

■2 群 (画像・音・言語) - 8 編 (音声・オーディオ符号化)

2 章 有線・汎用符号化方式

(執筆者：守谷健弘) [2009 年 12 月 受領]

■概要■

ITU-T, MPEG (Moving Picture Expert Group) などで制定されてきた音声やオーディオ信号の符号化方式を紹介する。これらの符号化方式の用途は必ずしも固定されていないが、多様な電話サービス, インターネットを通じた配信, デジタル放送, ディスクなどの蓄積媒体への記録などに必須の技術として貢献している。これらの用途では相互接続性, 互換性がきわめて重要であるため, 一部のデファクトとして普及したものを除いて国際標準が大きな役割を担っている。ITU-T では双方向の通信用途で圧縮率のほかに低遅延, 符号化復号化の演算量が評価基準となる。これに対し MPEG では片方向の放送や蓄積用途であるため, 圧縮率のほかに復号器の複雑さ, 広帯域, 多チャンネル化などが重要である。

このように 2 つの標準化規格は住み分けて制定されたが, 用途の多様化でその境界は, はっきりしなくなっている。

【本章の構成】

本章では, 2-1 節で各種標準化方式の仕様一覧を示す。2-2 節で ITU-T での電話帯域の音声符号化, 2-3 節で ITU-T での広帯域音声符号化を紹介する。2-4 節で ITU-T 以外で重要なアメリカ連邦政府標準の低ビット音声符号化を紹介し, 2-5 節で MPEG でのオーディオ符号化について紹介する。

■2 群 - 8 編 - 2 章

2-1 符号化方式の一覧

(執筆者：守谷健弘) [2009 年 12 月 受領]

有線・汎用符号化方式の一覧を表 2・1 に示す。

表 2・1 有線・汎用符号化方式の一覧

方式	ビットレート (kbit/s)	標本化 周波数 (kHz)	フレーム長 (msec)	標準化 機関	規格 番号	用途
PCM	64	8	0.125	ITU-T	G.711	公衆通信
ADPCM	16, 24, 32, 40	8	0.125	ITU-T	G.726	専用線 PHS など
エンベデッド ADPCM	16, 24, 32, 40	8	0.125	ITU-T	G.727	パケット
MP-MLQ/ACELP P	5.3/6.3	8	30	ITU-T	G.723.1	IP 電話
LD-CELP	16	8	0.625	ITU-T	G.728	専用線
CS-ACELP	8	8	20	ITU-T	G.729	IP 電話
エンベデッド CS-ACELP	8+	8/16	20	ITU-T	G.729.1	IP 電話
可変レート エンベデッド	8, 12, 16, 24, 32	8/16	20	ITU-T	G.718	IP 電話
PCM	64/80/96	8/16	5	ITU-T	G.711.1	会議電話
LLC	可変	8	5, 10, 20, 30, 40	ITU-T	G.711.0	IP 電話
SB-ADPCM	48, 56, 64	16	0.125	ITU-T	G.722	会議電話
MLT	24/32	16	20	ITU-T	G.722.1	会議電話
WB-ACELP	6.6~23.85	16	20	ITU-T	G.722.2	会議電話
-	32~128	48	40	ITU-T	G.719	会議電話
LPC-10	2.4	8	20	米国政府	FS-1015	専用線
CELP	4.8	8	20	米国政府	FS-1016	専用線
MELP	2.4	8	20	米国政府	FS-1017	専用線
iLBC	15.2/13.33	8	20/30	IETF	RFC3951	IP 電話

■2 群 - 8 編 - 2 章

2-2 音声符号化

(執筆: 守谷健弘) [2009 年 12 月 受領]

ITU-T で制定されてきた電話帯域 (0.3~3.4 kHz) の音声符号化方式を紹介する。信号処理プロセッサの能力向上やメモリ量の増加により、演算量や遅延の増加を伴いながら、情報の圧縮率が改善されてきた。基本技術はモバイル通信へと応用された。2000 年以降、ネットワークのブロードバンド化や IP 化が進行し、圧縮よりも互換性を保った高品質化を目的とした標準が制定されている^{1)~5)}。

2-2-1 G. 711

(執筆: 守谷健弘) [2009 年 12 月 受領]

G.711 は、図 2・1 のように 14 ビット相当の振幅値を対数に近い非線形なステップ幅で 8 ビットに量子化した圧伸 PCM (Pulse Code Modulation) で、1972 年に CCITT において標準化されている。日米で使われている μ 則と、欧州や中国で使われている A 則がある。いずれも、8 ビットは 1 ビットの極性、3 ビットのセグメント、4 ビットの線形部からなる。セグメントは指数的に折れ線の傾きが変わることになる。振幅の大きいサンプルの量子化誤差は大きいですが、元のアナログの振幅値との相対的誤差は振幅によらず一定となる。これは瞬時の振幅の聴感上許される量子化誤差は振幅に比例する性質があることを利用しており、8 ビットの振幅でもアナログ音声との品質の差はない。

このため電話の基幹回線伝送や VoIP (Voice over Internet Protocol) などに広く使われている。VoIP など、音声をパケットにより伝送する機会が増えたことに伴い、ITU-T において、PLC (Packet Loss Concealment: パケット消失隠蔽) 法が Annex I として 1999 年に標準化されている。

2-2-2 G. 726

(執筆: 守谷健弘) [2009 年 12 月 受領]

G.726 は 1 サンプルごとの後方適応型の適応予測と適応量子化ステップ幅制御を用いた ADPCM (Adaptive Differential PCM: 適応予測符号化方式) である。図 2・1 に示すように、予測は 2 次の全極型 (AR)、6 次の移動平均型 (MA) を用いる。サンプルごとに予測誤差に基づいて予測係数を更新する⁶⁾。予測誤差の量子化のステップ幅は過去の系列から適応的に更新する。予測誤差をサンプルあたり 4 ビットで量子化するので、情報量は G.711 の半分の 32 kbit/s となる。G.711 と比較した品質の低下はわずかで、有料品質 (Toll quality) の限界の基準ともなっている。PCM との ADPCM の多段の接続で劣化が累積しない工夫やモデム信号を伝送できるような適応化の工夫が盛り込まれている。

簡単な処理で遅延が 1 サンプルで情報量が半減できるので、PHS (Personal Handy Phone) や専用線多重化伝送などに使われている。なお、予測誤差のビット数が 2, 3, 5 で、それぞれ 16, 24, 40 kbit/s の規格も G.726 に含まれている。

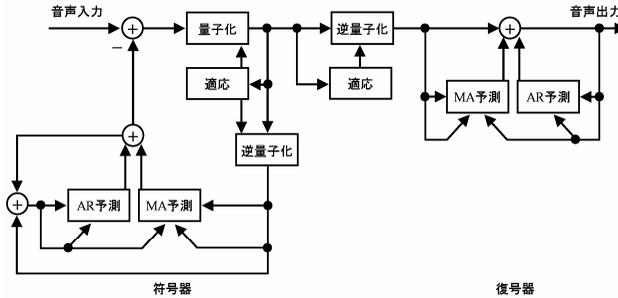


図 2-1 G.726 の符号器と復号器の構成

2-2-3 G. 727

(執筆者：守谷健弘) [2009年12月 受領]

G.727 は G.726 と同等の予測や適応化を行う予測符号化であるが、予測や適応化は振幅の量子化ビットの上位のビットだけを使って行うエンベデッド符号化である。サンプルごとの予測誤差の量子化ビットそのものは 2, 3, 4, 5 ビットで 16, 24, 32, 40 kbit/s に対応できるが、その量子化ビットをコアビットと拡張ビットに分ける。予測誤差は全部のビットを有効に利用して量子化誤差を小さくすることができるが、予測係数や量子化ステップサイズなどのすべての適応化はコアビットのみを使って行われる。

伝送や復号時には拡張したビットのみを別のパケットなどで伝送し、何らかの制約に応じて拡張ビットを廃棄しても、コアビットだけでそれに対応する波形が正しく再構成できる。コアビットであれば G.726 と等価になるが、全体で同じレートでは G.726 より適応効率が悪いので品質は低下する。

2-2-4 G. 723. 1

(執筆者：守谷健弘) [2009年12月 受領]

PSTN (Public Switched Telecommunication Network) 用テレビ電話のための音声符号化標準で、5.3 kbit/s の ACELP (Algebraic CELP: 代数 CELP) と 6.3 kbit/s の MPC-MLQ (Multi Pulse Coding-Maximum Likelihood Quantizer: 最尤量子化マルチパルス符号化) の 2 つのモードをもつ⁷⁾。

5.3 kbit/s モードは、フランステレコムと Sherbrook 大学、そして 6.3 kbit/s モードは、イスラエルの DSP グループ社の提案が基礎になっている。両者の違いは、雑音励振ベクトルの表現の部分のみである。フレームごとの線形予測係数は LSP パラメータに変換し高次と低次のベクトルをそれぞれ 9 ビットでベクトル量子化する。適応符号帳は 5 タップの予測フィルタを使う。ピッチ周期は第 1 第 3 フレームに整数周期で 7 ビット、第 2 第 4 副フレームに前の副フレームからの差分の 2 ビットで表す。一般的なフォルマント強調ポストフィルタと、時間方向の前後でピッチ合成フィルタを用いるピッチ強調ポストフィルタを使う。

2-2-5 G. 728

(執筆者：守谷健弘) [2009年12月 受領]

G.728 は 16 kbit/s のレートで遅延 2 ms 以下を達成する LD-CELP (Low delay CELP: 低遅延 CELP) 方式である。図 2-2 のように、5 サンプルごとの後方適応線形予測とその予測誤差信

号をベクトル量子化することにより、低ビットレートと低遅延の両立を実現している。

50 次の予測係数は、過去に復号された信号に窓をかけて相関関数を計算して求めるので、伝送する必要はない。予測誤差信号に対し、利得に 3 ビットの適応予測符号化を形状に 7 ビットのベクトル量子化を行う。すなわち、利得の値は対数領域での過去の利得から予測してその補正値を符号化し、形状の符号化として、利得で正規化した 5 サンプルからなるベクトルを 128 個の符号帳ベクトルから選択する⁸⁾。

この際の最小化する選択基準は、線形予測合成フィルタを通したあとの波形と入力波形との歪である。復号器でも高次の線形予測分析のために演算コストがかかることや、過去の復号信号に強く依存するので、符号誤りに対してはその影響が長く残るといった難点があり、普及は限定的である。

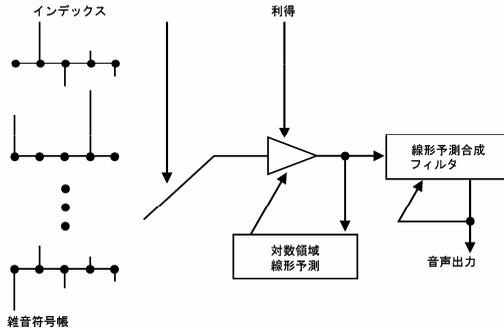


図 2・2 G.728 の復号器の構成

2-2-6 G. 729

(執筆者：守谷健弘) [2009 年 12 月 受領]

G.729 は図 2・3 のように Conjugate Structure ACELP (CS-ACELP) と呼ばれる 80 サンプル (10 ms) のフレームの 8 kbit/s を基本とする ACELP である。線形予測パラメータは、LSP パラメータを多段ベクトル量子化する。また、利得の符号化に 2 チャンネルの符号帳を使う共役ベクトル量子化を採用している。固定雑音励振信号は、40 サンプルごとに 4 本の 1 または -1 のパルスからなり、その位置と極性を 17 ビットで指定する⁹⁾。

アネックス A は低演算、B は音声検出と無音削除、C は浮動小数点記述、D は 6.4 kbit/s、E は 11.8 kbit/s、F、G、H、L は参照ソフト、を規定している。8 kbit/s でほぼ 32 kbit/s の ADPCM 相当の品質を達成し、符号誤りやパケットロスに耐性がある。このため、第 2 世代高品質符号化として日本の携帯電話や、VoIP での低ビット圧縮符号化のデフォルト符号化方式として世界に広く普及している。

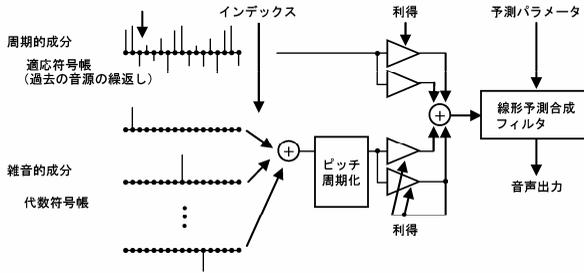


図 2-3 G.729 の復号器の構成

2-2-7 G. 729. 1

(執筆者：江原宏幸) [2009年12月 受領]

G.729.1: G.729 based embedded variable bit-rate coder: an 8~32 kbit/s scalable wideband coder bitstream interoperable with G.729

G.729.1 は 12 種類のビットレート (8~32 kbit/s) で復号可能なスケーラブル広帯域符号化方式で、2006年5月にITU-T 勧告として承認された。8 kbit/s で G.729 互換であり、14 kbit/s 以上のビットレートで広帯域信号に対応する。フレーム長は 20 ms である¹⁰⁾。8 kbit/s と 12 kbit/s で動作する階層型 CELP 符号化方式に 2 kbit/s の帯域拡張符号化方式と 18 kbit/s の MDCT 符号化方式を組み合わせている。QMF で入力信号を帯域分割し、低域成分を CELP 符号化方式と MDCT 符号化方式で、高域成分を帯域拡張符号化方式及び MDCT 符号化方式で、それぞれ符号化する。

階層型 CELP 符号化方式は、G.729 を基本アルゴリズムとする 8 kbit/s のコアレイヤと、固定符号帳を追加する拡張レイヤとからなる。帯域拡張符号化方式には、時間包絡と周波数包絡パラメータで信号を表現する TD-BWE (time-domain bandwidth extension) 方式を採用している。MDCT 符号化方式では、低域成分 (CELP 符号化残差) と高域成分を一括してサブバンドごとに利得-形状量子化する。サブバンドごとの利得をスペクトル包絡としてハフマン符号を用いた手法で量子化した後、量子化スペクトル包絡から各サブバンドのビット配分を決定して形状量子化を行う。また、フレーム消失耐性を高めるため、符号化された信号の分類情報、位相情報、エネルギー情報が、デコーダ側の FEC (frame erasure concealment) 処理のために符号化される。

G.729.1 は狭帯域 VoIP システムから広帯域 VoIP システムへの移行を促し、今後の高品質 VoIP 用音声符号化方式として期待されている。

2-2-8 G. 718

(執筆者：江原宏幸) [2009年12月 受領]

G.718: Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8~32 kbit/s

G.718 は 5 階層 (8, 12, 16, 24, 32 kbit/s) からなるスケーラブル広帯域符号化方式で、2008年6月にITU-T 勧告として承認された。フレーム長は 20 ms で 8 kbit/s と 12 kbit/s で動

作する階層型マルチモード CELP 符号化方式に MDCT 符号化方式を組み合わせている¹¹⁾。

階層型マルチモード CELP 符号化方式は、入力信号の特徴（無声、過渡、有声、その他）に応じて符号化モードを切り替える方式で、3GPP2 規格の VMR-WB 方式をベースとするコアレイヤと、適応符号帳と固定符号帳を追加して利得の再最適化を行う拡張レイヤとからなる。狭帯域信号と広帯域信号のそれぞれに適した符号化モードを備えており、どちらの信号が入力されても高品質に符号化することができる。

MDCT 符号化レイヤは 3 階層から構成され、MDCT 係数のベクトル量子化を行う。3 階層のうち 2 階層は、音声信号用と音楽信号用の符号化モードを備え、音声と音楽に対する符号化性能両立をはかっている。また、フレーム消失耐性を高めるため、G.729.1 と同様の FEC 情報も符号化する。さらに、G.722.2 の 12.65 kbit/s をコアレイヤとする G.722.2 互換モードも備えている。このモードでは、1 階層目のビットレートが 12.65 kbit/s となるので、全体で 4 階層 (12.65, 16, 24, 32 kbit/s) のスケーラブル符号化方式となる。

G.718 は 3GPP 規格の AMR-WB (ITU-T 勧告 G.722.2)、3GPP2 規格の VMR-WB と互換性があり、次世代携帯電話システムの高品質音声通信用符号化方式として期待されている。

2-2-9 G.711.1

(執筆: 佐々木茂明) [2009 年 12 月 受領]

ITU-T において 2008 年 3 月に国際標準となった広帯域音声符号化である。G.711 をコアレイヤとしたスケーラブルな構成により、再エンコードの処理なしに G.711 と接続できることを特徴とする¹²⁾。

サンプリングレートは 16 kHz、フレーム長は 5 ms、アルゴリズム遅延は 11.875 ms、演算量 (エンコーダ+デコーダ) は 8.7 WMOPS。広帯域入力信号を QMF で低域と高域に分割した後、低域信号をノイズシェーピング付きの G.711 で量子化し、コアレイヤ (64 kbit/s) とする (μ 則、A 則ともに適用可能)。さらに、量子化誤差を適応ビット割当を用いて 16 kbit/s で量子化し、低域拡張レイヤとする。高域信号は MDCT 係数に変換後、インターリーブ共役構造ベクトル量子化を用い、高域拡張レイヤとして 16 kbit/s で符号化する。各レイヤの組み合わせにより、4 モードのビットレートで利用できる (電話帯域: R1 = 64 kbit/s, R2a = 80 kbit/s, 広帯域: R2b = 80 kbit/s, R3 = 96 kbit/s)。主観品質において、広帯域 R2b モードで G.722 の 64kbit/s と同等以上、また、電話帯域 R1 (G.711 互換) モードでも従来の G.711 と比較しての品質向上 (特にフレーム誤り、A 則の低レベル入力時) が確認されている。

なお、上記において WMOPS は Weighted Million Operations Per Second の略で、ITU で定められた演算処理能力を表す単位である。

2-2-10 G.711.0

(執筆: 守谷健弘) [2009 年 12 月 受領]

VoIP で使われる G.711 (μ 則、A 則) のロスレス符号化で 2009 年 10 月に制定された。5, 10, 20, 30, 40 サンプルのフレーム単位で完結する低演算量 (符号化復号化合わせて平均 1 WMOPS, 最悪 1.7 WMOPS) の符号化である。音声通信で平均的な音声率 45% の音声や実際の通話音声に対しておおよそ 50% の圧縮が実現できる。

線形領域にマッピングした領域での PARCOR 係数を使った線形予測のほかに、信号系列の偏りのパターンに対する複数の圧縮手法が用意され最適なものを選択する。信号の偏りの例としてはフレーム内で振幅が一定であるとか、最大振幅が小さく限定されるとか、例外振幅

が1つであるとかである。端末間での圧縮符号化としてだけでなく、VoIPのルータ間の伝送、VoIP音声の蓄積などの効率化の用途などにも利用可能である。

■参考文献

- 1) ITU-T Recommendation Gシリーズ <http://www.itu.int/ITU-T/recommendations/index.aspx?ser=G>
- 2) TTC JT-G シリーズ <http://www.ttc.or.jp/cgi/document-db/docdb.cgi?cmd=s&sc=T16>
- 3) 守谷健弘, “音声符号化技術,” 信学誌, vol.84, no.11, pp.836-842, 2001.
- 4) 大室伸, 佐々木茂明, “音声通信用符号化技術と ITU-T 標準,” 電学誌, vol.127, no.7, pp.403-409, 2007.
- 5) A. Spanias, “Speech coding: A tutorial review,” Proceedings of the IEEE, vol.82, pp.1541-1582, October, 1994.
- 6) T. Nishitani, S. Aikoh, T. Araseki, K. Ozawa, and R. Maruta, “A 32 kb/s toll quality ADPCM codec using a single chip signal processor,” IEEE Proc. ICASSP, pp.960-963, 1982.
- 7) ITU-T, “G.723-Dual rate speech coder for multimedia communications transmitting at 5.3 & 6.3 kbit/s,” 1995.
- 8) J.H. Chen, “High-quality 16 kb/s speech coding with a one-way delay less than 2 ms,” IEEE Proc. ICASSP, pp.453-456, 1990.
- 9) R. Salami, C. Laflamme, J.P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, “Design and description of CS-ACELP: Toll quality 8 kb/s speech coder,” IEEE Trans. Speech and Audio Processing, vol.6, no.2, pp.116-130, 1998.
- 10) I. Varga, S. Proust, and H. Taddei, “ITU-T G.729.1 scalable codec for new wideband services,” IEEE Communication Magazine, Oct., pp.131-137, 2009.
- 11) M. Jelinek, T. Vaillancourt, and J. Gibbs, “G.718: A new embedded speech and audio coding standard with high resilience to error-prone transmission channels,” IEEE Communication Magazine, Oct., pp.117-123, 2009.
- 12) Y. Hiwasaki and H. Ohmuro, “ITU-T G.711.1: Extending G.711 to higher-quality wideband speech,” IEEE Communication Magazine, Oct., pp.110-116, 2009.

■2 群 - 8 編 - 2 章

2-3 広帯域音声符号化

(執筆著者：守谷健弘) [2009年12月 受領]

ITU-T で制定されてきた広帯域 (16 kHz サンプル) の音声符号化を紹介する。テレビ会議や電話会議ではハンズフリーの通信が普通で、その際には広帯域音声がよく使われる。狭帯域の音声符号化と同様に、低ビットが進展したが、2000 年以降は高品質化をめざしたステララブル符号化や、スーパーワイドバンド (超広帯域, 32 kHz サンプル)、フルバンド (全帯域, 48 kHz サンプル) に対応する規格も制定されている。

2-3-1 G.722

(執筆著者：守谷健弘) [2009年12月 受領]

G.722 は図 2.4 のように SB (Split Band) ADPCM で、帯域を 2 分割し、それぞれ 8 kHz サンプルの信号として低域は予測誤差 1 サンプルあたり 4, 5, 6 ビット (48, 56, 64 kbit/s に対応) の ADPCM, 高域は 2 ビットの ADPCM が使われている。7 kHz 帯域 (AM 放送程度)、標準化周波数 16 kHz で、情報量が 48, 56, 64 kbit/s の 3 種類のサブバンド ADPCM 方式である¹⁾。

まず、標準化周波数 16 kHz の音声に対して QMF (直交鏡像フィルタ) で帯域を完全に 2 つに分割する。低域 (8 kHz 以下) と高域でそれぞれ標準化周波数 8 kHz 相当の信号となる。それぞれを ADPCM で符号化する。64 kbit/s のモードではサンプル当たりの量子化ビット数は低域が 6, 高域が 2 である。また、56 kbit/s と 48 kbit/s のモード場合、それぞれ、低域の量子化ビット数が 5, 4 となる。さらに、いずれのモードにおいても測係数及びステップ幅の適応は低域で 4 ビットの精度、高域で 2 ビットの精度で行われる。

すなわち、G.727 と同様な 4 ビットをコアとするエンベデッド ADPCM である。このため、伝送路や復号器の都合で高域の 2 ビット、低域の 2 ビットに対応するビット系列を廃棄し低域の 4 ビットだけでも、32 kbit/s の ADPCM 符号系列として、音声信号を復号できる。

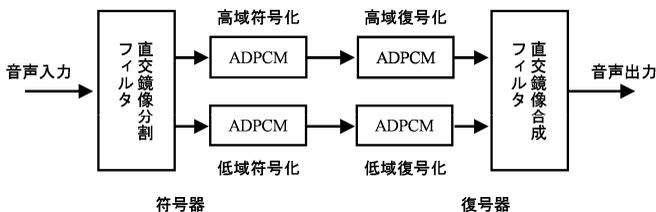


図 2.4 G.722 の復号器の構成

2-3-2 G.722.1

(執筆著者：伊藤博紀) [2009年12月 受領]

G.722.1 は、Polycom 社提案による DCT をベースとした周波数領域の符号化方式であり、1999 年に ITU-T で標準化選定された²⁾。2 種類のビットレート (24 kbit/s 及び 32 kbit/s) をサポートし、アルゴリズム遅延は 40 msec である。16 kHz サンプルングした 640 サンプルの信

号を 50% ずつオーバーラップさせて DCT を行う Modulated Lapped Transform (MLT) により、20 msec ごとに周波数領域の係数に変換される。このうち 7 kHz 以下に対応する MLT 係数のみがスカラー量子化された後、ハフマン符号化されて伝送される。エンコーダとデコーダ合わせた演算量は 11 WMOPS (Weighted Million Operations Per Second) 以下であり、aacPlus の約 1/4、AMR-WB+ の約 1/8 である。2006 年に、帯域を 14 kHz に、ビットレートを 3 種類 (24 kbit/s, 32 kbit/s 及び 48 kbit/s) に拡張した Annex C が追加された。

2-3-3 G. 722. 2

(執筆者：守谷健弘) [2009 年 12 月 受領]

G.722.2 はモバイル向け 3GPP 規格の AMR-WB と互換性があり、標本化周波数 16 kHz の入力に対し、12.8 kHz にダウンサンプルし、モバイル向けと同じアルゴリズムの AMR で符号化する。復号側では標本化周波数を 16 kHz に上げて、さらに高域をスペクトル包絡で補うことで 7 kHz 帯域信号を再構成する³⁾。

2-3-4 G. 719

(執筆者：佐々木茂明) [2009 年 12 月 受領]

G.719 は、ITU-T 初の 20 kHz 帯域まで再生できる音声符号化標準で、2008 年 6 月に承認された。主なスペックは、標本化周波数 48 kHz、フレーム長 20 ms、アルゴリズム遅延 40 ms、ビットレート 32~128 kbit/s、最大演算量 21 WMOPS (エンコーダとデコーダの合計)。

時間領域信号を MDCT 係数に変換後、符号化を行う変換符号化をベースに、適応変換長切り替え、適応ビット割当、ラティス VQ 等の技術が用いられている⁴⁾。ITU-T の品質確認試験では 32, 48, 64 kbit/s のビットレート時の符号化品質が、音声・音楽ともにそれぞれ LAME MP3 の 40, 56, 64 kbit/s と同等以上であることが確認されている。当初、7 kHz 広帯域音声符号化標準 G.722.1 の拡張方式として標準化が開始されたが、G.722.1 本体やその 14 kHz 帯域拡張である G.722.1 Annex C との互換性はない。

■参考文献

- 1) M. Taka, P. Combesure, P. Mermelstein, and F. Westall, "Overview of the 64 kbit/s (7 kHz) audio coding algorithm," IEEE Proc. Globecom, pp.599-604, 1986.
- 2) ITU-T Recommend, G.722.1, "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 1999.
- 3) ITU-T Recommend, G.722.2, "Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB)," 2002.
- 4) A. Taleb and S. Karapetkov, "G.719: The first ITU-T standard for high-quality conversational fullband audio coding," IEEE Communication Magazine, Oct., pp.124-130, 2009.

■2 群 - 8 編 - 2 章

2-4 低ビットレート音声符号化

(執筆著者：守谷健弘) [2009年12月 受領]

アメリカ政府のプライベート通信標準として 1015, 1016, 1017 がある。公衆通信でないこともあり、世界にさきがけて低ビット化のために先導的な技術が使われ、ソースコードが開示されたことで、その後の符号化技術の進展に貢献した。iLBC は VoIP に特化しパケット消失耐性が優れる IETF 標準である。

2-4-1 1015

(執筆著者：守谷健弘) [2009年12月 受領]

US Federal Standard (米国連邦政府標準) FS1015 規格は、もともと 1976 年に国防省 (Department of Defence) によって制定された LPC-10 と呼ばれる、情報量が 2.4 kbit/s の PARCOR ボコーダである¹⁾。1986 年に一部改善され LPC10e となり、FS1015 となっている。

従来のアナログ電話回線で低速モデムを通じて音声を送ることができ、秘話通信も容易に実現できる。スペクトル包絡の分析には共分散法を使い、特に有声フレームではピッチ周期の整数倍のフレーム長を使う。22.5 ms を 1 フレームとし、有声フレームではスペクトル包絡は 10 次の PARCOR 係数で量子化するが、無声フレームでは 4 次の係数を使う。ただし、いずれも 1 次と 2 次の係数は LAR に変換して量子化する。

2-4-2 1016

(執筆著者：守谷健弘) [2009年12月 受領]

国防省 (Department of Defence) が 1988 年に制定したものであり、情報量は 4.8 kbit/s で通称 DoD-CELP といわれている²⁾。制定に際しては技術コンテストが行われたが、AT&T ベル研究所の P. Kroon と B. Atal によって設計された CELP をもとに、国防省で最適化された方式が選定された。この規格は本来、機密通信用途で制定したものであるにもかかわらず、アメリカ政府によってソフトウェアが無償で世界中に配布されたことにより、低ビット符号化の研究用の基準としてもよく使われている。スペクトル包絡は LSP パラメータの次数間予測符号化量子化する。適応符号帳には非整数ピッチ周期を使う。雑音符号帳は 1 または -1 のスパースなパルス列からなる 1 つのベクトルを、2 サンプルずつシフトして別のベクトルとして扱う。これにより演算量とメモリ量を削減している。スペクトル包絡パラメータの線形補間や適応ポストフィルタも導入している。

2-4-3 1017

(執筆著者：守谷健弘) [2009年12月 受領]

2.4 kbit/s の新しい標準で MELP (Mixed Excitation Linear Prediction: 混合励振線形予測) に基づいている³⁾。先に標準化されていた 4.8 kbit/s の CELP(1016)の品質を上回る。有声音はパルス列、無声音は雑音を駆動音源とする全極型の線形予測ボコーダである。

有声音の場合、MBE と同様に有声無声の判定を 4 個の周波数帯域ごとに独立に行ない、さらに周期パルス列の位置を少し移動させたり、DFT の強度で駆動音源のスペクトル包絡の補正をしたり、パルス拡散フィルタを使ったりして、自然性の向上をはかっている。スペクトル包絡の量子化には 10 次の LSP パラメータの多段ベクトル量子化が使われている。

2-4-4 iLBC (RFC3951)

(執筆: 伊藤博紀) [2009年12月受領]

CELP 方式に基づきインターネットを用いた低ビットレート音声通信向けに開発された符号化である⁴⁾。2004年にIETFでRFC3951として採用された。

フレーム長は20 msecと30 msecの2種類あり、それぞれビットレートが15.2 kbit/sと13.33 kbit/sである。パケット損失による影響をフレームを超えて伝播させないことで、パケット損失耐性を強化している。主観評価試験によりITU-T G.729Aと比較して、パケット損失なしでは同等、パケット損失ありでは高い音質であることが示されている。

■参考文献

- 1) T.E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," Speech Technology Magazine, April, pp.40-49, 1982.
- 2) P. Campbell, P. Joseph, Jr., T.E. Tremain, and V.C. Welch, "The DoD 4.8 kbps Standard (Proposed Federal Standard 1016)," in Advances in Speech Coding, ed., Atal, Cuperman and Gersho, Kluwer Academic Publishers, 1991.
- 3) A. McCree, K. Truong, E. George, T. Barnwell, and V. Viswanathan, "A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard," IEEE Proc. ICASSP, pp.200-203, 1996.
- 4) RFC 3951 <http://www.rfc-editor.org/rfc/rfc3951.txt>

2-5-2 MPEG-レイヤ II

32 個のサブバンド信号ごとに 12 サンプルの小ブロックを 3 個集めて大ブロックを構成し、ブロック正規化を行う。正規化係数は以下の 4 種類のモードの 1 つを選択して量子化する。

- [モード 0] 小ブロックごとに 6 ビットで量子化 (計 18 ビット)
- [モード 1] 第 1 と第 2 小ブロックを併合して 6 ビットで量子化 (計 12 ビット)
- [モード 2] 大ブロック全体を 6 ビットで量子化 (計 6 ビット)
- [モード 3] 第 2 と第 3 小ブロックを併合して 6 ビットで量子化 (計 12 ビット)

この結果、入力信号の 1152 サンプルを 1 フレームとして扱うことになる。パワを正規化したサブバンド信号をレイヤ I の場合と同じ形の量子化器でスカラー量子化するが、ステップ数の選択の種類は低域サブバンドのみレイヤ I と同じ 4 ビットで、中域が 3 ビット (0 ビット以外で 7 種類)、高域は 2 ビット (0 ビット以外で 3 種類) に制限される。ただし、いずれの帯域でもステップ数は 3 から 65535 までとりうる。量子化ビットが 2, 3, 4 ビットの場合はサブバンド信号ごとに 3 サンプルを併合して量子化単位とし、6, 9, 12 ビットで量子化再現値を表す。レイヤ II はビデオ CD の音声、DVD の音声、欧州のデジタル放送に広く使われている。

2-5-3 MPEG-レイヤ III

32 サブバンドに分割された信号は、それぞれ通常 18 点の MDCT (入力窓長は 36) で周波数領域に変換される。この結果、等価的に $32 \times 18 = 576$ 点の周波数スペクトルが得られる。入力信号の 1152 サンプルを 1 フレームとし、576 サンプルずつオーバーラップさせながら処理することになる。また、過渡的な信号の場合には変換長が $1/3$ の 6 点の変換を 3 回行う。この切り替えは、振幅が急に大きくなる信号系列に対して生ずるプリエコーの歪みを軽減するため、聴感的エントロピーに基づいて適応的に切り替える。

過渡的な信号は、スペクトルが広がり聴感を考慮したエントロピーが大きくなることと対応する。言い換えれば過渡的な信号には多くのビット数を要することになる。レイヤ III では遅延が許されるならば、隣接する複数のフレーム間で量子化ビットをやりくりするビット貯蔵 (bit reservoir) 手法も使い、平均ビット数を削減する。パワを正規化したサブバンド信号をミッドトレッド型の非線形ステップでスカラー量子化する。ステップ幅は振幅の 0.75 乗に比例させる項と、量子化結果で制御する変数の項の積で表す。さらに、高域から 0 が連続する領域でランレングス符号化、絶対値が 1 を越えない領域で 2 次元、1 を越える領域 4 次元のサンプルをひとまとめにしてハフマン (Huffman) 符号化を行う。量子化値は雑音配分 (noise allocation) と呼ばれる 2 段階の処理で決める。

まず 1 段階では、利用できるビット数を越えないようなステップ幅変数を設定して量子化を行う。第 2 段階では、最初の量子化の結果、余ったビットを再配分する。この際、実際の量子化再現値に基づいて歪みを計算し、最も歪みの小さくなる組み合わせを探索する。またステレオ信号に対して、MS (和差) 符号化と高域のスペクトル包絡の強度だけを保存するジョイントステレオ符号化も取り入れている。

2-5-4 AAC (Advanced Audio Coding)

MPEG-2 は、後方互換性を保ったまま、マルチチャンネル対応と、低サンプリング周波数対応の拡張が定義された。後方互換のための構成と最高ビットレートの制約のために十分な品質が得られないという問題が生じた。特にマルチチャンネルの高品質化の要請により AAC の策定が開始され、1997 年に制定された。

基本構成は MDCT 係数を不均一な帯域に分割して適応的な正規化と量子化を行い、可変長符号化するものである。フレーム長の適応化、フレームをまたがるビットの貯蔵、ノイズシェイピングを初めとする数多くの要素技術が統合されている。聴覚歪を最小とするようエンコーダを制御する機能が柔軟に盛り込まれているため、MP3 より品質が高く、128 kbit/s で、ほぼ聴覚的な劣化のない CD 帯域のステレオ信号の符号化が可能になった。AAC の LC (Low Complexity) プロファイルはデジタル放送（日本）、携帯プレーヤに広く使われている。

2-5-5 HE-AAC (High Efficiency Advanced Audio Coding)

広帯域の入力信号に対し、低周波領域部分は通常の AAC で符号化し、高域部分は SBR (Spectral Bandwidth Replication: スペクトル帯域拡張) 技術を使って、スペクトルの包絡などの少数の補助情報で表現し、復号器側で高域側は低域のスペクトルと補助情報から波形を擬似的に再構成する符号化である。低域の AAC の部分は後方互換性を保つため、商品化されている従来の復号器でも低域のみの波形を再生できる。

一方、新たな拡張復号器ではわずかの補助情報を拡張ストリームから復号することにより高域の再生が可能で従来の AAC より 30~50%少ない情報量で広帯域信号の再生ができる。また、サブバンド領域の再構成の低演算モードも用意されている。このため、モバイル系など低ビットの音楽配信用途として有効で、日本でもワンセグ放送の音声伝送に採用されている。また、同様の処理は MP3 に適用され、MP3 プロとしても市場に投入されている。さらに、ステレオ信号をモノラル信号とサブバンド領域のパラメータで表現する PS (Parametric Stereo: パラメトリックステレオ) 技術と組み合わせられて、HE-AAC V2 プロファイルとして使われている。

2-5-6 MPEG-4 ロスレス

ロスレス符号化は、圧縮解凍あるいは符号化復号化によって全く歪を生じない符号化である。聴覚特性を利用する圧縮符号化ほど圧縮できないが、元の音源を変化させないという点で、特にサンプリング周期の高い音源の蓄積や伝送に重要である。MPEG では下記の 3 種類 (ALS, SLS, DST) が制定された。

ALS は時間領域の予測とエントロピー符号化に基づく簡易な符号化である。予測係数は PARCOR 係数で、予測残差はライス符号または算術符号が使われている。浮動小数点も含む幅広い入力信号に対応し、予測回数やフレーム長の選択範囲が広く、長期予測、ブロック長切り替え、マルチチャンネル符号化、ジョイントステレオ符号化などの付加機能を備えている。

このため、専門家向けのオーディオ信号の蓄積、編集、伝送、長期間のアーカイブなどの用途はもちろん、個人向けにも使われる可能性がある。また、生体や地震などの音楽以外の時系列信号に使われる可能性がある。

SLS は MPEG-4 で規定されている AAC などの周波数領域の圧縮符号化をベースにその誤

差信号をスケーラブルに符号化し、最高レートでは歪のないロスレス符号化となる符号化である。保存用には全体のビット列を使ってロスレス符号化で完全な波形の再構成を実現できるし、ビット列の一部から聴覚的に劣化のほとんどない信号を再構成できるので、携帯プレーヤでは保存容量の許す範囲で可能な限り高品質の符号化を実現するなど、用途によって柔軟に選択できる。

DST は SACD (Super Audio CD) の 1 ビットオーバーサンプルのフォーマット (direct stream digital: DSD) に対応したフレーム単位の線形予測に基づくロスレス圧縮符号化である。この 1 ビットのフォーマットはサンプリングレートと振幅分解能のトレードオフが選択できるなどの利点がある。圧縮符号化は +1 (符号としては 1) または -1 (符号としては 0) の値を線形予測し、0 または 1 の予測誤差を求める。予測が正しければ、誤差はほとんど 0 になるので、エントロピー符号化で圧縮が可能である。誤差の性質に合わせてエントロピー符号化のテーブルを切り替え、また予測係数は別途補助情報として伝送、記録する。これまでのディスクで使われている 64 倍オーバーサンプル規格用の圧縮と互換性を持ち、さらにマスター用の 128、256 倍のオーバーサンプルまでの信号に対応するので、ディスクだけでなく、音楽の編集などの中途段階のファイルやサーバー上でのアーカイブなど、広く利用される可能性がある。

2-5-7 MPEG サラウンド

MPEG Surround は 5 チャンネル信号と 1 または 2 チャンネル信号の対応づけを SBR と類似のサブバンド領域で行う技術である。サラウンドのスピーカ配置を想定して作られた 5 チャンネルの音源をステレオに混合し、例えば AAC でベースの信号として符号化する。同時に 5 チャンネル信号のもつ空間情報のパラメータ (レベル差, 位相差, 相関) をサブバンド領域ごとに抽出して、ステレオの波形情報に後方互換性を保ってわずかな情報量の補助情報として追加する。

デコーダでは 2 チャンネルの信号の再生が可能であり、同時に空間情報のパラメータから 2 チャンネル信号から原音に聴感が近い 5 チャンネル信号を再構成できる。補助情報は 0 ビットから完全な波形情報に近い情報量まで選択でき、ベースの信号は PCM でも可能である。また SBR と同じく、サブバンド領域のパラメータから低演算で 5 チャンネル信号を再構成するモードも用意されている。

2-5-8 MPEG-4 オーディオ

1999 年に第 1 版、2000 年に第 2 版が制定され、下記のリストのように、幅広いビットレート、帯域、そのスケーラビリティなどの機能が含まれている。ストリーミング、双方向通信、無線通信など柔軟で広範囲の用途への応用が期待されている。ただ、2009 年時点での市場展開は限定的である。

- 低ビット音声 HVXC (Harmonic Vector eXcitation Coder) : 正弦波符号化と CELP の切り替えによる 2 kbit/s での符号化。
- 中低ビット音声符号化 CELP : MP-CELP を基本に電話帯域、広帯域音声のスケーラブルにカバーする符号化。

- 低ビット楽音符号化 TwinVQ (Transform domain Weighted Interleave Vector Quantization) : MDCT 係数の量子化にインタリーブベクトル量子化, 包絡推定に線形予測を用いた 6 kbit/s までの低ビット用符号化.
- 楽音符号化 AAC : MPEG-2 AAC を基礎に LTP (Long Term Prediction), PNS (Perceptual Noise Substitution), スケーラブル構成などのツールが拡張された符号化.
- 小ステップスケーラブル楽音符号化 BSAC (Bit Slice Arithmetic Coding) : MDCT 係数の量子化に水平方向の算術符号を使った, スケーラブル符号化. 韓国での放送システムで利用される.
- パラメトリック楽音符号化 HILN (Harmonic Individual Line and Noise) : 正弦波, パルスと雑音による低ビット楽音符号化.
- 低遅延楽音 AAC-LD (Low Delay AAC) : 512 点のみの MDCT 係数を使う AAC で双方向通信とも可能.
- 誤り耐性枠組み : 誤り訂正符号やビットの並び替えを可能にするフォーマット.
- 3D 音場 : 再生時の両耳効果, 環境特性のたたみこみ.
- 構造化音響符号化 : MIDI の拡張などの合成楽音ツール群.
- 音声合成インタフェース : 韻律情報などを付加するための言語仕様.

■参考文献

- 1) ISO/IEC JTC1 標準:
<http://isotc.iso.org/livelink/livelink?func=ll&objId=327975&objAction=browse&sort=name>
- 2) ISO/IEC 11172-3:1993, Information technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s—Part 3: Audio. (MPEG-1 Audio)
- 3) ISO/IEC 13818-3:1998, Information technology—Generic coding of moving pictures and associated audio information—Part 3: Audio. (MPEG-2 Audio)
- 4) ISO/IEC 13818-7:2006, Information technology—Generic coding of moving pictures and associated audio information—Part 7: Advanced Audio Coding (AAC). (MPEG-2 AAC)
- 5) ISO/IEC 14496-3:2009 Information technology—Coding of audio-visual objects—Part 3: Audio. (MPEG-4 Audio)
- 6) ISO/IEC 23003-1:2007 Information technology—MPEG audio technologies—Part 1: MPEG Surround.
- 7) 守谷健弘, “オーディオ符号化技術と MPEG 標準,” 電気学会誌, vol.127, no.7, pp.407-410, 2007.
- 8) 亀山涉, 渡辺裕, 金子格, “そこが知りたい最新技術 オーディオ・ビデオ圧縮入門,” インプレス R&D, 2007.
- 9) 守谷健弘, “音声音響符号化における標準,” 日本音響学会学会誌, vol.64, no.2, pp.114-118, 2008.
- 10) 守谷健弘, “音声・オーディオの処理技術,” 画像電子学会学会誌年報特集,” vol.37, no.6, pp.894-897, 2008.
- 11) 守谷健弘, “音声・音響符号化技術と標準化動向,” 電子情報通信学会音声研究会, 5 月, 2009.