

音声符号化とその品質評価の国際標準化の動向

林 伸二

Trends in international standardization of speech coding and its quality assessment

HAYASHI Shinji

Abstract

This paper presents an overview on recent trends in international standardization of speech coding with very personal view on ITU-T Study-Group activities. During two decades of the end of 20th century, big tides of digitalization and open connection of telecommunication networks accelerated bit reduction in speech coding with help of rapid progress in processing speed and circuit scale of semiconductor devices. The speech coding technologies achieved around 8-times efficiency than basic PCM and main interests in industry shifted from compression rate to surrounding technologies, considering the vast expansion of Internet and the trends of Personalization and Mobilization. For future progress in speech coding, further studies not only in signal processing but also in hearing and human sensibilities will be necessary.

Key Word

Speech Coding / ITU-T / standardization / Internet

[抄録]

音声符号化の近年の動向を、ITU-T標準化の状況から私見によりまとめる。20世紀末の約20年間に通信のデジタル化と回線の開放という大変化が起き、半導体デバイスの高密度化という背景とあいまって、音声符号化を取り巻く状況が激変した。基本的なPCM符号化から約8倍程の圧縮に成功したあと、環境の変化から要求条件も変遷し、標準化そのものの位置づけも変わってきている。インターネットを利用した音声通信とパーソナル・モバイル化の進展が環境変化の中心であった。今後、圧縮符号化技術が一応の成熟に達したと見れば、研究対象は圧縮率よりも周辺技術に向かうと思われるが、新たな技術革新には、信号処理理論だけでなく聴覚や感性にまつわる基礎的研究が望まれる。

キーワード

音声符号化 ITU - T 標準化 インターネット

1 はじめに

音声符号化のここ数年の動向について、ITU-Tにおける国際標準化の状況から私見をもとにまとめた。まず、PCM符号化に始まる音声のデジタル符号化を振り返り、その進歩は何であったかを解釈した上で、その後の発展を概括し、その後個別に成立した勧告の要点に触れる。また、音声圧縮符号化に伴う品質劣化を聴感上正確に評価するための検討とその標準化についても関連して触れる。

2. 前世紀末までの発展

G.711 PCM方式⁽¹⁾は64 kbit/sのビットレートを持っている。当初実用化されたPCM24方式は8 kHzサンプリングで7ビット量子化のアナログ圧伸方式であったが、数回のタンデム接続(トランスコーディング、すなわち、いったんアナログに戻して再び符号化を繰り返す)時の品質を保つため、8ビット量子化が採用され国際標準となった。このときの品質がToll Quality(商用電話品質)と呼ばれ、電話網に適用される音声符号化はこの品質を条件とすることになった。その後、品質を保ったまま、ビットレートを1/2とする研究開発が進み、32 kbit/s ADPCM⁽²⁾のG.721(後に可変レート化されG.726)、16 kbit/s LD-CELP⁽³⁾のG.728、8 kbit/s CS-ACELP⁽⁴⁾のG.729へと標準化が進んだ。さらに4 kbit/s符号化方式の標準化が熱心に検討され⁽⁵⁾、おおむね要求条件を満たす標準候補案が出来たものの、提案組織間の思惑や地域標準化団体やキャリア、製造各社の利害の絡みなどが調整できなかったようであり、勧告化に至らず終わった。このため、低レート化の勧告は8 kbit/sで打ち止めとなり、G.729とこれを範にとったマルチレート符号化AMR方式⁽⁶⁾が一旦終着の方式となっている。

2.1 その後の時代の要請

4 kbit/sへの低レート化が沙汰止みとなったのは、表面的には先述の理由と思われるものの、その実、社会の要請の底流が変化したためと見なければならない。音声符号化を取り巻く主な情勢は次のものである。

2.1.1 モバイル・パーソナル化

携帯電話は1980年代後半から爆発的に普及した。このため限られた電波帯域を有効利用することが至上命令となり日本でもPDCハーフレート化が推し進められ、ビットレートは3.45 kbit/sまで落とされた⁽⁶⁾。その後無線変調方式の研究が進みCDMA方式では100 kbit/s程度まで利用可能となるなど、ビットレート削減の要求はやや弱まり、代わって音声品質の向上やサービスの多様化がモバイル網にも求められるようになった。

2.1.2 VOIP

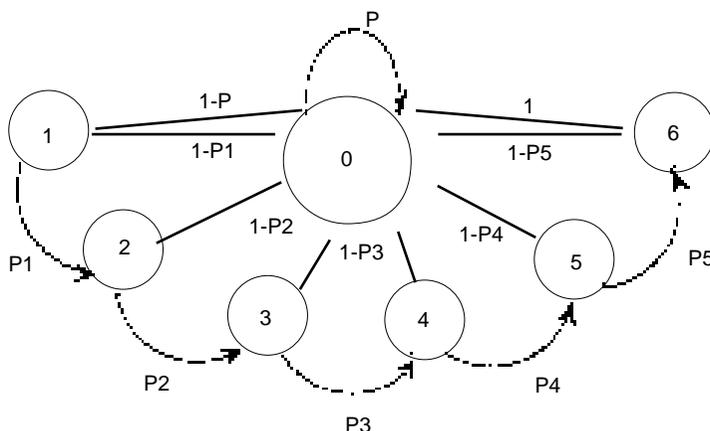
電話サービス事業は国営やそれに準じた独占的事業形態が効率的であるとする第二次世界

大戦後長く続いた常識が覆され、複数のコモンキャリアの競争によって良いサービスを提供させようという市場原理的傾向が世界的となってきた。複数キャリア間の相互接続や公衆網と専用網の相互乗り入れが制度的に推奨され、マルチメディアサービスの接続規約の勧告 H.323⁷⁾がITU-Tで勧告された。こうした環境整備により、インターネット上のIP電話と商用電話がIPゲートウェーで接続されることとなり、電話サービスをIPで行うVoIPが主流となる勢いとなった。元々データを運ぶパケットに実時間性の強い音声に乗せるという特徴により、音声符号化には従来と異なる要求が加えられることとなった。以上の情勢に関する技術的側面は次の点に集約される。

(1) 電波網の符号誤りとIPパケットロス

無線経路でしかも移動を伴う場合、伝送路符号誤りはランダム性よりもバースト性が主要となる。ITU-Tで早い時期に検討されたバースト誤りのモデルは図-1に示すものである。

図1 バースト誤り発生モデル



この条件下で当時検討下のG.728の品質を検討したところ伝送路ビット誤り訂正に頼る場合、品質を維持するためには保護に要するビット数が膨大となり、ビット低減アルゴリズムを採用する意味が薄れてしまうことが明らかとなった。このため、伝送路の符号誤り検出のみを用い、バースト誤り時はあえて訂正を試みず、誤り以前の健全なフレームの情報を外挿し、音声固有の性質を利用し、聴覚的に影響が隠蔽されるような後処理を加えることが試みられた。G.728は後方予測を多用するという特徴があるため、以前にPCM方式に提案された波形補間手法⁸⁾よりも隠蔽処理の効果が劇的であり、隠蔽処理をエラー訂正より優先する先鞭をつけるものとなった。

IPネットワークでは、移動無線通信路的な符号誤りは起こりにくいですがbest effort型パケット通信に固有の問題、すなわち、(i)パケット到達時間の揺らぎ、(ii)比較的長時間の遅延、(iii)

パケットの消失、が起こる。これらには、パケット通信システム設計による対策、音声符号化周辺システムによる対策と符号化アルゴリズムによる対策が採られる。(i)の時間的揺らぎすなわちジッタは受信システムに揺らぎを吸収するバッファを設けて対応する。必然的にバッファに伴う遅延を生ずるため、発生するジッタの性質に適応してバッファ長を調整する適応処理が用いられる。過大なバッファを要するような大きな遅延に対しては、パケット到着を待たず、消失と判定するのが合理的となる。(ii)の遅延時間は音声符号化に本質的に伴うアルゴリズム遅延時間、パケット構成時間、上記バッファ時間、ネットワーク伝送時間の総和である。旧来の電話網のように2線区間を含む伝送系では2線4線接続部でエコーを生じ、エコーが感知される場合はわずかな遅延でも大きな品質阻害要因となることが知られている。このためIPゲートウェイなどでは、エコーキャンセル処理を行う。また、会議通話のように音響結合が起こる場合、端末で音響結合によるエコーをキャンセルする処理が行われる。エコーの存在しない純粋な遅延については、許容範はかなり長時間(400 ms程度)となるが一定の制限は必要であり、ITU-TではG.114⁹⁾に勧告されている。この評価の元となったのは衛星通話とデジタル圧縮通話の品質が問題となり始めた1980年代に日本電信電話で行われた主観評価試験が貢献している¹⁰⁾。(iii)のパケットロス、は、網側の輻輳制御などの事情による場合と上記(ii)の遅延制限から受信端末で廃棄する場合がある。いずれにせよ、ここに利用されるのがFECと同様な処理であり、最後に到来した健全なパケットの情報からパラメータ外挿により音声を合成して、パケットロスを隠蔽する。これはPLC(Packet Loss Concealment)と呼ばれる。また、より正確な内挿が可能となるよう、符号化側で冗長な情報を加える手法も考案されている。

(2) スケーラビリティ

公衆電話網は一元的管理が行われる、整然とした設計に基づく網であって端末から加入者網、基幹網いずれも当初計画により整合が取れている。一方、IP網ではありとあらゆる設計による広範囲な性能の端末が一定のプロトコルを満たせば繋がることを許すため、ある意味自由で自然発生的な構成が起こりうる。こうした条件下で、たとえば複数地点間を結ぶ会議サービスを試みた場合、異なる仕様の端末間で最適な品質の通信をそれぞれの間で行いたいという要求が起きてくる。このような背景で、スケーラブル(階層的)音声符号化が考案された。スケーラブル符号化とは、階層構造からなる符号化ビットストリームを持ち、下位から任意の階層までを復号することができる特徴を持つ。最上の階層まで復号すると最も高い品質(周波数帯域と量子化精度)が得られ、端末の能力やネットワークの速度、輻輳の程度に応じて、符号化側に何の変更も求めず、ネットワークまたは端末側から復号品質を選ぶことができる。

(3) ソフトウェア実現

音声符号化アルゴリズムの実現は専用の通信端末機器ではDSP(Digital Signal Processing)チップへのプログラムで行われるが、PC(Personal Computer)の能力が向上してきたため、

PC上のアプリケーションプログラムとして、MPU（Micro Processor Unit）の動作で実現されることも普通となってきた。また、PDA（Personal Digital Assistance）や携帯ゲーム器なども実質小型PCのようにMPUでアプリケーションを動作させる構造となっている。このことは、音声符号化にも従来と異なった要求をもたらすこととなった。集約すると以下の3点があげられる。

i) 演算量の低減

DSPの飛躍的な能力向上により、音声符号化アルゴリズムは複雑さが増しても高度の圧縮率、高品質、高機能を追求することができた。しかし、MPUの能力の一部を用いた音声符号化が望まれるや、再び、演算量の削減は主要な要求に返り咲くこととなった。

ii) ソフトウェア動作

DSPは限られた回路規模で最大限の効力を発揮するために、演算精度を用途に合わせて限定しており、たとえば16 bit固定演算精度を採用し、音声符号化アルゴリズムもこれに合わせて規定される。PCは汎用性から浮動小数点演算が通常であり、固定小数点精度の規定と相互接続するためには別途配慮が必要となる。アルゴリズム規格の記述自体もテキストで表現するよりもコンピュータ言語で直接記述する方が合理的に思われる。アルゴリズムをコンパイル可能なソースコードで配布するのは著作権上の難点があるため、プラットフォーム限定ではあるが、実行に組み込み可能なライブラリとしてレファレンスを配布することも考えられる。

iii) 非標準の余地

前述の第ii)項に沿って考えると、PC上で動作する音声通信サービスでは、音声符号化アルゴリズムは、PCアプリケーションの一部となり、どんな符号化アルゴリズムを用いても同一アプリケーション同士の相互接続には何の問題もなく、また、スケラブル符号化により既存公衆網端末と相互接続可能な限り、内部の符号化アルゴリズムは自由となる。そこで既存標準を内包し、より高度な機能をオプションとするいわば「超標準」がデファクトスタンダードを目指して市場で競い合うという状況が予想される。

3 近年の新しい勧告

3.1 近年の個別勧告化動向

2006年勧告のG.729.1⁽¹¹⁾はG.729をコアとして、品質、帯域を階層的に拡張するもので、ビットレート増に応じて高品質化できるとともに、IP用途に向けてPLCを充実している。G.711をコアとして電話帯域内での品質向上、帯域の階層的拡張を図り、既存方式との対向可能性を残しながら、会議通話システムへの適用を図ったG.711.1⁽¹²⁾は2008年に勧告された。同年に勧告されたG.718⁽¹³⁾は8-32 kbit/sで電話帯域から7 kHz帯域の音声を階層的に符号化でき、既存のG.722.2⁽¹⁴⁾（3GPP AMR-WB）方式と相互接続可能なモードを含んでいる。このように、音声符号化の潮流は、圧縮率の向上が約8 kb/s程度で一旦止まり、網のIP化、モバイル化という環境変化に応じた柔軟かつロバスタな拡張に移行している。

広帯域符号化については、放送用途のコメントリック通信や、H.320⁽¹⁵⁾オーディオビジュアルシステムの遠隔会議通信をデジタル加入者線の64 kbit/sで伝送しようという需要に応じて、AM 放送品質の符号化G.722⁽¹⁶⁾が1988年に勧告され、7 KHzのオーディオ信号を64, 56, 48 kbit/sで符号化する。G.722の後継には、遠隔会議用を主目的とし、遅延時間の制約をやや緩めたビットレート半減版がMDCTを元に1999年にG.722.1⁽¹⁷⁾として勧告された。これに次ぐ、G.722.2は欧州の地域標準であるAMR-WBと互換性のある方式をITUとして勧告化したものである。

3.2 個別勧告

3.2.1 G.728におけるフレーム消失隠蔽

G.728 LD-CELP (Low-Delay Code Excited Linear Prediction) 方式は、16 kbit/sで商用電話品質の達成する。エコー制御の不要な短遅延時間とするため、フレームは5サンプル(0.625 ms)であり、線形予測は自己の復号信号から算出するバックワード型をとる。ハイブリッド窓、聴覚重み付け、VQターゲット計算等、その後の標準に取り入れられる技術が使われている。CELP符号化器の基本的な構成を図2に示す。

図2 CELP 符号化の基本構成

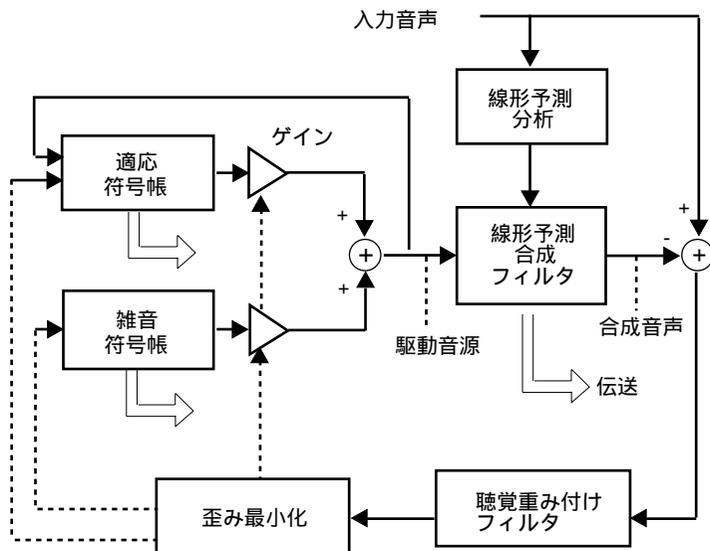


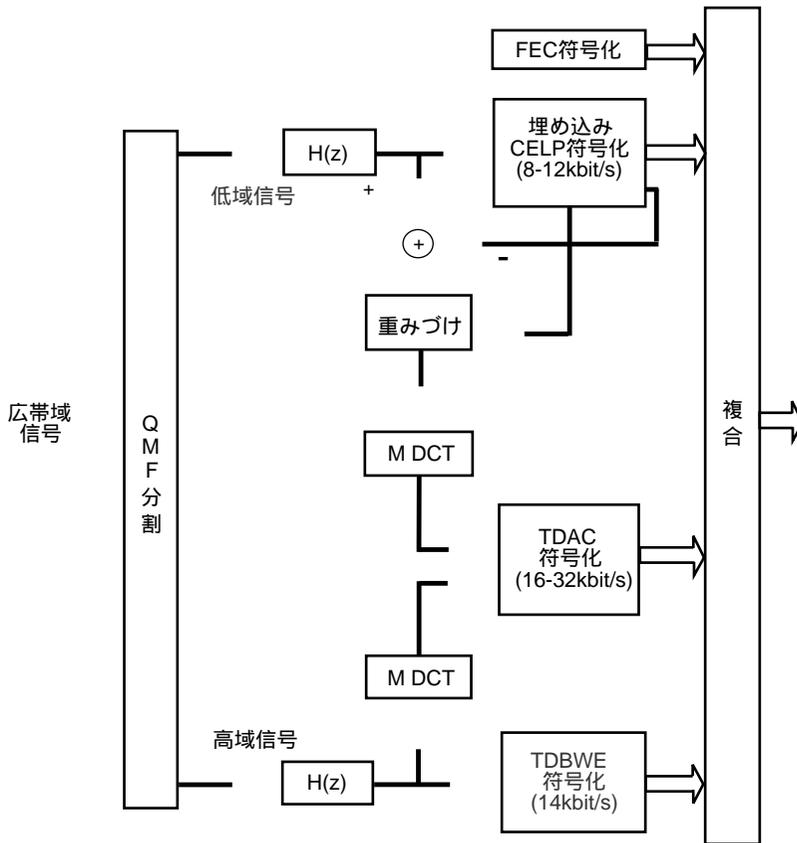
図1に示すような無線伝送路の符号誤りがバースト的に発生した場合、先述のように誤り訂正では対応しがたい。受信チャネルからバースト誤り検出情報を受けたとき、復号器は隠蔽動作を行う。これは、励振信号とLPCフィルタの外挿と、LPCと利得のバックワード適応化の連続性確保を基本とする。励振信号では、有声、無声の判定後有声ならば、最後の健常

フレームのピッチを保ち、振幅を緩やかに減衰させる。無声ならば、ピッチ性を避けるため過去の40サンプルから、利得調整した5サンプル組をランダムに使用する。LPC係数は最後の健常フレームの係数にQダンプ（帯域幅拡張）処理を施して外挿する。また、利得の予測には制限を施し、フレーム消失からの回復後に利得が跳躍することを防ぐ。このときの品質劣化の隠蔽度合いは筆者らが主観評価を実施し、その有効性を明らかにした⁽¹⁸⁾。

3.2.2 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「G.729を基にするエンベデッド可変レート符号化方式: G.729と相互接続可能な符号ビット列を含む階層構造8-32 kbit/s広帯域符号化方式」は、G.729が標準化当初主目的としたモバイル用途への適用から、むしろVoIPに多用され、電話のみならずPCを用いた拡声系、会議通話に利用される状況を踏まえ、旧来電話端末と対向可能としつつ、広帯域化に対応したいという要望に応えるべく設計されたエンベデッド（埋め込み）による階層構造広帯域符号化である。8, 12 kbit/sでは電話帯域を、14から32 kbit/sでは広帯域（50-7000 Hz）オーディオ信号を符号化し、電話帯域ではG.729およびそのAnnex A, Bと相互接続できる。符号化器の基本構成を図3に示す。QMFにより帯域の分割併合を行い、低域（電話帯域）はCELPを、高域（4-7 kHz）はTDBWE（Time Domain Bandwidth Extension）、全帯域の高精度化はTDAC（Time Domain Aliasing Cancellation）と称する変換（スペクトラム）を用いる。コアとなる電話帯域のCELP符号・復号化信号と原信号との時間波形の差分をMDCTを用いたスペクトル符号化により、追加ビットで順次高精度に符号化してゆく階層構造をとっている。また、パケットロスによる品質劣化を極力防ぐため、PLCに必要な情報を符号化時に追加している。これは、信号分類種別、エネルギー、ピッチパルス位置情報であり、正常フレームへの早期復帰と符号・復号間の状態乖離の緩和に役立つ。アルゴリズム遅延時間は48.9375 msである。その内訳は、フレーム構成（G.729フレームの2倍長でスーパーフレームと称する）に20 ms、MDCT分析に20 ms（先読み）、電話帯域の線形予測の先読みに5 ms、QMF分割・統合に3.9375 ms（782タップ）である。

図3 G.729.1 符号化器



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

G.718「フレーム誤りに強い8-32 kbit/s エンベッド可変レート音声、オーディオ符号化方式」は、IP、モバイル網への適用に柔軟に適應する階層構造をもつ可変帯域、可変ビットレート音声、オーディオ符号化方式であり、特にIP網で起こりやすいパケット消失への対応を強化し、既存のG.722.2方式との相互接続可能モードを有する符号化方式である。符号化器は表1に示すように5層の階層構造を有する。下位層1,2では電話帯域300-3400 Hzを8, 16 kHz サンプリングで入力し、内部サンプリングは12.8 kHzを使用する。上位層3-5では、下位層2の符号復号化の誤差信号を、広帯域(50-7000 Hz)でMDCTを用いて処理する。符号化器の構成を図4に示す。下位層1,2では超低域が除去されたあと、サンプリング変換、プリエンファシスを経てMSVQ (Multi Stage Vector Quantization) を含むCELP符号化が適用される。フレーム長は20 msであり、広帯域入力時のアルゴリズム遅延は42.875 msであり、10 msが符号器の先読み時間、10 msが復号器のMDCTオーバーラップ時間、1 msがポストフィルタ時間、1.875 msは入出力のサンプリング変換フィルタによる遅延である。狭帯域入力、狭帯域

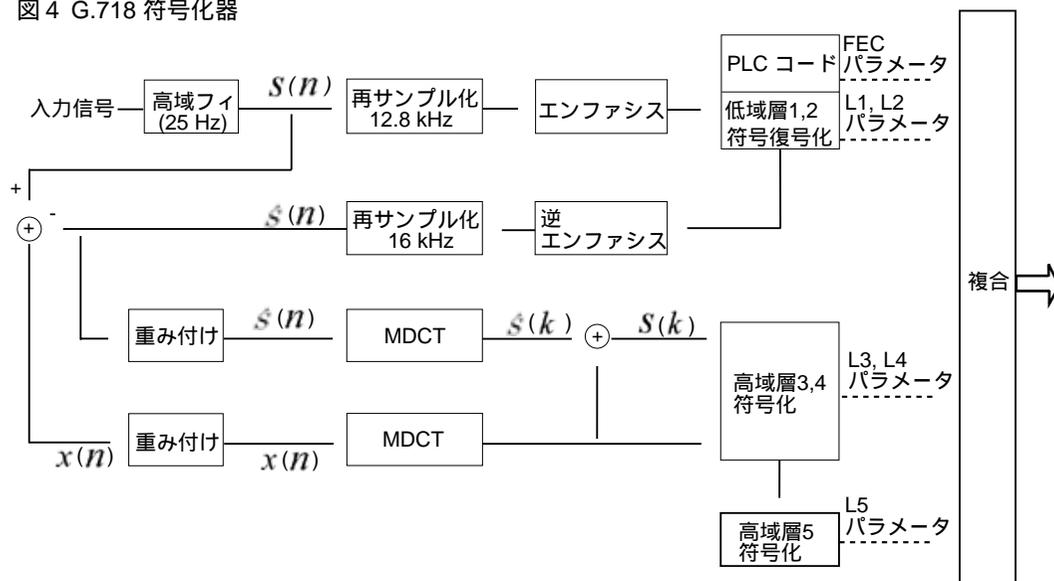
出力時の遅延時間は43.874 msであり、サンプリング変換に要する時間が増えている。MDCTを含まない下位層1, 2では遅延時間はこれより10 ms短い。この勧告の特徴は、無音圧縮機能DTX/SAD/CNGの強化による平均ビットレートの低減、G.729.1同様にFEC動作を助ける情報を符号器側から伝送し、パケットロスの品質劣化を最小に押える頑健性、また、ITU-T勧告G.722.2と共通テキストである3GPP AMR-WB、及び3GPP2 VMR-WBと相互接続可能なモード(最小ビットレート12.65 kbit/s)をオプションとして持つことであり、このとき標準構成との差異は僅かなものであり実現上の追加負担はほとんど無く、また、復号器はすべてのG.722.2のモードを復号できることにある。

表1 デフォルト構成の階層構造

階層	Bitrate	技術	サンプリング周波数
層1	8 kbit/s	分類に基づくコア層	12.8 kHz
層2	+4 kbit/s	ACELP エンハスメント	12.8 kHz
層3*	+4 kbit/s	FEC MDCT	12.8 kHz 16 kHz
層4*	+8 kbit/s	MDCT	16 kHz
層5*	+8 kbit/s	MDCT	16 kHz

* 狭帯域入出力には非適用

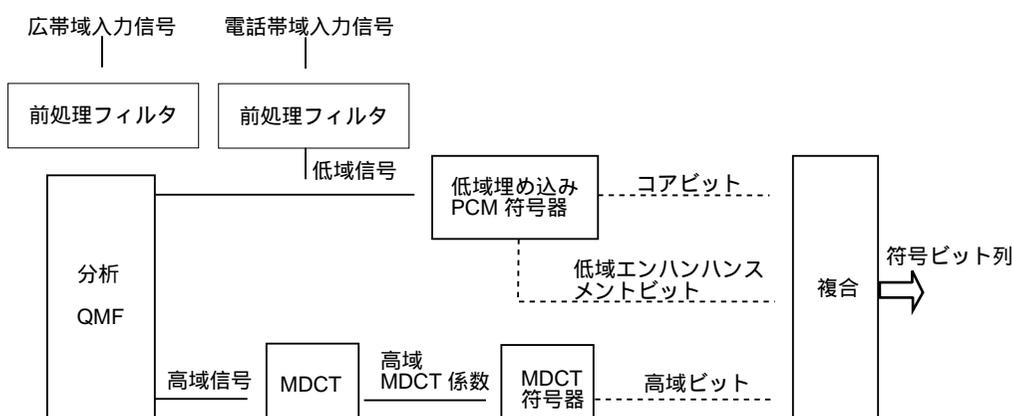
図4 G.718 符号化器



3.2.4 G.711.1 (Wideband embedded extension for G.711 pulse code modulation)

G.711はVoIPの必須符号化方式であり、電話端末は通常G.711により接続可能であるから、階層構造により高品質・広帯域化を図るとき、G.711をコアとすればもっとも多くの既存端末と相互接続可能となる。近年、IP接続の速度的ボトルネックであった加入者回線にも光やADSLが普及し、64 kbit/s以下に符号化速度を圧縮する要請は薄れてきた。一方、多地点会議通話で複数回線をブリッジ接続するには符号復号を複数こなさねばならず、PCで実現するには演算処理量を低減することに、より強い要求がある。そこで、IP会議通話を主目的とするG.711をコアとする階層構造による高品質・広帯域化方式の開発では、(1) 電話帯域音声の高品質化、(2) 7 kHz帯域への広帯域化、(3) 短遅延時間(5 msフレーム長)、(4) PCで実現可能な低演算量、(5) 頻繁な階層切り替えとパケット消失に対する頑健性、を条件として標準化が検討された。符号化の構成を図5に示す。サンプリング周波数は16 kHzであり、前処理は50 Hz以下を除く高域フィルタである。電話帯域のG.711(コア:階層0)符号復号信号と原音の誤差は階層1で16 kbit/sで符号化される。高域信号はMDCT変換され階層2で符号化される。復号器では、高域信号はiMDCTにより時間波形が再合成され、QMFにより帯域が統合される。QMF出力にはノイズゲートが適用され雑音が低減され、また、パケット消失時はPLC処理により音声信号が生成され品質劣化を最小限とする。電話帯域信号は8 kHzサンプリングで出力される。アルゴリズム遅延時間11.875 ms(190サンプル)である。

図5 G.711.1 符号化器



3.3 主観品質評価

前述のような音声符号化の背景となる通話環境や文化の変化に対応して、通話品質の評価法も通信網設計側の要求から、利用者満足度を明らかにする要求に応じられるよう変化してきたといえる。この観点から、ITU-Tにおける音声通話の主観品質評価手法の近年の傾向について述べる。環境変化とは、拡声系、マルチメディア系、広帯域系へとサービス内容が高度化すること、音声圧縮率が高く、使用条件が厳しい無線通話系と固定電話網と相互に通話が行われること、衛星通信、国際通信などの空間的拡がり、IP網とPSTNがゲートウェイで接続されて、サービス品質QoS（Quality of Service）非保証がありうること、もとよりPSTNを介さないIP通話がPC端末間で独自に行われることなどである。このため、通信サービス提供者からみたQoSよりも、利用者の品質感を反映するユーザ体感品質QoE（Quality of Experience）がより重視されるようになった。表2に拡張サービスと技術要素、特有の品質要因を示す。表中のVAD（Voice Activity Detection）は音声の休止区間を検出し、音声情報の伝送率を低下する、CNG（Comfort Noise Generator）は、VADにより省略された無音区間に雑音を付加し有音・無音の切り替えによる不自然さを低減する。

表2 拡張サービスと技術要素、特有の品質要因

通話系	技術要素	品質要素
拡声系	音響結合、音響エコーキャンセラ	エコー品質、エコーキャンセラ評価
マルチメディア系	画像との同期、協調	映像と音の時間差
広帯域系	帯域拡張、広帯域符号化	広帯域・高品質評価
無線モバイル網	高圧縮符号化、背景雑音、符号誤り、FERC、VAD/CNG	背景雑音時評価、FERC、VAD/CNG評価
IPゲートウェイVoIP	パケットジッタ、遅延時間、高圧縮符号化、背景雑音、符号誤り、PLC、VAD/CNG、QoS非保証混在	エコー品質、遅延時間、PLC、VAD/CNG評価

これらの要因はおおむねデジタル処理に関するものであり、新しい処理アルゴリズムが開発されるとき、その品質の特徴はユーザにとって未知であることが多い。そこで、これらの要因の評価手段として、QoEを表すのに適したオピニオン評価が主に用いられる。オピニオン評価の実施については、ITU-T勧告P.800⁽¹⁸⁾に整理された。基本的なオピニオン試験であるACR試験のほか、劣化の度合いを相対評価するためのDCR、CCR試験が用意されている。

3.3.1 ACR試験

オピニオン評価基準は、品質に応じて表3のように評点を配する。訳語に応じて微妙に意味が異なる可能性があり、国際間のMOS値の比較では、ずれが含まれることがある。

表3 ACRカテゴリ

評価カテゴリ	評点 Ci
Excellent 非常によい	5
Good よい	4
Fair まあよい	3
Poor 悪い	2
Bad 非常に悪い	1

3.3.2 DCR (劣化カテゴリ評価) 試験

P.800 Annex D に定義される。高品質な標準系に対し被試験系の劣化を劣化カテゴリで表す。被試験系の劣化が比較的小さく条件間に明確な差が得られないとき使用される。

3.3.3 CCR (比較カテゴリ評価) 試験

P.800 Annex E に定義される、DCR 試験法の変形である。提示されるカテゴリに標準系よりもよい評価が用意される。これは、被試験系に音声強調、雑音抑圧などの処理を含む場合に使われることがある。被験者には表のカテゴリと評点を示し、標準音の後に被試験音を一対にして提示し、後者が前者に比べてどうか意見を求める。

4 むすび

音声符号化の近年の動向を、ITU-T 標準化の状況から私見によりまとめた。音声符号化を取り巻く状況の変化から、要求条件も変遷し、標準化そのものの位置づけも変わってきている。圧縮符号化技術が一応の成熟に達したと見れば、産業界の関心はコンテンツ保護などの周辺技術に向かうと思われるが、新たなブレークスルーのもとになる技術のためにも、信号処理理論だけでなく、聴覚や感性にまつわる基礎的研究が望まれる。終わりに、本解説の執筆にあたりITU-Tの4 kbit/s 標準化活動以降の情勢について江原宏幸氏の学位論文⁽¹⁹⁾を参考にさせていただいた。

文献

- (1) CCITT Recommendation G.711, "Pulse Code Modulation (PCM) of Voice Frequencies", 1988
- (2) CCITT Recommendation G.726, "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)", 1990
- (3) CCITT Recommendation G.728, "Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction", 1992

- (4) ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction(CS-ACELP)", 1996
- (5) AMR speech codec, 3GPP 3G TS 26.09x V5.5.0.0, June 2002
- (6) 三樹、守谷、真野、大室、" ピッチ同期雑音励振源を持つ CELP 符号化 (PSI-CELP)、信学論A,Vol.J77-A, No.3, pp.314-324, 1994
- (7) ITU-T Recommendation H.323, "Packet-base multimedia communications system", June 2006
- (8) D. J. Goodman, G. B. Lockhart, O. J. Wasem, W-C. Wang, "Waveform Substitution Techniques for recovering missing speech segments in packet voice communication", IEEE Trans. On ASSP Vol.34, No.6, pp. 1440-1448, Dec. 1986
- (9) ITU-T Recommendation G.114, "One-way transmission time", May 2003
- (10) 伊藤憲三, 北脇信彦, " 会話音声の時間的特徴量に着目した遅延品質評価法," 日本音響学会誌, vol. 43, no. 11, pp. 851-857, 昭和62年11月.
- (11) ITU-T Recommendation G.729.1, "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729", 2006
- (12) ITU-T Recommendation G.711.1, "Embedded wideband G.711 extension", 2008
- (13) ITU-T Recommendation G.718, "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", 2008
- (14) ITU-T Recommendation G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", 2003
- (15) ITU-T Recommendation H.320, "Narrow-band visual telephone systems and terminal equipment", 2004
- (16) CCITT Recommendation G.722, "7 kHz Audio-Coding within 64 kbit/s", 1988
- (17) ITU-T Recommendation G.722.1, "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss", 2005
- (18) Shinji Hayashi, Mineyoshi Ogawa, "Standardization activity in ITU of extending 16-kbit/s LD-CELP for personal communication systems," IEEE Int. Conf. Univ. Personal Commn., Nov. 1995
- (19) 江原宏幸、" 高能率音声符号化方式とスケーラブル符号化に関する研究"、慶応義塾大学博士論文2008年2月