



14 kHz帯域音声符号化の 国際標準ITU-T G.711.1 Annex D

NTTメディアインテリジェンス研究所

ふく い まさひろ さ さ き しげあき ひ わ さ き ゆうすけ くりはら さちこ
福井 勝宏 / 佐々木 茂明 / 日和崎 祐介 / 栗原 祥子

NTTは、ITU-T標準G.711.1を拡張した14 kHz帯域スケーラブル音声符号化方式の開発に取り組み、国際標準の策定を主導してきました。ITU-T G.711.1 Annex Dとして成立した本方式は、双方向での臨場感に溢れた音声通信を可能とし、G.711/G.711.1とも相互接続できることから、幅広い利用が期待できます。

G.711.1 Annex D 標準化の背景

現在、電話帯域音声（300 Hz～3.4 kHz）のコーデック^{*1}であるITU-T（International Telecommunication Union - Telecommunication Standardization Sector）標準G.711^{*2}が、デジタル電話やVoIP電話に広く利用されています。G.711で符号化される電話帯域音声は、会話を行うには十分なものの、周波数帯域が狭いため、明瞭感や肉声感に欠けるものでした。これに対しNTT研究所は、NGN（Next Generation Network）商用サービスで広帯域音声（50 Hz～7 kHz）を用いたサービス展開を目指し、2008年に、G.711をコアとした広帯域スケーラブル符号化^{*3}方式をITU-T G.711.1^{*4}として成立させました。G.711.1は、G.711との相互互換性を確保しつつ、明瞭で聞き取りやすい広帯域音声を再生できます。また、最小5 ms単位での音声処理が可能のため、音声の遅延を抑えられる設計になっているのも特徴です。現在、G.711.1は、フレッツ光ネクスト/ライトの高音質電話サービスに採用され、ひかりフレッツフォンVP3000やひかりクリアフォンHQ-100などの端末に導入されています。

一方、テレプレゼンスやハイビジョン

通信会議などの高臨場な双方向通信サービスでは、人間の声や背景音を忠実に再現できる、より広い帯域の音声コーデックが必要とされています。これまで、放送やコンテンツ配信向けに超広帯域音声（50 Hz～14 kHz）の符号化技術は標準化されていましたが、既存サービスで利用されているコーデックと互換性がないという問題がありました。また、量子化した音声の品質が音楽と比べ比較的低いことも知られています。このため、ますます普及が見込まれる高臨場感通信サービスに適用可能な、音声に強く、超広帯域で後方互換性を有する符号化技術が次代の有望なコーデックとして期待されています。

標準成立までの経緯

2007年6月、France Telecom Orangeは、ITU-T専門家会合において、NG-DECT（New Generation Digital Enhanced Cordless Telecommunication）^{*5}規格向けに、「ITU-T G.722^{*6}とのスケーラビリティを保ちつつ、7 kHzを超える音声帯域に拡張する符号化方式」の標準化開始を要求する提案を行いました。これに対し、高臨場感双方向通信サービスの展開を目指していたNTT研究所は、2007年10月に、音質、周波数帯域、伝送遅延などの面でG.722と性質の似たG.711.1

も音声帯域拡張することを提案しました。両者の提案を踏まえ、G.711.1またはG.722のいずれかをコアとし、実装効率化のため、14 kHzまで音声帯域を共通のモジュールを用いて拡張する方式の標準化が開始されました。そしてそれ以降、NTT主導の下に新コーデックの標準化が進められました。NTTほか4社〔ETRI（韓国）、France Telecom Orange、Huawei Technologies（中国）、VoiceAge（カナダ）〕による最終共同提案は、各社の技術を統合して開発されたものです。ITU-Tで実施された主観品質評価による性能確認試験では、

- *1 コーデック：音・画像などの情報をデータに変換する符号化。そのデータを情報に戻す復号の処理を併せて行える装置、ソフト、方式。
- *2 G.711：ITU-Tで標準化された最初の音声符号化方式。8 kHzサンプリングされた信号を、サンプルごとに8 bit対数圧伸PCM量子化を行い、64 kbit/sで伝送できます。
- *3 スケーラブル符号化：ビット列の一部からでも信号が復号できるように、階層的に符号化を行います。
- *4 G.711.1：NTT研究所主導の下、ITU-Tで標準化された7 kHz広帯域音声符号化方式。G.711の符号データ（64 kbit/s）に、音声帯域を広げるための符号データ（16～32 kbit/s）を追加して伝送することにより、G.711との互換性を保ちつつ、高音質を実現できます。
- *5 NG-DECT：ETSI（European Telecommunications Standards Institute）が策定した次世代のデジタルコードレス電話規格。広帯域音声で通話ができます。
- *6 G.722：48 kbit/s、56 kbit/s、64 kbit/sの広帯域音声コーデックのITU-Tによる勧告。帯域分割ADPCM（Adaptive Differential PCM）に基づいて量子化されます。

品質要件 (Requirement) のすべてを満たすことが確認されました。この結果を踏まえ、2010年11月に正式な勧告として承認されました。

G.711.1 Annex Dの特徴

(1) 14 kHz帯域音声を再生可能

従来の電話音声や7 kHz広帯域音声よりもはるかに広い帯域の音声により、人の声のみならず、環境音や音楽まで忠実に再生することができ、相手に話し手の雰囲気まで伝えることが可能になります。

(2) 双方向通信に適した少ない遅延

携帯電話の符号化方式3GPP (3rd Generation Partnership Project) AMR^{*7}や放送用途の方式MPEG (Moving Picture Experts Group) AAC^{*8}では、音声信号を20 ms以上のブロック単位で処理するため、音声の遅れが避けられません。本方式は、G.711.1と同様に、最小5 ms単位での音声処理 (アルゴリズム遅延は12.8125 ms) が可能なため、音声の遅延を抑えられます。これにより、遅れの少ない自然な通話が可能となります。

*7 AMR: 3GPPで標準化された携帯電話向けの音声符号化方式。NTTドコモのFOMAなどW-CDMA (Wideband-Code Division Multiple Access) 対応の携帯電話で広く用いられています。

*8 AAC: ISO (International Organization for Standardization) で標準化された音声符号化方式。デジタル放送、コンテンツ配信・蓄積などに広く用いられています。

(3) 従来電話との相互接続が容易

従来、複数の地点を接続するような会議では、すべての端末が同じ音声コーデックを搭載している必要がありました。従来電話が一地点でも混在した場合、すべての端末のコーデックをG.711に合わせるか、装置間で符号データを変換する装置を別途設ける必要があります。本方式は、G.711、G.711.1で再生可能な符号データに、音声帯域を広げるための符号データ (14 kHz帯域拡張符号データ) を追加して伝送するため、それぞれの端末で再生できる符号データを取り出すだけで符号変換が可能です。した

がって、再エンコードなしにビットストリームの一部を切り捨てることにより、G.711/G.711.1と相互接続できます (図1)。

G.711.1 Annex Dの技術構成

表に、G.711.1 Annex Dの動作モード、ビットレートおよび拡張モジュールの関係を示します。表に示すように、G.711.1 Annex Dでは、ビットストリームの組合せにより、ビットレート (品質) と帯域がそれぞれ異なる8つのモードが利用できます。基本ブロック構

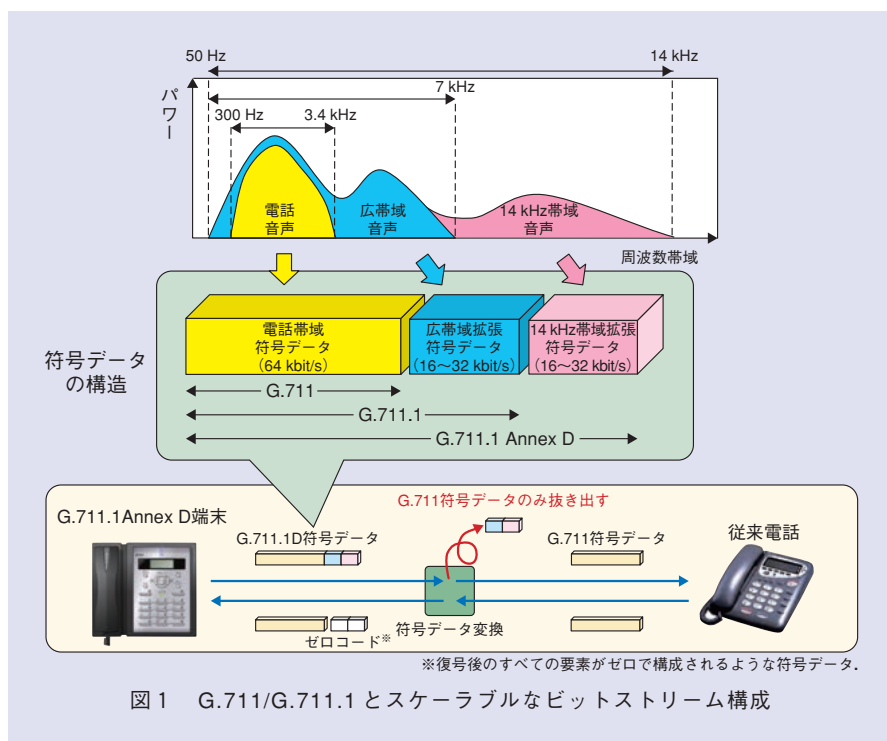


図1 G.711/G.711.1とスケーラブルなビットストリーム構成

表 モードごとのビットストリーム構成 (黄: G.711互換/青: G.711.1/赤: G.711.1 Annex D)

	モード	ビットレート (kbit/s)	サンプリング レート (kHz)	64 kbit/s	16 kbit/s	16 kbit/s	16 kbit/s	16 kbit/s
G.711.1	R1nm	64	8	L0 (G.711互換)				
	R2nm	80	8		L1			
	R2wm	80	16		L2			
	R3wm	96	16		L1	L2		
G.711.1 Annex D	R3sm	96	32		L2	G711 EL0	SWB L0	SWBL1
	R4sm	112	32				G711EL1	SWBL2
	R4*sm	112	32		L1	L2	G711 EL0	SWB L0
	R5*sm	128	32					SWBL1
							G711EL1	SWBL2

成を図2に示します。32 kHzサンプリングの入力信号はエンコーダで5 msフレームごとに処理されます。まず、プリプロセッシングフィルタで0~50 Hzの周波数成分を除去し、QMF (Quadrature Mirror Filter) で低域信号と高域信号に帯域分割します。低域信号は、コアとなるG.711.1エンコーダ〔表のL0 (G.711互換), L1 (狭帯域付加), L2 (広帯域付加) の3種類の符号データで構成〕, 4~8 kHz帯域の音質を改善するモジュールであるG711EL0エンコーダとG711EL1エンコーダにより符号化されます。G711EL0エンコーダは、G.711.1がエンコードしない7~8 kHzの帯域をベクトル量子化で符号化します。G711EL1エンコーダは、4~8 kHz帯域の量子化誤差が少なくなるよう動的ビット割当てとスカラ量子化を用いて符号化し、音質を改善します。

高域信号は、MDCT (Modified Discrete Cosine Transformation) 変換後、8~14.4 kHzまでの帯域が8サブバンドに分割され、SWBL0/SWBL1/SWBL2のマルチレイヤエンコーダで符号化されます。入力成分を概形〔サブバンドごとに計算されたRMS (Root Mean Square) 値〕と微細構造 (RMS値で正規化された入力MDCT係数) に分け、SWBL0で概形を、SWBL1/SWBL2で微細構造をそれぞれベクトル量子化します。

デコーダ側では、受信したビットストリームを復号し、得られた低域信号と高域信号をQMFで合成することで、出力信号を得ます。以上のようにエンコーダ、デコーダを構成することにより、G.711.1とスケーラブルな構成を実現しました。

本標準では、音声だけでなく音楽も含めたさまざまな信号を高品質に符号化することが求められました。次に、NTTから提案してSWBL1/SWBL2レイヤに採用された技術を紹介します。

従来方式と新方式の概念を図3に示します。AVQ (Algebraic Vector Quantization) は、入力を音声の母音区間

のように周波数要素の多くがゼロである信号 (疎な信号) と仮定してモデル化することで、コードブック (代表パターンの集まり) 探索を高速化しています。しかしながら、音楽など、非ゼロ要素がほとんどである信号 (密な信号) を符号

化すると、周波数成分が欠損し、しばしば聴感上の歪みが生じました。この問題を解決するため、本方式では、入力信号が疎な信号か密な信号かを判別し、方式を切り替えて符号化します。入力信号が密な信号の場合、入力の大きさの平均と

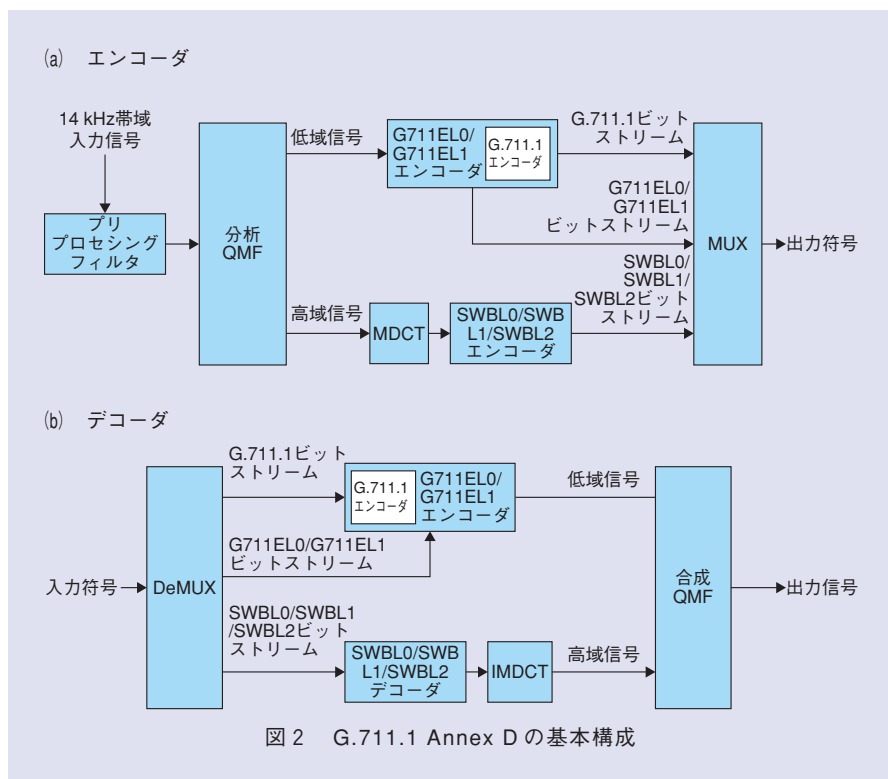


図2 G.711.1 Annex D の基本構成

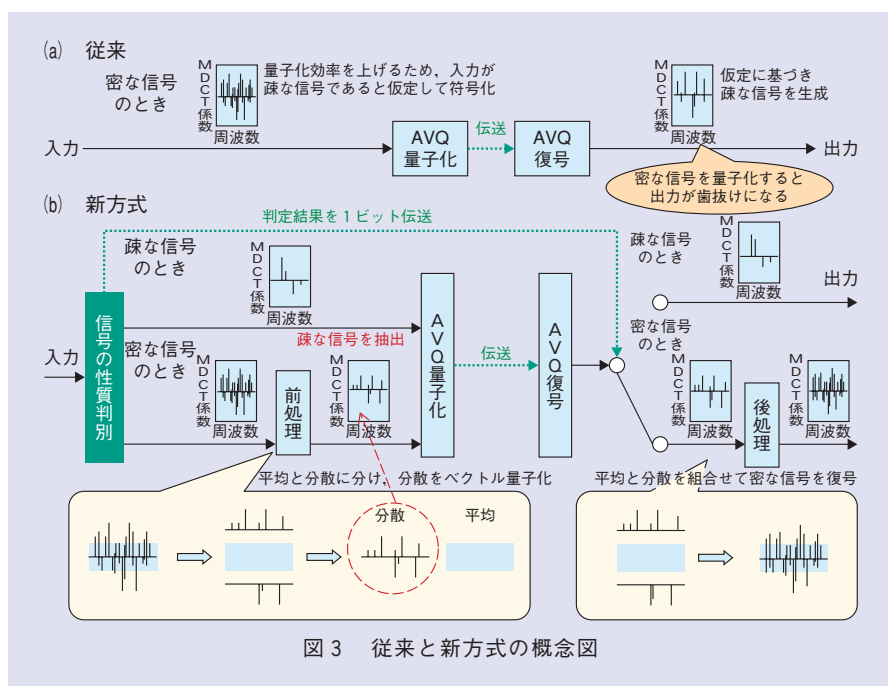


図3 従来と新方式の概念図

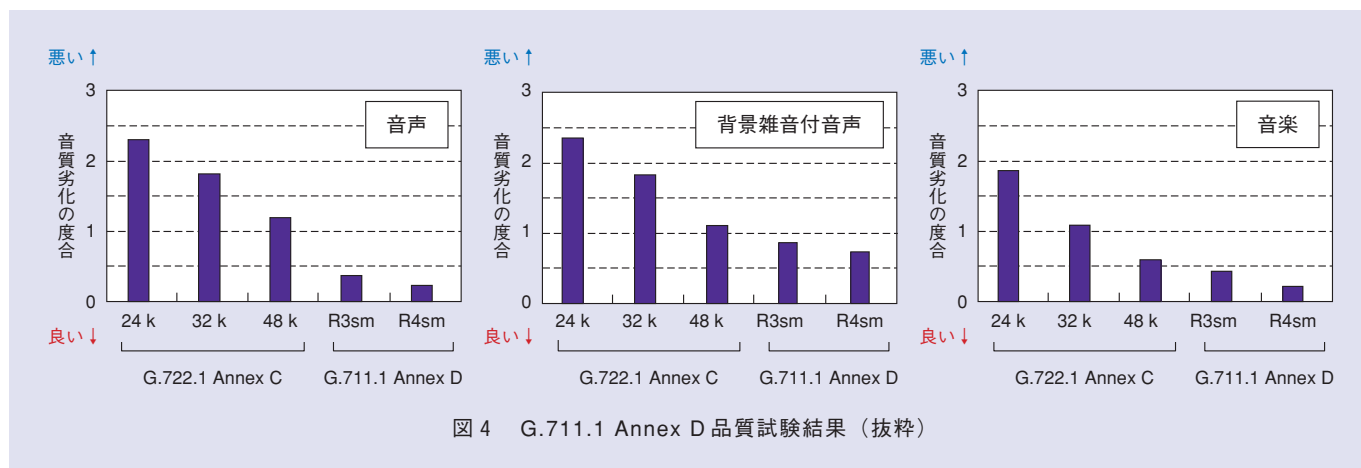


図4 G.711.1 Annex D 品質試験結果 (抜粋)

分散（入力の大きさから平均を引いた正の値の成分にMDCT係数の極性を付与して算出）に分け、分散をAVQで符号化します。平均は符号化せず、SWBL0で量子化した概形から計算します。デコーダ側では、復号した分散と平均を組み合わせて、分散の極性情報またはランダムに生成した極性情報を付与することで量子化出力とします。このようにして周波数成分の欠損を防ぐことで、従来AVQが不得手であった音源に対して高い音質を達成しています。

品質評価

ITU-Tでは、提案された符号化方式を勧告として承認する前に、品質要件を満たしているかどうか確認するため、主観評価試験を実施します。本標準でも品質要件として定められたさまざまな条件下（音声、背景雑音付音声、音楽）での主観評価試験が実施されました。品質評価の試験方法には、ITU-R (ITU-

Radiocommunication Sector) 標準 BS.1116-1が用いられました。BS.1116-1は、リファレンス音（原音）に対して、2つの評価対象音（1つはリファレンス音と同じ）から、どちらの音が符号化された音で、どれだけ劣化しているかを評価する試験方法で、劣化が少なく、原音との差が小さな音の評価に適しています。

図4は、公式試験結果⁽¹⁾からの抜粋です。グラフは、量子化による音質の劣化度合いを表し、品質が良いほど値が小さくなります。G.711.1 Annex Dにおいて、R3sm (96 kbit/s)、R4sm (112 kbit/s) の両モードで、音声、背景雑音付音声、音楽のすべてのカテゴリに対し、既存の14 kHz帯域音声符号化方式であるITU-T G.722.1 Annex C 48 kbit/sを上回ります。

その他の拡張標準

2011年の3月に、G.711.1 Annex DのMS (Mid-Side) ステレオ^{*9}拡張標準が、G.711.1 Appendix IVとして成立しました。また、さらなる拡張標準として、2012年5月のITU-T SG (Study Group) 16^{*10}会合にて、G.711.1 Annex Dの浮動小数点版^{*11}とレイヤードステレオ拡張版^{*12}を勧告化することの同意が得られました。これらは、それぞれG.711.1 Annex E, Annex Fとして、2012年10月に標準が成立する予定です。

今後の展開

今後NTT研究所では事業会社と連携し、これまで以上の高品質音声通信をお客さまに提供できるよう、本技術のサービスへの実装検討を行っていきます。これら実装検討と並行して、IETFにおいてG.711.1 Annex DのRTPペイロードフォーマットの登録を行い、他社の特許を含む本標準をワンストップで利用できるように仕組みを立ち上げる予定です。

参考文献

- (1) ITU-T WP3/16 TD289: "LS on G.722-SWB and G.711.1-SWB extension and G.718 post-processing," ITU-T Q7/SG12 rapporteurs, Geneva, Switzerland, July 2010.



(左から) 日和崎 祐介/ 福井 勝宏/
佐々木 茂明/ 栗原 祥子

ここで紹介した技術が、皆さんに広く利用していただけるようになれば幸いです。

◆問い合わせ先

NTTメディアインテリジェンス研究所
音声言語メディアプロジェクト
TEL 0422-59-2113
FAX 0422-60-7811
E-mail fukui.masahiro@lab.ntt.co.jp

*9 MSステレオ：既存のモノラル信号との互換性を保つために、主信号をモノラル音声 (L+R)、副信号を左右の音声の差 (L-R) とした信号。ステレオ音声には、 $\{(L+R)-(L-R)\}/2=R$, $\{(L+R)+(L-R)\}/2=L$ として変換されます。

*10 ITU-T SG16: 「マルチメディア端末・システムおよびアプリケーション」の標準化を担当する研究委員会。

*11 G.711.1 Annex Dの浮動小数点版：G.711.1 Annex Dは固定小数点で記述されており、浮動小数点版は別標準として勧告化されるのが通例。

*12 レイヤードステレオ拡張版：少ない情報を付加することでモノラル信号をステレオ化できる拡張標準。