



HOKKAIDO UNIVERSITY

Title	音声情報処理における適応ARMA分析手法の応用に関する研究
Author(s)	広重, 真人
Degree Grantor	北海道大学
Degree Name	博士(工学)
Dissertation Number	甲第3054号
Issue Date	1992-03-25
DOI	https://doi.org/10.11501/3060658
Doc URL	https://hdl.handle.net/2115/49763
Type	doctoral thesis
File Information	000000249604.pdf



音声情報処理における適応 ARMA 分析手法の
応用に関する研究

広 重 真 人

①

博士論文

音声情報処理における適応 ARMA 分析手法の
応用に関する研究

北海道大学大学院工学研究科
電子工学専攻

広 重 真 人

1991 年 12 月

目次

1	緒論	2
I	音声の分析と認識	10
2	適応信号処理手法	11
2.1	トランスヴァーサル型 ARMA スペクトル推定アルゴリズム	11
2.1.1	モデルの設定	11
2.1.2	重み付き RLS 法	12
2.1.3	モデル同定システム (MIS 法)	14
2.2	ARMA ラティスモデル	16
2.2.1	改良適応 ARMA ラティスモデル	16
2.2.2	規格化 ARMA ラティスモデル	20
2.2.3	コンパクト ARMA ラティスモデル	22
2.2.4	改良適応 ARMA ラティスと規格化 ARMA ラティスとのスペクトルの差異	24
2.2.5	符号化・合成と ARMA ラティスモデル	29
3	有声音の直接的分析手法	32
3.1	まえがき	32
3.2	直接的分析手法のアルゴリズム	33
3.3	モデル実験	36

目次	(ii)
4 重み付き RLS 法におけるλの周波数特性	40
4.1 まえがき	40
4.2 λ の周波数特性	41
5 重み付き RLS 法におけるスペクトル選択	48
5.1 スペクトル選択の必要性	48
5.2 スペクトル選択手法	49
5.2.1 if-then ルールによる選択システム	49
5.2.2 if-then ルールとニューラルネットワーク	51
5.2.3 ニューラルネットワークによる選択システム	54
5.2.4 ネットワーク構造に関する検討	56
5.3 選択実験結果	59
6 無声音声と有声音声の分離	66
6.1 分離の必要性	66
6.2 分離アルゴリズム	68
6.2.1 分離に用いるパラメータ	68
6.2.2 パラメータ-時間曲線の平滑化	70
6.2.3 分離点の決定	75
6.3 分離実験結果	79
7 アルゴリズム選択型分析システム	84
7.1 システム構成	84
7.1.1 本システムの特徴	84
7.1.2 インプリメント	87
7.2 スペクトル分析実験	87
8 連続音声のセグメンテーション候補検出	91
8.1 まえがき	91
8.2 セグメンテーションパラメータ	93
8.2.1 ケプストラム距離	93

目次	(iii)
----	-------

8.2.2 有声・無声分離アルゴリズムの出力点	95
8.2.3 MIS 法の $\hat{q}^2$ パラメータ	98
8.3 セグメンテーション候補検出実験	98
8.4 考察	100

II 音声の符号化・合成 105

9 音声符号化・合成システム 106

9.1 まえがき	106
9.2 音声符号化システムにおける適応 ARMA 分析の必要性	107
9.3 音声符号化システム	108
9.4 符号化・合成における音質評価	108

10 残差を線形量子化した符号化システム 111

10.1 実験システム	111
10.2 実験結果	112

11 長期予測フィルタと CELP 符号化方式 117

11.1 長期予測フィルタ	117
11.2 聴覚重み付けフィルタ	119
11.2.1 AR モデルの場合	119
11.2.2 ARMA モデルの場合	120
11.3 CELP 符号化方式	121
11.3.1 Open loop CELP 符号化方式	122
11.3.2 Closed loop CELP 符号化方式	122
11.4 ラティス係数から ARMA 係数への変換	127

12 Open loop CELP 符号化手法による実験 128

12.1 都道府県名データを用いた実験	128
12.2 短文データを用いた実験	131

目次	(iv)
13 Closed loop CELP 符号化手法による実験	140
13.1 短文データを用いた実験	140
13.1.1 STP 係数推定をラティス型で行う場合	140
13.1.2 STP 係数推定をトランスヴァーサル型で行う 場合	141
13.2 CD-ROM データを用いた実験	159
13.2.1 ARMA モデルと CELP 符号化手法の組み合わ せに関する考察	167
14 フォルマント情報を利用した CELP 符号化システム	169
14.1 まえがき	169
14.2 フォルマント抽出法	170
14.3 符号化・合成実験	171
15 結論	182
謝辞	190
参考文献	192
著者の発表論文	201

Chapter 1

緒論

緒論

人類が歴史の道に下りたのは、高度な知能を備えた動物の群である。人間は、自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。また、自然の環境を人間の活動によって変化させることができる。このように、自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。

そのような自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。また、自然の環境を人間の活動によって変化させることができる。このように、自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。

一方、人間は自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。また、自然の環境を人間の活動によって変化させることができる。このように、自然の環境と人間の活動との相互作用により、自然の環境を変化させることができる。

Chapter 1

緒論

人類が他の生物と区別される点の一つとして、高度な言語を持つことが挙げられる。人類は言語を用いて他の個体と意志を通じ合うことにより、高度な協調作業を行うことができる。また経験によって得た知識を言語によって記述し、後の世代へ伝承することにより、過去の蓄積の上にさらに発展を積み重ねることが可能である。

そのような言語の最も基本的な、かつ重要な表現形態として、音声による言語表現がある。音声による言語表現の起源は有史以前にさかのぼり、もう一つの言語表現である文字表現よりはるかに長い歴史を持つと考えられる。また音声による言語表現は、我々が誕生して成長していく過程で自然に体得するものであり、その意味で最も平易な言語表現であるといえる。

一方人類は直立歩行により自由になった2本の手を用いて、様々な道具を創造してきた。それらの道具は人間の限られた能力を補い、または増幅して人類の可能性を拡大してきた。人類が創造した「道具」の中で現在最も高度なものの一つに、「コンピュータ」が挙げられる。コンピュータは、デジタル情報処理技術を用いて人間の論理的な思考の一部をシミュレーションし、人間よりはるかに速く、正確に、大量に情報処理を行う。コンピュータの出現により、人間は従来ならば多くの時間と体力を費やさざるを得なかった大量の単純作業から解放され、人間にしかできない「創造的」な活動に多くの

時間を充てることが可能となった。

現在人間は、一種の「アシスタント」であるコンピュータに自らの意志を伝えるために、主に言語の文字表現を用いている。これは文字表現は論理的な思考により適しているという理由の他に、いくつかの技術的な理由がある。すなわち、(1) コンピュータに命令を与えるには、我々が普段話している自然言語よりも、必要な内容を短く効率よく表した文字による人工言語の方が、コンピュータ内部の処理が平易かつ能率的であること、(2) 現在の時点では、熟練したオペレータであれば、キーボードによる文字入力が最も効率の良いコンピュータへの情報入力方法であること、等である。

しかしこれらはある意味で、人間が訓練をすることによってコンピュータ側へ歩み寄っていることを示し、コンピュータが「難しい」ものになっている原因の一つと考えられる。コンピュータがより多くの人に使われ、その恩恵により、多くの人間が創造的な活動に従事するためには、コンピュータは現在より「易しい」ものになるべきであり、そのためには人間寄りのユーザ・インタフェースが開発されねばならない。

人間寄りのインタフェースで有力なものとして、一つは既にいくつかの実現例があるグラフィック・ユーザ・インタフェースであり、もう一つは、上述の「最も平易な言語表現」である音声によるインタフェースである。

音声によるユーザ・インタフェースを実現するには、まずコンピュータが人間の発音した音声を「聴き取る」ことができないとできない。つぎにこの聴き取った音声をコンピュータ内部で自在に処理し、必要に応じて再現できるように、効率よく「記録」できないとできない。この「記録」には、聴き取った音声を文字化して記録する方法と、抑揚や声質まである程度保存するように符号化して記録する方法がある。後者は、各個人の音声をデータとする音声データベースや、蓄積型の通信である電子メールにおいて送り主の

音声を送るボイスメール等で用いられるものである。

「聴き取り」および「文字化」のための技術が、音声の「認識」であり、個人性を保存した「記録・再生」のための技術が、「符号化・合成」である。またこれらの二つの分野に共通して用いられる基礎技術が、音声の「分析」である。これらの分野は上述の音声ユーザ・インタフェースの観点から密接な関係があり、音声情報処理の分野においても、音声生成モデルによって表されるスペクトルの推定と応用という点で表裏一体の関係がある。

音声情報処理で広く用いられている線形予測分析 (Linear Predictive Coding, LPC) に例をとってその関係を述べる。LPC は 1966 年に板倉、斉藤により⁽⁵⁾、1967 年に B.S.Atal, M.R.Schroeder により⁽⁶⁾、ほぼ時を同じくして開発された。開発当初は、少ないパラメータで音声の分析・合成系を表現できることから、音声の符号化・伝送に応用された。また 1969 年には板倉、斉藤により、ラティス型の LPC である PARCOR 分析・合成が提案されている⁽⁷⁾。当時音声認識の分野では、アナログ処理向きの帯域フィルタバンクによるスペクトル表現が用いられており、1970 年に迫江、千葉により提案された DP マッチングはこのスペクトル表現を用いている⁽⁴⁴⁾。1970 年代中頃に、B.S.Atal によって LPC ケプストラムが提案され⁽⁸⁾、計算機処理向きのスペクトル距離尺度である LPC ケプストラム距離が、話者認識や音声認識で用いられるようになった。1970 年代後半になると、計算機的能力向上に伴うデジタル信号処理技術の発展から、LPC を用いた音声認識の研究が盛んに行われた。このように、当初符号化のために開発された LPC が、音声認識の分野にも応用された経緯をみても、分析・認識・符号化の間の密接な関係が理解できる。

上述の LPC 分析は、音声の短区間定常性を仮定しているため、音声の過渡部分の急激なスペクトルの変化は表現しきれず、またモデルはフォルマントのみを表現できる全極型の AR モデルを用いていることから、アンチフォルマントを含む鼻子音等では十分なスペ

クトル精度が得られない。このことは音声認識の分野において十分な認識精度が得られない原因の一つと考えられている。

近年、一括型 AR モデルに基づく従来の LPC 分析手法よりも高精度なスペクトル分析が可能な適応 ARMA 分析アルゴリズムがいくつか提案されてきている。主なものとしては、SEARMA⁽¹⁰⁾⁽¹¹⁾、重み付き RLS⁽¹⁵⁾⁽¹⁶⁾、MIS⁽¹⁷⁾⁽¹⁶⁾等がある。これらの分析アルゴリズムでは適応的分析が行われるため、音声スペクトルが各サンプリング点に 1 組ずつ求められ、高い時間分解能を持つことができる。また音声生成モデルは ARMA モデルを推定可能であるため、フォルマントのみでなくアンチフォルマントも推定可能である。

これらの適応 ARMA 分析アルゴリズムは、精度の高い音声スペクトル分析が行えることが既に報告されているが、得られた高精度スペクトルを応用した音声認識システムや符号化・合成システムに関する研究は数少ない。

本論文では、音声によるユーザ・インタフェースを実現するための基礎研究として、上述の高精度適応 ARMA スペクトル分析アルゴリズムを音声認識と音声符号化・合成の分野に応用し、実音声を対象とする音声処理システムを構築して、それらの有効性を実験により検証する。すなわち本論文は、適応信号処理技術の音声情報処理分野への応用について論じたものである。

まず、認識・合成・符号化の基礎となる音声のスペクトル分析手法について、既存の適応 ARMA 分析アルゴリズムの問題点を改善する、アルゴリズム選択型スペクトル分析システムを構築する。次に音声認識の分野において、認識の基礎となるセグメンテーション候補点検出システムを構築する。符号化・合成の分野では適応 ARMA 分析アルゴリズムを導入した CELP 符号化システムを構築し、種々の条件のもとで実験・評価を行う。

本論文の構成を以下に示す。

本論文は 2 部から構成されている。第 1 部は、第 2 章から第 8 章

までにより構成され、分析アルゴリズムおよび音声認識について述べている。

第2章では、本論文全体で用いられる高精度適応 ARMA スペクトル推定手法について述べる。既に提案されている2つのトランスヴァーサル型推定アルゴリズム、すなわち重み付き RLS 法と MIS 法、および3つのラティス型モデル、すなわち改良適応 ARMA ラティスモデル⁽¹⁰⁸⁾、規格化 ARMA ラティスモデル⁽¹⁰⁸⁾⁽²¹⁾、コンパクト ARMA ラティスモデル⁽²¹⁾のそれぞれについて、アルゴリズムおよびモデルを表現する重要な式を示し、それらの手法の特徴について簡略に述べる。

第3章から第7章まででは、従来よりさらに高精度な音声分析システムを構築することを目的とする。また従来理論的・数式的であった信号処理に、経験的・知識工学的な手法を導入する。有声音声と無声音声はそれぞれに適したアルゴリズムで別個に分析されるべきであるという考えから、まず従来からピッチによる精度の変化の問題が指摘されている有声音声について、最適なアルゴリズムを検討する。次に有声音声と無声音声の分離を行い、有声・無声それぞれに最適なアルゴリズムを選択して分析する新しい音声分析システムを構築する。以下に個々の章について述べる。

第3章では、有声音声に対し理論的に高精度な直接的分析法⁽¹⁰¹⁾を提案する。この直接的分析法は、有声音声はその確率的な性質より信号そのものの示す性質が重要であるとの考えに基づき、相関操作を全く用いずに原信号から直接スペクトルを求める信号処理手法である。モデル実験の結果より、直接的分析法は理論的には高精度であるが、量子化雑音等のわずかな雑音にもきわめて敏感であることが示される。

第4章では、第3章で述べた直接適分析法よりも実際の有声音声分析に適する手法として、第2章で述べた重み付き RLS 法を取り上げ、そのアルゴリズム中に導入された重み係数 λ の周波数特性を新し

く明らかにする⁽¹⁰¹⁾。 λ の周波数特性から、重み付き RLS 法が有声音に適した分析法であることが確認される。

第 5 章では、重み付き RLS 法を用いて高精度な有声音分析を行うためには、分析結果を選択することが必要であることを述べ、if-then ルールを用いた選択システムと、ニューラルネット⁽³¹⁾のノードを用いた選択システムを構築し、それらのシステムによって良好なスペクトル選択が行えることを示す⁽¹⁰⁷⁾。

第 6 章では、有声音と無声音に別個の分析アルゴリズムを適用するために、有声音と無声音の分離を行う⁽¹¹⁰⁾。分離には、適応的に求められた 1 次の AR 係数が用いられ、分離点の決定には、いくつかの経験的ルールを用いる知識工学的手法が用いられる。実音声による実験結果を示し、提案する手法が有効であることが示される。

第 7 章では、第 2 章から第 6 章までで述べた手法を統合することにより、有声音・無声音を自動判別しそれぞれに適したアルゴリズムで高精度な適応 ARMA スペクトル分析を行う、新しい音声スペクトル分析システムを構築する⁽¹¹⁰⁾。

続く第 8 章では、適応的に求められた高精度 ARMA スペクトルを用いた音声認識への第一歩として、第 2 章で述べた MIS 法のスペクトル変化と、第 6 章で述べた有声音・無声音分離システムの出力結果を用いて、連続音声のセグメンテーション候補検出システムを作成する⁽⁵²⁾⁽¹¹⁶⁾。実音声による実験から、提案するセグメンテーション候補検出システムは音素の脱落の少ない良好な結果が得られることを示す。

以上が音声の分析と認識に関わる第 1 部の内容である。続く第 2 部は、第 9 章から第 14 章までにより構成され、音声の符号化・合成について述べている。すなわち、従来の音声符号化システムのスペクトル推定フィルタ (短期予測フィルタ、STP) 部に適応 ARMA スペクトル分析を導入し、より高精度なスペクトルを用いることによっ

て、低ビット音声符号化・合成システムの性能向上をめざす。

第9章では、まず音声符号化システムにおける適応 ARMA 分析の必要性について述べる。すなわち、高精度な調音スペクトルの推定によって残差信号の情報量が減少し、その結果同一のビットレートのもとでは高品質の合成音声を得られる可能性があることを述べる。次に一般的な符号化システムの構成を示し、本論文で導入する適応 ARMA 分析アルゴリズムの、符号化システムにおける位置を明確にする。一般的な音声符号化システムはスペクトルフィルタ部と残差符号化部により構成され、本論文では一貫してスペクトルフィルタ部に適応 ARMA スペクトルを用い、残差符号部に様々な符号化手法を用いて符号化・合成実験を行う。第9章の最後に、本論文での合成音声の品質評価法について述べ、本論文では主に客観評価尺度の一つであるセグメンタル S/N 比を品質評価に用いることを述べる。以下に続く各章では、残差符号化手法ごとの実験の詳細と結果について述べる。

第10章では、スペクトル推定フィルタの残差波形を単純に線形量子化した場合について述べる⁽¹⁰⁴⁾。実験により、適応 ARMA スペクトルを用いた場合、アンチフォルマントを含むと考えられる鼻子音等で、従来の一括型 AR スペクトルである PARCOR を用いた場合より良好な結果が得られることを示す。このことは第9章で述べた、符号化・合成における適応 ARMA 分析の必要性の裏付けとなる。

第11章では、以下に続く各章で用いる CELP 符号化手法⁽⁶³⁾と、CELP に付随する長期予測フィルタ (LTP)⁽⁶²⁾ および聴覚重み付けフィルタ (PWF)⁽⁶²⁾について述べる。

第12章と第13章では、第11章で述べた CELP 符号化手法に、適応 ARMA スペクトル分析アルゴリズムを導入した符号化システムを構築し、様々な条件下で行った多くの実験の詳細について述べる。第12章は Open loop CELP、第13章は Closed loop CELP を用いた実験について述べる。すべての実験の中で、スペクトル推定に AR

モデルの重み付き RLS、符号化手法に Closed loop CELP を用いた符号化システムが最も良い結果を示し、従来の PARCOR に基づく Closed loop CELP システムと比較して、平均セグメンタル S/N 比で 0.53~1.45dB の音質改善が得られることを示す。また、スペクトル推定に ARMA モデルを用いた場合、概して良好な結果が得られないことが示され、CELP 符号化方式と MA 部の関係からその原因について考察する。

第 2 部の最後となる第 14 章では、第 13 章で示された MA 部と CELP との不整合を避けるため、高精度 ARMA スペクトルから抽出したフォルマント情報を用いる CELP 符号化システムを構築し、実験によりその有効性を検討する⁽¹²⁰⁾。

第 15 章は結論として、本論文に述べられた研究を総括し、今後の課題を述べる。

Chapter 2

音声信号処理手法

Part I

音声の分析と認識

2.1.1 モデルの設定

本章では、音声信号をモデルとして表現する。モデルは、入力信号と出力信号の関係を示す。

モデルは、入力信号と出力信号の関係を示す。

$$y(n) = -\sum_{k=1}^M a_k y(n-k) + \sum_{k=0}^M b_k x(n-k) + w(n) \quad (2.1)$$

ここで、 $y(n)$ は出力信号、 $x(n)$ は入力信号、 $w(n)$ は雑音信号、 a_k と b_k はモデルのパラメータである。

モデルのパラメータは、入力信号と出力信号の関係を示す。

$$y(n) = -\sum_{k=1}^M a_k y(n-k) + \sum_{k=0}^M b_k x(n-k) + w(n) \quad (2.2)$$

ここで、 $y(n)$ は出力信号、 $x(n)$ は入力信号、 $w(n)$ は雑音信号、 a_k と b_k はモデルのパラメータである。

Chapter 2

適応信号処理手法

2.1 トランスヴァーサル型 ARMA スペクトル推定アルゴリズム

本節では、本論文で用いる高精度適応スペクトル推定手法について述べる。

2.1.1 モデルの設定

まず最初に推定の対象となる規範モデルを定義する。規範モデルは ARMA モデルとし、

$$y(k) = -\sum_{i=1}^n a_i(k)y(k-i) + \sum_{j=1}^m b_j(k)u(k-j) + u(k) \quad (2.1)$$

で与えられるとする。但し、 $y(k)$ は出力であり、 $u(k)$ は入力で、重み付き RLS 法の場合疑似パルス列、MIS 法の場合 $u(k)$ は白色ガウス過程である。

推定モデル (時刻 τ の推定値) を、

$$\hat{y}(\tau|k) = -\sum_{i=1}^s \hat{a}_i(k)y(\tau-i) + \sum_{j=1}^t \hat{b}_j(k)\hat{u}(\tau-j) \quad (2.2)$$

とする。ここで、

$$\hat{\mathbf{p}}(k)^T = [\hat{a}_1(k) \cdots \hat{a}_s(k) \hat{b}_1(k) \cdots \hat{b}_t(k)] \quad (2.3)$$

$$\mathbf{h}(k)^T = [-y(k-1) \cdots -y(k-s) \hat{u}(k-1) \cdots \hat{u}(k-t)] \quad (2.4)$$

とする。このとき、式 (2.2) は、

$$\hat{y}(\tau|k) = \mathbf{h}(\tau)^T \hat{\mathbf{p}}(k) \quad (2.5)$$

となる。

2.1.2 重み付き RLS 法

重み付き RLS 法については文献⁽¹³⁾⁽¹⁴⁾⁽¹⁵⁾において詳細が述べられているため、以下にそのアルゴリズムと特徴のみを述べる。

- アルゴリズム

- 漸化式

$$\nu(k) = y(k) - \mathbf{h}(k)^T \hat{\mathbf{p}}(k-1) \quad (2.6)$$

if $\nu^2(k) \geq C_0 \cdot \nu_s^2(k-1)$ and $k \geq t_i + t_0$ then

$$\hat{u}^P = \nu(k), \nu(k) = 0, t_i = k \quad (2.7)$$

else

$$\hat{u}^P = 0 \quad (2.8)$$

endif

$$\nu_s^2(k) = \nu_s^2(k-1) + \frac{\nu^2(k) - \nu_s^2(k-1)}{s} \quad (2.9)$$

$$\hat{\mathbf{p}}(k) = \hat{\mathbf{p}}(k-1) + \mathbf{F}(k-1)\mathbf{h}(k) \cdot \{\lambda + \mathbf{h}(k)^T \mathbf{F}(k-1)\mathbf{h}(k)\}^{-1} \nu(k) \quad (2.10)$$

$$\mathbf{F}(k) = \frac{\mathbf{F}(k-1) - \mathbf{F}(k-1)\mathbf{h}(k) \cdot \{\lambda + \mathbf{h}(k)^T \mathbf{F}(k-1)\mathbf{h}(k)\}^{-1} \cdot \mathbf{h}(k)^T \mathbf{F}(k-1)}{\lambda} \quad (2.11)$$

$$\hat{u}(k) = \hat{u}^P \quad (2.12)$$

○ 初期値

$$\hat{\mathbf{p}}(0) = 0, \mathbf{F}(0) = 100 \cdot \mathbf{I}, t_i = 100, \nu_s^2 = 0$$

○ 定数

$$t_0 = 20, s = 50, C_0 = 3.0$$

○ 評価関数

$$V(k) = \sum_{i=0}^k \lambda^{k-i} \{y(i) - \hat{y}(i|k)\}^2 \quad (2.13)$$

● 特徴

1. 入力はパルス列である。
2. 時変 ARMA パラメータの推定が可能である。
3. パルスの検出は予測誤差があるしきい値を越えた時に行なわれる。そのときのパルス値は同一時刻の予測誤差とする。
4. 現在の観測信号に重みをかけ、過去の観測信号の情報を徐々に忘却させる重み係数 λ を導入している。

5. 重み付き RLS 法は、重み係数 λ の働きにより時間的に狭い範囲で観測された信号を用いた推定を行なうため、1 周期内のデータからスペクトル推定を行なわなければならない確定信号に適している。

2.1.3 モデル同定システム (MIS 法)

MIS 法の詳細についても文献⁽¹⁸⁾⁽¹⁷⁾⁽¹⁶⁾に示されているため、そのアルゴリズムと特徴のみを以下に述べる。

- アルゴリズム

- 漸化式

$$\nu(k) = y(k) - \mathbf{h}(k)^T \hat{\mathbf{p}}(k-1) \quad (2.14)$$

$$\begin{aligned} &\text{if } \nu^2(k) \geq C \cdot \hat{\sigma}^2(k-1) \cdot \{1 + \mathbf{h}(k)^T \mathbf{F}(k-1) \mathbf{h}(k)\} \\ &\quad \text{and } k \geq t_i + t_0 \text{ then} \end{aligned} \quad (2.15)$$

$$\begin{aligned} \hat{q}^2(k-1) &= [\nu^2(k) - \hat{\sigma}^2(k-1) \\ &\quad \cdot \{1 + \mathbf{h}(k)^T \mathbf{F}(k-1) \mathbf{h}(k)\}] / \{\mathbf{h}(k)^T \mathbf{h}(k)\} \end{aligned} \quad (2.16)$$

$$\hat{u} = \nu(k), \nu(k) = 0, t_i = k \quad (2.17)$$

else

$$\hat{q}^2(k-1) = 0, \hat{u} = 0 \quad (2.18)$$

endif

$$\mathbf{F}(k|k-1) = \mathbf{F}(k-1) + \frac{\hat{q}^2(k-1)}{\hat{\sigma}^2(k-1)} \cdot \mathbf{I} \quad (2.19)$$

$$\hat{\mathbf{p}}(k) = \hat{\mathbf{p}}(k-1) + \mathbf{F}(k|k-1) \mathbf{h}(k) \cdot \{1 + \mathbf{h}(k)^T \mathbf{F}(k|k-1) \mathbf{h}(k)\}^{-1} \nu(k) \quad (2.20)$$

$$\hat{\mathbf{p}}^S(k) = \hat{\mathbf{p}}^S(k-1) + \frac{\hat{\mathbf{p}}(k) - \hat{\mathbf{p}}^S(k-1)}{s} \quad (2.21)$$

$$\begin{aligned} \mathbf{F}(k) = & \mathbf{F}(k|k-1) - \mathbf{F}(k|k-1)\mathbf{h}(k) \cdot \{1 + \mathbf{h}(k)^T \mathbf{F}(k|k-1)\mathbf{h}(k)\}^{-1} \\ & \cdot \mathbf{h}(k)^T \mathbf{F}(k|k-1) \end{aligned} \quad (2.22)$$

$$\alpha(k) = \min(k, \alpha) \quad (2.23)$$

$$\hat{\sigma}^2(k) = \hat{\sigma}^2(k-1) + \frac{\frac{\nu^2(k)}{1 + \mathbf{h}(k)^T \mathbf{F}(k|k-1)\mathbf{h}(k)} - \hat{\sigma}^2(k-1)}{\alpha(k)} \quad (2.24)$$

$$\hat{u}(k) = \frac{\nu(k)}{1 + \mathbf{h}(k)^T \mathbf{F}(k|k-1)\mathbf{h}(k)} + \hat{u} \quad (2.25)$$

○ 初期値

$$\hat{\mathbf{p}}(0) = 0, \mathbf{F}(0) = 100 \cdot \mathbf{I}, \hat{q}^2(0) = 0, \hat{\sigma}^2(0) = 0.1, t_i = 100$$

○ 定数

$$\alpha = 50, t_0 = 20, s = 50, C = 16$$

○ 評価関数

$$V = \frac{\sigma_u^2}{2\pi j} \oint \left| H(z^{-1}) \cdot \frac{1 + \sum_{i=1}^s \hat{a}_i z^{-i}}{1 + \sum_{j=1}^t \hat{b}_j z^{-j}} \right|^2 \frac{dz}{z} \quad (2.26)$$

● 特徴

1. 係数推定時における入力は白色ガウス過程である。
2. 時変 ARMA パラメータの推定が可能である。

3. パラメータの時変性は推定誤差の変化に従っており、誤差がガウス分布により決まるしきい値を越えた時パラメータが変化したと検出し、その変化量は最尤推定法により求めている。
4. 信号の2次モーメントからモデルの係数を推定するため、確率信号に適している。

2.2 ARMA ラティスモデル

本節では主に第2部の符号化・合成で用いる ARMA ラティスモデルについて述べる。本節では既に提案されているいくつかの ARMA ラティスモデルのうち、改良適応 ARMA ラティスモデル、規格化 ARMA ラティスモデル、コンパクト ARMA ラティスモデルの3種類の ARMA ラティスモデルについて述べる。

2.2.1 改良適応 ARMA ラティスモデル

文献⁽¹⁰⁸⁾に示される改良適応 ARMA ラティスモデルは、ARMA 確率過程の正確な表現が可能なモデルである。このモデルでは、以下に示す3つの誤差を用いる。

$$\nu_{s,t}(k+1|k) = y(k+1) - \hat{y}_{s,t}^\nu(k+1|k) \quad (2.27)$$

$$\gamma_{s,t}(k+1|k) = x(k+1-t) - \sum_{i=0}^{s-1} \hat{a}_i^\gamma(L) y(k+1-i) + \sum_{j=0}^{t-1} \hat{b}_j^\gamma(L) x(k+1-j) \quad (2.28)$$

$$\xi_{s,t}(k+1|k) = y(k+1-s) + \sum_{i=0}^{s-1} \hat{a}_i^\xi(L) y(k+1-i) - \sum_{j=0}^{t-1} \hat{b}_j^\xi(L) x(k+1-j) \quad (2.29)$$

ただし、

$$\hat{y}_{s,t}^{\nu}(k+1)(L) = -\sum_{i=1}^s \hat{a}_i^{\nu}(L)y(k+1-i) + \sum_{j=1}^t \hat{b}_j^{\nu}(L)x(k+1-j) \quad (2.30)$$

である。

式 (2.27), (2.28), (2.29) で与えられた各 ARMA 係数は、それぞれ以下の評価関数を最小にするように推定される。

$$V_{s,t}^{\nu}(k) = (\nu_{s,t}(i|k), \nu_{s,t}(i|k))_{i,k} = \sum_{i=0}^k \lambda_{s,t}^2(i|k) \quad (2.31)$$

$$V_{s,t}^{\gamma}(k) = (\gamma_{s,t}(i|k), \gamma_{s,t}(i|k))_{i,k} \quad (2.32)$$

$$V_{s,t}^{\xi}(k) = (\xi_{s,t}(i|k), \xi_{s,t}(i|k))_{i,k} \quad (2.33)$$

以上の誤差と評価関数より、改良適応 ARMA ラティスモデルの 3 つの基本区間の回帰式が以下のように求められる。

○ AR 型基本区間

$$\begin{aligned} \nu_{s+1,0}(k+1|k) &= \nu_{s,0}(k+1|k) - \mu_{s,0}^a(k)\xi_{s,0}(k|k-1) \\ \xi_{s+1,0}(k+1|k) &= \xi_{s,0}(k|k-1) - \mu_{s,0}^b(k)\nu_{s,0}(k+1|k) \end{aligned} \quad (2.34)$$

$$\alpha_{s+1,0}(k) = \alpha_{s,0}(k) + \frac{\xi_{s,0}^2(k|k-1)}{V_{s,0}^{\xi}(k-1)} \quad (2.35)$$

$$\begin{aligned} \mu_{s,0}^a(k) &= \frac{V_{s,0}^{\nu\xi}(k)}{V_{s,0}^{\xi}(k-1)} \\ \mu_{s,0}^b(k) &= \frac{V_{s,0}^{\nu\xi}(k)}{V_{s,0}^{\nu}(k)} \end{aligned} \quad (2.36)$$

$$\begin{aligned}
V_{s,0}^{\nu}(k+1) &= \lambda V_{s,0}^{\nu}(k) + \frac{\nu_{s,0}^2(k+1|k)}{\lambda + \alpha_{s,0}(k)} \\
V_{s,0}^{\xi}(k) &= \lambda V_{s,0}^{\xi}(k-1) + \frac{\xi_{s,0}^2(k|k-1)}{\lambda + \alpha_{s,0}(k)} \\
V_{s,0}^{\nu\xi}(k+1) &= \lambda V_{s,0}^{\nu\xi}(k) + \frac{\nu_{s,0}(k+1|k)\xi_{s,0}(k|k-1)}{\lambda + \alpha_{s,0}(k)}
\end{aligned} \tag{2.37}$$

ただし、

$$V_{s,0}^{\nu\xi}(k) = (\nu_{s,0}(i|k), \xi_{s,0}(i-1|k-1))_{i,k} \tag{2.38}$$

および $\alpha_{s,0}(k)$ は予測誤差の内積に対する補正項である。

○ AR 型 ARMA 基本区間

$$\begin{aligned}
\nu_{s+1,t}(k+1|k) &= \nu_{s,t}(k+1|k) - \eta_{s,t}^a \xi_{s,t}(k|k-1) \\
\xi_{s+1,t+1}(k+1|k) &= \xi_{s,t}(k|k-1) - \eta_{s,t}^b(k)(\nu_{s,t}(k+1|k) - x(k+1))
\end{aligned} \tag{2.39}$$

$$\alpha_{s+1,t}(k) = \alpha_{s,t}(k) + \frac{\xi_{s,t}^2(k|k-1)}{V_{s,t}^{\xi}(k-1)} \tag{2.40}$$

$$\begin{aligned}
\eta_{s,t}^a(k) &= \frac{V_{s,t}^{\nu\xi}(k)}{V_{s,t}^{\xi}(k-1)} \\
\eta_{s,t}^b(k) &= \frac{V_{s,t}^{\nu\xi}(k)}{V_{s,t}^{\nu}(k) - V_{0,0}^{\gamma}(k)}
\end{aligned} \tag{2.41}$$

$$\begin{aligned}
V_{s,t}^{\nu}(k+1) &= \lambda V_{s,t}^{\nu}(k) + \frac{\nu_{s,t}^2(k+1|k)}{\lambda + \alpha_{s,t}(k)} \\
V_{s,t}^{\xi}(k) &= \lambda V_{s,t}^{\xi}(k-1) + \frac{\xi_{s,t}^2(k|k-1)}{\lambda + \alpha_{s,t}(k)} \\
V_{s,t}^{\nu\xi}(k+1) &= \lambda V_{s,t}^{\nu\xi}(k) + \frac{\nu_{s,t}(k+1|k)\xi_{s,t}(k|k-1)}{\lambda + \alpha_{s,t}(k)}
\end{aligned} \tag{2.42}$$

○ MA 型 ARMA 基本区間

$$\begin{aligned}\nu_{s+1,t+1}(k+1|k) &= \nu_{s+1,t}(k+1|k) - \delta_{s+1,t}^a \gamma_{s+1,t}(k|k-1) \\ \gamma_{s+2,t+1}(k+1|k) &= \gamma_{s+1,t}(k|k-1) - \delta_{s+1,t}^b(k)(\nu_{s+1,t}(k+1|k) - x(k+1))\end{aligned}\quad (2.43)$$

$$\alpha_{s+1,t+1}(k) = \alpha_{s+1,t}(k) + \frac{\gamma_{s+1,t}^2(k|k-1)}{V_{s+1,t}^\gamma(k-1)} \quad (2.44)$$

$$\begin{aligned}\delta_{s+1,t}^a(k) &= \frac{V_{s+1,t}^{\nu\gamma}(k)}{V_{s+1,t}^\gamma(k-1)} \\ \delta_{s+1,t}^b(k) &= \frac{V_{s+1,t}^{\nu\gamma}(k)}{V_{s+1,t}^\nu(k) - V_{0,0}^\gamma(k)}\end{aligned}\quad (2.45)$$

$$\begin{aligned}V_{s+1,t}^\nu(k+1) &= \lambda V_{s+1,t}^\nu(k) + \frac{\nu_{s+1,t}^2(k+1|k)}{\lambda + \alpha_{s+1,t}(k)} \\ V_{s+1,t}^\gamma(k) &= \lambda V_{s+1,t}^\gamma(k-1) + \frac{\gamma_{s+1,t}^2(k|k-1)}{\lambda + \alpha_{s+1,t}(k)} \\ V_{s+1,t}^{\nu\gamma}(k+1) &= \lambda V_{s+1,t}^{\nu\gamma}(k) + \frac{\nu_{s+1,t}(k+1|k)\gamma_{s+1,t}(k|k-1)}{\lambda + \alpha_{s+1,t}(k)}\end{aligned}\quad (2.46)$$

入力推定

文献⁽¹⁰⁸⁾によると、入力が高ス過程の場合、最適推定入力は最終セクションでの前向き推定誤差で与えられる。高ス入力およびパルス入力を考慮した文献⁽¹⁵⁾による入力推定手法を用いると、改良適応ラティスモデルにおける入力推定は以下のようになる。

if $\nu_{n,m}^2(k+1|k) > C \cdot \sigma_x^2(k) \cdot \{1 + \alpha_{n,m}(k)\lambda^{-1}\}$ and $k > k_0 + T_0$ then

$$x(k+1) = \nu_{n,m}(k+1|k) \quad (2.47)$$

$$\nu_{n,m}(k+1|k) = 0, \quad k_0 = k \quad (2.48)$$

else

$$x(k+1) = \nu_{n,m}(k+1|k) \cdot \frac{\lambda}{\lambda + \alpha_{n,m}(k)} \quad (2.49)$$

endif

$$\sigma_x^2(k+1) = \sigma_x^2(k) + \frac{x^2(k+1) - \sigma_x^2(k)}{\min(k, T_\alpha)} \quad (2.50)$$

ここで C および T_0, T_α は任意定数である。

2.2.2 規格化 ARMA ラティスモデル

規格化 ARMA ラティスモデルは、エネルギー規格化誤差を用いることにより、モデルの表現に必要なパラメータが最少数となるように構築されたモデルである⁽²¹⁾。規格化 ARMA ラティスモデルで用いられるエネルギー規格化の誤差は次の式で与えられる。

$$\begin{aligned} \bar{\nu}_{s,t}(k|L) &= \frac{\nu_{s,t}(k|L)}{\sqrt{V_{s,t}^\nu(L)}} \\ \bar{x}(k|L) &= \frac{x(k)}{\sqrt{V_{0,0}^\gamma(L)}} \\ \bar{\nu}_{s,t}^x(k|L) &= \frac{\nu_{s,t}(k|L) - x(k)}{\sqrt{V_{s,t}^\nu(L) - V_{0,0}^\gamma(L)}} \\ \bar{\gamma}_{s,t}(k|L) &= \frac{\gamma_{s,t}(k|L)}{\sqrt{V_{s,t}^\gamma(L)}} \\ \bar{\xi}_{s,t}(k|L) &= \frac{\xi_{s,t}(k|L)}{\sqrt{V_{s,t}^\xi(L)}} \end{aligned} \quad (2.51)$$

AR 時数 n 、MA 次数 m について、 $n \geq m$ と仮定すると、以下の回帰式が得られる。

○ AR 型基本区間

$$\begin{aligned} \bar{\nu}_{s+1,0}(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\mu}_{s,0}^2(k)}} \cdot [\bar{\nu}_{s,0}(k+1|k) - \bar{\mu}_{s,0}(k)\bar{\xi}_{s,0}(k|k-1)] \\ \bar{\xi}_{s+1,0}(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\mu}_{s,0}^2(k)}} \cdot [\bar{\xi}_{s,0}(k|k-1) - \bar{\mu}_{s,0}(k)\bar{\nu}_{s,0}(k+1|k)] \end{aligned} \quad (2.52)$$

$$\bar{\mu}_{s,0}(k) = (\bar{\nu}_{s,0}(i|k), \bar{\xi}_{s,0}(i-1|k-1))_{i,k} \quad (2.53)$$

○ 追加型基本区間

$$\begin{aligned} \bar{\nu}_{n-m,0}^x(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\tau}_{n-m,0}^2(k)}} \cdot [\bar{\nu}_{n-m,0}(k+1|k) - \bar{\tau}_{n-m,0}(k)\bar{x}(k+1|k)] \\ \bar{\gamma}_{n-m+1,0}(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\tau}_{n-m,0}^2(k)}} \cdot [\bar{x}(k+1|k) - \bar{\tau}_{n-m,0}(k)\bar{\nu}_{n-m,0}(k+1|k)] \end{aligned} \quad (2.54)$$

$$\bar{\tau}_{n-m,0}(k) = (\bar{\nu}_{n-m,0}(i|k), \bar{x}(i))_{i,k} = \sqrt{\frac{V_{0,0}^\gamma(k)}{V_{n-m,0}^\nu(k)}} \quad (2.55)$$

○ AR 型 ARMA 基本区間

$$\begin{aligned} \bar{\nu}_{s+1,t}^x(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\eta}_{s,t}^2(k)}} \cdot [\bar{\nu}_{s,t}^x(k+1|k) - \bar{\eta}_{s,t}(k)\bar{\xi}_{s,t}(k|k-1)] \\ \bar{\xi}_{s+1,t+1}(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\eta}_{s,t}^2(k)}} \cdot [\bar{\xi}_{s,t}(k|k-1) - \bar{\eta}_{s,t}(k)\bar{\nu}_{s,t}^x(k+1|k)] \end{aligned} \quad (2.56)$$

$$\bar{\eta}_{s,t}(k) = (\bar{\nu}_{s,t}^x(i|k), \bar{\xi}_{s,t}(i-1|k-1))_{i,k} \quad (2.57)$$

○ MA 型 ARMA 基本区間

$$\begin{aligned} \bar{\nu}_{s+1,t+1}^x(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\delta}_{s+1,t}^2(k)}} \cdot [\bar{\nu}_{s+1,t}^x(k+1|k) - \bar{\delta}_{s+1,t}(k)\bar{\gamma}_{s+1,t}(k|k-1)] \\ \bar{\gamma}_{s+2,t+1}(k+1|k) &= \frac{1}{\sqrt{1 - \bar{\delta}_{s+1,t}^2(k)}} \cdot [\bar{\gamma}_{s+1,t}(k|k-1) - \bar{\delta}_{s+1,t}(k)\bar{\nu}_{s+1,t}^x(k+1|k)] \end{aligned} \quad (2.58)$$

$$\bar{\delta}_{s+1,t}(k) = (\bar{\nu}_{s+1,t}^x(i|k), \bar{\gamma}_{s+1,t}(i-1|k-1))_{i,k} \quad (2.59)$$

以上の回帰式で与えられる規格化 ARMA ラティスモデルの実現例として、各基本区間を縦続接続して実現した AR 4 次、MA 2 次のモデルを図 2.1 に示す。

実現される伝達関数

式 (2.52)～(2.59) で実現される規格化 ARMA ラティスモデルの伝達関数は、入力が $\bar{x}(k|L)$ で出力が $\bar{y}(k|L)$ であることから、

$$\bar{y}(k|L) = \frac{\hat{B}_m(z^{-1}, L)}{\hat{A}_n(z^{-1}, L)} \sqrt{\frac{V_{0,0}^\gamma(L)}{V_{0,0}^\nu(L)}} \bar{x}(k|L) \quad (2.60)$$

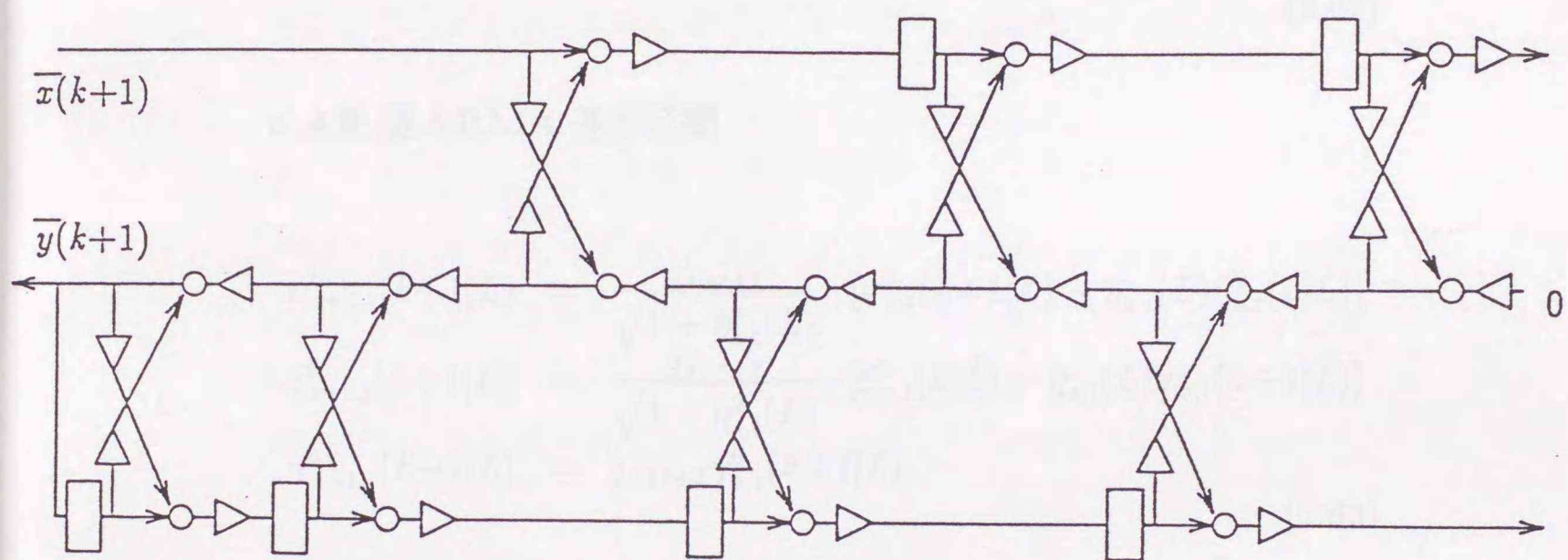
となり、 $\sqrt{\frac{V_{0,0}^\gamma(L)}{V_{0,0}^\nu(L)}}$ で表される利得項が付加される形となる。

2.2.3 コンパクト ARMA ラティスモデル

コンパクト ARMA ラティスモデルは、2.2.2 の規格化 ARMA ラティスモデルの各区間に利得を導入し、実現される伝達関数の利得を制御できるように一般化されたモデルである。コンパクト ARMA ラティスモデルの詳細は文献⁽²¹⁾に示されているため、本論文では回帰式を示すにとどめる。なお、以下の回帰式で、 $g_i = 1 (i = 1, 2, \dots, n+m+1)$ とすると、規格化 ARMA ラティスモデルが得られる。

○ AR 型基本区間

$$\begin{aligned} \nu_{s+1,0}^g(k+1|k) &= \frac{g_{s+1}}{\sqrt{1 - \bar{\mu}_{s,0}^2(k)}} \cdot [\nu_{s,0}^g(k+1|k) - \bar{\mu}_{s,0}(k) \xi_{s,0}^g(k|k-1)] \\ \xi_{s+1,0}^g(k+1|k) &= \frac{g_{s+1}}{\sqrt{1 - \bar{\mu}_{s,0}^2(k)}} \cdot [\xi_{s,0}^g(k|k-1) - \bar{\mu}_{s,0}(k) \nu_{s,0}^g(k+1|k)] \end{aligned} \quad (2.61)$$



A normalized ARMA lattice model with $n=4$ and $m=2$

図 2.1: 規格化 ARMA ラティスモデル

○ 追加型基本区間

$$\begin{aligned}
\nu_{n-m,0}^h(k+1|L) &= \frac{g_{n-m+1}}{\sqrt{1 - \bar{\tau}_{n-m,0}^2(L)}} \cdot [\nu_{n-m,0}^g(k+1|L) - \bar{\tau}_{n-m,0}(L)x^g(k+1|L)] \\
\gamma_{n-m+1,0}^h(k+1|L) &= \frac{g_{n-m+1}}{\sqrt{1 - \bar{\tau}_{n-m,0}^2(L)}} \cdot [x^g(k+1|L) - \bar{\tau}_{n-m,0}(L)\nu_{n-m,0}^g(k+1|L)] \\
\xi_{n-m,0}^h(k+1|L) &= g_{n-m+1}\xi_{n-m,0}^g(k+1|L)
\end{aligned} \tag{2.62}$$

○ AR 型 ARMA 基本区間

$$\begin{aligned}
\nu_{s+1,t}^h(k+1|L) &= \frac{g_{s+t+2}}{\sqrt{1 - \bar{\eta}_{s,t}^2(L)}} \cdot [\nu_{s,t}^h(k+1|L) - \bar{\eta}_{s,t}(L)\xi_{s,t}^h(k|L)] \\
\xi_{s+1,t}^h(k+1|L) &= \frac{g_{s+t+2}}{\sqrt{1 - \bar{\eta}_{s,t}^2(L)}} \cdot [\xi_{s,t}^h(k|L) - \bar{\eta}_{s,t}(L)\nu_{s,t}^h(k+1|L)] \\
\gamma_{s+1,t}^h(k+1|L) &= g_{s+t+2}\gamma_{s,t}^h(k+1|L)
\end{aligned} \tag{2.63}$$

○ MA 型 ARMA 基本区間

$$\begin{aligned}
\nu_{s+1,t+1}^h(k+1|L) &= \frac{g_{s+t+3}}{\sqrt{1 - \bar{\delta}_{s+1,t}^2(L)}} \cdot [\nu_{s+1,t}^h(k+1|L) - \bar{\delta}_{s+1,t}(L)\gamma_{s+1,t}^h(k|L)] \\
\gamma_{s+1,t+1}^h(k+1|L) &= \frac{g_{s+t+3}}{\sqrt{1 - \bar{\delta}_{s+1,t}^2(L)}} \cdot [\gamma_{s+1,t}^h(k|L) - \bar{\delta}_{s+1,t}(L)\nu_{s+1,t}^h(k+1|L)] \\
\xi_{s+1,t+1}^h(k+1|L) &= g_{s+t+3}\xi_{s+1,t}^h(k+1|L)
\end{aligned} \tag{2.64}$$

2.2.4 改良適応 ARMA ラティスと規格化 ARMA ラティスとのスペクトルの差異

2.2.1で述べた改良適応 ARMA ラティスモデルのラティス係数と、規格化 ARMA ラティスモデルのラティス係数の関係は、AR 型基本区間において式 (2.38) を考慮して、式 (2.36) と式 (2.53) より、

$$\bar{\mu}_{s,0}(k) = \sqrt{\mu_{s,0}^a(k) \cdot \mu_{s,0}^b(k)} \quad (2.65)$$

また AR 型 ARMA 基本区間において、式 (2.41) と式 (2.57) より、

$$\bar{\eta}_{s,t}(k) = \sqrt{\eta_{s,t}^a(k) \cdot \eta_{s,t}^b(k)} \quad (2.66)$$

同様に MA 型 ARMA 基本区間において、式 (2.45) と式 (2.59) より、

$$\bar{\sigma}_{s+1,t}(k) = \sqrt{\sigma_{s+1,t}^a(k) \cdot \sigma_{s+1,t}^b(k)} \quad (2.67)$$

となる。したがって規格化 ARMA ラティスモデルのラティス係数は、改良適応 ARMA ラティスモデルの 2 つのラティス係数を相乗平均したものとなっている。このため、これら 2 種類のラティスモデルによって表現されるスペクトルはわずかに異なったものとなる。

それぞれのラティスモデルのスペクトルの差異を示すために、モデル実験の例をいくつか示す。図 2.2 に示される AR 4 次、MA 2 次のモデルを周期 50 ポイントのパルス列で駆動した合成波形を、12 次の PARCOR、AR 8 次、MA 6 次の改良適応 ARMA ラティス、AR 8 次、MA 6 次の規格化 ARMA ラティスで分析した結果を、図 2.3 に示す。同じく図 2.2 に示されるモデルを白色ガウス雑音で駆動した合成波形を、図 2.3 と同様に PARCOR、改良適応及び規格化 ARMA ラティスで分析した結果を図 2.4 に示す。分析次数は図 2.3 の場合と同様である。図 2.3 と図 2.4 において、PARCOR のスペクトルは 120 ポイント毎、改良適応及び規格化 ARMA ラティスのスペクトルは 20 ポイント毎に表示してある。

図 2.3 の規格化 ARMA ラティススペクトルにおいて、周期的に平坦なスペクトルがみられるが、これは 100 ポイント毎に現れており、20 ポイントおきにスペクトルを取り出す際に、50 ポイント周期のパルスが入力された位置の不正確なスペクトルを取り出したものと考えられる。

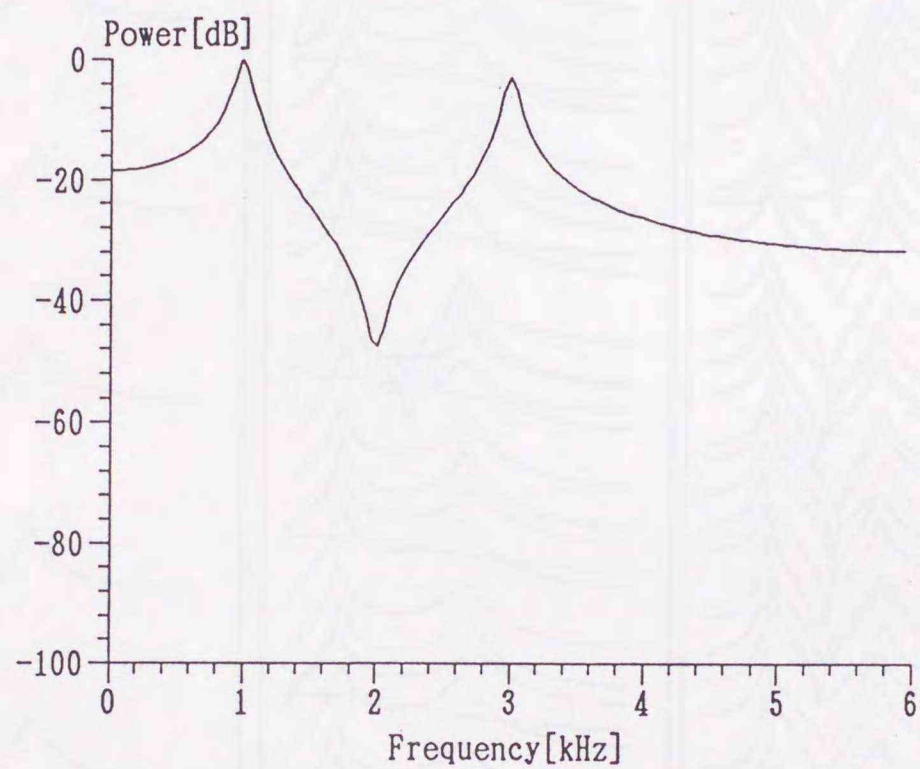
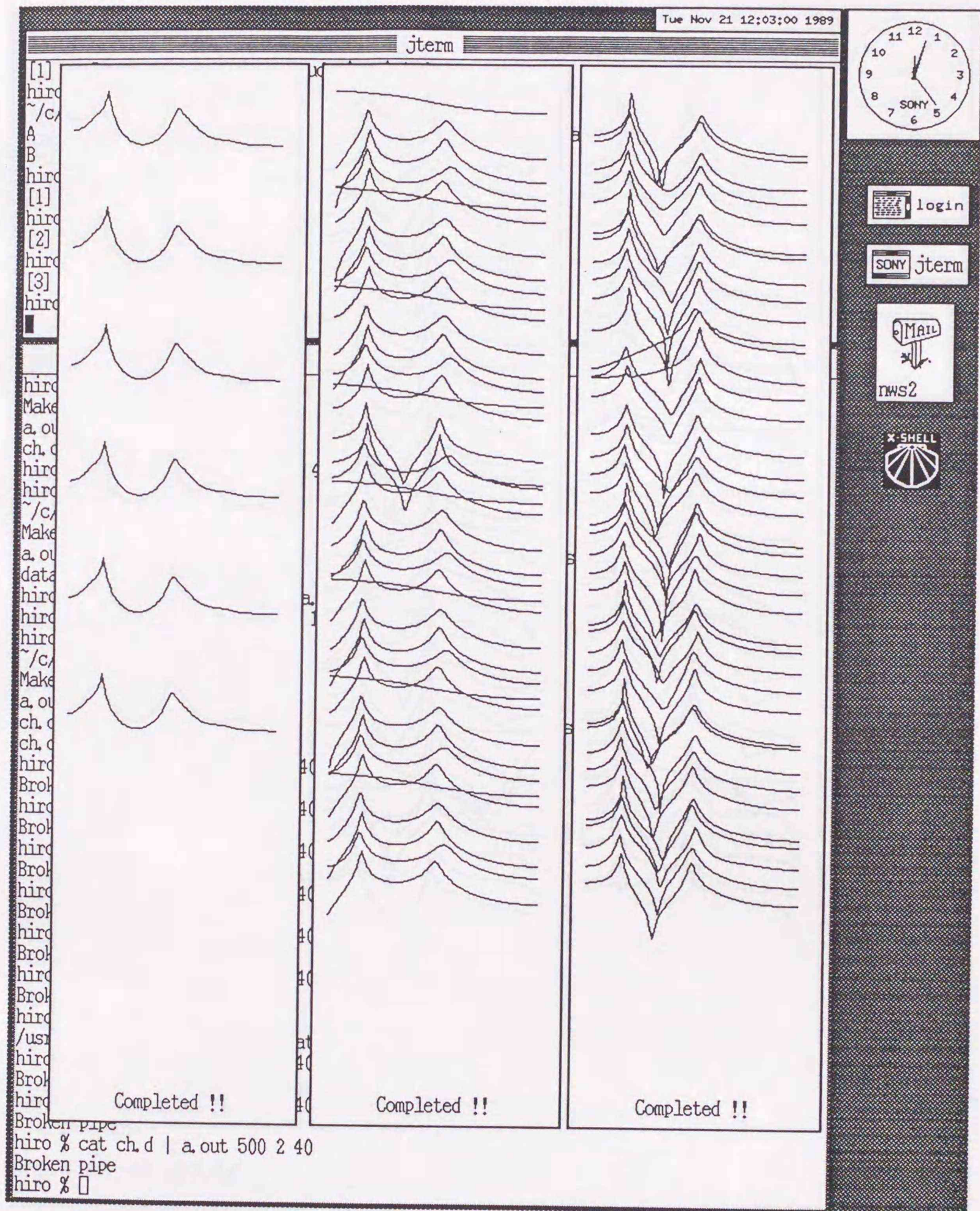


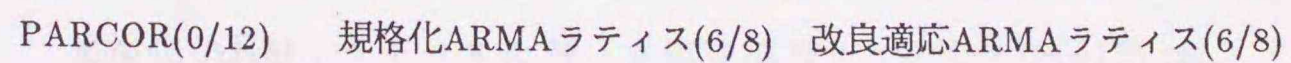
図 2.2: 比較実験に用いたモデルのスペクトル



PARCOR(0/12) 規格化ARMAラティス(6/8) 改良適応ARMAラティス(6/8)

図 2.3: PARCOR、改良適応 ARMA ラティス、規格化 ARMA ラティスのスペクトル比較

(インパルス駆動)



(ガウス雑音駆動)

図 2.3、2.4に示すように、改良 ARMA ラティスモデルと規格化 ARMA ラティスモデルのスペクトルはわずかに異なっており、どちらがより正確であるかについては、次数設定や表示スペクトルの選び方等によって変化するため一概には言明できない。

2.2.5 符号化・合成と ARMA ラティスモデル

本論文では第 2 部で、適応信号処理を用いた符号化・合成について述べるが、音声の合成・符号化でのスペクトルフィルタ、すなわち短期予測フィルタ (STP) を表すモデルを考える場合、規格化 ARMA ラティスモデル ($g_i = 1, i = 1, 2, \dots, n + m + 1$ のコンパクト ARMA ラティスモデルと等価) には、次のような利点が考えられる。

- 他の文献⁽¹⁹⁾に示されるいくつかの ARMA ラティスモデルに比べて、最少個数のパラメータでフィルタを構築可能である。
- ARMA 型であるので、フォルマントとアンチフォルマントの両方が表現できる。
- モデルが安定であるとき、各区間のパラメータの絶対値は 1 より小さくなるため、トランスヴァーサル型の ARMA 係数より量子化に向いている。
- 得られる ARMA ラティスパラメータは、トランスヴァーサル型 ARMA 係数と相互に変換可能である。

以上の利点より本論文では、量子化・伝送時の STP 係数は規格化 ARMA ラティス係数の形式で表現することとする。

量子化後の規格化 ARMA ラティスフィルタの残差計算

規格化 ARMA ラティスにおける残差信号 (入力信号) $\bar{x}(k|L)$ は、式 (2.49) および式 (2.51) によって与えられるが、各区間のラティス

係数が量子化されて、推定時と異なる値になった場合、推定アルゴリズムから得られる残差信号を入力としてラティスフィルタを駆動しても、もとの音声信号は得られない。したがって、量子化後のラティス係数を用いた逆フィルタを構成するか、もしくはそれと等価な方法によって、残差信号を新しく求め直すなくてはならない。

本論文では、ラティス型逆フィルタは構成せずに、逆フィルタと等価な方法を用いて量子化後の規格化 ARMA ラティスフィルタの残差信号を計算する。すなわち、ある時刻 k において各区間のラティス係数が求められており、各ディレイボックスに値が入っていると、まず入力 0 で規格化 ARMA ラティス順フィルタを駆動し出力信号 $\hat{y}(k)$ を得る。このときディレイボックスは更新せず、そのままの値を保つこととする。真の音声信号を $y(k)$ として、

$$\hat{u}_{q,temp}(k) = y(k) - \hat{y}(k) \quad (2.68)$$

によって得られる信号を残差信号とする。ただし残差信号を正しく求めるには利得の調整が必要である。利得の調整後の残差信号は、

$$d_i(k) = \sqrt{1 - \bar{\mu}_{i,0}^2(k)} \quad i = 0, 1, \dots, n - m - 1$$

として、

$$\hat{u}_q(k) = \frac{y(k) - \hat{y}(k)}{\bar{\tau}_{n-m,0}(k) \cdot \prod_{i=0}^{n-m-1} d_i(k)} \quad (2.69)$$

となる。この $\hat{u}_q(k)$ を再びラティスフィルタに入力し、各ディレイボックスを更新して、次の時刻 $(k+1)$ に進む。すなわち、残差信号を計算するラティスフィルタは、ゼロ入力を入れて残差を計算する部分を除けば、複号器側で音声信号を合成するフィルタと全く同じ動作をする事になる。

この疑似逆フィルタによる残差計算法は、STP 係数をはじめてから規格化 ARMA ラティス係数の形で推定するシステムで用いることと

Chapter 3

有声音の直接的分析手法

3.1 まえがき

有声音の音声生成モデルへの入力は疑似周期パルス列と見なすことができる。これは音声の生成機構を考慮しても妥当であると考えられ、また観測波形を見ても、音声生成モデルの入力は、白色ガウス過程と考えるよりはパルス列と考える方が合理的と考えられる。

従来広く行なわれている線形予測分析 (LPC⁽⁵⁾⁽⁶⁾⁽³⁾) は、

1. 入力は白色ガウス過程もしくは 1 個のインパルス
2. 分析フレーム内で定常

と仮定し、信号に相関操作を施してパラメータを推定している。

有声音では上述のように 1. の仮定は受け入れにくい。実際、有声音の音源パルスが推定結果に悪影響を及ぼすことが知られている。この問題に対処するため、分析区間を短くした LPC⁽²⁷⁾⁽²⁸⁾ や、ピッチ検出機構を持ち小さな時間範囲の信号を対象とした適応信号処理手法 (重み付き RLS 法⁽¹³⁾⁽¹⁴⁾⁽¹⁵⁾) が既に提案されている。しかしこれらのアルゴリズムはいずれも信号に対して相関操作を行ない、確率的な性質に注目して分析結果を得ている。

本来、有声音信号が疑似周期パルス列の応答であるならば、その信号の確率的な性質よりも、信号そのものが示す性質に意味があると考えべきである。

従来信号処理において、信号の確率的性質が用いられてきたのは、「実世界の信号はすべてなんらかの雑音を含むため、確率的取扱いをすることによりそれらの雑音の影響を除去できる」という考えからである。しかし、雑音に白色ガウス過程の性質を仮定しているために、理論的には、白色ガウス過程の性質を持つとみなすことのできるような雑音しか除去できない。

実際には、白色ガウス過程で記述できない突発性の雑音に対しても、ある時間範囲にわたる相関操作によってそれらの雑音の影響が平均化され、ある程度の精度を保つことが可能となっている。しかしこの場合、相関操作を行わずに求めたパラメータを時間平均することによっても、このような雑音の影響が除去できると考えられる。

また、ある時間範囲にわたった相関操作は、上述 2. の仮定を前提としている。2. の仮定が満たされないような「渡り」部分の音声信号に対しては、相関操作をすることは好ましくなく、信号から直接求めたパラメータをパラメータレベルで時間平均した方がよりよい時間応答性を持つと考えられる。

そこで本章では、相関操作を全く用いず、原信号から直接スペクトルパラメータを求める信号処理手法を導入し、その分析精度と雑音に対する推定精度の安定性について考察する⁽¹⁰¹⁾。

3.2 直接的分析手法のアルゴリズム

対象とする音声生成モデルを n 次の AR モデルとして、

$$y(k) + a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) = u(k) \quad (3.1)$$

と定義する。ただし $y(k)$ は時刻 k における出力、 $u(k)$ は時刻 k に

おけるモデルへの入力、 $a_1, a_2, \dots, a_n$ は AR 係数である。

このとき式 (3.1) より、

$$\begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-n+1) & y(k-n) \\ y(k+1) & y(k) & \cdots & y(k-n+2) & y(k-n+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y(k+n-1) & y(k+n-2) & \cdots & y(k) & y(k-1) \\ y(k+n) & y(k+n-1) & \cdots & y(k+1) & y(k) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix} = \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(k+n-1) \\ u(k+n) \end{bmatrix} \quad (3.2)$$

が得られる。

有声音における音声生成モデルの入力は疑似周期インパルス列であり、インパルスが入力された時刻以外では入力は 0 と仮定する。

時刻 k から $k+n$ までの範囲で、 n 次の線形自己回帰モデルが成立すると仮定すると、

$$\begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-n+1) & y(k-n) \\ y(k+1) & y(k) & \cdots & y(k-n+2) & y(k-n+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y(k+n-1) & y(k+n-2) & \cdots & y(k) & y(k-1) \\ y(k+n) & y(k+n-1) & \cdots & y(k+1) & y(k) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (3.3)$$

となる。式 (3.3) を解くことによって音声生成モデルの AR 係数が求められる。

ところが、時刻 k から $k+n$ までの範囲にパルスや雑音等によるなんらかの入力が存在する場合、式 (3.3) の右辺は $\mathbf{0}$ とならず、AR 係数推定に誤差を生じる。

$y(k-n)$ から $y(k+n-1)$ までの信号に対して AR 係数が正確に求められたとすると、推定誤差が生じるのは最も新しい信号 $y(k+n)$ の関係している項である。すなわち、

$$\begin{bmatrix} y(k) & y(k-1) & \cdots & y(k-n+1) & y(k-n) \\ y(k+1) & y(k) & \cdots & y(k-n+2) & y(k-n+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y(k+n-1) & y(k+n-2) & \cdots & y(k) & y(k-1) \\ y(k+n) & y(k+n-1) & \cdots & y(k+1) & y(k) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{a}_1^n \\ \vdots \\ \hat{a}_{n-1}^n \\ \hat{a}_n^n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ V_a^n(k+n) \end{bmatrix} \quad (3.4)$$

である。ただし $\hat{a}_1^n, \hat{a}_2^n, \dots, \hat{a}_n^n$ は n 次の推定 AR 係数である。

この式 (3.4) を解くと $\hat{a}_1^n, \hat{a}_2^n, \dots, \hat{a}_n^n$ と $V_a^n(k+n)$ が求められるが、この $V_a^n(k+n)$ は $y(k+n)$ になんらかの入力あるいは雑音が存在するとき 0 以外の値をとり、係数推定に誤差があることを示している。言い替えると $V_a^n(k+n) = 0$ となったとき、係数推定は正確に行なわれたことになる。この $V_a^n(k+n)$ は推定の評価値であり、他の LPC 等の評価関数とは異なり、1 点のみの誤差の値をそのまま用いていることになる。

本章で議論する直接的解析手法は、Levinson-Durbin 法を応用した回帰的手法で式 (3.4) を解くものである。解法の詳細は文献⁽⁷¹⁾に譲り、ここではその結果である推定モデルの次数を $n-1$ 次から n 次へ増加させる漸化式を示す。

$$\hat{a}_0^{n-1} = 1, \hat{b}_{-1}^{n-1} = 0, \hat{b}_m^m = 1$$

$$\begin{aligned} \hat{a}_{n-i+1}^n &= \delta_{n-1} \hat{a}_{n-i}^{n-1} + \hat{b}_{i-2}^{n-1} \quad (i = 1, 2, \dots, n) \\ \hat{b}_{i-1}^n &= \delta'_{n-1} \hat{a}_{n-i}^{n-1} + \hat{b}_{i-2}^{n-1} \quad (i = 1, 2, \dots, n) \end{aligned} \quad (3.5)$$

$$\delta_{n-1} = -\frac{V_b^{n-1}}{\tau_a}, \delta'_{n-1} = -\frac{\tau_b}{V_a^{n-1}} \quad (3.6)$$

$$\begin{aligned} V_a^n(k+n) &= \delta_{n-1} V_a^{n-1}(k+n-1) + \tau_b \\ V_b^n(k+n) &= \delta'_{n-1} \tau_a + V_b^{n-1}(k+n-1) \end{aligned} \quad (3.7)$$

$$\tau_a = \sum_{i=1}^n y(k-i) \hat{a}_{i-1}^{n-1}, \tau_b = \sum_{i=1}^n y(k+i) \hat{b}_{i-1}^{n-1} \quad (3.8)$$

ここで、各漸化式の初期値は、

$$V_a^0(k) = V_b^0(k) = y(k), \hat{a}_0^0 = \hat{b}_0^0 = 1 \quad (3.9)$$

である。

3.3 モデル実験

前節で述べた直接的分析アルゴリズムを用いて、合成信号の分析モデル実験を行なった。

使用した合成信号の特性を表 3.1、図 3.1 に示す。合成音声の特性は成人男性の母音に似せて設定した。分析次数は、精度の面からは高い次数が望ましいが、次数を高くするにつれ入力が存在しない分析点が少なくなるため、ここでは経験的に 35 次とした。

分析結果を図 3.2 に示す。図 3.2 では 1 周期分のすべてのサンプルに対するデータが示されている。(a) は $V_a(k+n)$ の値、(b)~(d) は直接的手法で求められたスペクトルであるが、(c) は $V_a(k+n)$ の 2 乗の個数 5 個の平均 $(\frac{1}{5} \cdot \sum_{i=k-4}^k (V_a(i))^2)$ が 2×10^4 以下のスペクトルのみを選択したもの、(d) はさらに選択されたスペクトルのみの間で個数 10 個で平均操作したものである。また比較のため、LPC スペクトル (自己相関法 12 次) を (e) に示す。このスペクトルは本来 1 フレーム内 (360 サンプル) で 1 つのパターンしか求められないため、フレームシフト幅を 1 サンプルとして、すべてのサンプルに対して表示した。

図 3.2 の実験結果より、分析点によっては LPC より高精度な推定スペクトルが得られるが、自由応答区間とみられる範囲内を推定パラメータレベルで平均操作してもなお、かなりの不安定さがみられる。この不安定さは種々の雑音 (量子化誤差をも含む) によるものと考えられる。

以上のことから、原信号から直接パラメータを求めるアルゴリズム

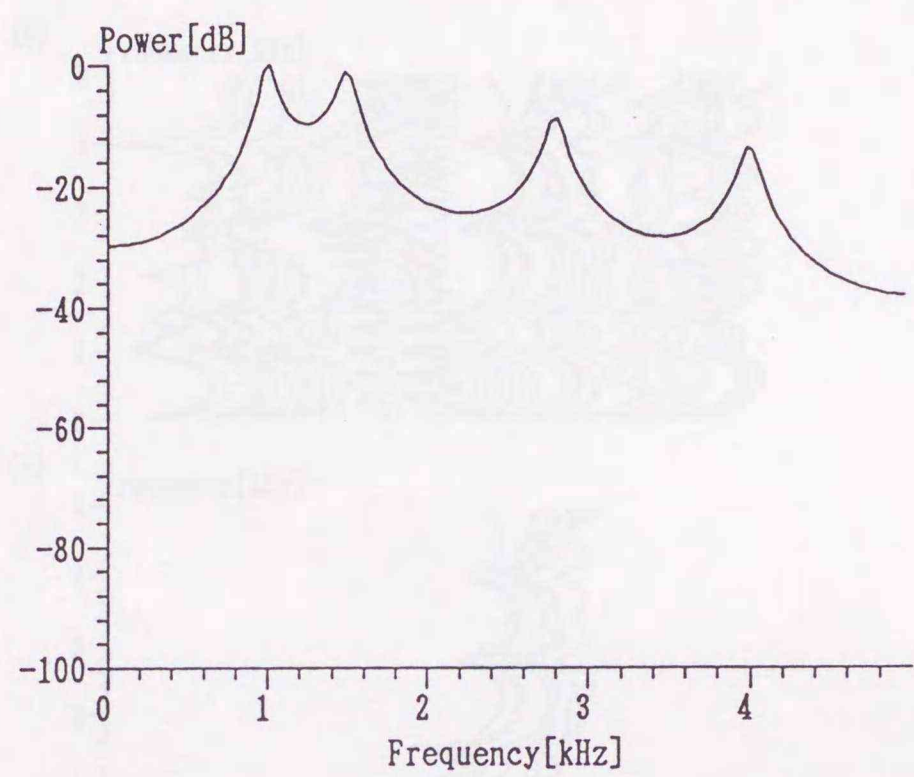


図 3.1: 合成信号の特性

サンプリング周波数の設定		10 kHz			
入力	種別	周期インパルス列			
	振幅	100			
	周期	50 サンプル			
フォルマント	周波数 (kHz)	1.0	1.5	2.8	4.0
	帯域 (kHz)	0.08	0.1	0.1	0.1

表 3.1: 合成信号の特性

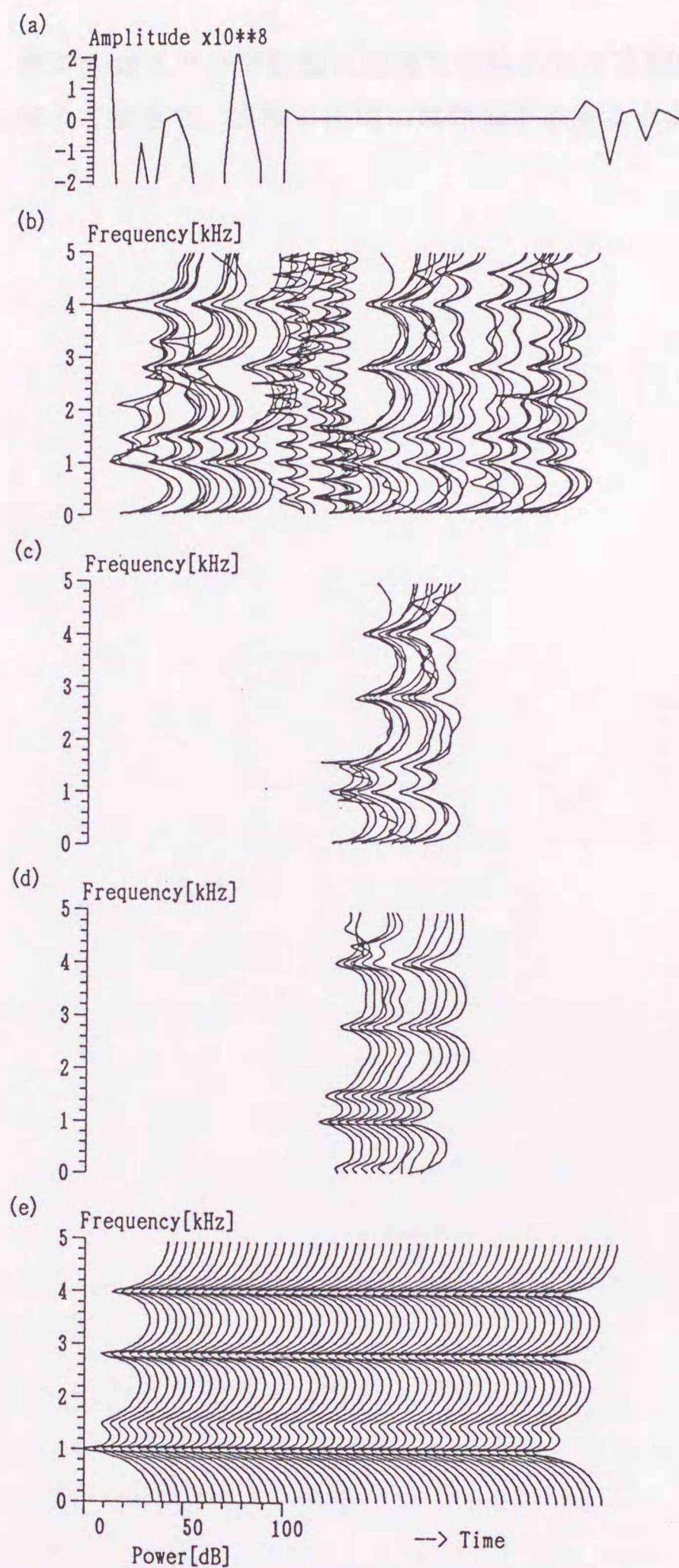


図 3.2: 直接的分析法による分析

(a) V_a (b) スペクトル (c) 選択されたスペクトル
(d) 平均したスペクトル (e) LPC

ムは、音声生成モデルの性質に忠実で理論的には高精度であるが、雑音に極めて敏感で、実際の使用には問題があることが示された。

Chapter 4

重み付きRLS法における入の周波数特性

4.1 まえがき

前章において信号に対して初期条件を与える方法、逐次最小二乗法について検討し、適用文法に必要となる計算量を評価した。本章では、逐次最小二乗法の応用であることが明らかになる。

本章で扱われる入の重み付きRLS法は、初期条件を与える方法を用いて、逐次最小二乗法の応用であることが明らかになる。この重み付きRLS法は信号に対して初期条件を与える方法を用いて、逐次最小二乗法の応用であることが明らかになる。この重み付きRLS法は信号に対して初期条件を与える方法を用いて、逐次最小二乗法の応用であることが明らかになる。

本章では、入の重み付きRLS法を用いて、逐次最小二乗法の応用であることが明らかになる。この重み付きRLS法は信号に対して初期条件を与える方法を用いて、逐次最小二乗法の応用であることが明らかになる。この重み付きRLS法は信号に対して初期条件を与える方法を用いて、逐次最小二乗法の応用であることが明らかになる。

Chapter 4

重み付き RLS 法における λ の周波数特性

4.1 まえがき

第 3 章において信号に対して相関操作を行なわない直接的分析法について検討し、高精度でかつ安定な分析結果を得るためには、ある程度の相関操作が必要であることが確かめられた。

既に提案されている重み付き RLS 法は、評価関数に重み係数 λ を導入し、過去の信号より現在の信号に大きな重みをかけてスペクトルパラメータを求める適応信号処理アルゴリズムである⁽¹³⁾⁽¹⁴⁾⁽¹⁵⁾。この重み付き RLS 法は信号に対して相関を計算する操作を行なうアルゴリズムであるが、 λ によって、現在時点の近傍の信号をスペクトル推定に用いることになり、ピッチ構造をもつ有声音に適することが知られている⁽¹⁶⁾。

本章では、 λ を窓関数と見なし、従来明らかにされていなかった λ の周波数特性について考察する⁽¹⁰¹⁾。すなわち、 λ の周波数特性と矩形窓関数の周波数特性との比較から、 λ が、対象とする信号の時間範囲を限定する有効な窓関数の働きを持つことを述べ、 λ の性質に関する従来の経験的な理解⁽²⁹⁾⁽³⁰⁾への裏付けを与える。

4.2 λ の周波数特性

重み付き RLS 法では、第 2 章式 (2.13) の評価関数

$$V(k) = \sum_{i=0}^k \lambda^{k-i} (y(i) - \hat{y}(i|k))^2 = \sum_{i=0}^k \lambda^{k-i} \nu^2(i|k) \quad (4.1)$$

を最小にするようにモデルの係数を決定している。 λ の周波数特性を調べるために次の関数を導入する。

$$f(i) = \begin{cases} \lambda^{k-i} & (0 \leq i \leq k) \\ 0 & (i < 0, i > k) \end{cases} \quad (4.2)$$

このとき、式 (4.1) は、

$$V(k) = \sum_{i=-\infty}^{\infty} f(i) \nu^2(i|k) \quad (4.3)$$

となる。

$f(i)$ の離散フーリエ変換を $F(\omega)$ とすると、

$$F(\omega) = \sum_{i=-\infty}^{\infty} f(i) e^{-j\omega T i} = \frac{\lambda^{k+1} - e^{-j\omega T(k+1)}}{\lambda - e^{-j\omega T}} \quad (4.4)$$

ただし、 T はサンプリング周期である。

$\nu^2(i|k)$ の離散フーリエ変換を $N(\omega)$ とすると、

$$N(\omega) = \sum_{i=-\infty}^{\infty} \nu^2(i|k) e^{-j\omega T i} \quad (4.5)$$

$x(i) = f(i) \nu^2(i|k)$ とおくと、 $x(i)$ の離散フーリエ変換 $X(\omega)$ は、

$$X(\omega) = \sum_{i=-\infty}^{\infty} f(i) \nu^2(i|k) e^{-j\omega T i} = \sum_{\Omega=-\infty}^{\infty} F(\Omega) N(\omega - \Omega) \quad (4.6)$$

すなわち、 $x(i)$ の周波数特性 $X(\omega)$ は、 $F(\omega)$ と $N(\omega)$ のたたみこみとなる。

ところで、重み付き RLS 法の評価関数 $V(k)$ は式 (4.3) より、

$$\begin{aligned}
 V(k) &= \sum_{i=-\infty}^{\infty} f(k) \nu^2(i|k) = \sum_{i=-\infty}^{\infty} x(i) \\
 &= \sum_{i=-\infty}^{\infty} f(i) \nu^2(i|k) e^{-j \cdot 0 \cdot T i} = X(0)
 \end{aligned} \tag{4.7}$$

したがって重み付き RLS 法の評価関数は、 $F(\omega)$ と $N(\omega)$ のたたみこみで得られる $X(\omega)$ の直流分 ($\omega = 0$) となる。

重み付き RLS 法はこの $V(k) = X(0)$ を最小にするよう係数を決定する。

ところで、

$$N(0) = \sum_{i=-\infty}^{\infty} \nu^2(i|k) \tag{4.8}$$

は、信号にエルゴート性を仮定すると、LPC における評価関数に対応している。したがって、重み付き RLS 法においてモデルの係数が正確に推定されるためには、関数 $F(\omega)$ が、 $X(0)$ に $N(0)$ を忠実に再現するような窓関数の特性を持っていなければならない。

なお、 $\lambda = 1$ のとき式 (4.2) は、

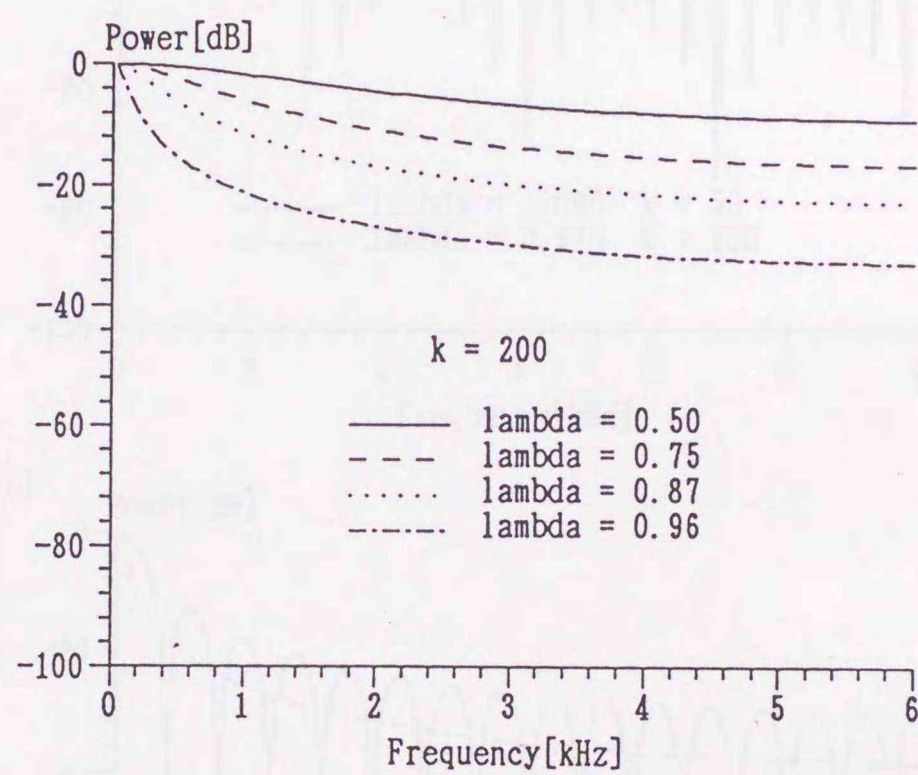
$$f(i) = \begin{cases} 1 & (0 \leq i \leq k) \\ 0 & (i < 0, i > k) \end{cases} \tag{4.9}$$

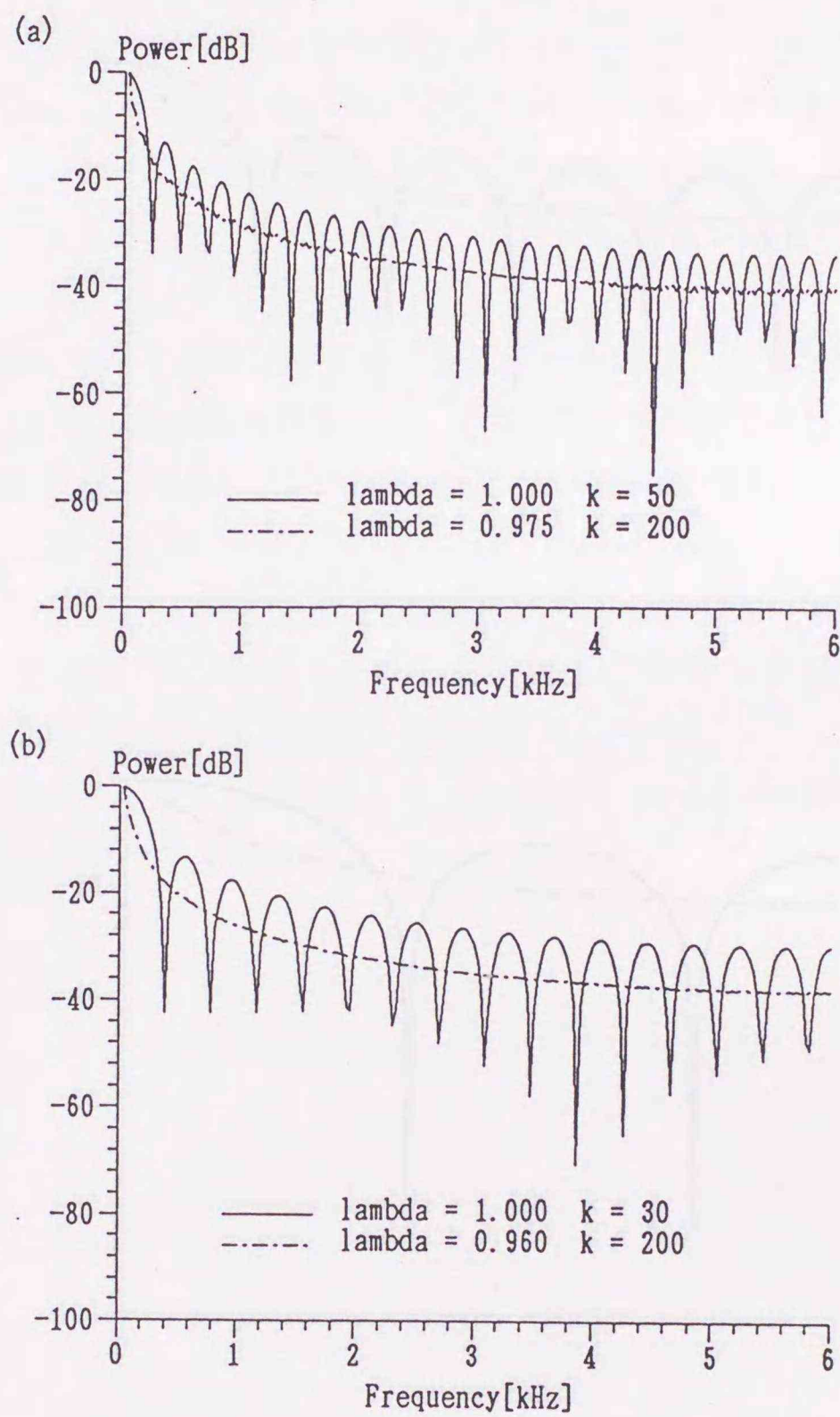
となり $f(i)$ は矩形窓関数となる。

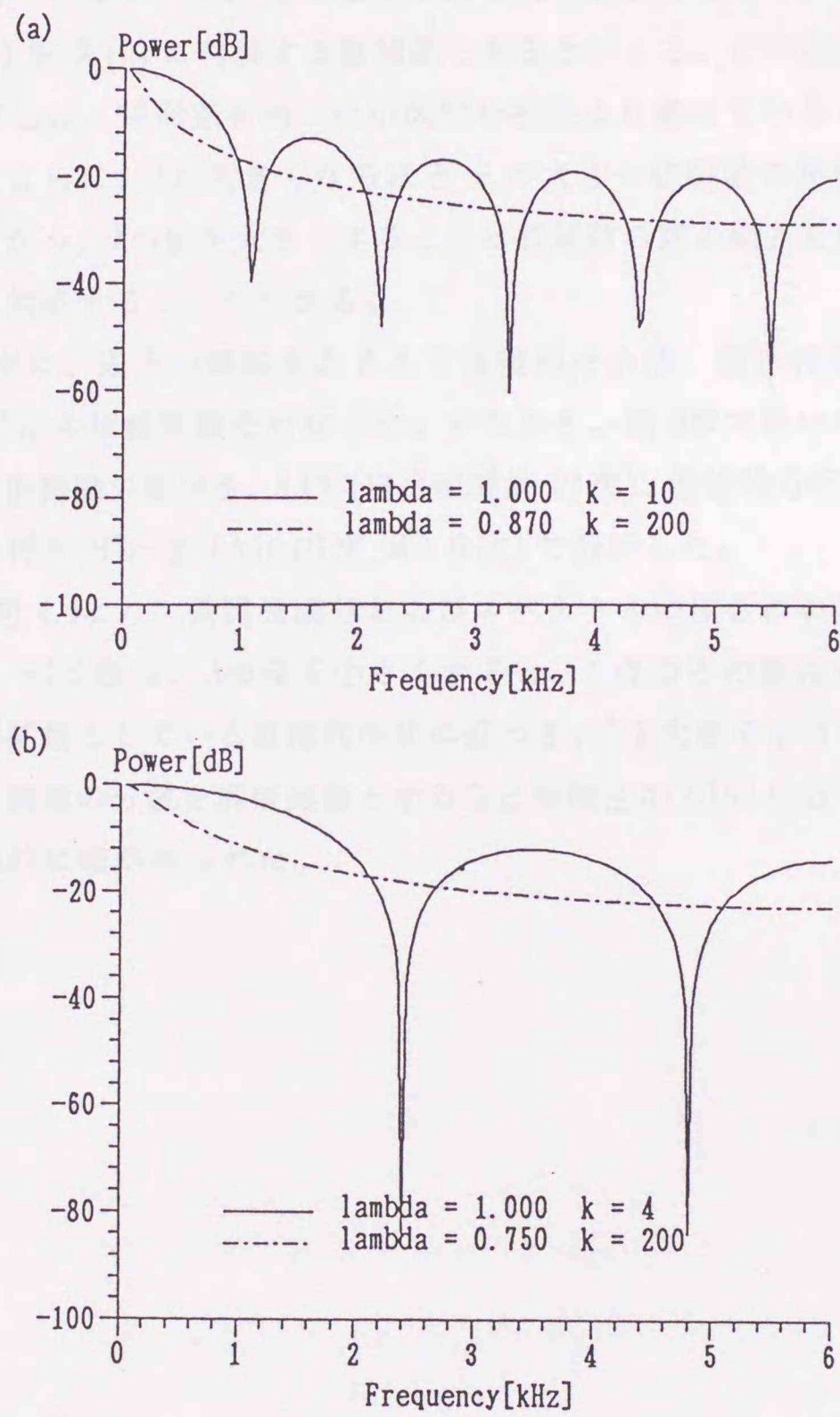
式 (4.4) より λ を様々に変化させて $F(\omega)$ のパワースペクトル $|F(\omega)|$ を求めたものを、図 4.1 に示す。

$\lambda = 1$ とし、 k を種々の値に変化させた矩形窓関数の特性と、 $\lambda < 1$ 、 $k = 200$ とした $f(i)$ の特性との比較を図 4.2、4.3 に示す。

$k = 50$ の矩形窓関数には、 $\lambda = 0.975$ の $f(i)$ が最もよく適合し、 $k = 30$ では $\lambda = 0.96$ 、 $k = 10$ では $\lambda = 0.87$ 、 $k = 4$ では $\lambda = 0.75$ がそれぞれよく適合している。ただし $\lambda < 1$ のときの k の値は、 $k > 200$ であればほとんどその周波数特性に影響を与えない。したがって $\lambda < 1$ における λ は、狭い範囲を取り出す窓関数とみなす

図 4.1: λ の周波数特性

図 4.2: λ の周波数特性と矩形窓の周波数特性の比較

図 4.3: λ の周波数特性と矩形窓の周波数特性の比較

ことができる。また、 λ の周波数特性は対応する矩形窓関数の特性に比べて $\omega = 0$ の近傍が急激に鋭く立ち上がっており、より忠実に $N(0)$ を $X(0)$ に再現する窓関数であるといえる。この点で重み付き RLS 法は、矩形窓を用いた小区間分析法より優れていることが分かる。さらに、 λ が大きくなるほど k の大きい窓関数の特性に近づくことから、 λ の値を大きくすることは窓関数の窓の範囲を広くすることに対応することが分かる。

次に、以上の議論をふまえて直接的分析法、重み付き RLS 法、LPC 法の比較実験を行なった。すなわち、第 3 章で用いた合成信号の自由振動の部分を、LPC(自己相関法 12 次)、直接的分析法 (35 次)、重み付き RLS 法 (AR:16 次, MA:0 次) で分析した。

図 4.4 より、低周波成分およびスペクトルの傾きにおいて、重み付き RLS 法は、 λ の値を小さくすると、1 点のみの誤差そのものを評価関数としている直接的手法に近づき、 λ を大きくすると、広い範囲の誤差の分散を評価関数とする自己相関法の LPC に近づくことが実験的に確かめられた。

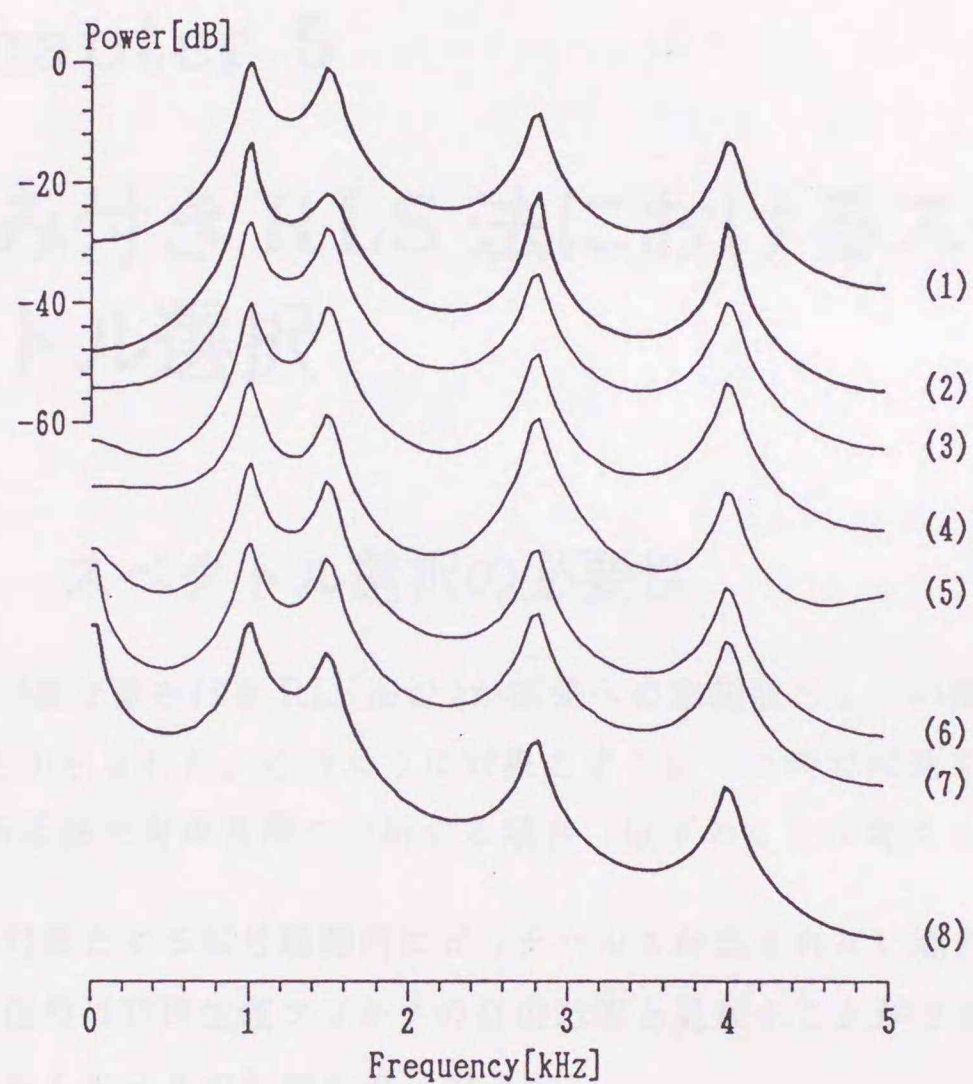


図 4.4: 直接的分析法、重み付き RLS 法、LPC の比較

(1) 合成信号 (2) LPC (3)-(7) 重み付き RLS
 $\lambda =$ (3) 0.95 (4) 0.90 (5) 0.80 (6) 0.65 (7) 0.50
 (8) 直接的分析法

Chapter 5

重み付き RLS 法におけるスペクトル選択

5.1 スペクトル選択の必要性

第 4 章で重み付き RLS 法の λ が誤差への窓関数としての働きを持つことが示された。このように対象とする信号の時間範囲を限定した分析手法で有声音声を分析する場合、以下のことが考えられる。

1. 対象とする信号範囲内にピッチパルスが含まれない場合、その信号は音声生成フィルタの自由応答と見なすことができ、精度良くモデルの係数を推定できる。
2. 対象とする信号範囲内にピッチパルスが含まれる場合、その信号は自由応答とは見なせないため、係数推定の精度は悪化する。

有声音声はピッチパルスが疑似周期的に現われるため、 $\lambda < 1$ とした重み付き RLS 法を用いると分析精度は原理的にピッチパルスに同期して上下することになる⁽¹⁵⁾。これは従来提案されている小区間分析法においても指摘されていること⁽²⁷⁾⁽²⁸⁾であり、本手法においても、よりよい分析スペクトルを得るためには、上述 1. の場合の分析結果のみを選択し抽出しなくてはならない。

実際に第 3 章で用いた合成音声を、 $\lambda = 0.80$ とした重み付き RLS 法 (AR:16 次, MA:0 次) を用いて時間的に連続して分析し、 $\lambda = 1$ の矩形窓関数の場合と比較したのが図 5.1 である。図 5.1 は 2 サンプルごとに表示している。

図 5.1 より、 $\lambda = 0.8$ とすると精度に周期的な変化がみられること、精度の良い部分では $\lambda = 1$ としたときよりも高精度であることが示され、 $\lambda < 1$ として誤差値が小さく安定している自由振動の部分を選択することによって、高精度な分析が行えることが確認された。

5.2 スペクトル選択手法

前節で述べたような分析スペクトル選択作業は、従来目視により手作業で行なわれていた。本節ではこの作業の自動化を以下の 2 つの手法それぞれについて試みる。すなわち、経験的 if-then ルールの蓄積による知識工学的手法、およびしきい値操作 if-then ルールと等価なニューラルネットワークのノードを学習させて選択する手法の 2 種類である。

5.2.1 if-then ルールによる選択システム

まず if-then ルールを経験的に蓄えて選択を行なうシステムについて考察する。

データ選択のために用いたルールは以下の通りである。

$$\overline{\nu_S^2(k)} = \frac{1}{10} \sum_{i=k-9}^k \nu^2(i) \quad (5.1)$$

$$\overline{\nu_L^2(k)} = \overline{\nu_L^2(k-1)} + \frac{\nu^2(k) - \overline{\nu_L^2(k-1)}}{50} \quad (5.2)$$

として、

[ルール 1] 予測誤差 ν の絶対値が 30 未満。

[ルール 2] $\overline{\nu_S^2}$ が $\overline{\nu_L^2}$ の 1/2 未満。

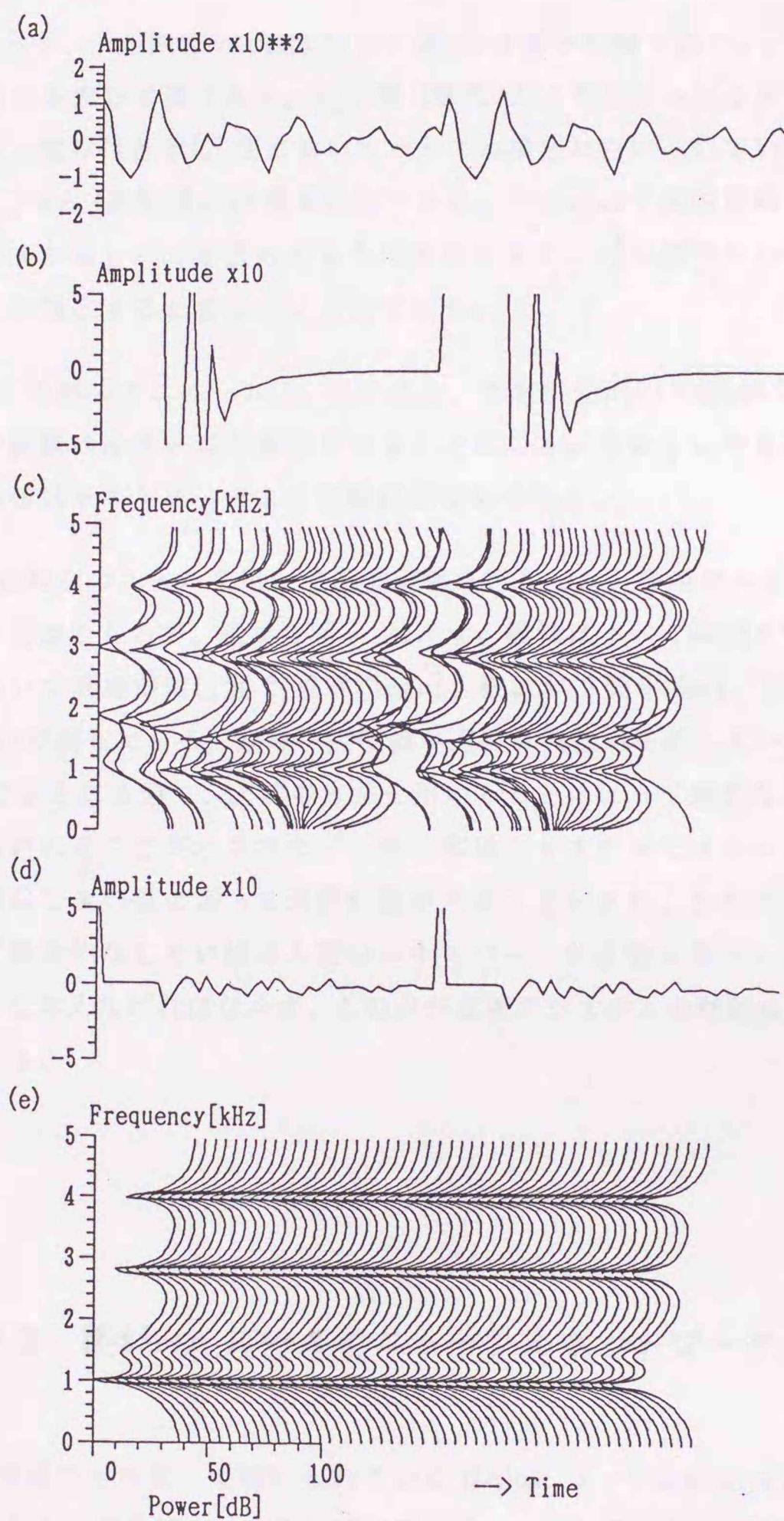


図 5.1: 重み付き RLS 法による分析 (合成信号)

(a) 波形 (b) $\lambda = 0.8$ の $\nu(k)$ (c) $\lambda = 0.8$ のスペクトル
 (d) $\lambda = 1.0$ の $\nu(k)$ (e) $\lambda = 1.0$ のスペクトル

$\overline{\nu_S^2}$ は、予測誤差 ν の短区間(10点)の2乗平均値であり、 ν の変動の度合を表わす量である。 $\overline{\nu_L^2}$ は第2章式(2.9)で求められるものと同じで、重み付き RLS 法においてパルス検出に用いられている量であり、 ν の(長区間の)2乗平均値である。この $\overline{\nu_L^2}$ は予測誤差の2乗のおおまかなレベルを表わすような変化をする。 $\overline{\nu_L^2}$ は信号のレベルの変化に対応するために用いられている。

したがって、ルール 1、2 により、予測誤差 $\nu(k)$ の値が小さく、かつ変動の小さい点を選択されることになる。なおここでも具体的な数値はテストデータより経験的に求められた。

これらのルールを用いて選択実験を行なった。使用データは第6章と同様のもので、結果を図5.2に示す。横軸はすべて時間 t で、(a)において実線が短区間2乗平均値 $\overline{\nu_S^2}$ 、破線が2乗平均 $\overline{\nu_L^2}$ 、(b)の実線は ν の変化である。縦の一点鎖線が選択された点を示している。図5.2に示される通り、上述の方法を用いることによって適切な選択が行なわれることが示された。しかし本項のシステムではルールの具体的なしきい値によって選択性能が大きく左右され、それにも関わらず具体的なしきい値は人間のエキスパートが経験に基づいてシステムに与えなければならず、この点が本項のシステムの問題点となっている。

5.2.2 if-then ルールとニューラルネットワーク

前項のシステムで用いられている if-then ルールの中には、条件部にしきい値を持つものが必ず含まれる。しきい値操作を行なうルールの典型的な形を以下に示す。

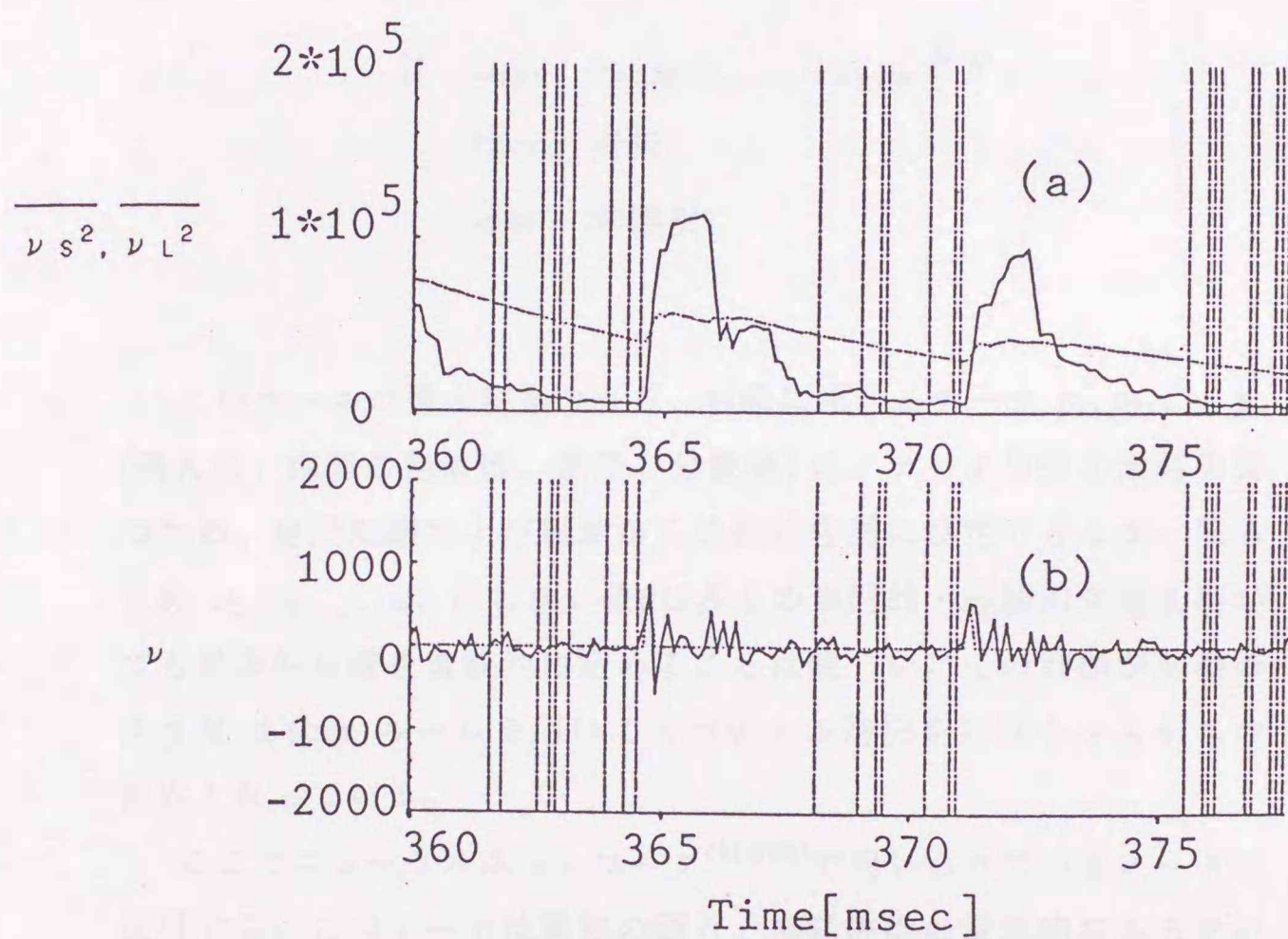


図 5.2: if-then ルールによる選択システムの選択結果

縦の一点鎖線：選択された点

(a) 実線： $\overline{\nu_s^2}$ 、破線： $\overline{\nu_L^2}$

(b) 実線： ν

$p_1, p_2, \dots, p_n$: 判断に用いるデータ
 $\omega_1, \omega_2, \dots, \omega_n$: 定数
 θ : しきい値

if $\omega_1 p_1 + \omega_2 p_2 + \dots + \omega_n p_n \geq \theta$
 then 選択;
 else 非選択;

このルールの構成要素のうち、判断に用いるデータ $p_1, p_2, \dots, p_n$ (例えば、誤差の絶対値、誤差の分散等) は、データ自体が意味を持つため、専門知識および経験から比較的容易に設定できるが、残る定数 $\omega_1, \omega_2, \dots, \omega_n$ としきい値 θ は多くの専門的・経験的知識を用いても具体的な値を客観的に定めることは難しい。この問題が前項のような if-then ルールを用いてスペクトル選択を行なうシステムの限界となっている。

ここでニューラルネットワーク⁽³¹⁾⁽³²⁾で用いられているノードに注目する。このノードは周知の通り、神経細胞の電氣的なふるまいをモデリングしたもので、以下のような構造を持つ。

$o_{in.1}, o_{in.2}, \dots, o_{in.n}$: 判断に用いるデータ
 $\omega_1, \omega_2, \dots, \omega_n$: 定数
 o_{out} : 出力
 f : 非線形関数

$$o_{out} = f(\omega_1 o_{in.1} + \omega_2 o_{in.2} + \dots + \omega_n o_{in.n})$$

非線形関数には、式 (5.3) で表わされるしきい値関数、あるいは式 (5.4) で表わされる sigmoid 関数等が用いられる。

$$f(i) = \begin{cases} 1 & i \geq \theta \\ 0 & i < \theta \end{cases} \quad (5.3)$$

$$f(x) = \frac{1}{1 + e^{-(x+\theta)}} \quad (5.4)$$

以上のようなノードと if-then ルールを比較すると、ニューラルネットに用いられている1個のノードはしきい値操作を行なう if-then ルールと等価であることが分かる。

そこで本項では、前節のスペクトル選択システムの if-then ルールをニューラルネットのノードに置き換え、定数 $\omega_1, \omega_2, \dots, \omega_n$ としきい値 θ を学習データからバックプロパゲーションアルゴリズム⁽³¹⁾を用いて獲得させることにより選択を行なう。ノードへの入力パラメータは経験的に、前節の if-then ルールによるシステムと同じものを用いる。このようにすることによって従来の if-then ルールを用いる方法より客観性の高い選択が可能であり、また入力パラメータの選択の段階で経験的知識が有効に盛り込まれるため、本来のニューラルネットのような多ノードネットワークを構成する必要がなくなり、効率のよい選択が可能となる。

本節ではノードの非線形関数として式 (5.4) の sigmoid 関数を用い、しきい値 θ を学習によって得るために常に 1 が入力される入力線を 1 本設定した。本節で用いるノードを図 5.3 に示す。

5.2.3 ニューラルネットワークによる選択システム

前項で述べたノードを用いて、 $\lambda = 0.8$ とした重み付き RLS 法 (AR:16 次, MA:0 次) で得られた母音スペクトルデータの選択実験システムを作成した。

ノードへの入力パラメータは 5.3.1 で述べた $\overline{\nu_S^2}$ と $\overline{\nu_L^2}$ 、および誤差 $\nu(k)$ の大きさを表わすパラメータとして $\nu^2(k)$ を用いた。それぞれのパラメータのダイナミックレンジは予測できず、ニューラルネットワークのノードは sigmoid 関数の性質から大きな入力では出力が

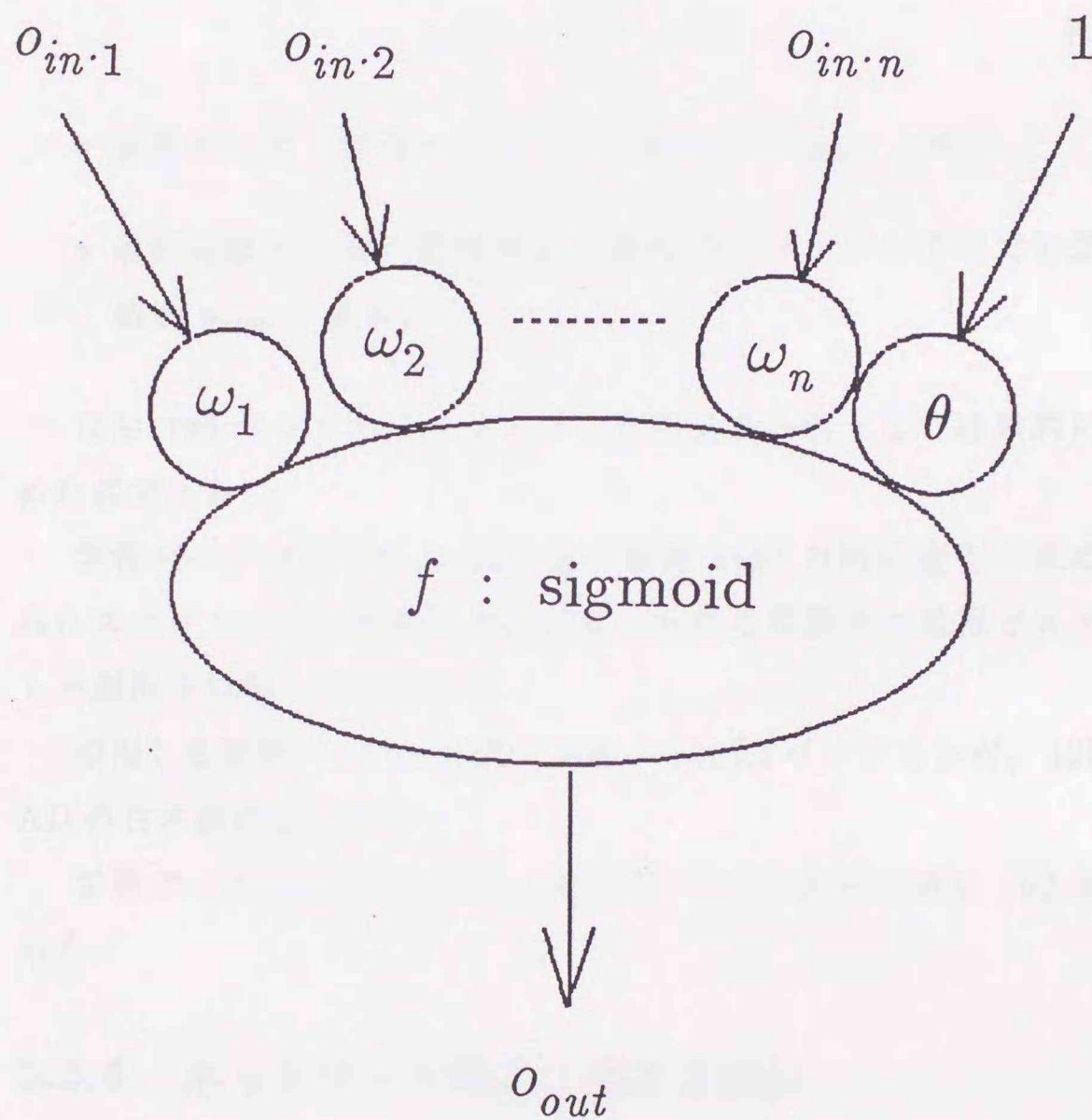


図 5.3: 本論文で用いるノードの構造

飽和してしまうため、ノードへ入力する際には以下の方法で最大値 2.0 へ規格化した。時刻 k におけるパラメータを $p(k)$ 、規格化されたパラメータを $p_n(k)$ とすると、

$$p_n(k) = \frac{p(k)}{p_{max}} \times 2.0 \quad (5.5)$$

- 学習のとき：学習サンプル内の最大値を p_{max} とする。
- 選択実験のとき：現時刻より過去 140 サンプルの中での最大値を p_{max} とする。

なお 140 サンプルは 1 ピッチより充分長い値として経験的に定めた値である。

学習データは重み付き RLS 法の誤差 $\nu(k)$ の時間変化と求められたスペクトルの時間変化を表示し、それらを参考に目視でスペクトル選択を行なって作成した。

使用した音声サンプルは成人男性、12kHz サンプリング、12bit-AD の日本語母音である。

学習データを図 5.4 に示す。学習データのパターン数は 162 個である。

5.2.4 ネットワーク構造に関する検討

前項で述べた実験システムにより選択実験を行なった。前項の入力パラメータの設定からノード 1 個で選択可能であることが予想されたが、パラメータの選択は経験的なものであり最適でない可能性もあるため、選択実験は以下の 3 種類のニューラルネットを用いて行なった (図 5.5)。

- A) 第 1 層 3 個、第 2 層 2 個の 2 層
- B) 第 1 層 3 個、第 2 層 1 個の 2 層

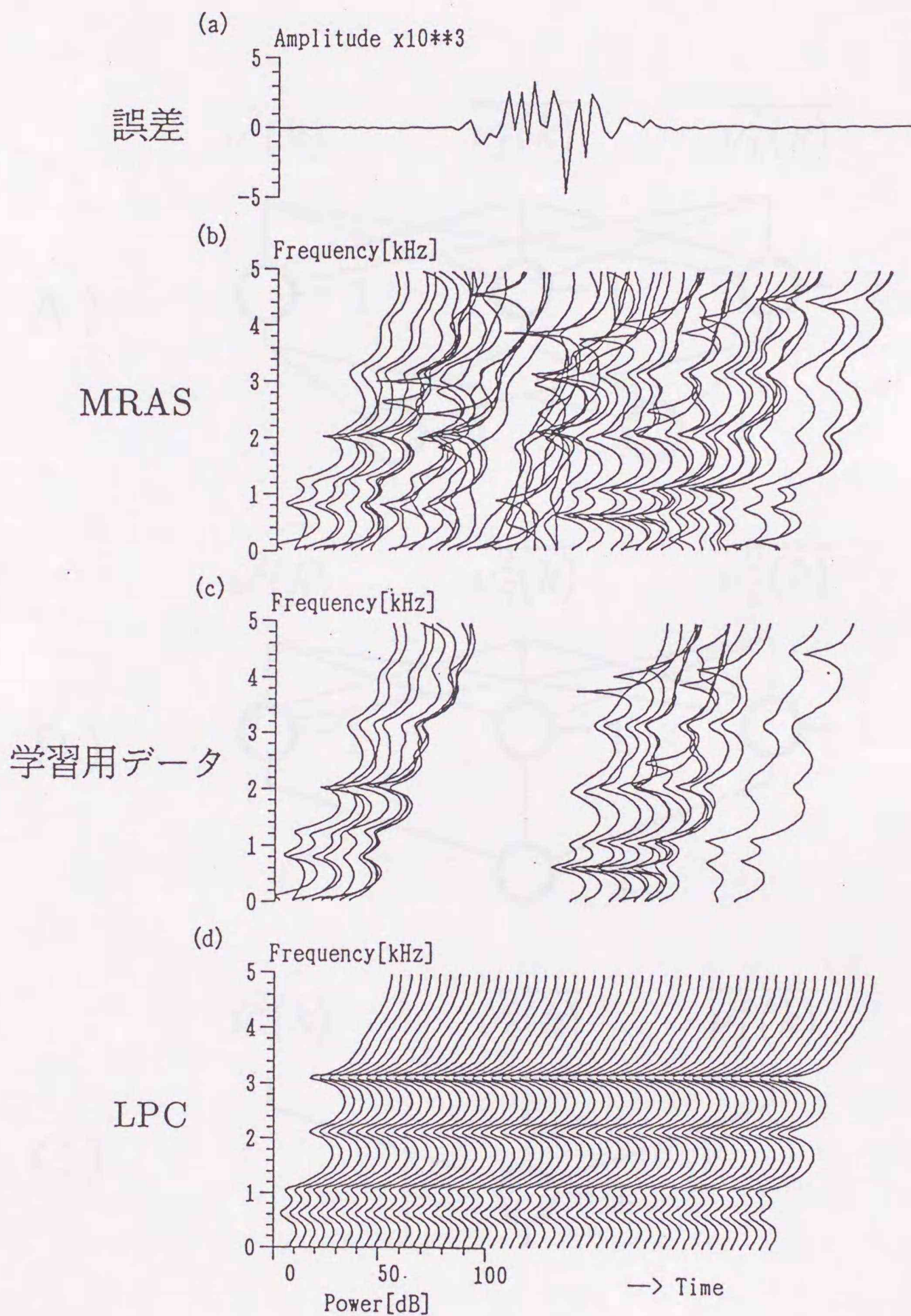


図 5.4: 学習用データ (話者 1/a/)

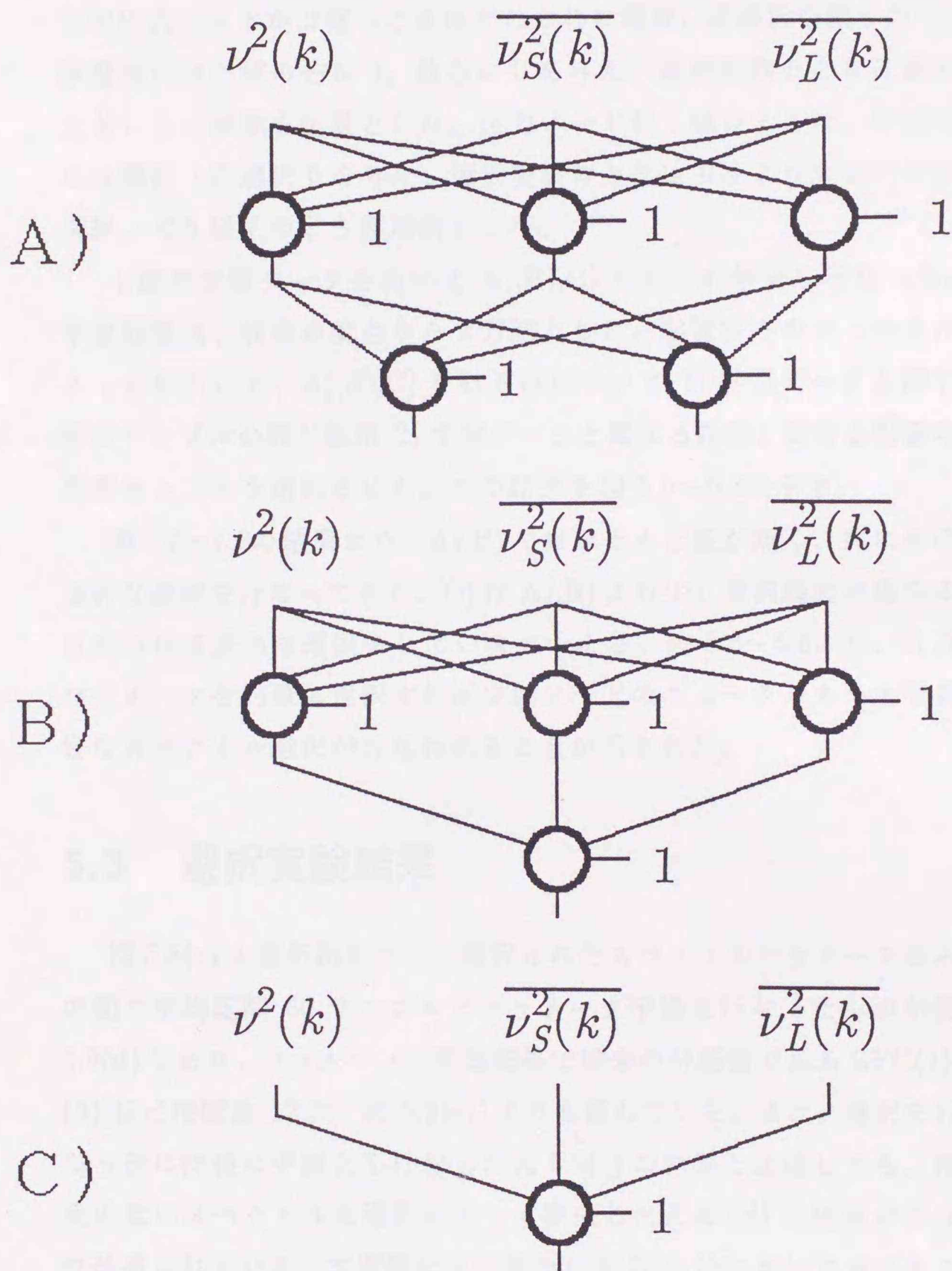


図 5.5: 3 種類のニューラルネットワークの構造図

C) 第1層1個の1層

なお出力ノードが2個のときはそれぞれに選択、非選択を割り当て、学習時には正解の方に1、他方に0を与え、選択実験のときは値の大きい方の事象を結果とした。出力ノードが1個のときは、学習時には選択1非選択0を与え、選択実験のときは0.5より大きいとき選択、0.5以下のとき非選択とした。

上述の学習データを用いて A), B), C) それぞれ学習を行なった。学習回数は、収束の度合から2万回とした。学習済みのニューラルネットを用いて、A), B), C) それぞれについて 1) 学習データと同じ音声サンプルの同じ箇所 2) 学習データと異なる話者、異なる母音の音声サンプルを選択させた。その結果を図5.6~5.8に示す。

図5.6~5.8の結果より、A), B) ではほとんど差が無く、共にほぼ良好な選択を行なっており、C) は A), B) より少し選択精度が落ちるもののほぼ妥当な選択をしているといえる。図5.6~5.8より、入力パラメータを的確に選択すれば少数ノードのニューラルネットで良好なスペクトル選択が行なわれることが示された。

5.3 選択実験結果

図5.8(e)の選択結果から、選択されたスペクトルパラメータのみの間で平均区間30サンプルでパラメータ平滑を行なったものが図5.9(d)であり、フォルマント帯域幅等で従来の分析法であるLPC(1)-(3)(自己相関法12次、図5.9(e))よりも優れている。また、選択を行なう前に同様に平滑化を行なった図5.9(c)の結果と比較しても、精度の低いスペクトルを選択によって除去したため(d)は結果がかなり改善されている。本実験により重み付きRLS法においてスペクトルを選択することの有効性が確認された。本論文ではこのニューラルネットワークの1個のノードによるシステムを、重み付きRLS法のスペクトル選択システムとして採用する。

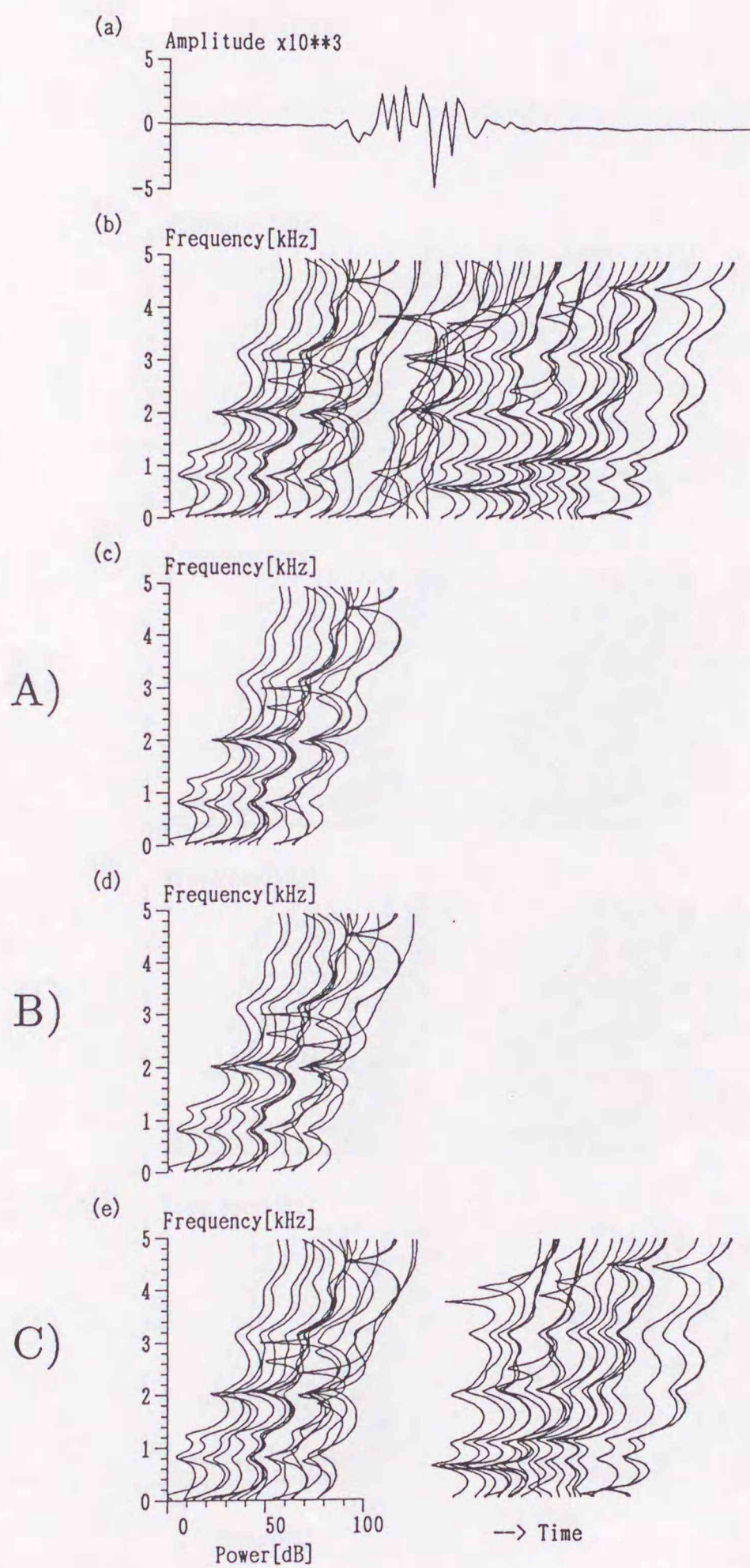


図 5.6: ニューラルネットのノードによる選択システムの選択結果
(話者 1/a/)

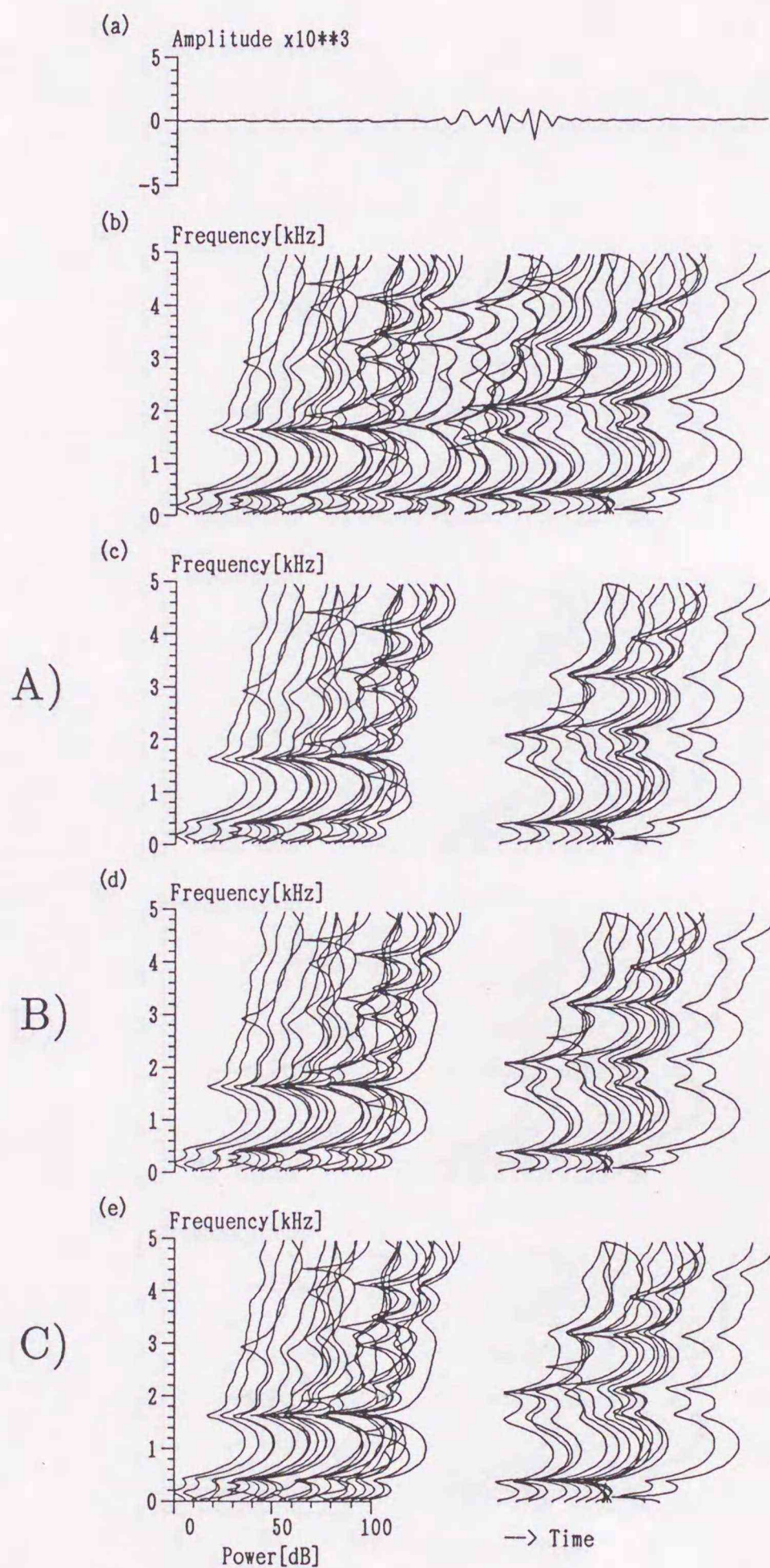


図 5.7: ニューラルネットのノードによる選択システムの選択結果
(話者 2/e/)

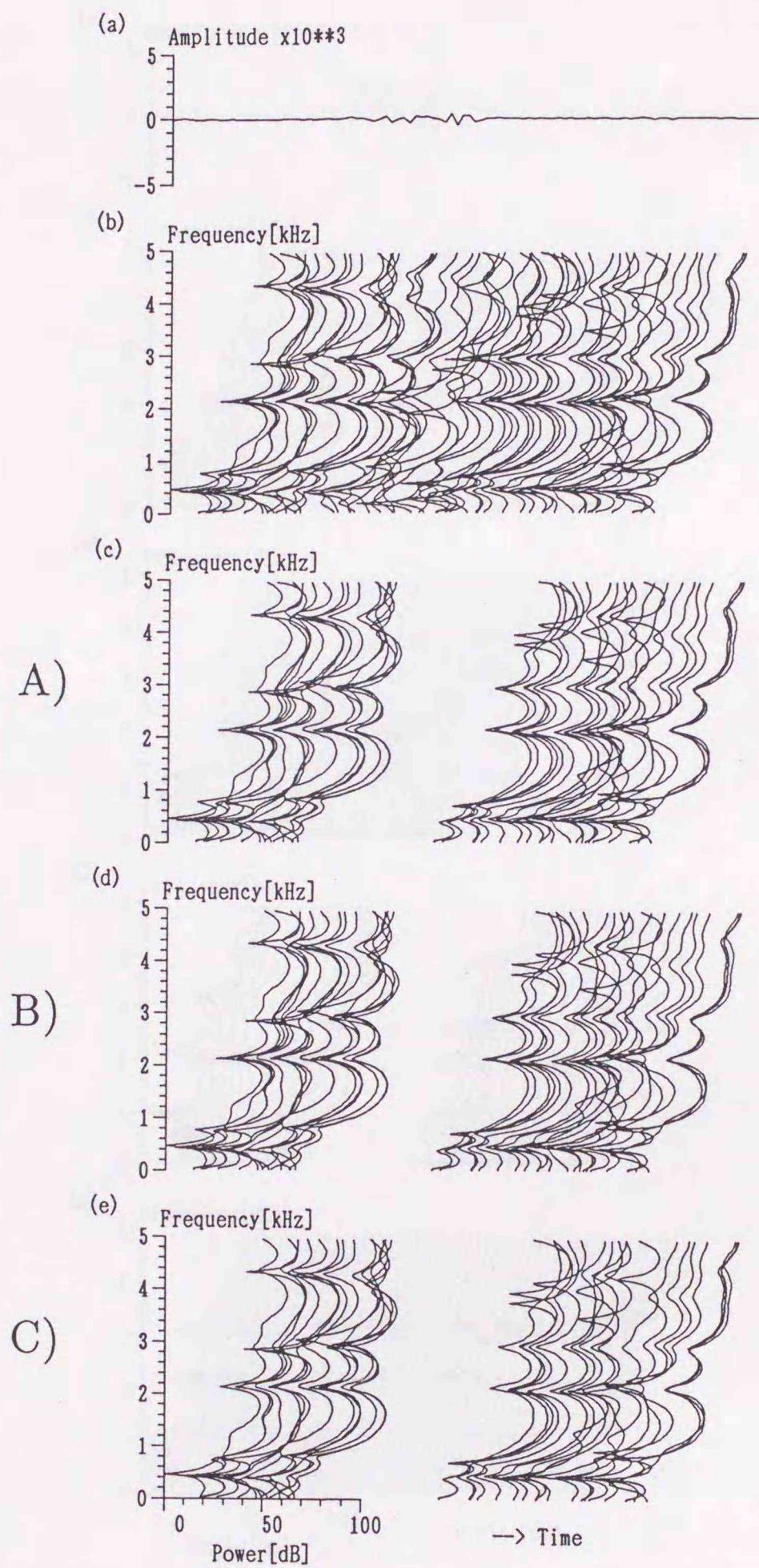


図 5.8: ニューラルネットのノードによる選択システムの選択結果
(話者 3/o/)

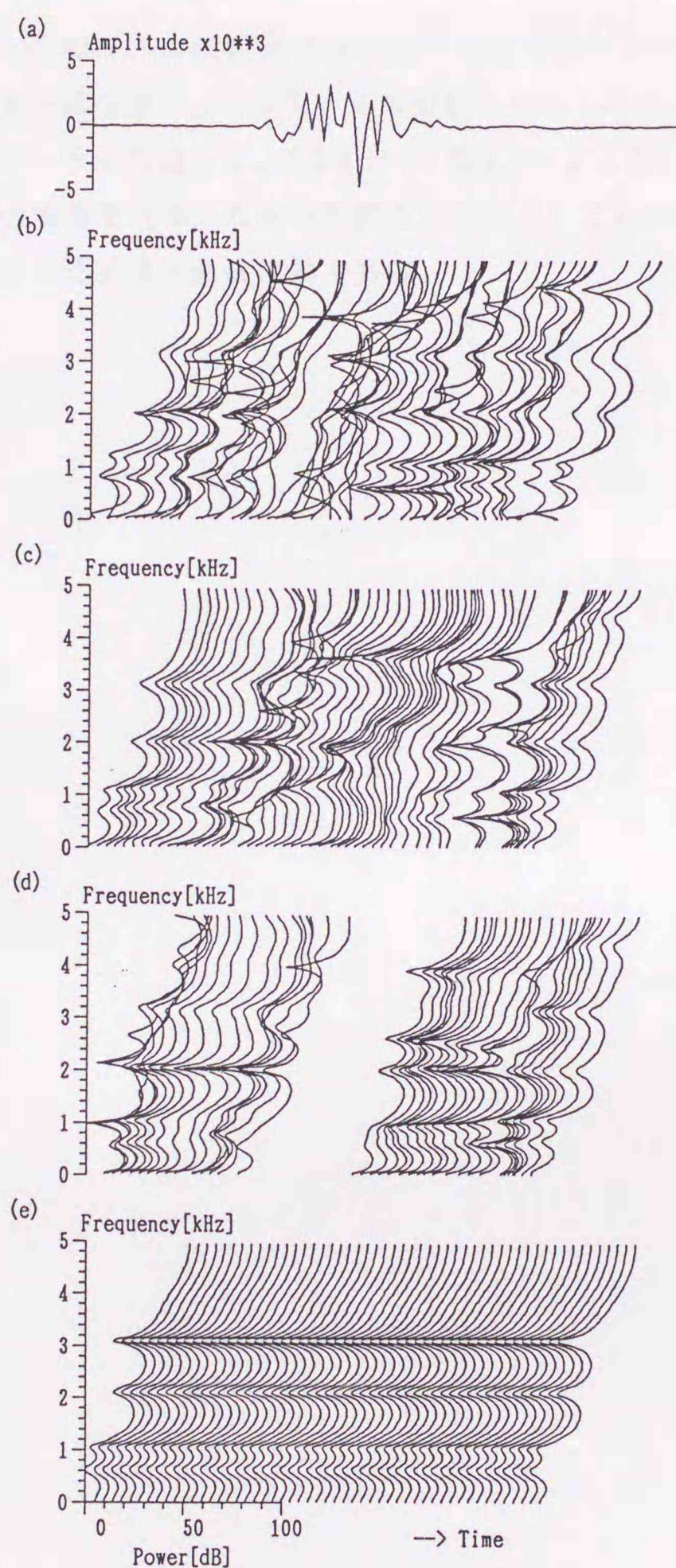


図 5.9: 重み付き RLS 法 (AR:16 次、MA:0 次) スペクトルの選択実験

(実音声 /a/)

(a) $\nu(k)$ (b) 平滑化前のスペクトル (c) 選択前に平滑化
 (d) 選択後に平滑化 (e) LPC

重み付き RLS 法の次数を AR:15 次、MA:4 次として母音以外の有声音にも対応できるようにし、その分析スペクトルから 5.2.3 と同様に学習データを作成して、5.2.4 の C) 型 (ノード 1 個) の選択システムで選択実験を行なったものが図 5.10 である。これが本章で述べたスペクトル選択法の最終結果である。

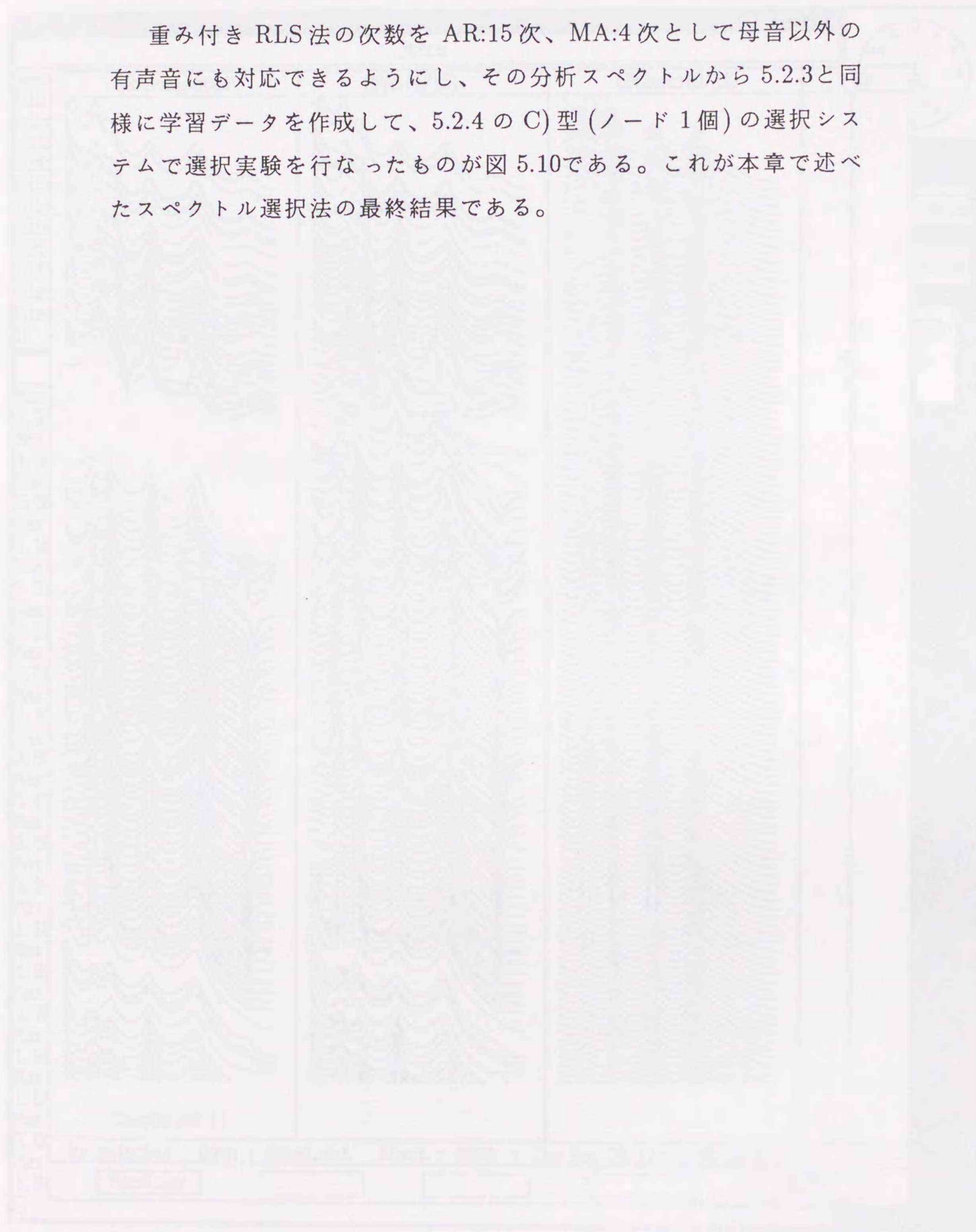


図 5.10. 重み付き RLS 法によるスペクトル選択の結果

図 5.10. 重み付き RLS 法によるスペクトル選択の結果

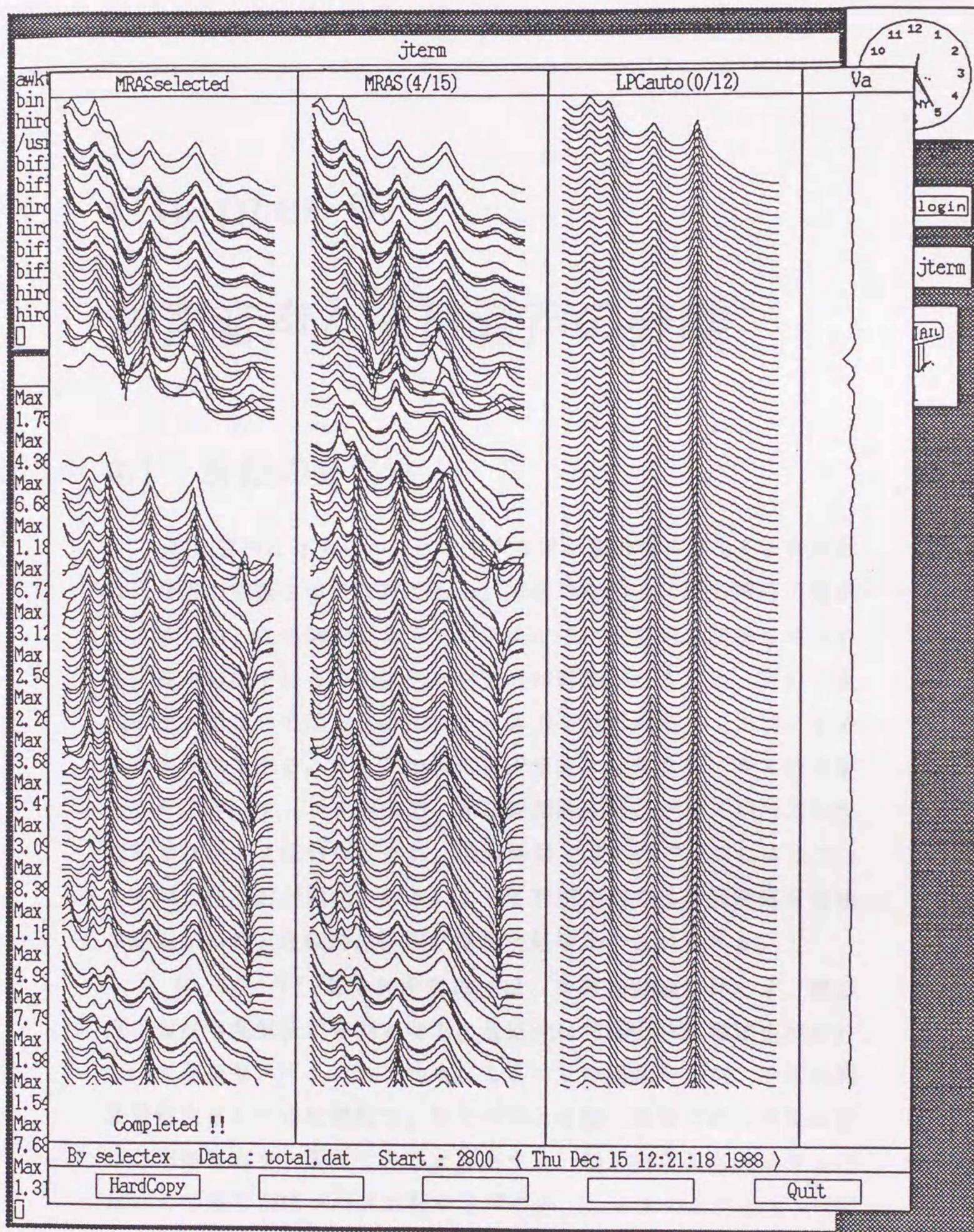


図 5.10: 重み付き RLS 法スペクトルの選択実験 (/a/)

(AR:15 次、MA:4 次)

Chapter 6

無声音声と有声音声の分離

6.1 分離の必要性

人間の音声には大きく分けて有声音声と無声音声がある。有声音声は声帯の振動を伴う音声である。声帯の振動とは即ち声帯の開閉であるから、この開閉によって呼気流はほぼ周期的に断続させられる。有声音声ではこの断続的な呼気流が音源となるため、音声の生成モデルにおいて図 6.1(a) のように入力を疑似周期パルス列とするのは合理的である。またこの様に音声生成フィルタをパルス列で励振した場合、フィルタの出力、即ち観測波形は、パルスが入力した時刻ではパルス成分が含まれ、他の時刻ではそのパルスによるフィルタの応答波形が観測される。よって有声音声では各観測信号の値に意味があり、近似的には確定信号と見なすことができる。

一方 $/s/$ 、 $/tʃ/$ 等の無声音声では、声帯の振動は伴わず、調音点における乱気流が音源となる。乱気流は可聴周波数帯域全域にわたった連続なスペクトルを持つ。ここでこの雑音音源を、平坦な周波数特性をもつ白色過程で、かつ平均と分散の情報だけで性質の記述が可能なガウス過程であると仮定する。無声音声の生成モデルでは、この様な白色ガウス過程を音声生成フィルタの入力とする (図 6.1(b))。入力がガウス過程であるから出力信号もガウス過程となり、観測波形は 2 次のモーメントに特徴を持つ確率信号となる。

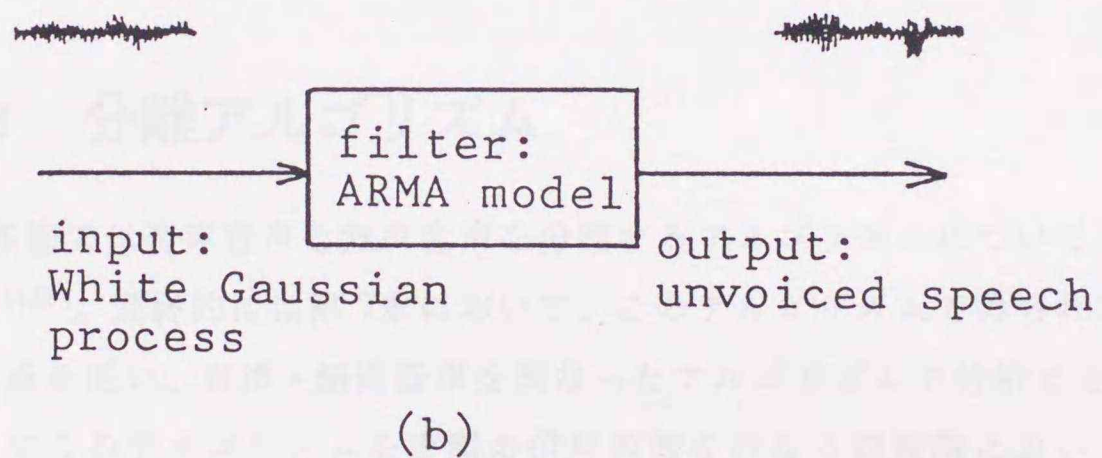
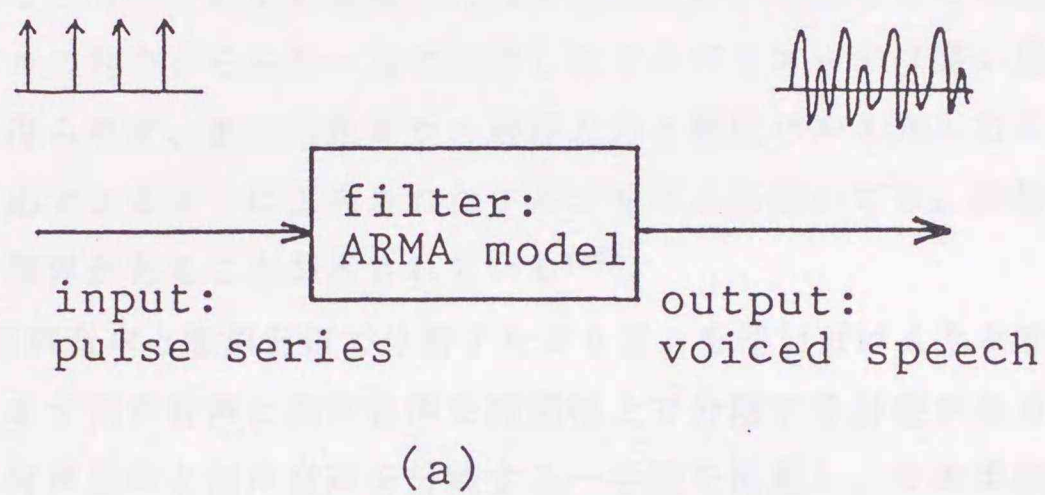


図 6.1: 音声の生成モデル

(a) 有声、(b) 無声

以上に述べたことから、有声音声と無声音声は、音声生成フィルタを励振する音源の性質が本質的に異なり、そのため観測される信号の性質も異なっている。従って音声を高精度に分析する場合、有声音声と無声音声では各々に適したアルゴリズムを使い分ける必要があると考えられる。実際、入力を白色ガウス過程もしくは疑似周期パルス列のどちらか一方に限定したアルゴリズムでは高い分析精度は得られず、また白色ガウス過程入力と疑似パルス列入力の両方に対応できるように工夫されたアルゴリズムを用いても、分析精度には限界があることが示されている⁽¹⁶⁾。

有声音声と無声音声で分析アルゴリズムを使い分けようとする場合、まず有声音声と無声音声を時間軸上で分離する必要がある。以下に有声音声と無声音声を分離する一手法を提案し、その手法による分離実験の結果を述べる。

6.2 分離アルゴリズム

本節では有声音声と無声音声を分離するアルゴリズムについて述べる⁽¹¹⁰⁾。最終的には第7章において、このアルゴリズムで得られた分離点を用い、有聲・無聲音声を異なったアルゴリズムで分析する。従ってこのアルゴリズムを実際の信号処理を行なう前段階において情報を得るという意味でパイロットアルゴリズム (PA) と名付ける。

6.2.1 分離に用いるパラメータ

有声音声と無声音声を分離するためには、有声音声と無声音声の違いを明確に表わすパラメータを見つけなければならない。本手法では、有声音声と無声音声のスペクトル包絡の大まかな傾きの違いに注目し、音声スペクトルの傾きを表わす AR 次数 1 次の時変 AR モデルを用いる。

このモデルの AR 係数は、第2章で述べた MIS 法⁽¹⁸⁾⁽¹⁷⁾⁽¹⁶⁾を書き換えることにより容易に求められる。ここでは生成フィルタの入

力を白色ガウス過程とし、1次 AR モデルを用いた場合のアルゴリズムを示す。

• 漸化式

$$\nu(k) = y(k) - y(k-1) \cdot \hat{a}_1(k-1) \quad (6.1)$$

$$f(k-1) = f(k-1|k-2) - \frac{y^2(k-2)f^2(k-1|k-2)}{1 + y^2(k-2)f(k-1|k-2)} \quad (6.2)$$

$$\text{if } \nu^2(k) \geq C \cdot \hat{\sigma}^2(k-1) \cdot \{1 + y^2(k-1) \cdot f(k-1)\} \\ \text{and } k \geq t_i + t_0 \text{ then} \quad (6.3)$$

$$\hat{q}^2(k-1) = \frac{\nu^2(k) - \hat{\sigma}^2(k-1) \cdot \{1 + y^2(k-1) \cdot f(k-1)\}}{y^2(k-1)} \quad (6.4)$$

$$\nu(k) = 0, t_i = k \quad (6.5)$$

else

$$\hat{q}^2(k-1) = 0 \quad (6.6)$$

endif

$$f(k|k-1) = f(k-1) + \frac{\hat{q}^2(k-1)}{\hat{\sigma}^2(k-1)} \quad (6.7)$$

$$\hat{a}_1 = \hat{a}_1(k-1) + \frac{y(k-1)f(k|k-1)}{1 + y^2(k-1) \cdot f(k|k-1)} \nu(k) \quad (6.8)$$

$$\alpha(k) = \min(k, \alpha) \quad (6.9)$$

$$\hat{\sigma}^2(k) = \hat{\sigma}^2(k-1) + \frac{\frac{\nu^2(k)}{1 + y^2(k-1) \cdot f(k|k-1)} - \hat{\sigma}^2(k-1)}{\alpha(k)} \quad (6.10)$$

- 初期値

$$\hat{a}_1(0) = 0, f(0| - 1) = 100, \hat{q}^2(0) = 0, \hat{\sigma}^2(0) = 0.1, t_i = 0$$

- 定数

$$\alpha = 50, t_0 = 20, s = 50, C = 16$$

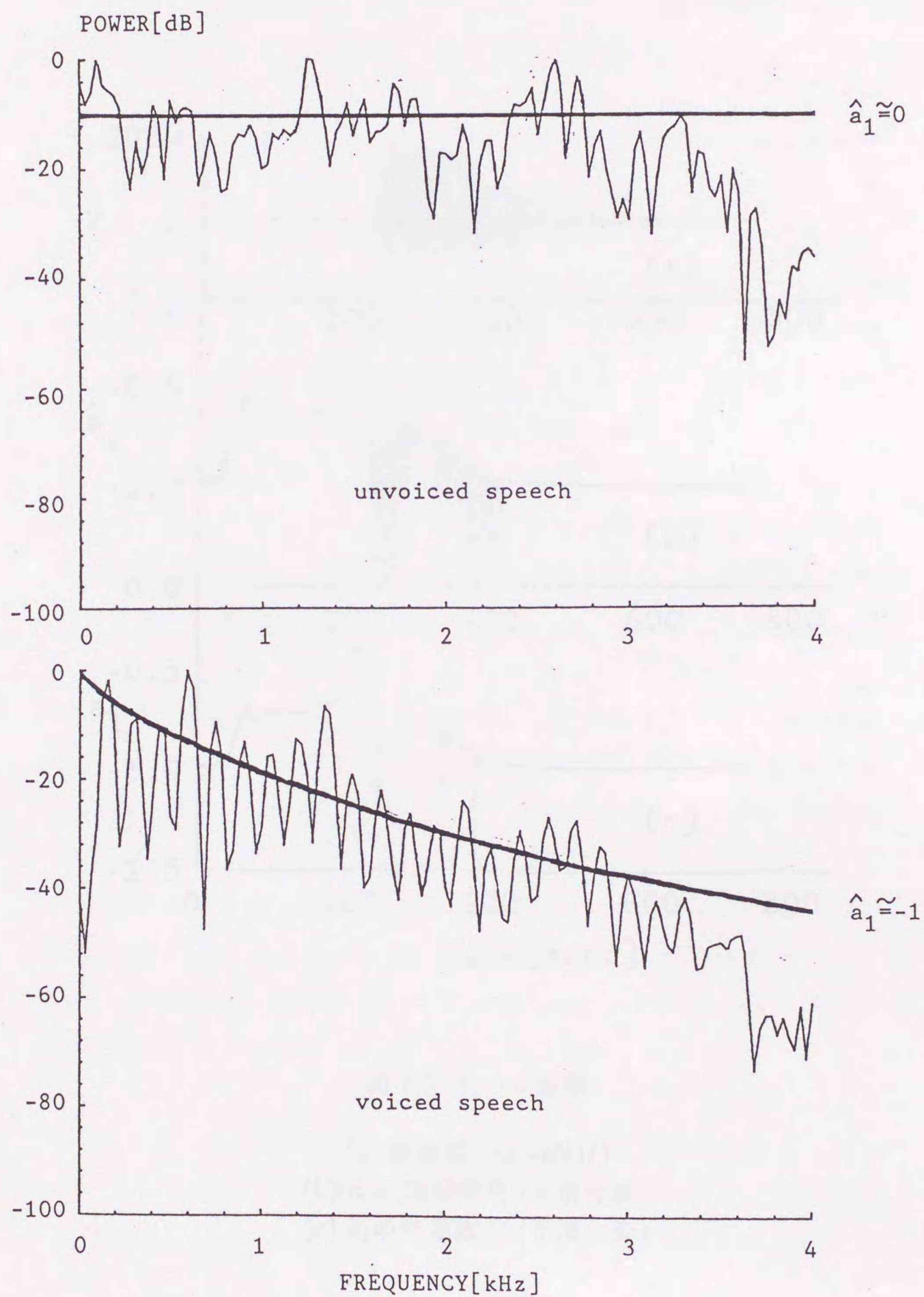
上式における $\hat{a}_1$ は次の評価関数を最小にするように求められる
(16)。

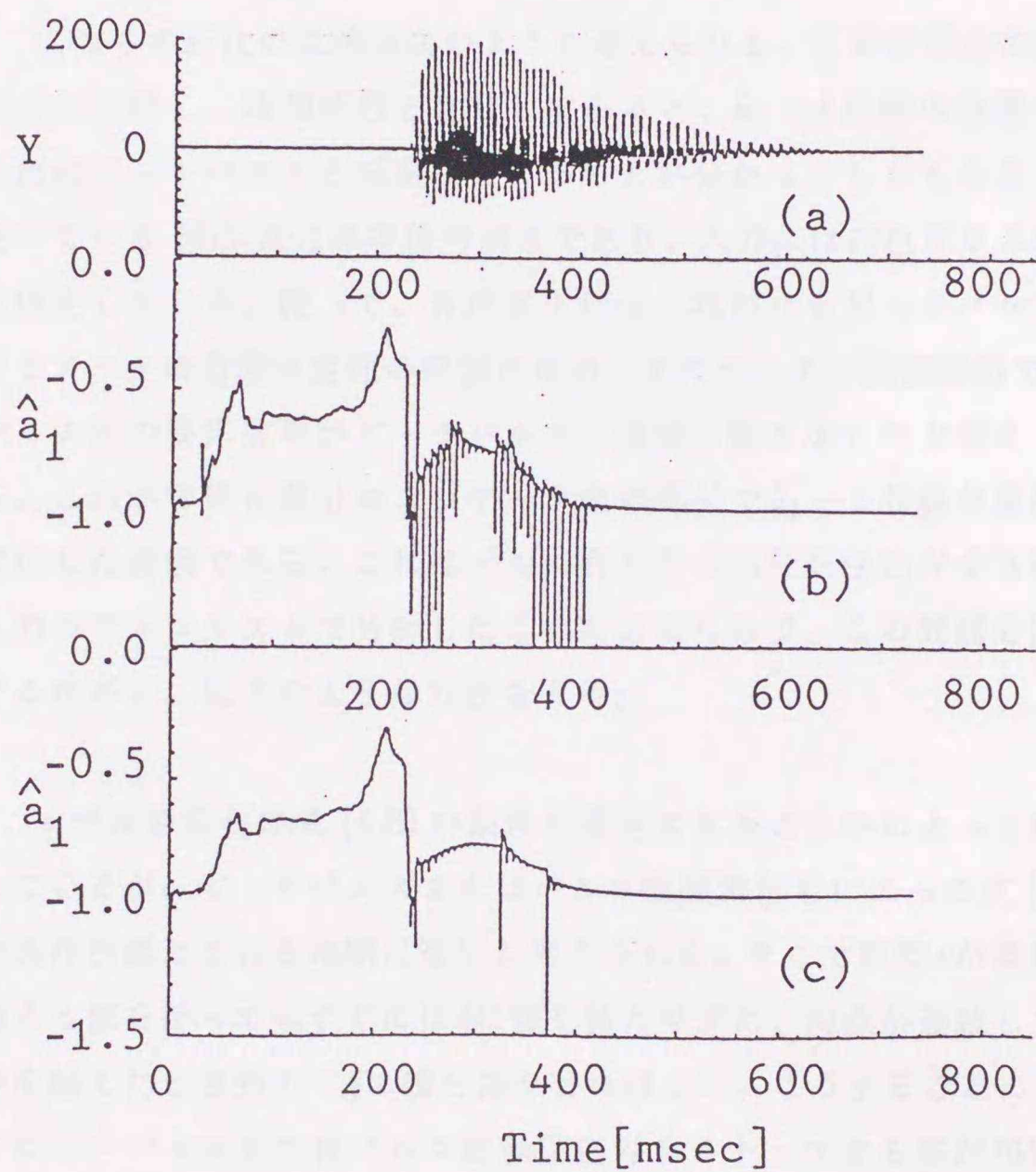
$$V = \frac{\sigma_u^2}{2\pi j} \oint \left| H(z^{-1}) \cdot (1 + \hat{a}_1 z^{-1}) \right|^2 \frac{dz}{z} \quad (6.11)$$

ここで $H(z^{-1})$ は音声生成モデルであり、 σ_u^2 はモデルの入力分散である。上式より音声信号の場合、音声の声帯および放射モデルを考慮すれば、 $\hat{a}_1$ は 0 から -1 までの値をとり、無声音声では 0 に、有声音声では -1 に近いことが予想される。物理的には $\hat{a}_1$ は分析スペクトルの傾きの概略を表わすことになる (図 6.2)。このアルゴリズムは適応型であるので、サンプルデータ 1 個につき 1 個の $\hat{a}_1$ が求められる。

6.2.2 パラメータ-時間曲線の平滑化

前項のアルゴリズムにテストデータをかけた例が図 6.3(b) である。縦軸が $\hat{a}_1$ 、横軸が時間 t である。使用したデータは成人男子の実音声 /3(saN)/ で、8kHz サンプルング、12bit A-D である。図 6.3(a) に示される原波形と比較すると、分離すべき無声音声 /s/ と有声音声 /a/ の境目では、パラメータ $\hat{a}_1$ の値が急激に増加・減少しており、明らかに他の点とは異なった特徴を示している。この特徴ある急激な増加・減少点を捕らえることにより無声音声と有声音声の分離が可能となる。しかし図 6.3(b) の $\hat{a}_1 - t$ 曲線はデータの先頭部分及び有声音部分で急に $\hat{a}_1$ が変化し元の値に戻る点がある。音声の生成モ

図 6.2: $\hat{a}_1$ パラメータの物理的意味

図 6.3: $\hat{a}_1 - t$ 曲線(a) 原波形 ($/3(\text{saN})/$)(b) $\hat{a}_1$ の時間変化 (平滑化前)(c) $\hat{a}_1$ の時間変化 (平滑化後)

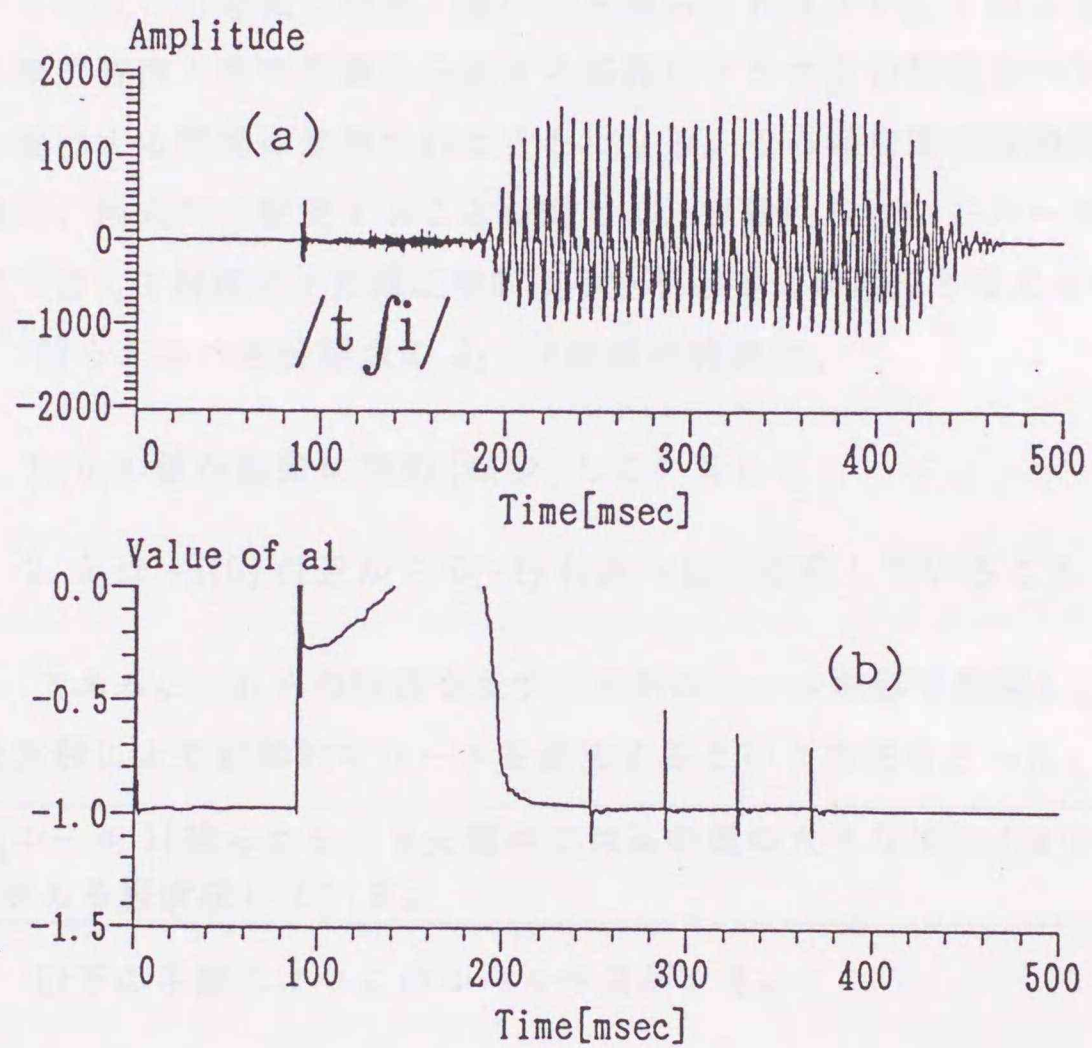
デルがこのように変化するとは考えにくいので、分離点の特徴を失わない程度に $\hat{a}_1 - t$ 曲線を平滑化する必要がある。

これらの変化の原因は次のように考えられる。有声音部分の変化部分に注目し、時間波形と比較してみると、 $\hat{a}_1 - t$ 曲線の急激な極小点はピッチパルスと同期していることが分かる。もともとここで用いている MIS 法は確率信号向きであり、入力には白色ガウス過程を仮定している。従って、有声音部分に現われるピッチパルスはパラメータの急激な変化の原因になる。またデータの先頭部分では、パルス性の異常信号がピッチパルスと同様な働きをしたと考えられる。これが有声音部分およびデータ先頭部分で $\hat{a}_1 - t$ 曲線が急激に変化した原因である。これはパルス列入力の部分を白色ガウス過程入力のアルゴリズムで分析したことによるため、この問題を回避するために、以下のような方法をとる。

モデルの変化は式 (6.3) の条件を満足するかどうかによって検出しているが、ピッチパルスまたはパルス性異常信号によって式 (6.3) の条件が満たされる時間は短いと考えられる。そこで誤差 ν が条件を越える値を持ってもすぐには $\hat{q}$ に値を持たせずに、何点か連続して条件を越えたとき初めて $\hat{q}$ に値を持たせればよい。こうすることによってピッチパルスまたはパルス性異常信号をスキップする事が可能である。具体的なスキップ点数は 5、6 個のテストデータによる実験から経験的に求めた結果、5 点が適当となった。

図 6.3(c) は以上のようにして求めた $\hat{a}_1 - t$ 曲線の例である。ピッチパルス及びパルス性異常信号をスキップすることによって分離点の特徴を失わずに比較的安定な $\hat{a}_1 - t$ 曲線を得ることができた。

同様のアルゴリズムで別のデータ (成人男子の実音声 /tʃi/, 12kHz サンプリング、12bitA-D) を分析した例を図 6.4 に示す。

図 6.4: $\hat{a}_1 - t$ 曲線(a) 原波形 (b) $\hat{a}_1$ の時間変化 (平滑化後)

6.2.3 分離点の決定

前項で求められた $\hat{a}_1 - t$ 曲線はほぼ予想通り無声子音部分では 0 に近く、有声音部分では -1 に近い値をとり、無声子音から有声音への変化点で急激に増加・減少している。無声音声と有声音声を分離するにはこの急激な増加・減少点を捕らえればよい。このことは即ち無声音声と有声音声を分離する問題がグラフ上の特定のパターンを抽出する問題に変換されたことになる。このような問題は純粋に理論的、数式的に解決することは難しく、経験的な法則をルール形式で蓄えて解決する知識工学的な手法をとるのが適当と考えられる。

捕らえるべき分離点の $\hat{a}_1 - t$ 曲線の特徴は、

1. $\hat{a}_1$ の値が急激に増加 (減少) していること
2. $\hat{a}_1$ は -1(0) 付近から 0(-1) 付近へ値が変化していること

である。これらの特徴をまず大まかなルールの形で表現し、その後実験により経験的なルールを追加するという方法をとった。

[ルール 1] 捕らえるべき分離点では $\hat{a}_1$ の値の大きな増加 (減少) がある程度続いている。

以下の手順によりこのルールを適用する。

1. $\hat{a}_1$ の差分 $\Delta\hat{a}_1$ をとる。
2. $\Delta\hat{a}_1$ を以下のように 3 段階に分類して記号化する。

a:	0.005	<	$\Delta\hat{a}_1$	
c:	-0.005	<	$\Delta\hat{a}_1$	< 0.005
e:			$\Delta\hat{a}_1$	< -0.005
3. 「a」(「e」) が 2 個連続した点を候補点とする。

正しい分離点を落とさないためには連続点数は経験的に 2 点となった。これは $\hat{a}_1 - t$ 曲線を目視した限りでは大きな増加・減少が (2 点よりも) 長く続いているように見えても、実際にはその間に「a」

(「e」)の条件を満たさない点が入っている場合があり、そのため「a」(「e」)は目視で感じられるほど長くは連続していないことによると考えられる。「a」(「e」)の連続点数を2点とすると関係のない候補点を多数挙げることになるが、これは後に他のルールで絞っていくことが可能であると考えられる。

[ルール 2] 候補点が入りも密集している場合には1つにまとめる。

目視では連続的に大きく増加・減少しているように見えるが実際には「a」(「e」)が連続していないような部分では、ルール1による候補点が入りも密集している。このような点を1つにまとめることによって、より目視に近い「増加」・「減少」の捕らえ方ができる。また「a」(「e」)が3個以上連続しているような場合ルール1では候補点が連続して複数出力されるが、実際の「増加」・「減少」は1個であるから、このような候補点もまとめる必要がある。そこで次のようなアルゴリズムで密集した候補点を統合する(図6.5)。

- i) 全ての候補点について、候補点に候補点マーク M、候補点の1点前に開始マーク S、1点後に終了マーク E を付ける。何もない点は不存在マーク F をつける。
- ii) データを時間軸上で頭から走査して、候補点に出会ったらその候補点 M1 および M1 に対応する開始点 S1 の位置を記憶する。
- iii) M1 から後方 120 点 (15msec) 以内に別の候補点が見つかった場合は、見つかった候補点のうち最後の候補点 M2 の終了点 E2 を記憶する。見つからなかった場合には M1 の次の点からさらに走査し初めて出会った候補点を新しい M1、対応する開始点を新しい S1 として記憶し iii) に戻る。
- iv) 記憶されている開始点 S1 と終了点 E2 の間は、すべて注目している「増加」・「減少」が起きていると見なして増加・減少存

i) $\dots F \boxed{SME} F F \boxed{SME} F \dots F \boxed{SME} F \dots$

ii) $\dots F \boxed{\textcircled{S}ME} F F \boxed{SME} F \dots F \boxed{SME} F \dots$
 $S_1 M_1$

iii) $\begin{array}{c} \leftarrow \text{-----120 points-----} \rightarrow \\ \dots F \boxed{\textcircled{S}ME} F F \boxed{SME} F \dots F \boxed{\textcircled{S}ME} F \dots \\ S_1 M_1 \qquad \qquad \qquad M_2 E_2 \end{array}$

iv) $\dots F \boxed{\textcircled{S}TTTTT\textcircled{M}T} \dots TTT\textcircled{E}F \dots$
 $S_1 \qquad \qquad \qquad M_3 \qquad \qquad \qquad E_2$

v) $\begin{array}{c} \leftarrow \text{---120 points-->} \\ \dots F \boxed{\textcircled{S}TTTTT\textcircled{M}T} \dots TTTEF \dots \\ S_1 \qquad \qquad \qquad M_1 \end{array}$
to iii)

図 6.5: 候補点統合アルゴリズム

在マーク T で埋め、S1 と E2 の中央の点を新しい候補点 M3 とする。

v) M3 を新しい M1、対応する開始点 S1 はそのまま S1 として記憶し、iii) に戻る。

vi) 走査がデータの最後まで到達したら終了する。

[ルール 3] 捕らえるべき分離点では、増加 (減少) 開始点と増加 (減少) 終了点の $\hat{a}_1$ の値の差が大きい。

増加・減少開始点及び終了点の $\hat{a}_1$ の値は、開始点 S の前、及び終了点 E の後ろそれぞれ 20 点の $\hat{a}_1$ の値を平均したものとした。このことによって開始、終了点の「付近」の値を表現した。この開始点と終了点の値の差があるしきい値以下の候補点は、ルール 3 を満たさないとして候補から外した。具体的なしきい値は実験データから経験的に、0.2(0.1) を採用した。比較的小さな値になったのは、正しい候補点を落とさないようにしたためである。

[ルール 4] 減少の場合、捕らえるべき分離点は、減少終了点の $\hat{a}_1$ の値が -1 の付近である。

無声音から有声音への変化点では、 $\hat{a}_1$ の値は理論的には 0 付近から -1 付近へ変化するはずであるが、実際のデータでは減少が 2 段階程度に分かれている例があるため、減少開始点の $\hat{a}_1$ の値は 0 付近にあるとは限らない。このためルール 4 は上記のようになった。終了点の値 (20 点の平均値) が、-0.8 以下であることを以て「-1 付近」であるとした。具体的な数字は他のルールと同様に実験データから経験的に求めた。なおこのルールは増加点には適用しない。

[ルール 5] $\hat{a}_1$ の値が 0(-1) にきわめて近い値を安定してとれば、明らかに無声音 (有声音) である。

「変化」のみに注目しているルール 1~4 の明らかな誤りを除くルールであり、 $\hat{a}_1$ の 3msec の平均が -0.1 以上 (-0.9 以下) ならば無声音 (有声音) とした。

6.3 分離実験結果

前節で述べたアルゴリズム及びルールで、ルール抽出に用いたものと同様のデータを分析し、分離実験結果を得た。音声データは比較的分離が難しいとされる破裂性の無声子音を含む単音節音声であり、3名の成人男性話者による $\{/p/, /t/, /k/\} + \{/a/, /i/, /u/, /e/, /o/\}$, $/t/ + \{/a/, /e/, /o/\}$, $/tʃi/$, $/tsu/$ (12kHz サンプリング、12bitA-D) の計 45 個である。表 6.1 に結果を示し、図 6.6~6.8 に結果の例を示す。図 6.6~6.8 において横軸はすべて時間 t で、(a) は原波形、(b) は $\hat{a}_1$ 、(c) はシステムが出力した分離実験結果である。45 サンプル中 19 サンプル (42%) ではば妥当な点を捕らえた (図 6.6)。25 サンプル (56%) については、破裂子音を対象としているため子音部がガウス性というよりはインパルス応答的な信号であり、そのため $\hat{a}_1$ の値にあまり変化がみられず、分離点は出力されなかった (図 6.7)。しかしこのような子音は、MIS 法ではなく重み付き RLS 法で分析されるべきであり、この意味でシステムは正しい判断をしたと考える。残りの 1 サンプル (2%) については、増加点に対応する減少点が捕らえられず、ルール 5 によっても正しく修正されなかったため、誤った分離結果を得た (図 6.8)。

以上の実験より、ルール抽出と同じデータではあるが、45 サンプル中 44 サンプル (98%) でシステムは妥当な分離を行なうことが確かめられ、前節で述べたアルゴリズム及びルールを用いることによって、無声音声と有声音声を良好に分離できることが示された。本実験において無声音声と有声音声の分離問題における、経験則を用いた知識工学的手法の有効性を示した。

話者 T.O.		話者 H.K.		話者 S.M.	
音節	結果	音節	結果	音節	結果
/ka/	C	/ka/	A	/ka/	A
/ke/	A	/ke/	A	/ke/	A
/ki/	A	/ki/	A	/ki/	A
/ko/	B	/ko/	B	/ko/	B
/ku/	B	/ku/	B	/ku/	B
/pa/	B	/pa/	B	/pa/	B
/pe/	B	/pe/	B	/pe/	B
/pi/	B	/pi/	A	/pi/	B
/po/	B	/po/	B	/po/	B
/pu/	A	/pu/	B	/pu/	B
/ta/	A	/ta/	B	/ta/	B
/te/	A	/te/	B	/te/	B
/tʃi/	A	/tʃi/	A	/tʃi/	A
/to/	A	/to/	B	/to/	B
/tsu/	A	/tsu/	A	/tsu/	A
A:8 B:6 C:1 (53%)(40%)(7%)		A:6 B:9 C:0 (40%)(60%)(0%)		A:5 B:10 C:0 (33%)(67%)(0%)	
A:19 B:25 C:1 (42%)(56%)(2%)					

表 6.1: 分離実験結果

(A:良好 B:無声音の特徴ほとんどなし C:失敗)

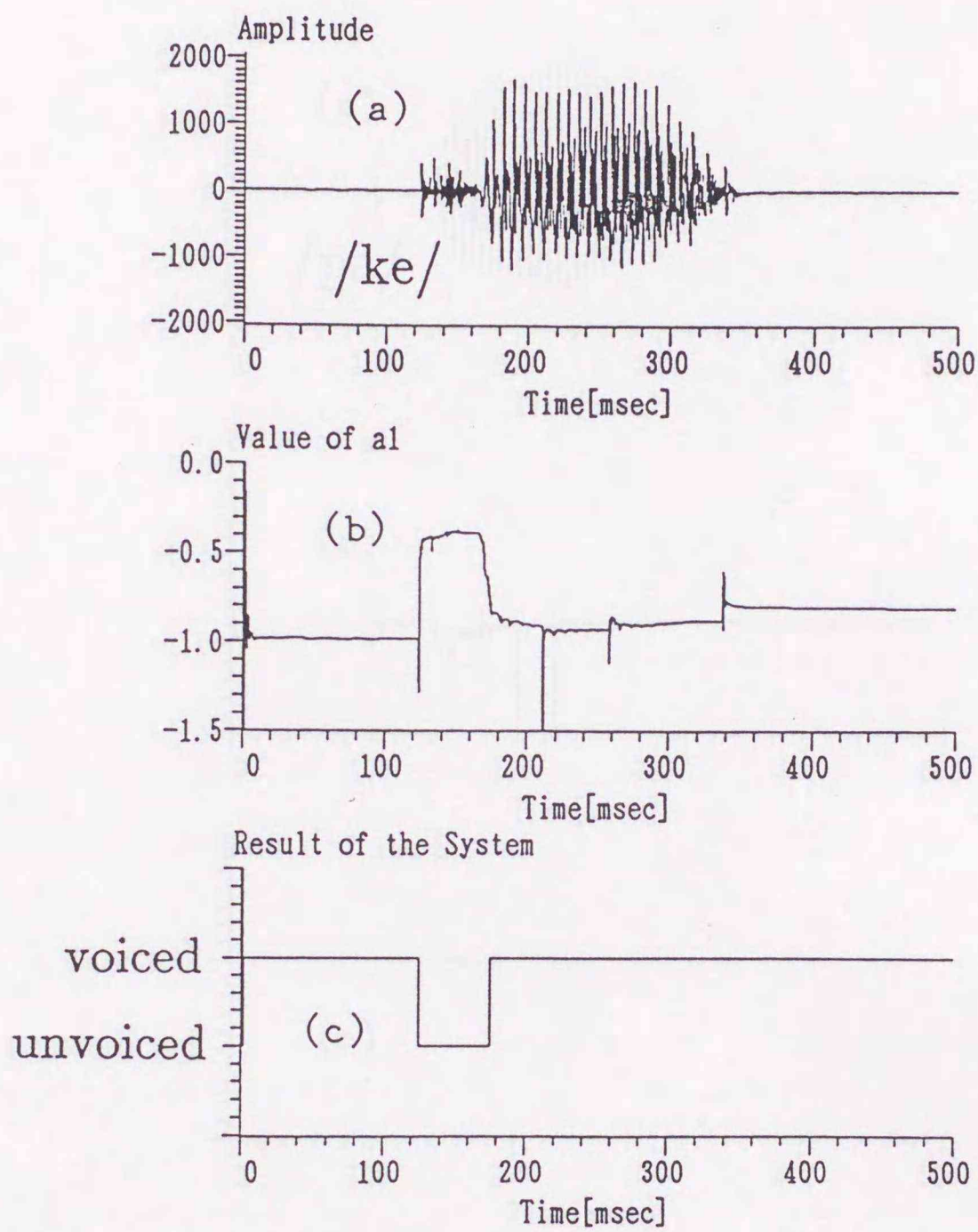


図 6.6: 分離実験結果

「A : 良好」の例

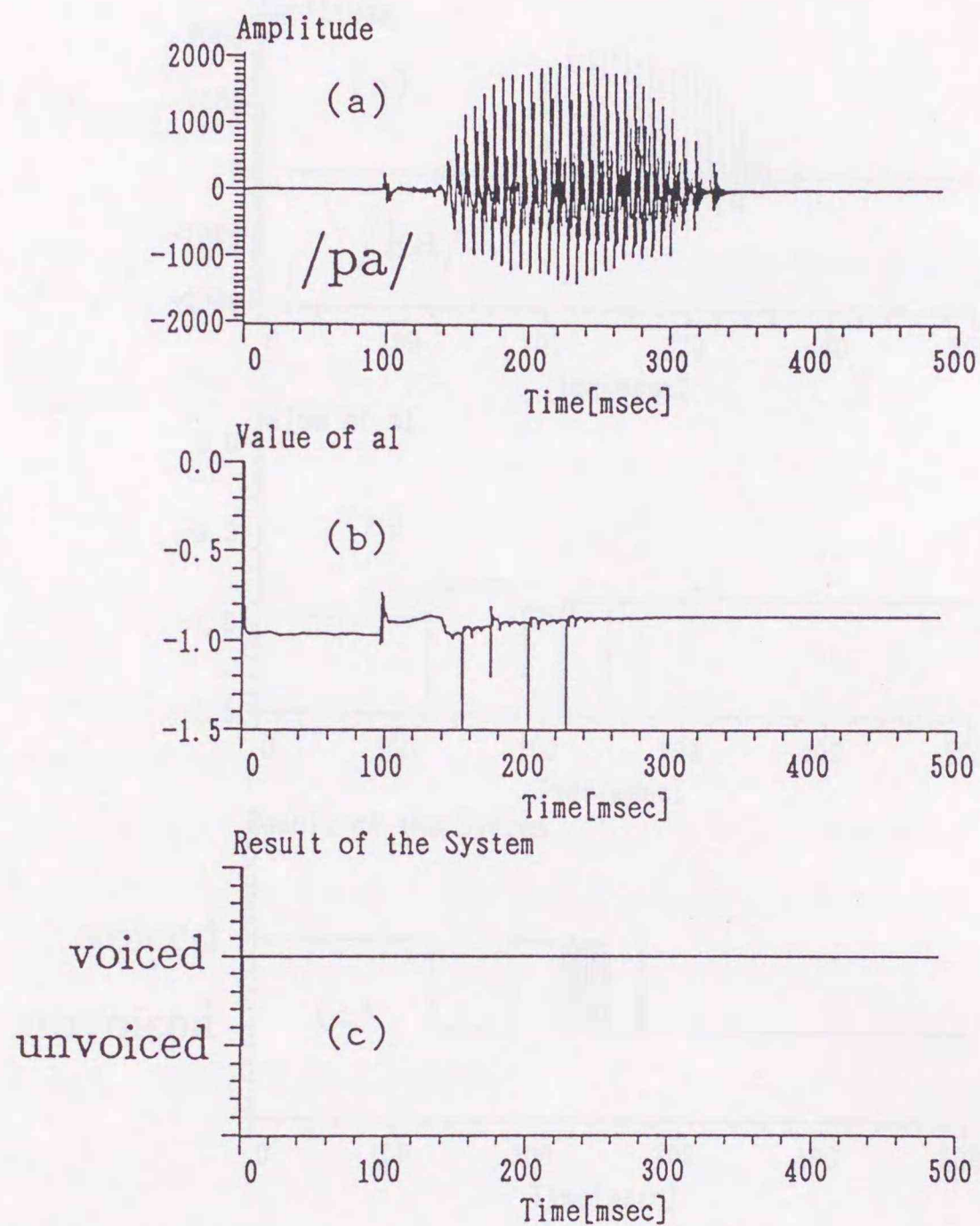


図 6.7: 分離実験結果

「B: 無声音の特徴がほとんど無い場合」の例

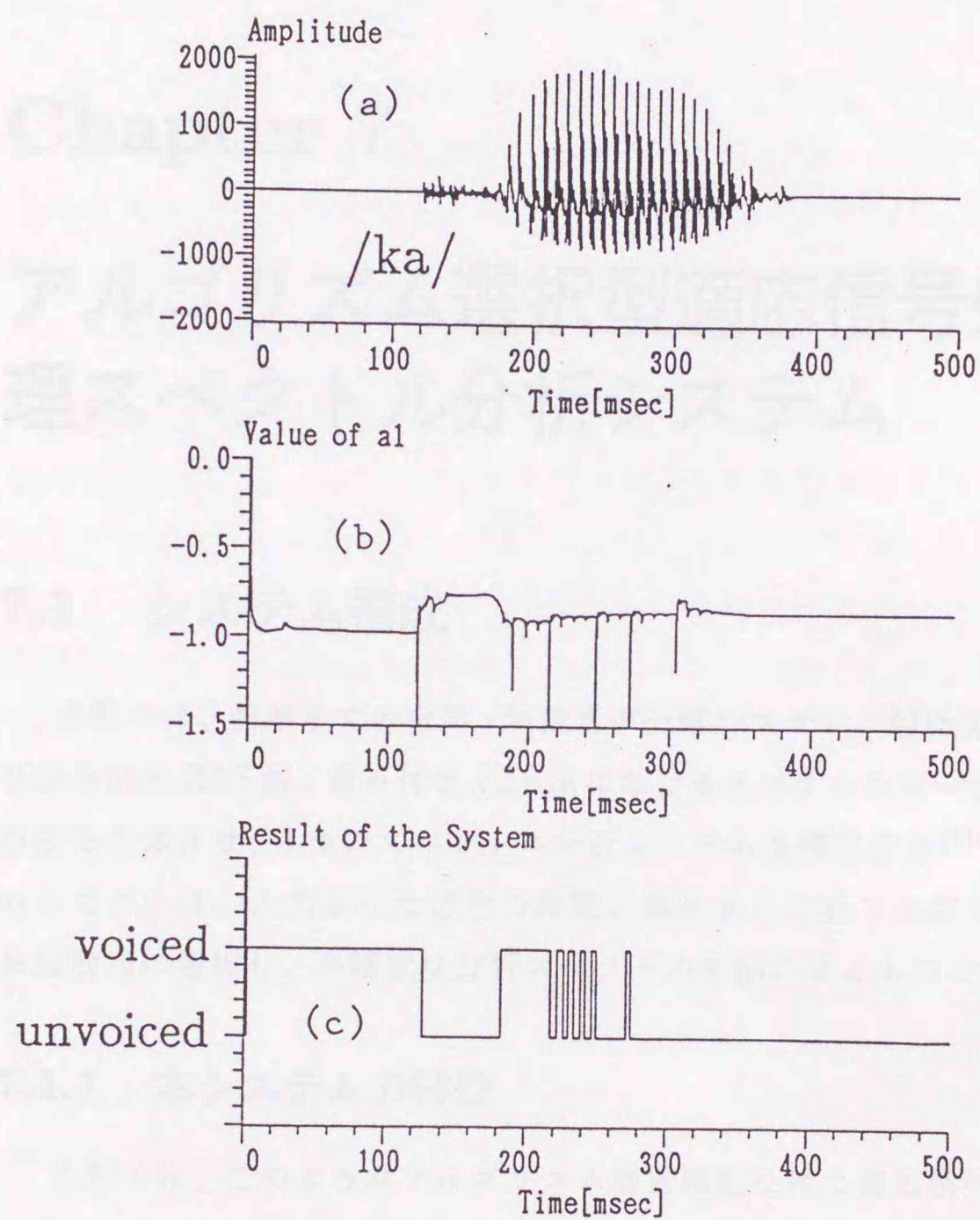


図 6.8: 分離実験結果

「C: 失敗」の例

Chapter 7

アルゴリズム選択型適応信号処理スペクトル分析システム

7.1 システム構成

本章では、前章までの有声・無声音声分離システム、MIS 法および重み付き RLS 法、重み付き RLS 法におけるスペクトルデータ選択手法を合体させ、1つのスペクトル分析システムを構成する⁽¹¹⁰⁾。このシステムは、入力された信号の性質に適合する分析アルゴリズムを自動的に選択し、高精度な分析スペクトルを出力するものである。

7.1.1 本システムの特徴

本節では、このようなアルゴリズム選択機能を持つ適応信号処理システムを構築するにあたり、まず本論文全体の背景となっている筆者らの DSP に対する基本的な考え方について述べる⁽¹¹⁴⁾。

従来の DSP の様子を図 7.1(a) に示す。DSP の対象となる信号には、信号生成モデルの入力が疑似周期インパルス列と見なせる確定信号、白色ガウス過程と考えられる確率信号、1個のインパルスによるインパルス応答等があり、また入力信号の無い無入力状態も考えられる。従来はこのような各種の信号の混在した入力に対して、なるべく多くの入力に対応できるような 1 個のアルゴリズムが各種開

発され、分析精度が競われてきた。しかし第 6 章でも述べたように、近年の研究によると 1 個のアルゴリズムでは限界があることが明らかにされた⁽¹⁶⁾。

そこで本論文では、それぞれの信号に適したアルゴリズムを入力に応じて切り替える方式を提案する。さらに、従来エキスパートである研究者が行なっていたアルゴリズム切り替えの部分に、経験則を用いたプロダクションシステム等を導入し、アルゴリズムの選択をも自動的に行なう適応信号処理システムを提案する。

その一般化した構成を図 7.1(b) に示す。入力信号は、まずその性質を識別するためのパラメータを抽出する比較的簡単なアルゴリズム (パイロットアルゴリズム PA) につけられる。PA からのデータは続くマネジメントシステム (最終的にはプロダクションシステム⁽⁴³⁾等によって実現) に入力され、エキスパートから抽出した経験的ルールベースに照らし合わせ、どの分析アルゴリズムが最適であるか判断する。それと平行して入力信号は各種の適応信号処理アルゴリズムによってスペクトル分析されており、出力されたスペクトルデータはマネジメントシステムからの指令によって時分割的に選択され、最終的な出力を得る。分析結果は人間のエキスパートが評価し、必要があればルールの追加、修正を行なう。PA 部および信号処理アルゴリズム部はポータブルな構造とし、必要があれば追加、変更が容易に可能とする。

本論文で実際にインプリメントしたシステムのブロック図を図 7.2 に示す。入力された音声波形は、まず MIS 法および重み付き RLS 法の両方でスペクトル分析される。重み付き RLS 法の出力スペクトルは第 4 章で述べた選択手法によって選択された後、スペクトル平滑化のため選ばれたスペクトルのみの間で個数 15 個で平均操作が施される。一方入力波形は第 6 章で述べた PA にも同時に入力され、その出力 $\hat{a}_1(k)$ がマネジメントシステムに入力される。マネジメントシステムは実際のシステムでは第 6 章で述べた有声・無声分離システ

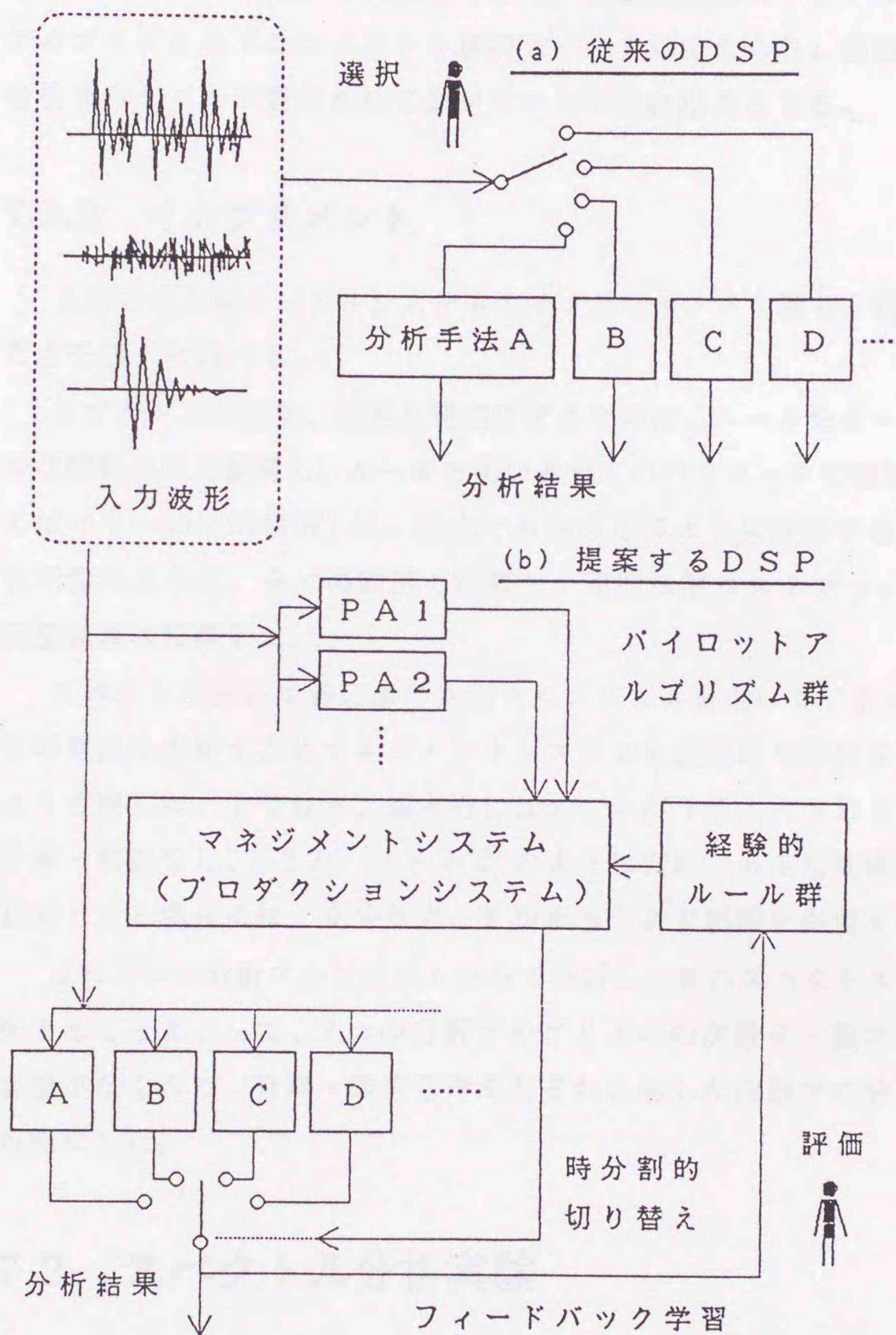


図 7.1: 従来の DSP と提案する DSP の概念図

ムによって実現され、蓄えられた if-then ルールを参照し $\hat{a}_1(k)$ の値の変化から有声・無声の判断を行なう。判断結果はスペクトル分析アルゴリズムの下でのスペクトル選択スイッチに伝えられ、高精度な分析スペクトルが選択されて本システムの最終出力となる。

7.1.2 インプリメント

上述のようなトータルシステムのインプリメントに際し、留意した点を以下に述べる。

まずルールの追加、変更を可能にするために、ルールをポータブルな関数の形で表現し、ルールで用いる全てのパラメータの値等(例えばパワーの平均値等)は、過去であればどのように参照することも可能なように、全ての時間で計算し、可能な限りスタティックな記憶領域に保存した。

スペクトル分析に適応信号処理アルゴリズムを用いているので、その利点を生かすためマネジメントシステムも適応的な動作をするよう考慮した。すなわち、基本的にはデータが1点入力される度に計算・判断をし、6.2.3の[ルール2]のように判断にある時間範囲を必要とする場合には、十分なデータがそろうまで判断を保留する。

また2つの分析アルゴリズム両方で分析した後にスペクトルを選択することによって、2つの分析アルゴリズムの次数を一致させる必要がなくなり、有声・無声音声それぞれに適した次数での分析を可能にした。

7.2 スペクトル分析実験

本システムに6.3で用いたものと同じ実音声データを入力し、分析結果を得た。結果の一例を図7.3に示す。

図7.3において、分析音声は/ta/であり、4枚のグラフは左からそれぞれ、MIS法(AR:15次,MA:6次)、比較のためのLPC(自己相

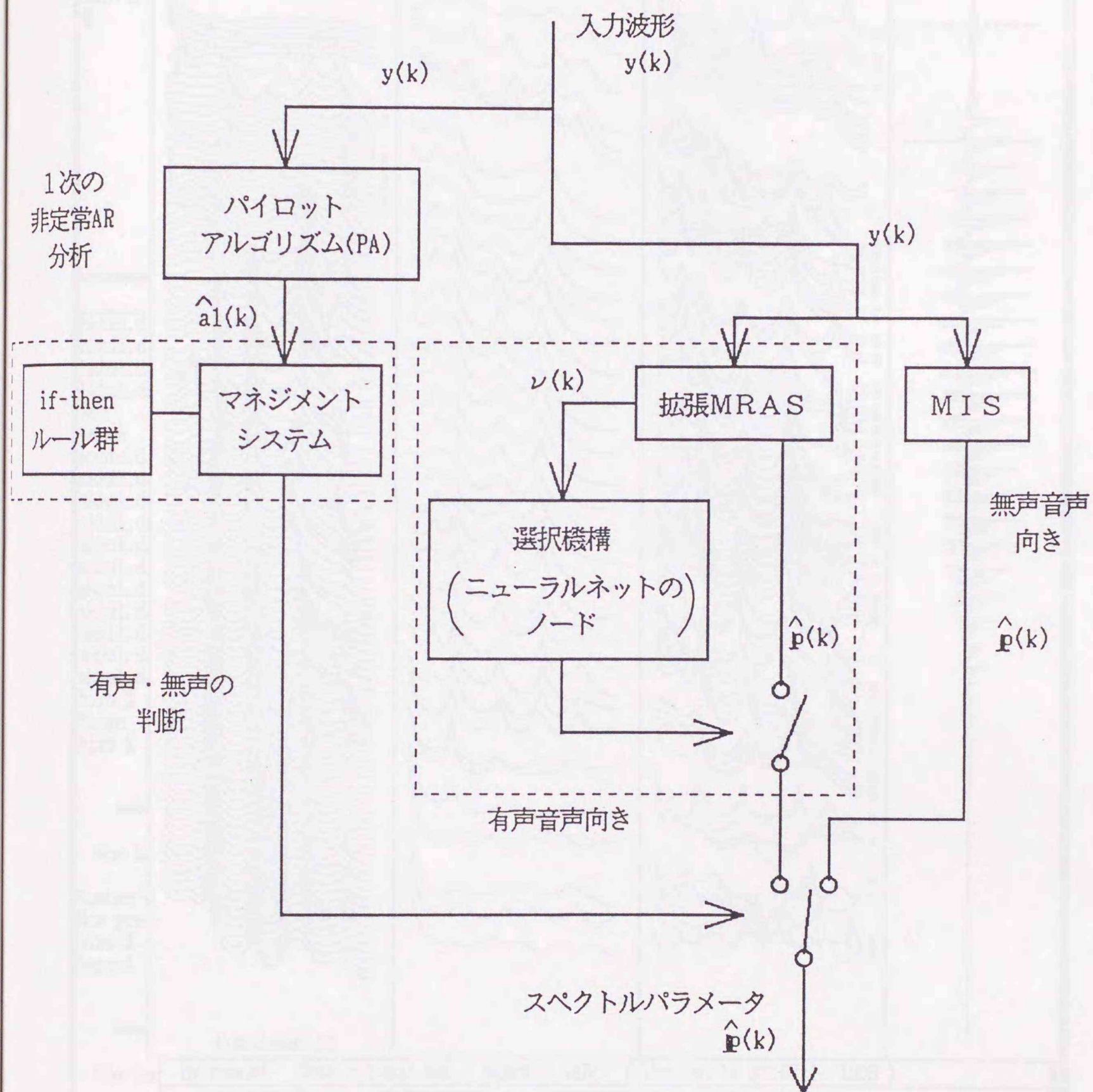


図 7.2: 本論文で実際に構築したシステムのブロック図

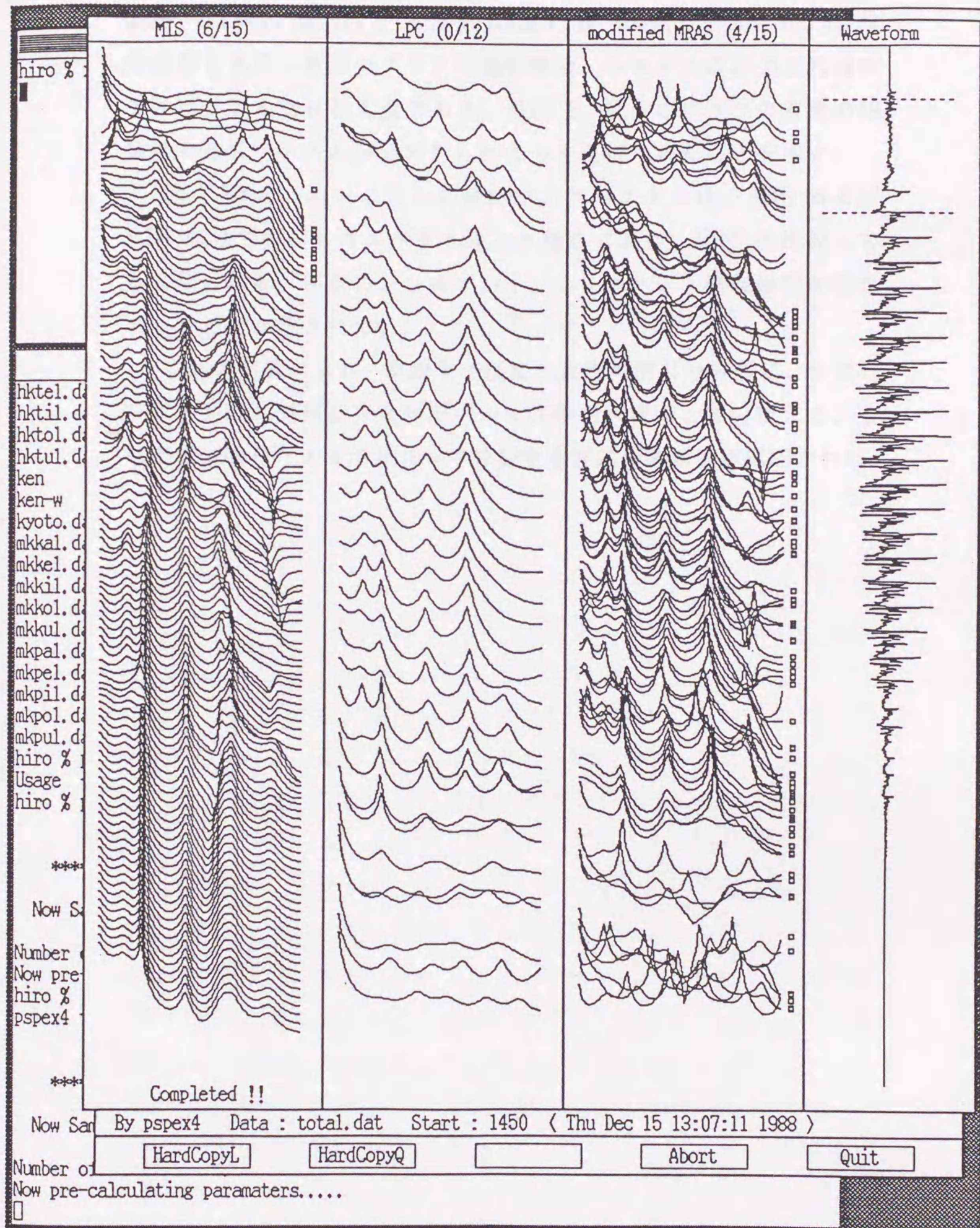


図 7.3: 本システムでのスペクトル分析実験 (/ta/)

関法、12次)、重み付き RLS 法 (AR:15 次, MA:4 次、 $\lambda = 0.9$)、信号波形である。波形のグラフの横破線は、システムが出力した分析アルゴリズム切り替え点であり、MIS と RLS のグラフの右側の四角はそのスペクトルがシステムによって選択されたことを示す。

図 7.3 より、本システムの最終出力スペクトルは、MIS および RLS 単独で分析したスペクトルより優れており、LPC と比較しても時間分解能、高次のフォルマント、フォルマント帯域幅等で優れていることが示された。

本章の実験により、本論文で提案した音声信号分析方式、すなわちガウス性の無声音声と周期パルス性の有声音声とを分離して、それぞれに適したアルゴリズムで分析する方式の有効性が確認された。

Chapter 8

連続音声のセグメンテーション 候補検出

8.1 まえがき

第7章までで構築されたような音声分析システムの重要な応用として、音声の認識がある。従来より音声認識の研究は盛んに行われている。当初単音素認識から始まった音声認識は、単音節認識、単語認識とレベルが向上し、近年では、認識の対象は日常われわれが話している連続音声とするのが主流となっている。連続音声の認識手段としていろいろな手法が既に提案されているが⁽⁴⁴⁾⁽³⁹⁾⁽⁴⁷⁾⁽⁴⁶⁾、それらは総じて、用意された何らかの標準パターンと、未知の入力音声との比較を行い、最も類似している標準パターンを決定することにより、選ばれた標準パターンに付加されているラベルを認識結果として出力するものである。標準パターンの単位としては、音素から単語まで様々な長さを持つものが提案されている。認識の対象が比較的短時間の単語で、しかも単語数も限られている場合(例えば、流通業務における行き先都道府県名など)は、対象単語そのものを標準パターンとして持つ DP マッチング⁽⁴⁴⁾等の手法が比較的有効であるが、対象単語を限定せず、しかも認識すべき発話の長さが通常 of 文章程度となると、これらの手法では記憶容量や計算量が莫大と

なり実用的でない。一方、音素を標準パターンとした認識手法では、上述のような発話長の増加による計算コストの増大は、DP 等と比較して少ない。そのため不特定単語を対象とした連続音声認識では、音素を単位とした認識が望まれる。しかし、音素を認識単位とした場合、標準パターンとマッチングをとるためには、入力未知音声之音素に分割(セグメンテーション)する必要がある、特に連続音声の場合、調音結合等が存在によりこのセグメンテーションは困難な課題となっている。

第 7 章までで構築した高精度適応 ARMA スペクトル分析システムでは、時間分解能の高い ARMA スペクトルが得られるため、従来の LPC により得られたスペクトルと比較して、上述のような調音結合等をより正確に表現できると考えられ、音素単位の連続音声認識におけるセグメンテーション等においても、有効であることが期待される。

そこで本章では、第 7 章で構築したスペクトル分析システムを用いた音声認識への第一歩として、上記システムの出力を利用して連続音声からセグメンテーション候補点を検出する簡単なシステムを作成し、実音声を用いて実験を行う⁽⁵²⁾⁽¹¹⁶⁾。

将来的には本セグメンテーション候補検出システムの後段に、ニューラルネット⁽³¹⁾等を用いた音素認識システムが接続され、セグメンテーション候補の統合・決定と音素認識が同時に行われる。さらにその認識結果は言語処理等を経て認識誤りが訂正され、最終的な認識結果となる。一般にこのような音素認識・言語処理においては、音素の付加は比較的容易に対応できるが、脱落は致命的な影響を及ぼす。このことをふまえ、本セグメンテーション候補検出システムは、多少冗長な候補点を出力しても正しい分離点は落とさないようなシステムを目標とする。

8.2 セグメンテーションパラメータ

本節ではセグメンテーション候補検出の手がかりとするパラメータについて検討する。セグメンテーションパラメータとして、以下の3種類を試みた。

1. ケプストラム距離
2. 6.2で述べた有声・無声分離アルゴリズムの出力点
3. 2.1.3で述べた MIS 法の式 (2.16) で与えられる $\hat{q}^2$ パラメータ

8.2.1 ケプストラム距離

音声のスペクトルがある時点で大きく変化した場合、その点で音素が変化した可能性が高いと考えられる。したがってスペクトルの変化を表現する量をセグメンテーションパラメータとするのは適切である。2つのスペクトル間の距離を表す量として、従来よりケプストラム距離がよく用いられているため、本論文でもケプストラム距離をセグメンテーションパラメータとして検討する。

ケプストラムを計算するに際して、AR モデルでは LPC ケプストラム等の効率の良い計算手法が提案されているが、本論文で用いる ARMA モデルの場合には上述のような手法はまだ提案されていない。そこで本論文では、図 8.1 に示す手法で適応的 ARMA 係数からケプストラムを計算する。

まず、MIS 法により各時刻で求められた ARMA 係数 (第 2 章式 (2.21) の $\hat{p}^s(k)$) よりインパルス応答を求め、その応答波形から FFT によってケプストラムを算出する。

2つのスペクトル間のケプストラム距離は一般に、 $c(n)$ を参照スペクトルのケプストラム、 $c'(n)$ を未知入力スペクトルのケプストラムとして、

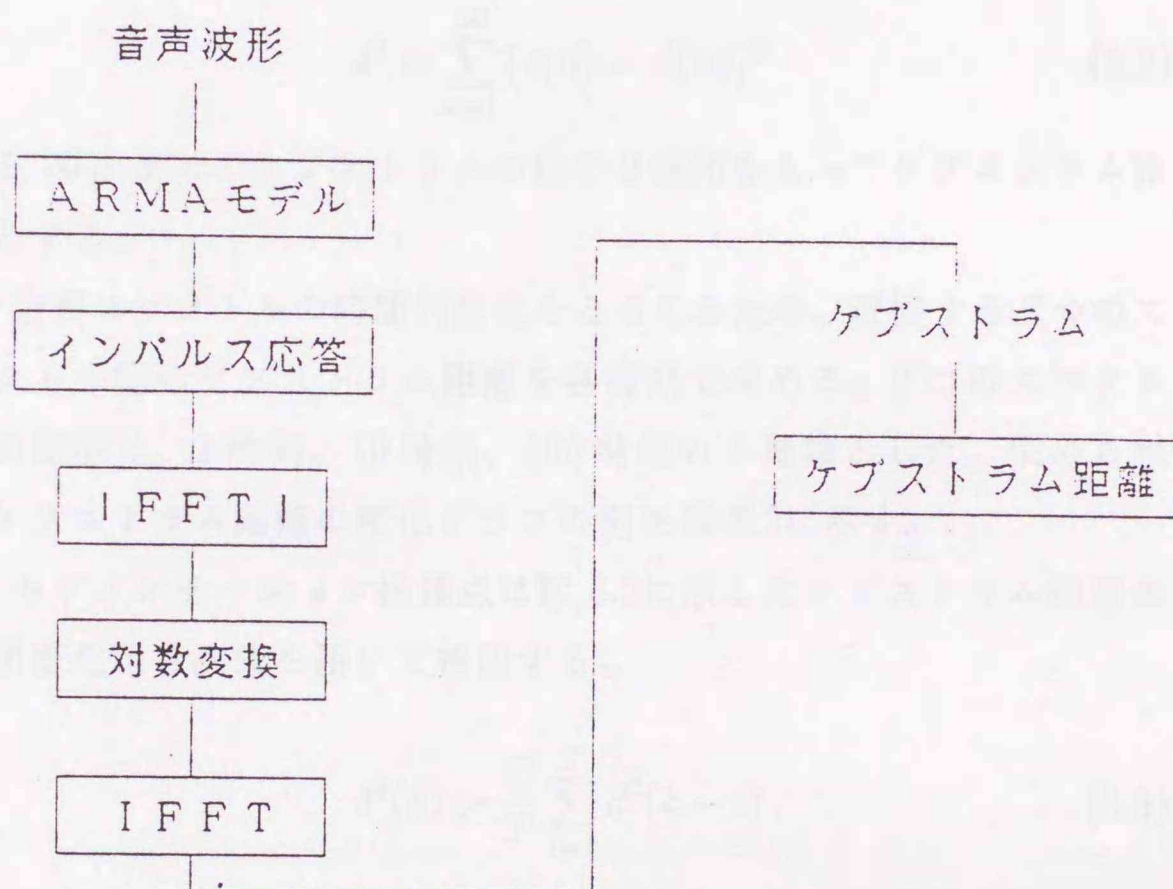


図 8.1: 適応的 ARMA モデルにおけるケプストラム計算手法

$$L^2 = \sum_{n=-\infty}^{\infty} \{c(n) - c'(n)\}^2 \quad (8.1)$$

と表される。このうち $c(0)$ はスペクトル全体の利得項でスペクトルの形状には関与しないことから無視し、また $c(-n) = c(n)$ なので $n < 0$ なる $c(n)$ も無視できる。したがって本論文では、

$$d^2 = \sum_{n=1}^{20} \{c(n) - c'(n)\}^2 \quad (8.2)$$

なる 20 次までのケプストラムの差の 2 乗和をもってケプストラム距離とする。

音声スペクトルの時間的变化をとらえるため、近接する 2 つのスペクトル間のケプストラム距離を各時刻で求める。2 つのスペクトルの間隔は、1 時刻、10 時刻、100 時刻の 3 種類とした。求められたケプストラム距離の変化グラフの例を図 8.2 に示す。

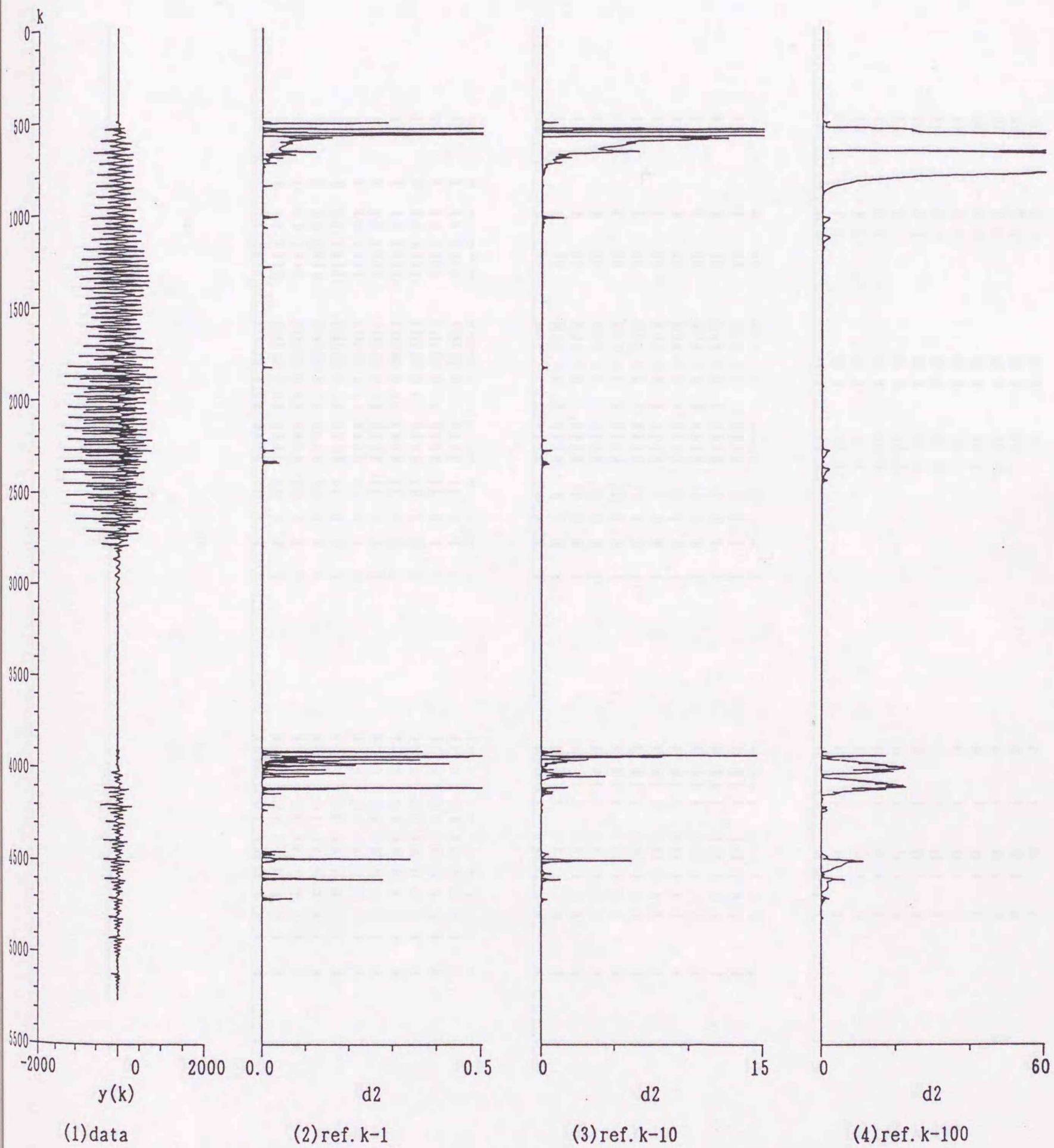
セグメンテーション候補点は図 8.2 に示したケプストラム距離の時間変化から次式を用いて検出する。

$$d^2(k) > \frac{C}{T} \sum_{i=1}^T d^2(k-i) \quad (8.3)$$

式 (8.3) において、 $C = 5.0, T = 50$ とした場合の候補点検出実験結果の例を図 8.3 に示す。図 8.3 において、波線が式 (8.3) を満たす点を示している。図 8.3 に例示される結果より、距離計算をするスペクトルの間隔が 10 時刻のものと 100 時刻のものがセグメンテーションパラメータとして適切であると判断した。間隔が 1 時刻のものは、定数 C や T の値を様々に調整しても、ピッチに依存する候補点を多数出力するため、セグメンテーションには不適切と判断した。

8.2.2 有声・無声分離アルゴリズムの出力点

本セグメンテーション候補検出システムでは、6.2 で述べた有声・無声分離アルゴリズムの出力点をそのまま利用する。この出力点の

図 8.2: ケプストラム距離 d^2 /iwate/

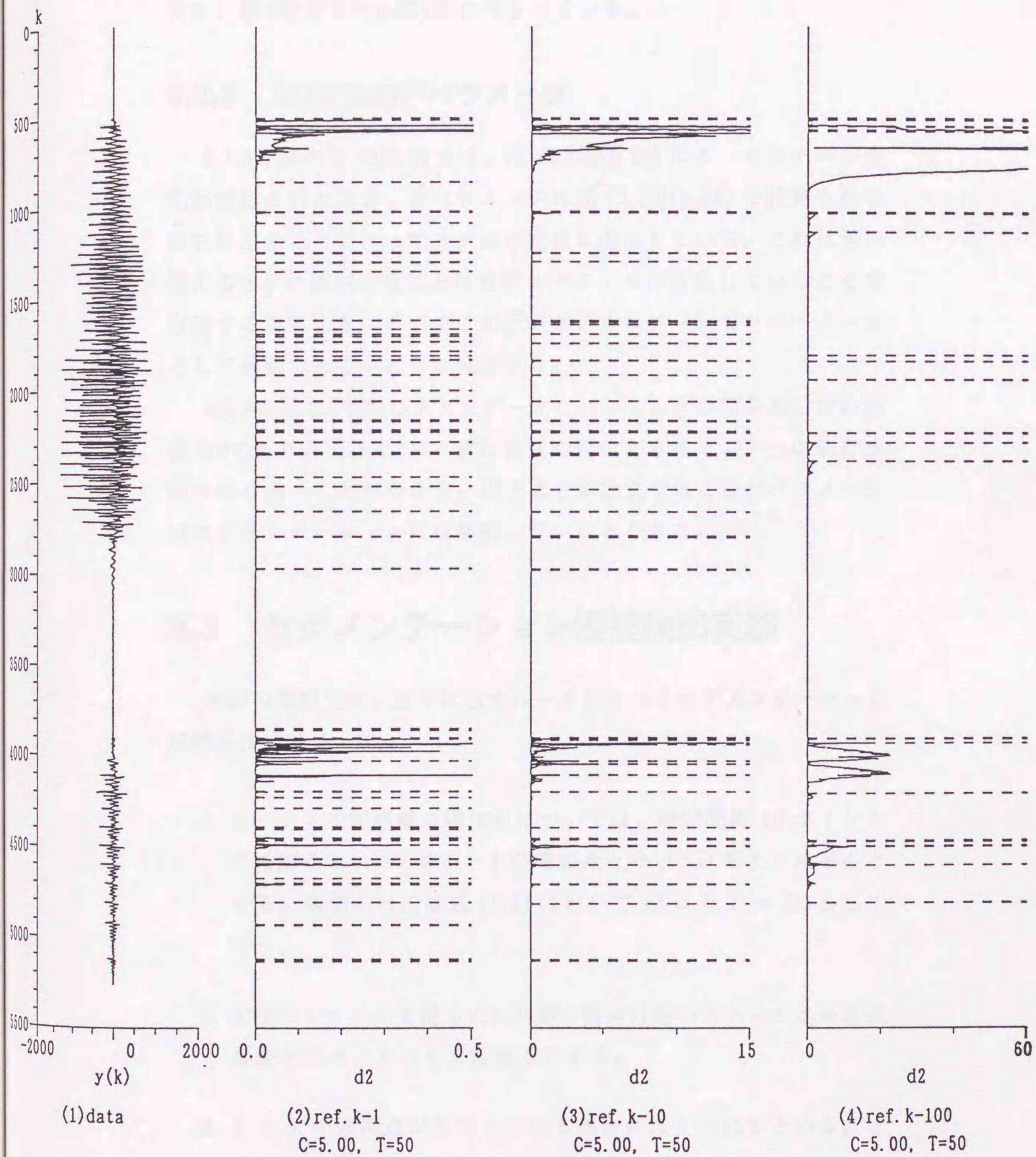


図 8.3: ケプストラム距離に基づく分離点検出例 /iwate/

例は、第 6 章図 6.6(p.81) 等に表示されている。

8.2.3 MIS 法の $\hat{q}^2$ パラメータ

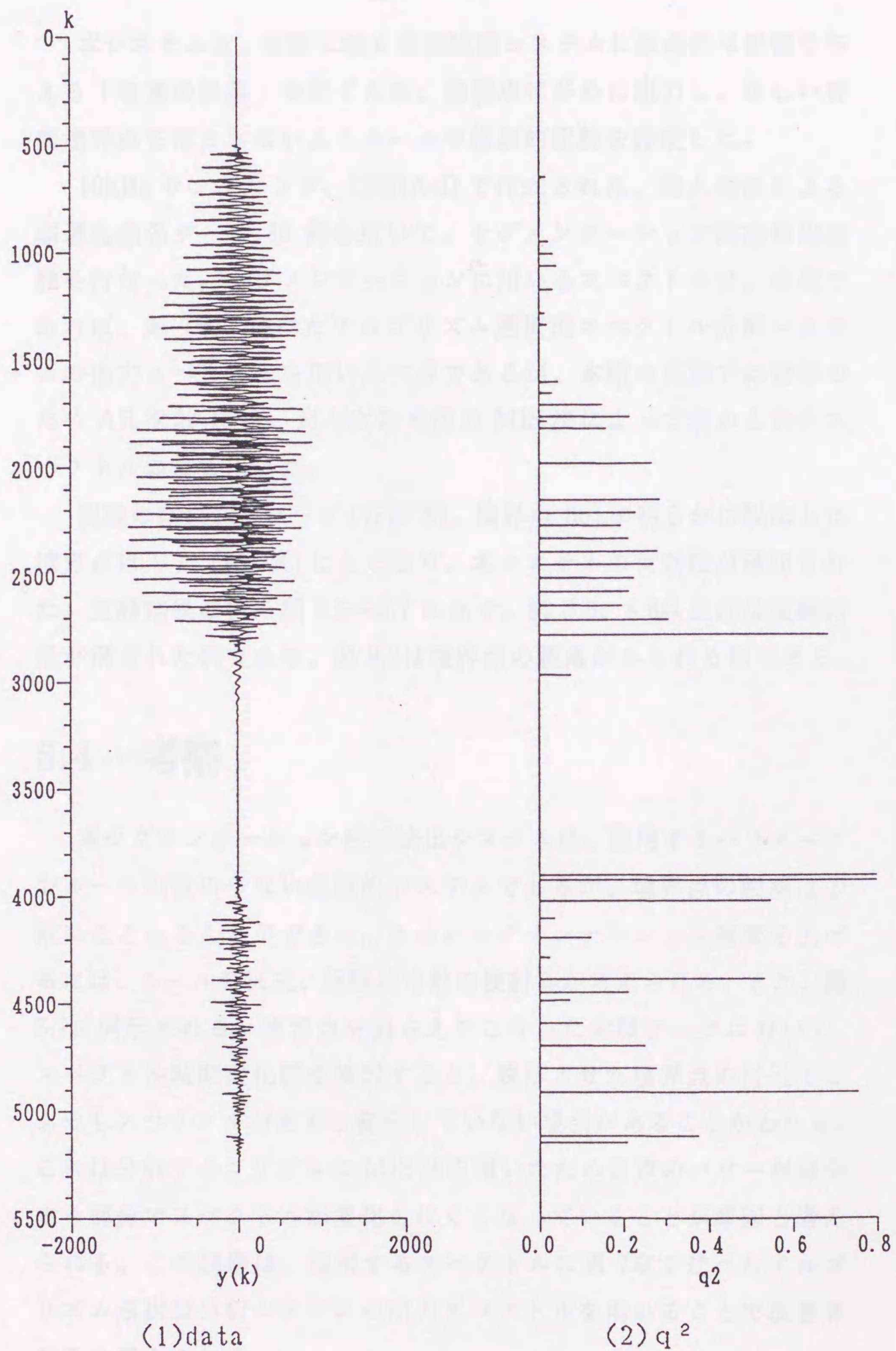
2.1.3で述べた MIS 法では、式 (2.15)(p.14) によってモデルの変化が検出されたとき、 $\hat{q}^2$ パラメータに式 (2.16)(p.14) で計算される値を与えることによってモデルの変化に追従している。これは言い替えると $\hat{q}^2$ に値があることは音声スペクトルが変化していることを意味すると言える。そこでこの $\hat{q}^2$ がセグメンテーションパラメータとして適切であるかどうか検討する。

8.2.1の図 8.2と同じテストデータについて、 $\hat{q}^2$ の値を示したのが図 8.4である。図 8.4より、 $\hat{q}^2$ は音素の変化よりはピッチへの依存度がかなり高いことが分かる。以上より本論文ではこの $\hat{q}^2$ パラメータはセグメンテーションには使用しないこととする。

8.3 セグメンテーション候補検出実験

本節の実験では、以下に示すルールによってセグメンテーション候補を決定する。

1. スペクトル間距離の増加点については、時間間隔 10 ポイントの増加点と、100 ポイントの増加点との AND をとり候補点とする。増加点検出は式 (8.3) において $C = 5, T = 50$ として行う。
2. 前段のシステムで得られた有声・無声分離パラメータ $\hat{a}_1$ の急激な変化点は、そのまま候補点とする。
3. 1. と 2. の候補点が密集している場合には 1 つにまとめる。すなわち 50 時刻中に複数の候補点が検出された場合にそれらをまとめることとする。

図 8.4: $\hat{q}^2$ の時間変化 /iwate/

本システムは、後段に続く音素認識システムに致命的な影響を与える「音素の脱落」を防ぐため、候補点は多めに出力し、正しい音素境界点を落とさないようルールや経験的定数を設定した。

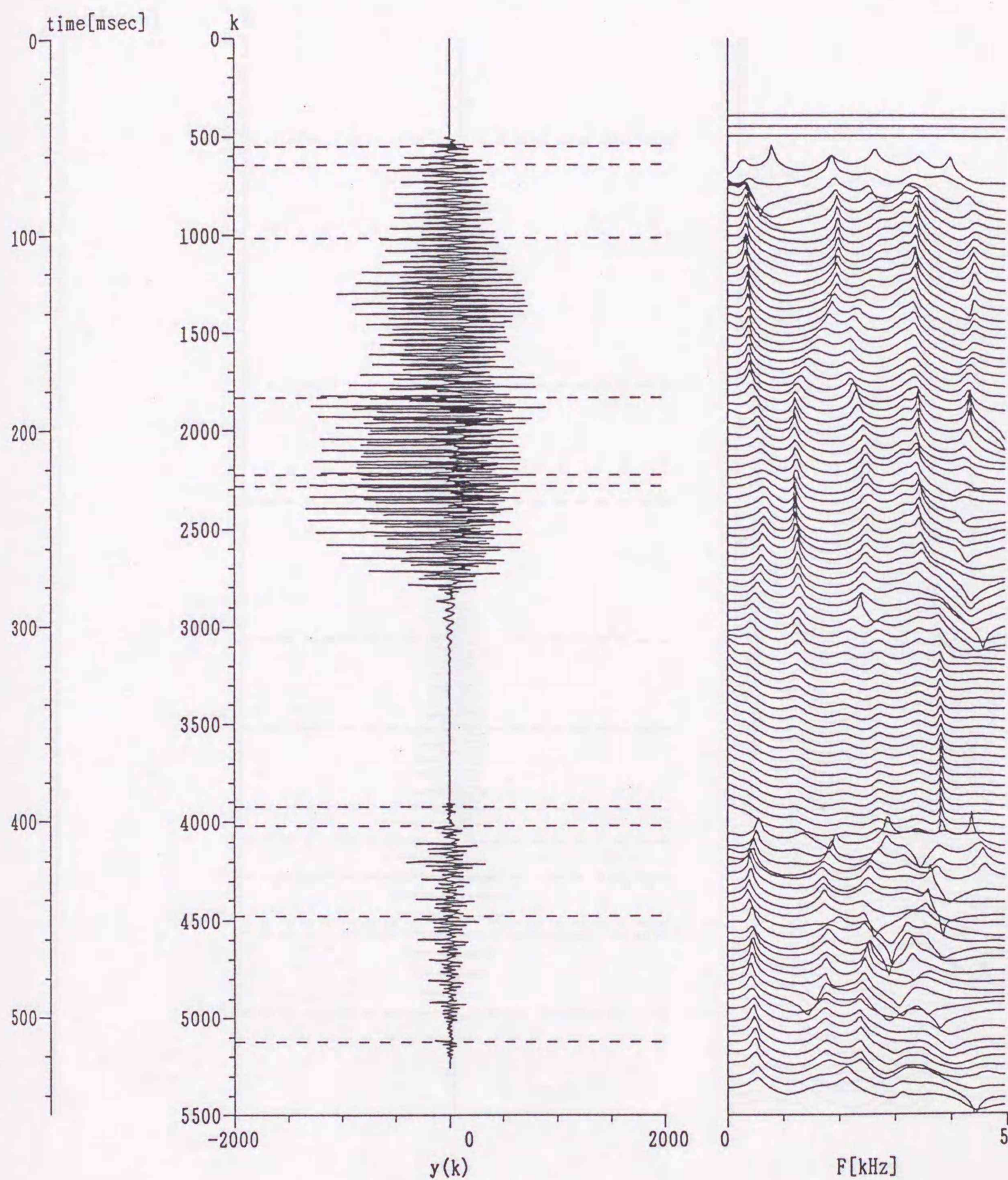
10kHz サンプリング、12bitA-D で作成された、成人男性による都道府県名データ 15 例を用いて、セグメンテーション候補検出実験を行なった。セグメンテーションに用いるスペクトルは、本来であれば、第 7 章で述べたアルゴリズム選択型スペクトル分析システムの出力スペクトルを用いるべきであるが、本節の実験では簡単のため AR 次数 15 次、MA 次数 6 次の MIS 法によって求められたスペクトルのみを用いた。

実験の結果、全データ (音素 81、境界点 66) 中明らかに脱落した境界点は 9 点 (13.6%) にとどまり、本システムの有効性が確認された。実験結果の例を図 8.5~8.7 に示す。図 8.5、8.6 は良好な実験結果が得られた例であり、図 8.7 は境界点の脱落がみられる例である。

8.4 考察

本セグメンテーション候補検出システムは、使用するパラメータやルールの数の少ない実験的システムであるが、境界点の脱落は少数にとどめることができた。さらにセグメンテーション精度を上げるには、ルールの拡充、経験的定数の検討等が考えられる。また、図 8.7 に例示される、境界点を捕らえそこなった実験データにおいて、スペクトル時間変化図を検討すると、脱落させた境界点の付近でそもそもスペクトルがあまり変化していない場合があることがわかる。これは分析アルゴリズムに MIS 法を用いたため音声のパワーが減少する部分でスペクトルが変化しにくくなっていることが原因と考えられる。この問題は、使用するスペクトルに第 7 章で述べたアルゴリズム選択型分析システムの出力スペクトルを用いることで改善されることが考えられる。

また本システムでは、ARMA モデルのケプストラムを計算する



(1) segmentation

d2(ref.k-10) : C=5.00, T=50

d2(ref.k-100): C=5.00, T=50

EXPOINT=50

(2) MIS

図 8.5: セグメンテーション実験結果 /iwate/

良好な結果が得られた例

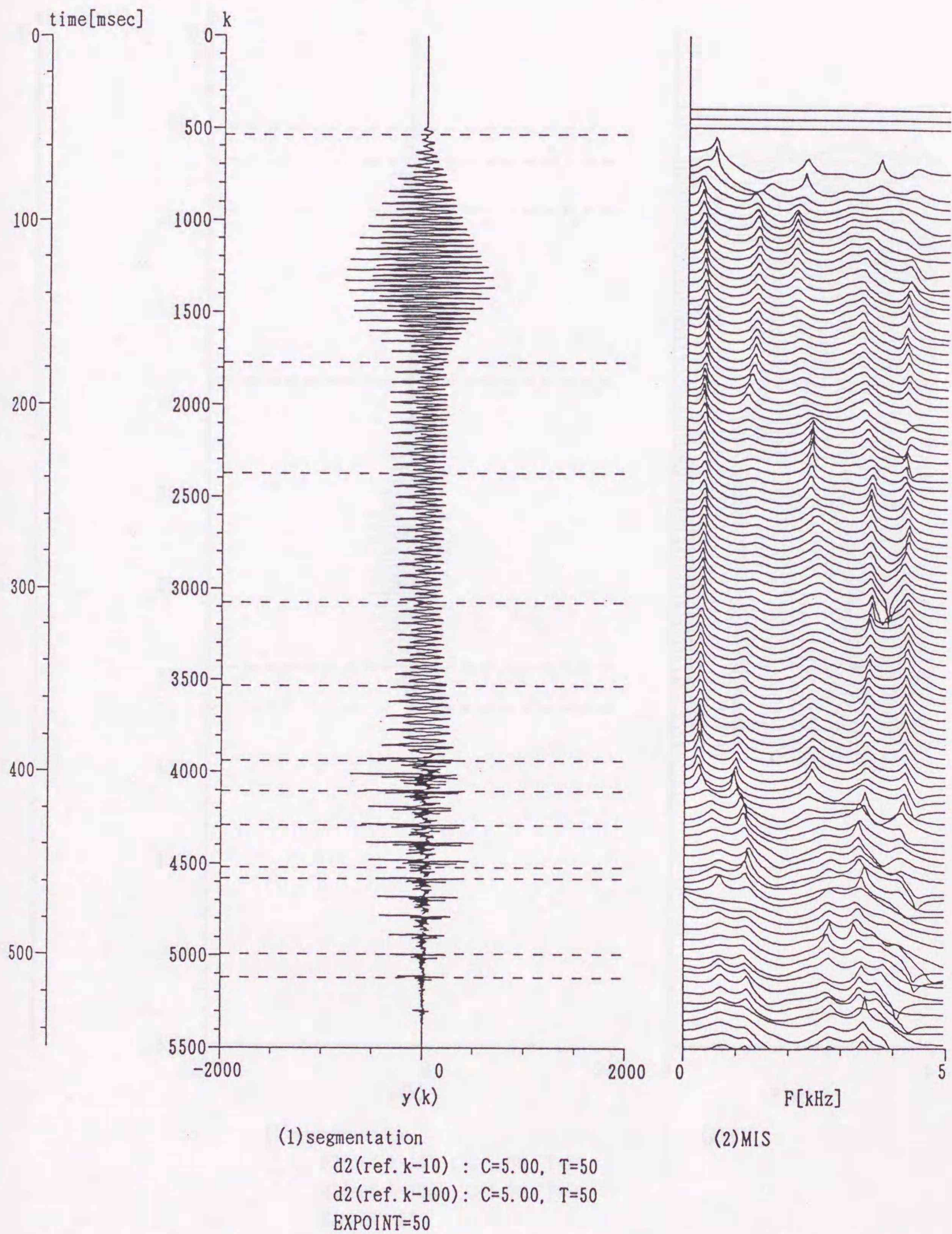
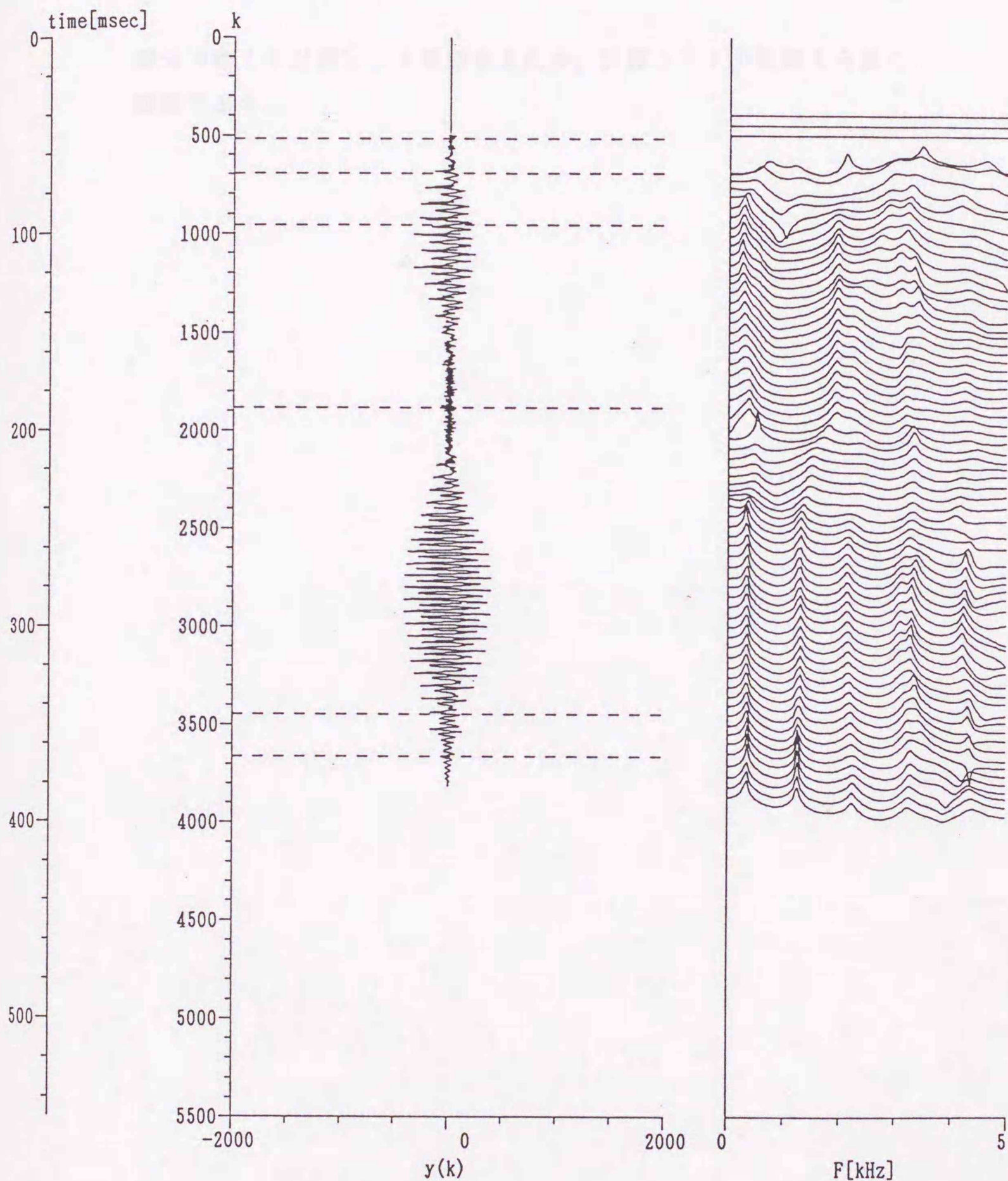


図 8.6: セグメンテーション実験結果 /gumma/

良好な結果が得られた例



(1) segmentation

d2(ref. k-10) : C=5.00, T=50

d2(ref. k-100): C=5.00, T=50

EXPPOINT=50

(2) MIS

図 8.7: セグメンテーション実験結果 /gifu/

境界点の脱落がみられる例

部分で多くの計算コストを要するため、計算コストの軽減も今後の課題である。

Chapter 9

音声符号化・合成システム

Part II

音声の符号化・合成

Chapter 9

音声符号化・合成システム

9.1 まえがき

前章までの第1部では、適応信号処理と知識工学的手法を融合し、従来より時間軸上でも、周波数軸上でも精度の高い、新しい音声スペクトル分析システムを構築した。その応用のひとつとして連続音声のセグメンテーション候補検出を行い、連続音声認識への応用の可能性を示した。

本章以降では第2部として、音声スペクトル推定システムのもうひとつの重要な応用である、音声の符号化・合成をとりあげる。第1章でも述べたように、音声情報処理の分野において認識と符号化・合成、もしくは圧縮は、音声調音モデルの記述であるスペクトルの工学的応用という点で密接な関係がある。

第1部で、適応 ARMA スペクトル推定システムは高精度なスペクトルを推定できることが示され、また音声認識の分野への応用の可能性が示されたが、このことは分析・認識と表裏一体をなす符号化・合成の分野でも、適応 ARMA スペクトル推定システムの有効性が期待できることを示している。

一方近年、移動体通信やボイスメール等の蓄積型通信の需要が急速に高まってきており、音声符号化・圧縮は最も注目される分野のひとつとなっている。

高精度適応 ARMA スペクトル推定アルゴリズムは、音声認識等の分野ではその有効性の研究が行われているが⁽¹⁰⁸⁾、符号化の分野での有効性の研究は数少ない⁽⁵³⁾⁽⁶⁰⁾。さらに、これらの研究では符号化・合成モデルの構築に主眼がおかれていると考えられ、実用的なビットレートを設定した音声符号化システムの研究はきわめて少ない。

そこで本論文第2部では、音声符号化システムのスペクトル推定部に、第1部で用いた高精度適応 ARMA スペクトル分析システムを導入する。残差信号の簡単な量子化により適応 ARMA 分析スペクトルの有効性を確認した後、CELP 符号化方式⁽⁶³⁾を用いて実用的なビットレートを持つ新しい音声符号化システムを構築し、その有効性について検討する。

9.2 音声符号化システムにおける適応 ARMA 分析の必要性

近年音声を高度に圧縮する MPC⁽⁵⁴⁾⁽⁵⁵⁾、CELP⁽⁶³⁾⁽⁶⁴⁾等の低ビット音声符号化方式が盛んに研究されているが、それらの符号化方式のほとんどは、音声スペクトルを抽出するアルゴリズムに一括型の AR モデルを仮定した LPC に基づく手法を採用している。

従来の一括型 AR モデルに基づいた手法では、音声スペクトルが急峻に変化する過渡部分や、アンチホルマントを含むと考えられる鼻子音等で十分なスペクトル推定精度が得られず、その結果本来スペクトルで表現されるべき声道調音情報の一部が残差信号の側に含まれてしまうと考えられる。すなわち従来手法では、残差信号に音源情報と一部の声道情報が含まれていると考えられる。

スペクトルが高精度に推定されれば、声道情報はスペクトルによってかなり正確に表現されるため、残差信号に漏れ出す声道情報が減り、その結果残差信号を表現するのに必要な情報量が削減され

ると考えられる。換言すれば、残差信号の表現に一定の情報量を割り当てた場合には、高精度スペクトル推定手法を用いることにより高品質な符号化・合成が行えると期待できる。

9.3 音声符号化システム

一般的な音声符号化システムの概略図を図 9.1 に示す。入力音声波形 $y(k)$ はスペクトル包絡推定フィルタ (短期予測逆フィルタ) を通過することによってスペクトルの包絡成分が除去され、残差信号 $x(k)$ となる。フィルタの係数は量子化され合成器側へ伝送される。残差信号 $x(k)$ は CELP⁽⁶³⁾ 等の適切な符号化手法で効率よく符号化された後、合成器側へ伝送される。図 9.1 において、11.1(p.117) で述べる長期予測フィルタは、残差符号化部に含めて考えるものとする。

従来の音声符号化システムのほとんどは、図 9.1 におけるスペクトル包絡推定フィルタに、PARCOR 等の AR モデルに基づくフィルタを用いている。本論文ではこのスペクトル推定フィルタに、第 2 章で述べた適応 ARMA ラティスフィルタや、重み付き RLS 法、MIS 法等の適応 ARMA スペクトル推定アルゴリズムを導入することにより、スペクトル推定の精度を向上し、再生された合成音声の品質向上をはかる。

9.4 符号化・合成における音質評価

符号化・合成システムによって得られた再生音声の品質評価方法は、大きく分けて主観評価法と客観評価法の 2 種類ある。評価の対象となる再生音声は、最終的には人間の耳で聴かれることから、複数の人間が実際に聴取して採点する主観評価法は有効な評価法であると考えられる。著名な主観評価法としては、シェッフエの一対比較法⁽⁵⁷⁾等がある。

しかし主観評価法はその名の通り評価に人間の主観を用いた

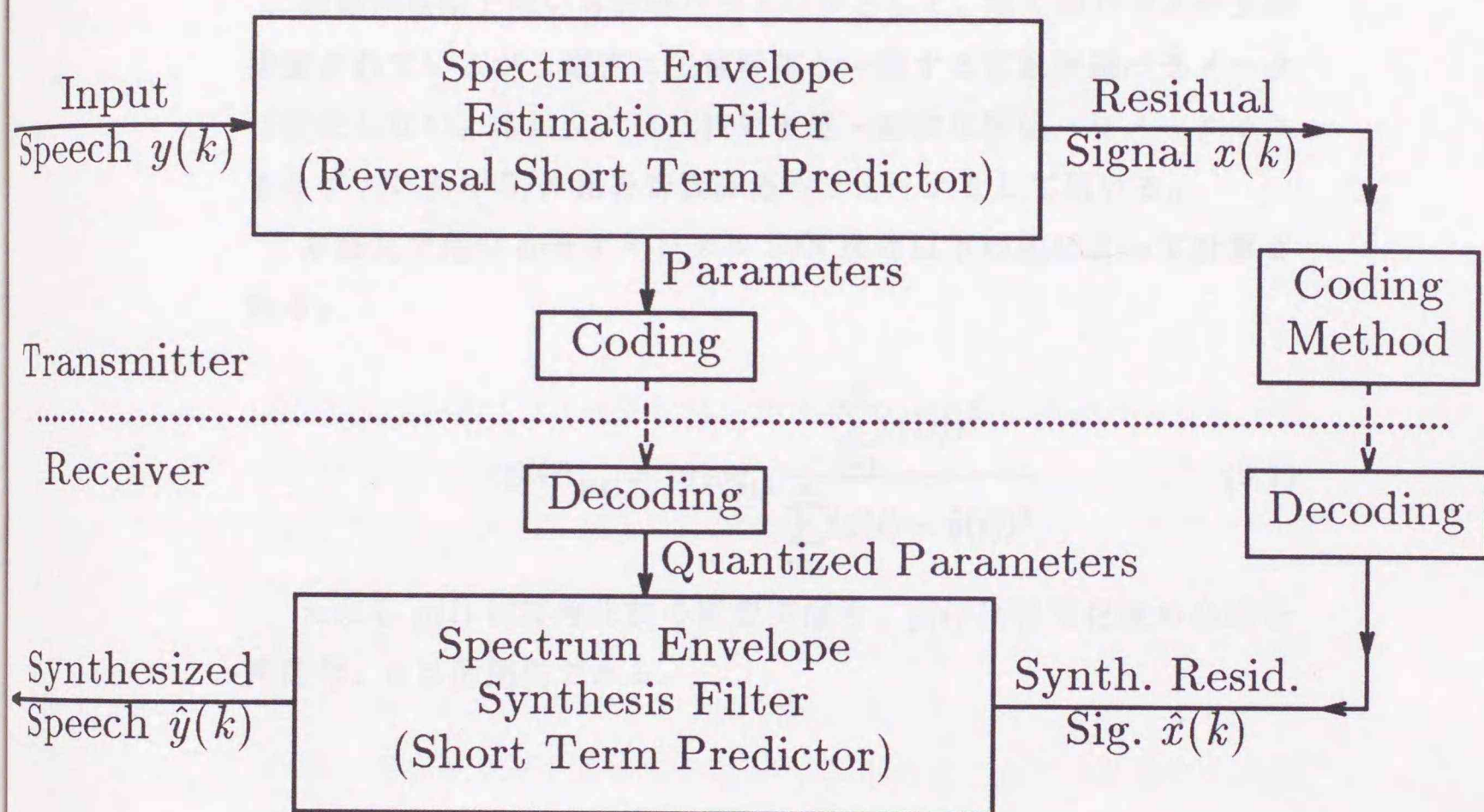


図 9.1: 一般的な音声符号化システム

め、評価を行う人間集団の訓練度等の違いによって評価に差が生じることも考えられる。また比較的多くの手間を必要とするため、大量の実験データの評価は困難を伴う。

そこで本論文では、再生音声の評価は主に客観評価法を用いて行い、主観評価は筆者らによる簡単な聴取試験にとどめる。

客観評価法で用いる評価パラメータとして、多くのパラメータが提案されているが、厳密に主観評価と一致する客観評価パラメータは存在しない。本論文では比較的敏感・厳密な評価パラメータであるセグメンタル S/N 比を音質評価パラメータとして用いる。

本論文で用いるセグメンタル S/N 比は以下の式によって計算される。

$$\text{SNR}_{\text{seg}} = 10 \log_{10} \frac{\sum_{i=1}^n (y(i))^2}{\sum_{i=1}^n (y(i) - \hat{y}(i))^2} \quad (9.1)$$

ただし $y(i)$ は符号化前の原音声信号、 $\hat{y}(i)$ は符号化後の合成音声信号、 n は区間長である。

Chapter 10

残差を線形量子化した符号化システム

本章では図 9.1 の残差符号化部で用いる符号化手法として、単純な線形量子化を用いた場合について考察する⁽¹⁰⁴⁾。スペクトル包絡推定アルゴリズムには、2.2.2 で述べた規格化 ARMA ラティスモデルの推定アルゴリズム⁽²¹⁾を用いる。規格化 ARMA ラティスフィルタから得られた残差波形と従来の PARCOR フィルタ⁽⁷⁾で求めた残差波形をそれぞれ同じビット数で線形量子化することにより、残差の持つ情報量を等しくすることができ、再生された合成音声の品質からスペクトル包絡推定アルゴリズムの性能を評価することが可能となる。

10.1 実験システム

本章で用いた実験システムを図 10.1 に示す。実験に用いた音声サンプルは成人男性話者 1 名によって発声された 47 都道府県名で、10 kHz サンプリング、12 bit-AD で作成されたものである。規格化 ARMA ラティス、PARCOR 共にパラメータの数は 8 個 (ARMA ラティス: AR 6 次、MA 1 次、追加型 1 次、PARCOR: AR 8 次) である。PARCOR のフレーム周期は 120 サンプル (12 msec) であり、

適応信号処理である規格化 ARMA ラティスの係数は、係数に関する情報量を PARCOR に合わせるため 120 サンプルの平均を用いる。

それぞれの係数は不均一ビット割当⁽⁵⁹⁾によって量子化される。本実験で使用したビット割当のパターンを表 10.1 に示す。

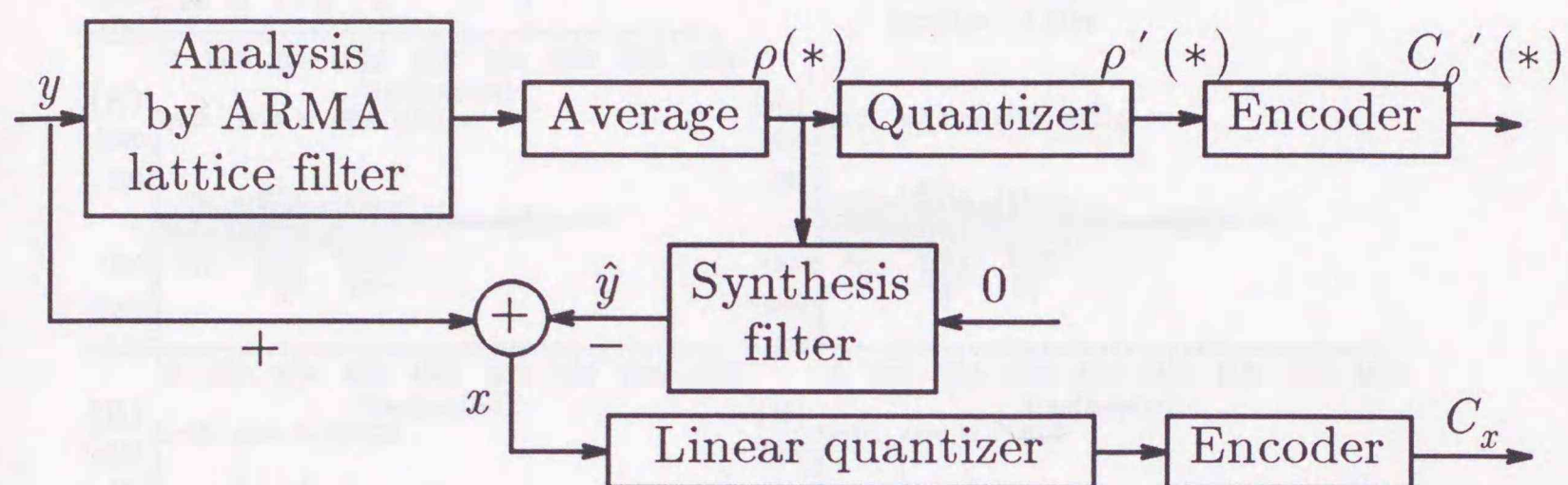
残差信号は 120 サンプル毎に求められるそれぞれのフィルタの係数を用いて、2.2.5(p.29) で述べた疑似逆フィルタによって計算される。本章の実験では、残差は量子化前のフィルタ係数から求められる。得られた残差信号は、短区間 (360 サンプル=36msec) の極大値によって正規化され、極大値は 6 ビット、正規化された信号は 3 ビットで量子化される⁽⁶⁰⁾。

10.2 実験結果

実験結果の例を図 10.2, 10.3 に示す。図 10.2, 10.3 において、(A) は規格化 ARMA ラティスによる合成音声、(B) は PARCOR による合成音声、(C) はセグメンタル SNR の比較グラフ、(D) は規格化 ARMA ラティスのセグメンタル SNR 値から PARCOR のそれを引いた値で、正であれば規格化 ARMA ラティスの方が SNR の面ですぐれていることを示す。

図 10.2, 10.3 によると、規格化 ARMA ラティスフィルタを用いた符号化・合成システムは、信号パワーの極めて小さい部分を除くほとんどの部分で、PARCOR と同等、もしくは PARCOR よりすぐれたセグメンタル SNR を示している。特に、アンチフォルマントを含んでいると考えられる /N/・/m/ 等の鼻子音部では、1 次の MA パラメータを導入したことによりスペクトルの傾きがより正確に表現され、高いセグメンタル SNR が得られている。また、係数の量子化ビット数を少なくすると (図 10.2, 10.3 の (b))、PARCOR との差はさらに大きくなり、より多くの点で ARMA ラティスがすぐれた SNR を示す傾向があることが分かる。このことから、規格化 ARMA ラティスフィルタは低ビット音声符号化に適していると考えられる。

(a) Coder part



(b) Decoder part

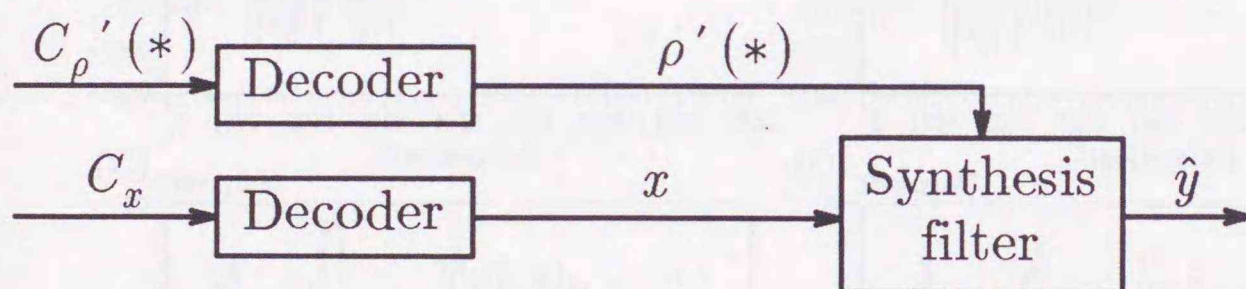


図 10.1: 残差の線形量子化による規格化 ARMA ラティス符号化システム

Name of bit pattern	Bit allocation	Total(bits/120points)
bits.6.1-3	9 7 6 6 5 5 5 5	48
bits.6.1-5	7 5 4 4 3 3 3 3	32

表 10.1: ラティス係数の量子化ビット割当

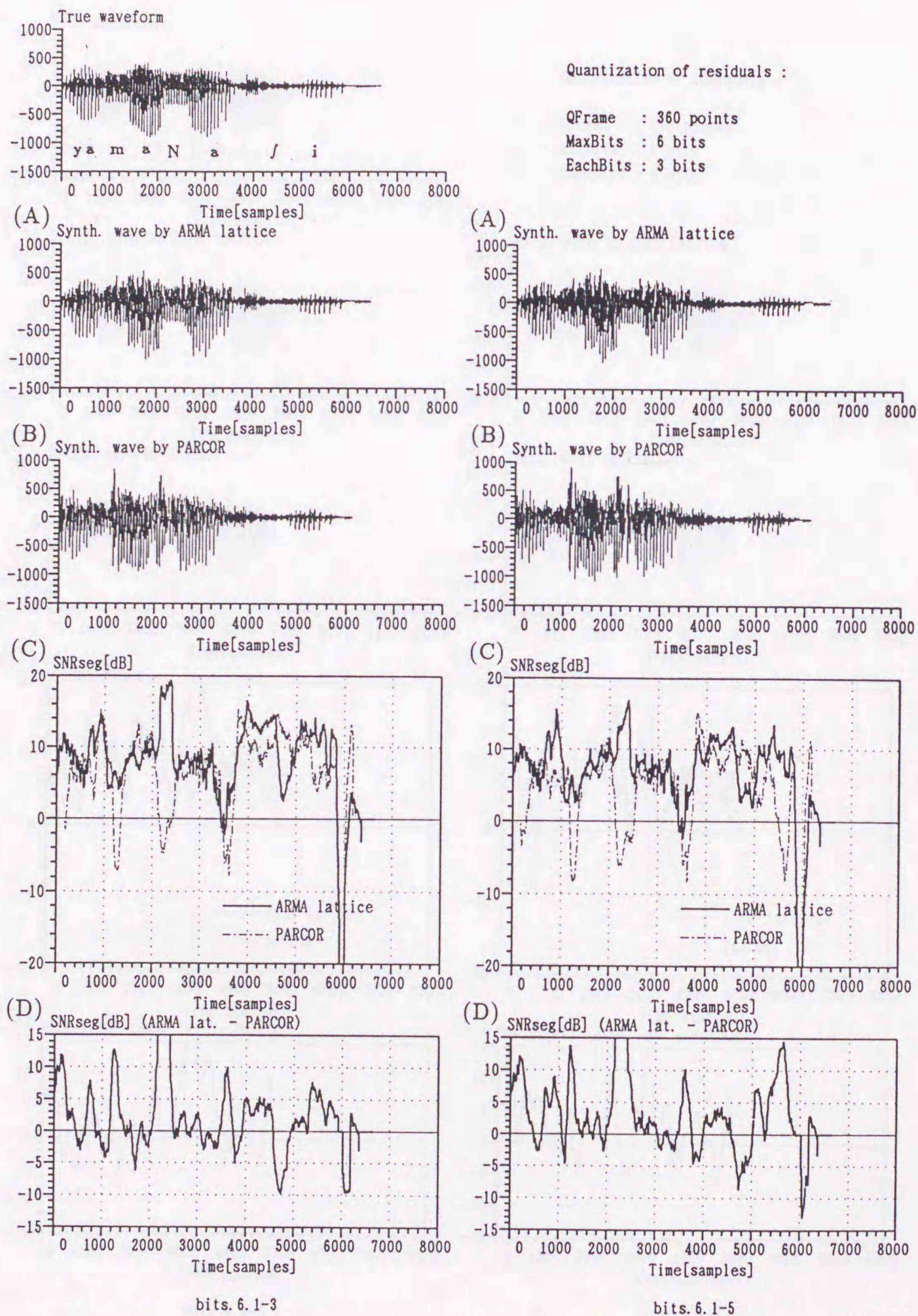


図 10.2: 残差の線形量子化による符号化システムの実験結果

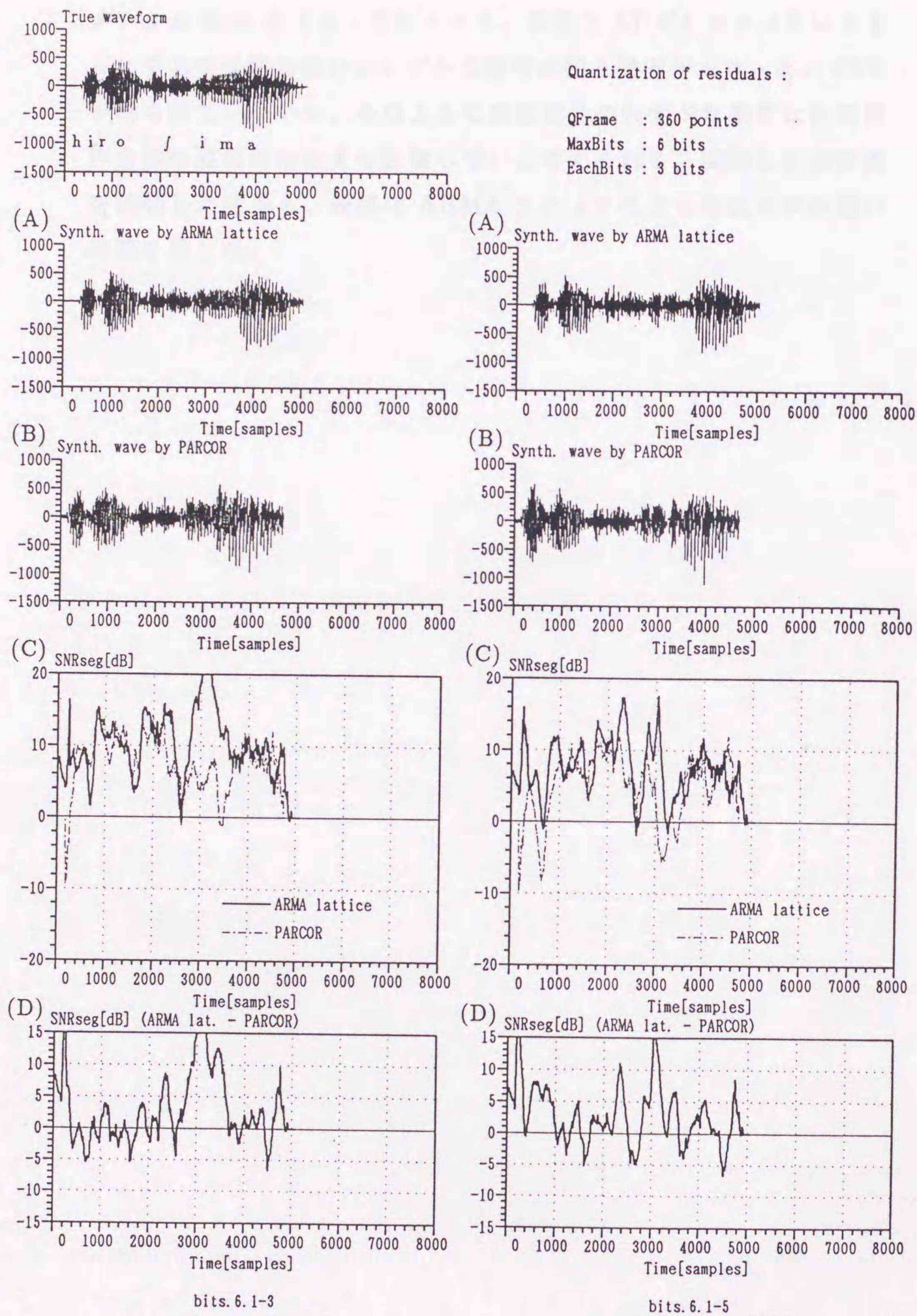


図 10.3: 残差の線形量子化による符号化システムの実験結果

信号パワーの極めて小さい無音部分等では、わずかな雑音もセグメンタル SNR を大きく悪化させる。規格化 ARMA ラティスによるシステムでは無音部分にわずかな信号が残る傾向があり、よい SNR が得られていないが、そのような無音部分のわずかな雑音は合成音声全体の品質にはあまり影響しないと考えられる。実際に合成音声を試聴した結果も、規格化 ARMA ラティスによる合成音声が高い品質を示した。

11.1 長期予測フィルタ

長期予測フィルタ (Long Term Prediction, LTP) は音声の周期性を利用して、過去の音声信号を現在の音声信号に予測するフィルタである。音声信号は周期的な成分を含むため、過去の音声信号を現在の音声信号に予測することで、音声信号の周期性を利用することができる。長期予測フィルタは、音声信号の周期性を利用して、過去の音声信号を現在の音声信号に予測するフィルタである。

$$x(n) = \frac{1}{L} \sum_{k=0}^{L-1} x(n-k) \quad (11.1)$$

ここで、 $x(n)$ は音声信号、 L は長期予測の遅延時間である。式 (11.1) は、音声信号の周期性を利用して、過去の音声信号を現在の音声信号に予測するフィルタである。

Chapter 11

長期予測フィルタと CELP 符号化方式

本論文ではより実用的な符号化システムとして、第 9 章図 9.1 で示した一般的な符号化システムの残差符号化部に、長期予測フィルタと CELP 符号化方式を導入する。本章では本論文のシステムで用いる長期予測フィルタ、聴覚重み付けフィルタ、および CELP 符号化手法について述べる。

11.1 長期予測フィルタ

長期予測フィルタ (Long Term Predictor, LTP)⁽⁶²⁾は波形の周期性を表現するフィルタであり、周波数領域では音声スペクトルの微細構造を表わしている。短期予測逆フィルタの出力である残差波形 $x(k)$ を長期予測逆フィルタに入力すると、残差波形に残っている周期性が取り除かれ、ガウス雑音に近い駆動波形 $d(k)$ が得られる。本報告で用いた長期予測フィルタの伝達関数は、

$$H_{LTP}(z) = \frac{1}{1 - bz^{-M}} \quad (11.1)$$

であり⁽⁶⁴⁾、この伝達関数にはピッチ周期 M とピッチ係数 b の 2 つのパラメータがある。

LTP を符号化システムに組み込んで用いる場合、この 2 つのパラメータの決定法について、決定の時期の違いから、Open loop 方式と Closed loop 方式の 2 種類の方式がある。

Open loop method 符号化に先立って、あらかじめ分析型の短期予測フィルタ (STP i.e., スペクトルフィルタ) の残差信号から 2 つのパラメータを推定しておく方式

Closed loop method STP の残差信号は用いず、符号化と同時に総当たり的な方法で 2 つのパラメータを推定していく方式

Closed loop 方式の詳細については後の 11.3.2 で述べるので、ここでは Open loop 方式符号化システムにおける LTP パラメータ決定法について述べる。

Open loop 方式での LTP 係数推定は、まず実音声 (男性) の基本ピッチの範囲を考慮して、ピッチ周期 M の値を $20 \leq M \leq 200$ (10kHz サンプルング時で 50~500Hz) 程度に制限する。このようにすることによって M のとりうる値は有限となり、すべての M について最適な b を推定することが可能となる。すなわち、STP 分析フィルタから求められる残差信号と、LTP の出力信号との誤差の 2 乗和を評価関数とし、それぞれの M の値について評価関数を最小にする b とそのときの評価関数値 V_M を求める。最後に V_M が最小となる M を選び出し、そのときの M と b を長期予測フィルタ係数の推定値とする。実際の符号化・合成システムでは、 M と b は量子化されて伝送される。

なお Open loop の場合、LTP 係数推定は符号化に先立って行われるが、本論文では LTP 係数推定時に用いる残差信号を計算する STP の係数は、量子化後のものを用いる。すなわち STP の残差信号は、STP 係数推定アルゴリズムから計算されるものは用いずに、量子化後の STP 係数から逆フィルタを構成して新しく計算し直したものを用いる。このことは Open loop の場合 STP 順フィルタのみなら

ず逆フィルタの安定性も必要になることを意味する。これは ARMA モデルで STP を表現した場合、MA 部も安定、すなわち零点も全て z 平面の単位円内に存在しなければならない。この制約が実際に実験を行うときにしばしば問題を起こすことがある。

11.2 聴覚重み付けフィルタ

聴覚重み付けフィルタ (Perceptual Weighting Filter, PWF)⁽⁶²⁾ はノイズシェイピングを行なうフィルタである。人間の聴覚のマスクング現象のため、音声の周波数成分のうち、エネルギーが集中して大きなパワーのある部分には比較的大きな雑音が存在しても知覚されにくく、逆にパワーの小さい周波数では小さな雑音でもよく知覚される。観測実音声と合成音声との誤差の 2 乗和を単純に最小化することは、誤差信号を白色のまま最小化することと等価であるが、上述の知覚特性を考慮すると望ましくない。そこで、原音声のスペクトルと同様のスペクトルをもつ逆フィルタに誤差信号を通過させ、その出力信号の 2 乗和を最小にすることによって、原音声と同様のスペクトルを持った誤差信号を最小化することができる。このフィルタが聴覚重み付けフィルタであり、原音声と合成音声との誤差パワーが比較的大きくても聴覚的に高品質な合成音声を得ることができる。

聴覚重み付けフィルタの周波数特性は、原音声のスペクトル包絡をそのまま用いるよりも、極の帯域を広げて鈍らせたものが効果的であることが知られている⁽⁶²⁾。

CELP 符号器における聴覚重み付けフィルタの原理的な位置は、後の 11.3 の図 11.1(p.123) に示されている。

11.2.1 AR モデルの場合

原音声のスペクトル包絡が AR モデルで表現されている場合、聴覚重み付けフィルタの伝達関数は、

$$H_{PWF\cdot AR}(z) = \frac{A(z)}{A(z/\alpha)} = \frac{1 + \sum_{i=1}^n \hat{a}_i z^{-i}}{1 + \sum_{i=1}^n \hat{a}_i \alpha^i z^{-i}} \quad (11.2)$$

で表わされる⁽⁶²⁾。ただし $\hat{a}_i (i = 1, \dots, n)$ は推定 AR 係数、 α は極の広帯域化の度合を表わす定数 ($\alpha \leq 1.0$) であり、サンプリング周波数等を考慮して設定される⁽⁶²⁾⁽⁶⁵⁾。

後の 11.3 で述べる実際の CELP 符号器では、計算コスト減少のため、聴覚重み付けフィルタは最適雑音コード探索ループの外側になるように移動されている (図 11.2(p.124), 図 11.3(p.126) 参照)⁽⁶⁴⁾。すなわち、STP と PWF の縦続接続の部分は、

$$H_{STP\cdot AR}(z) \cdot H_{PWF\cdot AR}(z) = \frac{1}{A(z)} \cdot \frac{A(z)}{A(z/\alpha)} = \frac{1}{A(z/\alpha)} \quad (11.3)$$

となり、STP を通した後真信号との誤差を計算をして PWF を通す図 11.1 の場合よりも、計算コストが軽減される。

11.2.2 ARMA モデルの場合

本節では ARMA モデルの聴覚重み付けフィルタを検討する。前節の AR 聴覚重み付けフィルタが、AR 逆フィルタと、極を広帯域化した AR 順フィルタとの縦続接続になっていることに注目し、ARMA 聴覚重み付けフィルタを、ARMA 逆フィルタと、極および零点を広帯域化した ARMA 順フィルタとの縦続接続とする。すなわち、

$$H_{PWF\cdot ARMA}(z) = \frac{A(z)}{B(z)} \cdot \frac{B(z/\alpha)}{A(z/\alpha)} = \frac{1 + \sum_{i=1}^n \hat{a}_i z^{-i}}{1 + \sum_{j=1}^m \hat{b}_j z^{-j}} \cdot \frac{1 + \sum_{j=1}^m \hat{b}_j \alpha^j z^{-j}}{1 + \sum_{i=1}^n \hat{a}_i \alpha^i z^{-i}} \quad (11.4)$$

である。ただし $\hat{a}_i (i = 1, \dots, n)$ は推定 AR 係数、 $\hat{b}_j (j = 1, \dots, m)$ は推定 MA 係数、 α は極および零点の広帯域化の度合を表わす定数 ($\alpha \leq 1.0$) である。

この ARMA 聴覚重み付けフィルタは ARMA 逆フィルタを含んでいるため、AR 部のみならず MA 部も安定 (z 平面で極および零点が単位円内) でなければならない。2.2.2 で述べた規格化 ARMA ラティスフィルタは、順フィルタの安定性は理論的に保証されているが、逆フィルタの安定性は保証されていない。実際の実験では零点が単位円外に出るスペクトルがかなりの数見られるため、この形の ARMA 聴覚重み付けフィルタは問題がある。

他の文献⁽⁶¹⁾等で用いられているような、ARMA 係数のうち AR 部のみを用いた、式 (11.2) と同じ形の聴覚重み付けフィルタを使用することも考えられるが、その場合は式 (11.3) で示される計算コストの軽減が計れないため、計算コスト上問題がある。

以上の問題は現時点では未解決の課題であるので、本論文では暫定的に、ARMA モデルを用いる場合には聴覚重み付けフィルタを用いずに実験を行うこととする。

11.3 CELP 符号化方式

長期および短期予測フィルタを縦続接続した合成フィルタに、前節までで得られた駆動波形 $d(k)$ を入力すると音声波形が再生されるが、合成フィルタのパラメータが量子化されている場合、符号化前の真の駆動波形 $d(k)$ を合成フィルタに入力しても真の音声波形は得られない。CELP 符号化方式⁽⁶³⁾では、合成フィルタへの入力駆動波形を、ガウス雑音から作られたコードブックによって全く新しく生成する。この新生駆動波形で合成フィルタを駆動し、真の音声波形に最も近い合成音声を生成するコードワードを最適駆動波形として、対応するコード番号を伝送する。この方式によって、真の駆動波形を用いて合成した合成音声よりも、より真音に近い合成音声を得る

ことができる。

CELP 符号化方式の原理図を図 11.1 に示す。

11.3.1 Open loop CELP 符号化方式

LTP 係数決定をあらかじめ行う Open loop 型の CELP 符号化方式において、符号化時に推定しなければならないパラメータは、雑音コードブックの形状 (雑音コード番号) k と利得 g の 2 つのパラメータである。これらの推定は、まずコードワード長に等しい長さのフレームを考え、フレーム内の観測音声信号 y を PWF に通して得られた聴覚重み付け観測信号 y' から、過去のフレームの影響を取り除き y_{cur} を得る。利得 g を 1 とし、コード番号 k の雑音コードワードを駆動波形として合成フィルタを駆動して、合成波形 $\hat{y}_{k0}$ を得る。聴覚重み付け観測信号 y' と合成信号 $\hat{y}_{k0}$ との誤差の 2 乗和を評価関数とし、 y_{cur} と $\hat{y}_{k0}$ から、評価関数を最小とする g と、そのときの評価関数の値 V_k が計算される。 g と V_k をすべての k について求め、 V_k が最も小さくなる k とそのときの g を最適パラメータとする⁽⁶⁵⁾。実際の符号化・合成システムでは、 g は量子化されて伝送される。本論文ではこの g の量子化は、最大・最小値を設定する固定量子化で行う。

本論文で用いる Open loop CELP 符号化方式のブロック図を図 11.2 に示す。

11.3.2 Closed loop CELP 符号化方式

Closed loop 型の CELP 符号化方式において、符号化時に推定しなければならないパラメータは、LTP のピッチ周期 M とピッチ係数 b 、雑音コードブックの形状 k と利得 g 、計 4 つのパラメータである。

Closed loop 型の CELP 符号化方式において LTP の機能は、過去の駆動波形をピッチ周期 M で繰り返した波形をコードワードに持つ波形コードブックと、そのコードブックの出力にピッチ係数 b を乗算する乗算器とで実現される。

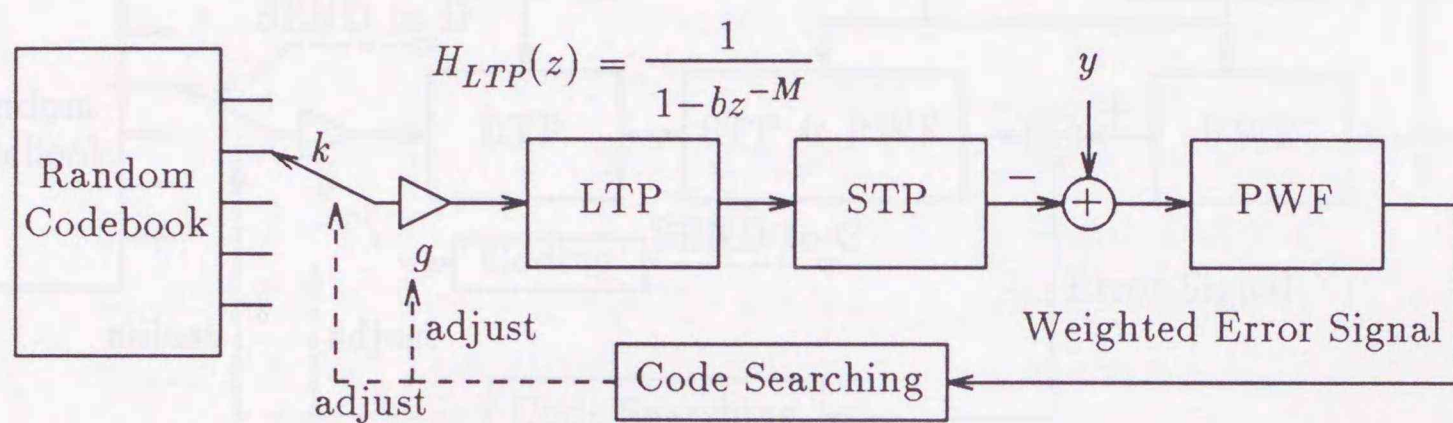
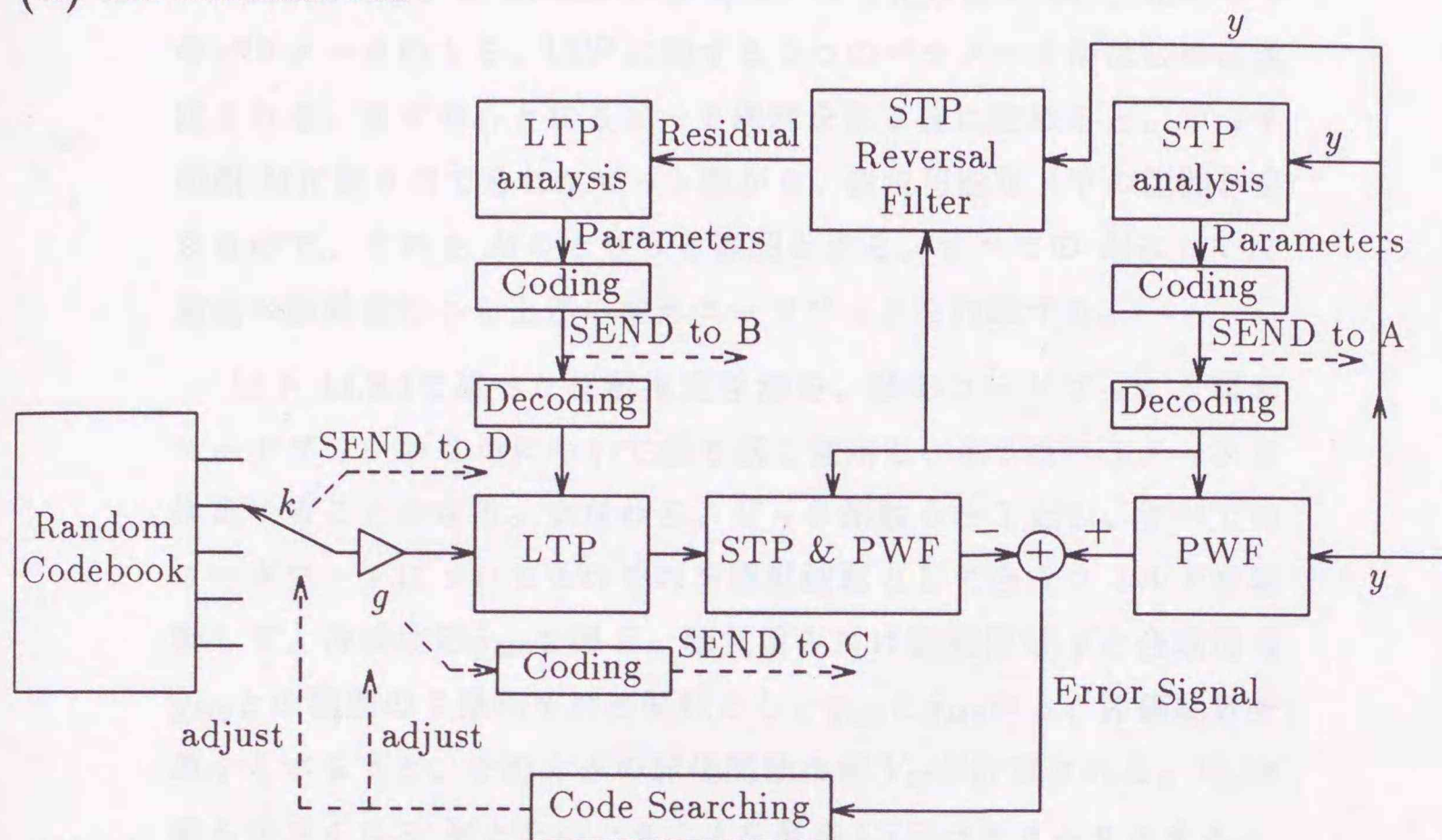
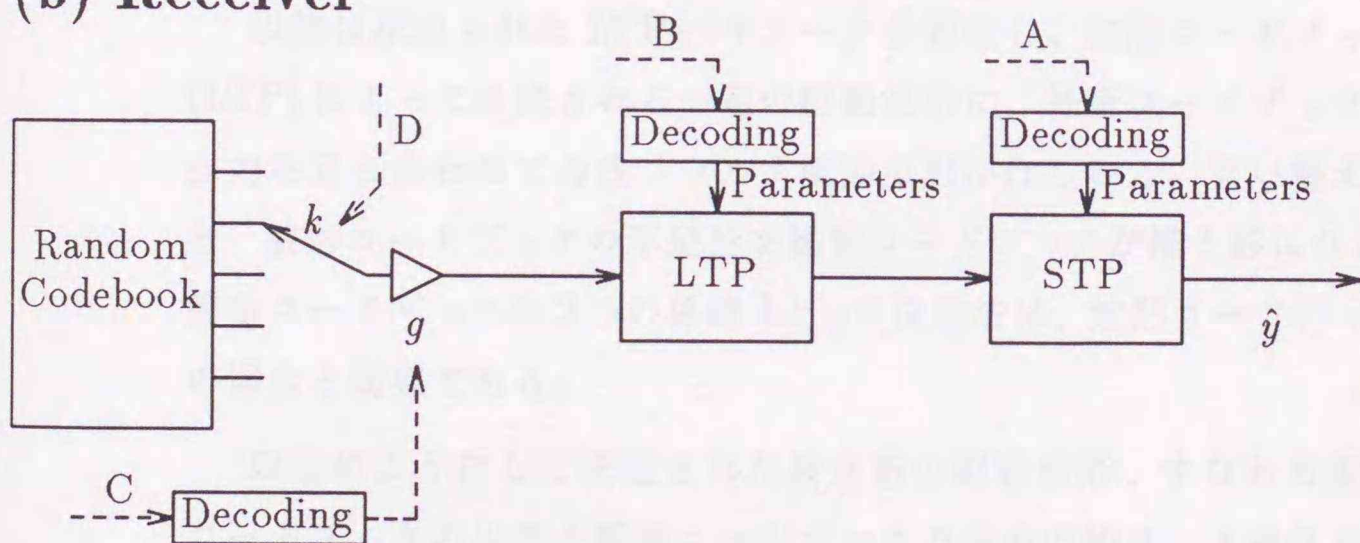


図 11.1: CELP 符号化方式の原理図

(a) Transmitter



(b) Receiver



STP: Short Term Predictor
 PWF: Perceptual Weighting Filter
 LTP: Long Term Predictor

$$H_{LTP}(z) = \frac{1}{1-bz^{-M}}$$

図 11.2: Open loop CELP 符号化方式のブロック図

本論文で用いる Closed loop CELP 符号化方式では、上述の 4 つのパラメータのうち、LTP に関する 2 つのパラメータがはじめに推定される。まず中心となるピッチ周期をある値に定めると、ピッチ周期 M に割り当てられたビット数から、表現可能な上下の範囲が定まるので、それを M のとりうる範囲とする。すべての M について過去の駆動波形から上述の波形コードブックを作成する。

以下 11.3.1 で述べた係数推定手法を、波形コードブックと雑音コードブックの 2 つについて繰り返し適用し、4 つのパラメータを推定することになる。すなわち、ピッチ係数 b を 1 とし、すべてのコードワードについてそれぞれを駆動波形として合成フィルタを駆動して、合成波形 $\hat{y}_{M0}$ を得る。聴覚重み付け観測信号 y' と合成信号 $\hat{y}_{M0}$ との誤差の 2 乗和を評価関数とし、 y_{cur} と $\hat{y}_{M0}$ から、評価関数を最小とする b と、そのときの評価関数の値 V_M が計算される。 V_M が最も小さくなる M とそのときの b を最適 LTP パラメータとする。

以後は求められた LTP パラメータを固定し、波形コードブック (LTP) によって生成される一定の駆動波形に、雑音コードブックの出力を足し合わせて合成フィルタ駆動波形が作られる。言い替えると、波形コードブックの不足分を雑音コードブックが補う形になる。雑音コードブックの 2 つの係数 k と g の決定法は、波形コードブックの場合と同様である。

以上のようにして決定された最適新生駆動波形、すなわち波形コードブックの出力と雑音コードブックの出力の和は、メモリに蓄えられ次の段階での波形コードブックの作成に使用される。

本論文で用いる Closed loop CELP 符号化方式のブロック図を図 11.3 に示す。

実際の CELP 符号器では、STP 係数推定フレーム長、STP 係数更新フレーム長、LTP 係数更新フレーム長 (波形コードワード長)、雑音コードワード長が異なる場合が多いので、それぞれの係数推定段階において矛盾が起こらないように、過去のフレームの影響等を

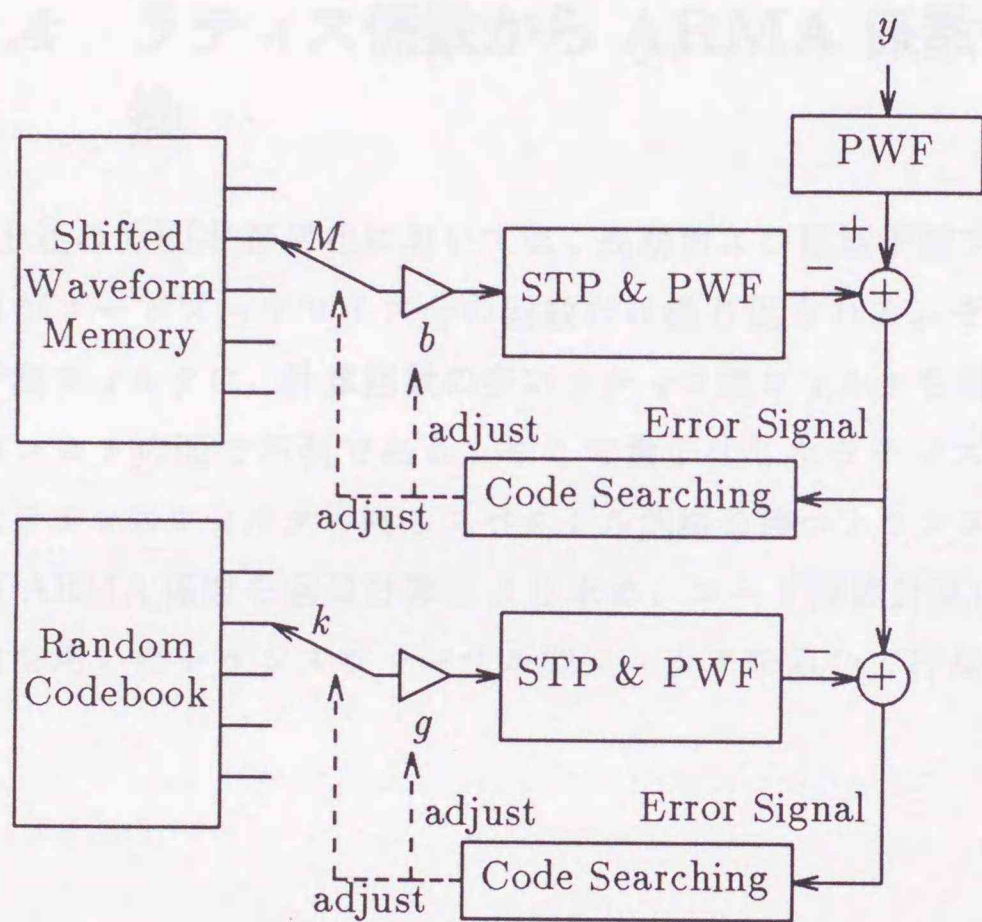


図 11.3: Closed loop CELP 符号化方式のブロック図

適宜除きながら、符号化を行う必要がある。また実際の符号化・合成システムでは、 g, b は量子化されて伝送される。Open loop の場合と同様に本論文では、この g, b の量子化は、最大・最小値を設定する固定量子化で行う。

11.4 ラティス係数から ARMA 係数への変換

上述の CELP 符号化においては、長期および短期予測フィルタの計算がコードブックサイズ分の回数だけ繰り返される。そのため短期予測フィルタに、計算回数の多いラティス型フィルタを用いるのは計算コストの面で不利である。そこで量子化したラティス係数を用いたラティスフィルタと同じスペクトル包絡を持つトランスヴァーサル型 ARMA 係数を回帰計算により求め、コード探索計算は ARMA 係数を用いたトランスヴァーサル型フィルタを用いて行なう。

Chapter 12

Open loop CELP 符号化手法 による実験

本章では、2.2.2(p.20) で述べた規格化 ARMA ラティスフィルタ係数を短期予測フィルタ (STP) 部に導入した Open loop CELP 符号化システムを構築し、その性能について実験により検討する。Open loop CELP 符号化方式については 11.3.1 に述べられている。以下の実験において、STP 係数の推定は規格化 ARMA ラティスフィルタ係数推定アルゴリズムによって直接行い、求められた係数は、10.1(p.111) と同様に PARCOR のフレーム周期に等しい長さで平均される。STP の残差信号は 2.2.5(p.29) で述べた疑似逆フィルタによる方法で計算する。

12.1 都道府県名データを用いた実験

まず、第 10 章で用いたものと同じ 47 都道府県名データ (10kHz サンプリング, 12bit-AD) を用いて符号化・合成実験を行った。

分析次数は、PARCOR 8 次、規格化 ARMA ラティス AR:6 次、MA:1 次であり、双方共に伝送すべき係数の個数は 8 個である。CELP 符号器のビット割当を表 12.1 に示す。

図 12.1, 12.2 に結果の例を示す。

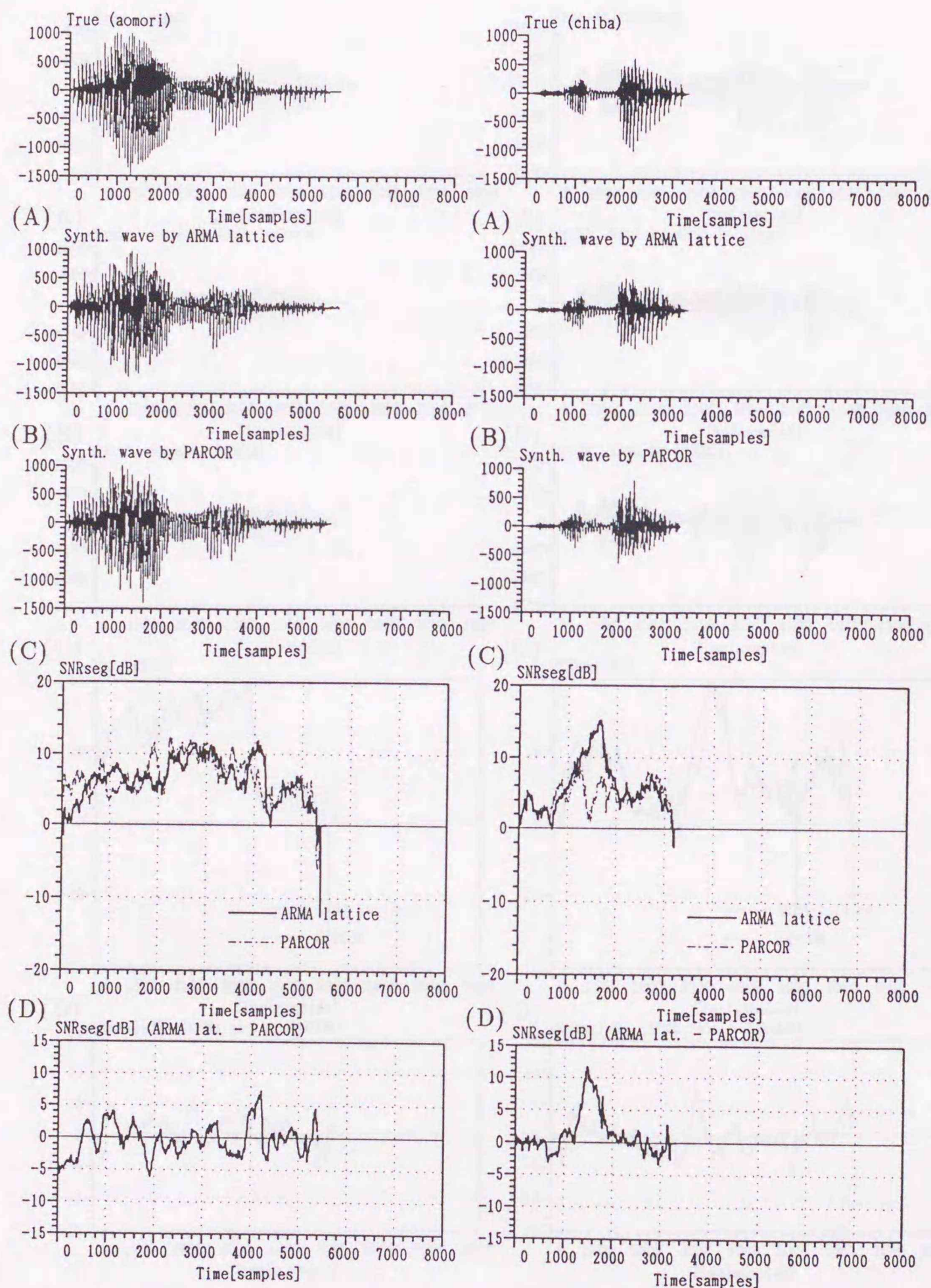


図 12.1: 都道府県名による Open loop CELP 符号化実験結果

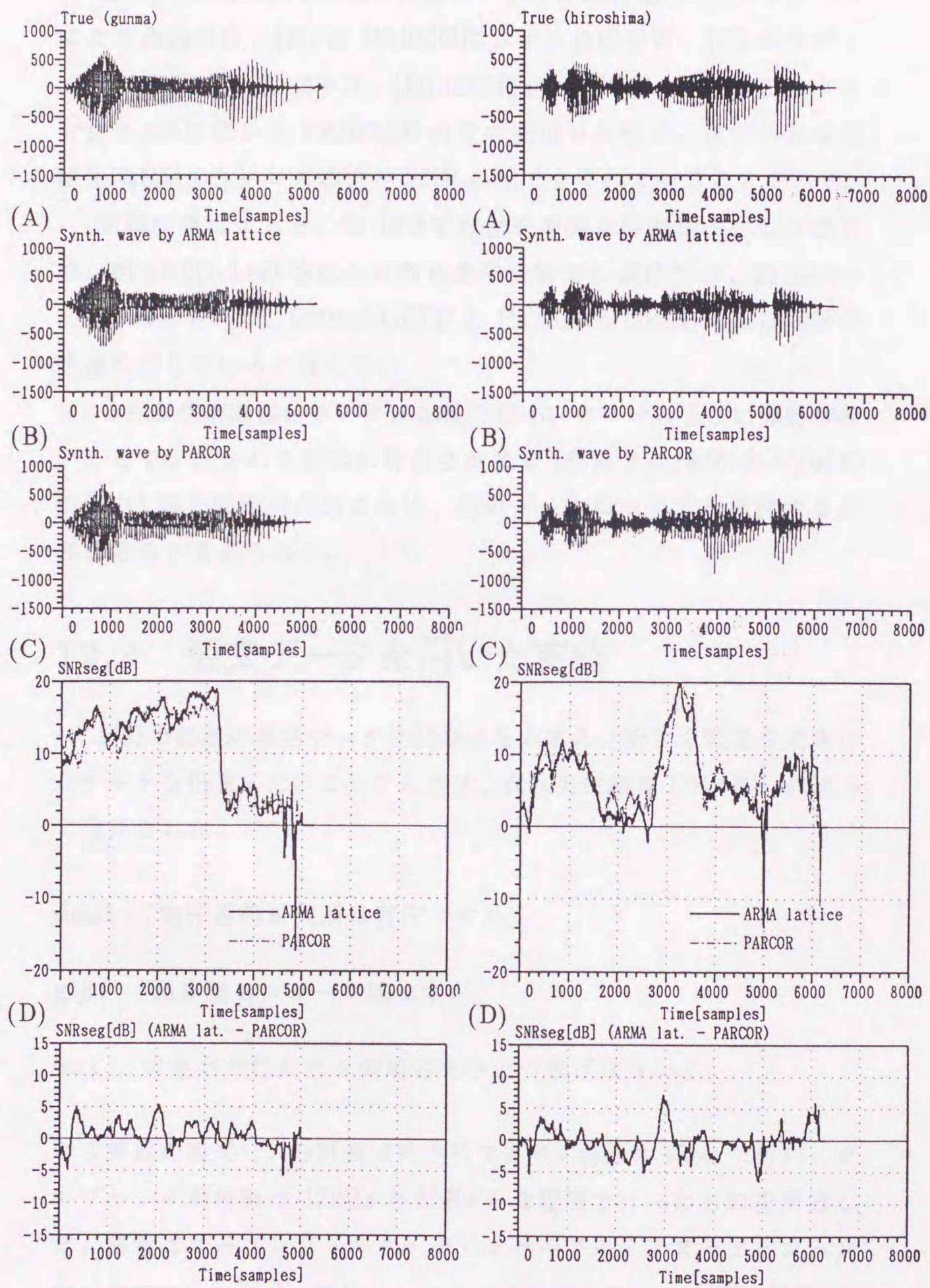


図 12.2: 都道府県名による Open loop CELP 符号化実験結果

図 12.1, 12.2 は図 10.2 等と同様に、(A) は規格化 ARMA ラティスによる合成音声、(B) は PARCOR による合成音声、(C) はセグメンタル SNR の比較グラフ、(D) は規格化 ARMA ラティスのセグメンタル SNR 値から PARCOR のそれを引いた値で、正であれば規格化 ARMA ラティスの方が SNR の面ですぐれていることを示す。

実験結果によると、第 10 章で残差を単純に線形量子化した場合の、図 10.2(p.114) 等にみられる鼻子音等での優位性が、図 12.1~12.2 にはみられず、Lattice-CELP と PARCOR-CELP はほぼ同等の性能を示していると言える。

しかし都道府県名データは比較的短く、パワーの小さい部分が多いことや、含まれる音素に片寄りがある(母音では /a/ が多く /u/ が少ない)等の問題点があるため、他のデータについても検討する必要があると考えられる。

12.2 短文データを用いた実験

前節の都道府県名データの問題点をふまえ、新しく短文を発声したデータを用意した。このデータは、成人男性話者 1 名 (EH) によって発声された、

den1 「電子通信研究所は官庁ですか」

giji1 「議事堂はスポーツ場ですか」

on1 「音楽会の行われる劇場名をひとつ挙げなさい」

の 3 種類の短文で、合計長は約 9 秒である。AD は 12bit で行い、サンプリング周波数は 12kHz と 8kHz の 2 種類で行ったものを用意した。録音はオープンリールオーディオテープにノイズリダクション無しで録音したため、多少のヒスノイズが乗っている。以下 8kHz のものはデータ名を **den1-8** のように記述する。処理の都合上各デー

タは、無音区間を境として 10000 ポイント程度の長さに分割した。データ名と発音内容の対応を表 12.2 に示す。

実験に際し、分析次数や CELP 符号器の定数を様々に変化させて符号化・合成を行った。使用した設定パターンを表 12.3 に示す。

データ **den1** について表 12.3(a)–(k) の条件で実験を行った結果を表 12.4 に示す。11.3.1 で述べたように、Open loop 型の CELP 符号化方式では、STP の残差を計算するため STP 逆フィルタが安定でなければならない。表 12.4 における “–” は MA 部等の影響により音声サンプルの一部で STP 逆フィルタが不安定になり、以後の残差計算が不可能になったものを示す。表 12.4 において PARCOR-CELP との差に注目すると、Lattice-CELP は PARCOR-CELP とほぼ同等か、わずかに劣る結果を示すことがわかる。最も良い結果を示した表 12.4(k) のパターンについて、全ての音声サンプルで実験を行った結果を表 12.5 に示す。さらに 8kHz の表 12.3(l) の条件で全てのサンプルについて実験した結果を表 12.6 に示し、セグメンタル SN 比の比較グラフを図 12.3～12.5 に示す。

表 12.5 および表 12.6 によると、セグメンタル SN 比で Lattice-CELP と PARCOR-CELP はほぼ同等の性能を示している。表 12.6 の結果について簡単な聴取比較実験を行ったが、両者の品質はほぼ同等であった。

Total Rate	STP Parameter	LTP Parameter		Stochastic Codebook	
		gain b	period M	gain g	shape k
8.4 kbps	32/100 (7544 3333)	3/50	7/50	6/50	10/50

(bits/samples)

表 12.1: Open loop CELP のビット割当

Sampling Freq.	Data Name	Contents
12kHz	den1-1	「電子通信」
	den1-2	「研究所は」
	den1-3	「官庁ですか」
	giji1-1	「議事堂は」
	giji1-2	「スポーツ場ですか」
	on1-1	「音楽会の」
	on1-2	「行われるげ」
	on1-3	「き場名を」
8kHz	on1-4	「ひとつ挙げなさい」
	den1-8-1	「電子通信研」
	den1-8-2	「究所は官庁ですか」
	giji1-8-1	「議事堂は」
	giji1-8-2	「スポーツ場ですか」
	on1-8-1	「音楽会の」
	on1-8-2	「行われる劇場名を」
	on1-8-3	「ひとつ挙げなさい」

表 12.2: 短文データ名と発音内容

Type		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
Sampl.Freq.		12											8
STP	Frame Length	360											240
	Frame Shift	120							240				160
	Order PAR.(AR)	8		8	10	12	8				12		
	Order Lat.(MA/AR)	1/6		1/8	1/10	4/12	1/6				1/10		
	Bit Pattern†	A			B			A	C			B	
LTP	Frame Shift	60					120					80	
	Gain Bits	3											
	Period Bits	7											
SCB	Block Length	60											40
	Shape Bits	10		8		10		8					
	Gain Bits	6						5					
PWF	α	1.0	0.98							0.95	0.81	0.95	

Type of Bit Pattern	Bit Pattern
† A	7544 3333 33
B	7654 3333 3322 2222 2
C	6433 2222

表 12.3: 短文による Open loop CELP 符号化の設定表

Type	Algorithm	Averaged SNRseg[dB]			Average [dB]
		den1-1	den1-2	den1-3	
(a)	Lattice	7.0	6.7	4.8	6.2
	PARCOR	7.3	7.7	5.4	6.8
	Difference	-0.3	-1.0	-0.7	-0.6
(b)	Lattice	7.3	6.8	4.9	6.3
	PARCOR	7.3	7.3	5.4	6.7
	Difference	-0.1	-0.5	-0.6	-0.4
(c)	Lattice	6.4	5.5	4.0	5.3
	PARCOR	6.4	6.9	4.7	6.0
	Difference	0.0	-1.4	-0.7	-0.7
(d)	Lattice	6.1	5.7	4.2	5.3
	PARCOR	6.4	6.9	4.7	6.0
	Difference	-0.3	-1.3	-0.6	-0.7
(e)	Lattice	6.8	6.7	4.7	6.1
	PARCOR	7.3	7.8	5.5	6.9
	Difference	-0.6	-1.0	-0.8	-0.8
(f)	Lattice	—	—	—	—
	PARCOR	7.1	7.3	5.3	6.6
	Difference	—	—	—	—
(g)	Lattice	5.8	4.8	2.4	4.3
	PARCOR	4.2	5.7	3.0	4.3
	Difference	1.6	-1.0	-0.6	0.0
(h)	Lattice	3.9	3.8	2.2	3.3
	PARCOR	4.2	4.2	2.9	3.8
	Difference	-0.3	-0.5	-0.6	-0.5
(i)	Lattice	3.4	3.7	2.0	3.0
	PARCOR	4.1	3.9	2.6	3.5
	Difference	-0.6	-0.2	-0.7	-0.5
(j)	Lattice	2.2	3.7	1.7	2.5
	PARCOR	3.3	3.3	2.2	2.9
	Difference	-1.1	0.4	-0.5	-0.4
(k)	Lattice	5.2	4.7	2.7	4.2
	PARCOR	4.0	5.2	2.0	3.7
	Difference	1.2	-0.5	0.6	0.4

表 12.4: 短文による Open loop CELP 符号化実験結果

Type	Algor.	Averaged SNRseg[dB]									Avr. [dB]
		den 1-1	den 1-2	den 1-3	giji 1-1	giji 1-2	on 1-1	on 1-2	on 1-3	on 1-4	
(k)	Lat	5.2	4.7	2.7	4.8	4.4	4.6	4.7	5.6	2.1	4.3
	PAR	4.0	5.2	2.0	6.1	4.0	4.8	6.2	5.4	3.3	4.6
	Diff	1.2	-0.5	0.6	-1.4	0.4	-0.1	-0.5	0.1	-1.2	-0.2

表 12.5: 短文による Open loop CELP 符号化実験結果

Type	Algor.	Averaged SNRseg[dB]							Avr. [dB]
		den 1-8-1	den 1-8-2	giji 1-8-1	giji 1-8-2	on 1-8-1	on 1-8-2	on 1-8-3	
(l)	Lat	—	3.3	4.7	3.2	3.7	6.9	2.7	4.1
	PAR	—	3.6	5.7	3.0	3.3	6.8	3.7	4.4
	Diff	—	-0.3	-1.0	0.3	0.4	0.1	-1.0	-0.2

表 12.6: 短文による Open loop CELP 符号化実験結果

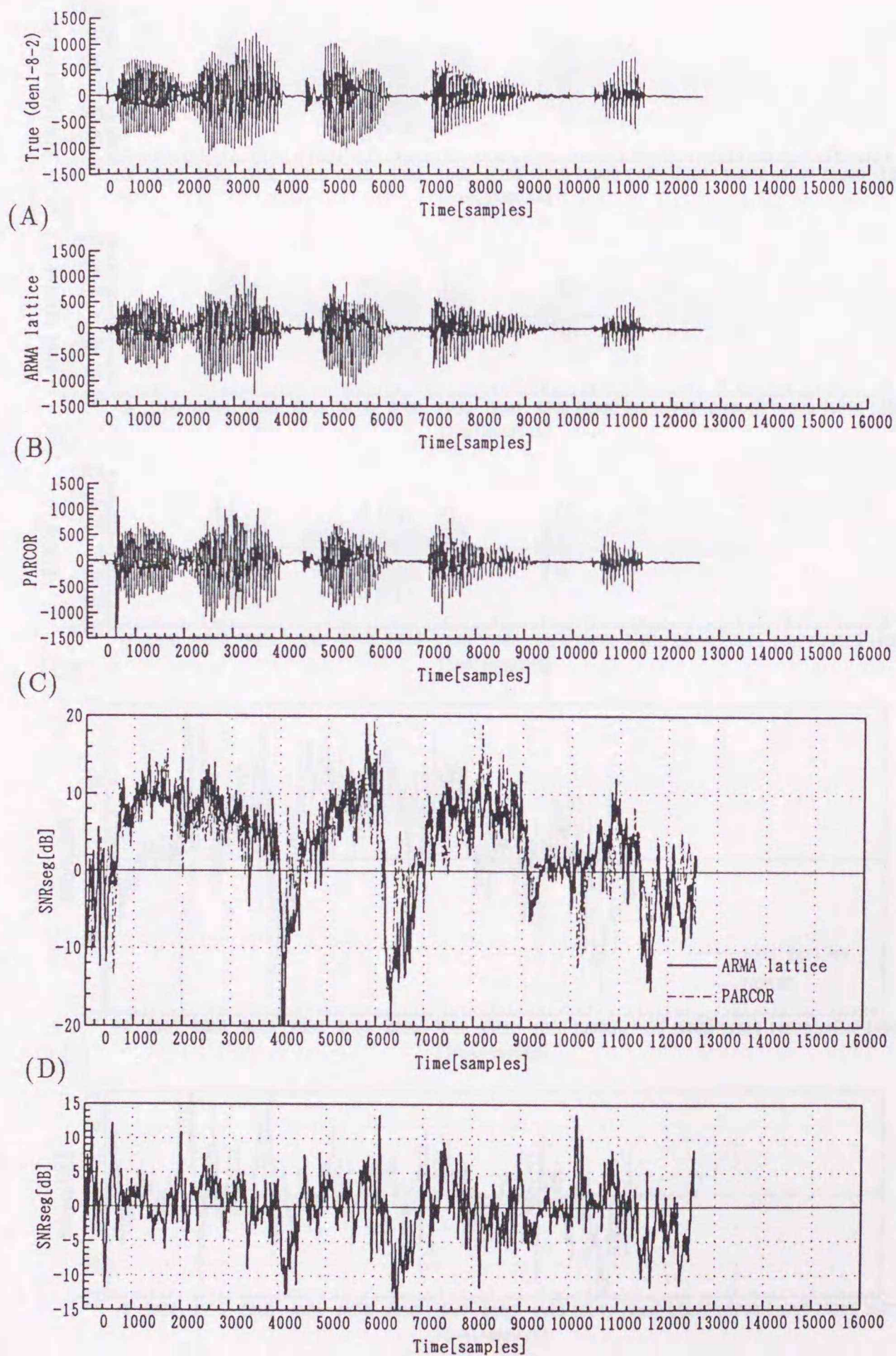


図 12.3: 短文による Open loop CELP 符号化実験結果例

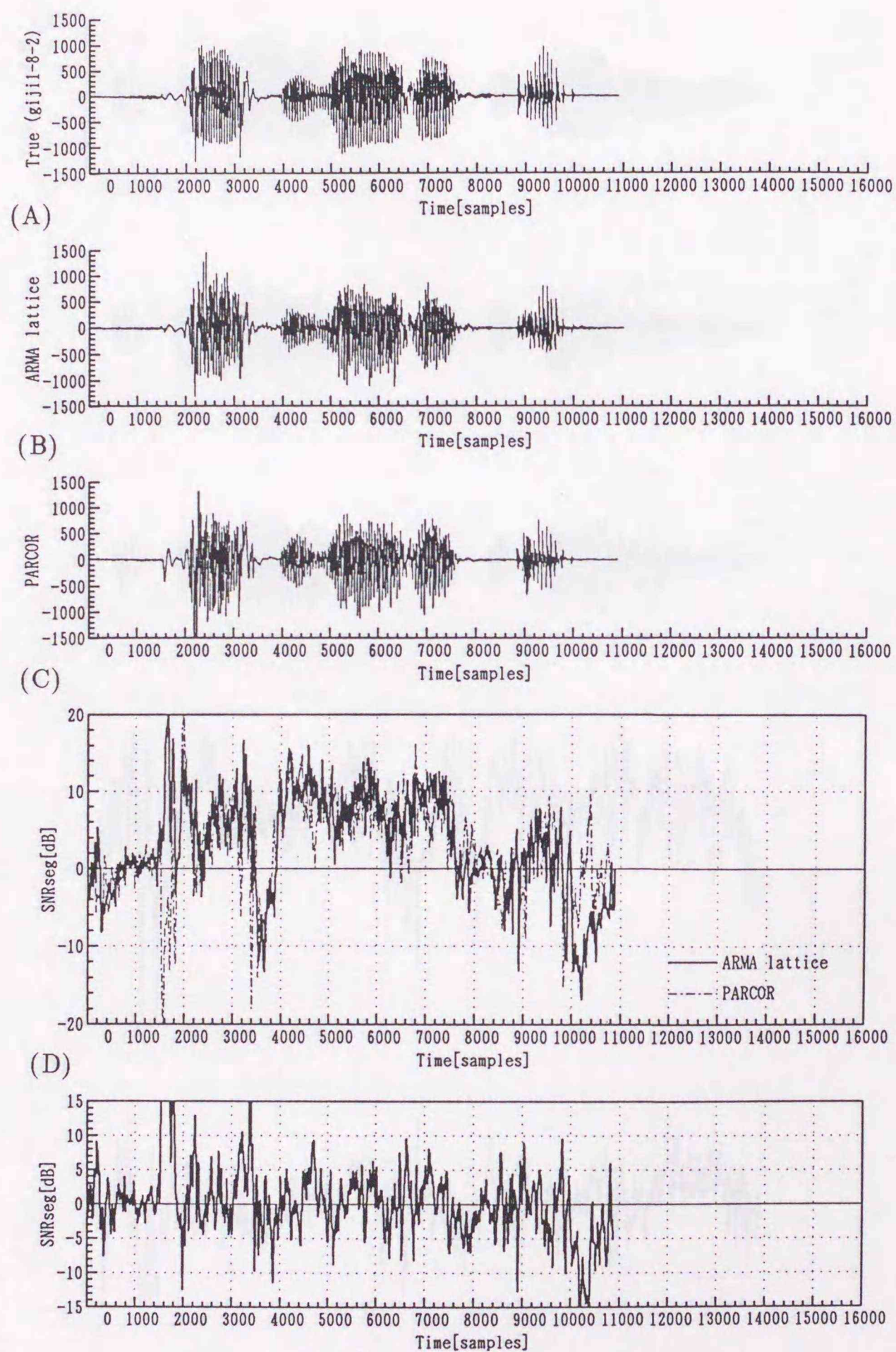


図 12.4: 短文による Open loop CELP 符号化実験結果例

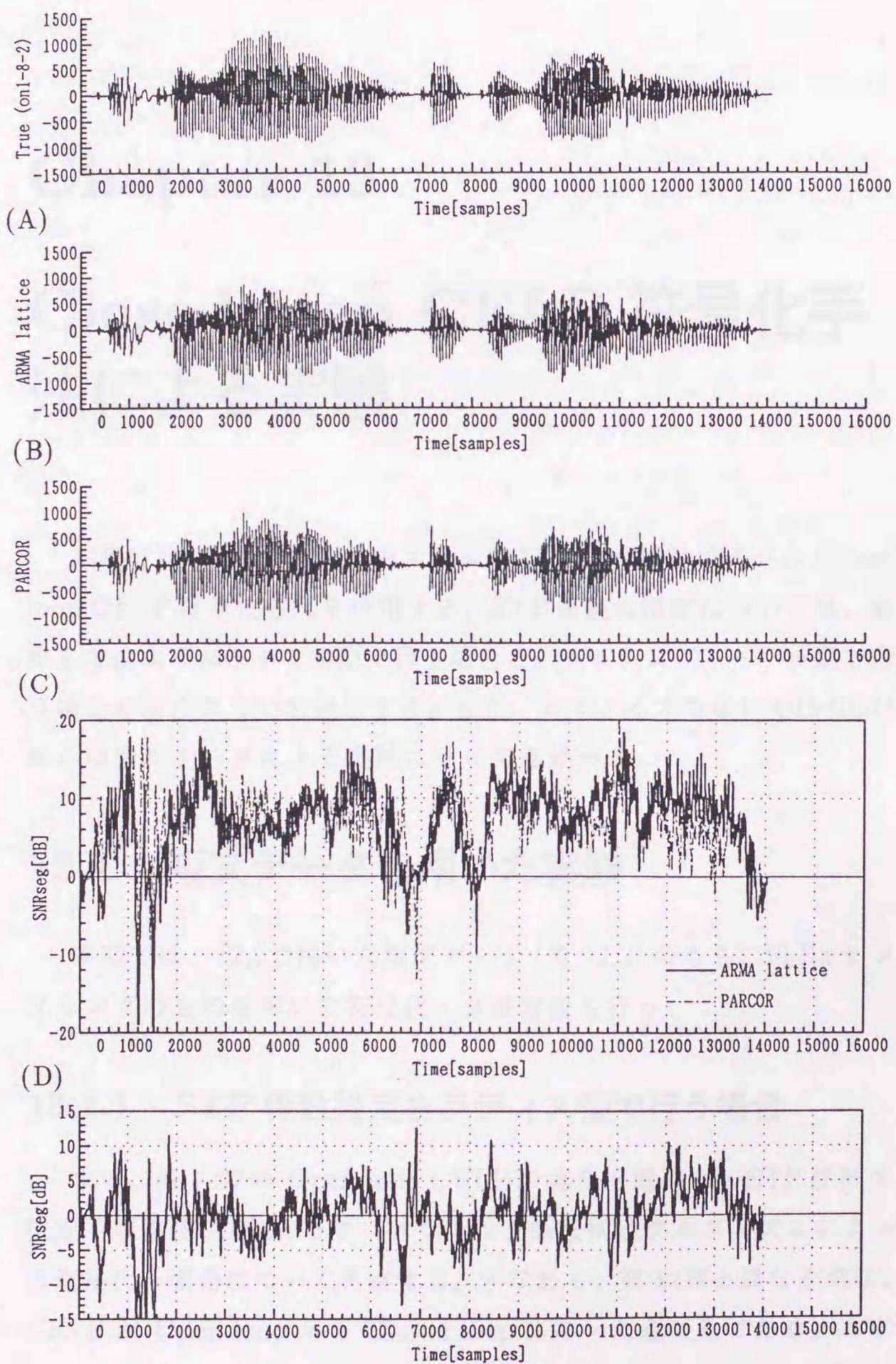


図 12.5: 短文による Open loop CELP 符号化実験結果例

Chapter 13

Closed loop CELP 符号化手法による実験

本章では、CELP 符号化システムに 11.3.2 (p.122) で述べた Closed loop CELP 符号化方式を採用する。STP 係数の推定については、前章までのようにラティス型で行う場合と、トランスヴァーサル型で行う場合の双方について検討する。また、ヒスノイズのない CD-ROM からの音声データによる実験についても述べる。

13.1 短文データを用いた実験

本節では、12.2 で用いた短文データ (表 12.2) のうち、8kHz サンプリングのものを用いて符号化・合成実験を行う。

13.1.1 STP 係数推定をラティス型で行う場合

まず、第 12 章の Open loop CELP の場合と同様に、STP 係数の推定を規格化 ARMA ラティスフィルタ係数推定アルゴリズムによって直接行う場合について考察する。すなわち、第 12 章と異なる点は、CELP が Open loop から Closed loop になった点のみである。本章では CELP が Closed loop なので、11.3.2 で述べたように、STP の残差信号を計算する必要はない。

実験に用いた CELP 符号器のビット割当を表 13.1 に示す。この割当は、第 12 章の Open loop CELP の表 12.3(1) の割当とほぼ同様の設定であり、次数が ARMA から AR になった点と、PWF の係数の値が表 12.3(1) の割当と異なっている。

実験結果を表 13.2 に示し、実験結果の例を図 13.1~13.3 に示す。例示したサンプルは第 12 章の Open loop CELP の実験で例示したサンプルと同じサンプルである。

第 12 章の Open loop 場合と比較して、CELP が Closed loop になったため、PARCOR, 規格化 ARMA ラティス共にセグメンタル S/N 比は向上しているが、PARCOR と規格化 ARMA ラティスの差をとった場合、Open loop の場合に比べて規格化 ARMA ラティスがより劣っている。このことは規格化 ARMA ラティスの推定精度と関係がある可能性があるため、より高精度なトランスヴァーサル型係数推定アルゴリズムを以下で導入する。

13.1.2 STP 係数推定をトランスヴァーサル型で行う場合

短期予測フィルタの構成

STP 係数は重み付き RLS 法 (p.12, 2.1.2) や MIS 法 (p.14, 2.1.3) を用いて、トランスヴァーサル係数の形で推定し、伝送時には規格化 ARMA ラティス係数を用いる短期予測フィルタ部の構成を図 13.4 に示す。

音声波形はまず重み付き RLS または MIS で分析され、1 サンプル点につき 1 組のトランスヴァーサル型 ARMA 係数が求められる。重み付き RLS の場合には、5.1(p.48) で述べた理由により、スペクトル選択を行わなくてはならない。これは本節では簡単に、予測誤差の短期平均値と長期平均値からしきい値操作を行うことによって、誤差が小さく安定している高精度な ARMA 係数を選出する。得られた ARMA 係数からモデルのインパルス応答波形を求め、インパ

Sampl.Freq.		8
STP	Frame Length	240
	Frame Shift	160
	Order PAR.(AR)	12
	Order Lat.(MA/AR)	0/12
	Bit Pattern	7654 3333 3322 2222 2
LTP	Frame Shift	80
	Gain Bits	3
	Period Bits	7
SCB	Block Length	40
	Shape Bits	8
	Gain Bits	5
PWF	α	0.92

表 13.1: Closed loop CELP 符号器の設定

Algor.	Averaged SNRseg[dB]							Avr. [dB]
	den 1-8-1	den 1-8-2	giji 1-8-1	giji 1-8-2	on 1-8-1	on 1-8-2	on 1-8-3	
Lat	5.4	4.3	6.4	4.4	4.7	7.9	2.4	5.1
PAR	5.7	5.0	6.6	4.7	5.0	7.5	3.1	5.4
Diff	-0.2	-0.7	-0.2	-0.3	-0.3	0.4	-0.7	-0.3

表 13.2: 係数推定を規格化 ARMA ラティスで行った Closed loop CELP の実験結果

CHAPTER 13. CLOSED LOOP CELP 符号化手法による実験 143

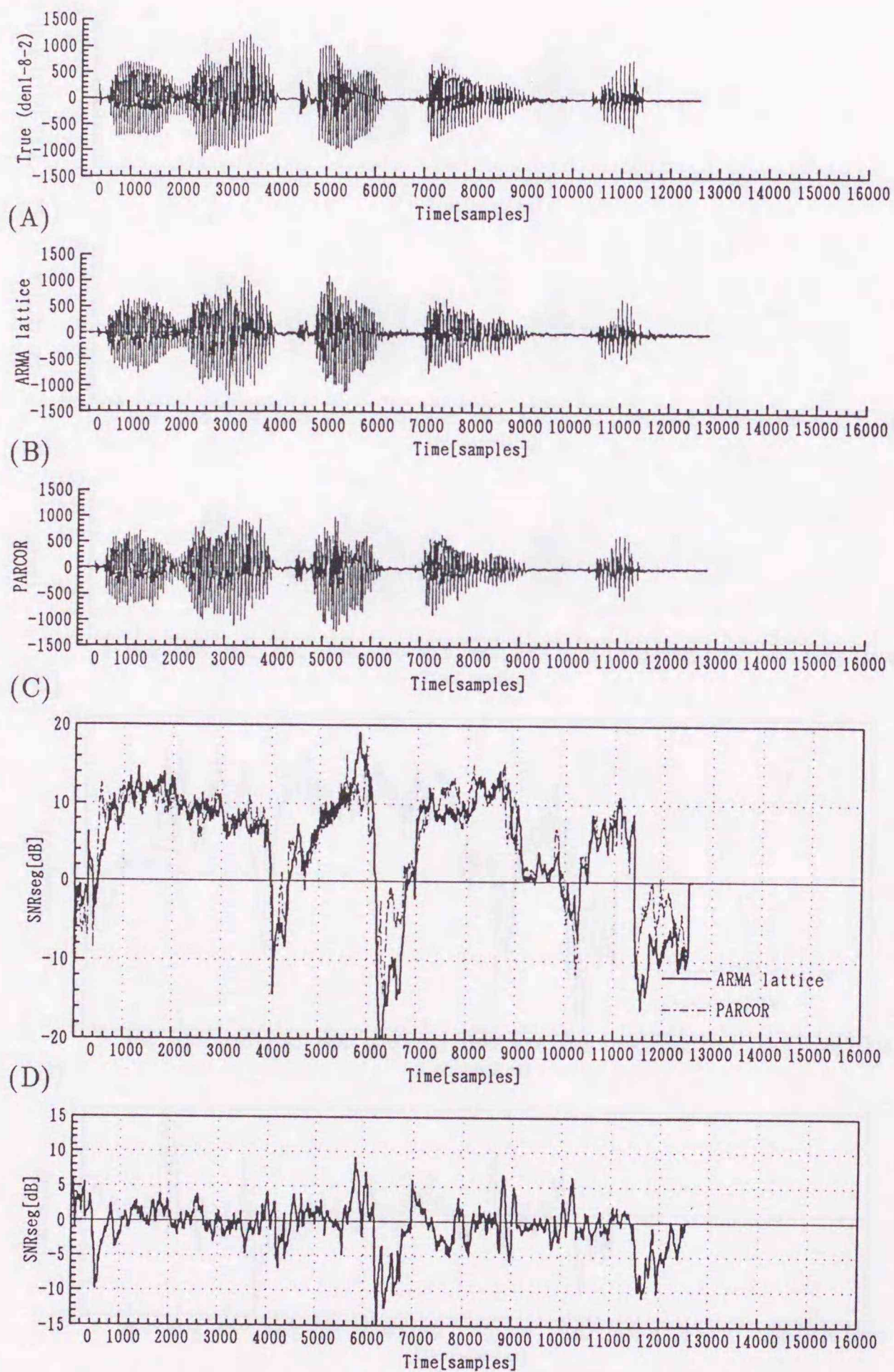


図 13.1: 係数推定を規格化 ARMA ラティスで行った Closed loop CELP の実験結果例

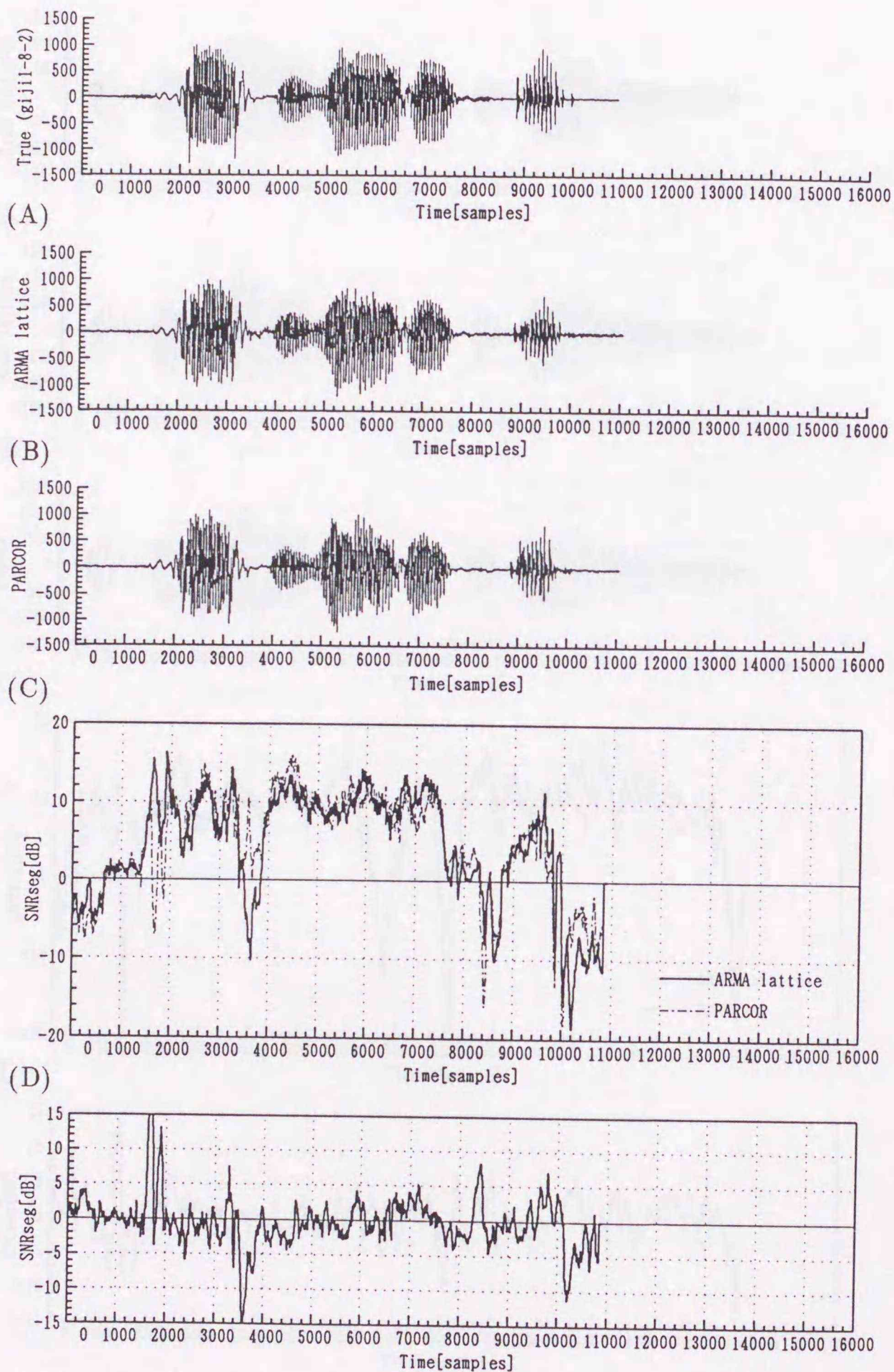


図 13.2: 係数推定を規格化 ARMA ラティスで行った Closed loop CELP の実験結果例

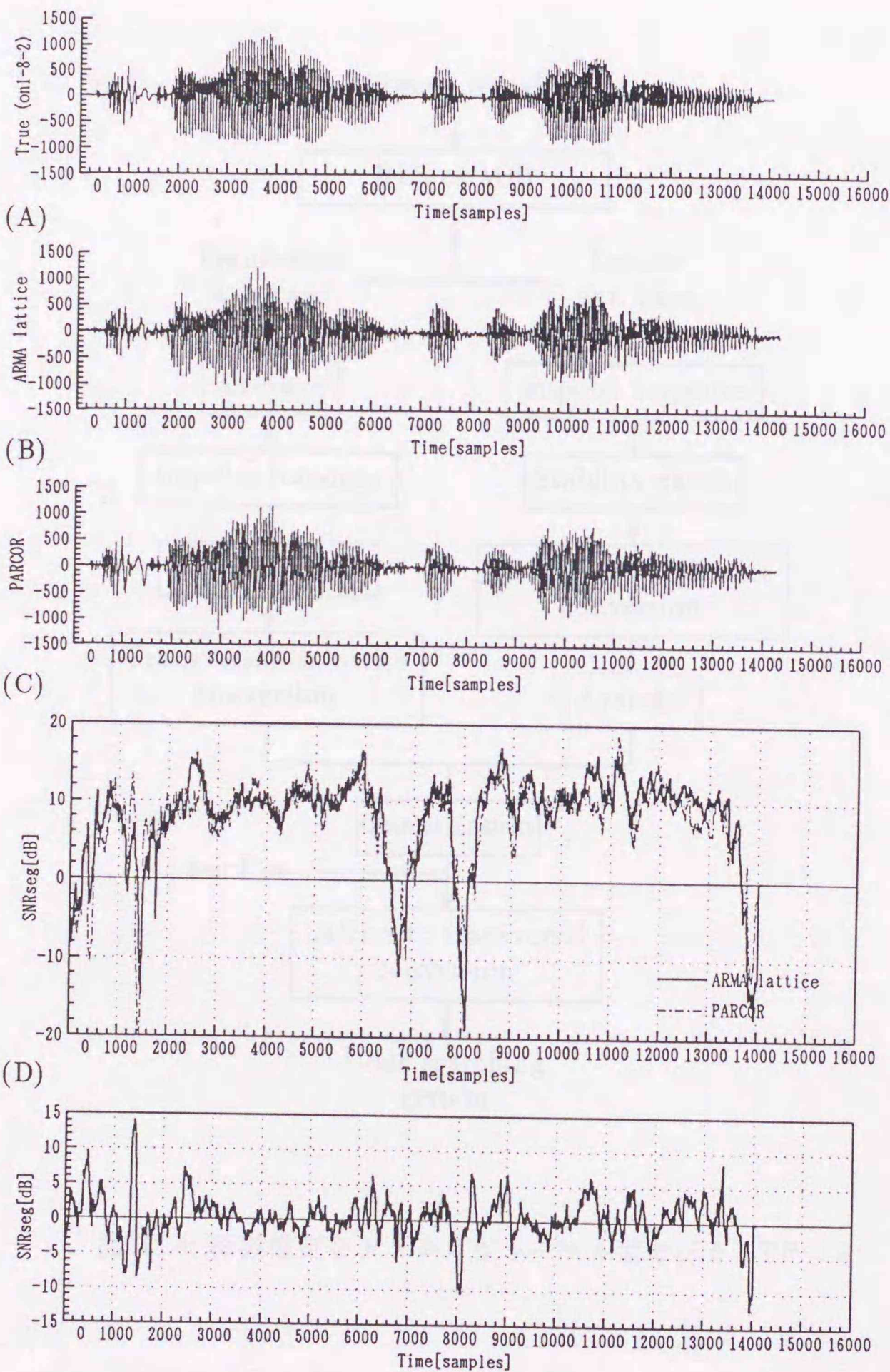


図 13.3: 係数推定を規格化 ARMA ラティスで行った Closed loop CELP の実験結果例

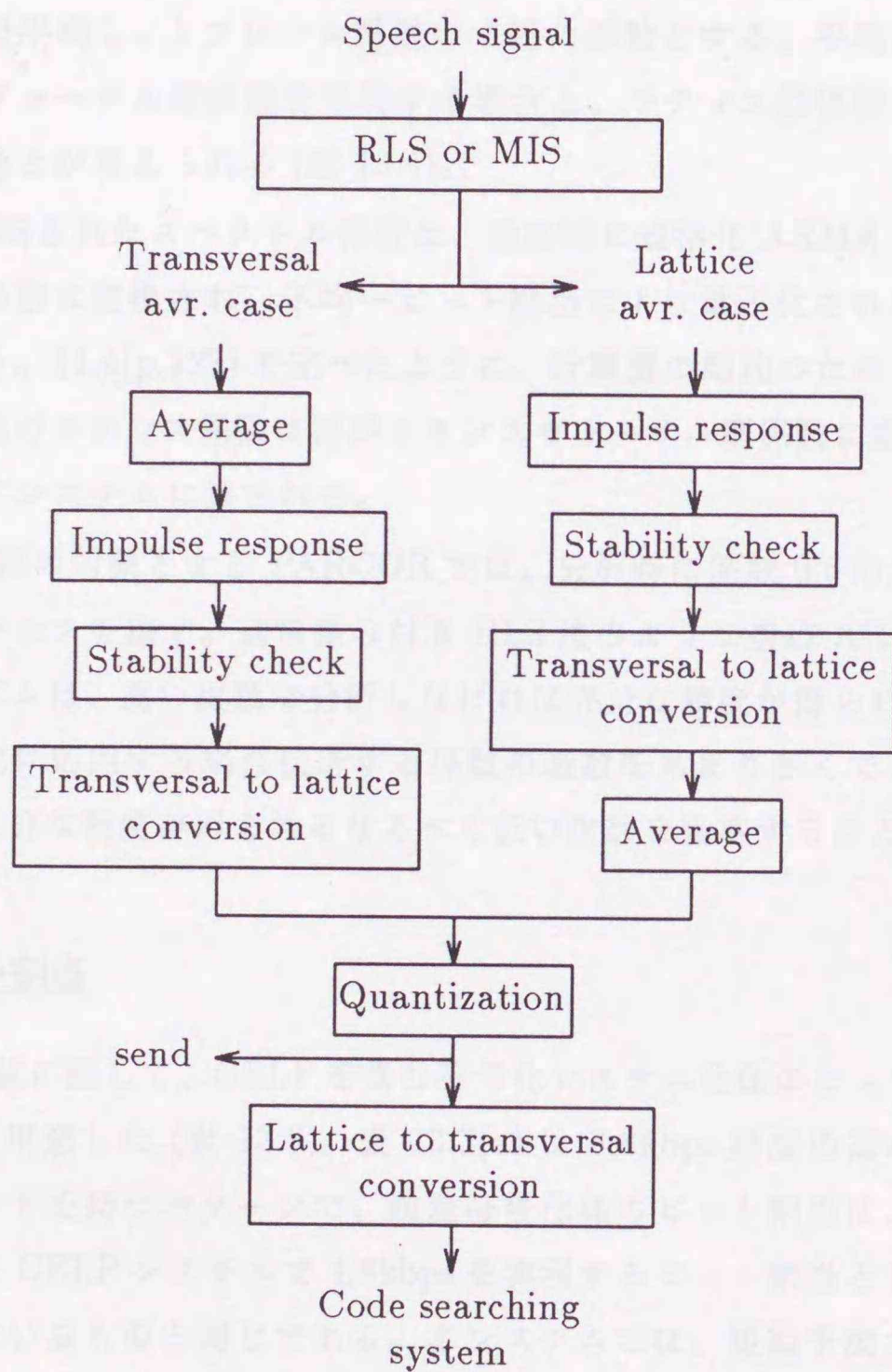


図 13.4: 係数推定をトランスヴァーサル型で行う STP

ルス応答が収束する安定な係数について等価な規格化 ARMA ラティス係数を求める。

求められたスペクトル係数は、LPC のフレーム周期に相当する区間で平均し、1 フレーム周期に 1 組の係数とする。平均は、トランスヴァーサル型係数を平均する場合と、ラティス型係数を平均する場合とが考えられる (図 13.4)。

平均されたスペクトル係数は、最終的に規格化 ARMA ラティス係数の形に変換され、不均一ビット割当により量子化された後伝送される。11.4(p.127) で述べたように、計算量の節約のため平均・量子化後のラティス係数は再びトランスヴァーサル型係数に変換され、CELP システムに渡される。

比較の対象とする PARCOR では、分析時に係数 0.6 のプリ・エンファシスを施す。通常重み付き RLS 法のような適応 ARMA アルゴリズムは、高い次数で分析しなければ十分な精度が得られないが、符号化に应用する場合伝送する係数の個数をあまり多くできないため、十分な精度が得られるなるべく低い次数を設定することとする。

ビット割当

実験に際して、CELP を含む符号化システム全体のビット割当を 2 種類用意した (表 13.3)。表 13.3(a) は 5.8kbps 程度の高めのビットレートを持つパターンで、残差符号化部のビット割当は、通常の多くの CELP システムで 4.8kbps を実現するビット割当として用いられているものと同じである。本システムでは、短期予測フィルタ部の次数と割当ビット数が通常の CELP システムより高めになっているため、総ビットレートは 5.8kbps 程度となっている。表 13.3(b) は 4.0kbps 程度の低めのビットレートを持つパターンである。このパターンでは、短期予測フィルタの推定アルゴリズムの違いがより明確になることを期待して、残差符号化部のビット割当を少なくし CELP の能力を大きく落としてある。

Type		(a)	(b)
Total Rate		5.8 ~ 6.1 kbps	4.0 ~ 4.3 kbps
STP Parameter		44 ~ 50/160 (7654 3333 3322 [222])	
LTP Parameter	gain b	3/80	3/80
	period M	7/80	5/80
Stochastic Codebook	gain g	5/40	4/80
	shape k	8/40	6/80

(bit/samples)

表 13.3: CELP 符号化システムのビット割当

平均・量子化後の推定スペクトルの例を図 13.5 に示す。図 13.5 において、用いた適応分析アルゴリズムは ARMA モデルによる MIS 法である。比較のため、PARCOR による量子化・デエンファシス後のスペクトルも示されている。文献⁽¹⁶⁾において、適応信号処理アルゴリズムで求められたモデルの係数を短区間で平均しても、一括型の LPC より高い時間分解能を得られることが示されているが、図 13.5 からそのことがわかる。

検討 1 — MA 部の有無について

短期予測フィルタを表現するモデルが、AR モデル (12 次) の場合と ARMA モデル (AR: 12 次、MA: 2 次) の場合とについて合成音声の品質を比較した。モデルの推定は MIS、重み付き RLS の双方で行った。図 13.6~13.9 に実験結果の例を示す。図 13.6~13.9 の左列は 10msec の範囲で求めたセグメンタル S/N 比の時間変化を 1 点ごとに示したグラフである。左列のそれぞれのグラフの右隣のグラフは PARCOR のセグメンタル S/N 比 (「(A)」のグラフ) との差を表し、値が正であれば対応する適応スペクトル推定手法が PARCOR よりも優れていることを意味する。図 13.6~13.9 において、(B) と (C) の比較が、MIS における AR と ARMA の比較に相当し、(D) と (E) の比較が、重み付き RLS における AR と ARMA の比較に相当する。

図 13.6~13.9 によると、MIS および重み付き RLS の双方共に、音声の多くの部分において、AR モデルを用いたものが ARMA モデルを用いたものよりも高いセグメンタル S/N 比を示している。また簡単な聴取テストにおいても、概して AR モデルによるものの方が ARMA モデルによるものより良好な音質を示した。

以上の結果より、実用的なビットレートを持つ CELP 符号化システムにおける ARMA モデルの導入は、現時点では困難があることが示された。MA 部の導入による音質劣化の理由の検討が必要である。

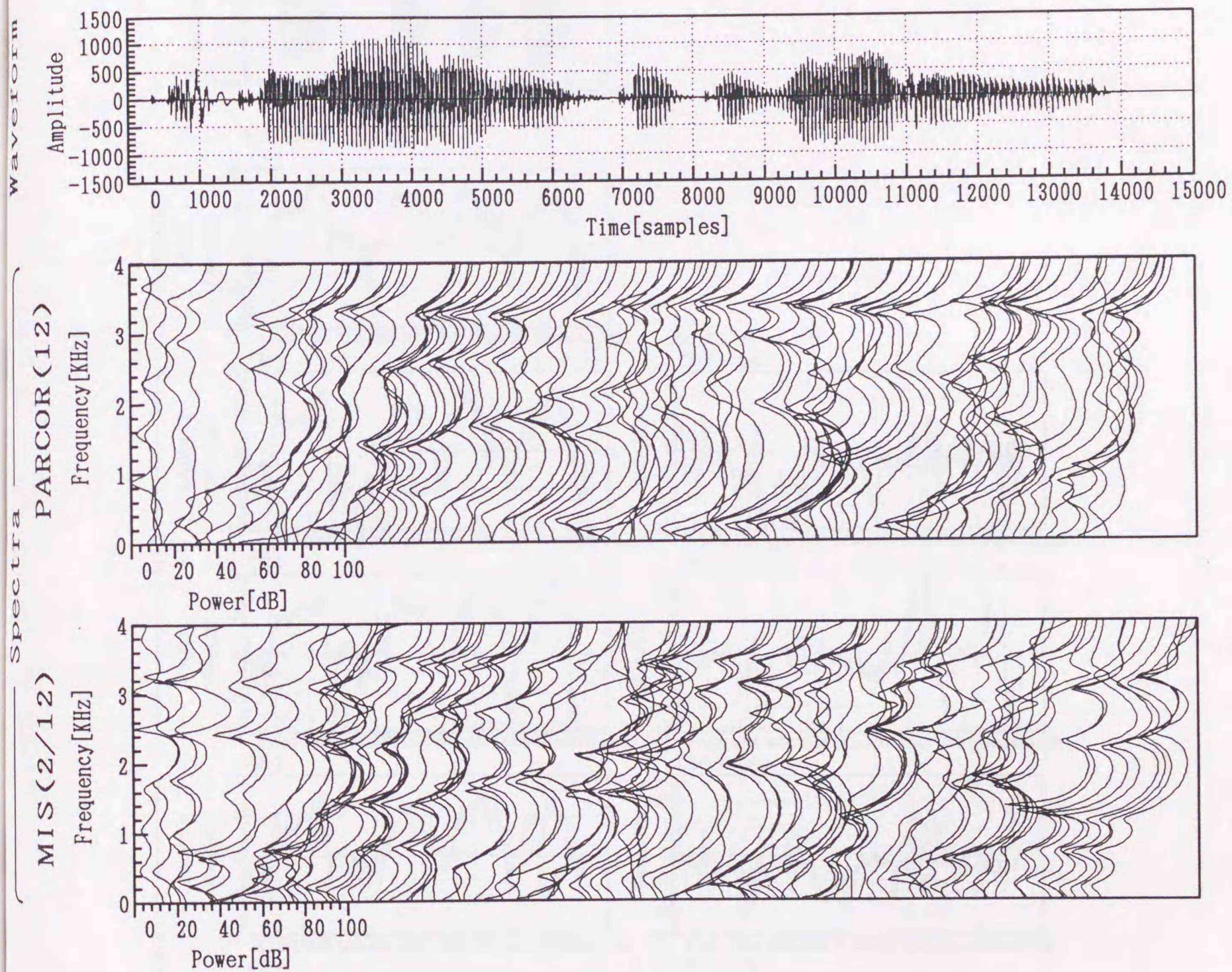


図 13.5: 係数推定をトランスヴァーサル型で行う STP のスペクトル

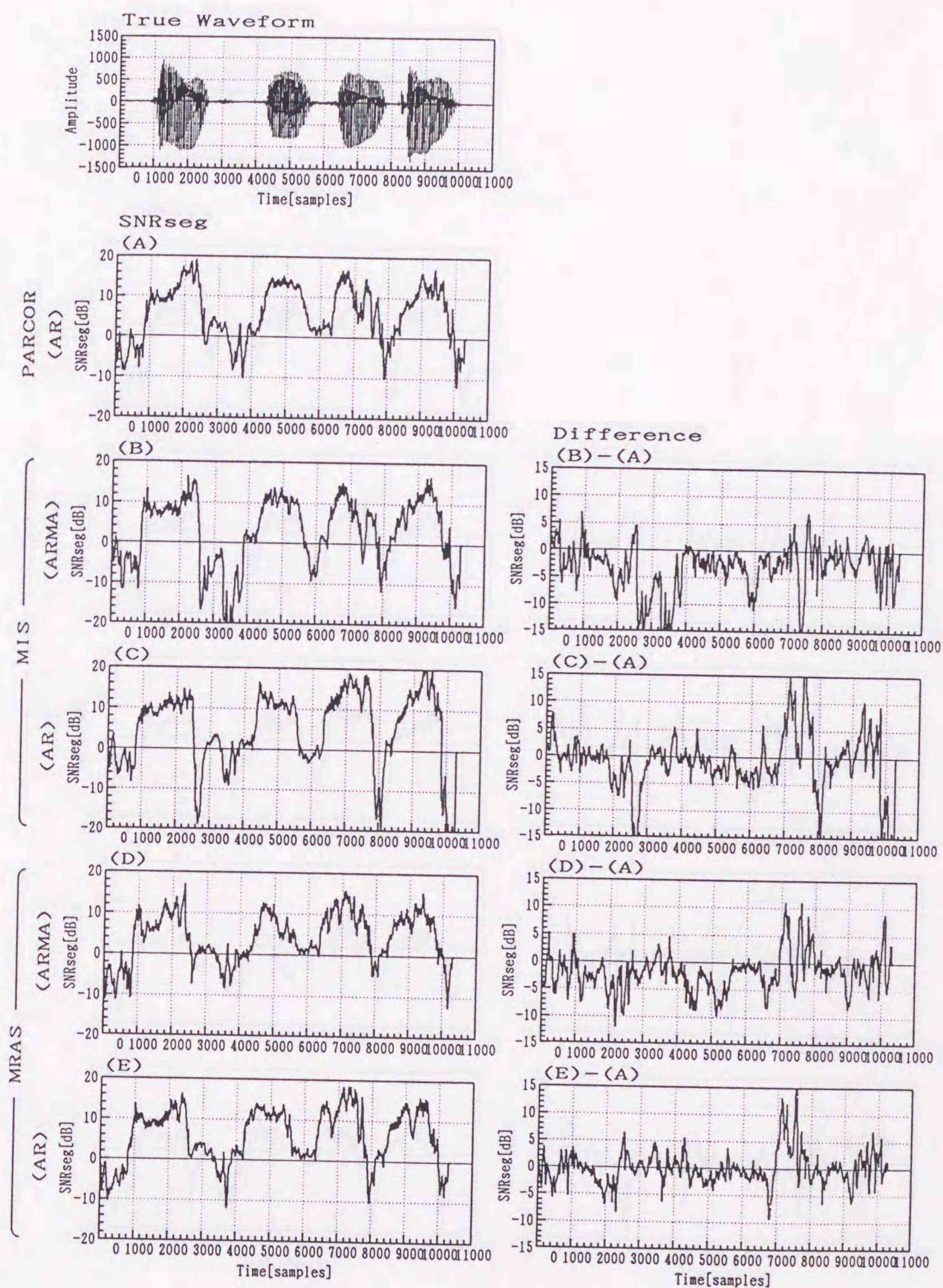


図 13.6: Closed loop CELP 比較実験結果 (検討 1~3)

ビットレート高 (表 13.3(a))

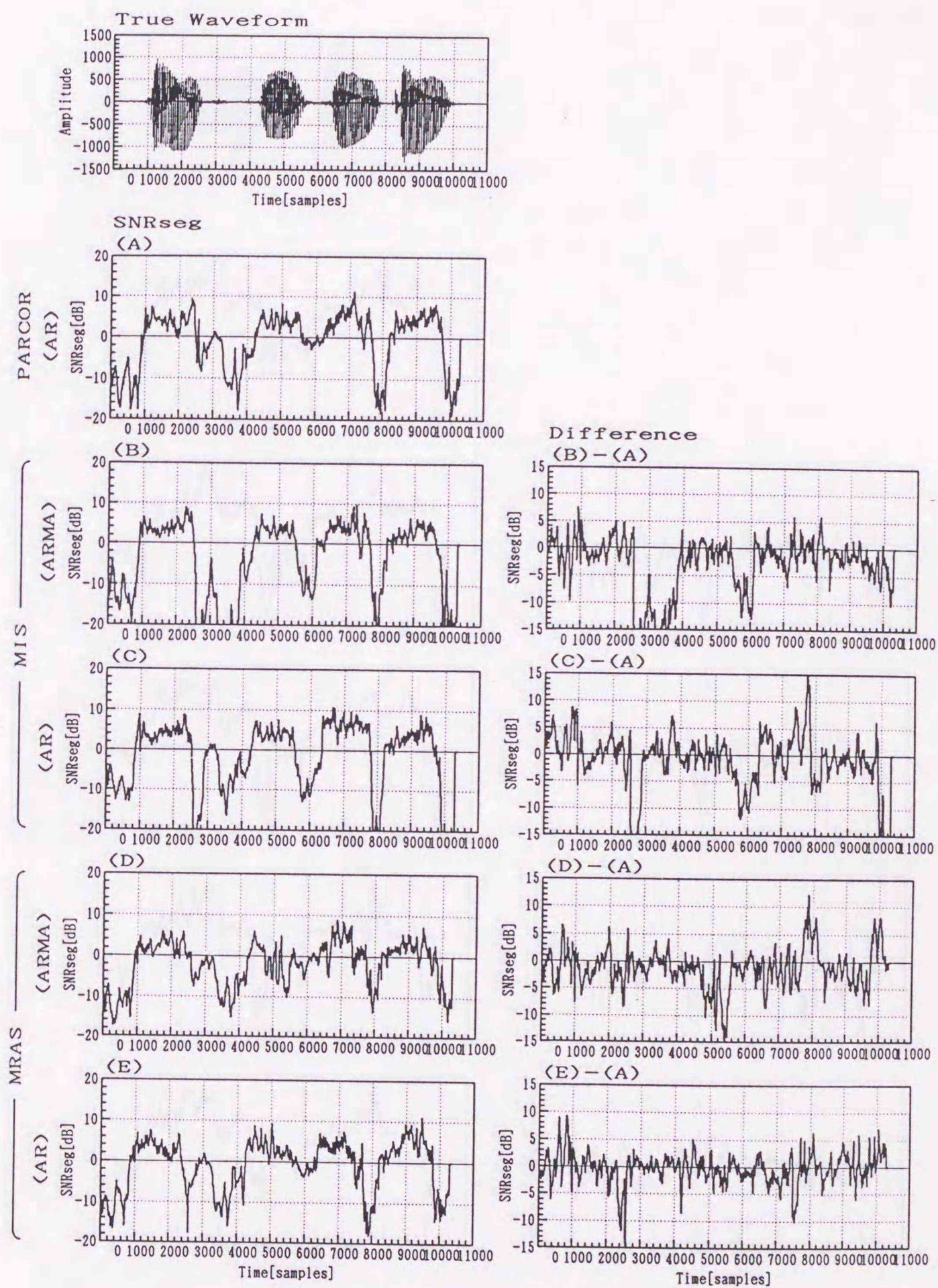


図 13.7: Closed loop CELP 比較実験結果 (検討 1~3)

ビットレート低 (表 13.3(b))

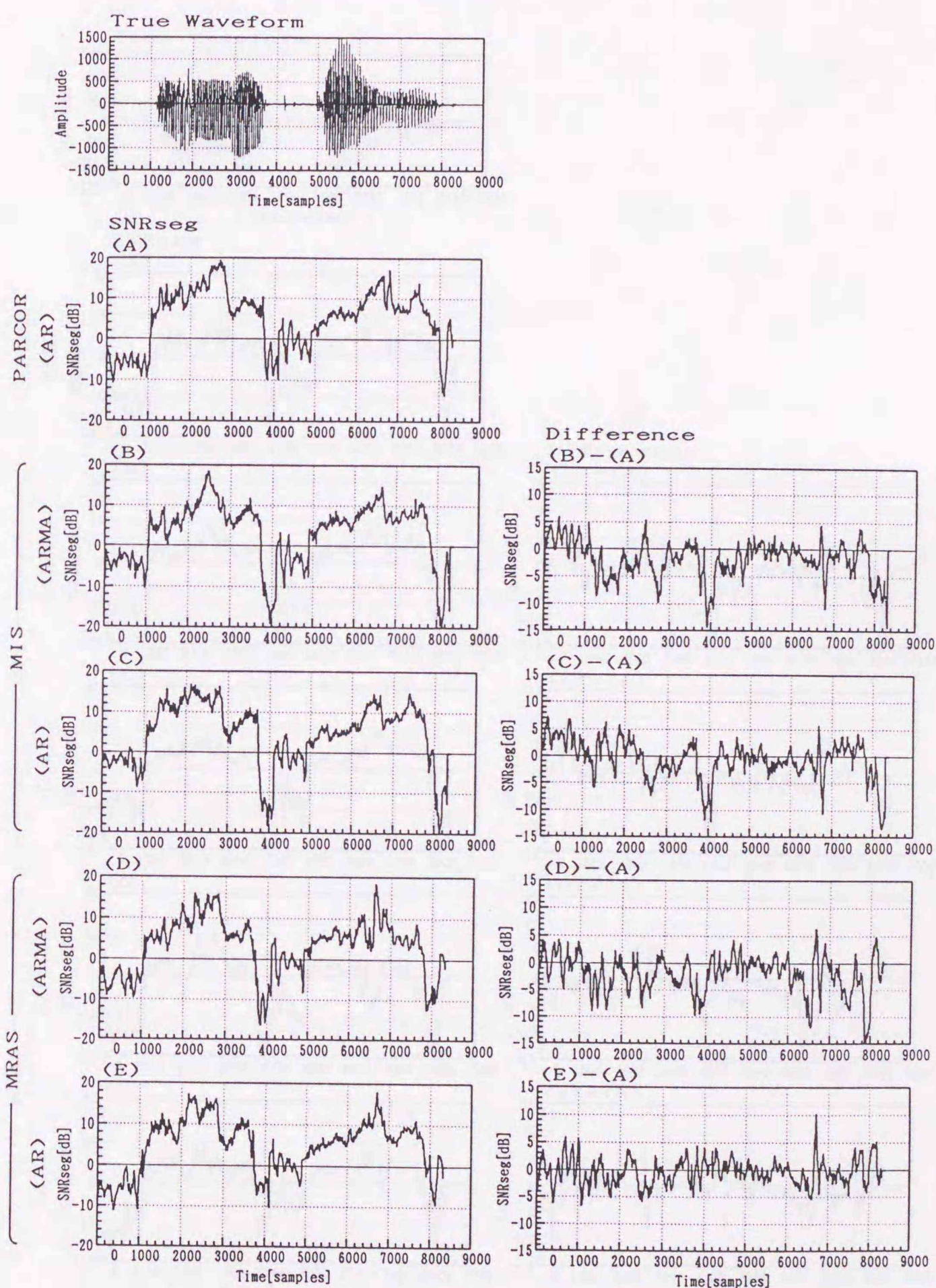


図 13.8: Closed loop CELP 比較実験結果 (検討 1~3)

ビットレート高 (表 13.3(a))

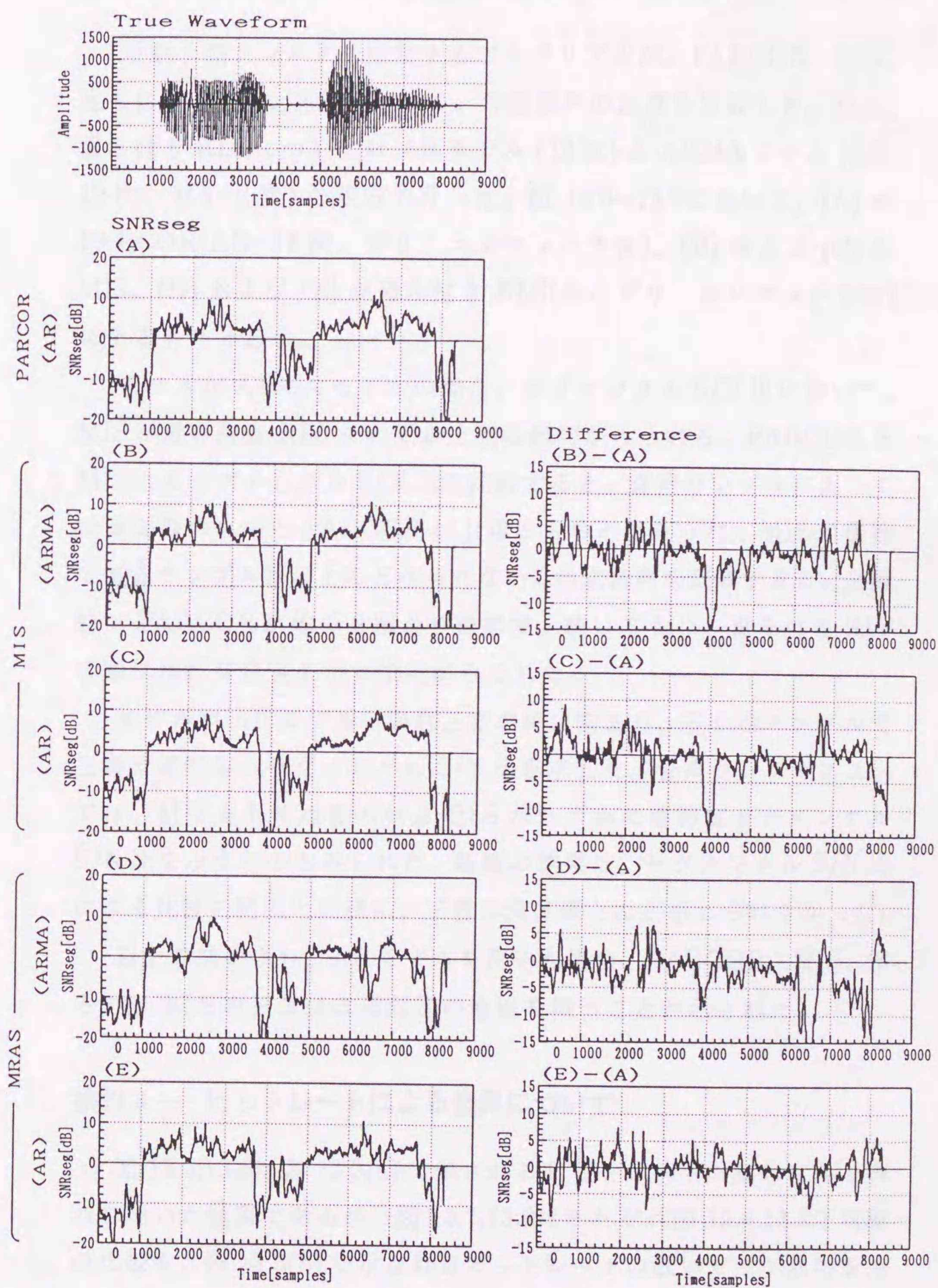


図 13.9: Closed loop CELP 比較実験結果 (検討 1~3)

ビットレート低 (表 13.3(b))

検討2 — 短期予測フィルタ推定アルゴリズムについて

短期予測フィルタを推定するアルゴリズムが、PARCOR、MIS、重み付き RLS の場合について、合成音声の品質を比較した。MIS、重み付き RLS については AR モデル (12 次) と ARMA モデル (AR: 12 次、MA: 2 次) の双方で行った。図 13.6~13.9 において、(A) が PARCOR (AR: 12 次、プリ・エンファシス有)、(B) および (C) が MIS、(D) および (E) が重み付き RLS (共にプリ・エンファシス無) によるものである。

モデルが ARMA モデルの場合、セグメンタル S/N 比において、MIS が重み付き RLS をわずかに上回る傾向がみられる。PARCOR と MIS とをセグメンタル S/N 比で比較すると、音声サンプルによってほぼ同等のものと PARCOR が上回るものとがみられ、MIS が優勢であるサンプルはほとんどみられない。合成音声を聴取すると、全般的に PARCOR が他の 2 種より優勢で、続いて MIS、重み付き RLS の順に良好な音質を示す傾向がみられた。

モデルが AR モデルの場合三者の差は縮まり、多くのサンプルで三者はほぼ同等のセグメンタル S/N 比を示した。音声サンプルによっては、MIS もしくは重み付き RLS がわずかに優勢なセグメンタル S/N 比を示すものもみられた。聴取の結果は、セグメンタル S/N 比による比較の結果と同様に、三者の差はほとんど感じられなかった。

以上の結果より、AR モデルを用いた場合、PARCOR、MIS、重み付き RLS の三者はほぼ同等の性能を持つことが示された。

検討3 — ビットレートによる差異について

図 13.6, 13.8 は表 13.3(a) で示されるビットレートの高いビット割当を用いた結果であるが、図 13.7, 13.9 はそれぞれ図 13.6, 13.8 と同様の比較を、表 13.3(b) で示されるビットレートの低いビット割当を用いて行ったものである。表 13.3(b) のビット割当は CELP の能力を落としてスペクトル推定アルゴリズムの差を強調することを狙った

ものであるが、図 13.7, 13.9 によると合成音声の音質の差の傾向は図 13.6, 13.8 とほぼ同様であり、聴取の結果も前節で述べた結果とほぼ同様であった。すなわち、表 13.3(b) の低ビットレートにおいても、3 つのアルゴリズムはほぼ同等の品質の合成音声を生成することが示された。

検討 4 — 聴覚重み付けフィルタの有無について

前節までの結果をふまえ、比較の対象を PARCOR と AR モデルの MIS に絞り、一般的に用いられている聴覚重み付けフィルタを導入して性能の変化を検討した。結果の例を図 13.10, 13.11 に示す。図 13.10, 13.11 において左列は聴覚重み付けフィルタ無、右列が有 ($\alpha = 0.92$) の結果で、最下行のグラフはセグメンタル S/N 比の差を表し、正ならば MIS、負ならば PARCOR が優れていることを意味する。なお聴覚重み付けフィルタは聴覚的な効果をもたらすものなので、セグメンタル S/N 比による比較は参考としてのみ意味を持つことになる。図 13.10, 13.11 によると、セグメンタル S/N 比においては、聴覚重み付けフィルタの導入により PARCOR と MIS との差が MIS に有利にごくわずかに変化する傾向があるが、全体的には目立った変化はない。聴取によると、聴覚重み付けを行った場合 MIS は PARCOR より明瞭性が増す傾向があり、本節の実験の中で最も良好な音質の合成音声を得られた。

以上の結果より、本節で検討した範囲では、適応信号処理を導入した CELP 符号化システムのうち、聴覚重み付けフィルタを導入した AR モデルに基づく MIS システムが最もよい性能を示すことがあきらかになった。従来型の PARCOR-CELP と比較すると、セグメンタル S/N 比ではほぼ同等、聴覚的には明瞭度でわずかに勝ることが示された。

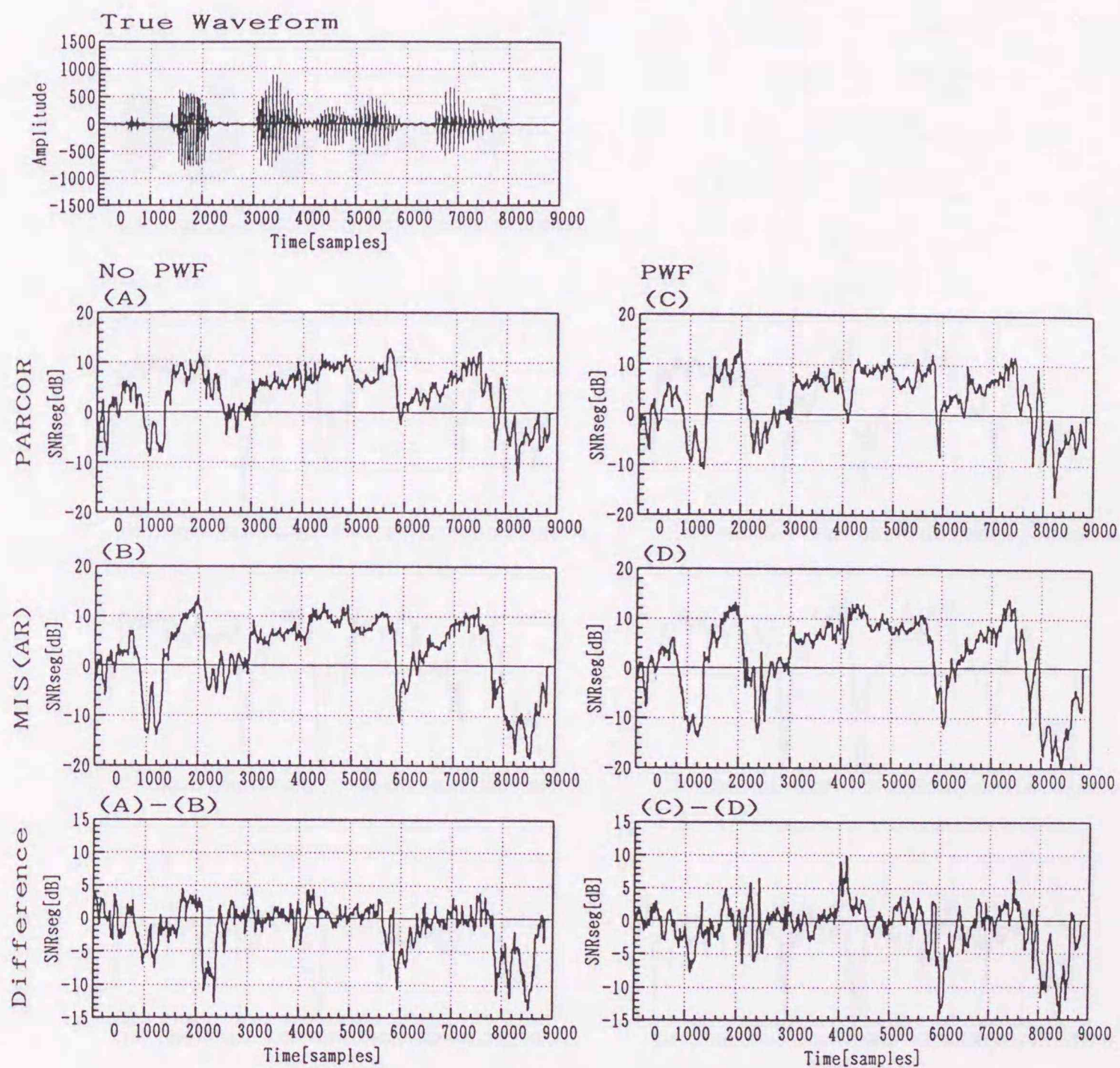


図 13.10: Closed loop CELP 比較実験結果 (検討 4)

ビットレート高 (表 13.3(a))

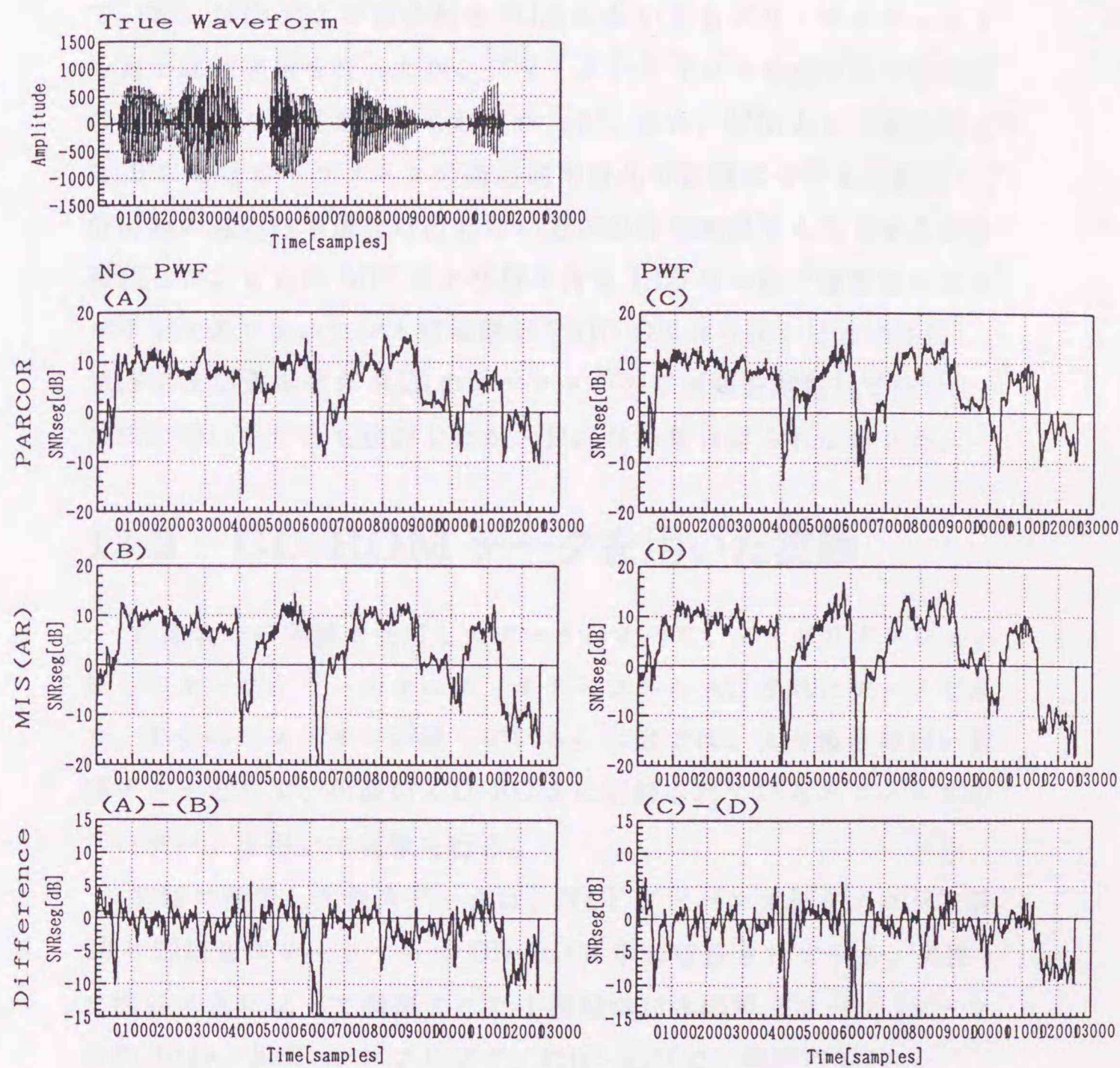


図 13.11: Closed loop CELP 比較実験結果 (検討 4)

ビットレート高 (表 13.3(a))

検討5 — プリ・エンファシス、適応的平均等

前項までの実験では PARCOR のみプリ・エンファシスが施されていた。MIS および重み付き RLS においてもプリ・エンファシスを施し比較実験を行ったが、プリ・エンファシスを施さない場合との間にほとんど差が認められなかった。また、MIS および重み付き RLS におけるパラメータの適応的平滑化の影響についても検討し、分析時の適応的平滑化は行わない方が良好な結果をもたらすことを確認した。さらに MIS および重み付き RLS は一般に推定されたスペクトルのフォルマント帯域幅が PARCOR より狭いことに注目し、MIS および重み付き RLS のスペクトルの帯域幅を調整して符号化・合成に用いることも検討したが、良好な結果は得られなかった。

13.2 CD-ROM データを用いた実験

前節までの実験で使ったデータはすべて、ノイズリダクション無しのオープンリールオーディオテープから AD されたデータであり、多少のヒスノイズが乗っている。本節では、より条件の良い音声データとして、市販の CD-ROM に記録されているヒスノイズのないデータを用いて実験を行う。

実験に使用した音声データは、NTT アドバンステクノロジー社製 20 ケ国語音声データベース CD-ROM 中に収録されている、男性・女性各 4 名によって発声された 1 種類の日本語短文データ (データ長約 10 秒、8kHz サンプリング、12bit-AD) で、発声内容は、

「彼は以前から／科学技術の進歩と人間の勇気が／
はるかな宇宙への旅を／
可能にしたのだと考えていました」

である。「／」は短いポーズを表す。

本節では CD-ROM に記録された信号のうち、44.1kHz でサンプリングされたオーディオ録音信号を用いる。このオーディオ信号は

通常の CD プレイヤーで再生し、AD 変換用のフィルタ・アンプに直結して、8kHz でサンプリングする。CD-ROM に記録されている原信号の S/N 比は 61dB である。

実験における STP 分析次数設定および CELP を含む符号化システム全体のビット割当を表 13.4 に示す。残差符号化部のビット割当は、通常の多くの CELP システムで 4.8kbps を実現するビット割当として用いられているものと同じである。13.1.2 で述べたように、本システムでは、短期予測フィルタ部の次数と割当ビット数が通常の CELP システムより高めになっているため、総ビットレートは 5.9kbps 程度となっている。なお 11.3.2 で述べたように、LTP と雑音コードブックの利得係数の量子化は、固定量子化で行う。

上述のシステムを用いて音声符号化・合成実験を行い、従来の PARCOR を用いたシステムと比較を行った。

重み付き RLS 法による実験

重み付き RLS 法を用いた CELP 符号化システムについて、種々の分析条件のもとで PARCOR による CELP 符号化システムとの比較実験を行った。注目した点は、(1) AR モデルを用いた場合と、ARMA モデルを用いた場合との差異、(2) STP 係数の平均をトランスヴァーサル型係数を用いて行った場合と、ラティス型係数を用いて行った場合の差異、(3) AR モデルの場合について、聴覚重み付けフィルタの有無による差異であり、計 6 種類の条件で実験を行った。

実験結果を図 13.12~13.14、表 13.5(a)~(f) に示す。図 13.12~13.14 において、最上段のグラフは原音声波形、(A) は重み付き RLS を用いたシステムで合成された音声波形について、10msec で求められたセグメンタル S/N 比を 10msec 毎に表示したものであり、(B) は PARCOR のシステムについての (A) と同様なセグメンタル S/N 比グラフである。最下段のグラフは各時刻において (A) から (B) を引いた値を表し、この値が正であれば重み付き RLS 法が PARCOR より

Model		AR	ARMA
Bit rate		5.9kbps	6.1kbps
STP order		12	AR:12,MA:2
STP		46/160 (7654 3333 3322 2)	50/160 (7654 3333 3322 222)
LTP	gain b	3/80	
	period M	7/80	
Stochastic	gain g	5/40	
Codebook	shape k	8/40	

(bits/samples)

表 13.4: CELP 符号化システムのビット割当

優れていることを示す。図 13.12~13.14 に示した実験結果例は、AR モデル、トランスヴァーサル型で係数平均、聴覚重み付けフィルタ有の場合 (表 13.5(b)) の話者 3f のデータによるものである。表 13.5 に示した値は、休止による無音区間を除いた部分について求めた、平均セグメンタル S/N 比である。無音区間の検出は、原音声のパワーについてしきい値を下回る区間が一定時間 (200msec) 以上続いたものを無音区間とした。

表 13.5(a)~(f) によると、AR モデルの場合、聴覚重み付けフィルタの有無、話者、スペクトル平均方法にかかわらず常に重み付け RLS 法による合成音声は PARCOR よりも高い平均セグメンタル S/N 比を示している。最も良い結果をあたえている表 13.5(e) では、話者別で 0.53~1.45dB、全話者平均で 0.86dB の平均セグメンタル S/N 比の改善がみられる。また図 13.12~13.14 に例示した S/N 比変化グラフを調べると、音声パワーの変化する点で重み付け RLS 法が高い S/N 比を示す傾向があることがわかり、フレーム周期で平均してもなお適応信号処理アルゴリズムが音声信号の時間変化に対して高精度なスペクトルを推定し、それが音声符号化において有効であることを示している。簡単な聴取実験においても、PARCOR による合成音声にときおり聴かれるパルス性の雑音が重み付け RLS 法による合成音声にはほとんど存在せず、重み付け RLS 法による合成音声は高い品質を示した。

しかし ARMA モデルの場合 (表 13.5(c),(f))、重み付け RLS 法は PARCOR よりも低い S/N を示している。聴取しても重み付け RLS 法による音声は変化の多いバックグラウンドノイズが多く存在し、PARCOR よりも品質が劣っている。このことについては後の節で考察する。

記号	モデル	平均 手法	PWF	手法	話者								平均
					1m	2m	3m	4m	1f	2f	3f	4f	
-	AR	-	無	PAR	5.48	2.24	4.60	5.28	3.06	5.24	6.31	2.44	4.33
-	AR	-	有	PAR	4.62	1.31	3.62	4.64	2.28	4.20	4.83	0.78	3.28
(a)	AR	T	無	RLS	5.61	2.82	4.71	5.68	3.30	5.46	7.08	2.83	4.69
				Diff	0.14	0.58	0.12	0.41	0.25	0.22	0.77	0.39	0.36
(b)	AR	T	有	RLS	5.11	2.21	4.07	5.34	2.62	4.81	6.01	2.10	4.03
				Diff	0.49	0.90	0.45	0.70	0.34	0.62	1.19	1.31	0.75
(c)	ARMA	T	無	RLS	4.19	0.86	3.42	3.57	1.70	4.37	5.41	1.16	3.08
				Diff	-1.29	-1.38	-1.18	-1.70	-1.36	-0.87	-0.90	-1.28	-1.25
(d)	AR	L	無	RLS	5.57	2.75	4.82	5.45	3.43	5.34	6.93	2.91	4.65
				Diff	0.09	0.50	0.22	0.17	0.38	0.10	0.62	0.47	0.32
(e)	AR	L	有	RLS	5.27	2.30	4.18	5.17	3.05	4.98	5.93	2.23	4.14
				Diff	0.65	0.99	0.56	0.53	0.77	0.78	1.11	1.45	0.86
(f)	ARMA	L	無	RLS	4.48	1.24	3.70	3.77	2.05	4.28	5.24	1.97	3.34
				Diff	-1.00	-1.00	-0.90	-1.50	-1.01	-0.96	-1.06	-0.47	-0.99
(g)	AR	T	無	MIS	4.46	1.31	3.77	4.09	2.18	4.72	6.23	1.89	3.58
				Diff	-1.02	-0.94	-0.82	-1.19	-0.87	-0.52	-0.07	-0.55	-0.75
(h)	AR	T	有	MIS	3.78	0.45	3.11	3.37	1.56	3.80	4.88	0.95	2.74
				Diff	-0.84	-0.86	-0.52	-1.27	-0.72	-0.40	0.05	0.17	-0.55
(i)	AR	L	無	MIS	5.40	2.56	4.53	4.82	3.01	5.21	6.23	2.37	4.27
				Diff	-0.08	0.32	-0.07	-0.46	-0.05	-0.03	-0.08	-0.07	-0.06
(j)	AR	L	有	MIS	4.95	2.32	4.07	4.52	2.60	4.92	5.18	1.33	3.73
				Diff	0.33	1.01	0.45	-0.12	0.31	0.72	0.35	0.55	0.45

注) 「平均手法」において T : トランスヴァーサル型で係数を平均

L : ラティス型で係数を平均

「手法」において

PAR : PARCOR

RLS : 重み付き RLS 法

MIS : MIS 法

Diff : PARCOR との差 (正なら当該欄の手法が勝る)

表 13.5: Closed CELP 符号化システムにおける平均セグメンタル
S/N 比の比較

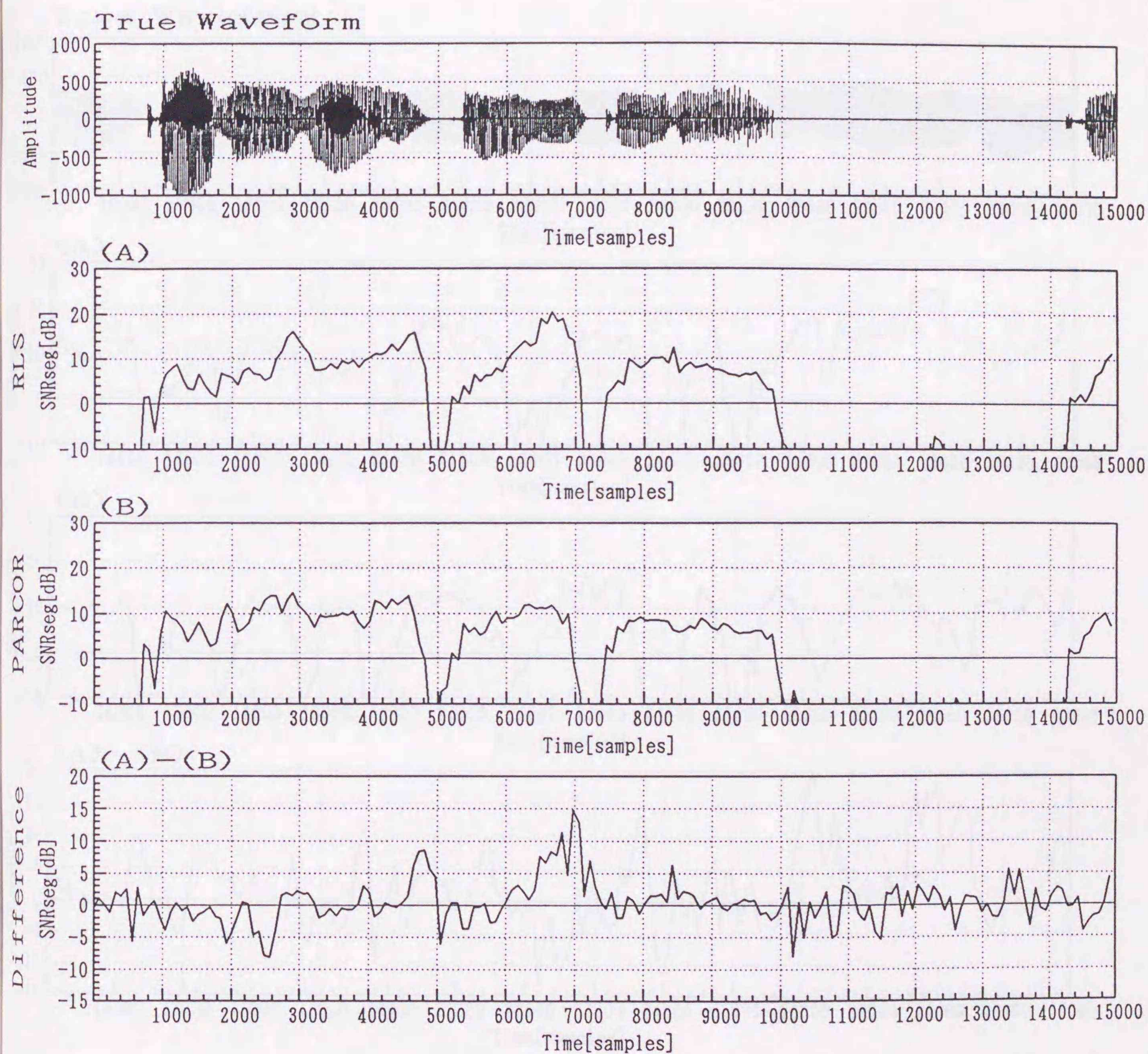


図 13.12: Closed CELP 符号化システムによる実験結果例

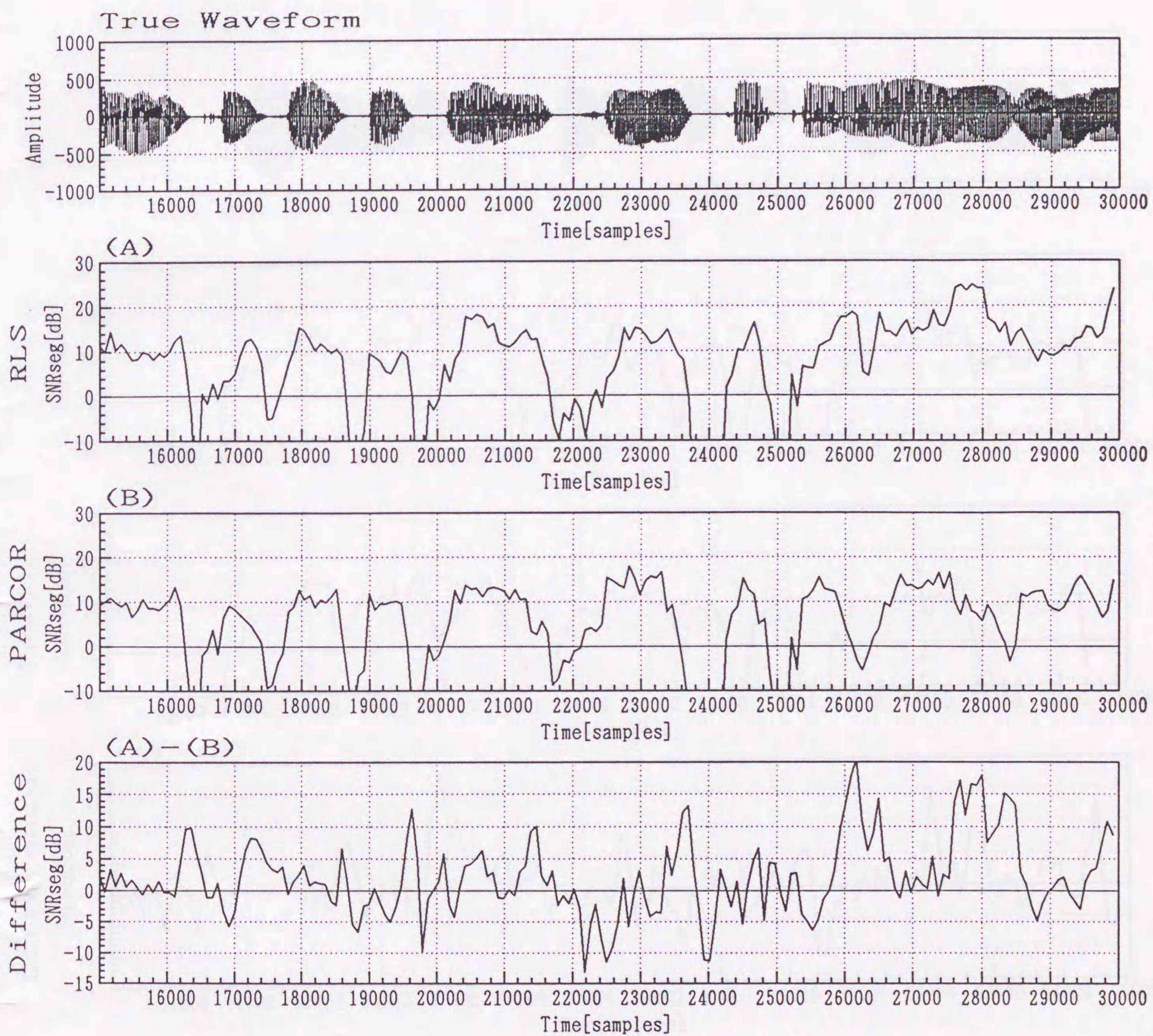


図 13.13: Closed CELP 符号化システムによる実験結果例

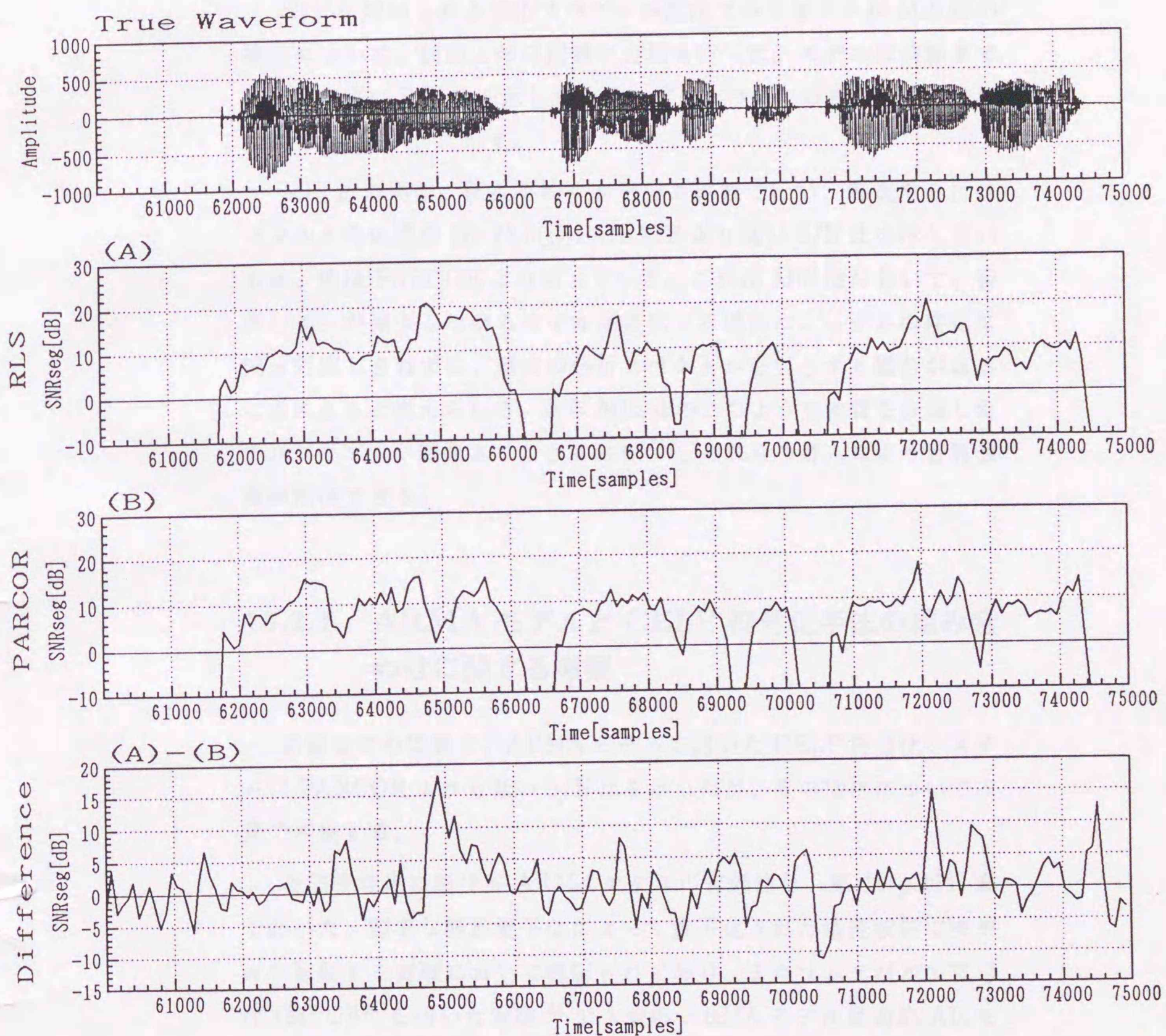


図 13.14: Closed CELP 符号化システムによる実験結果例

MIS 法による実験

STP に用いられる適応スペクトル推定アルゴリズムが MIS 法の場合について、前節とほぼ同様の実験を行った。モデルは前節までの実験で高い S/N 比を示した AR モデルについてのみ行った。結果を表 13.5(g)~(j) に示す。

MIS 法の場合、係数平均をラティス係数で行い、聴覚重み付けフィルタ有の場合 (表 13.5(j)) PARCOR より高い S/N 比を示しているが、他は PARCOR より劣っている。これは MIS 法において、音声パワーが減少しながらモデルが変化する場合に、モデルの変化に充分対応しきれずに、過去の分析スペクトルを引きずる場合があることによると考えられる。近年 MIS 法のこのような性質を改良したアルゴリズムが開発されてきており⁽⁶⁷⁾、それらの導入により音質改善が期待できる。

13.2.1 ARMA モデルと CELP 符号化手法の組み合わせに関する考察

前節までの実験で、ARMA モデルを用いた CELP 符号化システムは PARCOR よりも低い S/N 比を示したが、その理由について本節で考察する。

音声符号化における ARMA モデルの有効性は、第 10(p.111) 章で述べた、簡単な線形量子化によって量子化された残差波形でモデルを駆動する実験において確認されており、またマルチパルス符号化 (MPC)⁽⁵⁴⁾を用いた実験⁽⁶¹⁾でも適応 ARMA モデルは適応 AR モデルより優れた性質を示している。したがって本論文の ARMA モデル CELP 符号器が低い S/N 比を示すのは、ARMA モデルと CELP 符号器の組み合わせに原因があると考えられる。

ARMA 型音声生成モデル

$$y(k) = -\sum_{i=1}^n a_i(k)y(k-i) + u(k) + \sum_{j=1}^m b_j(k)u(k-j) \quad (13.1)$$

(y :出力、 u :入力、 a :AR 係数、 b :MA 係数、 n :AR 次数、 m :MA 次数) の MA 部は入力に関係しており、重み付き RLS 等の適応推定アルゴリズムは推定誤差を入力とみなしてモデルの推定を行っている。したがってそのようなアルゴリズムで求められた MA 係数は、推定アルゴリズムが出力する残差信号に対して最適な係数となっている。しかし CELP 符号器は推定アルゴリズムによって求められる残差波形とは異なる駆動波形を、コードブックの全探索によって全く新しく生成するため、既に求められている MA 係数と不一致を起し、合成音声の品質劣化を招くと考えられる。

マルチパルス符号化の場合も駆動信号を新しく生成するが、入力がパルスであるために零入力の時刻が多く、そのため MA 係数の影響が現れにくいと考えられ、MA 部の存在により正しく求められたスペクトルの傾きやホルマント位置等が符号化において有利に働いていると考えられる。しかし文献⁽⁶¹⁾に示されるマルチパルス符号化を用いた場合では、モデルを適応 AR とした場合でも性能が LPC に及ばないのに対し、本論文で提案する CELP を用いた手法ではモデルが AR の場合 LPC(PARCOR) よりも優れた性能を示している。

またオーディオテープに録音されたヒスノイズを含むデータを用いるよりも、ヒスノイズを含まない CD-ROM データのほうが適応分析システムにとって良い結果をもたらすことから、高精度適応分析がヒスノイズによって精度面で何らかの影響を受け、CELP 符号器と共に用いたとき合成音声の品質が劣化することが考えられる。これらの原因の究明と対策は今後の課題である。

Chapter 14

フォルマント情報を利用した CELP 符号化システム

14.1 まえがき

前章までで、適応信号処理アルゴリズムを導入した低ビット音声符号化システムを構築し、音声生成モデルとして AR モデルを用いた場合には、その有効性が確かめられた。しかし 13.2.1 で考察したように、ARMA モデルを用いた場合には、CELP で作られる新しい駆動信号と MA 部とがうまく適合せず、合成音声の品質低下をまねいた。また 13.1.2 でも述べたように、本来 MIS 法等の高精度適応 ARMA スペクトル推定アルゴリズムで十分な精度を得るためには、LPC の場合と比較してより高い次数で分析を行う必要があるが、前章までの音声符号化・合成システムでは、ビットレートの制限から十分な次数の確保は難しい。

実音声においてモデルの MA 部は、アンチフォルマントの表現よりはむしろスペクトル全体の傾きやフォルマントの正確な推定に貢献していると考えられる。そこで本章では、MA 部を含む高精度適応 ARMA スペクトルからフォルマント情報を抽出し、フォルマントを表さない極と零点は低次の AR モデルで近似する。これによりフォルマント精度を損なわずに ARMA スペクトルを AR スペク

トルに変換し、上述の問題の解決を試みる⁽¹²⁰⁾。

14.2 フォルマント抽出法

音声のスペクトルからフォルマントを抽出する方法は種々提案されているが、本章では次の2つの手法について考察する。

1. クラスタリング手法を応用した文献⁽⁷⁰⁾に示される手法
2. 伝達関数の極について、帯域幅の狭い順にある定まった個数を取り出し、それをフォルマントとみなす手法

予備実験によると、1. の手法⁽⁷⁰⁾は音声認識等でフォルマントの動き等を利用するような場合には有効であるが、しばしばフォルマントの連続性の観点から低域の大きなピークをフォルマントとして認識しないことがある。このような場合フォルマントと認識された極以外の極を低次モデルで近似してしまうと、もとのスペクトルとは大幅に異なったスペクトルが得られる場合がある。またこのような場合前後のスペクトルとの連続性がそこなわれることが多く、合成音声がざらつきの多い不安定なものになりやすい。

本章で論ずる符号化・合成では、受信側でもとのスペクトルを完全に再生しなければならないことや、決まった個数のパラメータを送受信する方がシステムが簡単になること等から、上述 2. の方法をフォルマント抽出手法として採用する。

音声信号はまず高次の MIS 法で分析され、短区間で平均される。ここで得られたスペクトルはキャンセレイションやフォルマントを表さない余計な極が含まれているため、平均後のパラメータについてモデルリダクションアルゴリズムを用いてより低次のスペクトルに変換し、余計な極・零点やキャンセレイションを取り除く。この係数に基づき極を求め、帯域幅の狭いものから順に決まった個数を選出しフォルマントとする。残った AR 係数と、全ての MA 係数は

スペクトルの傾きを表現すると考え、再びモデルリダクションアルゴリズムにかけて、1 次の AR モデルで近似する。求められたフォルマントの周波数と帯域幅、および 1 個の AR 係数は量子化され伝送される。残差信号の符号化は、11.3.2 で述べた Closed loop CELP 符号化方式によって行う。CELP の短期予測フィルタには、量子化後のフォルマントおよび傾きのパラメータから求められた AR 係数が用いられる。比較のため、PARCOR 分析から同様にフォルマント抽出・量子化を行う CELP 符号化システムも用意した。

フォルマント抽出手順をスペクトルの例で示したのが図 14.1 である。図 14.1(a) が AR 16 次 MA 6 次で求めたキャンセレーションを含む MIS スペクトル、(b) が (a) を AR 11 次 MA 2 次にリダクションしたスペクトル、(c) は (b) のうち帯域幅の狭い 4 個の極のみによるスペクトル、(d) は (b) のうち (c) に含まれない極と零点で構成されるスペクトル、(e) は (d) を AR 1 次にリダクションしたスペクトルで、最後の (f) は (c) と (e) を合成したスペクトルで、この (f) スペクトルが伝送され、CELP でも用いられる。(a),(b),(f) を比較すると、本手法がスペクトルのフォルマント特徴をそこなわずに情報圧縮していることがわかる。

14.3 符号化・合成実験

本章で用いる音声データは、12.2 および 13.1 で用いたオーディオテープ録音の短文データである。

実験システムの設定を表 14.1 (p.180) に示す。本章では表 14.1(a) と (b) に示される 2 つの設定を用意した。(a) は上述の本システムのねらいに沿った次数設定をしたもの、(b) はいくつかの実験のうち最も良い結果の得られた設定である。そのため (b) は上述のねらいとは完全には合致しない部分がある。

表 14.1(a) の設定で求められた STP スペクトルの時間推移を図 14.2 に示す。3 つあるスペクトル図のうち、右端が本章のフォルマン

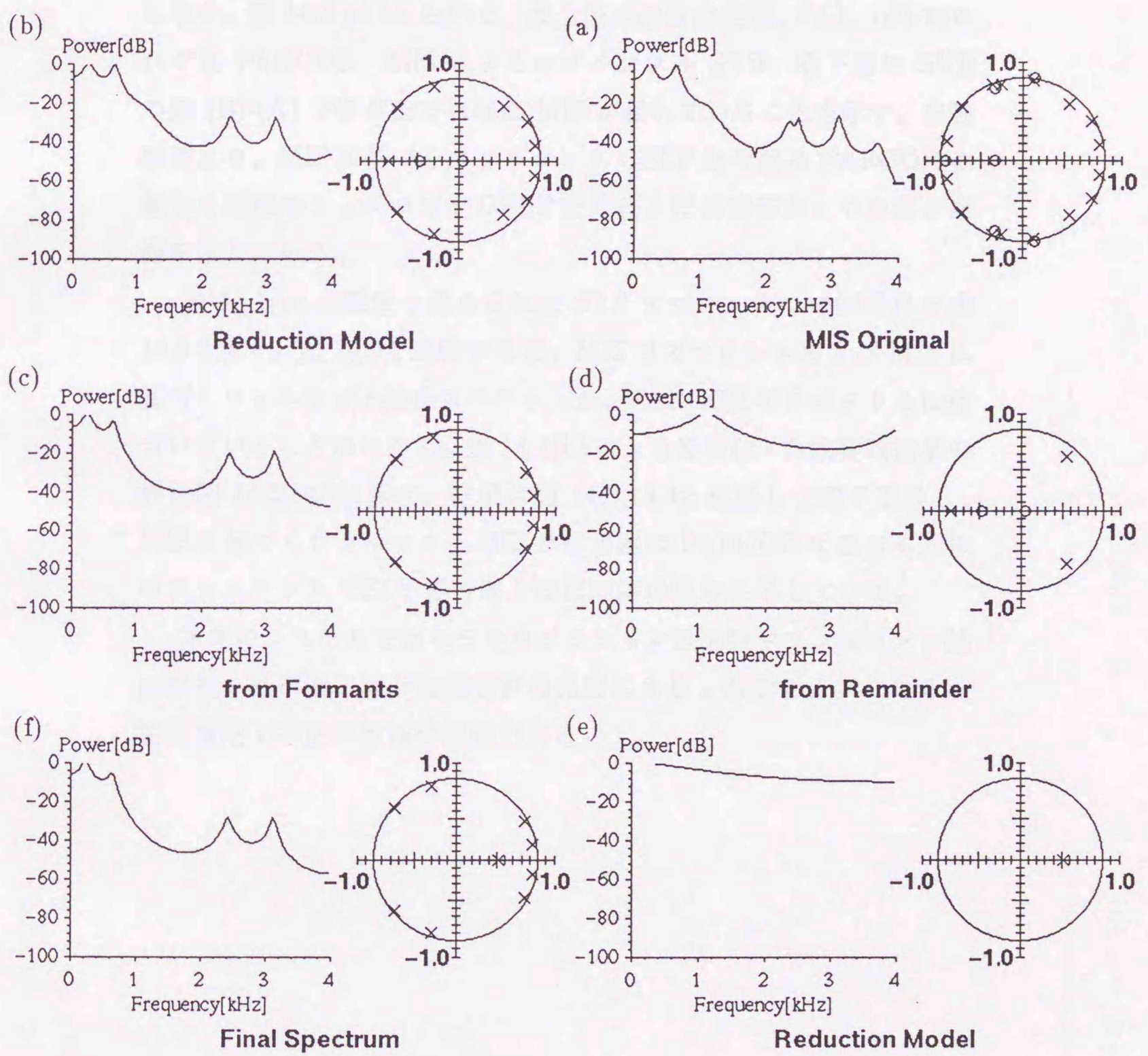


図 14.1: フォルマント抽出例

ト抽出手法によるスペクトル、中央がもともとの MIS スペクトル、左端が比較のための PARCOR による同様のフォルマント抽出スペクトルである。表 14.1(a) による符号化・合成実験結果の例を図 14.3, 14.4 に示す。図 14.3, 14.4 において、最上段が原音声波形、(A)、(B) はそれぞれ PARCOR、MIS によるセグメンタル SNR、最下段は SNR の差 $(B)-(A)$ で値が正であれば MIS が優れていることを示す。実験結果より、MIS に基づくフォルマント CELP 符号器は PARCOR に基づく同様のフォルマント CELP 符号器とほぼ同等か、やや劣る性能を示している。

表 14.1(b) の設定で求められた STP スペクトルの時間推移を図 14.5 に示す。図 14.2 と比較すると、MIS のスペクトルおよび MIS に基づくフォルマント抽出スペクトルが、PARCOR のスペクトルに近づいていることがわかる。表 14.1(b) による符号化・合成実験結果の例を図 14.6, 14.7 に示す。結果は図 14.3, 14.4 と比較して若干改善し、MIS に基づくフォルマント CELP 符号器は PARCOR に基づく同様のフォルマント CELP 符号器とほぼ同等の性能を示している。

本章のシステムではモデルリダクションの次数やフォルマント抽出個数、抽出手法等が合成音声の品質に大きく影響する。これらの定数設定・手法の改善が課題である。

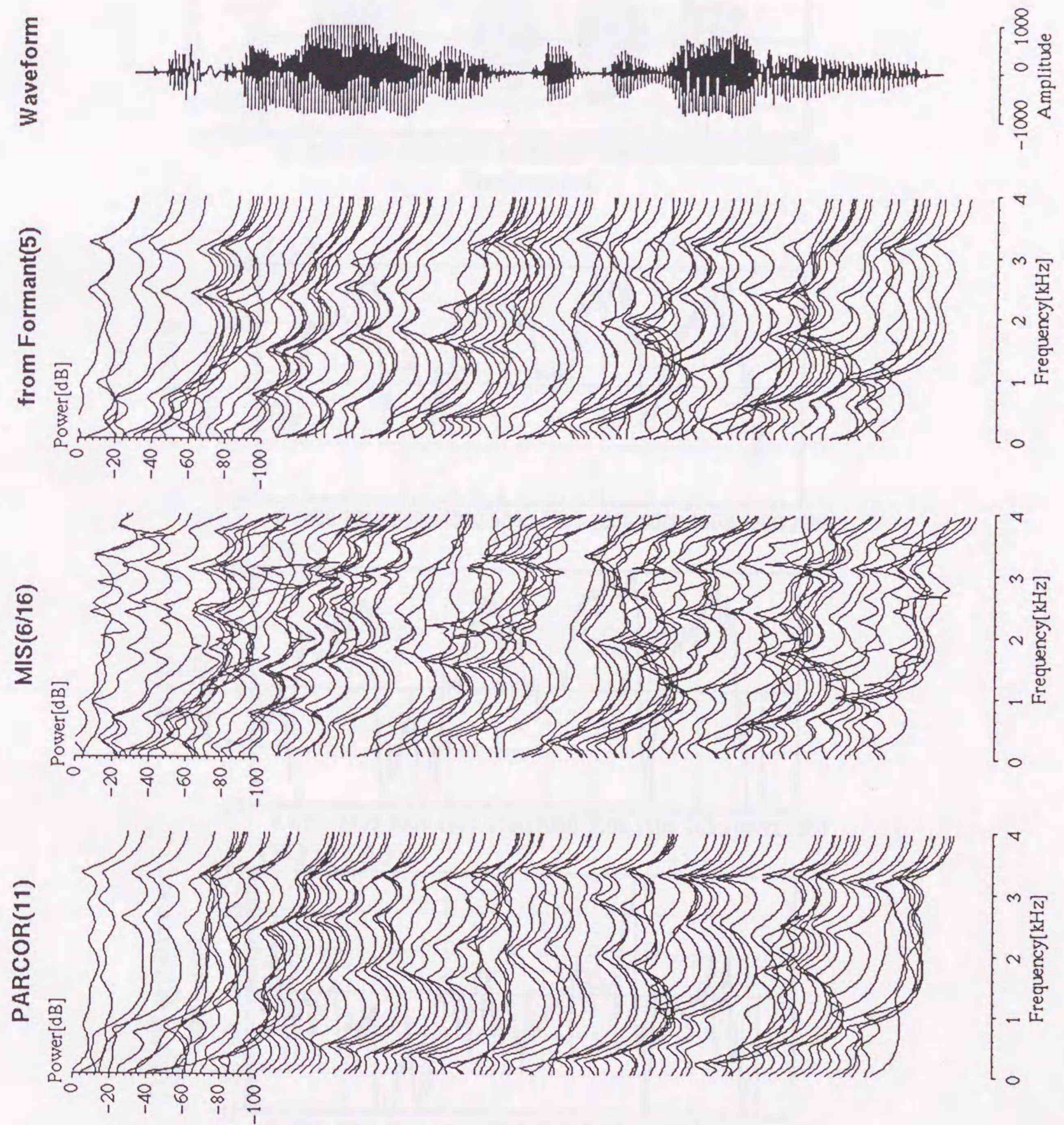


図 14.2: スペクトル比較 Type(a)

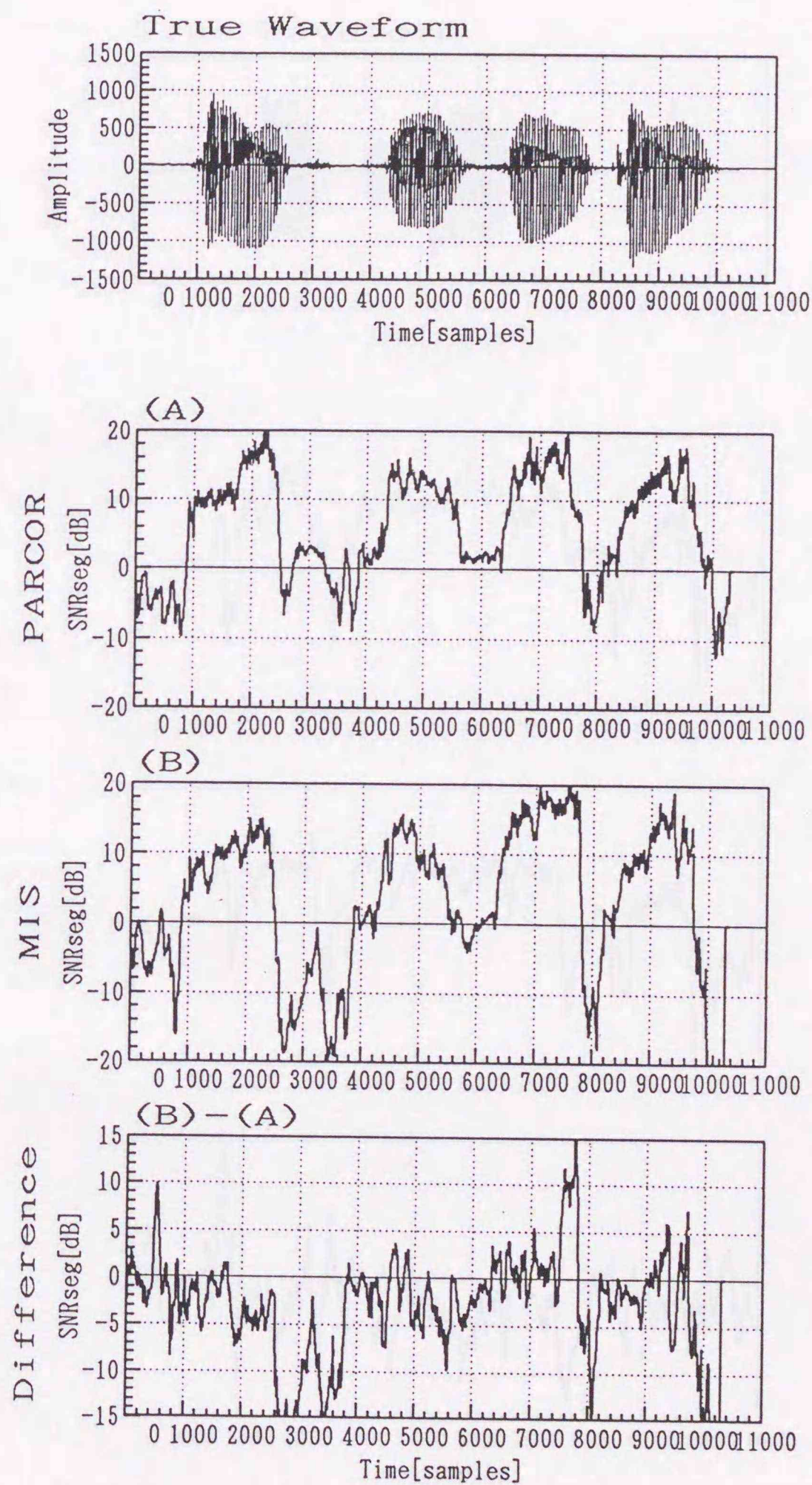


図 14.3: 実験結果例 (den1-8-1) Type(a)

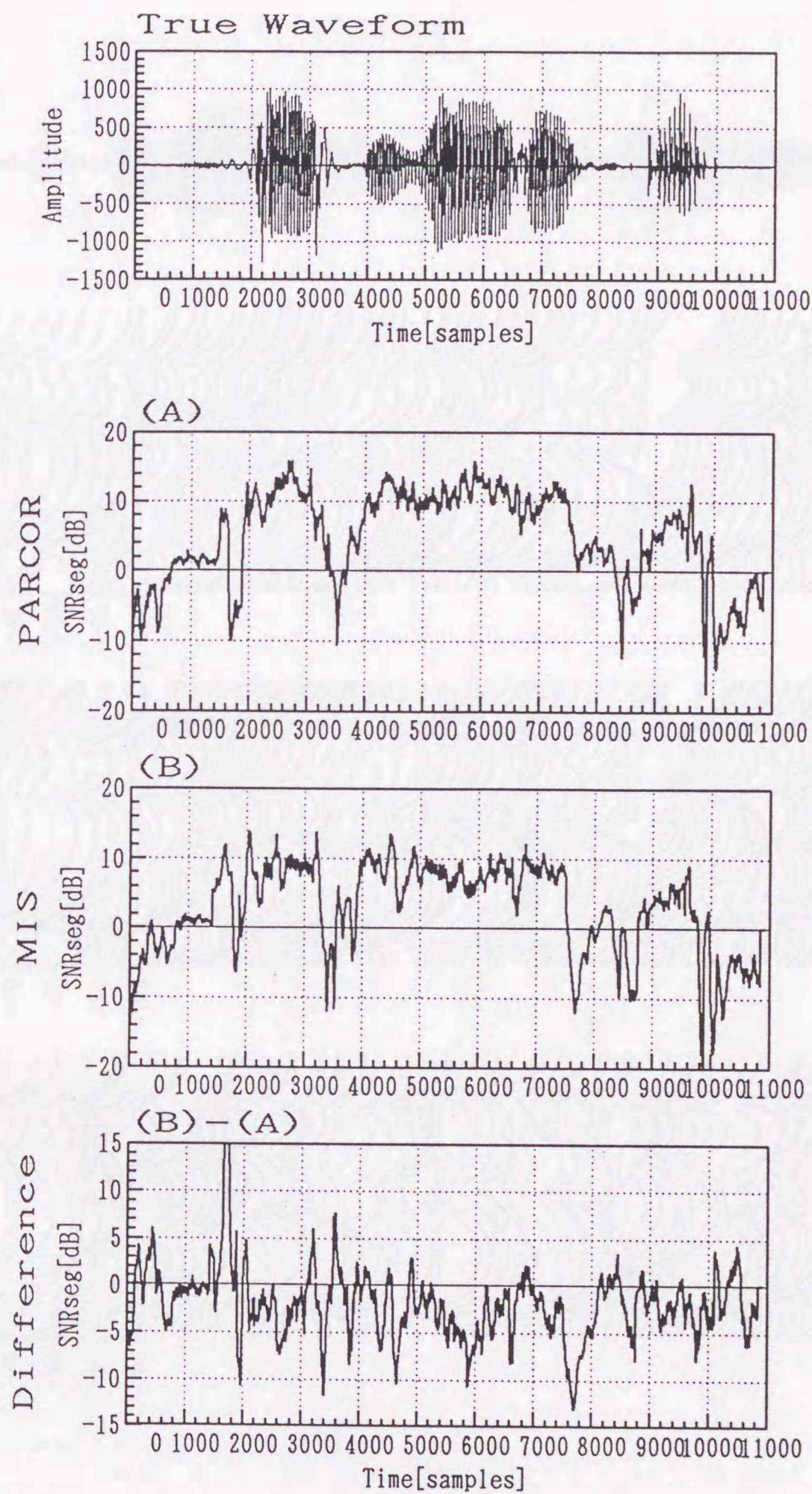


図 14.4: 実験結果例 (giji1-8-2) Type(a)

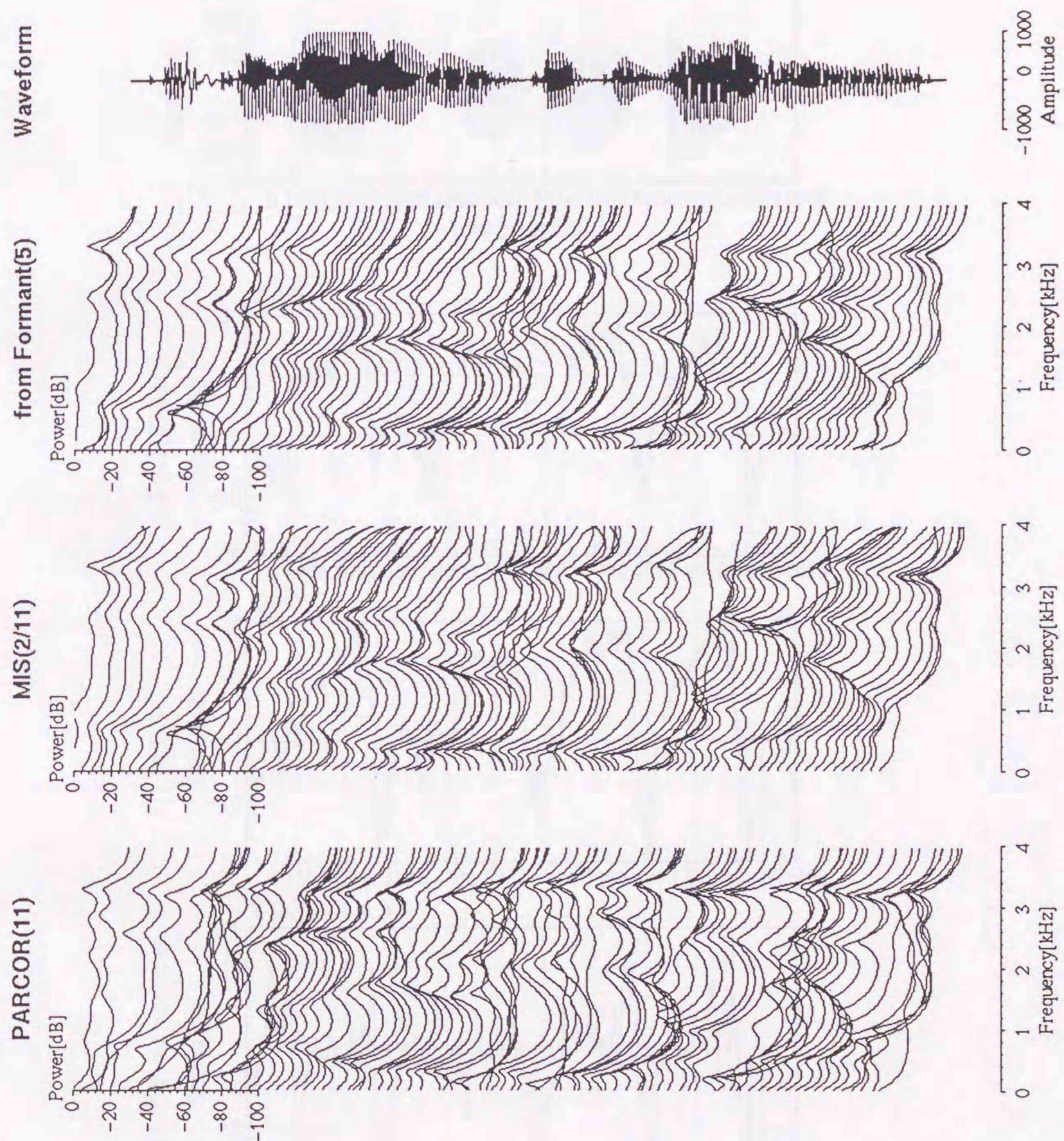


図 14.5: スペクトル比較 Type(b)

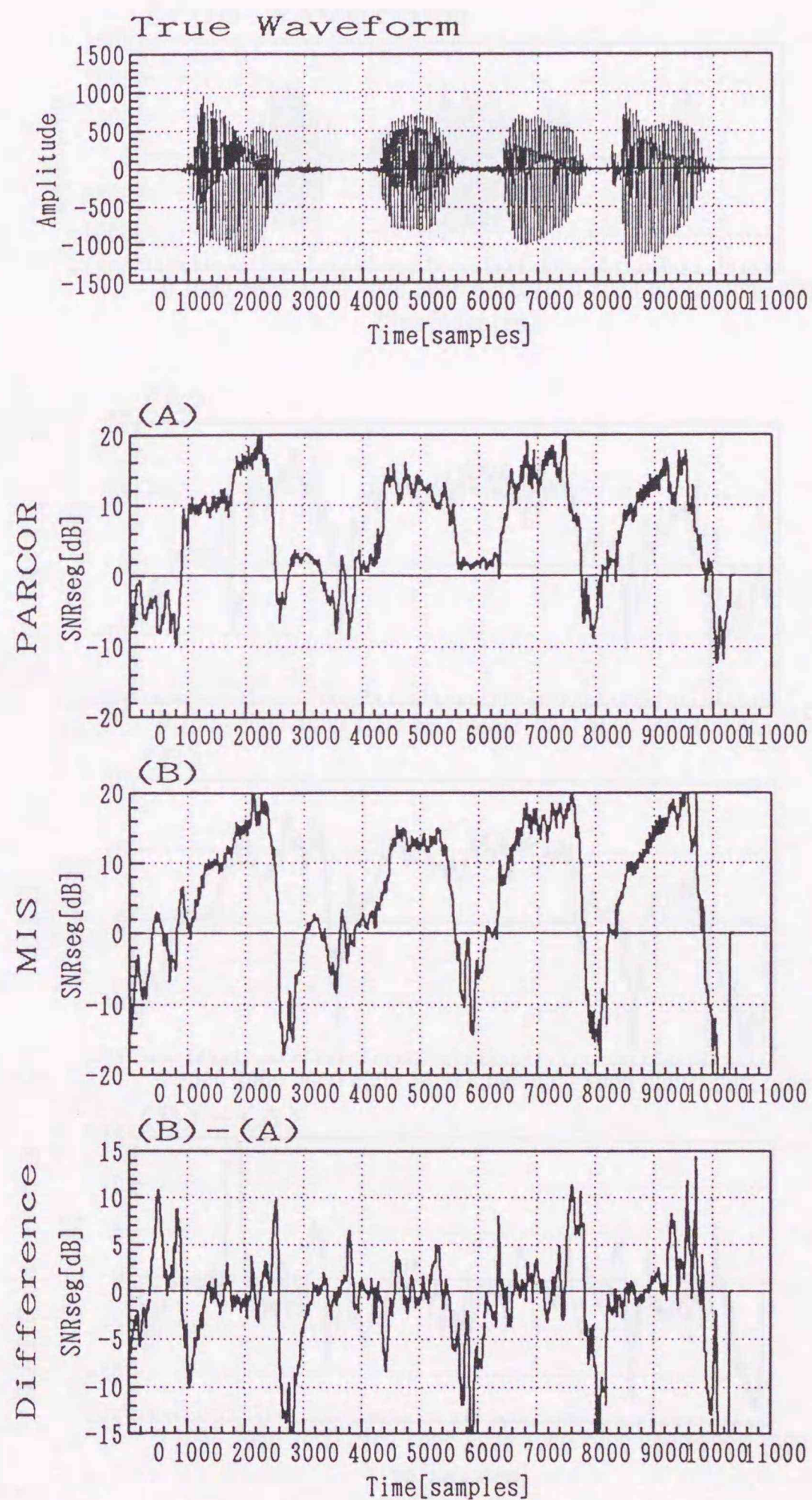


図 14.6: 実験結果例 (den1-8-1) Type(b)

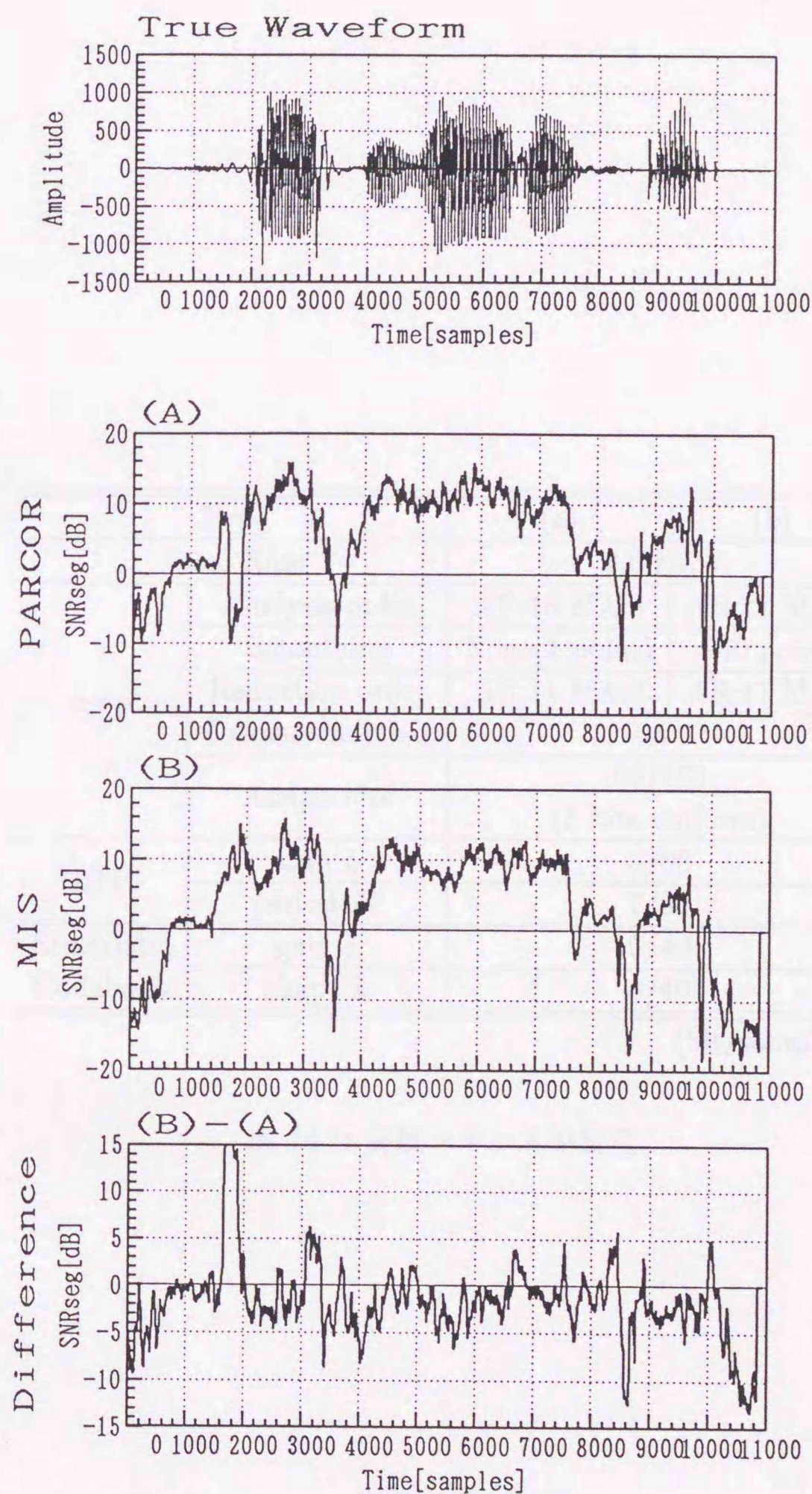


図 14.7: 実験結果例 (giji1-8-2) Type(b)

Type		(a)	(b)
Total Rate		8.0 kbps	
STP	Analysis order	AR:16 MA:6	AR:11 MA:2
	Smoothing	None(1 point)	240 points
	Reduction order	AR:11 MA:2	AR:11 MA:0
	Formant number	5	
	Assign bits	88/160 (8 bits uniform)	
LTP	gain b	3/80	
	period M	7/80	
Stochastic Codebook	gain g	5/40	
	shape k	8/40	

(bit/samples)

表 14.1: 実験システムの設定

Chapter 15

結論

結論

Chapter 15

結論

本論文では、適応信号処理アルゴリズムの音声情報処理への応用について論じた。すなわち、音声による人間とコンピュータのインタフェースを実現するための基礎研究として、高精度適応 ARMA スペクトル分析アルゴリズムを音声認識と音声符号化・合成の分野に応用し、実音声を対象とする音声処理システムを構築して、それらの有効性を実験により検証した。また本論文では、分析・認識の分野で有効なアルゴリズム選択型スペクトル分析システムを新しく提案し、その応用として連続音声のセグメンテーション候補検出システムを提案した。符号化・合成の分野では適応 ARMA スペクトル分析アルゴリズムを導入した CELP 符号化システムを提案し、良好な結果を得た。以下本論文で得られた結果を各章ごとに列記する。

第 2 章から第 8 章までが、音声の分析・認識に関する第 1 部となっている。

第 2 章の要約

1. 以下のトランスヴァーサル型適応 ARMA スペクトル推定アルゴリズムについて概説した。

(a) 重み付き RLS 法

(b) MIS 法

2. 以下の適応 ARMA ラティスモデルについて概説した。

(a) 改良適応 ARMA ラティスモデル

(b) 規格化 ARMA ラティスモデル

(c) コンパクト ARMA ラティスモデル

これらのアルゴリズムおよびモデルは、本論文全体にわたって用いられる。

第 3 章の要約

1. 有声音声において、相関操作を行う従来の方法に比べて、相関操作を行わない分析手法が理論的に高精度であることを述べた。
2. 相関操作を行わない有声音声向き直接的分析法を導出した。
3. 直接的分析法を用いて合成信号の分析実験を行い、直接的分析法は理論的には高精度であるが、雑音に敏感で実際の使用には問題があることが示された。

第 4 章の要約

1. 直接的分析法より実際の有声音声分析に向く重み付き RLS 法に注目し、評価関数内に導入された重み係数 λ の周波数特性を明らかにした。
2. λ の周波数特性の方形窓の周波数特性との比較から、重み付き RLS 法が有声音声に適していることの裏付けを得た。

第 5 章の要約

1. 重み付き RLS 法で有声音声を分析する際、得られたスペクトルデータを選択する必要があることを述べた。
2. 経験的 if-then ルールを用いた自動選択システムを作成した。
3. 学習により if-then ルールと同様の知識を獲得するニューラルネットのノードを用いた自動選択システムを作成した。
4. 重み付き RLS 法と自動選択システムとを用いることにより、高精度な有声音声分析スペクトルを得た。

第 6 章の要約

1. 有声音声と無声音声は異なるアルゴリズムで分析する必要があることを述べた。
2. 有声音声と無声音声を異なるアルゴリズムで分析するために、経験的ルールに基づく有声・無声分離手法を提案した。
3. 実音声を用いた実験により、提案する手法によって良好な有声・無声分離が行えることを示した。

第 7 章の要約

1. 新しいデジタル信号処理手法として、信号の性質に応じてアルゴリズムを選択して分析を行う方式を提案した。
2. 第 2 章から第 6 章までの手法を統合し、アルゴリズム選択型スペクトル分析システムを構築した。
3. 提案するシステムで、高精度なスペクトル分析が行えることを示した。

第 8 章の要約

1. 第2章で述べた適応信号処理アルゴリズムによるスペクトルと、第6章で述べた有声・無声分離システムの出力を用いて、連続音声のセグメンテーション候補点検出システムを提案した。
2. 分離に用いるいくつかのパラメータについて検討した。
3. 提案したセグメンテーション候補点検出システムを用いて実験を行い、良好なセグメンテーション候補検出が行えることを示した。

第9章から第14章までが、音声の符号化・合成に関する第2部となっている。

第9章の要約

1. 音声符号化システムにおける適応 ARMA 分析の必要性について述べた。
2. 符号化システムにおける適応 ARMA 分析アルゴリズムの位置を明らかにした。
3. 本論文で用いる合成音声の品質評価尺度である、セグメンタル S/N 比について述べた。

第10章の要約

1. 短期予測フィルタ (STP) の残差信号を単純な線形量子化によって量子化する実験システムを作成した。
2. 実験により、残差を単純に線形量子化する場合、鼻子音等で適応 ARMA 分析に基づく符号化システムが高い S/N 比を示すことが明らかになった。
3. 実験結果は符号化システムにおける適応 ARMA 分析の必要性・有効性の裏付けとなった。

第 11章の要約

1. 以下の章で用いる CELP 符号化手法について概説した。
2. 長期予測フィルタ (LTP) について概説した。
3. 聴覚重み付けフィルタ (PWF) について概説し、AR 型から ARMA 型へ拡張する際の問題点について述べた。

第 12章の要約

1. 規格化 ARMA ラティスモデルを用いた Open loop CELP 符号化システムを構築した。
2. 都道府県名データを用いて PARCOR との比較実験を行った。
3. 短文データを用いて PARCOR との比較実験を行った。
4. ARMA ラティスモデルを用いた CELP 符号化システムと、PARCOR による CELP システムとは、ほぼ同等の性能を示した。

第 13章の要約

1. 規格化 ARMA ラティスモデルおよび第 2章で述べたトランスヴァーサル型適応 ARMA 分析アルゴリズムを用いた Closed loop CELP 符号化システムを構築した。
2. 短文データを用いて PARCOR との比較実験を行い、提案するシステムは PARCOR とほぼ同等の性能を持つことが示された。
3. 高品質の CD-ROM 音声データを用いて PARCOR との比較実験を行い、AR モデルに基づく重み付き RLS 法を用いた CELP 符号化システムが最も良い結果を示すことを明らかにした。

4. 上述の AR モデル重み付き RLS 法による CELP 符号化システムでは、PARCOR と比較して、平均セグメンタル S/N 比で 0.53~1.45dB の音質改善が得られた。
5. ARMA モデルを用いた CELP 符号化システムは概して良好な結果を得られないことを示し、MA 部と CELP の関係からその原因を考察した。

第 14 章の要約

1. MA 部と CELP の不整合を避けることを目的として、高精度 ARMA スペクトルからフォルマント情報を抽出して符号化を行う CELP システムを作成した。
2. 短文データを用いて同様の PARCOR に基づくフォルマント符号化システムと比較実験を行った。
3. 適応 ARMA スペクトルに基づくフォルマント符号化システムは、PARCOR に基づく同様のシステムとほぼ同等の性能を示した。

本研究での成果をまとめると、以下のようになる。まず音声の分析において、信号の性質に応じて最適な適応 ARMA 分析アルゴリズムを選択して分析する方式の有効性が示された。認識の分野において、適応 ARMA 分析によって得られたスペクトルが、連続音声認識に有効である可能性があることを、セグメンテーション実験によって示した。符号化の分野では、AR モデルに基づく適応スペクトル推定アルゴリズムを用いた CELP 符号化システムが、音声符号化に有効であることを示した。以上より本研究によって、適応信号処理アルゴリズムの音声情報処理における有効性が示された。

適応信号処理の音声情報処理への応用について、今後の課題を述べる。

まず音声認識について、本論文の内容をさらに進め、適応信号処理手法に基づく連続音声認識システムの構築が望まれる。最終的な認識結果を得るまでには、自然言語の性質に深く関わった知識処理が必要と考えられ、信号処理と自然言語処理との融合が必要となる。

音声符号化・合成に関わる課題としては、本論文の符号化システムでは適応信号処理で各サンプル点で得られたスペクトルをフレーム周期で平滑化しているため、適応信号処理の分解能を十分に活用しているとは言えない。特定のフレームにとらわれない効率の良いスペクトルパラメータの符号化手法が望まれる。また ARMA スペクトルの精度を活用するために、ARMA スペクトルに適した新しい残差符号化手法の開発も必要である。より低ビットの音声符号化をめざす場合、保存すべき個人性とは何であるかについての検討も、重要であると考えられる。

適応 ARMA スペクトル分析アルゴリズムは一般に一括型 AR 分析よりも複雑であり、多くの計算コストを要する。また精度が高いため得られた結果を表現するには多くのパラメータを必要とし、それはコンピュータ上では多くのメモリを必要とすることを意味する。コンピュータの速度の向上やメモリのコスト減少を考慮にいれながら、並列処理等により計算コストの減少を計り、トータルな性能向上をめざす必要があると考えられる。また、より適応信号処理に適した情報処理の分野を開拓することも重要である。

本研究の目的である、音声による、人間にとって使いやすいコンピュータとのインタフェースを実現するには、結局人間の頭脳の中の情報処理に深く立ち入らねばならないと考えられる。人間の思考のメカニズムを解明することはたいへん興味深く、また工学的にもきわめて大きな成果をもたらすと予想されるが、人間の思考が完全に数理的にシミュレーションされ、さらに工学的にコントロール可能になる必要があるかどうかについては、疑問が多いと考える。「技術」は人間の思考そのもの、また精神状態に深く関与する。現在の

段階でも、コンピュータの導入が人間に新たなテクノストレスをもたらしている事実を考えると、将来出現するであろう人間の思考のメカニズムを取り入れた種々の「技術」は、単に興味や能率や経済性ばかりを追求するのではなく、人間がより「創造的」に、真の意味で人間らしく生きるために、活用されなければならない。

謝辞

本研究は、筆者が北海道大学大学院工学研究科電子工学専攻在学中に、北海道大学工学部電子工学科電子機器工学講座において、1987年より約5年間に行ったものである。

本研究の遂行にあたり、常に適切で温かいご指導を頂いた北海道大学工学部電子工学科電子機器工学講座、栃内香次教授に心より感謝いたします。

本研究の遂行にあたり、有益なご討論を頂いた、北海道大学工学部電子工学科電子回路工学講座、小川吉彦教授に深謝いたします。

本研究の遂行にあたり、学会等で温かくご指導頂いた、北海道大学応用電気研究所電子回路部門、永井信夫教授に深謝いたします。

本研究の遂行にあたり、ご討論・ご指導いただいた、北海道大学工学部情報工学科情報処理工学講座、新保勝教授に深謝いたします。

本研究の遂行にあたり、基礎から専門に至るまで終始温かくご指導頂いた北海道大学工学部電子工学科電子機器工学講座、宮永喜一助教授に深く感謝いたします。

本研究の端緒を与えて下さり、有益なご指導を頂いた釧路高等専門学校校長、永田邦一先生に深謝いたします。

筆者の研究期間中、公私にわたり温かいご指導、ご協力を頂いた北海道大学工学部電子工学科電子機器工学講座、故須戸満助手に感謝すると共に、ご冥福をお祈りいたします。

筆者の研究期間中、ご指導、ご協力を頂いた北海道大学工学部電子工学科電子機器工学講座、森伸夫助手に深謝いたします。

筆者の研究の遂行にあたり、コンピュータの保守やプログラミング等に関してご指導、ご協力を頂いた北海道大学工学部電子工学科電子機器工学講座、斎川勝男技官に深謝いたします。

ゼミ、学会等でご討論、ご指導を頂いた釧路高等専門学校、大槻典行氏に深謝いたします。

多忙の中基礎ゼミ等でご指導を頂いた日本電気(株)、舟木慶一氏に深謝いたします。

筆者の研究期間中、公私にわたりご協力頂いた北海道大学工学部電子工学科電子機器工学講座の大学院学生諸氏に厚く御礼申し上げます。特に、数式の理論的扱い等で、筆者を強力に援助して下さいった中垣淳氏に感謝いたします。

最後に筆者の研究生活において、温かい援助を与えて下さった両親および家族に、また心の支えとなって下さった多くの友人達に、深く感謝いたします。

参考文献

第 1 章

- (1) 斎藤収三、中田和男：“音声情報処理の基礎”，オーム社（1981-11）.
- (2) 古井貞熙：“デジタル音声処理”，東海大学出版会（1985-09）.
- (3) J.D.Markel and A.H.Gray Jr.：“Linear prediction of speech”，Springer-Verlag（1976）.
- (4) 藤村靖編：“音声科学”，東京大学出版会（1972-03）.
- (5) 板倉文忠、斎藤収三：“統計的手法による音声スペクトル密度とホルマント周波数の推定”，信学論（A），**53-A**, 1, pp.35-42（1970-01）.
- (6) B.S.Atal and Suzanne L. Hanauer：“Speech Analysis and Synthesis by Linear Prediction of the Speech Wave”，J. Acoust. Soc. Am., **50**, 2, pp.637-644（1971）.
- (7) 東倉洋一、板倉文忠：“PARCOR 帯域圧縮方式における音声品質向上”，信学論（A），**J61-A**, 3, pp.254-261（1978-03）.
- (8) B.S.Atal：“Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification”，J. Acoust. Soc. Am., **55**, 6, pp.1304-1312（Jun.1974）.

第2章

- (9) 深林太計志、鈴木久喜 : “極-零形の線形モデルによる音声分析”, 信学論 (A), **58-A**, 5, pp.270-277 (1975-05).
- (10) 森川博由、藤崎博也 : “AR・MA パラメータの同時推定法による音声分析”, 信学論 (A), **J61-A**, 3, pp.195-202 (1978-03).
- (11) H.Morikawa and H.Fujisaki : “Adaptive Analysis of Speech Based on a Pole-zero Representation”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-30**, 1, pp.77-88 (Feb.1982).
- (12) 森川博由、藤崎博也 : “SEARMA 法による音声分析における観測区間の適応的制御”, 音響誌, **39**, 8, pp.512-520 (1983-08).
- (13) 宮永喜一、三木信弘、永井信夫 : “ピッチ推定を含めた音声の ARMA パラメータの一推定法”, 信学論 (A), **63-A**, 11, pp.737-744 (1980-11).
- (14) 宮永喜一、三木信弘、永井信夫、羽鳥孝三 : “時変 ARMA パラメータの適応的同時推定”, 信学論 (D), **J64-D**, 4, pp.308-315 (1981-04).
- (15) Y.Miyanaga, N.Miki, N.Nagai and K.Hatori : “A Speech Analysis Algorithm Which Eliminates the Influence of Pitch Using Model Reference Adaptive System”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-30**, 1, pp.88-96 (Feb.1982).
- (16) 宮永喜一、三木信弘、永井信夫 : “音声の ARMA 分析に関する一考察”, 音響誌, **40**, 6, pp.397-406 (1984-06).
- (17) Y.Miyanaga, N.Miki and N.Nagai : “Adaptive Identification of Time-varying ARMA Speech Model”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-34**, 3, pp.423-433 (Jun.1986).

- (18) 宮永喜一、三木信弘、永井信夫：“自己回帰移動平均モデルの適応的次數推定”，信学論 (A), **J65-A**, 12, pp.1239-1246 (1982-12).
- (19) 宮永喜一、三木信弘、永井信夫：“線形予測理論に基づく ARMA デジタルラティスフィルタ”，信学論 (A), **J67-A**, 5, pp.487-494 (1984-05).
- (20) Y.Miyanaga, N.Ohtsuki and K.tochinai：“Adaptive ARMA Lattice Modeling for Speech Processing”，J. Acoust. Soc. Japan(E), **11**, 3, pp.173-182 (May 1989).
- (21) 宮永喜一、栃内香次：“ARMA ラティスモデルの等価回路記述について”，信学論 (A), **J73-A**, 6, pp.1048-1057 (1990-06).
- (22) 保原信、橋本清：“モデル規範適応制御を用いた音声符号伝送方式”，信学論 (A), **J70-A**, 8, pp.1195-1198 (1987-08).
- (23) Y.Miyanaga：“Estimation of ARMA Process and its Applications for Digital Signal Processing”，Ph.D. dissertation, Dep. Electron. Eng., Hokkaido Univ., Sapporo, Japan (Apr.1986).
- (24) 舟木慶一：“適応的手法を用いた母音音声分析に関する研究”，北海道大学修士論文 (1988-03).
- (25) 宮永喜一、栃内香次：“音声の等価回路モデルを用いた特徴記述”，信学技報, **DSP89-19**, pp.83-90 (1989-06).
- (26) 田中雅史、宮永喜一、栃内香次：“音声の最小実現回路モデルを用いた認識の実験的評価”，信学技報, **DSP89-19**, pp.91-97 (1989-06).

- (27) 河原英紀、栃内香次、永田邦一：“小区間の線形予測分析とその誤差評価”，音響誌，**33**, 9, pp.470-479 (1977-09).
- (28) 村上寛之、斎川勝男、栃内香次、永田邦一：“小区間線形予測分析法と母音音声の変動に関する検討”，北海道大学工学部研究報告，132, pp.119-128 (1986-07).

第 4 章

- (29) 深林太計志：“適応ラティス型分析によって得られるラティスパラメータの平滑化について”，信学論 (A), **J69-A**, 10, pp.1294-1297 (1986-10).
- (30) S.H.Ardalan and S.T.Alexander：“Fixed-Point Roundoff Error Analysis of the Exponentially Windowed RLS Algorithm for Time-Varying Systems”，IEEE Trans. Acoust., Speech & Signal Process., **ASSP-35**, 6, pp.770-783 (Jun.1987).

第 5 章

- (31) D.E.Rumelhart and J.L.McClelland：“Parallel Distributed Processing：Explorations in the Microstructure of Cognition Vol.1”，The MIT Press (Jan.1986).
- (32) R.P.Lippmann：“An Introduction to Computing with Neural Nets”，IEEE ASSP Magazine, pp.4-22 (Apr.1987).
- (33) T.Kohonen：“The “Neural” Phonetic Typewriter”，IEEE COMPUTER, pp.11-22 (Mar.1988).

第 6 章

- (34) V.W.Zue and L.F.Lamel：“An Expert Spectrogram Reader：A Knowledge-based Approach to Speech Recognition”，IEEE Proc. ICASSP-86, pp.1197-1200 (1986).

- (35) 溝口理一郎、角所収：“知識工学を応用した音声認識”，音響誌，**42**, 12, pp.942-947 (1986-12).
- (36) R.Miaoguchi, K.Tsujino and O.Kakusyo：“A Continuous Speech Recognition System Based on Knowledge Engineering Techniques”，IEEE Proc. ICASSP-86, pp.1221-1224 (1986).
- (37) 溝口理一郎、角所収：“知識工学と音響・音声処理”，音響誌，**42**, 5, pp.396-402 (1986-05).
- (38) 福田、溝口、角所：“知識工学的手法を用いた連続音声認識”，信学技報，**EA84-66**, pp.41-48 (1985-01).
- (39) 溝口、田中、福田、辻野、角所：“連続音声認識エキスパートシステム-SPREX-”，信学論 (D)，**J70-D**, 6, pp.1189-1198 (1987-06).
- (40) 辻野、櫻井、野村、千種、溝口、角所：“知識ベース構築支援環境を備えた連続音声認識エキスパートシステム-SPREXII-”，信学論 (D)，**J71-D**, 3, pp.531-542 (1988-03).
- (41) 森島繁生、原島博：“統計的手法に基づくプロダクションルールの自動抽出法とファジィ木探索”，信学論 (D)，**J69-D**, 11, pp.1754-1764 (1986-11).
- (42) 森島、原島：“音響処理と記号処理とを融合した単語音声認識システムの構成”，信学論 (D)，**J70-D**, 10, pp.1890-1901 (1987-10).

第7章

- (43) R.Davis, et al: “An overview of production systems”, Machine Intelligence, 8, pp.300-332 (1977).

第8章

- (44) 迫江、千葉：“動的計画法を利用した音声の時間正規化に基づく連続単語認識”，音響誌，**27**，9，pp.483-490 (1971-09).
- (45) 渡辺隆夫：“音声認識におけるパターンマッチングの手法”，音響誌，**42**，9，pp.725-730 (1986-09).
- (46) アレックス ワイベル：“時間遅れ神経回路網 (TDNN) による音韻認識”，信学技報，**SP87**-100 (1987-12).
- (47) 大河内正明：“Hidden Markov Model に基づいた音声認識”，音響誌，**42**，12，pp.936-941 (1986-12).
- (48) 大河内正明：“マルコフモデルによる音声認識”，信学誌，**70**，4，pp.352-358 (1987-04).
- (49) 今井、古市：“連続音声の音素的単位へのセグメンテーション”，信学論 (D-II)，**J72-D-II**，1，pp.11-21 (1989-01).
- (50) 畑崎香一郎、小森康弘、川端豪、鹿野清宏：“スペクトログラムリーディング知識を用いた音韻セグメンテーションエキスパートシステム”，信学論 (D-II)，**J73-D-II**，1，pp.1-9 (1990-01).
- (51) 小森康弘、畑崎香一郎、川端豪、鹿野清宏：“スペクトログラムリーディング知識とニューラルネットワークを用いた音韻認識エキスパートシステム”，信学論 (D-II)，**J73-D-II**，1，pp.10-18 (1990-01).
- (52) 松下充宏：“適応信号処理を用いた連続音声のセグメンテーション”，北海道大学卒業論文 (1989-03).

第9章

- (53) H.Morikawa and H.Fujisaki：“A Speech Analysis-Synthesis System Based on the ARMA Model and its Evaluation”，IEEE Proc. ICASSP-86，pp.1253-1256 (Apr.1986).

- (54) B.S.Atal and J.R.Remde : "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", IEEE Proc. ICASSP-82, pp.614-617 (Apr.1982).
- (55) 小澤一範、荒関卓 : "ピッチ情報を用いる 9.6~4.8kbit/s マルチパルス音声符号化方式", 信学論 (D-II), **J72-D-II**, 8, pp.1125-1132 (1989-08).
- (56) R.C.Rose and T.P.Barnwell : "Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders", IEEE Trans. Acoust., Speech & Signal Process., **38**, 9, pp.1489-1503 (Sep.1990).
- (57) Henry Scheffe : "An Analysis of Variance for Paired Comparisons", Am. Statis. Assoc. J., **47**, pp.381-400 (Sep.1952).
- (58) 境久雄、中山剛 : "聴覚と音響心理" (1978-05).

第 10 章

- (59) 北脇信彦、板倉文忠、斉藤収三 : "PARCOR 形音声分析合成系における最適符号構成", 信学論 (A), **J61-A**, 2, pp.119-126 (1978-02).
- (60) 王剛、三木信弘、永井信夫 : "ARMA モデルを用いた音声の一合成法に関する音質評価", 信学論 (A), **J71-A**, 12, pp.2111-2119 (1988-12).
- (61) 王剛 : "ARMA モデルによる音声の分析合成における音質評価の研究", 北海道大学工学博士学位論文 (1990-12).

第 11 章

- (62) B.S.Atal : "Predictive Coding of Speech at Low Bit Rates", IEEE Trans. Commun., **COM-30**, 4, pp.600-614 (Apr.1982).

- (63) M.R.Schroeder and B.S.Atal : "Code-excited Linear Prediction (CELP): High-quality Speech at Very Low Bit Rates", IEEE Proc. ICASSP-85, pp.937-940 (1985).
- (64) 熊野聡、森島繁生、原島博 : "CELP 音声符号化の改良の一検討", 第4回デジタル信号処理シンポジウム講演論文集, pp.297-302 (1989-12).
- (65) I.M.Trancoso and B.S.Atal : "Efficient Search Procedures for Selection the Optimum Innovation in Stochastic Coders", IEEE Trans. Acoust., Speech & Signal Process., **38**, 3, pp.385-396 (Mar.1990).

第 13 章

- (66) 堀田英輔、宮永喜一、栃内香次 : "2入力 ARMA モデルを用いた音声の特徴抽出", 信学論 (A), **J73-A**, 9, pp.1476-1488 (1990-09).
- (67) 堀田英輔、宮永喜一、栃内香次 : "ARMA-D モデルを用いた時変パラメータ推定法", 信学技報, **DSP91-2**, pp.9-16 (1991-05).

第 14 章

- (68) A.Erell, Y.Orgad and J.L.Goldstein : "JND's in the LPC Poles of Speech and Their Application to Quantization of the LPC Filter", IEEE Trans. Signal Process., **39**, 2, pp.308-318 (Feb.1991).
- (69) 芹沢昌宏、三木信弘、永井信夫 : "極座標上でのクラスタリング法を用いたフォルマントトラッキング", 昭 63 北海道連大, p.159 (1988-10).

- (70) 芹沢昌宏：“音声信号の動的特徴パラメータ抽出のための適応的 ARMA 係数推定法に関する研究”，北海道大学工学博士学位論文 (1989-12).
- (71) 広重真人：“アルゴリズム選択機能を持つ適応信号処理スペクトル分析システムに関する研究”，北海道大学修士論文 (1989-03).

著者の発表論文

A 学会誌

- (101) 広重真人、宮永喜一、栃内香次：“スペクトル選択機能を備えた有声音声向き適応信号処理システム”，信学論 (D-II), **J72-D-II**, 4, pp.476-486 (1989-04).
- (102) 広重真人、宮永喜一、栃内香次：“適応スペクトル推定アルゴリズムを導入した CELP 音声符号化システム”，信学論 (D-II), 投稿中.

B 国際学会

- (103) M.Hiroshige, Y.Miyanaga and K.Tochinai：“An Adaptive System of Multiple Algorithms for Speech Spectrum Estimation”，Proc. JTC-CSCC'89, pp.156-161 (Jun.1989).
- (104) M.Hiroshige, Y.Miyanaga and K.Tochinai：“A Study of a Speech Coding Method Using a Compact ARMA Lattice Filter”，Proc. IASTED Inter. Sympos. ADAPT. CONTR. & SIGNAL PROCESS., pp.32-35 (Oct.1990).
- (105) M.Hiroshige, Y.Miyanaga and K.Tochinai：“A Study of CELP Speech Coders Using ARMA Model and Adaptive Algorithms”，Proc. JTC-CSCC'91, pp.171-176 (Jul.1991).

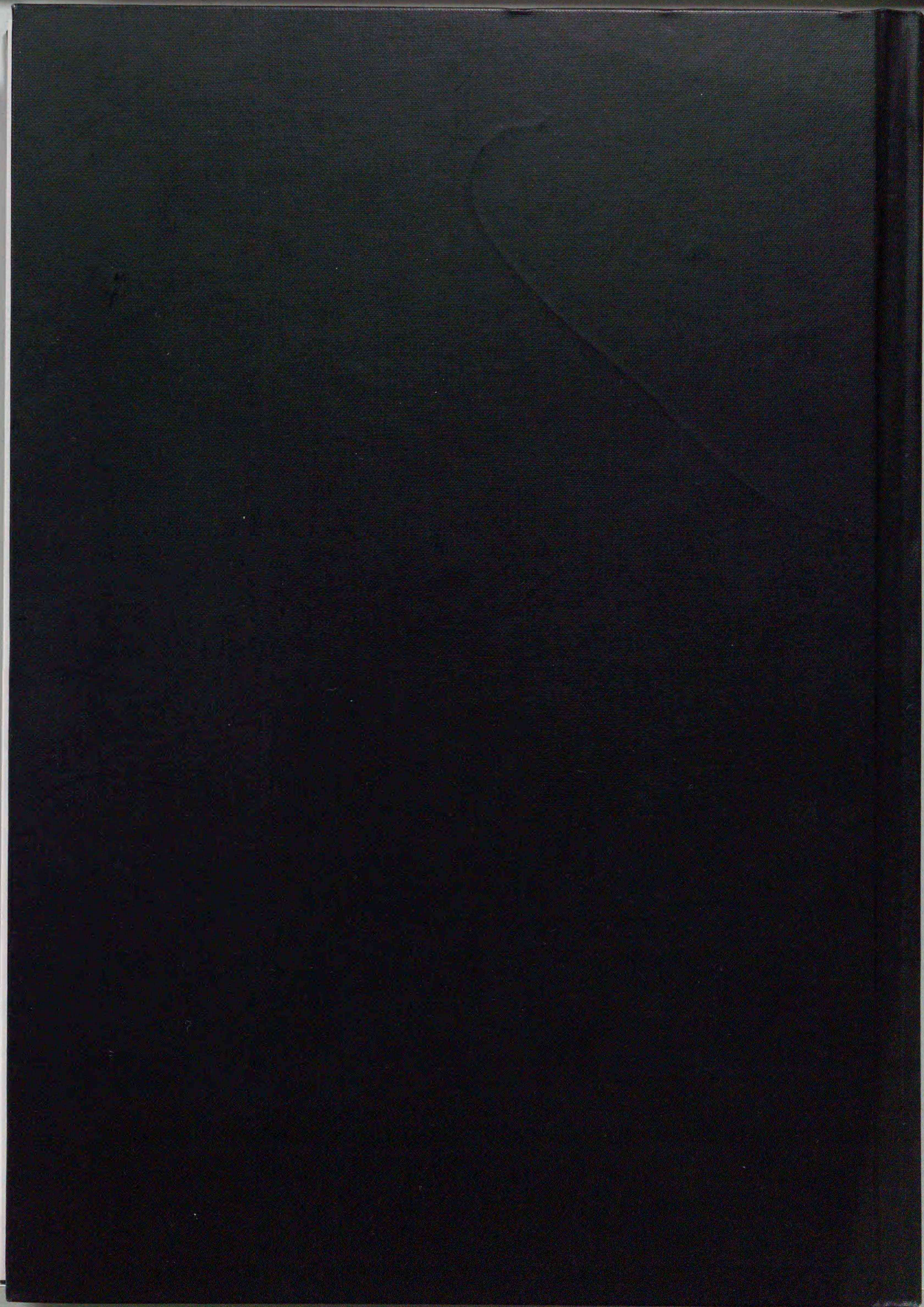
C 電子情報通信学会研究報告等

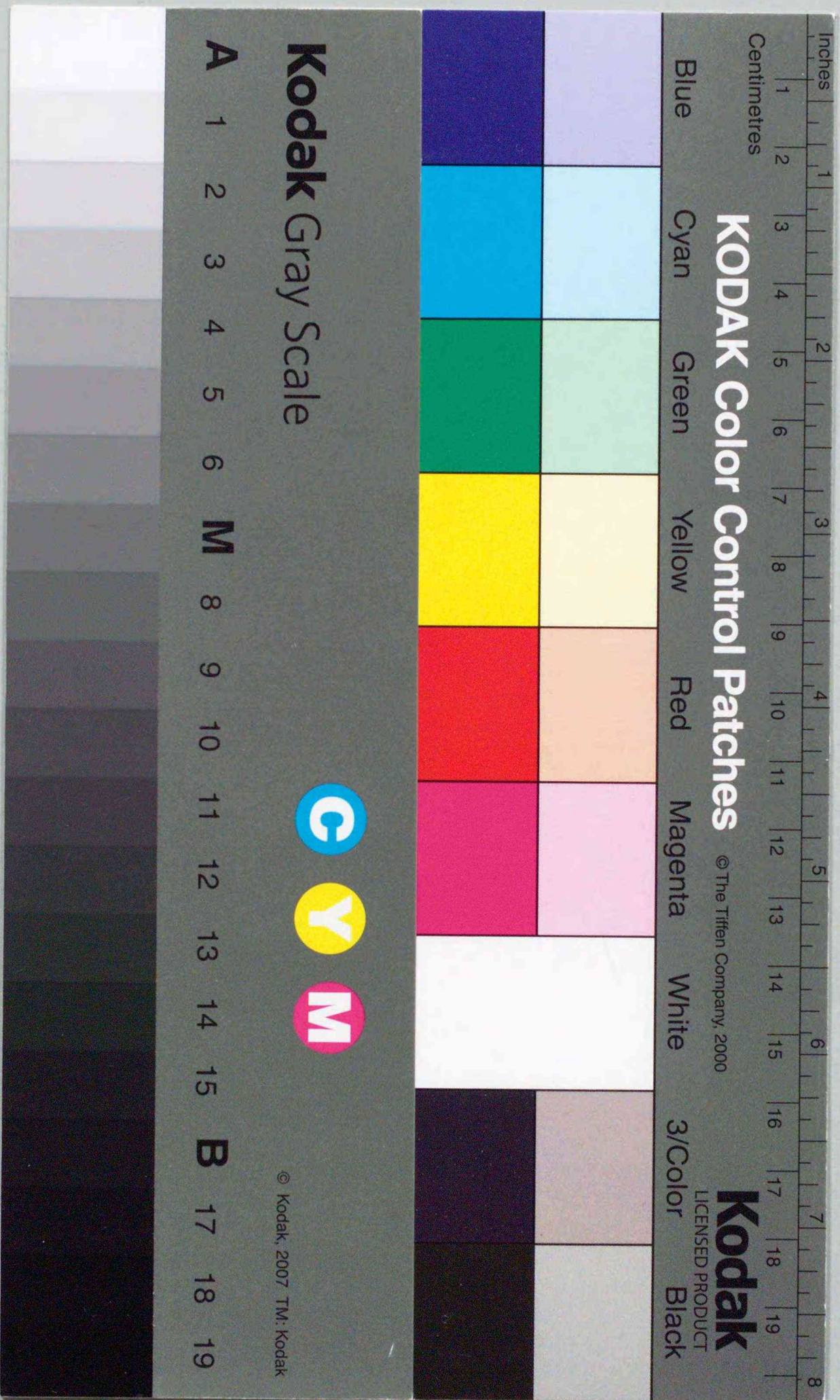
- (106) 広重真人、宮永喜一、栃内香次：“知識工学的手法を導入した適応信号処理”，第2回デジタル信号処理シンポジウム講演論文集，pp.337-342 (1987-12).
- (107) 広重真人、宮永喜一、栃内香次：“神経回路モデルを用いた高精度スペクトル選択法”，信学技報，**PRU88-52**，pp.31-37 (1988-09).
- (108) 宮永喜一、大槻典行、広重真人、栃内香次：“ARMA ラティスパラメータに基づく連続音声認識について”，信学技報，**SP88-118**，pp.15-22 (1989-01).
- (109) 広重真人、宮永喜一、栃内香次：“適応スペクトル推定手法を用いた CELP 符号化システムに関する検討”，信学技報，**DSP91-3**，pp.17-24 (1991-05).
- (110) 広重真人、宮永喜一、栃内香次：“アルゴリズム選択型適応信号処理スペクトル分析システム”，北海道大学工学部研究報告，146，pp.85-96 (1989-05).
- (111) 広重真人、宮永喜一、栃内香次：“音声符号化におけるコンパクト ARMA ラティスフィルタの有効性について”，北海道大学工学部研究報告，**153**，pp.91-102 (1990-11).

D 国内の学会発表

- (112) 広重真人、宮永喜一、栃内香次：“無声・有声音のスペクトル分析アルゴリズム”，昭 62 北海道連大，p.53 (1987-10).
- (113) 広重真人、宮永喜一、栃内香次：“非定常分析アルゴリズムによる確率信号と確定信号の分離について”，情処学第 35 回全大，pp.1949-1950 (1987-09).

- (114) 広重真人、宮永喜一、栃内香次：“経験的知識を用いた適応信号処理システムの構成に関する一考察”，昭 63 信学春全大，p.1=29 (1988-03).
- (115) 広重真人、宮永喜一、栃内香次：“ニューラルネットワークを導入した有声音のスペクトル選択法”，昭 63 北海道連大，p.155 (1988-10).
- (116) 広重真人、松下充宏、宮永喜一、栃内香次：“適応信号処理を用いた連続音声のセグメンテーション候補検出”，平元北海道連大，p.189 (1989-10).
- (117) 広重真人、宮永喜一、栃内香次：“ネストしたしきい値操作 if-then ルールにおけるしきい値の獲得手法”，平元北海道連大，p.322 (1989-09).
- (118) 広重真人、宮永喜一、栃内香次：“コンパクト ARMA ラティスフィルタの音声符号化への応用”，平 2 北海道連大，p.160 (1990-10).
- (119) 柴田克司、広重真人、宮永喜一、栃内香次：“適応音声分析を応用した連続音声認識システム”，平 2 北海道連大，p.163 (1990-10).
- (120) 広重真人、宮永喜一、栃内香次：“フォルマント情報を利用した低ビット音声符号化システム”，平 3 北海道連大，**158**，p.190 (1991-10).





Inches
Centimetres

KODAK Color Control Patches

© The Tiffen Company, 2000

Blue

Cyan

Green

Yellow

Red

Magenta

White

3/Color

Black

Kodak

LICENSED PRODUCT

Kodak Gray Scale



© Kodak, 2007 TM: Kodak

A

1

2

3

4

5

6

M

8

9

10

11

12

13

14

15

B

17

18

19