ITU-Tにおける音声符号化の最新標準化動向

佐々木 茂明（ささき しげあき）
NTTサイバースペース研究所

ITU-Tでは，音声通信を行ううえで欠かせない技術の１つである音声符号化において，電話向けの多数の標準方式を勧告化してきましたが，現在では市場の要求する製品・サービス向けに，再生できる音声帯域が広く，従来の電話標準とも互換性を有するG.711.1やG.729.1など，新たな国際標準方式の策定が積極的に進められています．

音声符号化と国際標準化

近年はフレッツ 光ネクストなど有線によるブロードバンドIPネットワークが広く普及し，携帯電話からのブロードバンドアクセスも提供されつつあります．また，それらIPネットワーク上での双方向音声通信サービスとして，「ひかり電話」等のVoIP（Voice over IP）が一般家庭やオフィスに普及しています．

アクセスネットワークがブロードバンド化されても，伝送帯域が限られたネットワーク上で，これらのサービスを多くのお客さまにすぐれた品質で提供するためには，信号を圧縮して伝送する符号化の技術が欠かせません．また標準の符号化方式を定め，標準に準拠した方式でサービスを提供することにより，音質，相互接続が保証されるというメリットがあります．

ITU-Tにおける標準化活動

ITU-T（International Telecommunication Union-Telecommunication Standardization Sector）とは，国際連合の専門機関の１つであるITUにおいて，電気通信の標準を策定する部門です．電話など双方向音声通信向けの音声符号化方式も，ITU-Tで標準化されてきました．

今会期（2009～2012年）中は，**図**の組織構成に示すように「マルチメディア符号化，システム及びアプリケーション」に関する標準化を担当するSG（Study Group）16以下に，各種「メディア符号化」を取りまとめるWP（Working Party）3，さらに検討課題ごとのQuestionがそれぞれ設置されています．これらのグループには多数の専門家が参加し，標準化活動を行っています．WP3では，主に以下のQuestionで音声メディアを取り扱っています．

- Q.8（汎用音声区間検出）：Q.8では，音声信号処理の前段として，音声区間を検出し，その区間の入力信号の種別（例えば，音声，音楽，雑音など）を判別できる方式を検討しています．音声符号化向けの標準G.720.1を策定した実績があります．
- Q.9（エンベデッド可変ビットレート音声符号化）：符号化のビットレートを可変にできる方式を検討するグループで，G.718を策定しました．目的を達成したため，2010年10月に活動を終了しています．
- Q.10（音声及びオーディオ符号化と関連ソフトウェアツール）：前会期（2005～2008年）のQ.10は既存標準のメンテナンスが主な役割でしたが，今会期からスコープが拡張され，ほぼすべての音声符号化の標準化を一手に引き受けるグループとなりました．また，標準化の作業に必要なツー

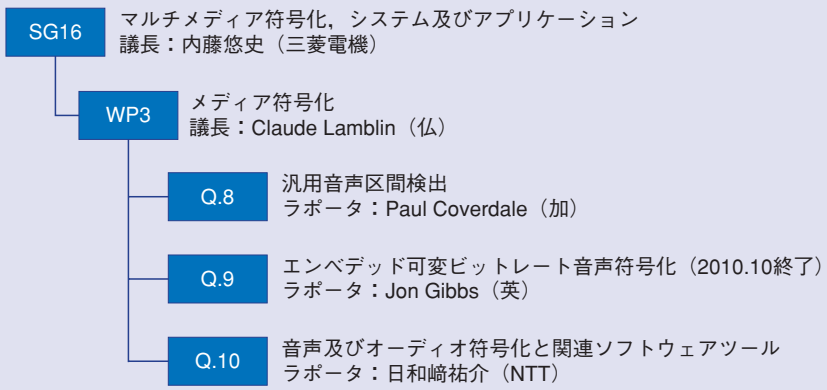

図　SG16における主な音声符号化関連課題

ル類をまとめたG.191の整備も重要な役割の1つです．

SG会合は，通常8～9カ月に1回の割合で開催されますが，市場の要求に迅速に対応できるよう，必要があればラポータ会合，WP会合などの中間会合を適宜開催し，標準成立までのスケジュールを短縮する努力がなされています．

音声符号化標準の最新動向

従来は，想定するサービス向けにターゲットとする音声帯域（電話帯域もしくは広帯域）とビットレートを定めて，標準化を行ってきました．しかし前会期以降，市場の要求する製品・サービスに，それに求められる要求条件を設定したうえで，標準化を進めるようになりました．その要求条件を達成するために，以下の要素が新たな特徴として標準化に取り入れられるようになりました．

- SWB音声，FB音声符号化：高音質，高臨場感を必要とする通信システム向けに，FM帯域相当のSWB（Super-WideBand, 50 Hz～14 kHz）音声やCD帯域相当のFB（Full-Band, 20 Hz～20 kHz）音声を符号化できます．
- スケーラブル符号化：ビットストリームを階層構造にすることで，伝送路の帯域や要求条件に応じて，ビットレートと音質を制御できます．従来方式のビットストリームをコアとして，帯域拡張や音質向上のためのビットストリームを追加できるよう，階層構造を設計すれば，コアのビットストリームを取り出すだけで，従来方式との相互接続が可能です．
- ロスレス符号化：圧縮しても原信号の情報を全く損なうことなく復号できる符号化方式です（従来の音声符号化は，高い圧縮率を得るために，圧縮による歪みを許容することから，ロッシー符号化とも呼ばれます）．

これらの特徴を有する，最近の標準方式を，標準化された順番に簡単に紹介します．

⑴ G.722.1 Annex* C

Polycom社の提案を基に，TV会議，音声会議向けに標準化された広帯域音声符号化G.722.1のアルゴリズムを拡張し，SWB音声を符号化できるように設計された方式です．ITU-Tで初めてのSWB音声符号化標準で，演算量が非常に軽いことが特徴です．基本的なアルゴリズムはG.722.1と共通ですが，ビットストリームの互換性はありません．

⑵ G.729.1

VoIP向け標準のG.729（8 kbit/s）をコアとしたスケーラブル符号化を採用し，広帯域音声を符号化できるとともに，コアのビットストリームをG.729で復号できます．8～32 kbit/sまで2 kbit/s単位で細かくビットレートを制御できることも特徴の1つです．France Telecom，VoiceAge，パナソニック，Siemensなど複数機関の共同提案に基づき標準化されました．

⑶ G.711.1

G.711がもっとも普及していることから，G.711との相互運用性に着目し，NTTの主導のもと，他4機関〔ETRI，France Telecom，華為（ファーウェイ）技術，VoiceAge〕との共同で提案した方式が，標準として採用されました．G.711をコアとしたスケーラブル符号化により，広帯域音声を再生でき，G.711との相互接続も容易です．64 kbit/sとビットレートの高いG.711をコアとするため，最大96 kbit/sとビットレートは高めですが，ブロードバンドネットワークでの利用に限定すれば圧縮率はさほど問題ではありません．G.711との後方互換性により本標準とG.711とが混在する環境での利便性が確保されるとともに，既存の広帯域音声符号化標準の中でもっとも優れた音質が得られています．また本標準は，「フレッツ 光ネクスト（ライト）」上で提供されているひかり電話の高音質電話サービスのコーデックとして，利用されています．

⑷ G.718

スケーラブル符号化の手法を用いて，8～32 kbit/sで可変ビットレートと広帯域符号化を実現しています．ビットレートやスケーラブル符号化の枠組みはG.729.1と同様ですが，8 kbit/sのコアレイヤは広帯域符号化として新規に設計されており，既存の電話標準との互換性という制約がない分，効率的な圧縮が可能といえます．これはEricsson，Motorola，Nokiaなどが参加するオープンコンソーシアムで検討された標準です．

⑸ G.719

ITU-Tで初めてFB音声に対応した標準です．G.722.1および同Annex Cをベースに，Polycom，Ericsson 2社による共同提案が採用されました．G.722.1や同Annex Cと同様に，会議システム向け用途であること，演算量が非常に軽いことが特徴です．

⑹ G.711.0

ITU-Tで初めてロスレス符号化の概念が導入された標準です．いったんG.711で符号化された符号自体を歪みなしに再度符号化します．つまり，G.711の音質を完全に保ったまま，さらに圧縮することが可能です．符号化対象信号（ここではG.711符号）の冗長性によって，歪みなく圧縮できる限度が異なるため，ビットレートを定めることはできませんが，平均で50％程

* Annex：勧告本体の一部となる付録勧告．勧告本体と何らかの関連があれば，本体の一部としてシリーズ化可能．

度に圧縮できることが確認されています．ルータ等ネットワーク機器に実装し，基幹網の複数チャネルのG.711符号に適用することで，音声に要する伝送帯域を節約することが期待できます．NTT，華為技術，Ciscoなどによる共同提案が標準として採用されています．

⑺ G.718-SWB/G.729.1-SWB

G.718とG.729.1とは，音声帯域，ビットレート，スケーラブルな構造などお互い共通な点を有しています．当初，SWB音声を符号化できる拡張方式の提案が別々になされましたが，特徴が似ているため，SWBに拡張するレイヤを共通化する試みがなされました．その結果，4～16 kbit/sのSWB共通レイヤが完成し，G.718と組み合わせたものがG.718 Annex B，G.729.1と組み合わせたうえにG.729.1専用の広帯域レイヤ（8または16 kbit/s）を追加したものがG.729.1 Annex Eの正式名称でそれぞれ成立しました．

⑻ G.711.1-SWB/G.722-SWB

G.718-SWB/G.729.1-SWBと同様に，広帯域符号化標準のG.711.1とG.722に，16 kbit/sの拡張レイヤを2段追加し，SWB音声の再生を目指します．2段のレイヤはともに，G.711.1，G.722に共通なSWBサブレイヤと，別々に設計された広帯域サブレイヤで構成されており，それぞれのサブレイヤには，第一レイヤで12.2 kbit/s（SWB）+3.8 kbit/s（広帯域），第二レイヤで8 kbit/s（SWB）＋8 kbit/s（広帯域）のビットが配分されています．これらをG.711.1に適用したものがG.711.1 Annex D，G.722に適用したものがG.722 Annex Bの正式名称で成立しました．本標準化は，G.711.1と同じくNTT主導のもと，ETRI，France Telecom，華為技術，VoiceAgeの5機関によって検討されました．

以上の標準について，用途，ビットレート，遅延，演算量などのスペックを**表**にまとめます．

さらに新たな動向として，ステレオ通信向けの標準も検討されています．具体的には，上記の⑺，⑻のSWB音声符号化の方式について，TV会議，テレプレゼンスなど高臨場を必要とする通信システムでの利用を想定し，ステレオに拡張する方式が，新たな試みとして進められています．⑺は2012年度以降，⑻については2011年度中の成立が予定されています．

ちなみに，これらの符号化標準は，電話機など端末を想定し，比較的安価なDSP（Digital Signal Processor）チップに実装することが考慮さ

表　前会期以降（2005年～）に成立した音声符号化標準

	G.722.1 Annex C	G.729.1	G.711.1	G.718	G.719	G.711.0	G.718 Annex B	G.729.1 Annex E	G.711.1 Annex D	G.722 Annex B
成立時期	2005.5	2006.4	2008.3	2008.6	2008.6	2009.9	2010.3		2010.11	
音声帯域	50 Hz～14 kHz	50 Hz～7 kHz	50 Hz～7 kHz	50 Hz～7 kHz	20 Hz～20 kHz	300 Hz～3.4 kHz	50 Hz～14 kHz		50 Hz～14 kHz	
ビットレート（kbit/s）	24, 32, 48	8～32	64, 80, 96	8～32	32～128	不定（平均して約50%の情報量に圧縮）	36～48	36～64	96, 112, 128	64, 80, 96
フレーム長（ms）	20	20	5	20	20	5～40	20	20	5	5
符号化遅延（ms）	40	48.9375	11.875	43.875	40	フレーム長と同じ	49.625	55.6875	12.8125	12.3125
演算量（WMOPS）	11	35.8	8.7	57	21	1.67	80	63	21.498	22.76
主な要素技術	MLT	G.729, MDCT, BWE, SVQ	G.711, MDCT, Interleave VQ	ACELP, MDCT, AVQ	Adaptive MDCT, FLVQ	Lossless compression	G.718, MDCT, Sinusoidal coding	G.729.1, MDCT, Sinusoidal coding	G.711.1, MDCT, BWE, AVQ	G.722, MDCT, BWE, AVQ
主な用途	会議システム，テレプレゼンス	VoIP	NGN	VoIP	会議システム，テレプレゼンス	基幹網の帯域圧縮	会議システム，テレプレゼンス		会議システム，テレプレゼンス	
その他の特徴		G.729と相互接続可	G.711と相互接続可			G.711符号ビットをロスレス圧縮	G.718と相互接続可	G.729, G.729.1と相互接続可	G.711, G.711.1と相互接続可	G.722と相互接続可

れており，そのアルゴリズムは固定小数点形式の演算で記述されています．固定小数点版の標準化が完了後，PCへの実装やアルゴリズム理解の手助けを目的とし，固定小数点版との互換性が保証された浮動小数点版をAnnexとして標準化することも，普及の一環として行われています（例えば，G.711.1の浮動小数点版はG.711.1 Annex Aとして標準化されました）．

その他の関連標準

前述したとおり，相互接続や互換性を保証することが通信の標準化の本来意図するところである一方，相互接続に影響しない技術でも，標準化されたものを実装すれば，品質，性能を容易に保証できるという考え方があります．PLC（Packet Loss Concealment），VAD（Voice Activity Detector）のように相互接続には影響しないものの，性能要求条件を満たすのに必要な技術は，符号化アルゴリズムの一部として組み込まれるのが一般的ですが，必要に応じて，技術単体でAppendix（勧告本体を補う付録勧告）として標準化されることもあります．

① PLC（パケットロス補償）：VoIPなど音声ビットストリームをパケットに格納して伝送する場合，何らかの原因でパケットが消失した場合，音の途切れとして聞こえてしまいます．そのような場合でも，音声の途切れを補間するのがPLCです．パケットでの音声伝送が想定されなかったころに標準化された方式にはPLCは含まれないため，その場合はAppendixとして標準化されます．近年策定されたものに，広帯域音声符号化標準G.722用のPLCであるG.722 Appendix III，G.722 Appendix IV（2006.11成立）があります．

② VAD（音声区間検出）：音声・無音など入力信号の状態を識別し，無音区間は情報を伝送しないなど，信号の特性にふさわしい符号化を行い，効率的な伝送を行うことが可能です．この音声区間かどうかを判別する技術がVADです．G.729 Annex BやG.722.2など従来のVAD付きの符号化標準では，その標準向けに特化したVADと判別情報に基づいた無音圧縮など符号化処理が併せて記述される一方，Q.8では，符号化とは独立かつ汎用的に利用できるVADをGSAD（General Sound Activity Detector）と定義し，華為技術の主導により，G.720.1（2010.1成立）を標準化しました．本標準では，音声・無音だけでなく，音声・音楽・雑音・無音の4つの状態を，10 ms単位で判別することが可能です．

また，方式を標準にしたもの以外に，標準化作業をサポートする目的で整備されたものもあります．標準化作業では，要求条件を満たすかどうかを確認するため，厳密な試験・測定が行われます．それらに必要なレベル調整，サンプリング変換，フィルタリング，エラーパターン作成，固定小数点化ライブラリなどの各種ツールをパッケージ化したものがG.191として標準化されています．これは定期的に更新されており，STL（Software Tools Library）-2009が現行版になります．

標準の利用について

最後に，ITU-Tの音声符号化標準の利用方法について述べます．発行準備中のものなど一部を除き，ITU-Tのサイト（http://www.itu.int/rec/T-REC/en）から標準勧告をダウンロードすることができます．音声符号化の標準の場合，勧告本体のテキストだけでなく，サンプルソースコードが添付されており，音質や相互接続性などの性能を手軽に確認することが可能です．

標準化作業自体や性能評価などの目的には，それら標準勧告を自由に利用することができます．ただし標準といえども，商用利用などの場合はパテントやソフトウェアコピーライトなど知的財産を保有するライセンサから利用許諾を得なければなりません．利用許諾を得るためには，ライセンサの定める適切な対価を支払う必要があるのが一般的です．前述したように，最近の標準は共同提案によるものが多く，複数のライセンサから許諾を得なければなりませんが，個別に契約することなく，ワンストップで利用許諾が得られるよう，ライセンサがパテントを持ち寄るパテントプールの仕組みが広く用いられています．

今後の取り組み

ITU-Tでは，今後も市場のニーズに基づき，マルチチャネル，NGN（Next Generation Network）・FMC（Fixed Mobile Convergence）など新たな音声符号化の課題について，標準策定に取り組むとともに，符号化に限らず，通信品質を向上させるための音声・音響技術の標準化を進めていきます．

NTT研究所もまた，NGNでのサービス向けにG.711.1の標準化を主導し，現在のQ.10ラポータを務めるなど音声符号化の標準化に貢献してきましたが，今後もお客さまへよりよい通信サービスを提供できるよう，積極的に標準化活動に取り組んでいきます．