

Title	複素ケプストラム分析を用いた残響音声の基本周波数推定に関する研究
Author(s)	細呂木谷, 敏弘
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3603
Rights	
Description	Supervisor: 鶴木 祐史, 情報科学研究科, 修士

修 士 論 文

複素ケプストラム分析を用いた残響音声の基本周
波数推定に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

細呂木谷 敏弘

2007 年 3 月

修 士 論 文

複素ケプストラム分析を用いた残響音声の基本周 波数推定に関する研究

指導教官 鵜木祐史 准教授

審査委員主査 鵜木祐史 准教授

審査委員 赤木正人 教授

審査委員 徳田功 准教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

510091 細呂木谷 敏弘

提出年月: 2007 年 2 月

概 要

音声の基本周波数は様々な音声信号処理において使われており、非常に重要な特徴である。そのため、多くの基本周波数推定法が提案されてきたが、残響に頑健な基本周波数推定法は報告されておらず、また非常に困難であると考えられている。そこで本論文では、残響音声から精度良く基本周波数を推定する手法を提案することを目的とする。残響音声から基本周波数を精度良く推定できれば、実環境での様々な音声信号処理に応用が可能である。

まず、従来の基本周波数推定法が残響環境下で機能するのを見極めるため、従来の基本周波数推定法を概説し、その中で代表的な10個の手法について耐残響性の評価を行った。評価は、人工的に非最小位相特性を持つ残響インパルス応答を作成し、計算機上で行った。その結果、従来の手法はいずれも残響の影響を大きく受け、ほとんど正しく基本周波数を推定できていないことが分かった。また、従来法の中ではケプストラム法が比較的精度が高かったことから、ケプストラム分析が残響に対して有効な処理であることが分かった。

この評価結果を踏まえ、複素ケプストラム分析を用いた基本周波数推定法を提案した。提案法は、複素ケプストラム分析を用いて音源情報と声道情報を分離し、声道情報を取り除くことで、同時に残響の影響を少なくし、残った音源情報の周期性や調波性から基本周波数を推定する。従来法の評価と同様の評価を行い、従来法と比較した結果、提案法は従来法を上回る推定精度であった。また、クリーンな状況での推定精度も高かったことから、提案法は幅広い環境で使える手法であることが分かった。従って、複素ケプストラム分析により抽出した音源情報は残響に対して頑健な特徴であることが分かった。

提案法は従来法に比べ、残響に対して頑健であったが、大幅な改善とは至らなかったの、さらに推定精度を上げるため、音声に大きく影響を与える残響の特徴量を調査した。その結果、残響の非最小位相の位相成分が音声に大きく影響を与えており、この成分を取り除くことができれば、さらに高精度に残響音声から基本周波数を推定できることが分かった。そこで、この残響の非最小位相成分をケプストラム平均除去法を用いて取り除くことを検討した。しかし、非最小位相ケプストラムに対しケプストラム平均を求めることには問題があったため、ケプストラム平均除去法では効果がなかった。

目次

第1章	序論	1
1.1	基本周波数推定の重要性	1
1.2	基本周波数推定の問題	2
1.3	本研究の目的	3
1.4	本論文の構成	3
第2章	従来の基本周波数推定法の耐残響特性	7
2.1	従来の基本周波数推定法	7
2.1.1	周期性の特徴を利用した手法	7
2.1.2	調波性の特徴を利用した手法	8
2.1.3	時間領域に現れる周期性と周波数領域に現れる調波性の両方を利用した手法	9
2.2	残響環境について	9
2.3	耐残響特性の評価	10
2.3.1	評価方法	11
2.3.2	評価結果	12
第3章	残響環境における基本周波数推定法の提案	15
3.1	残響環境における基本周波数推定法の構成	15
3.2	複素ケプストラム分析を用いた音源情報の抽出	15
3.2.1	複素ケプストラム分析	16
3.2.2	音源情報の抽出	17
3.3	基本周波数推定	18
第4章	残響環境における提案法の有効性の評価	24
4.1	評価方法	24
4.2	評価結果	24
4.3	考察	25
第5章	過去のフレームの影響を考慮した残響に頑健な基本周波数推定法の検討	27
5.1	長時間窓長での残響除去の検討	27
5.1.1	残響の影響の調査	27

5.1.2	位相と振幅の影響	28
5.1.3	最小位相成分と非最小位相成分の影響	28
5.2	残響の非最小位相成分除去の検討	28
第 6 章	結論	34
6.1	本稿で明らかにしたこと	34
6.2	今後の課題	35

第1章 序論

1.1 基本周波数推定の重要性

音声は肺から押し出された呼気が声帯や声道を通り，口唇から放射され作られる音波であり，一般に，声帯の振動を伴う有声音と振動を伴わない無声音に大別される．図 1.1 に発声区間の模式図を示す．有声音は，声帯の振動によって発生した呼気流の準周期的なパルスで声道が励振されて生じる．有声音には，図 1.2 のようなほぼ相似的な波の繰り返しがみられ，この繰り返しの周期を一般に基本周期と呼び，その逆数を基本周波数と呼んでいる．基本周期は声帯によって作られる励振波の周期と等しい．また，有声音の周波数スペクトルは，図 1.3 のような，基本周波数とその整数倍の周波数成分からなる調波構造をもっている．

基本周波数は聴感上は音の高さ（ピッチ）に深く関係している．従って，基本周波数の違いは男声女声や個人間の音色の違いとして現れ，その時間的变化はイントネーションやアクセントなどの音韻性以外の情報（韻律情報）を表現する．人間が音声を知覚する上で韻律情報が重要な役割を担っていることから，基本周波数は言語学的な研究に利用されているが，その他にも工学的な応用として，音声認識や話者認識，音声の分析合成など様々な音声信号処理において利用されている [1] ．

例えば，音声認識においては，音声の韻律的特徴を句や単語の細分化に利用したり，韻律辞書を作成し基本周波数の概形を利用することで，認識精度を向上することができる [2] ．またボコーダ型の音声分析合成においても基本周波数は重要な役割を担っている．この分析合成では，分析部において音源情報とスペクトル包絡情報に分け，低ビットレートで符号化伝送し，合成部で再合成を行っている．このとき，音源情報として基本周波数が必要になってくるため音声の分析合成にとって基本周波数抽出は重要な課題である．特に自然性の高い音声合成のためには高精度の基本周波数推定が必要である．声質変換や話速変換においても，自然性を損なわない変換をするためには，基本周波数の早く細かい変動に追従しなければならない．また，聴覚の情景解析の研究において，人間は様々な音の中から目的の音だけを聞き分けるときに基本周波数を利用していると考えられている．聴覚の情景解析，特に音源分離においては，雑音環境において目的音声の基本周波数を利用した計算論的聴覚情景解析の研究がなされている．

1.2 基本周波数推定の問題

基本周波数推定に関する研究は古くから行われてきたが、次のような理由からまだ根本的には解決していない [1, 3] .

1. 音声は口唇から放射された音波であるため、基本周波数を表す声帯振動を直接観測することができない。また音声波は声帯振動による励振波が声道と共振して生成されるが、音声波から声道の影響を取り除き、励振波のみを取り出すことは困難である。
2. 音声波形は完全な周期信号ではなく、振幅や周期が励振ごとに変化する準周期的な信号である。従って、励振波の変化に追従した推定を行わなければならない。
3. 観測される音声には、状況により様々な雑音や残響の影響を受けている。これらが音声の音源情報を表す特徴を歪ませ、正確な推定を困難にしている。

これまでの研究は主に 1 . と 2 . の問題に主眼を置いてきた。そのため過去 50 年にわたり相当数の基本周波数推定法が提案されてきた [4] が、これらの手法は音声はクリーンな状況で観測されるものと仮定されてきた。その成果として、近年では実際に計測された声帯振動（例えば EGG）の情報と比較して、高精度に基本周波数を推定できる手法も報告されてきている（例えば、TEMPO[5], YIN[6]）。しかし、今後実環境においての音声信号処理を行うには、上記の 3 . の問題への対応が重要になってくる。実環境下での音声信号処理は、雑音や残響の影響により大きく影響を受け性能が低下する。例えば、実環境での音声認識に置いては、周囲の雑音や部屋の残響によって認識精度が大幅に低下してしまう問題がある。もし、実環境での音声から基本周波数を推定できれば、様々な音声信号処理を実環境で行うことができる。例えば、音声認識においては、推定した基本周波数を句や単語の細分化に利用したり、特徴量の一つとして認識器に用いることで認識精度を向上させることができる [7]。また、音源分離システムを利用した補聴器が考えられているが、周囲の音から目的とする音声を強調するための手がかりとして基本周波数が使われている。

以上から、実環境での音声信号処理を行うには、雑音や残響に頑健な基本周波数推定法が必要不可欠である。そのためには従来の基本周波数推定法が雑音や残響の環境の中でどれだけ正しく機能するか見極め、さらに正しく機能しないのであれば、雑音や残響に頑健な基本周波数推定法を構築しなければならない。近年、この問題に対し、雑音環境については研究され始めている [8]。この研究によると、従来の代表的な手法はいずれも雑音の影響を大きく受け、雑音環境では機能していない。その上で、著者らは雑音に頑健な手法を提案し、その有効性を示している。また、他にも耐残響性のある手法が報告されてきている [9, 10, 11, 12]。しかし、残響環境に対しては、ほとんど議論されてない。残響に頑健な基本周波数推定法は報告されていないし、一般的に残響環境下では基本周波数を推定することは非常に困難であると考えられている。しかし、実環境において残響は必ず存在し、音声信号処理システムの性能はその影響を大きく受ける。例えば、音声認識において

は、一般的な住宅程度の残響であっても認識精度が著しく低下する．従って、実環境での音声信号処理のためには、残響に頑健な基本周波数推定法を構築することが必要である．

1.3 本研究の目的

実環境での音声信号処理のためには、雑音、残響の両方に頑健な基本周波数が必要である．しかし、残響に頑健な基本周波数推定法は提案されているが、残響に頑健な基本周波数推定法はまだない．残響に頑健な基本周波数推定法が構築できれば、雑音に頑健な手法と組み合わせたりすることで両方に対応できる手法を構築することができると考えられる．従って、残響に頑健な手法を構築することが急務である．そこで、本論文では残響に頑健な基本周波数推定法を構築することを目的とする．

この目的のためには、まず従来のクリーンな音声を対象とした手法や、雑音に頑健な手法が残響環境下で正しく機能するか見極める必要がある．その結果から、音声のどのような特徴が残響に頑健な特徴であるかを調査し、残響に頑健な基本周波数推定法を提案する．その後、この提案法の耐残響特性を評価することで有効性を示す．最後に、この評価結果から、残響音声に有効な音声の特徴とはどのような特徴であるか考察を行い、提案法の改良点を検討する．

1.4 本論文の構成

本論文は全 6 章から構成されている．本章では、残響環境での基本周波数推定の重要性を説明し、本論文の目的を述べた．

第 2 章では、従来の代表的な基本周波数推定法を概説する．残響特性についての説明と本論文での残響の取り扱い方を述べ、従来法の耐残響特性を評価する．

第 3 章では、従来法の評価結果を踏まえ、複素ケプストラム分析を用いた基本周波数推定法を説明する．

第 4 章では、提案法の耐残響特性の評価を行う．

第 5 章では、前章での結果を踏まえ、提案法の改良を検討する．

第 6 章では、本論文で得られた結果を要約し、今後の課題を述べる．

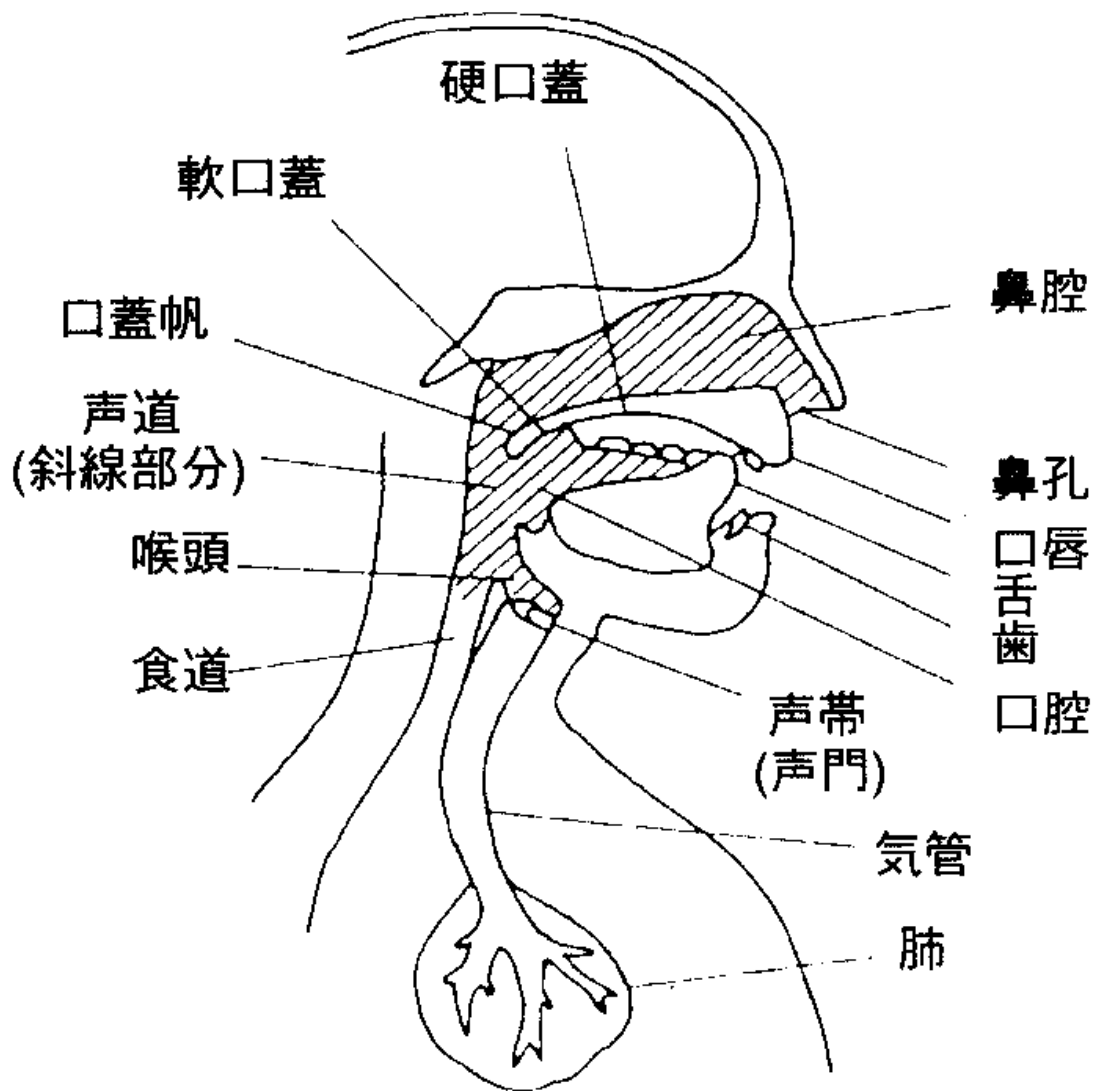


図 1.1: 発声器官の模式図

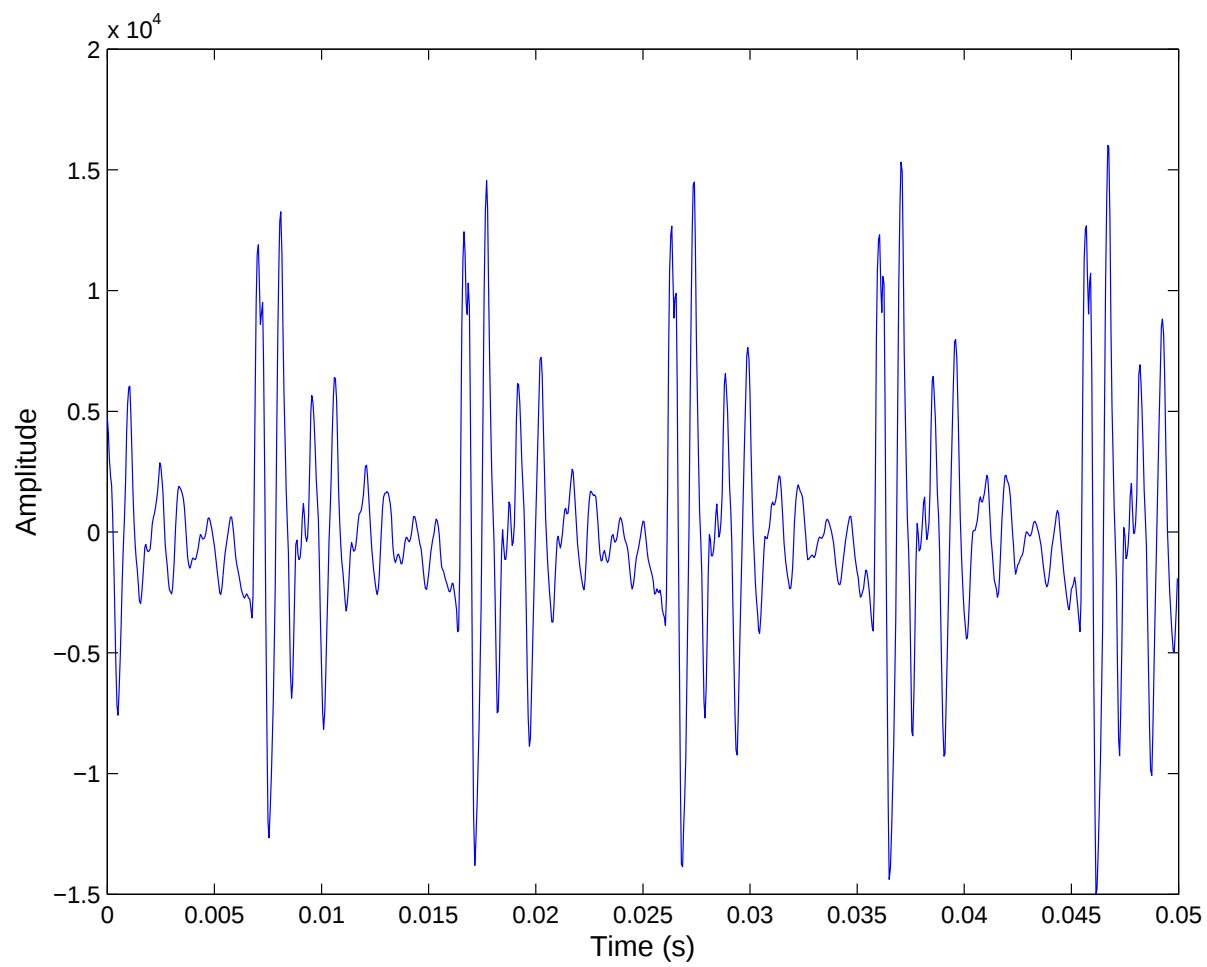


図 1.2: 有音声の時間波形 (男声/a/)

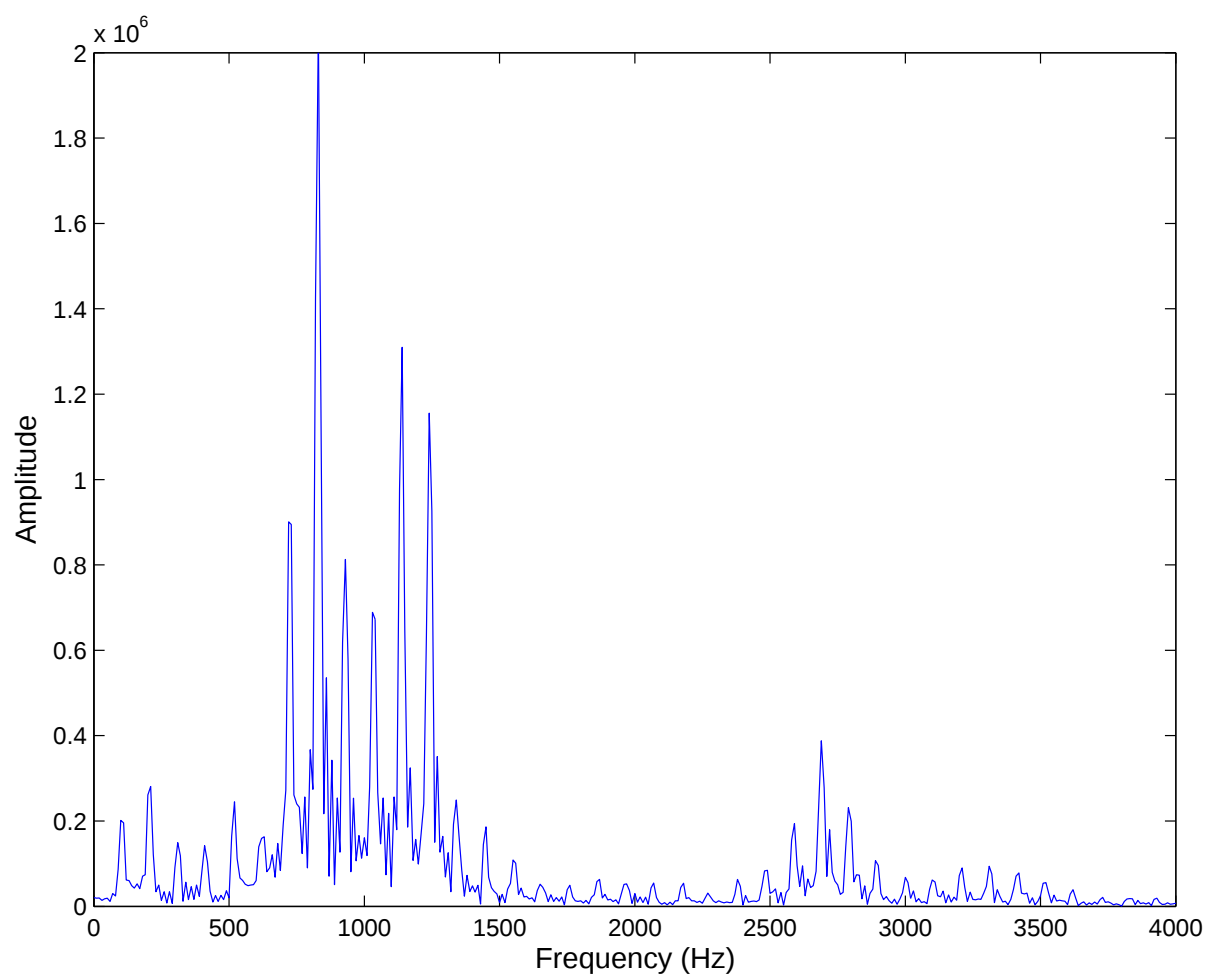


図 1.3: 図 1.2 に示した有声音の振幅スペクトル (男声/a/)

第2章 従来の基本周波数推定法の耐残響特性

2.1 従来の基本周波数推定法

従来の基本周波数推定法について述べる．従来の基本周波数は大別すると，

1. 時間領域に現れる周期性の特徴を利用した手法
2. 周波数領域に現れる調波性の特徴を利用した手法
3. 時間領域に現れる周期性と周波数領域に現れる調波性の両方を利用した手法

に分類できる．図 1.2 に示したように，有声音は，時間方向に基本周波数の逆数である基本周期を周期とした波形となるため，この周期を検出しようとするものが，時間領域に現れる周期性の特徴を利用した手法である．また，図 1.3 に示したように，周波数領域では，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のように現れるため，この調波成分の内の最低周波数や調波成分間隔を検出しようとするものが，周波数領域に現れる調波性の特徴を利用した手法である．時間領域に現れる周期性と周波数領域に現れる調波性の両方を利用した手法は，これらを統合し，両方の特徴を利用した手法である．

また，別の観点から眺めると，

1. 音源・フィルターモデルを仮定した手法
2. 音源・フィルターモデルを仮定しない手法

という分類もできる．音源・フィルターモデルとは，声帯振動を音源情報，声道特性をフィルター情報とし，音声音を音源情報とフィルター情報の畳み込みとして表現したモデルである．従って，音源・フィルターモデルを仮定した手法は，音源情報のみを抽出したり，フィルター情報を取り除いたりした後に基本周波数を推定する手法である．

2.1.1 周期性の特徴を利用した手法

基本周波数推定の初期の研究においては，零交差法 [1, 13] やピーク位置検出法 [1, 14] が利用された．零交差法とは，時間波形上で振幅が 0 となる点を微分処理などで検出し，

その時間間隔から基本周波数を求める手法である．ピーク位置検出法とは，同じく時間波形上の振幅のピーク値を微分処理で検出し，その時間間隔から基本周波数を求める手法である．これらの手法は実装が容易である利点があるが，精度が良くなく，実環境の影響を大きく受けやすい．

周期性を利用する手法としては，自己相関法 [4] も古くから用いられている．この手法は，自己相関係数の最大点から波形の周期間隔を求めている．この手法には，変形自己相関法やクリッピング処理（波形の時間平均値以下を 0 とする処理）を利用したり，様々な改良法がある．また，自己相関処理と関係して，自己相関係数の計算回数を減らすため，自己相関係数の代わりに平均振幅差関数（AMDF）を利用した手法 [1, 16] がある．近年では，複数の窓長を用いて自己相関関数を計算し，最適な基本周波数を選択することにより，誤推定を避けようとする手法（ACMWL 法 [15]）や，波形の振幅差関数を重み付けする手法（YIN[6]）などが提案されている．

音源・フィルターモデルを仮定する手法としては，LPC 残差法がある．この手法は，音声で LPC 分析することにより声道伝達特性を推定し，残差として残った音源情報から自己相関法を用いて周期性を検出する手法である．

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により，音声の時間 - 周波数表現を得ることができる．この時間 - 周波数表現は音声の調波構造を現している．短時間フーリエ変換を用いる手法には，対数スペクトルから自己相関法を用いて基本周波数を推定する手法 [1, 3, 4, 17] や，対数スペクトル上の調波性を Comb フィルタ（ある基本周波数成分とその調波成分が櫛の歯のように並んでいるフィルタ）を用いて検出する手法 [18] がある．

短時間フーリエ変換を利用し，音源・フィルターモデルを仮定する手法では，ラグ窓法 [19]，リフター法 [1, 4] がある．これらの手法はいずれも，対数振幅スペクトル上で，声道特性を推定し，その逆特性をかける手法がラグ窓法，声道特性を打ち消す手法がリフター法である．

調波性を利用する他の手法に，SHS 法 [20]，LPC-SHIFT 法 [21] がある．SHS 法は，対数周波数上の対数振幅スペクトルの調波性の荷重和から基本周波数を推定する手法である．LPC-SHIFT 法は，上述した LPC 分析を用いて，声道特性の逆フィルタをかけることで声道特性を打ち消し，基本周波数を推定する．

調波性を利用し音源・フィルターモデルを仮定した手法にケプストラム法 [22] がある．ケプストラムとは対数振幅スペクトルを逆フーリエ変換することで求めることができる．ケプストラムとは，ケフレンシーと呼ばれる時間領域の値であるが，スペクトル上の微細構造がケフレンシー上では高ケフレンシー部に基本周期に対応するピークとして現れ，スペクトル包絡が低ケフレンシー部に現れる．従って，ケフレンシー上で音源情報とフィルター情報を分離することができる．この高ケフレンシー部のピークを検出する手法がケプストラム法である．ケプストラム法には，クリッピング処理を行ってからケプストラム

を求めるクリップストラム法 [23] や、移動平均処理や帯域制限を追加した改良ケプストラム法 [24]、ハフ変換を行い時間連続性を考慮することで耐雑音性を高めた手法（ハフ変換法 [9]）などの改良法がある。

近年、音声の瞬時周波数から高精度に基本周波数を推定する手法が提案されている [5, 11, 12, 25]。代表的なものとして河原らの TEMPO [5] が挙げられる。この手法は、フィルタ中心周波数におけるフィルタ出力の瞬時周波数への写像の安定な不動点を検出することで高精度に基本周波数を推定している。

2.1.3 時間領域に現れる周期性と周波数領域に現れる調波性の両方を利用した手法

近年、耐残響特性を持つ手法として PHIA 法 [8] が提案されている。この手法は、まず瞬時周波数の周波数領域での周期性と時間領域での調波性から、雑音に頑健な基本周波数推定を行い、求めた基本周波数を利用して雑音抑圧を行い、最後に瞬時周波数を用いて高精度な推定を行う手法である。

2.2 残響環境について

残響とは、音源が発音を停止した後も音が響いて聞こえる現象である。人間が生活する空間は常に残響の影響があり、日常生活において重要な現象である。そのため、我々は無響室などの全く残響のない空間において発声された音声を聞くと非常に不自然に感じる。しかし、機械による音声信号処理のシステムにとっては、残響の影響が非常に大きく、問題となっている。閉空間で観測される音源は、音源からの直接と、壁や床、天井からの反射音から構成されるが、この反射音が残響である。現在、残響を定量的に現す物理量として残響時間が広く用いられている。残響時間とは、室内の音源が停止した後、音圧レベルが定常状態から 60 dB 減衰するまでの時間のことを言う。なお残響時間の例として、一般的な住宅では 0.5 s、コンサートホール等では 2 s 程度の残響時間となる。

Kuttruff [26] によると、統計的に残響を観測すると、残響の伝達特性とは、包絡線が指数関数的に減衰し、微細構造が白色雑音である以下のインパルス応答 $h(t)$ で表すことができる。

$$h_r(t) = a \exp\left(\frac{-6.9t}{T_R}\right) \cdot n(t) \quad (2.1)$$

$$a = \sqrt{\frac{1}{\int_0^T \exp\left(\frac{-13.8t}{T_R}\right) dt}} \quad (2.2)$$

ここで、 T_R はパワーが 60 dB 落ちるまでの残響時間、 $n(t)$ はランダム雑音、 a はパワーを正規化するための定数である。図 2.1 に残響時間が 1 s の場合の残響インパルス応答の例を示す。

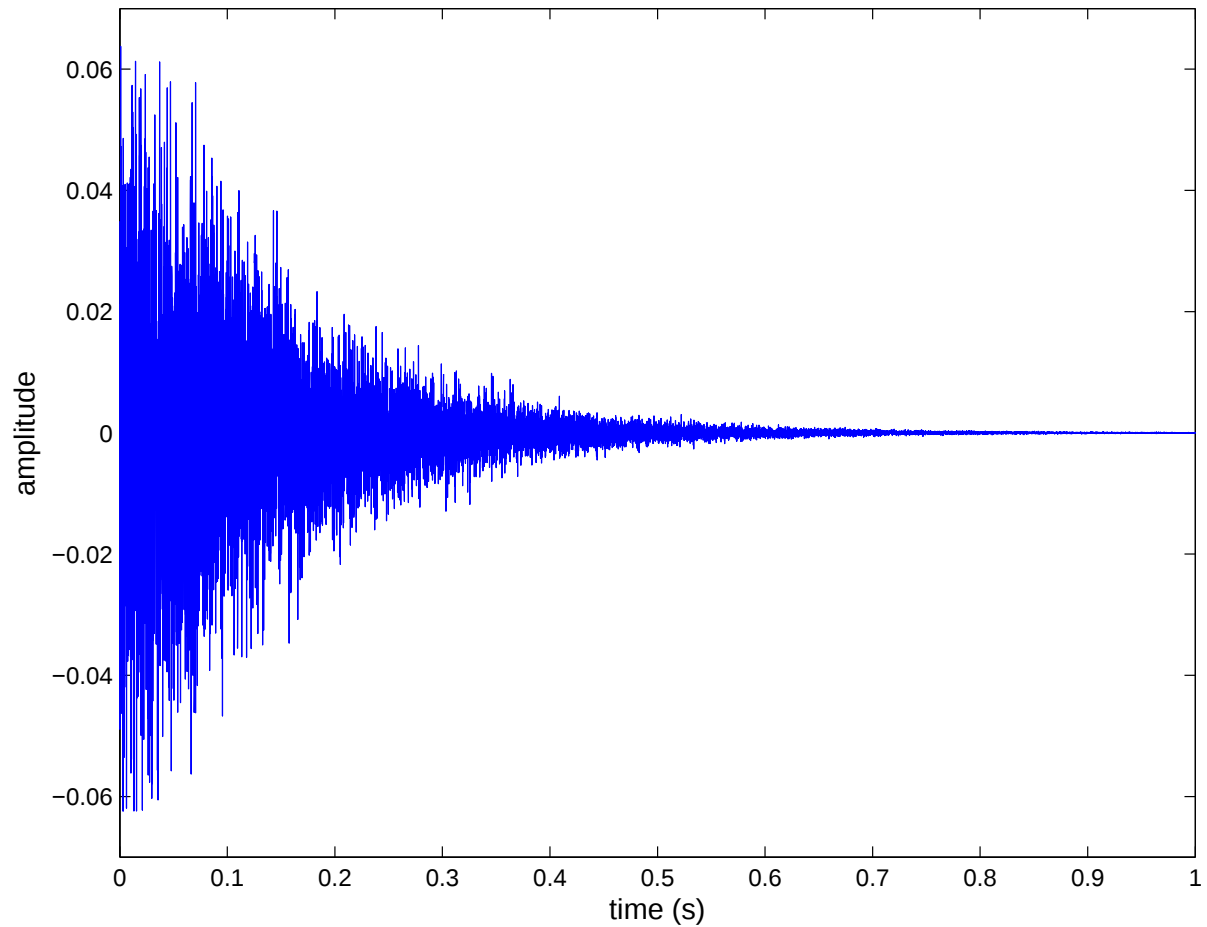


図 2.1: 非最小位相特性を持つ残響インパルス応答

2.3 耐残響特性の評価

本節では、従来の基本周波数の対残響特性の評価を行う．2.1 節で述べたように、従来の基本周波数推定法の数是非常に多いため、全ての手法について評価するのは困難である．そこで本論文では、利用している特長量が様々であることや比較的に良く知られ利用されていること、クリーンな状況で高精度であったり雑音に頑健などの特徴を持っていたりすることを基準に 10 個の手法を取り上げ耐残響特性の評価を行った．対象とした手法は

- 自己相関法
- 複数窓長を用いた自己相関法（ACMWL 法）
- AMDF 法
- LPC 残差法

- SHS 法
- STFT-Comb フィルター法
- リフター法
- ケプストラム法
- TEMPO
- PHIA 法

である．

2.3.1 評価方法

評価に使用したデータは，阿竹らによる音声と EGG が同時収録された音声データベース [12] を用いた．本稿では，そのデータベース内の男女各 12 名，4 文章を使用した．サンプリング周波数は 16 kHz である．本論文での音声への残響特性の付加は，2.1 節で説明した残響インパルス応答 $h(t)$ を計算機上で畳み込みを行って実現した．評価条件として，残響時間を $T_R = 0, 0.1, 0.3, 0.5, 1, 2$ s の 6 種類とした．

評価尺度は，正答率と SNR の二つの評価尺度を用いた．一般的には正答率のみで評価される場合もあるが，本論文では SNR も考慮した．正答率を求めるには，まず EGG 波形から TEMPO で基本周波数推定を行い，これを正解の基本周波数として $F_0(t)$ とする．次に各推定法にて音声波形から基本周波数推定を行い，これを $F'_0(t)$ とする．このとき， $F_0(t)$ に対して誤差 5 % 以内にある $F'_0(t)$ の区間の長さが，有声区間の長さに占める割合を正答率とする．正答率は次式で定義される．

$$\text{Correct Rate} = (N_c/N_v) \times 100 \quad (2.3)$$

N_v は有声区間の区間長， N_c は $F_0(t)$ と $F'_0(t)$ との誤差が 5 % 以内の区間長である．従って，正答率は有声区間内で，5 % の推定誤差の範囲内で正しく推定できた区間の割合である．正答率は正しく推定できた区間を表すことから，残響に対する頑健性を示す．

SNR とは，5 % の推定誤差の範囲にある区間において， $F_0(t)$ を信号， $F_0(t) - F'_0(t)$ を雑音とした時の信号対雑音比であり，次式で定義される．

$$\text{SNR} = 20 \log_{10} \frac{\int_0^T F_0(\tau) d\tau}{\int_0^T (F_0(\tau) - F'_0(\tau)) d\tau} \quad (\text{dB}) \quad (2.4)$$

ここで， τ は， $F'_0(t)$ が推定誤差 5 % 以内の範囲にある時間である．SNR は基本周波数を推定できた区間内における，推定した基本周波数の誤差を表すことから，推定精度を表す．

2.3.2 評価結果

図 2.2, 図 2.3 に各推定法の推定誤差 5%以内の正答率と SNR を示す．この結果から，どの手法も残響時間が長くなるにつれ推定精度が著しく低下し，残響時間が 2 s の場合では 50%を下回る正答率となっていることが分かる．残響時間が 0 s の場合，すなわち残響の影響がないクリーンな状況では TEMPO が 95 %を超える正答率で最も精度が高い．ケプストラム法はクリーンな状況でこそ TEMPO に劣るが，残響時間が長くなっても比較的正答率が低下していない．PHIA 法は雑音に対して頑健な手法であるが，残響に対しては他の手法と同様に大きく影響を受けている．SNR を見ると，ほぼ全ての手法が残響時間が長くなるにつれ，精度が低下している．残響時間が短い場合は TEMPO が他の手法よりも精度が高いが，残響時間が長くなるにつれ，他の手法と大差がない状況となっている．以上の結果から，従来法の中で最も残響に頑健な手法はケプストラム法であることが分かる．また，雑音と残響とでは違った対応をしていく必要があることが分かる．また，周期性と調波性の分類の観点から眺めてみても，どちらの特徴を利用しても残響には大きく影響を受けることが分かる．

以上の結果から，ケプストラム分析が残響に対して有効な処理であることが考えられる．この理由として，ケプストラム処理は畳み込みに対しての準同型変換であり，音声の反射音と直接音を同一に扱うことができ，遅延の影響をあまり受けなかったためであると考えられる．

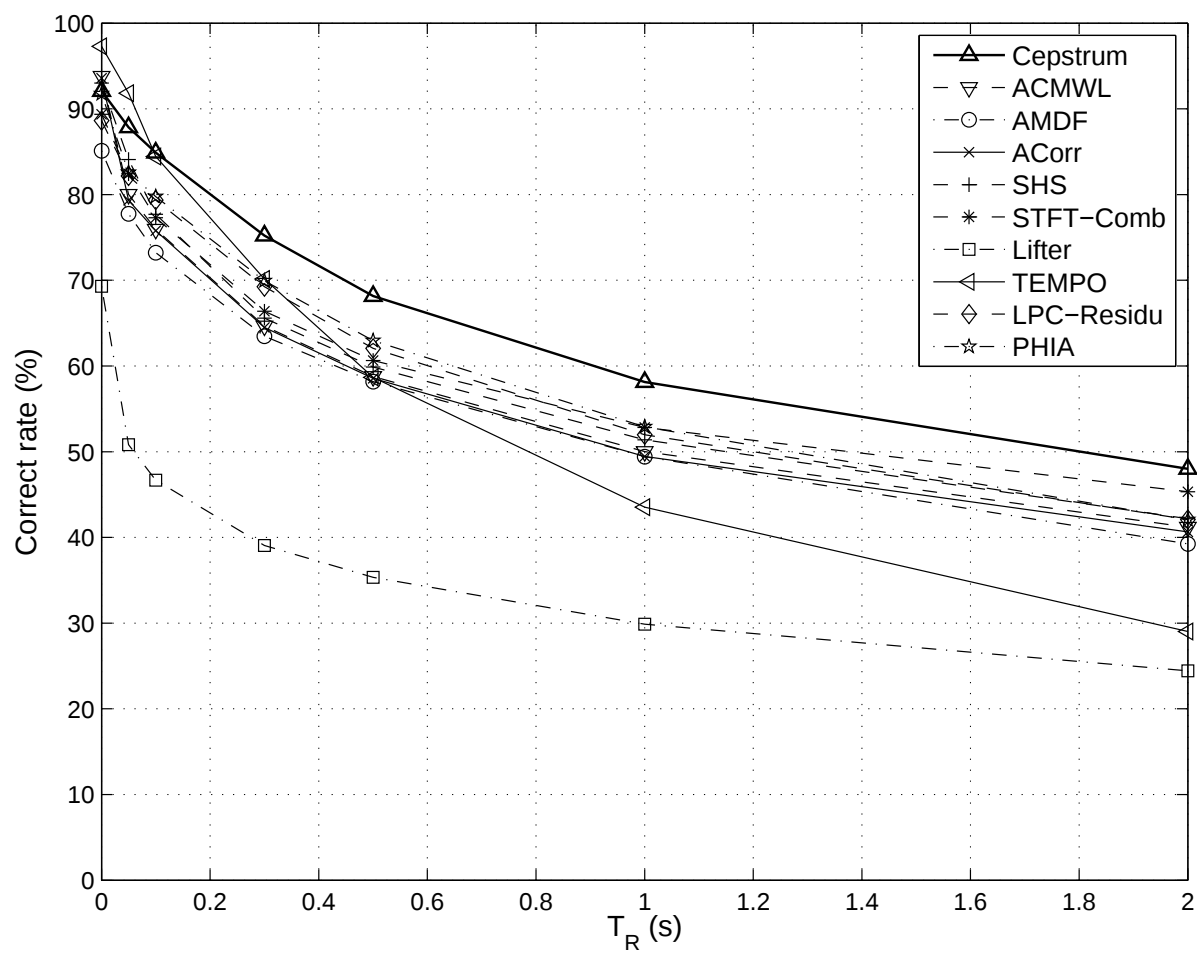


図 2.2: 従来法の評価 (正答率)

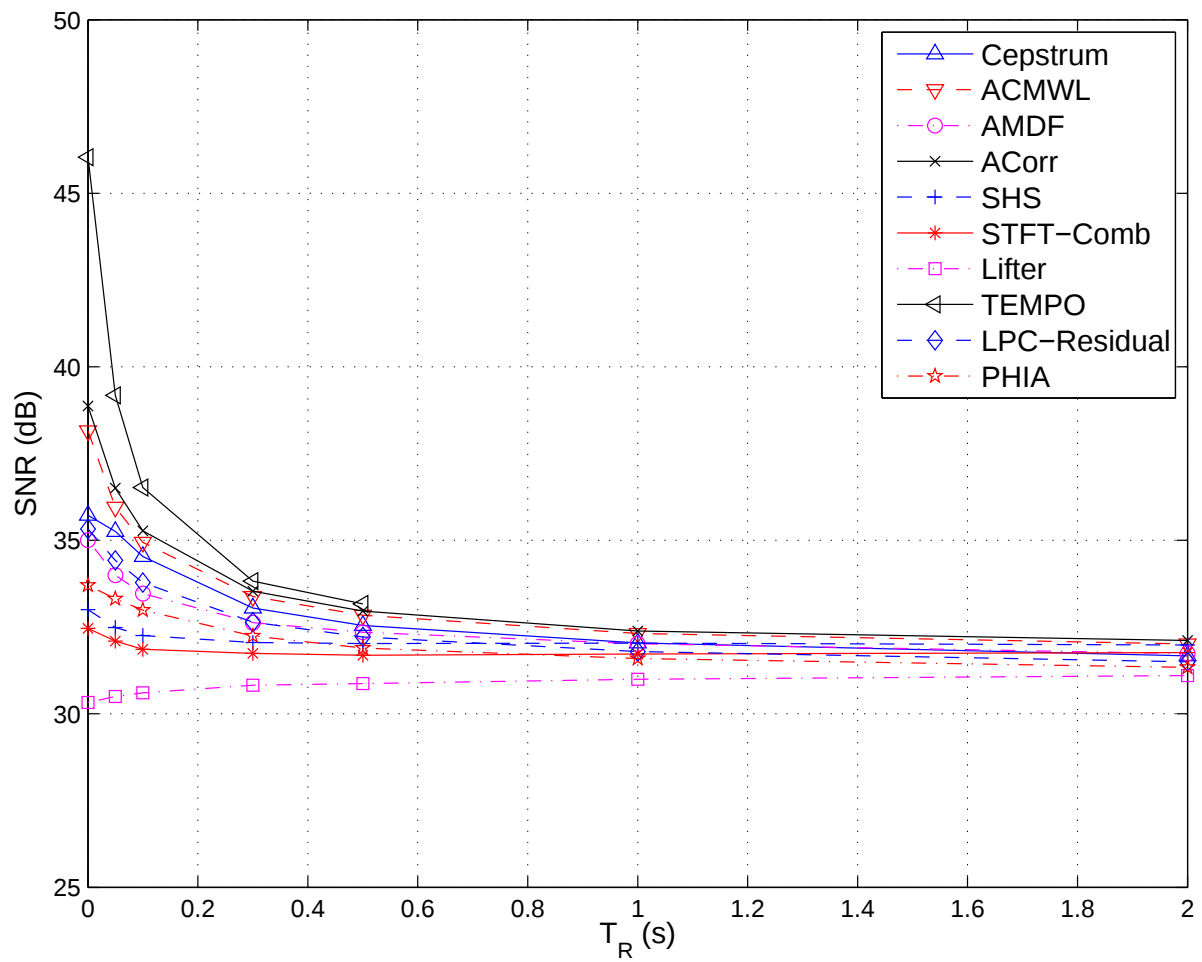


図 2.3: 従来法の評価 (SNR)

第3章 残響環境における基本周波数推定法の提案

本章では、従来法の評価結果を踏まえ、残響に頑健な基本周波数推定法を提案する。第1節において、提案法の考え方と構成を述べ、第2節、第3節において、その詳細を説明する。

3.1 残響環境における基本周波数推定法の構成

第2章の結果から、ケプストラム分析が残響に対して有効な処理であることが分かった。準同型処理の利点を活かすため、提案法ではケプストラムに基づいた処理をベースとする。ケプストラムは、その性質から低ケフレンシー部の振幅が大きく、ケフレンシーの増加につれて減衰する特性を持っている。したがって残響の伝達特性と音声のケプストラムは低ケフレンシー部にパワーが集中していると考えられる。ここで、音源・フィルターモデルを仮定すると、音声のケプストラムは、低ケフレンシー部に声道情報、高ケフレンシー部に音源情報とに分離する。従ってリフターにより高ケフレンシー部の成分だけを取り出せば、基本周波数推定に必要である音源情報のみを抽出することができる。同時にこの操作は、残響の低ケフレンシー部の成分を捨てることができ、残響の大きな成分を取り除くことにもなる。

この考えに基づき提案法を構築する。提案法は2つのブロックで構成される。前半部はリフタリングを行い、残響音声から低ケフレンシー部にある声道情報と残響の主要な成分を取り除く。後半部は残った音源情報から周期性や調波性を利用して基本周波数を推定する。

3.2 複素ケプストラム分析を用いた音源情報の抽出

本節では、提案法の前半部の処理である、残響音声からの音源情報の抽出の説明を述べる。まず、数学的な準備として複素ケプストラム分析の説明を行い、その後、提案法の前半部の説明を行う。

3.2.1 複素ケプストラム分析

観測信号を $x(t)$ ，複素スペクトルを $X(\omega)$ とすると，複素ケプストラム $C(q)$ は，対数複素スペクトル $\log X(\omega)$ のフーリエ逆変換として表される．なお $F[\bullet]$, $F^{-1}[\bullet]$ はそれぞれフーリエ変換，フーリエ逆変換を表す．

$$C(q) = F^{-1}[\log X(\omega)] \quad (3.1)$$

ここで，複素スペクトルを振幅スペクトル，位相スペクトルを用いて記述しなると

$$\begin{aligned} C(q) &= F^{-1}[\log\{|X(\omega)| \exp(j\phi(\omega))\}] \\ &= F^{-1}[\log |X(\omega)|] + F^{-1}[j\phi(\omega)] \\ &= F^{-1}[\log A(\omega)] + F^{-1}[j\phi(\omega)] \end{aligned}$$

となる．ここで，

$$\begin{aligned} C_A(q) &= F^{-1}[\log A(\omega)] \\ C_\phi(q) &= F^{-1}[j\phi(\omega)] \end{aligned}$$

とすると，複素ケプストラムは

$$C(q) = C_A(q) + C_\phi(q) \quad (3.2)$$

となり，振幅ケプストラム $C_A(q)$ と位相ケプストラム $C_\phi(q)$ の和で表現される．このように複素ケプストラムを用いることにより，振幅と位相の関係を，スペクトルでの積の表現から和の表現にかえて分析することができる．

複素スペクトルにおいて，振幅スペクトルと位相スペクトルによる表現の他に，最小位相特性と非最小位相特性による表現がある． $X_{\min}(\omega)$ を最小位相スペクトル， $X_{\text{all}}(\omega)$ を非最小位相スペクトルとすると，複素スペクトルは

$$X(\omega) = X_{\min}(\omega) \cdot X_{\text{all}}(\omega)$$

と表される．上式を複素ケプストラムを用いて表すと

$$\begin{aligned} C(q) &= F^{-1}[\log X(\omega)] \\ &= F^{-1}[\log X_{\min}(\omega)] + F^{-1}[\log X_{\text{all}}(\omega)] \end{aligned}$$

となり，これをまとめると

$$C(q) = C_{\min}(q) + C_{\text{all}}(q) \quad (3.3)$$

$$C_{\min}(q) = F^{-1}[\log X_{\min}(\omega)] \quad (3.4)$$

$$C_{\text{all}}(q) = F^{-1}[\log X_{\text{all}}(\omega)] \quad (3.5)$$

と表すことができる．ここで， $C_{\min}(q)$ は最小位相ケプストラム， $C_{\text{all}}(q)$ は非最小位相ケプストラムである．また，複素ケプストラムの偶関数，奇関数の関係から最小位相ケプストラムは

$$C_{\min}(q) = \begin{cases} 2C_A(q) & q > 0 \\ C_A(q) & q = 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

より求めることができる．

以上から，複素ケプストラムは，振幅，位相ケプストラムの和としての表現や，最小位相，非最小位相ケプストラムの和として表現できる．この関係を図 3.1 に示す．この図を眺めると，複素ケプストラムの最小位相，非最小位相特性と振幅，位相特性の対応関係を読みとることができる．

3.2.2 音源情報の抽出

音源・フィルターモデルを仮定すると，観測信号 $x(t)$ は，音源情報 $s(t)$ と声道伝達特性 $f(t)$ と残響伝達特性 $h(t)$ の畳み込みで表現される．

$$x(t) = s(t) * f(t) * h(t) \quad (3.7)$$

この式の周波数領域での表現は，

$$X(\omega) = S(\omega) \cdot F(\omega) \cdot H(\omega) \quad (3.8)$$

となり，時間領域での畳み込みの関係は周波数領域では積の関係となる．対数を取り，逆フーリエ変換をすることで，

$$C_x(q) = C_s(q) + C_f(q) + C_h(q) \quad (3.9)$$

となる．ここで， $C_x(q)$ 、 $C_s(q)$ 、 $C_f(q)$ 、 $C_h(q)$ はそれぞれ観測信号，音源情報，声道伝達特性，残響特性の複素ケプストラムである．このように，ケフレンシー領域では，信号を畳み込みの関係から和の関係で表現することができる．また，複素ケプストラムは振幅ケプストラムと位相ケプストラムとの和の関係にあるので，それぞれの複素ケプストラムは，

$$C_x(q) = C_{A,x}(q) + C_{\phi,x}(q)$$

$$C_s(q) = C_{A,s}(q) + C_{\phi,s}(q)$$

$$C_f(q) = C_{A,f}(q) + C_{\phi,f}(q)$$

$$C_h(q) = C_{A,h}(q) + C_{\phi,h}(q)$$

となる．従って，観測信号の振幅ケプストラムは

$$C_{A,x}(q) = C_{A,s}(q) + C_{A,f}(q) + C_{A,h}(q) \quad (3.10)$$

となる．ここで，振幅ケプストラムとは，対数スペクトルにおける周波数方向への振動を表すから，スペクトル上で緩やかな変動を示すフィルタの伝達特性（声道特性）と，速い振幅変動を示す音源情報（調波構造）は，それぞれ低ケフレンシー部，高ケフレンシー部に分離して現れる．従ってリフターにより高ケフレンシー部の成分だけを取り出せば，基本周波数推定に必要である音源情報のみを抽出することができる．この関係を図 3.2 に示す．また，残響特性のケプストラムは低ケフレンシー部にパワーが集中しており，ケフレンシーの増加につれてパワーが減衰していく特性を持っている．従って，観測信号の振幅ケプストラムには，低ケフレンシー部に声道特性と残響特性の主要な成分，高ケフレンシー部に音源情報と残響の残りの成分が存在している．つまり，リフターにより高ケフレンシー部の成分だけを取り出せば，声道特性と残響の主要な成分を取り除くことができる．提案法では以下のリフターを用いて音源情報を抽出した．

$$l(q) = \begin{cases} 0 & q \leq q_{\text{lif}} \\ 1 & q > q_{\text{lif}} \end{cases} \quad (3.11)$$

ここで，リフターのカットオフケフレンシーは 400 Hz 以下の基本周波数を通すように $q_{\text{lif}} = 2.5 \text{ ms}$ とした．リフタリング後の信号は，

$$C_{A,x'}(q) = C_{A,s}(q) + C_{A,h'}(q) \quad (3.12)$$

となる． $C_{A,h'}(q)$ は，低ケフレンシーの成分が取り除かれた残響の振幅ケプストラムである．

以上で音源情報の抽出は行えた．しかし，振幅ケプストラムだけであると，位相情報がないので，今後時間波形に戻すことができなくなってしまう．元の残響音声の位相情報を用いてもよいが位相情報も残響の影響を受けていると考えられる．そこで前節の式 (3.6) を用いて振幅ケプストラムから最小位相ケプストラムを求めることで位相情報を生成する．得られた最小位相ケプストラムは，

$$C_{\min,x'}(q) = C_{\min,s}(q) + C_{\min,h'}(q) \quad (3.13)$$

となる．各項の最小位相ケプストラムは，

$$C_{\min,x'}(q) = C_{A,x'}(q) + C_{\phi,\min,x'}(q)$$

$$C_{\min,s}(q) = C_{A,s}(q) + C_{\phi,\min,s}(q)$$

$$C_{\min,h'}(q) = C_{A,h'}(q) + C_{\phi,\min,h'}(q)$$

となり，振幅ケプストラムと最小位相ケプストラムの和となっている．

3.3 基本周波数推定

後半部の処理は，抽出した音源情報から基本周波数を推定する．この処理は音源情報の周期性を取り扱うか，調波性を取り扱うかによって様々な方法が考えられる．本論文で

は、周期性を取り扱う方法として時間領域での自己相関法を利用したもの、調波性を取り扱うものとして周波数領域での振幅スペクトルの comb フィルタ法を利用したものの2種類構築した。自己相関法を利用したアルゴリズムを図 3.3(a)、comb フィルタを利用したアルゴリズムを図 3.3(b) に示す。comb フィルタはガウス核をもつ 10 次までのフィルタで、指数関数的に減衰する。以下に用いたフィルタを示す。

$$c(\omega; F_0) = \sum_{n=1}^9 0.9^n \exp\left(-\frac{\omega - n \cdot F_0}{2\sigma}\right) \quad (3.14)$$

ここで、 F_0 は推定しようとする基準周波数、 σ はガウス核の標準偏差である。図 3.4 に基本周波数が 100 Hz の高調波成分を検出する Comb フィルタの例を示す。

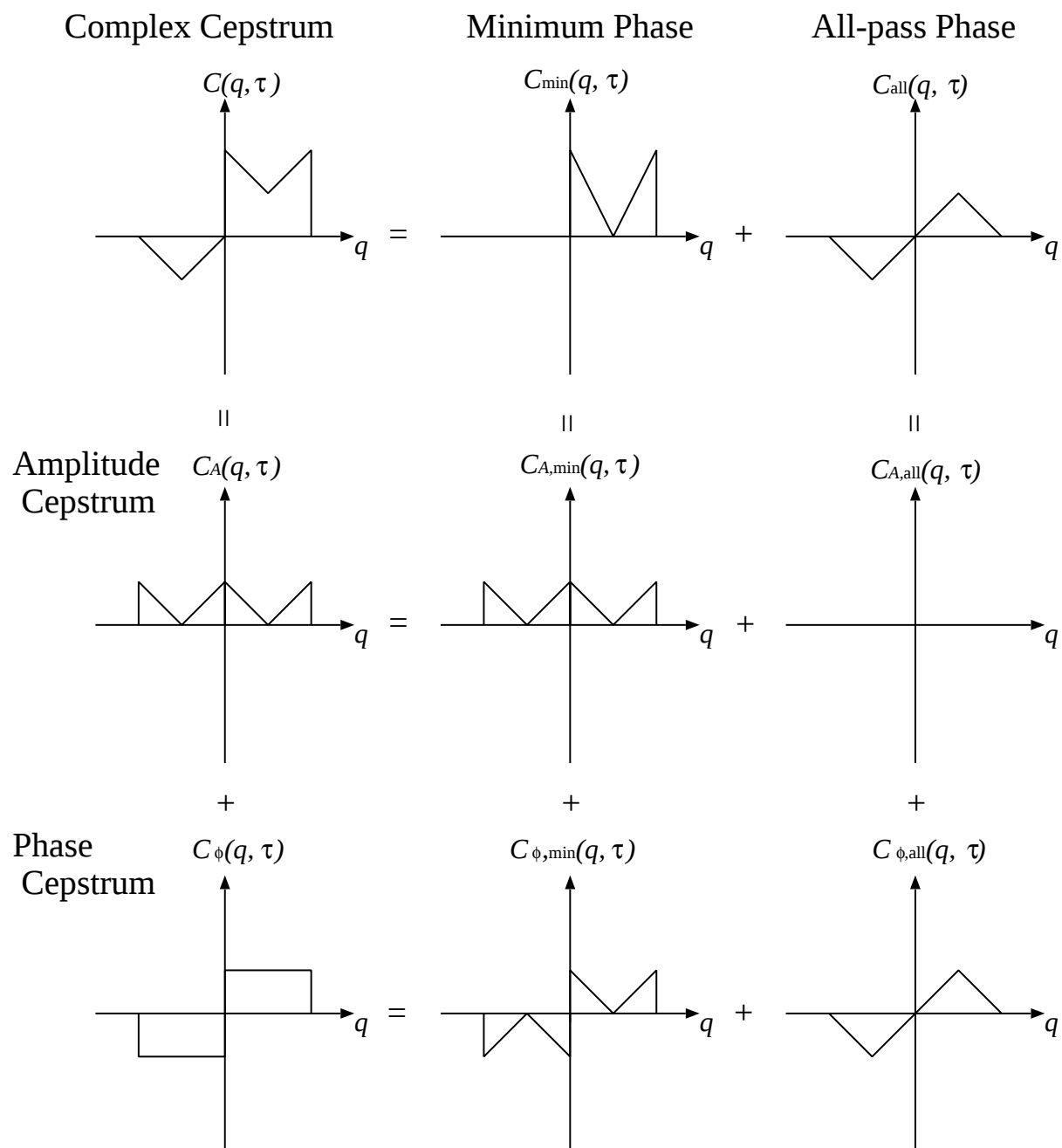


図 3.1: 複素ケプストラムにおける振幅，位相特性と最小位相，非最小位相特性の関係

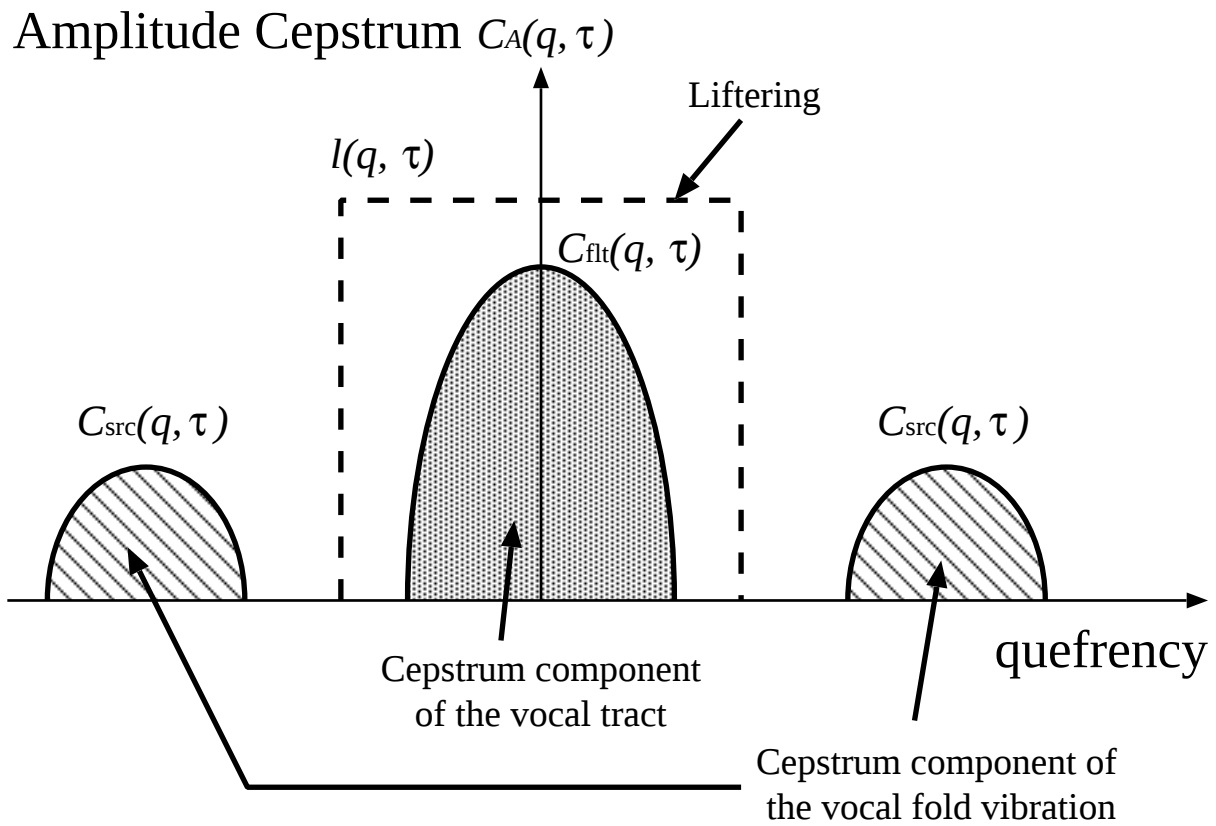


図 3.2: ケフレンシー領域における、音源情報と声道伝達情報の分離

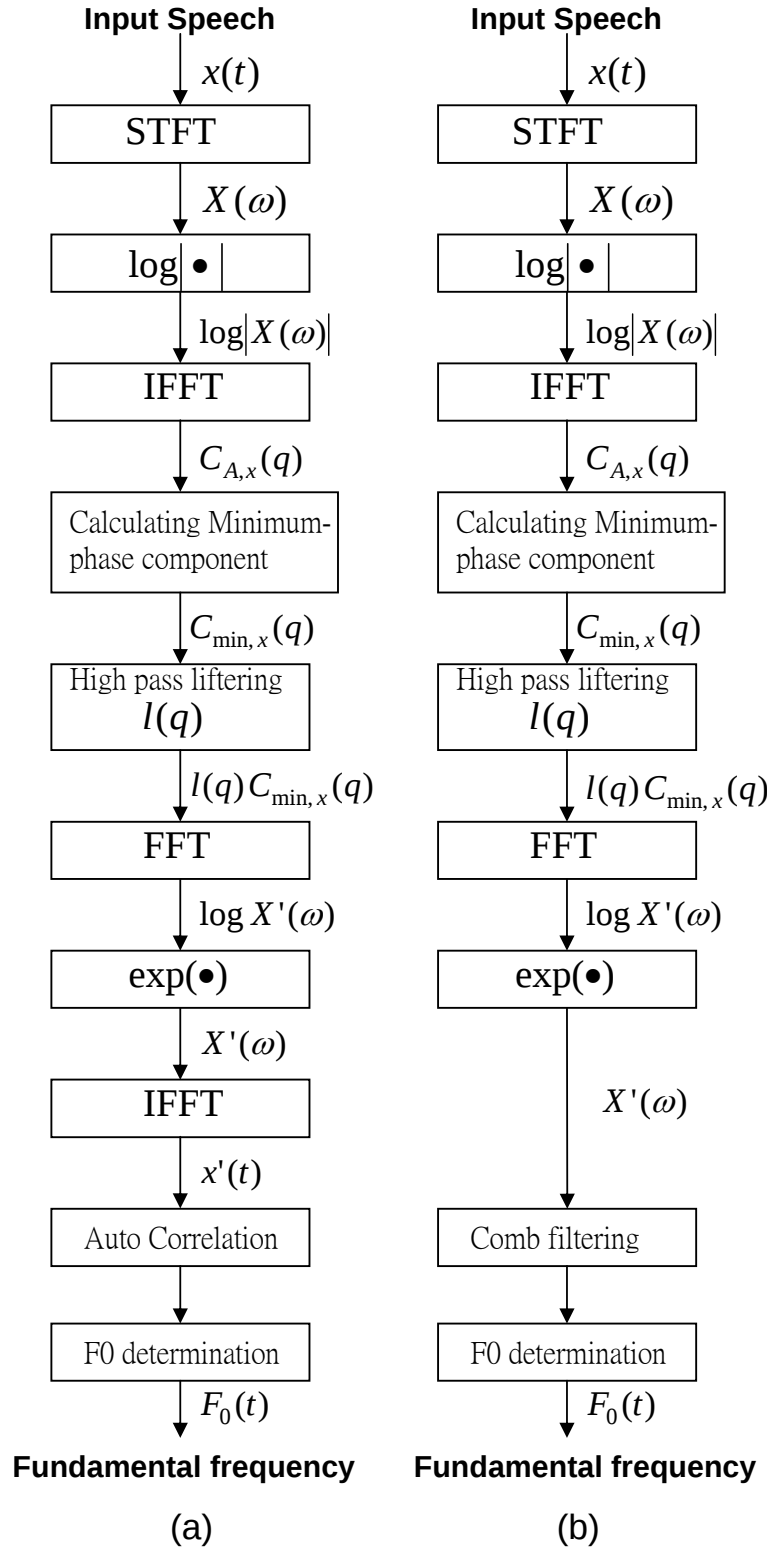


図 3.3: 提案手法のフローチャート. (a) 音源信号に対する自己相関法, (b) 音源の振幅スペクトルに対する comb フィルタ法

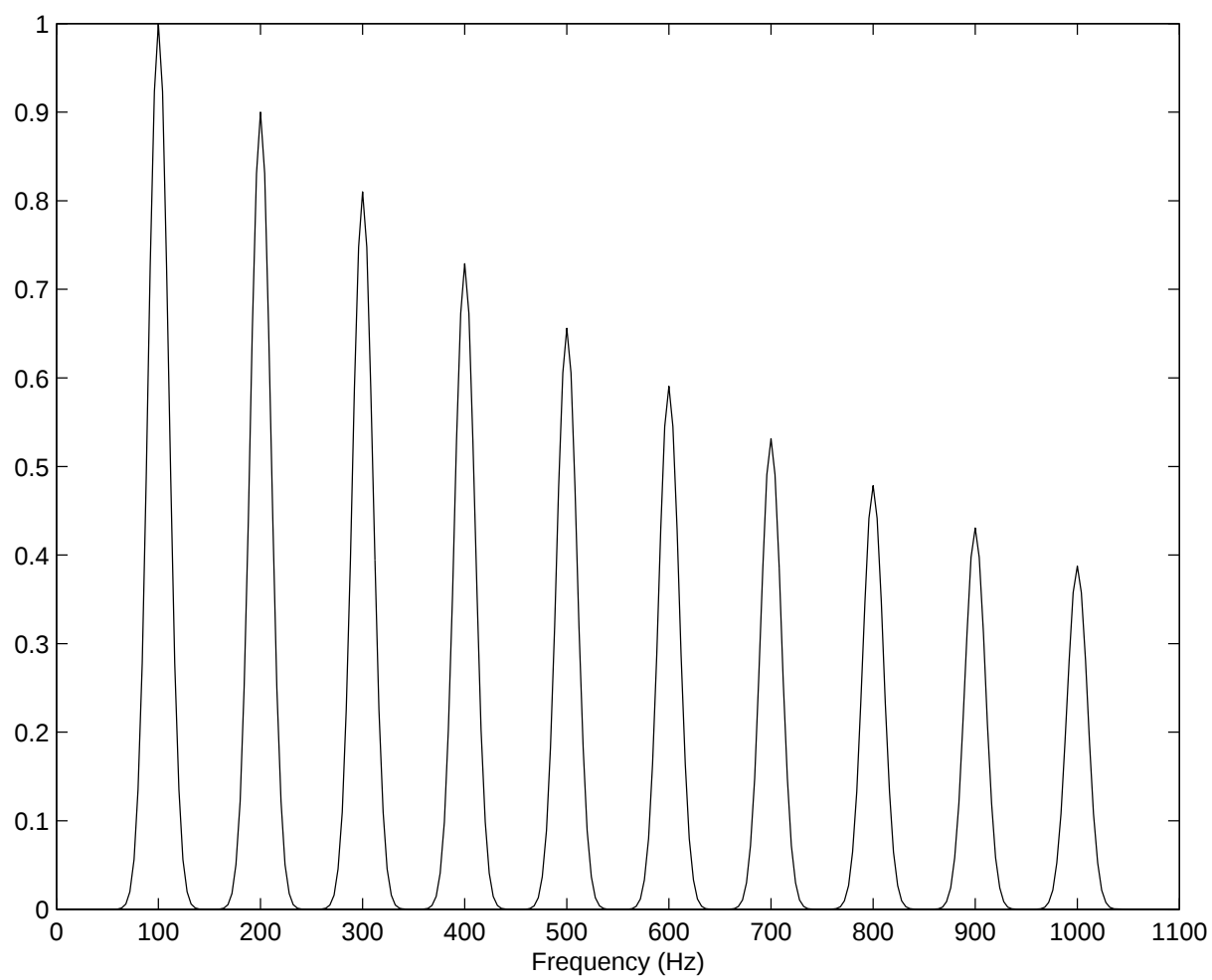


図 3.4: 基本周波数が 100 Hz の調波成分を検出するフィルタ

第4章 残響環境における提案法の有効性の評価

4.1 評価方法

本章では，本研究で提案する基本周波数推定法の耐残響性能を評価する．

残響の付加は第2章での従来法の評価と同じである．評価に使用した音声データは第2章と同じデータベースを用いた．評価尺度についても第2章と同じく，推定誤差5%以内の正答率とSNRを用いた．また，比較のため従来の手法のうち最も残響に頑健であったケプストラム法と，クリーンな状況では最も精度の高かったTEMPOも併せて示す．また，提案法の前半部の処理の残響に対する有効性を調べるため，自己相関法を用いた結果も併せて表示し，提案法の内，後半部の処理に自己相関法を用いた手法と比較する．

4.2 評価結果

図4.1，図4.2に評価結果を示す．まず，正答率を見ると，提案法の内，複素ケプストラム分析を用いて抽出した音源情報の振幅スペクトルからCombフィルターで基本周波数を推定した場合の正答率は，すべての時間においてケプストラム法を上回っている．残響時間が0.5 s以上の比較的長い場合では，3%程の改善となっている．また，残響時間が非常に短い場合（ほぼ残響の影響がない場合）でも，TEMPOより2%程度低いほどで，クリーンな状況においても推定精度の低下がほとんどないことが分かる．提案法の内，複素ケプストラム分析を用いて抽出した音源情報の時間波形から自己相関法を用いて基本周波数を推定した場合の正答率は，ケプストラム法に比べほぼすべての残響時間において3%ほど下回る性能であった．しかし，時間波形の自己相関法と比べると若干の改善があり，複素ケプストラム分析を用いた音源情報の抽出の効果があることが分かる．

SNRで見ると，残響時間が短い場合では，TEMPO，自己相関法よりも精度が悪いが，残響時間が長い場合は，提案法の内，両方の手法とも，従来法を僅かながら上回る精度であった．

以上のことから，提案法の内，複素ケプストラム分析により抽出した音源情報からCombフィルタを利用して推定する手法は従来法に比べ残響に対して頑健であり，また残響の有無に関わらず適用できる手法であることが分かった．また，提案法の内，複素ケプストラム分析により抽出した音源情報から自己相関法を利用して推定する手法と，時間波形か

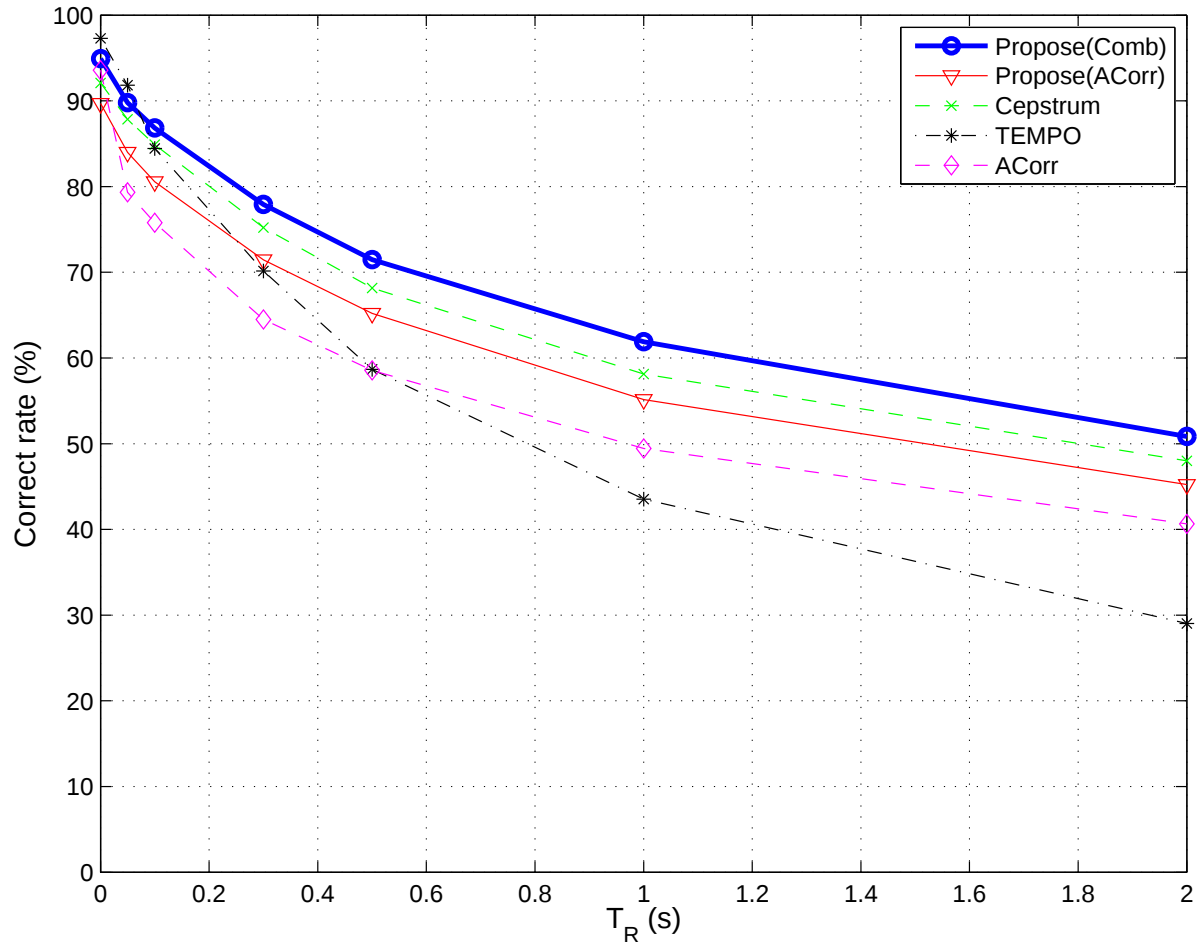


図 4.1: 提案法の評価 (正答率)

らの自己相関法とを比較すると，提案法のほうが残響に対して頑健であったことから，両者の違いである提案法の前半部の処理は残響に対して有効な処理であることが分かった．従って，音源・フィルターモデルを仮定し，ケプストラム上で音源情報を抽出する処理は残響の影響を取り除く有効な処理と言える．

4.3 考察

前節の結果から，提案法は従来法に比べて残響に頑健であることが分かった．しかし，提案法は従来法に比べて劇的に精度が良いわけではなかった．残響のインパルス応答は通常の基本周波数推定法で用いられる窓長よりも長い場合が多いため，過去のフレームの情報が現在のフレームに影響を与えているが，提案法はこの影響を考慮していない．従って，提案法は分析フレーム内の残響の影響はある程度取り除けてはいるが，過去のフレー

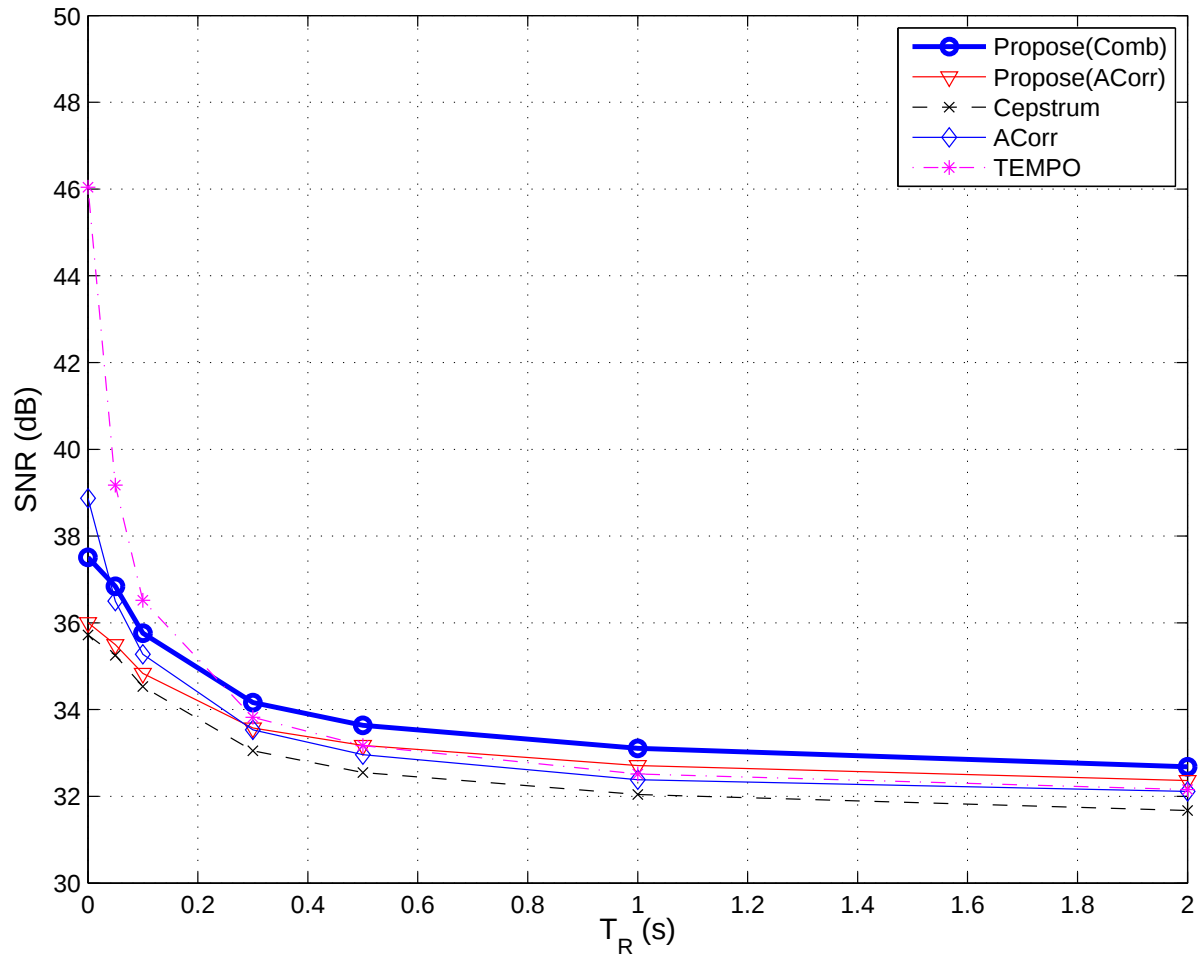


図 4.2: 提案法の評価 (SNR)

ムの影響が大きいため劇的な改善とはならなかったと考えられる。この影響を考慮するため、過去のフレームの情報を積極的に利用する処理や、残響のインパルス応答長よりも長いフレーム長での処理を検討することで大きく精度が向上すると考えられる。次章では、この過去のフレームの影響を考慮し、さらに頑健に残響音声から基本周波数を推定する手法を考察する。

第5章 過去のフレームの影響を考慮した 残響に頑健な基本周波数推定法の 検討

前章の考察から，提案法が従来法を大きく上回らなかった理由として，過去のフレームの情報が推定精度に大きく影響を与えていると考えられる．この影響を取り除くには，二つのアプローチが考えられる．

1. 残響のインパルス応答長が分析フレーム長を超えていることが問題なので，インパルス応答長よりも長い分析フレームを用いる．しかし，基本周波数推定のためにはフレーム長を長くすると，基本周波数の時間変化に追従できなくなる．そこで，残響インパルス応答長を超えるフレーム長で残響除去を行い，残響が除去された波形から基本周波数を推定する．
2. 残響のインパルス応答長が長い場合，現在の観測フレームには過去のフレームの音声が減衰し，歪んだものが重畳されいる．過去のフレームの音声は未知であるが，残響の影響を受けた過去のフレームの音声は既知である．従って，過去のフレームの情報をうまく用いれば，観測フレームに重畳されている過去のフレームの音声を取り除くことができる．このような処理を行った後，基本周波数推定を行う．

本論文では，一つ目のアプローチの検討を行う．

5.1 長時間窓長での残響除去の検討

まず，長時間窓長での残響除去を行うため，音声に対する残響の影響を調査し，どのような特徴量に残響の影響が大きいのかを調べる必要がある．その後，基本周波数推定に影響を与えている原因を切り分け，その影響を取り除く処理を検討する．

5.1.1 残響の影響の調査

基本周波数を推定する上で，残響が音声に与える影響を調べる．

残響とは位相遅れを与えるものと考え、残響伝達特性の位相成分が音声に大きく影響を与えていると考えられる。また、3.1 節の複素ケプストラム分析を用いると残響伝達特性の振幅と位相を最小位相、非最小位相成分に分けることができる。ここで、ケプストラム分析が、エコーのような最小位相特性を持つ単純な遅延を取り除けることから、非最小位相特性を持つ残響の影響は、複素ケプストラム上での非最小位相成分にあることが予想される。この予測を検証するため、シミュレーションを行った。

5.1.2 位相と振幅の影響

残響の位相と振幅のどちらが大きく影響を与えているか調べるため、次の波形を合成し、提案法で基本周波数推定を行い、推定結果を比較した。元音声、残響音声、残響インパルス応答は全て既知としている。

- 振幅を元音声、位相を残響音声のものをういて波形を合成
- 振幅を残響音声、位相を元音声のものをういて波形を合成

なお、用いた音声は女性話者 1 文章「コーヒーにミルクを入れますか」、残響時間は 1 s とした。結果を図 5.1, 5.2 に示す。この結果より、位相を元音声のものをういた場合が推定精度が良好である。従って、残響の位相情報に大きな影響があると考えられる。

5.1.3 最小位相成分と非最小位相成分の影響

前項の結果から残響の影響は位相成分にあることが分かった。次に、この位相成分の内、最小位相成分と非最小位相成分の影響を調べた。図 5.3 は、残響音声から残響インパルス応答の最小位相の位相成分を除去した波形から基本周波数推定を行った結果、図 5.4 は残響インパルス応答の非最小位相の位相成分を除去した波形から基本周波数推定を行った結果である。この結果を見ると、非最小位相の位相成分を除去した信号の基本周波数推定結果が良好であることが分かる。従って、残響の影響は非最小位相の位相成分に大きくあることが分かった。

5.2 残響の非最小位相成分除去の検討

前項までの結果から、残響の影響は非最小位相の位相成分にあることが分かった。従って、この成分を推定し、除去できれば残響の影響を取り除くことができ、高精度に基本周波数推定ができると考えられる。

本論文では、ケプストラム平均除去法を適用し、この成分の除去を検討した。ケプストラム平均除去法とは、音声認識においてマイクロホンなどの音響系の特性を除去するために用いられる。この処理は以下の二つの仮定をおいている。

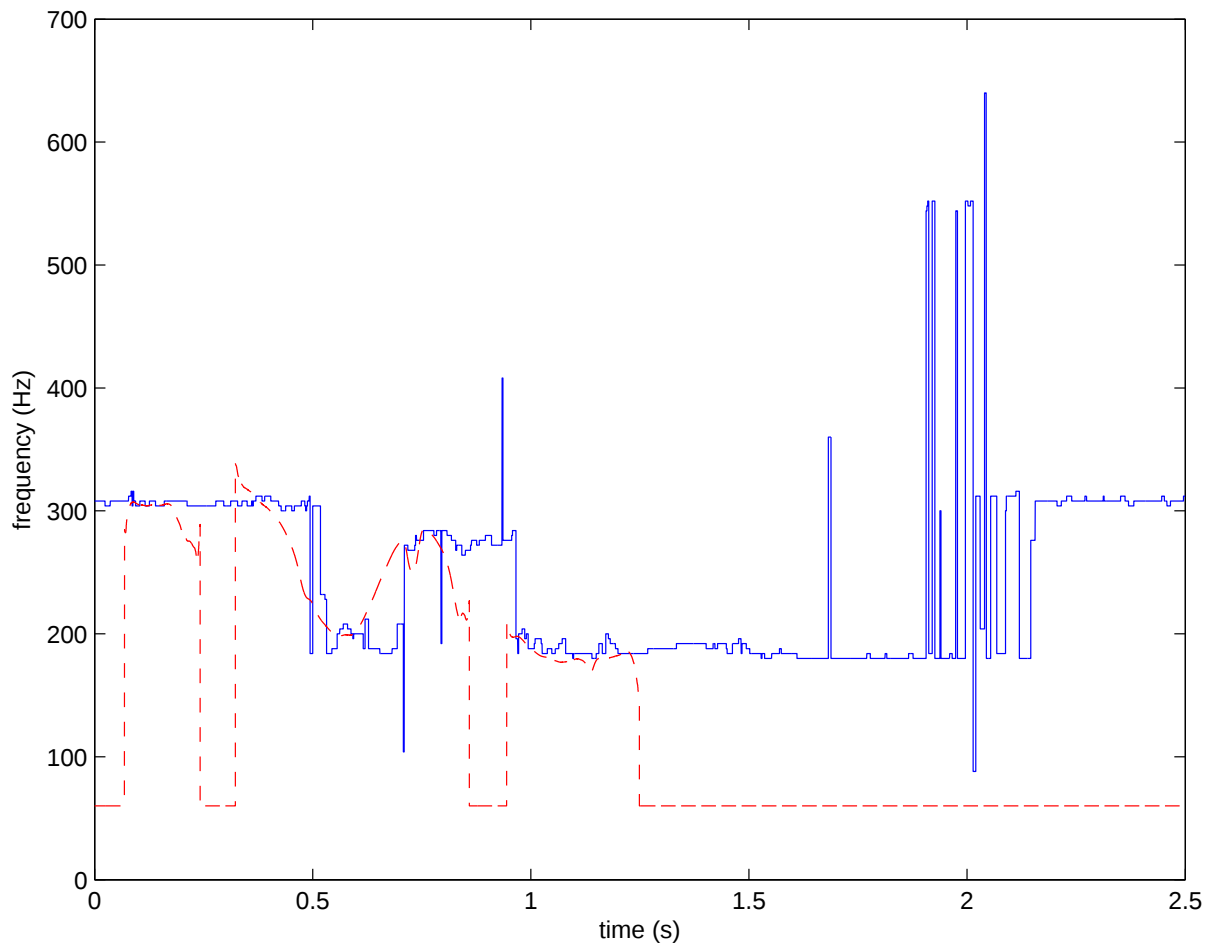


図 5.1: 基本周波数推定結果（振幅：元音声，位相：残響音声）

1. 音響系の特性は定常
2. 音声信号のケプストラム係数を比較的長時間にわたって平均すると一定の値に近づく

この二つを仮定した上で，ケプストラム上で減算を行う．この処理は音響系を対象としているが，残響特性も一文の発話内では定常と考えられるので，この処理が有効であると考えられる．

いま，音声信号に残響時間 1 s の残響特性 $h(t)$ が畳み込まれているとする．残響のインパルス応答長を超えるフレーム長で切り出された複素スペクトルは，

$$X(\omega, n) = S(\omega, n)H(\omega) \quad (5.1)$$

と表される．ここで， n はフレーム番号を表す． $S(\omega, n)$ は音声信号のスペクトル， $H(\omega)$ は残響のスペクトルを表す． $H(\omega)$ はフレームによらず一定であるので，フレーム番号の

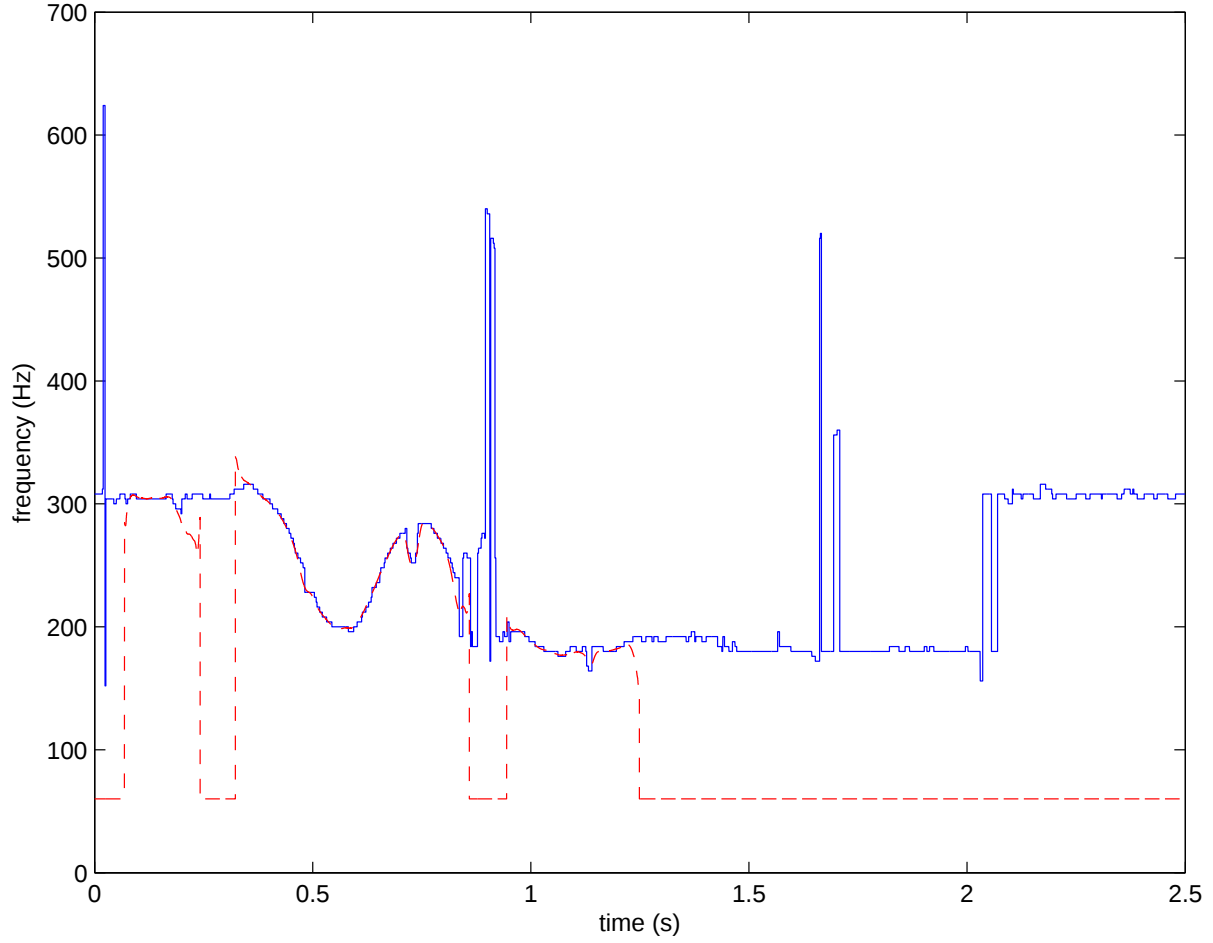


図 5.2: 基本周波数推定結果（振幅：残響音声，位相：元音声）

関数となっていない．また，切り出しに使ったフレーム長は 1.5 s としている．複素ケプストラム領域で表すと，

$$C_x(q, n) = C_s(q, n) + C_h(q) \quad (5.2)$$

となる．複素ケプストラム分析により非最小位相ケプストラムだけを抽出すると，

$$C_{all,x}(q, n) = C_{all,s}(q, n) + C_{all,h}(q) \quad (5.3)$$

となる．この音声区間に渡る時間平均は，

$$\frac{1}{N} \sum_{n=1}^N C_{all,x}(q, n) = \frac{1}{N} \sum_{n=1}^N C_{all,s}(q, n) + C_{all,h}(q) \quad (5.4)$$

である．この平均値を，フレームで切り出される前の長時間の残響音声の複素ケプストラムから減算すると，

$$C_x(q) - \frac{1}{N} \sum_{n=1}^N C_{all,x}(q, n) = C_s(q) + C_h(q) - \frac{1}{N} \sum_{n=1}^N C_{all,s}(q, n) - C_{all,h}(q) \quad (5.5)$$

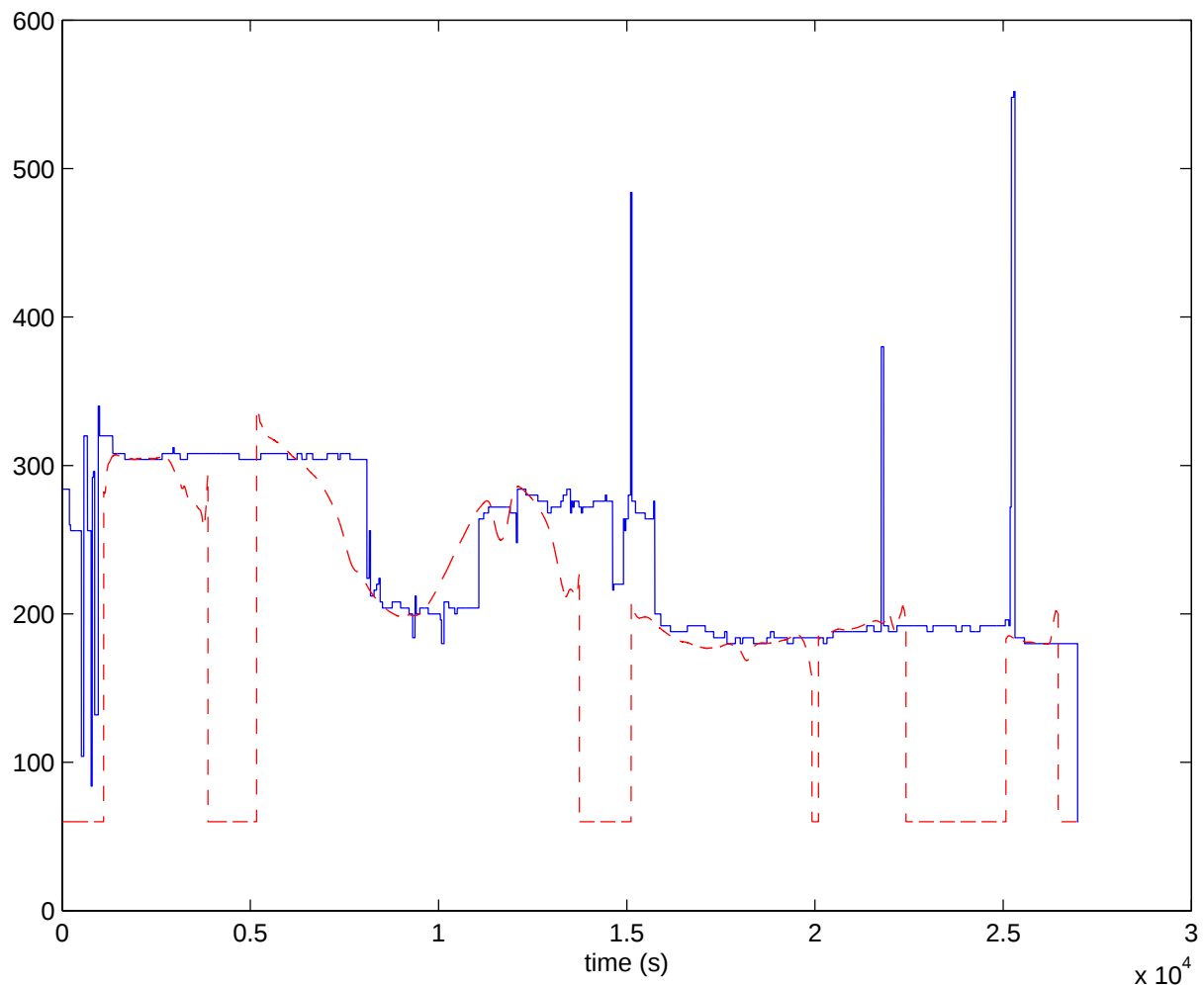


図 5.3: 基本周波数推定結果（最小位相の位相成分除去）

$$= C_s(q) + C_{\min,h}(q) - \frac{1}{N} \sum_{n=1}^N C_{\text{all},s}(q, n) \quad (5.6)$$

$$= C_s(q) + C_{\min,h}(q) - \overline{C_{\text{all},s}(q, n)} \quad (5.7)$$

となり，残響の非最小位相位相ケプストラムが除去されたものとなる．残響の最小位相ケプストラムは残っているが，前節の結果からその影響は少ないと考えられる．

女性話者 1 文章について，上記のケプストラム平均除去法を行い，提案法にて基本周波数推定を行った結果を図 5.5 に示す．この結果を見ると，ケプストラム平均除去法の効果はあまり見られなかった．

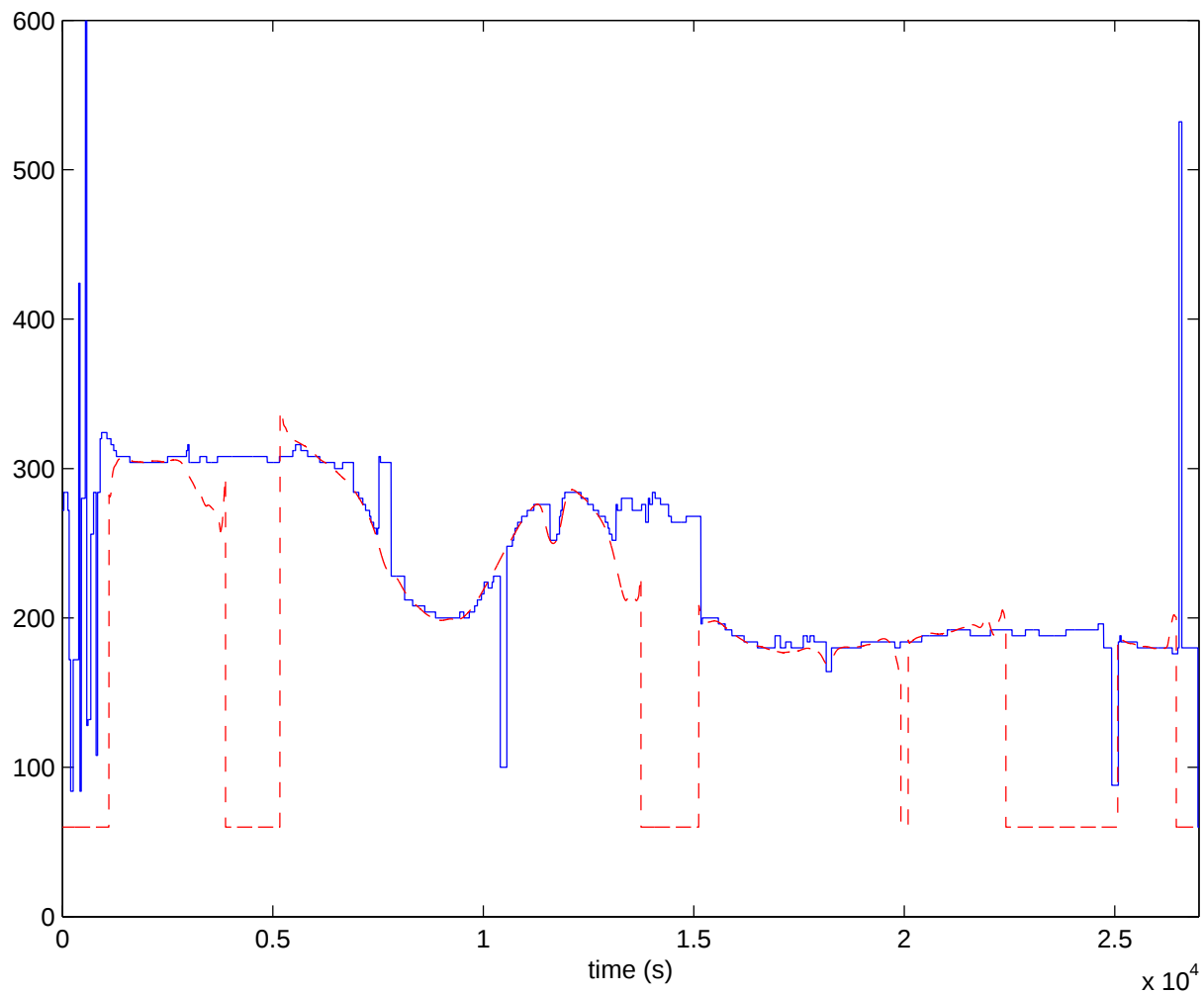


図 5.4: 基本周波数推定結果（非最小位相の位相成分除去）

考察

ケプストラム平均除去法の効果があまりなかった原因として、フレームごとに残響の非最小位相位相ケプストラムが一定ではなかったことが考えられる。実際に、音声と残響インパルス応答が既知である場合に、残響音声からフレームごとの残響の非最小位相ケプストラムを抽出し、フレーム間で比較を行ってみたが一致していなかった。しかし、同様の検証を振幅ケプストラムに対して行ってみたところ一致していた。このため、残響の振幅ケプストラムについてはケプストラム平均により推定できているが、非最小位相ケプストラムについては推定できていないと考えられる。この理由として、位相ケプストラムを求める際、複素対数を一意に定めるため、位相の不連続点をなくす処理が必要になるが、この処理が正しく機能していない可能性がある。また、別の可能性として、音声の振幅ケプストラムの平均は元の振幅ケプストラムに比べ小さい一定値に近づくが、音声の位相ケ

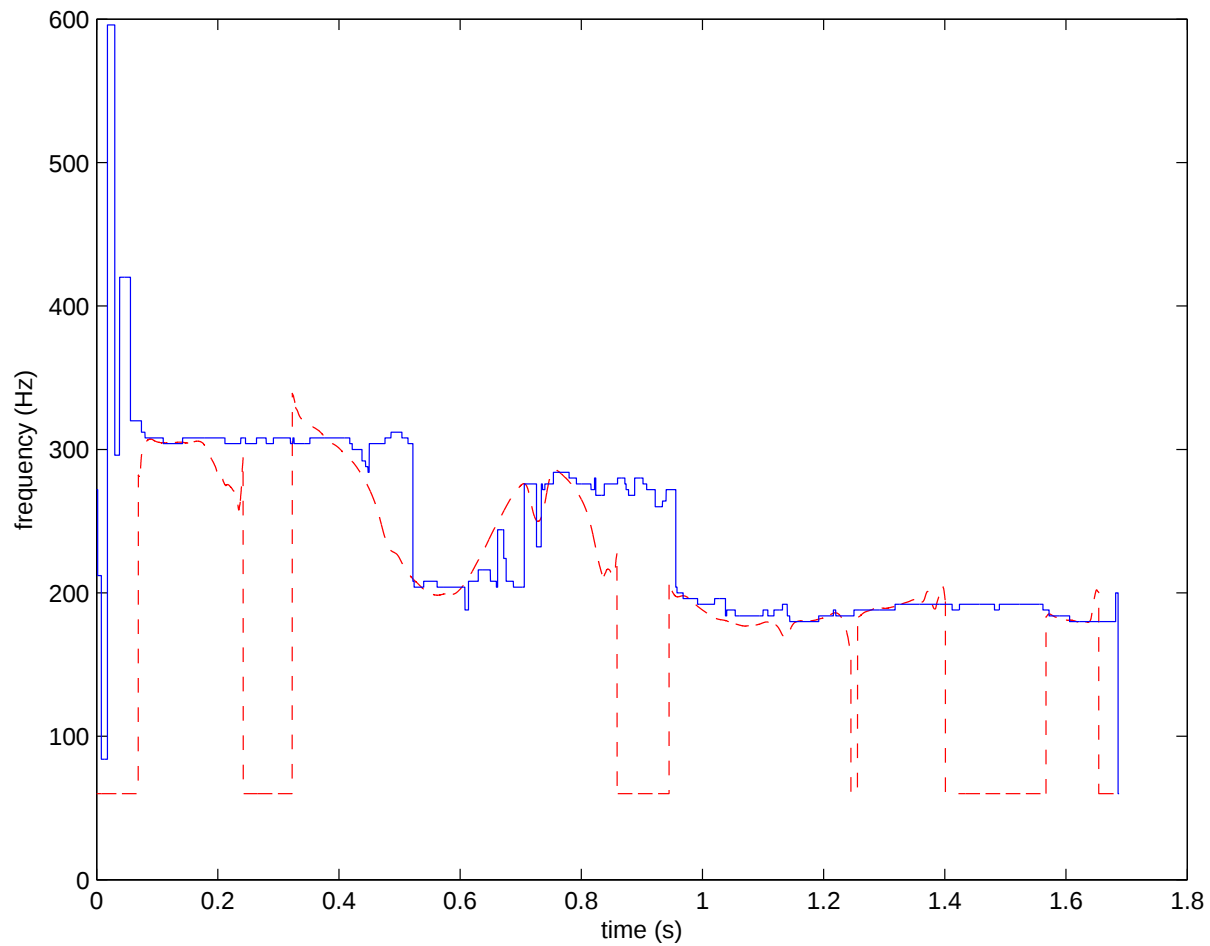


図 5.5: 基本周波数推定結果（ケプストラム平均除去法適用）

プストラムの平均をとり，ある一定値に近づいたとしても，位相の範囲は $-\pi$ から π までなので，平均値の影響は小さくなっていないと考えられる．

この問題を解決し，残響の非最小位相成分を除去できれば，提案法の基本周波数の推定精度を大幅に改善できる可能性がある．今後はこの問題をさらに検証し，解決する必要がある．

第6章 結論

6.1 本稿で明らかにしたこと

音声の基本周波数は様々な音声信号処理において使われており、非常に重要な特徴である。そのため、多くの基本周波数推定法が提案されてきたが、残響に頑健な基本周波数推定法は報告されておらず、また非常に困難であると考えられている。そこで本論文では、残響音声から精度良く基本周波数を推定する手法を提案することを目的とする。残響音声から基本周波数を精度良く推定できれば、実環境での様々な音声信号処理に応用が可能である。

まず、従来の基本周波数推定法が残響環境下で機能するのかを見極めるため、従来の基本周波数推定法を概説し、その中で代表的な10個の手法について耐残響性の評価を行った。評価は、人工的に非最小位相特性を持つ残響インパルス応答を作成し、計算機上で行った。その結果、従来の手法はいずれも残響の影響を大きく受け、ほとんど正しく基本周波数を推定できていないことが分かった。

この評価結果から、ケプストラム分析が残響に対して有効な処理であることが分かった。そこで、複素ケプストラム分析により音源情報と声道情報を分離し、声道情報を取り除くことで、同時に残響の影響を少なくし、残った音源情報の周期性や調波性から基本周波数を推定する手法を提案した。従来法の評価と同様の評価を行い、従来法と比較した結果、提案法は従来法を上回る推定精度であった。また、クリーンな状況での推定精度も高かったことから、提案法は幅広い環境で使える手法であることが分かった。従って、複素ケプストラム分析により抽出した音源情報は残響に対して頑健な特徴であることが分かった。

提案法は従来法に比べ、残響に対して頑健であったが、大幅な改善とは至らなかったの
で、さらに推定精度を上げるため、音声に大きく影響を与える残響の特徴量を調査した。
その結果、残響の非最小位相の位相成分が大きく影響を与えていることが分かった。そ
こで、この残響の非最小位相成分をケプストラム平均除去法を用いて取り除くことを検討
した。しかし、非最小位相ケプストラムに対しケプストラム平均を求めることには問題が
あったため、ケプストラム平均除去法では効果がなかった。この残響の非最小位相成分を
うまく取り除ければ残響に対して非常に頑健な基本周波数推定ができる。

6.2 今後の課題

残響の非最小位相成分が音声から基本周波数を推定する際に大きく影響を与えていることが分かった．そこで，この影響を取り除き，さらに残響に頑健な基本周波数推定法を構築しなければならない．

その上で，人工的な残響インパルス応答だけではなく，実環境で収録した残響インパルス応答を用いての評価を行い，残響環境下で正しく機能することを示す．

また，実環境とは一般的に雑音と残響の両方の影響があるので，雑音環境に対しても適応させていく必要がある．

参考文献

- [1] 古井 貞熙, “デジタル音声処理”, 東海大学出版会, 1985.
- [2] 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をその音声認識への適用”, 電子情報通信学会 技術研究報告, SP2002-13, pp. 37-42, 2002.
- [3] 鈴木 久喜, “ピッチ抽出の今昔”, 日本音響学会誌 56 巻 2 号, pp. 121-128, 2000.
- [4] M. J. Hess, “Pitch and Voicing Determination,” in Advances in speech signal processing, ed. Furui, Sondhi, pp. 3-48, Marcel Dekker. Inc. New York, 1992.
- [5] H.Kawahara, H.Katayose, A.de Cheveigné, R.D.Patterson, “Fixed Point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity”, Proc.Eurospeech99, pp. 2781-2784, 1999.
- [6] A.de Cheveigné, H.Kawahara, “Yin, a fundamental frequency estimator for speech and music”, J. Acoust. Soc. Am., vol. 111, no. 4, pp. 1917-1930, 2002.
- [7] 木下 慶介, 中谷 智広, 三好 正人, “マルチステップ線形予測を用いた 1ch 残響除去法の検討”, 日本音響学会講演論文集, pp. 511-512, 2006.
- [8] 石本 祐一, “時間情報と周波数情報を用いた雑音環境における音声の基本周波数推定に関する研究”, JAIST 博士論文, 2004.
- [9] 関 高浩, 岩野 公司, 古井 貞熙, “ハフ変換を用いた雑音中の音声からの基本周波数推定法”, 情報処理学会 研究報告, 2001-SLP-38-2, vol. 2001, no. 10, pp. 9-14, 2001.
- [10] T. Nakatani, and T. Irino, “Robust fundamental frequency estimation against background noise and spectral distortion,” Proc. ICSLP-2002, vol. 3, pp. 1733-1736, 2002.
- [11] 阿部 敏彦, 小林 隆夫, 今井 聖, “瞬時周波数に基づく雑音環境下でのピッチ推定,” 電子情報通信学会論文誌, vol. J79-D-2, no. 11, pp. 1771-1781, 1996.
- [12] 阿竹 義徳, 入野 俊夫, 河原 英紀, 陸 金林, 中村 哲, 鹿野 清宏, “調波成分の瞬時周波数を用いた基本周波数推定方法”, 電子情報通信学会誌 vol. J83-D-II, No. 11, pp. 2077-2086, 2000.

- [13] B. Gold, L. Rabiner, "Parallel processing techniques for estimating pitch period of speech in the time domain," J. Acoust. Soc. Am., vol. 46, no. 2, pp. 442-448, 1969.
- [14] D. M. Howard, "Peek-picking fundamental period estimation for hearing prostheses," J. Acoust. Soc. Am., vol. 86, no. 3, pp. 902-910, 1989.
- [15] T. Takagi, N. Seiyama, E. Miyasaka, "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.
- [16] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," IEEE Trans. Acoust., Speech, Signal Process. ASSP-22, pp. 353-361, 1974.
- [17] 小林 載, 島村 徹也, "対数スペクトルにクリッピングと帯域制限を用いる基本周波数抽出法," 電子情報通信学会論文誌, vol. J82-A, no.7, pp. 1115-1122, 1999.
- [18] T. Miwa, Y. Tadokoro and T. Saito, "The pitch estimation of different musical instruments sound using comb filter for transcription," IEICE, vol. J81-D-11, no. 9, pp. 1965-1974, Sept. 1998.
- [19] H. Singer and S. Sagayama, "Pitch dependent phone modelling for HMM-based speech recognition," J. Acoust. Soc. Jpn. (E) vol. 15, pp. 77-86, 1994.
- [20] D. J. Hermes, "Measurement of pitch by subharmonic summation," J. Acoust. Soc. Am., vol. 83, no. 1, pp. 257-264, 1988.
- [21] J. Markel, "The SIFT algorithm for fundamental frequency estimation," IEEE Trans, Audio, vol. AU-20, pp. 367-377, 1972.
- [22] A. M. Noll, "Short-time spectrum and cepstrum techniques for vocal-pitch detection," J. Acoust. Soc. Am., vol. 36, no. 2, pp. 226-302, 1964.
- [23] A. M. Noll, "Cepstrum pitch determination," J. Acoust. Soc. Am., vol. 44, no. 6, pp. 1585-1591, 1968.
- [24] A. V. Oppenheim, R. W. Schaffer, Homomorphic analysis of speech," IEEE Trans, Audio, Electroacoust., AU-16, no. 2, pp. 221-226, 1968.
- [25] T. Tanaka, T. Kobayashi, D. Arifianto, and T. Masko, "Fundamental frequency estimation based on instantaneous frequency amplitude spectrum," ICASSP 2002, vol. I, pp. 329-332, 2002.
- [26] H. Kuttruff, "Room Acoustics, Forth edition," Spon Press, 2000.

研究業績

1. 細呂木谷 敏弘, 鵜木 祐史, “残響環境における音声の基本周波数推定法に関する検討,” 日本音響学会 平成 19 年度春季研究発表会 講演論文集, 1-Q-1, March, 2007.
2. T. Hosorogiya, M. Unoki, “A study on the estimation method of the fundamental frequency from a reverberant speech,” NCSP07, March, 2007.