

JT-G722.2

適応マルチレート広帯域 (AMR-WB) 方式を用いた 16 kbit/s 程度の広帯域 音声符号化

WIDEBAND CODING OF SPEECH AT AROUND 16
KBIT/S USING ADAPTIVE MULTI-RATE WIDEBAND
(AMR-WB)

第 3.3 版

2007 年 5 月 31 日制定

社団法人
情報通信技術委員会

THE TELECOMMUNICATION TECHNOLOGY COMMITTEE

本書は、（社）情報通信技術委員会が著作権を保有しています。

内容の一部又は全部を（社）情報通信技術委員会の許諾を得ることなく複製、転載、改変、転用及びネットワーク上での送信、配布を行うことを禁止します。

目 次

<参考>	7
1. 本標準の規定範囲	9
2. 参照している勧告、標準等	9
3. 定義、記号、略語	10
3.1 定義	10
3.2 記号	11
3.3 略語	14
4. 概要	14
4.1 オーディオ部分の機能記述	15
4.2 音声サンプルの準備	15
4.3 適応マルチレート広帯域音声符号器の原理	15
4.4 適応マルチレート広帯域音声復号器の原理	20
4.5 符号化パラメータのシーケンスおよび主観的重要度	20
5. 符号器の詳細	22
5.1 前処理	22
5.2 線形予測分析と量子化	22
5.2.1 窓掛けおよび自己相関の算出	23
5.2.2 窓掛けおよび自己相関の算出	23
5.2.3 LPからISPへの変換	23
5.2.4 ISPからLPへの変換	25
5.2.5 ISP係数の量子化	25
5.2.6 ISPの補間	27
5.3 聴覚重み付け	27
5.4 開ループピッチ分析	27
5.4.1 6.60 kbit/sモード	28
5.4.2 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 および 23.85 kbit/sモード	29
5.5 インパルス応答の計算	29
5.6 ターゲット信号の計算	30
5.7 適応コードブック	30
5.8 代数的コードブック	32
5.8.1 コードブックの構成	32
5.8.1.1 23.85 および 23.05 kbit/sモード	32
5.8.1.2 19.85 kbit/s モード	33
5.8.1.3 18.25 kbit/sモード	33
5.8.1.4 15.85 kbit/sモード	34
5.8.1.5 14.25 kbit/sモード	34
5.8.1.6 12.65 kbit/sモード	35
5.8.1.7 8.85 kbit/sモード	35
5.8.1.8 6.60 kbit/sモード	35
5.8.2 パルスのインデックス化	36
5.8.3 コードブック探索	40

5. 9	適応および固定コードブック利得の量子化	44
5. 10	メモリ更新	45
5. 11	高域利得生成	45
6.	復号器の詳細	46
6. 1	復号と音声合成	46
6. 2	高域通過フィルタ、アップスケーリングと補間	49
6. 3	高周波数帯域	49
6. 3. 1	高域励振信号の生成	49
6. 3. 2	高周波数帯域のためのLPフィルタ	50
6. 3. 3	高域合成	51
7.	適応マルチレート広帯域コーデックの詳細なビット割当て	51
8.	ホーミングシーケンス	57
8. 1	機能説明	57
8. 2	定義	57
8. 3	符号器ホーミング	57
8. 4	復号器ホーミング	58
9.	音声検出(VAD)	59
9. 1	VAD記号	59
9. 1. 1	VAD変数	59
9. 1. 2	VAD定数	59
9. 1. 3	関数	60
9. 2	機能説明	60
9. 2. 1	フィルタバンクとサブバンドレベル計算	61
9. 2. 2	トーン検出	64
9. 2. 3	VAD判定	64
9. 2. 3. 1	ハングオーバー付加	65
9. 2. 3. 2	背景雑音推定	66
9. 2. 3. 3	音声レベル推定	68
10.	3GPP音声電話サービスにおける必須なAMR-WBコーデックモード	69
11.	参考文献(informative)	70
付属資料A 擬似背景雑音特性		
A. 1	本付属資料の規定範囲	71
A. 2	定義、記号、略語	71
A. 2. 1	定義	71
A. 2. 2	記号	71
A. 2. 3	略語	72
A. 3	概要	72
A. 4	送信側の機能	72
A. 4. 1	ISF算出	73
A. 4. 2	フレームエネルギー計算	74
A. 4. 3	背景雑音の変化と安定性の解析	74
A. 4. 4	SIDフレーム生成時の音声符号化アルゴリズムの修正	75

A. 4. 5	S I Dフレーム符号化	75
A. 5	受信側の機能	75
A. 5. 1	LPおよびエネルギーパラメータの平均化および復号	76
A. 5. 2	擬似背景雑音の生成および更新	77
A. 6	計算の詳細およびビット割り当て	77
付属資料B	ソースコントロールレート動作	78
B. 1	本付属資料の規定範囲	78
B. 2	定義、記号、略語	78
B. 2. 1	定義	78
B. 2. 2	記号	78
B. 2. 3	略語	78
B. 3	概要	79
B. 3. 1	構成概要	79
B. 4	AMR-WB S C R動作	79
B. 4. 1	送信 (T X) 側	79
B. 4. 1. 1	動作概要	80
B. 4. 1. 2	送信S C Rハンドラ機能	81
B. 4. 1. 3	ANの送信部	82
B. 4. 2	受信 (R X) 側	82
B. 4. 2. 1	動作概要	83
B. 4. 2. 2	受信S C Rハンドラへの要求	83
B. 4. 3	AMR-WB S I D情報フォーマット	84
付属資料C	固定小数点Cコード	85
C. 1	Cソースコードの構成	85
C. 1. 1	Cソースコードの内容	85
C. 1. 2	プログラムの実行	85
C. 1. 3	コード階層	85
C. 1. 4	変数、定数、テーブル	89
C. 1. 4. 1	Cコードで用いられている定数の詳細	90
C. 1. 4. 2	Cコードで用いられている固定的なテーブルの詳細	90
C. 1. 4. 3	Cコードで用いられるスタティック変数	92
C. 2	ホーミング処理	94
C. 3	ファイルフォーマット	94
C. 3. 1	音声ファイル (符号器の入力/復号器の出力)	94
C. 3. 2	モード制御ファイル (符号器の入力)	95
C. 3. 3	パラメータビットストリームファイル (符号器出力/復号器入力)	95
付属資料D	デジタルテストシーケンス	98
D. 1	本付属資料の規定範囲	98
D. 2	概要	98
D. 3	テストシーケンスのフォーマット	98

D. 3. 1	ファイルフォーマット	98
D. 3. 2	コーデックホーミング	98
D. 4	音声コーデックのテストシーケンス	99
D. 4. 1	コーデック構成	99
D. 4. 2	音声コーデックのテストシーケンス	99
D. 4. 2. 1	音声符号器のテストシーケンス	99
D. 4. 2. 2	音声復号器のテストシーケンス	100
D. 4. 2. 3	コーデックホーミングシーケンス	100
D. 5	ソースコントロールレート動作用テストシーケンス	101
D. 5. 1	コーデック構成	101
D. 5. 2	テストシーケンス	101
D. 5. 2. 1	背景雑音推定用テストシーケンス	101
D. 5. 2. 2	トーン信号検出用テストシーケンス	101
D. 5. 2. 3	実音声およびトーン	101
D. 5. 2. 4	信号対雑音比推定用テストシーケンス	101
D. 6	適応マルチレート音声符号器の 20ms フレーミング検出用シーケンス	102
D. 6. 1	ビット同期	102
D. 6. 2	フレーム同期	102
D. 6. 3	同期シーケンスのフォーマットおよびサイズ	103
付属資料 E	フレーム構造	104
E. 1	本付属資料の規定範囲	104
E. 2	定義と略語	104
E. 2. 1	定義	104
E. 2. 2	略語	104
E. 3	AMR-WB コーデックインタフェースフォーマット 1 (AMR-WB I F 1)	104
E. 3. 1	AMR-WB ヘッダと AMR-WB 補助情報	105
E. 3. 1. 1	フレームタイプ、モード表示とモード要求	105
E. 3. 1. 2	フレーム品質表示	106
E. 3. 1. 3	TX_TYPE と RX_TYPE へのマッピング	106
E. 3. 1. 4	コーデック CRC	106
E. 3. 2	AMR-WB コアフレーム	106
E. 3. 2. 1	音声ビットの AMR-WB コアフレーム：ビット順序	107
E. 3. 2. 2	音声ビットの AMR-WB コアフレーム：クラス分け	107
E. 3. 2. 3	擬似背景雑音ビットの AMR-WB コアフレーム	108
E. 3. 3	汎用 AMR-WB フレーム構成	108
E. 4	オクテット整合と AMR-WB インタフェースフォーマット 2 (AMR-WB I F 2)	111
E. 5	AMR-WB コアフレームビット順序表	114
付属資料 F	TTC 標準 JT-H 2 4 5 における AMR-WB 使用	121
F. 1	概要	121
F. 2	参考文献	121
F. 3	序論	121

F. 4	TTC標準JT-H245のためのTTC標準JT-G722.2本体(適応マルチレート広帯域)の能力定義	121
付録1	誤りフレームあるいは損失フレームの誤り補償	125
1. 1	本付録の規定範囲	125
1. 2	定義、略語	125
1. 2. 1	定義	125
1. 2. 2	略語	125
1. 3	概要	125
1. 4	要求	126
1. 4. 1	誤り検出	126
1. 4. 2	誤りあるいは損失音声フレーム	126
1. 4. 3	最初の損失SIDフレーム	126
1. 4. 4	後続の損失SIDフレーム	126
1. 5	ECU/BFHの解決例	126
1. 5. 1	ステートマシン	126
1. 5. 2	誤り/損失音声フレームの置換とミュートイング	127
1. 5. 2. 1	BFI=0、prevBFI=0、State=0あるいは1	127
1. 5. 2. 2	BFI=0、prevBFI=1、State=0から3	128
1. 5. 2. 3	BFI=1、prevBFI=0または1、State=1から6	128
1. 5. 2. 4	励振系列	132
1. 5. 2. 5	高域利得(23.85 kbit/sモード)	132
1. 5. 3	損失SIDフレームの置換とミュートイング	132
付録	用語対照表	133
付録	用語解説	136

<参考>

1. 国際勧告等との関連

本標準は、2001年11月にITU-T SG16でAAPに進むことが合意され、2002年1月に承認されたITU-T勧告G.722.2に準拠したものである。

本標準の付属資料A、B、C、D、Eおよび付録1は、2001年11月にITU-T SG16でAAPに進むことが合意され、2002年1月に承認されたITU-T勧告G.722.2 Annex A、B、C、D、EおよびAppendix Iに準拠したものである。

本標準の付属資料Fは、2002年10月にITU-T SG16でAAPに進むことが合意され、2002年11月承認されたITU-T勧告G.722.2 Annex Fに準拠したものである。

また、本標準および付録1は、2002年10月に承認されたITU-T勧告G.722.2に対するインプリメンターズガイドに準拠して改定されたものである。

本標準本体および付属資料Cは、2003年5月にITU-T SG16でAAPに進むことが合意され、2003年7月に承認されたITU-T勧告G.722.2（改定版）に準拠したものである。なお、ITU-T勧告G.722.2（改定版）の改定箇所のうち、Annex C.1.1およびC.2をはじめとする複数の箇所は、TTCからITU-Tへの提案により改定されたものであり、JT-G722.2第1版および第2版において、該当箇所の改定は既に反映済みである。

また、本標準は、2004年1月にITU-T SG16でAAPに進むことが合意され、2004年3月に承認された、ITU-T勧告G.722.2 Annex Cに対するCorrigendum 2に準拠して改定されたものである。

また、本標準は、2005年8月にITU-T SG16でAAPに進むことが合意され、2005年9月に承認された、ITU-T勧告G.722.2に対するCorrigendumに準拠して改定されたものである。

また、本標準は、2006年11月にITU-T SG16でAAPに進むことが合意され、2007年1月に承認された、ITU-T勧告G.722.2に対するCorrigendum2に準拠して改定されたものである。

2. 上記国際勧告等に対する追加項目等

2.1 オプション選択項目

なし

2.2 ナショナルマター決定項目

なし

2.3 その他

- (1) 本標準は、上記ITU-T勧告に対し、先行している項目はない。
- (2) 本標準は、上記ITU-T勧告に対し、追加した項目はない。
- (3) 本標準は、上記ITU-T勧告に対し、削除した項目はない。
- (4) 本標準は、上記ITU-T勧告に対し、変更した項目はない。

2.4 原勧告との章立て構成比較

上記国際勧告等との章立て構成の相違はない。

3. 改版の履歴

版 数	制 定 日	改 版 内 容
第1版	2002年11月28日	制定
第2版	2003年4月23日	付属資料A、B、C、D、E、Fおよび付録1の追加 本文の修正
第3版	2004年4月20日	I T U - T 勧告G. 7 2 2. 2の改定に伴う、標準本体および 付属資料Cの改定
第3.1版	2004年8月26日	Cコード改定に関する記述の追加
第3.2版	2006年2月28日	I T U - T 勧告G. 7 2 2. 2の改定に伴う標準本体の改定お よびCコード改定に関する記述の追加
第3.3版	2007年5月31日	Cコード改定に関する記述の追加

4. その他

(1) 参照している勧告、標準等

TTC標準： J T - G 7 2 2

I E T F : R F C 3 2 6 7

(2) TTC標準J T - G 7 2 2. 2は、I T U - T 勧告G. 7 2 2. 2に準拠しており、本標準中で言及しているCコードおよびテストシーケンスとは、I T U - T 勧告G. 7 2 2. 2のものをさし、日本I T U協会から入手可能である。

なお、2003年7月に承認されたI T U - T 勧告G. 7 2 2. 2の改定に伴い、Cコードの改定も行われている。

更に、2004年3月に再度のCコードの改定が行われており、2004年3月に承認されたI T U - T 勧告G. 7 2 2. 2 Annex Cに対するCorrigendum 2にその記述がある。

また、更に、2005年9月に承認されたI T U - T 勧告G. 7 2 2. 2の改定に伴い、Cコードの再度の改定も行われている。

更に、2007年1月に再度のCコードの改定が行われており、2007年1月に承認されたI T U - T 勧告G. 7 2 2. 2に対するCorrigendum 2にその記述がある。

1. 本標準の規定範囲

本標準は、16ビット均一PCMフォーマットで入力される320サンプルの音声ブロックから、132, 177, 253, 285, 317, 365, 397, 461 および 477 ビットの符号ブロックへ、そして132, 177, 253, 285, 317, 365, 397, 461 および 477 ビットの符号ブロックから再構成される320サンプルの出力音声ブロックへの、詳細なマッピングについて記述するものである。標準化周波数は16kHzであり、符号化ビット列のビットレートは、6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 および 23.85 kbit/s となる。マルチレート符号化モードの符号化手順は、いわゆる代数的符号励振線形予測符号化方式であり、以降ACELPと略称する。このマルチレートの広帯域ACELP符号器はAMR-WBと呼ばれている。本標準に記述されるコーデックでは、統合化された音声検出器(VAD)も利用されている。

本標準で想定するアプリケーションは、以下のようなものである：ボイスオーバーIP(VoIP)とインターネットアプリケーション、モバイルコミュニケーション、PSTNアプリケーション、ISDN広帯域電話、ISDNテレビ電話およびテレビ会議である。

TTC標準JT-G722.2本体で規定されるアルゴリズムに加えて、TTC標準JT-G722.2付属資料A、B、および付録1では、GSMおよび3GPPの無線システムとの互換性を可能にする補助的な機能を提供する。これらの機能は、本来それらのシステムのために開発されたものであるが、その利用については、モバイルアプリケーションに限定されない。TTC標準JT-G722.2付属資料DおよびEでは、それぞれテストベクトルおよびフレーム構造を記述する。これらの付属資料は、AMR-WBアルゴリズムを装備しているシステムの要求の違いにしたがって、本体の規定とは独立に実装してもよい。

- ・TTC標準JT-G722.2付属資料Aでは、ソースコントロールレート動作におけるAMR-WBアルゴリズムを利用するための擬似背景雑音の特性について記述する。この付属資料の実装は、GSMおよび3GPP無線システムとの互換性のために不可欠である。
- ・TTC標準JT-G722.2付属資料Bでは、AMR-WBのためのソースコントロールレート動作について記述する。この付属資料の実装は、GSMおよび3GPP無線システムとの互換性のために不可欠である。
- ・TTC標準JT-G722.2付属資料Dでは、デジタルテストシーケンスを記述する。これはAMR-WBコーデックの実装の正しさの指標を提供するための検証ツールである。
- ・TTC標準JT-G722.2付属資料Eでは、AMR-WBアルゴリズムの様々な動作モードで使用するための推奨フレーム構造について記述する。
- ・TTC標準JT-G722.2付録1では、AMR-WBの誤りあるいは消失したフレームの誤り補償のための、解決法の一例を記述する。

利便性のため、これらすべての機能を低レベルで記述するANSI Cコードが、TTC標準JT-G722.2付属資料Cという単一の付属資料に集められている。TTC標準JT-G722.2の記述と、TTC標準JT-G722.2付属資料Cの記述による実装との間に不一致があった場合には、いかなる場合においてもTTC標準JT-G722.2付属資料Cの記述が優先される。

第8章では、コーデックホーミングと呼ばれる、特有のリセット手順について記述する。これはコーデックをある既知の初期状態に導くのに便利な機能である(例えばテスト用途など)。第9章では、コーデックに用いられる音声検出器(VAD)について規定する。これはTTC標準JT-G722.2付属資料Bで規定されるソースコントロールレート動作(DTX)と同様に、本コーデックで用いられる。

第10章では、3GPP音声電話サービスにおけるAMR-WBをサポートするための最小要求条件に関する情報を提供する。

2. 参照している勧告、標準等

以下のTTC標準およびその他の標準は、本書での参照を通して、本標準の規定を構成する規定を含んでい

る。出版時においては下記の版が有効であった。すべての勧告およびその他の標準は改版される可能性がある。それゆえ本標準のすべての利用者は、下記に示された勧告および標準の最新版を適用する可能性を調べることが望ましい。現在有効なTTC標準のリストは定期的に発行されている。

[1] JT-G 7 2 2 64kbit/s 以下の7kHz オーディオ符号化方式

[2] RFC 3267 “A Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs”, June 2002.

3. 定義、記号、略語

3. 1 定義

本標準では、以下の定義を適用する。

適応コードブック： 適応コードブックは、サブフレームごとに適応される励振ベクトルからなる。適応コードブックは、長期フィルタの状態から得られる。そのラグの値が適応コードブックへのインデックスとみなされる。

代数的コードブック： 代数的コードが励振ベクトルを構成するために用いられる、固定のコードブック。励振ベクトルは、あらかじめ定義された組み合わせの位置の組をもつ、少数の非零パルスからなる。 k 番目の励振コードベクトルのそれぞれのパルスの振幅と位置は、物理ストレージを必要としないか、最小の物理ストレージに抑えるというルールを通して、インデックス k から導かれる。インデックスから関連するコードベクトルへのパスがルックアップテーブルを伴う確率的コードブックを用いる場合とは対照的である。

非スパース化処理： スパースな固定コードブックベクトルから聴覚的な不自然さを減少させるために、固定コードブックに適用される適応後処理手順。

閉ループピッチ解析： これが適応コードブック探索である。すなわち、重み付けされた入力音声と長期フィルタ状態から、ピッチ（ラグ）の値を推定する処理である。閉ループ探索では、ラグは誤差最小化ループ（合成による分析、A-b-S法）を用いて探索される。適応マルチレート広帯域コーデックにおいては、閉ループピッチ探索はサブフレームごとに実行される。

直接型係数： 短期フィルタのパラメータを格納するフォーマットの1つ。適応マルチレート広帯域コーデックにおいて、音声サンプルの補正に用いられるすべてのフィルタは、直接型係数を使用する。

固定コードブック： 固定コードブックは、音声合成フィルタのための励振ベクトルを包含する。このコードブックの内容は非適応型（つまり固定）である。適応マルチレート広帯域コーデックにおいては、固定コードブックは代数的コードブックを使用して実装される。

分数ラグ： サブサンプルの分解能をもつラグの値の集合。適応マルチレート広帯域コーデックにおいては、サブサンプルには1サンプルの1/2または1/4の分解能が用いられる。

フレーム： 20msの時間間隔（16kHzのサンプリングレートでは320サンプル）

イミタンススペクトル周波数：（イミタンススペクトル対を参照）

イミタンススペクトル対： LPCパラメータの変形。イミタンススペクトル対は、逆フィルタ伝達関数 $A(z)$ を分解して2つの伝達関数の対にすることによって得られる。2つの伝達関数のうち1つは偶対称、もう1つは奇対称となる。イミタンススペクトル対（イミタンススペクトル周波数とも呼ばれる）は、 z 単位円上の多項式の根である。

整数ラグ： 整数サンプルの分解能をもつラグの値の集合。

補間フィルタ： 整数サンプルの分解能をもつ入力を与えて、サブサンプルの分解能を推定するためのFIRフィルタ。この実装では、補間フィルタは低域通過フィルタの特性をもつ。したがって適応コードブックは、低域通過フィルタを通して補間された励振ベクトルからなる。

逆フィルタ： 音声信号から短期の相関を取り除くためのフィルタ。このフィルタは、声道の逆の周波数

応答をモデル化する。

ラグ： 長期フィルタの遅延。これは一般的には真のピッチ周期、またはその倍数もしくは約数である。

L P 分析窓： 各フレームに対し、短期フィルタ係数は、分析窓の中で高域通過フィルタを通った音声サンプルを用いて計算される。適応マルチレート広帯域コーデックにおいては、分析窓の長さは常に 384 サンプルである。全てのモードに対して、単一の L P 係数セットを生成するために単一の非対称窓が用いられる。5ms の先読みが分析に用いられる。

L P 係数： 線形予測 (L P) 係数 (線形予測符号化係数 (L P C) 係数ともいう) は、短期フィルタ係数に対する一般的表現である。

モード： 単独で用いられるときには、情報源コーデックのモード、つまり AMR-WB コーデックに使用される情報源コーデックの 1 つを指す。

開ループピッチ探索： 重み付けされた入力音声から直接、最適に近いラグの値を推定する処理。これは、開ループピッチの探索範囲を、開ループラグの推定値付近の数少ないラグの値に限定することで、ピッチ分析を簡単にするために行われるものである。適応マルチレート広帯域コーデックにおいては、開ループピッチ探索は各サブフレームで実行される。

残差： 逆フィルタの出力信号。

短期合成フィルタ： このフィルタは励振信号に対し、声道のインパルス応答をモデル化する短期相関を導入する。

聴覚重み付けフィルタ： コードブックの合成による分析 (A-b-S) 探索に使用されるフィルタ。このフィルタは、ホルマント (声道共鳴) のノイズマスキングの性質を利用する。つまり、誤差をホルマント周波数に近い帯域には少なく、遠い帯域には多く重み付けを行うというものである。

サブフレーム： 5ms の時間間隔 (16kHz のサンプリングレートでは 80 サンプル)。

ベクトル量子化： 複数のパラメータを単一のベクトルにグループ化して、まとめて量子化する手法。

零入力応答： 過去の入力に基づいた現在の内部状態で、零を入力したときのフィルタの出力。

零状態応答： 過去の入力は利用されず、つまりフィルタの内部状態は全て零という状況で、現在の値を入力したときのフィルタの出力。

3. 2 記号

本標準では、以下の記号を適用する。

$A(z)$	量子化されていない係数をもつ逆フィルタ
$\hat{A}(z)$	量子化された係数をもつ逆フィルタ
$H(z) = \frac{1}{\hat{A}(z)}$	量子化された係数をもつ音声合成フィルタ
a_i	量子化されていない線形予測パラメータ (直接型係数)
\hat{a}_i	量子化された線形予測パラメータ
m	L P モデルの次数
$W(z)$	聴覚重み付けフィルタ (量子化されていない係数)
γ_1	聴覚重み付け因子
T	サブフレームにおける開ループの分数ピッチラグに最も近い整数ピッチラグ
β	適応プリフィルタ係数 (量子化されたピッチ利得)
$H_{hl}(z)$	前処理の高域通過フィルタ
$w(n)$	L P 分析窓
L_1	L P 分析窓 $w(n)$ の 1 番目の部分の長さ
L_2	L P 分析窓 $w(n)$ の 2 番目の部分の長さ

$r(k)$	窓掛けされた音声 $s'(n)$ の自己相関
$w_{lag}(i)$	自己相関のためのラグ窓 (60Hz 帯域幅拡張)
f_0	帯域幅拡張周波数 [Hz]
f_s	サンプリング周波数 [Hz]
$r'(k)$	補正後の (帯域幅拡張された) 自己相関
$E(i)$	レビンソンアルゴリズムの i 番目の繰り返しにおける予測誤差
k_i	i 番目の反射係数
$a_j^{(i)}$	レビンソンアルゴリズムの i 番目の繰り返しにおける j 番目の直接型係数
$F_1'(z)$	対称 I S F 多項式
$F_2'(z)$	非対称 I S F 多項式
$F_1(z)$	多項式 $F_1'(z)$
$F_2(z)$	根 $z=1$ および $z=-1$ が除去された多項式 $F_2'(z)$
q_i	余弦領域のイミタンススペクトル対 (I S P)
\mathbf{q}	余弦領域の I S P ベクトル
$\hat{\mathbf{q}}_i^{(n)}$	フレーム n の i 番目のサブフレームにおける量子化された I S P ベクトル
ω_i	イミタンススペクトル周波数 (I S F)
$T_m(x)$	m 次のチェビシェフ多項式
$f_1(i), f_2(i)$	多項式 $F_1(z)$ および $F_2(z)$ の係数
$f_1'(i), f_2'(i)$	多項式 $F_1'(z)$ および $F_2'(z)$ の係数
$f(i)$	$F_1(z)$ または $F_2(z)$ のいずれかの係数
$C(x)$	チェビシェフ多項式の和多項式
x	角周波数 ω の余弦
λ_k	チェビシェフ多項式評価のための再帰係数
f_i	イミタンススペクトル周波数 (I S F) [Hz]
$\mathbf{f}^t = [f_1 f_2 \dots f_{16}]$	I S F のベクトル表現 [Hz]
$\mathbf{z}(n)$	フレーム n における平均除去 I S F ベクトル
$\mathbf{r}(n)$	フレーム n における I S F 予測残差ベクトル
$\mathbf{p}(n)$	フレーム n における予測 I S F ベクトル
$\hat{\mathbf{r}}(n-1)$	過去のフレームにおける量子化された残差ベクトル
$\hat{\mathbf{r}}_i^k$	量子化インデックス k における量子化 I S F サブベクトル i
d_i	イミタンススペクトル周波数 f_{i+1} と f_{i-1} との距離
$h(n)$	重み付き合成フィルタのインパルス応答
$H(z)W(z)$	重み付き合成フィルタ
T_1	直前の (1 番目または 3 番目の) サブフレームの分数ピッチラグに最も近い整数
$s'(n)$	窓掛けされた音声信号
$s_w(n)$	重み付けされた音声信号
$\hat{s}(n)$	再生音声信号
$x(n)$	適応コードブック探索のターゲット信号
$x_2(n), \mathbf{x}_2^t$	代数的コードブック探索のターゲット信号
$res_{LP}(n)$	L P 残差信号
$c(n)$	固定コードブックベクトル
$v(n)$	適応コードブックベクトル
$y(n) = v(n) * h(n)$	フィルタリングされた適応コードブックベクトル

$y_k(n)$	フィルタリングされた過去の励振ベクトル
$u(n)$	励振信号
$\hat{u}'(n)$	利得でスケーリングされた強調励振信号
T_{op}	最適開ループラグ
t_{min}	最小ラグ探索値
t_{max}	最大ラグ探索値
$R(k)$	適応コードブック探索における最大化相関項
$R(k)_t$	整数遅延 k および遅延の小数部分 t に対する $R(k)$ の補間値
A_k	インデックス k の代数的コードブック探索における最大化相関項
C_k	インデックス k における A_k の分子の相関
E_{Dk}	インデックス k における A_k の分母のエネルギー
$\mathbf{d}=\mathbf{H}'\mathbf{x}_2$	ターゲット信号 $x_2(n)$ とインパルス応答 $h(n)$ との相関、つまりバックワードフィルタを施されたターゲット信号
\mathbf{H}	$h(0)$ および $h(1), \dots, h(63)$ を対角要素にもつ、下三角テプリッツ畳み込み行列
$\Phi=\mathbf{H}'\mathbf{H}$	$h(n)$ の相関行列
$d(n)$	ベクトル \mathbf{d} の要素
$\phi(i, j)$	対称行列 Φ の要素
\mathbf{c}_k	励振ベクトル
C	A_k の分子の相関
m_i	i 番目のパルスの位置
g_i	i 番目のパルスの大きさ
N_p	固定コードブック励振ベクトルにおけるパルスの数
E_D	A_k の分母のエネルギー
$res_{LTP}(n)$	正規化された長期予測残差
$b(n)$	代数的コードブック探索において符号の初期設定に用いられる信号
$s_b(n)$	代数的コードブック探索のための符号
$d'(n)$	符号拡張され、バックワードフィルタを施されたターゲット信号
$\phi'(i, j)$	行列 Φ の補正された要素、符号情報を含む
$\mathbf{z}^t, z(n)$	$h(n)$ と畳み込まれた固定コードブックベクトル
$E(n)$	平均除去された励振エネルギー [dB]
\bar{E}	励振エネルギーの平均値
$\tilde{E}(n)$	予測エネルギー
$[b_1 b_2 b_3 b_4]$	MA 予測係数
$\hat{R}(k)$	サブフレーム k における量子化された予測誤差
E_t	励振エネルギーの平均値
$R(n)$	固定コードブック利得量子化の予測誤差
E_Q	固定コードブック利得量子化の量子化誤差
$e(n)$	合成フィルタ $1/\hat{A}(z)$ の状態
$e_w(n)$	合成による分析 (A-b-S) 探索の聴覚重み付け誤差
η	強調された励振ベクトルのための利得スケーリングファクタ
g_c	固定コードブック利得
g'_c	予測された固定コードブック利得
\hat{g}_c	量子化された固定コードブック利得

g_p	適応コードブック利得
\hat{g}_p	量子化された適応コードブック利得
$\gamma_{gc} = g_c / g'_c$	利得 g_c と予測利得 g'_c との補正係数
$\hat{\gamma}_{gc}$	γ_{gc} の最適値
γ_{sc}	利得スケーリング係数

3. 3 略語

本標準では、以下の略語を適用する。

ACELP	代数的符号励振線形予測 (Algebraic Code Excited Linear Prediction)
AGC	適応利得制御 (Adaptive Gain Control)
AMR	適応マルチレート (Adaptive Multi-Rate)
AMR-WB	適応マルチレート広帯域 (Adaptive Multi-Rate Wideband)
CELP	符号励振線形予測 (Code Excited Linear Prediction)
FIR	有限インパルス応答 (Finite Impulse Response)
ISF	イミタンススペクトル周波数 (Immittance Spectral Frequency)
ISP	イミタンススペクトル対 (Immittance Spectral Pair)
LP	線形予測 (Linear Prediction)
LPC	線形予測符号化 (Linear Predictive Coding)
LTP	長期予測器 (または長期予測) (Long Term Predictor, Long Term Prediction)
MA	移動平均 (Moving Average)
S-MSVQ	分割多段ベクトル量子化 (Split-MultiStage Vector Quantization)
WB	広帯域 (Wideband)
GERAN	GSM EDGE 無線アクセスネットワーク (GSM EDGE Radio Access Network)
GMSK	ガウス修正シフトキーイング (Gaussian Modified Shift Keying)
OTCH/F	オクタルTCH/フルレート (Octal-TCH/Full rate)、総ビットレート 68.4 kbit/s の GERAN-8PSK 無線チャンネル
OTCH/H	オクタルTCH/ハーフレート (Octal-TCH/Half rate)、総ビットレート 32.4 kbit/s の GERAN-8PSK の無線チャンネル
8PSK	8 値位相シフトキーイング (8 Phase Shift Keying)
TCH	通信回線 (Traffic Channel) (音声またはデータに対する専用無線チャンネル)
TCH/F	TCH/フルレート (TCH/Full rate)、総ビットレート 22.8 kbit/s の GERAN-GMSK 無線チャンネル
TCH/H	TCH/ハーフレート (TCH/Half rate)、総ビットレート 11.4 kbit/s の GERAN-GMSK 無線チャンネル
TFO	タンデムフリー動作 (Tandem Free Operation) (タンデムフリーは呼設定後のインバンドシグナリングを用いて実現される)
TrFO	トランスコーダフリー動作 (Transcoder Free Operation) (タンデムフリーは呼設定前のアウトバンドシグナリングを用いて実現され、トランスコーダは原則通信路には存在しない)
UTRAN	UMTS 陸上無線アクセスネットワーク (UMTS Terrestrial Radio Access Network)

4. 概要

本標準の構成は、次の通りである。

4. 1 節はA/DおよびD/A機能を含むオーディオ部分の機能記述が記載されている。4. 2 節はAMR

－WB符号器に対する入力フォーマットおよびAMR－WB復号器における出力フォーマットを記述している。4. 3節および4. 4節は、各々、AMR－WBコーデックの符号化処理および復号処理の原理に関する簡略な記述を示している。4. 5節では符号化パラメータのシーケンスおよび主観的重要度が示される。5章では、AMR－WBコーデックの符号化側の詳細を説明しており、6章では、復号側の詳細を説明している。7章では、AMR－WBコーデックの詳細なビット割当てが表に示されている。8章は、ホーミング処理に関して記述している。

4. 1 オーディオ部分の機能記述

アナログ・デジタル変換およびデジタル・アナログ変換は、原則として次の要素から構成される。

1) アナログから均一デジタルPCMへの変換 [1]

- － マイクロホン
 - － 入力レベル調整デバイス
 - － 入力折り返し阻止フィルタ
 - － 16kHz サンプリングのサンプルホールドデバイス
 - － アナログから 14 ビット表現での均一デジタル形式への変換
- 均一フォーマットは2の補数表現形式である。

2) 均一デジタルPCMからアナログへの変換 [1]

- － 14 ビット/16kHz の均一PCM形式からアナログへの変換
- － ホールドデバイス
- － $x/\sin(x)$ 補正を含む再生フィルタ
- － 出力レベル調整デバイス
- － イヤホンまたはスピーカ

端末装置においては、A/D機能は、14 ビット均一PCMフォーマットへの直接変換により実現される。D/A処理に対しては、逆の処理が行われる。

4. 2 音声サンプルの準備

符号器には 16 ビットワードに左詰めされた 14 ビット精度を有するサンプルで構成されるデータを与える。復号器は同一のフォーマットでのデータを出力する。トラフィックデータが違う形式の場合、音声コーデックの外側で、更なる処理を適用しなければならない。

4. 3 適応マルチレート広帯域音声符号器の原理

AMR－WBコーデックは、23.85, 23.05, 19.85, 18.25, 14.25, 15.85, 12.65, 8.85 および 6.60 kbit/s のビットレートを有する 9 つの情報源コーデックで構成される。

コーデックは、符号励振線形予測 (CELP) 符号化モデルに基づく。入力信号はフィルタ $H_{pre-emph}(z)=1-\mu z^{-1}$ を用いてプリエンファシスされる。そして、CELPモデルがプリエンファシスされた信号に適用される。16 次の線形予測 (LP)、すなわち短期の合成フィルタが使用され、それは次式で表される。

$$H(z) = \frac{1}{\hat{A}(z)} = \frac{1}{1 + \sum_{i=1}^m \hat{a}_i z^{-i}} \quad (1)$$

ここで、 $\hat{a}_i, i=1, \dots, m$ は (量子化された) 線形予測 (LP) パラメータであり、 $m=16$ は予測次数である。長期予測、すなわちピッチの合成フィルタは通常次式で表される。

$$\frac{1}{B(z)} = \frac{1}{1 - g_p z^{-T}} \quad (2)$$

ここで、 T はピッチ遅延であり、 g_p はピッチ利得である。ピッチ合成フィルタは、いわゆる適応コードブック手法により実現される。

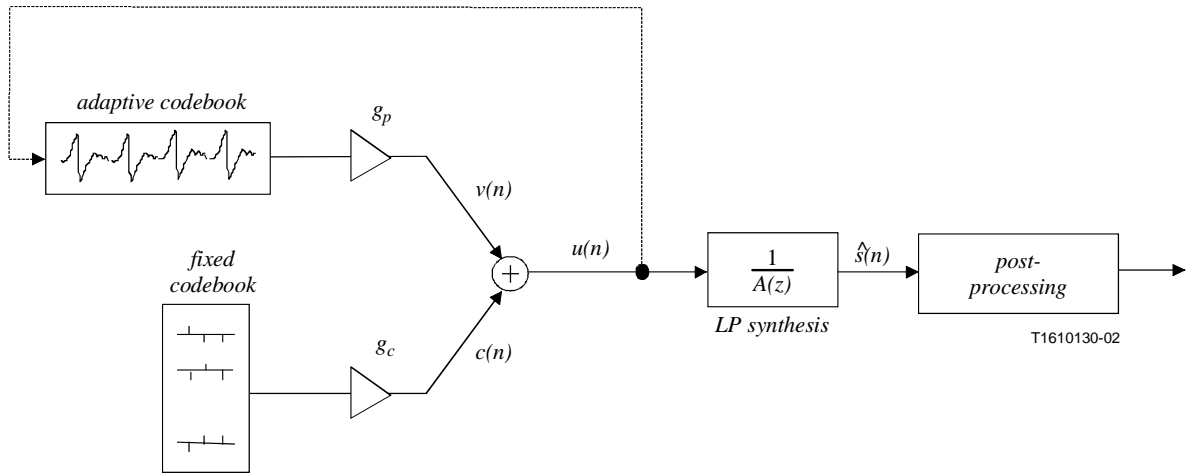


Figure 1 / JT-G722.2 Simplified block diagram of the CELP synthesis model

(ITU-T G.722.2)

CELP 音声合成モデルを Figure 1 / JT-G722.2 に示す。このモデルにおいて、短期 LP 合成フィルタへ入力される励振信号は、適応および固定コードブックからの 2 つの励振ベクトルを加算することによって構成される。音声は、短期合成フィルタに対して、これらのコードブックから適切に選択された 2 つのベクトルを与えることによって合成される。コードブック内の最適な励振シーケンスが、入力音声と合成音声の間の誤差を聴覚重み付き歪み尺度に基づいて最小にするという、合成による分析による探索手法を用いて選択される。

合成による分析による探索手法における聴覚重み付けフィルタは次式で与えられる。

$$W(z) = A(z/\gamma_1)H_{de-emph}(z) \quad (3)$$

ここで、 $A(z)$ は未量子化の LP フィルタ、 $H_{de-emph} = \frac{1}{1 - 0.68z^{-1}}$ 、そして $\gamma_1 = 0.92$ は聴覚重み付け係数である。重み付けフィルタは未量子化の LP パラメータを用いる。

符号器は、LPC、LTP および固定コードブックパラメータの分析を標準化周波数 12.8kHz で行う。符号化は 20ms の音声フレームに対して動作する。各々のフレームにおいて、音声信号は CELP モデルのパラメータ (LP フィルタ係数、適応および固定コードブックのインデックスおよび利得) を抽出するための分析が行われる。これらのパラメータに加えて、23.85 kbit/s モードにおいては、高域利得インデックスが算出される。これらのパラメータは符号化され送信される。復号器においては、これらのパラメータが復号されて、再生された励振信号に対して LP 合成フィルタによりフィルタ処理を行うことで、音声合成される。

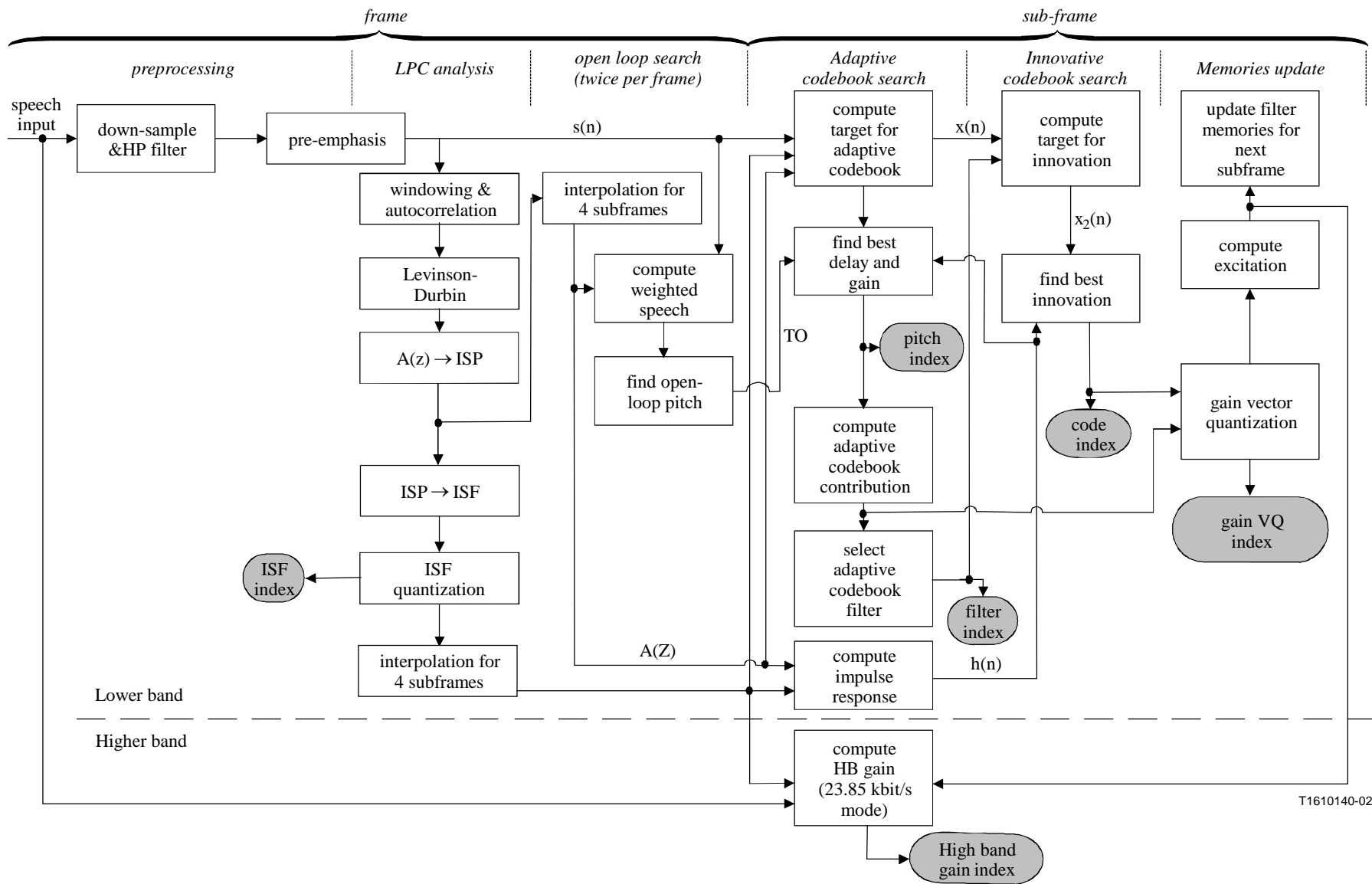


Figure 2 / JT-G722.2 Detailed block diagram of the ACELP encoder (ITU-T G.722.2)

符号化器における信号の流れを Figure 2 / JT-G722.2 に示す。間引きの後に、高域通過およびプリエンファシスのフィルタ処理が行われる。LP 分析は 1 フレームに 1 度行われる。LP パラメータのセットはイミタンススペクトル対 (ISP) に変換され、分割多段ベクトル量子化 (S-MSVQ) を用いてベクトル量子化される。音声フレームは 5ms 長 (標準化周波数 12.8kHz において 64 サンプル) の 4 サブフレームに分割される。適応および固定コードブックパラメータはサブフレーム毎に送信される。量子化および未量子化の LP パラメータ、あるいはその補間されたパラメータが、サブフレームに依存してそのいずれかが使用される。聴覚重み付き音声信号に基づき、開ループピッチラグが、1 サブフレームおき、あるいは 1 フレームに 1 度推定される。

次に、以下に示す処理がサブフレーム毎に反復される。

- LP 残差信号に対して、フィルタの初期状態として LP 残差信号と励振信号との誤差をフィルタ処理することにより更新された重み付き合成フィルタ $W(z)H(z)$ を用いたフィルタ処理によりターゲット信号 $x(n)$ が算出される。(これは、重み付き音声信号から重み付き合成フィルタの零入力応答を減算するという一般的な手法と等価である。)
- 重み付き合成フィルタのインパルス応答 $h(n)$ が算出される。
- 次に、ターゲット信号 $x(n)$ とインパルス応答 $h(n)$ を用いて、開ループピッチラグの周辺を探索することにより、(ピッチラグと利得を決定するための) 開ループピッチ分析が行われる。1/4 または 1/2 サンプル精度 (モードとピッチラグ値に依存) の分数ピッチが使用される。分数ピッチ探索における補間フィルタは、低域通過周波数応答を持つ。更に、適応コードブックには 2 つの潜在的な低域通過特性があり、その情報が 1 ビットで符号化される。
- ターゲット信号 $x(n)$ は適応コードブック寄与分 (フィルタリングされた適応コードブックベクトル) を除去することによって更新され、この新たなターゲット信号 $x_2(n)$ は最適な励振源を決定するために固定代数的コードブック探索に使用される。
- 適応コードブックおよび固定コードブックの利得は、(固定コードブック利得には移動平均 (MA) 予測が適用されて) 6 または 7 ビットでベクトル量子化される。
- 最後に、フィルタメモリが次のサブフレームのターゲット信号を算出するために、(決定された励振信号を用いて) 更新される。

AMR-WB コーデックのモード別のビット割当てを Table 1 / JT-G722.2 に示す。各 20ms 音声フレームにおいて、6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 および 23.85 kbit/s のビットレートに対応して、132, 177, 253, 285, 317, 365, 397, 461 および 477 ビットが生成される。コーデックパラメータ内の更なる詳細なビット割当ては、Table 1 2 a ~ 1 2 i / JT-G722.2 で与えられる。ここで、最上位ビット (MSB) が常に先に送信される。

Table 1 / JT-G722.2 Bit allocation of the AMR-WB coding algorithm for 20 ms frame
(ITU-T G.722.2)

Mode	Parameter	1st subframe	2nd subframe	3rd subframe	4th subframe	total per frame
23.85 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	88	88	88	88	352
	Codebook gain	7	7	7	7	28
	HB-energy	4	4	4	4	16
	Total					477
23.05 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	88	88	88	88	352
	Gains	7	7	7	7	28
		Total				
19.85 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	72	72	72	72	288
	Codebook gain	7	7	7	7	28
		Total				
18.25 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	64	64	64	64	256
	Gains	7	7	7	7	28
		Total				
15.85 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	52	52	52	52	208
	Gains	7	7	7	7	28
		Total				
14.25 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	44	44	44	44	176
	Gains	7	7	7	7	28
		Total				
12.65 kbit/s	VAD-flag					1
	ISP					46
	LTP-filtering	1	1	1	1	4
	Pitch delay	9	6	9	6	30
	Algebraic code	36	36	36	36	144
	Gains	7	7	7	7	28
		Total				
8.85 kbit/s	VAD-flag					1
	ISP					46
	Pitch delay	8	5	8	5	26
	Algebraic code	20	20	20	20	80
	Gains	6	6	6	6	24
		Total				
6.60 kbit/s	VAD-flag					1
	ISP					36
	Pitch delay	8	5	5	5	23
	Algebraic code	12	12	12	12	48
	Gains	6	6	6	6	24
		Total				

4. 4 適応マルチレート広帯域音声復号器の原理

復号器における信号の流れを Figure 3 / JT-G722.2 に示す。復号器においては、送信されたインデックスが受信したビットストリームから抽出される。このインデックスは各々の送信フレームにおけるコーデックパラメータを得るために復号される。これらのパラメータは、I S P ベクトル、4つの分数ピッチラグ、4つのLTPフィルタパラメータ、4つの励振コードベクトル、および4セットのベクトル量子化されたピッチおよび励振源利得である。23.85 kbit/s モードにおいては、高域利得インデックスも復号される。I S P ベクトルは、LPフィルタ係数に変換され、各サブフレームでのLPフィルタ係数を得るために補間される。次に、各々の64 サンプル長のサブフレームに対して下記の処理が行われる。

- ・ 適応および励振コードベクトルに対して各々の利得によりスケーリングされた信号が加算されて励振信号が再生される。
- ・ 励振信号に対してLP合成フィルタによるフィルタ処理により 12.8kHz の音声再生される。
- ・ 再生音声がデエンファシスされる。

最後に、再生された音声は、16kHz にアップサンプルされ、6kHz から 7kHz の周波数帯域に対して高域音声信号が加算される。

4. 5 符号化パラメータのシーケンスおよび主観的重要度

符号器は出力情報を一義的な固有のシーケンスおよび形式にて生成し、また復号器は同じ情報を同じ方法により受信できなければならない。各パラメータの出力ビットのシーケンスとビット割当てを Table 1 2 a ~ 1 2 i / JT-G722.2 に示す。

符号化音声の異なるパラメータおよびそれらのパラメータの個々のビットは、主観的品質に関して不均一な重要性を有している。

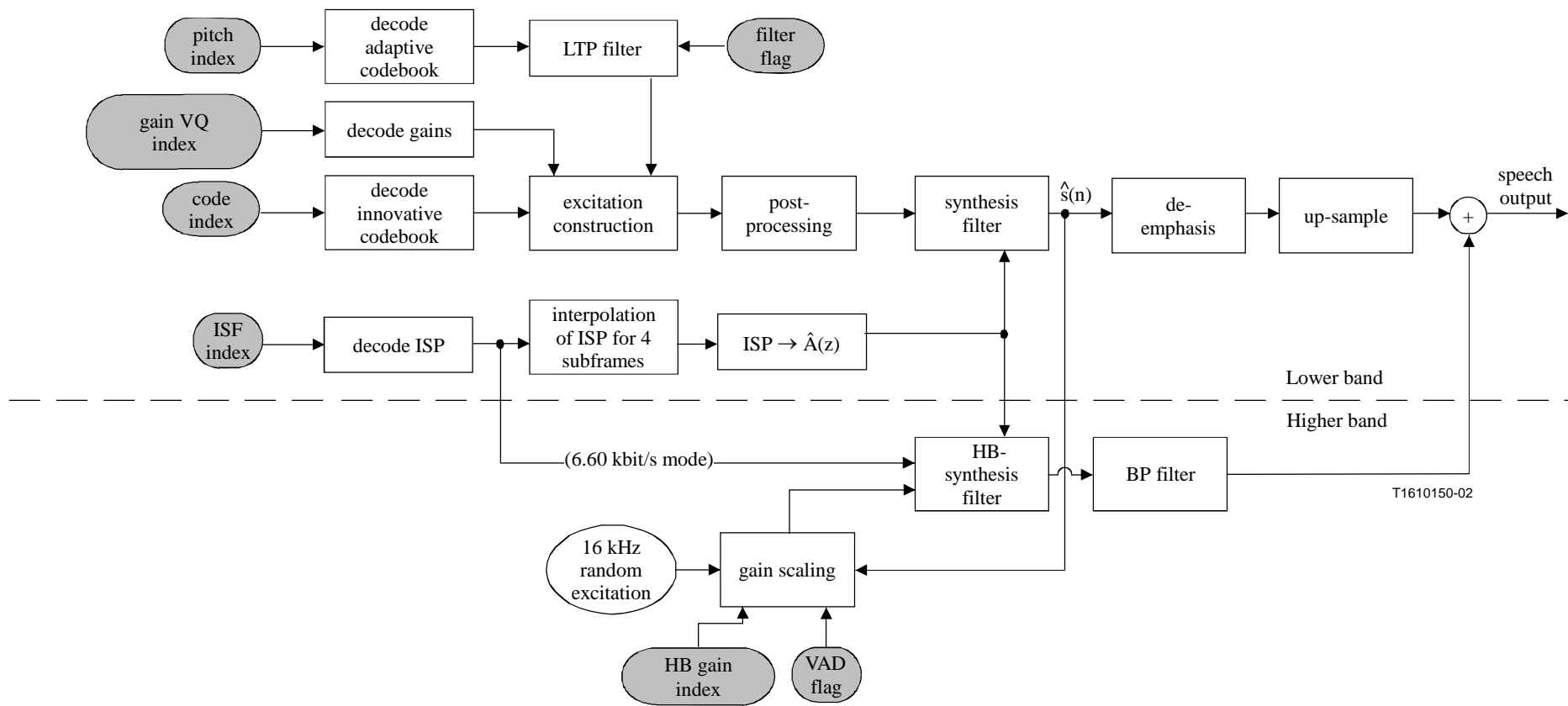


Figure 3 / JT-G722.2 Detailed block diagram of the ACELP decoder (ITU-T G.722.2)

5. 符号器の詳細

本章では、Figure 2 / JT-G722.2 に示された符号器のそれぞれの機能を記述する。

5. 1 前処理

符号器は、12.8kHz の標本化周波数で L P C、L T P および固定コードブックのパラメータの分析を行う。従って、入力信号は 16kHz から 12.8kHz へ間引きされる。間引きは、最初に 4 倍にアップサンプルし、その出力を 6.4kHz の遮断周波数を持つ低域通過 F I R フィルタ $H_{decim}(z)$ によりフィルタリングすることにより行われる。そして、その出力信号は 1/5 にダウンサンプルされる。フィルタリングによる遅延は、入力ベクトルの後尾に零を付加することによって補償される。

間引き処理後、符号化処理に先だって信号に対して、高域通過フィルタリングおよびプリアンファシス（および信号ダウンスケーリング）の 2 つの前処理が適用される。

（ダウンスケーリングは、固定小数点実装におけるオーバフローの可能性を減らすために、入力信号を 2 で除算する処理からなる。）

高域通過フィルタは、望ましくない低域周波数成分に対する対策として機能する。下式で与えられる遮断周波数 50Hz のフィルタが用いられる。

$$H_{hl}(z) = \frac{0.989502 - 1.979004z^{-1} + 0.989502z^{-2}}{1 - 1.978882z^{-1} + 0.979126z^{-2}} \quad (4)$$

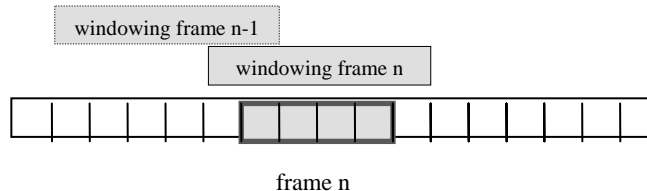
（ダウンスケーリングおよび高域通過フィルタ処理は、 $H_{hl}(z)$ の分子項の係数を 2 で除算することで併せて行われる。）

プリアンファシスでは、高域の周波数を強調するために 1 次の高域通過フィルタが用いられ、それは次式で与えられる。

$$H_{pre-emph}(z) = 1 - 0.68z^{-1} \quad (5)$$

5. 2 線形予測分析と量子化

短期予測（L P）分析が、30ms 非対称窓を用いた自己相関法により 1 音声フレーム当たり 1 度行われる。自己相関算出時には 5ms のオーバーヘッドが用いられる。フレーム構成は下図のように表される。



窓掛けされた音声に対する自己相関は、レビンソンーダービンアルゴリズムを用いて L P 係数に変換される。そして、L P 係数は量子化および補間を行うために I S P 領域へ変換される。補間された、量子化および未量子化のフィルタは、（各サブフレームでの合成および重み付けフィルタを構成するために）L P フィルタ係数に再度変換される。

5. 2. 1 窓掛けおよび自己相関の算出

LP分析は非対称窓を用いて1フレームに1度行われる。窓は、第4サブフレームにその重みが集中しており、第1の部分ハミング窓の半分の部分、第2の部分ハミング窓コサイン関数の1/4周期分という2つの部分からなる。窓は下式で与えられる。

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{2L_1 - 1}\right), \quad n = 0, \dots, L_1 - 1$$

$$= \cos\left(\frac{2\pi(n - L_1)}{4L_2 - 1}\right), \quad n = L_1, \dots, L_1 + L_2 - 1$$
(6)

ここで、 $L_1=256$ および $L_2=128$ が用いられる。

窓掛けされた音声 $s'(n), n=0, \dots, 383$ に対する自己相関は下式により算出される。

$$r(k) = \sum_{n=k}^{383} s'(n)s'(n-k), \quad k = 0, \dots, 16$$
(7)

そして、下式の窓[3]を用いた自己相関のラグ窓かけにより 60Hz の帯域幅拡張が行われる。

$$w_{lag}(i) = \exp\left[-\frac{1}{2} \left(\frac{2\pi f_0 i}{f_s}\right)^2\right], \quad i = 1, \dots, 16$$
(8)

ここで、 $f_0=60$ Hz は拡張帯域幅、 $f_s=12800$ Hz は標本化周波数である。更に $r(0)$ は -40dB の雑音フロアを加算することと等価である白色雑音補正係数 1.0001 で乗算される。

5. 2. 2 窓掛けおよび自己相関の算出

補正された自己相関 $r'(0) = 1.0001 r(0)$ および $r'(k) = r(k)w_{lag}(k), k = 1, \dots, 16$ を用いて、下記の等式を解くことによってLPフィルタ係数 $a_k, k=1, \dots, 16$ を得る。

$$\sum_{k=1}^{16} a_k r'(|i-k|) = -r'(i), \quad i = 1, \dots, 16$$
(9)

等式(9)は、レビンソン-ダービンアルゴリズム[3]を用いて解かれる。このアルゴリズムは、次の再帰式を利用する。

$$E(0) = r'(0)$$

For $i = 1$ to 16 do

$$k_i = -\left[r'(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} r'(i-j)\right] / E(i-1)$$

$$a_i^{(i)} = k_i$$

For $j = 1$ to $i-1$ do

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}$$

$$E(i) = (1 - k_i^2) E(i-1)$$

最終的な解は $a_j = a_j^{(16)}, j=1, \dots, 16$ で与えられる。

LPフィルタ係数は、量子化および補間処理のためにISP表現[5]へ変換される。ISP領域への変換、およびLPフィルタ領域への再変換は次の2つの節で説明される。

5. 2. 3 LPからISPへの変換

LPフィルタ係数 $a_k, k=1, \dots, 16$ は、量子化と補間を目的にISP表現に変換される。16次のLPフィルタに

対して、I S Pはそれぞれ和と差の多項式の根として定義される。

$$f_1'(z) = A(z) + z^{-16}A(z^{-1}) \quad (10)$$

$$f_2'(z) = A(z) - z^{-16}A(z^{-1}) \quad (11)$$

(多項式 $f_1(z)$ と $f_2(z)$ は、それぞれ対称と非対称である。) これらの多項式の根は、全て単位円上にあり、それらが互いに交番することが証明されている[6]。 $f_2(z)$ は、 $z=1$ ($\omega=0$)と $z=-1$ ($\omega=\pi$)の2つの根を持っている。この2つの根を除去するため新しい多項式を定義する。

$$f_1(z) = f_1'(z) \quad (12)$$

$$f_2(z) = f_2'(z)/(1-z^{-2}) \quad (13)$$

多項式 $f_1(z)$ と $f_2(z)$ は、それぞれ単一円上($e^{\pm j\omega_i}$)に8個と7個の共役根を持っている。従って、多項式は以下に示される通りである。

$$F_1(z) = (1+a[16]) \prod_{i=0,2,\dots,14} (1-2q_i z^{-1} + z^{-2}) \quad (14)$$

$$F_2(z) = (1-a[16]) \prod_{i=1,3,\dots,13} (1-2q_i z^{-1} + z^{-2}) \quad (15)$$

ここで、 $q_i=\cos(\omega_i)$ 、 ω_i はイミタンススペクトル周波数(I S F)であり $a[16]$ は最終予測係数である。I S Fは順序特性 $0 < \omega_1 < \omega_2 < \dots < \omega_{16} < \pi$ を満足する。 q_i は余弦領域のI S Pである。

両方の多項式 $f_1(z)$ と $f_2(z)$ が対称であるため、各多項式の最初の8個と7個の係数と最終予測係数だけを計算する必要がある。

これらの多項式の係数は、再帰的な関係式により求められる。

for $i=0$ to 7

$$\begin{aligned} f_1(i) &= a_i + a_{m-i} \\ f_2(i) &= a_i - a_{m-i} + f_2(i-2) \end{aligned} \quad (16)$$

$$f_1(8) = 2a_8$$

ここで、 $m=16$ は予測次数、および $f_2(-2) = f_2(-1) = 0$ である。

I S Pは、0と π の間を100等分し、その点について多項式 $F_1(z)$ と $F_2(z)$ の値を求めて極性変化を照合することによって求められる。極性変化は根の存在を意味し、極性変化の間隔を4分割して精度のよい根を突き止める。チェビシェフ多項式を用いて $F_1(z)$ と $F_2(z)$ の値を求める[7]。この方法により、直接、根が余弦領域 $\{q_i\}$ に求められる。 $z=e^{j\omega}$ で値を求めた多項式 $F_1(z)$ と $F_2(z)$ は、以下に示す通りである。

$$F_1(\omega) = 2e^{-j8\omega} C_1(x) \quad \text{and} \quad F_2(\omega) = 2e^{-j7\omega} C_2(x) \quad (17)$$

$$C_1(x) = \sum_{i=0}^7 f_1(i) T_{8-i}(x) + f_1(8)/2 \quad \text{and} \quad C_2(x) = \sum_{i=0}^6 f_2(i) T_{8-i}(x) + f_2(7)/2 \quad (18)$$

ここで、 $T_m=\cos(m\omega)$ は、 m 次のチェビシェフ多項式であり、 $f(i)$ は式(16)の方程式を使用して計算された $F_1(z)$ または $F_2(z)$ の係数である。多項式 $C(x)$ は、再帰的な関係式を使用して $x=\cos(\omega)$ として求められる。

```

for k = nf - 1 down to 1
  bk = 2xbk+1 - bk+2 + f(nf - k)
end
C(x) = xb1 - b2 + f(nf)/2

```

ここで、 $C_1(x)$ の場合 $n_f=8$ 、 $C_2(x)$ の場合 $n_f=7$ であり、初期値は $b_{n_f}=f(0)$ と $b_{n_f+1}=0$ である。チェビシェフ多項式計算方法の詳細は[7]に示す通りである。

5. 2. 4 I S P から L P への変換

一旦、I S P の量子化と補間を行い、それらを L P 係数領域 $\{a_k\}$ に変換する。L P 領域への変換は、以下に示す通り行われる。量子化され補間された I S P、 $q_i, i=0, \dots, m-1$ に $m=16$ を代入した方程式(14)と(15)を展開すると、 $F_1(z)$ と $F_2(z)$ の係数が求められる。以下の再帰的な関係式は、 $f_1(z)$ を計算するために使用される。

```

for i = 2 to m/2
  f1(i) = -2q2i-2f1(i-1) + 2f1(i-2)
  for j = i-1 down to 2
    f1(j) = f1(j) - 2q2i-2f1(j-1) + f1(j-2)
  end
  f1(1) = f1(1) - 2q2i-2
end
end

```

ここで、初期値は $f_1(0)=1$ と $f_1(1)=-2q_0$ である。係数 $f_2(i)$ は、 q_{2i-2} を q_{2i-1} 、 $m/2$ を $m/2-1$ に置換して初期化条件を $f_2(0)=1$ と $f_2(1)=-2q_1$ とすることで同様に計算される。

一旦、係数 $f_1(z)$ と $f_2(z)$ が得られたら、 $F_2(z)$ に $1-z^2$ を掛け合わせて $F'_2(z)$ を求める。すなわち、

$$\begin{aligned} f'_2(i) &= f_2(i) - f_2(i-2) & i &= 2, \dots, m/2-1 \\ f'_1(i) &= f_1(i) & i &= 0, \dots, m/2 \end{aligned} \quad (19)$$

さらに $F'_1(z)$ と $F'_2(z)$ にそれぞれ $1+q_{m-1}$ と $1-q_{m-1}$ を掛け合わせる。すなわち、

$$\begin{aligned} f'_2(i) &= (1 - q_{m-1})f'_2(i) & i &= 0, \dots, m/2-1 \\ f'_1(i) &= (1 + q_{m-1})f'_1(i) & i &= 0, \dots, m/2 \end{aligned}$$

(20)式より最終的に L P 係数が求められる。

$$\begin{aligned} a_i &= 0.5f'_1(i) + 0.5f'_2(i) & i &= 1, \dots, m/2-1 \\ &0.5f'_1(i) - 0.5f'_2(i) & i &= m/2+1, \dots, m-1 \\ &0.5f'_1(m/2) & i &= m/2 \\ &q_{m-1} & i &= m \end{aligned} \quad (20)$$

(20)式は $A(z) = (F'_1(z) + F'_2(z))/2$ の関係式と $F'_1(z)$ と $F'_2(z)$ がそれぞれ対称および非対称多項式であることを考慮すると直接導き出される。

5. 2. 5 I S P 係数の量子化

L P フィルタ係数は、周波数領域の I S P 表現を用いて量子化される。すなわち、

$$f_i = \frac{f_s}{2\pi} \arccos(q_i) \quad i = 0, \dots, 14$$

$$= \frac{f_s}{4\pi} \arccos(q_i) \quad i = 15$$
(21)

ここで、 f_i は [0,6400] Hz における I S F、 $f_s=12800$ はサンプリング周波数である。I S Fベクトルは、 $\mathbf{f}' = [f_0, f_1, \dots, f_{15}]$ である。ここで \mathbf{t} は転置を示す。

1 次のMA予測が適用されて、残差 I S Fベクトルは、分割ベクトル量子化 (S V Q) と多段ベクトル量子化 (M S V Q) の組み合わせを用いて量子化される。予測と量子化は、以下に示す通り実行される。 $\mathbf{z}(n)$ はフレーム n の平均除去 I S Fベクトルを表す。予測残差ベクトル $\mathbf{r}(n)$ は、以下に示す通りである。

$$\mathbf{r}(n) = \mathbf{z}(n) - \mathbf{p}(n)$$
(22)

ここで、 $\mathbf{p}(n)$ はフレーム n における予測 I S Fベクトルである。1 次の移動平均 (MA) 予測が用いられる。

$$\mathbf{p}(n) = \frac{1}{3} \hat{\mathbf{r}}(n-1)$$
(23)

ここで、 $\hat{\mathbf{r}}(n-1)$ は過去フレームにおける量子化残差ベクトルである。

I S F残差ベクトル \mathbf{r} は、分割多段ベクトル量子化 (S - M S V Q) を用いて量子化される。ベクトルは、それぞれ 9 次と 7 次の 2 つのサブベクトル $\mathbf{r}_1(n)$ および $\mathbf{r}_2(n)$ に分けられる。2 つのサブベクトルは、2 つの段階において量子化される。最初の段階において $\mathbf{r}_1(n)$ は 8 ビット、 $\mathbf{r}_2(n)$ は 8 ビットに量子化される。

8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 および 23.85 kbit/s モードに対して、量子化誤差ベクトル $\mathbf{r}_i^{(2)} = \mathbf{r}_i - \hat{\mathbf{r}}_i, i=1,2$ は、次の段階でそれぞれ 3 個と 2 個のサブベクトルに分けられる。サブベクトルは、Table 2 / JT-G722.2 において記述されるビットレートを用いて量子化される

Table 2 / JT-G722.2 Quantization of ISP vector for the 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 (ITU-T G.722.2) or 23.85 kbit/s modes

1. UNQUANTIZED 16-ELEMENT-LONG ISP VECTOR				
2. STAGE 1 (\mathbf{r}_1) 8 bits			2. STAGE 1 (\mathbf{r}_2) 8 bits	
3. STAGE 2 ($\mathbf{r}^{(2)}_{1,0-2}$) 6 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{1,3-5}$) 7 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{1,6-8}$) 7 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{2,0-2}$) 5 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{2,3-6}$) 5 bits

6.60 kbit/s モードに対して、量子化誤差ベクトル $\mathbf{r}_i^{(2)} = \mathbf{r}_i - \hat{\mathbf{r}}_i, i=1,2$ は、次の段階でそれぞれ 2 個と 1 個のサブベクトルに分けられる。サブベクトルは、Table 3 / JT-G722.2 において記述されるビットレートを用いて量子化される。

Table 3 / JT-G722.2 Quantization of ISP vector for the 6.60 kbit/s mode (ITU-T G.722.2)

1. UNQUANTIZED 16-ELEMENT-LONG ISP VECTOR		
2. STAGE 1 (\mathbf{r}_1) 8 bits		2. STAGE 1 (\mathbf{r}_2) 8 bits
3. STAGE 2 ($\mathbf{r}^{(2)}_{1,0-4}$) 7 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{1,5-8}$) 7 bits	3. STAGE 2 ($\mathbf{r}^{(2)}_{2,0-6}$) 6 bits

自乗誤差 I S P 歪み尺度が、量子化プロセスにおいて使用される。一般に入力 I S P または誤差残差サブベクトル $\mathbf{r}_i, i=1,2$ およびインデックス k の量子化ベクトル $\hat{\mathbf{r}}_i^k$ について、量子化は次式を最小にするインデックス k を求められることで実行される。

$$E = \sum_{i=m}^n [r_i - \hat{r}_i^k]^2 \quad (24)$$

ここで、 m と n は第 1 番目と最後のサブベクトル要素である。

5. 2. 6 I S P の補間

1～3 番目のサブフレームについては、隣接したフレームの線形補間パラメータを使用するのに対して、4 番目のサブフレームについては、量子化（および未量子化）LP パラメータのセットを使用する。補間は \mathbf{q} 領域の I S P について実行される。 $\hat{\mathbf{q}}_4^{(n)}$ は、フレームの 4 番目サブフレームにおける I S P ベクトルであり、 $\hat{\mathbf{q}}_4^{(n-1)}$ は過去フレーム $n-1$ の 4 番目サブフレームの I S P ベクトルである。1～3 番目のサブフレームの補間 I S P ベクトルは、以下に示される通りである。

$$\begin{aligned} \hat{\mathbf{q}}_1^{(n)} &= 0.55\hat{\mathbf{q}}_4^{(n-1)} + 0.45\hat{\mathbf{q}}_4^{(n)} \\ \hat{\mathbf{q}}_2^{(n)} &= 0.2\hat{\mathbf{q}}_4^{(n-1)} + 0.8\hat{\mathbf{q}}_4^{(n)} \\ \hat{\mathbf{q}}_3^{(n)} &= 0.04\hat{\mathbf{q}}_4^{(n-1)} + 0.96\hat{\mathbf{q}}_4^{(n)} \end{aligned}$$

量子化されていない I S P の補間についても同式が用いられる。補間された I S P ベクトルを用いて、5. 2. 4 節に説明する I S P から LP への変換方法により、各サブフレーム（量子化、未量子化の両方）の異なる LP フィルタを計算する。

5. 3 聴覚重み付け

従来の聴覚重み付けフィルタ $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ は、ホルマント構造および要求されるスペクトル傾斜を同時にモデリングする場合において本質的な制限を持っている。低周波数から高周波数の広いダイナミックレンジをもつ広帯域信号においては、スペクトル傾斜がより著しい。この問題は、入力にプリエンファシスフィルタを挿入し、プリエンファシスされた音声 $s(n)$ に基づいて LP フィルタ $A(z)$ を計算して、分母を固定して修正したフィルタ $W(z)$ を使用して解決される。この構造により、傾斜からホルマント重み付けが実質上分けられる。

$W(z) = A(z/\gamma_1)H_{de-emph}(z)$ 形式の重み付けフィルタを使用する。ここで $H_{de-emph} = \frac{1}{1-\beta_1 z^{-1}}$ 、および $\beta_1=0.68$ である。

プリエンファシスされた音声信号 $s(n)$ に基づいて $A(z)$ が計算されるので、 $A(z)$ が入力音声に基づいて計算される場合と比較して、フィルタ $1/A(z/\gamma_1)$ の傾斜が著しくない。復号器の最後にデエンファシスを実行することにより、伝達関数 $W^{-1}(z)H_{de-emph}(z)=1/A(z/\gamma_1)$ を持つフィルタが、量子化誤差スペクトルを形成することが示されている。このように量子化誤差スペクトルは、伝達関数が $1/A(z/\gamma_1)$ のフィルタにより形成される。ここで、 $A(z)$ はプリエンファシスされた音声信号に基づいて計算される。

5. 4 開ループピッチ分析

各フレームにおいてピッチラグの推定値を求めるために、モードに従って、開ループピッチ分析がフレームあたり 1 回（20ms 毎に 1 回）、またはフレームあたり 2 回（10ms 毎に 1 回）行われる。これはピッチ分析を簡略化して、閉ループピッチ探索を、開ループピッチ探索により推定されたラグ周辺の少数のラグに制限するためである。

開ループピッチ推定は、重み付けフィルタ $W(z) = A(z/\gamma_1)H_{de-emph}(z)$ (ここで、 $H_{de-emph} = \frac{1}{1-\beta_1 z^{-1}}$ 、および $\beta_1=0.68$) で入力音声信号をフィルタリングして求められる重み付け音声信号 $s_w(n)$ に基づく。すなわち、サイズ L のサブフレームにおいて、重み付けされた音声は、以下に示される通りである。

$$s_w(n) = s(n) + \sum_{i=1}^{16} a_i \gamma_1^i s(n-i) + \beta_1 s_w(n-1), n=0, \dots, L-1 \quad (25)$$

開ループピッチ分析は、信号を 1/2 に間引きして行う。4 次の F I R フィルタ $H_{decim2}(z)$ で $s_w(n)$ をフィルタリングして、その出力を 1/2 にダウンサンプリングして信号 $s_{wd}(n)$ を求めることにより、間引きされた信号が求められる。

5. 4. 1 6.60 kbit/s モード

各フレームのピッチラグの推定値を求めるために、開ループピッチ分析がフレームあたり 1 回 (20ms 毎に 1 回) 行われる。

開ループピッチ分析は次に示す通り行われる。最初に各ピッチのラグの値 d に対して、間引きされた重み付き音声の相関関数を以下のように決定する。

$$C(d) = \sum_{n=0}^{128} s_{wd}(n) s_{wd}(n-d) w(d), d=17, \dots, 115 \quad (26)$$

ここで、 $w(d)$ は重み付け関数である。推定されたピッチラグは、重み付けした相関関数 $C(d)$ を最大にするラグである。重み付けにより低ピッチラグの値を強調して、正しい遅延の倍数を選択する可能性を減少させる。重み付け関数は 2 つの部分から構成される。すなわち、低ピッチラグのエンファシス関数 $w_l(d)$ 、および前フレーム遅延の隣接エンファシス関数 $w_n(d)$ である。

$$w(d) = w_l(d) w_n(d) \quad (27)$$

低ピッチラグのエンファシス関数は、以下に示す通りである。

$$w_l(d) = cw(d) \quad (28)$$

ここで、 $cw(d)$ は固定小数演算記述のテーブルで定義される。前フレームラグの隣接エンファシス関数は、前音声フレームのピッチラグに依存する。

$$w_n(d) = \begin{cases} cw(|T_{old} - d| + 98) & v > 0.8 \\ 1.0 & otherwise \end{cases} \quad (29)$$

ここで、 T_{old} は有声ハーフフレームの過去 5 フレームにおけるピッチラグの中央値 (メディアンフィルタの出力) で、 v は適応パラメータである。もし、開ループゲイン $g > 0.6$ をもつ有声フレームとして分類されるならば、次フレームで v の値を 1.0 に設定する。そうでなければ、 v の値を $v = 0.9 v$ で更新する。開ループゲインは、以下に示す通りである。

$$g = \frac{\sum_{n=0}^{127} s_{wd}(n) s_{wd}(n-d_{max})}{\sqrt{\sum_{n=0}^{127} s_{wd}^2(n) \sum_{n=0}^{127} s_{wd}^2(n-d_{max})}} \quad (30)$$

ここで、 d_{max} は $C(d)$ を最大にするピッチ遅延である。メディアンフィルタは有声フレーム区間のみで更新され

る。重み付けは、過去のピッチラグの信頼性に依存する。もし前フレームが無声音または無音を含んでいるならば、パラメータ ν により重み付けが減衰される。

5. 4. 2 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 および 23.85 kbit/s モード

各フレームにおけるピッチラグの2つの推定値を求めるために、開ループピッチ分析がフレームあたり2回(10ms 毎に1回)行われる。

開ループピッチ分析は次に示す通り行われる。最初に各ピッチのラグの値 d に対して、間引きされた重み付き音声の相関関数を以下のように決定する。

$$C(d) = \sum_{n=0}^{63} s_{wd}(n) s_{wd}(n-d) w(d), d=17, \dots, 115 \quad (31)$$

ここで、 $w(d)$ は重み付け関数である。推定されたピッチラグは、重み付けした相関関数 $C(d)$ を最大にする遅延である。重み付けにより低ピッチラグの値を強調して、正しい遅延の倍数を選択する可能性を減少させる。重み付け関数は2つの部分から構成される。すなわち、低ピッチラグのエンファシス関数、 $w_l(d)$ 、および前フレームラグの隣接エンファシス関数 $w_n(d)$ である。

$$w(d) = w_l(d) w_n(d) \quad (32)$$

低ピッチラグのエンファシス関数は、以下に示す通りである。

$$w_l(d) = cw(d) \quad (33)$$

ここで、 $cw(d)$ は固定小数演算記述のテーブルで定義される。前フレームラグの隣接エンファシス関数は、前音声フレームのピッチラグに依存する。

$$w_n(d) = \begin{cases} cw(|T_{old} - d| + 98), & \nu > 0.8 \\ 1.0, & \text{otherwise} \end{cases} \quad (34)$$

ここで、 T_{old} は有声ハーフフレームの過去5フレームにおけるピッチラグの中央値(メディアンフィルタの出力)で、 ν は適応パラメータである。もし、開ループゲイン $g > 0.6$ をもつ有声フレームとして分類されるならば、次フレームで ν の値を1.0に設定する。そうでなければ、 ν の値を $\nu = 0.9 \nu$ で更新する。開ループゲインは、以下に示す通りである。

$$g = \frac{\sum_{n=0}^{63} s_{wd}(n) s_{wd}(n-d_{max})}{\sqrt{\sum_{n=0}^{63} s_{wd}^2(n) \sum_{n=0}^{63} s_{wd}^2(n-d_{max})}} \quad (35)$$

ここで、 d_{max} は $C(d)$ を最大にするピッチ遅延である。メディアンフィルタは有声フレーム区間のみで更新される。重み付けは、過去のピッチラグの信頼性に依存する。もし前フレームが無声音または無音を含んでいるならば、パラメータ ν により重み付けが減衰される。

5. 5 インパルス応答の計算

重み付き合成フィルタ $H(z)W(z) = A(z/\gamma_1)H_{de-emph}(z)/\hat{A}(z)$ のインパルス応答 $h(n)$ がサブフレームごとに計算される。このインパルス応答は、適応コードブックおよび固定コードブックの探索のために必要とされる。インパルス応答 $h(n)$ は、零要素で拡張したフィルタ $A(z/\gamma_1)$ の係数を持つベクトルを $1/\hat{A}(z)$ および $H_{de-emph}(z)$ の2

つのフィルタに通してフィルタリングすることによって計算される。

5. 6 ターゲット信号の計算

適応コードブック探索のためのターゲット信号は、通常、重み付き音声信号 $s_w(n)$ から重み付き合成フィルタ $H(z)W(z) = A(z/\gamma_1)H_{de-emph}(z)/\hat{A}(z)$ の零入力応答を差し引くことによって計算される。これは、サブフレーム単位で行われる。

ターゲット信号を計算する等価な手順は、LP 残差信号 $r(n)$ を、合成フィルタ $1/\hat{A}(z)$ および重み付きフィルタ $A(z/\gamma_1)H_{de-emph}(z)$ を組み合わせたものを通してフィルタリングするものであり、本コーデックではこの手順が用いられる。サブフレームの励振信号が決定された後に、これらのフィルタの初期状態が、LP 残差信号と励振信号との差分をフィルタリングすることによって更新される。これらのフィルタのメモリ更新については、5. 10 節で説明する。

ターゲットベクトルを調べるために必要とされる残差信号 $r(n)$ は、適応コードブック探索においても、過去の励振信号バッファを拡張するために用いられる。次節で説明するように、これによって、サブフレーム長である64よりも少ない遅延について、適応コードブック探索手順が簡略化される。LP 残差信号は以下の式で与えられる。

$$r(n) = s(n) + \sum_{i=1}^{16} \hat{a}_i s(n-i), n = 0, \dots, 63 \quad (36)$$

5. 7 適応コードブック

適応コードブック探索は、サブフレーム単位で行われる。適応コードブック探索は、閉ループピッチ探索を行うこと、その次に選択された分数ピッチラグで過去の励振信号を補間することによって適応コードベクトルを計算すること、から構成される。適応コードブックパラメータ（あるいはピッチパラメータ）は、ピッチフィルタの遅延および利得である。探索の段階においては、閉ループ探索を簡略化するために、励振信号はLP 残差信号によって拡張される。

12.65, 14.25, 15.85, 18.25, 19.85, 23.05, および 23.85 kbit/s モードでは、1 番目および3 番目のサブフレームについては、 $[34, 127 \frac{3}{4}]$ の範囲では $1/4$ の分解能で、 $[128, 159 \frac{1}{2}]$ の範囲では $1/2$ の分解能で分数ピッチ遅延が用いられ、 $[160, 231]$ の範囲では整数遅延のみが用いられる。2 番目および4 番目のサブフレームについては、 $[T_1-8, T_1+7 \frac{3}{4}]$ の範囲で $1/4$ のピッチ分解能が常に用いられる。ここで、 T_1 は前の（1 番目あるいは3 番目の）サブフレームの分数ピッチラグに最も近い整数である。

8.85 kbit/s モードでは、1 番目および3 番目のサブフレームについては、 $[34, 91 \frac{1}{2}]$ の範囲では $1/2$ の分解能で分数ピッチ遅延が用いられ、 $[92, 231]$ の範囲では整数遅延のみが用いられる。2 番目および4 番目のサブフレームについては、 $[T_1-8, T_1+7 \frac{1}{2}]$ の範囲で $1/2$ のピッチ分解能が常に用いられる。ここで、 T_1 は直前の（1 番目あるいは3 番目の）サブフレームの分数ピッチラグに最も近い整数である。

6.60 kbit/s モードでは、1 番目のサブフレームについては、 $[34, 91 \frac{1}{2}]$ の範囲では $1/2$ の分解能で分数ピッチ遅延が用いられ、 $[92, 231]$ の範囲では整数遅延のみが用いられる。2 番目、3 番目、および4 番目のサブフレームについては、 $[T_1-8, T_1+7 \frac{3}{4}]$ の範囲で $1/4$ のピッチ分解能が常に用いられる。ここで、 T_1 は1 番目のサブフレームの分数ピッチラグに最も近い整数である。

閉ループピッチ分析は、サブフレーム単位で、開ループピッチ推定値 (T_{op}) の周辺で行われる。8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, および 23.85 kbit/s モードでは、1 番目（および3 番目）のサブフレームで、 $T_{op} \pm 7$ の範囲が、34, ..., 231 に制限されて探索される。6.60 kbit/s モードでは、1 番目のサブフレームで、 $T_{op} \pm 7$ の範囲が、

34,...,231に制限されて探索される。すべてのモードについて、他のサブフレームでは、前に述べたように、前のサブフレームで選択された整数ピッチの周辺で閉ループピッチ分析が行われる。12.65, 14.25, 15.85, 18.25, 19.85, 23.05, および23.85 kbit/sモードでは、1番目および3番目のサブフレームでピッチ遅延が9ビットで符号化され、他のサブフレームの相対的な遅延が6ビットで符号化される。8.85 kbit/sモードでは、1番目および3番目のサブフレームでピッチ遅延が8ビットで符号化され、他のサブフレームの相対的な遅延が5ビットで符号化される。6.60 kbit/sモードでは、1番目のサブフレームでピッチ遅延が8ビットで符号化され、他のサブフレームの相対的な遅延が5ビットで符号化される。

閉ループピッチ探索は、入力音声と合成音声の重み付き自乗平均誤差を最小にすることによって行われる。これは、以下の項を最大にすることによって得られる。

$$T_k = \frac{\sum_{n=0}^{63} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{63} y_k(n)y_k(n)}} \quad (37)$$

ここで、 $x(n)$ はターゲット信号であり、 $y_k(n)$ は遅延 k における過去のフィルタリングされた励振信号 ($h(n)$ に畳み込まれた過去の励振信号) である。前に述べたように、探索範囲は閉ループピッチの周辺に制限されている。

畳み込み $y_k(n)$ は、探索範囲の最初の遅延について計算され、その他の遅延については、以下の再帰的な関係式を用いて更新される。

$$y_k(n) = y_{k-1}(n-1) + u(-k)h(n) \quad (38)$$

ここで、 $u(n)$, $n=-(231+17), \dots, 63$ は、励振信号バッファである。探索の段階においては、サンプル値 $u(n)$, $n=0, \dots, 63$ は未知であり、これらのサンプル値は64以下のピッチ遅延のために必要とされる。探索を簡略化するために、すべての遅延に式(38)の関係が有効になるように、LP残差が $u(n)$ にコピーされる。

最適な整数ピッチ遅延が決定されると、その整数の周辺の $-3/4$ から $3/4$ までの分数値が、 $1/4$ ずつのステップで調べられる。分数ピッチ探索は、式(37)における正規化相関関数を補間し、その最大値を探索することによって行われる。分数ピッチラグが決定されると、 $v(n)$ が、得られた位相(分数値)で過去の励振信号 $u(n)$ を補間することによって計算される。(補間は、2つのFIRフィルタ(ハミング窓をかけたsinc関数)を用いて行われる。その1つは、 ± 17 で打ち切られたsincを用いて式(37)の項を補間するためのもの、もう1つは、 ± 63 で打ち切られたsincを用いた、過去の励振信号を補間するためのものである。)これらのフィルタは、オーバーサンプルした領域で6000Hzのカットオフ周波数(-3dB)を持つ。これは、この補間フィルタが低域通過型の周波数応答を示すことを意味する。このように、ピッチ遅延が整数の場合においても、適応コードブックの励振信号は、得られた遅延で過去の励振信号を直接コピーしたものではなく、過去の励振信号の低域通過フィルタリングされたものから構成される。さらに、サブフレーム長よりも小さい遅延については、適応コードブック励振信号は、過去の励振信号を繰り返したものではなく、低域通過フィルタリングされて補間された過去の励振信号に基づいて計算される。

広帯域信号におけるピッチ予測性能を高めるために、周波数依存ピッチ予測器が用いられる。周期性は必ずしも周波数領域全体に広がるとは限らないので、この周波数依存ピッチ予測器は広帯域信号では重要である。本アルゴリズムでは、ピッチコードブックパラメータのセットそれぞれに関連した2つの信号のパスがある。その中で、それぞれの信号のパスは、ピッチコードブック探索手段から得られたピッチコードベクトルのピッチ予測誤差を計算するためのピッチ推定誤差計算手段を含む。

この2つのパスの1つは、ピッチコードベクトルをフィルタリングするための低域通過フィルタを含み、ピッチ予測誤差がこの2つの信号パスに対して計算される。最も小さいピッチ予測誤差が計算された信号のパスが、

関連するピッチ利得とともに選択される。

2番目のパスで用いられる低域通過フィルタは、 $B_{LP}(z)=0.18z+0.64+0.18z^{-1}$ の形である。1ビットが、選択されたパスを符号化するために用いられる。

このように、12.65, 14.25, 15.85, 18.25, 19.85, 23.05, および23.85 kbit/s モードについては、適応コードブック $v(n)$ を生成する2つの可能性がある。1番目のパスでは $v(n)=v'(n)$ であり、2番目のパスでは、 $v(n)=\sum_{i=1}^1 b_{LP}(i+1)v'(n+i)$ 、ここで $b_{LP}=[0.18, 0.64, 0.18]$ である。式(40)で定義されるターゲット信号 $x_2(n)$ の最小エネルギーを与えるパスが、フィルタリングされた適応コードブックベクトルとして選択される。6.60 および 8.85kbit/s モードについては、 $v(n)$ は常に $v(n)=\sum_{i=1}^1 b_{LP}(i+1)v'(n+i)$ で与えられる。

次に、適応コードブック利得は、以下の式で得られる。

$$g_p = \frac{\sum_{n=0}^{63} x(n)y(n)}{\sum_{n=0}^{63} y(n)y(n)}, \quad \text{bounded by } 0 \leq g_p \leq 1.2 \quad (39)$$

ここで、 $y(n)=v(n)*h(n)$ はフィルタリングされた適応コードブックベクトル ($v_i(n)$ に対する $H(z)W(z)$ の零状態応答) である。安定性を保証するために、前のサブフレームの適応コードブック利得が小さく、前のサブフレームのLPフィルタがほぼ不安定であった場合、適応コードブック利得 g_p は、0.95以下に制限される。

5. 8 代数的コードブック

5. 8. 1 コードブックの構成

このコードブックの構造は、インタリーブ単一パルス並び換え (ISPP) 設計に基づいている。コードベクトルの64の位置が、それぞれのトラックに16個の位置を配するように位置をインタリーブした4つのトラックに分割される。異なるレートで異なるコードブックは、トラック内に (トラックあたり1から6パルスの) 極性付きパルスを所定数配置することによって構成する。コードブックのインデックス、すなわち符号語は、それぞれのトラックにおけるパルスの位置と極性を示す。このように、復号器における励振ベクトルは、インデックス自身に含まれている情報を通して再構築することができる (テーブルを参照する必要がない) ので、コードブックを蓄積する必要がない。

この利用するコードブックの重要な特徴は、このコードブックが、合成音声の品質を向上させるために特別なスペクトル要素を強調する適応プリフィルタ $F(z)$ が後続する、代数的コードブックから構成される、動的なコードブックであることである。広帯域信号に関連したプリフィルタが用いられ、 $F(z)$ は周期性を強調する部分 $1/(1-0.85z^T)$ と、傾きの部分 $(1-\beta_1 z^{-1})$ の2つの部分からなる。ここで、 T はピッチラグの整数部分であり、 β_1 は前のサブフレームの有声性に関連があり、 $[0.0, 0.5]$ の範囲に制限されている。コードブック探索はコードブック探索に先立ってフィルタ $F(z)$ と重み付き構成フィルタを合成することによって、代数的領域で行われる。したがって、プリフィルタ $F(z)$ を含むようにインパルス応答 $h(n)$ を変更しなくてはならない。すなわち、 $h(n) \leftarrow h(n)*f(n)$ である。

異なるビットレートのコードブックの構造は以下に示される。

5. 8. 1. 1 23.85および23.05 kbit/sモード

このコードブックでは、励振ベクトルは24本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 4 / JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは6本のパルスを持つ。

Table 4 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 23.85 and 23.05 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	$i_0, i_4, i_8, i_{12}, i_{16}, i_{20}$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1, i_5, i_9, i_{13}, i_{17}, i_{21}$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	$i_2, i_6, i_{10}, i_{14}, i_{18}, i_{22}$	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	$i_3, i_7, i_{11}, i_{15}, i_{19}, i_{23}$	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の6本のパルスは、22ビットで符号化される。

これにより、代数的コードに(22+22+22+22)の88ビットが用いられる。

5. 8. 1. 2 19.85 kbit/s モード

このコードブックでは、励振ベクトルは18本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 5 / JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、最初の2つのトラックはそれぞれ5本のパルスを持ち、他のトラックはそれぞれ4本のパルスを持つ。

Table 5 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 19.85 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	$i_0, i_4, i_8, i_{12}, i_{16}$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1, i_5, i_9, i_{13}, i_{17}$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}, i_{14}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}, i_{15}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の5本のパルスは20ビットで符号化される。1トラック内の4本のパルスは16ビットで符号化される。

これにより、代数的コードに(20+20+16+16)の72ビットが用いられる。

5. 8. 1. 3 18.25 kbit/sモード

このコードブックでは、励振ベクトルは16本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 6 / JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは4本のパルスを持つ。

Table 6 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 18.25 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0, i_4, i_8, i_{12}	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9, i_{13}	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}, i_{14}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}, i_{15}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の4本のパルスは、16ビットで符号化される。

これにより、代数的コードに $(16+16+16+16)$ の64ビットが用いられる。

5. 8. 1. 4 15.85 kbit/sモード

このコードブックでは、励振ベクトルは12本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 7 / JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは3本のパルスを持つ。

Table 7 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 15.85 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0, i_4, i_8	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の3本のパルスは、13ビットで符号化される。

これにより、代数的コードに $(13+13+13+13)$ の52ビットが用いられる。

5. 8. 1. 5 14.25 kbit/sモード

このコードブックでは、励振ベクトルは10本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 8 / JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは2本あるいは3本のパルスを持つ。

Table 8 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 14.25 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0, i_4, i_8	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の2本のパルス位置は、8ビット（パルスごとの位置に4ビット）で符号化され、トラック内

の1番目のパルスの極性が1ビットで符号化される。

1トラック内の3本のパルスは、13ビットで符号化される。

これにより、代数的コードに(13+13+9+9)の44ビットが用いられる。

5. 8. 1. 6 12.65 kbit/sモード

このコードブックでは、励振ベクトルは8本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 9/JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは2本のパルスを持つ。

Table 9/JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 12.65 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0, i_4	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内の2本のパルス位置は、8ビット（合計32ビット、パルスごとの位置に4ビット）で符号化され、トラック内の1番目のパルスの極性が1ビット（合計4ビット）で符号化される。これにより、代数的コードに36ビットが用いられる。

5. 8. 1. 7 8.85 kbit/sモード

このコードブックでは、励振ベクトルは4本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 10/JT-G722.2に示すように、サブフレーム内の64の位置が4つのトラックに分割され、各トラックは1本のパルスを持つ。

Table 10/JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 8.85 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

1トラック内のパルス位置は、4ビットで符号化され、トラック内のパルスの極性は1ビットで符号化される。これにより、代数的コードに20ビットが用いられる。

5. 8. 1. 8 6.60 kbit/sモード

このコードブックでは、励振ベクトルは2本の非零のパルスを持つ。すべてのパルスは+1あるいは-1の振幅を持つことができる。Table 11/JT-G722.2に示すように、サブフレーム内の64の位置が2つのトラックに分割され、各トラックは1本のパルスを持つ。

Table 1 1 / JT-G722.2 Potential positions of individual pulses in the algebraic codebook, 6.60 kbit/s
(ITU-T G.722.2)

Track	Pulse	Positions
1	i_0	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62
2	i_1	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63

1トラック内のパルス位置は5ビットで符号化され、トラック内のパルスの極性は1ビットで符号化される。これにより、代数的コードに12ビットが用いられる。

5. 8. 2 パルスのインデックス化

前節では1トラック中のパルスを符号化するのに必要なビット数が与えられた。この節では、1トラックあたり1本から6本のパルスを符号化するための手続きについて述べる。以下の記述は1サブフレームあたり4トラックで、1トラックに16通りのパルス位置があり、パルス間隔は4の場合を対象とする（これは6.6 kbit/sモード以外のすべてのモードに対応する）。

1トラックあたり1本の極性付きパルスの符号化

パルスの位置インデックスは4ビット、極性インデックスは1ビットで符号化される。位置インデックスはパルス間隔で除算（整数除算）したときのサブフレーム内のパルス位置で与えられる。除算の剰余はトラックのインデックスを与える。例えば、位置31のパルスは $31/4=7$ の位置インデックスを持ち、インデックス3のトラック（すなわち第4番目のトラック）に属する。

ここでの極性インデックスは正の極性の場合に0、負の極性の場合に1と設定する。

極性付きパルスのインデックスは次式で与えられる。

$$I_{1p} = p + s \times 2^M$$

ここで、 p は位置インデックス、 s は極性インデックス、 $M=4$ はトラックあたりの位置インデックスのビット数である。

1トラックあたり2本の極性付きパルスの符号化

$K=2^M$ （ここでは $M=4$ ）通りの位置を取り得るトラックあたりに2本のパルスの場合、それぞれのパルスは極性に1ビット、位置に M ビット必要なため、合計 $2M+2$ ビットとなる。しかしながら、パルスの順序に重要性がないために多少冗長性が存在している。例えば、最初のパルスを位置 p に置き、第2番目のパルスを位置 q に置くことと、最初のパルスを位置 q に置き、第2番目のパルスを位置 p に置くことは等価である。1つの極性だけを符号化して第2番目の極性はインデックスの中の位置の順序から導き出すことにより1ビットを節約することができる。ここではインデックスは次式で与えられる。

$$I_{2p} = p_1 + p_0 \times 2^M + s \times 2^{2M}$$

ここで、 s は位置インデックス p_0 のパルスの極性インデックスである。もし2本のパルスの極性が等しければ小さなパルス位置は p_0 に、大きなパルス位置は p_1 に設定される。逆に2つの極性が等しくない場合、大きなパルス位置は p_0 に、小さなパルス位置は p_1 に設定される。復号器では位置 p_0 のパルスの極性は直ちに利用でき

る。第2番目の極性はパルスの順序から導き出される。もし p_0 が p_1 より大きい場合は位置 p_1 のパルスの極性は位置 p_0 のパルスの極性と逆であり、そうでない場合は2つの極性は同じに設定される。

1トラックあたり3本の極性付きパルスの符号化

トラックあたり3本のパルスの場合、2本のパルスの場合と同様の手法を使うことができる。 2^M 通りの位置を持つトラックには、 $3M+3$ ビットではなく $3M+1$ ビットが必要となる。パルスをインデックス化する簡単な方法は、トラック位置を2セクション（または半分）に分割し、少なくとも2本のパルスを含む方のセクションを識別することである。そのセクションにおける位置の数は $K/2 = 2^M/2 = 2^{M-1}$ であり、 $M-1$ ビットで表すことができる。少なくとも2本のパルスを含むセクションにある2本のパルスは2本の極性付きパルスの符号化手続きにより $2(M-1)+1$ ビットで符号化され、トラック内の（どちらかのセクションの）どこにでも存在し得る残りの1本のパルスは $M+1$ ビットで符号化される。最後に、2本のパルスを含むセクションのインデックスが1ビットで符号化される。従って、必要ビット数の合計は $2(M-1)+1 + M+1 + 1 = 3M+1$ となる。

2本のパルスの位置が同じセクションにあるかどうかをチェックする簡単な方法は、2本のパルスの位置インデックスの最上位ビット（MSB）が等しいかそうでないかをチェックすることである。MSBが0のときは位置がトラックの前半（0～7）に属することを意味し、MSBが1のときはトラックの後半（8～15）に属することを意味する。もし2本のパルスが後半に属する場合は、 $2 \times 3 + 1$ ビットを使って符号化する前に（0～7）の範囲にパルスをシフトする必要がある。これは $M-1$ 個の1から成る（この場合は数字の7に対応する）マスクで $M-1$ 個の最下位ビット（LSB）をマスクすることにより実現される。

3本の極性付きパルスのインデックスは次式で与えられる。

$$I_{3p} = I_{2p} + k \times 2^{2M-1} + I_{1p} \times 2^{2M}$$

ここで、 I_{2p} は同じセクションにある2本のパルスのインデックス、 k はセクションのインデックス（0または1）を示し、 I_{1p} はトラック内の第3番目のパルスのインデックスである。

1トラックあたり4本の極性付きパルスの符号化

長さ $K=2^M$ のトラック内の4本の極性付きパルスは $4M$ ビットで符号化できる。3本のパルスの場合と同様に、トラック内の K 個の位置は各セクションが $K/2=8$ 個の位置を含む2つのセクション（前半と後半）に分割される。ここでは、位置0から位置 $K/2-1$ までをセクションA、位置 $K/2$ から位置 $K-1$ までをセクションBとする。それぞれのセクションは0本から4本のパルスを含むことができる。以下のテーブルは、それぞれのセクションで取り得るパルス数を表す5つのケースを示している。

case	Pulses in Section A	Pulses in Section B	Bits needed
0	0	4	$4M-3$
1	1	3	$4M-2$
2	2	2	$4M-2$
3	3	1	$4M-2$
4	4	0	$4M-3$

ケース0または4では、長さ $K/2=2^{M-1}$ のセクションにおける4本のパルスは $4(M-1)+1=4M-3$ ビットで符号化できる（このことは後で説明する）。

ケース1または3では、長さ $K/2=2^{M-1}$ のセクションにおける1本のパルスは $M-1+1=M$ ビットで符号化でき、もう一方のセクションにある3本のパルスは $3(M-1)+1=3M-2$ ビットで符号化できる。このため合計は $M+3M-2$

= $4M-2$ ビットとなる。

ケース 2 では、長さ $K/2=2^{M-1}$ のセクションにおけるパルスは $2(M-1)+1 = 2M-1$ ビットで符号化できる。従って、両方のセクションに対しては $2(2M-1) = 4M-2$ ビットが必要とされる。

いま、ケース 0 と 4 を組み合わせると仮定すると、ケースのインデックスは 2 ビット（4 つの可能なケース）で符号化できる。その場合、ケース 1、2、3 で必要なビット数は $4M-2$ なので、合計は $4M-2+2 = 4M$ ビットとなる。ケース 0 または 4 では、どちらのケースであるかの識別に 1 ビット、セクション内の 4 本のパルスの符号化に $4M-3$ ビットが必要である。ケースのインデックスに対して必要な 2 ビットを加えると、合計は $1+4M-3+2 = 4M$ ビットとなる。

4 本の極性付きパルスのインデックスは次式で与えられる。

$$I_{4p} = I_{AB} + k \times 2^{4M-2}$$

ここで、 k はケースのインデックス（2 ビット）、 I_{AB} はそれぞれ個別のケースでの両セクションにおけるパルスのインデックスである。

ケース 0 と 4 では、 I_{AB} は次式で与えられる。

$$I_{AB,0,4} = I_{4p_section} + j \times 2^{4M-3}$$

ここで、 j は 4 本のパルスを持つセクションを識別する 1 ビットのインデックス、 $I_{4p_section}$ はそのセクションにおける 4 本のパルスのインデックス（ $4M-3$ ビット必要）である。

ケース 1 では、 I_{AB} は次式で与えられる。

$$I_{AB,1} = I_{3p_B} + I_{1p_A} \times 2^{3(M-1)+1}$$

ここで、 I_{3p_B} はセクション B における 3 本のパルスのインデックス（ $3(M-1)+1$ ビット）、 I_{1p_A} はセクション A における 1 本のパルスのインデックス（ $(M-1)+1$ ビット）である。

ケース 2 では、 I_{AB} は次式で与えられる。

$$I_{AB,2} = I_{2p_B} + I_{2p_A} \times 2^{2(M-1)+1}$$

ここで、 I_{2p_B} はセクション B における 2 本のパルスのインデックス（ $2(M-1)+1$ ビット）、 I_{2p_A} はセクション A における 2 本のパルスのインデックス（ $2(M-1)+1$ ビット）である。

最後に、ケース 3 では、 I_{AB} は次式で与えられる。

$$I_{AB,3} = I_{1p_B} + I_{3p_A} \times 2^M$$

ここで、 I_{1p_B} はセクション B における 1 本のパルスのインデックス（ $(M-1)+1$ ビット）、 I_{3p_A} はセクション A における 3 本のパルスのインデックス（ $3(M-1)+1$ ビット）である。

ケース 0 と 4 について、1 セクションにある 4 本のパルスは $4(M-1)+1$ ビットで符号化されると述べた。これは、セクションをさらに長さ $K/4=2^{M-2}$ （この場合 4）の 2 つのサブセクションに分割すること、少なくとも 2 本のパルスを含む方のサブセクションを識別すること、そのサブセクション内の 2 本のパルスを $2(M-2)+1=2M-3$ ビットで符号化すること、少なくとも 2 本のパルスを含むサブセクションのインデックスを 1 ビットで符号化すること、そして残りの 2 本のパルスはセクション内のどこにでも存在し得ると仮定して

$2(M-1)+1=2M-1$ ビットで符号化すること、でなされる。このため、合計は $(2M-3)+(1)+(2M-1)=4M-3$ ビットとなる。

1トラックあたり5本の極性付きパルスの符号化

長さ $K=2^M$ のトラック内の5本の極性付きパルスは $5M$ ビットで符号化できる。4本のパルスの場合と同様に、トラック中の K 個の位置は2つのセクションAとBに分割される。それぞれのセクションは0本から5本のパルスを含むことが可能である。5本のパルスを符号化するための簡単なアプローチは、少なくとも3本のパルスを含む方のセクションを識別し、そのセクションの3本のパルスを $3(M-1)+1=3M-2$ ビットで符号化し、トラック全体の中で残る2本のパルスを $2M+1$ ビットで符号化する。これで $5M-1$ ビットとなる。さらに、少なくとも3本のパルスを含むセクションの識別に1ビットが必要である。従って、5本の極性付きパルスを符号化するためには合計 $5M$ ビットが必要となる。

5本の極性付きパルスのインデックスは次式で与えられる。

$$I_{5p} = I_{2p} + I_{3p} \times 2^{2M} + k \times 2^{5M-1}$$

ここで、 k は少なくとも3本のパルスを含むセクションのインデックス、 I_{3p} はそのセクションにある3本のパルスのインデックス ($3(M-1)+1$ ビット)、 I_{2p} はトラック内の残りの2本のパルスのインデックス ($2M+1$ ビット) である。

1トラックあたり6本の極性付きパルスの符号化

長さ $K=2^M$ のトラック内の6本の極性付きパルスは $6M-2$ ビットで符号化される。5本のパルスの場合と同様に、トラック中の K 個の位置は2つのセクションAとBに分割される。それぞれのセクションは0本から6本のパルスを含むことが可能である。以下のテーブルは、それぞれのセクションで取り得るパルス数を表した7つのケースを示している。

case	Pulses in Section A	Pulses in Section B	Bits needed
0	0	6	$6M-5$
1	1	5	$6M-5$
2	2	4	$6M-5$
3	3	3	$6M-4$
4	4	2	$6M-5$
5	5	1	$6M-5$
6	6	0	$6M-5$

6本のパルスが違うセクションにあることを除けば、ケース0とケース6は類似している。同様に、ケース1と5、ケース2と4はより多くのパルスを含むセクションが違うだけである。従って、これらのケースは対にして、パルスをより多く含むセクションを識別するために特別な1ビットを割り当てるようにすることができる。これらのケースは元々 $6M-5$ ビットを要するので、セクションを識別するビットを加味すると、対にしたケースでは $6M-4$ ビット必要である。従って、(0,6)、(1,5)、(2,4)、(3)というケースの対の4状態があることになり、これには2ビットが必要となる。これにより、6本の極性付きパルスは合計 $6M-4+2=6M-2$ ビットとなる。

ケース0と6では、6本のパルスを含むセクションの識別に1ビットが必要とされる。セクション内の5本のパルスは $5(M-1)$ ビットで (このセクション内に限定なので) 符号化され、残りの1本のパルスは $(M-1)+1$ ビットで符号化される。従って、この対にしたケースでは合計 $1+5(M-1)+M=6M-4$ ビットが必要である。さらに対

にしたケースの状態の符号化に2ビット必要なので、合計は $6M-2$ ビットとなる。この場合の6本のパルスのインデックスは次式で与えられる。

$$I_{6p} = I_{1p} + I_{5p} \times 2^M + j \times 2^{6M-5} + k \times 2^{6M-4}$$

ここで、 k は対にしたケースのインデックス（2ビット）、 j は6本のパルスを含むセクションのインデックス（1ビット）、 I_{5p} はそのセクション内の5本のパルスのインデックス（ $5(M-1)$ ビット）、 I_{1p} はセクション内の残りの1本のパルスのインデックス（ $(M-1)+1$ ビット）である。

ケース1と5では、5本のパルスを含むセクションの識別に1ビットが必要とされる。セクション内の5本のパルスは $5(M-1)$ ビットで符号化され、もう一方のセクションにある1本のパルスは $(M-1)+1$ ビットで符号化される。この対にしたケースでは6本のパルスのインデックスは次式で与えられる。

$$I_{6p} = I_{1p} + I_{5p} \times 2^M + j \times 2^{6M-5} + k \times 2^{6M-4}$$

ここで、 k は対にしたケースのインデックス（2ビット）、 j は5本のパルスを含むセクションのインデックス（1ビット）、 I_{5p} はそのセクション内の5本のパルスのインデックス（ $5(M-1)$ ビット）、 I_{1p} はもう一方のセクション内の1本のパルスのインデックス（ $(M-1)+1$ ビット）である。

ケース2または4では、4本のパルスを含むセクションの識別に1ビットが必要とされる。セクション内の4本のパルスは $4(M-1)$ ビットで符号化され、もう一方のセクションにある2本のパルスは $2(M-1)+1$ ビットで符号化される。この対にしたケースでは6本のパルスのインデックスは次式で与えられる。

$$I_{6p} = I_{2p} + I_{4p} \times 2^{2(M-1)+1} + j \times 2^{6M-5} + k \times 2^{6M-4}$$

ここで、 k は対にしたケースのインデックス（2ビット）、 j は4本のパルスを含むセクションのインデックス（1ビット）、 I_{4p} はそのセクション内の4本のパルスのインデックス（ $4(M-1)$ ビット）、 I_{2p} はもう一方のセクション内の2本のパルスのインデックス（ $2(M-1)+1$ ビット）である。

ケース3では、それぞれのセクションにある3本のパルスがそれぞれ $3(M-1)+1$ ビットで符号化される。このケースでは6本のパルスのインデックスは次式で与えられる。

$$I_{6p} = I_{3pB} + I_{3pA} \times 2^{3(M-1)+1} + k \times 2^{6M-4}$$

ここで、 k は対にしたケースのインデックス（2ビット）、 I_{3pB} はセクションB内の3本のパルスのインデックス（ $3(M-1)+1$ ビット）、 I_{3pA} はセクションA内の3本のパルスのインデックス（ $3(M-1)+1$ ビット）である。

5. 8. 3 コードブック探索

代数的コードブックは、重み付き入力音声と重み付き合成音声との間の平均自乗誤差を最小化することにより探索される。閉ループピッチ探索で使用されたターゲット信号は、適応コードブック寄与分を減ずることにより更新される。

$$x_2(n) = x(n) - g_p y(n), \quad n = 0, \dots, 63 \quad (40)$$

ここで、 $y(n) = v(n) * h(n)$ はフィルタ処理された適応コードブックベクトルであり、 g_p は量子化されていない適応コードブック利得である。

行列 \mathbf{H} は、主対角成分に $h(0)$ を、それより下位の対角成分に $h(1), \dots, h(63)$ を持つ下三角テプリッツ畳込み行

列として定義される。 $\mathbf{d} = \mathbf{H}'\mathbf{x}_2$ はターゲット信号 $x_2(n)$ とインパルス応答 $h(n)$ との相関（バックワードフィルタ処理されたターゲットベクトルとしても知られる）、 $\Phi = \mathbf{H}'\mathbf{H}$ は $h(n)$ の相関からなる行列である。ベクトル \mathbf{d} の要素は次式で計算される。

$$d(n) = \sum_{i=n}^{63} x_2(i)h(i-n), \quad n = 0, \dots, 63 \quad (41)$$

対称行列 Φ の要素は次式で計算される。

$$\phi(i, j) = \sum_{n=j}^{63} h(n-i)h(n-j), \quad i = 0, \dots, 63, \quad j = i, \dots, 63 \quad (42)$$

\mathbf{c}_k をインデックス k の代数的コードベクトルとすると、代数的コードブックは次の探索基準を最大にする k を求めることによって探索される。

$$Q_k = \frac{(\mathbf{x}_2' \mathbf{H} \mathbf{c}_k)^2}{\mathbf{c}_k' \mathbf{H}' \mathbf{H} \mathbf{c}_k} = \frac{(\mathbf{d}' \mathbf{c}_k)^2}{\mathbf{c}_k' \Phi \mathbf{c}_k} = \frac{(R_k)^2}{E_k} \quad (43)$$

ベクトル \mathbf{d} と行列 Φ は通常、コードブック探索の前に計算される。

代数的構造を持つコードブックは、励振ベクトル \mathbf{c}_k が少数の非零のパルスしか含まないため、高速探索手順が可能である。式(43)の分子における相関は次式によって与えられる。

$$C = \sum_{i=0}^{N_p-1} a_i d(m_i) \quad (44)$$

ここで、 m_i は i 番目のパルスの位置、 a_i はその振幅、 N_p はパルスの数である。式(43)の分母におけるエネルギーは次式によって与えられる。

$$E = \sum_{i=0}^{N_p-1} \phi(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} a_i a_j \phi(m_i, m_j) \quad (45)$$

探索手順を簡略化するために、パルス振幅はあるリファレンス信号 $b(n)$ に基づいて前もって決定される。このいわゆる信号選択パルス振幅アプローチでは、位置 i におけるパルスの極性はその位置でのリファレンス信号の極性と同一に設定される。ここでは、リファレンス信号 $b(n)$ は次式で与えられる。

$$b(n) = \sqrt{\frac{E_d}{E_r}} r_{LTP}(n) + \alpha d(n) \quad (46)$$

ここで、 $E_d = \mathbf{d}'\mathbf{d}$ は信号 $d(n)$ のエネルギー、 $E_r = \mathbf{r}_{LTP}'\mathbf{r}_{LTP}$ は長期予測後の残差信号 $r_{LTP}(n)$ のエネルギーである。スケーリングファクタ α は、リファレンス信号の $d(n)$ への依存量を制御するもので、ビットレートの増加に応じて小さくされる。ここでは、6.6 と 8.85 kbit/s モードで $\alpha = 2$ 、12.65 と 14.25 と 15.85 kbit/s モードで $\alpha = 1$ 、18.25 kbit/s モードで $\alpha = 0.8$ 、19.85 kbit/s モードで $\alpha = 0.75$ 、23.05 と 23.85 kbit/s モードで $\alpha = 0.5$ とする。

探索を簡単にするため、信号 $d(n)$ と行列 Φ は前もって選定した極性を組み込むために修正される。いま $b(n)$ の極性を含むベクトルを $s_b(n)$ とすると、修正された信号 $d'(n)$ は次式で与えられる。

$$d'(n) = s_b(n)d(n), \quad n=0, \dots, N-1$$

また、修正された自己相関行列 Φ' は次式で与えられる。

$$\phi'(i, j) = s_b(i)s_b(j)\phi(i, j), \quad i=0, \dots, N-1, \quad j=i, \dots, N-1$$

こうして、探索基準 Q_k の分子における相関は次式で与えられる。

$$R = \sum_{i=0}^{N_p-1} d^i(m_i)$$

また、探索基準 Q_k の分母におけるエネルギーは次式で与えられる。

$$E = \sum_{i=0}^{N_p-1} \phi'(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \phi'(m_i, m_j)$$

ここでの探索の目標は、パルスの振幅が上述のようにして選択されたと仮定して、 N_p 本のパルスの位置の最適な組み合わせをもつコードベクトルを決定することである。基本的な選定の基準は上で述べた比 Q_k の最大化である。

探索の計算量を低減するため、縦形木探索として知られている高速探索手順が使用され、1度で N_m 本のパルスのパルス位置が決定される。より正確には、 N_p 本の利用可能なパルスが、それぞれ N_m 本のパルスから成る M 個の空でないサブセットに $N_1+N_2+\dots+N_m+\dots+N_M = N_p$ となるように分割される。初めに考慮される $J = N_1+N_2+\dots+N_{m-1}$ 個のパルス位置の特別な選択は、レベル m のパス、または長さ J のパスと呼ばれる。 J 個のパルス位置のパスの基本的な基準は、関係する J 本のパルスだけを考慮したときの比 $Q_k(J)$ である。

探索はサブセット#1 から開始し、木の第 m 番目のレベルでサブセット m が探索される木構造に従い、次のサブセットに進む。レベル1での探索の目的は、レベル1の木のノードである長さ N_1 の1個または数個の候補パスを決定するためのサブセット#1 に属する N_1 本のパルスとそれらの有効な位置を検討することである。レベル $m-1$ の各終端ノードのパスは、 N_m 本の新しいパルスとその有効位置を考慮することにより、レベル m で長さ $N_1+N_2+\dots+N_m$ に拡張される。レベル m のノードを構築するために1個または数個の拡張された候補パスが決定される。最適なコードベクトルは、レベル M の全ノードについて基準 $Q_k(N_p)$ を最大化する長さ N_p のパスに対応する。

ここでは、縦形木探索手順の特別な形を使うことにする。そこでは2本のパルスが同時に探索され（すなわち $N_m=2$ ）、これら2本のパルスは2つの連続するトラックに属している。さらに、行列 Φ が予め計算されて蓄積されると仮定すると $N \times N$ ワード ($64 \times 64 = 4k$ ワード)のメモリを要するが、これに代って、メモリ要求を低減するメモリ効率のよいアプローチが使用される。このアプローチでは、相関行列の必要とされる要素の部分だけを予め計算し蓄積する方法で探索が行われる。この部分というのは、 $\phi(j, j), j=0, \dots, N-1$ に対応する相関（すなわち行列 Φ の主対角の要素）だけでなく、連続するトラックで起こり得るパルス位置に対応するインパルス応答の相関に相当する。

計算量を低減するために、2本のパルスの可能な組み合わせを試す際に、第1のパルスの立ち得る位置について限定数だけ試行が行われる。さらには、多数本のパルスの場合、探索木のより高いレベルにある幾つかのパルスは固定される。第1のパルスに対し、どの潜在的なパルス位置が考慮されるかを効率的に推測するため、または幾つかのパルス位置を固定にするため、音声に関係する信号に基づいた「パルス位置尤度推定ベクトル」 \mathbf{b} が使用される。この推定ベクトル \mathbf{b} の第 p 番目の要素 $b(p)$ は、探索中の最適コードベクトルにおいて、ある1本のパルスが位置 p ($p = 0, 1, \dots, N-1$) を占有する確率を特徴づけている。ここで、推定ベクトル \mathbf{b} は振幅の予備選択で使用されたものと同一のベクトルで、式(46)で与えられている。

すべてのビットレートのモードに対し、探索手順は類似している。1度に2本のパルスが探索され、これら2本のパルスは常に連続したトラックに対応している。すなわち、2本の探索されたパルスはトラック $T_0-T_1, T_1-T_2, T_2-T_3, T_3-T_0$ のいずれかに存在する。

位置を探索する前に、パルスの立ち得る位置 n での極性は、その位置における $b(n)$ の極性と 同じに設定される。次に、前もって決定した極性を含めることにより、上述したように修正された信号 $d'(n)$ が計算される。

最初の2本のパルス(第1番目の木のレベル)について、探索基準の分子における相関は次式で与えられる。

$$R = d'(m_0) + d'(m_1)$$

また、探索基準 Q_k の分母におけるエネルギーは次式で与えられる。

$$E = \phi'(m_0, m_0) + \phi'(m_1, m_1) + 2\phi'(m_0, m_1)$$

ここで、相関 $\phi'(m_i, m_j)$ は位置 m_i と m_j において、予備選択された極性を含むように修正されたものである。

次のレベルでは、新しい2本のパルスの寄与分を加えることで分子と分母が更新される。ある木のレベルで、2つの連続するトラックからの位置 m_k と m_{k+1} を持つ2本の新しいパルスが探索されると仮定すると、 R の更新値は次式で与えられる。

$$R = R + d'(m_k) + d'(m_{k+1}) \quad (47)$$

また、更新されたエネルギーは次式で与えられる。

$$E = E + \phi'(m_k, m_k) + \phi'(m_{k+1}, m_{k+1}) + 2\phi'(m_k, m_{k+1}) + 2R_{hv}(m_k) + 2R_{hv}(m_{k+1}) \quad (48)$$

ここで $R_{hv}(m)$ は、インパルス応答 $h(n)$ と前で決定した位置に遅延させたインパルス応答の加算を含むベクトル $v_h(n)$ との相関である。すなわち、

$$v_h(n) = \sum_{i=0}^{k-1} h(n - m_i)$$

$$R_{hv}(m) = \sum_{n=m}^{N-1} h(n)v_h(n - m)$$

各木のレベルで、試行される2つのトラックのそれぞれの中で可能なすべての位置に対して $R_{hv}(m)$ の値がその都度計算される。式(48)からわかるように、行列 Φ の対角要素に対応する相関 $\phi'(m_k, m_k)$ (64ワード)とともに、2つの連続するトラックのパルス位置に対応する相関 $\phi'(m_k, m_{k+1})$ (4×16×16ワード)だけ蓄積が必要である。従って、ここでの代数的構造におけるメモリの要求は 64×64=4096ワードの代わりに 1088ワードとなる。

異なるビットレートモードで探索手順は類似している。その違いはパルス数にあり、従って木探索のレベル数に違いがある。異なるコードブック間で探索の計算量を同程度にするため、試行する位置の数は同程度にされる。

例として 12.65 kbit/s モードでの探索について述べる。このモードでは各トラックに2本のパルスが置かれ、長さ 64のサブフレームあたり合計8本のパルスとなる。一度に2本のパルスが探索され、これら2本のパルスは常に連続したトラックに対応している。すなわち、2本のパルスは $T_0-T_1, T_1-T_2, T_2-T_3, T_3-T_0$ のいずれかのトラックにある。木はこの場合4レベルある。第1番目のレベルで、パルス P_0 がトラック T_0 に、パルス P_1 がトラック T_1 に割り当てられる。このレベルでは探索は行われず、2つのパルス位置は各トラックでの $b(n)$ の最大値の位置に設定される。第2番目のレベルで、パルス P_2 がトラック T_2 に、パルス P_3 がトラック T_3 に割り当てられる。パルス P_2 の4つの位置がパルス P_3 の16通りすべての位置に対して試行される。試行されるパルス P_2 の4つの位置は、トラックの $b(n)$ の最大値に基づいて決定される。第3番目のレベルで、パルス P_4 がトラック T_1 に、パルス P_5 がトラック T_2 に割り当てられる。パルス P_4 の8つの位置がパルス P_5 の16通りすべての位置に対して試行される。前の探索レベルと同様に、試行されるパルス P_4 の8つの位置は、トラックの $b(n)$ の最大

値に基づいて決定される。第4番目のレベルで、パルス P_6 がトラック T_3 に、パルス P_7 がトラック T_0 に割り当てられる。パルス P_6 の8つの位置がパルス P_7 の16通りすべての位置に対して試行される。このように、試行される組み合わせの総数は $4 \times 16 + 8 \times 16 + 8 \times 16 = 320$ となる。この全体プロセスは、パルスを異なるトラックに割り当てることにより4回繰り返される。例えば第2番目の繰り返しでは、パルス P_0 からパルス P_7 は、それぞれトラック $T_1, T_2, T_3, T_0, T_2, T_3, T_0$ および T_1 に割り当てられる。こうして、試行される位置の組み合わせの総数は $4 \times 320 = 1280$ となる。

別な探索の例として、15.85 kbit/s モードでは各トラックに3本のパルスが置かれ、合計12本のパルスとなる。木探索は6レベルあり、各レベルで2本のパルスが探索される。最初の2つのレベルでは、4本のパルスが $b(n)$ の最大値の位置に設定される。次の4つのレベルでは、試行される組み合わせの数は、それぞれ $4 \times 16, 6 \times 16, 8 \times 16, 8 \times 16$ である。4回の繰り返しを用いることで、合計 $4 \times 26 \times 16 = 1664$ の組み合わせとなる。

5.9 適応および固定コードブック利得の量子化

適応コードブック利得（ピッチ利得）と固定（代数的）コードブック利得は、8.85 および 6.60 kbit/s モードでは6ビットの、他のすべてのモードでは7ビットのコードブックを用いてベクトル量子化される。

固定コードブック利得量子化は、固定係数によるMA予測を用いて実行される。4次のMA予測は、励振エネルギーに関して、次のように実行される。サブフレーム n における平均除去励振エネルギー (dB) を $E(n)$ とすると、

$$E(n) = 10 \log \left(\frac{1}{N} g_c^2 \sum_{i=0}^{N-1} c^2(i) \right) - \bar{E} \quad (49)$$

で与えられ、 $N=64$ はサブフレームサイズ、 $c(i)$ は固定コードブック励振信号、 $\bar{E} = 30$ dB は励振エネルギーの平均値である。予測エネルギーは、

$$\tilde{E}(n) = \sum_{i=1}^4 b_i \hat{R}(n-i) \quad (50)$$

で与えられ、 $[b_1 b_2 b_3 b_4] = [0.5, 0.4, 0.3, 0.2]$ はMA予測係数、 $\hat{R}(k)$ はサブフレーム k における量子化エネルギー予測誤差である。予測エネルギーは ($E(n)$ を $\tilde{E}(n)$ 、 g_c を g'_c とすることにより) 式(49)の予測固定コードブック利得 g'_c の計算に用いられ、次のようになる。まず、平均励振エネルギーは、

$$E_i = 10 \log \left(\frac{1}{N} \sum_{i=0}^{N-1} c^2(i) \right) \quad (51)$$

によって得られ、そのとき予測利得 g'_c は、

$$g'_c = 10^{0.05(\tilde{E}(n) + \bar{E} - E_i)} \quad (52)$$

で得られる。利得 g_c と推定後の利得 g'_c との間の補正係数は、

$$\gamma = g_c / g'_c \quad (53)$$

で与えられる。予測誤差は、

$$R(n) = E(n) - \tilde{E}(n) = 20 \log(\gamma) \quad (54)$$

で与えられる。

ピッチ利得 g_p と補正係数 γ は、8.85 および 6.60 kbit/s モードでは 6 ビットのコードブックを、その他のモードでは 7 ビットのコードブックを用いて、合わせてベクトル量子化される。利得コードブック探索は、入力音声と、以下で与えられる再生音声との間の、重み付き誤差の自乗平均を最小化することにより実行される。

$$E = x^t x + g_p^2 y^t y + g_c^2 z^t z - 2g_p x^t y - 2g_c x^t z + 2g_p g_c y^t z \quad (55)$$

ここで、 x はターゲットベクトル、 y はフィルタ処理後の適応コードブックベクトル、 z はフィルタ処理後の固定コードブックベクトルである。(コードブックの各利得ベクトルもまた、量子化エネルギー予測誤差相当の成分を持つ。) 選択済の利得と関連する量子化エネルギー予測誤差は、 $\hat{R}(n)$ の更新に使用される。探索では、量子化されていないピッチ利得 g_p に接近している 64 のコードベクトルだけが考慮される。

5. 10 メモリ更新

合成フィルタと重み付けフィルタの状態の更新は、次のサブフレーム内のターゲット信号を計算するために必要とされる。

2つの利得が量子化された後、現在のサブフレーム中の励振信号 $u(n)$ は、

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n), \quad n = 0, \dots, 63 \quad (56)$$

で求められ、ここで \hat{g}_p と \hat{g}_c はそれぞれ、量子化適応コードブック利得と量子化固定コードブック利得、 $v_i(n)$ は適応コードブックベクトル (補間された過去の励振信号)、 $c(n)$ は固定コードブックベクトル (ピッチシャープニングを含む代数的コード) である。フィルタの状態は、64 サンプルのサブフレーム、およびそのフィルタの状態を保存するためのフィルタ $1/\hat{A}(z)$ と $A(z/\gamma_1)H_{de-emph}(z)$ を通った信号 $r(n)-u(n)$ (残差信号と励振信号との間の差) のフィルタリングにより更新可能である。これには 3つのフィルタ処理を必要とする。ただ 1つのフィルタのみを必要とする単純な手法を以下に示す。局所合成音声 $\hat{s}(n)$ は、 $1/\hat{A}(z)$ を通してフィルタ処理した励振信号によって計算される。入力 $r(n)-u(n)$ によるフィルタ出力は $e(n) = s(n) - \hat{s}(n)$ と等価である。よって合成フィルタ $1/\hat{A}(z)$ の状態は、 $e(n)$, $n = 48, \dots, 63$ で与えられる。聴覚重み付き誤差 $e_w(n)$ を見つけるためのフィルタ $A(z/\gamma_1)H_{de-emph}(z)$ の状態更新は、このフィルタを通して誤差信号 $e(n)$ のフィルタ処理によって行うことができる。しかしながら、信号 $e_w(n)$ は、次の式によって同等に得ることが可能である。

$$e_w(n) = x(n) - \hat{g}_p v(n) + \hat{g}_c z(n) \quad (57)$$

信号 $x(n)$ 、 $y(n)$ と $z(n)$ が利用できるため、重み付けフィルタの状態は、 $n = 48, \dots, 63$ の間で、式(57)に示す $e_w(n)$ の計算によって更新される。これは 2つのフィルタ処理を省くことになる。

5. 11 高域利得生成

23.85 kbit/s モード用の高域利得の計算により、16kHz 入力音声は、通過帯域が 6.4k から 7kHz である帯域通過 FIR フィルタ $H_{HB}(z)$ を通ってフィルタ処理される。高域利得 g_{HB} は、

$$g_{HB} = \frac{\sum_{i=0}^{63} (s_{HB}(i))^2}{\sum_{i=0}^{63} (s_{HB2}(i))^2} \quad (58)$$

によって得られ、ここで $s_{HB}(i)$ は、帯域通過フィルタで処理された入力音声で、 $s_{HB2}(i)$ は、6. 3. 2. 2 節で述べる高域合成フィルタ $A_{HB}(z)$ で処理された高域励振信号 $u_{HB2}(i)$ によって得られた、高域音声合成である。

6. 復号器の詳細

復号器の機能は、送信されたパラメータ（LPパラメータ、適応コードブックベクトル、適応コードブック利得、固定コードブックベクトル、固定コードブック利得、および高域利得）を復号し、再生音声を得るための合成を実行することからなる。そして再生音声は、後処理とアップサンプル（とアップスケール）がなされる。最後に、6k から 7kHz の周波数帯域の高域信号が生成される。復号器の信号フローは Figure 3 / JT-G722.2 に示される。

6. 1 復号と音声合成

復号の過程は、以下の項目からなる。

LPフィルタパラメータの復号： I S P量子化の受信インデックスは、量子化 I S Pベクトルの再生に使用される。5. 2. 6節で述べた補間は、4つの補間した I S Pベクトル（4サブフレーム相当）を得るために実行される。各サブフレームに対して、補間した I S Pベクトルは、サブフレームの再生音声の合成に使用される、LPフィルタ係数の領域 a_k に変換される。

各サブフレームに対して、以下のステップが繰り返される。

1. 適応コードブックベクトルの復号： 受信したピッチインデックス（適応コードブックインデックス）は、ピッチラグの整数と分数の成分を得るために使用される。適応コードブックベクトル $v(n)$ は、5. 7節で述べた F I Rフィルタを用いた、過去の励振信号 $u(n)$ （ピッチ遅延による）の補間によって得られる。受信した適応フィルタインデックスは、フィルタ処理された適応コードブックが、 $v_1(n)=v(n)$ か、 $v_2(n)=0.18v(n)+0.64v(n-1)+0.18v(n-2)$ かどうかを見つけたすために使用される。
2. 励振ベクトルの復号： 受信した代数的コードブックインデックスは、励振パルスの位置と振幅（極性）を抽出し、代数的コードベクトル $c(n)$ を得るために使用される。ピッチラグの整数成分がサブフレームサイズの 64 未満の時、周期性強調部分 $1/(1-0.85z^{-T})$ と、傾斜部分 $(1-\beta_1z^{-1})$ の 2つの部分（ T はピッチラグの整数成分、 $\beta_1(n)$ は前のサブフレームの有声性と関係があり、 $[0.0, 0.5]$ の範囲を取る）からなる、適応プリフィルタ $F(z)$ を通してフィルタ処理することで、補正 $c(n)$ へ変換する、ピッチシャープニング手法が適用される。
3. 適応および励振コードブック利得の復号： 受信したインデックスは、固定コードブック利得補正因子 $\hat{\gamma}$ を与える。推定の固定コードブック利得 g'_c は 5. 8節の記述にて得られる。まず、全てのサブフレーム n の予測エネルギーは、

$$\tilde{E}(n) = \sum_{i=1}^4 b_i \hat{R}(n-i) \quad (59)$$

で得られ、その時の平均励振エネルギーは、

$$E_i = 10 \log \left(\frac{1}{N} \sum_{i=0}^{N-1} c^2(i) \right) \quad (60)$$

で得られる。予測利得 g'_c は、

$$g'_c = 10^{0.05(\tilde{E}(n) + \bar{E} - E_i)} \quad (61)$$

で得られる。量子化固定コードブック利得は、

$$\hat{g}_c = \hat{\gamma} g'_c \quad (62)$$

で与えられる。

4. 再生音声の計算： 次のステップでは、 $n = 0, \dots, 63$ である。励振信号の全体は、

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \quad (63)$$

で構成される。

音声合成の前に、励振信号成分の後処理が実行される。

5. 非スパース化処理 (6.60 および 8.85 kbit/s モード) : 適応的非スパース化後処理過程は、サブフレーム毎にいくつかの非零サンプルだけを備えた代数的固定コードブックベクトルの、スパース性に起因する聴覚的な不自然さを減少するために、固定コードブックベクトル $c(n)$ へ適用される。非スパース化処理は、インパルス応答と固定コードブックベクトルの巡回畳み込みから成る。事前に格納した3つのインパルス応答が使用され、番号 $impNr=0,1,2$ はそれらの1つを選択するために設定される。値2は補正無しに相当、値1は中程度の補正に相当、値0は強力な補正に相当する。インパルス応答の選択は、適応コードブック利得と固定コードブック利得から、適宜実行される。次の手順が使用される。

```

if  $\hat{g}_p < 0.6$  then
     $impNr = 0$ ;
else if  $\hat{g}_p < 0.9$  then
     $impNr = 1$ ;
else
     $impNr = 2$ ;

```

この固定コードブック利得と前回の固定コードブック利得とを比較することで、オンセット（音声の始まり）を検出する。もし現在の値が、前回の値の3倍を越えると、オンセットと検出される。

もしオンセットでなく、 $impNr=0$ の場合、現在と直前の4つの適応コードブック利得の中央値（メディアンフィルタの出力）が計算される。この値が0.6未満であると、 $impNr=0$ である。

もしオンセットでない場合、 $impNr$ 値は、以前のサブフレームより1ステップだけ増加するように制限される。

もしオンセットが宣言される場合、 $impNr$ 値は、それが2未満であると、1増加される。

8.85 kbit/s モードの場合、 $impNr$ 値は1増加される。

6. 雑音強調器： 非線形利得平滑化技法は、雑音中の励振信号を強調するために固定コードブック利得 \hat{g}_c へ適用される。音声部分の安定性と有声性に基づき、固定コードブックの利得は、定常的な信号の場合に、励振信号エネルギーの変動を低減するために、平滑化される。これは定常的な背景雑音における性能を改善する。

有声性因子は $\lambda = 0.5(1 - r_v)$ 、このとき $r_v = (E_v - E_c) / (E_v + E_c)$ で与えられ、 E_v と E_c はそれぞれ、スケーリングされたピッチコードベクトルとスケーリングされた励振コードベクトルのエネルギーである。 r_v の値は -1 と 1 の間であるため、 λ の値は 0 と 1 の間である。因子 λ は、純粋に有声な区間は値 0 に、純粋に無声な区間は値 1 とした時の、無声性の量に関連する。

安定性因子 θ は、隣接するLPフィルタ間の距離尺度を基に計算される。ここで、因子 θ はISP距離尺度と関連し、 θ がより大きい値の時、より安定な信号となり、 $0 \leq \theta \leq 1$ で制限される。

最後に、利得平滑化因子 S_m は、

$$S_m = \lambda \theta \quad (64)$$

で与えられる。

S_m の値は、定常な背景雑音信号といった無声や安定な信号の間は1に近づく。全くの有声信号、または不安定な信号では、 S_m の値は0に近づく。

初期補正利得 g_0 は、固定コードブック利得 \hat{g}_c と、前のサブフレームから初期補正利得により与えられるしきい値 g_{-1} との比較によって、計算される。もし \hat{g}_c が g_{-1} より大きいまたは同じ場合、 g_0 は $g_0 \geq g_{-1}$ に制限され、1.5dB ずつの \hat{g}_c の減算により計算される。もし \hat{g}_c が g_{-1} より小さい時、 g_0 は $g_0 \leq g_{-1}$ に制限され、1.5dB ずつの \hat{g}_c の加算により計算される。

最終的に、利得は、次のような平滑化利得の値で更新される。

$$\hat{g}_c = S_m g_0 + (1 - S_m) \hat{g}_c \quad (65)$$

7. ピッチ強調器： ピッチ強調器の手順は、低域周波数より高域周波数の周波数応答を強調し、係数を信号の周期性と関連づける予測励振フィルタを通して、固定コードブック励振信号のフィルタ処理により、励振信号 $u(n)$ 全体を補正することである。その形式のフィルタは、

$$F_{immo}(z) = -c_{pe} z + 1 - c_{pe} z^{-1} \quad (66)$$

であり、ここで $c_{pe} = 0.125(1+r_v)$ 、このとき $r_v = (E_v - E_c)/(E_v + E_c)$ である。フィルタ処理された固定コードブックベクトルは、

$$c'(n) = c(n) - c_{pe}(c(n+1) + c(n-1)) \quad (67)$$

で与えられ、更新された励振信号は、

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c'(n) \quad (68)$$

で与えられる。

上記手順は、次のように励振信号を更新することにより、1ステップで行うことができる。

$$u(n) = u(n) - \hat{g}_c c_{pe}(c(n+1) + c(n-1)) \quad (69)$$

8. 励振信号成分の後処理 (6.60 および 8.85 kbit/s モード)： 励振信号成分の後処理の手順は、適応コードブックベクトルの寄与分の強調により、励振信号 $u(n)$ 全体に適用される。

$$\hat{u}(n) = \begin{cases} u(n) + 0.25\beta\hat{g}_p v(n), & \hat{g}_p > 0.5 \\ u(n), & \hat{g}_p \leq 0.5 \end{cases} \quad (70)$$

適応利得制御 (AGC) は非強調励振信号 $u(n)$ と強調励振信号 $\hat{u}(n)$ との間の、利得差を補正するために利用される。強調励振信号の利得スケーリングファクタ η は、

$$\eta = \begin{cases} \sqrt{\frac{\sum_{n=0}^{63} u^2(n)}{\sum_{n=0}^{63} \hat{u}^2(n)}}, & \hat{g}_p > 0.5 \\ 1.0, & \hat{g}_p \leq 0.5 \end{cases} \quad (71)$$

で計算される。利得スケーリングした強調励振信号 $\hat{u}'(n)$ は、

$$\hat{u}'(n) = \hat{u}(n)\eta \quad (72)$$

で与えられる。サイズ 64 のサブフレームの再生音声は、

$$\hat{s}(n) = \hat{u}(n) - \sum_{i=1}^{16} \hat{a}_i \hat{s}(n-i), \quad n = 0, \dots, 63 \quad (73)$$

で与えられ、 \hat{a}_i は補間されたLPフィルタ係数である。
合成音声 $\hat{s}(n)$ は、次の節で記述される適応後処理を通される。

6. 2 高域通過フィルタ、アップスケーリングと補間

高域通過フィルタは、望ましくない低周波成分に対する予防策として機能する。信号は、高域通過フィルタ $H_{hi}(z)$ とデエンファシスフィルタ $H_{de_emph}(z)$ を通して、フィルタ処理される。

最終的に信号は、低域合成信号 $\hat{s}_{16k}(n)$ を得るために、16kHz にアップサンプルされる。 $\hat{s}_{16k}(n)$ は、最初に12.8kHzの低域合成信号 $\hat{s}_{12.8k}(n)$ を5倍にアップサンプリングし、その出力を $H_{decim}(z)$ を通してフィルタ処理し、最後に1/4にダウンサンプリングすることにより生成される。

(アップスケーリングは、前処理段階でのダウンスケーリングを補正するために、高域通過フィルタからの出力に2を乗じることから成る。)

6. 3 高周波数帯域

より高い周波数帯域(6.4k~7.0kHz)に対して、その周波数成分をモデル化するための励振信号が生成される。高周波数成分は、スペクトルの上位部分を、励振領域で適切にスケールされた白色雑音で満たすことにより生成され、ダウンサンプルした信号の合成に使用されたものと同じLP合成フィルタから得たフィルタで形成することにより音声領域に変換される。

6. 3. 1 高域励振信号の生成

高域励振信号は、まず、白色雑音 $u_{HB1}(n)$ を生成することによって得られる。高域励振信号のパワーは、低域励振信号のパワー $u_2(n)$ に合わせられる。これは、

$$u_{HB2}(n) = u_{HB1}(n) \sqrt{\frac{\sum_{k=0}^{63} u_2^2(k)}{\sum_{k=0}^{63} u_{HB1}^2(k)}} \quad (74)$$

を意味する。

最後に高域励振信号は、

$$u_{HB}(n) = \hat{g}_{HB} u_{HB2}(n) \quad (75)$$

で得られ、 \hat{g}_{HB} は利得因子である。

23.85 kbit/s モードでは、 \hat{g}_{HB} は受信した利得インデックスから復号される。

6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85 と 23.05 kbit/s モードでは、有声性情報を用いて、 g_{HB} は[0.1, 1.0]の範囲で制限される値として推定される。最初に、合成信号の傾き e_{nlt} は、

$$e_{nlt} = \frac{\sum_{n=1}^{63} \hat{s}_{lp}(n) \hat{s}_{lp}(n-1)}{\sum_{n=0}^{63} \hat{s}_{lp}^2(n)} \quad (76)$$

で得られ、 $\hat{s}_{lp}(n)$ は、カットオフ周波数が 400Hz の高域通過フィルタ処理された低域音声合成信号 $\hat{s}_{12.8k}(n)$ である。その時、 \hat{g}_{HB} は、

$$g_{HB} = w_{SP} g_{SP} + (1 - w_{SP}) g_{BG} \quad (77)$$

で得られ、 $g_{SP} = 1 - e_{ilt}$ は音声信号の利得、 $g_{BG} = 1.25g_{SP}$ は背景雑音信号の利得、 w_{SP} はVADがONの時に1、VADがOFFの時0に設定する重み付け関数である。 g_{HB} は[0.1, 1.0]の間に制限される。高周波数でのエネルギーがより少ない有声区間の場合、 e_{ilt} は1に近づき、結果として利得 g_{HB} は低くなる。これにより、有声区間の場合、生成された雑音のエネルギーが低減される。

6. 3. 2 高周波数帯域のためのLPフィルタ

6. 3. 2. 1 6.60 kbit/s モード

高域LP合成フィルタ $A_{HB}(z)$ は量子化されたISFベクトル \mathbf{f} を20次のISFベクトル \mathbf{f}_e に外挿することにより得られる。始めに、ISFベクトルの差分ベクトル $f_{\Delta}(i) = f(i+1) - f(i), i=1, \dots, 14$ の自己相関の最大値 $C_{\max}(i)$ が得られる。次に新しい16kHz ISFベクトル $f'_e(i)$ が以下のように計算される。

$$f'_e(i) = \begin{cases} f(i-1), & i = 1, \dots, 15 \\ f'_e(i-1) + f'_e(i - C_{\max}(i) - 1) - f'_e(i - C_{\max}(i) - 2), & i = 16, \dots, 19 \end{cases} \quad (78)$$

新しいISFベクトルの最終要素の近似値 f_{e19} は低域周波数係数に基づき更新される。新しく外挿されたISFベクトルの差分ベクトル $f'_{e\Delta}(i)$ は以下となる。

$$f'_{e\Delta}(i) = c_{scale}(f'_e(i) - f'_e(i-1)), \quad i = 16, \dots, 19 \quad (79)$$

ここで c_{scale} は、 f_{e19} が $f_e(19)$ と等しくなるように $f'_{e\Delta}(i)$ をスケールするための係数である。安定性を保証するため $f'_{e\Delta}(i)$ は以下のように拘束される。

$$f'_{e\Delta}(i) + f'_{e\Delta}(i-1) > 500, \quad i = 17, \dots, 19 \quad (80)$$

最後に、外挿されたISFベクトル \mathbf{f}_e は以下のように得られる。

$$f_e(i) = \begin{cases} f(i), & i = 1, \dots, 15 \\ f'_{e\Delta}(i) + f_e(i-1), & i = 16, \dots, 19 \\ f(16), & i = 20 \end{cases} \quad (81)$$

\mathbf{f}_e は標準化周波数16kHzで \mathbf{q}_e を得るために余弦領域に変換される。高域LP合成フィルタ $A_{HB}(z)$ は $m=20$ として5.2.4節で述べられているように \mathbf{q}_e をLPフィルタへ変換することにより得られる。

6. 3. 2. 2 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 または 23.85 kbit/s モード

高域LP合成フィルタ $A_{HB}(z)$ は重み付き低域LP合成フィルタである。

$$A_{HB}(z) = \hat{A}(z/0.8) \quad (82)$$

ここで $\hat{A}(z)$ は補間されたLP合成フィルタである。 $\hat{A}(z)$ は12.8kHzの標準化周波数で信号を分析することにより計算されるが、今は16kHzの信号に利用される。実質的に、このことは以下により $A_{HB}(z)$ の周波数応答 $FR_{16}(f)$ が得られることを意味する。

$$FR_{16}(f) = FR_{12.8}\left(\frac{12.8}{16}f\right) \quad (83)$$

ここで $FR_{12.8}(f)$ は $A(z)$ の周波数応答である。このことは12.8kHz領域中の帯域5.1-5.6kHzが16kHz領域における6.4-7.0kHz帯域にマッピングされることを意味する。

6. 3. 3 高域合成

$u_{HB}(n)$ は $A_{HB}(z)$ を通してフィルタリングされる。この高域合成の出力 $s_{HB}(n)$ は 6-7kHz の通過帯域を有する帯域通過 FIR フィルタを通してフィルタリングされる。最後に s_{HB} は合成された出力音声信号 $\hat{s}_{output}(n)$ を作るため合成された音声 $\hat{s}_{16k}(n)$ に加算される。

7. 適応マルチレート広帯域コーデックの詳細なビット割当て

Table 1 2 a ~ 1 2 i / JT-G722.2 に、適応マルチレート広帯域音声符号器のビットの詳細な割当てをモード毎に示す。これらの表は音声符号器により作られたビットのシーケンスを示す。各コーデックパラメータの最上位ビット (MSB) は常に最初に送られる。

Table 1 2 a / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 477 bits/20 ms, 23.85 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s68	Codebook Index1 for track 1
s69 – s79	Codebook Index1 for track 2
ss80 – s90	Codebook Index1 for track 3
s91 – s101	Codebook Index1 for track 4
s102 – s112	Codebook Index2 for track 1
s113 – s123	Codebook Index2 for track 2
s124 – s134	Codebook Index2 for track 3
s135 – s145	Codebook Index2 for track 4
s146 – s152	codebook gains
s153 – s156	High-band energy
subframe 2	
s157 – s162	adaptive codebook index (relative)
s163 – s262	same description as s57 – s156
subframe 3	
s263 – s371	same description as s48 – s156
subframe 4	
s372 – s477	same description as s157 – s262

Table 1 2 b / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech
(ITU-T G.722.2) frame of 461 bits/20 ms, 23.05 kbit/s mode

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s68	Codebook Index1 for track 1
s69 – s79	Codebook Index1 for track 2
ss80 – s90	Codebook Index1 for track 3
s91 – s101	Codebook Index1 for track 4
s102 – s112	Codebook Index2 for track 1
s113 – s123	Codebook Index2 for track 2
s124 – s134	Codebook Index2 for track 3
s135 – s145	Codebook Index2 for track 4
s146 – s152	codebook gains
subframe 2	
s153 – s158	adaptive codebook index (relative)
s159 – s254	same description as s57 – s152
subframe 3	
s255 – s359	same description as s48 – s152
subframe 4	
s360 – s461	same description as s153 – s254

Table 1 2 c / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech
(ITU-T G.722.2) frame of 397 bits/20 ms, 19.85 kbit/s mode

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s67	Codebook Index1 for track 1
s68 – s77	Codebook Index1 for track 2
s78 – s79	Pulse Selector for track 3
s80 – s81	Pulse Selector for track 4
s82 – s91	Codebook index2 for track 1
s92 – s101	Codebook index2 for track 2
s102 – s115	Codebook index for track 3
s116 – s129	Codebook index for track 4
s130 – s136	VQ gain
subframe 2	
s137 – s142	adaptive codebook index (relative)
s143 – s222	same description as s57 – s136
subframe 3	
s223 – s311	same description as s48 – s136
subframe 4	
s312 – s397	same description as s137 – s222

Table 1 2 d / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 365 bits/20 ms, 18.25 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s59	Pulse Selector for track 1
s60 – s61	Pulse Selector for track 2
s62 – s63	Pulse Selector for track 3
s64 – s65	Pulse Selector for track 4
s66 – s79	Codebook index for track 1
s80 – s93	Codebook index for track 2
s94 – s107	Codebook index for track 3
s108 – s121	Codebook index for track 4
s122 – s128	VQ gain
subframe 2	
s129 – s134	adaptive codebook index (relative)
s135 – s206	same description as s57 – s128
subframe 3	
s207 – s287	same description as s48 – s128
subframe 4	
s288 – s365	same description as s129 – s206

Table 1 2 e / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 317 bits/20 ms, 15.85 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s70	Codebook index for track 1
s71 – s83	Codebook index for track 2
s84 – s96	Codebook index for track 3
s97 – s109	Codebook index for track 4
s110 – s116	VQ gain
subframe 2	
s117 – s122	adaptive codebook index (relative)
s123 – s182	same description as s57 – s116
subframe 3	
s183 – s251	same description as s48 – s116
subframe 4	
s252 – s317	same description as s117 – s182

Table 1 2 f / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 285 bits/20 ms, 14.25 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s70	Codebook index for track 1
s71 – s83	Codebook index for track 2
s84 – s92	Codebook index for track 3
s93 – s101	Codebook index for track 4
s102 – s108	VQ gain
subframe 2	
s109 – s114	adaptive codebook index (relative)
s115 – s166	same description as s57 – s108
subframe 3	
s167 – s227	same description as s48 – s108
subframe 4	
s228 – s285	same description as s109 – s166

Table 1 2 g / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 253 bits/20 ms, 12.65 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s56	adaptive codebook index
s57	LTP-filtering-flag
s58 – s66	Codebook index for track 1
s67 – s75	Codebook index for track 2
s76 – s84	Codebook index for track 3
s85 – s93	Codebook index for track 4
s94 – s100	VQ gain
subframe 2	
s101 – s106	adaptive codebook index (relative)
s107 – s150	same description as s57 – s100
subframe 3	
s151 – s203	same description as s48 – s100
subframe 4	
s204 – s253	same description as s101 – s150

Table 1 2 h / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 177 bits/20 ms, 8.85 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s23	index of 3rd ISP subvector
s24 – s30	index of 4th ISP subvector
s31 – s37	index of 5th ISP subvector
s38 – s42	index of 6th ISP subvector
s43 – s47	index of 7th ISP subvector
subframe 1	
s48 – s55	adaptive codebook index
s56 – s60	Codebook index for track 1
s61 – s65	Codebook index for track 2
s66 – s70	Codebook index for track 3
s71 – s75	Codebook index for track 4
s76 – s81	VQ gain
subframe 2	
s82 – s86	adaptive codebook index (relative)
s87 – s112	same description as s56 – s81
subframe 3	
s113 – s146	same description as s48 – s81
subframe 4	
s147 – s177	same description as s82 – s112

Table 1 2 i / JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation within the speech frame of 132 bits/20 ms, 6.60 kbit/s mode (ITU-T G.722.2)

Bits (MSB-LSB)	Description
s1	VAD-flag
s2 – s9	index of 1st ISP subvector
s10 – s17	index of 2nd ISP subvector
s18 – s24	index of 3rd ISP subvector
s25 – s31	index of 4th ISP subvector
s32 – s37	index of 5th ISP subvector
subframe 1	
s38 – s45	adaptive codebook index
s46 – s57	Codebook Index
s58 – s63	VQ gain
subframe 2	
s64 – s68	adaptive codebook index (relative)
s69 – s86	same description as s46 – s63
subframe 3	
s87 – s109	same description as s64 – s86
subframe 4	
s110 – s132	same description as s64 – s86

8. ホーミングシーケンス

8. 1 機能説明

適応マルチレート広帯域音声コーデックは、正確な動作の一般的なテストに加えて簡単な認定試験を許容するため、ビットイグザクトな演算で記述される。

コーデックの内部状態変数が実験の初期値として予め定められた定数である場合のみ、予め定義された入力系列に対するコーデックの応答は予測可能である。それ故、ビットイグザクトなテストを実行する前に、コーデックをいわゆるホーム状態にする必要がある。これは通常リセット動作（コーデックの内部状態変数が初期値に設定される手続き）によりもたらされる。音声符号器と音声復号器のコーデックモードは、リセット時に外部手段によりテストコーデックモードに設定される。

遠隔地からコーデックのリセットを許可するために、例えばインバンドシグナリングによりコーデックのホーミングを可能にするような、符号器および復号器に対する特別なホーミングフレームが定義されている。

コーデックホーミングの手順は次のような方法で定義される。符号器もしくは復号器どちらにおいてもホーミング関数は、ホーミングフレームの処理後に呼び出される。したがって、最初のホーミングフレームに対応する出力は、そのフレームを受信した時に使用されているコーデックモードとコーデック状態に依存するため、通常未知である。以降のホーミングフレームに対する符号器の応答は、使用されているコーデックモードに対応する復号器のホーミングフレームの定義による。以降のホーミングフレームに対する復号器の応答は、符号器のホーミングフレームによる。適切なフレーミングを考慮し、ループバック構成が実現されている場合、この手順は、どちら側からでも符号器、復号器のいずれのホーミングを許可する。

8. 2 定義

符号器ホーミングフレーム：

符号器のホーミングフレームは、最下位ビットが「1」、その他すべてのビットが「0」に設定された、各々13ビット長の同一の320サンプルで構成される。左詰めで16ビットワードに書き込む場合、サンプルは「0x0008」の値を持つ。少なくとも2つの復号器ホーミングフレームが連続して復号器に入力した場合、復号器は2つ目以降の復号器ホーミングフレームに対する応答として、このフレームを生成しなければならない。符号器ホーミングフレームは全てのコーデックモードで同一である。

復号器ホーミングフレーム：

9つのAMR-WBコーデックモードに対応する、9つの異なる復号器ホーミングフレームが存在する。これらのコーデックモードのうち使用している1つのモードと一致する復号器ホーミングフレームは、少なくとも2つの符号器ホーミングフレームが連続して符号器に入力した場合、2つ目以降の符号器ホーミングフレームに対する音声符号器の応答となる。TTC標準JT-G722.2付属資料Cにて、各々の復号器ホーミングフレームパラメータ値が与えられる。

8. 3 符号器ホーミング

適応マルチレート広帯域音声符号器が、内部音声フレームセグメントに正確に位置調整された符号器ホーミングフレーム入力を受信する場合、以下に示すイベントが発生する。

第1ステップ：

音声符号器は、VADおよびSCRを含む通常動作を行い、使用されているコーデックモードに従った通常未知の音声パラメータフレームを生成する。しかし音声符号器が最初のフレームにおいてホーム状態であった場合、結果として生成される音声パラメータフレームは、使用されたコーデックモードに対応する復号器ホーミングフレームと同一のものとなる（これは復号器ホーミングフレームの生成方法である）。

第2ステップ：

処理の終了に成功した後、音声符号器はVADおよびSCRを含む全サブモジュールについてホーミング関数を起動し、全ての状態変数をそれらのホーム状態にする。次の入力フレームを受信する場合、音声符号

器はホーム状態から始動する。

注意：

N個の符号器ホームミングフレームのシーケンス入力は、少なくともN-1の復号器ホームミングフレームを音声符号器から出力させる。

8.4 復号器ホームミング

音声復号器が、使用しているコーデックモードに対応する復号器ホームミングフレーム入力を受信した場合、以下に示すイベントが発生する。

第1ステップ：

音声復号器は、通常動作を行い、通常未知の音声出力フレームを生成する。しかし、音声復号器がそのフレームの開始点においてホーム状態であった場合、結果として生じる音声フレームは符号器ホームミングフレームと置き換えられる。これは場合によらず、この定義により強制される。

第2ステップ：

その動作の終了が成功した後、音声復号器は擬似雑音発生器を含む全てのサブモジュールのホームミング関数を起動し、全ての変数をホーム状態に設定する。音声復号器は、次の入力フレーム受信時には、自身のホーム状態から動作開始する事になる。

注意1：

N個の復号器ホームミングフレームのシーケンス入力は、少なくともN-1の符号器ホームミングフレームを音声復号器から出力させる。

注意2：

定義により、各復号器のテストシーケンスの第1フレームと、復号器ホームミングフレームの間では、LPCパラメータと最初のサブフレームの内部で、少なくとも1ビットの位置で差異がなければならない。それ故、復号器がホーム状態である場合、次の復号器ホームミングフレームを検出するには、これらのパラメータのみを調査するだけで十分である。この定義は遅延が最適化された実現をサポートするために作られたものである。

9. 音声検出 (VAD)

VADアルゴリズムの関数は、各々20ms フレームが伝送されるべき音声・楽音・または情報トーンといった信号を含んでいるかどうかを示すためのものである。VADアルゴリズムの出力は、そのような信号の存在を示すブール型のフラグ (VAD_flag) である。このフラグは、AMR-WB 音声符号器と TTC 標準 JT-G 722. 2 付属資料Cにおいて使用される。

9. 1 VAD記号

本節では、以下の記号を適用する。

9. 1. 1 VAD変数

bckr_est[n]	周波数帯域 n における背景雑音推定
burst_count	VADハングオーバー付加部によって使用される、音声パースト長のカウンタ
hang_count	VADハングオーバー付加部によって使用される、ハングオーバーカウンタ
level[n]	周波数帯域 n における信号レベル
new_speech	音声符号器のポインタであり、音声フレームの最新受信サンプルを含むバッファを指し示す
noise_level	推定雑音レベル
pow_sum	入力電力
s(i)	入力フレームのサンプル
snr_sum	入力フレームと雑音推定値間の距離
speech_level	推定音声レベル
stat_count	定常性カウンタ
stat_rat	入力フレームの定常状態推定値
tone_flag	トーンを含むことを示すフラグ
vad_thr	VADしきい値
VAD_flag	ブール型VADフラグ
vadreg	VAD中間判定値

9. 1. 2 VAD定数

ALPHA_UP1	雑音推定値の更新のための定数
ALPHA_DOWN1	雑音推定値の更新のための定数
ALPHA_UP2	雑音推定値の更新のための定数
ALPHA_DOWN2	雑音推定値の更新のための定数
ALPHA3	雑音推定値の更新のための定数
ALPHA4	平均信号レベルの更新のための定数
ALPHA5	平均信号レベルの更新のための定数
BURST_HIGH	VADハングオーバー付加制御のための定数
BURST_P1	VADハングオーバー付加制御のための定数
BURST_SLOPE	VADハングオーバー付加制御のための定数
COEFF3	フィルタバンク係数
COEFF5_1	フィルタバンク係数
COEFF5_2	フィルタバンク係数
HANG_HIGH	VADハングオーバー付加制御のための定数

HANG_LOW	VADハングオーバー付加制御のための定数
HANG_P1	VADハングオーバー付加制御のための定数
HANG_SLOPE	VADハングオーバー付加制御のための定数
FRAME_LEN	音声フレーム長。256 サンプル (20ms)
MIN_SPEECH_LEVEL1	音声推定のための定数
MIN_SPEECH_LEVEL2	音声推定のための定数
MIN_SPEECH_SNR	VAD適応化のための定数
NO_P1	VAD適応化のための定数
NO_SLOPE	VAD適応化のための定数
NOISE_MAX	雑音推定値の最大値
NOISE_MIN	雑音推定値の最小値
POW_TONE_THR	トーン検出しきい値
SP_ACTIVITY_COUNT	音声推定のための定数
SP_ALPHA_DOWN	音声推定のための定数
SP_ALPHA_UP	音声推定のための定数
SP_CH_MAX	VAD適応化のための定数
SP_CH_MIN	VAD適応化のための定数
SP_EST_COUNT	音声推定のための定数
SP_P1	VAD適応化のための定数
SP_SLOPE	VAD適応化のための定数
STAT_COUNT	定常状態検出しきい値
STAT_THR	定常状態検出しきい値
STAT_THR_LEVEL	定常状態検出しきい値
THR_HIGH	VAD適応化のための定数
TONE_THR	トーン検出しきい値
VAD_POW_LOW	VADハングオーバー付加制御のための定数

9. 1. 3 関数

+	加算
-	減算
*	乗算
/	除算
$ x $	x の絶対値
AND	ブール型AND演算
OR	ブール型OR演算
$\sum_{n=a}^b x(n)$	$= x(a) + x(a+1) + \dots + x(b-1) + x(b)$
$\text{MIN}(x, y)$	$= \begin{cases} x, y \leq y \\ y, y < x \end{cases}$
$\text{MAX}(x, y)$	$= \begin{cases} x, y \geq y \\ y, y > x \end{cases}$

9. 2 機能説明

VADアルゴリズムのブロック図を Figure 4 / JT-G722.2 に示す。VADアルゴリズムは、ブール型のVAD

フラグ (VAD_flag) を計算するために音声符号器のパラメータを使用する。VADへの入力フレームは 12.8kHzで標本化され、256 サンプルで構成される。入力フレームのサンプル (s(i)) は、サブバンドに分離され、それぞれの帯域における信号レベル (level[n]) が計算される。音声符号器の開ループ分析によって算出される正規化開ループピッチ利得は、トーン検出関数に入力する。トーン検出関数は、シグナリングトーン、有声音、もしくは他の周期性の強い信号の存在を示すフラグ (tone_flag) を計算する。背景雑音レベル (bckr_est[n]) は、VAD判定、信号の定常状態、そしてトーンフラグに基づいて、それぞれの帯域で推定される。VAD中間判定値は、適応的しきい値と入力SNR (level[n]/bckr_est[n]) を比較することにより計算される。しきい値は、雑音と長区間音声推定に基づいて適応化させる。最後に、VAD中間判定値にハングオーバを加算することにより、VADフラグが計算される。

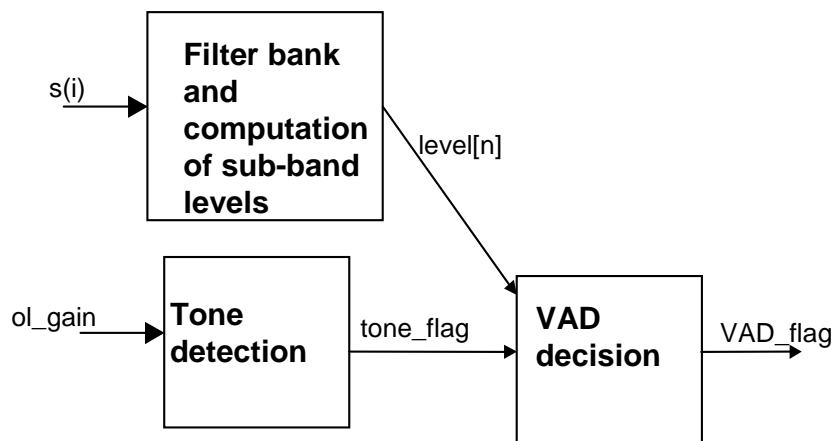


Figure 4 / JT-G722.2 Simplified block diagram of the VAD algorithm
(ITU-T G.722.2)

9. 2. 1 フィルタバンクとサブバンドレベル計算

入力信号は、フィルタバンクにより 12 帯域に分割される (Figure 5 / JT-G722.2)。フィルタバンクの遮断周波数を Table 1.3 / JT-G722.2 に示す。

Table 1 3 / JT-G722.2 Cut-off frequencies for the filter bank
(ITU-T G.722.2)

Band number	Frequencies
1	0 – 200 Hz
2	200 – 400 Hz
3	400 – 600 Hz
4	600 – 800 Hz
5	800 – 1200 Hz
6	1200 – 1600 Hz
7	1600 – 2000 Hz
8	2000 – 2400 Hz
9	2400 - 3200 Hz
10	3200 – 4000 Hz
11	4000 – 4800 Hz
12	4800 – 6400 Hz

フィルタバンクへの入力は、TTC標準JT-G722.2符号器のnew_speechポインタで指し示される音声フレームである。フィルタバンクへの入力値は、1ビットスケールダウンされる。これは、フィルタバンク計算におけるオーバーフローを回避するためである。

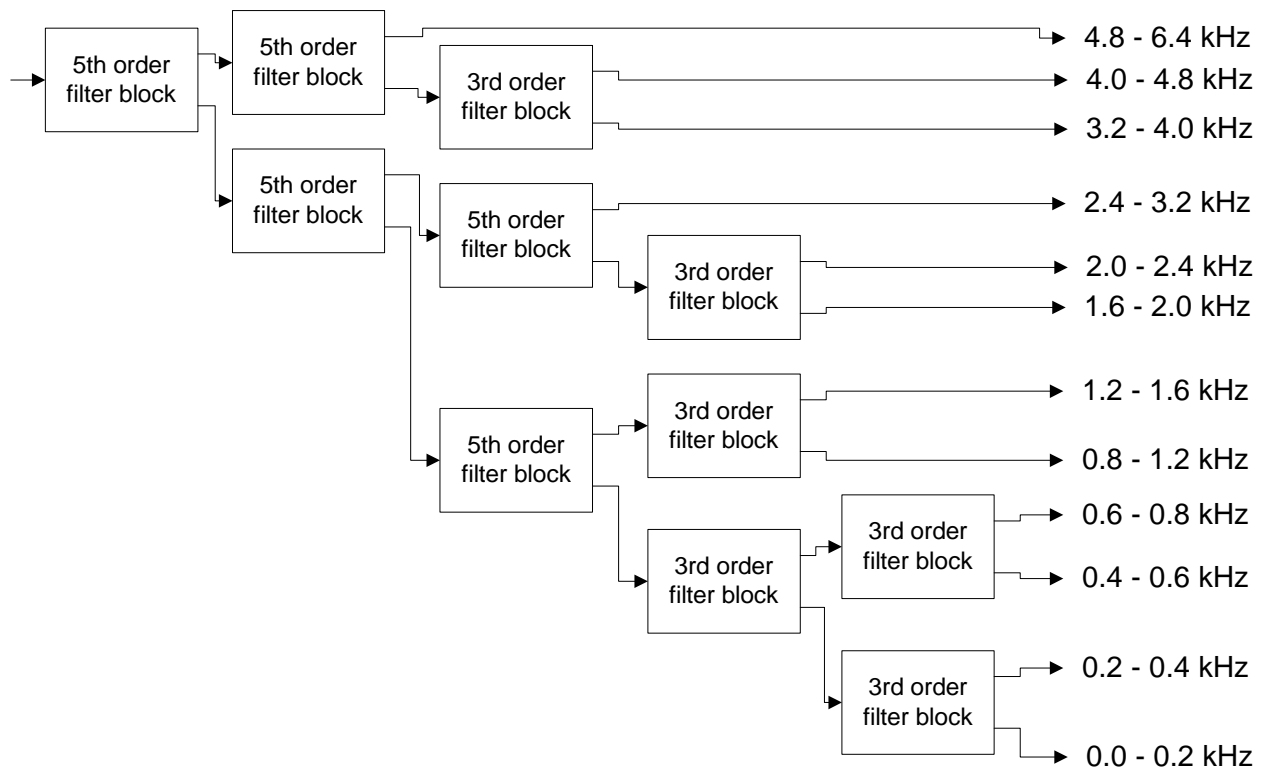


Figure 5 / JT-G722.2 Filter bank
(ITU-T G.722.2)

フィルタバンクは、5次及び3次のフィルタブロックで構成される。それぞれのフィルタブロックは、入力を高域と低域成分に分割し、標準化周波数を1/2に間引く。5次フィルタブロック計算を以下に示す。

$$x_{lp}(i) = 0.5 * (A_1(x(2*i)) + A_2(x(2*i+1))) \quad (84a)$$

$$x_{hp}(i) = 0.5 * (A_1(x(2*i)) - A_2(x(2*i+1))) \quad (84b)$$

ここで、

$x(i)$ フィルタブロックへの入力信号

$x_{lp}(i)$ 低域成分

$x_{hp}(i)$ 高域成分

3次フィルタブロック計算を以下に示す。

$$x_{lp}(i) = 0.5 * (x(2*i+1) + A_3(x(2*i))) \quad (85a)$$

$$x_{hp}(i) = 0.5 * (x(2*i+1) - A_3(x(2*i))) \quad (85b)$$

フィルタ $A_1()$ 、 $A_2()$ 、及び $A_3()$ は、1次の直接型全域通過フィルタで、その伝達関数は下式で与えられる。

$$A(z) = \frac{C + z^{-1}}{1 + C * z^{-1}} \quad (86)$$

ここで、 C はフィルタ係数である。

全域通過フィルタ $A_1()$ 、 $A_2()$ 、 $A_3()$ の係数は、それぞれ $COEFF5_1$ 、 $COEFF5_2$ 、及び $COEFF3$ である。

信号レベルは、フィルタバンクの各帯域出力から、以下のように計算される。

$$level(n) = \sum_{i=START_n}^{END_n} |x_n(i)| \quad (87)$$

ここで、

n 周波数帯域のインデックス

$x_n(i)$ 周波数帯域 n フィルタバンクのサンプル i 出力

$$START_n = \begin{cases} -6, & 1 \leq n \leq 4 \\ -12, & 5 \leq n \leq 8 \\ -24, & 9 \leq n \leq 11 \\ -48, & n = 12 \end{cases}$$

$$END_n = \begin{cases} 7, & 1 \leq n \leq 4 \\ 15, & 5 \leq n \leq 8 \\ 31, & 9 \leq n \leq 11 \\ 63, & n = 12 \end{cases}$$

$x_n(i)$ の負インデックス値は前フレームから参照する。

9. 2. 2 トーン検出

トーン検出機能の目的は、情報トーン、母音、及びその他の周期性信号を検出するためである。トーン検出には、正規化開ループピッチ利得 (ol_gain) を用いる。これは、音声符号器より受信する。ピッチ利得が定数 TONE_THR より大きいとき、トーンと判定しトーンフラグを設定する。

```
if (ol_gain > TONE_THR)
    tone_flag = 1
```

開ループピッチ探索、及び対応するトーンフラグはフレーム毎に2回計算される。但し、6.60 kbit/s モード時は1回である。

9. 2. 3 VAD判定

VAD判定アルゴリズムのブロック図を Figure 6 / JT-G722.2 に示す。

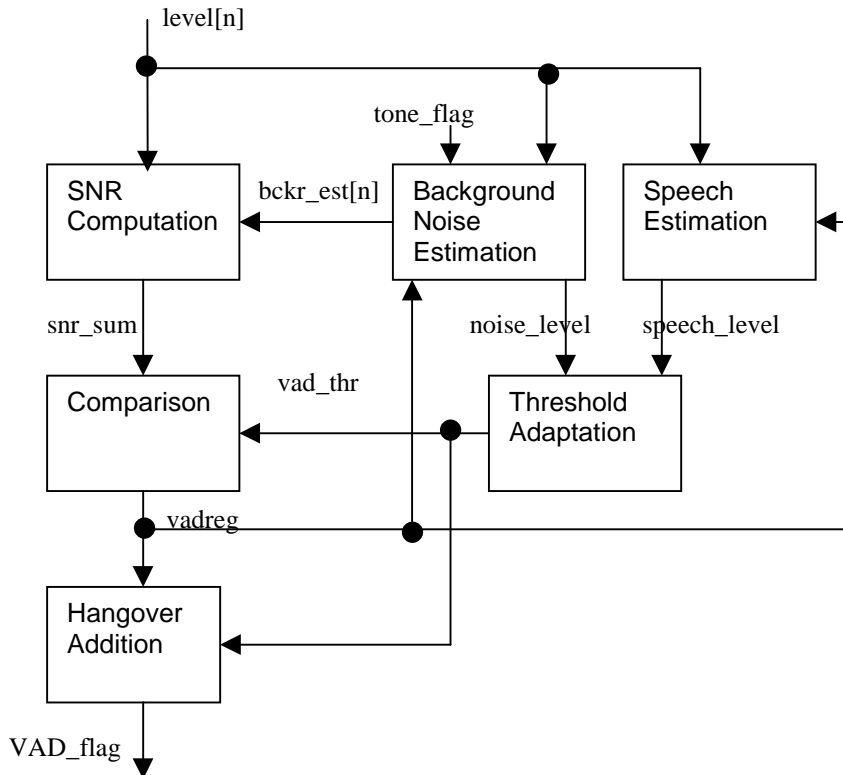


Figure 6 / JT-G722.2 Simplified block diagram of the VAD decision algorithm (ITU-T G.722.2)

入力フレームパワー計算を以下に示す。

$$frame_pow = \sum_{i=0}^{FRAME_LEN} s(i) * s(i) \quad (88)$$

ここで、入力フレームのサンプル s(i)は音声符号器の new_speech ポインタで指し示されるフレームから得られる。変数 pow_sum は、現フレームと前フレームのパワーの和である。この pow_sum が定数 POW_TONE_THR より小さいとき、トーンフラグを零に設定する。

入力フレームレベルと推定背景雑音レベルとの差分は、以下のように計算する。

$$snr_sum = \sum_{n=1}^{12} \text{MAX}(1.0, \frac{level[n]}{bckr_est[n]})^2 \quad (89)$$

ここで、

level[n] 帯域 n における信号レベル
bckr_est[n] 帯域 n における推定背景雑音レベル

VAD判定は、変数 snr_sum と閾値比較により計算される。この閾値 (vad_thr) は、想定される音声レベルと背景雑音レベルに依存する感度に応じて設定される。

平均背景雑音レベルは、最低帯域を除く各帯域の推定ノイズの総和計算である。

$$noise_level = \sum_{n=2}^{12} bckr_est[n] \quad (90)$$

SNRが閾値 (MIN_SPEECH_SNR) より小さいとき、音声レベルは以下のように増加する。

```
If (speech_level/noise_level < MIN_SPEECH_SNR)
    Speech_level = MIN_SPEECH_SNR * noise_level
```

推定雑音の対数値は以下のように計算する。

$$ilog2_noise_level = \log_2(noise_level) \quad (91)$$

推定音声レベルの対数値を計算する前に、低SNR時における音声レベルを補正するため、音声レベルから MIN_SPEECH_SNR*noise_level を減算する。

$$ilog2_speech_level = \log_2(speech_level - MIN_SPEECH_SNR * noise_level) \quad (92)$$

VAD判定閾値は以下のように計算される。

$$Vad_thr = NO_SLOPE * (ilog2_noise_level - NO_P1) + THR_HIGH + \text{MIN}(SP_CH_MAX, \text{MAX}(SP_CH_MIN, SP_CH_MIN + SP_SLOPE * (ilog2_speech_level - SP_P1))) \quad (93)$$

ここで、NO_SLOPE、SP_SLOPE、NO_P1、SP_P1、THR_HIGH、SP_CH_MAX 及び SP_CH_MIN は定数である。

変数 vadreg はVAD中間判定値であり、以下のように計算される。

```
if (snr_sum > vad_thr)
    vadreg = 1
else
    vadreg = 0
```

9. 2. 3. 1 ハングオーバー付加

最終VADフラグを設定する前に、ハングオーバーを付加する。このハングオーバー付加は、主観上重要だが検出が困難であるような音声バーストのパワーの低い語尾の検出を補助する。

判定値「1」が burst_len フレーム分連続して検出された以降は、判定値が「0」になっても、hang_len フレーム分経過するまでの期間、VAD_flag は「1」に設定される。変数 hang_len と burst_len は、vad_thr を用いて以下の様に計算される。

$$\text{hang_len} = \text{MAX}(\text{HANG_LOW}, (\text{HANG_SLOPE} * (\text{vad_thr} - \text{HANG_P1}) + \text{HANG_HIGH})) \quad (94)$$

$$\text{burst_len} = \text{BURST_SLOPE} * (\text{vad_thr} - \text{BURST_P1}) + \text{BURST_HIGH} \quad (95)$$

入力フレームパワーが閾値 (VAD_POW_LOW) より小さい時、VAD_flag を「0」に設定し、かつ、ハングオーバを付加しない。VAD_flag 計算を以下に示す。

```
Vad_flag = 0;
if (pow_sum < VAD_POW_LOW)
    burst_count = 0
    hang_count = 0
else
    if (vadreg = 1)
        burst_count = burst_count + 1
        if (burst_count >= burst_len)
            hang_count = hang_len
            VAD_flag = 1
    else
        burst_count = 0
        if (hang_count > 0)
            hang_count = hang_count - 1
            VAD_flag = 1
```

9. 2. 3. 2 背景雑音推定

背景雑音推定 ($bckr_est[n]$) は、前フレームの振幅レベルを用い更新される。したがって、この更新は1フレーム分遅延する。これは、非検出語頭による雑音推定エラーを避けるためである。現フレームに対する更新速度は、VAD中間判定値 (vadreg) と定常性カウンタ (stat_count) を用い、以下のように選択される。

```
if (vadreg for the last 4 frames has been zero)
    alpha_up = ALPHA_UP1
    alpha_down = ALPHA_DOWN1
else if (stat_count = 0)
    alpha_up = ALPHA_UP2
    alpha_down = ALPHA_DOWN2
else
    alpha_up = 0
    alpha_down = ALPHA3
```

変数 stat_count は、定常性を示し、この目的は本節で後述する。変数 alpha_up と alpha_down はそれぞれ増加方向と減少方向の更新速度を決定する。各帯域 n 毎の更新速度は以下のように選択される。

```
if ( $bckr\_est_m[n] < level_{m-1}[n]$ )
    alpha[n] = alpha_up
else
    alpha[n] = alpha_down
```

最終的に、推定雑音は以下のように更新される。

$$bckr_est_{m+1}[n] = (1.0 - alpha[n]) * bckr_est_m[n] + alpha[n] * level_{m-1}[n] \quad (96)$$

ここで、

- n 周波数帯インデックス
- m フレームインデックス

推定背景雑音 (bckr_est[n]) レベルは定数 NOISE_MIN、NOISE_MAX に制限される。

背景雑音が増加した時、vadreg は「1」と設定され、背景雑音は通常通りの増加更新はされない。この状態から回復するため、VAD 中間判定値 (vadreg) が長期間「1」で、かつ、スペクトラムが定常である時は、推定背景雑音を更新できるようにする。定常性推定値 (stat_rat) は、以下のように計算される。

$$stat_rat = \sum_{n=1}^{12} \frac{MAX(STAT_THR_LEVEL, MAX(ave_level_m[n], level_m[n]))}{MAX(STAT_THR_LEVEL, MIN(ave_level_m[n], level_m[n]))} \quad (97)$$

ここで、

- STAT_THR_LEVEL 定数
- n 周波数帯インデックス
- m フレームインデックス
- ave_level 入力信号の平均レベル

定常性推定値 (stat_rat) が、閾値より大きいとき、定常性カウンタ (stat_count) を初期値 STAT_COUNT に設定する。信号が定常的でなく、音声と判定されたとき (VAD 判定「1」)、stat_count は零になるまでフレーム毎に 1 カウント減算される。

```

if (5 last tone flags have been one)
    stat_count = STAT_COUNT
else
    if (8 last internal VAD decisions have been zero) OR (stat_rat > STAT_THR)
        stat_count = STAT_COUNT
    else
        if (vadreg) AND (stat_count ≠ 0)
            stat_count = stat_count - 1
    
```

平均信号レベル (ave_level[n]) は以下のように計算される。

$$ave_level_{m+1}[n] = (1.0 - alpha) * ave_level_m[n] + alpha * level_m[n] \quad (98)$$

上式における更新速度 (alpha) は、以下のように選択される。

```

if (stat_count = STAT_COUNT)
    alpha = 1.0
else if (vadreg = 1)
    
```

```

alpha=ALPHA5
else
alpha = ALPHA4

```

9. 2. 3. 3 音声レベル推定

はじめに、最低帯域を除く各帯域の入力レベル和として全帯域入力レベルを以下のように計算する。

$$in_level = \sum_{n=2}^{12} level[n] \quad (99)$$

入力レベルが MIN_SPEECH_LEVEL1 より十分に大きく、かつVAD中間判定値 (vadreg) が「1」に設定されるか、入力レベルが現在の音声レベル推定値より大きいかいずれかの場合、そのフレームは音声を含むとみなされる。その場合、最大レベル (sp_max) は、SP_EST_COUNT フレームから探索される。SP_EST_COUNT フレーム中に、SP_ACTIVITY_COUNT フレーム分の音声フレームがあるならば、音声レベルは最大信号レベル (sp_max) で更新される。音声レベル推定のための擬似コードを以下に示す。

```

If (SP_ACTIVITY_COUNT > SP_EST_COUNT - sp_est_cnt + sp_max_cnt)
  sp_est_cnt = 0
  sp_max_cnt = 0
  sp_max = 0
sp_est_cnt = sp_est_cnt + 1
if (in_level > MIN_SPEECH_LEVEL1) AND ((vadreg = 1) OR (in_level > speech_level))
  sp_max_cnt = sp_max_cnt + 1
  sp_max = MAX(sp_max, in_level)
if (sp_max_cnt > SP_ACTIVITY_COUNT)
  if (sp_max > MIN_SPEECH_LEVEL2)
    if (sp_max > speech_level)
      speech_level = speech_level + SP_ALPHA_UP * (sp_max - speech_level)
    else
      speech_level = speech_level + SP_ALPHA_DOWN * (sp_max - speech_level)
  sp_max_cnt = 0
  sp_max = 0
  sp_est_cnt = 0

```

10. 3GPP音声電話サービスにおける必須なAMR-WBコーデックモード

本章は3GPPの無線ネットワークとの相互接続が要求される場合に従われるべき章である。

3GPPシステムの回線交換型音声電話サービスにおけるAMR-WBの実装を容易にするために、必須なAMR-WBコーデックモードの数を5に削減した。これにより、端末及びネットワークにおける伝送路符号化に対する複雑さが軽減される。使用されるモードは、23.85、15.85、12.65、8.85 および 6.60 kbit/s である。音声電話サービスチャンネルに対する受聴試験結果により、高品質な音声電話サービスにはこれらの5つのモードで十分であると判断される。3GPPにおける他のサービスやアプリケーションに対しては、9つの全てのモードが使用される。すなわち、9つの全ての情報源コーデックは3GPP AMR-WBコーデックとして存在するが、音声電話サービスに使用されるモード数だけが5に制限される。

相互接続を改善するために、アクティブコーデックセット (ACS) 内の許容されるAMR-WBコーデックモード構成、すなわち、どのモードをモード適合の際に同時に使用するか、ということが更に制限される。次の3種類の構成が許容される。すなわち、構成A (6.60、8.85、12.65)、構成B (6.60、8.85、12.65、15.85) および構成C (6.60、8.85、12.65、23.85) である。これにより、3GPPの様々な音声サービス伝送チャンネルにおけるタンデムフリー動作/トランスコーダフリー動作 (TFO/TrFO) の相互接続性が改善される。なぜならば、これらの3つの構成に相互互換性があるからである。それらは、同じ3つの低レートなモードを含む。これにより、TFO/TrFOを即時に確立することができる。ここで、移動端末間の呼接続での2重のトランスコーディングが避けられる。この許容されるモード形態に対する制限は、シグナリング (例えば、呼設定時やハンドオーバー時における) も単純化し、更にはテストを非常に単純にできる。

Table 1 4/JT-G722.2 は、3GPP音声電話サービスにおけるAMR-WBコーデックモードのサポートに関する要求基準をまとめたものである。

Table 1 4/JT-G722.2 Minimum requirements for AMR-WB modes in 3GPP speech telephony service (ITU-T G.722.2)

Speech telephony service channel	Terminal	Network
TCH/F	Configuration A (6.60, 8.85 and 12.65)	Configuration A (6.60, 8.85 and 12.65)
O-TCH/H	Configuration A (6.60, 8.85 and 12.65)	Configuration A (6.60, 8.85 and 12.65)
O-TCH/F	Configurations A (6.60, 8.85, 12.65), B (6.60, 8.85, 12.65, 15.85) and C (6.60, 8.85, 12.65, 23.85)	Configuration A (6.60, 8.85 and 12.65)*
UTRAN	Configurations A (6.60, 8.85, 12.65), B (6.60, 8.85, 12.65, 15.85) and C (6.60, 8.85, 12.65, 23.85)	Configuration A (6.60, 8.85 and 12.65) *

*) Support for configurations B and C is optional

1 1. 参考文献 (informative)

- [2] M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High quality speech at very low bit rates," in *Proc. ICASSP'85*, pp. 937-940, 1985.
- [3] L.R. Rabiner and R.W. Schaefer. *Digital processing of speech signals*. Prentice-Hall Int., 1978.
- [4] F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, Supplement no. 1, S35, 1975.
- [5] Y. Bistriz and S. Pellerin, "Immittance Spectral Pairs (ISP) for speech encoding," in *Proc. ICASSP'93*, pp. II-9 - II-12.
- [6] K.K Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *IEEE Trans. Speech and Audio Processing*, vol. 1, no 1, pp. 3-14, 1993.
- [7] P. Kabal and R.P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials", *IEEE Trans. on ASSP*, vol. 34, no. 6, pp. 1419-1426, Dec. 1986.
- [8] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "GSM enhanced full rate speech codec", in *Proc. ICASSP'97*, pp. 771-774.
- [9] T. Honkanen, J. Vainio, K. Järvinen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "Enhanced full rate speech codec for IS-136 digital cellular system", in *Proc. ICASSP'97*, pp. 731-734.
- [10] R. Hagen, E. Ekudden, B. Johansson, and W.B. Kleijn, "Removal of sparse-excitation artifacts in CELP", in *Proc. ICASSP'98*, pp. I-145-I-148.
- [11] 3GPP TS 26.171 : "AMR Wideband Speech Codec; General description".
- [12] 3GPP TS 26.190 : "AMR Wideband Speech Codec; Transcoding functions".
- [13] 3GPP TS 26.173 : "AMR Wideband Speech Codec; ANSI-C code".
- [14] 3GPP TS 26.174 : "AMR Wideband Speech Codec; Test sequences".
- [15] 3GPP TS 26.193 : "AMR Wideband Speech Codec; Source Controlled Rate operation".
- [16] 3GPP TS 26.194 : "AMR Wideband Speech Codec; Voice Activity Detection (VAD)".
- [17] 3GPP TS 26.192 : "AMR Wideband Speech Codec; Comfort Noise Aspects".
- [18] 3GPP TS 26.191 : "AMR Wideband Speech Codec; Error Concealment of Lost Frames".
- [19] 3GPP TS 26.201 : "AMR Wideband Speech Codec; Frame Structure".
- [20] 3GPP TR 26.901 : "AMR Wideband Speech Codec; Performance characterisation".

付属資料A
(標準 J T - G 7 2 2 . 2 に対する)
擬似背景雑音特性

A. 1 本付属資料の規定範囲

本付属資料は、ソースコントロールレート (S C R) 動作における、AMR広帯域 (A M R - W B) 音声符号化方式のための背景音響雑音評価、雑音パラメータ符号化/復号、および擬似背景雑音生成の動作について記述するものである。

本付属資料の実装は、3 G P P システムとの互換性には必要であるが、その利用はモバイルアプリケーションには限定されない。

本付属資料のCコードによる実装には、標準 J T - G 7 2 2 . 2 付属資料CにおけるCコードの一部が利用可能である。本付属資料と、T T C 標準 J T - G 7 2 2 . 2 付属資料Cに含まれる固定小数点演算記述との間に不一致があった場合には、T T C 標準 J T - G 7 2 2 . 2 付属資料Cの記述が優先される。

A. 2 定義、記号、略語

A. 2. 1 定義

本付属資料では、以下の定義を適用する。

フレーム： 適応マルチレート広帯域音声符号器の処理時間単位に相当する 20ms の時間間隔。通信フレームの略語としても用いられる。

S I D フレーム： 特別な擬似背景雑音フレーム。音響背景雑音の情報を搬送したり、背景雑音発生の開始を復号器に知らせたりする。

音声フレーム： S I D フレームに属さない通信フレーム。

V A D フラグ： 音声検出フラグ。

T X _ T Y P E： 送信フレームの分類 (T T C 標準 J T - G 7 2 2 . 2 付属資料Bに定義される)

R X _ T Y P E： 受信フレームの分類 (T T C 標準 J T - G 7 2 2 . 2 付属資料Bに定義される)

本付属資料に用いられる他の用語は、T T C 標準 J T - G 7 2 2 . 2 および T T C 標準 J T - G 7 2 2 . 2 付属資料Bに現れる。S C R の全体的な動作は、T T C 標準 J T - G 7 2 2 . 2 付属資料Bに記述される。

A. 2. 2 記号

本付属資料では、以下の記号を適用する。ボールド体の記号はベクトル変数に用いられている。

$\mathbf{f}^T = [f_1 f_2 \dots f_{16}]$	量子化されていない I S F ベクトル
$\hat{\mathbf{f}}^T = [\hat{f}_1 \hat{f}_2 \dots \hat{f}_{16}]$	量子化された I S F ベクトル
$\mathbf{f}^{(m)}$	フレーム m の量子化されていない I S F ベクトル
$\hat{\mathbf{f}}^{(m)}$	フレーム m の量子化された I S F ベクトル
\mathbf{f}^{mean}	平均化された I S F パラメータベクトル
en_{\log}	対数フレームエネルギー
en_{\log}^{mean}	平均化された対数フレームエネルギー
e	I S F パラメータ予測残差

ϵ 量子化された I S F パラメータ予測残差

$$\sum_{n=a}^b x(n) = x(a) + x(a+1) + \dots + x(b-1) + x(b)$$

A. 2. 3 略語

本付属資料では、以下の略語を適用する。

AMR	適応マルチレート (Adaptive Multi-Rate)
AMR-WB	適応マルチレート広帯域 (Adaptive Multi-Rate Wideband)
CN	擬似背景雑音 (Comfort Noise)
I S F	イミタンススペクトル周波数 (Impedance Spectral Frequency)
I S P	イミタンススペクトル対 (Impedance Spectral Pair)
LP	線形予測 (Linear Prediction)
R S S	無線サブシステム (Radio Subsystem)
R X	受信 (Receive)
S C R	ソースコントロールレート動作 (Source Controlled Rate operation) (別名 不連続ソース伝送)
S I D	無音挿入記述子 (Silence Insertion Descriptor)
T X	送信 (Transmit)
U E	ユーザ装置 (User Equipment)
V A D	音声検出器 (Voice Activity Detector)

A. 3 概要

S C Rを使用する時の基本的な問題として、背景音響雑音は音声とともに伝送されるが、伝送が断たれた時には背景音響雑音も失われ、結果として背景雑音の不連続を引き起こすことが挙げられる。S C Rの切り替えが急速に行われるため、この影響は受聴者にとって非常に不快となりうることが知られている。特に自動車内の大きな背景雑音などは顕著であり、ひどい場合には音声がほとんど聞き取れなくなることもある。

本付属資料はこのような問題について、送信側の背景雑音に類似した合成雑音を、受信側で生成することによって解決する方法を定める。擬似背景雑音のパラメータは送信側で推定され、音声が存在しないときに一定のレートで受信側に伝送される。これにより、擬似背景雑音は送信側の雑音の変化に適応することが可能となる。

A. 4 送信側の機能

擬似背景雑音算出アルゴリズムには、T T C標準 J T - G 7 2 2 . 2 に定義されている AMR - W B 音声符号器の以下のパラメータを用いる。

- イミタンススペクトル対 (I S P) 表現を用いた、量子化されていない線形予測パラメータ。ここで、量子化されていないイミタンススペクトル周波数ベクトルは、 $f^T = [f_1 f_2 \dots f_{16}]$ で与えられる。

このアルゴリズムは、擬似背景雑音生成を補助する下記のパラメータを計算する。

- 重み付き平均化 I S F パラメータベクトル \mathbf{f}^{mean} (最新 8 フレームの I S F パラメータの重み付き平均)
- 平均化された対数フレームエネルギー en_{log}^{mean} (最新 8 フレームの対数エネルギーの平均)

これらのパラメータは、背景雑音のレベル (en_{log}^{mean}) およびスペクトル (\mathbf{f}^{mean}) の情報を与える。

算出された擬似背景雑音パラメータ (\mathbf{f}^{mean} および en_{log}^{mean}) は、無音挿入記述子 (S I D) フレームと呼ばれ

る特別なフレームに符号化され、受信側に伝送される。

無音挿入記述子フレームの品質を向上させるために、ハングオーバー論理が用いられる。符号器が、アクティブモードから非アクティブモードに切り替わるのを7フレームの間遅らせて、その間に復号器が、復号音声信号の対数フレームエネルギーと量子化されたISFから無音挿入記述子フレームを計算できるように、7フレームのハングオーバーをVADフラグに付加する。したがって、有音の後の最初のSIDフレームでは、擬似背景雑音記述は伝送されない。背景雑音が短期的な状態変化を持ち、符号器が非常に短い区間でアクティブモードと非アクティブモードが切り替わるようであれば、ハングオーバーは用いられない。その代わりに、以前に用いられた擬似背景雑音フレームが擬似背景雑音生成に用いられる。

最初のSIDフレームはまた、受信側での擬似背景雑音生成を開始するために与えられる。最初のSIDフレームはつねに音声バーストの最後、つまり伝送が終了する前に送信される。

ネットワークパス上のSIDフレームまたは音声フレームのスケジューリングは、TTC標準JT-G722.2付属資料Bに記述される。

A. 4. 1 ISF算出

SIDフレームに符号化される擬似背景雑音パラメータは、VAD=0で示された8連続フレームについて次のように計算される。

CN平均化区間でISFパラメータを平均化する前に、平均化されるISFパラメータのセットについて送信側での背景雑音の特徴を示さないパラメータを除去するために、中央値置換が行われる。初めに、それぞれのISFパラメータベクトル $\mathbf{f}(i)$ から、他のISFパラメータベクトル $\mathbf{f}(j)$, ($i=0, \dots, 7, j=0, \dots, 7, i \neq j$)までのCN平均化区間内のスペクトル距離が、次式によって近似される。

$$\Delta R_{ij} = \sum_{k=1}^{16} (f_i(k) - f_j(k))^2 \quad (1)$$

ここで、 $f_i(k)$ は、フレーム*i*におけるISFパラメータベクトル $\mathbf{f}(i)$ の*k*番目のISFパラメータである。

CN平均化区間内で、ISFパラメータベクトル $\mathbf{f}(i)$ から、他のすべての異なるフレーム $j=0, \dots, 7, j \neq i$ のISFパラメータベクトル $\mathbf{f}(j)$ までのスペクトル距離 ΔS_i を求めるために、スペクトル距離 ΔR_{ij} の総和が次式によって計算される。

$$\Delta S_i = \sum_{j=0, j \neq i}^7 \Delta R_{ij}, \quad i=0, \dots, 7, i \neq j \quad (2)$$

CN平均化区間内のすべてのISFパラメータベクトルの中で、最小のスペクトル距離 ΔS_i をもつISFパラメータベクトル $\mathbf{f}(i)$ は、その平均化区間の中央ISFパラメータベクトル \mathbf{f}_{med} とみなされ、そのスペクトル距離は ΔS_{med} で表される。中央ISFパラメータベクトルは、その平均化区間内のすべてのISFパラメータベクトルの中で、背景雑音の短期スペクトルが最も良く表現されたものとみなされる。ISFパラメータベクトル $\mathbf{f}(j)$ が、次式で示されるCN平均化区間にあるならば、多くとも2つのISFパラメータベクトル(TH_{med} を最も数多く超えるISFパラメータベクトル)が、平均化ISFパラメータベクトル \mathbf{f}^{mean} を計算する前に、中央ISFパラメータベクトルに置換される。

$$\frac{\Delta S_j}{\Delta S_{med}} > TH_{med} \quad (3)$$

ここで、 $TH_{med} = 2.25$ は、中央値置換の閾値である。

中央値置換の結果として得られた I S F パラメータベクトルのセットは $\mathbf{f}'(n-i)$ と表される。ここで n は現フレームのインデックスであり、 $i (i=0, \dots, 7)$ は平均化区間のインデックスである。

中央値置換がハングオーバー区間の最後（最初の CN 更新）で行われるとき、以前の 7 フレーム（ハングオーバー区間 $i=1, \dots, 7$ ）のすべての I S F パラメータベクトル $\mathbf{f}(n-i)$ は、量子化された値を持ち、一方、最新のフレーム n における I S F パラメータベクトルは、量子化されていない値を持つ。次の CN 更新では、ハングオーバー区間と重なっているフレーム内の CN 平均化区間の I S F パラメータベクトルは、量子化された値を持ち、一方、CN 平均化区間のそれ以降のフレームのパラメータベクトルは、量子化されていない値を持つ。最新の 8 フレームの区間がハングオーバー区間と重なっていないとき、I S F パラメータの中央値置換は、量子化されていないパラメータ値だけを用いて行われる。

フレーム n における平均化 I S F パラメータベクトル $\mathbf{f}^{mean}(n)$ は次式によって計算される。

$$\mathbf{f}^{mean}(n) = \frac{1}{8} \sum_{i=0}^7 \mathbf{f}'(n-i) \quad (4)$$

ここで $\mathbf{f}'(n-i)$ は、中央値置換後の最新 8 フレーム ($i = 0, \dots, 7$) のうち 1 つの I S F パラメータベクトルである。 i は平均化区間インデックス、 n はフレームインデックスである。

フレーム n における平均化 I S F パラメータベクトル $\mathbf{f}^{mean}(n)$ は、擬似背景雑音 I S F 量子化テーブルを用いて量子化される。量子化されるにあたり、平均除去された I S F ベクトルは、次式によって得られる。

$$\mathbf{r}(n) = \mathbf{f}^{mean}(n) - \bar{\mathbf{f}} \quad (5)$$

ここで $\mathbf{f}^{mean}(n)$ は、フレーム n における平均化 I S F パラメータベクトル、 $\bar{\mathbf{f}}$ は定数の平均 I S F ベクトル、 $\mathbf{r}(n)$ はフレーム n における平均除去された I S F ベクトル、 n はフレームインデックスである。

A. 4. 2 フレームエネルギー計算

フレームエネルギーは、VAD=0 で示された各フレームに対して、次式によって計算される。

$$en_{\log}(i) = \frac{1}{2} \log_2 \left(\frac{1}{N} \sum_{n=0}^{N-1} s^2(n) \right) \quad (6)$$

ここで $s(n)$ は、高域通過フィルタを通った現フレーム i の入力音声信号である。エネルギーはまた、示された音声モード能力にしたがって、擬似背景雑音から音声への高品位な遷移をもたらすように調整される。

平均化対数エネルギーは、次式によって計算される。

$$en_{\log}^{mean}(i) = \frac{1}{8} \sum_{n=0}^7 en_{\log}(i-n) \quad (7)$$

平均化対数エネルギーは、6 ビットの算術量子化器を用いて量子化される。6 ビットのエネルギーインデックスが S I D フレームで伝送される (Table A-1 / JT-G722.2 のビット割り当て参照)。

A. 4. 3 背景雑音の変化と安定性の解析

符号器は初めに、背景雑音が定常かどうかを決定する。非定常な背景雑音にはディザリングが使用される。ディザリングを使用するか否かの情報は、2 値情報 (CN_{dith} フラグ) を使用することで復号器に伝送される。

2 値の CN_{dith} フラグは、CN 平均化区間 (l_{dx}) 内のスペクトルパラメータベクトル $\mathbf{f}(i)$ からスペクトルパラメー

タ $\mathbf{f}(j)$ までのスペクトル距離 ΔS_i ($j=0, \dots, l_{dtx}-1, j \neq i$) を用いて得られる。スペクトル距離の計算は、A. 4. 1 節に記述されている。続いてスペクトル距離の総和 $D_S = \sum_{i=0}^7 \Delta S_i$ が計算される。 D_S が小さければ、 CN_{dih} フラグは0にセットされる。そうでないなら、 CN_{dih} フラグは1にセットされる。さらに、フレーム間のエネルギーの変化について調べてみる。 $en_{log}(i)$ の平均値 en_{log} からの絶対偏差の総和が計算される。総和が大きければ、以前のフラグが0にセットされていたとしても、 CN_{dih} フラグは1にセットされる。

A. 4. 4 S I Dフレーム生成時の音声符号化アルゴリズムの修正

TX_TYPE が SPEECH に等しくないとき、音声符号化アルゴリズムは以下の方法で修正される。

- 音声符号器のフィルタ $H(z)$ と $W(z)$ のフィルタ係数を得るために使用された、平均化されていないLPパラメータは量子化されない。
- 開ループピッチラグ探索は実行されるが、閉ループピッチラグ探索は非アクティブとなる。適応コードブックメモリは零にセットされる。
- 固定コードブック探索は行われない。
- 重み付けフィルタ $W(z)$ は零にセットされる。つまり $W(z)$ は更新されない。
- 通常のLPパラメータ量子化アルゴリズムは非アクティブとなる。平均化ISFパラメータベクトル \mathbf{f}^{mean} は、新しいSIDフレームが送信されるごとに計算される。このパラメータベクトルは、A. 4. 1 節に定義されたSIDフレームの中に符号化される。
- 通常のゲイン量子化アルゴリズムは非アクティブとなる。
- 通常のLPパラメータ量子化アルゴリズムの予測メモリは、音声再びアクティブになる際に量子化器が既知の初期状態から開始できるように、TX_TYPE が SPEECH でない場合に初期化される。

23.85 kbit/s モードにおいて、TX_TYPE が SPEECH に等しく、かつVADがOFFのとき、音声符号化アルゴリズムは以下の方法に修正される。

- 高域ゲイン g_{HB} の生成は、アクティブでない音声の区間、高域ゲインをスムーズに遷移させるために、ゲイン予測値に適応させることによって変更される。 g_{HB} は次式のようになる。

$$g_{HB} = \frac{hang_{DTX}}{7} g_{HB} + (1 - \frac{hang_{DTX}}{7}) g_{est} \quad (8)$$

ここで、 $hang_{DTX}$ はDTXカウンタである。

A. 4. 5 S I Dフレーム符号化

SIDフレーム中の擬似背景雑音ビットの符号化は、TTC標準JT-G722. 2付属資料Eに記述されている。またここで最初のSIDフレームの内容について記述されている。擬似背景雑音符号化におけるビット割り当ておよびビット列については、TableA-1/JT-G722.2に示されている。

A. 5 受信側の機能

擬似背景雑音が受信側で生成される状況については、TTC標準JT-G722. 2付属資料Bに定義されている。一般に擬似背景雑音の生成は、有効なSIDフレームが受信されるたびに開始または更新される。

A. 5. 1 LPおよびエネルギーパラメータの平均化および復号

音声フレームが復号器に受信される時、最新の7音声フレームのLPおよびエネルギーパラメータがメモリに保持される。復号器は、符号器によって最新のSIDフレームが更新され、RSSに通されてからのフレーム経過数をカウントする。このカウントに基づいて、復号器は、(TTC標準JT-G722.2付属資料Bに定義されている)音声バーストの最後におけるハングオーバー区間であるか否かを決定する。また補間因子がSID更新レートに適応する。

SIDフレームが受信されるとすぐに、擬似背景雑音が復号器で生成される。最初のSIDフレームパラメータは受信されず、ハングオーバー区間に保存されたパラメータから計算される。ハングオーバー区間が検出されなくなると、前回のSID更新からのパラメータが使用される。

最初のSIDフレームに対する擬似背景雑音パラメータを得るための平均化の手順は以下の通りである。

- 音声フレームが受信されると、ISFベクトルが復号されメモリに保存される。さらに復号された信号の対数フレームエネルギーが同様にメモリに保存される。
- 量子化されたISFベクトルの平均値と、復号されたフレームの対数フレームエネルギーの平均値が計算され、擬似背景雑音の生成に用いられる。

最初のSIDフレームに対するISFベクトルの平均値は、次式で与えられる。

$$\hat{\mathbf{f}}^{mean}(i) = \frac{1}{8} \sum_{n=0}^7 \hat{\mathbf{f}}(i-n) \quad (9)$$

ここで、 $\hat{\mathbf{f}}(i-n)$, $n > 0$ はハングオーバー区間の1フレームの量子化されたISFベクトルであり、 $\hat{\mathbf{f}}(i-0) = \hat{\mathbf{f}}(i-1)$ である。最初のSIDフレームに対する対数フレームエネルギーの平均値は、次式で与えられる。

$$\hat{en}_{\log}^{mean}(i) = \frac{1}{8} \sum_{n=0}^7 \hat{en}_{\log}(i-n) \quad (10)$$

ここで、 $\hat{en}_{\log}(i-n)$, $n > 0$ は復号されたフレームのために計算された、ハングオーバー区間の1フレームの対数ベクトルであり、 $\hat{en}_{\log}(i-0) = \hat{en}_{\log}(i-1)$ である。

通常のSIDフレームについては、ISFベクトルと対数フレームエネルギーはテーブル引きにより計算される。ISFベクトルは、復号された参照ベクトルと定数の平均ISFベクトルの和として与えられる。

擬似背景雑音生成の間、擬似背景雑音のスペクトルとエネルギーは、古いSIDフレームと新しいSIDフレームとの補間によって決定される。

ディザリングが使用される時、ISFベクトル \mathbf{f} は次式で修正される。

$$\mathbf{f}(i) = \mathbf{f}(i) + rand(-L(i), L(i)), \quad i = 1, \dots, 16 \quad (11)$$

ここで、 $L(i) = 100 + 0.8i \text{ Hz}$ であり、 $rand(-L(i), L(i))$ は、 $-L(i)$ と $L(i)$ の間の値を生成する乱数関数である。175Hzの最小ギャップが、 \mathbf{f} の各要素間に保証される。

エネルギーパラメータに対するディザリングの挿入は、スペクトルのディザリングと同様に、次式のように計算することができる。

$$en_{\log}^{mean} = en_{\log}^{mean} + rand(-L, L) \quad (12)$$

ここで $L=75$ および en_{\log}^{mean} は、擬似背景雑音励振のエネルギーをスケーリングするために使用されるエネルギー値である。

A. 5. 2 擬似背景雑音の生成および更新

擬似背景雑音の生成の手順は、TTC標準JT-G722.2に定義されている適応マルチレート広帯域(A MR-WB)音声復号アルゴリズムを用いる。

擬似背景雑音を生成するにあたり、様々な符号化パラメータが以下のようにセットされる。

各サブフレームにおいて、パルスの位置と励振の極性は、一様擬似乱数を用いて局所的に生成される。励振パルスは、擬似背景雑音が生成されるときに-2048から+2047までの値をとる。固定コードブック擬似背景雑音励振生成アルゴリズムは、下記のように動作する。

```
for (i = 0; i < 64; i++) u[i] = shr(random(),4);
```

ここで、 $u[0,\dots,63]$ は励振バッファであり、 $random()$ は-32768から+32767までの整数の一様乱数値を生成する。

励振ゲインは、対数フレームエネルギーパラメータの値を線形領域に変換することで計算される。

各サブフレームにおける適応コードブックゲインの値は0にセットされる。適応コードブックのメモリも同様に零にセットされる。

各サブフレームにおけるピッチ遅延値は、64にセットされる。

使用されるLPフィルタパラメータは、SIDフレームにおいて受信されたものである。

通常のLPパラメータアルゴリズムの予測メモリは、音声再びアクティブになる際に量子化器が与えられた初期状態から開始できるように、RX_TYPEがSPEECHでない場合に初期化される。以上のパラメータを用いて、音声復号器は、TTC標準JT-G722.2に記述されている標準的な動作を実行し、擬似背景雑音を合成する。高域生成は、CN生成中には、8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 kbit/sの各モードの音声期間中にあるように、推定された高域ゲインを用いて実行される。

擬似背景雑音パラメータ(エネルギーおよびLPフィルタパラメータ)の更新は、TTC標準JT-G722.2に記述されているように、有効なSIDフレームが受信されるたびに行われる。

擬似背景雑音を更新するとき、上述したパラメータは、スムーズに遷移させるために、SID更新区間で補間されるべきである。

A. 6 計算の詳細およびビット割り当て

擬似背景雑音の符号化および生成法のビットイグザクトな計算記述が、ANSI-Cソースコード形式でTTC標準JT-G722.2付属資料Cに記述される。

擬似背景雑音符号化におけるビット割り当ておよびビット列の詳細は、Table A-1/JT-G722.2に示される。

Table A-1/JT-G722.2 Source encoder output parameters in order of occurrence and bit allocation for (ITU-T G.722.2) comfort noise encoding

Bits (MSB-LSB)	Description
s1 - s6	index of 1st ISF subvector
s7- s12	index of 2st ISF subvector
s13 - s18	index of 3nd ISF subvector
s19 - s23	index of 4th ISF subvector
s24 - s28	index of 5th ISF subvector
s29 - s34	index of logarithmic frame energy
s35	dithering flag

付属資料B
(標準 J T - G 7 2 2 . 2 に対する)
ソースコントロールレート動作

B. 1 本付属資料の規定範囲

本付属資料は、適応マルチレート広帯域コーデックのソースコントロールレート（SCR）動作（不連続伝送、DTXとも呼ばれる）について記述する。本付属資料の実装は、3GPPシステムとの相互接続性のために必要であるが、その適用はモバイルアプリケーションに制限されない。

Figure B-1/JT-G722.2 のブロック図に従って構成を説明する。音声復号器の出力が同じである限り、各機能を各システム要素に分配したこの構成での実装は必須ではない。

B. 2 定義、記号、略語

B. 2. 1 定義

本付属資料では、以下の定義を適用する。

フレーム： 20ms の時間間隔。適応マルチレート広帯域コーデックの時間上の処理単位に相当する。また、通信フレームの略語としても用いられる。

通信フレーム： 音声伝送チャンネル上で送信される 132~477 情報ビットのブロック。

S I D フレーム： 音響背景雑音についての情報を伝送するフレーム。

音声フレーム： SPEECH_GOOD, SPEECH_BAD、または SPEECH_LOST フレームと分類された通信フレーム。

V A D フラグ： 論理フラグ。VAD アルゴリズムにより生成される。1 が有音、0 が無音の音声フレームを示す。

RX_TYPE： 受信フレームの分類。

TX_TYPE： 送信フレームの分類。

ハングオーバー区間： VAD フラグが 0 で、TX_TYPE が SPEECH_GOOD の場合、音声パーストの終了時に付加されるフレーム区間。この区間は、擬似背景雑音パラメータを生成する特別な窓を符号器に与える。

B. 2. 2 記号

本付属資料では、以下の定義を適用する。

N_{elapsed} 直前に更新された S I D フレーム後に経過したフレームの数。

B. 2. 3 略語

本付属資料では、以下の定義を適用する。

AN	アクセス網 (Access Network)
RAN	無線アクセス網 (Radio Access Network)
RX	受信 (Receive)
SCR	ソースコントロールレート動作 (Source Controlled Rate operation)
S I D	無音挿入記述子 (Silence Insertion Descriptor)
TX	送信 (Transmit)
UE	ユーザ端末 (User Equipment)
VAD	音声検出器 (Voice Activity Detector)

B. 3 概要

ソースコントロールレート動作（SCR）は、AMR広帯域音声コーデックのための仕組みであり、音声の無音性を考慮することによって平均的に低レートで入力信号を符号化する。SCRの仕組みは、次の目的で使用される。

- － ユーザ端末の電力を節約する。
- － 網における、全体的な相互干渉および負荷を削減する。

送信路（上り回線）のSCRは、網からSCRが制御された場合にUEで動作する。UEは、いつでも、送信路のSCRが制御されているかどうかに関係なく、受信路（下り回線）のSCRを操作することができる。

B. 3. 1 構成概要

本付属資料で記述するデフォルトのSCRの仕組みは、次の機能を必要とする。

- － 送信（TX）側の音声検出器（VAD）
- － 受信（RX）側に特性パラメータを伝送するための送信（TX）側での音響背景雑音の推定
- － 送信が停止している区間でのコンフォートノイズと呼ばれる擬似雑音の受信（RX）側での生成

音声検出器（VAD）は、TTC標準JT-G722.2本体で定義されており、擬似背景雑音機能はTTC標準JT-G722.2付属資料Aで定義されている。どちらの機能もTTC標準JT-G722.2で定義されている音声コーデックとその内部変数の一部に基づいている。

これらの機能に加えて、エラーにより受信側に到達したパラメータに破損が検出された場合は、受聴者側での不快な影響を回避するために、代替したデータから音声や擬似背景雑音を生成しなければならない。この機能は、TTC標準JT-G722.2付録1で定義されている。

1リンクのSCR動作の概観をFigure B-1/JT-G722.2に示す。

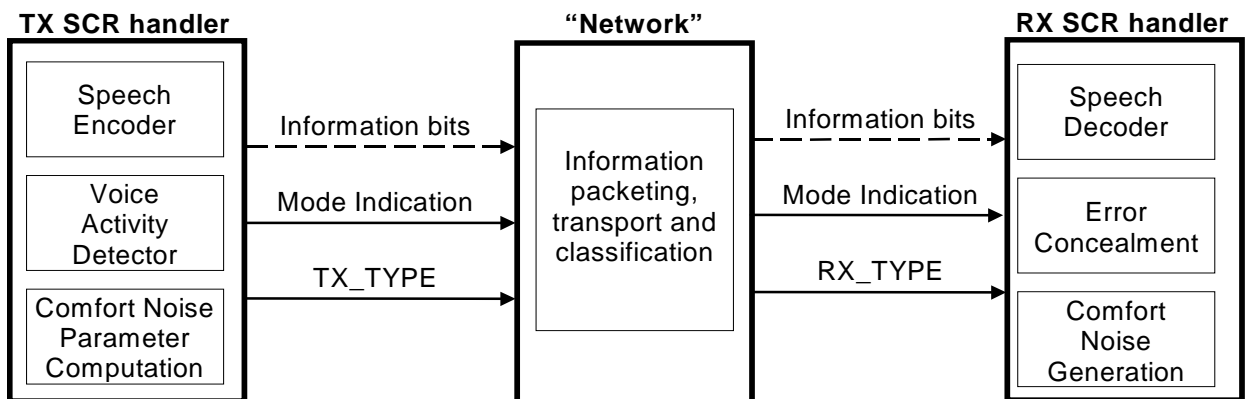
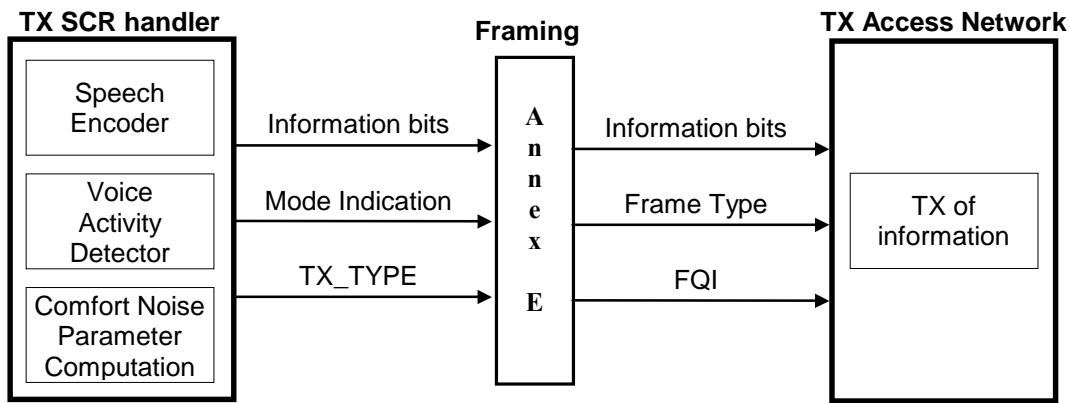


Figure B-1/JT-G722.2 Block diagram of one link SCR operation
(ITU-T G.722.2)

B. 4 AMR-WB SCR動作

B. 4. 1 送信（TX）側

Figure B-2/JT-G722.2に送信側SCR機能のブロック図を示す。



FigureB- 2 /JT-G722.2 Block diagram of SCR functions at the TX side
(ITU-T G.722.2)

B. 4. 1. 1 動作概要

送信SCRハンドラは、フレーミング処理ユニットに対して、個別に TX_TYPE を指定した通信フレームを送出する。各フレームは、情報ビット、コーデックモード指定、および TX_TYPE を含むビットフィールドから構成される。TX_TYPE は、フレームの内容を指定するために使用される。以下の表は、使用される TX_TYPE の一覧を示し、情報ビットおよびモード指示ビットフィールドにおいて必要な内容を説明する。

Table B- 1 /JT-G722.2 SCR TX_TYPE identifiers for AMR-WB and AMR-WB
(ITU-T G.722.2)

TX_TYPE	Information Bits	Mode Indication
SPEECH_GOOD	Speech frame, size 132..477 bits, depending on codec mode	Current codec mode
SPEECH_BAD	Corrupt speech frame (bad CRC), size 132..477 bits, depending on codec mode	Current codec mode
SPEECH_LOST	No useful information. (Note: If implementation does not support the SPEECH_LOST, SPEECH_BAD shall be used instead)	No useful information
SID_FIRST	Marker for the end of talkspurt, no further information, all 35 comfort noise bits set to "0"	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
SID_UPDATE	35 comfort noise bits	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
SID_BAD	Corrupt SID update frame (bad CRC)	The codec mode that would have been used if TX_TYPE had been "SPEECH_GOOD"
NO_DATA	No useful information, nothing to be transmitted	No useful information

TX_TYPE が SPEECH_LOST の場合、情報ビットとコーデックモードフィールドは全く有効なデータを含まない（ただし、ANに送信すべき）ことを示す。この TX_TYPE は、このフレームは送信されたが、前の段階において欠落が発生していることを示す。この TX_TYPE は、タンデムフリー動作とトランスコーダフリー動作においてのみ発生する。SPEECH_LOST を SPEECH_BAD に置換することが可能であるが、全く有効な情報を含まないフレームから受信パラメータの一部を誤り補償に使用するため、受信端末における誤り補償の品質を劣化させる。

TX_TYPE が NO_DATA の場合、情報ビットとコーデックモードフィールドが有効な情報を含まない（また ANに送信すべきでない）ことを示す。この TX_TYPE は、送信器とANとの網における伝送を節約するオプションを提供することを目的とするものである。SPEECH_BAD, SPEECH_LOST、または SID_BAD という TX_TYPE は、タンデムフリー動作やトランスコーダフリー動作において発生する。

送信SCRハンドラが、TX_TYPE フィールドを使用してANのフレーム伝送のスケジューリングを制御する。

B. 4. 1. 2 送信SCRハンドラ機能

送信SCR動作が許可されていない場合、送信SCRハンドラは、継続して音声フレーム、すなわち、TX_TYPE が SPEECH_GOOD であるフレームを生成する。

送信SCR動作が許可されている場合、VADフラグが、以下の段落に記述する送信SCRハンドラ動作を制御する。

B. 4. 1. 2. 1 AMR-WB SCRタイミング手順

送信SCRハンドラ機能の厳密な検証を行うため、システムをリセットする以前の全てのフレームについては、無限長の音声フレームが存在したものと仮定して扱う。従って、また受信SCR側の擬似背景雑音パラメータの正確な推定値を保証するために、リセット後またはSCR動作が可能になった後、最初の7フレームは、VADフラグが0の場合でも、常に TX_TYPE に SPEECH_GOOD が指定される（ハングオーバー区間、Figure B-3/JT-G722.2 を参照）。

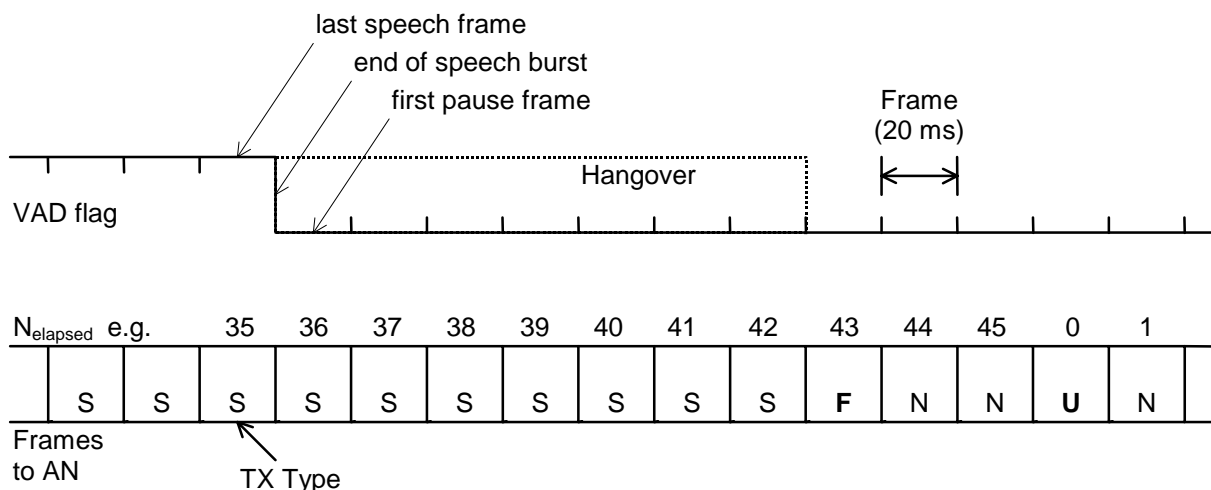
音声検出器（VAD）は、入力信号に音声が含まれるかどうかを判断するため常に動作する。フレーム単位に2値フラグ（VADフラグは1または0）を出力する（TTC標準JT-G722.2 本体を参照）。

VADフラグは、以下に説明される送信SCRハンドラ動作に従って、送信側のSCR動作全体を間接的に制御する。

VADフラグが1の場合は、常に TX_TYPE に SPEECH_GOOD が指定された音声符号器の出力フレームが、モード情報と一緒にANに直接送出される。

音声バーストの終了時（VADフラグが1から0に遷移）において、新たに更新されたSID分析を利用可能とするために、8連続のフレームが必要である（TTC標準JT-G722.2 付属資料A参照）。従って通常、音声バースト終了後、最初の7個の音声符号器出力フレームは、TX_TYPE が SPEECH_GOOD を指定した状態でANに直接送出される（ハングオーバー区間）。

TX_TYPE に SID_FIRST を指定した音声バーストの終了後8番目のフレームをANに送出することによって、音声バーストの終了が示される（Figure B-3/JT-G722.2 参照）。SID_FIRST フレームはデータを含まない。



TX Types: “S” = SPEECH; “F” = SID_FIRST; “U” = “SID_UPDATE; “N” = NO DATA
 N_{elapsed} : No. of elapsed frames since last SID_UPDATE

Figure B – 3 / JT-G722.2 Normal hangover procedure for AMR-WB ($N_{\text{elapsed}} > 23$)
 (ITU-T G.722.2)

しかしながら、音声バースト終了時において、最後に SID_UPDATE フレームが生成されてから 24 フレーム未満しか経過していない場合は、新たに更新された S I D 分析が可能になるまで (VAD フラグが 0 を指定した 8 連続フレーム)、SID_UPDATE フレームが生成される度に最後に更新された SID_UPDATE フレームが AN に送信される。これにより、S I D フレーム生成の為の “ハングオーバー” 待機を回避することで、短い背景雑音スパイクを音声とみなして動作する場合の網の負荷を軽減する。

一旦 SID_FIRST フレームが AN に送出されると、送信 S C R ハンドラは、VAD フラグが 0 である限り、一定間隔で最新の SID_UPDATE (擬似背景雑音) フレームを生成し、AN に送出する。SID_UPDATE フレームは 8 フレーム毎に生成される。最初の SID_UPDATE は、SID_FIRST フレームから 3 フレーム後に送信される。

符号器のすべての機能が擬似背景雑音パラメータ推定に必要ではないし、また擬似背景雑音パラメータはある一定間隔に生成されるだけであるので、音声符号器は、TX_TYPE が SPEECH_GOOD の場合は完全な音声モードで動作し、そうでない場合は簡易モードで動作する。

B. 4. 1. 3 AN の送信部

AN の送信部は、以下の全ての機能を持っている。話者が会話を中断した場合、SID_FIRST フレームを伝送した後送信を切断する。会話が途切れている間、一定間隔で SID_UPDATE の 1 フレーム送信を再開して、受信側の擬似背景雑音生成を更新する。AN の送信部の動作は TX_TYPE により送信 S C R ハンドラが制御する。

SPEECH_GOOD, SID_FIRST, SID_UPDATE と指定された全フレームが AN の送信部に伝送される。

B. 4. 2 受信 (RX) 側

受信側の S C R 動作のブロック図を、以下の Figure B – 4 / JT-G722.2 に示す。

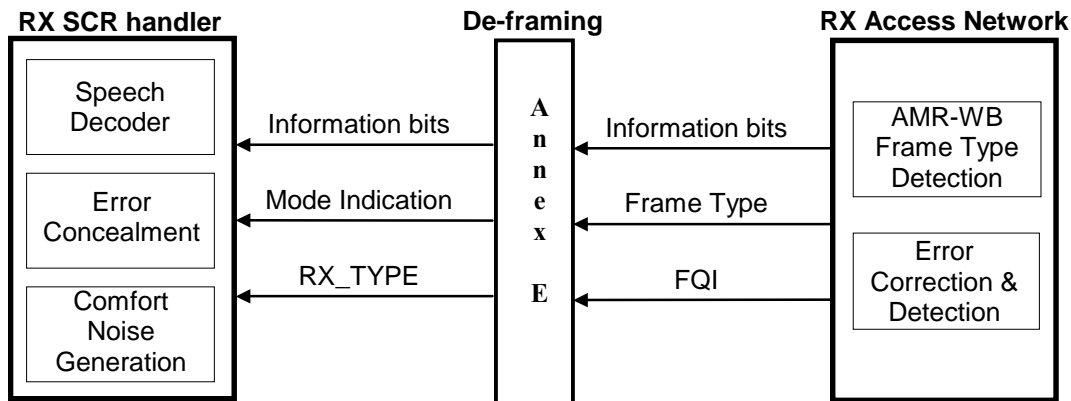


Figure B - 4 / JT-G722.2 Block diagram of the receive side SCR functions
(ITU-T G.722.2)

B. 4. 2. 1 動作概要

ANは、Table B - 2 / JT-G722.2 に記述されるように、RX_TYPE を用いて分類された全ての受信伝送フレームを受信SCRハンドラに送出する（TTC標準JT-G722.2 付属資料EおよびGを参照）。受信SCRはそれに応じて処理する。

Table B - 2 / JT-G722.2 RX_TYPE identifiers for AMR-WB
(ITU-T G.722.2)

RX_TYPE	Information Bits
SPEECH_GOOD	Speech frame without detected errors.
SPEECH_BAD	(likely) speech frame with bad CRC (or estimated to be very bad by the RX part of the AN)
SPEECH_LOST	No frame received. Indicates that this frame was transmitted, but never received.
SID_FIRST	This SID-frame marks the beginning of a comfort noise period.
SID_UPDATE	Correct SID update frame
SID_BAD	Corrupt SID update frame (bad CRC; applicable only for SID_UPDATE frames)
NO_DATA	Nothing useable was received. The synthesis mode of the previous frame type is used.

B. 4. 2. 2 受信SCRハンドラへの要求

受信SCRハンドラは、受信側の全てのSCR動作を請け負っている。音声モードと擬似背景雑音モードの2つのメインモードから構成される。初期モードは音声モードである。

受信側のSCR動作は、以下に示す通りである。

- SPEECH_GOOD に分類されたフレームを受信した場合、受信SCRハンドラは音声モードに遷移する。SPEECH_GOOD に分類されたフレームを受信した場合は常に、受信SCRハンドラは音声復号器に直接フレームを送出する。
- 受信SCRハンドラが音声モードの場合、SPEECH_BAD, SPEECH_LOST または NO_DATA に分類され

るフレームは、TTC標準JT-G722.2付録1で定義されるように置換されてミュートされる。NO_DATAに分類されるフレームは、有効な音声情報がないSPEECH_LOSTフレームと同様に処理される。

受信SCRハンドラの誤り補償処理が、RX_TYPEのSPEECH_LOSTをサポートしない場合、SPEECH_LOSTに分類されたフレームは、SPEECH_BADに置換される。

- － SID_FIRST, SID_UPDATE, SID_BADに分類されるフレームにより、受信SCRハンドラがCOMFORT_NOISEモードに遷移して、TTC標準JT-G722.2付属資料Aで定義される様に擬似背景雑音を生成する。SID_BADフレームは、TTC標準JT-G722.2付録1で定義されるように置換されてミュートされる。
- － COMFORT_NOISEモードにおいては、受信SCRハンドラは、全ての使用不可能なフレーム(NO_DATA, SPEECH_BAD)を無視する。擬似背景雑音はタイムアウトが適用されるまで生成され続ける(TTC標準JT-G722.2付録1)。

B. 4. 3 AMR-WB SID情報フォーマット

送信SCRハンドラが、網からAMR-WBモードにおいてSCR動作を可能にするように命令された場合、SID_UPDATEフレームフォーマットはTTC標準JT-G722.2付属資料Eに従う。これがデフォルトで唯一必須のSCRハンドラの動作モードである。

付属資料C
(標準JT-G722. 2に対する)
固定小数点Cコード

C. 1 Cソースコードの構成

本付属資料は、TTC標準JT-G722. 2本体、付属資料A（擬似背景雑音特性）、付属資料B（ソースコントロールレート動作）、および付録1（誤りフレームあるいは損失フレームの誤り補償）の正確な実装のためのビットイグザクトなCコードの構造の概観を示すものである。本付属資料は、Cコードの内容および構成の概略を示すものである。（付属資料A、B、および付録1を含め）TTC標準JT-G722. 2のいくつかの部分における記述とANSI Cソースコードとの間に相違がある場合、ANSI Cコードのアルゴリズム記述が優先される。

Cコードは、いくつかのシステム上で検証されている。（完全なリストは、`readme.txt` ファイルを参照。）

ANSI-Cは、移植性が望ましいので、プログラミング言語として選ばれた。

C. 1. 1 Cソースコードの内容

Cコードは、すべてのファイルをルートレベルに持つ。

拡張子“c”を持つファイルはソースコードであり、拡張子“h”を持つファイルはヘッダファイルである。ROM用のデータは、ほとんどが拡張子“tab”を持つファイルに含まれる。

最終的なビットイグザクト性の検証は、TTC標準JT-G722. 2付属資料Dで記述されている適応マルチレート広帯域テストシーケンスを用いて行わなくてはならない。

`Makefile` は、Cコード検証済みの上記プラットフォーム毎に提供されている。ソフトウェアがインストールされると、このディレクトリにコンパイル済みの符号器および復号器（音声コーデックのビットイグザクトなC実行ファイル）およびすべてのオブジェクトファイルが生成される。

C. 1. 2 プログラムの実行

適応マルチレート広帯域コーデックは、下の2つのプログラムで実現される。

- (encoder) 音声符号器
- (decoder) 音声復号器

プログラムは次のように呼び出される。

- `encoder` [符号器オプション] <音声入力ファイル> <パラメータファイル>
- `decoder` <パラメータファイル> <音声出力ファイル>

音声ファイルは、16ビット線形PCM符号化された音声サンプルを含み、パラメータファイルは、符号化された音声データおよびいくつかの付加的なフラグを含む。

符号器および復号器のオプションは、引数を何も与えないでプログラムを実行することによって説明される。符号器および復号器の実行方法についてのより詳しい情報は、ファイル `readme.txt` 参照されたい。

C. 1. 3 コード階層

Table C-1 ~ 3 / JT-G722.2 は、VAD、DTX、および擬似背景雑音生成を含んだ、音声コーデックにおいて用いられる関数を示すコールグラフである。

それぞれの列はコールレベルを表し、それぞれのセルは関数を表す。関数は右隣のセルの関数への呼び出しを持つ。コールグラフにおける時間的な順序は、フレームの処理が進むにしたがって上から下に向かう。`printf()`、`fwrite()`、などのすべての標準的なC関数は省略されている。また、`add()`、`L_add()`、`mac()`、といった基本演算や、倍精度拡張演算（たとえば `L_Extract()`）はグラフには記載されていない。スタティックRAMの初期化（す

なわち_init 関数の呼び出し) も省略されている。

基本演算は深さの拡張としてカウントされない。そのため、このソフトウェアのもっとも深いレベルは6である。

符号器のコールグラフは、Table C-1 / JT-G722.2 および Table C-2 / JT-G722.2 の2つのコールグラフに分割されている。

Table C – 1 / JT-G722.2 Speech encoder call structure

(ITU-T G.722.2)

coder	Copy					
	Decim_12k8	Down_samp	Interpol (function)			
		Copy				
	Set_zero					
	HP50_12k8					
	Scale_sig					
	wb_vad	Filter_bank	Filter5			
			Filter3			
			Level_calculation			
			vad_decision	llog2		
				Noise_estimate_update	update_cntrl	
		Estimate_Speech	hangover_addition			
	tx_dtx_handler					
	Parm_serial					
	Autocorr					
	Lag_window					
	Levinson					
	Az_isp	Chebbs2				
	Int_isp	Isp_Az	Get_isp_pol			
	Isp_isf					
	Gp_clip_test_isf					
	Weight_a					
	Residu					
	Deemph2					
	LP_Decim2					
	Scale_mem_Hp_wsp					
	Pitch_med_ol	Hp_wsp				
		lsqrt_n				
	wb_vad_tone_detection					
	Med_olag	median5				
	dtx_buffer	Copy				
	dtx_enc	Find_frame_indices				
		Aver_isf_history				
		Qisf_ns	Sub_VQ			
			Disf_ns	Reorder_isf		
		Parm_serial				
		Pow2				
		Random				
		Dot_product12				
		lsqrt_n				
		Isf_isp				
	Isp_Az	Get_isp_pol				
	Synthesis	Copy				
		Syn_filt_32				
		Deemph_32				
		HP50_12k8				
		Random				
		Scale_sig				
		Dot_product12				
		lsqrt_n				
HP400_12k8						
Weight_a						
Syn_filt						
Filt_6k_7k						
Reset_encoder	Set_zero					
	Init_gp_clip					
	Init_Phase_dispersion	Set_zero				
Qpif_2s_36b	VQ_stage1					
	Sub_VQ					
	Dpif_2s_36b	Reorder_isf				
Qpif_2s_46b	VQ_stage1					
	Sub_VQ					
	Dpif_2s_46b	Reorder_isf				
Syn_filt						
Preemph2						
Pitch_fr4	Norm_Corr	Convolve				
	Interpol_4	lsqrt_n				
Gp_clip						
Pred_lt4						
Convolve						
G_pitch	Dot_product12					
Updt_tar						
Preemph						
Pit_shrp						
Cor_h_x						
ACELP_2t64_fx	Dot_product12					
	lsqrt_n					
ACELP_4t64_fx	See Table 2					
Q_gain2	Dot_product12					
	Pow2					
Gp_clip_test_gain_pit						
voice_factor	Dot_product12					

Table C – 2 / JT-G722.2 ACELP_4t64_fx call structure

(ITU-T G.722.2)

ACELP_4t64_fx	Dot_product12			
	lsqrt_n			
	cor_h_vec			
	search_ixiy			
	quant_1p_N1			
	quant_2p_2N1			
	quant_3p_3N1	quant_2p_2N1		
		quant_1p_N1		
	quant_4p_4N	quant_4p_4N1	Quant_2p_2N1	
		quant_1p_N1		
		quant_3p_3N1	Quant_2p_2N1	
			Quant_1p_N1	
		quant_2p_2N1		
	quant_5p_5N	quant_3p_3N1	Quant_2p_2N1	
			Quant_1p_N1	
		quant_2p_2N1		
	quant_6p_6N_2	quant_5p_5N	Quant_3p_3N1	quant_2p_2N1
				Quant_1p_N1
			quant_2p_2N1	
		quant_1p_N1		
		quant_4p_4N	quant_4p_4N1	quant_2p_2N1
			quant_1p_N1	
			quant_3p_3N1	quant_2p_2N1
				quant_1p_N1
			quant_2p_2N1	
		quant_2p_2N1		
		quant_3p_3N1	quant_2p_2N1	
		Quant_1p_N1		

Table C – 3 / JT-G722.2 Speech decoder call structure

(ITU-T G.722.2)

decoder	Rx_dtx_handler			
	Dtx_dec	Copy		
		Disf_ns	Reorder_isf	
		Serial_parm		
		Pow2		
		Random		
		Dot_product12		
		lsqrt_n		
	Serial_parm			
	Isf_isp			
	Isp_Az	Get_isp_pol		
	Copy			
	Synthesis	Copy		
		Syn_filt_32		
		Deemph_32		
		HP50_12k8		
		Oversamp_16k	Copy	Interpol
		Up_samp		
		Random		
		Scale_sig		
		Dot_product12		
		lsqrt_n		
		HP400_12k8		
		Isf_Extrapolation	Isf_isp	
		Isp_Az	Get_isp_pol	
		Weight_a		
	Syn_filt			
	Filt_6k_7k	Copy		
	Filt_7k	Copy		
	Reset_decoder	Set_zero		
	Init_Phase_dispersion	Set_zero		
	Dpifs_2s_36b	Reorder_isf		
	Dpifs_2s_46b	Reorder_isf		
	Int_isp	Isp_Az	Get_isp_pol	
	Lagconc	insertion_sort	Insert	
	Random			
	Pred_lt4			
	Random			
	DEC_ACELP_2t64_fx			
	DEC_ACELP_4t64_fx	dec_1p_N1		
		add_pulses		
		dec_2p_2N1		
		dec_3p_3N1	Dec_2p_2N1	
		dec_4p_4N	dec_1p_N1	dec_2p_2N1
			dec_4p_4N1	dec_1p_N1
			Dec_3p_3N1	Dec_2p_2N1
			Dec_1p_N1	Dec_1p_N1
			Dec_2p_2N1	
		dec_5p_5N	dec_3p_3N1	Dec_2p_2N1
			Dec_1p_N1	Dec_1p_N1
		Dec_2p_2N1		
dec_6p_6N_2		Dec_5p_5N	dec_3p_3N1	Dec_2p_2N1
			Dec_1p_N1	Dec_1p_N1
		dec_1p_N1	dec_4p_4N1	dec_2p_2N1
		dec_4p_4N	dec_1p_N1	Dec_2p_2N1
			Dec_3p_3N1	Dec_1p_N1
		Dec_2p_2N1		
	dec_2p_2N1			
	dec_3p_3N1	Dec_2p_2N1		
		Dec_1p_N1		
Preemph				
Pit_shrp				
D_gain2	Dot_product12			
	lsqrt_n			
	Median5			
	Pow2			
Scale_sig				
voice_factor	Dot_product12			
Phase_dispersion	Set_zero			
Agc2	lsqrt	lsqrt_n		
Set_zero				
Dtx_dec_activity_update	Copy			

C. 1. 4 変数、定数、テーブル

固定小数点実装において用いられている変数およびテーブルのデータタイプは、2の補数表現の符号付き整数型であり、以下のように定義される。

- **Word16** 16ビット変数
- **Word32** 32ビット変数

C. 1. 4. 1 Cコードで用いられている定数の詳細

本節は、cnst.h で定義されているすべてのグローバル定数のリストを示す。Table C-4 / JT-G722.2 を参照。

Table C-4 / JT-G722.2 Global constants

(ITU-T G.722.2)

Constant	Value	Description
L_TOTAL	384	total size of speech buffer.
L_WINDOW	384	window size in LP analysis
L_NEXT	64	Look-ahead size
L_FRAME	256	frame size in 12.8 kHz
L_FRAME16k	320	frame size in 16 kHz
L_SUBFR	64	Subframe size in 12.8 kHz
L_SUBFR16k	80	Subframe size in 16 kHz
NB_SUBFR	4	Number of subframes
M16k	20	order of LP filter in high-band synthesis in 6.60 mode
M	16	order of LP filter
L_FILT16k	15	Delay of down-sampling filter in 16 kHz
L_FILT	12	Delay of down-sampling filter in 12.8 kHz
GP_CLIP	15565	Pitch gain clipping
PIT_SHARP	27853	pitch sharpening factor
PIT_MIN	34	minimum pitch lag (all modes)
PIT_FR2	128	Minimum pitch lag with resolution $\frac{1}{2}$
PIT_FR1_9b	160	Minimum pitch lag with resolution for 9 bit quantization
PIT_FR1_8b	92	Minimum pitch lag with resolution for 8 bit quantization
PIT_MAX	231	maximum pitch lag
L_INTERPOL	(16+1)	length of filter for interpolation
OPL_DECIM	2	Decimation in open-loop pitch analysis
PREEMPH_FAC	22282	preemphasis factor
GAMMA1	30147	Weighting factor (numerator)
TILT_FAC	22282	tilt factor (denominator)
Q_MAX	8	scaling max for signal
RANDOM_INITSEED	21845	random init value
L_MEANBUF	3	Size of ISF buffer
ONE_PER_MEANBUF	10923	Inverse of L_MEANBUF

C. 1. 4. 2 Cコードで用いられている固定的なテーブルの詳細

本節は、すべての固定的なテーブルを、ソースファイル名およびテーブルの名前でソートして示す。すべてのテーブルのデータは **Word16** として宣言されている。Table C-5 / JT-G722.2 を参照。

Table C – 5 / JT-G722.2 Fixed tables

(ITU-T G.722.2)

File	Table name	Length	Description
C4t64fx.c	Tipos	36	starting points of iterations
Cod_main.c	HP_gain	16	High band gain table for 23.85 kbit/s mode
Cod_main.c	Interpol_frac	4	LPC interpolation coefficients
Cod_main.c	Isp_init	16	isp tables for initialization
Cod_main.c	Isf_init	16	isf tables for initialization
D_gain2.c	cdown_unusable	7	attenuation factors for codebook gain in lost frames
D_gain2.c	cdown_usable	7	attenuation factors for codebook gain in bad frames
D_gain2.c	pdown_unusable	7	attenuation factors for adaptive codebook gain in lost frames
D_gain2.c	pdown_usable	7	attenuation factors for adaptive codebook gain in bad frames
D_gain2.c	Pred	4	algebraic code book gain MA predictor coefficients
Dec_main.c	HP_gain	16	High band gain table for 23.85 kbit/s mode
Dec_main.c	Interpol_frac	4	LPC interpolation coefficients
Dec_main.c	Isp_init	16	isp tables for initialization
Dec_main.c	Isf_init	16	isf tables for initialization
Decim54.c	fir_down	120	Downsample FIR filter coefficients
Decim54.c	fir_up	120	Upsample FIR filter coefficients
Dtx.c	en_adjust	9	Energy scaling factor for each mode during comfort noise
Grid100.tab	grid	101	grid points at wich Chebyshev polynomials
Ham_wind.tab	Window	384	LP analysis window
Hp400.c	A	3	HP filter coefficients (denominator) in higher band energy estimation
Hp400.c	B	3	HP filter coefficients (numerator) in higher band energy estimation
Hp50.c	A	3	HP filter coefficients (denominator) in pre-filtering
Hp50.c	B	3	HP filter coefficients (numerator) in pre-filtering
Hp6k.c	Fir_6k_7k	31	Bandpass FIR filter coefficients for higher band generation
Hp7k.c	Fir_7k	31	Bandpass FIR filter coefficients for higher band in 23.85 kbit/s mode
Hp_wsp.c	A	3	HP filter coefficients (denominator) in open-loop lag gain computation
Hp_wsp.c	B	3	HP filter coefficients (numerator) in open-loop lag gain computation
Isp_isf.tab	slope	128	table to compute cos(x) in Lsf_Isp()
Isp_isf.tab	Table	129	table to compute acos(x) in Lsp_Isf()
Lag_wind.tab	lag_h	16	high part of the lag window table
Lag_wind.tab	lag_l	16	low part of the lag window table
Lp_dec2.c	h_fir	5	HP FIR filter coefficients in open-loop lag search
Math_op.c	table_sqrt	49	table used in inverse square root computation
Math_op.c	table_pow2	33	table used in power of two computation
P_med_ol.tab	Corrweight	199	weighting of the correlation function in open loop LTP search
Ph_disp.c	ph_imp_low	64	phase dispersion impulse response
Ph_disp.c	ph_imp_mid	64	phase dispersion impulse response
Pitch_fr4.c	inter4_1	32	interpolation filter coefficients
Pred_lt4.c	inter4_2	128	interpolation filter coefficients
Q_gain2.c	pred	4	algebraic code book gain MA predictor coefficients
Q_gain2.tab	t_qua_gain6b	2*64	gain quantization table for 6-bit gain quantization
Q_gain2.tab	t_qua_gain7b	2*128	gain quantization table for 7-bit gain quantization
Qisf_ns.tab	dico1_isf_noise	2*64	1 st ISF quantizer for comfort noise
Qisf_ns.tab	dico2_isf_noise	3*64	2 nd ISF quantizer for comfort noise
Qisf_ns.tab	Dico3_isf_noise	3*64	3 rd LSF quantizer for comfort noise
Qisf_ns.tab	Dico4_isf_noise	4*32	4 th LSF quantizer for comfort noise
Qisf_ns.tab	Dico5_isf_noise	4*32	5 th LSF quantizer for comfort noise
Qisf_ns.tab	mean_isf_noise	16	ISF mean for comfort noise
Qpisf_2s.tab	dico1_isf	9*256	1 st ISF quantizer of the 1 st stage
Qpisf_2s.tab	Dico2_isf	7*256	2 nd ISF quantizer of the 1 st stage
Qpisf_2s.tab	Dico21_isf	3*64	1 st ISF quantizer of the 2 nd stage (not the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico21_isf_36b	5*128	1 st ISF quantizer of the 2 nd stage (the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico22_isf	3*128	2 nd ISF quantizer of the 2 nd stage (not the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico22_isf_36b	4*128	2 nd ISF quantizer of the 2 nd stage (the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico23_isf	3*128	3 rd ISF quantizer of the 2 nd stage (not the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico23_isf_36b	7*64	3 rd ISF quantizer of the 2 nd stage (the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico24_isf	3*32	4 th ISF quantizer of the 2 nd stage (not the 6.60 kbit/s mode)
Qpisf_2s.tab	Dico25_isf	4*32	5 th ISF quantizer of the 2 nd stage (not the 6.60 kbit/s mode)
Qpisf_2s.tab	Mean_isf	16	ISF mean

C. 1. 4. 3 Cコードで用いられるスタティック変数

本節では、音声符号器および復号器それぞれのスタティック変数を記載する2つのテーブルを示す。すべてのスタティック変数は、Cの構造体として宣言されている。Table C-6 / JT-G722.2 および Table C-7 / JT-G722.2 を参照。

Table C-6 / JT-G722.2 Speech encoder static variables
(ITU-T G.722.2)

Struct name	Variable	Type[Length]	Description
Coder_State	mem_decim	Word16[30]	Decimation filter memory
	mem_sig_in	Word16[6]	Prefilter memory
	mem_preemph	Word16	Preemphasis filter memory
	old_speech	Word16[128]	speech buffer
	old_wsp	Word16[115]	buffer holding spectral weighted speech
	old_exc	Word16[248]	excitation vector
	mem_levinson	Word16[18]	Levinson memories
	lspold	Word16[16]	Old ISP vector
	ispold_q	Word16[16]	Old quantized ISP vector
	past_isfq	Word16[16]	past quantized ISF prediction error
	mem_wsp	Word16	Open-loop LTP deemphasis filter memory
	mem_decim2	Word16[3]	Open-loop LTP decimation filter memory
	mem_w0	Word16	weighting filter memory (applied to error signal)
	mem_syn	Word16[16]	synthesis filter memory
	tilt_code	Word16	Preemphasis filter memory
	old_wsp_max	Word16	Open loop scaling factor
	old_wsp_shift	Word16	Maximum open loop scaling factor
	Q_old	Word16	Old scaling factor
	Q_max	Word16[2]	Maximum scaling factor
	gp_clip	Word16[2]	memory of pitch clipping
	qua_gain	Word16[4]	Gain quantization memory
	old_T0_med	Word16	weighted open loop pitch lag
	ol_gain	Word16	Open-loop gain
	ada_w	Word16	weighting level depending on open loop pitch gain
	ol_wght_flg	Word16	switches lag weighting on and off
	old_ol_lag	Word16[5]	Open loop lag history
	hp_wsp_mem	Word16[9]	Open-loop lag gain filter memory
	old_hp_wsp	Word16[243]	Open-loop lag
	vadSt	VadVars*	see below in this table
	dtx_encSt	dtx_encState*	see below in this table
	first_frame	Word16	First frame indicator
	lsfold	Word16[16]	Old ISF vector
	L_gc_thres	Word16	Noise enhancer threshold
	mem_syn_hi	Word16[16]	synthesis filter memory (most significant word)
	mem_syn_lo	Word16[16]	synthesis filter memory (least significant word)
	mem_deemph	Word16	Deemphasis filter memory
	mem_sig_out	Word16[6]	HP filter memory in the synthesis
	mem_hp400	Word16[6]	HP filter memory
	mem_oversamp	Word16[2*12]	Oversampling filter memory
	mem_syn_hf	Word16[16]	Higher band synthesis filter memory
mem_hf	Word16[30]	Estimated BP filter memory (23.85 kbit/s mode)	
mem_hf2	Word16[30]	Input BP filter memory (23.85 kbit/s mode)	
mem_hf3	Word16[30]	Input LP filter memory (23.85 kbit/s mode)	
seed2	Word16	Random generation seed	
disp_mem	Word16[8]	Phase dispersion memory	
vad_hist	Word16	VAD history	
Gain_alpha	Word16	Higher band gain weighting factor (23.85 kbit/s mode)	
dtx_encState	Isf_hist	Word16[128]	LSP history (8 frames)
	Log_en_hist	Word16[8]	logarithmic frame energy history (8 frames)
	Hist_ptr	Word16	pointer to the cyclic history vectors
	Log_en_index	Word16	Index for logarithmic energy
	Cng_seed	Word16	Comfort noise excitation seed
	D	Word16[28]	ISF history distance matrix
	sumD	Word16[8]	Sum of ISF history distances
	dtxHangoverCount	Word16	is decreased in DTX hangover period
decAnaElapsedCount	Word16	counter for elapsed speech frames in DTX	

Struct name	Variable	Type[Length]	Description
vadState1	bckr_est	Word16[12]	background noise estimate
	ave_level	Word16[12]	averaged input components for stationary estimation
	old_level	Word16[12]	input levels of the previous frame
	sub_level	Word16[12]	input levels calculated at the end of a frame (lookahead)
	a_data5	Word16[5][2]	memory for the filter bank
	a_data3	Word16[6]	memory for the filter bank
	burst_count	Word16	counts length of a speech burst
	Hang_count	Word16	hangover counter
	Stat_count	Word16	stationary counter
	Vadreg	Word16	15 flags for intermediate VAD decisions
	Tone_flag	Word16	15 flags for tone detection
	sp_est_cnt	Word16	Speech level estimation counter
	Sp_max	Word16	Maximum signal level
	sp_max_cnt	Word16	Maximum level estimation counter
	Speech_level	Word16	Speech level
	prev_pow_sum	Word16	Power of previous frame

Table C – 7 / JT-G722.2 Speech decoder static variables

(ITU-T G.722.2)

Struct name	Variable	Type[Length]	Description
Decoder_State	old_exc	Word16[248]	excitation vector
	ispold	Word16[16]	Old ISP vector
	ifold	Word16[16]	Old ISF vector
	isf_buf	Word16[48]	ISF vector history
	past_isfq	Word16[16]	past quantized ISF prediction error
	tilt_code	Word16	Preemphasis filter memory
	Q_old	Word16	Old scaling factor
	Qsubfr	Word16	Scaling factor history
	L_gc_thres	Word16	Noise enhancer threshold
	mem_syn_hi	Word16[16]	synthesis filter memory (most significant word)
	mem_syn_lo	Word16[16]	synthesis filter memory (least significant word)
	mem_deemph	Word16	Deemphasis filter memory
	mem_sig_out	Word16[6]	HP filter memory in the synthesis
	mem_oversamp	Word16[24]	Oversampling filter memory
	mem_syn_hf	Word16[20]	Higher band synthesis filter memory
	mem_hf	Word16[30]	Estimated BP filter memory (23.85 kbit/s mode)
	mem_hf2	Word16[30]	Input BP filter memory (23.85 kbit/s mode)
	mem_hf3	Word16[30]	Input LP filter memory (23.85 kbit/s mode)
	seed	Word16	Random code generation seed for bad frames
	seed2	Word16	Random generation seed for higher band
	old_T0	Word16	Old LTP lag (integer part)
	old_T0_frac	Word16	Old LTP lag (fraction part)
	lag_hist	Word16[5]	LTP lag history
	dec_gain	Word16[23]	Gain decoding memory
	seed3	Word16	Random LTP lag generation seed for bad frames
	disp_mem	Word16[8]	Phase dispersion memory
	mem_hp400	Word16[6]	HP filter memory
	prev_bfi	Word16	Previous BFI
	state	Word16	BGH state machine memory
	first_frame	Word16	First frame indicator
	dtx_decSt	dtx_decState*	see below in this table
Vad_hist	Word16	VAD history	
dtx_decState	Since_last_sid	Word16	number of frames since last SID frame
	true_sid_period_inv	Word16	inverse of true SID update rate
	log_en	Word16	logarithmic frame energy
	old_log_en	Word16	previous value of log_en
	isf	Word16[16]	ISF vector
	Isf_old	Word16[16]	Previous ISF vector
	Cng_seed	Word16	Comfort noise excitation seed
	Isf_hist	Word16[128]	ISF vector history (8 frames)
	Log_en_hist	Word16[8]	logarithmic frame energy history
	Hist_ptr	Word16	index to beginning of LSF history
	dtxHangoverCount	Word16	counts down in hangover period
	DecAnaElapsedCount	Word16	counts elapsed speech frames after DTX
	sid_frame	Word16	flags SID frames

Struct name	Variable	Type[Length]	Description
	valid_data	Word16	flags SID frames containing valid data
	log_en_adjust	Word16	mode-dependent frame energy adjustment
	dtxHangoverAdded	Word16	flags hangover period at end of speech
	dtxGlobalState	Word16	DTX state flags
	data_updated	Word16	flags CNI updates

C. 2 ホーミング処理

ホーミング処理の原理は、本標準の本体に述べられている。本節は、9つの復号器ホーミングフレームの詳細についてのみを含む。それぞれのAMR-WBのコーデックモードについて、対応する復号器のホーミングフレームは決まった一組のパラメータを持つ。そのシリアルフォーマットのパラメータは、最初のシリアルビットがMSBに挿入されるような15ビット長のフォーマットのパラメータにパッキングされる。この15ビット長のパラメータは、実音声パラメータを表現するものではなく、それと比較してメモリ消費が少ないものになっている。Table C-8/JT-G722.2は、異なるモードに対する15ビット長フォーマットのホーミングフレームを示す。復号器では、シリアルフォーマットの受信音声パラメータがまず15ビット長のフォーマットに変換される。次に、得られたパラメータはホーミングフレームのテーブル値 (Table C-8/JT-G722.2) と比較される。

Table C-8/JT-G722.2 Table values for the decoder homing frame in 15-bit-long format for different modes (ITU-T G.722.2)

Mode	Value (MSB=b0)
0	3168, 29954, 29213, 16121, 64, 13440, 30624, 16430, 19008
1	3168, 31665, 9943, 9123, 15599, 4358, 20248, 2048, 17040, 27787, 16816, 13888
2	3168, 31665, 9943, 9128, 3647, 8129, 30930, 27926, 18880, 12319, 496, 1042, 4061, 20446, 25629, 28069, 13948
3	3168, 31665, 9943, 9131, 24815, 655, 26616, 26764, 7238, 19136, 6144, 88, 4158, 25733, 30567, 30494, 221, 20321, 17823
4	3168, 31665, 9943, 9131, 24815, 700, 3824, 7271, 26400, 9528, 6594, 26112, 108, 2068, 12867, 16317, 23035, 24632, 7528, 1752, 6759, 24576
5	3168, 31665, 9943, 9135, 14787, 14423, 30477, 24927, 25345, 30154, 916, 5728, 18978, 2048, 528, 16449, 2436, 3581, 23527, 29479, 8237, 16810, 27091, 19052, 0
6	3168, 31665, 9943, 9129, 8637, 31807, 24646, 736, 28643, 2977, 2566, 25564, 12930, 13960, 2048, 834, 3270, 4100, 26920, 16237, 31227, 17667, 15059, 20589, 30249, 29123, 0
7	3168, 31665, 9943, 9132, 16748, 3202, 28179, 16317, 30590, 15857, 19960, 8818, 21711, 21538, 4260, 16690, 20224, 3666, 4194, 9497, 16320, 15388, 5755, 31551, 14080, 3574, 15932, 50, 23392, 26053, 31216
8	3168, 31665, 9943, 9134, 24776, 5857, 18475, 28535, 29662, 14321, 16725, 4396, 29353, 10003, 17068, 20504, 720, 0, 8465, 12581, 28863, 24774, 9709, 26043, 7941, 27649, 13965, 15236, 18026, 22047, 16681, 3968

C. 3 ファイルフォーマット

本節は、符号器および復号器プログラムによって用いられるファイルフォーマットについて述べる。テストシーケンスも、ここで述べるファイルフォーマットを用いる。

C. 3. 1 音声ファイル (符号器の入力/復号器の出力)

符号器によって読み込まれ、復号器によって書き出される音声ファイルは、各ワードが14ビットで左詰めの音声サンプルを持った16ビットワードからなる。バイト順はホストのアーキテクチャに依存する。(例えば、SUNワークステーションでは最上位バイトが最初になり、PCでは最下位バイトが最初になるなど。) 符号

器および復号器のプログラムともに（320 サンプルの）完全なフレームのみを処理する。つまり、符号器は、入力ファイルの長さが $n*320+k$ ワードの場合でも n フレームしか処理せず、復号器によって生成されたファイルは、常に $n*320$ ワードの長さを持つということになる。

C. 3. 2 モード制御ファイル（符号器の入力）

符号器プログラムは、処理される音声のそれぞれのフレームに対する符号化モードを指定するモード制御ファイルをオプションとして読み込むことができる。このファイルは1つの音声フレームあたり1つの数値を持つテキストファイルである。各行が0～8のモード番号を持つ。

C. 3. 3 パラメータビットストリームファイル（符号器出力／復号器入力）

音声符号器によって生成される／音声復号器に入力されるファイルは、下記のフォーマットで任意の数のフレームを持つ。

3 G P P（デフォルト）および I T Uのビットストリームフォーマットに関する注記：

I T Uのビットストリームフォーマットでは、ストリームの先頭において、NO_DATA と SID_FIRSTのフレームタイプを区別できる可能性が限定され得る。符号器と復号器の間で、最初のハングオーバーフレームのある区間がカットされるような非常にまれな場合（例えば、ハンドオーバーやストリームの編集）においては、ストリームフォーマットが I T Uかデフォルトである 3 G P Pかにより、復号器の出力が異なってくる。

3 G P Pフォーマット（デフォルト）：

これは、3 G P Pで用いられるデフォルトのフォーマットである。テストベクトルに対してコーデックのテストを実施する場合には、このフォーマットを用いなければならない。

TYPE_OF_FRAME_TYPE	FRAME_TYPE	MODE_INFO	B1	B2	...	Bnn
--------------------	------------	-----------	----	----	-----	-----

それぞれのフィールドは、ビットストリームファイルにおける1つの Word16 の値に相当し、合計でフレームあたり $3+nn$ ワードすなわち $6+2nn$ バイトである。ここで、 nn はフレームにおける符号化されたビットの数である。各フィールドは以下の意味を持つ。

TYPE_OF_FRAME_TYPE transmit frame type, which is one of

TX_TYPE (0x6b21)

RX_TYPE (0x6b20)

If TYPE_OF_FRAME_TYPE is TX_TYPE,

FRAME_TYPE transmit frame type, which is one of

TX_SPEECH (0x0000)

TX_SID_FIRST (0x0001)

TX_SID_UPDATE (0x0002)

TX_NO_DATA (0x0003)

If TYPE_OF_FRAME_TYPE is RX_TYPE,

FRAME_TYPE transmit frame type, which is one of

RX_SPEECH_GOOD (0x0000)

RX_SPEECH_PROBABLY_DEGRADED (0x0001)
RX_SPEECH_LOST (0x0002)
RX_SPEECH_BAD (0x0003)
RX_SID_FIRST (0x0004)
RX_SID_UPDATE (0x0005)
RX_SID_BAD (0x0006)
RX_NO_DATA (0x0007)

B0...B2nn speech encoder parameter bits (i.e. the bitstream itself). Each **Bx** either has the value 0x0081 (for bit 0) or 0x007F (for bit 1).

MODE_INFO encoding mode information, which is one of

6.60 kbit/s mode (0x0000)
8.85 kbit/s mode (0x0001)
12.65 kbit/s mode (0x0002)
14.25 kbit/s mode (0x0003)
15.85 kbit/s mode (0x0004)
18.25 kbit/s mode (0x0005)
19.85 kbit/s mode (0x0006)
23.05 kbit/s mode (0x0007)
23.85 kbit/s mode (0x0008)

上のC. 3. 1節で述べたように、バイト順はホストのアーキテクチャに依存する。

ITUフォーマット (コマンドラインパラメータ `-itu` により有効) :

SYNC_WORD	DATA_LENGTH	B1	B2	...	Bnn
-----------	-------------	----	----	-----	-----

それぞれのフィールドは、ビットストリームファイルにおける1つの Word16 の値に相当し、合計でフレームあたり 2+nn ワードすなわち 4+2nn バイトである。ここで、nn はフレームにおける符号化されたビットの数である。符号化ビットは、ビット0が 0x007f で、ビット1が 0x0081 で表現される。各フィールドは以下の意味を持つ。

SYNC_WORD Word to ensure correct frame synchronization between the encoder and the decoder. It is also used to indicate the occurrences of bad frames.

In the encoder output: (0x6b21)

In the decoder input: Good frames (0x6b21)

Bad frames (0x6b20)

DATA_LENGTH Length of the speech data. Codec mode and frame type is extracted in the

decoder using this parameter:

DATA_LENGTH	PREVIOUS FRAME	CODEC MODE	FRAMETYPE
0	RX_SPEECH_GOOD/ RX_SPEECH_BAD	DTX	RX_SID_FIRST
0	OTHER THAN RX_SPEECH_GOOD/ RX_SPEECH_BAD	DTX	RX_NO_DATA
35	-	DTX	RX_SID_UPDATE/ RX_SID_BAD
132	-	6.60 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
177	-	8.85 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
253	-	12.65 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
285	-	14.25 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
317	-	15.85 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
365	-	18.25 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
397	-	19.85 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
461	-	23.05 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD
477	-	23.85 kbit/s	RX_SPEECH_GOOD/ RX_SPEECH_BAD

MIME / ファイルストレージ フォーマット (コマンドラインパラメータ -mime により有効) :

AMR-WBの単一チャネルのMIME/ファイルストレージフォーマットに関する詳細は、参照標準[2](5.1節および5.3節)に記述がある。このフォーマットは、マルチメディアメッセージングサービス(MMS)等で使用される。

付属資料D
(標準 J T - G 7 2 2 . 2 に対する)
デジタルテストシーケンス

D. 1 本付属資料の規定範囲

本付属資料は、TTC標準 J T - G 7 2 2 . 2、その付属資料A、B、および付録1において規定される適応マルチレート広帯域 (AMR-WB) 音声コーデックのデジタルテストシーケンスを規定するものである。これらのシーケンスは、適応マルチレート広帯域 (AMR-WB) 音声コーデック本体、音声検出、擬似背景雑音、および、ソースコントロールレート動作のビットイグザクト実現を検査するために用いる。

D. 2 概要

デジタルテストシーケンスは、適応マルチレート広帯域 (AMR-WB) 音声コーデック本体、音声検出、擬似背景雑音、および、ソースコントロールレート動作のビットイグザクト実現を検査するために必要なものである。

また、このテストシーケンスはANSI Cコードのインストール検証にも使用できる。

D. 3節はデジタルテストシーケンスのファイルフォーマットを述べており、D. 4節は音声コーデック本体用のテストシーケンスを記述し、D. 5節ではVAD、擬似背景雑音、および、ソースコントロールレート動作のテストシーケンスを記述している。

D. 6節はテストシーケンスとテスト用の音声コーデックとの間で同期をとる方法について述べる。

D. 3 テストシーケンスのフォーマット

本節では、適応マルチレート広帯域 (AMR-WB) 音声コーデック本体、音声検出、擬似背景雑音、および、ソースコントロールレート動作のテストシーケンスのフォーマットに関する情報を提供する。

D. 3. 1 ファイルフォーマット

テストシーケンスはPCバイト順(リトルエンディアン)で書かれており、アーカイブファイル(ZIPフォーマット)にまとめられている。

解凍すると、以下の3タイプのファイルが提供される。

- | | |
|-----------------------------------|---------|
| — 音声符号器の入力用ファイル : | *.INP |
| — 音声符号器出力との比較用、および、音声復号器入力用ファイル : | *.COD |
| — 復号器出力との比較用ファイル : | *.OUT |
| — モード切り替えテスト用のモード制御ファイル : | T22.MOD |

すべてのファイルフォーマットはTTC標準 J T - G 7 2 2 . 2 付属資料Cに記述されている。

D. 3. 2 コーデックホーミング

それぞれの*.INPファイルは、テストシーケンスの初めの部分に2つのホーミングフレーム(TTC標準 J T - G 7 2 2 . 2 付属資料C参照)を含んでいる。これらのホーミングフレームの機能は、音声符号器の状態変数を初期値にリセットすることである。正しくインストールされたANSI-Cシミュレーションの場合、すべての音声符号器の出力フレームは*.CODファイルの対応したフレームと一致しなければならない。また、正しいハードウェア実現の試験段階では、音声符号器の初めの出力フレームは不定であるため、*.CODファイルの最初のフレームと一致する必要はないが、残りすべての音声符号器の出力フレームは、*.CODファイルの対応するフレームと一致しなければならない。

*.CODファイルの2つのホーミングフレームの機能は、音声復号器の状態変数を初期値にリセットすること

である。正しくインストールされたANSI-Cシミュレーションの場合、すべての音声復号器の出力フレームは、*.OUT ファイルの対応するフレームと一致しなければならない。また、正しいハードウェア実現の試験段階では、音声復号器の初めの出力フレームは不定であるため、*.OUT ファイルの最初のフレームと一致する必要はないが、残りすべての音声復号器の出力フレームは、*.OUT ファイルの対応するフレームと一致しなければならない。

D. 4 音声コーデックのテストシーケンス

本節では、適応マルチレート広帯域（AMR-WB）音声コーデック本体の動作確認用に設計された、テストシーケンスについて述べる。

D. 4. 1 コーデック構成

音声符号器は、ソースコントロールレート動作をしないように設定されなければならない。

D. 4. 2 音声コーデックのテストシーケンス

D. 4. 2. 1 音声符号器のテストシーケンス

23種類の符号器入力シーケンスが提供されている。入力シーケンスのT00.INPからT03.INPに対し、振幅値は14ビット精度で与えられている。有音レベルはdBovで与えられている。

- T00.INP - 合成ハーモニック信号。ピッチ遅延は34から231サンプルへ緩やかに変化。最小および最大振幅は-1475と+5952。
- T01.INP - 合成ハーモニック信号。ピッチ遅延は231から34サンプルへ緩やかに変化。振幅飽和点は-5386および+21707。
- T02.INP - 50 Hzから7000 Hzに変化する方形波掃引。振幅の範囲は±32767。
- T03.INP - 50 Hzから7000 Hzに変化するサイン波掃引。振幅の範囲は±6217。
- T04.INP - 女声、周囲ノイズ、有音レベル：-22.5 dBov、P.341 フィルタ処理。
- T05.INP - 男声、周囲ノイズ、有音レベル：-29.9 dBov、P.341 フィルタ処理。
- T06.INP - 女声と男声、周囲ノイズ、有音レベル：-36.1 dBov、P.341 フィルタ処理。
- T07.INP - 女声と男声、周囲ノイズ、有音レベル：-45.8 dBov、P.341 フィルタ処理。
- T08.INP - 女声と男声、周囲ノイズ、有音レベル：-7.7 dBov、P.341 フィルタ処理。
- T09.INP - 女声と男声、Hoth ノイズ、有音レベル：-37.4 dBov、P.341 フィルタ処理。
- T10.INP - 女声と男声、Hoth ノイズ、有音レベル：-27.3 dBov、P.341 フィルタ処理。
- T11.INP - 女声と男声、Hoth ノイズ、有音レベル：-16.9 dBov、P.341 フィルタ処理。
- T12.INP - 女声と男声、周囲ノイズ、有音レベル：-46.0 dBov、P.341 フィルタ処理。
- T13.INP - 音声、非常に高い car ノイズ、非常に低い car ノイズ、P.341 フィルタ処理。
- T14.INP - 女声と男声、周囲ノイズ、有音レベル：-26.0 dBov、P.341 フィルタ処理。
- T15.INP - 女声と男声、降雨ノイズ、有音レベル：-37.2 dBov、P.341 フィルタ処理。
- T16.INP - 女声と男声、降雨ノイズ、有音レベル：-26.5 dBov、P.341 フィルタ処理。
- T17.INP - 女声と男声、降雨ノイズ、有音レベル：-16.4 dBov、P.341 フィルタ処理。このファイルはホーミングフレーム試験を含む。
- T18.INP - 男声、有音レベル：-29.7 dBov、P.341 フィルタ処理、多数の零フレームあり。
- T19.INP - 子供の声、周囲ノイズ、有音レベル：-34.7 dBov、P.341 フィルタ処理。
- T20.INP - コーデックのLPCベクトル量子化コードブックとROMテーブルの試験用シーケンス。
- T21.INP - 零信号シーケンス。
- T22.INP - モード切り替え試験用の音声シーケンス。

これらの入力シーケンスを使用した出力は、試験する適応マルチレートモードによって異なる。以下で使用する表記<mode>は、試験するモードの番号、すなわち、2385, 2305, 1985, 1825, 1585, 1425, 1265, 885, 660 のどれか1つに置き換えるべきものである。

T00.INP と T01.INP のシーケンスは、適応マルチレート広帯域音声符号器のピッチラグの試験用に設計されている。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、それぞれ、T00_<mode>.COD と T01_<mode>.COD のシーケンスに指定されたものと一致しなければならない。

T02.INP と T03.INP のシーケンスは、L P C分析の試験に特に適したものであり、同時に、飽和の様々な問題を見つけるために適している。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、それぞれ、T02_<mode>.COD と T03_<mode>.COD のシーケンスに指定されたものと一致しなければならない。

T04.INP と T05.INP のシーケンスは、多くの低周波数成分を含んでいる。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、それぞれ、T04_<mode>.COD と T05_<mode>.COD のシーケンスに指定されたものと一致しなければならない。

T18.INP と T21.INP のシーケンスは、音声区間の間に“オールゼロ”のフレーム（無音）を含んでいる。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、それぞれ、T18_<mode>.COD と T21_<mode>.COD のシーケンスに指定されたものと一致しなければならない。

T20.INP のシーケンスは、コーデックのL P CコードのインデックスとR O Mテーブルのインデックスを試験するために設計されたものである。

T06.INP～T17.INP と T19.INP のシーケンスは、テストシーケンスのセットに、様々な入力特性（背景ノイズ）や入力レベルを導入する視点から選ばれたものである。T17.INP にはホーミングフレームの試験も含まれている。T17.INP には、32000 サンプル、16000 サンプル、48000 サンプルから始まる、それぞれ、320 サンプル、640 サンプル、960 サンプルの長さのホーミングフレームがある。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、それぞれ、T06_<mode>.COD～T17_<mode>.COD のシーケンスに指定されたものと一致しなければならない。

T22.INP のシーケンスは、符号器のモード切り替えを試験するために設計されたものである。モード切り替えを試験するために、このシーケンスは、モード制御ファイル T22.MOD と共に用いられる。モード制御ファイルのフォーマットについては、T T C標準 J T - G 7 2 2 . 2 付属資料Cを参照すること。正しく実現されていれば、結果として得られる音声符号器の出力パラメータは、T22.COD のシーケンスに指定されたものと一致しなければならない。T22.COD は様々なコーデックモードでのパラメータフレームを含んでいる。

D. 4. 2. 2 音声復号器のテストシーケンス

固定モードの試験用に、22×9種類の音声復号器入力シーケンス TXX_<mode>.COD (XX = 00..21, <mode> = {2385, 2305, 1985, 1825, 1585, 1425, 1265, 885 または 660})が提供されている。これらは、モード毎に1セットずつの、TXX.INP シーケンスに対応する出力である。正しく実現されていれば、結果として得られる音声復号器の出力は、対応する TXX_<mode>.OUT のシーケンスと一致しなければならない。

切り替え試験の復号器入力 T22.COD は、T22.OUT シーケンスと一致した復号器出力にならなければならない。復号器の切り替え試験では、ファイルフォーマットに従って、モード情報が.COD ファイルに含まれているので、特別なモード制御ファイルは必要ない。

D. 4. 2. 3 コーデックホーミングシーケンス

上述したテストシーケンスの他に、コーデック試験を支援するために、ホーミングシーケンスが提供されている。T23.INP は符号器ホーミングフレームを1つ含んでいる。シーケンス T23_<mode>.COD (<mode> = {2385, 2305, 1985, 1825, 1585, 1425, 1265, 885 または 660})は、対応するモード毎に、復号器ホーミングフレームを1つ

含んでいる。

すべてのファイルは、アーカイブ T.zip にまとめられている。

D. 5 ソースコントロールレート動作テストシーケンス

本節では、VADアルゴリズム、擬似背景雑音、および、ソースコントロールレート動作の試験用に設計されたテストシーケンスについて記述する。

テストシーケンス DTX1.*、DT2.*、DTX4.* および DTX5.*は、23.85 kbit/s の音声コーデックでだけ動作させなければならない。テストシーケンス DTX3.*は、すべての音声コーデックモードで動作させなければならない。

D. 5. 1 コーデック構成

VAD、擬似背景雑音、および、ソースコントロールレート動作は、音声コーデックと連動して試験しなければならない。音声符号器は、VADを使用したソースコントロールレートモードで動作する設定でなければならない。

D. 5. 2 テストシーケンス

DTXテストシーケンスはそれぞれ3つのファイルから構成される。

- 音声符号器の入力用ファイル： *.INP
- 音声符号器出力との比較用、および、音声復号器入力用ファイル： *.COD
- 復号器出力との比較用ファイル： *.OUT

*.COD と *.OUT のファイル名は DTXA_<mode>.*の形式であり、"A"はテストケース番号 (1, 2, 3, 4 または 5) であり、<mode>は音声コーデックモードを表す。

正しく実現されていれば、*.INP ファイルにより生成される音声符号器のパラメータは、*.COD ファイルで指定されるものと一致しなければならない。さらに、*.COD ファイルにより生成される音声復号器出力は、*.OUT ファイルで指定されるものと一致しなければならない。

D. 5. 2. 1 背景雑音推定用テストシーケンス

背景雑音推定アルゴリズムは、次のテストシーケンスにより試験される。

DTX1.*
DTX2.*

D. 5. 2. 2 トーン信号検出用テストシーケンス

トーン信号検出アルゴリズムは、次のテストシーケンスにより試験される。

DTX3.*

D. 5. 2. 3 実音声およびトーン

このテストシーケンスは、非常にクリーンな音声、辛うじて検出可能な音声、および、掃引された周波数トーンから構成されている。

DTX4.*

D. 5. 2. 4 信号対雑音比推定用テストシーケンス

フルレンジのSNR推定は、次のテストシーケンスを用いて試験される。

DTX5.*

D. 6 適応マルチレート音声符号器の 20ms フレーミング検出用シーケンス

符号器を試験する際、通常、符号器が音声入力の 20ms 区間をどこで開始するかについての情報は与えられていない。

以下に、特殊な同期シーケンスを用いて符号器の 20ms フレーミングを見つけるための手順を述べる。同期は 2つのステップにより実現される。初めに、ビット同期が検出されなければならない。次のステップで、フレーム同期を決めることができる。この手順は、最初のホームイングフレームを受信すると、コーデックが定義されたホーム状態になるという、適応マルチレートコーデックのコーデックホームイングの特徴を利用している。更にホームイングフレームを受信すると、コーデックの出力は所定信号となる。この出力は、他の動作を開始するために使用することができる。

D. 6. 1 ビット同期

音声符号器の入力は 14 ビット語長 (224 kbit/s, 14 bit 線形 PCM) の系列である。音声符号器の試験を開始する際、ビット同期 (符号器が、どこが最下位ビットで、どこが最上位ビットかを予見すること) について、何ら情報が与えられていない。

符号器ホームイングフレームは 320 サンプルから成り、すべて 16 進数の 0x0008 に設定されている。もし、そのような符号器ホームイングフレームが、2つ連続して符号器に入力されると、第 2 番目の符号器ホームイングフレームについての応答として、使用したコーデックモードに対応する復号器ホームイングフレームが、出力となるはずである。

ビット同期には 14 通りの可能性しかないので、コーデック各モードについて、最大 14 回の試行後に、ビット同期を獲得できる。それぞれの試行において、3つの連続した符号器ホームイングフレームが符号器に入力される。もし、対応する復号器ホームイングフレームが出力で検出されなければ、3つの入力フレームの相対的なビット位置が 1 ビットだけシフトされ、次の試行が行われる。使用されるコーデックモードの復号器ホームイングフレームが出力で検出されるとすぐに、ビット同期が検出され、第 1 のステップを終えることができる。

3つの連続した符号器ホームイングフレームが必要な理由は、この段階でフレーム同期が不明なためである。符号器が 2つの完全なホームイングフレームを読み込むことを確実にするためには、3つのフレームが入力されなければならない。符号器の 20 ms の区切りがどこであっても、符号器は常に、最低 2つの完全な符号器ホームイングフレームを読み込むことになる。

14 通りの異なる 3 連フレームの例は、シーケンス BITSYNC.INP で与えられる。

D. 6. 2 フレーム同期

ビット同期が見つかり、フレーム同期は、2つの同じフレームを連続して符号器に入力することにより検出することができる。符号器のフレーミングに関し、フレームのシーケンスの先頭が取り得る 320 通りの位置に依存して、320 通りの異なる出力シーケンスが存在する。

この特殊な同期シーケンスを符号器に入力する前に、1つの符号器ホームイングフレームにより、符号器は再度リセットされなければならない。もう 1つの符号器ホームイングフレームは、始動可能な復号器ホームイングフレームを出力で発生させるために必要である。この段階で、符号器のフレーミングは不明なので、符号器が最低 2つのホームイングフレームを確実に読み込むために、特殊な同期シーケンスよりも先に、3つの符号器ホームイングフレームが使用されなければならない。こうすることで、出力で最低 1つの復号器ホームイングフレームが生成され、出力データが記録可能な状態となる。

使用されたコーデックモードでの、最後の復号器ホームイングフレームの後は、その前の復号器ホームイングフレームとは異なる、2つの連続した出力フレームを検出することが要求される。

3つの符号器ホームイングフレームが前に付いた、特殊な同期シーケンスは SEQSYNC.INP で与えられる。

一般に、出力シーケンスは試験する適応マルチレート広帯域モードによって異なる。以下で使用する表記

<mode>は、試験するモードの番号、すなわち、2385, 2305, 1985, 1825, 1585, 1425, 1265, 885, 660 のどれか1つに置き換えるべきものである。

全320通りの出力シーケンスで、最後の復号器ホーミングフレームの後に続く、第2番目のフレームだけが、SYNC000_<mode>.COD～SYNC319_<mode>.COD で与えられている。これらの出力フレームはシーケンスSEQSYNC.INPを、初めのサンプルを零に設定して、位置0から319までシフトすることにより計算されたものである。各コーデックモードについて、この320種類の出力シーケンスの、それぞれ最後のフレームが、他のすべてのものと異なることは、最終的に確認済みである。

上記のファイル名にある3桁の数字は、符号器のフレーミングに関し、入力を遅延させたサンプル数を示している。逆方向への対応するシフトによって、使用したコーデックモードに関する符号器フレーミングとの位置あわせが達成できる。

D. 6. 3 同期シーケンスのフォーマットおよびサイズ

BITSYNC.INP:

このシーケンスは14通りの3連フレームから構成され、音声符号器入力テストシーケンスのフォーマットを持つ。したがって、そのサイズは以下ようになる。

$$\text{SIZE (BITSYNC.INP)} = 14 * 3 * 320 * 2 \text{ bytes} = 26880 \text{ bytes}$$

SYNCXXX_<mode>.COD:

これらのシーケンスは、それぞれ1つの符号器出力フレームから成り、音声符号器出力テストシーケンスのフォーマットを持つ。これらのフレーム内のTX/RX_TYPEの値は、送信フレームタイプを示すため固定されており、FRAME_TYPEおよびMODE_INFOフィールドは、送信フレームタイプとそれに対応する符号化モード情報に設定されている。

これらのサイズは以下ようになる：

$$\text{SIZE (SYNCXXX_2385.COD)} = (477 + 3) * 2 \text{ bytes} = 960 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_2305.COD)} = (461 + 3) * 2 \text{ bytes} = 928 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_1985.COD)} = (397 + 3) * 2 \text{ bytes} = 800 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_1825.COD)} = (365 + 3) * 2 \text{ bytes} = 736 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_1585.COD)} = (317 + 3) * 2 \text{ bytes} = 640 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_1425.COD)} = (285 + 3) * 2 \text{ bytes} = 576 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_1265.COD)} = (253 + 3) * 2 \text{ bytes} = 512 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_885.COD)} = (177 + 3) * 2 \text{ bytes} = 360 \text{ bytes}$$

$$\text{SIZE (SYNCXXX_660.COD)} = (132 + 3) * 2 \text{ bytes} = 270 \text{ bytes}$$

すべてのファイルは、アーカイブS.zipにまとめられている。

付属資料E
(標準 J T - G 7 2 2 . 2 に対する)
フレーム構造

E. 1 本付属資料の規定範囲

E. 3 節では、適応マルチレート広帯域 (AMR-WB) 音声コーデックに関する 3 G P P フレームフォーマット (以下、汎用フレームフォーマットと呼ぶ) について記述する。異なるシステム間、またシステムの要素間で音声フレームを接続する際、本フォーマットは、共通の参照ポイントとして使用されるべきである。本汎用フレームフォーマット間の適切なマッピングは、それぞれのシステム要素内と要素間で使用される。

E. 4 節では、AMR-WB フレームのオクテット整合が要求される場合 (例えば、パケットを基本としたアプリケーションのため) に使用される、第二のフレームフォーマットについて記述する。

E. 2 定義と略語

E. 2. 1 定義

本付属資料では、以下の用語と定義を適用する。

AMR-WBモード： 9つのAMR-WBコーデックビットレートの中の1つを意味する。また0から8のインデックスでも表現する。ここで0を6.60kbit/sモード、8を23.85kbit/sモードにマッピングする。

AMR-WBコーデックモード： AMR-WBモードと同義。

TX_TYPE： 送信フレームの分類 (T T C標準 J T - G 7 2 2 . 2 付属資料Bに定義される)。

RX_TYPE： 受信フレームの分類 (T T C標準 J T - G 7 2 2 . 2 付属資料Bに定義される)。

E. 2. 2 略語

本付属資料では、以下の略語を適用する。

C R C	巡回冗長検査 (Cyclic Redundancy Check)
F Q I	フレーム品質表示 (Frame Quality Indicator)
G S M	G S M (Global System for Mobile Communications)
L S B	最下位ビット (Least Significant Bit)
M S B	最上位ビット (Most Significant Bit)
R X	受信 (Receive)
S C R	ソースコントロールレート動作 (Source Controlled Rate (operation))
S I D	無音挿入記述子 (擬似背景雑音フレーム) (Silence Insertion Descriptor (Comfort Noise Frame))
T X	送信 (Transmit)

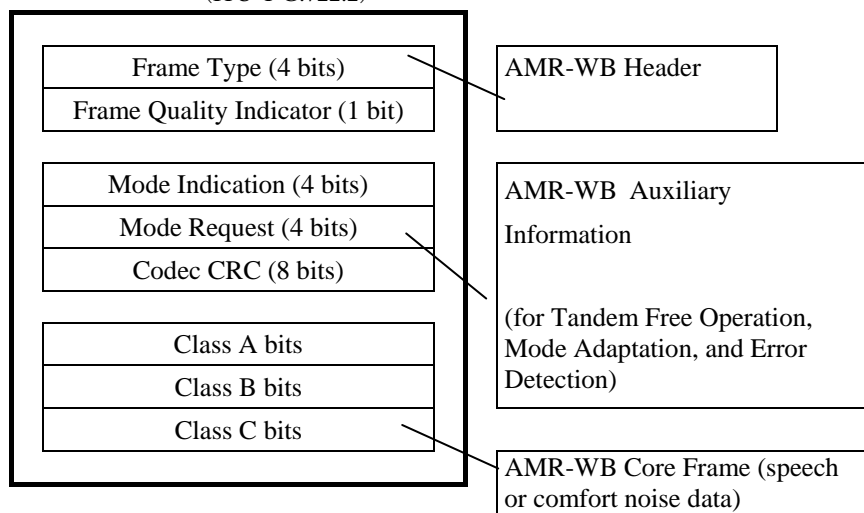
E. 3 AMR-WBコーデックインタフェースフォーマット1 (AMR-WB I F 1)

本節では、AMR-WB音声コーデックの音声、擬似背景雑音フレーム双方の汎用フレームフォーマットについて記述する。このフォーマットは、AMR-WBインタフェースフォーマット1 (AMR-WB I F 1) と呼ばれる。E. 4 節では、パケットを基本としたアプリケーション向けの、オクテット整合付きのAMR-WBインタフェースフォーマット2 (AMR-WB I F 2) について記述する。

各AMR-WBコーデックモードは、Figure E-1/JT-G722.2に示される、汎用フレーム構造に従う。このフレームは、AMR-WBヘッダ、AMR-WB補助情報と、AMR-WBコアフレームの3つの部分に分割される。AMR-WBヘッダ部は、フレームタイプとフレーム品質表示フィールドを含む。AMR-WB補助

情報部は、モード表示、モード要求とコーデックCRCフィールドを含む。AMR-WBコアフレーム部は、音声パラメータビットや、擬似背景雑音フレームの場合には、擬似背景雑音パラメータビットで構成される。擬似背景雑音フレームの場合、BクラスとCクラスのビットは省略され、擬似背景雑音パラメータがAMR-WBコアフレームのAクラスビットと入れ代わる。

Figure E-1 / JT-G722.2 Generic AMR-WB frame structure
(ITU-T G.722.2)



E. 3. 1 AMR-WBヘッダとAMR-WB補助情報

本節では、Figure E-1 / JT-G722.2 のAMR-WBヘッダについて記述する。補助情報フィールドは、ある種の無線システム（例えば、GSM）のインタフェースで応用される。

E. 3. 1. 1 フレームタイプ、モード表示とモード要求

Table E-1 a / JT-G722.2 は、4ビットフレームタイプフィールドを規定する。フレームタイプは、9つのAMR-WBコーデックモード、擬似背景雑音フレーム、損失音声フレーム、または空フレームのうちの1つから、その用途を示すことができる。さらに、4つのフレームタイプインデックスが、将来の利用のために予約されている。同表は、それぞれが4ビットのフィールドを持ち、9つのAMR-WBコーデックモードの中の1つを特定する0から8の範囲だけを定義する、モード表示やモード要求フィールド用にも再利用される。

Table E-1 a / JT-G722.2 Interpretation of Frame Type,
(ITU-T G.722.2) Mode Indication and Mode Request fields.

Frame Type Index	Mode Indication	Mode Request	Frame content (AMR-WB mode, comfort noise, or other)
0	0	0	AMR-WB 6.60 kbit/s
1	1	1	AMR-WB 8.85 kbit/s
2	2	2	AMR-WB 12.65 kbit/s
3	3	3	AMR-WB 14.25 kbit/s
4	4	4	AMR-WB 15.85 kbit/s
5	5	5	AMR-WB 18.25 kbit/s
6	6	6	AMR-WB 19.85 kbit/s
7	7	7	AMR-WB 23.05 kbit/s
8	8	8	AMR-WB 23.85 kbit/s
9	-	-	AMR-WB SID (Comfort Noise Frame)
10-13	-	-	For future use
14	-	-	Speech lost
15	-	-	No Data (No transmission/No reception)
	-	-	

E. 3. 1. 2 フレーム品質表示

フレーム品質表示フィールドの内容は、Table E-1 b/JT-G722.2 に規定される。フィールドの長さは1ビットである。フレーム品質表示は、そのフレームのデータに誤りを含んでいるかどうかを示す。

Table E-1 b/JT-G722.2 Definition of Frame Quality Indicator
(ITU-T G.722.2)

Frame Quality Indicator (FQI)	Quality of data
0	Bad frame or Corrupted frame (bits may be used to assist error concealment)
1	Good frame

E. 3. 1. 3 TX_TYPE と RX_TYPE へのマッピング

Table E-1 c/JT-G722.2 は、TTC標準 JT-G722.2 付属資料 B で定義された TX_TYPE と RX_TYPE フレームが、AMR-WBヘッダデータ (FQI やフレームタイプ) へどのようにマッピングされるかを示す。

Table E-1 c/JT-G722.2 Mapping of Frame Quality Indicator and Frame Type
(ITU-T G.722.2) to TX_TYPE and RX_TYPE, respectively

Frame Quality Indicator	Frame Type Index	TX_TYPE or RX_TYPE	Comment
1	0-8	SPEECH_GOOD	The specific Frame Type Index depends on the bit-rate being used.
0	0-8	SPEECH_BAD	The specific Frame Type Index depends on the bit-rate being used. The corrupted data may be used to assist error concealment.
0	14	SPEECH_LOST	No useful information. An erased or stolen frame with no data usable to assist error concealment.
1	9	SID_FIRST or SID_UPDATE	SID_FIRST and SID_UPDATE are differentiated using one Class A bit: STI.
0	9	SID_BAD	
1	15	NO_DATA	Typically a non-transmitted frame.

E. 3. 1. 4 コーデックCRC

無線システムにおける誤り検出の目的で、フレームタイプ0～9の汎用AMR-WBコーデックフレームは、8ビットCRCと統合される。Figure E-1/JT-G722.2 のAMR-WB補助情報のコーデックCRCフィールドは、このCRCの値を含む。これら8つのパリティビットは、巡回生成多項式で生成される。

$$G(x)=D^8 + D^6 + D^5 + D^4 + 1$$

上式は、AMR-WBコアフレームのクラスAビット全体にわたって計算される。フレームタイプ0～8のクラスAビットは、E. 3. 2. 2節 (音声ビットに関する) で、フレームタイプ9は、E. 3. 2. 3節 (擬似背景雑音ビットに関する) で規定される。

Table E-1 a/JT-G722.2 のフレームタイプインデックスが、14または15の時は、汎用AMR-WBフレームの中にCRCフィールドは含まれない。

E. 3. 2 AMR-WBコアフレーム

本節では、Figure E-1/JT-G722.2 のAMR-WBコアフレームを説明する。音声ビットと擬似背景雑音ビットに関するAMR-WBコアフレームを個別に説明する。

E. 3. 2. 1 音声ビットのAMR-WBコアフレーム：ビット順序

本節では、AMR-WBコアフレームがどのように符号化音声データを伝えるかを説明する。音声符号器により形成されたビットを $\{s(1),s(2),\dots,s(K)\}$ と表すと、 K は、Table E-2/JT-G722.2 に示した音声符号器によって形成されたビット数となる。 $s(i)$ の表記法は、TTC標準 JT-G 7 2 2. 2 本体に従う。音声符号器出力ビットは、その主観的重要度により、順序付けられる。例として、本ビット順序は、音声データを無線インタフェースへ受け渡す際に、誤り保護の目的で利用される。E. 5節の Table E-1 0 ~ E-1 8/JT-G722.2 は、全9つのAMR-WBコーデックモードに関する、AMR-WB I F 1 ビット順序を規定する。これらの表で、音声ビットは、TTC標準 JT-G 7 2 2. 2 本体の対応する表に記載のとおり、相当する音声符号器で生成された順序で番号付けされる。再編成したビットは、以下、 $\{d(0),d(1),\dots,d(K-1)\}$ のように、重要度の減少順に表示される。

順序付けのアルゴリズムを擬似コードで記述する。

```
for j = 0 to K-1
    d(j) := s(tablem(j)+1);
```

ここで $table_m(j)$ は、AMR-WBモード $m=0.8$ にしたがって、E. 5節の該当する表を参照する。E. 5節の表は、一行毎に左から右へ読む。表の第一の要素はインデックス0となる。

E. 3. 2. 2 音声ビットのAMR-WBコアフレーム：クラス分け

再編成したビットは、それらの主観的重要度にしたがって、3つのクラスに細分化される。その3つの異なる重要度クラスは、結果として、網の様々な誤り保護の対象になり得る。

その重要度クラスとは、クラスA、クラスB、とクラスCである。クラスAは、最も誤りに敏感で、これらのビットのいかなる誤りも、概して、適切な誤り訂正を適用せずには復号されるべきでない、不正な音声フレームとしての結果を招くビットを含む。このクラスは、AMR-WB補助情報の中のコーデックCRCで保護される。クラスBとCは、誤り率の増加と共に徐々に音声品質を劣化させるが、その誤った音声フレームにおいて、通常は不快な異音を発生させずに復号可能であるビットを含む。クラスBのビットは、クラスCのビットに比べ、誤りに対し、より敏感である。重要度の順序付けはまた、3つの各々のクラス内においても振り分けるが、クラスの境界の隣接ビット間の主観的重要度の変化には、大きな段階幅はない。

各AMR-WBモードに関する、各クラス（クラスA、クラスB、クラスC）の音声ビット数を、Table E-2/JT-G722.2 に示す。Table E-2/JT-G722.2 の分類と重要度順序 $d(j)$ は、共に、すべての音声ビットを正確なクラスに割り当てするには十分である。例えば、AMR-WBコーデックモードが6.60の時は、そのクラスAビットは $d(0)..d(53)$ 、クラスBビットは $d(54)..d(131)$ 、そしてクラスCビットは無い。

Table E-2/JT-G722.2 Number of bits in Classes A, B, and C for each AMR-WB codec mode
(ITU-T G.722.2)

Frame Type	AMR-WB codec mode	Total number of bits	Class A	Class B	Class C
0	6.60	132	54	78	0
1	8.85	177	64	113	0
2	12.65	253	72	181	0
3	14.25	285	72	213	0
4	15.85	317	72	245	0
5	18.25	365	72	293	0
6	19.85	397	72	325	0
7	23.05	461	72	389	0
8	23.85	477	72	405	0

E. 3. 2. 3 擬似背景雑音ビットのAMR-WBコアフレーム

Table E-1 a / JT-G722.2 のフレームタイプインデックス 9~15 の、付加的なフレームに関して、AMR-WB コアフレームの内容を本節に記述する。これらは主に、TTC標準 JT-G 7 2 2. 2 付属資料Bで規定された、ソースコントロールレート動作に関連するフレームから成る。

付加的なフレームタイプのデータ内容(擬似背景雑音ビット)は、AMR-WB コアフレームで伝えられる。擬似背景雑音ビットは、すべてAMR-WB コアフレームのクラスAにマッピングされ、クラスBとCは使われない。これは表記上の規約の意味だけであり、擬似背景雑音ビットに関するクラス区分は意味を持たない。

AMR-WB擬似背景雑音ビット(フレームタイプインデックス9)に関する各クラス(クラスA、クラスB、とクラスC)のビット数を Table E-3 / JT-G722.2 に示す。SID_UPDATE と SID_FIRST の内容は、TTC標準 JT-G 7 2 2. 2 付属資料Bで規定された、3つの部分(SIDタイプ表示(STI)、モード表示(mi(i))と擬似背景雑音パラメータ(s(i)))に分割される。

AMR-WB音声符号器によって生成される擬似背景雑音パラメータビットを $s(i) = \{s(1), s(2), \dots, s(35)\}$ と表記する。記号 $s(i)$ は TTC標準 JT-G 7 2 2. 2 付属資料Aのそれに従う。これらのビットは、何ら並び替えることなく、AMR-WB符号器によってそれらが生成された順序で、番号付けられる。これらのビットには、SIDタイプ表示STIとモード表示ビット $mi(i) = \{mi(0), mi(1), mi(2), mi(3)\} = \{\text{LSB} \dots \text{MSB}\}$ が続く。したがって、AMR-WB SIDや擬似背景雑音ビット $\{d(0), d(1), \dots, d(39)\}$ は、以下の擬似コードによる規定で形成される。

```

for j = 0 to 34;
    d(j) := s(j+1);
d(35) := STI;
for j = 36 to 39;
    d(j) := mi(39-j).

```

Table E-3 / JT-G722.2 Bit classification for Frame Type 9: AMR-WB SID (Comfort Noise Frame)
(ITU-T G.722.2)

Frame Type Index	FQI	AMR-WB TX_TYPE or RX_TYPE	Total number of bits	Class A			Class B	Class C
				SID Type Indicator STI	Mode Indication mi(i)	Comfort Noise Parameters(i)		
9	1	SID_UPDATE	40	1 (= "1")	4	35	0	0
9	1	SID_FIRST	40	1 (= "0")	4	35 (= "0")	0	0
9	0	SID_BAD	40	1	4	35	0	0

AMR-WB非伝送フレームタイプ(14または15)は、AMR-WB補助情報やAMR-WBコアフレームが省略され、AMR-WBヘッダ情報(Figure E-1 / JT-G722.2に規定される)からなる。AMR-WBヘッダは、対応するフレームタイプとフレーム品質表示(Table E-1 c / JT-G722.2に規定される)を含んでいる。

E. 3. 3 汎用AMR-WBフレーム構成

汎用AMR-WBフレームは、AMR-WBヘッダ、AMR-WB補助情報と、AMR-WBコアフレームの順に、連続して構成される。フレームタイプのMSBは、第一のオクテット(以下のTable E-4 / JT-G722.2の例を参照)のビット8に配置され、フレームタイプのLSBは、ビット5に配置される。そして、次のパラメータであるフレーム品質表示などが続く。コーデックCRCとAMR-WBコアフレームを、オクテット境界へ揃えるために、FQIの後に3つの予備ビットが挿入される。AMR-WBコアフレームの最初のビット

d(0)は、オクテット4のビット8に配置される。汎用AMR-WBフレームの最後のビットは、AMR-WBコアフレームの最後のビットであり、E. 3. 2. 1節とE. 3. 2. 3節に規定される、音声ビットの最後のビット、または擬似背景雑音ビットの最後のビットである。Table E-4/JT-G722.2は、コーデックモード12.65kbit/sの例に関する構成を示し、Table E-5/JT-G722.2は、AMR-WB SIDフレームに関する構成を示す。

Table E-4/JT-G722.2 Mapping of an AMR-WB speech coding mode into the generic AMR-WB frame, AMR-WB IF1,

(ITU-T G.722.2) Example: AMR-WB 12.65 kbit/s (Mode Indication = 3), "good frame", Mode Request = 1.

Octet	Mapping of bits AMR-WB 12.65							
	MSB bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	LSB bit 1
1	Frame Type (=3)				FQI	spare		
	0	0	1	1	1	0	0	0
2	Mode Indication (=3)				Mode Request (=1)			
	MSB	...	LSB	MSB	...	LSB		
	0	0	1	1	0	0	0	1
3	Codec CRC							
	CRC(7)	CRC(6)	CRC(5)	CRC(4)	CRC(3)	CRC(2)	CRC(1)	CRC(0)
4	AMR-WB Core Frame (octet 1)							
	d(0)	d(1)	d(2)	d(3)	d(4)	d(5)	d(6)	d(7)
5..34	AMR-WB Core Frame (octets 2 to 31)							
	d(8)
35	AMR-WB Core Frame (octet 32)						Undefined	
	d(248)	d(249)	d(250)	d(251)	d(252)			

Table E-5/JT-G722.2 Mapping of an AMR-WB SID frame into the generic AMR-WB frame, AMR-WB IF1,
(ITU-T G.722.2) Example: AMR-WB SID_Update, "good frame", Mode Indication = 3, Mode Request = 2.

Octet	Mapping of bits AMR-WB SID							
	MSB bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	LSB bit 1
1	Frame Type (=9)				FQI	spare		
	1	0	0	0	1	0	0	0
2	Mode Indication				Mode Request (=2)			
	undefined				MSB	...	LSB	
				0	0	1	0	
3	Codec CRC							
	CRC(7)	CRC(6)	CRC(5)	CRC(4)	CRC(3)	CRC(2)	CRC(1)	CRC(0)
4	AMR-WB Core Frame (octet 1)							
	d(0)=s(1)	d(1)=s(2)	d(2)	d(3)	d(4)	d(5)	d(6)	d(7)
5..7	AMR-WB Core Frame (octets 2 to 4)							
	d(8)
8	AMR-WB Core Frame (octet 5)			STI	Mode Indication (=3)			
	d(32)	d(33)	d(34) = s(35)	1	MSB	...	LSB	
			1	0	0	1	1	

Table E-6/JT-G722.2に、各フィールドのビット数の観点から、すべての可能なAMR-WBフレームタイプの組み合わせをまとめる。

Table E – 6 / JT-G722.2 Number of bits for different fields in different AMR-WB frame compositions
(ITU-T G.722.2)

Frame Type Index	Frame Type	Frame Quality Indicator	Mode Indication	Mode Request	Codec CRC	Class A	Class B	Class C	Total
						AMR-WB Core Frame			
0	4	1	4	4	8	54	78	0	153
1	4	1	4	4	8	64	113	0	198
2	4	1	4	4	8	72	181	0	274
3	4	1	4	4	8	72	213	0	306
4	4	1	4	4	8	72	245	0	338
5	4	1	4	4	8	72	293	0	386
6	4	1	4	4	8	72	325	0	418
7	4	1	4	4	8	72	389	0	482
8	4	1	4	4	8	72	405	0	498
9	4	1	4	4	8	40	0	0	61
10-13	Not used								
14	4	1	0	0	0	0	0	0	4
15	4	1	0	0	0	0	0	0	4

E. 4 オクテット整合とAMR-WBインタフェースフォーマット2 (AMR-WB IF2)

本節では、AMR-WBコーデックに関するオクテット整合されたフレームタイプを規定する。このフォーマットは、例えば、AMR-WBコーデックが、TTC標準JTHシリーズと共に使用される際に有効である。そのフォーマットは、AMR-WBインタフェースフォーマット2 (AMR-WB IF2) として参照される。

AMR-WB IF2フレームは、Figure E-2/JT-G722.2 に示すように、4ビットのフレームタイプフィールド (E. 3. 1. 1節のAMR-WB IF1に関する規定による)、1ビットのフレーム品質表示フィールド (E. 3. 1. 2節のAMR-WB IF1に関する規定による) と、AMR-WBコアフレーム (E. 3. 2節のAMR-WB IF1に関する規定による) の連続として構成される。AMR-WBコアフレームフィールドの長さは、特定のフレームタイプによって決まる。各モードのAMR-WB IF2音声フレームの総ビット数は、通常は8の倍数にはならず、オクテット構成を実現するために、ビット詰めが必要とされる。

Figure E-2/JT-G722.2 Frame structure for AMR-WB IF2
(ITU-T G.722.2)

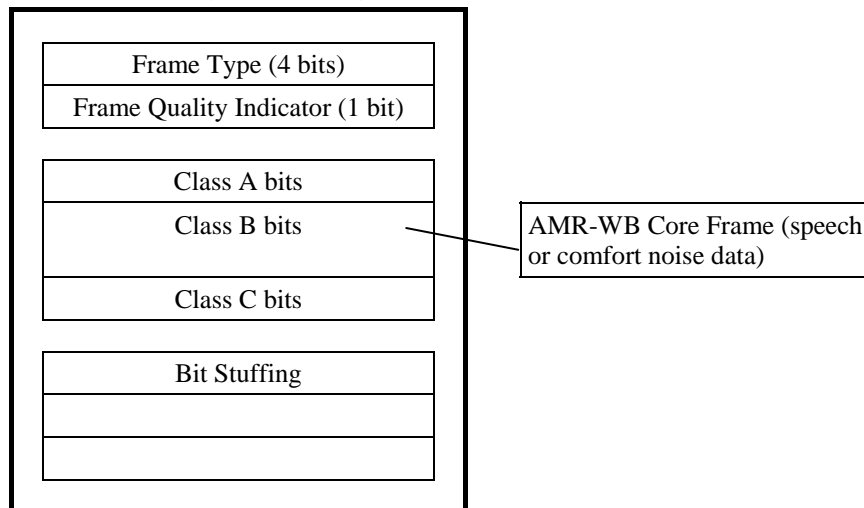


Table E-7 a/JT-G722.2 は、AMR-WB 8.85 kbit/s モードが、どのようにAMR-WB IF2へマッピングされるかを例示している。第一オクテット (オクテット1) の4MSBは、AMR-WB 8.85 kbit/s モード (AMR-WB IF1規格のTable E-1/JT-G722.2を参照) に関するフレームタイプ (=1) と、フレーム品質表示ビットから構成される。このフィールドには、AMR-WB IF1に関するTable E-2/JT-G722.2で記述されるとおり、64ビットのクラスAビットと113ビットのクラスBビットから成る、177ビットのAMR-WBコアフレーム音声ビット ($d(0)...d(176)$) が続く。結果として合計182ビットとなり、最も近い8の倍数である184ビットとするために、2ビットがビット詰めのために必要とされる。

Table E – 7 a /JT-G722.2 Example mapping of the AMR-WB speech coding mode 8.85kbit/s into AMR-WB IF2
(ITU-T G.722.2)

Octet	Mapping of bits AMR-WB 8.85 kbit/s							
	MSB bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	LSB bit 1
	Frame Type (=1)							
	MSB				LSB			
1	0	0	0	1	FQI	d(0)	d(1)	d(2)
2	d(3)	d(4)	d(5)	d(6)	d(7)	d(8)	d(9)	d(10)
3..22	d(11)
23	d(171)	d(172)	d(173)	d(174)	d(175)	d(176)	Stuffing bits	
							UB	UB

NOTE – The bits used for Bit Stuffing are denoted as UB (for "unused bit").

Table E – 7 b /JT-G722.2 に、Figure E – 2 /JT-G722.2 の各フィールドに関して、全フレームタイプが何ビット必要とするかの観点からの AMR – WB I F 2 フレームの構成を示す。

Table E – 8 /JT-G722.2 は、フレームタイプ 9 の AMR – WB コアフレーム擬似背景雑音ビットの、AMR – WB I F 2 へのマッピング方法を指定する。

Table E – 9 /JT-G722.2 は、空、または損失フレーム（“非伝送”または“音声損失”）に関するマッピングを指定する。

Table E – 7 b /JT-G722.2 Composition of AMR-WB IF2 Frames for all Frame Types
(ITU-T G.722.2)

Frame Type Index	Frame content	Number of bits in Frame Type	Number of bits in Frame Quality Indicator	Number of Bits in AMR-WB Core Frame	Number of Bits in Bit Stuffing	Number of octets (N)
0	AMR-WB 6.60 kbit/s	4	1	132	7	18
1	AMR-WB 8.85 kbit/s	4	1	177	2	23
2	AMR-WB 12.65 kbit/s	4	1	253	6	33
3	AMR-WB 14.25 kbit/s	4	1	285	6	37
4	AMR-WB 15.85 kbit/s	4	1	317	6	41
5	AMR-WB 18.25 kbit/s	4	1	365	6	47
6	AMR-WB 19.85 kbit/s	4	1	397	6	51
7	AMR-WB 23.05 kbit/s	4	1	461	6	59
8	AMR-WB 23.85 kbit/s	4	1	477	6	61
9	AMR-WB SID (Comfort Noise Frame)	4	1	40	3	6
10-13	For future use	–	–	–	–	–
14	speech lost	4	1	0	3	1
15	No Data (No transmission/ No reception)	4	1	0	3	1

Table E – 8 / JT-G722.2 Mapping of bits for Frame Type 9 (AMR-WB SID)
(ITU-T G.722.2)

Octet	Mapping of bits AMR-WB SID							
	bit 8	bit 7	bit 6	bit 5	bit 4	bit 3	bit 2	bit 1
	MSB				LSB			
	Frame Type (= 9)							
1	1	0	0	1	FQI	s1	s2	s3
2	s4	s5	s6	s7	s8	s9	s10	s11
3	s12	s13	s14	s15	s16	s17	s18	s19
4	s20	s21	s22	s23	s24	s25	s26	s27
5	s28	s29	s30	s31	s32	s33	s34	s35
	SID Type Indicator	Mode Indication mi(i)				Stuffing bits		
6	t1	mi(3)	mi(2)	mi(1)	mi(0)	UB	UB	UB

NOTE – (Bits s1 to s35 refer Annex A/G.722.2).

表にある無音挿入記述子に関して必要な、付加的な記述子のビット定義は次の通りである。SIDタイプ表示STIは{0=SID_FIRST, 1=SID_UPDATE}、音声モード表示(mi(0)-mi(3))は、Table E – 1 a / JT-G722.2の最初の9つの項目に準ずるAMR-WBコーデックモードである。パラメータmiにおいて、添え字3はMSBを指す。

Table E – 9 / JT-G722.2 Mapping of bit for Frame Type 14 (Speech Lost) and for Frame Type 15 (No Data)
(ITU-T G.722.2)

Transmitted Octets	Mapping of bits							
	MSB				LSB			
	Frame Type 14 = 1 1 1 0				Frame Type 15 = 1 1 1 1			
1	mi(3)	mi(2)	mi(1)	mi(0)	FQI	UB	UB	UB

E. 5 AMR-WBコアフレームビット順序表

本節は、様々なAMR-WBモードに対応する、AMR-WBコアフレーム音声ビットの順序付けに必要な表を含んでいる。これらの表は、E. 3. 2. 1節の $table_m(j)$ に相当し、ここで $m=0.8$ はAMR-WBモードである。表は、左から右へ読まれ、表の最初の要素（左最上端）をインデックス0、最後の要素（最後の行の最も右の要素）をインデックス $K-1$ とすると、 K は各モードにおける音声ビットの総数である。例えば、Table E-1 0/JT-G722.2 に記載のように、 $table_0(20)=60$ である。

Table E-1 0/JT-G722.2 Ordering of the speech encoder bits for the 6.60 kbit/s mode: $table_0(j)$
(ITU-T G.722.2)

0	5	6	7	61	84	107	130	62	85
8	4	37	38	39	40	58	81	104	127
60	83	106	129	108	131	128	41	42	80
126	1	3	57	103	82	105	59	2	63
109	110	86	19	22	23	64	87	18	20
21	17	13	88	43	89	65	111	14	24
25	26	27	28	15	16	44	90	66	112
9	11	10	12	67	113	29	30	31	32
34	33	35	36	45	51	68	74	91	97
114	120	46	69	92	115	52	75	98	121
47	70	93	116	53	76	99	122	48	71
94	117	54	77	100	123	49	72	95	118
55	78	101	124	50	73	96	119	56	79
102	125								

Table E-1 1/JT-G722.2 Ordering of the speech encoder bits for the 8.85 kbit/s mode: $table_1(j)$
(ITU-T G.722.2)

0	4	6	7	5	3	47	48	49	112
113	114	75	106	140	171	80	111	145	176
77	108	142	173	78	109	143	174	79	110
144	175	76	107	141	172	50	115	51	2
1	81	116	146	19	21	12	17	18	20
16	25	13	10	14	24	23	22	26	8
15	52	117	31	82	147	9	33	11	83
148	53	118	28	27	84	149	34	35	29
46	32	30	54	119	37	36	39	38	40
85	150	41	42	43	44	45	55	60	65
70	86	91	96	101	120	125	130	135	151
156	161	166	56	87	121	152	61	92	126
157	66	97	131	162	71	102	136	167	57
88	122	153	62	93	127	158	67	98	132
163	72	103	137	168	58	89	123	154	63
94	128	159	68	99	133	164	73	104	138
169	59	90	124	155	64	95	129	160	69
100	134	165	74	105	139	170			

Table E – 1 2 / JT-G722.2 Ordering of the speech encoder bits for the 12.65 kbit/s mode: *table₂(j)*

(ITU-T G.722.2)

0	4	6	93	143	196	246	7	5	3
47	48	49	50	51	150	151	152	153	154
94	144	197	247	99	149	202	252	96	146
199	249	97	147	200	250	100	203	98	148
201	251	95	145	198	248	52	2	1	101
204	155	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
156	31	102	205	9	33	11	103	206	54
157	28	27	104	207	34	35	29	46	32
30	55	158	37	36	39	38	40	105	208
41	42	43	44	45	56	106	159	209	57
66	75	84	107	116	125	134	160	169	178
187	210	219	228	237	58	108	161	211	62
112	165	215	67	117	170	220	71	121	174
224	76	126	179	229	80	130	183	233	85
135	188	238	89	139	192	242	59	109	162
212	63	113	166	216	68	118	171	221	72
122	175	225	77	127	180	230	81	131	184
234	86	136	189	239	90	140	193	243	60
110	163	213	64	114	167	217	69	119	172
222	73	123	176	226	78	128	181	231	82
132	185	235	87	137	190	240	91	141	194
244	61	111	164	214	65	115	168	218	70
120	173	223	74	124	177	227	79	129	182
232	83	133	186	236	88	138	191	241	92
142	195	245							

Table E – 1 3 / JT-G722.2 Ordering of the speech encoder bits for the 14.25 kbit/s mode: *table₃(j)*

(ITU-T G.722.2)

0	4	6	101	159	220	278	7	5	3
47	48	49	50	51	166	167	168	169	170
102	160	221	279	107	165	226	284	104	162
223	281	105	163	224	282	108	227	106	164
225	283	103	161	222	280	52	2	1	109
228	171	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
172	31	110	229	9	33	11	111	230	54
173	28	27	112	231	34	35	29	46	32
30	55	174	37	36	39	38	40	113	232
41	42	43	44	45	56	114	175	233	62
120	181	239	75	133	194	252	57	115	176
234	63	121	182	240	70	128	189	247	76
134	195	253	83	141	202	260	92	150	211
269	84	142	203	261	93	151	212	270	85
143	204	262	94	152	213	271	86	144	205
263	95	153	214	272	64	122	183	241	77
135	196	254	65	123	184	242	78	136	197
255	87	145	206	264	96	154	215	273	58
116	177	235	66	124	185	243	71	129	190
248	79	137	198	256	88	146	207	265	97
155	216	274	59	117	178	236	67	125	186
244	72	130	191	249	80	138	199	257	89
147	208	266	98	156	217	275	60	118	179
237	68	126	187	245	73	131	192	250	81
139	200	258	90	148	209	267	99	157	218
276	61	119	180	238	69	127	188	246	74
132	193	251	82	140	201	259	91	149	210
268	100	158	219	277					

Table E – 1 4 / JT-G722.2 Ordering of the speech encoder bits for the 15.85 kbit/s mode: *table_{4(f)}*
(ITU-T G.722.2)

0	4	6	109	175	244	310	7	5	3
47	48	49	50	51	182	183	184	185	186
110	176	245	311	115	181	250	316	112	178
247	313	113	179	248	314	116	251	114	180
249	315	111	177	246	312	52	2	1	117
252	187	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
188	31	118	253	9	33	11	119	254	54
189	28	27	120	255	34	35	29	46	32
30	55	190	37	36	39	38	40	121	256
41	42	43	44	45	56	122	191	257	63
129	198	264	76	142	211	277	89	155	224
290	102	168	237	303	57	123	192	258	70
136	205	271	83	149	218	284	96	162	231
297	62	128	197	263	75	141	210	276	88
154	223	289	101	167	236	302	58	124	193
259	71	137	206	272	84	150	219	285	97
163	232	298	59	125	194	260	64	130	199
265	67	133	202	268	72	138	207	273	77
143	212	278	80	146	215	281	85	151	220
286	90	156	225	291	93	159	228	294	98
164	233	299	103	169	238	304	106	172	241
307	60	126	195	261	65	131	200	266	68
134	203	269	73	139	208	274	78	144	213
279	81	147	216	282	86	152	221	287	91
157	226	292	94	160	229	295	99	165	234
300	104	170	239	305	107	173	242	308	61
127	196	262	66	132	201	267	69	135	204
270	74	140	209	275	79	145	214	280	82
148	217	283	87	153	222	288	92	158	227
293	95	161	230	296	100	166	235	301	105
171	240	306	108	174	243	309			

Table E – 1 5 / JT-G722.2 Ordering of the speech encoder bits for the 18.25 kbit/s mode: *table_{5(j)}*

(ITU-T G.722.2)

0	4	6	121	199	280	358	7	5	3
47	48	49	50	51	206	207	208	209	210
122	200	281	359	127	205	286	364	124	202
283	361	125	203	284	362	128	287	126	204
285	363	123	201	282	360	52	2	1	129
288	211	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
212	31	130	289	9	33	11	131	290	54
213	28	27	132	291	34	35	29	46	32
30	55	214	37	36	39	38	40	133	292
41	42	43	44	45	56	134	215	293	198
299	136	120	138	60	279	58	62	357	139
140	295	156	57	219	297	63	217	137	170
300	222	64	106	61	78	294	92	142	141
135	221	296	301	343	59	298	184	329	315
220	216	265	251	218	237	352	223	157	86
171	87	164	351	111	302	65	178	115	323
72	192	101	179	93	73	193	151	337	309
143	274	69	324	165	150	97	338	110	310
330	273	68	107	175	245	114	79	113	189
246	259	174	71	185	96	344	100	322	83
334	316	333	252	161	348	147	82	269	232
260	308	353	347	163	231	306	320	188	270
146	177	266	350	256	85	149	116	191	160
238	258	336	305	255	88	224	99	339	230
228	227	272	242	241	319	233	311	102	74
180	275	66	194	152	325	172	247	244	261
117	158	166	354	75	144	108	312	94	186
303	80	234	89	195	112	340	181	345	317
326	276	239	167	118	313	70	355	327	253
190	176	271	104	98	153	103	90	76	267
277	248	225	262	182	84	154	235	335	168
331	196	341	249	162	307	148	349	263	321
257	243	229	356	159	119	67	187	173	145
240	77	304	332	314	342	109	254	81	278
105	91	346	318	183	250	197	328	95	155
169	268	226	236	264					

Table E – 1 6 / JT-G722.2 Ordering of the speech encoder bits for the 19.85 kbit/s mode: *table_{6(f)}*

(ITU-T G.722.2)

0	4	6	129	215	304	390	7	5	3
47	48	49	50	51	222	223	224	225	226
130	216	305	391	135	221	310	396	132	218
307	393	133	219	308	394	136	311	134	220
309	395	131	217	306	392	52	2	1	137
312	227	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
228	31	138	313	9	33	11	139	314	54
229	28	27	140	315	34	35	29	46	32
30	55	230	37	36	39	38	40	141	316
41	42	43	44	45	56	142	231	317	63
73	92	340	82	324	149	353	159	334	165
338	178	163	254	77	168	257	153	343	57
248	238	79	252	166	67	80	201	101	267
143	164	341	255	339	187	376	318	78	328
362	115	232	242	253	290	276	62	58	158
68	93	179	319	148	169	154	72	385	329
333	344	102	83	144	233	323	124	243	192
354	237	64	247	202	209	150	116	335	268
239	299	188	196	298	94	195	258	123	363
384	109	325	371	170	370	84	110	295	180
74	210	191	106	291	205	367	381	377	206
355	122	119	120	383	160	105	108	277	380
294	284	285	345	208	269	249	366	386	300
297	259	125	369	197	97	194	286	211	281
280	183	372	87	155	283	59	348	327	184
76	111	330	203	349	69	98	152	145	189
66	320	337	173	358	251	198	174	263	262
126	241	193	88	388	117	95	387	112	359
287	244	103	272	301	171	162	234	273	127
373	181	292	85	378	302	121	107	364	346
356	212	278	213	65	382	288	207	113	175
99	296	374	368	199	260	185	336	331	161
270	264	250	240	75	350	151	60	89	321
156	274	360	326	70	282	167	146	352	81
91	389	266	245	177	235	190	256	204	342
128	118	303	104	379	182	114	375	200	96
293	172	214	365	279	86	289	351	347	357
261	186	176	271	90	100	147	322	275	361
71	332	61	265	157	246	236			

Table E – 1 7 /JT-G722.2 Ordering of the speech encoder bits for the 23.05 kbit/s mode: *table_{7(j)}*

(ITU-T G.722.2)

0	4	6	145	247	352	454	7	5	3
47	48	49	50	51	254	255	256	257	258
146	248	353	455	151	253	358	460	148	250
355	457	149	251	356	458	152	359	150	252
357	459	147	249	354	456	52	2	1	153
360	259	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
260	31	154	361	9	33	11	155	362	54
261	28	27	156	363	34	35	29	46	32
30	55	262	37	36	39	38	40	157	364
41	42	43	44	45	56	158	263	365	181
192	170	79	57	399	90	159	297	377	366
275	68	183	388	286	194	299	92	70	182
401	172	59	91	58	400	368	161	81	160
264	171	80	389	390	378	379	193	298	69
266	265	367	277	288	276	287	184	60	195
82	93	71	369	402	173	162	444	300	391
98	76	278	61	267	374	135	411	167	102
380	200	87	178	65	94	204	124	72	342
189	305	381	396	433	301	226	407	289	237
113	215	185	128	309	403	116	320	196	331
370	422	174	64	392	83	425	219	134	188
432	112	427	139	279	163	436	208	447	218
236	229	97	294	385	230	166	268	177	443
225	426	101	272	138	127	290	117	347	199
414	95	140	240	410	395	209	129	283	346
105	241	437	86	308	448	203	345	186	107
220	415	334	319	106	313	118	123	73	207
421	214	384	373	438	62	371	341	75	449
168	323	164	242	416	324	304	197	335	404
271	63	191	325	96	169	231	280	312	187
406	84	201	100	67	382	175	336	202	330
269	393	376	383	293	307	409	179	285	314
302	372	398	190	180	89	99	103	232	78
88	77	136	387	165	198	394	125	176	428
74	375	238	227	66	273	282	141	306	412
114	85	130	348	119	291	296	386	233	397
303	405	284	445	423	221	210	205	450	108
274	434	216	343	337	142	243	321	408	451
310	292	120	109	281	439	270	429	332	295
418	211	315	222	326	131	430	244	327	349
417	316	143	338	440	234	110	212	452	245
121	419	350	223	132	441	328	413	317	339
126	104	137	446	344	239	435	115	333	206
322	217	228	424	453	311	351	111	442	224
213	122	431	340	235	246	133	144	420	329
318									

Table E – 1 8 / JT-G722.2 Ordering of the speech encoder bits for the 23.85 kbit/s mode: *table_{8(j)}*

(ITU-T G.722.2)

0	4	6	145	251	360	466	7	5	3
47	48	49	50	51	262	263	264	265	266
146	252	361	467	151	257	366	472	148	254
363	469	149	255	364	470	156	371	150	256
365	471	147	253	362	468	52	2	1	157
372	267	19	21	12	17	18	20	16	25
13	10	14	24	23	22	26	8	15	53
268	31	152	153	154	155	258	259	260	261
367	368	369	370	473	474	475	476	158	373
9	33	11	159	374	54	269	28	27	160
375	34	35	29	46	32	30	55	270	37
36	39	38	40	161	376	41	42	43	44
45	56	162	271	377	185	196	174	79	57
411	90	163	305	389	378	283	68	187	400
294	198	307	92	70	186	413	176	59	91
58	412	380	165	81	164	272	175	80	401
402	390	391	197	306	69	274	273	379	285
296	284	295	188	60	199	82	93	71	381
414	177	166	456	308	403	98	76	286	61
275	386	135	423	171	102	392	204	87	182
65	94	208	124	72	350	193	313	393	408
445	309	230	419	297	241	113	219	189	128
317	415	116	328	200	339	382	434	178	64
404	83	437	223	134	192	444	112	439	139
287	167	448	212	459	222	240	233	97	302
397	234	170	276	181	455	229	438	101	280
138	127	298	117	355	203	426	95	140	244
422	407	213	129	291	354	105	245	449	86
316	460	207	353	190	107	224	427	342	327
106	321	118	123	73	211	433	218	396	385
450	62	383	349	75	461	172	331	168	246
428	332	312	201	343	416	279	63	195	333
96	173	235	288	320	191	418	84	205	100
67	394	179	344	206	338	277	405	388	395
301	315	421	183	293	322	310	384	410	194
184	89	99	103	236	78	88	77	136	399
169	202	406	125	180	440	74	387	242	231
66	281	290	141	314	424	114	85	130	356
119	299	304	398	237	409	311	417	292	457
435	225	214	209	462	108	282	446	220	351
345	142	247	329	420	463	318	300	120	109
289	451	278	441	340	303	430	215	323	226
334	131	442	248	335	357	429	324	143	346
452	238	110	216	464	249	121	431	358	227
132	453	336	425	325	347	126	104	137	458
352	243	447	115	341	210	330	221	232	436
465	319	359	111	454	228	217	122	443	348
239	250	133	144	432	337	326			

付属資料 F

(標準 JT-G 7 2 2. 2 に対する)

TTC 標準 JT-H 2 4 5 における AMR-WB 使用

F. 1 概要

本付属資料は、TTC 標準 JT-H 2 4 5 と共に使用するための、TTC 標準 JT-G 7 2 2. 2 適応マルチレート広帯域 (AMR-WB) 本体についてのパケットフォーマットと能力シグナリングを規定する。フォーマットと能力パラメータ両者は、シームレスな相互接続性を許容するため、AMR-WB RTP と完全に互換性がある。

F. 2 参考文献

[1] J. Sjöberg et al., "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate (AMR-WB) Audio Codecs", RFC 3267, June 2002.

[2] TTC 標準 JT-H 2 4 5 マルチメディア通信用制御プロトコル

F. 3 序論

リアルタイム伝送プロトコル (RTP) に含まれる適応マルチレート広帯域 (AMR-WB) 音声コーデックのパケットフォーマットと対応する能力シグナリングは、参考文献[1]にて規定される。このパケットフォーマットは、単に AMR-WB 音声フレームを RTP パケットのペイロードへカプセル化するための一方法ではなく、伝送エラーに対する耐性を強化したいくつかの選択可能なパケット化方法を導入している。さらに、パケット生成部能力のアウトバンドシグナリングのためのパラメータセットについても規定している。

本付属資料は参考文献[1]で定義されるパケットフォーマットの使用方法を規定すると共に、TTC 標準 JT-H 2 4 5 を共に使用する際の対応する能力の定義を提供するものである。本アプローチは TTC 標準 JT-H 2 4 5 と参考文献[1]で規定されたパラメータ (例: MIME シグナリング) を使用するシステム間の 1 対 1 マッピングを可能とし、完全な相互接続性を提供する。

F. 4 TTC 標準 JT-H 2 4 5 のための TTC 標準 JT-G 7 2 2. 2 本体 (適応マルチレート広帯域) の能力定義

Table F-1 / JT-G722.2 は、TTC 標準 JT-G 7 2 2. 2 本体 (AMR-WB) の能力識別子を定義している。Table F-2 / JT-G722.2 から F-10 / JT-G722.2 は、関係付けられた能力パラメータを定義している。オクテット構造への AMR-WB の音声フレームのパケット化とモードシグナリングの詳細は参考文献[1]で規定されている。TTC 標準 JT-H 2 4 5 [2]により提供されるリクエストモードの手順のため、参考文献[1]で規定されたインバンドモード要求処理[1]は使用されることはなく、パケットの CMR フィールドは常に「要求無し」を示す値 15 に設定される。

Table F-1 / JT-G722.2 能力識別子

(ITU-T G.722.2)

能力名:	JT-G 7 2 2. 2 (AMR-WB) オーディオ能力
能力クラス:	オーディオコーデック
能力識別子タイプ:	標準
能力識別子値:	{itu-t(0) recommendation(0)g(7)7222 generic-capabilities(1)0}
最大ビットレート:	239
NonCollapsingRaw:	本フィールドは含まない

transport :	本フィールドは含まない
-------------	-------------

Table F-2 / JT-G722.2 能力パラメータ — maxAI-sduFrames
(ITU-T G.722.2)

パラメータ名 :	MaxAI-sduFrames
パラメータ説明 :	本パラメータはCollapsing GenericParameterであり、AL-SDU毎のオーディオフレームの最大数を指定する。
パラメータ識別子値 :	0
パラメータステータス :	能力交換と論理チャネルングナリングを与える。モード要求は与えない。
パラメータタイプ :	UnsignedMin
Supersedes :	-

Table F-3 / JT-G722.2 能力パラメータ — bitRate
(ITU-T G.722.2)

パラメータ名 :	BitRate
パラメータ説明 :	本パラメータはnonCollapsing GenericParameterであり、AMRビットレートを指定する。本パラメータはモード要求とモード設定の定義 (Table F-5 / JT-G722.2参照) に使用される。0=6.60, 1=8.85, 2=12.65, 3=14.25, 4=15.85, 5=18.25, 6=19.85, 7=23.05, 8=23.85
パラメータ識別子値 :	1
パラメータステータス :	オプション
パラメータタイプ :	UnsignedMin
Supersedes :	-

Table F-4 / JT-G722.2 能力パラメータ — オクテット配列
(ITU-T G.722.2)

パラメータ名 :	OctetAlign
パラメータ説明 :	本パラメータはCollapsing GenericParameterであり、帯域効率モードもしくはオクテット配列モードが使用されることを指定する。パラメータが提供され、その値が真である場合、オクテット配列されたモードが使用され、その他の場合帯域効率モードが使用される。
パラメータ識別子値 :	2
パラメータステータス :	オプション
パラメータタイプ :	Logical
Supersedes :	-

Table F-5 / JT-G722.2 能力パラメータ — モード設定
(ITU-T G.722.2)

パラメータ名 :	ModeSet
パラメータ説明 :	本パラメータはCollapsing GenericParameterであり、提供される場合、現在のセッションに対して利用可能なモードのサブセットを指定する。提

	供されない場合、全てのモードが利用可能となる。
パラメータ識別子値：	3
パラメータステータス：	オプション
パラメータタイプ：	genericParameter パラメータ値 (unsignedMinのセット) は現在のセッションにおいてサポートされている JT-G 7 2 2 . 2 のモード (Table F - 3 / JT-G722.2 を参照) を示す。
Supersedes：	-

Table F - 6 / JT-G722.2 能力パラメータ - モード変更タイミング
(ITU-T G.722.2)

パラメータ名：	ModeChangePeriod
パラメータ説明：	本パラメータはCollapsing GenericParameterである。提供される場合、モード変更が許可されるインターバルN (フレーム数として) を指定する。初期フェーズでは任意であるが、モード変更はフレーム数Nの倍数の間隔で実行されなければならない。提供されない場合、モード変更はどのタイミングでも許可される。
パラメータ識別子値：	4
パラメータステータス：	オプション
パラメータタイプ：	UnsignedMin
Supersedes：	-

Table F - 7 / JT-G722.2 能力パラメータ - 隣接モード変更
(ITU-T G.722.2)

パラメータ名：	ModeChangeNeighbour
パラメータ説明：	本パラメータはCollapsing GenericParameterであり、提供される値が真ならば、指定されたモードセット内において隣接したモードへのモード変更は許可されることが指定される。値が提供されない、もしくは提供される値が偽である場合、指定されたモードセット内の全てのモードへのモード変更が許可される。
パラメータ識別子値：	5
パラメータステータス：	オプション
パラメータタイプ：	Logical
Supersedes：	-

Table F - 8 / JT-G722.2 能力パラメータ - crc
(ITU-T G.722.2)

パラメータ名：	Crc
パラメータ説明：	本パラメータはCollapsing GenericParameterであり、提供される値が真ならばフレームCRCがペイロードパケット内に含まれるということを指定する。値が提供されない、もしくは値が偽である場合、CRCは含まれない。CRCを有効とすることは、オクテット配列モード動作の使用を意味することに注意のこと。

パラメータ識別子値：	6
パラメータステータス：	オプション
パラメータタイプ：	Logical
Supersedes：	-

Table F-9/JT-G722.2 能力パラメータ – ロバストなソート処理
(ITU-T G.722.2)

パラメータ名：	RobustSorting
パラメータ説明：	本パラメータはCollapsing GenericParameterであり、提供される値が真ならば、ペイロード内に送られるフレームに対してロバストなソート処理が適用されるということを指定する。値が提供されない、もしくは値が偽である場合、単純なペイロードソート処理が使用される。ロバストなソート処理を有効とすることは、オクテット配列モード動作の使用を意味する。
パラメータ識別子値：	7
パラメータステータス：	オプション
パラメータタイプ：	Logical
Supersedes：	-

Table F-10/JT-G722.2 能力パラメータ – インタリーブ
(ITU-T G.722.2)

パラメータ名：	Interleaving
パラメータ説明：	本パラメータはCollapsing GenericParameterであり、提供される場合、フレームレベルのインタリーブが本セッションのペイロードに対して適用され、かつその値はインタリーブグループにおける最大フレーム数を指定する。提供されない場合、ペイロードにおけるフレームはインタリーブされない。インタリーブを有効とすることは、オクテット配列モード動作の使用を意味する。
パラメータ識別子値：	8
パラメータステータス：	オプション
パラメータタイプ：	UnsignedMin
Supersedes：	-

付録 1

(標準 J T - G 7 2 2 . 2 に対する)

誤りフレームあるいは損失フレームの誤り補償

1. 1 本付録の規定範囲

本付録は、ひとつあるいはそれ以上の誤り／損失音声フレームあるいは損失無音記述子フレーム (S I D) を受信した際に AMR - WB 音声コーデック受信端で用いられる、誤り補償の手法例を記述し、また、フレーム置換とミュートイング手法も記述する。

本付録に記述されているアルゴリズムは、TTC 標準 J T - G 7 2 2 . 2 付属資料 C における ANS I C コードの一部として利用可能である。本付録における記述と TTC 標準 J T - G 7 2 2 . 2 付属資料 C に含まれている本アルゴリズムの固定小数点演算記述との間に相違がある場合、TTC 標準 J T - G 7 2 2 . 2 付属資料 C の記述が優先される。

1. 2 定義、略語

1. 2. 1 定義

本付録における定義を以下に記述する。

N 点中央値処理： N 要素を含む中央値処理のセットにおいて、それらの値に応じ昇順に並べ、 $(\text{int}(N/2)+1)$ 番目に大きい値を中央値として選択する。

1. 2. 2 略語

本付録における略語を以下に記述する。

AMR - WB	適応マルチレート広帯域 (Adaptive Multi Rate - WideBand)
AN	アクセス網 (Access Network)
B F H	不正フレーム処理 (Bad Frame Handling)
B F I	AN からの不正フレーム表示 (Bad Frame Indication from AN)
B S I _ _ n e t w	AN インタフェース CRC チェックから得られる不正サブブロック表示 (Bad Sub-block Indication obtained from AN interface CRC checks)
C R C	巡回冗長検査 (Cyclic Redundancy Check)
E C U	誤り補償ユニット (Error Concealment Unit)
m e d i a n N	N 点中央値処理 (N-point median operation)
p r e v B F I	前フレームの不正フレーム表示 (Bad Frame Indication of previous frame)
R X	受信 (Receive)
S C R	ソースコントロールレート (動作) (Source Controlled Rate (operation))
S I D	無音挿入記述子 (背景雑音) (Silence Insertion Descriptor (Background noise))

1. 3 概要

誤り補償手法の目的は、誤りのある、あるいは損失した AMR - WB 音声フレームの影響を補償することである。誤り、あるいは損失が数フレーム発生した場合における出力ミュートイングの目的は、使用者に対しチャネルの切断を示し、また、誤り補償手法の結果として発生する可能性のある不快な音を回避することである。

網は、誤り／損失音声フレーム、あるいは損失 S I D フレームを、RX_TYPE 値 (TTC 標準 J T - G 7 2 2 . 2 付属資料 B) を SPEECH_BAD、SID_BAD あるいは SPEECH_LOST にセットすることで示す必要がある。こ

これらのフラグがセットされたならば、音声復号器は誤りを補償するためにパラメータ置換を行う必要がある。

1. 5 節にて提供される解決例は、完全な音声フレームベースにおける不正フレーム処理においてのみ適用できる。サブフレームベースの誤り補償は、同様な方法を用いて得る事ができるだろう。

1. 4 要求

1. 4. 1 誤り検出

AMR-WB 音声データの高感度ビットが誤りを持って受信された場合、網は、RX_TYPE=SPEECH_BAD (この場合 BFI フラグがセットされる) として示す必要がある。フレームを受信できなかった場合は、網は、RX_TYPE=RX_SPEECH_LOST (この場合も BFI フラグがセットされる) として示す必要がある。SID フレームが誤りを持って受信された場合、網は、RX_TYPE=SID_BAD として示す必要がある。

1. 4. 2 誤りあるいは損失音声フレーム

誤りのある、あるいは損失した音声フレームを通常通りに復号すると、とても不快な雑音が発生することがある。この主観品質を改善するために、誤り/損失音声フレームは、既に受信した正常な音声フレームの繰り返し、あるいは、外挿によって置換される必要がある。この置換によって、出力レベルは次第に減少し、最終的には無音となる。1. 5 節に解決例を示す。

1. 4. 3 最初の損失 SID フレーム

損失した SID フレームは、既に受信した有効な SID フレームからの SID 情報を用いて置換される必要がある。TTC 標準 JT-G 7 2 2. 2 付属資料 B に記述されている有効な SID フレームに対する手法が適用される。

1. 4. 4 後続の損失 SID フレーム

多くの後続する損失 SID フレームに対しては、擬似背景雑音に対してミュート法が適用される必要がある、これにより出力レベルが次第に減少する。後続する損失 SID フレームに対しては、出力のミュート法が維持される必要がある。1. 5 節に解決例を示す。

1. 5 ECU/BFHの解決例

1. 5. 1 ステートマシン

置換およびミュート法に対する解決例は、7 状態の状態遷移に基づいている (Figure 1-1/JT-G722.2)。

このシステムは状態 0 からスタートする。不正フレームが検出されるたびに、状態カウンタが 1 ずつ増加し、6 で飽和される。正常な音声フレームが検出されるたびに、状態カウンタは右に 1 ビットシフトされる。状態カウンタは、チャンネルの品質を示す：状態カウンタ値が大きい方が、チャンネル品質が悪い。状態遷移のコントロールフローは、以下の C コードによって記述することができる (BFI は不正フレーム記述子、State は状態値)。

```
if(BFI != 0)
    State = State + 1;
    if(State > 6)
        State = 6;
else
    State = State >> 1;
```

この状態遷移に加え、前フレームの Bad Frame Flag (prevBFI) がチェックされる。その処理は、State 変数の

値に依存する。状態0と6において、その処理は、BFI フラグに依存する。

状態遷移は、Figure 1 – 1 / JT-G722.2 に要約される。

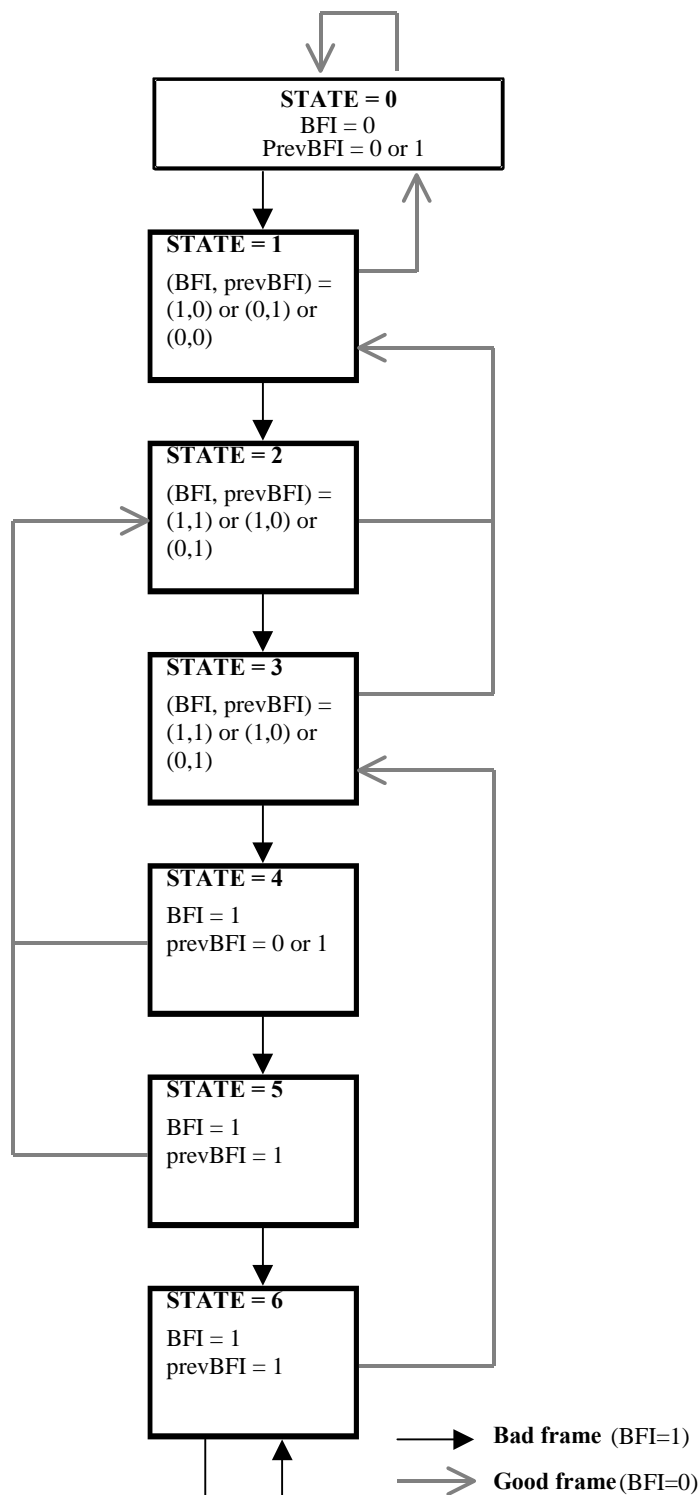


Figure 1 – 1 / JT-G722.2 : State machine for controlling the bad frame substitution (ITU-T G.722.2)

1. 5. 2 誤り／損失音声フレームの置換とミュートイング

1. 5. 2. 1 BFI=0、prevBFI=0、State=0あるいは1

現受信音声フレーム、あるいは前受信音声フレームに誤りが検出されなかった場合。受信した音声パラメー

タは、音声合成にそのまま使用される。現フレームの音声パラメータは保存される。

1. 5. 2. 2 BFI=0、prevBFI=1、State=0から3

現受信音声フレームに誤りが検出されなかったが、前受信音声フレームが不正だった場合。LTP利得は、音声合成にそのまま使用され、固定コードブック利得は、直近に受信した正常なサブフレームの値を用いて以下のように制限される。

$$g^c(n) = \begin{cases} g_{received}^c & , g_{received}^c \leq 100 \text{ or } g_{received}^c \leq g^c(n-1) \times 1.25 \\ 1.25 * g^c(n-1) & , otherwise \end{cases} \quad (1)$$

ここで、

$g_{received}^c$ = 現復号後固定コードブック利得

$g^c(n-1)$ = 最新の正常なサブフレーム (BFI=0) に使用された固定コードブック利得

$g^c(n)$ = 現フレームに対して使用される固定コードブック利得

その他の受信した音声パラメータは、音声合成にそのまま使用される。現フレームの音声パラメータは保存される。

1. 5. 2. 3 BFI=1、prevBFI=0または1、State=1から6

現受信音声フレームに誤りが検出された場合、置換とミュート処理が行われる。

1. 5. 2. 3. 1 RX_FRAME_TYPE=SPEECH_BAD時のLTP利得と固定コードブック利得の補償

LTP利得 g^p と固定コードブック利得 g^c は、前サブフレームから減衰させた値に置換される。

$$g^p = P^p(state) * median5(g^p(n-1), \dots, g^p(n-5)) \quad (2)$$

$$g^c = \begin{cases} P^c(state) * median5(g^c(n-1), \dots, g^c(n-5)) & , VAD_HIST \leq 2 \\ median5(g^c(n-1), \dots, g^c(n-5)) & , VAD_HIST > 2 \end{cases} \quad (3)$$

ここで、

g^p = 現復号後LTP利得

g^c = 現復号後固定コードブック利得

$g^p(n-1), \dots, g^p(n-5)$ = 最新5フレームに使用されたLTP利得

$g^c(n-1), \dots, g^c(n-5)$ = 最新5フレームに使用された固定コードブック利得

$median5()$ = 5点中央値処理

$P^p(state)$ = 減衰係数 ($P^p(1) = 0.98, P^p(2) = 0.96, P^p(3) = 0.75, P^p(4) = 0.23, P^p(5) = 0.05, P^p(6) = 0.01$)

$P^c(state)$ = 減衰係数 ($P^c(1) = 0.98, P^c(2) = 0.98, P^c(3) = 0.98, P^c(4) = 0.98, P^c(5) = 0.98,$

$$P^c(6) = 0.70)$$

state = 状態番号 {0...6}

VAD_HIST は連続した VAD=0 検出の数

状態値が高いほど、より利得が減衰される。予測固定コードブック利得のメモリも、メモリ内における過去の4値平均を用いて更新される。

$$ener(0) = \frac{1}{4} \left[\sum_{i=1}^4 ener(n-i) \right] - 3 \quad (4)$$

1. 5. 2. 3. 2 RX_FRAMETYPE=SPEECH_LOST 時の LTP 利得と固定コードブック利得の補償

LTP 利得 g^p と固定コードブック利得 g^c は、前サブフレームから減衰させた値に置換される。

$$g^p = P^p(state) * median5(g^p(n-1), \dots, g^p(n-5)) \quad (5)$$

$$g^c = \begin{cases} P^c(state) * median5(g^c(n-1), \dots, g^c(n-5)) & , VAD_HIST \leq 2 \\ median5(g^c(n-1), \dots, g^c(n-5)) & , VAD_HIST > 2 \end{cases} \quad (6)$$

ここで、

g^p = 現復号後 LTP 利得

g^c = 現復号後固定コードブック利得

$g^p(n-1), \dots, g^p(n-5)$ = 最新5フレームに使用された LTP 利得

$g^c(n-1), \dots, g^c(n-5)$ = 最新5フレームに使用された固定コードブック利得

median5() = 5点中央値処理

$P^p(state)$ = 減衰係数 ($P^p(1) = 0.95, P^p(2) = 0.90, P^p(3) = 0.75, P^p(4) = 0.23, P^p(5) = 0.05,$
 $P^p(6) = 0.01$)

$P^c(state)$ = 減衰係数 ($P^c(1) = 0.50, P^c(2) = 0.25, P^c(3) = 0.25, P^c(4) = 0.25, P^c(5) = 0.15,$
 $P^c(6) = 0.01$)

state = 状態番号 {0...6}

VAD_HIST は連続した VAD=0 検出の数

状態値が高いほど、より利得が減衰される。予測固定コードブック利得のメモリも、メモリ内における過去の4値平均を用いて更新される。

$$ener(0) = \frac{1}{4} \left[\sum_{i=1}^4 ener(n-i) \right] - 3 \quad (7)$$

1. 5. 2. 3. 3 I S F補償

I S Fは、その部分的適応平均値に向かって漸近する。

$$ISF_q(i) = \alpha * past_ISF_q(i) + (1-\alpha) * ISF_{mean}(i) \quad i=0..16 \quad (8)$$

ここで、

$$\alpha = 0.9$$

$ISF_q(i)$ は現フレームに対する I S Fベクトル

$past_ISF_q(i)$ は前フレームからの I S Fベクトル

$ISF_{mean}(i)$ ベクトルは、I S Fベクトルの適応平均と固定平均から、下式によって得られる。

$$ISF_{mean}(i) = \beta * ISF_{const_mean}(i) + (1-\beta) * ISF_{adaptive_mean}(i) \quad i=0..16 \quad (9)$$

ここで、

$$\beta = 0.75$$

$$ISF_{adaptive_mean}(i) = \frac{1}{3} \sum_{i=0}^2 past_ISF_q(i) \quad (\text{BFI}=0 \text{の時更新される})$$

$ISF_{const_mean}(i)$ は I S Fベクトルの長期平均を含むベクトル

1. 5. 2. 3. 4 L T Pラグ補償

過去5個の正常なL T PラグとL T P利得の履歴は、更新に対する最良の方法を見つけるために用いられる。

1. 5. 2. 3. 4. 1 RX_FRAME TYPE=SPEECH_BAD時のL T Pラグ補償

受信したL T Pラグの有用性は、以下のように定義される（受信したラグ値が、送信した値に近く、これを用いても不快な音を発生させないものであるかどうかを推測する）。

$$Q_{lag} = \begin{cases} 1 & ,T_{dif} < 10 \text{ and } T_{min} - 5 < T_{received} < T_{min} + 5 \\ 1 & ,g^p(n-1) > 0.5 \text{ and } g^p(n-2) > 0.5 \text{ and } T(n-1) - 10 < T_{received} < T(n-1) + 10 \\ 1 & ,g_{min}^p < 0.4 \text{ and } g^p(n-1) = g_{min}^p \text{ and } T_{min} < T_{received} < T_{max} \\ 1 & ,T_{dif} < 70 \text{ and } T_{min} < T_{received} < T_{max} \\ 1 & ,T_{mean} < T_{received} < T_{max} \\ 0 & ,otherwise \end{cases} \quad (10)$$

ここで、

$T(n-1)$ は前正常フレームからのL T Pラグ

$$T_{dif} = |T_{received} - T(n-1)|$$

$$T_{min} = \min(T_{buffer})$$

$$T_{max} = \max(T_{buffer})$$

$T_{received}$ は受信したラグ

$$g_{min}^p = \min(g_{buffer}^p)$$

g^p は現フレームの L T P 利得

$g^p(-1)$ は前正常フレームの L T P 利得

$g^p(-2)$ は前々正常フレームの L T P 利得

$$T_{mean} = average(T_{buffer})$$

現フレーム L T P ラグ値は、以下のように定義される。

$$T = \begin{cases} T_{received} & , Q_{lag} = 1 \\ \frac{1}{3} \sum (T_{max} + T_{max-1} + T_{max-2}) + RND(T_{max} - T_{max-2}) & , Q_{lag} = 0 \end{cases} \quad (11)$$

ここで、

$$T_{max} = \max(T_{buffer})$$

T_{max-1} は T_{buffer} 内の 2 番目に大きい値

T_{max-2} は T_{buffer} 内の 3 番目に大きい値

$RND(x)$ は $\left[-\frac{x}{2}, +\frac{x}{2}\right]$ の範囲にて発生する乱数

1. 5. 2. 3. 4. 2 RX_FRAME TYPE=SPEECH_LOST 時の L T P ラグ補償

最新の正常なフレームからの L T P ラグの有用性は、以下のように定義される (受信したラグ値が、送信した値に近く、これを用いても不快な音を発生させないものであるかどうかを推測する)。

$$Q_{lag_t-1} = \begin{cases} 1 & , g_{min}^p > 0.5 \text{ and } T_{dif} < 10 \\ 1 & , g^p(n-1) > 0.5 \text{ and } g^p(n-2) > 0.5 \\ 0 & , otherwise \end{cases} \quad (12)$$

ここで、

$$g_{min}^p = \min(g_{buffer}^p)$$

$g^p(n-1)$ は前正常フレームの L T P 利得

$g^p(n-2)$ は前々正常フレームの L T P 利得

現フレーム L T P ラグ値は、以下のように定義される。

$$T = \begin{cases} T(n-1) & , Q_{lag_t-1} = 1 \\ \frac{1}{3} \sum (T_{max} + T_{max-1} + T_{max-2}) + RND(T_{max} - T_{max-2}) & , Q_{lag_t-1} = 0 \end{cases} \quad (13)$$

ここで、

$T(n-1)$ は前正常フレームからのLTPラグ

$T_{\max} = \max(T_{\text{buffer}})$

$T_{\max-1}$ は T_{buffer} 内の2番目に大きい値

$T_{\max-2}$ は T_{buffer} 内の3番目に大きい値

$RND(x)$ は $\left[-\frac{x}{2}, +\frac{x}{2}\right]$ の範囲にて発生する乱数

1. 5. 2. 4 励振系列

RX_FRAMETYPE=SPEECH_BAD の時、誤りのあるフレームから受信した固定コードブック励振パルスは、受信したまま使用される。

RX_FRAMETYPE=SPEECH_LOST の時、誤りのあるフレームから受信した固定コードブック励振パルスは使用されず、固定コードブック励振ベクトルは乱数で埋められる（値は $[-1, +1]$ の範囲に制限される）。

1. 5. 2. 5 高域利得 (23.85 kbit/s モード)

RX_FRAMETYPE=SPEECH_BAD または RX_FRAMETYPE=SPEECH_LOST の場合には、受信した高域エネルギーパラメータは使用されず、代わりに高域利得の推定値が用いられる。これは、誤り／損失音声フレームにおいては、高域の生成は全てのモードに対して同様な方法で行われることを意味する。

1. 5. 3 損失SIDフレームの置換とミュートイング

音声復号器において、SID_BAD と分類されたひとつのフレームは、最新の有効なSIDフレーム情報によって置換される必要があり、有効なSIDフレームに対する手法が適用される。SID情報の更新間隔（更新は、SID_UPDATE または場合によってはSID_FIRSTの受信によって示される）が1秒以上の場合、減衰処理が必要となる。

付録
(標準 J T - G 7 2 2 . 2 に対する)
用語対照表

英語	T T C 標準用語
active codec sets	アクティブコーデックセット
adaptive codebook	適応コードブック
adaptive multi-rate wideband	適応マルチレート広帯域
adaptive gain control	適応利得制御
algebraic codebook	代数的コードブック
ambient noise	周囲ノイズ
analysis-by-synthesis	合成による分析
anti-sparseness processing	非スパース化処理
bad speech frame	不正音声フレーム
bandwidth expansion	帯域幅拡張
basic operation	基本演算
channel coding	伝送路符号化
code excited linear prediction	符号励振線形予測
closed-loop	閉ループ
codebook	コードブック
codec	コーデック
comfort noise	擬似背景雑音
cyclic generator polynomial	巡回生成多項式
decoder	復号器
depth-first tree search	縦形木探索
encoder	符号器
error concealment	誤り補償
excitation	励振
fixed codebook	固定コードブック
fixed point	固定小数点
fractional lags	分数ラグ
fractional pitch	分数ピッチ
Frame Quality Indicator	フレーム品質表示
frame structure	フレーム構造
frame type	フレームタイプ
gain	利得
generic frame format	汎用フレームフォーマット
hangover	ハングオーバー
hangover period	ハングオーバー区間
homing	ホームイング
immittance spectrum pairs	イミタンススペクトル対
immittance spectral frequency	イミタンススペクトル周波数

impulse response	インパルス応答
innovation vector	励振ベクトル
interleaved single-pulse permutation	インタリーブ単一パルス順列
inverse filter	逆フィルタ
lag	ラグ
Levinson-Durbin algorithm	レビンソン-ダービンアルゴリズム
linear prediction	線形予測
long term prediction	長期予測
lost speech frame	損失音声フレーム
median	中央値
mode indication	モード表示
moving average	移動平均
most significant bit	最上位ビット
multi-stage vector quantization	多段ベクトル量子化
noise floor	雑音フロア
onset	オンセット
open-loop	開ループ
perceptual weighting	聴覚重み付け
pitch delay	ピッチ遅延
pitch lag	ピッチラグ
prediction error	予測誤差
pseudocode	擬似コード
radio subsystem	無線サブシステム
Real-Time Transport Protocol	リアルタイム伝送プロトコル
residual	残差
sampling frequency	標本化周波数
short term	短期
short term synthesis filter	短期合成フィルタ
sign	極性
silence insertion descriptor	無音挿入記述子
sinc function	sinc 関数
source coding	情報源符号化
source controlled rate operation	ソースコントロールレート動作
spectral distances	スペクトル距離
speech telephony service	音声電話サービス
state machine	ステートマシン
stochastic codebooks	確率的コードブック
subframe	サブフレーム
subjective importance	主観的重要度
tandem free operation	タンデムフリー動作
target signal	ターゲット信号
test sequence	テストシーケンス
track	トラック

traffic channel	通信回線
traffic frame	通信フレーム
transcoder free operation	トランスコーダフリー動作
unvoiced	無声の
vector quantization	ベクトル量子化
voice activity detection	音声検出
voiced	有声の
white noise correction	白色雑音補正
zero input response	零入力応答
zero-state response	零状態応答

付録
(標準 J T - G 7 2 2 . 2 に対する)
用語解説

合成による分析 (A - b - S) (analysis-by-synthesis)

信号生成モデルを用いて合成した音声と実際の信号との誤差を最小とるように、パラメータを変化させて求める分析方法。

不正フレーム (bad frame)

無線伝送やパケット伝送により、音声フレームにビット誤りや損失が発生したフレーム。このようなフレームを復号側で受信した場合には、誤り補償が行われる。

コーデックホーミング (codec homing)

コーデックをある既知の初期状態に導くための特有のリセット手順。コーデックの符号器および復号器に対して予め定義した既知のフレームデータを与えることで、符号器、復号器を既知の初期状態にする。

符号励振線形予測 (C E L P) (code excited linear prediction)

音声信号の短期および長期相関を取り除いた残差信号をあらかじめ作成されたコードブックを用いてベクトル量子化し、このコードブックのインデックスと補助情報を用いて音声を符号化する方式。コードブックの探索は、合成による分析 (A - b - S) 法を用いる。

擬似背景雑音 (comfort noise)

ソースコントロールレート動作において、無音区間中に、送信側の背景音響雑音に類似した合成雑音を受信側で生成することができるように、送信側の背景音響雑音から推定した擬似的な雑音信号。

分数ピッチ遅延 (fractional pitch delay)

適応コードブックパラメータ (またはピッチパラメータ) の 1 つであり、有声音の基本周期を示す。本標準では分析精度を上げるため、分数値を用いている。

レビンソン-ダービンアルゴリズム (Levinson-Durbin algorithm)

対象とする信号の相関係数から再帰的に線形予測係数と反射係数を求める効率的な方法。

線形予測分析 (linear prediction analysis)

現時刻におけるサンプル値を、過去のサンプル値の線形結合として予測する線形予測モデルを用いて、現サンプルの観測値と予測値の誤差が最小になるように線形式の係数を求め、それを分析値とする音声信号の分析方法。

リアルタイム伝送プロトコル (R T P) (Real-Time Transport Protocol)

IP ネットワーク上で音声や映像等の各種リアルタイムメディアをリアルタイムに伝送する方式を定めた技術規約。

ソースコントロールレート動作 (source controlled rate operation)

入力音声信号に対して有音無音判定を行い、有音区間に対してのみ通常の音声符号化を行い、無音区間で

は背景雑音情報を少ない情報量で間欠的に送信することで、効率的な伝送を実現する伝送制御手法。

分割多段ベクトル量子化（S-MSVQ）（split-multistage vector quantization）

ベクトルを複数のサブベクトルに分割して量子化する分割ベクトル量子化と、ベクトルを複数の段階（階層）に分けて量子化する多段ベクトル量子化を組み合わせたベクトル量子化手法。多段量子化の各段階に対して、ベクトルを複数のサブベクトルに分割して各々のサブベクトル単位で量子化する。

音声検出（voice activity detection）

無音圧縮を行うために、入力音声を有音と無音に分類するための判定。この音声検出の結果に基づきソースコントロールレート動作が行われる。