}}) \prod_{j=1}^{3} \{1+\cos(2\pi f_j i_{\mathrm{d}} T)\} \\ \qquad\qquad , f_3 < f_{\mathrm{c}} \\ R_{\mathrm{S}}(i_{\mathrm{d}})\{1+\cos(2\pi f_1 i_{\mathrm{d}} T)\} \\ \qquad\qquad , \text{otherwise} \end{cases} \tag{3.58}$$

ここで，$R_{\mathrm{CM}}(i_{\mathrm{d}})$ はコサイン変調した雑音低減後の ACF を，$R_{\mathrm{S}}(i_{\mathrm{d}})$ は雑音低減した（コサイン変調前）ACF を，T はサンプリング周期をそれぞれ表す．

コサイン変調の後に，平滑化 $R_{\mathrm{CM}}(i_{\mathrm{d}})$ は以下のようになる．

$$R_{\mathrm{CMS}}(i_{\mathrm{d}}) = \sum_{\zeta=-\gamma}^{\gamma} R_{\mathrm{CM}}(i_{\mathrm{d}}+\zeta)/(2\gamma+1) \tag{3.59}$$

ここで，$R_{\mathrm{CMS}}(i_{\mathrm{d}})$ は平滑化後の雑音低減した変調 ACF を示す．

図 3.10 はコサイン変調とその処理前の平滑化した雑音低減 ACF の例を表す．図のフレームは，おおよそ 3.5 ms の基本周期である．図からコサイン変調後の雑音低減 ACF から正しい推定が可能なことが確認できるが，コサイン変調前の ACF（ACS 後の ACF）から 4.0 ms を推定してしまうことを確認できる．この例

のフレームは条件式 (3.57) を満たし，式 (3.58) の f_1，f_2 と f_3 による変調を行っている．

以上の処理を組み合わせた手法が ACS-CM となり，得られた $R_{\mathrm{CMS}}(i_{\mathrm{d}})$ を用いて AUTOC と同様に F0 を算出する．

第4章　雑音混入音声における検証実験

4.1　はじめに

提案したACS-CMのF0推定精度を検証するため，コンピュータを用いたシミュレーション実験を行った．実験で設定した条件と用いた音声サンプルの説明をする．そして，実験で設定するパラメータを決めるため，予備実験を行った．これらの後に，雑音の種類によるそれぞれの結果を示す．また，比較的新しく提案された雑音混入音声のF0推定の中で特に，白色雑音混入に頑健とされる手法と走行自動車内雑音に頑健とされる手法について比較を行い，ACS-CMの有効性と問題点を示す．

4.2　実験条件

実験で用いたパラメータを表4.1に示す．F0の探索範囲については，日本人の平均的な会話音声を考えて，50–500 Hzとした．そこから分析フレーム長について設定している．また，F0の推定率は，基準の±5 %以内を正解とした．ここでの基準とは，クリーンな音声データから目視により求めたものである．また，音声データの無音や有声／無声はあらかじめ目視により判別した．

推定したF0の評価方法と，その評価で用いる基準のF0の求め方を以下で述べる．

評価方法

F0推定の評価には，Gross F0 errorとFine F0 errorを用いた[53].

表 4.1: 実験条件

サンプリング周波数	10 kHz
フレーム長	40 ms
シフト幅 (フレーム周期)	10 ms
窓関数	ハミング窓
DFT(FFT) ポイント数	1024

Gross F0 error は，有声区間から求めた基準の F0 と ±5 % 以上異なる場合のフレーム数から算出する．推定した F0 の $\hat{F}(n)$ と基準 F0 の $F_{\text{true}}(n)$ の差を

$$e(n) = \hat{F}(n) - F_{\text{true}}(n) \tag{4.1}$$

としたとき，有声音区間のフレーム数 N_v において，$|e(n)| \leq 0.05F_{\text{true}}(n)$ の条件を満たすフレーム数 N_e を引いた割合

$$\text{Gross F0 error} = \frac{100(N_v - N_e)}{N_v} \quad [\%] \tag{4.2}$$

によって求められる．このため，F0 が推定できなかったフレームを表す．

Fine F0 error は，正解したフレーム内での標準偏差となるように定義した．F0 が推定できたフレームでの F0 の誤差を表す．

F0 の基準抽出

F0 の推定評価を行うために，基準となる F0 の値が必要である．さらに，有声音のフレームについて評価を行う．そこで，実験に用いるクリーンな音声の時間波形から目視によって判断した．音声休止区間，有声音，無声音の中で，有声音のフレームを分別し，有声音のフレームについては相似な波形を目視で求めて，周期を得た．その周期の逆数を F0 の基準とする．

表 4.2: 文章 1 の各話者の発話時間

話者	女性 A	女性 B	女性 C	男性 A	男性 B	男性 C
発話時間 [s]	5.96	5.72	6.10	5.72	6.52	5.94

表 4.3: 文章 2 の各話者の発話時間

話者	女性 D	女性 E	女性 F	女性 G	女性 H	女性 I
発話時間 [s]	3.20	3.80	3.04	3.83	3.03	3.75
話者	男性 D	男性 E	男性 F	男性 G	男性 H	男性 I
発話時間 [s]	2.68	2.50	3.45	2.72	2.26	2.71

4.3 音声サンプル

提案法の有効性を検討するため，音声サンプルでの実験を行った．傾向やパラメータなどを調べるため単母音と文章 1 を用いて，さらに文章 2 を用いた実験を行った．

- 単母音：/a/, /i/, /u/, /e/, /o/
- 文章 1:”創造とは，今までにない新しいものをつくり生みだすことである．”
- 文章 2：”静岡では，まもなく天気が回復するでしょう．”

各話者が発話した文章 1 の発話時間を表 4.2 に，文章 2 の発話時間を表 4.3 にそれぞれ示す．文章 1 の平均発話時間は 5.99 s で，文章 2 の平均発話時間は 3.08 s である．実音声の詳細を以下で説明する．

実音声

実際に人が発する音声は，合成音のように定常な周期をもつ音声ではない．そこで，実音声についての検討を行う．実験に用いた実音声サンプルで，日本語の単母音と”創造とは，今までにない新しいものをつくり生みだすことである．”

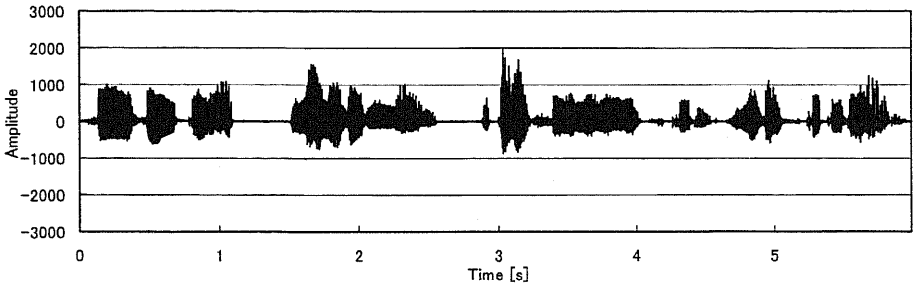
表 4.4: 実音声のフレーム数と F0 平均

		母音	文章 1	文章 2
フレーム数	女声	280	1163	1292
	男声	310	1070	801
F0 平均 [Hz]	女声	267	236	240
	男声	132	123	149

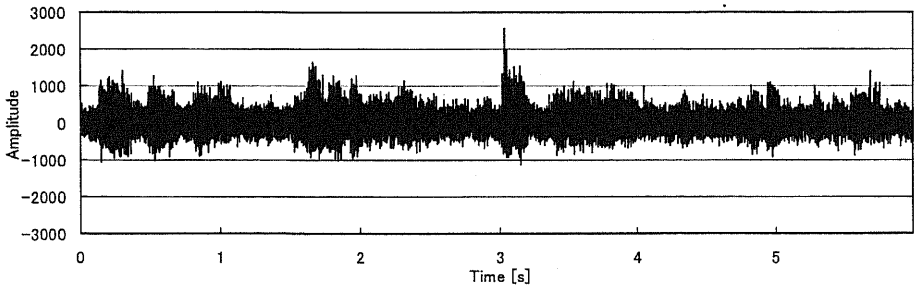
(/soozoo.../) は，無響室で 1 インチマイクロホンを用いた録音である．そして，45 Hz の遮断周波数を持つハイパスフィルタと，4.5 kHz の遮断周波数を持つローパスフィルタに通し，12 bit で量子化したデータを研究室のデータベースから得た．実験には，サンプリング周波数が 10 kHz の男女各 3 名が発声した 6 名の音声を用いた．また，実験の精度を上げるためさらに別の話者によって発話された文章の音声サンプルも用いた．"静岡では，まもなく天気が回復するでしょう．" (/sizuoka.../) を男女各 6 名の計 12 名がそれぞれ発話した文章である．音声サンプルは音声資源コンソーシアム [54] のデータベースを利用した．ダイナミック型の単一指向性のマイクロホン，三研 MU-2C によって収録され，発話者の口とマイクロホンの間の距離は約 20 cm である．その中で有声音として扱った女声と男声の各フレーム数，F0 平均を表 4.4 に示す．実験に用いた音声サンプルの F0 平均は，女声と男声ともに日本人の平均的な F0 に近いことを表から確認できる．

これらの実音声を対象とした実験を行った．また，音声サンプルの音声休止区間や有声／無声については，あらかじめ目視によりフレームごとの判別をした．

実音声の文章 1 と文章 2 のクリーンな波形と，文章 1 に白色雑音を付加した波形と文章 2 に走行自動車内雑音を付加した波形をそれぞれ図 4.1 と図 4.2 に示す．実音声についても，音声の振幅が大きな部分と音声休止区間付近ともに雑音が影響していることを確認できる．

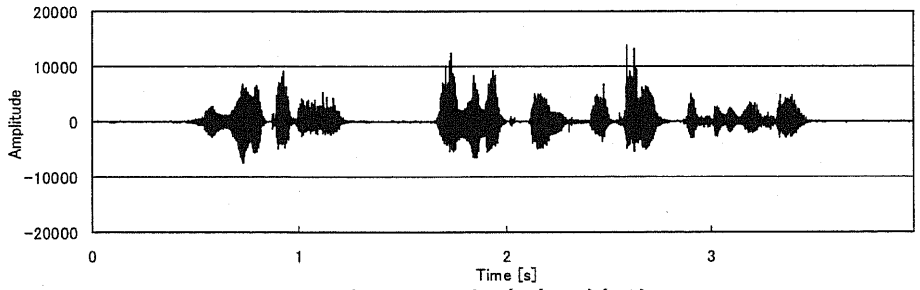


(a) クリーンな音声の波形

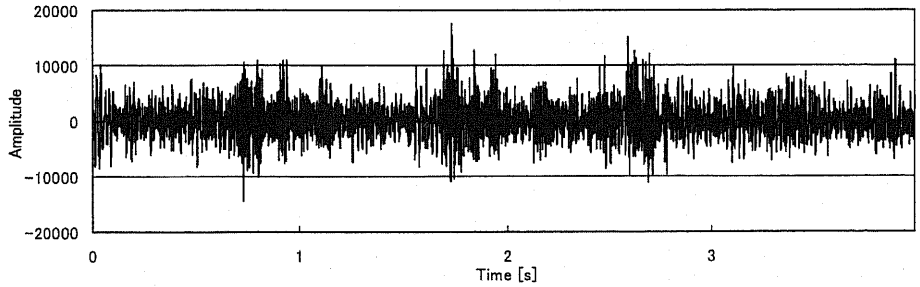


(b) 有色雑音混入音声の波形 (SNR 0 dB)

図 4.1: 男声/soozoo.../の波形



(a) クリーンな音声の波形



(b) 走行自動車内雑音混入音声の波形 (SNR 0 dB)

図 4.2: 男声/sizuoka.../の波形

付加雑音

雑音混入音声のサンプルは，クリーンな音声に雑音を付加することで準備した．

SNR B_{K} は，雑音に対する変数 K を用いることで決めることができる．ここで，目的信号のパワーを P_{S}，雑音信号のパワーを P_{N} とすると観測信号は

$$B_{\mathrm{K}} = 10\log_{10}\left(\frac{P_{\mathrm{S}}}{K^2 P_{\mathrm{N}}}\right) \tag{4.3}$$

となる．そこで，変数 K を次式から求める．

$$K = \sqrt{\frac{P_{\mathrm{S}}}{10^{\frac{B_{\mathrm{K}}}{10}} P_{\mathrm{N}}}} \tag{4.4}$$

この変数 K を用いた雑音信号 Y_{N} を音声サンプルに付加することで，観測信号を生成すると

$$X(z) = S(z) + KY_{\mathrm{N}}(z) \tag{4.5}$$

となる．P_{S} は，あらかじめ目視によって求めたクリーンな文章の有声区間開始部分から終了部分までを用い，間に音声休止区間も含んでいる．そのためフレームごとで考えた場合，各フレームでの音声信号の振幅によって実際の SNR は異なる．

雑音の影響が比較的大きい SNR については，F0 推定が困難になる．そこで SNR については，0 dB，-5 dB に主な重点を置いた．SNR が -5 dB については，雑音の影響が大きいため極端に F0 推定が困難になる．

4.4 予備実験

ACS-CM の各パラメータは予備実験によって決定し，各処理の有効性を検討した．ここでは，4.3 節の各母音にコンピュータで生成した白色雑音を付加したデータを用いた．

単母音は，コンピュータで合成した白色雑音によって劣化させた．独立に 2000 回合成した雑音をそれぞれに混入させた各母音について，2000 データを用いた．それらのデータで，それぞれ中央フレームで信号の SNR が -5 dB か 0 dB となるように設定した中央フレームについてを予備実験で用いた．ここでの予備実験の評価は，Gross F0 error に着目した．

表 4.5: 帯域制限パラメータ b_{d} による Gross F0 error

b_{d}	1.0	0.8	0.6
Gross F0 error [%]	21.18	17.30	17.81

表 4.6: K_{M} の調波成分の割合 [%]

	Harmonics						
dB	1st	2nd	3rd	4th	5th	Others	Rest
0	19.93	30.57	12.15	23.25	3.33	3.41	7.38
−5	20.05	30.21	11.43	22.34	3.19	4.08	8.72

表 4.5 で，式 (3.25) の帯域制限パラメータ b_{d} による Gross F0 error を表す．予備実験の結果から ACS-CM の b_{d} は，0.8 に設定する．表 4.6 は，上記の単母音データで調波成分となる式 (3.26) の K_{M} の周波数について割合を示す．ここで，この周波数は，K_{M} の間隔の周波数の調波成分と調波が調波の 5 %以内の調波成分であると考えられる．表 4.6 は，これらの 90 %以上が調波成分であることを表す．

表 4.7 にコサイン変調と ACS の有効性を示す．結果は，SNR が −5 dB の雑音混入単母音の場合である．表 4.7 で，ケース (1) の (a)+(b) は，図 3.1 で (a) と (b) の処理を実行し，(c) と (d) と (e) と (f) の処理を実行しないことを意味する．AUTOC でのケース (1) の比較結果によって，帯域制限スペクトルの変調が有効であることを確認できる．ケース (2) の結果から，ACF のコサイン変調が有効であることを確認できる．ケース (3) の結果から，それらのコサイン変調が有効であることを確認できる．AUTOC でのケース (4) の結果から，ACS が有効であることを確認できる．ケース (4) と (5) の結果から，BSS が有効であることを確認できる．ケース (6) と (7) と (8) の結果から，ACS-CM でそれぞれ帯域制限のスペクトルの変調，雑音低減した ACF のコサイン変調，BSS が有効であることを確認できる．

表 4.7: ACS-CM の各処理の効果についての実験結果で求められた Gross F0 error [%]. ACS-CM：(a)+(b)+(c)+(d)+(e)+(f)，ケース (1)：(a)+(b)，ケース (2)：(f)，ケース (3)：(a)+(b)+(f)，ケース (4)：(c)+(e)，ケース (5)：(c)+(d)+(e)，ケース (6)：(c)+(d)+(e)+(f)，ケース (7)：(a)+(b)+(c)+(d)+(e)，ケース (8)：(a)+(b)+(c)+(e)+(f).

ケース	女声	男声	平均
AUTOC	35.21	22.60	28.90
ACS-CM	21.67	12.93	17.30
(1)	34.52	18.36	26.44
(2)	32.10	17.90	25.00
(3)	30.81	17.03	23.92
(4)	27.93	20.99	24.46
(5)	27.27	19.59	23.43
(6)	25.02	14.66	19.84
(7)	25.89	14.11	20.00
(8)	23.04	13.97	18.50

4.5 合成雑音混入音声の場合

付加する雑音の種類については，白色雑音と，スペクトルの傾きを持つ雑音を用いた．ここで，高域で約 −5 dB/oct の傾きを持つ信号を有色雑音と呼ぶことにする．

白色雑音

白色雑音とは，観測周波数帯域において，均一なパワースペクトルをもつ雑音である [50]．実験に用いた白色雑音は，ボックス・ミューラーの正規乱数発生法

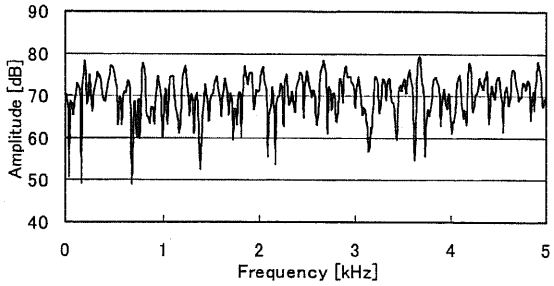


図 4.3: 白色雑音のスペクトル

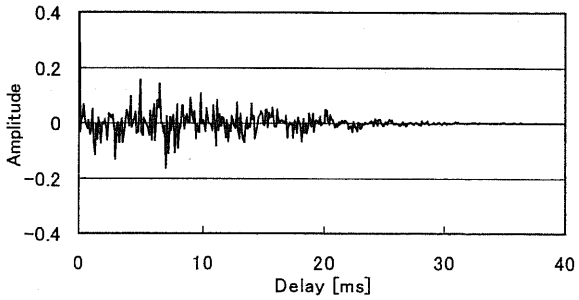


図 4.4: 白色雑音の ACF

で生成することで得た．ここで，r を一様分布としたとき正規乱数は，

$$\begin{aligned} r^2 &= -2\log r_1 \\ \theta &= 2\pi r_2 \end{aligned} \tag{4.6}$$

より，2つの一様乱数 r_1，r_2 から互いに独立に標準正規分布に従う2つの正規乱数

$$\begin{aligned} z_1 &= \sqrt{-2\ln r_1}\cos\left(2\pi r_2\right) \\ z_2 &= \sqrt{-2\ln r_1}\sin\left(2\pi r_2\right) \end{aligned} \tag{4.7}$$

を求める．

実験に用いた白色雑音のスペクトルを図 4.3 に示す．すべての帯域で，ほぼスペクトルが偏在しないことを確認できる．また，図 4.3 の ACF について図 4.4 に示す．白色雑音については周期性が無いため，顕著な振幅が見られない．遅延 0 付近では，大きな振幅を持っていることを確認できる．

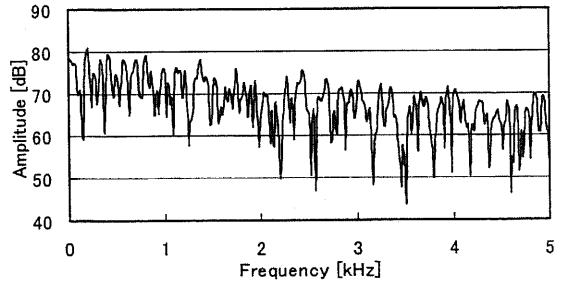


図 4.5: 有色雑音のスペクトル

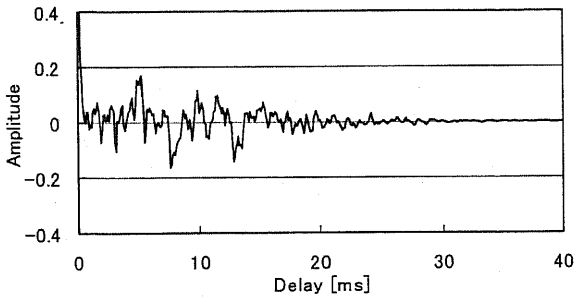


図 4.6: 有色雑音の ACF

有色雑音

有色雑音とは，パワースペクトルが周波数軸上で傾きを持つ雑音である．実験に用いた有色雑音は，白色雑音を次式 (4.8) のフィルタに通して生成した高域で約 −5 dB/oct の傾きを持つ信号である．

$$C(z) = \left[\frac{1 - e^{-\pi B_1 T} z^{-1}}{1 - e^{-\pi B_2 T} z^{-1}}\right]^4 \tag{4.8}$$

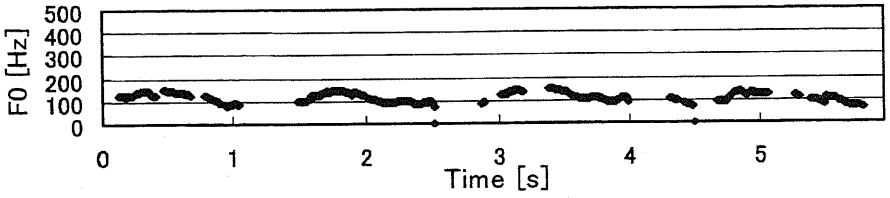
ここで，B_1 は 10000 Hz，B_2 は 5000 Hz，T は 0.0001 s である．この雑音は，周波数軸上で傾きを持つため，白色雑音と比べて実環境中に生じる雑音に近いと考えられる．実験に用いた有色雑音のスペクトルを図 4.5 に示す．図 4.3 と比べて図 4.5 のスペクトルは傾きを確認できる．また，図 4.5 の ACF を図 4.6 に示す．白色雑音の ACF（図 4.3）と比べて，遅延 0 付近以外にも比較的大きな振幅を確認できる．

実験結果

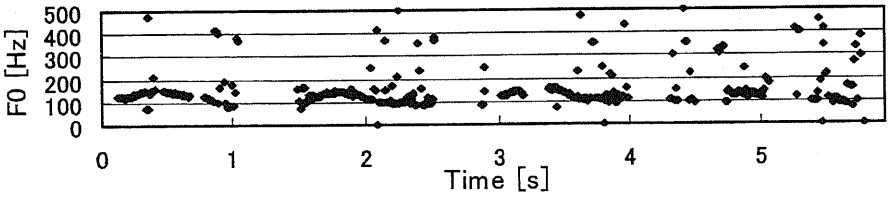
図 4.7 に図 4.1 の文章/soozoo.../の F0 推定結果の男声例（約 6 s）を示す．フレームごとに推定された F0 を点印で表している．目視で得た無声や音声休止区間のフレームは，点印が無い部分である．図 4.7(a) は AUTOC で推定されたクリーンな音声の F0 結果で，Gross F0 error は約 99.3 %である．図 4.7 の (b) と (c) は，それぞれ AUTOC と ACS-CM で推定された雑音混入音声の F0 結果である．AUTOC と比べて ACS-CM の誤りが少ないことを確認できる．しかし，どちらも語頭や語尾で多くの推定誤りがある．明瞭な周期波になり難い部分であることや音声波形の振幅が小さな部分のため，雑音の影響が大きいと考えられる．これは，図 4.7(d) の参照によって得られた True の D_m からも確認できる．図 4.7(d) から ACS-CM で用いている D_m は，おおよそ推定できていることが確認できる．

図 4.8 に ACS-CM と AUTOC と BPPAS と CEPST から得られる話者 6 名の Gross F0 error と Fine F0 error を示す．ここで，推定した各フレームごとの F0 を用いて評価した．つまり，各フレームで F0 推定された値を用いた平滑化は行っていない．また，BPPAS は [45] において，F0 探索範囲を 50–400 Hz に設定されている．そのため，同様の探索範囲で行った BPPAS(50–400 Hz) についても比較する．BPPAS と BPPAS(50–400 Hz) の雑音は，あらかじめ得た音声休止区間を用いて推定している．CEPST での F0 推定は，白色雑音の影響によって精度が良くないことを確認できる．白色雑音混入音声での Gross F0 error では，ACS-CM と BPPAS で大きく精度向上となっている．Fine F0 error では SNR が −5 dB の場合に，ACS-CM で精度向上となっている．Gross F0 error と Fine F0 error とも ACS-CM の推定精度が高いことを確認できる．

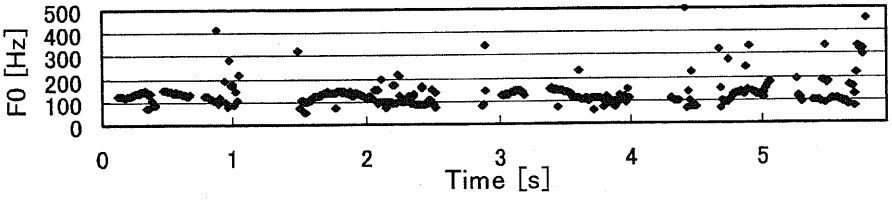
合成した白色雑音混入と有色雑音混入の場合に F0 推定精度の向上が確認できた．このため，合成雑音混入において ACS-CM は有効であると考えられる．そこで次に，実環境で収録された雑音を実雑音として実験する．



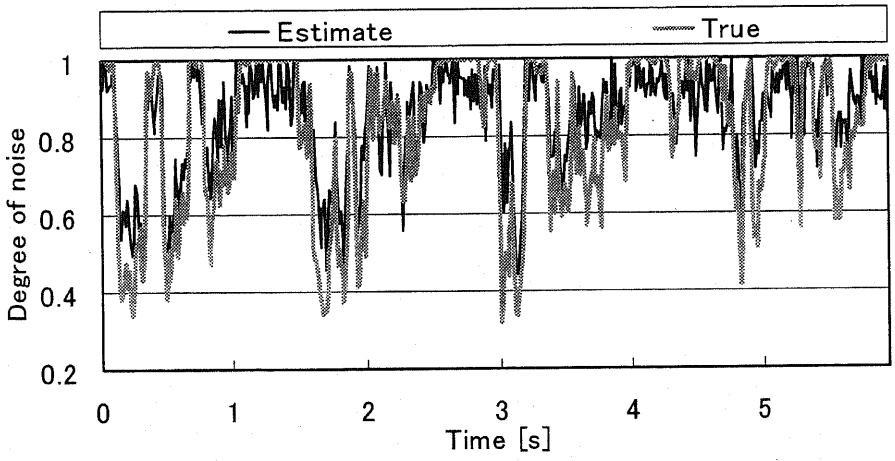
(a) AUTOC (クリーン音声)



(b) AUTOC (雑音混入音声)

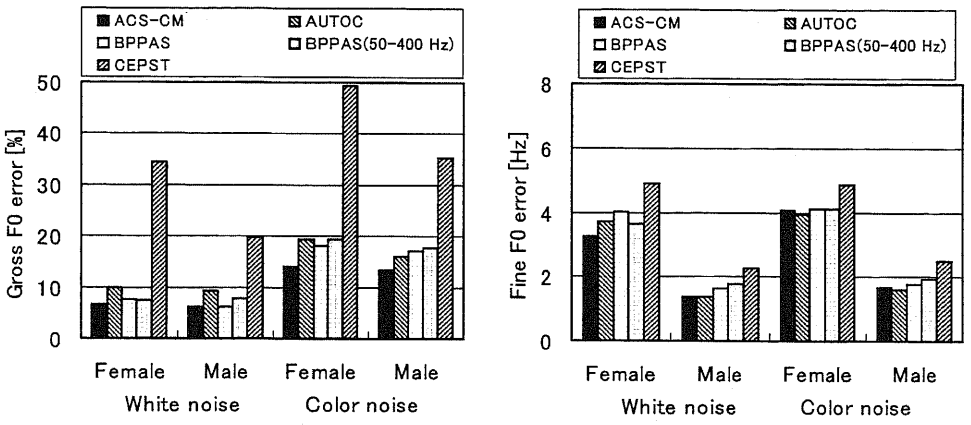


(c) ACS-CM (雑音混入音声)



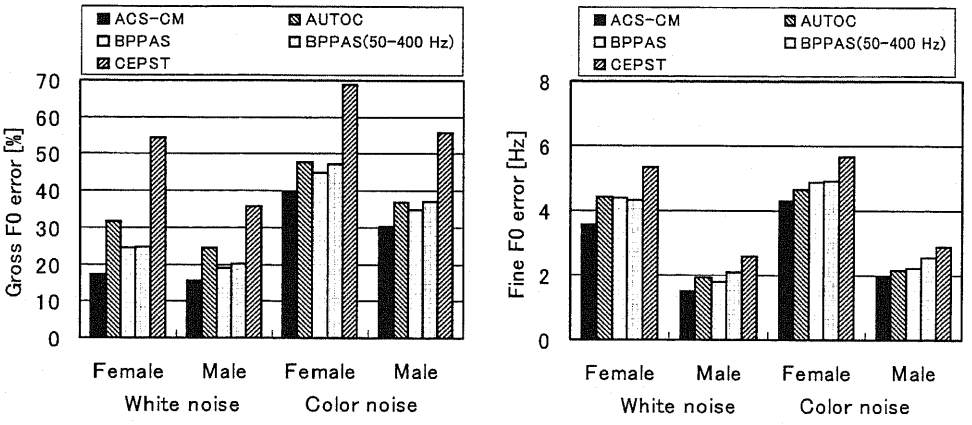
(d) 推定された雑音の程度 D_m

図 4.7: AUTOC と ACS-CM から得られた F0 推定の結果（図 4.1 と同一の音声サンプル）



Gross F0 errors　　Fine F0 errors

(a) SNR 0 dB



Gross F0 errors　　Fine F0 errors

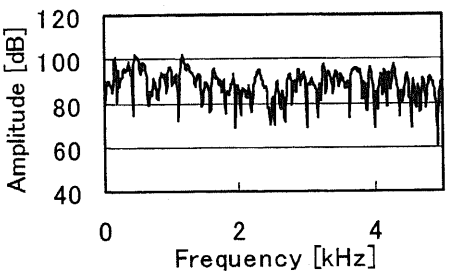
(b) SNR −5 dB

図 4.8: 白色雑音混入音声と有色雑音混入音声の場合のそれぞれの Gross F0 error と Fine F0 error

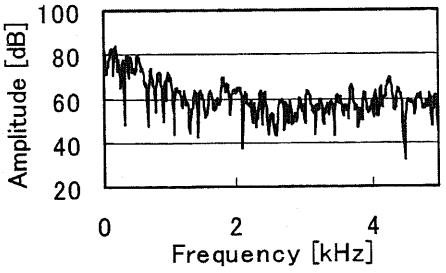
4.6 実雑音混入音声の場合

実雑音混入音声の実験について示す.

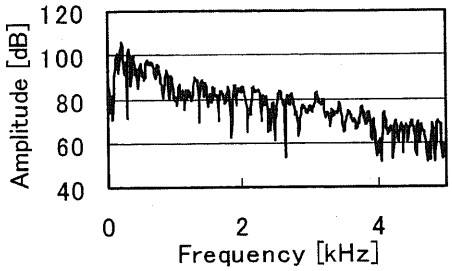
実環境の雑音データは，データベース（www.milab.is.tsukuba.ac.jp/corpus /noise_db.html）を利用した．実験に用いた実雑音データの 1 フレームのスペクトル例を図 4.9 でそれぞれ示す．時間によっても変化があるためフレームによる



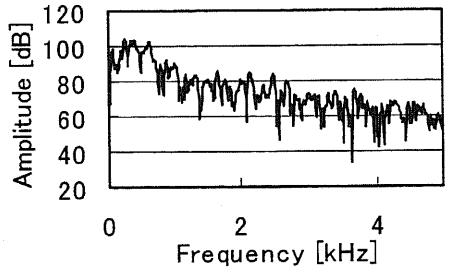
(i) 工場（板金）



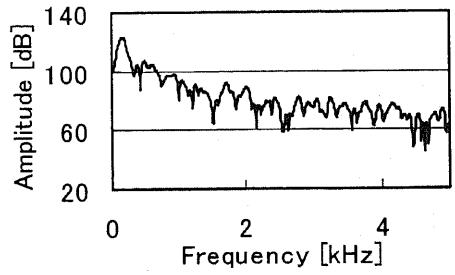
(ii) 仕分け処理場



(iii) 計算機室（ワークステーション）



(iv) 展示会場（ブース内）



(v) 走行自動車内

図 4.9: 実雑音のスペクトル

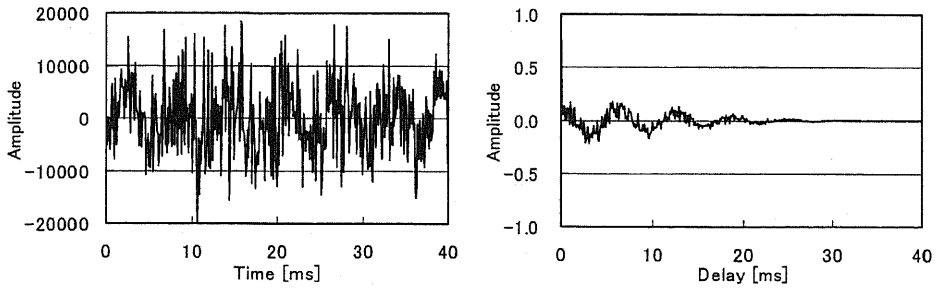
違いもあるが，特定の周波数成分が強く見られる雑音もある．

図 (i) のスペクトルは，どの帯域にも同じようなパワーを持っているため，比較的白色雑音に近いと考えられる．また，(ii) では低域にある程度大きなパワーを持ち，(iii) と (iv) は全体的に傾きを持ったスペクトルである．(v) では低域に顕著なパワーを持つスペクトルで，他と比べて特定の周波数帯域に大きなパワーが偏在した雑音である．

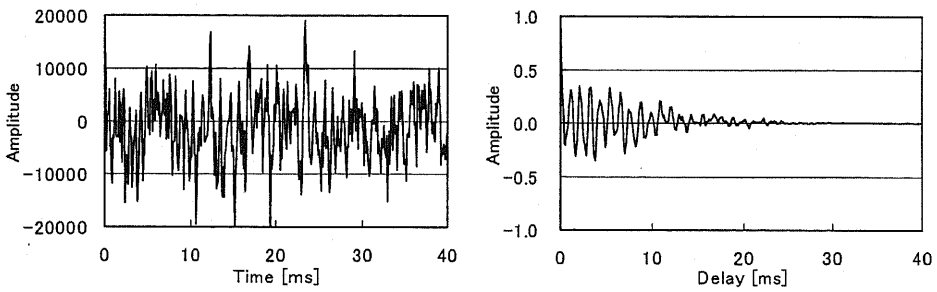
図 4.9 の雑音の時間波形と ACF を図 4.10 でそれぞれ示す．雑音の種類によっ

て波形の振幅に違いがあり，図 (i) の ACF で遅延 0 以外の振幅は小さく，余り相関が確認できない．(ii) では，振幅に細かな起伏が確認でき，(i) と比べて相関を持つ遅延がある．(iii)，(iv) と (v) では，特定の遅延で比較的大きな振幅が確認でき，(v) の遅延が 0 付近の F0 探索範囲にも大きな振幅が確認できる．

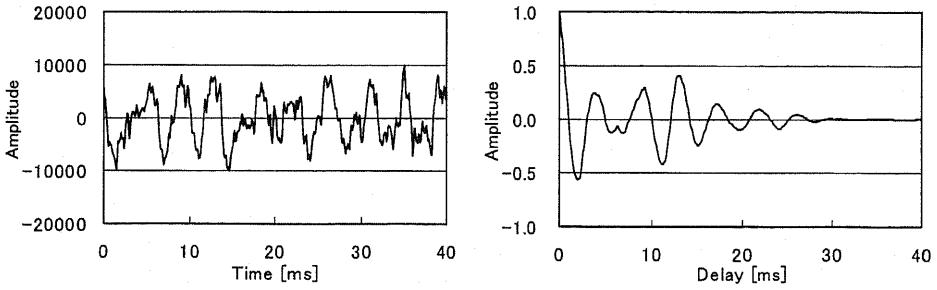
各文章（/shizuoka…/）にそれぞれの雑音を付加した音声サンプルを実雑音混入音声として，次の実験で用いた．



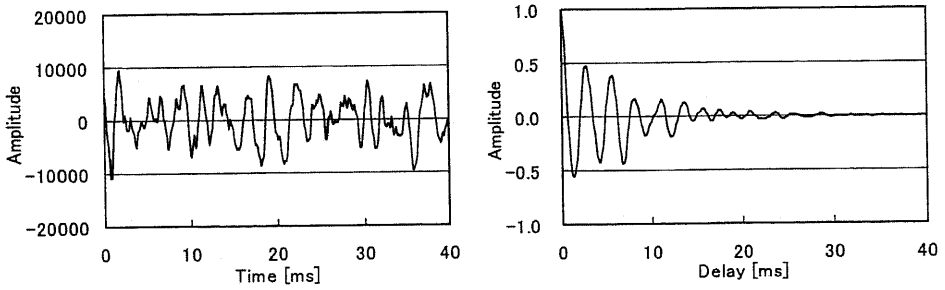
(i) 工場（板金）



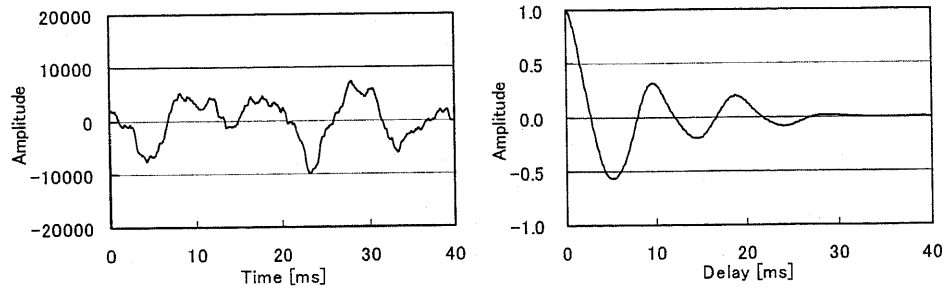
(ii) 仕分け処理場



(iii) 計算機室（ワークステーション）



(iv) 展示会場（ブース内）



(v) 走行自動車内

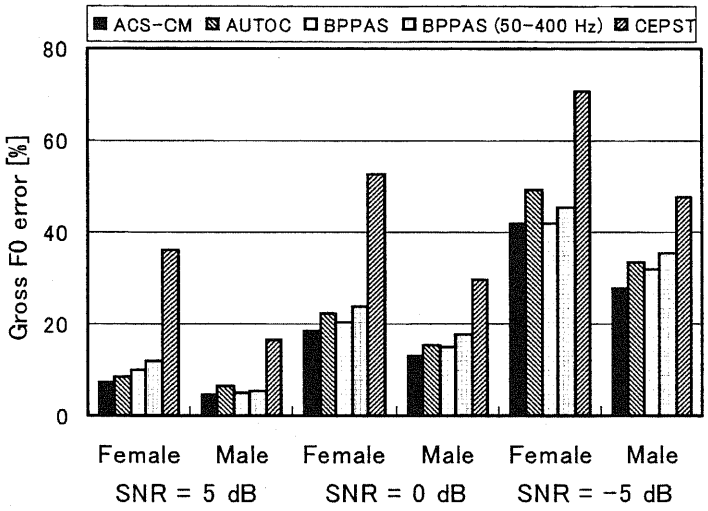
図 4.10: 実雑音の波形（左）と ACF（右）

実験結果

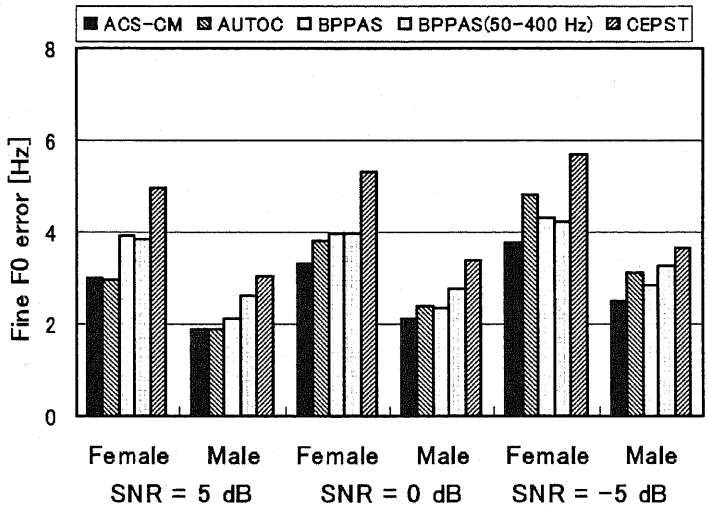
話者 12 名（女性 6 名，男性 6 名）がそれぞれ発話した文章/shizuoka.../の Gross F0 error と Fine F0 error を図 4.11 で以下の雑音が混入した音声についてそれぞれ示す．

- (a) 工場（板金）
- (b) 仕分け処理場
- (c) 計算機室（ワークステーション）
- (d) 展示会場（ブース内）

この音声サンプルは，予備実験とは異なる話者が発話した文章である．図から ACS-CM の F0 推定結果を確認できる．比較のため AUTOC と BPPAS と BPPAS(50–400 Hz) と CEPST の結果についても示す．他の F0 推定と比較して ACS-CM のほ



Gross F0 errors

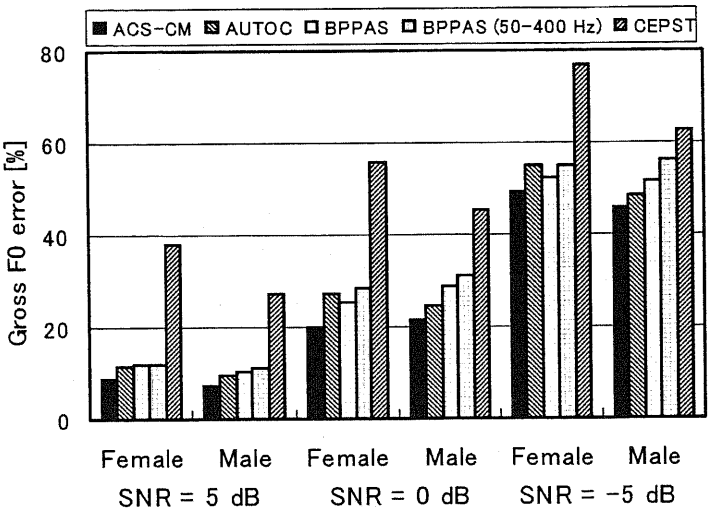


Fine F0 errors

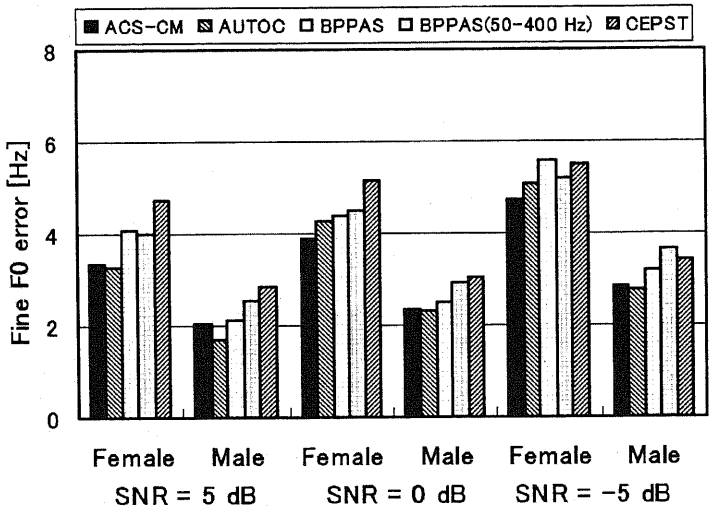
(a) 工場（板金）

とんどのGross F0 errorが小さなことが分かる．特に図4.11(a)男声のSNR=−5 dBの場合では，他のF0推定のGross F0 errorと比べて改善が確認できる．

しかし，図4.11(c)では他の方法と比べてGross F0 errorの大きな改善は確認できなかった．計算機室（ワークステーション）の雑音スペクトルでは，低域の



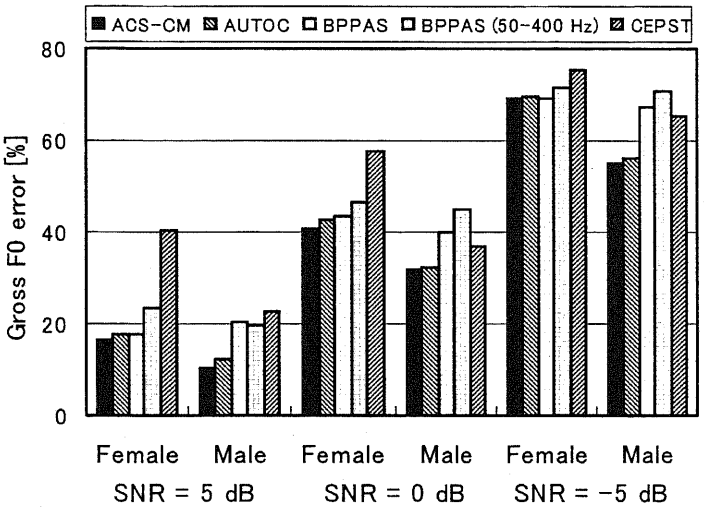
Gross F0 errors



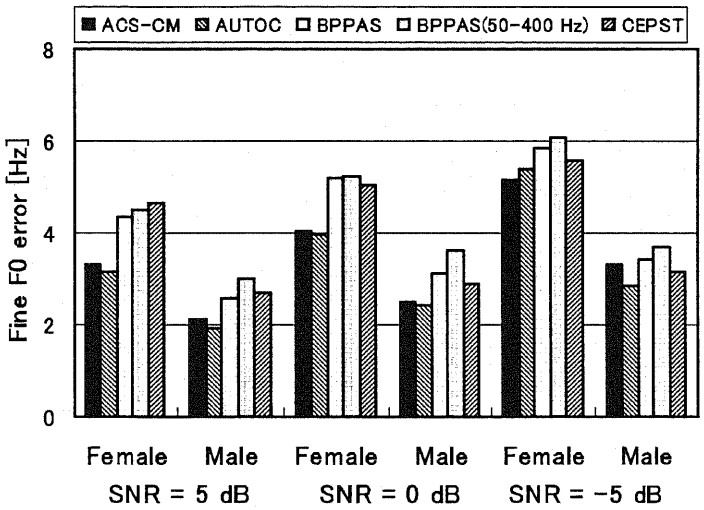
Fine F0 errors

(b) 仕分け処理場

周波数でスペクトルが大きなパワーを持っている．そのためACS-CMで用いられる第1ピークと第2ピークの周波数の多くが，調波成分ではない可能性が考えられる．また，SNR=5 dBで比較した場合にACS-CMよりAUTOCのFine F0 errorが小さい結果となった．



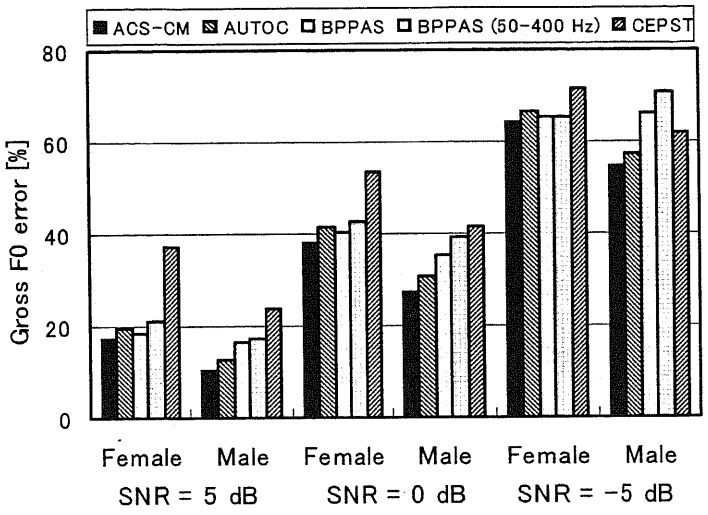
Gross F0 errors



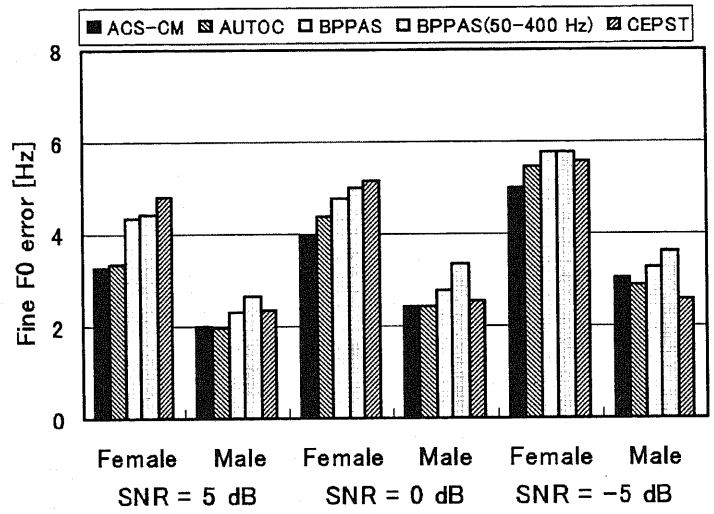
Fine F0 errors

(c) 計算機室（ワークステーション）

実験結果から，特定の帯域に雑音が偏在し大きなパワーを持つ雑音混入音声の場合にACS-CMのF0推定で良くない傾向が確認できる．そこで，走行自動車内などのスペクトルが偏在する帯域を持つ雑音混入音声で，F0推定精度が高いとされるPHIA[41]とEWPH[46]について比較実験を行った．



Gross F0 errors



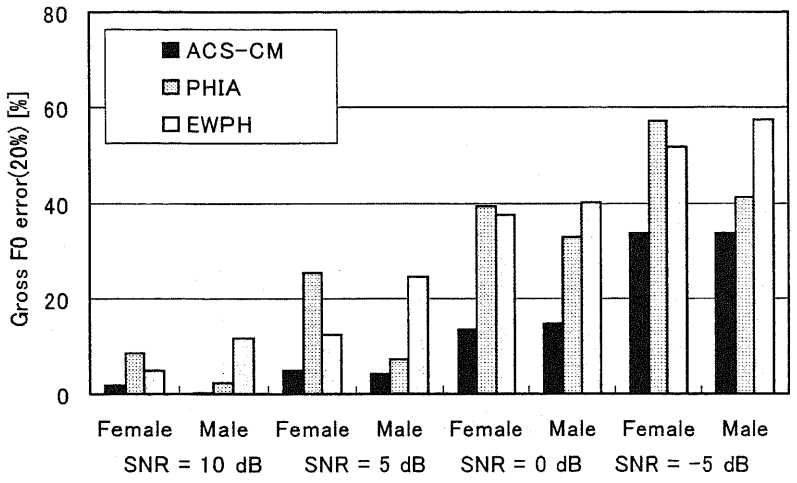
Fine F0 errors

(d) 展示会場（ブース内）

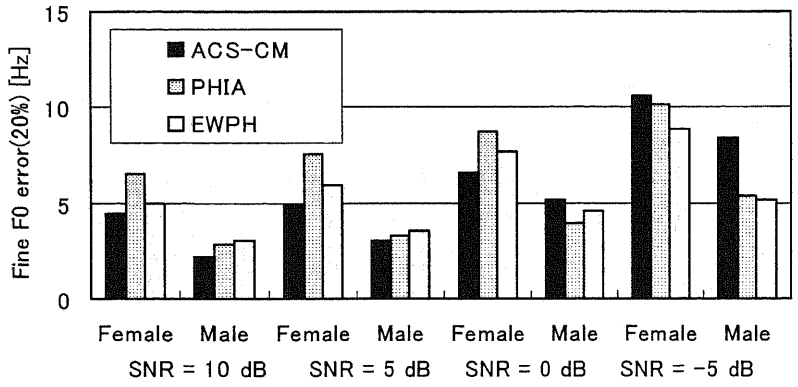
図 4.11: 文章/shizuoka.../の Gross F0 error と Fine F0 error

この F0 推定の結果を図 4.12 で以下の雑音が混入した音声についてそれぞれ示す．ここでの評価は，比較のために基準 F0 の 20 %以内を正解とする．

- (a) 工場（板金）



Gross F0 errors

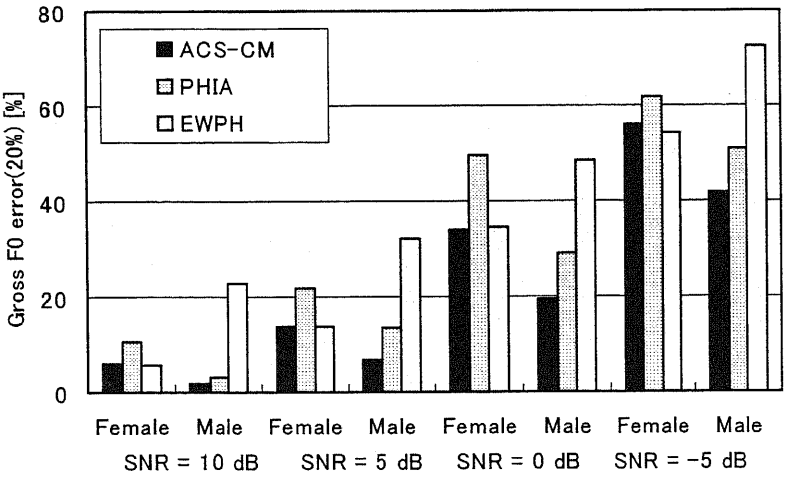


Fine F0 errors

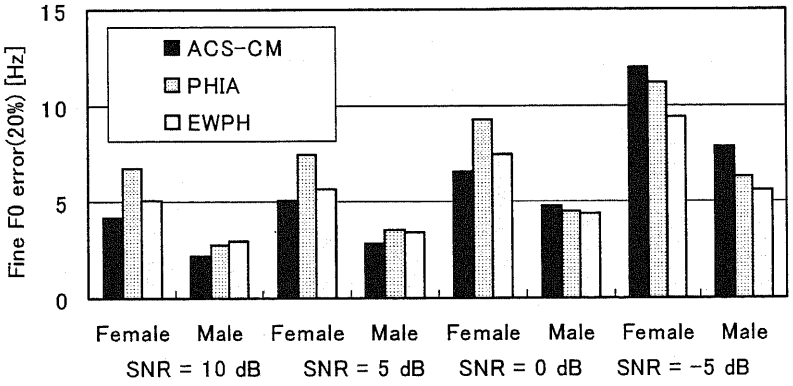
(a) 工場（板金）

- (b) 展示会場（ブース内）
- (c) 走行自動車内

PHIA や EWPH は，比較的雑音の影響が少ない場合にも F0 推定精度が高いとされるため，SNR を 10 dB から −5 dB までの 5 dB ごととした．また，PHIA の定 Q-gammatone フィルタバンクを 7 チャネルにして中心周波数を約 2 kHz から約 4.5 kHz に設定した．そして，EWPH の定 Q-gammatone フィルタバンクを 60 チャネルにして中心周波数を 60 Hz から約 4.5 kHz に設定した．



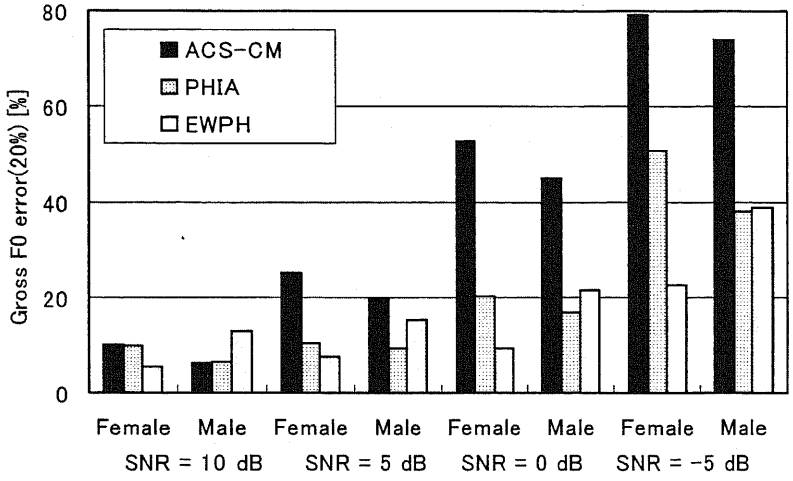
Gross F0 errors



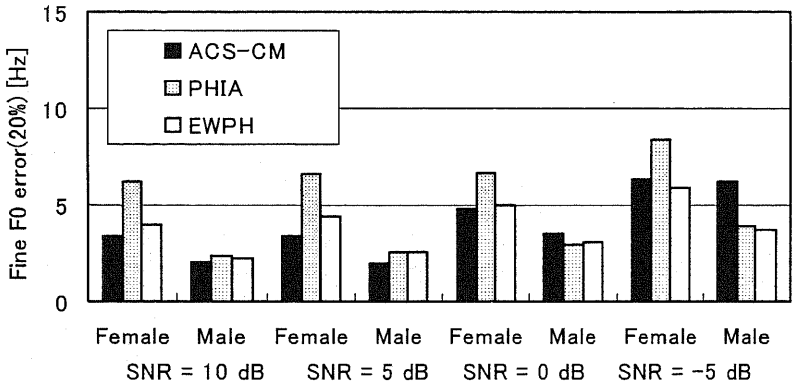
Fine F0 errors

(b) 展示会場（ブース内）

図 4.12(a) から白色雑音のスペクトルに比較的近い雑音の混入では，PHIA や EWPH と比較して ACS-CM の Gross F0 error(20%) が低いことを確認できる．Fine F0 error(20%) は，Fine F0 error と比べて F0 探索範囲が広くなったために全体で大きくなっている．また，SNR=−5 dB では ACS-CM よりも PHIA や EWPH が良い結果となっている．(b) では，SNR=−5 dB となるようにした Female で EWPH の Gross F0 error(20%) が ACS-CM よりも小さいことを確認できる．(a) と比べてスペクトルに傾きがあることから，ACS-CM は大きなパワーが偏在する帯域を持つ雑音が混入することで F0 推定率が下がると考えられる．(c) では，そ



Gross F0 errors



Fine F0 errors

(c) 走行自動車内

図 4.12: 文章/shizuoka.../の Gross F0 error(20%) と Fine F0 error(20%)

の問題がさらに大きく現れて SNR=−5 dB の ACS-CM で F0 をほとんど推定できていない.

4.7 まとめ

ACS-CM による雑音混入音声の F0 推定実験を示した. 実験条件について説明し, 予備実験で各処理の効果と ACS-CM の実験結果を示した. 合成した白色雑

音や有色雑音では，Gross F0 error と Fine F0 error から ACS-CM の有効性を示せた．また，比較実験での AUTOC と BPPAS と CEPST の結果からも有効性を確認できる．

実雑音混入音声では，スペクトルの傾きが小さな場合の Gross F0 error について ACS-CM の有効性を示せた．しかし，偏在した雑音において問題があることが分かった．そのため次章では，振幅調節と変調を施した振幅スペクトルを用いることで F0 推定法を改良する．

第5章　振幅調節と変調を施した振幅スペクトルを用いた基本周波数推定

5.1　はじめに

混入した雑音が特定の周波数帯域に偏在する場合は，ACS-CM の F0 推定に問題があった．そのため，雑音が偏在する場合を考慮した F0 推定の改良を行う．この章で振幅調節と変調を施した振幅スペクトルを用いた F0 推定を説明して，実験によって改良法の有効性を確認する．

5.2　帯域に偏在する雑音混入音声にも対応した基本周波数推定

改良法では ACS-CM にはなかった振幅調節の前処理を導入し，音声のある帯域に大きなパワーを持つ雑音が混入した場合にも対応した．ACM-CM においては 1 点のみの変調周波数点を用いて変調を行っているが，改良法は複数点の変調周波数点を用いて変調することを導入している．ACS-CM と改良法の主な処理の違いを表 5.1 に示す．

改良法の概略を図 5.1 に示す．振幅スペクトルの振幅の調節が必要であるか判断するために，振幅スペクトルの全帯域正規化振幅分散を用いる．振幅調節は 2 段階で行う．最初は線形予測分析を応用して，バンド幅拡大を施した線形予測係数で構成される逆フィルタを通すことで行う．雑音混入音声スペクトルの大まかな傾きを含め，ホルマントや偏在する大きなパワーを持つ雑音によるスペクトルの起伏を緩やかに（抑圧）できる．次に後述するように F0 探索範囲（50–500 Hz）を考慮した帯域幅 600 Hz 程度の振幅スペクトルの平均を使って各周波数成分の

表 5.1: ACS-CM と改良法の各処理の違い

	ACS-CM	改良法
帯域制限	あり	なし
振幅調節	なし	あり
スペクトルの変調	変調周波数点を 1 点	変調周波数点を複数点，振幅を反復変調
雑音推定	隣接調波間を利用した推定，BSS と雑音の程度	隣接調波間を利用した推定
雑音低減	ACS	SS
ACF の変調	変調周波数点を 1 点	変調周波数点を複数点，反復変調

振幅を調節する．その後，振幅スペクトルの反復変調を行うことで，低域を含め，調波構造の明瞭な帯域を増やす．変調後のスペクトルにおいて，隣接調波成分間は雑音であると仮定 [51, 52] して，雑音を推定し，スペクトルサブトラクション（SS : spectrum subtraction）する．この操作によってさらに調波構造は明瞭になる．そして，このスペクトルに対する自己相関関数（ACF）を求める．振幅スペクトルを変調した同じ周波数（点）で ACF を反復変調を行った後，この ACF から基本周波数を推定する．

5.2.1 全帯域正規化振幅分散

音声のある帯域に大きなパワーを持つ雑音が偏在する場合にはそのままの振幅スペクトルを用いると，雑音の影響が大きく，また変調の効果も小さい．雑音の影響を抑圧して，変調の効果がより大きく出るようにするために前処理として，振幅調節を行う．一方，白色雑音混入音声の場合の振幅調節は逆に雑音を強調することになり，効果はない．そのため振幅調節をするかどうかのパラメータが必要である．そのパラメータとして，スペクトルの傾きや起伏，スペクトル全帯域の調波構造の明瞭性が反映される振幅スペクトルの全帯域正規化振幅分散を用いる．

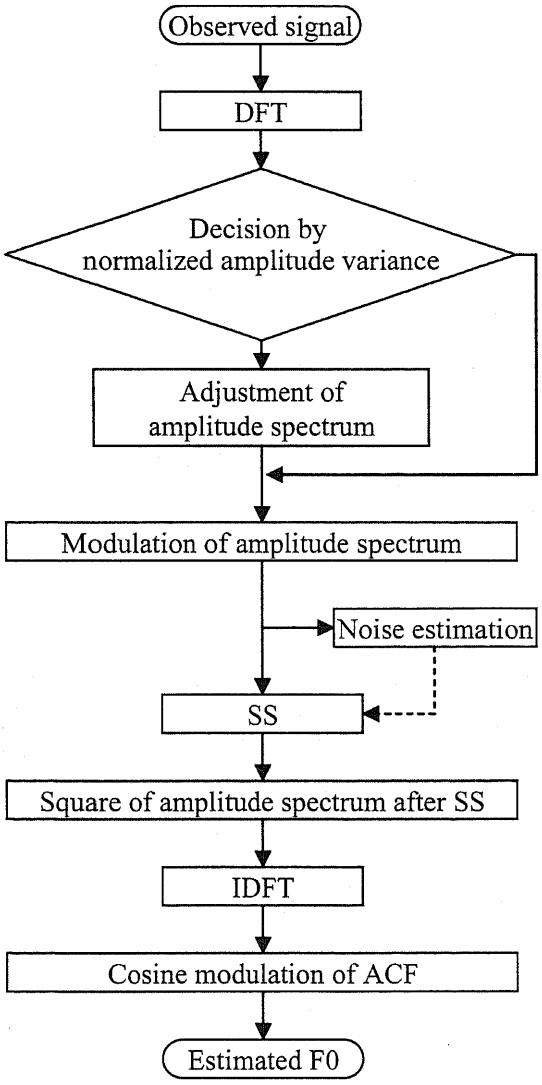


図 5.1: 改良した F0 推定法

全帯域正規化振幅分散 σ^2 は次式で表せる．

$$\sigma^2 = \frac{1}{\bar{S}^2}\left\{\frac{1}{N}\sum_{h=0}^{N-1}\left(S(f_h) - \bar{S}\right)^2\right\} \tag{5.1}$$

ここで，

$$\bar{S} = \frac{1}{N}\sum_{h=0}^{N-1} S(f_h) \tag{5.2}$$

$S(f_h)$ は観測信号の振幅スペクトルである．N は DFT のポイント数である．

表 5.2 に全帯域正規化振幅分散の例を示す．白色雑音が混入した場合と走行自

表 5.2: クリーン母音と雑音混入母音（SNR=0 dB）の全帯域正規化振幅分散 σ^2 の例

母音	クリーン音声	混入雑音		
		白色雑音	有色雑音	走行自動車内雑音
/a/	5.40	0.73	1.22	6.19
/i/	10.22	0.88	1.34	8.59
/u/	15.00	1.00	1.67	12.10
/e/	5.29	0.78	1.29	5.90
/o/	10.56	0.85	1.46	9.85

動車内雑音が混入した場合は明らかに全帯域正規化振幅分散が異なることが分かる．クリーン母音や走行自動車内雑音混入母音では，明瞭な調波成分となる帯域が含まれているためと考えられる．

5.2.2 振幅スペクトルの振幅調節

振幅調節は音声のある帯域に大きなパワーをもつ雑音が偏在し，雑音の影響で雑音混入音声の振幅スペクトルの最大値が元のクリーンな音声の振幅スペクトルの最大値程度以上に大きい場合や低域の雑音がホルマント程度以上の大きさのパワーを持つ場合等に必要である．変調は 5.2.3 節で述べるように振幅スペクトルの最大値の周波数を変調周波数として用いて，最大値付近の周波数帯を移動することになるので，このような場合振幅調節をしないで，変調を行うと，振幅の大きい雑音を多く含む帯域をそのまま移動することになり，その影響は大きい．振幅調節はこのような雑音を含む帯域の振幅を抑え，元々振幅は小さいが，調波構造の明瞭な帯域の振幅を相対的に大きくして，変調の効果がより出るように，振幅スペクトルの傾きを緩やかにすることや振幅スペクトルの起伏を緩やかにすることを意味する．振幅調節は 2 段階で行う．1 段階目は線形予測分析を応用して，バンド幅拡大操作を施した線形予測係数 [55] で構成される逆フィルタを通すことで行う振幅調節（振幅調節 1）[56] である．線形予測係数を α_i とすると，バンド

幅拡大操作を施した線形予測係数 β_i は次式で表せる．

$$\beta_i = \alpha_i e^{-\pi i BT},\ 1 \leq i \leq p \tag{5.3}$$

ここで，B はバンド幅拡大幅で 1000 Hz とし，p は線形予測分析の分析次数で 4 とした．T はサンプリング周期で 0.1 ms である．

1 段階目の処理は次式で表せる．

$$S_1(f_h) = |H(f_h)|S(f_h),\ 0 \leq h < N \tag{5.4}$$

ここで，$H(h)$ は逆フィルタの周波数応答で次式で表せる．

$$H(f_h) = 1 + \sum_{i=1}^{p} \beta_i e^{-j2\pi f_h iT} \tag{5.5}$$

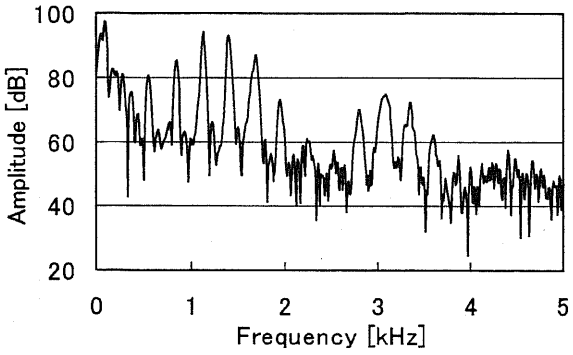
2 段階目の振幅調節（振幅調節 2）は次式で表せる．

$$S_2(f_h) = \frac{S_1(f_h)}{2k+1} \sum_{i=-k}^{k} \frac{1}{\bar{S}_1(f_{h+i})},\quad 0 \leq h < \frac{N}{2} \tag{5.6}$$

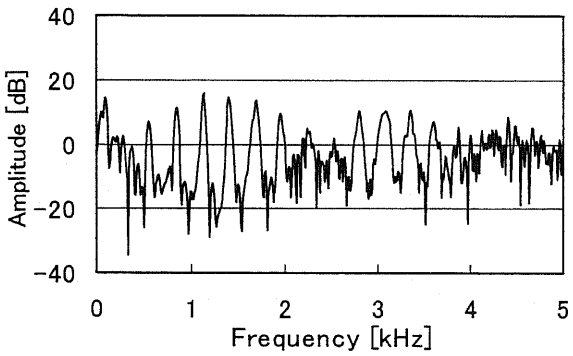
ここで，

$$\bar{S}_1(f_h) = \begin{cases} \dfrac{1}{2k+1}\displaystyle\sum_{i=0}^{2k} S_1(f_i), & -k \leq h < k \\ \dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k} S_1(f_{h+i}), & k \leq h < \frac{N}{2} - k \\ \dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k} S_1(f_{\frac{N}{2}-k-1+i}), & \frac{N}{2} - k \leq h < \frac{N}{2} + k \end{cases} \tag{5.7}$$

と表せ，$2k+1$ は基本周波数探索範囲 (50–500Hz) を考慮して選ばれる周波数軸のサンプル点数（ポイント数）である．10 kHz サンプリングにおいては，63 点（k =31）（男声，女声すべて同一）とした．式 (5.7) の平均においては，平均する両端の振幅値は，端が調波成分の場合は大きく，端が調波成分と調波成分の間であれば小さい，このような両端の影響を少なくするために式 (5.6) において $\sum$ を取っている．$\sum$ を取ると重み付きではあるが $\sum$ をとる前の約 2 倍の帯域の情報で



(a) 振幅調節前の振幅スペクトル

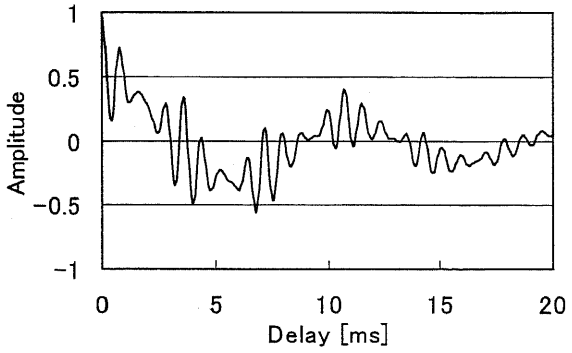


(b) 振幅調節後の振幅スペクトル

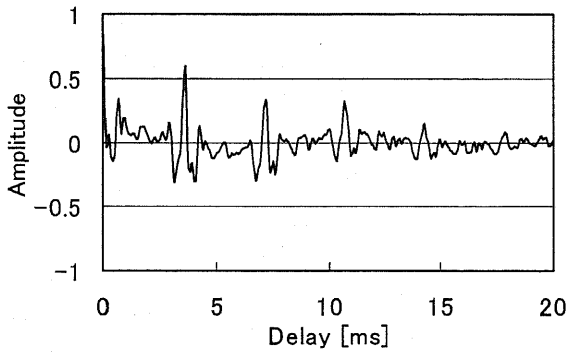
振幅調節をすることになるが，予備実験でその効果が若干あることを確認している．図5.2に振幅調節前と振幅調節後（振幅調節1と振幅調節2を行った後）の振幅スペクトルとそれらから求まるACFを示す．このフレームのF0は，約250 Hzである．(a)では，振幅の最大が雑音の影響によって現れた低域に確認できる．そのため，(c)のACFでは音声の基本周期に相当する遅延付近の振幅が抑圧されている．(b)は振幅調節によって低域の雑音が抑圧され，(d)のACFにおいて基本周期の遅れの点，4 ms付近で大きな振幅を持つ例を示している．

5.2.3 部分帯域正規化振幅分散と変調周波数点

変調は調波構造の明瞭な帯域を変調周波数（変調周波数点）を用いて低域に移動することを含め，調波構造の明瞭な帯域を増やすために用いる．そのため変調周波数点は調波成分にほぼ一致することが望ましい．必ずしもすべての変調周波



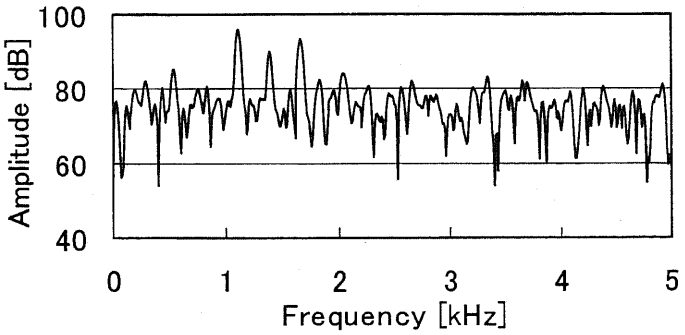
(c) (a) から求まる ACF



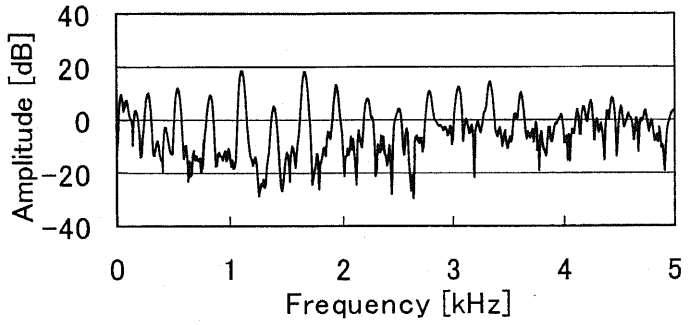
(d) (b) から求まる ACF

図 5.2: 走行自動車内雑音混入母音/a/（SNR −5 dB）の振幅を調節した振幅スペクトルとその ACF

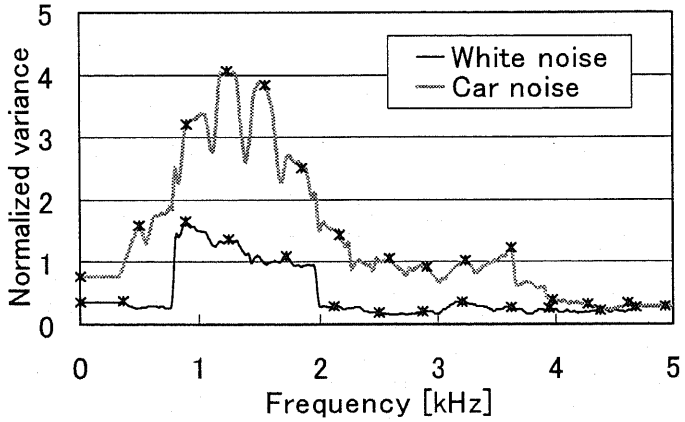
数点が調波成分にほぼ一致するとは限らないので，その一致性の程度を表すパラメータとして導入したものが部分帯域正規化振幅分散である．後で述べるように，変調周波数点は対象とする部分帯域において，振幅の最大である点としている．図 5.3 に部分帯域正規化振幅分散の例を示す．(a) と (b) を比べた場合，(a) では帯域の全部で調波成分が雑音の影響を受けている．そのため，全体的に部分帯域正規化振幅分散の値が小さくなっている．(b) では，雑音の影響をあまり受けていない帯域について部分帯域正規化振幅分散の値が大きくなっていることを確認できる．図 5.4 から/a/以外の単母音についても，同様の状態が確認できる．式 (5.6) の $2k+1$ に対応する帯域幅の部分帯域 h の正規化振幅分散 σ_h^2 は次式で表せる．



(a) 白色雑音混入音声スペクトル

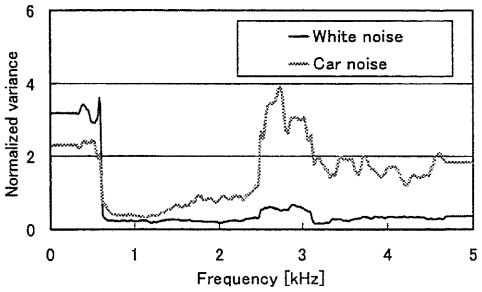


(b) 振幅調節後の走行自動車内雑音混入音声スペクトル

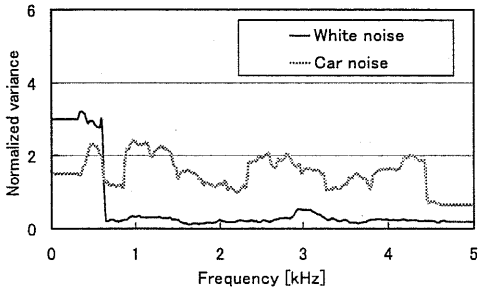


(c) (a) と (b) に対する部分帯域正規化振幅分散 σ_h^2 と選択された部分帯域の中心 (∗印)

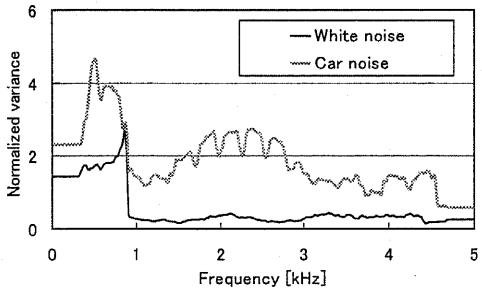
図 5.3: 雑音混入母音/a/（SNR 0 dB）の部分帯域正規化振幅分散



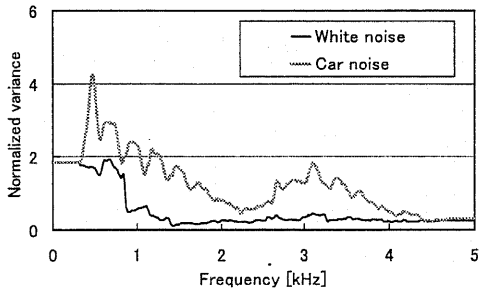
(a) /i/



(b) /u/



(c) /e/



(d) /o/

図 5.4: 雑音混入母音（SNR 0 dB）の部分帯域正規化振幅分散

$$\sigma_h^2 = \begin{cases} \dfrac{1}{\bar{S}_h^2}\left\{\dfrac{1}{2k+1}\displaystyle\sum_{i=0}^{2k}\left(S_a(f_i)-\bar{S}_h\right)^2\right\}, & 0 \le h \le k \\ \dfrac{1}{\bar{S}_h^2}\left\{\dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k}\left(S_a(f_{h+i})-\bar{S}_h\right)^2\right\}, & k < h < \frac{N}{2}-k \\ \dfrac{1}{\bar{S}_h^2}\left\{\dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k}\left(S_a(f_{\frac{N}{2}-k-1+i})-\bar{S}_h\right)^2\right\}, & \frac{N}{2}-k \le h < \frac{N}{2} \end{cases} \tag{5.8}$$

ここで，

$$\bar{S}_h = \begin{cases} \dfrac{1}{2k+1}\displaystyle\sum_{i=0}^{2k} S_a(f_i), & 0 \le h \le k \\ \dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k} S_a(f_{h+i}), & k < h < \frac{N}{2}-k \\ \dfrac{1}{2k+1}\displaystyle\sum_{i=-k}^{k} S_a(f_{\frac{N}{2}-k-1+i}), & \frac{N}{2}-k \le h < \frac{N}{2} \end{cases} \tag{5.9}$$

である．式 (5.8) の σ_h^2 から分かるように，図 5.3 の低域と高域の周波数帯域で振幅は一定である．式 (5.9) における $S_a(f_h)$ は，式 (5.1) の全帯域正規化振幅分散 σ^2 の大きさで判断する．振幅調節をしない場合は式 (5.1) における観測信号の振幅スペクトル $S(f_h)$ であり，振幅調節が必要な場合は振幅調節後の式 (5.6) の $S_2(f_h)$ である．式で表すと次のようになる．

$$S_a(f_h) = \begin{cases} S(f_h), & 0 \le h < \frac{N}{2}, \sigma^2 < G \\ S_2(f_h), & 0 \le h < \frac{N}{2}, \sigma^2 \ge G \end{cases} \tag{5.10}$$

式 (5.10) の G の値は後述する予備実験で決めた．G は 2.0 程度が適切であった．

変調周波数点を求めるにはまず部分帯域正規化振幅分散 σ_h^2 の大きさを基に全帯域を σ_h^2 が最大である帯域から順に，帯域の重複が 1/2 ポイント以上にならないように選択する．図 5.3 に選択された部分帯域の中心点を $*$ 印で示している．変調周波数点はそれらの部分帯域で振幅が最大である点である．

5.2.4 振幅スペクトルの反復変調

変調は5.2.3で述べたように調波構造の明瞭な帯域を低域に移動することを含め，調波構造の明瞭な帯域を増やすために用いる．そのため変調周波数点は調波成分にほぼ一致することが望ましい．必ずしもすべての変調周波数点が調波成分にほぼ一致するとは限らないので，複数点で変調を行い，その一致性の程度を表すσ_i用いて，この後で示す式(5.14)のように重み付き和をとっている．

変調周波数点iによるj回の反復変調で得られる振幅スペクトル$M_{ij}(h)$は次式で表せる．

$$M_{i0}(f_h) = S_a(f_h) \tag{5.11}$$

$$M_{ij}(f_h) = M_{ij-1}(f_h) + X_{ij}(f_h), \quad 1 \leq j \leq J \tag{5.12}$$

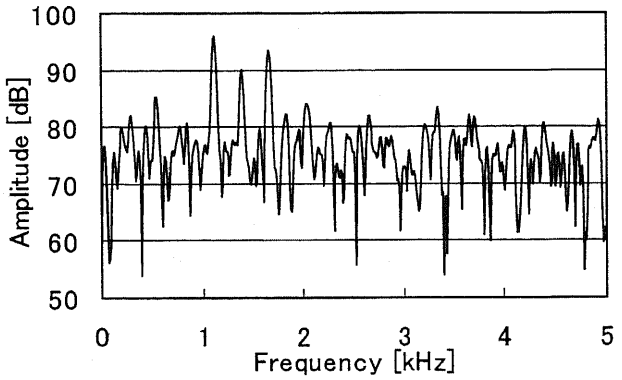
ここで，

$$X_{ij}(f_h) = \begin{cases} \dfrac{M_{ij-1}(f_{|h-i|}) + M_{ij-1}(f_{h+i})}{2}, & 0 \leq h < \frac{N}{2} - i \\ \dfrac{M_{ij-1}(f_{|h-i|})}{2}, & \frac{N}{2} - i \leq h < \frac{N}{2} \\ 0, & h = \frac{N}{2} \\ X_{ij}(f_{N-h}), & \frac{N}{2} < h < N \end{cases} \tag{5.13}$$

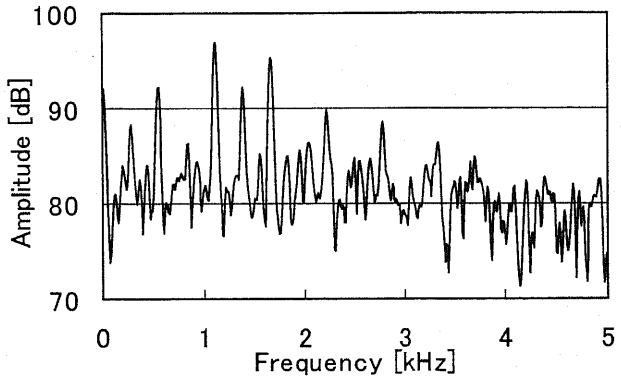
である．前節で得られるすべての変調周波数点iでJ回反復変調した振幅スペクトルのσ_i^2（中心i点の部分帯域正規化振幅分散）での重み付け和$Y(f_h)$をここでは変調スペクトルとして利用する．$Y(f_h)$は次式で表せる．

$$Y(f_h) = \frac{\sum_i \sigma_i^2 M_{iJ}(f_h)}{\sum_i \sigma_i^2} \tag{5.14}$$

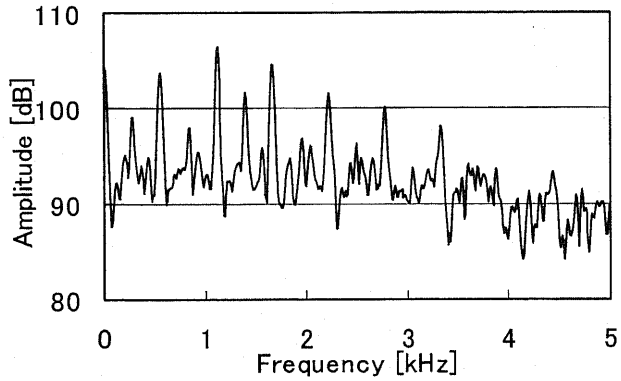
図5.5にJが1と3の場合の変調スペクトルを示す．このように変調すると調波構造が明瞭な帯域が増えることが分かる．(a)では，音声のスペクトルでピーク付近の振幅が雑音に埋もれている．(b)では，低域や高域の音声のスペクトルが存在すべき帯域付近に明瞭な振幅を確認できるようになり，(c)でより顕著になっている．



(a) 変調前の振幅スペクトル



(b)1 回反復変調した変調スペクトル



(c)3 回反復変調した変調スペクトル

図 5.5: 反復変調した変調スペクトル（白色雑音混入母音/a/（SNR 0 dB））

5.2.5 雑音の推定と雑音低減

スペクトルサブトラクション（SS）

観測信号から雑音成分を減算することで目的信号を強調する手法のSSについて説明する．

まず，雑音が定常であることを利用して，既知である音声休止区間の信号より雑音の特徴量を推定する．これを用いて，雑音を含む音声から雑音成分を取り除く処理をする．しかし，これには事前の情報が必要である．

ここで，時間tにおける観測信号を$x(t)$とすると，$x(t)$は目的信号$s(t)$と雑音$y(t)$の和で

$$x(t) = s(t) + y(t) \tag{5.15}$$

と表せる．これを，目的信号について示すと

$$s(t) = x(t) - y(t) \tag{5.16}$$

のようになる．また，フーリエ変換によって

$$S(e^{j\omega}) = X(e^{j\omega}) - Y(e^{j\omega}) \tag{5.17}$$

と示すことができる．ここで，$S(e^{j\omega})$，$X(e^{j\omega})$，$Y(e^{j\omega})$は，それぞれ$s(t)$，$x(t)$，$y(t)$のフーリエ変換である．

次に，事前情報がない場合を考える．そこで，推定した雑音を用いた処理を行う．周波数領域においてスペクトルの減算をすることで，雑音除去を行う手法について述べる．雑音の統計的性質を用いることで，目的以外の信号による影響を低減する．

これを実験に用いる場合の処理手順においては，以下のようになる．まず，推定雑音スペクトルを求める．そして，観測スペクトルから推定雑音スペクトルを減算する．そこから，雑音の影響が少ない目的スペクトルを推定できる．周波数点をf_hとすると，N点DFTで求めた観測信号スペクトル$S(f_h)$は，次式(5.18)のとおりである．

$$Y_{\mathrm{C}}(f_h) = S(f_h) - Y_{\mathrm{Z}}(f_h)\ , \quad 0 \leq h < N/2 \tag{5.18}$$

ここで，$Y_C(f_h)$ は減算後の推定目的スペクトル，$Y_Z(f_h)$ は推定雑音スペクトルである．

これは，推定した雑音スペクトルを用いて減算を行うことになる．しかし，正確な雑音スペクトルのみを減算することは，困難である．従って，推定した雑音スペクトルが実際の雑音スペクトルと異なる場合は，推定に影響を及ぼすことがある．

この SS を利用した雑音低減を改良法で用いる．雑音の推定を使った詳細を以下で述べる．

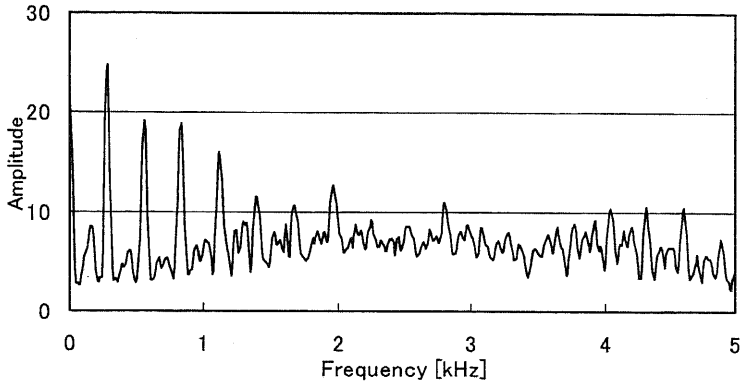
推定した雑音を用いたスペクトルサブトラクション（SS）

有声音の場合，そのスペクトルは調波構造を持つ．雑音混入音声の場合はその隣接間の成分のほとんどは雑音とみなすことができる [51, 52]．この仮定に基づいて雑音を以下の手順で推定する．(1) 式 (5.14) の変調スペクトルの極小点を求める．(2) 各隣接極小点間を線形補間する．(3) 周波数軸上で 3 点の移動部分平均をとって平滑化する．このように求めた雑音を $Z_A(f_h)$ とすると，この雑音を減算 (SS) して得られる雑音の低減されたスペクトル $S_e(f_h)$ は次式で表せる．

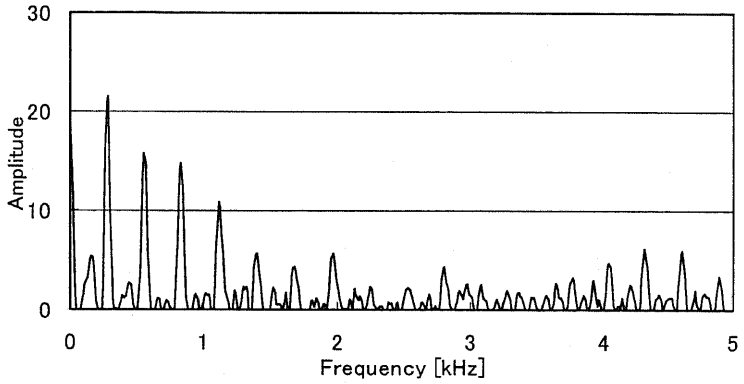
$$S_e(f_h) = \begin{cases} Y(f_h) - Z(f_h), & Y(f_h) \geq Z(f_h) \\ 0, & Y(f_h) < Z(f_h) \end{cases} \tag{5.19}$$

図 5.6 に SS 前と SS 後のスペクトルを示す．このフレームは，女性が発話したおおよそ 278 Hz の F0 を持つ文章で，その整数倍付近に調波成分を確認できる．(b) は，調波間の雑音成分と推定される振幅が，(a) と比べて減算されていることを確認できる．

図 5.7 に SS を行わない場合と SS を行った場合の反復変調後の ACF を示す．図 5.6 と同じ音声サンプルのフレームである．図 5.7(a) では，基準 F0 の 1/2 倍に相当する遅延部分に探索範囲内のピークが確認できる．そのため，このフレームは誤った F0 が推定される．(b) では，全体的に振幅が大きくなり，基準 F0 に相当する遅延 3.6 ms 付近の振幅は顕著になることで正しい推定がされる．そのため，SS の効果が確認できる．



(a)SS 前の振幅スペクトル



(b)SS 後の振幅スペクトル

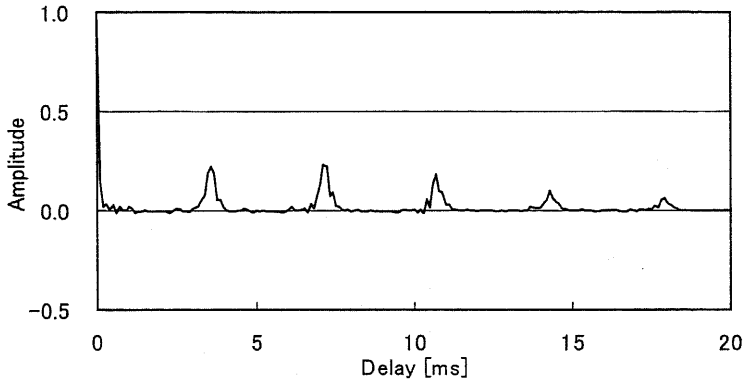
図 5.6: 工場（板金）雑音混入文章/shizuoka.../（SNR 0 dB）の振幅スペクトル

5.2.6 自己相関関数の反復変調

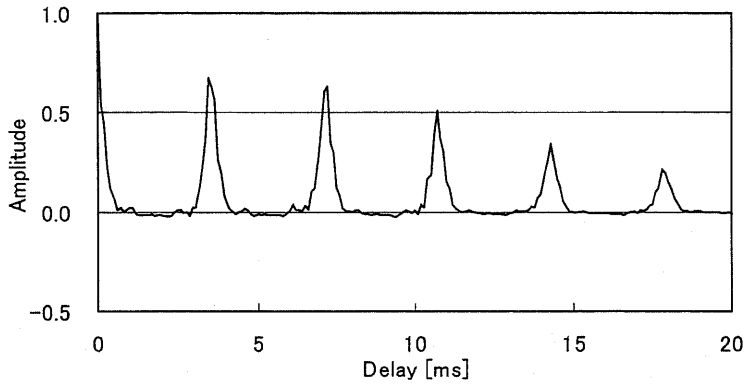
残留雑音の影響を抑圧するために，式(5.19)から求まるACFの$R(n)$にもL回の反復変調を施す．この変調は前で述べた振幅スペクトルの変調をACF上で実現するものである．変調周波数点は振幅スペクトルの変調で用いた変調周波数点と同じである．変調周波数点iによるl回の反復変調で得られるACFは次式で表せる．

$$R_{i0}(n) = R(n) \tag{5.20}$$

$$R_{il}(n) = R_{il-1}(n)\{1 + \cos(2\pi\frac{ni}{N})\}, \quad 1 \leq l \leq L \tag{5.21}$$



(a)SS を行わない場合の ACF の反復変調後の ACF



(b)SS を行った場合の ACF の反復変調後の ACF

図 5.7: 工場（板金）雑音混入文章/shizuoka.../（SNR 0 dB）の ACF

振幅スペクトルの変調と同様にすべての変調周波数点 i で L 回変調した ACF の σ_i^2 での重み付け和 $U(n)$ をここでは変調 ACF として利用する．$U(n)$ は次式で表せる．

$$U(n) = \frac{\sum_i \sigma_i^2 R_{iL}(n)}{\sum_i \sigma_i^2} \tag{5.22}$$

F0 推定には式 (5.22) を遅れ n 軸上で 3 点の移動平均を取ることによる平滑化（式 (3.59)）した ACF を用いる．ACS-CM をここまでの処理で改良して，そこから得た ACF の探索範囲内の最大振幅となる周波数点でサンプリング周波数を割り，F0 を推定する．この F0 推定が改良法である．

5.3 予備実験

改良法の各パラメータは予備実験によって決定し，各処理の有効性を検討した．ここで，音声データは4.3節の各母音にコンピュータで生成した白色雑音と走行自動車内雑音を付加したデータを用いた．音声の振幅はフレームごとに異なる．そのため，図4.7の(d)からも確認できるように，フレームによって実際のSNRは異なる．

前処理として振幅スペクトルの振幅調節が必要であるか判断することを全帯域正規化振幅分散，式(5.1)のσ^2を使って行うが，その閾値，式(5.10)のGの適切な値を求めるために予備実験を行った．表5.3にその結果を示す．Gは2.0程度が適切であることが分かる．また，振幅調節1をしないで振幅調節2を行う場合と振幅調節1と振幅調節2を行う場合を比較すると振幅調節1も行う方がよいことが分かる．

表5.4に調波成分の割合を示す．改良法では振幅のピークを各部分帯域から求めている．そこで，(b)の調波成分では，部分帯域正規化振幅分散が最大となった部分帯域からの振幅のピークで求めた割合である．実際の改良法では，部分帯域正規化振幅分散が最大以外の帯域の情報も利用している．雑音の影響によって，ACS-CMでは調波成分ではない振幅を求めていることが確認できる．Improvedでは，ACS-CMと比較して調波成分を求められる．そのため，変調の精度が上がることでF0の推定率の向上を考えられる．

表5.5に振幅スペクトルの反復変調回数とACFの反復変調回数による結果を示す．これから振幅スペクトルについては3回程度の反復変調が，ACFについては2回程度の反復変調が効果があることが分かる．これらの実験結果から，以下の実験では$G \geq 2.0$のときは振幅調節1と振幅調節2の振幅調節を行い，振幅スペクトルの反復変調回数は3，ACFの反復変調回数は2に設定し，合成雑音混入音声の場合の実験と実雑音混入音声の場合の実験を行った．

振幅調節と変調の効果を調べるために行った実験結果，振幅調節「あり」,「なし」と変調「あり」,「なし」によるGross F0 errorを表5.6に示す．ここで，振幅調節「あり」の場合の「すべて」は，全帯域のスペクトルの正規化振幅分散に関係なく，すべてのフレームで,「$G \geq 2.0$」は$G \geq 2.0$のとき，振幅調節の振幅

表 5.3: 式 (5.10) の G の値による Gross F0 error [%]

(a) 振幅調節 1 をしないで振幅調節 2 を行う場合

G	混入雑音					
	白色雑音			走行自動車内雑音		
	SNR= −5 dB	SNR= 0 dB	SNR= 5 dB	SNR= −5 dB	SNR= 0 dB	SNR= 5 dB
0.0	28.98	14.75	5.42	30.00	20.00	14.92
1.0	27.12	13.56	5.59	30.00	20.00	14.41
2.0	27.12	13.56	5.42	30.00	20.34	14.41
3.0	27.12	13.56	5.42	30.17	20.51	14.75
4.0	27.12	13.56	5.42	31.02	22.54	16.61

(b) 振幅調節 1 と振幅調節 2 を行う場合

G	混入雑音					
	白色雑音			走行自動車内雑音		
	SNR= −5 dB	SNR= 0 dB	SNR= 5 dB	SNR= −5 dB	SNR= 0 dB	SNR= 5 dB
0.0	28.14	13.90	5.42	22.03	17.80	14.41
1.0	27.12	13.90	5.42	22.03	17.80	14.41
2.0	27.12	13.56	5.42	22.03	18.14	14.41
3.0	27.12	13.56	5.42	22.20	18.47	14.92
4.0	27.12	13.56	5.42	23.22	20.68	16.61
∞	27.12	13.56	5.59	64.41	44.24	21.86

調節 1 と振幅調節 2 を行っている．変調「あり」の場合は，振幅スペクトルの反復変調（$L = 2$）と ACF の反復変調（$J = 3$）を行っている．走行自動車内雑音混入音声の場合，振幅調節を行うだけでも，Gross F0 error は少なくなり，さらに変調の効果も大きいことが分かる．白色雑音混入音声の場合は振幅調節を行うと Gross F0 error が多くなることを示している．「$G \geq 2.0$」では，$G \geq 2.0$ 以上で振幅調節を行うので，フレームによって振幅調節される場合とされない場合が

表 5.4: 走行自動車内雑音混入音声における調波成分の割合 [%]

(a) ACS-CM

	Harmonics						
dB	1st	2nd	3rd	4th	5th	Others	Rest
0	16.81	18.45	3.06	9.35	0.57	1.24	50.53
−5	2.89	1.11	0.12	0.29	0.01	0.03	95.56

(b) Improved

	Harmonics						
dB	1st	2nd	3rd	4th	5th	Others	Rest
0	0.63	19.19	5.06	24.61	4.90	23.81	21.79
−5	0.24	14.40	3.36	24.58	3.70	25.70	28.04

表 5.5: 振幅スペクトルの反復変調回数（J）と ACF の反復変調回数（L）による Gross F0 error [%]（混入雑音は白色雑音，SNR 0 dB）

反復回数 J または L	$L=0$ の場合	$J=0$ の場合	$L=2$ の場合	$J=3$ の場合
0	14.41	14.41	13.56	12.88
1	13.90	13.73	13.56	12.71
2	13.22	13.56	13.22	12.37
3	12.88	13.90	12.37	12.88
4	13.05	14.92	13.07	13.22
5	13.05	14.92	13.90	12.88

ある．表 5.7 から分かるように走行自動車内雑音混入音声の場合は，全帯域正規化振幅分散の平均が 7.17 で，ほとんどのフレームにおいて振幅調節が行われている．白色雑音混入音声の場合には，全帯域正規化振幅分散の平均が 0.81 ですべてのフレームにおいて振幅調節が行われていない．

表 5.6: 振幅調節「あり」,「なし」と変調「あり」,「なし」による Gross F0 error [%]（白色雑音混入，走行自動車内雑音混入いずれの場合も SNR 0 dB）

混入雑音	振幅調節				
	あり			なし	
	すべて		$G \geq 2.0$		
	変調		変調	変調	
	あり	なし	あり	あり	なし
白色雑音	13.90	15.25	13.56	13.56	14.41
走行自動車内雑音	17.80	35.42	18.14	44.24	57.46

表 5.7: 全帯域正規化振幅分散の平均と振幅調節されるフレーム数の割合（白色雑音混入，走行自動車内雑音混入いずれの場合も SNR 0 dB）

	混入雑音	
	白色雑音	走行自動車内雑音
全帯域正規化振幅分散の平均	0.81	7.17
振幅調節されるフレーム数の割合 [%]	0.00	98.64

5.4 実験結果

雑音が偏在する場合に，ACS-CM と比べて改良法が有効な F0 推定法であるか検討する．そこで，ACS-CM と同様な実験を行い，既存の F0 推定法についても比較を行う．

実験はコンピュータを用いて実行している．ここで，改良法と ACS-CM は複数の処理を組み込んでいるため BPPAS，AUTOC，CEPST と比べて計算量を必要とする．また，EWPH では処理の前半で多くのチャネルを用いた手法であるため，さらに計算量が必要である．そこで，白色雑音混入音声と走行自動車内雑

表 5.8: F0 推定の実行時間

F0 推定法	AUTOC	Improved	ACS-CM	BPPAS	EWPH′
実行時間比	1.0（基準）	2.1	2.2	1.1	58.8

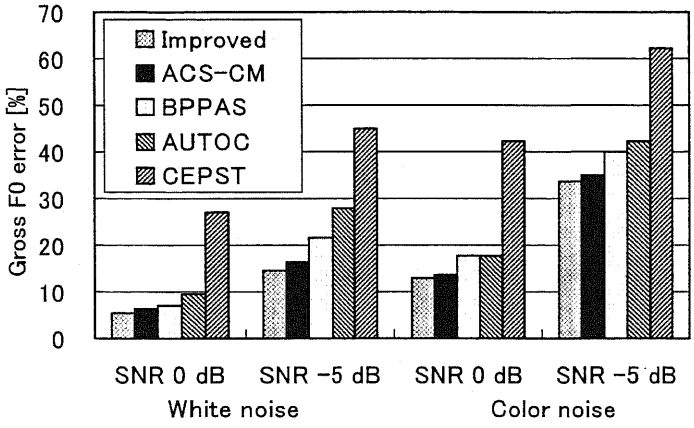
音混入音声の同程度のフレーム数を用いて，実行時間を求めることで計算量を検討した．AUTOC を基準とした場合，各 F0 推定法の実行時間比を表 5.8 に示す．ここで，EWPH′ の実行時間比は EWPH から最終推定部の STRAIGHT-TEMPO を除いた値である．

5.4.1 合成雑音混入音声の場合の実験結果

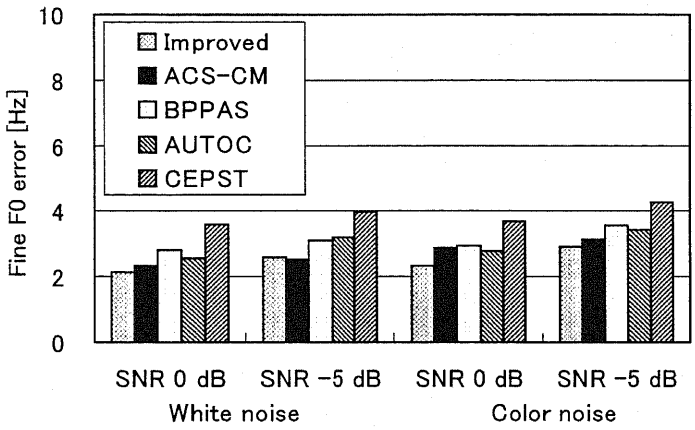
実験の設定や音声サンプルと合成雑音は，4.5 節の実験と同じである．改良法は，ACS-CM と同様に推定した各フレームごとの F0 を用いて評価した．ここで，BPPAS で用いる帯域制限は今回の実験で設定した F0 探索範囲を考慮して 50–500 Hz とする．

図 5.8 に改良法の F0 推定結果を示す．比較のため，ACS-CM，BPPAS，AUTOC，CEPST についても同様に示す．白色雑音または有色雑音を付加した場合ともに改良法の Gross F0 errors が少ないことが分かる．Fine F0 errors では，白色雑音を SNR −5 dB となるように付加した場合の結果を除き改良法での誤りが低減していることを確認できる．このため，改良法は合成雑音混入音声に比較的有効な F0 推定法である．

図 5.9 に AUTOC で推定されたクリーン文章/soozoo.../の基本周期を示す．図 5.10 に男性が発話した文章/soozoo.../の各時間のフレームで求められた ACF を示す．この音声サンプルは，図 5.9 の基本周期を持つ音声に白色雑音を SNR が −5 dB となるように付加した．図で色の濃さは ACF の振幅を表し，色が薄いほど −1.0 に近く，濃いほど 1.0 に近い振幅である．遅延時間は，F0 推定の探索範囲内に相当する範囲で表示している．そこで，各時間のフレームで最も濃い遅延部分が最大振幅として，推定された基本周期になる．(a) では振幅が負となる遅延点が確認できるが，(b) では減少する．しかし，振幅の正となる部分は，AUTOC より明瞭に表示されている．これは，探索範囲内の最大振幅から F0 を推定して



(a) Gross F0 error



(b) Fine F0 error

図 5.8: 合成雑音混入音声の F0 推定結果

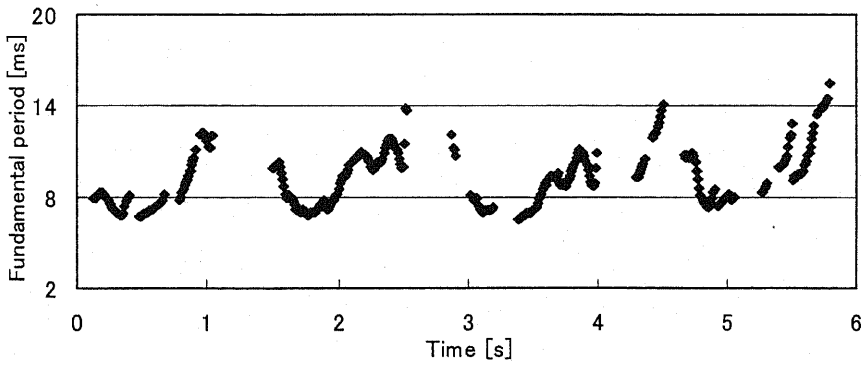
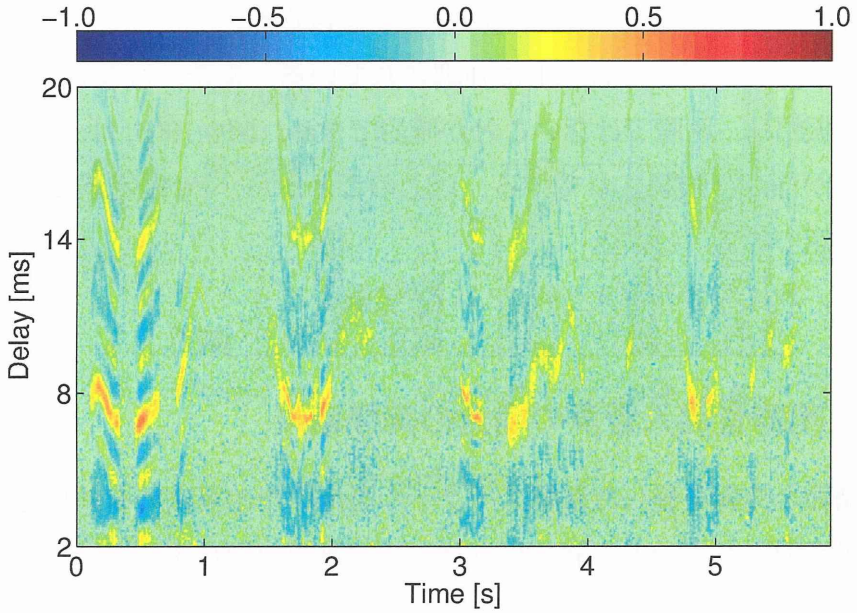
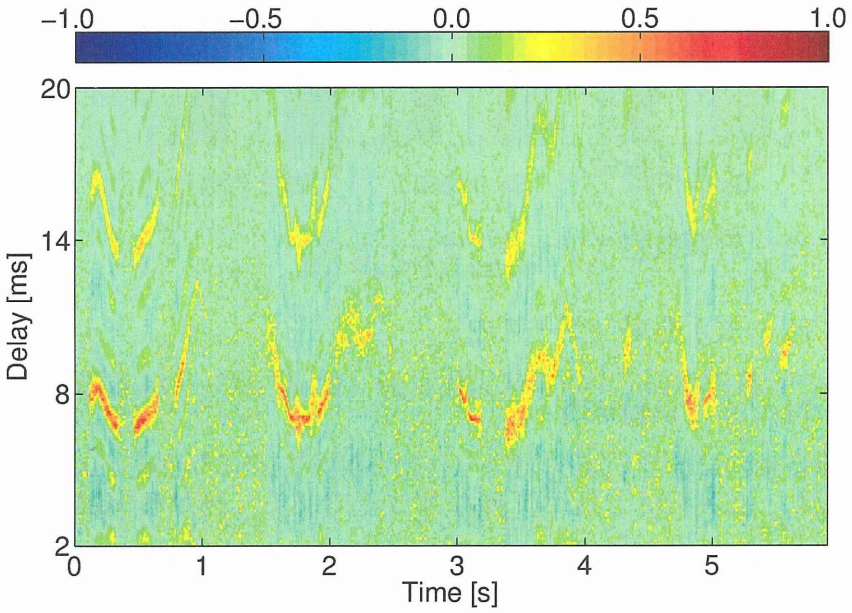


図 5.9: クリーン文章/soozoo.../の基本周期



(a) AUTOC



(b) Improved

図 5.10: 雑音混入文章/soozoo.../の ACF

いるため，推定に必要となる基本周期に相当するの部分が強調されていると判断できる．また，基本周期の整数倍に相当する付近も，強調されている．しかし，基本周期部分は同様かそれ以上に強調されている．また，図5.10(a)の単語の語尾付近の振幅は，明瞭ではないことが確認できる．図5.10(b)は，(a)と比べて語尾の部分が僅かに強調されている．以上のことから，F0推定誤りを低減できたことが確認できる．

そこで，次に実雑音混入音声について実験する．

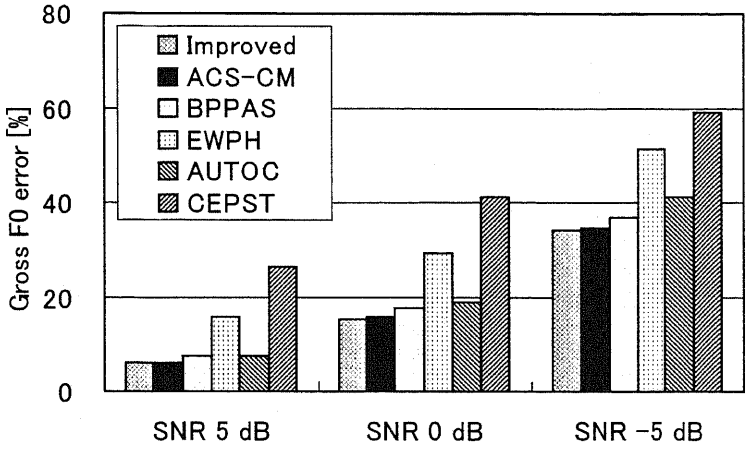
5.4.2 実雑音混入音声の場合の実験結果

振幅調節と変調を施した振幅スペクトルを用いた改良法の有効性を検討するために実雑音混入音声での実験を行った．特に雑音が偏在する場合の雑音混入音声のF0推定について検討した．

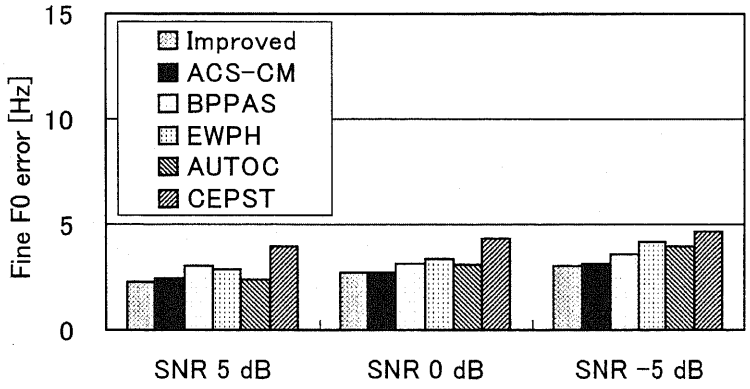
F0推定結果のGross F0 errorとFine F0 errorを図5.11–5.14で以下の雑音が混入した音声についてそれぞれ示す．比較のため，ACS-CM，BPPAS，EWPH，AUTOC，CEPSTについても同様に示す．実験の設定と音声サンプルは，4.6節と同じである．

- 図5.11 ：工場（板金）
- 図5.12 ：計算機室（ワークステーション）
- 図5.13 ：展示会場（ブース内）
- 図5.14 ：走行自動車内

図5.11では，ACS-CMと比べて僅かな改善である．図から実雑音混入音声についても改良法が有効であると確認できる．特にACS-CMで問題になった，走行自動車内雑音混入音声で誤りの低減が確認できる．図5.14のSNR 0，−5 dBでは，ACS-CMと比べて改良法のGross F0 errorが大きく改善できている．SNR 0 dBのときは30 %程度の改善である．このため，特定の帯域に偏在する雑音などにも対応可能なF0推定法であることが確認できる．また，改良法はFine F0 errorについてもACS-CMと比べて僅かな改善を確認できる．



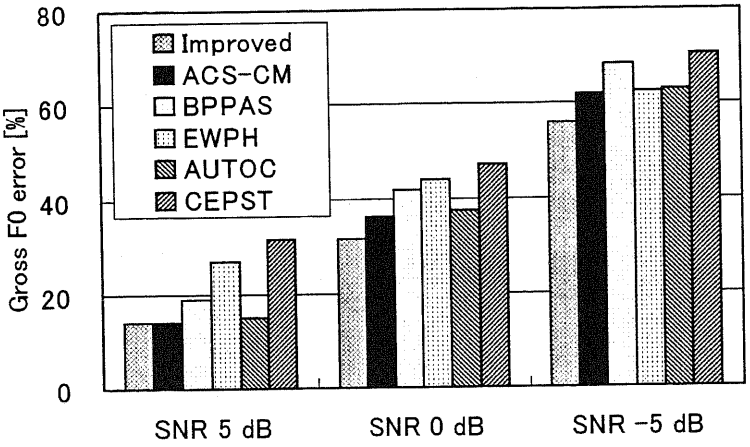
(a) Gross F0 error



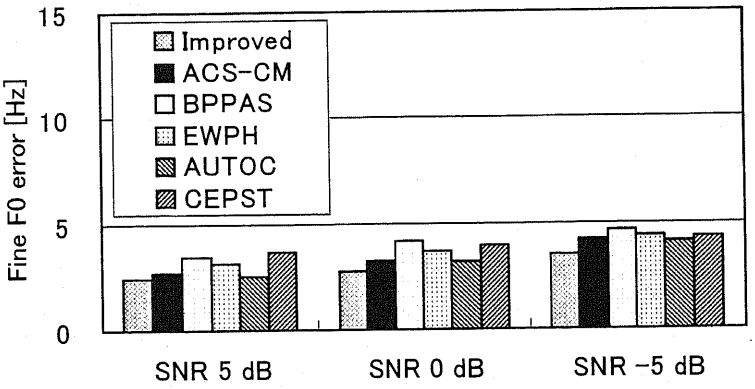
(b) Fine F0 error

図 5.11: 工場（板金）雑音混入音声の F0 推定結果

改良法による F0 推定の実験で得られた F0 の例を走行自動車内雑音混入音声について図 5.15 で示す．比較のためクリーンな音声について AUTOC を用いて推定した F0 を図 5.15(a) に示す．(b) では，走行自動車内雑音の影響によって高域や低域部分に多くの F0 誤りが確認できる．(c) では，さらに多くの F0 誤りが確認できる．これは，走行自動車内雑音混入音声では雑音が偏在し大きなパワーを持つため，ACS-CM のスペクトルの最大振幅となる周波数点を用いた処理で，雑音による誤った情報を用いるためである．そして，(d) では，音声休止区間付近



(a) Gross F0 error



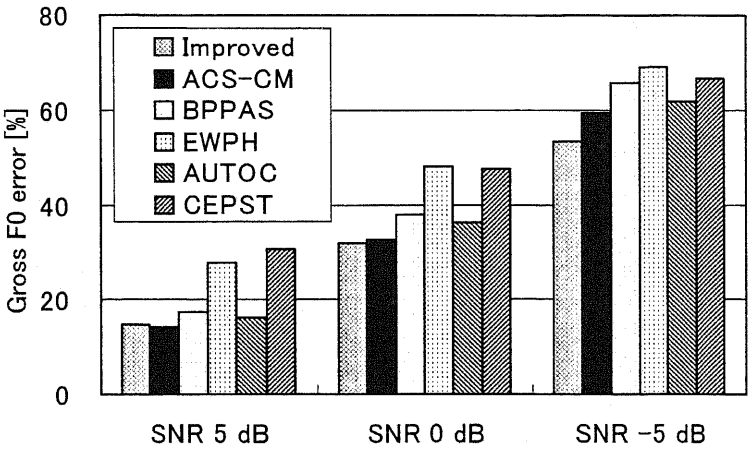
(b) Fine F0 error

図 5.12: 計算機室（ワークステーション）雑音混入音声の F0 推定結果

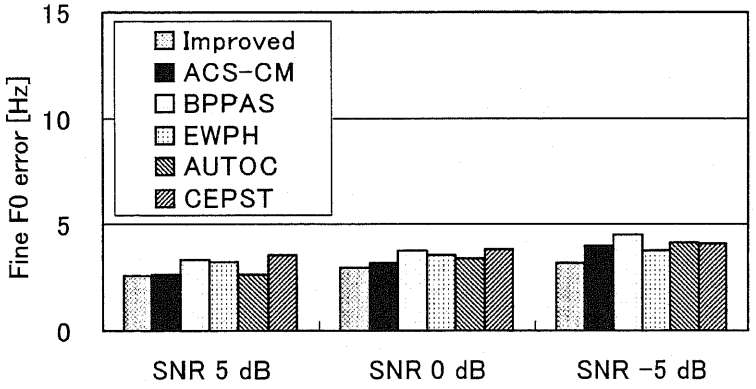
の F0 以外の多くが (a) に近づき推定の改善を確認できる．

図 5.16 に AUTOC で推定されたクリーン文章/shizuoka.../の基本周期を示す．図 5.17 に男性が発話した文章/shizuoka.../の各時間のフレームで求められた ACF を示す．この音声サンプルは，図 5.16 の基本周期を持つ音声に走行自動車内雑音を SNR が −5 dB となるように付加した．

図 5.17(b) の音声休止区間で振幅の大きな箇所が多く確認できる．そのため，走行自動車内雑音の相関が大きなことが分かる．(a) と比べて (b) は全体的に振幅



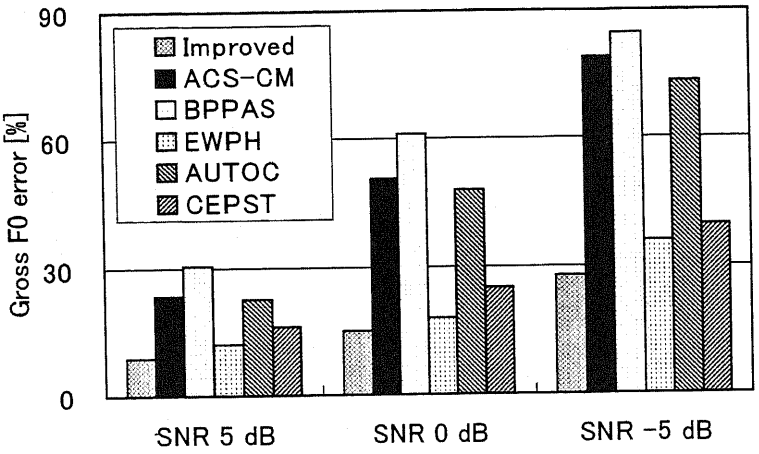
(a) Gross F0 error



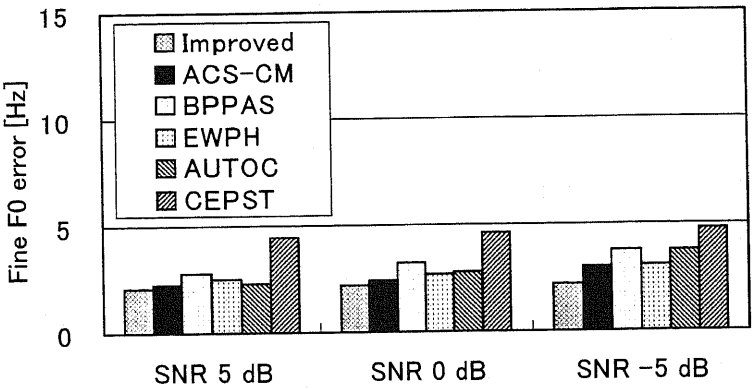
(b) Fine F0 error

図 5.13: 展示会場（ブース内）雑音混入音声の F0 推定結果

が小さくなっているが，精度の良い推定に必要な図 5.16 の基本周期付近はあまり抑圧されていない．そのため，(a) の基本周期の推定は (b) と比べて誤りを低減する結果になったと考えられる．また，図 5.17 のそれぞれの処理後から得られるスペクトログラムを図 5.18 に示す．(a) では，走行自動車内雑音の影響によって低域にパワーが集中し，僅かな周波数帯域で音声のパワーが確認できる．(b) では，(a) と比較して雑音の影響が減少し，高域の部分まで音声のパワーがシフトされていることを確認できる．



(a) Gross F0 error



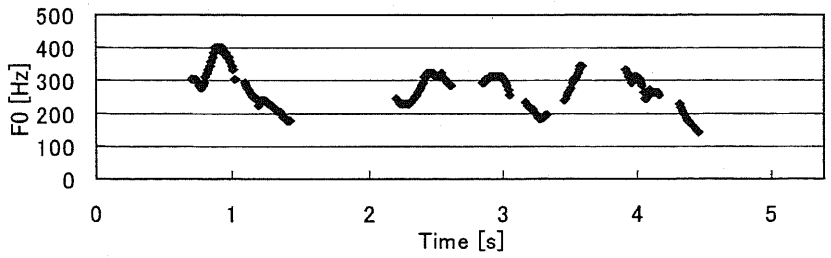
(b) Fine F0 error

図 5.14: 走行自動車内雑音混入音声の F0 推定結果

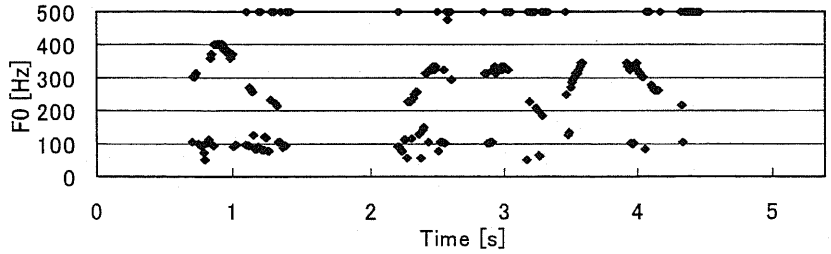
5.4.3 考察

ACS-CM では，走行自動車内雑音のような特徴を持つ雑音混入に対応できなかった．しかし，改良法の実験結果では走行自動車内雑音混入音声についても F0 推定の誤り率を低減できた．そのため，振幅調節と変調の処理は有効であると考えられる．

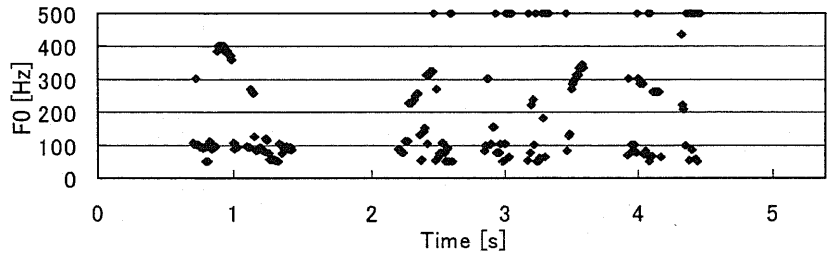
改良法は，白色雑音，有色雑音，工場（板金）雑音，計算機室（ワークステーション）雑音，展示会場（ブース内），走行自動車内雑音に対応でき，多くの種



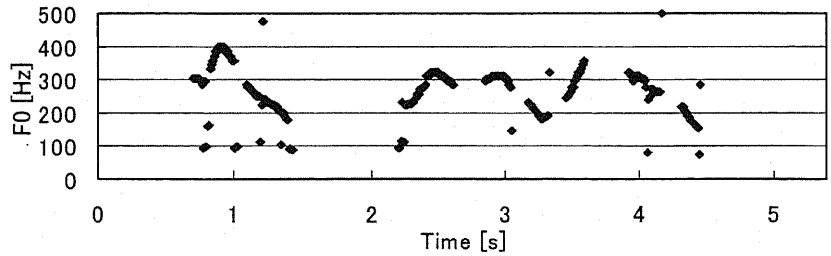
(a) AUTOC (クリーン音声)



(b) AUTOC (走行自動車内雑音混入音声，SNR −5 dB)



(c) ACS-CM (走行自動車内雑音混入音声，SNR −5 dB)



(d) Improved (走行自動車内雑音混入音声，SNR −5 dB)

図 5.15: 文章/shizuoka.../の推定した F0

類の実雑音に有効であると考えられる．そこで，実環境の様々な場所で混入する雑音でも精度の良いF0推定が期待できる．

5.5 まとめ

混入雑音が白色雑音のような広帯域雑音か，走行自動車内雑音のようなある帯域に雑音が偏在し，大きなパワーを持つかを，振幅スペクトルの全帯域正規化振幅分散で判断し，全帯域正規化振幅分散が大きい場合は振幅調節を組み入れた雑音混入音声のF0推定法を提案した．自己相関を用いたF0推定法のGross F0 errorと比べて走行自動車内雑音混入音声でSNRが0 dBの場合，改良法は30 %の低減を実現した．低SNRの走行自動車内雑音混入音声の場合，従来の変調を組み入れた方法(ACS-CM)に比べて，大きな改善がみられた．また，走行自動車内雑音に強いといわれているEWPHに比べてもよい結果を得た．

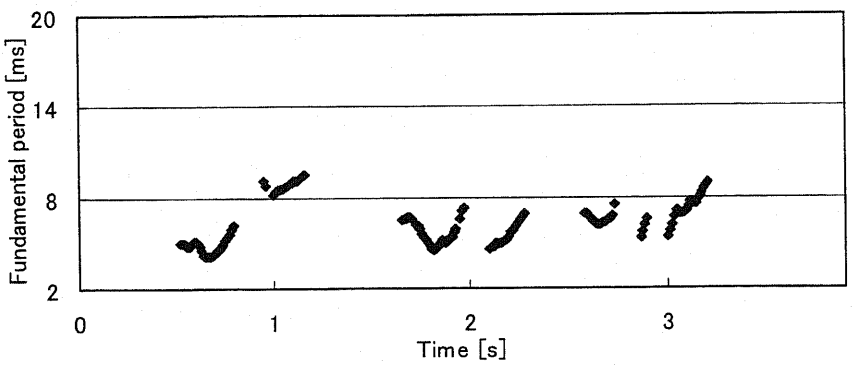
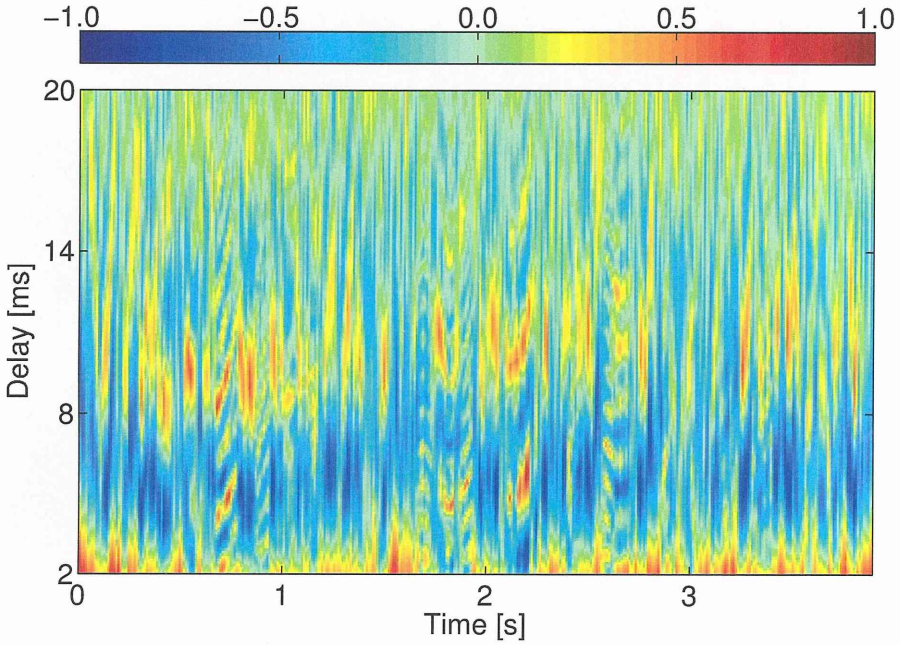
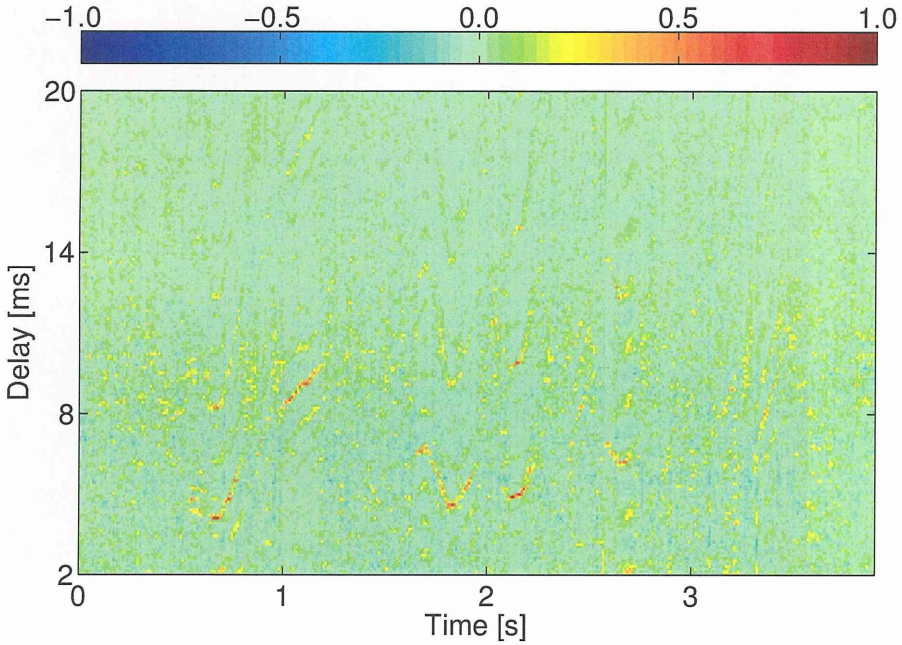


図 5.16: AUTOCで推定されたクリーン文章/shizuoka.../の基本周期

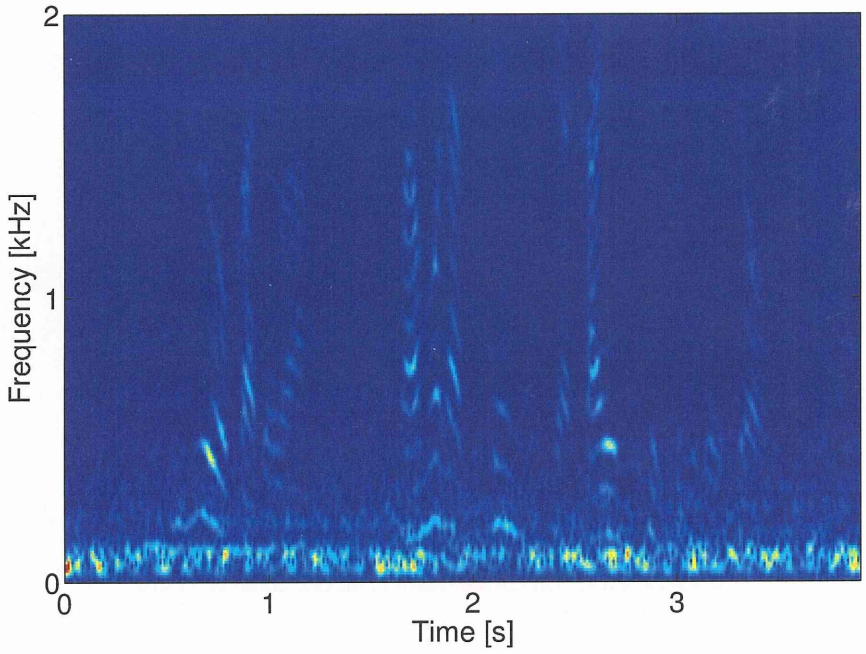


(a) AUTOC

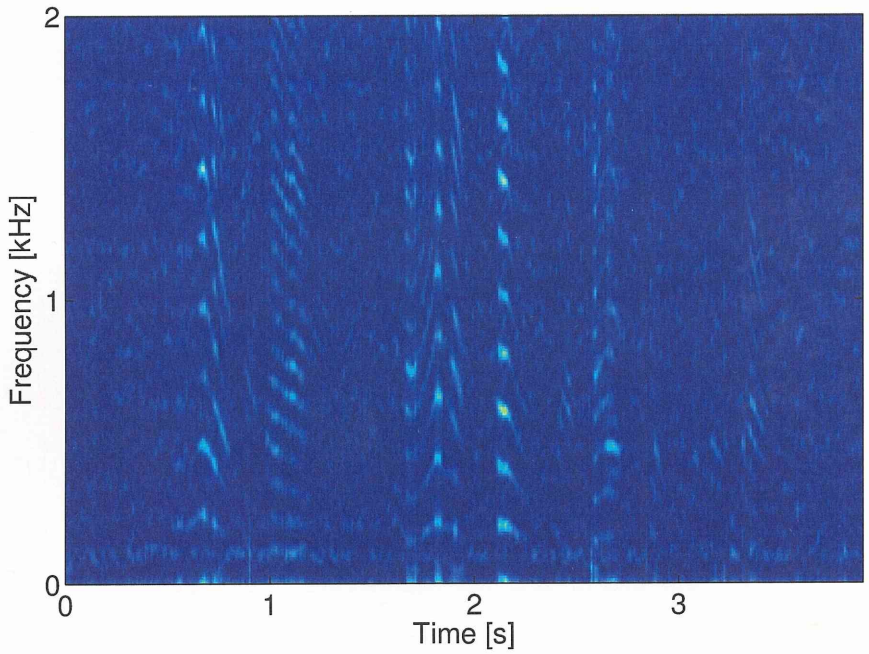


(b) Improved

図 5.17: 雑音混入文章/shizuoka.../の処理後の ACF



(a) AUTOC



(b) Improved

図 5.18: 雑音混入文章/shizuoka.../の処理後のスペクトログラム

第6章　結論

6.1　本論文の要約

雑音混入音声からF0推定するために，ACS-CMによる推定法を提案した．シングルチャネル入力の雑音混入音声を観測信号として，雑音の情報を与えていない状態で観測信号から処理フレーム内でF0を推定する手法である．比較的スペクトルが特定の帯域に偏在しない雑音の混入で，ACS-CMの有効性を示した．白色雑音に比較的近い雑音混入音声では，ACS-CMでF0の高い推定率を得ることができた．しかし，スペクトルの特長によっては効果を得ることができない結果となった．そこで，混入雑音が白色雑音のような帯域全体に均一なパワーを持つ雑音か，走行自動車内雑音のようなある帯域に雑音が偏在し，大きなパワーを持つかを振幅スペクトルの全帯域正規化振幅分散で判断し，全帯域正規化振幅分散が大きい場合は振幅調節を組み入れた雑音混入音声のF0推定法を提案した．調波構造の強調や雑音成分の低減を目指し，

- 振幅スペクトルの振幅調節
- 振幅スペクトルの反復変調
- SSによる雑音低減
- ACFの反復変調

を組み合わせたF0推定法を構築した．低SNRの走行自動車内雑音混入音声の場合，ACS-CMに比べて，大きな改善がみられた．また，白色雑音に強いといわれているBPPASと比べてもよい結果を得た．そして，走行自動車内雑音に強いといわれているEWPHと比べてもよい結果を得た．改良法では，ACS-CMで問題となった特徴を持つ雑音においても実験結果によって有効性を示すことができた．

AUTOCのGross F0 errorと比べて走行自動車内雑音混入音声のSNRが0 dBの場合，改良法は30 %程度の誤り低減を実現した．本論文の雑音混入音声のF0推定法において，以下の優れた点が挙げられる．

- シングルチャネル入力の信号から雑音混入音声のF0を推定
- 雑音の情報を事前に与えない観測信号からフレーム内で処理
- 白色雑音混入や走行自動車内雑音混入の場合ともに少ないF0推定誤り
- 比較的雑音の影響が大きな場合にも頑健
- 時間によってスペクトルが変化する実雑音混入にも有効

このように，雑音の特徴に合わせたパラメータなどの変更がない場合においても，F0推定の誤りを低減できる．実環境の雑音混入音声でF0推定の誤りを低減できる本研究の成果は，音声情報処理の幅広い分野において役立てることができる．

6.2 今後の課題

実際のシステムで利用するには，いくつかの問題が考えられる．

スペクトルの正規化振幅分散で，より精度良く雑音の影響が少ない帯域を求める必要がある．そして，音声のスペクトルに似た特徴を持つ雑音やスペクトルで大きな正規化振幅分散を持つ雑音については，対応が必要である．また，音声信号と雑音が無相関である仮定によって手法を決めてきたが，実際には相関を持つ場合も考えられるため対応が必要である．目的音声を発話した人だけを考えた場合，他の音声は雑音であると考えることができる．会話では複数の話者が存在するため，異なる話者との発話が重なる部分の観測信号について処理が必要である．この場合，分離などによって目的音声の成分を推定するなどの問題があり，F0推定はさらに困難となる．2つの信号のF0推定法はいくつか提案されているので，対応が必要である．

提案法を評価するため，本研究ではシミュレーションによって実験を行った．あらゆる方向から到来する雑音や残響が存在するため，実際の雑音環境を再現す

ることは，一般に困難である．そこで，評価用音声データの生成には次のような方法が挙げられる．

- コンピュータを用いて，プログラムによりクリーン音声へ実雑音を重畳することで，擬似的に加法性雑音を生成する．
- 複数個のスピーカから雑音を再生することで雑音空間を作り，その中で被験者に発声させることで，擬似的な実雑音環境を生成する．

本研究ではクリーン音声に実雑音を重畳した信号を，実環境で収録された雑音混入音声であると仮定した．そのため，実際の雑音環境下での収録音声を用いた場合と違い評価に限界がある．また，実際の雑音環境下での収録と異なり，実験のために予めSNRを設定した評価用音声データの生成が可能である．実際の雑音環境下での収録では，雑音の影響によって発話が通常と異なり，音声の特徴変化が起こる．パワーの増大，ホルマントの変化，F0の上昇については探索範囲内であればF0推定が可能であると考えられる．そこで，ロンバード効果の問題[57, 58, 59]についても対応が期待できる．さらに，実環境では雑音のほかに残響の問題がある．有声音の影響による残響で周期性を持ったような信号が表れる．そのため，音声休止区間や無声音のフレームが残響の影響を受けることで，そのフレームにF0のような周期性が表れ，音声区間と音声休止区間の推定や有声／無声区間推定での影響が考えられる．また，音声波形は時間によって徐々に変化するため，残響によって相関の高い信号が僅かに異なって重なり，音声波形を歪ませる．さらに，目的の音声の調波成分が残響の影響を受けることで，調波構造が不明瞭になる．そこで，調波成分の利用を慎重に行う処理が必要になる．今回は残響環境についての実験を行っていない．雑音の場合と比べて，残響環境では目的音声との相関を持ってしまうため，提案法に新たな処理を加える必要があると考える．

さらに，実際のシステムにおいて提案法を用いることで，F0推定誤りの低減を確かめる実験が必要である．そして，システム全体の精度向上に関係することを確認する．また，使用目的を決めたシステムについて，提案法を組み込む実験も必要である．

付録

A 既存の雑音混入音声のF0推定法

クリーンな音声からのF0推定と同様に，雑音の混入した音声についても様々なF0推定法が提案されている．そこで，処理領域によって大きく分類する．雑音混入音声のF0推定の中で主な手法を表A.1に示す．

表 A.1: 雑音混入音声のF0推定の主な従来法

主な処理領域	特徴	推定法	参考文献
自己相関	複数の窓幅を利用		[13]
	スペクトル包絡候補を利用		[34]
	対数スペクトルのACFを利用	ACLOS	[38]
	調波構造強調処理を利用		[44]
	振幅スペクトルのべき乗を利用	BPPAS	[45]
	適応分析窓長を利用		[60]
	複素音声分析を利用		[61]
ケプストラム	クリッピングと帯域制限を利用	MCEP	[35]
	ハフ変換を利用	hough	[36]
	MCEPに予測残差信号を利用	ARMC	[62]
瞬時周波数	調波成分の占有度を利用		[33]
	瞬時振幅に表れる音声の周期性と調波性を利用	PHIA	[41]
	周期性と調波性に対してエントロピーによる重み付けを利用	EWPH	[42]
	瞬時周波数の不動点を利用	STRAIGHT -TEMPO	[47]
	Comb Filterを利用		[63]

謝辞

本研究を行うにあたり，終始懇切なるご指導を賜った静岡大学工学部 深林太計志 教授に深く感謝致します．

本論文をまとめるにあたり，ご助言や査読を頂き，実験に必要なデータ等についての援助を頂きました静岡大学 中井孝芳 教授，北澤茂良 教授，杉浦敏文 教授に心より感謝いたします．

そして，貴重なご助言や多面に渡って励ましていただいた，Bangladesh の Rajshahi 大学 Dr. Mohammad Ekramul Hamid に心より感謝いたします．

また，田中宏和 氏と古田直哉 氏をはじめ静岡大学深林研究室のみなさまには，日々の研究に関しまして多大なる協力を頂きました．

さらに，筆者の家族には学生生活を送る上での理解と応援を受けました．

以上の方々に改めて感謝の意を表します．

EWPH の MATLAB プログラムを快く提供くださった東京工科大学の石本祐一 博士と STRAIGHT-TEMPO の MATLAB プログラムを快く提供くださった和歌山大学の河原英紀 教授に感謝いたします．

参考文献

[1] 日本音響学会編，音響工学講座，音声，オーム社，1977.

[2] 斎藤 収三，田中 和男，音声情報処理の基礎，オーム社，1981.

[3] W. J. Hess, *Pitch determination of speech signals* (*Algorithms and devices*), Springer-Verlag, 1983.

[4] 古井 貞熙，音声情報処理，森北出版，1998.

[5] H. Fujisaki and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," Journal of the Acoustical Society of Japan (E), vol. 5, no. 4, pp. 233–242, 1984.

[6] 森山 高明，小川 均，天白 成一，"大阪方言合成のための基本周波数生成手法，"電子情報通信学会技術研究報告，vol. 98，no. 423，pp. 25–32，1998.

[7] M. Akagi and T. Ienaga, "Speaker individuality in fundamental frequency contours and its control," Journal of the Acoustical Society of Japan (E), vol. 18, no. 2, pp. 73–80, 1997.

[8] 大野 宏，赤木 正人，"文音声中の基本周波数パターンに含まれる個人性の検討，"電子情報通信学会技術研究報告，vol. 97，no. 586，pp. 89–96，1998.

[9] 浅見 太一，岩野 公司，古井 貞熙，"雑音に頑健な話者照合のための基本周波数情報の利用，"電子情報通信学会技術研究報告，SP2004-15，vol. 104，no. 87，pp. 1–6，2004.

[10] 岩野 公司，関高 浩，古井 貞熙，"雑音に頑健な基本周波数抽出法とその音声認識への適用，"電子情報通信学会技術研究報告，SP2002-13，vol. 102，no. 35，pp. 37–42，2002.

[11] S. G. Knorr, "Reliable voiced/Unvoiced decision," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no. 3, pp. 263–267, 1979.

[12] C. K. Un and H. H. Lee, "Voiced/Unvoiced/Silence discrimination of speech by delta modulation," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, no. 4, pp. 398–407, 1980.

[13] 都木 徹, 清山 信正, 宮坂 栄一, "複数の窓幅から得られた自己相関関数を用いる音声基本周期抽出法," 電子情報通信学会論文誌, vol. J80-A, no. 9, pp. 1341–1350, 1997.

[14] 斎藤 収三, 加藤 勝洋, 寺西 昇, "音声の基本周波数の特性について," 日本音響学会誌, vol. 14, no. 2, pp. 111–116, 1958.

[15] 猪股 修二, "電子計算機による新しい音声の基本周期抽出法の提案," 日本音響学会誌, vol. 16, no. 4, pp. 283–285, 1960.

[16] C. A. Mcgonegal, L. R. Rabiner and Aaron E. Rosenberg, "A semiautomatic pitch detector (SAPD)," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-23, no. 6, pp. 570–574, 1975.

[17] D. H. Friedman, "Pseudo-maximum-likelihood speech pitch extraction," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 3, pp. 213–221, 1977.

[18] 安広 輝夫, "自乗音声波のピーク検出による基本周波数抽出," 日本音響学会誌, vol. 36, no. 10, pp. 487–495, 1980.

[19] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGONEGAL, "A comparative performance study of several pitch detection algorithms," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-24, no.5, pp. 399–418, 1976.

[20] 板倉 文忠, 東倉 洋一, "音声の特徴抽出と情報圧縮," 情報処理, vol. 19, no. 7, pp. 644–656, 1978.

[21] B. GOLD and L. R. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," The Journal of the Acoustical Society of America, vol. 46, no. 2, pp. 442–448, 1969.

[22] N. J. Miller, "Pitch Detection by Data Reduction," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-23, no. 1, pp. 72–79, 1975.

[23] N. C. Geçkinli and D. Yavuz, "Algorithm for pitch extraction using zero-crossing interval sequence," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-25, no. 6, pp. 559–564, 1997.

[24] 中村 正孝, 大元 芳尚, 大和 俊孝, 高山 一男, "波形処理による音声信号の基本周波数の抽出," 電子情報通信学会技術研究報告, vol. 98, no. 557, pp. 23–31, 1999.

[25] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-25, no. 1, pp. 24–33.

[26] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg and H. J. Manley, "Average magnitude difference function pitch extractor," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-22, no. 5, pp. 353–362, 1974.

[27] A. M. Noll, "Cepstrum Pitch Determination," The Journal of the Acoustical Society of America, vol. 41, no. 2, pp. 293–309, 1967.

[28] M. R. Schroeder, "Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement," The Journal of the Acoustical Society of America, vol. 43, no. 4, pp. 829–834, 1968.

[29] 漢野 救泰, 下平 博, "低域スペクトルの予測残差を利用した非定常高騒音環境での有声音区間の検出," 電子情報通信学会論文誌, vol. J80-D-2, no. 1, pp. 26–35, 1997.

[30] 宮林 穎夫, 船田 哲男, “ピッチ乱れ, 波形変動及び雑音付加に対するBPFPピッチ抽出法の性能評価, ”電子情報通信学会論文誌, vol. J85-A, no. 3, pp. 271–281, 2002.

[31] D. A. Krubsack and R. J. Niederjohn, “An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech,” IEEE Transactions on Signal Processing, vol. 39, no. 2, pp. 319–329, 1991.

[32] I. Zeljkovic and Y. Stylianou, “Single complex sinusoid and ARHE model based pitch extractors,” Eurospeech'99, Budapest, Hungary, 1999.

[33] 中谷 智広, 入野 俊夫, “占有度を用いた耐雑音性の高い基本周波数推定法, ”電子情報通信学会技術研究報告, SP2001-138, vol. 101, no. 744, pp. 21–28, 2002.

[34] 柳沢 浩一, 田中 京子, 山浦 逸雄, “スペクトル包絡の時間的連続性を利用した基本周期の検出法, ”電子情報通信学会論文誌, vol. J83-D-II, no. 11, pp. 2087–2098, Nov. 2000.

[35] 小林 載, 島村 徹也, “対数スペクトルにクリッピングと帯域制限を用いる基本周波数抽出法, ”電子情報通信学会論文誌, vol. J82-A, no. 7, pp. 1115–1122, 1999.

[36] 関 高浩, 岩野 公司, 古井 貞熙, “ハフ変換を用いた雑音中の音声からの基本周波数抽出法, ”日本音響学会秋季研究発表会講演論文集, vol. 2001, no. 2, pp. 209–210, 2001.

[37] K. Nishi and A. Shigeru, “An optimal comb filter for time-varying harmonics extraction,” IEICE transactions on fundamentals of electronics, communications and computer sciences, vol. E81-A, no. 8, pp. 1622–1627, 1998.

[38] 國枝 伸行, 島村 徹也, 鈴木 誠史, “対数スペクトルの自己相関関数を利用したピッチ抽出法, ”電子情報通信学会論文誌, vol. J80-A, no. 3, pp. 435–443, 1997.

[39] 阿部 敏彦，小林 隆夫，今井 聖，“瞬時周波数に基づく雑音環境下でのピッチ推定，”電子情報通信学会論文誌，vol. J79-D-2，no. 11，pp. 1771–1781，1996.

[40] 阿竹 義徳，入野 俊夫，河原 英紀，陸 金林，中村 哲，鹿野 清宏，“調波成分の瞬時周波数を用いた基本周波数推定方法，”電子情報通信学会論文誌，vol. J83-D-II，no. 11，pp. 2077–2086，2000.

[41] Y. Ishimoto, M. Unoki and M. Akagi, “A fundamental frequency estimation method for noisy speech based on instantaneous amplitude and frequency,” Proc. Eurospeech2001, vol. 4, pp. 2439–2442, 2001.

[42] 石本 祐一，石塚 健太郎，相川 清明，赤木 正人，“エントロピーによる重み付けを用いた雑音環境下での基本周波数推定，”電子情報通信学会技術研究報告，SP2002-53，vol. 102，no. 247，pp. 13–18，2002.

[43] 長嶋 一将，深林 太計志，“調波構造強調処理を用いた雑音環境下でのピッチ抽出”，電子情報通信学会総合大会講演論文集，基礎・境界，A-4-21，p. 131，2001.

[44] 長嶋 一将，深林 太計志，“調波構造強調処理によるピッチ抽出の耐雑音性向上”，電子情報通信学会総合大会講演論文集，基礎・境界，SA-3-3，p. S-17，2003.

[45] 島村 徹也，高木 浩司，“帯域制限をかけた振幅スペクトルのべき乗に基づく基本周波数抽出法，”電子情報通信学会論文誌（A），vol. J86-A，no. 11，pp. 1097–1107，2003.

[46] Y. Ishimoto, K. Ishizuka, K. Aikawa, M. Akagi, “Fundamental frequency estimation for noisy speech using entropy-weighted periodic and harmonic features,” IEICE transactions on information and systems, vol. E87-D, no. 1, pp. 205–214, 2004.

[47] H. Kawahara, H. Katayose, Alain de Cheveigne, Roy D. Patterson, “Fixed point analysis of frequency to instantaneous frequency mapping for accurate

estimation of F0 and periodicity," Proc. Eurospeech'99, Budapest, Hungary, vol. 6, pp. 2781–2784, 1999.

[48] 深林 太計志，森田 優一郎，“ブラインド信号分離の周波数領域処理または時間領域処理による音声信号の雑音低減，”日本音響学会春季研究発表会講演論文集，3-4-7，pp. 615–616，2002.

[49] M. E. Hamid, K. Ogawa and T. Fukabayashi, "Improved Single-Channel Noise Reduction Method of Speech by Blind Source Separation", Acoustical science and technology, Japan, 2007.

[50] 社団法人音響学会，新版音響用語辞書，コロナ社，2003.

[51] R. Martin, "Spectral subtraction based on minimum statistics," EUSIPCO'94, pp. 1182–1185, 1994.

[52] 細谷 進一，伊藤 憲三，“DSP 処理を目的とした簡便な雑音抑圧処理に関する検討，”日本音響学会秋季研究発表会講演論文集，pp. 147–148，2000.

[53] W. J. Hess, "Pitch and voicing determination, in *Advances in speech signal processing*," Eds. Furui and Sondhi, Marcel Dekker, Inc., 1992, pp. 3–48.

[54] Speech corpus "North wind and short sentences" by Grant-in-aid for scientific research on priority area "Spoken language" and by GSR on developmental scientific research on "Speech database", http://research.nii.ac.jp/src/eng/org/index.html, 音声資源コンソーシアム，国立情報学研究所. — 板橋 秀一：音声情報処理研究用日本語音声データベースの作成，1992.

[55] J. Makhoul, "Spectral analysis of speech by linear prediction, "IEEE Transactions on Audio and Electroacoustics, vol. AU-21, vol. 3, pp. 140–148, 1973.

[56] 深林 太計志，宮下 勝好，鶴木 覚，“母音スペクトルの正規化と個人差を除去する照合方式，”電子情報通信学会技術研究報告，vol. EA82-6，pp. 1–8，1982.

[57] W. Van Summers, David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow, and Michael A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," The Journal of the Acoustical Society of America, vol. 84, no. 3, pp. 917–928, 1988.

[58] 若尾 淳，武田 一哉，板倉 文忠，"種々の定常雑音下における Lombard 音声の認識法の検討，"電子情報通信学会論文誌，vol. J80-D-2，no. 7，pp. 1643–1650，1997.

[59] 小川 哲司，勘場 智之，小林 哲則，"シミュレーションに基づく音声認識システム評価の妥当性の検証，"電子情報通信学会技術研究報告，vol. 106，no. 123，pp. 1–6，2006.

[60] 橋本 論，中村 正孝，"適応分析窓長を持つ自己相関による基本周波数抽出：周波数スペクトル・ベースクリッピングによる雑音耐性の向上，"電子情報通信学会技術研究報告，vol. 105，no. 482，pp. 19–25，2005.

[61] 金城 竜彦，舟木 慶一，"複素音声分析を用いた音声の基本周期推定に関する一検討，"電子情報通信学会技術研究報告，vol. 106，no. 95，pp. 25–30，2006.

[62] 小林 載，島村 徹也，鈴木 誠史，"予測残差信号を利用した改良ケプストラム法による基本周波数の抽出，"日本音響学会春季研究発表会講演論文集，3-7-11，pp. 275–276，1998.

[63] 石本 祐一，赤木 正人，"雑音が付加された音声の基本周波数推定と雑音抑圧，"電子情報通信学会技術研究報告，vol. 99，no. 679，pp. 17–24，2000.

本研究に対する発表論文

論文

[1] 小川 啓太，深林 太計志，“雑音を低減した自己相関関数からの音声のピッチ抽出，”静岡大学大学院電子科学研究科研究報告，第28号，pp. 33–39，Mar. 2007.

[2] M. E. Hamid, K. Ogawa and T. Fukabayashi, "Improved Single-Channel Noise Reduction Method of Speech by Blind Source Separation", Acoustical Science and Technology, Japan, vol. 28, no. 3, pp. 153–164, May. 2007.

[3] K. Ogawa and T. Fukabayashi, "Fundamental Frequency Estimation of Single-channel Noisy Speech using Autocorrelation Subtraction and Cosine Modulation", Journal of Signal Processing, vol. 12, no. 5, Sep. 2008. (採録決定)

[4] 小川 啓太，深林 太計志，“振幅調節と変調を施した振幅スペクトルを用いた雑音混入音声の基本周波数推定，”日本音響学会誌. (2008年5月，条件付掲載可)

国際会議

[5] K. Ogawa, M. E. Hamid and T. Fukabayashi, "Pitch Extraction of Speech from Noise-Reduced Autocorrelation Function using Estimated Noise", Proc. WESPAC IX, Seoul, SP-P-2, 249, Jun. 2006.

[6] M. E. Hamid, K. Ogawa and T. Fukabayashi, "Wide Band Noise Reduction of Speech using Noise Subtraction and Blind Source Separation", Proc. WESPAC IX, Seoul, SP-P-2, 219, Jun. 2006.

[7] M. E. Hamid, K. Ogawa and T. Fukabayashi, "Noise estimation for Speech Enhancement by the Estimated Degree of Noise without Voice Activity Detection", Proceeding Signal and Image Processing-2006, pp. 420–424, Hawaii, Aug. 2006.

国内口頭発表

[8] 小川 啓太，鶴田 哲也，深林 太計志，立蔵 洋介，“ブラインド分離による雑音低減処理を用いた雑音を含む音声のピッチ抽出，”日本音響学会春季研究発表会講演論文集，2-7-8，pp.279–280，Mar. 2004.

[9] 小川 啓太，深林 太計志，立蔵 洋介，“変調処理を用いた雑音を含む音声のピッチ抽出，”電気関係学会東海支部連大，O-215，Sep. 2004.

[10] 小川 啓太，深林 太計志，“雑音低減処理と自己相関係数列のコサイン変調を用いた雑音を含む音声のピッチ抽出，”電子通信情報学会技術報告，EA2004-114，pp. 65–70，Dec. 2004.

[11] 小川 啓太，深林 太計志，“コサイン変調と雑音抑圧を施した自己相関関数を用いた音声のピッチ抽出，”日本音響学会秋季研究発表会講演論文集，2-6-1，pp.259–260，Sep. 2005.

[12] 小川 啓太，M. E. Hamid，深林 太計志，“推定した雑音を用いて雑音成分を低減した自己相関関数からの音声のピッチ抽出，”電子通信情報学会技術報告，SP2005-154，pp. 25–30，Jan. 2006.

[13] 小川 啓太，深林 太計志，“推定した雑音を用いて雑音成分を低減した自己相関関数からの音声のピッチ抽出，”日本音響学会春季研究発表会講演論文集，2-11-5，pp.311–312，Mar. 2006.

[14] 小川 啓太，深林 太計志，“推定した雑音成分を低減した自己相関関数からの音声のピッチ抽出の改良，”日本音響学会秋季研究発表会講演論文集，2-6-13，pp.209–210，Sep. 2006.