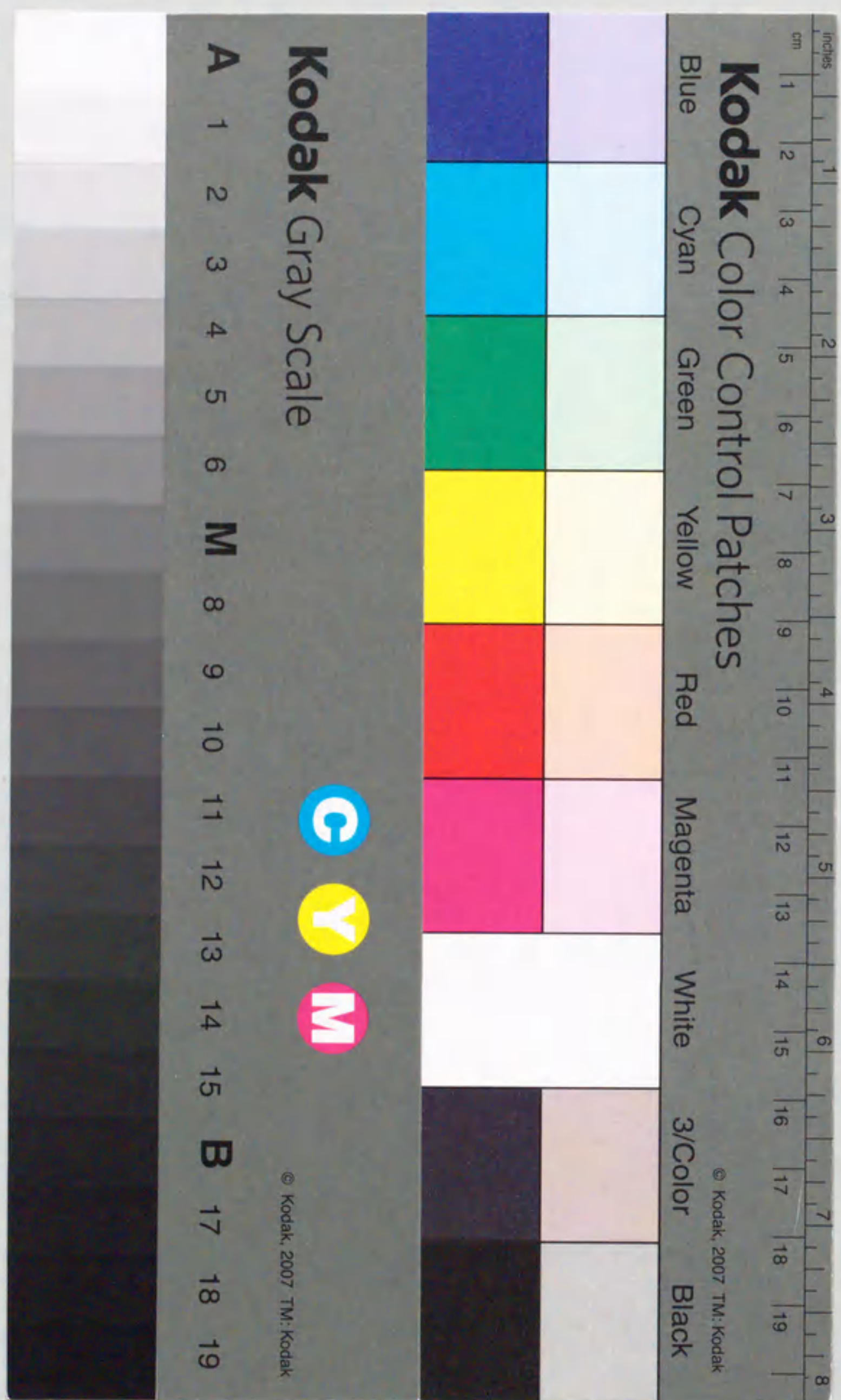


報告番号 乙第 5290 号

マルチパルスとベクトル量子化に基づく  
高能率音声符号化方式の研究

小澤 一範



①

マルチパルスとベクトル量子化に基づく  
高能率音声符号化方式の研究

小澤一範

## 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	本研究の背景	1
1.2	応用分野	2
1.3	従来の高能率音声符号化技術	3
1.3.1	音声符号化技術の分類	3
1.3.2	分析合成符号化	3
1.3.3	波形符号化	7
1.3.4	ハイブリッド符号化	9
1.4	音声符号化の標準化動向	15
1.5	音声符号化に対する要求条件	15
1.6	本研究の目的と課題	17
1.6.1	マルチパルス音源における課題	17
1.6.2	ベクトル量子化音源における課題	18
1.6.3	マルチパルスとベクトル量子化を結合した音源における課題	19
1.7	本論文の構成	19
1.8	本論文で提案する方式の伝送速度と音質の関係	21
<b>2</b>	<b>相関演算に基づくマルチパルスの逐次高速探索アルゴリズム</b>	<b>23</b>
2.1	まえがき	23
2.2	マルチパルス駆動形音声符号化法 (MPELP)	24
2.2.1	マルチパルス音源による音声合成モデル	24

2.2.2	マルチパルス探索アルゴリズム	24
2.3	4種のマルチパルス探索アルゴリズムの提案と比較	28
2.3.1	効率的なマルチパルス探索法(方式1)	28
2.3.2	パルス探索法の改良と4種のパルス探索法の比較	34
2.3.3	振幅調整法(方式2)	36
2.3.4	演算量低減型振幅調整法(方式2-1, 方式2-2)	37
2.4	実験	38
2.4.1	評価尺度	38
2.4.2	予備実験	40
2.4.3	パルス数と $SNR_{seg}$ の関係(方式1)	41
2.4.4	伝送情報の量子化	41
2.4.5	4種のパルス探索法の $SNR_{seg}$ 性能と演算量の比較	44
2.4.6	4種のパルス探索法の音質比較	46
2.4.7	パルス探索法2-2による波形例	46
2.4.8	考察	47
2.5	マルチパルス音声符号化装置の試作	50
2.6	むすび	53
3	ピッチ予測を用いるマルチパルス符号化	55
3.1	まえがき	55
3.2	ピッチ予測マルチパルス符号化方式	56
3.2.1	ピッチフィルタのインパルス応答をパルス探索に含めるフレーム内ピッチ予測法(P-1法)	57
3.2.2	フレーム内・フレーム間ピッチ予測法(P-2法)	60
3.2.3	サブフレーム間ピッチ予測法(P-3法)	65
3.3	評価実験	67
3.3.1	実験条件	67
3.3.2	客観評価	69

3.3.3	主観評価	71
3.3.4	考察	72
3.4	むすび	73
4	マルチモードとピッチ同期マルチパルスによる 4.8 kb/s 符号化	75
4.1	まえがき	75
4.2	提案方式の構成	76
4.2.1	モードと音源信号の種類	76
4.2.2	音響的特徴	77
4.2.3	ピッチ同期マルチパルス音源(PSMP)	78
4.2.4	マルチパルス音源(MP)、乱数コードブック音源(SC)	83
4.2.5	送信側、受信側の構成	83
4.3	実験	85
4.3.1	シミュレーション条件	85
4.3.2	客観評価	86
4.3.3	主観音質評価	87
4.3.4	考察	90
4.4	むすび	91
5	マルチモードとベクトル量子化を用いる CELP	93
5.1	まえがき	93
5.2	M-LCELP 音声符号化	94
5.2.1	マルチモード符号化	96
5.2.2	フィルタ係数の効率的なベクトル量子化	97
5.2.3	ピッチラグの差分符号化	102
5.2.4	スパース構造ベクトル量子化音源コードブックの開発と従来構造との比較	107
5.2.5	適応型コムフィルタ	112
5.3	実験	113

5.3.1	実験条件	113
5.3.2	誤り訂正の付加	114
5.3.3	主観音質評価	115
5.3.4	考察	115
5.4	DSPによる音声CODECの試作	117
5.5	むすび	119
<b>6</b>	<b>マルチパルスベクトル量子化音源と高速探索に基づくMP-CELP</b>	<b>121</b>
6.1	まえがき	121
6.2	MP-CELP音声符号化	124
6.2.1	モード判別	125
6.2.2	マルチパルスベクトル量子化音源	125
6.2.3	組合せ探索による振幅ベクトル量子化の性能改善	129
6.2.4	分割探索によるマルチパルス位置の高速探索	132
6.2.5	ゲイン量子化	135
6.2.6	LPC分析の次数	136
6.3	評価実験	137
6.3.1	実験条件	137
6.3.2	客観評価	139
6.3.3	予備音質評価	139
6.3.4	音質評価	141
6.3.5	考察	142
6.4	演算量	144
6.5	むすび	144
<b>7</b>	<b>マルチパルス符号化の符号誤りに対する評価</b>	<b>147</b>
7.1	まえがき	147
7.2	マルチパルス符号化法	148
7.2.1	アルゴリズム	148

7.2.2	パルス振幅の簡易調整	148
7.2.3	分析条件とビット割当	148
7.3	実験	149
7.3.1	伝送符号誤り	149
7.3.2	客観、主観評価尺度	150
7.3.3	ランダム誤りに対する性能評価	151
7.3.4	バースト誤りに対する性能評価	156
7.4	むすび	160
<b>8</b>	<b>高能率音声符号化装置の実用化例</b>	<b>163</b>
8.1	まえがき	163
8.2	企業内通信向けマルチパルス音声符号化装置	163
8.3	ネットワーク向けマルチメディア多重化装置	167
8.4	移動体通信向け音声コーデックチップ	168
8.5	むすび	172
<b>9</b>	<b>結論と今後の課題</b>	<b>173</b>
9.1	結論	173
9.2	本研究の意義と貢献	177
9.3	今後の研究課題	178
	謝辞	181
	参考文献	183
	著者業績	201
	略語一覧	207

## 図目次

1.1	音声符号化技術の分類 . . . . .	4
1.2	伝送速度と音質の大まかな関係 . . . . .	5
1.3	線形予測分析に基づくボコーダの音声合成モデル . . . . .	6
1.4	ADPCM 符号化の原理 . . . . .	8
1.5	A-b-S (合成による分析) に基づく音声符号化の原理 . . . . .	10
1.6	聴感重み付けによる周波数特性の例 . . . . .	12
1.7	マルチパルス符号化の原理 . . . . .	13
1.8	CELP (Code Excited LPC) 符号化の原理 . . . . .	14
1.9	本論文で提案する種々の方式の伝送速度と音質の大まかな関係 . . . . .	22
2.1	マルチパルス音源を用いる音声合成モデル . . . . .	25
2.2	マルチパルス音声符号化のブロック図 . . . . .	26
2.3	A-b-S (合成による分析) に基づくマルチパルスの探索 . . . . .	27
2.4	パルス探索の波形例, (a) $R(m)$ , (b) 第 1 パルス, (c) 第 1 パルスによる $R(m)$ の修正, (d) 第 2 パルス, (e) 第 2 パルスによる $R(m)$ の修正, (f) 第 3 パルス, (g) 第 3 パルスによる $R(m)$ の修正, (h) 第 4 パルス . . . . .	33
2.5	パルス探索部の構成 . . . . .	34
2.6	フィルタの次数と $SNR_{seg}$ , $CD$ との関係 . . . . .	42
2.7	フレーム長と $SNR_{seg}$ , $CD$ との関係 . . . . .	42
2.8	パルスの個数と $SNR_{seg}$ との関係 (方式 1) . . . . .	43
2.9	4 種のパルス探索法の $SNR_{seg}$ と演算量の比較 . . . . .	45
2.10	4 種のパルス探索法の音質比較 . . . . .	47

2.11	方式 2-2 による 16 kb/s の波形の例, (a) 原音声波形 (女性話者 200 ms), (b) 再生波形, (c) 誤差波形。各図の縦軸は (a) の最大値、最小値で正規化されている。 . . . . .	48
2.12	方式 2-2 による 9.6 kb/s の波形の例, (a) 原音声波形 (女性話者 200 ms), (b) 再生波形, (c) 誤差波形。各図の縦軸は (a) の最大値、最小値で正規化されている。 . . . . .	49
2.13	試作した 16 kb/s, 9.6 kb/s マルチパルス音声符号化ボードの外観 . . . . .	52
3.1	有声区間での波形例, (a) 有声区間のピッチ周期波形, (b)(a) の区間で求めたマルチパルス . . . . .	56
3.2	ピッチ予測マルチパルス符号化 (P-1 法) を示すブロック図 . . . . .	57
3.3	ピッチ予測マルチパルス符号化 (P-1 法) による処理の波形例, (a) フレーム (20 ms) の音声波形, (b) 式 (3.4) から求めた縦続接続合成フィルタのインパルス応答, (c) 式 (3.2) を最小化するように求めたマルチパルス, (d)(c) とピッチ再生フィルタ (LTP) から求めた励振音源, (e) 合成音声。ただし, 図の縦軸は各図毎に正規化されている。 . . . . .	61
3.4	ピッチ予測マルチパルス符号化 (P-1 法) の送信側、受信側の構成を示すブロック図 . . . . .	62
3.5	ピッチ予測マルチパルス符号化 (P-2 法) のブロック図 . . . . .	64
3.6	STP を共通化し LTP のみ 2 種類用いる P-2 法のブロック図 . . . . .	64
3.7	ピッチ予測マルチパルス符号化 (P-3 法) のブロック図 . . . . .	66
3.8	9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化の $SNR_{seg}$ 特性 . . . . .	70
3.9	9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化の音質評価結果 . . . . .	72
4.1	ピッチ同期マルチパルス音源 (PSMP) を用いる音声合成モデル . . . . .	79
4.2	ピッチ同期マルチパルス音源 (PSMP) の処理例, (a) フレームの音声波形, (b) 代表区間の候補区間で求めたマルチパルス (4 パルス), (c) 代表区間の一候補区間のマルチパルスと位相, 振幅補正係数を用いてフレーム全体で復元した音源信号, (d)(c) の音源信号により合成した波形 . . . . .	82

4.3	提案方式の送信側、受信側のブロック図 . . . . .	84
4.4	提案法と 2 モード PI 法の S/N の時間変化の例, (a) は男性話者, (b) は女性話者。各図の上段はパワー, 下段は S/N (dB) を示す。各図の横軸は時間を示し、全長は 2 秒である。 . . . . .	88
4.5	音質一対評価結果 . . . . .	89
5.1	M-LCELP のブロック図, (a) 送信側, (b) 受信側 . . . . .	95
5.2	LSP の補間込み 2 段分割ベクトル量子化のブロック図 . . . . .	98
5.3	適応コードブックピッチラグの時間変化の例, (a) 開ループ-閉ループ併用法 (従来法 1) のラグ変化, (b) 提案法のラグ変化。(a),(b) 共に、縦軸はラグの値, 横軸は時間を示す。 . . . . .	106
5.4	スパースコードブックの学習による Weighted SNR の向上 . . . . .	110
5.5	スパース音源コードブックの波形例, (a) は初期乱数コードベクトル, (b) は 30 回学習後のコードベクトル (パルスの個数は 8 である)。いずれも 5 ベクトル分を示す。 . . . . .	111
5.6	試作した M-LCELP 音声符号化装置 . . . . .	118
6.1	MP-CELP エンコーダのブロック図 . . . . .	124
6.2	マルチパルスベクトル量子化音源を用いるエンコーダの原理図 . . . . .	127
6.3	位置の候補数と $SNR_{seg}$ の関係 . . . . .	130
6.4	パルス位置の分割探索 (Q=2) . . . . .	134
7.1	パルス振幅の簡易調整による $SNR_{seg}$ の改善 . . . . .	149
7.2	16 kb/s, 9.6 kb/s マルチパルス符号化のランダム誤りに対する $SNR_{seg}$ 特性 . . . . .	152
7.3	16 kb/s マルチパルス符号化において各伝送パラメータにランダム誤りを発生させたときの再生音声の $SNR_{seg}$ . . . . .	153
7.4	16 kb/s マルチパルス符号化において各伝送パラメータにランダム誤りを発生させたときの再生音声のケプストラム距離 $CD$ . . . . .	153

7.5	3.2 kb/s の誤り訂正を付加した 16 kb/s マルチパルス符号化のランダム誤りに対する $SNR_{seg}$ 特性 . . . . .	155
7.6	16 kb/s, 9.6 kb/s マルチパルス符号化のランダム誤りに対する音質 . . . . .	155
7.7	16 kb/s, 9.6 kb/s マルチパルス符号化のバースト誤りに対する $SNR_{seg}$ 特性 . . . . .	158
7.8	16 kb/s, 9.6 kb/s マルチパルス符号化のバースト誤りに対する音質 . . . . .	159
8.1	企業内通信向け音声符号化装置の接続の例, LS: Line Switch, TLS: Toll Line Switch, MMM: MultiMedia Mux . . . . .	164
8.2	ユニバーサルトランスコードの外観 . . . . .	165
8.3	ユニバーサルトランスコードの音声符号化パッケージ . . . . .	166
8.4	マルチメディア多重化装置によるネットワーク構築例 . . . . .	168
8.5	9.6 kb/s, 8 kb/s LCELP 音声符号化ボード . . . . .	169
8.6	移動通信用 6.4 kb/s 音声コーデックチップの外観 . . . . .	170

## 表目次

1.1	音声符号化の伝送速度と応用分野の関係 . . . . .	2
1.2	ITU-T での音声符号化方式の標準化動向と応用分野 . . . . .	16
1.3	携帯電話向けの主要な音声符号化方式の標準化動向と応用分野 . . . . .	16
2.1	4 種のパルス探索法の比較 . . . . .	35
2.2	4 種のパルス探索法の演算量の比較 . . . . .	36
2.3	5 段階品質評価尺度 . . . . .	40
2.4	伝送パラメータの量子化ビット配分 . . . . .	44
2.5	マルチパルス符号化装置に使用した DSP ( $\mu$ PD7720) の特徴 . . . . .	51
2.6	マルチパルス符号化装置における DSP の処理分担 . . . . .	51
3.1	9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化 (P-1,P-2,P-3) の符号化条件 . . . . .	68
3.2	一対評価 (i, j) における 5 段階評定尺度 . . . . .	71
4.1	モードと音源信号の分類 . . . . .	76
4.2	4.8 kb/s ビット配分 (ビット / フレーム) . . . . .	85
4.3	客観評価結果 . . . . .	86
4.4	音質一対評価 (i, j) に用いた 5 段階評定尺度 . . . . .	87
5.1	LSP 量子化ビット配分 / 40 ms フレーム . . . . .	101
5.2	対数スペクトル歪性能の比較 . . . . .	101
5.3	ラグ探索の性能比較 . . . . .	105
5.4	ラグ予備選択の性能 . . . . .	105



5.5	コードブックの構造による $SNR_{seg}$ 性能、処理量、メモリ量の比較 . . .	112
5.6	4.1 kb/s M-LCELP のビット配分 . . . . .	114
5.7	MOS 音質評価結果 . . . . .	116
5.8	試作した M-LCELP 音声 CODEC 装置の処理量、メモリ量 . . . . .	117
6.1	位置の候補数と振幅コードブックの探索演算量 . . . . .	132
6.2	5パルス /5 ms のときの各パルスのとりうる位置の例 . . . . .	133
6.3	位置の分割探索における分割数と $SNR_{seg}$ 性能、演算量の関係 . . . . .	135
6.4	ゲイン量子化時の候補数と $SNR_{seg}$ 性能 . . . . .	136
6.5	LPC 分析の次数と性能の関係 . . . . .	136
6.6	11 kb/s MP-CELP のビット配分 (ビット /10 ms フレーム) . . . . .	138
6.7	母音区間, 子音区間のパルスのとりうる位置 . . . . .	138
6.8	11 kb/s MP-CELP と 11 kb/s M-LCELP の $SNR_{seg}$ 比較 . . . . .	139
6.9	10.7 kb/s MP-CELP と 11 kb/s M-LCELP の音質比較 . . . . .	140
6.10	5 段階聴取努力評価尺度 (LES ACR) . . . . .	142
6.11	音質評価結果 . . . . .	143
6.12	MP-CELP エンコーダの演算量, メモリ量 . . . . .	144
7.1	16 kb/s, 9.6 kb/s マルチパルス符号化の分析条件とビット配分 . . . . .	150
8.1	マルチパルス音声符号化装置ファミリー . . . . .	167
8.2	マルチメディア多重化装置向け 9.6 kb/s, 8 kb/s LCELP 音声符号化諸元	167
8.3	NEC DSP $\mu$ PD77018 の主要諸元 . . . . .	171
8.4	移動体通信用 6.4 kb/s 音声コーデックの諸元 . . . . .	171
8.5	移動体通信用 6.4 kb/s 音声コーデックチップのメモリ使用量 . . . . .	171

## 第 1 章

### 序論

#### 1.1 本研究の背景

音声通信はコミュニケーションの基本的な手段である。人間の発声した音声信号をデジタル的に圧縮、符号化して伝送あるいは蓄積し、受信側で復号する高能率音声符号化方式は、通信のデジタル化の進展に伴い、研究開発がこれまでに盛んに行なわれてきた。近年は実用化レベルに到達しており、無線通信では、最近、加入者数が急増しているデジタル携帯電話・自動車電話、PHS などで中核技術の一つになっている。さらに、業務用デジタル移動体通信や、デジタル移動体衛星通信にも浸透しつつある。有線通信では、デジタル基幹伝送、企業内通信、テレビ会議、モデムによるパソコン通信などに使用されており、最近ではインターネットへの応用も増えている。これらの分野では、移動体通信における電波の有効利用のために、あるいは、通信のマルチメディア化において音声と他のマルチメディア情報を同時に伝送するために、音声の高能率符号化の重要性は今後もますます増大してくると予想される。一方で、信号処理 LSI 技術の急速な進展に伴い、これまでより比較的複雑な処理を有する方式もコンパクトに装置化することが可能になってきており、より低い伝送速度で、より良好な音質を提供できる音声符号化技術の研究開発が加速されている。

表 1.1: 音声符号化の伝送速度と応用分野の関係

伝送速度 (kb/s)	応用分野
32 ~ 16	基幹伝送、パーソナル通信 (PHS)
16 ~ 8	企業内通信、移動体通信 (フルレート)、FPLMTS
8 ~ 4	移動体通信 (フル・ハーフレート)、PC 通信、FPLMTS、携帯テレビ電話
4 ~ 2	移動体通信 (ハーフレート)、移動体衛星通信、音声応答、無線専用網

PHS: Personal Handyphone System

FPLMTS: Future Public Land Mobile Telecommunication Systems

PC: Personal Computer

## 1.2 応用分野

本論文では、伝送速度が 32 kb/s 以下の音声符号化を高エネルギー音声符号化と呼ぶ。高エネルギー音声符号化の伝送速度と応用分野の関係を表 1.1 にまとめる [138], [141]。

32 ~ 16 kb/s の音声符号化は、デジタル基幹伝送や、PHS (Personal Handyphone System) などのパーソナル通信に適用されている。16 ~ 8 kb/s では、特に、8kb/s 音声符号化がデジタル携帯電話などのフルレート移動体通信や、企業内通信での重要技術となっている。また、16 kb/s、8 kb/s は、将来的には、次世代移動体通信システム FPLMTS (Future Public Land Mobile Telecommunication Systems) などでの採用が期待されている。8 ~ 4 kb/s は、デジタル携帯電話 (フルレート、ハーフレート) や、モデムを使用したパーソナルコンピュータ (PC) 間の通信などへの応用で、現在、需要が最も急速に伸びている。将来的には、FPLMTS や携帯テレビ電話などへの応用が予想される。4 kb/s 以下は、ハーフレート移動体通信、移動体衛星通信や音声応答への応用がある。また、無線専用網などでは、できる限り伝送速度の低い方式が望まれている。

## 1.3 従来の高エネルギー音声符号化技術

### 1.3.1 音声符号化技術の分類

音声符号化技術は、波形符号化 (Waveform coding)、分析合成符号化 (Analysis-synthesis coding)、ハイブリッド符号化 (Hybrid coding) の 3 種のカテゴリに分類される。これを図 1.1 に示す。各カテゴリの符号化の特徴は 1.3.2 節から 1.3.4 節で述べる。カテゴリ毎に伝送速度 (ビットレート) と音質の大まかな関係を図 1.2 に示す。波形符号化は、32 kb/s の音質は極めて良好であるが、16 kb/s 以下に下げると音質が劣化してくる。代表的な方式は、PCM、ADPCM (Adaptive Differential PCM) や APC (Adaptive Predictive Coding) である。分析合成符号化は、2 kb/s 以下まで伝送速度を低減化できるが、逆に伝送速度を上げても音質は飽和する。代表的な方式はボコーダ (Vocoder) である。ハイブリッド符号化は、波形符号化と分析合成符号化の長所を組み合わせ、16 kb/s ~ 4 kb/s で音質の大幅な改善を狙った方式である。代表的な方式には、マルチパルス符号化 (MPELP: Multi-Pulse Excited Linear Prediction) や、CELP (Code Excited Linear Prediction) などがある。本論文で研究する符号化は、ハイブリッド符号化に属する。

### 1.3.2 分析合成符号化

分析合成符号化の研究は、1939 年の Bell 研究所の H.Dudley らによるボコーダ、ボコーダ [21] の研究にはじまると考えられる。これは、音声を合成モデルにあてはめ、モデルのパラメータを抽出して音声を合成する。当初は機械による音声合成が主な目的であったが、その後、符号化伝送も考慮するようになり、ボコーダの改良が続けられた [27], [150]。

特に、1.3.3 節で述べる線形予測分析手法を用いる最尤ボコーダ [41]、PARCOR (Partial Autocorrelation) ボコーダ [43]、共分散ボコーダ [5]、LSP (Line Spectrum Pair) ボコーダ [160] などの研究により、性能が改善された。線形予測分析法に基づくボコーダにおける音声合成モデルを図 1.3 に示す。このモデルでは、音源信号を有声区間と無声区間で切り替える。有声区間では、ピッチ周期に等しい時間間隔を有する

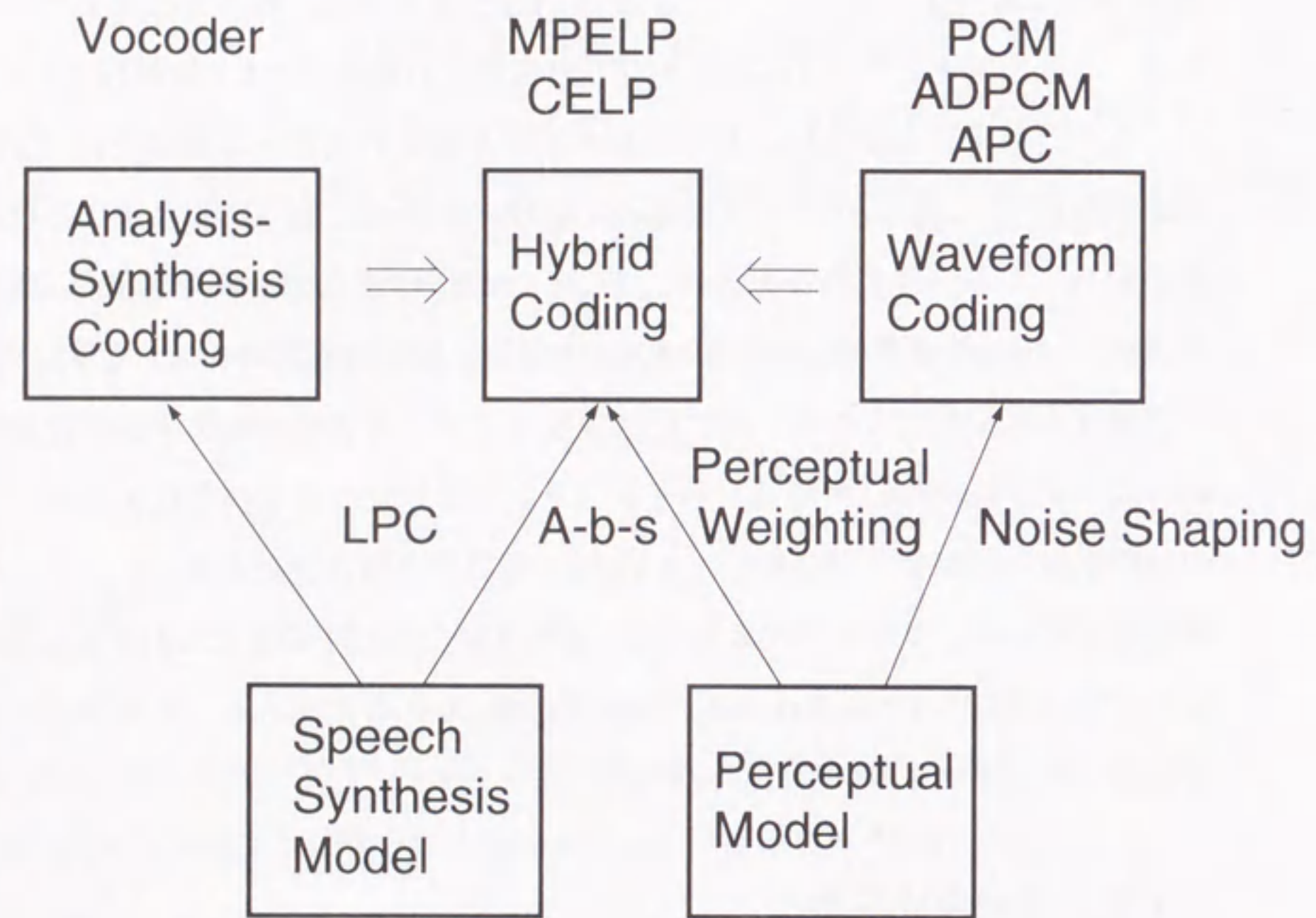


図 1.1: 音声符号化技術の分類

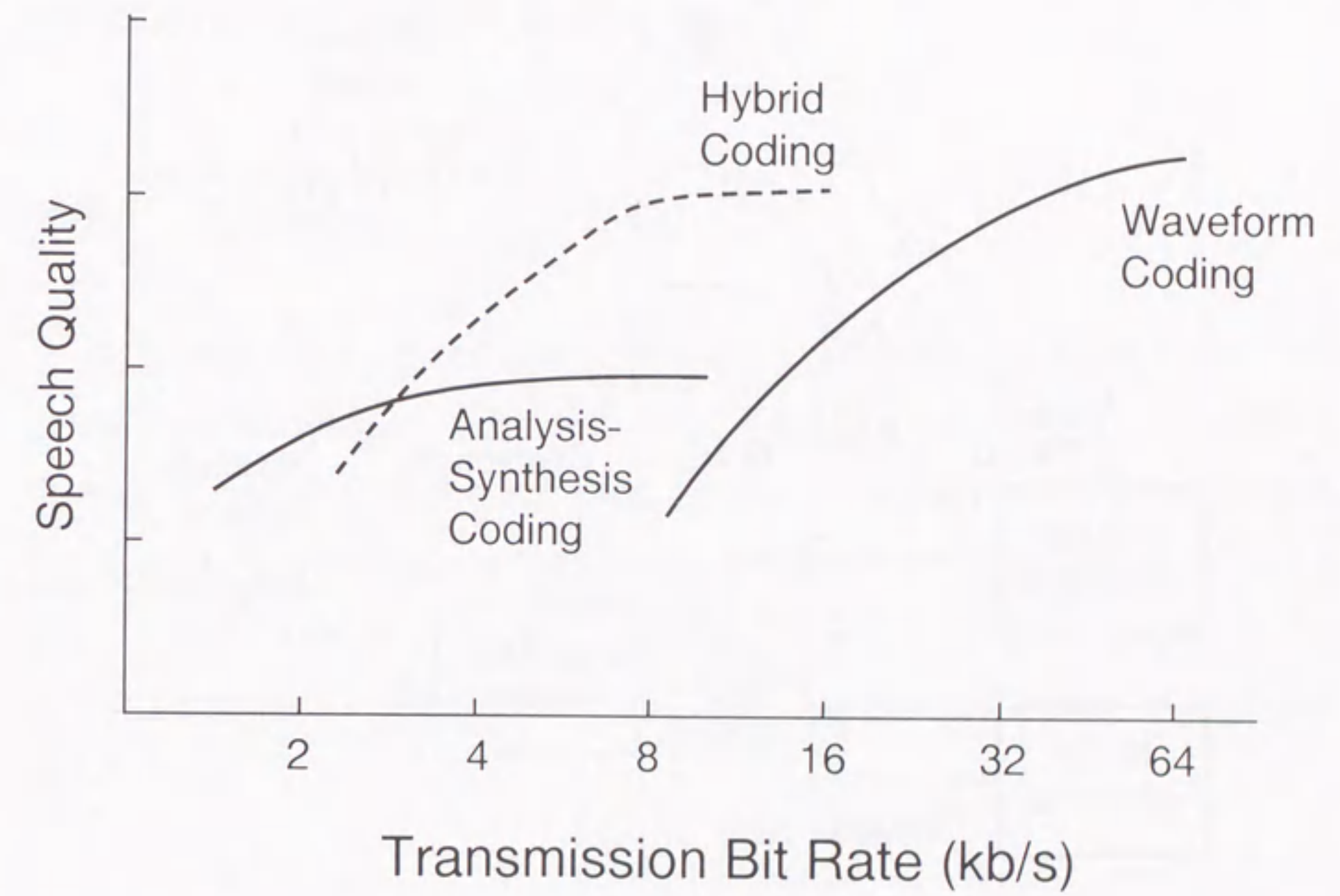


図 1.2: 伝送速度と音質の大まかな関係

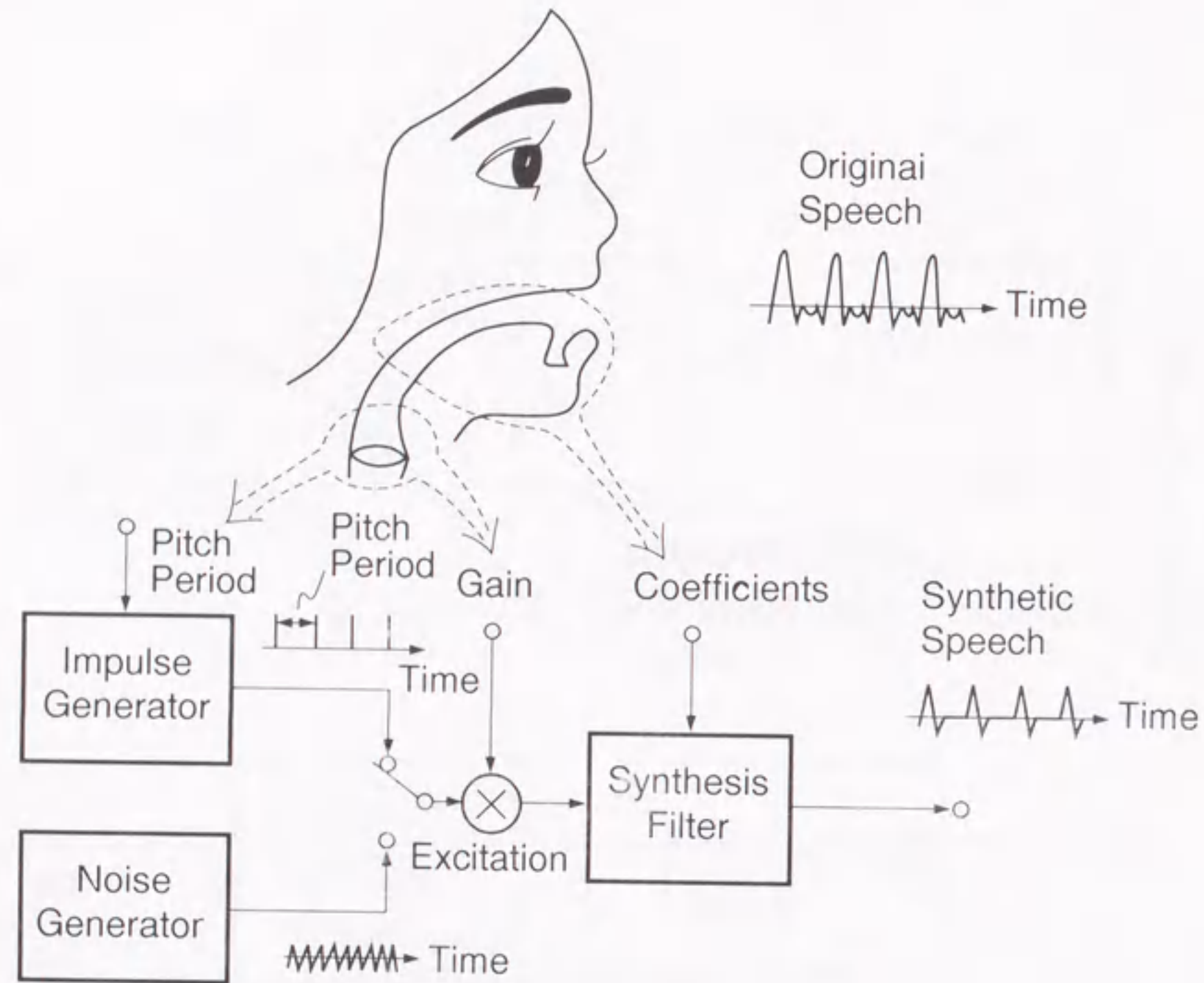


図 1.3: 線形予測分析に基づくボコーダの音声合成モデル

振幅一定の周期的なインパルス列を用いる。無声区間では、乱数音源が使用される。この合成モデルはパラメータが少なく、伝送情報量が少ないため、音質的に十分とはいえないものの、伝送速度を2 kb/s以下まで低減化することができる。しかしながら、モデルを単純化しすぎているために、逆にビットレートを高くしても、高品質の音声を合成するには限界があった。また、有声/無声判別の誤り、ピッチ抽出の誤りなどに起因して音質が劣化することがあった。

### 1.3.3 波形符号化

波形符号化方式は、音声波形をできるだけ忠実に再現することを目的とする。PCM (Pulse Code Modulation) 方式が1930年代に考案され、現在広く普及している瞬時圧伸PCMが1950年代に開発された[159]。さらに、予測を加えた差分符号化(DPCM: Differential PCM)が提案され[78]、予測による波形符号化の基本的な枠組が確立された。この後、DPCMをベースに、予測器及び量子化器に逐次適応化処理を加えることで性能を改善した適応差分符号化(ADPCM: Adaptive Differential PCM)が提案された[54]。

### ADPCM

ADPCMの原理を図1.4に示す。予測器において入力信号をサンプル毎に適応予測し、予測残差をサンプル毎に適応量子化して伝送することで効率を高めている。ADPCMは、精力的に改良が行なわれ[98]、1984年に、CCITT(国際電信電話諮問委員会)において、G.721(32 kb/s ADPCM)として国際標準化された(その後、32 kb/s ADPCMはG.726に吸収されG.721は欠番となった[14])。32 kb/s ADPCMは、音質が良好で通話上問題がないこと、有線電話網などですでに広く使用されていることから、高能率音声符号化における音質の基準として、近年の符号化方式の国際標準化などでは、達成すべき音質目標に挙げられている。

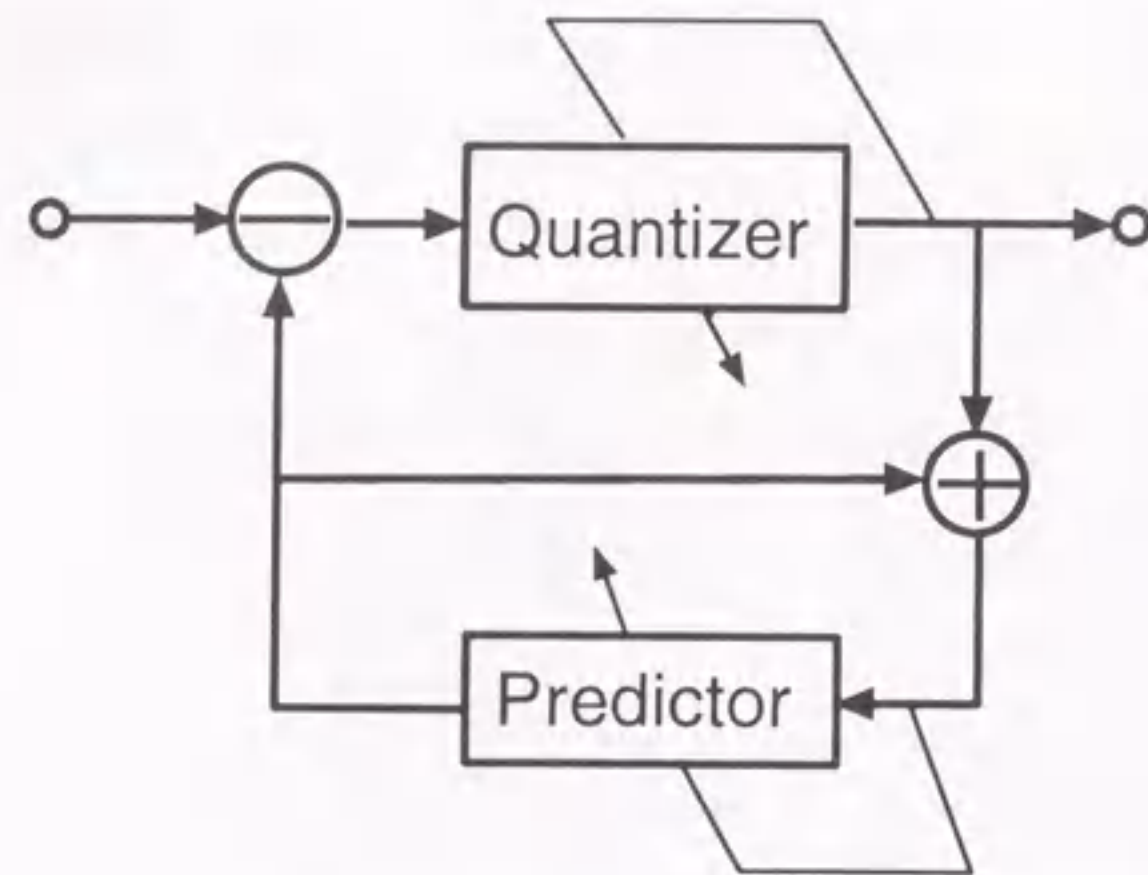


図 1.4: ADPCM 符号化の原理

#### 線形予測分析にもとづく線形予測符号化

音声分析の分野では、線形予測分析手法の提案が、音声のデジタル処理に画期的な進展をもたらした。1968年に線形予測分析(LPC: Linear Predictive Coding) [3], [41], [42], [5], その後、偏自己相関分析(PARCOR: Partial Autocorrelation) [44]、続いて線スペクトル対分析(LSP: Line Spectrum Pair) [45]が提案され、音声の声道伝達関数のパラメトリックモデルによるデジタル表現を可能にした。

線形予測分析法は、波形符号化における予測器の係数計算に適用され、以後、線形予測分析に基づく線形予測符号化法が主流となった。1968年に適応予測や適応ピッチ予測を用いる適応予測符号化(APC: Adaptive Predictive Coding)が開発された[3], [4]。APCでは、音声信号をほぼ定常と考えられる時間長(10-20 ms)のフレームに区切り、フレーム毎に線形予測分析により予測器の係数を分析し予測信号を計算し、これを入力音声から減算し予測残差信号を求める。さらに、予測残差信号から音声の長期相関に基づきピッチ予測を施し、ピッチ予測残差を求め、これを適応量子化して伝送する。APCではその後、聴感上のS/Nを改善するために、送信側で量子化雑音の周波数特性を制御するノイズシェイピング法[6]が組み込まれ、音質が向上した。

APCの開発により、線形予測分析に基づく適応予測符号化技術、ピッチ予測技術、

#### 1.3. 従来の高能率音声符号化技術

ノイズシェイピング技術が確立された。伝送速度的には、16 kb/s ~ 9.6 kb/sで有効であるが、これよりも低い伝送速度では音質が劣化する。

APCに帯域分割と、時間・周波数方向の適応ビット割り当てを導入して音質を改良した方式として、適応ビット割り当て適応予測符号化(APC-AB: APC with Adaptive Bit Allocation) [36],[37]などが提案されている。

#### 1.3.4 ハイブリッド符号化

ハイブリッド符号化は、分析合成符号化と波形符号化(特に適応予測符号化)の両者の利点を組み合わせた方式と考えることができる。つまり、音声の合成モデルを用いているが、合成モデルなどのパラメータは、入力波形と合成波形との波形上の歪みを最小化するように抽出する。線形予測分析に基づく代表的な符号化方式としては、マルチパルス駆動形線形予測符号化(MPELP: Multi-Pulse Excited Linear Prediction) [7]と、符号励振型線形予測符号化(CELP: Code Excited Linear Prediction) [152], [9]があり、16 kb/s ~ 4 kb/sでは、近年の音声符号化の主流となっている。

これらの方式は、音源信号の表現が異なる以外は、基本的には同一の構成である。共通の主要要素技術は、(i)音源と合成フィルタから構成される音声合成モデルを有し、合成フィルタの係数を線形予測分析により計算、(ii)合成による分析手法(A-b-S: Analysis-by-Synthesis)を用いて波形歪み最小化基準による音源パラメータの決定[7]、(iii)聴感上のS/Nを向上させるための聴感重み付け技術[6], [7]、などである。以下では、このような特徴を有する符号化方式をA-b-Sに基づく音声符号化と呼ぶ。

#### A-b-S (Analysis-by-Synthesis) に基づく音声符号化

A-b-Sに基づく音声符号化の原理を図1.5に示す。合成モデルのパラメータ(特に音源信号のパラメータ)は、入力信号と合成信号との聴感重み付け波形歪みを最小にするように決定される[7]。合成フィルタ、聴感重み付けフィルタの係数は、線形予測分析から計算される。

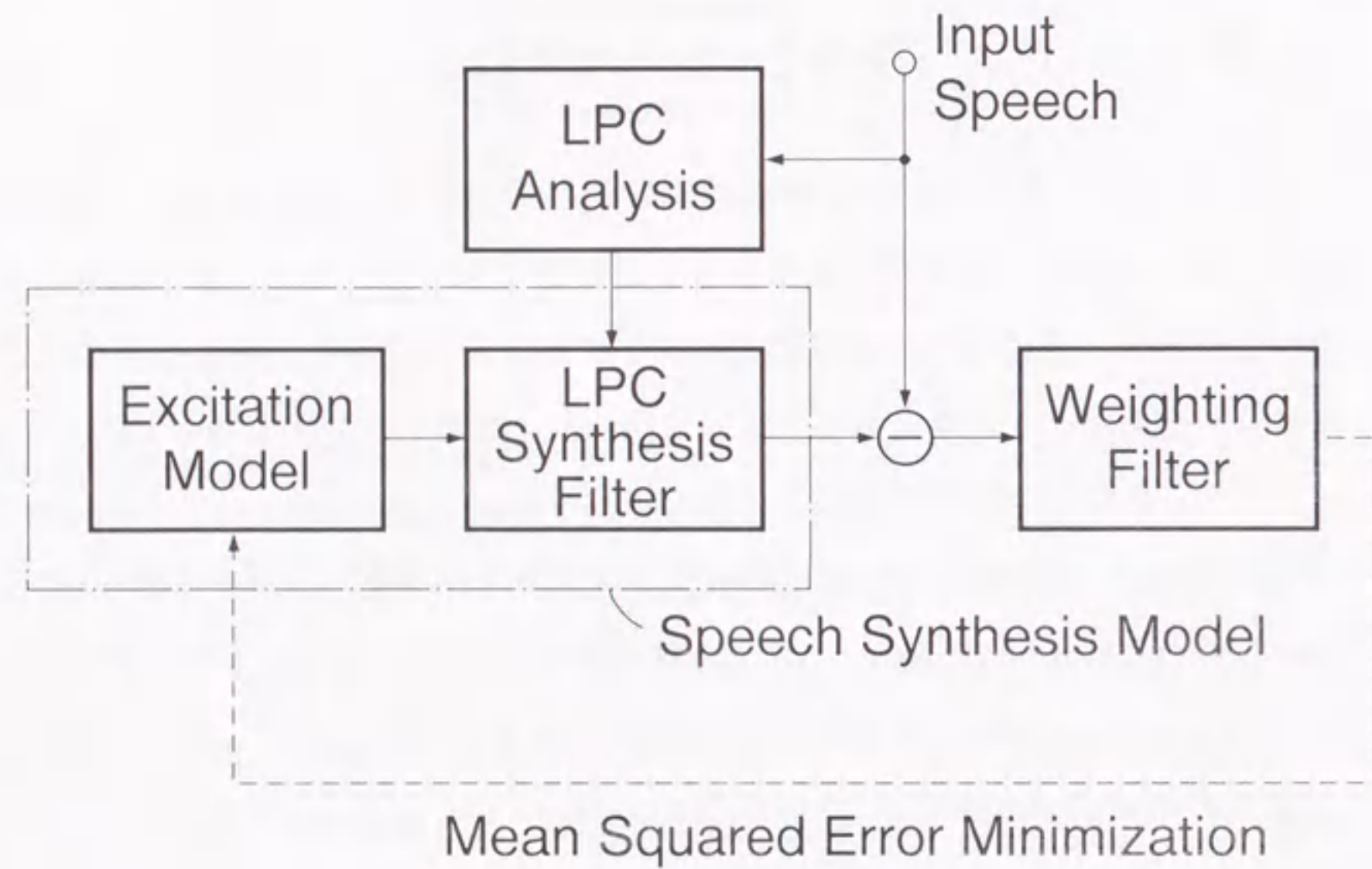


図 1.5: A-b-S (合成による分析) に基づく音声符号化の原理

### 聴感重み付けによる音質の改善

聴感重み付けは、聴感上の S/N を改善し音質を向上させる [6], [151], [7], [10]。A-b-S 探索過程において、聴感重み付けフィルタを通した後で 2 乗誤差を計算し、これを最小化するようにパラメータを決定する。聴感重み付けフィルタの伝達関数は、 $z$  変換上で式 (1.1) のように表せる。

$$W(z) = \frac{1 - \sum_{i=1}^P \alpha_i \gamma_1^i z^{-i}}{1 - \sum_{i=1}^P \alpha_i \gamma_2^i z^{-i}} \quad (1.1)$$

ここで、 $\alpha_i$  は、フレーム毎に線形予測分析から求めた  $i$  番目の線形予測係数、 $P$  は線形予測分析の次数である。 $\gamma_1$ 、 $\gamma_2$  は、聴感重み付け量を制御する定数であり、 $0 < \gamma_2 < \gamma_1 \leq 1$  に選ぶ。

母音区間フレームでの音声信号のスペクトルと聴感重み付けによる周波数特性の一例を図 1.6 に示す。聴感重み付けフィルタは、音声のピークの部分 (フォルマント) を抑圧し、フォルマント間の谷の部分を増幅するような周波数特性を有している (図 1.6 の破線部)。このような重み付け処理を施した上で音源等のパラメータを探索することにより、最終的に合成信号に重畳するモデル化誤差信号の短時間スペクトルを図の一点鎖線で示すように制御することができる。このようにするのは、人間の聴覚のマスク作用により、フォルマント部分の誤差は入力音声にマスクされて耳につきにくい、フォルマント間や高域の誤差は、入力音声にマスクされにくく、耳につきやすいためである。

### 音源をパルスの組合せで表すマルチパルス符号化

線形予測符号化において、予測残差信号を振幅と位置の異なる少ない個数のパルスの組合せ (マルチパルス) で表現することにより、伝送速度の低減化が可能となる。なぜなら、予測残差信号を時間的に不均一に間引くことに相当するため、予測残差信号を送信するためのビットレートを低減化することが可能になるからである。しかし、予測残差信号のサンプルを単純に不均一に間引いたのでは、特にビットレートが低い場合に、良好な音質を確保することが困難となる。

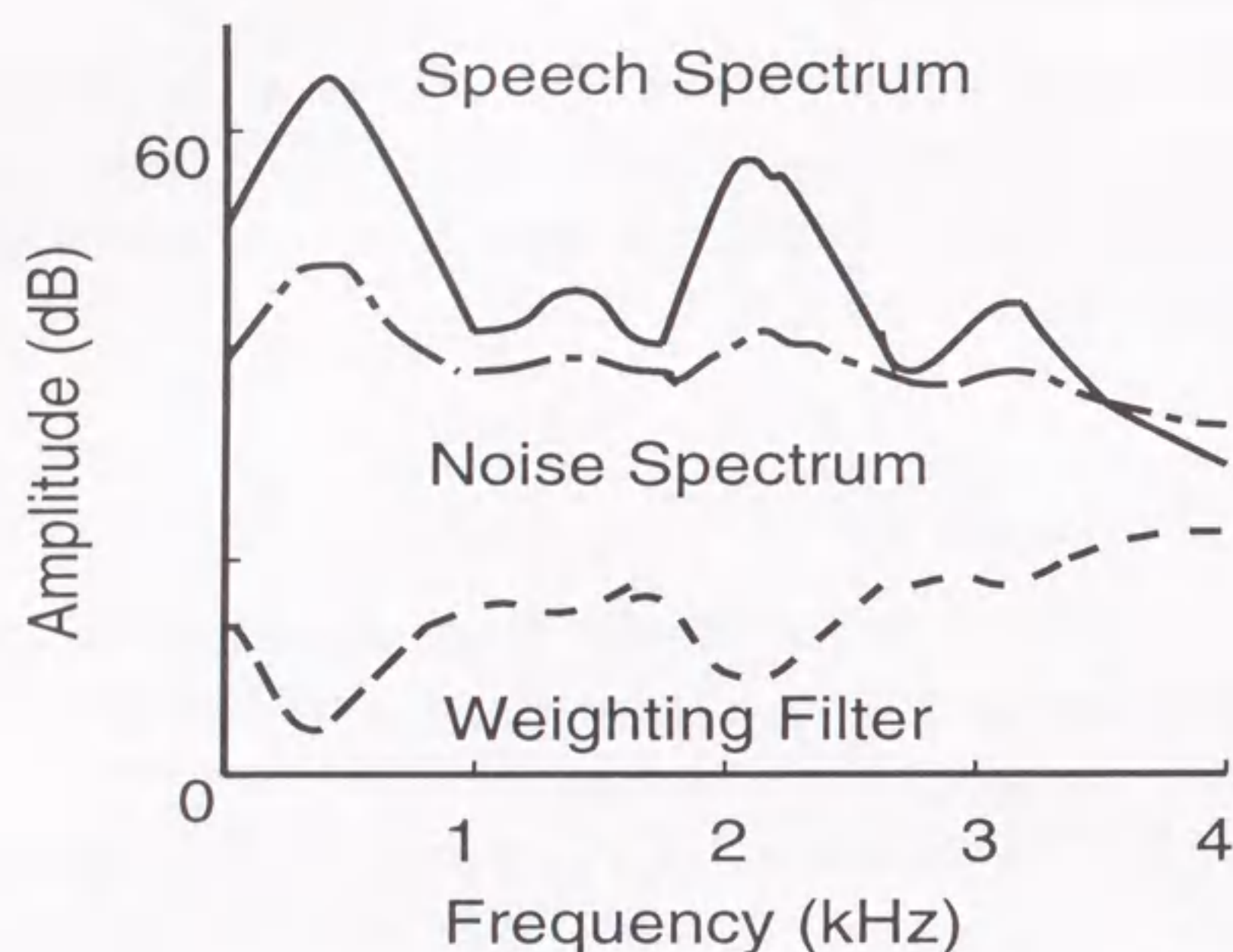


図 1.6: 聴感重み付けによる周波数特性の例

そこで、これらを解決するために、マルチパルスの振幅、位置の決定に A-b-S の手法が導入された [7]。また、これにより、従来の分析合成符号化の問題点であった有声/無声の判別誤りやピッチ抽出誤りに起因した音質劣化を防ぐことができ、より高い音質が得られるようになった。マルチパルス符号化の原理を図 1.7 に示す。マルチパルス音源と合成フィルタから構成される合成モデルを送信側で有し、マルチパルスの各パルスで合成フィルタを駆動して合成音声を一旦再生し、入力音声との聴感重み付け波形歪みを測定し、これを最小化するようにパルスの位置と振幅を探索する。

マルチパルス音源は次の長所がある。(1) 従来の分析合成符号化のように有声/無声判別の誤りに起因する音質の劣化がないこと、(2) 音声の時間的に急激な変化にも追従可能であること、(3) 後述の CELP に比べ、音源信号の探索に必要な演算量を削減できること、(4) 音源信号の格納が不要であること、(5) 音源信号がパルスの振幅と位置のパラメータに分解されて表現されており比較的制御しやすいこと、などであ

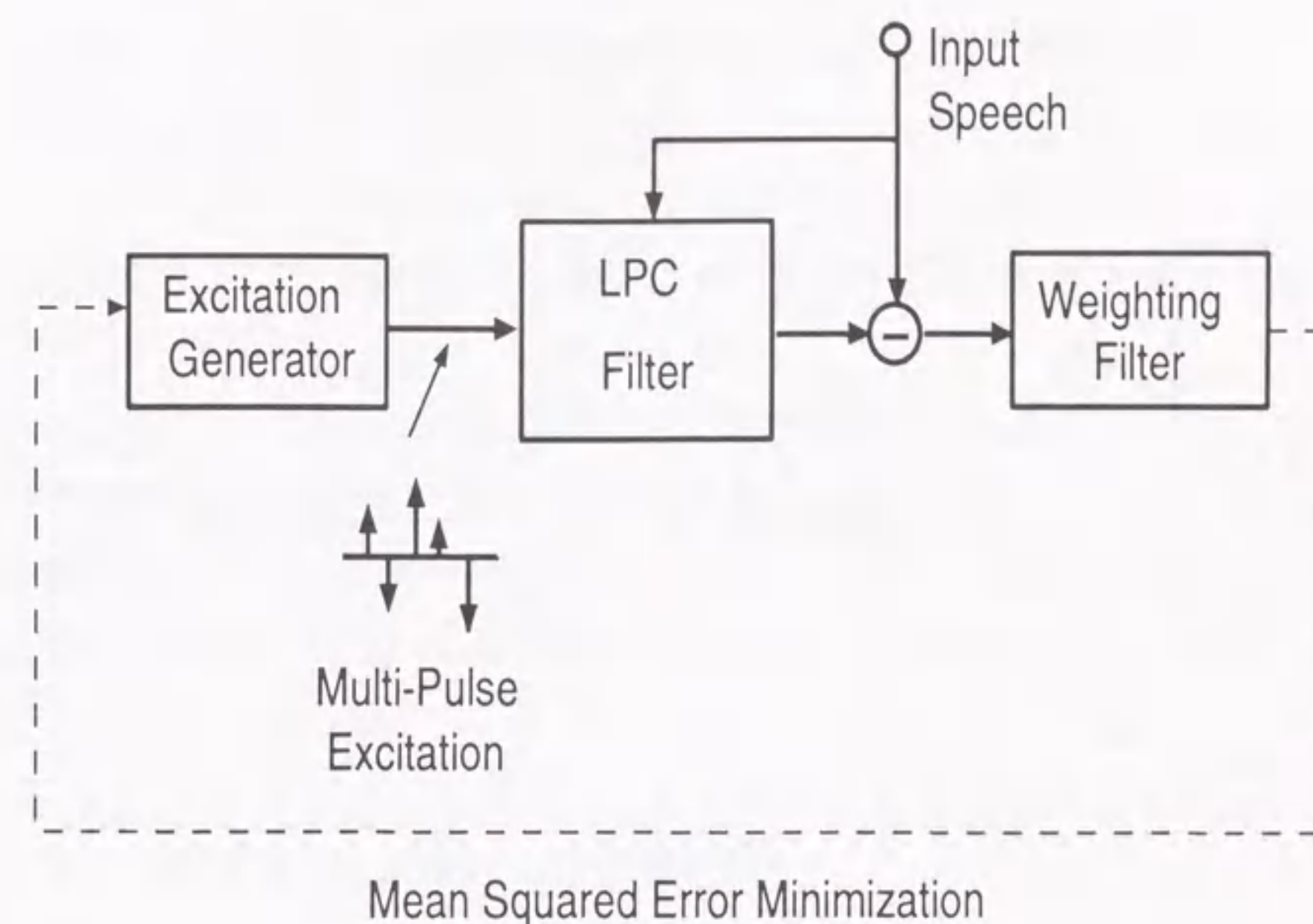


図 1.7: マルチパルス符号化の原理

る。

しかし、マルチパルスには次の問題点があった。(1) パルスの振幅と位置を A-b-S 手法により直接的に計算すると、探索回数だけ合成処理が必要で演算量が膨大となる。(2) マルチパルスの位置と振幅の伝送には、パルスの個数が多いほど多くのビット数が必要であり、伝送速度の低減が容易ではない。これは、各パルスの振幅をスカラ量子化していたこと、フレームなどの処理単位全体における全ての位置の組合せを符号で表して伝送していたためである。これらを解決するために効率的なパルス探索法や伝送情報の低減が必要である。

マルチパルス音源に基づく方式として、演算量を低減化した RPE-LTP (Regular-Pulse Excitation with Long-Term Prediction) [95], [67] や、近年は、MP-MLQ (Multi-Pulse Maximum Likelihood Quantization) [49]、ACELP (Algebraic Code Excited Linear Prediction) [147] や、CS-ACELP (Conjugate Structure Algebraic CELP)

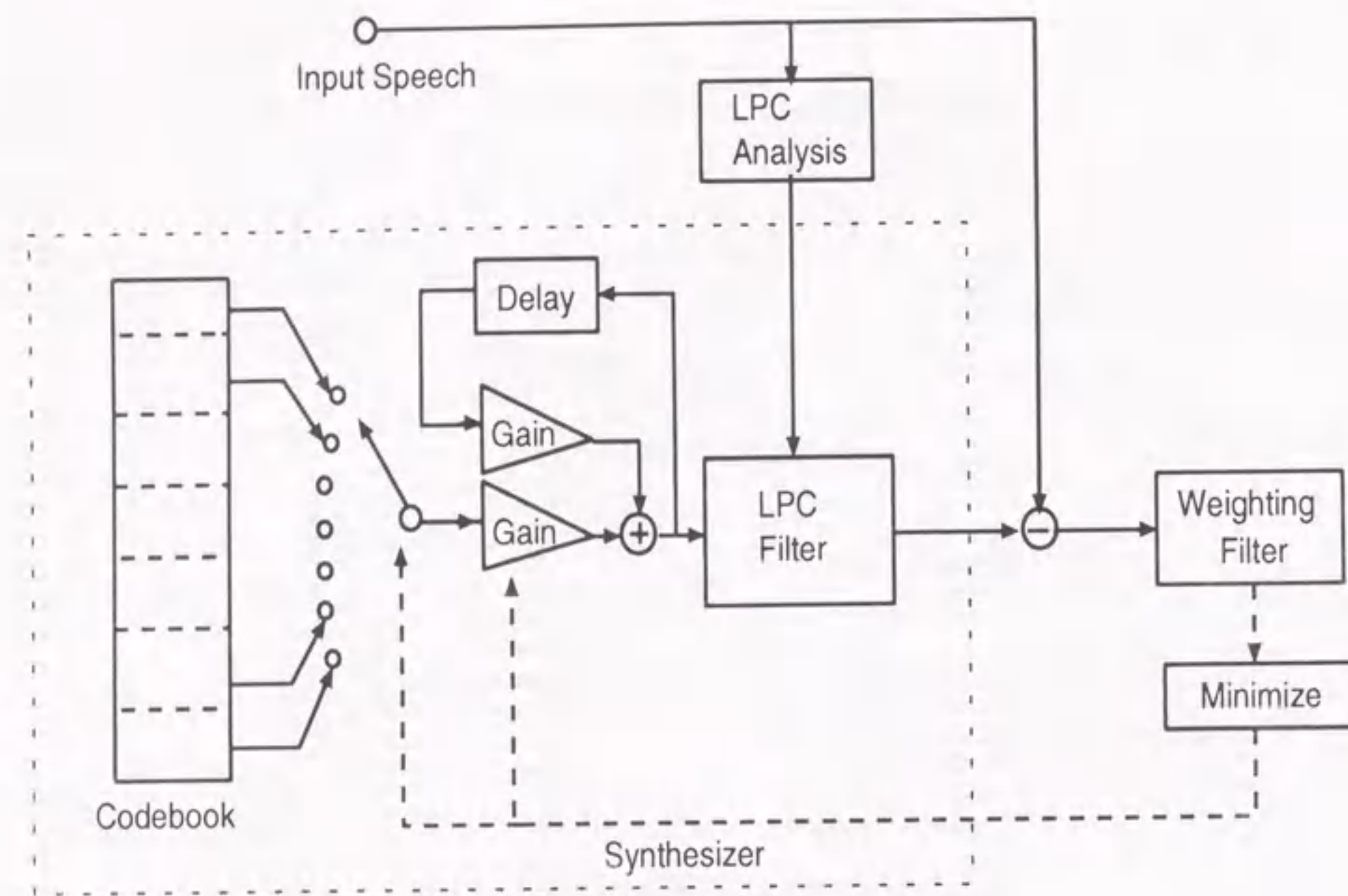


図 1.8: CELP (Code Excited LPC) 符号化の原理

[148] などが開発されている。

音源をベクトル量子化コードブックで表す CELP (Code Excited Linear Prediction)

符号励振型線形予測符号化 (CELP: Code Excited Linear Prediction) は、マルチパルス符号化に比べ、音源信号の量子化効率を高め、一層のビットレート低減化をはかるために、マルチパルス音源をベクトル量子化による音源に置き換えた方式である。符号帳 (Codebook) と呼ばれるベクトル量子化音源コードブックから最適な音源コードベクトルを探索して伝送する [152]。原理を図 1.8 に示す。

マルチパルスにおける聴感重み付け、A-b-S 探索技術は、コードブックの探索にそのまま使用されている。しかし、比較的ビット数の多いコードブック (例えば 5 ms 当たり 10 ビット) から、最適なコードベクトルを A-b-S 法により探索すると多大な演

算量を必要とする。また、コードブックを送信側、受信側で格納するのに多くのメモリ容量を必要とする。これらは、音声符号化方式を装置化する際に問題となって来る。特に、携帯電話や携帯端末などの用途においては、装置をできる限り小型化する必要があるのできわめて重要な課題となる。これらの問題点を解決するために、演算量の少ないコードブック探索法や、演算量、メモリ量の少ない音源コードブックの構造の比較検討が重要である。

CELP は、主として移動体通信応用に開発が加速されており、近年では、VSELP (Vector Sum Excited LP) [31]、PSI-CELP (Pitch Synchronous Innovation CELP) [80], [106] や、音源にマルチパルスを用いる方式として ACELP (Algebraic CELP) [147]、CS-ACELP (Conjugate Structure Algebraic CELP) [148] などが開発されている。

#### 1.4 音声符号化の標準化動向

1984 年に CCITT で ADPCM が標準化されて以来、ITU-T (International Telecommunications Union-Telecommunications Standards Sector) などの国際標準化機関や、携帯電話向けでは欧州 ETSI (European Telecommunications Standards Institute)、北米 TIA (Telecommunications Industry Association)、日本 ARIB (Association of Radio Industries and Businesses)(旧 RCR) などで方式の標準化が行なわれている。ITU-T での標準化動向を表 1.2 に、ETSI、TIA、ARIB での携帯電話向けの標準化動向を表 1.3 にそれぞれ示す [141], [142], [17]。特に、携帯電話などの移動体通信への応用が見え出してきた 1980 年代末頃から標準化が急速に進展している。

#### 1.5 音声符号化に対する要求条件

実用化レベルでの要求条件として、伝送速度、音質、演算量、処理遅延が挙げられる。これらに対する目標を次に示す。

(1) 伝送速度: 表 1.1 に示したようにアプリケーションにより異なるが、特に移動体通信では、無線周波数容量が限られているために、音声符号化の伝送速度を低減させ



表 1.2: ITU-T での音声符号化方式の標準化動向と応用分野

方式 (勧告番号)	音声伝送速度 (kb/s)	機関 / 時期 (年)	応用分野
SB-ADPCM (G.722)	64/56/48	CCITT/1988	7 kHz 帯域の音声による TV 会議
ADPCM (G.726)	32	CCITT/1984	PHS, 基幹伝送, 音声応答
LD-CELP (G.728)	16	CCITT/1992	有線伝送, FPLMTS など
CS-ACELP (G.729)	8	ITU-T/1995	FPLMTS など
低演算量 CS-ACELP (G.729AnnexA)	8	ITU-T/1996	音声とデータの同時伝送 (PC)
MP-MLQ/ACELP (G.723.1)	6.3/5.3	ITU-T/1995	テレビ電話 (PC), インターネット

ADPCM: Adaptive Differential PCM

LD-CELP: Low Delay Code Excited LPC

ACELP: Algebraic CELP

FPLMTS: Future Public Land Mobile Telecommunication Systems

SB-ADPCM: Sub-band ADPCM

CS-ACELP: Conjugate Structure Algebraic CELP

MP-MLQ: Multi-Pulse Maximum Likelihood Quantization

表 1.3: 携帯電話向けの主要な音声符号化方式の標準化動向と応用分野

方式 (勧告番号)	音声伝送速度 (kb/s)	機関 / 時期 (年)	応用分野
RPE-LTP	13	欧州 ETSI/1988	GSM フルレート携帯電話
EFR	13	欧州 GSM/1996	GSM フルレート (音質改良版)
PCS1900 EFR	13	北米 TIA/1996	PCS 向けフルレート (音質改良版)
QCELP (IS-96)	8.5/4/2/0.8	北米 TIA/1994	北米 CDMA 用
VSELP (IS-54B)	7.95	北米 TIA/1989	北米フルレート
VSELP (STD-27C)	6.7	日本 ARIB/1990	日本 PDC フルレート
PSI-CELP (STD-27C)	3.45	日本 ARIB/1993	日本 PDC ハーフレート

RPE-LTP: Regular Pulse Excitation with Long-Term Prediction

QCELP: Qualcomm CELP

PSI-CELP: Pitch Synchronous Innovation CELP

GSM: Global System for Mobile Communication

ARIB: Association of Radio Industries and Businesses

PCS: Personal Communications Systems

EFR: Enhanced Full-Rate

VSELP: Vector Sum Excited Linear Prediction

ETSI: European Telecommunications Standards Institute

TIA: Telecommunications Industry Association

PDC: Personal Digital Cellular

る要請が強い。しかしながら、音質を犠牲にしないように伝送速度を削減するために、効率的な音源信号のモデル化や、効率的な量子化技術などが必要である。

(2) 音質: 音質はできる限り高い方が望ましい。目標は 32 kb/s ADPCM と同等の音質である。しかし、伝送速度が低いほど目標の達成は容易ではない。

(3) 演算量: 特に移動体通信などにおいて、携帯電話や携帯端末などに搭載する場合は、演算量、メモリ量を削減する必要がある。符号化・復号化処理がデジタル信号処理用 LSI (DSP) 1 チップで実現できるのが望ましい。これらの削減は音質を損なわずに行なわなければならない。また、パーソナルコンピュータ (PC) 応用で、PC の CPU を用いて符号化、復号化を行なうソフトウェアコーデックを実現する場合は、演算量の削減がとりわけ重要である。

(4) 処理遅延: タンデム接続や長距離伝送などで発生する遅延を考慮すると、符号化・復号化に必要な処理遅延はできる限り短い方が望ましい。しかしながら、4 kb/s といった低伝送速度では、音質を確保するためにフレーム長を長くするので、遅延を長くせざるを得ない。

以上の要求条件及び、1.1 節の背景を考慮して、本研究における目的と課題を設定する。

## 1.6 本研究の目的と課題

本論文は、16 kb/s ~ 4 kb/s の範囲の伝送速度において、A-b-S 探索による線形予測音声符号化を検討する。

本研究の目的は、高音質化と、演算量、メモリ量の低減化を両立させるための音声符号化アルゴリズムを提供することにある。本論文では、特に、音質、演算量、メモリ量と極めて密接な関係を有する音源信号の構造に焦点をあて、マルチパルス音源とベクトル量子化音源に基づく符号化方式を対象にし、次の課題について研究を行う。

### 1.6.1 マルチパルス音源における課題

第1の課題は、マルチパルスの位置及び振幅を A-b-S に基づき直接探索すると、合成処理を探索回数だけ繰り返す必要があるため、非常に多くの演算量を必要とし実

現困難となる。このため、効率的な探索法の研究が重要である。

本研究は、相関演算に基づきパルスの位置、振幅を逐次的に計算する効率的な探索法を提案する。提案法では、パルス探索の過程で信号の合成処理が不要となること、相関係数の修正と比較によりパルスの位置が計算できることから、従来法に比べ極めて高速な探索を可能にしている。また、同一の誤差評価基準から、さらに性能の高い3種のパルス探索法を提案している。

第2の課題は、音質を保持しながら伝送速度を低減化することである。

このために、本研究は、ピッチ情報を用いて伝送速度を低減化するための2種の方法を提案する。まず最初に、ピッチ予測について検討する。APCなどで使用されている従来の適応ピッチ予測とは異なり、基本的にはマルチパルスの探索ループ内にピッチ合成処理を含め、ピッチ合成を行ないながらパルスを最適に探索する方法を検討する。次に、マルチモード化による音源信号の切替えと、母音区間でピッチ同期マルチパルス音源を用いる符号化について述べる。また、この方法によれば、ピッチ予測に比べ、伝送速度の低い領域で有効であることを明らかにする。

### 1.6.2 ベクトル量子化音源における課題

第1の課題は、演算量、メモリ量を削減できるコードブックの構造及び探索法の検討である。

本研究は、音源コードブックの複数の異なる構造について、性能と探索に必要な演算量・メモリ量とを比較検討し、最良のものとしてマルチパルスから構成されるスパース構造を選択する。

第2の課題は、4 kb/s程度までビットレートを低減化しても良好な音質を維持できる符号化方式の検討である。

本研究は、先に選択した音源コードブック構造を基本とし、マルチモード化の考えによるM-LCELP (Multi-mode Learned CELP) を提案する。ここでは、入力音声の種類を複数種のモードに分類し、モードに応じてコードブックを切り替えることにより、音声の特徴に依らず良好な性能を得ることを目的とする。

第3の課題は、特に携帯電話などの移動体通信の場合に重要となる、伝送路誤り

に対するアルゴリズム自身の頑健性である。

本研究は、伝送路誤りに対する頑健性を高めるための検討を行う。

### 1.6.3 マルチパルスとベクトル量子化を結合した音源における課題

これまでに検討したマルチパルスとベクトル量子化の長所を結合して音源信号を表現する。

第1の課題は、マルチパルス音源の効率的な量子化である。

マルチパルス音源は、音源信号の時間的に急激な変化に追従できる長所があり、さらに、本研究で提案する相関演算による探索法を用いることで、ベクトル量子化音源に比べ、探索に要する演算量を大幅に削減できるが、従来、位置、振幅情報をスカラ量子化しており伝送に多くのビット数が必要であった。そこで、本研究は、位置を限定して位置を表す量子化ビットを削減し、さらに、パルスの振幅をまとめてベクトル量子化することで、マルチパルス音源の伝送情報量の低減化を可能にする。

第2の課題は高音質化である。

本研究は、性能を改善するために、位置の複数候補に対し振幅ベクトル量子化器を組合せ探索する方法を提案する。また、候補の個数と性能の改善度、演算量との関係を明らかにする。

## 1.7 本論文の構成

本論文の構成は次の通りである。

第2章は、マルチパルス音源を用いる線形予測音声符号化法において、効率的なパルス探索法を提案する。ここでは、まず、A-b-S (Analysis-by-Synthesis) 法に基づきパルスの位置、振幅を求める従来の基本的な探索法について述べる。次に、聴感重み付け誤差最小化基準から、相関演算による演算量の極めて少ない効率的な逐次探索法を導出し、有効性を実験的に示す。さらに、音質を一層改善するために、同一の誤差基準から、パルス探索の改良法を3種類提案し、これらについて、必要な演算量と符号化音質との関係を比較検討する。また、以上の改良により、良好な音質を維持し

ながら、伝送速度を 16 kb/s ~ 9.6 kb/s 程度まで低減化可能であることを音質評価実験の結果により明らかにする。

次に、マルチパルス符号化の伝送速度を 4.8 kb/s まで低減化するために、ピッチ情報を用いる方法を2種類検討する。

3章は、ピッチ予測マルチパルス法について論じる。従来の APC など既知に知られている適応ピッチ予測法とは異なり、ピッチ合成処理をパルス探索のループ内に組み込む方法を提案する。また、実験を行い、伝送速度を 8 kb/s 程度まで低減化可能であることを示す。

4章は、ピッチ情報を利用するもう一つの方法として、マルチモード符号化の考えによる音源信号の切替え及び、母音モードでのピッチ同期マルチパルス音源 (PSMP: Pitch Synchronous Multi-Pulse) を提案する。伝送速度を 4.8 kb/s 以下まで低減しても比較的良好な音質が得られることを実験的に示す。

第5章は、ベクトル量子化音源を用いる従来の CELP 方式の音質改良として、M-LCLEP (Multi-mode Learned Code Excited Linear Prediction) 方式を提案する。第4章のマルチモード符号化の考えを採用し、モードに応じてコードブックを切替えることで性能を改善する。また、この中で、3種の音源コードブックの構造について、演算量、メモリ量、音質の観点から比較検討し、最良の構造として、マルチパルス列から構成されるスパース構造を選び出す。さらに、伝送路誤りに対する頑健性についても考察する。移動体通信での種々の環境条件を考慮した実験により、本方式が 4 kb/s でも、従来の 8 kb/s CELP 方式の音質と同等という良好な音質を提供できること、適切な誤り訂正符号化を付加することで、伝送路誤りに頑健であることを示す。

第6章はマルチパルス音源とベクトル量子化の長所を結合する。5章の M-LCLEP をベースにして、高音質化及び、演算量、メモリ量の一層の削減を狙い、マルチパルスの振幅をベクトル量子化により効率的に量子化する MP-CELP (Multi-Pulse based CELP) 符号化を提案する。また、パルス位置の総組合せ探索に必要な演算量を大幅に削減するために、パルス位置の高速探索法を提案する。5章の M-LCLEP 方式と音質、演算量、メモリ量について比較検討し、演算量が同程度の場合、より少ないメモリ量で音質が大幅に改善されることを明らかにする。さらに、移動体通信での種々の

環境条件において音質評価を行ない、極めて良好な音質が得られることを明らかにする。

第7章は、2章で検討したマルチパルス音声符号化法のデジタル移動体衛星通信や、デジタル移動通信などへの適用性を調査するために、伝送符号誤りに対する品質を客観的、主観的に評価する。まず、各伝送パラメータの誤りに対する感度を調べる。16 kb/s の場合、誤りに敏感なパラメータを一部保護することで、450 b/s の冗長ビット数の追加で、誤り率が  $10^{-3}$  のときに、保護なしに比べ、 $\mu$ -law PCM 換算で 0.5 ~ 1 ビット分相当の音質が改善されることを明らかにする。

第8章は、本論文で検討した高能率音声符号化方式が実際の製品として実用化された例について示す。

第9章は、本研究を総括し、結論、本研究の意義と貢献、今後の研究の課題と展望について述べる。

## 1.8 本論文で提案する方式の伝送速度と音質の関係

本論文の第2章~第6章の各々で提案するマルチパルス符号化(方式2)(MPC)、ピッチ予測マルチパルス符号化(MPC with Pitch Prediction)、ピッチ同期マルチパルス(PSMP)、マルチモード CELP(M-LCELP)、マルチパルス CELP(MP-CELP)の伝送速度と音質の大まかな関係を図1.9に示す。

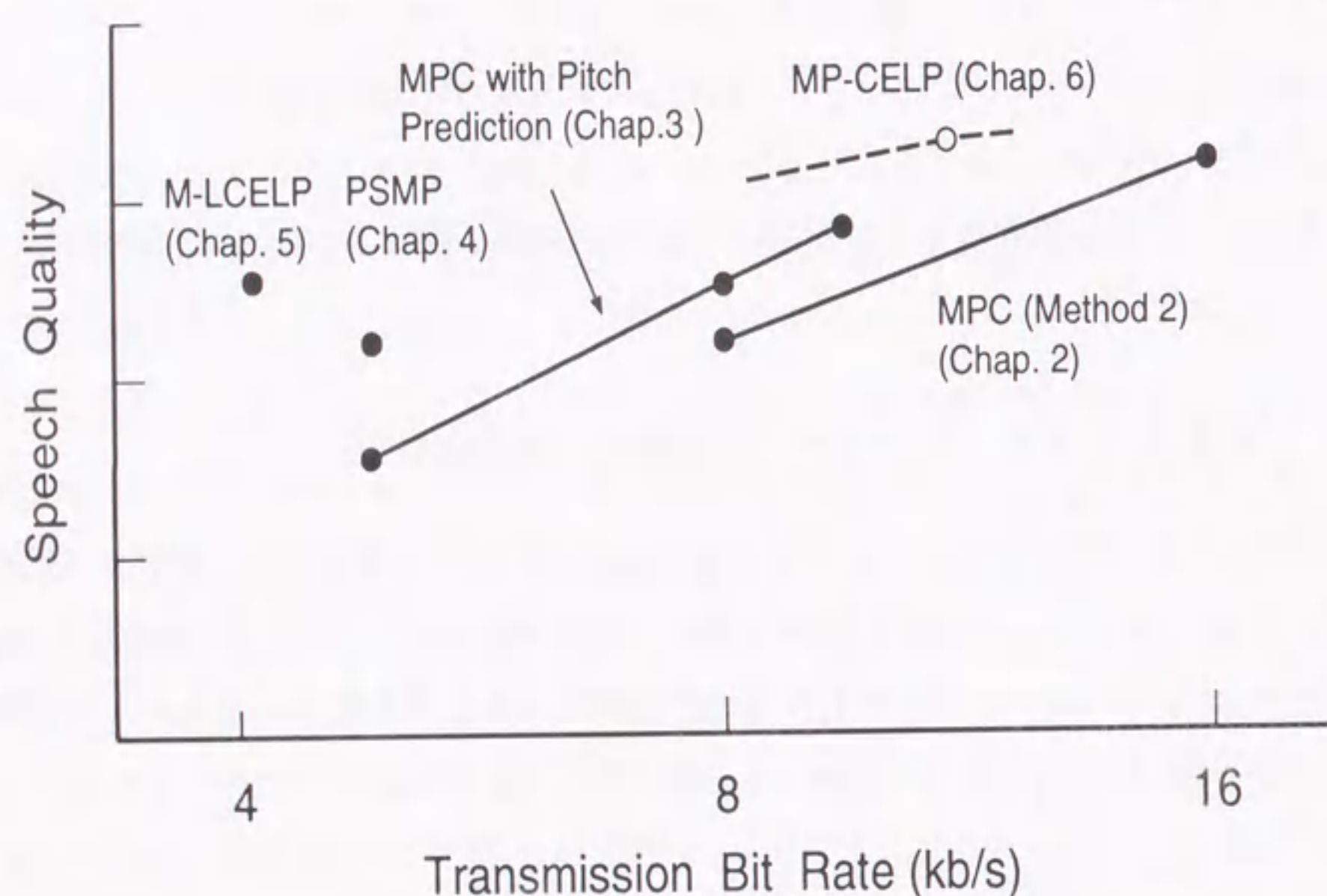


図 1.9: 本論文で提案する種々の方式の伝送速度と音質の大まかな関係

## 第2章

### 相関演算に基づくパルスの逐次高速探索アルゴリズムを用いるマルチパルス音声符号化法

#### 2.1 まえがき

音声信号を 16 kb/s ~ 9.6 kb/s のビットレートで高品質に符号化する音声符号化方式は、伝送容量の拡大や蓄積時間の拡大などが可能となるため、有線、無線などの通信分野、音声応答、ボイスメールなどの音声蓄積の分野などから方式の確立が期待されている。この伝送帯域においては、内外を問わず種々の方式の研究が精力的に進められている [18], [7], [37], [180], [29], [89]。

この伝送帯域で有効な方式として、マルチパルス駆動形線形予測音声符号化法 (MP-CELP: Multi-Pulse Excited Linear Prediction) [7] が提案されている。マルチパルス音声符号化は、駆動音源信号を位置と振幅の異なる複数個のインパルスの組合せ (以下マルチパルスと呼ぶ) からなるマルチパルス音源モデルで表し、そのような音源を用いて LPC 合成フィルタを駆動して得られた合成波形が入力音声波形と最も近くなるようなパルス列を探索し符号化する。しかしながら、A-b-S (Analysis-by-Synthesis) 手法により、信号を合成しながらパルスの位置、振幅を計算するために、信号の合成処理が探索回数だけで膨大な演算量を必要とするという問題点があった。

パルスの位置、振幅の探索法は、符号化全体の音質及び演算量を左右する重要な技術である。本章は、マルチパルス音声符号化法において、マルチパルス音源の位置、振幅を逐次探索する探索アルゴリズムとして、4 種類の方法を提案し、比較評価を行

なう。まず最初に、相関演算に基づく演算量の少ない効率的なパルス探索アルゴリズムを提案する。次に、音質をさらに改善するために、3種の改良法について検討する。4種のパルス探索法について、演算量、性能、音質を比較評価した結果を示す。

## 2.2 マルチパルス駆動形音声符号化法 (MPELP)

### 2.2.1 マルチパルス音源による音声合成モデル

第1章1.3.4節でも述べたように、従来の線形予測ボコーダに用いられた音声合成モデルにおける問題点を解決し、高品質な音声を合成可能であり、ビットレートを低減化できる可能性を有する音源モデルとして、マルチパルス音源モデルが提案された[7]。マルチパルス音源による音声合成モデルを図2.1に示す。

マルチパルス音源発生器は、互いに異なるサンプリング位置  $m_1, m_2, \dots, m_K$  に、振幅  $g_1, g_2, \dots, g_K$  の  $K$  個のインパルス列 (マルチパルス) を発生する。ここで  $K$  は、あらかじめ定められた時間長を有するフレーム内のパルスの個数である。マルチパルスは線形予測による合成フィルタに入力され、フィルタの出力に音声が合成される。合成フィルタの係数は、パルス探索の前に入力音声からフレーム毎に線形予測分析により計算される。

マルチパルス音声合成モデルに基づく音声符号化方式のブロック図を図2.2に示す。送信側は、線形予測分析部とマルチパルス探索部から構成される。フレーム毎に線形予測係数、マルチパルスの位置と振幅が伝送される。受信側は、マルチパルス発生部と合成フィルタ部から構成される。

### 2.2.2 マルチパルス探索アルゴリズム

マルチパルス音声符号化では、良好な音質を得るためには、マルチパルスの位置と振幅を計算する探索法が重要な役割を占める。AT&Tベル研究所のAtalらは、A-b-S (Analysis-by-Synthesis) の手法を用いてパルスを探索する方法を提案した[7]。図2.3にブロック図を示すように、マルチパルス音源によりLPC合成フィルタを駆動し

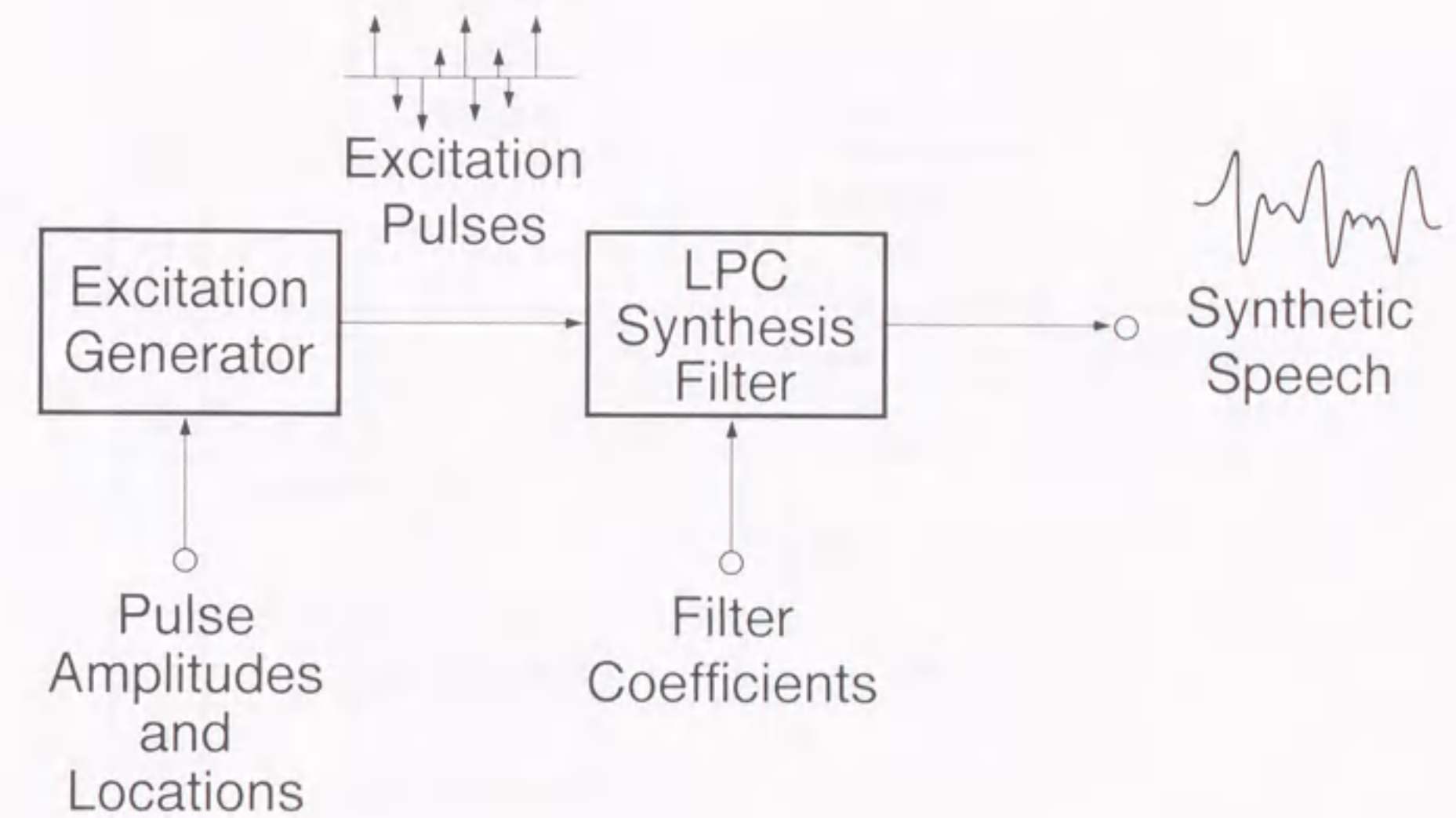


図 2.1: マルチパルス音源を用いる音声合成モデル

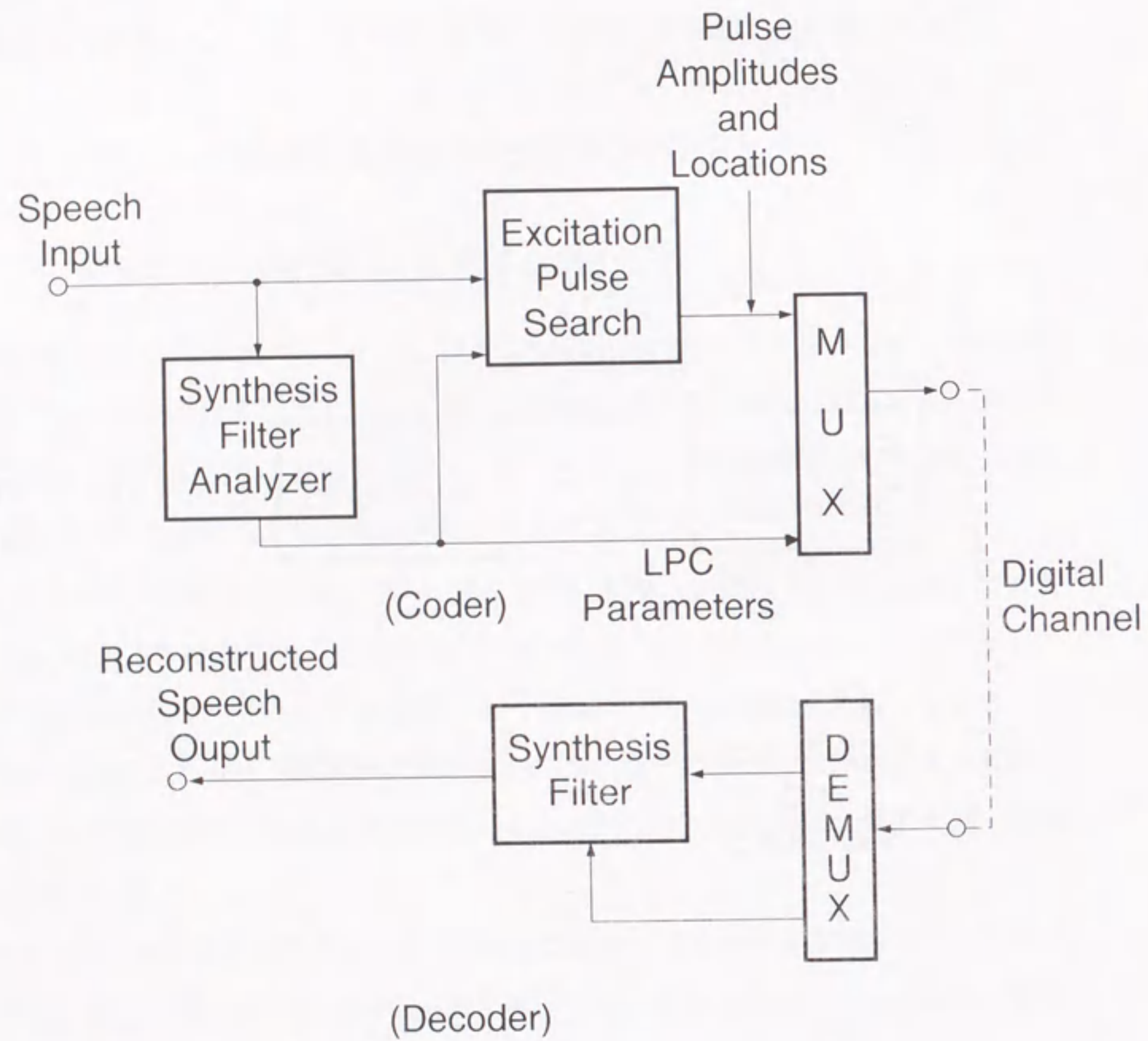


図 2.2: マルチパルス音声符号化のブロック図

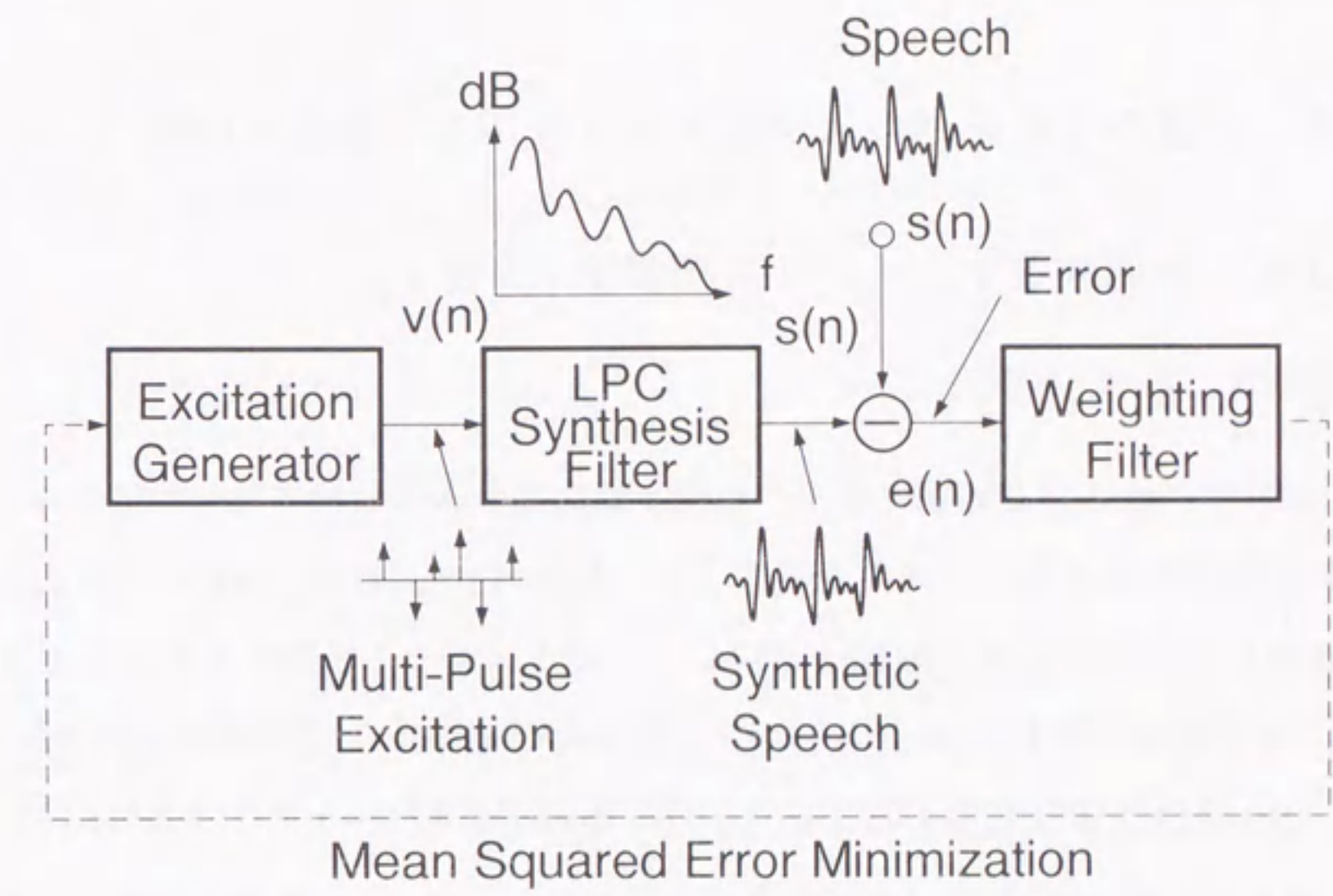


図 2.3: A-b-S (合成による分析) に基づくマルチパルスの探索

て信号を合成し、合成音声と入力音声との聴感重み付け2乗誤差を最小化するようにパルスの位置と振幅を求める。全てのパルスの位置と振幅を同時に決定するのが理想であるが、非線形最適化問題となり膨大な演算量を要する。そこで、彼らは、1パルスずつ逐次的にA-b-Sにより探索する方法を示した[7]。しかしながら、逐次解法であるものの、探索回数だけ合成処理を伴うため、依然として演算量が多いという問題が残されていた。

### 2.3 4種のマルチパルス探索アルゴリズムの提案と比較

#### 2.3.1 効率的なマルチパルス探索法(方式1)

##### 相関演算に基づく定式化

信号の合成処理を必要とせず、相関演算に基づききわめて効率的にパルスの位置、振幅を探索するアルゴリズムを提案する[108], [110], [113], [116]。本方式では、過去に求めたパルスの位置、振幅は固定し、1パルスずつ逐次的に求めることを特徴とする。逐次探索の考えはAtalらと同じであるが、探索時に信号の合成処理が不要なため、本方式の方が演算量ははるかに少なくなっている。

フレーム当たりK個のパルスを求めることとし、 $m_i$ 、 $g_i$ をそれぞれ、 $i$ 番目のパルスの位置、振幅とすると、マルチパルス音源信号 $v(n)$ は下式のように表せる。

$$v(n) = \sum_{i=1}^K g_i \delta(n - m_i) \quad (2.1)$$

ここで、 $\delta$ はクロネッカーのデルタ関数であり、

$$\delta(n - m_i) = \begin{cases} 1 & (n = m_i) \\ 0 & (n \neq m_i) \end{cases}$$

とかける。

過去のフレームからの応答信号があらかじめ除去されている場合、 $v(n)$ を合成フィルタに通して得られる合成音声 $s(n)$ は、合成フィルタのインパルス応答を $h(n)$ とすると次式のように表せる。

$$s(n) = \sum_{i=1}^K g_i h(n - m_i), \quad (0 \leq m_i \leq N - 1) \quad (2.2)$$

従って、入力音声と合成音声との誤差信号は

$$e(n) = x(n) - s(n) \quad (2.3)$$

となる。誤差信号に聴感重み付けを行った後の聴感重み付け誤差信号 $e_w(n)$ は

$$\begin{aligned} e_w(n) &= e(n) * w(n) \\ &= \left[ x(n) - \sum_{i=1}^K g_i h(n - m_i) \right] * w(n) \\ &= x_w(n) - \sum_{i=1}^K g_i h_w(n - m_i) \end{aligned} \quad (2.4)$$

と表せる。ここで、記号 $*$ は畳み込み演算を示す。 $w(n)$ は式(1.1)の伝達特性を有する聴感重み付けフィルタのインパルス応答を示す。また、

$$x_w(n) = x(n) * w(n) \quad (2.5)$$

$$h_w(n) = h(n) * w(n) \quad (2.6)$$

である。ここで、 $x_w(n)$ 、 $h_w(n)$ は聴感重み付け音声信号、聴感重み付けインパルス応答をそれぞれ示す。

聴感重み付け後の重み付け誤差電力 $E$ は

$$\begin{aligned} E &= \sum_{n=0}^{N-1} e_w^2(n) \\ &= \sum_{n=0}^{N-1} \left[ x_w(n) - \sum_{i=1}^K g_i h_w(n - m_i) \right]^2 \end{aligned} \quad (2.7)$$

となり、パルスの振幅、位置は、式(2.7)を最小化するように求めることができる。ここで、 $N$ はパルスの位置を探索する区間長を示し、本章ではフレーム長にとる。

$K$ 個のパルス全てについて、式(2.7)を同時に最小化するのは、一つのパルスにつきパルス位置の自由度が $N$ あるため、 $\frac{N!}{K!(N-K)!}$ の組合せを調べる必要がある。これは区間長 $N$ やパルス数 $K$ が大きくなるほど現実的には探索不可能となる。そこで、この問題を避けるために、パルスを1パルスずつ逐次的に探索する。さらに、すでに過去に求めた $i-1$ 個のパルスの位置と振幅は固定されており、 $i$ 番目のパルスの探索により変化しないと仮定する。この仮定により、一つのパルスのみの位置と振幅を一度に探索すればよく、問題を単純化できる。

今、1個のパルスのみが存在するとすれば、式(2.7)をパルスの振幅 $g_1$ で偏微分し0とおくと次式が導かれる。

$$E_1 = \sum_{n=0}^{N-1} x_w^2(n) - \frac{[R_{hx}(m_1)]^2}{R_{hh}(m_1, m_1)} \quad (2.8)$$

$$g_1(m_1) = \frac{R_{hx}(m_1)}{R_{hh}(m_1, m_1)} \quad (2.9)$$

となる。ここで、

$$R_{hx}(m_i) = \sum_{n=m_i}^{N-1} x_w(n) h_w(n - m_i) \quad (2.10)$$

$$R_{hh}(m_i, m_j) = \sum_{n=\max(m_i, m_j)}^{N-1} h_w(n - m_i) h_w(n - m_j) \quad (2.11)$$

である。式(2.10)は聴感重み付け信号 $x_w(n)$ と聴感重み付けインパルス応答 $h_w(n)$ との相互相関、式(2.11)は聴感重み付けインパルス応答 $h_w(n)$ の共分散を示す。

従って、パルスの位置 $m_1$ は、式(2.8)を最小化するように求めればよい。ここで、右辺第1項は重み付け信号 $x_w(n)$ の電力であり、フレーム内で一定であるため、右辺第2項を最大化する位置 $m_1$ を計算すればよい。このときの最適な振幅 $g_1(m_1)$ は、 $m_1$ を式(2.9)に代入して計算できる。

2番目以降のパルスは、1番目のパルスの振幅、位置を固定して考える。式(2.7)を $g_2$ について最小化すると、

$$E_2 = \sum_{n=0}^{N-1} x_w^2(n) - \frac{[R_{hx}(m_2) - g_1 R_{hh}(m_1, m_2)]^2}{R_{hh}(m_2, m_2)} \quad (2.12)$$

$$g_2(m_2) = \frac{R_{hx}(m_2) - g_1 R_{hh}(m_1, m_2)}{R_{hh}(m_2, m_2)} \quad (2.13)$$

となる。

以上より、 $k$ 番目のパルスの探索には式(2.14)、式(2.15)が導かれる。まず、位置は式(2.14)を最大化するように求める。

$$D_k = \frac{[R_{hx}(m_k) - \sum_{i=1}^{k-1} g_i R_{hh}(m_k, m_i)]^2}{R_{hh}(m_k, m_k)}, \quad (0 \leq m_k \leq N-1) \quad (2.14)$$

このときの位置 $m_k$ を使用して、振幅は式(2.15)から計算する。

$$g_k = \frac{R_{hx}(m_k) - \sum_{i=1}^{k-1} g_i R_{hh}(m_k, m_i)}{R_{hh}(m_k, m_k)} \quad (2.15)$$

これらの式から、式(2.10)、式(2.11)に示した2種類の相関係数 $R_{hx}(m_k)$ 、 $R_{hh}(m_k, m_i)$  ( $0 \leq m_k \leq N-1$ )をフレーム毎にあらかじめ計算しておけば、式(2.14)、式(2.15)により、パルスの位置と振幅が逐次演算により求められることがわかる。

#### 自己相関近似による演算量の低減化

共分散 $R_{hh}(m_k, m_i)$ の計算は演算量を必要とするので、演算量をさらに低減化するために、これを自己相関 $R_{hh}(|m_k - m_i|)$ で近似し、さらに、2乗計算を絶対値計算で置き換える。ここで、入力信号が処理区間では定常でインパルス応答系列が充分長ければ共分散を自己相関で近似することの誤差は少ないと考えられる。



このとき式(2.14)は式(2.16)のように簡略化できる。

$$D'_k = \frac{|R_{hx}(m_k) - \sum_{i=1}^{k-1} g_i R_{hh}(|m_k - m_i|)|}{R_{hh}(0)}, \quad (0 \leq m_k \leq N-1) \quad (2.16)$$

ここで分母はフレーム内で一定なので、位置は式(2.16)の分子のみを最大化するように探索すればよく、演算量がさらに削減される。ここで、自己相関は式(2.17)に従い計算する。

$$R_{hh}(|m_k - m_i|) = \sum_{n=0}^{N-1-|m_k - m_i|} h_w(n)h_w(n + |m_k - m_i|) \quad (2.17)$$

式(2.16)から求めた位置  $m_k$  を使用して、振幅は式(2.18)から計算することができる。

$$g_k = \frac{R_{hx}(m_k) - \sum_{i=1}^{k-1} g_i R_{hh}(|m_k - m_i|)}{R_{hh}(0)} \quad (2.18)$$

#### パルス逐次探索の波形例

式(2.16)の分子と式(2.18)によるパルスの逐次探索の波形例を図2.4に示す。まず、式(2.16)の分子の絶対値の中身を評価関数と考え  $R(m)$  とする。パルスをたてる前はこの値は  $R_{hx}(m)$  に等しい。図2.4(a)は  $R(m)$  を示す。(b)は1番目のパルスを決め、(c)は1番目のパルスにより式(2.16)の分子を修正することで、 $R(m)$  を修正した様子を示す。(d)は、 $|R(m)|$  の最大値を探索し、2番目のパルスの位置を求め、式(2.18)から最適な振幅を求めている。(e)~(h)では、以上の処理を繰り返し、定められた個数(ここでは4)のパルスを決定している。

#### 効率的なパルス探索法を用いるマルチパルス符号化の構成

効率的なパルス探索法(方式1)を用いるマルチパルス符号化のパルス探索部の構成を図2.5に示す。2.3.2節~2.3.4節以降に示す他のパルス探索法の構成も図2.5と

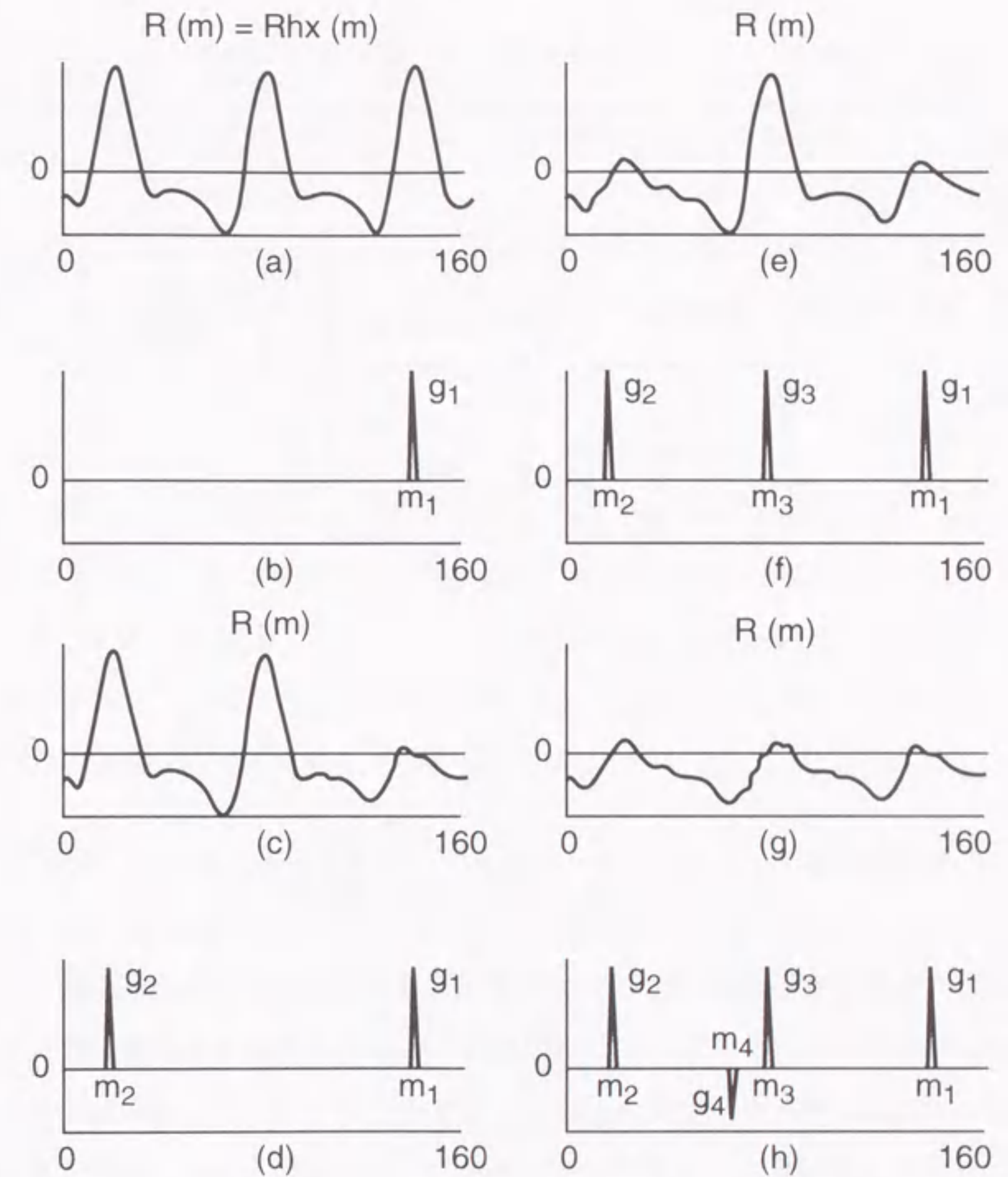


図2.4: パルス探索の波形例, (a) $R(m)$ , (b)第1パルス, (c)第1パルスによる  $R(m)$  の修正, (d)第2パルス, (e)第2パルスによる  $R(m)$  の修正, (f)第3パルス, (g)第3パルスによる  $R(m)$  の修正, (h)第4パルス

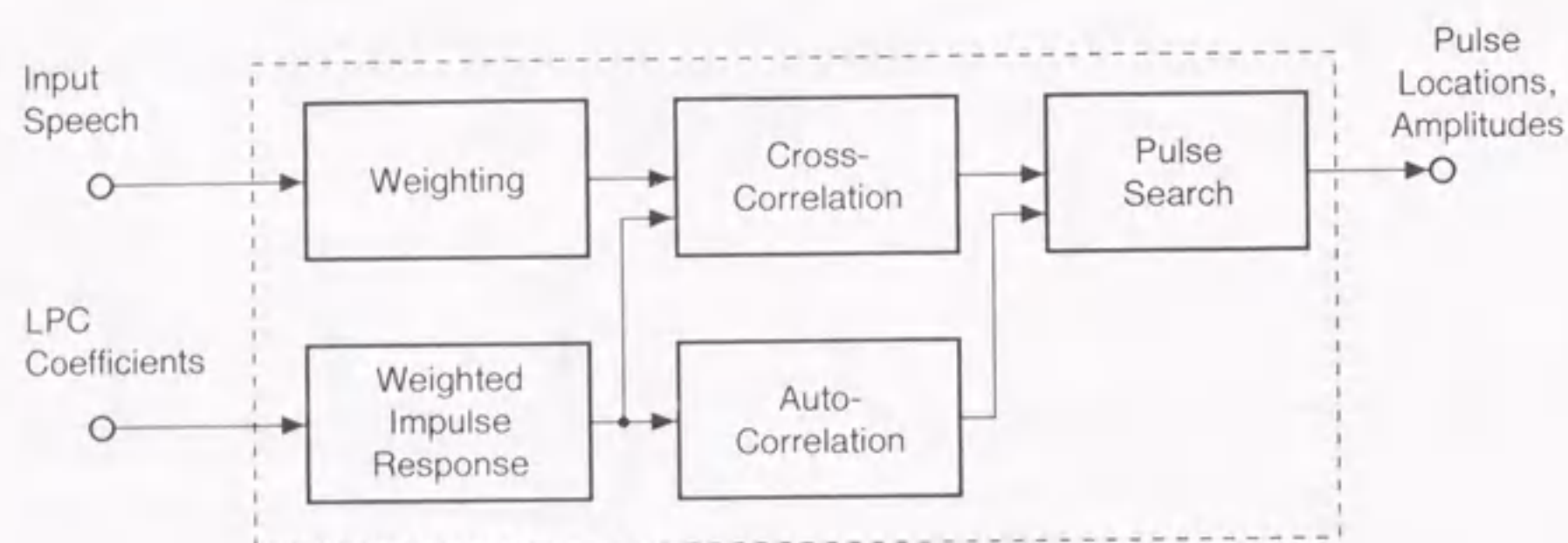


図 2.5: パルス探索部の構成

ほぼ同じである。聴感重み付け部は、音声信号  $x(n)$  及び線形予測係数  $\alpha_i$  を入力して式 (2.5) の計算をフィルタリングを用いて行い  $x_w(n)$  を出力する。重み付けインパルス応答計算部は、線形予測係数  $\alpha_i$  を入力して式 (2.6) に基づき合成フィルタと聴感重み付けフィルタを縦続接続したフィルタのインパルス応答  $h_w(n)$  を計算する。相互相関計算部、自己相関計算部はそれぞれ、式 (2.10)、式 (2.17) の計算を行い、 $R_{hx}(m)$ 、 $R_{hh}(m)$  を出力する。パルス探索部は、式 (2.16) の分子、式 (2.18) を用いてパルスの探索を行い、パルスの位置と振幅を出力する。

### 2.3.2 パルス探索法の改良と4種のパルス探索法の比較

以下では、性能をさらに向上させるために、パルス探索法の改良について示す。いずれも、聴感重み付け誤差電力最小化基準により導出される。

表 2.1: 4種のパルス探索法の比較

探索法	パルス位置探索時の過去のパルス振幅	パルス位置決定後のパルス振幅の調整
方式1	全て固定	新たに求めたパルスに対してのみ振幅計算
方式2	全て固定	パルスを求める毎に全てのパルスの振幅を再計算
方式2-1	全て固定	新たに求めたパルスとその近傍1パルスの振幅を再計算
方式2-2	全て固定	新たに求めたパルスとその近傍2パルスの振幅を再計算

今まで述べた効率的なパルス探索法 (方式1) と、2.3.3節～2.3.4節で説明する3種の改良法 (方式2, 方式2-1, 方式2-2) との相違点を表 2.1 にまとめる。いずれの方法も、新たなパルスの位置を探索する時は、すでに求めた過去のパルスの位置は固定しているが、新たなパルスの位置を探索した後でパルスの振幅を再計算する際に違いがある。この違いにより、性能と演算量が異なってくる。

また、図 2.5 に示したパルス探索部において4種のパルス探索に必要な演算量を表 2.2 で比較する。方式1では、フレーム当たり、 $R(m)$  の修正及び位置の探索に  $2LK$  回の演算、パルスの振幅の計算に  $K$  回の演算と、きわめて少ない演算量で実現することができる。ここで、聴感重み付け、インパルス応答、相互相関、自己相関の計算は各方法とも共通であるため、表 2.2 では除外してある。これらは、それぞれ、フレーム当たり、 $2PN$ 、 $PN$ 、 $NL$ 、 $L$  ( $L < N$ ) 回の演算を必要とする。ただし、 $N$ 、 $P$ 、 $L$  はそれぞれ、フレームのサンプル数、線形予測分析の次数、重み付けインパルス応答の点数を示す。

また、表 2.2 には一例として、フレーム長を 20 ms ( $N = 160$ )、パルス数 ( $K$ ) を 16、インパルス応答長 ( $L$ ) を 40 としたときの1秒当たりの積和回数を MOPS (Million Operation Per Second) で示す。

表 2.2: 4 種のパルス探索法の演算量の比較

探索法	パルス位置決定 (積和回数)	パルス振幅計算 (積和回数)	R(m) の修正 (積和回数)	16 パルスの数値例 (MOPS)
方式 1	-	$K$	$2LK$	0.26
方式 2	-	$K^3/3$	$K^2L$	2.83
方式 2-1	-	$8K$	$4LK$	0.52
方式 2-2	-	$33K$	$6LK$	0.79

$K$ : フレーム内のマルチパルスの個数

$N$ : フレームのサンプル数

$L$ : 重み付けインパルス応答のサンプル数

### 2.3.3 振幅調整法 (方式 2)

方式 1 では、新たなパルスを求める際に、過去に定めたパルスの位置と振幅は固定していた。パルスの個数が少なくパルス間の間隔が広い場合は問題が少ないが、パルスの個数が増してくると、パルス間の間隔は接近し、重み付けインパルス応答によるパルス間の相互作用が無視できなくなり、性能が伸びにくくなる。

方式 2 は、過去に求めたパルスの位置は固定として考えるが、パルスの振幅については、新たなパルスの位置を決定する毎に、すでに求めた全てのパルスの振幅を式 (2.7) を最小化するように計算しなおすことを特徴とする [112],[157]。この特徴を前述の表 2.1 にまとめる。

式 (2.7) において、パルスの位置は固定と考え、パルスの振幅で偏微分して 0 とおくと、

$$\sum_{i=1}^K g_i R_{hh}(m_i, m_j) = R_{hz}(m_j), \quad (1 \leq j \leq K) \quad (2.19)$$

が得られる。新たなパルスの位置を求める毎に、すでに位置が決定されている全てのパルスに対して、式 (2.19) を解いてパルスの振幅を計算し直す。

### 2.3. 4 種のパルス探索アルゴリズムの提案と比較

演算量の比較を表 2.2 に示す。方式 1 に比べ、式 (2.19) の解法のために演算量が増大する。パルスの振幅の再計算には、式 (2.19) を解くのに  $K^3/3$  の積和回数が必要である。また評価関数  $R(m)$  を修正するのに  $K^2L$  だけ必要となる。

#### 2.3.4 演算量低減型振幅調整法 (方式 2-1, 方式 2-2)

方式 2 は、方式 1 に比べ、パルスの個数が増えた時に性能の改善が大きいと考えられるが、パルスの振幅を再計算するための演算量が増大する。

そこで、方式 2 において、性能をほとんど損なわずに演算量を低減化する方法として、方式 2-1、方式 2-2 を提案する [113], [116]。これらの方式は、パルスの振幅の調整に要する演算量を低減化するために、新たなパルスの位置が決定された際に、新たなパルスとそこから近い距離にある少数のパルスのみを調整する。これは、聴感重み付けインパルス応答は比較的早く減衰するために、距離の近いパルス同志のみが影響を及ぼしあうと考えられるからである。また、再計算するパルスの個数は常に一定とする。これらにより、演算量を大幅に削減することが可能となる。

方式 2-1 は、振幅を調整するパルスの個数を常に 2 としている。つまり、新たなパルスを求めると、そのパルスの振幅と、そのパルスの位置に最も近い過去の 1 パルスの振幅とを、式 (2.19) において  $K = 2$  として計算する。今、新たに求めたパルスの位置を  $m_s$ 、その位置に最も近いパルスの位置を  $m_t$  とすると、各々のパルスの振幅  $g_s$ 、 $g_t$  は次式から計算できる。

$$\begin{bmatrix} R_{hh}(0) & R_{hh}(|m_s - m_t|) \\ R_{hh}(|m_t - m_s|) & R_{hh}(0) \end{bmatrix} \begin{bmatrix} g_s \\ g_t \end{bmatrix} = \begin{bmatrix} R_{hz}(m_s) \\ R_{hz}(m_t) \end{bmatrix} \quad (2.20)$$

方式 2-1 の特徴を表 2.1 に、演算量を表 2.2 に示す。振幅計算には、式 (2.20) の計算に  $8K$  必要である。また、評価関数  $R(m)$  の修正には、2 個のパルスを用いて修正を行なうので、方式 1 の 2 倍の  $4LK$  の演算量が必要となる。しかし、方式 2 に比べると大幅に低減化されていることがわかる。

方式 2-2 は、振幅を調整するパルスの個数を 3 としている。つまり、新たなパルスを求めると、そのパルスの振幅と、そのパルスの位置の近傍の過去の 2 パルスの振幅

とを、式(2.19)において  $K = 3$  とおいて計算する。今、新たに求めたパルスの位置を  $m_s$ 、その位置に最も近い2つのパルスの位置をそれぞれ、 $m_t$ 、 $m_u$  とすると、各々のパルスの振幅  $g_s$ 、 $g_t$ 、 $g_u$  は次式から計算できる。

$$\begin{bmatrix} R_{hh}(0) & R_{hh}(|m_s - m_t|) & R_{hh}(|m_s - m_u|) \\ R_{hh}(|m_t - m_s|) & R_{hh}(0) & R_{hh}(|m_t - m_t|) \\ R_{hh}(|m_u - m_s|) & R_{hh}(|m_u - m_t|) & R_{hh}(0) \end{bmatrix} \begin{bmatrix} g_s \\ g_t \\ g_u \end{bmatrix} = \begin{bmatrix} R_{hz}(m_s) \\ R_{hz}(m_t) \\ R_{hz}(m_u) \end{bmatrix} \quad (2.21)$$

特徴を表2.1に、演算量を表2.2にまとめる。方式2-2の演算量は、方式2と比較すると大幅に少なく、方式2-1よりは若干多いが、性能は2.4.5節、2.4.6節で述べるように、方式2に近い性能が得られる。

## 2.4 実験

本章で提案した4種のパルス探索法を用いたマルチパルス音声符号化法について、実験により性能、音質を比較した。

### 2.4.1 評価尺度

評価尺度には、符号化の物理的な性能を測定する客観評価尺度と、音質を評価する主観評価尺度の2種類を用いた。パラメータを変化させたときの性能の測定に客観評価尺度を使用し、最終的な音質判断に主観評価を使用する。将来的には、主観評価値と極めて高い精度で対応づけられる客観評価手法が開発されれば、労力を必要とする主観評価は不要となろうが、現状では精度が不十分なため、客観評価と主観評価を併用する必要がある。

### 客観評価尺度

客観評価尺度としては、時間領域の尺度としてセグメンタル SNR ( $SNR_{seg}$ : Segmental Signal to Noise Ratio) を、周波数領域の尺度としてケプストラム距離 ( $CD$ : Cepstrum Distance) を使用した [179]。これらの尺度は、他の尺度に比べて、主観音質評価値との対応が比較的良好なことが報告されている [179], [60]。

(1) セグメンタル SNR ( $SNR_{seg}$ ): 32 ms のセグメント毎に、入力信号のパワと量子化雑音のパワの比として  $SNR$  を dB 値で求め、入力音声の  $RMS$  (Root Mean Square) があらかじめ定められたしきい値を越えた区間について  $SNR$  を下式により平均することにより算出した。このような処理を行なうのは、 $RMS$  が高い区間には音声の母音部などが含まれ音質への影響が大きいと考えられ、逆に  $RMS$  が低い区間は無音や子音部で音質への影響が少ないと考えられるためである。

$$SNR_{seg} (dB) = \frac{1}{M} \sum_{j=1}^M SNR_j \quad (2.22)$$

$$SNR_j (dB) = 10 \log_{10} \frac{\sum_n x_j^2(n)}{\sum_n [x_j(n) - s_j(n)]^2} \quad (2.23)$$

ここで、 $x_j(n)$ 、 $s_j(n)$  はそれぞれ、 $j$  番目のセグメントの入力音声、 $j$  番目のセグメントの再生音声を示す。

(2) ケプストラム距離 ( $CD$ ): 入力音声と符号化再生音声のスペクトル包絡の違いを評価する尺度であり、下式により求めた。

$$CD (dB) = \frac{10}{\ln 10} \sqrt{2 \sum_{n=1}^{16} [c(n) - c'(n)]^2} \quad (2.24)$$

ここで、 $c(n)$ 、 $c'(n)$  はそれぞれ、入力音声及び符号化音声のケプストラム係数である。

表 2.3: 5段階品質評価尺度

評点	絶対品質尺度 (ACR) の評定値	妨害品質尺度 (DCR) の評定値
5	非常に良い	レファレンスに比べ劣化が全く認められない
4	よい	レファレンスに比べ劣化が認められるが気にならない
3	まあよい	レファレンスに比べ劣化がわずかに気になる
2	悪い	レファレンスに比べ劣化が気になる
1	非常に悪い	レファレンスに比べ劣化が非常に気になる

#### 主観評価尺度

主観評価尺度としては、5段階のオピニオン評価尺度を用いた。オピニオン評価には絶対品質尺度 (ACR: Absolute Category Rating) と妨害品質尺度 (DCR: Degradation Category Rating) などがある。5段階評価尺度の一例を表 2.3 に示す。

ここで行なうのは音質の良さの評価であるため、ACR を用いた。各被験者に符号化音声ランダムな順番で聴取させ、その音質の良さを5段階のスケールで評価させてオピニオン値を求め、オピニオン値を被験者の人数で平均し、符号化方式毎に平均オピニオン値 (MOS: Mean Opinion Score) を算出した [179]。なお、これらの評価尺度は、ITU-T (International Telecommunications Union-Telecommunications Standards Sector) において標準方式が P.80 として勧告化されている [46]。

#### 2.4.2 予備実験

合成フィルタの次数、フレーム長について、 $SNR_{seg}$ 、 $CD$  を用いて予備実験を行った。なお、予備実験では伝送情報の量子化は行っていない。

#### 合成フィルタの次数

フレーム長は 20 ms とし、合成フィルタの次数を変化させた。線形予測分析は自己相関法 [146] を使用した。結果を図 2.6 に示す [108]。図から、フィルタ次数が 12 ~ 16 では  $SNR_{seg}$ 、 $CD$  共に大きな変化はないが、次数を 10 以下とすると  $CD$  は大きく劣化する。これは、 $CD$  がスペクトル包絡の誤差に重きをおく評価尺度であることによる。この結果から、フィルタ次数が 8 程度では音質の劣化が予想されるため、以降では 12 次を選んだ。

#### フレーム長

合成フィルタの次数は 12 とし、フレーム長を変化させた。フレーム当りのサンプル数とパルスの個数との比率は一定とした。結果を図 2.7 に示す [108]。フレーム長を変えても  $SNR_{seg}$ 、 $CD$  共にほとんど差がないことがわかる。フィルタの伝送ビットレートはフレーム長が長いほど少なくなるが、符号化・復号化の処理遅延が長くなるので、以降ではフレーム長として 20 ms を選択した。

#### 2.4.3 パルス数と $SNR_{seg}$ の関係 (方式 1)

基本特性として、パルス探索法に方式 1 を用いたときの、フレーム当りのパルス数と  $SNR_{seg}$  の関係を図 2.8 に示す [112]。フレーム長は 20 ms である。図から、パルス数の増加による  $SNR_{seg}$  の増加は、パルス数が 16 以下のときの方が、16 以上のときよりも大きいことがわかる。これは、方式 1 は、パルスの振幅を再調整していないために、個数が多くなると性能がのびにくくなるためと考えられる。なお、2.4.4 節で示す伝送情報の量子化を施すと、30 パルスは 16 kb/s、16 パルスは 9.6 kb/s の伝送速度に相当する。

#### 2.4.4 伝送情報の量子化

ビットレートとして、16 kb/s と 9.6 kb/s の 2 種類を選んだ。伝送情報には表 2.4 に示す様な量子化を施した。フレーム長は 20 ms、合成フィルタの次数は 12 とした。式 (1.1) における聴感重み付けの係数は  $\gamma_1 = 1.0$ 、 $\gamma_2 = 0.9$  とした。フィルタ係数を

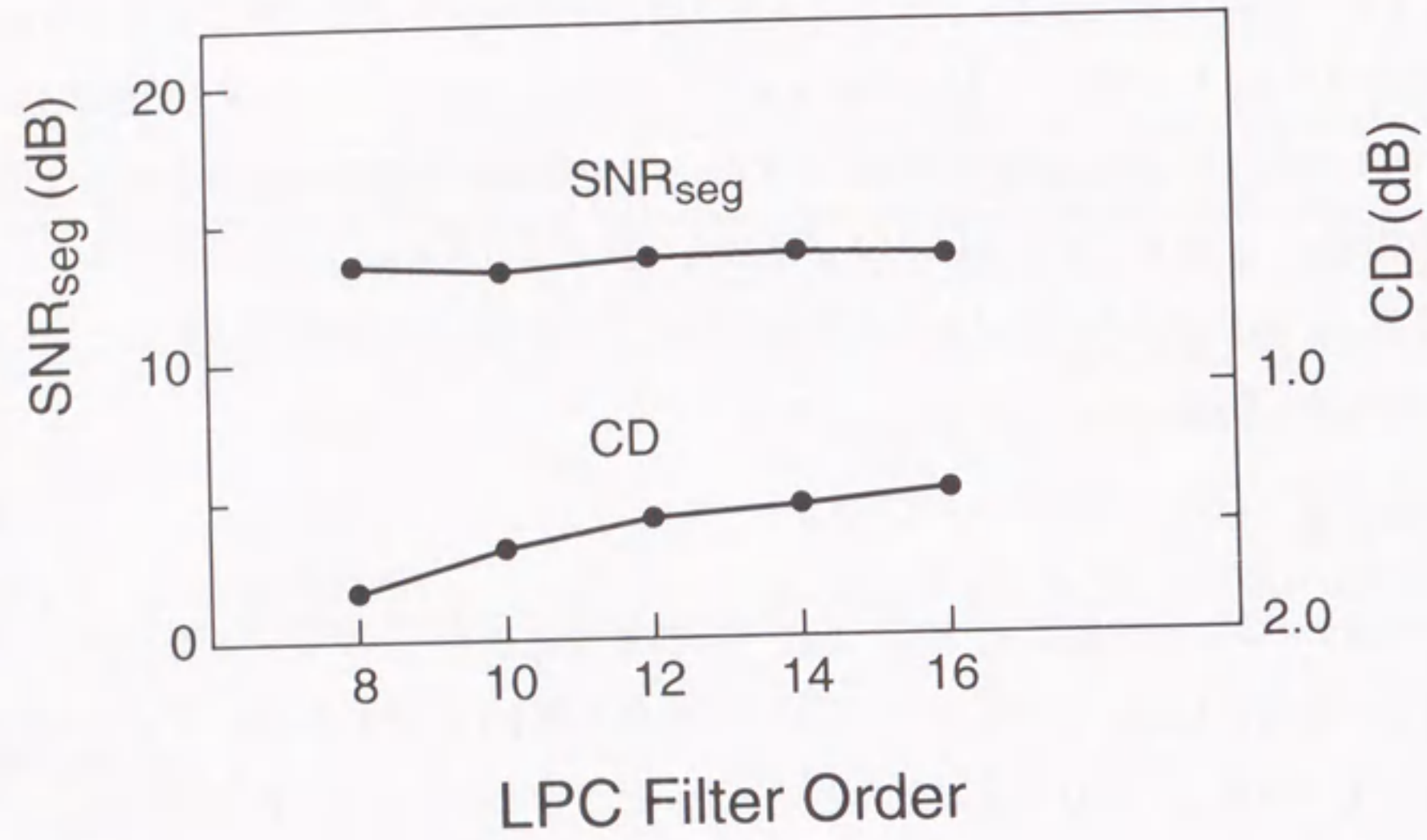
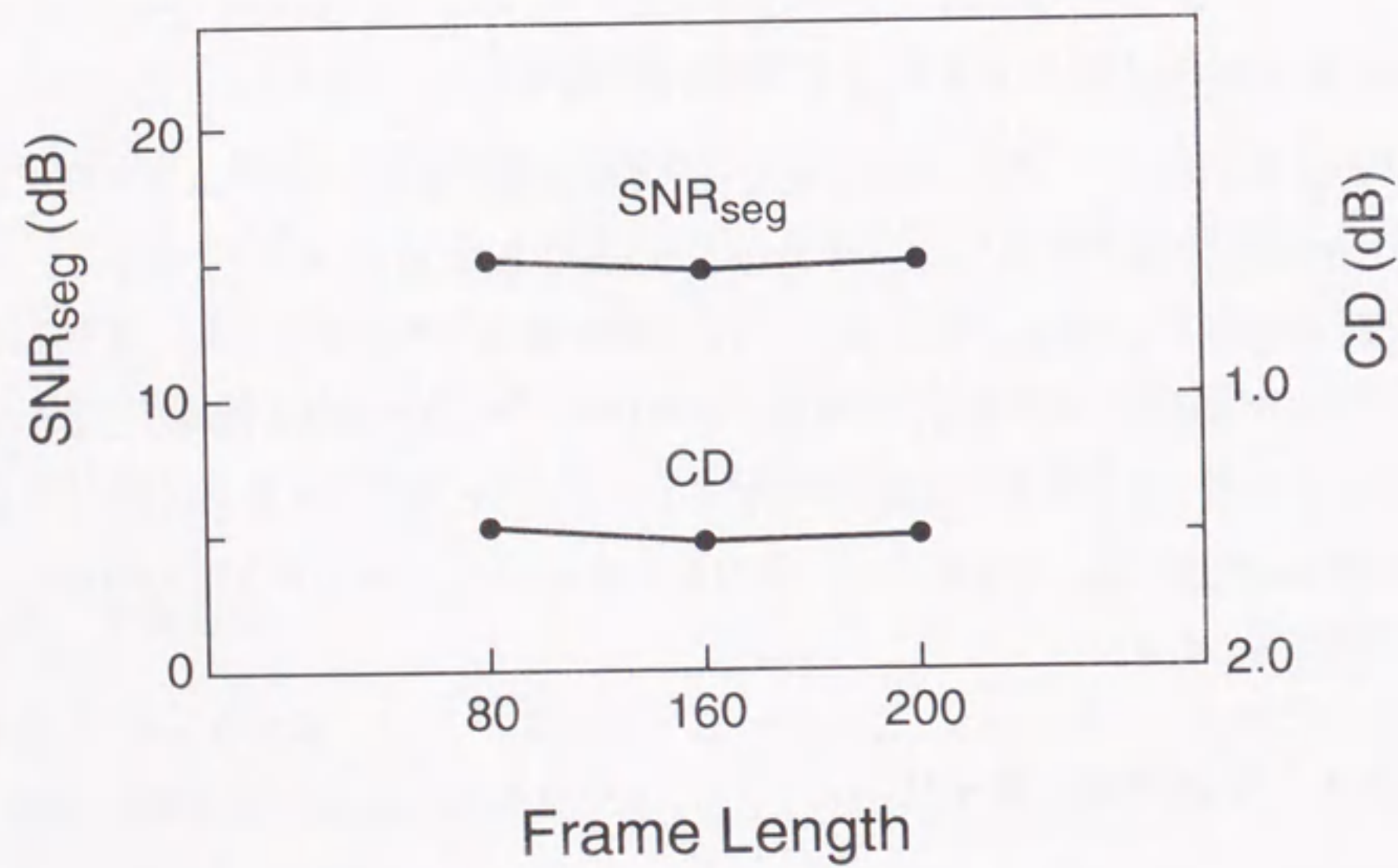
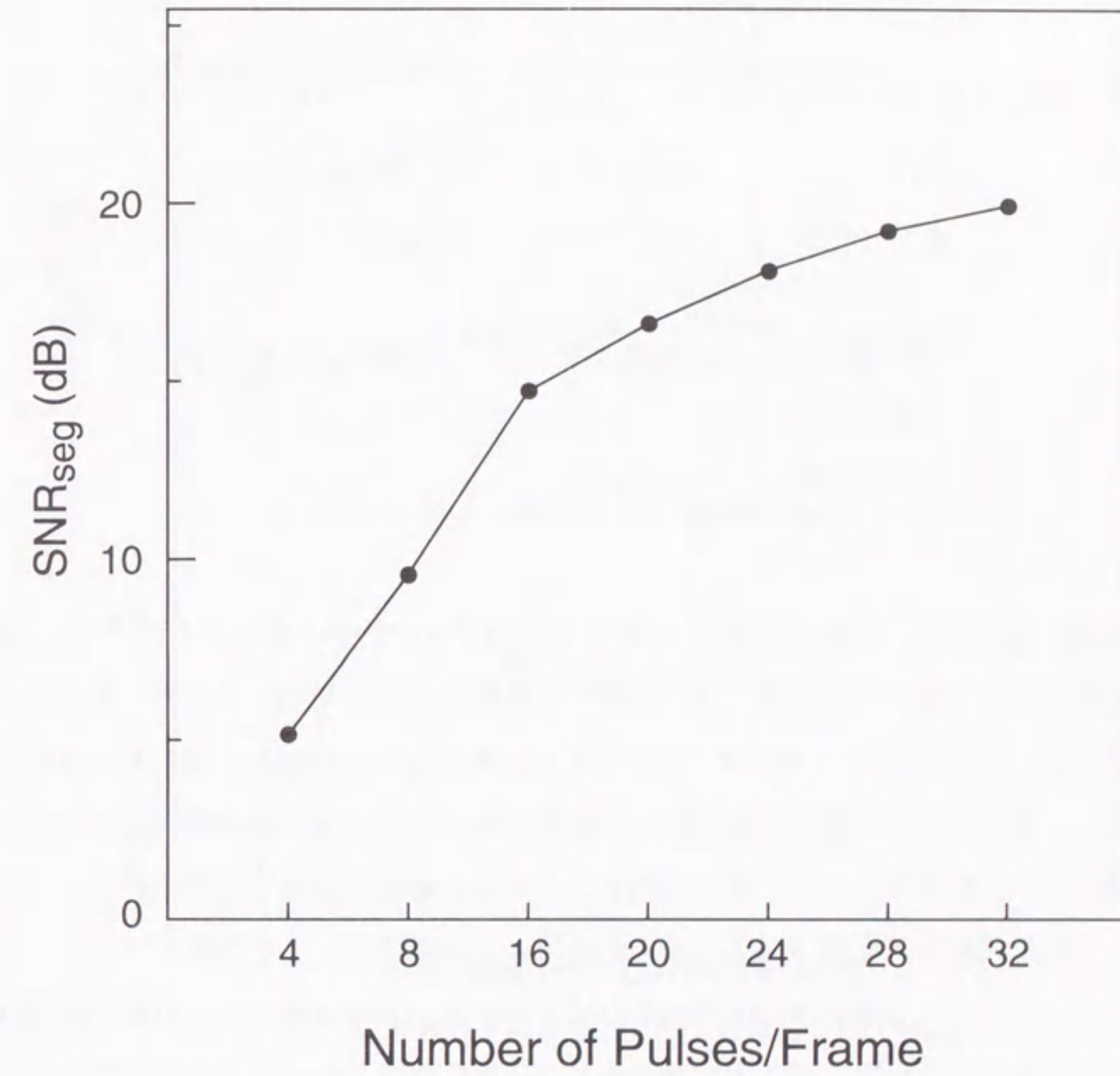
図 2.6: フィルタの次数と  $SNR_{seg}$ ,  $CD$  との関係図 2.7: フレーム長と  $SNR_{seg}$ ,  $CD$  との関係図 2.8: パルスの個数と  $SNR_{seg}$  との関係 (方式 1)

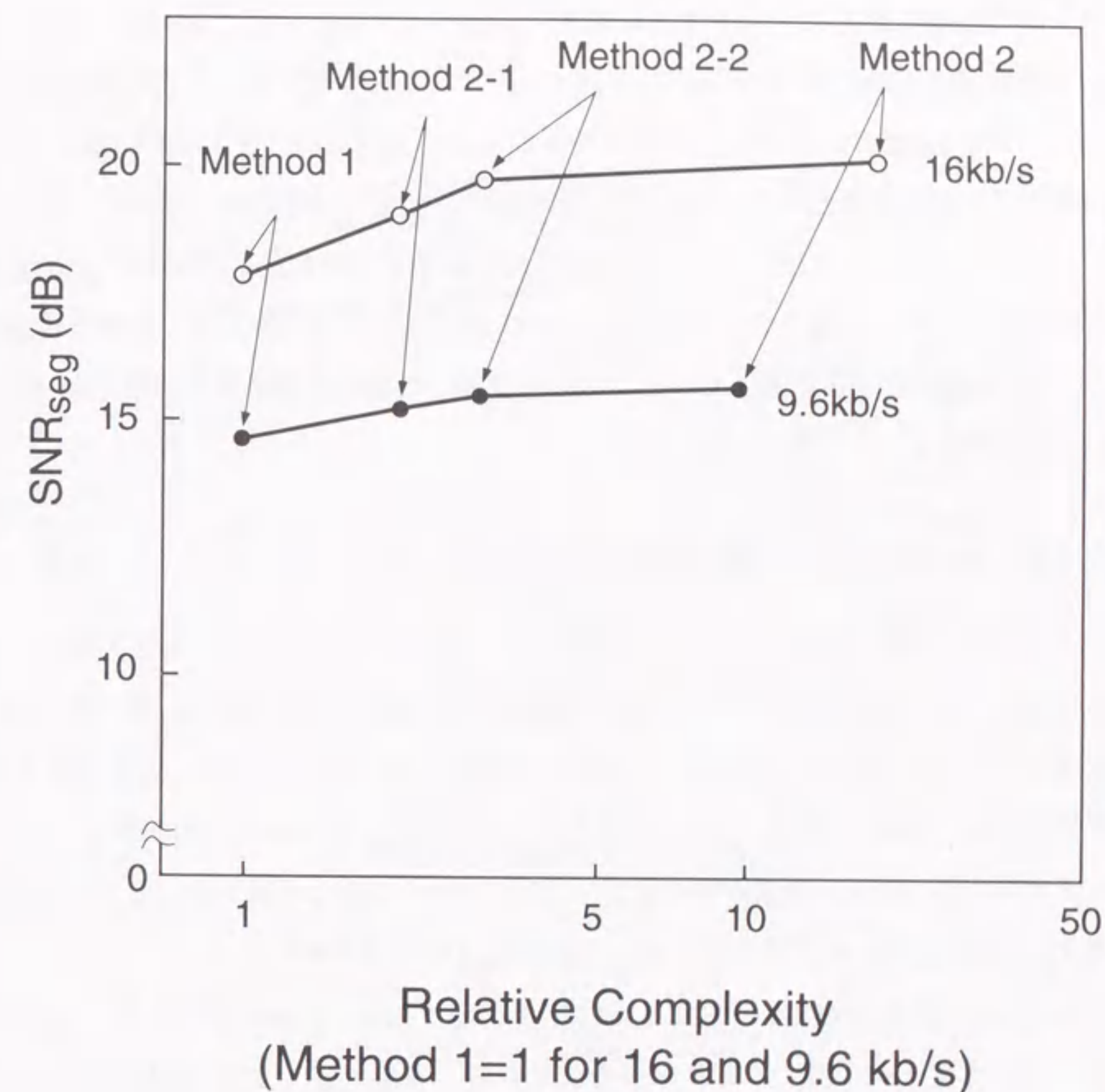
表 2.4: 伝送パラメータの量子化ビット配分

パラメータ	ビット数	
	16 kb/s	9.6 kb/s
K パラメータ		
K1 ~ K4	7,6,5,4 ビット	7,6,5,4 ビット
K5 ~ K12	各 3 ビット	3,3,2,2,2,2,2,2 ビット
パルス数	30	16
パルス最大値	対数圧伸 7 ビット	対数圧伸 7 ビット
パルス振幅	各 5 ビット	各 4 ビット
パルス位置	各 4 ビット	各 5 ビット

直接量子化すると 1 係数当たり 8 ~ 10 ビットのビット数が必要であるため [5]、効率的な量子化が可能な K パラメータ (PARCOR 係数 [44] と等価) に変換した後にスカラ量子化した。低次の K パラメータは量子化に敏感なため感度を考慮した非線形変換を施した後に分布を考慮して量子化し、高次の K パラメータは分布を考慮した量子化を施し、さらに不均一ビット割当を施すことにより効率的な量子化を実現した [59]。パルスの振幅の量子化にはスカラ量子化を用いた。量子化により振幅の大きなパルスがクリップされて性能が劣化するのを防ぐために、まず、フレーム内の絶対値振幅最大のパルスを対数圧伸 7 ビットで量子化して正規化係数とし、各パルスの振幅を正規化係数で正規化した後に、表 2.4 に示すビット数で量子化した [116]。パルスの位置は、各パルス間の距離を表 2.4 に示すビット数で表した。

#### 2.4.5 4 種のパルス探索法の SNRseg 性能と演算量の比較

2.3 節で述べた 4 種のパルス探索法について、 $SNR_{seg}$  性能と演算量を比較した。図 2.9 に示す [113]。図において、各探索法の演算量は、方式 1 の演算量で正規化して

図 2.9: 4 種のパルス探索法の  $SNR_{seg}$  と演算量の比較

表示してある。評価用音声には日本語6短文を使用した。

図から、16 kb/sでは、パルス探索法が方式1から2へ改善されるに従い、 $SNR_{seg}$ は大幅に改善されており、方式2は方式1に比べ $SNR_{seg}$ が約2 dB高い。方式2の演算量は方式1の約20倍である。方式2-2、方式2-1の演算量は、それぞれ方式1の3倍、2倍であるが、 $SNR_{seg}$ はそれぞれ約2 dB、約1 dB改善されている。従って、演算量と性能を考慮すると、方式2-2、方式2-1が良好であるといえよう。

一方、9.6 kb/sでは、パルス探索法の改良による $SNR_{seg}$ の改善は、16 kb/sほど顕著ではない。方式2による改善は約0.7 dB、方式2-2、方式2-1ではそれぞれ、約0.7 dB、約0.5 dBであった。ここでも演算量と性能を考慮すると方式2-2、方式2-1が有力である。

#### 2.4.6 4種のパルス探索法の音質比較

5段階のACR尺度によるMOS評価を行い、パルス探索法の音質を比較した。基準音質を与える符号化として、 $\mu$ 法則対数圧伸( $\mu$ -law)PCM 64, 56, 48, 40, 32 kb/sを用い、これらを含めて評価した。被験者は男女合計10名である。結果を図2.10に示す[116]。16 kb/sでは、方式2-2と方式2は同等で、 $\mu$ -law PCM 56 kb/s (7ビット/サンプル)に近い音質が得られている。方式2、方式2-2は方式1よりもMOSで約0.3改善される。方式2-1と方式1の音質はほぼ同等である。

9.6 kb/sでも方式2-2と方式2は同等の音質であり、 $\mu$ -law PCM 52 kb/s (6.5ビット/サンプル)程度の音質が得られている。また方式1よりもMOSで約0.1改善される。

以上より、演算量の増加をできるだけ抑えながら最大の改善効果を得るには、方式2-2が有効と考えられる。

#### 2.4.7 パルス探索法2-2による波形例

パルス探索法として方式2-2を用いたときの波形の一例を図2.11と図2.12に示す。図2.11は16 kb/s、図2.12は9.6 kb/sの例である。いずれの図も(a)は原波形(女声話者200 ms)、(b)は再生波形、(c)は誤差波形である。図から、16 kb/sの場合

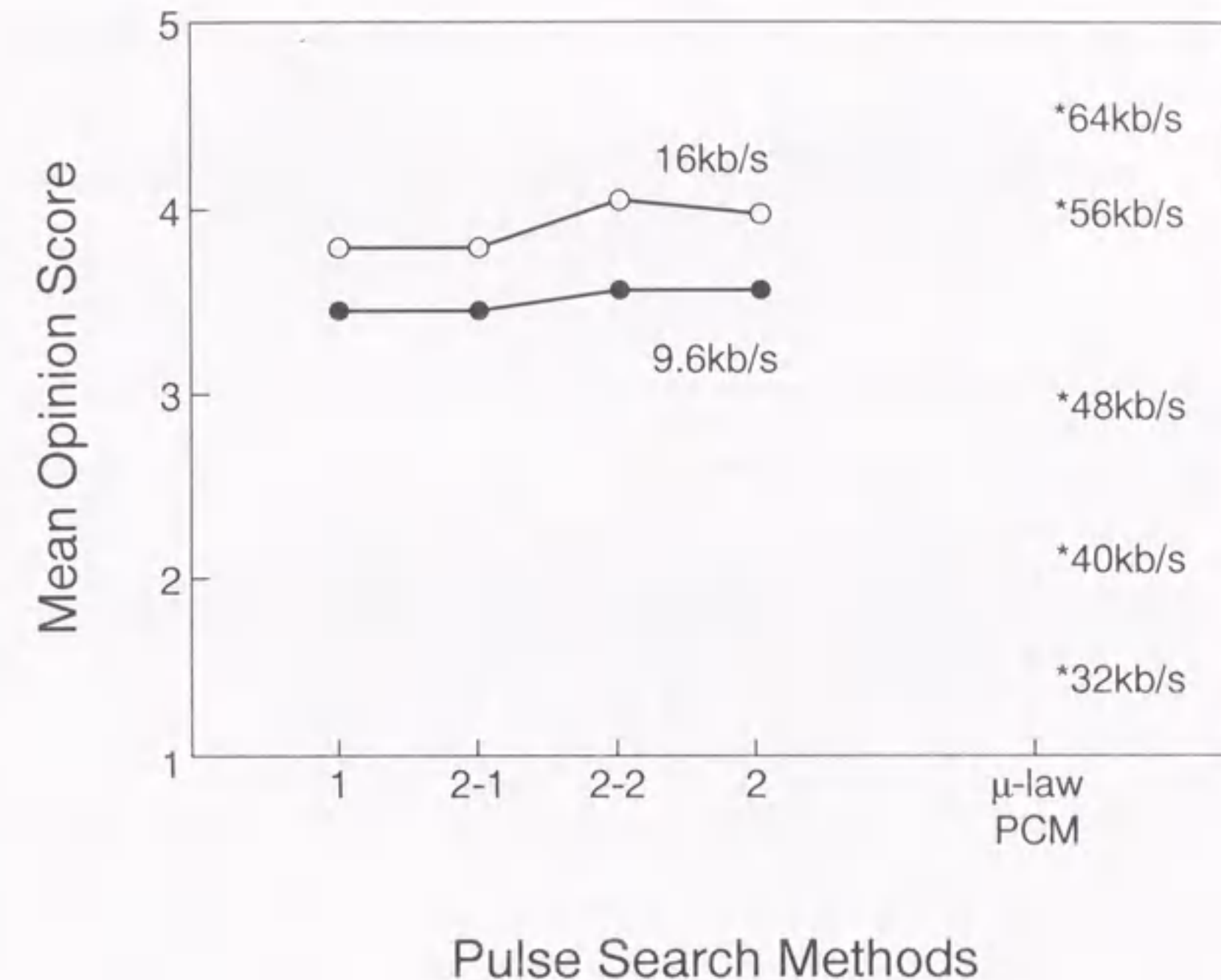


図 2.10: 4種のパルス探索法の音質比較

は誤差が十分小さくなっているが、9.6 kb/sでは誤差に低い周波数成分が残存しており、このような女声話者に対しては、パルスの個数が十分でないことを示している。

#### 2.4.8 考察

実験の結果、以下の点が明らかとなった。

(1) 方式1を用いてフレーム当りのパルス数を増加させていくと、特に20パルス以上において $SNR_{seg}$ の増加がゆるやかになる。これは、パルス数の増加に伴い、パルスの間隔が短くなり、パルス間の相互作用が無視できなくなることに起因すると考えら



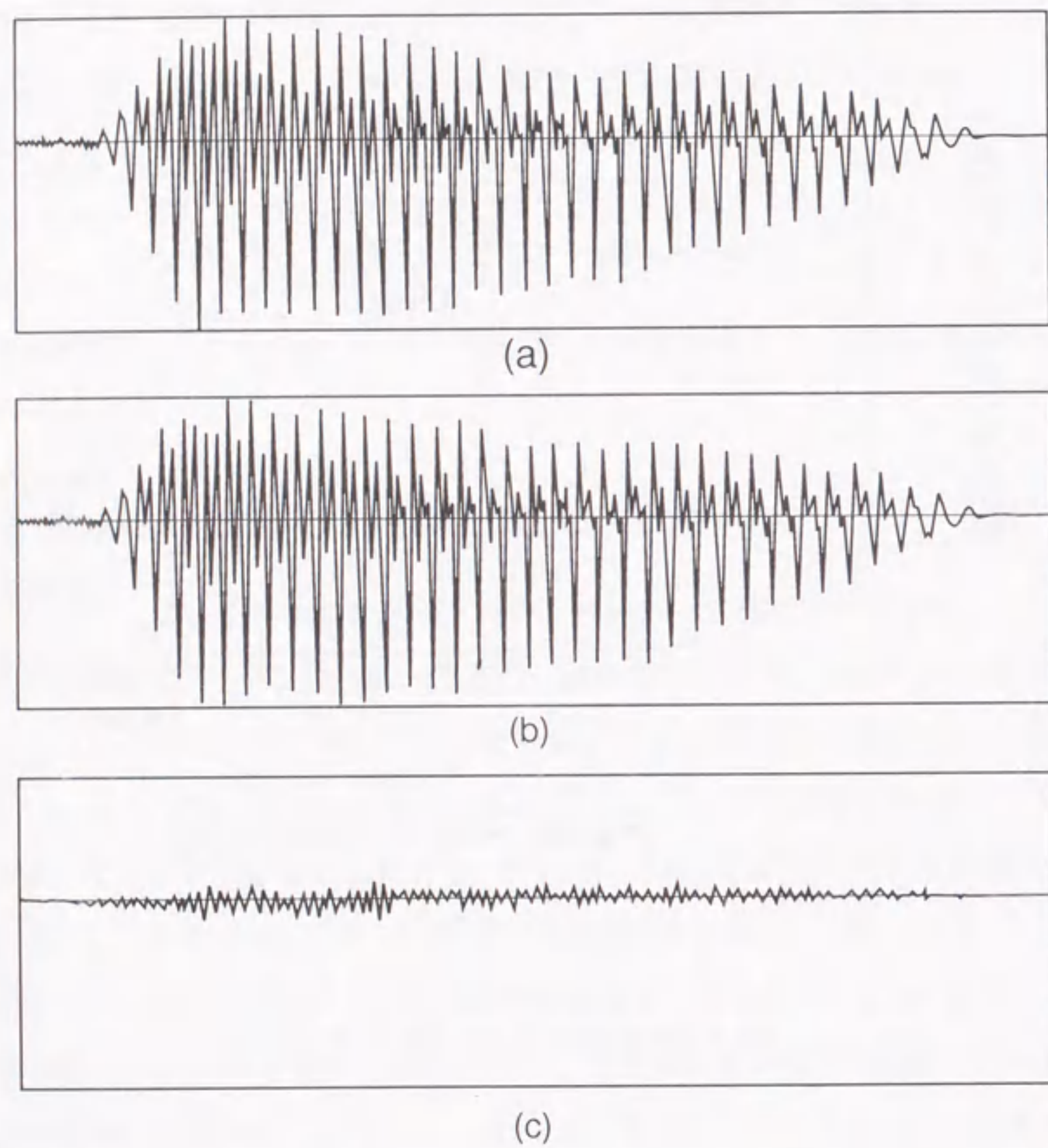


図 2.11: 方式 2-2 による 16 kb/s の波形の例, (a) 原音声波形 (女性話者 200 ms), (b) 再生波形, (c) 誤差波形。各図の縦軸は (a) の最大値、最小値で正規化されている。

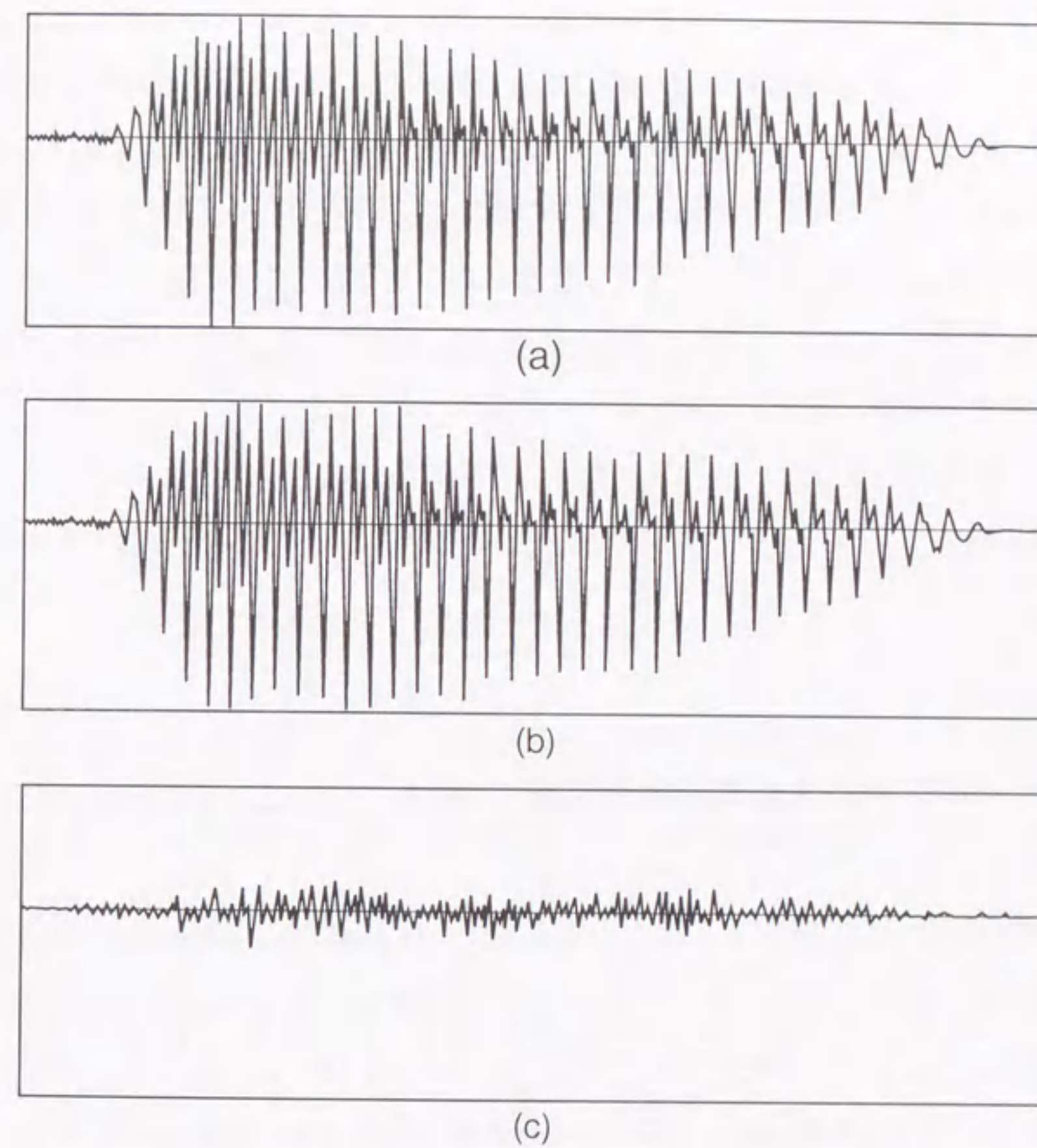


図 2.12: 方式 2-2 による 9.6 kb/s の波形の例, (a) 原音声波形 (女性話者 200 ms), (b) 再生波形, (c) 誤差波形。各図の縦軸は (a) の最大値、最小値で正規化されている。

れる。

(2) 方式1による音質は、16 kb/sで $\mu$ -law PCM 54 kb/s相当、9.6 kb/sで $\mu$ -law PCM 51 kb/s相当の音質が得られた。

(3) パルス探索法の改良による改善効果は、フレーム当りのパルス数が多いほど顕著となる。方式2-2と方式2はほぼ同じ性能であった。16 kb/sの場合（フレーム当り30パルス）では、方式1に比べ、 $SNR_{seg}$ で約2 dB、音質MOS値で約0.3と、大幅に改善された。これらの音質は $\mu$ -law PCM 56 kb/s相当であった。方式2-1は、方式1より $SNR_{seg}$ は向上したが、音質は改善されなかった。一方、9.6 kb/s（フレーム当り16パルス）では、方式2-2及び2による改善は $SNR_{seg}$ では0.7 dB、音質MOS値では0.1の改善であり、改善度は十分でなかった。

(4) 演算量を比較すると、16 kb/sでは、方式2は方式1に比べ約20倍、方式2-2は約3倍である。従って、性能と演算量を考えた場合、方式2-2が有効であると考えられる。

## 2.5 マルチパルス音声符号化装置の試作

NEC社製の汎用16ビット固定小数点DSP (Digital Signal Processor) LSI ( $\mu$ PD7720) を複数チップ使用してマルチパルス音声符号化装置を試作した。9.6 kb/sと16 kb/sの2種類のビットレートを切替えて使用できることを特徴とする[175]。使用したDSPの特徴[98]を表2.5に示す。DSPの実行速度や内蔵メモリ容量の制約から、フレーム長を20 msとして処理を複数のDSPに分担させた。送信側は4チップ、受信側は1チップを使用した。演算量の制約から音源パルスの探索には演算量の最も少ない方式1を使用し、さらに、フレームを5 msのサブフレームに分割しサブフレーム毎にパルスの探索を行なった。これによりパルス探索部はDSP 1チップで実現できた。各DSPの処理分担を表2.6にまとめる。また、試作したボードの外観を図2.13に示す。ほぼA4サイズとコンパクトな構成になっている。

今後は、多重化通信や企業内デジタル通信などへの応用などが期待される。

表 2.5: マルチパルス符号化装置に使用した DSP ( $\mu$ PD7720) の特徴

命令実行速度	250 ns
演算精度	16 ビット固定小数点
命令 ROM 容量	512 ワード
データ ROM 容量	510 ワード
データ RAM 容量	128 ワード
並列乗算器	16 ビット x 16 ビット

表 2.6: マルチパルス符号化装置における DSP の処理分担

DSP (S) 1	LPC 分析、PARCOR 係数、自己相関関数
DSP (S) 2	相互相関関数
DSP (S) 3	音源パルスの探索
DSP (S) 4	PARCOR 係数及び音源パルスの符号化
DSP (R) 1	受信データ復号、音声合成

S は送信側、R は受信側を示す

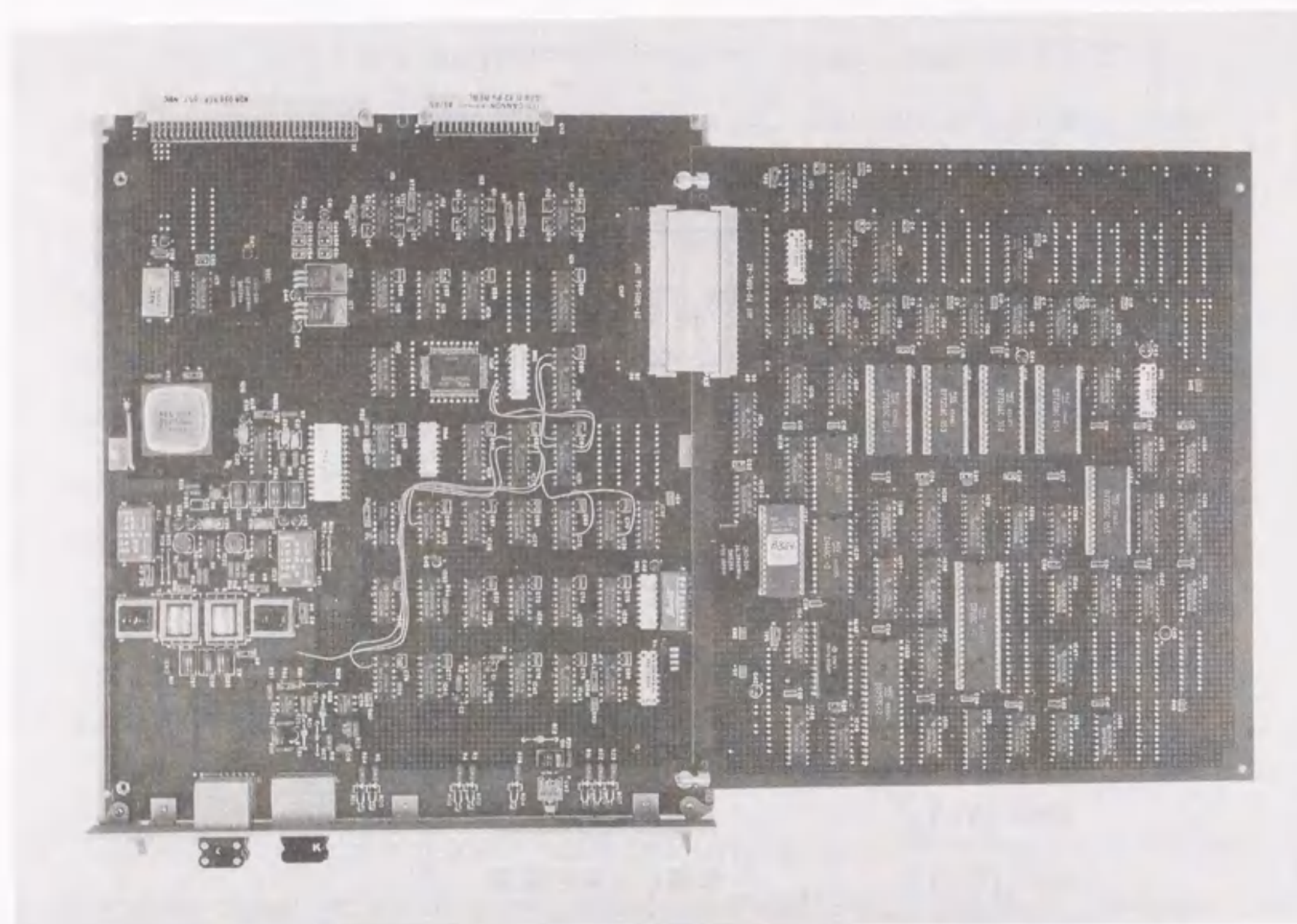


図 2.13: 試作した 16 kb/s, 9.6 kb/s マルチパルス音声符号化ボードの外観

## 2.6 むすび

マルチパルス音源の振幅と位置を逐次探索する方法として、相関演算に基づく演算量の極めて少ない効率的な方法から、さらに性能を改善する方法まで、4種類の方法（方式1、方式2、方式2-1、方式2-2）を提案した。これらの方法はいずれも、マルチパルスにより合成した音声と入力音声との聴感重み付け誤差電力の最小化基準から導かれることを示した。これらのパルス探索法を用いるマルチパルス音声符号化を構築し、16 kb/s、9.6 kb/sにおいて、 $SNR_{seg}$  性能、音質と演算量の比較検討を行った。16 kb/sではパルス探索法の改善による効果は大きく、方式2-2、方式2によれば、 $\mu$ -law PCM 56 kb/sに近い極めて良好な音質が得られた。9.6 kb/sでは、この改善度は充分ではなかったが、方式2-2、方式2によれば、 $\mu$ -law PCM 52 kb/s相当の音質が得られた。16 kb/sの演算量は、方式1に比べ、方式2-2は約3倍、方式2は約20倍であった。これらから、性能と演算量を考慮すると方式2-2が有効であると考えられる。また、16ビット固定小数点DSPを複数チップ使用して16 kb/s、9.6 kb/sデュアルレートのマルチパルス符号化装置をA4サイズのボード上に試作した。

9.6 kb/s以下で音質をさらに改善するためには、パルス数の不足を補う必要がある。これについては、3章、4章において、ピッチ情報を利用することによりさらに改善をはかりたい。

## 第3章

### ピッチ予測を用いる 9.6 kb/s ~ 8 kb/s マルチパルス 音声符号化

#### 3.1 まえがき

第2章では、マルチパルス音声符号化方式において、4種のパルス探索アルゴリズムを提案した。そして、16 kb/s と比較的ビットレートが高く、フレーム当りのパルス数が多い場合に改善効果が大きいことを示した。しかし、ビットレートを 9.6 kb/s 以下に低減化すると、パルス探索アルゴリズムによる改善効果は充分ではなかった。また、マルチパルスの個数が不足してくるために、特にピッチ周期の短い女性話者の音声に対して、ビットレートの低下に伴い、音質が急速に劣化するという問題点があった。

本章と次章は、ビットレートの低減化を目的とし、これらの問題点を改善するために、音声信号のピッチ周期毎の準周期性を利用して音源信号を効率的に表す、ピッチ情報を用いるマルチパルス音声符号化方式を提案する。ピッチ情報の適用の仕方として、本章はピッチ予測を検討し、次章はピッチ同期マルチパルス音源を考察する。本章で提案するピッチ予測の基本方式(3.2.1節のP-1法)は、従来のAPC[8]などに使用されているピッチ予測とは構成が異なり、また、ほぼ同時期に他の機関で独立に検討されたピッチ予測マルチパルス方式[157]とも異なる独自のものである。

以降では、9.6 kb/s から 8 kb/s で有効な方式として、3種類のピッチ予測マルチパルス符号化方式を提案し、性能を客観的、主観的に比較評価する。

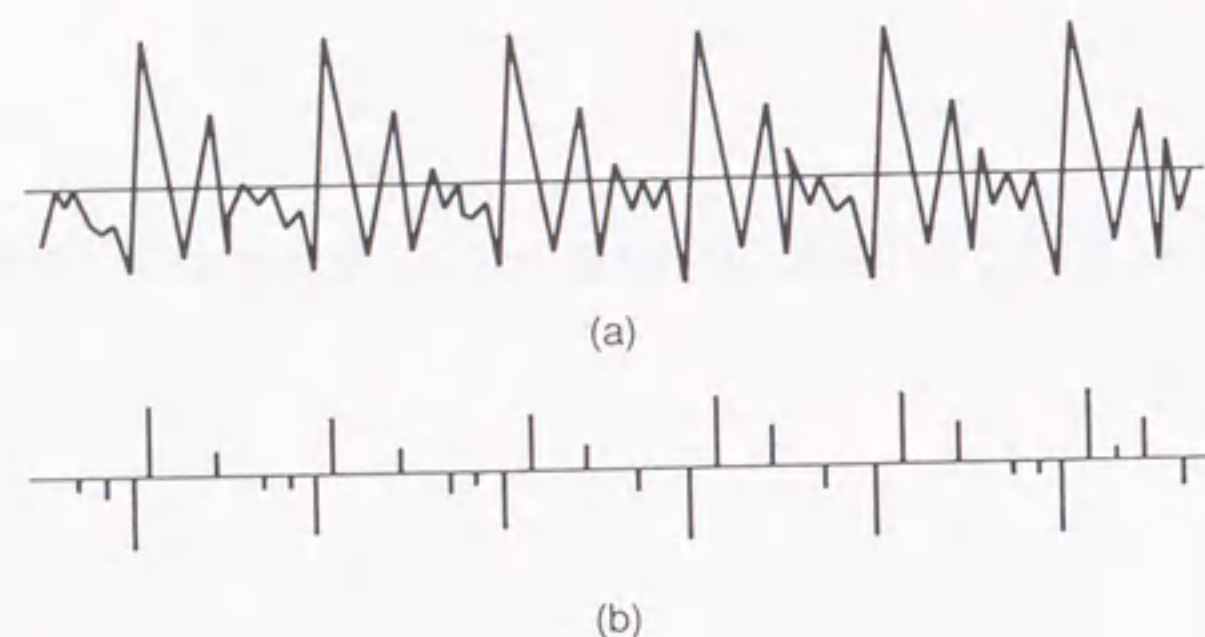


図 3.1: 有声区間での波形例, (a) 有声区間のピッチ周期波形, (b)(a) の区間で求めたマルチパルス

### 3.2 ピッチ予測マルチパルス符号化方式

有声区間では、音声信号にはサンプル値間の近接相関のみならず、図 3.1(a) に示すようにピッチ周期毎の長期相関が観測される。このような区間でマルチパルスを求めると、図 3.1(b) に示すように、振幅の大きなパルスはピッチ周期毎にほぼ周期的にならぶ。従って、このようなパルスの準周期性を利用したピッチ予測を行うことにより、効率的な符号化が可能となる。以下では、3 種類のピッチ予測マルチパルス符号化法を提案し、比較検討する。

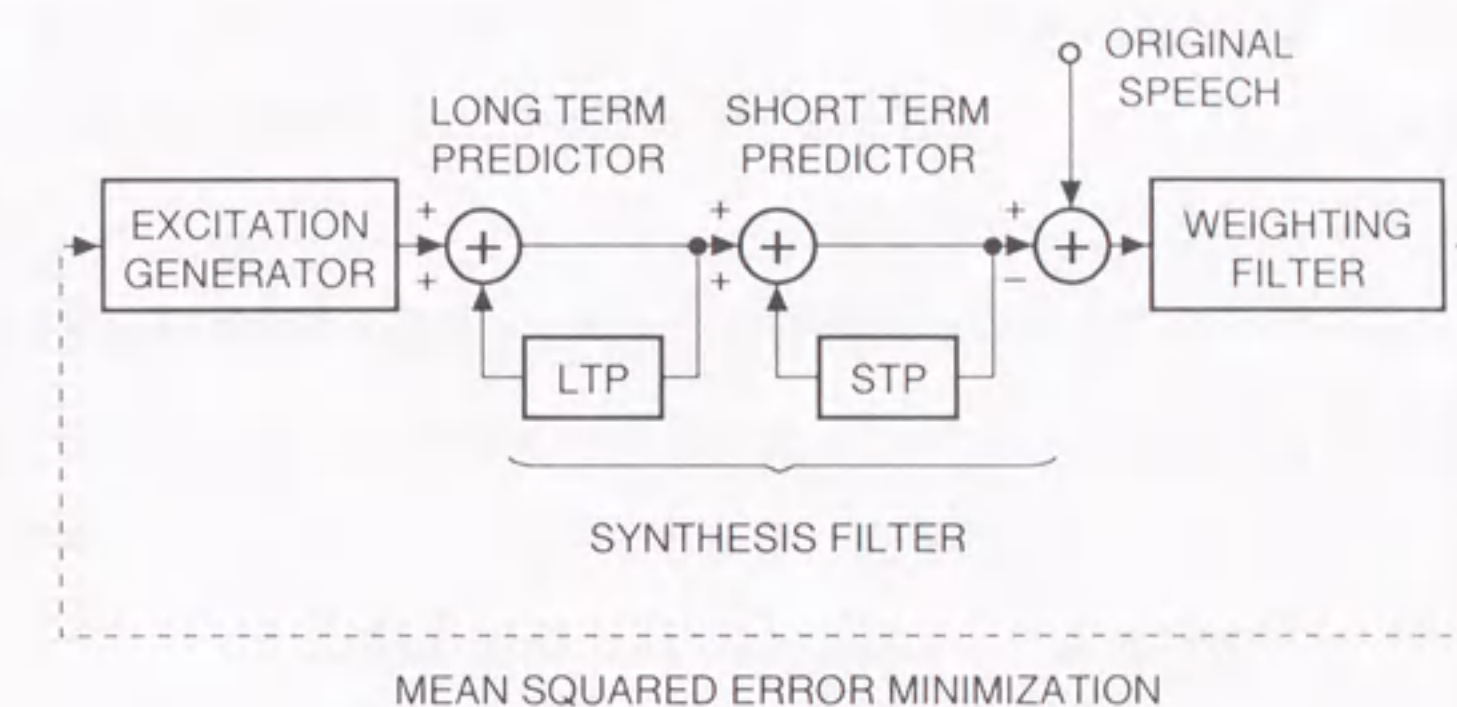


図 3.2: ピッチ予測マルチパルス符号化 (P-1 法) を示すブロック図

#### 3.2.1 ピッチフィルタのインパルス応答をパルス探索に含めるフレーム内ピッチ予測法 (P-1 法)

##### 原理

本方式の送信側のブロック図を図 3.2 に示す [109],[116]。フレーム長は 20 ms 程度を考えているため、フレーム内には複数個のピッチ波形が存在する。フレーム区間でピッチ予測を行うために、本方式では、合成フィルタをマルチパルスのピッチ周期成分を生成するピッチ再生フィルタ (LTP: Long Term Predictor) と音声のスペクトル包絡を合成するスペクトル包絡合成フィルタ (STP: Short Term Predictor) との縦続接続により構成する [109],[116],[157]。本手法が APC などで使用されている従来のピッチ予測 [8] や、ほぼ同時に独立に他の機関で検討された方式 [157] と異なる

のは、ピッチ再生フィルタを含む合成フィルタをパルス探索の誤差最小化ループ内に配置し、LTP の特性と STP の特性の両方を合成フィルタのインパルス応答に含めた上でマルチパルスの探索を行う点である。インパルス応答に LTP 成分を含めることにより、マルチパルスの探索時に、ピッチ周期だけ離れた位置にパルスが生成され、パルスの個数が等価的に増大することになる。信号の再生時には、同一の LTP を接続して信号を再生するので、自動的にピッチ周期だけ離れた位置にパルスが生成される。

ここで、合成フィルタ全体の伝達特性  $H(z)$  は、 $z$  変換を使用して次式のように表すことができる。

$$H(z) = \frac{1}{1 - \sum_{i=-[Q/2]}^{[Q/2]} b_i z^{-(T+i)}} \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}} \quad (3.1)$$

ただし、式 (3.1) 右辺の第 1 項、第 2 項はそれぞれ、LTP、STP の伝達特性を示す。 $b_i$ 、 $T$ 、 $a_i$  はそれぞれ、LTP の係数、ピッチ周期、STP の係数を示す。 $Q$  は LTP の次数を示す。 $[Q/2]$  は整数値に切捨てを表す。

$b_i$ 、 $T$  の抽出法には、入力音声あるいは予測残差信号の自己相関から求める開ループ (open-loop) 法と、過去に求めた音源信号に基づく再生音声と入力音声との相互相関から求める閉ループ (closed-loop) 法がある [157],[68]。後者は前者より性能が良好であるが、演算量的に多いこと、P-1 法では、パルスを探索する際に LTP の特性を含め誤差を最小化するために、ピッチ抽出の細かな誤差はパルスの探索時に補正されることから、ここでは開ループ法を用いた。

#### 定式化

P-1 法では、LTP におけるピッチ再生処理は 1 フレーム内に限定して行うこととした [109],[116],[114]。以下で式を導出する。

LTP、STP を縦続接続した合成フィルタをマルチパルス音源により駆動して得た合成音声  $\tilde{x}(n)$  と原音声  $x(n)$  との聴感重み付け誤差の 2 乗平均値  $E$  は式 (3.2) のようにかける。

$$E = \sum_{n=0}^{N-1} [x(n) - \tilde{x}(n)]^2 \quad (3.2)$$

$$\tilde{x}_w(n) = \sum_{i=1}^K g_i h_w(n - m_i) \quad (3.3)$$

ここで、 $N$ 、 $K$ 、 $g_i$ 、 $m_i$  はそれぞれ、パルス探索の区間長、探索区間内のマルチパルスの個数、 $i$  番目のパルスの振幅と位置を示す。

聴感重み付け後の縦続接続合成フィルタのインパルス応答  $h_w(n)$  は、聴感重み付けフィルタのインパルス応答を  $w(n)$  とすると、次式のように表せる。

$$\begin{aligned} h_w(n) &= h(n) * w(n) \\ &= [h_s(n) * h_p(n)] * w(n) \\ &= [h_s(n) * w(n)] * h_p(n) \end{aligned} \quad (3.4)$$

$$h_{sw}(n) = h_s(n) * w(n) \quad (3.5)$$

$h_s(n)$ 、 $h_p(n)$  は、それぞれ、STP のインパルス応答、LTP のインパルス応答を示す。記号  $*$  は畳み込み演算を示す。

式 (3.5)、式 (3.4) は、STP のインパルス応答  $h_s(n)$  に聴感重み付けフィルタのインパルス応答  $w(n)$  を畳み込んで  $h_{sw}(n)$  を計算し、これに、LTP のインパルス応答  $h_p(n)$  を畳み込めば  $h_w(n)$  が求まることを意味する。ここで、聴感重み付けフィルタには、第 1 章の式 (1.1) で述べたものと同一のものを使用した。

$h_p(n)$  を畳み込む代わりに、 $h_{sw}$  から次式のようにしてインパルス応答  $h_w(n)$  を求めることもできる。簡単のために、LTP の次数を 1 とし、係数を  $b$  とすると、

$$h_w(n) = h_{sw}(n) + b h_{sw}(n - T) + b^2 h_{sw}(n - 2T) + \dots \quad (3.6)$$

となる。

P-1法では、縦続型合成フィルタのインパルス応答により、ピッチ周期離れた位置にパルスが生成され、等価的にパルスの個数を増加させる。式(3.4)や式(3.6)に基づきインパルス応答  $h_w(n)$  を求めることにより、パルス探索には第2章2.3節で提案した4種のパルス探索法をそのまま使用することができる。ここでは、ピッチ予測によりパルスの個数が増えるのでパルス相互の振幅を調整する必要があるためと最良の性能を引き出すため、パルス探索法として、第2章2.3.2節で述べた方式2を用いた。また、パルスを探索する区間長  $N$  に関しては、ピッチ周期  $T < N$  の話者に対してピッチ予測の効果が得られること、 $N$  が長い方がパルスの位置の自由度が増え性能が向上することが予想されることから、1フレーム全体とした。

#### パルス探索の波形例

P-1法によるパルス探索の波形例を図3.3に示す。(a)はフレーム(20ms)の音声波形、(b)は式(3.4)から求めたLTP、STP縦続接続合成フィルタのインパルス応答、(c)は式(3.2)を最小化するように求めたマルチパルス(フレーム当たり11本)、(d)は(c)とLTPから求めた励振音源、(e)は(d)による合成音声  $\hat{x}(n)$  を示す。(c)から、LTPによりフレームのパルスの個数が増加していることがわかる。

#### 送信側・受信側の構成

図3.4は、P-1法によるマルチパルス符号化の送信側、受信側のブロック図である。送信側から伝送する情報は、LTPのピッチ係数、周期、STPの係数と、音源パルスの振幅と位置である。

### 3.2.2 フレーム内・フレーム間ピッチ予測法(P-2法)

#### 原理

P-1法ではピッチ予測処理を1フレーム内に限定していた。この場合、ピッチ周期  $T < N$  の話者に対してはピッチ予測は効果的であるが、そうでない話者に対しては効

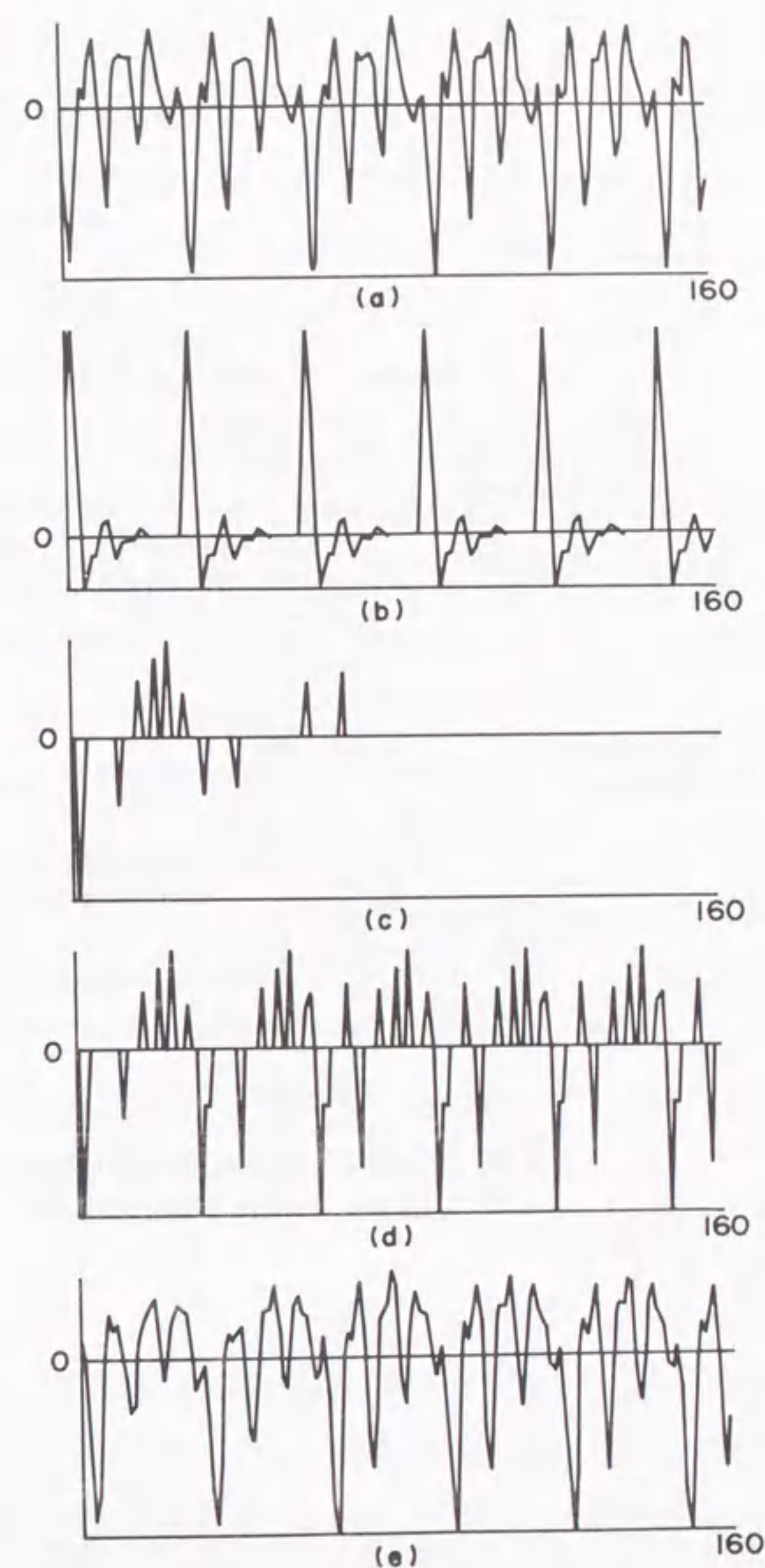


図3.3: ピッチ予測マルチパルス符号化(P-1法)による処理の波形例, (a) フレーム(20ms)の音声波形, (b) 式(3.4)から求めた縦続接続合成フィルタのインパルス応答, (c) 式(3.2)を最小化するように求めたマルチパルス, (d) (c)とピッチ再生フィルタ(LTP)から求めた励振音源, (e) 合成音声。ただし、図の縦軸は各図毎に正規化されている。

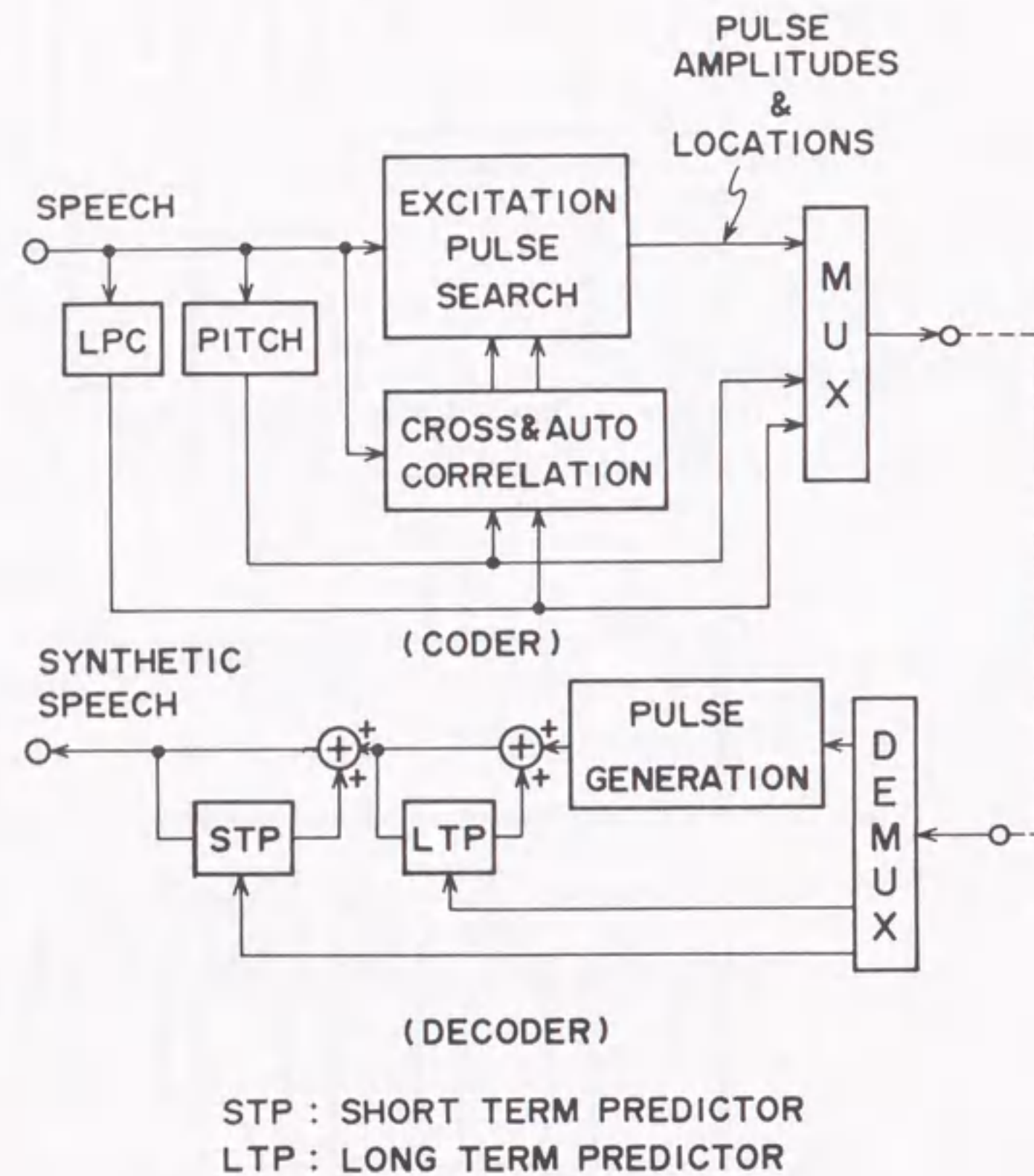


図 3.4: ピッチ予測マルチパルス符号化 (P-1 法) の送信側、受信側の構成を示すブロック図

果がなかった。

P-2 法は、この点を改善するために、P-1 法でのフレーム内ピッチ予測に加えて、過去のフレームの音源信号からのフレーム間ピッチ予測も併用した [114],[116]。

P-2 法の原理を示すブロック図を図 3.5 に示す。図は、一例として、フレーム内予測用、フレーム間予測用の 2 セットの縦続型合成フィルタ (F-0 及び F-1) を使用する例を示している。まず F-1 では、過去のフレームの音源信号を用いてフレーム間ピッチ予測が行われる。1 フレーム過去に求めた STP 入力音源信号を F-1 に通し、現在のフレームの音声信号  $x(n)$  のピッチ予測値を計算し、これを入力音声から減算し、フレーム全体で残差信号  $y(n)$  を求める。この考えは、処理区間長が異なるものの、[157] の手法や、近年、CELP において使用されている適応コードブック [62] の考えとほぼ同一である。

一方、フレーム長を 20 ms と比較的長くとしているため、F-1 によるフレーム間ピッチ予測のみでは、残差信号  $y(n)$  にはピッチ成分が残存してしまうことが予想される。そこで、この残存ピッチ成分をフィルタ F-0 で表す。フィルタ F-0 は、3.2.1 節の P-1 法による縦続接続型合成フィルタと同じ構成であり、フレーム内ピッチ予測を行う。マルチパルスとフィルタ F-0 により再生した信号と、残差信号  $y(n)$  との聴感重み付け誤差電力を最小化するように、フレーム内でマルチパルスを求める。パルスの探索には P-1 法と同様に方式 2 を用いた。

#### STP の共通化

P-2 法において、STP 係数、LTP 係数、ピッチ周期は F-0、F-1 の 2 セット分必要である。これら以外にもマルチパルスの位置、振幅を送送する必要がある。しかし、STP、LTP の係数を 2 セット分伝送するとサイド情報量が増大する。そこで予備実験の結果から、伝送情報量の多い STP の係数は、F-0、F-1 共に共通として入力音声信号  $x(n)$  から LPC 分析により求めたものを使用した。一方、情報量は少ないが性能改善に寄与する LTP のピッチ周期、係数は、F-0、F-1 別々に求めて伝送することにした。このときの LTP をそれぞれ、LTP-0、LTP-1 とする。この構成を示すブロック図を図 3.6 に示す。



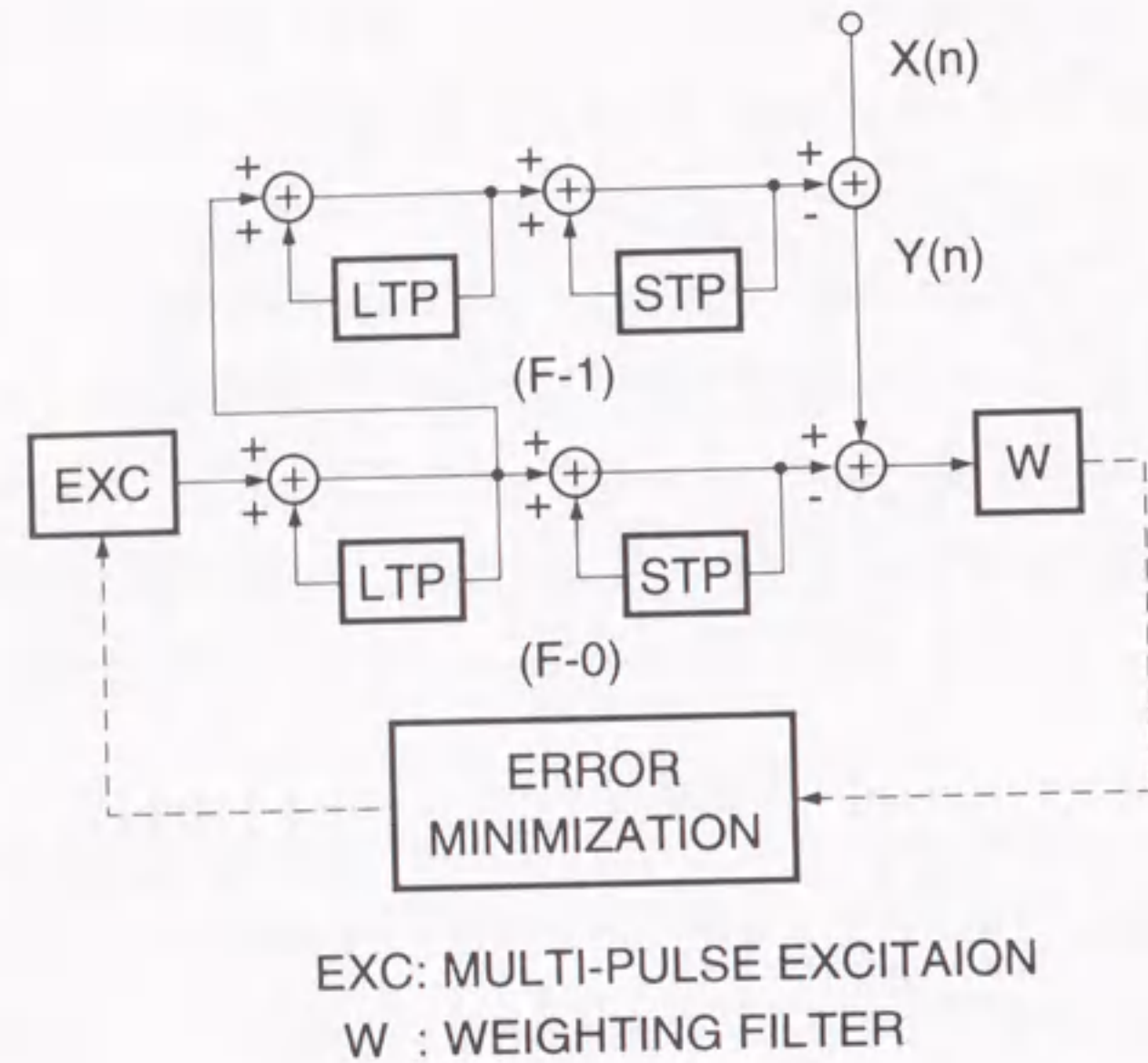


図 3.5: ピッチ予測マルチパルス符号化 (P-2 法) のブロック図

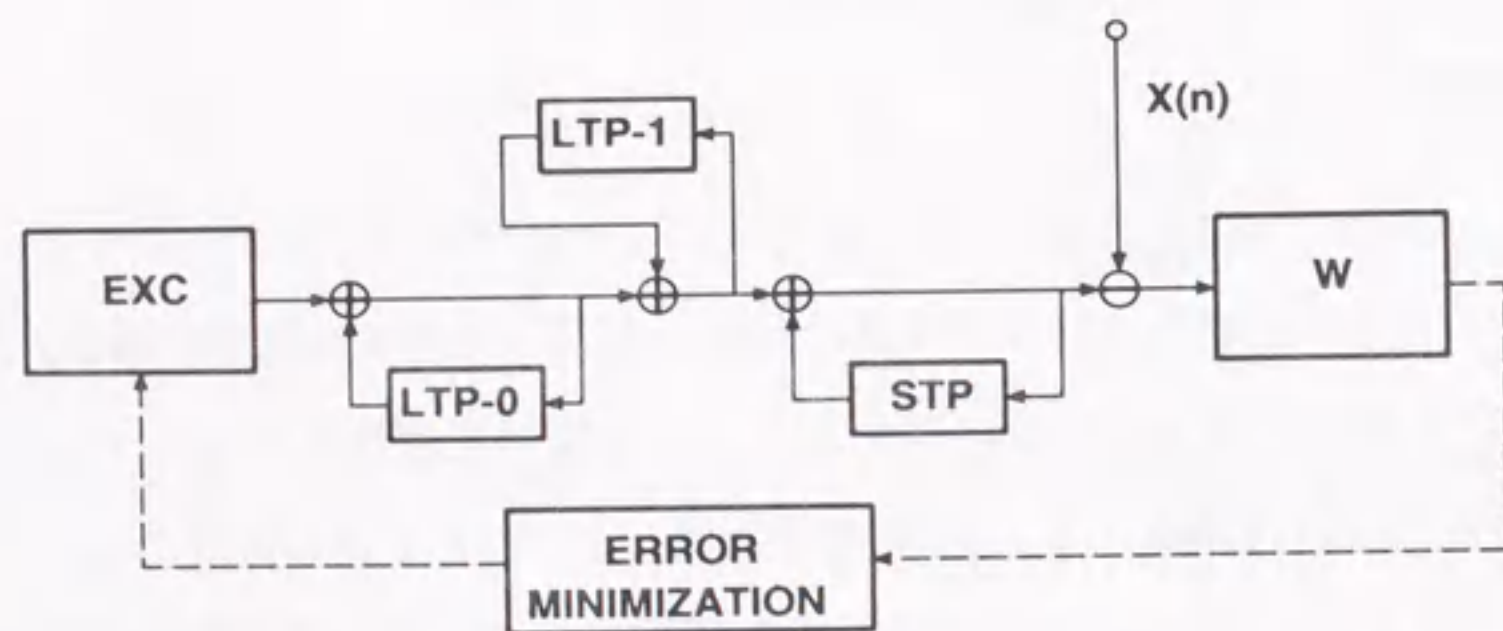


図 3.6: STP を共通化し LTP のみ 2 種類用いる P-2 法のブロック図

LTP-1 のピッチ周期、係数は、過去の音源信号による再生信号と入力音声  $x(n)$  との閉ループ探索法から求めた。また、LTP-0 のピッチ周期、係数は、LTP-1 によるピッチ予測値を減算した予測残差信号  $y(n)$  から閉ループ法により求めた。なお、これら 2 種の LTP のピッチ周期は相関が高いので、サイド情報の伝送量を低減化するために、LTP-0 の周期は LTP-1 の周期との差分で伝送した。

### 3.2.3 サブフレーム間ピッチ予測法 (P-3 法)

#### 原理

P-1 法、P-2 法では、LTP をパルス探索の誤差最小化ループの内側に配置していた。従って、合成フィルタは LTP と STP との縦続接続となり、縦続接続合成フィルタのインパルス応答は、LTP のインパルス応答を畳み込んだものとなるので、インパルス応答長を長くする必要があった。また、フレーム全体にわたりフレーム内ピッチ予測の効果を上げようとする、原理的にはフレーム長と等しい長さのインパルス応答長が必要であった。さらに、パルスの探索法として、方式 2 を使用していたため演算量が多かった。

P-3 法では、演算量を削減するため、パルスを探索する処理単位をピッチ周期に等しい長さで分割したサブフレームとする。これにより、サブフレームには常に一つのピッチ波形のみが入ることになる。サブフレーム間ピッチ予測によりサブフレームのピッチ波形はあらかじめ除去されるため、サブフレーム内ピッチ予測は不要となる。従って、合成フィルタとしては STP のみの構成でよく、インパルス応答長も短くて済む [114],[116]。

P-3 法の原理を示すブロック図を図 3.7 に示す。まず、入力音声信号  $x(n)$  から閉ループ法によりピッチ周期  $T$  を抽出し、フレームをピッチ周期  $T$  の長さに等しい長さのサブフレームに分割する。

次に、過去のフレームでの再生音声  $x_d(n)$  (STP の出力) をもとに、LTP においてピッチ周期  $T$  を用いて式 (3.7) に従いフレーム全体でピッチ予測を行ない、残差信号  $y(n)$  を一旦求める。

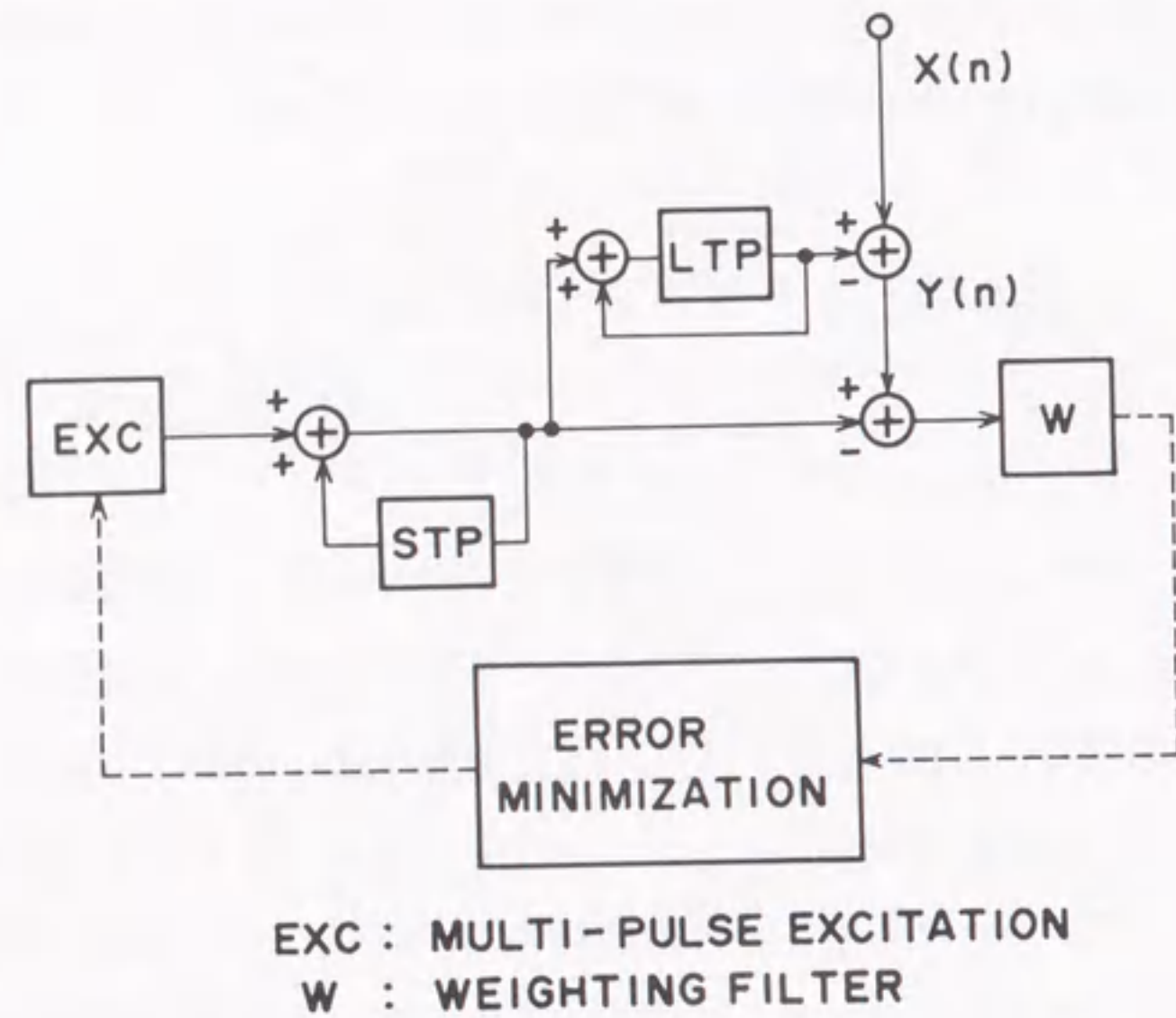


図 3.7: ピッチ予測マルチパルス符号化 (P-3 法) のブロック図

$$y(n) = x(n) - b x_d(n-T) \quad (3.7)$$

ここで、 $b$  はピッチ係数であり、式 (3.8) を最小化するように閉ループ法により求めた。

$$\begin{aligned} E &= \sum_{n=0}^{N-1} y^2(n) \\ &= \sum_{n=0}^{N-1} [x(n) - b x_d(n-T)]^2 \end{aligned}$$

$$= \sum_{n=0}^{N-1} x^2(n) - \frac{\left[ \sum_{n=0}^{N-1} x(n) x_d(n-T) \right]^2}{\sum_{n=0}^{N-1} x_d^2(n-T)} \quad (3.8)$$

式 (3.8) を  $b$  で偏微分して 0 とおくことにより、ピッチ係数  $b$  は下式から計算される。

$$b = \frac{\sum_{n=0}^{N-1} x(n) x_d(n-T)}{\sum_{n=0}^{N-1} x_d^2(n-T)} \quad (3.9)$$

ここで、 $N$  はフレーム長に等しい長さとした。

次に、ピッチ周期  $T$ 、ピッチ係数  $b$  を用いてサブフレーム単位でピッチ予測を実施しピッチ予測残差  $y(n)$  を求め、 $y(n)$  に対しサブフレーム単位でマルチパルスの探索を行う。マルチパルスの探索法には、合成フィルタが STP のみでインパルス応答が短時間で減衰すること、サブフレームに割り当てるパルスの個数が少ないことから、第 2 章 2.3.1 節で述べた演算量の少ない方式 1 を用いた。このため前述の P-1 法、P-2 法に比べ、演算量を大幅に低減できる。ただし、パルスの探索区間はサブフレーム区間となるので、ピッチ周期が短くサブフレーム長が短いときには、フレーム全体での探索と比べ性能が若干低下することが予想される。

本方式において伝送する情報は、フレーム毎に、ピッチ係数、ピッチ周期、STP の係数、サブフレーム毎に、マルチパルスの振幅と位置である。

### 3.3 評価実験

#### 3.3.1 実験条件

これまでに説明した 3 種類のピッチ予測マルチパルス符号化方式 (P-1 法、P-2 法、P-3 法) を比較評価した。ビットレートは 9.6 kb/s 及び 8 kb/s とした。

符号化条件を表 3.1 に示す。STP の係数は第 2 章 2.4.4 節と同様に、PARCOR 係数に変換した後にスカラ量子化した。LTP の次数はいずれの方式も 1 とし、ピッチ

表 3.1: 9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化 (P-1,P-2,P-3) の符号化条件

項目	P-1	P-2	P-3
フレーム長 (ms)	20	20	20
STP			
次数	12	12	12
ビット数	44	44	44
LTP			
次数	1	LTP-0: 1, LTP-1: 1	1
周期 (ビット)	7	LTP-0: 6, LTP-1: 7	7
係数 (ビット)	3	LTP-0: 3, LTP-1: 3	3
パルス数	11 (15)	10 (14)	11 (15)
パルス最大値 (ビット)	6	6	6
パルス振幅 (ビット)	4	4	4
パルス位置 (ビット)	5	5	5

( )内は 9.6 kb/s の条件を示す。

周期は 7b で整数値表現した。但し、P-2 法のみは、2 種のピッチフィルタ LTP-0、LTP-1 のピッチ周期のうち、LTP-1 は 7b とし、LTP-0 は LTP-1 のピッチ周期と差分符号化し 6b で表した。LTP-0、LTP-1 のピッチ係数はそれぞれ 3b でスカラ量子化した。マルチパルスの振幅、位置は、第 2 章の 2.4.4 節で述べた手法によりスカラ量子化した。第 2 章で述べたピッチ予測を用いないマルチパルス符号化方式 (MPC) (パルス探索法は方式 1 を使用) [108] も基準方式として評価に含めた。MPC では、パルスの個数を 9.6 kb/s で 16、8 kb/s で 12 とし、LPC の分析次数を両ビットレートともに 12 とした。

### 3.3.2 客観評価

客観評価尺度には第 2 章 2.4.1 節で述べたセグメンタル S/N ( $SNR_{seg}$ ) [179] を用いた。評価には 8 種の日本語短文 (合計約 30 sec) を用いた。話者は男女合計 6 名 (男性 3 名、女性 3 名) である。これらはダイナミックマイクにより収録し、3.4 kHz で低域濾波し、標本化周波数 8 kHz、精度 12 ビットの A/D 変換器を用いてデジタル化した。

9.6 kb/s, 8 kb/s における平均  $SNR_{seg}$  を図 3.8 に示す。

ピッチ予測の導入により、ピッチ予測のない MPC に比べ、 $SNR_{seg}$  が 1.3 dB から 2.0 dB 改善されている。本章で提案した 3 種類のピッチ予測法の比較では、8 kb/s では、P-2 法、P-1 法、P-3 法の順に性能がよく、 $SNR_{seg}$  はそれぞれ、14.2 dB、13.8 dB、13.6 dB であった。話者の性別にみると、男性では P-2 法、P-3 法、P-1 法の順であり、 $SNR_{seg}$  はそれぞれ、13.3 dB、12.6 dB、12.2 dB である。一方、女性話者では P-1 法が最もよく、続いて P-2 法、P-3 法であり、それぞれ、15.7 dB、15.2 dB、14.8 dB であった。

9.6 kb/s では、P-1 法、P-2 法は同等の性能で 15.7 dB、P-3 法は 15.4 dB であった。男性話者に対しては、P-2 法、P-1 法、P-3 法の順で、それぞれ、14.9 dB、14.3 dB、14.2 dB の  $SNR_{seg}$  であった。女性話者に対しては、P-1 法が最も良く、P-2 法、P-3 法は同等であり、それぞれ、17.1 dB、16.6 dB、16.6 dB の  $SNR_{seg}$  を得た。

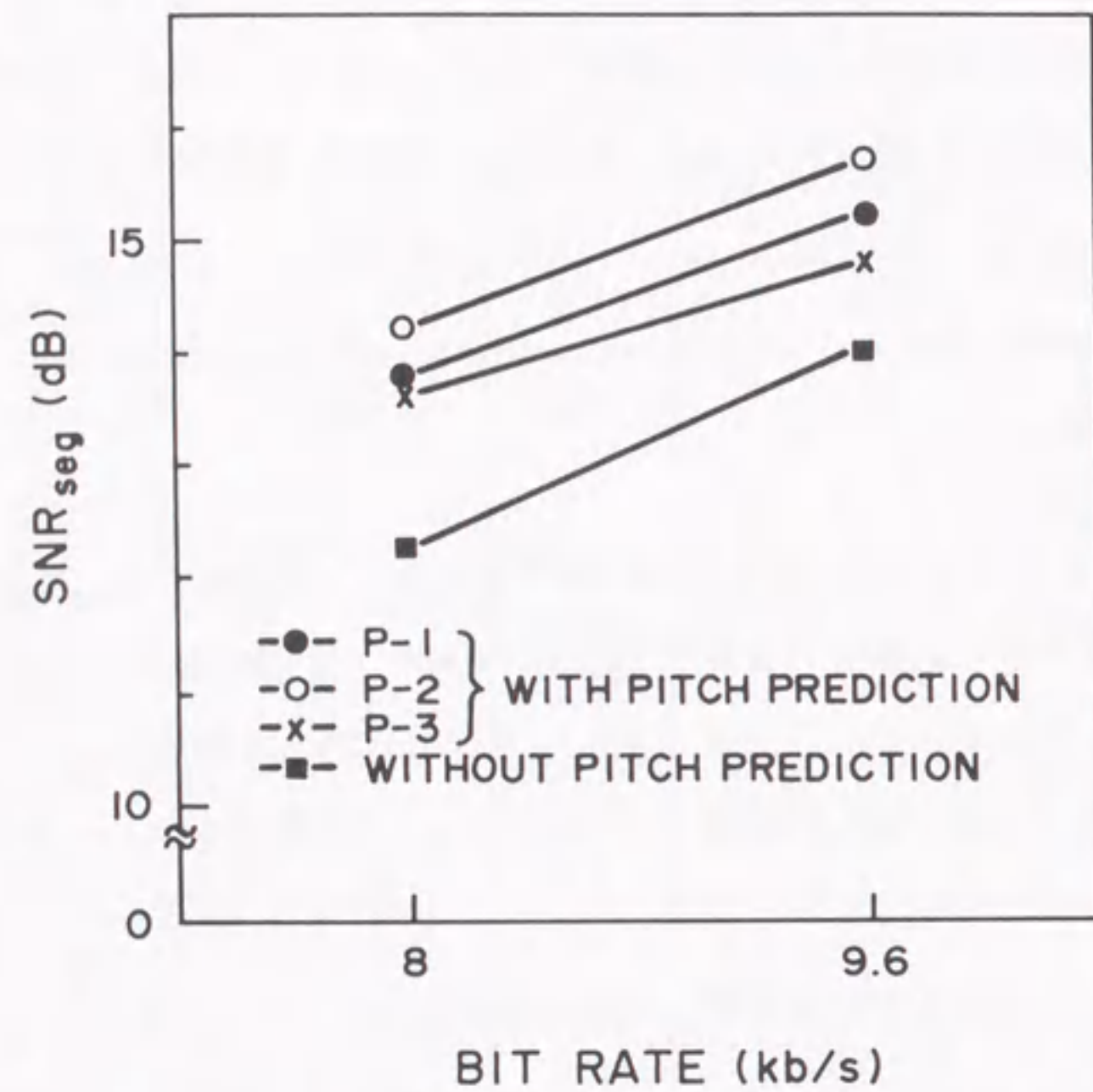


図 3.8: 9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化の  $SNR_{seg}$  特性

表 3.2: 一対評価 (i, j) における 5 段階評定尺度

評価語	評点
i は j よりかなり良い	2
i は j よりやや良い	1
i と j は差がない	0
j は i よりやや良い	-1
j は i よりかなり良い	-2

### 3.3.3 主観評価

一対比較法を用いて主観評価を行った。評価した方式は、3種のピッチ予測マルチパルス法と、ピッチ予測のない MPC とし、基準方式として 56 kb/s、48 kb/s、40 kb/s の  $\mu$ -law PCM (以下それぞれ、PCM 7、PCM 6、PCM 5 と記す) も加えた。一対比較には Scheffe の方法 [149], [34] を用い、表 3.2 に示す 5 段階の評定尺度を使用した。なお、この尺度は、ITU-T など CCR (Comparative Category Rating) 尺度としても知られている。音声資料には、客観評価に用いた資料のうちの 4 種の短文 (男性 2 名、女性 2 名) を使用した。被験者は音声を専門としない女性 6 名である。実験は防音室内にてヘッドホンから両耳に音声を提示した。比較対内、対間の時間間隔はそれぞれ 1 sec、3 sec とした。

結果を図 3.9 に示す。横軸は、最も良く評価された方式を 100、最も悪く評価された方式を 0 と記してある。

$SNR_{seg}$  の結果とはやや異なり、8 kb/s では P-1 法が最も良好で、続いて P-3 法、P-2 法の順であった。P-1 法は  $\mu$ -law PCM 6.3 ビット相当の音質であり、MPC に対する音質改善度は、 $\mu$ -law PCM 0.3 ビット相当であった。また、P-3 法、P-2 法は、 $\mu$ -law PCM 6 ビットと同等の音質であった。一方、9.6 kb/s では、P-1 法、P-3 法がほぼ同じ音質であり、P-2 法はこれらに比べ評価は劣った。P-1 法、P-3 法の音質は  $\mu$ -law PCM 6.6 ビット相当であり、MPC に対する音質改善度は、 $\mu$ -law PCM

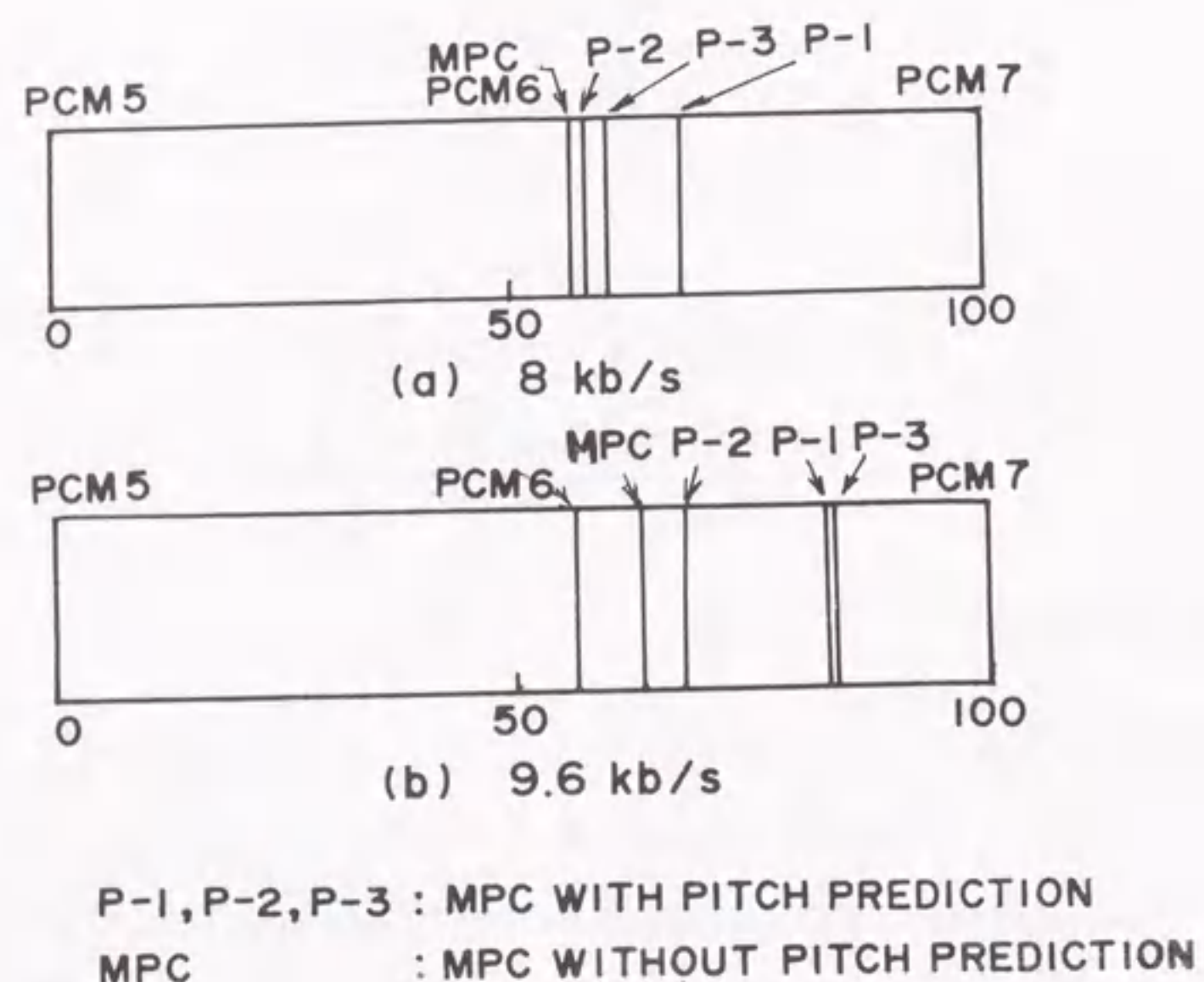


図 3.9: 9.6 kb/s, 8 kb/s ピッチ予測マルチパルス符号化の音質評価結果

0.4 ビット相当であった。

### 3.3.4 考察

以上の実験の結果、次の点が明らかとなった。

(1)  $SNR_{seg}$  評価では、男性話者はピッチ周期が長いので、フレーム内・フレーム間併用ピッチ予測 (P-2 法) や、サブフレーム間ピッチ予測 (P-3 法) が有効である。女性話者に対しては、フレーム内ピッチ予測 (P-1 法) が有効である。女声話者の場合に、P-2 法が P-1 法に比べ  $SNR_{seg}$  がやや低下しているのは、ピッチ周期の時間的なゆら

ぎ等のために、20 ms という比較的長いフレーム全体では、P-2 法のフレーム間ピッチ予測が必ずしも良好に動作しないためと考えられる。また、P-3 法の  $SNR_{seg}$  が他に比べ低下しているのは、パルスの探索をサブフレーム単位に実施しており、サブフレーム長が比較的短いこと、パルス探索に効率法 (方式 1) を用いているためと考えられる。

(2) 音質評価結果によれば、8 kb/s では P-1 法が最も良く、ついで P-3 法であった。9.6 kb/s では、P-1 法、P-3 法はほぼ同等であり、P-2 法はそれらよりも悪かった。話者毎に調べると、女性の方が男性より良好であった。男性では P-2 法、P-1 法、P-3 法の順に良いが、女性に対しては P-1 法が最も良く、ついで P-3 法、P-2 法であった。P-2 法が女声話者で悪い理由は、(1) でも述べたように、フレーム間ピッチ予測を 20 ms フレーム全体で行っているために、フレーム内でのピッチ周期のゆらぎや変化などを十分に表せず、これが一種の周期的な雑音となる。振幅が小さいために  $SNR_{seg}$  には反映されにくい、人間には知覚されやすいため音質を損なっていると考えられる。

(3) ピッチ予測の付加による音質の改善度は、P-1 法によれば、ピッチ予測のない MPC に対して、8 kb/s では  $\mu$ -law PCM 0.3 ビット相当であり、9.6 kb/s では  $\mu$ -law PCM 0.4 ビット相当であった。

### 3.4 むすび

有声区間において、音声波形のピッチ毎の周期性を利用したピッチ予測マルチパルス符号化方式を提案し、伝送速度を 8 kb/s まで低減化した。ピッチ予測技術として、フレーム内ピッチ予測 (P-1 法)、フレーム内・フレーム間ピッチ予測併用型 (P-2 法) 及び、サブフレーム間ピッチ予測 (P-3 法) の 3 種類を提案し、9.6 kb/s、8 kb/s において、客観的、主観的に比較評価した。特に、P-1 法は、従来の APC などに使用されているピッチ予測法や、他の機関でほぼ同時期に独立に検討された方式や、CELP に使用されている適応コードブックとは、構造の異なる独自の手法である。主観音質評価結果によれば、9.6 kb/s では P-1 法、P-3 法がほぼ同等の音質であり、 $\mu$ -law PCM 6.6 ビット (52.8 kb/s) 相当の音質を得た。8 kb/s では P-1 法が最も良好で、 $\mu$ -law

PCM 6.3 ビット (50.4 kb/s) 相当の音質を得た。ピッチ予測の導入により、ピッチ予測のない場合に比べ、9.6 kb/s、8 kb/s ではそれぞれ、 $\mu$ -law PCM 0.4 ビット (3.2 kb/s) 相当、 $\mu$ -law PCM 0.3 ビット (2.4 kb/s) 相当の音質改善度が得られた。また  $SNR_{seg}$  評価によれば、ピッチ予測の導入により、 $SNR_{seg}$  が 1.3 dB ~ 2 dB 改善された。

今後は、音質を維持しながら、伝送速度の一層の低減化が課題である。

## 第4章

### マルチモードによる音源切り替えとピッチ同期マルチパルス音源に基づく 4.8 kb/s 音声符号化

#### 4.1 まえがき

前章までの検討で、マルチパルス音声符号化方式は、ピッチ予測を用いることにより、8 kb/s 程度までは良好な音質を提供できることが明らかとなった。しかし、ビットレートを 4.8 kb/s 程度に低減すると、音質が急激に低下していた。主な原因は、ビットレートの低減に伴い、フレーム当りのマルチパルスの個数が減少することが挙げられる。例えば、フィルタ係数や音源情報の量子化に前章までで述べたスカラ量子化を使用する場合、4.8 kb/s ではパルスの個数は 20 ms フレーム当り 4 パルス程度となる。このため、有声区間ではピッチ予測を用いても音源波形のピッチ成分の近似度が不十分となり、特にピッチ周期の短い女性話者において劣化が顕著であった。また、子音区間では、特に摩擦音などでは音源に周期性がみられず雑音的となるため、少ない個数のマルチパルスでは音源波形を良好に表すのは困難であり、パルスの音質になっていた。本章は、ビットレートが低くパルスの個数が少ない場合に、ピッチ予測に比べ音質を改善する方法として、マルチモードによる音源信号の切り替えとピッチ同期マルチパルス音源に基づく音声符号化を提案する。

本方式は次の特徴を有する。(1) フレームの音声を複数種類のモード (マルチモード) に分類し、モードにより適切な音源信号を選択して使用する (マルチモード符号化)。(2) 母音モードでは、音声のピッチ毎の準周期性を最大限に利用したピッチ同

表 4.1: モードと音源信号の分類

モード	選択する音源信号
母音	ピッチ同期マルチパルス (PSMP)
鼻音	マルチパルス (MP)
破裂・過渡	マルチパルス (MP)
子音・無音	乱数コードブック (SC)

期マルチパルス音源を用いる。これは、フレームをピッチ周期に等しい長さのピッチ区間に分割し、代表的な一つのピッチ区間 (代表区間) においてのみマルチパルスを求め、音源信号として伝送する。同一フレーム内の他のピッチ区間では、代表区間で求めたマルチパルスをピッチ同期で適用するための振幅と位相の補正係数を求めて伝送する。代表区間のフレーム内の位置及び、振幅、位相補正係数は、フレーム全体の波形歪みを最小にすように計算する。これらの構成により、音源信号伝送に必要な情報量を削減している。(3) 子音モードでは、雑音性の音源を良好に表現するために、CELP などで使用されている乱数コードブック音源を使用する。過渡モードでは、マルチパルス音源を使用する。

これらの構成により、4.8 kb/s という低ビットレートでも、音声信号の種類によらず、従来方式よりも音質の良好な合成音声を得ることができることを示す。

## 4.2 提案方式の構成

### 4.2.1 モードと音源信号の種類

フレームの音声信号から開ループで種々の音響的特徴を抽出し、これらを使用して複数種類のモードに分類し、各モードに適切な音源信号を切り替える。分類に使用した音響的特徴については4.2.2節で説明する。モードの種類とそれに対応する音源信号の関係を表4.1に示す [123]。ここではモードの個数は4とした。

### 4.2. 提案方式の構成

各モードに用いる音源信号は、周囲雑音などにより、モードの選択を誤まっても、音質の致命的な劣化を避けられるように考慮して、表4.1に示すものを選んだ [123], [124]。

母音モードでは、音源信号の周期性を最大限に利用してパルスの個数を低減することが可能なピッチ同期マルチパルス音源 (PSMP: Pitch Synchronous Multi-Pulse) を用いる。鼻音モードでは、LPC 予測利得が高いため、予測残差に明確なピッチ周期性が見られないことが多いので、PSMP は使用せず、従来のマルチパルス音源 (MP: Multi-Pulse) を用いる。破裂・過渡モードは有声の場合が多いが、ピッチ周期などの音源信号の特性がフレーム内で時間的に急速に変化するので、マルチパルス音源 (MP) を用いる。子音・無音モードでは、音源信号は雑音的となるため、少ない個数のマルチパルスで良好に表すことは困難である。従ってこのような音源信号を良好に表すために、CELP 方式 [152] に使用されている乱数コードブック音源 (SC: Stochastic Codebook) を用いる。

モードの分類及び音源信号の選択には、別の方法として、全ての音源信号を並列に使用し各音源信号に対して合成音声一旦求め、原音声との間で波形歪を評価し、歪みを小さくするような音源信号を選択するという閉ループ選択法も考えられる [164] が、本章では、次の理由により、この方法は採用しない。(a) 閉ループ選択法では、モード毎に信号の合成と歪計算が必要なため、演算量は、開ループ選択に比べはるかに多く、モードの種類に比例して増大する。(b) 音源信号の選択に用いる歪評価尺度が問題となる。単純に誤差電力や S/N などにより評価するのは最良ではなく、聴感と対応づいた評価尺度が望ましい。現状ではこのような評価尺度は明らかになっていない。

### 4.2.2 音響的特徴

モードの分類に用いる音響的特徴には、音声認識などで用いているもの [97] や、[176]などを参考にして、次の4種の特徴を用いた。フレーム内 RMS (RMS)、フレームを分割したサブフレーム毎に求めた RMS のサブフレーム間での変化 (RMS-D)、フレームピッチ予測ゲイン (P)、LPC 分析によるフレーム短期予測ゲイン (G) であ

る。

これらの値を組み合わせて次のように分類を行なった。RMS、P、Gが予め定められたしきい値よりも大きく、かつRMS-Dがしきい値より小さいときは母音モードとした。母音モードにおいてGが鼻音モードを示すしきい値を越えている場合は鼻音モードとした。RMS、P、Gがしきい値よりも小さく、かつRMS-Dがしきい値よりも大きいときは破裂・過渡モードとした。以上のいずれにも属さないときには子音・無音モードとした。

#### 4.2.3 ピッチ同期マルチパルス音源 (PSMP)

##### 原理

母音モードでは、音源信号のピッチ毎の準周期性を最大限に利用して音源の伝送情報量を低減化するために、ピッチ同期マルチパルス音源 (PSMP: Pitch Synchronous Multi-Pulse) を提案する。ピッチ同期マルチパルス音源 (PSMP) を用いる音声合成モデルを図4.1に示す。フレームをピッチ周期に等しい長さのピッチ区間に分割し、一つのピッチ区間 (代表区間) の音源信号のみを伝送する。伝送する区間の音源信号はマルチパルスで表す [118], [124]。母音区間において自然性の知覚を改善するには、ピッチ波形のピッチ毎の位相変化を精度よく求めることが重要であると報告されている [181]。そこで、本方法では、同一フレーム内の代表区間以外のピッチ区間において、音源信号の振幅、位相が時間的に滑らかに変化していると仮定する。代表区間の音源をピッチ同期に繰り返しながら、位相、振幅を最適に補正するために位相補正係数、振幅補正係数を代表区間以外のピッチ区間毎に求め、これらを補助情報として伝送する [122], [123], [124]。これらの情報を用いてフレーム全体の音源信号を復元する。

##### 位相、振幅補正係数の計算法

位相、振幅補正係数の計算は次のように行なう。代表区間以外のピッチ区間では、まず代表区間の音源信号をピッチ同期で繰り返し合成フィルタを駆動して該当するピッチ区間の合成音声を求める。次に合成音声と入力音声信号との聴感重み付け誤差電力

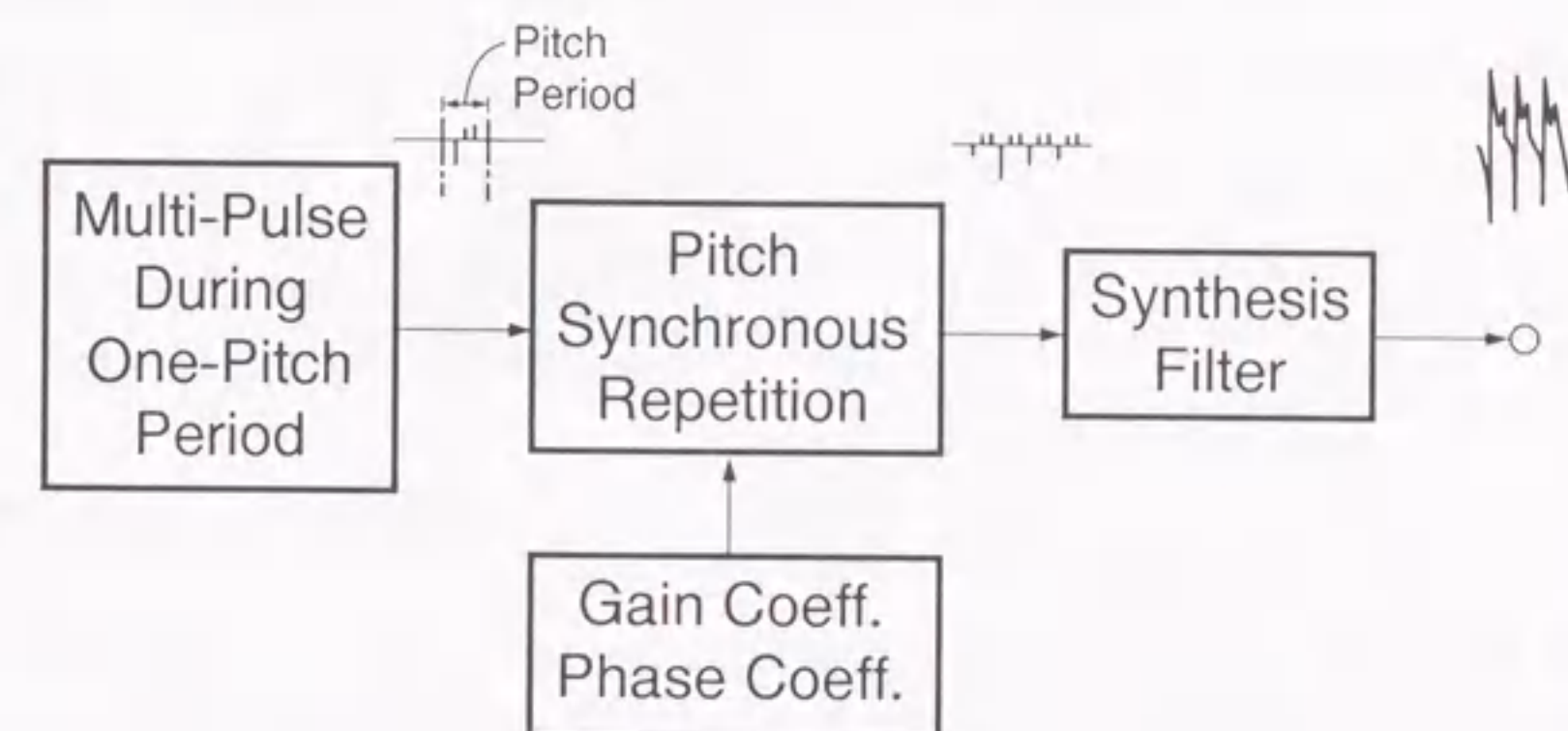


図4.1: ピッチ同期マルチパルス音源 (PSMP) を用いる音声合成モデル

を最小化するように補正係数を決定する [122], [123], [124]。

具体的な計算法を次に示す。まず、代表区間の候補を複数区間選択し、候補区間の各々に対し、 $K$  個のマルチパルスを計算する。ここで、 $i$  番目のマルチパルスの振幅、位置をそれぞれ、 $g_i$ 、 $m_i$  とする。同一フレームで代表区間以外の  $j$  番目のピッチ区間における振幅補正係数、位相補正係数をそれぞれ、 $c_j$ 、 $d_j$  とする。ピッチ区間  $j$  において、 $c_j$ 、 $d_j$  を用いて音源信号を補正し、その区間の合成音声  $c_j s_j(n - d_j)$  を式(4.1)に従い求める。



$$c_j s_j(n - d_j) = c_j \sum_{i=1}^K g_i h(n - m_i - L \cdot T - d_j), \quad (L \text{は整数}) \quad (4.1)$$

ここで、 $T$ 、 $h(n)$ はそれぞれ、フレームの平均ピッチ周期、合成フィルタのインパルス応答を示す。ピッチ区間  $j$  における原音声と合成音声との聴感重み付け誤差電力は式(4.2)で表される。

$$E_j = \sum_n \left[ x_{wj}(n) - c_j \cdot s_{wj}(n - d_j) \right]^2 \quad (4.2)$$

$$x_{wj}(n) = x_j(n) * w(n) \quad (4.3)$$

$$s_{wj}(n) = s_j(n) * w(n) \quad (4.4)$$

ここで、 $w(n)$ は聴感重み付けフィルタのインパルス応答を示す。記号\*は畳み込み演算を示す。

補正係数  $c_j$ 、 $d_j$ は、式(4.2)を最小化するように求める。 $c_j$ と $d_j$ を同時に求めるのは演算量が多いため、ここでは、準最適法を用いる。式(4.1)と式(4.4)を式(4.2)に代入して $c_j$ で偏微分して0とおくことにより次式を得る。

$$E_j = \sum_n x_{wj}^2(n) - \frac{\left[ \sum_n x_{wj}(n) s_{wj}(n - d_j) \right]^2}{\sum_n s_{wj}^2(n - d_j)} \quad (4.5)$$

$$c_j = \frac{\sum_n x_{wj}(n) s_{wj}(n - d_j)}{\sum_n s_{wj}^2(n - d_j)} \quad (4.6)$$

式(4.5)の右辺第1項はフレーム内で一定であるから、式(4.5)を最小化するには右辺第2項を最大化すればよい。

従って、式(4.5)の右辺第2項を最大化する $d_j$ を求め、式(4.6)から $c_j$ を計算することにより、位相補正係数 $d_j$ 、振幅補正係数 $c_j$ が求められる。以上の処理を代表区間以外のピッチ区間に対して繰り返す。

#### 処理の波形例

以上説明した処理の波形例を図4.2に示す。(a)はフレーム(20 ms)の音声波形(ここでは女性話者の波形)を示す。(b)はフレーム平均ピッチ周期 $T$ を用いてフレームをピッチ周期に等しい長さのピッチ区間に分割し、代表区間の候補区間で求めたマルチパルス(この場合は4パルス)を示す。(c)は代表区間以外のピッチ区間で代表区間の音源をピッチ同期で繰り返した後に、式(4.5)、式(4.6)に従い位相補正係数、振幅補正係数を求め、これらを用いて音源に補正を加え、フレーム全体の音源信号を復元した波形、(d)は(c)に示す音源信号により合成フィルタを駆動して求めたフレーム全体の合成音声波形である。

#### 代表区間の選択

母音遷移部などのように、フレーム内で音声信号の特性が変化している場合でも良好な合成音質を保持するために、フレーム毎に、代表区間の候補をあらかじめ複数区間選択し、これらの各々に対し、上述の方法で補正係数を計算してフレーム全体の合成音声を求め、フレーム全体の波形歪みを求め、これを最小にする候補を代表区間として選択する。

具体的には、代表区間の候補 $l$ に対して、他のピッチ区間において振幅、位相補正係数を求め、フレーム全体の音源信号を復元して合成音声 $\tilde{x}_l(n)$ を一旦求める。次式に従い、フレーム全体で聴感重み付け誤差電力を計算する。

$$E_l = \sum_{n=0}^{N-1} \left[ x(n) - \tilde{x}_l(n) * w(n) \right]^2 \quad (4.7)$$

代表区間の複数候補の各々に対して式(4.7)を計算し、フレーム全体の波形歪みを最小にする区間を代表区間として選択する。

#### 位相、振幅補正係数の量子化法

位相補正係数の量子化は次のように行なう。代表区間を除くピッチ区間毎に、フレームの平均ピッチ周期と補正係数値の差分を量子化する。ここで、少ないビット数

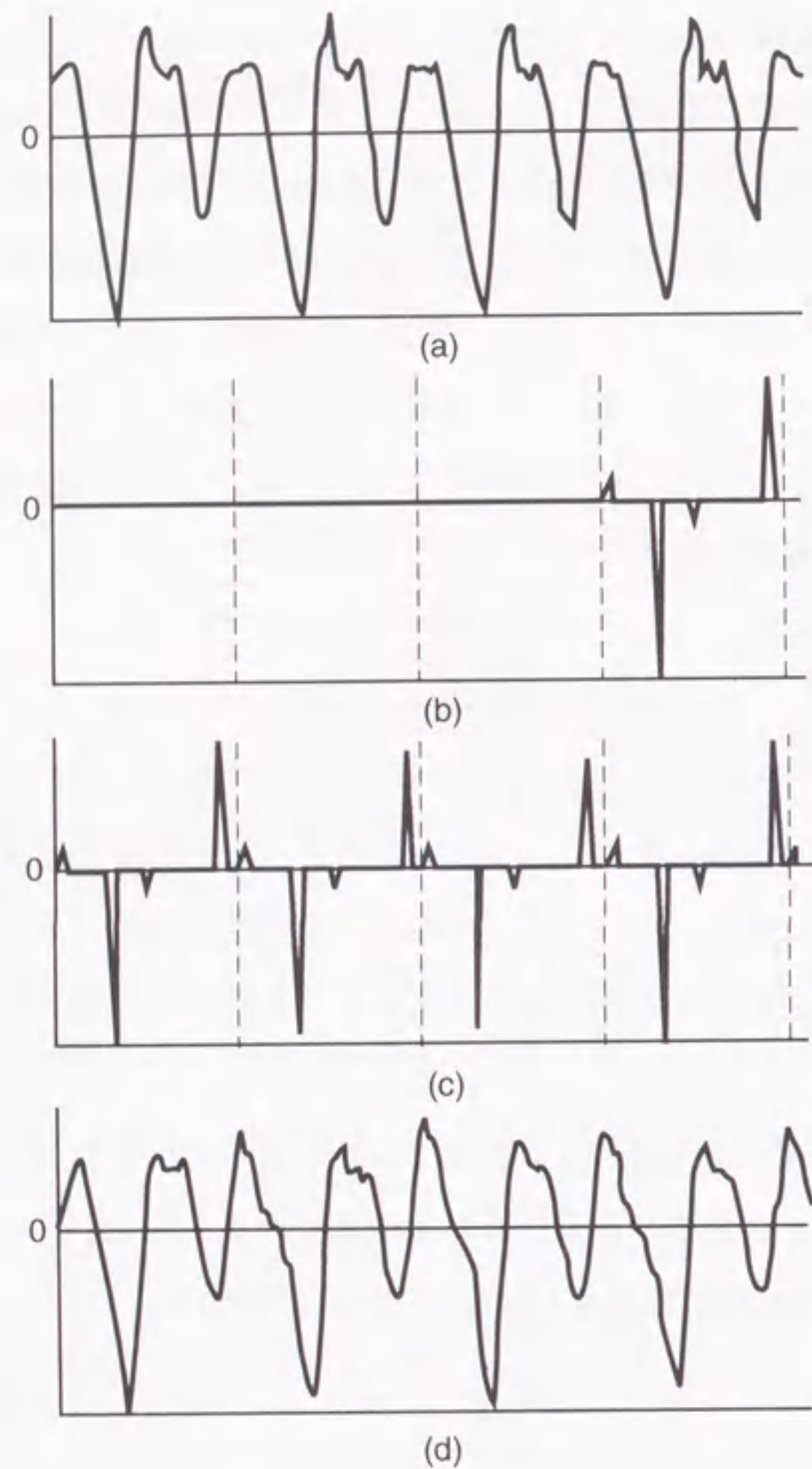


図 4.2: ピッチ同期マルチパルス音源 (PSMP) の処理例, (a) フレームの音声波形, (b) 代表区間の候補区間で求めたマルチパルス (4 パルス), (c) 代表区間の候補区間のマルチパルスと位相, 振幅補正係数を用いてフレーム全体で復元した音源信号, (d)(c) の音源信号により合成した波形

で効率的に量子化するために、量子化範囲の異なる量子化パターンを複数種類用意し、最も良好なパターンをフレーム毎に選択する [126]。具体的には、全ての量子化パターンにより補正係数を一旦量子化し、量子化誤差電力を最小にする量子化パターンを選択する。

振幅補正係数は、代表区間を除くピッチ区間毎に、補正係数値を直接スカラ量子化する [126]。

#### 4.2.4 マルチパルス音源 (MP)、乱数コードブック音源 (SC)

マルチパルス音源を用いるモードでは、フレーム全体に対して予め定められた個数のマルチパルスの振幅、位置を計算する [116]。

乱数コードブック音源を用いるモードでは、CELP 方式 [152] と同様に、ガウス性の乱数コードブックを音源として採用し、波形歪みを最小にするコードベクトルを選択する。

#### 4.2.5 送信側、受信側の構成

本方式の送信側、受信側のブロック図を図 4.3 に示す。送信側では、フレームの音声信号から LPC 分析、ピッチ抽出を行なう。モード判別部では、4.2.1 節、4.2.2 節で述べた方法によりモードを判別する。音源切り替え部では、モード情報により音源信号を選択する。各音源部では、選択された音源信号のパラメータを求める。ここで音源に関する伝送情報は、ピッチ同期マルチパルス音源 (PSMP) のときは、ピッチ周期、代表区間のマルチパルスの振幅と位置、代表区間のフレーム内での位置、同一フレームの他のピッチ区間で求めた振幅補正係数、位相補正係数である。マルチパルス音源 (MP) のときは、マルチパルスの振幅と位置である。乱数コードブック音源 (SC) のときは、選択されたコードベクトルを示すインデックス、音源ゲインである。また、これら以外に、合成フィルタの係数、モード情報を伝送する。

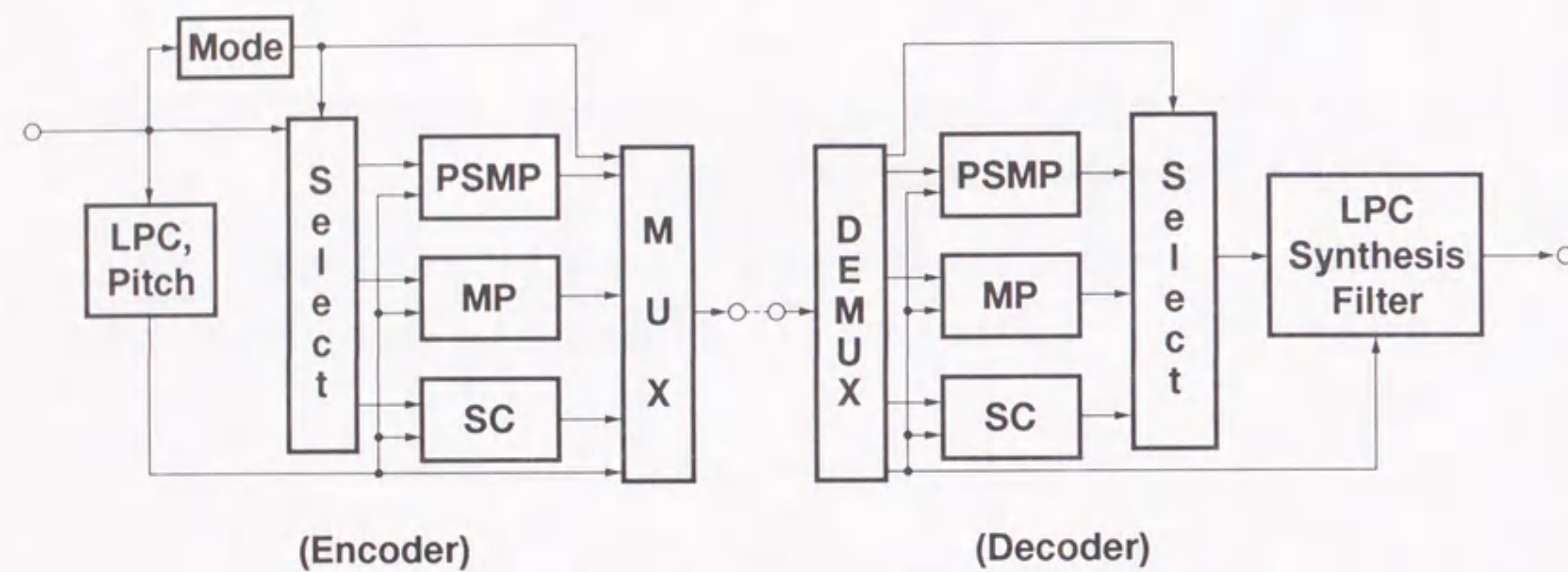


図 4.3: 提案方式の送信側、受信側のブロック図

表 4.2: 4.8 kb/s ビット配分 (ビット / フレーム)

パラメータ	ビット数		
	PSMP	MP	SC
LPC (10次)	36	36	36
モード	2	2	2
ピッチ周期	7	-	-
パルス	4パルス: 37	6パルス: 54	-
代表区間位置	2	-	-
補正係数	12	-	-
乱数コードブック	-	-	10 x 4
音源ゲイン	-	-	4 x 4
合計	96	92	94

## 4.3 実験

## 4.3.1 シミュレーション条件

4.8 kb/s のビット配分を表 4.2 に示す。フレーム長は 20 ms とした。フィルタ係数の量子化法として、今回は第 2 章 2.4.4 節と同様に PARCOR 係数をスカラ量子化した。LSP 係数に変換した上でベクトル量子化 [72] することにより、さらにビットレートを削減できると予想される。ピッチ同期マルチパルス音源 (PSMP) において、振幅補正係数は、代表区間以外のピッチ区間毎にスカラ量子化した。ピッチ周期が長いときは 3 ビット、短いときは 2 ビットを割り当てた。位相補正係数の量子化には、フレーム全体で 2 ビットの量子化パターンを用いた。乱数コードブックの探索は、5 ms サブフレーム毎に行なった。

シミュレーションに用いた音声は、日本語短文 8 種 (男性、女性話者各 3 名発声) である。これらは、標本化周波数 8 kHz、遮断周波数 3.4 kHz で低域濾波され、12

表 4.3: 客観評価結果

方式	$SNR_{seg}$ (dB)	ケプストラム距離 (dB)
4.8 kb/s 提案方式	9.2	1.9
4.8 kb/s 2モードPI法	9.0	2.1
4.8 kb/s ピッチ予測 (P-1法)	8.5	2.0

提案法：4モードとし母音モードでは振幅、位相補正係数の伝送あり

PI法：2モードとし母音モードでは振幅、位相補正係数の伝送なし

ビットのA/D変換器によりデジタル化された。

#### 4.3.2 客観評価

本方式の性能を調べた。第2章2.4.1節に示したセグメンタルSNR ( $SNR_{seg}$ )とケプストラム距離 (CD) を使用した。 $SNR_{seg}$ の平均値を表4.3に示す。比較のために、2種の方式を追加した。一つは、第3章3.2.1節で検討したピッチ予測マルチパルス符号化 (P-1法) を4.8 kb/sに低減化したもの、もう一つは、提案方式において、モードを母音と子音の2モードとし、ピッチ同期マルチパルス音源と乱数コードブック音源を切り替え、ピッチ同期マルチパルス音源において振幅、位相補正係数を伝送せずに代表区間の音源同士を単純に直線補間する方法 (PI) [118], [119], [120], [128] である。

表から、提案方式では、2モードのPI法よりも0.2 dB、ピッチ予測マルチパルス (P-1法) よりも0.7 dB、 $SNR_{seg}$ が向上していることがわかる。提案法の $SNR_{seg}$ がPIよりも向上しているのは、主に、母音モードで、振幅、位相補正係数を伝送している点が寄与していると考えられる。

次に、ケプストラム距離によりスペクトル歪を調べた。ケプストラムの次数は10次とした。結果を表4.3に示す。表から、3方式間の差は少なかった。これは3種の方式ともに合成フィルタの次数を10次としているためと考えられる。

表 4.4: 音質一対評価 (i,j) に用いた5段階評定尺度

iはjより非常に良い	2
iはjよりやや良い	1
iとjは差がない	0
jはiよりやや良い	-1
jはiより非常に良い	-2

提案方式及び、2モードのPI法におけるS/Nの時間的変化の例を図4.4に示す。図4.4 (a)、(b)は、それぞれ男声話者、女声話者の例を示す。各図の上段は、入力音声から20 ms毎に求めたパワの時間変化、下段はS/Nの時間変化を示す。図において、提案方式のS/Nは実線、PI方式のS/Nは破線で示してある。(a)の文章全体の $SNR_{seg}$ は、提案方式が8.4 dB、PI方式が7.4 dBである。(b)では、提案方式が10.8 dB、PI方式が10.5 dBである。提案方式によれば、音声パワの大きな箇所でもS/Nが改善されている。これは、主に、母音モードでの振幅、位相補正係数の伝送がS/Nの改善に貢献しているためと考えられる。

#### 4.3.3 主観音質評価

一対比較法を用いて4.8 kb/s提案方式の音質評価を行った。これ以外に、4.8 kb/sの2モードPI法、4.8 kb/sピッチ予測マルチパルス (P-1法) 及び、基準方式として48 kb/s (PCM 6)、40 kb/s (PCM 5)  $\mu$ -law PCM方式を評価に含めた。一対比較には第3章3.3.3節で使用したScheffeの方法 [149] を用い、表4.4に示す5段階の評定尺度を使った。

音声資料には、 $SNR_{seg}$ の測定に用いた資料のうちの4種の短文 (男性2名、女性2名) を用いた。被験者は、音声を専門としない女性6名である。評価は無響室内にてヘッドホンから両耳に音声を提示した。比較対内、対間の時間間隔はそれぞれ1 sec、3 secとした。

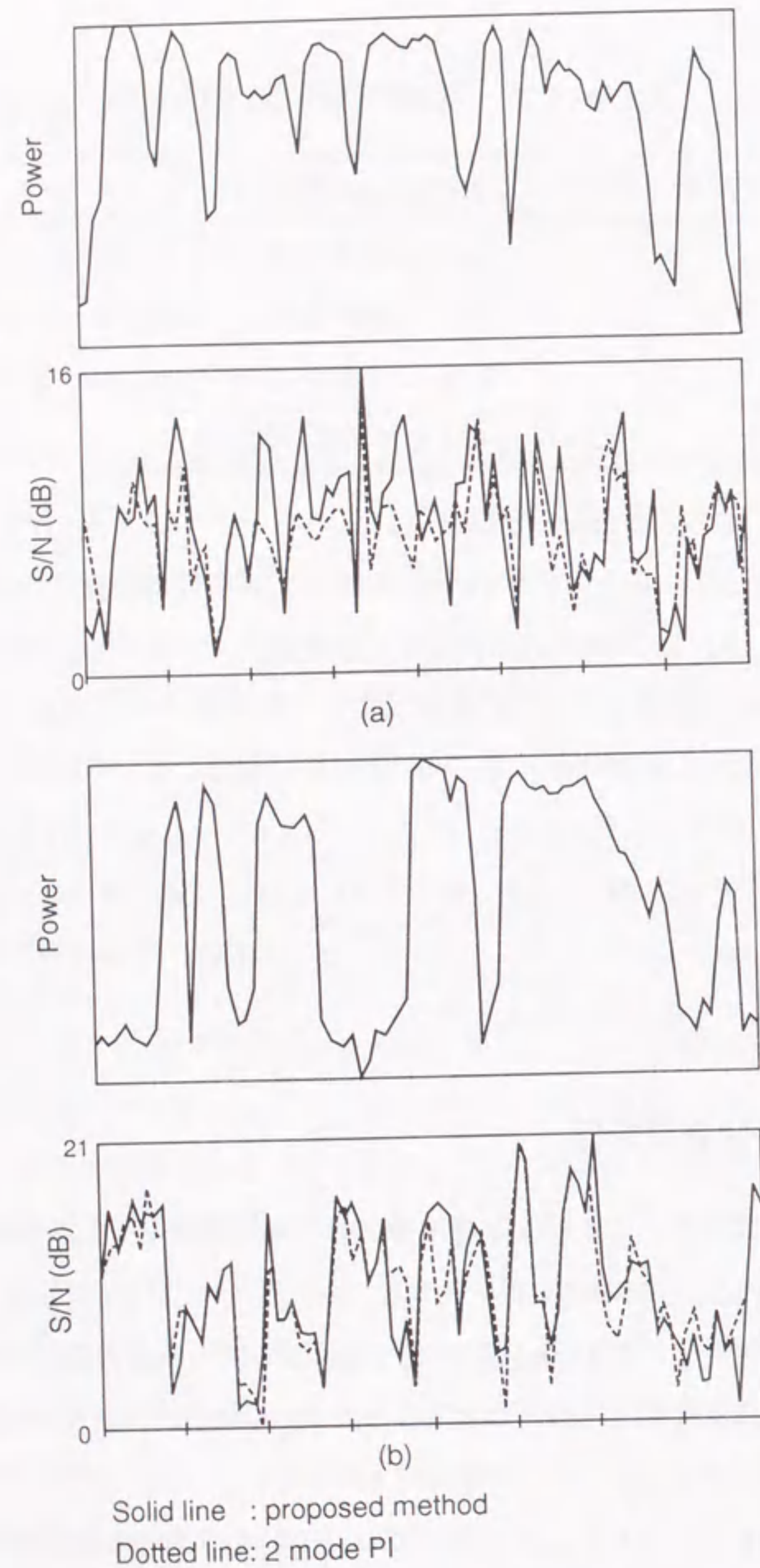


図 4.4: 提案法と 2 モード PI 法の S/N の時間変化の例, (a) は男性話者, (b) は女性話者。各図の上段はパワ, 下段は S/N (dB) を示す。各図の横軸は時間を示し、全長は 2 秒である。

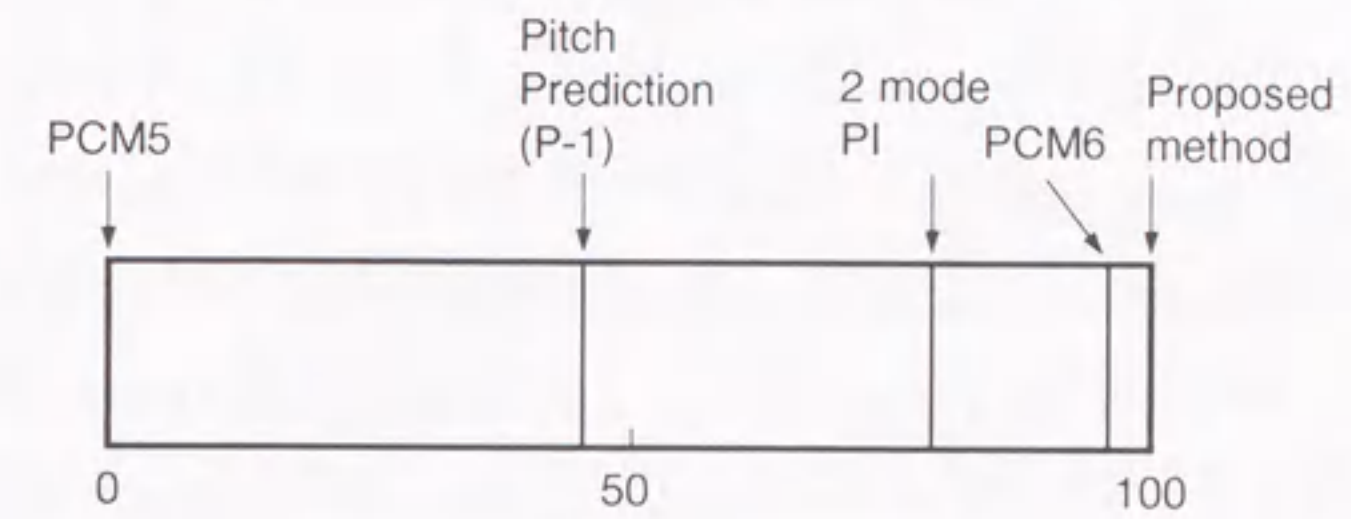


図 4.5: 音質一対評価結果

評価結果を図 4.5 に示す。便宜的に、最も良く評価された方式を 100、最も悪く評価された方式を 0 と記してある。図から、提案方式は 3 種の符号化方式のうちで音質が最も良好で、48 kb/s  $\mu$ -law PCM (PCM 6) 方式と同等の音質が得られた。本方式に続き、2 モード PI 法、ピッチ予測マルチパルス (P-1 法) の順であった。統計検定の結果、提案方式と PCM 6 は有意差がなかった。提案方式と 2 モード PI 法との差異は、危険率 5% で有意、P-1 法との差異は、危険率 1% で有意であった。なお、提案方式と PI 法の音質を比較すると、提案方式は特に、母音区間、過渡区間で明らかに音質が改善されていることを確認した。これは、マルチモード化、母音モードでの

PSMP 音源における位相、振幅補正係数の伝送と、過渡モードでのマルチパルス音源が改善に寄与しているためと考えられる。

#### 4.3.4 考察

実験の結果、次の点が明らかとなった。

- (1)  $SNR_{seg}$  の評価結果から、4.8 kb/s の提案方式は、2モードのPIよりも0.2 dB、ピッチ予測マルチパルス (P-1法) よりも0.7 dB、 $SNR_{seg}$  が向上した。
- (2) 提案方式とピッチ予測マルチパルス法との比較により、4種類のモード毎に音源信号を切り替える点、母音部でピッチ同期マルチパルス音源 (PSMP) を用いる点が有効なことがわかった。特に、ピッチ予測マルチパルス法では、4.8 kb/s とビットレートが低くなると、パルス数が十分でなく、各ピッチ区間の音源を良好に表すことが困難となる。しかしながら、ピッチ同期マルチパルス音源 (PSMP) では、母音モードで、代表区間におけるマルチパルス音源と補正係数によりフレームの音源を表しているため、音源の近似精度を高め、音質を向上させることが可能となった。
- (3) 提案方式と2モードPI方式との比較から、提案方式において、モードの種類を2から4に増やしている点、母音モードのピッチ同期マルチパルス音源において、代表区間以外のピッチ区間で位相補正係数、振幅補正係数を伝送している点が有効であることが示された。後者の効果は、図4.4において、音声パワの大きい箇所では  $SNR_{seg}$  が改善されていることから明らかである。
- (4) 音質評価結果から、提案方式は、2モードPI方式、ピッチ予測マルチパルス (P-1法) よりも音質が良好であり、 $\mu$ -law PCM 6ビット (48 kb/s) 相当の音質を得ている。この音質は、第3章3.3.3節での実験結果から、ピッチ予測を使用しない8 kb/s マルチパルス符号化 (MPC) と同等であるが、8 kb/s ピッチ予測マルチパルス (P-1法) と比較すると、 $\mu$ -law PCM 0.3ビット相当だけ低下している。

#### 4.4 むすび

第3章で検討したピッチ予測マルチパルス符号化は、ビットレートを4.8 kb/s程度まで低下すると、フレームでのパルスの個数が不足するため、母音部、子音部共に音質が劣化してくる。そこで、4.8 kb/s以下のビットレートで有効な音声符号化方式として、マルチモードによる音源信号の切替えとピッチ同期マルチパルス音源 (PSMP: Pitch Synchronous Multi-Pulse) を組み合わせた4.8 kb/s 符号化を提案した。提案法では、フレーム毎に、音声信号から音響的特徴を求め、開ループにて4種のモードに分類し、モード毎に適切な音源信号を選択した。母音モードではピッチ同期マルチパルス音源を開発した。これは、フレームの音源信号を少ない個数のマルチパルスにより極めて効率的に表現するために、代表的な一つのピッチ区間でマルチパルス音源を求め、同一フレームの他のピッチ区間では代表区間のマルチパルスをピッチ同期で繰り返しながら、振幅、位相補正係数により振幅、位相を調整する。この構成に加え、フレーム内で、波形歪みを最小にする代表区間を探索することにより、フレーム内での音源信号の時間変化にも対応できるようにした。一方、過渡モードや鼻音モードではマルチパルス音源、子音モードでは乱数コードブック音源に切り替えた。実験結果によれば、4.8 kb/sの音質は、 $\mu$ -law PCM 6ビット (48 kb/s) 相当と良好な音質を得た。この音質は、ピッチ予測を用いない8 kb/s マルチパルス符号化と同等であった。提案法は、モードの個数を2に限定し母音モードで振幅、位相補正係数を伝送しないPI法、ピッチ予測マルチパルス法 (P-1法) のいずれよりも良好であった。また、平均  $SNR_{seg}$  は9.2 dBであった。

今後の課題は、フィルタ係数を量子化効率がKパラメータに比べさらに高いLSP係数に変換してベクトル量子化しビット数の削減をはかること、振幅補正係数にベクトル量子化を導入することでこれらの情報量を減らしピッチ同期マルチパルス音源のパルス数を増すことである。これらにより、8 kb/s ピッチ予測マルチパルス符号化の音質に近づけることが可能となる。さらに、マルチパルス音源の振幅、位置情報を一層効率的に量子化することも重要である。これには、振幅、位置をコードブックにより同時にベクトル量子化する方法、いずれか片方のみをベクトル量子化し残りをスカ

ラ量子化する方法などが考えられる。ベクトル量子化音源や、マルチパルス音源へのベクトル量子化の適用については、それぞれ、5章、6章で検討したい。

## 第5章

### マルチモードとベクトル量子化音源を用いる CELP 音声符号化

#### 5.1 まえがき

前章までに、マルチパルス音源を用いる線形予測音声符号化方式を検討してきた。これらの方式は、効率的なパルスの探索法を用いることで演算量やメモリ量が少ないことを特徴とする。パルス探索法の改良や、ピッチ予測、マルチモード符号化、ピッチ同期マルチパルス音源などの技術の開発により、16 kb/s ~ 4.8 kb/s のビットレートで比較的良好な音質が得られた。しかし、量子化法としてスカラ量子化をベースにしていたため、量子化効率についてはさらに改善の余地があり、ベクトル量子化などの導入により、特に4.8 kb/s 程度の低いビットレートにおいて一層の高効率化をはかれる可能性があった。

マルチパルス符号化と同じ A-b-S に基づく線形予測音声符号化の枠組みで、音源信号をベクトル量子化 (VQ: Vector Quantization) し、量子化効率の改善を狙った有力な方式として、第1章 1.3.4 節で紹介した CELP (Code Excited Linear Prediction) 符号化 [152] がある。CELP は1980年代末頃から、とりわけ携帯電話などの移動体通信応用向けに研究開発が急速に進展しているが [152], [31], [32], [63], [164]、ビットレートを4.8 kb/s 以下に低減化すると、コードブックのビット数が不十分となり、音質が急速に劣化していた。これを防ぐために、ベクトル量子化の次元数を増やしコードブックのビット数を増加させ量子化性能をあげようとする、コードブックの探索

に必要な演算量や、格納に必要なメモリ量が指数的に増大し、装置化が容易ではなくなっていた。

本章は、ベクトル量子化音源の検討を行なうと共に、CELPにおいて、低ビットレートでの音質改善と、演算量、メモリ量の削減をはかるために、マルチモード符号化とベクトル量子化音源を用いる M-LCELP (Multi-mode Learned Code Excited Linear Prediction) 音声符号化 [135], [136] を提案する。本方式は、モードに応じて、ベクトル量子化音源の種類やパラメータのベクトル量子化コードブックを切り替えることにより、量子化効率の改善を可能とする。4 kb/s あるいはそれ以下のビットレートでは、ピッチに同期させたベクトル量子化励振源を用いる PSI-CELP (Pitch Synchronous Innovation CELP) 方式 [80], [106], [74] などが開発されているが、本方式は、マルチモードによるベクトル量子化コードブックの切り替えを行う点や、音源コードブックの構造などが異なっている。

本符号化の主要な応用分野として、ハーフレート携帯電話などのデジタル移動体通信を考えた場合、音声符号化に要求される条件として、(1) 4 kb/s 程度のビットレートで誰の声に対しても良好な音質を確保すること、(2) さまざまな環境条件や周囲雑音環境下でも音質を維持すること、(3) 無線伝送路で発生するバースト誤りに対して頑健であること、(4) 近年の DSP の処理能力の向上から、既存の 8 kb/s フルレート音声 CODEC の 2 倍以内の演算量におさえること、などがあげられる [76]。

以下では、まず M-LCELP 方式のアルゴリズムを示す。次に音声のビットレートを 4 kb/s とし音質評価を中心とした実験結果について述べ、M-LCELP が上記の要求条件を満足することができることを明らかにする。また、試作装置の演算量についてふれる。

## 5.2 M-LCELP 音声符号化

M-LCELP 音声符号化の送信側、受信側のブロック図をそれぞれ、図 5.1 (a),(b) に示す。

前記 (1)-(4) の要求条件を満たすために、主として、マルチモード符号化、マルチパルスから構成されるスパース構造のベクトル量子化音源コードブックを開発した。

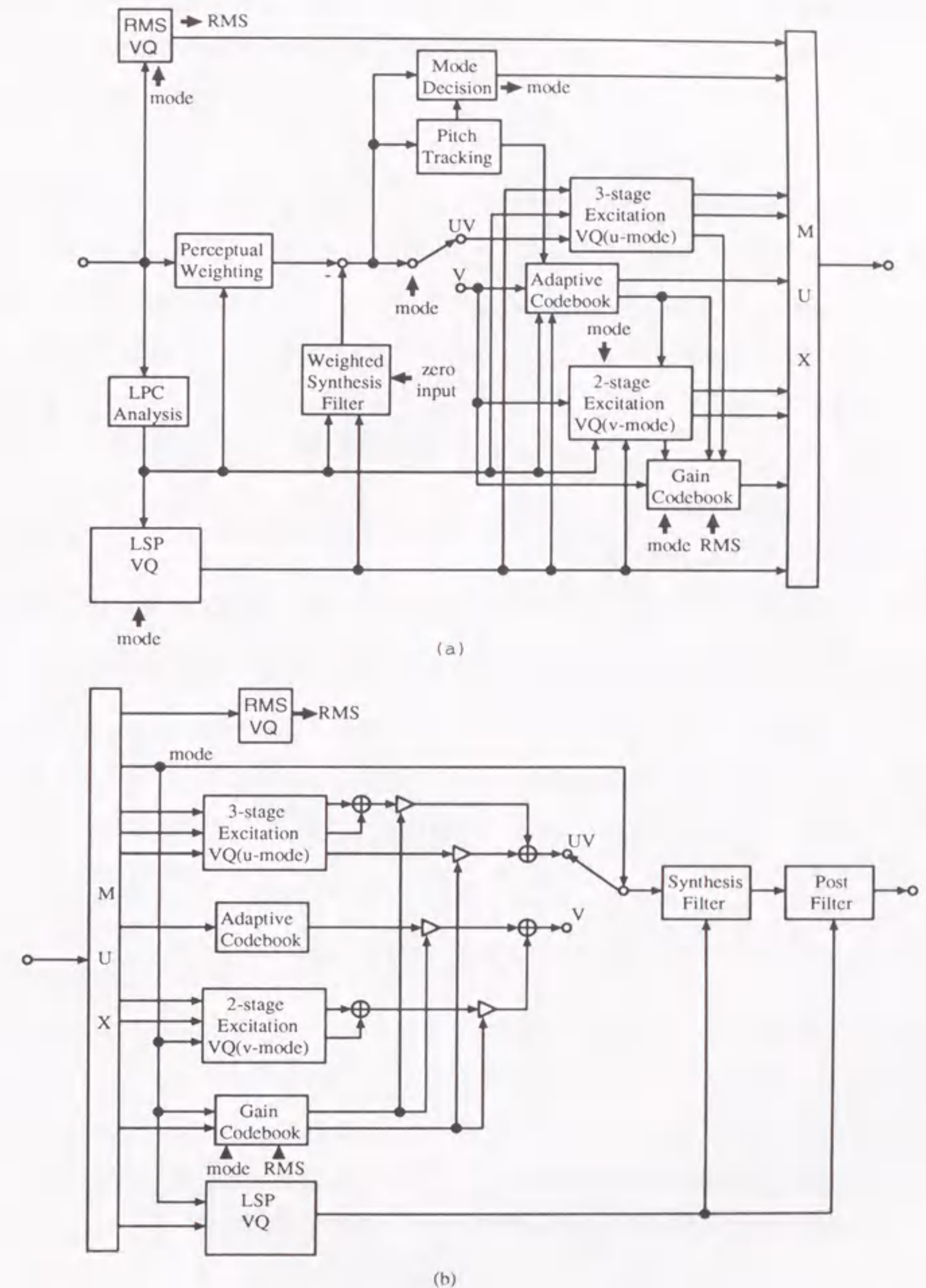


図 5.1: M-LCELP のブロック図, (a) 送信側, (b) 受信側



またこれら以外にも、フィルタ係数の効率的なベクトル量子化、適応コードブックの差分符号化、適応型コムフィルタ技術を開発した。これらを以下で説明する。

### 5.2.1 マルチモード符号化

種々の音声を効率よく良好に符号化するために、第4章で述べたマルチモード符号化を導入する。モード判別部においてフレーム区間の音声を4つのモード(モード0, 1, 2, 3)の一つに分類し、ベクトル量子化音源コードブックならびにパラメータ量子化用のコードブックをモード毎に最適なものに切り替えた。さらに、符号化のアルゴリズムもモードに応じて変え、ピッチ予測を行う適応コードブックをモード1-3のみに使用した。ここで、以下の分類手法によりモード分類を行なうと、モード0は無音/無声部におおむね対応し、モード1, 2, 3は、それぞれ、過渡部、有声の弱定常部、有声の定常部にほぼ対応する。

モードの分類は、第4章4.2節で述べた方法を簡略化し、音響的特徴としてフレームの音声のピッチ周期性の強さ  $G$  のみを用いて行なう。 $G$  は式(5.1)を用いて計算する。

$$G = \frac{1}{M} \sum_{i=1}^M [10 \log_{10} GP(i)] \quad (5.1)$$

ここで、 $M$  はフレーム内のサブフレームの個数、 $GP(i)$  はフレーム内の  $i$  番目のサブフレームで求めたピッチ予測ゲインである。 $GP(i)$  は次式から求める。

$$GP(i) = \frac{\sum_{n=0}^{N-1} x_w^2(n)}{A} \quad (5.2)$$

$$A = \sum_{n=0}^{N-1} x_w^2(n) - \frac{\left[ \sum_{n=0}^{N-1} x_w(n)x_w(n-T) \right]^2}{\sum_{n=0}^{N-1} x_w^2(n)} \quad (5.3)$$

式(5.2)、式(5.3)において、 $x_w(n)$ 、 $T$  はそれぞれ、聴感重み付け入力音声、サブフレーム毎に開ループ探索で求めたピッチ周期を示す。聴感重み付けには式(5.4)の伝

達特性を有する聴感重み付けフィルタを使用した。ここで、2~4章で用いた聴感重み付けとの違いは、重み付けの精度を向上させるために量子化をする前のLPC係数を使用している点である[15]。

$$H(z) = \frac{1 - \sum_{i=1}^P \alpha_i \gamma_1^i z^{-i}}{1 - \sum_{i=1}^P \alpha_i \gamma_2^i z^{-i}} \quad (5.4)$$

上式で、 $\gamma_1$ 、 $\gamma_2$  は、聴感重み付けを制御する定数であり、 $0 < \gamma_2 < \gamma_1 \leq 1$  に選ぶ。 $P$  は聴感重み付けフィルタの次数を示す。

### 5.2.2 フィルタ係数の効率的なベクトル量子化

#### 構造

M-LCELPでは、ビットレートを4 kb/sに下げするために、5.3.1節で示すようにフレーム長を40 msecと長目を選んでいく。フレーム毎にフィルタ係数を一括してベクトル量子化し伝送することにより、フィルタ係数の伝送に必要なビット数を低減できる。しかしその反面、フレーム長が長いことでフィルタ係数を時間軸方向に粗く量子化することになるため、係数の時間変化が大きいフレームでは、粗い量子化のために音質の劣化が予想される。この劣化を防ぐために、補間込み2段分割ベクトル量子化法を開発した[100]。本方法のブロックを図5.2に示す。

本方法では、フレームを前半部と後半部に分け、まずフレーム後半部のフィルタ係数をベクトル量子化する。次にフレーム前半部では補間係数コードブックと誤差コードブックを用いて量子化を行なう。フィルタの次数  $P$  は10とし、フィルタ係数を量子化による劣化の少ない線スペクトル対(LSP: Line Spectrum Pair)係数[45], [160]に一旦変換した上で量子化した。ここで、LSPは、4章までで使用した  $K$  パラメータと比較し次の長所がある。(a)周波数上のパラメータであり、隣接する線スペクトルの接近度によりスペクトル包絡の共振特性を表すため、スペクトル包絡との対応が直接的である。(b)LSPの順序関係は  $0 < \omega_1 < \omega_2 < \dots < \omega_p < \pi$  であり、この場合

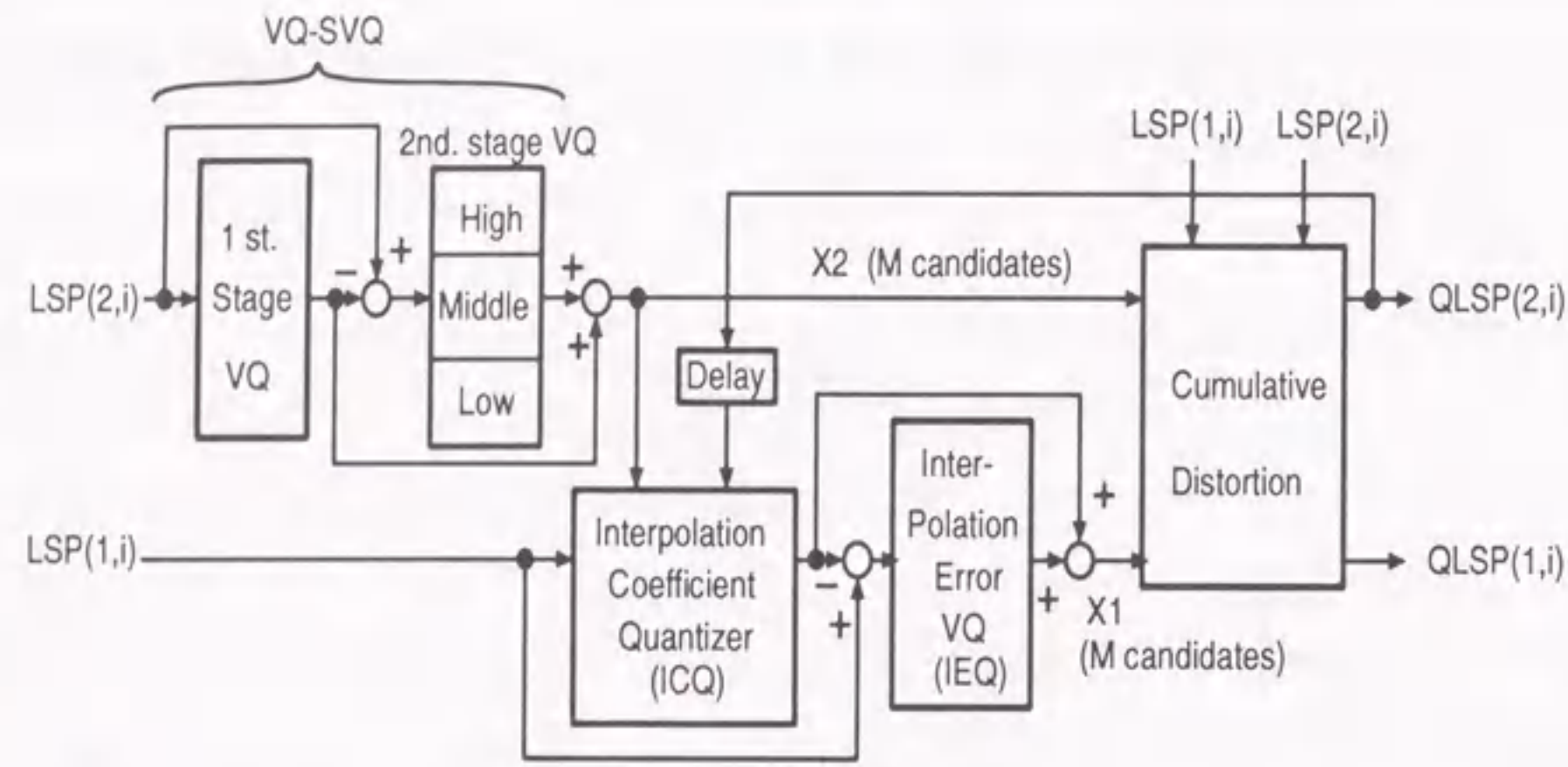


図 5.2: LSP の補間込み 2 段階ベクトル量子化のブロック図

に合成フィルタの安定性が保証される。(c) 量子化感度がほぼ一様化していると共にパラメータの分布範囲が比較的狭い。(d) パラメータの時間的な補間特性が良好である。これらの特徴から K パラメータに比べ量子化ビット数を 20% 以上削減可能と報告されている [160]。特に移動体通信の分野では、伝送路誤りにより LSP の伝送インデクスが誤っても合成フィルタの安定性の判別が容易であることから有効性が高い。

#### (1) フレーム後半部

LSP を量子化するには通常、フレーム当たり 20 ~ 30 ビットのビット数を必要とする。これを 1 段のベクトル量子化器で表すと量子化器のサイズが大きくなりすぎ、5.1 節の条件 (4) を満足できない。そこで、ベクトル量子化器を 2 段に分割し、さらに、2 段目を LSP の次数方向に 3 分割した [100], [91], [143]。1 段目で 10 次の LSP を一括してベクトル量子化し、その誤差を 2 段目で次数方向に 3 分割してベクトル量子化する。LSP は次数間の相関が高いことが知られているが、2 段目は誤差信号を量子化するため、相関はかなり減少していることが予想されるので、2 段目を分割しても損失は少ないと考えられる。各段での LSP コードベクトルの選択は式 (5.5) の歪を最小化するように行う。

$$D_j = \sum_{i=1}^{10} C(i) \omega(i) [LSP(2,i) - QLSP_j(2,i)]^2 \quad (5.5)$$

ここで、 $LSP(2,i)$ 、 $QLSP_j(2,i)$  はそれぞれ、フレーム後半部での量子化前の LSP、LSP コードブックの j 番目のコードベクトルを用いて量子化した LSP 値を示す。 $C(i)$  は聴覚の特性を考慮し次の値を用いた [143]。

$$c(i) = \begin{cases} 1.0, & \text{for } 1 \leq i \leq 8 \\ 0.8, & \text{for } i=9 \\ 0.4, & \text{for } i=10 \end{cases}$$

$\omega(i)$  は、隣接の LSP が接近するほど強い重みをかけ量子化精度を向上させるために導入した重み係数である [145]。

$$\omega(i) = \frac{1}{LSP(i) - LSP(i-1)} + \frac{1}{LSP(i+1) - LSP(i)} \quad (5.6)$$

ベクトル量子化の際に、1 段目と 2 段目の分割による損失を防ぐために、1 段目において、式 (5.5) が小さい順に複数個の候補を出力し、候補の各々に対して 2 段目を探索して、歪が小さい順に組み合わせを M 個選択する。

#### (2) フレーム前半部

図 5.2 の補間係数コードブック (ICQ: Interpolation Coefficient Quantizer) において、現フレームの後半部で量子化した LSP と過去のフレームの後半部で量子化した LSP とを補間係数  $\eta$  を用いて補間し現フレーム前半部の補間値を求める。さらに、前半部の補間値と真の LSP との差分を誤差コードブック (IEQ: Interpolation Error Quantizer) を用いて量子化する。従って、前半部での量子化値  $QLSP(1,i)$ 、前半部での真の LSP は、それぞれ式 (5.7)、式 (5.8) で表される。

$$QLSP(1,i) = \eta QLSP_p(2,i) + (1-\eta) QLSP(2,i) + E(i) \quad (5.7)$$

$$LSP(1,i) = QLSP(1,i) + q(i) \quad (5.8)$$

ここで、 $\eta$ 、 $QLSP_p(2,i)$ 、 $QLSP(2,i)$ 、 $E(i)$ 、 $q(i)$ はそれぞれ、補間係数、過去のフレームの後半部で量子化したLSP、現フレームの後半部で量子化したLSP、誤差コードベクトル、前半部での量子化誤差を示す。

さらに、後半部と前半部とで分割して量子化することによる損失を防ぐために、以下の処理を行う。フレーム後半部で選択されたM個の量子化候補の各々に対して、式(5.7)により前半部のLSPを量子化し、前半部、後半部における真のLSPとの歪の累積を式(5.9)により計算し、これを最小にする組合せを選択する。

$$D_j = \sum_{l=1}^2 \sum_{i=1}^{10} c(i) \omega(l,i) \left[ LSP(l,i) - QLSP_j(l,i) \right]^2 \quad (5.9)$$

ここで、 $\omega(l,i)$ は、式(5.6)と同様にして求める。

#### 性能

本量子化法と従来の手法（ベクトル-スカラ量子化と補間の併用 [100]）とを比較した。評価尺度としては原信号と符号化信号の対数スペクトル上での歪みを測定する対数スペクトル歪み尺度 [160] を用いた。式(5.10)に示す。

$$SD(dB) = \sqrt{\frac{1}{N_f} \sum_{m=1}^{N_f} \left[ \frac{1}{L} \sum_{k=0}^{L-1} \left( \log S_m(k) - \log S'_m(k) \right)^2 \right]} \quad (5.10)$$

ここで、 $S_m(k)$ 、 $S'_m(k)$ 、 $N_f$ はそれぞれ、第mセグメントにおける量子化前のLSPを線形予測係数に変換しFFTして求めたスペクトル包絡、量子化後のLSPを線形予測係数に変換しFFTして求めたスペクトル包絡、評価信号に含まれるセグメントの個数を示す。

LSPの量子化ビット数は40msフレーム当たり32ビット(800b/s)とした。量子化ビット配分を表5.1に示す。ここで、SQはスカラ量子化(Scalar Quantization)、SVQは分割ベクトル量子化(Split Vector Quantization)を示す。

対数スペクトル歪を表5.2に示す。表から、本手法によれば、特にフレーム前半部の歪が大幅に低減化され、それに伴いフレーム平均の歪も減少していることがわかる。

表 5.1: LSP 量子化ビット配分 /40 ms フレーム

項目	従来方式	提案方式
1 段目	VQ 7 ビット	VQ 7 ビット
2 段目	SQ 23 ビット	SVQ 6 ビット x3
補間コードブック (ICQ)	2 ビット	2 ビット
誤差コードブック (IEQ)	-	5 ビット
合計	32 ビット	32 ビット

SQ Scalar Quantization

SVQ Split Vector Quantization

ICQ Interpolation Coefficient Quantizer

IEQ Interpolation Error Quantizer

表 5.2: 対数スペクトル歪性能の比較

項目	平均 (dB)	フレーム前半部 (dB)	フレーム後半部 (dB)
提案方式	1.51	1.76	1.22
従来方式	1.76	2.25	1.05

## 5.2.3 ピッチラグの差分符号化

音声の有声区間では、波形にピッチ周期毎の周期性がみられる。ピッチ予測、あるいは適応コードブック [62] は、ピッチ周期性を利用してピッチ予測を行い音声を効率的に符号化する手段である。このためにはピッチ周期に対応するピッチラグを伝送する必要がある。しかし、ピッチラグは、伝送路誤りに対し感度が高いという問題がある。また、適応コードブックのピッチラグは、女性音声に対して十分な時間分解能を得るために、整数値ではなく補間して小数値で表すことが多い [69], [70]、量子化精度としてサブフレーム当たり 8 ビット程度必要である。そのため、ラグ伝送に必要なビット数やラグ探索の演算量が比較的多くなる。そこで、M-LCELP では、ピッチラグをサブフレーム間で差分符号化 [154], [136] することにより、性能を落とさずにピッチラグ伝送に必要なビット数を従来に比べ約 30% 低減し、これにより誤りに対する頑健性を向上させた。また、ラグ探索に必要な演算量を約 30% 削減した。

ラグの差分符号化の具体的な手法を以下に示す。

## 開ループ (open-loop) ピッチトラッキング

サブフレーム毎に独立に求めたピッチラグは、後述の図 5.3 (a) に示すように、必ずしも時間的に滑らかな変化をしていないため、サブフレーム間でラグを単純に差分符号化すると性能が劣化する。そこで、まず、開ループピッチトラッキング手法により、サブフレーム間で時間的に滑らかに変化するようなピッチラグの遷移パスをフレーム全体で 1 種類求める。ピッチトラッキングの方向は、時間順方向、逆方向の双方向とした。まず、聴感重み付け信号  $x_w(n)$  に対して、式 (5.2) に従いサブフレーム毎にピッチ予測ゲインを求める。次に、ピッチ予測ゲインを最大とするサブフレームにおいて、予測ゲインの大きい順にラグの候補を複数個 (例えば L 個) 出力する。ラグの候補の各々に対して、近隣サブフレームで順次ラグを探索し、トラッキングパス候補を L 種類求める。ここで、近隣サブフレームでのラグの探索範囲は、ラグを差分符号化する場合の量子化ビット数で定められる範囲とした。また、女声や子供の声で精度を向上させるために、ラグは整数値ではなく、補間フィルタを用いて補間し小数値 [69], [75], [70] を使用した。最後に、候補パスの各々について、式 (5.11) に従い、フ

レーム全体での累積歪を計算し、累積歪を最小化するパスを 1 種類選択する。

$$D_j = \sum_{l=1}^M D_{l,j} \quad (5.11)$$

上式で、 $M$ 、 $j$  はそれぞれ、フレーム内のサブフレームの個数、ピッチトラッキングパスの候補番号を示す。

## 閉ループ (closed-loop) ラグ探索

開ループピッチトラッキングにより、フレーム全体でピッチトラッキングパスが 1 種類求めた後に、第 1 サブフレームから順に閉ループでのラグ探索を行う。ここで、伝送路誤りの影響を 1 フレーム内におさえるために、ラグ符号化において第 1 サブフレームのみはラグをそのまま伝送し、第 1 サブフレーム以外では前サブフレームとのラグの差分を符号化する。ピッチトラッキングパスで示されるラグの値の近傍の区間と、一つ過去のサブフレームで閉ループ探索で求められたラグの近傍区間との重複領域を探索し、式 (5.12) の歪みを最小化するラグを選択する。ここで、近傍区間のサイズは、ラグ差分符号化のビット数で規定される。例えば、差分符号化が 5 ビットのときは、近傍区間のサイズは 32 レベルである。

$$D_d = \sum_{n=0}^{N-1} [x_w(n) - g_d S_d(n)]^2 \quad (5.12)$$

式 (5.12) で、 $g_d$  は、閉ループラグ探索における最適ゲインを示す。 $S_d(n)$  はラグ  $d$  から切り出した音源信号による聴感重み付け再生信号である。式 (5.12) を  $g_d$  で偏微分して 0 とおくことにより、次式が導ける。

$$D_d = \sum_{n=0}^{N-1} x_w^2(n) - \frac{\left[ \sum_{n=0}^{N-1} x_w(n) S_d(n) \right]^2}{\sum_{n=0}^{N-1} S_d^2(n)} \quad (5.13)$$

式 (5.13) を最小化する、つまり、右辺第 2 項を最大化するラグ  $d$  を選択すればよい。

## ラグの予備選択による演算量の削減

式(5.13)の計算を探索範囲の全てのラグについて行うのではなく、予備選択した複数個の候補のみに限定することにより、さらに演算量を低減化できる [154], [136]。予備選択は、近似的に、式(5.14)を最大化するラグを複数候補選択することで行なう。その後、選択された複数候補の各々に対して式(5.13)右辺第2項を最大化するラグを1種類選択する。

$$CC_d = \left[ \sum_{n=0}^{N-1} x_w(n) S_d(n) \right]^2 \quad (5.14)$$

## 性能

提案手法と従来手法との性能の比較を表5.3に示す。表は、ラグ伝送に必要なビットレート、 $SNR_{seg}$ 、演算量（積和回数をMOPSで表示）を比較している。提案法は式(5.14)によるラグの予備選択を含めてある（予備選択数は3）。従来法としては、開ループと閉ループの併用法（従来法1）[133]、フル閉ループ法（従来法2）[62]を示してある。なお、これらの従来法ではラグの差分符号化は行っていない。フレーム長、サブフレーム長はそれぞれ、40 msec、8 msec としてある。

表から、提案法はラグ伝送速度を30%低減化し、演算量を従来法1に比べ約30%削減可能であるにもかかわらず、 $SNR_{seg}$ の低下はわずかであることがわかる。

次に、閉ループ探索において、式(5.14)の予備選択による演算量の低減と $SNR_{seg}$ 性能との関係を表5.4に示す。

表から、候補数が3以上で性能が飽和することがわかる。そこで、閉ループ探索の予備選択数は3に設定した。これによる演算量低減は約30%である。

ピッチラグの時間的な変化を図5.3で比較する。図の横軸は時間、縦軸がピッチラグの値である。(a)は従来方式（従来法1）で求めたラグ、(b)は提案手法により求めたラグである。ここで、モード0の区間のラグは0とした。本手法によれば、ピッチトラッキング及び差分符号化を用いることにより、従来法と比べ、整数倍のラグが誤って求まる頻度が大幅に減少し、時間的に滑らかなラグが求まっていることがわかる。これは、女性話者の音質の向上に有効と考えられる。

表 5.3: ラグ探索の性能比較

項目	ラグ伝送速度 (kb/s)	$SNR_{seg}$ (dB)	演算量 (MOPS)
提案方式 (予備選択数=3)	0.7	10.27	4.9
従来方式 1	1.0	10.40	7.2
従来方式 2	1.0	10.49	42

従来方式 1: 開ループ - 閉ループ併用法

従来方式 2: フル閉ループ法

表 5.4: ラグ予備選択の性能

予備選択候補数	$SNR_{seg}$ (dB)	演算量 (MOPS)
1	10.18	4.7
2	10.25	4.8
3	10.27	4.9
4	10.27	5.0
全候補	10.27	7.2

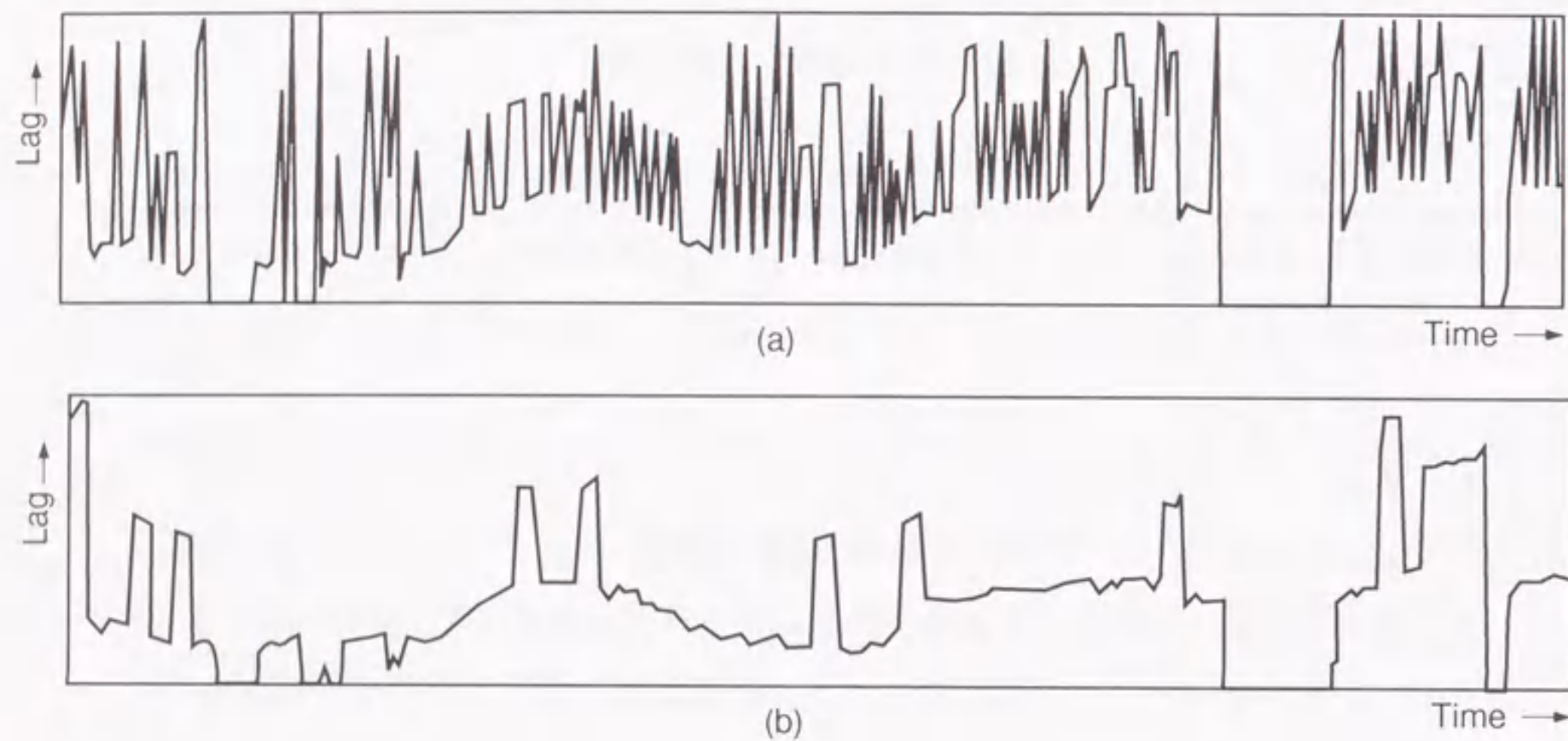


図 5.3: 適応コードブックピッチラグの時間変化の例, (a) 開ループ-閉ループ併用法 (従来法1) のラグ変化, (b) 提案法のラグ変化。 (a),(b) 共に、縦軸はラグの値, 横軸は時間を示す。

#### 5.2.4 スパース構造ベクトル量子化音源コードブックの開発と従来構造との比較

CELP 系の音声符号化では、一般に、音源コードブックの探索に多くの演算量を必要とする。また、音源コードブックの格納のためにデータメモリが多く必要である。M-LCELP では、これらを考慮し、音源信号及び音源コードブックの切り替えは、有声区間 (モード 1、2、3) と無声区間 (モード 0) の 2 種類とした。さらに、有声区間では、探索に必要な演算量と格納に必要なメモリ量の両者を削減するために、マルチパルス音源の考え方を応用したスパース構造の音源コードブック [165], [57], [19] を開発した。

一方、無声区間 (モード 0) は、摩擦性の音声や背景雑音などを表せるように、極性分離の 3 段フルパルス乱数構造を採用した。極性ビットを別に分離することにより、音源コードブックのメモリ量、演算量を  $1/2$  に低減化した。

#### 構造

スパース構造音源コードブックでは、サブフレーム内にあらかじめ定められた個数の振幅が非零のマルチパルスをたてる。第 2 章で述べたマルチパルスと異なるのは、符号化の際に 1 パルスずつ位置と振幅を探索するのはなく、マルチパルスの位置と振幅は符号化に先だてあらかじめ設計してコードブックとして固定化しておく点である。パルスの位置と振幅は、同時にベクトル量子化され、コードベクトルのパターン種類だけ用意される。符号化時には、各コードベクトルにおいて、振幅が 0 でないパルスのみを探索すればよいこと、振幅が 0 でないパルスの位置と振幅のみを格納すればよいので、従来のコードブック構造と比較して演算量、メモリ量の削減が可能となる。

各パルスの振幅と位置をあらかじめどのように決定しておくかで性能が異なる。文献 [19] と提案方式との違いは、文献 [19] では、乱数信号から振幅の大きなパルスを取りだし、それ以外を振幅 0 としてコードブックを作成しているのに対して、提案法では音声データベースを用いてあらかじめ学習により、最適なパルスの振幅と位置

を計算しコードブックを作成する点である。提案法の方がより高い SNR 性能が得られる。また、この考えは、ほぼ同時期に独立して検討された文献 [57] とほぼ同じである。

#### 演算量の削減

スパースコードブックにおいて、演算量、メモリ量をさらに低減するため、2段構成 [56], [91], [86] とし、予備選択 - 本選択の2段探索 [91], [86], [135] を用いる。2段構成の場合、音源信号  $v(n)$  は式 (5.15) のように表せる。

$$v(n) = g_e \left[ \sum_{i=1}^K g_{1i} \delta(n - m_i) + \sum_{j=1}^K g_{2j} \delta(n - m_j) \right], \quad (n = 0, \dots, N-1) \quad (5.15)$$

ここで、 $K$  はコードベクトル当たりのパルスの個数、 $g_{1i}$ 、 $g_{2j}$  は、それぞれ、1段目、2段目のスパースコードベクトルにおけるパルスの振幅、 $m_i$ 、 $m_j$  はそれぞれ、1段目、2段目のパルスの位置、 $g_e$  は音源のゲインを示す。

#### (1) 予備選択:

1段目、2段目のコードブックから、それぞれ式 (5.16)、式 (5.17) を最大化する順に、複数候補 ( $Q$  個) を予備選択する。

$$D_{ex1} = \left[ \sum_{i=1}^K g_{1i} \psi(n - m_i) \right]^2 \quad (5.16)$$

$$D_{ex2} = \left[ \sum_{j=1}^K g_{2j} \psi(n - m_j) \right]^2 \quad (5.17)$$

ここで、式 (5.16)、式 (5.17) の積和計算は、パルスのたつ (振幅が零でない) サンプル位置のみを対象とすればよい。また

$$\psi(n) = \sum_{l=n}^{N-1} z_w(l) h_w(l - n), \quad n = 0, \dots, N-1 \quad (5.18)$$

$$z_w(n) = x_w(n) - g_a S_d(n) \quad (5.19)$$

である。式 (5.18)、式 (5.19) で  $g_a$ 、 $S_d(n)$ 、 $h_w(n)$  はそれぞれ、適応コードブックのゲイン、ピッチラグ  $d$  の適応コードベクトル聴感重み付け再生信号、聴感重み付け合成フィルタのインパルス応答である。

#### (2) 本選択:

予備選択で選択された1段目、2段目の候補の組み合わせについて、式 (5.20) の歪を最小化する組み合わせを  $N_e$  ( $N_e \leq Q$ ) 種類選択する。

$$D_{ex} = \sum_{n=0}^{N-1} \left[ x_w(n) - g_a S_d(n) - g_e \left[ \sum_{i=1}^K g_{1i} h_w(n - m_i) + \sum_{j=1}^K g_{2j} h_w(n - m_j) \right] \right]^2 \quad (5.20)$$

#### スパースコードブックの学習

1段目、2段目のスパースコードブックのマルチパルスの振幅と位置は、多量の音声データベースを用いて、一般化ロイドアルゴリズムに基づき [72], [30]、聴感重み付け距離を最小化するように、あらかじめ閉ループ学習により設計する。コードブック設計の際に、マルチパルスの振幅と位置を計算する方法は、オフライン学習のため、演算量的に多くても性能のよい方が望ましいので、第2章 2.3.3 節に記述した方式2を使用した。クラスタリングの後、各クラスタ毎に、定められた個数のパルスの位置と振幅を方式2により計算する。

コードブック学習の際の学習曲線を図 5.4 に示す。学習により性能 (Weighted SNR : 聴感重み付け SNR) が向上しており、30 回学習することで、初期乱数コードブックに比べ 1.7 dB 改善されている。

学習後の音源コードベクトルのパルスの例を図 5.5 に示す。ここでは、5 ベクトル分示している。サブフレーム長は 64 サンプル (8 ms) で、サブフレーム当たりのパルスの個数は 8 としている。(a) には学習をする前の初期ガウス乱数コードベクトルを示し、(b) には 30 回学習後のパルスの様子を示す。

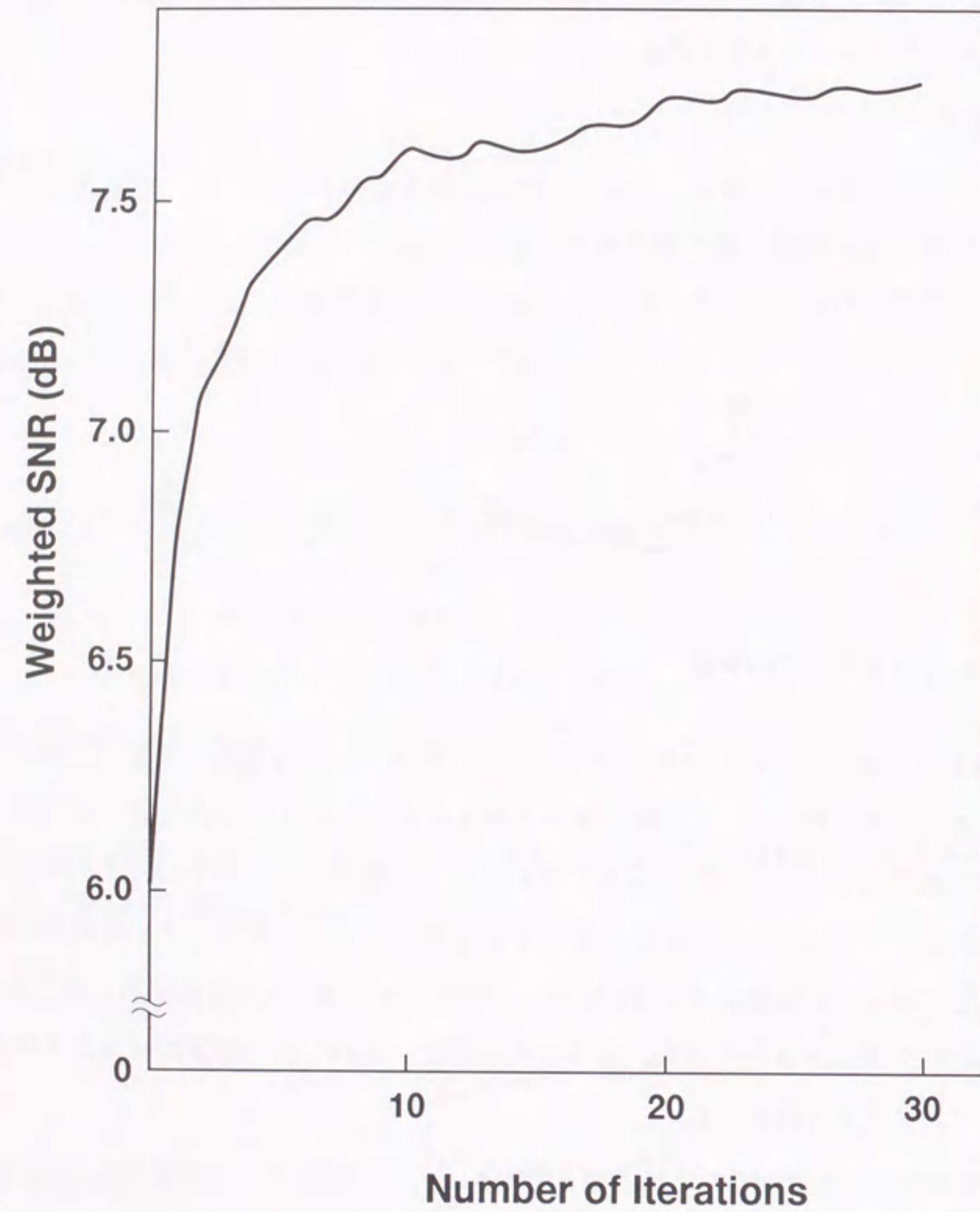


図 5.4: スパースコードブックの学習による Weighted SNR の向上

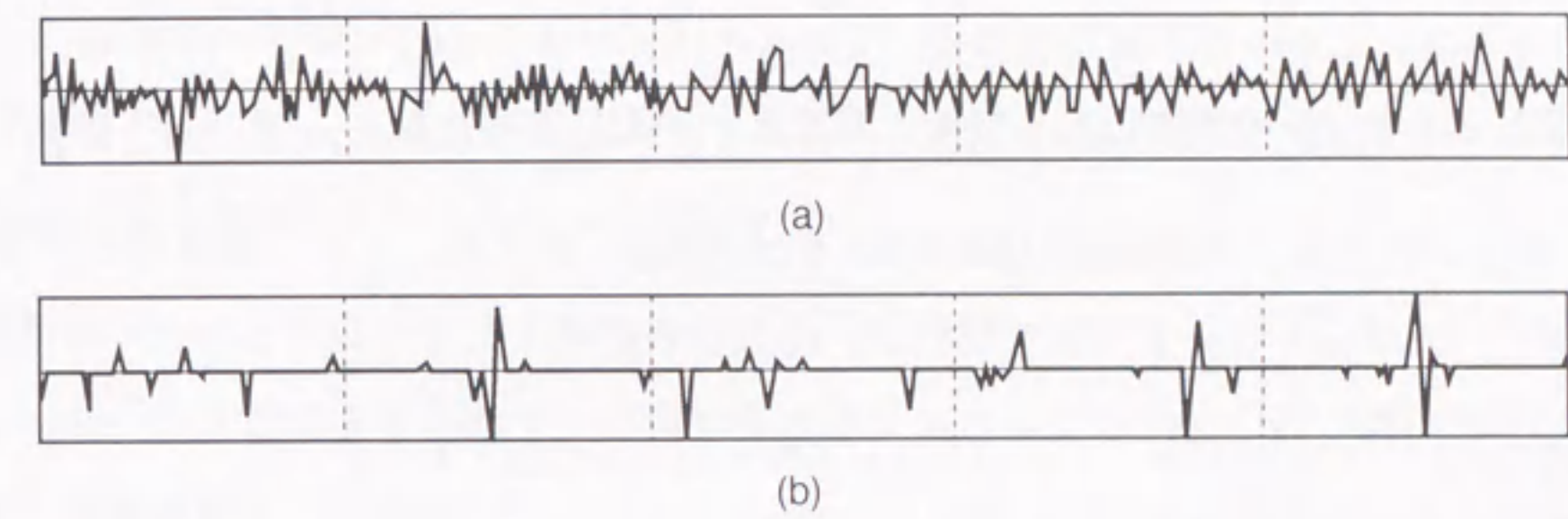


図 5.5: スパース音源コードブックの波形例, (a) は初期乱数コードベクトル, (b) は 30 回学習後のコードベクトル (パルスの個数は 8 である)。いずれも 5 ベクトル分を示す。



表 5.5: コードブックの構造による  $SNR_{seg}$  性能、処理量、メモリ量の比較

項目	$SNR_{seg}$ (dB)	演算量の比	メモリ量の比
提案法 (2段スパース構造)	9.33	0.63	0.25
2段レギュラーパルス構造	9.34	0.63	0.5
2段フルパルス構造	9.26	1.0	1.0

## 音源コードブックの構造による性能、演算量、メモリ量の比較

音源コードブックの構造の違いによる  $SNR_{seg}$ 、演算量、メモリ量の比較を表 5.5 に示す。スパースコードブック以外に、従来型の構造として、レギュラーパルス構造 [68], [67], [20], [2], [87] やフルパルス構造の 2 段コードブックを比較している。いずれもサブフレーム長は 64 サンプル (8 ms)、ビット数は、1 段目、2 段目それぞれ、5 ビット、4 ビットとし、同一の音声データベースを用いて 30 回学習した後に測定した。表において、演算量、メモリ量は、従来型 2 段フルパルス構造の値を 1.0 として正規化してある。表から、提案のスパースコードブックによれば、 $SNR_{seg}$  性能は他の従来方式とほぼ同等ながら、演算量を従来フルパルスの 60% 程度に、メモリ量をフルパルス、レギュラーパルスに比べ、それぞれ、25%、50% に削減できることがわかる。

## 5.2.5 適応型コムフィルタ

ビットレートが 4 kb/s と低く、サブフレーム長が 8 ~ 10 msec と長い場合、ピッチ周期の短い女声話者や子供の声では、サブフレーム内に複数個のピッチが含まれるため、適応コードブックのみでは音声のピッチ周期性を十分良好に表すことが困難となる。このため、適応コードブックでピッチ予測した成分を入力音声から減算した後の信号  $z_w(n)$  にはピッチ周期成分が残ることになり、音源コードブックでは周期性を充分表せないために、これらの声に対しては音質が顕著に劣化することになる。

これを防ぐために、有声区間では、音源コードブックの探索時に、ピッチ周期性

を強調する方法がとられる。例えば、音源コードベクトルの先頭からピッチ周期の長さだけ切り出した波形をサブフレーム内でピッチ周期毎に繰り返して周期化する方法 [80], [106] や、自己回帰・移動平均型 (極零型とも呼ぶ) (ARMA: Auto-Regressive and Moving-Average) コムフィルタを用いてピッチ周期性を強調する方法 [177] などが提案されている。

コムフィルタはピッチの調波成分に重畳する量子化雑音を抑圧する働きを行うが、音源コードベクトルとの整合性が重要なため、M-LCELP では、式 (5.20) で示したように、音源コードブックの探索で選択された  $N_c$  個の音源コードベクトルの各々を移動平均型 (MA: Moving-Average) コムフィルタに通し [135], [136]、式 (5.21) の歪  $D_c$  を最小化する音源コードベクトルを選択する。

$$D_c = \sum_{n=0}^{N-1} [x_w(n) - g_a S_d(n) - g_c S_c(n)]^2 \quad (5.21)$$

$$S_c(n) = S_{ex1}(n) + S_{ex2}(n) + \epsilon [S_{ex1}(n-d) + S_{ex2}(n-d)] \quad (5.22)$$

ここで、 $S_c(n)$  は 2 段の音源信号をピッチラグ  $d$  の遅延を有するコムフィルタに通した後の聴感重み付け合成信号である。 $\epsilon$  は MA 型コムフィルタの係数でありモード毎に異なる定数値を用いた。MA 型コムフィルタを用いた理由は、ARMA 型のようにフィードバックループを持たないために、伝送路誤りの影響を受けにくいためである。

MA 型コムフィルタの組み込みにより、 $SNR_{seg}$  が平均で約 0.2 dB 改善され、特に女性話者に対して音質の向上が確認された。

## 5.3 実験

## 5.3.1 実験条件

以上説明した M-LCELP の音質を評価するために実験を行なった。ハーフレート移動体通信への適用を考慮し音声符号化のビットレートは 4.1 kb/s とした。ビット配分を表 5.6 に示す。フレーム長は 40 msec、サブフレーム長は 8 msec である。LSP の量子化では、補間コードブックに 2 ビット、誤差コードブックに 4 ビットを使用し

表 5.6: 4.1 kb/s M-LCELP のビット配分

パラメータ	ビット数 / フレーム	
	母音区間	子音区間
LSP (10 次)	28	30
モード	2	2
RMS	6	6
適応コードブック	28 (8+5+5+5+5)	-
音源コードブック	(7+6) x 5	(7+6+6) x 5
ゲインコードブック	7 x 5	6 x 5

た。LSP コードブック、フレーム内平均レベルを符号化する RMS (Root Mean Square) コードブック、音源コードブックは、有声モード (モード 1, 2, 3) と無声モード (モード 0) で切り替えた。ゲインコードブックは、各モード毎に切り替えた。また、デコーダ出力には、雑音感を低減するために、文献 [16] に示すようなポストフィルタを付加した。

なお、背景雑音対策として、モード 0 のフレームでは、デコーダで RMS 復号値を 3 dB 低減させるという簡単な抑圧処理を付加した。

全てのベクトル量子化コードブックは、約 300,000 ベクトルを含む多量の音声データベース (約 40 分) を用いてあらかじめ学習して作成した。音質評価には、学習に使用していない約 8 分の音声信号を使用した。

### 5.3.2 誤り訂正の付加

誤りなしの条件に加え、北米などでのハーフレート移動通信環境での伝送路誤りに対する評価も行うために、誤り訂正として、2.4 kb/s の軟判定誤り訂正符号 FEC (Forward Error Correction) [38] を付加し、総ビットレートを 6.5 kb/s に設定した。ここで、M-LCELP のモードビットは誤り感度が高いので特に強く保護した。さらに、

主観的に重要なビットについては誤り検出符号を付加した上で、フレーム当たり 164 ビットのうちの 78 ビットを FEC で保護した。FEC には畳み込み符号化 / 軟判定ビット復号化アルゴリズムを用いた [38]。また、バースト誤り条件に対処するため、バースト誤りを拡散させるためのインターリーブを M-LCELP に付加するが、遅延の点から 2 フレーム構成とした。デコーダで訂正できなかった誤りを検出した場合は、過去の正しく受信したフレームのパラメータを繰り返し、音質の劣化を防いだ [38]。

### 5.3.3 主観音質評価

北米などのハーフレート移動体通信を想定し、次の条件 [168] について、第 2 章 2.4.1 節で述べた ACR 尺度による音質 MOS (Mean Opinion Score) 評価を行なった。評価した条件は、(a) 2 種の入力特性：フラット特性と、電話器の送話側の周波数特性を模擬した IRS (Intermediate Reference System) 特性 [13]、(b) 子供の声を含む話者依存性条件、(c) タンデム接続条件、(d) 2 種の背景雑音条件：車走行雑音 (S/N=10dB) とバブル雑音 (S/N=15dB)、(e) 伝送路誤り条件 (1% 及び 3% バースト誤り、各誤り率共に車速 3 種類) である。

話者依存性条件は、他の条件と異なる男性、女声話者を用い、子供の声も評価に含めた。基準方式として、北米フルレート携帯電話標準方式の 13 kb/s VSELP (Vector Sum Excited LPC) [22], [31] (音声符号化に 7.95 kb/s、誤り訂正に 5.05 kb/s) も評価した。評価には、第 2 章 2.4.1 節で示した 5 段階 ACR (Absolute Category Rating) 尺度 [46] を使用した。言語は英語とし、英米人 20 人が評価に参加した。

結果を表 5.7 に示す。

表から、全条件の平均では、M-LCELP は、ビットレートが 2 倍の VSELP と同等という良好な音質が得られていることがわかる。以上により、5.1 節で述べた要求条件のうち、(1)、(2)、(3) の条件を満足できることが明らかとなった。

### 5.3.4 考察

実験の結果、次の点が明らかとなった。

(1) 入力特性条件 (フラット、IRS)、タンデム接続条件では、M-LCELP は VSELP

表 5.7: MOS 音質評価結果

条件	M-LCELP	VSELP
フラット特性	3.61	3.60
IRS 特性	3.34	3.40
話者依存性	3.21	3.40
タンデム接続	2.86	2.94
車走行雑音 (S/N=10 dB)	2.26	2.10
バブル雑音 (S/N=15 dB)	2.44	2.45
1% バースト誤り	2.95	2.82
3% バースト誤り	2.64	2.63
平均	2.91	2.91

M-LCELP: 音声 4.1 kb/s + FEC: 2.4 kb/s

VSELP: 音声 7.95 kb/s + FEC 5.05 kb/s

とほぼ同等の音質を得た。

(2) 話者依存性条件では、VSELP の音質に到達しなかった。これは、特にピッチの短い子供の声で音質が不十分であったためである。

(3) 車走行雑音 (S/N=10dB) では、M-LCELP の方が良好であった。これは、デコーダに付加した簡単な抑圧処理が雑音感の低減に効果的であることを示している。

(4) 伝送路誤り条件では、M-LCELP は VSELP よりも一貫して良好であった。これは符号化アルゴリズム自体の誤りに対する頑健性と適切な FEC の付加による効果と考えられる。

表 5.8: 試作した M-LCELP 音声 CODEC 装置の処理量、メモリ量

項目	MOPS	MIPS	RAM (kワード)	データ ROM (kワード)
M-LCLEP	18.2	32.2	5.7	15.9
VSELP	10.0	18.0	2.5	8.0

#### 5.4 DSP による音声 CODEC の試作

32ビット浮動小数点 DSP (命令実行速度 40 ns) を用いて M-LCELP 音声 CODEC を試作した。エンコーダ、デコーダにそれぞれ、1チップの DSP を使用している。試作した装置の外観を図 5.6 に示す。

CODEC の演算量 (MOPS: Million Operation Per Second)、DSP での命令数 (MIPS: Million Instruction Per Second)、RAM 容量、コードブック格納に必要なデータ ROM 容量を表 5.8 にまとめる。

表から、M-LCLEP の MOPS 数、MIPS 数は VSELP の 2 倍以下であり、処理能力が 35 ~ 40 MIPS 程度の最新の DSP を用いれば、1チップで実現可能であることがわかる。これにより、M-LCLEP は、5.1 節の要求条件 (4) も満足でき、全ての条件を満足することができることが明らかとなった。しかし、コードブックを格納するためのデータ ROM 容量が VSELP のほぼ 2 倍とやや多いので、これをさらに削減するのが今後の課題である。

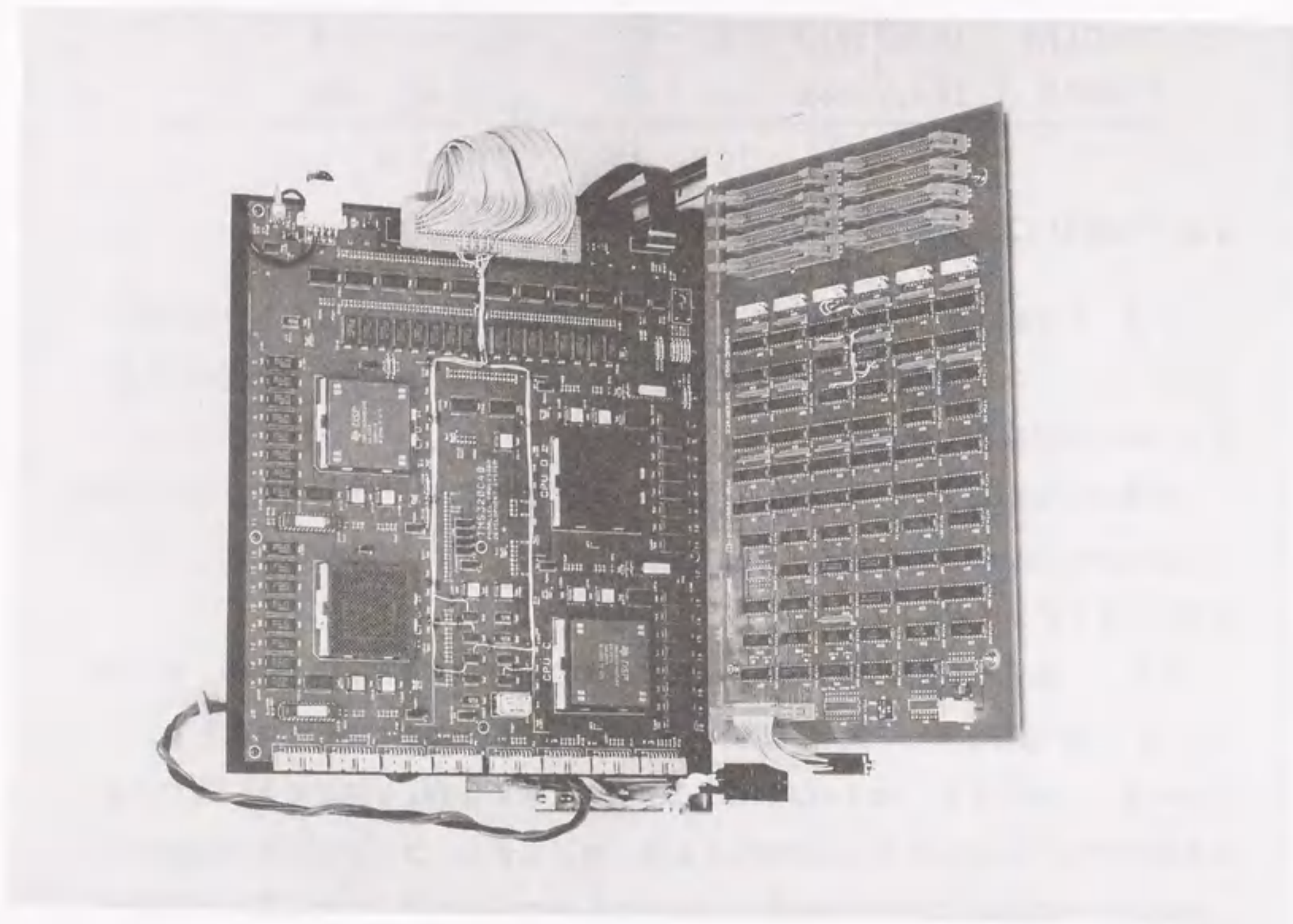


図 5.6: 試作した M-LCELP 音声符号化装置

## 5.5 むすび

本章は、線形予測分析と A-b-S に基づく音声符号化の枠組はそのままとし、低伝送速度での音質をさらに改良するために、ベクトル量子化音源について検討した。ここでは、従来の CELP (Code Excited Linear Prediction) 方式の低伝送速度における音質改良も目的とし、4 kb/s M-LCELP (Multi-mode Learned CELP) 音声符号化を提案した。本方式は、第4章で提案した音声のモード毎に音源信号を切り替えるマルチモード符号化を CELP に適用し、モードによりベクトル量子化コードブックを切替えることにより、量子化の効率化をはかった。さらに、ベクトル量子化音源において、3種の音源コードブック構造について、 $SNR_{seg}$  性能、演算量、メモリ量を比較検討し、マルチパルスから構成されるスパース音源構造を採用した。スパース音源におけるパルスの振幅、位置は、あらかじめコードブック学習により、各コードベクトル毎に決定した。30回の学習後、重み付け SNR を 1.7 dB 向上させることができた。さらに、適応コードブックのラグの差分符号化技術などを開発した。音質 MOS 評価結果によれば、誤り無し、背景雑音や誤り有りなどの種々の評価条件において、ビットレートが2倍の北米フルレート標準 VSELP と同等という良好な音質が得られた。また、VSELP の2倍以下の演算量で試作 CODEC が実現できた。これらにより、提案方式は、5.1節で設定した4種の要求条件を満足できることが明らかになった。

なお、本方式は、移動体通信に限らず、マルチメディア通信、パソコン通信、音声蓄積など、様々な分野に適用することが可能である。

今後は、音質をさらに改善すること、移動体携帯端末などでの実現の容易性を考慮し、演算量、メモリ量を一層削減化することが必要であり、これらの実現のために、音源モデルの改善が課題である。また、フレーム長が 40 ms と長いため、遅延の低減化も重要である。

## 第6章

### マルチパルスベクトル量子化音源と高速探索に基づく MP-CELP 音声符号化

#### 6.1 まえがき

デジタル移動体通信の分野では、限られた無線周波数容量の中でチャンネル数を拡大するために、1980年代末から主に音声 CODEC のビットレートを低減化する方向で方式の標準化が行われてきた。しかし、今後は、無線伝送速度の高速化とともに、マルチメディア化や、有線と無線通信のシームレス接続が浸透していくことが予想される。このような背景の中で、音声符号化への要求は、有線通信と同等の高品質を確保する方向に変化していくと予想される [138], [141]。

北米の PCS (Personal Communications Systems) では、音質が ITU-T (International Telecommunications Union-Telecommunications Standards Sector) 標準 G.726 (32 kb/s ADPCM) [12] と同等の 13 kb/s 音声 CODEC が採用される予定である [23]。日本の PHS (Personal Handyphone Systems) では、音質を重視して 32 kb/s ADPCM が採用されている。このような背景の中で、欧州、米では、すでに商用化されているフルレート携帯電話の品質を有線電話品質と同等に高め、ユーザ数を増大させるために、13 kb/s あるいは 8 kb/s の EFR (Enhanced Full-Rate) 音声 CODEC を採用する予定である。また、これらのシステムは将来的には、FPLMTS (Future Public Land Mobile Telecommunication Systems) や UMTS (Universal Mobile Telecommunication Systems) などの次世代移動体通信システムに移行していくことが予想さ

れる [138]。

本章はこれらの状況を考慮し、欧州 ETSI (European Telecommunications Standards Institute) で定められた EFR の条件 [24] をもとに次のような品質要求条件を設定し、これらを満たすような音声符号化方式を検討する。

(1) 音質は、伝送路誤りなしのクリーン音声に対する基本特性に対して、有線通信において現在最良で 32 kb/s ADPCM を越える音質を有する ITU-T 標準 G.728 (16 kb/s LD-CELP) [15] と同等の音質を達成する。伝送路誤りの条件に対しては、誤り訂正込で既存の同一ビットレートの標準符号化に大幅に勝る誤り耐性を確保する。また、背景雑音に対しても G.728 と同等の音質を得るのが望ましい。

(2) 処理遅延は、タンデム接続などを考慮すると、エンコーダとデコーダ合計の一方遅延が 40 ms 以下 (ITU-T 勧告 G.174) を満足できるよう、遅延の低減化をはかるのが望ましい。

(3) 演算量に関しては、移動体携帯端末などでの実現を考慮すると、DSP 1 チップに納まる処理量が望ましい。

前章までに、マルチパルス音源を用いる符号化、マルチモードとベクトル量子化を用いる M-LCELP (Multi-mode Learned CELP) 方式を検討した。マルチパルス符号化においては、演算量は CELP に比べ少ないものの、音源信号などの量子化にスカラ量子化を用いたため、量子化効率の改善の余地があった。また、M-LCELP では、ベクトル量子化音源を用いたために、量子化効率は高いが、探索に必要な演算量とコードブック格納に必要なメモリ量が、マルチパルス符号化に比べかなり多く、演算量とメモリ量の一層の削減が必要であった。一般に、CELP 系の音声符号化では、音源コードブックの探索に多くの演算量が必要である。この問題は、ビットレートが高くなるほどコードブックのビット数が増加するために、顕著となっていた。

近年、CELP において、演算量、メモリ量を削減するために、比較的簡便な方法として、有声区間の 1 つのピッチ区間の音源信号を単一のパルスで表す方法 [33] や、パルス位置の異なる単一パルスをパルスコードブックとしてもち、有声区間では、ピッチ周期だけ離れた位置にパルスを繰り返しながら使用する手法 [184], [163], [55], [153] が提案されている。基本的には、1 つのピッチ区間の音源を 1 パルスで表現すること、

ピッチ周期が誤った場合に補正が困難なことなどのために、音源波形を良好に表すには充分ではなかった。一方、演算量、メモリ量の削減のみならず、音質も改善するために、マルチパルス音源に改良を加え、CELP に組合せた方式として、ACELP (Algebraic CELP) [147],[71]、1995 年に ITU-T 標準に選定された G. 729 CS-ACELP (Conjugate Structure-ACELP) [148] や G. 723.1 MP-MLQ (Multi-Pulse Maximum Likelihood Quantization) [49] などが開発されている。ACELP、CS-ACELP や MP-MLQ では、マルチパルスの振幅を極性のみで表すことや、ACELP、CS-ACELP ではさらに、探索時に各パルスの位置を限定することでパルス伝送に必要なビット数を低減化している。

しかしながら、これらの方式では、パルスの振幅を極性のみで表しているため、性能改善の余地が残されていた。また、パルスの位置の探索法は、性能と演算量を決める重要な技術であり、今までに ACELP や CS-ACELP において、しきい値を用いて探索回数を制限することで演算量をおさえる手法が提案されているが [71]、しきい値処理を必要とするため、別の手法についても検討する余地があった。

これらを考慮して、本章は、従来の CELP に比べ、比較的少ない演算量、メモリ量を維持しながら、高音質化を狙った方式として、MP-CELP (Multi-Pulse based CELP) 方式を提案する [139], [140]。提案方式では、5 章の M-LCELP をもとに、音源信号として、マルチパルス音源にベクトル量子化を組合せて使用する。マルチパルスの振幅をまとめてベクトル量子化することで効率を高め、複数セットの位置候補と振幅ベクトル量子化コードブックを組合せて探索することにより、性能を改善する。各パルスのとりうる位置は、ACELP や CS-ACELP と同様の代数構造 [148] を用いて限定し、ビット数を低減する。さらに、位置を総組合せで最適に探索する場合に、しきい値処理の不要なパルス位置の高速探索法を提案し、演算量を大幅に削減しても性能の低下が少ないことを示す。これらにより、従来の CELP に比べ、高音質化と、演算量及びメモリ量の削減が可能であることを示す。

以下では、提案方式の特徴を述べ、性能改善量と演算量の関係を示す。次に、音質評価結果を示す。欧州 GSM フルレート携帯電話の音質改善を狙った 11 kb/s MP-CELP の音質は、演算量を同程度におさえた M-LCELP [137] よりも良好で、誤りな

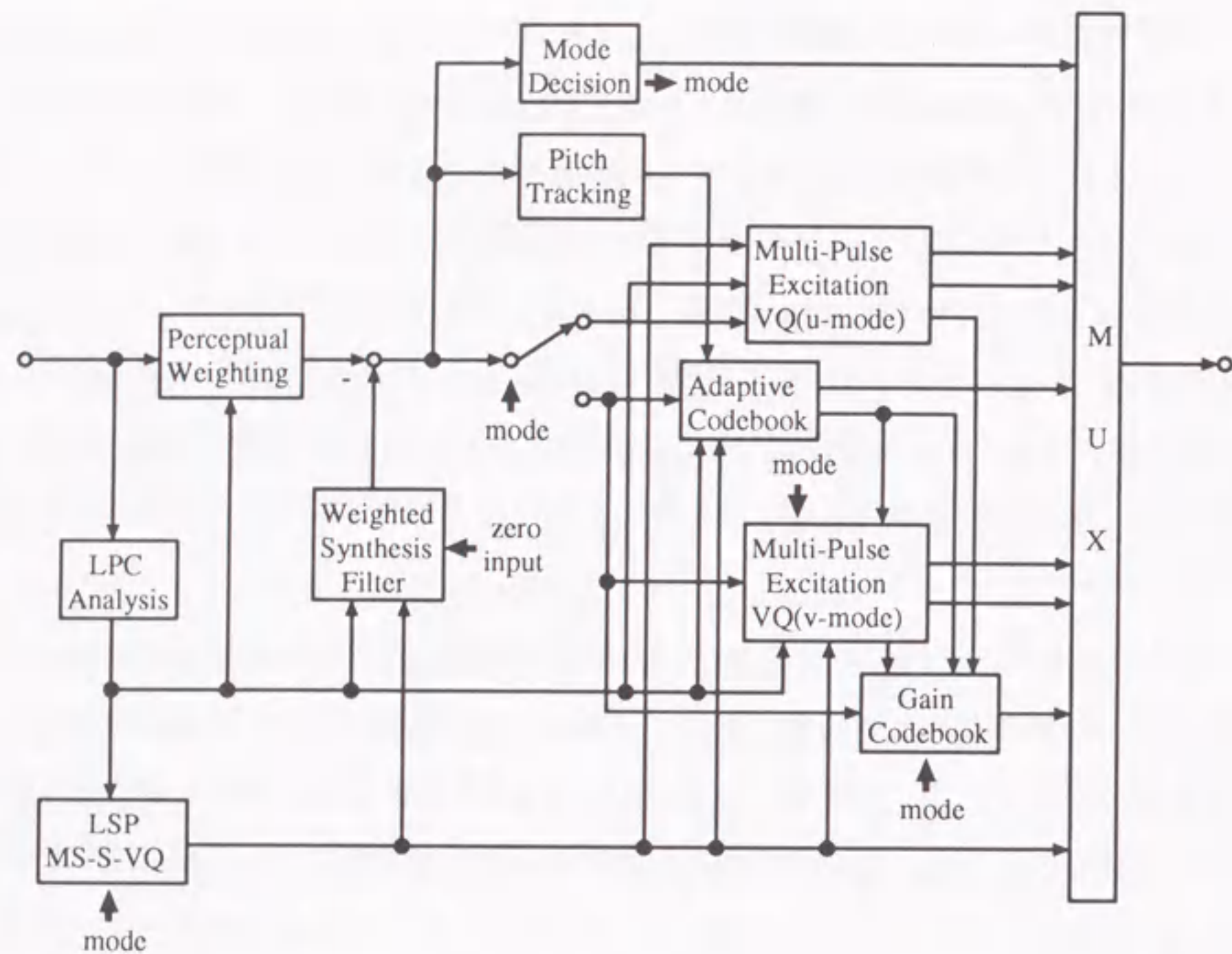


図 6.1: MP-CELP エンコーダのブロック図

し的时候に、ITU-T 標準 G. 728 (16 kb/s LD-CELP) [15] と同等であることを明らかにする。

## 6.2 MP-CELP 音声符号化

本方式のブロック図を図 6.1 に示す。5 章の M-LCELP [136] におけるマルチモード、マルチコードブックの構成を採用している。モード判別部は、5 章の M-LCELP と同一の方法を用いており、6.2.1 節の方法により、フレームの音声をモードに分類する。モードの個数は 4 としている。モード 0 は無声部に、モード 1、2、3 は有声部にほぼ対応する。LSP の VQ コードブックは、有声部、無声部で切り替える。有声部では、移動平均 (MA: Moving Average) 型予測ベクトル量子化を導入している [92], [107], [101]。適応コードブックは有声部のみに使用する。適応コードブックのラグは、

## 6.2. MP-CELP 音声符号化

M-LCELP と同様にフレーム全体で開ループピッチトラッキングした後に、サブフレーム間で閉ループ差分符号化する [154]。音源には、6.2.2 節で述べるマルチパルスベクトル量子化音源を使用する。ゲインコードブックは、4 種のモードの各々について切り替える。

### 6.2.1 モード判別

モードは、フレーム音声の平均ピッチ予測ゲイン  $G$  を用いて判別する。 $G$  は次式により計算する。

$$G = \frac{\sum_{l=1}^{N_s} 10 \log_{10}(GP_l)}{N_s} \quad (6.1)$$

$GP_l$ ,  $N_s$  は、第  $l$  サブフレームのピッチ予測ゲイン、フレーム内のサブフレームの個数をそれぞれ示す。 $GP_l$  は次式から求める。

$$GP_l = \frac{\sum_{n=0}^{N-1} x_w^2(n)}{A} \quad (6.2)$$

$$A = \sum_{n=0}^{N-1} x_w^2(n) - \frac{\left[ \sum_{n=0}^{N-1} x_w(n)x_w(n-T) \right]^2}{\sum_{n=0}^{N-1} x_w^2(n-T)} \quad (6.3)$$

ここで、 $x_w(n)$ 、 $T$  は、それぞれ、聴感重み付け後の音声信号、サブフレームのピッチ予測ゲインを最大にするピッチ周期を示す。

なお、フレーム長が 10 ms と短い場合に、モード判別誤りを低減するために、過去のフレームのモードとの時間相関を利用して、モード判別誤りを補正した [166]。

### 6.2.2 マルチパルスベクトル量子化音源

2 ~ 4 章で述べたマルチパルス音声符号化では、マルチパルスの各パルスの振幅、位置ともに、独立にスカラ量子化していたため、量子化に多くのビット数が必要であっ

た。さらに、各パルスの位置は、処理単位に属する全てのサンプル位置を候補としていたため、量子化ビット数の低減が困難であった。

これらのビット数を低減化する方法として、ベクトル量子化 (VQ: Vector Quantization) を導入する。マルチパルス音源に対するベクトル量子化の適用法としては、振幅情報あるいは位置情報のどちらか片方、もしくは両方が考えられる。後者については、第5章で述べた M-LCELP で導入したスパース音源コードブックが、振幅、位置を同時にベクトル量子化する場合に相当する。

#### 振幅のベクトル量子化

本章は、次の理由により、振幅のベクトル量子化を検討する。(a) パルスの位置の探索には、6.2.4 節で述べる高速探索法を使用できるため、位置の探索の後で振幅を VQ する方が演算量の削減が可能となる。(b) 複数セットの位置の候補と振幅コードブックとを組合せ探索することで、振幅の量子化性能を改善することができる。(c) 位置をベクトル量子化することで、振幅の大きなピッチパルスの位置が正しい位置からずれると、音質が劣化することが予想される。(d) パルスの位置は、ACELP と同様の手法を用いてとり得る位置を限定することにより、性能をほとんど落さずに伝送ビット数を低減できる。

振幅をベクトル量子化するマルチパルスベクトル量子化音源信号  $v(n)$  は、式 (6.4) のように表わせる。

$$v(n) = \gamma \sum_{k=1}^M g'_{kj} \delta(n - m_k) \quad (6.4)$$

ここで、 $M$  は、処理単位にたてるマルチパルスの個数、 $g'_{kj}$ 、 $m_k$  は、それぞれ、 $j$  番目のパルス振幅コードベクトルにおける  $k$  番目のパルスの振幅、 $k$  番目のパルスの位置を示す。 $\delta$  はクロネッカーのデルタ、 $\gamma$  は音源のゲインを示す。

マルチパルスベクトル量子化音源を用いるエンコーダの原理を図 6.2 に示す。パルスの位置、振幅コードブックは、演算量削減のため、以下に示す方法により 2 段階で探索される。また、パルスの位置は、6.2.4 節の方法により高速探索される。

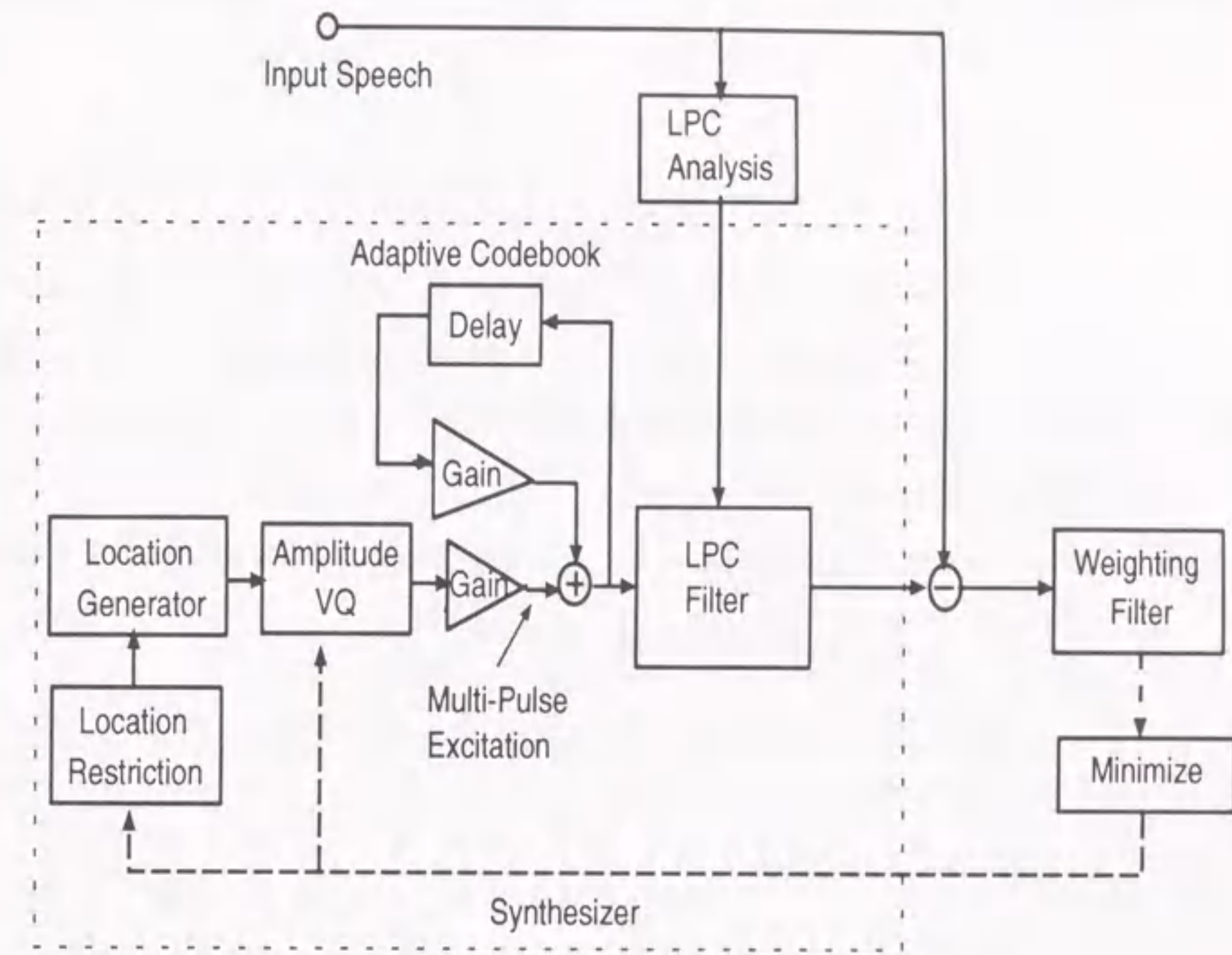


図 6.2: マルチパルスベクトル量子化音源を用いるエンコーダの原理図

#### 2 段階探索法

パルス振幅コードベクトルの各々に対し  $M$  個の位置を探索し、最適な組合せを決定するのが理想であるが、ビットレートが比較的高くパルス数が多い場合は振幅コードブックのサイズが大きいため、位置の高速探索法などを適用しても演算量が増大する。

そこで、演算量を低減化する探索法として、準最適であるが、2 段階探索法を提案する。まず、6.2.4 節の方法で、振幅を固定しながら位置を探索し、歪みの小さい順に複数セットの位置の候補を出力する。次に、これらの位置候補に対して振幅コードブックを組合せ探索し、歪みを最小にする組合せを決定する。位置の探索に 6.2.4 節で述べる高速探索を用いることで、演算量を大幅に削減することができる。



位置が与えられている場合、振幅ベクトル量子化コードブックの探索は、式(6.5)を最小化するように行なう。

$$D_j = \sum_{n=0}^{N-1} \left[ z(n) - \gamma \sum_{k=1}^M g'_{kj} h_w(n - m_k) \right]^2 \quad (6.5)$$

$$z(n) = x_w(n) - \beta b_w(n) \quad (6.6)$$

ここで、 $\beta$ ,  $b_w(n)$ ,  $z(n)$ ,  $h_w(n)$  は、適応コードブックのゲイン、適応コードブック聴感重み付け再生信号、適応コードブック減算後の聴感重み付け信号、聴感重み付け合成フィルタのインパルス応答をそれぞれ示す。

式(6.5)の $\gamma$ に関する最小化は次式の最大化に等価となる。

$$D_j = \frac{C_j^2}{E_j} \quad (6.7)$$

ここで、パルスの位置を最適に求めるために、第2章で述べた逐次探索でなく、位置の総組合せを探索することになると、

$$C_j = \sum_{k=1}^M g'_{kj} \psi(m_k) \quad (6.8)$$

$$E_j = \sum_{k=1}^M g'_{kj}{}^2 \phi(m_k, m_k) + 2 \sum_{k=1}^{M-1} \sum_{i=k+1}^M g'_{kj} g'_{ij} \phi(m_k, m_i) \quad (6.9)$$

$$\psi(n) = \sum_{i=n}^{N-1} z(i) h_w(i - n), \quad n = 0, \dots, N-1 \quad (6.10)$$

$$\phi(p, q) = \sum_{n=\max(p,q)}^{N-1} h_w(n-p) h_w(n-q), \quad p, q = 0, \dots, N-1 \quad (6.11)$$

となる。式(6.10)、式(6.11)は、それぞれ、第2章の式(2.10)、式(2.11)と同一であるが、本章では位置を総組合せで探索するため、 $C_j$ 、 $E_j$ の計算法が第2章とは異なっている。

#### ピッチフィルタのパルス探索への組み込み

有声部において、ピッチ周期が処理単位よりも短い場合は、第3章3.2.1節のピッチ予測マルチパルス(P-1法)の考えを導入する。パルス探索の際に、式(6.12)の伝達

特性を有するピッチフィルタを聴感重み付け合成フィルタのインパルス応答  $h_w(n)$  に含め、インパルス応答を第3章の式(3.4)もしくは式(3.6)を用いて計算する。

$$P(z) = \frac{1}{1 - \beta' z^{-T}} \quad (6.12)$$

ここで、 $T$ ,  $\beta'$  はそれぞれ、ピッチ周期、ピッチゲインを示し、適応コードブックで求めた値を使用する。但し、ピッチゲインに関しては前サブフレームの値を使用する。

これにより、ピッチ周期離れた位置にパルスを発生させながらパルスの探索を行なうことと等価となり [112], [117], [148]、ピッチが短い話者に対する音質を改善することができる。合成時にはピッチフィルタをLPCフィルタに縦続に接続して信号を合成する。

#### 6.2.3 組合せ探索による振幅ベクトル量子化の性能改善

振幅ベクトル量子化の性能改善のため、ディレイドデシジョンの考え [73], [1], [26] を導入し、複数セットのパルス位置の候補と振幅コードベクトルとを組合せ探索し、式(6.7)を最大化する位置の候補と振幅コードベクトルの組合せを選択する。ここで、パルスの位置の探索は、6.2.4節で述べるように、演算量削減のために振幅を固定し、CS-ACELP [148]と同様の手法であらかじめ抽出した極性を用いて行なった。

パルス振幅コードブックは、式(6.5)の歪み尺度を用いて、閉ループ学習により、あらかじめ設計しておく [72], [30]。

なお、振幅コードブックのビット数が大きい場合には、第5章5.4節で述べた予備選択を導入することで、演算量を削減できる。例えば、位置候補の各々について、式(6.7)の分子のみを最大化する振幅コードベクトルを複数種予備選択する。これらに対してのみ、式(6.7)の分母の計算を行ない、式(6.7)を最大化するものを選択すればよい。

#### 位置の候補数と $SNR_{seg}$ 性能との関係

組合せ探索に使用する位置の候補数と  $SNR_{seg}$  の関係を図6.3に示す。ここでは、EFRなどへの適用を考慮し、マルチパルスの個数  $M$  は5msサブフレーム当たり7

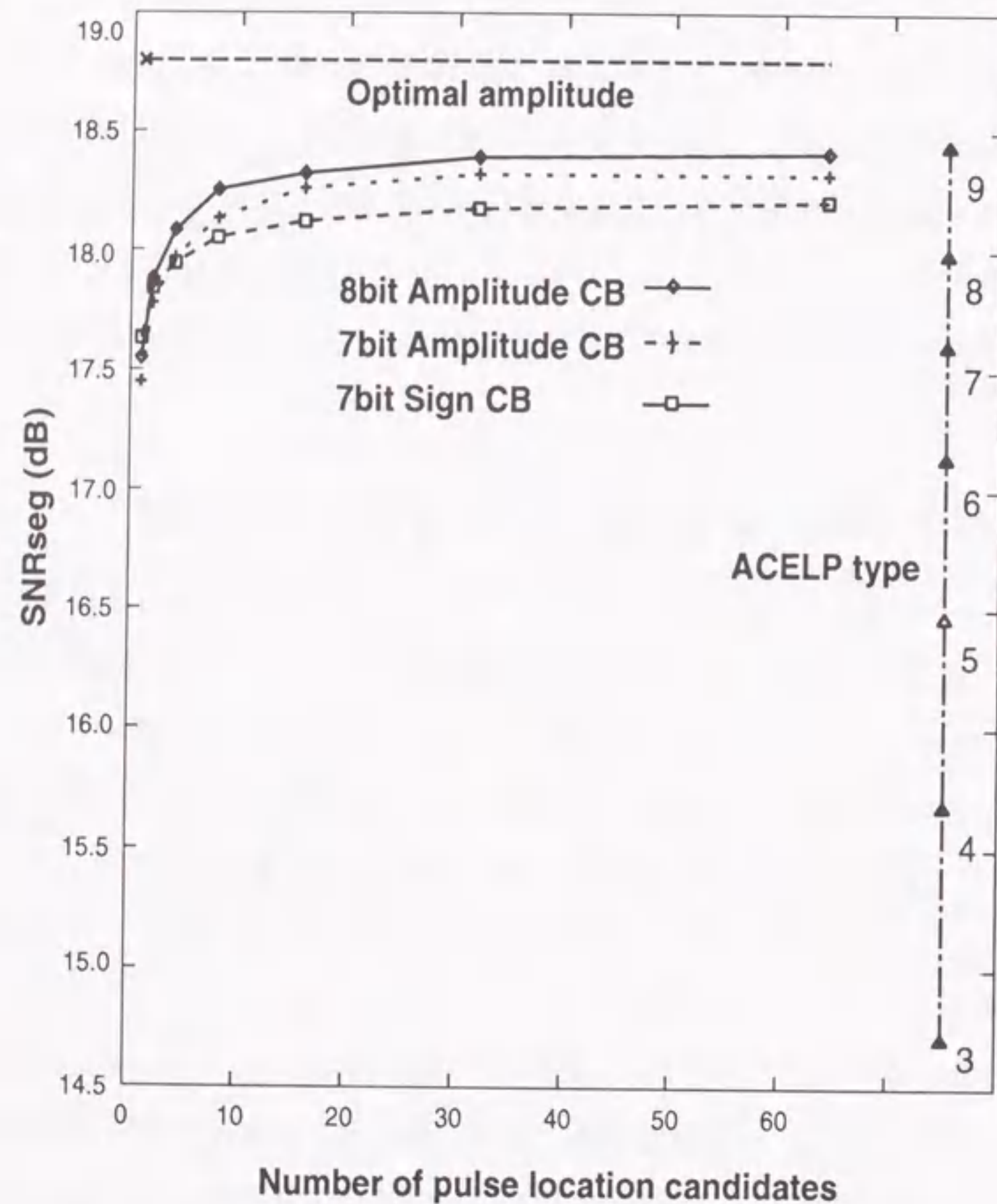


図 6.3: 位置の候補数と  $SNR_{seg}$  の関係

とし、振幅コードブックには7次元7ビット及び、7次元8ビットを使用した。位置の探索には6.2.4節で述べる2分割探索を使用した。また、振幅コードブックの代わりに7パルス分7ビットの極性コードブックを用いて位置の候補数を変えた場合の性能も示した。これらの手法の限界値として、振幅を量子化せずに各パルス毎に最適振幅を与えたときの性能も示す。レファレンスとして、パルスの振幅を極性のみで表す ACELP タイプの性能も示す。ここで、ACELP タイプとは、MP-CELP と同様に2分割探索を用いあらかじめ抽出した極性を用いて位置を探索するが、位置の候補数を1に絞る点、位置の探索に用いた極性を伝送するため振幅CBや極性CBは用いない点がMP-CELPと異なる。ACELP タイプのパルス数は3~9とした。なお、いずれの方式も、LPC分析の次数は6.2.6節の検討から14とし量子化を施した。ゲインは量子化していない。コードブック学習外の音声(約160秒)を使用し  $SNR_{seg}$  を測定した。

図 6.3 から次の諸点が見られる。

- (1) 組合せ探索により、位置の候補数の増加に伴い  $SNR_{seg}$  性能が向上し、候補数が16以上で性能が飽和傾向になる。7パルスの ACELP タイプと比較した場合、候補数16での  $SNR_{seg}$  の改善量は、振幅コードブックが8ビット、7ビットでそれぞれ、0.72 dB、0.66 dBである。また、最適振幅による限界値に比べ、それぞれ0.48 dB、0.54 dBの低下である。従って、今後の実験では、振幅コードブックを8ビット、候補数を16とする。
- (2) 位置の候補数が4、64の性能は、それぞれ、ACELP タイプの8パルス、9パルスに相当する。これらの性能改善量をビットレートに換算すると、ビットレートをそれぞれ、0.4 kb/s、1.0 kb/s 増加したことに相当する。
- (3) 候補数が1では、7パルスの ACELP タイプに比べ  $SNR_{seg}$  がやや下回っているが、これは、(i) 極性を用いて位置を探索し、その後で振幅CBにより振幅を量子化しているため、候補数が1のときは振幅CBに対して最適な位置が求まっているわけではないこと、(ii) 振幅CB自体の性能が十分でないこと、などが考えられる。後者に関しては、コードブック設計法の改善が今後の課題である。
- (4) 振幅コードブックのかわりに7ビット極性コードブックを用いても、組合せ探索

表 6.1: 位置の候補数と振幅コードブックの探索演算量

位置の候補数	演算量 (MOPS)
1	0.056
2	0.112
4	0.224
8	0.448
16	0.896
32	1.792
64	3.584

時の位置の候補数の増加に伴い性能が向上するが、候補数が8以上では、振幅コードブックを用いる方が性能がやや高い。

#### 位置の候補数と振幅コードブック探索演算量の関係

位置の候補数とパルス振幅コードブック探索に要する演算量 (積和回数を MOPS で表示) の関係を表 6.1 に示す。候補数を 16 としても、約 0.9 MOPS の演算量ですむ。

#### 6.2.4 分割探索によるマルチパルス位置の高速探索

##### 位置の探索における問題点

第2～4章で検討したマルチパルス符号化では、フレームなどの処理単位全体にわたりパルスの位置を探索していた。最適なパルスの位置を位置の総組合せにより探索すると、組合せ回数は、処理単位のサンプル数を  $N$  とすると、 $\frac{N!}{M!(N-M)!}$  のオーダーとなり、 $N$  の値とパルス数  $M$  の増加につれ膨大となる。これを防ぐため、第2章では、位置を1パルスずつ決定する逐次パルス探索法を使用していた [7],[108],[116]。さらに、位置を表すために比較的多くのビット数が必要であった。

表 6.2: 5 パルス / 5 ms のときの各パルスのとりうる位置の例

パルス番号	位置
1	0, 5, 10, 15, 20, 25, 30, 35
2	1, 6, 11, 16, 21, 26, 31, 36
3	2, 7, 12, 17, 22, 27, 32, 37
4	3, 8, 13, 18, 23, 28, 33, 38
5	4, 9, 14, 19, 24, 29, 34, 39

##### とりうる位置の限定

本章は、高音質化を狙うために、位置を総組合せで探索する。さらに前述の問題点を解決し、位置の探索に要する演算量及び、位置を表すビット数を低減化するために、ACELP や CS-ACELP と同様の代数構造 [147],[148],[71] を導入し、各パルスのとりうる位置を限定する。例えば、5 ms ( $N = 40$  サンプル) サブフレームに 5 本のマルチパルスをたてる ( $M = 5$ ) 場合を考えると、各パルスのとりうる位置の例は表 6.2 のようになる。このような限定により、各パルスのとりうる位置は  $\frac{N}{M}$  種類となるので、組合せ回数は、 $\left(\frac{N}{M}\right)^M$  のオーダーに低減される。しかし、それでも、ビットレート増加に伴いパルスの個数  $M$  が増えると、探索演算量は増大する。

これらを解決するために、高速探索法として、しきい値処理の不要な分割探索法を提案し、性能と演算量の関係を調べる。提案法によれば、 $M$  個のパルスを  $Q$  種のグループに分割し、各グループに属するパルスのとりうる位置について総組合せによる最適探索を行なう [102]。位置探索時の演算量削減化のために、パルスの振幅は固定し、CS-ACELP [148] と同様の手法であらかじめ抽出した極性を用いた。つまり、位置探索のときは、式 (6.8)、式 (6.9) において振幅値  $g_{kj}$  のかわりに極性  $sign[\psi(m_k)]$  を与え、式 (6.7) を最大化する位置を探索した。

一例として、分割数  $Q$  が 2、各グループでのパルス数が 2、各パルスのとりうる位置の種類が 3 の場合を図 6.4 に示す。

第  $l$  グループに属するパルスの個数を  $M_l$  とすると、位置探索の組合せ回数  $R$  は、

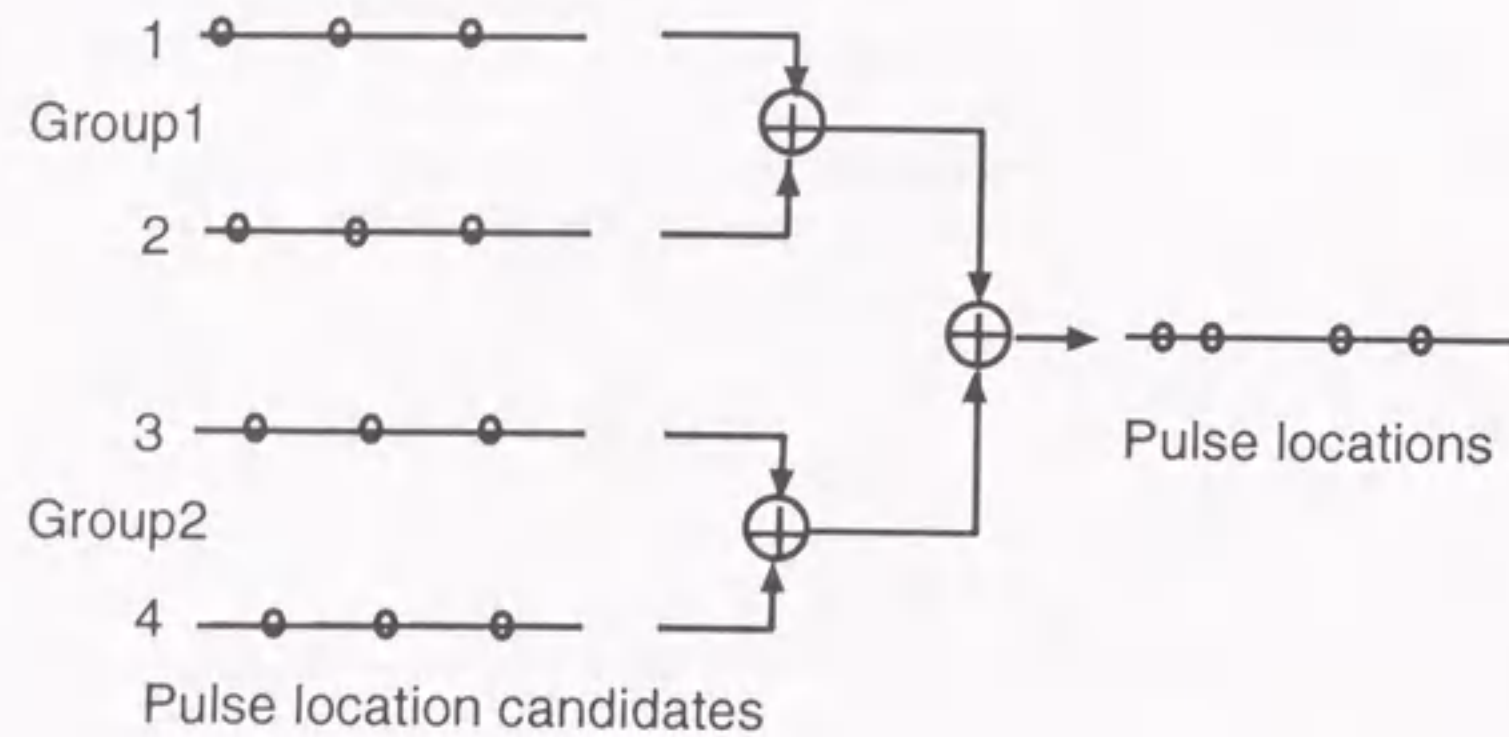


図 6.4: パルス位置の分割探索 (Q=2)

$$R = \sum_{i=1}^Q \left(\frac{N}{M}\right)^{M_i} \quad (6.13)$$

のオーダに削減される。

また、パルスの位置を表すのに必要なビット数はパルス当たり、 $\log_2\left(\frac{N}{M}\right)$  に低減される。

#### 分割数による性能と演算量の関係

分割数  $Q$  と  $SNR_{seg}$  性能及び演算量の関係を表 6.3 に示す。サブフレーム長は 5 ms、パルス数は 7 とした。演算量は式 (6.8)、式 (6.9) に従い計算した。但し、式 (6.10)、式 (6.11) の演算はここでは除外してある。比較のために、第 2 章 2.3.2 節で述べた逐次パルス探索法 (方式 2) についても  $SNR_{seg}$  と演算量を示す。演算量は 2 章の表 2.2 に従い計算した。

表から、分割数を増やしても劣化は少ない。2 分割では、演算量は総組合せに比べ、0.15% に削減され、5.47MOPS となるが、 $SNR_{seg}$  の劣化はほとんどない。4 分割では、演算量は 0.0059% の 0.22 MOPS となるが、 $SNR_{seg}$  は 0.11 dB の低下にすぎない。一方、2 章で述べた逐次パルス探索法 (方式 2) によれば、分割数 3 の約 2 倍

表 6.3: 位置の分割探索における分割数と  $SNR_{seg}$  性能, 演算量の関係

分割数	パルス数	$SNR_{seg}$ (dB)	演算量 (MOPS)
なし	7	18.33	3766
2	4+3	18.32	5.47
3	3+2+2	18.26	0.49
4	2+2+2+1	18.22	0.22
2 章の逐次探索 (方式 2)	7	17.35	1.17

の演算量を必要とするが、 $SNR_{seg}$  性能は分割数 3 に比べ約 0.9 dB 低下する。

なお、文献 [71] のしきい値を用いる位置探索法によれば、位置の探索演算量を 0.03% に低減すると、 $SNR_{seg}$  の低下は 0.7 dB である。ビットレート、信号の帯域などの条件が異なるため、単純に比較はできないが、提案法はしきい値処理が不要、しきい値処理に比べ演算量の変動がないなどの利点がある。

#### 6.2.5 ゲイン量子化

パルス振幅コードブックを探索してパルスの振幅を量子化した後に、ゲインコードブックを用いてゲインをベクトル量子化する。有声部では、ゲインコードブックは、適応コードブックのゲインとマルチパルスのゲインの 2 次元構成とする。

歪み評価には式 (6.14) を用いる。

$$D_r = \sum_{n=0}^{N-1} \left[ x_w(n) - \beta'_r b_w(n) - \gamma'_r \sum_{k=1}^M g'_{kj} h_w(n - m_k) \right]^2 \quad (6.14)$$

ここで、 $(\beta'_r, \gamma'_r)$  は、ゲインコードベクトルの  $r$  番目の要素である。

ゲイン量子化の性能を改善するために、パルス振幅コードベクトルと位置の組合せの複数候補に対して、ゲインコードベクトルを組合せ探索し、歪みを最小化する組合せを選択する。音源の候補数と性能の関係を表 6.4 に示す。ここで、有声部のゲインコードブックは 7 ビットを使用している。表から、候補数が 8 以上で性能がほぼ飽

表 6.4: ゲイン量子化時の候補数と  $SNR_{seg}$  性能

音源の候補数	$SNR_{seg}$ (dB)
1	17.27
2	17.38
4	17.58
8	17.68
16	17.73

表 6.5: LPC 分析の次数と性能の関係

次数	$SNR_{seg}$ (dB)	
	LSP 量子化なし	LSP 量子化有り
10	19.44	18.71
12	19.57	-
14	19.69	18.80
16	19.78	-

和することがわかる。なお、候補数が 16 のとき、ゲイン量子化なしに比べ、ゲイン量子化による  $SNR_{seg}$  の低下は 0.59dB である。

### 6.2.6 LPC 分析の次数

有声部において、音源をマルチパルスで良好に表すためには、音声波形の近接相関が充分除去され、LPC 予測残差がなるべくパルスのようになることが望ましい。そこで、LPC 分析次数を変化させて性能を調べた。LPC 分析の次数と  $SNR_{seg}$  性能との関係を表 6.5 に示す。ここで、量子化は全てなしとし、フレーム長は 10 ms、LPC 分析の窓長は 22.5 ms とした。

表 6.5 から、次数が 14 次以上では、改善が少なくなることがわかる。次に、10 次

と 14 次の場合に、LSP 係数に変換して量子化した結果も表 6.5 に示す。ここで、LSP の量子化には、6.3.1 節で述べる移動平均 (MA) 型予測 2 段分割ベクトル量子化を用い、量子化ビット数はいずれもフレーム当たり 22 ビットとした。量子化ありの場合にも 14 次の方がより高い  $SNR_{seg}$  が得られ、これは簡単な視聴により聴感的にも裏づけられたので、次数は 14 を用いる。

## 6.3 評価実験

提案した MP-CELP 方式の性能、音質を評価した。

### 6.3.1 実験条件

EFR や、PHS のハーフレート化などへの適用を考慮して、ビットレートを 11 kb/s とした。処理遅延の低減化を狙うために、フレーム長を 10 ms、サブフレーム長を 5 ms とした。ビット配分を表 6.6 にまとめる。フレーム長が短いため、音声の定常部等では LSP 係数は時間的に相関が高いと考えられる。そこで、LSP の量子化には、伝送路誤りに強い構造をもち過去のフレームから予測を行なう移動平均 (MA) 型予測 2 段分割ベクトル量子化 [92],[107],[101] を用いた。LSP のコードブックは、1 段目は 14 次元、2 段目は 3 分割し、それぞれ、5、5、4 次元とした。ゲインコードブックは、モード 1、2、3 のいずれも 2 次元、モード 0 では 1 次元とした。ゲインコードブックは 4 種のモード毎に切替えた。マルチパルスの位置探索における分割数は 2 とし、パルス数の分割はモード 1、2、3 共通とモード 0 でそれぞれ、4+3、5+5 とした。モード 1、2、3 とモード 0 で、サブフレーム毎に各パルスのとりうる位置を表 6.7 に示す。

コードブックの学習には約 90 分の音声信号を使用し、評価には学習外の約 160 秒の音声を使用した。

表 6.6: 11 kb/s MP-CELP のビット配分 (ビット /10 ms フレーム)

パラメータ	ビット数	
	モード 1,2,3	モード 0
LSP (14 次)	22	22
RMS	5	5
モード	2	2
適応コードブック	8+5	-
マルチパルス個数	7	10
位置	19 x 2	25 x 2
振幅	8 x 2	10 x 2
ゲイン	7 x 2	5 x 2
合計	110	109

表 6.7: 母音区間, 子音区間のパルスのとりうる位置

パルス番号		位置
モード 1,2,3	モード 0	
1, 6	1, 6	0, 5, 10, 15, 20, 25, 30, 35
2, 7	2, 7	1, 6, 11, 16, 21, 26, 31, 36
3	3, 8	2, 7, 12, 17, 22, 27, 32, 37
4	4, 9	3, 8, 13, 18, 23, 28, 33, 38
5	5, 10	4, 9, 14, 19, 24, 29, 34, 39

表 6.8: 11 kb/s MP-CELP と 11 kb/s M-LCELP の  $SNR_{seg}$  比較

条件	$SNR_{seg}$ (dB)	
	MP-CELP	M-LCELP
基本 (M-IRS 特性)	13.91	13.29
話者依存性 (フラット特性)	17.73	16.12

### 6.3.2 客観評価

11 kb/s MP-CELP の学習外音声に対する  $SNR_{seg}$  性能を表 6.8 に示す。比較のために、5 章の M-LCELP を 11 kb/s とし、演算量を MP-CELP と同程度におさえたもの (フレーム長 10 ms, サブフレーム長 2.5 ms) [137], [166], [167] についても示す。M-LCELP が MP-CELP と異なる点を次に示す。サブフレーム長が 2.5 ms、音源コードブックは母音区間が 6 ビット + 6 ビットの 2 段 8 パルス学習スパースコードブック、子音区間は 8 ビット + 7 ビットの 2 段乱数コードブックを使用し、いずれも予備選択を行なっている。LPC 分析の次数は 10 である。

電話器送話側の周波数特性を模擬した M-IRS (Modified-IRS) 特性 [13], [48] の音声を用いる基本条件と、周波数的にフラットな音声を用いる話者依存性条件に対する  $SNR_{seg}$  性能を表 6.8 に示す。

表から、11 kb/s MP-CELP は、同一ビットレートの従来 M-LCELP に比べ、 $SNR_{seg}$  が 0.6 dB から 1.6 dB 改善されることがわかる。

### 6.3.3 予備音質評価

#### 評価条件

誤りなしの条件に対して、MP-CELP と第 5 章の M-LCELP との音質比較を行なった。ここでは、LPC 分析の次数は両者ともに 10 としたため、ビットレートは M-LCELP の 11 kb/s に対し、MP-CELP は 10.7 kb/s である。

評価条件は、表 6.8 と同一の基本条件と話者依存性条件を用いた。言語は英語を用

表 6.9: 10.7 kb/s MP-CELP と 11 kb/s M-LCELP の音質比較

条件	MOS 値 (括弧内はオピニオン等価 Q 値)	
	MP-CELP	M-LCELP
基本 (M-IRS 特性)	3.57 (23.74)	3.35 (21.32)
話者依存性 (フラット特性)	4.29	4.04

い、前者は、男性 2 名、女性 2 名が発声した 8 文章を用いた。後者は前者とは異なる男性 4 名、女性 4 名及び子供 2 名が発声した 20 文章を使用した。被験者は英米人 24 人である。

#### 評価法

EFR 評価法 [144] に従い、基本条件は 5 段階 ACR (Absolute Category Rating) 尺度、話者依存性は 5 段階 DCR (Degradation Category Rating) 尺度を用いた [46]。DCR はレファレンス音として原音を先に聞かせ、次にテスト音として符号化音を聞かせた。なお、ACR、DCR の違いは第 2 章の表 2.3 にまとめた。

音質評価試験における試験条件の枠組、環境条件や被験者の違いに起因する評価の不安定さを除くために、MNRU (Modulated Noise Reference Unit) 標準系との相対評価による等価品質評価も含めた [47], [61]。この場合、符号化音声の音質は、符号化音声の MOS 値と等しい MOS 値を有する標準系の MNR 信号の S/N で表現される。これはオピニオン等価 Q 値と呼ばれ、dB で表示される。

#### 結果

MOS 評価結果を表 6.9 に示す。基本条件でのオピニオン等価 Q 値を括弧内に示す。これらから、MP-CELP は M-LCELP に比べビットレートが 0.3 kb/s 低いにも拘らず、MOS 値で 0.2 以上、Q 値で 2.4 dB 高い音質が得られる。

表 6.9 では、MP-CELP と M-LCELP の相違は音源構造とサブフレーム長のみであることから、今回の実験条件では、演算量を同程度とすると、MP-CELP のマルチ

パルスベクトル量子化音源は、M-LCELP のスパース学習音源よりも良好といえよう。

#### 6.3.4 音質評価

##### 評価した符号化方式

表 6.6 に示したビット配分に従う 11 kb/s の MP-CELP について、より詳細な音質評価を行なった。ここでは、欧州 ETSI (European Telecommunications Standards Institute) / GSM (Global System for Mobile Communication) フルレート移動電話 [95] と同一の伝送路を想定し、伝送路誤りに対処するために、MP-CELP に 11.8 kb/s の軟判定 FEC を付加し、総ビットレートを GSM フルレートと同一の 22.8 kb/s に設定した。軟判定 FEC には文献 [39] の方法を適用した。評価した方式は、MP-CELP 1 (音声 11 kb/s) の他に、MP-CELP 2 (音声 10.7 kb/s)、レファレンス符号化として、ITU-T 標準 G.728 (16 kb/s LD-CELP) [15]、ITU-T 標準 G.729 (8 kb/s CS-ACELP) [148]、GSM フルレート (FEC 込で 22.8 kb/s) [95] も評価に含めた。ここで、MP-CELP 1 と MP-CELP 2 の違いは、LPC 分析の次数と量子化ビット数であり、後者はそれぞれ 10 次、19 ビットである。

##### 評価条件

評価条件は、ETSI/GSM の定めた EFR 評価条件 [144] に従った。誤りなしとして、基本条件、話者依存性、2 タンデム接続を用い、背景雑音として、車走行背景雑音 (S/N = 10 dB)、バイオリン雑音 (S/N = 20 dB) を用いた。誤りありの条件として、4%、2 タンデム 4%、7%、13% の伝送路誤りを用いた。話者依存性は男性 4 名、女性 4 名及び子供 2 名が発声した 20 文章を、背景雑音は男性 2 名、女性 2 名が発声した 16 文章を、他の条件は男性 2 名、女性 2 名が発声した 8 文章を使用した。

##### 評価法

EFR 評価法 [144] に従い、背景雑音及び話者依存性条件のみ 5 段階 DCR [46] 尺度を使用し、13% 伝送路誤り条件のみは、5 段階 Listening Effort Scale ACR 尺度 (LES ACR) [46] を用いた。これは、評価文章の意味を理解する場合の聴取努力の程度を評

表 6.10: 5 段階聴取努力評価尺度 (LES ACR)

評点	絶対品質尺度の評定値
5	意味を理解するのに努力は全く必要ない
4	注意が必要であるがとりわけ努力はいらない
3	やや努力を必要とする
2	かなりの努力を必要とする
1	努力しても意味が理解できない

価するものである。評定値を表 6.10 に示す。他の条件では 5 段階 ACR 尺度を使用した [144]。また、6.3.3 節で述べた等価品質評価法も含めた。言語は英語を使用し、英米人 24 名が評価した。

#### 結果

評価結果を表 6.11 に示す。話者依存性、背景雑音以外の条件でのオピニオン等価 Q 値を括弧内に示す。

#### 6.3.5 考察

以上の結果から次の点が明らかとなった。

- (1)MP-CELP 1 (音声 11 kb/s) は、13% 誤りを除く全ての条件で、MP-CELP 2 (音声 10.7 kb/s) を上回った。MOS 値の改善は、特にバイオリン雑音で大きく、0.6 であった。これは、LPC 分析次数を 14 次としたのが貢献していると考えられる。13% 誤りでは、FEC の冗長ビット数の多い MP-CELP 2 の方が若干有利であった。
- (2)MP-CELP 1 を ITU-T 標準 G.728 (16 kb/s LD-CELP) と比較すると、背景雑音条件を除き、同等の音質が得られた。しかし、背景雑音条件では、G.728 の音質に到達しておらず、さらに改良が必要である。
- (3)MP-CELP 1 を GSM フルレートと比較すると、全ての条件で大幅な向上が見られ

表 6.11: 音質評価結果

条件	MOS 値 (括弧内はオピニオン等価 Q 値)				
	MP-CELP 1	MP-CELP 2	ITU-T G.728	ITU-T G.729	GSM
基本	3.58 (24.30)	3.54 (23.81)	3.40 (22.29)	3.35 (21.90)	2.85 (18.13)
話者依存性 †	4.46	4.26	4.48	4.50	-
2 タンデム接続	3.29 (21.36)	3.00 (19.12)	3.10 (19.87)	3.27 (21.18)	-
車走行雑音 †	4.13	3.98	4.48	3.77	4.06
バイオリン雑音 †	4.10	3.48	4.52	3.52	3.31
4% 誤り	3.63 (24.84)	3.54 (23.81)	-	-	3.10 (19.87)
2 タンデム 4% 誤り	3.52 (23.57)	3.08 (19.72)	-	-	2.17 (13.70)
7% 誤り	3.52 (23.57)	3.33 (21.72)	-	-	2.50 (15.85)
13% 誤り ‡	3.25 (12.79)	3.31 (13.34)	-	-	1.79 (0.81)

MP-CELP 1: 音声 11 kb/s + FEC 11.8 kb/s

MP-CELP 2: 音声 10.7 kb/s + FEC 12.1 kb/s

ITU-T 標準 G.728: 音声 16 kb/s

ITU-T 標準 G.729: 音声 8 kb/s

GSM (フルレート): 音声 13 kb/s + FEC 9.8 kb/s

無印の条件: ACR 尺度で評価

†の条件: DCR 尺度で評価

‡の条件: LES ACR 尺度で評価



表 6.12: MP-CELP エンコーダの演算量, メモリ量

パルス位置探索の分割数	積和回数 (MOPS)	データ ROM 容量 (k ワード)
2	13.71	4.4
3	8.72	4.4
4	8.45	4.4

た。特に、基本特性、バイオリン雑音が MOS 値でそれぞれ、0.7、0.8 の改善であった。さらに、7%、13% 誤りが 1.0 ~ 1.5 の改善であった。

#### 6.4 演算量

11 kb/s MP-CELP エンコーダの積和回数、データ ROM 容量を計算した。表 6.12 に示す。パルス位置探索の分割数が 2 の場合でも比較的少ない演算量で実現可能であることがわかる。また、分割数を 3 以上に増やすことで、性能をほとんど落さずに、エンコーダの演算量を大幅に低減化できる。また、デコーダは分割数によらず、1.5 MOPS 程度で実現できる。これらから、提案方式のエンコーダ・デコーダを DSP 1 チップで実現することは容易である。

#### 6.5 むすび

本章は、マルチパルス音源とパルス振幅のベクトル量子化を結合し、5 章の M-LCELP に比べ、高音質化及び、演算量、メモリ量の一層の低減化を狙った MP-CELP (Multi-Pulse based CELP) 方式を提案した。さらに、パルス振幅のベクトル量子化による性能を向上させるため、複数セットの位置候補と振幅コードベクトルを組合せ探索する方法を提案し、位置の候補数を 16 に増やすと、約 0.9 MOPS の演算量の増加で、 $SNR_{seg}$  が 0.77 dB 改善されることを明らかにした。また、振幅をあらかじめ抽出した極性のみとした ACELP タイプに比べ約 0.7 dB の改善を得た。パルス位置を総組合せで最適に求める場合の演算量を削減するために、パルスを複数のグループに分割

て位置を高速探索する分割探索法を提案し、演算量と性能の関係を示した。分割数を 4 まで増やし演算量を大幅に削減しても、 $SNR_{seg}$  の低下は 0.11 dB と少なかった。また、予測残差をよりパルス化させるために、LPC 分析の次数について検討を加えた。

これらの提案法を組み込んだ 11 kb/s MP-CELP の音質を実験により評価した。予備実験の結果、演算量を同程度に設定した M-LCELP に比べ良好な音質が得られた。欧州 GSM フルレート移動電話の伝送路を想定し、GSM フルレートとビットレートを合わせるために、11.8 kb/s 軟判定 FEC を組み込んだ。評価結果によれば、本方式は、背景雑音条件を除く基本特性、話者依存性、2 タンデム接続条件などで、32 kb/s ADPCM を越える音質を有する ITU-T 標準 G.728 (16 kb/s LD-CELP) と同等の音質を示した。また、既存の GSM フルレートとの比較では、全ての条件で大幅な音質の向上が見られた。特に、基本特性、バイオリン雑音条件がそれぞれ MOS 値で 0.7、0.8 の改善、7%、13% 誤りが 1.0 ~ 1.5 の改善であった。

本方式のエンコーダの演算量は、分割探索の分割数を 3 とすると、約 9 MOPS と少なく、コードブック格納に必要なデータ ROM 容量も 5 k ワード以下と M-LCELP に比べ大幅に低減されている。また、フレーム長が 10 ms のため、遅延も少なくできる。

これらから、6.1 節で述べた要求条件のうち、背景雑音条件を除く全ての条件を満たすことがわかった。提案方式は、演算量、メモリ量の削減が可能で、比較的短い遅延で、高音質を提供できる音声符号化として、移動体通信の分野などで有効になると考えられる。今後は、背景雑音条件に対しても ITU-T G.728 と同等の音質を目標に、さらに改良を施したい。

## 第7章

### 16 kb/s, 9.6 kb/s マルチパルス音声符号化方式の符号誤りに対する品質評価

#### 7.1 まえがき

移動体衛星通信や移動体通信の分野においては、チャンネル容量の拡大や、通信の秘匿の確保、マルチメディア化などの種々のサービスの付加のために、デジタル化が急速に進行している。デジタル音声伝送に関しては、無線周波数の有効利用、衛星や移動体のアンテナの小型化、送信電力の低減化のために、ビットレートの低い高能率音声符号化方式の導入が必要である。高能率音声符号化方式を移動体衛星通信に適用する場合、移動体衛星通信特有の伝送路符号誤りが品質に及ぼす影響を詳細に検討する必要がある。

本章は、2章で述べたマルチパルス音声符号化方式について、デジタル移動体衛星通信への適用性を調べるために、16 kb/s、9.6 kb/sにおいて、伝送路の符号誤りを計算機シミュレーションにより疑似的に発生させ、品質を客観的、主観的に評価した結果について述べる。また、伝送パラメータの誤りに対する感度に応じた効率的な誤り保護法を示し、これによる品質改善度についても述べる [129]。

## 7.2 マルチパルス符号化法

### 7.2.1 アルゴリズム

マルチパルス符号化法のアルゴリズムとしては、最も基本的な方式として第2章2.3.1.節で述べた効率的なマルチパルス探索法(方式1)を使用する。パルスの位置及び振幅の探索には、第2章の式(2.16)及び式(2.18)を用いた。

### 7.2.2 パルス振幅の簡易調整

性能を向上させるために、パルスの位置が全て求まった後に、簡易型の振幅調整を行なった[129], [171]。これは、第2章2.3.1節の式(2.16)、式(2.18)を用いて全てのパルスの位置及び振幅を一旦求めた後に、各パルスの位置について、式(2.16)の右辺の分子の値 $R(m_k)$ を調べ、この値が大きい位置から順に、式(2.18)を用いて振幅を求め直し、これらの処理を繰り返すものである。この振幅調整による性能の改善効果を図7.1に示す。ここでは、パルス振幅の量子化ビット数は4ビットまたは5ビットとし、パルスの個数は24としている。図から、振幅調整の回数を30回とすると、振幅量子化ビット数が5ビット、4ビットについてそれぞれ、約0.8 dB、約0.5 dBの $SNR_{seg}$ 改善が得られている。なお、簡易調整による演算量の増加はわずかである。

### 7.2.3 分析条件とビット割当

16 kb/s、9.6 kb/sのマルチパルス符号化の分析条件、ビット割当を表7.1に示す。

パルスの振幅は、分析フレーム内で絶対値が最大の振幅を正規化係数として、表に示したビット数の $\mu$ -lawPCMで符号化した後に、各パルスの振幅を正規化し表に示したビット数で符号化した[116]。パルスの位置は、隣接したパルスの位置の間隔を表に示した位置符号化ビット数で符号化した。パルスの間隔がビット数で示される符号化可能な距離を越えるときは、量子化最大距離の位置に、量子化最小レベルの振幅を有するダミーパルスをたて、その代わり絶対値振幅が最小のパルスを削除する処理を繰り返した。この処理により、位置を4~5ビットで符号化することによる $SNR_{seg}$ の低下を0.5 dB以内に抑えることができた。

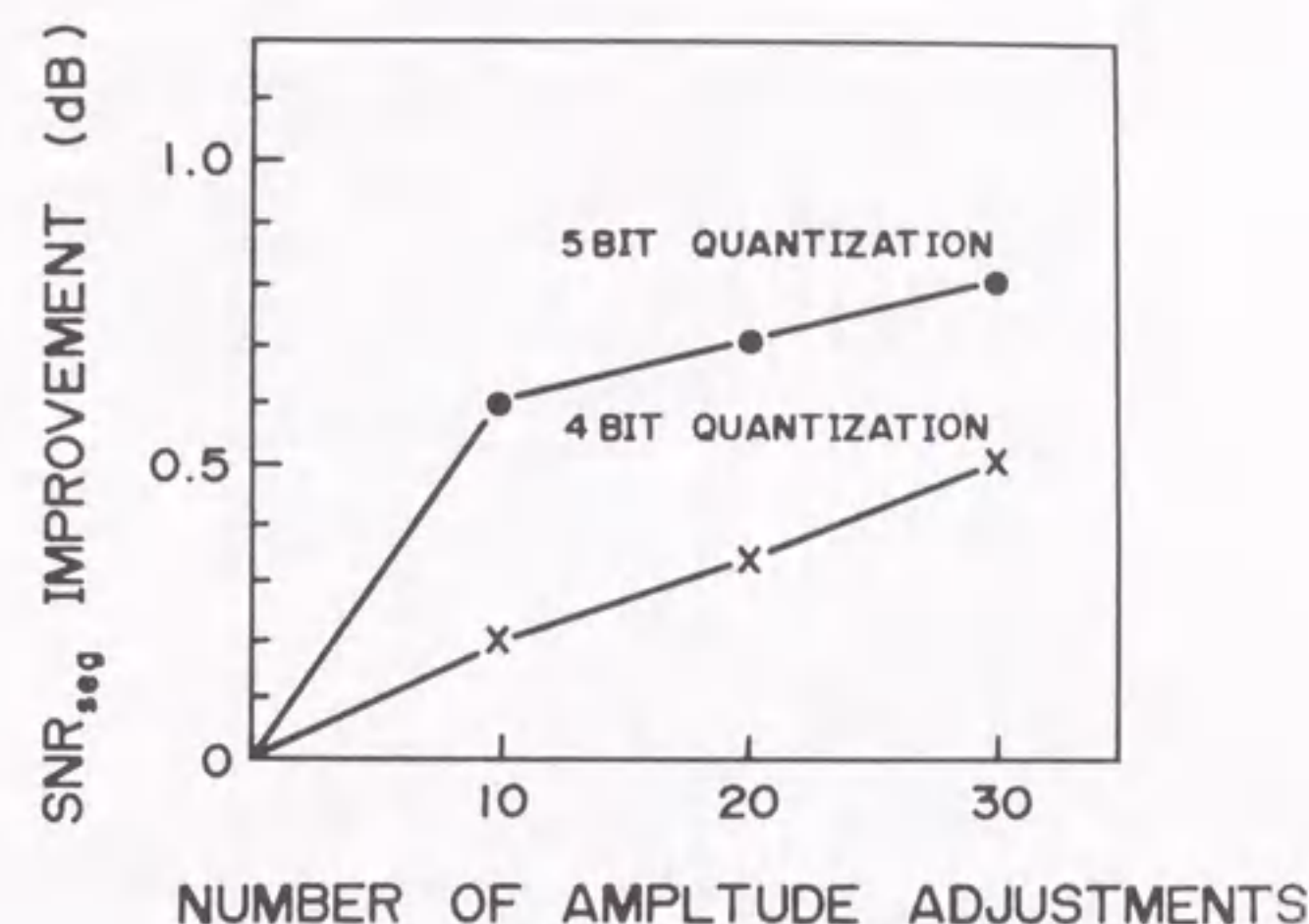


図7.1: パルス振幅の簡易調整による $SNR_{seg}$ の改善

## 7.3 実験

デジタル移動体衛星通信への適用性を調べるために、16 kb/s、9.6 kb/sのマルチパルス符号化法に伝送符号誤りを発生させ品質を客観的、主観的に評価した。

### 7.3.1 伝送符号誤り

伝送符号誤りとしては、ランダム誤りと、デジタル移動体衛星通信で誤り訂正符号化を導入したときに復号側で誤り訂正復号後に残存するバースト誤りの2種を仮定した。デジタル移動体衛星通信システムでは、通常、音声信号と制御信号などを含めたシステム全体の伝送符号に対して誤り訂正符号化(FEC: Forward Error Correction)を行なうため、音声符号化のみの部分には専用に誤り訂正を付加していない。

表 7.1: 16 kb/s, 9.6 kb/s マルチパルス符号化の分析条件とビット配分

	16 kb/s	9.6 kb/s
フレーム長	20 ms	20 ms
LPC 分析次数	12	12
ビット配分	7,6,5,4,3,3, 3,3,3,3,2,2	7,6,5,4,3,3, 3,3,3,3,2,2
パルス数	29	15
最大値	7 ビット	6 ビット
パルス振幅	5 ビット	4 ビット
パルス位置	4 ビット	5 ビット
パルス振幅簡易調整	30 回	30 回
フレーム同期ビット	8 ビット	7 ビット

このとき、復号側で FEC 復号後の誤りパターンがバースト的になることが知られている [182]。そこで、FEC 復号後に残存するバースト誤りを検討の対象にした。本実験では、システム全体の FEC には、符号化率 1/2 の畳み込み符号 / ヴィタビ復号 [182] を用いることにし、復号後のバースト誤りパターンを文献 [182] を参照して計算機により近似的に発生した。文献 [182] では、バースト誤りパターンは畳み込み符号 / ヴィタビ復号回路、QPSK (Quadrature Phase Shift Keying) モデムを用い伝送路にランダム雑音を発生し実測したものである。誤り率は  $10^{-4}$ 、 $10^{-3}$  とした。平均バースト長は文献 [182] に従い、誤り率  $10^{-4}$  で 7、 $10^{-3}$  で 9 とした。

### 7.3.2 客観、主観評価尺度

客観評価尺度としては、第 2 章 2.4.1 節で示した時間波形の歪を測るセグメンタル S/N ( $SNR_{seg}$ ) [179] とスペクトル上での歪を測るケプストラム距離 (CD) [179] を用いた。音声資料には、男女各 3 名により発声された日本語短文 10 種 (合計で約 32

sec) を用いた。

主観評価には第 2 章 2.4.1 節に示した 5 段階の ACR 尺度を用いた。符号化方式は、マルチパルス符号化のほかに、基準方式として 64 kb/s (PCM 8) から 16 kb/s (PCM 2) までの 7 種の  $\mu$ -law PCM も評価した。評価には客観評価に用いた短文のうちの 2 種 (男性 1 種、女性 1 種) を用いた。被験者は女性 6 名である。

### 7.3.3 ランダム誤りに対する性能評価

#### 客観評価

16 kb/s、9.6 kb/s マルチパルス符号化方式のランダム誤りに対する  $SNR_{seg}$  特性を図 7.2 に示す。

いずれのビットレートでも  $10^{-3}$  までは劣化は小さいが、 $10^{-2}$  になると特性は急激に劣化する。また、16 kb/s に比べ 9.6 kb/s の方が  $10^{-2}$  での劣化が大きい。これは 9.6 kb/s の方が誤りに敏感な PARCOR 係数の全体のビット数に対する割合が高いためと考えられる。

#### 伝送パラメータの誤りに対する特性

16 kb/s マルチパルス符号化のランダム誤りに対する劣化度を各伝送パラメータ毎に調べた結果を図 7.3 と図 7.4 に示す。各伝送パラメータ毎に所定の誤り率のランダム誤りを発生して合成した合成音声と原音声との違いをそれぞれ、 $SNR_{seg}$ 、ケプストラム距離 (CD) を用いて評価した。ここで、図 7.3 では  $SNR_{seg}$  の値の低いほど、図 7.4 では CD 値の大きいほど、誤りによる劣化が大きいことを示している。

図から、伝送パラメータにより符号誤りに対する感度が大きく異なり、1~5 次の PARCOR 係数やパルスの位置の劣化度が大きいことがわかる。一方、パルスの振幅と 6~10 次の PARCOR 係数の劣化度は小さい。これは LPC 分析を用いた APC などの符号化にも同様の傾向がある。伝送パラメータの感度に応じて強度の異なる誤り訂正 [81] を施すことにより、誤り訂正に必要な冗長ビット数を最小限に抑えた効率的

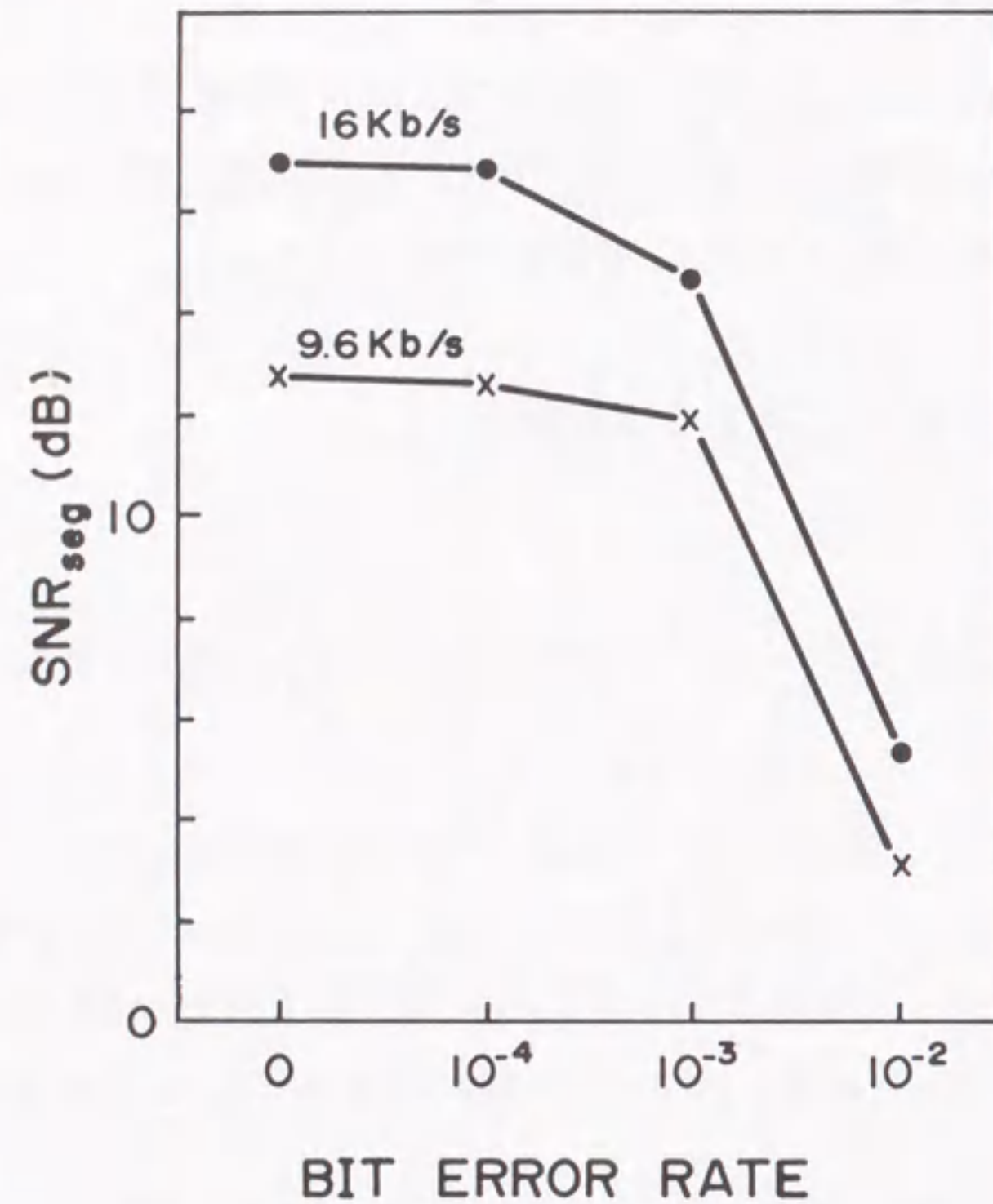


図 7.2: 16 kb/s, 9.6 kb/s マルチパルス符号化のランダム誤りに対する  $SNR_{seg}$  特性

な誤り訂正を行なうことができる。

#### 効率的な誤り訂正の効果

16 kb/s マルチパルス符号化において、図 7.3、図 7.4 の結果に基づき伝送パラメータの誤りに対する感度を考慮して、PARCOR 係数、パルスの位置、振幅、最大値に対して訂正能力の異なる 5 種類の誤り訂正符号を付加した。誤り訂正符号にはリードソロモン (RS: Reed Solomon) 符号 [40] を使用し、図 7.3 において  $10^{-2}$  の誤りで CD が 2.5 を越えるパラメータを劣化の大きなものから順に重点的に保護した。具体的には、1、2 次の PARCOR 係数の上位 2 ビットは多数決判別を行なう MLD (Majority

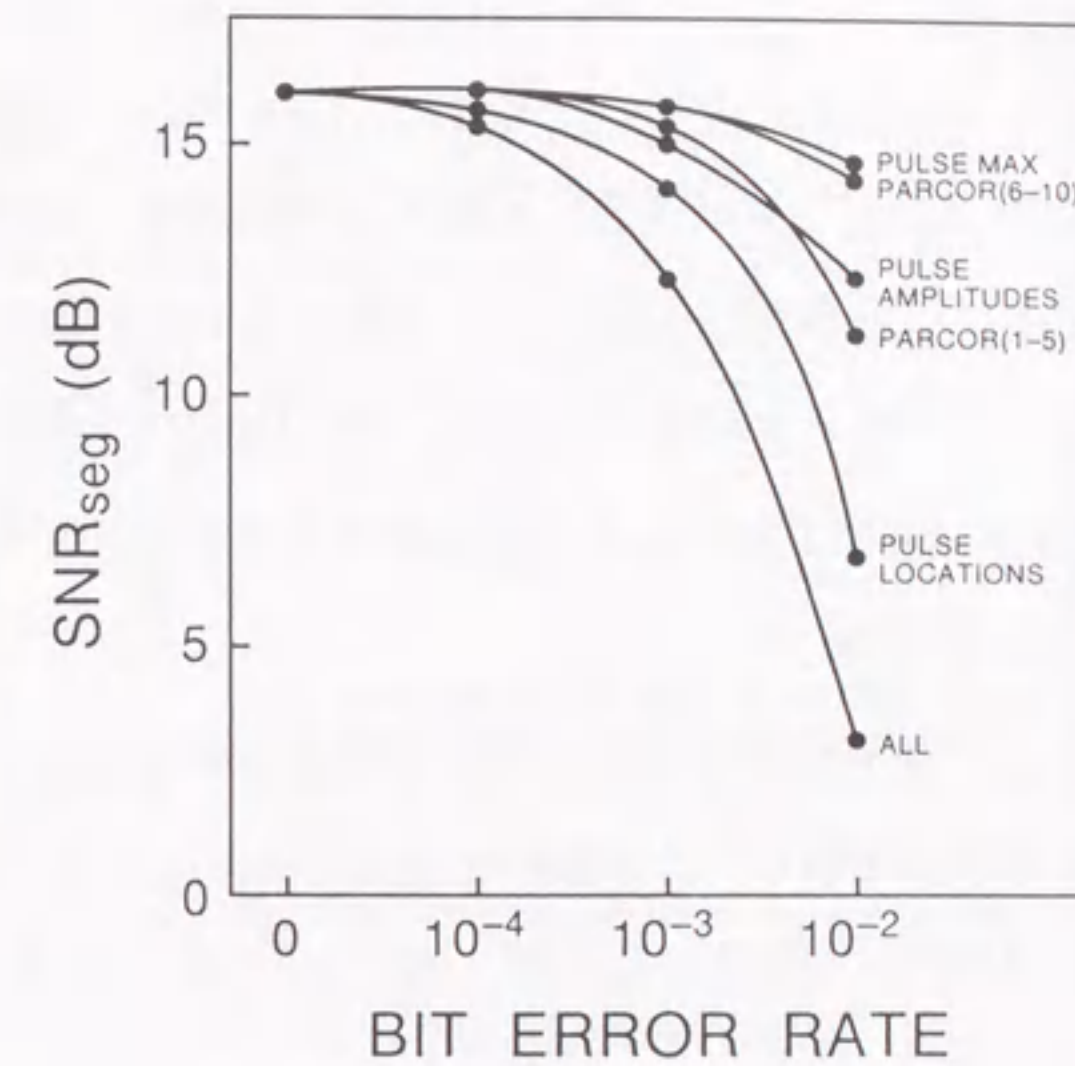


図 7.3: 16 kb/s マルチパルス符号化において各伝送パラメータにランダム誤りを発生させたときの再生音声の  $SNR_{seg}$

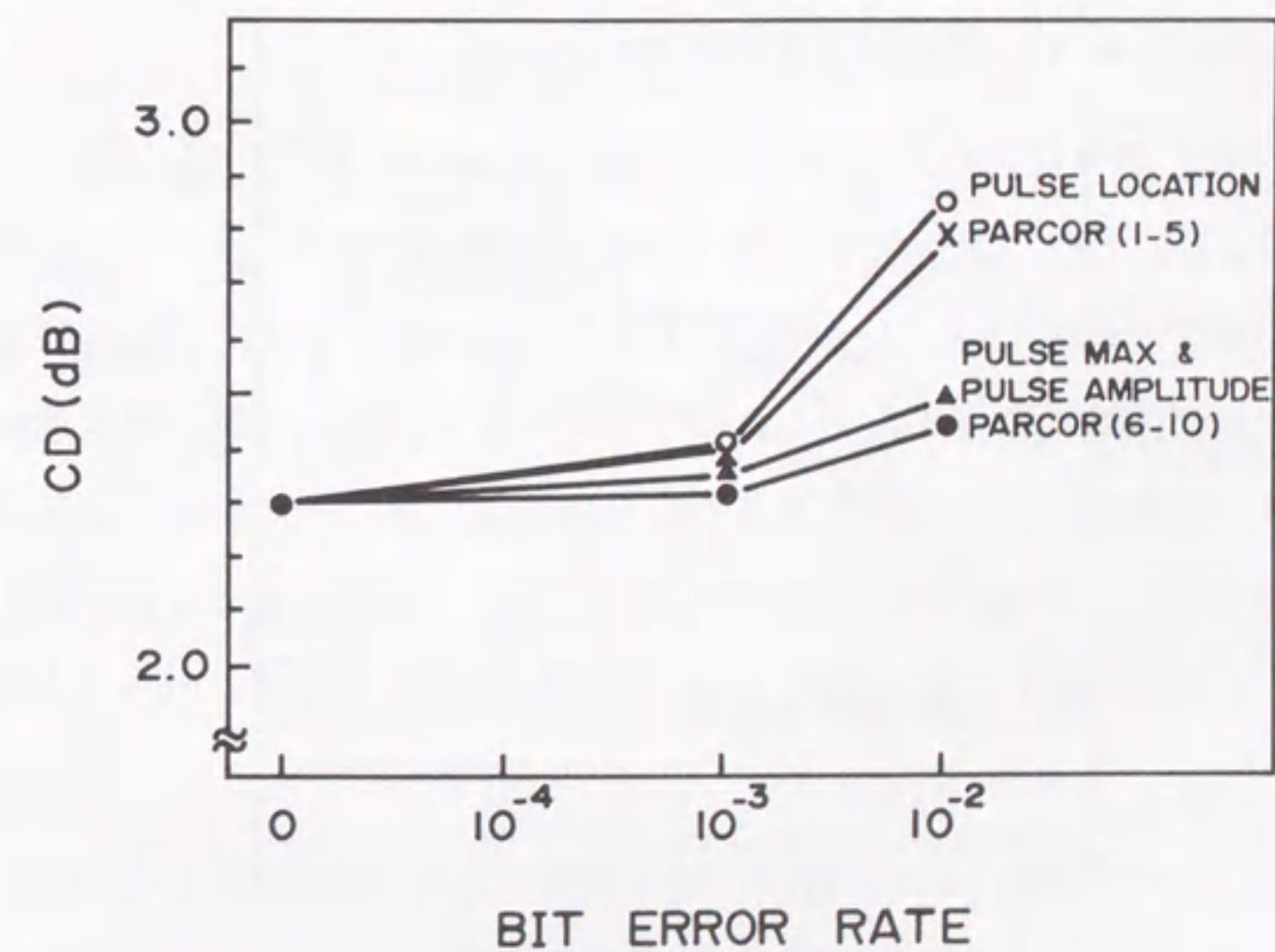


図 7.4: 16 kb/s マルチパルス符号化において各伝送パラメータにランダム誤りを発生させたときの再生音声のケプストラム距離  $CD$

Logic Detection) 符号の MLD(3,1) を用い、他の PARCOR 係数の上位ビットは RS(10,6) と、RS(13,9) の一部を用いた。また、パルスの振幅、位置の上位ビットは前記 RS(13,9) を共通に用いると共に、これ以外に RS(8,6), RS(15,11) を用いて保護した。この結果、誤り訂正符号化に必要な冗長ビットは合計 3.2 kb/s となるため、音声を 12.8 kb/s に減らし、合計で 16 kb/s にした [170]。音声のビットレートを 3.2 kb/s 分低減するために、パルス数、振幅の符号化ビット数をそれぞれ、25 パルス、4 ビットに変更した。

$SNR_{seg}$  特性を図 7.5 に示す。図には 3.2 kb/s 誤り訂正を付加する前の 16 kb/s マルチパルスの特性ものせてある。

図から、誤り率が  $10^{-3}$  より高い場合に誤り訂正の効果が現れることがわかる。誤り訂正の付加により、誤り率が  $10^{-4}$  より低いときは、誤り訂正を用いない場合に比べ  $SNR_{seg}$  が低下するものの、 $10^{-2}$  では大幅に向上し、10.2 dB の  $SNR_{seg}$  が得られる。

#### 主観評価

16 kb/s、9.6 kb/s の主観評価結果を図 7.6 に示す。

16 kb/s では、誤り訂正なしの方式では、誤りが 0 のときに PCM 6.5 ビット相当の音質を得ている。また、誤りが  $10^{-4}$  では劣化がわずかであり、 $10^{-3}$  では PCM 5.5 ビット相当の音質を得ている。しかし、 $10^{-2}$  では劣化が大きく、了解性はあるものの、耳障りな音が頻繁に発生する。これは PARCOR 係数やパルスの最大値が頻繁に誤り、スペクトルやゲインが変化することに起因している。一方、誤り訂正を施した 16 kb/s では、誤りが 0 のときの音質は PCM 6.2 ビット相当であり、誤り訂正なしの 16 kb/s よりも音質は低下しているが、 $10^{-2}$  では、PCM 5 ビット相当と大幅に向上している。

一方、9.6 kb/s の誤り訂正なしでは、誤りが 0 のときは PCM 6 ビット相当、 $10^{-3}$  では PCM 5 ビット相当の音質を得ているが、 $10^{-2}$  では急速に劣化し、PCM 3 ビット相当の音質になっている。

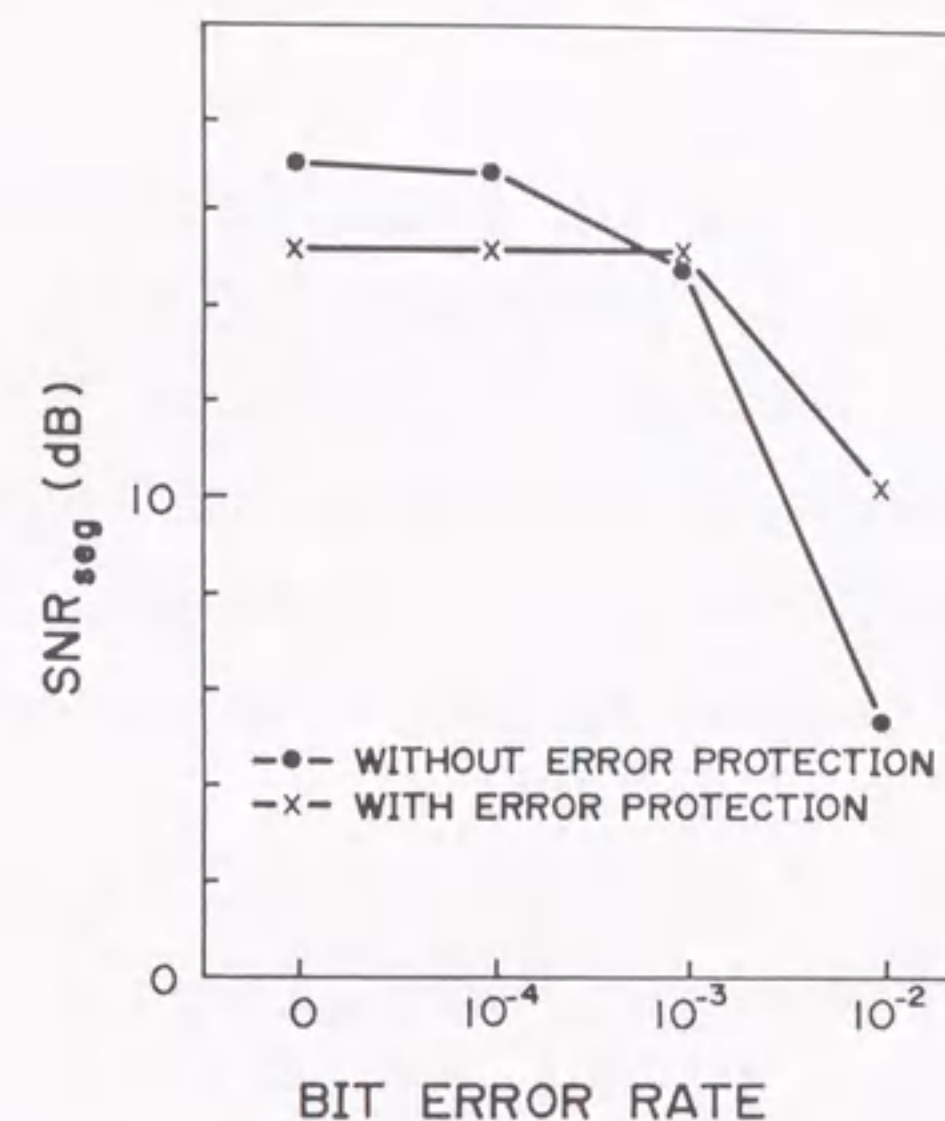


図 7.5: 3.2 kb/s の誤り訂正を付加した 16 kb/s マルチパルス符号化のランダム誤りに対する  $SNR_{seg}$  特性

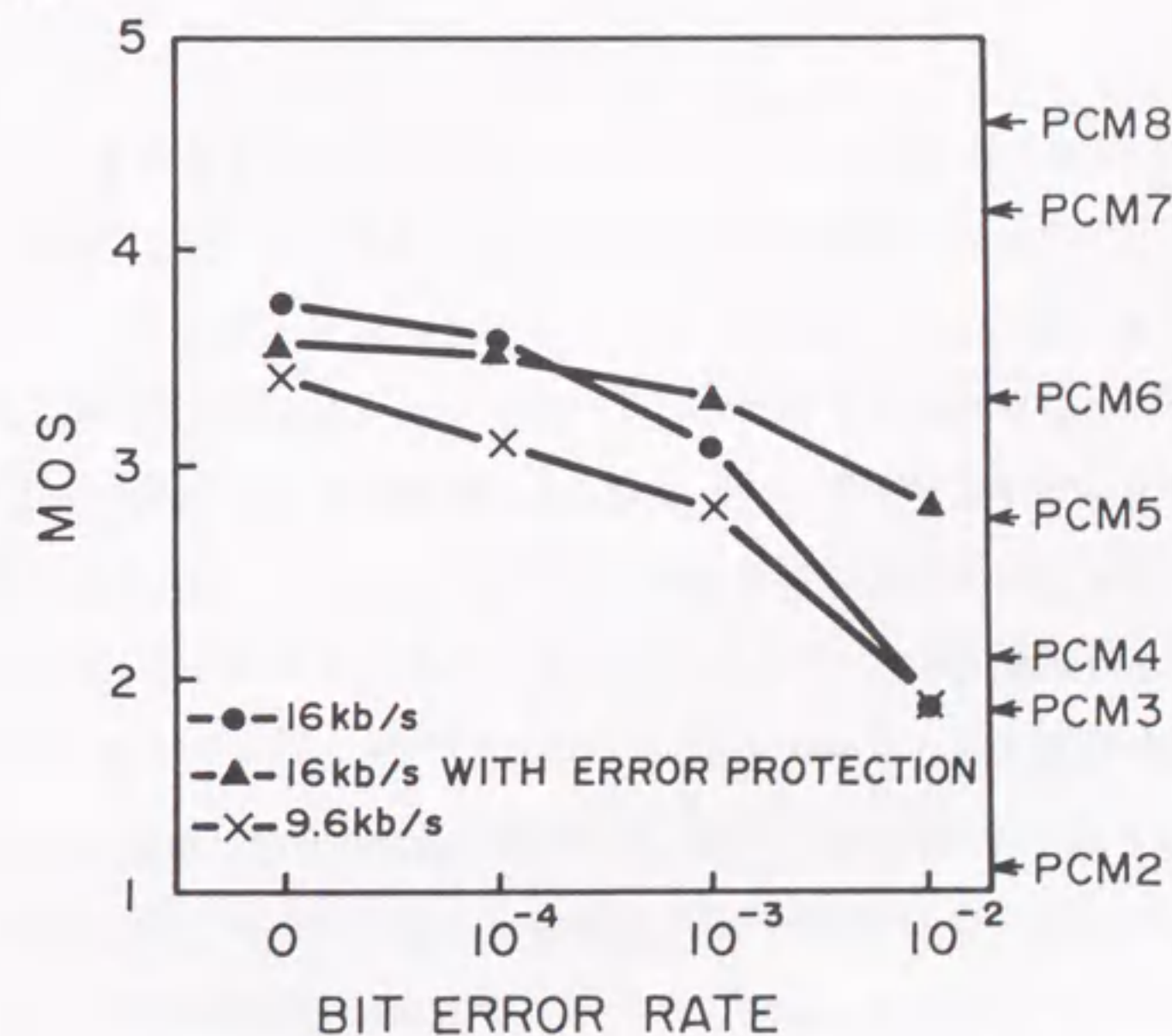


図 7.6: 16 kb/s、9.6 kb/s マルチパルス符号化のランダム誤りに対する音質

## ランダム誤りに対する考察

以上から、次の点が明らかになった。

- (1) ランダム誤りに対しては、16 kb/s、9.6 kb/sともに、誤り訂正なしでも誤りが $10^{-3}$ までは音質の劣化は少ないことが明らかになった。しかし、 $10^{-2}$ では劣化が大きく、了解性はあるもののかなり耳障りになった。
- (2) 16 kb/sにおいて、伝送パラメータの符号誤りに対する感度を考慮した誤り保護を付加することにより、3.2 kb/sの冗長ビットの付加で、誤りが0のときは若干音質が低下するものの、 $10^{-2}$ での音質が大幅に改善され、PCM 5ビット相当の音質が得られた。
- (3)(2)に示した誤り保護法は、伝送パラメータを全て同一の強度で保護する誤り保護法に比べ、伝送路の誤りが $10^{-2}$ 程度までで、全体のビットレートに制約がある場合には有効な方法であると考えられる。

## 7.3.4 バースト誤りに対する性能評価

## 客観評価

デジタル移動体衛星通信システムで用いられる誤り訂正符号をマルチパルス符号化に導入したときに、復号側で誤り訂正復号後に残存するバースト誤りが再生音声の品質に及ぼす影響について調べた。バースト誤りに対処するために、PARCOR係数の各ビットはフレーム内でインタリーブした。また、これ以外に、図7.3、図7.4に示した符号誤りに対する感度を考慮した簡単な誤り保護を更に付加した方式についても検討した。デジタル移動体衛星通信システムでは、システム全体の伝送符号に対しては既にFECが導入されているので、極めて少量の冗長ビットの追加ですむように、PARCOR係数の1～5次の上位ビットの合計9ビットとパルスの最大値の上位3ビットの合計12ビットのみをBCH (Bose-Chaudhuri-Hocquenghem) 符号 [40] を用い、BCH(21,12)による2ビット誤り訂正符号を用いて保護した。この方法では、誤り保護のために、冗長ビット数がフレーム当たり9ビット増えるので、音声に割り当てるビット数をフレーム当たり9ビット低減する必要がある。ここでは、フレーム

当たりのパルス数を表7.1に示した誤り保護なしの方式に比べ1パルス分低減することにより対処した。

誤り訂正復号後の誤り率が $10^{-4}$ 、 $10^{-3}$ のときの $SNR_{seg}$ 性能を図7.7に示す。ビットレートは16 kb/s、9.6 kb/sとした。図において、MPC161、MPC961はそれぞれ、システム全体のFECに上述のインタリーブのみを付加した16kb/s、9.6 kb/sである。MPC162、MPC962は、それぞれ、MPC161、MPC961に使用したFEC及びインタリーブに加え、前述のBCH(21,12)による誤り保護を追加した16 kb/s、9.6 kb/sをそれぞれ示す。

図から、16 kb/sではMPC161、MPC162ともに $10^{-4}$ の誤りでは誤り0と比べほとんど劣化はない。しかし、 $10^{-3}$ では特性は劣化し、MPC161では $SNR_{seg}$ が誤り0に比べ1.4 dB劣化し15.5 dBとなる。MPC162では $10^{-3}$ の誤りでMPC161よりも $SNR_{seg}$ が0.2 dB向上する。

一方、9.6 kb/sでは、誤りが $10^{-4}$ の場合劣化はほとんどないが、 $10^{-3}$ ではMPC961は約1.0 dB低下する。MPC962は $10^{-3}$ の誤りでMPC961に比べ $SNR_{seg}$ が0.2 dB向上する。

## 主観評価

主観評価結果を図7.8に示す。16 kb/sのMPC161では、誤りが $10^{-4}$ のときはPCM 6.5ビット相当の音質が得られている。しかし、 $10^{-3}$ では音質が急激に劣化しPCM 5ビット相当になる。これに対し、MPC162は $10^{-3}$ での $SNR_{seg}$ の改善はわずかであったものの、PCM 5.5ビット以上の音質を得ており、MPC161に対する音質改善度はPCM 0.5ビット相当以上である。

9.6 kb/sでは、誤りが $10^{-4}$ ではMPC961、MPC962ともにPCM 5.5ビット相当以上の音質が得られる。 $10^{-3}$ では、MPC962はMPC961に比べ音質が大きく改善され、PCM 5ビット相当の音質が得られる。MPC962による音質の改善度は16 kb/sのときよりも大きく、PCM 1ビット相当の改善度が得られている。

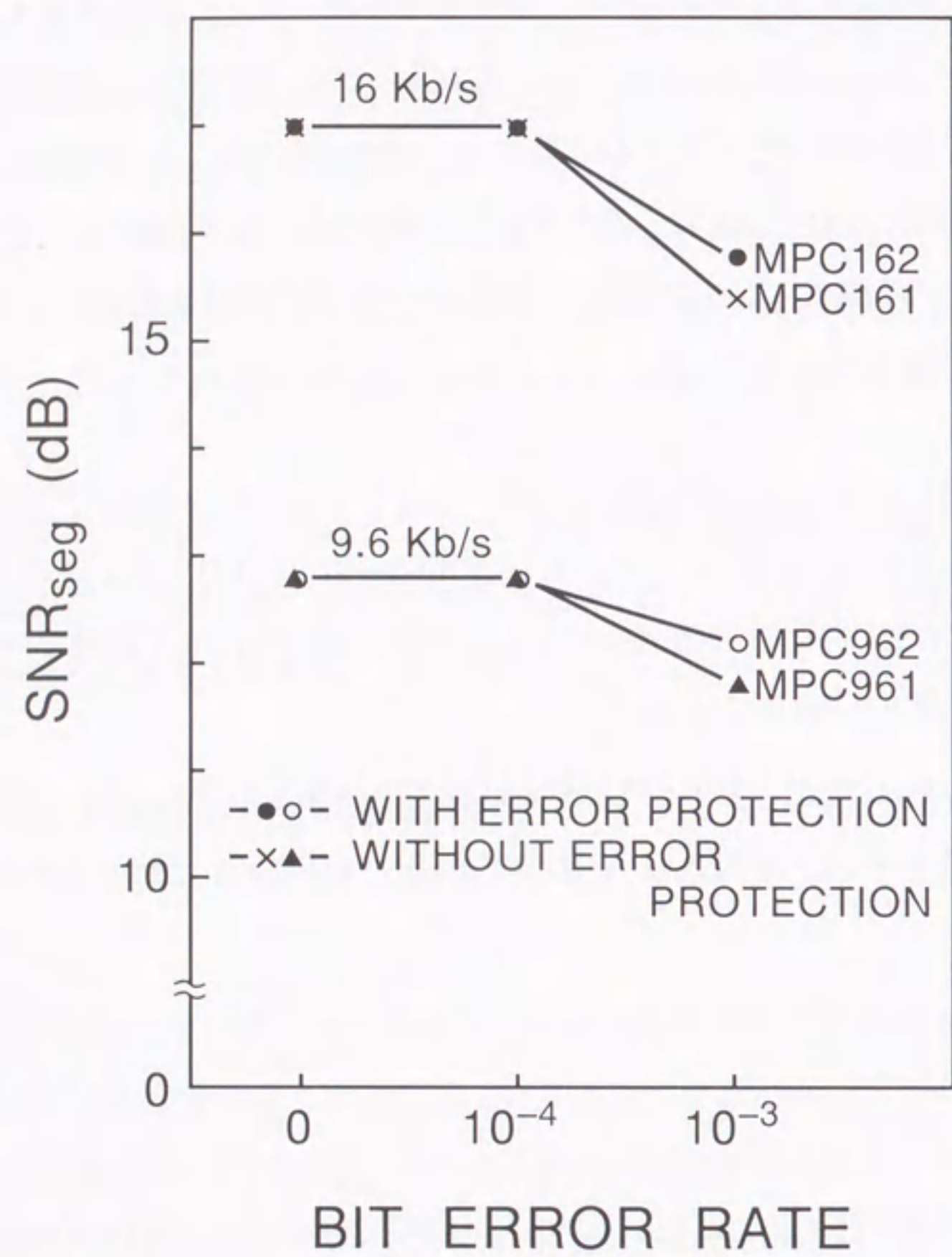
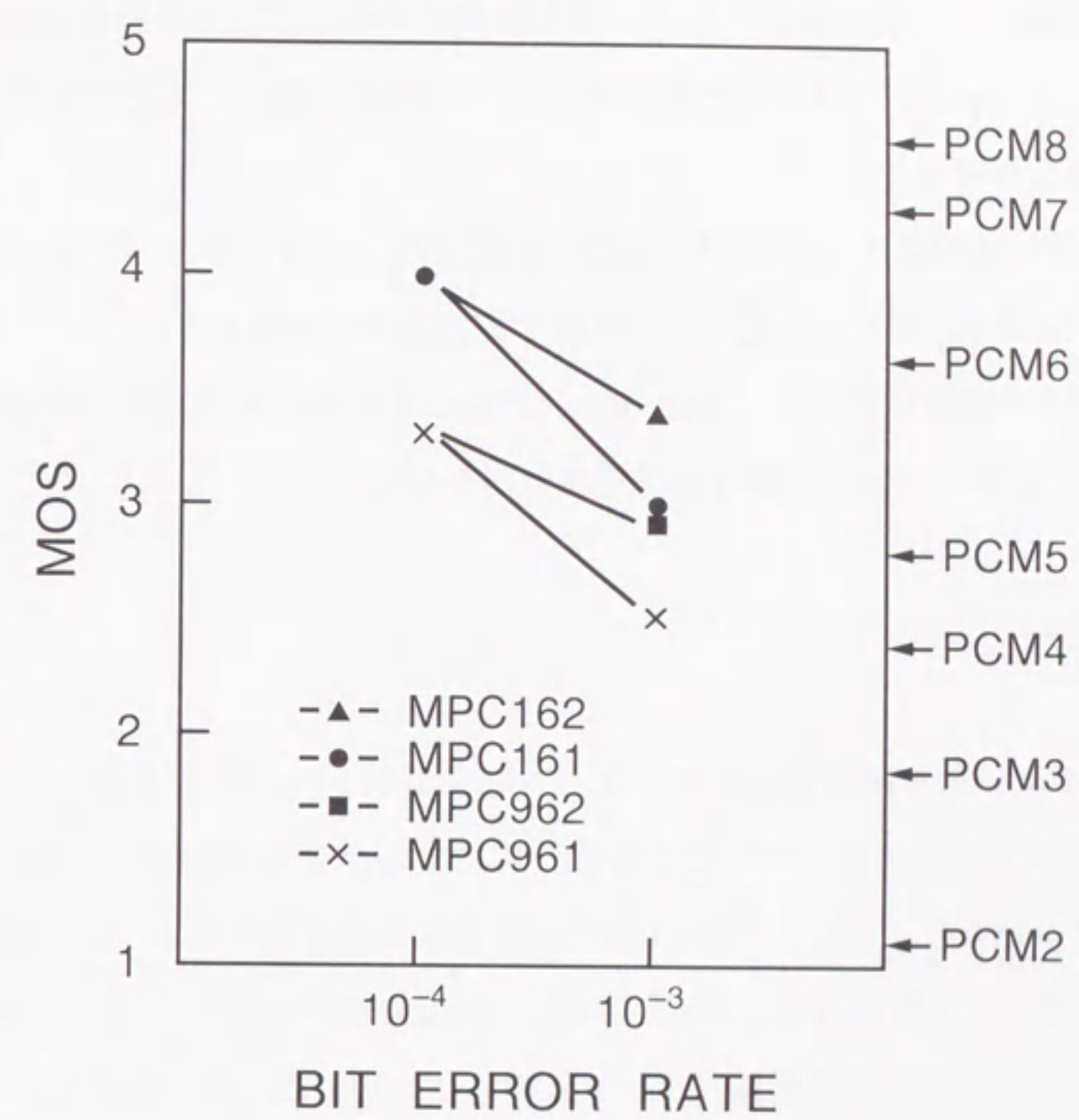
図 7.7: 16 kb/s, 9.6 kb/s マルチパルス符号化のバースト誤りに対する  $SNR_{seg}$  特性

図 7.8: 16 kb/s, 9.6 kb/s マルチパルス符号化のバースト誤りに対する音質



## バースト誤りに対する考察

以上から、次の点が明らかになった。

- (1) デジタル移動体衛星通信において、システム全体の伝送符号に付加される FEC を復号側で復号した後に残存するバースト誤りをマルチパルス符号化に適用した場合、復号後の誤り率が  $10^{-4}$  以下のときは再生音の音質劣化はほとんどないが、 $10^{-3}$  以上となると音質が劣化した。
- (2) 符号誤りに敏感な 1～5 次の PARCOR 係数の上位ビットと、パルスの最大値の上位ビットの合計 12 ビットを更に誤り保護する方法が有効であった。
- (3)(2) の誤り保護法によれば、450 b/s の冗長ビットの追加で、誤りが  $10^{-3}$  のときに PCM 0.5～1 ビット相当と良好な音質改善が得られた。

## 7.4 むすび

デジタル移動体衛星通信やデジタル移動通信への適用を考慮して、16 kb/s、9.6 kb/s マルチパルス符号化の伝送路誤りに対する品質を評価した。符号誤りとしては、ランダム誤りと、デジタル移動体衛星通信で通常用いられている畳み込み符号/ヴィタビ復号による誤り訂正符号 (FEC) を復号側で復号したさいに残存するバースト誤りを用いた。実験の結果、ランダム誤りに対しては、誤り訂正なしでも  $10^{-3}$  までは劣化が少ないが、 $10^{-2}$  にすると劣化が大きく、PCM 3 ビット相当の音質に低下した。これに対して、誤りに敏感なパラメータを感度に応じて効率的に保護するために、3.2 kb/s の冗長ビットを含めた 16 kb/s の符号化は、 $10^{-3}$  のときの音質が PCM 5 ビット相当と大幅に改善された。

一方、FEC の復号時に残存するバースト誤りに対しては、復号後の誤りが  $10^{-4}$  までは音質劣化はほとんどないが、 $10^{-3}$  のときに音質が劣化し、16 kb/s で PCM 5 ビット相当、9.6 kb/s で PCM 4 ビット相当に低下した。これに対し、符号誤りに敏感な PARCOR 係数の上位ビットとパルス振幅最大値の上位ビットをフレーム当たり合計で 12 ビットさらに保護することにより、450 b/s という極めて少ない冗長ビット数の追加で、誤り率が  $10^{-3}$  での音質が PCM 5.5～5 ビット相当に改善されることがわ

かった。

## 第8章

### 高能率音声符号化装置の実用化例

#### 8.1 まえがき

本論文で研究した高能率音声符号化方式を NEC にて製品化した例を紹介する。有線伝送などの分野において、企業内の通信ネットワークなどで、音声信号を圧縮・符号化し、他のマルチメディア信号と多重化・伝送するための音声符号化装置として、第2章及び第3章で述べたマルチパルス符号化方式やピッチ予測マルチパルス符号化方式を用いたもの、第5章で述べたベクトル量子化に基づく M-LCELP 方式を用いたものを示す。また、移動体通信向けの音声コーデックとして、M-LCELP 方式をもとに開発した LSI チップを紹介する。

#### 8.2 企業内通信向けマルチパルス音声符号化装置

高速デジタル専用線を用いて企業内通信のネットワークを効率的に構築するために、第2章2.5節に記載したマルチパルス試作装置 [175] をベースにして音声符号化装置ユニバーサルトランスコーダを開発した [173]。

本装置は、図8.1に示すように、デジタル電話交換機 (LS: Line Switch) またはデジタル回線中継交換機 (TLS: Toll Line Switch) とマルチメディア多重化装置 (MMM: MultiMedia Mux) との間に接続され、デジタル電話交換機から入力される 64 kb/s PCM 信号を 32 kb/s、16 kb/s、8 kb/s に圧縮符号化する。符号化方式としては 32 kb/s ADPCM 及び、16 kb/s と 8 kb/s マルチパルス方式 (MPC) を採用して

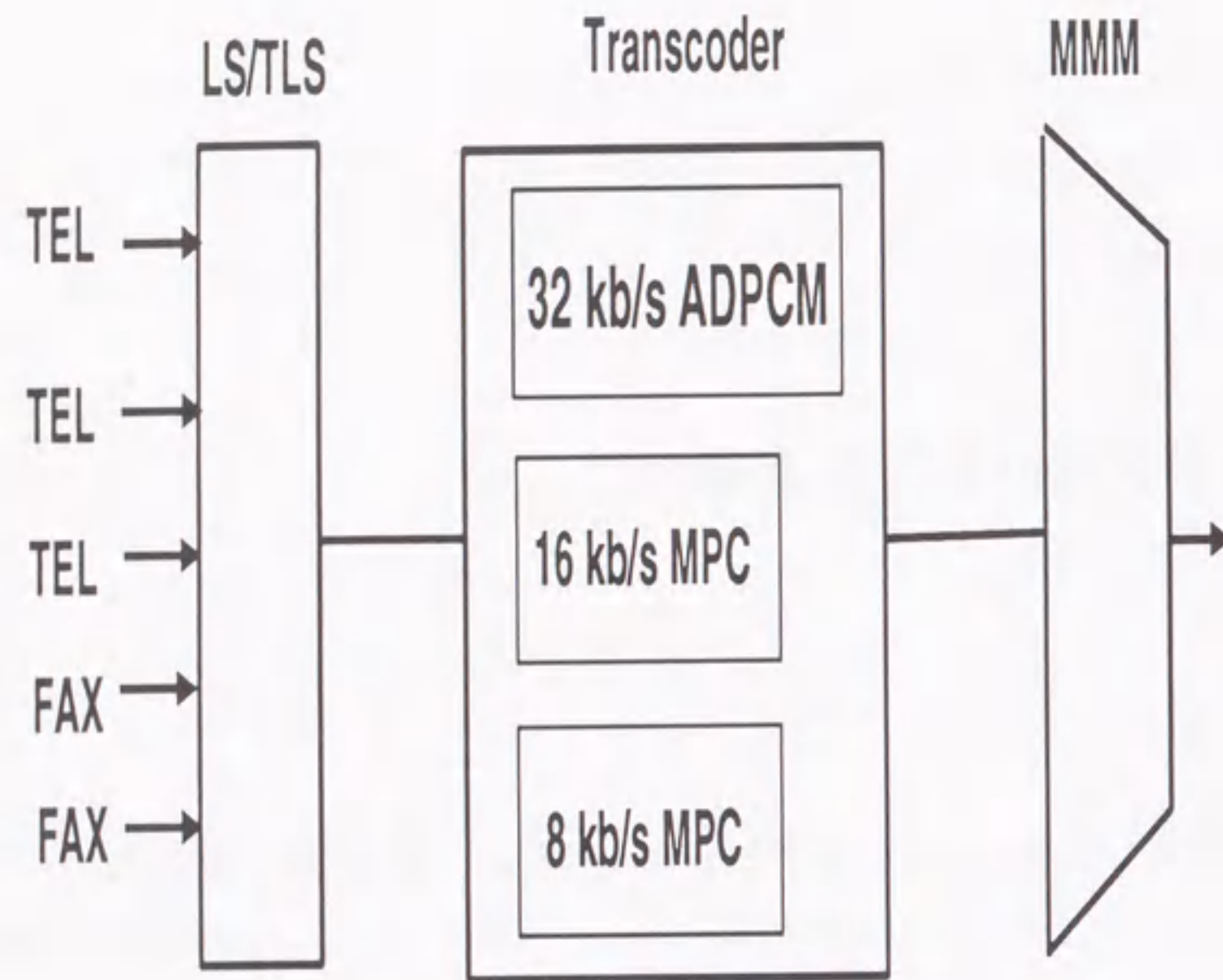


図 8.1: 企業内通信向け音声符号化装置の接続の例, LS: Line Switch, TLS: Toll Line Switch, MMM: MultiMedia Mux

いる。16 kb/s、8 kb/s マルチパルス符号化には、それぞれ、第 2 章 2.3.1 節の方式 1、第 3 章 3.2.3 節のピッチ予測マルチパルス (P-3 法) を採用している。

本装置は 32 kb/s、16 kb/s、8 kb/s の符号化パッケージを最大 30 チャンネルまで実装することができる。従来の PCM に比べ、16 kb/s で 4 倍、8 kb/s で 8 倍のチャンネル容量を確保できるので、通信コストの削減や他の信号との多重化伝送に貢献することができる。ユニバーサルトランスコーダの外観を図 8.2 に示す。

マルチパルス符号化部は、NEC 製の浮動小数点 DSP ( $\mu$ PD 77230R) [99] を用いて実現されている。装置に実装されている音声符号化パッケージを図 8.3 に示す。

マルチパルス音声符号化装置は、上記の装置以外にも表 8.1 のようなファミリーが

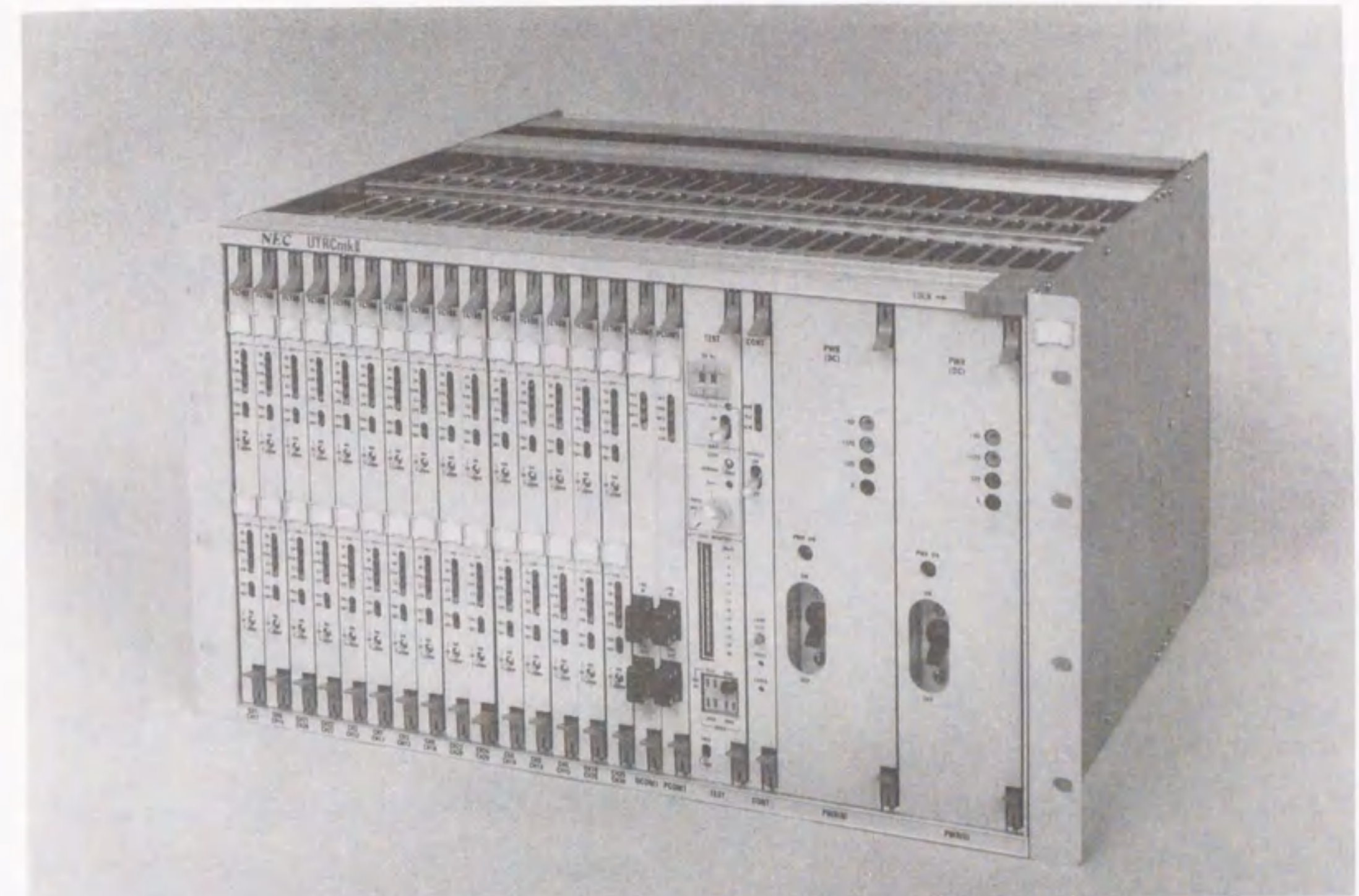


図 8.2: ユニバーサルトランスコーダの外観

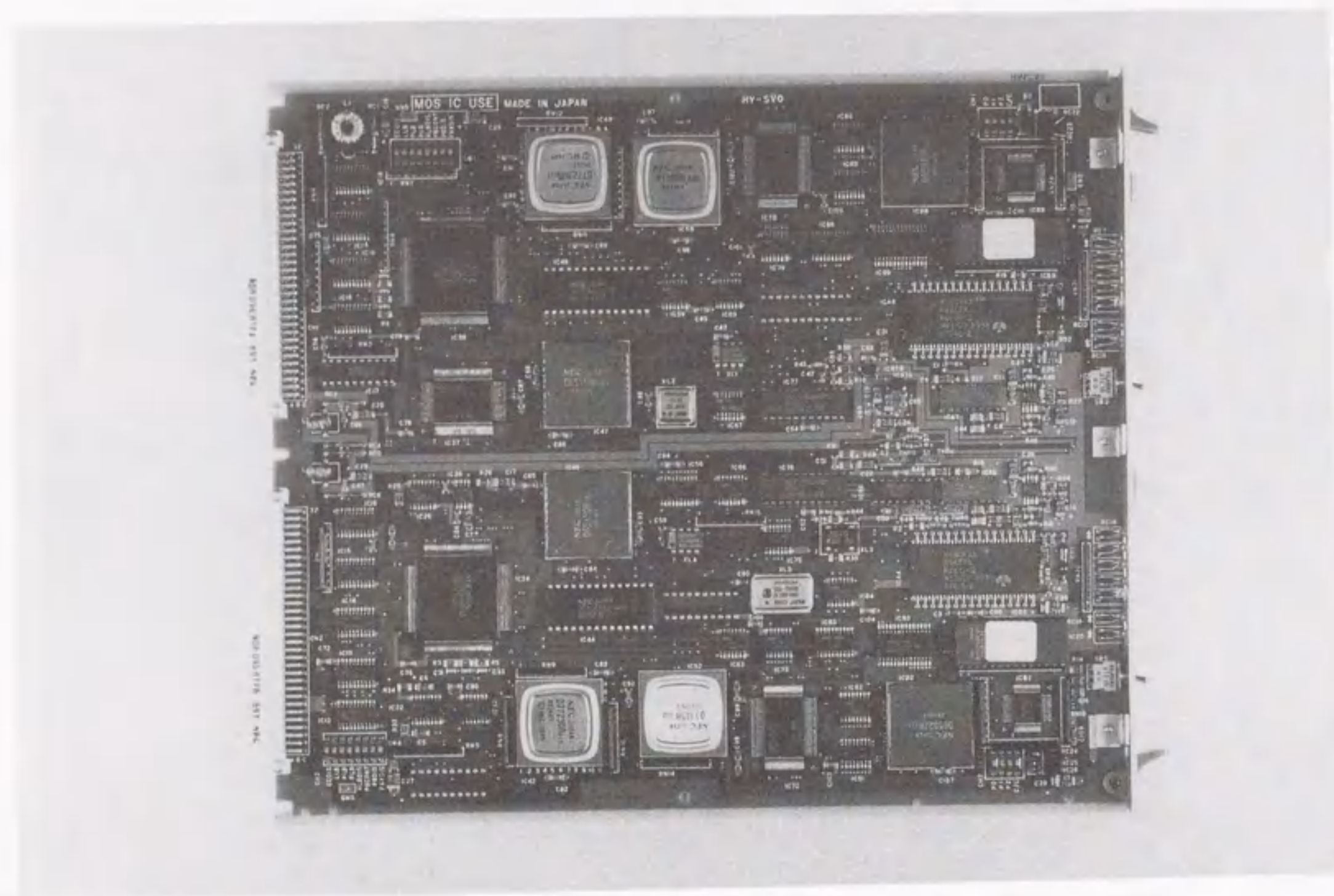


図 8.3: ユニバーサルトランスコーダの音声符号化パッケージ

表 8.1: マルチパルス音声符号化装置ファミリー

製品名	特徴		
	符号化出力	音声入力	符号化速度 (kb/s)
MPC9600	中継交換機	回線交換機	9.6
MPC1600	中継交換機	回線交換機	16
ネオコーダ MP	RS-232C	電話機	8, 9.6, 16

表 8.2: マルチメディア多重化装置向け 9.6 kb/s, 8 kb/s LCELP 音声符号化諸元

項目	仕様
エンコーダ	$\mu$ PD 77230 及び DSP16A
デコーダ	$\mu$ PD 77230
入力フォーマット	64 kb/s $\mu$ -law PCM
フレーム長, サブフレーム長	20 ms, 5 ms
エンコーダ実行時間	$\mu$ PD 77230: 8.8 ms/ フレーム DSP16A: 16.2 ms/ フレーム
処理遅延	エンコーダ: 60 ms, デコーダ: 20 ms

製品化されており [77], [183]、企業内ネットワークなどで活躍している。

### 8.3 ネットワーク向けマルチメディア多重化装置

高速デジタル回線、ISDN (Integrated System Digital Network)、国際デジタル回線、衛星回線や、今後の B-ISDN (Broadband-ISDN) などに適用して効率のよいネットワークを構築するために、マルチメディア多重化装置を開発した。本装置によるネットワーク構築例を図 8.4 に示す。

音声符号化には、32 kb/s ADPCM、16 kb/s マルチパルス符号化、ITU-T 標準 16 kb/s LD-CELP [15] に加え、9.6 kb/s および 8 kb/s LCELP (Learned Code Excited LPC) 符号化方式 [82], [83], [84], [85], [172] を採用した [169]。LCELP 符号化

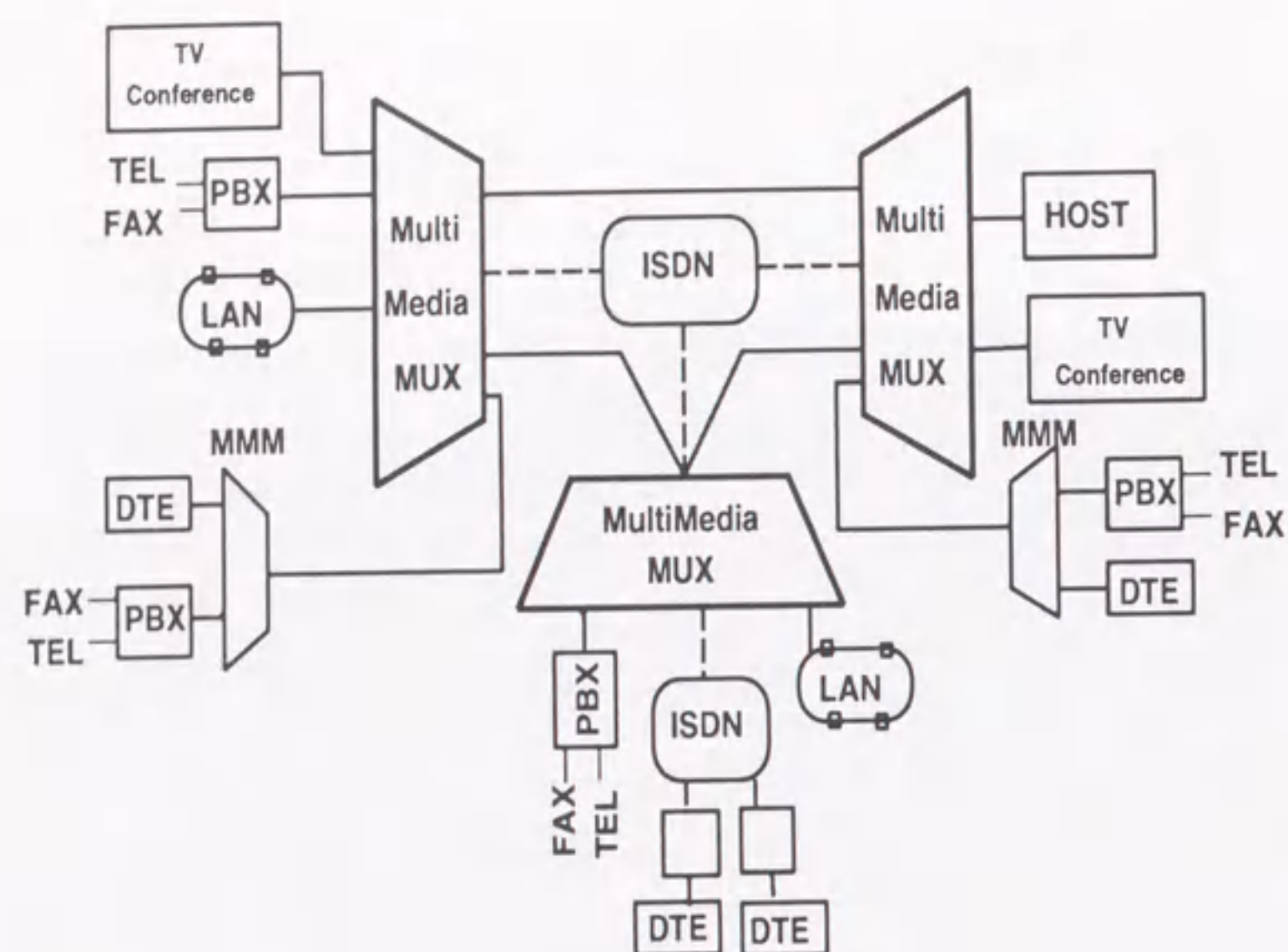


図 8.4: マルチメディア多重化装置によるネットワーク構築例

は、第 5 章の M-LCELP 符号化の機能を一部簡略化しているが、音源信号をベクトル量子化している点は同じである。LCELP 音声符号化部の諸元を表 8.2 に示す [169]。

エンコーダは、演算精度が必要な LPC 分析部と LSP 変換部を NEC 社製の浮動小数点 DSP ( $\mu$ PD 77230) により実現し、高速演算が必要なピッチ探索と音源コードブックの探索には、AT&T 社製 16 ビット固定小数点 DSP (DSP16A) を使用した。これらの処理はそれぞれ、1 フレーム以内に終了する構成としているため、エンコーダは合計で 3 フレーム分の処理遅延 (60ms) を必要とする。LCELP 音声符号化ボードの外観を図 8.5 に示す。

#### 8.4 移動体通信向け音声コーデックチップ

デジタル移動体通信向けの音声コーデックチップを開発した [94]。符号化方式は第 5 章の M-LCELP をベースに改良を加えた ELCELP 方式 [28] を採用した。全体のビットレートは 6.4 kb/s とし、7% までのバースト伝送路誤りに頑健にするために、

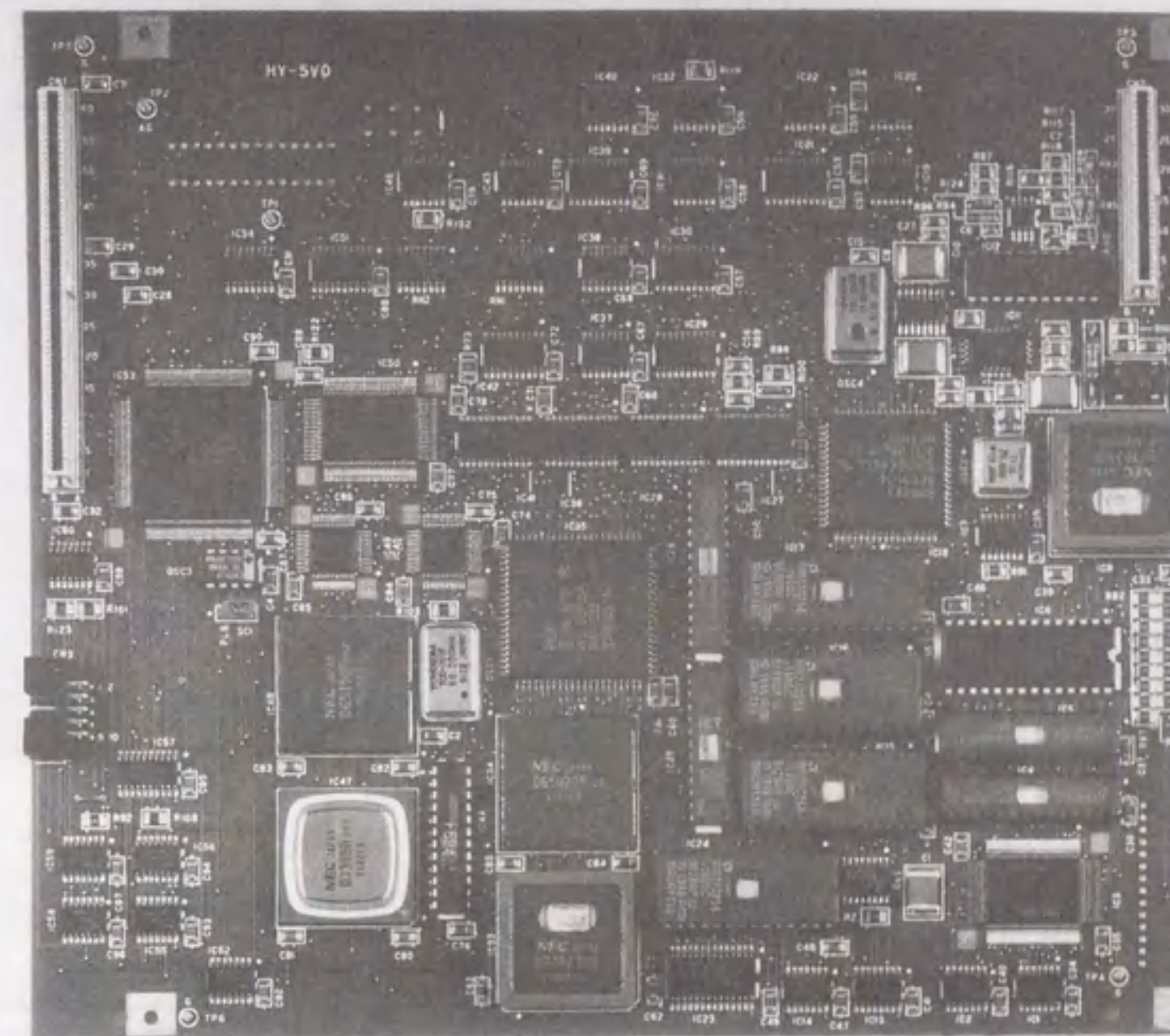


図 8.5: 9.6 kb/s, 8 kb/s LCELP 音声符号化ボード

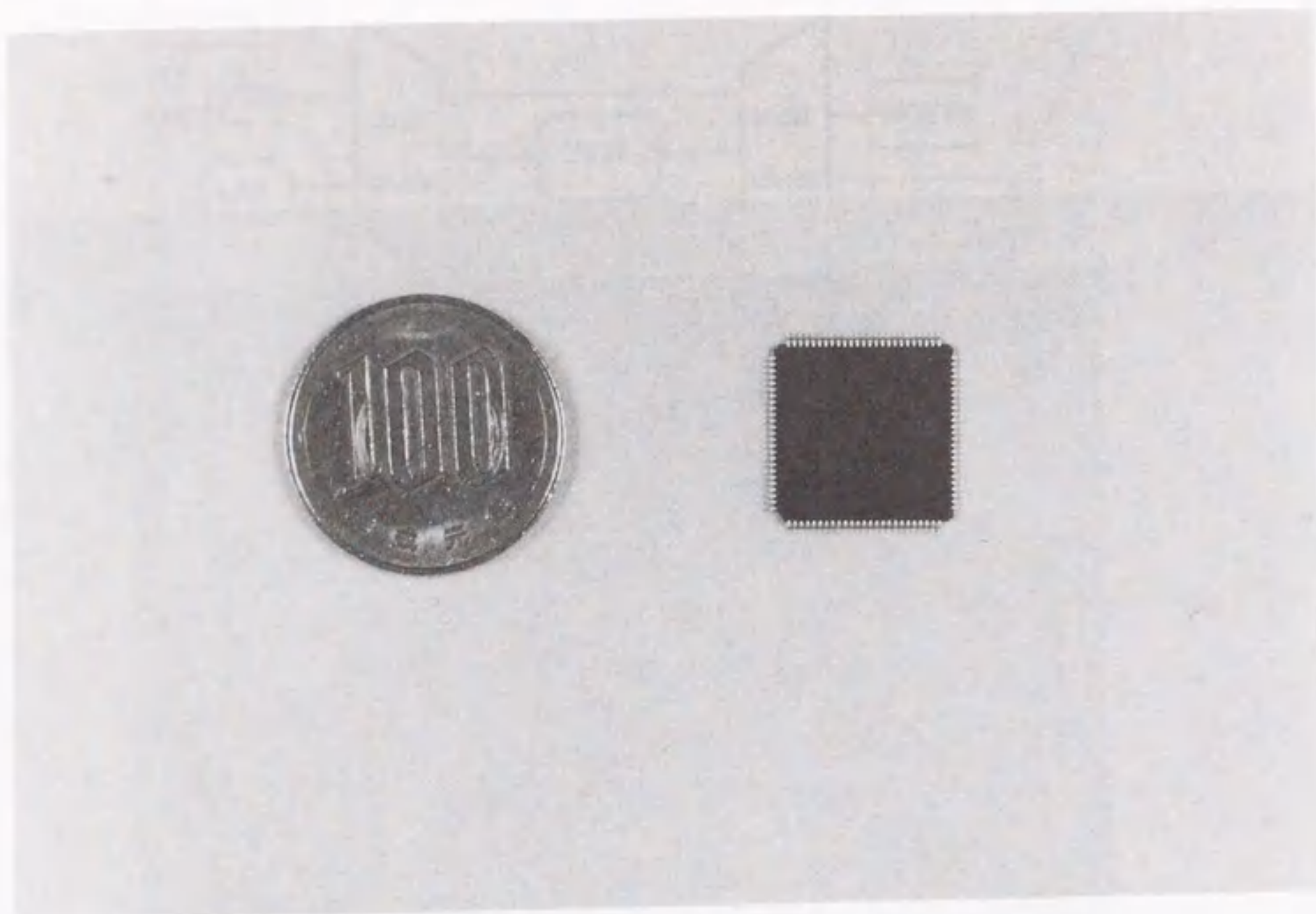


図 8.6: 移動通信用 6.4 kb/s 音声コーデックチップの外観

音声符号化を 3.2 kb/s に低減し、残りの 3.2 kb/s を誤り訂正符号化に使用した。誤り訂正符号化は、第 5 章 5.3.2 節で述べた M-LCELP 向けの FEC [39] を基に、冗長ビット数を増やし頑健性を高めた。DSP チップには NEC 社製 16 ビット固定小数点 DSP ( $\mu$ PD77018) [96] を使用し、1 チップ上にエンコーダとデコーダの両者を実現した。本 DSP の主要諸元を表 8.3 に示す。

6.4 kb/s 音声コーデックチップの外観を図 8.6 に示す。また、コーデックの諸元を表 8.4 に、本コーデックのメモリ使用量を表 8.5 に示す。

表 8.3: NEC DSP  $\mu$ PD77018 の主要諸元

項目	仕様
インストラクションサイクル	30 ns
マルチプライアキュムレータ	16 bit x 16 bit + 40 bit $\rightarrow$ 40 bit
汎用レジスタ	8 本 x 40 bit
データ RAM	X 領域: 3 kW x 16 bit, Y 領域: 3 kW x 16 bit
データ ROM	X 領域: 12 kW x 16 bit, Y 領域: 12 kW x 16 bit
命令 ROM	24 kW x 32 bit
電源	2.7 ~ 3.6 v

表 8.4: 移動体通信用 6.4 kb/s 音声コーデックの諸元

項目	仕様
DSP	NEC 製 16 ビット固定小数点 $\mu$ PD77018
フレーム長, サブフレーム長	40 ms, 8 ms
ビットレート	音声: 3.2 kb/s, 誤り訂正: 3.2 kb/s
音声符号化方式	ELCELP

表 8.5: 移動体通信用 6.4 kb/s 音声コーデックチップのメモリ使用量

項目	内容
データ RAM (X 領域)	2.937 kW
データ RAM (Y 領域)	3.043 kW
データ ROM (X 領域)	6.321 kW
データ ROM (Y 領域)	5.768 kW
命令 ROM	22.701 kW

### 8.5 むすび

本論文の第2章で検討したマルチパルス音声符号化方式、第3章で検討したピッチ予測マルチパルス符号化方式や、第5章で述べたM-LCELP音声符号化方式にもとづく製品の例を紹介した。これらは、有線通信の分野での企業内ネットワークにおいて、音声信号を64 kb/s PCMの1/8程度まで圧縮し、他のマルチメディア信号と多重化して効率的に伝送することで、通信コストの削減などに貢献することができる。また、無線伝送路の誤りに頑健な6.4 kb/s音声コーデックチップは今後、デジタル移動体通信の分野で重要になると考えられる。

## 第9章

### 結論と今後の課題

#### 9.1 結論

本論文は、16 kb/s ~ 4 kb/sにおけるA-b-S探索に基づく線形予測音声符号化方式の研究開発に関する著者の研究成果をまとめたものである。この研究の背景には、特に、近年のデジタル移動体通信分野における音声符号化に対する需要の急速な増大と実用化の進展がある。

本研究の目的は、16 kb/s ~ 4 kb/sの伝送速度において、高音質化と、演算量、メモリ量の低減化を両立させるための音声符号化アルゴリズムを提供することにある。本論文では、特に、音質、演算量、メモリ量と極めて密接な関係を有する音源信号の構造に焦点をあて、マルチパルス音源とベクトル量子化音源に基づく符号化方式を対象に研究を行なった。

第1章は、研究の背景、従来の音声符号化技術の概要を述べ、音声符号化に対する要求条件、本研究の目的と課題を明らかにした。

第2章~4章は、マルチパルス音源を用いるマルチパルス音声符号化(MPELP: Multi-Pulse Excited Linear Prediction)について、主として、マルチパルス音源の探索アルゴリズムの検討とピッチ情報の利用による伝送速度の低減化の検討を行なった。

第2章では、マルチパルスの位置と振幅を一パルスずつ逐次的に探索する方法として、聴感重み付フィルタを通した後の聴感重み付け誤差電力最小化基準に基づき、

相関演算に基づく演算量の少ない極めて効率的な方法から、さらに性能を改良する方法まで、4種類の方法を提案した。これらのパルス探索法を用いるマルチパルス音声符号化を構築し、16 kb/s、9.6 kb/sにおいて、これらの探索法の性能、音質と演算量を比較検討した。16 kb/sでは、パルス探索法を改善することによる効果は大きく、 $\mu$ -law PCM 56 kb/sに近い極めて良好な音質が得られた。9.6 kb/sでは、この改善度は16 kb/sほど高くはなかったが、 $\mu$ -law PCM 52 kb/s相当の良好な音質が得られた。音質と演算量の関係を考慮して、実用的な探索法を選択した。また、16ビット固定小数点DSP(命令実行速度250 ns)を複数チップ使用して16 kb/s、9.6 kb/sデュアルレートマルチパルス符号化装置を試作した。

第3章では、マルチパルス符号化をベースに、有声区間において音声波形のピッチ周期性を利用し、ピッチ予測技術を導入することで、伝送速度を8 kb/sまで低減化した。ピッチ予測技術として、ピッチフィルタのインパルス応答をパルス探索における合成フィルタのインパルス応答に含めるフレーム内ピッチ予測(P-1法)、フレーム間・フレーム内ピッチ予測併用型(P-2法)及び、サブフレーム間ピッチ予測(P-3法)の3種類の方式を提案し、9.6 kb/s、8 kb/sにおいて、主観的、客観的に比較評価した。特に、P-1法は、従来のAPCなどに使用されているピッチ予測法や、他の機関ではほぼ同時期に独立に検討された方式や、CELPに使用されている適応コードブックとは、構造の異なる独自の方法である。主観音質評価結果によれば、9.6 kb/s、8 kb/sの両伝送速度において、総合的に、フレーム内ピッチ予測法(P-1法)が最も良好であった。9.6 kb/sの音質は $\mu$ -law PCM 6.6ビット(52.8 kb/s)相当であり、8 kb/sでは $\mu$ -law PCM 6.3ビット(50.4 kb/s)相当と、いずれも良好であった。ピッチ予測の導入による音質改善度は、ピッチ予測のない場合に比べ、9.6 kb/s、8 kb/sではそれぞれ、 $\mu$ -law PCM 0.4ビット相当、 $\mu$ -law PCM 0.3ビット相当であった。また、 $SNR_{seg}$ 評価によれば、ピッチ予測の導入により、 $SNR_{seg}$ が1.3 dB ~ 2 dB改善された。以上により、本方式の有効性が明らかになった。

なお、P-1法によるピッチ予測技術を第6章のMP-CELPにおいても採用した。

第4章では、マルチパルス符号化において、伝送速度を4.8 kb/sまで低減化するために、音源信号の効率的な表現を一層進めた。このために、マルチモード化による

音源信号の切替えと、母音モードでのピッチ同期マルチパルス音源(PSMP: Pitch Synchronous Multi-Pulse)を提案した。フレーム毎に、音声信号から音響的特徴を求め、開ループにてモードの分類を行なった。ピッチ同期マルチパルス音源(PSMP)では、フレームの音源信号を少ない個数のマルチパルスにより極めて効率的に表現するために、代表的な一つのピッチ区間でマルチパルス音源を求め、同一フレームの他のピッチ区間では代表区間のマルチパルスをピッチ同期で繰り返しながら、振幅・位相補正係数により振幅、位相を調整した。この構成に加え、フレーム内で、波形歪みを最小にする代表区間を探索し、その位置を伝送することにより、フレーム内での音源信号の時間変化にも対応できるようにした。一方、過渡モードや鼻音モードでは、マルチパルス音源、子音モードでは乱数コードブック音源に切り替えた。実験結果によれば、提案方式の4.8 kb/sの音質は、 $\mu$ -law PCM 6ビット(48 kb/s)相当と良好な音質を得た。この音質は、ピッチ予測を用いない8 kb/sマルチパルス符号化と同等であった。

第5章は、線形予測分析とA-b-Sに基づく音声符号化の枠組はそのままとし、量子化効率を高め、低伝送速度での音質をさらに改良するために、ベクトル量子化音源について検討した。ここでは、従来のCELP(Code Excited Linear Prediction)方式の低伝送速度における音質改良も考慮し、第4章で述べたマルチモード符号化の考えとベクトル量子化音源を組合せたM-LCELP(Multi-mode Learned CELP)音声符号化方式を提案した。本方式によれば、モードに応じて、ベクトル量子化音源とパラメータ量子化のコードブックを切り替えることにより、量子化の効率化をはかった。さらに、ベクトル量子化音源において、3種のコードブック構造について、 $SNR_{seg}$ 性能、演算量、メモリ量を比較検討し、マルチパルス列から構成されるスパース音源コードブック構造を採用した。スパース音源におけるパルスの振幅、位置は、音声信号データベースを用いてコードブック学習により、各コードベクトル毎にあらかじめ決定した。このときのパルスの計算には第2章で提案した探索法を使用した。音声のビットレートを4.1 kb/sとし、北米のハーフレート移動電話環境などにおける種々の条件で評価するために、2.4 kb/sの誤り訂正FEC(Forward Error Correction)を付加した。音質MOS評価結果によれば、誤り無し、背景雑音や、誤り有りなどの種々の評価条件において、ビットレートが2倍の北米移動電話フルレート標準方式13 kb/s



VSELP (Vector Sum Excited Linear Prediction) と同等という良好な音質が得られた。また、浮動小数点 DSP を使用して、VSELP の 2 倍以下の演算量 (18.2 MOPS) で試作 CODEC を実現することができた。しかしながら、移動体端末などでの実現の容易性を考慮すると、演算量、メモリ量をさらに低減化することが望ましく、このために、音源モデルの一層の改善が課題であった。また、フレーム長が 40 ms と長い場合、遅延の低減化も重要であった。

第6章は、マルチパルス音源とベクトル量子化の結合をはかった。5章で検討した M-LCELP に対して、高音質化及び、演算量、メモリ量の一層の低減化を狙い、マルチパルス音源とパルス振幅のベクトル量子化を組み込んだ MP-CELP (Multi-Pulse based CELP) 方式を提案した。さらに、パルス振幅のベクトル量子化による性能を向上させるため、複数セットの位置候補と振幅コードベクトルを組合せ探索する方法を提案し、性能改善量を明らかにした。パルス位置を総組合せで最適に求める場合の演算量を削減するために、パルスをグループに分割して位置を高速探索する分割探索法を提案し、演算量と性能の関係を示した。分割数を4まで増やし演算量を大幅に削減しても、 $SNR_{seg}$  の低下は少なかった。これらの提案法を組み込んだ 11 kb/s MP-CELP を構築した。欧州 GSM フルレートの伝送路を想定し、11.8 kb/s の軟判定 FEC を組み込んだ。音質評価結果によれば、本方式は、背景雑音条件を除く基本特性、話者依存性、2 タンデム接続条件などで、32 kb/s ADPCM を越える音質を有する ITU-T 標準 G.728 (16 kb/s LD-CELP) と同等の音質を示した。また、従来の GSM フルレート符号化との比較では、全ての条件で大幅な音質の向上が見られた。特に、基本特性、バイオリン雑音条件がそれぞれ MOS 値で 0.7、0.8 の改善、7%、13% 誤りが 1.0 ~ 1.5 の改善であった。提案方式のエンコーダの演算量は、分割探索数を3にした場合、約 9 MOPS と、従来の CELP に比べ少なく、コードブック格納に必要なメモリ量も 5 k ワード以下と低減できた。また、フレーム長が 10 ms のため、遅延も少なくできる。今後は、特に移動体通信の分野などで有効になると考えられる。

第7章は、デジタル移動体衛星通信やデジタル移動通信への適用を考慮して、16 kb/s、9.6 kb/s マルチパルス符号化の伝送路誤りに対する品質を評価した。ランダム誤りに対しては、誤り訂正なしでも  $10^{-3}$  までは劣化が少ないが、 $10^{-2}$  にすると

劣化が大きいことがわかった。また、移動体衛星通信システムで通常使用される誤り訂正 FEC の復号後に残存するバースト誤りに対しては、符号誤りに敏感な PARCOR 係数の上位ビットとパルス振幅最大値の上位ビットをフレーム当たり合計で 12 ビット保護することにより、450 b/s という極めて少ない冗長ビット数の追加で、誤り率が  $10^{-3}$  での音質が PCM 換算で 0.5 ビット ~ 1 ビット相当改善されることがわかった。

第8章は、本論文の第2章で検討したマルチパルス音声符号化方式、第3章で検討したピッチ予測マルチパルス符号化方式や、第5章で述べた M-LCELP 音声符号化方式にもとづく製品の例を紹介した。有線通信の分野での企業内ネットワークにおいて、音声信号を 64 kb/s PCM の 1/8 程度まで圧縮し、他のマルチメディア信号と多重化して効率的に伝送することで、通信コストの削減などに貢献している。また、無線伝送路の誤りに頑健な 6.4 kb/s 音声コーデックチップは、今後移動体通信の分野で重要になると考えられる。

## 9.2 本研究の意義と貢献

本研究を通して研究開発した高能率音声符号化アルゴリズムは、それぞれのビットレートで、先駆的な研究の意義をもつ点、従来の方式に比べより高い音質を提供することができる点で画期的である。特に、第2章2.3節で提案した相関演算に基づくマルチパルス音源の探索アルゴリズムや、第3章3.2.1節のピッチフィルタをインパルス応答に含めるピッチ予測アルゴリズム (P-1法) は、1995年にITU-Tで標準化された G.729 (8 kb/s CS-ACELP) [148], [58], [50]、G.723.1 (6.3 kb/s MP-MLQ 及び 5.3 kb/s ACELP) [49]、ACELP [147]、1996年に欧州 ETSI/GSM で選定された音質改善 13 kb/s フルレート (EFR) [25] や、PCS 1900 向けに北米 TIA で選定された 13 kb/s EFR [23] などの音源パルス探索部に同種の手法が採用され、高い評価を得ている。また、上記マルチパルス探索アルゴリズムは、音声符号化の分野のみならず、音源信号からの話者認識、音声合成 [174], [35]、オーディオ符号化 [158] や、画像処理 [105] など、種々の分野で使用され、研究の発展に貢献している。また、第4章4.2.3節で述べたピッチ同期マルチパルス音源 (PSMP) では、フレームで代表的な一つのピッチ区間の音源信号を伝送するという基本的な考え方が、PWI (Pitch Wave-

form Interpolation) 音声符号化 [64] などに適用され、2.4 kb/s 程度の伝送速度で有力な方式となっている。さらに、第4章4.2.1節や第5章5.2.1節に示した音響的特徴から閉ループで音源信号を切替えるマルチモード化技術や、第5章5.2.4節のスパース構造ベクトル量子化技術なども重要度を増してきている [88], [161]。第6章のマルチパルスベクトル量子化音源を用いる MP-CELP 符号化アルゴリズムは、高音質化と、演算量、メモリ量の低減化を両立することができる点で、実用的に極めて有効な方式と考えられ、音声符号化の今後の進展に貢献する意義は大きい。

### 9.3 今後の研究課題

将来的な方向としては、移動体通信の分野では、音声、画像、オーディオ、データ信号などを含むマルチメディア情報を伝送したり、有線とのシームレスな接続が増大することが予想される。また、パーソナルコンピュータ (PC) やインターネットの急速な浸透により、PC 間のマルチメディア通信、PC 間の会議、IC カード等による音声の蓄積などが盛んになることが予想される [138], [141]。

このような動向を考慮して、今後重点的に検討すべき課題として、次のものが考えられる。

#### (1) 4 kb/s 以下の伝送速度における一層の高音質化:

21 世紀初頭には実用化される予定のマルチメディア次世代移動体通信 FPLMTS (Future Public Land Mobile Telecommunication Systems) や UMTS (Universal Mobile Telecommunication Systems) などでは、16 kb/s ~ 4 kb/s の広い範囲にわたり、32 kb/s ADPCM と同等以上の音質が必要とされる。また、移動体携帯テレビ電話や、モデム伝送を用いる TV 電話などへの応用も考えられている。現状では、8 kb/s まで良好な音質を確保できるので、今後は、まずは 4 kb/s の伝送速度で、その後は 3 ~ 2 kb/s で、32 kb/s ADPCM と同等の音質を確保することが重要である。

これらに対して、技術的には、分析法の改良 [65], [66], [155] や、一層効率的な音源モデルの検討が必要であろう。特にマルチパルスをベースにした音源モデルの場合は、位置の量子化効率の一層の改善や、文献 [11], [10] に示唆されたように、マルチパルス音源と主観音質との対応関係をより明確化する必要がある。これら以外に、音

声の知見の積極的な導入や、A-b-S 分析での歪み尺度として聴感と対応づけのよい歪み評価尺度の研究 [151], [10], [178], [162] などますます重要となろう。

ITU-T では、FPLMTS を主要な応用として、4 kb/s で 32 kb/s ADPCM と同等の音質を有する符号化を 2000 年頃までに標準化する予定であり [52]、積極的に貢献していきたい。

一方、ISO (International Standardization Organization) では、MPEG (Moving Picture coding Expert Group を最近 Coding of Moving Pictures and Audio に改称) において、MPEG-4 の一つのツールとして、音質を確保しながら、音声帯域の信号に対して伝送速度を 4 kb/s 以下から 16 kb/s 程度まで可変にできる符号化方式の標準化を目指しており、NEC は第6章の MP-CELP を基本にした方式を提案している [103], [104]。標準化終了は 1998 年秋の予定である。

#### (2) 背景雑音対策:

移動体通信の分野では、背景雑音に対する頑健性が要求される。第5章5.3節に示したように無音箇所を判別し信号のレベルを抑圧する手法や、ノイズキャンセラを併用して雑音を消去する方法なども考えられる。一方で ITU-T などでは、背景雑音は重要な環境情報を含んでおりできる限り忠実に伝送したいという考え方が強く、方式標準化で主観評価を行なうときに、背景雑音条件では、原音声と符号化音声の DCR (第6章6.3.3節で使用) 評価や、第3章3.3.3節で使用した一対比較と同種の CCR (Comparison Category Rating) 評価を検討している [52]。

線形予測符号化では、線形予測分析に基づく音声合成モデルを用いているので、特に、伝送速度が 8 kb/s 以下と低い場合は、予測利得の低い背景雑音に対して音質が劣化してくる。これを改善するために、背景雑音に適したモデルを開発し音声合成モデルと切替えて使用することや、処理法を検討する必要がある。

#### (3) 広帯域音声・音楽符号化の伝送速度低減化:

PC 会議やテレビ会議では、臨場感の向上のために、従来の電話帯域の信号ではなく、2 倍の帯域 (帯域幅は 50 ~ 7000 Hz) を有する広帯域信号の符号化が重要である。PC 会議などで、電話モデムによる伝送などを考えると、伝送速度はまずは 16 kb/s、その後は 9.6 kb/s などが望ましい。これら以外にも、将来的には、広帯域電話器など

や移動体通信での広帯域音声サービスなども実用化されると予想される。音質は、ITU-T 標準 G.722 サブバンド ADPCM [79] の 48 kb/s ~ 56 kb/s 相当の高音質が目標である。また音声のみならず、音楽信号も良好に符号化する必要がある。音楽信号に対しては、直交変換に基づくアプローチも研究が進んでいるが [90], [53]、線形予測符号化の枠組を用いる場合は、従来使用されている音声合成モデルは音楽には整合しにくいので、音楽に適した新たなモデルを研究する必要がある。

この帯域の符号化は、ITU-T において、伝送速度を 16, 24, 32 kb/s として、1995 年から方式標準化活動が開始されており [51]、1999 年までに標準化を終了する予定である。NEC も第 6 章の MP-CELP に基づく方式を提案している [93], [156]。

## 謝辞

本研究は、NEC (株) 中央研究所において、1982 年より行なってきたもので、今日まで、多くの方々のご指導、ご助言、ご協力を賜った。ここに深く感謝の意を表する。

故加藤康雄専務取締役、石黒辰雄取締役、飯沼一元博士、丹羽邦彦博士、後藤敏博士、山本昌弘博士には、研究所長あるいは、研究部長として、本研究の機会を与えて下さり、終始ご支援、ご助言を頂いた。

落合和雄氏、荒関卓氏には、本研究前半での上司として、的確なご指導、ご鞭撻と暖かい援助を頂いた。

西谷隆夫博士、渡辺隆夫博士、亘理誠夫氏、村木一至氏には、本研究後半の上司として、ご指導、ご助言、ご理解を頂いた。

芹沢昌宏博士、小野茂博士、田海真一氏、野村俊之氏、池川将夫氏、宮野俊樹氏には、共同研究者として、多大なご協力とご討論を頂いた。

日本電気情報システムズの熊谷弘己氏には、ソフト開発でご協力いただいた。

田中俊二氏、安永智氏、和気靖浩氏、海野義博氏、富依牧生氏、早田利浩氏、長崎真由実氏、舟木慶一氏、吉田和永氏には、音声符号化装置の試作にご支援をいただいたり、事業部にて製品化していただいた。

本論文をまとめるにあたっては、名古屋大学大学院工学研究科 電子情報専攻 板倉文忠 教授より、懇切なるご指導を賜った。また、谷本正幸 教授、武田一哉 助教授より有益なご教示を賜った。深く感謝します。

## 参考文献

- [1] J. B. Anderson and J. B. Bodie, "Tree encoding of speech," *IEEE Trans. IT*, vol.IT-21, pp.379-387, 1975.
- [2] M. Akamine, K. Miseki, "CELP coding with an adaptive density pulse excitation model," *IEEE Proc. ICASSP*, pp. 29-32, 1990.
- [3] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals," *Reports of 6th Int. Cong. Acoust.*, C-5-4, 1968.
- [4] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals," *Bell System Technical Journal*, pp. 1973-1986, Oct., 1970.
- [5] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction," *J. Acoust. Soc. Am.*, vol.50, pp.637-655, 1971.
- [6] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust. Speech and Signal Processing*, vol. ASSP-27, No.3, pp.247-254, 1979.
- [7] 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*, pp. 614-617, April, 1982.
- [8] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. COM-30, no.4, pp.600-614, 1982.

- [9] B. S. Atal, "High-quality speech at low bit rates: Multi-Pulse and stochastically excited linear predictive coders," *IEEE Proc. ICASSP*, pp.1681-1684, 1986.
- [10] B. S. Atal, "Speech coding and human speech perception," Eds. B. A. G. Elsendoorn and H. Bouna, *Working Models of Human Perception*, pp. 101-125, Academic Press, 1989.
- [11] B. Caspers and B. S. Atal, "Role of multi-pulse excitation in synthesis of natural-sounding speech," *IEEE Proc. ICASSP*, pp.2388-2391, 1987.
- [12] CCITT, "32 kbps adaptive differential pulse code modulation," in *CCITT Red Book III*, Oct. 1984.
- [13] CCITT Recommendation P.48, "Specification for an intermediate reference system," Volume V, 1988.
- [14] CCITT Recommendation G.726, "40, 32, 24, 16 kbit/s Adaptive differential Pulse Code Modulation (ADPCM)," 1990.
- [15] J-H. Chen, R. V. Cox, Y-C. Lin, N. Jayant and M. J. Melchner, "A low-delay CELP coder for the CCITT 16 kb/s speech coding standard," *IEEE J. Sel. Areas Commun.*, vol. 10, No. 5, pp.830-849, June 1992.
- [16] J-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. Speech and Audio Proc.*, vol.3, no.1, Jan., pp.59-71, 1995.
- [17] R. V. Cox, "Speech coding standards," in *Speech Coding and Synthesis* (W. B. Kleijn and K. K. Paliwal, eds.), pp.49-78, Elsevier Science B.V., 1995.
- [18] R. E. Crochiere, R. V. Cox and J. D. Johnston, "Real-time speech coding," *IEEE Trans. Commun.*, vol. COM-30, pp. 621-634, Apr. 1982.

- [19] G. Davidson and A. Gersho, "Complexity reduction methods for vector excitation coding," *IEEE Proc. ICASSP*, pp. 3055-3058, 1986.
- [20] M. Delprat, M. Lever and C. Gruet, "A 6 kbps regular pulse CELP coder for mobile radio communications," *IEEE Workshop on Speech Coding for Telecommunications*, pp.29-30, 1989.
- [21] H. Dudley, "The vocoder," *Bell Labs Record*, vol.18, no.4, pp. 122-126, 1939.
- [22] EIA/TIA Interim Standard, "Cellular system dual-mode mobile station-base station compatibility standard IS-54-B," TIA, April 1992.
- [23] EIA/TIA Standards proposal No.3612-A, "PCS1900 Enhanced full rate codec US1," August 1996.
- [24] ETSI, "Report on the meeting about speech quality selection criteria," ETSI/SMG2 SEG TD95/19/Rev. 2, June, 1995.
- [25] ETSI, "Digital cellular telecommunications system; Enhanced full rate (EFR) speech transcoding (GSM 06.60)," Version 2.0.0, January 1996.
- [26] H. G. Fehn and P. Noll, "Multipath search coding of stationary signals with applications to speech," *IEEE Trans. Commun.*, vol.COM-30, no.4, pp.687-701, 1982.
- [27] J. L. Flanagan, "Systems for analysis-synthesis telephony," in *Speech Analysis, Synthesis and Perception*, Chapter 8, Springer-Verlag, 1972.
- [28] 舟木慶一, 吉田和永, 小澤一範, 奥村剛, "4 kbps 高圧縮率音声応答システム," *NEC 技報*, Vol. 48, No. 6, pp.10-13, 1995.
- [29] A. Gersho and V. Cuperman, "Adaptive differential vector coding of speech," *IEEE Proc. GLOBECOM*, pp.1092-1095, 1982.

- [30] A. Gersho and R. M. Gray, "Vector quantization and signal compression," Kluwer Academic Publishers, 1992.
- [31] I. A. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8kbps," IEEE Proc. ICASSP, pp.461-464, 1990.
- [32] I. A. Gerson and M. A. Jasiuk, "Techniques for improving the performance of CELP-Type speech coders," IEEE Journal Sel. Areas Commun., vol.10, no.5, pp.858-865, June, 1992.
- [33] W. Granzow, B. S. Atal, K. K. Paliwal and J. Schroeter, "Speech coding at 4 kb/s and lower using single-pulse and stochastic models of LPC excitation," IEEE Proc. ICASSP, pp. 217-220, 1991.
- [34] 肥田野直, 瀬谷正敏, 大川信明, "心理教育統計学," 第9章, 培風館, 1984.
- [35] 平井宏, 斎藤尚宏, 堤喜代司, "マルチパルス法により合成されるピッチ揺らぎに関する考察," 日本音響学会誌, vol. 51, no. 10, pp. 774-775, 1995.
- [36] 益田雅彰, 板倉文忠, "音声の適応ビット割当予測符号化," 音響学会音声研資, S80-2, 1980.
- [37] M. Honda and F. Itakura, "Bit allocation in time and frequency domains for predictive coding of speech," IEEE Trans. Acoust. Speech, Signal Proc., vol. ASSP-32, pp. 465-473, June 1984.
- [38] M. Ikekawa, T. Miyano, T. Nomura and K. Ozawa, "Channel coding for low-bit rate speech coding," Proc. SITA, pp.247-250, 1993.
- [39] M. Ikekawa, T. Miyano, S. Taumi and K. Ozawa, "Effective channel coding combined with low bit-rate speech coding for digital cellular system," IEEE Proc. VTC, Vol. 3, pp.1685-1689, 1994.
- [40] 今井秀樹, "符号理論," 第7章, 電子情報通信学会, 1990.

- [41] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," Reports of the 6th Int. Cong. Acoust., C-5-5, 1968.
- [42] 板倉文忠, 齊藤収三, "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, J53-A, pp. 35-42, 1970.
- [43] F. Itakura, S. Saito, Y. Koike, H. Sawabe and M. Nishikawa, "An audio response unit based on partial correlation," IEEE Trans. Commun., vol. COM-20, pp. 792-797, 1972.
- [44] 板倉文忠, "新しい音声分析合成方式 PARCOR," 日経エレクトロニクス, 2.12, pp. 58-75, 1973.
- [45] 板倉文忠, "線形予測係数の線スペクトル表現," 日本音響学会音声研資, S75-34, 1975.
- [46] ITU-T Recommendation P. 80, "Methods for subjective determination of transmission quality," March 1993.
- [47] ITU-T Revised Recommendation P. 81, "Modulated noise reference unit," COM 12-44-E, March 1995.
- [48] ITU-T Users Group on Software Tools, "ITU-T software tool library manual," Feb. 1995.
- [49] ITU-T Recommendation G. 723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," COM 15-153-E, July 1995.
- [50] ITU-T Recommendation G. 729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," COM 15-152-E, July 1995.
- [51] ITU-T, "Subjective Qualification test plan for the ITU-T wideband (7 kHz) speech coding algorithm," TD 56 (2/15), November 1995.

- [52] ITU-T, "Subjective Qualification test plan for the ITU-T 4-kbit/s speech coding algorithm," TD 83 (2/15), November 1995.
- [53] 岩上直樹, 守谷健弘, 三樹聡, "周波数重み付けインタリーブベクトル量子化 (Twin-VQ) によるオーディオ符号化," 音学講論, pp.339-340, Oct. 1994.
- [54] N. S. Jayant, "Adaptive quantization with a one-word memory," Bell Systems Technical Journal, pp.1119-1144, Sept. 1973.
- [55] 実広貴敏, 板倉文忠, "パルス駆動音源と後向きピッチ予測器を併用した LD-CELP 方式," 音学講論, pp. 275-276, Oct. 1992.
- [56] B. Juang and A. Gray, Jr., "Multiple stage vector quantization for speech coding," IEEE Proc. ICASSP, pp. 597-600, 1982.
- [57] 片岡章俊, 守谷健弘, "スパースな学習符号帳を用いた CELP 符号化," 音学講論, pp. 305-306, March 1994.
- [58] 片岡章俊, 林伸二, "ITU-T 標準 8 kbit/s 音声符号化 (CS-ACELP)," 音学講論, pp.299-300, Sept. 1995.
- [59] 北脇信彦, 板倉文忠, "PARCOR 係数の非線形量子化と不均一標本化による音声の高効率符号化," 信学論 A, J-61A, pp. 543-550, 1978.
- [60] 北脇信彦, 伊藤憲三, "LPC ケプストラム距離尺度による符号化非直線歪みの客観的品質推定," 音学講論, pp.415-416, Oct. 1986.
- [61] 北脇信彦 編著, "音のコミュニケーション工学 - マルチメディア時代の音声・音響技術 -, " 第 5 章, コロナ社, 1996.
- [62] W. B. Kleijn, D. J. Krasinski and R. H. Ketchum, "Improved speech quality and efficient vector quantization in SELP," IEEE Proc. ICASSP, pp.155-158, 1988.

- [63] W. B. Kleijn and P. Kroon, "A 5.85 kb/s CELP algorithm for cellular applications," IEEE Proc. ICASSP, pp.596-599, 1993.
- [64] W. B. Kleijn, "Encoding speech using prototype waveforms," IEEE Trans. Speech and Audio Proc., vol. 1, No. 4, pp.386-399, Oct. 1993.
- [65] 小石田和人, 徳田恵一, 小林隆夫, 今井聖, "メル一般化ケプストラム分析に基づく CELP 符号化," 信学技報, SP 95-51, Sept. 1995.
- [66] 小石田和人, 徳田恵一, 小林隆夫, 今井聖, "メル一般化ケプストラム分析に基づく CELP 符号化系とその評価," 音学講論, pp. 257-258, March 1996.
- [67] P. Kroon, Ed. F. Deprettere and R. J. Sluyter, "Regular-pulse excitation - A novel approach to effective and efficient multipulse coding of speech," IEEE Trans. Acoust. Speech and Signal Proc., vol. ASSP-34, No.5, pp.1054-1063, Oct. 1986.
- [68] P. Kroon and Ed F. Deprettere, "A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8 and 16 kbits/s," IEEE J. Sel. Areas Commun., vol.6, no.2, pp.353-363, Feb. 1988.
- [69] P. Kroon and B. S. Atal, "Pitch predictors with high temporal resolution," IEEE Proc. ICASSP, pp.661-664, 1990.
- [70] P. Kroon and B. S. Atal, "On the use of pitch predictors with high temporal resolution," IEEE Trans. Signal Proc., vol.39, No.3, pp.733-735, 1991.
- [71] C. Laflamme, J-P. Adoul, R. Salami, S. Morissete and P. Mabillean, "16 kbps wideband speech coding technique based on Algebraic CELP," IEEE Proc. ICASSP, pp.13-16, 1991.
- [72] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol. COM-28, No. 1, pp.84-95, Jan. 1980.

- [73] K. Mano and T. Moriya, "4.8 kbit/s delayed decision CELP coder using tree coding," IEEE Proc. ICASSP, pp.21-24, 1990.
- [74] K. Mano, T. Moriya, S. Miki, H. Ohmuro, K. Ikeda and K. Ikedo, "Design of a pitch synchronous innovation CELP coder for mobile communications," IEEE J. Sel. Areas on Commun., vol. 13, No.1, pp. 31-41, Jan. 1995.
- [75] J. S. Marques, J. M. Tribolet, I. M. Trancoso and L. B. Almeida, "Pitch prediction with fractional delays in CELP coding," European Conf. on Speech Communication and Technology, vol.2, pp.509-512, 1989.
- [76] 増井文男, 小口 "デジタル自動車電話システム用ハーフレート音声 CODEC の選定活動," 信学技報, RCS93-77, 1993.
- [77] 松本美明, 川谷内登, 中西良信, 松村尚己, 和気靖浩, 田中俊二, 安永智, "高能率音声符号化装置," NEC 技報, Vol. 41, No. 9, pp.9-18, 1988.
- [78] R. A. McDonald, "Signal-to-noise and idle channel performance of differential pulse code modulation systems-particular application to voice signals," Bell System Tech. J., vol.45, pp.1123-1151, 1966.
- [79] P. Mermelstein, "G.722, a new CCITT coding standard for digital transmission of wideband audio signals," IEEE Communications Magazine, vol. 26, no. 1, pp. 8-15, Jan. 1988.
- [80] 三樹聡, 守谷健弘, 間野一則, 大室伸, "ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP)," 信学論 A, vol. J77-A, pp. 314-324, March 1994.
- [81] T. Miki and H. Suda, "A channel/source coding for digital voice transmission in land mobile radio," Proc. Second Nordic Seminar on Digital Land Mobile Radio Commun., pp. 59-64, 1986.
- [82] 宮野俊樹, 小澤一範, "学習コードブックによる 8 kb/s CELP の改良 (LCELP)," 信学春季全大, SA-5-4, March 1990.

- [83] 宮野俊樹, 小澤一範, "学習コードブックを用いた 8 kb/s CELP (LCELP) の改善," 音学講論, pp.187-188, Sept. 1990.
- [84] T. Miyano and K. Ozawa, "Improvement on 8 kb/s CELP using learned codebook: LCELP," Proc. ICSLP, pp.97-100, 1990.
- [85] 宮野俊樹, 小澤一範, "学習コードブックによる 8 kb/s CELP の改良 (LCELP)," 信学秋季全大, 428, Sept. 1991.
- [86] T. Miyano, M. Serizawa, J. Takizawa, S. Ikeda and K. Ozawa, "Improved 4.8 kb/s CELP coding using two-stage vector quantization with multiple candidates," IEEE Proc. ICASSP, pp. I-321-324, 1992.
- [87] 宮野俊樹, 芹沢昌宏, 野村俊之, 小澤一範, "3.6 kbps LCELP 音声符号化," 信学春季全大, SA-5-10, March 1993.
- [88] 森井利幸, 田中直也, 吉田幸司, "音声の短時間的特徴に対応したマルチモード CELP 符号化," 信学技報, SP95-80, Nov. 1995.
- [89] 守谷健弘, 菅田雅彰, "残差信号の周波数領域でのベクトル量子化," 日本音響学会音声研資 S82-47, Nov. 1982.
- [90] T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," IEEE J. Sel. Areas Commun., vol. 6, No. 2, pp.425-430, 1988.
- [91] T. Moriya, "Two-channel conjugate vector quantizer for noisy channel speech coding," IEEE J. Sel. Areas Commun., vol.10, No.5, pp.866-874, 1992.
- [92] 守谷健弘, 大室伸, "フレーム相関を用いる LSP パラメータの符号化," 音学講論, 2-5-12, pp. 259-260, Oct. 1992.
- [93] 村島淳, 芹沢昌宏, 小澤一範, "16 kbps 全帯域型 MP-CELP 広帯域音声符号化," 信学春季全大, SD-5-2, March 1997.



- [94] 長崎真由美, 海野義博, 富依牧生, 柳本雅俊, 小澤一範, 吉田和永, "6.4 kbps EL-CELP 音声 CODEC チップの開発," NEC 技報, vol.49, No.11, pp.48-54, 1996.
- [95] J. Natvig, "Pan-European speech coding standard for digital mobile radio," Speech Communication, pp. 113-123, 1988.
- [96] NEC, "μPD7701x ファミリ デジタルシグナルプロセッサ ユーザーズマニュアル,"
- [97] 新実康永, "音声認識," 共立出版, 1980.
- [98] T. Nishitani, S. Aikoh, T. Araseki, K. Ozawa and R. Maruta, "A 32 kb/s toll quality ADPCM codec using a single chip signal processor," IEEE Proc. ICASSP, pp. 960-963, May 1982.
- [99] T. Nishitani, I. Kuroda, Y. Kawakami, H. Taka and T. Nukiyama, "Advanced single-chip signal processor," IEEE Proc. ICASSP, 8.7, 1986.
- [100] T. Nomura, K. Ozawa, M. Serizawa and T. Miyano, "LSP coding using VQ-SVQ with interpolation in 4.075 kbps M-LCELP speech coding," Proc. Mobile Multimedia Commun. workshop, pp.B2.5-1-2.5-4, Dec. 1993.
- [101] 野村俊之, 芹沢昌宏, 小澤一範, "適応予測を用いた LSP パラメータのベクトル量子化," 音学講論, 1-5-7, pp. 245-246, Oct. 1994.
- [102] 野村俊之, 小澤一範, 芹沢昌宏, "CELP におけるパルス励振源の効率的な探索法," 音学講論, pp. 311-312, March 1996.
- [103] T. Nomura, K. Ozawa, M. Serizawa and M. Iwadare, "Proposal of compression algorithm with rate control for MPEG-4/Audio core experiments," ISO/IEC JTC1/SC29/WG11, MPEG96/1509, November 1996.
- [104] 野村俊之, 芹沢昌宏, 小澤一範, "ビットレート制御可能な MP-CELP 音声符号化方式," 信学春季全大, SD-5-3, Mar. 1997.

- [105] 小田弘, 田中初一, "1次元マルチパルスモデル用修正 AIC の高速計算法," 信学論 (D-II), vol.J76-D-II, no.6, pp.1315-1316, June 1993.
- [106] 大矢智之, 須田博人, 三木俊雄, "Pitch synchronous innovation CELP (PSI-CELP) - PDC ハーフレート音声 CODEC -," 信学技報, RCS93-78, Nov. 1993.
- [107] 大室伸, 守谷健弘, 間野一則, 三樹聡, "移動平均フレーム間予測を用いる LSP パラメータのベクトル量子化," 信学論 (A), vol. J77-A, no. 3, pp. 303-313, March 1994.
- [108] 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の検討," 信学技法, CS82-16, March 1983.
- [109] 小澤一範, 荒関卓, "ピッチ情報を用いたマルチパルス駆動型音声符号化法," 音響学会秋季講論, 2-2-14, Oct. 1983.
- [110] 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の改良と主観評価," 信学春季全大, 2088, 1984.
- [111] 小澤一範, 荒関卓, "波形の周期性に着目したマルチパルス駆動型音声符号化法," 信学秋季全大, 372, 1984.
- [112] 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の品質改善," 信学技報, EA83-62, Jan. 1984.
- [113] 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法におけるパルス探索法の検討," 信学春季全大, pp. 301-302, March 1985.
- [114] 小澤一範, 荒関卓, "ピッチ情報を用いる低ビットレートマルチパルス音声符号化法," 音響学会電音研資, EA85-57, Jan. 1986.
- [115] 小澤一範, 荒関卓, "4.8 kb/s マルチパルス音声符号化法," 音学講論, pp.203-204, Sept. 1985.

- [116] K. Ozawa, S. Ono and T. Araseki, "A Study on pulse search algorithms for multi-pulse excited speech coder realization," IEEE J. Sel. Areas Commun., vol. SAC-4, No. 1, pp.133-141, Feb. 1986.
- [117] K. Ozawa and T. Araseki, "High quality multi-pulse speech coder with pitch prediction," IEEE Proc. ICASSP, pp. 1689-1692, April, 1986.
- [118] K. Ozawa and T. Araseki, "Low bit rate multi-pulse speech coder with natural speech quality," IEEE Proc. ICASSP, pp. 457-460, April 1986.
- [119] 小澤一範, "4.8 kb/s マルチパルス音声符号化法とその音質評価," 信学春季全大, 1401, 1986.
- [120] 小澤一範, 荒関卓, "マルチパルス音源のピッチ補間法に基づく低ビットレート音声符号化法," 信学秋季全大, 187, 1986.
- [121] 小澤一範, "ピッチ予測マルチパルス符号化法の演算量低減化の検討," 信学秋季全大, A-8, 1988.
- [122] 小澤一範, "種々の音源を用いる 4.8 kb/s マルチパルス音声符号化," 音学講論, 3-7-2, March 1989.
- [123] 小澤一範, "種々の音源を用いる 4.8 kb/s マルチパルス音声符号化 (SPMEX)," 信学技報, SP89-2, May 1989.
- [124] 小澤一範, "種々の音源を用いる低ビットレートマルチパルス音声符号化," 信学春季全大, A-1, March 1989.
- [125] 小澤一範, 花田英輔, "改良ピッチ補間を用いる 8 kb/s マルチパルス音声符号化方式の改善," 信学秋季全大, A-1, 1989.
- [126] 小澤一範, "4.8 kb/s SPMEX 方式の音質改善検討," 音学講論, pp. 189-190, Oct. 1989.

- [127] K. Ozawa, "High-quality speech coding using multiple types of excitation signals at 4.8 kb/s and below," IEEE Workshop on Speech Coding for Telecommunications, pp.57-58, 1989.
- [128] 小澤一範, 荒関卓, "ピッチ情報を用いる 9.6 ~ 4.8 kbit/s マルチパルス音声符号化方式," 信学論文誌 (D-II), vol.J72-D-II, no.8, pp.1125-1132, August 1989.
- [129] 小澤一範, "16, 9.6 kbit/s マルチパルス音声符号化方式の符号誤りに対する品質評価," 信学論文誌 (B-II), vol. J72-B-II, no.7, pp.316-322, July, 1989.
- [130] 小澤一範, 熊谷弘己, "マルチパルスと CELP を用いる 3.2 kb/s 音声符号化方式," 信学春季全大, SA-5-5, 1990.
- [131] K. Ozawa, "A hybrid speech coding based on multi-pulse and CELP at 3.2 kb/s," IEEE Proc. ICASSP, pp.677-680, 1990.
- [132] 小澤一範, "応用面からみた音声符号化の現状と将来," 音学講論, pp.179-180, Sept. 1990
- [133] K. Ozawa, T. Miyano and M. Serizawa, "4 kb/s improved CELP coder with efficient vector quantization," IEEE Workshop on Speech Coding for Telecommunications, pp.39-40, 1991.
- [134] K. Ozawa and T. Miyano, "4 kb/s improved CELP coder with efficient vector quantization," IEEE Proc. ICASSP, pp.213-216, 1991.
- [135] 小澤一範, 芹沢昌宏, 宮野俊樹, 野村俊之, "マルチコードブックによる 4 kb/s M-LCELP 音声符号化," 信学秋季全大, A-141, 1993.
- [136] K. Ozawa, M. Serizawa, T. Miyano, T. Nomura, M. Ikekawa and S. Taumi, "M-LCELP speech coding at 4 kb/s with Multi-Mode and Multi-Codebook," IEICE Trans. Commun., vol.E77-B, No.9, pp.1114-1121, Sep. 1994.

- [137] 小澤一範, 田海真一, 熊谷弘巳, 野村俊之, 芹沢昌宏, "13 kb/s 低遅延 M-LCELP 音声符号化方式," 信学春季全大, D-702, Mar. 1995.
- [138] 小澤一範, "移動体通信における音声符号化の現状と将来," 音学講論, pp. 271-272, Sept. 1995.
- [139] 小澤一範, 田海真一, 野村俊之, "マルチパルスベクトル量子化音源と高速探索に基づく MP-CELP 音声符号化," 信学会論文誌 (A), vol.J79-A, no.10, pp.1655-1663, Oct. 1996.
- [140] 小澤一範, 田海真一, 野村俊之, 熊谷弘巳 "マルチパルスベクトル量子化と位置の高速探索に基づく MP-CELP 音声符号化," 音学講論, pp. 261-262, March 1996.
- [141] 小澤一範, "音声符号化技術とその応用," 信学技報, SP95-110, Jan. 1996.
- [142] 小澤一範, "音声の符号化," 共立出版 bit, vol.29, No.2, pp.62-69, Feb. 1997.
- [143] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," IEEE Proc. ICASSP, pp.661-664, 1991.
- [144] D. Pascal, "A subjective pre-selection test plan for the enhanced full rate speech coding," ETSI/GSM SEG-4 (v 0. 0), 1995.
- [145] N. Phamdo, N. Farvardin and T. Moriya, "Combined source-channel coding of LSP parameters using multi-stage vector quantization," Proc. IEEE Workshop on Speech Coding for Telecommunications, pp.36-38, 1991.
- [146] L. R. Rabiner and R. W. Schafer, "Linear predictive coding of speech," Chapter 8, Digital Processing of Speech Signals, Prentice-Hall, 1978.
- [147] R. Salami, C. Laflame and J-P. Adoul, "8 kbit/s ACELP coding of speech with 10 ms speech-frame: A candidate for CCITT standardization," IEEE Proc. ICASSP, pp. 97-100, 1994.

- [148] R. Salami, C. Laflame, J-P. Adoul, A. Kataoka, S. Hayashi, C. Lamblin, D. Massaloux, S. Proust, P. Kroon and Y. Shoham, "Description of the proposed ITU-T 8 kb/s speech coding standard," IEEE Workshop on Speech Coding for Telecommunications, pp. 3-4, Sept. 1995.
- [149] H. Scheffe, "An analysis of variance for paired comparisons," Am. Statis. Assoc., J., pp. 381, 1952.
- [150] M. R. Schroeder, "Vocoders: Analysis and synthesis of speech," Proc. IEEE, vol.54, No.5, pp.720-734, 1966.
- [151] M. R. Schroeder, B. S. Atal and J. L. Hall, "Optimizing digital coders by exploiting masking properties of the human ear," J. Acoust. Soc. Am., vol.66, pp.1647-1652, 1979.
- [152] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," IEEE Proc. ICASSP, pp. 937-940, April, 1985.
- [153] 関根英敏, 高島卓, 浅川吉章, 松田俊幸, 岡本亮, 村松隆二郎, "全検索型パルス音源を用いた CELP 方式," 信学春季全大, SA-5-3, March 1993.
- [154] 芹沢昌宏, 小野茂, 小澤一範, "CELP 音声符号化方式における低ビットレートピッチ伝送方式," 信学春季全大, SA-5-11, March 1993.
- [155] M. Serizawa and A. Gersho, "Joint optimization of LPC and closed-loop pitch parameters in CELP coders," submitted for publication to IEEE Signal Processing Letter, 1996.
- [156] M. Serizawa, "Proposal for mode A wideband speech coding algorithm and its subjective qualification test results," ITU-T D.26, March 1997.
- [157] S. Singhal and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," IEEE Proc. ICASSP, pp. 1.3.1-1.3.4, April 1984.

- [158] S. Singhal, "High quality audio coding using multi-pulse LPC," IEEE Proc. ICASSP, pp.1101-1104, 1990.
- [159] B. Smith, "Instantaneous companding of quantized signals," Bell Systems Technical Journal, vol.36, No.3, pp. 653-709, May 1957.
- [160] 菅村昇, 板倉文忠, "線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮," 信学論 A, J64-A, pp.599-606, August 1981.
- [161] 鈴木政直, 大田恭上, 谷口智彦, 田中良紀, "モード切替えを適用した低ビットレート音声符号化方式," 第10回デジタル信号処理シンポジウム, pp.123-128, Nov. 1995.
- [162] J. Takizawa and K. Ozawa, "2.4 kbps M-LCELP speech coding with psychoacoustic measure," IEEE Workshop on Speech Coding for Telecommunication, pp. 109-110, 1995.
- [163] 田中直也, 板倉文忠, "CELP 音声符号化方式におけるパルス音源導入による音声品質の向上," 信学技報, EA92-24, 1992.
- [164] T. Taniguchi, S. Unagami and R.M.Gray, "Multimode coding: application to CELP," IEEE Proc. ICASSP, pp.156-159, 1989.
- [165] 田海真一, 芹沢昌宏, 宮野俊樹, 小澤一範, "M-LCELP における音源コードブックの比較評価," 音学講論, pp.263-264, March 1994.
- [166] 田海真一, 小澤一範, 芹沢昌宏, 野村俊之, "13 kbps 低遅延 M-LCELP 音声符号化方式," 音学講論, pp.247-248, March 1995.
- [167] 田海真一, 小澤一範, 野村俊之, "13 kbps 低遅延 M-LCELP 音声符号化方式における伝送路誤り特性," 音学講論, pp. 291-292, Sept. 1995.
- [168] TIA, "Half-rate speech codec test plan v6.0," TR45.3.5/93.05.19.01, 1993.

- [169] 友竹世光, 薄克典, 中川博明, 和気靖浩, 安永智, 芹沢昌宏, "8 kb/s LCELP の実時間ハードウェア," 信学秋季全大, B-570, 1992.
- [170] 海野義博, 小澤一範, "マルチパルス音声符号化法の伝送路誤り発生時の特性と評価," 信学春季全大, A-25, 1988.
- [171] Y. Unno, N. Shigemoto and K. Ozawa, "16 kbps multi-pulse speech codec for digital mobile radio," NEC Research and Development, vol. 30, No. 1, pp.39-44, Jan. 1989.
- [172] Y. Unno, M. Nakamura, T. Sato, T. Miyano and K. Ozawa, "Learned codebook excited linear predictive (LCELP) speech codec for digital cellular system," NEC Research and Development, vol. 32, No. 4, pp.549-556, Oct. 1991.
- [173] 薄克典, 荒田勝利, 和気靖浩, "高能率音声符号化装置ユニバーサル・トランスコーダ (UTRC mkII)," 日本電気エンジニアリング NEEC Review, no.16, 1991.
- [174] A. Varga and F. Fallside, "A technique for using multipulse linear predictive speech synthesis in text-to-speech-type systems," IEEE Trans. ASSP, vol. ASSP-35, no. 4, pp.586-587, April 1987.
- [175] Y. Wake, S. Tanaka, K. Ozawa and T. Araseki, "A multi-pulse LPC speech coder using digital signal processors," IEEE Proc. ICASSP, p.1429-1432, March 1985.
- [176] S. Wang and A. Gersho, "Phonetically-based vector excitation coding of speech at 3.6 kbps," IEEE Proc. ICASSP, pp. 49-52, 1989.
- [177] S. Wang and A. Gersho, "Improved excitation for phonetically-segmented VXC speech coding below 4 kb/s," IEEE Proc. GLOBECOM, pp.946-950, 1990.
- [178] S. Wang, A. Sekey and A. Gersho, "An objective measure for predicting subjective quality of speech coders," IEEE J. Sel. Areas Commun., vol. 10, No. 5, pp.819-829, June 1992.

- [179] 渡辺俊朗, 伊藤憲三, 北脇信彦, “種々の評価尺度による符号化音声品質の比較,” 日本音響学会音声研資, S82-48, Nov. 1982.
- [180] S. G. Wilson and S. Husain, “Adaptive tree encoding of speech at 8000 bps with a frequency-weighted error criterion,” IEEE Trans. Commun., vol. COM-27, pp. 165-170, Jan. 1979.
- [181] 八幡, 伊福部, 星宮, “音声の自然性における位相と波形ゆらぎの役割 - 電気式人工喉頭の基礎研究,” 音学会聴覚研資, H-87-5, 1987 年
- [182] Y. Yasuda, Y. Hirata, K. Nakamura and S. Ohtani, “Development of variable-rate viterbi decoder and its performance characteristics,” Proc. Digisat, XII-24-XII-31, 1983.
- [183] 安永智, 和気靖浩, 田中俊二, 佃井, “マルチパルス音声符号化装置ネオコーダMP,” 信学春季全大, 2000, March 1986.
- [184] 吉田幸司, 吉田正, 本間光一, “低ビットレート CELP 符号化へのパルス音源探索の適用,” 信学技報, SP91-68, 1991.

## 著者業績

## 論文

- 1 K. Ozawa, S. Ono and T. Araseki, “A Study on pulse search algorithms for multi-pulse excited speech coder realization,” IEEE J. Sel. Areas Commun., vol. SAC-4, No. 1, pp.133-141, Feb. 1986.
- 2 小澤一範, “16, 9.6 kbit/s マルチパルス音声符号化方式の符号誤りに対する品質評価,” 電子情報通信学会論文誌 (B-II), vol. J72-B-II, no.7, pp.316-322, July 1989.
- 3 小澤一範, 荒関卓, “ビット情報を用いる 9.6 ~ 4.8 kbit/s マルチパルス音声符号化方式,” 電子情報通信学会論文誌 (D-II), vol.J72-D-II, no.8, pp.1125-1132, August 1989.
- 4 K. Ozawa and J. S. Logan, “Perceptual evaluation of two speech coding methods by native and non-native speakers of English,” Computer Speech and Language, vol. 3, pp. 53-59, 1989.
- 5 K. Ozawa, M. Serizawa, T. Miyano, T. Nomura, M. Ikekawa and S. Taumi, “M-LCELP speech coding at 4 kb/s with Multi-Mode and Multi-Codebook,” IEICE Trans. Commun., vol.E77-B, No.9, pp.1114-1121, Sep. 1994.
- 6 小澤一範, 田海真一, 野村俊之, “マルチパルスベクトル量子化音源と高速探索に基づく MP-CELP 音声符号化,” 電子情報通信学会論文誌 (A), vol.J79-A, no.10, pp.1655-1663, Oct. 1996.

## 国際会議

- 1 K. Ozawa, T. Araseki and Y. Itoh, "An adaptive echo canceller using digital signal processor LSI chips," IEEE Proc. ICASSP, pp. 466-469, April 1983.
- 2 K. Ozawa and T. Araseki, "High quality multi-pulse speech coder with pitch prediction," IEEE Proc. ICASSP, pp. 1689-1692, April 1986.
- 3 K. Ozawa and T. Araseki, "Low bit rate multi-pulse speech coder with natural speech quality," IEEE Proc. ICASSP, pp. 457-460, April 1986.
- 4 K. Ozawa, "High-quality speech coding using multiple types of excitation signals at 4.8 kb/s and below," IEEE Workshop on Speech Coding for Telecommunications, pp.57-58, 1989.
- 5 K. Ozawa, "A 4.8 kb/s high-quality speech coding using multiple types excitation signals," Proc. EUROSPEECH, pp.306-309, 1989.
- 6 K. Ozawa, "A hybrid speech coding based on multi-pulse and CELP at 3.2 kb/s," IEEE Proc. ICASSP, pp.677-680, 1990.
- 7 K. Ozawa, T. Miyano and M. Serizawa, "4 kb/s improved CELP coder with efficient vector quantization," IEEE Workshop on Speech Coding for Telecommunications, pp.39-40, 1991.
- 8 K. Ozawa and T. Miyano, "4 kb/s improved CELP coder with efficient vector quantization," IEEE Proc. ICASSP, pp.213-216, 1991.
- 9 K. Ozawa, M. Serizawa, T. Miyano and T. Nomura, "M-LCELP speech coding at bit-rates below 4 kb/s," Proc. EUROSPEECH, pp. 51-54, 1993.
- 10 K. Ozawa, M. Serizawa, T. Miyano and T. Nomura, "M-LCELP speech coding at 4 kb/s," IEEE Proc. ICASSP, pp.269-272, 1994.

## 研究会

- 1 小澤一範, 荒関卓, "シグナルプロセッサを用いたエコーキャンセラ," 電子通信学会通信方式研究会, CS82-65, Sept. 1982.
- 2 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の検討," 電子通信学会通信方式研究会, CS82-161, March 1983.
- 3 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の品質改善," 電子通信学会電気音響研究会, EA83-62, Jan. 1984.
- 4 小澤一範, 荒関卓, "ピッチ情報を用いる低ビットレートマルチパルス音声符号化法," 電子通信学会電気音響研究会, EA85-57, Jan. 1986.
- 5 小澤一範, "種々の音源を用いる 4.8 kb/s マルチパルス音声符号化 (SPMEX)," 電子情報通信学会音声研究会, SP89-2, May 1989.
- 6 小澤一範, "音声符号化技術とその応用," 電子情報通信学会音声研究会, SP95-110, Jan. 1996.

## 講演論文

- 1 小澤一範, 荒関卓, "ピッチ情報を用いたマルチパルス駆動型音声符号化法," 音響学会秋季研究発表会, pp. 159-160, Oct. 1983.
- 2 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法の改良と主観評価," 電子通信学会春季全国大会, 2088, 1984.
- 3 小澤一範, 荒関卓, "波形の周期性に着目したマルチパルス駆動型音声符号化法," 電子通信学会秋季全国大会, 372, 1984.
- 4 小澤一範, 小野茂, 荒関卓, "マルチパルス駆動型音声符号化法におけるパルス探索法の検討," 電子通信学会春季全国大会, pp.301-302, March 1985.

- 5 小澤一範, 荒関卓, “4.8 kb/s マルチパルス音声符号化法,” 音響学会秋季研究発表会, pp.203-204, Sept. 1985.
- 6 小澤一範, “4.8 kb/s マルチパルス音声符号化法とその音質評価,” 電子通信学会春季全国大会, 1401, 1986.
- 7 小澤一範, “16 k ビット / 秒以下でも音質の劣化が少ないマルチパルス音声符号化法を用いた実時間コーデック,” 日経エレクトロニクス, no. 397, pp. 185-215, June 1986.
- 8 小澤一範, 荒関卓, “マルチパルス音源のピッチ補間法に基づく低ビットレート音声符号化法,” 電子通信学会秋季全国大会, 187, 1986.
- 9 小澤一範, “ピッチ予測マルチパルス符号化法の演算量低減化の検討,” 電子情報通信学会秋季全国大会, A-8, 1988.
- 10 小澤一範, J. S. Logan, “英語の母語話者と非母語話者による符号化音声の知覚,” 音響学会春季研究発表会, pp. 293-294, March 1988.
- 11 小澤一範, J. S. Logan, “心理的負荷を用いる符号化音声の評価,” 音響学会春季研究発表会, pp. 337-338, March 1989.
- 12 小澤一範, “種々の音源を用いる 4.8 kb/s マルチパルス音声符号化,” 音響学会春季研究発表会, 3-7-2, March, 1989.
- 13 小澤一範, “種々の音源を用いる低ビットレートマルチパルス音声符号化,” 電子情報通信学会春季全国大会, A-1, March 1989.
- 14 小澤一範, “4.8 kb/s SPMEEX 方式の音質改善検討,” 音響学会秋季研究発表会, pp. 189-190, Oct. 1989.
- 15 小澤一範, 花田英輔, “改良ピッチ補間を用いる 8 kb/s マルチパルス音声符号化方式の改善,” 電子情報通信学会秋季全国大会, A-1, 1989.

- 16 小澤一範, 熊谷弘己, “マルチパルスと CELP を用いる 3.2 kb/s 音声符号化方式,” 電子情報通信学会春季全国大会, SA-5-5, 1990.
- 17 小澤一範, “LSP パラメータの効率的なベクトル量子化法の検討,” 電子情報通信学会秋季全国大会, A-130, 1990.
- 18 小澤一範, “応用面からみた音声符号化の現状と将来,” 音響学会秋季研究発表会, pp.179-180, Sept. 1990.
- 19 小澤一範, 芹沢昌宏, 宮野俊樹, 野村俊之, “マルチコードブックによる 4 kb/s M-LCELP 音声符号化,” 電子情報通信学会秋季全国大会, A-141, 1993.
- 20 小澤一範, 田海真一, 熊谷弘己, 野村俊之, 芹沢昌宏, “13 kb/s 低遅延 M-LCELP 音声符号化方式,” 電子情報通信学会春季全国大会, D-702, Mar. 1995.
- 21 小澤一範, 熊谷弘己, “LSP 量子化候補と適応コードブックラグの組合せ探索に基づく 13 kb/s 低遅延 M-LCELP 音声符号化,” 電子情報通信学会秋季全国大会, D-249, Sept. 1995.
- 22 小澤一範, “移動体通信における音声符号化の現状と将来,” 音響学会秋季研究発表会, pp. 271-272, Sept. 1995.
- 23 小澤一範, 田海真一, 野村俊之, 熊谷弘己, “マルチパルスベクトル量子化と位置の高速探索に基づく MP-CELP 音声符号化,” 音響学会春季研究発表会, pp. 261-262, March 1996.
- 24 小澤一範, 田海真一, 野村俊之, 芹沢昌宏, 熊谷弘己, “MP-CELP 音声符号化のビットレート低減化の検討,” 電子情報通信学会秋季全国大会, D-467, Sept. 1996.

## 著書

- 1 “デジタル移動体通信のための高能率音声符号化技術,” トリケップス刊, 1992.

- 2 “ハーフレート (5.6 kb/s) 音声符号化,” 安田浩 編著, MPEG/ マルチメディア符号化の国際標準, 第 7 章, 丸善 (株), 1994.

## 略語一覧

A-b-S	Analysis-by-Synthesis (合成による分析)
ACELP	Algebraic Code Excited Linear Prediction (代数構造符号励振型線形予測符号化)
ACR	Absolute Category Rating (絶対品質尺度)
ADPCM	Adaptive Differential Pulse Code Modulation (適応差分符号化)
APC	Adaptive Predictive Coding (適応予測符号化)
APC-AB	APC with Adaptive Bit Allocation (適応ビット割り当て適応予測符号化)
ARIB	Association of Radio Industries and Businesses (日本電波産業ビジネス協会)
ARMA	Auto-Regressive and Moving-Average (自己回帰移動平均過程)
BCH	Bose-Chaudhuri-Hocquenghem (BCH 符号)
CCR	Comparison Category Rating (比較品質尺度)
CD	Cepstrum Distance (ケプストラム距離)
CELP	Code Excited Linear Prediction (符号励振型線形予測符号化)
CS-ACELP	Conjugate Structure ACELP (共役構造 ACELP)
DCR	Degradation Category Rating (妨害品質尺度)
DSP	Digital Signal Processor (信号処理プロセッサ)
EFR	Enhanced Full-Rate (音質改善型フルレート)
ETSI	European Telecommunications Standards Institute (欧州電気通信標準化協会)
FEC	Forward Error Correction (前方向誤り訂正)
FPLMTS	Future Public Land Mobile Telecommunication Systems (次世代公衆陸上移動体通信システム)
GSM	Global System for Mobile Communication (欧州移動体通信標準方式)
IRS	Intermediate Reference System (中間標準系)
ISDN	Integrated System Digital Network (デジタル統合ネットワークシステム)
ISO	International Standardization Organization (国際標準化機構)
ITU-T	International Telecommunications Union-Telecommunications Standards Sector (国際電気通信連合電気通信標準化部門)



LD-CELP	Low Delay CELP (低遅延 CELP)
LES ACR	Listening Effort Scale ACR (聴取努力評価尺度)
LPC	Linear Predictive Coding (線形予測分析)
LSP	Line Spectrum Pair (線スペクトル対)
LTP	Long Term Predictor (長期予測器)
MA	Moving-Average (移動平均過程)
MIPS	Million Instruction Per Second (百万命令回数 / 秒)
M-LCELP	Multi-mode Learned CELP (マルチモード学習符号帳 CELP)
MMM	MultiMedia Mux (マルチメディア多重化装置)
MLD	Majority Logic Detection (多数決判別符号)
MNRU	Modulated Noise Reference Unit (変調された雑音による基準単位)
MOPS	Million Operation Per Second (百万処理回数 / 秒)
MOS	Mean Opinion Score (平均オピニオン値)
MP-CELP	Multi-Pulse based CELP (マルチパルスベース CELP)
MPEG	Moving Picture coding Expert Group (動画画像専門家グループ) 最近は Coding of Moving Pictures and Audio に改称
MPELP	Multi-Pulse Excited Linear Prediction (マルチパルス駆動型線形予測符号化)
MP-MLQ	Multi-Pulse Maximum Likelihood Quantization (マルチパルス最尤量子化)
PARCOR	Partial Autocorrelation (偏自己相関関数)
PCM	Pulse Code Modulation (パルス符号変調)
PCS	Personal Communications Systems (パーソナル通信システム)
PDC	Personal Digital Cellular (日本セルラー電話システム)
PHS	Personal Handyphone System (パーソナルハンディフォン)
PSI-CELP	Pitch Synchronous Innovation CELP (ピッチ同期励振 CELP)
PSMP	Pitch Synchronous Multi-Pulse (ピッチ同期マルチパルス)
PWI	Pitch Waveform Interpolation (ピッチ波形補間)
QPSK	Quadrature Phase Shift Keying (直交位相シフトキー方式)
RMS	Root Mean Square (実効値)
RPE-LTP	Regular Pulse Excitation with Long-Term Prediction (長期予測つきレギュラーパルス駆動音声符号化)
RS	Reed-Solomon (リードソロモン符号)

SC	Stochastic Codebook (統計符号帳)
$SNR_{seg}$	Segmental Signal to Noise Ratio (セグメンタル信号対雑音比)
SQ	Scalar Quantization (スカラー量子化)
STP	Short Term Predictor (短期予測器)
SVQ	Split Vector Quantization (分割ベクトル量子化)
TIA	Telecommunications Industry Association (北米電気通信産業協会)
UMTS	Universal Mobile Telecommunications System (世界移動体通信システム)
VQ	Vector Quantization (ベクトル量子化)
VSELP	Vector Sum Excited Linear Prediction (ベクトル加算励振型線形予測符号化)



