

(千葉大学学位申請論文)

# 雑音混入音声の音質改善を目的とした 雑音除去方式の研究

2006 年 1 月

千葉大学大学院自然科学研究科  
情報科学専攻情報システム科学講座

野村 行弘

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
1.1	研究の背景	1
1.2	適応ノイズキャンセラ	4
1.2.1	適応ノイズキャンセラの構成	4
1.2.2	線形システムに対応した適応フィルタ	4
1.2.3	非線形システムに対応した適応フィルタ	7
1.3	スペクトルサブトラクション	11
1.4	研究の目的および本論文の構成	13
<b>第2章</b>	<b>並列リカレントニューラルフィルタを利用した適応ノイズキャンセラ</b>	<b>22</b>
2.1	はじめに	23
2.2	PRNFを利用した適応ノイズキャンセラ	24
2.2.1	提案方法のモデル	24
2.2.2	PRNFのモデル	25
2.3	シミュレーション	29
2.3.1	RNFによるシミュレーション結果	29
2.3.2	PRNFの有効性	34
2.3.3	PRNFの学習方法の比較	36
2.3.4	実際の音声への適用	38
2.3.5	他のフィルタとの比較	40
2.4	まとめ	44
<b>第3章</b>	<b>雑音量に依存しない音声領域と雑音領域の判別法を利用したスペクトルサブトラクション</b>	<b>48</b>

---

3.1	はじめに . . . . .	49
3.2	従来方法 . . . . .	49
3.3	提案方法 . . . . .	55
3.4	性能評価 . . . . .	57
3.4.1	シミュレーション条件 . . . . .	58
3.4.2	判別結果の比較 . . . . .	62
3.4.3	板倉-斉藤ひずみ距離および segmental SNR . . . . .	66
3.4.4	MOS テスト . . . . .	69
3.5	まとめ . . . . .	70
<b>第4章</b>	<b>スペクトログラム上の特徴量に基づく音声領域と雑音領域の判別法を利用したスペクトルサブトラクション</b>	<b>72</b>
4.1	はじめに . . . . .	73
4.2	従来方法の問題点 . . . . .	74
4.2.1	変位値に基づく雑音スペクトルの推定 . . . . .	74
4.2.2	音声領域と雑音領域の判別法を用いたスペクトルサブトラクション . . . . .	75
4.3	提案方法 . . . . .	75
4.3.1	提案方法の構成 . . . . .	75
4.3.2	音声領域と雑音領域の判別 . . . . .	76
4.3.3	判別後の処理 . . . . .	79
4.4	性能評価 . . . . .	80
4.4.1	シミュレーション条件 . . . . .	80
4.4.2	音声波形およびスペクトログラム . . . . .	83
4.4.3	Segmental SNR . . . . .	85
4.4.4	板倉-斉藤ひずみ距離 . . . . .	87
4.4.5	MOS テスト . . . . .	89
4.5	まとめ . . . . .	90
<b>第5章</b>	<b>モルフォロジー処理を用いたスペクトルサブトラクションにおけるミュージャカルノイズ除去</b>	<b>92</b>

---

5.1	はじめに . . . . .	93
5.2	従来方法 . . . . .	94
5.3	提案方法 . . . . .	98
5.3.1	提案方法の構成 . . . . .	98
5.3.2	モルフォロジー処理 . . . . .	100
5.4	性能評価 . . . . .	103
5.4.1	シミュレーション条件 . . . . .	103
5.4.2	計算回数の比較 . . . . .	103
5.4.3	ミュージカルノイズの除去および弱い音声成分の保存につい ての検討 . . . . .	104
5.4.4	スペクトログラムおよび segmental SNR の改善度 . . . . .	105
5.4.5	MOS テスト . . . . .	112
5.5	まとめ . . . . .	113
<b>第6章</b>	<b>結論</b>	<b>115</b>
	<b>謝辞</b>	<b>118</b>
	<b>本研究に関する参考資料</b>	<b>119</b>



# 目 次

1.1	適応ノイズキャンセラの基本構成 . . . . .	5
1.2	適応フィルタの構成 . . . . .	6
1.3	適応 Volterra フィルタを用いた非線形システム同定のブロック図 . . . . .	7
1.4	ニューラルフィルタの構成 . . . . .	9
1.5	音声のスペクトログラム . . . . .	13
1.6	本論文の構成 . . . . .	14
2.1	PRNF を利用した雑音キャンセラの基本構成 . . . . .	25
2.2	PRNF の構成 . . . . .	28
2.3	パスが指数関数モデルの場合の RNF の学習 . . . . .	31
2.4	パスが線形特性の場合の RNF の出力波形 . . . . .	32
2.5	パスが指数関数モデルの場合の RNF の出力波形 . . . . .	32
2.6	RNF の入出力 SNR の関係 . . . . .	33
2.7	パスが指数関数モデルの場合の PRNF の出力波形 . . . . .	35
2.8	パスが指数関数モデルの場合の PRNF の入出力 SNR の関係 . . . . .	35
2.9	パスが指数関数モデルの場合の PRNF の学習曲線 . . . . .	37
2.10	パスが NARMAX モデルの場合の PRNF の出力波形 . . . . .	39
2.11	各フィルタの入出力 SNR の関係 . . . . .	42
2.12	PRNF と AVF の収束特性 . . . . .	43
3.1	Yoon & Yoo の方法の構成 . . . . .	50
3.2	Yoon & Yoo の方法による class の分類 . . . . .	51
3.3	提案方法の構成 . . . . .	55
3.4	提案方法における class の分類の例 . . . . .	56
3.5	パラメータ $d$ に対する板倉-斉藤ひずみ距離 . . . . .	60

---

3.6	パラメータ $d$ に対する segmental SNR の改善度	61
3.7	従来方法と提案方法の判別結果の比較	64
3.8	入力 SNR における音声領域の判別率	65
3.9	入力 SNR における板倉-斉藤ひずみ距離	67
3.10	入力 SNR における segmental SNR の改善度	68
4.1	提案方法の構成	76
4.2	観測信号のスペクトルの標準偏差	77
4.3	$SD(i, r)$ のヒストグラム	79
4.4	バンド幅 ( $BW$ ) に対する客観的評価の結果	82
4.5	パラメータ $q$ に対する客観的評価の結果	82
4.6	処理信号の時間波形とスペクトログラム	84
4.7	入力 SNR における segmental SNR の改善度	86
4.8	入力 SNR における板倉-斉藤ひずみ距離	88
5.1	Goh らの方法の構成図	94
5.2	音声のスペクトログラム	97
5.3	提案方法の構成図	99
5.4	$3 \times 3$ の窓によるモルフォロジー処理	100
5.5	モルフォロジー処理の例	101
5.6	提案方法で用いる長方形窓	102
5.7	モルフォロジー処理の結果	107
5.8	入力 SNR における (a) ミュージカルノイズ除去率, (b) 弱い音声成分 の保存率	108
5.9	音声のスペクトログラム	109
5.10	音声のスペクトログラム	110
5.11	入力 SNR における segmental SNR の改善度	111

# 表 目 次

1.1	単一マイクロホンによる雑音除去方式の比較 . . . . .	3
2.1	1回の学習にかかる計算回数 . . . . .	34
2.2	各学習方法による収束回数 . . . . .	37
2.3	PRNF, NF および AVF の1回の学習にかかる計算回数 . . . . .	42
3.1	STFTの周波数バンドとクリティカルバンドの関係 . . . . .	54
3.2	従来方法および提案方法のパラメータ . . . . .	59
3.3	MOSテストの結果 (SNR=5, 10 dB) . . . . .	69
4.1	MOSテストの結果 (SNR=5 dB) . . . . .	89
5.1	計算回数の比較 . . . . .	104
5.2	MOSテストの結果 (SNR=5, 10 dB) . . . . .	112

---

# 第1章

---

## はじめに

### 1.1 研究の背景

近年の携帯電話の普及や音声の符号化、音声認識システムなどのデジタル音声信号処理技術の発展に伴い、雑音除去方式の需要が高まっている。例えば、携帯電話のハンズフリー機能や、テレビ会議システムなどの拡声通話系では周囲の雑音による通話品質の劣化が問題である。そのため、観測信号から雑音の除去を行い、通話品質を改善することが必要である [1]。また、音声認識システムにおいては、雑音環境下では認識率が大幅に低下することが知られており、前処理として雑音除去を行うことにより、認識率の向上を図っている [2]。このように、雑音除去方式は、現在のデジタル音声信号処理において欠くことのできない技術となっており、今後その重要性は益々高まると考えられる。

デジタル音声信号に対する雑音除去方式に関する研究は非常に長い歴史を持っており、古くは低域フィルタや帯域制限フィルタなどのフィルタリングによって行われていた。フィルタリングによる方法では、ハムノイズなどのように雑音が特定の周波数にしか存在しない場合や雑音の周波数帯域が音声の可聴周波数帯域外にある場合などにおいては有効である。しかし、多くの雑音は周波数成分も広帯域でありかつ、ランダム性を有しているため、フィルタリングによる方法では対処療法的な処理をしたにすぎない。

1970年代後半以降、デジタル音声信号の雑音除去方式に関する研究が盛んになった。デジタル音声信号の雑音除去方式に関する研究が盛んになった背景には、

- 雑音環境, 電磁環境の悪化による音声通信の品質劣化
- 高速フーリエ変換 (FFT), 線形予測をはじめとする音声に関する信号処理技術の進歩
- IC, LSI などの素子の進歩によりハードウェア製作の可能性が大きくなったこと
- 音声認識装置, 低ビット・レート伝送方式におけるニーズ

などが同時期に生じたことによることが大きい [3].

デジタル音声信号の雑音除去方式は, 観測信号を複数のマイクロホンを用いて取得する方式と, 単一のマイクロホンを用いて取得する方式の2つに分類される.

前者の方法としては, 適応ノイズキャンセラ [4-7] や, マイクロホンアレー [8,9] などが提案されている. 適応ノイズキャンセラはシステム同定の手法で雑音を推定し, 観測信号から減算して目的の信号を取り出すための適応フィルタである. しかし, 適応ノイズキャンセラは観測信号が入力される主入力端子のほかに, 雑音を取得する参照入力端子が必要である. マイクロホンアレーは, 単一のマイクロホンでは得ることができない信号の到来方向という空間情報を利用して雑音除去を行う方法である. 具体的には, 雑音の到来方向にマイクロホンの指向特性の死角を形成することにより, 雑音の低減を行う. しかし, マイクロホンアレーではマイクロホンの数を雑音源の数より多くする必要があるため, 雑音源の数が既知である必要がある. さらに, 小型の携帯型機器への適用を考慮した場合, 複数のマイクロホンを用いる方式の適用は機器の小型化に制限を与える.

一方, 単一のマイクロホンによる方式にはスペクトルサブトラクション (spectral subtraction, SS) [10-17], くし形フィルタ [18-20], SPAC (speech processing system by use of auto-correlation function) [21-24], カルマンフィルタを用いた方法 [25-28], ウィナーフィルタを用いた方法 [29,30] などがある. この中で, SS は雑音が定常であることを仮定し, 観測信号から雑音のスペクトルの推定を行い, それを差し引くことにより雑音除去を行う方法である. SS は計算が比較的簡単であるという利点から, 広く用いられている. しかし, 雑音のスペクトルの推定誤差などが原因で処理後の音声信号にミュージカルノイズが発生するという問題点があり, この問題点を改良する方法が数多く提案されている [13,14,16,17]. くし形フィルタは, 音声信号のピッチ周期 (基本周波数) および高調波を通過するフィルタを通すことにより雑音の除去を図る方法である. しかし, 観測信号から高精度にピッチ周期を推定する必

表 1.1 単一マイクロホンによる雑音除去方式の比較

Table 1.1 Comparison of noise reduction system for single microphone.

方式	使用する情報	雑音抑圧効果	ピッチ抽出	計算量
スペクトルサブトラクション	雑音の減算	中	不要	小
くし形フィルタ	音声の周期性	大	要	小
SPAC	相関関数	中	要	小
カルマンフィルタ	AR モデル	中	不要	大
ウィーナーフィルタ	雑音の抑圧	中	不要	小

要があるという問題点がある。SPACは、音声信号の有する周期性に着目して、音声波形を自己相関領域でつなぎ合わせる方法である。SPACでは、ピッチ周期の推定はくし形フィルタの場合のように精度を必要とせず、優れた雑音除去効果が得られることが知られている。しかし、2乗ひずみや高調波ひずみが生じるために明瞭度が減少するという問題点がある。カルマンフィルタを用いる方法は、音声信号に自己回帰 (auto regressive, AR) モデルを用いてモデル化を行い、カルマンフィルタによって観測信号から原音声の推定を行う。この方法は信号処理分野における理論的な背景が確立されているものの、計算回数が多いという問題点がある。ウィーナーフィルタによる方法は、原音声および雑音のパワースペクトルが既知の条件下で平均2乗誤差を最小化するフィルタ処理を行う。ウィーナーフィルタの設計方法としては、音声信号のARスペクトルを用いて反復的にウィーナーフィルタを設計する方法 [29] やSSの原理を利用して、原音声および雑音のパワースペクトルを算出する方法 [30] などがある。

表 1.1 に単一マイクロホンによる雑音除去方式の比較表を示す。表 1.1 より、各方法にはそれぞれ一長一短があり、用途に応じて適切な方法を選択する必要がある。

本論文では以上のような背景から、雑音混入音声の音質改善を実現するために、特に適応ノイズキャンセラおよびスペクトルサブトラクションによる雑音除去について研究を進めたものである。

## 1.2 適応ノイズキャンセラ

本節では、複数のマイクロホンによる雑音除去方式である、適応ノイズキャンセラについて概説する。また、適応ノイズキャンセラに用いられる適応フィルタについても概説する。

### 1.2.1 適応ノイズキャンセラの構成

観測信号に雑音が含まれている場合、システム同定の手法で雑音を推定し、観測信号から減算して目的の信号を取り出す方法として Widrow らが提案した適応ノイズキャンセラがある [4-7]。図 1.1 に適応ノイズキャンセラの基本構成を示す。適応ノイズキャンセラは主入力端子と少なくとも 1 個の参照入力端子を持つ。主入力端子の信号は目的の信号  $s(k)$  と雑音源  $x(k)$  がパスを介して得られる雑音  $d(k)$  の和からなる観測信号である。一方、参照入力端子の信号には雑音源からの信号  $x(k)$  が入力される。適応フィルタは雑音源から主入力端子までのパスを推定するために動作する。適応フィルタの出力を  $y(k)$  とすると、システムの入力  $z(k)$  は

$$z(k) = s(k) + d(k) - y(k) \quad (1.1)$$

となる。ここで、適応フィルタで推定したパスが雑音源から主入力端子までのパスと等しい場合、 $d(k) = y(k)$  となり、観測信号に含まれる雑音を完全に除去できる。

### 1.2.2 線形システムに対応した適応フィルタ

図 1.2 にこの基本構成を示す。ただし、 $z^{-1}$  は 1 サンプル時間の時間遅れを表す。時刻  $k$  で信号  $x(k)$  を観測するとき、適応フィルタは時刻  $k$  から時刻  $k - N + 1$  までの  $N$  個の信号  $x(k), x(k-1), \dots, x(k-N+1)$  に、それぞれフィルタ係数  $w_0, w_1, \dots, w_{N-1}$  を乗じた信号の和  $y(k)$  を出力する。このような構造をもつフィルタのことをトランスバーサルフィルタとよぶ。このフィルタの働きを定式化すると、

$$y(k) = \sum_{n=0}^{N-1} w_n x(k-n) = \mathbf{w}_k^T \boldsymbol{\phi}_k = \boldsymbol{\phi}_k^T \mathbf{w}_k \quad (1.2)$$

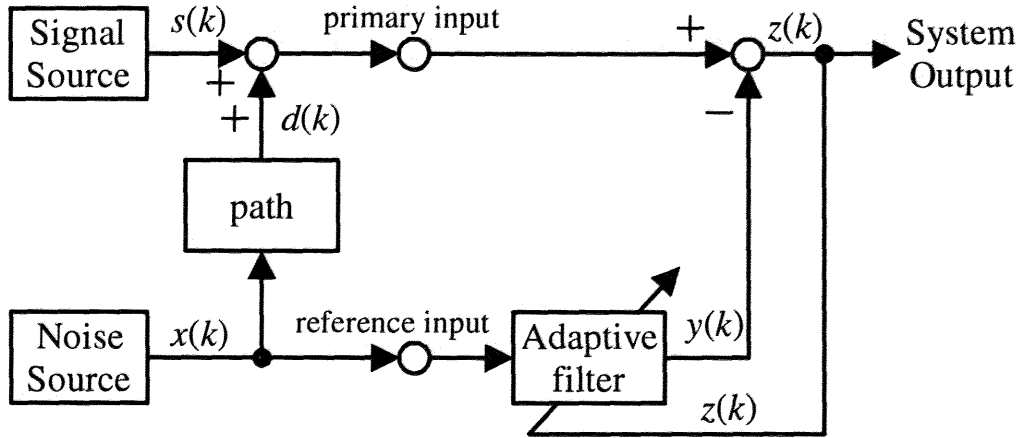


図 1.1 適応ノイズキャンセラの基本構成

Fig. 1.1 Configuration of adaptive noise canceller

となる。ただし、 $N$ はフィルタ次数、 $\phi_k, \mathbf{w}_k$ はそれぞれ時刻 $k$ におけるフィルタ入力ベクトル、フィルタ係数ベクトルであり、

$$\phi_k = [x(k), x(k-1), \dots, x(k-N+1)]^T \quad (1.3)$$

$$\mathbf{w}_k = [w_{0,k}, w_{1,k}, \dots, w_{N-1,k}]^T \quad (1.4)$$

と定義される。適応フィルタでは、フィルタ出力 $y(k)$ があらかじめ決めておいた目的の信号 $d(k)$ に近づくように、つまり、誤差

$$e(k) = d(k) - y(k) = d(k) - \phi_k^T \mathbf{w}_k \quad (1.5)$$

が小さくなるように、 $\mathbf{w}_k$ が時々刻々と調節される。その調整手続きを行うのが適応信号処理アルゴリズムであり、その具体的な方法の1つがLMS(least mean square)法である。

LMS法は、2乗平均誤差を最急降下法に基づいて最小にする方式で、その適応性能のよさ、計算量の少なさ、理解のしやすさなどから、現在最も広く用いられている [31, 32]。

図 1.2において、 $x(k), d(k)$ を平均0の定常な確率変数として、評価関数を $J(\mathbf{w}) = E[e(k)^2]$ で与えたとき、

$$J(\mathbf{w}) = E[d(k)^2] - 2\mathbf{p}^T \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (1.6)$$



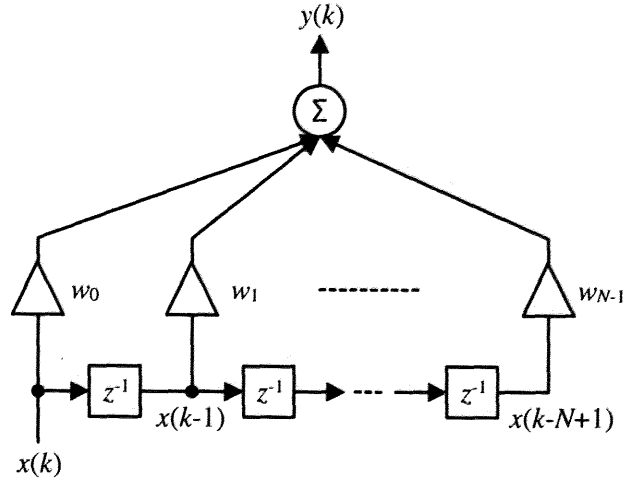


図 1.2 適応フィルタの構成

Fig. 1.2 Configuration of adaptive filter

が成り立つ。ただし、 $\mathbf{R}$ はフィルタ入力ベクトル  $\phi_k$  の相関行列、 $\mathbf{p}$ は  $\phi_k$  と  $d(k)$  の相互相関ベクトルであり、

$$\mathbf{R} = E[\phi_k \phi_k^T] \quad (1.7)$$

$$\mathbf{p} = E[d(k) \phi_k] \quad (1.8)$$

と定義される。  $J(\mathbf{w})$  を最小にする係数ベクトルを求めるために、最急降下法を適用すると、  $J(\mathbf{w})$  の  $\mathbf{w} = \mathbf{w}_k (= [w_{0,k}, w_{1,k}, \dots, w_{p-1,k}]^T)$  における勾配ベクトルは、

$$\nabla J(\mathbf{w}_k) = \left( \frac{\partial E[e(k)^2]}{\partial \mathbf{w}} \right)_{\mathbf{w}=\mathbf{w}_k} = -2\mathbf{p} + 2\mathbf{R}\mathbf{w}_k \quad (1.9)$$

となり、  $\mathbf{w}$  の更新値  $\mathbf{w}_{k+1}$  は次の繰り返し計算によって求まる。

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \mu \nabla J(\mathbf{w}_k) \\ &= (\mathbf{I} - 2\mu\mathbf{R})\mathbf{w}_k + 2\mu\mathbf{p} \end{aligned} \quad (1.10)$$

ただし、  $\mu$  はステップサイズと呼ばれる小さな正の定数である。

式(1.9)の勾配ベクトル  $\nabla J(\mathbf{w}_k)$  において、期待値  $E[e(k)^2]$  を瞬時値  $e(k)^2$  に置き換えると、

$$\hat{\nabla} J(\mathbf{w}_k) = \left( \frac{\partial e(k)^2}{\partial \mathbf{w}} \right)_{\mathbf{w}=\mathbf{w}_k} = 2 \left( e(k) \frac{\partial e(k)}{\partial \mathbf{w}} \right) = -2e(k)\phi_k \quad (1.11)$$

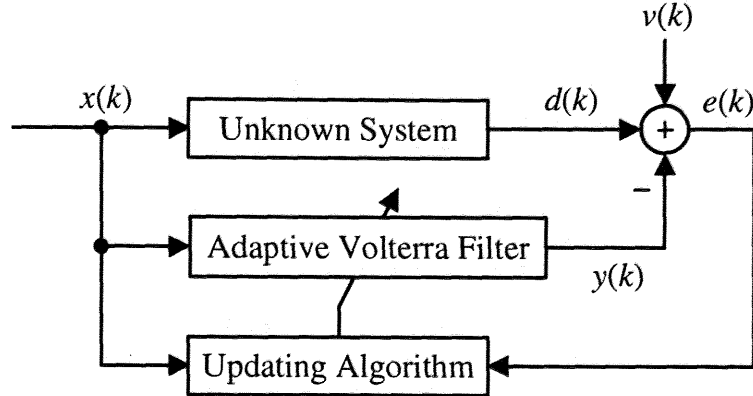


図 1.3 適応 Volterra フィルタを用いた非線形システム同定のブロック図

Fig. 1.3 Block diagram of nonlinear system identification using adaptive Volterra filter

と近似できる。ただし,

$$e(k) = d(k) - \phi_k^T \mathbf{w}_k \quad (1.12)$$

である。このとき、式 (1.10) の最急降下法の繰り返し計算式は

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \mu \hat{\nabla} J(\mathbf{w}_k) \\ &= \mathbf{w}_k + 2\mu e(k) \phi_k \end{aligned} \quad (1.13)$$

となる。式 (1.13) は Widrow-Hoff の LMS アルゴリズムまたは適応 LMS アルゴリズムと呼ばれている。

LMS 法は適応性能のよさ、計算量の少なさなどから、広く用いられている。しかし、実環境で利用する場合、性能が大きく劣化することも多々発生する。その要因の一つとしては、対象システムが非線形特性を持っていることが考えられる。適応ノイズキャンセラにおいても雑音源から観測信号までの伝達特性を線形システムと仮定したことが多いが [4]- [5], 状況によっては伝達特性は線形特性だけでなく非線形特性を考慮する必要がある [7, 33]. その場合、線形システムに対応した適応フィルタでは性能に限界があり、非線形システムに対応したものを利用する必要がある。

### 1.2.3 非線形システムに対応した適応フィルタ

非線形システムに対応した適応フィルタとして Volterra 級数展開 [34] を利用した適応 Volterra フィルタ (adaptive Volterra filter, AVF) [34, 35] およびニューラルネッ

トワークを利用したニューラルフィルタ (neural filter, NF) [36, 37] が挙げられる。

### A. 適応 Volterra フィルタ

図 1.3 に非線形システムを AVF で同定する場合の基本的なブロック図を示す。ただし、 $v(k)$  は外乱、 $e(k)$  は誤差信号である。このとき、AVF の基となる Volterra 級数展開は次式で定義される。

$$\begin{aligned}
 y(k) = & h_0 + \sum_{n_1=0}^{N_1-1} h_1(n_1)x(k-n_1) \\
 & + \sum_{n_1=0}^{N_2-1} \sum_{n_2=0}^{N_2-1} h_2(n_1, n_2)x(k-n_1)x(k-n_2) \\
 & + \dots
 \end{aligned} \tag{1.14}$$

ここで、 $x(k)$  と  $y(k)$  はそれぞれ時刻  $k$  での入力信号と出力信号を表している。また、 $h_l(n_1, \dots, n_l)$  は  $l$  次の離散 Volterra 核である。ここで、 $h_l(n_1, \dots, n_l)$  は一般的に対称性の性質をもつ。すなわち、いかなる  $n_1, \dots, n_l$  の順列の入れ替えを行っても  $h_l(n_1, \dots, n_l)$  は一般性を失うことなく不変である。式 (1.14) において、定数  $h_0$  はオフセット (直流) の項であり、 $h_1(n_1)$  は、無限長の線形インパルス応答であり、 $h_l(n_1, \dots, n_l)$  はシステムの非線形特性を特徴付ける  $l$  次のインパルス応答である。

AVF の入出力関係は式 (1.14) に示したように、出力がフィルタ係数 (Volterra 核) に対して線形である。そのため、AVF のフィルタ係数の更新に従来の線形適応フィルタで用いられている適応アルゴリズムを容易に適用できる。AVF ではフィルタ係数更新アルゴリズムとして LMS 法、RLS (recursive least squares) 法などがよく用いられている。

AVF における LMS 法の更新式は線形システムの場合と同様に、

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k)\mathbf{x}(k) \tag{1.15}$$

$$e(k) = d(k) - \mathbf{h}^T(k)\mathbf{x}(k) \tag{1.16}$$

と表せる。ただし、 $\mathbf{h}(k), \mathbf{x}(k)$  はそれぞれ時刻  $k$  におけるフィルタ係数ベクトル、入力信号ベクトルである。

AVF は非線形特性をもつシステムに対応する有効な適応フィルタの 1 つであるが、フィルタ係数更新アルゴリズムの計算回数およびフィルタ規模の 2 点で問題がある。

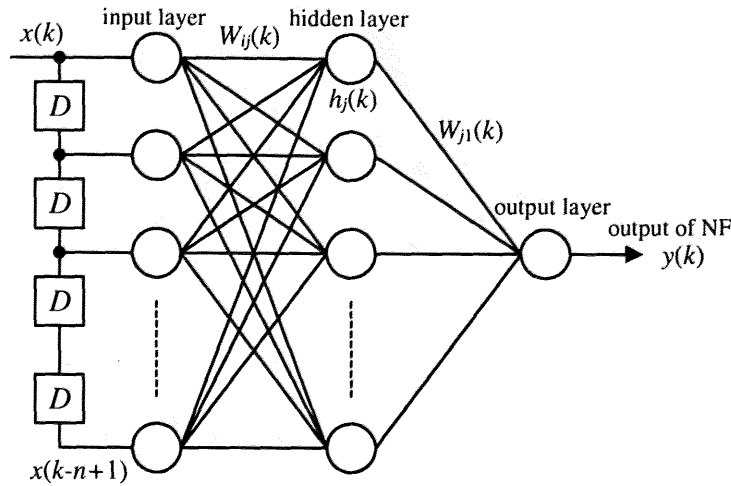


図 1.4 ニューラルフィルタの構成

Fig. 1.4 Configuration of Neural filter

まず、更新アルゴリズムの計算回数の点については、更新アルゴリズムにLMS法を用いた場合は計算回数が $O(N^2)$  ( $N$ はフィルタの記憶長)と少ないが、収束速度がRLS法に比べて明らかに遅い。一方、RLS法は非常に良好な収束特性を有するが、計算回数が $O(N^4)$ と膨大になる。

フィルタの規模の点では、実際の実システムでは高次の非線形項が存在するため、同定精度の向上には高次のAVFが必要となる。しかし、AVFはフィルタの次数を上げるに従いフィルタの規模は指数関数的に増大するため、現在のハードウェア性能では2次AVFが限界と考えられている。

## B. ニューラルフィルタ

NFは、非線形の入出力関係を持つニューラルネットワークを適応フィルタに応用したものである。NFはニューラルネットワークの特徴である、有限個のパターンから学習した内容を生かして未知の信号に対応するという汎化能力があるため、高次非線形性のフィルタリングを比較的小規模な構成で行うことができる。

図1.4にNFの構成を示す。NFは入力層、中間層および出力層の3層構造である。ただし、 $\boxed{D}$ は遅延素子を表す。NFの入力には $n$ 個の信号 $x(k), x(k-1), \dots, x(k-n+1)$ が入力される。

ここで、各層の働きを説明する。入力層では入力された信号をそのまま出力される。

次に、中間層には入力層の出力  $x(k-i)$  ( $0 \leq i \leq n-1$ ) と結合係数  $W_{ij}$  の重み付け和

$$h_j(k) = \sum_{-i=0}^{n-1} W_{ij}x(k-i) \quad (1.17)$$

が入力され、入出力関数  $f_{mid}()$  を通したものが出力となる。入出力関数  $f_{mid}()$  にはシグモイド関数や線形関数などが用いられる。

最後に、出力層には中間層の出力と結合係数  $W_{j1}$  の重み付け和

$$o(k) = \sum_{-j=0}^{n-1} W_{j1}x(k-j) \quad (1.18)$$

が入力され、入出力関数  $f_{out}()$  を通したもの

$$y(k) = f_{out}(o(k)) \quad (1.19)$$

を出力し、これがNFの出力となる。

NFの結合係数の学習にはバックプロパゲーション法が用いられる。バックプロパゲーション法はRumelhartらにより提案された教師信号付き学習方法である [38]。バックプロパゲーション法はNFの出力  $y(k)$  と教師信号  $T(k)$  の2乗誤差

$$J(k) = \frac{1}{2} \{T(k) - y(k)\}^2 \quad (1.20)$$

を最小にするように結合係数を修正していく。中間層と出力層との間の結合係数  $W_{j1}$  の修正量  $\Delta W_{j1}$  は

$$\begin{aligned} \Delta W_{j1}(k+1) &= -\eta \frac{\partial J(k)}{\partial W_{j1}(k)} + \alpha \Delta W_{j1}(k) \\ &= -\eta (T(k) - y(k)) f'_{out}(o(k)) h_j(k) + \alpha \Delta W_{j1}(k) \end{aligned} \quad (1.21)$$

となる。ここで、 $\eta$  は学習係数といい、1回あたりどのくらい結合係数を変化させるかを表す値である。 $\alpha$  は慣性係数といい、 $0 \leq \alpha < 1$  とする。また、 $\alpha \Delta W(k)$  のことをモーメント項とよぶ。同様に、中間層と出力層との間の結合係数  $W_{ij}$  の修正量  $\Delta W_{ij}$  は次式で与えられる。

$$\begin{aligned} \Delta W_{ij}(k+1) &= -\eta \frac{\partial J(k)}{\partial W_{ij}(k)} + \alpha \Delta W_{ij}(k) \\ &= -\eta (T(k) - y(k)) f'_{out}(o(k)) h_j(k) W_{j1} f'_{mid}(h_j(k)) x(k-i) \\ &\quad + \alpha \Delta W_{ij}(k) \end{aligned} \quad (1.22)$$

結合係数の修正値は式 (1.21), 式 (1.22) より,

$$W_{j1}(k+1) = W_{j1}(k) + \Delta W_{j1}(k+1) \quad (1.23)$$

$$W_{ij}(k+1) = W_{ij}(k) + \Delta W_{ij}(k+1) \quad (1.24)$$

となる.

NFは高次非線形性のフィルタリングを比較的小規模な構成で行うことが可能であるが, 学習に時間がかかるという問題点を持っている. これはNFの学習で用いられているバックプロパゲーション法の持つ問題点といえる. NFを実際のシステムに適用することを考えた場合, 学習時間を減らすことが必要となる. この問題を解消する方法として, バックプロパゲーション法を改良して学習の収束を速める方法が提案されている [39–43]. 具体的には, 学習が早く収束するように学習係数および慣性係数を動的に変更する方法 [39], ニューラルネットワークの情報伝達構造を用いて結合係数の初期値を設定する方法 [40], 非線形最適化法を利用する方法 [41], 誤差逆伝搬量の特異点解消を行う方法 [42], 入出力関数を動的に変化させる方法 [43] などがある.

### 1.3 スペクトルサブトラクション

本節では, 単一のマイクロホンによる雑音除去方式である, スペクトルサブトラクション (SS) について概説する.

観測信号を  $y(k)$  とすると,  $y(k)$  は原音声  $s(k)$  と雑音  $n(k)$  との和, すなわち

$$y(k) = s(k) + n(k) \quad (1.25)$$

$$Y(\omega, r) = S(\omega, r) + N(\omega, r) \quad (1.26)$$

となる. ここで,  $Y(\omega, r)$ ,  $S(\omega, r)$  および  $N(\omega, r)$  はそれぞれ  $r$  番目のフレームにおいて  $y(k)$ ,  $s(k)$  および  $n(k)$  を短時間フーリエ変換 (STFT) したものである. ここで, 原音声  $s(k)$  と雑音  $n(k)$  に相関がないと仮定する. もし, 雑音のスペクトル  $|N(\omega, r)|$  が  $|\hat{N}(\omega, r)|$  と推定される場合, SSによる目的の信号のスペクトルの推定値  $|\hat{S}(\omega, r)|$  は以下の式で得られる.

$$|\hat{S}(\omega, r)| = \begin{cases} \left( |Y(\omega, r)|^2 - \alpha |\hat{N}(\omega, r)|^2 \right)^{1/2} & \text{if } |Y(\omega, r)|^2 > \alpha |\hat{N}(\omega, r)|^2 \\ 0 & \text{otherwise} \end{cases} \quad (1.27)$$

ただし、 $\alpha$ はサブトラクション係数で、 $1 \leq \alpha$ の範囲で設定する。 $\alpha$ は観測信号から雑音スペクトルを減算する割合をコントロールする役割を持っている。 $|\hat{S}(\omega, r)|$ が得られれば、短時間逆フーリエ変換 (ISTFT) を行うことにより処理した音声  $\hat{s}(k)$  を得る。

$$\hat{s}(k) = \text{IFFT}[|\hat{S}(\omega, r)| \cdot e^{j\arg(Y(\omega, r))}] \quad (1.28)$$

雑音スペクトル  $|\hat{N}(\omega, r)|$  を推定する方法は様々あり、一般的には無音区間における観測信号のスペクトルの平均を雑音スペクトルとする方法 [10] が用いられる。その他の方法としては、過去のフレームの観測信号のスペクトルの最小値を雑音スペクトルと推定する最小統計法 [12]、過去のフレームにおける観測信号のスペクトルの変位値に基づいて雑音スペクトルを推定する方法 [15] などが知られている。

図 1.5(a) に原音声（女性が「休み無く打ち寄せてはさっと引いていく白い波」と発声したもの）のスペクトログラム、図 1.5(b) に原音声に SNR=10 dB でホワイトノイズを付加したときの観測信号のスペクトログラム、図 1.5(c) に図 1.5(b) の観測信号に対して  $\alpha = 1.8$  を用いた SS によって処理した音声のスペクトログラムをそれぞれ示す。図 1.5 においては濃い点ほどパワーが強いことを示している。図 1.5(c) より、SS は 1 入力で雑音の除去が行えるものの、処理後のスペクトログラムには複数の孤立点が現れていることがわかる。これをミュージカルノイズといい、SS 特有の問題点である。SS では一般的に無音区間のスペクトルの平均を雑音スペクトルの推定値  $|\hat{N}(\omega, r)|$  としている。そのため、実際の雑音スペクトル  $|N(\omega, r)|$  と  $|\hat{N}(\omega, r)|$  との間には誤差が発生する。よって、SS では  $\alpha$  が雑音除去に重要な役割を果たす。 $\alpha$  が小さい場合、SS 処理後の音声には雑音スペクトルの引き残しが存在し、図 1.5(c) のように引き残しの部分が孤立点として現れる。さらに、孤立点が現れる周波数が短時間で変化するため、観測信号に含まれている雑音よりも耳障りに聞こえる。ミュージカルノイズを発生させない最も簡単な方法は、 $\alpha$  を大きくして SS を行い、雑音スペクトルの引き残しをなくすことである。図 1.5(d) に図 1.5(b) の観測信号に対して  $\alpha = 16$  を用いた SS によって処理した音声のスペクトログラムを示す。図 1.5(d) より、大きい  $\alpha$  で SS を行うことによりミュージカルノイズが発生していないことがわかる。しかし、弱い音声成分まで除去されてしまう。そのため、弱い音声成分を保存しつつミュージカルノイズが発生しない方法が望まれており、SS によって処理した音声に後処理を施すことによりミュージカルノイズを除去する方法 [13]、人間の

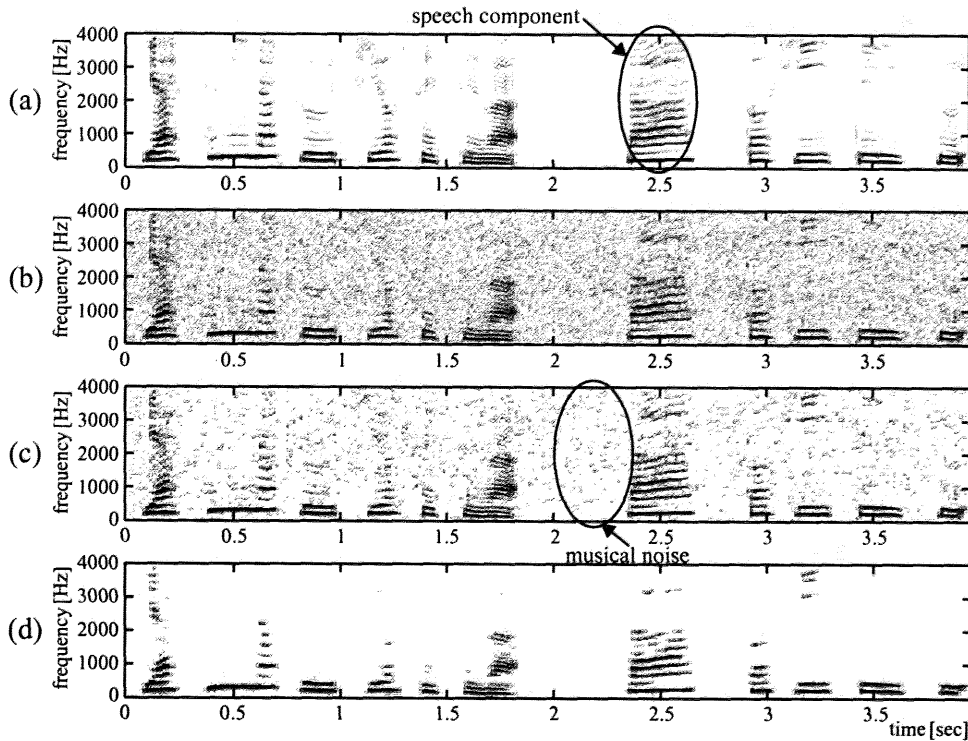


図 1.5 音声のスペクトログラム, (a) 原音声, (b) 観測信号 (ホワイトノイズ, SNR=10 dB), (c)  $\alpha = 1.8$  を用いた SS によって処理した音声, (d)  $\alpha = 16$  を用いた SS によって処理した音声

Fig. 1.5 Spectrograms of (a) Clean speech, (b) Observation signal (additive white noise with SNR=10 dB), (c) Enhanced speech by spectral subtraction for  $\alpha = 1.8$ , (d) Enhanced speech by spectral subtraction for  $\alpha = 16$ .

聴覚特性を利用して雑音スペクトル減算時の係数調整を行う方法 [14] や、音声領域と雑音領域との判別を行う方法 [16, 17] が提案されている。

## 1.4 研究の目的および本論文の構成

デジタル音声信号処理技術の発展に伴い、雑音除去方式の需要が高まり、第 1.1 節に示したように様々な雑音除去方式が提案されてきた。本論文では、数ある雑音除去方式のうち複数マイクロホンを用いる方式である適応ノイズキャンセラおよび、単一のマイクロホンで実現できるスペクトルサブトラクションを対象として、両方法のさらなる音質改善を試みたものである。



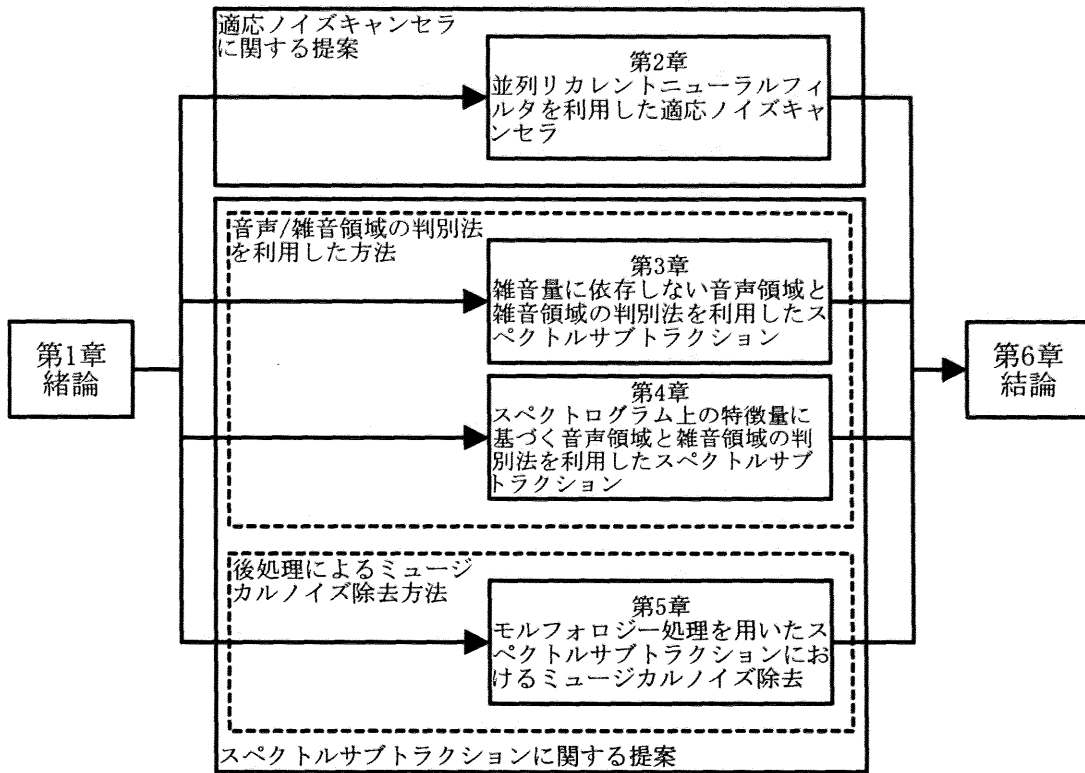


図 1.6 本論文の構成

Fig. 1.6 Configuration of this dissertation.

図 1.6 に本論文の構成を示す。本論文は、全 6 章により構成されており、各章の概要を以下に述べる。

第 1 章では、研究の背景と目的、および本論文の概要について述べている。

第 2 章では、非線形特性のパスを持ったシステムに対応できる適応フィルタとして並列リカレントニューラルネットフィルタ (parallel recurrent neural filter, PRNF) を利用した適応ノイズキャンセラを提案する。PRNF はリカレントニューラルネットワークを用いたリカレントニューラルフィルタを多重分割して並列化したものであり、これによりフィルタの計算量の削減を図っている。さらに、PRNF の学習に学習係数を動的に変化させる方法を使用して学習を安定させ、かつ収束を速めることにより、全体として計算回数の削減を図る。計算機シミュレーションの結果より、提案方法がパスの特性が線形・非線形にかかわらず十分に雑音が除去できることを示す。また、提案方法と線形適応フィルタの LMS 法、非線形システムに対応した適応フィルタであるニューラルフィルタおよび適応 Volterra フィルタとの比較を行い、

提案方法が非線形特性を持ったパスの雑音除去に有効な方法の1つであることを示す。

第3章では、雑音量に依存しない音声領域と雑音領域の判別方法を利用したスペクトルサブトラクションを提案する。提案方法では、観測信号が音声と雑音の和であることを仮定し、判別のしきい値を雑音量によって適応的に変化させることにより、判別時の雑音の影響を低減させる。このことにより、提案方法は音声領域と雑音領域の判別が雑音量に依存せず一定の性能を保って行うことが可能となるため、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少を図ることができる。性能評価の結果より、提案方法は従来方法より音声領域と雑音領域の判別が雑音量に依存せず正確に行われ、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少できることを示す。さらに、提案方法によって処理した音声のひずみは雑音量に関係なくほぼ一定であることを示す。

第4章では、雑音の事前情報を用いない音声領域と雑音領域との判別方法として、複数の短時間フーリエ変換の周波数から構成されるバンドにおける観測信号のスペクトログラム上の特徴量に着目し、各バンド内の標準偏差を利用した音声領域と雑音領域との判別方法を提案する。バンド内の成分が音声と雑音で構成される場合は、音声成分と雑音成分との周波数成分での特徴の違いから観測信号のスペクトルの標準偏差は高くなる。一方、バンド内の成分が雑音のみであれば、雑音成分の周波数成分での特徴は音声成分のものに比べて一様であるため、観測信号のスペクトルの標準偏差は低くなる。したがって、バンド毎の観測信号のスペクトルの標準偏差に適切なしきい値を設定することにより、音声領域と雑音領域との判別が可能となる。性能評価の結果より、提案方法が従来方法より処理した音声のミュージカルノイズや音声ひずみを減少できることを示す。そして、提案方法が雑音の事前情報を用いない1入力システムの雑音除去方法として有効な方法であることを示す。

第5章では、スペクトルサブトラクションの問題点である、処理後の音声信号に発生するミュージカルノイズに対して、モルフォロジー処理を用いたミュージカルノイズの除去方法を提案する。提案方法では、ミュージカルノイズがスペクトログラム上で孤立点として現れることに注目し、モルフォロジー処理の1つであるオープニングが孤立点除去に向いていることを利用してミュージカルノイズの除去を行う。また、提案方法ではミュージカルノイズ検出処理を必要とせず、かつモルフォ

ロジック処理は比較演算のみで行える。そのため、提案方法は従来方法と比較してシステム的设计が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。性能評価の結果より、提案方法は従来方法に比べて少ない計算回数で雑音除去を行うことができ、かつミュージカルノイズの除去性能が優れていることを示す。

最後に、第6章は結論であり、本論文の内容を総括している。

## 第1章の参考文献

- [1] 古井貞熙, 音響工学, 近代科学社, 東京, 1992.
- [2] 中川聖一, “音声認識研究の動向,” 電子情報通信学会論文誌 D-II, vol.J-83D-II, no.2, pp.433-457, Feb. 2000.
- [3] 鈴木誠史, “S/N の低い音声信号から雑音を減らす最近の信号処理技術,” 日経エレクトロニクス, no.281, pp.128-154, Jan. 1982.
- [4] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr. and R.C. Goodlin, “Adaptive noise cancelling : Principles and application,” Proc. IEEE, vol.63, no.12, pp.1692-1716, Dec. 1975.
- [5] 久保田一, 古川利博, 板倉秀清, “前処理を含むノイズキャンセラのアルゴリズムとその性能評価,” 電子情報通信学会論文誌 A, vol.J69-A, no.5, pp.584-591, May 1986.
- [6] J.E. Greenberg, “Modified LMS algorithms for speech processing with an adaptive noise canceller,” IEEE Trans. Speech and Audio Process., vol.6, no.4, pp.338-351, July 1998.
- [7] J.P. Costa, L.Pronzato and E. Thierry, “Nonlinear prediction by kriging with application to noise cancellation,” Signal Process., vol.80, no.4, pp.553-566, Apr. 2000.
- [8] J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, “Computer-steered microphone arrays for sound transduction in large rooms,” J. Acoust. Soc. Am., vol.78, no.5, pp.1508-1518, Nov. 1985.
- [9] 金田豊, “アダプティブマイクロホンアレー,” 電子情報通信学会論文誌 B-II, vol.J-75B-II, no.11, pp.742-748, Nov. 1992.
- [10] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Process., vol.ASSP-27, no.2, pp.113-120, Apr. 1979.

- 
- [11] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," *Speech Commun.*, vol.11, no.2-3, pp.215-228, June 1992.
- [12] R. Martin, "Spectral subtraction based on minimum statistics," *Proc. of EUSIPCO'94*, pp.1182-1185, Sep. 1994.
- [13] Z. Goh, K.C. Tan and T.G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," *IEEE Trans. Speech and Audio Process.*, vol.6, no.3, pp.287-292, May 1998.
- [14] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech and Audio Process.*, vol.7, no.2, pp.126-137, Mar. 1999.
- [15] V. Stahl, A. Fischer, and R. Bippusuchi, "Quantile based noise estimation for spectral subtraction and Wiener filtering," *Proc. of IEEE ICASSP 2000*, pp.1875-1878, Istanbul, Turkey, June 2000.
- [16] S. Yoon and C.D. Yoo: "Speech enhancement based on speech/noise-dominant decision," *IEICE Trans. Inf. & Syst.*, vol.E85-D, no.4, pp.744-750, Apr. 2002.
- [17] H. Nakashima, Y. Chisaki and T. Usagawa, "Spectral subtraction based on statistical criteria of the spectral distribution," *IEICE Trans. Fundamentals*, vol.E85-A, no.10, pp.2283-2292, Oct. 2002.
- [18] R. Frazier, S. Samsam, L. Braida, and A. Oppenheim, "Enhancement of speech by adaptive filtering," *Proc. of IEEE ICASSP'76*, vol.1, pp.251-253, Apr. 1976.
- [19] J. Lim, A. Oppenheim, and L. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-26, no.4, pp.354-358, Aug. 1978.
- [20] D. Malah and R. Cox, "A generalized comb filtering technique for speech enhancement," *Proc. of IEEE ICASSP'82*, vol.7, pp.160-163, May. 1982.

- 
- [21] J. Suzuki, "Speech processing by splicing of autocorrelation function," Proc. of IEEE ICASSP'76, vol.1, pp.713-716, Apr. 1976.
- [22] 高杉敏男, 鈴木誠史, 田中良二, "SPAC (自己相関関数を利用した音声処理方式) の機能と基本特性," 電子通信学会論文誌 A, vol.J62-A, no.3, pp.175-182, Mar. 1979.
- [23] 吉谷清澄, 鈴木誠史, "自己相関関数を利用した音声処理方式 (SPAC) の SN 比改善特性," 電子通信学会論文誌 A, vol.J61-A, no.3, pp.217-223, Mar. 1978.
- [24] 高杉敏男, 鈴木誠史, "SPAC (自己相関関数を利用した音声処理方式) の雑音低減効果," 日本音響学会誌, vol.35, no.7, pp.361-369, Sep. 1979.
- [25] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," Proc of IEEE ICASSP'87, vol.12, pp.177-180, Apr. 1987.
- [26] J.D. Gibson, B. Koo and S.D. Gray, "Filtering of colored noise for speech enhancement and coding," IEEE Trans. Signal Process., vol.39, no.8, pp.1732-1742, Aug. 1991.
- [27] S. Gannot, D. Burshtein and E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms," IEEE Trans. Speech Audio Process., vol.6, no.4, pp.373-385, Apr. 1998.
- [28] Z. Goh, K. Tan. and B.T.G. Tan, "Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model," IEEE Trans. Speech Audio Process., vol.7, no.5, pp.510-524, May 1999.
- [29] J. Lim and A. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-26, no.3, pp.197-210, Aug. 1978.
- [30] S.V. Vaseghi, Advanced digital signal processing and noise reduction, 2nd edition, Wiley, New York, 2000.
- [31] 谷萩隆嗣, デジタル信号処理の理論 3, コロナ社, 東京, 1986.

- [32] 飯國洋二, 適応信号処理アルゴリズム, 培風館, 東京, 2000.
- [33] S.W. Piché, "Steepest descent algorithms for neural network controllers and filters," IEEE Trans. Neural Networks, vol.5, no.2, pp.198-212, Mar. 1994.
- [34] S.Y. Fakhouri, "Identification of the Volterra kernels of nonlinear systems," IEE Proc., vol.127D, no. 6, pp.296-304, Nov. 1980.
- [35] 梶川嘉延, "適応 Volterra フィルタの現状と展望," 電子情報通信学会論文誌 A, vol.J82-A, no.6, pp.759-768, June 1999.
- [36] 曹建庭, 谷萩隆嗣, 呂建明, "リカレントニューラルネットワークを用いた並列非線形適応デジタルフィルタ," 電子情報通信学会論文誌 A, vol.J79-A, no.4, pp.868-877, Apr. 1996.
- [37] 柳坂和秀, 関文隆, 梶川嘉延, 野村康雄, "入力信号のパワー変動を考慮したニューラルフィルタ," 電子情報通信学会論文誌 A, vol.J83-A, no.3, pp.253-262, Mar. 2000.
- [38] 谷萩隆嗣, ニューラルネットワークとファジィ信号処理, コロナ社, 東京, 1998.
- [39] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink and D.L. Alkon, "Accelerating the convergence of the back-propagation method," Biological Cybernetics, vol.59, no.3, pp.257-263, 1988.
- [40] 賈棋, 戸田尚宏, 臼井支朗, "ニューラルネットにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察," 電子情報通信学会論文誌 D-II, vol.J73-D-II, no.8, pp.1179-1185, Aug. 1990.
- [41] 高木英行, 坂上茂生, 戸川隼人, "ニューラルネット学習における非線形最適化手法の効果," 電子情報通信学会論文誌 D-II, vol.J74-D-II, no.4, pp.528-535, Apr. 1991.
- [42] 田中哲夫, 古村光夫, "逆誤差伝搬量の特異点解消による学習の高速化," 電子情報通信学会論文誌 D-II, vol.J75-D-II, no.5, pp.1000-1008, May 1992.

- [43] 折川典生, 原田豊, “伝達関数の改良による誤差伝搬学習,” 電子情報通信学会論文誌 D-II, vol.J83-D-II, no.2, pp.852-854, Feb. 2000.



---

## 第2章

---

# 並列リカレントニューラルフィルタを 利用した適応ノイズキャンセラ

### ●● 本章概要 ●●

本章では、非線形特性のパスを持ったシステムに対応できる適応フィルタとして並列リカレントニューラルネットフィルタ (PRNF) を利用した適応ノイズキャンセラを提案した。PRNF はリカレントニューラルネットワークを用いたリカレントニューラルフィルタを多重分割して並列化したもので、これによりフィルタの計算量の削減を図っている。さらに、PRNF の学習に学習係数を動的に変化させる方法を使用して学習を安定させ、かつ収束を速めることにより、全体として計算回数の削減を図る。計算機シミュレーションの結果より、提案方法がパスの特性が線形・非線形にかかわらず十分に雑音が除去できることを示した。また、提案方法と線形適応フィルタの LMS 法、非線形システムに対応した適応フィルタであるニューラルフィルタおよび適応 Volterra フィルタとの比較を行い、提案方法が非線形特性を持ったパスの雑音除去に有効な方法の1つであることを示した。

## 2.1 はじめに

複数マイクロホンによる雑音除去方式の1つとして、適応ノイズキャンセラ [1-4] がある。適応ノイズキャンセラはシステム同定の手法で雑音を推定し、観測信号から減算して目的の信号を取り出すための適応フィルタである。雑音源から観測信号までのパスを線形システムと仮定した場合の適応ノイズキャンセラの方法は既によく知られている [1-3]。しかし、状況によってはパスは線形特性だけでなく非線形特性を考慮する必要がある [4,5]、線形のパスを仮定した適応ノイズキャンセラではその性能に限界がある。したがって、非線形特性を持ったパスに対応できる適応ノイズキャンセラの設計は重要な研究課題である。

非線形特性に対応した適応フィルタとして適応 Volterra フィルタ (AVF) [6, 7]、ニューラルフィルタ (NF) [8,9] が挙げられる。AVF は Volterra 級数展開に基づいて構成されており、精度の高さが利点であるが、高次の非線形特性をフィルタリングする場合に構造が複雑になり、計算回数が膨大になる問題点がある。一方、NF は非線形の入出力関係を持つニューラルネットワークの特徴である、有限個のパターンから学習した内容を生かして未知の信号に対応するという汎化能力を適応フィルタに応用したもので、精度は AVF より多少劣るが高次非線形性のフィルタリングを比較的小規模な構成で行うことができる。

ニューラルネットワークの中でも静的なパターンの学習だけでなく、動的なパターンである時系列を扱えるものとしてリカレントニューラルネットワーク (recurrent neural network, RNN) が提案されている [10]。RNN には出力層から入力層へのフィードバックループをもたせる方法 (Jordan のネットワーク) [11]、中間層から入力層へのフィードバックループをもたせる方法 (Elman のネットワーク) [12] などがある。一方、ニューラルネットワークは学習に時間がかかるという問題点を持っている。これはニューラルネットワークの学習で用いられているバックプロパゲーション法の持つ問題点といえる。ニューラルネットワークを実際のシステムに適用することを考えた場合、学習時間を減らすことが必要となる。この問題を解消する方法として、バックプロパゲーション法を改良して学習の収束を速める方法が提案されている [13-17]。以上より、優れた学習能力を持っており、時系列を扱える RNN を非線形特性に対応する適応フィルタとして適用することは有効であると考えられるが、同時に計算時間削減の工夫が必要である。

本章では、非線形特性のパスを持ったシステムに対応できる適応フィルタとして並列リカレントニューラルフィルタ (parallel recurrent neural filter, PRNF) を利用した適応ノイズキャンセラを提案する。PRNF は RNN を用いたリカレントニューラルフィルタ (recurrent neural filter, RNF) を多重分割して並列化したもので、これにより RNF の計算を高速化させることが可能である。さらに、文献 [13] で提案されている学習係数および慣性係数を動的に変化させる方法を適用して、学習を安定させ収束を速めることにより、全体として計算回数の削減を図る。計算機シミュレーションを行い、提案方法がパスの特性が線形・非線形にかかわらず十分に雑音が除去できることを示す。また、提案方法と線形適応フィルタの LMS 法、非線形システムに対応した適応フィルタである NF および AVF との比較を行い、提案方法が非線形特性を持ったパスの雑音除去に有効な方法の 1 つであることを示す。

## 2.2 PRNF を利用した適応ノイズキャンセラ

### 2.2.1 提案方法のモデル

図 2.1 に本章で提案する雑音キャンセラの構成を示す。雑音キャンセラは主入力端子と少なくとも 1 個の参照入力端子を持つ。主入力端子の信号は目的の信号  $s(k)$  と雑音源  $x(k)$  がパス  $f(\cdot)$  を介して得られる雑音  $d(k)$  の和からなる観測信号である。一方、参照入力端子の信号には雑音源からの信号  $x(k)$  が入力される。PRNF はパス  $f(\cdot)$  を推定するための適応フィルタである。PRNF は動的なパターンである時系列を扱える RNN を使用したニューラルフィルタであり、非線形の入出力関係を持ち優れた学習能力を持っているので、非線形パスに対応することができる。PRNF の詳細については次節で述べる。PRNF の出力を  $y(k)$  とすると、システムの出カ  $z(k)$  は

$$z(k) = s(k) + d(k) - y(k) \quad (2.1)$$

となる。ここでは、パスを推定するための PRNF の学習は、 $s(k)$  が入力される前に前処理として学習する。つまり、 $s(k) = 0$  として学習を行う。

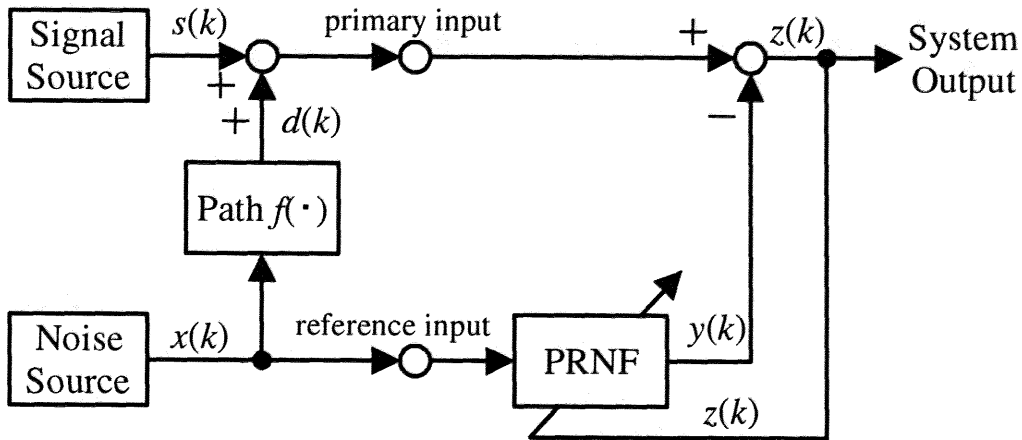


図 2.1 PRNF を利用した雑音キャンセラの基本構成

Fig. 2.1 Configuration of the noise canceller using PRNF

### 2.2.2 PRNF のモデル

図 2.2 に PRNF のモデルを示す. PRNF は RNF を多重分割して並列接続したものである. ただし, 図 2.2 の  $M$  は分割数,  $n$  は入力層および中間層のニューロン数である. また, 提案方法における RNF を分割する際の方針は, 分割された RNF $m$  ( $m = 1, 2, \dots, M$ ) は同一構造の RNF であり,  $n/M$  は整数であるとする. もし,  $n/M$  が整数でない場合には  $n$  を増加させて  $n$  が  $M$  の整数倍となるようにしておくものとする. なお,  $M = 1$  のとき, PRNF は RNF と一致する. PRNF ( $M \geq 2$ ) は RNF と比較して各ニューロン間の結合数の減少により計算回数が減少し高速な処理が可能となる. RNF に用いる RNN は図 2.2(b) のような構造を持つ中間層を自己フィードバックしたものを用いる. この RNN は本章で提案するシステムにおいて予備実験により調べた結果, Elman ネットワークとほぼ同程度の能力を持ち, かつフィードバックが中間層の自己フィードバックのみなので Elman ネットワークより計算回数が減少される. なお, RNF $m$  の  $j$  番目の中間層への入力  $h_j^{(m)}(k)$  は

$$h_j^{(m)}(k) = \sum_{i=\frac{n(m-1)}{M}}^{\frac{mn}{M}-1} W_{ij}^{(m)}(k)x(k-i) + \tau f_{mid} \left( h_j^{(m)}(k-1) \right) \quad (2.2)$$

となる。ここで、 $\tau$ は忘却係数である。ただし、 $k = 0$ の場合、第2項は0とする。また、PRNFの出力 $y(k)$ は次式で与えられる。

$$y(k) = \sum_{m=1}^M f_{out}(o^{(m)}(k)) \quad (2.3)$$

ただし、 $o^{(m)}(k)$ はRNF $m$ の出力層への入力である。入力層への入力は $x(k)$ を単位時間素子 $[D]$ を通過させたものとし、 $n + 1$ 時刻前までの履歴を使用している。自己フィードバックの機能のみに頼らず過去の履歴を直接与えることにより、時間的に近いデータに対する学習が明確になり、学習能力が向上される [19]。中間層の入出力関数は出力が $[-1, 1]$ の範囲内で単調非減少のシグモイド関数

$$f_{mid}(x) = \frac{1 - \exp(-\gamma_1 x)}{1 + \exp(-\gamma_1 x)} \quad (2.4)$$

を用いる。出力層の入出力関数は線形関数

$$f_{out}(x) = \gamma_2 x \quad (2.5)$$

を用いる。ただし、 $\gamma_1 > 0, \gamma_2 > 0$ はそれぞれ入出力関数の傾きである。

RNFの学習にはバックプロパゲーション法を用いる。このとき、結合係数の修正量は

$$\Delta W^{(m)}(k+1) = -\eta \frac{\partial J(k)}{\partial W^{(m)}(k)} + \alpha \Delta W^{(m)}(k) \quad (2.6)$$

$$J(k) = \frac{1}{2} \{T(k) - y(k)\}^2 \quad (2.7)$$

となる。ただし、 $T(k)$ は教師信号、 $\eta > 0$ は学習係数、 $\alpha$ は慣性係数であり、 $0 \leq \alpha < 1$ とする。

さらに、PRNFの学習を安定でかつ収束を速くするために文献 [13] で提案されている学習係数および慣性係数を動的に変化させる方法を適用する。文献 [13] では一般的なニューラルネットワークを使ったパターン認識問題に対して利用しており、そのままでは本章で提案するシステムで扱う音声信号（時系列データ）に利用できない。そのため、次のようにして音声信号で適用できるようにする。

まず、PRNFの学習回数 $L$ 回ごとに次式を用いて教師信号 $T(k)$ とPRNFの出力との平均2乗誤差(MSE)を計算する。

$$MSE(i) = \frac{1}{L} \sum_{k=L(i+1)}^{L(i+1)} [T(k) - y(k)]^2 \quad (2.8)$$

なお本論文では、ある時刻  $k$  において学習を1回だけ行い、これを1回の学習とする。すなわち、時刻  $k$  から  $k+t$  まで学習を行った場合、 $t+1$  回学習したことになる。次に、 $MSE(i)$  と  $MSE(i-1)$  を比較し、以下の条件で学習係数  $\eta(i)$  および慣性係数  $\alpha(i)$  を更新する。

$$\begin{cases} \eta(i) = \phi\eta(i-1) & \text{if } MSE(i) \leq MSE(i-1) \\ \eta(i) = \beta\eta(i-1) & \text{if } MSE(i) > MSE(i-1) \end{cases} \quad (2.9)$$

$$\begin{cases} \alpha(i) = A & \text{if } MSE(i) \leq MSE(i-1) \\ \alpha(i) = 0 & \text{if } MSE(i) > MSE(i-1) \end{cases} \quad (2.10)$$

ただし、 $\phi, \beta, A$  はそれぞれ  $\phi > 1, 0 < \beta < 1, 0 < A < 1$  とする。

式(2.9)のように学習係数を変化させることにより、学習のはじめのうちは  $\eta(i)$  が大きくなるため学習が促進され、学習がある程度進むと  $\eta(i)$  が小さくなり、安定した学習が行えるので、より速い学習の収束と安定した学習が期待できる。また、式(2.10)のように慣性係数を変化させることで、学習が進んでいるときはモーメント項  $\alpha(i)\Delta W(k)$  からの修正量により学習が促進される。一方、パスが変化するなどして学習が進まない場合は  $\alpha(i) = 0$  となり、モーメント項からの影響を受けない。

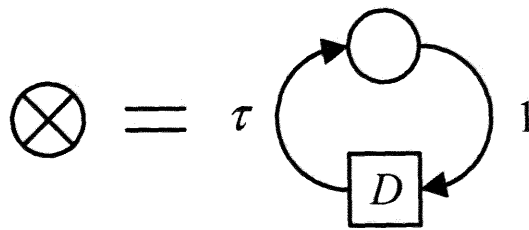
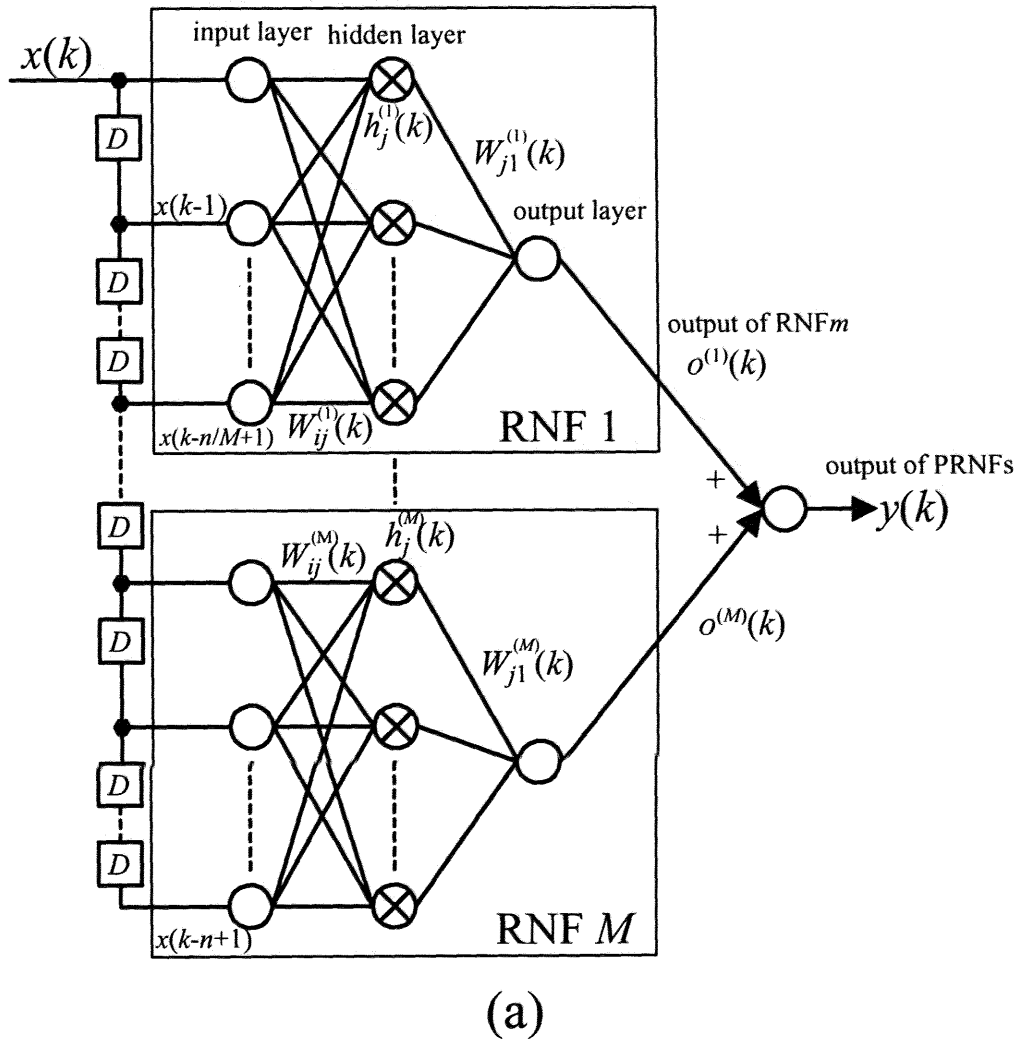


図 2.2 PRNF の構成

Fig. 2.2 Configuration of the PRNF

## 2.3 シミュレーション

### 2.3.1 RNFによるシミュレーション結果

まず、本章で使用するRNFについて計算機シミュレーションを行った。RNFの入力層および中間層のニューロン数は  $n = 2, 4, 8, 16, 32$  とし、出力層のニューロン数は1とする。RNFの学習は目的の信号  $s(k)$  が  $s(k) = 0$  の状態、すなわちオフラインの状態で行うので教師信号  $T(k)$  は  $T(k) = d(k)$  となる。目的の信号  $s(k)$  は

$$s(k) = 0.17 \sin(0.05k) - 0.65 \cos(0.04k - 30) + 0.34 \cos(0.03k - 5) - 0.3 \sin(0.02k + 25) + 0.21 \cos(0.01k + 45) \quad (2.11)$$

とする。また、雑音  $x(k)$  は白色雑音とする。雑音源からのパス  $f(\cdot)$  については線形特性と非線形特性の両方について検討を行う。線形特性のパスは

$$d(k) = 0.02x(k-1) - 0.21x(k-2) - 0.09x(k-3) + x(k-4) + 0.43x(k-5) + 0.16x(k-6) + 0.07x(k-7) \quad (2.12)$$

とし、非線形特性のパスは次式で定義される指数関数モデルとする。

$$d(k) = x(k) + 0.5 \exp \left[ \frac{-\{x(k-1) - 1\}^2}{0.67} \right] - 0.5 \exp \left[ \frac{-\{x(k-1) + 1\}^2}{0.67} \right] \quad (2.13)$$

また、RNFの各パラメータは予備実験を結果より、 $\gamma_1 = 0.5$ 、 $\gamma_2 = 1.0$ 、 $\tau = 0.25$ 、 $\eta(0) = 0.2$ 、 $L = 100$ 、 $\phi = 1.2$ 、 $\beta = 0.75$ 、 $A = 0.5$  とした。

図2.3にパスが式(2.13)の指数関数モデルの場合のRNFの学習曲線を示す。なお、RNFのニューロン数は入力層8、中間層8、出力層1の場合である。PRNFの学習精度の指標としては、文献[9]で用いられているReductionを用いる。Reductionは次式で定義され、値が大きければ大きいほど精度よく学習されていることを意味する。

$$\text{Reduction} = 10 \log_{10} \left[ \frac{\sum y(k)^2}{\sum \{T(k) - y(k)\}^2} \right] \quad (2.14)$$

図2.3より、RNFは学習によってReductionで約25 dBの精度を得ていることがわかる。よって、学習の段階では十分同定できているといえる。

図2.4にパスを線形特性とした場合のシミュレーション結果、図2.5に非線形特性である指数関数モデルとした場合のシミュレーション結果を示す。これらの図は縦



軸が出力値あるいは入力値であり、横軸は時刻  $k$  である。また、各図の (a) は観測信号  $s(k) + d(k)$ , (b) は目的の信号  $s(k)$  とシステム出力  $z(k)$  を重ねて表示したものである。なお、入力信号の SNR は  $SNR_{in} = 10$  dB とした。入力信号および出力信号の SNR はそれぞれ次式で定義する。

$$SNR_{in} = 10 \log_{10} \left[ \frac{\sigma_s^2}{\sigma_x^2} \right] \quad (2.15)$$

$$SNR_{out} = 10 \log_{10} \left[ \frac{\sigma_s^2}{\sigma_e^2} \right] \quad (2.16)$$

ただし、 $\sigma_s^2$  は  $s(k)$  の分散、 $\sigma_x^2$  は  $x(k)$  の分散、 $\sigma_e^2$  は  $z(k) - s(k)$  の分散である。

図 2.4 および図 2.5 より、RNF はパスの特性が線形・非線形にかかわらず十分に雑音が除去できていることがわかる。

図 2.6 に入力層および中間層のニューロン数を変化させた場合の入出力の SNR の関係を示す。図 2.6 より、RNF のニューロン数を多くするほど雑音除去性能が良くなっていることがわかる。しかし、ニューロン数が多くなりすぎると性能の向上の割合が小さくなり、線形モデルでは  $n = 16$  以上、指数関数モデルでは  $n = 8$  以上の場合ではほぼ同じ性能になっている。これらのことより、ニューロン数は多い方が良いが、計算回数および性能の向上の割合を考えると適切なニューロン数を決定することが望ましい。また、最適なニューロン数はパスに依存している。以上の結果を踏まえて、今後のシミュレーションでは RNF のニューロン数をそれぞれ入力層 8, 中間層 8, 出力層 1 とする。

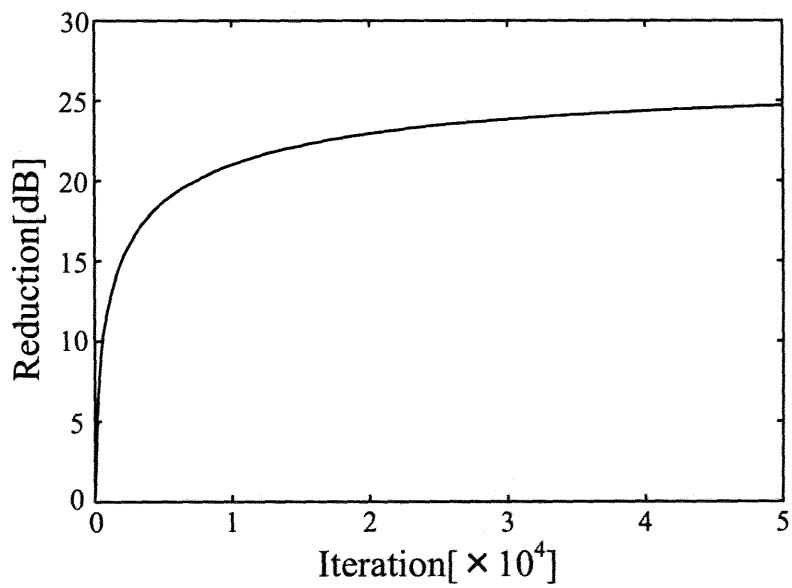


図 2.3 パスが指数関数モデルの場合の RNF の学習

Fig. 2.3 Learning curve of RNF in the exponential model.

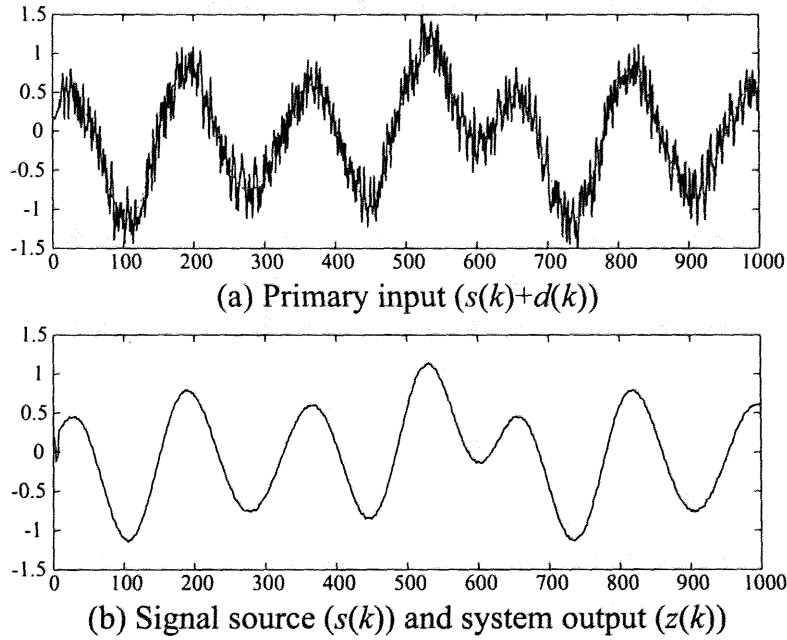


図 2.4 パスが線形特性の場合の RNF の出力波形

Fig. 2.4 Signal waveforms using RNF in the linear path for  $n = 8$  and  $SNR_{in} = 10$  dB.

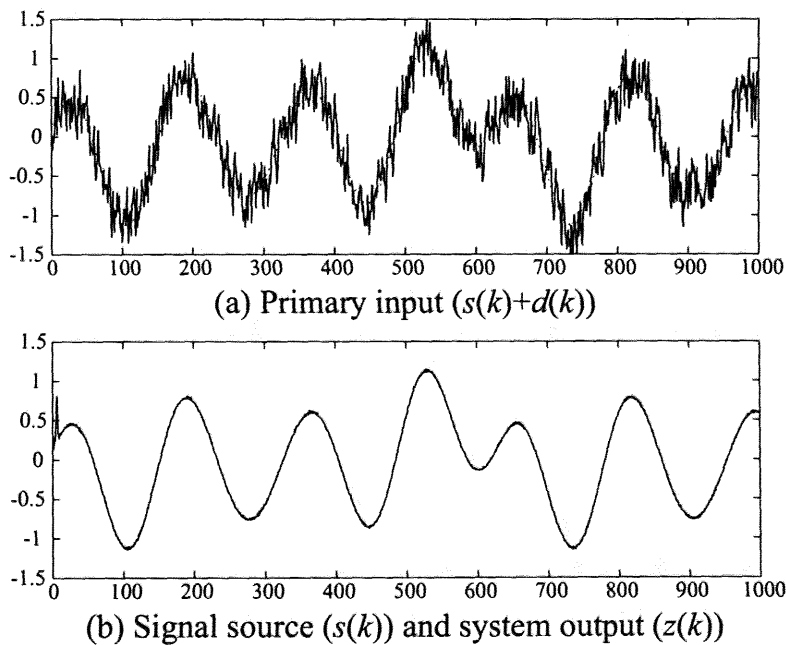


図 2.5 パスが指数関数モデルの場合の RNF の出力波形

Fig. 2.5 Signal waveforms using RNF in the exponential model for  $n = 8$  and  $SNR_{in} = 10$  dB.

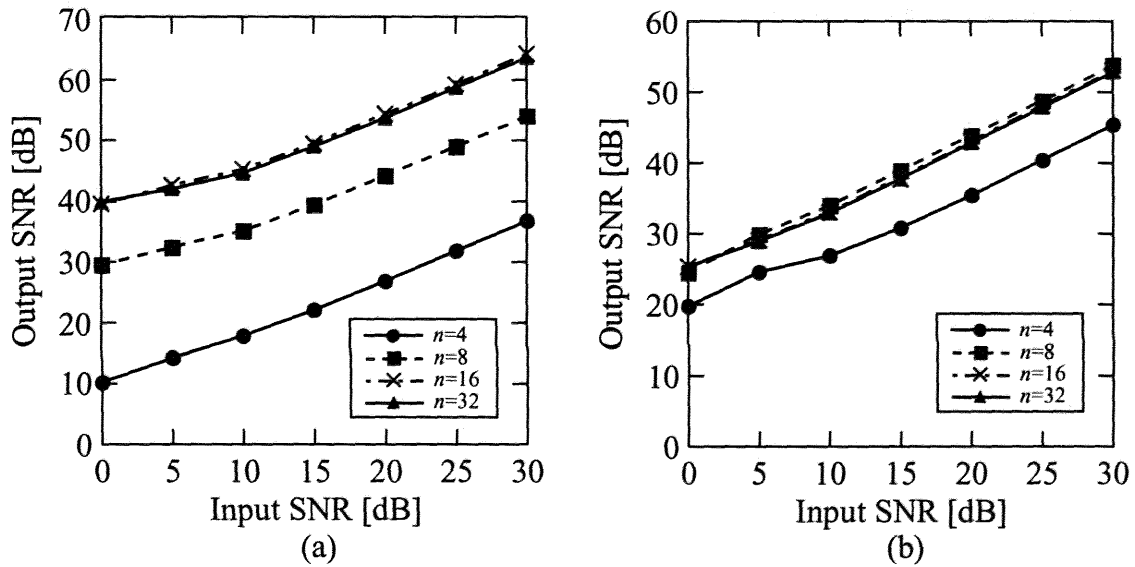


図 2.6 RNF の入出力 SNR の関係 ((a) : 線形特性, (b) : 指数関数モデル)

Fig. 2.6 Input-output relation of SNR using RNF in (a) the linear path, (b) the exponential model.

表 2.1 1回の学習にかかる計算回数  
Table 2.1 Number of calculations for each learning

Type	Addition	Multiplication	exp
RNF	248	347	8
PRNF( $M=2$ )	154	220	8
PRNF( $M=4$ )	104	158	8

### 2.3.2 PRNFの有効性

ここではRNFとPRNFの雑音除去性能について比較を行う。ここで使用するPRNFは図2.2のものであり、ニューロン数は前節の結果より入力層8、中間層8、出力層1とし、分割数は $M=2,4$ とする。シミュレーション条件は前節の条件と同一とし、パスは式(2.13)の指数関数モデルとする。図2.7に $M=2$ のPRNFを用いたときのシミュレーション結果を示す。ただし、入力SNRは10 dBとした。図2.7より、PRNFを用いても十分に雑音が除去され、良好な結果が得られることがわかる。また、図2.5および図2.7より、PRNFの出力結果はRNFの出力結果とほぼ同じである。また、図2.8にRNFおよびPRNFの入出力のSNRの関係を示す。図2.8より、 $M=2$ のPRNFの場合はRNFと比べて雑音除去の性能にほとんど差がないものの、 $M=4$ の場合は出力SNRがRNFより5 dB程度下がっている。

また、表2.1にRNFとPRNFの計算回数を示す。表2.1の値はシミュレーションで使用している入力層8、中間層8、出力層1のRNFを分割した場合の1回あたりの学習にかかる加算回数、乗算回数および指数関数の計算回数である。なお、本章ではある時刻 $k$ においてバックプロパゲーション法による学習を1回だけ行い、これを1回の学習とする。すなわち、時刻 $k$ から $k+t$ まで学習を行った場合、 $t+1$ 回学習したことになる。表2.1より、 $M=4$ のPRNFの計算回数はRNFと比べて加算回数で約41%、乗算回数で約45%となっていることがわかる。以上のことから、PRNFは計算回数低減の面では有効な方法であるが、ノイズ除去性能と計算回数(分割数)との間にトレードオフの関係があるので、適切な分割数を選ぶ必要がある。

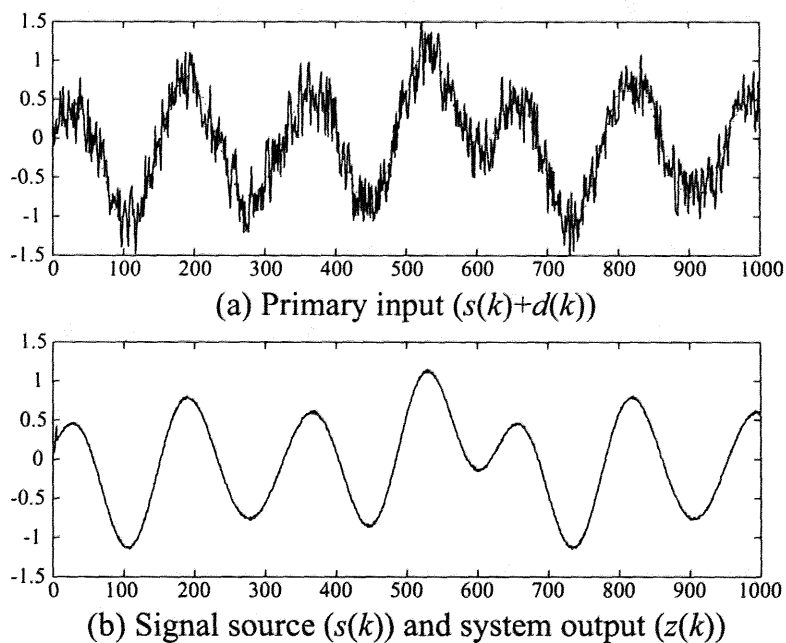


図 2.7 パスが指数関数モデルの場合の PRNF の出力波形

Fig. 2.7 Signal waveforms using PRNF in the exponential model for  $M = 2$  and  $SNR_{in} = 10$  dB.

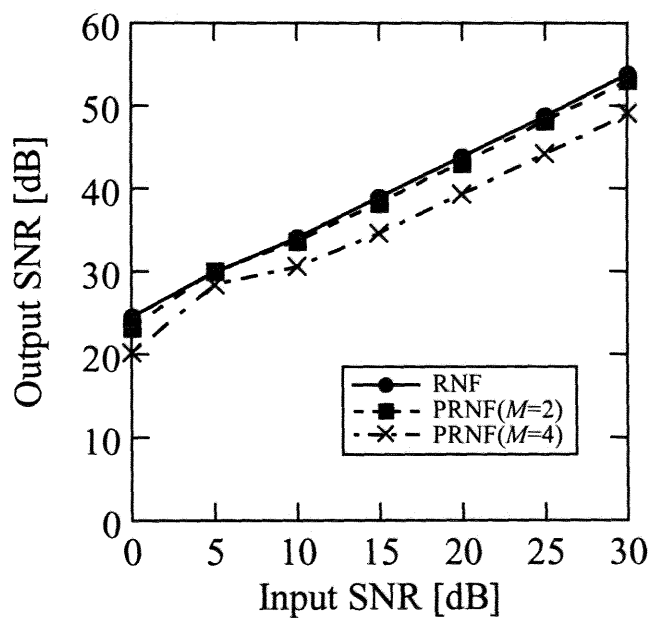


図 2.8 パスが指数関数モデルの場合の PRNF の入出力 SNR の関係

Fig. 2.8 Input-output relation of SNR using PRNF in the exponential model.

### 2.3.3 PRNF の学習方法の比較

ここでは本章で適用する PRNF の学習方法の有効性について検証する。PRNF の分割数は前節の結果より  $M = 2$  とし、その他のシミュレーション条件は前節と同じとする。図 2.9 に PRNF の学習曲線を示す。図 2.9 の曲線は学習係数が  $\eta = 0.05$  の場合、 $\eta = 0.2$  の場合、本章で適用する学習係数および慣性係数の動的変更を行った場合 ( $\eta(0) = 0.2, L = 100, \phi = 1.2, \beta = 0.75, A = 0.5$ ) の 3 種類である。なお、学習係数を固定している場合の慣性係数は  $\alpha = 0.5$  とする。まず、学習係数を  $\eta = 0.05$  で固定した場合は曲線の立ち上がりが緩やかであり、学習に時間がかかっている。次に、学習係数を  $\eta = 0.2$  で固定した場合は曲線の立ち上がりが急になっているが、学習回数が 2 万回を過ぎたあたりから Reduction が小さくなってしまふ。これは、学習が発散してしまったためである。また、パスの種類によっては十分に学習されないうちに発散してしまうことがあった。しかし、係数の動的変更を行った場合、曲線の立ち上がりが急でかつ収束すると値が安定している。これは、学習のはじめのうちは学習係数の値が大きいので学習が速く進み、学習がある程度進むと学習係数の値が小さくなって学習が安定するためである。さらに、動的変化を行った場合の方が精度高く学習できている。

表 2.2 にそれぞれの学習方法による収束回数を示す。表 2.2 の値は線形・非線形の各モデルについて収束回数を求め、平均したものである。なお、収束回数の判定条件は 100 回毎に Reduction を計算し、

$$\frac{\text{Reduction}(i) - \text{Reduction}(i - 1)}{\text{Reduction}(i)} < 0.001 \quad (2.17)$$

が成立した場合とする。表 2.2 より、動的変化を使用した場合は平均で 8500 回であり、8 kHz サンプリングのデータで考えると約 1 秒で収束できることがわかる。一方、 $\eta = 0.2$  で固定した方が動的変化を使用した場合と比べて収束にするまでの回数が 300 回少ない。これは学習係数を動的に変化させた場合、学習が進むと  $\eta$  が小さくなるためである。しかし、 $\eta = 0.2$  で固定した場合は収束する前に発散する場合があります。学習精度は動的変化を使用した場合と比べて低くなっている。以上のことより、本章で適用した学習方法が有効であると考えられる。

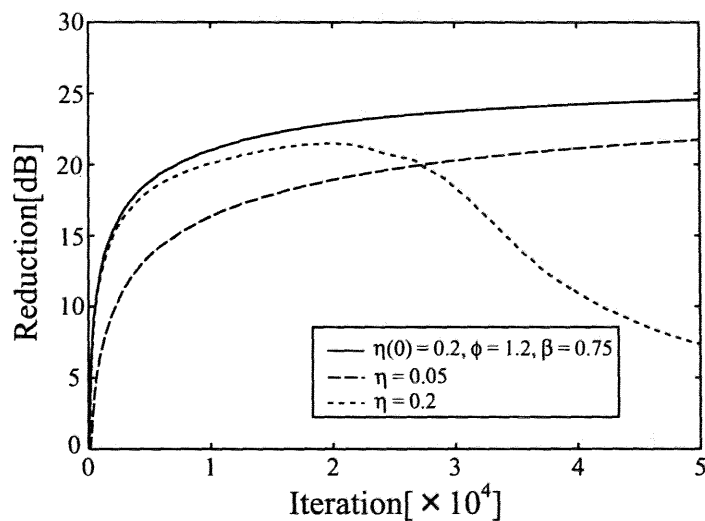


図 2.9 パスが指数関数モデルの場合の PRNF の学習曲線

Fig. 2.9 Learning curves of PRNF in the exponential model.

表 2.2 各学習方法による収束回数

Table 2.2 Number of iterations to converge

Type	Iteration
$\eta=0.05$	12000
$\eta=0.2$	8200
$\eta(0) = 0.2, \phi = 1.2, \beta = 0.75$	8500



### 2.3.4 実際の音声への適用

ここでは、目的の信号  $s(k)$  が音声信号の場合について検証する。目的信号の  $s(k)$  としては、女性が「森田村への誘致企業はこれで23社になりますが」と発声したものを用いる。なお、サンプリング周波数は8 kHz、量子化レベルは8 bit である。また、ここで用いるパス（非線形特性）は、次式で定義される NARMAX モデルを用いる [20].

$$\begin{aligned}
 d(k) = & \sum_{i=0}^{m_x-1} a(i)x(k-i) + \sum_{i=0}^{m_y-1} b(j)d(k-j) + \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_x-1} c(i,j)x(k-i)x(k-j) \\
 & + \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} u(i,j)x(k-i)d(k-j) + \sum_{i=0}^{m_y-1} \sum_{j=0}^{m_y-1} v(i,j)d(k-i)d(k-j) \\
 & + \dots
 \end{aligned} \tag{2.18}$$

このように、NARMAX モデルは出力のフィードバックが含まれている。シミュレーションで使用するパスは次式の通りである。

$$d(k) = x(k) + 0.5x(k-2)x(k-3)^2 - 0.2d(k-1)d(k-2)^2 \tag{2.19}$$

PRNF のシミュレーション条件は第2.3.3節のもの同一とした。図2.10にパスが NARMAX モデルの場合の (a) 目的の信号  $s(k)$ , (b) 観測信号  $s(k) + d(k)$ , (c) システム出力  $z(k)$  の波形をそれぞれ示す。なお、入力 SNR は 10 dB とした。図2.10より、提案方法は実際の音声の場合においても雑音を低減できることがわかる。また、処理結果は聴覚上、十分に良好な雑音低減を実現している。

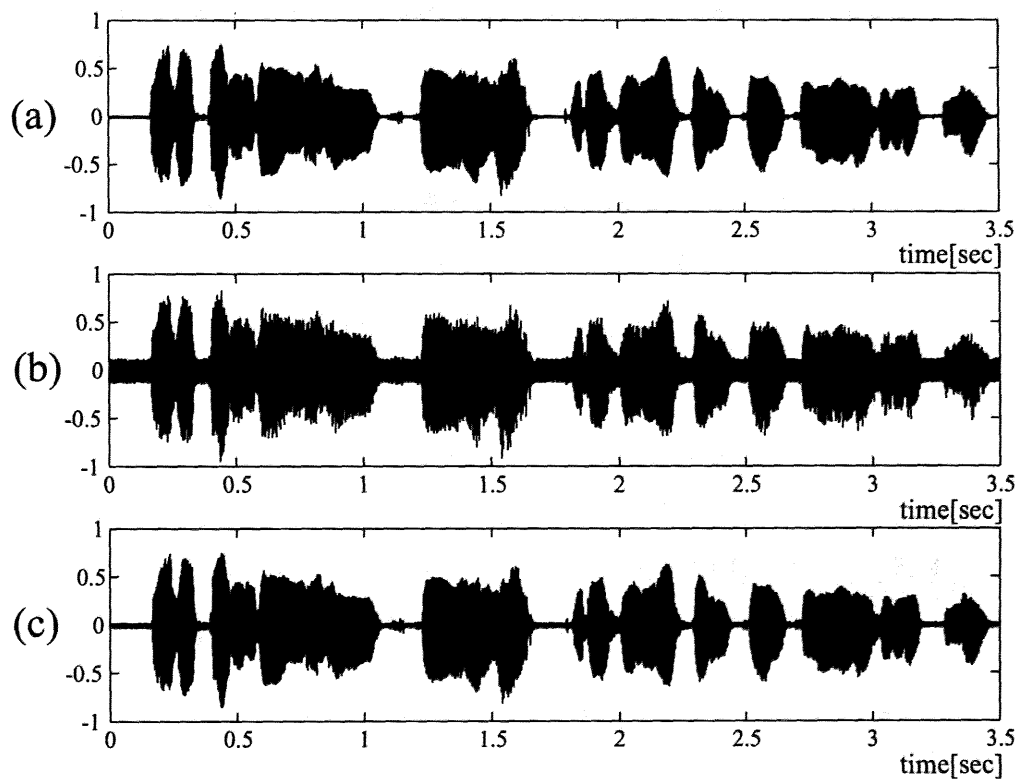


図 2.10 パスが NARMAX モデルの場合の PRNF の出力波形  
 ((a): $s(k)$ , (b): $s(k) + d(k)$ , (c): $z(k)$ )

Fig. 2.10 Signal waveforms using PRNF in NARMAX model for real voice((a): $s(k)$ , (b): $s(k) + d(k)$ , (c): $z(k)$ ).

### 2.3.5 他のフィルタとの比較

ここでは、本章で提案する方法の有効性を示すために他のフィルタとの比較検討を行う。比較するフィルタとしては、線形適応フィルタである LMS 法、非線形特性に対応した適応フィルタである AVF および RNF のフィードバックループを持たない NF を使用する。NF のニューロン数は入力層 8、中間層 8、出力層 1 とし、その他の条件は RNF と同一とする（ただし、 $\tau = 0$  である）。

LMS 法はフィルタ長  $N = 8$  とし、ステップサイズは  $\mu = 0.05$  とする。また、AVF はフィルタ長  $N_1 = N_2 = 8$  の 2 次 AVF とし、更新アルゴリズムとして NLMS アルゴリズムを使用する。なお、NLMS アルゴリズムのフィルタ係数の更新式は

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \frac{\mu}{\|\mathbf{x}(k)\|^2} \{d(k) - y(k)\} \mathbf{x}(k) \quad (2.20)$$

$$\|\mathbf{x}(k)\|^2 = \mathbf{x}^T(k) \mathbf{x}(k) \quad (2.21)$$

となる。ただし、 $\hat{\mathbf{h}}(k)$  は時刻  $k$  におけるフィルタ係数ベクトル、 $\mathbf{x}(k)$  は入力信号ベクトルであり、 $\mu = 0.1$  とする。図 2.11(a) にパスが式 (2.13) の指数関数モデルの場合の PRNF と各フィルタでの入出力の SNR の関係を、2.11(b) にパスが式 (2.19) の NARMAX モデルの場合の PRNF と各フィルタでの入出力の SNR の関係を示す。図 2.11 より、LMS 法は両方のモデルにおいて他の非線形フィルタと比較して雑音除去性能が劣っている。PRNF と NF と指数関数モデルの場合では雑音除去性能はほぼ同程度の性能であり、NARMAX モデルの場合は PRNF の方が若干性能が優れている。これは、PRNF が自己フィードバックを持っているためだと考えられる。また、PRNF と AVF とでは両方のモデルにおいて雑音除去性能はほぼ同程度である。

また、表 2.3 に PRNF、NF および AVF の 1 回の学習にかかる計算回数を示す。PRNF および NF の指数関数 ( $\exp(-x)$ ) の計算回数については Padé 近似 [21] を用いて次式より加算回数および乗算回数を算出した。

$$\exp(-x) = \frac{1 - \frac{1}{2}x + \frac{1}{2}\left(\frac{x}{2}\right)^2}{1 + \frac{1}{2}x + \frac{1}{2}\left(\frac{x}{2}\right)^2} \quad (2.22)$$

表 2.3 より、PRNF の計算回数は NF と比べて加算回数で約 68%、乗算回数で約 66% となっていることがわかる。また、AVF と比べて加算回数で約 86%、乗算回数で約 71% となっていることがわかる。図 2.12 に PRNF および AVF の収束特性を示

す。なお、図 2.12 の縦軸は式 (2.13) および式 (2.19) の非線形モデルについてそれぞれフィルタの初期値を変えて Reduction を 10 回計算し、平均した値である。図 2.12 より、PRNF と AVF の収束回数がほぼ同じであり、さらに収束時の学習精度もほぼ同程度であることがわかる。以上の結果より、PRNF と NF とでは PRNF の方がパスにフィードバックが含まれている場合の雑音除去性能が優れており、1 回の学習にかかる計算回数が少ないことがわかる。また、PRNF と AVF とでは雑音除去性能と収束回数に関してはほぼ同等であるが、1 回の学習にかかる計算回数に関しては PRNF の方が優れていることがわかる。従って、本章で提案する方法が非線形特性を有するパスの雑音除去を行うための有効な方法の 1 つであることがわかる。

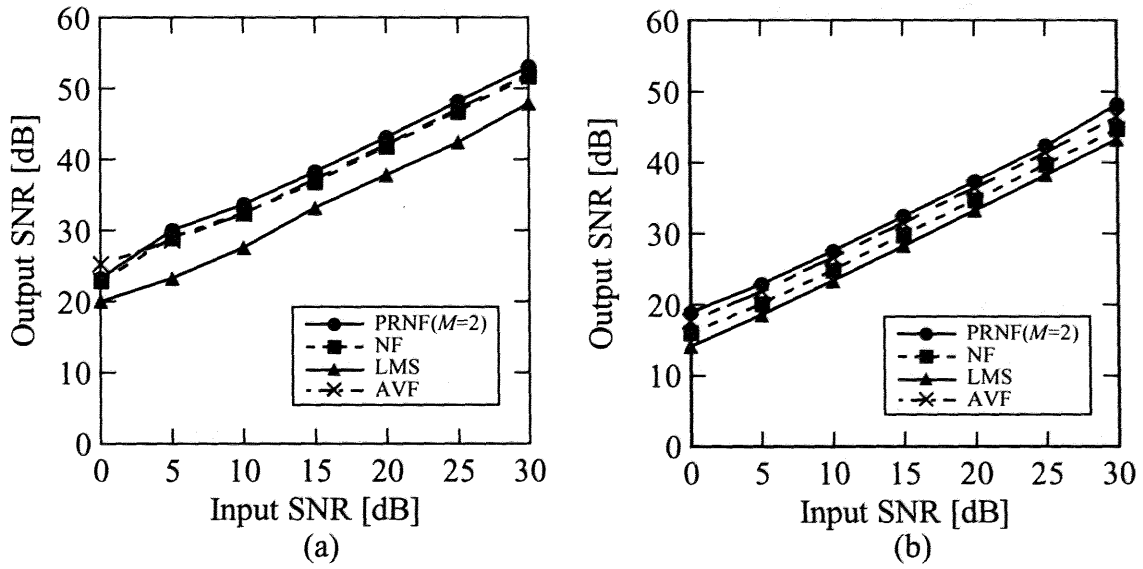


図 2.11 各フィルタの入出力 SNR の関係 ((a):指数関数モデル, (b): NARMAX モデル)

Fig. 2.11 Input-output relation of SNR using PRNF and other filters in (a) the exponential model, (b) the NARMAX model.

表 2.3 PRNF, NF および AVF の 1 回の学習にかかる計算回数

Table 2.3 Number of calculations for each learning in PRNF, NF and AVF.

Type	Addition	Multiplication
PRNF( $M=2$ )	186	252
NF	272	379
AVF	214	352

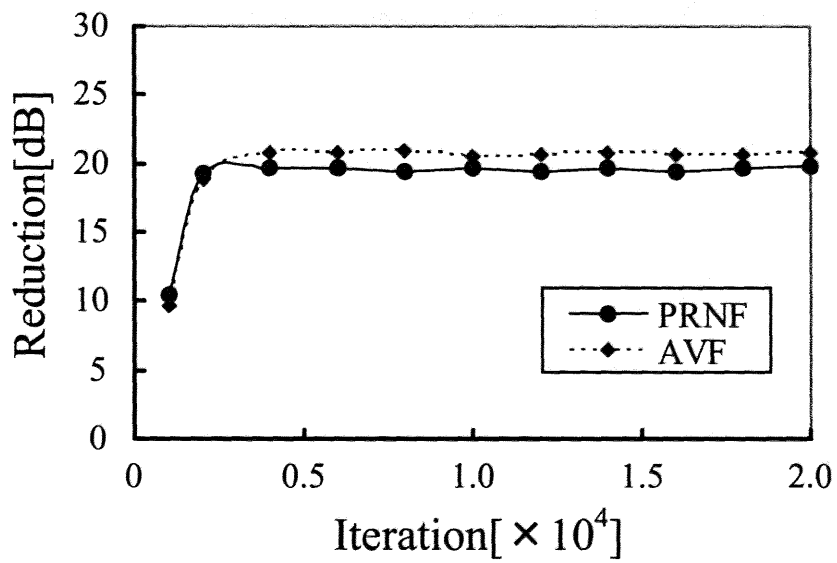


図 2.12 PRNF と AVF の収束特性

Fig. 2.12 Learning curves of PRNF and AVF.

## 2.4 まとめ

本章では、非線形特性のパスを持ったシステムに対応できる適応フィルタとして並列リカレントニューラルネットフィルタ (PRNF) を利用した適応ノイズキャンセラを提案した。PRNF はリカレントニューラルネットワークを用いたリカレントニューラルフィルタを多重分割して並列化したもので、これによりフィルタの計算量の削減を図った。さらに、PRNF の学習に学習係数を動的に変化させる方法を使用して学習を安定させ、かつ収束を速めることにより、全体として計算回数の削減を図った。計算機シミュレーションの結果より、提案方法がパスの特性が線形・非線形にかかわらず十分に雑音が除去できることを示した。また、提案方法と線形適応フィルタの LMS 法、非線形システムに対応した適応フィルタであるニューラルフィルタおよび適応 Volterra フィルタとの比較を行い、提案方法が非線形特性を持ったパスの雑音除去に有効な方法の1つであることを示した。

## 第2章の参考文献

- [1] B. Widrow, J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr. and R.C. Goodlin, "Adaptive noise cancelling : Principles and application," Proc. IEEE, vol.63, no.12, pp.1692-1716, Dec. 1975.
- [2] 久保田一, 古川利博, 板倉秀清, "前処理を含むノイズキャンセラのアルゴリズムとその性能評価," 電子情報通信学会論文誌 A, vol.J69-A, no.5, pp.584-591, May 1986.
- [3] J.E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," IEEE Trans. Speech and Audio Process., vol.6, no.4, pp.338-351, July 1998.
- [4] J.P. Costa, L.Pronzato and E. Thierry, "Nonlinear prediction by kriging with application to noise cancellation," Signal Process., vol.80, no.4, pp.553-566, Apr. 2000.
- [5] S.W. Piché, "Steepest descent algorithms for neural network controllers and filters," IEEE Trans. Neural Networks, vol.5, no.2, pp.198-212, Mar. 1994.
- [6] S.Y. Fakhouri, "Identification of the Volterra kernels of nonlinear systems," IEE Proc., vol.127D, no. 6, pp.296-304, Nov. 1980.
- [7] 梶川嘉延, "適応 Volterra フィルタの現状と展望," 電子情報通信学会論文誌 A, vol.J82-A, no.6, pp.759-768, June 1999.
- [8] 曹建庭, 谷萩隆嗣, 呂建明, "リカレントニューラルネットワークを用いた並列非線形適応デジタルフィルタ," 電子情報通信学会論文誌 A, vol.J79-A, no.4, pp.868-877, Apr. 1996.
- [9] 柳坂和秀, 関文隆, 梶川嘉延, 野村康雄, "入力信号のパワー変動を考慮したニューラルフィルタ," 電子情報通信学会論文誌 A, vol.J83-A, no.3, pp.253-262, Mar. 2000.



- [10] B.A. Pearlmutter, "Gradient calculations for dynamic recurrent neural networks: a survey", IEEE Trans. Neural Networks, vol.6, no.5, pp.1212-1228, Nov. 1995.
- [11] M.I. Jordan, "Serial order : A parallel distributed processing approach," ICS-Report 8604 Institute for Cognitive Science, University of California, San Diego La Jolla, California 92903.
- [12] J.L. Elman, "Finding structure in time," Cognitive Science, vol.14, pp.179-211, 1990.
- [13] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink and D.L. Alkon, "Accelerating the convergence of the back-propagation method," Biological Cybernetics, vol.59, no.3, pp.257-263, 1988.
- [14] 賈棋, 戸田尚宏, 臼井支朗, "ニューラルネットにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察," 電子情報通信学会論文誌 D-II, vol.J73-D-II, no.8, pp.1179-1185, Aug. 1990.
- [15] 高木英行, 坂上茂生, 戸川隼人, "ニューラルネット学習における非線形最適化手法の効果," 電子情報通信学会論文誌 D-II, vol.J74-D-II, no.4, pp.528-535, Apr. 1991.
- [16] 田中哲夫, 古村光夫, "逆誤差伝搬量の特異点解消による学習の高速化," 電子情報通信学会論文誌 D-II, vol.J75-D-II, no.5, pp.1000-1008, May 1992.
- [17] 折川典生, 原田豊, "伝達関数の改良による誤差伝搬学習," 電子情報通信学会論文誌 D-II, vol.J83-D-II, no.2, pp.852-854, Feb. 2000.
- [18] 谷萩隆嗣, ニューラルネットワークとファジィ信号処理, コロナ社, 東京, 1998.
- [19] 菊地進一, 中西正和, "短期記憶ニューラルネットワークと高速な構造学習法," 電子情報通信学会論文誌 D-II, vol.J84-D-II, no.1, pp.159-169, Jan. 2001.
- [20] S. Chen, S. Billings, "Representation of nonlinear systems : the NARMAX model," Int. J. Control, vol.49, no.3, pp.1013-1032, 1989.

[21] 谷萩隆嗣, デジタル信号処理の理論 2, コロナ社, 東京, 1985.

---

## 第3章

---

# 雑音量に依存しない音声領域と 雑音領域の判別法を利用した スペクトルサブトラクション

### ●● 本章概要 ●●

スペクトルサブトラクションにおいて音声の雑音除去と明瞭性の向上を図る方法として、音声領域と雑音領域の判別を行う方法がある。本章では、雑音量に依存しない音声領域と雑音領域の判別方法を利用したスペクトルサブトラクションを提案する。提案方法では、音声領域と雑音領域の判別のしきい値を雑音量によって適応的に変化させることにより、判別時の雑音の影響を低減させる。性能評価の結果より、提案方法は従来方法より音声領域と雑音領域の判別が雑音量に依存せず、正確に行われることを示す。さらに、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少できることを示す。

### 3.1 はじめに

近年、スペクトルサブトラクション (SS) において、音声の雑音除去と明瞭性の向上の両者を実現する方法として、音声領域と雑音領域の判別を利用した SS が提案された [1,2]. Nakashima らの方法 [1] では、音声領域と雑音領域の判別に雑音の統計量に関する事前情報が必要である。一方、Yoon & Yoo の方法 [2] では、雑音の統計量に関する事前情報を必要とせず音声領域と雑音領域の判別を行うことが可能である。また、彼らの方法では人間の聴覚特性を考慮した雑音除去を行うためにクリティカルバンド [3] 毎に音声領域と雑音領域の判別を行う。しかし、彼らの方法では、低 SNR において処理した音声にひずみが発生する。これは、低 SNR では雑音量の増加の影響により音声領域を雑音領域と誤判別する割合が増加するためである。雑音領域と誤判別されると、雑音成分を完全に除去しようとするため、雑音だけでなく音声成分も低減してしまい、処理した音声には音声ひずみが発生する。したがって、処理した音声のひずみの低減を図るために、雑音量に依存しない音声領域と雑音領域の判別方法が必要となる。

本章では、雑音量に依存しない音声領域と雑音領域の判別方法を利用した SS を提案する。提案方法では、観測信号が音声と雑音の和であることを仮定し、判別のしきい値を雑音量によって適応的に変化させることにより、判別時の雑音の影響を低減させる。このことにより、提案方法は音声領域と雑音領域の判別が雑音量に依存せず一定の性能を保って行うことが可能である。そのため、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少を図ることができる。提案方法について6種類の雑音による性能評価を行う。その結果、提案方法によって処理した音声は従来方法によって処理した音声と比較して雑音除去性能を維持しながら音声ひずみを減少できることを示す。

### 3.2 従来方法

図 3.1 に Yoon & Yoo の方法の構成図を示す。この方法は、音声領域と雑音領域の判別の際に雑音の統計量に関する事前情報を必要としない点に特徴がある。具体的には、以下の手順により雑音除去を行う。

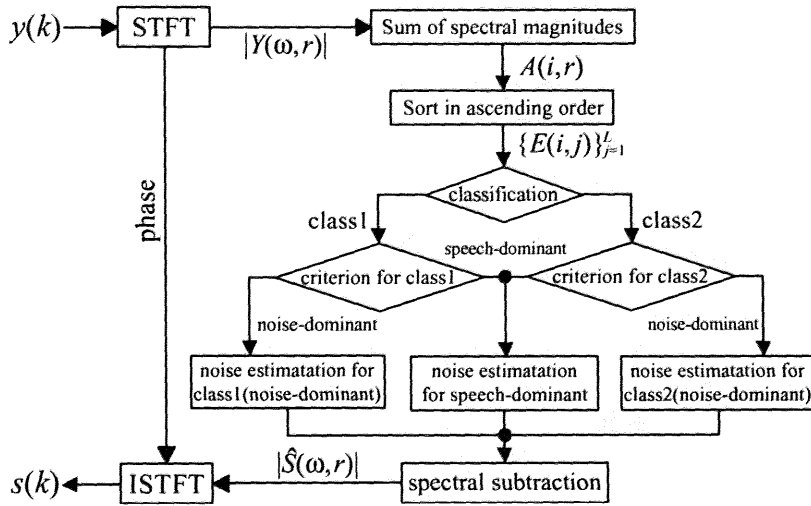


図 3.1 Yoon & Yoo の方法の構成

Fig. 3.1 Overall flow of the conventional method.

まず、観測信号を  $y(k)$  とすると、 $y(k)$  は原音声  $s(k)$  と雑音  $n(k)$  の和、すなわち

$$y(k) = s(k) + n(k) \tag{3.1}$$

$$Y(\omega, r) = S(\omega, r) + N(\omega, r) \tag{3.2}$$

となる。ここで、 $Y(\omega, r)$ ,  $S(\omega, r)$  および  $N(\omega, r)$  はそれぞれ  $r$  番目のフレームにおいて  $y(k)$ ,  $s(k)$  および  $n(k)$  を短時間フーリエ変換 (STFT) したものである。次に、聴覚特性に人間の聴覚特性を考慮した雑音除去を行うためにクリティカルバンド毎に音声領域と雑音領域の判別を行う。そこで、クリティカルバンド毎に観測信号のスペクトルの和  $A(i, r)$  を次式より求める。

$$A(i, r) = \sum_{\omega \in C_i} |Y(\omega, r)| \tag{3.3}$$

ここで、 $|Y(\omega, r)|$  は周波数  $\omega$ ,  $r$  番目のフレームにおける観測信号のスペクトルであり、 $C_i$  は  $i$  番目のクリティカルバンドに属する周波数を表す。ここで、表 3.1 に STFT の周波数バンドとクリティカルバンドの関係を示す。なお、表 3.1 は文献 [4] のクリティカルバンドの構成方法に基づいており、各クリティカルバンドを複数の STFT の周波数バンドで構成している。そして、表 3.1 のクリティカルバンドは各バンドの STFT の周波数バンドが重複しないように周波数分割を考えたときのクリティカルバンドである。次に、 $A(i, r)$  を用いて、領域  $(i, r)$  が音声領域か雑音領域

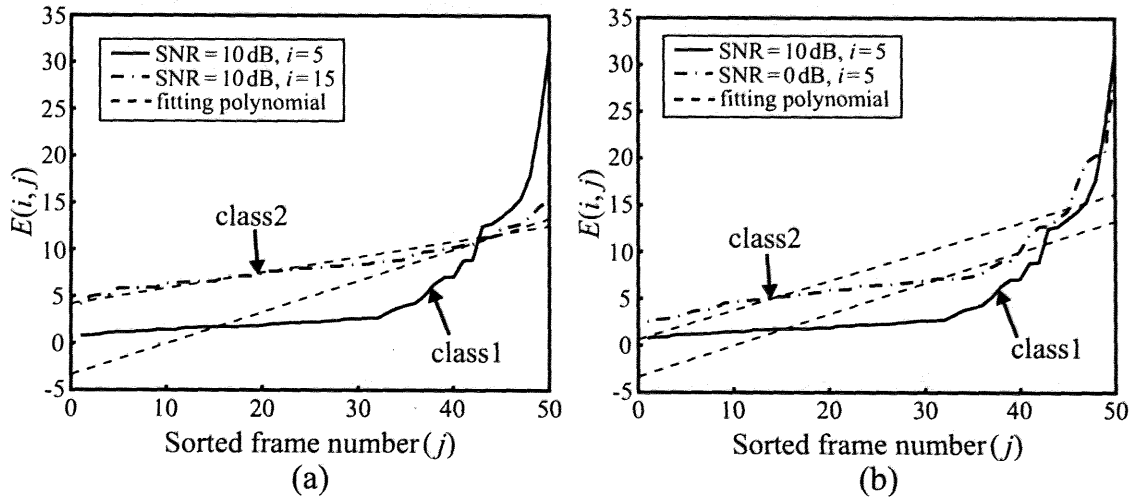


図 3.2 Yoon & Yoo の方法による class の分類 (ホワイトノイズ,  $L = 50$ ) ((a) : 分類の例, (b) : 誤分類の例)

Fig. 3.2 The classification into two classes by the conventional method for white noise where  $L = 50$ . (a) An example of classification. (b) An example of incorrect classification by the influence of noise.

かを判別する. そのために, 各クリティカルバンドにおいて, 過去  $L$  フレーム分の  $A(i, r)$  のデータ列  $\{A(i, j)\}_{j=r-L}^{r-1}$  を生成し, その要素を昇順に並べ替えることにより, データ列  $\{E(i, j)\}_{j=1}^L$  を得る. ここで,  $E(i, p)$  はデータ列  $\{A(i, j)\}_{j=r-L}^{r-1}$  の中で  $p$  番目に小さい要素を表す. そして,  $\{E(i, j)\}_{j=1}^L$  に対して直線近似を行い, 近似した直線の切片の正負により 2 つの class に分類する.

図 3.2(a) に class の分類の例を示す. もし, 近似直線の切片が 0 より小さい場合, 領域  $(i, r)$  は過去  $L$  フレーム中に音声成分が含まれているフレーム数が多いと判断され, 図 3.1 の class 1 に分類される. 一方, 近似直線の切片が 0 以上の場合, 領域  $(i, r)$  は過去  $L$  フレーム中に音声成分が含まれているフレーム数がほとんどないと判断でき, 図 3.1 の class 2 に分類される. class の分類後, それぞれの class において音声領域と雑音領域の判別を行う. 音声領域と雑音領域とを判別するための基準は, 過去  $L$  フレームのスペクトルの分布に基づいたものであり, 各 class で異なる. class 1 の判別基準では領域  $(i, r)$  内に音声成分と雑音成分が含まれている可能性が高いと仮定して, 次式のように音声領域と雑音領域の判別を行う.

$$\left\{ \begin{array}{ll} (i, r)\text{-region is in noise-dominant} & \text{if } (A(i, r) < E(i, [L \cdot a]) \text{ and } i \leq 17) \\ & \text{or } (A(i, r) < \overline{E(i)} \text{ and } i > 17) \\ (i, r)\text{-region is in speech-dominant} & \text{otherwise} \end{array} \right. \quad (3.4)$$

$$\overline{E(i)} = \frac{1}{L} \sum_{j=1}^L E(i, j) \quad (3.5)$$

ここで,  $0 \leq a < 1$ である。また,  $[X]$ は  $X$ の小数点以下を切り上げる。一方, class 2の判別基準では領域  $(i, r)$ 内がほとんど雑音成分のみであると仮定して, 次式のように音声領域と雑音領域の判別を行う。

$$\left\{ \begin{array}{ll} (i, r)\text{-region is in noise-dominant} & \text{if } (A(i, r) < E(i, [L \cdot b]) \text{ and } i \leq 17) \\ & \text{or } i > 17 \\ (i, r)\text{-region is in speech-dominant} & \text{otherwise} \end{array} \right. \quad (3.6)$$

ここで,  $a < b \leq 1$ である。以上の音声領域と雑音領域の判別後, それぞれの領域において雑音スペクトルの推定を行う。

$$|\hat{N}(i, r)| = \left\{ \begin{array}{ll} E(i, [L \cdot low])/B_i & \text{if } (i, r)\text{-region is in speech-dominant} \\ E(i, [L \cdot high])/B_i & \text{if } (i, r)\text{-region is in noise-dominant} \\ & \text{and class 1} \\ c \cdot E(i, L)/B_i & \text{if } (i, r)\text{-region is in noise-dominant} \\ & \text{and class 2} \end{array} \right. \quad (3.7)$$

ただし,  $0 \leq low < 1$ ,  $low < high \leq 1$ ,  $c \geq 1$ である。また,  $B_i$ は  $i$ 番目のクリティカルバンドに属するSTFTのバンド数を表す。式(3.7)より, 領域  $(i, r)$ が音声領域と判別された場合は, 音声成分を保存するために雑音スペクトルは小さく推定される。一方, 領域  $(i, r)$ が雑音領域と判別された場合は, 雑音成分を完全に除去するために雑音スペクトルは大きく推定される。ここで,  $|\hat{N}(i, r)|$ が得られれば, SSを行うことにより, 原音声のスペクトル推定値  $|\hat{S}(\omega, r)|$ が得られる。

$$|\hat{S}(\omega, r)| = \max(|Y(\omega, r)| - |\hat{N}(i, r)|, 0) \quad \omega \in C_i \quad (3.8)$$

最後に, 式(3.8)より得られた  $|\hat{S}(\omega, r)|$ と観測信号の位相を用いて短時間逆フーリエ変換(ISTFT)を行い, 処理した音声  $\hat{s}(k)$ を得る。

しかし、この方法では音声領域と雑音領域の判別方法では処理した音声にひずみが生じ、低SNRにおいて特に顕著である。低SNRでは、雑音量の増加の影響によりclassの分類に誤りが発生する。図3.2(b)に雑音量の増加の影響による誤分類の例を示す。このフレームでは $j > 35$ で $E(i, j)$ が急激に大きくなっており、この部分に音声成分が含まれていると考えられる。つまり、音声領域と判定したい。そのためにはclassの分類でclass 1に分類される必要がある。SNR=10 dBの場合、領域 $(i, r)$ はclass 1と分類される。しかし、SNR=0 dBの場合、領域 $(i, r)$ はclass 2と誤分類されてしまう。これは、雑音量の増加によってデータ列 $\{E(i, j)\}_{j=1}^L$ が正の方向に平行移動したためである。class 2と誤分類されると、雑音領域であることを前提に音声領域と雑音領域の判別を行うため、誤判別されることが多い。雑音領域と判別されると、雑音スペクトルは大きく推定され、SSでは雑音だけでなく音声成分も低減してしまう。よって、処理した音声には音声ひずみが発生する。以上の問題点を解決するために、雑音量に依存しない音声領域と雑音領域の判別方法が必要である。



表 3.1 STFT の周波数バンドとクリティカルバンドの関係

 (サンプリング周波数 16 kHz, フレームサイズ  $N = 512$ )

 Table 3.1 Mapping from STFT binds to critical bands at a sampling frequency of 16kHz and a frame size  $N = 512$ .

critical band number $i$	STFT binds		real frequency [Hz]
	Intervals	Number of binds	
1	1-3	3	0-94
2	4-6	3	94-187
3	7-10	4	187-312
4	11-13	3	312-406
5	14-16	3	406-500
6	17-20	4	500-625
7	21-25	5	625-781
8	26-29	4	781-906
9	30-35	6	906-1094
10	36-41	6	1094-1281
11	42-47	6	1281-1469
12	48-55	8	1469-1719
13	56-64	9	1719-2000
14	65-74	10	2000-2312
15	75-86	12	2312-2687
16	87-100	14	2687-3125
17	101-118	18	3125-3687
18	119-140	22	3687-4375
19	141-169	29	4375-5281
20	170-204	35	5281-6375
21	205-246	42	6375-7687
22	247-256	10	7687-8000

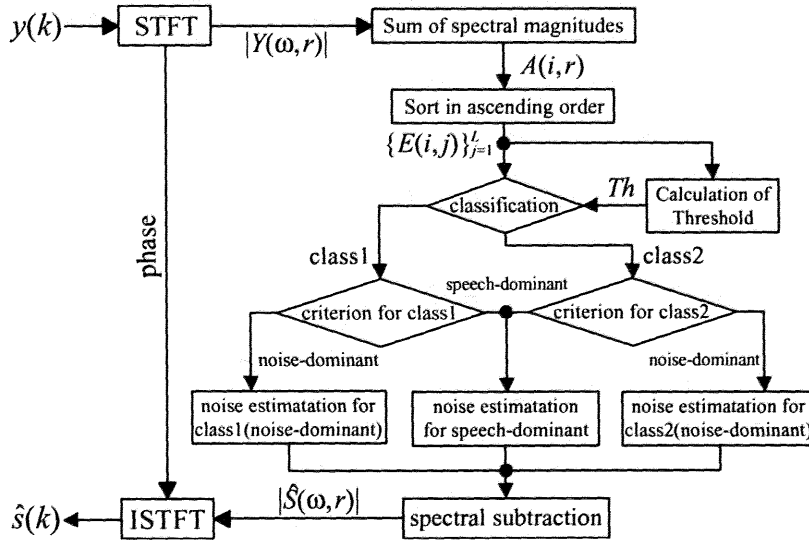


図 3.3 提案方法の構成

Fig. 3.3 Overall flow of the proposed method.

### 3.3 提案方法

図 3.3 に提案方法の構成図を示す。提案方法では、Yoon & Yoo の方法における class の分類のしきい値を雑音量によって適応的に変化させることにより、分類時の雑音の影響を低減させ、雑音量に依存しない音声領域と雑音領域の判別を実現する。これは、従来の方法が高 SNR の場合では雑音の影響が少ないために、class の分類が正確に行われ、結果として満足できる処理結果が得られていることおよび、観測信号が音声と雑音の和であることの 2 点を利用している。まず、領域  $(i, r)$  の雑音成分  $N_c(i, r) = E(i, 1)$  を求める。これは、観測信号が音声と雑音の和であることから、過去  $L$  フレーム分の  $A(i, r)$  を並べ替えて生成したデータ列  $\{E(i, j)\}_{j=1}^L$  の最小値である  $E(i, 1)$  は雑音のみが含まれている可能性が最も高いためである。次に、以下のように class 1 と class 2 の分類を行う。

$$\begin{cases} (i, r)\text{-region is in class 1} & \text{if } v < Th \\ (i, r)\text{-region is in class 2} & \text{if } v \geq Th \end{cases} \quad (3.9)$$

$$Th = N_c(i, r) - d \quad (3.10)$$

ただし、 $v$  は近似直線の切片であり、 $d$  は class の分類を正確に行うためのパラメータである。ここで、 $d = 0$  すなわち、 $Th = N_c(i, r)$  の場合は、class の分類を行うと非

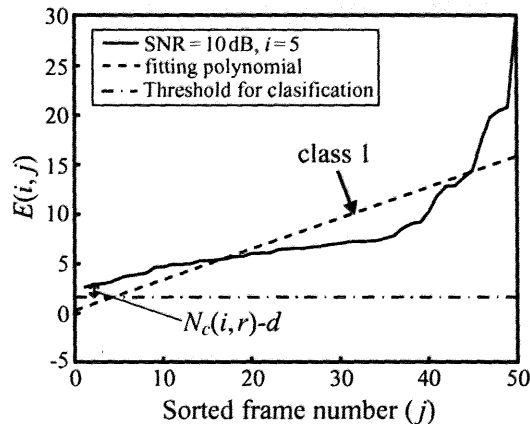


図 3.4 提案方法における class の分類の例 (ホワイトノイズ,  $L = 50$ )

Fig. 3.4 Example of the classification by the proposed method for white noise where  $L = 50$ .

音声領域でも近似直線の切片が  $Th$  より小さくなり, class 1 と誤分類してしまう. しかし,  $d$  を大きくしすぎると, 音声領域でも近似直線の切片が  $Th$  以上になり, class 2 に誤分類してしまう. すなわち, class 1 と class 2 を正確に分類するための  $d$  が存在し, その値は予備実験により求める. class の分類後, それぞれの class において, 式 (3.4) および式 (3.6) を用いて音声領域と雑音領域の判別を行う. さらに, 式 (3.7) および式 (3.8) を用いて雑音スペクトルの推定および SS を行い, 処理した音声を得る. 図 3.4 に提案方法による class の分類の例を示す. 図 3.4 より, class の分類のしきい値  $Th$  を雑音量によって適応的に変化させることにより, 高 SNR の場合の class の分類に近い状態が得られる. そのため, 低 SNR においても高 SNR の場合と同様に class の分類が正確に行えることが期待でき, 音声領域と雑音領域の判別が雑音量に依存せず, 正確に行えることが期待できる. その結果, 処理した音声のひずみが低減できると考えられる.

### 3.4 性能評価

計算機シミュレーションにより提案方法の性能評価を行う。本章では、Yoon & Yooの方法を従来方法として比較を行う。まず、音声のスペクトログラムと判別結果との対比および音声領域の判別率を比較することにより、各方法の音声領域と雑音領域の判別精度を比較する。なお、判別率を算出する際に基にした音声領域については、性能評価に使用する信号のラベル情報がないため、元信号のスペクトログラムよりスペクトログラムの解析に長けている第三者5名に目視してもらい、その結果の平均を取った。しかし、前者は主観的な評価方法であり、また後者は基となる音声領域を第三者に目視によって決定してもらっているため、完全に客観的な評価方法とは言えない。そのため、本章では各判別方法を用いて処理した音声に対して、音質の客観的評価方法である板倉-斉藤ひずみ距離 (Itakura-Saito distortion measure, IS) [5, 6] および segmental SNR の改善度を算出する。さらに、音質の主観的評価方法である MOS(mean opinion score) テスト [6] も行う。そして、以上の5種類の評価結果を総合的に判断することにより、提案する判別方法の有効性を評価する。このうち、ISは処理後の信号のスペクトルと原音声のスペクトルとの間の相違を表す指標であり、次式で定義される。

$$\begin{aligned}
 D_{IS} &= \frac{1}{M} \sum_{r=1}^M d_{IS}(r) \\
 &= \frac{1}{M} \sum_{r=1}^M \left\{ \frac{1}{N} \sum_{\omega=1}^N \frac{S(\omega, r)}{\hat{S}(\omega, r)} - \ln \frac{S(\omega, r)}{\hat{S}(\omega, r)} - 1 \right\} \quad (3.11)
 \end{aligned}$$

ここで、 $S(\omega, r)$  および  $\hat{S}(\omega, r)$  はそれぞれ原音声および処理した音声のスペクトルである。また、 $N$  はFFTの分析フレーム長であり、 $M$  は全フレーム数である。ただし、非現実的に高いスペクトルの距離値の影響をなくすために、 $d_{IS}(r)$  の上位5%はISの計算から除外した [7]。音声処理において、音声のひずみは音声成分が除去されることによって発生する。よって、処理した音声と原音声との相違を表すISがひずみの度合いを示す数値として考えることもできる。そして、音声のひずみが小さい、すなわち音声成分が保存されていれば、式(3.11)よりISは小さくなる。そのため、ISを音声成分が保存されているかの指標として考えることができる。一方、segmental SNRの改善度  $G_{SNR}$  は雑音が観測信号に比べどれほど減少したかを示す

数値であり、次式で定義される。

$$G_{SNR} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{k=Nm}^{Nm+N-1} n^2(k)}{\sum_{k=Nm}^{Nm+N-1} \{\hat{s}(k) - s(k)\}^2} \quad (3.12)$$

従来方法および提案方法では雑音領域が音声領域と誤判別されると、その後の雑音スペクトルの推定およびSSでは、雑音の除去は十分でないものの、音声成分を十分保存される。そのため、処理した音声は音声領域の雑音が十分に除去されない分、segmental SNRの改善度は低減するものの、音声成分が保存されるためISは小さくなる。一方、音声領域が雑音領域と誤判別されると、雑音は十分に除去されるものの、音声成分も除去される。そのため、雑音除去性能すなわちsegmental SNRの改善度は大きいものの、ISも大きくなる。以上のことから、ISおよびsegmental SNRの改善度の結果を組み合わせることにより、音声領域と雑音領域の判別精度を評価することができる。

### 3.4.1 シミュレーション条件

原音声としては女声の「休み無く打ち寄せてはさっと引いていく白い波」と発生したものを用いる。雑音はNOISEX-92データベース [8] よりホワイトノイズ、ピンクノイズ、F16 コックピットノイズ、バブルノイズの4種類、電子協騒音データベース（日本電子工業振興協会）より交差点雑音および計算機室（中型）雑音の2種類の計6種類を用いる。ただし、原音声および雑音のサンプリング周波数は16 kHz、量子化レベルは16 bit とする。STFTの分析フレーム長は $N = 512$ とし、1/2 オーバーラップとする。また、窓関数にはハミング窓を用いる。従来方法のその他のパラメータについては、予備実験を行い、適切なパラメータを設定する。一方、提案方法のその他のパラメータは $d$ を除いて従来方法で設定したものと同一とし、表3.2に示す。

提案方法におけるパラメータ $d$ については予備実験により求める。予備実験で使用した音声信号は研究用連続音声データベース（日本音響学会）の男性話者2名（can0001, mit0001）、女性話者2名（can1001, mit1001）による発話文5文（a01～a05）の計20文である。図3.5に $d$ の変化とISとの関係を、図3.6に $d$ の変化とsegmental SNRの改善度との関係を示す。図3.5より、ISについては $d$ が大きくな

表 3.2 従来方法および提案方法のパラメータ

Table 3.2 The parameters of the conventional method and the proposed method.

parameter	value
$L$	50
$a$	0.6
$b$	0.9
$c$	1.5
<i>high</i>	0.8
<i>low</i>	0.3

ると IS も大きくなり、高 SNR ほどこの現象が顕著である。これは、 $d$  を大きくしすぎることにより音声成分が class 2 へ分類される割合が大きくなり、その後の判別で雑音領域と誤判別されやすくなる。その結果、処理した音声は雑音だけでなく弱い音声成分も除去され音声にひずみが生じるためであると考えられる。このことから、音声ひずみの点からは  $d$  をできるだけ小さくする必要があることがわかる。一方、図 3.6 より、低 SNR、高 SNR とともに  $d$  が大きくなると segmental SNR の改善量が大きくなり、 $d$  が 1.0 より大きくなると改善量がほぼ一定になっていることがわかる。これは、 $d$  を大きくすることで雑音領域が class 2 へ分類される割合が増え、その結果雑音が十分に除去されるためだと考えられる。このことから、雑音の除去性能の観点から  $d \geq 1.0$  の範囲で設定すればよいことがわかる。本章では、以上の結果より  $d = 1.0$  と設定する。

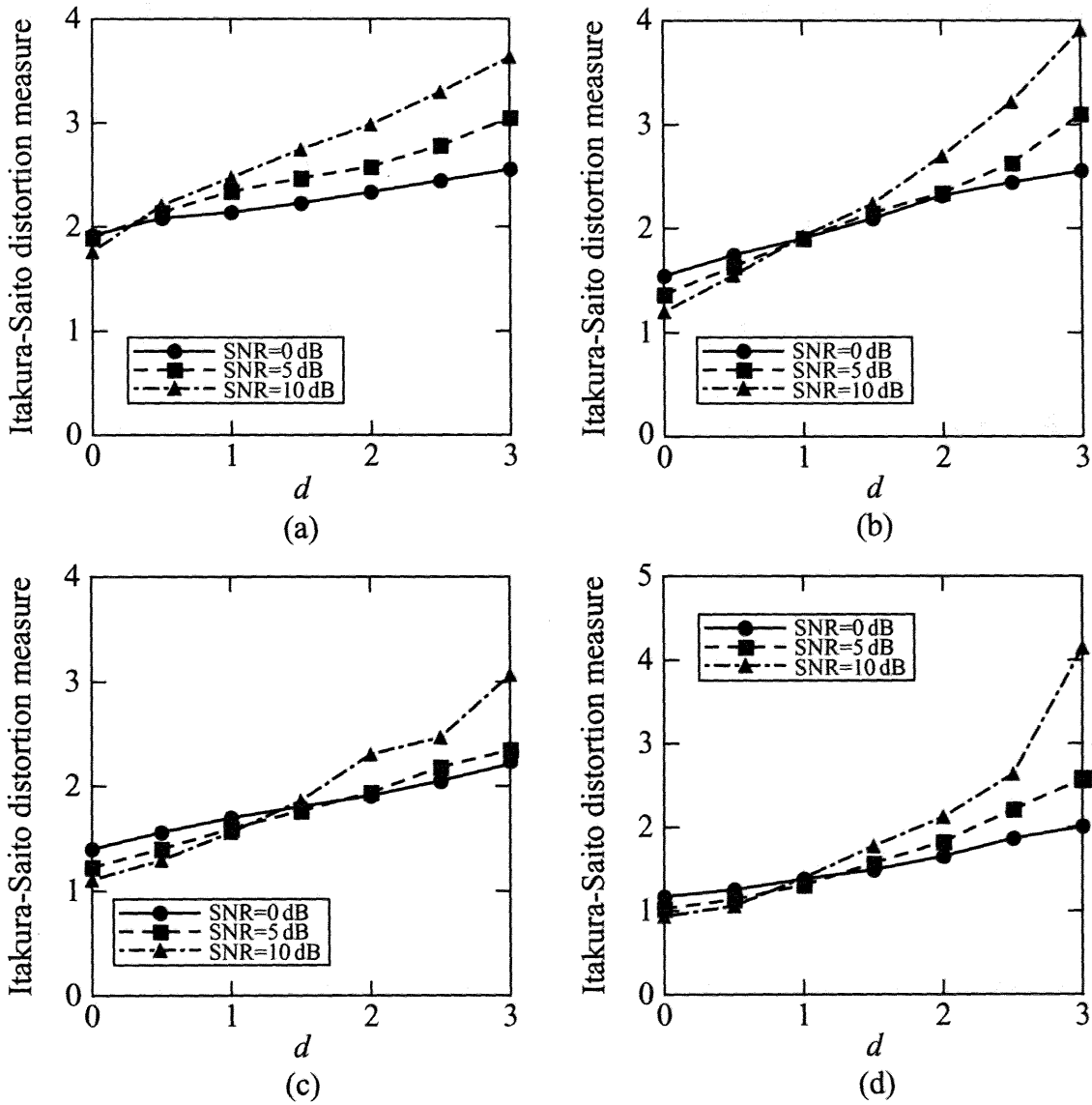


図 3.5 パラメータ  $d$  に対する板倉-斉藤ひずみ距離 ((a): ホワイトノイズ, (b): ピンクノイズ, (c): F16 コックピットノイズ, (d): バブルノイズ)

Fig. 3.5 Itakura-Saito distortion measure for parameter  $d$ ((a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise).

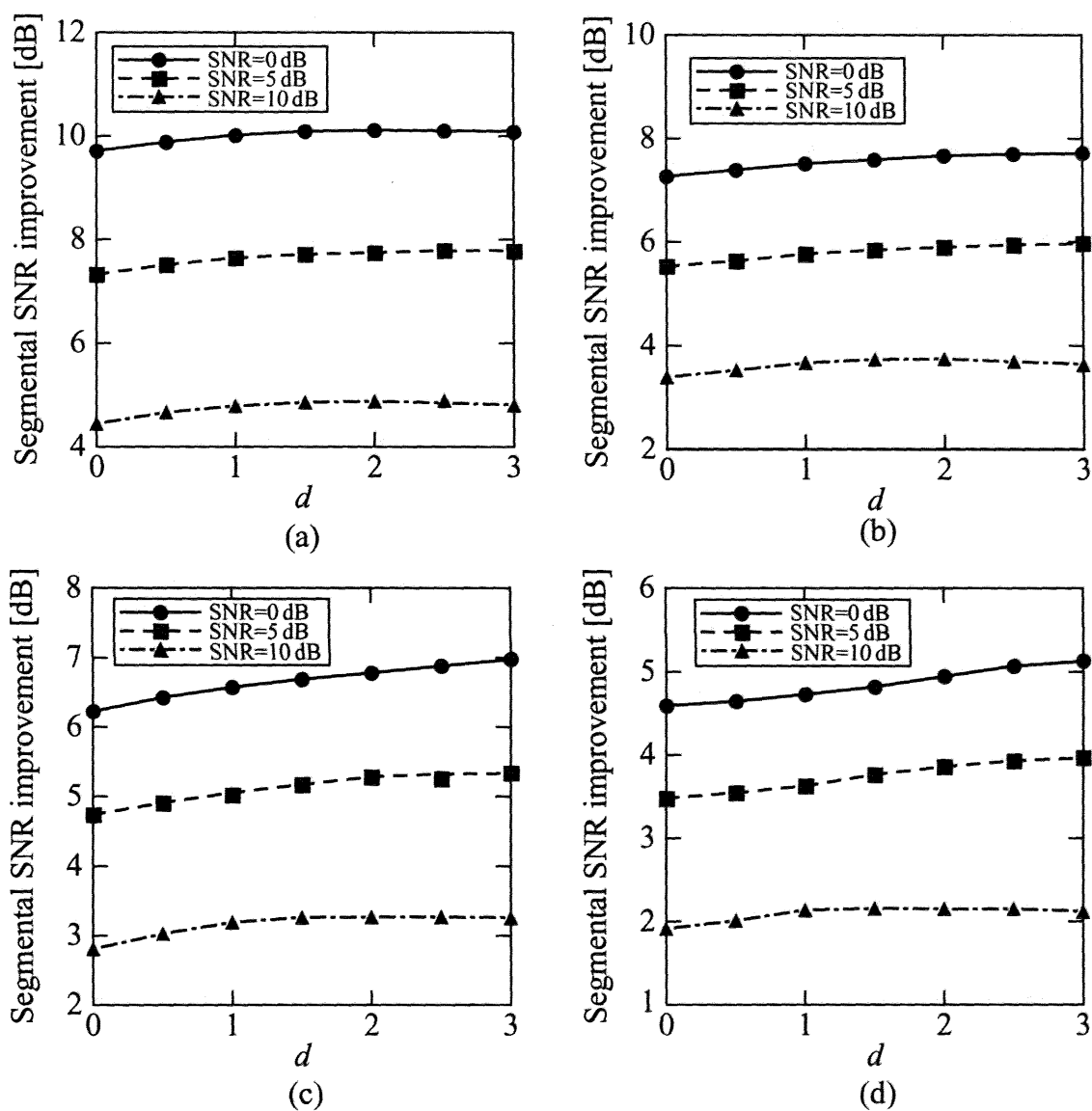


図 3.6 パラメータ  $d$  に対する segmental SNR の改善度 ((a) : ホワイトノイズ, (b) : ピンクノイズ, (c) : F16 コックピットノイズ, (d) : バブルノイズ)

Fig. 3.6 Segmental SNR improvement for parameter  $d$  ((a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise).



### 3.4.2 判別結果の比較

図 3.7(a) および (b) に原音声および観測信号のスペクトログラムをそれぞれ示す。図 3.7(a) および (b) は濃淡分布であり、濃い点ほどスペクトルのパワーが強いことを示している。図 3.7(c) および (d) に従来方法と提案方法の判別結果を時間一周波数領域で表示したものをそれぞれ示す。図 3.7(c) および (d) は白と黒の 2 値分布であり、黒い部分は音声領域と判別された領域である。一方、白い部分は雑音領域と判別された領域である。また、プロットが低周波数にいくほど細かくなっているのは、従来方法および提案方法の両者とも音声領域と雑音領域の判別を表 3.1 のクリティカルバンド毎に行っているためである。図 3.7(a), (b) および (c) より、従来の判別方法では音声成分を雑音領域と誤判別しており、特に高周波成分において顕著であることがわかる。このことは、雑音量の増加の影響により class の分類で class 2 への誤分類が発生し、それが音声領域と雑音領域の誤判別につながったためであると考えられる。一方、提案する判別方法では図 3.7(a),(b) および (d) より、全体的に音声成分を雑音領域と誤判別する割合が減少されていることがわかる。特に、図 3.7(d) の丸く囲った部分のように、従来方法では音声成分を雑音領域と誤判別していた部分が正しく音声領域と判別されていることがわかる。このことは、提案方法では高 SNR の場合に近い状態で class の分類を行っているため、class 2 への誤分類が少なくなり、その結果音声成分の誤判別が減少したためだと考えられる。

図 3.8 に従来方法と提案方法における音声領域の判別率のグラフを示す。判別率を算出する際に基にした音声領域については、性能評価に使用する原音声のラベル情報がないため、図 3.7(a) の原音声のスペクトログラムよりスペクトログラムの解析に長けている第三者 5 名に目視してもらい、その結果の平均を取った。雑音がホワイトノイズ、ピンクノイズ、F16 コックピットノイズおよび計算機室雑音の場合は図 3.8(a),(b),(c) および (e) より、提案方法が従来方法より判別率が高いことがわかる。これは、提案方法では高 SNR の場合に近い状態で class の分類を行っているため、class 2 への誤分類が少なくなり、その結果、音声成分の誤判別が減少したためだと考えられる。また、従来方法および提案方法ともに全体に判別率が低い理由としては、無声音などの弱い音声成分の判別が有声音の場合と比較して難しいためであると考えられる。さらに、低 SNR になるにつれて判別率の低下がみられるのは、低 SNR では音声雑音が雑音に埋もれてしまうために、音声領域の判別自体が難しく

なっているためだと考えられる。一方、雑音がバブルノイズおよび交差点雑音の場合は、図 3.8(d) および (f) より高 SNR の部分では従来方法の方が判別率が高くなっている。これは、バブルノイズが複数の音声から構成されている非定常雑音であり、高 SNR になると過去  $L$  フレームのデータの中に雑音に加わっている部分と雑音に加わっていない部分が存在する。そのため、 $N_c(i, r)$  がほぼ 0 となり、class 分類のしきい値は  $Th \approx -d$  となる。その場合、class の分類はデータ列  $\{E(i, j)\}_{j=1}^L$  にさらに雑音を加わった状態と同じになるため、従来方法と比べて class 2 と誤分類する割合が増える。その結果、提案方法の音声領域の判別率が従来方法より低くなったと考えられる。ここで、パラメータ  $d$  について考えると、予備実験の結果より設定した  $d = 1.0$  を用いることにより、低 SNR においても高 SNR の場合の class の分類に近い状態が得られる。そのため、従来方法に比べて音声領域と雑音領域を正確に判別することができる。そして、図 3.5 および図 3.6 より、予備実験に使用した 4 種類の性質の異なる雑音において、パラメータ  $d$  の変化に対する IS および segmental SNR の改善度の変化の傾向が同じである。そのため、予備実験の結果より設定した  $d = 1.0$  は他の雑音においても有効であると考えられる。以上のことから、パラメータ  $d$  は雑音の種類や SNR に依存せずに用いることができるため、Nakashima らの方法に対して指摘した事前情報にあたらぬ。

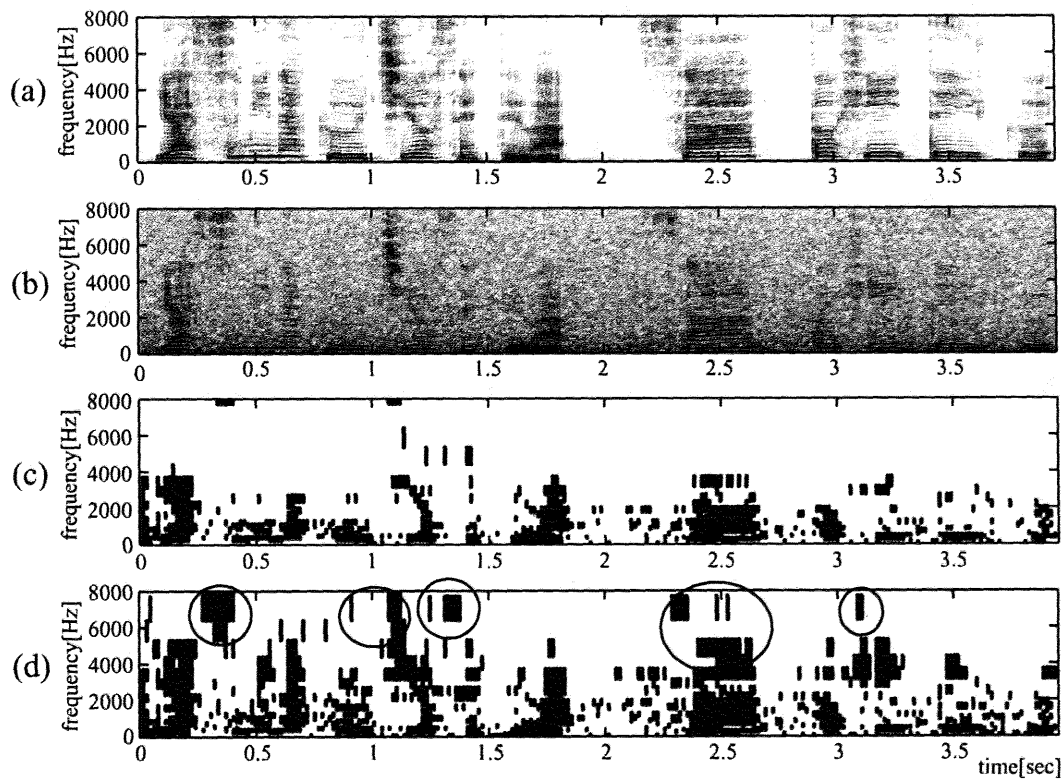


図 3.7 従来方法と提案方法の判別結果の比較 ((a):原音声のスペクトログラム, (b):観測信号のスペクトログラム (ピンクノイズ, SNR=10 dB), (c):従来方法の判別結果, (d):提案方法の判別結果)

Fig. 3.7 Spectrograms of (a) Clean speech, (b) Observation signal (additive pink noise with SNR=10 dB). The plot shows speech dominant part in time-frequency domain as (c) the conventional method, (d) the proposed method.

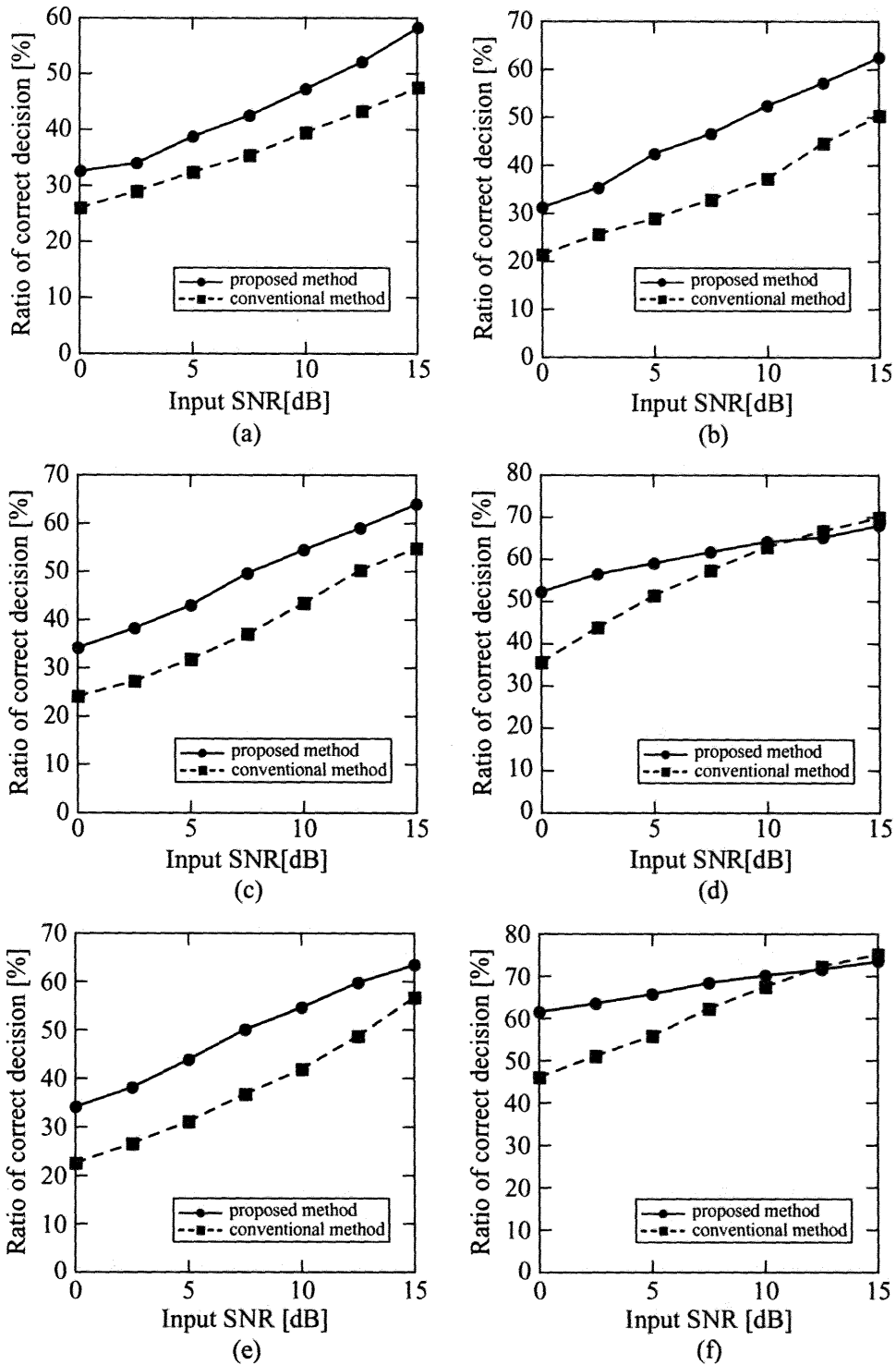


図 3.8 入力 SNR における音声領域の判別率 ((a): ホワイトノイズ, (b): ピンクノイズ, (c): F16 コックピットノイズ, (d): バブルノイズ, (e): 計算機室雑音, (f): 交差点雑音

Fig. 3.8 Ratio of correct decision of the speech component for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

### 3.4.3 板倉－斉藤ひずみ距離および segmental SNR

図 3.9 に従来方法と提案方法における IS のグラフを示す。雑音がホワイトノイズ、ピンクノイズ、F16 コックピットノイズおよび計算機室雑音の場合は図 4.8(a),(b),(c) および (e) より、提案方法が従来方法より優れている特性を示すことがわかる。また、提案方法は雑音量に関係なく、一定のひずみを保っていることがわかる。提案方法では雑音成分を低減させたデータ列を class の分類に使用しているため、雑音量に依存しない class の分類が行える。そのため、図 3.8 より、提案方法は従来方法と比較して音声成分を雑音領域と誤判別する割合が少ない。その結果、音声成分が保存され、処理した音声のひずみが抑えられると考えられる。雑音がバブルノイズおよび交差点雑音の場合は、図 3.9(d) および (f) より高 SNR の部分で従来方法の方が IS が小さくなっている。バブルノイズおよび交差点雑音は非定常雑音であり、高 SNR になると過去  $L$  フレームのデータの中に雑音に加わっていない部分が存在する。よって、class の分類のしきい値は  $Th \approx -d$  となる。その場合、class の分類はデータ列  $\{E(i, j)\}_{j=1}^L$  にさらに雑音に加わった状態と同じになるため、従来方法と比べて class 2 と誤分類する割合が増える。このことは音声の判別率より確認した結果、図 3.8(d) および (f) のように提案方法の音声領域の判別率が従来方法より低くなる。以上の結果から、提案方法が従来方法より処理した音声のひずみを抑えるのに有効であることがわかる。

図 3.10 に従来方法と提案方法における segmental SNR の改善度のグラフを示す。雑音がホワイトノイズの場合は図 3.10(a) より、提案方法は従来方法とほぼ同等の雑音除去性能を示していることがわかる。一方、雑音がピンクノイズ、F16 コックピットノイズ、バブルノイズ、計算機室雑音および交差点雑音の場合は図 3.10(b),(c),(d),(e) および (f) より、低 SNR の部分で従来方法の方が優れていることがわかる。そして、提案方法と従来方法との差はほとんどの部分において 0.5 dB 未満である。これは、低 SNR の場合、従来方法は雑音の影響により音声領域を class 2 と分類され、class 2 における音声領域と雑音領域の判別で音声領域を雑音領域と誤判別する割合が多くなるために、処理した音声は弱い音声成分を除去しながらも雑音の除去を十分に行うためである。

IS および segmental SNR の改善度の結果より、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少できることがわかる。

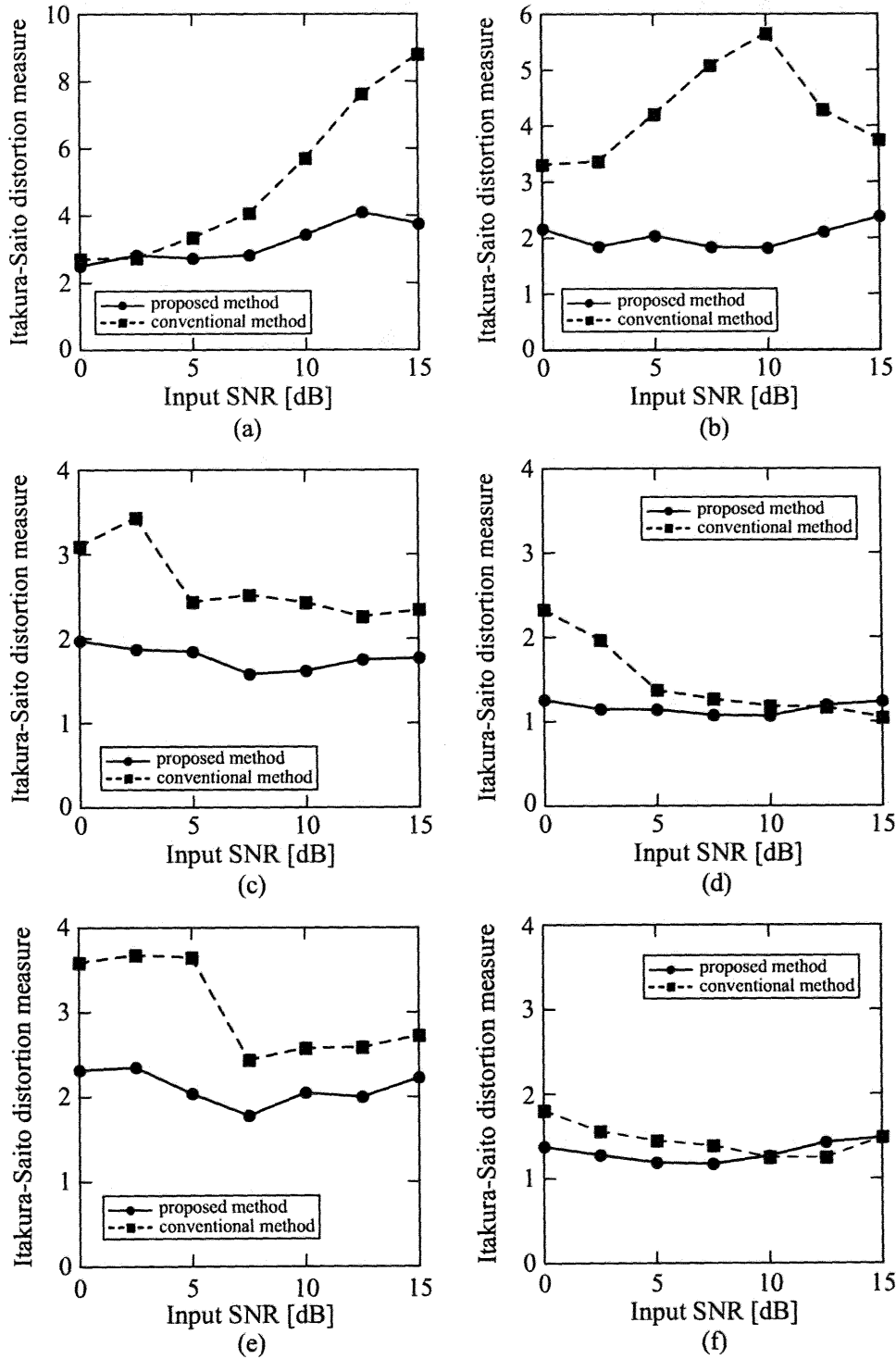


図 3.9 入力 SNR における板倉-斉藤ひずみ距離 ((a): ホワイトノイズ, (b): ピンクノイズ, (c): F16 コックピットノイズ, (d): バブルノイズ, (e): 計算機室雑音, (f): 交差点雑音)

Fig. 3.9 Itakura-Saito distortion measure for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

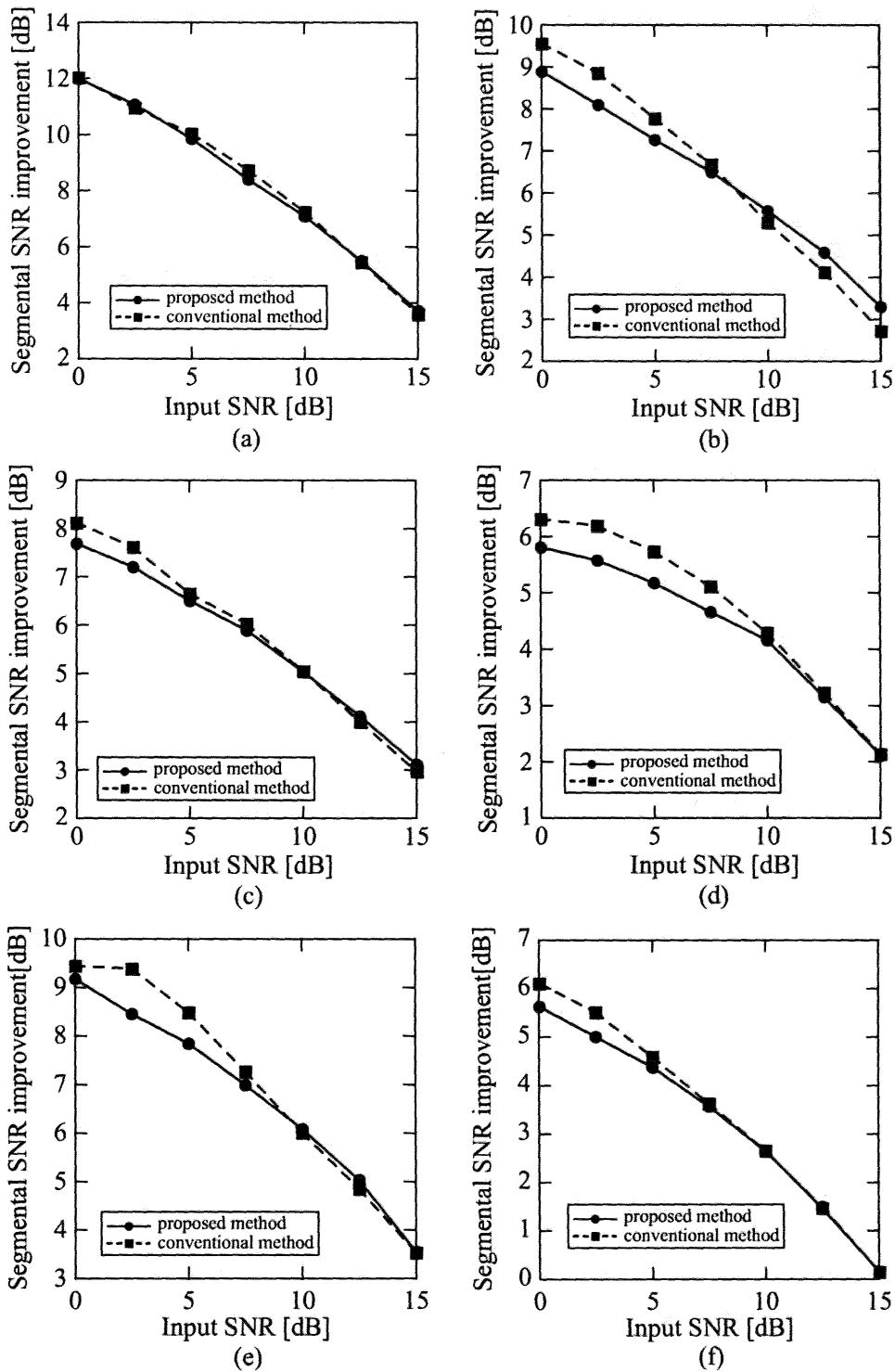


図 3.10 入力 SNR における segmental SNR の改善度 ((a) : ホワイトノイズ, (b) : ピンクノイズ, (c) : F16 コックピットノイズ, (d) : バブルノイズ, (e) : 計算機室雑音, (f) : 交差点雑音)

Fig. 3.10 Segmental SNR improvement for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

表 3.3 MOS テストの結果 (SNR=5, 10 dB)

Table 3.3 Comparison of MOS tests

SNR [dB]	method	white	pink	F16	babble
5	conventional	2.23	2.47	2.27	1.93
	proposed	2.40	2.63	2.47	1.73
10	conventional	3.63	3.53	3.20	2.73
	proposed	3.73	4.00	3.33	2.80

### 3.4.4 MOS テスト

試聴テストの1つであるMOSテストを行い、提案方法について主観評価を行った。各テストとも被験者は原音声および雑音を加えた観測信号を1回ずつ聞いた後に、従来方法および提案方法により処理した音声をそれぞれ3回ずつ聞く。ただし、処理した音声の順番はランダムにする。処理した音声を聞いた後に、1~5の5段階で評価を行う。なお、MOSテストのスコアは1が最も悪く、5が最も良い値となる。本章では被験者は10名とし、各被験者から得られたスコアの平均を取る。

表3.3にSNR=5, 10 dBにおけるMOSテストの結果を示す。雑音がホワイトノイズ、ピンクノイズおよびF16コックピットノイズの場合は、表3.3より提案方法で処理した音声のスコアが従来方法に比べて同等もしくは良くなっている。3種類の雑音における提案方法のスコアの上昇は、図3.9および図3.10より、処理した音声の雑音除去性能を保ちつつ、音声ひずみの発生を抑えているためであると考えられる。雑音がバブルノイズの場合は、表3.3よりSNR=10 dBの場合は従来方法と提案方法とが同等のスコアであるものの、SNR=5 dBの場合は従来方法のスコアが提案方法のスコアより良くなっている。これは、図3.9(d)および図3.10(d)より、SNR=5 dBでは音声のひずみが同程度であるが、segmental SNRの改善度は従来方法の方が優れているために、雑音除去性能が良い従来方法の方が聞きやすいと判断されたためだと考えられ。

以上の結果から、主観的評価においても本章で提案した判別方法が有効であることが言える。



### 3.5 まとめ

本章では、雑音量に依存しない音声領域と雑音領域の判別方法を利用したスペクトルサブトラクションを提案した。提案方法では、音声領域と雑音領域の判別のしきい値を雑音量によって適応的に変化させることにより、判別時の雑音の影響を低減させるため、雑音量に依存しない判別が可能となる。性能評価の結果より、提案方法は従来方法より音声領域と雑音領域の判別が雑音量に依存せず、正確に行われ、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少できることを示した。さらに、処理した音声のひずみは雑音量に関係なくほぼ一定であることを示した。

### 第3章の参考文献

- [1] H. Nakashima, Y. Chisaki and T. Usagawa, "Spectral subtraction based on statistical criteria of the spectral distribution," IEICE Trans. Fundamentals, vol.E85-A, no.10, pp.2283-2292, Oct. 2002.
- [2] S. Yoon and C.D. Yoo: "Speech enhancement based on speech/noise-dominant decision," IEICE Trans. Inf. & Syst., vol.E85-D, no.4, pp.744-750, Apr. 2002.
- [3] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag, Berlin, Germany, 1990.
- [4] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Process., vol.7, no.2, pp.126-137, Mar. 1999.
- [5] A. El-Jaroudi and J. Makhoul, "Discrete All-Pole Modeling," IEEE Trans. on Signal Process., vol.39, no.2, pp.411-423, Feb. 1991.
- [6] Jr. J.R. Deller, J.Hansen, and J.G. Proakis, Discrete-time processing of speech signals, IEEE Press, New York, 2000.
- [7] Y. Hu and P.C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," in Proc. of IEEE ICASSP 2002, Orlando, USA, May 2002.
- [8] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun., vol.12, no.3, pp.247-251, July 1993.

---

## 第4章

---

# スペクトログラム上の特徴量に基づく 音声領域と雑音領域の判別法を利用 したスペクトルサブトラクション

### ●● 本章概要 ●●

雑音の事前情報を用いない音声領域と雑音領域との判別方法として、複数の短時間フーリエ変換の周波数から構成されるバンドにおける観測信号のスペクトログラム上の特徴量に着目し、各バンド内の標準偏差を利用した音声領域と雑音領域との判別方法を提案した。音声領域と雑音領域の判別の後、それぞれの領域において異なるパラメータを用いたスペクトルサブトラクションを行った。性能評価の結果より、提案方法が従来方法より処理した音声のミュージカルノイズや音声ひずみを減少できることを示した。そして、提案方法が雑音の事前情報を用いない1入力システムの雑音除去方法として有効な方法であることを示した。

## 4.1 はじめに

近年、スペクトルサブトラクション (SS) において、雑音の統計量に関する事前情報を必要とせず雑音除去を行う方法が提案された [1, 2]. Stahl らの方法 [1] では観測信号のスペクトルの変位値に基づいた雑音スペクトルの推定を行う. しかし, この方法では処理した音声にはミュージカルノイズが発生する. ミュージカルノイズが発生する原因としては, スペクトルの不適切な減算処理や不十分なフィルタリングによって発生するスペクトルの不連続性が挙げられる. 一方, Yoon & Yoo の方法 [2] では過去のフレームの観測信号を用いた音声領域と雑音領域の判別を行い, それぞれの領域において Stahl らの方法に基づいた雑音スペクトルの推定が行われる. しかし, 彼らの方法では過去のフレームの観測信号の影響により音声領域と雑音領域の誤判別が発生するため, 雑音スペクトルの推定に誤差が生じ, SS 後の処理した音声にミュージカルノイズやひずみが発生する. したがって, 処理した音声のミュージカルノイズやひずみを減少させるために, 過去の観測信号に依存しない音声領域と雑音領域とを判別する精度の高い方法が必要となる.

本章では, 複数の短時間フーリエ変換 (STFT) の周波数から構成されるバンドにおける観測信号のスペクトログラム上の特徴量に着目し, 各バンド内の標準偏差を利用した音声領域と雑音領域の判別方法を提案する. バンド内の成分が音声と雑音で構成される場合は, 音声成分と雑音成分との周波数成分での特徴の違いから観測信号のスペクトルの標準偏差は高くなる. 一方, バンド内の成分が雑音のみであれば, 雑音成分の周波数成分での特徴は音声成分のものに比べて一様であるため, 観測信号のスペクトルの標準偏差は低くなる. したがって, バンド毎の観測信号のスペクトルの標準偏差に適切なしきい値を設定することにより, 音声領域と雑音領域の判別が可能である. 音声領域と雑音領域の判別の後, それぞれの領域において SS を行う. 提案方法は従来方法と同様に雑音の統計量に関する事前情報を必要としない. さらに, 提案する判別方法は過去のフレームの観測データに依存しないため, 従来の判別方法で発生していた過去のフレームのデータの影響による誤判別がない. また, 観測信号のスペクトログラム上の特徴を利用することにより, 正確に音声領域と雑音領域の判別を行うことが可能である. そのため, 雑音スペクトルの推定誤差が減少し, 処理した音声のミュージカルノイズや音声ひずみが減少する. 提案方法について 6 種類の雑音による性能評価を行う. その結果, 提案方法が従来方法より

処理した音声のミュージカルノイズや音声ひずみを減少できることを示す。そして、提案方法が雑音の事前情報を用いない1入力システムの雑音除去方法として有効な方法であることを示す。

## 4.2 従来方法の問題点

### 4.2.1 変位値に基づく雑音スペクトルの推定

Stahlらの方法では、観測信号のスペクトルの時間的な変位値を利用して雑音スペクトルの推定を行う。この方法は、雑音の統計量に関する事前情報を必要としない点の特徴である。雑音スペクトルの推定は以下の手順によって行われる。まず、STFTの周波数 $\omega$ において、観測信号のスペクトル $|Y(\omega, r)|$  ( $r = 0, 1, \dots, R$ ) を以下のように昇順に並べ替える。

$$|Y(\omega, r_0)| \leq |Y(\omega, r_1)| \leq \dots \leq |Y(\omega, r_R)| \quad (4.1)$$

ただし、 $r$ は短時間フーリエ変換のフレームの番号を、 $R$ は現在のフレームの番号を表す。次に、雑音スペクトルの推定値は各 $\omega$ において少なくとも $[qR]$ フレーム<sup>1</sup>分は雑音であると仮定し、

$$|\hat{N}(\omega, r)| = |Y(\omega, r_{[qR]})| \quad (4.2)$$

となる。ここで、 $0 \leq q \leq 1$ である。例えば、 $q = 0$ の場合は並べ替えた観測信号のスペクトルの最小値、 $q = 1$ の場合は最大値をそれぞれ取る。この方法を用いてロバストな雑音スペクトルの推定を行うためには、 $q$ は中間値に近い値、すなわち $q \approx 0.5$ にするのが望ましい。Stahlらの方法では、以上の手順で推定した雑音スペクトルを用いてSSを行うことにより、処理した音声を得る。しかし、Stahlらの方法で処理した音声にはミュージカルノイズが発生する。ミュージカルノイズが発生する原因としては、雑音領域において雑音が不連続的に残されるためであると考えられる。

<sup>1</sup> $[X]$ は $X$ の小数点以下を切り上げる。例えば、 $[1.3] = 2$ である。

## 4.2.2 音声領域と雑音領域の判別法を用いたスペクトルサブトラクション

雑音の統計量に関する事前情報を必要とせず、音声領域と雑音領域の判別を行うことができる方法として Yoon & Yoo の方法が提案された。彼らの方法の概説は第 3.2 節に示したとおりである。彼らの方法は、表 3.1 のクリティカルバンド毎に音声領域と雑音領域とを判別し、判別結果を用いて雑音のスペクトルの推定を行った上で SS を行う。また、Stahl らの方法と同様に、雑音スペクトルの推定の際に雑音の統計量に関する事前情報を必要としない。しかし、彼らの方法では処理した音声にミュージカルノイズやひずみが生じてしまう。このことは雑音のスペクトルの推定誤差によるものが大きい。Yoon & Yoo の方法で雑音のスペクトルの推定誤差が発生する原因としては、過去  $L$  フレームの観測信号の影響により class の分類の際に誤分類が発生し、それにより音声領域と雑音領域の判別に誤りが発生する場合が考えられる。誤判別が発生する原因としては、具体的には次のような場合が考えられる。まず、領域  $(i, r)$  が雑音領域で、過去  $L$  フレーム中に音声成分が含まれているフレーム数が多い場合は、class の分類で class 1 と誤分類される。class 1 と誤分類されると、音声領域であることを前提に音声領域と雑音領域の判別を行うため、音声領域と誤判別されることが多くなる。また、領域  $(i, r)$  が音声領域で、過去  $L$  フレーム中に音声成分が含まれているフレーム数が少ない場合にも同様の誤判別が生じることが多くなる。これらの誤判別は、音声領域と雑音領域が切り替わった後の数フレームにおいて特に多く発生する。以上の問題点を解決するために、過去の観測信号に依存しない音声領域と雑音領域の判別方法が必要となる。

## 4.3 提案方法

### 4.3.1 提案方法の構成

図 4.1 に本章で提案する方法の構成図を示す。提案方法は以下の手順により雑音の除去を行う。

- 1) 観測信号  $y(k)$  を STFT により周波数領域へ変換する。

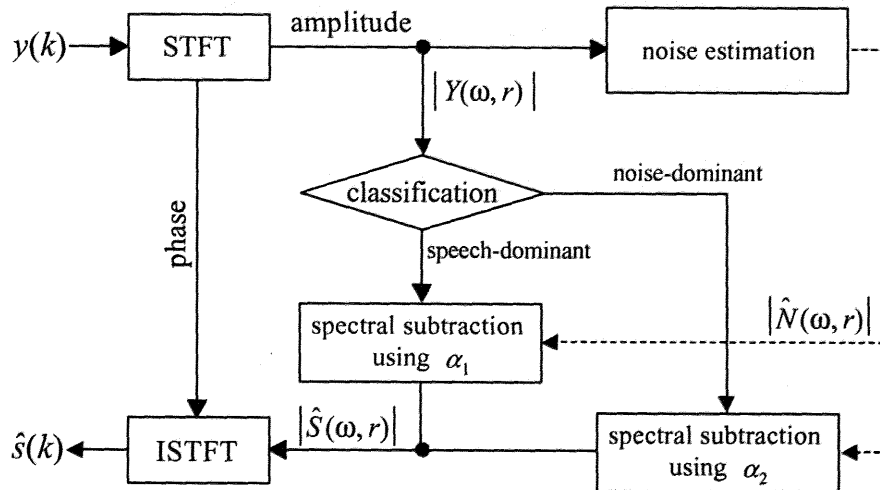


図 4.1 提案方法の構成

Fig. 4.1 Overall flow of the proposed method.

- 2) 複数の STFT の周波数から構成されるバンド毎に観測信号のスペクトルの標準偏差を用いて音声領域と雑音領域の判別を行う。ただし、各バンドの幅  $BW$  は同一とする。
- 3) 式 (4.2) を用いて STFT の周波数毎に雑音スペクトルの推定を行う。
- 4) 音声領域と雑音領域の判別の後、SS を行う。ただし、音声領域および雑音領域においてサブトラクション係数をそれぞれ設定する。
- 5) 4) より得られた  $|\hat{S}(\omega, r)|$  と観測信号の位相より短時間逆フーリエ変換を行い、処理した音声  $\hat{s}(k)$  を得る。

### 4.3.2 音声領域と雑音領域の判別

図 4.2(a) に原音声（女性が「休み無く打ち寄せてはさっと引いていく白い波」と発声したもの）のスペクトログラム、図 4.2(b) に原音声に SNR=5 dB でホワイトノイズを付加したときの観測信号のスペクトログラム、図 4.2(c) に図 4.2(b) の 0~250 Hz の部分 ( $BW = 8$  の場合の 1 番目のバンドに相当) を拡大表示したものをそれぞれ示す。スペクトログラム上での音声成分は図 4.2(a) および (c) より、一定の周波数間隔でパワーの強い部分とほぼ 0 である部分とがストライプ状に存在していることがわかる。この構造のことを調波構造といい、有声音特有の構造である。調波構

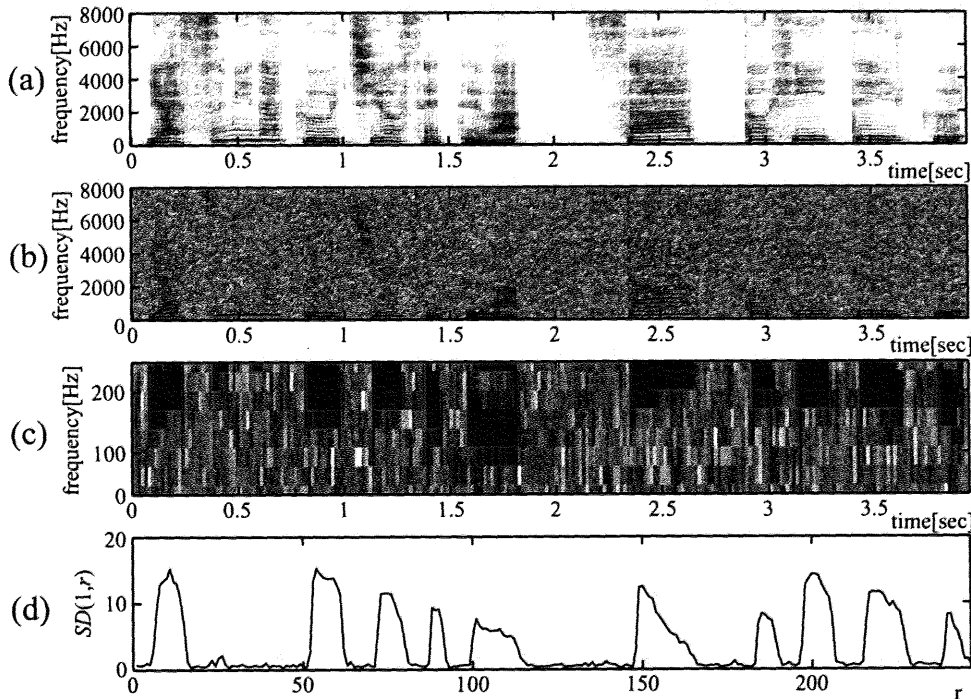


図 4.2 観測信号のスペクトルの標準偏差 ((a):原音声のスペクトログラム, (b):観測信号のスペクトログラム (ホワイトノイズ, SNR= 5 dB), (c):観測信号のスペクトログラム (周波数 : 0~250 Hz), (d): $SD(i, r)$  ( $i=1, BW=8$ ))

Fig. 4.2 Example of speech/noise-dominant classification using the spectrum of observation signal (a) Spectrogram of clean speech, (b) Spectrogram of observation signal (additive white noise with SNR=5 dB), (c) Spectrogram of observation signal (range : 0-250 Hz), (d) The output of  $SD(i, r)$  ( $i=1, BW=8$ )

造は基本周波および複数の高調波から構成されている。なお、音声の基本周波数は男性の場合は平均 125 Hz、標準偏差 20.5 Hz であり、女性の場合はそれぞれ男性の約 2 倍に等しい [3]。一方、雑音成分は図 4.2(b) および (c) より、灰色から白のまだらな様子になっていることがわかる。以上のことから、スペクトログラム上でみられる音声成分と、ホワイトノイズによる雑音成分はスペクトルの周波数成分での特徴が異なることがわかる。また、他の定常性雑音でも同様の特徴があると確認できる。そのため、スペクトログラムを画像として捉え、画像の特徴量を抽出することにより、音声領域と雑音領域の判別が行えると考えられる。本章では画像の局所の特徴量を表す指標として用いられる標準偏差を利用する。標準偏差（分散）は画像の濃淡に関する特徴量としてパターンの分割やテクスチャ解析などに利用されてい



る [4-6]. 本章では, 標準偏差は複数の STFT の周波数から構成されるバンド毎に計算する. 領域  $(i, r)$  における標準偏差  $SD(i, r)$  は以下の式によって求める.

$$SD(i, r) = \sqrt{\frac{1}{BW} \sum_{\omega=BW(i-1)+1}^{BW*i} \{|Y(\omega, r)| - \overline{|Y(i, r)|}\}^2} \quad (4.3)$$

$$\overline{|Y(i, r)|} = \frac{1}{BW} \sum_{\omega=BW(i-1)+1}^{BW*i} |Y(\omega, r)| \quad (4.4)$$

ここで,  $i$  はバンドの番号を表す. 領域  $(i, r)$  について考えると, 領域  $(i, r)$  の中に音声成分と雑音成分で構成されている場合は, 領域内に音声の調波構造によるパワーの強い部分が一定周波数間隔毎に存在することにより標準偏差  $SD(i, r)$  が大きくなる. 一方, 領域  $(i, r)$  が雑音のみの場合は, 図 4.2(b) および (c) より, 雑音成分のスペクトルの変動は音声成分の調波構造による変動に比べれば小さいため,  $SD(i, r)$  は音声成分が含まれる場合よりも小さくなる. 図 4.2(d) に  $BW = 8$  の場合の 1 番目のバンド (周波数: 0~250 Hz) における  $SD(i, r)$  の出力を示す. 図 4.2 より, 音声成分があるところでは  $SD(i, r)$  の値は大きくなり, 雑音成分のみのところでは  $SD(i, r)$  の値が小さくなることがわかる. また, 図 4.3 に原音声に SNR=5 dB でホワイトノイズを付加したときの 1 番目のバンドの  $SD(i, r)$  のヒストグラムを示す. なお, このヒストグラムを作成する際に使用した原音声は, 研究用連続音声データベース (日本音響学会) の男性話者 4 名 (can0001, can0002, mit0001, mit0002), 女性話者 4 名 (can1001, can1002, mit1001, mit1002) による発話文 25 文 (a01~a25) の計 200 文である. 図 4.3 のヒストグラムは 0 付近の急激なピークと 4 を中心とする緩やかな分布の 2 つに分けることができる. 図 4.2 および図 4.3 より, 前者はバンド内に雑音しか含まれていない場合であり, 後者はバンド内に雑音成分と音声成分の両者が含まれている場合であると考えられる. また, 他の定常性雑音でも同様の特徴があると確認できる. 以上のことから,  $SD(i, r)$  に対して適切なしきい値  $Th$  を設定することにより, 以下のように音声領域と雑音領域の判別が可能である.

$$\begin{cases} (i, r)\text{-region is in speech-dominant} & \text{if } SD(i, r) > Th \\ (i, r)\text{-region is in noise-dominant} & \text{if } SD(i, r) \leq Th \end{cases} \quad (4.5)$$

ただし, 音声領域と雑音領域の判別を行う際のしきい値  $Th$  は小さすぎると雑音を音声領域と誤判別する割合が多くなり, 大きすぎると音声成分を雑音領域と誤判別

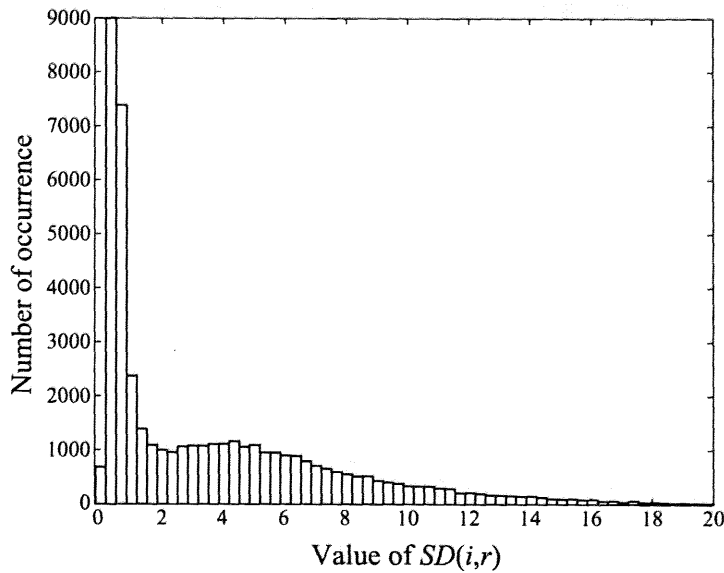


図 4.3  $SD(i, r)$  のヒストグラム ( $i=1, BW=8, SNR=5$  dB)

Fig. 4.3 Histogram of  $SD(i, r)$  ( $i=1, BW=8, SNR=5$  dB).

する割合が多くなる。提案する判別方法では、観測信号のスペクトログラム上の特徴を利用することにより、正確に音声領域と雑音領域の判別を行うことが可能である。また、提案方法では、過去のフレームの観測信号のスペクトルに依存しないため、従来の判別方法で発生していた過去のフレームのデータの影響による誤判別がなくなる。特に、音声領域と雑音領域が切り替わった後の数フレームにおいては判別精度の向上が期待できる。そのため、雑音スペクトルの推定における誤差が減少し、処理した音声のミュージカルノイズやひずみを減少することができる。

### 4.3.3 判別後の処理

前節の方法により音声領域と雑音領域の判別を行った後、SSを行う。サブトラクション係数は図 4.1 のように、音声領域と雑音領域とでそれぞれ別の値を設定する。音声領域のサブトラクション係数  $\alpha_1$  は、音声成分を保存しつつ雑音の除去を行える値に設定する。一方、雑音領域のサブトラクション係数  $\alpha_2$  は、雑音の除去を十分に行うために  $\alpha_1$  より大きい値に設定する。

## 4.4 性能評価

計算機シミュレーションにより提案方法の性能評価を行う。本章では、比較対象として Stahl らの方法および Yoon & Yoo の方法を用いる。なお、Stahl らの方法は提案方法のシステムより音声領域と雑音領域の判別を部分を除いたものに相当する。評価は音声の時間波形およびスペクトログラムの比較、segmental SNR の改善度、板倉－斉藤ひずみ距離 (IS) [7,8] および MOS テスト [8] の4つの方法で行う。

### 4.4.1 シミュレーション条件

原音声としては、女性が「休み無く打ち寄せてはさっと引いていく白い波」と発声したものを用いる。雑音は NOISEX-92 データベース [9] よりホワイトノイズ、ピンクノイズ、F16 コックピットノイズ、バブルノイズの4種類、電子協騒音データベース（日本電子工業振興協会）より交差点雑音および計算機室（中型）雑音の2種類の計6種類を用いる。なお、原音声および雑音のサンプリング周波数は 16 kHz、量子化レベルは 16 bit である。STFT の分析フレーム長は  $N = 512$  とし、1/2 オーバーラップとする。また、窓関数にはハミング窓を用いる。

提案方法の各パラメータについては、雑音がホワイトノイズの場合において予備実験を行い、その結果より適切な値を選ぶ。まず、音声領域と雑音領域の判別を行うバンドの幅  $BW$  についてであるが、 $BW$  が小さすぎると、低 SNR の場合に雑音成分の標準偏差が高 SNR の場合より大きくなるため、雑音領域を音声領域と誤判別する可能性が大きくなる。一方、 $BW$  が大きすぎると、高周波数領域において無声音成分による周波数成分の特徴の変化を捉えることができないため、音声領域を雑音領域と誤判別する可能性が大きくなる。これらのことから、精度良く判別できる適切なバンド幅が存在すると考えられる。図 4.4(a) に  $BW$  の変化と segmental SNR の改善度との関係を、図 4.4(b) に  $BW$  の変化と IS との関係を示す。音声領域と雑音領域の判別が正確に行われていれば、音声成分が保存されかつ、雑音が十分に減少するため、segmental SNR の改善度が大きくなり、IS は小さくなる。SNR = -5 dB においては、図 4.4(a) より、 $BW$  を小さくすると、segmental SNR が下がっていることがわかる。このことは、 $BW$  を小さくしすぎると、低 SNR における雑音成分の標準偏差の上昇の影響により、雑音領域を音声領域と誤判別する可能性が大き

なり、スペクトルサブトラクションにおいて十分に雑音が除去されないためだと考えられる。一方、 $BW$  大きくすると図 4.4 より、segmental SNR は改善されているが、IS は値が大きくなる。このことは、 $BW$  大きすぎると音声成分を雑音領域と誤判別する割合が多くなり、スペクトルサブトラクションにおいて雑音だけでなく音声成分も除去され音声にひずみが生じるためだと考えられる。SNR=0 dB および SNR=5 dB の場合は、図 4.4 より、segmental SNR および IS とともに SNR=-5 dB の場合ほど大きな変化がみられないが、 $BW=4$  もしくは  $BW=8$  のとき、segmental SNR および IS が良い特性を示していることがわかる。以上のことから、雑音の除去性能と音声ひずみのバランスを考慮して、本章では  $BW=8$  と設定する。

また、式 (4.2) 中の雑音スペクトルの推定のパラメータ  $q$  についてであるが、システムの性能が他のパラメータより  $q$  の変化に敏感である。図 4.5(a) に  $q$  の変化と segmental SNR の改善度との関係を、図 4.5(b) に  $q$  の変化と IS との関係を示す。図 4.5(a) より、低 SNR、高 SNR とともに  $q$  を大きくすることで、segmental SNR が改善されていることがわかる。これはミュージカルノイズが減少しているためであると考えられる。一方、IS については図 4.5(b) より、 $q$  が 0.4 から 0.5 の範囲であればさほど変化していないことがわかる。しかし、 $q$  が 0.5 を超えたところから値が急激に大きくなり、高 SNR ほどこの現象が顕著である。これは、 $q$  を大きくしすぎることにより SS で減算する量が大きくなり、雑音だけでなく弱い音声成分も除去され音声にひずみが生じるためであると考えられる。図 4.5 より、 $q$  が雑音除去性能および音声ひずみに大きく関与しており、用途に応じて適切な  $q$  を設定する必要があることがわかる。ここでは、雑音の除去性能と音声ひずみのバランスを考慮して  $q=0.5$  と設定する。音声領域と雑音領域の判別を行う際のしきい値  $Th$  は小さすぎると雑音を音声領域と誤判別する割合が多くなり、大きすぎると音声成分を雑音領域と誤判別する割合が多くなる。そのため、本章では  $Th=2.5$  とする。SS のパラメータである  $\alpha_1$  および  $\alpha_2$  については、それぞれ  $\alpha_1=2.5$ 、 $\alpha_2=5$  とする。

Stahl らの方法のパラメータについては提案方法と同様に  $q=0.5$ 、SS のサブトラクション係数を  $\alpha=2.5$  とする。また、Yoon & Yoo の方法のパラメータについては、文献値を基に提案方法と同様の予備実験を行い、実験結果より適当な値を選ぶ。

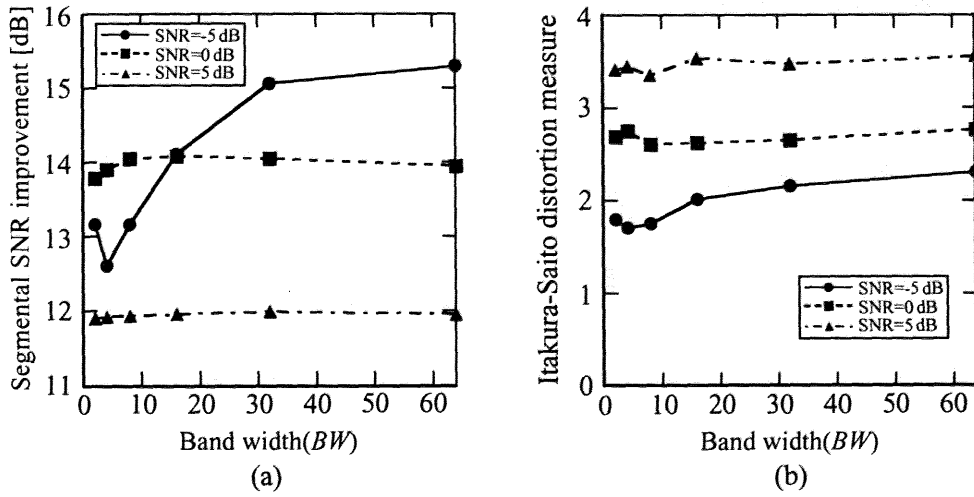


図 4.4 バンド幅 ( $BW$ ) に対する (a):segmental SNR の改善度, (b):板倉-斉藤ひずみ距離

Fig. 4.4 (a) Segmental SNR improvement, (b) Itakura-Saito distortion measure for band width( $BW$ ).

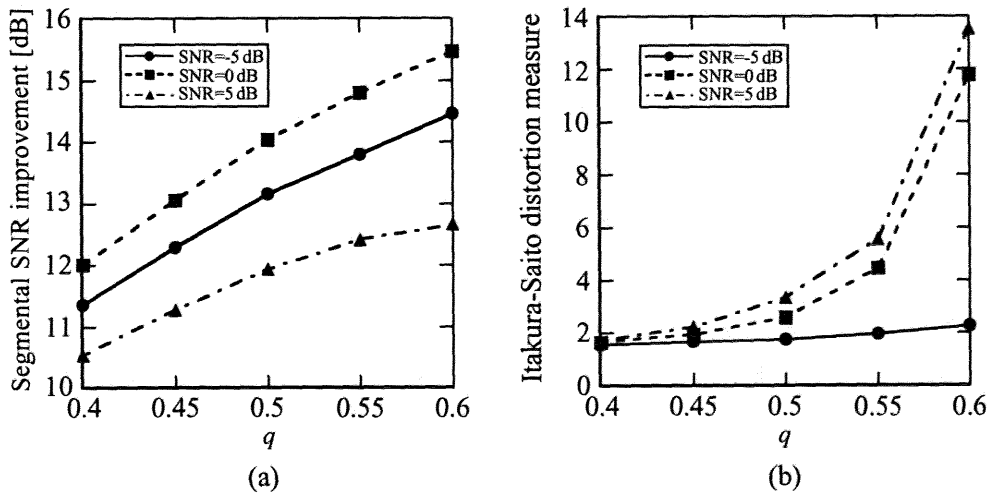


図 4.5 パラメータ  $q$  に対する (a):segmental SNR の改善度, (b):板倉-斉藤ひずみ距離

Fig. 4.5 (a) Segmental SNR improvement, (b) Itakura-Saito distortion measure for parameter  $q$ .

#### 4.4.2 音声波形およびスペクトログラム

図 4.6 に従来方法と提案方法により処理をした音声の波形およびスペクトログラムを示す。図 4.6(1), (2), (3), (4), (5) はそれぞれ原音声, 観測信号 (F16 コックピットノイズ, SNR=10 dB), Stahl らの方法で処理した音声, Yoon & Yoo の方法で処理した音声, 提案方法で処理した音声である。また, 図 4.6(a) は時間波形, (b) は (a) のスペクトログラムである。図 4.6(3)(b) より, Stahl らの方法で処理した音声は音声成分が十分に保存されているものの, 雑音除去が十分でないためミュージカルノイズが多いことがわかる。また, 図 4.6(3)(b) および (4)(b) より, Yoon & Yoo の方法で処理した音声は Stahl らの方法よりミュージカルノイズが減少していることがわかる。しかし, 低周波数部分, 特に 1kHz 以下の部分でミュージカルノイズが集中してみられる。また, 図 4.6(1)(b) の原音声のスペクトログラムと比較して音声成分が削られており, 高周波成分において顕著であることがわかる。このことから, Yoon & Yoo の方法で処理した音声にはひずみが発生していると考えられる。これらのことは, Yoon & Yoo の方法の音声領域と雑音領域の判別が低周波数成分では雑音領域を音声領域と誤判別し, 高周波数成分では音声領域を雑音領域と誤判別している割合が多いからであると考えられる。一方, 提案方法で処理した音声は図 4.6(3)(b), (4)(b) および (5)(b) より, ミュージカルノイズが従来方法と比較して低減していることがわかる。また, 図 4.6(1)(b), (4)(b) および (5)(b) より, Yoon & Yoo の方法より音声成分が保存されていることがわかる。これらのことは, 提案する判別方法が Yoon & Yoo の判別方法より誤判別が少ないためであると考えられる。以上のことより, 提案する判別方法は Yoon & Yoo の判別方法より優れていることがわかる。

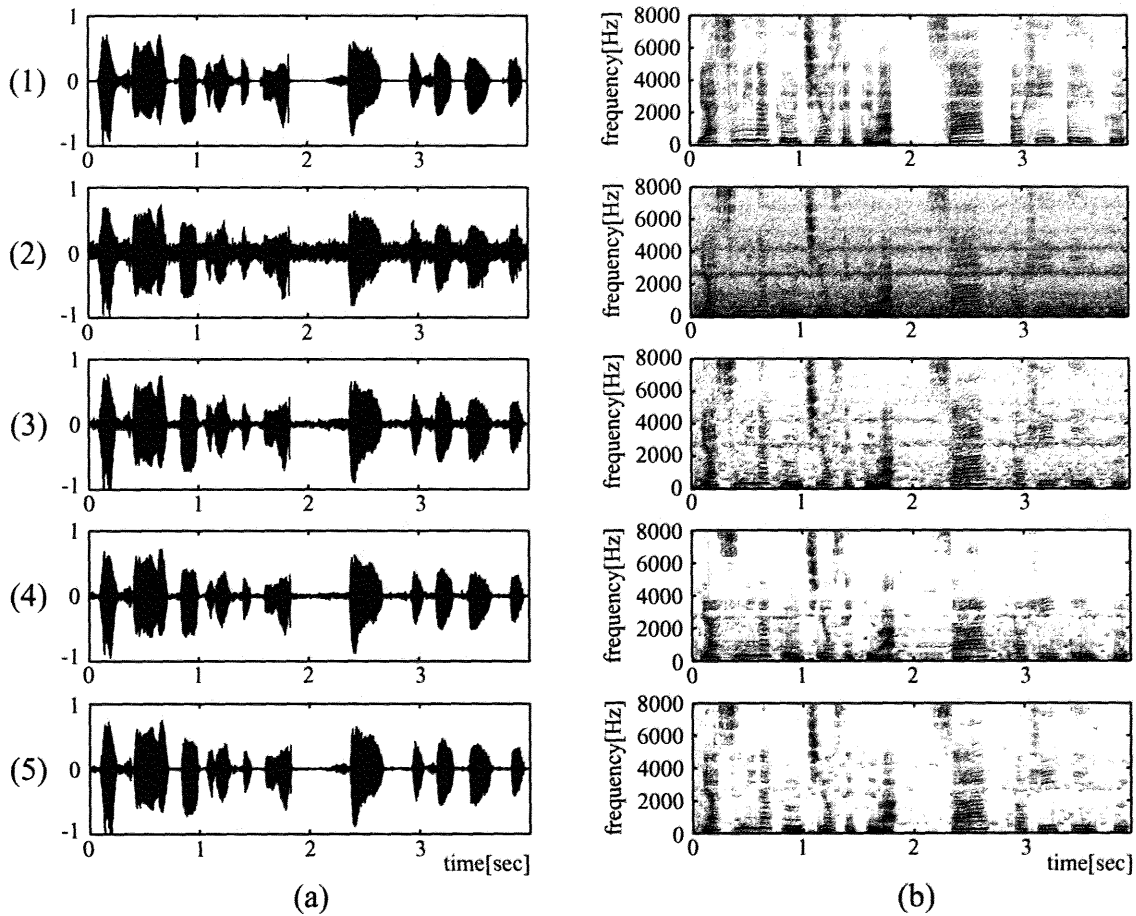


図 4.6 処理信号の時間波形とスペクトログラム ((1) : 原音声, (2) : 観測信号 (F16 コックピットノイズ, SNR=10 dB), (3) : Stahl らの方法による処理信号, (4) : Yoon & Yoo の方法による処理信号, (5) : 提案方法による処理信号, (a) : 時間波形, (b) : スペクトログラム)

Fig. 4.6 Waveform (a) and spectrograms (b) of the (1) Clean speech, (2) Observation signal (additive F16 cockpit noise with SNR=10 dB), (3) Enhanced speech by the method by Stahl, (4) Enhanced speech by the method by Yoon & Yoo, (5) Enhanced speech by the proposed method.

### 4.4.3 Segmental SNR

図 4.7 に従来方法と提案方法における segmental SNR の改善度のグラフを示す。図 4.7 より、Yoon & Yoo の方法および提案方法は Stahl らの方法より優れていることがわかる。これは、Stahl らの方法で処理した音声は雑音除去が不十分なためにミュージカルノイズが多く残っているためだと考えられる。このことから、音声領域と雑音領域の判別がミュージカルノイズの除去に有効であることがいえる。また、Yoon & Yoo の方法と提案方法について比較してみると、加えた雑音がホワイトノイズ、ピンクノイズ、F16 コックピットノイズおよび計算機室雑音の場合は、図 4.7(a),(b),(c) および (e) より全般的に提案方法が Yoon & Yoo の方法より優れていることがわかる。提案する判別方法は Yoon & Yoo の判別方法より正確に音声領域と雑音領域の判別が行われていることにより、雑音を音声として扱うことが少なくなる。その結果、不連続的に雑音が残されることによって発生するミュージカルノイズが低減したと考えられる。一方、バブルノイズおよび交差点雑音の場合は、図 4.7(d) および (f) より低 SNR では Yoon & Yoo の方法の方が優れていることがわかる。これは、バブルノイズはおよび交差点雑音が非定常雑音であるため、提案する判別方法において雑音領域を音声領域と誤判別されやすくなるためであると考えられる。以上の結果から、本章で提案する方法は雑音のスペクトログラム上の特徴が音声と異なる場合、雑音の事前情報を用いない 1 入力の雑音除去方法として有効であることが言える。



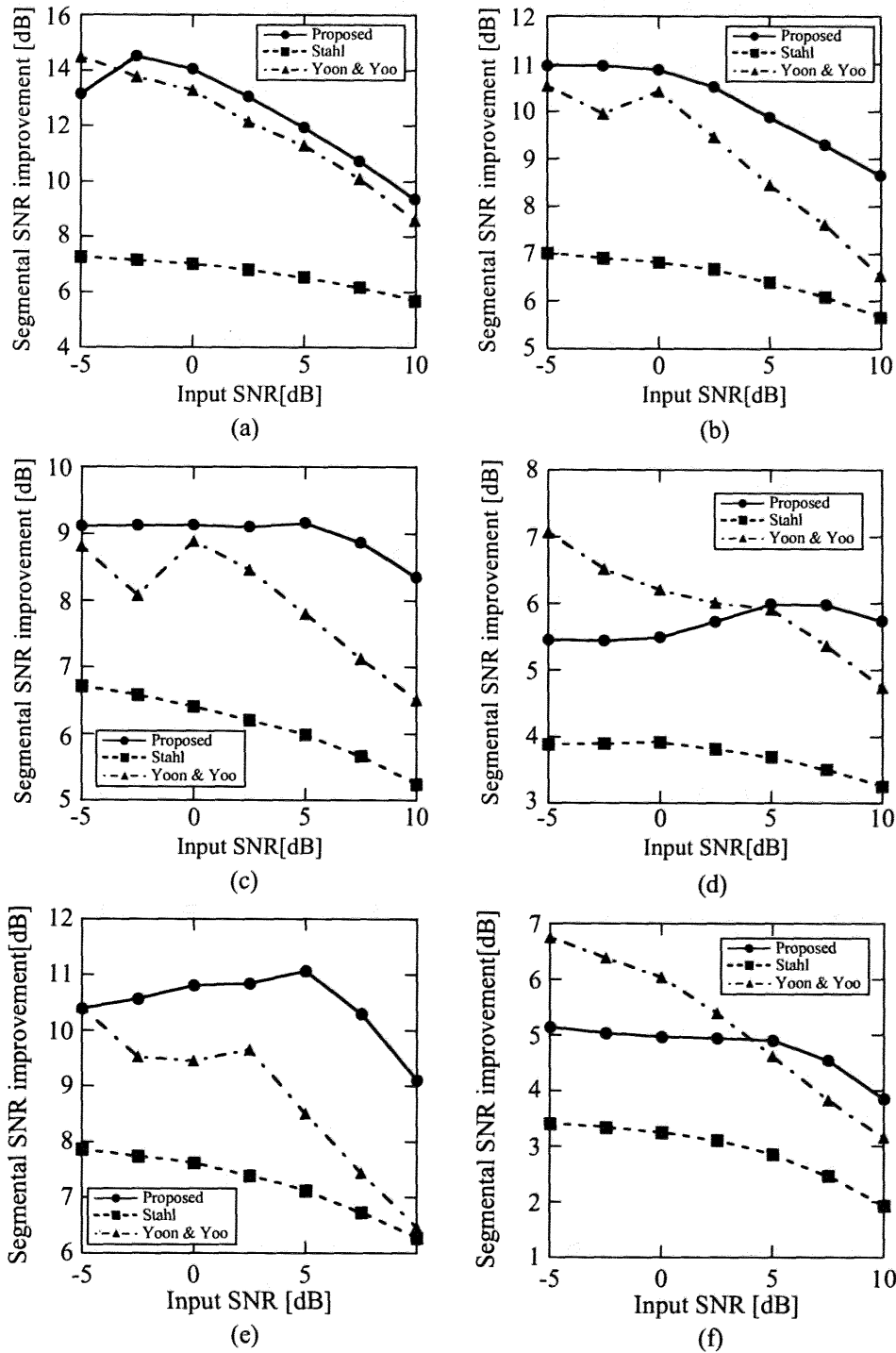


図 4.7 入力 SNR における segmental SNR の改善度 ((a) : ホワイトノイズ, (b) : ピンクノイズ, (c) : F16 コックピットノイズ, (d) : バブルノイズ, (e) : 計算機室雑音, (f) : 交差点雑音)

Fig. 4.7 Segmental SNR improvement for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

#### 4.4.4 板倉－斉藤ひずみ距離

図4.8に従来方法と提案方法におけるISのグラフを示す。図4.8より、Stahlらの方法はYoon & Yooの方法および提案方法と同等もしくは優れている特性を示すことがわかる。これは、Stahlらの方法によって処理された音声は雑音が十分に除去されていないことによって、音声成分が保存できているためであると考えられる。また、Yoon & Yooの方法と提案方法について比較してみると、図4.8より全てのノイズにおいて提案方法が従来方法と同等もしくは優れている特性を示すことがわかる。提案する判別方法はYoon & Yooの判別方法と比較して音声領域を雑音領域と誤判別する割合が少ないことにより、音声を雑音として扱うことが少なくなる。その結果、音声成分が保存され、ひずみが抑えられると考えられる。以上の結果から、提案方法がYoon & Yooの方法より処理した音声のひずみを抑えるのに有効であることがわかる。

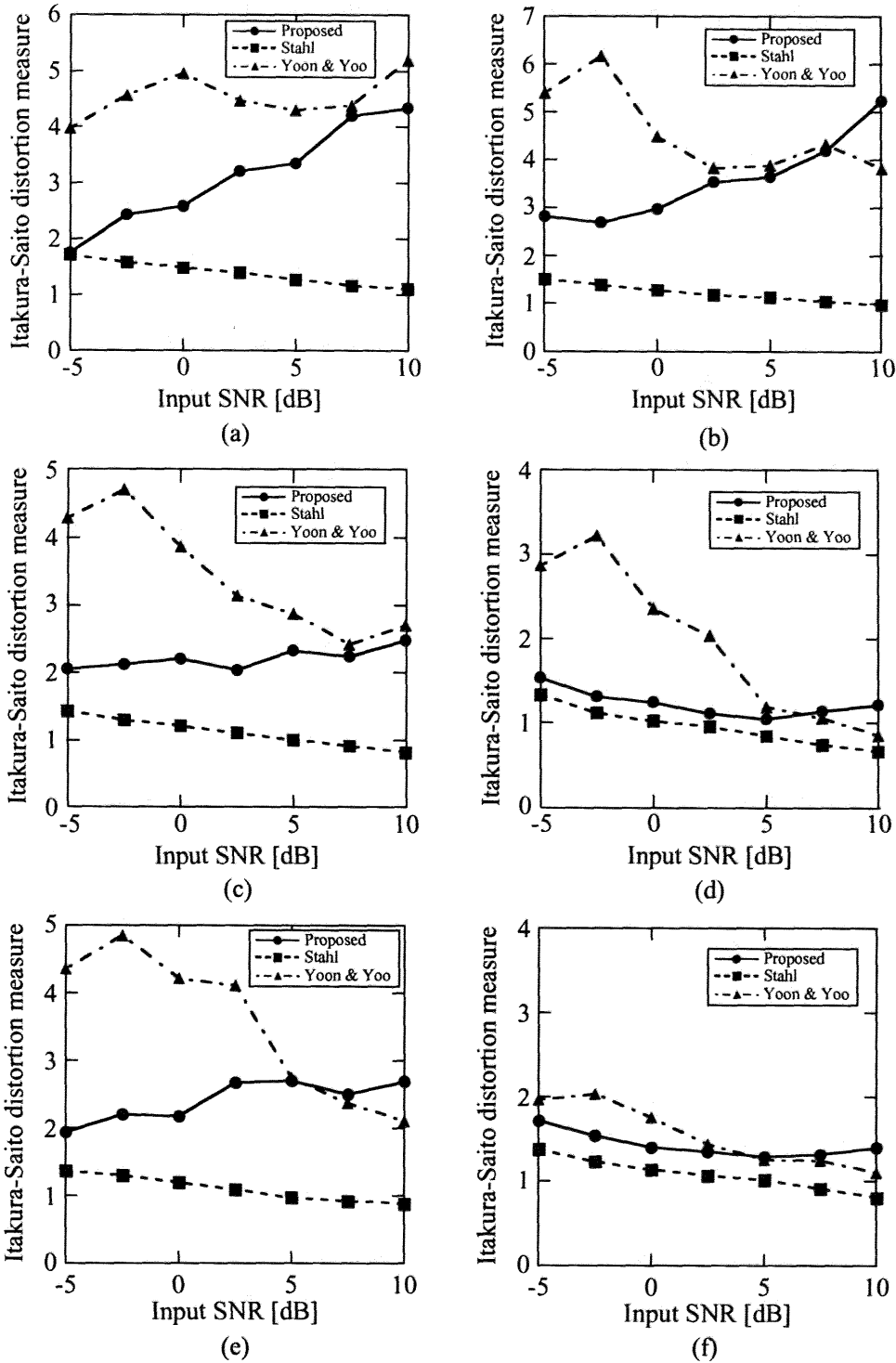


図 4.8 入力 SNR における板倉-斉藤ひずみ距離 ((a): ホワイトノイズ, (b): ピンクノイズ, (c): F16 コックピットノイズ, (d): バブルノイズ, (e): 計算機室雑音, (f): 交差点雑音)

Fig. 4.8 Itakura-Saito distortion measure for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

表 4.1 MOS テストの結果 (SNR=5 dB)

Table 4.1 Comparison of MOS tests for SNR=5 dB

	white	pink	F16	babble
proposed method	3.26	4.16	3.73	2.87
method by Stahl	1.60	2.13	2.50	2.17
method by Yoon & Yoo	2.93	2.83	2.83	2.87

#### 4.4.5 MOS テスト

試聴テストの1つである MOS テストを行い、提案方法について主観評価を行った。なお、被験者は10名とし、各被験者から得られたスコアの平均を取る。

表 4.1 に SNR=5 dB における MOS テストの結果を示す。表 4.1 より、Stahl らの方法で処理した音声は他の方法に比べてスコアが低い。一方、客観評価の結果では図 4.7 および図 4.8 より、音声のひずみについては Stahl らの方法が他の方法より優れているものの、ミュージカルノイズの低減については文献 Stahl らの方法よりも他の方法の方が優れている。これらのことから、Stahl らの方法で処理した音声にはミュージカルノイズが多く含まれているためにスコアが低かったと考えられる。次に、Yoon & Yoo の方法で処理した音声は Stahl らの方法によるものと比較して良いスコアになっている。これは Yoon & Yoo の方法が Stahl らの方法よりもミュージカルノイズが低減されているためであると考えられる。提案方法で処理した音声のスコアは、全ての雑音において従来方法に比べて良くなっている。雑音がホワイトノイズ、ピンクノイズおよび F16 コックピットノイズの場合は表 4.1 より、提案方法のスコアが Yoon & Yoo の方法より格段に良くなっている。これらの雑音は図 4.7(d) および図 4.8(d) より、ミュージカルノイズの低減および音声のひずみの両者について提案方法の方が優れている。したがって、提案方法によって処理した音声はミュージカルノイズおよび音声のひずみの発生が抑えられているため、スコアが格段に良くなったと考えられる。雑音がバブルノイズの場合は表 4.1 より、提案方法のスコアは Yoon & Yoo の方法と同等である。これは、客観評価の結果では図 4.7(d) および図 4.8(d) より、提案方法の音声のひずみおよびミュージカルノイズの低減が Yoon & Yoo の方法と比較してほぼ同等であるためだと考えられる。以上の結果から、主観的評価においても本章で提案した方法が雑音の事前情報を用いない1入力の雑音除去方法として有効であることが示される。

## 4.5 まとめ

本章では、複数の短時間フーリエ変換の周波数から構成されるバンドにおける観測信号のスペクトログラム上の特徴量に着目し、各バンド内の標準偏差を利用した音声領域と雑音領域の判別方法を提案した。音声領域と雑音領域の判別の後、それぞれの領域において異なるパラメータを用いたスペクトルサブトラクションを行った。性能評価の結果より、提案方法は従来方法と比較して正確に音声領域と雑音領域の判別を行えるため、処理した音声のミュージカルノイズやひずみを減少できることを示した。そして、提案方法が1入力システムの雑音除去方法として有効な方法であることを示した。

## 第4章の参考文献

- [1] V. Stahl, A. Fischer, and R. Bippusuchi, "Quantile based noise estimation for spectral subtraction and Wiener filtering," Proc. of IEEE ICASSP 2000, pp.1875-1878, Istanbul, Turkey, June 2000.
- [2] S. Yoon and C.D. Yoo: "Speech enhancement based on speech/noise-dominant decision," IEICE Trans. Inf. & Syst., vol.E85-D, no.4, pp.744-750, Apr. 2002.
- [3] 古井貞熙, 音響工学, 近代科学社, 東京, 1992.
- [4] 笠井聡, 藤田広志, 原武史, 畑中裕司, 遠藤登喜子, "マンモグラム上の腫瘍陰影自動検出アルゴリズムにおける索状の偽陽性候補陰影の削除," コンピュータ支援画像診断学会誌, vol.3, no.2, pp.1-7, Apr. 1999.
- [5] 長野智章, 伊東正安, "Morphology 演算を用いた可変ブロック法による医用超音波画像の領域分割," 電子情報通信学会論文誌 A, vol.J84-A, no.12, pp.1444-1451, Dec. 2001.
- [6] 呂建明, 藤本稔, 谷萩隆嗣, "ニューラルネットワークを用いた劣化画像の雑音除去," 電気学会論文誌 C, vol.122, no.8, pp.1301-1308, Aug. 2002.
- [7] A. El-Jaroudi and J. Makhoul, "Discrete All-Pole Modeling," IEEE Trans. on Signal Process., vol.39, no.2, pp.411-423, Feb. 1991.
- [8] Jr. J.R. Deller, J.Hansen, and J.G. Proakis, Discrete-time processing of speech signals, IEEE Press, New York, 2000.
- [9] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun., vol.12, no.3, pp.247-251, July 1993.

---

## 第5章

---

# モルフォロジー処理を用いた スペクトルサブトラクションに おけるミュージカルノイズ除去

### ●● 本章概要 ●●

本章では、モルフォロジー処理を用いたスペクトルサブトラクションにおけるミュージカルノイズ除去方法を提案する。提案方法では、スペクトログラム上でミュージカルノイズが孤立点として現れることに注目し、モルフォロジー処理が孤立点除去に向いていることを利用してミュージカルノイズの除去を行う。また、提案方法ではミュージカルノイズ検出の際のしきい値を必要とせず、かつモルフォロジー処理は比較演算のみで行える。そのため、提案方法は従来方法と比較してシステムの設計が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。性能評価の結果より、提案方法は従来方法に比べて少ない計算回数で雑音除去を行うことができ、かつミュージカルノイズ除去性能が優れていることを示す。

## 5.1 はじめに

音声の雑音除去方法の1つとして、スペクトルサブトラクション(SS) [1-5]が注目されている。SSは観測信号のスペクトルから雑音のスペクトルを推定し、それを観測信号のスペクトルから減算することにより雑音の除去を行う方法である。SSは1入力で行うことができ、計算が比較的簡単であるという利点がある。しかし、雑音の推定誤差などが原因で処理後の音声信号にミュージカルノイズが発生するという問題点がある [2, 4, 5]。

その問題点を解決する方法として、人間の聴覚特性を利用して雑音スペクトル減算時の係数調整を行う方法 [4] や、SSによって処理した音声のスペクトログラムを画像として捉え、後処理によってミュージカルノイズを除去する方法 [5] が提案されている。Viragの方法 [4] は係数調整が複雑であるという問題点がある。Gohらの方法 [5] では、ミュージカルノイズがスペクトログラム上で孤立点として現れることに着目し、16種類の方向窓を用いたミュージカルノイズ検出処理を行い、メディアンフィルタを用いてミュージカルノイズの除去を行う。しかし、この方法ではミュージカルノイズ検出処理を行う必要がある。さらに、16種類の方向窓で各窓内の分散を計算しなくてはならないため、計算回数が大きいという問題点もある。

本章では、モルフォロジー処理を用いたミュージカルノイズの除去方法を提案する。提案方法では、ミュージカルノイズがスペクトログラム上で孤立点として現れることに注目し、モルフォロジー処理の1つであるオープニングが孤立点除去に向いていることを利用してミュージカルノイズの除去を行う。また、提案方法ではミュージカルノイズ検出処理を必要とせず、かつモルフォロジー処理は比較演算のみで行える。そのため、提案方法は従来方法と比較してシステムの設計が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。提案方法について6種類の雑音による性能評価を行う。その結果、提案方法は従来方法に比べて少ない計算回数で雑音除去を行うことができ、かつミュージカルノイズの除去性能が優れていることを示す。



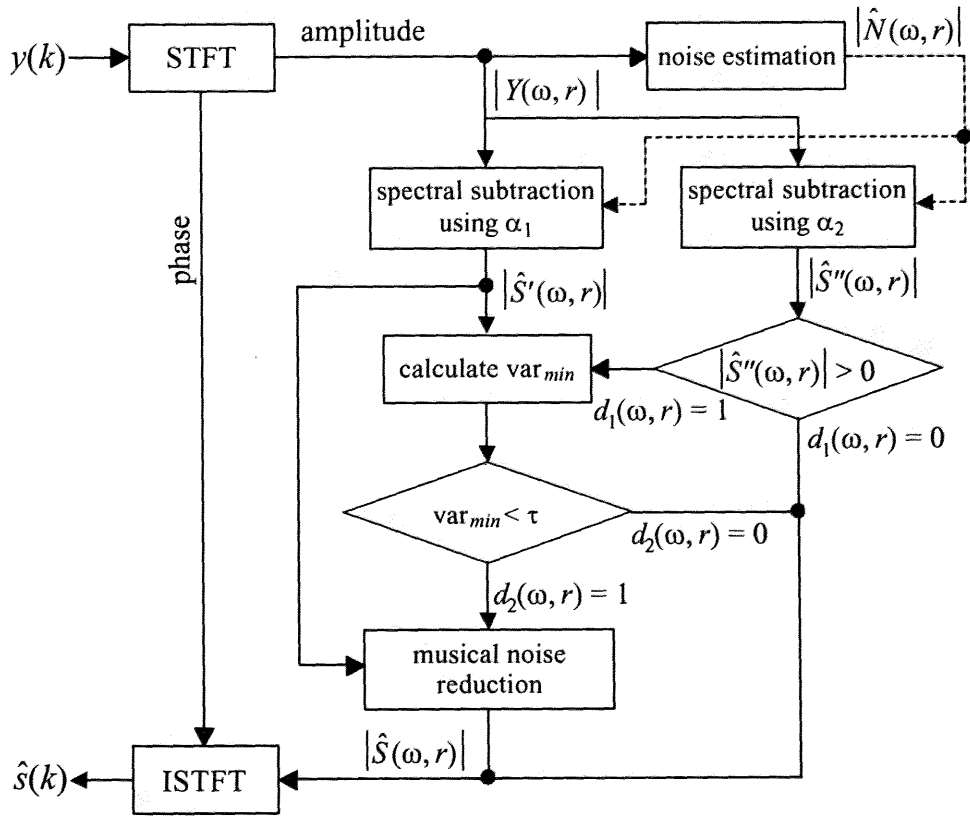


図 5.1 Goh らの方法の構成図

Fig. 5.1 Overall flow of the method by Goh.

## 5.2 従来方法

図 5.1 に Goh らの方法 [5] の構成図を示す。この方法では、SS によって処理した音声のスペクトログラム上でミュージカルノイズが孤立点として現れることを利用してミュージカルノイズの除去を行う。Goh らの方法におけるミュージカルノイズの除去は以下の手順によって行われる。

まず、2 種類のサブトラクション係数  $\alpha_1$  および  $\alpha_2$  を用いた SS をそれぞれ行い、処理した音声のスペクトル  $|\hat{S}'(\omega, r)|$  および  $|\hat{S}''(\omega, r)|$  を得る。ここで、 $\alpha_1$  は雑音が完全に除去されなくても音声成分が十分に保存される値に設定する。一方、 $\alpha_2$  はたとえ弱い音声成分が除去されても雑音の除去を十分に行うために  $\alpha_2 > \alpha_1$  とする。よって、 $|\hat{S}''(\omega, r)|$  は強い音声成分のみが残ることになる。つまり、 $|\hat{S}''(\omega, r)|$  を用い

て強い音声成分とそれ以外の領域とを以下のように判別できる。

$$d_1(\omega, r) = \begin{cases} 0 & \text{if } |\hat{S}''(\omega, r)| > 0 \\ 1 & \text{if } |\hat{S}''(\omega, r)| = 0 \end{cases} \quad (5.1)$$

$d_1(\omega, r) = 1$  である  $(\omega, r)$  において、 $|\hat{S}'(\omega, r)|$  は強い音声成分以外の領域、つまりミュージカルノイズ、弱い音声成分、もしくは無音領域であることを示している。弱い音声成分を保存しつつミュージカルノイズを除去するためには、 $d_1(\omega, r) = 1$  となる  $(\omega, r)$  における  $|\hat{S}'(\omega, r)|$  を用いてミュージカルノイズ検出処理を行う必要がある。そのため、スペクトログラム上における音声成分とミュージカルノイズの性質を利用したミュージカルノイズ検出処理を行う。図 5.2(a) に原音声（女性が「休み無く打ち寄せてはさっと引いていく白い波」と発声したもの）のスペクトログラム、図 5.2(b) に原音声に SNR=10 dB でホワイトノイズを付加したときの観測信号のスペクトログラム、図 5.2(c) に図 5.2(b) の観測信号に対して  $\alpha = 1.8$  を用いた SS によって処理した音声のスペクトログラムをそれぞれ示す。スペクトログラム上での音声成分は図 5.2(a) より、一定の周波数間隔でパワーの強い部分とほぼ 0 である部分とがストライプ状に存在していることがわかる。この構造のことを調波構造といい、有声音特有の構造である。一方、ミュージカルノイズは図 5.2(c) より孤立点として現れており、図 5.2(b) のように一様に分布している観測信号の雑音とは異なる性質を示していることがわかる。そこで、式 (5.1) において  $d_1(\omega, r) = 1$  となる  $(\omega, r)$  に対して、16 種類の方向窓を用意する。なお、窓の長さはミュージカルノイズより長く、音声成分より短くなるように設定する。次に、各方向窓の分散をそれぞれ求める。そして、16 種類の方向窓の中から最も分散の小さい方向窓を選択し、その分散を  $\text{var}_{min}$  とする。領域  $(\omega, r)$  が音声成分の一部である場合は、分散が最も小さい方向窓と同じ方向に音声成分があるため、窓内の要素はスペクトルのパワーの近い成分で構成される。そのため、 $\text{var}_{min}$  は小さくなる。一方、領域  $(\omega, r)$  がミュージカルノイズの一部である場合は、方向窓の長さをミュージカルノイズより長く設定しているため、窓内にミュージカルノイズに相当する孤立点のエッジが含まれる。そのため、 $\text{var}_{min}$  は大きくなる。以上のことから、 $\text{var}_{min}$  に対して適切なしきい値  $\tau$  を設定することにより、以下のようにミュージカルノイズを検出することが可能である。

$$d_2(\omega, r) = \begin{cases} 0 & \text{if } \text{var}_{min} < \tau \\ 1 & \text{if } \text{var}_{min} \geq \tau \end{cases} \quad (5.2)$$

式 (5.2) において,  $d_2(\omega, r) = 1$  となる  $(\omega, r)$  が最終的にミュージカルノイズまたは無音領域として検出される. ミュージカルノイズまたは無音領域として検出された点については, 分散の最も小さい方向窓においてメディアンフィルタを施すことにより, ミュージカルノイズが除去された音声を得ることができる.

しかし, この方法では 16 種類の方向窓が必要で, 各窓内の分散を計算しなくてはならないため計算回数が多いという問題点がある. さらに, ミュージカルノイズ検出処理の際に用いるしきい値  $\tau$  が小さすぎると弱い音声成分をミュージカルノイズと誤検出する割合が増加し, 処理後の音声はミュージカルノイズだけでなく弱い音声成分も除去される. また,  $\tau$  が大きすぎるとミュージカルノイズの未検出の割合が増加し, 処理後の音声にミュージカルノイズが残る. そのため,  $\tau$  の設定にはミュージカルノイズの除去と弱い音声成分の保存との間にトレードオフが存在し, 適切な  $\tau$  を設定することは困難である. 以上の問題点を解決するために, システムの設計が容易で, かつ計算回数の少ないミュージカルノイズの除去方法が必要である.

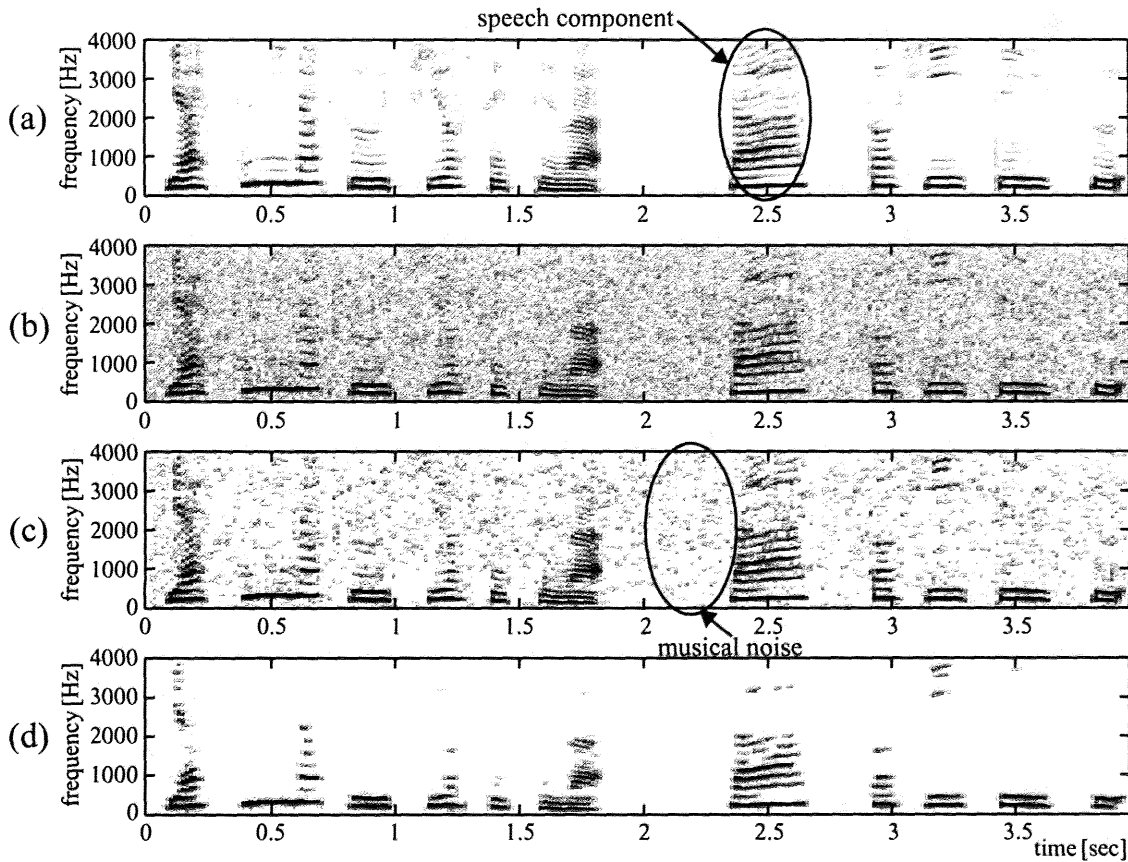


図 5.2 音声のスペクトログラム, (a) 原音声, (b) 観測信号 (ホワイトノイズ, SNR=10 dB), (c)  $\alpha = 1.8$  を用いた SS によって処理した音声, (d)  $\alpha = 16$  を用いた SS によって処理した音声

Fig. 5.2 Spectrograms of (a) Clean speech, (b) Observation signal (additive white noise with SNR=10 dB), (c) Enhanced speech by spectral subtraction for  $\alpha = 1.8$ , (d) Enhanced speech by spectral subtraction for  $\alpha = 16$ .

## 5.3 提案方法

### 5.3.1 提案方法の構成

図 5.3 に提案方法の構成図を示す。提案方法は Goh らの方法と同様に SS によって処理した音声に対して後処理によってミュージカルノイズの除去を行う。提案方法におけるミュージカルノイズの除去手順は以下の通りである。

- 1) Goh らの方法と同様に、2 種類のサブトラクション係数  $\alpha_1$  および  $\alpha_2$  を用いた SS をそれぞれ行い、処理した音声のスペクトル  $|\hat{S}'(\omega, r)|$  および  $|\hat{S}''(\omega, r)|$  をそれぞれ得る。
- 2) 音声成分の位置を示す 2 値データ  $m_s(\omega, r)$  およびミュージカルノイズと弱い音声成分の位置を示す 2 値データ  $m_n(\omega, r)$  を以下の式のように作成する。

$$m_s(\omega, r) = \begin{cases} 1 & \text{if } |\hat{S}''(\omega, r)| > 0 \\ 0 & \text{if } |\hat{S}''(\omega, r)| = 0 \end{cases} \quad (5.3)$$

$$m_n(\omega, r) = \begin{cases} 1 & \text{if } |\hat{S}'(\omega, r)| > 0 \text{ and } |\hat{S}''(\omega, r)| = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.4)$$

- 3) 式 (5.4) で求めた  $m_n(\omega, r)$  に対してモルフォロジー処理を行い、2 値データ  $m_m(\omega, r)$  を得る。
- 4)  $m_s(\omega, r)$  および  $m_m(\omega, r)$  を用いてミュージカルノイズ除去処理を以下のように行い、ミュージカルノイズ除去後の音声スペクトル  $|\hat{S}(\omega, r)|$  を得る。

$$|\hat{S}(\omega, r)| = \begin{cases} |\hat{S}'(\omega, r)| & \text{if } m_s(\omega, r) = 1 \text{ or } m_m(\omega, r) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.5)$$

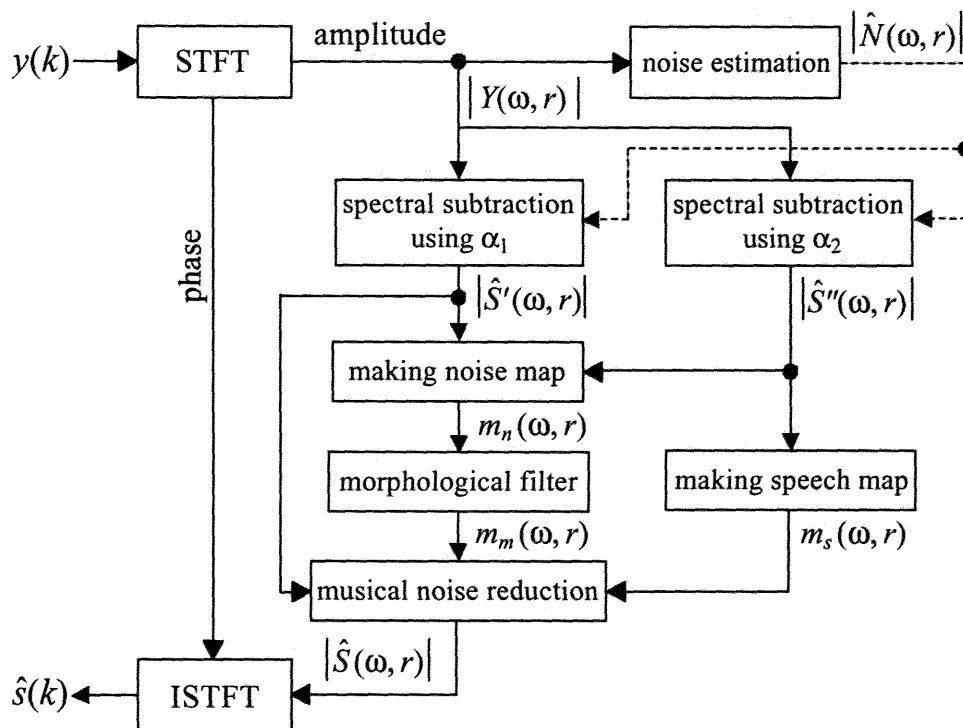


図 5.3 提案方法の構成図

Fig. 5.3 Overall flow of the proposed method.

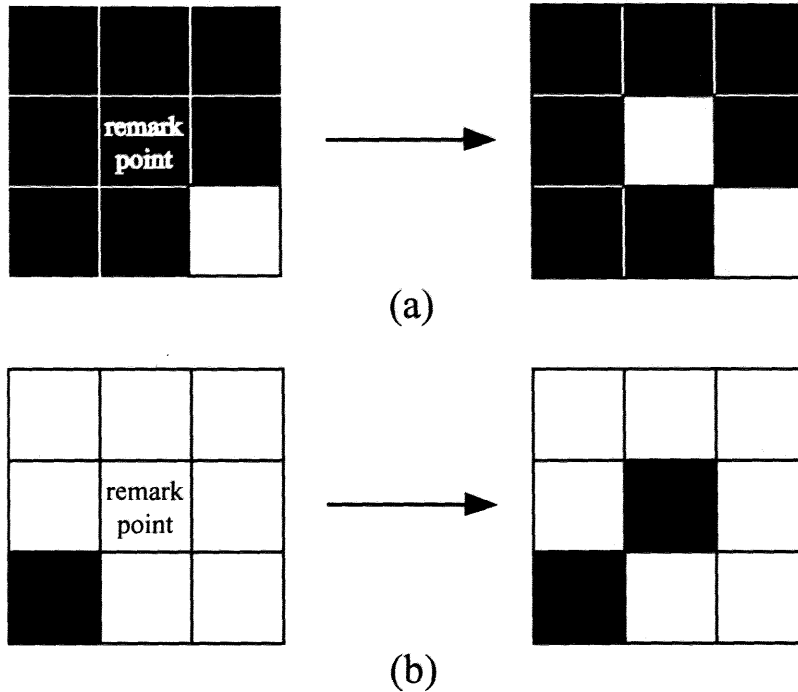


図 5.4 3 × 3 の窓によるモルフォロジー処理, (a) 収縮, (b) 膨張

Fig.5.4 Morphological process using 3 × 3 window for (a) erosion, (b) dilation.

### 5.3.2 モルフォロジー処理

音声成分を保存しつつミュージカルノイズを除去するためには、ミュージカルノイズと弱い音声成分のスペクトログラム上の位置を示す  $m_n(\omega, r)$  からミュージカルノイズと弱い音声成分とを判別する必要がある。提案方法では、両者のスペクトログラム上での性質の違いを利用し、モルフォロジー処理 [6] により両者の判別を行う。

モルフォロジー処理は2値画像処理の1つであり、収縮と膨張の2つの処理から成り立っている。図5.4(a)に3 × 3の正方形窓による収縮を示す。収縮は図5.4(a)のように、窓内の画素を調べて0（白）となる画素が1つでもある場合、注目画素を0に置き換える処理である。収縮によって孤立点や細かい線が除かれる。一方、図5.4(b)に3 × 3の正方形窓による膨張を示す。膨張は図5.4(b)のように、窓内の画素を調べて1（黒）となる画素が1つでもある場合、注目画素を1に置き換える処理である。モルフォロジー処理では収縮と膨張は一般的に組み合わせて用いられる。その中で、収縮－膨張の順番で処理を行うことをオープニングという。オープニングを施すことにより孤立点や細かい線が除去される。図5.5にモルフォロジー処理の例を

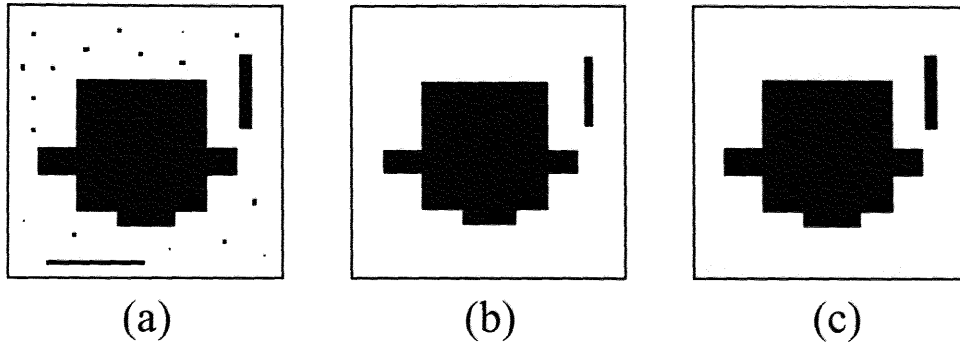


図 5.5 モルフォロジー処理の例, (a) 原画像, (b) 収縮, (c) オープニング  
 Fig. 5.5 Examples of morphological process (a) Original image, (b) Image after erosion, (c) Image after opening operation.

示す. 図 5.5(b) に図 5.5(a) に対して収縮を行った結果を示す. 図 5.5(c) に図 5.5(a) に対してオープニングを行った結果を示す. 図 5.5(b) より, 収縮を行うことにより, 図 5.5(a) 全体に存在する孤立点だけでなく, 図 5.5(a) 左下に存在する細い線が取り除かれていることがわかる. そして, 孤立点や細い線以外の領域は図 5.5(a) と比較して縮小されていることがわかる. 一方, オープニング処理を施した場合は図 5.5(c) より, 孤立点や細い線は除去されておりかつ, 孤立点や細い線以外の領域は図 5.5(a) と大きさが変わっていない. このことから, オープニングが孤立点除去に向いていることがわかる.

図 5.2(a) よりスペクトログラム上では音声成分はストライプ状の調波構造を有しているのに対し, 図 5.2(c) よりミュージカルノイズは孤立点として現れている. そのため,  $m_n(\omega, r)$  に対してオープニングを施すことで  $m_n(\omega, r)$  よりミュージカルノイズの領域を取り除くことができると考えられる.

モルフォロジー処理では, 一般的に図 5.4 のような  $a \times a$  の正方形窓を用いる. しかし, 正方形窓を用いた場合, 図 5.5(c) のように,  $a$  より狭い幅の線は除去される. そのため, 収縮時にスペクトログラム上でストライプ状に存在する音声成分が除去されてしまう可能性がある. そのため, 提案方法では  $1 \times b$  の長方形窓を用いることによって, 収縮時に音声成分が除去されることを防ぐ. 図 5.6 に本章で使用する長方形窓を示す. 従って,  $m_n(\omega, r)$  に対してオープニング処理を行った後の出力  $m_m(\omega, r)$  は

$$m_m(\omega, r) = (A \ominus B) \oplus B \quad (5.6)$$



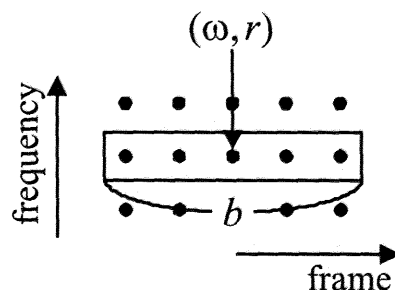


図 5.6 提案方法で用いる長方形窓

Fig. 5.6 Rectangle window using the proposed method.

$$A = \left\{ m_n(\omega, u) : r - \frac{b-1}{2} \leq u \leq r + \frac{b-1}{2} \right\} \quad (5.7)$$

となる。ただし、 $\oplus$  および  $\ominus$  は Mathematical Morphology の表記法に従い、それぞれ膨張および収縮を意味する。また、 $B$  は  $1 \times b$  の長方形窓を表す。 $b$  は小さすぎると窓内の全要素がミュージカルノイズのみで構成されるため、ミュージカルノイズが除去されない。一方、 $b$  が大きすぎると弱い音声成分が除去されてしまう。ここで、 $b$  は予備実験により決定する。本章では、予備実験の結果より  $b = 7$  とする。また、 $B$  については、Goh の方法と同様に方向窓を併用した場合についても検討を行った。その結果、方向窓を追加することにより長方形窓のみの場合よりも弱い音声成分が保存されるものの、オープニング処理による比較演算の回数が窓の種類に比例して増加する。さらに、窓の種類を増やすことによりミュージカルノイズの除去性能も低下する。そのため、本章では方向窓を併用した場合よりも弱い音声成分が保存されないものの、計算回数が少なく、かつミュージカルノイズの除去性能が高い長方形窓のみとした。

提案方法では、ミュージカルノイズがスペクトログラム上で孤立点として現れていることに注目し、オープニングが孤立点除去に向いていることを利用することにより、ミュージカルノイズの除去が行えると考えられる。また、提案方法ではミュージカルノイズ検出処理を必要としない。さらに、提案方法で用いているオープニングは  $2b$  回の比較演算で行える。そのため、提案方法は従来方法と比較してシステムの設計が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。

## 5.4 性能評価

### 5.4.1 シミュレーション条件

計算機シミュレーションにより提案方法の性能評価を行う。シミュレーション条件は以下のとおりである。STFTの分析フレーム長は $N = 256$ とし、 $1/4$ オーバーラップとする。また、窓関数にはハミング窓を用いる。サブトラクション係数 $\alpha_1$ および $\alpha_2$ はGohらの方法と同様にそれぞれ $\alpha_1 = 1.8$ ,  $\alpha_2 = 16$ とする。また、原音声は研究用連続音声データベース（日本音響学会）の男性話者4名（can0001, can0002, mit0001, mit0002）、女性話者4名（can1001, can1002, mit1001, mit1002）による発話文25文（a01～a25）の計200文である。雑音はNOISEX-92データベース [7]よりホワイトノイズ、ピンクノイズ、F16コックピットノイズ、バブルノイズの4種類の雑音、電子協騒音データベース（日本電子工業振興協会）より交差点雑音および計算機室（中型）雑音の2種類の計6種類を用いる。なお、原音声および雑音のサンプリング周波数は8 kHz、量子化レベルは16 bitである。本章では、比較対象としてViragの方法、Gohらの方法、およびSSを用いる。なお、SSの $\alpha$ は予備実験の結果より、 $\alpha = 4.0$ とする。性能評価は計算回数、スペクトログラムの比較、segmental SNRの改善度およびMOSテスト [8]の4つの方法により行う。

### 5.4.2 計算回数の比較

表5.1に領域 $(\omega, r)$ におけるGohらの方法と提案方法の計算回数を示す。ここで、MPYは乗算回数、ADDは加算回数、CMPは比較演算の回数、ANDはANDを取った回数、ORはORを取った回数である。表5.1より、Gohらの方法と提案方法の計算回数を比較すると、提案方法の計算回数が低減しており、特に加算と乗算の回数が大幅に低減している。これは、Gohらの方法では16種類の方向窓において分散の計算をそれぞれ行う必要があるためである。以上のことから、提案方法はGohらの方法よりも少ない計算回数で実現できることがわかる。

表 5.1 計算回数の比較

Table 5.1 Comparison of computational efforts.

Method	Process	MPY	ADD	CMP	AND	OR
Method by Goh	Speech component detection	0	0	1	0	0
	Musical noise detection	256	176	17	0	0
	Musical noise reduction	0	0	22	0	0
Proposed method	Making speech map	0	0	1	0	0
	Making noise map	0	0	2	1	0
	Morphology process	0	0	14	0	0
	Musical noise reduction	0	0	1	0	1

### 5.4.3 ミュージカルノイズの除去および弱い音声成分の保存についての検討

図 5.7(a) に図 5.2(a) と同一の原音声のスペクトログラム, 図 5.7(b) に図 5.7(a) の原音声に SNR=10 dB でホワイトノイズを付加したときの観測信号のスペクトログラム, 図 5.7(c) にミュージカルノイズの位置を示す 2 値データ  $m_{MN}(\omega, r)$  をプロットしたものを示す. なお,  $m_{MN}(\omega, r)$  は以下のように作成した.

$$m_{MN}(\omega, r) = \begin{cases} 1 & \text{if } |\hat{S}'(\omega, r)| > 0 \text{ and } |S(\omega, r)| = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.8)$$

図 5.7(c) より, ミュージカルノイズが孤立点として現れていることがわかる. また, 図 5.7(d) に  $m_{MN}(\omega, r)$  に対して,  $1 \times 7$  の長方形窓によってオープニング処理を行った結果を示す. 図 5.7(c) および (d) より, 孤立点として現れているミュージカルノイズがほとんど除去されていることがわかる. 図 5.7(e) に弱い音声の位置を示す 2 値データ  $m_{WS}(\omega, r)$  をプロットしたものを示す. なお,  $m_{WS}(\omega, r)$  は以下のように作成した.

$$m_{WS}(\omega, r) = \begin{cases} 1 & \text{if } |S(\omega, r)| > 0 \text{ and } |\hat{S}''(\omega, r)| = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.9)$$

図 5.7(a) および (e) より, 弱い音声成分は有声音の部分と無声音の部分で特徴が異なることがわかる. 有声音の部分については調波構造に相当する弱い音声成分が水平もしくは斜めの線として現れている. 一方, 無声音の部分にみられる弱い

音声成分は有声音のように特徴はなく、ランダムに存在していることがわかる。図 5.7(f) に図 5.7(e) に対して、 $1 \times 7$  の長方形窓によってオープニング処理を行った結果を示す。さらに、図 5.7(g) に図 5.7(e) と図 5.7(f) の差分、すなわちオープニング処理によって除去された部分を示す。図 5.7(e) および (f) より、有声音の調波構造のうち、水平方向のものは長方形窓と同じ方向であるため保存されていることがわかる。しかし、図 5.7(e) および (g) より、調波構造の斜め方向の成分は除去されていることがわかる。さらに、図 5.7(a),(e) および (g) より、有音区間と無音区間とが切り替わる部分において音声成分が除去されていることがわかる。これは、オープニング処理が窓内の要素に 1 つでも 0 が含まれる場合、領域  $(\omega, r)$  を 0 とするためである。

図 5.8(a) に  $m_{MN}(\omega, r)$  に対してオープニング処理を行った際の除去率を示す。図 5.8(a) より、除去率がほぼ 100% を示していることがわかる。これは、図 5.7(c) および (d) より、孤立点として現れているミュージカルノイズがほとんど除去されているためである。図 5.8(b) に  $m_{WS}(\omega, r)$  に対してオープニング処理を行った際の保存率を示す。図 5.8(b) より、除去率が SNR=10 dB で約 70% を示していることがわかる。これは、図 5.7(e) および (g) より、調波構造の斜め方向の成分だけでなく、有音区間と無音区間とが切り替わる部分においても音声成分が除去されているためだと考えられる。

以上のことから、提案方法では長方形窓によるモルフォロジー処理によって、スペクトログラム上で孤立点として現れているミュージカルノイズを除去できることがわかる。また、弱い音声成分については、斜め方向の成分が除去されるものの、水平方向の有声音の弱い音声成分を保存できることがわかる。

#### 5.4.4 スペクトログラムおよび segmental SNR の改善度

図 5.9(a) に図 5.7(b) の観測信号に対して  $\alpha = 4.0$  を用いた SS によって強調した音声のスペクトログラム、図 5.9(b) に Virag の方法によって強調した音声のスペクトログラム、図 5.9(c) に Goh らの方法によって強調した音声のスペクトログラム、図 5.9(d) に提案方法によって強調した音声のスペクトログラムをそれぞれ示す。また、図 5.10 に図 5.9 のスペクトログラムのうち 1.5 秒から 2.5 秒の部分を拡大表示し

たものを示す。  $\alpha = 4.0$  を用いた SS で強調した音声は図 5.9(a) および図 5.10(a) の丸く囲った部分より、音声成分が十分に保存されていることがわかる。しかし、図 5.10(a) の四角で囲った部分のように、雑音除去が十分でないためミュージカルノイズが発生していることがわかる。Virag の方法は図 5.10(b) より、  $\alpha = 4.0$  を用いた SS と同様に音声成分が保存されているものの、雑音スペクトルの引き残しがみられる。これは、Virag の方法の雑音スペクトル減算時の係数調整が聴覚特性に基づいて雑音を引きすぎないように行われているためである。Goh らの方法では図 5.10(c) より、SS と同様に音声成分が保存されているものの、ミュージカルノイズの除去性能は十分とは言えない。これは、ミュージカルノイズ検出処理に用いるしきい値  $\tau$  の設定にはミュージカルノイズの除去と弱い音声成分の保存との間にトレードオフが存在し、適切な  $\tau$  を設定することは困難であるためだと考えられる。一方、提案方法は図 5.10(d) より、ミュージカルノイズが十分に除去されていることがわかる。これは図 5.7 より、提案方法ではモルフォロジー処理によってスペクトログラム上で孤立点として現れているミュージカルノイズを除去できるためである。

図 5.11 に従来方法と提案方法における segmental SNR の改善度のグラフを示す。図 5.11 より、提案方法が各方法の中でも最も優れていることがわかる。これは、図 5.7、図 5.8 および図 5.10(d) より、提案方法ではモルフォロジー処理によってスペクトログラム上で孤立点として現れているミュージカルノイズを除去できているためである。

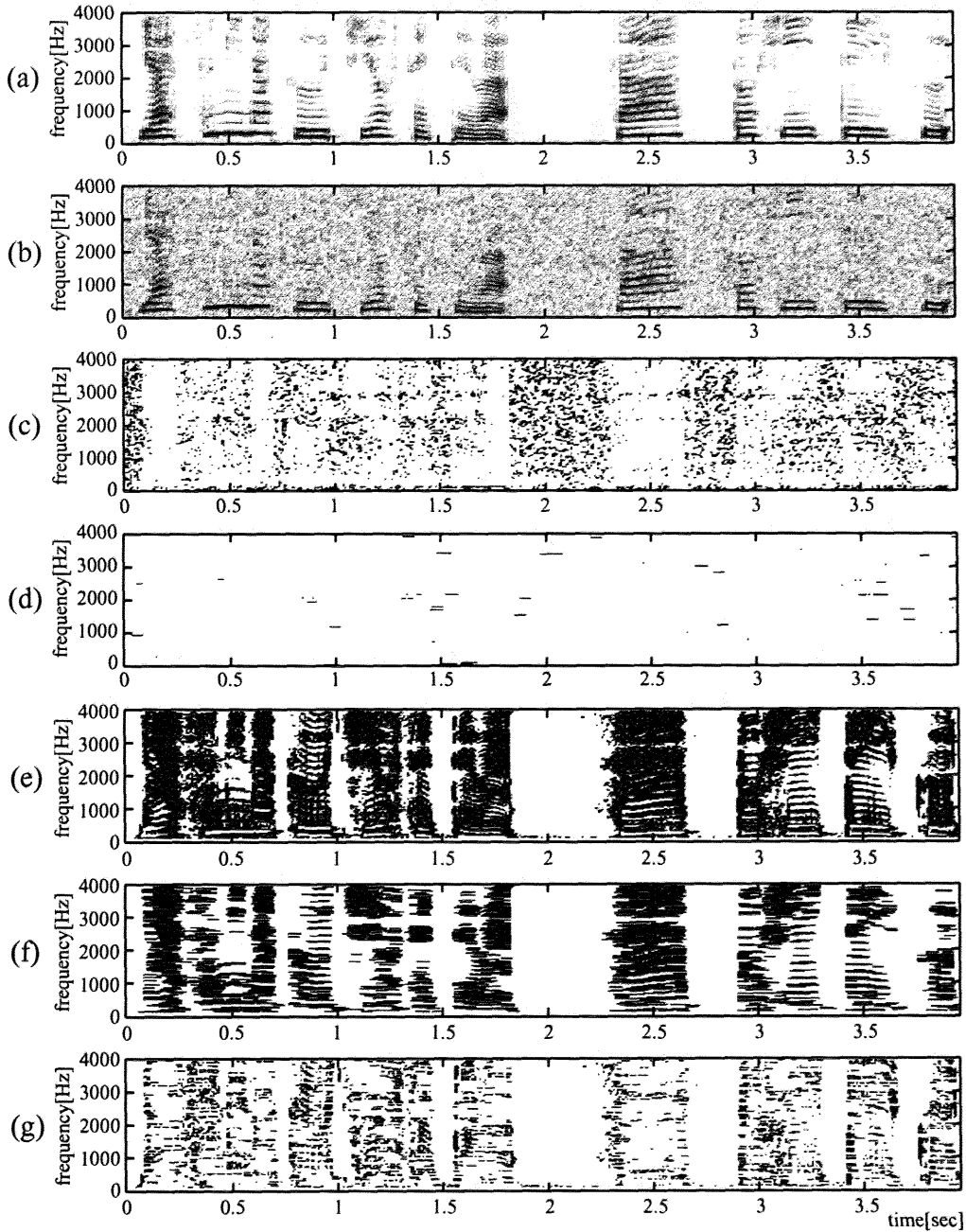


図 5.7 モルフォロジー処理の結果 ((a) 原音声のスペクトログラム, (b) 観測信号のスペクトログラム (ホワイトノイズ, SNR=10 dB), (c)  $m_{MN}(\omega, r)$ , (d)  $m_{MN}(\omega, r)$  に対するオープニング処理結果, (e)  $m_{WS}(\omega, r)$ , (f)  $m_{WS}(\omega, r)$  に対するオープニング処理結果, (g) 差分表示

Fig. 5.7 Spectrogram of (a) clean speech and (b) observation signal (additive white noise with SNR=10 dB). (c) The plot of  $m_{MN}(\omega, r)$ , (d) Result of opening operation for  $m_{MN}(\omega, r)$ , (e) The plot of  $m_{WS}(\omega, r)$ , (f) Result of opening operation for  $m_{WS}(\omega, r)$ , (g) Difference map.

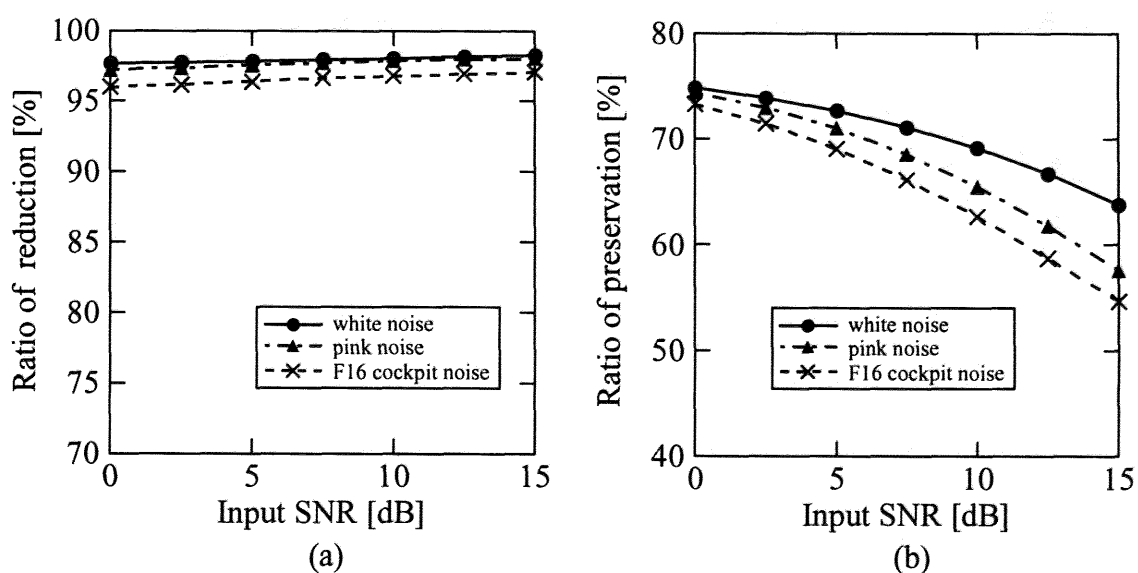


図 5.8 入力 SNR における (a) ミュージカルノイズ除去率, (b) 弱い音声成分の保存率

Fig. 5.8 (a) Ratio of musical noise reduction, (b) Ratio of weak speech preservation for Input SNR.

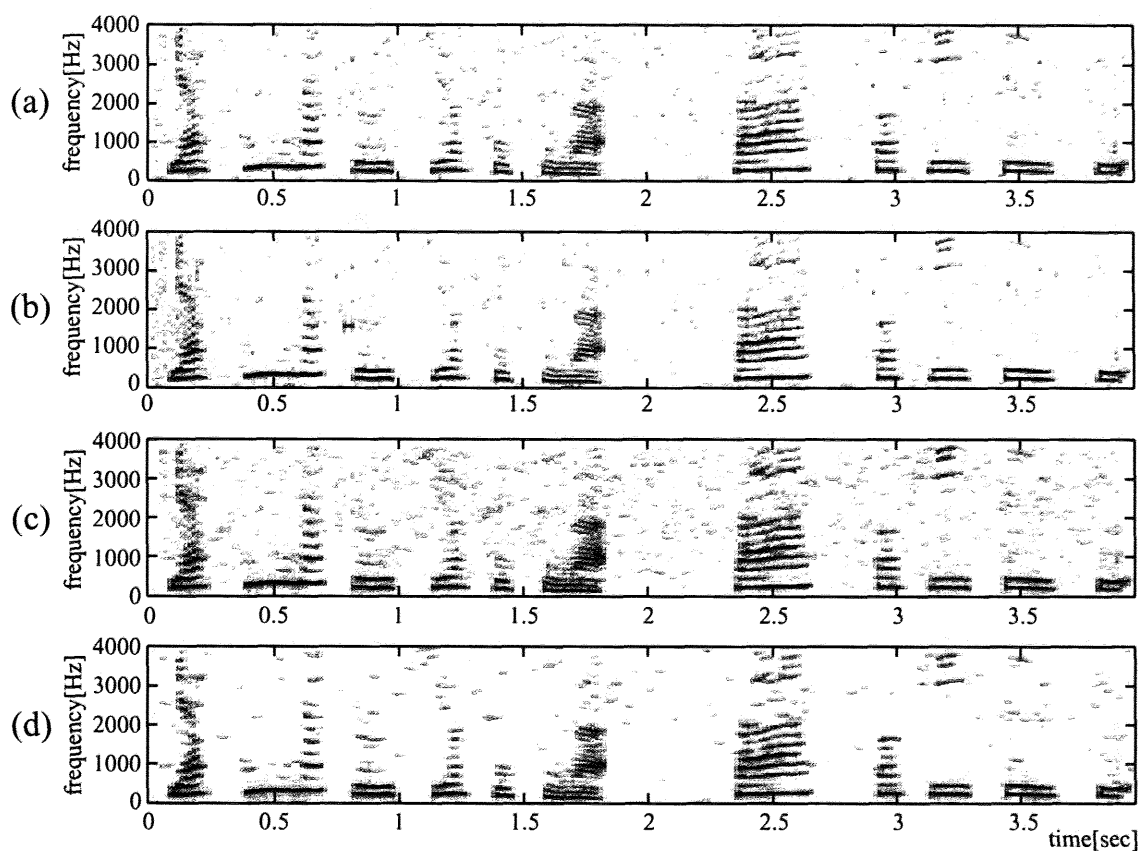


図 5.9 音声のスペクトログラム, (a) $\alpha = 4.0$ を用いたSSによって強調した音声, (b)Viragの方法を用いて強調した音声, (c)Gohらの方法を用いて強調した音声, (d)提案方法を用いて強調した音声

Fig. 5.9 Spectrograms of (a) Enhanced speech by spectral subtraction for  $\alpha = 4.0$ . (b) Enhanced speech by the method by Virag, (c) Enhanced speech by the method by Goh, (d) Enhanced speech by the proposed method.



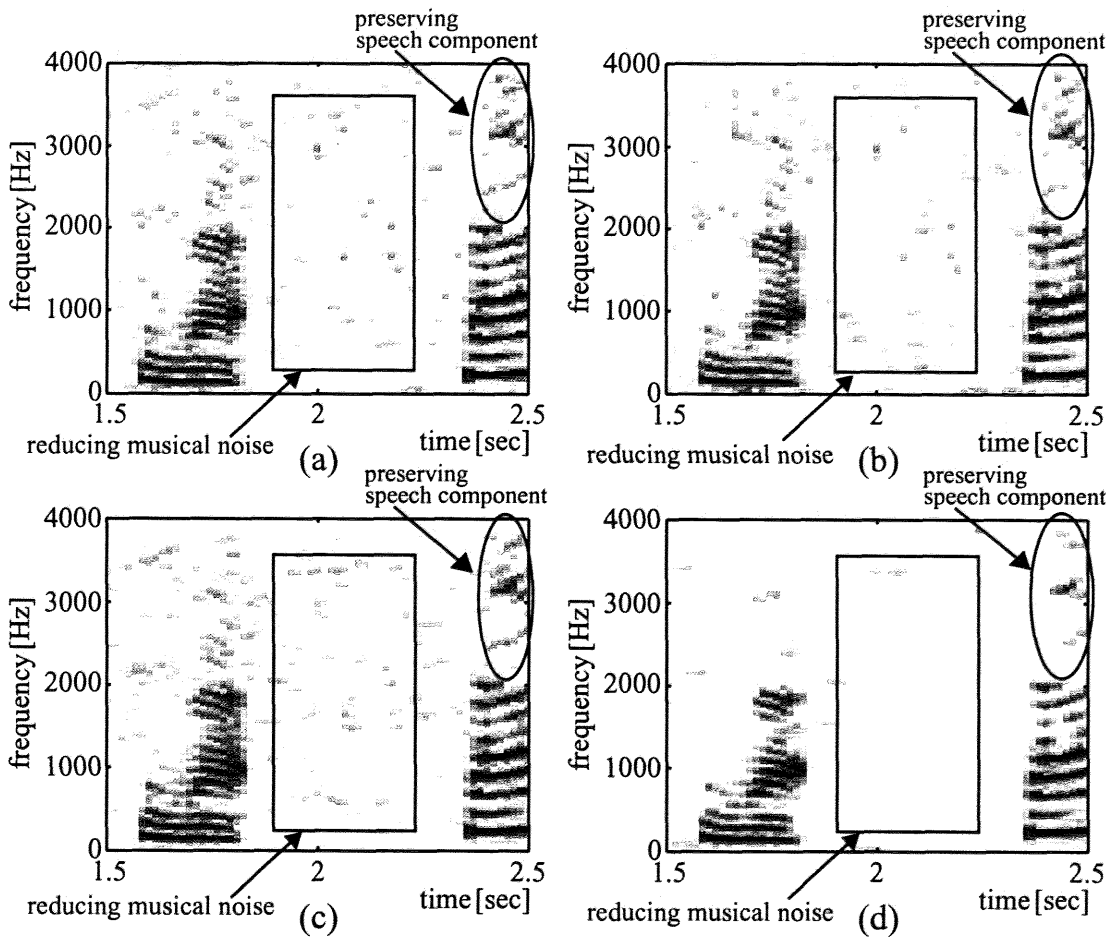


図 5.10 音声のスペクトログラム, (a) $\alpha = 4.0$ を用いたSSによって強調した音声, (b)Viragの方法を用いて強調した音声, (c)Gohらの方法を用いて強調した音声, (d)提案方法を用いて強調した音声

Fig. 5.10 Spectrograms of (a) Enhanced speech by spectral subtraction for  $\alpha = 4.0$ . (b) Enhanced speech by the method by Virag, (c) Enhanced speech by the method by Goh, (d) Enhanced speech by the proposed method.

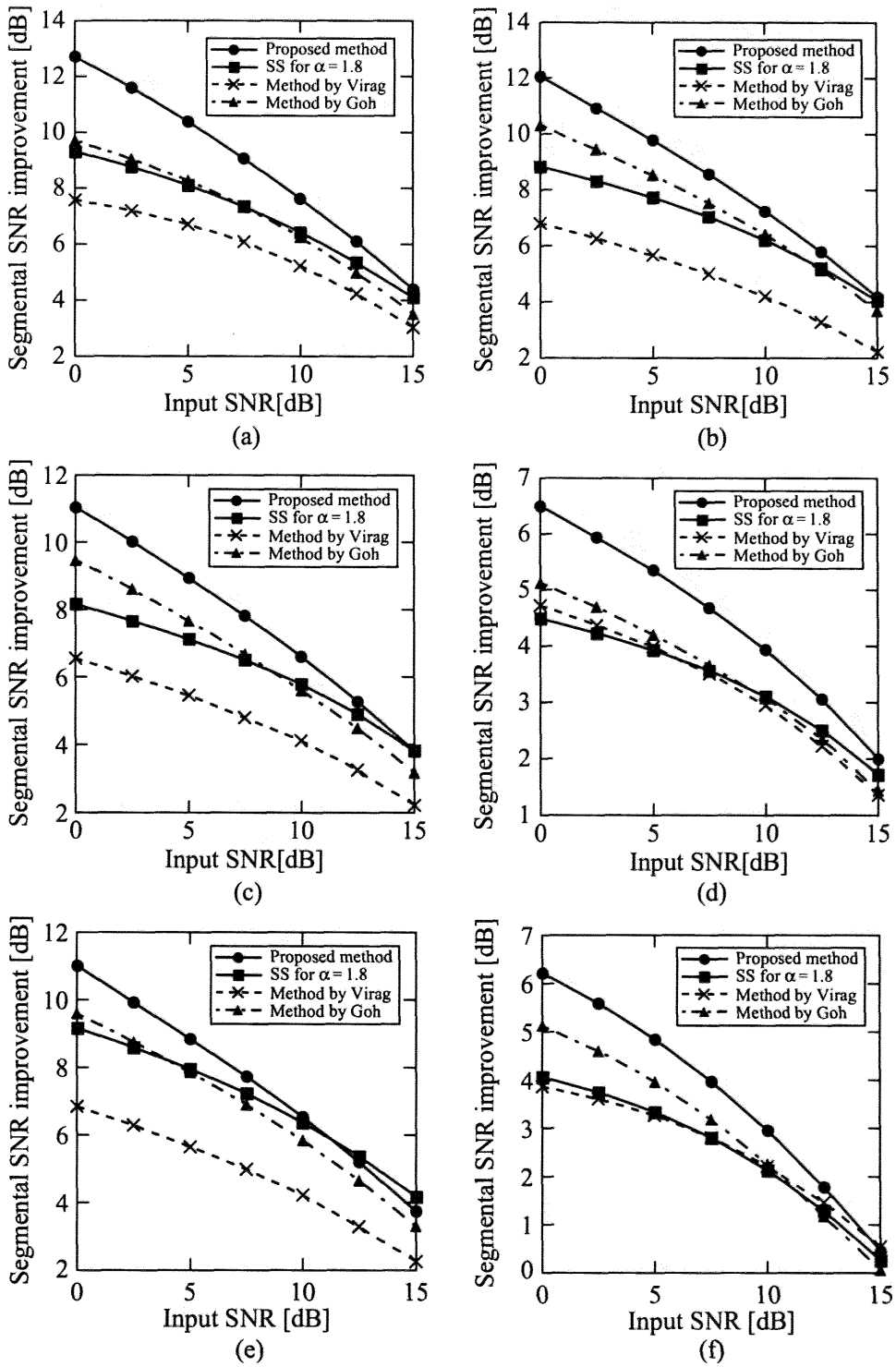


図 5.11 入力 SNR における segmental SNR の改善度 ((a) : ホワイトノイズ, (b) : ピンクノイズ, (c) : F16 コックピットノイズ, (d) : バブルノイズ, (e) : 計算機室雑音, (f) : 交差点雑音)

Fig. 5.11 Segmental SNR improvement for (a) White noise, (b) Pink noise, (c) F16 cockpit noise, (d) Babble noise, (e) Computer room noise, (f) Cross noise.

表 5.2 MOS テストの結果 (SNR=5, 10 dB)

Table 5.2 Comparison of MOS tests.

SNR [dB]	method	White	Pink	F16	Babble
5	SS for $\alpha = 4$	2.11	2.78	2.81	2.33
	Method by Virag	1.78	2.44	2.67	2.45
	Method by Goh	2.45	2.95	3.09	2.41
	Proposed method	3.24	3.51	3.71	2.92
10	SS for $\alpha = 4$	2.78	3.39	3.38	2.89
	Method by Virag	2.57	3.14	3.18	3.03
	Method by Goh	3.17	3.52	3.48	3.06
	Proposed method	3.87	4.19	4.22	3.59

### 5.4.5 MOS テスト

試聴テストの1つである MOS テストを行い、提案方法について主観評価を行った。MOS テストに使用した音声信号は研究用連続音声データベース（日本音響学会）のうち、男声 (can0001\_a22, can0002\_a13, can0002\_a18, mit0001\_a01, mit0002\_a03) および、女声 (can1001\_a16, can1001\_a24, can1002\_a25, mit1001\_a21, mit1002\_a05) の計 10 文である。なお、被験者は 10 名とし、各被験者から得られたスコアの平均を取る。

表 5.2 に SNR=5, 10 dB における MOS テストの結果を示す。表 5.2 より、提案方法で処理した音声のスコアはすべての雑音において従来方法に比べて良くなっている。提案方法のスコアの上昇は、図 5.9, 図 5.10 および図 5.11 より、処理した音声は音声成分を保存しつつミュージカルノイズを除去しているためであると考えられる。このことから、主観評価においても提案方法がミュージカルノイズの除去に対して有効であることが言える。

以上の結果より、提案方法は計算回数およびミュージカルノイズ除去性能の両面から従来方法より有効であることが言える。

## 5.5 まとめ

本章では、モルフォロジー処理を用いたミュージカルノイズの除去方法を提案した。提案方法では、ミュージカルノイズがスペクトログラム上で孤立点として現れることに注目し、モルフォロジー処理の1つであるオープニングが孤立点除去に向いていることを利用してミュージカルノイズの除去を行う。また、提案方法ではミュージカルノイズ検出の際のしきい値を必要とせず、かつモルフォロジー処理は比較演算のみで行える。そのため、提案方法は従来方法と比較してシステム的设计が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。性能評価の結果より、提案方法は従来方法に比べて少ない計算回数で雑音除去を行うことができ、かつミュージカルノイズの除去性能が優れていることを示した。

## 第5章の参考文献

- [1] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol.ASSP-27, no.2, pp.113-120, Apr. 1979.
- [2] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," Speech Commun., vol.11, no.2-3, pp.215-228, June 1992.
- [3] V. Stahl, A. Fischer, and R. Bippusuchi, "Quantile based noise estimation for spectral subtraction and Wiener filtering," Proc. of IEEE ICASSP 2000, pp.1875-1878, Istanbul, Turkey, June 2000.
- [4] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Process., vol.7, no.2, pp.126-137, Mar. 1999.
- [5] Z. Goh, K.C. Tan and T.G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Trans. Speech and Audio Process., vol.6, no.3, pp.287-292, May 1998.
- [6] 小畑秀文, モルフォロジー, コロナ社, 東京, 1996.
- [7] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun., vol.12, no.3, pp.247-251, July 1993.
- [8] Jr. J.R. Deller, J.Hansen, and J.G. Proakis, Discrete-time processing of speech signals, IEEE Press, New York, 2000.

---

## 第6章

---

### 結論

近年の携帯電話の普及や音声の符号化、音声認識システムの発展に伴い、デジタル音声信号に含まれる雑音を低減する技術が望まれている。本論文では、このような背景と要求から、雑音混入音声の音質改善を目的として、特に適応ノイズキャンセラおよびスペクトルサブトラクションについての研究を行った。

第1章は研究の背景と目的、および本論文の概要について述べている。第2章はパスが非線形特性を有する場合に対応するために、並列リカレントニューラルフィルタを利用した適応ノイズキャンセラを提案している。第3章および第4章は音声領域と雑音領域の判別を利用したスペクトルサブトラクションに関する提案である。第5章はスペクトルサブトラクションにおける後処理によるミュージカルノイズ低減方法に関する提案である。本研究の主要な研究成果を以下に列挙する。

(1) 第2章では、非線形特性のパスを持ったシステムに対応できる適応フィルタとして並列リカレントニューラルネットフィルタ (PRNF) を利用した適応ノイズキャンセラを提案した。PRNF はリカレントニューラルネットワークを用いたリカレントニューラルフィルタを多重分割して並列化したものあり、これによりフィルタの計算量の削減を図った。さらに、PRNF の学習に学習係数を動的に変化させる方法を使用して学習を安定させ、かつ収束を速めることにより、全体として計算回数の削減を図った。計算機シミュレーションの結果より、提案方法がパスの特性が線形・非線形にかかわらず十分に雑音が除去できることを示した。また、提案方法と線形適応フィルタの LMS 法、非線形システムに対応した適応フィルタであるニューラルフィルタおよび適応 Volterra フィルタとの比較を行い、提案方法が非線形特性を持ったパスの雑音除去に有効な方法の1つであることを示した。

(2) 第3章では、雑音量に依存しない音声領域と雑音領域の判別方法を利用したスペ

クトルサブトラクションを提案した。提案方法では、音声領域と雑音領域の判別のしきい値を雑音量によって適応的に変化させることにより、判別時の雑音の影響を低減させるため、雑音量に依存しない判別が可能となる。性能評価の結果より、提案方法は従来方法より音声領域と雑音領域の判別が雑音量に依存せず、正確に行われ、提案方法によって処理した音声は雑音除去性能を維持しながら音声ひずみを減少できることを示した。さらに、提案方法によって処理した音声のひずみは雑音量に関係なくほぼ一定であることを示した。

(3) 第4章では、雑音の事前情報を用いない音声領域と雑音領域との判別方法として、複数の短時間フーリエ変換の周波数から構成されるバンドにおける観測信号のスペクトログラム上の特徴量に着目し、各バンド内の標準偏差を利用した音声領域と雑音領域との判別方法を提案した。音声領域と雑音領域の判別の後、それぞれの領域において異なるパラメータを用いたスペクトルサブトラクションを行った。性能評価の結果より、提案方法が従来方法より処理した音声のミュージカルノイズや音声ひずみを減少できることを示した。そして、提案方法が雑音の事前情報を用いない1入力システムの雑音除去方法として有効な方法であることを示した。

(5) 第5章では、モルフォロジー処理を用いたミュージカルノイズの除去方法を提案した。提案方法では、ミュージカルノイズがスペクトログラム上で孤立点として現れることに注目し、モルフォロジー処理の1つであるオープニングが孤立点除去に向いていることを利用してミュージカルノイズの除去を行う。また、提案方法ではミュージカルノイズ検出の際のしきい値を必要とせず、かつモルフォロジー処理は比較演算のみで行える。そのため、提案方法は従来方法と比較してシステムの設計が容易で、かつ少ない計算回数でミュージカルノイズの除去が行える。性能評価の結果より、提案方法は従来方法に比べて少ない計算回数で雑音除去を行うことができ、かつミュージカルノイズの除去性能が優れていることを示した。

以上述べたように、本論文では雑音混入音声の音質改善を実現するために、特に適応ノイズキャンセラおよびスペクトルサブトラクションについての研究を行った。上記提案技術を適用することにより、従来方法よりも雑音混入音声の品質改善を図ることが可能となる。したがって、これらの提案技術が今後のデジタル音声処理技術のさらなる発展に貢献することを確信している。

提案した雑音除去方式の実用性をより高めるためには、非定常雑音、特にドアの開閉音や足音などの一般の生活環境に多く存在する、突発的に発生する雑音に対応できる方式を提案する必要があると考える。このことにより、提案方法の有用性はさらに向上し、提案方法が携帯電話や音声認識システムの前処理などのように現段階で早急に雑音除去方式を必要とする分野のみならず、現在研究が進められている様々な音声処理技術の将来的な実用化にも大きく貢献できるであろうと期待できる。



## 謝辞

本論文の執筆にあたり，御懇切な御指導，御鞭撻，ならびに様々な御配慮を賜った千葉大学大学院自然科学研究科情報科学専攻 谷萩隆嗣教授，呂建明助教授，関屋大雄助手に心より感謝し，厚く御礼を申し上げます。また，日頃から大変有意義な御助言を頂きました千葉大学工学部都市環境システム学科 山本一雄助手にも深く感謝致します。さらに，本論文をまとめるにあたり多大なる御助言，御討論を頂いた千葉大学大学院自然科学研究科情報科学専攻 市川熹教授，千葉大学フロンティアメディカル工学研究開発センター 蜂屋弘之教授に深く感謝致します。

また，MOSテストやプログラム開発等の際にご協力頂いた千葉大学谷萩・呂・関屋研究室の大学院生，学部生，研究生の方々に深く感謝致します。

最後に，長年私を支えて下さった家族に心より感謝申し上げます。

2006年1月 野村行弘

# 本研究に関する参考資料

## 本研究に関連する原著論文

- [1] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “リカレントニューラルネットワークを利用したノイズキャンセラの一設計法,” 信号処理, vol.6, no.6, pp.401-410, Nov. 2002.
- [2] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “観測信号のスペクトログラム上の特徴量に基づく音声領域と雑音領域との判別を用いた音声強調,” 電気学会論文誌 C, vol.124-C, no.11, pp.2310-2319, Nov. 2004.
- [3] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “雑音量に依存しない音声領域と雑音領域の判別法を利用した音声強調の改良,” 日本音響学会誌, vol.62, no.1, pp.12-22, Jan. 2006.
- [4] 野村 行弘, 斗澤 秀亮, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “モルフォロジー処理を用いたスペクトルサブトラクションにおけるミュージカルノイズ除去,” 電子情報通信学会論文誌 D (採録決定).

## 本研究に関連する国際会議

- [1] Yukihiro Nomura, Jianming Lu, Hiroo Sekiya and Takashi Yahagi, “A design method for noise canceller using parallel recurrent neural filters,” 2002 International Symposium on Nonlinear Theory and its Applications (NOLTA2002), pp.1005-1008, Xi' an, China, Oct. 2002. (発表者：野村 行弘)

- [2] **Yukihiro Nomura**, Jianming Lu, Hiroo Sekiya and Takashi Yahagi, "Spectral subtraction based on speech/noise-dominant classification," 2003 International Workshop on Acoustic Echo and Noise Control (IWAENC2003), pp.127-130, Kyoto, Japan, Sep. 2003. (発表者：野村 行弘)
- [3] **Yukihiro Nomura**, Jianming Lu, Hiroo Sekiya and Takashi Yahagi, "Speech/noise-dominant decision regardless of SNR for speech enhancement," 2004 RISP International Workshop on Nonlinear Circuits and Signal Processing (NCSP'04), pp.415-418, Hawaii, USA, Mar. 2004. (発表者：野村 行弘)
- [4] **Yukihiro Nomura**, Hideaki Tozawa, Noritaka Yamashita, Jianming Lu, Hiroo Sekiya and Takashi Yahagi, "Musical noise reduction by spectral subtraction using morphological filter," 2005 RISP International Workshop on Nonlinear Circuits and Signal Processing (NCSP'05), pp.415-418, Hawaii, USA, Mar. 2005. (発表者：野村 行弘, Recieved NCSP'05 sutudent paper award)
- [5] **Yukihiro Nomura**, Jianming Lu, Hiroo Sekiya and Takashi Yahagi, "Quantile based speech/noise-dominant decision for wavelet based speech enhancement," International Workshop on Nonlinear Signal and Image Processing (NSIP 2005), pp.513-516, Sapporo, Japan, May 2005. (発表者：野村 行弘)

## 本研究に関連する学会・研究会等

- [1] **野村 行弘**, 呂 建明, 関屋 大雄, 谷萩 隆嗣, "リカレントニューラルネットワークを利用したノイズキャンセラの一設計法," 2001年電子情報通信学会総合大会, vol.A-4-20, pp.120, Mar. 2001. (発表者：野村 行弘)
- [2] **野村 行弘**, 呂 建明, 関屋 大雄, 谷萩 隆嗣, "並列リカレントニューラルネットワークを利用したノイズキャンセラ的设计法," 2002年電子情報通信学会総合大会, vol.A-4-41, pp.151, Mar. 2002. (発表者：野村 行弘)

- [3] 趙 雪琴, 呂 建明, 野村 行弘, 谷萩 隆嗣, “並列型 Volterra フィルタを利用したノイズキャンセラの一設計法,” 第 18 回デジタル信号処理シンポジウム, Nov. 2003. (発表者: 趙 雪琴)
- [4] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “雑音量に依存しない音声領域と雑音領域との判別を用いた音声強調,” 電子情報通信学会応用音響(音声)研究会, EA2004-6(SP-2004-6), pp.29-34, Apr. 2004. (発表者: 野村 行弘)
- [5] 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “雑音量に依存しない音声領域と雑音領域との判別方法,” 第 47 回 自動制御連合講演会, Nov. 2004. (発表者: 野村 行弘)
- [6] 斗澤 秀亮, 野村 行弘, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “反復ウェーブレット雑音除去を用いた音声強調,” 第 47 回 自動制御連合講演会, Nov. 2004. (発表者: 斗澤 秀亮)
- [7] 斗澤 秀亮, 野村 行弘, 山下 哲孝, 呂 建明, 関屋 大雄, 谷萩 隆嗣, “モルフォロジー処理を用いたスペクトルサブトラクションにおけるミュージカルノイズ除去,” 第 18 回 回路とシステム軽井沢ワークショップ, pp.159-164, Apr. 2005. (発表者: 野村 行弘)