

## JT-G729.1

# JT-G729 ベースのエンベデッド 可変ビットレート符号化: JT-G729 とビット列互換な 8-32kbit/s スケーラブル広帯域符号化

G.729 based Embedded Variable bit-rate coder:  
An 8-32 kbit/s scalable wideband coder bitstream  
interoperable with G.729

第5版

2013年11月14日制定

一般社団法人  
情報通信技術委員会

THE TELECOMMUNICATION TECHNOLOGY COMMITTEE

本書は、一般社団法人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一般社団法人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> .....	8
1. 本標準の規定範囲 .....	10
2. 参照文献 .....	10
3. 略語 .....	10
4. 表記法と規約 .....	12
5. 符号器の概要記述 .....	14
5. 1 符号器 .....	15
5. 2 復号器 .....	16
5. 3 フレーム消失補償 (F E C) .....	17
5. 4 符号化モード .....	17
5. 5 符号化パラメータに対するビット配分およびビット列のレイヤフォーマット .....	18
5. 6 アルゴリズム遅延 .....	20
5. 8 符号器の記述 .....	20
6 符号器の機能記述 .....	20
6. 1 QMF分析フィルタバンク、信号のダウンスケーリングおよび高域のスペクトル折り返し .....	20
6. 2 低域の前処理 .....	22
6. 3 狭帯域エンベデッド符号器 (レイヤ1 およびレイヤ2) .....	22
6. 3. 1 線形予測分析と量子化 .....	22
6. 3. 2 聴覚重み付け .....	22
6. 3. 3 開ループピッチ分析 .....	22
6. 3. 4 インパルス応答の計算 .....	23
6. 3. 5 ターゲット信号の計算 .....	23
6. 3. 6 適応コードブック探索 .....	23
6. 3. 7 8kbit/sにおける固定コードブック — 構造と探索 .....	23
6. 3. 8 8 kbit/sにおける利得の量子化 .....	25
6. 3. 9 12kbit/sにおける固定コードブック — 構造と探索 .....	25
6. 3. 10 12 kbit/s 固定コードブック利得の量子化 .....	28
6. 3. 11 F E Cについての信号分類 .....	29
6. 3. 12 メモリ更新 .....	33
6. 4 高域のための前処理 .....	34
6. 5 TDBWE符号器 (レイヤ3) .....	34
6. 5. 1 時間包絡の計算 .....	34
6. 5. 2 周波数包絡の計算 .....	34
6. 5. 4 F E Cのための位相情報 .....	37
6. 6 TDAC符号器 (レイヤ4から12) .....	38
6. 6. 1 F E Cのためのエネルギー情報 .....	38
6. 6. 2 C E L P差分信号の聴覚重み付け .....	39
6. 6. 3 M D C T .....	40
6. 6. 4 サブバンド分割 .....	41
6. 6. 5 スペクトル包絡計算 .....	41
6. 6. 6 スペクトル包絡符号化 .....	41
6. 6. 7 聴覚重要度によるサブバンド順序付け .....	43

6. 6. 8	分割球状ベクトル量子化のためのビット割当て	43
6. 6. 9	MDC T係数の量子化	44
6. 6. 10	TDACパラメータの多重化	48
7.	復号器の機能説明	48
7. 1	狭帯域エンベデッドCELP復号器（レイヤ1および2）	48
7. 1. 1	LPフィルタパラメータの復号	48
7. 1. 2	パリティビットの計算	48
7. 1. 3	適応コードブックベクトルの復号	48
7. 1. 4	8kbit/s 固定コードブックベクトルの復号	49
7. 1. 5	8kbit/s 適応および固定コードブック利得の復号	49
7. 1. 6	12kbit/s 固定コードブックベクトルの復号	49
7. 1. 7	12kbit/s 固定コードブック利得の復号	49
7. 1. 8	8kbit/s または 12kbit/s における再生音声の計算	49
7. 2	TDBWE復号器（レイヤ3）	49
7. 2. 1	パラメータ復号処理	50
7. 2. 2	励振信号の生成	50
7. 2. 3	時間包絡形成処理	53
7. 2. 4	周波数包絡整形	54
7. 2. 5	適応振幅圧縮による時間領域後処理	55
7. 3	TDAC復号器（レイヤ4から12）	56
7. 3. 1	MDC T正規化係数の復号化	56
7. 3. 2	スペクトル包絡復号化	57
7. 3. 3	聴覚重要度によるサブバンド順序付け	57
7. 3. 4	MDC T係数量子化におけるビット割当て	57
7. 3. 5	MDC T係数の復号化とスペクトル逆正規化	57
7. 3. 6	欠落した高域サブバンドの推定と推定されたサブバンドのレベル調整	59
7. 3. 7	復号された高域バンドの後処理	59
7. 3. 8	スペクトル分離、逆MDC Tとオーバーラップ加算	60
7. 3. 9	聴覚重み付け逆フィルタ	61
7. 3. 10	プリエコー処理	61
7. 4	低域バンドのポストフィルタリング	64
7. 4. 1	適応ポストフィルタ	64
7. 4. 2	高域通過フィルタ	66
7. 5	高域バンドのスペクトル折り返し処理、信号アップスケール処理、QMF合成フィルタバンク	66
7. 6	フレーム損失補償	68
7. 6. 1	減衰係数の決定	68
7. 6. 2	復号器における分類	69
7. 6. 3	消失したスーパーフレームのクラスの決定	71
7. 6. 4	励振信号の周期的な区間の生成	72
7. 6. 5	フレーム消失補償処理におけるピッチ予測	75
7. 6. 6	励振信号の乱数成分の生成	76
7. 6. 7	スペクトル包絡の補償、合成及び更新	76

7. 6. 8	フレーム消失補償処理後の通常処理への復帰	77
7. 6. 9	復号器における長期予測 (LTP) 利得の制限	80
7. 7	ビットレート切り換え	81
7. 7. 1	低域側後処理のクロスフェーディング	81
7. 7. 2	狭帯域から広帯域への切り換え後の高域におけるフェードイン	81
8	伝送パラメータ・インデックスの記述	83
9	JT-G729. 1の符号器のビットイグザクト記述	85
9. 1	シミュレーションソフトウェアの使い方	85
9. 2	シミュレーションソフトウェアの構成	86
付属資料A	パケットフォーマット、能力識別子および能力パラメータ	96
A. 1	参考文献	96
A. 2	JT-G729. 1フレームに対するパケット構成	96
A. 3	TTC標準JT-H245で用いる能力識別子およびパラメータ	96
A. 4	JT-G729との相互接続性	98
付属資料B	JT-G729. 1に対する浮動小数点演算での実装	99
B. 1	本付属資料の規定範囲	99
B. 2	参考文献	99
B. 3	概要	99
B. 4	アルゴリズムの記述	99
B. 5	ANSI Cコード	99
付属資料C	DTX/CNG手法	102
C. 1	本標準の規定範囲	102
C. 2	参考文献	102
C. 3	定義	102
C. 4	略語と頭字語	102
C. 5	表記法	103
C. 6	DTX、SID、CNGの構成要素の説明	106
C. 6. 1	無音圧縮符号器	107
C. 6. 2	無音圧縮復号器	107
C. 6. 3	符号化モード	108
C. 6. 4	ビットアロケーションフォーマット	109
C. 6. 5	アルゴリズム遅延	110
C. 6. 6	計算量、要求される記憶容量、DTXの性能	110
C. 6. 7	符号器	111
C. 7	無音圧縮符号器に関する機能記述	111
C. 7. 1	オプションのVAD	112
C. 7. 2	ハングオーバ	112
C. 7. 3	低域側のDTX	112
C. 7. 4	高域側のDTX	112
C. 7. 5	結合されたDTX	113
C. 7. 6	低域パラメータ推定	114
C. 7. 7	高域パラメータ推定	116
C. 7. 8	低域パラメータ量子化	117

C. 7. 8. 2 低域スペクトル量子化.....	117
C. 8 無音圧縮復号器の機能記述.....	118
C. 8. 1 低域復号とCNG.....	118
C. 8. 1. 2 エネルギー減衰.....	120
C. 8. 1. 3 励振生成.....	120
C. 8. 2 高域の復号とCNG.....	122
C. 8. 3 SIDスーパーフレーム消失補償.....	122
C. 8. 4 ビットレート切り替え.....	123
C. 8. 4. 1 狭帯域から広帯域への切り替えに対する帯域幅フェードイン.....	123
C. 8. 4. 2 広帯域から狭帯域への切り替えに対する帯域幅フェードアウト.....	123
C. 9 メモリの更新.....	124
C. 10 伝送パラメータインデックスの詳細.....	124
C. 11 JT-G729. 1無音圧縮のビットイグザクト詳細.....	124
C. 11. 1 シミュレーションソフトウェアの使用法.....	125
C. 11. 2 シミュレーションソフトウェアの構成.....	125
付属資料D JT-G729. 1付属資料C (DTX/CNG) に対する浮動小数点演算での実装.....	130
D. 1 適用範囲.....	130
D. 2 参考文献.....	130
D. 3 概要.....	130
D. 4 アルゴリズム記述.....	130
D. 5 ANS I-Cコード.....	130
付属資料E 超広帯域スケーラブル拡張.....	133
E. 1 適用範囲.....	133
E. 2 概要.....	133
E. 3 略語.....	133
E. 4 数学的表現.....	134
E. 5 コーデックの概要.....	134
E. 5. 1 入力/出力サンプリングレート.....	135
E. 5. 2 アルゴリズム遅延.....	135
E. 5. 3 演算量とメモリ量.....	135
E. 6 符号化器の機能記述.....	135
E. 6. 1 サンプリング変換.....	136
E. 6. 2 WB符号化.....	137
E. 6. 3 MDCT領域への変換.....	137
E. 6. 4 トーナリティの推定.....	141
E. 6. 5 汎用モードの符号化.....	145
E. 6. 6 正弦波モードの符号化.....	152
E. 6. 7 拡張レイヤ符号化手法の決定.....	154
E. 6. 8 正弦波拡張レイヤの符号化.....	155
E. 6. 9 WB拡張.....	157
E. 7 復号器の機能記述.....	170
E. 7. 1 WB復号.....	170
E. 7. 2 MDCT領域合成WB成分の取得.....	172

E. 7. 3	汎用モード復号レイヤ6m o .....	172
E. 7. 4	正弦波モード復号レイヤ6m o .....	174
E. 7. 5	正弦波改善レイヤ復号 .....	176
E. 7. 6	WB改善レイヤ復号 .....	177
E. 7. 7	フレーム消失補償 .....	178
E. 7. 8	SWB出力とWB出力の切替.....	181
E. 7. 9	時間領域への信号変換 .....	185
E. 7. 10	ミュージックエンハンスメント.....	189
E. 7. 11	後処理.....	196
E. 7. 12	広帯域信号再標本化 .....	197
E. 8	伝送パラメータインデックスの記述.....	198
E. 9	SWB拡張のビットイクザクト記述.....	200
	参考文献.....	201
付録	用語対照表.....	202

## <参考>

### 1. 国際勧告等との関連

本標準は、2006年4月にITU-T SG16でAAPに進むことが合意され、2006年5月に承認されたITU-T勧告G.729.1に準拠したものである。

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月および2007年2月に承認された、ITU-T勧告G.729.1に対するAmendment1およびAmendment2に準拠して改定されたものである。さらに、本標準は、2007年7月にITU-T SG16でAAPに進むことが合意され、2007年8月に承認された、ITU-T勧告G.729.1に対するAmendment3に準拠して改定されたものである。また、本標準は、2008年6月および2008年12月に承認された、ITU-T勧告G.729.1に対するAmendment4およびAmendment5に準拠して改定されたものである。また、本標準は、2009年8月に承認されたITU-T勧告G.729.1に対するCorrigendum1に準拠して改定されたものである。また、本標準は、2010年3月に承認された、ITU-T勧告G.729.1に対するAmendment6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	2007年3月15日	制定
第2版	2007年11月26日	ITU-T勧告G.729.1の改定に伴う、標準本体の改定および付属資料A、Bの追加
第3版	2010年5月26日	ITU-T勧告G.729.1の改定に伴う、標準本体の改定および付属資料C、Dの追加
第4版	2011年11月16日	Cコード改定に関する記述の追加
第5版	2013年11月14日	ITU-T勧告G.729.1の改定に伴う、標準本体の改定および付属資料Eの追加



#### 4. 工業所有権

本標準に関わる「工業所有権の実施の権利に係る確認書」の提出状況は、TTCホームページでご覧になれます。

#### 5. その他

##### (1) 参照している勧告、標準等

TTC標準： JT-G729、JT-G729 付属資料A、JT-G729 付属資料B、JT-G729.2.2、JT-H245、JT-H225.0  
ITU-T勧告： G.191、G.192  
IETF： RFC4729

(2) TTC標準JT-G729.1は、ITU-T勧告G.729.1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、ITU-T勧告G.729.1のものをさし、ITU-TのWebサイトから入手可能である。

なお、2007年1月および2007年8月に承認されたITU-T勧告G.729.1の改定に伴い、標準本体に対するCコードの改定が行われている。

また、2008年6月および2008年12月に承認されたITU-T勧告G.729.1の改定に伴い、標準本体および付属資料Bに対するCコードの改定が行われている。

また、標準本体、付属資料B、CおよびDにて参照しているCコードの改定に関して、2009年8月に承認されたITU-T勧告G.729.1に対するCorrigendum1に記述がある。

#### 6. 標準作成部門

メディア符号化専門委員会

## 1. 本標準の規定範囲

本標準では、狭帯域および広帯域の音声およびオーディオ信号に対し8～32kbit/sでのスケーラブル符号化を実現するためのTTC標準JT-G729の拡張アルゴリズムを記述する。

本標準は以下のように構成される。2章、3章および4章で、本標準で用いられる参考文献、略語および表記法がそれぞれ定義される。5章ではJT-G729.1のアルゴリズムの概要を示す。そして6章および7章において、JT-G729.1の符号器および復号器の動作原理をそれぞれ記述する。8章では伝送パラメータについて説明する。9章では、本標準の符号器を16-32ビットの固定小数点演算で定義したソフトウェアについて記述する。

## 2. 参考文献

下記のTTC標準およびITU-T勧告は、本標準での参照を通して本標準の規定を構成するものである。全ての標準および他の参考文献は、改定に従うものとする。従って、本標準のユーザには、以下のTTC標準やその他の参照すべき文献について、最新の版の適用の可能性を調査するよう奨励される。現在有効なTTC標準およびITU-T勧告のリストは定期的な出版されている。

本標準内での文書の参照は、単独の文書としては、それを標準の扱いとはしない。

### (1) TTC標準JT-G729

8kbit/s CS-ACELP を用いた音声符号化方式

### (2) TTC標準JT-G729 附属資料A

低演算量版 8kbit/s CS-ACELP 音声コーデック

### (3) TTC標準JT-G729 附属資料B

ITU-T 勧告 V.70 端末に適した標準 JT-G729 に対する無音圧縮手法

### (4) ITU-T 勧告 G.191

Software Tools for Speech and Audio Coding Standards

### (5) ITU-T 勧告 G.192

A Common digital parallel interface for speech standardization activities

## 3. 略語

本標準で使われる頭字語を Table1/JT-G729.1 に示す。

Table 1/JT-G729.1 Glossary of acronyms  
(ITU-T G.729.1)

Acronym	Description
ACELP	Algebraic CELP
BWE	BandWidth Extension
CELP	Code-Excited Linear-Prediction
CNG	Comfort Noise Generator
DCME	Digital Circuit Multiplication Equipment
DEMUX	DEMUltipleXer
DTX	Discontinuous Transmission
FEC	Frame Erasure Concealment
FIR	Finite Impulse Response
FFT	Fast Fourier Transform
HB	Higher Band
HPF	High Pass Filter
IIR	Infinite Impulse Response
IP	Internet Protocol
IPBX	Internet Private Branching eXchange
LB	Lower-Band
LP	Linear Prediction
LPF	Low Pass Filter
LSB	Least Significant Bit
LSF	Line Spectrum Frequency
LSP	Line Spectrum Pair
LTP	Long-Term Prediction
MDCT	Modified Discrete Cosine Transform
MSB	Most Significant Bit
MUX	MUltipleXer
PCM	Pulse Coded Modulation
PSTN	Public Switched Telephone Network
QMF	Quadrature Mirror Filterbank
SID	Silence Insertion Descriptor
TDAC	Time Domain Aliasing Cancellation
TDBWE	Time-Domain BandWidth Extension
VoIP	Voice over IP
VQ	Vector Quantization
WB	Wideband
WMOPS	Weighted Million Operations Per Second
xDSL	any type of Digital Subscriber Lines

#### 4. 表記法と規約

J T - G 7 2 9 標準の記述との整合を取るため、J T - G 7 2 9 . 1 で使われる 20ms フレームをスーパーフレームと呼び、C E L P 符号化処理に関わる 10ms フレームおよび 5ms サブフレームは、それぞれフレームおよびサブフレームと呼ぶ。

本標準全般にわたり、J T - G 7 2 9 の表記法が利用される。表記に関する規約の詳細は以下の通り。

- (1) コードブックは、カリグラフ文字で記述する(例  $\mathcal{C}$ )。
- (2) 時間領域の信号は、そのシンボルと丸括弧で括られたサンプル番号で記述する(例  $s(n)$ )。変数  $n$  は、サンプル番号である。
- (3) 周波数領域に変換された信号は、対応する時間領域の信号を大文字に変えることにより記述する(例  $S(k)$  は  $s(n)$  の変換)。変数  $k$  は、係数の番号である。
- (4) 丸括弧で括られた上付きの添字は、時間に依存する変数に用いる(例  $g^{(m)}$ )。変数  $m$  はその前後関係よりフレーム番号、あるいはサブフレーム番号に対応している。
- (5) 再帰を示す添字は、角括弧で括られた上付きで記述する(例  $E^{[k]}$ )。
- (6) 下付きの添字は、係数配列の各要素を示す。
- (7) 記号  $\hat{\cdot}$  は量子化されたパラメータを示す(例  $\hat{g}_c$ )。
- (8) パラメータの範囲は、角括弧で括られた値で記述する。この値は境界値を含む(例  $[0.6, 0.9]$ )。
- (9) 関数  $\text{int}()$  は、切り捨てによる整数値への変換を示す。
- (10) 関数  $\text{even}()$  は、引き数が偶数の整数値の場合 1 を、そうでない場合 0 を返す。
- (11) 関数  $\text{round}()$  は、最も近い整数値への丸めを示す。
- (12) 使用される 10 進の浮動小数点値は、16 ビット固定小数点 A N S I C での実現に使用された値を丸めたものである。

本標準全般にわたる主要なシンボルを、Table2/JT-G729.1 に示す。

Table2/JT-G729.1 Glossary of most relevant symbols  
(ITU-T G.729.1)

Type	Name	Description
Filters	$1/\hat{A}(z)$	Quantized LP synthesis filter
	$H_{h1}(z)$	Lower-band high-pass filter
	$H_p(z)$	Long-term postfilter
	$H_f(z)$	Short-term postfilter
	$H_t(z)$	Tilt-compensation filter
	$H_{h2}(z)$	Higher-band low-pass filter
	$H_1(z)$	QMF low pass analysis filter
	$H_2(z)$	QMF high pass analysis filter
	$G_1(z)$	QMF low pass synthesis filter
	$G_2(z)$	QMF high pass synthesis filter
	$P(z)$	Pre-filter for fixed codebook in JT-G729
	$W(z)$	Weighting filter in embedded CELP encoder
	$W_{LB}(z)$	Lower-band difference weighting filter
Signals	$\mathcal{C}(n)$	12 kbit/s layer fixed-codebook codevector
	$s_{WB}(n)$	Wideband input signal

Type	Name	Description
	$s_{LB}^{qmf}(n)$	QMF low pass analysis filter output signal after decimation
	$s_{LB}(n) = s(n)$	Pre-processed lower band signal
	$\hat{s}_{enh}(n)$	Local synthesis of the CELP encoder at 12 kbit/s
	$d_{LB}(n)$	Lower band difference signal between $s(n)$ and $\hat{s}_{enh}(n)$
	$d_{LB}^w(n)$	Lower-band difference weighting filter output signal
	$s_{HB}^{fold}(n)$	QMF analysis filter output signal after decimation and spectral folding
	$s_{HB}(n)$	Pre-processed higher band signal
	$\hat{s}(n)$	8 kbit/s CELP decoded signal
	$\hat{s}_{enh}(n)$	12 kbit/s CELP decoded signal
	$\hat{s}_{LB}(n)$	Lower-band decoded signal before postfiltering
	$\hat{s}_{LB}^{post}(n)$	Postfiltered lower-band decoder signal
	$\hat{s}_{LB}^{hpf}(n)$	High-pass filtered postfiltered lower band reconstructed signal
	$\hat{s}_{LB}^{qmf}(n)$	Decoded lower-band signal (input to QMF synthesis filterbank)
	$\hat{s}_{HB}^{qmf}(n)$	Decoded higher-band signal (input to QMF synthesis filterbank)
	$\hat{s}_{HB}^{bwe}(n)$	TDBWE decoded higher band signal
	$\hat{s}_{HB}^{qmf}(n)$	Decoded higher-band signal (input to QMF synthesis filterbank)
	$\hat{s}_{HB}(n)$	Decoded higher band signal
	$\hat{d}_{LB}^w(n)$	Decoded lower-band weighted difference signal
	$\hat{d}_{LB}(n)$	Decoded lower-band difference signal
	$\hat{s}_{HB}^{fold}(n)$	Decoded folded higher band signal
	$D_{LB}^w(k)$	MDCT spectrum of $d_{LB}^w(n)$
	$S_{HB}(k)$	MDCT spectrum of $s_{HB}(n)$
	$\hat{D}_{LB}^w(k)$	MDCT spectrum of $\hat{d}_{LB}^w(n)$
	$\hat{S}_{HB}^{bwe}(k)$	MDCT spectrum of $\hat{s}_{HB}^{bwe}(n)$

Type	Name	Description
	$\hat{S}_{HB}(k)$	MDCT spectrum of $\hat{S}_{HB}(n)$
Parameters	$\hat{a}_i$	Quantized lower-band LP coefficients
	$\alpha_{enh}$	Tri-pulse parameter
	$\alpha$	FEC attenuation factor
	$\tau$	Position of the last glottal pulse

## 5. 符号器の概要記述

J T-G 7 2 9. 1 符号器は、J T-G 7 2 9 標準に対する 8~32 kbit/s のスケーラブル広帯域(50-7000Hz) 拡張である。デフォルトでは、符号器への入力と復号器の出力は 16000Hz で標本化された信号である。符号器により生成されるビット列はスケーラブルで 12 のエンベデッドレイヤからなり、それらはレイヤ 1 からレイヤ 12 と呼ぶ。レイヤ 1 は、8 kbit/s のビットレートに対応するコアレイヤである。このレイヤは J T-G 7 2 9 のビット列に準拠しており、これにより J T-G 7 2 9. 1 を J T-G 7 2 9 と相互接続可能にする。レイヤ 2 は、4 kbit/s を付加する狭帯域のエンハンスメントレイヤであり、一方、レイヤ 3 から レイヤ 12 は 2 kbit/s ステップで 20 kbit/s を付加する広帯域のエンハンスメントレイヤである。

本符号器は、符号器への入力として 16000Hz で標本化され 16 ビット線形 P C M に変換されたデジタル信号に対して処理を行うように設計されている。しかしながら、8000 Hz の入力標本化周波数もサポートする。同様に、復号器出力のフォーマットは、標本化周波数が 8000 Hz または 16000 Hz の 16 ビット線形 P C M である。他の入力/出力形式の場合は、符号化に先立ち標本化周波数が 8000 Hz または 16000 Hz の 16 ビット線形 P C M に変換する、あるいは、復号後に 16 ビット線形 P C M から適切なフォーマットに変換されるべきである。符号器から復号器へのビット列は本標準で定義される。

J T-G 7 2 9. 1 符号器は、エンベデッド符号帳駆動線形予測 (Code-Excited Linear-Prediction (C E L P)) 符号化、時間領域帯域拡張 (Time-Domain Bandwidth Extension (T D B W E)) 符号化および時間領域折り返し歪打消し (Time-Domain Aliasing Cancellation (T D A C)) と呼ばれる予測変換符号化の 3 段階の構成で構築されている。階層化された C E L P では、8 kbit/s および 12 kbit/s で狭帯域 (50-4000 Hz) の合成信号を生成する レイヤ 1 およびレイヤ 2 のビット列を生成する。T D B W E では、レイヤ 3 のビット列を生成し、それにより 14 kbit/s で広帯域 (50-7000 Hz) 出力信号を生成する。T D A C では、変形離散コサイン変換 (Modified Discrete Cosine Transform (MDCT)) の領域で動作し、レイヤ 4 からレイヤ 12 のビット列を生成し、14 ~ 32 kbit/s のビットレートで音声品質を改善する。T D A C 符号化では、50-4000 Hz 帯域の重みつき C E L P 符号化誤差信号と 4000-7000 Hz 帯域の入力信号を対象に符号化を行う。

J T-G 7 2 9. 1 符号器は 20ms フレームで動作する。しかし、エンベデッド C E L P では、J T-G 7 2 9 と同様 10ms フレームで動作する。従って、20ms フレーム当たり 2 フレーム分の 10ms の C E L P フレームが処理される。以降、J T-G 7 2 9 標準の記述との整合を取るため、J T-G 7 2 9. 1 で使われる 20ms フレームをスーパーフレームと呼び、C E L P 符号化処理に関わる 10ms フレームおよび 5ms サブフレームは、それぞれフレームおよびサブフレームと呼ぶ。

5. 1 節および 5. 2 節では、符号部および復号部の概要記述を示す。また、5. 3 節ではフレーム消失補償に関して記述する。J T-G 7 2 9. 1 は非常に柔軟性のある構造であり、復号側のみならず符号化側で複数のモードが許容される。異なるモードに関しては 5. 4 節にて記述される。ビット配分、遅延および演算量については 5. 5 節および 5. 6 節にて示される。

## 5. 1 符号器

符号化側の機能ブロック図を Figure1/JT-G729.1 に示す。符号器は、20ms のスーパーフレーム入力を単位に動作する。デフォルトでは、入力信号  $s_{WB}(n)$  は 16000Hz で標本化されたものである。従って、入力のスーパーフレームのサイズは 320 サンプルである。

入力信号  $s_{WB}(n)$  は、まず、フィルタ  $H_1(z)$  および  $H_2(z)$  で定義される QMF フィルタバンクを用いて 2 つのサブバンドに分割される。間引き後に得られる低域の入力信号  $s_{LB}^{qmf}(n)$  は遮断周波数 50Hz の高域通過フィルタ  $H_{h1}(z)$  による前処理が行われる。その結果得られる信号  $s_{LB}(n)$  は 8-12 kbit/s の狭帯域エンベデッド CELP 符号器により符号化される。JT-G729 標準との整合を取るため、信号  $s_{LB}(n)$  は  $s(n)$  とも表記される。信号  $s(n)$  と 12kbit/s での CELP 符号器の局部合成信号  $\hat{s}_{enh}(n)$  との差信号  $d_{LB}(n)$  に対して聴覚重みフィルタ  $W_{LB}(z)$  が施される。 $W_{LB}(z)$  のパラメータは、CELP 符号器の量子化された LP (線形予測) 係数から求められる。更に、フィルタ  $W_{LB}(z)$  には、 $W_{LB}(z)$  の出力  $d_{LB}^w(n)$  と高域の入力信号  $s_{HB}(n)$  とのスペクトルの連続性を保証するための利得補償を含む。そして、重み付けされた差信号  $d_{LB}^w(n)$  は MDCT により周波数領域へ変換される。

間引きされ、 $(-1)^n$  によって折り返された高域の入力信号  $s_{HB}^{fold}(n)$  に対して、遮断周波数 3000Hz の低域通過フィルタ  $H_{h2}(z)$  による前処理が行われる。その結果得られる信号  $s_{HB}(n)$  は、TDBWE 符号器により符号化される。信号  $s_{HB}(n)$  もまた MDCT により周波数領域へ変換される。

2 セットの MDCT 係数  $D_{LB}^w(k)$  および  $S_{HB}(k)$  は、最後に TDAC 符号器により符号化される。

また、いくつかのパラメータは、ビット列にパラメータレベルの冗長性を持たせるために、フレーム消失補償 (FEC) 符号器により伝送される。この冗長性により、スーパーフレームの消失が生じた際の音声品質が改善される。

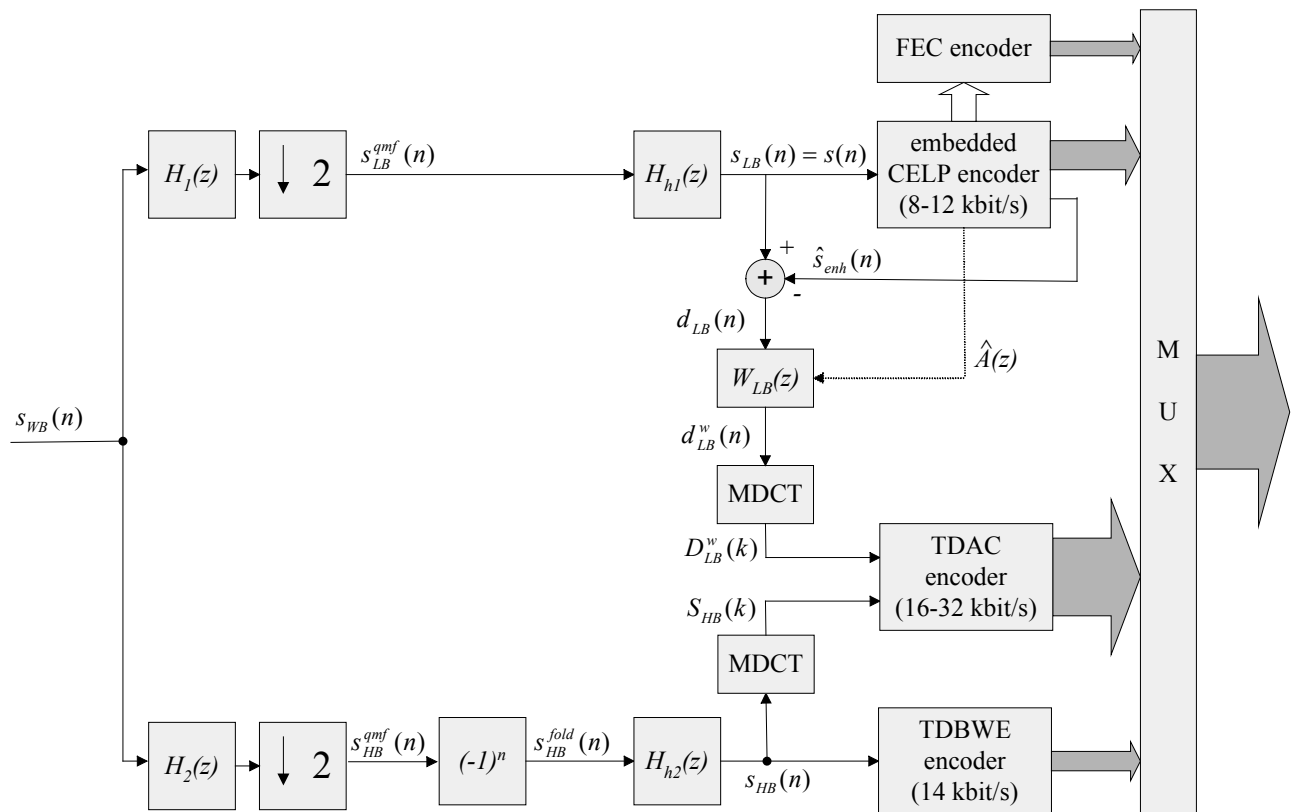


Figure1/JT-G729.1 High-level block diagram of the encoder (ITU-T G.729.1)

## 5. 2 復号器

復号側の機能ブロック図を Figure2/JT-G729.1 に示す。フレーム消失補償の動作はこの図では考慮されておらず、それは 5. 3 節で示される。復号処理は、実際に受信されたレイヤの数、すなわち受信されたビットレートに依存する。各受信ビットレート別の動作を以下に示す。

- 8 kbit/s (レイヤ 1): コアレイヤがエンベデッド C E L P 復号器により復号され、 $\hat{s}_{LB}(n) = \hat{s}(n)$  が得られる。そして、 $\hat{s}_{LB}(n)$  に対してポストフィルタ処理が行われ、得られた信号  $\hat{s}_{LB}^{post}(n)$  に対して高域通過フィルタ (H P F) 処理が行われ、 $\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{hpf}(n)$  が得られる。フィルタ  $G_1(z)$  および  $G_2(z)$  で定義される QMF 合成フィルタバンクにより、高域側周波数の合成信号  $\hat{s}_{HB}^{qmf}(n)$  が零にセットされた出力信号が生成される。
- 12 kbit/s (レイヤ 1 および 2): コアレイヤと狭帯域エンハンスメントレイヤがエンベデッド C E L P 復号器により復号されて  $\hat{s}_{LB}(n) = \hat{s}_{enh}(n)$  が得られ、 $\hat{s}_{LB}(n)$  に対してポストフィルタ処理が行われ、得られた信号  $\hat{s}_{LB}^{post}(n)$  に対して高域通過フィルタ (H P F) 処理が行われ、 $\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{hpf}(n)$  が得られる。QMF 合成フィルタバンクにより、高域側周波数の合成信号  $\hat{s}_{HB}^{qmf}(n)$  が零にセットされた出力信号が生成される。
- 14 kbit/s (レイヤ 1 から 3): 狭帯域の C E L P 復号および低域の適応ポストフィルタ処理に加えて、T D B W E 復号器が、高域の合成信号  $\hat{s}_{HB}^{bwe}(n)$  を生成し、M D C T により周波数領域に変換し、その変換された  $\hat{S}_{HB}^{bwe}(k)$  の 3000Hz 以上の周波数帯域の成分を零にする。そして得られたスペクトル  $\hat{S}_{HB}(k)$  は、逆 M D C T により時間領域に変換され、重ね合わせ加算後に  $(-1)^n$  によりスペクトル折り返し処理がなされる。QMF 合成フィルタバンクにおいて、生成された高域信号  $\hat{s}_{HB}^{qmf}(n)$  は高域通過フィルタ処理を行わずに 12 kbit/s で生成された対応する低域信号  $\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{post}(n)$  と結合される。
- 14 kbit/s を超えるビットレート (レイヤ 1 から 4(またはそれ以上)): 狭帯域の C E L P 復号および T D B W E 復号に加えて、T D A C 復号器が、低域 (0-4000 Hz) の復号重み付き差信号および高域 (4000-7000 Hz) の復号信号に対応する M D C T 係数  $\hat{D}_{LB}^w(k)$  および  $\hat{S}_{HB}(k)$  を生成する。ここで、高域では、非受信のサブバンドおよび割り当てが 0 ビットのサブバンドは、レベル調整後の  $\hat{S}_{HB}^{bwe}(k)$  のサブバンドに置き換えられる。 $\hat{D}_{LB}^w(k)$  および  $\hat{S}_{HB}(k)$  は共に逆 M D C T により時間領域に変換され重ね合わせ加算が行われる。そして、低域信号  $\hat{d}_{LB}^w(n)$  は聴覚重み付け逆フィルタ  $W_{LB}(z)^{-1}$  による処理が行われる。変換符号化による異音を抑圧するために、低域信号  $\hat{d}_{LB}(n)$  および高域信号  $\hat{s}_{HB}(n)$  共に、プリ/ポストエコーを検出し低減する。低域合成信号  $\hat{s}_{LB}(n)$  はポストフィルタ処理が行われ、一方で、高域合成信号  $\hat{s}_{HB}^{fold}(n)$  は、 $(-1)^n$  によりスペクトル折り返し畳み処理がなされる。信号  $\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{post}(n)$  および  $\hat{s}_{HB}^{qmf}(n)$  は、QMF 合成フィルタバンクにおいて、結合されアップサンプリングされる。



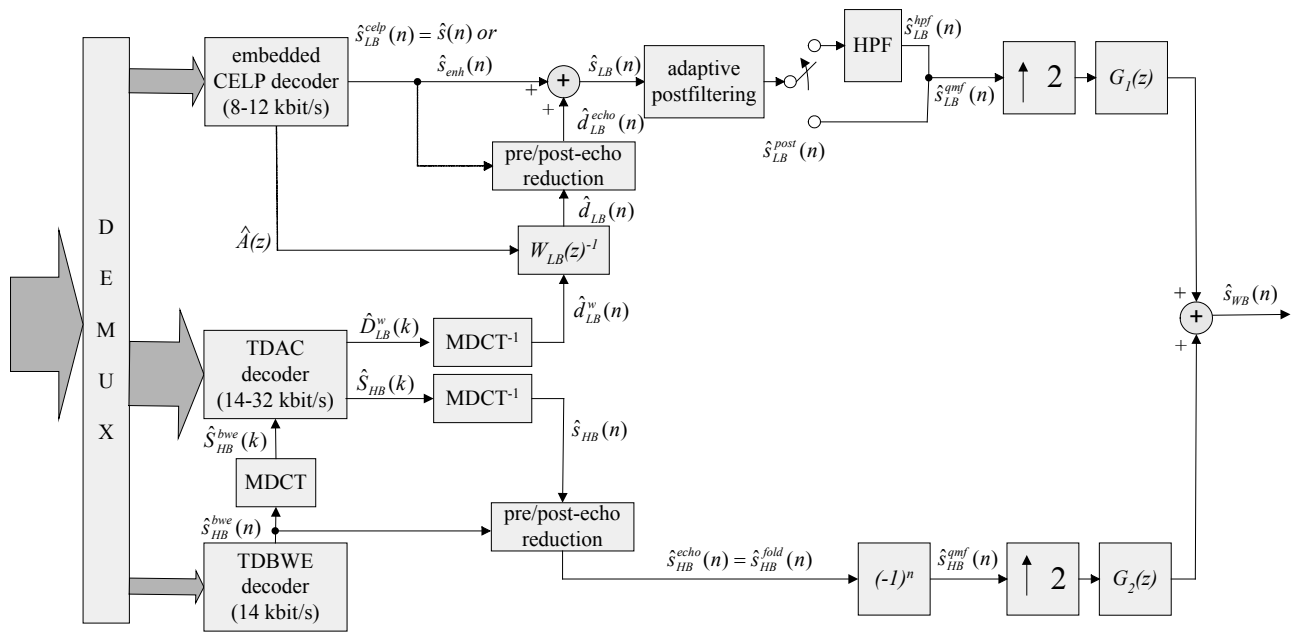


Figure2 / JT-G729.1 High-level block diagram of the decoder (ITU-T G.729.1)

### 5. 3 フレーム消失補償 (FEC)

消失フレーム後の収束・回復およびフレーム消失補償の性能の向上を図るために、少数の補助的な補償／回復用パラメータを決定し伝送する。この符号化側の冗長性により、特に、消失補償後に通常の処理に回復した際の、復号信号の本来の信号への収束性の改善や符号器と復号器の間でのミスマッチの影響の軽減により、復号器でのフレーム消失補償と回復が、著しく改善される。

補償／回復用パラメータには信号分類情報、エネルギー情報、位相情報（前スーパーフレームでの最後の声門パルスの推定位置）を含む。これらのFECパラメータは低域の入力信号のみを用いて推定される。分類情報には2ビット必要であり、これはレイヤ2にて伝送される。位相情報は7ビットで量子化され、レイヤ3にて伝送される。エネルギー情報は5ビットで量子化され、レイヤ4にて伝送される。分類、位相およびエネルギーの各情報の算出は、6. 3. 1 1 節、6. 5. 4 節および6. 6. 1 節でそれぞれ記述される。

ここで、レイヤ1（コアレイヤ）のみが受信された場合においては、分類情報は利用できない。そのような場合には、信号分類情報は復号器側で推定される（7. 6. 1 節参照）。

復号器側でのフレーム消失補償は、伝送された補償／回復用パラメータを用い、また復号器側での余分なスーパーフレーム遅延を利用して行われる。声門パルス再同期、エネルギー制御および擬似立ち上がり再生を含む効率的な補償および回復技術が用いられる。これらは7. 6 節で詳述される。

### 5. 4 符号化モード

JT-G729.1 符号器は、本来非常に柔軟な構造である帯域分割符号化手法に基づいている。本符号化は、QMF分析・合成フィルタバンクの利点を活かすことで、標準化周波数が16000Hzのみならず8000Hzの入力・出力信号を扱うことが可能である。Table3/JT-G729.1に、JT-G729.1で動作可能なモードの一覧を示す。DEFAULTモードは、JT-G729.1のデフォルトの動作モードに対応し、その場合は、入出力は標準化周波数16000Hzの信号である。

2つの付加的な符号化モードがある。

- NB\_INPUTモードは、8000Hzで標準化された信号を符号化入力とする仕様であり、この場合は、QMF分析フィルタバンクを迂回する。
- G729\_BSTモードは、8kbit/sで符号器を動作させ10msフレームのJT-G729フォーマットでのピ

ット列を生成する。符号化への入力はデフォルトでは標準化周波数 16000Hz の信号である。もし、NB\_INPUT モードも同時に設定された場合、入力は標準化周波数 8000Hz の信号となる。

一方、復号器には3つのモードがある。

- NB\_OUTPUT モードは、8000Hz で標準化された信号を復号出力とする仕様であり、この場合は、QMF 合成フィルタバンクを迂回する。
- G729B\_BST モードでは、復号器は J T - G 7 2 9 付属資料Bの符号化フレームを読み込み復号することができる。
- LOW\_DELAY モードは狭帯域での使用に適用される。この場合、復号のビットレートは 8-14 kbit/s に制限され、逆MDCTおよび重ね合わせ加算処理をスキップすることにより全体のアルゴリズム遅延を低減することができる。

G729B\_BST または LOW\_DELAY モードでは、復号器の出力はデフォルトでは標準化周波数 16000Hz の信号である。もし、NB\_OUTPUT モードも同時に設定された場合、出力は標準化周波数 8000Hz の信号となる。

なお、低遅延モードでは、14 kbit/s を超えるビットレートへのスケーラビリティが使用できないため、広帯域の音声品質に制限が生じる。更に、誤りなし条件において、14 kbit/s での低遅延モードでは、14 kbit/s のデフォルトモードと比較してわずかな品質劣化が生じ得る。

Table3/JT-G729.1 Encoder/decoder modes  
(ITU-T G.729.1)

Mode	Encoder operation	Decoder operation
DEFAULT	16000 Hz input	16000 Hz output
NB_INPUT	8000 Hz input	N/A
G729_BST	bit rate limited to 8 kbit/s, output JT-G729 bitstream	N/A
NB_OUTPUT	N/A	8000 Hz output
G729B_BST	N/A	read and decode G729B bitstream
LOW_DELAY	N/A	bit rate limited to 8-14 kbit/s, low delay

### 5. 5 符号化パラメータに対するビット配分およびビット列のレイヤフォーマット

符号器のビット配分を Table 4/JT-G729.1 に示す。本表は各々の異なるレイヤに対応するように構成されている。所定のビットレートに対して、対応するレイヤのビット列を連結することによりビット列が得られる。例えば、24 kbit/s のビットレート（これはスーパーフレーム当たり 480 ビットに相当する）においては、ビット列は、レイヤ 1 (160 ビット) + レイヤ 2 (80 ビット) + レイヤ 3 (40 ビット) + レイヤ 4 から 8 (200 ビット) で構成される。

J T - G 7 2 9 . 1 のビット列フォーマットを Figure3/JT-G729.1 に示す。

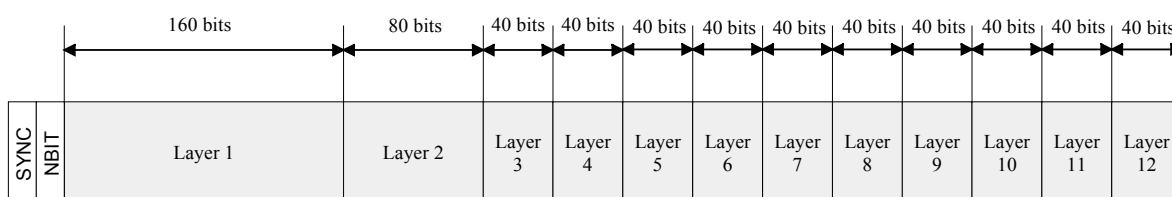


Figure3/JT-G729.1 JT-G729.1 bitstream format (compliant with G.192).  
(ITU-T G.729.1)

TDAC符号器はスペクトル包絡エントロピー符号化および適応サブバンドビット配分を行うため、TDACパラメータは可変ビット数で符号化される。しかしながら、14 kbit/s を超えるビット列は2 kbit/s のレイヤにフォーマットすることができる。これは、TDAC符号器は常に最大符号化ビットレート (32 kbit/s) を基本にビット割り当てが行われ、TDAC復号器は任意の位置でのビット列の打ち切りが可能だからである。

Table4/JT-G729.1 Bit allocation (per 20 ms superframe).  
(ITU-T G.729.1)

Parameter	Codeword	Number of bits				Total per superframe
<b>Layer 1 - core layer (narrowband embedded CELP)</b>						
		10 ms frame 1		10 ms frame 2		
Line spectrum pairs	$L0, L1, L2, L3$	18		18		36
		subframe1	subframe2	subframe1	subframe2	
Adaptive-codebook delay	$P1, P2$	8	5	8	5	26
Pitch-delay parity	$P0$	1		1		2
Fixed-codebook index	$C1, C2$	13	13	13	13	52
Fixed-codebook sign	$S1, S2$	4	4	4	4	16
Codebook gains (stage 1)	$GA1, GA2$	3	3	3	3	12
Codebook gains (stage 2)	$GB1, GB2$	4	4	4	4	16
<b>8 kbit/s core total</b>						<b>160</b>
<b>Layer 2 – narrowband enhancement layer (embedded CELP)</b>						
2 <sup>nd</sup> Fixed-codebook index	$C1, C2$	13	13	13	13	52
2 <sup>nd</sup> Fixed-codebook sign	$S1, S2$	4	4	4	4	16
2 <sup>nd</sup> Fixed-codebook gain	$G1, G2$	3	2	3	2	10
FEC bits (class information)	$CL1, CL2$		1		1	2
<b>12 kbit/s layer total</b>						<b>80</b>
<b>Layer 3 - wideband enhancement layer ( TDBWE)</b>						
Time envelope mean	$MU$	5				5
Time envelope split VQ	$T1, T2$	7+7				14
Frequency envelope split VQ	$F1, F2, F3$	5+5+4				14
FEC bits (phase information)	$PH$	7				7
<b>14 kbit/s layer total</b>						<b>40</b>
<b>Layers 4 to 12 – wideband enhancement layers ( TDAC)</b>						
FEC bits (energy information)	$E$	5				5
MDCT norm	$N$	4				4
HB spectral envelope	$RMS2$	variable number $nbits\_HB$				$nbits\_HB$
LB spectral envelope	$RMS1$	variable number $nbits\_LB$				$nbits\_LB$
fine structure (VQ of subbands coefficients)	$VQ1$ to $VQ18$	$nbits\_VQ = 351 - nbits\_HB - nbits\_LB$				$nbits\_VQ$
<b>16-32 kbit/s layer total</b>						<b>360</b>
<b>TOTAL</b>						<b>640</b>

## 5. 6 アルゴリズム遅延

デフォルトの動作 (DEFAULT モード) では、J T-G 7 2 9. 1 符号器のアルゴリズム遅延は 48.9375 ms (標本化周波数 16000Hz で 783 サンプル) である。遅延の寄与を以下に示す。

- 入力スーパーフレームの 20ms
- M D C T 分析 (先読み) のための 20ms
- 狭帯域 L P C 分析 (先読み) のための 5ms
- Q M F 分析-合成フィルタバンクのための 3.9375 ms

ここで、符号器が NB\_INPUT モードで、復号器が NB\_OUTPUT かつ LOW\_DELAY モードの場合は、アルゴリズム遅延は 25ms に低減する。また、符号器が DEFAULT モードで、復号器が 14kbit/s での LOW\_DELAY モードの場合は、アルゴリズム遅延は 28.9375ms に低減する。

## 5. 7 演算量および所要記憶容量

J T-G 7 2 9. 1 符号器 (符号器+復号器) の観測された最悪値の演算量は 35.15WMOPS である。これは、I T U-T ソフトウェアツールライブラリ STL2005 v2.1 (G.191 に基づく) の基本演算子に基づくものである。J T-G 7 2 9. 1 の 16-bit kword 単位での所要記憶容量を Table5/JT-G729.1 に示す。

Table5/JT-G729.1 Complexity figures of the JT-G729.1 coder (encoder/decoder).  
(ITU-T G.729.1)

Computational complexity (WMOPS)	35.15
Static RAM (kwords)	4.2
Scratch RAM (kwords)	4.6
Data ROM (kwords)	8.3
Program ROM (basic ops+function calls)	8325

## 5. 8 符号器の記述

本標準の符号化アルゴリズムは、ビットイグザクトな固定小数点算術演算で記述されている。9章で示されるANSI Cコードは、本標準の必須部分を構成するものであり、このビットイグザクトな固定小数点での記述を反映している。符号器 (6章) および復号器 (7章) の算術的な記述は、他の方法でも実装し得るが、本標準に準拠しないコーデックを実装することになってしまう可能性がある。

したがって、不一致が生じた場合には、6章および7章の算術的な記述よりも、9章のANSI Cコードによるアルゴリズム記述の方が優先される。ANSI Cコードと共に用いられるテスト信号の非網羅的なセットは、ITU-TのWebサイトから入手可能である。

## 6 符号器の機能記述

### 6. 1 Q M F 分析フィルタバンク、信号のダウンスケーリングおよび高域のスペクトル折り返し

標本化周波数 16000Hz の入力信号  $s_{WB}(n)$  を標本化周波数 8000 Hz の 2 つの信号、低域信号  $s_{LB}^{qmf}(n)$  および高域信号  $s_{HB}^{qmf}(n)$  に分割するために、Figure 1/JT-G729.1 に示すように Q M F 分析フィルタバンクが用いられる。使用される Q M F フィルタバンクにおいては、低域通過フィルタ  $H_1(z)$  は 64 個の係数を有する対称型の F I R フィルタである。高域通過フィルタ  $H_2(z)$  は、 $H_1(z)$  の係数を用いて下式で表される。

$$h_2(n) = (-1)^n h_1(n) \quad (1)$$

フィルタ  $H_1(z)$  との周波数応答は Figure4/JT-G729.1 に示される。

フィルタは対称であるため、係数の半分のみが記憶される。 $h_1(j)$ ,  $j=0,\dots,31$ , をフィルタ係数の後半の半分の係数とする（全体 ( $j=-32$  から  $j=31$  まで) のフィルタは、 $h_1(j)=h_1(-j-1)$ ,  $j=0,\dots,31$  により展開される)。標本化周波数 16000Hz の低域の信号は下式で表される。

$$y_1(n) = \sum_{j=0}^{31} h_1(j) [s_{WB}(n+1+j) + s_{WB}(n-j)] \quad (2)$$

式(2)と同様に、標本化周波数 16000Hz の高域信号は下式で表される。

$$y_2(n) = \sum_{j=0}^{31} h_2(j) [s_{WB}(n+1+j) + s_{WB}(n-j)] \quad (3)$$

式(1)の関係および  $h_2(j)$  は反対称 ( $h_2(j) = -h_2(-j-1)$ ) であることから、式(3)の高域信号は下記のように表せる。

$$y_2(n) = \sum_{j=0}^{31} h_1(j) (-1)^j [s_{WB}(n+1+j) + s_{WB}(n-j)] \quad (4)$$

標本化周波数 8000Hz の低域側および高域側の帯域信号は単に 2 サンプルから 1 サンプルを落とす、すなわち、

$$s_{LB}^{qmf}(n/2) = y_1(n) \quad \text{および} \quad s_{HB}^{qmf}(n/2) = y_2(n)$$

とすることで得られる。

したがって、式(2)および式(4)において、 $s_{LB}^{qmf}(n/2)$  および  $s_{HB}^{qmf}(n/2)$  を直接得るためには、信号  $y_1(n)$  および  $y_2(n)$  は、 $n=0,2,4,\dots$  の値においてのみ計算すればよい。

なお、演算量削減のため、式(2)と式(4)の間の対称性を利用して、 $y_1(n)$  および  $y_2(n)$  を同時に算出することができる。これは、係数の偶数次と奇数次での値に分けて計算することで実現される。まず、中間信号  $A(n)$  および  $B(n)$  を下記のように求める。

$$A(n) = \sum_{j=0}^{15} (h_1(2j)s_{WB}(n+1+2j) + h_1(2j+1)s_{WB}(n-2j-1)) \quad (5)$$

$$B(n) = \sum_{j=0}^{15} (h_1(2j)s_{WB}(n-2j) + h_1(2j+1)s_{WB}(n+1+2j+1)) \quad (6)$$

そして、信号  $y_1(n)$  および  $y_2(n)$  を以下のように算出する。

$$y_1(n) = A(n) + B(n)$$

$$y_2(n) = A(n) - B(n)$$

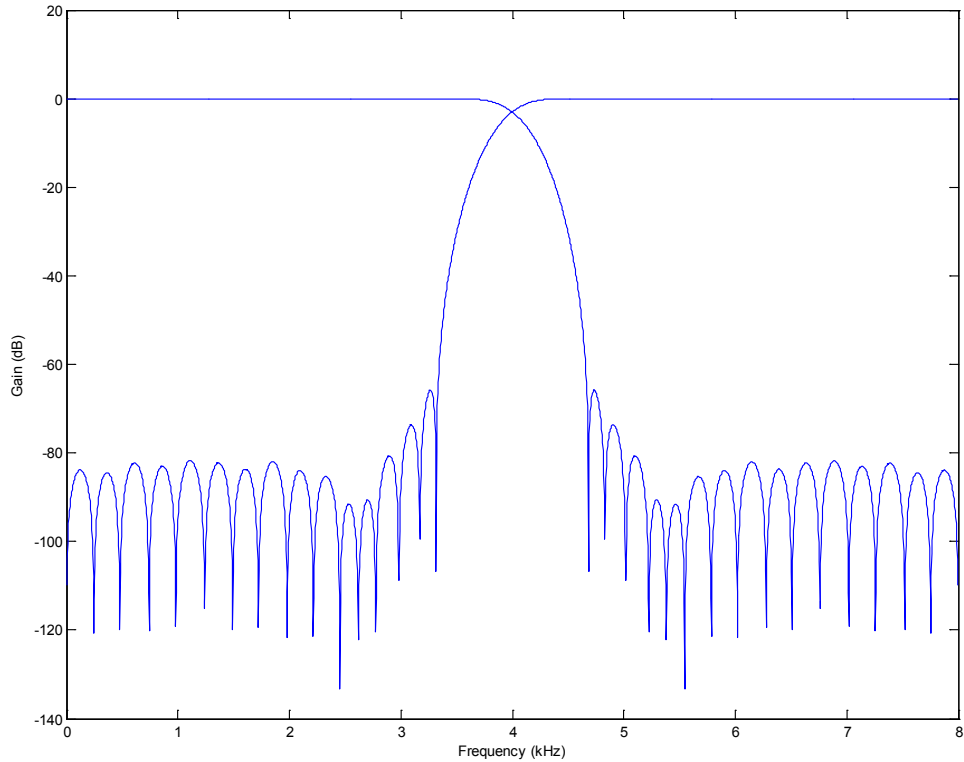


Figure4/JT-G729.1 Frequency responses of  $H_1(z)$  and  $H_2(z)$ .  
(ITU-T G.729.1)

なお、QMF分析フィルタバンクは信号のダウンスケーリング（ファクター2の）処理を含む。高域信号  $s_{HB}^{qmf}(n)$  は、次のように周波数的に折り返される。

$$s_{HB}^{fold}(n) = (-1)^n s_{HB}^{qmf}(n), \quad n = 0, \dots, 159$$

## 6. 2 低域の前処理

低域信号は、50 Hz以下の周波数成分を除去するため、2次の楕円型高域通過フィルタ  $H_{h1}(z)$  によりフィルタリングされる。フィルタ  $H_{h1}(z)$  は下記で定義される。

$$H_{h1}(z) = \frac{0.95551031152729 - 1.91102039813878 z^{-1} + 0.95551031152729 z^{-2}}{1 - 1.96646455789804 z^{-1} + 0.9671820760729101 z^{-2}}$$

## 6. 3 狭帯域エンベデッド符号器（レイヤ1およびレイヤ2）

### 6. 3. 1 線形予測分析と量子化

3. 2節/JT-G729と同一である。

なお、補間され量子化された線形予測フィルタ  $\hat{A}(z)$  は、TDAC符号器にも必要である。CELP符号器による線形予測分析、量子化および補間の結果は、TDAC符号器に供給される。

### 6. 3. 2 聴覚重み付け

3. 3節/JT-G729と同一である。

### 6. 3. 3 開ループピッチ分析

平滑化された開ループピッチの時系列は、聴覚的音声品質の安定化に役立つ。特に、平滑化されたピッチの時系列は、復号側でのフレーム消失補償アルゴリズムにおいて、ピッチ予測（消失フレームにおけるピッチ推定）をより容易にする。JT-G729の開ループピッチ分析アルゴリズムは、倍数ピッチラグを避け

るために現フレームのみを利用している。ここで記述される平滑化アルゴリズムは、前フレームの開ループピッチを利用することで、JT-G729のアルゴリズムを改善する。

JT-G729開ループピッチ分析アルゴリズムは以下のように修正される。10ms フレーム毎に、初期開ループピッチ分析により、下記の3つの探索範囲から3つの候補が与えられる。

$$\{t_1, R'(t_1)\}, \{t_2, R'(t_2)\}, \{t_3, R'(t_3)\}$$

ここで $t_1 > t_2 > t_3$  は開ループピッチ候補であり、 $R'(t_1)$ 、 $R'(t_2)$  および  $R'(t_3)$  は対応する長期相関値である。 $R'(t_1)$  の定義は、3.4節/JT-G729を参照のこと。 $T_{op,old}$  を前フレームの開ループピッチ、そして  $\delta$  を3候補からひとつを選択するための判定に影響を与える閾値とすると、下記に示す3ステップの手順により、前フレームピッチ  $T_{op,old}$  に近いピッチ候補が選ばれ易くする。

ステップ1： 最大ピッチ相関を見つけることで初期候補を選択する。

$$R'_{\max} = \max_{i=1,2,3} R'(t_i)$$

$T_{op}$  を上記の最大相関値に対応する開ループピッチ候補とする。

ステップ2： if  $t_2 < T_{op}$

if  $|T_{op,old} - t_2| < 10$ ,  $\delta = 0.7$  else  $\delta = 0.9$

if  $R'_{\max} \delta < R'(t_2)$ ,  $R'_{\max} = R'(t_2)$  and  $T_{op} = t_2$

ステップ3： if  $t_3 < T_{op}$

if  $|T_{op,old} - t_3| < 5$ ,  $\delta = 0.7$  else  $\delta = 0.9$

if  $R'_{\max} \delta < R'(t_3)$ ,  $T_{op} = t_3$

$T_{op}$  を最終的に選択された現フレームの開ループピッチとする。

#### 6.3.4 インパルス応答の計算

3.5節/JT-G729と同一である。

#### 6.3.5 ターゲット信号の計算

3.6節/JT-G729と同一である。

#### 6.3.6 適応コードブック探索

3.7節/JT-G729と同一である。

#### 6.3.7 8kbit/sにおける固定コードブック — 構造と探索

固定コードブックの構造は、3.8節/JT-G729と同一である。探索法も、本節に記述されるいくつかの箇所を除いては、JT-G729の固定コードブック探索と同様である。

#### 固定コードブック探索の直交化

固定コードブック探索は、フィルタリングされた固定コードブックベクトルがフィルタリングされた適応コードブックベクトルに対して直交化されるような直交化探索を用いて行われる。

直交化探索を実現するために、式(50/JT-G729)と式(51/JT-G729)は、次のように置き換えられる。まず、式(50/JT-G729)は下式で置き換えられる。

$$x'(n) = \left( \sum_{i=0}^{39} y(i)y(i) \right) x(n) - \left( \sum_{i=0}^{39} x(i)y(i) \right) y(n) \quad n = 0, \dots, 39 \quad (7)$$

そして、式(51/JT-G729)は下式で置き換えられる。

$$\phi(i, j) = \left( \sum_{n=0}^{39} y(n)y(n) \right) \left( \sum_{n=j}^{39} h(n-i)h(n-j) \right) - y'(i)y'(j) \quad i = 0, \dots, 39 \quad j = i, \dots, 39 \quad (8)$$

相関信号  $y'(n)$  は、フィルタリングされた適応コードブックベクトル  $y(n)$  とインパルス応答  $h(n)$  とから下式で求められる。

$$y'(39-i) = \sum_{j=0}^i h(j)y(39-i+j) \quad i = 0, \dots, 39 \quad (9)$$

上記のターゲットベクトル  $x'(n)$  と行列  $\Phi_{ij}$  を用い、下記に示す手法により直交化探索が行われる。

#### 全パルス置換を用いた高速コードブック探索

本節に記載の高速探索手法は、パルス位置にのみ関するものである。パルス振幅（極性）は、3. 8. 1 節/J T - G 7 2 9 に記載の手順を用いて予め選択される。

固定コードブックは全パルス置換手順に基づき探索される。この手順は、初期符号ベクトルの決定とパルス置換の2段階で構成される。初期符号ベクトルは、各トラックにおいて下式で与えられるパルス位置尤度推定ベクトル  $b(n)$  を最大にするようなパルス位置により決定される。

$$b(n) = \frac{d(n)}{\sqrt{\sum_{i=0}^{39} d(i)d(i)}} + \frac{r_{LTP}(n)}{\sqrt{\sum_{i=0}^{39} r_{LTP}(i)r_{LTP}(i)}} \quad n = 0, \dots, 39 \quad (10)$$

ここで、 $r_{LTP}(n)$  は下記に示す長期予測残差信号である。

$$r_{LTP}(n) = r(n) - g_p v(n) \quad n = 0, \dots, 39 \quad (11)$$

ここで、 $r(n)$  は入力信号  $s(n)$  の線形予測残差信号、 $g_p$  は式(43/JT-G729)で計算されたピッチ利得、 $v(n)$  は式(40/JT-G729)で算出される適応コードブックベクトルである。

初期符号ベクトルの決定後に、各トラックの初期パルスを式(53/JT-G729)の  $C_k^2/E_k$  を最大にする新たなパルスに置換することにより最適符号ベクトルが探索される。パルス置換手順は4回反復されるか、あるいは、 $C_k^2/E_k$  が増加しないときに終了する。例えば、初期符号ベクトルを  $(m_{01}, m_{11}, m_{21}, m_{31})$  と決定したと仮定する。ここで、記号  $m_{xy}$  の  $x$  はトラック番号を、 $y$  は各トラックのパルス位置を示す。第1ステップにおいて、以下の符号ベクトルでの  $C_k^2/E_k$  を最大にするような新たな符号ベクトルが探索される。

$$\begin{aligned} \text{Track } 0 & (m_{02}, m_{11}, m_{21}, m_{31}), (m_{03}, m_{11}, m_{21}, m_{31}), \dots, (m_{08}, m_{11}, m_{21}, m_{31}) \\ \text{Track } 1 & (m_{01}, m_{12}, m_{21}, m_{31}), (m_{01}, m_{13}, m_{21}, m_{31}), \dots, (m_{01}, m_{18}, m_{21}, m_{31}) \\ \text{Track } 2 & (m_{01}, m_{11}, m_{22}, m_{31}), (m_{01}, m_{11}, m_{23}, m_{31}), \dots, (m_{01}, m_{11}, m_{28}, m_{31}) \\ \text{Track } 3 & (m_{01}, m_{11}, m_{21}, m_{32}), (m_{01}, m_{11}, m_{21}, m_{33}), \dots, (m_{01}, m_{11}, m_{21}, m_{316}) \end{aligned}$$



もし、符号ベクトル  $(m_{03}, m_{11}, m_{21}, m_{31})$  での  $C_k^2/E_k$  が最大であり、かつ初期符号ベクトル  $(m_{01}, m_{11}, m_{21}, m_{31})$  でのそれより大きい場合には、符号ベクトル  $(m_{03}, m_{11}, m_{21}, m_{31})$  が新たな符号ベクトルとして選択される。最初のステップの結果として、初期符号ベクトルの  $m_{01}$  が  $m_{03}$  に置換される。第 2 ステップにおいて、トラック 0 を除く下記の符号ベクトルに対して同様な手順が繰り返される。

*Track 1:*  $(m_{03}, m_{12}, m_{21}, m_{31}), (m_{03}, m_{13}, m_{21}, m_{31}), \dots, (m_{03}, m_{18}, m_{21}, m_{31})$

*Track 2:*  $(m_{03}, m_{11}, m_{22}, m_{31}), (m_{03}, m_{11}, m_{23}, m_{31}), \dots, (m_{03}, m_{11}, m_{28}, m_{31})$

*Track 3:*  $(m_{03}, m_{11}, m_{21}, m_{32}), (m_{03}, m_{11}, m_{21}, m_{33}), \dots, (m_{03}, m_{11}, m_{21}, m_{316})$

第 1 ステップと同じ手順により新たな符号ベクトルが選択される。パルス置換手順が反復されるため、置換された符号ベクトルの  $C_k^2/E_k$  が増加する。この新しい手法においては、テストされるパルス位置組み合わせの回数が少ないため演算量は少ない。

### 6. 3. 8 8 kbit/s における利得の量子化

3. 9 節/J T-G 7 2 9 と同一である。

### 6. 3. 9 12kbit/s における固定コードブック — 構造と探索

12 kbit/s エンハンスメントレイヤでは、追加の固定コードブックにより 8 kbit/s コアレイヤの励振信号と組合わされ、より表現力のある励振信号を生成する。特に、この追加の固定コードブックは励振信号の高域周波数をよりよく表すことができる。高域周波数を強調するために修正聴覚フィルタを使うことでこの効果はより大きくなる。

#### 12 kbit/s ターゲット信号の算出

12 kbit/s FCB 探索のためのターゲット信号  $x_{enh}(n)$  は、線形予測残差  $r(n)$  と 8 kbit/s レイヤにおける式 (75/JT-G729) の励振信号  $u(n)$  との差信号を、フィルタ  $1/\hat{A}(z)$  および  $W(z)$  によりフィルタリングすることにより得られる。

$$x_{enh}(z) = \frac{W(z)}{\hat{A}(z)}(r(z) - u(z)) \quad (12)$$

#### コードブックの構造

12 kbit/s 固定コードブック  $\mathcal{C}$  は、中央のパルス (+1) と反対の極性で低振幅 ( $-\alpha_{enh}$ ) の 2 つのサイドパルスを有するトライパルスパターン  $-\alpha_{enh}z^{-1} + 1 - \alpha_{enh}z$  で定義される。

各符号ベクトルは、このパターンに極性  $\pm 1$  を付与したものを 4 つ足し合わせるにより得られる。パターンの中央は、Table 6/JT-G729.1 に示す J T-G 7 2 9 の固定 ACELP コードブックパルスと同じ位置セットに対応する。

Table 6/JT-G729.1 Structure of the extra fixed codebook.  
(ITU-T G.729.1)

Pattern occurrence	Sign	Positions
$i_0$	$s'_0: \pm 1$	$m'_0: 0, 5, 10, 15, 20, 25, 30, 35$
$i_1$	$s'_1: \pm 1$	$m'_1: 1, 6, 11, 16, 21, 26, 31, 36$
$i_2$	$s'_2: \pm 1$	$m'_2: 2, 7, 12, 17, 22, 27, 32, 37$
$i_3$	$s'_3: \pm 1$	$m'_3: 3, 8, 13, 18, 23, 28, 33, 38$ 4, 9, 14, 19, 24, 29, 34, 39

コードブックベクトル  $c'(n)$  は、40 次元の零ベクトルにパタンを 4 回足し込むことによって作り出される。ブロックサイズに合わせるため、必要に応じて境界におけるパタンは打ち切る。また、対応する極性を各パタンに乗じる。

$$c'(n) = \sum_{i=0}^3 s'_i \times (-\alpha_{enh} \delta(n-m'_i+1) + \delta(n-m'_i) - \alpha_{enh} \delta(n-m'_i-1)) \quad n=0, \dots, 39 \quad (13)$$

ここで、 $\delta(n)$  は単位パルスである。

なお、コアレイヤの固定コードブック探索での適応プリフィルタと類似のように、トライパルスパタンは、以降に示すとおり、通常の代数的コードブック探索手順を用いるために聴覚重み付けフィルタのインパルス応答と組み合わせることができる（式(13)から、トライパルスパタンは適応プリフィルタ  $P'(z) = -\alpha_{enh}z + 1 - \alpha_{enh}z^{-1}$  と等価である）。

#### トライパルスパタンの適応化

係数  $\alpha_{enh}$  は信号の性質に対して適応化される。 $\alpha_{enh}$  を変更することによりトライパルス形状を変えられる。各 5 ms サブフレームに対して  $\alpha_{enh}$  は下記で与えられる。

$$\alpha_{enh} = 0.17(1+r_v) \quad (14)$$

ここで、 $r_v$  は信号の有声性に関連するものであり、下記で与えられる。

$$r_v = (E_v - E_c)/(E_v + E_c) \quad (15)$$

ここで、 $E_v$  および  $E_c$  はそれぞれ、スケーリングされたピッチ符号ベクトル  $\hat{g}_p v(n)$  およびスケーリングされた固定符号ベクトル  $\hat{g}_c c(n)$  のエネルギーである。 $r_v$  の値は-1 から 1 の間であるため、 $\alpha_{enh}$  の値は 0 から 0.34 となる。すなわち、係数  $\alpha_{enh}$  は有声度に関連し、完全な無声セグメントでは値 0 を、完全な有声セグメントでは 0.34 の値を持つ。したがって、12 kbit/s 固定コードブックが有する高域周波数成分は、有声セグメントでより多く、無声セグメントでは少ない。

#### コードブック探索

12 kbit/s 固定コードブックは、ターゲット信号  $x_{enh}(n)$  と高域周波数を強調するための修正聴覚フィルタによりフィルタリングされたコードブック励振音源との間の平均自乗誤差を最小化するように探索される。聴覚

フィルタは、高域通過フィルタ  $-0.15z^{-1} + 1 - 0.15z$  を  $W(z)$  に適用することにより修正される。  
この聴覚フィルタの修正は、ターゲットベクトルと重み付けフィルタインパルス応答を修正することによりなされる。

ターゲット信号の修正は以下により得られる。

$$\begin{aligned} x'_{enh}(n) &= -0.15x_{enh}(n-1) + x_{enh}(n) - 0.15x_{enh}(n+1) \quad n = 0, \dots, 39 \\ \text{where } x_{enh}(-1) &= x_{enh}(40) = 0 \end{aligned} \quad (16)$$

修正重み付けフィルタのインパルス応答  $h_{enh}(n)$  は、下記により得られる。

$$\begin{aligned} h_{enh}(n) &= -0.15h(n-1) + h(n) - 0.15h(n+1) \quad n = 0, \dots, 39 \\ \text{where } h(-1) &= h(40) = 0 \end{aligned} \quad (17)$$

ここで、 $h(n)$  は 3.5 節/J T-G 7 2 9 で算出されるインパルス応答である。探索手順は、トライパルスパタンが J T-G 7 2 9 固定コードブックパルスと同じ位置に中心を置くということを利用する。トライパルスパタンの影響をインパルス応答  $h_{enh}(n)$  に転換することが可能である。サイドパルスの振幅値は中央パルスより十分低い値であるので、ブロック境界での符号ベクトルの打ち切りが生じてもその影響は無視され得る。そのような修正により、12 kbit/s 固定コードブック探索は、8 kbit/s 固定コードブック探索に非常に類似したものになる。

上記を実現するため、インパルス応答  $h_{enh}(n)$  は以下に従って修正される。

$$\begin{aligned} h'_{enh}(n) &= -\alpha_{enh} h_{enh}(n-1) + h_{enh}(n) - \alpha_{enh} h_{enh}(n+1) \quad n = -1, \dots, 39 \\ \text{where } h_{enh}(-2) &= h_{enh}(-1) = h_{enh}(40) = 0 \end{aligned} \quad (18)$$

行列  $\mathbf{H}_{enh}$  は、畳み込み行列として下記で定義される。

$$\mathbf{H}_{enh} = \begin{bmatrix} h'_{enh}(0) & h'_{enh}(-1) & 0 & \dots & 0 \\ h'_{enh}(1) & h'_{enh}(0) & h'_{enh}(-1) & \ddots & \vdots \\ h'_{enh}(2) & h'_{enh}(1) & h'_{enh}(0) & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & h'_{enh}(-1) \\ h'_{enh}(39) & h'_{enh}(38) & h'_{enh}(37) & \dots & h'_{enh}(0) \end{bmatrix} \quad (19)$$

行列  $\mathbf{\Theta}_{enh} = \mathbf{H}_{enh}^t \mathbf{H}_{enh}$  には、 $h_{enh}'(n)$  の相関を含み、行列要素  $\Theta_{enh} = \{\phi_{enh}(i, j)\}$  は以下で与えられる。

$$\begin{aligned} \phi_{enh}(0, 0) &= \sum_{n=0}^{39} h'_{enh}(n)^2 \\ \phi_{enh}(i, j) &= \phi_{enh}(j, i) = \sum_{n=-1}^{39-j} h'_{enh}(n+j-i) \times h'_{enh}(n) \quad i = 0, \dots, 39; \quad j = i, \dots, 39 \end{aligned} \quad (20)$$

相関信号  $d_{enh}(n)$  は、ターゲット信号  $x_{enh}(n)$  と修正インパルス応答  $h'_{enh}(n)$  から以下のように得られる。

$$\begin{aligned} d_{enh}(0) &= \sum_{n=0}^{39} x'_{enh}(n) h'_{enh}(n) \\ d_{enh}(i) &= \sum_{n=-1}^{39-i} x'_{enh}(n+i) h'_{enh}(n) \quad i = 1, \dots, 39 \end{aligned} \quad (21)$$

コードブックは、4つのパタンの位置  $\{m'_i, i=0, \dots, 3\}$  および極性  $\{s'_i, i=0, \dots, 3\}$  により定義される符号ベクトルの中から、下記の項を最小化されるように探索される。

$$\frac{C'^2}{E'} = \frac{\left( \sum_{i=0}^3 s'_i \times d_{enh}(m'_i) \right)^2}{\sum_{i=0}^3 \phi_{enh}(m'_i, m'_i) + 2 \sum_{i=0}^2 \sum_{j=i+1}^3 s'_i \times s'_j \times \phi_{enh}(m'_i, m'_j)} \quad (22)$$

12 kbit/s 固定コードブック探索は、8 kbit/s 固定コードブック探索と同じ方法により行われる。

### インデクス算出

符号ベクトルのインデクスと極性は、8 kbit/s の固定コードブックパルスの極性  $s_i$  および位置  $m_i$  を、生成するパタンの極性  $s'_i$  および中央位置  $m'_i$  に置き換えた 8 kbit/s の固定コードブックのインデクスと極性（3.8.2節/J T-G 7 2 9で定義される）により符号化される。

### 6. 3. 1 0 12 kbit/s 固定コードブック利得の量子化

12 kbit/s 固定コードブック利得の量子化は、ターゲット信号  $x_{enh}(n)$  とフィルタリングされたコードブック励振信号  $z_{enh}(n)$  の間の重み付き平均自乗誤差を最小に行われる。ここで、 $z_{enh}(n)$  は、固定コードブックベクトル  $c'(n)$  に  $h_{enh}(n)$  を畳み込んだものである。

$$z_{enh}(n) = \sum_{k=0}^{39} c'(k) \times h_{enh}(n-k) \quad (23)$$

式(13)により、これは下記と等価である。

$$z_{enh}(n) = \sum_{i=0}^3 s'_i \times h'_{enh}(n-m'_i) \quad (24)$$

J T-G 7 2 9と同様に、コードブック探索後にこのベクトルが得られる。最初に、相互相関項とエネルギー項  $xy$  および  $yy$  が下記により算出される。

$$xy = \sum_{i=0}^{39} z_{enh}(n) \times x'_{enh}(n) \quad (25)$$

$$yy = \sum_{i=0}^{39} z_{enh}(n)^2 \quad (26)$$

12 kbit/s 固定コードブック利得は、8 kbit/s 量子化固定コードブック利得  $\hat{g}_c$  に対して相対的にスカラ量子化される。ビット数はサブフレームインデクスに依存する。

- 各 10ms フレームの第 1 サブフレームに対しては、比率  $\hat{g}_{enh} / \hat{g}_c$  に対してロイドマックス (Lloyd-Max) 量子化器を用い、重み付き平均自乗誤差規準により 3 ビットでスカラ量子化され量子化値  $\hat{g}_{enh} / \hat{g}_c$  を得る。最適な比率  $\hat{g}_{enh} / \hat{g}_c$  は以下を最小化するものである。

$$-2 xy \times \frac{\hat{g}_{enh}}{\hat{g}_c} \times \hat{g}_c + yy \times \left( \frac{\hat{g}_{enh}}{\hat{g}_c} \right)^2 \times (\hat{g}_c)^2 \quad (27)$$

- 各 10ms フレームの第 2 サブフレームに対しては、比率  $g_{enh}/\hat{g}_c$  は、平均自乗誤差規準を用いて[-6 dB, 3 dB]の範囲を dB 単位で 2 ビットで均一量子化される。

### 6. 3. 11 FEC についての信号分類

消失したスーパーフレームにおける信号の再構成のために音声の分類を用いることの背景にある基本的な考え方は、理想的な補償の手法が、準定常的な音声セグメントに対する場合と迅速に変わる特性を持つ音声セグメントに対する場合とで異なるという事実にある。非定常的な音声セグメントにおける消失したスーパーフレームのもっともよい処理は迅速なエネルギーの減衰にまとめることができるのに対して、準定常的な信号の場合は、音声符号化パラメータは、大きくは変わらないので、減衰する直前のいくつかの消失したスーパーフレームの間実際上変わらない状態を保った状態にすることができる。このように、スーパーフレームの消失したブロックに続く信号復元のための最適な手法は、音声信号の分類によって変化する。

この情報は、8kbit/s コアレイヤのみが受信された場合には利用できないことに注意が必要である。この場合、7. 6. 1 節に述べるように、信号分類は復号器で実施される。

音声信号は、有声音、無声音、および無音に粗く分類することができる。有声音は周期的成分を含んでおり、さらに、有声音立上り、有声音セグメント、有声音過渡、有声音立下り、というカテゴリに分割される。有声音立上りは、無音あるいは無声音セグメントのあとの有声音セグメントの開始と定義される。有声音セグメント区間では、音声信号パラメータ（スペクトル包絡、ピッチ周期、周期的成分と非周期成分の比、およびエネルギー）は、スーパーフレーム間でゆっくり変化する。有声音過渡は、たとえば母音間の遷移のように、有声音の迅速な変化によって特徴付けられる。有声音立下りは、有声音セグメントの末端において、エネルギーや有声音性が徐々に低減することによって特徴付けられる。

信号の無声音部分は、周期成分をまったく含まず、さらに、非定常スーパーフレームおよび定常スーパーフレームに分けることができる。非定常スーパーフレームでは、エネルギーとスペクトルが急激に変化し、定常スーパーフレームでは、これらの特徴が比較的定常である。残りのスーパーフレームは、無音に分類される。無音スーパーフレームは、背景雑音が存在する場合の雑音のみのスーパーフレームのような、有音以外のすべてのスーパーフレームを含む。

上記で述べたクラスは、各々別々の補償を行う必要はない。信号の分類のいくつかは、グループ化される。

#### 信号分類

スーパーフレームの分類は、補償および回復の方策を考慮して行われる。つまり、あらゆるスーパーフレームが、次のスーパーフレームが失われた場合に補償が最適になるように、あるいは前のスーパーフレームが失われた場合に回復が最適化できるように分類される。FEC 処理で用いられる分類のいくつかは、復号器であいまいなところなく推定することができるので、送信する必要がない。ここでは、5つの分類が用いられ、次のように定義される。

- UNVOICED クラスは、すべての無声音スーパーフレームおよびすべての有音を含まないスーパーフレームからなる。有声音立下りスーパーフレームでも、その終端が無声音の傾向がある場合、UNVOICED クラスに分類され、次のスーパーフレームが消失した場合に無声音スーパーフレーム用に設計された補償が用いられることがあり得る。
- UNVOICED TRANSITION クラスは、終端で適切な有声音立上りを持つ無声音スーパーフレームからなる。しかし、その立上りは、非常に短期間であるか、あるいは十分に確立されていないので、有声音スーパーフレーム用に設計された補償を利用することができない。UNVOICED TRANSITION クラスは、UNVOICED クラスあるいは UNVOICED TRANSITION クラスに分類されたスーパーフレームにのみ続くことができる。

- VOICED TRANSITION クラスは、比較的弱い有声音の特徴をもつ有声音スーパーフレームからなる。この有声音スーパーフレームは、(母音間の遷移のような) 迅速に特徴量に変化する典型的な有声音スーパーフレーム、あるいは、スーパーフレーム全体で続いている有声音立下りスーパーフレームである。VOICED TRANSITION クラスは、VOICED TRANSITION クラス、VOICED クラス、あるいは ONSET クラスに分類されたスーパーフレームにのみ続くことができる。
- VOICED クラスは、定常的な特徴量をもつ有声音スーパーフレームからなる。このクラスは、VOICED TRANSITION クラス、VOICED クラス、あるいは ONSET クラスに分類されたスーパーフレームにのみ続くことができる。
- ONSET クラスは、UNVOICED クラスあるいは UNVOICED TRANSITION クラスに分類されたスーパーフレームに続く定常的な特徴量をもつすべての有声音スーパーフレームからなる。ONSET クラスに分類されたスーパーフレームは、消失した有声音スーパーフレーム用に設計された補償での利用のために立上りがすでに十分に確立された有声音立上りスーパーフレームに対応する。ONSET クラスに続くスーパーフレーム消失用に用いられるこの補償技術は、VOICED クラスに続くスーパーフレーム消失の場合と同一である。違いは、復元の手法にある。立上りスーパーフレームが消失した場合(すなわち、消失の直後に正常な有声音スーパーフレームが到着したが、消失前最後の正常なスーパーフレームが UNVOICED クラスの場合)は、消失した立上りスーパーフレームを擬似的に再構成するために特別な技術を用いることができる。この擬似的な立上り再生技術は、7. 6. 7 節で述べる。一方、正常な ONSET クラスが消失の直後に到着し、消失前最後の正常なスーパーフレームが UNVOICE の場合は、立上りが失われない(消失したスーパーフレームにない)ので、この特別な処理は不要である。

UNVOICED TRANSITION クラスおよび VOICED TRANSITION クラスは、復号器で明確に分離することができる (UNVOICED TRANSITION クラスは UNVOICED クラスか UNVOICED TRANSITION クラスにのみ続くことができ、VOICED TRANSITION クラスは ONSET クラス、VOICED クラス、あるいは VOICED TRANSITION クラスにのみ続くことができる) ので、両者はグルーピングできることに注意すること。つまり、5つのクラスがあるが、必要なビット数は2ビットのみである。

### クラス情報

クラス情報の2ビットは、VOICED、UNVOICED、ONSET、TRANSITION の4つのクラスを示す。復号器では、TRANSITION クラスが受信された場合は、直前のスーパーフレームのクラスが UNVOICED クラスか UNVOICED TRANSITION クラスの場合は UNVOICED TRANSITION クラスと解釈され、それ以外の場合は VOICED TRANSITION と解釈される。

### 信号分類パラメータ

符号器において、正規化相関係数  $r_x$ 、スペクトル傾斜尺度  $e_l$ 、信号対雑音比  $snr$ 、ピッチ定常性カウンタ  $pc$ 、現在のスーパーフレームの末端における信号の相対的スーパーフレームエネルギー  $E_s$ 、およびゼロ交差カウンタ  $zc$ 、といったパラメータが分類に用いられる。信号を分類するために用いられるこれらのパラメータの計算方法を以下に述べる。

### 正規化相関係数

正規化相関係数  $r_x$  は、10ms ごと (つまりスーパーフレームあたり2回) に開ループピッチ推定値を出力する開ループピッチ探索の一部として計算される。この探索部が正規化相関係数尺度を出力するためにも用いられる。この正規化相関係数は、開ループピッチ遅延における現在の重み付け音声信号  $s_w(n)$  上で計算される。平均相関係数  $\bar{r}_{sw}$  は、下記のように定義される。

$$\bar{r}_{sw} = 0.5(r_{sw}(1) + r_{sw}(2)) \quad (28)$$

ここで、 $r_{sw}(1)$ 、 $r_{sw}(2)$  は、それぞれ、第1および第2番目の10msのフレームの正規化相関係数である。正規化相関係数は、下記のように計算される。

$$r_{sw}(i) = \frac{\sum_{n=0}^{79} s_w(80(i-1)+n)s_w(80(i-1)+n-T_{op}(i))}{\sqrt{\sum_{n=0}^{79} s_w^2(80(i-1)+n) \sum_{n=0}^{79} s_w^2(80(i-1)+n-T_{op}(i))}} \quad (29)$$

相関係数  $r_{sw}(i)$  は重み付け音声信号  $s_w(n)$  を用いて計算される。 $T_{op}(i)$  は、6.3.3節で述べたように、 $i$  番目の10msフレームで計算された開ループピッチである。

#### スペクトル傾斜

スペクトル傾斜は、音声信号の正規化された1次の自己相関係数（すなわちLP分析中に得られた1次の反射係数）として推定される。LP分析はスーパーフレームあたり2回（すなわち10msフレームあたり1回）実行されるので、スペクトル傾斜はその二つのLP分析から得られた1次の反射係数の平均として計算される。つまり、以下のようになる。

$$e_i = -0.5(k_1(1) + k_1(2)) \quad (30)$$

ここで、 $k_1(i)$  は、TTC標準JT-G729本体の3.2.2節にあるような方法で計算された  $i$  番目のフレームにおけるLP分析から得られた1次の反射係数である。

#### 信号対雑音比

信号対雑音比（SNR）測定は、一般的な波形マッチング型の符号器に関しては、人間の音声についてのSNRがより高いという事実を利用する。 $snr$  パラメータは、以下の関係を利用して、スーパーフレーム全体について計算される。

$$snr = \frac{E_s}{E_d} \quad (31)$$

ここで  $E_s$  は現在のスーパーフレーム内の音声信号  $s(n)$  のエネルギーであり、 $E_e$  は音声信号  $s(n)$  と、現在のスーパーフレーム内の12kbit/sでの局所的な合成信号  $\hat{s}_{enh}(n)$  との差分  $d_{LB}(n)$  のエネルギーである。

#### ピッチ定常性カウンタ

ピッチ定常性カウンタ  $pc$  は、ピッチ周期の変化量を評価する。次のように計算する。

$$pc = |T_2(2) - T_1(2)| + |T_1(2) - T_2(1)| \quad (32)$$

ここで、値  $T_1(i)$  および  $T_2(i)$  は、それぞれ、 $i$  番目の10msフレームの第1および第2番目の閉ループピッチラグに対応する。すなわち、 $pc$  は、最後の3つの5ms長サブフレームの閉ループピッチラグで計算される。

#### 相対エネルギー

相対スーパーフレームエネルギー  $E_s$  は、スーパーフレームの dB 単位のエネルギーと、その長期間平均との差分として計算される。

$$E_s = E_t - E_{lt} \quad (33)$$

ここで、スーパーフレームエネルギー  $E_t$  は、窓掛け入力信号の dB 単位のエネルギーである。

$$E_t = 10 \log_{10} \left( \frac{1}{160} \sum_{n=0}^{159} [s(n)w_{Hanning}(n)]^2 \right) \quad (34)$$

ここで、 $w_{Hanning}(n)$  は 160 サンプル長のハニング窓である。すなわち、

$$w_{Hanning}(n) = \frac{1}{2} \left( 1 - \cos \left( \frac{2\pi n}{159} \right) \right), \quad n = 0, \dots, 159 \quad (35)$$

である。長期平均エネルギーは、次の関係を用いて、有音のスーパーフレーム中に更新される。

$$E_{lt} = 0.99E_{lt} + 0.01E_t \quad (36)$$

#### ゼロ交差

最後のパラメータは、現在の音声スーパーフレーム上で計算されるゼロ交差パラメータ  $zc$  である。ここでは、ゼロ交差カウンタ  $zc$  は、現在の音声スーパーフレーム内で信号の極性が正の値から負の値に変化する回数をカウントする。

#### 分類手順

分類パラメータは、評価関数  $f_m$  を定義するために用いられる。この目的のために、それぞれの分類パラメータが無声音で 0 および有声音で 1 に変換されるように、分類パラメータは 0 から 1 の値にスケールされる。それぞれのパラメータ  $p_x$  は、次のような線形関数によってスケールされる。

$$p_x^s = k_x p_x + c_x \quad (37)$$

そして、0 から 1 の間の値になるように、クリッピングされる。ただし、相対エネルギーのみ 0.5 から 1 の間の値になるようにクリッピングされる。スーパーフレーム消失が存在する場合に用いられる補償および復元技術による信号のひずみが最小になるように、この関数の係数  $k_x$  および  $c_x$  は、それぞれのパラメータに対して、経験的に求められている。用いられる値は、Table7/JT-G729.1 にまとめられている。

Table7/JT-G729.1 Scaling function coefficients of the classification parameters.  
(ITU-T G.729.1)

Parameter	Meaning	$k_x$	$c_x$
$\bar{r}_{sw}$	Normalized correlation	0.91743	0.26606
$e_t$	Spectral tilt	2.5	-1.25
$snr$	Signal to noise ratio	0.09615	-0.25
$pc$	Pitch stability counter	-0.1176f	2.0
$E_s$	Relative frame energy	0.05	0.45
$zc$	Zero crossing counter	-0.067	2.613



評価関数は次のように定義されている。

$$f_m = \frac{1}{7}(2\bar{r}_{sw}^s + e_t^s + 1.2snr^s + pc^s + E_s^s + zc^s) \quad (38)$$

ここで、肩文字  $s$  はパラメータのスケーリングされたバージョンを示す。

そして、評価関数は、 $E_s^s \leq 0.5$  の場合に 1.05 でスケーリングされ、 $E_s^s > 0.75$  の場合に 1.25 でスケーリングされる。さらに、メリットの関数は、瞬間的な相対エネルギーの変化と長期相対エネルギーの変化との違いをチェックする状態機械に基づく要素  $f_E$  によってもスケーリングされる。これは、背景雑音が存在する場合に信号分類を改善するために追加される。このため、相対エネルギー変化パラメータ  $E_{var}$  は次のように更新される。

$$E_{var} = 0.05(E_s - E_{prev}) + 0.95E_{var} \quad (39)$$

ここで  $E_{prev}$  は前のスーパーフレームの  $E_s$  である。

if  $|E_s - E_{prev}| < (E_{var} + 6)$  and  $(class_{old} = UNVOICED)$   $f_E = 0.8$   
 else  
     if  $(E_s - E_{prev}) > (E_{var} + 3)$  and  $(class_{old} = UNVOICED \text{ or } TRANSITION)$   $f_E = 1.1$   
 else  
     if  $(E_s - E_{prev}) < (E_{var} - 5)$  and  $(r_{sw}(1) \text{ } class_{old} = VOICED \text{ or } ONSET)$   $f_E = 0.6$ .

ここで  $class_{old}$  は前のスーパーフレームのクラスである。

そして、評価関数  $f_m$  を用いて、かつ Table8/JT-G729.1 にまとめたようなルールに従って、分類が実施される。

Table8/JT-G729.1 Signal classification rules at the encoder.  
(ITU-T G.729.1)

Previous frame class	Rule	Current superframe class
ONSET VOICED VOICED TRANSITION	$f_m \geq 0.68$	VOICED
	$0.56 \leq f_m < 0.68$	VOICED TRANSITION
	$f_m < 0.56$	UNVOICED
UNVOICED TRANSITION UNVOICED	$f_m > 0.64$	ONSET
	$0.64 \geq f_m > 0.58$	UNVOICED TRANSITION
	$f_m \leq 0.58$	UNVOICED

FECの分類のという目的のために、有音でないスーパーフレームは UNVOICED クラスに分類される。相対エネルギーが（音声の非有音性を簡単に測るものとして） $-10\text{dB}$  未満の場合、スーパーフレームは直接 UNVOICED クラスに分類される。

### 6. 3. 12 メモリ更新

TTC標準JT-G729本体の3.10節と同一である。

加えて、次のサブフレームにおける 12kbit/s の目標信号のために、合成フィルタおよび重み付けフィルタ

の状態が更新される。このメモリ更新は、8kbit/s で再構成された励振信号の代わりに 12kbit/s で再構成された励振信号に基づいているという点を除き、JT-G 7 2 9 本体と同一である。

#### 6. 4 高域のための前処理

元の信号における 7000~8000Hz の帯域に対応する 3000~4000Hz 帯域の周波数成分を除去するために、高域信号は、4 次の楕円型の低域通過フィルタ  $H_{h2}(z)$  によってフィルタリングされる。フィルタ  $H_{h2}(z)$  は次のように定義される。

$$H_{h2}(z) = \frac{0.3500277721 + 1.3045646694 z^{-1} + 1.9127698530 z^{-2} + 1.3045646694 z^{-3} + 0.3500277721 z^{-4}}{1 + 1.79857371201 z^{-1} + 1.69962113314 z^{-2} + 0.70669663302 z^{-3} + 0.16954708937 z^{-4}} \quad (40)$$

#### 6. 5 TDBWE 符号器 (レイヤ 3)

TDBWE 符号器を Figure5/JT-G729.1 に示す。時間領域帯域拡張 (TDBWE) 符号器は、前処理を施され、ダウンサンプルされた高域信号  $s_{HB}(n)$  から、非常に粗いパラメトリックな記述を抽出する。このパラメトリックな記述は、時間包絡およびスペクトル包絡のパラメータを含む。それぞれの包絡の計算については 6. 5. 1 節および 6. 5. 2 節を、またこのパラメータの量子化手順の記載については 6. 5. 3 節を参照すること。

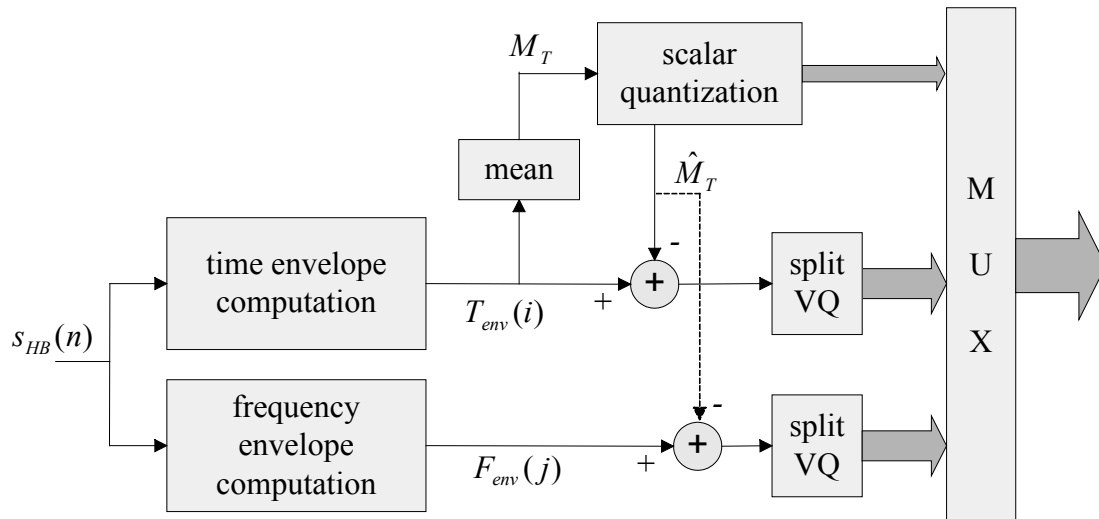


Figure5/JT-G729.1 High-level block diagram of the TDBWE encoder (ITU-T G.729.1)

##### 6. 5. 1 時間包絡の計算

20ms の入力音声スーパーフレーム  $s_{HB}(n)$  は、それぞれは 1.25ms の長さの 16 の領域に分割される。つまり、それぞれの領域は 10 サンプルを含むことになる。16 個の時間包絡パラメータ  $T_{env}(i)$ ,  $i = 0, \dots, 15$  は、次のように、対数的なサブフレームエネルギーとして計算される。

$$T_{env}(i) = \frac{1}{2} \log_2 \left( \sum_{n=0}^9 s_{HB}^2(n+i \cdot 10) \right), \quad i = 0, \dots, 15 \quad (41)$$

##### 6. 5. 2 周波数包絡の計算

12 個の周波数包絡パラメータ  $F_{env}(j)$ ,  $j = 0, \dots, 11$  の計算のために、信号  $s_{HB}(n)$  はわずかに非対称な分析窓  $w_F(n)$  によって窓掛けされる。この窓は、次のように、128 タップ (16ms) 長であり、144 タップのハンニング窓の立上り傾斜、それに続く 112 タップのハンニング窓の立ち下がり傾斜から構成される。

$$w_F(n) = \begin{cases} \frac{1}{2} \left( 1 - \cos \left( \frac{2\pi n}{143} \right) \right), & n = 0, \dots, 71 \\ \frac{1}{2} \left( 1 - \cos \left( \frac{2\pi (n-16)}{111} \right) \right), & n = 72, \dots, 127 \end{cases}$$

窓  $w_F(n)$  の最大値は、現スーパーフレームの 2 番目の 10ms フレームの中心に位置する。2 番目の 10ms フレームにわたって窓掛けされた信号は、以下のように与えられる。

$$s_{HB}^w(n) = s_{HB}(n) \cdot w_F(n+31), \quad n = -31, \dots, 96 \quad (42)$$

窓  $w_F(n)$  は、周波数包絡の計算が 16 サンプル (2ms) の先読みおよび 32 サンプル (4ms) の後読みを持つように構成される。Figure6/JT-G729.1 は、窓関数  $w_F(n)$  を描いたものである。

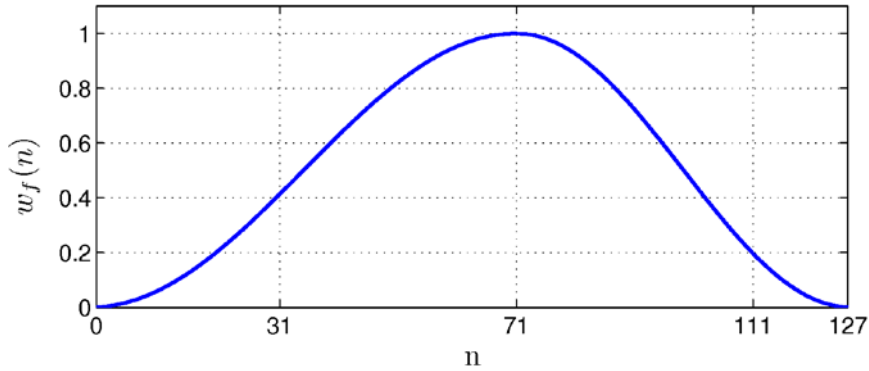


Figure6/JT-G729.1 Window for the frequency envelope computation (ITU-T G.729.1)

窓掛けされた信号  $s_{HB}^w(n)$  は、64 の長さの F F T によって変換される。この F F T 長で、128 タップの F F T の偶数成分がポリフェーズの構成を用いて以下のように計算される。

$$S_{HB}^{ff}(k) = FFT_{64} \left( s_{HB}^w(n) + s_{HB}^w(n+64) \right), \quad k = 0, \dots, 63, \quad n = -31, \dots, 32 \quad (43)$$

最後に、この周波数包絡パラメータのセットが、12 の F F T の領域における等間隔で同じ幅にオーバーラップしたサブバンドについて、対数領域の重み付けされたサブバンドのエネルギーとして計算される。

$$F_{env}(j) = \frac{1}{2} \log_2 \left( \sum_{k=2j}^{2(j+1)} W_F(k-2j) \cdot |S_{HB}^{ff}(k)|^2 \right), \quad j = 0, \dots, 11. \quad (44)$$

j 番目のサブバンドは 2 j のインデックスの F F T の成分で始まり、F F T 成分の 3 倍の帯域を持つ。これは、Table9/JT-G729.1 に示すように、物理的なサブバンド分割に対応する。この表記における 0 Hz は元の広帯域周波数領域における 4000Hz の周波数に対応することに注意すること。

Table9/JT-G729.1 Subband division for the frequency envelope  
(ITU-T G.729.1)

Subband index $j$	Subband frequency range [F F T bins] $k = 2j, \dots, 2(j+1)$	Subband frequency range [Hz] $f = \max\left(0, \frac{k-0.5}{64} \cdot 8000\text{Hz}\right) \dots \frac{k+0.5}{64} \cdot 8000\text{Hz}$
0	0,...,2	0 ... 312.5
1	2,...,4	187.5 ... 562.5
2	4,...,6	437.5 ... 812.5
3	6,...,8	687.5 ... 1062.5
4	8,...,10	937.5 ... 1312.5
5	10,...,12	1187.5 ... 1562.5
6	12,...,14	1437.5 ... 1812.5
7	14,...,16	1687.5 ... 2062.5
8	16,...,18	1937.5 ... 2312.5
9	18,...,20	2187.5 ... 2562.5
10	20,...,22	2437.5 ... 2812.5
11	22,...,24	2687.5 ... 3062.5

物理的な帯域幅はそれぞれの（最初以外の）サブバンドに対して  $\Delta f = 375\text{Hz}$  である。周波数領域の重み付け窓は以下のように与えられる。

$$W_F(0) = 0.5, \quad W_F(1) = 1, \quad \text{and} \quad W_F(2) = 0.5 \quad (45)$$

### 6. 5. 3 TDBWEパラメータの量子化

TDBWEパラメータ  $T_{env}(i)$ ,  $i = 0, \dots, 15$  および  $F_{env}(j)$ ,  $j = 0, \dots, 11$  は、平均を除去した分割ベクトル量子化によって量子化される。まず、時間包絡の平均が次のように計算される。

$$M_T = \frac{1}{16} \sum_{i=0}^{15} T_{env}(i) \quad (46)$$

次に、平均値  $M_T$  は、対数領域において、3 dB の均一なステップを用いて5ビットでスカラー量子化される。この量子化は、量子化された値  $\hat{M}_T$  を与える。次に、以下のように、量子化された平均値が差し引かれる。

$$T_{env}^M(i) = T_{env}(i) - \hat{M}_T, \quad i = 0, \dots, 15 \quad (47)$$

および

$$F_{env}^M(j)_i = F_{env}(j) - \hat{M}_T, \quad j = 0, \dots, 11 \quad (48)$$

平均が除去された時間包絡パラメータのセットは、以下のように、2つの8次元のベクトルに分割される。

$$\mathbf{T}_{env,1} = (T_{env}^M(0), T_{env}^M(1), \dots, T_{env}^M(7)) \quad \text{and} \quad \mathbf{T}_{env,2} = (T_{env}^M(8), T_{env}^M(9), \dots, T_{env}^M(15)) \quad (49)$$

一方、周波数包絡パラメータは、以下のように3つの4次元のベクトルに分割される。

$$\begin{cases} \mathbf{F}_{env,1} = (F_{env}^M(0), F_{env}^M(1), F_{env}^M(2), F_{env}^M(3)) \\ \mathbf{F}_{env,2} = (F_{env}^M(4), F_{env}^M(5), F_{env}^M(6), F_{env}^M(7)) \\ \mathbf{F}_{env,3} = (F_{env}^M(8), F_{env}^M(9), F_{env}^M(10), F_{env}^M(11)) \end{cases} \quad (50)$$

最後に、事前に学習された量子化テーブルを用いたベクトル量子化が、Table10/JT-G729.1 に示すように適用される。ベクトル  $\mathbf{T}_{env,1}$  および  $\mathbf{T}_{env,2}$  は必要なメモリ量 (storage requirements) の削減のために、同一のベク

トル量子化コードブックを共有する。 $\mathbf{T}_{env,1}/\mathbf{T}_{env,2}$ ,  $\mathbf{F}_{env,1}$ ,  $\mathbf{F}_{env,2}$  および  $\mathbf{F}_{env,3}$  のためのコードブック (あるいは量子化テーブル) は、二つのセントロイド間の距離の最小値が実証できるように一般化ロイドマックスによるセントロイドを修正することによって生成された。コードブック修正は、ロイドマックスによるセントロイドを対数領域における 6 dB のステップサイズを持った矩形のグリッド上に丸めることによって行われる。

Table10/JT-G729.1 Time and frequency envelope quantization  
(ITU-T G.729.1)

Parameter vector	Quantized vector	Dimension	Number of allocated bits
$M_T$	$\hat{M}_T$	1	5
$\mathbf{T}_{env,1}$	$\hat{\mathbf{T}}_{env,1}$	8	7
$\mathbf{T}_{env,2}$	$\hat{\mathbf{T}}_{env,2}$	8	7
$\mathbf{F}_{env,1}$	$\hat{\mathbf{F}}_{env,1}$	4	5
$\mathbf{F}_{env,2}$	$\hat{\mathbf{F}}_{env,2}$	4	5
$\mathbf{F}_{env,3}$	$\hat{\mathbf{F}}_{env,3}$	4	4

#### 6. 5. 4 FECのための位相情報

6. 3節で述べたように、補償・復元パラメータは、信号分類情報、エネルギー、および位相情報を含んでおり、異なるビットストリームレイヤで送信される。この位相情報は、直前のスーパーフレームにおける最後の声門パルスの位置からなるが、レイヤ 3 (14kbit/s) で送信される。

消失したスーパーフレームのブロックのあとは、復号器のメモリは符号器のメモリと同期していないため、有声音の消失したセグメントのあとの復元中は、位相のコントロールが特に重要である。送信された位相情報は、直前のスーパーフレームにおける声門パルスの位置と極性からなる。したがって、この情報は、7.

6. 7節で述べるように、有声音立上りが消失したあとの復元に用いられる。より重要な点として、この情報は、正しく受信された連続的なスーパーフレームにおける収束を早めたり、エラーが伝搬するのを低減するために、消失した励振信号を再び同期させるためにも用いられる。

(デフォルトモードにおける) JT-G729.1 復号器では、MDCT再構成におけるオーバーラップ加算には、ひとつのスーパーフレームの遅れが復号器で利用可能である。このように、ひとつのスーパーフレームが消失した場合、追加のスーパーフレームの遅延のため、後続のスーパーフレームのパラメータが利用可能である。この場合、消失したスーパーフレームの末端の最大のパルスの位置と極性が、後続のスーパーフレームから利用可能である。すなわち、ピッチ励振信号は、この最後の最大パルスを下に説明するようにこの後続のスーパーフレームで受信された位置に合わせるといった方法で、補償される。

最大パルスの探索は、このように、低域通過フィルタを適用された線形予測残差に対して行われる。

$$r^{lpf}(n) = 0.25r(n-1) + 0.5r(n) + 0.25r(n+1) \quad (51)$$

最後の声門パルス  $\tau$  の位置は、このスーパーフレームにおける、低域フィルタを適用された残差  $r^{lpf}(n)$  の最後の  $T_0$  サンプルの中で振幅が最大の絶対値を持つサンプルを求めることによって探索される。ここで、 $T_0$  は最後のサブフレームの丸められた閉ループピッチラグ  $T_2(2)$  である。つまり  $T_0 = \text{int}(T_2(2))$  となる。 $\tau$  がスーパーフレームの末端に関連した位置であることに注意すること。

最後の声門パルス  $\tau$  の位置は、以下のように、6 ビットを用いて符号化される。最初の声門パルスの位置を符号化するために用いられる精度は、 $T_0$  に依存する。これが可能なのは、この値が符号器および復号器の

双方で既知であり、またひとつあるいは複数のスーパーフレーム消失後でもエラー伝搬の影響を受けないためである。もし  $T_0 < 64$  の場合、スーパーフレームの末端に関連した最後の声門パルスの位置は、1 サンプルの精度で直接符号化される。  $64 \leq T_0 < 128$  の場合は、スーパーフレームの末端に関連する最後の声門パルスの位置は、たとえば  $\tau/2$  のように単純な整数分割を用いて 2 サンプルの精度で符号化される。もし  $T_0 \geq 128$  の場合は、スーパーフレームの末端に関連する最後の声門パルスの位置は、 $\tau$  をさらに 2 で分割することで、4 サンプルの精度で符号化される。

振幅が最大の絶対値を持つパルスの極性は、1 ビットで送信される。これにより、位相情報のために 7 ビットが用いられることになる。声門パルスの形状は、通常、極性の異なる二つの大きなパルスを含むので、極性が位相の再同期のために重要である。極性を無視すると、その位置において小さなドリフトが発生したり、再同期処理の性能が低減されてしまったりする可能性がある。

## 6. 6 TDAC符号器 (レイヤ4から12)

時間領域折り返し歪み打ち消し (TDAC) 符号器を Figure7/JT-G729.1 に図示する。TDAC符号器は、利得形状ベクトル量子化により、2つに分割されたMDC Tスペクトル  $D_{LB}^w(k)$  と  $S_{LB}(k)$  を結合して表現する。結合されたスペクトルはサブバンドに分割される。各サブバンドの利得は、スペクトル包絡を規定する。各サブバンドの形状は、学習された順列符号を用いたエンベデット球面ベクトル量子化により符号化される。

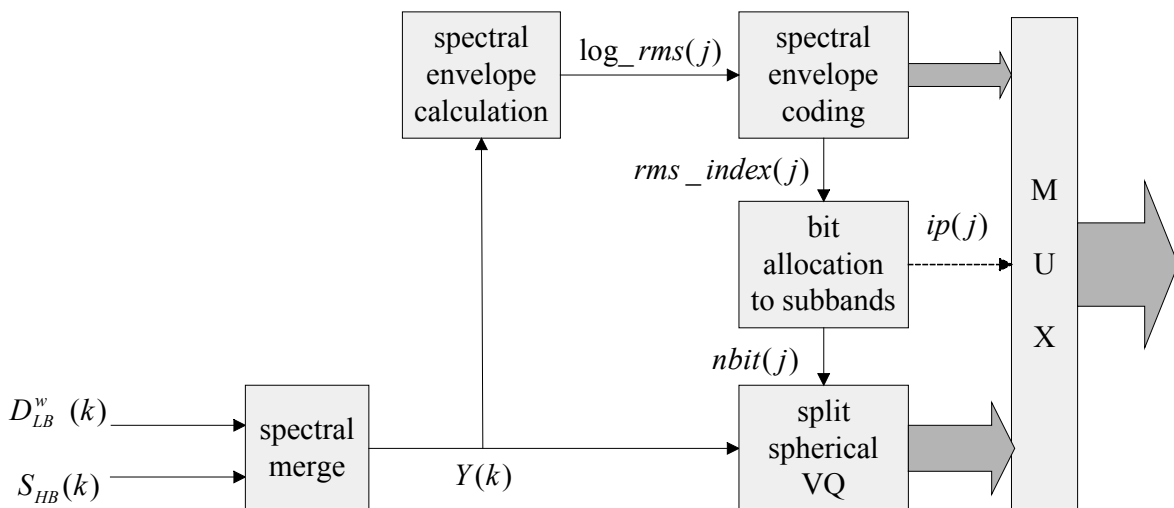


Figure7/JT-G729.1 High-level block diagram of the TDAC encoder (ITU-T G.729.1)

### 6. 6. 1 FECのためのエネルギー情報

3つ目の隠蔽/回復パラメータは、符号器における 12kbit/s 時の局部合成信号  $\hat{s}_{enh}(n)$  に基づいて計算されるエネルギー情報である。エネルギー情報は、レイヤ4 (16kbit/s) で伝送される。従って、レイヤ4が受信されれば、フレーム消失補償を改善するためにこの情報を利用することができる。そうでない場合は、エネルギーは復号器側で推定される。

エネルギー情報  $E$  は、VOICED もしくは ONSET として分類されたスーパーフレームの最大サンプルエネルギー、もしくは他のスーパーフレームのサンプルあたりの平均エネルギーである。VOICED もしくは ONSET スーパーフレームに対しては、最大サンプルエネルギーは以下のようにスーパーフレームの最後にピッチに同期して計算される：

$$E = \max_{n=160-t_g \dots 159} \hat{s}_{enh}^2(n) \quad (52)$$

ここで、 $t_E$  は第 2 フレームの最後のサブフレームの閉ループピッチラグ  $T_2(2)$  に基づいて計算される。 $T_2(2) > 40$  であれば  $t_E = \text{round}(T_2(2))$  とし、そうでなければ  $t_E = 2 \text{round}(T_2(2))$  とする。他の分類に対しては、 $E$  は第 2 の 10ms フレームのサンプルあたりの平均エネルギーであり、 $t_E = 80$  として以下のように計算される：

$$E = \frac{1}{t_E} \sum_{n=160-t_E}^{159} \hat{s}_{enh}^2(n) \quad (53)$$

ステップサイズが 3.1dB で、範囲が 0 から 96dB の 5 ビット一様量子化器が用いられる。量子化インデックスは以下により与えられる。

$$i = \text{int} \left( \frac{10 \log_{10}(E + 0.001)}{3.1} \right) \quad (54)$$

ここで、インデックスは  $0 \leq i \leq 31$  の範囲に制限される。

### 6. 6. 2 CELP 差分信号の聴覚重み付け

エンベデット CELP 符号器の入力  $s(n)$  と 12kbit/s の局部合成  $\hat{s}_{enh}(n)$  との差分  $d_{LB}(n)$  は、以下で定義される聴覚重み付けフィルタ  $W_{LB}(z)$  により処理される：

$$W_{LB}(z) = \text{fac} \frac{\hat{A}(z/\gamma_1')}{\hat{A}(z/\gamma_2')} \quad (55)$$

ここで、 $\text{fac}$  は以下で与えられる利得補正である。

$$\text{fac} = \frac{\left| \sum_{i=0}^{10} (-\gamma_2')^i \hat{a}_i \right|}{\left| \sum_{i=0}^{10} (-\gamma_1')^i \hat{a}_i \right|} \quad (56)$$

そして、 $\hat{a}_i$  はエンベデット CELP 符号器の下式から得られる量子化された線形予測フィルタ  $\hat{A}(z)_i$  の係数である。

$$\hat{A}(z) = \hat{a}_0 + \hat{a}_1 z^{-1} + \dots + \hat{a}_{10} z^{-10} \quad (57)$$

係数  $\hat{a}_i$  は、3. 2. 5 節/J T-G 7 2 9 および 3. 2. 6 節/J T-G 7 2 9 に説明されている通り、各 5ms サブフレームで更新される。パラメータ  $\gamma_1' = 0.94$  と  $\gamma_2' = 0.6$  は定数である。等価的に、係数  $\text{fac}$  は、低域 (4000Hz) のナイキスト周波数における  $\hat{A}(z/\gamma_1)/\hat{A}(z/\gamma_2)$  の利得の逆数として定義される。

$$\text{fac} = \frac{\left| \hat{A}(z/\gamma_2') \right|}{\left| \hat{A}(z/\gamma_1') \right|} \quad (58)$$

ここで、 $z = -1$  である。利得補正係数は、 $W_{LB}(z)$  の出力  $d_{LB}^v(n)$  と、隣接の高域の信号  $s_{HB}(n)$  との間でスペクトルの連続性を保証することが必要とされる。

フィルタ  $W_{LB}(z)$  は、短期逆周波数マスキングカーブを形成し、平均自乗誤差の尺度で最適化された MDCT 符号化の適用を可能にする。そしてまた、差分信号  $d_{LB}(n)$  を 8 および 12kbit/s で使用される CELP ター

ゲット領域と同様の重み付け領域に写像する。

### 6. 6. 3 MDCT

低域、高域信号  $d_{LB}^w(n)$  と  $s_{HB}(n)$  は、スーパーフレーム長が 20ms で窓長が 40ms の変形離散コサイン変換 (MDCT) により周波数領域に変換される。40ms の正弦窓掛けによる信号  $d_{LB}^w(n)$  の MDCT 出力  $D_{LB}^w(k)$  は、以下のように与えられる：

$$D_{LB}^w(k) = \frac{\sqrt{2}}{160} \sum_{n=0}^{319} w_{TDAC}(n) \cos\left(\frac{\pi}{160}(n+80.5)(k+0.5)\right) d_{LB}^w(n), \quad k=0, \dots, 159 \quad (59)$$

ここで、 $w_{TDAC}(n)$  は分析重み窓である。

$$w_{TDAC}(n) = \sin\left(\frac{\pi}{320}(n+0.5)\right), \quad n=0, \dots, 319 \quad (60)$$

160 個の MDCT 係数の計算は以下のようになされる。

最初に、複素数  $z(n) = z_R(n) + j z_I(n)$ ,  $n=0, \dots, 79$  の実部と虚部が、下式により計算される。

$$\begin{aligned} z_R(n) &= w_{TDAC}(2n) d_{LB}^w(2n) - w_{TDAC}(159-2n) d_{LB}^w(159-2n) \\ z_I(n) &= w_{TDAC}(319-2n) d_{LB}^w(319-2n) + w_{TDAC}(160+2n) d_{LB}^w(160+2n) \end{aligned} \quad (61)$$

複素信号  $z(n)$  は、その後  $z'(n)$  に変換される：

$$z'(n) = (W_{320})^n \times z(n), \quad n=0, \dots, 79 \quad (62)$$

ここで、 $W_m = e^{2\pi j/m}$  は単位円の  $m$  次根を表す。

$z'(n)$  の逆 FFT は、係数  $Z'(k)$ ,  $k=0, \dots, 79$  を得るために計算される。これらの係数は以下のように変換される：

$$Z(k) = (-1)^{k+1} \times (W_8)^{-1} \times (W_{1280})^{4k+1} \times Z'(k) \quad k=0, \dots, 79 \quad (63)$$

ここで、 $Z_R(k)$  と  $Z_I(k)$  は、 $Z(k)$  の実部と虚部である。MDCT 係数は下式により与えられる。

$$\begin{cases} D_{LB}^w(2k) = Z_I(k) \\ D_{LB}^w(159-2k) = -Z_R(k) \end{cases} \quad \text{for } k=0, \dots, 79 \quad (64)$$

$s_{HB}(n)$  の MDCT 出力  $S_{HB}(k)$  は、同様の手法で計算される。得られた低域、高域 MDCT スペクトル  $D_{LB}^w(k)$  と  $S_{HB}(k)$  は、全域スペクトル  $Y(k)$  として結合される。

$$[Y(0) Y(1) \dots Y(319)] = [D_{LB}^w(0) D_{LB}^w(1) \dots D_{LB}^w(159) S_{HB}(0) S_{HB}(1) \dots S_{HB}(159)] \quad (65)$$

精度を最大とするためおよび固定小数点演算のオーバーフローを避けるため、信号  $d_{LB}^w(n)$  と  $s_{HB}(n)$  は MDCT の前に正規化されることに注意されたい。TDAC 符号器は、低域と高域の 2 つの MDCT に基づくため、2 つの正規化係数  $norm\_lo$  と  $norm\_hi$  が入力をスケールングするために計算される。しかし、ただ 1 つの正規化係数のみが 4 ビットで復号器に伝送される：

$$norm\_MDCT = \min(norm\_lo, norm\_hi) \quad (66)$$



#### 6. 6. 4 サブバンド分割

0-7000Hz 帯域のMDCT係数は、18のサブバンドに分割される。Table11/JT-G729.1は、サブバンドの境界と大きさを定義する。 $j$ 番目のサブバンドは、 $sb\_bound(j) \leq k < sb\_bound(j+1)$ である $nb\_coef(j)$ 個の係数 $Y(k)$ から成る。最初の17サブバンドは16個の係数(400Hz)から成り、最後のサブバンドは8個の係数(200Hz)からなる。

Table11/JT-G729.1 Subband boundaries and number of coefficients per subband in the TDAC coder  
(ITU-T G.729.1)

$j$	$sb\_bound(j)$	$nb\_coef(j)$
0	0	16
1	16	16
2	32	16
3	48	16
4	64	16
5	80	16
6	96	16
7	112	16
8	128	16
9	144	16
10	160	16
11	176	16
12	192	16
13	208	16
14	224	16
15	240	16
16	256	16
17	272	8
18	280	-

#### 6. 6. 5 スペクトル包絡計算

スペクトル包絡は、18サブバンドの対数領域における平均自乗根 (rms) として定義される：

$$\log\_rms(j) = \frac{1}{2} \log_2 \left[ \frac{1}{nb\_coef(j)} \sum_{k=sb\_bound(j)}^{sb\_bound(j+1)-1} Y(k)^2 + \varepsilon_{rms} \right], \quad j=0, \dots, 17 \quad (67)$$

ここで、 $\varepsilon_{rms} = 2^{-24}$ である。

#### 6. 6. 6 スペクトル包絡符号化

スペクトル包絡は、一様スカラー量子化により5ビットで量子化され、得られた量子化インデックスは2モードの2進数符号器を用いて符号化される。5ビット量子化は、以下のようにインデックス $rms\_index(j)$ 、 $j=0, \dots, 17$ 、を計算することである：

$$rms\_index(j) = round(2 \log\_rms(j)) \quad (68)$$

このとき、次のような制約がある。

$$-11 \leq rms\_index(j) \leq +20 \quad (69)$$

すなわち、インデックスは-11 から+20 (32 のとり得る値) の間に制限される。得られた量子化された全帯域の包絡は、2つのサブベクトルに分割される：

-低域スペクトル包絡： $(rms\_index(0), rms\_index(1), \dots, rms\_index(9))$

-高域スペクトル包絡： $(rms\_index(10), rms\_index(11), \dots, rms\_index(17))$

これらの2つのサブベクトルは、差分ハフマン符号化（モード0）と2進数符号化（モード1）を適応的に切替える2つのモードの可逆符号器を用いて独立に符号化される。差分ハフマン符号化は、平均ビット数を最小化するために用いられ、一方、2進数符号化は、差分ハフマン符号化により飽和された信号の包絡（例えば正弦波）を正しく符号化すると同時に、ビット数の最悪値を制限するために用いられる。スペクトル包絡復号器に対して選択されたモードを示すために1ビットが使われる。

### 低域包絡の符号化

差分ハフマン符号器（モード0）は、4つのステップからなる：

ステップ0：最初のインデックス  $rms\_index(0)$  の2進数符号化。 $[-11, +20]$ から成るインデックス  $rms\_index(0)$  は、5ビットで符号化される。

ステップ1：差分インデックスと飽和フラグ  $satur$  の計算。

$$diff\_index(j) = rms\_index(j) - rms\_index(j-1), \quad j = 1, \dots, 9 \quad (70)$$

$$satur = \begin{cases} 0 & \text{if } |diff\_index(j)| \leq 12 \text{ for } j = 1, \dots, 9 \\ 1 & \text{otherwise} \end{cases} \quad (71)$$

2進フラグ  $satur$  は、 $diff\_index(j)$  が $[-12, +12]$ の範囲外である場合を検出するために用いられる。

ステップ2： $satur = 0$  の場合は、差分インデックスに差分ハフマン符号化が適用され、 $mode$  が0に設定される。 $j = 1 \dots 9$  に対する差分インデックス  $diff\_index(j)$  は、テーブル参照によりハフマン符号化される。Table12/JT-G729.1 は、関連するハフマン符号と符号長を示す。

ステップ3： $satur = 1$  もしくは差分ハフマン符号化による使用ビット数が45を超える場合、ステップ0と同様な  $rms\_index(1) \dots rms\_index(9)$  の2進数符号化が適用され、 $mode$  が1に設定される。

Table12/JT-G729.1 Huffman codes used in the TDAC spectral envelope encoder (ITU-T G.729.1)

$diff\_index(j)$	Huffman code	Length (bits)	$diff\_index(j)$	Huffman code	length (bits)
-12	00000000110	11	1	111	3
-11	00000000111	11	2	1101	4
-10	00000010010	11	3	00011	5
-9	0000001000	10	4	000011	6
-8	000000101	9	5	0000010	7
-7	00000011	8	6	00000001	8
-6	0000011	7	7	000000000	9

-5	000010	6	8	0000000010	10
-4	00010	5	9	000000100111	12
-3	1100	4	10	0000001001100	13
-2	001	3	11	00000010011011	14
-1	01	2	12	00000010011010	14
0	10	2			

その結果、スペクトル包絡符号化モード（0もしくは1）の選択は、各モードで使用されるビット数と同様に2進フラグ *satur* の値に基づく。*satur*=1 もしくはモード0よりもモード1の消費ビット数が少ない場合は、モード1が選択される。それ以外の場合は、モード0が選択される。選択されたモードを復号器に知らせるために1ビットが使用される。通常、このビットは次のように設定される：0 → 差分ハフマン符号化、1 → 2進数符号化。

### 高域包絡の符号化

高域スペクトル包絡は同様な手法、すなわち差分ハフマン符号化と（直接）2進数符号化を切替えることで符号化される。ハフマンテーブルは同一（Table12/JT-G729.1）である。選択されたモードを復号器に知らせるために1ビットが使用される。

### 6. 6. 7 聴覚重要度によるサブバンド順序付け

各サブバンドの聴覚重要度  $ip(j)$ 、 $j=0...7$  は、以下のように定義される：

$$ip(j) = \frac{1}{2} \log_2 (rms\_q(j)^2 \times nb\_coef(j)) + offset \quad (72)$$

ここで、 $rms\_q(j) = 2^{1/2 rms\_index(j)}$  は量子化された平均自乗根であり、 $rms\_q(j)^2 \times nb\_coef(j)$  は量子化されたサブバンドエネルギーに相当する。従って、聴覚重要度は、サブバンド対数エネルギー（オフセットは置いておいて）と等価である。この情報は、次のような量子化されたスペクトル包絡である：

$$ip(j) = \frac{1}{2} [rms\_index(j) + \log_2(nb\_coef(j))] + offset \quad (73)$$

オフセット値は、 $ip(j)$  の式を更に簡単化するために導入される。 $offset = -2$  を用いることにより、聴覚重要度は以下ようになる：

$$ip(j) = \begin{cases} \frac{1}{2} rms\_index(j) & \text{for } j = 0, \dots, 16 \\ \frac{1}{2} (rms\_index(j) - 1) & \text{for } j = 17 \end{cases} \quad (74)$$

サブバンドは、聴覚重要度の降順に並べ替えられる。その結果、サブバンド  $j$  が  $(ord\_ip(j)+1)$  番目に大きい聴覚重要度であることを示す各サブバンドのインデックス  $0 \leq ord\_ip(j) < 18$ 、 $j=0, \dots, 17$  である。この順序付けは、ビット割当ておよびベクトル量子化インデックスの多重化のために用いられる。

### 6. 6. 8 分割球状ベクトル量子化のためのビット割当て

各サブバンドに割当てられるビット数は、TDA C復号器においても計算される聴覚重要度  $ip(j)$ 、 $j=0...17$

を用いることにより決定される。その結果、復号器は補助情報なしに同じ手法を実行することができる。

可能なビット割当てを Table13/JT-G729.1 に記載する。ここで、次元（8または16）は、各サブバンドにおけるMDCT係数の数に相当する。

Table13/JT-G729.1 Possible bit allocations for embedded spherical vector quantization (ITU-T G.729.1)

Dimension	Set of possible bit allocation (in bits)
8	$\mathbf{R}_8 = \{0, 7, 10, 12, 13, 14, 15, 16\}$
16	$\mathbf{R}_{16} = \{0, 9, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32\}$

合計のビット割当て量は、 $nbits\_VQ = 351 - nbits\_HB - nbits\_LB$  である。ここで、 $nbits\_LB$  および  $nbits\_HB$  は、それぞれ低域と高域のスペクトル包絡を符号化するために使われるビット数に相当する。各サブバンドに割当てられるビット数  $nbit(j)$ 、 $j=0, \dots, 17$  は、逆注水定理を使用した二分探索アルゴリズムを用いて Table13/JT-G729.1 の中から探索される。

二分探索アルゴリズムにより、以下のような“水位”  $\lambda_{opt}$  を見つけ出す。

$$\begin{cases} nbit(j) = \arg \min_{r \in \mathbf{R}_{nb\_coef(j)}} |nb\_coef(j) \times (ip(j) - \lambda_{opt}) - r| & j = 0, \dots, 17 \\ \sum_{j=0}^{17} nbit(j) \approx nbits\_VQ \end{cases} \quad (75)$$

ここで、 $\mathbf{R}_{nb\_coef(j)}$  は可能なビット割当てを含む。

$\lambda_{opt}$  の探索区間は、以下のように制限される：

$$\begin{cases} \lambda_0 = \max_{j=0, \dots, 17} (ip(j)) \\ \lambda_1 = \min_{j=0, \dots, 17} (ip(j)) - 4 \end{cases} \quad (76)$$

ここで、 $\lambda_0$  は零ビット割当てに相当し、 $\lambda_1$  は聴覚的に最も重要度の低いサブバンドに対してサンプルあたり4ビットのビット割当てに相当する。10回の繰返しの後、ビット割当ては次のように計算される：

$$nbit(j) = \arg \min_{r \in \mathbf{R}_{nb\_coef(j)}} |nb\_coef(j) \times (ip(j) - \lambda_{opt}) - r| \quad (77)$$

割当てられる合計のビット数は、ビット割当て量を超えることはない（適切に初期化された探索区間により）。しかし、ビット割当て量よりも少なくなる場合がある。この場合、残ったビット割当て量は、聴覚重要度の降順にさらに各サブバンドに分配される（この手順はインデックス  $ord\_ip(j)$  に基づく）。

### 6. 6. 9 MDCT係数の量子化

次元  $nb\_coef(j)$  の各サブバンド  $j=0, \dots, 17$  は、球面ベクトル量子化により  $nbit(j)$  ビットで符号化される。この手順は、2つのステップに分割される：最適コードベクトルの探索と選択されたコードベクトルのインデックス化である。基礎となるコードブックが次元  $nb\_coef(j)$  の単位球面上に配置されるコードベクトルを持つため、量子化は球面である。次元8においては、 $\mathbf{Q}_r^8, r \in \mathbf{R}_8$  かつ  $r > 0$  で記述される7つのコードブックがあり、次元16においては、 $\mathbf{Q}_r^{16}, r \in \mathbf{R}_{16}$  かつ  $r > 0$  で記述される19のコードブックがある。効率性の目的のために、コードブックは次のような特性を持つ：

- タイプII順列符号の結合である

- ちょうど大ききで埋め込まれる
- ほとんどのコードベクトルは零で補完された低次元コードベクトルからなる

### タイプII 順列符号

各コードブック  $\mathbf{Q}_{nb_{bit}(j)}^{nb_{coef}(j)}$  は、タイプII 順列符号の結合からなる。タイプII 順列符号においては、全ての極性の組合せを以下のようなリーダと呼ばれる次元  $N$  のコードベクトル  $\mathbf{y}_0$  の要素に順序を変えて設定することによりコードベクトルが生成される。

$$\mathbf{y}_0 = \left( \begin{array}{c} \xrightarrow{w_{lead}^0} \xrightarrow{w_{lead}^1} \xrightarrow{w_{lead}^{q-1}} \\ \zeta_0, \dots, \zeta_0, \zeta_1, \dots, \zeta_1, \dots, \zeta_{q-1}, \dots, \zeta_{q-1} \end{array} \right) \quad (78)$$

一般化する目的で、 $\mathbf{y}_0$  の成分を  $\{\zeta_0, \zeta_1, \dots, \zeta_{q-1}\}$  とする。ここで、 $q$  はアルファベットサイズである。重み  $w_{lead}^i$  (ここで、 $i=0\dots q-1$ ) は、 $\mathbf{y}_0$  における  $\zeta_i$  の反復回数である。重み  $w_{lead}^i$  は、 $\sum_{i=0}^{q-1} w_{lead}^i = N$  を満たす正の整数である。通常、 $\mathbf{y}_0$  の要素は、位置 0 から  $N-1$  の降順で保存される： $\zeta_0 > \zeta_1 > \dots > \zeta_{q-1} \geq 0$ 。一般に、順列  $M$  の総数は、以下で与えられる：

$$M = 2^{nb\_sign(\mathbf{y}_0)} \frac{N!}{\prod_{i=0}^{q-1} w_{lead}^i!} \quad (79)$$

ここで、 $nb\_sign(\mathbf{y}_0)$  は、 $\mathbf{y}_0$  の非零要素数である。

例えば、次元  $N = nb\_coef(17) = 8$  のベクトル量子化器のリーダを列挙する Table14/JT-G729.1 の中の最初のリーダ  $\mathbf{y}_0 = (1,0,0,0,0,0,0,0)$  は、次のような特性を持つ： $q=2$ 、 $w_{lead}^0=1$ 、 $w_{lead}^1=7$ 、 $nb\_sign(\mathbf{y}_0)=1$  かつ  $\mathbf{y}_0$  の  $M=16$  極性付順列がある。

コードブックは、そのリーダとして定義される。次元 8 における 7 つのコードブックの (非正規化) リーダは、Table14/JT-G729.1 に列挙される。例えば、コードブック  $\mathbf{Q}^7$  (7 ビット) は、2 つのリーダ  $(1,0,0,0,0,0,0,0)$  と  $(1,1,0,0,0,0,0,0)$  の全ての極性付組み合わせを含む。全部で 33 のリーダがある。実際は、これらのリーダは正規化される。すなわち、単位球面上に配置させるために自身の大ききの逆数でスケールされる。

### ちょうど大ききで埋め込まれたコードブック

2 つのサブバンド次元 (8 と 16) の各々に対して、コードブックはちょうど大ききで埋め込まれる：

$$\begin{aligned} \mathbf{Q}_7^8 &< \mathbf{Q}_0^8 < \mathbf{Q}_2^8 < \mathbf{Q}_3^8 < \mathbf{Q}_4^8 < \mathbf{Q}_5^8 < \mathbf{Q}_6^8 \\ \mathbf{Q}_6^6 &< \mathbf{Q}_4^6 < \mathbf{Q}_6^6 < \dots < \mathbf{Q}_9^6 < \mathbf{Q}_{30}^6 < \mathbf{Q}_{31}^6 < \mathbf{Q}_{32}^6 \end{aligned}$$

例えば、コードブック  $\mathbf{Q}_0^8$  (10 ビット) は、3 つの追加リーダの全ての極性付順列とコードブック  $\mathbf{Q}^7$  (7 ビット) から成る。低分解能 (サイズ) コードブックの全てのリーダは、高分解能 (サイズ) の全てのコードブックのリーダである。例えば、Table 1 4/JT-G729.1 において、 $\mathbf{Q}^7$  の 2 つのリーダは、 $\mathbf{Q}_0^8, \mathbf{Q}_2^8, \dots, \mathbf{Q}_6^8$  のリーダでもある。

### 低次元辞書

リーダは、MDC T 係数ベクトル上での学習手続きにより構成される。どのサブバンド次元に対しても、通常、リーダは次元よりも少ない非零要素を持つ。すなわち、 $1 \leq nb\_sign(\mathbf{y}_0) \leq nb\_coef(j)$  となる。 $nb\_sign(\mathbf{y}_0)$  で与えられる非零要素数は、1 から  $nb\_coef(j)$  の間で変化する。各リーダの零要素は保存され

ない；代わりに、リーダーは、次元  $nb\_sign(\mathbf{y}_0)$  (1 から 16) の 16 辞書に分類されて保存される。低次元  $nb\_sign(\mathbf{y}_0)$  の辞書からのコードベクトルの最後部に  $nb\_coef(j) - nb\_sign(\mathbf{y}_0)$  個の零を挿入することにより、 $\mathbf{y}_0$  のリーダーは次元  $nb\_coef(j)$  に拡張される。

例えば、コードブック  $\mathbf{Q}_2^8$  は、11 の順列符号の集合である。1 つのリーダー (1,0,0,0,0,0,0) は、7 つの零が挿入された次元 1 の辞書のリーダー (1) から構成され、6 つのリーダーは、6 つの零が挿入された次元 2 の辞書の 6 つのリーダー { (1,1), (2,1), (5,1), (4,3), (7,5), (3,2) } から構成され、1 つのリーダー (1,1,1,0,0,0,0) は、5 つの零が挿入された次元 3 の辞書のリーダー (1,1,1) から構成され、等々。ただ 1 つのリーダー (1,1,1,1,1,1,1) は次元 8 である。このリーダーは、次元 8 の辞書に保存され、また 8 つの零を挿入することにより 1 つの 16 次元リーダーを構成するためにも使われる。

Table14/JT-G729.1 List of leaders in codebooks of demension 8 (33 leaders)  
(ITU-T G.729.1)

Non-normalized leader $\mathbf{y}_0$	$nb\_sign(\mathbf{y}_0)$	$offset\_lead(\mathbf{y}_0)$	$\mathbf{Q}_1^8$	$\mathbf{Q}_0^8$	$\mathbf{Q}_2^8$	$\mathbf{Q}_3^8$	$\mathbf{Q}_4^8$	$\mathbf{Q}_5^8$	$\mathbf{Q}_6^8$
1 0 0 0 0 0 0 0	1	0	×	×	×	×	×	×	×
1 1 0 0 0 0 0 0	2	16	×	×	×	×	×	×	×
2 1 0 0 0 0 0 0	2	128		×	×	×	×	×	×
5 1 0 0 0 0 0 0	2	352		×	×	×	×	×	×
1 1 1 0 0 0 0 0	3	576		×	×	×	×	×	×
4 3 0 0 0 0 0 0	2	1024			×	×	×	×	×
7 5 0 0 0 0 0 0	2	1248			×	×	×	×	×
3 2 0 0 0 0 0 0	2	1472			×	×	×	×	×
1 1 1 1 1 1 1 1	8	1696			×	×	×	×	×
1 1 1 1 0 0 0 0	4	1952			×	×	×	×	×
1 1 1 1 1 1 1 0	7	3072			×	×	×	×	×
3 1 1 1 1 1 1 1	8	4096				×	×	×	×
4 1 1 1 1 1 1 1	8	6144				×	×	×	×
4 1 0 0 0 0 0 0	2	8192					×	×	×
5 2 0 0 0 0 0 0	2	8416					×	×	×
5 3 0 0 0 0 0 0	2	8640					×	×	×
6 1 0 0 0 0 0 0	2	8864					×	×	×
3 3 1 1 1 1 1 1	8	9088					×	×	×
2 2 1 0 0 0 0 0	3	16256						×	×
1 1 1 1 1 0 0 0	5	17600						×	×
1 1 1 1 1 1 0 0	6	19392						×	×
2 2 2 2 2 2 2 1	8	21184						×	×
5 4 3 0 0 0 0 0	3	23232						×	×
5 5 3 3 0 0 0 0	4	25920						×	×
5 4 0 0 0 0 0 0	2	32640							×
8 1 0 0 0 0 0 0	2	32864							×

5	2	2	2	2	2	2	2	2	8	33088									×
7	1	1	1	1	1	1	1	1	8	35136									×
8	2	1	0	0	0	0	0	0	3	37184									×
7	5	2	0	0	0	0	0	0	3	39872									×
6	6	1	1	0	0	0	0	0	4	42560									×
5	2	2	2	2	2	2	2	2	8	49280									×
2	1	1	1	1	0	0	0	0	5	56448									×

次元 16 においては、64 の可能なリーダー  $\mathbf{y}_0$  と 19 の非零ビット割当てがある。このリーダーとビット割当てが多いため、これらのリーダーはここには記載されていない。しかし、原理は次元 8 の場合と同じである。

### 最適コードベクトルの探索

非零ビット割当て  $nb_{ii}(j)$  の各サブバンド  $j=0, \dots, 17$  は、順列符号の集合によって符号化される。関連するコードブック  $\mathbf{Q}_{nb_{ii}(j)}^{nb\_coef(j)}$  が球面であるため、結局は、最適コードベクトルの探索はサブバンド係数とコードベクトルとの内積を最大化することになる。基礎となるタイプ II 順列符号の構造と低次元辞書を利用することにより、探索はさらに最適化される。

$\mathbf{Q}_{nb_{ii}(j)}^{nb\_coef(j)}$  の中で選択されたベクトル  $\mathbf{y} = (y_0, \dots, y_{N-1})$  は、リーダー  $\mathbf{y}_0$  の順列である。

### 選択されたコードベクトルのインデックス化

$\mathbf{y}$  のインデックスは次のように計算される：

$$index(\mathbf{y}) = offset\_lead(\mathbf{y}_0) + rank(\mathbf{y}|\mathbf{y}_0) \times 2^{nb\_sign(\mathbf{y}_0)} + sign\_bits(\mathbf{y}) \quad (80)$$

ここで、

- $offset\_lead(\mathbf{y}_0)$  は、 $\mathbf{y}_0$  によって定義される順列符号の基点となるオフセットである
- $rank(\mathbf{y}|\mathbf{y}_0)$  は、その順列符号の中の  $\mathbf{y}$  の順列階数である
- $nb\_sign(\mathbf{y}_0)$  は、 $\mathbf{y}_0$  の非零要素数である
- $sign\_bits(\mathbf{y})$  は、 $\mathbf{y}$  の非零要素の印を表す LSB を含む整数である

基点となるオフセットはテーブル参照により求められ、さらに順列階数は Schalkwijk の公式により計算される。後者は辞書式順序を仮定する。 $\mathbf{y}$  の階数の計算は、 $\mathbf{y}$  に関連するベクトル  $\mathbf{D} = (d_0, d_1, \dots, d_{n-1})$  の階数の計算と等価である。ここで、 $y_k = \zeta_{d_k}$  の場合に限り  $d_k = d$  となる。 $\mathbf{y}$  と  $\mathbf{D}$  の階数は同一である。しかし、 $\mathbf{D}$  の定義は、階数計算を  $\{0, 1, \dots, q-1\}$  の中の値を持つ列  $\mathbf{D}$  の場合に低下させることを許す。

$\mathbf{y}$  と  $\mathbf{D}$  の重みは同一である。部分的な重み  $(w_{lead,k}^0, w_{lead,k}^1, \dots, w_{lead,k}^{q-1})$  は、下式のように重み  $(y_k, y_{k+1}, \dots, y_{N-1})$  として定義される。

$$w_{lead,k}^d = \sum_{i=k}^{N-1} \delta(y_i, a_d) = \sum_{i=k}^{N-1} \delta(d_i, d) \quad (81)$$

ここで、 $x=y$  であれば  $\delta(x,y)=1$ 、そうでなければ  $\delta(x,y)=0$  である。

定義により： $(w_{lead,0}^0, w_{lead,0}^1, \dots, w_{lead,0}^{q-1}) = (w_{lead}^0, w_{lead}^1, \dots, w_{lead}^{q-1})$  となる。

$\mathbf{y}$  の階数  $rank(\mathbf{y}|\mathbf{y}_0)$  は、次の組み合わせ表現式を用いて計算される。

$$\text{rank}(\mathbf{y}|\mathbf{y}_0) = \sum_{k=0}^{N-1} I_k^{d_k} \quad \text{ここで} \quad I_k^{d_k} = \frac{(N-1-k)!}{\prod_{i=0}^{q-1} (w_{lead,k}^i)!} \left( \sum_{d=0}^{d_k-1} w_{lead,k}^d \right) \quad (82)$$

通常、 $\sum_{d=0}^{-1} = 0$  かつ  $(-1)! = \infty$  である。

項  $I_k^{d_k}$  は階数  $k$  の部分階数である。 $I_k^{d_k}$  は、 $k = N-1$  から  $k = 0$  の全ての  $k$  位置に対して計算される。部分階数の実際の計算は、それらの素因数分解に依存する。この素因数分解は 32 ビット固定小数点演算の Schalkwijk の公式の実装のために必須である。 $I_k^{d_k}$  の素因数分解は、全ての部分項  $(N-1-k)!$ 、 $\sum_{d=0}^{d_k-1} w_{lead,k}^d$  および  $\prod_{i=0}^{q-1} (w_{lead,k}^i)!$  の素因数分解から計算される。整数  $\{k\}_{k=1,\dots,16}$  と  $\{k!\}_{k=0,\dots,16}$  の素因数分解はメモリに保存される。このようにして最初の 2 つの部分項 ( $I_k^{d_k}$  の分子項:  $(N-1-k)!$ 、 $\sum_{d=0}^{d_k-1} w_{lead,k}^d$ ) の素因数分解は、メモリから直接読み込まれる。最終項の素因数分解は、次のような再帰式を用いて計算される:

$$\prod_{i=0}^{q-1} (w_{lead,k}^i)! = w_{lead,k}^j \times \prod_{i=0}^{q-1} (w_{lead,k+1}^i!)$$

$I_k^{d_k}$  の分母 ( $\prod_{i=0}^{q-1} (w_{lead,k}^i)!$ ) の素因数分解は、素因数分解  $w_{lead,k}^j$  の指数を部分項  $\prod_{i=0}^{q-1} (w_{lead,k+1}^i!)$  の素因数分解の指数に加えることにより計算される。これら 3 つの部分項の素因数分解から、 $I_k^{d_k}$  の素因数分解は、最初の 2 つ (分子項) の指数を加え、最後の 1 つ (分母項) の指数を減ずることにより計算される。最後に、 $I_k^{d_k}$  はその素因数分解から計算される。

## 6. 6. 10 TDACパラメータの多重化

高域スペクトル包絡符号化に関連するビットは、低域スペクトル包絡符号化に関連するビットの前に多重化される。さらに、サブバンド量子化インデックスは、聴覚重要度の降順に多重化される。聴覚的に、より重要なサブバンド (すなわち、最も聴覚重要度の大きい  $ip(j)$ ) は、ビット列の最初に書き込まれる。

その結果、復号器において符号化スペクトル包絡のちょうどその部分が受信されれば、低域のその前に高域包絡の復号が可能である。この特性は、TDAC復号器において高域MDC Tスペクトルの部分レベル調整を実行するために用いられる。

## 7. 復号器の機能説明

復号処理は受信されるビットレートに依存している。8kbit/s、12kbit/s では CELP での復号、14kbit/s では CELP と TDBWE での復号、14kbit/s を越えるレートでは CELP、TDBWE、及び TDAC での復号という構成になっている。送信されるパラメータについては 8 章で述べる。

### 7. 1 狭帯域エンベデッド CELP 復号器 (レイヤ 1 および 2)

#### 7. 1. 1 LP フィルタパラメータの復号

4. 1. 1 節 / JT-G 7 2 9 と同一である。

#### 7. 1. 2 パリティビットの計算

4. 1. 2 節 / JT-G 7 2 9 と同一である。

#### 7. 1. 3 適応コードブックベクトルの復号

4. 1. 3 節 / JT-G 7 2 9 と同一である。



#### 7. 1. 4 8kbit/s 固定コードブックベクトルの復号

4. 1. 4 節/JT-G729 と同一である。

#### 7. 1. 5 8kbit/s 適応および固定コードブック利得の復号

4. 1. 5 節/JT-G729 と同一である。

#### 7. 1. 6 12kbit/s 固定コードブックベクトルの復号

5ms のサブフレーム毎に、パターン係数  $\alpha_{enh}$  がエンコーダで行われたのと同じ方法で最初に計算される。それから、4組のトライパルス  $-\alpha_{enh}z^{-1}+1-\alpha_{enh}z$  の位置  $\{m'_0, m'_1, m'_2, m'_3\}$  と極性  $\{s'_0, s'_1, s'_2, s'_3\}$  を抽出するために、それぞれ受信したコードブックインデックス、及び極性符号語が使われる。これは 8kbit/s 固定コードブックベクトルの復号処理と同様の方法で行われる。一度、位置と極性が復号されると、各トライパルスパターンは、中心位置と極性に従って配置され、固定コードブックベクトル  $c'(n)$  は式(13)のように 4 組を加算する事によって構築される。

#### 7. 1. 7 12kbit/s 固定コードブック利得の復号

12kbit/s のコードブック利得  $\hat{g}_{enh}$  は、第1サブフレームが 3 ビット、第2サブフレームが 2 ビットのスカラ逆量子化器を使って、6. 3. 10 節の逆の手順によって復号される。

#### 7. 1. 8 8kbit/s または 12kbit/s における再生音声の計算

8kbit/s は 4. 1. 6 節/JT-G729 と同一である。12kbit/s においては、8kbit/s の励振信号  $u(n)$  と 12kbit/s で追加された固定コードブック励振信号  $\hat{g}_{enh} \times c'(n)$  が式(83)のように結合されて、LP 合成フィルタ  $\hat{A}(z)$  の入力として使われる。

$$u_{enh}(n) = u(n) + \hat{g}_{enh} \times c'(n) \quad (83)$$

サブフレームの再生音声は式(84)によって与えられる。

$$\hat{s}_{enh}(n) = u_{enh}(n) - \sum_{i=1}^{10} \hat{a}_i \hat{s}_{enh}(n-i), \quad n=0, \dots, 39 \quad (84)$$

ここで、 $\hat{a}_i$  は現サブフレームの  $\hat{A}(z)$  の補間された LP フィルタ係数である。

### 7. 2 TDBWE 復号器 (レイヤ3)

TDBWE 復号器モジュールの原理を Figure8/JT-G729.1 に示す。7. 2. 1 節のパラメータ抽出手順によって計算された TDBWE 受信パラメータは、所望の時間包絡、周波数包絡である  $\hat{T}_{env}(i)$ 、 $\hat{F}_{env}(j)$  に従って、人工的に生成された励振信号  $\hat{s}_{HB}^{exc}(n)$  を形成するために使われる。これに続いて時間領域での後処理が行われる。

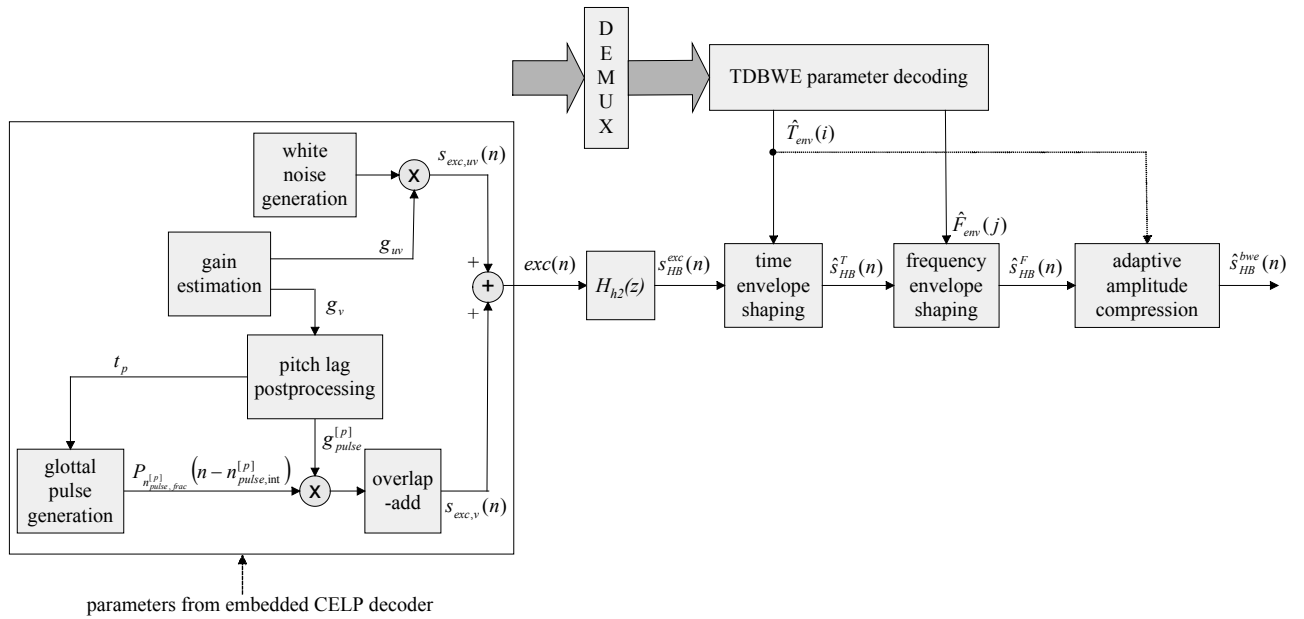


Figure8/JT-G729.1 High-level block diagram of the TDBWE decoder (ITU-T G.729.1)

### 7. 2. 1 パラメータ復号処理

量子化パラメータセットは、 $\hat{M}_T$  という値と、 $\hat{T}_{env,1}$ 、 $\hat{T}_{env,2}$ 、 $\hat{F}_{env,1}$ 、 $\hat{F}_{env,2}$ 、 $\hat{F}_{env,3}$  というベクトルによって構成される。分割ベクトルは式(49)、(50)によって定義される。

量子化された平均時間包絡  $\hat{M}_T$  は、式(85)(86)のように、個々のベクトル要素から時間包絡と周波数包絡パラメータを再構築するために使われる。

$$\hat{T}_{env}(i) = \hat{T}_{env}^M(i) + \hat{M}_T, \quad i = 0, \dots, 15 \quad (85)$$

$$\hat{F}_{env}(j) = \hat{F}_{env}^M(j) + \hat{M}_T, \quad j = 0, \dots, 11 \quad (86)$$

### 7. 2. 2 励振信号の生成

TDBWE励振信号  $exc(n)$  は、ビットストリームのレイヤ1、2の中で送られるパラメータに基づいて、5msのサブフレーム毎に生成される。具体的には以下のパラメータが使われる：サブフレームに依存した整数ピッチラグ  $T_0 = \text{int}(T_1) \text{ or } \text{int}(T_2)$ 、分数ピッチラグ  $frac$ 、固定コードブック寄与分エネルギー  $E_c = \sum_{n=0}^{39} (\hat{g}_c \cdot c(n) + \hat{g}_{enh} \cdot c'(n))^2$ 、適応コードブック寄与分エネルギー  $E_p = \sum_{n=0}^{39} (\hat{g}_p \cdot v(n))^2$ 。

励振信号生成のパラメータは5msサブフレーム毎に計算される。励振信号生成は以下の手順からなる。

- (1)最終励振信号  $exc(n)$  への有声音と無声音寄与分に対応する2つの利得  $g_v$  と  $g_{wn}$  の推定。
- (2)ピッチラグ後処理
- (3)有声音寄与分の生成
- (4)無声音寄与分の生成
- (5)低域通過フィルタ処理

#### 有声音/無声音寄与分に対する利得推定

適応、及び固定コードブック励振信号（追加されたコードブックを含む）のエネルギー比は、サブフレーム毎に以下の式によって計算される。

$$\xi = \frac{E_p}{E_c} \quad (87)$$

無声音時のこの比 $\xi$ を小さくするため、下記の Wiener フィルタ特性が適用される。

$$\xi_{post} = \xi \cdot \frac{\xi}{1 + \xi} \quad (88)$$

これによって、より一貫性のある無声音となる。 $exc(n)$ の有声音、及び無声音寄与分の利得は、以下の手順で決定される。まず、中間の有声音利得 $g'_v$ が式(89)によって計算される。

$$g'_v = \sqrt{\frac{\xi_{post}}{1 + \xi_{post}}} \quad (89)$$

続いて、最終有声音利得を得るためにわずかに平滑化される。

$$g_v = \sqrt{\frac{1}{2} \left( g_v'^2 + g_{v,old}'^2 \right)} \quad (90)$$

ここで、 $g'_{v,old}$ は前サブフレームの $g'_v$ の値である。

$g_v^2 + g_{uv}^2 = 1$ という制約を満たすため、無声音利得が式(91)で与えられる。

$$g_{uv} = \sqrt{1 - g_v^2} \quad (91)$$

### ピッチラグ後処理

励振信号 $exc(n)$ の中で一貫性のあるピッチ構造を生成するためには、音声生成プロセスにおける基本ピッチラグの適切な推定が必要である。ビットストリームのレイヤ1の中では、整数、及び分数ピッチラグ $T_0$ 、及び $frac$ が、現スーパーフレームの4つの5msサブフレームで利用できる。各サブフレームにおける $t_0$ の推定は、これらのパラメータに基づく。

J T - G 7 2 9 符号器側でのピッチ探索手順の目的は、L T P 残差信号のパワーを最小にするピッチラグを見つける事である。つまり、L T P ピッチラグは必ずしも $t_0$ と一致しなくても良いが、有声音の成分を簡潔に再生するためには一致することが必要条件となる。最も典型的なピッチ推定の逸脱は、倍ピッチ推定や半ピッチ推定といった推定エラー、すなわち、L T P ラグに対応する周波数が、元の基本音声周波数の半分、または倍になるという事である。とくに、ピッチが倍になる（または3倍など）誤りは厳密に避けなければならない。それゆえ、以下のL T P ラグ情報の後処理が使われる。

最初に、オーバーサンプルされたタイムスケールに対応したL T P ピッチラグが $T_0$ 、 $frac$ から再生され、帯域拡張係数2が考慮される。

$$t_{LTP} = 2 \cdot (3 \cdot T_0 + frac) \quad (92)$$

現フレームの観測されたL T P ラグ $t_{LTP}$ と、前サブフレームの後処理されたピッチラグ $t_{post,old}$ の間の（整数化された）係数が式(93)によって計算される。

$$f = \text{int} \left( \frac{t_{LTP}}{t_{post,old}} + 0.5 \right) \quad (93)$$

もし、係数 $f$ が2、...、4に入るなら、下記の相対誤差が評定される。

$$e = 1 - \frac{t_{LTP}}{f \cdot t_{post,old}} \quad (94)$$

この相対誤差の大きさが閾値  $\varepsilon = 0.1$  より小さい場合は、現在の L T P ラグは、ピッチが倍（または 3 倍など）になるエラーが始まった結果と見なされる。それゆえに、ピッチラグは整数係数  $f$  によって除算される事で補正され、それによって前のピッチラグに対して連続的なピッチラグのふるまいを実現する事ができる。

$$t_{post} = \begin{cases} \text{int}\left(\frac{t_{LTP}}{f} + 0.5\right) & |e| < \varepsilon, f > 1, f < 5 \\ t_{LTP} & \text{otherwise} \end{cases} \quad (95)$$

このピッチラグは、さらに式(96)のように平滑化される。

$$t_p = \frac{1}{2} \cdot (t_{post,old} + t_{post}) \quad (96)$$

この移動平均によって、実質的な精度はサンプルの 1/3 から 1/6 に向上する。

最後に、後処理されたピッチラグ  $t_p$  は整数部と分数部に分解される。

$$t_{0,int} = \text{int}\left(\frac{t_p}{6}\right) \quad \text{and} \quad t_{0,frac} = t_p - 6 \cdot t_{0,int} \cdot \quad (97)$$

#### 有声音寄与分の生成

T D B W E 励振信号の有声音成分  $s_{exc,v}(n)$  は整形され重み付けられた声門パルスとして表現される。それゆえに、 $s_{exc,v}(n)$  は単一パルス寄与分のオーバーラップ加算によって生成される。

$$s_{exc,v}(n) = \sum_p g_{Pulse}^{[p]} \times P_{n_{Pulse,frac}^{[p]}}(n - n_{Pulse,int}^{[p]}) \quad (98)$$

ここで、 $n_{Pulse,int}^{[p]}$  は各パルスのパルス位置、 $P_{n_{Pulse,frac}^{[p]}}(n - n_{Pulse,int}^{[p]})$  は各パルスのパルス形状、 $g_{Pulse}^{[p]}$  は各パルスの利得係数である。これらのパラメータは以下のように求められる。

後処理されたピッチラグ  $t_{0,int}$  と  $t_{0,frac}$  は、式(99)に従ってパルス間隔、つまりパルス位置を決定する。

$$n_{Pulse,int}^{[p]} = n_{Pulse,int}^{[p-1]} + t_{0,int} + \text{int}\left(\frac{n_{Pulse,frac}^{[p-1]} + t_{0,frac}}{6}\right), \quad (99)$$

ここで、 $p$  はパルスカウンタである。つまり、 $n_{Pulse,int}^{[p]}$  は現在のパルスの（整数）位置であり、 $n_{Pulse,int}^{[p-1]}$  は前パルスの（整数）位置である。

パルス位置の分数部である式(100)は、パルス形状選択のインデックスとなる。

$$n_{Pulse,frac}^{[p]} = n_{Pulse,frac}^{[p-1]} + t_{0,frac} - 6 \cdot \text{int}\left(\frac{n_{Pulse,frac}^{[p-1]} + t_{0,frac}}{6}\right) \quad (100)$$

プロトタイプパルス形状  $P_i(n)$  ( $i=0, \dots, 5, n=0, \dots, 56$ ) は Figure9/JT-G729.1 にプロットされたルックアップテーブルから取得される。これらのパルス形状は、あるスペクトル形状、つまり高域に向かって有声音の励振信号成分が滑らかに減衰し、最後には収束するように、またピッチラグ情報の全サブサンプル解像度が利用されるように設計される。さらに、励振信号の頂点が強く抑制され、主観品質が改善される。各パルス  $g_{Pulse}^{[p]}$

は有声音利得パラメータ  $g_v$  と、ピッチラグパラメータから算出される。

$$g_{Pulse}^{[p]} = (2 \cdot \text{even}(n_{Pulse,int}^{[p]}) - 1) \cdot g_v \cdot \sqrt{6t_{0,int} + t_{0,frac}} \quad (101)$$

ここで、パルス間隔が増加する事で、含まれるエネルギーが減少しない事が保証される。even() 関数は、引数が偶数なら 1 を返し、それ以外の時は 0 を返す。

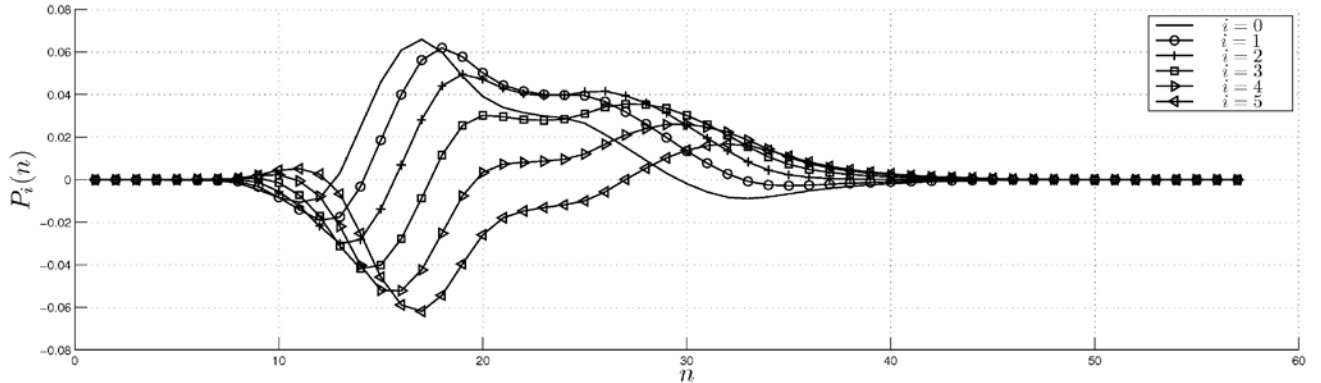


Figure9/JT-G729.1 Pulse shape lookup table  
(ITU-T G.729.1)

### 無声音寄与分の生成

無声音寄与分  $s_{exc,uv}(n)$  は、白色雑音発生器のスケーリングされた出力を使って生成される。

$$s_{exc,uv}(n) = g_{uv} \cdot \text{random}(n), \quad n=0, \dots, 39. \quad (102)$$

乱数発生器は、4. 4. 4 節/JT-729 と同一であり、単位分散を持つ信号を出力する。

### 低域通過フィルタ処理

有声音寄与分  $s_{exc,v}(n)$  と無声音寄与分  $s_{exc,uv}(n)$  を使って、最終的な励振信号  $s_{HB}^{exc}(n)$  は  $exc(n) = s_{exc,v}(n) + s_{exc,uv}(n)$  に対し低域通過フィルタ処理を行うことによって得られる。

低域通過フィルタは 3000Hz のカットオフ周波数を持ち、その実装は 6. 4 節で記述されるような高域信号に対する前処理としての低域通過フィルタと同一である。

### 7. 2. 3 時間包絡形成処理

励振信号  $s_{HB}^{exc}(n)$  の時間包絡の形成は、復号された時間包絡パラメータ  $\hat{T}_{env}(i)$  ( $i=0, \dots, 15$ ) を使って、符号器側の高域信号  $s_{HB}(n)$  の時間包絡とはほぼ一致する時間包絡  $\hat{s}_{HB}^T(n)$  を得る。これは単純なスカラ乗算によって実行される。

$$\hat{s}_{HB}^T(n) = g_T(n) \cdot s_{HB}^{exc}(n), \quad n=0, \dots, 159 \quad (103)$$

利得関数  $g_T(n)$  を決定するため、励振信号  $s_{HB}^{exc}(n)$  は符号器でのパラメータ抽出として 6. 5. 1 節に記述されているのと同じ方法で分割され、分析される。得られた分析結果は、再び時間包絡パラメータ  $\tilde{T}_{env}(i)$  ( $i=0, \dots, 15$ ) となる。これらは  $s_{HB}^{exc}(n)$  の観測された時間包絡を記述する。そして予備の利得係数が式(104)のように計算される。

$$g_T^i(i) = 2^{\hat{T}_{env}(i) - \tilde{T}_{env}(i)}, \quad i = 0, \dots, 15 \quad (104)$$

インデックス  $i=0, \dots, 15$  の各区間において、これらの利得係数は“フラットトップ”ハンニング窓を使って補間される。

$$w_t(n) = \begin{cases} \frac{1}{2} \cdot \left[ 1 - \cos\left( (n+1) \cdot \frac{\pi}{6} \right) \right] & n = 0, \dots, 4 \\ 1 & n = 5, \dots, 9 \\ \frac{1}{2} \cdot \left[ 1 - \cos\left( (n+9) \cdot \frac{\pi}{6} \right) \right] & n = 10, \dots, 14 \end{cases} \quad (105)$$

この窓関数  $w_t(n)$  は Figure10/JT-G729.1 に示される。この補間処理は最終的に以下の所望の利得関数を導き出す。

$$g_T(n+i \cdot 10) = \begin{cases} w_t(n) \cdot g'_T(i) + w_t(n+10) \cdot g'_T(i-1) & n = 0, \dots, 4 \\ w_t(n) \cdot g'_T(i) & n = 5, \dots, 9 \end{cases} \quad (106)$$

ここで、 $g'_T(-1)$  は前スーパーフレームの最後の 1.25ms 区間からの保存された利得係数  $g'_T(15)$  として定義される。

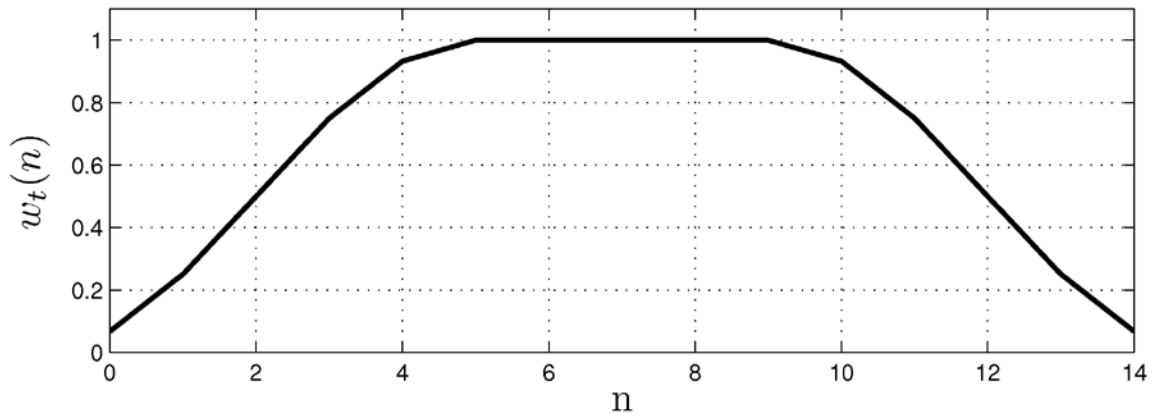


Figure10/JT-G729.1 “Flat-top” Hanning window for the time envelope shaping  
(ITU-T G.729.1)

#### 7. 2. 4 周波数包絡整形

復号された周波数包絡パラメータ  $\hat{F}_{env}(j)$  ( $j=0, \dots, 11$ ) は 20ms のスーパーフレーム中の 2 番目の 10ms フレームの代表値である。最初の 10ms のフレームは、現在のパラメータセットと前スーパーフレームのパラメータセット  $\hat{F}_{env,old}(j)$  との補間によって補われる。

$$\hat{F}_{env,int}(j) = \frac{1}{2} \left( \hat{F}_{env,old}(j) + \hat{F}_{env}(j) \right) \quad j = 0, \dots, 11 \quad (107)$$

$\hat{s}_{HB}^T(n)$  のスーパーフレームは、6. 5. 2 節の記述に従ってスーパーフレーム毎に 2 回ずつ分析される。これは現スーパーフレーム中の 1 番目 ( $l=1$ )、及び 2 番目 ( $l=2$ ) で行われ、2 つの観測された周波数包絡パラメータセット  $\tilde{F}_{env,l}(j)$  ( $j=0, \dots, 11$ 、フレームインデックスとして  $l=1, 2$ ) を生成する。ここで、式(108)のように、サブバンド毎の補正利得係数が第 1、第 2 フレームにおいて決定される。

$$G_{F,1}(j) = 2^{\hat{F}_{env,int}(j) - \tilde{F}_{env,1}(j)} \quad \text{and} \quad G_{F,2}(i) = 2^{\hat{F}_{env}(j) - \tilde{F}_{env,2}(j)}, \quad j = 0, \dots, 11. \quad (108)$$

これらの利得はフィルタバンクイコライザのチャネルを制御する。フィルタバンクイコライザは、個々のチャネルが Table 9/JT-G729.1 で与えられるサブバンド分割に適合するように設計される。それはフィルタのインパルス応答  $h_F^{(i)}(n)$  ( $i=0, \dots, 11, n=0, \dots, 32$ ) と補完された高域寄与分  $h_{HP}(n)$  ( $n=0, \dots, 32$ ) として定義され、 $h_F^{(i)}(n)$  と  $h_{HP}(n)$  はそれぞれ群遅延 2ms (16 サンプル) を持つ FIR フィルタを構成する。この遅延は符号器側でのパラメータ抽出 (6.5.2 節参照) によって生じた先読みに正確に一致する。それぞれのフィルタバンク設計における周波数応答は、Figure11/JT-G729.1 に表される。

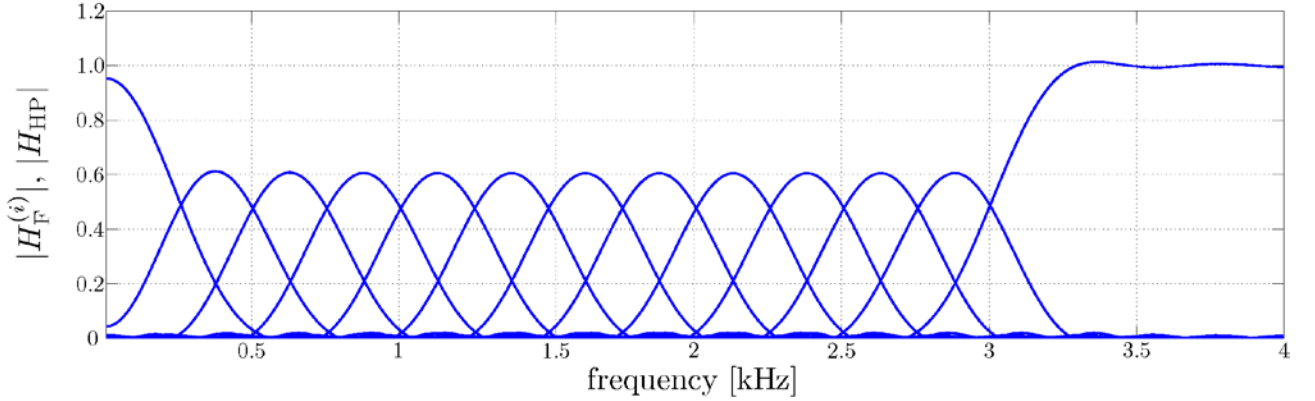


Figure11/JT-G729.1 Filter-bank design for the frequency envelope shaping (ITU-T G.729.1)

周波数包絡の整形を実現するために、2つの FIR フィルタがスーパーフレーム毎に構築される。

$$h_{F,l}(n) = \sum_{i=0}^{11} G_{F,l}(i) \cdot h_F^{(i)}(n) + 0.1 \cdot h_{HP}(n), \quad n = 0, \dots, 32, \quad l = 1, 2 \quad (109)$$

これらの2つの FIR 補正フィルタは、信号  $\hat{s}_{HB}^T(n)$  に適用されて  $\hat{s}_{HB}^F(n)$  を得る。

$$\hat{s}_{HB}^F(n) = \sum_{m=0}^{32} \hat{s}_{HB}^T(n-m) h_{F,1}(m), \quad n = 0, \dots, 79 \quad (110)$$

$$\hat{s}_{HB}^F(n) = \sum_{m=0}^{32} \hat{s}_{HB}^T(n-m) h_{F,2}(m), \quad n = 80, \dots, 159 \quad (111)$$

### 7. 2. 5 適応振幅圧縮による時間領域後処理

信号  $\hat{s}_{HB}^F(n)$  は、所望の時間包絡、及び周波数包絡に従って、(CELP 復号器によって低域で推定されたパラメータから生成された) 励振信号  $s_{HB}^{exc}(n)$  を整形することによって得られた。一般的には、この励振信号と、関連する包絡形状  $\hat{T}_{env}(i)$ 、 $\hat{F}_{env}(j)$  とが結合されることはない。結果として、いくらかのクリックノイズが信号  $\hat{s}_{HB}^F(n)$  に現れるかもしれない。このノイズを抑制するため、適応振幅圧縮が  $\hat{s}_{HB}^F(n)$  に用いられる。i 番目の 1.25ms 区間の  $\hat{s}_{HB}^F(n)$  における各サンプルは、復号された時間包絡  $\hat{T}_{env}(i)$  と比較され、 $\hat{s}_{HB}^F(n)$  の振幅は、この包絡との大きな偏差を抑制するために圧縮される。

圧縮関数は Figure12/JT-G729.1 に示される。 $\sigma$  は i 番目の 1.25ms 区間内の復号された時間包絡の現在の値  $\hat{T}_{env}(i)$  を指す事に注意されたい。圧縮の出力は以下の式で与えられる。

$$\hat{s}_{HB}^{bwe}(n) = \begin{cases} \hat{s}_{HB}^F(n) & \text{if } |\hat{s}_{HB}^F(n)| < \hat{T}_{env}(i) \\ \hat{s}_{HB}^F(n)/2 & \text{if } \hat{T}_{env}(i) \leq |\hat{s}_{HB}^F(n)| \leq 4\hat{T}_{env}(i) \\ \hat{s}_{HB}^F(n)/16 & \text{if } |\hat{s}_{HB}^F(n)| > 4\hat{T}_{env}(i) \end{cases}$$

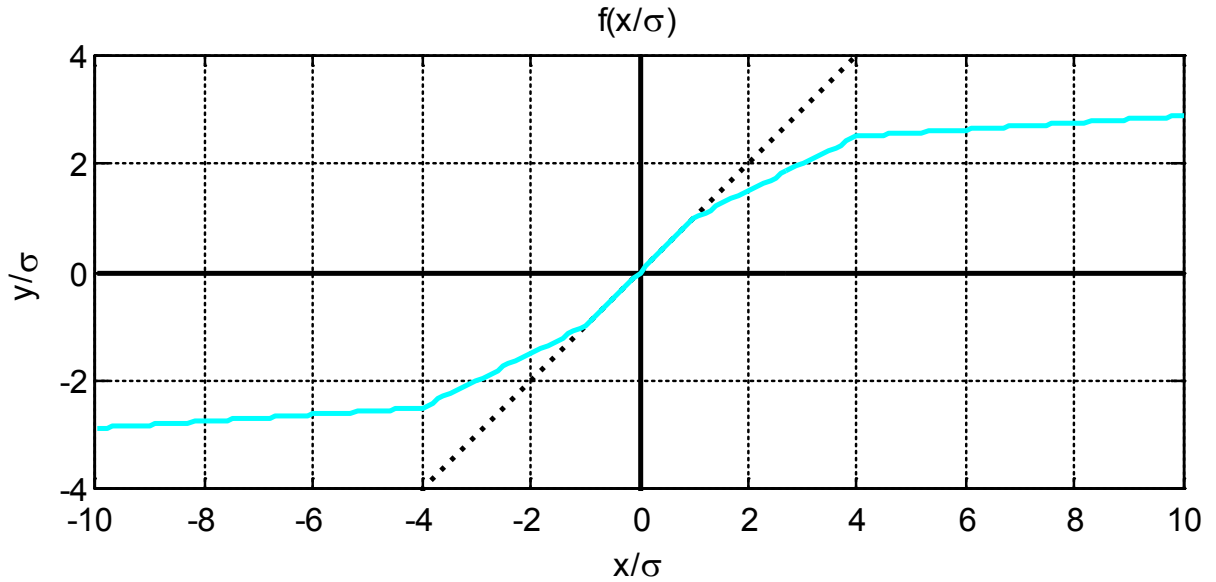


Figure12/JT-G729.1 Adaptive amplitude compression function (ITU-T G.729.1)

TDBWE合成  $\hat{s}_{HB}^{bwe}(n)$  はMDCTによって  $\hat{s}_{HB}^{bwe}(k)$  に変換される。このスペクトルは欠如しているサブバンドを外挿するためにTDAC復号器によって使用される。信号  $\hat{s}_{HB}^{bwe}(n)$  は、6.6.3節で記述された正規化と同様の方法で、MDCTの前に係数  $norm\_bwe$  によって正規化される。

### 7.3 TDAC復号器 (レイヤ4から12)

TDAC復号器を Figure13/JT-G729.1 に図示する。

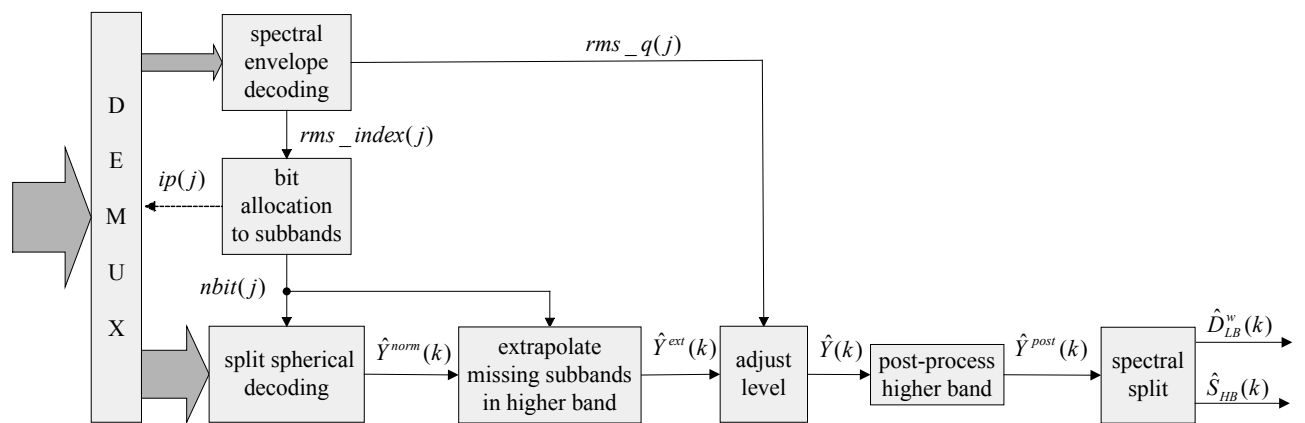


Figure13/JT-G729.1 Block diagram of the TDAC decoder (ITU-T G.729.1)

#### 7.3.1 MDCT正規化係数の復号化

符号器が伝送した4ビットの受信正規化係数 ( $norm\_MDCT$ ) は、MDCT係数をスケーリングするため



にTDAC復号器において使用される。その係数を使用して2つの逆MDCTにおいて復元される信号をスケールリングする。

### 7. 3. 2 スペクトル包絡復号化

最初に、高域スペクトル包絡が復号される。符号器において選択された符号化モードを示すビットは、0→差分ハフマン符号化、1→2進符号化である。

符号化モード0が選択された場合、5ビットを復号することにより、[-11,+20]の範囲のインデックス  $rms\_index(10)$  が得られる。次に、差分インデックス  $diff\_index(j)$   $j=11,\dots,17$  に関するハフマン符号が復号される。インデックス  $rms\_index(j)$   $j=11,\dots,17$  は、下式のように復元される。

$$rms\_index(j) = rms\_index(j-1) + diff\_index(j) \quad (112)$$

符号化モード1が選択された場合、 $8 \times 5$ ビットを復号することにより、[-11,+20]の範囲のインデックス  $rms\_index(j)$   $j=10,\dots,17$  が得られる。

高域スペクトル包絡を完全に復号するためにビット数が足りない場合、復号されたインデックス  $rms\_index(j)$  は、復号された高域スペクトルの部分的なレベル調整を行い続ける。

低域バンドに関するビット、すなわち、 $rms\_index(j)$   $j=0,\dots,9$  は、符号化モード0または1を選択する1ビットを含めて高域バンドと同じ方法で復号される。

復号されたインデックスを一次元ベクトル  $[rms\_index(0) rms\_index(1) \dots rms\_index(17)]$  に結合し、対数領域における復元されるスペクトル包絡を表す。線形領域におけるこの包絡は、下式のように変形される。

$$rms\_q(j) = 2^{\frac{1}{2} rms\_index(j)} \quad (113)$$

### 7. 3. 3 聴覚重要度によるサブバンド順序付け

6. 6. 7節と同様である。

スペクトル包絡が完全に復号されない場合、サブバンド順序付けは行われない。

### 7. 3. 4 MDCT係数量子化におけるビット割当て

6. 6. 8節と同様である。

スペクトル包絡が完全に復号されない場合、ビット割当ては行われない。

### 7. 3. 5 MDCT係数の復号化とスペクトル逆正規化

スペクトル包絡が完全に復号されない場合、MDCT係数は受信されずにMDCT係数の復号は行われな

い。  
ベクトル量子化インデックスは、聴覚重要度に応じてTDACビット列から読み出される。サブバンド  $j$  に全くビットが割当てられない、すなわち、 $nbit(j)=0$  の様な場合、または、対応するベクトル量子化が受信されない場合に、この段階において係数に零が設定される。

#### ベクトル量子化インデックスの復号化

次元  $nb\_coef(j)$  および非零ビット割当て  $nbit(j)$  の  $j$  番目のサブバンドにおいて、ベクトル量子化インデックスは、リーダー  $\mathbf{y}_0$  の極性順列となるコードベクトル  $\mathbf{y}$  と特定される。このインデックスは、式(80)のように展開することができる。

$j$  番目のサブバンドにおいて、 $index(\mathbf{y})$  の復号化は、次のステップで構成される。

- (1) テーブル参照により、リーダー  $\mathbf{y}_0$  を特定する
- (2)  $index(\mathbf{y}) - offset\_lead(\mathbf{y}_0)$  から  $\mathbf{y}$  の極性ビットと順列階数  $rank(\mathbf{y}|\mathbf{y}_0)$  を抽出する (単一ビット操作によ

り)

- (3) 順序階数  $\text{rank}(\mathbf{y}|\mathbf{y}_0)$  を復号して、 $\mathbf{y}_0$  を並替える
- (4) 極性ビットに従って極性を設定するにより、 $\mathbf{y}$  を復元する

順序階数  $\text{rank}(\mathbf{y}|\mathbf{y}_0)$  の復号は、Shalkwijk の公式に基づく。

$$\text{rank}(\mathbf{y}|\mathbf{y}_0) = \sum_{k=0}^{N-1} I_k^{d_k} \quad \text{ここで、} \quad I_k^{d_k} = \frac{(N-1-k)!}{\prod_{i=0}^{q-1} (w_{\text{lead},k}^i)!} \left( \sum_{d=0}^{d_k-1} w_{\text{lead},k}^d \right)$$

$\mathbf{y}$  の復号は、順序階数から  $\mathbf{y}$  に関するベクトル  $\mathbf{D} = (d_0, d_1, \dots, d_{N-1})$  を決定する。Shalkwijk の公式に基づき逐次探索を行う。

- 最初に、次の不等式を用いて  $d_0$  を決定する

$$I_0^{d_0} \leq \text{rank}(\mathbf{y}|\mathbf{y}_0) < I_0^{d_0+1}$$

$$\text{ここで } I_0^{d_0} = \frac{(N-1)!}{\prod_{i=0}^{q-1} (w_{\text{lead},0}^i)!} \left( \sum_{d=0}^{d_0-1} w_{\text{lead},0}^d \right) \quad \text{かつ} \quad I_0^{d_0+1} = \frac{(N-1)!}{\prod_{i=0}^{q-1} (w_{\text{lead},0}^i)!} \left( \sum_{d=0}^{d_0} w_{\text{lead},0}^d \right)$$

- 次に、以下の不等式を用いて  $d_1$  を決定する

$$I_1^{d_1} \leq \text{rank}(\mathbf{y}|\mathbf{y}_0) - I_0^{d_0} < I_1^{d_1+1} \quad \text{ここで、} \quad I_1^{d_1} = \frac{(N-2)!}{\prod_{i=0}^{q-1} (w_{\text{lead},1}^i)!} \left( \sum_{d=0}^{d_1-1} w_{\text{lead},1}^d \right)$$

- 同様に、次の不等式を用いて  $d_k$  を決定する

$$I_k^{d_k} \leq \text{rank}(\mathbf{y}|\mathbf{y}_0) - \sum_{j=0}^{k-1} I_j^{d_j} < I_k^{d_k+1}$$

6. 6. 9節で説明される様に、部分階数の実際の計算（および部分階数の分子と分母の項の計算）は、素因数分解に基づく。分母  $I_k^{d_k}$  の素因数分解は、次の漸化式によって計算されることに注意されたい。

$$\prod_{i=0}^{q-1} (w_{\text{lead},k}^i)! = \frac{\prod_{i=0}^{q-1} (w_{\text{lead},k-1}^i)!}{w_{\text{lead},k-1}^i}$$

$\prod_{i=0}^{q-1} (w_{\text{lead},k}^i)!$  の素因数分解は、素因数分解した  $w_{\text{lead},k-1}^i$  の指数部を素因数分解した部分項  $\prod_{i=0}^{q-1} (w_{\text{lead},k-1}^i)!$  の指数部から減算して計算される。

復元された  $j$  番目のサブバンドは、 $[\hat{Y}^{\text{norm}}(sb\_bound(j)+k)]_{k=0, \dots, nb\_coef(j)-1} = \mathbf{y}$  で与えられる。

ベクトル量子化インデックスを復号した後、各サブバンドは、単位球面上のコードベクトル  $\mathbf{y}$ 、または NULL ベクトルとして復元される。 $nbit(j) > 0$  の非零サブバンドは、復号されたサブバンド平均自乗根  $rms\_q(j)$  に従ってスケールされる。

$$\hat{Y}(sb\_bound(j)+k) = rms\_q(j) \times \sqrt{nb\_coef(j)} \times \hat{Y}^{\text{norm}}(sb\_bound(j)+k) \quad (114)$$

ここで、for  $k=0, \dots, nb\_coef(j)-1$

### 7. 3. 6 欠落した高域サブバンドの推定と推定されたサブバンドのレベル調整

高域バンドスペクトル (サブバンド  $j=10, \dots, 17$ ) において、非受信サブバンド、および、 $nbit(j)=0$  のサブバンドは、TDBWE 合成の MDCT において等価なサブバンドに置換される。すなわち、 $\hat{Y}^{ext}(sb\_bound(j)+k) = \hat{S}_{HB}^{bwe}(sb\_bound(j)-160+k)$ ,  $k=0, \dots, nb\_coef(j)-1$  である。

この高域バンドの包絡情報が利用できる場合、推定されたサブバンドの平均自乗根に  $rms\_q(j)$  を設定する。このレベル調整により受信した TDAC レイヤの数によって段階的に音質を改善することができる。

### 7. 3. 7 復号された高域バンドの後処理

低域バンドは、従来の時間領域の手法による後処理を適応するため、MDCT 係数の後処理は、高域バンドのみ適用される。高域バンドにおいて LPC 係数は復号器に伝送されない。TDAC 後処理は、復号器側で利用可能な MDCT 係数に関して行われる。

160 個の高域バンド MDCT 係数が存在し、 $\hat{Y}(k)$ ,  $k=160, \dots, 319$  とする。

この特別な後処理において、高域バンドは 16 個の MDCT 係数をもつ 10 個のサブバンドに分割される。各サブバンドの平均値は、以下の包絡で定義される。

$$env(j) = \sum_{k=0}^{15} |\hat{Y}(160+16j+k)|, \quad j=0, 1, \dots, 9 \quad (115)$$

後処理は、2つのステップから構成される。最初のステップは、*包絡後処理* (短期後処理に対応) であり、包絡を修正する処理である。2番目のステップは、*微細な構造の後処理* (長期後処理に対応) であり、各サブバンドの各係数の値を強調する。基本的な概念は、低い値は相対的により小さく設定する、その場合の符号化誤差は、値が大きき場合よりもより一層大きくなる。

包絡を修正するアルゴリズムを以下に記述する。最大包絡値は下式となる。

$$env_{\max} = \max_{j=0, \dots, 9} env(j) \quad (116)$$

包絡に適用する利得係数を下式で算出する。

$$fac_1(j) = \alpha_{ENV} \frac{env(j)}{env_{\max}} + (1 - \alpha_{ENV}), \quad j=0, \dots, 9 \quad (117)$$

ここで、 $\alpha_{ENV}$  ( $0 < \alpha_{ENV} < 1$ ) は、ビットレートに依存する。ビットレートが高くなるに従って、定数  $\alpha_{ENV}$  は、より小さい値となる。係数  $fac_1(j)$  が決定されると、修正された包絡は、下式で表される。

$$env'(j) = g_{norm} fac_1(j) env(j), \quad j=0, \dots, 9 \quad (118)$$

ここで、 $g_{norm}$  は、全体のエネルギーを保持する利得である。

$$g_{norm} = \frac{\sum_{k=0}^9 env(j)}{\sum_{k=0}^9 fac_1(j) env(j)} \quad (119)$$

各サブバンド内の微細な構造修正は、上記の包絡後処理と同様である。利得係数の値は、下式で算出される。

$$fac_2(j,k) = \beta_{ENV} \frac{|\hat{Y}(160+16j+k)|}{Y_{\max}(j)} + (1-\beta_{ENV}), \quad k=0,\dots,15, \quad (120)$$

ここで、サブバンド内の最大値  $Y_{\max}(j)$  は、

$$Y_{\max}(j) = \max_{k=0,\dots,15} |\hat{Y}(160+16j+k)| \quad (121)$$

となり、 $\beta_{ENV}$  ( $0 < \beta_{ENV} < 1$ ) は、ビットレートに依存する。ビットレートが高くなるに従い、 $\beta_{ENV}$  はより小さな値となる。包絡後処理と微細な構造の後処理を結合することにより、最終的に後処理された高域バンド MDCT 係数は下式となる。

$$\hat{Y}^{post}(160+16j+k) = g_{norm} fac_1(j) fac_2(j,k) \hat{Y}(160+16j+k), \quad j=0,\dots,9 \quad k=0,\dots,15 \quad (122)$$

低域バンドは、後処理されないことに注意されたい。従って、 $\hat{Y}^{post}(k) = \hat{Y}(k)$ ,  $k=0,\dots,159$  となる。

### 7. 3. 8 スペクトル分離、逆MDCTとオーバーラップ加算

復元されたスペクトル  $\hat{Y}^{post}(k)$  を低域バンドスペクトル  $\hat{D}_{LB}^w(k)$  と高域バンドスペクトル  $\hat{S}_{HB}(k)$  に分離する。両バンドは、逆MDCT変換において時間領域に変換される。

$$d_{LB}^{cur}(n) = \sqrt{\frac{2}{160}} \sum_{m=0}^{159} \cos\left(\frac{\pi}{160}(m+0.5)(n+80.5)\right) \hat{D}_{LB}^w(m) \quad (123)$$

低域バンドの差分信号は、以下のオーバーラップ加算処理によって算出される。

$$\hat{d}_{LB}^w(n) = w_{TDAC}(n+160) d_{LB}^{prev}(n) + w_{TDAC}(n) d_{LB}^{cur}(n), \quad n=0,\dots,159 \quad (124)$$

ここで、 $w_{TDAC}(n)$  は、合成重み窓である。

$$w_{TDAC}(n) = \sin\left(\frac{\pi}{320}(n+0.5)\right), \quad n=0,\dots,319 \quad (125)$$

$d_{LB}^{prev}(n)$  は、直前の逆MDCT変換により得られ、下式のように更新される。

$$d_{LB}^{prev}(n) = d_{LB}^{cur}(160+n), \quad n=0,\dots,159 \quad (126)$$

同様に、高域バンドにおいて、

$$s_{HB}^{cur}(n) = \sqrt{\frac{2}{160}} \sum_{m=0}^{159} \cos\left(\frac{\pi}{160}(m+0.5)(n+80.5)\right) \hat{S}_{HB}(m) \quad (127)$$

$$\hat{s}_{HB}(n) = w_{TDAC}(n+160) s_{HB}^{prev}(n) + w_{TDAC}(n) s_{HB}^{cur}(n), \quad n=0,\dots,159 \quad (128)$$

ここで、 $w_{TDAC}(n)$  は、合成重み窓であり、 $s_{HB}^{prev}(n)$  は、直前の逆MDCT変換により得られ、下式のように更新される。

$$s_{HB}^{prev}(n) = s_{HB}^{cur}(160 + n) \text{ for } n = 0, \dots, 159 \quad (129)$$

逆MDCTは、復号されたMDCT正規化係数を含む（MDCT正規化係数は4ビットで伝送される）。

### 7. 3. 9 聴覚重み付け逆フィルタ

逆フィルタ  $W_{LB}(z)^{-1}$  を下式に定義する。

$$W_{LB}(z)^{-1} = \frac{1}{fac} \frac{\hat{A}(z/\gamma_2')}{\hat{A}(z/\gamma_1')}, \quad (130)$$

ここで、 $1/fac$  は利得補正係数である。

$$\frac{1}{fac} = \frac{\left| \sum_{i=0}^p (-\gamma_1')^i \hat{a}_i \right|}{\left| \sum_{i=0}^p (-\gamma_2')^i \hat{a}_i \right|}, \quad (131)$$

$\hat{a}_i$  は、4. 1. 1 節/JT-G729の狭帯域エンベデッドCELP復号器から得られる復号された線形予測フィルタ  $\hat{A}(z)$  の係数である。符号器と同様に、これらの係数は、5msのサブフレーム毎に更新される。 $W_{LB}(z)^{-1}$  の役割は、低域バンドのTDAC復号器によって付加される符号化雑音を低減することである。係数  $1/fac$  は、 $\hat{d}_{LB}(n)$  と  $\hat{s}_{LB}(n)$  のスペクトル連続性を保証するように最適化される。

### 7. 3. 10 プリエコー処理

信号エネルギーが急峻に増加する、例えば、音声の立ち上がりや打楽器音のような場合、プリエコーとして知られる変換符号化による典型的な異音が観測される。プリエコーの原因は、次の通りである。周波数領域の量子化雑音は、逆MDCT変換や加算/オーバーラップ処理によって時間領域に変換され、MDCT合成窓(JT-G729.1では40ms)において一様に分散する。音声の立ち上がりにおいて、多くの場合、立ち上がり直前の入力信号部分は、直後の部分のエネルギーと比較して非常にエネルギーが低くなる、その一方で、合成窓全体において量子化雑音レベルが非常に高くなる。このような場合、エネルギーが低い部分において信号と雑音の比率は、非常に小さくなる（しばしばマイナスの値となる）。プリエコーと呼ばれる余分な人工的な信号のように、音声の立ち上がり直前に量子化雑音を聞き取ることができる。同様な現象が、急激に信号にオフセットを生じた後にも存在し、信号にオフセットが生じた直後にポストエコーと呼ばれる量子化雑音を聞き取ることができる。

エコー異音を防止するためにエコー低減が必要となる。

- プリエコーについては、合成窓の幾つかの部分において著しいエネルギー増加が認められる場合、ただし、エコー低減は、合成窓のエネルギーが低い部分に限定されなければならない。
- ポストエコーについては、直前に復号されたスーパーフレームのエネルギーが現在のスーパーフレームのエネルギーよりも著しく大きい場合である。

エコー低減方法の主な特徴は、スケーラブル構造において時間領域の符号化を利用するところである。レイヤ1から3(8~14kbit/s)までの復号された信号は、プリ/ポストエコーに関係しない。従ってMDCT

レイヤの時間包絡を制限し、関係する時間包絡を使用して各MDC Tレイヤ（低域バンドと高域バンドの両方において）のエコーを低減する。この制限をエコーが存在する区間のみ実行し、エコーの存在しない区間においては禁止する。一方、14kbit/s の時間包絡は、誤りが多いため、信号の高エネルギー部分において時間包絡を制限することにより劣化が生じる。このように、逆MDC Tのみを使用してエコー／非エコー区間の識別を行い、最終的な判断は2つの基準に基づく。

#### エコー／非エコー区間の識別手順

最初に、逆MDC Tに基づいてエコー区間と非エコー区間を識別する手順を実行する。低域バンドと高域バンドについて同じ操作を行う。

低域バンドについて、現在のスーパーフレームにおける非エコー区間の開始および終了を示すインデックス  $ind_{LB}^1$  と  $ind_{LB}^2$  を識別手順に従って算出する。 $ind_{LB}^1 > ind_{LB}^2$  の場合、スーパーフレーム全体をエコー区間と見なす。以下に説明する手順によりインデックスが得られる。

信号  $\hat{d}_{LB}^w(n)$  と  $w_{TDAC}(n) d_{LB}^{prev}(n)$  を連結して、現在のMDC T合成窓に相当する 320(40ms)の長さの補助信号  $d_{LB}^{conc}(n)$  を形成する。連結した信号を 40 サンプル(5ms)の8つのサブフレームに分割してサブフレームのエネルギーを計算する。信号  $d_{LB}^{prev}(n)$  が対称であるため、6つのエネルギーだけが相違することに注意されたい。

$$En_{LB}(i) = \sum_{n=40i}^{40(i+1)-1} [d_{LB}^{conc}(n)]^2, \quad i = 0, \dots, 5 \quad (132)$$

$$En_{HB}(i) = \sum_{n=40k}^{40(i+1)-1} [s_{HB}^{conc}(n)]^2, \quad i = 0, \dots, 5 \quad (133)$$

低域バンドについて、下式の手順を継続する。

$$\begin{aligned} Max_{LB}^1 &= \max_{i=0, \dots, 3} En_{LB}(i) \\ Max_{LB}^2 &= \max_{i=0, \dots, 5} En_{LB}(i) \\ Min_{LB} &= \min_{i=0, \dots, 3} En_{LB}(i) \end{aligned} \quad (134)$$

以下の最大値、最小値を探索する。

直前のスーパーフレームで蓄積された  $Max_{LB}^{prev}$  を用いることで、インデックス  $ind_{LB}^1$  と  $ind_{LB}^2$  は下式のように求められる。

$$\begin{aligned} &\text{if } Max_{LB}^{prev} > 16 Max_{LB}^1, \quad ind_{LB}^1 = 160, ind_{LB}^2 = 159 \quad (\text{a post-echo situation}) \\ &\text{else if } Max_{LB}^2 < 16 Min_{LB}, \quad ind_{LB}^1 = 0, ind_{LB}^2 = 159 \quad (\text{no significant energy increase}) \\ &\text{else } \quad ind_{LB}^1 = \arg \max_{k=0, \dots, 5} (En_{LB}(k)) \times 40 \text{ and } ind_{LB}^2 = \min(ind_{LB}^1 + 80, 159) \end{aligned}$$

最大エネルギーをもつスーパーフレームの最初のサンプルインデックスを高エネルギー区間の開始と考える。

高エネルギー区間は、最大2つのスーパーフレームに及ぶ（これが時間包絡の誤差問題を処理する際に重要となる）。 $ind_{LB}^1 > 159$  の場合、次のスーパーフレームは最大値となり、現在のスーパーフレーム全体が、エコー区間として見なされることに注意されたい。

$Max_{LB}^{prev}$  の値は、下式のように更新される。

$$Max_{LB}^{prev} = Max_{LB}^1 \quad (135)$$

高域バンドについては、添え字 LB を HB に置換して同様の方法を繰り返す。

### エコー低減

次のようにエコー低減を実行する。最初に時間領域レイヤを使用してエコー検出を行い、利得減衰を算出する。次に  $ind_{LB}^1$  と  $ind_{LB}^2$  の識別区間を使用してエコーが存在する区間を再決定する。最後に平滑化を利得減衰に適用する。

低域バンドにおいて、低域バンドの TDAC 出力  $\hat{d}_{LB}(n)$  を逆聴覚重み付けした出力に CELP レイヤ  $\hat{s}_{LB}^{celp}(n)$  の出力を加算することでエコー低減を実行する。傾き補償について、フィルタ  $H_{ii}(z) = 1 - z^{-1}$  により信号がフィルタリングされる。

$$\hat{s}^{ii}(z) = H_{ii}(z) \hat{s}(z) \quad (136)$$

$$\hat{d}_{LB}^{ii}(z) = H_{ii}(z) \hat{d}_{LB}(z) \quad (137)$$

出力信号を 40 サンプル (5 ms) の 4 つのサブフレームに分割して、各サブフレームの時間包絡をサブフレームのエネルギーとして算出する。

$$Env_{LB}^{celp}(i) = \sum_{n=40i}^{40(i+1)-1} \hat{s}^{ii}(n)^2 \quad (138)$$

$$Env_{LB}^{dac}(i) = \sum_{n=40i}^{40(i+1)-1} \hat{d}_{LB}^{ii}(n)^2 \quad (139)$$

TDAC レイヤ  $Env_{LB}^{dac}(i)$  の  $i$  番目のサブフレームにおける時間包絡と、CELP レイヤ  $Env_{LB}^{celp}(i)$  の対応する時間包絡との比率を 1 と比較する。 $\frac{Env_{LB}^{dac}(i)}{Env_{LB}^{celp}(i)} > 1$  の場合、エコーが検出される。初期スケーリング係数  $g_{LB}(n)$  を各サブフレームの各サンプルについて算出する。

$$g_{LB}(n) = \begin{cases} \sqrt{\frac{Env_{LB}^{celp}(i)}{Env_{LB}^{dac}(i)}} & \text{if } \frac{Env_{LB}^{dac}(i)}{Env_{LB}^{celp}(i)} > 1 \\ 1 & \text{otherwise} \end{cases} \quad n = 40i, \dots, 40(i+1)-1, \quad i = 0, \dots, 3 \quad (140)$$

$ind_{LB}^1$  と  $ind_{LB}^2$  に範囲で限定された低域バンドの非エコー区間において、初期スケーリング係数を 1 に設定する。

$$g_{LB}(n) = 1, \quad n = ind_{LB}^1, \dots, ind_{LB}^2$$

最後に、下式により  $g_{LB}(n)$  を平滑化する。

$$g'_{LB}(n) = 0.85g'_{LB}(n-1) + 0.15g_{LB}(n) \quad (141)$$

最終的な低域バンド出力は、CELP 出力と  $g'_{LB}(n)$  で重み付けられた TDAC 低域レイヤ出力信号を加算して得られる。

$$\hat{s}_{LB}(n) = \hat{s}(n) + g'_{LB}(n) \hat{d}_{LB}(n) \quad (142)$$

高域バンドにおいて、TDACの出力  $\hat{s}_{HB}(n)$  についてエコー低減を同様の方法で行う。高域バンドのTDAC出力信号とTDBWE出力信号  $\hat{s}_{HB}^{dbwe}(n)$  を40サンプル(5ms)の4つのサブフレームに分割して、各サブフレームの時間包絡をサブフレームのエネルギーとして算出する。

$$Env_{HB}^{dbwe}(i) = \sum_{n=40i}^{40(i+1)-1} \hat{s}_{HB}^{dbwe}(n)^2 \quad (143)$$

$$Env_{HB}^{dac}(i) = \sum_{n=40i}^{40(i+1)-1} \hat{s}_{HB}(n)^2 \quad (144)$$

TDACレイヤ  $Env_{HB}^{dac}(i)$  の  $i$  番目の時間包絡とTDBWEレイヤ  $Env_{HB}^{dbwe}(i)$  の対応する時間包絡との比率を0.81と比較する。 $\frac{Env_{HB}^{dac}(i)}{Env_{HB}^{dbwe}(i)} > 0.81$  の場合、エコーが検出される。初期スケーリング係数  $g_{HB}(n)$  を各サブフレームの各サンプルについて算出する。

$$g_{HB}(n) = \begin{cases} \sqrt{\frac{Env_{HB}^{dbwe}(i)}{Env_{HB}^{dac}(i)}} & \text{if } \frac{Env_{HB}^{dac}(i)}{Env_{HB}^{dbwe}(i)} > 0.81 \\ 1 & \text{otherwise} \end{cases} \quad n = 40i, \dots, 40(i+1)-1, \quad i = 0, \dots, 3 \quad (145)$$

$ind_{HB}^1$  と  $ind_{HB}^2$  の範囲で限定された高域バンドの非エコー区間において、初期スケーリング係数を1に設定する。

$$g_{HB}(n) = 1, \quad n = ind_{HB}^1, ind_{HB}^2$$

最後に、下式により  $g_{HB}(n)$  を平滑化する。

$$g'_{HB}(n) = 0.85g'_{HB}(n-1) + 0.15g_{HB}(n) \quad (146)$$

スペクトル折り返し以前の最終的な高域バンド出力は、 $g'_{HB}(n)$  で重み付けられたTDAC高域レイヤ出力信号として得られる。

$$\hat{s}_{HB}^{fold}(n) = g'_{HB}(n) \hat{s}_{HB}(n) \quad (147)$$

#### 7. 4 低域バンドのポストフィルタリング

4. 2節/JTG729に記載される様に、JTG729の復号器は、適応ポストフィルタ処理、高域通過フィルタ処理、信号アップスケーリング処理に分けられる後処理を含む。同様にJTG729. 1の復号器は、低域バンドの後処置を含む。しかしながら、この手順は、適応ポストフィルタ処理と高域通過フィルタ処理に限定される。JTG729. 1の復号器において、信号のアップスケーリング処理は、QMF合成フィルタバンクにより処理される。

##### 7. 4. 1 適応ポストフィルタ

JTG729. 1の適応ポストフィルタは、JTG729ポストフィルタに直接由来している。3つのフィルタが直列に接続されている。長期ポストフィルタ  $H_p(z)$ 、短期ポストフィルタ  $H_f(z)$ 、傾き補償フィルタ  $H_t(z)$  であり、これらは適応利得制御手順に従う。JTG729の適応ポストフィルタからの変更点は、以下のみである。



- JT-G 7 2 9 の長期および短期ポストフィルタのパラメータ  $\gamma_p, \gamma_m, \gamma_d$  は、復号器ビットレート（8 または 12kbit/s またはそれ以上）に依存する
- JT-G 7 2 9 の適応利得制御は、無音セグメント（8 および 12kbit/s のみ）において量子化誤差を低減するように変更される

これらの2つの変更点の詳細を次に示す。

#### 適応ポストフィルタパラメータの適応動作

長期および短期ポストフィルタの  $\gamma_p, \gamma_n$  と  $\gamma_d$  のパラメータ値を Table15/JT-G729.1 に示す。12 kbit/s において、 $\gamma_n$  と  $\gamma_d$  のパラメータ値は、係数  $0 \leq Th \leq 1$  に依存し、10ms のフレームエネルギーに基づき、5タップのメディアンフィルタにより平滑化される。

Table15/JT-G729.1 Parameters of the adaptive postfilter depending on bit rate.  
(ITU-T G 729.1)

Bit rate (kbit/s)	$\gamma_p$	$\gamma_n$	$\gamma_d$
8	0.5	0.55	0.7
12		$Th \times 0.7 + (1 - Th) \times 0.55$	$Th \times 0.75 + (1 - Th) \times 0.7$
14 and above		0.7	0.75

#### 適応利得制御手順の変更（8 および 12kbit/s のみ）

無音区間では、サンプル値が非常に小さくかつ量子化誤差が相対的に非常に大きくなる。固定小数点演算では、無音区間の相対量子化誤差が非常に大きくなり、本来の無音入力エネルギーと比較して出力エネルギーが大きくなる。無音信号の絶対エネルギーは低いけれども、明瞭に聞き取ることができる。この問題を解決する1つの簡単な方法は、低レベルの無音信号を検出後、復号器側でエネルギーを低減することである。手順を次に示す。

後処理前の無音信号レベルを次のように定義する（4. 2. 4節/JT-G 7 2 9参照）

$$g_{in} = \sum_{n=0}^{39} |\hat{s}_{LB}(n)| \quad (148)$$

ここで、 $\hat{s}_{LB}(n)$  は、後処理を行う前の信号である。平滑化レベルは、次のように算出される。

$$g_{in}^{sm} = 0.75g_{in,prev}^{sm} + 0.25g_{in} \quad (149)$$

ここで、 $g_{in,prev}^{sm}$  は、直前のサブフレームの  $g_{in}^{sm}$  の値である。無音検出は、信号レベルとスペクトル傾きパラメータに基づく（第一反射係数と呼ばれ  $k'_1$  で記述される。式(87) / JT-G 7 2 9において算出される）。

$$S_{dec} = (g_{in}^{sm} < 1024) \text{ and } (g_{in} < 2g_{in}^{sm}) \text{ and } (k'_1 < 0.015625) \quad (150)$$

$S_{dec} = 1$  を満足する場合、低いレベルの無音が検出され、 $g_{in}^{sm} < g_{in}$  の場合、参照信号レベルは平滑化した参照信号レベルに置換される。

$$g_{in} = g_{in}^{sm} \quad (151)$$

4. 2. 4節/JT-G 7 2 9参照) で定義された信号  $sf(n)$ ,  $n=0, 1, \dots, 39$  が、利得調整前のポストフィ

ルタ信号となる。初期ポストフィルタ信号レベルは、次の通りである。

$$g_{out} = \sum_{n=0}^{39} |sf(n)| \quad (152)$$

ポストフィルタエネルギーを調整するために使用する初期利得を次に示す。

$$G = \frac{g_{in}}{g_{out}} \quad (153)$$

低いレベルの無音が検出された場合（例えば  $S_{dec} = 1$ ）、利得係数  $g'$  を最適化することにより、上記の利得は低減される。

$$g' = kk \times (g_{in}^{sm} / 1024) + (1 - kk) \quad (154)$$

ここで、 $0 \leq g' \leq 1$  と  $0 \leq kk \leq 1$  は、 $k'_1$  の関数である。

$$kk = \min\left(\frac{0.015625 - k'_1}{0.0624695}, 1\right) \quad (155)$$

4. 2. 4 節/J T-G 7 2 9 と同様に、スケーリングされた信号  $sf'(n)$   $n=0, \dots, 39$  は、 $sf(n)$  に  $g^{(n)}$  を乗算することにより得られる。ここで、

$$g^{(n)} = 0.95 g^{(n-1)} + 0.05 G \times g' \quad (156)$$

である。

無音利得の低減は、8 および 12 kbit/s のみで実行される。狭帯域と広帯域をスイッチする間、利得係数  $g'$  は徐々に変化する。

#### 7. 4. 2 高域通過フィルタ

4. 2. 5 節/J T-G 7 2 9 と同様である。

この高域通過フィルタは 8 および 12 kbit/s のみで使用されることに注意されたい。

#### 7. 5 高域バンドのスペクトル折り返し処理、信号アップスケール処理、QMF 合成フィルタバンク

高域バンドの合成  $\hat{s}_{HB}^{fold}(n)$  は、次のようにスペクトル領域において折り返し処理される。

$$\hat{s}_{HB}^{qmf}(n) = (-1)^n \hat{s}_{HB}^{fold}(n), \quad n = 0, \dots, 159$$

8000Hz でサンプリングされた合成用低域バンド信号  $\hat{s}_{LB}^{qmf}(n)$  と高域バンド信号  $\hat{s}_{HB}^{qmf}(n)$  が与えられ、Figure 2/JT-G729.1 に示すような QMF 合成フィルタバンクを用いて、16000Hz の出力信号が算出される。フィルタ  $G_1(z)$  と  $G_2(z)$  は、 $G_1(z) = H_1(z)$  と  $G_2(z) = -H_2(z)$  であることに注意されたい。

QMF 合成を簡潔に実施するため、次の処理を行う。 $\hat{s}_{LB}^{qmf}(n)$  と  $\hat{s}_{HB}^{qmf}(n)$  を係数 2（零挿入）でアップサンプリングすることで 16000Hz の  $\hat{y}_1(n)$  と  $\hat{y}_2(n)$  が得られ、次にフィルタ  $G_1(z)$  と  $G_2(z)$  で個別にフィルタリングされる。最後に、 $\hat{s}_{WB1}(n)$  と  $\hat{s}_{WB2}(n)$  を加算して出力信号  $\hat{s}_{WB}(n)$  が得られる。

式(2)、式(3)式(4)を同様に、フィルタリングされた信号  $\hat{s}_{WB1}(n)$  と  $\hat{s}_{WB2}(n)$  を次に与える。

$$\begin{aligned}\hat{s}_{WB1}(n) &= \sum_{j=0}^{31} g_1(j) [\hat{y}_1(n-j) + \hat{y}_1(n+1+j)] \\ &= \sum_{j=0}^{31} h_1(j) [\hat{y}_1(n-j) + \hat{y}_1(n+1+j)]\end{aligned}\quad (157)$$

$$\begin{aligned}\hat{s}_{WB2}(n) &= \sum_{j=0}^{31} g_2(j) [\hat{y}_2(n-j) + \hat{y}_2(n+1+j)] \\ &= \sum_{j=0}^{31} h_1(j) (-1)^j [\hat{y}_2(n-j) - \hat{y}_2(n+1+j)]\end{aligned}\quad (158)$$

QMF 分析と同様に、 $\hat{s}_{WB1}(n)$  と  $\hat{s}_{WB2}(n)$  の計算を一緒に行うことで演算量を削減する。ここでは、 $n$  が奇数値の場合、 $\hat{y}_1(n)$  と  $\hat{y}_2(n)$  が、零に等しいことを利用する。

式(157)と式(158)より、 $\hat{s}_{WB1}(n)$  と  $\hat{s}_{WB2}(n)$  を加算し、奇数および偶数フィルタ係数に応じて総和を分けることにより、出力信号  $\hat{s}_{WB}(n)$  は、次のように与えられる。

$$\begin{aligned}\hat{s}_{WB}(n) &= \sum_{j=0}^{31} h_1(j) [\hat{y}_1(n-j) + \hat{y}_1(n+1+j)] + \sum_{j=0}^{31} h_1(j) (-1)^j [\hat{y}_2(n-j) - \hat{y}_2(n+1+j)] \\ &= \sum_{k=0}^{15} h_1(2k) [\hat{y}_1(n-2k) + \hat{y}_1(n+1+2k)] + \sum_{k=0}^{15} h_1(2k) [\hat{y}_2(n-2k) - \hat{y}_2(n+1+2k)] \\ &\quad + \sum_{k=0}^{15} h_1(2k+1) [\hat{y}_1(n-2k-1) + \hat{y}_1(n+2k+2)] - \sum_{k=0}^{15} h_1(2k+1) [\hat{y}_2(n-2k-1) - \hat{y}_2(n+2k+2)]\end{aligned}\quad (159)$$

$N$  が奇数値の場合、 $\hat{y}_1(n)$  と  $\hat{y}_2(n)$  が、零に等しくなることから、式(145)の半数の項は、 $n$  が偶数値の場合に零となり、残りの半数の項も  $n$  が奇数値の場合に零となる。すなわち、 $n$  が偶数値の場合、以下のようになる。

$$\begin{aligned}\hat{s}_{WB}(n) &= \sum_{k=0}^{15} h_1(2k) [\hat{y}_1(n-2k) + \hat{y}_2(n-2k)] + \sum_{k=0}^{15} h_1(2k+1) [\hat{y}_1(n+2k+2) + \hat{y}_2(n+2k+2)] \\ &= \sum_{k=0}^{15} h_1(2k) \hat{y}_{sum}(n-2k) + \sum_{k=0}^{15} h_1(2k+1) \hat{y}_{sum}(n+2k+2)\end{aligned}\quad (160)$$

$n$  が奇数値の場合、以下のようになる。

$$\begin{aligned}\hat{s}_{WB}(n) &= \sum_{k=0}^{15} h_1(2k) [\hat{y}_1(n+1+2k) - \hat{y}_2(n+1+2k)] + \sum_{k=0}^{15} h_1(2k+1) [\hat{y}_1(n-2k-1) - \hat{y}_2(n-2k-1)] \\ &= \sum_{k=0}^{15} h_1(2k) \hat{y}_{diff}(n+1+2k) + \sum_{k=0}^{15} h_1(2k+1) \hat{y}_{diff}(n-1-2k)\end{aligned}\quad (161)$$

ここで、 $\hat{y}_{sum}(n) = \hat{y}_1(n) + \hat{y}_2(n)$  と  $\hat{y}_{diff}(n) = \hat{y}_1(n) - \hat{y}_2(n)$  である。

$\hat{y}_{sum}(n)$  と  $\hat{y}_{diff}(n)$  は、アップサンプリングされた信号であることに注意されたい。

$$x_{sum}(n) = \hat{s}_{LB}(n) + \hat{s}_{HB}(n) \quad \text{かつ} \quad x_{diff}(n) = \hat{s}_{LB}(n) - \hat{s}_{HB}(n).$$

このように、式(160)と(161)は以下のように記述することができる。

$$\hat{s}_{WB}(n) = \begin{cases} \sum_{k=0}^{15} h_1(2k)x_{sum}(m-k) + \sum_{k=0}^{15} h_1(2k+1)x_{sum}(m+k+1) & \text{if } n \text{ is even} \\ \sum_{k=0}^{15} h_1(2k)x_{diff}(m+1+k) + \sum_{k=0}^{15} h_1(2k+1)x_{diff}(m-k) & \text{if } n \text{ is odd} \end{cases} \quad (162)$$

ここで  $m = n/2$  である。最後に出力信号は、係数 2 でスケーリングされる。

QMF 合成フィルタバンクは、符号器で (2) でダウンサンプリングした信号を補償するために、加算信号を (2) でアップスケーリングすることに注意されたい。

## 7. 6 フレーム損失補償

損失補償処理は、消失したスーパーフレームのパラメータを、直前と後続のスーパーフレームから適切に置換する処理と、エネルギーを注意深く調整することで消失したスーパーフレームを合成する処理とから成り立つ。損失補償処理は、消失したスーパーフレームのクラスに依存し、位相情報と利得情報を含む他の伝達されたパラメータを用いる。このことにより、声門パルスの再同期やエネルギー制御、及び疑似立ち上がり再生を含む、効率的な補償技術と回復技術を用いることが出来る。

### 7. 6. 1 減衰係数の決定

損失補償処理の間、信号は通常、最新の正常に受信したスーパーフレームクラスのパラメータと、連続して消失したスーパーフレームの数によって決まる、減衰係数  $\alpha$  に基づいて減衰される。さらに、減衰係数  $\alpha$  は無声スーパーフレームに対する線形フィルタの安定性にも依存する。一般に、最新の正常に受信したスーパーフレームが安定した区間にあるなら減衰速度は遅く、スーパーフレームが遷移区間にあるなら減衰速度は速い。 $\alpha$  の値を Table16/JT-G729.1 にまとめる。

Table16/JT-G729.1 Values of FEC attenuation factor  $\alpha$ .  
(ITU-T G.729.1)

Last good received superframe	Number of successive erased superframes	$\alpha$
VOICED	1	$\beta$
	2, 3	$\bar{g}_p$
	> 3	0.4
ONSET	1	$0.8\beta$
	2, 3	$\bar{g}_p$
	> 3	0.4
ARTIFICIAL ONSET	1	$0.6\beta$
	2, 3	$\bar{g}_p$
	> 3	0.4
VOICED TRANSITION	$\leq 2$	0.8
	$> 2$	0.2
UNVOICED TRANSITION		0.88
UNVOICED	= 1	0.95
	2, 3	$0.6\theta + 0.4$
	> 3	0.4

Table16/JT-G729.1 において、 $\bar{g}_p$  はサブフレーム当たりのピッチ利得の平均値で、次式で与えられる。

$$\bar{g}_p = 0.1g_p^{(0)} + 0.2g_p^{(1)} + 0.3g_p^{(2)} + 0.4g_p^{(3)} \quad (163)$$

ここで、 $g_p^{(i)}$  はサブフレーム  $i$  のピッチ利得である。さらに  $\beta$  は次式で与えられる。

$$\beta = \sqrt{g_p} \quad \text{bounded by } 0.85 \leq \beta \leq 0.98 \quad (164)$$

$\theta$  は、隣接した LP フィルタの間の距離から計算される安定性係数である。 $\theta$  の値は、 $0 \leq \theta \leq 1$  の区間で制限され、 $\theta$  が大きくなればなるほど信号が安定していることを示す。この結果、安定した無声区間で単独のスーパーフレームの損失が発生したとき、エネルギーとスペクトル包絡のふらつきが減少する。安定性係数  $\theta$  は次式で与えられる。

$$\theta = 1.25 - \frac{1}{1.4} \sum_{i=0}^9 (LSP_i - LSPold_i)^2 \quad \text{bounded by } 0 \leq \theta \leq 1. \quad (165)$$

ここで、 $LSP_i$  は、現在のスーパーフレームにおける 2 番目の 10ms フレームの LSP であり、 $LSPold_i$  は直前のスーパーフレームにおける 2 番目の 10ms フレームの LSP である。ただし、LSP は余弦領域(-1 から 1) の値である。

## 7. 6. 2 復号器における分類

分類ビットはレイヤ 2 (12 kbit/s) で伝送される。従って、8 kbit/s のコアレイヤのみを受信した場合、スーパーフレーム分類は復号器にて処理される必要がある。この手順は、6.3.1.1 項に記載された分類法に類似している。

復号器における分類のため、正規化された相関関数  $r_x$ 、スペクトル傾き係数  $e_t$ 、ピッチ安定性カウンタ  $pc$ 、相対スーパーフレームエネルギー  $E_s$ 、及び零交差カウンタ  $zc$  の 5 つのパラメータを用いる。

復号器では重み付け音声信号が得られないため、6.3.1.1 項の記載とは逆に、正規化された相関関数  $r_x$  は CELP 合成の段階で計算される。

正規化された相関関数  $r_x$  はピッチに同期して以下の式で計算される。

$$r_x = \frac{\sum_{n=0}^{T-1} \hat{s}_{LB}^{celp}(t+n) \hat{s}_{LB}^{celp}(t+n-T)}{\sqrt{\sum_{n=0}^{T-1} [\hat{s}_{LB}^{celp}(t+n)]^2 \sum_{i=0}^{T-1} [\hat{s}_{LB}^{celp}(t+n-T)]^2}} \quad (166)$$

ここで、 $T$  は直前のサブフレームのピッチラグであり、時刻  $t=160-T$  である。

$T > 60$  の場合、 $T$  は直前の 2 サブフレームのピッチラグの平均値が設定される。ピッチラグがサブフレームサイズよりも小さい ( $T < 40$ ) 値では、正規化された相関関数は時刻  $t=160-T$  及び  $t=320-T$  において 2 度計算され、 $r_x$  はこの 2 つの値の平均値として与えられる。

スペクトル傾きパラメータ  $e_t$  は、最新の 3 サブフレームの合成信号  $\hat{s}_{LB}^{celp}(n)$  の 1 次正規化自己相関関数として計算される。

$$e_t = \frac{\sum_{n=40}^{319} \hat{s}_{LB}^{celp}(n) \hat{s}_{LB}^{celp}(n-1)}{\sum_{n=40}^{319} [\hat{s}_{LB}^{celp}(n)]^2} \quad (167)$$

ピッチ安定性カウンタ  $pc$  は、ピッチ周期の変化量を評価するもので、次式に基づき復号器にて計算される。

$$pc = |T_2(2) - T_1(2)| + |T_1(2) - T_2(1)| \quad (168)$$

ここで、 $T_1(i)$  と  $T_2(i)$  はそれぞれ、 $i$  番目の 10ms フレームの第 1 及び第 2 の閉ループピッチラグに対応する。即ち  $pc$  は、最新の 3 つの 5ms サブフレームの閉ループピッチラグを用いて計算する。また、相対スーパーフレームエネルギーは、dB 単位で表現された現在のスーパーフレームエネルギーと長区間平均エネルギーとの差分として計算される。

$$E_s = E_t - E_{it} \quad (169)$$

ここで、スーパーフレームエネルギー  $E_t$  は合成信号のエネルギー(dB 単位)で、スーパーフレームの終端でピッチに同期して計算される。

$$E_t = 10 \log_{10} \left( \frac{1}{160} \sum_{n=0}^{159} [\hat{s}_{LB}^{celp}(n+160-T)]^2 \right) \quad (170)$$

ここで、 $T$  は最新の 2 サブフレームのピッチラグの平均値である。短いピッチラグ値( $T < 40$ )については、2 倍のピッチ周期を用いて、即ち  $T$  を  $2T$  に設定してエネルギーを計算する。

長区間平均エネルギーは、有音音声スーパーフレームについて、次式を用いて更新される。

$$E_{it} = 0.99E_{it} + 0.01E_t \quad (171)$$

最後のパラメータである零交差パラメータ  $zc$  は、1 スーパーフレームの合成信号を用いて計算される。零交差カウンタ  $zc$  は、この区間において信号の極性が正から負へ変化した回数をカウントする。

分類パラメータは評価関数  $f_m$  を定義するために用いられる。パラメータ  $p_x$  は次式で求められる。

$$p^s = k_p p_x + c_p \quad (172)$$

正規化されたピッチ安定性パラメータは、0 と 1 の間に制限される。また、正規化された相関パラメータは、正值の場合は 2 倍される。係数  $k_p$  及び  $c_p$  は、スーパーフレームの消失が起きたときに用いられる補償及び回復技術に起因する信号歪を最小にするよう、それぞれ実験的に見出されたものである。用いる値を Table17/JT-G729.1 に示す。

Table17/JT-G729.1 Signal classification parameters at the decoder and the coefficients.  
(ITU-T G.729.1)

Parameter	Meaning	$k_p$	$c_p$
$r_x$	Normalized correlation	2.857	-1.286
$\bar{e}_t$	Spectral tilt	0.8333	0.2917
$pc$	Pitch stability counter	-0.0588	1.6468
$E_s$	Relative superframe energy	0.57143	0.85741
$zc$	Zero crossing counter	-0.067	2.613

評価関数は次式で定義される。

$$f_m = \frac{1}{6} (2\bar{r}_x^s + \bar{e}_t^s + pc^s + E_s^s + zc^s) \quad (173)$$

ここで、右肩付の  $s$  は、パラメータの正規化されたバージョンであることを示す。この評価関数  $f_m$  を用い、Table18/JT-G729.1 に示す規則に従って分類がなされる。

Table18/JT-G729.1 Signal classification rules at the decoder.  
(ITU-T G.729.1)

Previous super frame class	Rule	Current superframe class
ONSET VOICED VOICED TRANSITION	ARTIFICIAL ONSET $f_m \geq 0.63$	VOICED
	$0.39 \leq f_m < 0.63$	VOICED TRANSITION
	$f_m < 0.39$	UNVOICED
UNVOICED TRANSITION UNVOICED	$f_m > 0.56$	ONSET
	$0.56 \geq f_m > 0.45$	UNVOICED TRANSITION
	$f_m \leq 0.45$	UNVOICED

### 7. 6. 3 消失したスーパーフレームのクラスの決定

後続のスーパーフレームの分類情報が利用できない場合、クラスは最新の正常に受信したスーパーフレームのクラスと同じクラスに設定する。クラス情報が後続のスーパーフレームで有効である場合、消失したスーパーフレームのクラスは、後続のスーパーフレームのクラスと、最新の正常に受信したスーパーフレームのクラスに基づいて推測される。後続のスーパーフレームのレイヤ2が受信された場合（後続のスーパーフレームのビットレートが 8kbit/s 以上かつ消失していない）、後続のスーパーフレームのクラスは有効である。

$class_{old}$  を、最新の正常に受信されたスーパーフレームのクラス、 $class_{new}$  を後続のスーパーフレームのクラスとし、また、推定すべき消失したスーパーフレームのクラスを  $class_{lost}$  と定義する。

まず、 $class_{lost} = class_{old}$  とする。後続のスーパーフレームが有効ならば、そのクラス情報を復号し  $class_{new}$  として定義する。次に、 $class_{lost}$  の値を以下の手順に従って更新する。

$class_{new} = \text{VOICED}$  かつ  $class_{old} = \text{ONSET}$  ならば、 $class_{lost} = \text{VOICED}$  とする。

$class_{new} = \text{VOICED}$  かつ  $class_{old} = \text{ONSET}$  または  $\text{VOICED}$  ならば、 $class_{lost} = \text{VOICED}$  とする。

$class_{new} = \text{UNVOICED}$  かつ  $class_{old} = \text{VOICED}$  ならば、 $class_{lost} = \text{UNVOICED TRANSITION}$  とする。

$class_{new} = \text{VOICED}$  または  $\text{ONSET}$  かつ  $class_{old} = \text{UNVOICED}$  ならば、 $class_{lost} = \text{ARTIFICIAL ONSET(onset reconstruction)}$  とする。

#### 7. 6. 4 励振信号の周期的な区間の生成

$class_{lost} = \text{UNVOICED}$  または  $\text{UNVOICED TRANSITION}$  ならば、励振信号の非周期的成分が生成される。他のクラスの場合、以下の手順に従い励振信号の周期的成分が生成される。

##### 最新のピッチ周期の繰り返し

直前のスーパーフレームの最新のピッチ周期が繰り返しコピーされる。最新のピッチサイクルを選択するために用いられるピッチ周期  $T_c$  は、倍ピッチあるいはハーフピッチを回避するように定義される。ピッチ周期  $T_c$  を以下の通り定義する。

$$\begin{aligned} & \text{if } (T_0 < 1.8T_s \text{ and } T_0 > 0.6T_s) \text{ or } (T_{cnt} \geq 30) \text{v} \\ & \quad T_c = T_0 \\ & \text{else} \\ & \quad T_c = T_s. \end{aligned}$$

ここで、 $T_0$  は正常に受信された最新のスーパーフレームの第4サブフレームの丸められたピッチ周期、 $T_{cnt}$  は最後に  $T_s$  が更新されてからのスーパーフレーム数をカウントしたもの、 $T_s$  は、コヒーレントピッチ推定で、最新の正常に受信された、安定した有声スーパーフレームの第4サブフレームの丸められた予測ピッチ周期である。ここで、安定した有声スーパーフレームとは、有音タイプのスーパーフレーム (“VOICED TRANSITION”、“VOICED”、“ONSET”) に続く”VOICED”スーパーフレームとして定義される。ピッチ安定性は、閉ループピッチ推定が適度に近いかどうか吟味すること、即ち、最終のサブフレームのピッチと、第2サブフレームのピッチと、直前のスーパーフレームの最終サブフレームのピッチとの間の比率が [0.7, 1.4] の範囲に収まっているかどうかによって評価する。複数のスーパーフレームが消失した場合は、 $T_0$  は最後に補償されたスーパーフレームの第4サブフレームのピッチ周期を丸めて計算する。

$T_c$  の計算の背景となる基本原理は以下の通りである。もし最新の正常受信スーパーフレームの最終のピッチと、最新の安定したスーパーフレームのピッチがお互いに近接していれば、最新の正常受信スーパーフレームのピッチが用いられる。逆に、音声の立ち上がり部においては、このピッチは信頼できないと見なし、誤ったピッチ評価による影響を回避するため、代わりに最新の安定したスーパーフレームのピッチを用いる。しかしながらこの論理は、最新の安定したセグメントが余り遠くない過去にある場合しか有効とはならない。 $T_{cnt} \geq 30$ 、即ち、最新の  $T_s$  が更新されてから少なくとも 30 スーパーフレーム経過している場合、機械的に最新の正常に受信されたスーパーフレームのピッチを用いる。安定したセグメントを検出する度に  $T_{cnt}$  が 0 にリセットされ、 $T_s$  が更新される。周期  $T_c$  は、連続する消失したブロックについて補償処理を行っている間は、定数として扱われる。

“UNVOICED”判定以外の、正常に受信されたスーパーフレームの後に続く消失スーパーフレームについては、CELP 励振信号は周期的な成分のみを用いて更新される。この更新信号は、次のスーパーフレームでピッチ符号帳励振信号を生成するために用いられる。

##### 声門パルスの再同期

励振信号を生成するために用いられるピッチ周期が符号器のピッチ周期と異なる可能性があるため、上記の手順では声門パルス位置がドリフトしてしまう恐れがある。その結果、適応符号帳 (あるいは直前の CELP 励振信号) が、実際の CELP 励振信号から同期外れを起こしてしまう。従って、消失スーパーフレームに続き正常スーパーフレームが受信された場合、ピッチ励振信号 (あるいは適応符号帳励振信号) が後続する数フレームに影響を及ぼしかねない誤りを内包し、正常に受信されたスーパーフレームの音声品質に悪影響を及ぼしかねない。



この問題を克服し復号器の収束を改善するため、実際の声門パルス位置に同期するように、補償されたサブフレームの最新の声門パルス位置を調整する、パルス位置の再同期の手法が用いられる。再同期の手順は、補償するスーパーフレームにおける最新の声門パルスの本当の位置を表す位相情報に基づいてなされる。位相情報は、後続のスーパーフレームのレイヤ3（伝送速度 14 kbit/s）で伝送される。しかしながら、後続のスーパーフレームからの情報が無効である場合、復号器にて、過去の最新の声門パルス位置を、声門パルスの本当の位置を表す位相情報と見なす。この手法は、後続のスーパーフレームがレイヤ1または2（伝送速度が8または12 kbit/s）しか受信できない場合、または後続のスーパーフレームも連続して消失してしまった場合に用いられる。

消失したスーパーフレームにおけるピッチ励振信号は、長さ  $T_c$  の最新のピッチ周期を繰り返すことによって生成される。現在のスーパーフレームが、正常に受信されたスーパーフレームの直後の最初の消失スーパーフレームである場合、このピッチ周期はまずローパスフィルタに掛けられる。ここで用いられるフィルタは、単純な3タップの線形位相 FIR フィルタで、係数は  $\{0.18, 0.64, 0.18\}$  である。フィルタ計算は次式の通りである。

$$\begin{aligned} u(n) &= 0.18u(n-T_c-1) + 0.64u(n-T_c) + 0.18u(n-T_c+1), & n=0, \dots, T_c-1 \\ u(n) &= u(n-T_c), & n=T_c, \dots, 199 \end{aligned} \quad (174)$$

ここで、 $u(n)$  は励振信号である。現在のスーパーフレームが最初の消失スーパーフレームではない場合、補償する励振信号は単純に次の式を用いて生成される。

$$u(n) = u(n-T_c), \quad n=0, \dots, L+N-1 \quad (175)$$

以下に示すように再同期手順を補助するため、エキストラサブフレームについても補償された励振信号が計算されることに注意されたい。

補償された励振信号  $u(n)$  が生成されたら、再同期手順が実行される。後続のスーパーフレームが有効、かつ声門パルス情報を含んでいる場合、この情報は復号される。この情報は、絶対値が最大のパルスについて、スーパーフレームの終端からの位置  $\tau$  及びパルスの極性で構成される。 $T_0 < 64$  である場合、受信した量子化位置  $\tau$  がそのまま用いられる。 $64 \leq T_0 \leq 128$  である場合、受信した量子化位置  $\tau$  は2倍され1加算される。また、 $T_0 \geq 128$  である場合、受信した量子化位置  $\tau$  は4倍され2加算される。

絶対値最大パルスの実際の位置は、次式で与えられる。

$$P_{last} = 160 - \tau \quad (176)$$

そして、（ローパスフィルタリングされた励振信号を基に）スーパーフレームの先頭から、補償された励振信号  $u(n)$  における復号された極性情報と同じ極性を持つ最大パルス位置が決定される。復号された最大パルス位置が正の場合、スーパーフレームの先頭から補償された励振信号における正の最大パルス位置が決定される。その逆の場合は、負の最大パルス位置が決定される。 $T(0)$  が補償された励振信号における第1最大パルスである場合、他の最大パルス位置は次式で与えられる。

$$T(i) = T(0) + iT_c, \quad i=1, \dots, N_p-1 \quad (177)$$

ここで、 $N_p$  はパルス数である（後続のスーパーフレームにおける第1パルスを含む）。

実際のパルス  $P_{last}$  に最も近いパルス  $T(i)$  を探索することによって、スーパーフレームにおける最後の補償されるパルスのパルス位置での誤差値が判る。誤差値は次式で与えられる。

$$T_e = P_{last} - T(k), \quad (178)$$

ここで、 $k$  は  $P_{last}$  に最も近いパルスのインデックスである。

$T_e = 0$  のとき、再同期処理は不要である。 $T_e \geq 0$  のとき、 $T_e$  サンプル挿入する必要がある。 $T_e \leq 0$  のとき、 $T_e$  サンプル分除去する必要がある。さらに、再同期処理は、 $T_e < 40$  かつ  $T_e < N_p \times T_{diff}$  のときのみ実行する。ここで  $T_{diff}$  は、 $T_c$  と後続のスーパーフレームにおける第 1 サブフレームのピッチラグとの差分の絶対値である。

挿入されるか、あるいは除去される必要があるサンプルは、スーパーフレームのピッチ周期全体に配分される。異なるピッチ周期での最小エネルギー領域が決定され、そしてサンプル挿抜がそれらの領域で実行される。スーパーフレームにおける、パルス位置  $T(i), i=0, \dots, N_p-1$  におけるピッチパルス数を  $N_p$  とする。最小エネルギー領域の数は  $N_p-1$  となる。最小エネルギー領域はスライドする 5 サンプル長の窓関数を用いてエネルギーを計算することにより決定される。最小エネルギー位置は、エネルギーが最小となる窓関数の中央に設定される。位置  $T(i)$  及び  $T(i+1)$  で与えられる 2 つのピッチパルスの間でなされる探索は、 $T(i)+T_c/4$  と  $T(i+1)-T_c/4$  との間に制限される。

サンプル挿抜処理は、 $T_{min}(i)$  周辺でなされる。ここで、 $T_{min}(i), i=0, \dots, N_{min}-1$  は上記の最小パルス位置であり、 $N_{min} = N_p - 1$  は最小エネルギー領域の数である。挿入されるか、あるいは除去されるサンプルは、次のように異なるピッチ周期全体に配分される。

$N_{min} = 1$  のとき、唯一の最小エネルギー領域が存在し、 $T_e$  個の全パルスに対して、位置  $T_{min}(0)$  にてサンプル挿抜される。 $N_{min} > 1$  のとき、各ピッチ周期において挿抜されるサンプル数を決定するのに用いられる単純なアルゴリズムを用いる。それは、スーパーフレームの先頭ではサンプル挿抜数は少なめとし、スーパーフレームの終端に向かって徐々に挿抜数を増やすというものである。挿抜するパルス数の合計を  $T_e$ 、最小エネルギー領域の数を  $N_{min}$  とすると、ピッチ周期当たり挿抜するサンプル数  $R(i), i=0, \dots, N_{min}-1$  は、次の漸化式を用いて計算される。

$$R(i) = \text{round} \left( \frac{(i+1)^2}{2} f - \sum_{k=0}^{i-1} R(k) \right) \quad (179)$$

ここで、 $f = \frac{2|T_e|}{N_{min}^2}$  である。

各ステージにおいて、 $R(i) < R(i-1)$  となる場合は  $R(i)$  と  $R(i-1)$  の値は置き換えられることに注意されたい。値  $R(i)$  はスーパーフレームの先頭から始まるピッチ周期に対応する。 $R(0)$  は  $T_{min}(0)$  に、 $R(1)$  は  $T_{min}(1)$  に、 $\dots$ 、 $R(N_{min}-1)$  は  $T_{min}(N_{min}-1)$  に、それぞれ対応する。 $R(i)$  は単調増加するため、スーパーフレームの終端に向かって徐々に挿抜サンプル数が増える。

サンプルの除去は単純である。サンプルの挿入は、前回の  $R(i)$  サンプルを 20 で割り極性を反転させたものを複写することによってなされる。例えば、位置  $T_{min}(0)$  において 5 サンプル挿入する必要がある場合、次式の通りとなる。

$$u(T_{min}(0)+i) = -u(T_{min}(0)+i-R(3))/20, \quad i=0, \dots, 4 \quad (180)$$

上記の手順を用いて、補償される励振信号における最新の最大パルスは、後続のスーパーフレームとして伝送される、スーパーフレームの終端における実際の最大パルス位置に強いて揃えようとする。

もし後続のスーパーフレームが無効ならば、消失したスーパーフレームのピッチ周期が推定され、その後、サブフレーム毎のピッチラグを推定するために、直前のピッチ周期と補間される。次に補償されたスーパー

フレームで用いられた最新のピッチ、及びサブフレーム毎に推定されたピッチの両者について、補償されたスーパーフレームにおける全ピッチ周期の合計の遅延時間が計算される。これら2つの合計遅延時間の差は、このスーパーフレームにおいて最近補償された最大パルスと推定されたパルスとの差分の推定値を与える。これら2つのパルスは、上記に基づき再同期処理がなされる。

直前のスーパーフレームの励振信号の最新のパルスが周期的区間の生成のために用いられるのと同様に、その利得も補償されるスーパーフレームの先頭でおよそ正しいと仮定して、1に設定される。次に利得値は、スーパーフレームの終端での値が $\alpha$ となるよう、スーパーフレーム全体にわたってサンプル毎に線形的に減衰される。

$\alpha$ の値は、有声区間のエネルギー変化を考慮した Table 1 6 / JT-G729.1 に対応する。最新の正常スーパーフレームの各サブフレームのピッチ励振利得値を用いることによって、この変化分が外挿される。通常、利得値が1以上の場合、信号エネルギーは単調増加とし、1未満の場合は単調減少とする。次に、先に述べたように $\alpha$ を $\beta = \sqrt{g_p}$ に設定する。急激なエネルギーの増大や減少が起きないようにするため、 $\beta$ は0.98と0.85の間に制限される。

UNVOICED 以外のスーパーフレームを受信した後に続く消失スーパーフレームについては、励振信号バッファは、(再同期処理及び利得調整処理を行った後) 励振信号の周期的成分のみで更新される。この更新は、次のスーパーフレームにおいてピッチ符号帳励振信号を生成するのに用いられる。

#### 7. 6. 5 フレーム消失補償処理におけるピッチ予測

位相情報が無効のとき(例えば、フレーム消失が2スーパーフレームに渡って連続して起きたとき)、FECにおけるピッチ予測が有効である。有音区間において1或いはそれ以上のスーパーフレームが消失したとき、消失した現在のスーパーフレームを再生するために必ず過去のピッチ情報が用いられる。現在の推定されたピッチの精度は、原信号への位相合わせに直接影響を与えるため、現在の消失したスーパーフレーム、及びその消失したスーパーフレームに続く正常受信されたスーパーフレームの再生信号の品質を左右する。直前の1ピッチラグを繰り返し用いるよりも、直前の複数のピッチラグを用いた方が、統計学的に良いピッチ推定ができる。JT-G729.1の符号器では、FECにおけるピッチ予測は、直前の5つのピッチに基づく線形予測によりなされる。ただし、直前の5つのピッチ値を $P(i)$ ,  $i=0, 1, 2, 3, 4$ . とし、 $P(0)$ を最も過去のものとする。予測モデルは次式で定義される。

$$P'(i) = a + ib \quad (181)$$

新たに予測される現在のピッチ値は次式の通りである。

$$P'(5) = a + 5b \quad (182)$$

係数 $a$ 及び $b$ を決めるため、次の誤差最小化の手法を用いる。

$$\begin{aligned} E &= \sum_{i=0}^4 [P'(i) - P(i)]^2 \\ &= \sum_{i=0}^4 [(a + b \cdot i) - P(i)]^2 \end{aligned} \quad (183)$$

式(183)において、下記の通り設定する。

$$\frac{\partial E}{\partial a} = 0 \quad \text{and} \quad \frac{\partial E}{\partial b} = 0 \quad (184)$$

式(183)、(184)より、以下の式が得られる。

$$a = \frac{3\sum_{i=0}^4 P(i) - \sum_{i=0}^4 iP(i)}{5} \quad \text{and} \quad b = \frac{\sum_{i=0}^4 iP(i) - 2\sum_{i=0}^4 P(i)}{10} \quad (185)$$

### 7. 6. 6 励振信号の乱数成分の生成

励振信号のイノベーション（非周期）成分は、ほぼ一様分布と見なせる単純な乱数生成器を用いてランダムに生成される。非周期成分をランダムに生成し、いくつかの参照値を用いてスケーリングし、サンプル当たりの単位エネルギーを確定した後、イノベーション成分の利得を調整する。

消失したブロックの先頭にて、最新の正常受信されたスーパーフレームの各サブフレームの励振信号のイノベーション成分利得を用いて、イノベーション成分の利得  $g_s$  が初期化される。

$$g_s = 0.1g(0) + 0.2g(1) + 0.3g(2) + 0.4g(3) \quad (186)$$

ここで、 $g(0)$ 、 $g(1)$ 、 $g(2)$  及び  $g(3)$  は、正常受信した最新のスーパーフレームにおける、固定符号帳、つまりイノベーション成分の4サブフレーム分の利得値である。イノベーション成分利得は、以下の通り減衰される。

$$g_s^1 = \alpha g_s^0 \quad (187)$$

ここで、 $g_s^1$  は、次のスーパーフレームの先頭におけるイノベーション成分の利得値、 $g_s^0$  は現在のスーパーフレームの先頭におけるイノベーション成分の利得値である。また、 $\alpha$  は Table 1 6 / JT-G729.1 で定義される係数である。励振信号の周期的成分を減衰させた方法と同様に、利得値は、スーパーフレーム全体にわたって、 $g_s^0$  から開始し、次のスーパーフレームの先頭における利得である  $g_s^1$  に達するようサンプル毎に線形的に減衰される。

最後に、最新の正常受信（誤り無く受信したかまたは消失していない）スーパーフレームが”UNVOICED”で無い場合、励振信号のイノベーション成分は、係数が-0.0125, -0.109, 0.7813, -0.109, -0.0125である、線形位相 FIR ハイパスフィルタにてフィルタリングされる。有音区間における雑音成分を緩和するため、これらのフィルタ係数に、 $0.75 - 0.25r_v$  で表現される適応成分を乗じる。ここで  $r_v$  は、式(15)で与えられる有声性に関する係数で、-1 から 1 の間の値をとる。励振信号の乱数成分は、励振信号の適応成分と加算され、トータルの励振信号となる。最新の正常受信スーパーフレームが”UNVOICED”の場合、励振信号のイノベーション成分のみが用いられ、さらに減衰係数が 0.8 で減衰される。この場合励振信号の周期的成分は得られないため、直前の励振信号バッファは励振信号のイノベーション成分のみで更新される。

### 7. 6. 7 スペクトル包絡の補償、合成及び更新

合成音声は、励振信号を線形予測合成フィルタに通すことによって得られる。フィルタ係数は LSF 表現から計算され、通常の符号化処理でなされると同様に、各サブフレーム（1スーパーフレーム当たり4回）について補間処理が施される。

後続のスーパーフレームが無効である場合、消失したスーパーフレームの LSF パラメータは、単純に直前のスーパーフレームと同じ値が設定される。

後続のスーパーフレームが有効である（単一のスーパーフレームが消失した）場合、後続するスーパーフレームと直前のスーパーフレームにおける情報を用いて LSP 値を内挿することによって、サブフレーム毎の線形予測フィルタパラメータが得られる。

補間は 3 段階で行われる。まず、消失したスーパーフレームの 2 番目の量子化係数セット  $\hat{l}_i^{(1)}$  を、後続するスーパーフレームと直前のスーパーフレームにおける量子化係数を用いて推定する。

$$\hat{l}_i^{(1)} = 0.8\hat{l}_i^{(0)} + 0.2\hat{l}_i^{(2)}, \quad i = 1, \dots, 10 \quad (188)$$

ここで、 $\hat{l}_i^{(0)}$  は、直前のスーパーフレームの 2 番目の量子化係数セット、 $\hat{l}_i^{(2)}$  は、後続のスーパーフレームの 2 番目の量子化係数セットである（ $\hat{l}_i$  の定義は、3. 2. 4 節/JT-G729 参照）。そして、現在の消失フレームの 2 番目のセットの推定 LSF  $\hat{\omega}_i^{(1)}$  は、式(20/JT-G729)と同様にして  $\hat{l}_i^{(1)}$  から生成される。推定された LSF は余弦領域に変換され、 $q_i^{(current)}$  を得る。最後に、各サブフレームに対する LSP は下記の関係を用いて得られる。

$$\begin{aligned} \text{Subframe 1: } & q_i^{(1)} = 0.45q_i^{(current)} + 0.55q_i^{(previous)} & i = 1, \dots, 10 \\ \text{Subframe 2: } & q_i^{(2)} = 0.8q_i^{(current)} + 0.2q_i^{(previous)} & i = 1, \dots, 10 \\ \text{Subframe 3: } & q_i^{(3)} = 0.96q_i^{(current)} + 0.04q_i^{(previous)} & i = 1, \dots, 10 \\ \text{Subframe 4: } & q_i^{(4)} = q_i^{(current)} & i = 1, \dots, 10 \end{aligned}$$

ここで、 $q_i^{(previous)}$  は、直前のスーパーフレームの最終サブフレームの量子化 LSP である。

イノベーション成分利得量子化及び LSF 量子化はともにフレーム間予測が用いられているため、通常処理が再開された直後は、予測メモリの内容は更新されない。この影響を抑えるため、消失した各々のスーパーフレームの終端にて、量子化器のフレーム間予測メモリの内容を推定し更新する。

## 7. 6. 8 フレーム消失補償処理後の通常処理への復帰

### 疑似立ち上がり再生

CELP 復号器において、長期予測の使用に関して最も複雑な状況となるのは、有声区間の立ち上がり部分が失われたときである。立ち上がり部が消失するということは、音声の立ち上がり部分が消失したブロックの区間のどこかで起きていることを意味する。この場合、最新の正常受信されたスーパーフレームは無声区間であるため、励振信号バッファに周期的成分がない。ところが、消失した直後に最初に受信した正常スーパーフレームは有声区間であるため、符号器の励振信号バッファは周期性が強く、励振信号の適応符号帳成分が、直前の周期的な励振信号を用いて符号化されている。復号器においては、この励振信号の周期的成分が完全に失われているため、この消失から回復するために複数のスーパーフレームが必要となる。

ONSET スーパーフレームが消失した場合（すなわち、消失スーパーフレームの直前の正常スーパーフレームが“UNVOICED”でありながら、消失スーパーフレームの直後に、正常な“VOICED”スーパーフレームを受信した場合）、消失した立ち上がり部を疑似的に再生し、有声部の合成のきっかけを生成するための特別な手段を用いる。補償されたスーパーフレームにおける最新の声門パルス位置は、後続のスーパーフレームより再生することができる（後続のスーパーフレームが消失せず、過去のスーパーフレームと関連する位相情報がその後続のスーパーフレームにおいて受信されている場合）。この場合、消失したスーパーフレームの補償処理は通常通りなされる。しかしながら、消失スーパーフレームの最新のパルスは、後続のスーパーフレームから得られる有効なパルス位置情報と極性情報に基づいて、疑似的に再生される。この情報は、スーパーフレームの終端からの最大パルス位置とその極性情報を含んでいる。このパルスにローパスフィルタを掛けることにより、消失スーパーフレームにおける最新の声門パルスが疑似的に再生される。パルスの極

性が正の場合、インパルス応答が  $h_{low} = \{-0.0125, 0.109, 0.7813, 0.109, -0.0125\}$  である、単純な線形位相 FIR フィルタがローパスフィルタとして用いられる。また、パルスの極性が負の場合、インパルス応答が  $-h_{low}$  である線形位相 FIR フィルタがローパスフィルタとして用いられる。

ローパスフィルタを通過したパルスは、ローパスフィルタのインパルス応答を（過去に零に初期化された）励振信号バッファの適応符号帳成分のメモリに置くことによって実現する。ローパスフィルタリングされた声門パルス（ローパスフィルタのインパルス応答）は、復号位置  $P_{last}$ （後続のスーパーフレームのビットストリーム内にて伝送される）の中心に置く。次に正常受信されたスーパーフレームを復号するとき、通常の復号処理が再開される。ローパスフィルタリングされた声門パルスを、補償されたスーパーフレームの終端の適切な位置に置くことにより、後に連続する正常受信スーパーフレームの音声品質が著しく向上し、実際の復号器の状態に収束する速度が速まる。

疑似立ち上がり部の励振信号の周期的成分のエネルギーは、フレーム損失補償区間の量子化され伝送されたエネルギーに相当する利得値を乗じ、その後線形予測合成フィルタの利得で除する。線形予測合成フィルタ利得は次の式で計算する。

$$g_{LP} = \sqrt{\sum_{n=0}^{39} h(n)} \quad (189)$$

ここで、 $h(n)$  は線形予測フィルタのインパルス応答である。最後に、周期成分に 0.96 を乗ずることにより、疑似立ち上がり部の利得を減衰させる。

出力音合成のための線形予測フィルタは、疑似立ち上がり再生時においては内挿処理を行わない。その代わりに、スーパーフレーム全体の音声合成に当たっては、受信した線形予測パラメータが用いられる。

## エネルギー制御

合成された音声信号のエネルギーを適切に制御することが、消失したスーパーフレームの直後の信号を回復させるための最も重要な処理となる。有声区間でスーパーフレームの消失が起きたとき、エネルギー制御は最も重要となる。有声のスーパーフレームの直後にスーパーフレームの消失が発生するとき、補償区間では、最新の正常受信スーパーフレームにおける励振信号に所定の減衰処理を施して用いられる。新しい線形予測フィルタが、消失の直後の正常受信スーパーフレームに達するとき、励振信号エネルギーと新しい線形予測合成フィルタの利得との間で不整合を起こす可能性がある。新しい合成フィルタは、最新合成された消失スーパーフレームのエネルギーとも、原信号のエネルギーとも、大幅に異なるエネルギーを持つ合成信号を生成する可能性がある。消失スーパーフレーム直後の正常受信スーパーフレームの区間におけるエネルギー制御は、以下のようにまとめることができる。合成された信号は、極端なエネルギーの増加を回避するため、そのエネルギーが直前の消失したスーパーフレームの終端の合成音声信号のエネルギー、および直後の正常受信スーパーフレームの先頭の合成信号のエネルギーに近い値になるようにスケーリングされることで、スーパーフレームの終端に向かって伝送エネルギーが収束する。

エネルギー制御は合成音声信号の範囲で行われる。励振信号もまた、後続するスーパーフレームのために長期予測メモリとしての役割を果たすことから、スケーリングされる必要がある。こうして再合成することによりスムーズに遷移する。 $g_0$  を、現在のスーパーフレームの第 1 サンプルのスケーリングに用いられた利得を、 $g_1$  を、スーパーフレームの終端のサンプルのスケーリングに用いた利得を示すこととする。励振信号は次式のようにスケーリングされる。

$$u_s(n) = g_{AGC}(n)u(n), \quad n = 0, \dots, 159 \quad (190)$$

ここで、 $u_s(n)$  はスケーリングされた励振信号、 $u(n)$  はスケーリングされる前の励振信号を、 $g_{AGC}(n)$  は、 $g_0$  から始まり、指数関数的に  $g_1$  に収束する利得である。

$$g_{AGC}(n) = f_{AGC}g_{AGC}(n-1) + (1-f_{AGC})g_1, \quad n = 0, \dots, 159 \quad (191)$$

なお、 $g_{AGC}(n)$  は  $g_{AGC}(-1) = g_0$  で初期化される。ここで、 $f_{AGC}$  は値が 0.98 に設定される減衰係数である。この値は、一方は（消失した）過去のスーパーフレームから滑らかに遷移させることと、もう一方では、正しい（送信された）値にできるだけ近くなるように、現在のスーパーフレームの直前のピッチ周期をスケーリングさせることの両方を満足させるよう、実験的に求められたものである。伝送されるエネルギー値がスーパーフレームの終端においてピッチに同期して推定されるため、このことは重要である。利得  $g_0$  と  $g_1$  は次のように定義される。

$$g_0 = \sqrt{E_{-1}/E_0}, \quad \text{and} \quad g_1 = \sqrt{E_q/E_1} \quad (192)$$

ここで、 $E_{-1}$  は過去の（消失した）スーパーフレームの終端において計算されるエネルギー、 $E_0$  は現在の（再生された）スーパーフレームの先頭にて計算されるエネルギー、 $E_1$  は現在のスーパーフレームの終端において計算されるエネルギー、そして  $E_q$  は符号器で計算され送信された、現在のスーパーフレームの終端における、量子化エネルギー情報である。 $E_{-1}$  と  $E_1$  は、合成された音声信号  $\hat{s}_{LB}^{celp}(n)$  を用いて計算されるという点を除いては、ともに同じような手法で計算される。 $E_{-1}$  はピッチ周期  $T_c$  の補償を用いて、ピッチ周期に同期して計算されるのに対し、 $E_1$  は直前のサブフレームの丸められたピッチ  $T_3$  を用いる。 $E_0$  も同様に、式(53)を修正したものを適用（"VOICED"及び"ONSET"スーパーフレームについて）し、丸められたピッチ周期  $T_0$  を用いて計算される。

$$E = \max_{i=0, \dots, t_E} \left[ \hat{s}_{LB}^{celp}(i) \right]^2 \quad (193)$$

ここで、 $T_0 < 64$  のとき、 $t_E = T_0$  または  $t_E = 2T_0$  である。他のスーパーフレームでは式(193)を下記の通り修正する。

$$E = \frac{1}{t_E} \sum_{i=160-t_E}^{159} \left[ \hat{s}_{LB}^{celp}(i) \right]^2 \quad (194)$$

ここで、 $t_E = 80$  である。さらに、利得  $g_0$  および  $g_1$  はエネルギー値が極端に大きくなるのを防ぐため、所定の値を上限として値を制限される。上限値は 1 に設定される。

$E_q$  が受信されない場合、 $E_q$  は  $E_1$  に設定される。有声音声区間において消失が発生した場合（例えば、消失の起きる直前の正常受信スーパーフレームと、消失直後の正常受信スーパーフレームが共に"VOICED TRANSITION"、"VOICED"、または"ONSET"のいずれかに区分されるものであった場合）、励振信号エネルギーと、先に述べた線形予測フィルタ利得との間に不整合が生じるため、さらに予防策が講じられる。消失スーパーフレームに続き最初に正常に受信されたスーパーフレームの線形予測フィルタの利得が、スーパーフレーム消失区間における最後の消失スーパーフレームの線形予測フィルタの利得よりも大きい場合、特に危険な状況が生じる。その特定の場合、最初に受信された正常なスーパーフレームにて復号器にて生成される、線形予測フィルタリングされた励振信号のエネルギーは、次の関係式を用いて、最初に受信された正常なスーパーフレームの線形予測フィルタの利得に補正される。

$$E_q = E_1 \frac{E_{LP0}}{E_{LP1}} \quad (195)$$

ここで、 $E_{LP0}$  は、消失直前の正常受信スーパーフレームの線形予測フィルタのインパルス応答のエネルギー、 $E_{LP1}$  は消失直後に正常受信されたスーパーフレームの線形予測フィルタのエネルギーである。ここで、線形予測フィルタには最終サブフレームのものを用いる。最後に、値  $E_q$  が、この場合だと  $E_1$  に制限される（伝送される  $E_q$  の情報なしに有声区間が消失した場合）。

さらに、音声信号の状態遷移に係わる全ての箇所について、 $g_0$  の計算に代わり、以下の例外処理を実行する。現在のスーパーフレームにおいて疑似立ち上がり部が用いられる場合、立ち上がり部のエネルギーが徐々に増加するよう、 $g_0$  を  $g_1/2$  に設定する。

消失直後の正常受信スーパーフレームが”ONSET”に分類される場合、利得  $g_0$  は  $g_1$  より大きくならないようにする。スーパーフレームの先頭（少なくともおそらくまだ部分的に無声である区間）において積極的に利得補正を行うことで、有声の立ち上がり部（スーパーフレームの終端）の増幅を回避するため、この予防策を実行する。

最後に、有声から無声に遷移する間（例えば、最新の正常スーパーフレームが”VOICED TRANSITION”、”VOICED”、または”ONSET”に分類され、かつ現在のスーパーフレームが”UNVOICED”に分類される場合）、 $g_0$  は  $g_1$  に設定される。

有声区間が消失した場合、消失直後の正常受信スーパーフレームに続くスーパーフレームにおいても、不適当なエネルギー値の問題に対する影響が現れることがある。このことは、たとえ消失直後の正常受信スーパーフレームのエネルギーに上記の補正を施したとしても起こりうる。この問題を減らすため、エネルギー制御は有声区間の終端まで続けられる。

#### 7. 6. 9 復号器における長期予測（LTP）利得の制限

スーパーフレームが消失した後、符号器と復号器の間の励振信号メモリの同期性が失われたり、一方よりも大きい値を持つピッチ利得が長時間続いたりすることにより、復号器側において、出力信号の飽和が引き起こされてしまうことがある。これを回避するため、TTC標準JT-G729符号器の抑圧手法と同様、飽和する危険性が高まったときに、長期予測利得を1に制限する手法を、復号器においても実行する。

直前のスーパーフレーム消失に起因する長期予測フィルタにおける累積的な誤差は、次の誤差表示関数により継続的に計算される。

$$x_i(n) = e_i(n) + g_i x_i(n-P) \quad (196)$$

ここで、：

- 受信したスーパーフレームにおいては、
  - ・  $g_i$  は復号したピッチ利得と同値とする。
  - ・  $e_i(n)$  は0に設定する。
  - ・  $P$  は復号されたピッチ周期の整数部とする。
- 受信されないスーパーフレームにおいては、
  - ・  $g_i$  は、FECアルゴリズムにより得られたピッチ利得と同値とする。
  - ・  $e_i(n)$  は1に設定する。
  - ・  $P$  は、FECアルゴリズムにより得られたピッチ周期の整数部とする。

スーパーフレームを受信したとき、各サブフレームについて誤り表示関数を用いて、誤り表示パラメータ  $S_i$  を求める。



$$S_i = \min\left(\sum_{n=0}^{39} x_i(n), 120\right) \quad (197)$$

$g_i > 1$ かつ $S_i > 80$ の場合は、ピッチ利得は式(198)に従い減衰され、その新たなピッチ利得 $g'_i$ が、与えられたサブフレームにおける励振信号の生成に用いられる。

$$g'_i = 1 + (g_i - 1)(120 - S_i)/40 \quad (198)$$

ここで、 $80 < S_i \leq 120$ であるので、新たなピッチ利得は $g'_i < g_i \leq 1$ となる。

## 7. 7 ビットレート切り換え

J T-G 7 2 9. 1 符号器は、狭帯域ビットレート (8 と 12 kbit/s) と広帯域ビットレート (14 kbit/s 以上) で動作する。適切な方法を用いずに、ビットレートの二つのセット間を高速に切り換えると、以下のような結果をもたらす。

- 低域合成信号  $\hat{s}_{LB}^{post}(n)$  は、8 と 12 kbit/s においては I I R 後処理により高域通過フィルタリングされており、14 kbit/s 以上においては無処理のため、狭帯域と広帯域ビットレート間の遷移において、明らかな位相問題を引き起こす。
- 狭帯域と広帯域の切り換え時に、高域信号  $\hat{s}_{HB}(n)$  が、急激に出現、消失するのは、非常に耳障りである。

これらの異音を回避するため、復号器はビットレート切り換えを操作する二つの特殊な機構を搭載している。それは、低域側の高域通過後処理のクロスフェーディングと、狭帯域から広帯域への緩やかな遷移 (1 秒) をもたらすフェードインである。

### 7. 7. 1 低域側後処理のクロスフェーディング

復号器において、低域側のクロスフェーディングは、現スーパーフレームのビットレートが前ビットレートと異なっている場合にのみ、適用される。但し、現スーパーフレームが消失した場合、そのビットレートは、前ビットレートと同一と仮定し、クロスフェーディングは適用されない。

クロスフェーディングの詳細を、Table19/JT-G729.1 に示す。クロスフェーディング窓は、パートレット窓 (三角窓) である。

Table19/JT-G729.1 Cross-fading operation  
(ITU-T G.729.1)

		current bit rate	
		8 or 12 kbit/s	14 kbit/s or above
Previous bit rate	8 or 12 kbit/s	$\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{hpf}(n)$	$\hat{s}_{LB}^{qmf}(n) = \frac{n}{159}\hat{s}_{LB}^{post}(n) + (1 - \frac{n}{159})\hat{s}_{LB}^{hpf}(n)$
	14 kbit/s or above	$\hat{s}_{LB}^{qmf}(n) = \frac{n}{159}\hat{s}_{LB}^{hpf}(n) + (1 - \frac{n}{159})\hat{s}_{LB}^{post}(n)$	$\hat{s}_{LB}^{qmf}(n) = \hat{s}_{LB}^{post}(n)$

### 7. 7. 2 狭帯域から広帯域への切り換え後の高域におけるフェードイン

狭帯域から広帯域への高域側遷移は、Figure14/JT-G729.1 に示されるフェードイン・ゲイン  $gain\_att$  に

よって制御される。8 と 12 kbit/s で復号されたスーパーフレームの後、14 kbit/s 以上で復号された 50 連続の 20ms スーパーフレームが、本来の広帯域出力を得るのに必要とされる。Figure14/JT-G729.1 に示されるように、狭帯域出力は、数スーパーフレームの間維持され、その後、広帯域出力への緩やかな遷移が得られる。

カウンター  $count\_rcv$  は、以下のように定義される。

- 復号器が 8 と 12 kbit/s で動作する毎、 $count\_rcv$  は 0 に設定される。
- 復号器の現在のビットレートが 14 kbit/s 以上の場合、 $count\_rcv$  は以下のように更新される。

$$count\_rcv = \min(count\_rcv + 1, COUNT\_RCV\_MAX)$$

ここで、 $COUNT\_RCV\_MAX = 50$  である。

スーパーフレームが消失した場合、カウンター  $count\_rcv$  は、更新されない。但し、広帯域から狭帯域へは、即時に遷移する。

ゲイン減衰  $gain\_att$  は、 $\hat{s}_{HB}^{gmf}(n)$  に適用される。

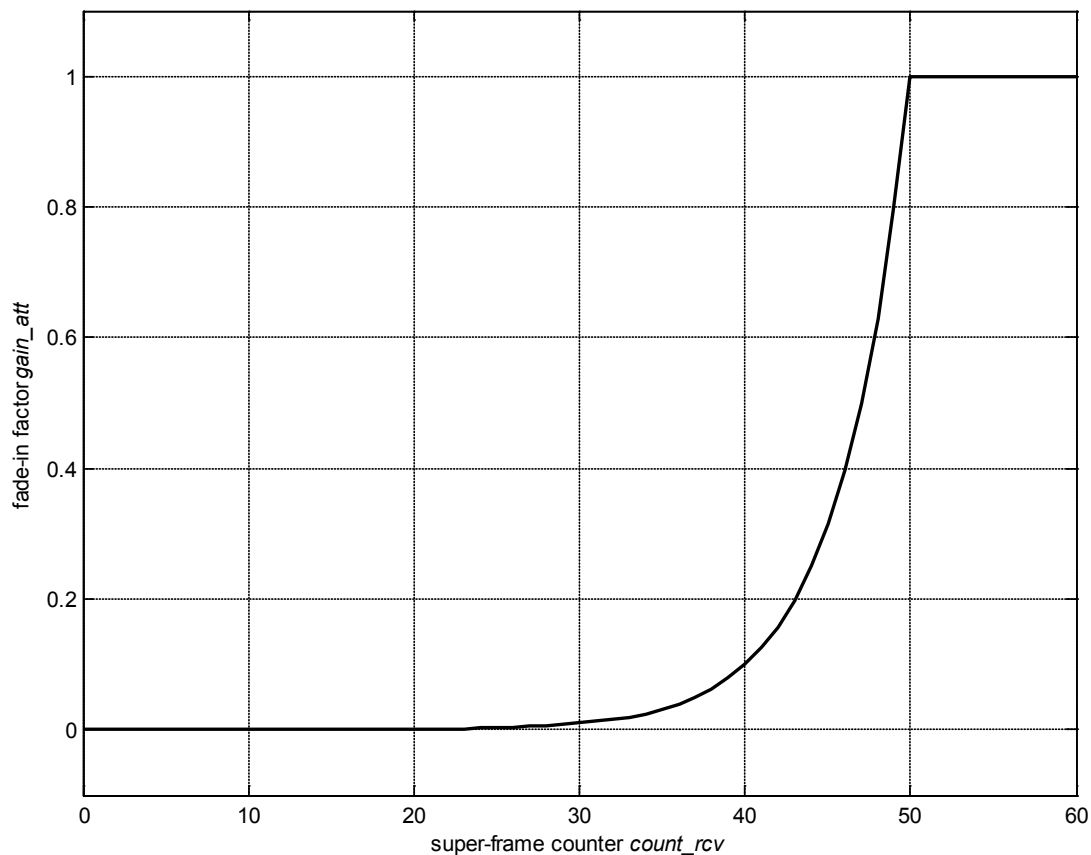


Figure14 /JT-G729.1 Fade-in factor applied on higher band after narrowband to wideband switch. (ITU-T G.729.1)

## 8 伝送パラメータ・インデックスの記述

ビットストリームの順番は、テーブル内の順番が反映される。MSB が最初に送出される。

Table20/JT-G729.1 Description of transmitted parameter indices.  
(ITU-T G.729.1)

Symbol	Description	Bits
<i>L0_1</i>	CELP8k – 1 <sup>st</sup> frame - Switched MA predictor of LSP quantizer	1
<i>L1_1</i>	CELP8k – 1 <sup>st</sup> frame - First stage vector of quantizer	7
<i>L2_1</i>	CELP8k – 1 <sup>st</sup> frame - Second stage lower vector of LSP quantizer	5
<i>L3_1</i>	CELP8k – 1 <sup>st</sup> frame - Second stage higher vector of LSP quantizer	5
<i>P1_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Pitch delay	8
<i>P0_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Parity bit for pitch delay	1
<i>C1_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Fixed codebook	13
<i>S1_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Signs of fixed-codebook pulses	4
<i>GA1_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Gain codebook index (stage 1)	3
<i>GB1_1</i>	CELP8k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Gain codebook index (stage 2)	4
<i>P2_1</i>	CELP8k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Pitch delay	5
<i>C2_1</i>	CELP8k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Fixed codebook	13
<i>S2_1</i>	CELP8k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Signs of fixed-codebook pulses	4
<i>GA2_1</i>	CELP8k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Gain codebook (stage 1)	3
<i>GB2_1</i>	CELP8k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Gain codebook (stage 2)	4
<i>L0_2</i>	CELP8k – 2 <sup>nd</sup> frame - Switched MA predictor of LSP quantizer	1
<i>L1_2</i>	CELP8k – 2 <sup>nd</sup> frame - First stage vector of quantizer	7
<i>L2_2</i>	CELP8k – 2 <sup>nd</sup> frame - Second stage lower vector of LSP quantizer	5
<i>L3_2</i>	CELP8k – 2 <sup>nd</sup> frame - Second stage higher vector of LSP quantizer	5
<i>P1_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Pitch delay	8
<i>P0_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Parity bit for pitch delay	1
<i>C1_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Fixed codebook	13
<i>S1_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Signs of fixed-codebook pulses	4
<i>GA1_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Gain codebook (stage 1)	3
<i>GB1_2</i>	CELP8k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Gain codebook (stage 2)	4
<i>P2_2</i>	CELP8k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Pitch delay	5
<i>C2_2</i>	CELP8k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Fixed codebook	13
<i>S2_2</i>	CELP8k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Signs of fixed-codebook pulses	4
<i>GA2_2</i>	CELP8k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Gain codebook (stage 1)	3
<i>GB2_2</i>	CELP8k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Gain codebook (stage 2)	4

Symbol	Description	Bits
<i>C'1_1</i>	CELP12k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe – Extra fixed codebook	13
<i>S'1_1</i>	CELP12k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe - Signs of extra fixed codebook pulses	4
<i>G'1_1</i>	CELP12k – 1 <sup>st</sup> frame - 1 <sup>st</sup> subframe – Extra gain index	3
<i>C'1_2</i>	CELP12k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Extra fixed codebook	13
<i>S'1_2</i>	CELP12k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Signs of extra fixed codebook pulses	4
<i>G'1_2</i>	CELP12k – 1 <sup>st</sup> frame - 2 <sup>nd</sup> subframe - Extra gain index	2
<i>CL1</i>	CELP12k – 1 <sup>st</sup> frame – class bit for FEC	1
<i>C'2_1</i>	CELP12k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Extra fixed codebook	13
<i>S'2_1</i>	CELP12k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Signs of extra fixed codebook pulses	4
<i>G'2_1</i>	CELP12k – 2 <sup>nd</sup> frame - 1 <sup>st</sup> subframe - Extra gain index	3
<i>C'2_2</i>	CELP12k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Extra fixed codebook	13
<i>S'2_2</i>	CELP12k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Signs of extra fixed codebook pulses	4
<i>G'2_2</i>	CELP12k – 2 <sup>nd</sup> frame - 2 <sup>nd</sup> subframe - Extra gain index	2
<i>CL2</i>	CELP12k – 2 <sup>nd</sup> frame - class bit for FEC	1
<i>MU</i>	TDBWE – time envelope mean	5
<i>T1</i>	TDBWE – time envelope 1 <sup>st</sup> stage vector	7
<i>T2</i>	TDBWE – time envelope 2 <sup>nd</sup> stage vector	7
<i>F1</i>	TDBWE – frequency envelope 1 <sup>st</sup> stage vector	5
<i>F2</i>	TDBWE – frequency envelope 2 <sup>nd</sup> stage vector	5
<i>F3</i>	TDBWE – frequency envelope 3 <sup>rd</sup> stage vector	4
<i>PH</i>	TDBWE – phase information for FEC	7

Symbol	Description	Bits
<i>E</i>	TDAC – energy information for FEC	5
<i>N</i>	TDAC – MDCT normalization factor	4
<i>RMS2</i>	TDAC – HB spectral envelope	<i>nbits_HB</i>
<i>RMS1</i>	TDAC – LB spectral envelope	<i>nbits_LB</i>
<i>VQ1</i>	TDAC – 1 <sup>st</sup> subband quantization index	<i>variable</i>
<i>VQ2</i>	TDAC – 2 <sup>nd</sup> subband quantization index	<i>variable</i>
<i>VQ3</i>	TDAC – 3 <sup>rd</sup> subband quantization index	<i>variable</i>
<i>VQ4</i>	TDAC – 4 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ5</i>	TDAC – 5 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ6</i>	TDAC – 6 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ7</i>	TDAC – 7 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ8</i>	TDAC – 8 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ9</i>	TDAC – 9 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ10</i>	TDAC – 10 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ11</i>	TDAC – 11 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ12</i>	TDAC – 12 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ13</i>	TDAC – 13 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ14</i>	TDAC – 14 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ15</i>	TDAC – 15 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ16</i>	TDAC – 16 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ17</i>	TDAC – 17 <sup>th</sup> subband quantization index	<i>variable</i>
<i>VQ18</i>	TDAC – 18 <sup>th</sup> subband quantization index	<i>variable</i>

## 9 JT-G729. 1の符号器のビットイグザクト記述

JT-G729. 1符号器を16ビット固定小数点でシミュレートするANSI Cコードは、ITU-TのWebサイトから入手できる。以下の節で、シミュレーションコードの使い方、ソフトウェアの構成を要約している。

### 9. 1 シミュレーションソフトウェアの使い方

Cコードは、符号器メインG729EV\_MAIN\_Encoder.cと復号器メインG729EV\_MAIN\_Decoder.cの二つから構成されている。

符号器のコマンドラインは、以下の通り。

#### **encoder [-options] inputfile bitstreamfile**

符号器オプションを以下に示す。

-rXXXXX : run encoder at XXXXX bit/s (by default: 32000)

-f8 : 8000 Hz sampled input

-g729\_bst : run encoder at 8 kbit/s and generate bitstream with JT-G729 format

デフォルトでは、入力サンプリングレートは16000Hzで、そのビットストリームは20msフレーム（あるいは、スーパーフレーム）に分割される。

復号器のコマンドラインは、以下の通り。

**decoder [-options] bitstreamfile outputfile**

復号器オプションを以下に示す。

- rXXXXXX : run decoder with maximal bit rate XXXXXX bit/s (by default: 32000)
- f8 : 8000 Hz sampled output
- ld : low-delay mode (the decoder bit rate must be limited to 8, 12 or 14 kbit/s)
- g729b\_bst : read and decode G729B bitstream

デフォルトでは、出力サンプリングレートは 16000Hz で、そのビットストリームは 20ms フレーム（あるいは、スーパーフレーム）に分割される。

入力ファイル、出力ファイルは 16 ビット PCM 信号でのデータファイルである。符号化ビットストリームのマッピングテーブルは、シミュレーションソフトウェアに含まれる。

## 9. 2 シミュレーションソフトウェアの構成

Table21/JT-G729.1 Tables of G729EV\_G729\_TAB\_LD8K.c  
(ITU-T G.729.1)

Table name	Size in 16-bit word	Description
hamwindow	240	LPC analysis window
lag_h	10	Lag window for bandwidth expansion (high part)
lag_l	10	Lag window for bandwidth expansion (low part)
table	65	Lookup table in LSF to LSP conversion and vice versa
slope_cos	64	Line slopes in LSP to LSP conversion
lspcb1	128 x 10	LSP quantizer (first stage)
lspcb2	32 x 10	LSP quantizer (second stage)
fg	2 x 4 x 10	MA predictors in LSP VQ
fg_sum	2 x 10	Used in LSP VQ
fg_sum_inv	2 x 10	Used in LSP VQ
grid	61	Grid points in LP to LSP conversion
inter_3	13	FIR filter for interpolating the correlation
inter_3l	31	FIR filter for interpolating the correlation
pred	4	MA gain prediction coefficients
gbk1	8 x 2	Codebook GA in gain VQ
gbk2	16 x 2	Codebook GB in gain VQ
map1	8	Used in gain VQ
map2	16	Used in gain VQ
coef	2 x 2	Used in gain VQ
L_coef	2 x 2 x 2	Used in gain VQ
thr1	4	Used in gain VQ
thr2	8	Used in gain VQ
imap1	8	Used in gain VQ

imap2	16	Used in gain VQ
tab_hup_s	28	Upsampling filter for postfilter
tab_hup_1	112	Upsampling filter for postfilter
tabpow	33	Lookup table in $2^x$ computation
tablog	33	Lookup table in base 2 logarithm computation
tabsqr	49	Lookup table in inverse square root computation
tab_zone	153	Table for taming procedure
freq_prev_reset	10	Used for previous LSP

Table22/JT-G729.1 Tables of G729EV\_MAIN\_table.c  
(ITU-T G.729.1)

Table name	Size in 16-bit word	Description
G729EV_MAIN_lp3k_b_hi	5	3kHz low-pass filter coefficients (high part)
G729EV_MAIN_lp3k_b_lo	5	3kHz low-pass filter coefficients (low part)
G729EV_MAIN_lp3k_a_hi	5	3kHz low-pass filter coefficients (high part)
G729EV_MAIN_lp3k_a_lo	5	3kHz low-pass filter coefficients (low part)
G729EV_MAIN_hp50_b	3	50Hz high-pass filter coefficients
G729EV_MAIN_hp50_a	3	50Hz high-pass filter coefficients
G729EV_MAIN_b100	3	High-pass postfilter coefficients
G729EV_MAIN_a100	3	High-pass postfilter coefficients
G729EV_MAIN_qmf_J64D	32	QMF filter coefficients
G729EV_MAIN_switching_gain	51	Fade-in attenuation of higher band for bitrate switching
G729EV_MAIN_tab_gain_enha	8	Quantization table for fixed-codebook gain quantization at 12 kbit/s
G729EV_MAIN_tab_gain_enha2bits	4	Quantization table for fixed-codebook gain quantization at 12 kbit/s
Sqrt_han_wind8k	81	Used in FEC
G729EV_FEC_h_low	5	Used in FEC (low part)
G729EV_FEC_h_high	5	Used in FEC (high part)
interpol_frac1	4	Used in FEC
interpol_frac2	4	Used in FEC
Sqi	7	Used in FEC
inv_sqi	7	Used in FEC
G729EV_TDAC_sb_bound	19	TDAC subband division
G729EV_TDAC_nb_coef	18	Number of MDCT coefficients by subband
G729EV_MAIN_maskBit	17	Masks for bitstream packing
G729EV_TDAC_len_huff_diff	25	Length of Huffman codes used by spectral envelope coding
G729EV_TDAC_code_huff_diff	25	Huffman codes used by spectral envelope coding
G729EV_TDAC_xcos	25	Table for MDCT computation
G729EV_TDAC_xsin	25	Table for MDCT computation
G729EV_TDAC_tab_map	32	Table for MDCT computation
G729EV_TDAC_tab_map2	32	Table for MDCT computation
G729EV_TDAC_tab_rev_ipp	6	Table for MDCT computation
G729EV_TDAC_tab_rev_i	6	Table for MDCT computation
G729EV_TDAC_rw1	80	Table for MDCT computation
G729EV_TDAC_rw2	80	Table for MDCT computation
G729EV_TDAC_wcos	80	Table for MDCT computation
G729EV_TDAC_wsin	80	Table for MDCT computation
G729EV_TDAC_wetr	80	Table for MDCT computation
G729EV_TDAC_weti	80	Table for MDCT computation



G729EV_TDAC_wetrm1	80	Table for MDCT computation
G729EV_TDAC_wetim1	80	Table for MDCT computation
G729EV_TDAC_h	320	Table for MDCT computation
G729EV_TDAC_nb_coef_sqrt	18	Square root of number of MDCT coefficients per band
G729EV_TDAC_nb_coef_div	18	Square root of number of MDCT coefficients per band
G729EV_MAIN_NbDic	17	Number of possible rates for VQ
G729EV_MAIN_dim_leader	84	Leaders dimension
G729EV_MAIN_list_lead8	33	List of dimension 8 leaders
G729EV_MAIN_Rate8	8	Possible bit allocations in dimension 8
G729EV_MAIN_NbLeadpRat8	8	Number of leaders of dimension 8 per rate
G729EV_MAIN_OffsetLead8	34 x 2	Offset for dimension 8 leaders
G729EV_MAIN_list_lead16	64	List of dimension 16 leaders
G729EV_MAIN_NbLeadpRat16	20	Number of leaders of dimension 16 per rate
G729EV_MAIN_Rate16	20	Possible bit allocations in dimension 16
G729EV_MAIN_OffsetLead16	65 x 2	Offset for dimension 16 leaders
G729EV_MAIN_leader1_norm	1	Leader of dimension 1
G729EV_MAIN_leader2_norm	28	Leaders of dimension 2
G729EV_MAIN_leader3_norm	33	Leaders of dimension 3
G729EV_MAIN_leader4_norm	44	Leaders of dimension 4
G729EV_MAIN_leader5_norm	40	Leaders of dimension 5
G729EV_MAIN_leader6_norm	18	Leaders of dimension 6
G729EV_MAIN_leader7_norm	14	Leaders of dimension 7
G729EV_MAIN_leader8_norm	80	Leaders of dimension 8
G729EV_MAIN_leader9_norm	18	Leaders of dimension 9
G729EV_MAIN_leader10_norm	10	Leaders of dimension 10
G729EV_MAIN_leader11_norm	22	Leaders of dimension 11
G729EV_MAIN_leader12_norm	12	Leaders of dimension 12
G729EV_MAIN_leader13_norm	13	Leaders of dimension 13
G729EV_MAIN_leader14_norm	28	Leaders of dimension 14
G729EV_MAIN_leader15_norm	30	Leaders of dimension 15
G729EV_MAIN_leader16_norm	224	Leaders of dimension 16
G729EV_MAIN_adLeader_norm	17	List of leaders
G729EV_MAIN_adListLead	17	List of leaders per dimension
G729EV_MAIN_delta_leader	84	Used in MDCT VQ
G729EV_MAIN_adRate	17	Used in MDCT VQ
G729EV_MAIN_adOffsetLead	17	Used in MDCT VQ
G729EV_MAIN_adNbLeadpRat	17	Used in MDCT VQ
G729EV_TDAC_tab_nb_prm_fact	17	Used in MDCT VQ
G729EV_TDAC_delta_decomp_val	17	Used in MDCT VQ
G729EV_TDAC_tab_decomp_val	102	Used in MDCT VQ

G729EV_TDAC_tab_decomp_fact	102	Used in MDCT VQ
G729EV_TDAC_tab_pow13	3	Used in MDCT VQ
G729EV_TDAC_tab_pow11	3	Used in MDCT VQ
G729EV_TDAC_tab_pow7	4	Used in MDCT VQ
G729EV_TDAC_tab_pow5	5	Used in MDCT VQ
G729EV_TDAC_tab_pow3	8	Used in MDCT VQ
G729EV_table_isqrt	49	Used in FEC
G729EV_TDBWE_MEAN_TIME_ENV_cb	32	Mean time envelope quantization
G729EV_TDBWE_TIME_ENV_cb	1024	Time envelope quantization
G729EV_TDBWE_FREQ_ENV_NUMBER_ENTRIES	3	Frequency envelope quantization
G729EV_TDBWE_FREQ_ENV_BITS	3	Frequency envelope quantization
G729EV_TDBWE_FREQ_ENV_cb1	128	Frequency envelope quantization
G729EV_TDBWE_FREQ_ENV_cb2	128	Frequency envelope quantization
G729EV_TDBWE_FREQ_ENV_cb3	64	Frequency envelope quantization
G729EV_TDBWE_FREQ_ENV_cb	3	Frequency envelope quantization
G729EV_TDBWE_center_bins	12	Frequency envelope extraction
G729EV_TDBWE_subband_window	3	Frequency envelope extraction
G729EV_TDBWE_frequency_envelope_window	128	Frequency envelope extraction
G729EV_TDBWE_fft_phs_tbl	64	Used for FFT computation
G729EV_TDBWE_fes_coeffs_matrix	221	Frequency envelope shaping
G729EV_TDBWE_pulse_shapes_frac6	342	Excitation generation
G729EV_TDBWE_TIME_ENVELOPE_SHAPING_WINDOW	5	Time envelope shaping

Table23/JT-G729.1 Tables of G729EV\_G729B\_TAB\_DTX.c  
(ITU-T G.729.1)

Table name	Size	Description
noise_fg_sum	20	Quantization of LSF vector
PtrTab_1	32	Quantization of LSF vector
PtrTab_2	32	Quantization of LSF vector
fact	3	Quantization of SID gain
marg	3	Quantization of SID gain
tab_Sidgain	32	Quantization of SID gain

Table 24/JT-G729.1 Summary of encoder specific routines.  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_acelp_ca.c	G729EV Search fixed codebook
G729EV_G729_lpc.c	G729 LP analysis
G729EV_G729_pitch.c	G729EV pitch search
G729EV_G729_pwf.c	G729 computation of perceptual weighting coefficients
G729EV_G729_qua_gain.c	G729 gain quantizer
G729EV_G729_qua_lsp.c	G729 LSP quantizer
G729EV_TDBWE_encoder.c	TDBWE encoder routine
G729EV_TDBWE_vector_quantization.c	TDBWE vector quantization
G729EV_TDAC_encod.c	TDAC encoder routine
G729EV_MAIN_encod.c	MAIN encoder routine
G729EV_CELP2S_acelp_ca.c	CELP2S fixed codebook search
G729EV_CELP2S_encod.c	CELP2S encoder routine
G729EV_FEC_ferenc.c	FEC encoder routine

Table 25/JT-G729.1 Summary of decoder specific routines.  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_de_acelp.c	G729 algebraic codebook decoding
G729EV_G729_dec_gain.c	G729 gain decoding
G729EV_G729_dec_lag3.c	G729 adaptive-codebook index decoding
G729EV_G729_lspdec.c	G729 LSP decoding
G729EV_G729_pst.c	G729 postfilter routines
G729EV_G729B_calcexc.c	G729B CNG decoder
G729EV_G729B_dec_sid.c	G729B SID decoder
G729EV_G729B_qlsidgain.c	G729B SID quantization
G729EV_G729B_tab_dtx.c	G729B DTX tables
G729EV_G729B_util.C	G729B utility routines
G729EV_TDBWE_compression.c	TDBWE post processing
G729EV_TDBWE_decoder.c	TDBWE decoder routine
G729EV_TDBWE_fir.c	TDBWE filter functions
G729EV_TDBWE_frequency_envelope_shaping.c	TDBWE frequency envelope shaping functions
G729EV_TDBWE_generate_excitation.c	TDBWE excitation generation functions
G729EV_TDBWE_time_envelope_shaping.c	TDBWE time envelope shaping functions
G729EV_TDAC_decod.c	TDAC decoder routine
G729EV_TDAC_post.c	TDAC post processing
G729EV_MAIN_decod.c	MAIN decoder routine
G729EV_MAIN_envadaption.c	MAIN pre/post echo reduction routines
G729EV_CELP2S_decod.c	CELP2S decoder routine
G729EV_CELP2S_syn.c	CELP2S core synthesis functions
G729EV_CELP2S_post.c	CELP2S postfiltering
G729EV_FEC_clasdec.c	FEC signal classification routines
G729EV_FEC_decbfi.c	FEC Frame Erasure Concealment functions
G729EV_FEC_ferdec.c	FEC information decoder
G729EV_FEC_onset.c	FEC codebook reconstruction routines
G729EV_FEC_pit_updt.c	FEC pitch update functions

Table 26/JT-G729.1 Summary of general routines.  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_error.c	G729 codebook error computation
G729EV_G729_filter.c	G729 filter functions
G729EV_G729_gainpred.c	G729 gain predictor
G729EV_G729_lpcfunc.c	G729 miscellaneous routines related to LP filter
G729EV_G729_lspgetq.c	G729 LSP quantizer
G729EV_G729_p_parity.c	G729 pitch parity computation
G729EV_G729_pred_lt3.c	G729 generation of adaptive codebook
G729EV_G729_tab_ld8k.c	G729 tables
G729EV_G729_util.c	G729 utility functions
G729EV_TDBWE_extract_frequency_envelope.c	TDBWE frequency envelope extraction function
G729EV_TDBWE_extract_time_envelope.c	TDBWE time envelope extraction functions
G729EV_TDBWE_fft.c	TDBWE FFT computation functions
G729EV_TDAC_bitalloc.c	TDAC bit allocation functions
G729EV_TDAC_lib_vq.c	TDAC VQ routines
G729EV_TDAC_mdct.c	TDAC IMDCT / MDCT functions
G729EV_TDAC_spectenv.c	TDAC spectral envelope coding functions
G729EV_TDAC_tfr.c	TDAC FFT computation functions
G729EV_TDAC_util.c	TDAC miscellaneous routines
G729EV_TDAC_vq.c	TDAC VQ routines
G729EV_MAIN_dspfunc.c	MAIN mathematical operations
G729EV_MAIN_filt.c	MAIN filter functions
G729EV_MAIN_oper_32b.c	MAIN 32 bits operations
G729EV_MAIN_prm.c	MAIN bit manipulation
G729EV_MAIN_table.c	MAIN codec tables
G729EV_CELP2S_put4pat.c	CELP2S tri-pulse pattern routine
G729EV_CELP2S_qua_gain.c	CELP2S gain quantizer
G729EV_FEC_ferutil.c	FEC miscellaneous functions
G729EV_FEC_tools.c	FEC tools library
G729EV_FEC_voicefac.c	FEC voicing factor

## LIST OF TABLES

Table 1: Glossary of acronyms	7
Table 2: Glossary of most relevant symbols	8
Table 3: Encoder/decoder modes	14
Table 4: Bit allocation (per 20 ms superframe)	16
Table 5: Complexity figures of the JT-G729.1 coder (encoder/decoder)	17
Table 6: Structure of the extra fixed codebook	23
Table 7: Scaling function coefficients of the classification parameters	29
Table 8: Signal classification rules at the encoder	30
Table 9: Subband division for the frequency envelope	33
Table 10: Time and frequency envelope quantization	34
Table 11: Subband boundaries and number of coefficients per subband in the TDAC coder.	38
Table 12: Huffman codes used in the TDAC spectral envelope encoder	40
Table 13: Possible bit allocations for embedded spherical vector quantization	41
Table 14: List of leaders in codebooks of dimension 8 (33 leaders)	43
Table 15: Parameters of the adaptive postfilter depending on bit rate	63
Table 16: Values of the FEC attenuation factor	66
Table 17: Signal classification parameters at the decoder and the coefficients	68
Table 18: Signal classification rules at the decoder	69
Table 19: Cross-fading operation	79
Table 20: Description of transmitted parameter indices	81
Table 21 : Tables of G729EV_G729_TAB_LD8K.c	85
Table 22: Tables of G729EV_MAIN_table.c	86
Table 23 : Tables of G729EV_G729B_TAB_DTX.c	88
Table 24: Summary of encoder specific routines	89
Table 25: Summary of decoder specific routines	90
Table 26: Summary of general routines	90

## LIST OF FIGURES

Figure 1: High-level block diagram of the encoder	11
Figure 2: High-level block diagram of the decoder	13
Figure 3: JT-G729.1 bitstream format (compliant with G.192)	15
Figure 4: Frequency responses of $H1(z)$ and $H2(z)$	19
Figure 5: High-level block diagram of the TDBWE encoder	31
Figure 6: Window for the frequency envelope computation	32
Figure 7: High-level block diagram of the TDAC encoder	35
Figure 8: High-level block diagram of the TDBWE decoder	47
Figure 9: Pulse shape lookup table	50
Figure 10: “Flat-top” Hanning window for the time envelope shaping	51
Figure 11: Filter-bank design for the frequency envelope shaping	52
Figure 12: Adaptive amplitude compression function	53
Figure 13: Block diagram of the TDAC decoder	54
Figure 14: Fade-in factor applied on higher band after narrowband to wideband switch	80

## 付属資料A パケットフォーマット、能力識別子および能力パラメータ

(標準 J T - G 7 2 9 . 1 に対する)

本付属資料は、T T C 標準 J T - H 2 4 5 で定義されたプロトコルを実装する端末で使用される場合の、J T - G 7 2 9 . 1 オーディオ符号化に対する R T P ペイロードフォーマットおよび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は、シームレスな相互接続を許容するため、対応する J T - G 7 2 9 . 1 R T P の規定と完全に互換性がある。

### A. 1 参考文献

下記の T T C 標準および他の参照すべき文献は、本標準での参照を通して本標準の規定を構成するものである。全ての標準および他の参照文献は、改定に従うものとする。従って、本標準のユーザには、以下の T T C 標準や他の参照すべき文献について、最新の版の適用の可能性を調査するよう奨励される。現在有効な T T C 標準および他の参照すべき文献は、定期的な出版されている。本標準内での文書の参照は、単独の文書としては、それを標準の扱いとはしない。

- (1) T T C 標準 J T - G 7 2 9 . 1 “JT-G729 ベースのエンベデッド可変ビットレート符号化:JT-G729 とビット列互換な 8-32kbit/s スケーラブル広帯域符号化”
- (2) T T C 標準 J T - H 2 4 5 “マルチメディア通信用制御プロトコル”
- (3) IETF RFC4749 “RTP Payload Format for the G.729.1 Audio Codec”
- (4) T T C 標準 J T - H 2 2 5 . 0 “パケットに基づくマルチメディア通信システムのためのシグナリングプロトコル とメディア信号のパケット化”
- (5) T T C 標準 J T - G 7 2 9 “8kbit/s CS-ACELP を用いた音声符号化方式”

### A. 2 J T - G 7 2 9 . 1 フレームに対するパケット構成

J T - G 7 2 9 . 1 オーディオコーデックに対する R T P ペイロードフォーマットは R F C 4 7 4 9 で規定されている。

ペイロードヘッダで定義されている M B S フィールドにより、受信可能な最大のコーデックビットレートのインバンドシグナリングが可能である。これは、J T - G 7 2 9 . 1 を双方向通信として使用するユニキャスト R T P セッションでのみ有用である。

J T - H 2 4 5 の Flow Control コマンドが、シグナリング経路上でのものであるが、同様な機能を提供している。したがって、完全な J T - H 2 4 5 環境においては、各 R T P パケットの M B S フィールドは 15 (=NO\_MBS) に設定すべきであり、ビットレート制御は Flow Control コマンドを通して実行されるべきである。

R F C 4 7 4 9 (例えば、S I P / S D P) を用いる非 J T - H 2 4 5 システムとの相互接続性を保証するために、M B S シグナリングが優先されなければならない。この状況は、S I P エンドポイントが、JT-H323-SIP ゲートウェイ<sup>1</sup>を用いて J T - H 3 2 3 端末に発呼した場合またはその逆の場合に起こりうる。R T P セッションの一方方向でインバンドの M B S が使用された場合には、逆方向でもそれが使用されなければならない。言い換えれば、エンドポイントが、有効かつ NO\_MBS とは異なる値の M B S 値を受信した場合には、その端末は、ビットレートの変更を要求する場合には、送信するパケットの M B S フィールドを使用しなければならない。

### A. 3 T T C 標準 J T - H 2 4 5 で用いる能力識別子およびパラメータ

<sup>1</sup>この規定により、メディア処理が不要となるため、JT-H323-SIP ゲートウェイの処理量を削減できる。



J T-G 7 2 9. 1 能力は、J T-H 2 4 5 の付録 VII に従い、J T-H 2 4 5 の generic capability として定義される。Table A.1/JT-G729.1 は J T-G 7 2 9. 1 の能力識別子を定義している。2 つの J T-G 7 2 9. 1 能力パラメータは、Table A.2/JT-G729.1 および Table A.3/JT-G729.1 にて定義される。

Table A.1/JT-G729.1 J T-G 7 2 9. 1 能力識別子  
(ITU-T G.729.1)

Capability name	JT-G729.1
Capability class	オーディオ
Capability identifier type	標準
Capability identifier value	{ itu-t (0) recommendation (0) g (7) 7291 generic-capabilities (1) }
maxBitRate	存在しなければならない。 フレーム内の実際のビットレートおよび MBS で指示されたビットレートのいずれもが超えてはならないビットレートを指定する。 とりうる値は、80, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, および 320 (100 bits/s 単位)である。 maxBitRate に対する一般的な値は 320 (= 32 kb/s)である。
collapsing	本フィールドは下記に示す JT-G729.1 の能力パラメータを含まなければならない
nonCollapsing	本フィールドは含まれてはならない
nonCollapsingRaw	本フィールドは含まれてはならない
Transport	本フィールドは含まれてはならない

Table A.2/JT-G729.1 J T-G 7 2 9 能力パラメータ - maxAL-sduAudioFrames  
(ITU-T G.729.1)

Parameter name	maxAL-sduAudioFrames
Parameter description	本パラメータは collapsing GenericParameter であり、AL-SDU 当たりの最大オーディオフレーム数を指定する。
Parameter identifier value	1
Parameter status	必須
Parameter type	unsignedMin
Supersedes	-

Table A.3/JT-G729.1 JT-G729能力パラメータ – mbs  
(ITU-T G.729.1)

Parameter name	mbs
Parameter description	本パラメータは collapsing GenericParameter であり、受信側での対応可能な現在の最大コーデックビットレートを指定する。  とりうる値は、80, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, および 320 (100 bits/s 単位)である。ただし、maxBitRate パラメータの値を超えてはならない。
Parameter identifier value	2
Parameter status	オプション  本パラメータが省略された場合には、maxBitRate と同じ値をとる。  本パラメータは RFC4749 との互換性のために定義され、Fast Connect 手順を用いるときに、SIP/SDP ベースのシステムとの互換性のために用いられる。  JT-H245 システムにおいては、Flow Control コマンドが代わりに用いられるべきである。
Parameter type	unsignedMin
Supersedes	-

#### A. 4 JT-G729との相互接続性

JT-G729.1のコア部はJT-G729と完全な相互接続性があるため、JT-G729.1能力に加えて、JT-G729能力も宣言することを推奨する。これにより、JT-G729.1エンドポイントとJT-G729エンドポイントとの間の相互接続性を保証する。

唯一の拘束条件は、JT-G729のフレーム長が10msであるのに対して、JT-G729.1は20msであることである。

優先順位としては、JT-G729.1はJT-G729の前に現れるべきである。その理由は、JT-G729.1は8kbit/sのビットレートでJT-G729と同等の音声品質であるが、8から32kbit/sまでのスケーラブルな音声品質改善を提供できるからである。

基本となるJT-G729能力はJT-H245のAudioCapability構造で定義される。JT-G729.1は、g729, g729AnnexA, g729wAnnexB および g729AnnexAwAnnexB と相互接続性がある (JT-G729.1は符号器側ではJT-G729.1付属資料Bで規定されたフレームは生成しないが、復号器ではそれを適切に復号できる)。

RFC4749で定義されるRTPペイロードフォーマットは、2つのJT-G729.1のエンドポイント間での通信での使用のみに限定されなければならない。JT-G729エンドポイントでは正しく認識されない。JT-G729.1エンドポイントとJT-G729エンドポイントとの間でオーディオチャネルを確立したい場合には、JT-H225.0で定義される、JT-G729ペイロードフォーマットを持つJT-G729のチャンネルでオープンする。

## 付属資料B JT-G729.1に対する浮動小数点演算での実装

(標準JT-G729.1に対する)

### B. 1 本付属資料の規定範囲

本付属資料は、浮動小数点演算に基づくJT-G729.1の別の実現方法についての記述であり、標準JT-G729.1と完全に相互接続性のあるものである。

本付属資料に対応する浮動小数点演算に基づく参照Cコードは、ITU-TのWebサイトから入手可能である。テストベクトルセットの設計は今後の課題である。

### B. 2 参考文献

下記のTTC標準および他の参照すべき文献は、本標準での参照を通して本標準の規定を構成するものである。全ての標準および他の参照文献は、改定に従うものとする。従って、本標準のユーザには、以下のTTC標準や他の参照すべき文献について、最新の版の適用の可能性を調査するよう奨励される。現在有効なTTC標準および他の参照すべき文献は、定期的に出版されている。

- (1) TTC標準JT-G729.1 “JT-G729 ベースのエンベデッド可変ビットレート符号化:JT-G729とビット列互換な8-32kbit/s スケーラブル広帯域符号化”
- (2) TTC標準JT-G729 “8kbit/s CS-ACELP を用いた音声符号化方式”

### B. 3 概要

TTC標準JT-G729.1は、JT-G729とビット列互換な8-32kbit/s スケーラブル広帯域符号化アルゴリズムをビットイグザクトな固定小数点で規定している。その仕様の詳細は、ビットイグザクトな固定小数点演算のCコードとしてITU-TのWebサイトから入手可能である。本付属資料は、浮動小数点演算に基づく標準JT-G729.1の別の実現方法を記述および定義するものである。

### B. 4 アルゴリズムの記述

標準JT-G729.1の浮動小数点版は、固定小数点版と同一のアルゴリズムステップを踏んでいる。同様に、浮動小数点版のビット列は、標準JT-G729.1の固定小数点版と全く同一である。アルゴリズムの詳細に関しては、標準JT-G729.1を参照すること。

### B. 5 ANSI Cコード

本付属資料にて定義されている標準JT-G729.1浮動小数点版をシミュレートするANSI Cコードは既に開発され本付属資料の対応物として入手可能である。本ANSI Cコードは本付属資料を規定するものである。Cコードによるアルゴリズム記述は、標準JT-G729.1のテキスト記述よりも優先される。JT-G729.1の改定により、より最新のバージョンが入手可能となる可能性があるため、最新の入手可能なバージョンを使用するように留意すること。

浮動小数点ソースコードの構成は、対応する固定小数点ソースコードと関連付けられている。CodecTypedef.hファイルに、全ての浮動小数点変数および定数を倍精度型または単精度型として定義した記述が含まれている。ソフトウェアファイル名とその概要のリストを、TableB.1/JT-G729.1からTableB.3/JT-G729.1に示す。基本演算や数値演算に関連するファイルは浮動小数点演算では使用されないことに留意すること。また、floatをshortに変換するルーチンがファイルG729EV\_TDAC\_util.cに追加されている。

Table B.1/JT-G729.1 Summary of encoder specific routines  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_ACELP_CA.c	G729EV Search fixed codebook
G729EV_G729_LPC.c	G729 LP analysis
G729EV_G729_PITCH.c	G729EV pitch search
G729EV_G729_PWF.c	G729 computation of perceptual weighting coefficients
G729EV_G729_QUA_GAIN.c	G729 gain quantizer
G729EV_G729_QUA_LSP.c	G729 LSP quantizer
G729EV_TDBWE_encoder.c	TDBWE encoder routine
G729EV_TDBWE_vector_quantization.c	TDBWE vector quantization
G729EV_TDAC_encod.c	TDAC encoder routine
G729EV_MAIN_encod.c	MAIN encoder routine
G729EV_CELP2S_acelp_ca.c	CELP2S fixed codebook search
G729EV_CELP2S_encod.c	CELP2S encoder routine
G729EV_FEC_ferenc.c	FEC encoder routine

Table B.2/JT-G729.1 Summary of decoder specific routines  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_DE_ACELP.c	G729 algebraic codebook decoding
G729EV_G729_DEC_GAIN.c	G729 gain decoding
G729EV_G729_DEC_LAG3.c	G729 adaptive-codebook index decoding
G729EV_G729_LSPDEC.c	G729 LSP decoding
G729EV_G729_PST.c	G729 postfilter routines
G729EV_G729B_CALCEXC.c	G729B CNG decoder
G729EV_G729B_DEC_SID.c	G729B SID decoder
G729EV_G729B_QSIDGAIN.c	G729B SID quantization
G729EV_G729B_TAB_DTX.c	G729B DTX tables
G729EV_G729B_UTIL.c	G729B utility routines
G729EV_TDBWE_compression.c	TDBWE post processing
G729EV_TDBWE_decoder.c	TDBWE decoder routine
G729EV_TDBWE_fir.c	TDBWE filter functions
G729EV_TDBWE_frequency_envelope_shaping.c	TDBWE frequency envelope shaping functions
G729EV_TDBWE_generate_excitation.c	TDBWE excitation generation functions
G729EV_TDBWE_time_envelope_shaping.c	TDBWE time envelope shaping functions
G729EV_TDAC_decod.c	TDAC decoder routine
G729EV_TDAC_post.c	TDAC post processing
G729EV_MAIN_decod.c	MAIN decoder routine
G729EV_MAIN_EnvAdaption.c	MAIN pre/post echo reduction routines
G729EV_CELP2S_decod.c	CELP2S decoder routine

Filename	Description
G729EV_CELP2S_syn.c	CELP2S core synthesis functions
G729EV_CELP2S_post.c	CELP2S postfiltering
G729EV_FEC_clasdec.c	FEC signal classification routines
G729EV_FEC_decbf.c	FEC Frame Erasure Concealment functions
G729EV_FEC_ferdec.c	FEC information decoder
G729EV_FEC_onset.c	FEC codebook reconstruction routines
G729EV_FEC_pit_updt.c	FEC pitch update functions

Table B.3/JT-G729.1 Summary of general routines  
(ITU-T G.729.1)

Filename	Description
G729EV_G729_ERROR.c	G729 codebook error computation
G729EV_G729_FILTER.c	G729 filter functions
G729EV_G729_GAINPRED.c	G729 gain predictor
G729EV_G729_LPCFUNC.c	G729 miscellaneous routines related to LP filter
G729EV_G729_LSPGETQ.c	G729 LSP quantizer
G729EV_G729_P_PARITY.c	G729 pitch parity computation
G729EV_G729_PRED_LT3.c	G729 generation of adaptive codebook
G729EV_G729_TAB_LD8K.c	G729 tables
G729EV_G729_UTIL.c	G729 utility functions
G729EV_TDBWE_extract_frequency_envelope.c	TDBWE frequency envelope extraction function
G729EV_TDBWE_extract_time_envelope.c	TDBWE time envelope extraction functions
G729EV_TDBWE_fft.c	TDBWE FFT computation functions
G729EV_TDAC_bitalloc.c	TDAC bit allocation functions
G729EV_TDAC_lib_vq.c	TDAC VQ routines
G729EV_TDAC_mdct.c	TDAC IMDCT / MDCT functions
G729EV_TDAC_spectenv.c	TDAC spectral envelope coding functions
G729EV_TDAC_tfr.c	TDAC FFT computation functions
G729EV_TDAC_util.c	TDAC miscellaneous routines
G729EV_TDAC_vq.c	TDAC VQ routines
G729EV_MAIN_filt.c	MAIN filter functions
G729EV_MAIN_prm.c	MAIN bit manipulation
G729EV_MAIN_Table.c	MAIN codec tables
G729EV_CELP2S_put4pat.c	CELP2S tri-pulse pattern routine
G729EV_CELP2S_qua_gain.c	CELP2S gain quantizer
G729EV_FEC_ferutil.c	FEC miscellaneous functions
G729EV_FEC_tools.c	FEC tools library
G729EV_FEC_voicefac.c	FEC voicing factor

## 付属資料C DTX/CNG手法

(標準JT-G729.1に対する)

### C.1 本標準の規定範囲

本付属資料は、JT-G729.1における無音圧縮拡張のためのエンベデッド不連続伝送(DTX)、無音挿入記述子(SID)、擬似背景雑音発生器(CNG)について取り扱っている。

本付属資料の構成は次の通りである。本付属資料を通して使用されている文献、用語の定義、略語、表記法については、C.2節、C.3節、C.4節、C.5節のそれぞれで定義している。C.6節では無音圧縮手法の概略について述べている。無音圧縮符号化/復号の原理についてはC.7節、C.8節のそれぞれで述べている。符号化メモリの更新についてはC.9節で概説し、伝送パラメータについてはC.10節で述べている。C.11節では、本付属資料を16-32ビット固定小数点演算により定義したソフトウェアについて記載している。

### C.2 参考文献

下記のTTC標準およびITU-T勧告は、本標準での参照を通して本標準の規定を構成するものである。出版された時点でその版が適用され、全ての標準および他の参照文献は、改定に従うものとする。従って、本標準のユーザには、以下のTTC標準やその他の参照すべき文献について、最新の版の適用の可能性を調査するよう奨励される。現在有効なTTC標準およびITU-T勧告のリストは定期的には出版されている。本標準内での文書の参照は、単独の文書としては、それを標準の扱いとはしない。

- (1) ITU-T勧告G.191  
Software tools for speech and audio coding standardization
- (2) ITU-T勧告G.192  
A common digital parallel interface for speech standardization activities
- (3) TTC標準JT-G729  
8kbit/s CS-ACELP を用いた音声符号化方式
- (4) TTC標準JT-G729付属資料A  
低演算量版 8 kbit/s CS-ACELP 音声コーデック
- (5) TTC標準JT-G729付属資料B  
ITU-T勧告V.70 端末に適した標準JT-G729に対する無音圧縮手法
- (6) TTC標準JT-G729.1  
JT-G729ベースのエンベデッド可変ビットレート符号化：JT-G729とビット列互換な  
8-32kbit/s スケーラブル広帯域符号化
- (7) TTC標準JT-G722.2  
適応マルチレート広帯域 (AMR-WB) 方式を用いた 16kbit/s 程度の広帯域音声符号化

### C.3 定義

本標準では特に用語の定義を行っていない。

### C.4 略語と頭字語

本付属資料では、Table C.1/JT-G729.1 で定義される略語と頭字語を使用している。本付属資料で使用されている記号については後述のC.5節を参照すること。

Table C.1/JT-G729.1 – Glossary of acronyms  
(ITU-T G.729.1)

Acronym	Description
ACELP	Algebraic CELP
CELP	Code-Excited Linear Prediction
CNG	Comfort Noise Generator
DEMUX	DEMULTipleXer
DTX	Discontinuous Transmission
FEC	Frame Erasure Concealment
FIR	Finite Impulse Response
HB	Higher Band
LB	Lower Band
LPC	Linear Prediction Coding
LSF	Line Spectrum Frequency
LSP	Line Spectrum Pair
MDCT	Modified Discrete Cosine Transform
MSB	Most Significant Bit
MUX	Multiplexer
NT	Non Transmission
PCM	Pulse Code Modulation
QMF	Quadrature Mirror Filterbank
SID	Silence Insertion Descriptor
TDAC	Time-Domain Aliasing Cancellation
TDBWE	Time-Domain BandWidth Extension
VQ	Vector Quantization
WB	Wideband
WMOPS	Weighted Million Operations Per Second

### C. 5 表記法

J T - G 7 2 9 標準の記述との整合を取るため、J T - G 7 2 9 . 1 の無音圧縮で用いられる 20ms フレームをスーパーフレームと呼び、狭帯域処理での 10ms フレームをフレームと呼ぶ。

本標準では、J T - G 7 2 9、J T - G 7 2 9 付属資料B、J T - G 7 2 9 . 1 の表記法を、適用できる場合はそのまま使用している。表記法の詳細は次の通りである。

- (1) コードブックは、カリグラフ文字で記述する(例  $\mathcal{C}$ )。
- (2) 時間領域の信号は、そのシンボルと丸括弧で括られたサンプル番号で記述する(例  $s(n)$ )。変数  $n$  は、サンプル番号である。
- (3) 周波数領域に変換された信号は、対応する時間領域の信号を大文字に変えることにより記述する (例  $S(k)$ は  $s(n)$ の変換)。変数  $k$  は、係数の番号である。
- (4) 丸括弧で括られた上付きの添字は、時間に依存する変数に用いる(例  $g^{(m)}$ )。変数  $m$  はその前後関係よりフレーム番号、あるいはサブフレーム番号に対応している。
- (5) 再帰を示す添字は、角括弧で括られた上付きで記述する(例  $E^{[k]}$ )。
- (6) 下付きの添字は、係数配列の各要素を示す。

- (7) 記号  $\hat{\cdot}$  は量子化されたパラメータを示す(例  $\hat{g}_c$ )。
- (8) パラメータの範囲は、角括弧で括られた値で記述する。この値は境界値を含む(例  $[0.6, 0.9]$ )。
- (9) 関数  $int()$ は、切り捨てによる整数値への変換を示す。
- (10) 関数  $even()$ は、引き数が偶数の整数値の場合 1 を、そうでない場合 0 を返す。
- (11) 関数  $round()$ は、最も近い整数値への丸めを示す。
- (12) 記号  $\lfloor \cdot \rfloor$  は小数部の切り捨てを示す。
- (13) 使用される 10 進の浮動小数点値は、16 ビット固定小数点 ANS I C での実現に使用された値を丸めたものである。

Table C.2/JT-G729.1 は、本標準で用いられている主な記号のリストである。

Table C.2/JT-G729.1 – Glossary of most relevant symbols  
(ITU-T G.729.1)

Name	Description
$dtx_{lb}^s$	LB DTX indicator
$dtx_{1st}^f$	DTX indicator of first frame in one superframe
$dtx_{2nd}^f$	DTX indicator of second frame in one superframe
$Ftyp_t^s$	Superframe type
$Ftyp_t$	Frame type
$dtx_{hb}^s$	HB DTX indicator
$flag_{cmb}$	Combined DTX flag indicates if combined DTX is needed
$dtx^s$	Combined DTX indicator
$C_f$	Weighted Itakura distance
$C_g$	Weighted energy difference
$d$	Combined difference measure
$count\_fr$	Counter of elapsed superframes since the last SID
$\tilde{T}^{sid}$	Time envelope of the latest SID superframe
$\tilde{T}^m$	The smoothed time envelope of the superframe $m$
$\tilde{F}_{env}^{sid}(i)$	The $i^{th}$ frequency envelope of the last SID superframe
$\tilde{F}_{env}^m(i)$	The smoothed $i^{th}$ frequency envelope of the superframe $m$
$T^{ltx}$	Last transmitted time envelope parameter
$F_{env}^{ltx}(j)$	Last transmitted frequency envelope parameter
$r^f(j)$	Average frame autocorrelation function
$R_a(j)$	Autocorrelation of the SID filter coefficients
$E_t$	Residual energy of the frame
$r'_{m,k}(j)$	Modified autocorrelation function of $k^{th}$ frame in $m^{th}$ superframe
$\bar{R}_p(j)$	Averaged autocorrelation function for past five superframes
$E_{sid}^{lt}$	SID energy parameter
$r_n(j)$	Latest $n^{th}$ frame autocorrelation function
$Norm_n$	2-norm of latest $n^{th}$ frame autocorrelation function



Name	Description
$r_{mid1}(j)$	First of selected frame autocorrelation functions with the middle 2-norm value
$r_{mid2}(j)$	Second of selected frame autocorrelation functions with the middle 2-norm value
$E_{1st}^f$	Residual energy of the first 10 ms frame in one superframe
$E_{2nd}^f$	Residual energy of the second 10 ms frame in one superframe
$r_{1st}^f(j)$	Autocorrelation functions of the first 10 ms frame in one superframe
$r_{2nd}^f(j)$	Autocorrelation functions of the second 10 ms frame in one superframe
$r^s(j)$	Autocorrelation functions of the processing superframe
$E^s$	Superframe energy parameter
$F_{env}^m(j)$	Frequency envelope parameters at superframe $m$
$T_{env}^m(i)$	Time envelope parameters at superframe $m$
$\hat{\omega}_i$	Quantized LSF vector
$\omega'_i$	Target quantization vector for third stage LSF quantization
$Y_3$	Third stage codebook vector
$Y_2$	Mapped codebook vector from G.729 second stage
$H$	Scalar factor codebook for stage 3 quantization
$E_{idx}$	Energy quantization index
$E_{LSF3}(m)$	Weighted quantization error
$W_i$	Weighting vector
$\hat{\omega}_i^{lt}$	Long term averaged reconstructed spectrum parameters
$Gf^{lt}$	Long term fixed codebook gain of hangover
$\hat{g}_c$	8 kbit/s fixed codebook gain
$\hat{g}_{enh}$	12 kbit/s fixed codebook gain
$E_{1st}$	Residual energy of first frame in first SID after hangover
$E_{sid}$	Residual energy parameter derived from latest SID
$\hat{\omega}_i^{sid}$	Spectrum parameters derived from latest SID
$d_{sid}^{lt}$	Long term SID interval in number of frames
$d_{sid}$	Distance between latest two SID in number of frames
$\hat{E}(k)$	Reconstructed frame energy parameter of $k^{th}$ frame
$ACT_{sw}^{lt}$	Averaged activity changing rate
$ACT_{sw}$	Number of activity indication changes in two seconds
$Hg_E$	Distance between current superframe and first SID after hangover
$pitch_{rnd}$	Random pitch lag for CNG
$pitch_{old}$	Pitch lag of previous frame
$pitch_{new}$	Pitch lag of current frame
$gp_{rnd}$	Random pitch gain
$gp_{old}$	Pitch gain of previous frame
$gp_{new}$	Pitch gain of current frame

Name	Description
$exc_{new}$	Reconstructed excitation
$exc_{pre}$	Excitation of previous frame
$exc_{cur}$	Transition excitation for first frame after hangover
$\hat{T}_{env}^{lt}$	Long term time envelope parameter for CNG
$\hat{F}_{env}^{lt}$	Long term frequency envelope parameter for CNG
$\hat{F}_{env}^w(j)$	Windowed frequency envelope parameter
$\hat{T}_{env}^{sid}$	Time envelope derived from SID update
$\hat{F}_{env}^{sid}(j)$	Frequency envelope derived from SID update
$\hat{T}_{env}^{sp}$	Time envelope of active superframe
$\hat{F}_{env}^{sp}(j)$	Frequency envelope of active superframe
$\hat{s}_l(n)$	Reconstructed lower band components
$\hat{T}_{env}^{att}$	Attenuated time envelope after FER
$\hat{F}_{env}^{att}(j)$	Attenuated frequency envelope after FER
$Par^k$	HB parameter of $k^{th}$ superframe after bandwidth switching occurs, either time envelope or frequency envelope
$Par_{pre}$	HB parameter of previous SID, either time envelope or frequency envelope
$Par_{sid}$	HB parameter of latest SID, either time envelope or frequency envelope
$\hat{F}_{env}^k(j)$	Fade-out processed frequency envelopes of $k^{th}$ superframe after bandwidth switching occurs
$\bar{A}_p(z)$	Past LPC filter
$A_t(z)$	Current superframe LPC filter
$A_{sid}(z)$	LPC filter will be transmitted in SID
$H_{sh}(z)$	Excitation shaping filter

### C. 6 DTX、SID、CNGの構成要素の説明

J T-G 7 2 9. 1 符号器は、J T-G 7 2 9 の 8-32kbit/s スケーラブル広帯域 (50-7000Hz) 拡張である。デフォルトでは符号器の入力と復号器の出力は 16000Hz にサンプリングされている。符号器が出力するビット列はスケーラブルであり、2つの狭帯域レイヤと 10 個の広帯域拡張レイヤから成る。ここで、第1のレイヤ (コアレイヤとも書く) は 8kbit/s のビットレートに相当し、J T-G 7 2 9 のビット列に従う。

本付属資料では、J T-G 7 2 9. 1 の無音圧縮手法に用いられる DTX、SID、CNG の構成要素について詳述している。ここで J T-G 7 2 9. 1 は、J T-G 7 2 9 付属資料 B との互換性を持つコアレイヤを有している。

J T-G 7 2 9. 1 と同様に、この無音圧縮手法は、16-bit リニア PCM への変換によりサンプリングされた 16000Hz のデジタル信号を符号器の入力として、処理を行うように設計されている。但し、8000Hz サンプリングの入力もまたサポートされている。同様に復号器の出力のフォーマットは、8000Hz もしくは 16000Hz サンプリングの 16-bit リニア PCM である。入出力のフォーマットがその他のフォーマットである場合、符号化の前に 8000Hz もしくは 16000Hz サンプリングの 16-bit リニア PCM に変換したり、復号の後に 16-bit リニア PCM から所望のフォーマットに変換したりすることが必要となる。符号器から復号器に渡すビット列については本付属資料で定義されている。

J T-G 7 2 9. 1 と同様に、無音圧縮手法は 20ms のフレーム単位で処理が行われる。但し、狭帯域では J T-G 7 2 9 付属資料 B のように 10ms のフレーム単位で処理が行われる。C. 6. 1 節、C. 6. 2 節では、無音圧縮符号器／復号器の全般的な特徴を説明している。C. 6. 3 節では、その他の異なる処理モードについて記述している。ビットアロケーション、遅延、演算量については C. 6. 4 節、C. 6. 5 節、C. 6. 6 節で詳述している。

### C. 6. 1 無音圧縮符号器

J T-G 7 2 9. 1 の無音圧縮符号器の機能説明を Figure C.1/JT-G729.1 に示す。有音、無音での符号化手法を含む、完全な無音圧縮手法の全体的な説明については Figure B.1/JT-G729 を参照すること。

入力信号は、QMF フィルタにより低域 (LB) 成分と高域 (HB) 成分とに分割される。LB、HB のパラメータが抽出され、LB の DTX 尺度と HB の DTX 尺度を生成する目的で用いられる。両方の DTX 尺度と付加的方法により得られる DTX 判定を結合したものが、最終的な DTX 判定となる。

背景雑音特性の更新を伝送する必要がある場合、結合された DTX 判定を 1 にセットして指示し、LB、HB パラメータを別々に量子化し、それらを結合して S I D フレームを生成し、これを復号器に送る。

一方、DTX 判定がリセットされた (すなわち 0 にセットされた) 場合、符号器は復号器に何ら情報の伝送を行わない (非伝送 (NT)) 。

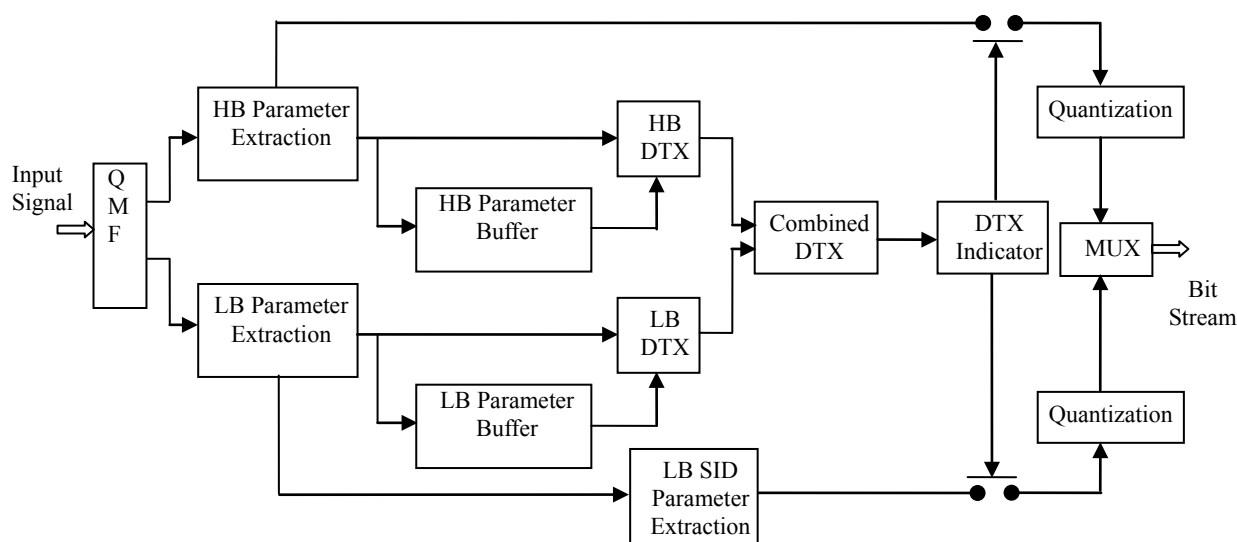


Figure C.1/JT-G729.1– High-level block diagram of the silence compression encoder (ITU-T G.729.1)

### C. 6. 2 無音圧縮復号器

J T-G 7 2 9. 1 の無音圧縮復号器の機能説明を Figure C.2/JT-G729.1 に示す。復号は、符号器より S I D フレームを受け、LB ビット列と HB ビット列に分割する。これらは別々に復号され、それぞれ LB と HB の擬似背景雑音を生成するために用いられる。LB、HB の擬似背景雑音 (Figure 2/JT-G729.1 中の  $\hat{S}_{LB}^{qm f}(n)$  と  $\hat{S}_{HB}^{qm f}(n)$ ) は、出力信号を生成するためにアップサンプリング処理と QMF フィルタ処理が施される。

S I D フレームが受信されない (非伝送) 場合、復号器は以前の伝送パラメータから背景雑音パラメータを推定する。

無音圧縮復号器は、最初の S I D フレームの損失に対する処理とビットレートスイッチングに対する処理

が組み込まれている。

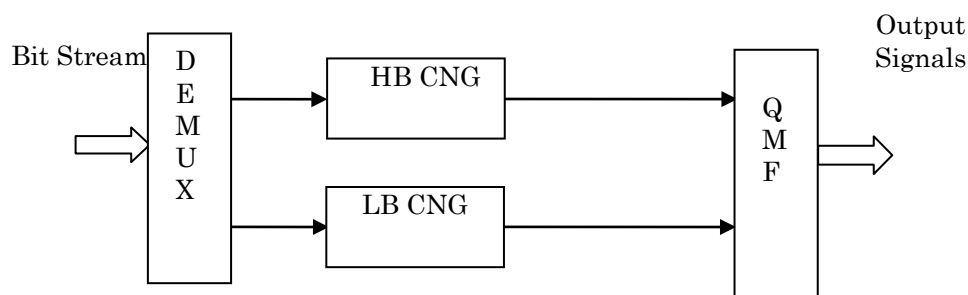


Figure C.2/JT-G729.1 – High-level block diagram of the silence compression decoder (ITU-T G.729.1)

### C. 6. 3 符号化モード

J T - G 7 2 9 . 1 D T X / C N G 符号器は、柔軟なアーキテクチャを有する J T - G 7 2 9 . 1 本体をベースとしている。この符号器は、J T - G 7 2 9 . 1 と同様に 16000Hz サンプリングの信号だけではなく、8000Hz サンプリングの信号も入出力として扱うことができる。Table C.3/JT-G729.1 は、J T - G 7 2 9 . 1 D T X / C N G で取り扱うことが可能なモードを示している。DEFAULT モードは J T - G 7 2 9 . 1 D T X / C N G がデフォルトで行う処理モードであり、入力信号と出力信号は 16000Hz にサンプリングされている。

以下の付加的な符号化モードがある：

- (1) NB\_INPUT モードは、符号器の入力が 8000Hz サンプリングであり、QMF 分析フィルタバンクは迂回される。
- (2) DTX モードは、DTX/CNG の手法を用いて符号化処理を行う。3 レイヤを含む無音のスケラブルビット列が生成される。
- (3) G729B\_BST モードは、符号器は 8000Hz で動作し、10ms のフレームを用いて J T - G 7 2 9 付属資料 B のビット列を生成する。デフォルトでは符号器の入力は 16000Hz サンプリングである。NB\_INPUT モードが同時にセットされている場合、入力信号は 8000Hz サンプリングになる。

以下の三つの復号モードは本付属資料でも利用可能である：

- (1) NB\_OUTPUT モードは、復号の出力が 8000Hz サンプリングであり、QMF 分析フィルタバンクは迂回される。
- (2) G729B\_BST モードは、復号は J T - G 7 2 9 付属資料 B のフレームを復号する。
- (3) LOW\_DELAY モードは、狭帯域のユースケースのために提供されている。ここでは、逆MDCTと重ね合わせ加算処理のスキップによる全体アルゴリズムの遅延削減のために、復号器のビットレートは 8-14 kbit/s に制限されている。

G729B\_BST モードと LOW\_DELAY モードは、復号器の出力はデフォルトでは 16000Hz サンプリングである。NB\_OUTPUT モードが同時にセットされている場合、出力は 8000Hz サンプリングになる。

Table C.3/JT-G729.1– Encoder/decoder modes  
(ITU-T G.729.1)

Mode	Encoder operation	Decoder operation
DEFAULT	16000 Hz input	16000 Hz output
NB_INPUT	8000 Hz input	N/A
DTX	Enable DTX/CNG scheme	N/A
G729B_BST	Bit rate limited to less than 8 kbit/s, output G.729B bitstream	N/A
NB_OUTPUT	N/A	8000 Hz output
G729B_BST	N/A	Read and decode G729B bitstream
LOW_DELAY	N/A	Bit rate limited to 8-14 kbit/s, low delay

#### C. 6. 4 ビットアロケーションフォーマット

符号器のビットアロケーションは Table C.4/JT-G729.1 に示されている。JT-G729.1 DTX/CNG は、3つの拡張レイヤを含むSIDスーパーフレームを生成する。

Table C.4/JT-G729.1 – SID bit allocation (per 20 ms superframe)  
(ITU-T G.729.1)

	Parameter description	Bits
LB core layer	Switched predictor index of LSF quantizer	1
	First stage vector of LSF quantizer	5
	Second stage vector of LSF quantizer	4
	Gain (Energy)	5
LB enhancement layer	The index of third stage vector of LSF quantizer	6
	Gain (Energy parameter)	3
HB layer	Time envelope mean	5
	Frequency envelope split VQ	14

JT-G729.1 DTX/CNGにおけるスーパーフレームSIDフォーマットは Figure C.3/JT-G729.1 に示される。

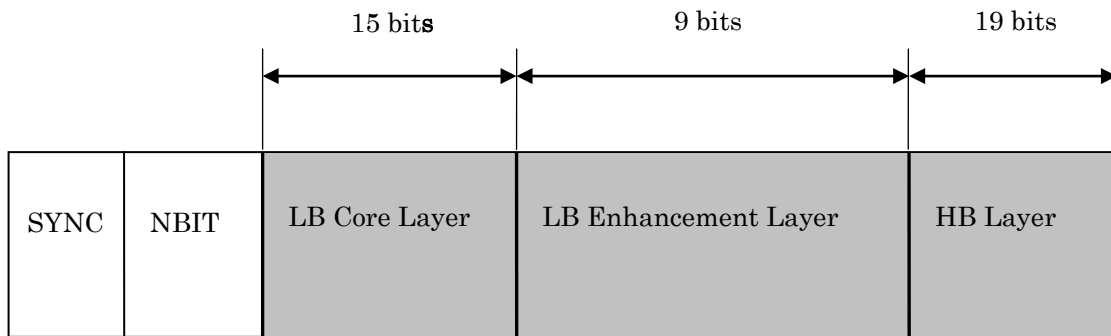


Figure C.3/JT-G729.1 – JT-G729.1 DTX/CNG SID bitstream format (compliant with G.192) (ITU-T G.729.1)

符号器側では、SIDの最大符号化レートは、Table C.5/JT-G729.1 に示されているように音声スーパーフレームの符号化レートにより決定される。

Table C.5/JT-G729.1 – Mapping between SID and speech superframe (ITU-T G.729.1)

SID structure	Speech superframe structure
Low band core layer	8 kbit/s
Low band enhancement layer	12 kbit/s
HB layer	≥ 14 kbit/s

#### C. 6. 5 アルゴリズム遅延

DTX/CNGアルゴリズムでは、JT-G729.1 に対して更なるアルゴリズム遅延の増加は無い。

#### C. 6. 6 計算量、要求される記憶容量、DTXの性能

JT-G729.1 DTX/CNG全体（符号器と復号器）の計算量は、セレクション/オプティマイゼーションフェーズで使用された中国語データベースに基づき、ITU-TソフトウェアツールライブラリSTL2005 v 2.1（ITU-T勧告G.191）の基本演算子を用いて見積もられた。有音音声での計算量の増加はピーク値で0.27WMOPS、無音送信での計算量の増加は最大で16.06WMOPSであった。Table C.6/JT-G729.1 は、JT-G729.1 DTX/CNGの計算量、16-bit kwords でのSRAM、DRAM、DROM容量、同様にSTLツールの基本演算子により見積もられたPROMの測定結果を示している。

Table C.6/JT-G729.1 – Complexity figures of the JT-G729.1 DTX/CNG (encoder/decoder) (ITU-T G.729.1)

Items	G.729.1	G.729.1 DTX/CNG
Computational complexity (WMOPS)	35.15	16.06 (inactive)
		36.06 (active)
Static RAM (kwords)	4.2	G.729.1+0.34
Dynamic (scratchpad) RAM (kwords)	4.6	G.729.1+0.03
Data ROM (kwords)	8.3	G.729.1+0.287
Program ROM (ops+function-call)	8325	9557

J T - G 7 2 9 . 1 D T X / C N G アルゴリズムは、スーパーフレーム数が 2 の最小 S I D 間隔と、ユーザの設定する最大 S I D 間隔（デフォルトのスーパーフレーム数は 25）を持つ。Table C.7/JT-G729.1 は中国語データベースでの D T X の性能の統計データを示している。

Table C.7/JT-G729.1 – DTX Efficiency for the Chinese Database (ITU-T G.729.1)

Conditions	Clean speech	Office noise	Babble noise (128)	Babble noise (40)
Number of superframes	12500	12500	12500	12500
Percentage of active superframes	56.18	65.06	56.20	56.80
Percentage of hangover frames	4.56	9.27	3.63	4.11
Percentage of inactive superframes	39.26	25.66	40.17	39.09
Percentage of SID frames	4.22	2.90	3.44	4.31
Percentage of NT superframes	35.04	22.76	36.72	34.78
Overall bitrate (kbit/s)	18.708	22.181	18.567	18.844
Average SID distance (superframe)	10.953	15.649	13.870	10.576

### C. 6. 7 符号器

本標準の符号化アルゴリズムは、ビットイグザクトな固定小数点算術演算で記述されている。C. 1 1 節で示される A N S I C コードは、本標準の必須部分を構成するものであり、このビットイグザクトな固定小数点での記述を反映している。符号器 (C. 7 節) および復号器 (C. 8 節) の算術的な記述は、他の方法でも実装し得るが、本標準に準拠しないコーデックを実装することになってしまう可能性がある。

したがって、不一致が生じた場合には、C. 7 節および C. 8 節の算術的な記述よりも、C. 1 1 節の A N S I C コードによるアルゴリズム記述の方が優先される。A N S I C コードと共に用いられるテスト信号の非網羅的なセットは、I T U - T の Web サイトから入手可能である。

### C. 7 無音圧縮符号器に関する機能記述

本節では、オプションである V A D、D T X、及び S I D フレームにおけるパラメータ算出や量子化を含む、無音圧縮符号器について記述する。C. 7. 1 節では、オプションである J T - G 7 2 2 . 2 の V A D の使用方法を、C. 7. 2 節では、ハングオーバー区間での使用について記述する。C. 7. 3 節では、低域側の D T X、C. 7. 4 節では高域側の D T X、C. 7. 5 節では結合された D T X 判定について言及する。C. 7. 6 節では、L B パラメータの推定、C. 7. 7 節では、H B パラメータの推定について説明する。S I D の L B パラメータ量子化については C. 7. 8 節で、H B パラメータ量子化については C. 7. 9 節で記述する。

### C. 7. 1 オプションのVAD

J T-G 7 2 2. 2 で記述されているVADは、J T-G 7 2 9. 1 の無音圧縮方式の開発と評価に利用された。無音圧縮方式のビットイグザクト検証用の評価ベクトルは、J T-G 7 2 2. 2 のVADを使って生成され、無音圧縮方式として報告されたDTXの性能は、そのVADを使って得られたものである。しかしながら、無音圧縮方式として報告された処理負荷とメモリ量は、VADアルゴリズムの実装に必要な処理負荷、またはメモリ量を含んでおらず、無音圧縮符号器の最大性能を発揮するためには、提供されたDTXやSIDアルゴリズムと共に、どんなVADが無音圧縮符号器で使われても良い。

### C. 7. 2 ハングオーバー

いったんVADが無音を検出すると、符号器はC. 7. 7 節で述べられるように、背景雑音の特性を学習するために6つの追加のスーパーフレームを利用する。これらの6つのハングオーバースーパーフレームの期間は、ビットレートが14 kbit/s より大きい場合には符号器は14 kbit/s のパケットを生成し、それ以外の場合は8 kbit/s のパケットを生成する。このハングオーバーは実装されたVADに関わらず使われる。

### C. 7. 3 低域側のDTX

無音区間の20msのスーパーフレームでは、LB DTXモジュールは、無音区間での低域側における聴感上の変化を測定することによって、無音信号パラメータ更新の送信が必要かを知らせる。無音区間の各10msのフレームでは、アルゴリズムはJ T-G 7 2 9 BのDTXと似た方法に基づく。

各20msのスーパーフレームの先頭10ms フレームでは、現在のLPCフィルタと前回のSID LPCフィルタが著しく異なる場合や、現在の励振エネルギーと前回のSIDエネルギーが著しく異なる場合には、最初のフレームでのLB DTXインジケータ $dtx_{1st}^f$ は1に設定され、それ以外は0に設定される。各20msスーパーフレームにおける2番目の10ms フレームにおいても、2番目のフレームのLB DTXインジケータ $dtx_{2nd}^f$ を決定するために同様の処理が行われる。各10ms フレームにおいては、DTXインジケータは、若干の違いはあるが、J T-G 7 2 9 BのB. 4. 1 節と同様の方法で計算される。1つ目の違いとしては、パラメータ抽出手順がJ T-G 7 2 9 Bとは若干異なる(C. 7. 6. 2 節参照)。2つ目の違いは、板倉距離の比較閾値 $thr1$ として、J T-G 7 2 9 Bの1.20226の代わりに、1.342676475 が設定されること、そして、エネルギー差分閾値 $thr2$ がJ T-G 7 2 9 Bの2 dBの代わりに3 dB が設定されることである。

現スーパーフレームの狭帯域DTXインジケータ $dtx_{ib}^s$ は次のように計算される。

$$dtx_{ib}^s = dtx_{1st}^f \parallel dtx_{2nd}^f \quad (C.1)$$

言い換えると、現スーパーフレームのLB DTXインジケータは、最初のフレームか2番目のフレームのDTXインジケータが設定された場合に設定される。20ms スーパーフレームのタイプは、J T-G 7 2 9 Bで使われている10ms フレームのタイプ $Ftyp_i$ と同様に、 $Ftyp_i^s$ と表示される(C. 7. 5 節参照)。入力信号が狭帯域である場合(C. 1 1. 1 節参照)、スーパーフレームタイプ $Ftyp_i^s$ はLB DTXインジケータのみに基づく。すなわち、J T-G 7 2 9 BのB. 4. 1. 2 節と同様、もし $dtx_{ib}^s$ が1なら $Ftyp_i^s = 2$ で、それ以外は $Ftyp_i^s = 0$ となる。もし入力信号が広帯域の場合、スーパーフレームタイプ $Ftyp_i^s$ は、C. 7. 5 節で述べられる結合されたDTX判定として決められる。

### C. 7. 4 高域側のDTX

HB DTXインジケータ $dtx_{hb}^s$ は、無音区間での高域側における聴感上の変化を測定することによって、無音パラメータ更新の送信が必要かを知らせる。HB DTXインジケータは、各スーパーフレーム $m$ における時間包絡パラメータ(6. 5. 1 節参照)を使って評価された、以下の条件を満たす場合に1に設定される。



$$|\tilde{T}^m - \tilde{T}^{m-1}| > \frac{1.4}{6.020599913} \quad (\text{C.2})$$

or

$$|\tilde{T}^m - T^{ltx}| > \frac{3.2}{6.020599913} \quad (\text{C.3})$$

ここで、 $T^{ltx}$  は最後に送信された時間包絡パラメータを表す。フィルタ処理された平均時間包絡パラメータ  $\tilde{T}^m$  はC. 7. 7節に従って計算される。

HB DTXインジケータの設定は、以下の条件が満たされれば、周波数包絡パラメータ(6. 5. 2節参照)の評価からも許される。

$$\frac{1}{12} \sum_{j=0}^{11} |\tilde{F}_{env}^m(j) - \tilde{F}_{env}^{m-1}(j)| > \frac{1.4}{6.020599913} \quad (\text{C.4})$$

or

$$\frac{1}{12} \sum_{j=0}^{11} |\tilde{F}_{env}^m(j) - F_{env}^{ltx}(j)| > \frac{3.2}{6.020599913} \quad (\text{C.5})$$

ここで、 $F_{env}^{ltx}(j)$  は最後に送信された周波数包絡パラメータを表す。フィルタ処理された周波数包絡パラメータ  $\tilde{F}_{env}^m(j)$ ,  $j=0 \dots 11$  はC. 7. 7節に従って得られる。

#### C. 7. 5 結合されたDTX

LB DTXインジケータとHB DTXインジケータは独立に計算されるため、全帯域で明らかな聴感上の違いが知覚されない場合でも、DTXインジケータが一つの帯域では設定され他方の帯域では設定されないということがあり得る。それゆえ、結合DTXモジュールにより、最終的なスーパーフレームDTX判定が行われる。スーパーフレームでの判定は、JT-G729BのB. 4. 1. 2節で使われる10msフレームタイプ  $Ftyp_t$  と同様、 $Ftyp_t^s$  で表されるスーパーフレームタイプによって示される。

もし現スーパーフレームがハングオーバー区間後の最初のスーパーフレームであった場合、スーパーフレームタイプはSIDフレーム  $Ftyp_t^s = 2$  として設定され、結合されたDTX判定が完了する。

一方、結合されたDTXフラグ  $flag_{cmb}$  は、Table C.7-bis/JT-G729.1 に従って、LB DTXインジケータ  $dtx_{lb}^s$  とHB DTXインジケータ  $dtx_{hb}^s$  によって決定される。

Table C.7-bis/JT-G729.1 – Combined DTX indicator  
(ITU-T G.729.1)

$dtx_{lb}^s$	$dtx_{hb}^s$	$flag_{cmb}$	$dtx^s$
0	0	0	0
1	0	1	–
0	1	1	–
1	1	0	1

LB DTXインジケータとHB DTXインジケータが同じ場合は、スーパーフレームDTX判定は  $dtx^s = dtx_{lb}^s (= dtx_{hb}^s)$  によって設定され、結合されたDTX判定が完了する。

一方、 $flag_{cmb} = 1$  の場合、結合されたDTX判定は、次のように、結合された距離尺度  $d$  を計算することによって生成される。

$$d = w_1 \cdot \left| \tilde{T}^{sid} - \tilde{T}^m \right| + w_2 \cdot \sum_{i=1}^{12} \left| \tilde{F}_{env}^{sid}(i) - \tilde{F}_{env}^m(i) \right| + w_3 \cdot C_f + w_4 \cdot C_g \quad (C.6)$$

ここで、 $w_1 = 0.07142857$ 、 $w_2 = 0.107142857$ 、 $w_3 = 0.45$ 、 $w_4 = 0.3$ である。また、 $\tilde{T}^{sid}$ は直近のSIDスーパーフレームの時間包絡、 $\tilde{T}^m$ は現スーパーフレームの時間包絡であり(C.7.7節中の式(C.18)参照)、 $\tilde{F}_{env}^{sid}(i)$ は最後のSIDスーパーフレームの $i$ 番目の周波数包絡、 $\tilde{F}_{env}^m(i)$ は現スーパーフレームの $i$ 番目の周波数包絡である(C.7.7節中の式(C.20)参照)。結合されたDTXがSIDを送出すべきであると示した時点で、 $\tilde{T}^{sid}$ は $\tilde{T}^m$ に、 $\tilde{F}_{env}^{sid}$ は $\tilde{F}_{env}^m$ に更新される。パラメータ $C_f$ は前回のSID LPCフィルタと現LPCフィルタの板倉距離の重み付けされた比率であり、次のように表される。

$$C_f = \frac{\sum_{j=0}^{11} R_a(j) \cdot r^f(j)}{E_t \cdot thr1} \quad (C.7)$$

ここで、重み $thr1 = 1.342676475$ である。パラメータ $r^f(j)$ はC.7.6.2.1節で定義され、 $R_a(j)$ はJT-G729Bの式(B.13)のように計算される。また、パラメータ $C_g$ は次式により計算される、 $E_t$ と $\hat{E}_{pre}^{sid}$ の重み付き距離である。

$$C_g = \frac{\left| E_t - \hat{E}_{pre}^{sid} \right|}{thr2} \quad (C.8)$$

ここで重み $thr2 = 3.0$ であり、 $\hat{E}_{pre}^{sid}$ は最後に送信されたSIDの量子化されたエネルギーパラメータ、 $E_t$ は現フレームの残差エネルギーである(C.7.6.2.1節参照)。もし、 $d < 1$ の場合、結合されたDTX判定 $dtx^s$ は、0に再設定され、それ以外の場合は1に設定される。

結合されたDTX判定 $dtx^s$ が決定される時点で、スーパーフレームタイプは次式により得られる。

$$\left. \begin{array}{l} count\_fr \geq N_{min} \\ dtx^s = 1 \end{array} \right\} \Rightarrow Ftyp_t^s = 2$$

$$count\_fr \geq N_{max} \Rightarrow Ftyp_t^s = 2 \quad (C.9)$$

$$\text{Otherwise} \Rightarrow Ftyp_t^s = 0$$

ここで、カウンタ $count\_fr$ は、最後のSIDスーパーフレームが送信されてからいくつのスーパーフレームが経過したかを表す。値 $N_{min} = 2$ は、各SIDスーパーフレーム間の最小間隔が2スーパーフレームであることを保証する。もし $count\_fr$ が $N_{max}$ より大きい場合、フレームタイプは強制的に2にされる。ここで $N_{max}$ は符号器のパラメータである。

$Ftyp_t^s = 2$ の場合、SIDはエネルギーを記述し、スペクトルパラメータが送られる。そのようなスーパーフレームはSID\_UPDATEスーパーフレームと呼ばれる。

$Ftyp_t^s = 0$ の場合、何も送信されず、これらのスーパーフレームは、NO\_DATAスーパーフレームを生成する非伝送(NT)スーパーフレームと呼ばれる。

### C.7.6 低域パラメータ推定

JT-G729Bと同様、LB SIDパラメータはエネルギーとスペクトルパラメータである。これらのパラメータは、JT-G729Bと同様に、10ms フレーム単位で抽出される。しかしながら、JT-G729.1のスーパーフレームは20msなので、JT-G729.1のSIDで使われるパラメータは、2つの

10ms フレームのパラメータの結合となる。結合の重みは、2つの 10ms フレームの安定性に依存する。

#### C. 7. 6. 1 最初のスーパーフレームSIDにおけるパラメータ推定

最初のSIDスーパーフレームはハングオーバー区間の後で発生する。平滑化された残差エネルギーとスペクトルパラメータを得るために、ハングオーバー区間の6スーパーフレームの間とハングオーバー後の最初の無音スーパーフレームにおいてそれらの推定が行われる。現SIDに先行する  $N_{sum}$  スーパーフレームから構築された過去のLPCフィルタ  $\bar{A}_p(z)$  は、以下の自己相関の総和を使って計算される。

$$\bar{R}_p(j) = \frac{1}{2 \cdot N_{sum}} \sum_{m=t-N_{sum}+1}^t \sum_{k=0}^1 r'_{m,k}(j), \quad j=0\dots10 \quad (C.10)$$

ここで  $r'_{m,k}(j)$  は、 $m$  番目のスーパーフレームの  $k$  番目のフレームの補正された自己相関関数であり(J T-G 7 2 9の式(7)参照)、 $N_{sum}=5$  である。つまり5スーパーフレームでの10個の自己相関関数が平均を求める際に使われることを意味する。平均自己相関関数  $\bar{R}_p(j)$  を入力としたレビンソン-ダービン手順は、残差エネルギー  $E^s$  とLPC係数を算出する。より安定した残差エネルギーパラメータを得るために、残差エネルギー  $E^s$  は次のように平滑化される。

$$E_{sid}^l = \alpha \cdot E_{sid}^l + (1-\alpha) \cdot E^s \quad (C.11)$$

長期平均残差エネルギー  $E_{sid}^l$  は、後に量子化されるSIDエネルギーパラメータとなる。長期平均残差エネルギーは0に初期化され、有音または無音のスーパーフレーム両方を含む各スーパーフレームにおいて更新される。 $\alpha$  という値は、スーパーフレームタイプに依存する。現スーパーフレームが有音か、または最初のハングオーバースーパーフレームである場合、 $\alpha$  は0.1である。現スーパーフレームがその後の5つのハングオーバースーパーフレームである場合、 $\alpha$  は0.5であり、その他の場合、 $\alpha$  は0.9である。これは、ノイズ性の高いスーパーフレームであるほど更新速度が遅く、ノイズ性が低いスーパーフレームであるほど更新速度が速いことを意味する。

#### C. 7. 6. 2 その他のスーパーフレームSIDにおけるパラメータ推定

本節では、無音区間の最初のスーパーフレームSID以外のSIDで使われるパラメータ推定について記述する。

##### C. 7. 6. 2. 1 フレームパラメータ推定

4つの連続フレームにおける自己相関関数  $r'_n(j)$ ,  $n=0,\dots,3$ ,  $j=0,\dots,10$  (J T-G 7 2 9の式(7)参照)、が計算で使われる。ノイズパラメータの平滑化を改善するため、これらの4つの自己相関関数の中から2つが選ばれる。最初に、これらの4つの自己相関関数の2-ノルムが計算される。

$$Norm_n = \sqrt{\sum_{j=0}^{10} (r'_n(j))^2} \quad (C.12)$$

それから、最大と最小の2-ノルム値となる自己相関関数が捨てられ、中間の2-ノルム値( $r_{mid1}(j)$ 、 $r_{mid2}(j)$ と表される)となる2つの自己相関関数を使って、現フレームの平均自己相関関数が次のように計算される。

$$r^f(j) = 0.5 \cdot r_{mid1}^f(j) + 0.5 \cdot r_{mid2}^f(j), \quad j = 0 \dots 10 \quad (C.13)$$

平均自己相関関数  $r^f(j)$  を入力とするレビンソン-ダービン手順を使って、スーパーフレーム残差エネルギーと L P C 係数が算出される。C. 7. 6. 1 節で記述されたのと同じ残差エネルギー平滑化手順が、そのフレームの残差エネルギーを平滑化するために使われる。計算された残差エネルギー  $E_t$  とスペクトルパラメータが、C. 7. 3 節と C. 7. 5 節でそれぞれ説明された L B D T X と結合された D T X 両方において使われる。

#### C. 7. 6. 2. 2 スーパーフレームパラメータ推定

スーパーフレームの残差エネルギーと自己相関関数は次のように計算される。

$$E^s = \phi \cdot E_{1st}^f + (1 - \phi) \cdot E_{2nd}^f \quad (C.14)$$

$$r^s(j) = \phi \cdot r_{1st}^f(j) + (1 - \phi) \cdot r_{2nd}^f(j), \quad j = 0, \dots, 10 \quad (C.15)$$

ここで  $E_{1st}^f$  は最初の 10ms フレームの残差エネルギー、 $E_{2nd}^f$  は 2 番目の 10ms フレームの残差エネルギー、 $r_{1st}^f(j)$  は最初の 10ms フレームの自己相関関数、 $r_{2nd}^f(j)$  は 2 番目の 10ms フレームの自己相関関数である。重みパラメータ  $\phi$  は次のように与えられる。

$$\begin{aligned} dtx_{1st}^f = 0 \text{ and } dtx_{2nd}^f = 1 &\Rightarrow \phi = 0.1 \\ \text{Otherwise} &\Rightarrow \phi = 0.5 \end{aligned} \quad (C.16)$$

現スーパーフレームの L P C フィルタ  $A_t(z)$  は、 $r^s(j)$  を入力としたレビンソン-ダービン手順によって算出される。残差エネルギー  $E^s$  はスーパーフレームエネルギーパラメータとして量子化される。S I D の中で送出されるフィルタパラメータは、現スーパーフレームパラメータと、式(C.10)の  $\bar{R}_p(j)$  から計算された過去の平均フィルタ  $\bar{A}_p(z)$  との比較によって選択される。それから S I D L P C フィルタが次のように得られる。

$$A_{sid}(z) = \begin{cases} A_t(z) & \text{if distance}(A_t(z), \bar{A}_p(z)) \geq thr3 \\ \bar{A}_p(z) & \text{otherwise} \end{cases} \quad (C.17)$$

閾値  $thr3$  は 1.0966466 に固定され、現 L P C フィルタと過去の平均フィルタの距離は、J T - G 7 2 9 B の B. 4. 1. 3 節と同じ方法で計算される。フィルタ  $A_{sid}(z)$  の L S F パラメータは C. 7. 8 で述べられる方法で量子化され、S I D フレームの中で送出される。

#### C. 7. 7 高域パラメータ推定

H B D T X パラメータは、フィルタ処理された平均時間包絡とフィルタ処理された周波数包絡パラメータから構成される。スーパーフレーム  $m$  のフィルタ処理された平均時間包絡  $\tilde{T}^m$  は次のように算出される。

$$\tilde{T}^m = \alpha_{tenv} \cdot \bar{T}^m + (1 - \alpha_{tenv}) \cdot \tilde{T}^{m-1}, \quad (C.18)$$

where

$$\bar{T}^m = \frac{1}{16} \sum_{i=0}^{15} T_{env}^m(i), \quad (C.19)$$

ここで  $\alpha_{tenv}$  は 0.25 に設定される。また、 $\tilde{T}^{m-1}$  はハングオーバー区間の 6 スーパーフレームに渡る平均  $T_{env}(i)$

(6. 5. 1 節の式(41)参照)によって初期化される。また、 $T_{env}^m(i)$ はスーパーフレーム  $m$  における時間包絡パラメータ  $T_{env}(i)$ を表す。

周波数包絡パラメータのフィルタ処理は次のように行われる。

$$\tilde{F}_{env}^m(j) = \alpha_{fenv} \cdot F_{env}^m(j) + (1 - \alpha_{fenv}) \cdot \tilde{F}_{env}^{m-1}(j), \quad (C.20)$$

ここで  $\alpha_{fenv}$  は 0.25 に設定され、 $\tilde{F}_{env}^{m-1}(j)$  はハングオーバー区間の 6 スーパーフレームに渡る平均  $F_{env}(j)$  (6. 5. 2 節の式(44))によって初期化される。また、 $F_{env}^m(j)$  はスーパーフレーム  $m$  における周波数包絡パラメータ  $F_{env}(j)$ を表す。

### C. 7. 8 低域パラメータ量子化

本節では、LB 残差エネルギーとスペクトルパラメータの量子化について記述する。

#### C. 7. 8. 1 低域エネルギー量子化

本節では、JT-G 7 2 9. 1 SID スーパーフレームのコアレイヤとエンハンスメントレイヤ両方における残差エネルギー量子化について記述する。

##### C. 7. 8. 1. 1 コアレイヤにおけるエネルギー量子化

コアレイヤでのエネルギー量子化は、JT-G 7 2 9 B の B. 4. 2. 1 節と同じ方法を使って、20ms スーパーフレーム毎に  $E_{sid}^l$  を量子化する。

##### C. 7. 8. 1. 2 エンハンスメントレイヤにおけるエネルギー量子化

LB エンハンスメントレイヤでは、より高精度な量子化のために 3 ビットが追加される。ここで、エンハンスメントレイヤでの各エネルギー量子化間隔のステップは、コアレイヤのエネルギー量子化間隔の 1/8 である。エンハンスメントレイヤにおけるエネルギー量子化の範囲は -16 dB から 67 dB であり、-12 dB から 66 dB の範囲を持つコアレイヤに対して、低い方で 4dB、高い方で 1dB 拡張されている。67 dB から 16dB の間では 0.25dB という均一の量子化ステップサイズが使われる。また、16dB から -4dB の間では 0.5dB の量子化ステップサイズが使われる。-4dB 未満の場合は、1つのステップサイズが使われて、最小量子化レベルは -16dB となる。量子化は直接行われ、量子化テーブルを必要としない。

#### C. 7. 8. 2 低域スペクトル量子化

本節では、JT-G 7 2 9. 1 の SID のコアレイヤとエンハンスメントレイヤ両方におけるスペクトルパラメータの量子化について記述する。 $r^f(j)$  (C. 7. 6. 2. 1 節参照) から計算される LSF パラメータは、20ms スーパーフレーム毎に量子化される。

##### C. 7. 8. 2. 1 コアレイヤにおける LSF 量子化

コアレイヤによる LSF ベクトルの量子化は、JT-G 7 2 9 B の B. 4. 2. 2 節に記載された LSF 量子化に従う。しかしながら、このコアレイヤ量子化は、JT-G 7 2 9 B での各 10ms フレームの代わりに、各 20ms スーパーフレームにおいて実行される。

##### C. 7. 8. 2. 2 エンハンスメントレイヤにおける LSF 量子化

LB エンハンスメントレイヤは、LSF 係数のより高精度な量子化のために 6 ビットを追加している。JT-G 7 2 9 B で使われる 2 番目の LSF コードブックのように、エンハンスメントレイヤの 3 段目のコー

ドブックは、J T-G 7 2 9 の 2 段目のコードブックからインデックスマッピングすることによって得られる。主な修正は、コードブックベクトルを調整するために適応的な倍率が使われる点である。さらに、エンハンスメントレイヤのインデックスマッピングテーブルは、J T-G 7 2 9 B では 16 ペアのエントリだったのに対して、64 ペアのエントリを持つ。これは、エンハンスメントレイヤの 3 段目のコードブックが 6 ビットのコードブックであることを意味する。エンハンスメントレイヤの 3 番目のコードブックのターゲットベクトルは  $\omega'_i$  ( $i=1,\dots,10$ ) であり、再構築された L S F ベクトル  $\hat{\omega}_i$  ( $i=1,\dots,10$ ) (コアレイヤの 2 段量子化によって量子化されたもの) と量子化されていない L S F ベクトル  $\omega_i$  ( $i=1,\dots,10$ ) の距離として次式により与えられる。

$$\omega'_i = \omega_i - \hat{\omega}_i, \quad i=1,\dots,10 \quad (\text{C.22})$$

3 段目のベクトル  $Y_3$  は、次のように計算される。

$$Y_3 = \gamma \cdot Y_2 \quad (\text{C.23})$$

ここで  $Y_2$  はエンハンスメントレイヤでのインデックスマッピングにより J T-G 7 2 9 の 2 番目のコードブックから選択されたベクトルであり、 $\gamma$  はエネルギー量子化インデックスの上位 3 ビット(次式参照)により選択された倍率である。

$$\gamma = H[E_{idx} \gg 2] \quad (\text{C.24})$$

ここで  $H$  は倍率として 8 つのエントリを持つスカラコードブック、 $E_{idx}$  はコアレイヤの量子化エネルギーインデックスである。

3 段目のコードブックは、次式で算出される  $m$  に渡る重み付き誤差  $E_{LSF3}(m)$  を最小化することによって探索される。

$$E_{LSF3}(m) = \sum_{i=1}^{10} W_i \left( \omega'_i - Y_3^{(m)} \right)_i^2 \quad (\text{C.25})$$

ここで  $W_i$  は重みベクトルである(J T-G 7 2 9 の 3. 2. 4 節中の式(22)参照)。

## C. 8 無音圧縮復号器の機能記述

本節では、パラメータ再構成、CNG アルゴリズム、フレーム消失補償、ビットレート切り替え処理を含む無音圧縮復号器について記述する。C. 8. 1 節では低域パラメータと低域 CNG の復号について、C. 8. 2 節では高域パラメータと高域 CNG の復号について記述する。最初の S I D 消失の処理は C. 8. 3 節に、ビットレート切り替え処理については C. 8. 4 節に記述される。

### C. 8. 1 低域復号と CNG

本節では、L B パラメータの復号及び擬似背景雑音の生成について記述する。

### C. 8. 1. 1 低域パラメータの再構成

#### C. 8. 1. 1. 1 最初のSIDスーパーフレームパラメータの再構成

ハングオーバー期間後の最初のスーパーフレームにおける最初の 10ms フレームに対し、ハングオーバー期間からのパラメータを用いてLBパラメータが計算される。

J T-G. 7 2 9 同様、復号器にて再構成される量子化されたLSF係数は $\hat{\omega}_i, i=0, \dots, 9$ で表記される。有音期間の間、スペクトラムパラメータの長期平均 $\hat{\omega}_i^l$ は、以下のようにフレーム毎に更新される。

$$\hat{\omega}_i^l = 0.5 \cdot \hat{\omega}_i^l + 0.5 \cdot \hat{\omega}_i \quad i=0, \dots, 9 \quad (\text{C.26})$$

ここで、 $\hat{\omega}_i^l$ は、平坦なスペクトラム特性を現すような値の組に初期化される。ベクトル $\hat{\omega}_i^l$ は、最初のSIDスーパーフレームの最初のフレームのLSPとして使われる。

同様に、有音期間の間、固定コードブック利得の長期平均 $Gf^l$ は、以下のようにフレーム毎に更新される。

$$Gf^l = 0.5 \cdot Gf^l + 0.5 \cdot \sqrt{\hat{g}_c^2 + \hat{g}_{enh}^2} \quad (\text{C.27})$$

ここで、 $\hat{g}_c$ は生成された8kbit/s固定コードブック利得(3. 8. 2節/J T-G. 7 2 9参照)、 $\hat{g}_{enh}$ は生成された12kbit/s固定コードブック利得(6. 3. 10節参照)であり、 $Gf^l$ はゼロに初期化される。

その後、 $Gf^l$ を用いて以下のように最初のフレームの残差エネルギーパラメータ $\hat{E}_{lsp}$ が得られる。

$$\hat{E}_{lsp} = 0.4 \cdot Gf^l \quad (\text{C.28})$$

平均スペクトラムパラメータ $\hat{\omega}_i^l$ は、 $\hat{\omega}_i(0)$ を初期化するために用いられ、再構成されたエネルギーパラメータ $\hat{E}_{lsp}$ は、 $\hat{E}(0)$ を初期化するために用いられ、これらはC. 8. 1. 1. 2節で説明される通り後続のフレームパラメータの再構成に用いられる。

最初のSIDスーパーフレームの二番目のフレームに対し、SIDパラメータが、 $\hat{\omega}_i^{sid}$ および $\hat{E}_{sid}$ で表記される量子化値から得られる。これらのパラメータは、現フレームからそれに続く $d_{sid}$ フレームを線形外挿する外挿アルゴリズムに用いられる。 $d_{sid}$ よりも前に新たなSIDスーパーフレームが受信された場合、このスーパーフレームまでの外挿された値が次の期間の外挿のための初期値として使われる。 $d_{sid}$ フレームの外挿が終了しその時までには新たなSIDスーパーフレームが受信されなかった場合、その値が最後の外挿値として保持される(これは最後に受信されたSID再構成値と同じ)。通常のSIDスーパーフレームに対しては、 $d_{sid}$ は前回のSIDフレームと現在のSIDフレームとの間の距離として設定される。このとき、 $d_{sid}$ はスーパーフレームではなくフレーム単位で距離数をカウントすることに注意が必要である。最初のSIDスーパーフレームの二番目のフレームに対しては、 $d_{sid}$ は適宜 $2 \cdot DTX\_HANG\_CONST - 1$ に初期化される。ここで、 $DTX\_HANG\_CONST = 6$ はハングオーバー長である。外挿手順は、後続のC. 8. 1. 1. 2節に詳細が記述されている。

#### C. 8. 1. 1. 2 NO\_DATAスーパーフレームパラメータの再構成

NO\_DATAスーパーフレームは復号器が更新パラメータを受信しなかったスーパーフレームである。狭帯域CNGパラメータは、最後のSID更新フレームからのフレームインデックスを表すインデックス $k$ を用いて線形外挿補間により再構成される。このとき、 $k$ はスーパーフレームではなくフレームのインデックスであることに注意が必要である。

$k$ 番目のフレームのフレームエネルギーパラメータ $\hat{E}(k)$ は、以下のように生成される。

$$\hat{E}(k) = \hat{E}(k-1) + \frac{\hat{E}_{sid} - \hat{E}(k-1)}{|k - d_{sid}| + 1} \quad (\text{C.29})$$

そして、 $k$ 番目のフレームのフレームスペクトラムパラメータ $\hat{\omega}_i(k)$ は、以下のように生成される。

$$D_{lsp}(i) = \frac{\hat{\omega}_i^{sid} - \hat{\omega}_i(k-1)}{|k - d_{sid}| + 1} \quad i=0, \dots, 9 \quad (\text{C.30})$$

$$\hat{\omega}_i(k) = \hat{\omega}_i(k-1) + D_{lsp}(i) + \text{rand}\left(-\frac{D_{lsp}(i)}{2}, \frac{D_{lsp}(i)}{2}\right) \quad i = 0, \dots, 9 \quad (\text{C.31})$$

ここで、 $\hat{\omega}_i^{sid}$  は、最新の S I D\_\_UPDATE スーパーフレームから得られた量子化 L S F ベクトルである。そして、関数  $\text{rand}(a,b)$  は、 $a$  と  $b$  の間の乱数を生成する。

### C. 8. 1. 1. 3 S I D\_\_UPDATE スーパーフレームパラメータの再構成

ハングオーバー後の最初の S I D 以外の S I D\_\_UPDATE スーパーフレームに対し、最初のフレームは C. 8. 1. 1. 2 節に記載されるような N O\_\_DATA フレームとして扱われる。二番目のフレームの間、 $\hat{\omega}_i^{sid}$  と  $E_{sid}$  は導出された L S P パラメータとエネルギーパラメータにより更新される。その後、C. 8. 1. 1. 2 節に記載されるような、N O\_\_DATA フレームに対するものと同じ外挿補間法が使われる。

S I D スーパーフレームが受信されたとき、次節 C. 8. 1. 2 のエネルギー減衰に使われる長期 S I D 区間  $d_{sid}^l$  が下式により更新される。

$$d_{sid}^l = 0.95 \cdot d_{sid}^l + 0.05 \cdot d_{sid} \quad (\text{C.32})$$

ここで、 $d_{sid}^l$  は  $2 \cdot DTX\_HANG\_CONST - 1$  に初期化される。

### C. 8. 1. 2 エネルギー減衰

適応エネルギー減衰はターゲット励振利得に対して以下のように適用される。S I D フレーム間の区間が 400ms 以上の場合、背景雑音が定常的であることを示し、以下のように C N G の聴覚品質を向上するためにエネルギー減衰が再構成されたエネルギーパラメータ  $\hat{E}(k)$  に対して適用される。

$$\hat{E}(k) = \hat{E}(k) \cdot \left( 0.9 + 0.1 \cdot \frac{30 \cdot ACT_{sw}^l + Hg_E}{30 \cdot ACT_{sw}^l + MAX\_ATT\_HO + d_{sid}} \right) \quad (\text{C.33})$$

平均有音性変化率  $ACT_{sw}^l$  は、1 に初期化され以下のように更新される。

最初に、有音性変化数  $ACT_{sw}$  を計測するために観測窓が使われる。観測窓長は 2 秒である。観測窓の最初で  $ACT_{sw}$  はゼロに設定される。観測窓の間、有音性変化が生じた場合は  $ACT_{sw}$  は 1 だけインクリメントされる。観測窓の最後で、 $ACT_{sw}^l$  は以下のように  $ACT_{sw}$  で更新される。

$$ACT_{sw}^l = 0.5 \cdot ACT_{sw}^l + 0.5 \cdot ACT_{sw} \quad (\text{C.34})$$

パラメータ  $Hg_E$  はデクリメントカウンタで、最初の S I D フレームで  $MAX\_ATT\_HO = 150$  に初期化され、その後無音フレームごとに 1 だけデクリメントされる。

### C. 8. 1. 3 励振生成

この節は、L B C N G に対する励振の生成について記述する。

#### C. 8. 1. 3. 1 新しい励振の計算

励振の生成は、J T - G 7 2 9 B の B. 4. 4 節とほぼ同じである。違いは、C E L P パラメータが J T - G 7 2 9 B のように完全には乱数的に生成されないことである。ピッチラグ  $pitch_{rnd}$  は区間 20 から 143 の間で乱数的に選定され、最後のフレームのピッチラグ  $pitch_{old}$  とで以下のように平滑化される。

$$pitch_{new} = pitch_{rnd} \& 0x000F + pitch_{old} \& 0x0030 \quad 20 \leq pitch_{new} \leq 143 \quad (\text{C.35})$$

最後の適応コードブック利得  $gp_{old}$  は、ガウス乱数適応コードブック利得  $gp_{rnd}$  ( J T - G 7 2 9 B の式(B.22) 参照) を平滑化するために以下のように使われる。

$$gp_{new} = 0.5 \cdot gp_{rnd} + 0.5 \cdot gp_{old} \quad (\text{C.36})$$

生成された新しい励振  $exc_{new}$  は、ハングオーバー期間後の最初の S I D を除き現フレームの励振として使われ



る。

### C. 8. 1. 3. 2 励振の遷移

ハングオーバー期間の最後で、有音／無音の切り替えにより生成される不快な事象を避けるため、有音スーパーフレームから無音スーパーフレームへの切り替えのときに励振遷移が使われる。最初の無音スーパーフレームに対する励振遷移は以下のように計算される。

$$exc_{cur}(n) = \frac{n}{160} \cdot exc_{new}(n) + \frac{160-n}{160} \cdot exc_{pre}(n) \quad n=1, \dots, 160 \quad (C.37)$$

ここで、 $exc_{new}$  は J T - G 7 2 9 の B. 4. 4 節で計算される現スーパーフレームの励振であり、 $exc_{pre}$  は、最初の無音スーパーフレームにおける前回の有音スーパーフレームの励振の外挿補間である。励振遷移は、有音区間後の最初の無音スーパーフレームの両 10ms フレームに適用される。

その後、有音期間後の最初のこの無音スーパーフレームにおける各フレームで、利得制御アルゴリズムが使われる。 $exc_{cur}(n)$  に対する利得制御は、以下のように表記される。

$$exc_{cur}(n) = exc_{new}(n) \cdot \frac{\sqrt{\frac{1}{80} \sum_{i=0}^{79} (exc_{cur}(i))^2}}{\sqrt{\frac{1}{80} \sum_{i=0}^{79} (exc_{new}(i))^2}} \quad (C.38)$$

### C. 8. 1. 3. 3 励振のシェーピング

励振  $exc_{cur}$  は、合成信号の聴覚品質を向上するためにシェーピングフィルタ  $H_{sh}(z)$  によりフレーム毎にシェーピングされ、以下のように表記される。

$$H_{sh}(z) = \frac{1}{\sum_{k=0}^6 |h_k|} \cdot \left( h_0 + \sum_{k=1}^6 h_k z^{-k} \right) \quad (C.39)$$

励振シェーピングフィルタの係数を、Table C.8/JT-G729.1 に示す。

Table C.8/JT-G729.1 – Excitation reshaping filter coefficients  
(ITU-T G.729.1)

Name	Constant
$h_0$	0.002144660940673
$h_1$	0.027883661103
$h_2$	0.07285533905933
$h_3$	0.797116338897
$h_4$	0.07285533905933
$h_5$	0.027883661103
$h_6$	0.002144660940673

### C. 8. 1. 3. 4 低域擬似背景雑音の生成

LPCパラメータと励振信号が完成するとすぐ、擬似背景雑音を生成するために合成フィルタリングが使用される (J T - G 7 2 9 の式(77)参照)。

### C. 8. 2 高域の復号とCNG

HB DTXの復号とCNGは7. 2節に従う。DTX工程の間、励振信号は7. 2. 2節のごとく生成され、無声部寄与分の生成とローパスフィルタリングのみが、実行される処理ステップである。

滑らかなHB成分の信号を得るため、平滑化手法がSIDから導出された周波数包絡 $\hat{F}_{env}^{sid}$ および時間包絡 $\hat{T}_{env}^{sid}$ に適用される。最初に、バイアス余弦窓が周波数包絡を形成するために以下のように使われる。

$$\hat{F}_{env}^w(j) = \hat{F}_{env}^{sid}(j) \cdot W_{\cos}(j) \quad j=0, \dots, 11 \quad (C.40)$$

ここで、 $W_{\cos}(j) = 0.8 + 0.2 \cdot \cos(j\pi/12)$   $j=0, \dots, 11$ は長さ12の余弦窓であり、 $\hat{F}_{env}^w(j), j=0, \dots, 11$ は窓掛け後の周波数包絡パラメータである。

その後、SID受信時に、一次ARフィルタが時間包絡と周波数包絡双方に適用される。

$$\begin{aligned} \hat{T}_{env}^{lt} &= 0.75 \cdot \hat{T}_{env}^{lt} + 0.25 \cdot \hat{T}_{env}^{sid} \\ \hat{F}_{env}^{lt}(j) &= 0.75 \cdot \hat{F}_{env}^{lt}(j) + 0.25 \cdot \hat{F}_{env}^w(j) \quad j=0, \dots, 11 \end{aligned} \quad (C.41)$$

ここで、 $\hat{T}_{env}^{lt}$ と $\hat{F}_{env}^{lt}$ は、高域成分を生成するために再構成パラメータとして使われる(7. 2節参照)。有音期間の間、同様な一次ARフィルタが長期時間包絡および長期周波数包絡を更新するために以下のように使われる。

$$\begin{aligned} \hat{T}_{env}^{lt} &= 0.5 \cdot \hat{T}_{env}^{lt} + 0.5 \cdot \hat{T}_{env}^{sp} \\ \hat{F}_{env}^{lt}(j) &= 0.5 \cdot \hat{F}_{env}^{lt}(j) + 0.5 \cdot \hat{F}_{env}^{sp}(j) \quad j=0, \dots, 11 \end{aligned} \quad (C.42)$$

ここで、 $\hat{T}_{env}^{sp}$ は有音スーパーフレームの時間包絡であり、 $\hat{F}_{env}^{sp}(j), j=0, \dots, 11$ は有音スーパーフレームの周波数包絡である。

### C. 8. 3 SIDスーパーフレーム消失補償

SIDスーパーフレーム消失は、無音信号期間の最初のSIDに対してのみ対処されることを必要とする。なぜならば、無音期間の中間のSIDの消失は、復号器において単に非伝送スーパーフレームと認識されるからである。JT-G729Bは、G. 192に規定されるビット列構造を用いて、失われたSIDフレームに続くNTフレームの検出により消失した最初のSIDの検出を行う。しかし、JT-G729. 1のための実際の通信アプリケーションにおいては、もし最初のSIDが失われると、復号器は失われたスーパーフレームがSIDスーパーフレームか音声スーパーフレームかを見分けることができず、SID更新スーパーフレームが受信されるまで音声スーパーフレーム消失のためのスーパーフレーム消失補償アルゴリズムを用いる(7. 6節参照)。SID更新スーパーフレームが到着した時点で、再構成された信号は無音に等しい極低レベルに減衰されている。なぜならば、JT-G729. 1におけるスーパーフレーム消失補償アルゴリズムは長期間のスーパーフレーム消失に対する減衰を用いるからである。このような場合の特別な取り扱いを除き、新しいSIDスーパーフレームが受信されると、背景雑音レベルは直ちに通常レベルに増加され、知覚される擬似背景雑音における異音感を生成する。このため、特別なエネルギー遷移手順が用いられる。

最初に、消失補償アルゴリズムの期間中、消失スーパーフレームに対する減衰係数 $\alpha$ が保持される。減衰係数の計算は、7. 6. 1節に記述される。SIDスーパーフレームが受信されるとすぐに、 $\alpha$ はフレーム毎に0.00390625のステップで最大値1.0まで増加される。CNG手順の後、再構成されたLB成分 $\hat{s}_l(n)$ は $\alpha$ が乗じられる。

$$\hat{s}_l(n) = \alpha \cdot \hat{s}_l(n) \quad (C.43)$$

HB成分に対する計算上の効果的手順のために、減衰係数は再構成された時間包絡 $\hat{T}_{env}^{lt}$ および周波数包絡 $\hat{F}_{env}^{sp}(j), j=0, \dots, 11$ に適用される。

$$\hat{T}_{env}^{att} = \alpha \cdot \hat{T}_{env}^{lt} \quad (C.44)$$

$$\hat{F}_{env}^{att}(j) = \alpha \cdot \hat{F}_{env}^{lt}(j) \quad j=0, \dots, 11 \quad (C.45)$$

$\alpha$  が 1.0 に達するとき、それは遷移期間の終了を意味し、その後の減衰処理は不要である。減衰された時間包絡  $\hat{T}_{env}^{att}$  および周波数包絡  $\hat{F}_{env}^{att}(j), j=0, \dots, 11$  は、7.2 節に記述されるように HB 成分の生成に用いられる。

#### C. 8. 4 ビットレート切り替え

JT-G 729.1 の音声および SID スーパーフレームの階層的構造のおかげで、実際の通信システムのアプリケーションにおいては、無音期間においても、帯域幅は広帯域から狭帯域もしくは狭帯域から広帯域に切り替えることができる。帯域幅切り替えによって生じるであろう聴覚劣化を避けるため、特別な処理が必要である。

DTX/CNG モード中のビットレート切り替えメカニズムは下式を用いて有音期間中に計算されるビットレート切り替え指標  $wbStat$  により制御される。

$$wbStat(m) = \begin{cases} wbStat(m-1) \cdot \frac{m-1}{m} & \text{if } m^{th} \text{ superframe is narrowband} \\ wbStat(m-1) \cdot \frac{m-1}{m} + \frac{1}{m} & \text{if } m^{th} \text{ superframe is wideband} \end{cases} \quad (C.46)$$

ここで、 $m$  は有音スーパーフレームのインデックスを表す。無音期間中に用いられる  $wbStat$  の値は、最後に計算された  $wbStat(m)$  の値である。

CNG 中の狭帯域から広帯域復号への切り替えの処理は、前回のスーパーフレームが狭帯域で生成された CNG スーパーフレームもしくは狭帯域ハングオーバースーパーフレームのときに実行され、現在のスーパーフレームは広帯域 SID スーパーフレームでありビットレート切り替え指標は  $wbStat > 0.5$  を満たす。

同様に、広帯域から狭帯域復号および CNG への切り替え処理は、前回のスーパーフレームが広帯域で生成された CNG スーパーフレームもしくは広帯域ハングオーバースーパーフレームのときに実行され、現在のスーパーフレームは狭帯域 SID スーパーフレームでありビットレート切り替え指標  $wbStat$  は 0.5 以下である。

##### C. 8. 4. 1 狭帯域から広帯域への切り替えに対する帯域幅フェードイン

狭帯域から広帯域への切り替え処理は、7.7.2 節のごとく、Table C.9/JT-G729.1 の減衰係数を信号  $\hat{s}_{HB}^{qmf}(m)$  に適用することにより実行される。

Table C.9/JT-G729.1 – DTX fade-in attenuation coefficients  
(ITU-T G.729.1)

Superframe counter	Attenuation Coefficient
1	0.0
2	0.09525986892242
3	0.19753086419753
4	0.36595031245237
5	0.62429507696997
6	1.0

##### C. 8. 4. 2 広帯域から狭帯域への切り替えに対する帯域幅フェードアウト

広帯域から狭帯域への切り替えの処理は以下のように実行される。最初に、下式により HB パラメータが推定される。

$$Par^k = \left(1 - \frac{n}{N}\right) \cdot Par_{sid} + \frac{n}{N} \cdot Par_{pre\_sid} \quad (C.47)$$

ここで、 $Par^k$  は帯域幅切り替えが生じた後の  $k$  番目のスーパーフレームの推定 HB パラメータであり、 $Par_{sid}$  は、直近に受信された SID の HB パラメータであり、 $Par_{pre}$  は以前に受信された SID の HB パラ

メータである。HBパラメータは、TDBWE時間包絡もしくはTDBWE周波数包絡のどちらでもよい。その後、帯域幅フェードアウトを実行するために時変矩形窓がTDBWE周波数包絡に適用される。その結果、帯域幅は以下のように滑らかに広帯域から狭帯域に切り替えられる。

$$\hat{F}_{env}^k(j) = \begin{cases} \hat{F}_{env}^l(j) & j \leq \left\lfloor \frac{(L_{fad} - k) \cdot J}{L_{fad}} \right\rfloor \\ LOW\_LEV & otherwise \end{cases} \quad (C.48)$$

ここで、 $LOW\_LEV = -15.95541$ であり、周波数包絡の最下値である。 $\hat{F}_{env}^k(j), j=0, \dots, 11$ は帯域幅切り替えが生じてからのk番目のスーパーフレームのフェードアウト周波数包絡であり、 $\hat{F}_{env}^l(j), j=0, \dots, 11$ は、C.8.2節に記載されている無音スーパーフレームのための再構成された周波数包絡である。 $J=12$ は周波数包絡ベクトルの大きさであり、フェードアウト処理は $L_{fad} = 50$ スーパーフレームに対して実行される。

### C.9 メモリの更新

標準JT-G729.1の符号器において、有音スーパーフレーム処理に対して、2つの合成フィルタが使用される。一つは、8kbit/sであり、もう一つは12kbit/sである。無音スーパーフレーム処理に対しては、8kbit/s合成フィルタのみ適用される。符号器と復号器のCELPの同期を保証するため、ハングオーバー期間中に、符号化レートが8kbit/s以上の場合は、8kbit/sに対する合成フィルタの状態は、12kbit/sの合成フィルタの状態を更新される。そして、符号化レートが8kbit/sの場合は、有音スーパーフレームに対して更新されると同様に無音区間に対しても、12kbit/sに対する合成フィルタ状態は、8kbit/sの合成フィルタの状態を更新される。

### C.10 伝送パラメータインデックスの詳細

ビット列の順序は、Table C.10/JT-G729.1に示される順番である。それぞれのパラメータは、MSBから伝送される。

Table C.10/JT-G729.1 Description of transmitted parameter indices  
(ITU-T G.729.1)

Symbol	Description	Bits
<i>L0</i>	LB CORE LAYER – Switched MA predictor of LSP quantizer	1
<i>L1</i>	LB CORE LAYER – First stage vector of LSP quantizer	5
<i>L2</i>	LB CORE LAYER – Second stage vector of LSP quantizer	4
<i>G1</i>	LB CORE LAYER – Gain(energy)	5
<i>L3</i>	LB ENHANCEMENT LAYER – Third stage vector of LSP quantizer	8
<i>G2</i>	LB ENHANCEMENT LAYER – Second stage Gain	1
<i>MU</i>	TDBWE – Time envelope	5
<i>F1</i>	TDBWE – Frequency envelope 1 <sup>st</sup> stage vector	5
<i>F2</i>	TDBWE – Frequency envelope 2 <sup>nd</sup> stage vector	5
<i>F3</i>	TDBWE – Frequency envelope 3 <sup>rd</sup> stage vector	4

### C.11 JT-G729.1無音圧縮のビットイグザクト詳細

16ビット固定小数点のJT-G729.1DTX/CNGをシミュレートするANSI-Cコードは、ITU-TのWebサイトから入手可能である。以下の節では、そのシミュレーションコードの使用方法和、そのソフトウェアの構成を概説する。

### C. 11. 1 シミュレーションソフトウェアの使用方法

Cコードは、符号器と復号器をシミュレートする2つのメインプログラム G729EV\_MAIN\_Encoder.c と G729EV\_MAIN\_Decoder.c から構成されている。

符号器に対するコマンドラインは、以下の通りである。

```
encoder [-options] [<vadfile>] inputfile bitstreamfile
```

符号器に対して利用可能なオプションは、以下の通りである。

dtx: run encoder with DTX/CNG enabled; the [<vadfile>] has to be specified;

rXXXXX: run encoder at XXXXX bit/s (by default: 32000);

mXX: the maximum SID interval (by default: 25);

f8: 8000 Hz sampled input;

g729b\_bst: run encoder at 8 kbit/s and generate bitstream with G.729B format

デフォルトでは、入力サンプリングレートは、16000Hzで、そのビット列は、20ms フレーム（あるいはスーパーフレーム）に分割される。

復号器に対するコマンドラインは、以下の通りである。

```
decoder [-options] bitstreamfile outputfile
```

復号器に対して利用可能なオプションは、以下の通りである。

– rXXXXX: run decoder with maximal bit rate XXXXXX bit/s (by default: 32000);

– f8: 8000 Hz sampled output;

– ld: low-delay mode (the decoder bit rate must be limited to 8 or 12 kbit/s);

– g729b\_bst: read and decode [G.729B] bitstream

デフォルトでは、出力サンプリングレートは、16000Hzで、そのビット列は、20ms フレーム（あるいはスーパーフレーム）に分割される。

入力ファイルと出力ファイルは 16-bit PCM 信号を含むサンプルデータファイルである。

符号化ビット列のマッピングテーブルは、シミュレーションソフトウェアに含まれている。

### C. 11. 2 シミュレーションソフトウェアの構成

ソースコードは、"src" ディレクトリに格納されている。Microsoft Visual C6.0 ワークスペースファイルは、"workspace/VC6.0"に格納されている。g729ev.dsw を開くと、JT-G729.1 DTX/CNGのCソースコードとプロジェクトが開く。

Table C.11/JT-G729.1 JT-G729.1 DTX/CNG encoder state memory (structure CODSTATMAIN)  
(ITU-T G.729.1)

Member	Words (16-bit)	Description
g729ev_dtx	1	Dtx flag
tdbwe_parameters	13	HB parameter

Table C.12/JT-G729.1 JT-G729.1 DTX/CNG encoder state memory (G729EV\_G729\_CODSTAT)  
(ITU-T G.729.1)

Member	Words (16-bit)	Description
Vad	1	VAD flag
pastVad	1	VAD flag of previous frame
ppastVad	1	VAD flag of previous of past frame
lsp_mean	10	Spectrum parameters
Ga_gain_mean	1	Adaptive codebook gain
Gf_gain_mean	1	Fixed codebook gain
seed	1	Random generator seed
hangover_count	1	Hangover counter
cng_nb	1	LB CNG flag
dtx_max_fr	1	Maximum distance between SID
flag_chang_wb	1	DTX indicator of HB
counter_k	1	Frame counter since SID
counter_n	1	Interval between last two SIDs
seed2	1	Random generator seed
RCoeff	11	Past autocorrelation functions
pastCoeff	11	Past filter coefficients
sh_RCoeff	1	Scalar
Acf	44	Buffered ACF
sh_Acf	4	Scalar
sumAcf	55	Summed ACF
sh_sumAcf	5	Scalar
ener	4	Buffered energy
sh_ener	4	Scalar
fr_cur	1	Frame counter
pre_gain	1	Energy
nb_ener	1	Energy
sid_gain	1	SID energy
prev_energy	1	Energy
count_fr0	1	Frame counter
pre_lsp	10	Spectrum parameters
sid_lsp	10	Spectrum parameters
old_pitch	1	Pitch information
old_gp	1	Pitch gain
Nb_First_change_flag	1	LB DTX indicator
Nb_Second_change_flag	1	LB DTX indicator
SuperFrame_CurAcf	22	Buffered ACF
sh_SuperFrame_CurAcf	2	Scalar
SuperFrame_Ener	2	Buffered energy
sh_SuperFrame_Ener	2	Scalar
energy_lt	1	Energy
sh_energy_lt	1	Scalar
step_counter	1	Counter
VADf_lag_change_counter	1	VAD changing counting
VAD_flag_chang_ecounter_move	1	Long term VAD changing counting

gain_change_hangover	1	Counter
transition_flag	1	Excitation transition flag
subfr_num	1	Frame number in one superframe
g729_bst	1	Flag
filt_temp	6	Reshaping filter stat
counter_n_lt	1	Long term SID interval

Table C.13/JT-G729.1 JT- G729.1 DTX/CNG decoder state memory (DECSTATMAIN)  
(ITU-T G.729.1)

Member	Words (16-bit)	Description
wbStat	1	HB counter
count_rv_dtx	1	Fade-in counter
SpToSil	1	Speech to noise flag
first_sp_seg	1	First speech flag
old_SID_rate	1	Previous SID rate
prev_ftyp_null	1	Previous frame type
SID_parameters_tdbwe	28	High band parameters
layer_pre	1	Previous superframe layer number
layer_cur	1	Current superframe layer number
SIDpre_parameters_tdbwe	1	Previous SID HB parameters
fade_out_count	1	Fade-out counter
fade_out_flag	1	Fade-out flag
parameters_tdbwe_longmove	28	Long term HB parameters

Table C.14/JT-G729.1 JT-G729.1 DTX/CNG decoder state memory (G729EV\_G729\_DECSTAT)  
(ITU-T G.729.1)

Member	Words (16-bit)	Description
sid_lsp	10	Spectrum parameters
old_pitch	1	Pitch information
old_gp	1	Pitch gain
Nb_First_change_flag	1	LB DTX indicator
Nb_Second_change_flag	1	LB DTX indicator
SuperFrame_CurAcf	22	Buffered ACF
sh_SuperFrame_CurAcf	2	Scalar
SuperFrame_Ener	2	Buffered energy
sh_SuperFrame_Ener	2	Scalar
energy_lt	1	Energy
sh_energy_lt	1	Scalar
step_counter	1	Counter
VADf_lag_change_counter	1	VAD changing counting
VAD_flag_chang_eounter_move	1	Long term VAD changing counting
gain_change_hangover	1	Counter
transition_flag	1	Excitation transition flag
subfr_num	1	Frame number in one superframe
g729_bst	1	Flag
filt_temp	6	Reshaping filter stat
counter_n_lt	1	Long term SID interval

Table C.15/JT-G729.1 JT-G729.1 DTX/CNG table ROM  
(ITU-T G.729.1)

Member	Words (16-bit)	Description
Tabinv	65	Used for division
PtrTab_3	128	Third stage codebook mapping index
ratio_cb2	8	Scalar codebook
noise_fg_sum_inv	20	The inverse of noise_fg_sum
Adjust_fact	4	For energy quantization
tab_Sidgain2	8	For 2nd stage energy quantization
tab_Sidgain3	8	For 3rd stage energy quantization
Mp	2	For spectrum parameter quantization
G729EV_DTX_switching_gain	6	For bandwidth switching
Shaping_filter	7	Reshaping filter
SmoothWindow	12	HB smoothing window
G729EV_DTX_LSP_INIT	10	Initialized LSP

Table C.16/JT-G729.1 は、JT-G729.1 から修正した JT-G729.1 DTX/CNG における C コードのソースファイルのリストである。

Table C.17/JT-G729.1 は、JT-G729.1 DTX/CNG で追加された新ファイルのリストである。



Table C.16/JT-G729.1 JT-G729.1 DTX/CNG modified source files  
(ITU-T G.729.1)

File name	Description
G729EV_CELP2S_decod.c/h	G.729.1 CELP decode functions
G729EV_CELP2S_encod.c/h	G.729.1 CELP encode functions
G729EV_CELP2S_post.c	CELP post functions
G729EV_FEC_decbfi.c	Bfi decoding functions
G729EV_G729_codstat.h	Structure for G.729 encoder declare
G729EV_G729_decstat.h	Structure for G.729 decoder declare
G729EV_G729_defines.h	Macro defines for G.729
G729EV_G729_ld8k.h	G.729 functions declare
G729EV_G729_lpc.c	LPC functions
G729EV_G729_lpcfunc.c	LPC functions
G729EV_G729_pst.c	G.729 post processing functions
G729EV_G729_tab_ld8k.c/h	Table for G.729
G729EV_G729_util.c	Tool functions
G729EV_G729B_calcexc.c	Excitation computing functions
G729EV_G729B_dec_sid.c	SID decoding functions
G729EV_G729B_defines.h	Macro defines for G.729B
G729EV_G729B_dtx.h	Head file for DTX
G729EV_G729B_qlsidgain.c	Noise energy quantization
G729EV_G729B_sid.h	Head for SID
G729EV_G729B_tab_dtx.c/h	Table for DTX
G729EV_G729B_util.c	Tool functions
G729EV_MAIN_decod.c/h	Root routine of decoding
G729EV_MAIN_encod.c/h	Root routine of encoding
G729EV_MAIN_decoder.c	Decoder main function
G729EV_MAIN_encoder.c	Encoder main function
G729EV_MAIN_defines.h	Macro defines for main functins
G729EV_MAIN_dspfunc.c/h	Integrated DSP functions
G729EV_MAIN_prm.c/h	Bitstream read/write functions
G729EV_MAIN_table.c	Table for main function

Table C.17/JT-G729.1 JT-G729.1 DTX/CNG added source files  
(ITU-T G.729.1)

File name	Description
G729EV_G729EVDTX_sp2no_trans.c/h	Excitation transition functions
G729EV_G729B_dtx.c	DTX functions
G729EV_G729B_enc_sid.h	SID encoding head file
G729EV_G729B_qlsidlsf.c	Spectrum parameters quantization functions

## 付属資料D JT-G729. 1 付属資料C (DTX/CNG) に対する浮動小数点演算での実装

(標準 JT-G729. 1 に対する)

### D. 1 適用範囲

本付属資料は、浮動小数点演算に基づく、JT-G729. 1 付属資料Cの別の実現方法を記述している。本付属資料は、JT-G729. 1 付属資料Cの固定小数点版と完全に相互接続性のあるものである。本付属資料における浮動小数点演算に基づく参照Cコードは、ITU-Tの Web サイトから入手可能である。テストベクトルセットの設計は、今後の課題である。

### D. 2 参考文献

JT-G729. 1 本体の2章を参照のこと。

### D. 3 概要

JT-G729. 1 付属資料Cは、JT-G729とビット列互換な 8-32kbit/s スケラブル広帯域符号器の、不連続伝送 (DTX) と擬似背景雑音発生器 (CNG) アルゴリズムをビットイグザクトな固定小数点で規定している。これらの仕様の正確な記述は、JT-G729. 1の一部として統合され、利用可能なビットイグザクトかつ固定小数点Cコードが提供される。本付属資料は、浮動小数点演算によるJT-G729. 1 付属資料Cの別の実現方法を記述し、規定するものである。

### D. 4 アルゴリズム記述

JT-G729. 1 付属資料Cの浮動小数点版は、固定小数点版と同じアルゴリズムステップを踏んでいる。同様に、浮動小数点版のビット列は、JT-G729. 1 付属資料Cの固定小数点版と同一である。アルゴリズムの詳細は、JT-G729. 1 付属資料Cの記述を参照のこと。

### D. 5 ANSI-Cコード

ANSI-Cコードが、本付属資料の標準的な記述として付属資料Dの電子的付録として入手可能である。この浮動小数点による実現方法は、JT-G729. 1 付属資料Cの一部として承認された固定小数点による実現方法に基づいている。Cコードによるアルゴリズム記述は、JT-G729. 1 付属資料Dに含まれるテキスト記述より優先される。最新版は、ITU-TG. 729に対する *Corrigenda* ないしは *Amendment* として入手可能となる可能性があるので、ITU-Tの Web サイトから入手可能な最新のバージョンを使用すること。

浮動小数点ソースコードの構造は、対応する固定小数点ソースコードと関連づけられている。CodecTypedef.h ファイルに、全ての浮動小数点変数および定数を倍精度型または単精度型として定義した記述が含まれている。ソフトウェアファイル名とその概要のリストを、JT-G729. 1 付属資料Dの Table D.1 から Table D.3 に示す。ここで、基本演算や数値演算に関するファイルは浮動小数点演算では使用されないことに留意すること。また、float から short への変換ルーチンがファイル G729EV\_TDAC\_util.c に加えられている。

Table D.1/JT-G729.1 Summary of encoder specific routines  
(ITU-T G.729.1)

<b>Filename</b>	<b>Description</b>
G729EV_CELP2S_acelp_ca.c	CELP2S fixed codebook search
G729EV_CELP2S_encod.c	CELP2S encoder routine
G729EV_FEC_ferenc.c	FEC encoder routine
G729EV_G729_acelp_ca.c	G729EV fixed codebook search
G729EV_G729_lpc.c	G729 LP analysis
G729EV_G729_pitch.c	G729EV pitch search
G729EV_G729_pwf.c	G729 computation of perceptual weighting coefficients
G729EV_G729_qua_gain.c	G729 gain quantizer
G729EV_G729_qua_lsp.c	G729 LSP quantizer
G729EV_G729B_dtx.c	DTX/SID routine with G.729B embedded
G729EV_G729B_qsidlsf.c	LSF vector quantization routine
G729EV_MAIN_encod.c	MAIN encoder routine
G729EV_TDAC_encod.c	TDAC encoder routine
G729EV_TDBWE_encoder.c	TDBWE encoder routine
G729EV_TDBWE_vector_quantization.c	TDBWE vector quantization

Table D.2/JT-G729.1 Summary of decoder specific routines  
(ITU-T G.729.1)

Filename	Description
G729EV_CELP2S_decod.c	CELP2S decoder routine
G729EV_CELP2S_syn.c	CELP2S core synthesis functions
G729EV_CELP2S_post.c	CELP2S postfiltering
G729EV_FEC_clasdec.c	FEC signal classification routines
G729EV_FEC_decbfi.c	FEC Frame Erasure Concealment functions
G729EV_FEC_ferdec.c	FEC information decoder
G729EV_FEC_onset.c	FEC codebook reconstruction routines
G729EV_FEC_pit_updt.c	FEC pitch update functions
G729EV_G729_de_acealp.c	G729 algebraic codebook decoding
G729EV_G729_dec_gain.c	G729 gain decoding
G729EV_G729_dec_lag3.c	G729 adaptive-codebook index decoding
G729EV_G729_lspdec.c	G729 LSP decoding
G729EV_G729_pst.c	G729 postfilter routines
G729EV_G729B_dec_sid.c	G729B SID decoder
G729EV_G729B_util.C	G729B utility routines
G729EV_MAIN_decod.c	MAIN decoder routine
G729EV_MAIN_envadaption.c	MAIN pre/post echo reduction routines
G729EV_TDAC_decod.c	TDAC decoder routine
G729EV_TDAC_post.c	TDAC post processing
G729EV_TDBWE_compression.c	TDBWE post processing
G729EV_TDBWE_decoder.c	TDBWE decoder routine
G729EV_TDBWE_fir.c	TDBWE filter functions
G729EV_TDBWE_frequency_envelope_shaping.c	TDBWE frequency envelope shaping functions
G729EV_TDBWE_generate_excitation.c	TDBWE excitation generation functions
G729EV_TDBWE_time_envelope_shaping.c	TDBWE time envelope shaping functions

付属資料E 超広帯域スケーラブル拡張  
(標準JT-G729.1に対する)

E. 1 適用範囲

この付属資料は、36-64 kbit/s のビットレートで動作し、JT-G729 および JT-G729.1 と相互接続できるスケーラブル超広帯域 (SWB, 50-14000 Hz) 音声およびオーディオ符号化アルゴリズムを記述する。

本勧告は、次のように構成されている。本勧告で使用される略語と数式は、E.3 および E.4 にそれぞれ定義されている。E.5 は、アルゴリズムの概要を記載している。符号化器と復号器の原理は、E.6 および E.7 にそれぞれ記述されている。伝送パラメータは、E.8 に記載されている。そして E.9 は、本符号化器を定義する 16-32 ビットの固定小数点演算ソフトウェアについて記述している。

E. 2 概要

JT-G729.1 SWB 符号化器の出力の帯域幅は 50 - 14000Hz である。符号化器は 20 ms のフレーム単位で動作し、アルゴリズム遅延は 55.6875 ms である。デフォルトで、符号化器の入力および復号器の出力のサンプリング周波数は 32kHz である。

超広帯域符号化器は、JT-G729.1 の 12 レイヤに加えて、36-64 kbit/s のビットレートに対応する 5 レイヤのエンベディッドビットストリームを生成する。このビットストリームは、復号器側もしくは通信システムの任意の場所で切り詰められることにより、アウトバンド信号を使用することなく、所望のビットレートに瞬時に調整することができる。ビットレートが 32 kbit/s の場合、JT-G729.1 SWB は、JT-G729.1 と完全に相互接続が可能である。

基本的なアルゴリズムは 2 段階の符号化構造、すなわち 7000-14000Hz の帯域に対する修正離散コサイン変換(MDCT)領域の帯域拡張、および 50-7000Hz の帯域に対する MDCT 係数の誤差成分のベクトル量子化に基づいている。

この勧告は、ANSI C ソースコードおよびテストベクトルを含む。テストベクトルは、ITU-T テスト信号データベース(<http://www.itu.int/net/ITU-T/sigdb/speaudio/Gseries.htm#G.729.1>)で参照することができる。

E. 3 略語

この勧告で使用される略語を、Table E.1/JT-G729.1 にまとめて示す。

Table E.1/JT-G729.1 – Glossary of acronyms

Acronym	Description
ACELP	Algebraic CELP
CELP	Code-Excited Linear Prediction
DFT	Discrete Fourier Transform
FEC	Frame Erasure Concealment
GLCVQ	Gosset Low Complexity Vector Quantization
HF	High Frequency (7-14 kHz)
IDFT	Inverse Discrete Fourier Transform
IIR	Infinite Impulse Response
iMDCT	Inverse MDCT
LF	Low Frequency (0-7 kHz)
MDCT	Modified Discrete Cosine Transform
NB	Narrowband
SVQ	Spherical Vector Quantization

Table E.1/JT-G729.1 – Glossary of acronyms

Acronym	Description
SWB	Superwideband
TB	Transition Band
VQ	Vector Quantization
WB	Wideband
WMOPS	Weighted Million Operations Per Second

#### E. 4 数学的表現

本勧告においては、数学的な表現を以下のように定義する。

- $\lfloor x \rfloor$  は  $x$  以下の最大の整数値を表す。  $\lfloor 1.1 \rfloor = 1$ ,  $\lfloor 1.0 \rfloor = 1$  および  $\lfloor -1.1 \rfloor = -2$ ;
- $\lceil x \rceil$  は  $x$  以上の最小の整数値を表す。  $\lceil 1.1 \rceil = 2$ ,  $\lceil 2.0 \rceil = 2$  および  $\lceil -1.1 \rceil = -1$ ;
- $|x|$  は  $x$  の絶対値を表す。  $|17| = 17$ ,  $|-17| = 17$ ;

$\min(x_0, x_1, \dots, x_{N-1})$  は  $x_0, x_1, \dots, x_{N-1}$  の最小値を表す。ここで  $N$  は要素数を表す。

$\max(x_0, x_1, \dots, x_{N-1})$  は  $x_0, x_1, \dots, x_{N-1}$  の最大値を表す。

- $\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{otherwise} \end{cases}$

–  $\mathbf{A}^T$  は行列  $\mathbf{A}$  の転置を表す。

–  $x \bmod y$  は  $x$  割る  $y$  の剰余を表す。  $x \bmod y = x - (y \lfloor x/y \rfloor)$ ;

–  $\text{round}(x)$  は  $x$  の丸め処理を表す。  $\text{round}(x) = \text{sgn}(x) \cdot \lfloor |x| + 0.5 \rfloor$ ;

–  $\exp(x)$  は  $e^x$  と等価であり、 $e$  は自然対数の底を表す。

–  $\sum$  は総和を表す。

–  $\prod$  は総乗を表す。

–  $\wedge$  は論理積を表す。

–  $\vee$  は論理和を表す。

–  $\text{Re}(z)$  は複素数  $z$  の実部を表す。

–  $\text{Im}(z)$  は複素数  $z$  の虚部を表す。

特に記述がない限り、本勧告において  $\log(x)$  は 10 を底とする対数を表すことにする。

#### E. 5 コーデックの概要

JT-G729.1 の超広帯域拡張は、JT-G729.1 の広帯域符号化を超広帯域に拡張する、ビットレートが 36、40、48、56 および 64 kbit/s の 5 つのレイヤにより構成される。第 1 の SWB 拡張レイヤは 2 モードの符号化、すなわち汎用モードあるいは正弦波モードによる高周波の符号化、により構成される。汎用モードでは、符号化された広帯域信号の適応的な複製に基づき符号化を行う。正弦波モードの符号化では、高周波成分に正弦波を加える。モード選択は、入力信号の特性に依存する。第 2 の拡張レイヤは低周波と高周波の間で適応ビット配分を行い品質を改善し、第 3 のレイヤは正弦波を追加することにより高周波の品質を改善する。第 4 と第 5 の拡張レイヤは広帯域信号を改善する。

拡張レイヤで用いられる技術およびビットレートを Table E.2/JT-G729.1 に示す。

Table E.2/JT-G729.1 Layer structure of JT-G729.1 superwideband extension

Layer	Bit rate	Technology
JT-G729.1 codec	32 kbit/s	Embedded NB/WB codec comprising 12 layers
Layer 6mo	+4 kbit/s	Generic mode/Sinusoidal mode HF coding
Layer 7mo	+4 kbit/s	Additional HF sinusoids/Wideband improvement
Layer 8mo	+8 kbit/s	Additional HF sinusoids
Layer 9mo	+8 kbit/s	Wideband improvement
Layer 10mo	+8 kbit/s	Wideband improvement

### E. 5. 1 入力/出力サンプリングレート

JT-G729.1 SWB 拡張は、SWB 符号化のために 32kHz でサンプリングされた信号を処理する。そして SWB 拡張が使用されない場合には、16 もしくは 8kHz でサンプリングされた広帯域(WB)信号あるいは狭帯域(NB)信号を JT-G729.1 として処理する。同様に、32kHz サンプリングの SWB 変換を用いて、復号器出力は 32、16、8kHz でサンプリングされた信号が出力される。

### E. 5. 2 アルゴリズム遅延

入力信号は、JT-G729.1 WB 符号化に対応して 20 ms のフレーム長で処理される。信号のサンプリングレート変換および後処理に起因する追加アルゴリズム遅延は、6.75 ms である。

### E. 5. 3 演算量とメモリ量

テストベクトルを用いて算出された JT-G729.1 SWB 符号化および復号化の演算量の最悪値は、ビットレートスイッチング条件における 62.55WMOPS である。Table E.3/JT-G729.1 は、32kHz サンプリングの JT-G729.1 出力(32 kbit/s)へのスイッチングを含む様々なビットレート条件の演算量を要約したものである。

Table E.3/JT-G729.1 Complexity (in WMOPS)

Option	Bit rate	Encoder	Decoder	Total
SWB	32 kbit/s	22.31	20.36 (Note)	42.67
	36 kbit/s	33.61	22.35 (Note)	55.96
	40 kbit/s	34.36	22.51 (Note)	56.87
	48 kbit/s	34.95	19.14 (Note)	54.09
	56 kbit/s	38.70	20.49	59.19
	64 kbit/s	38.67	21.06 (Note)	59.73
Bit-rate switching	–	38.72	24.55	63.27
NOTE – Calculated with 15% frame erasure rate.				

Table E.4/JT-G729.1 は、SWB 拡張のメモリ量を示している。テーブル ROM は合計で約 13.2 kWords、スタティック RAM は合計で約 5.7 kWords である。符号化器と復号器で使用されるダイナミック RAM の最悪値は、約 9.9 kWords となる。

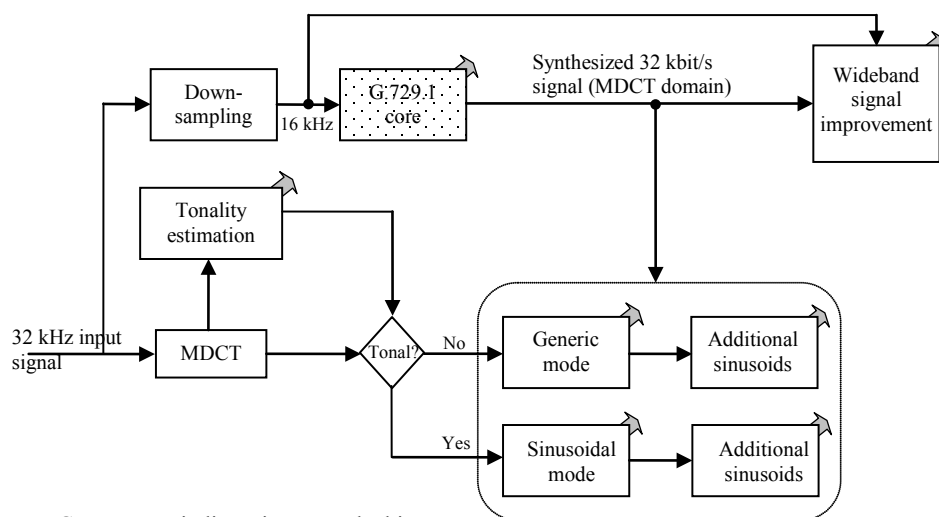
Table E.4/JT-G729.1 Memory

Memory	Encoder	Decoder	Common	Total
ROM (kWord)	–	4.3	8.9	13.2
RAM (kWord)	1.6	4.1	–	5.7

### E. 6 符号化器の機能記述

符号化器の構成を Figure E.1/JT-G729.1 に示す。この図から、JT-G729.1 コアコーデックは 16kHz の信号上

で動作するのに対し、SWB 拡張は 32kHz の信号を使用することが分かる。SWB の符号化は MDCT 領域上で行われる。選択される 2 つのモード、汎用モードおよび正弦波モードが、SWB 拡張の最初のレイヤとして用いられる [b-SWB]。このモードの決定は、入力信号のトーナリティの推定値に基づいて行われる。高位の SWB レイヤは、高域成分の品質を改善する正弦波信号の追加、あるいは広帯域成分の知覚品質を改善する広帯域信号の高品質化のいずれかを用いて符号化を行う。



Note: Grey arrows indicate inputs to the bitstream.

Figure E.1/JT-G729.1 – Structural block diagram of the encoder

### E. 6. 1 サンプリング変換

JT-G729.1 コア符号化器に対して、32kHz の入力信号は 16kHz にダウンサンプリングされる。その変換は、20 ms の入力フレーム単位に行われる。まず、低周波成分を抑圧するために、入力信号  $s_{inp}(n)$  をカットオフ周波数が 50Hz の 2 次の高域通過フィルタに通す。次にその出力信号  $s_{32}(n)$  を、2 段の無限インパルス応答 (IIR) フィルタに通してダウンサンプリングを行う。IIR フィルタは、以下に示す 1 段目の数式、

$$s_{stage1}(n) = \sum_{i=0}^5 b_{i,stage1} s_{32}(n-i) - \sum_{j=1}^5 a_{j,stage1} s_{stage1}(n-j) \quad n = 0, \dots, 639 \quad (E-1)$$

そして、2 段目の数式で表される。

$$s_{stage2}(n) = \sum_{i=0}^4 b_{i,stage2} s_{stage1}(n-i) - \sum_{j=1}^4 a_{j,stage2} s_{stage2}(n-j) \quad n = 0, \dots, 639 \quad (E-2)$$

ここで、負値のインデックスで示される信号は、メモリに記憶されている前フレームの信号を表す。フィルタ係数  $b_{k,stage1}$  および  $a_{k,stage1}$  は、Table E.5/JT-G729.1 に示される。

最後に、ダウンサンプル信号  $s_{16}(n)$  は次のように求められる。

$$s_{16}(n) = s_{stage2}(k) \quad n = 0, \dots, 319 \quad k = 0, 2, \dots, 638 \quad (E-3)$$



Table E.5/JT-G729.1 – Downsampling filter coefficients

$b_{k,stage1}$	$a_{k,stage1}$	$b_{k,stage2}$	$a_{k,stage2}$
0.0581908	1.0	0.0708131	1.0
0.1273200	-1.462269006103	0.1446910	-1.308055753351
0.2007510	2.194832312468	0.1941500	1.509406808301
0.2007510	-1.638502434011	0.1446910	-0.8377333289311
0.1273200	0.926185061962	0.0708131	0.2783957005103
0.0581908	-0.2701781844527		

### E. 6. 2 WB符号化

WB符号化は、ここで定義されたダウンサンプル後の入力信号  $s_{16}(n)$  に対して、TTC規格 JT-G729.1 に従って行われる。

### E. 6. 3 MDCT領域への変換

入力される 32kHz の信号は MDCT 領域へ変換される。窓は WB 符号化器の MDCT と同期させなければならない。つまり、JT-G.729.1 符号化器および 16kHz へのダウンサンプリング処理に起因する遅延を考慮して、入力フレームを遅延させる。

変換の長さは、32kHz 超広帯域信号に対し  $N=1280$  (前フレームの  $M=640$  ( $=N/2$ ) サンプルと現フレームの 640 サンプルにより構成) である。

$$x(n) = s'_{32}(n - N/2) \quad n = 0, \dots, N-1 \quad (\text{E-4})$$

ここで、 $s'_{32}(n)$  は、JT-G729.1 コア符号化を実施するために  $s_{32}(n)$  を遅延させた信号である。

順変換は、次の方程式によって公式化される。

$$M_{32}(2k) = A \cdot \text{Re} \left\{ b \cdot W_N^{k+0.5} \sum_{n=0}^{N/4-1} u(n) \cdot a \cdot W_N^{n+0.5} W_{N/4}^{nk} \right\} \quad k = 0, \dots, N/4-1 \quad (\text{E-5})$$

$$M_{32}(N/2-1-2k) = A \cdot \text{Im} \left\{ b \cdot W_N^{k+0.5} \sum_{n=0}^{N/4-1} u(n) \cdot a \cdot W_N^{n+0.5} W_{N/4}^{nk} \right\} \quad k = 0, \dots, N/4-1 \quad (\text{E-6})$$

ここで、 $W_N = e^{-2\pi j/N} = \cos(2\pi/N) - j \sin(2\pi/N)$  と  $u(n)$  は窓掛けし、回旋した(twiddling)後のサンプルを表す。

順変換は 4 ステップに分割することができる。

#### 1) 前処理

前処理は、前処理信号  $u(n)$  を得るために窓掛けし回旋する処理を含む。

$$y(n) = h(n)x(n) \quad n = 0, \dots, N-1 \quad (\text{E-7})$$

ここで、 $h(n)$  は分析窓を表す。合成窓が  $g(n)$  と定義される場合、分析窓および合成窓は完全再構成条件を満足しなければならない。

$$h(n)g(n) + h(n + N/2)g(n + N/2) = 1, n = 0, \dots, N - N/2 - 1 \quad (\text{E-8})$$

変換係数の数を変えずに変換符号化時に生じる遅延を 10 ms 短くするために、非対称窓が使用される。従来の MDCT 窓とは異なり、この窓は対称ではない。窓の後半部は、前半部を時間軸上で逆順にしたものではない。

非対称分析窓の形状は、16kHz の入力信号用に設計されており、次の方程式に示されるように 6.11.2 節 [b-JT-G718]に基づいている。

$$w_a(n) = \frac{w_i(n)}{\sqrt{D(n)}} \quad 0 \leq n < M \quad (\text{E-9})$$

ここで、

$$w_i(n) = \begin{cases} \sin \left[ \left( n + \frac{1}{2} \right) \frac{\pi}{\left( M - \frac{M}{8} \right)} \right] & 0 \leq n < M - \frac{M}{8} \\ 0, & M - \frac{M}{8} \leq n < M \end{cases} \quad (\text{E-10})$$

そして、 $D(n)$ は、

$$D(n) = w_i(n)w_i(M-1-n) + w_i\left(n + \frac{M}{2}\right)w_i\left(\frac{M}{2} - 1 - n\right) \quad 0 \leq n < \frac{M}{2} \quad (\text{E-11})$$

$$D\left(n + \frac{M}{2}\right) = D(n) \quad 0 \leq n < \frac{M}{2} \quad (\text{E-12})$$

と定義される。ここで、 $M/8$  は補充されるゼロの数を表す。

32kHz でサンプリングされた入力信号を扱うために、この窓は、MDCT のサイズが  $M/2$  ではなく  $M$  になるように更新および補間される。従って、 $N$  サンプルの窓は 2 つの分離された配列に格納される。最初の配列は、6.11.2 節 [b-JT-G718]で行われているように、非対称窓  $w_a(n)$  から直接得られた  $M$  タップを含む。第 2 の配列は、補間窓  $M$  タップを含む。補間処理は、窓  $w_a(n)$  の各係数間に係数を挿入し、元の窓の 2 倍の窓長がある新しい窓を定義する。

補間窓は次のように構成される。

偶数項は、6.11.2 節 [b-JT-G718]で定義される非対称窓のタップを直接用いる。

$$\begin{cases} h(2n) = w_a(n) \\ h(M+2n) = w_a(M/2+n) \end{cases} \quad 0 \leq n < M/2 \quad (\text{E-13})$$

そして奇数項は、 $w_a(M)=0$  として線形に補間される。

$$\begin{cases} h(2n+1) = P_n \cdot [w_a(n) + w_a(n+1)] \\ h(M+2n+1) = P_n \cdot [w_a(M/2+n) + w_a(M/2+n+1)] \end{cases} \quad (\text{E-14})$$

ここで修正係数  $P_n$  は、MDCT の段階で完全再構成を保証する。

$$P_n = [1 + w_a(n+1) \cdot w_a(M-1-n) + w_a(M/2+n+1) \cdot w_a(M/2-1-n)]^{-1} \quad (\text{E-15})$$

このようにサイズが  $N=2M$  の更新された窓に対して最終的に補充されるゼロの数は、 $M_4=N/8$  となる。

窓掛け後に、 $y(n)$  の要素は次の処理のようにして回旋される。

$$\begin{cases} z(n + N/4) = y(n) - y(N/2 - 1 - n) \\ z(N/4 - 1 - n) = -y(N - 1 - n) - y(N/2 + n) \end{cases} \quad n = 0, \dots, N/4 - 1 \quad (\text{E-16})$$

および、

$$u(n) = z(2n) + jz(N/2 - 1 - 2n) \quad n = 0, \dots, N/4 - 1 \quad (\text{E-17})$$

となる。ここで、 $u(n)$  は前処理後のデータ、そしてそれは複素数であり、 $z(2n)$  は実部、 $z(N/2 - 1 - 2n)$  は虚部を表す。

## 2) 前回転

回転処理は回旋されたデータに対して行われる。

$$v(n) = u(n) \cdot a \cdot W_N^{n+0.5} \quad n = 0, \dots, N/4 - 1 \quad (\text{E-18})$$

ここで、 $a$  は回転係数  $W_N^{n+0.5}$  へのスカラ値、 $W_N = e^{-2\pi j/N} = \cos(2\pi/N) - j \sin(2\pi/N)$  を表す。演算量を削減するため、スカラ値は回転係数へ反映させる。ここでは、 $a = \sqrt{2}/\sqrt[4]{N}$  を用いる。回転係数  $W_N^{n+0.5}$  の実部および虚部は次の条件を満たすので、

$$\cos\left(\frac{2\pi}{N}(n+0.5)\right) = \sin\left(\frac{2\pi}{N}(L-1-n+0.5)\right) \quad (\text{E-19})$$

スカラ値を反映した回転係数の実部および虚部は等しく、実部(または虚部)  $\cos\left(\frac{2\pi}{N}(n+0.5)\right)/\sqrt[4]{320}$   $n = 0, \dots, 319$  は  $N/4$  点のテーブルとして記憶される。

## 3) 離散フーリエ変換(DFT)

長さ  $L$  ( $L = N/4 = 320$ ) の複素 DFT を回転後のデータ  $v(n)$  に適用する。

$$r(k) = \sum_{n=0}^{L-1} v(n) W_L^{nk} \quad k = 0, \dots, L-1 \quad (\text{E-20})$$

ここで、単純な 2 のべき乗の DFT は適切ではない。よって、次に示す低演算量の 2 次元 ( $L = N/4 = P \times Q$ ) DFT を適用する。ここで、 $P=64$  および  $Q=5$  は素因数である。

演算量を削減するため、アドレステーブルを用いる。アドレステーブルは次のように算出される。

$$I(n_1, n_2) = (K_1 \times n_1 + K_2 \times n_2) \bmod L \quad \begin{array}{l} n_1 = 0, \dots, P-1 \\ n_2 = 0, \dots, Q-1 \end{array} \quad (\text{E-21})$$

ここで  $K_1, K_2$  は素因数を表し、条件  $(K_1 K_2) \bmod L = 0$  を満たす。

ここでは、 $K_1 = 65, K_2 = 256$  を用いる。アドレステーブル  $I$  は低演算量 2 次元 DFT のために記憶しておき、次のように、どのサンプルが P ポイント DFT もしくは Q ポイント DFT に用いるかを示すために使用される。

a) アドレステーブル  $I$  を用いた Q 回 P ポイント DFT の  $v(n)$  への適用

$i$  番目 ( $i = 0, \dots, Q-1$ ) の P ポイント DFT への入力データは、アドレステーブル  $I$  に格納されたアドレスを探すことにより見つけられる。 $i$  番目の P ポイント DFT に対して、入力データのアドレスは、テーブル  $I$  の  $i * P$  s 要素から開始する P 個の連続的な要素として表される。P ポイント DFT の各々に対し、算出されるデータに  $x$  ステップの巡回シフトを適用する必要がある。ここで  $x$  は再配列インデックスを表し、 $(x \cdot ((K_1^2 / Q) \bmod P)) \bmod P = 1$  を満足する。

ステップ a) の出力は

$$w(k) = \text{DFT}_P(v(I + iP))_x \quad i = 0, \dots, Q-1 \quad (\text{E-22})$$

となる。

$i$  番目 ( $i = 0, \dots, 4$ ) の 64 ポイント DFT に対し、入力データのアドレスは  $I[64i]$  から始まる 64 個の連続する要素となり、その結果、 $x=5$  で巡回シフトされる。ここで巡回シフトの例を示す。元のベクトルを  $Z = [z_0 \ z_1 \ z_2 \ z_3 \ z_4]$  とすると、巡回シフト 2 による新しいベクトルは 2 つの循環シフトを備えた新しいベクトルは  $Z = [z_0 \ z_2 \ z_4 \ z_1 \ z_3]$  となる。

b) アドレステーブル  $I$  を用いた P 回 Q ポイント DFT の  $w(k)$  への適用

$i$  番目 ( $i = 0, \dots, P-1$ ) の Q ポイント DFT への入力データは、アドレステーブル  $I$  に格納されたアドレスを探すことにより見つけられる。 $i$  番目の Q ポイント DFT に対して、入力データのアドレスは、テーブル  $I$  の  $i$  番目の要素から開始する Q 個の要素として表され、各要素は P 個おきに分離される。Q ポイント DFT の各々に対し、算出されるデータに  $y$  ステップの巡回シフトを適用する必要がある。ここで  $y$  は再配列インデックスを表し、 $(y \cdot ((K_2^2 / P) \bmod Q)) \bmod Q = 1$  を満足する。

ステップ b) の出力は、

$$r(k) = \text{DFT}_Q(w(I + i))_y \quad i = 0, \dots, P-1 \quad (\text{E-23})$$

となる。

$i$  番目 ( $i = 0, \dots, 63$ ) の 5 ポイント DFT に対し、入力データのアドレスは  $I[i]$  から始まる 5 個の連続する要素となり、各々はステップ 64 により分離される。その結果、 $y=4$  で巡回シフトされる。

4) 後回転およびスペクトル係数の出力

$r(k)$  に対する DFT 後に、後回転が施される。

$$l(k) = b \cdot W_N^{k+0.5} r(k) \quad k = 0, \dots, N/4 - 1 \quad (\text{E-24})$$

ここで、 $b$  は回転係数に対するスカラー値である。ここでは、 $b = \sqrt{2/4N}$  が用いられる。

最後に、スペクトル係数が次のように算出される。

$$M_{32}(2k) = A \cdot \text{Re}\{l(k)\} \quad k = 0, \dots, N/4 - 1 \quad (\text{E-25})$$

$$M_{32}(N/2 - 1 - 2k) = A \cdot -\text{Im}\{l(k)\} \quad k = 0, \dots, N/4 - 1 \quad (\text{E-26})$$

ここで  $A$  は正規化係数を表し、単純化のため、ここでは 1 に設定される。

#### E. 6. 4 トーナリティの推定

第 1 の SWB レイヤ(レイヤ 6mo)の符号化モードは、トーナリティ推定値に基づいて選択される。その選択は、MDCT 領域での現フレームと前フレームの間の対数領域エネルギーの比較により行われる。

トーナリティの推定は、現フレームと前フレーム間のスペクトルピークの相関分析に基づいている。推定アルゴリズムには、高周波数帯域の MDCT スペクトルの対数エネルギーが入力される。対数エネルギーは次式で定義される。

$$E_{dB}(k - 280) = 10 \log[M_{32}^2(k)] \quad k = 280, \dots, 559 \quad (\text{E-27})$$

ここで、 $M_{32}(k)$  は式 E-25 および E-26 で定義される。この節では、上式で定義されるように、「スペクトル」という用語は MDCT スペクトルの対数エネルギーを指すことにする。推定アルゴリズムは、4 段階により構成される。

##### 1) スペクトルフロアの算出

ここではスペクトルフロアを算出し、スペクトルから減じる。これは移動平均(MA)フィルタの適用により行われる。フィルタの長さは 31 サンプルで、これは実験的に見つけたものである。スペクトルフロアの最初の値は、最初の 31 サンプルの中央のビンである位置  $i=15$  で計算される。以降、 $LMA=15$  はフィルタ長さの半分を表す。

$$E_{fl}(L_{MA}) = \frac{1}{2L_{MA} + 1} \sum_{i=0}^{2L_{MA}} E_{dB}(i) \quad (\text{E-28})$$

スペクトルフロアの次のサンプルは、次式のように再帰的に計算される。

$$E_{fl}(k) = E_{fl}(k-1) + \frac{1}{2L_{MA} + 1} (E_{dB}(k + L_{MA}) - E_{dB}(k - L_{MA} - 1)), \quad k = L_{MA} + 1, \dots, 279 - L_{MA} \quad (\text{E-29})$$

ここまでで算出されない部分、つまり、最初および最後の  $LMA$  サンプルは以下のように外挿される。

$$\begin{aligned} E_{fl}(k) &= 0.9E_{fl}(k+1) + 0.1E_{dB}(k) & k = L_{MA} - 1, \dots, 0 \\ E_{fl}(k) &= 0.9E_{fl}(k-1) + 0.1E_{dB}(k) & k = 280 - L_{MA}, \dots, 279 \end{aligned} \quad (\text{E-30})$$

これにより、原スペクトルから減じられる 280 サンプルのスペクトルフロアが算出される。

$$E_{res}(k) = E_{dB}(k) - E_{fl}(k) \quad k = 0, \dots, 279 \quad (\text{E-31})$$

## 2) 振幅の平滑化と周波数圧縮

スペクトルフロアの減算後、スペクトル(振幅)は次の方法を用いて、X 軸(周波数ビン)方向に平滑化および 1/2 に圧縮される。圧縮はスペクトルの開始部および終了部にて行われる。中央部は平滑化されるのみである。開始部は、圧縮した領域の中の 1 から 44 のビンの範囲で占められる。周波数ビンの圧縮および振幅平滑化は、次のように行われる。

$$X_{dB}(k) = 0.3[0.65E_{res}(2k-2) + E_{res}(2k-1) + E_{res}(2k) + 0.65E_{res}(2k+1)] \quad k = 1, \dots, 44 \quad (\text{E-32})$$

中央部は 45 から 56 のビンにより構成され、この範囲は圧縮を行わない。次のように平滑化のみが行われる。

$$X_{dB}(k) = 0.33[E_{res}(k+42) + E_{res}(k+43) + E_{res}(k+44)] \quad k = 45, \dots, 56 \quad (\text{E-33})$$

終了部は 57 から 145 のビンを含み、開始部と同一の方法で次のように計算される。

$$X_{dB}(k+57) = 0.3[0.65E_{res}(2k+99) + E_{res}(2k+100) + E_{res}(2k+101) + 0.65E_{res}(2k+102)] \quad k = 0, \dots, 88 \quad (\text{E-34})$$

インデックス 0 のサンプルは、圧縮されていないスペクトルからコピーされる。すなわち、 $X_{dB}(0) = E_{res}(0)$  である。インデックス 14 のサンプルは以下のように計算される。

$$X_{dB}(146) = 0.33[E_{res}(277) + E_{res}(278) + E_{res}(279)] \quad (\text{E-35})$$

従って、圧縮されたスペクトルの全長は  $N_{tot}=147$  ビンとなる。XdB のビンがすべて負だった場合、次のステップは実行されず、非トータルと最終判断される。

## 3) 局所最小値の探索

ここでは、次式のようにループ処理により圧縮スペクトルの局所最小値のインデックスを探索し、このインデックスをバッファ  $i_{min}$  に格納する。

$$i_{min} = (\forall i: (X_{dB}(i-1) > X_{dB}(i)) \wedge (X_{dB}(i) < X_{dB}(i+1))) \quad i = 1, \dots, 145 \quad (\text{E-36})$$

もし  $X_{dB}(0) < X_{dB}(1)$  の場合、インデックス 0 は  $i_{min}$  に加算される。同様に、もし  $X_{dB}(146) < X_{dB}(145)$  の場合、インデックス 146 は  $i_{min}$  に加算される。ここで、最小値の総数が  $N_{min}$  だけ見つかったものとする。

#### 4) 相関マップの算出

ここでは、相関マップと長期相関マップの算出を行う。これは区分的処理である。従って、最小値がピークの範囲を特定するので、これはピークごとに実行される。以下、圧縮スペクトル  $X_{dB}$  中の隣接する 2 つの最小値間の区間を示すために用語「ピーク」を用いる。

以下、前フレームの圧縮スペクトルを  $X_{dB}^{[-1]}(k)$  と表す。現フレームの圧縮スペクトル中のピークそれぞれについて、前フレームの圧縮スペクトルを用いて正規化相関を算出する。相関処理は、2 つの隣接する最小値で特定される区間の全てのピークのインデックス(ビン)を考慮する。すなわち、次のような特定の最小値の開始インデックスを、

$$i_{st} = i_{\min}(x) \quad (\text{E-37})$$

そしてその最小値の終了インデックスを

$$i_{end} = i_{\min}(x+1) \quad (\text{E-38})$$

と表す。そのとき、そのピークに対する相関は以下のように表される。

$$M_{cor}(k) = \frac{\left( \sum_{j=i_{st}}^{i_{end}} X_{dB}(j) X_{dB}^{[-1]}(j) \right)^2}{\sum_{j=i_{st}}^{i_{end}} (X_{dB}(j))^2 \sum_{j=i_{st}}^{i_{end}} (X_{dB}^{[-1]}(j))^2} \quad k = i_{st}, \dots, i_{end} - 1 \quad (\text{E-39})$$

次に、上記 3 式を全ての最小値に対して繰り返す、すなわち  $x=0, \dots, N_{\min}-2$  に対して繰り返す。開始ビンおよび終了ビンの  $M_{cor}$  は次のようにゼロに設定する。

$$\begin{aligned} M_{cor}(k) &= 0 & k &= 0, \dots, i_{\min}(0) - 1 \\ M_{cor}(k) &= 0 & k &= i_{\min}(N_{\min} - 1), \dots, 146 \end{aligned} \quad (\text{E-40})$$

そして、式 E-43 による相関マップの総和算出に向け、圧縮スペクトルおよび非圧縮スペクトルの重みを補償するため、相関の中央部を 1/2 にする。

$$M_{cor}(k) = 0.5 \cdot M_{cor}(k) \quad k = 45, \dots, 56 \quad (\text{E-41})$$

ここで配列  $M_{cor}$  を相関マップと呼ぶ。相関マップは、 $X_{dB}$  と同数のビンを持つ。相関マップのイラストを Figure E.2/JT-G729.1 に示す。

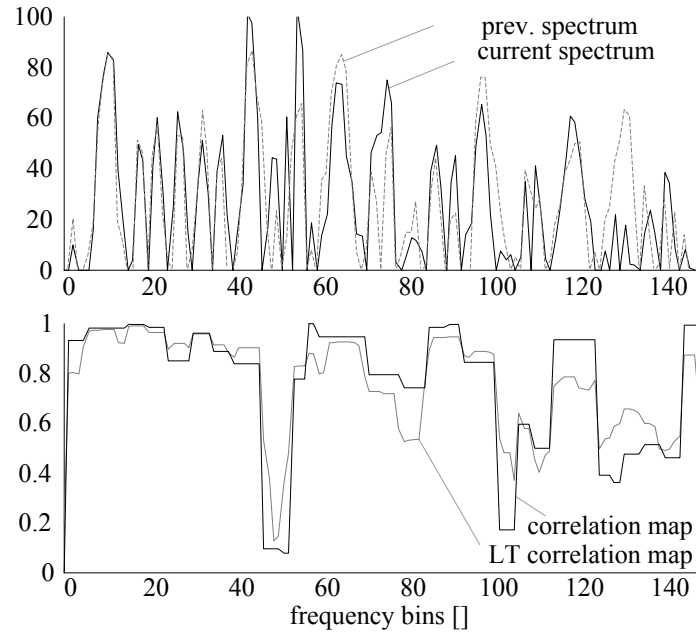


Figure E.2/JT-G729.1 – Calculation of the correlation map and the long-term correlation map

現フレームの相関マップは、次式に示されるように、その長期的な値を更新するために使用される。

$$\bar{M}_{cor}(k) = \alpha_{map} \bar{M}_{cor}(k) + (1 - \alpha_{map}) M_{cor}(k) \quad k = 0, \dots, 146 \quad (\text{E-42})$$

ここで、 $\alpha_{map} = 0.5$  とする。長期相関マップは全ての  $k$  に対してゼロで初期化される。長期相関マップのイラストは Figure E.2/JT-G729.1 に示される。

最後に、 $\bar{M}_{cor}(k)$  の全てのピンの合計を求める。

$$m_{sum} = \sum_{j=0}^{146} \bar{M}_{cor}(j) \quad (\text{E-43})$$

トーンリティは、 $m_{sum}$  と適応閾値  $th_{tonal}$  とに基づいて決定される。この閾値  $th_{tonal}$  は 52.5 に初期化され、フレーム毎に以下のように更新される。なお、上限値は 57.5、および下限値は 47.5 とする。

$$\begin{aligned} &\text{if } (m_{sum} > 52.5) \\ &\quad th_{tonal} = th_{tonal} - 0.5 \\ &\text{else} \\ &\quad th_{tonal} = th_{tonal} + 0.5 \end{aligned}$$

従って、相関が相対的にアクティブ(トナル)な区間を示す場合に閾値は減少し、逆の場合には閾値は増加する。閾値が低い場合、多くのフレームがトナルと判定される。特にアクティブ区間の終了時にそうなる。従って、この適応閾値はハングオーバーとみることできる。

トーンリティ推定は、現フレームがトナルであるかどうかという情報を出力する。もし、 $m_{sum}$  が  $th_{tonal}$  より大きい場合、現フレームはトナル( $tonal=1$ )、そうでなければ非トナル( $tonal=0$ )とする。トーンリテ



ィ情報もまた、復号器に伝送されるビットストリームに含まれる。トーンリティの決定は、バイナリ変数 *tonal* としてコーデックの中で使用される。もし現フレームがトーンルに分類される場合、正弦波モードが使用され、そうでなければ汎用モードが使用される。

#### E. 6. 5 汎用モードの符号化

入力フレームがトーンルと考えられない場合(*tonal*=0)、汎用モードが使用される。汎用モードは、高周波を符号化するために、JT-G729.1 により符号化された MDCT 領域成分  $\hat{M}_{16}(k)$  を利用する。高周波バンド(7-14kHz)は4つのサブバンドに分割され、各サブバンドにおいて、選択された類似度基準の下、符号化されかつ包絡により正規化された広帯域成分と最も類似度の高いバンドが探索される。最も類似度の高いバンドの成分を2つのスケーリング係数にてスケーリングを行い、高周波成分を合成する。第1のスケーリング係数は線形領域、第2のスケーリング係数は対数領域で表される。この高周波成分は、第1のSWB拡張レイヤ(レイヤ6mo)および追加の正弦波レイヤによる追加的な正弦波信号でさらに改善される。

##### E. 6. 5. 1 MDCT領域の合成されたWB成分の獲得

SWB拡張は、32 kbit/s の JT-G729.1 コアコーデックにて量子化された MDCT 領域成分  $\hat{M}_{16}(k)$ ,  $k=0, \dots, 279$  を利用する。

##### E. 6. 5. 2 包絡正規化

低周波成分の MDCT 領域の包絡はしばしば急激に変化し、それは高周波成分に対する良いマッチングを見つけることを困難にする。加えて、スペクトルのダイナミックスの大きな違いがあるかもしれず、それがスケーリング係数の量子化をより困難にする。これら問題を回避するために、処理を進める前に、量子化された WB 成分  $\hat{M}_{16}(k)$  の包絡を正規化する。この正規化は、対数領域で行われる。

まず、量子化された WB 成分  $\hat{M}_{16}(k)$  を 35 個のサブバンドに分割する。各サブバンドは、8 個の MDCT 係数から成る。

$$SB_k(i) = \hat{M}_{16}(8k+i) \quad i=0, \dots, 7; \quad k=0, \dots, 34 \quad (\text{E-44})$$

各サブバンドの前半部の算術平均  $M_a^1(k)$  および後半部の算術平均  $M_a^2(k)$  を以下のように算出する。

$$M_a^1(k) = \frac{\sum_{i=0}^3 |SB_k(i)|}{4} \quad k=0, \dots, 34 \quad (\text{E-45})$$

$$M_a^2(k) = \frac{\sum_{i=4}^7 |SB_k(i)|}{4} \quad k=0, \dots, 34$$

次に、サブバンド毎に、 $M_a^1(k)$  および  $M_a^2(k)$  の対数領域での幾何平均  $L(k)$  を以下のように求める。

$$L(k) = 20 \cdot \log_{10}(M_a^1(k) \cdot M_a^2(k)) \cdot 0.5 \quad k=0, \dots, 34 \quad (\text{E-46})$$

ここでは、線形領域での平方根の計算を、対数領域にて 0.5 を乗ずることにより実現している。

対数関数へのゼロ値の入力を避けるため、 $M_a^1(k) \cdot M_a^2(k)$  がゼロであり、かつ  $(M_a^1(k))^2 + (M_a^2(k))^2$  がゼロを超える場合に、 $L(k)$  は以下のように算出される。

$$L(k) = 20 \cdot \log_{10} \left( (M_a^1(k))^2 + (M_a^2(k))^2 \right) \cdot 0.5 \quad k = 0, \dots, 34 \quad (\text{E-47})$$

もし  $(M_a^1(k))^2 + (M_a^2(k))^2$  がゼロとなる場合、 $L(k)$  は固定小数点実装での 0x0001 の対数値として算出される。

35 個の幾何平均、 $L(k)$ ,  $k = 0, \dots, 34$ , は 7 サンプル幅の移動平均によって平滑化される。平滑化された包絡  $\tilde{L}(k)$  は以下のように求められる。

$$\tilde{L}(k) = \begin{cases} L(k), & k = 0 \\ \frac{L(k-1) + L(k) + L(k+1)}{3}, & k = 1 \\ \frac{L(k-2) + \dots + L(k+2)}{5}, & k = 2 \\ \frac{L(k-3) + \dots + L(k+3)}{7}, & k = 3, \dots, 31 \\ \frac{L(k-2) + \dots + L(k+2)}{5}, & k = 32 \\ \frac{L(k-1) + L(k) + L(k+1)}{3}, & k = 33 \\ L(k), & k = 34 \end{cases} \quad (\text{E-48})$$

平滑化された包絡の逆数  $\tilde{L}(k)$  は、次のようにして対数領域から線形領域へ変換される。

$$\tilde{L}'(k) = 10^{(-1.0) \cdot 0.05 \cdot (\tilde{L}(k))} \quad k = 0, \dots, 34 \quad (\text{E-49})$$

量子化 WB 係数  $\hat{M}_{16}(k)$  に平滑化された包絡の逆数  $\tilde{L}'(k)$  を乗じて、正規化 WB 係数  $\tilde{M}_{16}(k)$  を求める。

$$\tilde{M}_{16}(k) = \hat{M}_{16}(k) \cdot \tilde{L}'(k') \quad k = 8 \cdot k', \dots, 8 \cdot k' + 7 \quad k' = 0, \dots, 34 \quad (\text{E-50})$$

### E. 6. 5. 3 サブバンド探索

高域(HF(7-14kHz))バンド  $M_{32}(k)$  は、以下のように、4 個のサブバンドに分割される。

$$\begin{aligned} M_{32}^0(k) &= M_{32}(k+280), \quad k = 0, \dots, 39 \\ M_{32}^1(k) &= M_{32}(k+320), \quad k = 0, \dots, 69 \\ M_{32}^2(k) &= M_{32}(k+390), \quad k = 0, \dots, 69 \\ M_{32}^3(k) &= M_{32}(k+460), \quad k = 0, \dots, 99 \end{aligned} \quad (\text{E-51})$$

サブバンド  $d^j$  ( $j = 0, \dots, 3$ ) の幅は、MDCT 領域で各々 40, 70, 70, 100 サンプルである。各々のサブバンド  $M_{32}(k)$  については、最適位置が探索される探索バンド  $k = k^j, \dots, k^j + w^j - 1$  があり、バンド  $j$  については、 $w^j$  の異なる位置がある。

偶数サブバンドの探索の際の開始位置は固定されているのに対し、奇数サブバンドの開始位置はより低域のサブバンドの最適位置に依存して適応的に変化する。

開始位置  $k^j$  および各サブバンドの探索範囲  $w^j$  は、以下のように設定される。

$$k^j = \begin{cases} 0 & j=0 \\ \text{BestIdx}^{j-1} + d^{j-1} - \frac{w^j}{2} & j=1 \\ 0 & j=2 \\ \text{BestIdx}^{j-1} + d^{j-1} - \frac{w^j}{2} & j=3 \end{cases} \quad (\text{E-52})$$

$$w^j = \begin{cases} 240 & j=0 \\ 128 & j=1 \\ 210 & j=2 \\ 128 & j=3 \end{cases} \quad (\text{E-53})$$

ここで  $\text{BestIdx}^j$  は、次の手順によって決定されるバンド  $j$  の最適位置を表す。式 E 52 において、 $k^j$  が負値を取る場合、 $k^j$  は 0 にする。また、 $k^j$  が  $280 - d^j - w^j$  より大きくなる場合、 $k^j$  は  $280 - d^j - w^j$  とする。

各サブバンドの最適位置は、以下のように探索される。インデックス  $k$  の相関値は

$$\text{corr}(k') = \sum_{k=0}^{k=d^j-1} M_{32}^j(k) \tilde{M}_{16}(k^j + k' + k) \quad k'=0, \dots, w^j - 1 \quad (\text{E-54})$$

と計算される。同様に、インデックス  $k$  のエネルギーは

$$\text{Ene}(k') = \sum_{k=0}^{k=d^j-1} M_{16}(k^j + k' + k)^2 \quad k'=0, \dots, w^j - 1 \quad (\text{E-55})$$

と求められる。実際に用いられる類似度は次式で表される。

$$S(k') = \left| \frac{\text{corr}(k')}{\sqrt{\text{Ene}(k')}} \right| \quad k'=0, \dots, w^j - 1 \quad (\text{E-56})$$

この処理は、 $S(k)$  を最大とする  $k$  を見つけ出すことを目的とする。 $S(k)$  の 2 乗を算出することにより、絶対値や分母の平方根の算出を回避し、この処理の演算量を削減することができる。サブバンド  $j$  に対する最適位置は、次のようにして、効率的に探索される。

```

BestIdxj = 0
lagCorr = 0
lagEnergy = 1e30
for k' = 0 to wj - 1
    if (Ene(k') > 0)
        if (lagCorr2 Ene(k') < corr(k')2 lagEnergy)
            BestIdxj = k'
            lagCorr = corr(k')
            lagEnergy = Ene(k')
        end
    end
end
end

```

最適位置のインデックス  $\text{BestIdx}^j$  はパラメータ  $\text{LagIndex}^j$  としてビットストリームに詰められる。

#### E. 6. 5. 4 第1スケーリング

サブバンド探索において最適位置が見つかったならば、原信号のサブバンド  $M_{32}^j(k)$  と聴覚的に最も類似するように、2つのスケーリング係数が用いられる。第1スケーリング係数  $\alpha_1(j)$  は、スペクトルの中の高振幅ピークに匹敵するように線形領域で動作する。第1スケーリング係数は、サブバンド  $j$  に対して、

$$\alpha_1(j) = \frac{\text{corr}(\text{BestIdx}^j)}{\text{Ene}(\text{BestIdx}^j)} \quad j = 0, \dots, 3 \quad (\text{E-57})$$

と求められる。ここで  $\text{BestIdx}^j$  は、サブバンド  $j$  に対する最適位置を表す。スケーリング係数  $\alpha_1(j)$  は正値および負値の両者を取りえる。符号およびスケーリング係数の絶対値は、別々に量子化される。符号値は、

$$\text{Sign\_subband}(j) = \begin{cases} 0, & \alpha_1(j) \geq 0 \\ 1, & \text{otherwise} \end{cases} \quad (\text{E-58})$$

と定義される。

スケーリング係数の絶対値は、対数領域で量子化される。対数領域への変換は次のように行われる。

$$\alpha_1^{\text{Log}}(j) = \begin{cases} 0, & \alpha_1(j) = 0 \\ \log_{10}(|\alpha_1(j) + 1e-15|), & \text{otherwise} \end{cases} \quad (\text{E-59})$$

4つの対数領域のスケーリング係数は、2つのベクトルにグループ化される。

$$\begin{aligned} V_1 &= [\alpha_1^{\text{Log}}(0) \quad \alpha_1^{\text{Log}}(2)] \\ V_2 &= [\alpha_1^{\text{Log}}(1) \quad \alpha_1^{\text{Log}}(3)] \end{aligned} \quad (\text{E-60})$$

#### E. 6. 5. 5 第1成分で並び替えられたコードブックによるベクトル量子化

この2つのベクトルは、8ビットのコードブック  $C\_2\_8$  を用いた2次元ベクトル量子化によって量子化される。最近傍探索法は、第2のスケーリング係数(対数領域スケーリング係数)および後節で述べられる正弦値の量子化と共通の手法である。ここで、最近傍探索法の説明を行う。

入力ベクトル  $\mathbf{v}$  を考えたとき、二分探索法を用いてコードブックを探索する。このコードブックは、第1成分から順に増分するように並び替えられている。二分探索法は、入力ベクトルの第1成分よりも大きい第1成分を持つコードベクトルに対応する最小インデックス  $\text{idx1}$  をコードブックから返す。二分探索より与えられるインデックスが表すコードベクトルから開始して、仮に、第1成分に対してのみ算出される部分歪みが現時点での最小歪みよりも大きい場合に探索を停止するように、修正部分歪み探索が実行される。順序付けられたコードブックの探索の方向は、インデックス  $\text{idx1}$  からは上向きに、 $\text{idx1}-1$  からは下向きとなる。もし、 $\text{idx1}$  がコードベクトルの数と等しい場合、下向きの探索のみが行われる。逆に、 $\text{idx1}$  が0となる場合、上向きの探索のみが行われる。両方向の探索が行われる場合、最適コードベクトルは、両方向の修正部分探索歪みから求められる2つのコードベクトルから選択される。

次節で述べられるように、この手法は1次元、3次元、および4次元コードブックにも適用される。さらに、4ビットのコードブック  $E\_1\_4$  を用いる1次元量子化器に対しては、最適なインデックスを選択するために、 $\text{idx1}-1$  と  $\text{idx1}$  のコードブックインデックスに対する歪みのみ計算すれば良い。また、8ビット量子化器のコードブック  $E\_3\_8$  と  $E\_4\_8$  を用いる3次元VQと4次元VQの一つに対しては、その汎用的な手法が2回実行される。まず、前半の128個のコードベクトルの中から最適コードベクトルを選択し、次に後半の

128 個のコードベクトルの中から最適コードベクトルを選択する。最終的な最適コードベクトルは、2 つの汎用探索より求められるコードベクトルの中から選択される。

量子化の結果、対数領域パラメータの量子化値  $\hat{\alpha}_1^{Log}(j)$  が得られる。これらは再び、次式を用いて線形領域に変換される。

$$\hat{\alpha}_1(j) = 10^{\hat{\alpha}_1^{Log}(j)} \quad (\text{E-61})$$

そして、元の符号が次のように反映される。

$$\hat{\alpha}_1(j) = \begin{cases} -\hat{\alpha}_1(j), & \text{if } (\text{Sign\_subband}(j) \neq 0) \\ \hat{\alpha}_1(j), & \text{otherwise} \end{cases} \quad (\text{E-62})$$

#### E. 6. 5. 6 第2スケーリング

第2のスケーリング係数は対数領域上で適用され、より適切なエネルギーおよび対数領域での形状を得るために用いられる。第2のスケーリング係数算出のために、原信号の高域成分  $M_{32}(k)$  と暫定高域成分の両者の対数領域の表現が必要である。

サブバンド  $j=0, \dots, 3$  各々の暫定高域成分は

$$\dot{M}_{32}^j(k) = \tilde{M}_{16}(k^j + k) \hat{\alpha}_1(j) \quad k = 0, \dots, d^j - 1 \quad (\text{E-63})$$

と求められる。ここで、 $k^j$  はサブバンド探索(E.6.5.3 節)で決定されるパラメータ  $\text{LagIndex}^j$  により、次のように生成される。

$$k^j = \begin{cases} \text{LagIndex}^j & j = 0, 2 \\ \min(\max(0, \text{LagIndex}^{j-1} + d^{j-1} - \frac{w^j}{2}), 280 - d^j - w^j) + \text{LagIndex}^j & j = 1, 3 \end{cases} \quad (\text{E-64})$$

4 サブバンドで構成される暫定高域成分は次のように表される。

$$\dot{M}_{32}(k) = \begin{cases} \dot{M}_{32}^0(k-280), & k = 280, \dots, 319 \\ \dot{M}_{32}^1(k-320), & k = 320, \dots, 389 \\ \dot{M}_{32}^2(k-390), & k = 390, \dots, 459 \\ \dot{M}_{32}^3(k-460), & k = 460, \dots, 559 \end{cases} \quad (\text{E-65})$$

偶数値  $k$  については、原信号の高域部は次式のように対数領域に変換される。

$$\hat{M}_{32}(k) = 20 \cdot \log_{10}(|M_{32}(k) + 1e-15|) \quad k = 280, 282, \dots, 558 \quad (\text{E-66})$$

全ての  $\hat{M}_{32}(k)$  は次のように4個のサブバンドにより構成される。

$$\begin{aligned} \hat{M}_{32}^0(k) &= \hat{M}_{32}(k+280), & k &= 0, 2, \dots, 38 \\ \hat{M}_{32}^1(k) &= \hat{M}_{32}(k+320), & k &= 0, 2, \dots, 68 \\ \hat{M}_{32}^2(k) &= \hat{M}_{32}(k+390), & k &= 0, 2, \dots, 68 \\ \hat{M}_{32}^3(k) &= \hat{M}_{32}(k+460), & k &= 0, 2, \dots, 98 \end{aligned} \quad (\text{E-67})$$

偶数値  $k$  に対して、合成高域成分も同様に対数領域に変換され、

$$\hat{M}_{32}(k) = 20 \cdot \log_{10}(|\dot{M}_{32}(k) + 1e-15|) \quad k = 280, 282, \dots, 558 \quad (\text{E-68})$$

これは次式と等しい。

$$\hat{M}_{32}^j(k) = 20 \cdot \log_{10}(|\dot{M}_{32}^j(k) + 1e-15|) \quad j = 0, \dots, 3 \quad k = 0, 2, \dots, d^j - 2 \quad (\text{E-69})$$

対数領域の各サブバンドに対して、最大値が探索される。

$$m^j = \max\left(-10000, \max_k\left(\hat{M}_{32}^j(k)\right)\right) \quad j = 0, \dots, 3 \quad k = 0, 2, \dots, d^j - 2 \quad (\text{E-70})$$

各サブバンドに対して、式 E-70 を満たす位置は  $k_{\max}^j$  として示される。相関値とエネルギーは対数領域のサブバンドに対して計算され、最大値が減じられる。この処理は、対数領域の暫定高域成分の偶数サンプルに対して行われる。

$$\text{LogCorr} = \sum_k \left( \hat{M}_{32}^j(k) - m^j \right) \left( \hat{M}_{32}^j(k) - m^j \right) \quad j = 0, \dots, 3 \quad k = 0, 2, \dots, d^j - 2 \quad (\text{E-71})$$

$$\text{LogEne} = \sum_k \left( \hat{M}_{32}^j(k) - m^j \right)^2 \quad j = 0, \dots, 3 \quad k = 0, 2, \dots, d^j - 2 \quad (\text{E-72})$$

ここで、もし  $\text{LogEne} > 0$  ならば、第2 スケーリング係数は次のようにして得られる。

$$\alpha_2(j) = \frac{\text{LogCorr}}{\text{LogEne}} \quad (\text{E-73})$$

もし  $\text{LogEne} = 0$  ならば、 $\alpha_2(j)$  は 1 に設定される。加えて、

if  $\alpha_2(j) < 0$   
 $\alpha_2(j) = 0$

End

のように  $\alpha_2(j)$  は強制的に正値とする。

第2 スケーリング係数は、ベクトル量子化を用いて量子化される。まず、ベクトルが次のように生成される。

$$V = [\alpha_2(0) \alpha_2(1) \alpha_2(2) \alpha_2(3)] \quad (\text{E-74})$$

4次元ベクトル  $V$  は、第1要素にて昇順に並べられた8ビットのコードブック(D\_4\_8)を用いて量子化される。

E.6.5.5 節に記載の手法と同様に、最近傍探索が行われる。

量子化の結果、量子化パラメータ  $\hat{\alpha}_2(j)$  が得られる。ここで、2つのスケーリング係数より生成される高

域成分  $\hat{M}_{32}(k)$  が計算される。

まず、選択フラグ  $F_m(k)$  が次式に従い定義される。

$$F_m(k) = \begin{cases} 1 & \text{if } ((P_m(j) - R_m(j) \leq k \leq P_m(j) + R_m(j)) \\ & \text{or } (k = 280, 282, \dots, 558)) \quad j = 0, \dots, 3 \\ 0 & \text{otherwise} \end{cases} \quad (\text{E-75})$$

ここで、 $P_m(j)$  は

$$P_m(j) = \begin{cases} 280 + k_{\max}^j & j = 0 \\ 280 + \sum_{i=0}^{j-1} d^i + k_{\max}^j & j = 1, \dots, 3 \end{cases} \quad (\text{E-76})$$

と定義される。 $R_m(j)$  は、あらかじめ定められた定数の配列である。

$$R_m(j) = \begin{cases} 5 & j = 0 \\ 15 & j = 1, 2 \\ 21 & j = 3 \end{cases} \quad (\text{E-77})$$

ここで、各サブバンドの選択フラグは、

$$\begin{aligned} F_m^0(k) &= F_m(k + 280), & k = 0, \dots, 39 \\ F_m^1(k) &= F_m(k + 320), & k = 0, \dots, 69 \\ F_m^2(k) &= F_m(k + 390), & k = 0, \dots, 69 \\ F_m^3(k) &= F_m(k + 460), & k = 0, \dots, 99 \end{aligned} \quad (\text{E-78})$$

と定義される。次に、選択フラグ  $F_m^j(k)$  が 1 となる暫定高域成分の奇数サンプルが、次のように対数領域に変換される。

$$\hat{M}_{32}(k) = 20 \cdot \log_{10}(|\hat{M}_{32}^j(k) + 1e-15|) \quad \text{if } (F_m^j(k) = 1), \quad j = 0, \dots, 3 \quad k = 1, 3, \dots, d^j - 1 \quad (\text{E-79})$$

$F_m^j(k)$  が 1 に等しいサンプルに対し、第 1 スケーリング後の MDCT 係数の符号が保存される。

$$\text{Sign}^j(k) = \begin{cases} 1, & \hat{M}_{32}^j(k) < 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{if } (F_m^j(k) = 1) \quad j = 0, \dots, 3 \quad k = 0, \dots, d^j - 1 \quad (\text{E-80})$$

次に、合成信号が次のように求められる。

$$\hat{M}_{32}^j(k) = \begin{cases} 10^{0.05 \cdot \{\hat{\alpha}_2(j) (|\hat{M}_{32}^j(k)|^{-m^j}) + m^j\}} & \text{if } (F_m^j(k) = 1 \text{ and } \text{Sign}^j(k) = 0) \\ -10^{0.05 \cdot \{\hat{\alpha}_2(j) (|\hat{M}_{32}^j(k)|^{-m^j}) + m^j\}} & \text{else if } (F_m^j(k) = 1 \text{ and } \text{Sign}^j(k) = 1) \\ \hat{M}_{32}^j(k) & \text{otherwise} \end{cases} \quad (\text{E-81})$$

最後に、4つのサブバンドで構成される全帯域の合成信号は次のように表される。

$$\ddot{M}_{32}(k) = \begin{cases} \ddot{M}_{32}^0(k-280), & k = 280, \dots, 319 \\ \ddot{M}_{32}^1(k-320), & k = 320, \dots, 389 \\ \ddot{M}_{32}^2(k-390), & k = 390, \dots, 459 \\ \ddot{M}_{32}^3(k-460), & k = 460, \dots, 559 \end{cases} \quad (\text{E-82})$$

#### E. 6. 5. 7 正弦波符号化による品質向上

2つのスケーリング係数を用いた上述した符号化は、正弦波成分の追加により更に改善される。ビット割当量は、これまでの第1の4 kbit/s SWB レイヤに2つの正弦波を追加することを可能にする。

すべての追加される正弦波の位置は、元の高域成分と合成高域成分の間の絶対値差に基づいて選択される。

$$D(k) = \left| \ddot{M}_{32}(k) - M_{32}(k) \right| \quad k = 280, \dots, 559 \quad (\text{E-83})$$

正弦波の位置の探索に用いられるアルゴリズムは、E.6.61 節に与えられる。2つの追加正弦波の探索トラックの開始位置は、合成高域信号のサブバンドエネルギーに基づいて選択される。合成されたサブバンドのエネルギーは、

$$SbE(k) = \sum_{n=0}^{n=31} \ddot{M}_{32}(k \times 32 + n)^2 \quad k = 0, \dots, 7 \quad (\text{E-84})$$

と得られる。ここで、 $k$  はサブバンドインデックスを表し、各サブバンドは32個のMDCT係数により構成される。最大エネルギーを有するサブバンドが、正弦波符号化の探索トラックとして選択される。この節でのトラックは、刻み幅が1の32個の位置より構成される。これはすなわち、サブバンドである。

2つの正弦波の振幅は、E.6.5.5 節で述べられた手法のように、それぞれ4ビット1次元のコードブックE\_1\_4によって量子化される。

#### E. 6. 6 正弦波モードの符号化

正弦波モードの符号化は、トータルに分類されたフレームに対して用いられる。正弦波モードにおいて、HF信号は、有限の正弦波成分をHFスペクトルに加えることにより生成される。正弦波の総数は10個であり、周波数範囲7000 – 8600 Hzの中に4個の正弦波、周波数範囲8600 – 10200 Hzの中に4個の正弦波、周波数範囲10200 – 11800 Hzの中に1個の正弦波、周波数範囲11800 – 12600 Hzに1個の正弦波が配置される。

##### E. 6. 6. 1 正弦波の位置

すべての追加される正弦波の位置は、元の高域成分と合成高域成分の間の絶対値差に基づいて選択される。

$$D(k) = \left| \ddot{M}_{32}(k) - M_{32}(k) \right| \quad k = 280, \dots, 559 \quad (\text{E-85})$$

レイヤ6moでは合成されたHF信号は0値の信号であり、また、正弦波は原信号の最大値の場所に配置される。この節の探索範囲は、さらに $k = 280, \dots, 503$ に制限される。他の符号化レイヤとモードでは、探索の上限値は異なる場合がある。

セグメント $D_j(k)$ は、差ベクトル $D(k)$ から生成される。ここで $j = 0, \dots, 5$ はトラックインデックスを表す。各セグメントに、あらかじめ定められた数 $N_j$ 個の正弦波が配置される。正弦波の位置は、ベクトル $D_j(k)$ の



内、最も大きな  $N_j$  個の値を見つけることにより選択される。レイヤ 6mo のこの符号化モードで使用されるセグメントもしくはトラック  $D_j(k)$  は、Figure E.3/JT-G729.1 に記載されている。第 0 および第 1 トラックはインターリーブされ、同様に、第 2 および第 3 トラックはインターリーブされる。Table E.6/JT-G729.1 は、32kHz の MDCT スペクトルに対するトラック開始位置、刻み幅および長さを示している。

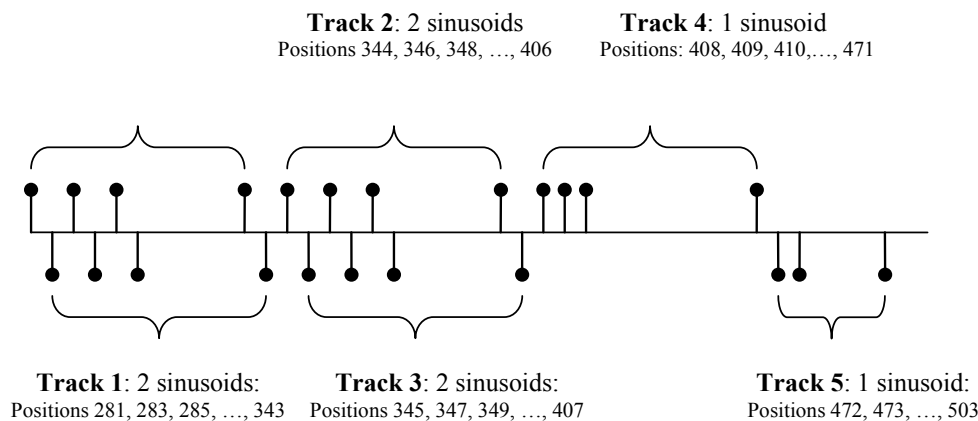


Figure E.3/JT-G729.1 – Sinusoid track structure in Layer 6mo illustrating the allowed sinusoid positions. Positions 280-503 correspond to frequencies 7000-12600 Hz

Table E.6/JT-G729.1 – Sinusoid track structure in Layer 6mo (sinusoidal mode)

Track	Num of sinusoids	Starting position	Position step size	Length
0	2	280	2	32
1	2	281	2	32
2	2	344	2	32
3	2	345	2	32
4	1	408	1	64
5	1	472	1	32

### E. 6. 6. 2 正弦波の量子化

上述したように、各トラック  $j$  の  $N_j$  個の正弦波が、それらの位置  $pos_j(l)$  ( $l=0, \dots, N_j$ ) を決定することによって選択される。そしてその位置は、対応するトラックの開始位置からの相対値として表される。

次に、正弦波の振幅  $c_j(l)$  は、

$$c_j(l) = \log\left(|D_j(pos_j(l))|\right) \quad (\text{E-86})$$

と得られる。式 E-85 に見られるように、符号情報は失われている。よって、正弦波の符号  $Sign\_sin_j(l)$  は別途保存され、量子化される。

$$Sign\_sin_j(l) = \begin{cases} 1 & D_j(pos_j(l)) \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (\text{E-87})$$

$N_j=2$  のとき、一方の符号(トラックの第 1 符号)のみそのトラック用に伝送される。他方の正弦波の符号情報は、2つの正弦波の相対位置に含めて符号化する。

```

if  $pos_j(0) < pos_j(1)$  AND  $Sign\_sin_j(0) \neq Sign\_sin_j(1)$  OR
    $pos_j(0) > pos_j(1)$  AND  $Sign\_sin_j(0) = Sign\_sin_j(1)$ 
    $pos\_tmp = pos_j(0)$ ,  $pos_j(0) = pos_j(1)$ ,  $pos_j(1) = pos\_tmp$ 
    $Sign\_tmp = Sign\_sin_j(0)$ ,  $Sign\_sin_j(0) = Sign\_sin_j(1)$ ,  $Sign\_sin_j(1) = Sign\_tmp$ 
    $c\_tmp = c_j(0)$ ,  $c_j(0) = c_j(1)$ ,  $c_j(1) = c\_tmp$ 
end

```

正弦波  $c_j(l)$  は、E.6.5.5 節に示された処理に従い、2、3 または 4 次元のベクトル量子化器を用いて量子化される。7 および 8 ビットの場合は同一のコードブックが用いられ、7 ビットコードブックは 8 ビットコードブックの後半部で構成される。8 ビットコードブックの前半部および後半部の各々は、第 1 成分に対して昇順に並べられている。最近傍探索が 8 ビットコードブックに対して行われる場合、E.6.5.5 節の処理が前半部および後半部に連続して適用され、入力に最も近いコードベクトルが 2 つの最近傍値から選択される。量子化された正弦波  $\hat{c}_j(l)$  は、拡張レイヤの新しい目標信号を見つけるために、HF 信号スペクトルに加えられる。

#### E. 6. 7 拡張レイヤ符号化手法の決定

第一の拡張レイヤ(レイヤ 7mo)において、下記両方に対して適応ビット配分が用いられる。

- JT-G729.1 のコア符号化器では符号化されない MDCT 係数のベクトル量子化によって 4-7kHz 周波数帯を MDCT 係数のベクトル量子化によって品質向上する。
- HF スペクトルにおける正弦波要素を符号化する。

レイヤ 7mo に対するビット割当ては 80 ビットである。このビット配分は、以下の手順にて行われる。

- JT-G729.1 のコア符号化器(6.6.8 節を参照)ではゼロである  $nbit(j)$  のビット配分に対して、4000-7000Hz のサブバンドには 9 ビットが配分される。すなわち、 $j=0, \dots, 9$ ; に対しては、 $nbit\_enhanced(j)=0$  であり、 $j=10, \dots, 17$  に対しては、

```

if  $nbit(j)=0$ 
    $nbit\_enhanced(j)=9$ 
else
    $nbit\_enhanced(j)=0$  for  $j=10, \dots, 17$ 
end

```

$$nbit\_enhanced\_tot = \sum_{j=10}^{17} nbit\_enhanced(j) \quad (E-88)$$

である。

- 残余ビットがある場合には、これらのビットは、HF スペクトルにおける正弦波成分の符号化に使われる。正弦波成分の数はコア符号化器の WB 拡張に対して用いられるビット数に依存する  $N_{sin}$  は、Table E.7/JT-G729.1 によるように、コア符号化器の WB 拡張に対して用いられるビット数に依存する。

Table E.7/JT-G729.1 – Number of sinusoidal components in Layer 7mo

<i>Nbit_enhanced_tot</i>	<i>Nsin</i>
0	10
1 ≤ ... ≤ 19	8
20 ≤ ... ≤ 30	6
31 ≤ ... ≤ 49	4
50 ≤ ... ≤ 60	2
61 ≤ ... ≤ 80	0

このビット配分に用いた手法は、残余ビットを用いて HF スペクトルを向上させるのと同時に、符号化器に対して異なる信号にも適用され、コア符号化器の品質を向上させる。4-7kHz 帯域に対する MDCT 係数は、E.6.9.6 節にて規定される Gosset 低演算量ベクトル量子化(GLCVQ)を用いて配分された *nbit\_enhanced(j)* に応じて量子化される。復号出力は、 $\hat{Y}_1(k)$  で与えられる。

### E. 6. 8 正弦波拡張レイヤの符号化

レイヤ 7mo と 8mo は、E.6.6 節で記述した正弦波モード符号化手法に基づいている。しかしながら、各レイヤの正確な機能は、第一 SWB 拡張レイヤ 6mo において用いられる符号化モードに依存する。下節にて、両レイヤを詳細に記述する。

#### E. 6. 8. 1 レイヤ 7mo

汎用モードフレームにおいて、HF スペクトルに加えられる正弦波成分の数 *Nsin* は、E.6.7 節にて定義したビット割り当てに基づき、0, 2, 4, 6, 8, 10 のいずれかに設定される。*Nsin* が 0 の場合、HF スペクトルには正弦波要素成分が加えられない。レイヤ 6mo からの合成 HF 要素成分のサブバンドエネルギーに基づき、正弦波符号化に対するトラックが選択される。

周波数帯 7000Hz-13400Hz の合成 HF 要素は、8 つのサブバンドに分割される。それぞれのサブバンドは、それぞれ 32 個の MDCT 係数で構成され、そのサブバンドエネルギーは、下式にて算出される。

$$SbE_{6mo}(k) = \sum_{n=0}^{n=31} \dot{M}_{32}^{6mo}(k \times 32 + n)^2 \quad k = 0, \dots, 7 \quad (\text{E-89})$$

ここで、 $\dot{M}_{32}^{6mo}(k)$  はレイヤ 6mo の後の合成 HF 信号である。正弦波符号化に対するトラックは、E.6.6.1 節に記述したソートアルゴリズムを用いて *Nsin/Nsin\_track* の最大サブバンドエネルギーを探索することで選択される。*Nsin\_track* は、トラックあたりの正弦波成分の数であり、ここでは 2 に設定される。サブバンド毎に選択されるそれぞれの *Nsin/Nsin\_track* は、正弦波符号化に用いられるトラックに直接対応する。例えば、*Nsin* が 4 の場合、最初の 2 つの正弦波は、最大サブバンドエネルギーのサブバンドに配置され、残りの 2 つの正弦波は、2 番目に大きいエネルギーのサブバンドに配置される。正弦波符号化に対するトラック位置は、利用可能なビット割り当てと HF 信号のエネルギー特性に依存し、フレーム毎に異なる。

正弦波モードフレームでは、10 個の正弦波要素が、HF 信号に以下のように付加される。最初の 4 つの正弦波は、2 つずつ 2 つのトラックにグループ化され、残りの 6 つの正弦波は、2 つずつ 3 つのトラックにグループ化される。最初の 4 つの正弦波は、9400Hz から 11000Hz の間に挿入され、残りの (final)6 つの正弦波は、11000Hz から 13400Hz の間に挿入される。Table E.8/JT-G729.1 にトラックの開始位置、ステップサイズ、長さをまとめる。

Table E.8/JT-G729.1 – Sinusoid track structure in Layer 7mo (sinusoidal mode frames)

Track	Num of sinusoids	Starting position	Position step size	Length
0	2	376	2	32
1	2	377	2	32
2	2	440	3	32
3	2	441	3	32
4	2	442	3	32

### E. 6. 8. 2 レイヤ8mo

レイヤ8moにおいて、さらに20個の正弦波が、2ステップでHF信号に付加される。そのトラック構造は、汎用モードや正弦波モードのフレームとは異なる。

汎用モードのフレームでは、正弦波符号化トラックの開始位置は、レイヤ7moの正弦波パルス数 $N_{sin}$ に依存する。 $N_{sin}$ が設定閾値より小さい場合は、正弦波パルスは、HF信号周波数帯の低域部分に配置される。 $N_{sin}$ が閾値以上の場合、多くの正弦波はHF信号周波数帯の高域部分に配置される。この閾値には、8を用いる。

最初のステップでは、以下のように10個の正弦波がHFスペクトルに付加される。はじめに、6つの正弦波が、周波数帯7000Hz-9400Hz、あるいは、9750Hz-12150Hzに2つずつ3つのトラックにグループ化される。次に、残りの4つの正弦波が、周波数帯9400Hz-11000Hz、あるいは、12150Hz-13750Hzに2つずつ2つのトラックにグループ化される。第2ステップでは、残りの10個の正弦波が、以下のように付加される。はじめに、6つの正弦波が、3周波数帯7800Hz-10200Hz、9400Hz-11800Hz、あるいは、8600Hz-11000Hzのいずれか1つの帯域に、2つずつ3つのトラックにグループ化される。最後の4つの正弦波が、3周波数帯10200Hz-11800Hz、11800Hz-13400Hz、あるいは、11000Hz-12600Hzのいずれか1つの帯域に、2つずつ2つのトラックにグループ化される。Table E.9/JT-G729.1に汎用モードフレームに対する、開始位置、ステップサイズ、トラック長をまとめる。各トラックは、2つの正弦波要素を含んでいる。

Table E.9/JT-G729.1 – Sinusoid track structure in Layer 8mo (generic mode frames)

$N_{sin}$	First starting position	Second starting position	Position step size	Length
0, 2	280	312	3	32
	376	408	2	32
4, 6	280	376	3	32
	376	472	2	32
8, 10	390	344	3	32
	486	440	2	32

正弦波モード符号化を用いるフレームでは、10個の正弦波の最初のセットが、以下のように付加される。はじめに、6つの正弦波が、周波数帯7000Hz-9400Hzに2つずつ3つのトラックにグループ化される。次に、残りの4つの正弦波が、周波数帯11000Hz-12600Hzに2つずつ2つのトラックにグループ化される。10個の正弦波の第2のセットは、以下のように付加される。はじめに、4つの正弦波が、周波数帯9400Hz-11000Hzに2つずつ2つのトラックにグループ化される。残りの6つの正弦波が、周波数帯11000Hz-13400Hzに、2つずつ3つのトラックにグループ化される。Table E.10/JT-G729.1及びTable E.11/JT-G729.1に正弦波モードフレームに対する、開始位置、ステップサイズ、トラック長をまとめる。

Table E.10/JT-G729.1 – Sinusoid track structure in Layer 8mo: 1st set (sinusoidal mode frames)

Track	Number of sinusoids	Starting position	Position step size	Length
0	2	280	3	32
1	2	281	3	32
2	2	282	3	32
3	2	440	2	32
4	2	441	2	32

Table E.11/JT-G729.1 – Sinusoid track structure in Layer 8mo: 2nd set (sinusoidal mode frames)

Track	Number of sinusoids	Starting position	Position step size	Length
0	2	376	2	32
1	2	377	2	32
2	2	440	3	32
3	2	441	3	32
4	2	442	3	32

### E. 6. 9 WB拡張

2つの拡張レイヤ9moと10moは、スペクトルのWB部分を品質向上させるように構成されている。本レイヤは、TDAC入力 $Y(k)$  (6.6.3節を参照)とJT-G729.1復号出力 $\hat{Y}(k)$  (7.3.5節を参照)のWB残差信号を符号化する。

$$Y_{err}(k) = Y(k) - \hat{Y}(k) - \hat{Y}_1(k) \quad k = 0, \dots, 319 \quad (\text{E-90})$$

ここで、 $\hat{Y}_1(k)$  は最初のWB拡張(E.6.7節を参照)の復号出力である。

#### E. 6. 9. 1 サブバンド分割

6.6.4節に記述したように、0-7000Hz帯のMDCT係数は、18のサブバンドに分割される。Table E.12/JT-G729.1にサブバンド境界と大きさを規定する。 $j$ 番目のサブバンドは $sb\_bound(j) \leq k < sb\_bound(j+1)$ の範囲の $nb\_coef(j)$ 個の係数の $Y_{err}(k)$ から構成されている。最初の17サブバンドは、16係数(400Hz)から構成され、最後のサブバンドは8係数(200Hz)から構成されている。

Table E.12/JT-G729.1 – Sub-band boundaries and number of coefficients per sub-band in the TDAC coder

$j$	$sb\_bound(j)$	$nb\_coef(j)$
0	0	16
1	16	16
2	32	16
3	48	16
4	64	16
5	80	16
6	96	16
7	112	16
8	128	16
9	144	16
10	160	16

Table E.12/JT-G729.1 – Sub-band boundaries and number of coefficients per sub-band in the TDAC coder

$j$	$sb\_bound(j)$	$nb\_coef(j)$
11	176	16
12	192	16
13	208	16
14	224	16
15	240	16
16	256	16
17	272	8
18	280	-

### E. 6. 9. 2 スペクトル包絡計算

残差信号  $Y_{err}(k)$  のスペクトル包絡は、18 サブバンドの対数領域における平均二乗根(rms)で定義される。

$$\log\_rms\_err(j) = \frac{1}{2} \log_2 \left[ \frac{1}{nb\_coef(j)} \sum_{k=sb\_bound(j)}^{sb\_bound(j+1)-1} Y_{err}(k)^2 + \epsilon_{rms} \right] \quad j=0,\dots,17 \quad (\text{E-91})$$

ここで

$$\epsilon_{rms} = 2^{-24}$$

である。

### E. 6. 9. 3 スペクトル包絡符号化

$Y_{err}(k)$  の 18 個の残差サブバンドのスペクトル包絡の符号化のために、JT-G729.1 のコア復号出力  $\hat{Y}(k)$  ( $\log\_rms(j)$ 、6.6.5 節を参照)のスペクトル包絡と、本ステップにて割り当てられるビット数( $nb\_bits(j)$ 、6.6.8 節を参照)から予測値が算出される。すなわち、

$$\log\_rms\_err\_predicted(j) = \log\_rms(j) - \frac{nb\_bits(j)}{nb\_coef(j)} \quad j=0\dots17 \quad (\text{E-92})$$

である。

残差信号  $Y_{err}(k)$  のスペクトル包絡から予測したスペクトル包絡が差し引かれる。この差分を  $delta\_log\_rms(j)$  と呼び、

$$delta\_log\_rms(j) = \log\_rms\_err(j) - \log\_rms\_err\_predicted(j), \quad j=0\dots17 \quad (\text{E-93})$$

と求められる。

この値は、 $delta\_log\_rms\_q(j)$  に丸められ、単一 2 ビット線形量子化器により量子化され、0、1、2、3 の値のいずれかを取る。すなわち、

$$\begin{aligned}
& \text{if}(\text{delta\_log\_rms}(j) < 0) \\
& \quad \text{delta\_log\_rms\_q}(j) = 0 \\
& \text{else if}(\text{delta\_log\_rms}(j) > 3) \\
& \quad \text{delta\_log\_rms\_q}(j) = 3 \\
& \text{else} \\
& \quad \text{delta\_log\_rms\_q}(j) = \text{round}(\text{delta\_log\_rms}(j))
\end{aligned} \tag{E-94}$$

である。

$Y_{err}(k)$  のスペクトル包絡の符号化には、36 ビットを必要とする。

#### E. 6. 9. 4 聴覚重要度によるサブバンドの順序付け

— 4-7kHz 周波数のサブバンドに対して、JT-G729.1 コアレイヤのスペクトル包絡からマスキング閾値が算出される。 $\hat{\sigma}^2$  と拡散関数  $B(\nu)$  を用い、

$$\text{mask}(j) = \sum_{k=10}^{17} \hat{\sigma}^2(k) \times B(\nu_j - \nu_k) \quad j = 10, \dots, 17 \tag{E-95}$$

となる。ここで、 $\hat{\sigma}^2(j) = \text{rms\_q}(j)^2 \times \text{nb\_coef}(j)$  であり、 $\text{rms\_q}(j)$  は 6.6.7 節で与えられる。 $\nu_j$  は、Bark スケールにおける  $j$  番目のサブバンドの中心周波数である。

— 聴覚重要度  $ip\_mask(j)$  は、以下のように上記マスク閾値を用いて修正される。

$$ip\_mask(j) = ip(j) - [\log_2(\text{mask}(j)) - \text{normfac}] \tag{E-96}$$

ここで、 $ip(j)$  は、いかなるマスキング効果を持たない信号に対する JT-G729.1 コアレイヤにて算出される聴覚重要度である。

$$ip(j) = \begin{cases} \frac{1}{2} \text{rms\_index}(j) & j = 0, \dots, 16 \\ \frac{1}{2} (\text{rms\_index}(j) - 1) & j = 17 \end{cases} \tag{E-97}$$

ここで、 $\text{rms\_index}(j) = \text{round}\left(\frac{1}{2} \log_{10} \text{rms}(j)\right)$  であり、 $\text{normfac}$  は、以下のように定義される正規化ファクタである。

$$\text{normfac} = \log_2 \left[ \sum_{j=9}^{17} \hat{\sigma}^2(j) \times B(\nu_9 - \nu_k) \right] \tag{E-98}$$

$\text{normfac}$  は、実際のところ、低周波数 3.6-4kHz の最後の周波数帯のマスク閾値である。この差分は、ビット配分に対する聴覚重要度計算において、低周波数(いかなるマスキング効果を持たない)と高周波数スペクトルの連続性を与える。

マスク閾値の計算における再帰性の理由から、拡散関数  $B(\nu)$  は、下記乗算として定義される。

$$\text{if } k > j, \quad B(\nu_j - \nu_k) = \prod_{i=0}^{k-j-1} \text{att}_1(16 - j - i)$$

$$\text{if } k < j, \quad B(v_j - v_k) = \prod_{i=0}^{j-k-1} att_2(j-1-i)$$

$$\text{if } k = j, \quad B(v_j - v_k) = 1$$

ここで、 $att_1(k)$  と  $att_2(k)$  は Table E.13/JT-G729.1 に定義される。例えば、

$$B(v_{11} - v_{17}) = att_1(5)att_1(4)att_1(3)att_1(2)att_1(1)att_1(0),$$

$$B(v_{12} - v_{10}) = att_2(11)att_2(10)$$

である。

Table E.13/JT-G729.1 – Spreading coefficients used for computing mask (Q15)

$k$	$att_1(k)$	$att_2(k)$
1	0.20669556	0
2	0.14553833	0
3	0.12698364	0
4	0.10147095	3.0518E-05
5	0.07800293	0.00024414
6	0.05715942	0.00100708
7	0.03939819	0.00308228
8	0.02514648	0.00732422
9	0.01452637	0.01452637
10	0.00732422	0.02514648
11	0.00308228	0.03939819
12	0.00100708	0.05715942
13	0.00024414	0.07800293
14	3.0518E-05	0.10147095
15	0	0.12698364
16	0	0.14553833
17	0	0.20669556

— 前出レイヤを考慮し、聴覚重要度は前出のビット配分と共に減じられる。すなわち、

$$ip\_mask(j) = ip\_mask(j) - nb\_bits\_enhanced(j) \quad (\text{E-99})$$

であり、ここで、 $nb\_bits\_enhanced(j)$  は、コアレイヤにおけるサブバンド  $j$  と最初の拡張レイヤ  $7mo$  を符号化するためのビット数である。

これらサブバンドは、聴覚重要度が減少する順に降順となるよう順位づけられる。その結果は、サブバンド  $j$  が  $(ord\_ip(j)+1)$  番目に大きい聴覚重要度であることを表す各サブバンドに対する  $0 \leq ord\_ip(j) < 18$   $j=0, \dots, 17$  の範囲のインデックスとなる。この順位は、ビット配分とベクトル量子化インデックスの多重化とに用いられる。

#### E. 6. 9. 5 ビット配分

各サブバンドに配分するビット数は、聴覚重要度を用いて決定される。これは、復号器でも算出できるため、補助情報なしに、復号器は同一処理が可能である。



Table E.14/JT-G729.1 に、ビット配分セットを示す。表中の次数 8 あるいは 16 は、各サブバンドにおける MDCT 係数の数に対応する。最大ビット配分は、サンプル当たり 2 ビットに制限される。

Table E.14/JT-G729.1 – Possible bit allocations for embedded spherical vector quantization

Dimension	Set of possible bit allocation (in bits)
8	$\mathbf{R}_8 = \{0, 7, 10, 12, 13, 14, 15, 16\}$
16	$\mathbf{R}_{16} = \{0, 9, 16, 21, 23, 26, 28, 30, 32\}$

総ビット数は、 $nbits\_err\_VQ = 284$  である。各サブバンドに配分されるビット数  $nbit\_err(j)$ ,  $j = 0, \dots, 17$  は、逆 waterfilling 原理による二分法探索アルゴリズムを用いて算出される。

二分法探索アルゴリズムは、以下のように water level  $\lambda_{opt}$  を算出する。

$$\begin{cases} nbit\_err(j) = \arg \min_{r \in \mathbf{R}_{nb\_coef(j)}} |nb\_coef(j) \times (ip\_mask(j) - \lambda_{opt}) - r| & j = 0, \dots, 17 \\ \sum_{j=0}^{17} nbit\_err(j) \approx nbits\_err\_VQ \end{cases} \quad (\text{E-100})$$

ここで、 $\mathbf{R}_{nb\_coef(j)}$  はビット配分を含んでいる。

$\lambda_{opt}$  に対する探索間隔は、以下の間隔となる。

$$\begin{cases} \lambda_0 = \max_{j=0, \dots, 17} (ip\_mask(j)) \\ \lambda_1 = \min_{j=0, \dots, 17} (ip\_mask(j)) - 4 \end{cases} \quad (\text{E-101})$$

ここで、 $\lambda_0$  はゼロビット配分に対応し、 $\lambda_1$  は聴感上もっとも重要でないサブバンドに対してサンプル当たり 4 ビットの配分に対応する。10 回の反復計算後、ビット配分は下式によって算出される。

$$nbit\_err(j) = \arg \min_{r \in \mathbf{R}_{nb\_coef(j)}} |nb\_coef(j) \times (ip\_mask(j) - \lambda_{opt}) - r| \quad (\text{E-102})$$

総ビット数は、(適切に初期化された探索間隔により)ビット割当量を超過しない。しかしながら、この手法はビット割当量を全て使い切らない可能性がある。この場合、残りのビット割当量は、聴覚重要度に対して降順となるように低い順に各サブバンドに振り分けられる(この方法は  $ord\_ip(j)$  に基づく)。

#### E. 6. 9. 6 欠落バンドと WB 誤差に対する量子化

$j$  番目のサブバンド次数が 8 の場合、 $nbit(j)$  ビットのエンベディッド球状ベクトル量子化(SVQ)で量子化される(6.6.9 節を参照のこと)。  $j$  番目のサブバンド次数が 16 の場合、 $nbit(j)$  ビットの Gosset 低演算量ベクトル量子化(GLCVQ)で量子化される。ここで、 $nbit(j)$  は、符号化レイヤに依存し、 $nbit\_enhanced(j)$  ないしは  $nbit\_err(j)$  で決定される。

##### E. 6. 9. 6. 1 球状ベクトル量子化

6.6.9 節に準ずる。

##### E. 6. 9. 6. 2 Gosset低演算量ベクトル量子化

Gosset 低演算量ベクトル量子化(GLCVQ)は、量子化サンプルで信号サンプルを表現する手法である。

GLCVQにおいて、信号サンプルはE格子[b-Gosset]ポイントに対応する信号に量子化される。量子化サンプルは、クラスリーダールートベクトルの代表値とその符号分布によって特定される。この時のE格子次数は、16である。

### E. 6. 9. 6. 2. 1 GLCVQ コードブック

次数が 16 の場合、 $\mathbf{Q}_r^{16}$ ,  $r \in \mathbf{R}_{16}$  and  $r > 0$  と記述される 8 個のコードブックがある。 $\mathbf{R}_{16}$  は、GLCVQ に対するビット配分(Table E.14/JT-G729.1 で規定)である。各コードブックは、異なるクラスリーダールートベクトルを有する。

効率化のため、コードブックは以下の特徴を有する。

- コードベクトルは、符号分布を含むクラスリーダールートベクトルの順列としての代表値で表される。
- クラスリーダールートベクトルは、重みベクトルと振幅ベクトルから構成される。
- 重みベクトルと振幅ベクトルは、特徴量ベクトルセットとマッピングアドレスの代表値によって表される。

クラスリーダールートベクトル

E 格子は、Gosset 格子から任意次数  $L_v$  への一般化として表される。

$$E_{L_v} = D_{L_v} \cup (D_{L_v} + \mathbf{v}); \mathbf{v} = [\frac{1}{2} \cdots \frac{1}{2}] \quad (\text{E-103})$$

ここで、

$$D_{L_v} : y \in \left\{ [y_0 \ y_1 \ \cdots \ y_{L_v-1}] : y_j \in \mathbb{Z}, \sum_{j=0}^{L_v-1} y_j \equiv 0 \pmod{2} \right\} \quad (\text{E-104})$$

は E 格子と同一次数  $L_v$  のチェッカーボード格子、あるいは D 格子である。原点に対して同一距離の格子ベクトルは、その格子のシェル  $S$  を定義する。一般化 Gosset 格子、あるいは E 格子  $E_{L_v}$  に対して、原点への平方距離が  $2\mu$  のシェルは、 $S_\mu^{(E_{L_v})}$  と表記する。E 格子シェルの半径を、 $R_\mu^{(E_{L_v})}$  と表記する。各 E 格子球状ベクトルコードブックは、原点から距離  $u$  の格子ポイントで、平方半径が  $2u$  のシェルを半径 1 へ正規化し、このシェルに関連する全てのベクトルから構成される。

$$\tilde{\chi}_\mu^{(E_{L_v})} : \tilde{c} = \frac{y}{\sqrt{2\mu}} \quad \forall y \in S_\mu^{(E_{L_v})} \quad (\text{E-105})$$

低演算量 Gosset SVQ は、Gosset 格子あるいは E 格子における、以下の特徴を十分に引き出している。Gosset 格子球状コードブックの全てのコードベクトルは、カテゴリ A と B に分類される。カテゴリ A のコードベクトルは、 $D_{L_v}$  に属するコードベクトルである。カテゴリ B のコードベクトルは、 $D_{L_v}$  格子のシフト版によって規定される条件に属するコードベクトルである。

正しいカテゴリ A の球状コードベクトル  $\tilde{c}^{(A)}$  を考えると、座標位置  $j_0$  の符号が変わる時、その結果として得られるベクトル  $\tilde{c}_A^*$  は、正しいベクトルで有り続ける。なぜならば、

$$\sum_{j=0}^{L_v-1} \tilde{c}_j^{(A)} - 2 \cdot \tilde{c}_{j_0}^{(A)} \equiv 0 \pmod{2} \quad (\text{E-106})$$

であり、ベクトル  $\tilde{c}_A^*$  のそれぞれの要素の絶対値は変化しない。例えば、 $\{2,1,1,0,0,0,0,0,0,0,0,0,0,0,0\}$  は Gosset 格子球状コードブック  $\mathbf{Q}_{21}^{16}$  における  $\mu=3$  でのコードベクトルであり、座標値  $j_0=0$  の符号が変わる場合、 $\sum_{j=0}^{15} \tilde{c}_j^{(A)} - 2 \cdot \tilde{c}_{j_0}^{(A)} = 4 - 2 \cdot 2 = 0 \equiv 0 \pmod{2}$  を満足し、その結果として得られるベクトル  $\tilde{c}_A^*$  は、 $\{1,1,0,0,0,0,0,0,0,0,0,0,0,0,-2\}$  である。

同様に、正しいカテゴリ B の球状コードベクトル  $\tilde{c}^{(B)}$  を考えると、 $j_0$  と  $j_1$  の二つの符号が変わる場合、下式を満足するのでベクトル  $\tilde{c}_B^*$  は正しいベクトルである。

$$\sum_{j=0}^{L_v-1} \tilde{c}_j^{(B)} - 2 \cdot \tilde{c}_{j_0}^{(B)} - 2 \cdot \tilde{c}_{j_1}^{(B)} \equiv \sum_{j=0}^{L_v-1} \tilde{c}_j^{(B)} \pmod{2} \quad (\text{E-107})$$

ここで、また、ベクトル  $\tilde{c}_B^*$  のそれぞれの要素の絶対値は変化しない。例えば、 $\{3/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, -1/2\}$  は Gosset 格子球状コードブック  $\mathbf{Q}_{21}^{16}$  と  $\sum_{j=0}^{15} \tilde{c}_j^{(B)} = 8$  における  $\mu=3$  でのコードベクトルであり、座標値  $j_0=0$ 、 $j_1=0$ 、 $\sum_{j=0}^{L_v-1} \tilde{c}_j^{(B)} - 2 \cdot \tilde{c}_{j_0}^{(B)} - 2 \cdot \tilde{c}_{j_1}^{(B)} = 8 - 2 \cdot \frac{1}{2} - 2 \cdot \frac{1}{2} = 6$  の符号が変わる場合、 $\sum_{j=0}^{L_v-1} \tilde{c}_j^{(B)} - 2 \cdot \tilde{c}_{j_0}^{(B)} - 2 \cdot \tilde{c}_{j_1}^{(B)} \pmod{2} = \sum_{j=0}^{L_v-1} \tilde{c}_j^{(B)} \pmod{2} = 0$  を満足し、その結果としてのベクトル  $\tilde{c}_B^*$  は、 $\{3/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, 1/2, -1/2, -1/2, -1/2\}$  である。さらに、符号パリティの制限が、ベクトル  $\tilde{c} = [\tilde{c}_0 \tilde{c}_1 \cdots \tilde{c}_{L_v-1}]$  に対して、以下のように導入される。

$$\text{parity}(\tilde{c}) = \left( \sum_{j=0}^{L_v-1} \text{sign}(\tilde{c}_j) \right) \pmod{2} \quad (\text{E-108})$$

これは、以下のように偶数、奇数に分けられる。

$$\text{sign}(\tilde{c}_j) = \begin{cases} 0 & \text{if } \text{sgn}(\tilde{c}_j) = 1 \\ 1 & \text{if } \text{sgn}(\tilde{c}_j) = -1 \end{cases} \quad (\text{E-109})$$

以上から、Gosset 格子に対する等価クラスは、 $0 \leq m' < M_F^{(E_{L_v})}$  であるところの、ひとつのクラスリーダールートベクトル  $\tilde{c}_{F, P_{m'}}$  を有する  $M_F^{(E_{L_v})}$  ルートクラス  $F_{m'}$  に分類することができる。クラスリーダールートベクトルから、全てのクラスリーダーベクトルは符号置換によって構成される。

クラスリーダールートベクトルは、 $\mu_l \in Z$  の異なる値  $L_F$  から構成される。ここで、それぞれは  $w_l$  回発生し、 $w_l$  は  $\mu_l$  の重みに影響される。

$$\begin{aligned} \tilde{c}_{F, P_{m'}} &= [\tilde{c}_{F, P_{m'}, 0} \quad \tilde{c}_{F, P_{m'}, 0} \quad \cdots \quad \tilde{c}_{F, P_{m'}, L_v-1}]^T \\ &= \left[ \begin{array}{ccccccc} \leftarrow w_0 \rightarrow & \leftarrow w_1 \rightarrow & & \leftarrow w_{L_F-1} \rightarrow & & & \\ \mu_0 & \mu_0 & \mu_1 & \mu_1 & \cdots & \mu_{L_F-1} & \mu_{L_F-1} \end{array} \right]^T \end{aligned} \quad (\text{E-110})$$

ここで、全てのクラスリーダールートベクトルに対して、 $\mu_l \geq 0$  かつ  $\mu_0 > \mu_1 > \cdots > \mu_{L_F-1}$  である。

$m'$  インデックスを持つカテゴリ A のクラスリーダールートベクトルを考えた時、構成されるクラスリーダーベクトルの数は、 $N_m^{(F)}$  である。これは、非ゼロ座標値  $\mu_l \neq 0$  の重み  $w_l$  と関係する符号置換可能な数として決定される。

$$N_{m'}^{(F)} = \prod_{l=0; \mu_l \neq 0}^{L_F-1} (w_l + 1) \quad (\text{E-111})$$

カテゴリ B に対して、符号群は一对毎に置換されるので、符号のパリティ制限が満足される。カテゴリ B のクラスリーダールートベクトルを考えた時、偶数または奇数の符号制限のものと符号置換数であり、かつカテゴリ B のクラスリーダールートベクトルに関連するクラスリーダーベクトル偶数ないしは、奇数の符号制限のものと符号置換数は、下式で与えられる。

$$N_{m'}^{(F)} = \begin{cases} \frac{\left( \prod_{l=0; \mu_l \neq 0}^{L_F-1} (w_l + 1) \right) + 1}{2} & \text{for even sign parity} \\ \frac{\left( \prod_{l=0; \mu_l \neq 0}^{L_F-1} (w_l + 1) \right) - 1}{2} & \text{for odd sign parity} \end{cases} \quad (\text{E-112})$$

例えば、クラスリーダールートベクトルと対応する  $\mathbf{Q}_{21}^{16}$  コードブックにおけるクラスリーダーベクトルを、Table E.15/JT-G729.1 に示す。全てのクラスリーダールートベクトルと対応するクラスリーダーベクトルは、各要素が整数であることを保証するために 2 単位でスケールされる。

Table E.15/JT-G729.1 – Class-leader root vectors and corresponding class-leader vector

Class-leader root vectors	Class-leader vectors
{4,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0} (category A)	{4,2,2,0,0,0,0,0,0,0,0,0,0,0,0,0} {4,2,0,0,0,0,0,0,0,0,0,0,0,0,0,-2} {4,0,0,0,0,0,0,0,0,0,0,0,0,0,-2,-2} {2,2,0,0,0,0,0,0,0,0,0,0,0,0,-4} {2,0,0,0,0,0,0,0,0,0,0,0,0,0,-2,-4} {0,0,0,0,0,0,0,0,0,0,0,0,0,-2,-2,-4}
{2,2,2,2,2,2,0,0,0,0,0,0,0,0,0,0} (category A)	{2,2,2,2,2,2,0,0,0,0,0,0,0,0,0,0} {2,2,2,2,0,0,0,0,0,0,0,0,0,0,-2} {2,2,2,2,0,0,0,0,0,0,0,0,0,-2,-2} {2,2,2,0,0,0,0,0,0,0,0,0,0,-2,-2,-2} {2,2,0,0,0,0,0,0,0,0,0,0,-2,-2,-2,-2} {2,0,0,0,0,0,0,0,0,0,-2,-2,-2,-2,-2} {0,0,0,0,0,0,0,0,0,-2,-2,-2,-2,-2,-2}

Table E.15/JT-G729.1 – Class-leader root vectors and corresponding class-leader vector

Class-leader root vectors	Class-leader vectors
{3,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1} (category B, odd parity)	{3,1,1,1,1,1,1,1,1,1,1,1,1,1,1,-1} {3,1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-1} {3,1,1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1} {3,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1} {3,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1} {3,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1} {3,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1} {3,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1} {1,1,1,1,1,1,1,1,1,1,1,1,1,1,-3} {1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-3} {1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-3} {1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-3} {1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-3} {1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-3} {1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-3} {1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-3}

クラスリーダーベクトルから、全てのコードベクトルは、位置置換によって構成される。

#### コードブック生成

コードブック用メモリを削減するため、特徴ベクトルセットと、そのマッピングアドレスが、クラスリーダールートベクトルの代わりに格納される。そして、GLCVQ に対するコードブックは、以下のように生成される。

- 特徴ベクトルセットの格納  
特徴ベクトルセットには、相互に異なる振幅ベクトルと GLCVQ コードブックのクラスリーダールートベクトルに対応する重みベクトルを含んでいる。  
全てのクラスリーダールートベクトルは、振幅ベクトルと重みベクトルから構成される。その振幅ベクトルは、クラスリーダールートベクトルの異なる非ゼロ値から構成される。そして、重みベクトルは、振幅ベクトル内における非ゼロの異なる値の数に対応する重みから構成される。
- 特徴ベクトルセットに対するクラスリーダールートベクトルに対応する振幅ベクトルと重みベクトルのマッピングアドレスを格納する。
- 特徴ベクトルセットとマッピングアドレスに対する GLCVQ コードブックを生成する。

例えば、クラスリーダールートベクトルが{4,2,2,0,0,0,0,0,0,0,0,0,0,0,0}の時、振幅ベクトルは{4,2}で、重みベクトルは{1,2}である。別の例として、クラスリーダールートベクトルが{4,2,2,2,2,0,0,0,0,0,0,0,0,0,0}の時、振幅ベクトルは{4,2}で、重みベクトルは{1,4}である。コードブックに対するメモリを削減するため、特徴ベクトルセット{ {4,2}, {1,2}, {1,4} }が格納され、振幅ベクトルと重みベクトルのマッピングアドレスは、{0,0}と{1,2}としてそれぞれ格納される。特徴ベクトルセットとマッピングアドレスに従って、全てのクラスリーダールートベクトルが構成される。その時、全てのクラスリーダーベクトルは、符号分布に応じて構成され、全てのコードベクトルは、位置置換のクラスリーダーベクトルから構成される。

#### E. 6. 9. 6. 2. 2 量子化手法

量子化手法の演算処理量は、全てのクラスリーダーベクトルの代わりにクラスリーダールートベクトルのみ GLCVQ で探索し、全てのクラスリーダーベクトルで代替するので、量子化手法の演算処理量は大幅に削

減される。各量子化信号は、符号分布を含み、決定されるクラスリーダールートベクトルの代表値を有している。

- ベクトル  $x = [x_0 \ x_1 \ \dots \ x_{L_v-1}]$  を量子化するために、はじめに、半径 1 のベクトル  $c = [c_0 \ c_1 \ \dots \ c_{L_v-1}]$  に正規化される。

$$c = \frac{x}{\|x\|} \quad (\text{E-113})$$

正規化ベクトルの各要素は、振幅ベクトルと符号ベクトルとそれぞれ相互に変換可能な振幅と符号に分割される。

$$c_{mag} = [|c_0| \ |c_1| \ \dots \ |c_{L_v-1}|]^T \quad (\text{E-114})$$

$$c_{sgn} = [\text{sgn}(c_0) \ \text{sgn}(c_1) \ \dots \ \text{sgn}(c_{L_v-1})]^T \quad (\text{E-115})$$

- 振幅ベクトルは、降順に要素を並べるために、置換行列  $P_c$  によって置換される。

$$c_{sgn}^* = P_c \cdot c_{sgn} \quad (\text{E-116})$$

また、以下のように符号ベクトルが置換される。

$$c_{sgn}^* = P_c \cdot c_{sgn} \quad (\text{E-117})$$

ここで、 $P_c$  置換行列である。

- 次に、全てのクラスリーダールートベクトルに対して、ベクトル  $c_{mag}^*$  に対する距離が、二乗誤差メトリックとして計算される。

$$D_{m'} = \|c_{mag}^* - \tilde{c}_{F,P_m}\|^2 \quad (\text{E-118})$$

- ・ カテゴリ **A** のクラスリーダーベクトルを考えた時、算出される二乗誤差メトリックは、全ての対応するクラスリーダーベクトルの中で最小誤差となる。これは、全ての対応するクラスリーダーベクトルの置換された入力ベクトルの符号が、クラスリーダーベクトルのそれと同一である場合に限る。

$$D_{m'}^{(A)} = D_{m'} \quad (\text{E-119})$$

- ・ クラスリーダールートベクトルが、最小メトリックの場合、最小誤差となるクラスリーダーベクトルは、ベクトル  $c_{sgn}^*$  の符号歪分布によって単独で決定される。

- ・ カテゴリ B のクラスリーダールートベクトル  $\tilde{c}_{F, P_m}^{(B)}$  を考えた時、誤差メトリックは、符号が除かれた入力ベクトルに基づき、式 E-118 に従い算出される。この場合、全ての符号組合せが可能なのわけではない。したがって、最小誤差  $D_m^{(B)}$  を算出するために、二つのケースを区分しなければならない。

Case 1:

$$\text{parity}(c_{\text{sgn}}^*) = \text{parity}(\tilde{c}_{F, P_m}^{(B)}) \quad (\text{E-120})$$

入力ベクトルの符号パリティが、クラスリーダールートベクトルの符号パリティと同一である場合、算出される誤差は、全てのそれぞれの対応するクラスリーダーベクトルによって最小値に到達できる。

$$D_m^{(B)} = D_m \quad (\text{E-121})$$

- ・ クラスリーダールートベクトルが、最小メトリックの場合、最小誤差となるクラスリーダーベクトルは、ベクトル  $c_{\text{sgn}}^*$  の符号歪分布によって単独で決定される。

Case 2

$$\text{parity}(c_{\text{sgn}}^*) \neq \text{parity}(\tilde{c}_{F, P_m}^{(B)}) \quad (\text{E-122})$$

- ・ 入力ベクトルの符号パリティが、クラスリーダールートベクトルの符号パリティと異なる場合、クラスリーダールートベクトルの未知インデックス  $j_0$  のある座標の符号は、 $c_{\text{sgn}}^*$  の符号と異なる。訂正後の誤差は、以下のようになる。

$$\begin{aligned} D_m^{(B)} &= D_m - \left( c_{\text{mag} \square j_0}^* - \tilde{c}_{F, P_m, j_0}^{(B)} \right)^2 + \left( c_{\text{mag} \square j_0}^* + \tilde{c}_{F, P_m, j_0}^{(B)} \right)^2 \\ &= D_m + 4 \cdot c_{\text{mag} \square j_0}^* \cdot \tilde{c}_{F, P_m, j_0}^{(B)} \end{aligned} \quad (\text{E-123})$$

最小誤差とするために、未知インデックス  $j_0$  は、以下のようになる。

$$j_0 = \underset{0 \leq j < L_v}{\text{argmin}} c_{\text{mag}, j}^* \cdot \tilde{c}_{F, P_m, j}^{(B)} \quad (\text{E-124})$$

置換後符号ベクトル  $c_{\text{sgn}}^*$  は、位置  $j_0$  で置換後符号ベクトル要素が 1(モジュロ 2)ずつ増加するように訂正される。そして、case 1 と同じように、最小誤差となるクラスリーダーベクトルを決定するために使われる。

全ての  $M_F^{(E_{L_v})}$  クラス中、最小メトリック値のものは、以下に示す量子化手順で最終的に抽出される。

$$m'_Q = \operatorname{argmin}_{0 \leq m' < M_F^{(E_{L_v})}} D_{m'}^{(A/B)} \quad (\text{E-125})$$

### E. 6. 9. 6. 2. 3 コードベクトルのインデックスへのマッピング

探索後の最適化球状コードベクトルのインデックスをベクトル再生のために復号器へ伝送するインデックスへ変換するため、置換行列  $P_c$ 、符号ベクトル  $c_{sgn}^*$  とルートクラス  $m'_Q$  を考えた時、以下の3ステップが実行される。

- Step 1:* ルートクラス  $m'_Q$  の予備選択と対応する等価クラス。ルートクラス  $m'_Q$  の選択に従い、クラスリーダールートベクトルが構成される。
- Step 2:* 符号ベクトル  $c_{sgn}^*$  に基づき、等価クラスインデックス  $m'_Q$  と全ての対応する等価クラス  $\varepsilon_{m_Q}$ 。インデックスオフセット  $i_{off, \varepsilon_{m_Q}}$  を読み、クラスリーダーベクトル選択が構成される。
- Step 3:*  $P_c$  行列に関する位置置換のインデックス  $i_p$  を計算する。

最終的な量子化インデックスは下式となる。

$$i_Q = i_p + i_{off, \varepsilon_{m_Q}} \quad (\text{E-126})$$

演算量削減のために、高速インデックス符号化手法が、探索後最適化コードベクトルの位置インデックス  $i_p$  を符号化するために適用される。

#### 高速インデックス符号化法

探索後最適化コードベクトルは、高速インデックス符号化法の入力ベクトルである。最適コードベクトルは、選択クラスリーダーベクトルの置換として決定される。選択クラスリーダーベクトルに対応する振幅ベクトルは、 $[u_0 u_1 \dots u_{L_p-1}]$  として定義される。ここで、 $L_p$  は振幅ベクトルの異なる値の数である。重みベクトルは、 $[w_0 w_1 \dots w_{L_p-1}]$  として定義され、最適化コードベクトルは、 $\tilde{c} = [c_0 c_1 \dots c_{L_p-1}]$  として定義される。

- 1) 最適化コードベクトルの振幅ベクトルと重みベクトルを取り出す。選択クラスリーダーベクトルの置換として、最適化コードベクトルが決定されるので、選択コードベクトルの振幅ベクトルと重みベクトルは、選択クラスリーダーベクトルの振幅ベクトルと重みベクトルと等価である。  
例えば、最適コードベクトルが  $\tilde{c} = [00020-20000000000]$  の場合、その振幅ベクトルは  $[u_0 u_1 \dots u_{L_p-1}] = [2 \ 0 \ -2]$  であり、重みベクトルは  $[w_0 w_1 \dots w_{L_p-1}] = [1 \ 14 \ 1]$   $L_p = 3$  である。
- 2) 降順に重みベクトルを並べ替える。  $[w'_0 w'_1 \dots w'_{L_p-1}] \quad w'_0 \geq w'_1 \geq \dots \geq w'_{L_p-1}$   
例えば、並べ替えた重みベクトルが  $[w'_0 w'_1 \dots w'_{L_p-1}] = [14 \ 1 \ 1]$  の場合、その振幅ベクトルは  $[u'_0 u'_1 \dots u'_{L_p-1}] = [0 \ 2 \ -2]$  である。
- 3) 並べ替えた振幅ベクトルと重みベクトルに従い、探索後最適コードベクトルの位置インデックスを得る。要素移動に対する最適順序に基づき、並べ替えた振幅ベクトルと重みベクトルに従い、探索後最適コードベクトルにおける各エレメント位置に対する位置置換符号化を用いて、位置インデックスを得る。本方法は、以下の手順で動作する。



- Step A: 最適コードベクトルは、並べ替えた振幅ベクトルの要素に従い、 $L_p$  レベルに再生される。最上位レベルベクトルは、元の最適コードベクトルである。
- Step B:  $n$  でインクリメントする。レベル  $n$  ( $0 < n < L_p$ ) で、新ベクトルは、上位レベルベクトルで最も頻繁に使用される要素  $\mu'_{n-1}$  を取り除くことによって得られる。残りの要素は、現行レベルで、新ベクトルに組み込まれる。すなわち、 $n-1$  レベルベクトルの位置順である。残要素の全ての位置値は、位置ベクトル生成に使われる。  
例えば、最上位レベルベクトルが、元の最適コードベクトル [00020-2000000000] の場合、レベル 1 において、レベル 0 ベクトルから  $\mu'_{n-1} = \mu'_0 = 0$  (元の最適コードベクトル) を取り除き、レベル 1 に対する新ベクトルは [2 -2] で、位置ベクトルは [3 5] となる。レベル 2 において、レベル 1 ベクトルから  $\mu'_{n-1} = \mu'_1 = 2$  を取り除き、レベル 2 に対する新ベクトルは [-2] で、位置ベクトルは [1] となる。
- Step C: 上位レベル(level  $n-1$ )ベクトルに関連する現レベル(level  $n$ )ベクトルの位置ベクトルは、置換・組合せ機能に基づきインデックスが付与される。インデックスの結果を、 $mid\_index_n$  と呼ぶ。現レベルにおける新ベクトルに対して、その位置ベクトルインデックスは、以下のように計算される。

$$mid\_index_n = C_{m_{n-1}}^{m_n} - C_{m_{n-1}-p_0}^{m_n} + \sum_{i=1}^{i < m_n} (C_{m_{n-1}-p_{i-1}-1}^{m_n-i} - C_{m_{n-1}-p_i}^{m_n-i}) \quad (E-127)$$

$$final\_index = final\_index * C_{m_{n-1}}^{m_n} + mid\_index_n \quad (E-128)$$

$final\_index$  は、本手順の初め、すなわち step A の前に 0 に初期化される。要素  $p_0, p_1, p_2, \dots$  は、これら  $n$  レベルにおける左から右への位置ベクトル要素値である。 $m_{n-1}$  は、上位レベル(level  $n-1$ )ベクトルの次元であり、 $m_n$  は、現レベル(level  $n$ )ベクトルの次元であり、 $C_p^m$  は置換・組合せ式  $C_p^m = \frac{p!}{m!(p-m)!}$  を表す。ここで、 $p, m = \{1, \dots, 16\}$ , および  $and p > m$  である。 $C_p^m$  に対する値の全ては、階乗計算を避けるため 1 つのテーブルに格納される。 $n-1$  レベルの最終インデックスは、現レベルで取り得るインデックス値の数  $C_{m_{n-1}}^{m_n}$  を乗じ、現レベルのインデックス  $mid\_index_n$  へ加え、現レベルの最終インデックスを得る。

- Step D: ここまで、ステップ B と C を繰り返すと、現新ベクトルにおいて、たった 1 つの要素タイプのみが残る。最下位レベルに対する  $final\_index$  は、最適ベクトルの位置インデックスである。  
例えば、レベル 1 において、上位レベル(level 0)ベクトルの次元は、16 である。すなわち、 $m_{n-1} = 16$  である。現行レベル(level 1)の新ベクトルの次元が 2 の場合は、 $m_n = 2$  である。

$$\begin{aligned} mid\_index_1 &= C_{m_{n-1}}^{m_n} - C_{m_{n-1}-p_0}^{m_n} + \sum_{i=1}^{i < m_n} (C_{m_{n-1}-p_{i-1}-1}^{m_n-i} - C_{m_{n-1}-p_i}^{m_n-i}) \\ &= C_{16}^2 - C_{16-p_0}^2 + \sum_{i=1}^{i < 2} (C_{16-p_{i-1}-1}^{2-i} - C_{16-p_i}^{2-i}) \\ &= C_{16}^2 - C_{16-p_0}^2 + C_{16-p_0-1}^1 - C_{16-p_1}^1 \\ &= C_{16}^2 - C_{16-3}^2 + C_{16-3-1}^1 - C_{16-5}^1 \\ &= 43 \end{aligned}$$

$$\text{final\_index} = \text{final\_index} * C_{16}^2 + \text{mid\_index}_1 = 0 * C_{16}^2 + 43 = 43$$

レベル 2 において、上位レベル(level 1)ベクトルの次元は 2 であり、 $m_{n-1} = 2$  である。現レベル(level 2)の新ベクトルは 1 であり、 $m_n = 1$  である。

$$\text{mid\_index}_2 = C_2^1 - C_{2-p_0}^1 = C_2^1 - C_{2-1}^1 = 2 - 1 = 1$$

$$\text{final\_index} = \text{final\_index} * C_2^1 + \text{mid\_index}_2 = 43 * C_2^1 + 1 = 87$$

そして、現新ベクトルにおいてたった 1 つの要素タイプが残り、最適コードベクトルの位置インデックスは 87 となる。

## E. 7 復号器の機能記述

JT-G729.1 SWB 拡張復号器の構造を、Figure E.4/JT-G729.1 に示す。JT-G729.1 コアコーデックは 16kHz 信号を復号し、SWB 拡張は 32kHz 出力を提供するために高い周波数を復号する。SWB 復号は、ほとんど MDCT 領域で実行される。最初に復号されるトーナリティの指標に従って、二つの選ぶべきモード、すなわち汎用モードと正弦波モードが、最初のレイヤの復号に使われる。WB 品質向上と付加正弦波とにビットを配分するために、第 2 のレイヤは、符号化器と同じビット割り当てを用いる。第 3 の SWB レイヤは、高い周波数成分の品質を向上させる付加正弦波モード符号化からなる。第 4、第 5 の拡張レイヤは、ワイドバンド信号を改善する。時間領域における合成 SWB 信号を改善するために、後処理が用いられる。

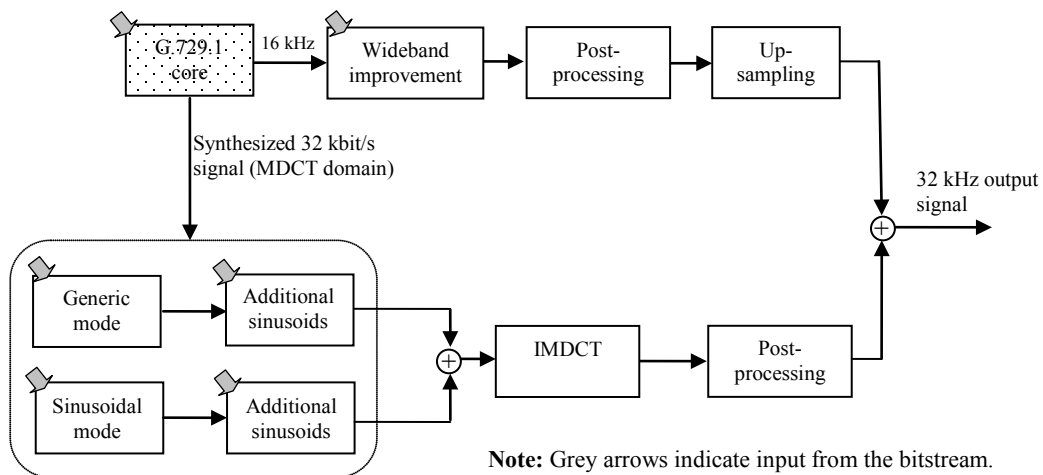


Figure E.4/JT-G729.1 – Structural block diagram of the decoder

### E. 7. 1 WB復号

HF 復号の開始前に、WB 信号の合成が必要である。これは、JT-G729.1 復号器の機能記述に従って実行される。デフォルトの工程において、共通後処理機能の適用に先立つ 32kbit/s WB 合成が、HF 信号の復号に用いられる。

#### E. 7. 1. 1 4-7 kHzにおけるゼロビットサブバンドの充填

JT-G729.1 のコア符号化において、4-7 kHz の周波数のスペクトル包絡は、ビットレートが 16kbit/s 以上の場合に符号化される。しかし、時々この周波数範囲におけるエネルギーの小さいサブバンドのスペクトル微細構造が符号化されず(例えば、全てのバンドを符号化するための利用可能な符号化ビット量が十分でない場合、もしくは復号ビットレートが低い場合など)、復号器において、利用可能な情報を用いて生成されなければなら

らない。JT-G729.1のコアにおいて、これらの欠落した周波数バンドの表現は、TDBWEスキームを用いてなされる。JT-G729.1 SWB復号器における符号化されないサブバンドのスペクトル微細構造の生成処理は、出力サンプリングレートが32kHzのときのみ、「ゼロビットサブバンド充填」と呼ばれる手法を用いてなされる。

サンプリングレート16kHzにおいて32kbit/s以下のレートに対しては、JT-G729.1とのビットイクザクタ性が保たれ、TDBWEアルゴリズムは3つの機能を実行するために用いられる。1つ目の機能は、14kbit/s出力レイヤを生成するものであり、2つ目の機能は、いくつかの低いエネルギーのサブバンドのスペクトル微細構造が符号化されずに符号化器から送られてこない場合に、4-7kHzにおけるゼロビットサブバンドを充填するものであり、3つ目の機能は、ビット列パッケージが転送の途中で失われた際に、4-7kHzスペクトルを生成するものである。

サンプリングレート32kHzにおいて32kbit/s以上のレートに対しては、前記の1つ目の機能はもはや必要なく、2つ目と3つ目の機能は、E.7.7.1節に記載される、より簡易なアルゴリズムに置き換えられる。JT-G729.1 コアコーデックにおいて、後処理を除く0-4kHzのNB時間領域出力は次のように表される。

$$\hat{s}_{LB}(n) = \hat{s}_{LB}^{celp}(n) + \hat{d}_{LB}(n) \quad (\text{E-129})$$

もしくは、

$$\hat{s}_{LB}(n) = \hat{s}_{LB}^{celp}(n) + \hat{d}_{LB}^{echo}(n) \quad (\text{E-130})$$

ここで、 $\hat{s}_{LB}^{celp}(n)$ はCELP出力、 $\hat{d}_{LB}(n)$ は、時間領域における原信号とCELP出力の量子化差分を表すMDCT拡張レイヤの出力、 $\hat{d}_{LB}^{echo}(n)$ は、エコー削減処理が行なわれた後の変形 $\hat{d}_{LB}(n)$ である(7.3.10節参照)。重み付け時間領域信号は、信号を重み付けフィルタ $W_{LB}(z)$ でフィルタリングすることにより得られ(6.6.2節参照)、式E-129は以下ようになる。

$$\hat{s}_{LB}^w(n) = \hat{s}_{LB}^{celp,w}(n) + \hat{d}_{LB}^w(n) \quad (\text{E-131})$$

MDCT周波数領域においては、式E-131は更に以下ようになる。

$$\hat{S}_{LB}^w(k) = \hat{S}_{LB}^{celp,w}(k) + \hat{D}_{LB}^w(k) \quad (\text{E-132})$$

ここで、 $\hat{S}_{LB}^{celp,w}(k)$ はCELPコーデック出力からのものである。また、 $\hat{D}_{LB}^w(k)$ は、MDCT拡張レイヤの出力からのものであり、より雑音となるように、原参照信号とCELPコーデック出力との誤差を補償するために用いられる。ここで、 $\hat{S}_{LB}^{celp,w}(k)$ はハーモニック要素からなり、 $\hat{D}_{LB}^w(k)$ は雑音要素からなるとみなせる。4-7kHzにおけるいくつかのサブバンド(ゼロビットサブバンド)のスペクトル微細構造が復号器において利用可能でない場合、これらのサブバンドは、NB情報を用いて以下のように充填される。

- 1) 信号の周期性をチェック。周期性は、 $G_p = E_p / (E_c + E_p)$ 、 $0 < G_p < 1$ と標記される正規化有声係数で表記され、CELPアルゴリズムの第2ステージから得られる。 $E_c$ と $E_p$ はそれぞれ、固定符号帳寄与のエネルギー、適応符号帳寄与のエネルギーである。

$$E_c = \sum_{n=0}^{39} (\hat{g}_c \cdot c(n) + \hat{g}_{enh} \cdot c'(n))^2 \quad E_p = \sum_{n=0}^{39} (\hat{g}_p \cdot v(n))^2$$

あるサブフレームから次のサブフレームへの平滑化有声係数は、 $\bar{G}_p \leftarrow 0.75 \bar{G}_p + 0.25 G_p$ と表記される。

- 2) 周期性が十分高いとき( $\overline{G}_p > 0.5$ )、式 E-132 における 0-3kHz のスペクトル係数  $\{\hat{S}_{LB}^w(k)\}$  は、ゼロビットサブバンドに対して、単純に 4-7kHz に複製される。これは、拡張ゼロビットサブバンドは、スペクトル包絡が適用される前に重み付き低域  $S_{BWE}(k) = \hat{S}_{LB}^w(k)$  に設定されることを意味する。
- 3) 周期性が低いとき( $\overline{G}_p \leq 0.5$ )は、拡張スペクトルは以下に設定される。

$$S_{BWE}(k) = g_h \cdot \hat{S}_{LB}^{cep,w}(k) + g_n \cdot \hat{D}_{LB}^w(k) \quad (\text{E-133})$$

ここで、 $g_h = 1 - 0.9(0.5 - \overline{G}_p)/0.5$ 、 $g_n = 1$  である。 $\hat{D}_{LB}^w(k)$  は、複雑さを抑えるための雑音性要素とみなされ、 $\hat{S}_{LB}^{cep,w}(k)$  と  $\hat{D}_{LB}^w(k)$  は既に復号器にて利用可能であるため、低域信号と拡張高域信号の同期が保たれる。

その後、JT-G729.1 標準と同様に、スペクトル包絡シェイピングが 4-7kHz の MDCT 係数に対して適用される(7.3.6 節参照)。TDBWE に送られた時間領域の包絡シェイピング情報は、4-7kHz の最終時間領域高域出力をシェイピングするために用いられる。

### E. 7. 2 MDCT 領域合成 WB 成分の取得

HF 信号の復号は、JT-G729.1 の WB 復号からの合成 MDCT 領域表現  $\hat{M}_{16}(k)$  を得ることから始める。MDCT 領域 WB 成分は、一般符号化フレームにおける HF 信号復号のために必要である。ここで、HF 信号は、WB 周波数範囲からの符号化サブバンドの適応的な複製によって構成される。

### E. 7. 3 汎用モード復号レイヤ 6 m o

汎用モードにおいて、HF 信号は、適応サブバンド複製によって構成される。さらに、二つの正弦波成分が、最初の 4kbit/sSWB 拡張レイヤのスペクトルに加算される。汎用モードと正弦波モードは、正弦波モード符号化技術に基づく同様の改善レイヤを利用する。改善レイヤの復号は E.7.5 節に詳述される。

#### E. 7. 3. 1 包絡正規化

WB 信号包絡正規化は、E.6.5.2 節に記載されている符号化器と同じように実行される。結果として、包絡正規化 WB 信号  $\tilde{M}_{16}(k)$  が得られる。

#### E. 7. 3. 2 サブバンド複製

最初に、4 つのサブバンドの各々に対する低周波数領域の開始位置がビット列から復号される。サブバンド 0、2 に対して、開始位置  $k^j$  が、ビット列における  $LagIndex^j$  によって決定される。サブバンド 1、3 に対して、開始位置  $k^j$  が、開始位置  $LagIndex^{j-1}$  と幅  $d^{j-1}$  の加算であるサブバンド 2、4 の終了位置近傍から下式のように算出される。

$$k^j = \begin{cases} LagIndex^j & j=0,2 \\ \min(\max(0, LagIndex^{j-1} + d^{j-1} - \frac{w^j}{2}), 280 - d^j - w^j) + LagIndex^j & j=1,3 \end{cases} \quad (\text{E-134})$$

ここで、 $w^j$  は式 E-53 で定義される各サブバンドに対する探索範囲である。HF サブバンドは、包絡正規化 WB 信号  $\tilde{M}_{16}(k)$  から合成される。

以下の節では、スケーリング操作とサブバンド複製の詳細を記述する。

### E. 7. 3. 2. 1 第1スケーリング

最初のスケーリング操作は、線形領域で実行される。まず利得係数符号  $Sign\_subband(j)$  と符号帳インデックスがビット列から得られる。インデックスは利得値  $\hat{\alpha}_1^{Log}(j)$  に対応し、下式を用いて線形領域に変換される。

$$\hat{\alpha}_1(j) = 10^{\hat{\alpha}_1^{Log}(j)} \quad (E-135)$$

さらに、元の符号が下式により復号される。

$$\hat{\alpha}_1(j) = \begin{cases} -\hat{\alpha}_1(j), & \text{if } (Sign\_subband(j) \neq 0) \\ \hat{\alpha}_1(j), & \text{otherwise} \end{cases} \quad (E-136)$$

合成 HF 成分は、下式のようにサブバンド  $j(j=0, \dots, 3)$  ごとに得られる。

$$\hat{M}_{32}^j(k) = \tilde{M}_{16}(k^j + k) \hat{\alpha}_1(j) \quad j=0, \dots, 3 \quad k=0, \dots, d^j - 1 \quad (E-137)$$

偶数  $k$  に対して、合成 HF 成分は以下のように対数領域に変換される。

$$\hat{M}_{32}^j(k) = 20 \cdot \log_{10} \left( \left| \hat{M}_{32}^j(k) + 1e-15 \right| \right) \quad j=0, \dots, 3 \quad k=0, 2, \dots, d^j - 2 \quad (E-138)$$

### E. 7. 3. 2. 2 第2スケーリング

元の HF スペクトルエネルギーへ適合させる第二のスケーリングは、対数領域で算出される。単独符号帳インデックス  $\alpha_2'$  がビット列から得られ、その後、対応する4つの利得  $\hat{\alpha}_2(j)$  がテーブル参照により得られる。符号化器同様、各対数領域サブバンドに対して、偶数  $k$  に対する最大値が探索される。

$$m^j = \max \left( -10000, \max_k \left( \hat{M}_{32}^j(k) \right) \right) \quad j=0, \dots, 3 \quad k=0, 2, \dots, d^j - 2 \quad (E-139)$$

式 E-78 にて定義される選択フラグ  $F_m^j(k)$  が1であるサンプルに対しては、第一のスケーリング後の MDCT 係数の符号は以下のように設定される。

$$Sign^j(k) = \begin{cases} 1, & \text{if } \hat{M}_{32}^j(k) < 0 \\ 0 & \text{otherwise} \end{cases}, \quad \text{if } (F_m^j(k) = 1) \quad j=0, \dots, 3 \quad k=0, \dots, d^j - 1 \quad (E-140)$$

合成された高域成分の奇数サンプルは、選択フラグ  $F_m^j(k)$  が1のときは対数領域に変換される。

$$\hat{M}_{32}^j(k) = 20 \cdot \log_{10} \left( \left| \hat{M}_{32}^j(k) + 1e-15 \right| \right), \quad \text{if } (F_m^j(k) = 1) \quad j=0, \dots, 3 \quad k=1, 3, \dots, d^j - 1 \quad (E-141)$$

その後、合成信号が以下のように得られる。

$$\hat{M}_{32}^j(k) = \begin{cases} 10^{0.05 \left\{ \hat{\alpha}_2(j) \left( \left| \hat{M}_{32}^j(k) \right|^{-m^j} \right)^{m^j} \right\}} & \text{if } (F_m^j(k) = 1 \text{ and } Sign^j(k) = 0) \\ -10^{0.05 \left\{ \hat{\alpha}_2(j) \left( \left| \hat{M}_{32}^j(k) \right|^{-m^j} \right)^{m^j} \right\}} & \text{else if } (F_m^j(k) = 1 \text{ and } Sign^j(k) = 1) \\ \hat{M}_{32}^j(k) & \text{otherwise} \end{cases} \quad \begin{matrix} j=0, \dots, 3 \\ k=0, \dots, d^j - 1 \end{matrix} \quad (E-142)$$

全体の合成信号は、以下のような4つのサブバンドからなる。

$$\hat{M}_{32}(k) = \begin{cases} \hat{M}_{32}^0(k-280), & k = 280, \dots, 319 \\ \hat{M}_{32}^1(k-320), & k = 320, \dots, 389 \\ \hat{M}_{32}^2(k-390), & k = 390, \dots, 459 \\ \hat{M}_{32}^3(k-460), & k = 460, \dots, 559 \end{cases} \quad (\text{E-143})$$

### E. 7. 3. 3 正弦波改善復号

一般符号化モードにおいては、再構成された HF スペクトル全体に対して2つの正弦波成分が付加される。正弦波は、位置、符号、振幅で表現される。正弦波の位置  $pos_{Gen}(l)$ ,  $l=0,1$  がビット列から得られると、正弦波符号が復号される。単独符号インデックス  $Sign\_ind$  (0 or 1) がビット列から復号され、第一の正弦波成分の符号  $Sign\_sin_{Gen}(0)$  (-1 or 1) を与える。第一の正弦波成分の位置  $pos_{Gen}(0)$  が第二の正弦波成分の位置  $pos_{Gen}(1)$  以下ならば、第二の正弦波成分は第一の正弦波成分の符号を継承する。そうでない場合は、第二の正弦波成分は反対の符号を与えられる。言い換えると以下の通りである。

$$\begin{aligned} Sign\_sin_{Gen}(1) &= Sign\_sin_{Gen}(0) && \text{if } pos_{Gen}(0) < pos_{Gen}(1) \\ Sign\_sin_{Gen}(1) &= -Sign\_sin_{Gen}(0) && \text{otherwise} \end{aligned}$$

正弦波成分の振幅  $amp_{Gen}(l)$  を復号するために、2つの符号帳インデックスがビット列から復号される。そして正弦波成分  $\hat{c}_0(l)$  が以下のように計算される。

$$\hat{c}_0(l) = -Sign\_sin_{Gen}(l) 10^{amp_{Gen}(l)} \quad l = 0,1 \quad (\text{E-144})$$

下式で与えられるように、HF スペクトルは2つの正弦波成分を付加することにより改善される。

$$\hat{M}_{32}(pos_{Gen}(l) + start\_ind_{Gen}) = \hat{M}_{32}(pos_{Gen}(l) + start\_ind_{Gen}) + \hat{c}_0(l) \quad l = 0,1 \quad (\text{E-145})$$

ここで、 $start\_ind_{Gen}$  は、最も大きいサブバンドエネルギーを持つサブバンドインデックスから得られるトラックの開始位置である。

$$\begin{aligned} SbE_{max} &= MAX(SbE(k)), && k = 0, \dots, 7 \\ K_{max} &= index(SbE_{max}) \\ start\_ind_{Gen} &= K_{max} \times 32 \end{aligned} \quad (\text{E-146})$$

ここで、 $SbE(k)$  は、 $\hat{M}_{32}(k)$  により式 E-48 から計算されるサブバンドエネルギーであり、 $K_{max}$  は、最も高いサブバンドエネルギー  $SbE_{max}$  を持つサブバンドのインデックスである。

#### E. 7. 3. 3. 1 FECパラメータ更新

正弦波改善成分の絶対位置はフレーム消失の改善のために保存され、保存された正弦波成分の数が更新される。

$$\begin{aligned} pos_{FEC}(l) &= pos_{Gen}(l) + start\_ind_{Gen}, \quad l = 0,1 \\ n_{FEC} &= 2 \end{aligned} \quad (\text{E-147})$$

### E. 7. 4 正弦波モード復号レイヤ6mo

正弦波モードにおいては、HF 信号は、限定された正弦波成分のセットから生成される。レイヤ6moにおいては、正弦波の総数は10であり、4つの正弦波が周波数範囲7000-8600 Hzに、4つの正弦波が周波数範囲8600-10200 Hzに、1つの正弦波が周波数範囲10200-11800 Hzに、1つの正弦波が周波数範囲11800-12600 Hzに配置される(E6.6節に説明されている通り)。

#### E. 7. 4. 1 正弦波位置

10 個の正弦波位置  $pos_j(l)$  を得るために、ビット列が復号される(E.6.6.1 節参照)。最初の 4 つの正弦波は、2 つの正弦波ごとの 2 つのトラックに分類される。同様に、続く 4 つの正弦波も 2 つごとの 2 つのトラックに分類される。最後の 2 つの正弦波は、別々のトラックとなる。ビット列から得られる位置情報は、これら 6 つのトラックに関連する。Figure E.3/JT-G729.1 に、レイヤ 6mo のトラック構造を絶対位置で示す。

##### E. 7. 4. 1. 1 FECパラメータ更新

正弦波成分の数とそれらの位置が、フレーム消失補償パラメータとして保存される。トラック  $j$  の開始インデックスの始まり  $start\_ind_j$ 、各トラック  $j$  の正弦波の数  $N_j$  として、全体の HF スペクトルに対して位置が保存される。

```

n_FEC = 0
for j = 0, ..., 5
  for l = 0, ..., N_j - 1
    pos_FEC(n_FEC + l) = pos_j(l) + start_ind_j
    n_FEC = n_FEC + 1
  end
end
end

```

#### E. 7. 4. 2 正弦波振幅

正弦波成分の振幅は、振幅と符号から構成される。6 つのトラック各々における最初の正弦波成分の 6 つの符号インデックス  $ind_p$  (0 or 1) を復号することにより、まず符号がビット列から得られる。符号  $Sign\_sin_j(l)$  (-1 or 1) は以下のように復号される。

```

for j = 0...5
  Sign_sin_p(0) = 2 * (ind_p - 0.5)
end

```

その後、2 つの正弦波を持つ最初の 4 つのトラックに対し、2 つ目の正弦波の符号が E.7.3.3 節の記述に基づき得られる。

```

for j = 0...3
  if pos_j(0) < pos_j(1)
    Sign_sin_j(1) = Sign_sin_j(0)
  else
    Sign_sin_j(1) = -Sign_sin_j(0)
  end
end
end

```

3 つの 8 ビット振幅符号帳インデックスがビット列から復号され、各々の符号帳から 3 つ、3 つ、4 つの振幅エントリ  $amp_j(l)$  を読み出すために使われる。

正弦波成分が以下のように再生され、

$$\hat{c}_j(l) = -Sign\_sin_j(l) 10^{amp_j(l)} \quad j = 0, \dots, 5 \quad l = 0, \dots, N_j - 1, \quad l = 0, 1 \quad (\text{E-148})$$

以下のように HF 信号に加算される。

$$\hat{M}_{32}(pos_j(l) + start\_ind_j) = \hat{M}_{32}(pos_j(l) + start\_ind_j) + \hat{c}_j(l) \quad , j = 0, \dots, 5 \quad l = 0, \dots, N_j - 1 \quad (\text{E-149})$$

## E. 7. 5 正弦波改善レイヤ復号

正弦波モード符号化技術は、汎用モードフレームと正弦波モードフレームの最上位の SWB 改善レイヤに利用される。それらは、以下の節にて記述されるレイヤ 7mo(4kbit/s)とレイヤ 8mo(8kbit/s)の 2 つの改善レイヤである。

### E. 7. 5. 1 レイヤ 7mo

最初の SWB レイヤであるレイヤ 7mo は、正弦波モードフレームにおける HF 信号スペクトルに対して 10 個以上の正弦波成分を付加する。汎用モードフレームでは、付加される正弦波成分の数は、LF と HF の改善の間での適応ビット配分に応じて設定される。

復号は、E.7.4 節に記載の正弦波モードと同様の原理に従う。最初に、正弦波位置がビット列から得られる。そして、伝送された符号インデックスと振幅符号帳インデックスを読み出すために、ビット列が復号される。その処理は、E.7.4.2 に詳述されているものと同様である。適応ビット配分に加えて、伝送されたトラックの符号インデックス及び位置は、正弦波モード 6mo と 7mo における正弦波改善では異なる。

汎用モードフレームにおいては、正弦波の数  $N_{sin}$  は、HF 符号化のビット数に応じて 6 つの値(0, 2, 4, 6, 8, 10)のうちの一つが設定される。正弦波符号化に対するトラックは、E.6.8.1 に詳述されるがごとく、位置  $start\_ind_j(l)$  から開始するトラックを与え、合成 HF スペクトルの最大サブバンドエネルギー  $N_{sin}/N_{sin\_track}$  を見つけ出すことにより選択される。

正弦波モードフレームにおいては、最初の 4 つの正弦波は 2 つずつの 2 つのグループに分類されるが、最後の 6 つの正弦波は 2 つずつの 3 つのグループに分類される。最初の 4 つの正弦波は周波数範囲 9400-11000Hz にあり、最後の 6 つの正弦波は周波数範囲 11000-13400Hz にある。

復号は、E.7.4 節の正弦波モードの記述に基づき処理する。それぞれのトラックに関連する 10 個の正弦波の位置インデックスは、最初にビット列から読み出される。そして、10 個の正弦波の符号が復号される。最後に振幅(3 つの 8 ビット符号帳インデックス)が復号され、正弦波成分が式 E-148 に従って得られる。

HF 信号スペクトルは、復号された正弦波成分  $\hat{c}_j(l)$  を付加することにより、以下のように更新される。

$$\hat{M}_{32}(pos_j(l)+start\_ind_j)=\hat{M}_{32}(pos_j(l)+start\_ind_j)+\hat{c}_j(l), \quad l=0,1 \quad (\text{E-150})$$

#### E. 7. 5. 1. 1 FEC パラメータ更新

レイヤ 7mo の正弦波の位置は FEC パラメータとして保存され、現フレームの正弦波の数はインクリメントされる。

for  $j=0,\dots,4$

for  $l=0,\dots,N_j-1$

$pos_{FEC}(n_{FEC}+l)=pos_j(l)+start\_ind_j$

$n_{FEC}=n_{FEC}+1$

end

end

### E. 7. 5. 2 レイヤ 8mo

レイヤ 8mo においては、別の 20 個の正弦波が HF 信号に付加される。トラック構造は、汎用モードフレームと正弦波モードフレームで異なる。汎用モードにおいては、レイヤ 7mo で得られる正弦波の数は、E.6.8.2 に記述されるようにこのレイヤのトラック位置に影響を与える。

復号は、前節の記述に基づき処理される。それぞれのトラックに関連する、10 個の正弦波の最初の組の位



置インデックスは、ビット列から最初に読み出される。そして、10 個の正弦波に対する符号が復号される。最後に振幅(3 つの 8 ビット符号帳インデックス)が復号され、この手順が 10 個の正弦波の第 2 の組に対して繰り返される。

HF 信号スペクトルは、復号された正弦波成分  $\hat{c}_j(l)$  を付加することにより、以下のように更新される。

$$\hat{M}_{32}(pos_j(l)+start\_ind_j)=\hat{M}_{32}(pos_j(l)+start\_ind_j)+\hat{c}_j(l) , j=0,\dots,9 \quad l=0,\dots,N_j-1 \quad (\text{E-151})$$

#### E. 7. 5. 2. 1 FECパラメータ更新

レイヤ 8mo の 20 個の正弦波の位置は FEC パラメータとして保存され、現フレームの正弦波の数はインクリメントされる。

```

for j = 0, ..., 9
  for l = 0, ..., N_j - 1
    pos_FEC(n_FEC + l) = pos_j(l) + start_ind_j
    n_FEC = n_FEC + 1
  end
end
end

```

#### E. 7. 6 WB改善レイヤ復号

##### E. 7. 6. 1 スペクトル包絡復号

##### E. 7. 6. 2 聴覚重要度によるサブバンド順序付け

E.6.9.4 節と同じ。

##### E. 7. 6. 3 ビット配分

E.6.9.5 節と同じ。

##### E. 7. 6. 4 MDCT係数復号とスペクトル逆正規化

$j$  番目のサブバンドの次元が 8 ならば、エンベディッド球面ベクトル量子化(SVQ)により復号される。一方、 $j$  番目のサブバンドの次元が 16 ならば、Gosset 低演算ベクトル量子化(GLCVQ)により復号される。

##### E. 7. 6. 4. 1 SVQに対するベクトル量子化インデックスの復号

7.3.5 節と同じ。

##### E. 7. 6. 4. 2 GLCVQに対するベクトル量子化インデックスの復号

GLCVQ インデックス  $i_Q$  の復号は、以下のステップからなる。

- 1) ビット列から  $i_Q$  を抽出す
- 2) 受信した  $i_Q$  に従い、テーブル引きと比較によりルートクラス  $m'_Q$  を特定する
- 3) 受信した  $i_Q$  とルートクラス  $m'_Q$  に応じて、テーブル参照と比較によりクラス  $m_Q$  を特定する
- 4)  $i_Q$  からインデックス  $i_{P_c}$  を抽出す

$$i_{P_c} = i_Q - i_{off, \varepsilon_{m_Q}} \quad (\text{E-152})$$

ここで、 $i_{off, \varepsilon_{m_Q}}$  は選択されたクラスリーダーベクトルのインデックスオフセットに相当する。

- 5) インデックス  $i_{P_c}$  に応じて、最適コードベクトルを復号する。インデックス  $i_{P_c}$  の復号方法を以下に記す。  
ステップ 1: 振幅ベクトルと重みベクトルを得る。振幅ベクトルと重みベクトルは、選択されたクラスリーダーベクトルに応じて構成される。

ステップ2：重みベクトルを降順に並べ  $[w'_0 w'_1 \dots w'_{L_p-1}]$ 、それにならって振幅ベクトルを並べ替える。  
 例えば、 $L_p=3$ において、オリジナルの振幅ベクトルが  $[u_0 u_1 \dots u_{L_p-1}] = [2 \ 0 \ -2]$ 、重みベクトルが  $[w_0 w_1 \dots w_{L_p-1}] = [1 \ 6 \ 1]$  のとき、並べ替えられた重みベクトルは  $[w'_0 w'_1 \dots w'_{L_p-1}] = [6 \ 1 \ 1]$ 、並べ替えられた振幅ベクトルは  $[u'_0 u'_1 \dots u'_{L_p-1}] = [0 \ 2 \ -2]$  となる。

ステップ3：並べ替えられた振幅ベクトルと並べ替えられた重みベクトルに応じて、最適コードベクトルを再生する。

A.  $mid\_index_n$  を得る：最適ベクトルインデックスは、最低レベルから最高レベルの各レベルに対していくつかの中間インデックスに分解される。最適ベクトルインデックスは、最低レベルに対する開始値である。各最低レベルの  $mid\_index_n$  は、可能なインデックス値カウント  $C_{m_{n-1}}^{m_n}$  でインデックスを除することにより得られ、商は次の低レベルのインデックスとなる。余りは現レベルの  $mid\_index_n$  となる。そして、 $n$  ( $L_p > n > 0$ ) をデクリメントする。

$$mid\_index_n = index \bmod C_{m_{n-1}}^{m_n} \quad (E-153)$$

$$index = \lfloor index / C_{m_{n-1}}^{m_n} \rfloor \quad (E-154)$$

ここで、 $m_{n-1}$  は上位レベル(レベル  $n-1$ )ベクトルの次元であり、 $m_n$  は現レベル(レベル  $n$ )ベクトルの次元である。

$$m_n = \sum_{i=n}^{L_p-1} w'_i \quad (E-155)$$

B. 位置ベクトルの復号：各低レベルの  $mid\_index_n$  は、順列組合せ関数に基づいて復号され、各々の上位レベルベクトルに関する低レベルベクトルの位置ベクトルが得られる。

各低レベルの中間インデックスから位置ベクトルを得るため、アルゴリズムは、位置列を評価するための順列組合せ関数を用いる。評価手順は以下の通りである。

- 中間インデックスが  $C_{m_{n-1}}^{m_n} - C_{m_{n-1}-pos}^{m_n}$  よりも小さい間、ゼロから開始する  $pos$  値をインクリメントする。
- $p_0 = pos - 1$  を最初の位置とし、中間インデックスから  $C_{m_{n-1}}^{m_n} - C_{m_{n-1}-p_0}^{m_n}$  を減じる。
- 中間インデックスが  $C_{m_{n-1}-p_{i-1}-1}^{m_n-i} - C_{m_{n-1}-pos}^{m_n-i}$  よりも小さい間、 $p_{i-1} + 1$  から開始する  $pos$  をインクリメントする。ここで、 $p_{i-1}$  は前ステップで復号された位置である。
- $p_i = pos - 1$  を位置番号  $i$  とし、中間インデックスから  $C_{m_{n-1}-p_{i-1}-1}^{m_n-i} - C_{m_{n-1}-p_i}^{m_n-i}$  を減じる。
- 現レベルの位置列から全ての位置が復号されるまで、ステップ c) と d) を繰り返す。

C. 最適コードベクトル復号：最後に、最低レベルから最高レベルまで一つずつ、各低レベルベクトルは位置パラメータに応じて上位レベルベクトル要素を部分的に置き換えるために使われる。

最低レベルにおいては、ベクトルは、並べ替えられた振幅ベクトルから得られる値である一つのタイプの要素のみを包含する。最低レベルベクトルは次のレベルに渡され、次のステップにおいて別のタイプの要素が加えられる。この新しい要素は並べ替えられた振幅ベクトルから得られる。この手順は、最高レベルに達するまで繰り返される。

## E. 7. 7 フレーム消失補償

### E. 7. 7. 1 4-7 kHzにおけるMDCT符号化の低演算量FECアルゴリズム

本節では、出力サンプリングレートが 32kHz のときの JT-G729.1 の変更について記述する。14kbit/s よりも高いレイヤは MDCT ベースの符号化アルゴリズムにより符号化されるが、E.7.1.1 節にて記述したように、JT-G729.1 における TDBWE アルゴリズムの三番目の機能は、14kbps レイヤのみならずそれよりも高

レイヤに対しても高域 [4-7kHz] のフレーム消失補償(FEC)をするためのものである。

サンプリングレート 32kHz における 32kbit/s 以上の符号化レートに対して、14kbit/s のレイヤは不要となる。TDBWE の二番目の機能は E.7.1.1 節記載のより簡易なアルゴリズムに置き換えられ、TDBWE の三番目の機能もまた本節にて記載されるより簡易なアルゴリズムに置き換えられる。

サンプリングレート 32kHz における 32kbit/s 以上のレートに対する高域 [4-7kHz] の FEC アルゴリズムは、MDCT ベースのコーデックアルゴリズムの特性を調査する。パケットロスが発生した場合の FEC は 2 つの主機能からなり、1 つは MDCT 領域の係数生成、もう 1 つは時間領域における高域信号の時間領域エネルギー形状修正である。二つの主機能の詳細を以下に記述する

ー パケットロス時の MDCT 領域係数の推定

パケットロス時の単純な解決策は、MDCT 領域係数を前フレームから現フレームに複製することである。しかし、このような単純な前 MDCT 係数の繰り返しは、不自然な音もしくは過度な周期性(高すぎる調波性)を生じることがある。信号の周期性と音の自然性を制御するために、前フレームから複製された高域[4-7kHz]の MDCT  $\hat{S}_{HB}^{old}(k)$  係数に、ランダムノイズ成分  $N(k)$  が適応的に加算される。

$$\hat{S}_{HB}(k) = g_1 \cdot \hat{S}_{HB}^{old}(k) + g_2 \cdot N(k) \quad (\text{E-156})$$

[7-8kHz]の全ての MDCT 係数は、コーデック定義の観点からゼロに設定される。20 個の MDCT 係数ごとに 1 サブバンドが定義され、4-8kHz は 8 つのサブバンドとなる。[7-8kHz]の最後の二つのサブバンドはゼロに設定される。乱数雑音係数  $N(k)$  のエネルギーは、各サブバンドにおいて最初に  $\hat{S}_{HB}^{old}(k)$  のエネルギーに正規化される。式 E-156 における  $g_1$  及び  $g_2$  は、FEC 中における前フレームと比較した適切な全エネルギー減衰を維持しながら、 $\hat{S}_{HB}^{old}(k)$  と  $N(k)$  との間のエネルギー比を制御するために推定された利得である。 $\bar{G}_p, 0 \leq \bar{G}_p \leq 1$  が信号の周期性を測るために定義されたパラメータであり、 $\bar{G}_p = 0$  は周期性がないことを意味し、 $\bar{G}_p = 1$  は完全な周期性を表す。そして、 $g_1$  及び  $g_2$  は以下のように定義される。

$$g_1 = g_r \cdot \bar{G}_p \quad (\text{E-157})$$

$$g_2 = g_r \cdot (1 - \bar{G}_p) \quad (\text{E-158})$$

ここで、 $g_r = 0.9$  は前フレームの一つよりも低い現フレームエネルギーを維持するための MDCT 領域における利得減衰係数である。実際は、この段階におけるエネルギー制御はそれほど重要ではない。なぜならば、時間領域において後に時間領域のエネルギー形状が修正されるからである。 $\bar{G}_p$  は直近の平滑化された有声係数であり、ある受信サブフレームから次の受信サブフレームで  $\bar{G}_p \leftarrow 0.75 \bar{G}_p + 0.25 G_p$  のように表現される。そして  $G_p$  は、他節で記述した受信サブフレームと同じ定義である。

$$G_p = \frac{E_p}{E_p + E_c} \quad (\text{E-159})$$

FEC フレームの間、更に FEC フレームが連続して生じた時に周期性を低減していくために、 $\bar{G}_p$  は現フレームから次のフレームにかけて係数 0.75 で減少される ( $\bar{G}_p \leftarrow 0.75 \bar{G}_p$ )。生成された MDCT 係数  $\hat{S}_{HB}(k)$  が決定された後、それらは時間領域に逆変換される。推定高域信号  $\hat{s}_{HB}(n)$  を得るために、逆変換の間、現 MDCT 窓における寄与成分は一つ前の MDCT 窓の寄与成分と補間される。

- ー 高域[4-7kHz]と低域[0-4kHz]のエネルギー比に基づく FEC の時間領域エネルギー制御  
8kHz でサンプリングされた低域及び高域の時間領域合成信号は、それぞれ  $\hat{s}_{LB}(n)$  及び  $\hat{s}_{HB}(n)$  と表

記される。誤差のない状況においては、 $\hat{s}_{LB}(n)$  は CELP 出力と MDCT 拡張レイヤ出力の組み合わせとなり ( $\hat{s}_{LB}(n) = \hat{s}_{LB}^{celp}(n) + \hat{d}_{LB}^{echo}(n)$ )、通常は CELP 出力  $\hat{s}_{LB}^{celp}(n)$  が主要成分となる。 $\hat{s}_{HB}(n)$  は、 $\hat{S}_{HB}(k)$  の逆 MDCT を実行することにより得られる。16kHz でサンプリングされた最終出力  $\hat{s}_{WB}(n)$  は、 $\hat{s}_{LB}(n)$  及び  $\hat{s}_{HB}(n)$  をアップサンプリングし、アップサンプリングされた信号を QMF 合成フィルタバンクでフィルタリングすることにより計算される。時間領域信号  $\hat{s}_{HB}(n)$  が  $\hat{S}_{HB}(k)$  の逆 MDCT により得られたとき、 $\hat{s}_{HB}(n)$  の最初の FEC フレームに対する正確な時間領域の包絡形状が直近の受信 TDBWE パラメータから得られるよう、 $\hat{s}_{HB}(n)$  は直近の受信 CELP フレームもしくは時間領域における TDBWE フレームと比してちょうど 1 フレーム遅延している。時間領域のエネルギー包絡を求めるため、ひとつの 20ms フレームは 8 個のサブセグメント(1 サブセグメントは 2.5ms)に分割される。そして、時間領域のエネルギー包絡は、各サブセグメントのエネルギーを表す  $Tenv(i)$ ,  $i=0,1,..,7$  として表記される。最初の  $\hat{s}_{HB}(n)$  の FEC フレームに対して、 $Tenv(i)$  は直近の受信 TDBWE パラメータの復号することによって、単純に得られる。そして、対応する低域 CELP 出力  $\hat{s}_{LB}^{celp}(n)$  は直近の受信 CELP パラメータを復号することによって正確である。しかし、MDCT 拡張レイヤからの寄与  $\hat{d}_{LB}^{echo}(n)$  は一部分しか正しくなく、最初の FEC フレームから 2 番目の FEC フレームまでにゼロに減衰されなければならない。

$\hat{s}_{HB}(n)$  の二番目の FEC フレーム及びそれ以降の FEC フレームにおいては、JT-G729.1 の対応する低域出力  $\hat{S}_{LB}(n)$  を回復するために、注意深く厳密に設計された FEC アルゴリズムを用いる。高域信号  $\hat{s}_{HB}(n)$  は、式 E-156 で表現される、 $\hat{S}_{HB}(k)$  の逆 MDCT によって最初に推定される。 $\hat{s}_{LB}(n)$  と  $\hat{s}_{HB}(n)$  は各々異なる手法による異なる道筋で推定されるので、これらの相対的なエネルギー関係は、聴覚的な観点では重要であるが最適ではない。 $\hat{s}_{HB}(n)$  のエネルギーは、 $\hat{s}_{LB}(n)$  のエネルギーに比べると、時間領域においては高すぎるかもしくは低すぎる。この問題を解決するための一つの方法は、 $\hat{s}_{HB}(n)$  の最後の受信フレームもしくは最初の FEC フレームから  $\hat{s}_{LB}(n)$  と  $\hat{s}_{HB}(n)$  のエネルギー比を得、このエネルギー比を後続の FEC フレームのために保持することである。

逆 MDCT は 1 フレームの遅延を生じるが、 $\hat{s}_{HB}(n)$  の最初の FEC フレームの間は、低域信号と高域信号のエネルギー比の良好な推定値が計算される。低域エネルギーは、JT-G729.1 復号器の低域信号  $\hat{s}_{LB}(n)$  から得られる。高域エネルギーは、直近の受信 TDBWE パラメータから求められる時間領域のエネルギー包絡  $Tenv(i)$  パラメータの単純な和である。エネルギー比は以下のように定義される。

$$Ratio = \frac{E_{HB}}{E_{LB}} = \frac{\sum_i Tenv(i)}{\|\hat{s}_{LB}(n)\|^2} \quad (\text{E-160})$$

式 E-160 は、時間領域フレーム全体の平均エネルギー比を表す。

$\hat{s}_{HB}(n)$  の最初の FEC フレームに対しては、時間領域のエネルギー包絡  $Tenv(i)$  は、各高域サブセグメント  $\hat{s}_{HB}^i(j) = \hat{s}_{HB}(20 \cdot i + j)$  に利得係数  $g_f(i)$  を乗じることにより直接的に適用される。

$$g_f(i) = 0.9 \sqrt{\frac{Tenv(i)}{\sum_{j=0}^{20} |\hat{s}_{HB}(i \cdot 20 + j)|^2}} \quad i = 0,1,..,7 \quad (\text{E-161})$$

上記利得係数は、利得係数乗算の間、サンプルごとに平滑化される。

$$\bar{g}_f(j) = 0.95 \cdot \bar{g}_f(j-1) + 0.05 \cdot g_f(i) \quad (\text{E-162})$$

$$\hat{s}_{HB}(i \cdot 20 + j) = \hat{s}_{HB}(i \cdot 20 + j) \cdot \bar{g}_f(j) \quad (\text{E-163})$$

式 E-161、E-162、E-163 において、 $i$  はサブセグメントのインデックス、 $j$  はサンプルのインデックスである。

$\hat{s}_{HB}(n)$  の二番目の FEC フレーム及びそれ以降の FEC フレームに対しては、各フレームは 8 つ個の細かいサブセグメントに分割される。エネルギー比修正は、各々の細かいサブセグメントにおいて実行される。 $i$  番目のサブセグメントに対するエネルギー修正利得係数  $g_i$  は、以下のように計算される。

$$g_i = \sqrt{\text{Ratio} \cdot \frac{\|\hat{s}_{LB}^i(j)\|^2}{\|\hat{s}_{HB}^i(j)\|^2}} \quad (\text{E-164})$$

if  $g_i > 1, \quad g_i = 1$

式 E-164 において、 $\|\hat{s}_{LB}^i(j)\|^2$  と  $\|\hat{s}_{HB}^i(j)\|^2$  は、各々  $i$  番目の低域信号  $\hat{s}_{LB}^i(j) = \hat{s}_{LB}(20 \cdot i + j)$  のエネルギー、高域信号  $\hat{s}_{HB}^i(j) = \hat{s}_{HB}(20 \cdot i + j)$  のエネルギーを表す。式 E-164 で定義される修正利得は、あるセグメントから次のセグメントまでサンプル毎に平滑化されるが、最終的に  $i$  番目のサブセグメント  $\hat{s}_{HB}^i(j)$  にまで適用される。

$$\bar{g}_i(j) = 0.95 \cdot \bar{g}_i(j-1) + 0.05 \cdot g_i \quad (\text{E-165})$$

$$\hat{s}_{HB}^i(j) = \hat{s}_{HB}^i(j) \cdot \bar{g}_i(j) \quad (\text{E-166})$$

最後のステップにおいて、修正された高域信号  $\hat{s}_{HB}(n)$  のエネルギー及び低域信号  $\hat{s}_{LB}(n)$  のエネルギーは、最終的な広帯域出力信号  $\hat{s}_{WB}(n)$  を形成するために、アップサンプリングされ、QMF フィルタバンクでフィルタリングされる。

### E. 7. 7. 2 HF 信号の FEC

SWB の FEC は、直前のフレームの HF 信号にスケールングを適用することにより、フレーム消失の間 HF 信号を減少させる。正弦波成分は、消失フレームにおいてトーン信号の特性を保持するために、全体の包絡よりも強く維持される。

二つのスケールング係数  $\beta_{FEC}$  と  $\beta_{FEC, \sin}$  が FEC 処理において用いられる。直前のフレームが汎用モードフレームの場合、スケールング係数は各々 0.5 と 0.6 に設定される。直前のフレームが正弦波モードフレームの場合、各々 0.0 と 0.8 に設定される。HF 信号は、直前のフレームの HF 信号をスケールングすることにより、以下のように生成される。

$$\begin{aligned} \hat{M}_{32}(k) &= \beta_{FEC} \hat{M}_{32, \text{prev}}(k) & k &= 280, \dots, 559 \\ \hat{M}_{32}(\text{pos}_{FEC}(k)) &= \beta_{FEC, \sin} \hat{M}_{32, \text{prev}}(\text{pos}_{FEC}(k)) & k &= 0, \dots, n_{FEC} - 1 \end{aligned} \quad (\text{E-167})$$

### E. 7. 8 SWB 出力と WB 出力の切替

コーデックビットレートが周期的に切り替わるような場合、SWB 拡張に割り当てられるビットレートに対して、短い期間ゼロに落としその後再現することが可能である。このような場合、7-14kHz 周波数成分のオンとオフが切り替わることが、聴覚上不快に感じられることが良く知られている。そのような合成されたバンド幅の切り替えの影響を緩和するため、7-14kHz 成分周波数を符号化するためのビットレートが不十分のときは、失われた成分を近似するためにゼロビット帯域拡張アルゴリズムが動作する。このアルゴリズムは、復号器にて利用可能な広帯域周波数成分(50-7000Hz)の分析に基づく。このアルゴリズムは二つの動作モードを持ち、一つは、7-14kHz 成分周波数が活発に生成されるアクティブ処理、もう一つは、逆 MDCT 合成バッファが残りのデータを持つときにエネルギー予測が更新されて 7-14kHz 成分周波数が単に生成されるスタ

ンバイ処理である。これらの二つのモードにおける処理手順は、以降の節に記述される。

### E. 7. 8. 1 アクティブ処理

#### E. 7. 8. 1. 1 遷移帯域分析

20ms フレームサイズかつ 16kHz サンプリングに対しては、遷移帯域 MDCT 係数のインデックス範囲は 160 から 279 である。遷移帯域(TB)分析ブロックにおいて、120 個の係数(4-7kHz)で表現されるスペクトルは、スペクトル包絡及び励振信号のスペクトルを抽出するために分析される。

スペクトル包絡を抽出するため、絶対値演算子が、振幅スペクトルを得るために最初に適用される。次に、振幅スペクトル中のゼロ値セクションが特定される。このようなセクションは、符号化器での MDCT 係数量子化の結果である。ゼロ振幅は、係数(例えば 0.2)でスケールされた補間振幅で置き換えられる。補間は非ゼロ境界振幅で線形に行われる。この操作は、振幅スペクトルのダイナミックレンジを低減し、より正確な包絡計算結果をもたらす。その後、振幅スペクトルは dB 領域に変換され、ピボットの役割を果たすインデックス 279 でのスペクトル折り返しにより、160 から 319 の範囲をカバーするために拡張される。この拡張された dB スペクトルは対数スペクトルの正の半分として扱われ、その後包絡を抽出するためにケプストラム手法が適用される。適用されるケプストラム手法は、逆離散フーリエ変換(IDFT)、DC 係数を含む最初の 8 個のケプストラム係数のみを保持するためのフィルタリング、離散フーリエ変換を包含する(DFT)。結果として得られる。インデックス範囲 160 から 279 に対応する包絡は 1/2 にダウンサンプリングされ、dB 領域における TB スペクトル包絡として扱う。TB 励振信号スペクトルを得るため、TB MDCT スペクトルは、TB スペクトル包絡を用いて平坦化される。これは、包絡値の全てを得るために TB スペクトル包絡において 1:2 の線形補間を実行し、対応する dB 包絡値で振幅スペクトルから減じ、得られた励振信号振幅を線形領域に変換し、オリジナルの MDCT スペクトルから適切な符号を割り当てることにより、dB 領域において簡単に実現される。振幅スペクトルのゼロ値セクションに対応する TB 励振信号スペクトルにおけるどのような穴も、予め保存された擬似乱数雑音系列の MDCT スペクトルの値で満たされる。フレームレートにおける雑音 MDCT スペクトルの周期的な繰り返しに起因する人工感を緩和及び最小限にするために、この二つの予め保存された雑音 MDCT スペクトルは、交互に使われる。

#### E. 7. 8. 1. 2 高域励振信号生成器

20ms フレームサイズかつ 32kHz サンプリングに対しては、高域スペクトルに対するインデックス範囲は 280 から 559 である(7-14kHz)。この範囲に対応する励振信号スペクトルは、TB 励振信号スペクトル及びピッチ遅延情報から以下のように生成される。最初に、ピッチ遅延がピッチ周波数に変換される。各フレームにおける最後の(5ms)サブフレームに対応するピッチ遅延が、この目的のために用いられる。続いて、対応する MDCT インデックス領域において定義される 120 以下の整数周波数遅延  $D$  を得るために、このピッチ周波数の最も大きい整数倍数が特定される。その後、 $D$  を用いた TB 励振信号スペクトルの周期的繰り返しを利用して、HB 励振信号スペクトルが以下のように生成される。

$$\hat{M}_{exc}(i) = \hat{M}_{exc}(i - D) \quad i = 280, \dots, 559, D \leq 120 \quad (\text{E-168})$$

ここで、 $\hat{M}_{exc}$  は MDCT 領域における HB 励振信号スペクトルである。このアプローチは高域に対する TB 励振信号スペクトルの再利用を保証し、遷移帯域と高域との間の MDCT スペクトルの調波関係を維持する。

#### E. 7. 8. 1. 3 高域エネルギー推定器

高域エネルギーを推定するために、遷移帯域エネルギー  $E_{tb}$  が、TB MDCT 係数の二乗和として最初に計算される。

$$E_{tb} = 10 \log \left[ \sum_{280}^{559} \left( \hat{M}_{exc}(i) \right)^2 \right] \quad (\text{E-169})$$

dB 値  $E_{ib}$  により、高域エネルギーの dB 値  $E_{hb0}$  が以下のように推定される。

$$E_{hb0} = \alpha E_{ib} + \beta \quad (\text{E-170})$$

ここで、 $\alpha$  と  $\beta$  は、学習用データベースの多数のフレームに対して高域エネルギーの真値と推定値の平均二乗誤差を最小にするように選択される。正規化ゼロクロスパラメータ  $zc$  (範囲: 0 から 1) や TB スペクトル包絡形状  $TBEnv$  のような付加パラメータからの情報を用いることにより、高域エネルギー推定の正確さが改善される。 $zc$  -  $TBEnv$  パラメータ平面は、 $zc$  の 8 つのスカラ量子化レベル及び  $TBEnv$  の 8 つのベクトル量子化(VQ)形状に対応した、最大で 64 のパーティションに分割される。関連のない係数  $\alpha$  と  $\beta$  は、学習/設計段階の間、パラメータ平面の各セグメントに対して選択される。更に、エネルギー適応で使用されるため、各セグメントに対応する推定誤差の標準偏差  $\sigma$  が計算され、予め保存される。異なるセグメントに対しては、 $\sigma$  の値はおよそ 4dB から 8dB の間で変動し、平均はおよそ 5.9dB である。

#### E. 7. 8. 1. 4 高域エネルギー適応器

推定された高域エネルギーは、人工感を最小化するために以下の記述のように適応され、その結果、出力 SWB 音声の品質が向上する。高域エネルギーの推定は誤差が生じやすい。過剰推定は人工感に通ずるため、推定された高域エネルギーは、以下のように推定誤差の標準偏差に比例した量によって小さくする方向に偏らせる。

$$E_{hb1} = E_{hb0} - \lambda \cdot \sigma \quad (\text{E-171})$$

ここで、 $E_{hb1}$  は適応された高域エネルギーの dB 値であり、 $\lambda \geq 0$  は比例係数である。上述の推定された高域エネルギーの“下方偏向”により、エネルギーの過剰推定の確率は低減され、人工感の頻度も低減される。また、推定されたエネルギーが低減される量は、いかに良い推定かどうか按比例し、信頼性の高い(すなわち低い  $\sigma$  値である)推定は、信頼性の低い推定よりも小さい量で低減される。この推定されたエネルギーの“下方偏向”は、有声音フレームに対して、高域スペクトル包絡形状の推定における誤差により生じる“雑音性”の人工感を隠蔽するという付加的な恩恵を持つ。しかし、無声音フレームに対しては、推定された高域エネルギーにおける低減が高すぎる場合、出力音声は超広帯域音声のように聴こえない。これに対処するため、推定された高域エネルギーは、有声音レベル  $\nu$  に依存して更に以下のように適応される。

$$E_{hb2} = E_{hb1} + (1 - \nu) \cdot \delta_1 + \nu \cdot \delta_2 \quad (\text{E-172})$$

ここで、 $E_{hb2}$  は有声音レベルで適応された高域エネルギーの dB 値であり、 $\delta_1$  と  $\delta_2$  ( $\delta_1 > \delta_2$ ) は dB 値の定数である。 $\delta_1$  と  $\delta_2$  の選定は“下方偏向”で使われる  $\lambda$  の値に依存し、最適に聴こえる出力音声となるよう、実験的に決定される。典型的な  $\lambda$  と  $\delta_1$  と  $\delta_2$  の値は、それぞれ、1.2、3.0、-3.0 である。有声音レベル  $\nu$  自体は、正規化ゼロクロスパラメータ  $zc$  と二つの閾値  $ZC_{low}$  と  $ZC_{high}$  から推定される。 $zc$  が  $ZC_{low}$  以下であれば  $\nu$  は 1、 $zc$  が  $ZC_{high}$  以上であれば 0、そうでなければ、 $ZC_{low}$  と  $ZC_{high}$  の間の範囲は、 $\nu$  は 0 から 1 の範囲で線形に写像される。最後に、推定された高域エネルギーは、パラメータ  $d$  により示される立ち上がり/破裂音の発生に依存して適応される。立ち上がり/破裂音は、以下の理由により特別な問題を提起する。a) 立ち上がり/破裂音近傍の高域エネルギー推定は困難である、b) 典型的なブロック処理が用いられるため、出力音声にプリエコー型の人工感が生じる。もし、先行するフレームの広帯域音声エネルギー  $E_{wb}$  がある閾値以下であり、現フレームと先行するフレームの間の広帯域エネルギー差が他の閾値を超える場合、もしくは同様な状況が  $E_{ib}$  との関連で存在するならば立ち上がり/破裂音が検知される。立ち上がり/破裂音の検知( $d=1$ )されたときの高域エネルギー適応は、以下のように実行される。

$$\begin{aligned}
E_{hb}(k) &= E_{\min} & k &= 1, \dots, K_{\min} \\
E_{hb}(k) &= E_{hb2}(k) - \Delta & k &= K_{\min} + 1, \dots, K_T \\
E_{hb}(k) &= E_{hb2}(k) - \Delta + \Delta_T(k - K_T) & k &= K_T + 1, \dots, K_{\max}
\end{aligned} \tag{E-173}$$

立ち上がり/破裂音が検出されたフレーム( $k=1$ )から始まる最初の $K_{\min}$ フレームに対しては、高域エネルギーは最小値 $E_{\min}$ に設定される。それに続くフレーム(即ち、 $k = K_{\min} + 1, \dots, K_{\max}$ )に対しては、フレームの有声音レベル $v(k)$ が閾値 $V_1$ を超えた場合のみエネルギー適応がなされる。この範囲の中のフレームの有声音レベルが $V_1$ 以下となる場合はいつでも、立ち上がり/破裂音エネルギー適応は直ちに停止される。この特性は、有声音の立ち上がりなどの短い区間のエネルギー適応を実行する。有声音レベル $v(k)$ が $V_1$ よりも大きい場合、 $k = K_{\min} + 1, \dots, K_T$ に対しては、高域エネルギーは固定値 $\Delta$ で低減される。 $k = K_T + 1, \dots, K_{\max}$ に対しては高域エネルギーは、予め特定された数列 $\Delta_T(k - K_T)$ を用いて、 $E_{hb2}(k)$ に向かって $E_{hb2}(k) - \Delta$ により徐々に増加され、 $k = K_{\max} + 1$ において、 $E_{hb2}(k)$ に設定される。立ち上がり/破裂音が検出されない場合( $d = 0$ )、最後の適応された高域エネルギー推定 $E_{hb}$ は、 $E_{hb2}$ に設定される。適応パラメータの典型的な値は、 $K_{\min} = 2$ 、 $K_T = 3$ 、 $K_{\max} = 5$ 、 $V_1 = 0.9$ 、 $\Delta = -12$  dB、 $\Delta_T(1) = 6$  dB and  $\Delta_T(2) = 9.5$  dB である。

### E. 7. 8. 1. 5 高域包絡切替器

適応された高域エネルギー $E_{hb}$ に基づき、高域スペクトル包絡 $HB\ Env$ は、包絡形状のルックアップテーブルを用いて選択される。ルックアップテーブルは、以下のように生成される。32kHzでサンプリングされた超広帯域音声の大きな学習用データベースをはじめとし、MDCTスペクトルが各フレームに対して計算される。そして、振幅スペクトルが絶対値演算子を用いて計算され、dB領域に変換される。4000-16000Hz周波数帯域に対応するdB値振幅スペクトルから、ケプストラム手法を用いて、スペクトル包絡が計算される。このケプストラム手法は、使われるケプストラム係数が24であるという違いはあるが、遷移帯域包絡を計算するために用いられるのと同様手法である。各フレームの超広帯域スペクトル包絡から、7000-16000Hzに対応する高域包絡が抽出され、7000Hzのスペクトル振幅で全体が正規化される。正規化された高域スペクトル包絡は7kHzで0dBとなる。各正規化された高域包絡に対応する高域エネルギーが次に計算される。高域スペクトル包絡の選択が、高域エネルギーに基づいて区切られる。例えば、0.7dB異なる標準エネルギー値の系列が、全体の範囲をカバーするように選択され、標準値の0.35dB以内のエネルギーの全ての包絡がグループ化される。このように形作られた各グループに対し、平均の高域スペクトル包絡形状が計算され、続いて対応する高域エネルギーが計算される。このように形作られた64の高域スペクトル包絡形状のセットから、1番目、11番目、22番目、33番目、44番目、54番目、64番目の形状(以下、予め計算された形状)がルックアップテーブルに保存される。残りの57形状は保存を必要とせず、直近の予め計算された形状の間での単純な線形補間(dB領域で)により得られる。これらの形状のエネルギーは、1番目の形状に対する約8.5dBから64番目の形状に対する約50.5dBまで、平均のエネルギー分解能約0.67dBで変化する。フレームに対する高域エネルギーを考えると、最も一致する高域スペクトル包絡形状を選択することは簡単な問題である。この高域包絡形状選択の方法は、高域スペクトル包絡形状の小さな変化に対応する高域エネルギーの小さな変化を保証する。これは、高域エネルギーの時間変化の制御により、高域スペクトル包絡形状の時間変化の明確な制御を可能とする。少なくとも明らかな音声セグメントにおいては、高域スペクトルの滑らかな変化は、自然に聴こえる高品質出力SWB音声を保証するために重要である。

### E. 7. 8. 1. 6 高域MDCT生成器

高域励振信号スペクトル $HB\ Exc$ と高域スペクトル包絡 $HB\ Env$ は、HB MDCT係数を形成するために結合される。高域励振信号スペクトルは、MDCT係数が平均エネルギー1.0を持つように最初に正規化される。その後、高域スペクトル包絡は線形領域に変換され、MDCTスペクトルを形成するために励振信号スペクトルからの対応する係数が乗じられる。高域信号成分は、7-14kHz帯域に対応するものを除き全ての係数を初期化し逆MDCTを適用することにより得られる。その後、出力SWB音声を生成するために、アップサンプリン



グされた広帯域音声に高域信号成分が加算される。

### E. 7. 8. 2 スタンバイ処理

このモードは、7-14kHz 成分周波数が符号化された信号に存在するとき、

- ・アクティブモードに切り戻す必要がある場合にエネルギー予測を維持するため
- ・前アクティブモードからの残差データの逆 MDCT 合成バッファをフラッシュするため

に用いられる。

スタンバイ処理の間、広帯域音声エネルギー  $E_{wb}$  が各復号フレームに対して計算され、これは、信号レベル及び背景雑音レベルの局所推定を更新するために用いられる。前述の遷移帯域エネルギー  $E_b$  は、同様に各復号フレームに対して計算される。 $E_{wb}$  と共にこの値は、アクティブ処理が次の復号フレームで再開される場合上記の立ち上がり/破裂音検出スキームを可能とするために必要である。前フレームの 7000-14000Hz 帯域が前項記載のアルゴリズムで合成される事象において、必要に応じて逆 MDCT オーバーラップ加算バッファも同様にこのスタンバイ処理の間フラッシュされる。

### E. 7. 9 時間領域への信号変換

合成 MDCT 領域信号  $\hat{M}_{32}(k)$  が得られれば、時間領域に変換することができる。この変換の前に、7kHz 未満と 14kHz 以上を超える以下の周波数はゼロにセットされる。

$$\hat{M}_{32}(k) = \begin{cases} 0 & k = 0, \dots, 279 \\ \hat{M}_{32}(k) & k = 280, \dots, 559 \\ 0 & k = 560, \dots, 639 \end{cases} \quad (\text{E-174})$$

これに加え、14kHz より低い最大の周波数はわずかに減衰される。

$$\begin{aligned} \hat{M}_{32}(556) &= 0.5 \hat{M}_{32}(556) \\ \hat{M}_{32}(557) &= 0.25 \hat{M}_{32}(557) \\ \hat{M}_{32}(558) &= 0.125 \hat{M}_{32}(558) \\ \hat{M}_{32}(559) &= 0.0625 \hat{M}_{32}(559) \end{aligned} \quad (\text{E-175})$$

逆変換は次のステップを含む。

1) 復号スペクトル係数の回転

$$u(k) = \hat{M}_{32}(2k) + j\hat{M}_{32}(\frac{N}{2}-1-2k) \quad k = 0, \dots, \frac{N}{4}-1 \quad (\text{E-176})$$

2) 事前回転

フォワード変換(E.6.3 節)における step 2) のように、事前回転が回転データ  $u(k)$  に対して行なわれる。

$$s(k) = c \cdot W_N^{k+0.5} u(k) \quad k = 0, \dots, \frac{N}{4}-1 \quad (\text{E-177})$$

ただし、 $c$  は回転因子  $W_N^{k+0.5}$  に対応するスカラーであり、簡単のため  $\frac{\sqrt{2}}{\sqrt[4]{N}}$  とし、これはフォワード変換のステップ 2 と類似している。格納されたテーブルもここで再使用される。

3) N/4 点 DFT

この部分はフォワード変換(E.6.3節)のステップ3)と同じである。

$$r(n) = \sum_{k=0}^{N/4-1} s(k) W_N^{nk} \quad n=0, \dots, N/4-1 \quad (\text{E-178})$$

4) 事後回転

$$w(n) = r(n) \cdot d \cdot W_N^{n+0.5} \quad n=0, \dots, N/4-1 \quad (\text{E-179})$$

事後回転後のデータは

$$y(n) = \begin{cases} B \cdot (-j) \cdot w(n) & n=0, \dots, N/8-1 \\ B \cdot w(n) & n=N/8, \dots, N/4-1 \end{cases} \quad (\text{E-180})$$

ただし  $B$  は正規化係数,  $d$  は回転因子  $W_N^{n+0.5}$  のスカラ

フォワード変換と同様に,  $d$  は  $\frac{\sqrt{2}}{\sqrt[4]{N}}$  と設定される。、 $B$  は回転因子に統合され、簡単のため1に設定される。

5) 窓かけと回転による合成信号  $\hat{s}_{32\_SWB}(n)$

$$\begin{cases} \hat{s}_{32\_SWB}(2n) = m(2n)g(N/2+2n) + \text{Re}\{u(n)\}g(2n) \\ \hat{s}_{32\_SWB}(N/2-1-2n) = m(N/2-1-2n)g(N-1-2n) - \text{Re}\{u(n)\}g(N/2-1-2n) \end{cases} \quad n=0, \dots, N/4-1 \quad (\text{E-181})$$

ただし

$$\begin{cases} u(n) = y(n + N/8) \\ u(n + N/8) = y(n) \end{cases} \quad n=0, \dots, N/8-1 \quad (\text{E-182})$$

$g(n)$  は合成窓で,  $m(n)$  はバッファリングされた一時データである。

$$\begin{cases} m(N/4-1-2n) = m(N/4+2n) = \text{Im}\{u(n)\} & n=0, \dots, N/8-1 \\ m(2n - N/4) = m(3N/4-1-2n) = \text{Im}\{u(n)\} & n=N/8, \dots, N/4-1 \end{cases} \quad (\text{E-183})$$

変換と逆変換で次の関係が保たれている必要がある。

$$A \cdot a \cdot b \cdot B \cdot c \cdot d = 4/N \quad (\text{E-184})$$

これは次のようにすることで可能となる。

$$a = b = c = d = \frac{\sqrt{2}}{\sqrt[4]{N}}; \quad A = B = 1 \quad (\text{E-185})$$

これにより、定数 a、b、c、d、A、B は ROM テーブルに格納された回転因子へ統合される。逆変換の出力は合成 SWB 信号  $\hat{s}_{32\_SWB}(t)$  となる。

### E. 7. 9. 1 プリエコー減衰

信号エネルギーが音声立ち上がりやパーカッション音楽のように急に変化する場合、変換符号化における典型的な異音として知られるプリエコーが観察される。プリエコーの原因は次のように説明される。周波数領域中の量子化雑音は逆 MDCT および加算/オーバーラップ演算によって時間領域に変換される。したがって、量子化雑音は、MDCT 合成窓で様に広げられる。立ち上がりの場合には、立ち上がりに先行する入力信号の部分が、しばしば非常に低いエネルギーを立ち上がり部分のエネルギーに対して持つ。量子化雑音レベルはフレームの平均エネルギーに依存するので、全体の合成窓で高くなりえる。この場合、信号対雑音比は低エネルギー部分において非常に低く(しばしば負に)なる。この人工的な信号がプリエコーと呼ばれ、量子化雑音が立ち上がりの前に聞こえる。プリエコー異音を防ぐために、合成ウィンドウのある部分に大きなエネルギー増加(立ち上がり)がある場合、これを減衰する必要がある。また、立ち上がりに先行する合成ウィンドウの低いエネルギー部分でプリエコー抑圧を適用する必要がある。下記では、立ち上がりに先行するこの低いエネルギー部分を、「エコーゾーン」と呼ぶことにする。

エコー/非エコーゾーンの識別は逆 MDCT に基づいて決定される。また、プリエコー抑圧はエコーゾーンで行われる。

エコー/非エコーゾーンの識別手順

エコーゾーン/非エコーゾーン間の識別手順は、逆 MDCT の 2 つの信号に基づく。すなわち、合成 SWB 信号  $\hat{s}_{32\_SWB}(n)$ 、および更新された窓かけなしで記憶されている逆 MDCT  $m(n)$  であり、これらは次のフレームで、SWB 信号を合成するためのオーバーラップ加算において使われる。 $m(n)$  が対称的であることに注意が必要である。

まず、信号  $\hat{s}_{32\_SWB}(n)$  および  $m(n)$  は、現在の MDCT 合成ウィンドウの長さ 1280(40ms)の補助信号  $d_{32\_SWB}^{conc}(n)$  を形成するために連結される。

$$\begin{aligned} d_{32\_SWB}^{conc}(n) &= \hat{s}_{32\_SWB}(n) & n = 0, \dots, 639 \\ d_{32\_SWB}^{conc}(n+640) &= m(n) \end{aligned} \quad (E-186)$$

連結された信号は、40 点(1.25ms)の 32 のサブフレームに分割される。また、この信号のテンポラル時間領域の包絡  $Es_{MDCT}(i)$  が連続するサブフレームエネルギーとして計算される。信号  $m(n)$  の対称性により最初の 24 のサブフレームエネルギーだけが異なる(計算される)ことに注意が必要である。

$$Es_{MDCT}(i) = \sum_{n=40i}^{40(i+1)-1} [d_{32\_SWB}^{conc}(n)]^2 \quad i = 0, \dots, 23 \quad (E-187)$$

ここでは、インデックス  $n$  はサンプル数に使用される。また、インデックス  $i$  はサブフレームに使用される。最大のエネルギーを持つサブフレームも  $Max_{Es} = \max_{i=0, \dots, 23} Es_{MDCT}(i)$  として探索される。高いエネルギー部分への遷移はサブフレームで検出され、その指標  $Maxind_{Es}$  は次の通りである。

$$Maxind_{Es} = \min \left( \arg \max_{i=0, \dots, 23} (Es_{MDCT}(i)), 16 \right) \quad (E-188)$$

最大エネルギーのサブフレームが将来のフレーム(連結した信号のメモリ部)に存在する場合、 $MaxindEs$  はメモリ部分の最初で 16 にセットされることに注意が必要である。これはさらなる計算量削減のために行われる。すなわち遷移サブフレームが将来のフレームにある場合、その正確な位置は現在のフレームにおいて重要ではない。最大のエネルギー  $Max_{Es}$  は前のサブフレームのそれと比較される。

$$r_{Es}(i) = \frac{Max_{Es}}{Es_{MDCT}(i)} \quad i = 0, \dots, Maxind_{Es} - 1 \quad (E-189)$$

遷移サブフレームに先行するサブフレームで  $r_{Es}(i) > 8$  となる場合、エコーゾーンとして決定される。プリエコー減衰は、合成 SWB 信号  $\hat{s}_{32\_SWB}(n)$  に減衰関数  $g_{pre}(n)$  をかけることにより作られる。エコーゾーンサブフレームのために、最初のプリエコー減衰利得関数  $g_{pre}'(n)$  が、連結した信号  $d_{32\_SWB}^{conc}(n)$  のスペクトル包絡の関数として決定される、すなわち、サブフレームの各サンプルについて、この利得は  $r_{Es}(i) > 16$  の場合 0.2 に、そうでない場合は 0.5 にセットされる。

非エコーゾーンについては、最初の利得は 1 にセットされる。しかしながら、最初のプリエコー減衰利得は、さらに前フレームの構造のエネルギーに依存する、すなわち、各エコーゾーンサブフレームの最小の利得も、背景雑音エネルギーを保存するために減衰されたサブフレームエネルギーが前フレームの構造のエネルギーほど低くなりえないように決定される。

$$g_{pre}'(n) = \begin{cases} \max\left(0.2, \min\left(\sqrt{\frac{Es_{prev}}{Es_{MDCT}(i)}}, 1\right)\right) & \text{if } r_{Es}(i) > 16 \\ \max\left(0.5, \min\left(\sqrt{\frac{Es_{prev}}{Es_{MDCT}(i)}}, 1\right)\right) & \text{if } r_{Es}(i) > 8 \\ 1 & \text{if } r_{Es}(i) \leq 8 \text{ or } n \geq 40Maxind_{Es} \end{cases} \quad (E-190)$$

with:  $i = 0, \dots, Maxind_{Es} - 1; \quad n = 40i, \dots, 40i + 39$

ただし、 $Es_{prev}$  は前フレームで次のように求められる。

$$Es_{prev} = \sum_{i=0}^{15} Es_{MDCT}(i) g_{pre}'(40i) \quad (E-191)$$

最終的なプリエコー減衰利得  $g_{pre}(n)$  は初期プリエコー減衰利得  $g_{pre}'(n)$  を平滑化することによって得られる。

$$g_{pre}(0) = 0.85g_{pre\_old} + 0.15g_{pre}'(0) \quad (E-192)$$

$$g_{pre}(n) = 0.85g_{pre}(n-1) + 0.15g_{pre}'(n) \quad n = 1, \dots, 639$$

ただし、 $g_{pre\_old}$  は前のプリエコー減衰利得で前フレームで適用されたものである。値は次フレームのため次のように更新される。

$$g_{pre\_old} = g_{pre}(639) \quad (E-193)$$

最終的なプリエコーは減衰された合成 SWB 信号  $\hat{s}_{32\_SWB}'(n)$  であり、合成 SWB 信号  $\hat{s}_{32\_SWB}(n)$  の  $g_{pre}(n)$  による重みづけとして次のように表される。

$$\hat{s}_{32\_SWB}'(n) = g_{pre}(n)\hat{s}_{32\_SWB}(n) \quad (\text{E-194})$$

## E. 7. 10 ミュージックエンハンスメント

### E. 7. 10. 1 信号分類

ミュージックエンハンサーの目的は、音声の質に影響せずに、許容可能な水準の音楽の品質を得るため、符号化および復号過程に作成された音に含まれるトーン間の量子化雑音を減らすことである。そのために、このモジュールが音楽においては中の準最適に、一方で音声はほぼ透過となるように信号タイプ分別器が設計され、調整される。

Figure E.5/G.729.1 のブロック図は、ミュージックエンハンサー内の分別器の構造を示す。

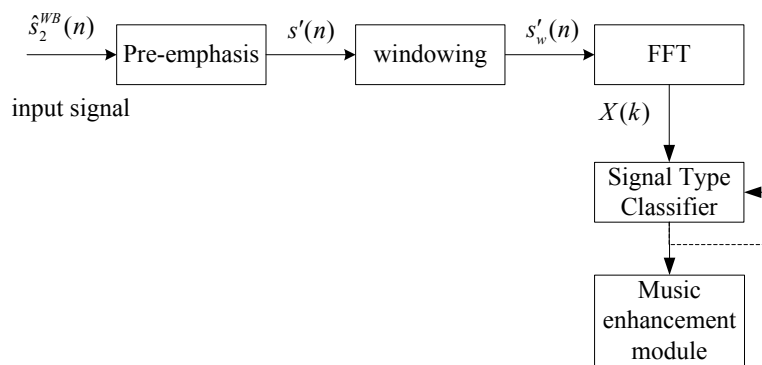


Figure E.5/G.729.1 – Overview of signal classification and music enhancement

### E. 7. 10. 1. 1 信号分析

#### E. 7. 10. 1. 1. 1 プリエンファシス

スペクトル分析を行なう前に、JT-G729.1 復号器(E.7.1 節)の関数記述によって得られた WB 合成信号  $\hat{s}_2^{WB}(n)$  は、最初に、プリエンファシスされる。プリエンファシスでは、1 次の高域通過フィルタが、高域周波数を強調するために使用される。

$$H_{pre-emph}(z) = 1 - 0.68z^{-1} \quad (\text{E-195})$$

プリエンファシスされた信号は、以降  $s'(n)$  と表記する。

#### E. 7. 10. 1. 1. 2 スペクトル分析

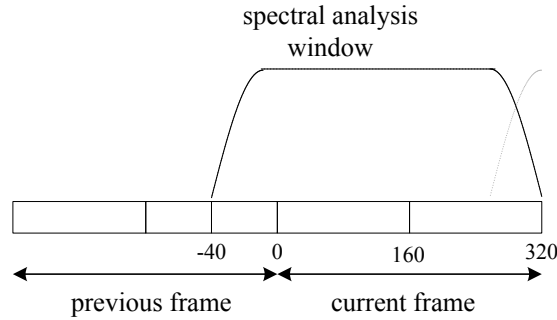


Figure E.6/G.729.1 – Spectral analysis window for music enhancement

離散フーリエ変換(DFT)は信号を周波数領域に変換するために使用される。スペクトル分析は、1/9(Figure E.6/G.729.1 の中で図示)のオーバーラップと共に 256 ポイントの高速フーリエ変換(FFT)を使用して、1 フレーム当たり一度行なわれる。360 サンプルの分析ウィンドウは現フレームおよび過去合成信号の 40 サンプルをカバーする。用いられる窓は、次のように中央付近が平らで端が正弦波関数となる。

$$w_{FFT}(n) = \begin{cases} \sin\left(\frac{\pi n}{2L_{window}/9}\right) & n = 0, \dots, L_{window}/9 - 1 \\ 1 & n = L_{window}/9, \dots, 8L_{window}/9 - 1 \\ \sin\left(\frac{\pi(n - L_{window}/9)}{2L_{window}/9}\right) & n = 8L_{window}/9, \dots, L_{window} - 1 \end{cases} \quad (\text{E-196})$$

ただし、 $L_{window} = 360$  は窓サイズである。FFT は 512 点であることに注意が必要である。つまり、窓かけされた信号は 152 点のゼロづめがされる。トーン間の雑音低下に使用される窓かけ信号は次のように得られる。

$$s'_w(n) = \begin{cases} w_{FFT}(n)s'(n - 40) & n = 0, \dots, L_{window} - 1 \\ 0 & n = L_{window}, \dots, 512 \end{cases} \quad (\text{E-197})$$

スペクトルパラメータを一組得るために窓かけ信号に FFT が適用される。

$$X(k) = \sum_{n=0}^{511} s'_w(n) e^{-j2\pi \frac{kn}{512}} \quad k = 0, \dots, 512 \quad (\text{E-198})$$

FFT によって得られるスペクトルの実部と虚部はそれぞれ、 $\text{Re}(X(k))$ ,  $k = 0, \dots, 256$ ,  $\text{Im}(X(k))$ ,  $k = 1, \dots, 255$  と表される。 $\text{Re}(X(0))$  は 0 Hz (直流成分) に対応するスペクトルであり、 $\text{Re}(X(128))$  は 8000 Hz のスペクトルに対応する。これらのスペクトルは実部のみであり、以降の分析では通常無視される。

FFT 分析の後、スペクトルは 臨界帯域に分割され、これには Table E.16/G.729.1 に示すような上限を持つインターバル(0-8000 Hz の範囲で 21 帯域)を用いて行う。

Table E.16/G.729.1 – Critical band upper limits

Upper limit of the band (Hz)				
100	200	300	400	510
630	770	920	1080	1270
1480	1720	2000	2320	2700
3150	3700	4400	5300	6700
8000				

512 点の FFT の周波数分解能は 31.25 Hz (8000/256) である。スペクトルの直流成分を除いた後のそれぞれの臨界帯域の周波数ビン数は  $M_{CB} = \{3, 3, 3, 3, 3, 4, 5, 4, 5, 6, 7, 7, 9, 10, 12, 14, 17, 22, 28, 44, 41\}$  である、ただし、周波数分解能は 32Hz で近似している。

### E. 7. 10. 1. 1. 3 エネルギー分析

エネルギー分析モジュールは、いくつかのエネルギーに関連するパラメータを計算する。周波数ビンごとのエネルギー  $E_{BIN}(k)$  は、次のように計算される。

$$E_{BIN}(k) = \text{Re}(X(k))^2 + \text{Im}(X(k))^2 \quad k = 0, \dots, 255 \quad (\text{E-199})$$

臨界帯域内の平均エネルギーは次のように計算される。

$$E_{CB}(i) = \frac{1}{\left(\frac{L_{FFT}}{2}\right)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} (\text{Re}(X(k+j_i))^2 + \text{Im}(X(k+j_i))^2) \quad i = 0, \dots, 20 \quad (\text{E-200})$$

ただし、 $X(k)$  は  $k$  番目の周波数ビンで、 $j_i$  は  $i$  番目の臨界帯域内の最初のビンのインデックスである。この場合  $i$  番目の臨界帯域内の最初のビンは次のように与えられる。

$$j_i = \{1, 4, 7, 10, 13, 16, 20, 25, 29, 34, 40, 47, 54, 63, 73, 85, 99, 116, 138, 166, 210\} \quad (\text{E-201})$$

最後に、スペクトル分析モジュールは、平均臨界帯域エネルギー  $E_{CB}$  の足し合わせにより 20ms フレーム内の FFT 分析ごとの全エネルギーを次のように計算する。

$$E_i = 10 \log \sum_{i=0}^{20} E_{CB}(i) \quad (\text{E-202})$$

スペクトル分析モジュールの出力パラメータ(つまり臨界帯域ごとのエネルギー)、周波数ビンごとのエネルギーおよび全エネルギーは、トーン間雑音低減アルゴリズムおよび利得修正モジュールにおいて使用される。

### E. 7. 10. 1. 2 信号種別分類

信号種別分類モジュールは、どの音がトーン間雑音低減アルゴリズム(例えば調波性音楽)によく適しているか、また、どの音がそうではないか(例えば音声)を特定することにより、トーン間雑音低減器の効率をより最大化するように設計されている。信号種別分類モジュールは、音声の劣化の可能性を最小限にするため音響信号をカテゴリに分類しトーン間雑音抑圧低減モジュールに情報を付与する。信号種別分類モジュールは可能な限り単純に保たれる。モジュールへの主要な入力全フレームエネルギー  $E_i$  で式 E-202 のように示

される。

まず、全フレームエネルギー変動の過去 40 個の値の平均は次のようになる。

$$\bar{E}_\Delta = \frac{1}{40} \sum_{i=-40}^{-1} E_\Delta^{[i]} \quad (\text{E-203})$$

ただし、

$$E_\Delta^{[i]} = E_t^{[i]} - E_t^{[i-1]} \quad \text{for } i = -40, \dots, -1 \quad (\text{E-204})$$

上付きの  $i$  は過去のフレームを表す。そして、全エネルギー変動の過去 15 個の値と 40 個の値の平均の統計的な偏差が計算される。すなわち、

$$E_{dev} = 0.7745967 \sqrt{\frac{\sum_{i=-15}^{-1} (E_\Delta^{[i]} - \bar{E}_\Delta)^2}{15}} \quad (\text{E-205})$$

得られたエネルギーの偏差は、トーン間ノイズ抑圧低減の特定のフレームに対する効率を決定するため、4 つの閾値と比較される。信号種別分類モジュールの出力は、5 つのカテゴリの中の 1 つに対応する 0-45 のインデックスで、それぞれはトーン間雑音低減の設定に関連づけられている。最初の種別 (カテゴリ 0) は音声のような非トーン性の信号に対応し、トーン間雑音低減アルゴリズムの影響を受けない。この種別の音響信号には大きな統計的な偏差が一般にある。3 つの中間のカテゴリ (1 から 3) は異なる種別の統計的な偏差の音を含んでいる。カテゴリ 1 はカテゴリ 0 に次いで最も大きなエネルギー偏差を示し、2000 から 8000 Hz の帯域のトーン間雑音低減を許容し、最大許容抑圧は 6dB である。カテゴリ 2 は 1270 から 8000 Hz の雑音低減を許容し、最大 9dB である、またカテゴリ 3 は 770 から 8000 Hz の帯域の雑音低減を許容し、最大 12 dB である。最後のカテゴリ (カテゴリ 4) は最小の統計的な偏差を示す音を含んでいる。この場合、トーン間雑音低減アルゴリズムは、400 から 8000 Hz の帯域の品質向上を最大許容低減抑圧 12dB で行うことを許容されている。トーン間雑音低減カテゴリは表 E.17 のようにまとめられる。

Table E.17/G.729.1 – Inter-tone noise reduction categories

Category	Enhanced band (Hz)	Allowed reduction (dB)
0	N/A	0
1	[2000, 8000]	6
2	[1270, 8000]	9
3	[770, 8000]	12
4	[400, 8000]	12

分類間違いを防ぐため閾値は適応的である。典型的には、音楽のようなトーン性の音は、声のような非トーン性信号よりもはるかに低い統計的な偏差を示す。しかし、音楽でも高い統計的な偏差を含むことがあり得、また、同様に、音声も低い統計的な偏差を含むこともあり得る。フレーム内で音声と音楽が変化することはあまりない。

連続するカテゴリの 2 つのカウンタがそれぞれの閾値を上げ下げするのに用いられる。最初のカウンタ



は、カテゴリ 3 か 4 が選ばれた場合フレーム単位で増加される。このカウンタはカテゴリ 0 が選ばれるとゼロにセットされ、その他の場合変わらない。もう一方のカウンタは逆の効果を持っている。カテゴリ 0 が選ばれた場合増加され、カテゴリ 3 か 4 が選ばれた場合ゼロにセットされ、それ以外では変わらない。両カウンタの初期値はゼロである。カテゴリ 3 かカテゴリ 4 のカウンタが 30 になると、より多くのフレームがカテゴリ 4 になるように、すべての閾値が 0.15625 だけ増加される。一方で、カテゴリ 0 のカウンタが 30 になると、より多くのフレームがカテゴリ 0 に分類されるよう全ての閾値が 0.15625 だけ下げられる。

音響種別分類が固定のカテゴリにロックしないように閾値は最大値と最小値で制限される。閾値の初期値、最小値、最大値の値を以下に示す。

$$\begin{array}{lll}
 M^{[0]} = 2.5 & M_{\min}^{[0]} = 1.875 & M_{\max}^{[0]} = 3.125 \\
 M^{[1]} = 1.875 & M_{\min}^{[1]} = 1.25 & M_{\max}^{[1]} = 2.8125 \\
 M^{[2]} = 1.5625 & M_{\min}^{[2]} = 0.9375 & M_{\max}^{[2]} = 2.1875 \\
 M^{[3]} = 1.3125 & M_{\min}^{[3]} = 0.625 & M_{\max}^{[3]} = 1.875
 \end{array}$$

ただし、上付き  $[j] = 0, \dots, 3$  はカテゴリ  $j$  を表す。

カテゴリは、統計的な偏差の計算値、 $E_{dev}$  と 4 つの閾値との比較に基づいて選択される。選択アルゴリズムを次に示す。

```

if ( $E_{dev} < M^{[3]}$ ) AND (Categoryprev  $\geq 3$ )
    select Category 4
else if ( $E_{dev} < M^{[2]}$ ) AND (Categoryprev  $\geq 2$ )
    select Category 3
else if ( $E_{dev} < M^{[1]}$ ) AND (Categoryprev  $\geq 1$ )
    select Category 2
else if  $E_{dev} < M^{[0]}$ 
    select Category 1
else
    select Category 0

```

(E-206)

フレーム消失の場合、全ての閾値は最小値にリセットされ、消失フレーム後の連続 2 フレーム(消失フレームを含めて 3 フレーム)は、分類結果は強制的にカテゴリ 0 にセットされる、その期間は消失フレーム後の連続 2 フレーム(消失フレームを含めて 3 フレーム)である。音響活性度検出器 (SAD)からの情報が使え、それが音響活性度ゼロを示すか( $f_{SAD} = 0$ )、もしコーデックが DTX モードでデコードビット列が SID フレームまたは CNG フレームに対応する場合、信号種別分類はカテゴリ 0 になる。

#### E. 7. 10. 2 知覚マスキングに基づく周波数領域後処理

通常の音声信号に対し、CELP の 後処理フィルタは効果的である。しかしながら音楽信号に対しては、周波数領域後処理のほうが品質が上がる。現在の復号器では、周波数範囲 [0-7 kHz] の MDCT 係数が重みつき領域で利用可能で、トータルで 280 係数、 $\hat{M}_{16}(i)$ ,  $i=0,1,\dots,279$  を持っている。これらが時間領域に変換される前に、これらの周波数係数は音楽信号の周波数領域後処理を行うために使用することができる。1.0 付近の適切な利得係数が知覚的に全体的な性能を改善するために各周波数係数に適用される。明らかに、利得係数推定アルゴリズムは周波数領域の後処理の重要な鍵となる部分である。この節では、知覚マスキングの原理に基づく利得係数推定アルゴリズムについて述べる。

知覚重み付けフィルタを使用して、時間領域中の信号を符号化する場合、復号された信号の周波数係数は知覚的に重要な領域では相対的に品質が良いのに対し、知覚に重要ではない領域では相対的に品質が劣化する。同様に、知覚マスキングモデルを使用して符号化器が周波数係数を量子化する場合、復号された周波数係数の知覚的な品質は等しく(一様に)分配されない。十分な品質がある周波数は、1 よりわずかに大きな利

得係数を乗じることで振幅を増すことができる。一方、品質が悪い周波数には、1未満の利得係数を掛けるか、推定マスキング閾値より低いレベルに下げることができる。品質を判断するために、3パラメータが定義される。それぞれ局所マスキング振幅  $M_0(i)$ 、局所被マスキング振幅  $M_1(i)$ 、全体平均振幅  $M_{av}$  である。3パラメータは周波数係数を使用して推定され、特に  $M_0(i)$  および  $M_1(i)$  の推定は知覚マスキング効果に基づく。

原理上、周波数トーンがマスキングトーンとして働く場合、マスキングトーンは、下側エリアより上側エリアにより多くの影響を及ぼす。つまり、マスキングトーンの影響範囲は、高い周波数の方が、低い周波数よりも大きい。通常、実際の信号は単なるトーンから構成されない。スペクトルエネルギーが関連する帯域に存在する場合、ある周波数位置  $i$  の"知覚ラウドネス"は位置  $i$  のエネルギーだけでなくその位置のまわりのエネルギー分配にも依存する。局所マスキング振幅  $M_0(i)$  は位置  $i$  の"知覚ラウドネス"としてみなされ、その周辺のスペクトル振幅の重み付き和をとることで次のように推定される。

$$M_0(i) = \sum_k w_0^i(k) \cdot |F_0(i+k)| \quad (\text{E-207})$$

ここで、 $F_0(i) = \hat{M}_{16}(i)$  は後処理前の周波数係数を表す。重みづけ窓  $w_0^i(k)$  は非対称であり、窓の端部は  $i$  の右側より左側の方が長く、窓サイズの合計は低い周波数エリアより高い周波数エリアの方が大きい。理論上、窓はすべての  $i$  で異なるべきだが、簡単のため、窓は、20個の周波数係数の小さな間隔に対して同じに設定されている。すべての窓係数は事前に計算され、正規化され、テーブルに保存される。局所被マスク振幅、 $M_1(i)$ 、は局所的な"知覚誤差下限"と考えられる。符号化器は知覚領域で信号を符号化するため、復号器側での高いエネルギー周波数係数は通常、相対誤差が低く、絶対誤差が高い。反対に、復号器側での低いエネルギー周波数係数は相対誤差が高く絶対誤差が低い。異なる周波数の誤差も、通常の信号のマスキング効果と同様に、知覚的にお互いに影響しあう。要するに、局所被マスク振幅  $M_1(i)$  は  $M_0(i)$  と同様に次のように推定される。

$$M_1(i) = \sum_k w_1^i(k) \cdot |F_0(i+k)| \quad (\text{E-208})$$

しかしながら重みづけ窓  $w_1^i(k)$  の形は、 $w_0^i(k)$  よりも平らで長い。 $w_0^i(k)$  と同様に、窓  $w_1^i(k)$  は理論的には  $i$  毎に異なるが、簡単のため窓は20個の周波数係数の間隔に対して同じに設定される。全ての窓係数は事前に計算され、正規化され、テーブルに保存される。比率  $M_0(i)/M_1(i)$  は位置  $i$  での局所的な相対知覚品質を反映することができる。全体平均振幅の影響を考えれば、周波数に沿った利得係数の推定を初期化する1つの方法は、大きな利得係数を回避するために局所被マスク振幅を全体平均振幅と比較することである。それは次のように表される。

$$Gain_0(i) = \frac{M_0(i)}{\alpha \cdot M_1(i) + (1-\alpha) \cdot M_{av}} \quad (\text{E-209})$$

ただし  $\alpha=15/16$  は1に近い値で、平均エネルギー振幅の影響を制御するために用いられる。

$$M_{av} = \sum_i |F_0(i)| / N_F \quad (\text{E-210})$$

$N_F=280$  は周波数係数の総数である。後処理の後の全体エネルギーの大きな変化を避けるため、後処理の後、

利得正規化が適用される。スペクトル帯域全体は、4 サブバンドに分割され、係数  $Norm$  を乗じることで、70 個の係数のそれぞれのサブバンドに利得正規化が適用される。

$$Gain_1(i) = Gain_0(i) \cdot Norm \quad (E-211)$$

各サブバンドの正規化係数は次のように定義される。

$$Norm = \sqrt{\frac{\sum_i |F_0(i)|^2}{\sum_i |Gain_0(i) \cdot F_0(i)|^2}} \quad (E-212)$$

式 E-211 で推定された利得係数は、強い後処理が必要であることを仮定している。実際の応用では、単に弱い、あるいは、後処理なしが復号信号品質によって必要となる時もある。そこで、後処理の全体制御が、制御パラメータ  $\beta$  ( $0 \leq \beta \leq 1$ ) により導入され、 $\beta = 0$  で後処理なし、 $\beta = 1$  で完全後処理とする。 $\beta$  のセッティングは、E.7.10.1.2 に述べられていた信号種別分類の出力による。

```

if (Category=0) { //Speech
     $\beta = 0$ ;
}
else if (Category<3) {
     $\beta = 0.5 \beta_0$ ;
}
else if (Category=4) { //music
     $\beta = 1.1 \beta_0$ ;
}

```

$\beta_0$  はおよそ 0.5 の定数であり、カテゴリ決定アルゴリズムは E.7.10.1.2 に記載されている。 $\beta$  は次のように若干縮小される。

```

if (Sharpness>0.18 or Voicing>0.8) {
     $\beta \leftarrow 0.4 \beta$ ;
}
else if (Sharpness>0.17 or Voicing>0.7) {
     $\beta \leftarrow 0.5 \beta$ ;
}
else if (Sharpness>0.16 or Voicing>0.6) {
     $\beta \leftarrow 0.65 \beta$ ;
}
else if (Sharpness>0.15 or Voicing>0.5) {
     $\beta \leftarrow 0.8 \beta$ ;
}

```

$Voicing$  は正規化された CELP の有声係数をスムージングしたものである。

$$Voicing \leftarrow 0.5 Voicing + 0.5 G_p$$

$$G_p = E_p / (E_p + E_c)$$

$Sharpness$  は、周波数サブバンドにおいて、平均振幅とピーク振幅の間の比率として定義されたスペクトル先鋭パラメータである。典型的な音楽信号では、 $Sharpness$  と  $Voicing$  の値は小さい。 $Voicing$  が 1 に近い時、

CELP コーデック線形予測符号化がうまく機能していることを意味する。*Sharpness* が大きい時、復号信号のスペクトルが雑音的な可能性がある。現在と過去の制御パラメータによるスムージング処理が次のように追加される。すなわち、 $\bar{\beta} \leftarrow 0.75 \bar{\beta} + 0.25 \beta$  である。利得係数は平滑化された制御パラメータにより次のように調整される。

$$Gain_2(i) = \bar{\beta} \cdot Gain_1(i) + (1 - \bar{\beta}) \quad (E-213)$$

調整された利得係数はさらに現在と過去の利得係数により次のように平滑化される

$$\overline{Gain}(i) = 0.25 \overline{Gain}(i) + 0.75 Gain_2(i) \quad (E-214)$$

最後に、決定された利得係数は周波数係数  $F_0(i) = \hat{M}_{16}(i)$  に乗じられ、後処理された周波数係数  $F_1(i)$  が次のように得られる。

$$F_1(i) = F_0(i) \cdot \overline{Gain}(i) \quad i = 0, 1, 2, \dots, 279 \quad (E-215)$$

## E. 7. 1 1 後処理

### E. 7. 1 1. 1 復号器における短ピッチラグ補正

調波性音楽信号または歌声の信号のピッチラグが、CELP アルゴリズム (JT-G729 3.7 節参照) で定義された最小ピッチラグ限界  $P\_MIN$  よりも小さい場合、伝送されるピッチラグは実際のピッチラグの 2 倍、3 倍、またはその他の倍数になりうる。その結果、伝送されたピッチラグでピッチポストフィルタされた信号のスペクトルは、知覚される歪が生じ、ポストフィルタなしの場合よりも知覚的に悪く聞こえる場合がある。下記方法を用いて、ピッチ後処理を行う前に実際の調波ピークを強調するため CELP 復号器でピッチラグを修正する。まず、CELP 出力信号のピッチ修正が伝送されたピッチラグの周辺、および伝送されたピッチラグの  $1/2, 1/3, 1/m(m>3)$  の周辺で推定される。

$$R(P) = \frac{\sum_n \hat{s}(n) \cdot \hat{s}(n-P)}{\sqrt{\sum_n \|\hat{s}(n)\|^2 \cdot \sum_n \|\hat{s}(n-P)\|^2}} \quad (E-216)$$

ただし、 $R(P)$  は伝送されたピッチラグ  $P$  の正規化ピッチ相関である。信号  $\hat{s}(n)$  はカットオフ 4kHz でローパスフィルタされた信号で相関推定の安定性を高める。式 E-216 の平方根を避けるため、相関は  $R^2(P)$  で表現され負の  $R(P)$  はゼロとする。すなわち  $R(P) \leftarrow R^2(P)$  とする。相関値最大化の探索の最中に式 E-216 の分母は、さらなる演算量削減のため 1 に固定される。フレームごとに、第 3 または第 4 サブフレームの伝送されたピッチラグがピッチ相関の推定と比較を行うためのフレーム代表として選ばれる。 $P$  が伝送されたピッチラグ、 $P_2$  は  $P/2$  付近で選ばれた整数値であって、相関値  $R(P_2)$  を最大化し、 $P_3$  は  $P/3$  付近で選ばれた整数値で相関値  $R(P_3)$  を最大化し、 $P_m$  は  $P/m$  付近で選ばれた整数値で相関値  $R(P_m)$  を最大化する場合、 $R(P_2)$  または  $R(P_m)$  が  $R(P)$  に比べて十分大きい場合、かつ、この現象が所定の時間より長くまたは 1 フレームよりも多くのフレームで起きる場合、ピッチ後処理を行う前に、 $P$  は  $P_2$  または  $P_m$  に置き換えられる。

$$\begin{aligned}
& \text{if} ( R(P_2) > C \cdot R(P) \ \& \ P_2 \approx P_{old} ) & P = P_2 \\
& \quad \vdots \\
& \text{if} ( R(P_m) > C \cdot R(P) \ \& \ P_m \approx P_{old} ) & P = P_m
\end{aligned}$$

ただし  $P_{old}$  は前フレームのピッチ候補で  $P_{MIN}$  よりも小さい。  $P_{old}$  は次のフレームのため以下のように更新される。

$$\begin{aligned}
& \text{initial } P_{old} = P ; \\
& \\
& \text{if} ( R(P_2) > C \cdot R(P) \ \& \ P_2 < P_{MIN} ) , P_{old} = P_2 ; \\
& \quad \vdots \\
& \text{if} ( R(P_m) > C \cdot R(P) \ \& \ P_m < P_{MIN} ) , P_{old} = P_m ;
\end{aligned}$$

ここで  $C$  は 1 より小さい重み係数である。

原理的には、短ピッチラグ ( $<P_{MIN}$ ) 検出はスペクトル範囲  $[0, F_{MIN}]$  のエネルギーが十分小さいかどうかを調べることでより信頼性が高まる。ただし、  $F_{MIN} = F_s / P_{MIN}$  で  $F_s$  は標本化周波数である。復号器を詳細に述べると、  $mdct(i) = \hat{M}_{16}(i)$  が低域信号の MDCT 係数、  $i_{max}$  が  $mdct(i) = \hat{M}_{16}(i)$  のピーク振幅位置 で、  $i=17$  から  $i=17+23=40$  と仮定した場合、

$$|mdct(i_{max})|^2 = MAX\{|mdct(i)|^2 \ i = 17, 18, \dots, 40\} \quad (E-217)$$

2 つのエネルギーを次のように定義し、

$$low\_erg = \sum_{i=0}^{15} |mdct(i)|^2 + \sum_{i=i_{max}+4}^{i=i_{max}+15} |mdct(i)|^2 \quad (E-218)$$

$$max\_erg = |mdct(i_{max})|^2 \quad (E-219)$$

比較の比は次のように定義される。

$$sharp\_ratio = 16 \cdot low\_erg / max\_erg \quad (E-220)$$

上記の比とその平滑化された値が 1 よりかなり小さい場合、短ピッチラグが存在する可能性があり、そうでなければ短ピッチラグはない。

### E. 7. 12 広帯域信号再標本化

合成された 32 kHz 出力信号を作成するため、WB 符号化器の出力  $\hat{s}_{16}(n)$  は 32kHz で再標本化されなければならない。まず、2 ステージの IIR フィルタの入力が次のように得られる。

$$\begin{aligned}
\hat{s}_{IIR}(k) &= \hat{s}_{16}(n) \\
\hat{s}_{IIR}(k+1) &= 0 \quad n = 0, \dots, 319 \quad k = 0, 2, \dots, 638
\end{aligned} \quad (E-221)$$

2 ステージ IIR フィルタの第 1 ステージを以下に示す。

$$\hat{s}_{stage1}(n) = \sum_{i=0}^5 b_{i, stage1} \hat{s}_{IIR}(n-i) - \sum_{j=1}^5 a_{j, stage1} \hat{s}_{stage1}(n-j) \quad n = 0, \dots, 639 \quad (E-222)$$

アップサンプルされた信号  $\hat{s}_{32\_WB}(n)$  を得るための第 2 ステージを以下に示す。

$$\hat{s}_{32\_WB}(n) = \sum_{i=0}^4 b_{i,stage2} \hat{s}_{stage1}(n-i) - \sum_{j=1}^4 a_{j,stage2} \hat{s}_{32\_WB}(n-j) \quad n = 0, \dots, 639 \quad (E-223)$$

ただし、処理信号の負のインデックスは前のフレームのメモリに対応する。フィルタ係数を Table E.18/JT-G729.1 に示す。

Table E.18/JT-G729.1 – Upsampling filter coefficients

$b_{k,stage1}$	$a_{k,stage1}$	$b_{k,stage2}$	$a_{k,stage2}$
0.1163816	1.0	0.0708131	1.0
0.2546400	-1.462269006103	0.1446910	-1.308055753351
0.4015020	2.194832312468	0.1941500	1.509406808301
0.4015020	-1.638502434011	0.1446910	-0.8377333289311
0.2546400	0.926185061962	0.0708131	0.2783957005103
0.1163816	-0.2701781844527		

最終出力は、アップサンプリングされた WB 信号 と SWB 信号を加算し次のようになる。

$$\hat{s}_{32}(n) = \hat{s}_{32\_WB}(n) + \hat{s}_{32\_SWB}(n) \quad (E-224)$$

## E. 8 伝送パラメータインデックスの記述

Table E.19/JT-G729.1 から Table E.28/JT-G729.1 に拡張レイヤごとの伝送パラメータのためのビット割り当てを示す。レイヤ 8mo はエンベディッドビット列のさらなるグラニュラリティを許容できるように設計されている。最初の 80 ビットおよび残りの 80 ビットは、HF スペクトルの 2 つの独立した強調を含む。

Table E.19/JT-G729.1 – Bit allocation for Layer 6mo generic mode

Parameter	Bits, per transmitted index	Bits, total
SWB/Stereo	1	1
Generic/Sinusoidal	1	1
Sub-band lags	8 + 7 + 8 + 7	30
Gain signs	1 + 1 + 1 + 1	4
Gains, 1st scaling	8 + 8	16
Gains, 2nd scaling	8	8
Sinusoidal positions	5 + 5	10
Sinusoidal signs	1	1
Sinusoidal amplitudes	4 + 4	8
Reserved	1	1

Table E.20/JT-G729.1 – Bit allocation for Layer 6mo sinusoidal mode

Parameter	Bits, per transmitted index	Bits, total
SWB/Stereo	1	1
Generic/Sinusoidal	1	1
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 6 + 5	51
Sinusoidal signs	1 + 1 + 1 + 1 + 1 + 1	6
Sinusoidal amplitudes	7 + 7 + 7	21

Table E.21/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 10$ 

Parameter	Bits, per transmitted index	Bits, total
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5	50
Sinusoidal signs	1 + 1 + 1 + 1 + 1	5
Sinusoidal amplitudes	8 + 8 + 8	24
Reserved	1	1

Table E.22/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 8$ 

Parameter	Bits, per transmitted index	Bits, total
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5 + 5 + 5	40
Sinusoidal signs	1 + 1 + 1 + 1	4
Sinusoidal amplitudes	8 + 8	16
WB enhancement/VQ of sub-bands coefficients	9 + 9	18
Reserved	1	1

Table E.23/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 6$ 

Parameter	Bits, per transmitted index	Bits, total
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5	30
Sinusoidal signs	1 + 1 + 1	3
Sinusoidal amplitudes	8 + 8	16
WB enhancement/VQ of sub-bands coefficients	9 + 9 + 9	27
Reserved	1	1

Table E.24/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 4$ 

Parameter	Bits, per transmitted index	Bits, total
Sinusoidal positions	5 + 5 + 5 + 5	20
Sinusoidal signs	1 + 1	2
Sinusoidal amplitudes	8	8
WB enhancement/VQ of sub-bands coefficients	9 + 9 + 9 + 9 + 9	45
Reserved	1	1

Table E.25/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 2$ 

Parameter	Bits, per transmitted index	Bits, total
Sinusoidal positions	5 + 5	10
Sinusoidal signs	1	1
Sinusoidal amplitudes	8	8
WB enhancement/VQ of sub-bands coefficients	9 + 9 + 9 + 9 + 9 + 9	54
Reserved	1	1

Table E.26/JT-G729.1 – Bit allocation for Layer 7mo,  $N_{sin} = 0$ 

Parameter	Bits, per transmitted index	Bits, total
WB enhancement/VQ of sub-bands coefficients	9 + 9 + 9 + 9 + 9 + 9 + 9 + 9	72
Reserved	1	1

Table E.27/JT-G729.1 – Bit allocation for Layer 8mo

Parameter	Bits, per transmitted index	Bits, total
SWB/Stereo	1	1
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5	50
Sinusoidal signs	1 + 1 + 1 + 1 + 1	5
Sinusoidal amplitudes	8 + 8 + 8	24
Reserved: SWB/Stereo	1	1
Sinusoidal positions	5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5	50
Sinusoidal signs	1 + 1 + 1 + 1 + 1	5
Sinusoidal amplitudes	8 + 8 + 8	24

Table E.28/JT-G729.1 – Bit allocation for Layer 9mo and 10mo

Parameter	Bits, per transmitted index	Bits, total
Error spectral envelope	2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 2	36
Error fine structure (VQ of sub-bands coefficients)	$nbits\_err\_VQ = 284$	284

### E. 9 SWB拡張のビットイクザクト記述

JT-G729.1 SWB 拡張をシミュレートする ANSI C コードはこの付録の一部である。このアネックス付録と C コードの記述の間に相違があった場合、C コードの記述が優先するものとする。



参考文献

- [b-ITU-T G.718] Recommendation ITU-T G.718 (2008), Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s.
- [b-Gosset] Gosset, T. (1990), *On the regular and semi-regular figures in space of  $n$  dimensions*, Messenger of Mathematics 29, pp 43-48, Macmillan.
- [b-SWB] Tammi, M., Laaksonen, L., Rämö, A., and Toukoma, H. (2009), *Scalable superwideband extension for wideband coding*, Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing.

付録 用語対照表

(標準 J T - G 7 2 9 . 1 に対する)

英 語	T T C 標準用語
adaptive codebook	適応コードブック
adaptive postfiltering	適応ポストフィルタ処理
algebraic codebook	代数的コードブック
artificial onset reconstruction	擬似立ち上がり再生
binary coding	2進数符号化
bit-exact	ビットイグザクトな
bitstream	ビット列
codebook	コードブック
Code-Excited Linear-Prediction	符号励振線形予測
coder	符号器
codevector	符号ベクトル
concealment	補償
core layer	コアレイヤ
decoder	復号器
differential Huffman coding	差分ハフマン符号化
elliptic	楕円型の
embedded spherical vector quantization	エンベデッド球面ベクトル量子化
embededed	エンベデッド
encoder	符号器
enhancement layer	エンハンスメントレイヤ
envelope	包絡
erasure	消失
excitation	励振信号
excitation signal	励振信号
filter bank	フィルタバンク
fine structure	微細構造
fixed codebook	固定コードブック
fixed-point arithmetic	固定小数点演算
fixed-point mathematical operations	固定小数点算術演算
fold	折り返す
fractional part	分数部
frame erasure concealment	フレーム消失補償
frequency responses	周波数応答
gain	利得
gain attenuation	利得減衰
gain-shape	利得形状
glottal pulse	声門パルス
higher band	高域
inverse filter	逆フィルタ

inverse MDCT	逆 MDCT
layer	レイヤ
leader	リーダー
level-adjustment	レベル調整
linear-predictive filter	線形予測フィルタ
log-energy	対数エネルギー
long-term postfilter	長期ポストフィルタ
long-term post-processin	長期後処理
lookahead	先読み
lower band	低域
LP residual	線形予測残差
MDCT	変形離散コサイン変換
mean-squared error	平均自乗誤差
median filter	メディアンフィルタ
modified discrete cosine transform	変形離散コサイン変換
narrowband	狭帯域
Nyquist frequency	ナイキスト周波数
onset	立ち上がり
overlap-add	重ね合わせ加算
partial ranks	部分階数
perceptual importance	聴覚重要度
perceptual weighting	聴覚重み付け
permutation codes	順列符号
permutation rank	順列階数
pitch multiples	倍数ピッチ
post-echo	ポストエコー
postfilter	ポストフィルタ
post-processing	後処理
pre-echo	プリエコー
prime decomposition	素因数分解
principle of reverse waterfilling	逆注水定理
prototype pulse	プロトタイプパルス
QMF analysis	QMF 分析
QMF synthesis	QMF 合成
quantization error	量子化誤差
residual	残差
root mean square	平均自乗根
sampling	標本化
scalable	スケーラブル
scalar	スカラ
short-term postfilter	短期ポストフィルタ
short-term post-processing	短期後処理
sign	極性

silence detection	無音検出
spectral envelope	スペクトル包絡
subband	サブバンド
superframe	スーパーフレーム
target signal	ターゲット信号
target vector	ターゲットベクトル
tilt compensation	傾き補償
time envelope	時間包絡
Time-Domain Aliasing Cancellation	時間領域折り返し歪打消し
Time-Domain Bandwidth Extension	時間領域帯域拡張
track	トラック
transform coding	変換符号化
tri-pulse	トライパルス
unit sphere	単位球面
vector quantization	ベクトル量子化
water level	水位
white noise	白色雑音
wideband	広帯域