博士論文

振幅変調特性に着目した
雑音残響に頑健な基本周波数推定法

三輪　賢一郎

主指導教員　鵜木　祐史

北陸先端科学技術大学院大学
情報科学研究科

平成３０年１２月

要 旨

18 世紀末の機械的音声合成機の発明により始まった音声・音楽情報処理技術は，長足の進歩を遂げ，コミュニケーションの形態を大きく変えてきた．今や音声コミュニケーションはヒトとヒトの直接的な形態には限られず，携帯電話システムのようにヒトとヒトとの間に機械が介在する音声コミュニケーションが普通に見られるようになっている．このような機械が介在する音声コミュニケーションを含めた様々な音声情報処理のシーンにおいて，基本周波数の情報は非常に重要な役目を果たしており，基本周波数を正確に捉えることができるかどうかが，それら各種音声・音楽情報処理の処理精度を左右するものとなる．

基本周波数数推定の研究の歴史は半世紀に及ぶが，その中で数多くの方法が提案されてきた．それらは一般に時間領域での周期的な特徴を利用するもの，周波数領域における調波性を利用して処理するもの，もしくは両者を併用するものとに大別される．これらのほとんどが，時間領域における周期性を利用して自己相関法などにより処理するもの，または周波数領域における調波性を利用して櫛形フィルタなどにより処理するものである．加えて，音源フィルタモデルを仮定した上で，声道フィルターによる影響を取り除くことで，音源である声帯振動の情報を抽出する方法も提案されてきた．近年，静音環境において精密に F0 を推定する方法は確立されてきており，推定の正確性という問題はほぼ解決されてきた．しかしながら，雑音や残響等の外乱への対応という部分では，依然として課題を残している．雑音環境に頑健な基本周波数推定手法も数多く研究されてきているものの，それらの手法は耐残響性については考慮がなされていない．耐残響性が考慮された手法は非常に少ない上，まだ高精度といえるレベルにはなく，それらは耐雑音性については考慮がなされていない．つまり，雑音と残響の両方に対して総合的に対応できる手法は未だ無いのが現状である．

以上の問題意識を受けて，本研究では，雑音・残響を伴う実環境における基本周波数推定の確立に向けた課題解決に取り組むこととした．様々な手法が存在する中で，耐雑音性と耐残響性の両方を兼ね備えた F0 推定法は未だ実現できておらず，雑音と残響が常に混在する実環境への適応という点で閉塞状況にある．従来法

の延長で検討している限り，恐らくこの問題の解決は難しいと考えられる．したがって，従来法とは全く違う新たなアプローチでこの課題に取り組む必要がある．

そこで本研究では，ヒトの AM 音のピッチ知覚の挙動からヒントを得て，音信号の変調成分に着目した上で，振幅変調の復調技術を用いてヒトのピッチ知覚を模擬するという今までに無いアプローチにより，雑音や残響に頑健な新しい F0 推定法を提案した．音信号の変調成分に着目することは，いくつかのメリットがある．振幅変調信号を変調伝達関数（Modulation Transfer Function : MTF）の観点から考えると，雑音や残響などの外乱による影響は，全て変調度（Modulation Index）の低下という単純な図式に落とし込める．よって，時間波形の自己相関を調べたり，周波数スペクトル情報を調べたりするといった複雑な信号処理を経ること無く，変調度の観測から外乱の影響を把握することができる．また，外乱により変調成分の振幅は小さくなるものの，変調成分の周期は保持されるため，音信号の変調成分自体は外乱に強い信号成分であると同時に，必要に応じ変調度をパラメータとした波形回復機構を付加することも可能である．

シミュレーションの結果から，提案法が時変の調波信号に対して確実に対応でき，加えて頑健性を備えていることを確認した．さらに提案法の特性に合致した応用先の一例として楽器音の音高推定を考え，提案法が楽器音のような信号に対しては十分適用が可能であることを示した．一方で，提案法は，現状ではその扱える信号に制約が存在することも明らかとなった．現提案法は，理想的な調波信号に対しては頑健に適応できるものの，音声信号に対しては頑健性の点で優位性は確認することは出来なかった．これら音声信号に対する頑健性の課題に対しては，復調に適切な調波を選択的に利用する機構や，周波数領域に応じて適切な分析窓長を選択する機構，FM 復調の適用等が有効と考えられる．

本研究の全体を通して，ヒトの AM 音のピッチ知覚からもヒントを得て，信号の AM 成分に着目して振幅変調の復調技術を応用するという今までに無いやり方で，雑音や残響に頑健な F0 推定法が構築可能であることを示した．このことにより，半世紀以上もの長きにわたって解決が困難であった，雑音残響に頑健な F0 推定法を考える上で，ひとつの方向性を示すことができた．

現状の提案法には依然として課題が存在するものの，特に理想的な調波信号に対しては非常に頑健な手法であることから，その貢献分野は確実に存在すると考

えられる．例えば，他の F0 推定法の一部機能として提案法を実装することにより，他の F0 推定法を頑健性の面で補完するような役割が期待できる．このように，提案法のエッセンスは将来の音楽／音声情報処理技術に大いに貢献するものと考えられる．

目 次

図 目 次

第 1章

序論

1.1 はじめに

我々は日常生活の中で，音の高さの情報を常に意識している．例えば，音楽を鑑賞するときは，歌手や様々な楽器から発せられる音の高さの時間的変化を，我々はメロディとして楽しんでいる．歌を歌う場合は，自身の口から発せられる声の高さと伴奏されるメロディの音の高さの両方の情報を，自分の耳で聞いて確認しながら，正確な音程で歌えるように声の高さを意識的に調節して歌っている．会話する場合，聞き手に話の真意がよく伝わるようにイントネーションやアクセントを付加するが，巧みに声の高さを調節しながら発話を行っている．一方，聞き手側は，相手の声の高さの変化に現れる抑揚やイントネーション，アクセントなど（韻律情報）が付加された発話文を聞くことで，単語の特定，強調部分の認識，感情の把握など，相手の真意を正確に汲み取ることができる．我々が普段何気なく行っている日常の営みには，音の高さの情報，つまりピッチ (pitch) の情報が大いに関係している．

ピッチはヒトの聴覚器官において，聴神経の神経興奮パターンのピーク（場所情報）として知覚されるとともに，聴神経の発火間隔（時間情報）としても知覚される [1]．つまりピッチは，音の高さ（時間情報）に加え，音の周期性（場所情報）や音調性（場所情報）によってもヒトに知覚される [2]．

これらピッチと強く関連している物理量が，基本周波数（Fundamental frequency: F0）である．楽器音を例に取ると，国際標準化機構（International Organization for Standardization：ISO）において楽器の中央ハの一点イの周波数は 440 Hz と定められているが [3]，この 440 Hz が中央ハのイ音の基本周波数に相当する．この基本周波数の時間的な変化が，すなわちメロディを表すことになる．一方，ヒトの発する音声の場合は，1 秒あたりに声帯が振動する回数が基本周波数に相当する．この声帯振動によってもたらされる有声音の基本周波数の時間的変化によって，抑揚やアクセントやイントネーションが表現されることになる．さらに，声帯振動数の時間的変化のパターンには，男女差や個人差が存在し，性別や個人の話し方の癖を特徴付けるものともなる．

ここで，基本周波数をもつ音の物理的な構成について考える．音声や楽器音等の周期的な音信号においては，その基本周期の周波数成分とその整数倍の周波数

成分から構成される特徴的な構造（調波構造）を形成している．図 1.1 は男声による音声信号であるが，その周波数構造は基本周期であるおおよそ 100 Hz とその整数倍から成る調波構造を形成していることが観測できる．この基本周期の周波数成分（100 Hz）をすなわち基本周波数と呼んでいる．その音信号が複数の純音（pure tone）で構成された複合音 (complex tone) であれば，基本周波数は最も低い周波数成分である基音 (fundamental tone) の周波数のことを指す．

18 世紀末の機械的音声合成機の発明により始まった音声・音楽情報処理技術は，長足の進歩を遂げ，コミュニケーションの形態を大きく変えてきた [4]．今や音声コミュニケーションはヒトとヒトの直接的な形態には限られず，携帯電話システムのようにヒトとヒトとの間に機械が介在する音声コミュニケーションが普通に見られるようになっている．このような機械が介在する音声コミュニケーションを含めた様々な音声情報処理のシーンにおいて，基本周波数の情報は非常に重要な役目を果たしており，基本周波数を正確に捉えることができるかどうかが，それら各種音声・音楽情報処理の処理精度を左右するものとなる．

基本周波数の解析を伴う音声・音楽情報処理技術が利用されるシーンの例を図 1.2 に示す．例えば，音声認識技術においては，音声中の基本周波数から得られた韻律情報を積極的に利用すると，その音声認識精度も向上することが報告されている [5]．音声分析合成技術においては，音声を，声帯振動を指す音源情報と声道フィルターによるスペクトル包絡情報とに分離した上で音声合成が行われる．その合成音声の品質の向上には音源情報がきわめて重要であり，そのためには正確な基本周波数の抽出が必要である [6]．一例として移動体通信の分野では，音声符号化技術に音声分析合成が積極的に使われており，基本周波数推定精度が伝送音声品質を大きく左右するだけに，基本周波数の正確な推定は非常に重要である [6, 7]．音楽の現場においては，異種複数楽器による混合音和音を含む楽器音の音高推定・自動採譜技術に基本周波数推定技術が用いられている [8, 9, 10, 11]．実際にはこれらの音声・音楽情報処理を効果的に実施するために，雑音や残響を除去することで音声を強調するプロセスが前処理段や後処理段に置かれる．この音声強調処理において，雑音や残響の影響を除去するために基本周波数の情報がやはり利用されており [12, 13]，フロントエンドにおける基本周波数の正確な推定が音声強調処理精度の向上につながる．また聴覚情景解析技術の一例として，音源分離を計算

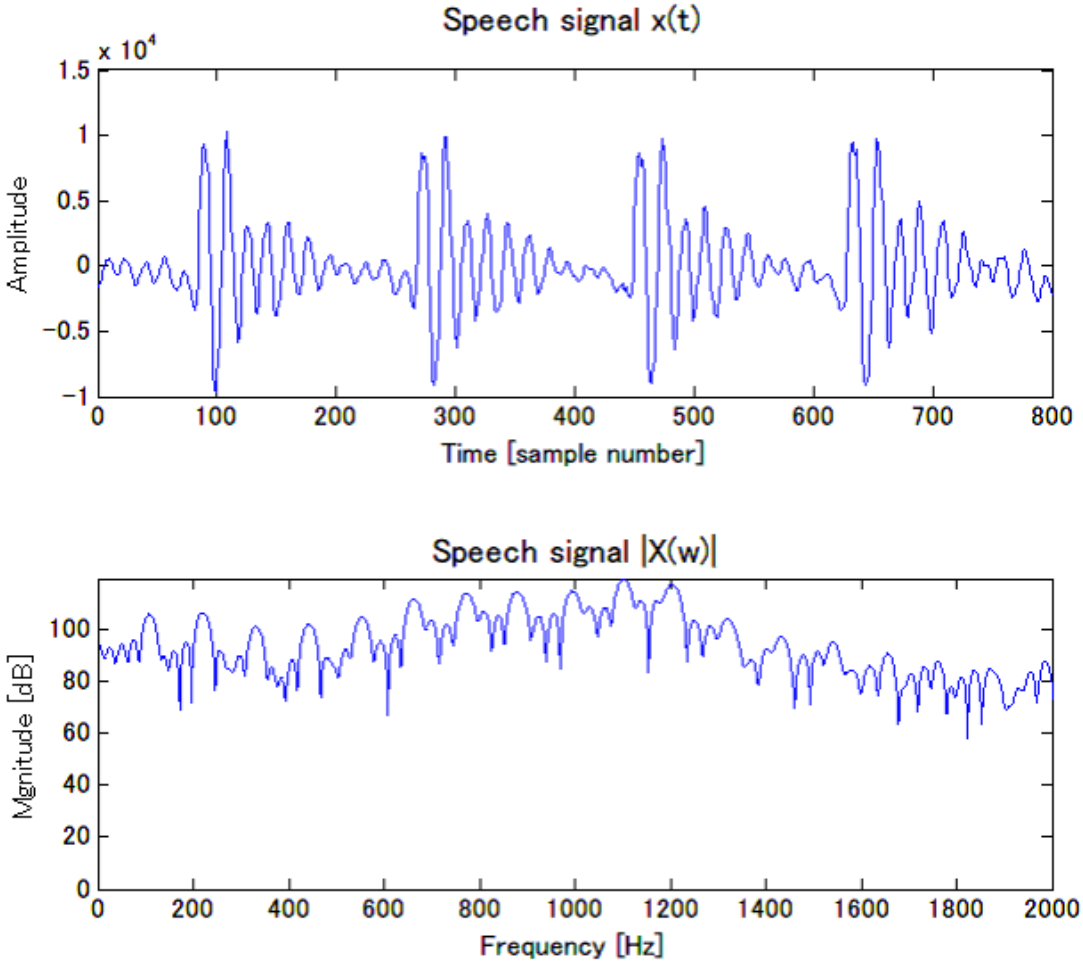

図 1.1: 音声の信号波形と周波数構造

論的聴覚情景分析を適用して解く場合には，個々の構成音の瞬時特徴成分としての基本周波数の解析がやはり鍵となる [14, 15, 16].

このように，基本周波数を用いた音声・音楽信号処理の工学的応用範囲は大変多岐にわたっており，基本周波数はきわめて重要不可欠な特徴量であるといえる．これら，音声信号処理に関連する音声工学分野のさらなる発展は，基本周波数推定技術にかかっていると言っても過言ではない.

1.2 F0 推定技術の課題

F0 推定技術には，大きく分けて下記の 3 つの課題がある.

（1）正確性

（2）頑健性

（3）即時性

これら課題について，ここでは音声の場合を例にとって考えてみる.

まず（1）の正確性を考える．ヒトの音声の発生器官の構造を図 1.3 に示す．ヒトが音声を生成する際には，肺からの空気流による声帯振動が音源となり，声帯音源が声道フィルタを通り，さらに口唇から放射されて生成される．そのため，音声から直接声帯音源の振動数を観測することはできない．音声の基本周波数を知るためには，観測した信号から声道特性や口唇からの放射特性を取り除いたものから推定しなければならない．このように基本周波数の正確な推定は実際容易ではない.

加えて（2）の頑健性としては，観測された音声信号には雑音や残響等の外乱の影響が混入しており，音源情報の特徴抽出を困難なものにする.

さらに（3）の即時性として，リアルタイム処理を行うには，基本周波数の推定アルゴリズムの計算コストや処理遅延は極力小さいものが望ましく．加えて目的やアルゴリズムに合った適切なハードウェアを選定する必要もある.

基本周波数数推定の研究の歴史は半世紀に及ぶが，その中で数多くの方法が提案されてきた [7, 17, 18, 19, 20, 21, 22]．それらは一般に時間領域での周期的な特徴を利用するもの，周波数領域における調波性を利用して処理するもの，もしくは両者を併用するものとに大別される．これらのほとんどが，時間領域における

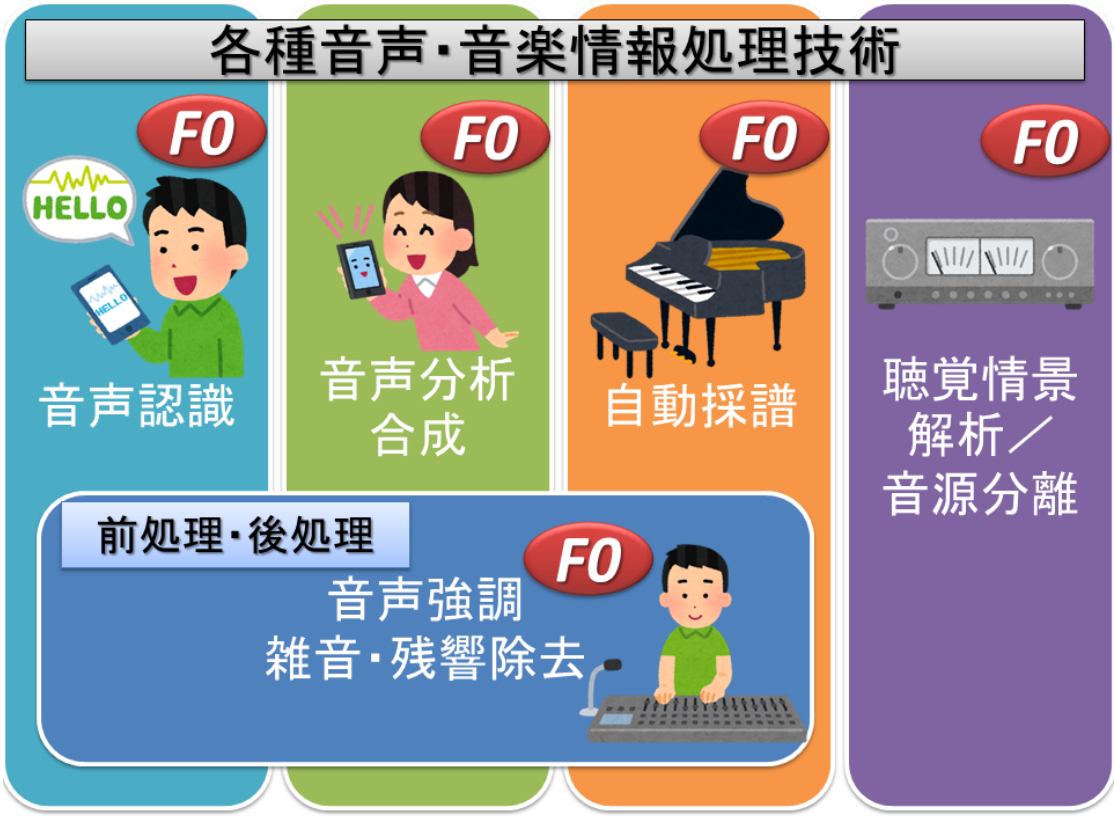

図 1.2: 基本周波数（F0）推定技術の応用先

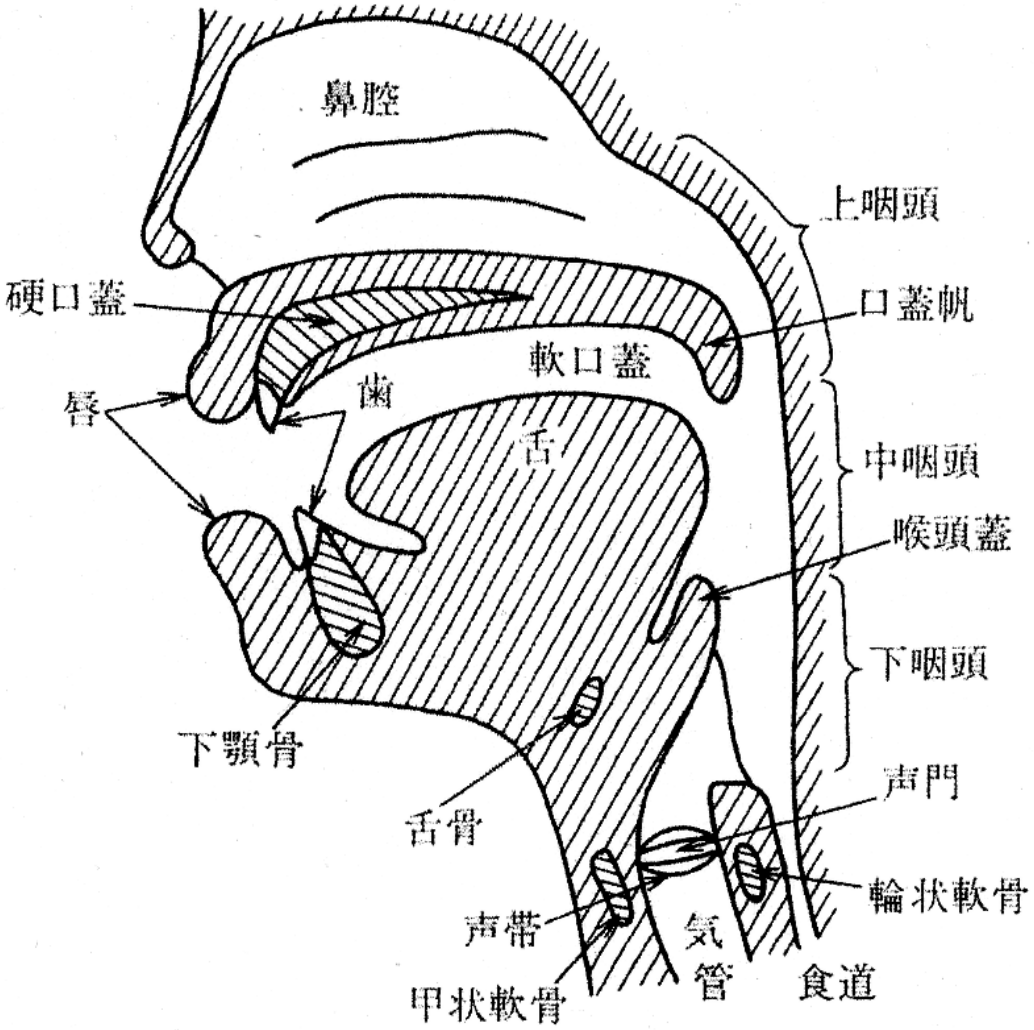

図 1.3: ヒトの音声の発声器官（鏑木）[6]

周期性を利用して自己相関法などにより処理するもの，または周波数領域における調波性を利用して櫛形フィルタなどにより処理するものである．加えて，音源フィルタモデルを仮定した上で，声道フィルターによる影響を取り除くことで，音源である声帯振動の情報を抽出する方法も提案されてきた．

近年，静音環境において精密にF0を推定する方法は確立されてきており，推定の正確性という問題はほぼ解決されてきた．しかしながら，雑音や残響等の外乱への対応という部分では，依然として課題を残している．雑音環境に頑健な基本周波数推定手法も数多く研究されてきているものの，それらの手法は耐残響性については考慮がなされていない．耐残響性が考慮された手法は非常に少ない上，まだ高精度といえるレベルにはなく，それらは耐雑音性については考慮がなされていない．つまり，雑音と残響の両方に対して総合的に対応できる手法は未だ無いのが現状である．一方で，計算量を低減した方式も検討されてきているが，それらは多くの場合外乱の無い環境を前提としている．

以上のように，F0推定の研究開発において，雑音と残響を含む外乱への対応は現在においても未解決の課題である．

1.3　本研究の動機と目的

前述の通り，基本周波数推定技術は様々な音声・音楽信号処理に不可欠な技術である．実環境では雑音や残響が存在しており，これらは対象信号を複雑に歪ませるため，基本周波数の推定を困難なものにさせる．実環境に適応可能な基本周波数推定技術が確立されない限り，音声・音楽信号処理技術の発展，ひいては音声工学の発展は望めない．したがって，実環境における外乱に頑健な基本周波数推定問題は解決されなければならない課題であり，それら課題解決のための研究開発は我々に課せられた使命でもある．

以上の問題意識を受けて，本研究では，雑音・残響を伴う実環境における基本周波数推定の確立に向けた課題解決に取り組むこととした．様々な手法が存在する中で，耐雑音性と耐残響性の両方を兼ね備えたF0推定法は未だ実現できておらず，雑音と残響が常に混在する実環境への適応という点で閉塞状況にある．従来法の延長で検討している限り，恐らくこの問題の解決は難しいと考えられる．した

がって，従来法とは全く違う新たなアプローチでこの課題に取り組む必要がある．

そこで本研究では，ヒトのピッチ知覚に立ち返ってこの問題を検討する．特に頑健性に着目して検討を実施し，雑音や残響などの外乱に頑健なF0推定法の確立を目指す（図1.4）．

本研究により頑健で正確な基本周波数推定法を確立することができれば，実環境で利用する様々な応用技術への貢献が期待できる．特に基本周波数の情報を利用する音声強調技術や音源分離技術等は，聴覚情景解析や音声認識技術と密接な関係があり，本研究はそれら技術の処理性能や効率の飛躍的な向上に貢献できる．また，実環境下で正確な韻律情報を必要とする話者認識技術や感情認識技術などの性能向上にも寄与できる．さらに，今日の移動体通信においては，基本周波数を重要な特徴量として用いる音声分析合成が利用されているが，雑音残響にロバストな基本周波数推定法は通話品質の向上に大いに貢献できる可能性がある．

1.4　本論文の構成

本論文は全部で6章から構成される．

第1章では，本研究で対象とする分野の背景や関連する応用分野などを簡単に述べる．次いで，現状の課題や本研究の目的を述べる．

第2章では，従来からの基本周波数推定法に関して述べる．また，代表的な従来手法を用いた検証シミュレーションを実施し，その内容と問題点を概観する．

第3章では，ヒトのピッチ知覚から得られる知見に着目し，それから着想を得た振幅変復調技術を用いたアプローチについて述べる．

第4章では，時変信号に提案法を適用していくために必要な提案法の拡張について述べ，改良後の提案法の頑健性を考察する．また，提案法の応用の一例として，楽器音に対する適性を確認する．

第5章では，音声を模擬した時変信号に対する提案法の挙動を考察し，基本周波数推定法としての提案法の現状での性能限界を明らかにする．

第6章では，本研究で明らかにしたことを述べ，残された課題ならびに将来に向けた課題を述べる．最後に提案法の将来展望について私見を述べる．

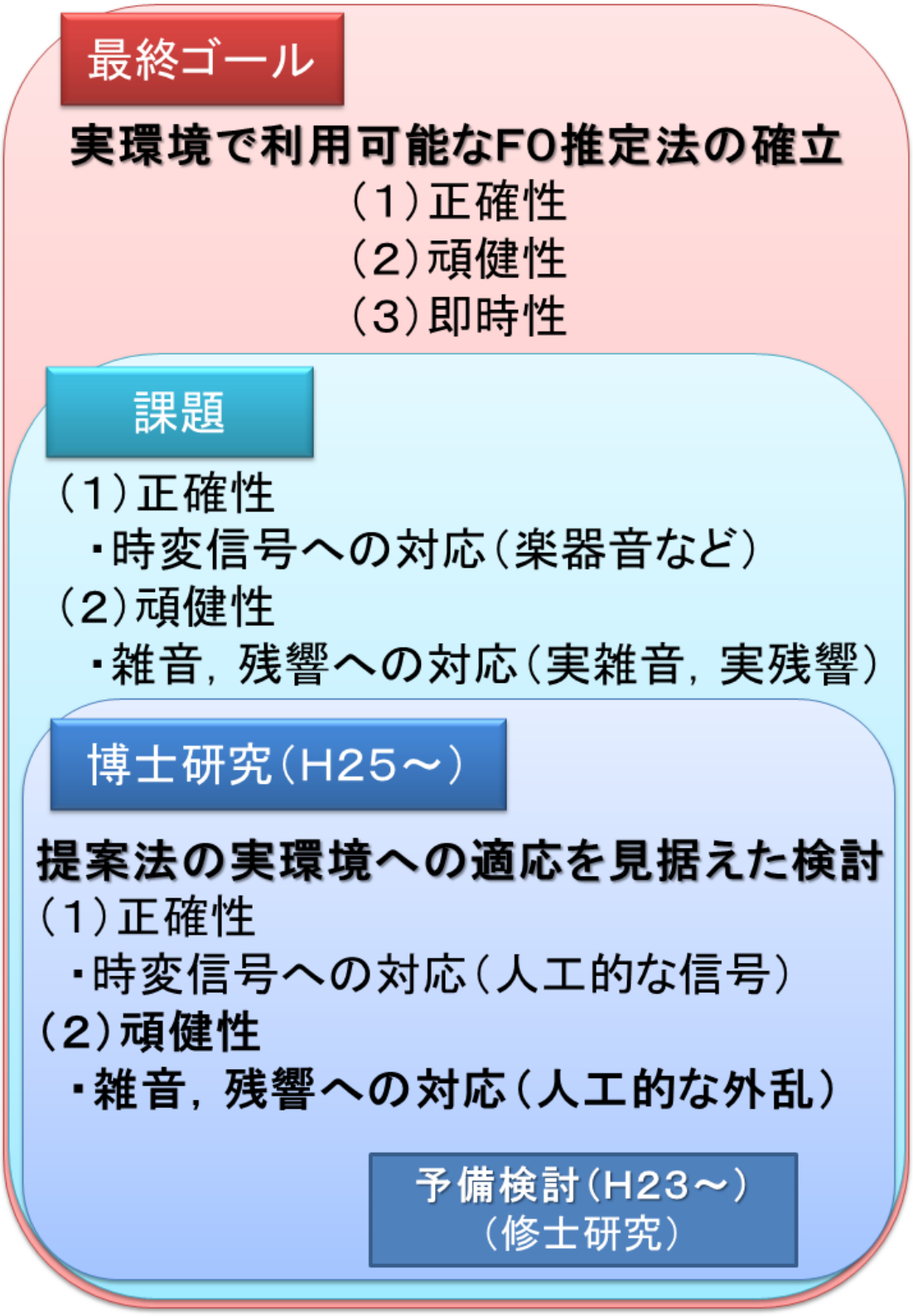

図 1.4: 本研究の位置づけ

第 2章

従来の基本周波数推定法

本章では，従来の基本周波数（F0）推定法について概観する．

2.1　従来法の種類と特長

基本周波数の推定については，半世紀以上にわたり様々な手法が提案されており [18, 19]，おおむね次の 2 種類に大別される．

1. 時間領域に現れる周期性の特徴（時間情報）を利用した手法
2. 周波数領域に現れる調波性の特徴（周波数情報）を利用した手法

音声（有声音）は，時間方向に基本周期を持つ波形となるため，この周期性の検出を行うものが 1 の手法である．対して，音声を周波数分解した場合に，周波数方向に基本周波数とその整数倍の周波数帯域に調波成分が櫛の歯状に出現するが，この調波成分のうちの最低周波数成分や調波成分の周波数間隔の検出を行うものが 2 の手法である．また，周期性と調波性の両方に着目した手法も存在する．

2.1.1　周期性の特徴を利用した手法

周期性の特徴を利用する手法はいくつか存在するが，音声信号の時間波形からダイレクトに推定する方法や，音源・フィルタモデルを仮定する手法などが代表的である．以下に，主な手法について簡単に示す．

音声信号の時間波形を基に推定する方法

まず古典的な手法として**ゼロ交差法** [23] は，時間波形上で振幅が 0 となる点を微分処理等で検出し，その時間間隔から基本周波数を求める手法である．一方，**ピーク検出法** [24] は，時間波形上の振幅のピーク値を微分処理等で検出し，その時間間隔から基本周波数を求める手法である．これらの手法は計算量が少なく実時間処理が可能だが，信号波形そのものを観測するものであることから，雑音に弱いという欠点を持っていた．対して，信号波形から自己相関関数を求め，その相関値のピーク時刻を基に波形の周期間隔を求めて基本周波数を抽出することとした手法が**自己相関法** [25] である．自己相関法は，ピーク検出法などと比べて比較的

雑音の影響を受けにくいという性質を持つものの，窓長の設定によっては倍ピッチの誤推定（真値の 1/2 倍，もしくは 2 倍誤り）が見られたり，あるいは推定が行えないケースも存在する．**多重窓長自己相関法：ACMWL（Auto-Corrlation Multiple Window Length）法** [26] は，複数以上の異なる幅の分析窓を用いて自己相関関数を計算し，それらいくつかの候補の中から最適な窓長と基本周波数を選択することで，それら倍ピッチ等の誤推定の低減を図ったものである．ただし，複数以上の分析窓を用いることから，計算量は分析窓の数だけ増大することとなる．また，**YIN 法** [22] は，自己相関関数のボトム値に着目して，周期間の振幅の変化の影響を受けにくい振幅差関数の重み付けにより周期性を検出する手法である．**PYIN（Probabilistic YIN）法** [28] は，YIN の改良法であるが，複数以上の予測候補を用いつつ，最終的な軌跡の決定には HMM も併用することで F0 推定精度を高めたものである．ただし，YIN 法，PYIN 法はいずれも処理が複雑で，計算量は大きい．これら自己相関関数法よりも演算処理を軽くすることを目的として考え出されたのが，平均振幅差関数 AMDF の距離尺度を利用して周期性を検出する **AMDF(Average Magnitude Difference Function) 法** [29] である．これら相関処理は，比較的雑音には強いが，周期性雑音には弱く，フォルマントの影響も受けやすいという性質がある．

以上の手法は分析フレーム内の平均を取って基本周波数を推定しているため，その推定結果は常に誤差を伴っている．**基本波フィルタリング（VFWF）法** [30] は，帯域通過フィルタを適応的に利用して基本波付近の成分を抽出した中から F0 を連続的に推定可能な手法である．ただしその機構上，前処理として，フィルタの特性を決めるために従来法を用いて F0 推定を行うため，プラスアルファでの計算コストを常に伴う．その発展形である **DIO 法** [31] は，複数以上のフィルタを用いて得られた複数以上の基本波から F0 を求めることにより計算量を低減した手法である．また離散ウェーブレット変換を基本波フィルタリングに用いることによって極力計算量を低減した手法 [32] も提案されている．

ソース・フィルタモデルを仮定する方法

音源フィルタモデルを仮定した上で，声道フィルタによる影響を取り除くことで，音源である声帯振動の情報を抽出する方法である．**LPC(Linear Predictive**

Coding) 法 [33] は，線形予測分析（LPC 分析）により得られた残差信号の周期性を自己相関処理により検出する手法である．**LPC-SIFT 法** [34] は，LPC 分析による逆フィルタにより声道特性の影響を取り除いて得た信号に対して，自己相関処理により周期性を検出する手法である．これらの手法は，比較的短い分析窓長が使えることと，少ない計算量で処理できることなどが特長である．ただし，LPC 分析は雑音の影響を受けやすいという問題があり，複素 LPC 残差を用いることで耐雑音性を高めた手法 [35] も提案されている．

2.1.2 調波性の特徴を利用した手法

短時間フーリエ変換などから得られる音声の時間・周波数表現から，音声の調波構造を分析することで基本周波数を得る手法である．フーリエ変換によりその調波構造を調べる手法や，ソース・フィルタモデルを仮定する方法などが存在する．以下に，主な手法について簡単に記す．

フーリエ変換等により調波構造を分析する方法

STFT（Short-time Fourier Trabsform）法は，短時間フーリエ変換などから得られる対数スペクトル上に現れる調波構造から基本周波数を求めるものであり，比較的雑音に強いという特長がある．**自己相関関数を用いる STFT 法** [36, 37]，耐雑音性を高めるために **Comb フィルタを用いる STFT 法** [11, 38]，基本波と高調波に対応する周波数成分の荷重和を部分的に検出することで調波成分の一部欠落にも対応できる **SHS(Sub-Harmonic Summation) 法** [39]，さらに音源・声道フィルターモデルを仮定した上で音源情報を抽出してその周期性・調波性などから基本周波数を推定する**ラグ窓法** [40] や**リフター法** [41] などがある．この他，耐雑音性を高めるために，周波数ドメインにて雑音レベルに応じたべき乗処理を施す方法 [42] もある．近年，振幅スペクトルとのこぎり波のスペクトルとの相互相関関数により調波構造を分析する **SWIPE 法** [43, 44] が，耐雑音性を持つ高精度な手法として知られているが，誤差を低減するために複雑な演算を行うことから，計算コストは大きい [45]．

ソース・フィルタモデルを仮定する方法

ケプストラム法 [46, 47] は，対数振幅スペクトラムを逆フーリエ変換して求めたケプストラムのうち，音源情報のケフレンシー領域を観測して得られた基本周期に対応するピーク位置から基本周波数を求めるもの．改良法の中には，逆フーリエ変換の前処理としてクリッピング処理を追加した**クリップストラム法** [48] や，帯域制限を追加した**改良ケプストラム法** [49]，クリッピング処理と帯域制限により調波構造をクリアにすることでさらに耐雑音性を高めた手法 [50] 等がある．さらに**複素ケプストラム法** [51, 52] は，位相情報を包含する複素ケプストラムを用いることで耐残響性を高めた方式である．これらケプストラム分析による方法は，一般的にフォルマントの影響は受けにくいが，雑音に対しては弱いという性質を持つ．

瞬時周波数の調波性を観測する方法

TEMPO 法 [53, 54] は，音声の瞬時周波数を利用し，調波周波数におけるフィルタ出力での安定な不動点を抽出することで基本周波数を推定する手法で，静音環境では極めて正確な推定が可能である．これらの手法は，耐雑音性に課題があったが，フロントエンドで非周期性の検出を行って F0 の軌跡を修正することで耐雑音性を高めた**改良法** [55] や，帯域幅方程式を利用した重み付けも併用することで耐雑音性を高めた **IFHC 法** [56] なども提案されている．

2.1.3 周期性と調波性の特徴を利用した手法

PHIA 法 [57] は，瞬時振幅の周期性と調波性からそれぞれ基本周波数の候補を求め，最終段で瞬時周波数を用いる手法である．雑音抑圧処理の導入によって耐雑音性の大幅な向上が図られているが，処理が多段に渡る分計算コストも大きい．

2.1.4 先行研究から言えること

以上のように，基本周波数の推定に関してはありとあらゆる方法・手段が試されてきた中で，それら手法による傾向は下記のように要約できる．

- ゼロ交差やピーク検出を用いる古典的な手法は，計算量は小さいが，雑音には弱い．
- 相関処理は，比較的雑音には強いが，フォルマントの影響は受けやすい．
- ケプストラム処理は，フォルマントの影響は受けにくいが，雑音には弱い．
- 各手法において耐雑音性を向上させるとそれなりに処理も複雑となり，耐雑音性と計算量はトレードオフの関係がある．

加えて，耐雑音性を向上させるための研究は比較的多く提案されているが，耐残響性を指向した研究はきわめて少ないという事実もある．つまり，処理がシンプルで計算量が小さく，かつ耐雑音性と耐残響性をも備えた手法は，現段階では存在しない，と結論づけられる．

2.2 従来法の F0 推定精度

ここでは，上記従来法の中から代表的な 25 手法に着目し，実際にその推定精度を試験により確認することで，F0 推定技術の現状を把握する．

今回試験に用いた音声信号は，ATR データベースの男声と女声それぞれ 2 名による三連続母音/aoi/である．試験に用いた環境は，静音環境に加え，ホワイトノイズによる 5 種類（20，10，0，−5，−10 [dB]）の SNR による雑音環境，統計的室内インパルス応答（Schroeder のインパルス応答 [58]）による 5 種類（0.1，0.3，0.5，1.0，2.0 [sec]）の残響時間による残響環境を用いた．雑音環境と残響環境はいずれもランダム関数により各コンディションにつき 10 回ずつ試験を実施し，正答率（許容誤差 5 %）により評価を行った．以下に，その結果をまとめて示す．なお，結果の詳細は付録に示す．

まず，クリーンな環境におけるシミュレーション結果を図 2.1 に示す．今回用いた 25 種類の手法のうち，静音環境において 4 種類の音声に対する F0 推定の平均正答率が 90%を超えたのは，自己相関法，多重窓長自己相関法（ACMWL），LPC を用いた平均振幅差関数法（AMDF-LPC），ラグ窓を用いた短時間フーリエ変換（STFT-Lag），SWPE’，オリジナルのケプストラム法，改良ケプストラム法，LPC

法，TEMPO 法，IFHC，PHIA の計 11 手法であった（ただし今回の基準値とした TEMPO2 は除く）．この結果から，静音環境においても，音声の正確な F0 推定に利用可能な手法は比較的限られているということがあらためて明らかとなった．

次に，比較的雑音が少ない（高 SNR）環境（20 及び 10 dB）と，残響時間が比較的短い場合（0.1，0.3，0.5 [sec]）におけるそれぞれの正答率の平均値を図 2.2 に示す．耐雑音性という点では，正答率が 80 %を超えたものが 15 手法確認できた．一方で，耐残響性という面では，残響時間は比較的短いにもかかわらず，正答率が 80 %を超えた手法は確認できなかった．以上から，ある程度の雑音に対応できる手法は比較的数多く存在するが，短時間の残響環境に対応できる手法はほぼ皆無であることが改めて判明した．

次に，更に厳しい雑音環境（SNR=0，−5，−10 dB），ならびに残響時間が長い場合（1.0，2.0 [sec]）における正答率の平均値を，図 2.3 に示す．耐雑音性という点では，正答率が 80 %を超えるものは存在しないが，70 %を超える手法は，SWIPE’，複素ケプストラム法，PHIA の 3 手法が確認できた．うち PHIA は雑音抑圧機構を備えているから当然として，特に耐雑音性は考慮されていない複素ケプストラム法が健闘している．一方で，耐残響性という面では，正答率が 50 %を超えた手法さえ確認できず，正答率が 40 %台に届いたものも，櫛形フィルタを用いた短時間フーリエ変換法だけであった．以上から，厳しい雑音環境に対応できる手法は若干存在するが，長時間の残響環境に対応できる手法は皆無であることが改めて明らかとなった．

今回の試験の総合結果を図 2.4 として示す．ここでは，全雑音環境（SNR=20，10，0，−5，−10 dB）と全残響環境（0.1，0.3，0.5，1.0，2.0 [sec]）のそれぞれにおける平均正答率を 2 次元平面上にプロットしている．耐雑音性という尺度（横軸）では，正答率が 80 %を超えるものは SWIPE’ のみとなっている．一方で，耐残響性という尺度（縦軸）では，正答率が 60 %を超えた手法は存在せず，50 %に届いたものも 3 手法（SWIPE’，複素ケプストラム法，櫛形フィルタを用いた短時間フーリエ変換法）にとどまっている．このことは，耐雑音性と耐残響性の両方を兼ね備える手法は皆無であることを示している．

以上の結果から，F0 推定法の現況は下記のように要約される．

- 静音環境において，全ての手法が音声の F0 推定を高精度に行えるわけでは

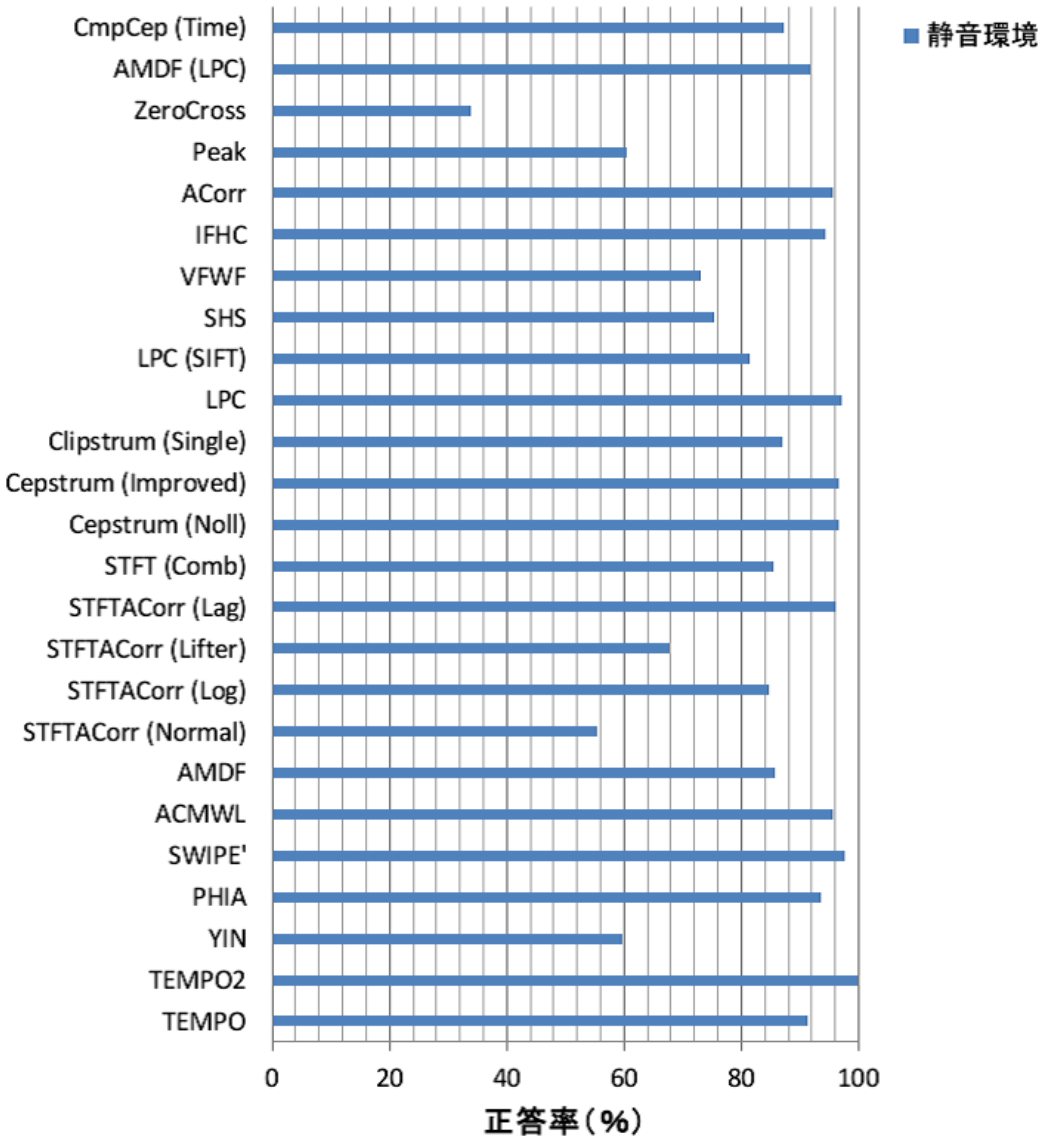

図 2.1: 静音環境における各手法毎の平均正答率（%）

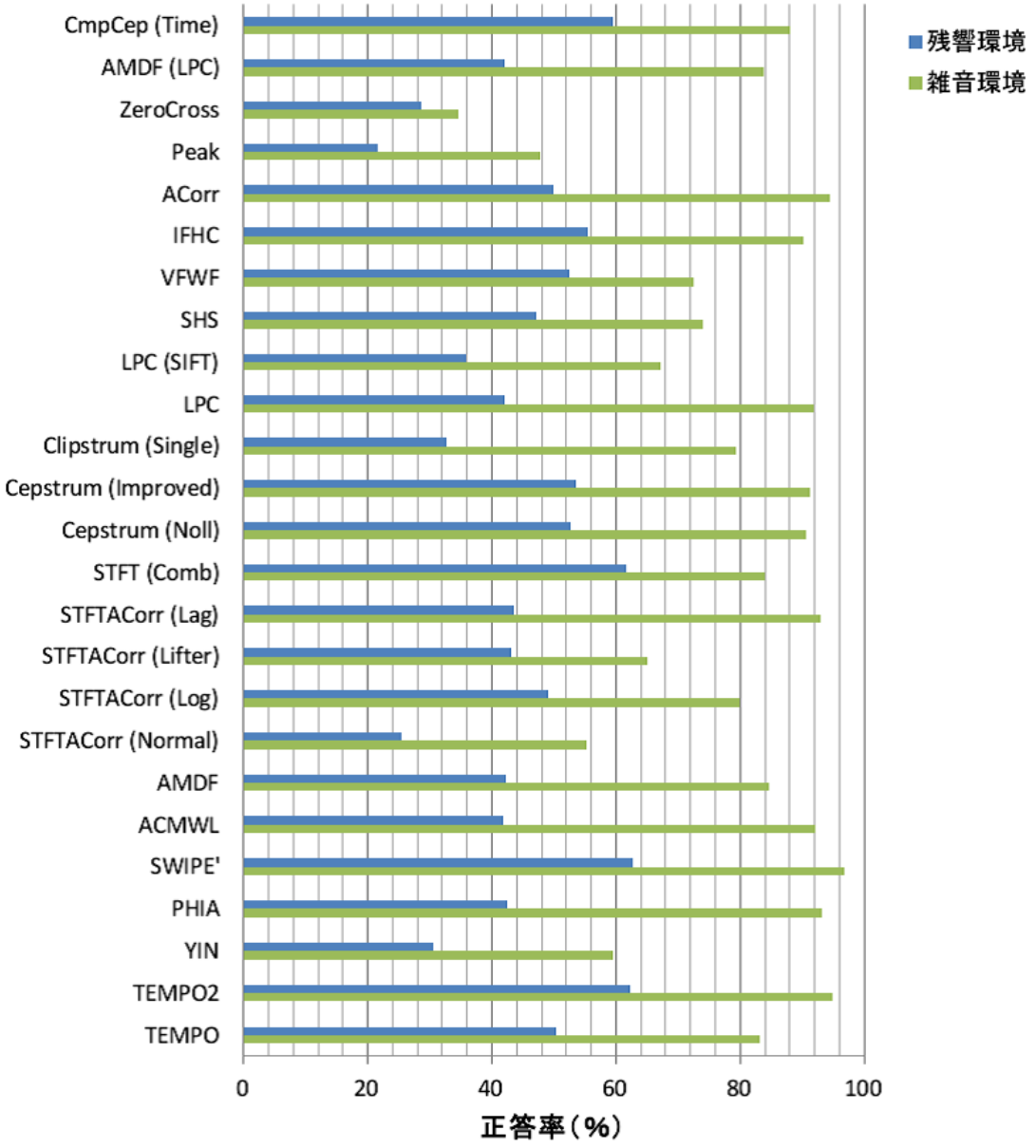

図 2.2: 雑音環境（高 SNR）及び残響環境（短時間）における各手法毎の平均正答率（%）

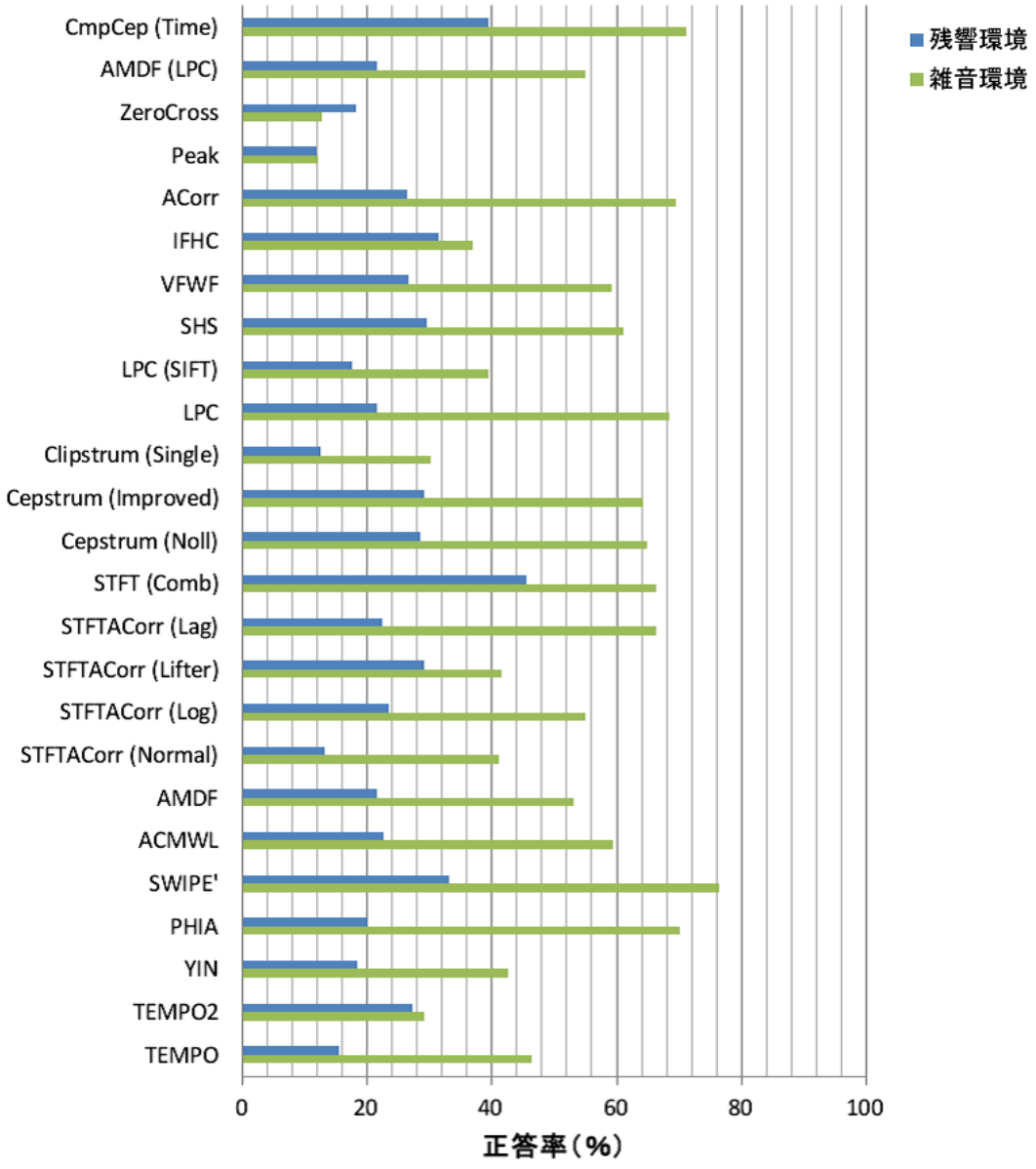

図 2.3: 雑音環境（低 SNR）及び残響環境（長時間）における各手法毎の平均正答率（%）

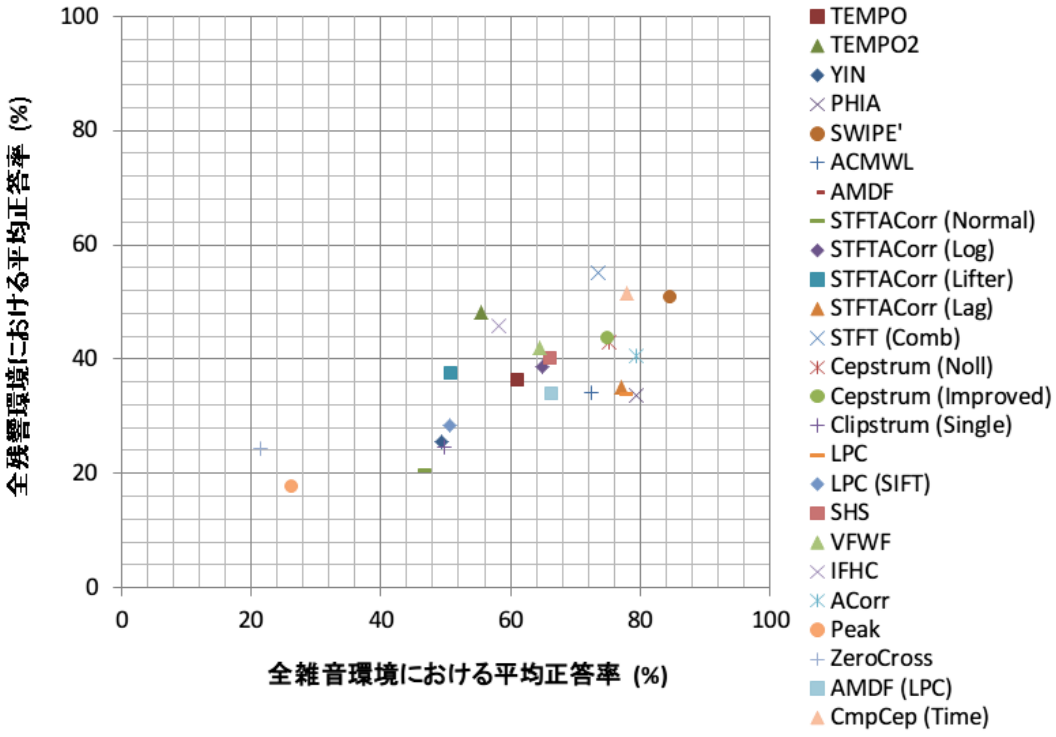

図 2.4: 全雑音環境及び全残響環境における各手法毎の平均正答率（%）

ない．

- 雑音に対する耐性を備えた手法はいくつか存在するものの，低 SNR の環境にも対応できる手法はごく一部である．
- 残響に対する耐性を持つものも若干存在するが，いずれも高精度といえるレベルにはない．

これらのことから，雑音と残響が混在する実環境に対応でるような頑健な手法は，残念ながら存在しないという現状があらためて浮き彫りとなった．

2.3 課題

基本周波数の推定に関してはありとあらゆる方法・手段が試されてきた中で，シンプルもしくは古典的な手法は概して雑音等の外乱に弱く，そのため耐雑音性を向上させる提案も数多く提案されてきた．しかしながら，耐雑音性を向上させるためにはそれなりに信号処理も複雑なものとなり，実時間処理には不向きなものとなってくる．加えて，耐雑音性を向上させるための研究は比較的多く提案されているが，耐残響性を指向した研究はきわめて少ないという現状がある．

また，今回の事前試験を通して，静音環境においてさえ，音声信号に有効な手法はごく限られていることが改めて判明した．従来法のうちのごく一部ではあるが，厳しい雑音環境下でも高い推定精度を示す手法も存在しており，耐雑音性という部分では解決がなされてきているとも言える．一方，残響環境において高い推定精度を示した手法は皆無であった．さらに進んで，耐雑音性と耐残響性の両方を兼ね備えた手法となると，もはや絶望的な状況であり，本当の意味で実環境に対応できる手法は存在しない．

このように，様々な手法が存在する中で，耐雑音性と耐残響性の両方を兼ね備えた F0 推定法は未だ実現できておらず，雑音と残響が常に混在する実環境への適応という点で閉塞状況にある．従来法の延長線上で検討している限り，恐らくこの問題の解決は難しいと考えられる．したがって，従来法とは全く違う新たなアプローチでこの課題に取り組む必要がある．そこで本研究では，ヒトのピッチ知覚に立ち返ってこの問題を検討することとする．

第 3章

ピッチ知覚から着想を得たF0推定法の提案

本章ではまずヒトのピッチ知覚について述べ，そこから得られた知見を基にした提案法を概説する．

3.1　ヒトのピッチ知覚からの知見

我々は経験的に，日常生活において常に様々な音信号のピッチを知覚している．その我々が日常生活を送っている音環境は，常に雑音や残響が存在している世界である．そのような環境の中で我々がピッチを難なく知覚できているという事実は，ヒトのピッチ知覚が雑音や残響の影響に頑健な機構となっていることを意味する．我々は，F0 推定を考える上で，ここでもう一度ヒトのピッチ知覚に立ち返って見つめ直す必要があると思われる．

そこで本節では，普段ヒトが雑音や残響を伴う環境でもピッチを知覚しており [59, 60]，それを手がかりに音信号を知覚できていることに着目し [61, 62]，ヒトのピッチ知覚のメカニズム [63] に立ち返って F0 推定法の検討を行う．

3.1.1　ピッチ知覚のメカニズム

ヒトの聴覚系は，外耳・中耳・内耳からなる聴覚器官，聴神経から大脳の聴覚皮質に至る神経経路である聴覚神経系，大脳の聴覚皮質とから成る（図 3.1）．

まず外耳の耳介により集音された音波は外耳道を通って中耳の鼓膜に到達する．音波による振動は鼓膜により 3 つの耳小骨（つち骨，きぬた骨，あぶみ骨）に伝えられ，それら耳小骨でインピーダンス変換された振動は，内耳の蝸牛に直結した前庭窓に伝えられる．前庭窓の振動は，蝸牛内のリンパ液に圧力差を生じさせ，蝸牛内の基底膜を振動させる．基底膜の振動は蝸牛内の 3,500 個の内有毛細胞に伝えられ，振動によりこの内有毛細胞から化学伝達物質が放出される．聴神経の末端器官においてその化学伝達物質を受領し，ある閾値以上になると電気的な神経インパルス（神経発火）が起こる．この電気信号は，聴覚神経系を通じて大脳の聴覚皮質に至り，聴覚野において音として知覚される [1]．

これら聴覚系においてピッチ知覚がどう処理されているかに関しては，19 世紀から様々な説が提唱されてきた [2]．以下に，代表的な学説を簡単に示す．

時間説

元々はSeeBeckが19世紀半ばに提唱した説で，聴神経発火の間隔がピッチとして知覚されるという説である．

場所説（周波数説）

古くはOhmによって提唱され，後にHelmholtzによって拡張されたもので，音信号中の基本周波数の正弦波成分そのものを知覚しているという説．基底膜上に配置された聴神経の発火位置 (場所) で説明されることから，一般的に「場所説」と称される．

差音説

ヒトはF0が欠落した音信号でもピッチを知覚できることから，近傍の周波数成分同士の周波数差分 (差音) を手がかりにピッチを知覚しているという説である．Helmholtzと Fletcherにより提唱された．

微細構造説

信号波形の微細構造に着目し，微細構造の山と山との間の時間間隔がピッチ知覚と関係しているとする説である．Shoutenらによって提唱されたもの．

この他にも様々な学説が提唱されているが，どれか一つの学説だけでヒトのピッチ知覚をすべて説明することは困難である．ヒトは，ピッチ知覚を行う際に，時間情報や場所情報，差音や微細構造を含め様々な情報を総動員して，またある時はそれらの情報を使い分けながらピッチを知覚している，と考えるのが自然である [64]．

これらヒトのピッチ知覚のメカニズムにヒントを得たF0推定法もいくつか既に存在する．例えばMeddisらのグループは，ヒトの蝸牛が帯域通過フィルタの集合としての特性を持つこと，つまり「場所」説を前提にしつつ，蝸牛内での聴神経発火計算モデルの発火頻度の確率，つまり「時間」説に基づいた自己相関関数を

導入することでヒトのピッチ知覚をモデル化している [65]．石本らも同様に，それら時間情報と周波数情報の両方を利用することで耐雑音性の向上に成功している [57]．

3.1.2 ミッシングファンダメンタルと AM 音の知覚

ヒトのピッチ知覚現象の特に注目すべき点として，ヒトは F0 が欠落した音信号でもピッチを知覚できることが挙げられ，この現象はミッシングファンダメンタル（missing fundamental）と呼ばれる．ミッシングファンダメンタルから，ヒトは AM（Amplitude modulation：振幅変調）音の変調音を知覚できることも以前から知られている．Schouten は，周波数 f の正弦波を周波数 g の正弦波で振幅変調した $f-g$, f, $f+g$ Hz の三つの周波数成分を有する AM 音を，ヒトがピッチ周波数 g として知覚することを実験により示した（Schouten の実験 [66]）．Shouten の実験を簡単に図示すると，キャリア周波数が 1,000 Hz， 振幅変調周波数が 200 Hz，変調度が 1 である AM 音は図 3.2(a) に示す波形のようになり，その周波数構造は図 3.2(b) に示すとおり 800 Hz， 1,000 Hz， 1,200 Hz の 3 本からなる．この信号には図 3.2(b) に点線で示す 200 Hz の F0 成分は存在していないが，ヒトはこの AM 信号のピッチを 200 Hz と知覚する．この 200 Hz のことを，low pitch と呼ぶ．この例では，この low pitch と，800 Hz， 1,000 Hz， 1,200 Hz の 3 本とが基本周波数とその調波の関係にあるので，3 本の調波のことをマッチング音， 基本周波数 200 Hz が知覚されることをピッチマッチングと呼ぶ [2]．

Shouten は続く実験で， 850 Hz， 1,050 Hz， 1,250 Hz の 3 本からなる AM 音を使って知覚実験を行った．差音説によれば 200 Hz が知覚されるはずであるが，実際に知覚されるピッチは 210 Hz であり，差音説を否定する結果となった．この例では差音と実際に知覚されたピッチとは 10 Hz の開き（ピッチシフト）が存在するが，この現象は微細構造説で説明がなされる．これらの現象からは，ピッチが g Hz として知覚されるためには周波数 f は周波数 g の整数 n 倍でなければならないことが分かる．ちなみにこのピッチシフトは計算による算出が可能であり，f から Δf だけ高くずらした AM 音の low pitch は，g に $\Delta f/n$ がプラスされた周波数でピッチ知覚されることになる．この場合は n が 5 で，Δf が 50 Hz であること

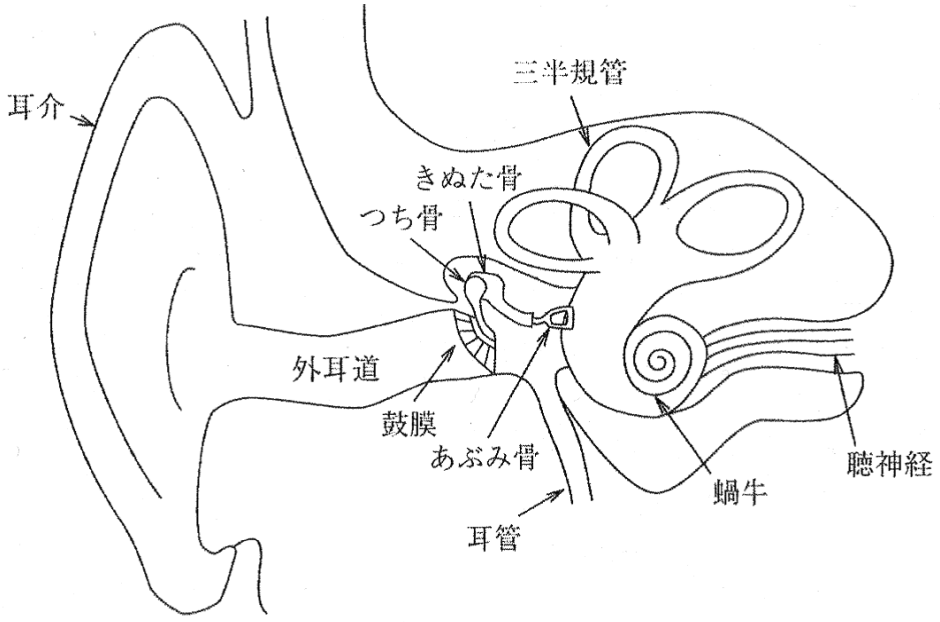

図 3.1: 聴覚末梢系の構造（大串）[1]

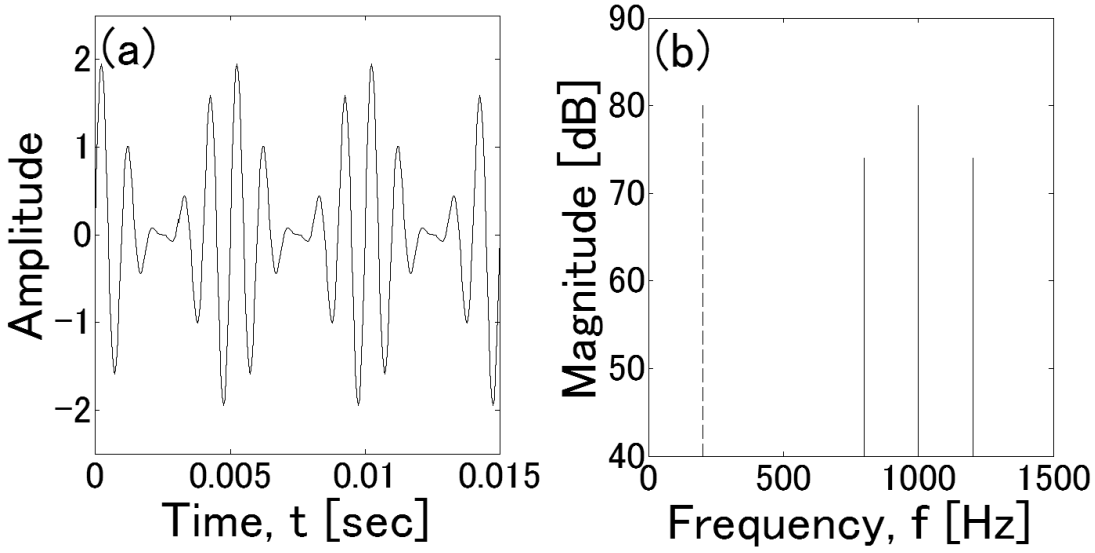

図 3.2: ピッチ知覚の事例「ミッシングファンダメンタル」: (a) AM 信号の時間波形と (b) その周波数構造

から，ピッチシフト $\Delta f/n$ は 10 Hz となり，知覚される low pitch は 210 Hz となる [2].

その後，Ritsma は AM 音のピッチ知覚に関してさらに研究を進め，low pitch が知覚される周波数領域はある程度限られていること，またその領域は変調度が低くなるほど狭くなることを明らかにしている [2].

このように，ヒトがある条件下では，AM 音の変調成分をピッチとして知覚できるという事実は，基本周波数推定法を考える上で大きなヒントとなる．

3.1.3 F0 推定へのアプローチ

ここで図 3.2 の信号にもう一度着目してみる．この信号を，無線技術で用いられる振幅変復調として眺めてみると，これは丁度 1,000 Hz の搬送波が 200 Hz のメッセージ信号で振幅変調されている信号に相当する．つまり，200 Hz の音声もしくは音楽を，1,000 Hz の電波（搬送波）に乗せている信号であるとも言い換えられる．AM 放送においては，この AM 波を受信機内の復調回路で復調処理することで，元の音声ないしは音楽である 200 Hz のメッセージ信号を取り出してスピーカー等で再生している．

前節で触れたヒトが AM 音の変調成分をピッチとして知覚している現象から考えて，AM 放送で用いられる振幅変調技術の復調処理を利用することで，ヒトのピッチ知覚を計算機上で人工的に模擬できるはずである．つまり，実環境に存在する音信号を AM 調波複合音とみなせば，雑音残響環境で観測された音信号から任意の隣り合う 3 本の調波成分 f_1, f_2, f_3 を抽出し，そこから変調信号（時間包絡線）を復調処理により取り出して，その周期を特定することで F_0 を推定することが可能であると考えられる．

音信号の変調成分に着目することは，いくつかのメリットがある．振幅変調信号を変調伝達関数（Modulation Transfer Function : MTF）の観点から考えると，雑音や残響などの外乱による影響は，全て変調度（Modulation Index）の低下という単純な図式に落とし込める [67]．ここで，変調度が低下するということは，AM 信号の変調成分の振幅が低下することを意味している．よって，時間波形の自己相関を調べたり，周波数スペクトル情報を調べたりするといった複雑な信号処理

を経ること無く，変調度の観測から外乱の影響を把握することができる．また，外乱により変調成分の振幅は小さくなるものの，変調成分の周期は保持されるうえ，必要に応じ変調度をパラメータとした波形回復機構を付加することも可能である．

これらのポイントを根拠に，ヒトの AM 音のピッチ知覚現象から着想を得て，それを計算機上で模擬するべく，振幅変調技術の復調処理を用いた F0 推定法を提案する．

3.2 振幅変調特性に着目した F0 推定法「*FreeDAM*」

ここでは，前述のヒトの AM 音のピッチ知覚にヒントを得た F0 推定法を提案する．その原理は，実環境に存在する音信号を AM 調波複合音とみなし，雑音残響環境で観測された音信号から任意の隣り合う 3 本の調波成分 f_1, f_2, f_3 を抽出し，そこから変調信号を復調処理により取り出して，その周期を特定することで音信号の F0 を推定するというものである．

3.2.1 問題設定

ここで対象とする解析的な時不変調波複合音 $x(t)$ を，次式のように表現する [68]．

$$x(t) = \sum_{k \in K} a_k \exp(j\omega_k t + j\theta_k) \tag{3.1}$$

ただし，a_k は振幅，θ_k は位相，k は調波の次数，K は調波の数である $(k = 1, 2, \cdots, K)$．ω_k は $2\pi k F_0$ であり，基本周波数 F_0 は一定である．

3.2.2 振幅変調・復調の理論

ヒトのピッチ知覚にヒントを得た提案法は，音信号を AM 音の集まり（AM 調波複合音）として捉えて音信号の基本周波数を推定するものであるので，AM 信号の復調技術の導入が不可欠である．

AM 波の復調方法は「同期検波方式」と「非同期検波方式」とに大別される．前者の「同期検波方式」とは，受信機側に設けた局部発信器において生成した信号

をアンテナから受信したAM信号と混合して，さらに不要な成分をローパスフィルタで除去することで復調を行う方式の総称である．特にその局部発信器において，搬送波と同周期かつ同相の信号を生成して混合する復調方式は「ホモダイン方式」もしくは「ダイレクトコンバージョン方式」と呼ばれており，中間周波数を用いるスーパーヘテロダイン方式に比べて実装が容易であることから，近年再び脚光を浴びている方式である [69, 70]．提案法では，いくつかある同期検波方式の中から，信号処理を前提としたソフトウェア無線の分野でも広く用いられているこの「ダイレクトコンバージョン方式」を採用することとした [69].「非同期検波方式」については，事前の評価試験において雑音に対する耐性に難があったことから，提案法への採用は見送った [71]．以下及び図 3.3 において，振幅変調における同期検波による復調過程の原理について説明する．

最初に，メッセージ信号 $m(t)$ を次式で表すこととする．

$$m(t) = \cos(\omega_m t) \tag{3.2}$$

また，振幅変調に用いる搬送波 $c(t)$ を，次式で表すこととする．ただし ω_c は搬送波の周波数である．

$$c(t) = A\cos(\omega_c t) \tag{3.3}$$

すると，振幅変調したAM信号 $x_{\mathrm{AM}}(t)$ は次式で表される．このときの周波数スペクトルは，図 3.3(a) のようになる．ただし，M は変調度である．(変調度とは，搬送波の振幅値に対する信号波の振幅値の比であり，振幅変調波の山と谷の比率に相当する.)

$$\begin{aligned} x_{\mathrm{AM}}(t) &= A\{1 + Mm(t)\}c(t) \\ &= A\{1 + M\cos(\omega_m t)\}\cos(\omega_c t) \end{aligned} \tag{3.4}$$

ここで， 変調度 M は通常は 1 以下の値を取るが，M が 1 を超えると過変調となる．上式を三角関数の公式により変形すると次式のようになる．

$$\begin{aligned} & x_{\mathrm{AM}}(t) \\ &= \frac{AM}{2}\{\cos((\omega_c - \omega_m)t) + \cos((\omega_c + \omega_m)t)\} \\ &\quad + A\cos(\omega_c t) \end{aligned} \tag{3.5}$$

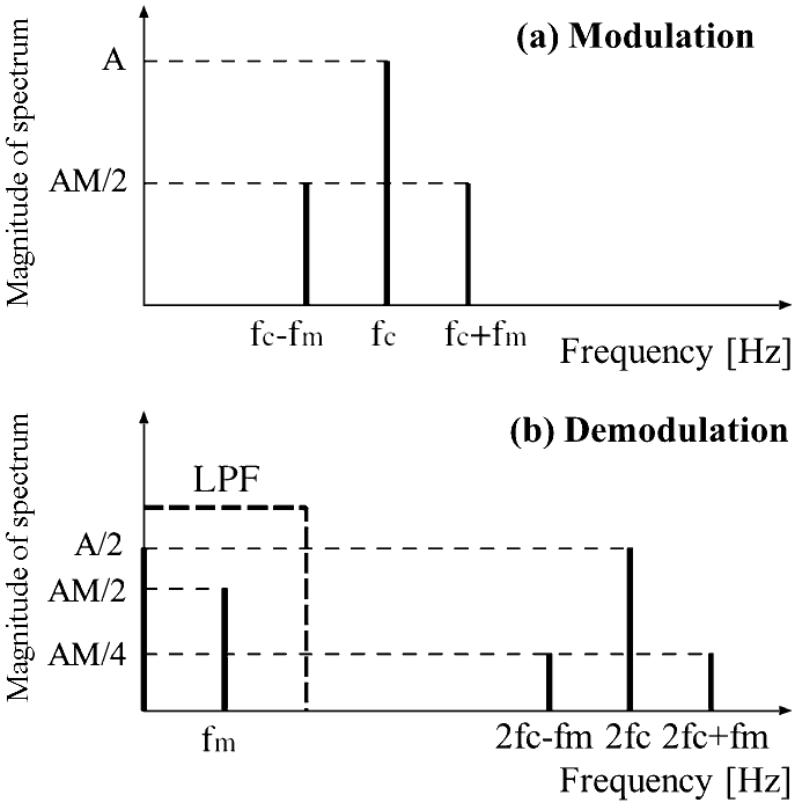

図 3.3: 振幅変調と復調過程：(a) AM 音のスペクトル，(b) 同期検波による復調時のスペクトル

次いで，$x_{\mathrm{AM}}(t)$ を同期検波方式で復調するために，局部発信器においてキャリア信号と同周期かつ同位相の信号 $\cos(\omega_c t)$ を生成し，これを AM 信号 $x_{\mathrm{AM}}(t)$ に乗ずると次式のようになる．

$$\begin{aligned} & x_{\mathrm{AM}}(t)\cos(\omega_c t) \\ = \; & \frac{AM}{4}\left\{\cos((2\omega_c-\omega_m)t)+\cos((2\omega_c+\omega_m)t)\right\} \\ & +\frac{A}{2}\cos(2\omega_c t)+\frac{AM}{2}\cos(\omega_m t)+\frac{A}{2} \end{aligned} \tag{3.6}$$

ここでローパスフィルタにて第 1 項から第 3 項までの高周波成分を除去し，次いで第 5 項の直流成分を除去すれば次式が得られる（図 3.3(b)）．

$$\frac{AM}{2}\cos(\omega_m t) \tag{3.7}$$

ここで 式 (3.7) は，式 (3.2) の振幅の大きさを変えたものに過ぎず，よって次式に示すように 式 (4.2) 式から元のメッセージ信号 $m(t)$ が復調できることとなる．

$$\begin{aligned} m(t) &= \frac{2}{AM}\left(\mathrm{LPF}[x_{\mathrm{AM}}(t)\cos(\omega_c t)]-\frac{A}{2}\right) \\ &= \cos(\omega_m t) \end{aligned} \tag{3.8}$$

提案法では，以上の原理を基本周波数の推定に応用している．

3.2.3 振幅変調・復調の F0 推定への応用

図 3.4 に，AM 信号の復調過程を利用した提案法の概略イメージを示す．まず音声信号中の隣り合う周波数として仮定する f_1, f_2, f_3 の 3 本 1 組を抽出するために，帯域通過フィルタのカットオフ周波数を適切に設定する．ここで，f_1, f_2, f_3 間の周波数間隔は f であるものとする．帯域通過フィルタを通して抽出された信号は，搬送波 f_c とその両側波 $f_c - f_m, f_c + f_m$ から成る振幅変調信号（AM 信号）とみなせることから，これを同期検波により復調して信号 f_m を取り出すことを考える．そこで，局部発信器において f_2 と等しい局部発信周波数 f_c を生成し，フィルタを通して抽出した音声信号に混合する．混合された信号には高調波が含まれているため，低域通過フィルタで高調波成分を除去し，併せて直流成分を除去する．このようにして得られた復調信号の波形周期が最初に仮定した f の周期と等しけ

れば，基本周波数 F_0 は f であると推定する．以上のプロセスをあらかじめ想定する基本周波数の範囲内において繰り返し，判定基準に合致するものを基本周波数の値として出力することになる．

3.2.4 提案法の特長

耐雑音性

以上のように，提案法は前章で示した分類に従えば，音声の調波性に着目した手法であると言える．音声の調波構造が雑音に対して頑健であることは知られており [56, 57]，提案法は雑音に対してロバストな手法であると考えられる．

耐残響性

次いで，提案法の耐残響性について再度変調伝達関数（Modulation Transfer Function：MTF）の観点から考える．まず，変調伝達関数は，音声明瞭度を評価する上での重要な指標である．音場内において音声波形の時間的な包絡線情報が残響や雑音によってどう変形するかを，100%振幅変調した正弦波を利用してその変調度の減衰量からある程度予測することが可能である [72]．図 3.5 は，残響の影響と変調度との関係性の概要を示したものである．(a) では，入力信号 $x(t)$ が，残響の影響を受けた結果，$y(t)$ のような波形として出力される様子を示している．両者の波形を比較すると，残響の影響により信号の変調成分の振幅自体は小さくなっており，入力信号 $x(t)$ を AM 信号とみなせば，即ち変調度は減少しているが，包絡線を形成する変調信号の周波数 f_m の周期はそのまま保持されていることが見て取れる．このことは，仮に f_m が基本周波数であるとすれば，たとえ音声信号が残響の影響を受けても f_m は十分復調可能であると考えられ，つまり残響環境下でも提案法を適用した基本周波数推定は十分可能であろうと考えられる．

残響環境にける変調度の減衰量を示す変調伝達関数は，理論上は次式で表される [72]．

$$m(f_m) = \frac{1}{\sqrt{1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2}} \tag{3.9}$$

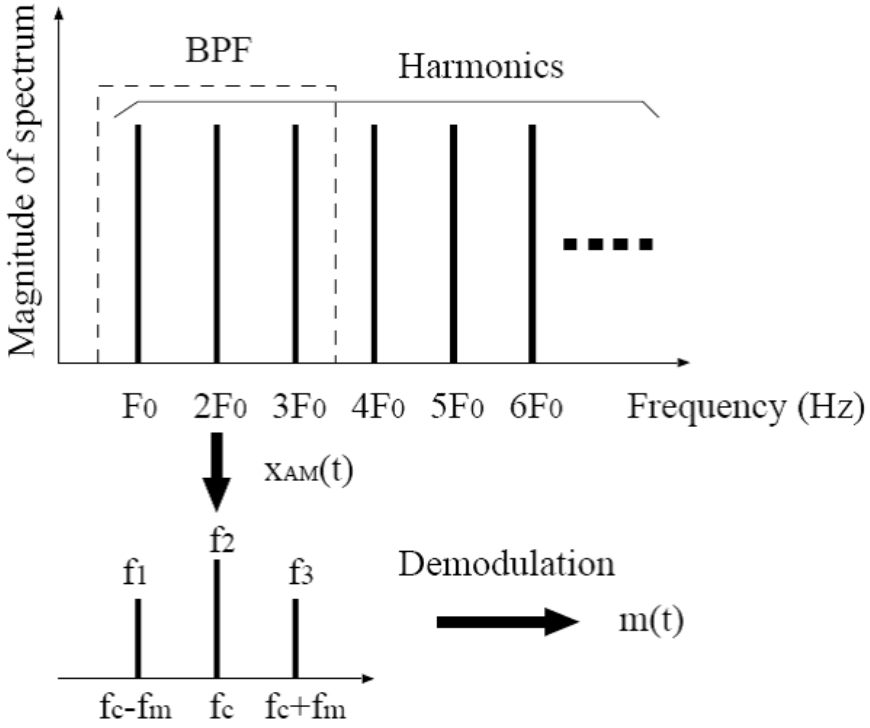

図 3.4: AM 音のピッチ知覚にヒントを得た F0 推定の概要

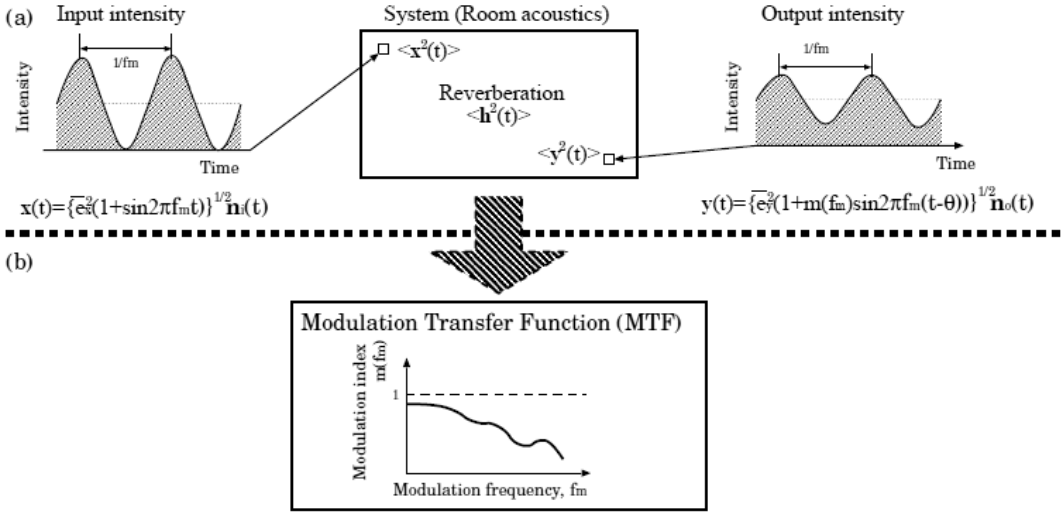

図 3.5: 残響の影響と変調度との関係性 (a) 残響の影響と入出力信号波形 (b) 変調度の推移（鵜木） [72]

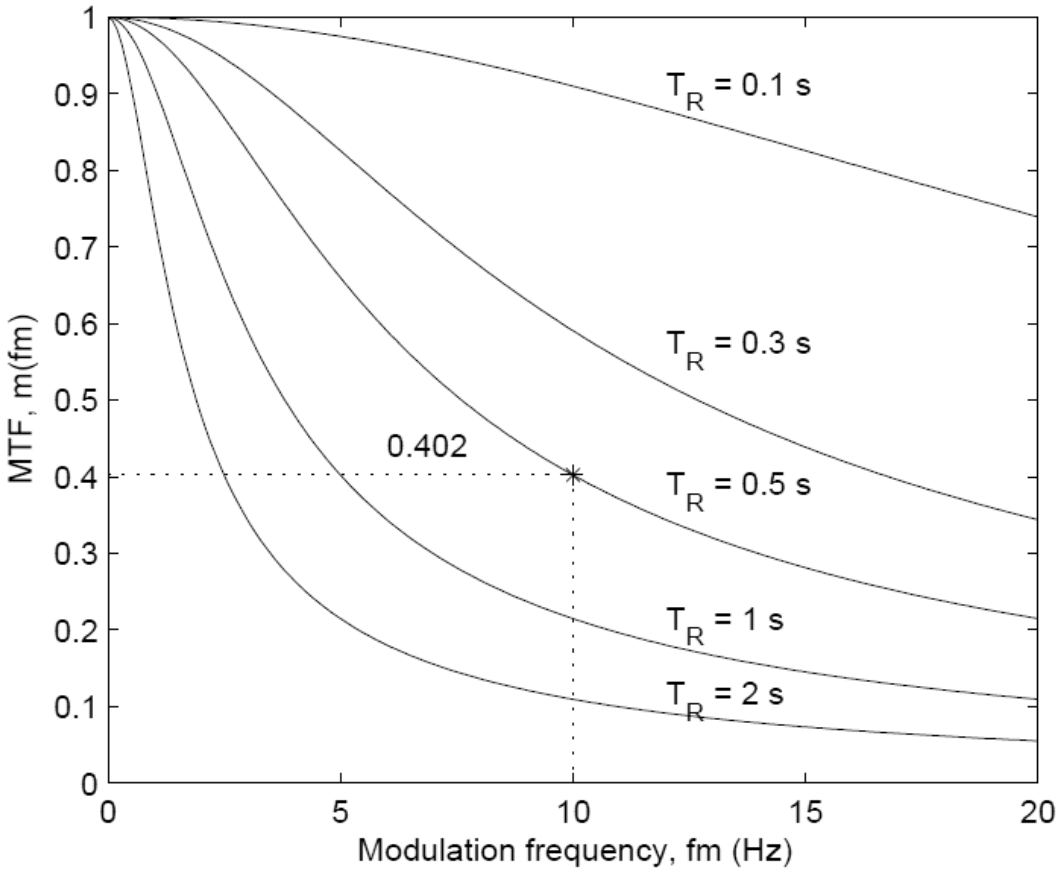

図 3.6: 変調周波数ならびに残響時間と変調度との関係性（鵜木）[72]

式 3.9 によれば，変調度 m は残響時間 T_R もしくは変調周波数 f_m が増加すると減衰していくことが分かる．これを図示すると図 3.6 のようになり，基本的に変調周波数 f_m が高いと変調度 m は 0 に漸近していくが，残響時間 T_R が大きいほどその漸近が顕著になることが見て取れる．このことは，提案法においても，残響環境において基本周波数が高い場合や，残響時間が大きい場合には，基本周波数の推定精度に影響が及ぶことをあらかじめ念頭においておく必要がある．

3.2.5 提案法のアルゴリズム

図 3.7 に，提案法の処理フローを示す．処理フローは，(1) フレーム処理により切り出した信号から，帯域通過フィルタで隣り合う 3 本の調波を抽出し，(2) 同期検波を用いた復調過程により主要な変調信号の波形を取り出し，(3) 取り出した波形の周期を求めて F0 の推定を行う，という以下の流れで処理される．

(1) 帯域通過フィルタによる 3 本の調波の抽出

入力信号 $y(t)$ からフレーム処理により分析区間を切り出し，帯域通過フィルタにて AM 信号 $x_{\mathrm{AM}}(t)$ に見立てた隣り合う 3 本 1 組の調波 (f_1，f_2，f_3) を抽出する．通過帯域幅の下限ならびに上限は，それぞれ，$f_c - f_m(= f_1)$，$f_c + f_m(= f_3)$ に設定される．

(2) 同期検波による復調

局部発信器にて，キャリア周波数（この場合では $f_c(= f_2)$）と同周期で同相の信号 $\cos(\omega_c t)$ を生成して，抽出された 3 本 1 組の調波信号 $x_{\mathrm{AM}}(t)$ と乗算する．ここで，同相の生成信号とするために，あらかじめ $x_{\mathrm{AM}}(t)$ との相互相関係数を求めることで，位相差を補償する処理を行う．その後で，低域通過フィルタで高調波成分を除去することで，復調されたメッセージ信号 $m(t)$ を得る．

(3) 基本周波数 F_0 の決定

得られた復調信号 $m(t)$ の主要な周期ないし周波数を，自己相関関数や高速フーリエ変換等を用いて求める．これにより得られた主要な周波数 $\hat{F_0}$ が，(1) で抽出した 3 本の調波間の間隔，即ち $f_2 - f_1$（または $f_3 - f_2$）と全く等しければ，この周波数 $\hat{F_0}$ は基本周波数 F_0 と決定され，等しくなければ $\hat{F_0}$ は F_0 の候補から外される．

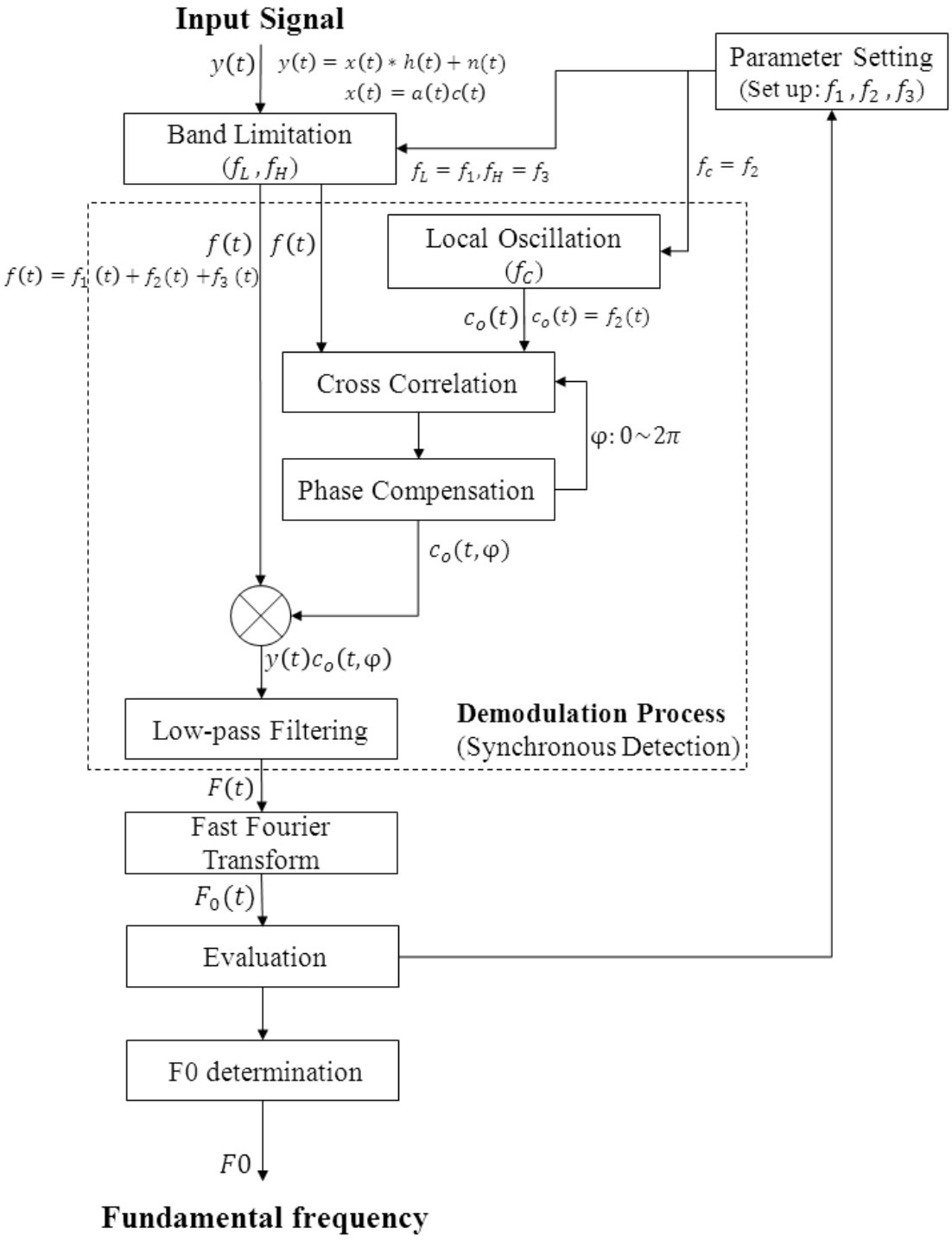

図 3.7: 提案法の処理フロー

以上の三つの処理 (1)(2)(3) は，推定の対象とする全周波数範囲にわたって順に繰り返される．

このように提案法は，振幅変調技術の復調過程を利用して F0 を推定するものであり，以後，*FreeDAM* (Fundamental fRequency Estimation mEthod using Demodulation of Amplitude Modulation) という名称も併用する．

3.2.6 各機構の詳細

ここでは，提案法で用いている特記すべき処理について触れる．

局部発信器と位相補償機構

提案法に用いている同期検波方式（ホモダイン方式）は，AM 信号の搬送波と同周期かつ同相の信号を局部発信器で生成する必要があるが，搬送波に相当する ω_c の周期も位相も未知である．そのうち，周期（周波数）については全レンジにわたってサーチを行うことで対応できるが，位相差については，観測信号の位相を調べた上で局部発信信号の位相を観測信号のそれと等しくする処理が必要である．次式は AM 信号の同期検波を表現しているものであるが，入力信号中に未知の位相差 θ が存在したとすれば，局部発信器の信号の位相差 ϕ も θ と同じ値に設定する必要がある．

$$\cos(\omega_c t + \phi)\{A\{1 + M\cos(\omega_m t)\}\cos(\omega_c t + \theta)\} \tag{3.10}$$

提案法では，位相差 ϕ を求める方法として，入力信号と局部発信器の信号との相互相関の大小に着目する．具体的には，同期検波の前段として，位相差 ϕ の値を 0 から 2π まで順に変化させながら両信号間の相互相関係数を求めて，係数の値が最大になる ϕ の値をあらかじめ見つけておき，同期検波プロセスでは求めた ϕ の値をあらためて代入した上で局部発信信号を生成して復調を実施している．

基本周波数の決定プロセス

処理フローのところで触れたように，提案法では入力信号に対して順次帯域通過フィルターのカットオフ周波数を変更させつつ，かつそれに対応した局部発信

信号も生成しながら，あらかじめ想定する全周波数レンジにわたって F_0 をサーチする処理を行っている．これは入力信号のある一区間の基本周波数を求めるのに，何百通りもの計算が行われることを意味しており，何百という候補の中から一番確からしい値を推定された基本周波数として返すことになる．

ここで用いている判定方法は，最初に 3 本の調波を抽出すために設定した帯域通過フィルタのカットオフ周波数で決定される 3 本の調波間の間隔，即ち $f_2 - f_1$(または $f_3 - f_2$) が，復調の結果出力される信号の周期と全く等しければ，$f_2 - f_1$(または $f_3 - f_2$) がこの場合の基本周波数 F_0 と決定される．決定された 3 本の調波の間隔に相当する周波数と，復調信号の基本周期とが全く一致する候補が無い場合には，もっとも値の近い候補が F_0 として選択されるようになっている．

3.2.7 提案法による F0 推定の例

図 3.8 に，FreeDAM による F0 推定の実例を示す．まずここでは図 3.8(a) のような信号が観測されたと仮定し，その周波数構造は図 3.8(b) のように F_0 が 100 Hz でかつ 10 次の調波から成る時間長 1.0 s の信号であるとする．次に，その F_0 の予測値を 100 Hz と仮定した上で，(a) の信号から帯域通過フィルタにより任意の隣り合う 3 本の調波（ここでは第 5，6，7 次である 500 Hz，600 Hz，700 Hz）を抽出するとその周波数構造は図 3.8(d) のようになり，その時間信号波形は図 3.8(c) のようになる．この信号を検波処理により復調すると図 3.8(e) のような信号が得られ，この信号の周期を特定すると図 3.8(f) に示すように 100 Hz となり，最初に仮定した予測値と等しくなることから，100 Hz が F_0 の推定値となる．この値は観測信号の F_0 と一致しており，正しく F0 が推定できていることが確認できる．

一方，同じ入力信号にて F_0 の予測値を 150 Hz として推定したときの結果を図 3.9 に示す．帯域通過フィルタは 600 Hz を中心として 150 Hz 間隔で隣り合う計 3 本の調波（450 Hz，600 Hz，750 Hz）を抽出する設定となっているが，図 3.9(b) に示すように 500，600，700 Hz が取り出され，その結果同期検波による復調信号波形の周期は図 3.9(c) 及び図 3.9(d) に示すように 100 Hz となり，予測値の 150 Hz と一致しないことから，150 Hz の仮定は棄却される．

以上のように，FreeDAM は任意の隣り合う 3 本の調波を取り出して F0 推定を

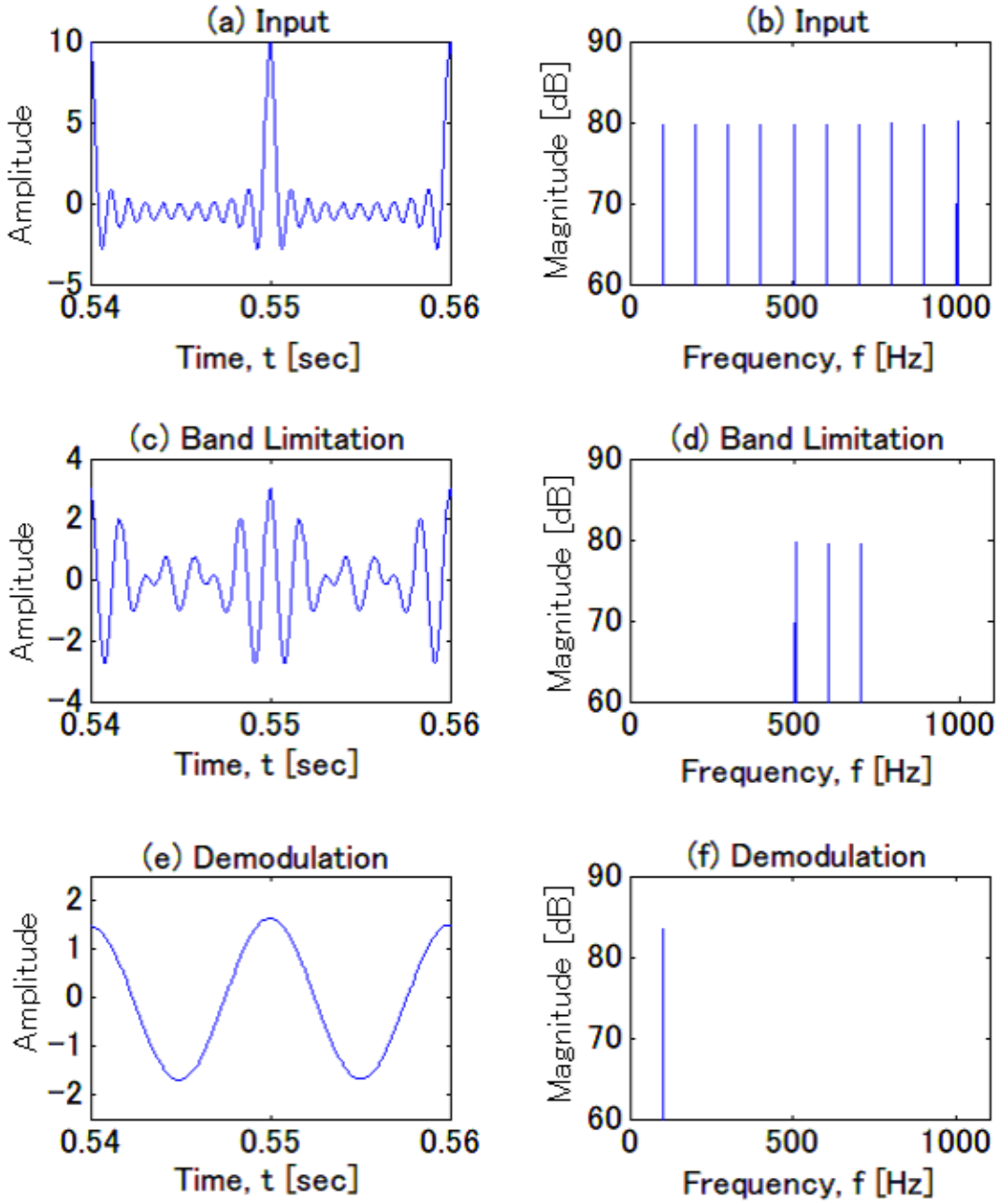

図 3.8: FreeDAM による F0 推定の例：(a) 観測信号波形，(b) 観測信号成分，(c) 抽出信号波形，(d) 抽出信号成分，(e) 復調信号波形，(f) 復調信号成分

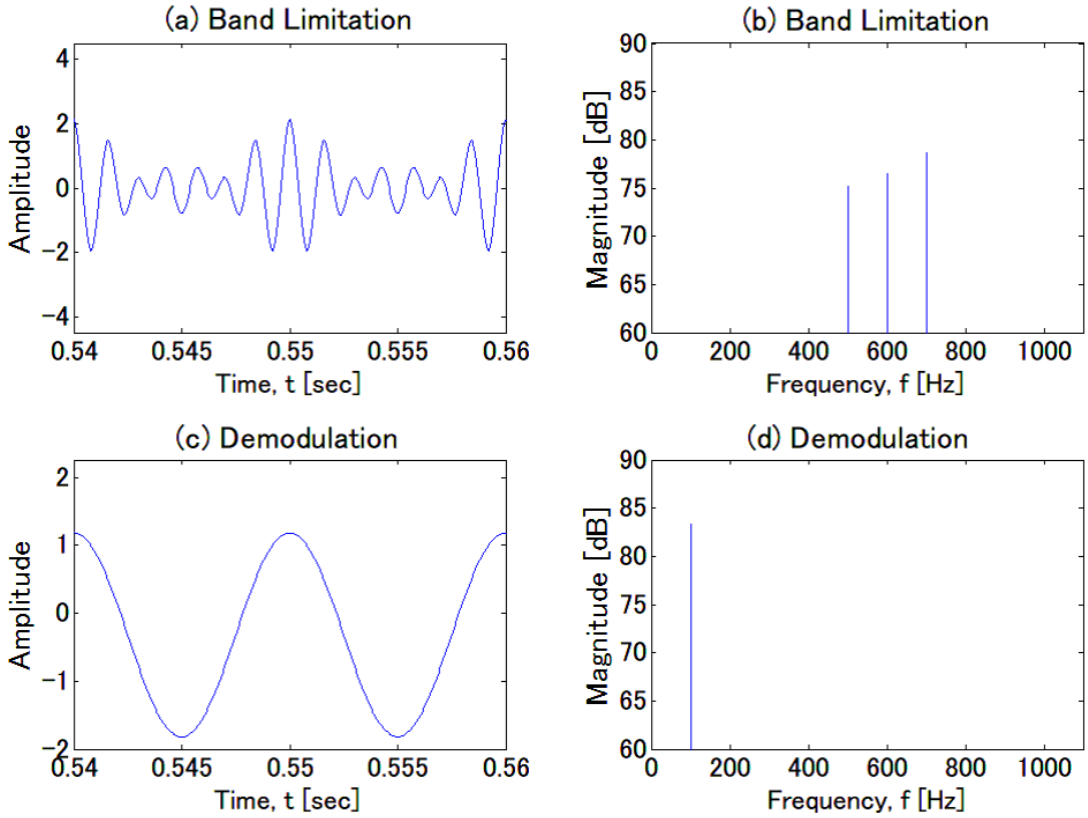

図 3.9: FreeDAM による F0 推定 (棄却) の例：(a) 抽出信号波形，(b) 抽出信号成分，(c) 復調信号波形，(d) 復調信号成分

行うことができ，また，誤った推定結果の棄却も行える．FreeDAM ではこれらを巧みに利用して，帯域通過フィルタの通過周波数を順次変化させつつ，復調が確実に可能なポイントを探索しながら，確実な F0 推定を実現している．

3.3 基本原理の確認

3.3.1 推定精度の確認

入力信号 $x(t)$ として 10 次の調波複合音を仮定し，各倍音のレベルは基音も含めて全て同一とした． 入力信号中の F_0 は，60〜600 Hz の範囲で 5 Hz 間隔に変化させた計 108 種類とし，信号長は 1.0 sec で F0 は変化しないものとした．ここで，FreeDAM の分析窓長は，信号長と同一の 1.0 sec としている．評価尺度としては，真値に対する推定値の許容誤差を 5 % とした正答率（%）を用いた．

10 次の調波成分のうち，第 1 次，第 2 次，第 3 次の調波を抽出する設定として FreeDAM を動作させた時の検証結果を図 3.10(a) に示す．横軸の F0 の真値に対して縦軸の F0 の出力値は常に対応が取れており，その正答率は 100 %となっている．また，一般によく見られるような真値の 2 倍もしくは 1/2 倍を返す誤推定も観測されていない．次に，第 4 次，第 5 次，第 6 次の調波を抽出する設定として FreeDAM を動作させた時の検証結果を図 3.10(b) に示す．こちらも正答率は 100 %で，横軸の真値と縦軸の推定値との対応がすべて取れていることが確認できる．また，ここでは割愛するが，他の任意の 3 本の部分調波を利用した場合でも，同様に良好な結果が得られている．

以上から，FreeDAM の原理が，抽出する調波成分の位置に関わりなく，対象とする 60〜600 Hz の全範囲に渡って問題なく動作することが明らかとなった．

3.3.2 雑音残響に対する基礎的耐性の確認

雑音・残響に対する基礎的な耐性について，信号の F0 が長時間にわたって一定であるという条件のもとでシミュレーションによる評価を行った．検証内容としては，F_0 が 60〜600 Hz の範囲をとる人工的な調波複合音に対して人工的な雑音や残響を付加した上で，信号に対する提案法の F0 推定精度を調べる形で実施した．

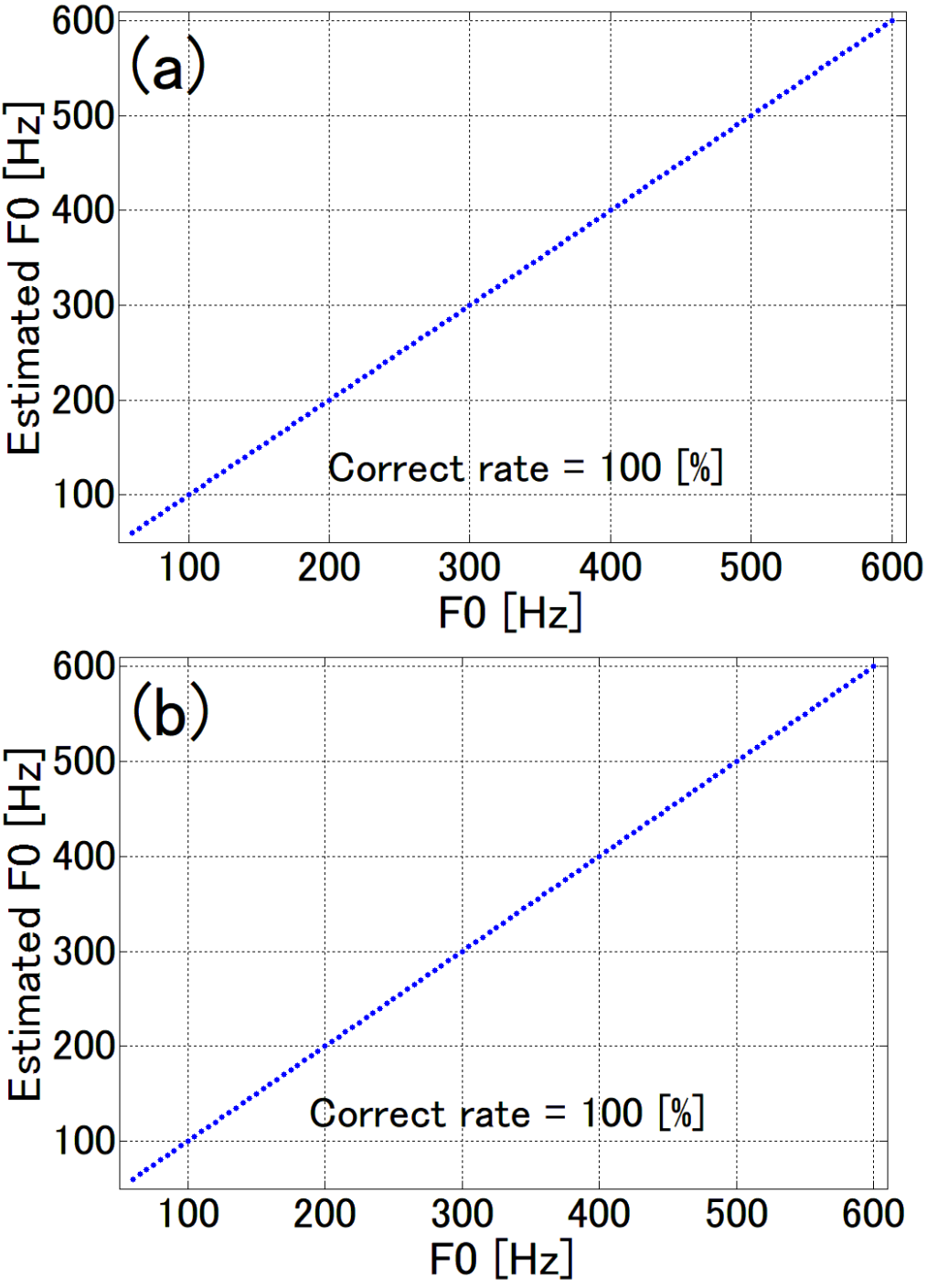

図 3.10: FreeDAM の動作検証：(a) 1 次～3 次調波を利用した場合, (b) 4 次～6 次調波を利用した場合

比較対象は，代表的な従来法である TEMPO [53]，YIN [27]，SWIPE'[43, 44] のほか，耐雑音性を持つ PHIA [57]，耐残響性を持つ複素ケプストラム法（CmpCep）[51] とした．基本的に従来法のパラメータ設定については，その性能が最も発揮できたデフォルト値を利用した．ただし PHIA と SWIPE' のフレーム長とシフト長については，FreeDAM と同じ 1.0 sec とした．信号の F0 は，60〜600 [Hz] の範囲において 5 [Hz] 間隔で設定し，合計 108 種類の信号を用いた．

背景雑音は白色雑音とし，SNR を，20，10，0，−10 dB の 4 種類とした．

残響環境についても人工的なものとし，入力信号に対し次式に示す統計的室内インパルス応答（Schroeder のインパルス応答）[58] を畳み込むことによって実現した．

$$h(t) = a \exp\left(\frac{-6.9t}{T_R}\right) n(t) \tag{3.11}$$

ただし，$n(t)$ は白色雑音であり，定数 a は次式の値である．

$$a = \sqrt{\frac{1}{\int_0^\infty \exp\left(\frac{-13.8t}{T_R}\right)}} \tag{3.12}$$

残響時間 T_R については 0.1，0.3，0.5，1.0，2.0 sec の 5 種類とした．

評価尺度としては，次式に示す正答率 [%] を用いた．

$$\text{Correct rate} = \frac{N_{F_{0,\text{Est}}}(E)}{N_{F_{0,\text{Ref}}}} \times 100 \tag{3.13}$$

ここで，分母の $N_{F_{0,\text{Ref}}}$ は，既に定義した F0 が含まれる入力信号の個数であり，今回のケースではこの値は 108 である．それに対する分子の $N_{F_{0,\text{Est}}}(E)$ は，許容誤差を E [%] とした場合に正しく F0 を推定できたものの個数である．今回は許容誤差 E を 5 [%] として評価を行った．

図 3.11 は，雑音残響信号に対する F0 推定精度を上記の正答率 %で示したものである．残響の影響が無い場合，TEMPO と複素ケプストラム法は SNR が低下すると正答率は著しく低下するが（図 3.11(a)，図 3.11(c)），FreeDAM では全て 75 %以上の高い精度が維持されている (図 3.11(f))．逆に，雑音の影響が無い場合で残響時間 T_R を増加させた場合には，TEMPO，PHIA，YIN，SWIPE' では正答率が徐々に低下していくのに対して（図 3.11(a)，図 3.11(b)，図 3.11(d)，図 3.11(e)），FreeDAM では 95 %以上の精度が維持できている（図 3.11(f)）．さらに，雑音・残

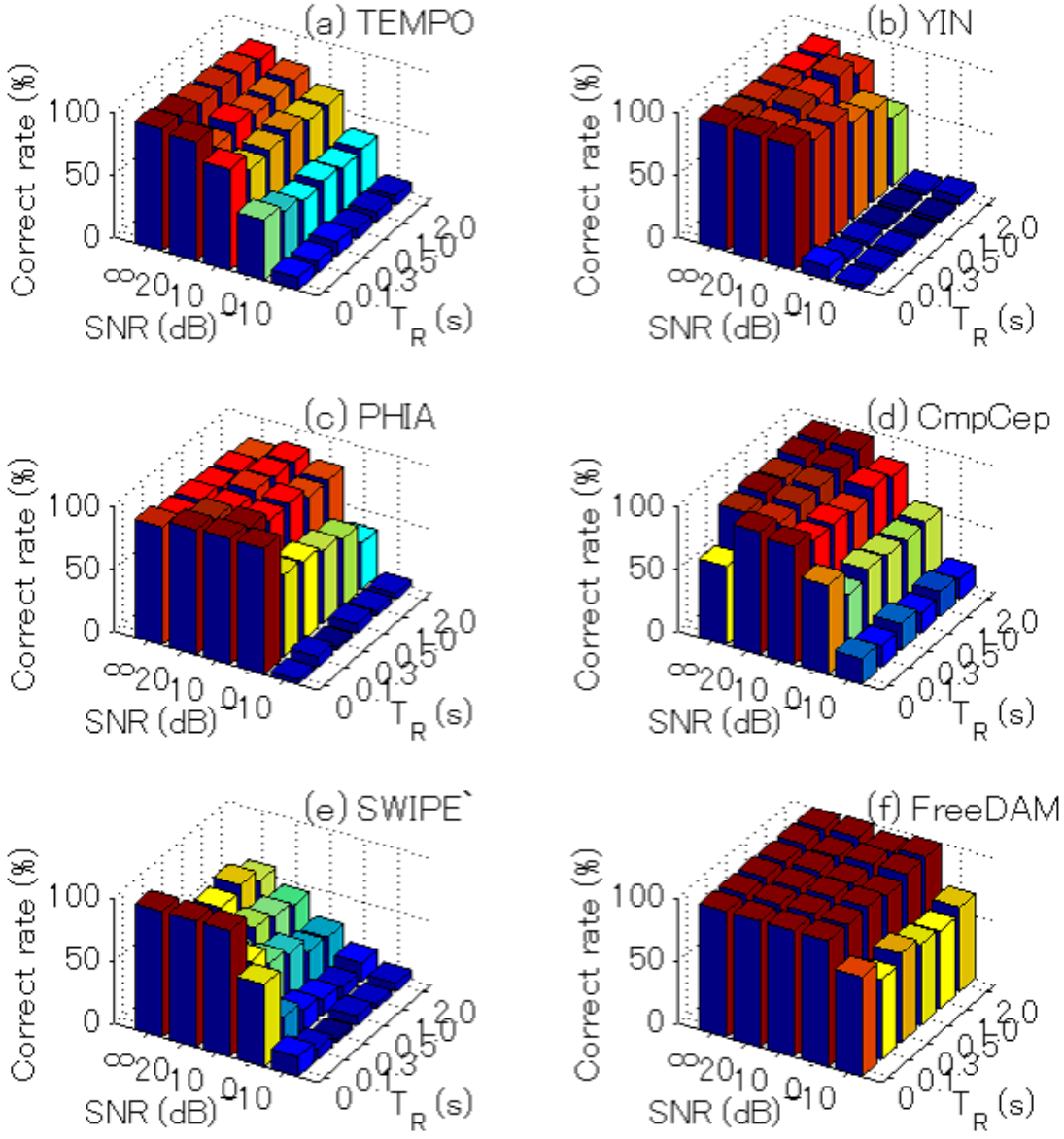

図 3.11: 雑音残響環境における F0 推定正答率 %：(a) TEMPO, (b) YIN, (c) PHIA, (d) 複素ケプストラム法，(e) SWIPE’，(f) FreeDAM（Proposed）

響が混在した場合，SNR が 0 dB でかつ残響時間 T_R が 2.0 sec の条件では，従来法はほとんど正しく推定できていないのに対し（図 3.11(a)-図 3.11(e)），FreeDAM は 90 %以上の精度で推定できている（図 3.11(f)）．

以上から，F0 が一定な時不変の長時間信号という限定された条件において，提案法が従来法と比較して雑音残響に耐性を持つことが明らかとなった．

3.4 課題整理

まず，提案法の原理が，抽出する調波成分の位置に関わりなく，対象とする基本周波数の全範囲に渡って問題なく動作することが明らかとなった．

また，雑音残響に対する基礎的耐性の考察の結果から，耐雑音性，耐残響性に加え，雑音残響に対する耐性を備えていることが明らかとなった．

一方で，提案法を最終的に実環境・実音声に適応していくことを考えると，現段階において提案法は次に示すいくつかの課題を抱えており，これらを本研究で解決しておく必要がある．

3.4.1 時変信号への対応

現状では，提案法は，時不変で長時間の調波複合音にのみ対応する実装となっている．これを時間と共に変動する信号に対応できるように，提案法の実装を拡張する必要がある，

3.4.2 評価指標等から得られる情報の統合

現在の実装において，F0 の決定処理において，復調信号成分から得られる F0 候補決定のための評価指標の取り扱いに関しては，経験に頼りながら調整を行っていた．経験に頼るやり方では限界が見えていることから，複数以上に独立した評価指標から得られる情報を統合するための手法として，「Dempster と Shafer の結合規則」[76] を提案法に導入することとする．この結合規則は，相互に関連が無い別個独立な事象の確率同士を統合するのに有効な手段である．

3.4.3　外乱が AM 信号に及ぼす影響

提案法においては，フィルタで隣合う 3 本の調波を抽出することによって，観測信号から AM 信号を取り出している．雑音環境においては，周波数軸上で考えると，雑音によるスペクトルが信号の調波構造以外の部分に出現する．残響環境においても同様に，過去の信号の周波数成分が，現在時の信号の調波構造以外の部分に出現する．これらの影響により，フィルタで抽出された信号は完全な AM 波形から崩れて，いろいろな周期の信号が入り込んだいびつな波形となる．崩れた AM 波形を使って復調を行うならば，時間包絡線を巧く抽出することができず，F0 決定プロセスの際の周期の特定が巧く行えなくなり，F0 推定精度に悪影響を与えることになる．そのため，常に完全に近い AM 波形を復調に使えるようにするための工夫が必要である．

3.4.4　外乱が復調信号に及ぼす影響

提案法においては，振幅変調の復調技術を使って，AM 信号の時間包絡線 (復調信号) を抽出している．雑音や残響の影響を受けると，取り出された AM 信号の時間波形の振幅は小さくなり，それを復調した信号（時間包絡線）の周期の特定がやはり難しくなる．そのため，復調信号の振幅を，より本来の大きさに近い値に回復するための工夫が必要である．

3.4.5　外乱やフォルマントが音声の調波構造に及ぼす影響

観測された音声中の母音の種類により，特定の複数箇所以上の周波数が声道による共鳴により強調される (フォルマント周波数)．また，雑音や残響は音声の比較的低域の部分に特に影響を与える [73]．一方で高域になるほど音声の調波成分の振幅が小さくなるとともに，調波構造自体も不明瞭となり，やはり雑音の影響を受ける場合がある [74]．そのため，音声の周波数構造を考えたとき，その AM 成分を抽出する箇所も含め，より適切な情報を選択する必要がある．

現状の実装では，3 本の隣り合う調波から成る AM 成分の抽出処理を，最も低域側の F0，2F0，3F0 の組だけを対象としているため，実環境や実音声に適用し

た場合に，雑音や残響はもちろん，フォルマント周波数の影響を強く受けてしまうことが考えられる．そこで，低域だけに限らず，中域から高域の調波まで含め，複数通り以上の 3 本の調波の組を扱えるように，提案法を拡張する必要がある．

上記で述べた課題の解決も含め，実環境・実音声に提案法を適応していくための検討を行うのが本研究のテーマであり，次章以降でさらに掘り下げて検討する．

第 4章

提案法の拡張

本章では，提案法を実環境・実音声に適応していくために，解決すべき課題をいくつか検討する．

4.1 時変信号への対応

時不変信号のみに対応していた提案法を，将来的な実音声への適用を見据え，時変信号へ対応できるように拡張を行う．具体的には，時間と共に信号の分析範囲をタイムシフトできる一定の窓長の時間窓を実装した．窓長は任意の長さに設定変更が可能であり，窓関数は，サイドローブが最小限に抑えられ，かつダイナミックレンジも広いブラックマンーハリス窓を採用した [75]．また，分析窓のシフトタイムも，任意の時間に設定が可能な実装とした．

4.1.1 問題設定

ここで対象とする解析的な時変調波複合音 $x(t)$ を，次式のように表現する [68]．

$$x(t) = \sum_{k \in K} a_k(t) \exp(j\omega_k(t)t + j\theta_k(t)) \tag{4.1}$$

ただし，$a_k(t)$ は瞬時振幅，$\theta_k(t)$ は瞬時位相，k は調波の次数，K は調波の数である $(k = 1, 2, \cdots, K)$． $\omega_k(t)$ は $2\pi k F_0(t)$ であるので，基本周波数 $F_0(t)$ は瞬時周波数となり，$x(t)$ から抽出される瞬時値である．ここでは，基本周波数 $F_0(t)$ をある時間間隔で一定な $\overline{F_0(t)}$ と仮定し，一定時間間隔で変動するものとする．

4.2 復調波形の評価指標と情報の統合

提案法は，波形の AM 成分を復調処理して取り出した時間包絡線の周期を特定することで，基本周波数を推定するものである．特に F0 決定処理の中では，様々に出力される復調波形の中から，もっとも基本周波数の特徴が出ているものを選択する必要がある．この処理をより正確に行うために，復調信号波形の評価指標を追加して，決定処理のさらなる精緻化をねらう．

4.2.1 設定周波数との一致率による指標の拡張

提案法の F0 決定処理においては，予め設定した F0 候補値と復調信号の基本周期の逆数との一致率を評価して F0 を決定している．復調信号の基本周期の特定には，自己相関関数，もしくは FFT を用いているが，そのそれぞれの特性をここで把握しておく必要がある．

図 4.1 は横軸は周波数の候補値（Hz) で，縦軸はその候補値と自己相関関数により特定された基本周波数との差分（割合）である．この場合の F_0 は 200 Hz であるが，静音環境の図 4.1(a) では確かに 200 Hz で差分が極大値を迎えており，評価指標としては有効であることが分かる．ただし，雑音環境（0 dB，図 4.1(b)）や，残響環境（1.0 sec，図 4.1(c)）においては，F_0 の正解値 200 Hz の特定が不明瞭になっていることが確認できる．

次に，FFT による基本周期特定にかかる特性を観測する．図 4.2 は横軸は周波数の候補値（Hz) で，縦軸はその候補値と FFT の出力の最大スペクトル周波数との差分（割合）である．この場合の F_0 は 200 Hz であるが，静音環境の図 4.2(a) では確かに 200 Hz で差分が極大値を迎えており，評価指標としては有効であることが分かる．雑音環境（0 dB，図 4.2(b)），残響環境（1.0 sec，図 4.2(c)）においても，200 Hz の位置には極大値が必ず出現することが確認できる．一方で，図 4.2(a)〜(c) を通じて，正解値以外の周波数にも極大値が出現しており，これらは偽候補であるため，FFT 出力の情報だけで F0 を特定することは難しいと考えられる．

以上から，自己相関による指標は外乱に弱く，FFT による指標は外乱には比較的強いが偽候補が出現しやすいという特徴が判明した．従って，自己相関関数による指標と FFT による指標と併用するのがベターであると考えられる．

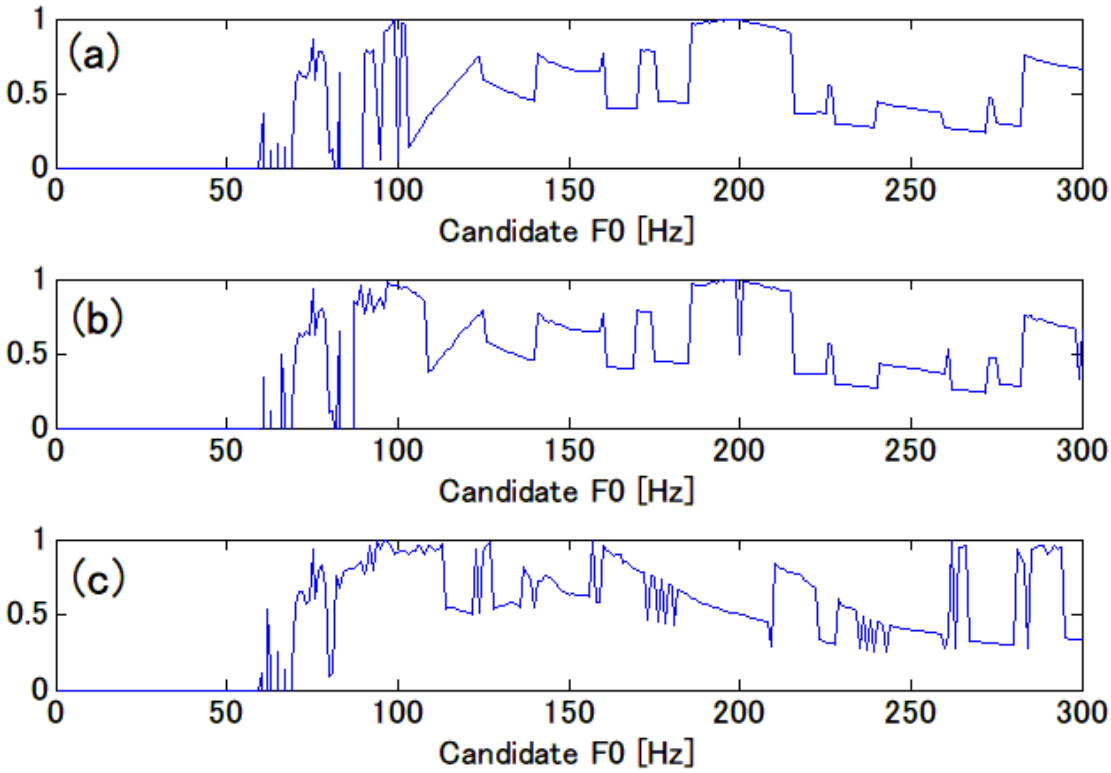

図 4.1: 自己相関関数を用いた F0 候補の特定：(a) 静音環境，(b) 雑音環境，(c) 残響環境

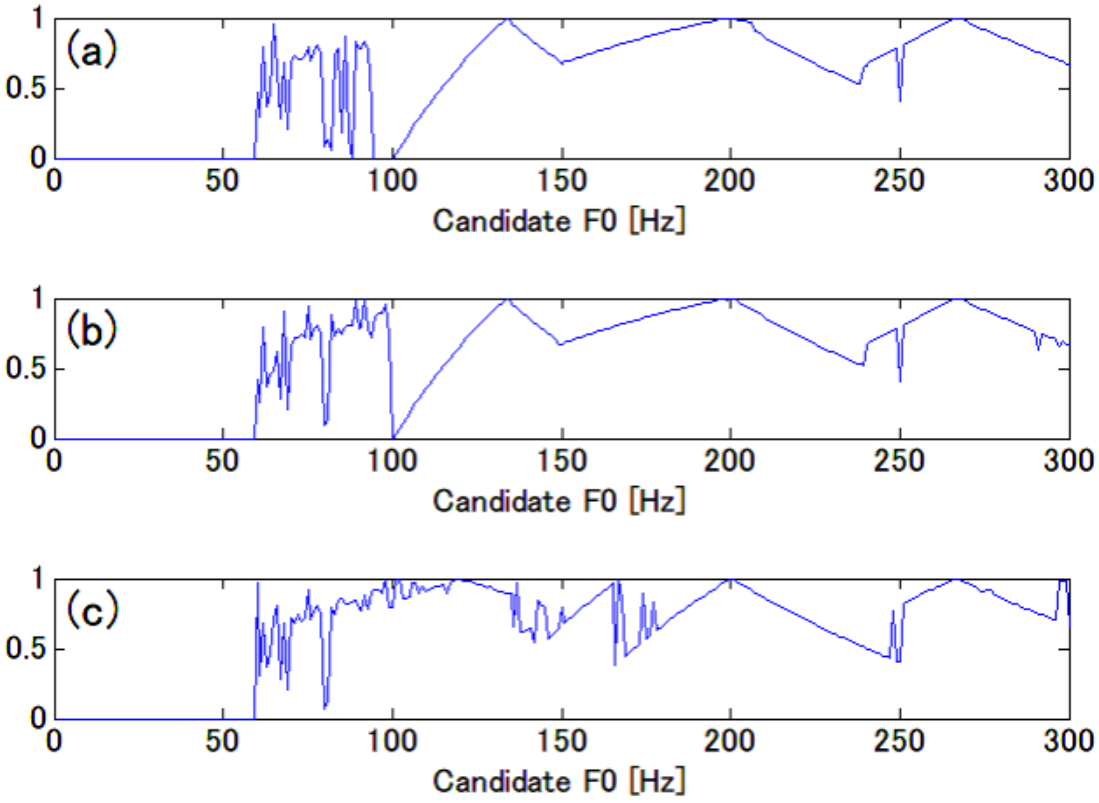

図 4.2: FFT を用いた F0 候補の特定：(a) 静音環境，(b) 雑音環境，(c) 残響環境

4.2.2 直流成分による指標

復調の理論によれば，同期検波により復調が正しく行われた際には直流成分が出現する．

$$\begin{aligned} & x_{\mathrm{AM}}(t)\cos(\omega_c t) \\ = \; & \frac{AM}{4}\left\{\cos((2\omega_c-\omega_m)t)+\cos((2\omega_c+\omega_m)t)\right\} \\ & +\frac{A}{2}\cos(2\omega_c t)+\frac{AM}{2}\cos(\omega_m t)+\frac{A}{2} \end{aligned} \tag{4.2}$$

上式は同期検波による復調の理論式（再掲）であるが，うち最終項の $A/2$ が直流成分に相当するが，通常の復調プロセスでは直流成分はカットして捨てている．提案法では，復調信号の時間波形の平均値，つまり直流成分を評価指標として積極的に利用することを考える．

図 4.3 は横軸は周波数の候補値（Hz) で，縦軸は直流成分の振幅値である．この場合の F_0 は 200 Hz であるが，静音環境（図 4.3(a)）および雑音環境（0 dB，図 4.3(b)）においては，確かに 200 Hz で直流成分の極大値が出現しており，評価指標としては先ずは有効であることが分かる．ただし，正解値以外の箇所にも一定間隔で極大値が出現しており，これらは偽候補である．また，残響環境（1.0 sec，図 4.3(c)）においては正解の 200 Hz を特定することは難しい．よって，この指標は雑音環境における正解値の特定に有用な指標として利用することにする．

4.2.3 復調信号波形の形状による指標

同期検波による復調プロセスが復調に成功した場合は，復調信号波形の形状としては原理的には次式（再掲）に示すとおり純粋な正弦波となる．

$$\begin{aligned} m(t) &= \frac{2}{AM}\left(\mathrm{LPF}[x_{\mathrm{AM}}(t)\cos(\omega_c t)]-\frac{A}{2}\right) \\ &= \cos(\omega_m t) \end{aligned} \tag{4.3}$$

つまり，復調に成功したかどうかを評価するには，復調信号波形の形状が正弦波に近いかどうかを観測すれば評価が可能である．具体的には，復調信号と正弦波との相互相関係数を求めることで，復調信号が正弦波に近いかどうかの評価が行える．

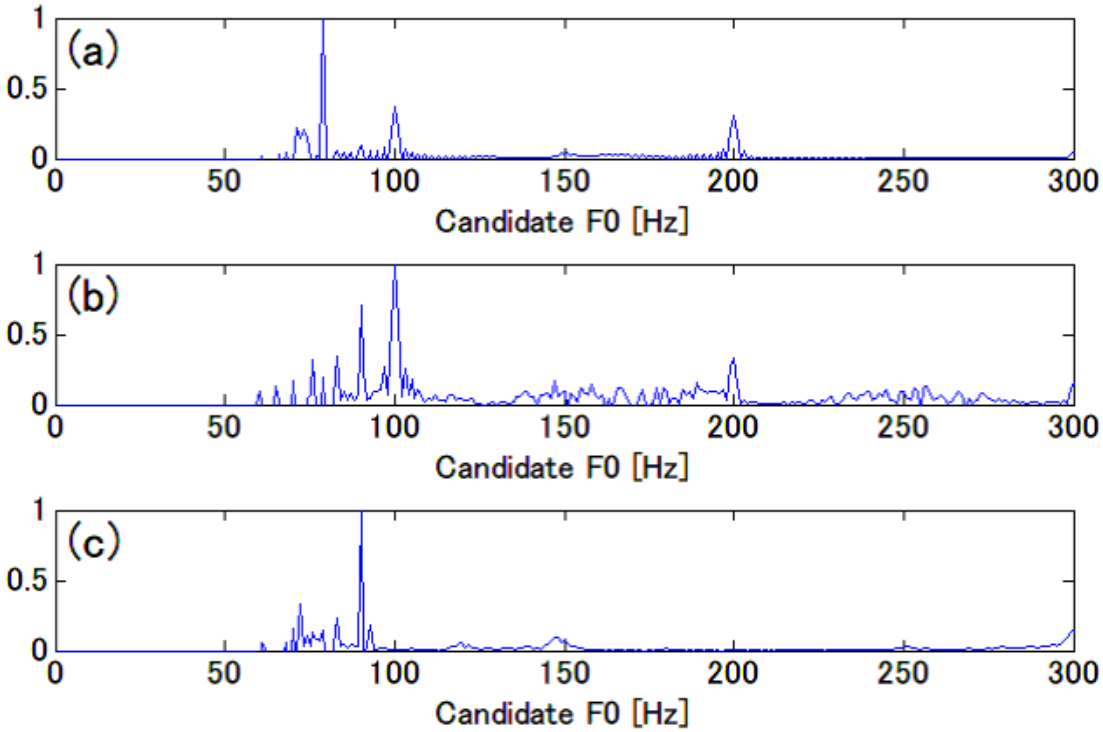

図 4.3: 平均値指標（直流成分）を用いた F0 候補の特定：(a) 静音環境，(b) 雑音環境，(c) 残響環境

図 4.4 は横軸は周波数の候補値（Hz) で，縦軸は復調信号と正弦波との相互相関係数である．この場合の F_0 も 200 Hz であるが，静音環境（図 4.4(a)），雑音環境（0 dB，図 4.4(b)），残響環境（1.0 sec，図 4.4(c)）のいずれにおいても，確かに 200 Hz で相互相関係数が極大値を迎えており，評価指標として有効であることが分かる．特に残響環境においても有効な指標であることから，復調信号波形の評価指標として非常に有用であると言える．

4.2.4 各評価指標のまとめ

検討した復調信号波形の各評価指標の特性は，下記のように特徴付けられる．

- 自己相関関数による指標：雑音・残響等の外乱に弱い．
- FFT による指標：偽候補は出やすいが，雑音や残響等の外乱には頑健．
- 直流成分による指標：雑音環境下で有用．
- 波形形状による指標：雑音環境下及び残響環境下で有効．

上記 4 指標にはそれぞれ一長一短あるものの，概ね相補的な関係にあることから，復調信号の評価指標として活用することとする．

4.2.5 各評価指標の有機的統合

前節までで検討した 4 種類の評価指標を，F0 決定プロセスに効果的に取り入れることを考える．ここで，証拠理論の分野で使われる「Dempster と Shafer の結合規則」[76] を提案法に導入する．同結合規則は次式で表される．

$$m(\omega_\kappa) = \sum_{\omega_i \cap \omega_j = \omega_k} m^\alpha(\omega_i) m^\beta(\omega_j) \tag{4.4}$$

上式では，
（１）独立した情報源 α，β から得られた基本確率 m^α，m^β を掛け合わせる
（２）（１）より得られた結果の総和を順に足していく
ことにより，基本確率 m が求められることを表している．

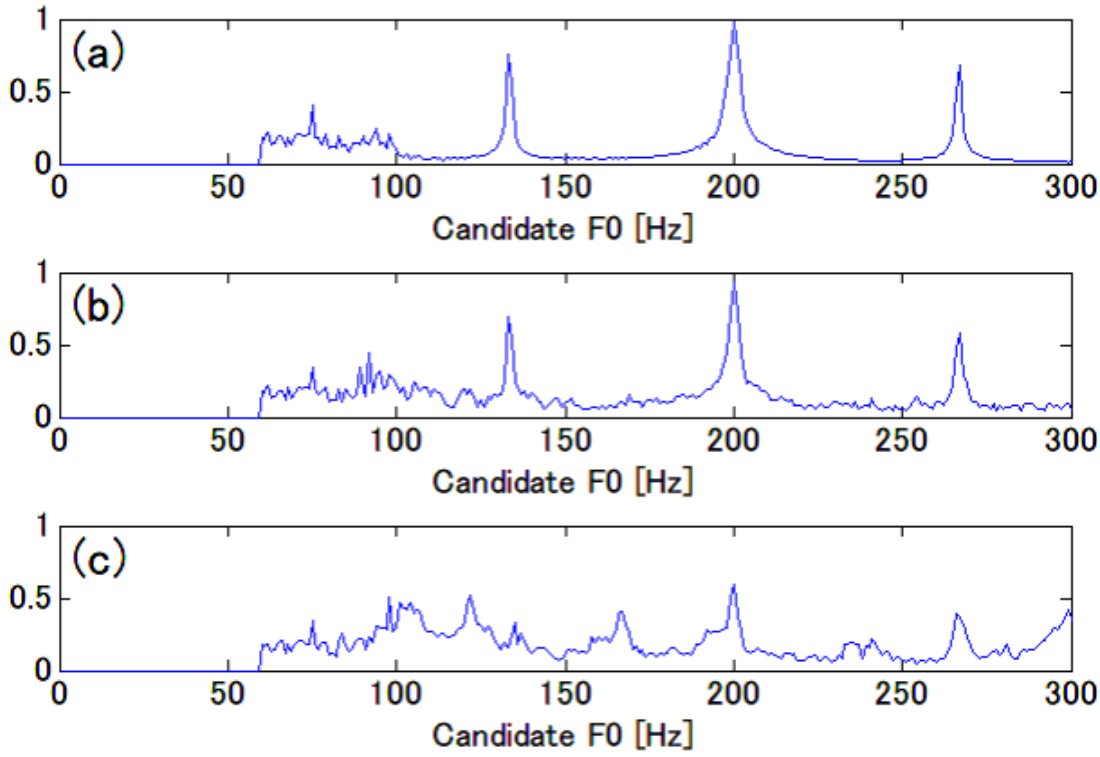

図 4.4: 相関値指標を用いた F0 候補の特定：(a) 静音環境，(b) 雑音環境，(c) 残響環境

ここでは，まず上記結合則の（１）を各指標に当てはめる．代入した指標は，以下に示す 4 指標である．

・設定値と自己相関係数との一致率による指標：A（図 4.5(a)）

・設定値と FFT 出力との一致率による指標：F（図 4.5(b)）

・直流成分による指標：D（図 4.5(c)）

・復調信号波形による形状：W（図 4.5(d)）

以上の A，F，D，W を乗算すると，図 4.5(e) の総合評価値が導き出される．この例では，総合評価値から 200 Hz が F_0 であると導き出される．

4.3　AM 信号中の外乱除去機構

提案法では，隣り合う 3 本の調波を抽出して得られた AM 信号を復調することで，F0 の基本周期を取り出している．しかし，実際の観測信号は実環境では雑音や残響の影響を受ける．そのため，そこから取り出された AM 信号もやはり雑音や残響の影響を受け，理想的な AM 波形からは崩れてくるため，最終的な復調にも悪影響を及ぼす．従って，外乱の影響を受けた AM 波形に対して，信号処理により不用成分を除去して，できるだけ理想的な AM 信号に近づけてやることが必要である．

ここでは，周波数ドメイン上で不用スペクトルをキャンセルする手法を採用した．図 4.6(a) が外乱の影響を受けた AM 信号である．この全スペクトルの中央値を求め，その中央値以下のスペクトルを削除することで，図 4.6(b) に示すように，理想的な AM 成分に近づける処理を行っている．

4.4　外乱の影響を受けた復調信号の波形回復機構

提案法においては，振幅変調の復調技術を使って，AM 信号の時間包絡線（復調信号）を抽出している．雑音や残響の影響を受けると，取り出された AM 信号の時間波形の振幅が小さくなるため，それを復調した信号（時間包絡線）の周期の特定が難しくなる．そのため，復調信号の振幅を，より本来の大きさに近い値に回復するための処理が必要である．

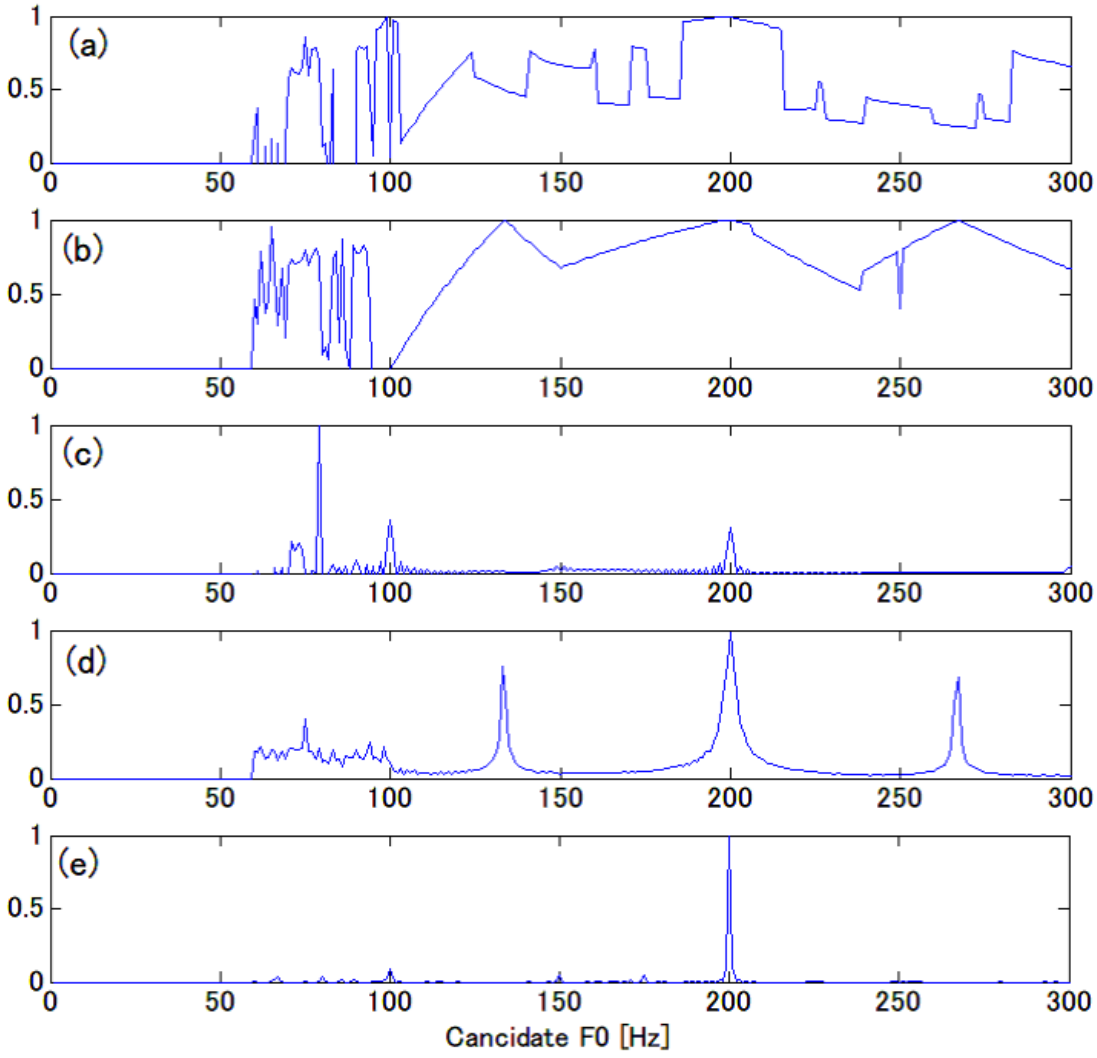

図 4.5: 各評価指標の統合例：(a) 指標 A，(b) 指標 F，(c) 指標 D，(d) 指標 W，(e) 総合評価値 R

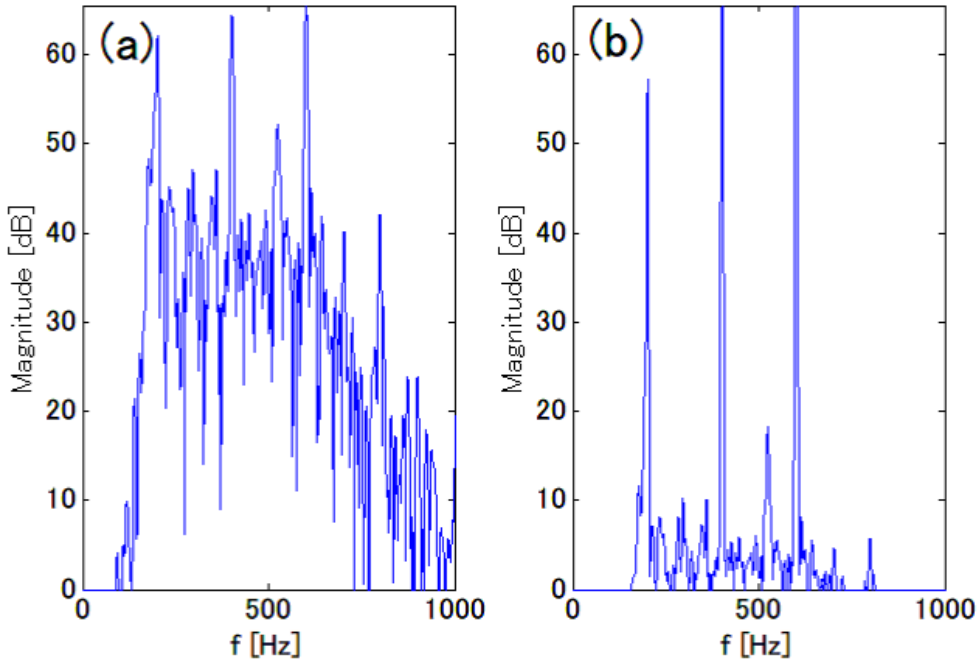

図 4.6: 周波数ドメイン上での不用成分の除去：(a) 除去処理前，(b) 除去処理後

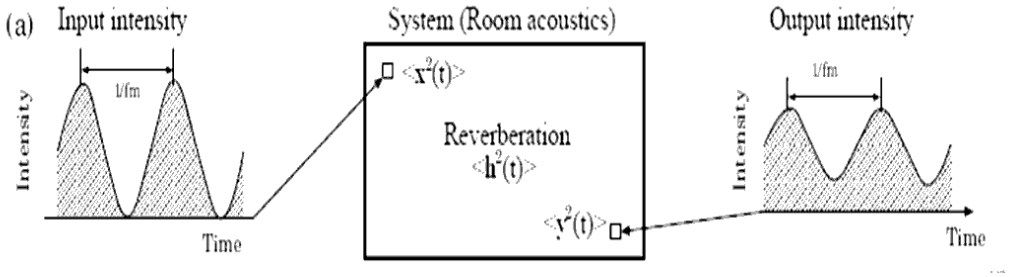

図 4.7: 残響が AM 成分に与える影響（鵜木） [72]

そこで，MTF のコンセプトに基づく波形回復手法 [72] を導入する．図 4.7 に残響が AM 成分に与える影響を示す．左側の入力側の AM 波形が残響の影響を受けると，右側の出力波形のように，振幅が抑圧された波形となる．つまり，振幅変調で言う変調度で表現すると，入力側の波形の変調度を 1 とすると，出力側の変調度は 1 以下となる．ここで，出力側の AM 信号の変調度を 1 に戻すように逆フィルタ処理を施すことにより，AM 波形を右側の入力波形に戻してやることを考える．出力信号 $E_y(z)$ を入力信号 $E_x(z)$ に回復してやるための逆フィルタリングの式は次式で表される．

$$E_x(z) = \frac{E_y(z)}{a^2}\left(1 - \exp\left(-\frac{13.8}{\hat{T_R} f_s}\right) z^{-1}\right). \tag{4.5}$$

ただし係数 a は次式で与えられる．

$$\hat{a} = \sqrt{\frac{1}{\int_0^T \exp\left(\frac{-13.8t}{\hat{T_R}}\right) dt}} \tag{4.6}$$

ただし T_R は未知であるが，変調度については波形から類推することができる．今回の実装においては，$E_x(z)$ の変調度を観測しながら T_R パラメータを順に変化させつつ逆フィルタリングを施し，変調度が 1.0 に到達した時点を最適な回復ポイントとみなして逆フィルタリング処理を終了させることとする．図 4.8 に波形回復処理による波形回復の状況を示す．図 4.8(a) は，残響の影響を受けた 200 Hz 復調信号波形である．それに対し，図 4.8(b) は，逆フィルタリング処理を施した後の復調信号波形であり，確かに 200 Hz の復調波形が回復されている様子が確認できる．

4.5 調波構造を考慮した多数決処理の検討

音声信号においては，発話される母音の種類により，特定の箇所の周波数が声道による共鳴により強調されるフォルマント周波数と呼ばれる現象が出現する．加えて，雑音や残響の影響が，音声の比較的低域の部分に現れる [73]．一方で，高域になるほど音声の調波のレベルが下がり，またその調波構造も不明瞭になるため，やはり雑音の影響を受けやすくなる場合もある [74]．そのため，音声の周波数構造

を考えたとき，その AM 成分を抽出する箇所の考慮が必要である．低域の成分だけを抽出すれば，フォルマントや外乱の影響が避けられず，正確な基本周波数推定が難しくなる．高域の部分だけを抽出するなら，調波構造が不明瞭な情報だけを使っていることになり，やはり正確な基本周波数推定には好ましくない．そこで，低域から高域までもれなく信号を抽出するための改良を行い，これら観測信号の情報をフルに活用することを考える．具体的に 10 次の調波構造を想定するとすると，f_1 から f_{10} までの隣り合う 3 本一組の調波の全ての組み合わせは 8 通り存在する．これら 8 通りの組み合わせにより調波の様々な箇所から AM 信号を抽出して得られたそれぞれの結果（前節の総合評価値 R，つまり R_1〜R_8）を，「Dempster と Shafer の結合規則」の（２）の総和則を適用して情報統合する．

$$m(\omega_\kappa) = \sum_{i=1}^{8} R_i \tag{4.7}$$

図 4.9 にその一例を図示する．図 4.9(a)〜(h) は，3 本一組の調波の 8 通りの組み合わせにより得られた総合評価値であり，図 4.9(i) はそれら 8 通りの結果の総和を正規化して示した最終総合評価値である．この例ではこの最終総合評価値から 200 Hz が F_0 であると導き出される．

以上の改良を加えた提案法の処理フローを，図 4.10 に示す．

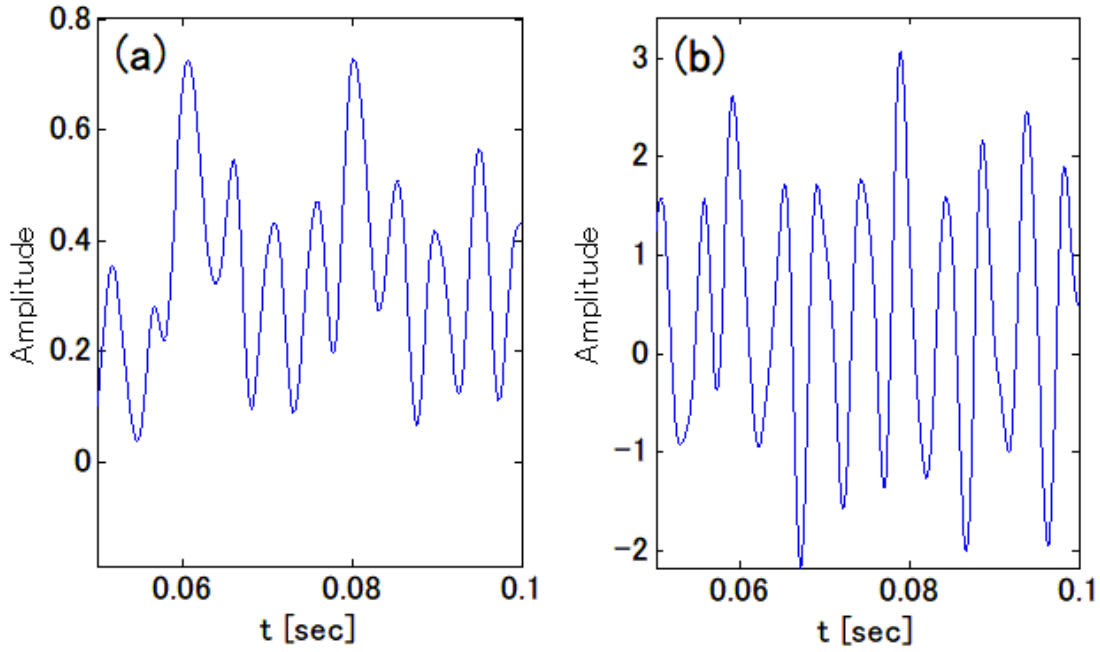

図 4.8: MTF ベースの波形回復処理：(a) 回復処理前，(b) 回復処理後

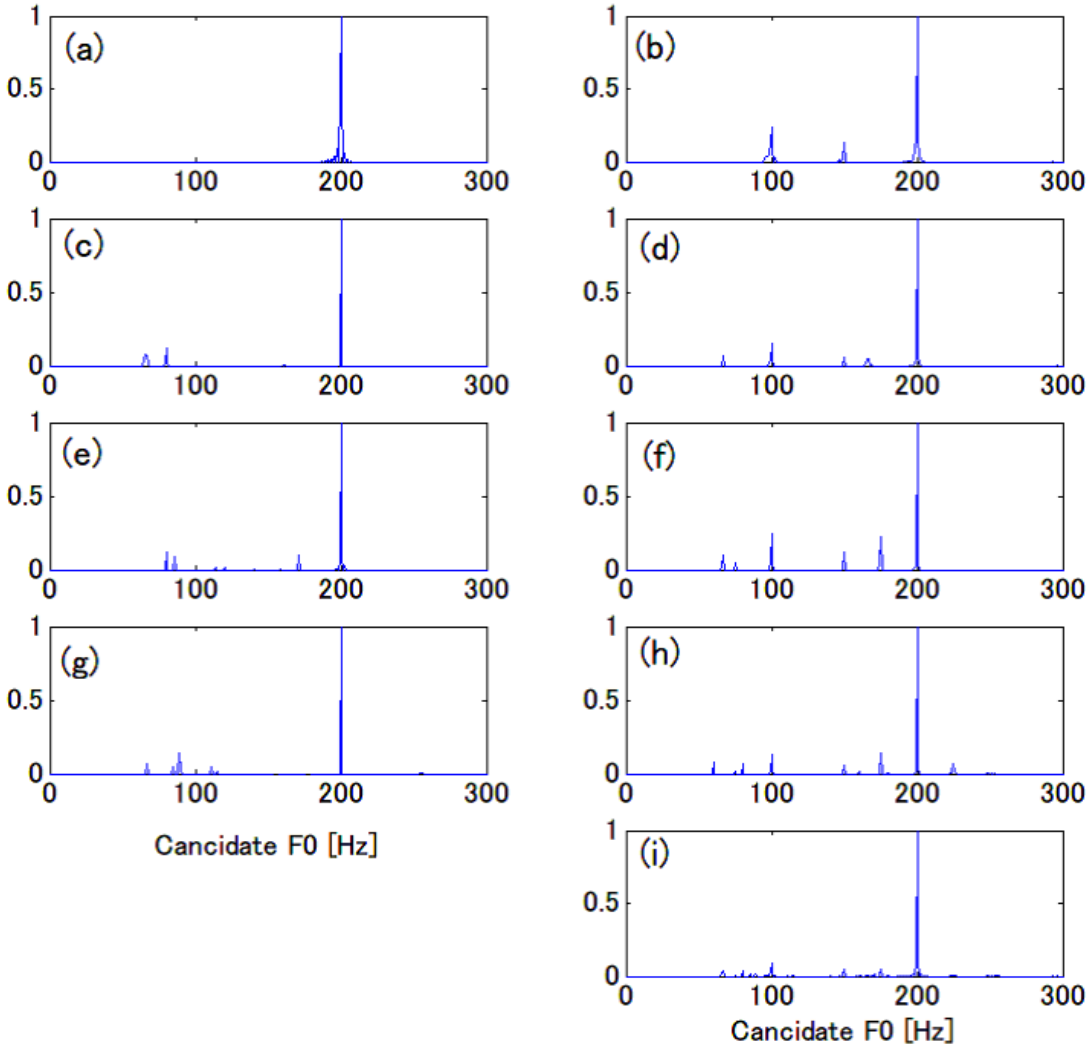

図 4.9: 総合評価指標の統合：(a) 総合評価値 R_1（f_1，f_2，f_3），(b) 総合評価値 R_2（f_2，f_3，f_4），(c) 総合評価値 R_3（f_3，f_4，f_5），(d) 総合評価値 R_4（f_4，f_5，f_6），(e) 総合評価値 R_5（f_5，f_6，f_7），(f) 総合評価値 R_6（f_6，f_7，f_8），(g) 総合評価値 R_7（f_7，f_8，f_9），(h) 総合評価値 R_8（f_8，f_9，f_{10}），(i) 最終総合評価値

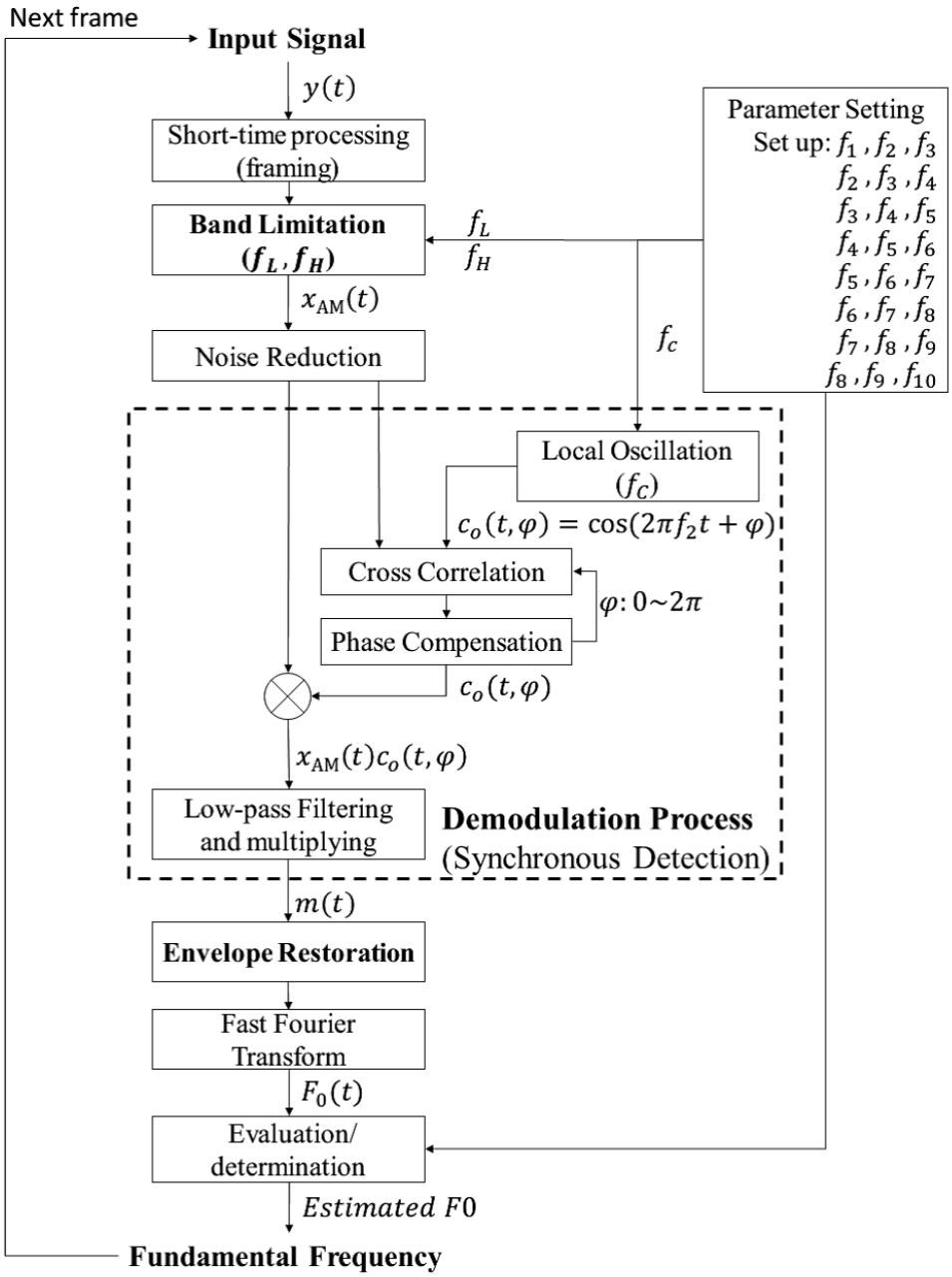

図 4.10: 提案法の処理フロー（拡張改良後）

4.6　基礎評価

ここでは，改良後の提案法が時変信号に対応できるかを確認する．

4.6.1　評価方法

入力信号と外乱の決定

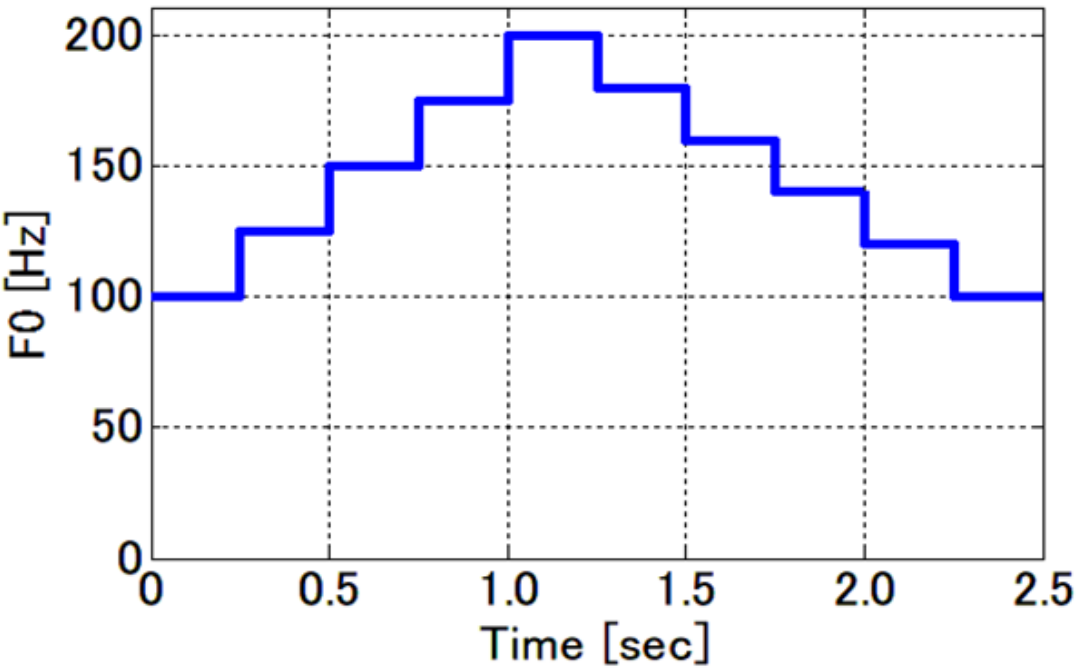

図 4.11: 時変入力信号の F0 の軌跡

時変の入力信号として信号長が 250 msec である 10 次の調波複合音を連続 10 フレーム準備した．ただし，調波複合音のレベルは基音を含めすべて同一とした．信号中の F_0 は図 4.11 に棒グラフで示すように時刻と共に階段状にへの字を描くものを仮定した．

提案法の分析窓長は 250 msec とした上で，静音環境，雑音環境，残響環境，雑音残響環境を設定し，前述の信号に対する提案法の F0 推定精度を調べる形で実施した．なお，提案法による F0 推定の実行にあたっては，直前のフレームの推定値を参酌する処理を加えた上で実行することとした．

雑音環境の背景雑音は白色雑音とし，SNR は，20，10，0，−5，−10 dB の 5 種類とした．残響環境についても人工的なものとし，入力信号に対し次式に示す統計的室内インパルス応答（Schroeder のインパルス応答 [58]）を畳み込むことによって実現した．

$$h(t) = a \exp\left(\frac{-6.9t}{T_R}\right) n(t) \tag{4.8}$$

ただし，$n(t)$ は白色雑音であり，定数 a は次式の値である．

$$a = \sqrt{\frac{1}{\int_0^\infty \exp\left(\frac{-13.8t}{T_R}\right)}} \tag{4.9}$$

残響時間 T_R については 0.1，0.3，0.5，1.0，2.0 sec の 5 種類とした．

評価尺度

評価指標として，Fine pitch error[80] と Gross pitch error[27] が一般的に知られている．Fine pitch error[80] は，真値との誤差が 20 %以内の区間のうち，真値に対する誤差率平均を表すもので，推定精度の指標である．一方，Gross pitch error[27] は，真値との誤差が 20 %以上の区間の存在割合であり，頑健性の指標である．本検証試験では，これらの評価指標に加えて，次式によって定義される許容誤差率を 5 %以内とした正答率 [%] を併せて用いることとした [56].

$$\text{Correct rate} = \frac{N_{F_{0,\text{Est}}}(E)}{N_{F_{0,\text{Ref}}}} \times 100 \tag{4.10}$$

ここで，分母の $N_{F_{0,\text{Ref}}}$ は，既に定義した F0 が含まれる入力信号の個数である．それに対する分子の $N_{F_{0,\text{Est}}}(E)$ は，許容誤差を E [%] とした場合に正しく F0 を推定できたもの個数である．ここでは許容誤差 E を 5 [%] として評価を行った．

比較対象手法

比較対象は，代表的な従来法から 6 手法を選んだ．静音環境で定評のある手法の中からは，TEMPO [53] と YIN [27] の 2 手法を，耐雑音性を持つ手法の中からは PHIA [57] と SWIPE[43, 44] の 2 手法を，耐残響性を持つ手法として複素ケプストラム法（CmpCep）[51] を，加えてフレーム処理の代表的な手法として短時間

フーリエ変換法 [36] をそれぞれ選び，比較対象とした．基本的に従来法のパラメータ設定については，その性能が最も発揮できると考えられるデフォルト値を利用した．

4.6.2 評価結果

耐雑音性能

最初に耐雑音性について，入力信号に白色雑音を付加してシミュレーションを行った．今回は 10 種類の白色雑音を準備し，5 種類の各 SNR 毎に 10 回ずつ，計 50 回試験を行った．図 4.12 に，SNR に対応する各手法の正答率を示す．横軸は SNR であり，SNR が∞表示の軸上で静音環境の結果を併せて示す．表す．まず TEMPO と YIN は，低雑音な環境では正確に推定が行えるが，SNR が 0 dB より悪くなると，途端に推定精度が低下する．PHIA と SWIPE および短時間フーリエ変換については，SNR が 0 dB においてもほぼ正確に推定が行える．複素ケプストラム法は，0 dB で 80 %の推定精度を維持しており，PHIA と SWIPE' に次ぐ耐雑音性が確認できる．一方，提案法は全雑音環境において正確に推定が行えており，従来法と比べて雑音耐性を持つことが確認できた．

耐残響性能

次に，耐残響性について，10 種類の白色雑音を基に生成した統計的室内インパルス応答 [58] を用いて畳み込んだ残響環境にて評価を行った．5 種類の各残響時間に対してそれぞれ 10 回ずつ，計 50 回試験を行った．図 4.13 に，残響時間に対応する各手法の正答率を示す．横軸は残響時間であるが，T_R が 0 表示の軸上で静音環境の結果を併せて示す．残響時間が増大すると正答率が順次低下するが，提案法は残響時間が 0.5 s までは 100 %を維持しており，従来法に比して残響に対する耐性を持っていることが確認できた．ただし，残響時間が 2.0 s まで増大すると，提案法の正答率は従来法と同レベルまで低下しており，長時間の残響に対する提案法の耐性は従来法並みであることも判明した．

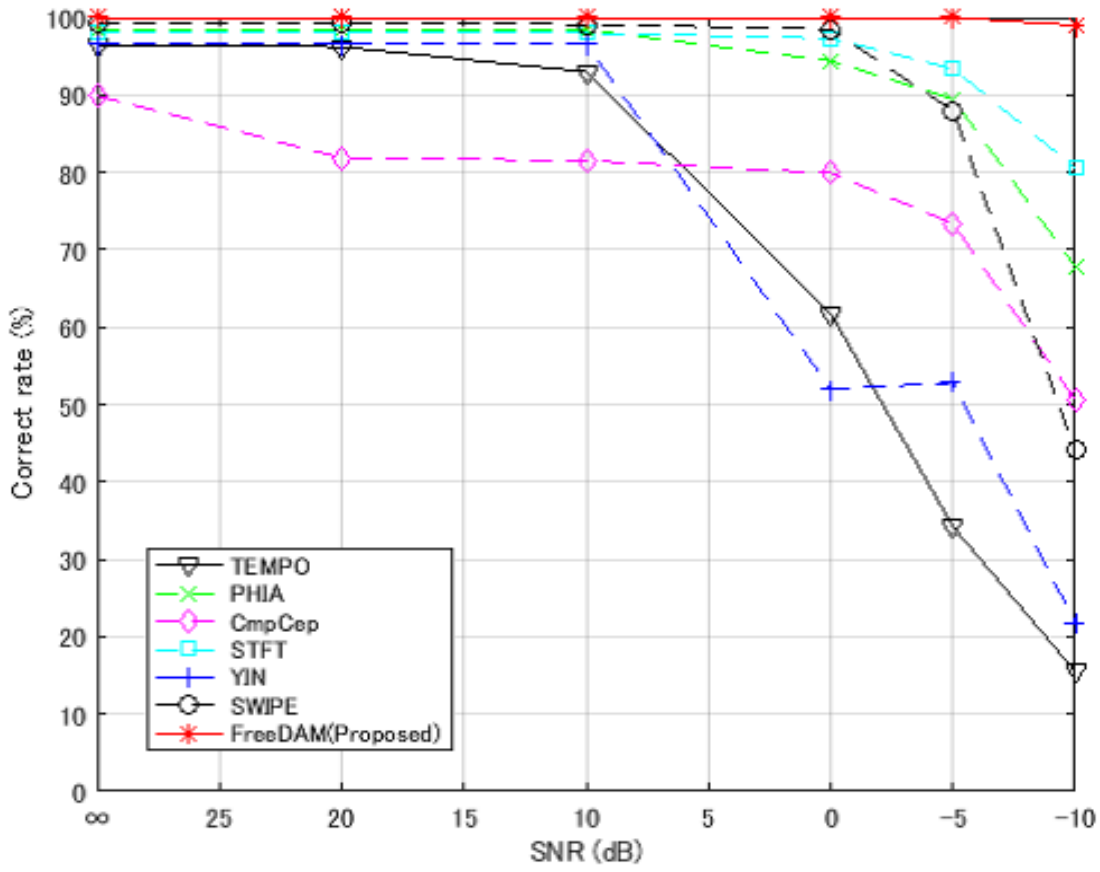

図 4.12: 雑音環境における推定精度

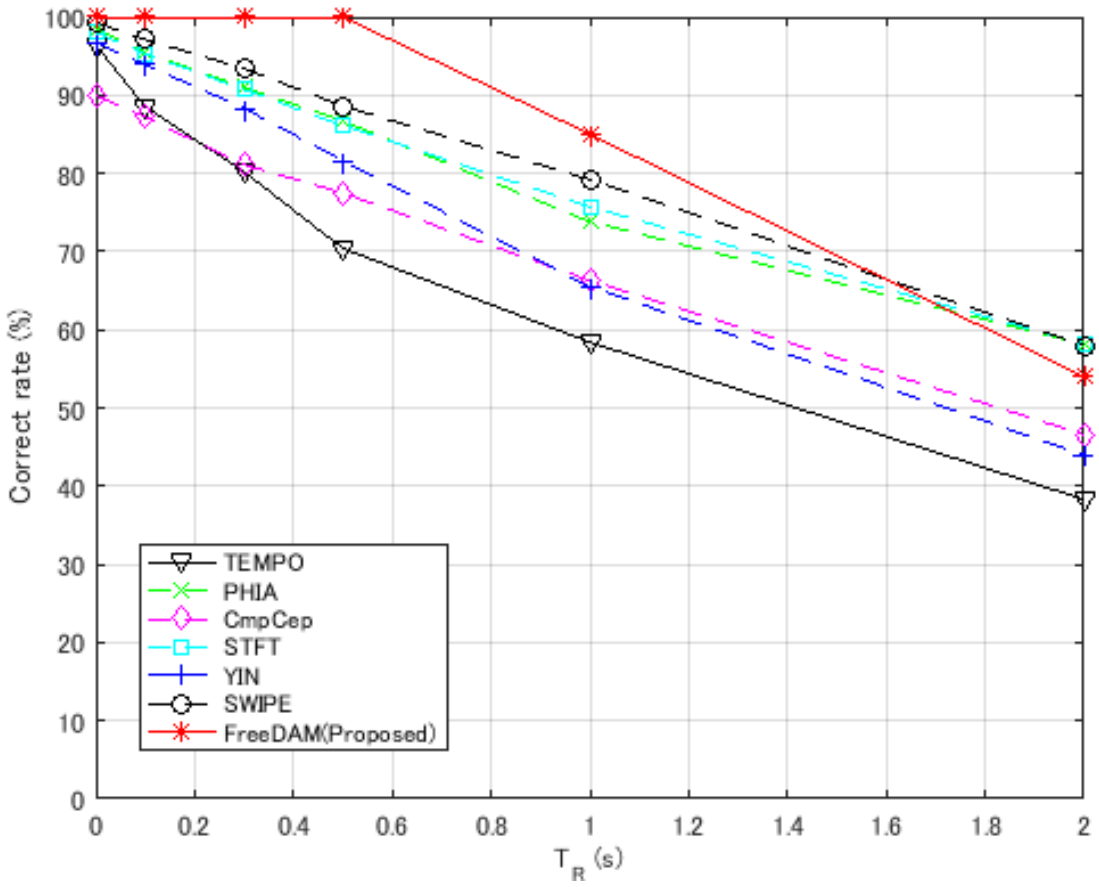

図 4.13: 残響環境における推定精度

耐雑音残響性能

次に，雑音と残響が混在した環境での評価を行った．5 種類の白色雑音と人工インパルス応答による 5 種類の残響時間とを用いて，25 通りの雑音残響環境を生成し，それぞれ 10 回ずつ，計 250 回試験を行った．図 4.14 に，各 SNR と残響時間とに対応する各手法の正答率（許容誤差　5 %）を示す．TEMPO（図 4.14(a)）と YIN（図 4.14(e)）においては，低雑音かつ短時間の残響では高推定精度であるが，雑音や残響時間が増大すると著しく推定性能が低下する．これらと比較して，PHIA（図 4.14(b)），複素ケプストラム法（図 4.14(c)），短時間フーリエ変換（図 4.14(d)），SWIPE'（図 4.14(f)）は雑音残響環境下のほぼ全域で安定した性能を発揮するが，SNR が −10 dB の条件では揃って性能が低下する．それに対して提案法（図 4.14(g）は，特に SNR が −5 dB までで T_R が 0.5 s 以内の条件に限れば，100 %の正答率を維持しており，高い外乱耐性を持っていることが確認できる．

次に，外乱に対する頑健性を確認するために，図 4.15 に各 SNR と残響時間とに対応する各手法の Gross pitch error を示す．TEMPO（図 4.15(a)），PHIA（図 4.15(b)），複素ケプストラム法（図 4.15(c)）においては，ほぼ全コンディションにおいて Gross pitch error が出現しており，外乱に対する頑健性は高くない．短時間フーリエ変換（図 4.15(d)）と YIN（図 4.15(e)）は，雑音と残響がごく小さい条件では頑健性は高いが，外乱が増大すると頑健性は低下する．SWIPE'（図 4.15(f)）は，SNR が 10 dB 以上の条件においては，残響時間の大小にかかわらず頑健である．それに対して提案法（図 4.15(g）は，特に SNR が −5 dB までで T_R が 0.5 s 以内の条件に限れば，Gross pitch error はほぼ 0 %となっており，従来法と比べて高い頑健性を持っていることが確認できる．

次に，外乱に対する正確性を確認するために，図 4.16 に各 SNR と残響時間とに対応する各手法の Fine pitch error を示す．TEMPO（図 4.16(a)）については，雑音又は残響時間が増大すると，正確性は低下することが確認できる．PHIA（図 4.16(b)），複素ケプストラム法（図 4.16(c)），短時間フーリエ変換（図 4.16(d)），YIN（図 4.16(e)），SWIPE'（図 4.16(f)）5 手法は，残響時間が増大するにつれて揃って正確性が低下することが確認できる．それに対して提案法（図 4.16(g）は，T_R が 0.5 s 以内の条件に限れば，残響時間に関わりなく Fine pitch error はほぼ 0 %となっており，従来法と比べて正確性にも優れることが確認できる．ただし，残

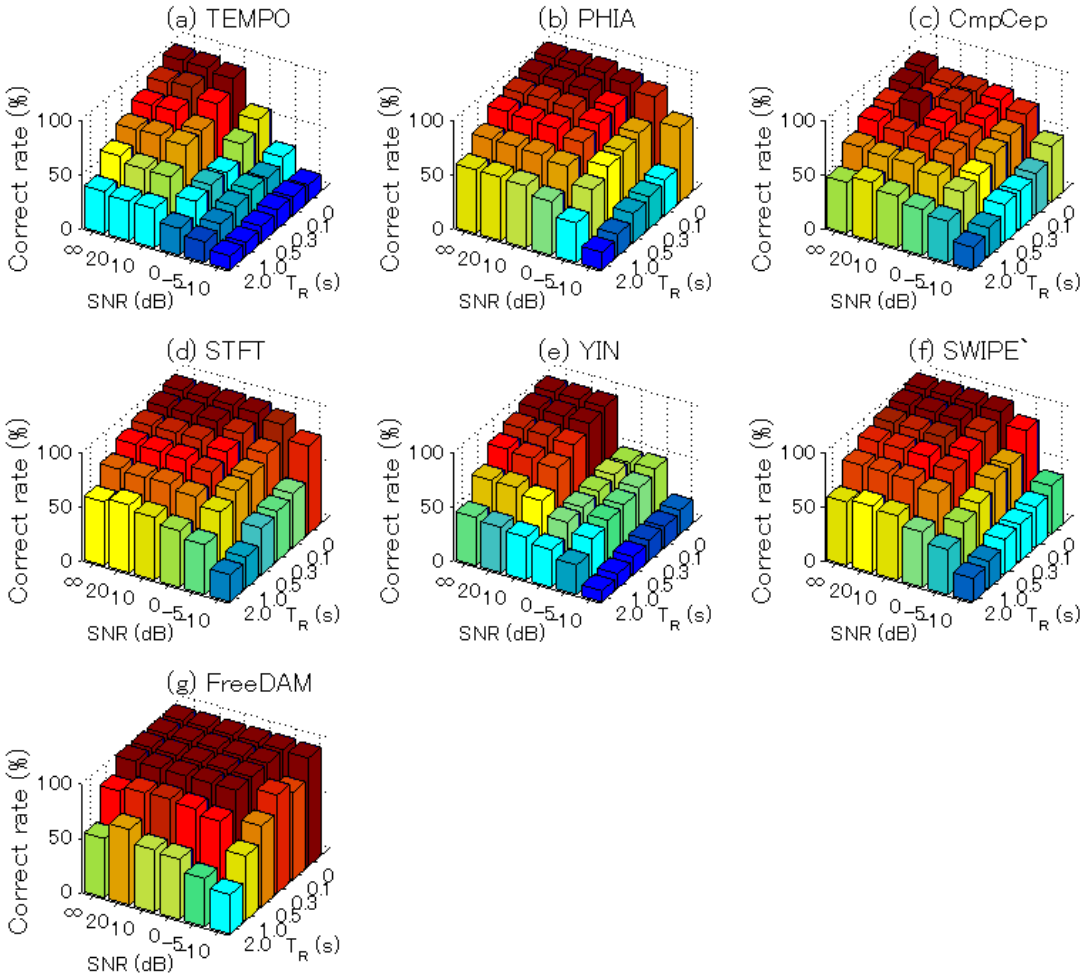

図 4.14: 雑音残響環境における正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

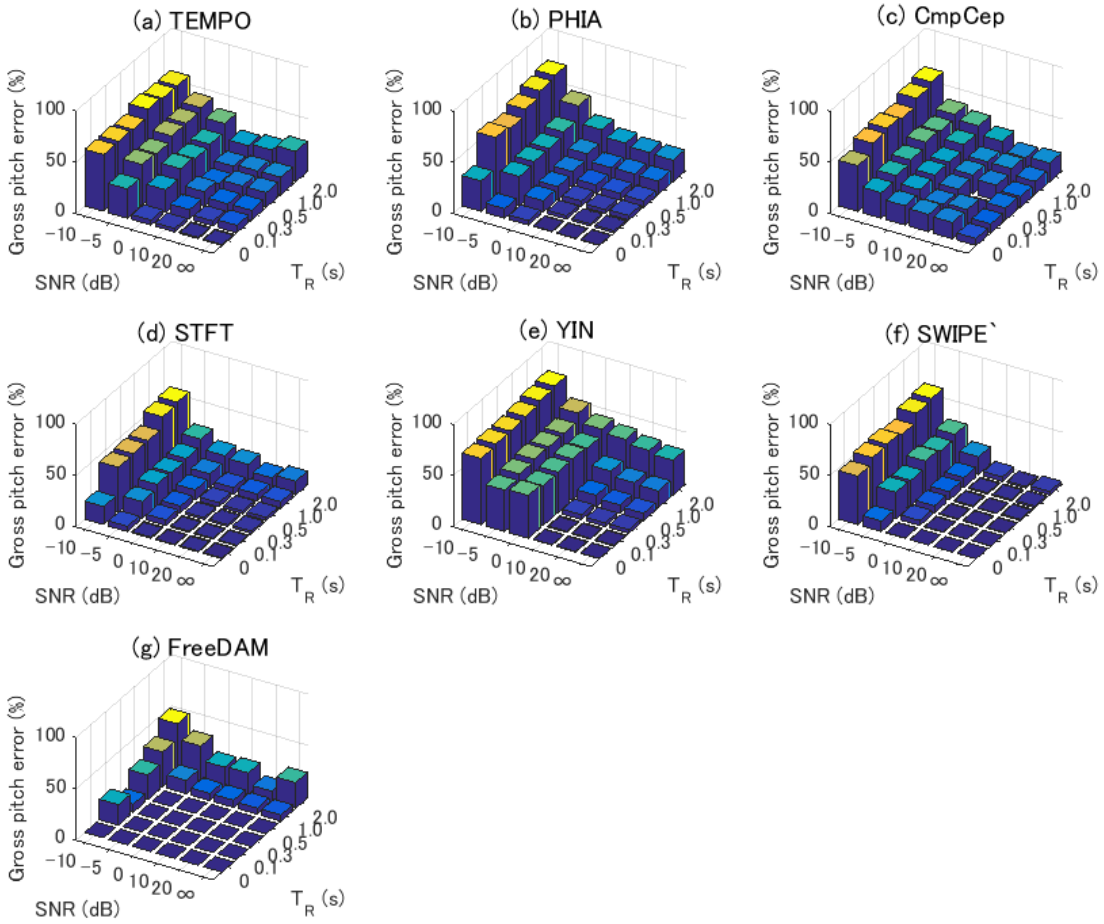

図 4.15: 雑音残響環境における Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

響時間 T_R が 2.0 s の条件においては，提案法の正確性は一部の従来法よりも低下する結果となった．

図 4.17 に，雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における F_0 推定軌跡の一例を示す．従来法はいずれも，真値から大きく外れる場合が散見されるが（図 4.17(a)-(f)），提案法（図 4.17(g)）は真値から外れることなく推定が行えていることが分かる．

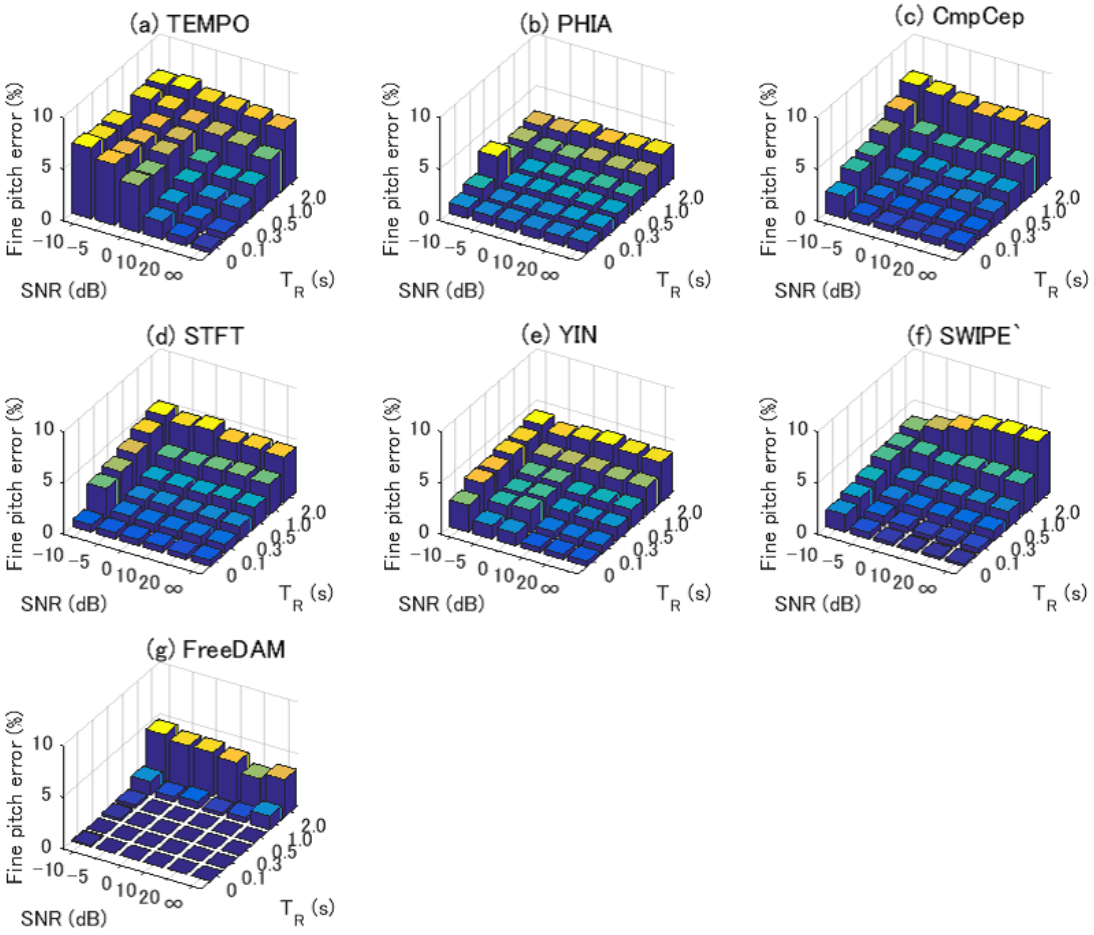

図 4.16: 雑音残響環境における Fine pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

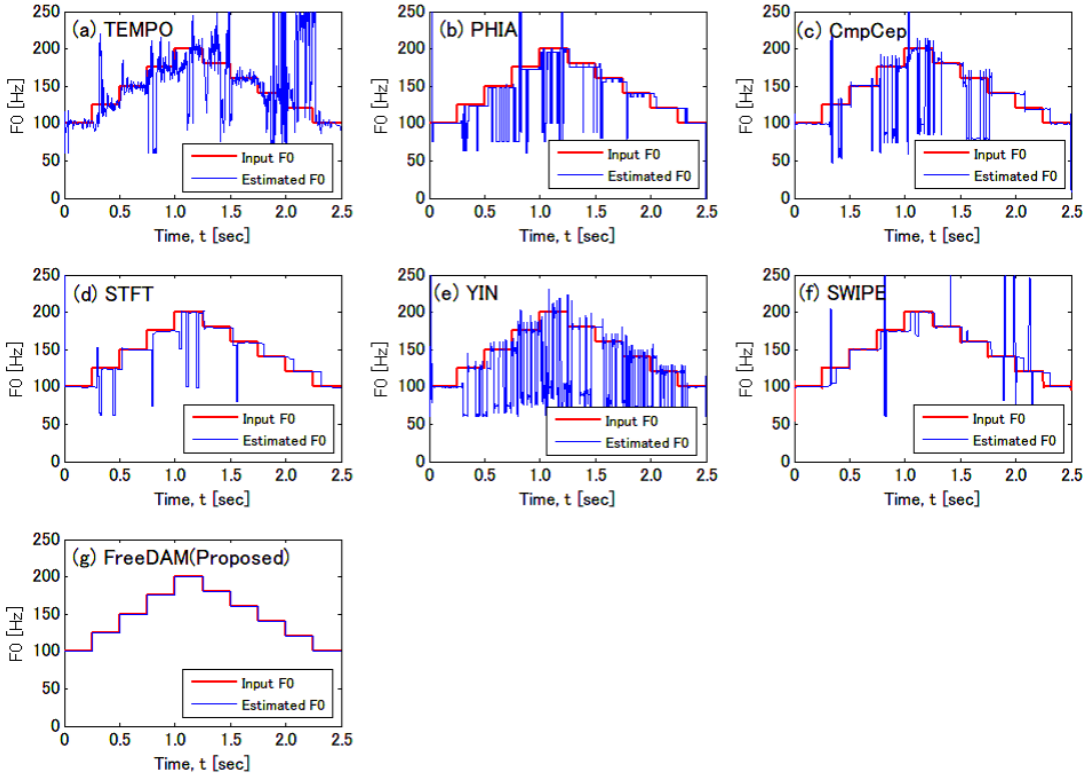

図 4.17: 雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における時変信号の F_0 推定軌跡の一例： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

追加確認シミュレーション

次に，提案法の分析窓が入力信号のステップ間に跨がる場合の確認シミュレーションを実施した．ここでは，同じ入力信号に対し，提案法の分析窓長は同じく 250 msec とし，フレームシフトだけを 125 msec に変更することで，確実にステップ間を跨ぐ分析箇所を発生させることで実施した．試験結果の正答率を図 4.18 に，Gross pitch error を図 4.19 に，Fine pitch error を図 4.20 にそれぞれ示す．

提案法 (g) の正答率（図 4.18）については全コンディションで 80 %以下にとどまっているが，これはフレームシフトが比較的長めであることに起因するものであり，シフト長をさらに短時間とすれば正答率は向上すると考えられる．Gross pitch error（図 4.19）を見ると，提案法 (g) は一部の高雑音なコンディションを除き非常に安定しており，頑健性を備えていると言える．Fine pitch error（図 4.20）については，提案法 (g) は従来法と比べて Fine pitch error が小さいというわけではないが，雑音や残響の大小にかかわらず安定はしており，シフト長をさらに短くすることで正確性は向上すると考えられる．

図 4.21 に，雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における F_0 推定軌跡の一例を示す．提案法（図 4.17(g)）は，フレームシフト長を大きくとっているため，F_0 が階段状に変化する境界付近で真値との若干のズレは見られるものの，非常に安定した推定が行えていることが分かる．

以上の結果から，提案法の分析窓が入力信号のステップ間に跨がるケースにおいても，その動作は概ね良好であることが示され，フレームシフトタイムをさらに短く採ることで恐らく正確性も向上できると期待できる．

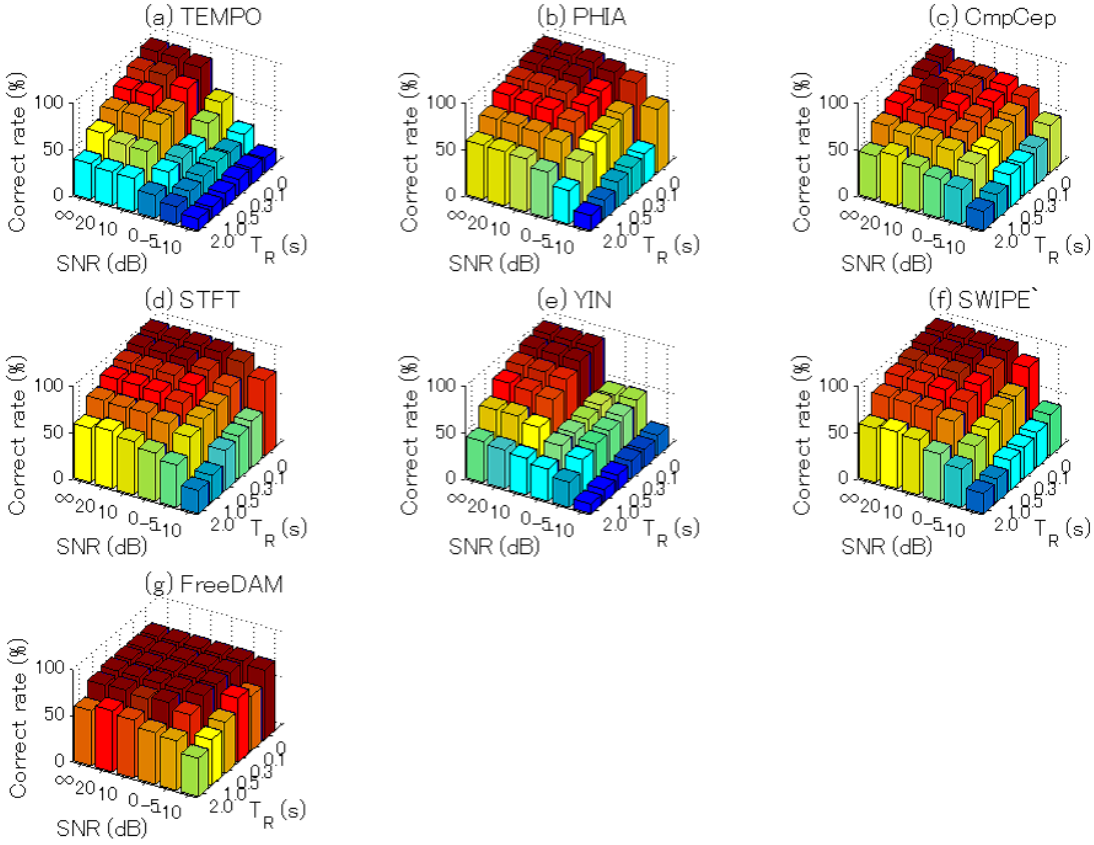

図 4.18: 雑音残響環境における正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

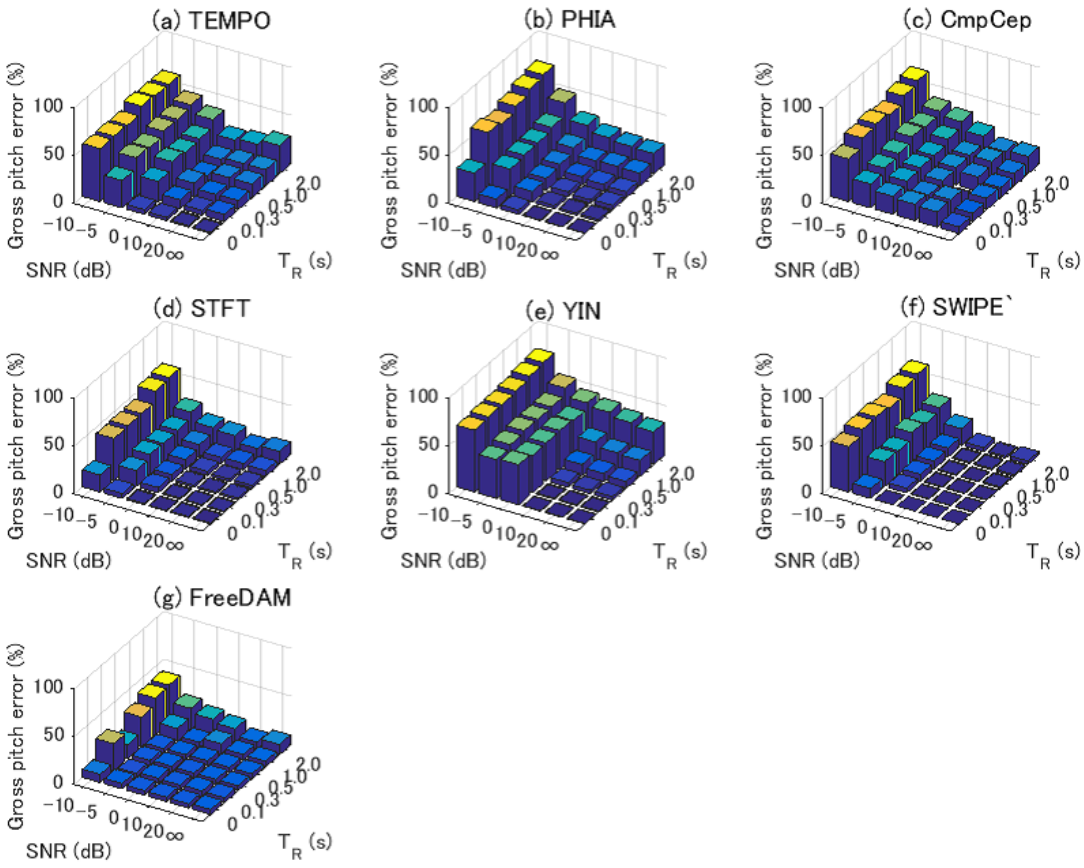

図 4.19: 雑音残響環境における Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

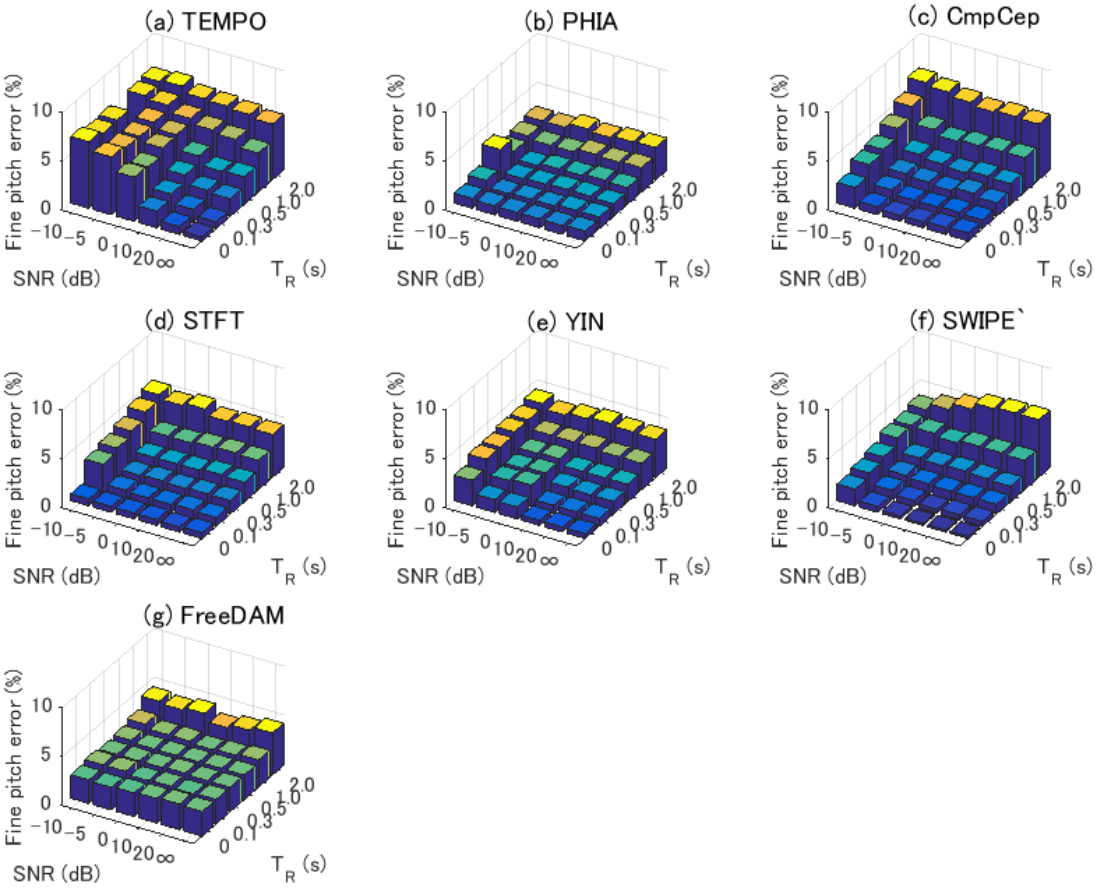

図 4.20: 雑音残響環境における Fine pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

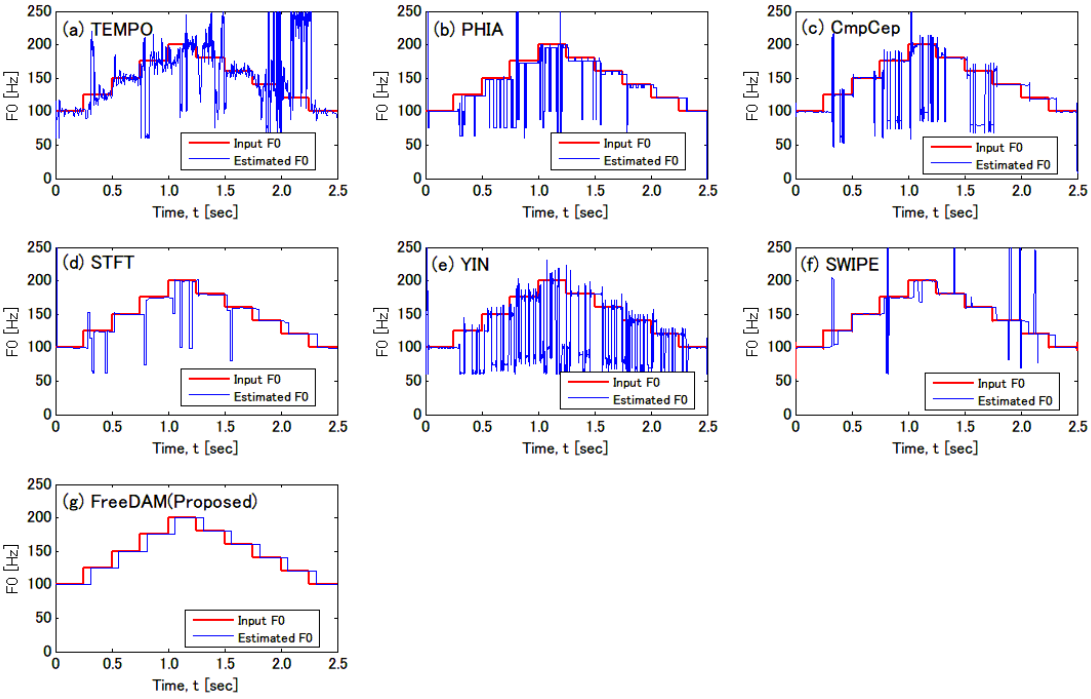

図 4.21: 雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における時変信号の F_0 推定軌跡の一例： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

4.7　楽器音への適用

ここでは，提案法のユースケースの一例として楽器音の音高推定を考える．

提案法は，時間フレーム内でのF0が一定との仮定の下にAM復調を実行する方式であることから，提案法は楽器音の音高推定に一定の適性を持つと考えられる．そこで，楽器音の音高推定への提案法の適用可能性について，以下シミュレーションにより検証する．

4.7.1　評価方法

ここで分析対象とする楽器音は，北村の制作によるYAMAHAのMIDI音源[9, 77]の中から，ピアノ音を用いることとした．そのメロディは図4.22に示すもので，C4（262 Hz）からG4(392 Hz)の範囲で音高が変動するものである．速度表記は四分音符が120回／分であることから，四分音符あたり0.5秒の長さを持つ．

提案法の分析窓長は250 msecとした上で，静音環境，雑音環境，残響環境，雑音残響環境を設定し，前述の信号に対する提案法のF0推定精度を調べる形で実施した．前回と同様に，提案法によるF0推定の実行にあたっては，直前のフレームの推定値を参酌する処理を加えた上で実行することとした．ただし今回は，ピアノ音の周波数構造を考慮し，観測信号の分析に用いる周波数の上限は5次の調波までとした．雑音環境の背景雑音は10種類の白色雑音とし，SNRは，20，10，0，-5，-10 dBの5種類とし，計50回試験を行った．残響環境については，入力信号に対し10種類の白色雑音を基に生成した統計的室内インパルス応答（Schroederのインパルス応答[58]）を畳み込むことによって実現したものを用い，残響時間T_Rについては0.1，0.3，0.5，1.0，2.0 secの5種類とした．こちらについても計50回試験を行った．雑音残響環境については，5種類の白色雑音と人工インパルス応答による5種類の残響時間とを用いて，25通りの雑音残響環境を生成し，それぞれ10回ずつ，計250回試験を行った．比較対象とした従来法は，前節と同じ6手法である．

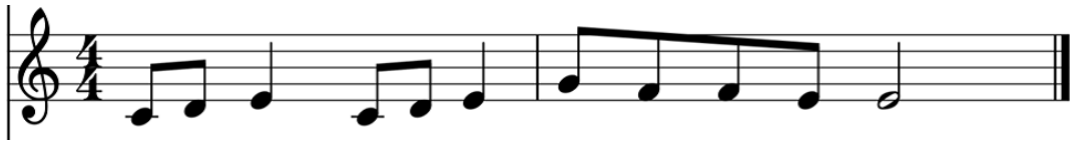

図 4.22: 試験に用いた楽器音のメロディ（北村）[9, 77]

4.7.2 評価結果

図 4.23 に，各 SNR と残響時間とに対応する各手法の正答率（許容誤差　5 %）を示す．TEMPO（図 4.23(a)）においては，静音環境では正確な推定が行えるが，外乱が加わると推定精度は低下する．YIN（図 4.23(e)）においては，低雑音では高推定精度であるが，雑音が増大すると著しく推定性能が低下する．PHIA（図 4.14(b)），複素ケプストラム法（図 4.14(c)），短時間フーリエ変換（図 4.23(d)）も同様な傾向を示すが，雑音に対する耐性は YIN より高いことが見て取れる．一方，SWIPE'（図 4.23(f)）と提案法（図 4.23(g)）は，特に SNR が 10 dB まででかつ T_R が 0.5 s 以内の条件に限れば，ほぼ 100 %の正答率を維持しており，高い外乱耐性を持つことが確認できる．

図 4.24 に，各 SNR と残響時間とに対応する各手法の Gross pitch error を示す．提案法（図 4.24(g)）は，SNR が 10 dB 以上でかつ T_R が 0.5 s 以内の条件に限れば，Gross pitch error はほぼ 0 %であり，SWIPE'（図 4.24(f)）とともに高い頑健性を示す．

図 4.25 に，各 SNR と残響時間とに対応する各手法の Fine pitch error を示す．従来法はいずれも雑音もしくは残響が増大するにつれて Fine pitch error も大きくなる傾向があるが，提案法（図 4.25(g)）は，SNR が 0 dB 以上でかつ T_R が 0.5 s 以内の条件に限れば Fine pitch error は 1 %以内に収まっており，一定の条件に限れば提案法は正確性も備えていると言える．

図 4.26 に，雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における F_0 推定軌跡の一例を示す．従来法はいずれも，真値から大きく外れる場合が散見されるが（図 4.26(a)-(f)），提案法（図 4.26(g)）は，1 箇所だけ誤推定が発生しているものの，非常に安定した推定が行えていることが分かる．

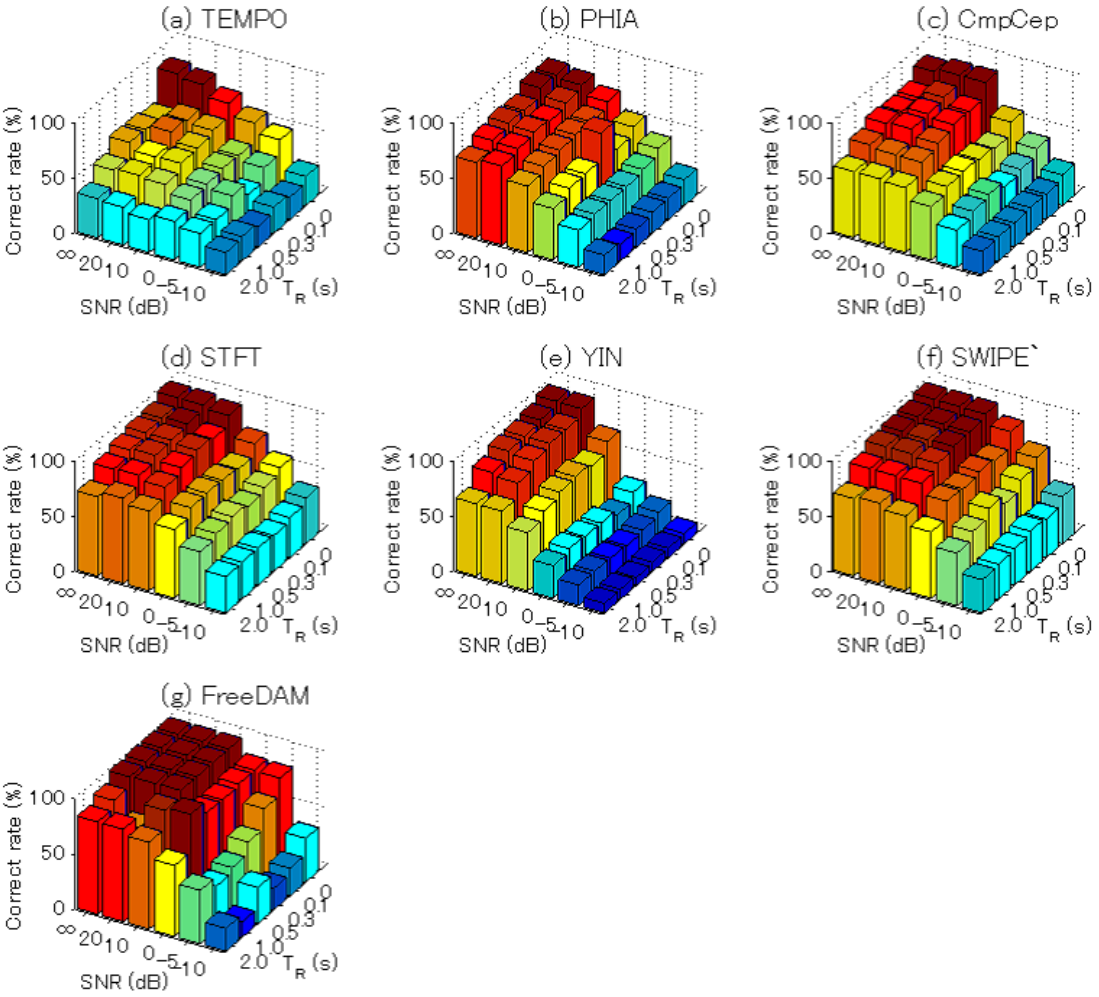

図 4.23: 雑音残響環境における楽器音の正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

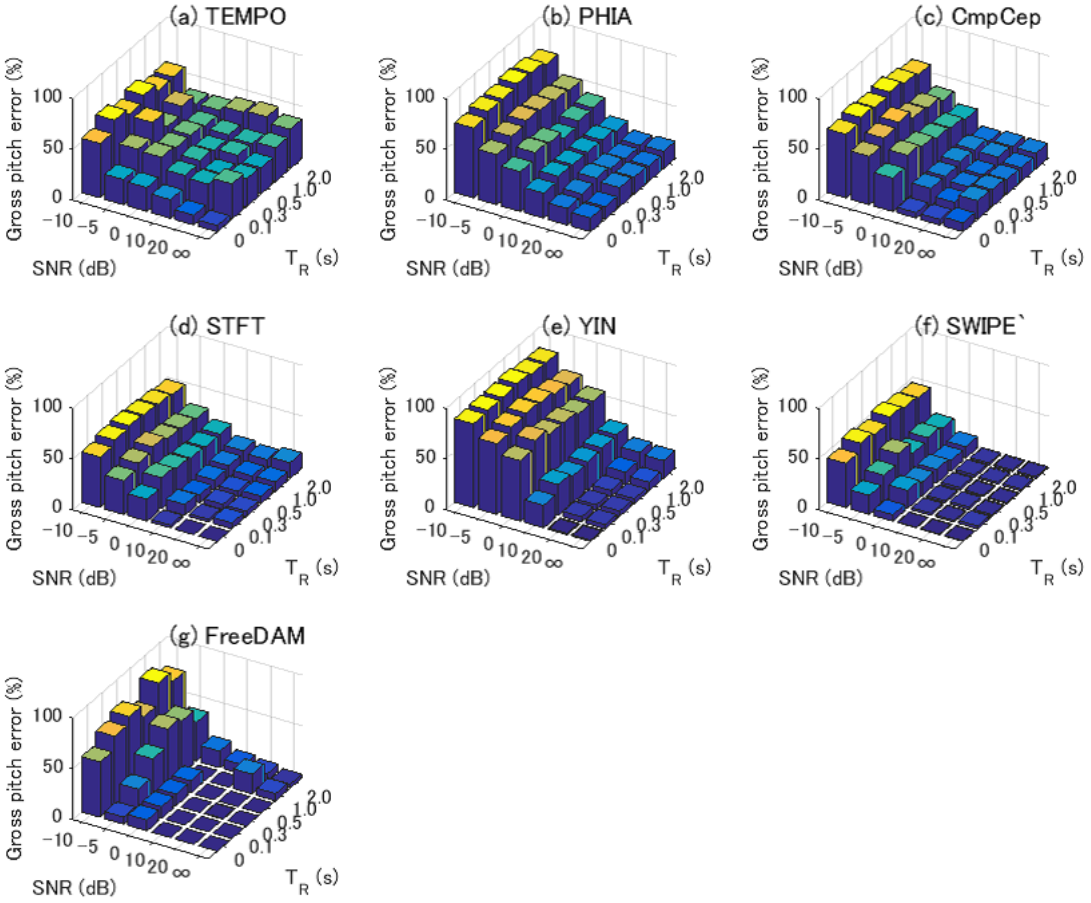

図 4.24: 雑音残響環境における楽器音の Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

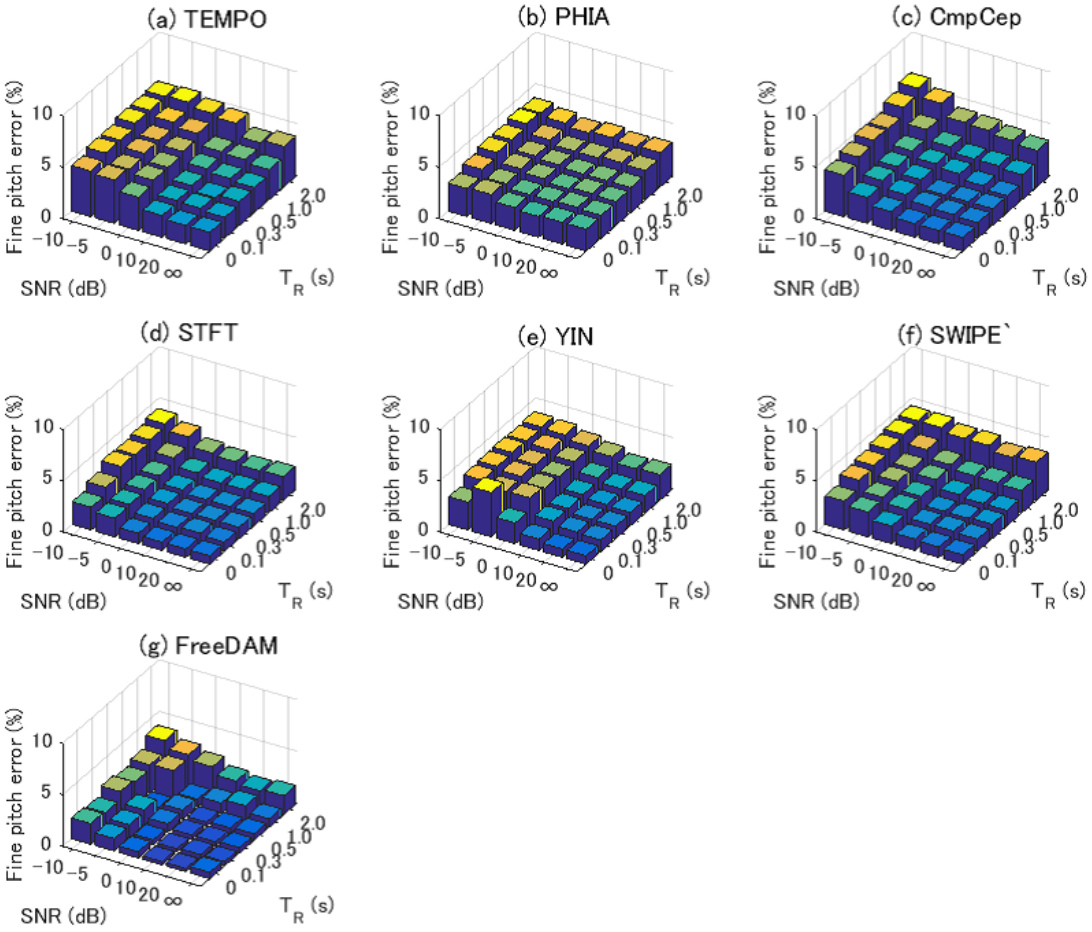

図 4.25: 雑音残響環境における楽器音の Fine pitch error：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE'，(g) FreeDAM（Proposed）

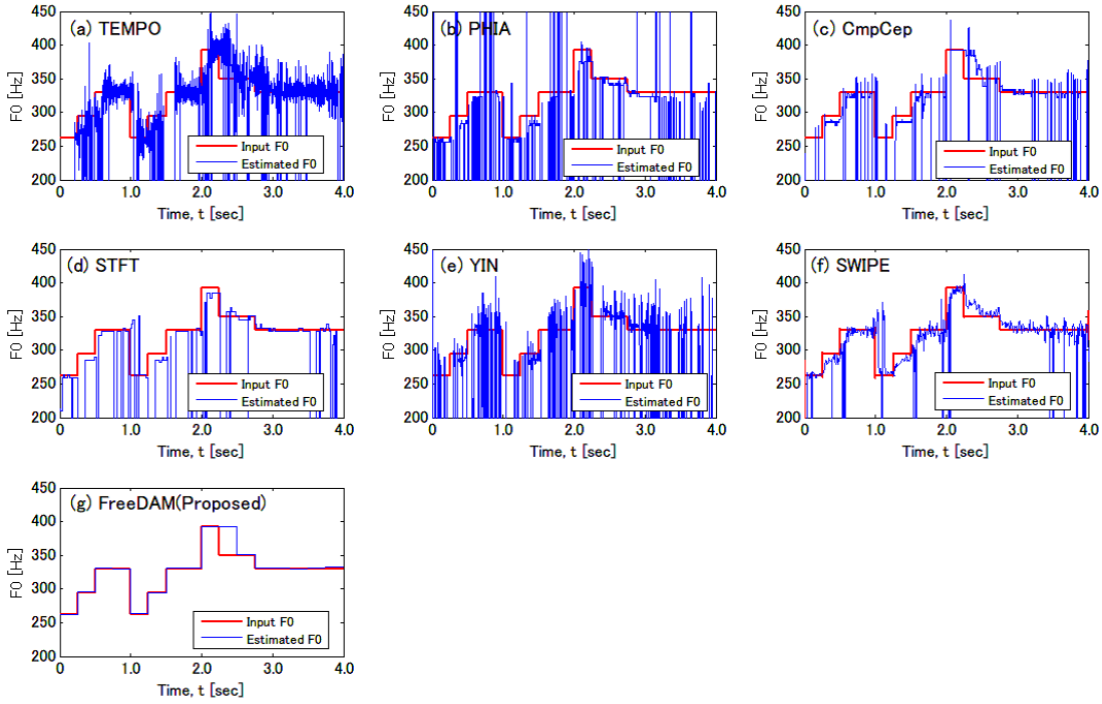

図 4.26: 雑音残響環境（SNR = 0 dB，TR = 1.0 sec）における楽器音（ピアノ）の F_0 推定軌跡の一例：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE'，(g) FreeDAM（Proposed）

4.7.3 考察

図 4.23(g）の提案法の正答率に再度着目してみると，特に SNR が 0 dB までで T_R が 1.0 s 以内の条件では，一部を除き 85 %以上の高い推定精度を維持していることが確認できる．F0 が一定時間不変であるという特性を持つ楽器音に対しては，AM 復調を用いる提案法の特性がうまくマッチしていることが考えられる．また，楽器音の音高指定を行う場合，それほど高雑音な環境は想定しなくてもよいことから，0 dB という数値は SNR として十分に条件を満たすたすものであり，また 1.0 s という残響時間も，一般的なホールにおける条件を満たすものである．よって提案法は，音楽ホールでの楽器ソロ演奏における音高推定用途に十分適用できる可能性を示していると言える．

4.8 まとめ

本章では，提案法を時変信号へ対応できるようにするための拡張を実施した．具体的には，復調波形の評価指標を複数導入し，加えて雑音の抑圧機構，残響に影響を受けた波形の回復機構を追加した．さらに，調波構造に着目した多数決処理を導入し，できるだけ多くの調波成分を用いて F0 の決定を行えるようにした．

また，拡張改良後の提案法に対して評価試験を実施し，提案法が時変の信号に対して確実に対応でき，加えて従来法を上回る頑健性を備えていることを確認した．さらに提案法の特性に合致した応用先の一例として楽器音の音高推定を考え，提案法が楽器音のような F0 が一定時間不変の信号に対しては十分適用が可能であることを示した．

第 5章

音声信号への適用可能性の検討

本章では，将来的に提案法を音声信号に適用していくために，時刻と共に F_0 が変化する様々な時変信号に対して，雑音残響環境のもとで提案法が適用できるかどうかを検証する．

5.1 評価に用いる信号の検討

ここでは，評価に用いる刺激となる入力信号を検討する．本節では，実際の音声信号を基に作成した，時刻と共にステップ状に F_0 が変化する時変調波複合音をベースとして考える．この時変調波複合音と実際の音声信号との間には，F_0 の時間変化の仕方や調波構造など，大きな隔たりがある．この時変調波複合音と音声信号とのギャップとの中間に位置づけられる人工的な信号を用いた試験も併せて実施することにより，提案法の音声信号に対する適性や限界を明らかにすることを狙う．

ステップ的に F_0 が変化する調波複合音

図 5.1 に，ATR デジタル音声データベースによる実際の男声の音声信号/aoi/の F_0 の軌跡を示す．この音声信号に対して，分析フレーム長を 40 ms とし，さらに同フレーム区間内においては F_0 を平均化して一定として処理を行うと，図 5.2 のような時刻と共にステップ状に F_0 が変化する軌跡が得られる．この各フレーム毎の F_0 に対し，それぞれ 30 次まで調波を生成することで，時刻と共にステップ的に F_0 が変化する調波複合音を作成している．その先頭フレームの周波数構造を，図 5.3 に示す．この信号では，各調波の大きさは全て均一となっており，フォルマント周波数の影響などは一切考慮していないものになる．これをスペクトログラムで示したものが図 5.4 であり，常に均一な調波が時刻と共にステップ的に不連続で変化するものとなっている．

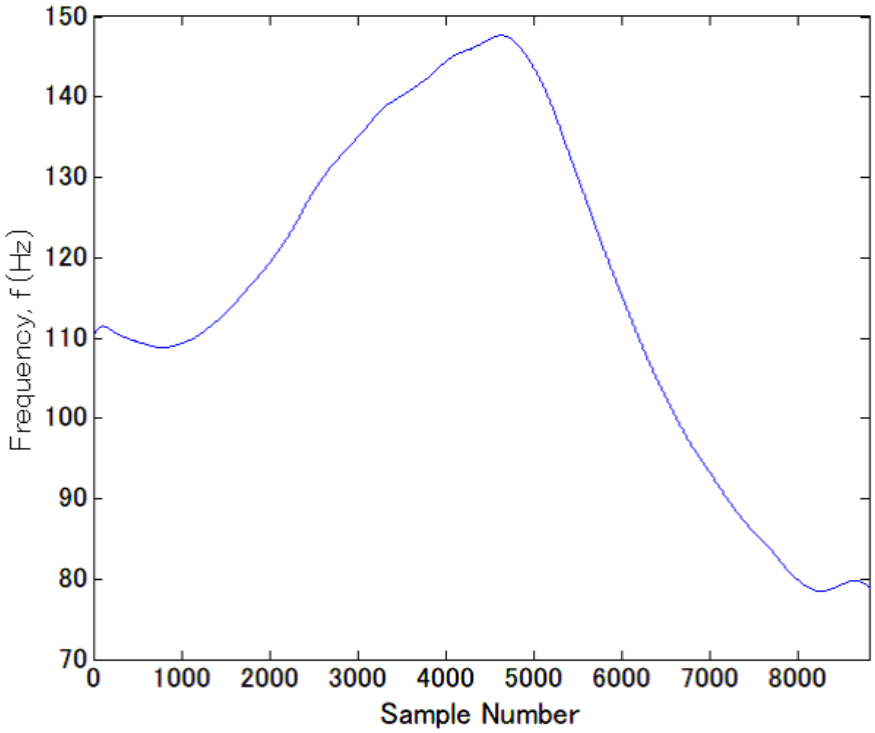

図 5.1: 実音声信号（男声）の F_0 の軌跡

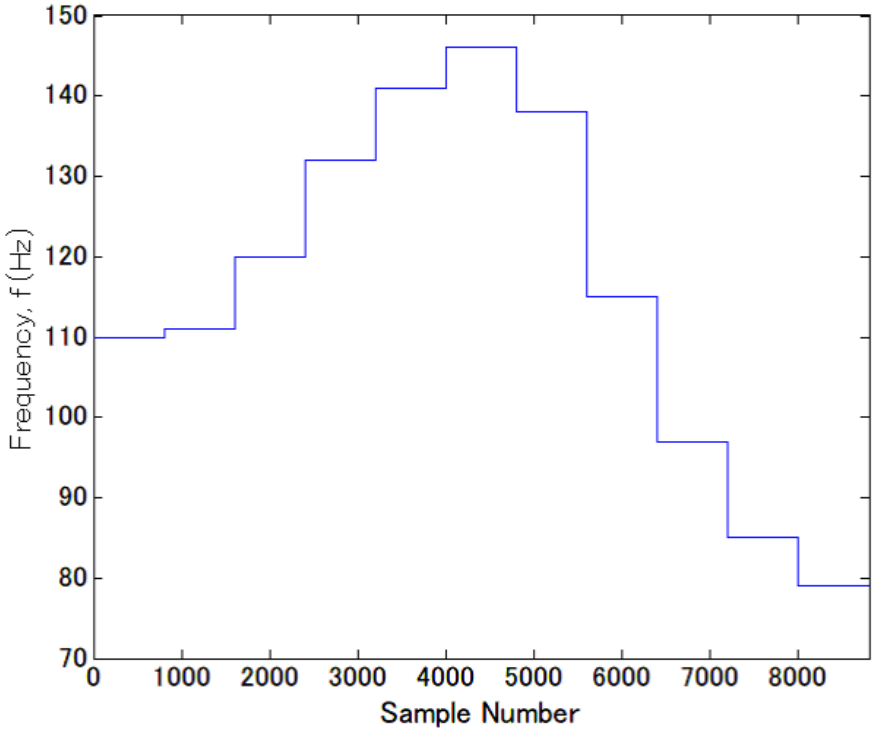

図 5.2: 時変調波複合音の F_0 の軌跡

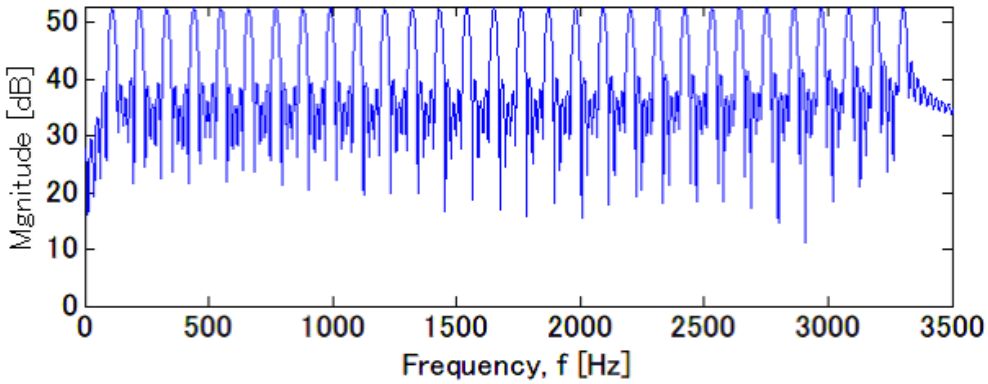

図 5.3: 時変調波複合音の調波構造

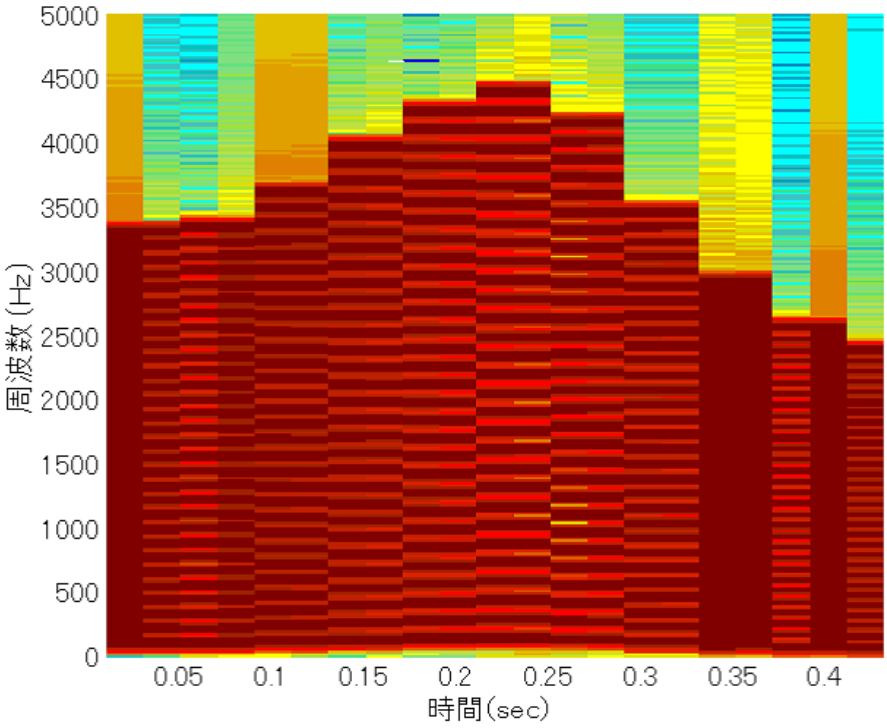

図 5.4: 時変調波複合音のスペクトログラム

ステップ的に F_0 が変化する合成音

ここでは，前述の時変調波複合音に，音声信号と同様な調波構造を反映することを考える．図 5.5(a) は音声信号/aoi/の一区間の周波数構造であるが，このスペクトル包絡を線形予測（LPC：Linear Predictive Coding）分析により求めると図 5.5(b) のようになる [17, 19, 78, 79]．このスペクトル包絡情報を，合成フィルタにより図 5.5(c) の調波複合音に適用すると，図 5.5(d) のような合成音が得られる．これをスペクトログラムで示したものが図 5.6 であり，スペクトル包絡情報が反映された調波が時刻と共にステップ的に不連続で変化するものとなっている．

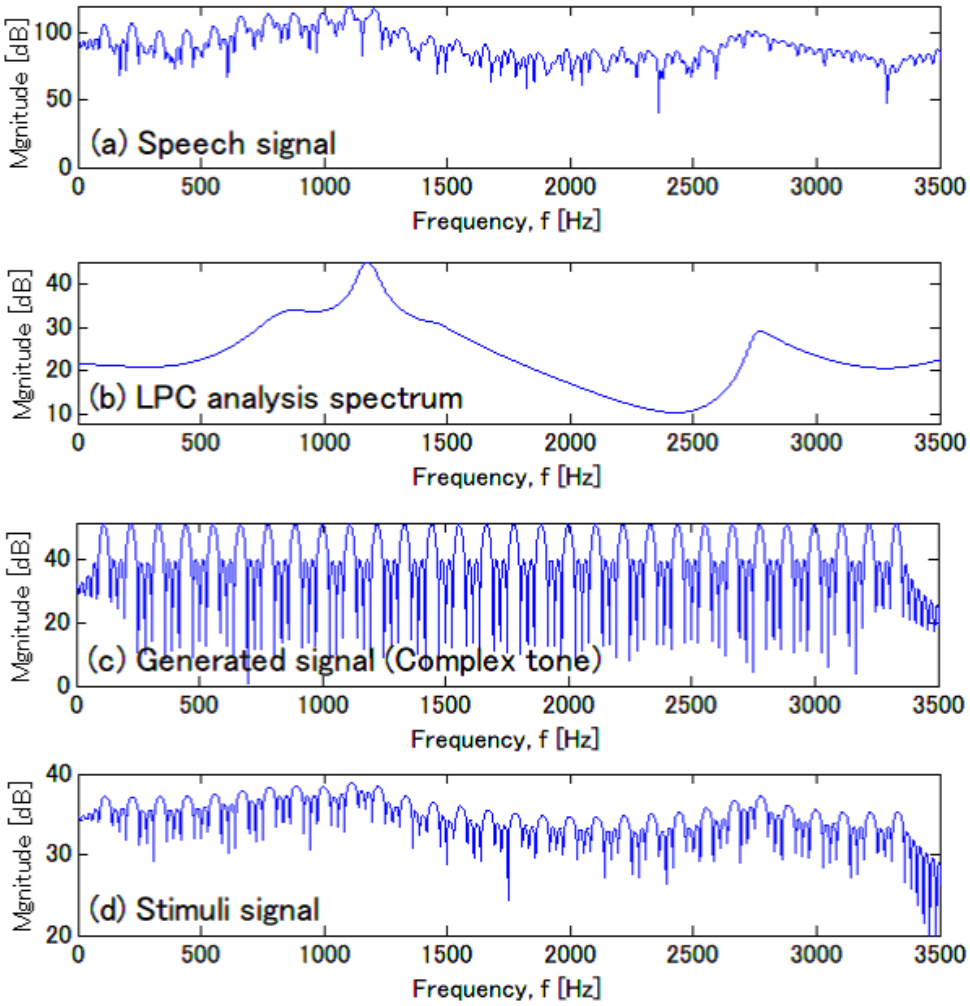

図 5.5: ステップ的に F_0 が変化する合成音の作成：(a) 音声信号，(b) 音声信号のスペクトル包絡，(c) 調波複合音，(d) 合成音

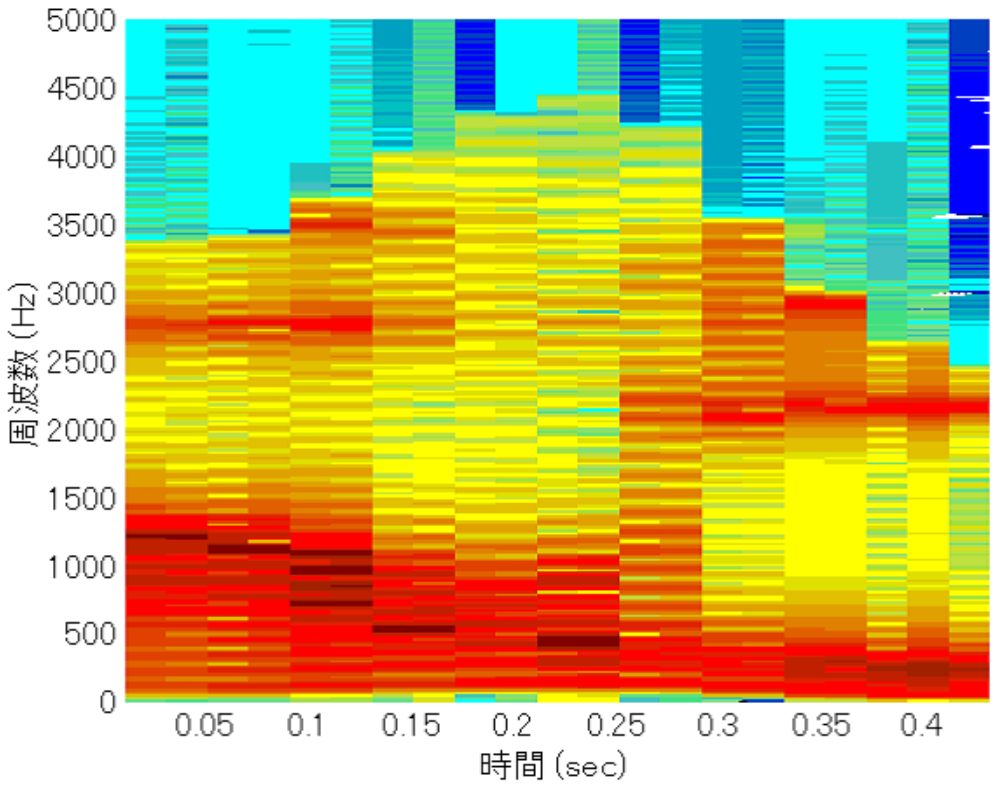

図 5.6: ステップ的に F_0 が変化する合成音のスペクトログラム

連続的に F_0 が変化する調波複合音

ベースとした時変調波複合音は，各フレーム内では F_0 及び調波構造が常に一定であったが，ここでは各フレーム内で F_0 及び調波構造を連続的に変化する信号を考える．具体的には，各フレーム内において，最初に求めた F_0 の一定値をキャリア周波数と置き，オリジナルの音声信号中の F_0 値と F_0 の一定値との差分を変調周波数として周波数変調を実施することで，図 5.7 に示すような F_0 の軌跡を得る．同様に，全ての調波（倍音）に対しても同様に周波数変調を施した上で，これをスペクトログラムで示したものが図 5.8 である．調波としては均一であるが，各フレーム内では全調波が時刻と共に連続的に変化するが，フレーム間は不連続な信号である．

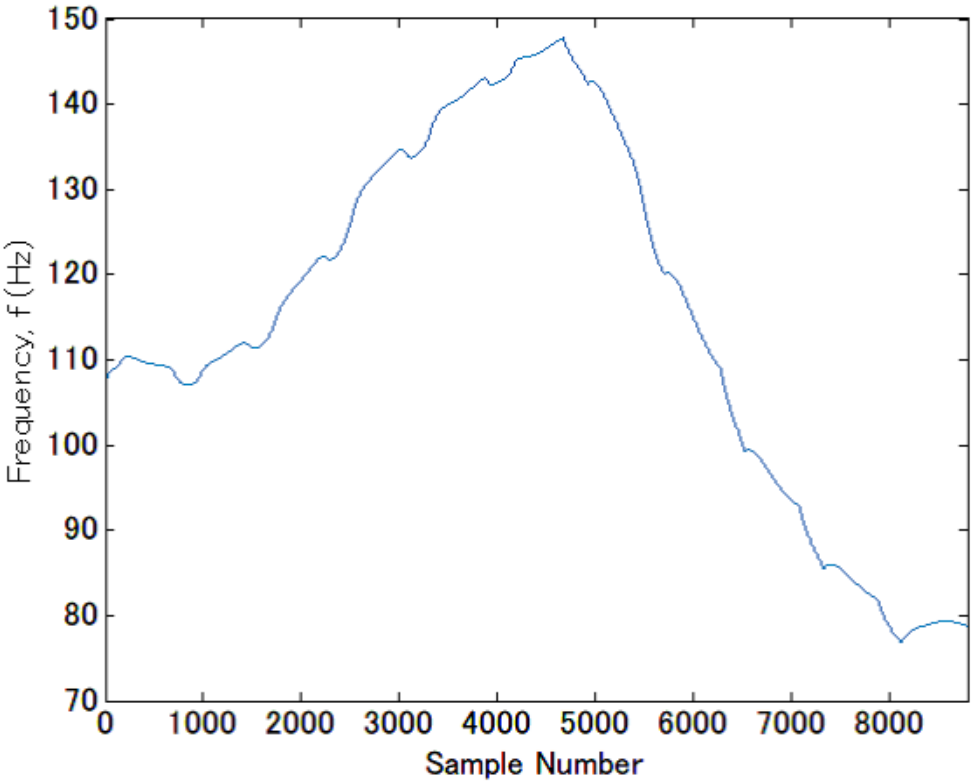

図 5.7: 連続的に F_0 が変化する時変調波複合音の F_0 の軌跡

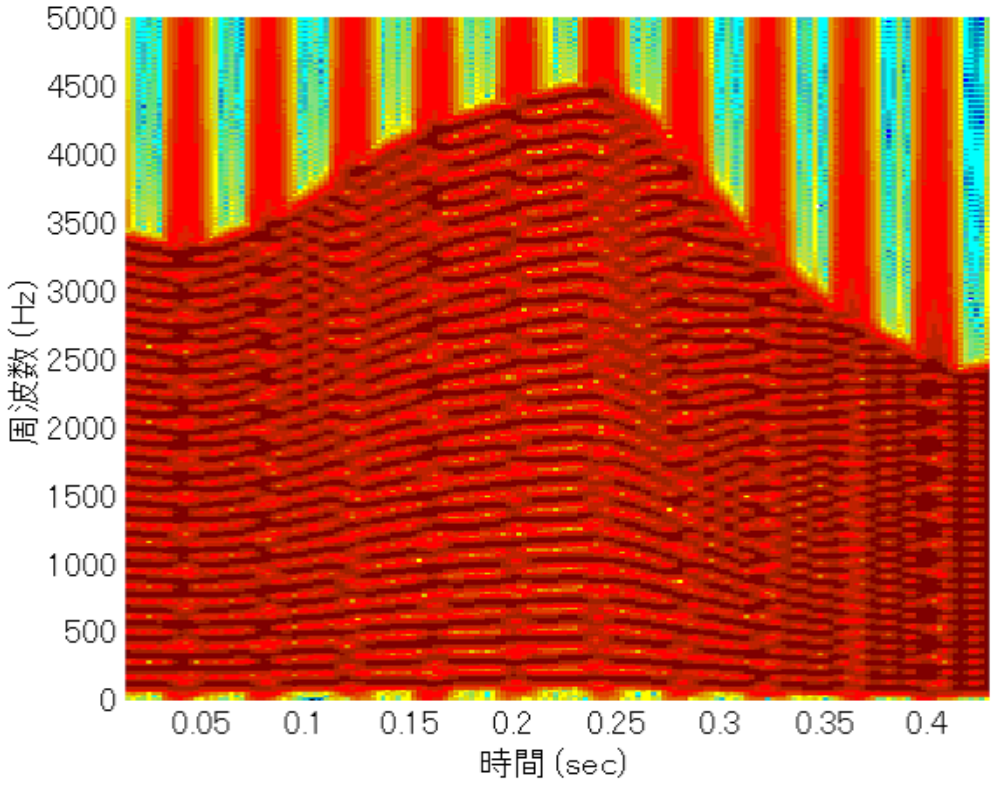

図 5.8: 連続的に F_0 が変化する時変調波複合音のスペクトログラム

連続的に F_0 が変化する合成音

ここでは，前述の時変調波合成音に，音声信号と同様な調波構造を反映することを考える．図 5.9(a) は音声信号/aoi/の一区間の周波数構造であるが，このスペクトラム包絡を LPC 分析により求めると図 5.9(b) のようになる．このスペクトラム包絡情報を，合成フィルタにより図 5.9(c) の調波複合音に適用すると，図 5.9(d) のような合成音が得られる．これをスペクトログラムで示したものが図 5.10 であり，スペクトラム包絡情報が反映された調波が時刻と共に連続的に変化するものとなっている．

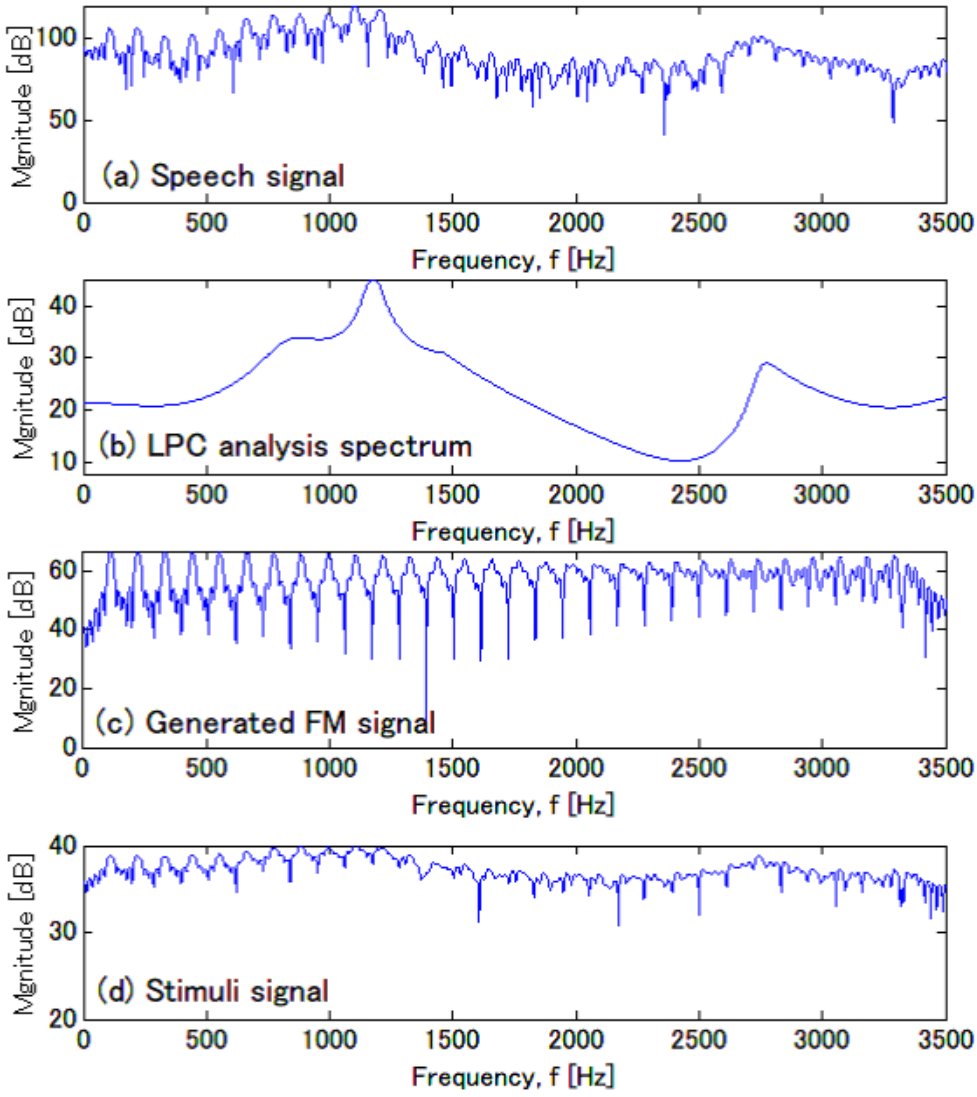

図 5.9: 連続的に F_0 が変化する合成音の作成：(a) 音声信号，(b) 音声信号のスペクトル包絡，(c) 調波複合音，(d) 合成音

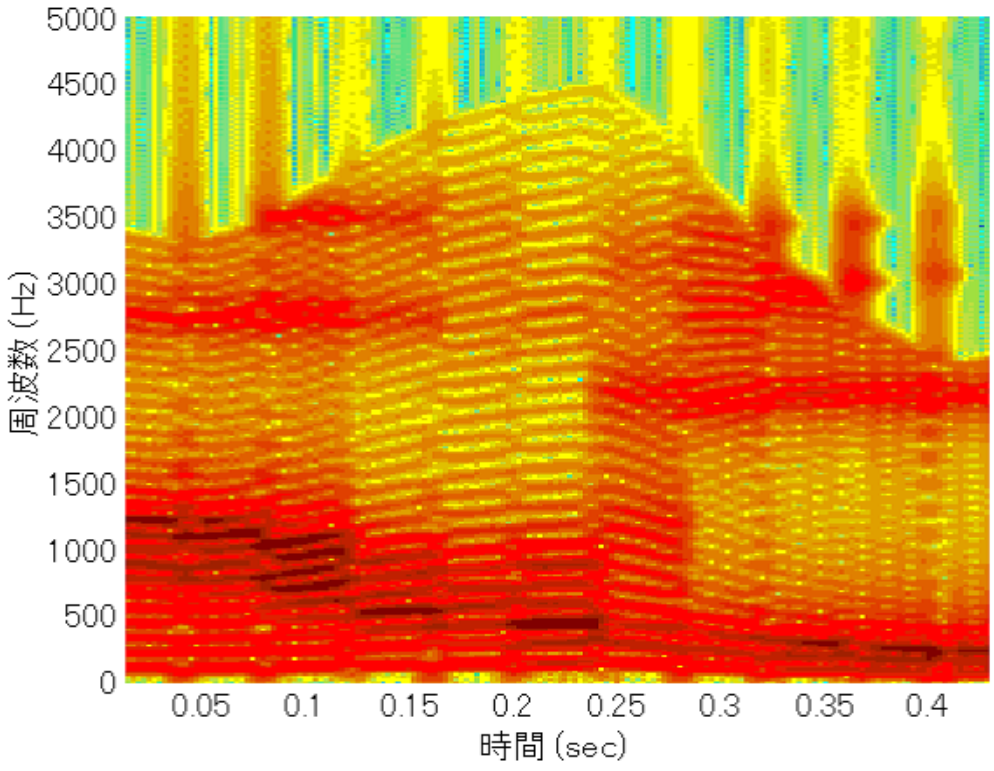

図 5.10: 連続的に F_0 が変化する合成音のスペクトログラム

実音声信号

以上の 4 種類の人工音に加え，作成した人工音のベースとした実音声信号も評価に用いることとする．男声/aoi/の実音声信号のスペクトログラムを図 5.11 に示す．

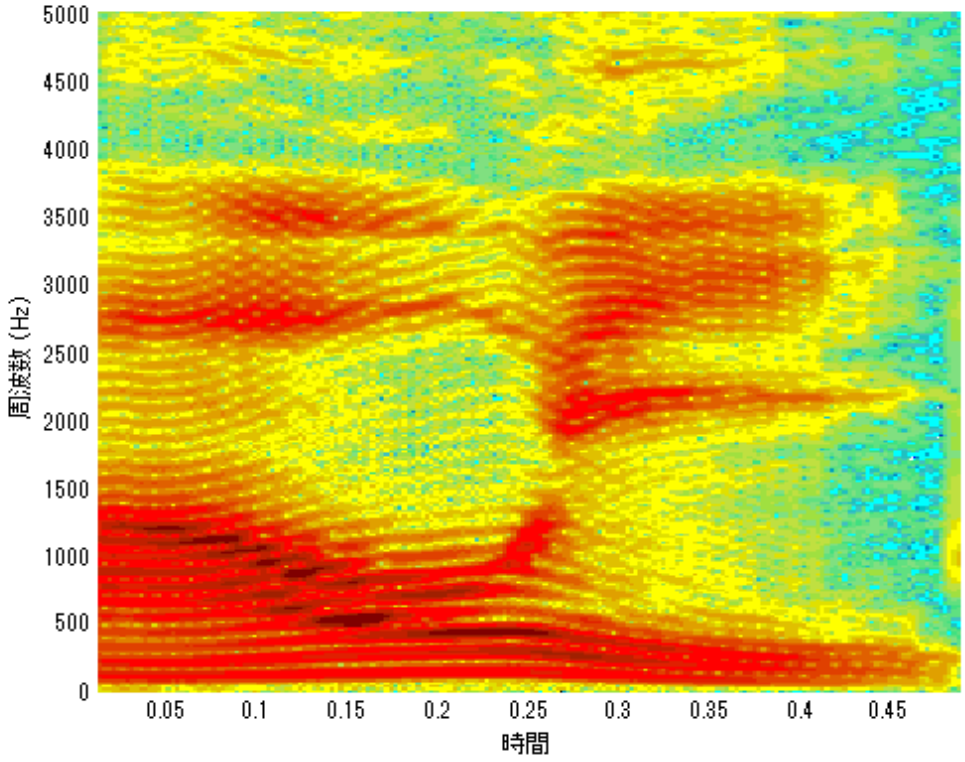

図 5.11: 実音声信号（男声）のスペクトログラム

5.2 評価結果

5.2.1 評価条件

今回は，提案法の分析窓長は 100 msec（シフト長 10 msec）とした上で，静音環境，雑音環境，残響環境，雑音残響環境を設定し，前述の信号（男声ベース）に加え，女声ベースの各種信号に対する提案法の F0 推定精度を調べる形で評価を実施した．なお，提案法による F0 推定の実行にあたっては，直前のフレームの推定値を参酌する処理を加えた上で実行することとした．

雑音環境の背景雑音は 10 種類の白色雑音とし，SNR は，20，10，0，−5，−10 dB の 5 種類とし，計 50 回試験を行った．残響環境については，入力信号に対し 10 種類の白色雑音を基に生成した統計的室内インパルス応答（Schroeder のインパルス応答 [58]）を畳み込むことによって実現したものを用い，残響時間 T_R については 0.1，0.3，0.5，1.0，2.0 sec の 5 種類とした．こちらについても計 50 回試験を行った．雑音残響環境については，5 種類の白色雑音と人工インパルス応答による 5 種類の残響時間とを用いて，25 通りの雑音残響環境を生成し，それぞれ 10 回ずつ，計 250 回試験を行った．

比較対象は，代表的な従来法から 6 手法（TEMPO [53]，YIN [27]，PHIA [57]，SWIPE[43, 44]，複素ケプストラム法（CmpCep）[51]，短時間フーリエ変換法 [36]）を採用した．基本的に従来法のパラメータ設定については，その性能が最も発揮できると考えられるデフォルト値を利用した．

評価指標としては，今回は許容誤差率を 10 %以内とした正答率 [%][56] に加え，Fine pitch error[80] と Gross pitch error[27] を用いた．以下，試験結果について主に正答率を用いて考察を行う．（試験結果のうち Fine pitch error と Gross pitch error については付録に示す．）

5.2.2 ステップ的に F_0 が変化する調波複合音

図 5.4 に示す入力信号を用いて試験を行った結果を，図 5.12 に正答率（許容誤差 10%）で示す．提案法（図 5.12(g)）は，雑音環境では比較的安定しているものの，残響が加わるとその耐性は大きく低下し，従来法に対する優位性を確認するこ

とは出来なかった．前章における試験との違いは，分析窓長が 250 msec から 100 msec に変更されている点にあるが，外乱に来する耐性を維持するには現状では長時間の分析窓を必要とすることが明らかとなった．

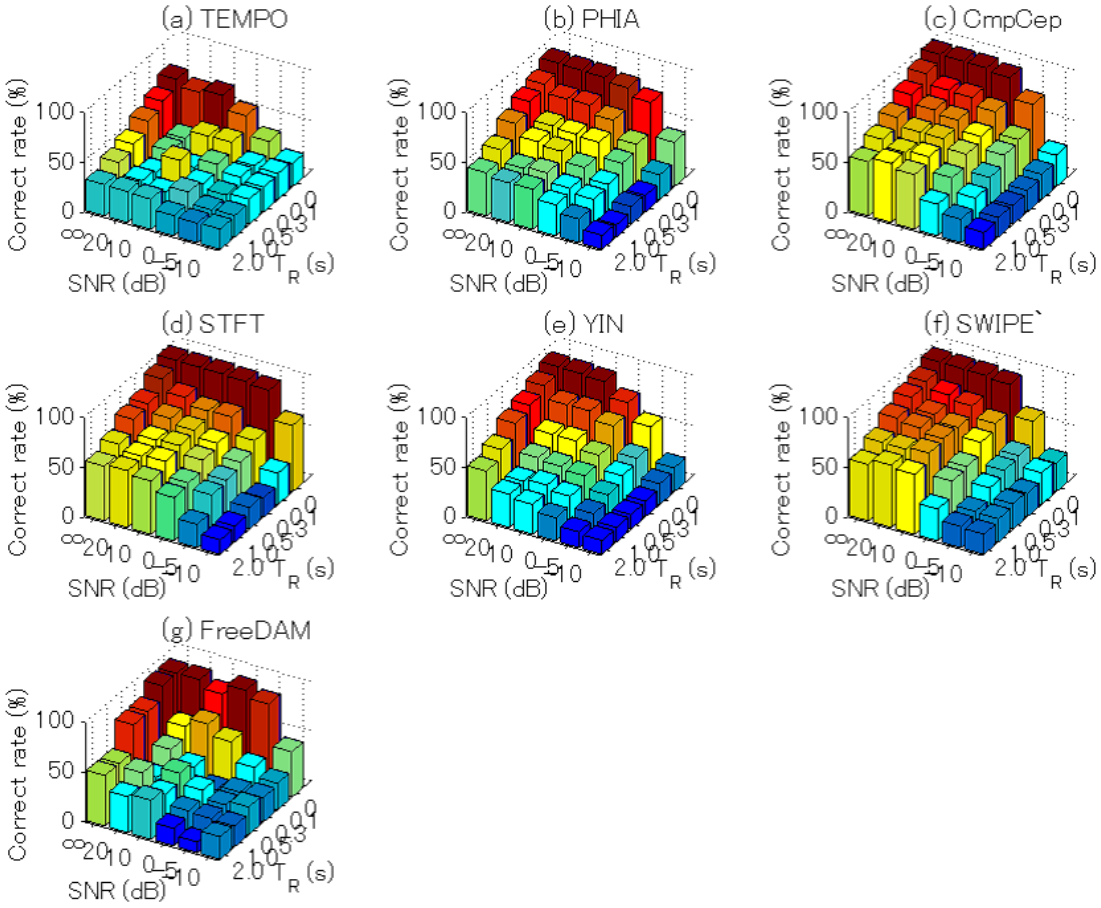

図 5.12: 雑音残響環境における調波複合音（ステップ）の正答率：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

5.2.3　ステップ的に F_0 が変化する合成音

ここでは，フォルマント周波数が提案法に与える影響を確認する．

図 5.6 に示す入力信号を用いて試験を行った結果を，図 5.13 に正答率（許容誤差 10%）で示す．提案法（図 5.13(g)）は，静音環境には対応できるものの，雑音環境，残響環境，雑音残響環境では耐性が低下する結果となった．このことは，フォルマント周波数の影響を受けて各調波の振幅値が不均一となった信号に対しては，提案法の外乱に対する耐性が低下することを示している．

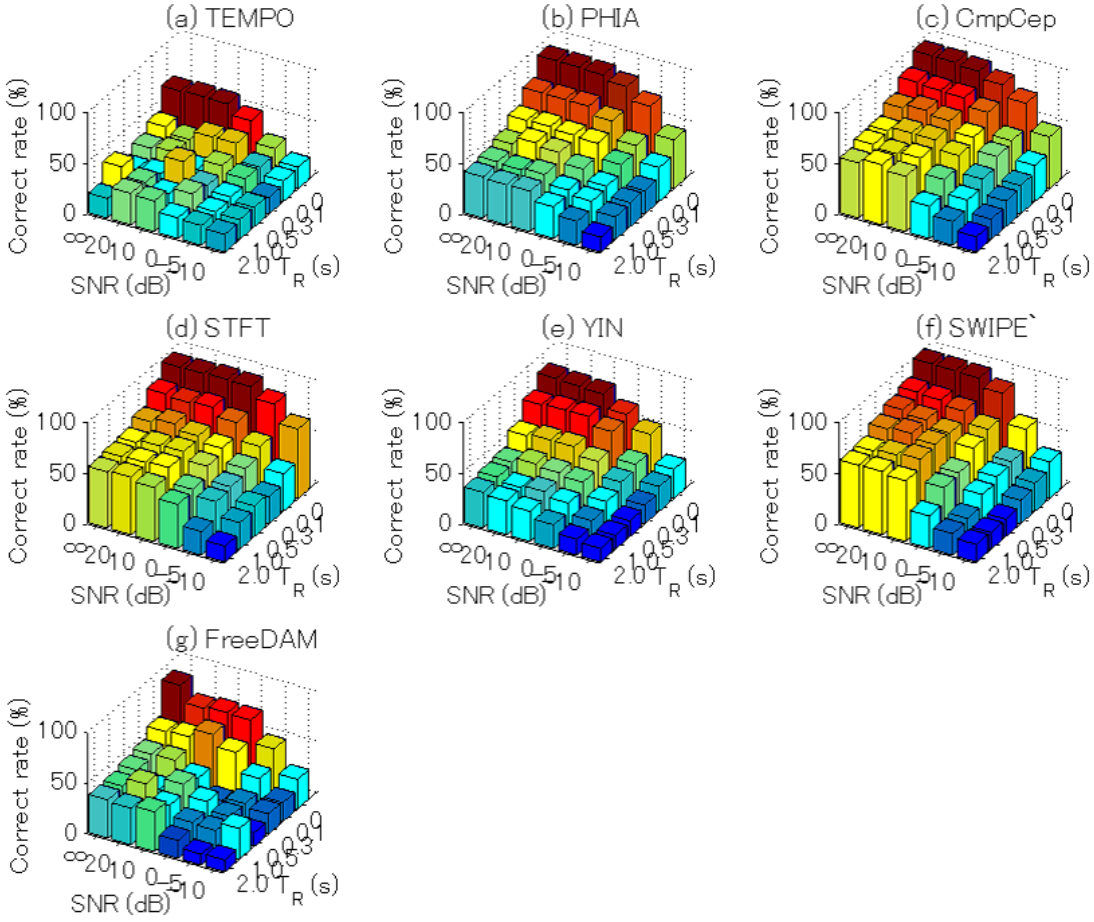

図 5.13: 雑音残響環境における調波合成音（ステップ）の正答率：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

5.2.4　連続的に F_0 が変化する調波複合音

ここでは，F_0 が分析フレーム内で変動する信号に対する提案法の挙動を確認する．

図 5.8 に示す入力信号を用いて試験を行った結果を，図 5.14 に正答率（許容誤差 10%）で示す．提案法（図 5.14(g)）は，雑音環境には比較的対応できるものの，残響が加わるとその耐性は大きく低下し，従来法に対する優位性を確認することは出来なかった．このことは，分析フレーム内で F_0 が変動する信号に対しては，提案法の外乱に対する耐性が低下することを示している．

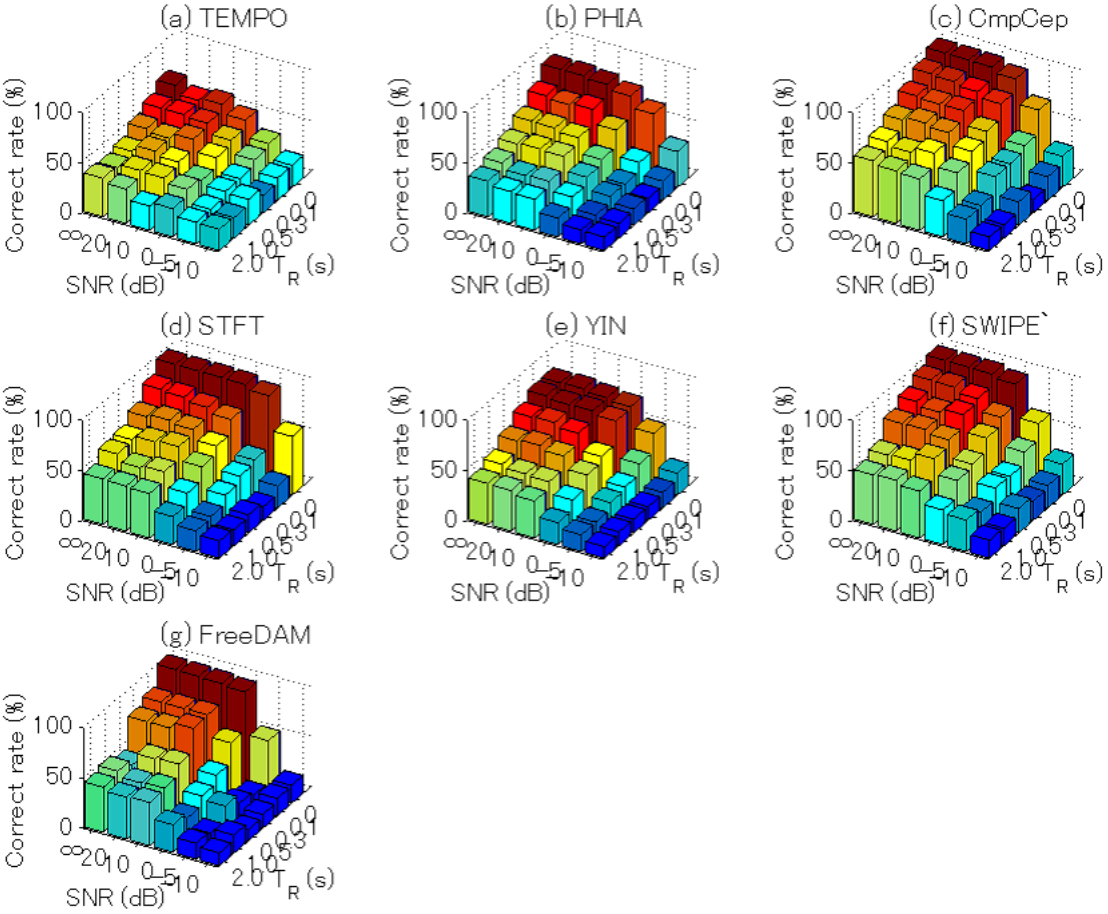

図 5.14: 雑音残響環境における調波複合音（連続）の正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

5.2.5　連続的に F_0 が変化する合成音

ここでは，フォルマント周波数が提案法に与える影響に加え，F_0 が分析フレーム内で変動する信号に対する提案法の挙動を確認する．

図 5.10 に示す入力信号を用いて試験を行った結果を，図 5.15 に正答率（許容誤差 10%）で示す．提案法（図 5.15(g)）は，静音環境には対応できるものの，雑音環境，残響環境，雑音残響環境では耐性が著しく低下する結果となり，従来法に対する優位性を確認することは出来なかった．このことは，フォルマント周波数の影響を受け，なおかつ分析フレーム内で F_0 が変動する信号に対しては，提案法の外乱に対する耐性はほぼ失われることを示している．

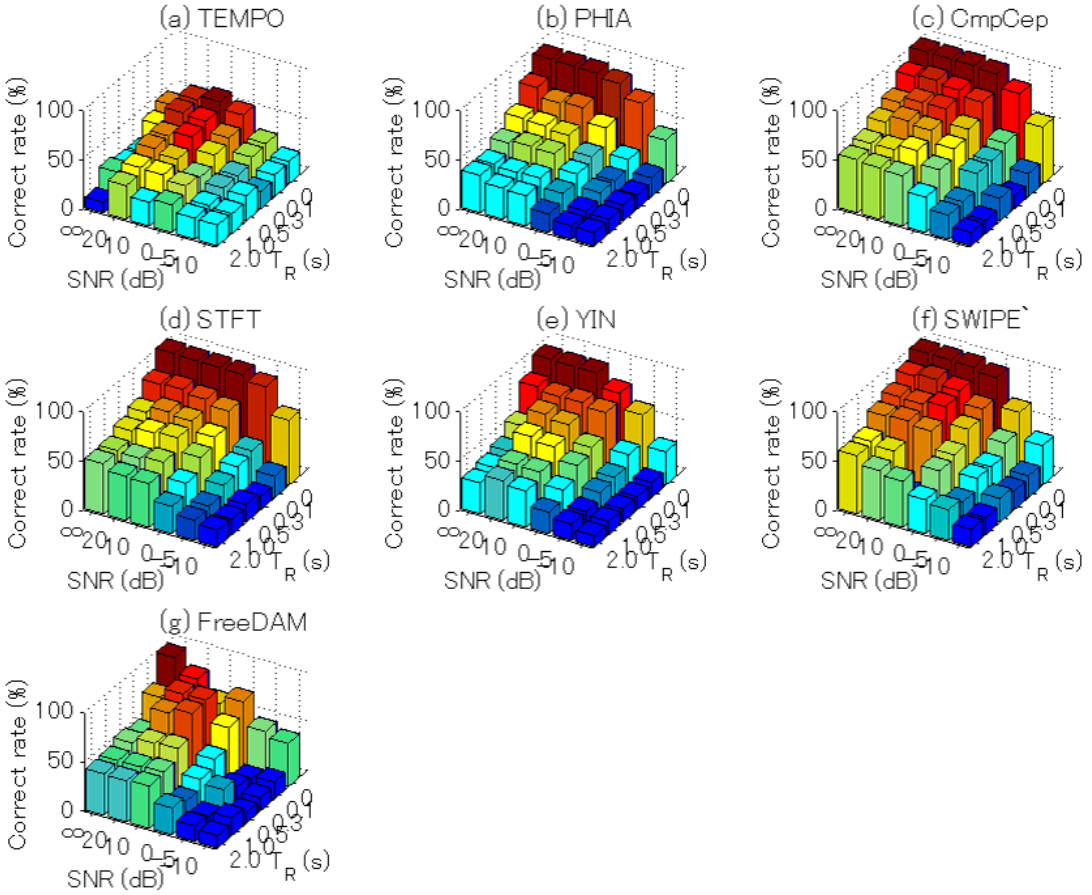

図 5.15: 雑音残響環境における調波合成音（連続）の正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

5.2.6　実音声信号

図 5.11 に示す実音声信号を用いて試験を行った結果を，図 5.16 に正答率（許容誤差 10%）で示す．提案法（図 5.16(g)）は，雑音環境には比較的対応できるものの，残響が加わるとその耐性は低下し，全体としては複素ケプストラム法（図 5.16(c)）, 短時間フーリエ変換（図 5.16(d)）, SWIPE’（図 5.16(f)）とほぼ同等の結果となったが，提案法の優位性を示すまでには至らなかった．

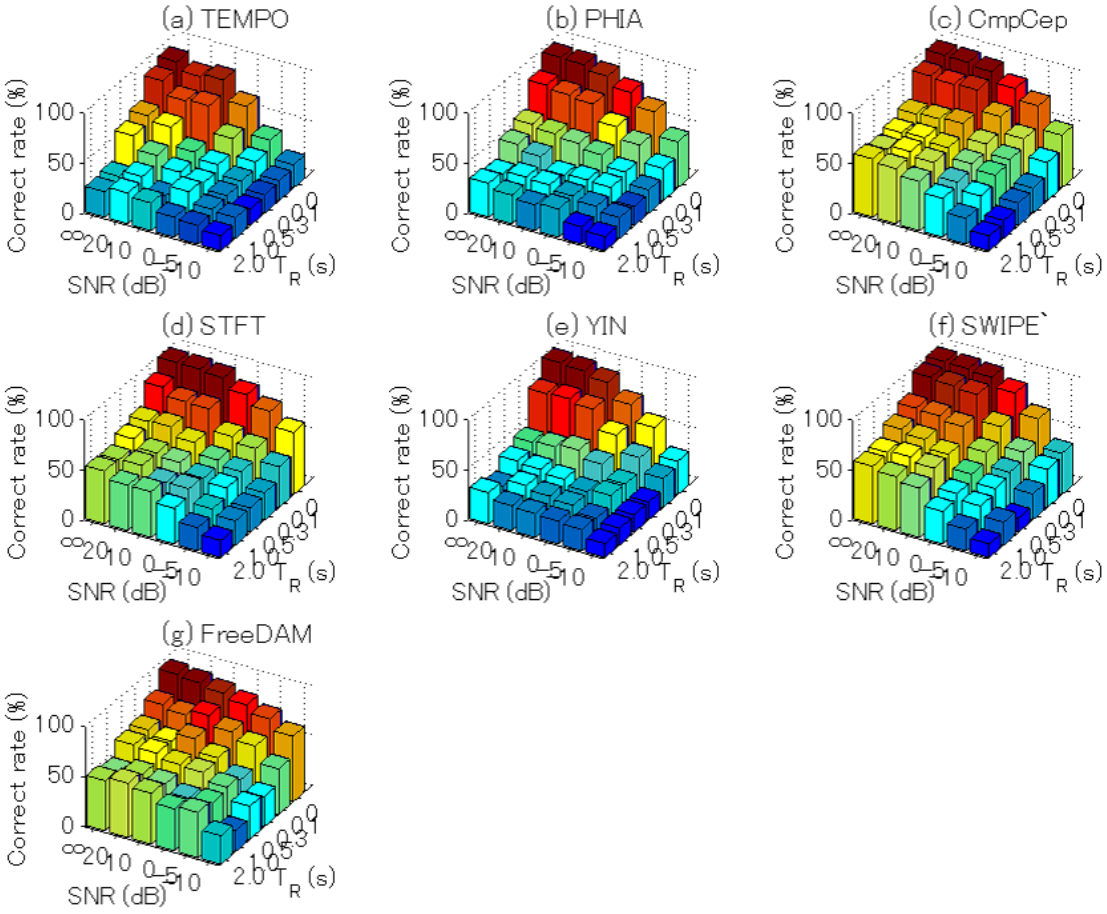

図 5.16: 雑音残響環境における実音声/aoi/の正答率： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’，(g) FreeDAM（Proposed）

5.3 まとめ

音声信号を模擬した人工信号，ならびに実音声信号用いた試験を通じて導き出される提案法の限界は次のようにまとめられる．

まず，提案法は，残響に頑健なF0推定を行うには，現状では長時間の分析フレーム長を必要とする．加えて，F_0 が分析フレーム内で変動する信号の場合は，強雑音のような特に厳しい環境への対応は難しくなる．また，フォルマント周波数の影響を受けた信号，つまり各調波の振幅が不均一な信号に対しては，雑音残響に対する耐性を維持できなくなる．よって，長時間の分析フレームでかつその間は F_0 が変動せず，かつ各調波の振幅がほぼ均一な信号が，現状の提案法で対応可能な信号である．すなわち，前章で考慮した楽器音のような信号が，現状の提案法が扱える信号の限界であると言える．

この背景として，提案法は振幅変調の復調プロセスを利用しており．したがって，分析フレーム内では F_0 が一定であること，また各調波成分の振幅が全て均一であることが前提となっている．一方で音声信号は，その F_0 は時刻と共に連続的に変動する上，その調波成分の振幅はフォルマント周波数の影響により常に不均一である．つまり，上記前提の下で提案法を現状のまま音声信号のF0推定に適用することはやはり難しいと考えられ，今後さらなる改良検討が必要である．

例えば，現在の提案法ではフォルマント周波数への対応措置として，抽出する隣り合う3本の周波数の組み合わせを8通り利用することにより解決を試みているが，より復調に最適な振幅値の揃った調波の組み合わせだけを採用して復調に利用するなどの機構を付加することなどが考えられる．また，提案法の分析フレーム長については現状100 msecとしているが，窓長としては比較的長めであり，フレーム内での F_0 の値に変動幅が生じるため，特に外乱が加わった場合に復調に失敗する原因となる．ただし提案法の場合，原理上は復調に利用する調波の場所は任意であることから，高い周波数領域の調波を重点的に活用することで，短い分析フレーム長を採ることが可能となると考えられる．このように，活用する周波数領域に応じて適切な分析窓長を適用する機構を付加することで，動作の安定化と推定精度の向上が望めると考えられる．加えて，分析区間内でのキャリア周波数や変調周波数の変動が無いことを前提としている振幅変調の復調処理だけで対

応するのには根本的な限界があり，キャリア周波数の変動を許容するような復調方式，例えば周波数変調（FM) の適用なども併せて考えていく必要がある．

第 6 章

結論

6.1 本研究で明らかにしたこと

本研究では，ヒトのピッチ知覚に原点回帰することで，雑音や残響に頑健なF0推定法の検討を行った．具体的には，ヒトがF0が欠落した信号からでもピッチを知覚できるという点にまず着目し，加えてヒトのAM音のピッチ知覚からもヒントを得て，振幅変調の復調技術を用いてヒトのピッチ知覚を模擬するアプローチを採った．このように本研究では，信号のAM成分に着目して振幅変調の復調技術を応用するという今までに無いやり方で，雑音や残響に頑健なF0推定法が構築可能であることを示した．評価試験の結果からも，調波複合音や楽器音のような理想的な調波信号に対して，提案法が従来法を上回る頑健性を備えていることが明らかとなった．

一方で，提案法を音声信号に適用させるためには，頑健性の点でいくつかの課題を残していることも明らかとなった．ただし，これら課題に対しては，有効と考えられる対応策も複数存在している．

本研究の全体を通して，雑音残響環境での基本周波数推定に関し，当初の予測どおり，信号中のAM成分に着目するという方法の有効性が明らかとなった．このことにより，半世紀以上もの長きにわたって解決が困難であった，雑音残響に頑健なF0推定法を考える上で，ひとつの方向性を打ち出すことが出来た．提案法自体にはまだ課題が存在するものの，提案法のエッセンスは，雑音残響に頑健なF0推定法を指し示すものとなった．

6.2 残された課題

次に，提案法の残された課題について述べる．

まず頑健性については，現提案法は，理想的な調波信号に対しては頑健に適応できるものの，音声信号には対応できていない．これら音声信号に対する頑健性の課題に対しては，復調に適切な調波を選択的に利用する機構や，周波数領域に応じて適切な分析窓長を選択する機構，FM復調の適用等が有効と考えられる．

頑健性のその次に考えるべき課題に，即時性の問題がある．提案法は本質的に検索処理とともに検波のための位相同期処理を伴うため，その計算量は大きい．現在の実装では原理レベルで$O(N^2)$の計算量となっており，リアルタイム処理はま

ず不可能である．したがって，実用化を考える際には，アルゴリズムの改良により計算量の大幅な低減を図るとともに，処理の軽い検波方式も併せて検討していく必要がある．

また，今回の評価シミュレーションでは，白色雑音による雑音環境と，白色雑音を利用したインパルス応答による残響環境とを利用しているが，これらはいずれも人工的に生成された外乱である．提案法が実環境で使えることを主張していくためには，実雑音，実残響を用いた評価試験が必須であり，評価のメニューを再検討していく必要がある．

さらに将来的な課題として，現在の提案法には，基本周波数推定の機能以外は無く，音声が存在するか否かの判定のための機能（VAD) や，有声区間／無声区間の判定（UV 判定）のための機能は含まれていない．これら機能を提案法に取り込んで初めて，実環境における F0 推定法が確立できることから，これら付加機能の導入も併せて考えていく必要がある．

6.3 展望

最後に，提案法の展望について述べる．

まず提案法は，半世紀以上もの長きにわたって解決が困難であった，雑音残響に頑健な F0 推定法を考える上で，ひとつの方向性を示したものであると言える．現状の提案法自体にはまだ課題が存在するものの，少なくとも理想的な調波信号に対しては非常に頑健な手法であることから，その貢献分野は確実に存在すると考えられる．第 4 章で触れた楽器音の音高推定もその一つであり，例えば残響時間の比較的長い音楽ホールで，かつ観客のノイズが含まれる中でのソロ楽器のライブ録音の音源から，雑音除去や強調処理，音源分離等を実施する際の要素技術として，この方法が貢献できる可能性がある．また，提案法は任意の調波の組から F0 を推定できることから，何らかの理由により F0 成分を含む低域が欠落した楽音から F0 を推定・復元・強調するための，音楽編集ツールの一機能としての活用も考えられる．

さらには，他の F0 推定法の一部機能として提案法を実装することにより，他の F0 推定法を頑健性の面で補完するような役割も大いに期待できる．具体的には，

他の F0 推定法の前処理/後処理部分に提案法を配置することで，雑音や残響に対する耐性を向上させることなどが考えられる．

このように，提案法のユニークなエッセンスは，将来の音声/音楽情報処理の発展に大いに貢献するものとなることが期待できる．

謝辞

本研究を進めるにあたり，不肖の弟子に対し終始熱意あるご指導を賜った，北陸先端科学技術大学院大学 鵜木祐史 教授に深く感謝いたします.

また，本研究の要所要所で貴重なご助言を賜った，北陸先端科学技術大学院大学 赤木正人 教授，党 建武 教授，そして長谷川忍 准教授に，心より感謝いたします.

さらに，社会人大学院生として勉学を進める中で，数多くの知見を賜った情報通信研究機構の戸室知二氏に厚くお礼申し上げます.

参考文献

[1] 大串健吾, “音のピッチ知覚,” 音響サイエンスシリーズ１５ , コロナ社, 2016.

[2] 森周司, 香田徹, “音の高さのモデル,” 音響サイエンスシリーズ 9 聴覚モデル, 森周司, 香田徹（編）, 第 1 章, コロナ社, 2011.

[3] ISO 16:1975, “Acoustics – Standard tuning frequency (Standard musical pitch),” International Organization for Standardization, 1975.

[4] 斎藤収三, 中田和男, “音声情報処理の基礎,” オーム社, 1981.

[5] 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,” 信学技報, SP, 音声 102(35), 37–42, 2002.

[6] 北脇信彦, “音のコミュニケーション工学,” 音響テクノロジーシリーズ１ , コロナ社, 1996.

[7] 鈴木久喜, “ピッチ抽出の今昔,” 日本音響学会誌, Vol. 56, No. 2, pp121–128, 2000.

[8] 北原鉄朗, “2-1 音響特徴抽出,” 電子情報通信学会知識ベース，2 群-9 編-2 章, 2010.

[9] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, “Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 654–669, April 2015.

[10] 後藤真?, “音楽音響信号を対象としたメロディーとベースの音高推定,”信学論 (D), vol. J84-D-II, no. 1, pp. 12–22, Jan. 2001.

[11] 三輪多恵子, 田所嘉昭, 斎藤努, “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定, ” 信学論 (D), vol. J81-D-II, no. 9, pp. 1965–1974, Sept. 1998.

[12] Kinoshita, K., Nakatani, T., and Miyoshi, M., “Harmonicity based dereverberation for improving automatic speech recognition performance and speech intelligibility,” IEICE Trans. Fundamentals, E88-A(7), pp.1724–1731, 2005.

[13] 中谷智広, 三好正人, 木下慶介, “調波構造に基づくモノラル音声信号のブラインド残響除去,” 信学論, vol. J88-D-II, no. 3, pp. 509–520, Mar., 2005.

[14] 中谷智広, 奥乃博, 川端豪, “マルチエージェントによる音環境理解のための音響ストリーム分離,” 人工知能誌, vol. 10, no. 2, pp. 232–241, Mar. 1995.

[15] 鵜木祐史, 赤木正人, “聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法,” 信学論, vol. J82-A, no. 10, pp. 1497–1507, Oct., 1999.

[16] Cooke, M., Ellis, D. P. W., “The auditory organization of speech and other sources in listeners and computational models,” Speech Communication, vol.35, pp.141–177, 2001.

[17] 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, July 1978.

[18] 鵜木祐史, 石本祐一, 赤木正人, “残響音声からの基本周波数推定に関する検討,” JAIST Research Report, IS-RR-2005-007, March 2005.

[19] 古井貞熙, “ディジタル音声処理,” 東海大学出版会, 1985.

[20] Hess, W. J., “Pitch Determination of Speech Signals: Algorithms and Devices,” Springer-Verlag, New York, 1983.

[21] Hess, W. J., “Pitch and Voicing Determination,” in Adcanced in speech signal processing, ed. S. Furui and M.M. Sondhi, pp. 3-48, Marcel Dekker. Inc. New York 1992.

[22] Cheveigné, A., Kawahara, H., “Comparative evaluation of F0 estimation algorithms,” Proc. Interspeech2001, pp. 2451–2454, 2001.

[23] Gold, B. and Rabiner, L., “Parallel processing techniques for estiming pitch periods of speech in the time domain,” J. Acoust. Soc. Am., vol. 46, no. 2, pp. 442-448, Aug. 1969.

[24] N. C. Geckinli, D. Yavuz, “Algorithm for pitch extraction using zero-crossing interval sequence.” IEEE Trans. Acoust., Speech and Signal Process., vol. 25, no. 6, 1977.

[25] Hess, W. J. “Pitch and Voicing Determination,” in Adcanced in speech signal processing, ed. S. Furui and M.M. Sondhi, pp. 3-48, Marcel Dekker. Inc. New York 1992.

[26] Takagi,T., Seiyama, N., and Miyasaka, E., “A Method for pitch extraction of speech signal using autocorrelation functions through multiple window-length,” IEICE vol. J80-A, no. 9, pp. 1341-1350, Sept. 1997 (in Japanese).

[27] Cheveigné, A., Kawahara, H., “Yin, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Am., vol. 111, no. 4, pp. 1917–1930, 2002.

[28] Mouch, M. and Dixon, S., “PYIN: A fundamental frequency estimator using probabilistic threshold distributions”, in Proc. ICASSP2014, pp.659–1652. 2008.

[29] Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J., “Average magnitude difference function pitch extractro,” IEEE Trans. Acoust., Speech, Signal Process. ASSP-22, pp. 353-361, 1974.

[30] 大村浩, 田中和世, “基本波フィルタリング法による精細ピッチパターンの抽出,” 音響誌，vol. 51, no. 7, pp. 509–518, 1995.

[31] 森勢将雅, 河原英紀, 西浦敬信, “基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法,” 信学論 (D)，vol.J93-D, no. 2, pp. 109–117, 2010.

[32] 佐宗晃, 中村向五, “ウェーブレット変換を用いたピッチ抽出法の一方法,” 信学論 (A)，vol.J80-A, no. 11, pp. 1848–1856, Nov. 1997.

[33] B. S. Atal, S. L. Hanauer,”Speech analysis and synthesis by linear prediction of the speech wave,” J. Acoust. Soc. Am., vol. 50, no. 2, pp. 637–655, 1971.

[34] Markel, J., “ The SIFT algorithm for fundamental frequency estimation, ” IEEE Trans. Audio vol. AU-20, pp. 367-377, 1972.

[35] 金城竜彦, 舟木慶一, “複素 LPC 音声分析を用いたロバスト F0 推定, ” 信学技報 SP2006-36, 2006.

[36] M. Lahat, R.J. Niederjohn, and D.A. Krubsack, “A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech, ” IEEE Trans. Acoust., Speech, Signal Process, ASSP-35, no. 6, pp. 741–750, June 1987.

[37] 国枝伸行, 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, ” 信学論 (A), vol. J80-A, no. 3, pp. 435–443, March 1997.

[38] Nishi, K. and Ando, S., “An optimal comb filter for timevarying harmonics extraction,” IEICE Trans. Fundamentals, vol. E81-A, no. 8, pp. 1622-1627, Aug. 1998.

[39] Hermes, D. J., “ Measurement of pitch by subharmonic summation, ” J. Acoust. Soc. Am., vol. 83, no. 1, pp. 257-264, Jan. 1988.

[40] Singer, H. and Sagayama, S.,“ Pitch dependent phone modelling for HMM-based speech recognition, ” J. Acoust. Soc. Jpn. (E) vol 15, pp. 77-86, 1994.

[41] Suzuki, H.,“A story of old-and news of pitch extracton in speech technology, ” J. Acoust. Soc. Jpn. vol. 56, no. 2 pp. 121-128, 2000.

[42] 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,” 信学論（A), vol. J86-A, no. 11, pp. 1097–1107, Nov. 2003.

[43] A. Camacho, “SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech And Music,” Ph.D. Thesis, University of Florida, 2007.

[44] Camacho, A., Harris, J. G. “A sawtooth waveform inspired pitch estimator for speech and music,” J. Acoust. Soc. Am., vol.124, no. 3, pp. 1638–1652, 2008.

[45] 森勢将雅, “2-2 基本周波数推定（歌声研究に関する視点から）,” 電子情報通信学会知識ベース，2 群-9 編-2 章, 2010.

[46] Noll, A. M.“Short-time spectrum and“cepstrum”techniques for vocal-pitch detection,” J. Acoust. Soc. Am., vol. 36, no. 2, pp. 226-302, Feb. 1964.

[47] Noll, A. M.“Cepstrum pitch determination,”J. Acoust. Soc. Am., vol. 41, no. 2, pp. 293-309, Aug. 1966.

[48] Noll, A. M.“Clipstrum pitch determination,”J. Acoust. Soc. Am., vol. 44, no. 6, pp. 1585-1591, July. 1968.

[49] Oppenheim, A. V.“Speech analysis-synthesis system based on homomorphic filtering,” J. Acoust. Soc. Am., vol. 45, no. 2, pp. 458-465, 1969.

[50] 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,” 信学論（A), vol. J82-A, no. 7, pp. 1115–1122, July 1999.

[51] Unoki, M., Hosorogiya, T., and Ishimoto, Y., “Comparative evaluations of robust and accurate F0 estimates in reverberant environments,” Proc. ICASSP2008, pp. 4569–4572, 2007.

[52] Unoki, M., Hosorogiya, T., “Estimation of fundamental frequency of reverberant speech by utilizing complex cepstrum analysis,” Journal of Signal Processing, vol. 12, no. 1, pp. 31-44, Jan. 2008.

[53] Kawahara, H., Masuda-Katsuse, I. and Cheveigné, A., “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an

instantaneous-frequency-based F0 extraction: Possible role of a reptitive structure in sounds," Speech Communication, vol. 27, pp.187-207, 1999.

[54] Kawahara, H., Katayose, H., Cheveigné, A., and Patterson, R. D., "Fixed Point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," Proc. Eurospeech'99, vol. 6, pp. 2781–2784, 1999.

[55] H. Kawahara, Y. Agiomyrgiannakis, and H. Zen, "Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis," in 9th ISCA Workshop on Speech Synthesis, 2016.

[56] 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," 信学論（D), vol. J83-D-2, no. 11, pp. 2077–2086, 2000.

[57] Ishimoto, Y., Unoki, M., and Akagi, M., "A Fundamental Frequency Estimation Method for Noisy Speech Based on Instantaneous Amplitude and Frequency," Proc. EuroSpeech2001, pp. 2439–2442, 2001.

[58] Schroeder, M. R., "Modulation Transfer Functions: Definition and Measurement," Acustica, Vol. 49, pp. 179–182, 1981.

[59] Andrew J. Oxenham, "Pitch Perception," The Journal of Neuroscience, vol. 32 no. 39, pp. 13335–13338 Sept. 2012.

[60] Mark Sayles and Ian M. Winter, "Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds," Neuron 58, 789–801, June 12, 2008.

[61] Kate Helms Tillery, Christopher A. Brown, and Sid P. Bacon, "Comparing the effect of reverberation and of noise on speech recognistion in smiulated electric-acoustic listening," J. Acoust. Soc. Am., vol. 131, no. 1, pp. 416-423, 2012.

[62] Marianne van Zyl, Johan J. Hanekom, “Speech perception in noise: A comparison between sentence and prosody recognition,” J. Hearing Science, vol. 1, no. 2, pp. 54-56, 2011.

[63] Bregman, A. S., “Auditory Scene Analysis: The Perceptual Organization of Sound.,” MIT Press, Cambridge, MA, 1990.

[64] Robert J. Zatorre, “Pitch perception of complex tones and human temporal-lobe function,” J. Acoust. Soc. Am. 84 (2). August 1988.

[65] Meddis, R., Hewitt, M. J., “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. 1: pitch identification,” J. Acoust. Soc. Am., vol. 89, no. 6, pp. 2866–2822, 1991.

[66] Schouten, J.F., “The Residue, a new Component in Subjective Sound Analysis,” Proc. Koninkl. Ned. Akad. Wetenschap. vol. 43, pp. 356-365, 1940.

[67] 鵜木祐史, 山崎悠, 赤木正人, “雑音残響環境下における MTF ベース・パワーエンベロープ回復処理の検討,” 日本音響学会春季講演論文集, pp.853–856, 2010.

[68] Unoki, M., Akagi, M., “A method of signal extraction from noisy signal based on auditory scene analysis,” Speech Communication, vol. 27, pp. 261–279, April 1999.

[69] 藤井義巳, “ソフトウェア無線（SDR）技術の最新動向と将来展望,” ITU ジャーナル, vol. 47, no. 11, pp. 17–21, 日本 ITU 協会 Nov. 2007.

[70] 谷本洋, “ダイレクトコンバージョン受信機用ミクサの研究開発動向,” 信学論（C), vol. J84-C, no. 5, pp. 337–348, May 2001.

[71] 三輪賢一郎, “雑音残響音声の基本周波数推定の基礎的検討,” 北陸先端科学技術大学院大学　情報科学研究科　修士論文, Sep. 2013.

[72] 鵜木祐史, “変調伝達関数に基づく音声信号処理 (1)　パワーエンベロープ逆フィルタ処理の原理とその応用について,” 信号処理, 12(5), pp. 339-348, 信号処理学会, 2008.

[73] Rico Petrick, Masashi Unoki, Anish Mittal, Calros Segura, and Ruediger Hoffmann, “A Comprehensive Study on the Effects of Room Reverberation on Fundamental Frequency Estimation,” Proc. Interspeech2008, pp. 131-134, Brisbane, Australia, Sept. 2008.

[74] 柳沢浩一, 田中京子, 山浦逸雄, “ディジタル補聴器のためのスペクトル調波構造に基づく連続音声に対する雑音抑圧処理,” 電学論 C, vol. 120-C, no. 3, pp. 382–389, 2000.

[75] 飯田一博, “音響工学基礎論,” コロナ社, 2012.

[76] 浅沼克紀, 大西正輝, 小島篤博, 福永邦雄, “色情報と領域追跡情報を用いた人物の顔と手の領域の抽出,” 電学論 C, vol. 119-C, no. 11, pp. 1351–1358, 1999.

[77] Daichi Kitamura, “Open Daraset,” http://d-kitamura.net/dataset.htm, (2018-08-28 閲覧)

[78] 板倉文忠, “音声の線形予測分析,”信学技報 SP98-55, Sep. 1998.

[79] 青木直史, “ディジタル・サウンド処理入門,” CQ 出版社, 2006.

[80] 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 acoustic, speech, and signal processing, vol.ASSP-24, no.5, pp.399-418, 1976.

[81] Un, C. K., and Yang, S. C. “A pitch extraction algorithm based on LPC inverse filtering and AMDF,”IEEE Trans. Acoust., Speech, Signal Process, ASSP-25, pp. 565-572, 1977.

本研究に関する研究業績

論文

[1] 三輪賢一郎, 鵜木祐史,“振幅変調音のピッチ知覚に基づいた調波複合音の基本周波数推定法,” 電子情報通信学会論文誌A, Vol. J98-A, No.12, pp. 668-679, 2015 年 12 月.

国際会議

[2] Miwa, K., Unoki, M., “Study on method for estimating F0 of steady complex tone in noisy reverberant environments,” Proc. IIHMSP2013, pp. 456–459, Oct. 2013.

[3] Miwa, K., Unoki, M., “Robust method for estimating F0 of complex tone based on pitch perception of amplitude modulated signal,” Proc. INTERSPEECH 2017, pp. 2311–2315, Aug. 2017.

研究会

[4] 三輪賢一郎, 鵜木祐史，“振幅変調音のピッチ知覚に基づいた基本周波数推定法の検討,” 日本音響学会聴覚研究会資料, vol. 45, no. 8, H-2015-112, 甲州市勝沼 ぶどうの丘，Nov. 2015.

口頭発表

[5] 三輪賢一郎, 鵜木祐史,“信号の AM 成分に着目した雑音残響にロバストな基本周波数推定法の検討,” 日本音響学会 2013 年度秋季研究発表会講演論文, 1-7-11, pp. 277-278, 豊橋, Sept. 2013.

[6] 三輪賢一郎, 鵜木 祐史, “音声の AM 成分に着目した基本周波数推定法の検討,” 日本音響学会 2014 年度春季研究発表会講演論文, 2-6-10, pp. 315-316, 御茶ノ水, March 2014.

付録 A

各手法の F0 推定精度（第 2 章関連）

ゼロ交差法 [23]

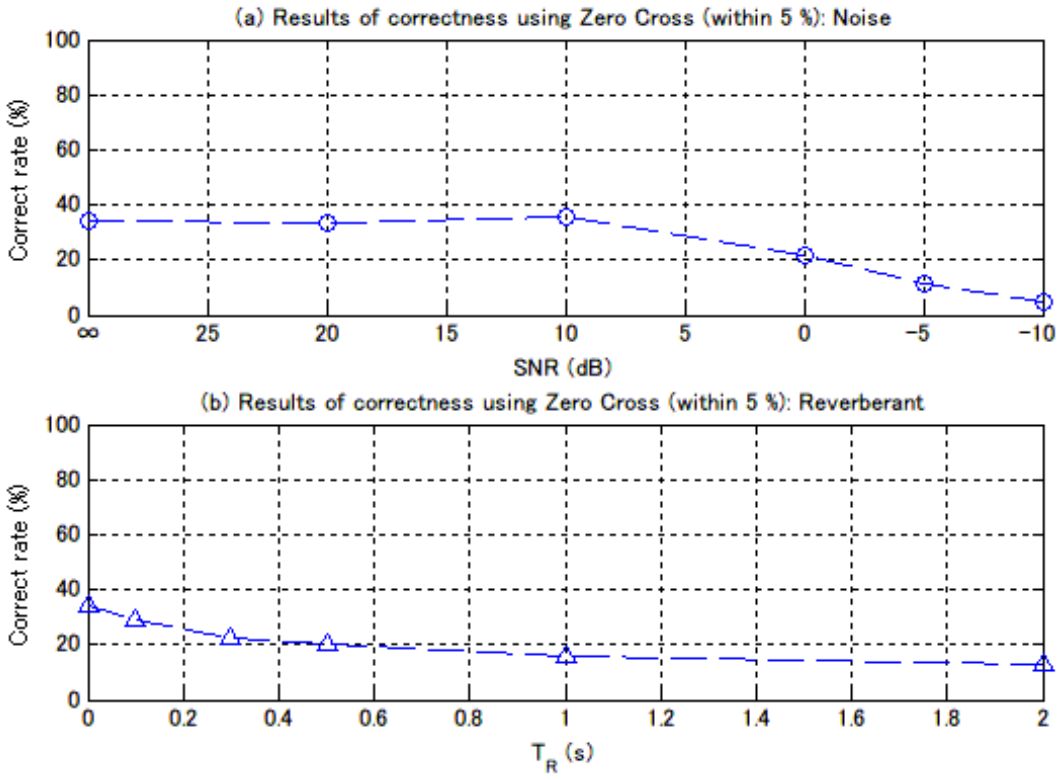

図 A.1: ゼロ交差法の F0 推定性能： (a) 雑音環境，(b) 残響環境

ピーク検出法 [24]

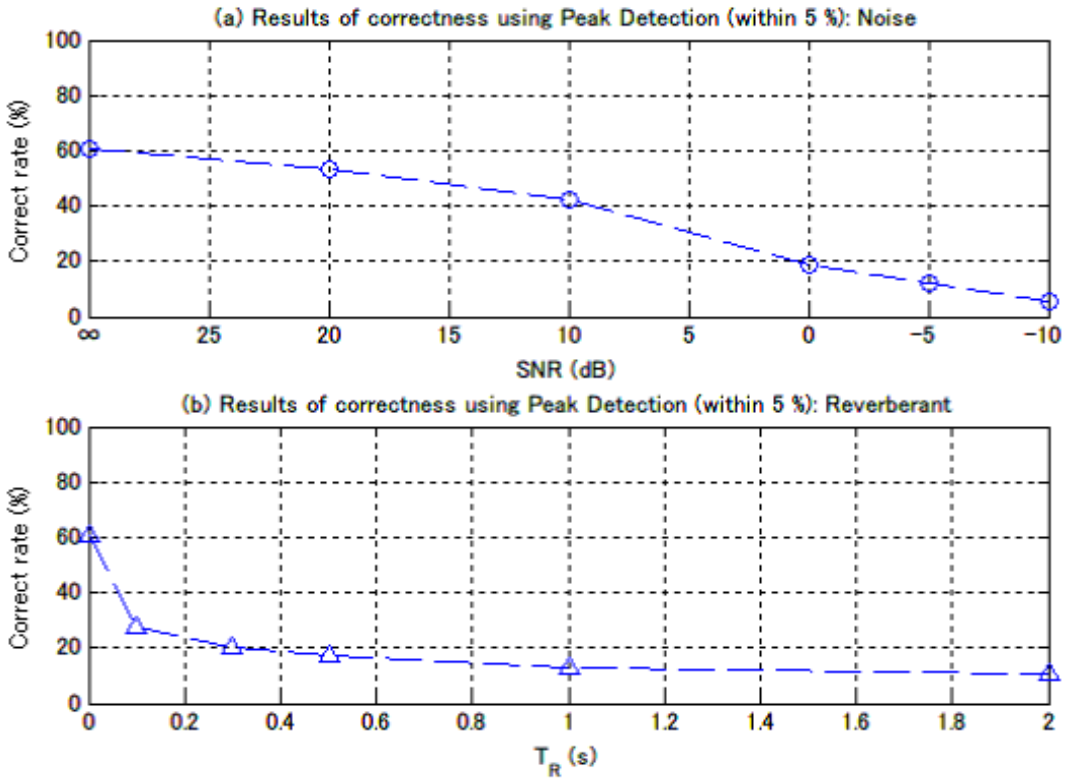

図 A.2: ピーク検出法の F0 推定性能： (a) 雑音環境，(b) 残響環境

自己相関法 [25]

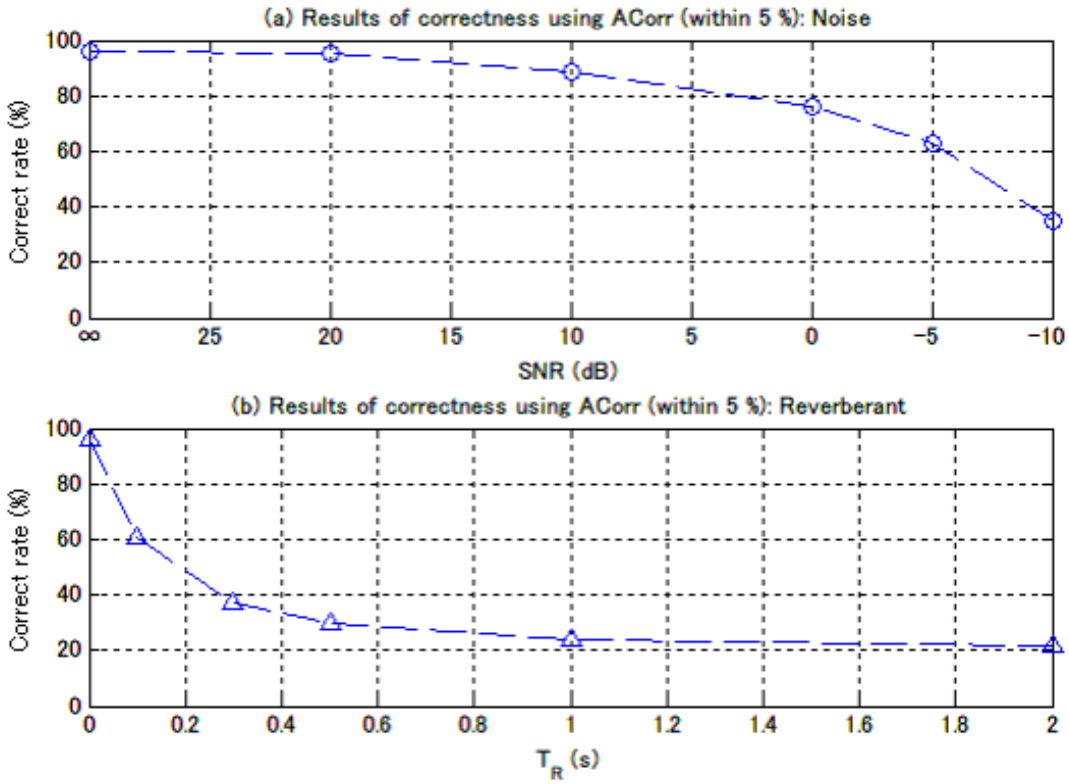

図 A.3: 自己相関法の F0 推定性能： (a) 雑音環境，(b) 残響環境

YIN[27]

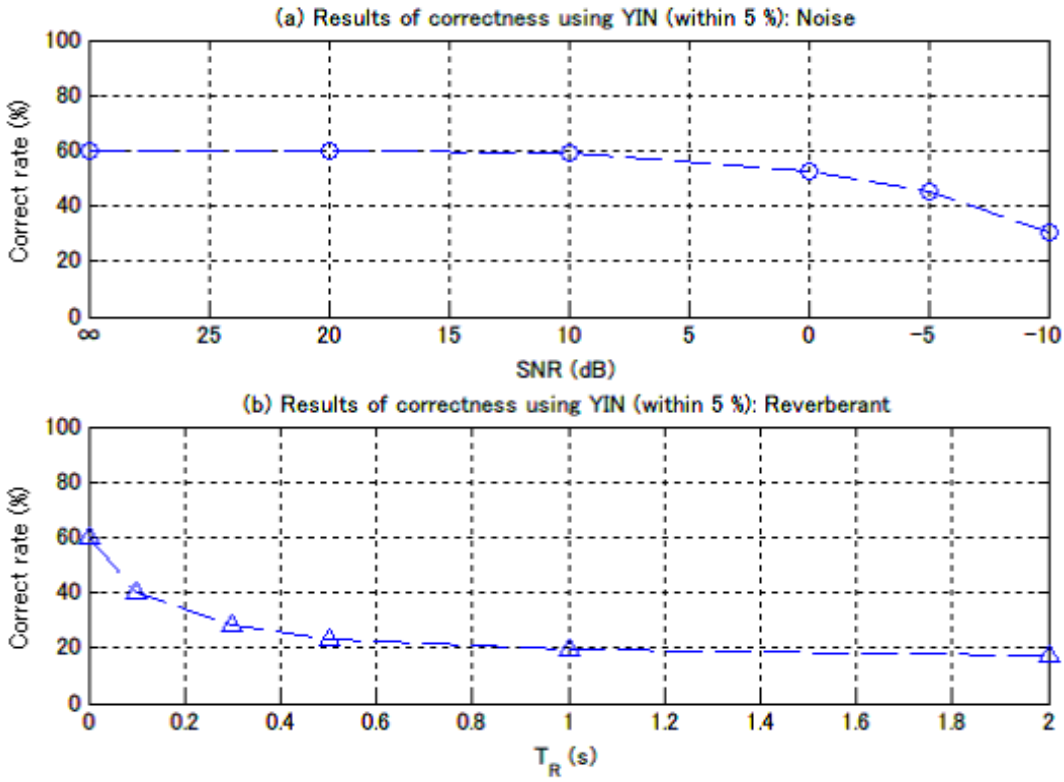

図 A.4: YIN の F0 推定性能： (a) 雑音環境，(b) 残響環境

多重窓長自己相関法（ACMWL）[26]

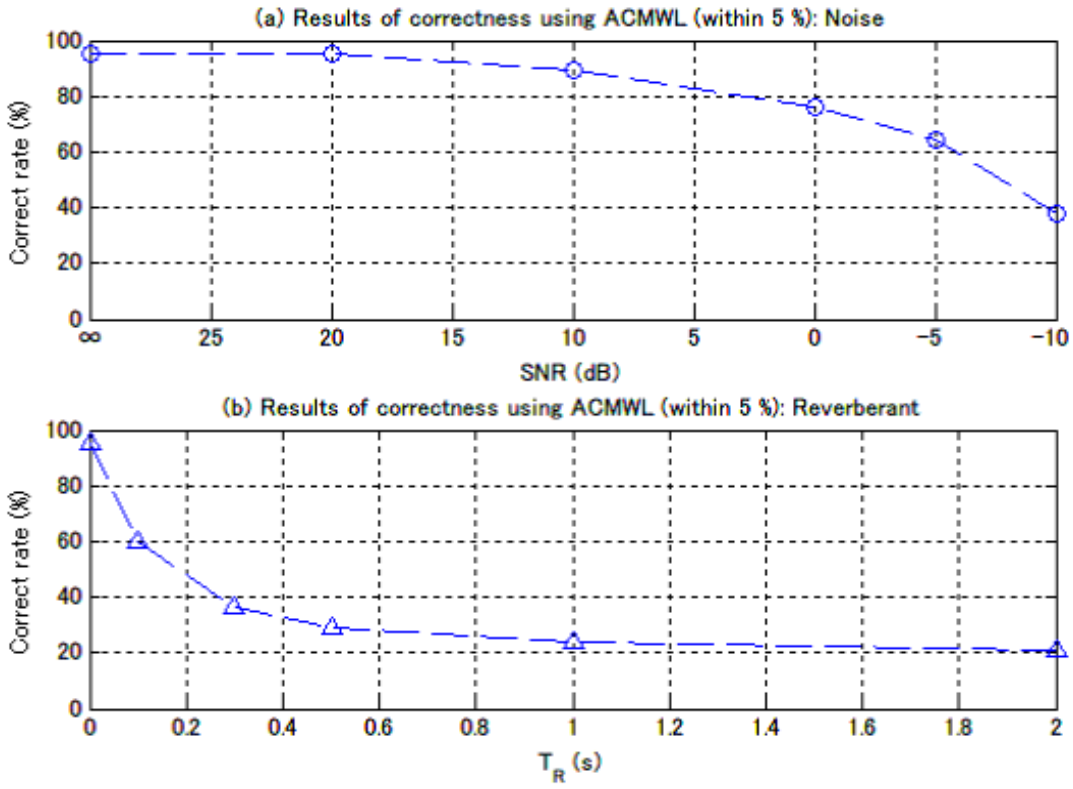

図 A.5: 多重窓長自己相関法（ACMWL）の F0 推定性能： (a) 雑音環境，(b) 残響環境

平均振幅差関数法（AMDF）[29]

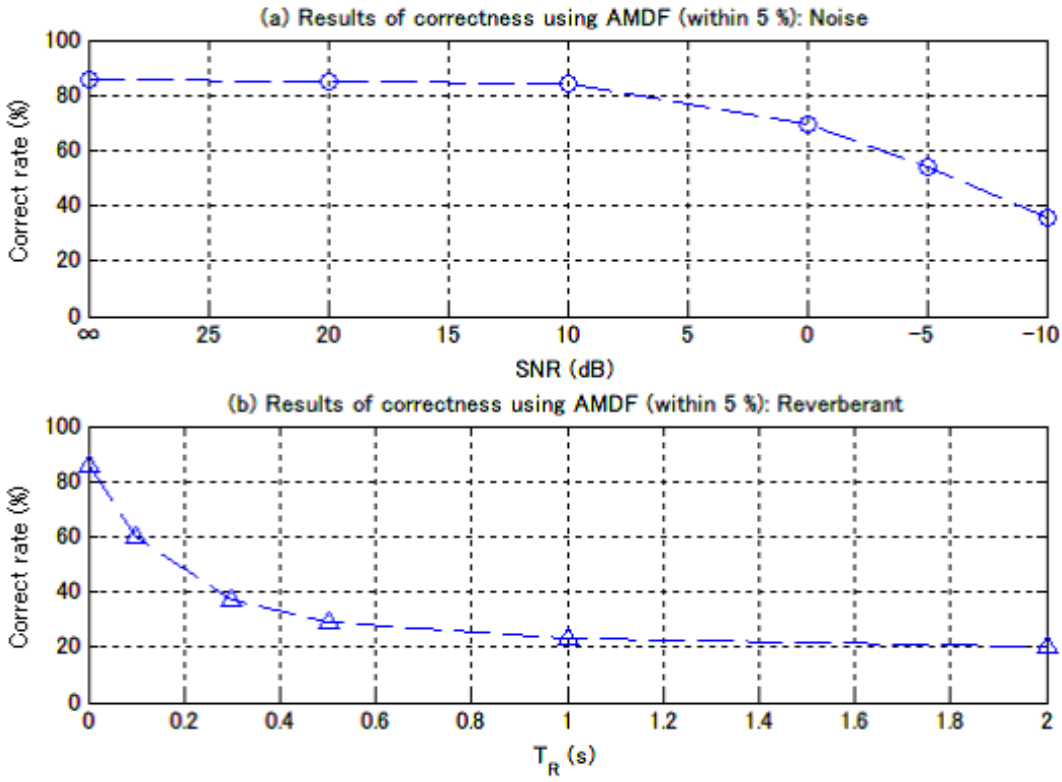

図 A.6: 平均振幅差関数法（AMDF）の F0 推定性能： (a) 雑音環境，(b) 残響環境

平均振幅差関数法（AMDF-LPC）[81]

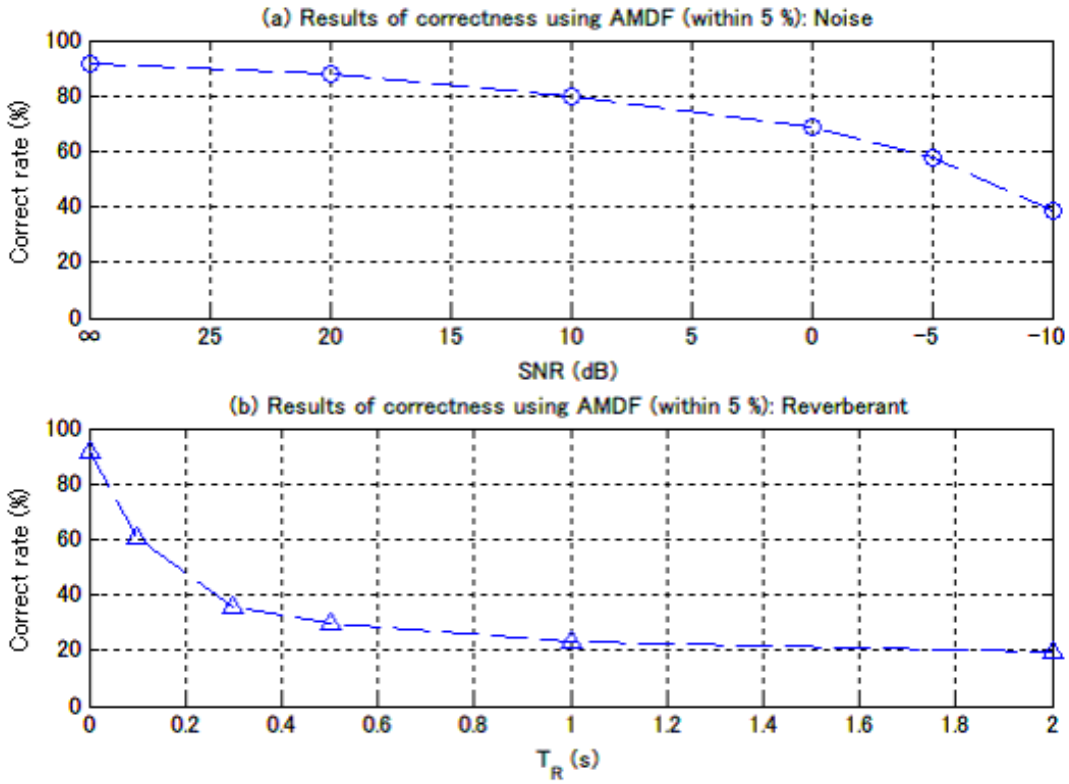

図 A.7: 平均振幅差関数法（AMDF-LPC）の F0 推定性能： (a) 雑音環境，(b) 残響環境

短時間フーリエ変換（**STFT**）[**36**]

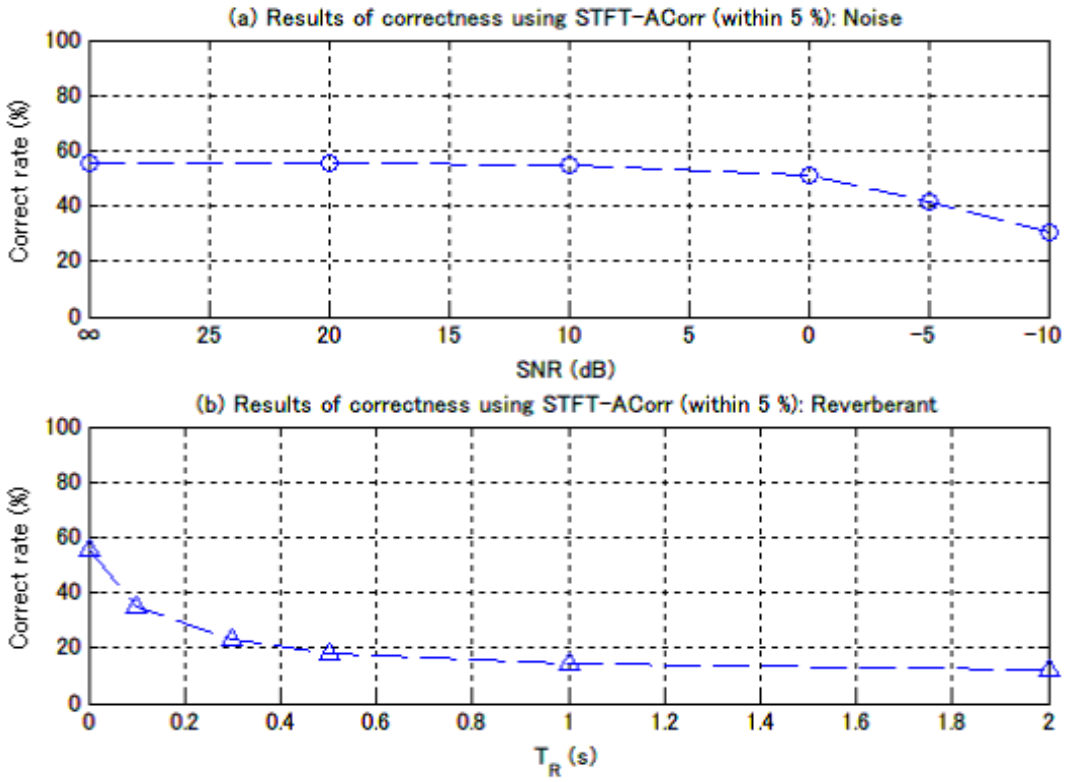

図 A.8: 短時間フーリエ変換（STFT）の F0 推定性能：(a) 雑音環境，(b) 残響環境

短時間フーリエ変換・対数振幅スペクトル（**STFT-log**）[**37**]

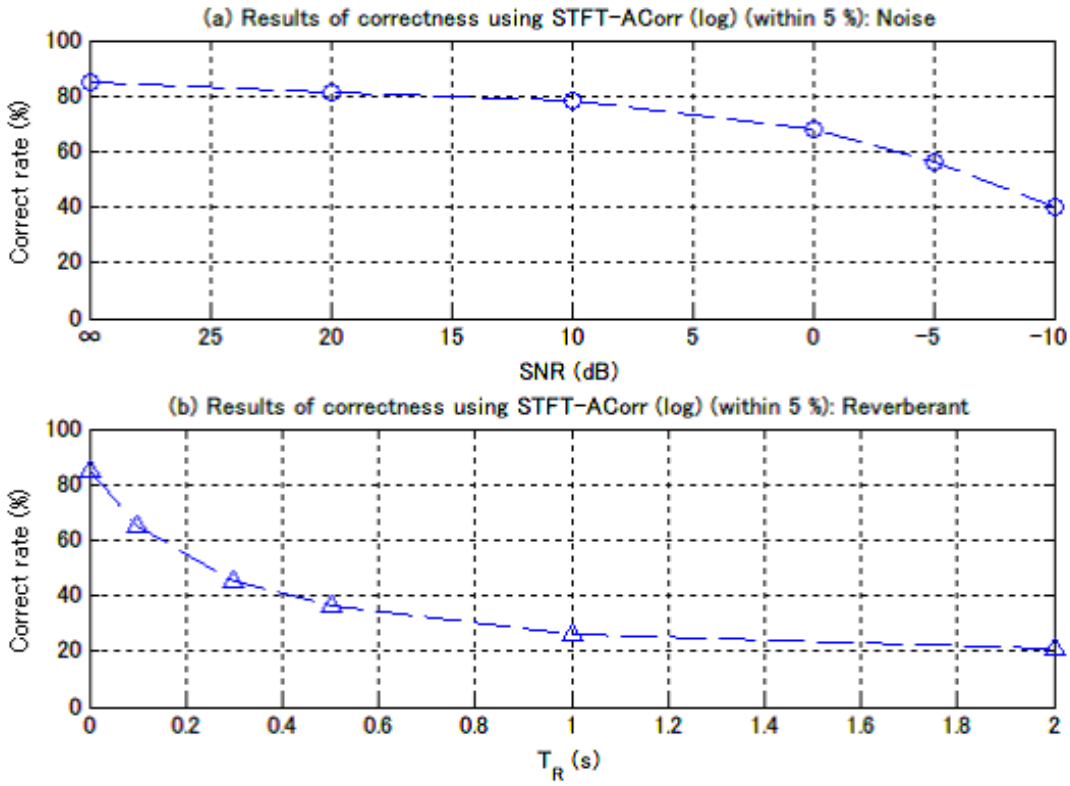

図 A.9: STFT-log の F0 推定性能： (a) 雑音環境，(b) 残響環境

短時間フーリエ変換・リフター処理（**STFT-Lifter**）[41]

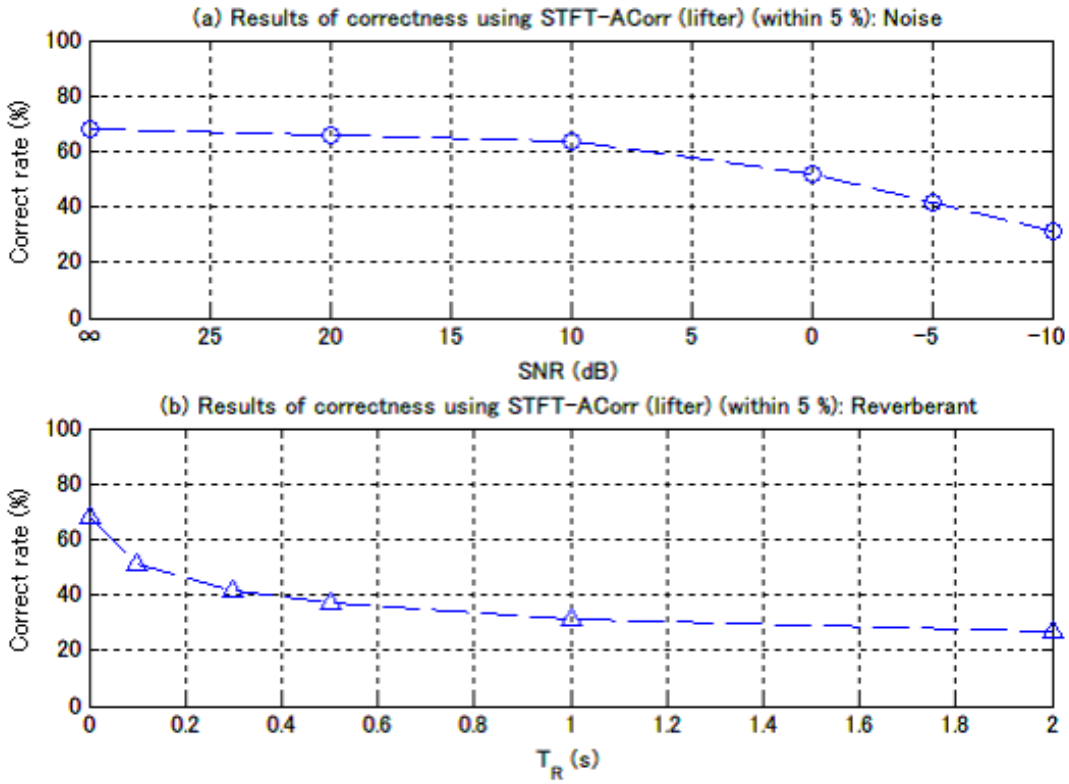

図 A.10: STFT-Lfter の F0 推定性能： (a) 雑音環境，(b) 残響環境

短時間フーリエ変換・ラグ窓（**STFT-Lag**）[40]

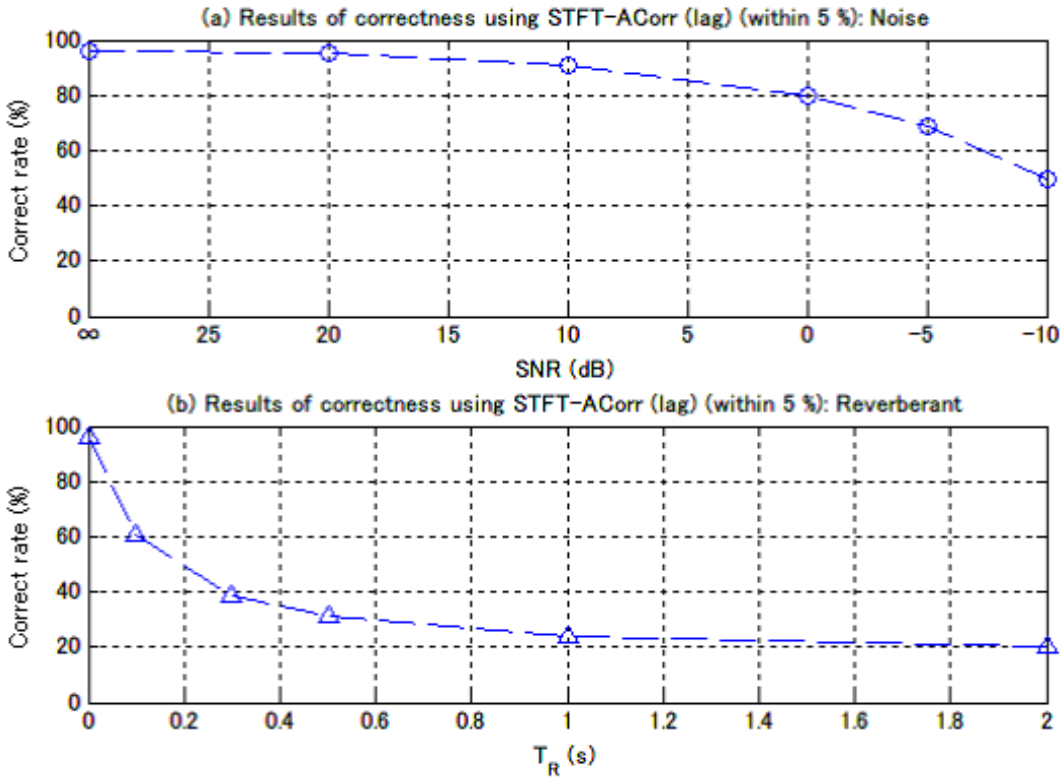

図 A.11: STFT-Lag の F0 推定性能： (a) 雑音環境，(b) 残響環境

短時間フーリエ変換・Comb フィルタ（STFT-Comb）[11]

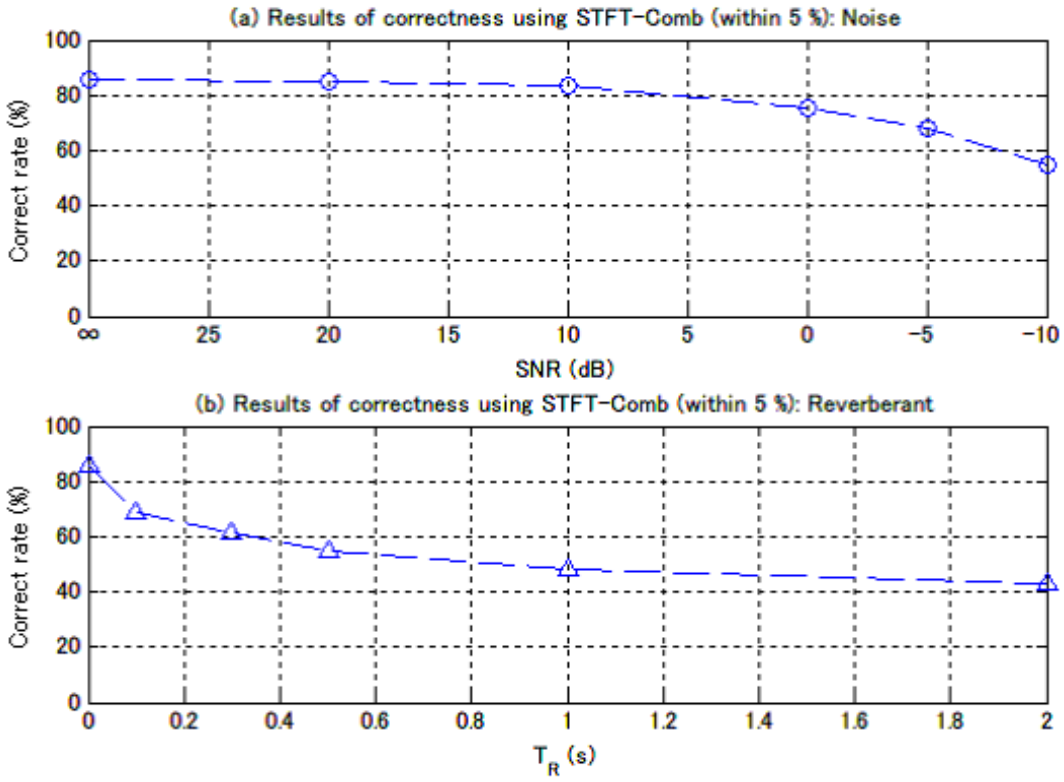

図 A.12: STFT-Comb の F0 推定性能： (a) 雑音環境，(b) 残響環境

SHS[39]

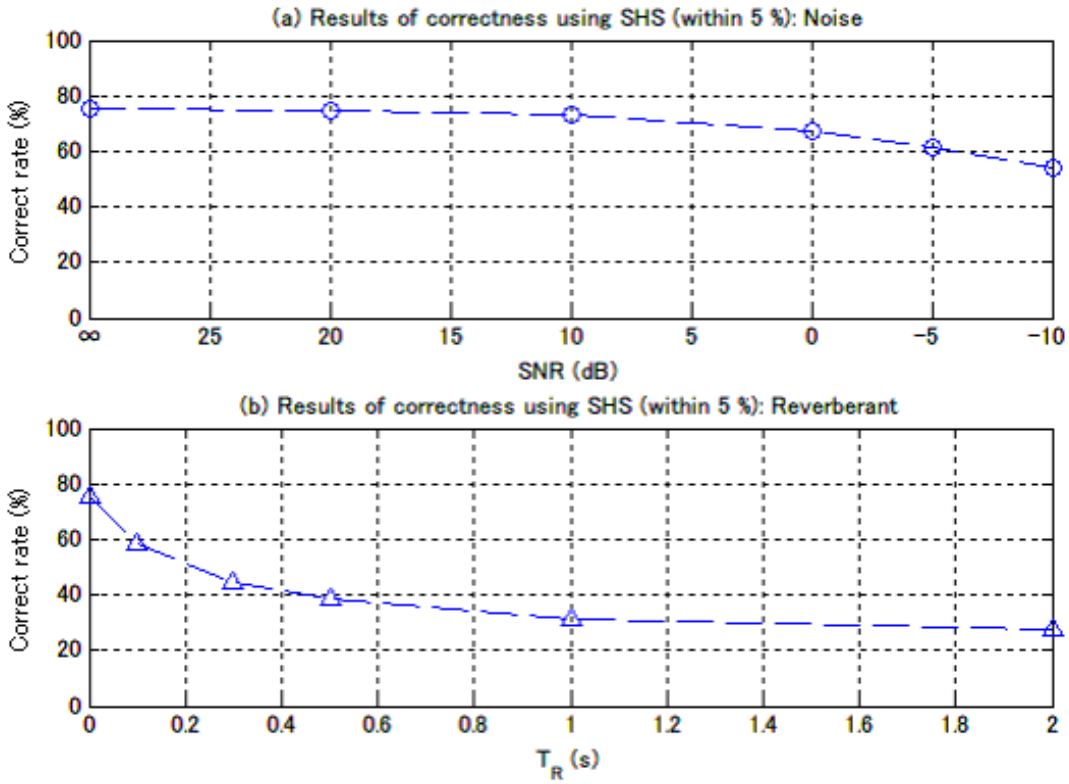

図 A.13: SHS の F0 推定性能： (a) 雑音環境，(b) 残響環境

SWIPE'[43, 44]

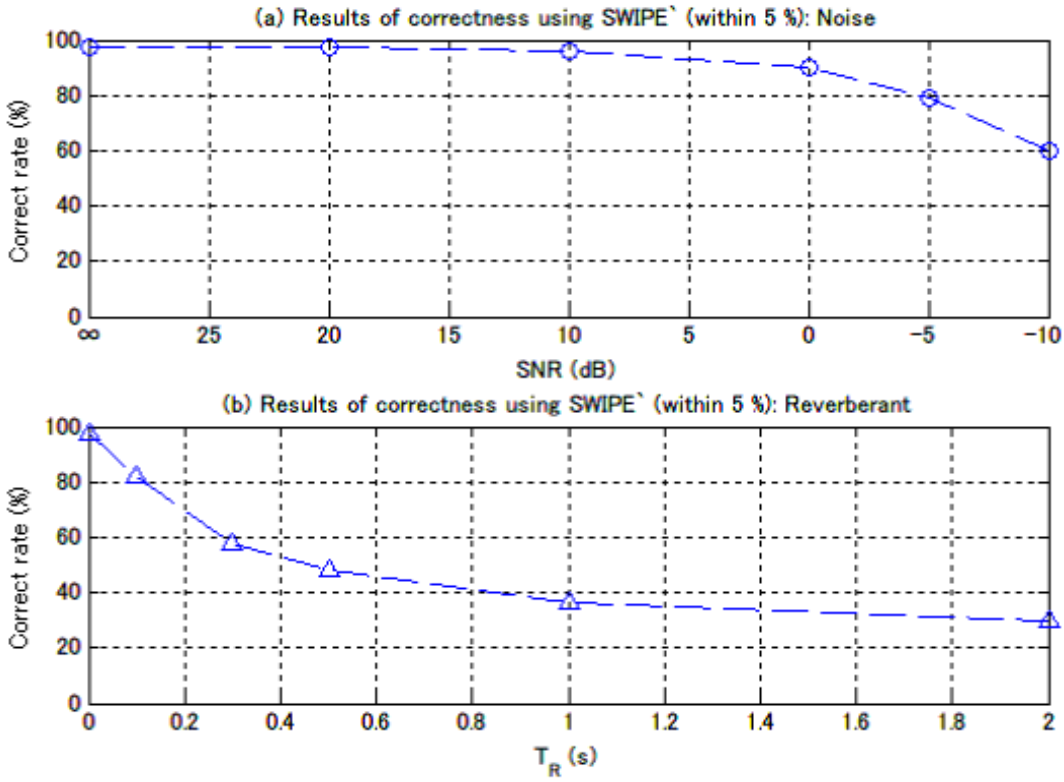

図 A.14: SWIPE' の F0 推定性能： (a) 雑音環境，(b) 残響環境

Cepstrum 法 [46, 47]

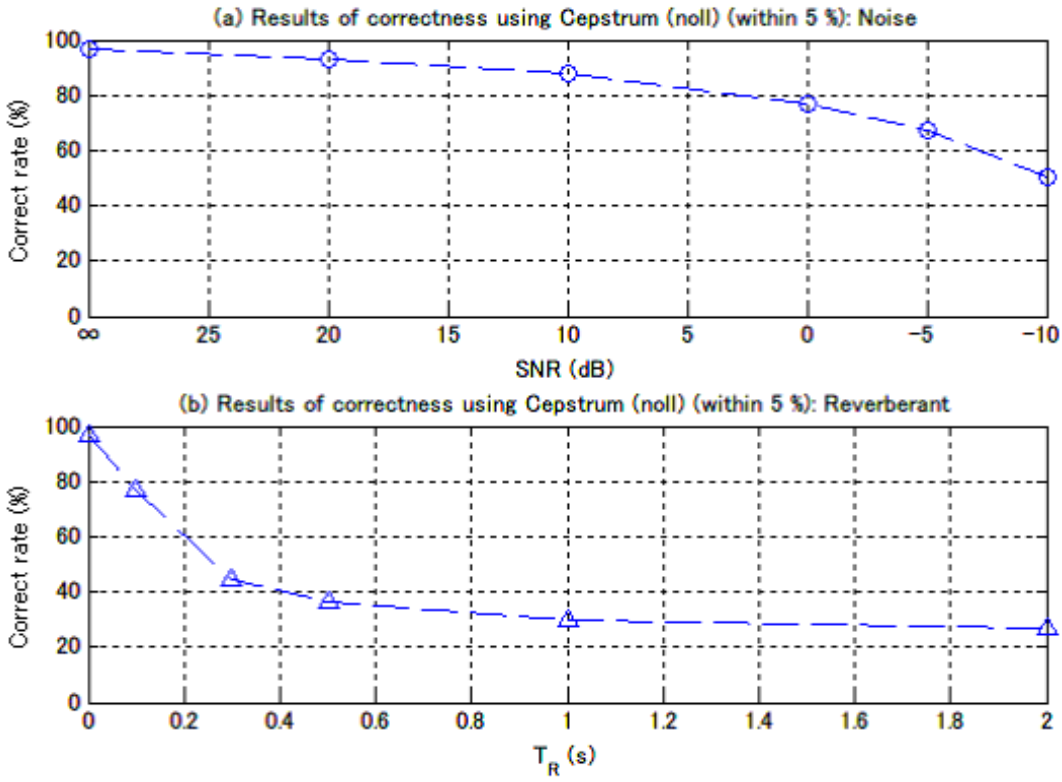

図 A.15: Cepstrum 法の F0 推定性能： (a) 雑音環境，(b) 残響環境

改良 Cepstrum 法 [49]

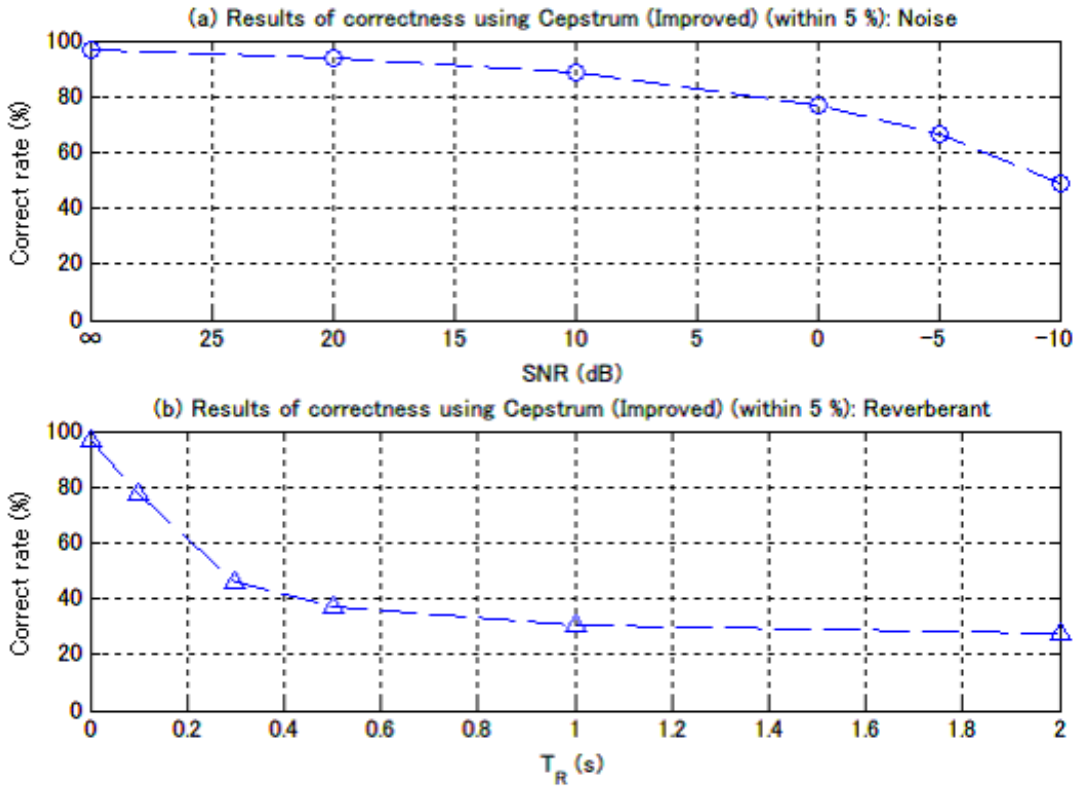

図 A.16: 改良 Cepstrum 法の F0 推定性能： (a) 雑音環境，(b) 残響環境

Clipstrum 法 [48]

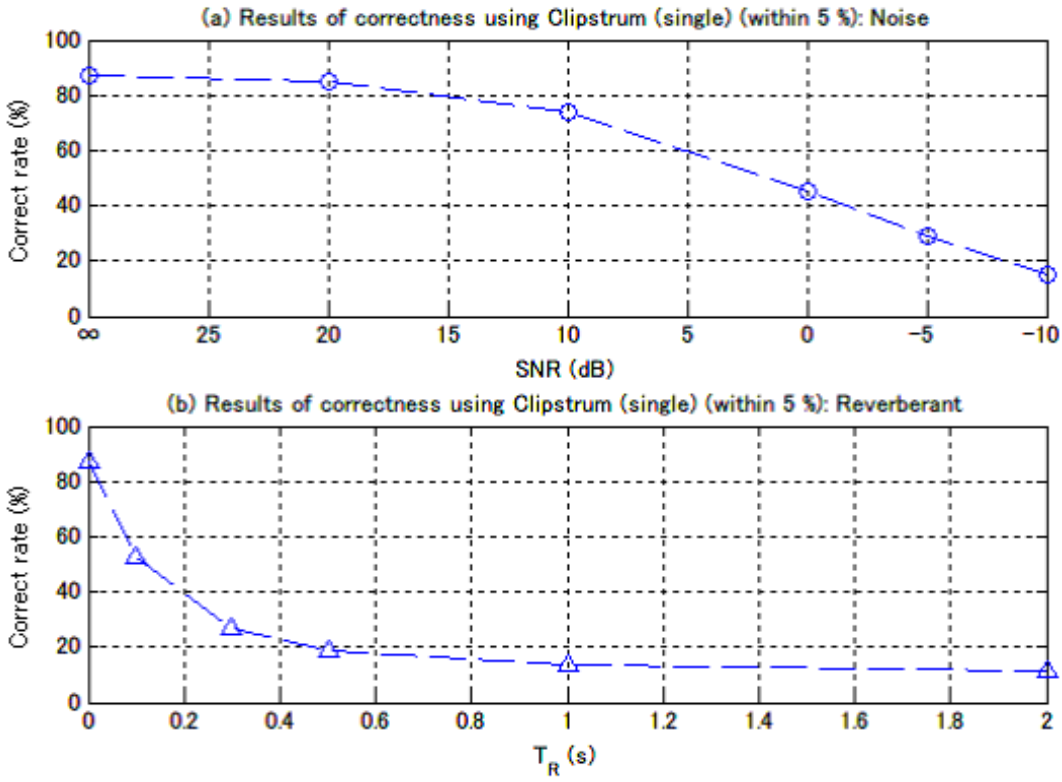

図 A.17: Clipstrum 法の F0 推定性能： (a) 雑音環境，(b) 残響環境

複素ケプストラム法 [51, 52]

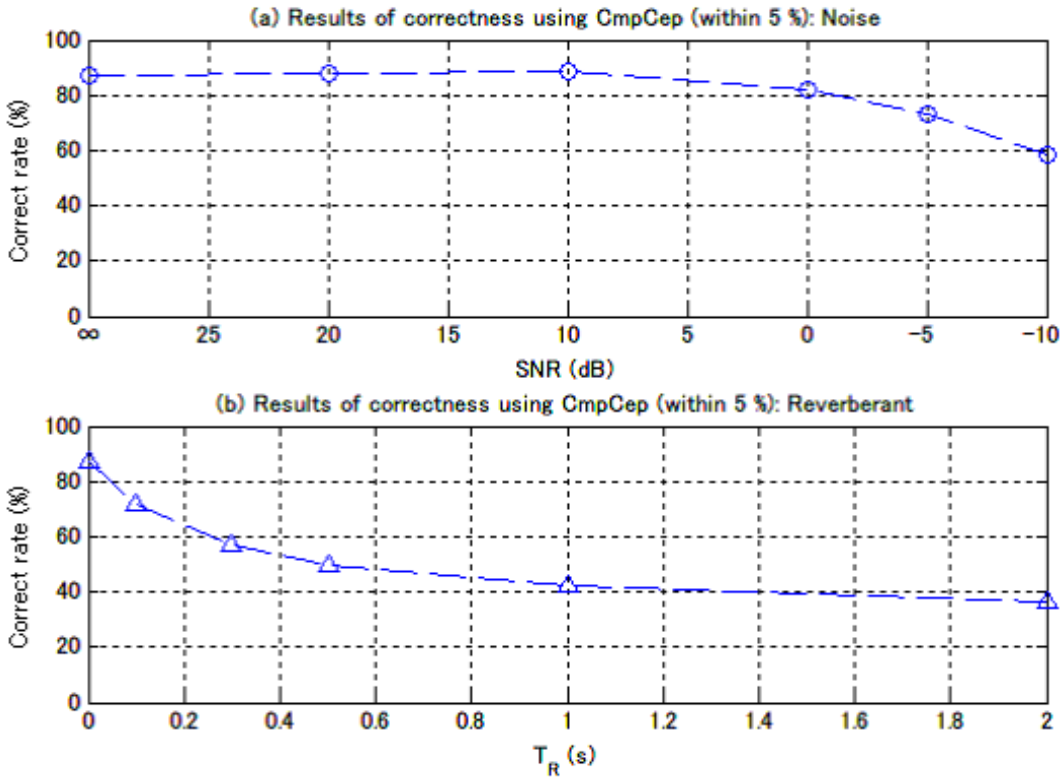

図 A.18: 複素ケプストラム法の F0 推定性能： (a) 雑音環境，(b) 残響環境

LPC 法 [33]

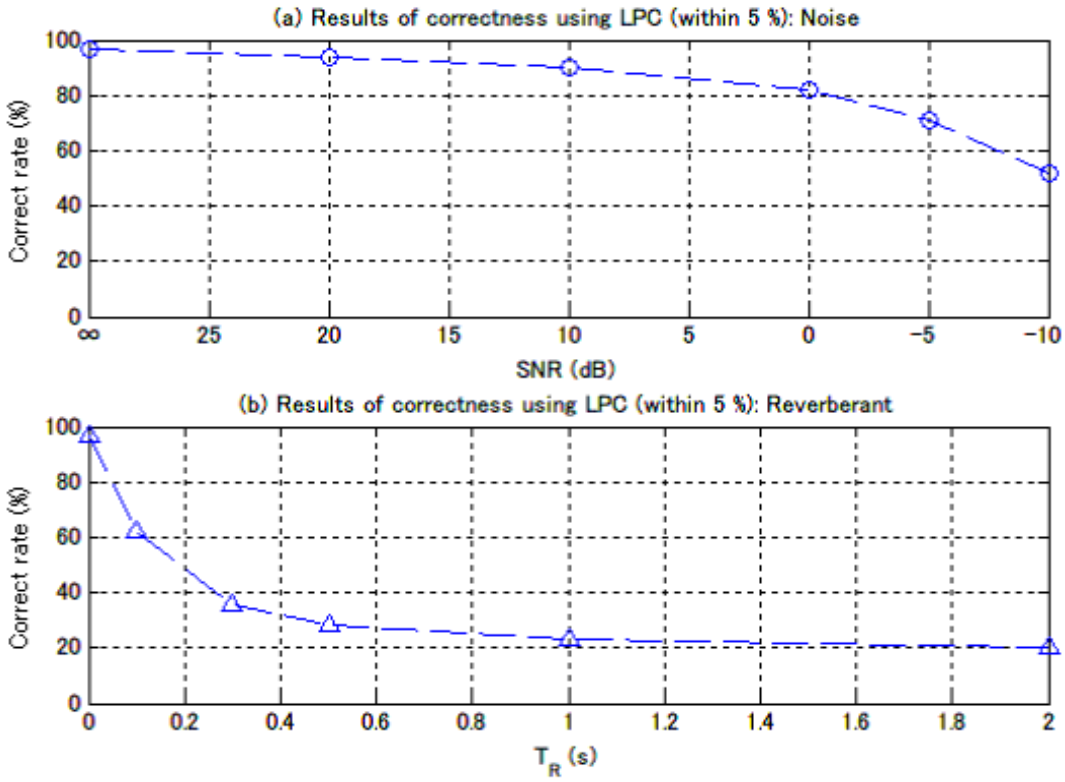

図 A.19: LPC 法の F0 推定性能： (a) 雑音環境，(b) 残響環境

LPC-SIFT 法 [34]

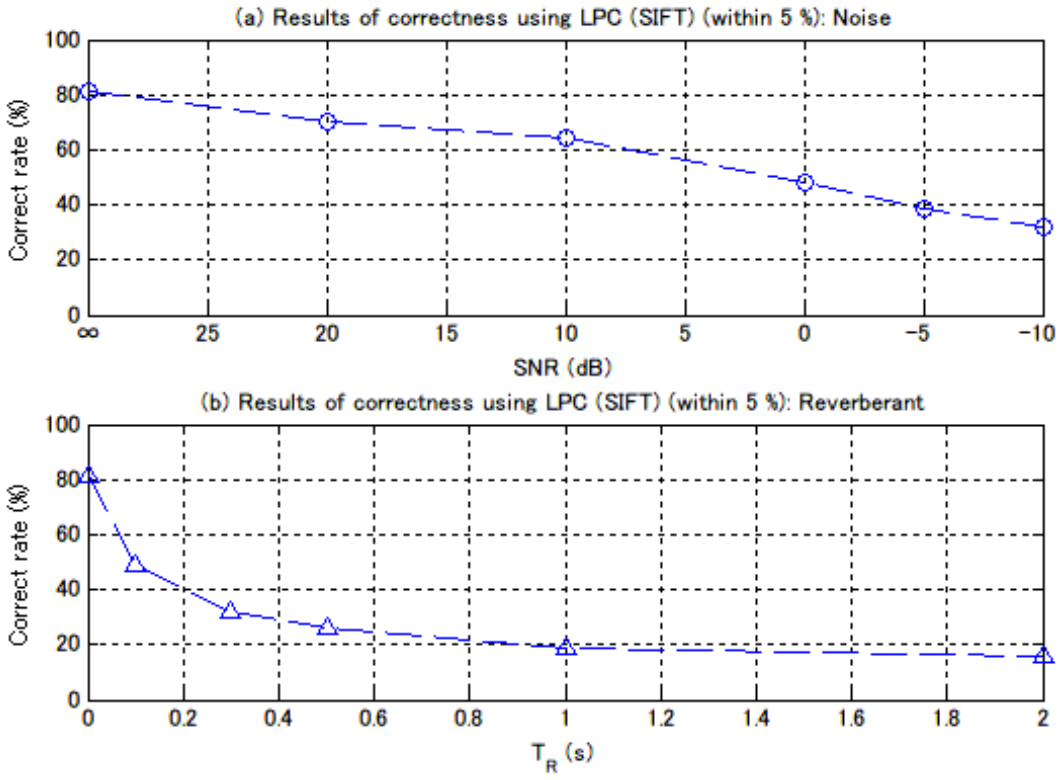

図 A.20: LPC-SIFT の F0 推定性能： (a) 雑音環境，(b) 残響環境

TEMPO 法 [53]

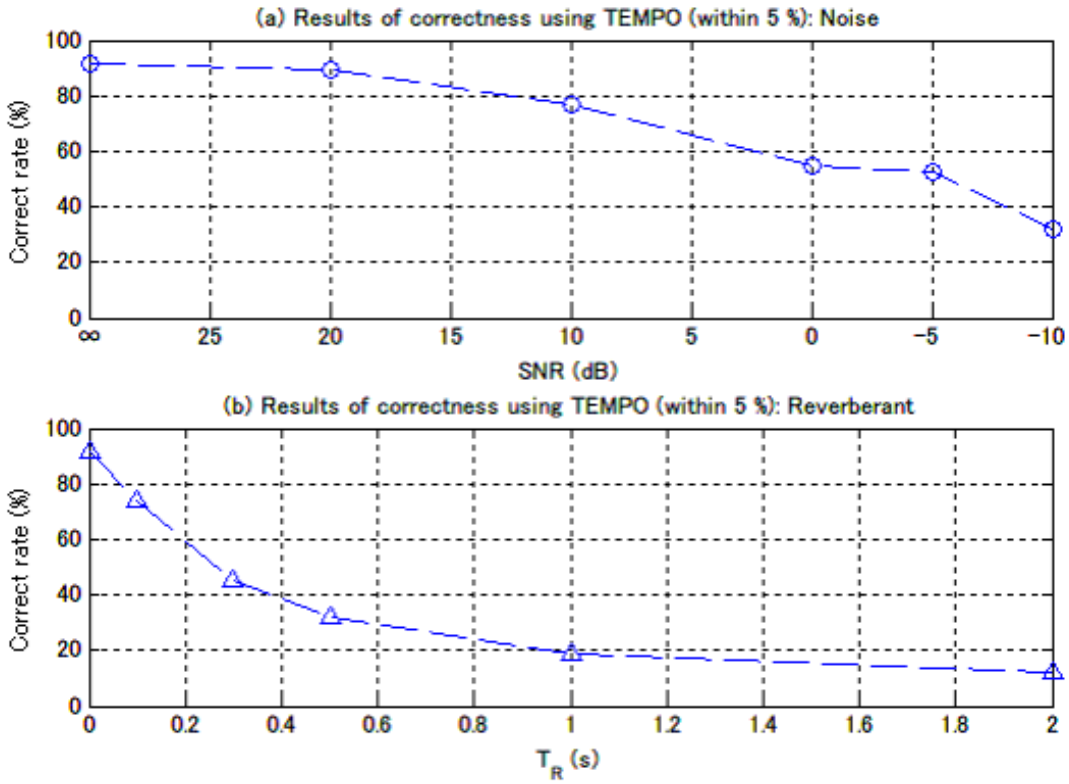

図 A.21: TEMPO 法の F0 推定性能： (a) 雑音環境，(b) 残響環境

TEMPO2[54]

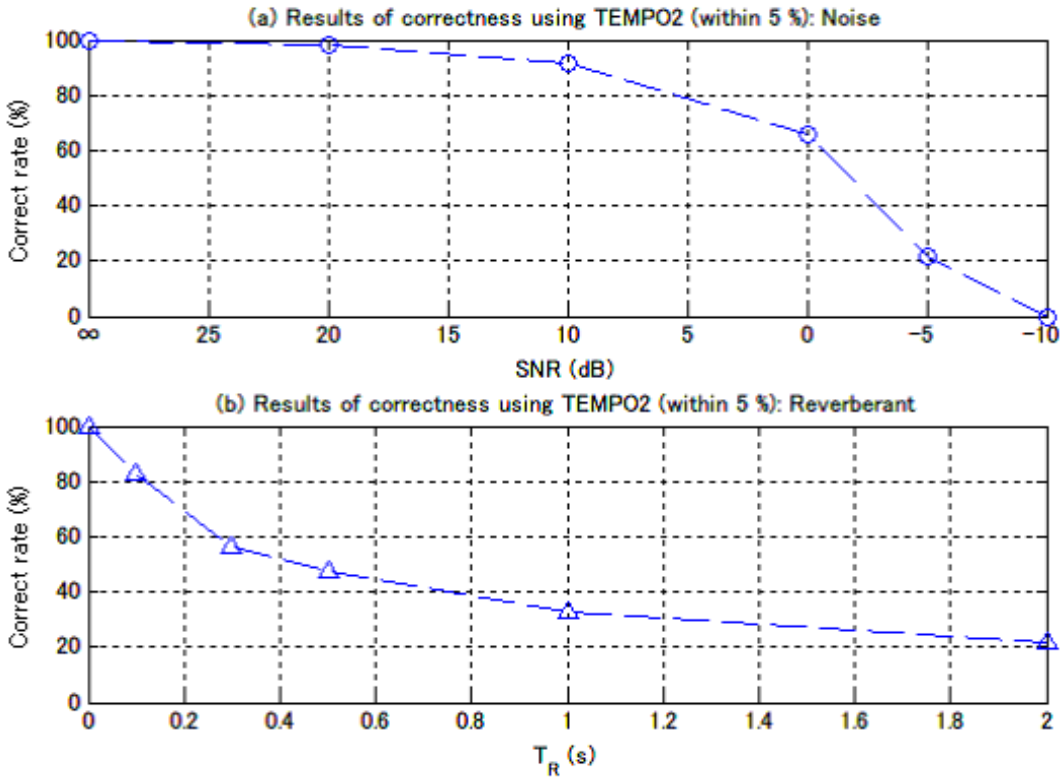

図 A.22: TEMPO2 の F0 推定性能： (a) 雑音環境，(b) 残響環境

IFHC[56]

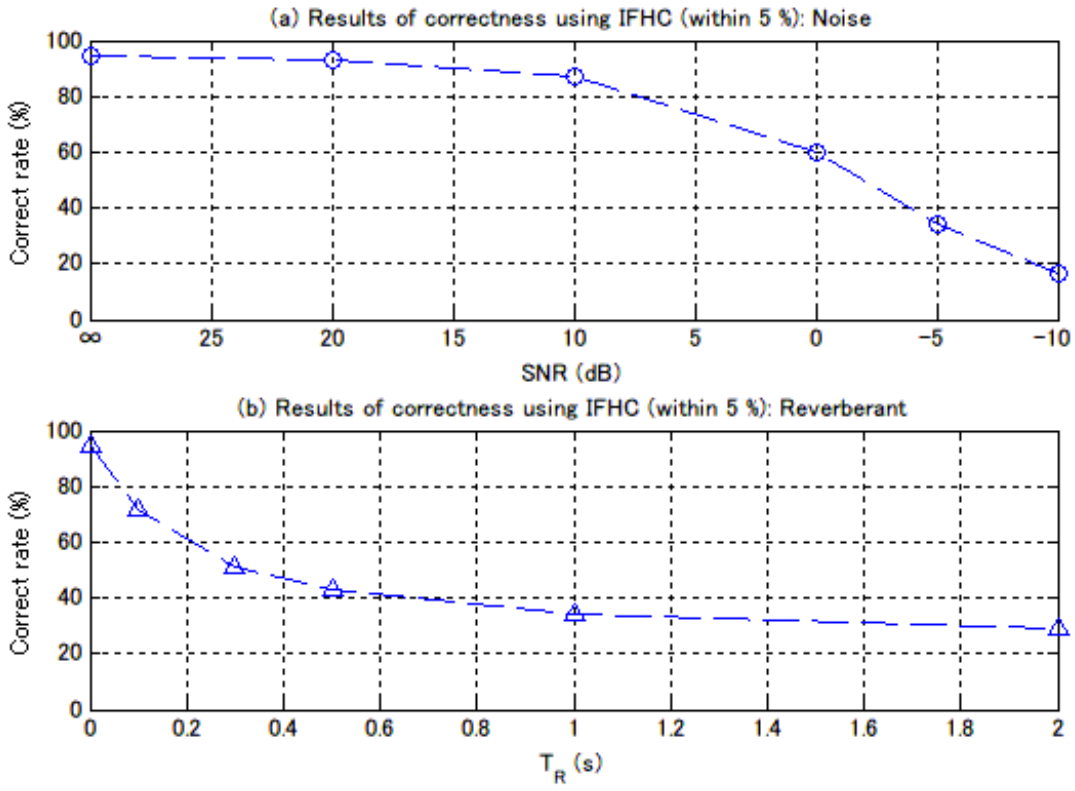

図 A.23: IFHC の F0 推定性能： (a) 雑音環境，(b) 残響環境

基本波フィルタリング法 [30]

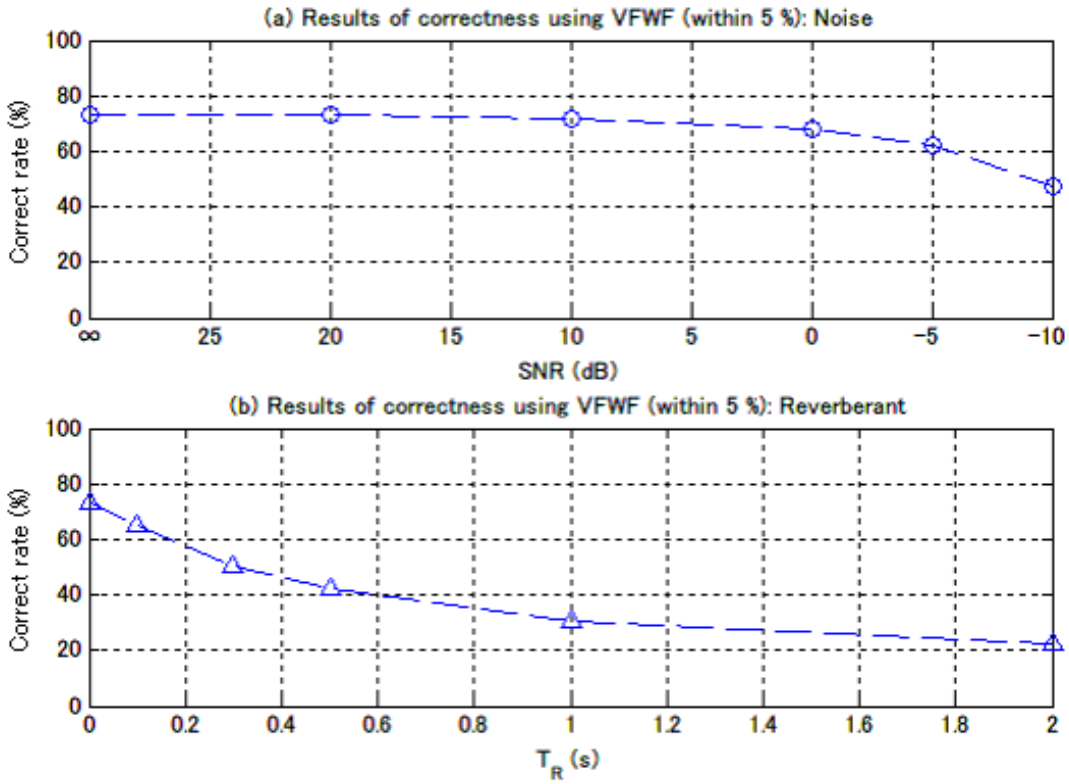

図 A.24: 基本波フィルタリング法の F0 推定性能： (a) 雑音環境，(b) 残響環境

PHIA[57]

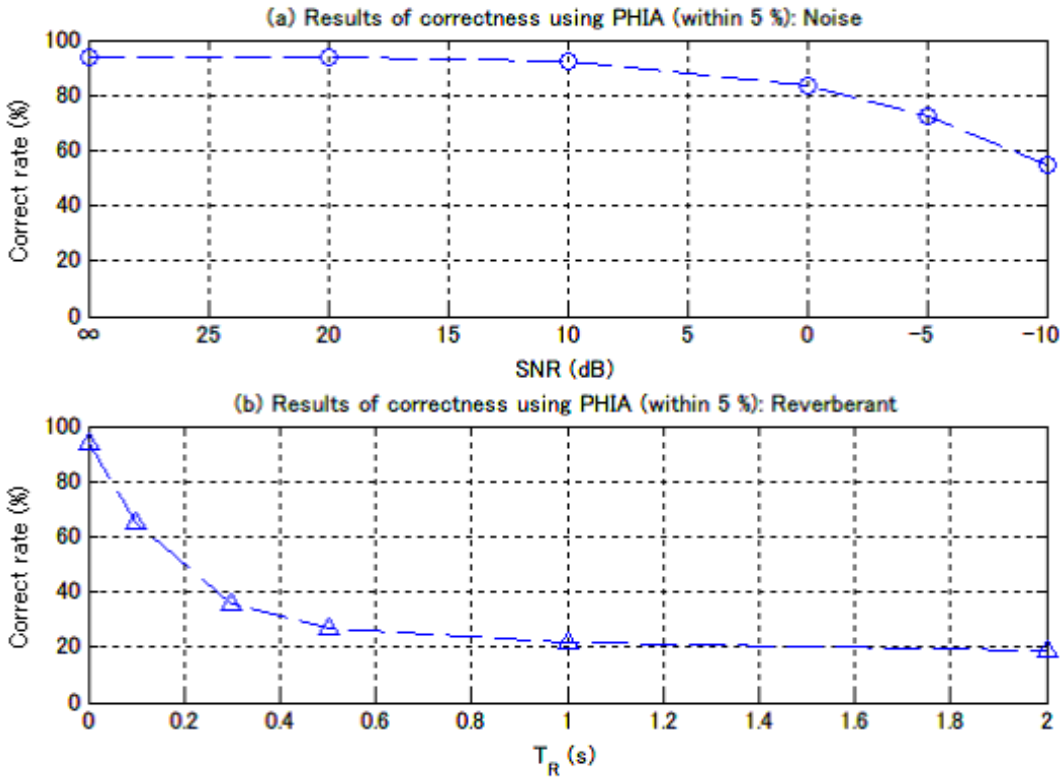

図 A.25: PHIA の F0 推定性能： (a) 雑音環境，(b) 残響環境

付録 B

シミュレーション結果補足（第 5 章）

ステップ的に F_0 が変化する調波複合音

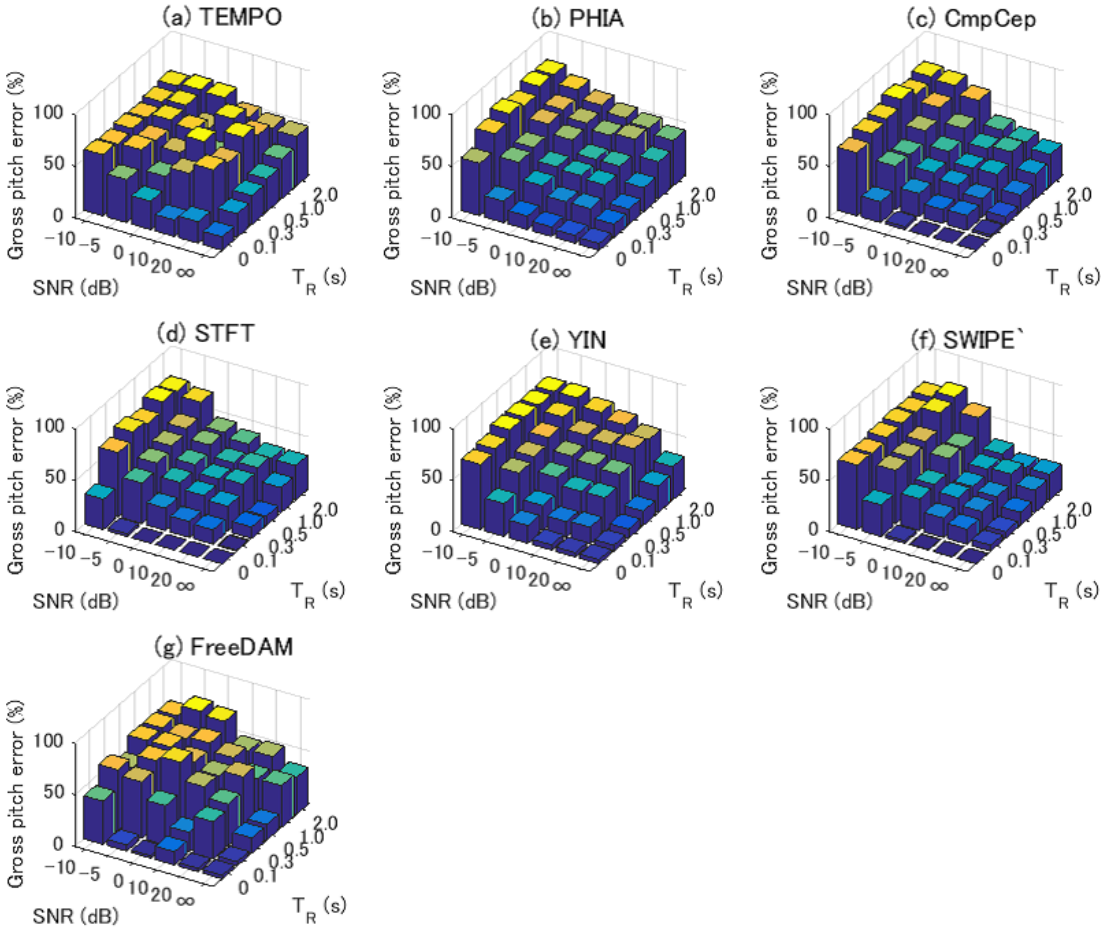

図 B.1: 雑音残響環境における調波複合音（ステップ）の Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN, (f) SWIPE’, (g) FreeDAM（Proposed）

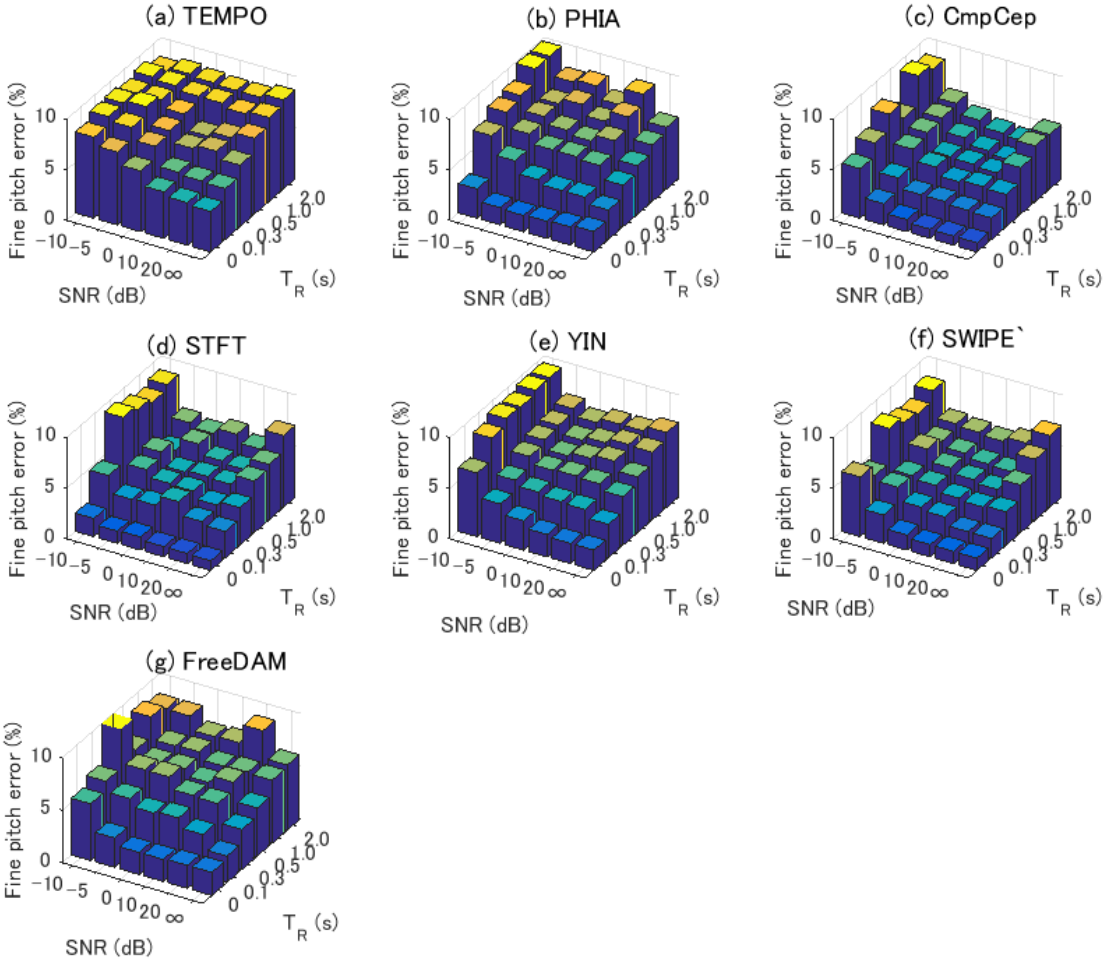

図 B.2: 雑音残響環境における調波複合音（ステップ）の Fine pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN, (f) SWIPE’, (g) FreeDAM（Proposed）

ステップ的に F_0 が変化する合成音

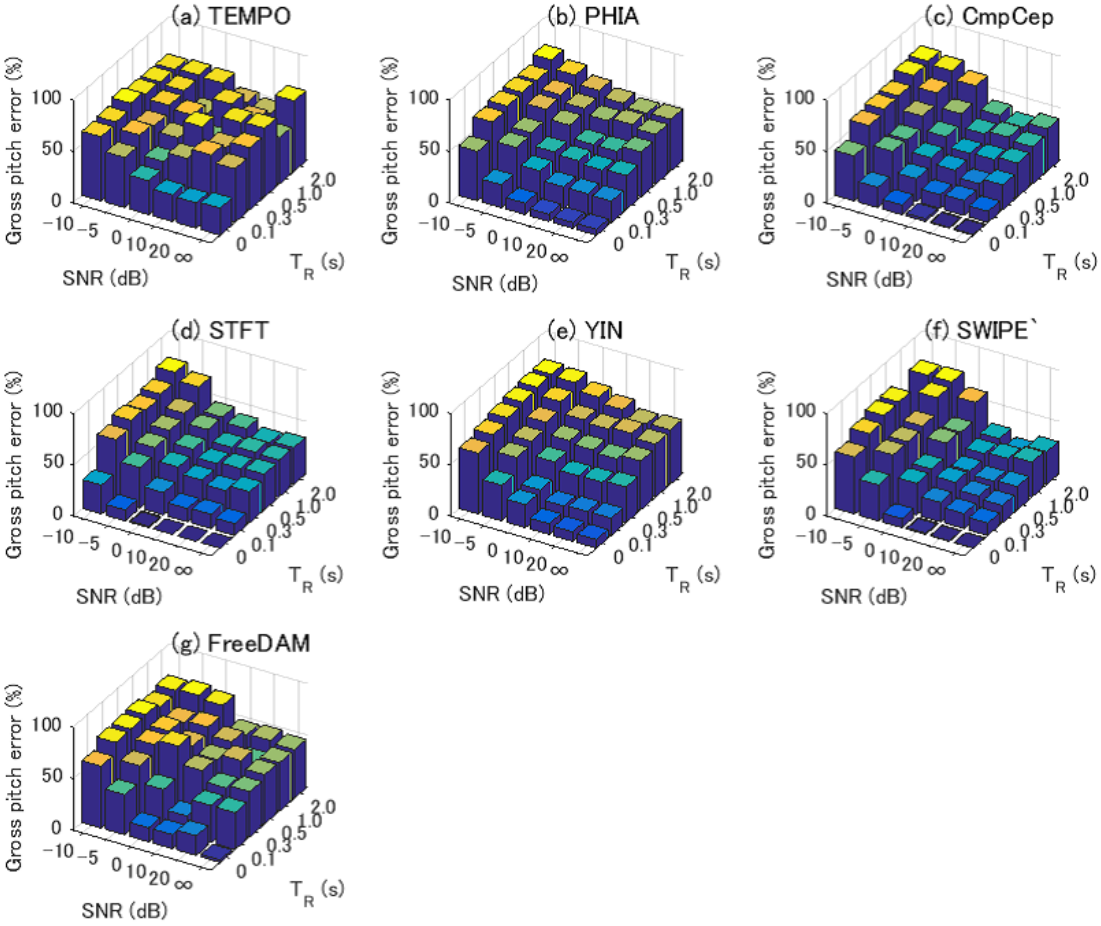

図 B.3: 雑音残響環境における調波合成音（ステップ）の Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN, (f) SWIPE’, (g) FreeDAM（Proposed）

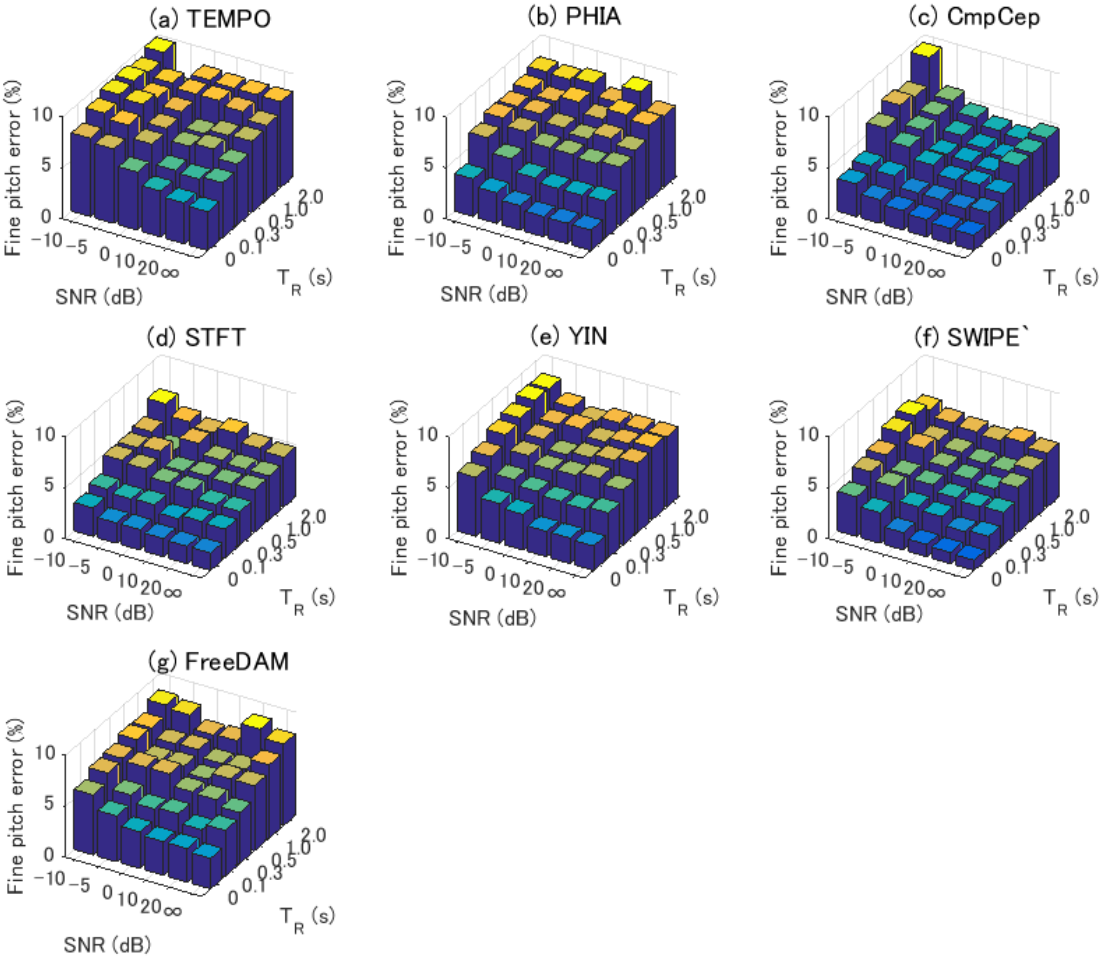

図 B.4: 雑音残響環境における調波合成音（ステップ）の Fine pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN, (f) SWIPE’, (g) FreeDAM（Proposed）

連続的に F_0 が変化する調波複合音

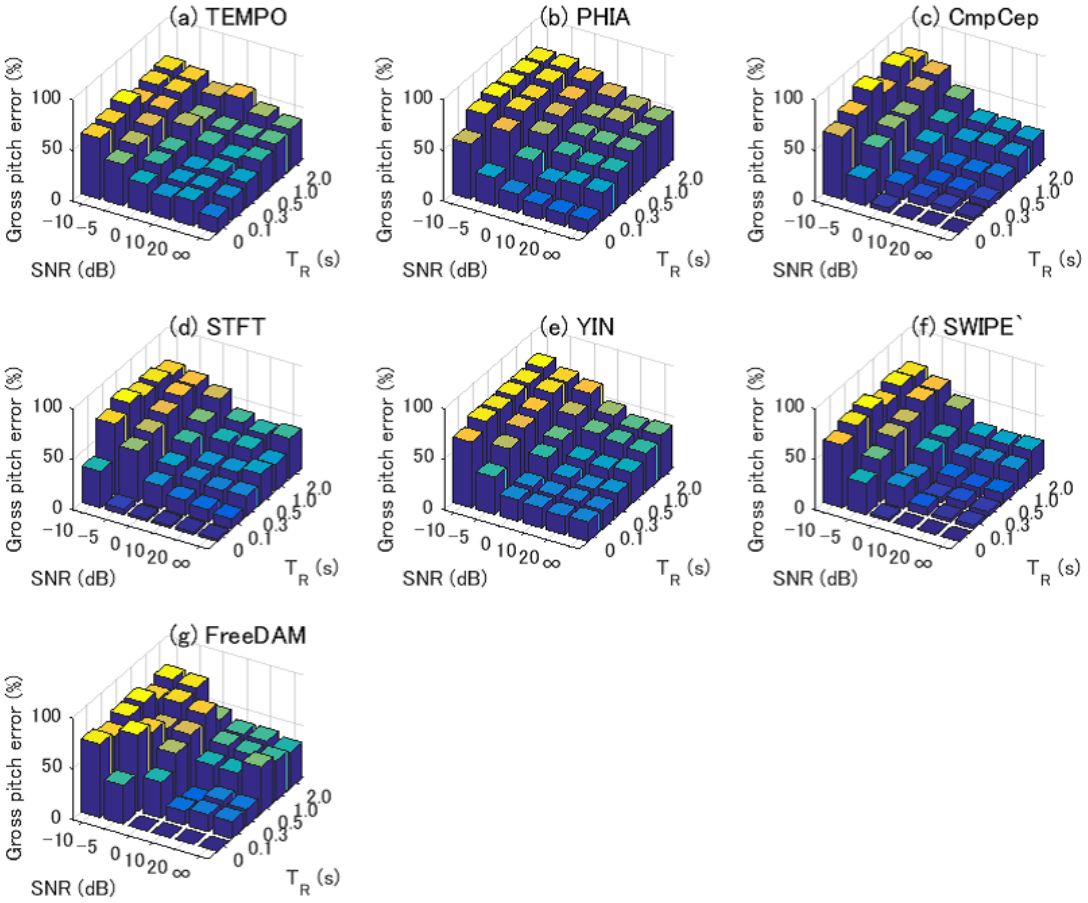

図 B.5: 雑音残響環境における調波複合音（連続）の Gross pitch error: (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

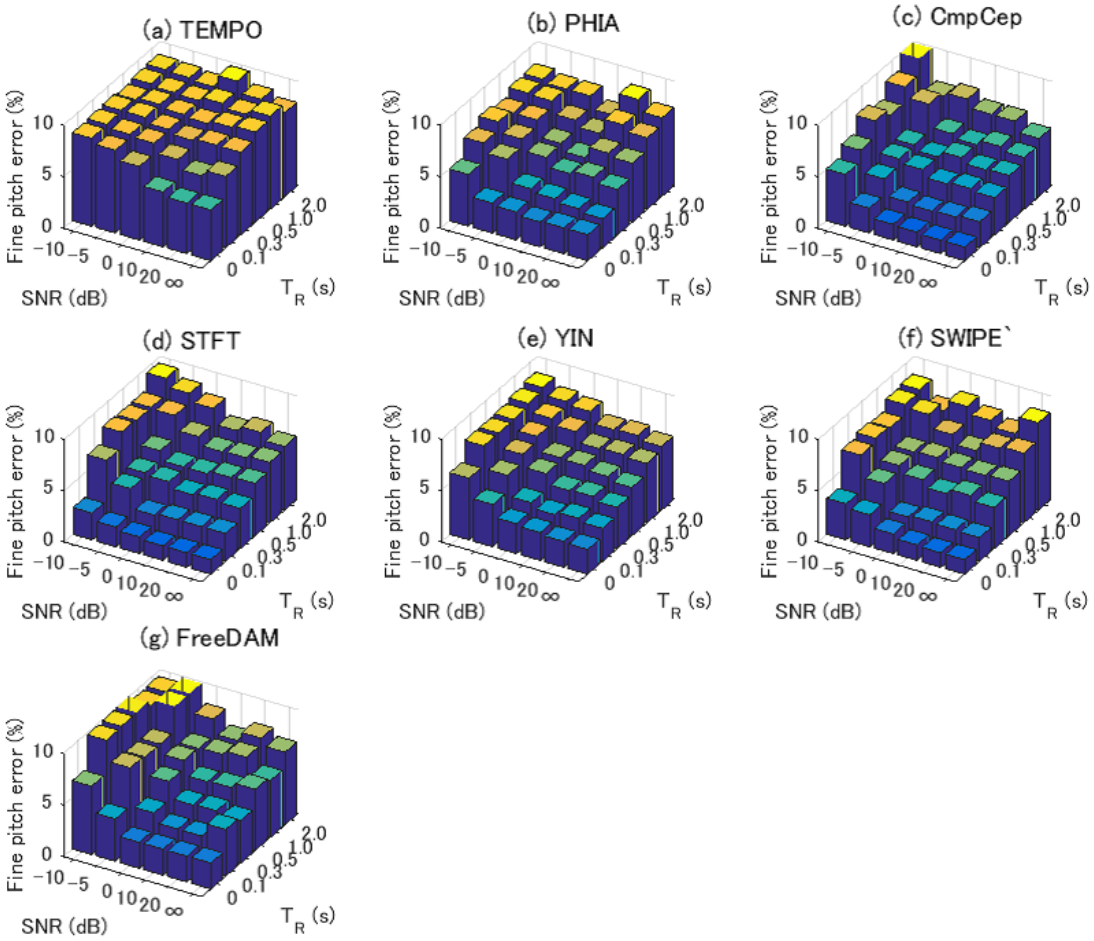

図 B.6: 雑音残響環境における調波複合音（連続）の Fine pitch error：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

連続的に F_0 が変化する合成音

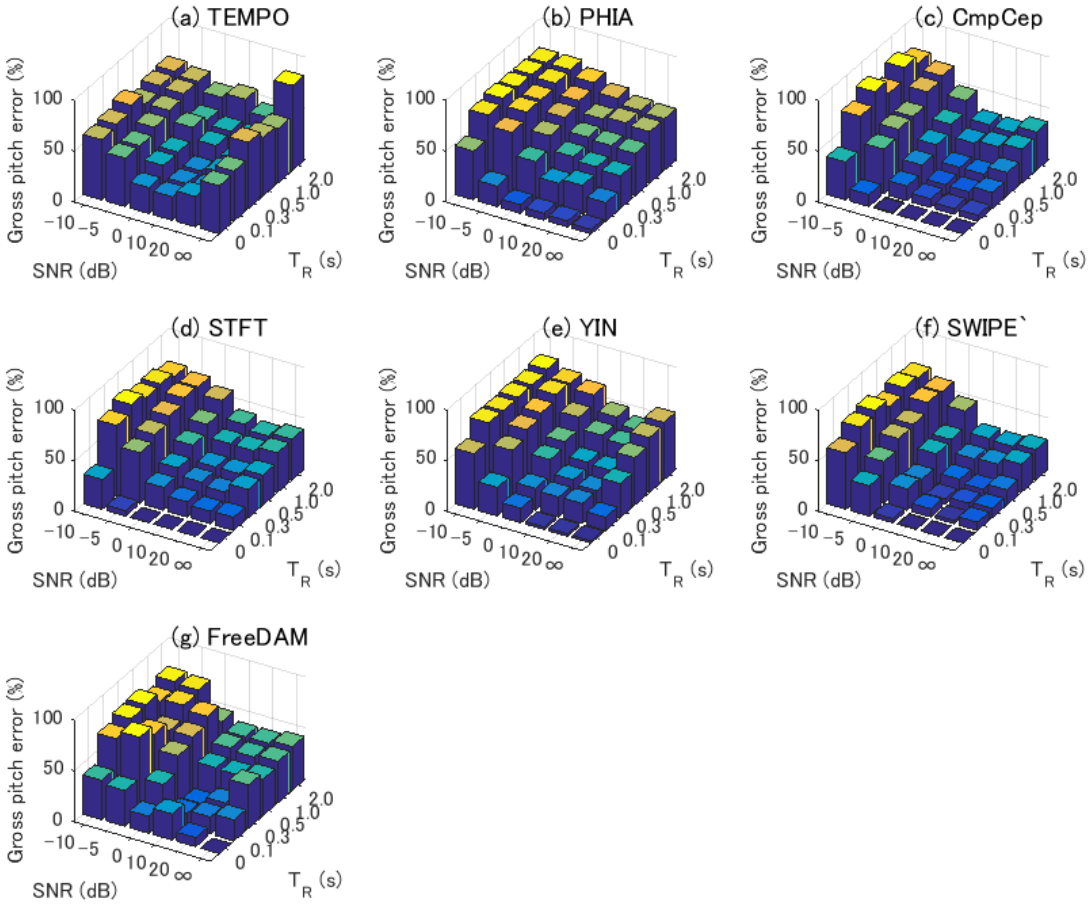

図 B.7: 雑音残響環境における調波合成音（連続）の Gross pitch error: (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

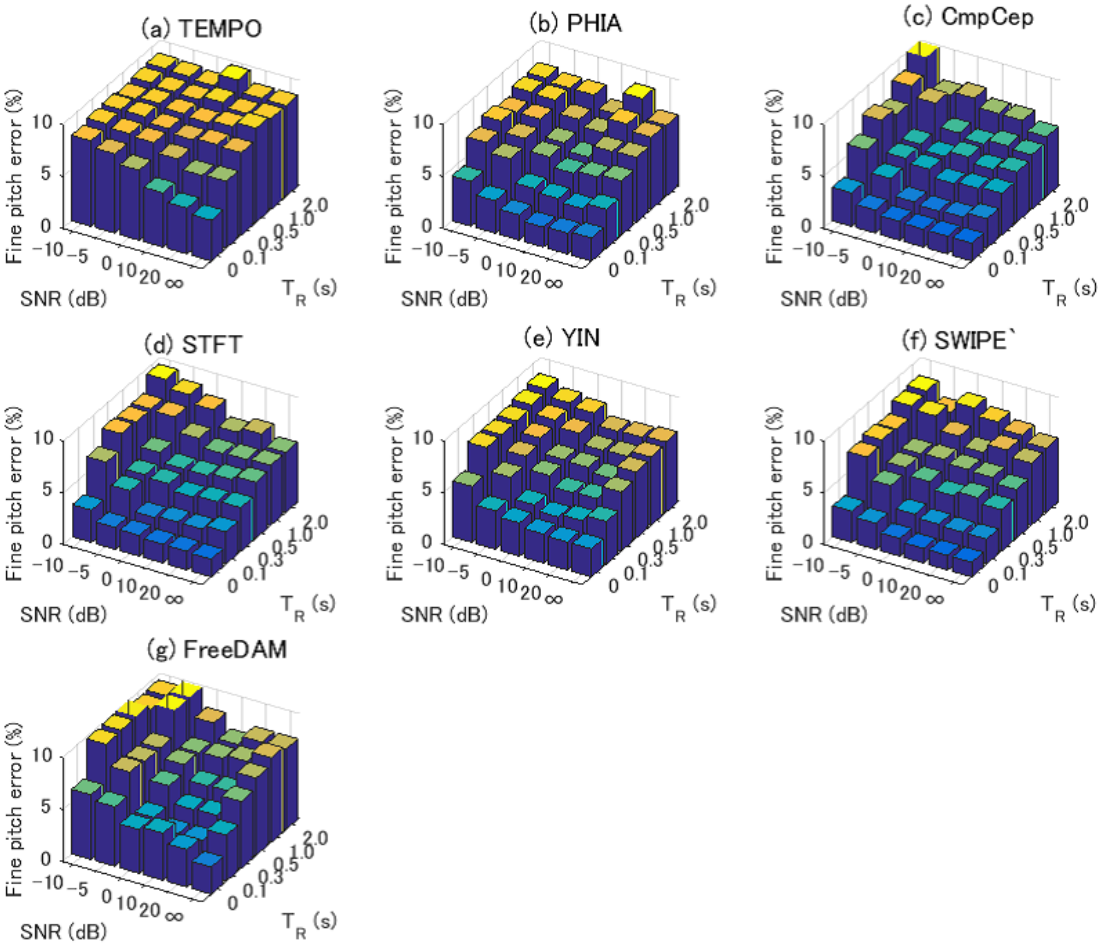

図 B.8: 雑音残響環境における調波合成音（連続）の Fine pitch error：(a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

実音声信号

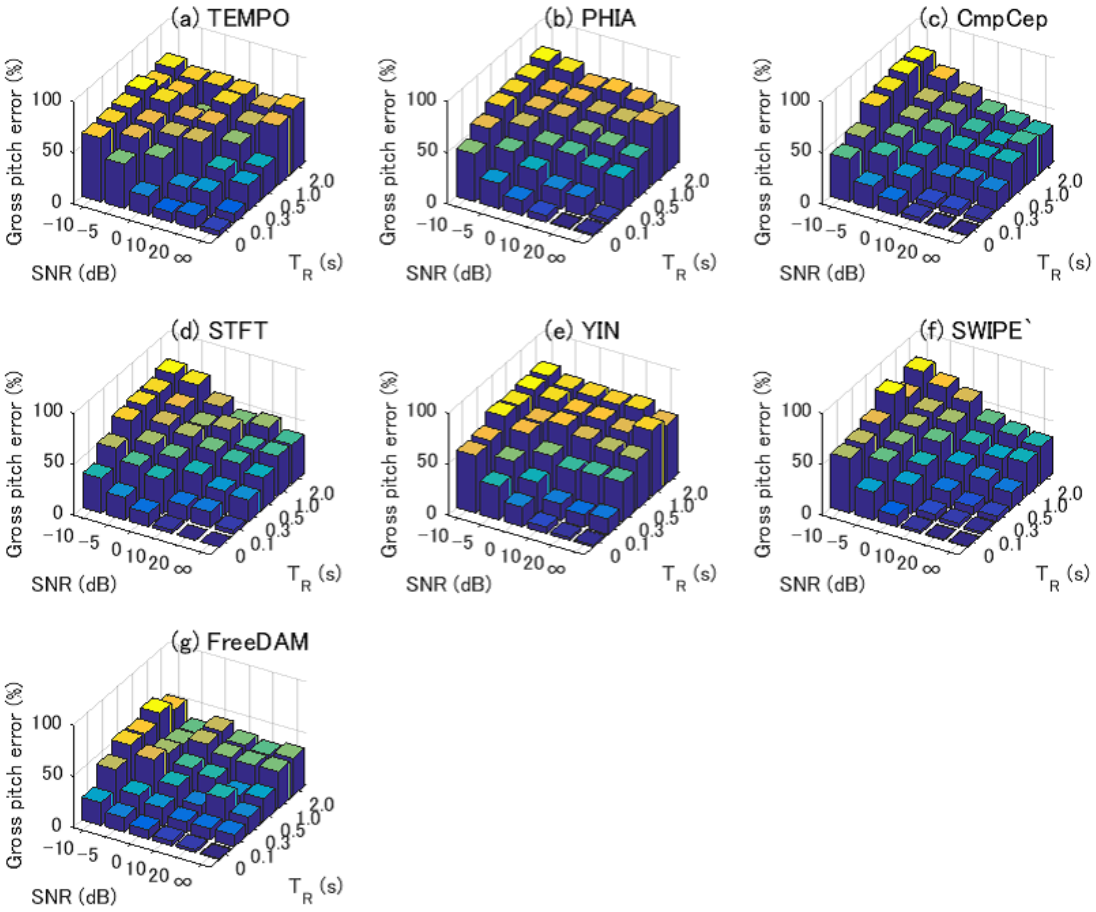

図 B.9: 雑音残響環境における実音声/aoi/の Gross pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）

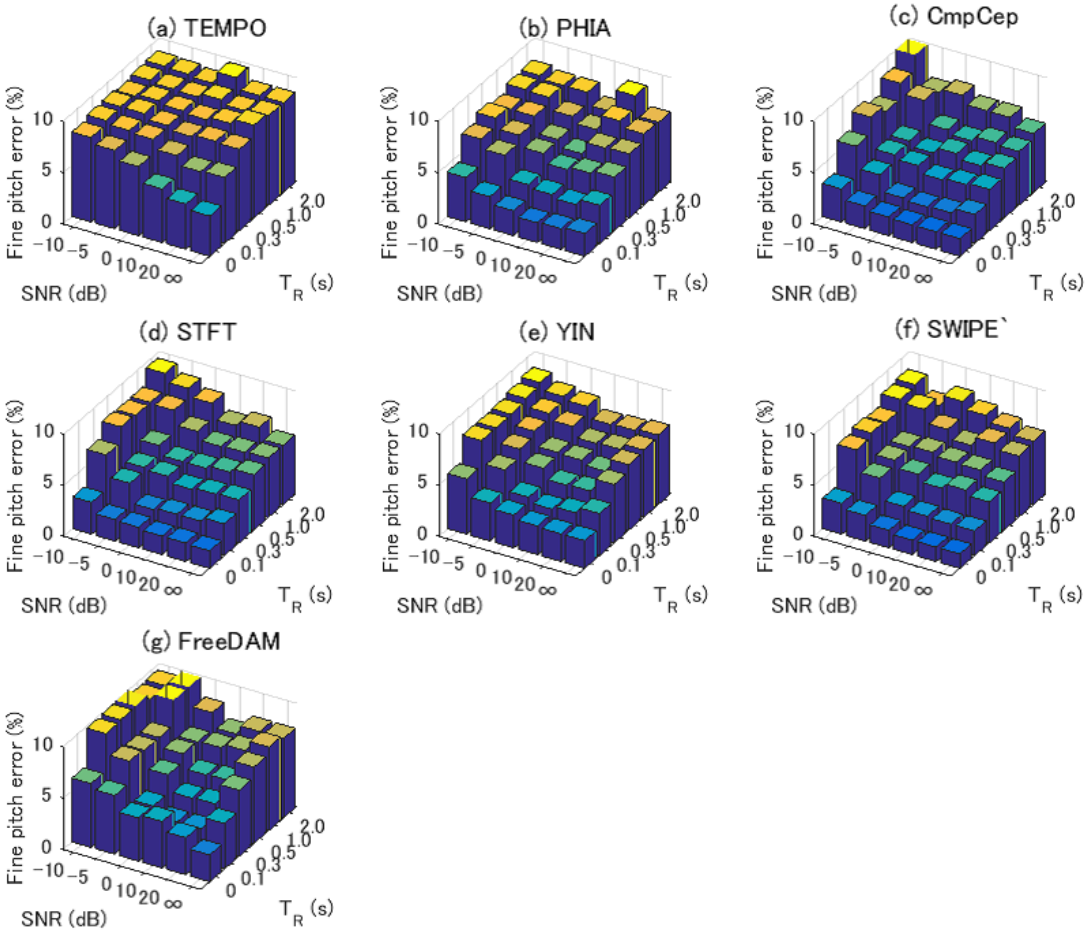

図 B.10: 雑音残響環境における実音声/aoi/の Fine pitch error： (a) TEMPO, (b) PHIA, (c) 複素ケプストラム法, (d) 短時間フーリエ変換, (e) YIN，(f) SWIPE’, (g) FreeDAM（Proposed）