音声音響符号化技術と3GPPでの標準化

守谷 健弘（もりや たけひろ）
NTTコミュニケーション科学基礎研究所　守谷特別研究室長　NTTフェロー

EVS（Enhanced Voice Services）規格は将来のVoLTE（Voice over Long Term Evolution）を想定し，2014年末に有力12機関によって共同開発された3GPPの音声音響符号化標準です．品質評価の結果，EVS規格は従来の携帯電話とほぼ同じ情報量で，従来の4倍の音声帯域を出力でき，さらに音楽の品質の格段の向上が確かめられました．このためEVSはVoLTEのみならず多様な音声サービスへの応用が期待されています．本稿では，音声音響符号化技術の概要とEVS規格の符号化技術を紹介します．

音声音響符号化技術

音声音響符号化技術はさまざまな基盤技術に支えられつつ，すでに多くの実用用途に使われてきています．1つは携帯電話やIP電話向けの双方向通信用の音声符号化です．これは通常低ビット，低遅延，音声専用の符号化で帯域は狭いものです．もう1つは放送や蓄積のための片方向通信用の音響符号化です．これは一般にビットレートが高く，遅延も大きいですが，音声帯域も広く，音楽なども含めた高品質の符号化が中心です．これらの2つの符号化は圧縮率優先の高圧縮符号化で，品質はできるだけ劣化させないことをねらいますが，これとは別にロスレス符号化のカテゴリがあります．これは歪を許さないという拘束の中で情報量を削減することをねらうものです．これらの音声音響符号化技術の開発経過を**図1**に示します．NTT研究所では40数年前から世界に先駆けて音声のデジタル信号処理の研究に取り組み，

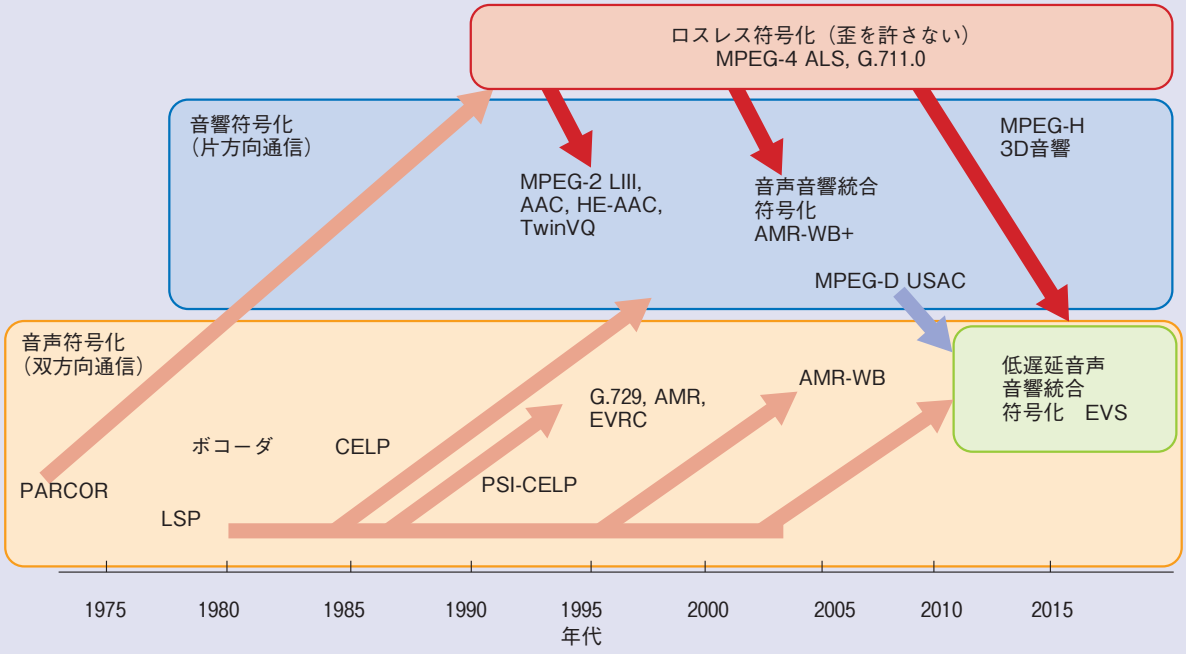

図1　音声音響符号化の開発経過

PARCOR（Partial Auto Correlation）やLSP（Line Spectrum Pair）といった世界的に優れた技術を考案してきました．中でもLSPは現在でも世界のほとんどの携帯電話に使われ，2014年には電気電子通信分野の技術の世界遺産に相当するIEEE Milestone[1]に選定されました．

1990年代にはITU-Tなどで音声符号化の国際標準化が盛んに行われ，デジタル携帯電話，IP電話の実用化が進展しました．音響符号化の分野でも1990年代にMPEGで国際標準化が行われ，デジタル放送，音楽配信，プレーヤなどで広く使われるようになりました．

21世紀になると歪のない符号化の国際標準化も行われ，長期保存や配信に使われ始めています．2014年にARIB規格，総務省令で，日本の超高精細度テレビジョン放送の高音質サービス用途に音響ロスレス符号化の国際標準であるMPEG-4 ALS（Audio LoSsless Coding）[2]が選定されました．情報量は約半分にしか圧縮できませんが，放送スタジオで丹念につくり込まれた音声や音楽がありのまま，衛星放送やIPTVで家庭に届けられるようになる見込みです．またALSはサンプリングレート，遅延，チャネル数が自由に選択できるので，放送だけでなくさまざまな高品質コンテンツの伝送蓄積に柔軟に利用できます．また，医療用信号，環境センサ信号，変調後の電波のデジタル信号の圧縮にも利用できます．

さらに21世紀以降，3GPPのAMR-WB+[3]やMPEG-D USAC[4]といった音声音響の統合符号化が標準化されるようになりました．ただいずれも長いフレーム長を要する遅延の大きい音響符号化に音声符号化を統合したものです．これに対し，2014年に完成した3GPP EVS（Enhanced Voice Services）は双方向通信に使える遅延の短い音声符号化に，音響符号化を統合し，双方向通信に使える低遅延音声音響統合符号化になりました．

EVSの紹介

■EVSの経緯

現在世界で使われている大多数の携帯電話の音声符号化方式は1995年ごろの技術による標準化方式がそのまま使われており，音声帯域は3.5 kHzまで，音楽に対して不愉快な出力となるといった問題がありました．音声符号化方式を更改することはシステム全体の大幅な変更や円滑な移行手順が伴うことになり，容易ではありません．携帯電話が3Gの回線交換からLTEになる機会に音声符号化も新たな統一規格の制定が望まれてきました．3GPPでは2010年より，EVSを目標に活発な活動が開始され2014年に完成しました[5], [6]．

■EVSの特徴

まず，世界の有力機関12社による統一規格ができたことが大きな特徴です．さらに代表的な国際標準の仕様比較（**表1**）のように，EVSはこれまでの標準規格で達成されなかった，4つの符号化性能のすべてを満たす最初の標準となりました．すなわち，ビットレートは低いものまで対応する“高圧縮”，双方向通信を実現する“低遅延”，高いサンプリング周波数まで対応できる“広帯域”，音楽も含めた再生音声の品質が高い“高音質”のすべてが満たされていることになります．なお表1のオレンジ色の枠は好ましい仕様を示しています．

表1　EVSの達成

	電話用 符号化	放送用 符号化	音声音響 符号化	低遅延音響 符号化	VoLTE用 符号化	
レート（kbit/s）	10	48	16	24	5.9-128	高圧縮
遅延（ms）	30	80	80	30	32	低遅延
帯域（kHz）	4	24	24	24	24	広帯域
音楽品質	劣化	良好	良好	良好	良好	高音質
標準例	ITU-T G.729 3GPP AMR	MPEG MP3 MPEG AAC	AMR-WB+ MPEG USAC	AAC- ELD	3GPP EVS	

■国際協調

結果的にEVSは有力12社による共同案が標準化仕様となりました．この12社の中には移動通信サービスを行っているOrange，NTT DOCOMO，通信機器を扱うEricsson，Huawei，Nokia，ZTE，電話機やチップを製造するQualcomm，Samsung，Panasonic，研究開発機関であるFraunhofer，VoiceAge，NTTが含まれています．通常はこれらの組織のエンジニアはライバルであり，各社の利害は激しく対立し，符号化技術の策定は難航しましたが，世界最高の品質を世界の人たちにタイミング良く提供する熱意という点では結束することができました．

■技術の統合（広帯域）

入出力の信号帯域が広がり，サンプリング周波数が高くなると，単純に符号化情報量が増えてしまいます．情報量をあまり増やさずに広帯域化するには，低域の波形符号化と高域のスペクトル符号化の統合が有効な手段で，2000年以降，盛んに研究が進みました．別の言葉では高域の信号の位相成分を無視することに相当し，聴覚心理の知見を活用していることになります．

■技術の統合（高圧縮化）

高圧縮符号化のさらなる効率化のために，歪のない符号化で盛んに使われている可変長符号化が組み込まれています．可変長符号は別名エントロピー符号とも呼ばれ，パラメータなどの情報の冗長性を排除して本来の情報量に近い情報量までの圧縮を実現します．一方，符号の単位の長さ（ビット数）が固定でなくなり，符号誤りが混入すると，符号の単位の境界も誤り，ビット列全体の復号結果は無意味になってしまいます．従来の音声符号化は回線交換を前提として，デジタル圧縮ビット列に符号誤りが生じることは避けられなかったため，可変長符号は利用されることはありませんでした．一方LTEではパケットベースでの伝送であるため，上位のレイヤで誤りが制御されます．したがって，パケットの中の符号列に誤りが含まれることはなくなり，可変長符号を使って圧縮率を高めることが可能となりました．

■技術の統合（低遅延化）

音声入力に対しては時間領域のCELP（Code Excited Linear Prediction）方式をベースとするほうが品質が高く，音響信号が入力の場合はMDCT（Modified Discrete Cosine Transform）を使った周波数領域の符号化の品質が高くなります．これらの符号化を統合するには，適応的分類，モード切替，なめらかな遷移などの工夫が必要になります．さらにこれまでの周波数領域の音響符号化では品質を維持するため，例えば80 msのフレーム長が使われていましたが，EVSではフレーム長20 ms，サンプルの先読み，オーバラップも含めて符号化のアルゴリズム遅延が32 ms以内であることが必須の要求条件と決められています．このような低遅延化によって特に調波成分がはっきりする音楽での劣化が著しくなり，これを解決するために，周波数領域での量子化歪削減，調波成分の効率的表現などの技術が考案されました．

■EVSのその他の機能

EVS規格には，将来のVoLTEを想定した実用的要求条件が設定されており，すべてが機能しています（**表2**）．

EVSの評価

帯域の異なる3つの標準化符号化の主観評価結果を**図2**に示します．横軸はビットレート，縦軸は一般の評価者16名の平均主観評点です．雑音下の音声（女声3，男声3）を入力としました．結果の1番目は3G（第3世代移動通信システム）で広く使われている狭帯域（8 kHzサンプル）AMR[7]，2番目は現在VoLTEで広く使われている広帯域（16 kHzサンプル）AMR-WB[8]，3番目が超広帯域（32 kHzサンプル）のEVSです．12～13 kbit/s

表2　EVSの実用的特徴

- 広範囲のビットレートに対応
  −5.9 kbit/s から128 kbit/s の12種（帯域による）
- AMR-WB互換モード（品質は一部改善）
- DTX（無音圧縮）符号化
- 瞬時のレート切替，帯域切替に対応
- パケット消失時の品質劣化隠ぺい機能
- パケットジッタ用バッファ制御機能
- 実用的演算量範囲内

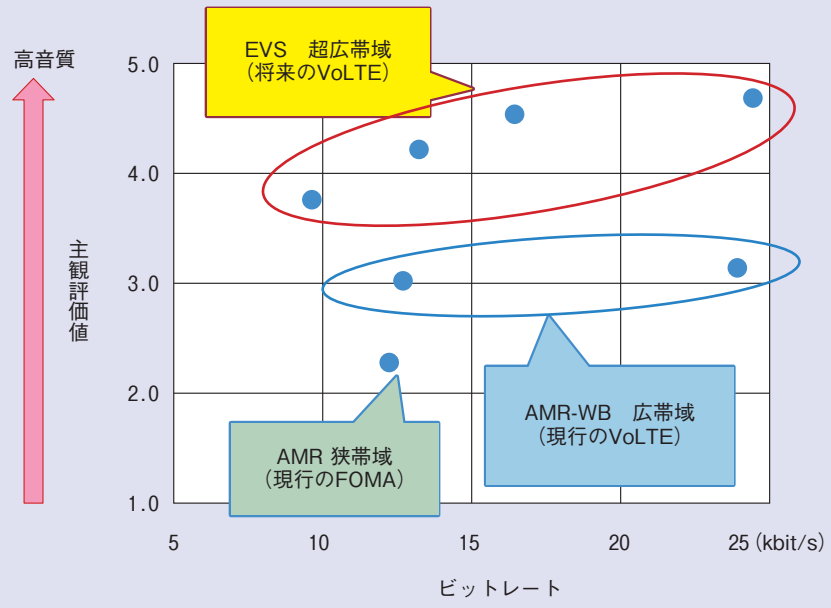

図2　EVSの主観品質評価結果例

付近の類似のビットレートで比較すると，EVSの品質の高さが明らかです．この傾向は音楽や音声音楽の混合入力などでも確認されています．

応用用途

電話の長い歴史で音声の周波数帯域を振り返ってみたいと思います．1876年の電話の発明時からおよそ100年はアナログ伝送で，音声帯域はほぼ3.5 kHzでした．その後伝送や交換がデジタル化されるなど，さまざまな技術の進歩がありましたが，音声帯域は変わりませんでした．技術的には1980年代には広帯域符号化が標準化されましたが，対応する電話器が一部のTV電話だけにとまっていました．自分の電話機が広帯域になっても相手の電話機が対応していないと有効に使えないことが普及の障害になったと考えられます．EVS規格という国際統一規格で，情報量をほとんど増やさないで超広帯域化ができ，比較的短時間で買い替えられるスマートフォンに搭載されることで初めて，超広帯域の電話をたくさんの人に使ってもらえる可能性が出てきました．電話の発明以来，100数十年で大きく変わることになります．

EVSはLTEを目標に設計，最適化された符号化ですが，音声や音楽の圧縮伝送蓄積の要素技術として幅広く有効利用できます．これらの用途の中には移動通信に限らず，一般の通信，TV電話，Web会議，ゲーム，音楽関係のアプリなどが考えられます．

今後の展開

本稿では，これまでの音声音響符号化の集大成的なEVS規格を紹介しました．将来のVoLTEはもちろんさまざまな音声サービスやアプリのかたちで世界中で有効利用されることが期待されます．2020年ごろには世界の移動通信で，多くの人たちが高品質広帯域の通話を使っていただけることを願っています．

■参考文献

(1) 守谷：“高圧縮音声符号化の必須技術：線スペクトル対（LSP），” NTT技術ジャーナル，Vol.26，No.9，pp.58-60，2014.
(2) ISO/IEC 14496-3：“Information technology - Coding of Audiovisual Objects - Part 3: Audio，” Subpart 11, Audio Lossless Coding，2005.
(3) 3GPP TS 26.290：“Audio codec processing functions - Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions，” 2005.
(4) ISO/IEC 23003-3：“Information technology - MPEG audio technologies - Part 3: Unified speech and audio coding，” 2011.
(5) 3GPP TS 26.445：“Codec for Enhanced Voice Services (EVS) - Detailed Algorithm Description (Release12)，” 2014.
(6) 守谷・鎌本・原田・菊入・仲・堤・大崎・江原・三田・河嶋・中尾：“3GPP標準EVSコーデックの概要～VoLTE用高性能音声音響符号化～，” 信学技報，Vol.114，No.475，pp.25-30，2015.
(7) 3GPP TS 26.071：“Mandatory speech CODEC speech processing functions; AMR speech CODEC，” 1999.
(8) 3GPP TS 26.190：“Speech codec speech processing functions - Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions，” 2002.

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
守谷特別研究室
TEL　046-240-3141
FAX　046-240-3145
E-mail　moriya.takehiro@lab.ntt.co.jp