

Title	雑音環境における基本周波数推定とこれを用いた雑音抑圧に関する研究
Author(s)	石本, 祐一
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1362
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修士論文

雑音環境における基本周波数推定と これを用いた雑音抑圧に関する研究

指導教官 赤木 正人 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

石本 祐一

2000年2月15日

目次

1	序論	1
1.1	研究の背景	2
1.1.1	時間情報を用いた雑音抑圧	2
1.1.2	雑音環境における基本周波数推定	3
1.2	本研究の目的・特色	4
2	雑音抑圧アルゴリズムの概要	5
3	帯域幅可変くし形フィルタによる雑音抑圧	7
3.1	基本周期を用いた雑音推定の定式化	7
3.2	雑音スペクトルの推定	8
3.3	基本周期一定のための波形伸縮	8
3.4	雑音抑圧アルゴリズムの評価	11
4	雑音環境における基本周波数推定法	17
4.1	従来の基本周波数推定法の耐雑音性能	17
4.2	帯域幅可変くし形フィルタを用いる基本周波数推定法	22
5	雑音抑圧シミュレーション	27
5.1	シミュレーションの目的	27
5.2	白色雑音を付加された音声の抽出	27
5.2.1	単母音	27
5.2.2	連続母音	30
5.3	ピンク帯域雑音を付加された音声の抽出	33
5.3.1	単母音	33
5.3.2	連続母音	35

5.4	二重母音の音声の抽出	38
5.5	考察	40
6	結論	42
6.1	まとめ	42
6.2	今後の課題	43

目次

1.1	Neural cancellation model	3
2.1	雑音抑圧アルゴリズムの概要	6
3.1	雑音抑圧システムの振幅特性 (基本周期 5ms) : 通常のくし形フィルタ (左上)、 $\alpha=0.2$ (右上)、 $\alpha=0.5$ (左下)、 $\alpha=0.8$ (右下)	9
3.2	基本周期一定のための時間軸での波形伸縮の模式図	10
3.3	波形伸縮前 : 音声波形 (上)、基本周波数 (下)	11
3.4	波形伸縮後 : 音声波形 (上)、基本周波数 (下)	11
3.5	雑音抑圧例 : 目的音 $f_{su} / a/$ (左上)、混合音 SNR=5dB(右上)、目的音の基本周波数 (左下)、雑音抑圧音声 (右下)	12
3.6	雑音抑圧アルゴリズムの SNR による評価 (白色雑音)	13
3.7	雑音抑圧アルゴリズムの SNR による評価 (ピンク帯域雑音)	13
3.8	雑音抑圧アルゴリズムの SD による評価 (白色雑音)	14
3.9	雑音抑圧アルゴリズムの SD による評価 (ピンク帯域雑音)	14
3.10	雑音抑圧アルゴリズムの ASD による評価 (白色雑音)	15
3.11	雑音抑圧アルゴリズムの ASD による評価 (ピンク帯域雑音)	15
4.1	合成音声 $/a/$: 波形 (上) と基本周波数 (下)	18
4.2	雑音付加した合成音声 $/a/$ (SNR=5dB): 白色雑音 (上)、ピンク帯域雑音 (下)	18
4.3	基本周波数推定結果 $/a/$ (白色雑音、SNR=5dB) : 自己相関法 (左上)、残差相関法 (右上)、ケプストラム法 (左中)、平均振幅差関数 (右中)、瞬時振幅の Comb Filtering による方法 (左下)、瞬時周波数を基にした方法 (右下)	19
4.4	基本周波数推定結果 $/a/$ (ピンク帯域雑音、SNR=5dB) : 自己相関法 (左上)、残差相関法 (右上)、ケプストラム法 (左中)、平均振幅差関数 (右中)、瞬時振幅の Comb Filtering による方法 (左下)、瞬時周波数を基にした方法 (すべての時点で推定できていない)(右下)	20

4.5	白色雑音付加音声における基本周波数推定精度	21
4.6	ピンク帯域雑音付加音声の基本周波数推定精度	22
4.7	基本周波数推定アルゴリズム	23
4.8	基本周波数推定結果 /a/ (白色雑音、SNR=5dB) : 瞬時振幅の Comb Filtering による方法 (上)、TEMPO2(中)、本方法 (下)	24
4.9	基本周波数推定結果 /a/ (ピンク帯域雑音、SNR=5dB) : Comb Filter による方法 (上)、TEMPO2(すべての時点において推定できていない)(中)、本方法 (下)	25
4.10	雑音付加音声における提案手法の基本周波数推定精度 : 白色雑音 (上)、ピンク帯域雑音 (下)	26
5.1	雑音抑圧例 : 目的音 fsu /a/(左上)、混合音:白色雑音 SNR=5dB(右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)	28
5.2	白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SNR)	29
5.3	白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SD)	29
5.4	白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (ASD)	30
5.5	雑音抑圧例 : 目的音 fsu /aoi/(左上)、混合音:白色雑音 SNR=5dB(右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)	31
5.6	白色雑音を付加された連続母音/aoi/の雑音抑圧精度 (SNR)	31
5.7	白色雑音を付加された連続母音/aoi/の雑音抑圧精度 (SD)	32
5.8	白色雑音を付加された連続母音/aoi/の雑音抑圧精度 (ASD)	32
5.9	雑音抑圧例 : 目的音 fsu /a/(左上)、混合音:ピンク帯域雑音 SNR=5dB(右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)	33
5.10	ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SNR)	34
5.11	ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SD)	34
5.12	ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (ASD)	35
5.13	雑音抑圧例 : 目的音 fsu /aoi/(左上)、混合音:ピンク帯域雑音 SNR=5dB(右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)	36
5.14	ピンク帯域雑音を付加された連続母音/aoi/の雑音抑圧精度 (SNR)	36
5.15	ピンク帯域雑音を付加された連続母音/aoi/の雑音抑圧精度 (SD)	37
5.16	ピンク帯域雑音を付加された連続母音/aoi/の雑音抑圧精度 (ASD)	37
5.17	雑音抑圧例 : 目的音 fsu /a/(左上)、混合音:fsu /a/ + mht /a/ SNR=5dB(右上)、混合音から推定した fsu /a/の基本周波数 (左下)、雑音抑圧音声 (右下)	38

5.18	二重母音の雑音抑圧精度 (SNR)	39
5.19	二重母音の雑音抑圧精度 (SD)	39
5.20	二重母音の雑音抑圧精度 (ASD)	40

第 1 章

序論

音声は人間が用いるコミュニケーション手段のなかで基本的かつ効果的なものである。そのため、デジタル信号処理の発展に伴い、音声の特徴情報を抽出することによる音声符号化、音声分析合成、音声合成、音声認識、話者認識等の研究が幅広く行われてきている [1]。また、信号処理技術の応用としてデジタル補聴器のような製品も開発されており [2]、人間-機械のコミュニケーションシステムの実現へ向けて期待が寄せられている。

しかし、これらの研究のうち、完全な実用段階に至っているものは少ない。

音声認識は一部実用化されているものの、雑音が存在しない環境において使用することを前提とした認識アルゴリズムによる研究がほとんどである。そのため、雑音のない環境で発声した音声に対する音声認識システムの認識精度が 95% であるのに対し、18dB の信号対雑音比下で発声した音声では誤り率が一桁大きくなると報告されている [3]。また、雑音が定常的であれば雑音が付加された音声から参照パターンを作成することにより、雑音による認識精度の低下は非常に小さくできるが、突発的な雑音には対処できない。さらに、実環境では参照パターンの学習時とテスト時の条件が同一であることは滅多にないため、雑音環境においても頑健な性能の音声認識システムはまだ実用化されていない。

音声信号処理のなかで実用化されているひとつとしてデジタル補聴器が挙げられる。デジタル補聴器では補聴器の調整及び音響処理をディジタル信号処理で行なっている。補聴器は難聴者が聴きたい会話音を難聴者の聴きやすい範囲に増幅するために使用されるが、外界からの音をそのまま増幅したのでは雑音も増幅してしまうことになり、騒音により補聴器使用者にうるさい感覚を引き起こし、また、騒音が語音を遮蔽してしまうため会話理解を妨害する。これは、ノンリニア増幅により周波数帯域ごとに利得を変えることで対処できるが、環境によって利得調整を行なう必要がある [2]。

このように雑音環境における音声信号処理では、雑音を抑圧し音声波形や音声の特徴を

抽出することが重要となる。よって、本研究では、音声認識や補聴器の前処理等に利用できるような音声信号処理として、雑音環境における音声の基本周波数推定法及び雑音抑圧法を提案し、その有用性を検討する。

1.1 研究の背景

1.1.1 時間情報を用いた雑音抑圧

複数の音や雑音がある環境で特定の音源から生成される音を聴き分けることは人間には容易なことである。「カクテルパーティ効果」と呼ばれる聴覚のこの機能を工学的に実現することができれば、音声認識を様々な環境で安定して動作させるために極めて有効であるし、補聴器等の前処理に用いることで雑音による会話理解の妨害を防ぐことが可能となる。

複数の音の中から目的の音を取り出す音源分離(雑音抑圧)の研究は古くから行なわれているが、それらは受音源の数で2種類に分けられる。1つは受音源の数を複数にすることで音源の位置情報を利用して音を取り出すというものである。受音源としてマイクロホンアレーを使い、受音の指向性を鋭くすることで特定の位置にある音源からの音のみを受音する方法[4]や音の到来方向時間差から雑音スペクトルを推定しスペクトルサブトラクションによって目的の音を得る方法[5]等がある。もう1つは受音源の数を1つのみとして受音した音から何らかの制約条件を用い目的音や雑音を推定するというものである。これには、CookeやBrownのガンマトーンフィルタバンクを用いた音響事象に基づいたモデル[6, 7, 8, 9]やEllisの蝸牛殻フィルタバンクを用いた心理音響学的グルーピングの規則に基づいたモデル[10, 11]、中谷らのマルチエージェントシステムによる音響ストリーム分離モデル[12, 13]等がある。

しかし、これらの手法には、それぞれ問題点がある。マイクロホンアレーで指向特性を鋭くするにはある程度の広さのマイクロホン間隔をとらなければならないため[14]、受音源が複数ある方法が補聴器への応用に適しているとは言えない。また、受音源が単一の手法は上述のものを含めてほとんどが音声分析法として振幅(あるいはパワー)スペクトルを用いているために音本来の情報である基本周期や位相等の時間情報の多くが潰されてしまい利用されない。このため、位相情報が失われてしまうので完全な波形の復元を行なうことができない。

時間情報を用いた研究としてはde Cheveignéが調波構造を持つ混合音の基本周期を推定するキャンセレーションモデルを提案している[15, 16]。このモデルは1つの音の基本

周期を推定し、基本周期に合わせた遅延回路によってその音を混合音から除去することで残った信号の基本周期を推定している。この遅延線と抑制性シナプスからなる遅延回路は猫や豚等の多くの動物の神経回路にも存在することが報告されている。しかし、de Cheveigné は図 1.1 に示すような神経回路の処理の定式化を考慮しており、分離を行なっているのは聴神経線維によるスパイク列である。工学的にはくし形フィルタによる基本周期推定の提案に留まっており、音声波形の分離の可能性を示唆しているものの波形レベルでの分離は行っていない。

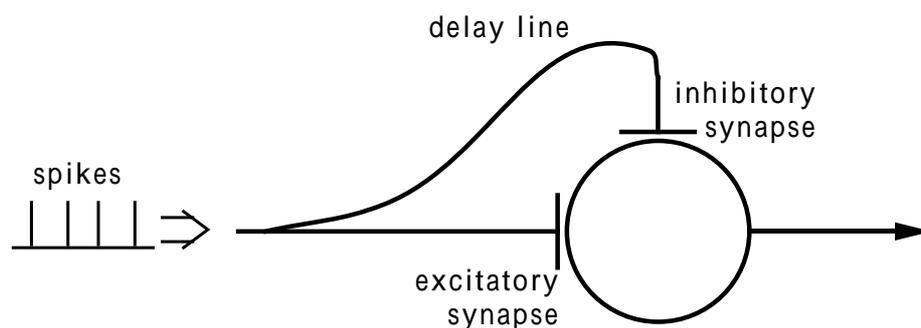


図 1.1: Neural cancellation model

1.1.2 雑音環境における基本周波数推定

有声/無声判定や音量等とともに基本周波数は音源情報のひとつであり、音声の符号化伝送上の役割は大きい。音声の認識や理解、話者識別、音声分析合成などには声道情報と音源情報の処理が必要であるが、基本周波数は韻律性だけでなく実際は音韻情報を担うスペクトルパラメータとも相関があるので、音韻認識にも効果的に利用することができる [1, 17]。また、聴覚の情景解析 [18] に関する研究では、基本周波数が重畳する音を別々の音の流れであると認識するための手掛りのひとつであるとしている。

このように基本周波数は音声の信号処理上重要であるため、基本周波数推定は音声分析の研究が始まって以来の研究課題であり、これまでも様々な方法が提案されている。代表的なものとして、音声波形の自己相関による方法、LPC 分析の残差信号の自己相関による方法、パワースペクトルの対数の逆フーリエ変換によりスペクトル包絡と微細構造を分離して微細構造から基本周波数を求める方法 [19]、平均振幅差関数 (AMDF) によって周期性を検出する方法 [16] 等がある。

しかし、音声完全に周期的な波ではなく準周期的であることや状況に応じて雑音が含ま

まれるために基本周波数の抽出誤差が生まれる等の問題があるため、まだ決定的な方法は確立していない[17]。

短時間フーリエ変換の瞬時周波数には音源情報が含まれていることに着目し、河原らはフィルタの中心周波数からフィルタ出力の瞬時周波数への写像の不動点を用いて基本周波数を抽出する方法を提案している[20]。河原らのこの基本周波数抽出法(TEMPO2)はクリーンな音声から高精度の基本周波数を得ることができる。しかし、雑音のある環境においては推定精度が低下するという問題がある。

雑音に強い基本周波数推定法として、鷓木は瞬時振幅に対して通過量を最大となるような周波数軸上の Comb Filtering による方法を提案している[22]。この鷓木の瞬時振幅の Comb Filtering による基本周波数推定法は信号対雑音比が 0dB 程度の雑音が付加された音声に対しても基本周波数を推定することができるが、クリーンな音声に対しては TEMPO2 よりも推定精度の点で劣る。

また、阿部らは瞬時周波数に基づく雑音環境における基本周波数抽出法を提案している[23, 24]。これは、瞬時周波数に関する振幅スペクトルに対してガウス関数で重み付けすることにより倍音成分を強調し、Comb Filter と動的計画法によって基本周波数を求めるものである。

1.2 本研究の目的・特色

前節までに述べたように、実環境における音声認識システムの推定精度の大幅な低下や補聴器使用時の雑音の増幅による不快感、その他音声分析合成・音源分離等から雑音の存在する環境における音源情報抽出や雑音抑圧が求められている。

本研究は、有声音のように調波構造を持つ対象として、雑音環境における頑健で高精度な基本周波数推定法及び雑音抑圧アルゴリズムを構築することを目的とする。

また、本研究の特色は、

- 帯域幅可変くし形フィルタで雑音を推定・除去することにより、位相情報を保ったままの波形レベルでの雑音抑圧が可能となること
- 帯域幅可変くし形フィルタによる雑音抑圧と雑音には弱いが高精度な基本周波数推定法である TEMPO2、雑音に強い Comb Filter による基本周波数推定法を組み合わせることにより、頑健で高精度な基本周波数推定が可能になること

である。

第 2 章

雑音抑圧アルゴリズムの概要

ここで、本研究で提案する雑音抑圧アルゴリズムについて概説する。

雑音抑圧アルゴリズムの処理の流れを図 2.1 に示す。

調波構造をもつ目的音 $s(t)$ と雑音 $n(t)$ が混ざり合った混合音 $x(t)$ が、ただ一つの受音源で観測される。

基本周波数推定部において、

1. 雑音に強い瞬時振幅の Comb Filtering による基本周波数推定法 [22] を用いて、観測された混合音 $x(t)$ から目的音の基本周波数を大まかに推定 (\bar{F}_0) する
2. 推定された基本周波数に合わせた帯域幅可変くし形フィルタによって推定雑音を抽出し、推定雑音を元の混合音から引き去ることにより雑音の抑圧を行なう
このとき、基本周波数を一定にするような波形伸縮を行なうことにより、雑音抑圧時の誤差を低減させる
3. 雑音を抑圧した信号 $\hat{s}'(t)$ に対して、瞬時周波数を基にした基本周波数推定法 [20] を用いる

という操作によって高精度の基本周波数 F_0 が推定できる。また、さらに次の雑音抑圧部において、

4. この高精度の F_0 を用いた帯域幅可変くし形フィルタで雑音を推定し除去する

ことで、雑音抑圧した目的音 $\hat{s}(t)$ を得ることができる。

帯域幅可変くし形フィルタによる雑音抑圧部については第 3 章、基本周波数推定部について第 4 章で詳しく解説する。

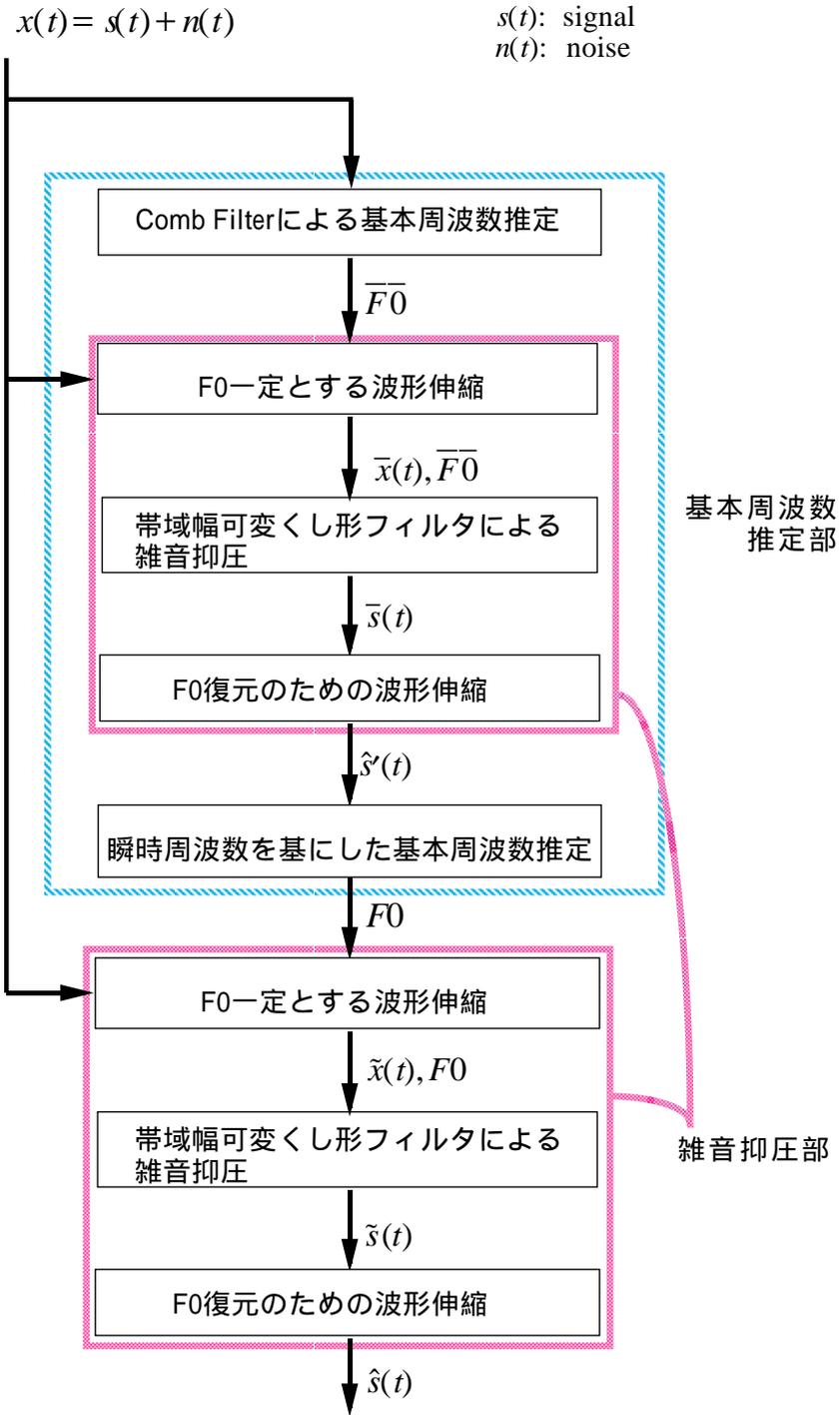


図 2.1: 雑音抑圧アルゴリズムの概要

第 3 章

帯域幅可変くし形フィルタによる雑音抑圧

3.1 基本周期を用いた雑音推定の定式化

目的音 $s(t)$ を基本周期 $T(t)$ の調波複合音とすれば、雑音を $n(t)$ として、混合音 $x(t)$ は、

$$x(t) = s(t) + n(t) \quad (3.1)$$

$$= \sum_m a_m e^{j(m\omega_0(t)t + \theta_m)} + \sum_k b_k e^{j(\omega_k t + \theta_k)} \quad (3.2)$$
$$(\omega_0(t) = 2\pi/T(t))$$

と表わされる。 $T(t)$ を一定値 $T (= 2\pi/\omega_0)$ と仮定し、式 (3.2) を $\pm T$ だけ時間軸でずらし、混合音から引き去る信号 $g(t)$ を計算すれば、

$$g(t) = \frac{2x(t) - x(t - T) - x(t + T)}{4} \quad (3.3)$$

$$= \sum_k b_k e^{j(\omega_k t + \theta_k)} \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (3.4)$$

となる。 $n(t)$ のフーリエ変換を $N(\omega_k)$ とすれば、 $g(t)$ のフーリエ変換 $G(\omega_k)$ は、

$$G(\omega_k) = N(\omega_k) \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (3.5)$$

となる。よって、雑音スペクトル $N(\omega_k)$ は、

$$N(\omega_k) = G(\omega_k) / \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (3.6)$$

となる。 $N(\omega_k)$ を逆フーリエ変換した雑音 $n(t)$ を、元の混合音から引き去ることにより雑音が除去され、目的音 $s(t)$ が推定できる。

しかし、このままでは、

- ω_k/ω_0 が整数のとき、式(3.6)の雑音スペクトル $N(\omega_k)$ が無限大となる
- 実音声では時間的に変動する基本周期 $T(t)$ を式(3.3)において一定と仮定しているため、推定される雑音に誤差が生じる

という問題がある。

そこで、これらの問題点について次のような対策を試みる。

3.2 雑音スペクトルの推定

式(3.6)では、 ω_k/ω_0 が整数のときに雑音スペクトル $N(\omega_k)$ が無限大になってしまう。そこで、実際の使用にはある値 ϵ を設定し、

$$\tilde{N}(\omega_k) = \begin{cases} G(\omega_k)/\sin^2 \frac{\omega_k}{\omega_0} \pi & \left| \sin \frac{\omega_k}{\omega_0} \right| \geq \epsilon \\ G(\omega_k) & \left| \sin \frac{\omega_k}{\omega_0} \right| < \epsilon \end{cases} \quad (3.7)$$

とする。

混合音を入力、目的音を出力とし基本周期 5ms(基本周波数 200Hz) に合わせたシステムの振幅特性を図 3.1に示す。

通常のくし形フィルタの場合は通過帯域が非常に狭くなるが、本研究で用いる方法では ϵ が小さければ通過帯域が狭く、 ϵ が大きければ通過帯域が広がる。すなわち、パラメータ ϵ の値によって通過帯域幅が制御できるくし形フィルタとなっている。

3.3 基本周期一定のための波形伸縮

3.1節では基本周期 T を一定と仮定して計算を行なっているが、実音声では基本周期は時間的に変動する。そのため、このままでは実音声において推定雑音 $\tilde{n}(t)$ に誤差が生じてしまう。

そこで、図 3.2に示すような基本周期を一定とするための音声波形の時間軸での伸縮を行なう。音声波形はあるサンプリング周波数 $1/T_s$ [Hz] でサンプリングされているものとして、波形伸縮処理の流れを以下に示す。

1. 音声の基本周期 $T(t)$ を求める

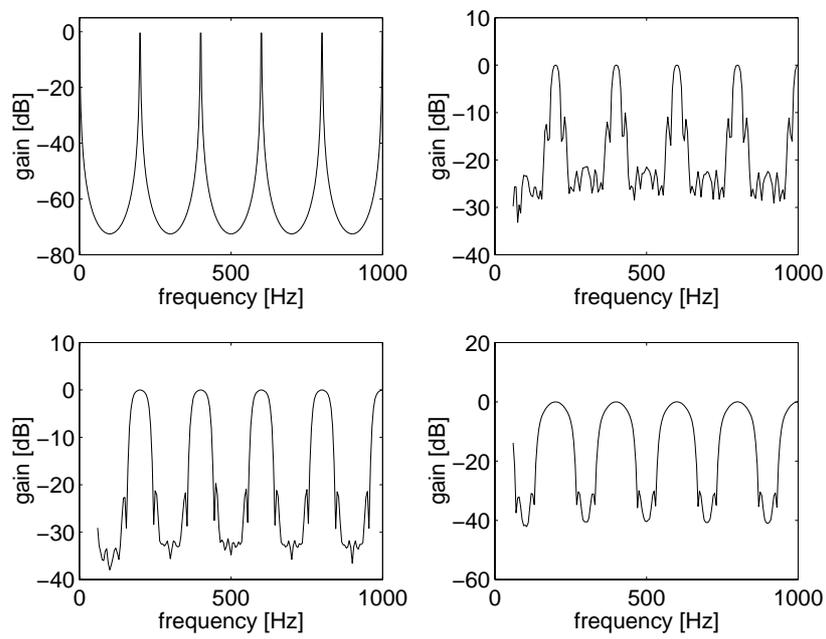


図 3.1: 雑音抑圧システムの振幅特性 (基本周期 5ms) : 通常のくし形フィルタ (左上)、 $\alpha=0.2$ (右上)、 $\alpha=0.5$ (左下)、 $\alpha=0.8$ (右下)

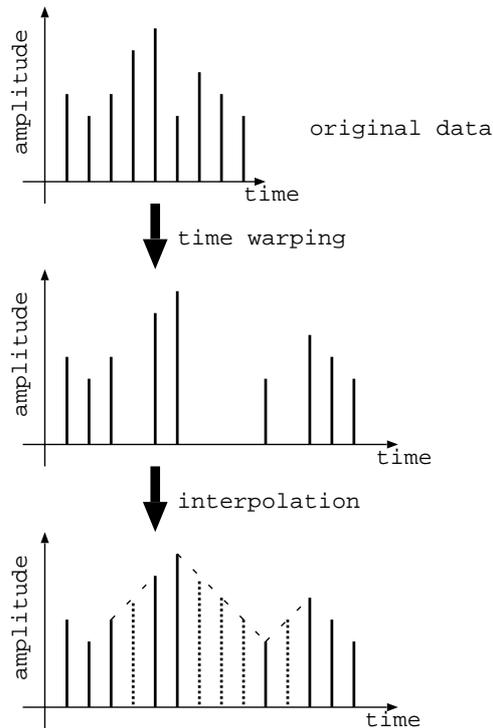


図 3.2: 基本周期一定のための時間軸での波形伸縮の模式図

2. あるサンプリング点 n における基本周期 $T[n]$ と音声全体の平均基本周期 T_{ave} との比 $T_{ave}/T[n]$ を求める
3. $T_{ave}/T[n]$ によってそのサンプリング点とひとつ前のサンプリング点の時間間隔 $T'_s[n]$ を

$$T'_s[n] = T_s \times \frac{T_{ave}}{T[n]} \quad (3.8)$$

とする。これにより時間軸での波形の伸縮が起こる

4. 伸縮した波形に対して、新たにサンプリング間隔 T_s 毎に値を持つよう線形補間を行なう

図 3.3 に表わす音声波形に上記の波形伸縮操作を行なったものを図 3.4 に示す。波形伸縮操作によって音声の基本周波数がほぼ一定となっていることが分かる。

雑音抑圧後の音声に対しては、先に施した波形伸縮操作と逆の操作を行なうことにより、元の基本周波数を持つ音声に戻すことができる。

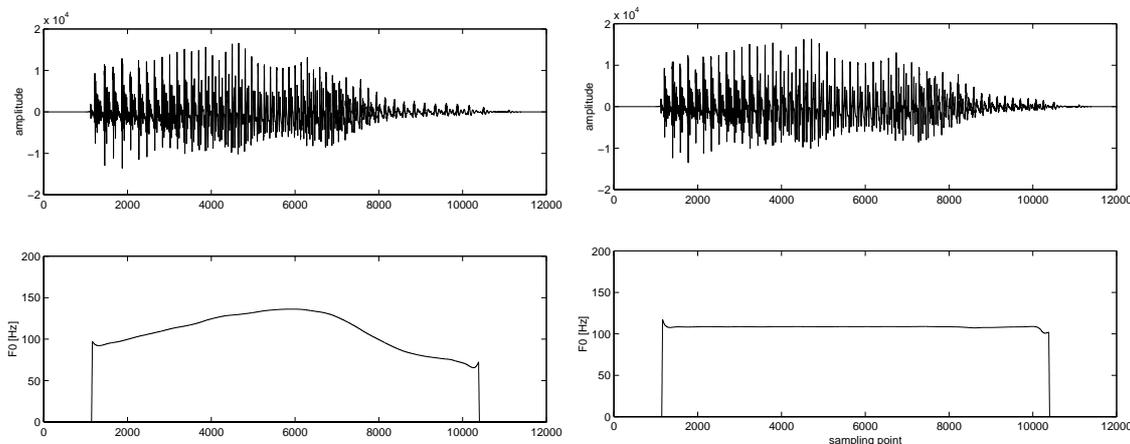


図 3.3: 波形伸縮前：音声波形 (上)、基本周波数 (下) 図 3.4: 波形伸縮後：音声波形 (上)、基本周波数 (下)

3.4 雑音抑圧アルゴリズムの評価

本雑音抑圧アルゴリズムがどの程度雑音を抑圧できるか評価するために、評価実験を行なう。

評価に利用する音声として、ATR 音声データベースの男性話者 mht と女性話者 fsu の単母音 (/a//i//u//e//o/) を用いる。また、雑音として白色雑音と 60 ~ 6000Hz に帯域制限されたピンク雑音を利用する。信号対雑音比 (SNR) は 0dB から 20dB まで 5dB 刻みで変化させる。

基本周波数はクリーンな音声から瞬時周波数を基にした基本周波数抽出法 [20] によってあらかじめ得ている。短時間フーリエ変換の窓として三角窓を使用し、フレーム長 1024point、フレーム周期 256point とした。

帯域幅可変くし形フィルタのパラメータ ϵ は、音声の調波成分の幅が時間や音韻等によって異なるため最適値としてひとつに定めることはできないが、ここでは $\epsilon=0.5$ とした。

評価尺度として、時間領域で歪みを評価するために音声全体の信号対雑音比 (SNR)

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=1}^N x^2[n]}{\sum_{n=1}^N \{x[n] - y[n]\}^2} \quad (3.9)$$

と、周波数領域で歪みを評価するためにスペクトル歪み尺度 (SD)

$$SD = \left[\frac{1}{W} \sum_{\omega=1}^W \{S_x(\omega) - S_y(\omega)\}^2 \right]^{\frac{1}{2}} \quad (3.10)$$

そして、聴覚特性を考慮した歪み評価尺度 (ASD)[25] を用いる。ここで、式 (3.9) の $x[n]$ 、 $y[n]$ は信号のサンプル値、 N は測定区間のサンプル数であり、式 (3.10) の $S_x(\omega)$ 、 $S_y(\omega)$ は信号の対数スペクトルを表わしている。

雑音抑圧の例を図 3.5 に示す。図 3.5(左上) に示すような fsu の単母音 /a/ に SNR が 5dB となるような白色雑音を付加すると、図 3.5(右上) に示す混合音となる。図 3.5(左下) に示す基本周波数を用いて雑音抑圧を行なうと図 3.5(右下) のように混合音から SNR が 11.2dB の音声抽出できる。

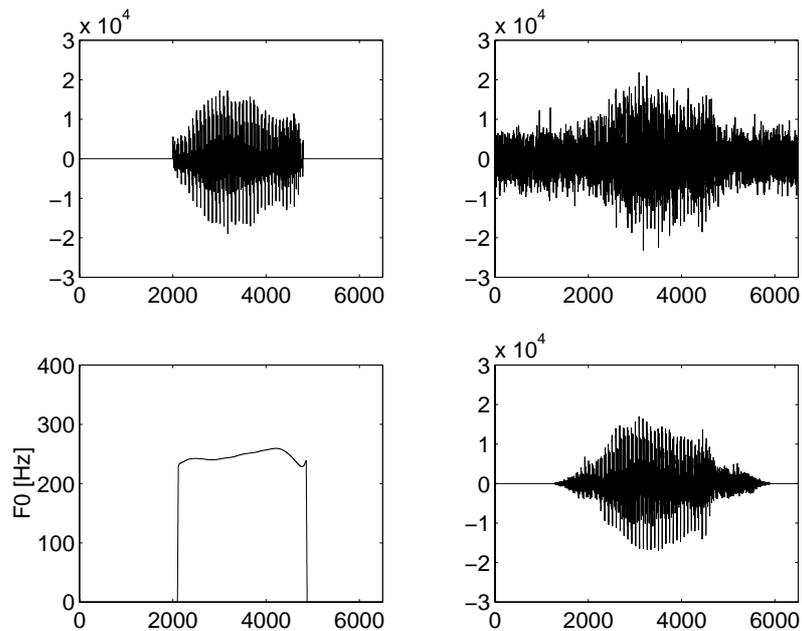


図 3.5: 雑音抑圧例：目的音 fsu /a/(左上)、混合音 SNR=5dB(右上)、目的音の基本周波数 (左下)、雑音抑圧音声 (右下)

雑音抑圧前後の比較を行なった結果、それぞれの値の平均と標準偏差として図 3.6 ~ 3.11 を得た。これらの図の横軸は雑音抑圧前の SNR を表わす。また図 3.6、図 3.7において縦軸は雑音抑圧後の SNR を表わす。

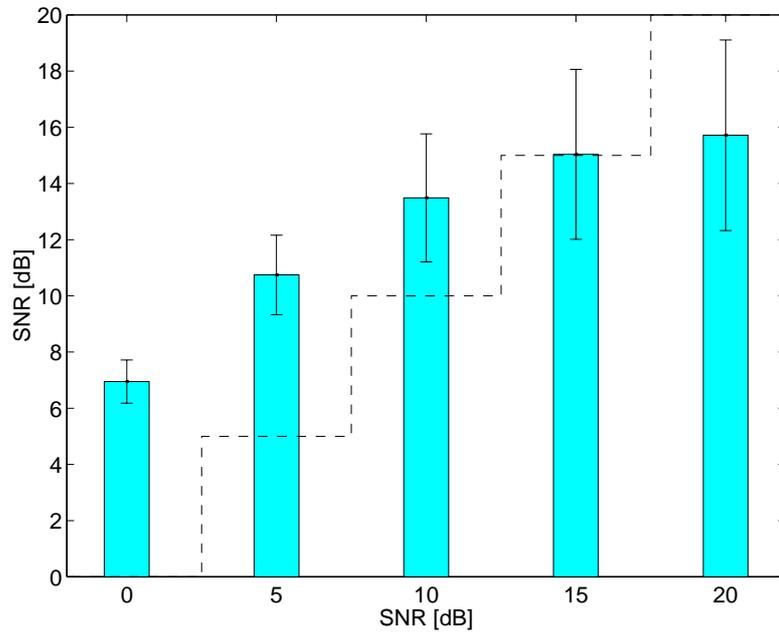


図 3.6: 雑音抑圧アルゴリズムの SNR による評価 (白色雑音)

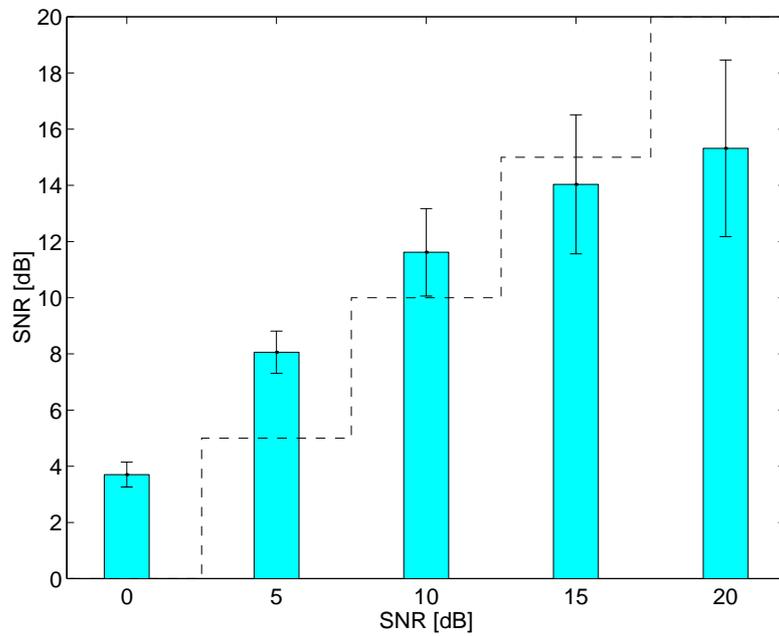


図 3.7: 雑音抑圧アルゴリズムの SNR による評価 (ピンク帯域雑音)

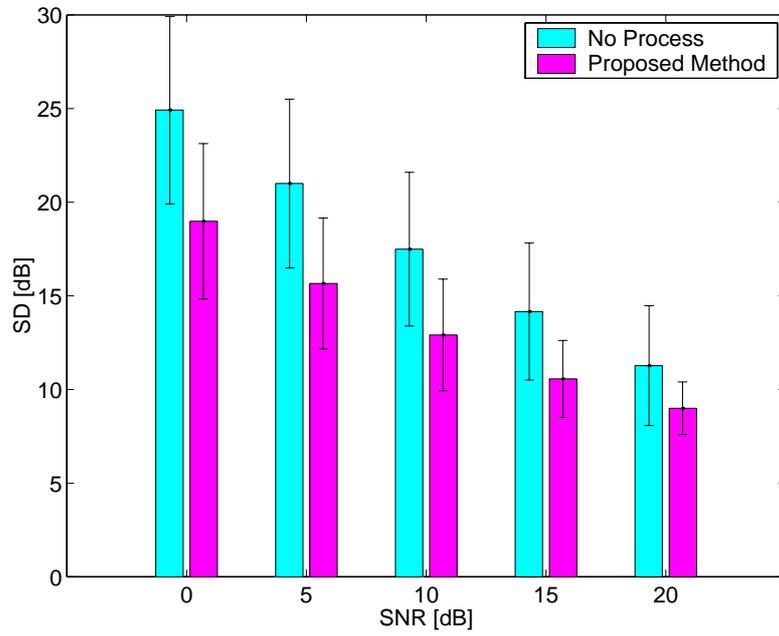


図 3.8: 雑音抑圧アルゴリズムの SD による評価 (白色雑音)

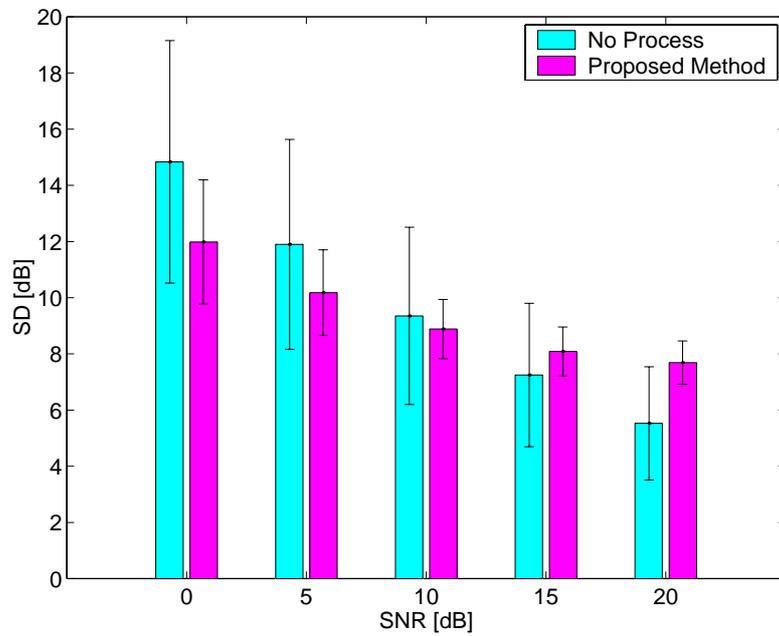


図 3.9: 雑音抑圧アルゴリズムの SD による評価 (ピンク帯域雑音)

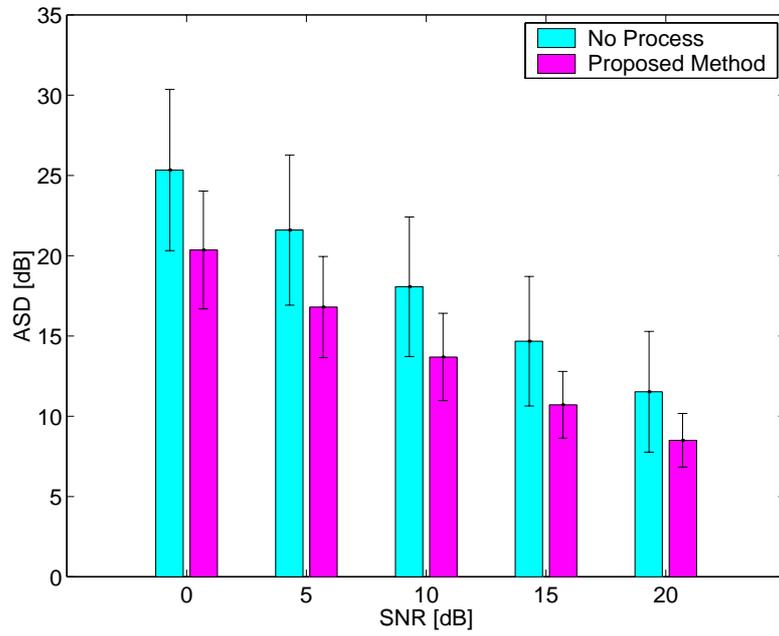


図 3.10: 雑音抑圧アルゴリズムの ASD による評価 (白色雑音)

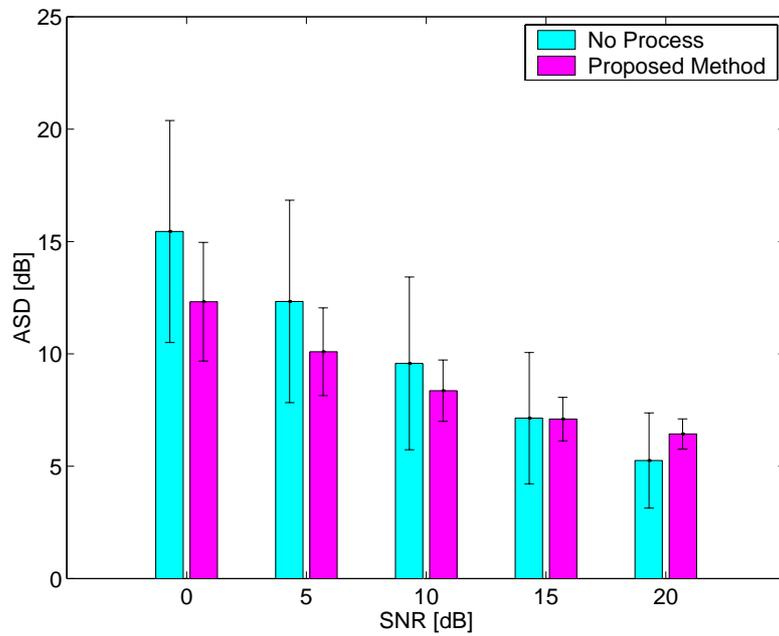


図 3.11: 雑音抑圧アルゴリズムの ASD による評価 (ピンク帯域雑音)

図 3.6,3.7から、雑音が大きい場合、雑音抑圧前に比べて SNR が 5 ~ 7dB 程度向上することが分かる。また、図 3.8 ~ 3.11からスペクトル構造に関しても雑音抑圧による改善が見られることが分かる。

雑音が小さい場合、SNR は雑音抑圧前よりも低下するという結果がみられた。これは、帯域幅可変くし形フィルタによって雑音と共に目的音の成分の一部も除去してしまうため、雑音が小さいときは除去する雑音成分よりも多く目的音の成分を除去してしまうことが原因である。今回の評価実験では ϵ の値を一定としたが、音声によって ϵ の値を変えて最適な通過帯域を選ぶことである程度改善できる。

また、SNR が雑音抑圧前よりも低下していても、音色はやや変わって聴こえることがあるものの、聴感上の雑音感は減少している。これは、聴覚特性を考慮した評価尺度である ASD(図 3.10) では SNR が 20dB でも雑音抑圧後の目的音の精度が向上していることと対応している。聴感上の雑音感の減少は本研究の応用として補聴器の前処理に雑音抑圧アルゴリズムを用いる上で重要である。

図 3.9、図 3.11に示すように、雑音が小さなピンク帯域雑音であるとき、SD、ASD でも抽出精度は低下している。これは、ピンク帯域雑音では周波数の低域のパワーが強いため、高域では雑音除去量よりも目的音の除去量が多くなることに加え、帯域幅可変くし形フィルタでは目的音の基本周波数や高調波と同じ周波数帯域の雑音が除去できずに残ることによると考えられる。

以上の結果から、目的音の基本周波数を正確に推定することができれば本雑音抑圧アルゴリズムによって SNR が 15dB 以下では十分に雑音抑圧効果が得られるといえる。

第 4 章

雑音環境における基本周波数推定法

第 3 章において、目的音の基本周波数はあらかじめわかっているものとしたが、実際の使用においては雑音のある環境から目的音の基本周波数を推定しなければならない。本章では、雑音環境における基本周波数推定法について述べる。

4.1 従来の基本周波数推定法の耐雑音性能

雑音環境で基本周波数を推定するため、まず、従来手法の耐雑音性能を調べる。調査に使用する基本周波数推定法は次の 6 つである。

- 自己相関法 (AC)[19]
- 残差相関法 (RAC)[19]
- ケプストラム法 (NC)[21]
- 平均振幅差関数 (AMDF)[16]
- 瞬時振幅の Comb Filtering による基本周波数推定法 (CF)[22]
- 瞬時周波数を基にした基本周波数抽出法 (TEMPO2)[20]

使用する音声データは、図 4.1 に示すような振幅をほぼ一定とした合成音声 /a//i//u//e//o/ の 5 種類である。音声データの基本周波数は 150 ~ 200Hz 間で 50Hz/200ms で直線状に変動する。サンプリング周波数は 20kHz とした。

雑音は白色雑音及びピンク帯域雑音 (60 ~ 6000Hz) を用いた。音声データに雑音を付加したときの音声波形の例を図 4.2 に示す。

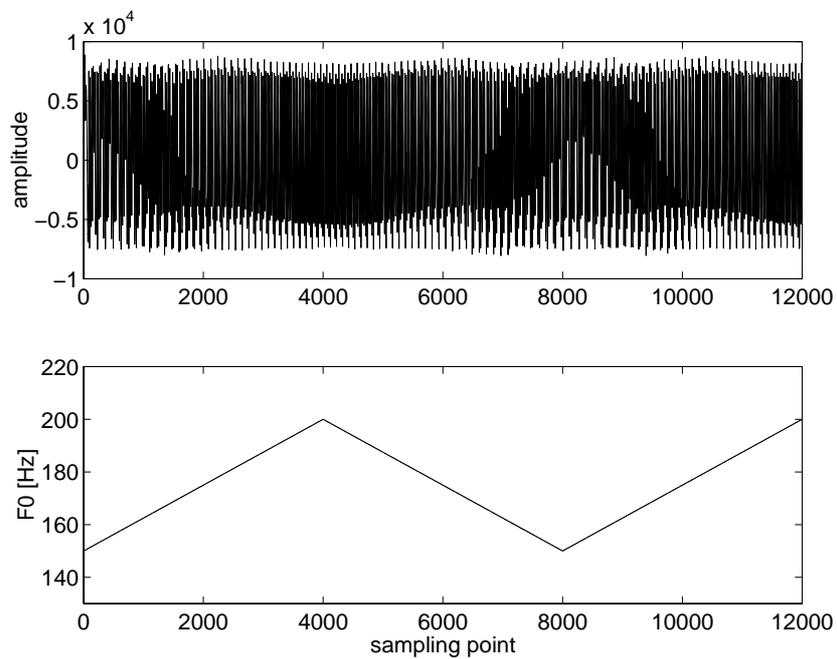


図 4.1: 合成音声 /a/ : 波形 (上) と基本周波数 (下)

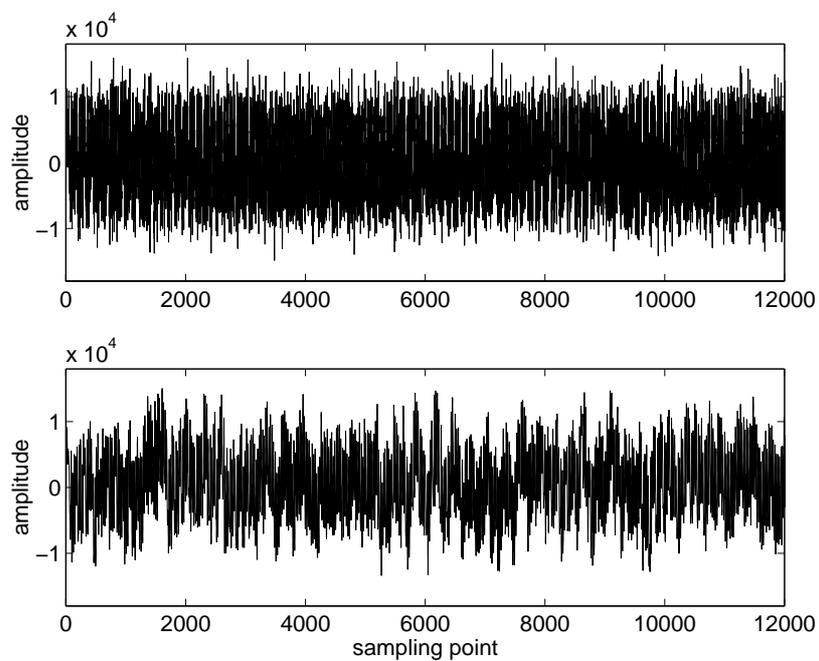


図 4.2: 雑音付加した合成音声 /a/ (SNR=5dB): 白色雑音 (上)、ピンク帯域雑音 (下)

また、白色雑音を付加した音声の基本周波数推定結果の例を図 4.3、ピンク帯域雑音を付加した音声の基本周波数推定結果の例を図 4.4に示す。なお、図 4.4において瞬時周波数を基にした方法ではすべての時点で基本周波数推定ができていない。

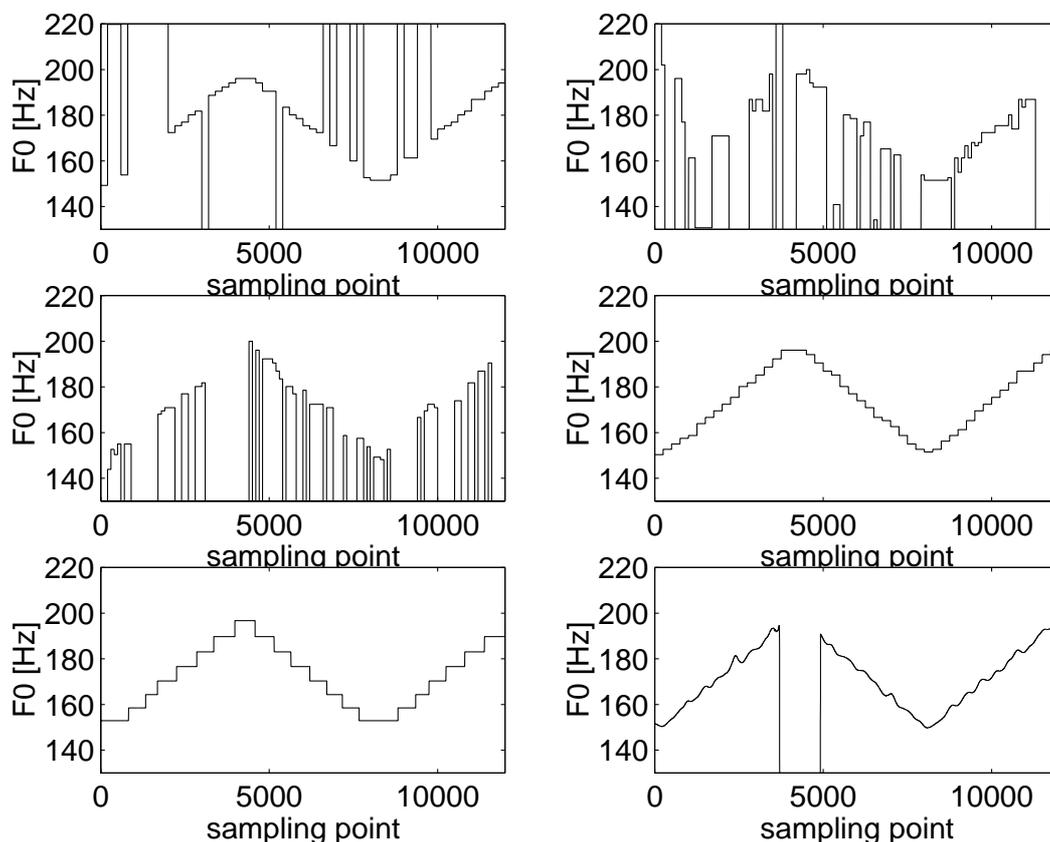


図 4.3: 基本周波数推定結果 /a/ (白色雑音、SNR=5dB) : 自己相関法 (左上)、残差相関法 (右上)、ケプストラム法 (左中)、平均振幅差関数 (右中)、瞬時振幅の Comb Filtering による方法 (左下)、瞬時周波数を基にした方法 (右下)

調査の結果、図 4.5、図 4.6に示すような推定精度 (推定誤差の標準偏差) と雑音強度の関係が得られた。なお、推定精度の計算時において、推定値が 100Hz 以下の値は 100Hz、250Hz 以上の値は 250Hz とし、推定できなかった場合は 100Hz とする異常値処理を行っている。

雑音が白色雑音の場合は、SNR が 10dB 以上のときに瞬時周波数を基にした基本周波数推定法 (TEMPO2) が最も精度が良く、SNR が 0 ~ 5dB のときは TEMPO2 程ではない

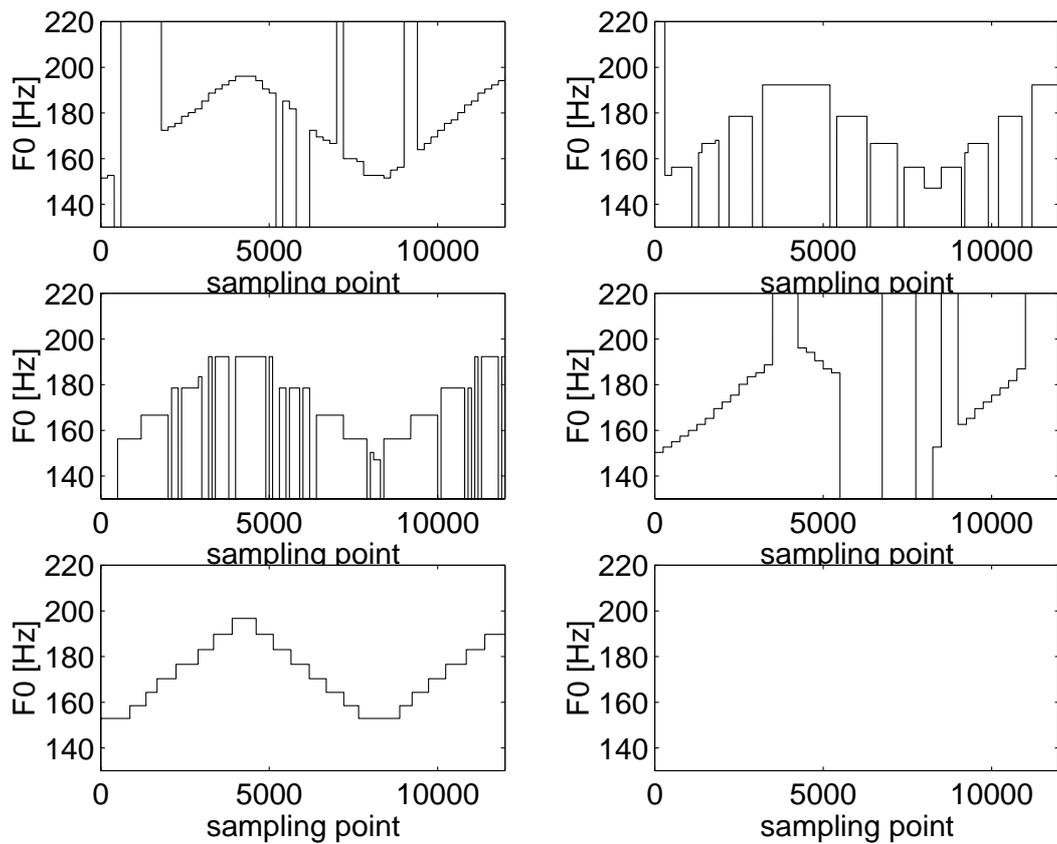


図 4.4: 基本周波数推定結果 /a/ (ピンク帯域雑音、SNR=5dB) : 自己相関法 (左上)、残差相関法 (右上)、ケプストラム法 (左中)、平均振幅差関数 (右中)、瞬時振幅の Comb Filtering による方法 (左下)、瞬間周波数を基にした方法 (すべての時点で推定できていない)(右下)

が瞬時振幅の Comb Filtering による方法の精度が良い。SNR が 0dB よりも小さいような雑音が非常に大きいときはどの基本周波数推定法も十分な精度が得られない。

雑音がピンク帯域雑音の場合は、SNR が 20dB 以上のときに瞬時周波数を基にした基本周波数推定法 (TEMPO2) が最も精度が良く、SNR が 5~15dB のときは Comb Filter による方法の精度が良い。SNR が 0dB のときも、十分な精度が得られているとはいえないが、瞬時振幅の Comb Filtering による方法である程度の基本周波数を推定できる。

以上の結果から、雑音が白色雑音、ピンク帯域雑音のどちらの場合にも共通の傾向が見られる。それは、

- 雑音が小さい場合は TEMPO2 が最も精度が良いが、雑音が大きくなると推定精度が急激に低下する
- 雑音大きい場合は瞬時振幅の Comb Filtering による基本周波数推定法が精度が良いが雑音が小さい場合では TEMPO2 程の推定精度が得られない

というものである。

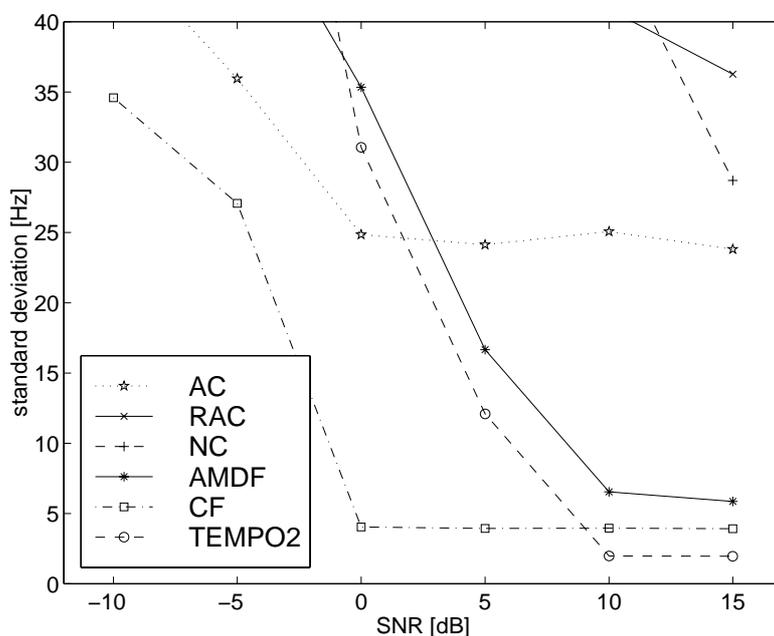


図 4.5: 白色雑音付加音声における基本周波数推定精度

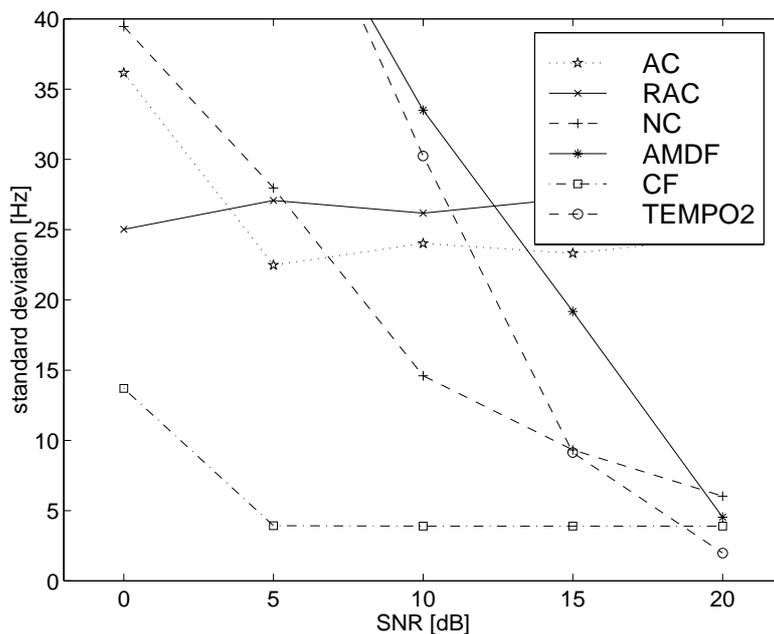


図 4.6: ピンク帯域雑音付加音声の基本周波数推定精度

4.2 帯域幅可変くし形フィルタを用いる基本周波数推定法

雑音小さい環境では TEMPO2 によって高精度の基本周波数を推定できる。雑音大きい環境では瞬時振幅の Comb Filtering による方法が頑健性を示した。そこで、本研究では第 3 章で述べた雑音抑圧アルゴリズムを用いて、雑音環境において頑健で高精度な基本周波数推定法を作成する。本方法のアルゴリズムは、図 4.7 のように TEMPO2 と瞬時振幅の Comb Filtering による方法の 2 種類の基本周波数推定法と帯域幅可変くし形フィルタから構成される。

本方法の基本周波数推定の手順を以下に述べる。

1. まず、雑音に頑健な基本周波数推定法である瞬時振幅の Comb Filtering による方法で混合音 $x(t)$ からある程度の精度をもつ基本周波数 $\bar{F}0$ を得る
2. その基本周波数 ($\bar{F}0$) を基にした帯域幅可変くし形フィルタを用い第 3 章で述べた雑音抑圧アルゴリズムによって雑音を抑圧する

ここで、帯域幅可変くし形フィルタのパラメータ ϵ を調節して通過帯域幅を制御することにより音声の基本周波数成分を抑圧しないようにする

3. そして、雑音が抑圧された音声に対して、雑音に弱い高精度の基本周波数抽出法である TEMPO2 を用いることにより、雑音中の音声から高精度の基本周波数を推定する

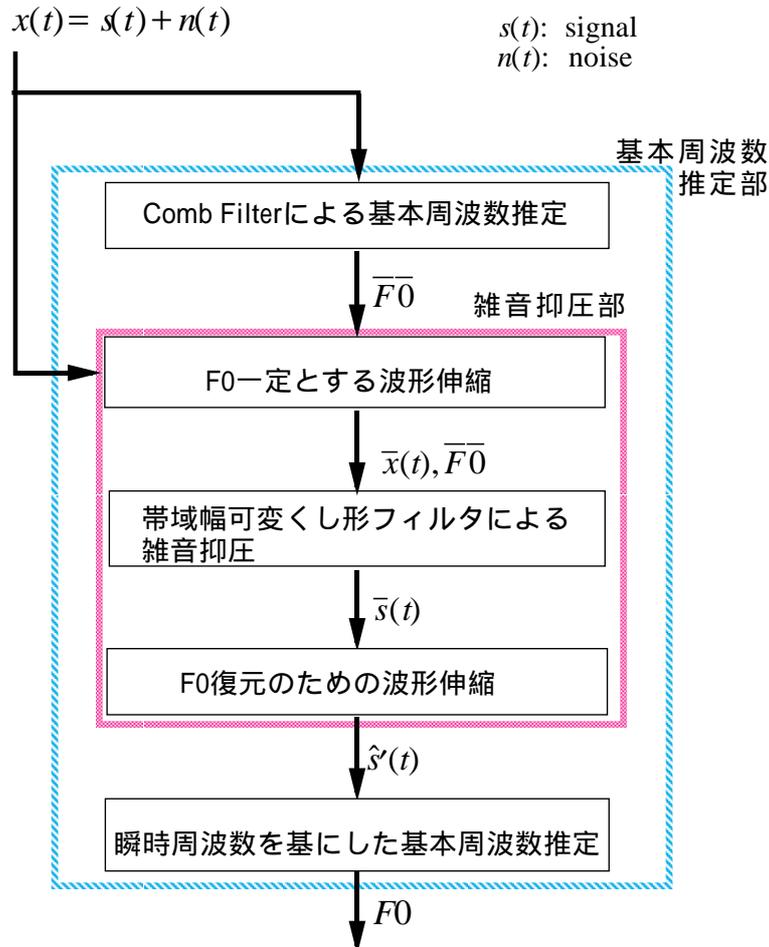


図 4.7: 基本周波数推定アルゴリズム

雑音の付加された音声を用いて本方法の耐雑音性能を調べる。帯域幅可変くし形フィルタのパラメータ $\epsilon = 0.5$ とし、音声、雑音共に 4.1節と同じものを用いた。

白色雑音を付加した音声の基本周波数推定結果の例を図 4.8、ピンク帯域雑音を付加した音声の例を図 4.9に示す。図 4.9(中) の TEMPO2 のみによる推定ではすべての時点で基本周波数が推定できていないが、図 4.9(下) の本方法では基本周波数が得られている。

図 4.10に推定精度 (推定誤差の標準偏差) と雑音強度の関係を示す。比較のため、瞬時

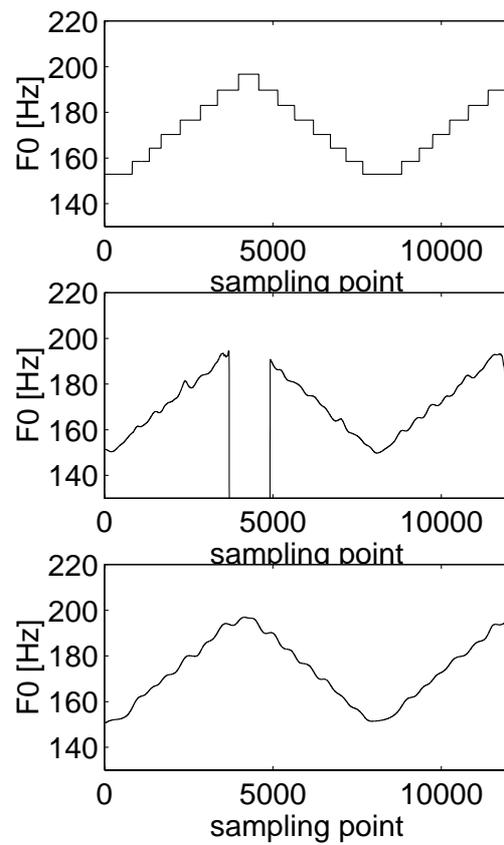


図 4.8: 基本周波数推定結果 /a/ (白色雑音、SNR=5dB) : 瞬時振幅の Comb Filtering による方法 (上)、TEMPO2(中)、本方法 (下)

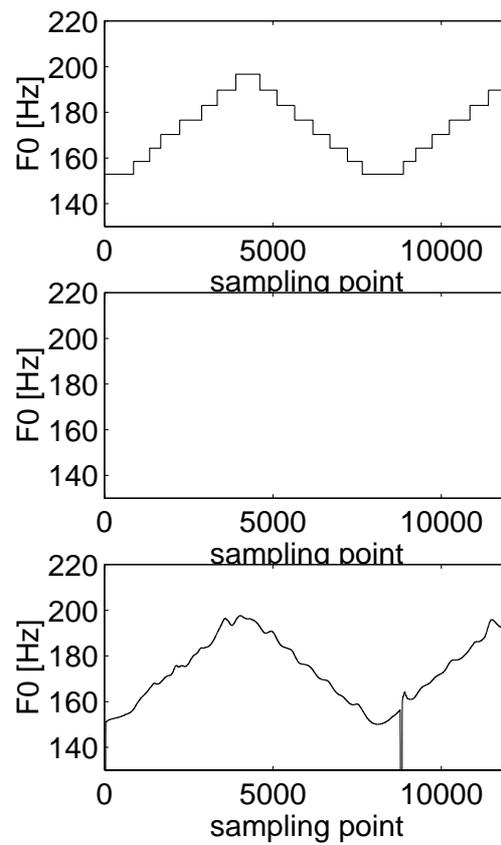


図 4.9: 基本周波数推定結果 /a/ (ピンク帯域雑音、SNR=5dB) : Comb Filter による方法 (上)、TEMPO2(すべての時点において推定できていない)(中)、本方法 (下)

振幅の Comb Filtering による方法のみの場合の推定精度と、TEMPO2 のみの場合の推定精度も示す。

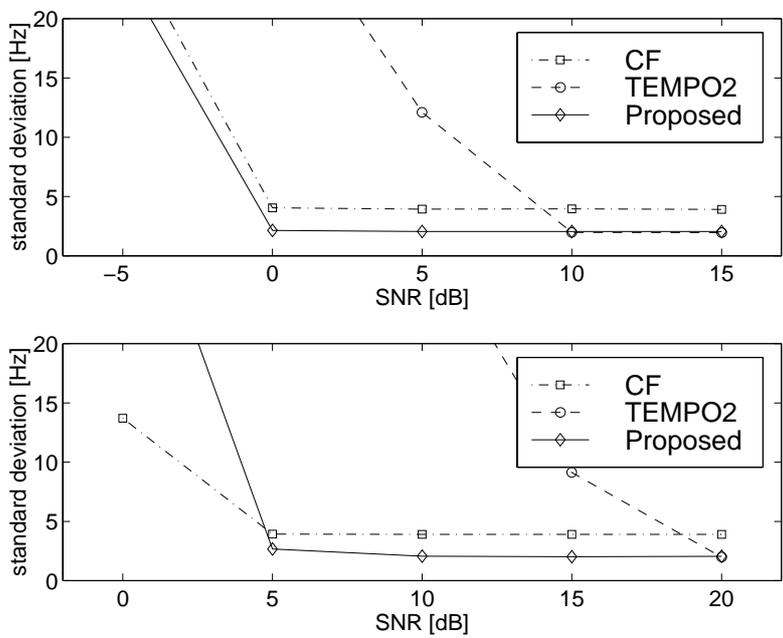


図 4.10: 雑音付加音声における提案手法の基本周波数推定精度：白色雑音 (上)、ピンク帯域雑音 (下)

図 4.10(上) から雑音が白色雑音の場合は、SNR が 0dB 以上である音声に対しても、本方法によって TEMPO2 と同じ精度で基本周波数を推定できることが分かる。

そして、図 4.10(下) から雑音がピンク帯域雑音の場合は、SNR が 5dB 以上である音声に対しても、本方法によって TEMPO2 とほぼ同程度の精度で基本周波数を推定できることが分かる。

これらの結果から、本方法が、SNR が 5dB と大きな雑音が存在する環境においても雑音のない環境と同様の基本周波数を推定することができる、頑健で高精度な基本周波数推定法であるといえる。

また、本方法の耐雑音性能は、図 4.10 で表わされているように瞬時振幅の Comb Filtering による方法の耐雑音性能に大きく依存しているため、雑音抑圧の前処理の基本周波数推定法の改良によりさらなる性能向上が可能であると考えられる。

第 5 章

雑音抑圧シミュレーション

5.1 シミュレーションの目的

雑音抑圧シミュレーションの目的は、本研究で提案した雑音環境における基本周波数推定法及び雑音抑圧法の有効性を調べることである。

ここでは典型的な雑音として白色雑音、ピンク帯域雑音、音声の 3 種類を用いる。

また、評価尺度として、時間領域での歪みを評価するために音声全体の信号対雑音比 (SNR)、周波数領域での歪みを評価するためにスペクトル歪み尺度 (SD) と聴覚特性を考慮した歪み評価尺度 (ASD)[25] を用いる。

5.2 白色雑音を付加された音声の抽出

音声データとして ATR 音声データベースの男性話者 2 名 mht, mau と女性話者 2 名 fsu, fkn の単母音 (/a//i//u//e//o/) と連続母音 (/aoi/) を用いる。また、雑音として白色雑音を利用する。SNR は 0dB から 20dB まで 5dB 刻みで変化させる。

サンプリング周波数 20kHz、フレーム長 1024points、フレーム周期 256points、帯域幅可変くし形フィルタのパラメータは常に $\epsilon = 0.5$ とした。

5.2.1 単母音

白色雑音を付加された単母音に本雑音抑圧アルゴリズムを用いた場合の雑音抑圧例を図 5.1 に示す。

雑音抑圧後の音声全体の SNR の平均と標準偏差を図 5.2 に、雑音抑圧前後の SD、ASD の平均と標準偏差を図 5.3、図 5.4 に示す。

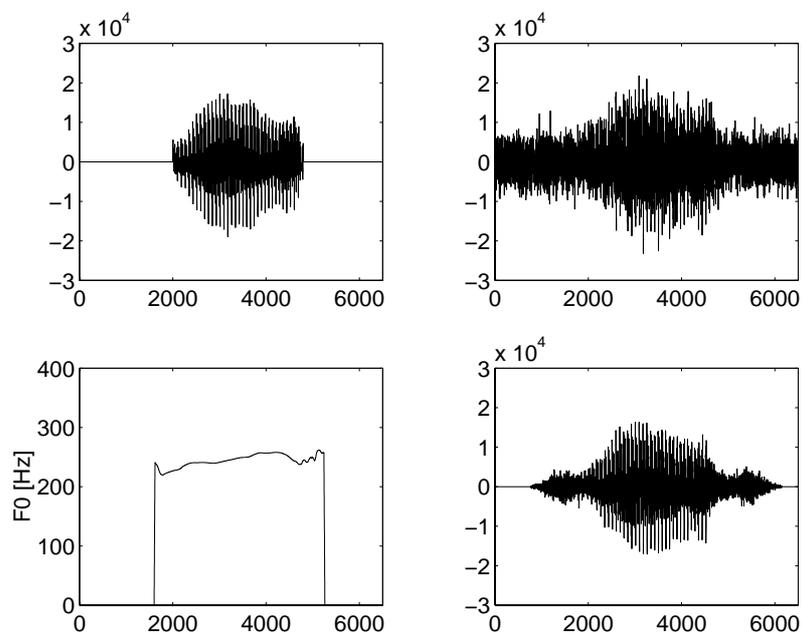


図 5.1: 雑音抑圧例 : 目的音 fsu /a/ (左上)、混合音: 白色雑音 SNR=5dB (右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)

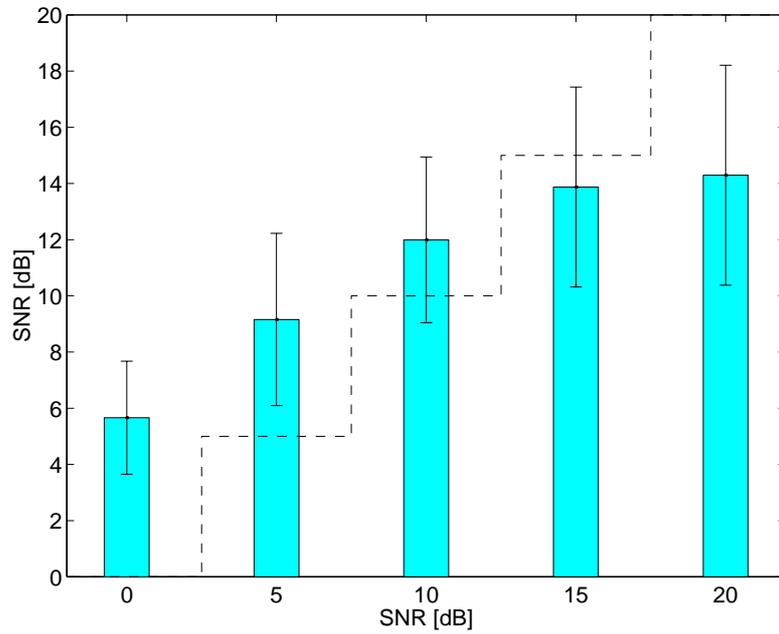


図 5.2: 白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SNR)

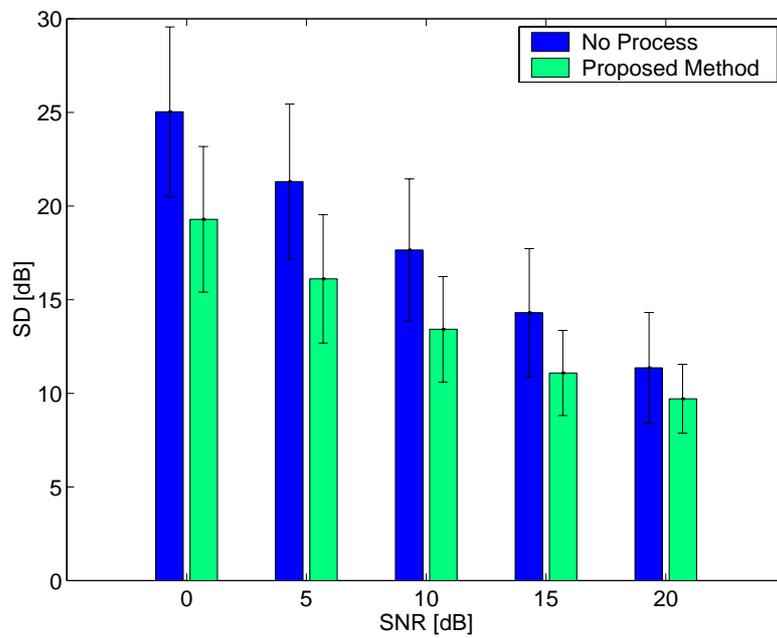


図 5.3: 白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SD)

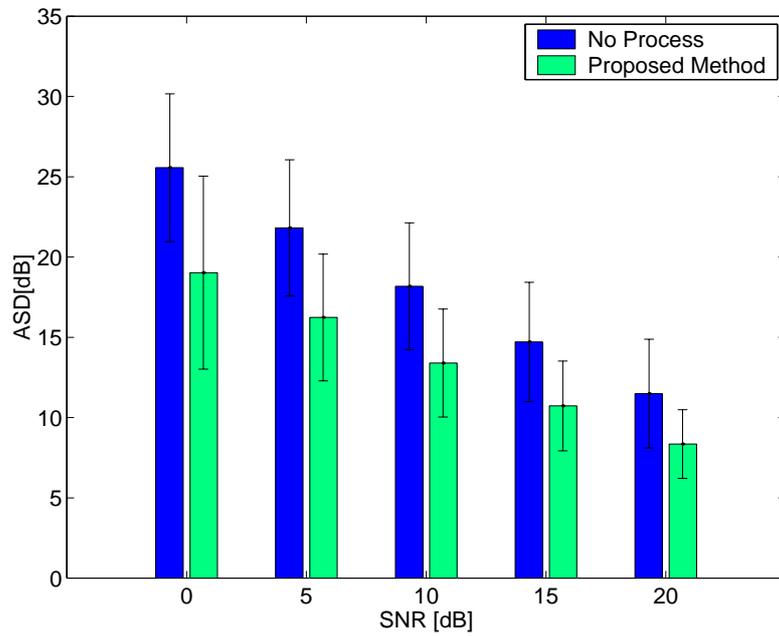


図 5.4: 白色雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (ASD)

5.2.2 連続母音

白色雑音を付加された連続母音に本雑音抑圧アルゴリズムを用いた場合の雑音抑圧例を図 5.5 に示す。

雑音抑圧後の音声全体の SNR の平均と標準偏差を図 5.6 に、雑音抑圧前後の SD、ASD の平均と標準偏差を図 5.7、図 5.8 に示す。

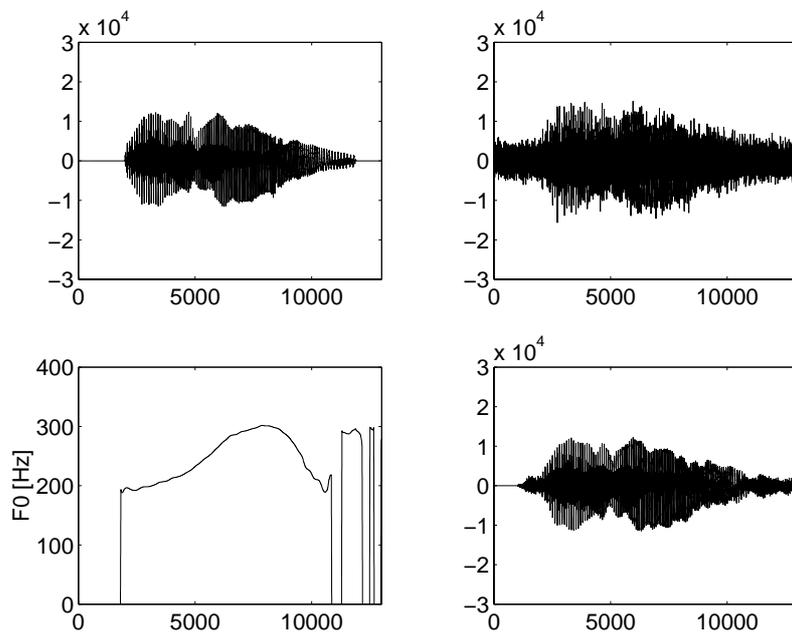


図 5.5: 雑音抑圧例: 目的音 fsu /a oi/ (左上)、混合音: 白色雑音 SNR=5dB (右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)

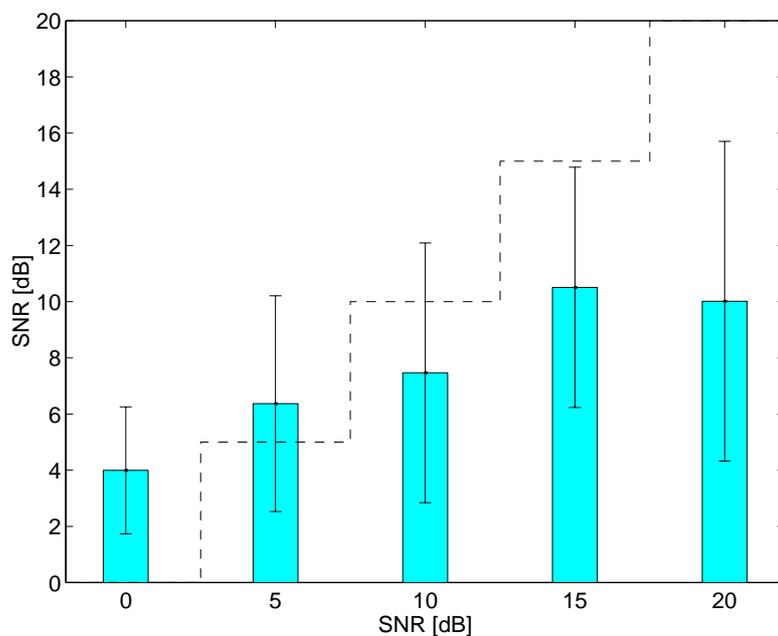


図 5.6: 白色雑音を付加された連続母音/a oi/の雑音抑圧精度 (SNR)

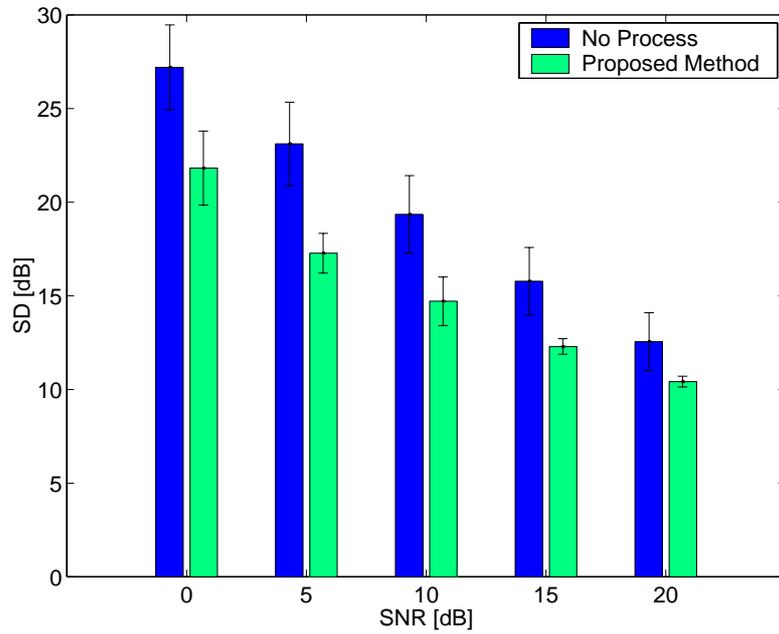


図 5.7: 白色雑音を付加された連続母音/a*oi*/の雑音抑圧精度 (SD)

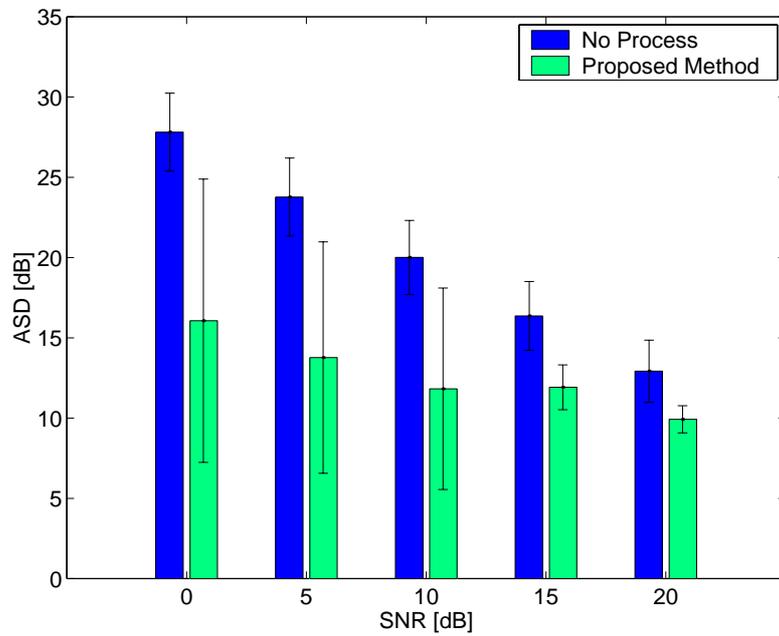


図 5.8: 白色雑音を付加された連続母音/a*oi*/の雑音抑圧精度 (ASD)

5.3 ピンク帯域雑音を付加された音声の抽出

音声データとして ATR 音声データベースの男性話者 2 名 mht,mau と女性話者 2 名 fsu,fkn の単母音 (/a//i//u//e//o/) と連続母音 (/aoi/) を用いる。また、雑音として 60 ~ 6000Hz に帯域制限されたピンク雑音を利用する。SNR は 0dB から 20dB まで 5dB 刻みで変化させる。

サンプリング周波数 20kHz、フレーム長 1024points、フレーム周期 256points、帯域幅可変くし形フィルタのパラメータは常に $\epsilon = 0.5$ とした。

5.3.1 単母音

ピンク帯域雑音を付加された単母音に本雑音抑圧アルゴリズムを用いた場合の雑音抑圧例を図 5.9 に示す。

雑音抑圧後の音声全体の SNR の平均と標準偏差を図 5.10 に、雑音抑圧前後の SD、ASD の平均と標準偏差を図 5.11、図 5.12 に示す。

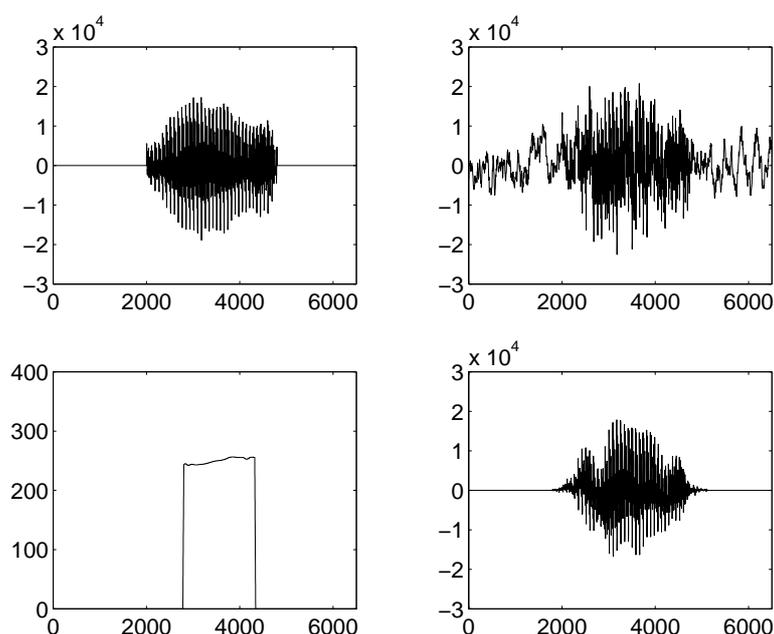


図 5.9: 雑音抑圧例 : 目的音 fsu /a/(左上)、混合音:ピンク帯域雑音 SNR=5dB(右上)、混合音から推定した基本周波数(左下)、雑音抑圧音声(右下)

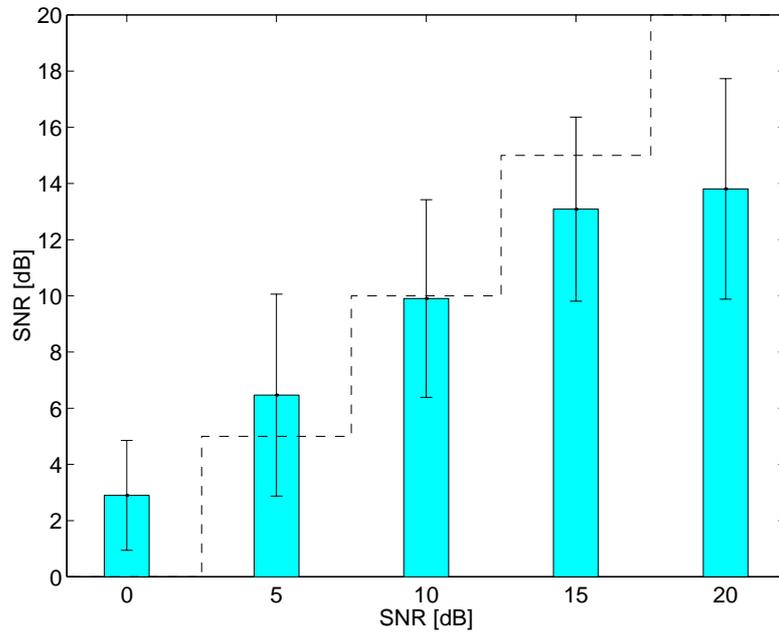


図 5.10: ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SNR)

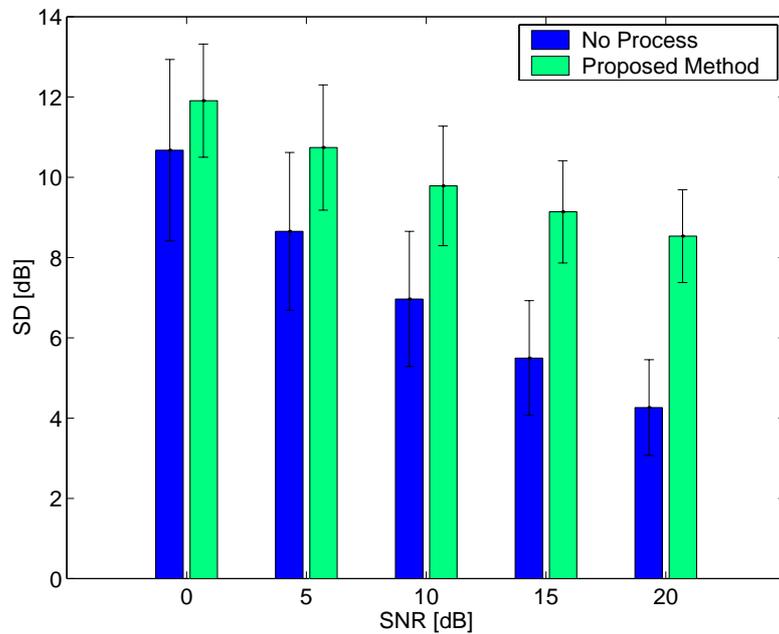


図 5.11: ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (SD)

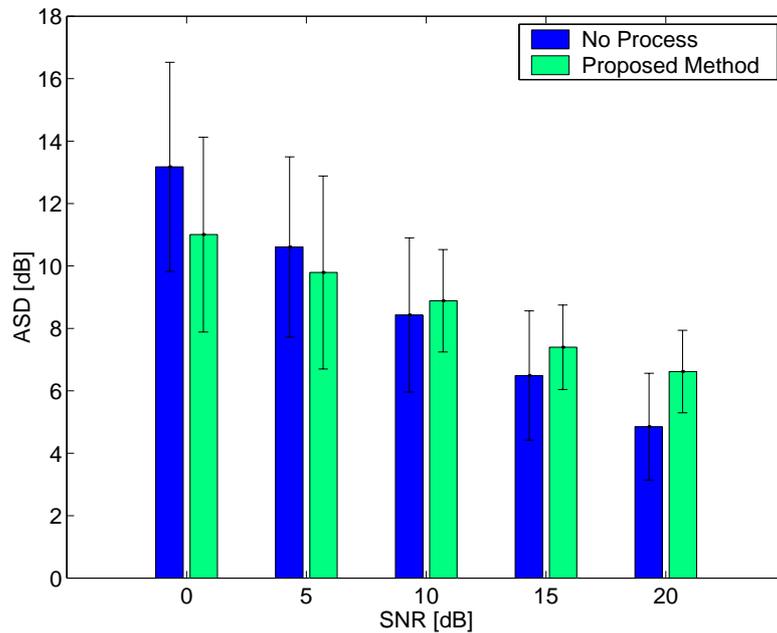


図 5.12: ピンク帯域雑音を付加された単母音/a//i//u//e//o/の雑音抑圧精度 (ASD)

5.3.2 連続母音

ピンク帯域雑音を付加された連続母音に本雑音抑圧アルゴリズムを用いた場合の雑音抑圧例を図 5.13 に示す。

雑音抑圧後の音声全体の SNR の平均と標準偏差を図 5.14 に、雑音抑圧前後の SD、ASD の平均と標準偏差を図 5.15、図 5.16 に示す。

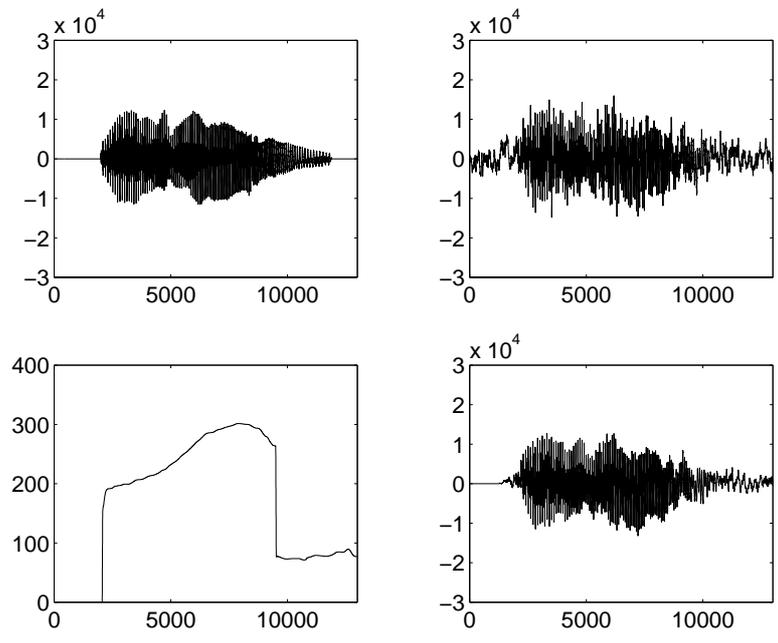


図 5.13: 雑音抑圧例 : 目的音 fsu /aoi/ (左上)、混合音: ピンク帯域雑音 SNR=5dB (右上)、混合音から推定した基本周波数 (左下)、雑音抑圧音声 (右下)

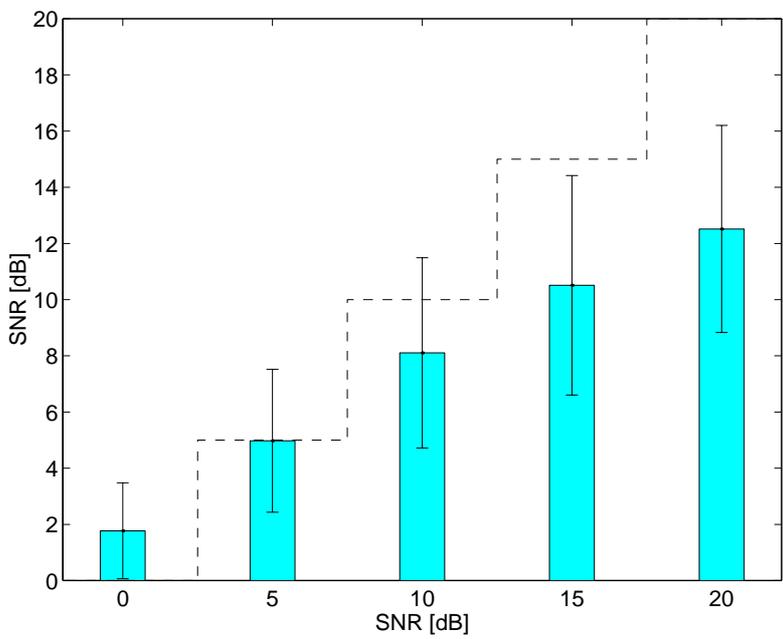


図 5.14: ピンク帯域雑音を付加された連続母音 /aoi/ の雑音抑圧精度 (SNR)

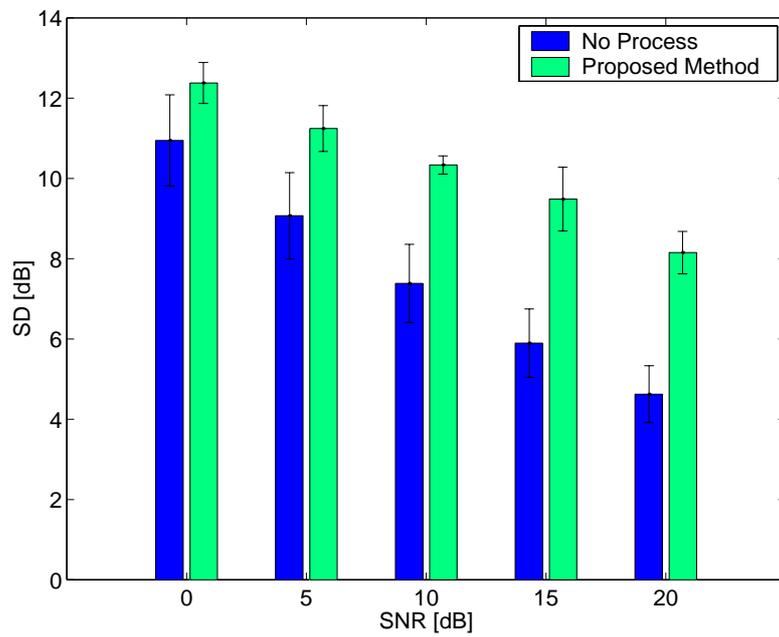


図 5.15: ピンク帯域雑音を付加された連続母音/aoi/の雑音抑圧精度 (SD)

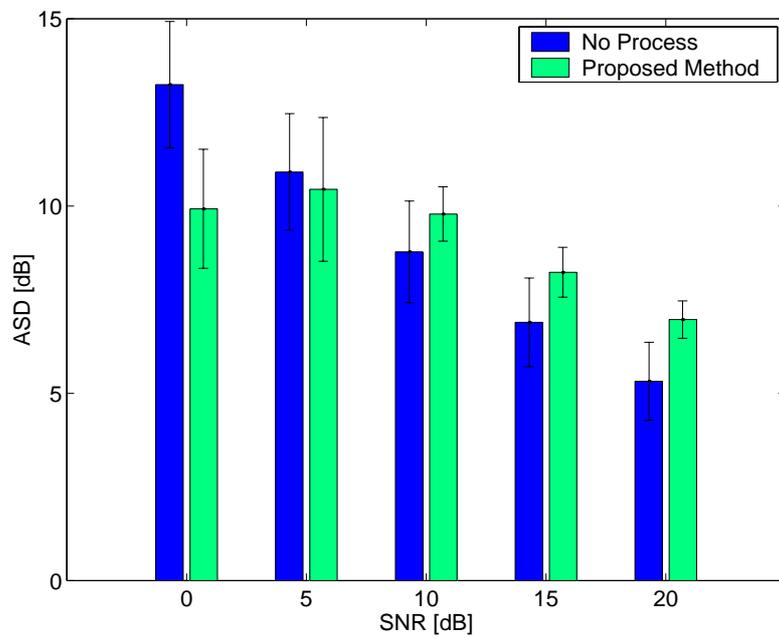


図 5.16: ピンク帯域雑音を付加された連続母音/aoi/の雑音抑圧精度 (ASD)

5.4 二重母音の音声の抽出

目的音として ATR 音声データベースの女性話者 fsu の単母音 (/a/) を用いる。また、雑音として ATR 音声データベースの男性話者 mht の単母音 (/a/) を用いる。目的音と雑音を SNR が 0dB から 20dB まで 5dB 刻みで変化するように足し合わせた音声 (二重母音) から目的音の抽出を試みた。

二重母音に本雑音抑圧アルゴリズムを用いた場合の雑音抑圧例を図 5.17 に示す。

雑音抑圧後の目的音と妨害音が重なっていた区間の SNR の平均を図 5.18 に、雑音抑圧前後の SD、ASD の平均を図 5.19、図 5.20 に示す。

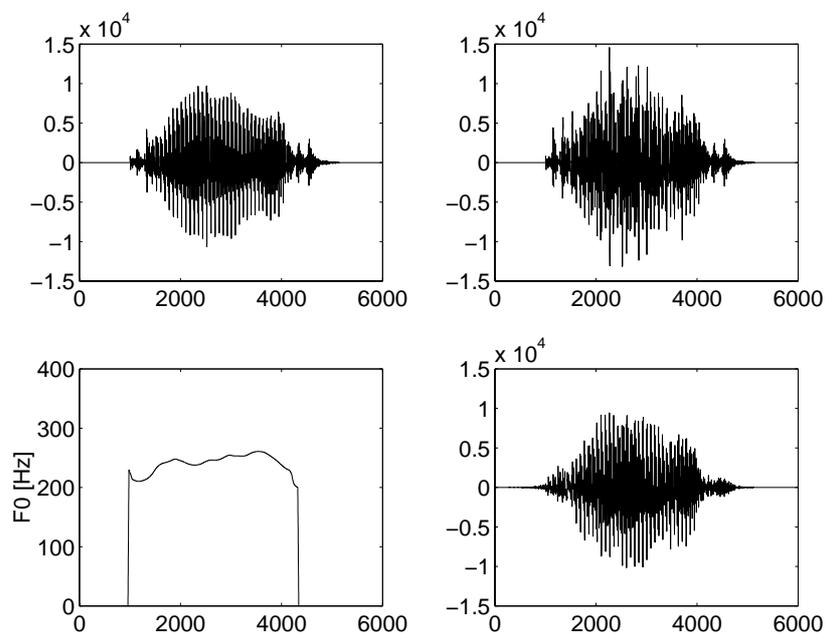


図 5.17: 雑音抑圧例: 目的音 fsu /a/ (左上)、混合音: fsu /a/ + mht /a/ SNR=5dB (右上)、混合音から推定した fsu /a/ の基本周波数 (左下)、雑音抑圧音声 (右下)

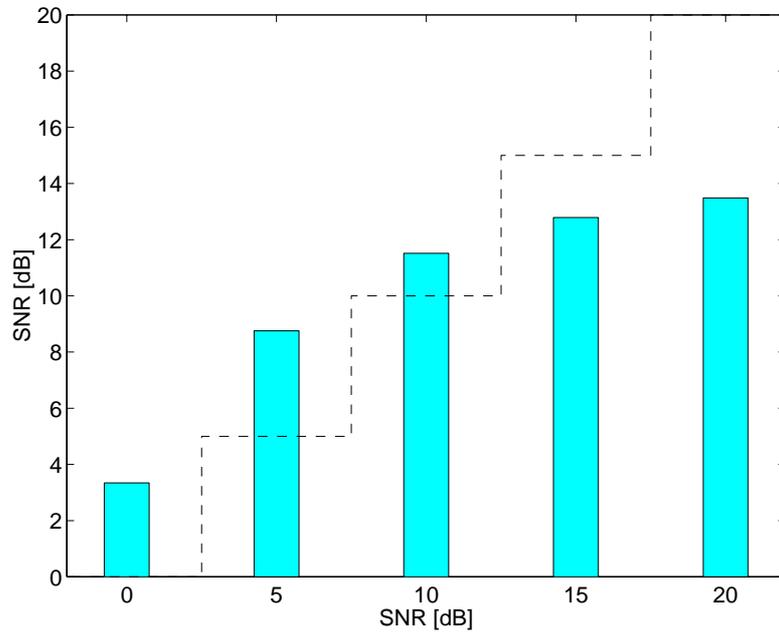


図 5.18: 二重母音の雑音抑圧精度 (SNR)

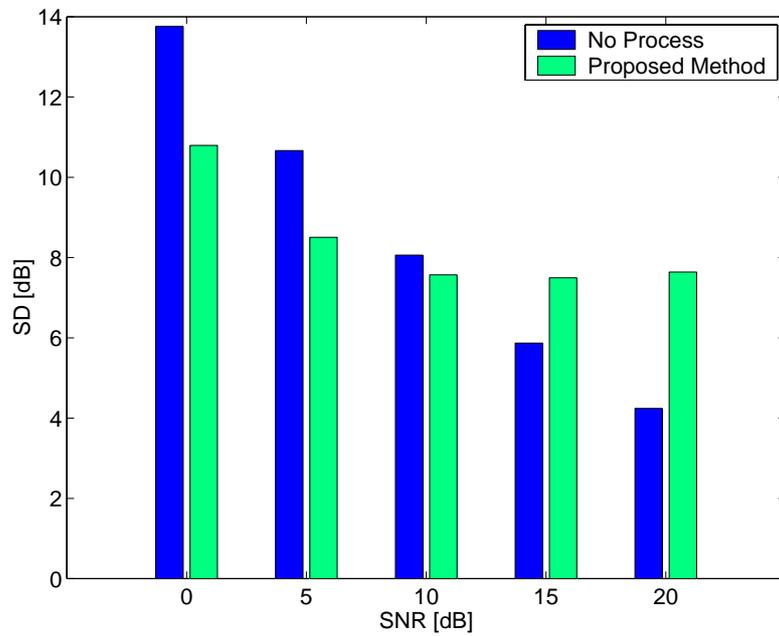


図 5.19: 二重母音の雑音抑圧精度 (SD)

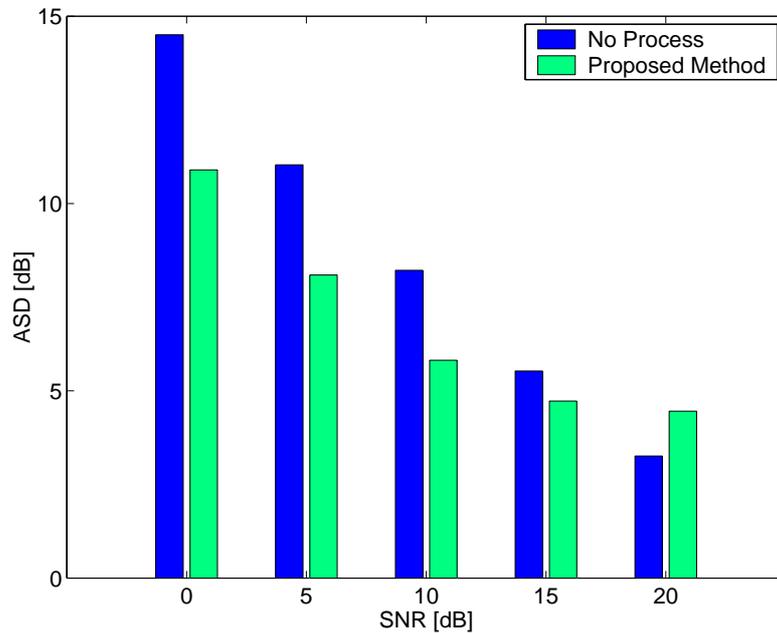


図 5.20: 二重母音の雑音抑圧精度 (ASD)

5.5 考察

まず、白色雑音を付加された音声の雑音抑圧結果から考える。

単母音では SNR、SD、ASD 全ての評価方法で第 3 章のクリーンな音声から得られた基本周波数による雑音抑圧の場合とほぼ同様の結果が得られた。第 3 章のものよりも SNR はやや精度が下がっているが、これは音声の立ち上がり/立ち下がりでは振幅が小さくなるためにその区間の SNR が小さくなり、音声の立ち上がり/立ち下がりの基本周波数が正確に得られないことによるものであると考えられる。抑圧前の SNR が 15dB 以上のときに抑圧後の SNR が悪くなるのは第 3 章で述べたように雑音と共に音声成分の一部も除去してしまうことが原因である。

連続母音では単母音に比べて大きく SNR の精度が低下している。今回のシミュレーションでは帯域幅可変くし形フィルタのパラメータは常に $\epsilon = 0.5$ と一定の値としたが、実際には最適な SNR が得られる ϵ の値は音韻や高調波の時間的変動によって異なる。特に高調波成分が周波数軸上で幅広くなっていると高調波成分が多く除去されてしまうことで SNR の悪化が生じると考えられる。除去される高調波成分が韻律性に強く関わらない部分であれば、やや音色が変わって聞こえることがあるものの聴感上は雑音感の軽減が感じられる。これは聴覚特性を考慮した尺度である ASD では SNR と異なり精度の向上が見

られることに対応していると考えられる。

次に、ピンク帯域雑音を付加された音声の雑音抑圧結果について考察する。

単母音については雑音抑圧前の SNR が小さい場合は SNR、ASD において雑音抑圧処理による改善が見られる。目的音の成分の一部を除去してしまうために SD は悪くなっているが、聴覚マスキングによって聴こえない部分であるため ASD では改善を示していると考えられる。処理前の SNR が 10dB 以上の音声に対しては雑音抑圧処理によって評価値が悪くなってしまう。これも音声の立ち上がり/立ち下がりにおける基本周波数推定エラーによるものと高調波成分の除去によるものであると考えられる。特にピンク帯域雑音では音声と同じ低周波数域でのパワーが強いため、目的音の基本周波数やその高調波と同じ周波数帯域の雑音が除去できないことが SNR や SD、ASD の悪化に繋がっている。

連続母音についてもほぼ同様の傾向が見られる。

二重母音の雑音抑圧結果からは、雑音が大きい場合は SNR、SD、ASD 共に改善が見られる。特に二重母音では白色雑音やピンク帯域雑音に比べ、聴感上の雑音感がかなり軽減しているように感じられる。これは SD が悪化していても ASD が改善されていることに対応しており、二重母音に対しても本研究が有効であると考えられる。

以上の結果から、本研究で提案した雑音環境における基本周波数推定法及び雑音抑圧法が、雑音が白色雑音であるときや二重母音の場合において、有効であることが分かった。しかし、雑音がピンク帯域雑音の場合には、音声の高調波成分の除去や目的音と同じ周波数帯域の雑音の影響により、十分な性能は見られなかった。

第 6 章

結論

6.1 まとめ

本研究では、雑音環境における頑健で高精度な基本周波数推定法及び雑音抑圧法を提案した。

本基本周波数推定法は、帯域幅可変くし形フィルタによる雑音抑圧と雑音環境において頑健な瞬時振幅の Comb Filtering による基本周波数推定法、雑音には弱いが高精度の瞬時周波数を基にした基本周波数推定法を組み合わせることにより、SNR が 5dB と大きな雑音が存在する環境においても雑音のない環境と同様の基本周波数を推定することができた。

また、本方法の耐雑音性能は、瞬時振幅の Comb Filtering による推定法の耐雑音性能に大きく依存しているため、雑音抑圧の前処理の基本周波数推定法がより頑健であればさらなる性能向上が可能であると考えられる。

基本周波数に合わせた帯域幅可変くし形フィルタによる雑音抑圧法では、雑音抑圧前に比べて SNR で 5 ~ 7dB 程度の向上が得られた。ASD による評価で表わされたように、特に白色雑音や二重母音に対する雑音抑圧では、聴感上の雑音感の減少がみられた。聴感上の雑音感の減少は本研究の応用として補聴器の前処理を考える上で重要である。今後、高調波成分の保持や目的音と同じ周波数帯域の雑音の除去等について改善していくことにより、本方法が音声認識や補聴器等の前処理に応用できる可能性を持つと考えている。

6.2 今後の課題

本研究で提案した雑音環境における基本周波数推定法では、音声の立ち上がり/立ち下がりのように雑音に対して振幅が小さい区間では高精度に基本周波数を得ることができないため、低 SNR 時における頑健性をより高めなければならない。また、別の手法で音声の始まりと終わりを検出し、本手法では推定できなかった部分を補間することも考えられる。

雑音抑圧アルゴリズムも SNR が 15dB 以上の環境においては雑音抑圧による音質の低下がみられるが、特に補聴器の前処理として用いるためには音質を損なわないように、高調波成分の保持や目的音の基本周波数や高調波と同じ周波数帯域の雑音の除去を考慮に入れて、基本周波数以外の制約条件を取り入れる等、アルゴリズムを改良する必要がある。

謝辞

本論文の作成にあたり、熱心に御指導して頂きました赤木 正人 教授に厚く御礼を申し上げます。

また、本研究を進める上で、熱心な御議論、多大な御助言をして頂いた鷓木 祐史 氏、水町 光徳 氏、伊藤 一仁 氏を始めとする赤木研究室の皆様にも御礼申し上げます。

最後に、本研究に関わったすべての皆様に感謝致します。

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] 小寺 一興, 補聴器フィッティングの考え方, 診断と治療社, 1999.
- [3] Lawrence Rabiner, Biing-Hwang Juang(古井 貞熙監訳), 音声認識の基礎, NTT アドバンステクノロジー, 1995.
- [4] 阿部 正人, 多数センサによる音源推定, 日本音響学会誌, Vol.51, No.5, pp.384-389, 1999.
- [5] 水町 光徳, 赤木 正人, マイクロホン対を用いたスペクトルサブトラクションによる雑音除去法, 電子情報通信学会論文誌 (A), Vol.J82-A, No.4, pp.503-512, 1999.
- [6] M. P. Cooke, Modelling Auditory Processing and Organization, Ph. D. Thesis, University of Sheffield, 1991.
- [7] G. J. Brown, Computational Auditory Scene Analysis : A Representational Approach, Ph. D. Thesis, University of Sheffield, 1992.
- [8] M. P. Cooke and G. J. Brown, Computational auditory scene analysis : Exploiting principles of perceived continuity, Speech Communication, pp.391-399, North Holland, 13, 1993.
- [9] G. J. Brown and M. P. Cooke, Computational auditory scene analysis, Computer Speech and Language, pp.297-336, 8, 1994.
- [10] D. P. W. Ellis, A Computer Implementation of Psychoacoustic Grouping Rules, Proc. 12th Int. Conf. on Pattern Recognition, 1994.
- [11] D. P. W. Ellis, Prediction-driven computational auditory scene analysis, Ph. D. Thesis, MIT Media Lab., 1996.

- [12] 中谷 智広, 川端 豪, 奥野 博, 計算論的アプローチによる音響ストリームの分離, 日本音響学会聴覚研究会資料, H-93-83, 1993.
- [13] 中谷 智広, 後藤 真孝, 川端 豪, 奥野 博, 調波構造と方向同定に基づく音響ストリーム分離, 日本音響学会秋季講演論文集, 2-3-10, Sep., 1995.
- [14] 大賀 寿郎, 山崎 芳男, 金田 豊, 音響システムとデジタル処理, 電子情報通信学会, 1995.
- [15] Alain de Cheveigné, Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing J. Acoust. Soc. Am., 93(6), pp.3271-3290, June, 1993.
- [16] Alain de Cheveigné, Cancellation model of pitch perception, J. Acoust. Soc. Am., 103(3), pp.1261-1271, March, 1998.
- [17] 鈴木 久喜, ピッチ抽出の今昔, 日本音響学会誌, vol.56, No.2, pp.121-128, 2000.
- [18] A. S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, 1990.
- [19] L. R. Rabiner, R. W. Schafer (鈴木 久喜訳), 音声のデジタル信号処理, コロナ社, 1983.
- [20] 河原 英紀, Parham Zolfaghari, Alain de Cheveigné, Roy D. Patterson, 周波数から瞬時周波数への写像の不動点を用いた音源情報の抽出について, 電子情報通信学会技術報告, SP99-40, July 1999.
- [21] A. M. Noll, Cepstrum pitch determination, J.Acoust.Soc.Am., Vol.41, p.293, 1964.
- [22] 鷓木 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, 電子情報通信学会論文誌 (A), vol.J82-A, No.10, pp.1497-1507, 1999.
- [23] 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 (D-II), vol.J79-D-II, No.11, pp.1771-1781, 1996.
- [24] Toshihiko Abe, Takao Kobayashi and Satoshi Imai, Robust Pitch Estimation with Harmonics Enhancement in Noisy Environments Based on Instantaneous Frequency, ICSLP'96, pp.1277-1280, 1996.

- [25] Mitsunori Mizumachi, Masato Akagi, An Objective Distortion Estimator for Hearing Aids and its Application to Noise Reduction, EUROSPEECH'99, pp.2619-2622, 1999.

学会発表リスト

1. 石本 祐一, 赤木 正人, 雑音が付加された音声の基本周波数推定と雑音抑圧, 電子情報通信学会技術報告, 2000 年 3 月発表予定.
2. 石本 祐一, 赤木 正人, 雑音中の音声基本周波数推定法の提案, 日本音響学会 平成 12 年春期研究発表会, 2-7-7, 2000.