

携帯電話や放送をきれいな音で伝えるための音声音響符号化技術

現在のスマートフォンどうしの通話では多くの場合VoLTE方式が使われており、これまでの固定電話や3G携帯電話（フィーチャーフォン）よりも、より自然な音声で会話を楽しむことができるようになってきました。また、IPTVをはじめとする放送においても音声の高品質化が計画されており、より臨場感の高いコンテンツを楽しむことができるようになります。本稿ではこのような高品質な音源を伝送するための基礎技術である音声音響符号化技術について紹介します。

かまもと ゆたか もりや たけひろ
鎌本 優 / 守谷 健弘
はらだ のぼる
原田 登

NTTコミュニケーション科学基礎研究所

近年の音声音響符号化

音声符号化や音響符号化は、低いビットレートでそれなりの音質で伝送するものから、高いビットレートで原音と全く同じ音質で伝送するものまで、用途に応じて選択できるようにさまざまな標準方式が策定されています。ネットワークに関する技術の改善により伝送帯域は増加してきており、4K放送やハイレゾ音源などの高いビットレートのデータを伝送できるようになってきています。一方、電話のように最優先で送らなければならない情報は、多くの方が利用しても輻輳しないようにしなければならないため、できるだけ低いビットレートで伝送する必要があります。また、通話のような双方向通信では遅延が長くなってしまうと会話が成り立たなくなりますが、放送のような一方通行の通信では遅延が長くてもそれほど問題が生じないため、遅延を活かした信号処理により効率良く圧縮することができます。さらに、音声に特化したモデルを用いた方式の場合は音楽の表現が苦手だったり、逆に、音楽の圧縮を目的とした方式で音声を圧縮すると機械がしゃべっているような声に聞こえたりする

ことがあります。このようなトレードオフの関係を克服するために世界中の研究者が集まり、短い遅延という厳しい条件の下でも音声と音楽の両方を満足できるような音質を達成する符号化方式の検討をしてきました。また、音質を重視し全く劣化させない符号化方式も標準化されました。

本稿では、新たに標準化された携帯電話向け音声音響符号化3GPP Enhanced Voice Services (EVS) コーデックと、今後数年での普及が見込まれるハイレゾ放送向け音響ロスレス符号化MPEG-4 Audio Lossless Coding (ALS) について紹介します。

携帯電話向け音声音響符号化 EVSコーデック

移動通信システムの国際標準化団体である3GPP (3rd Generation Partnership Project) では将来の携帯電話やVoLTE (Voice over Long Term Evolution) 向け音声音響統合符号化方式として、EVSコーデックを新たに策定しました^{(1),(2)}。これまでの携帯電話向け音声符号化は人間の発声メカニズムにならったCELP (Code Excited Linear Prediction) という方式が用いられ、人間の声を低いピッ

トレートにもかかわらず高品質で伝送してきました。EVSコーデックではCELPに加えて新たに開発された低遅延の音楽向け符号化を組み合わせることで、これまで実現できなかった背景雑音や背景音楽を含む音声信号や音楽信号そのものを高音質のまま低遅延で伝送することが可能となっています^{(3)~(5)}。

EVSコーデックは帯域拡張技術によって効率良く符号化を行うことができるため、これまで固定電話や3G携帯電話で使われていた8 kHzサンプリングレートの音だけではなく、VoLTEで使われている16 kHzサンプリングレートの音 (AMラジオ相当) や32 kHzサンプリングレートの音 (FMラジオ相当)、さらにTV放送で使われている48 kHzサンプリングレートの音というように幅広い周波数帯域をサポートしています (図1)。1フレーム20 msごとに伝送し、アルゴリズム遅延32 msという通話可能な遅延で処理を行います。また、VoLTE向けということもあり、これまでの3G携帯電話で必要だったビット列の誤り耐性 (0と1が反転してしまうような状況への対策) ではなく、パケット消失のあったフレームの音声

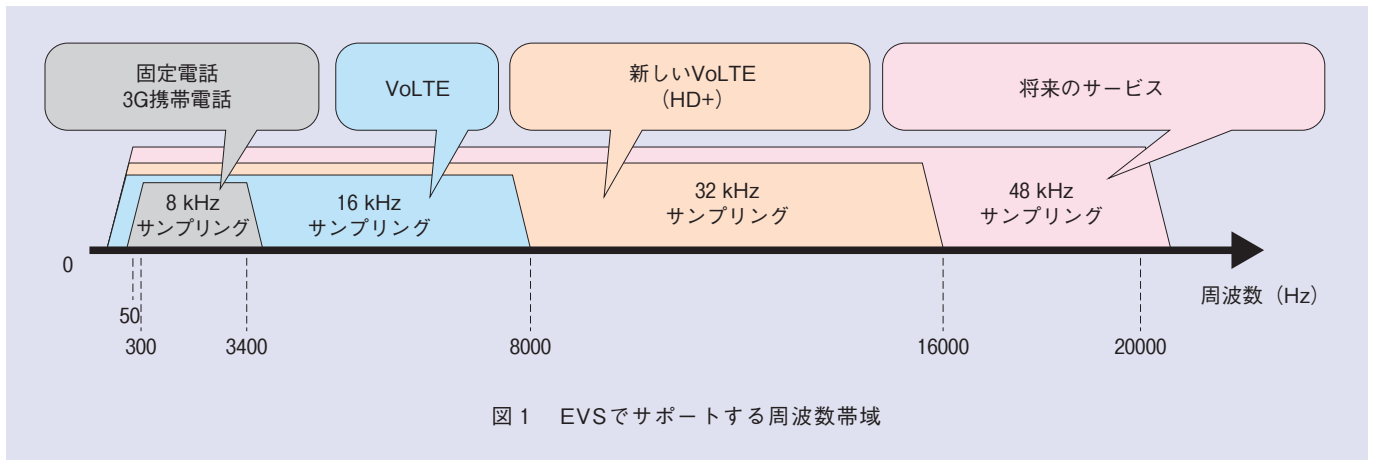


図1 EVSでサポートする周波数帯域

をそれよりも前に送られたフレームの情報から復元することにより違和感を小さくする方策がとられ、LTEによる伝送に適した方式となっています。さらに、対応するビットレートも5.9 kbit/sから128 kbit/sというように幅広く対応し、任意の周波数帯域・任意のビットレートにフレームごとに切り替えることができ、ネットワークの状況に柔軟に対応できます。そして、すでに普及しているVoLTEサービスで使われているAMR-WB (Adaptive Multi-Rate WideBand) との互換性も具備するため、コーデックの置き換えもスムーズに進めることが可能です。

標準化の過程において、さまざまな条件・音源・言語での大規模な主観品質評価実験が第三者機関によって行われ、実験結果が評価レポートに示されており、従来の音声符号化方式・音響符号化方式よりも高音質であることが確認されています⁽⁶⁾。また、同様の手

順で日本語に特化した主観品質評価実験も行われ、ほかの言語と同様にEVSコーデックが高音質であるという結果を得ることができています⁽⁷⁾。通信会社やメーカなど12の有力な企業・研究機関が協力して技術を提案し(日本からはNTT, NTTドコモ, パナソニックが参画)、従来の符号化方式よりも高性能であるEVSコーデックを開発・策定することに貢献しました。

実際にNTTドコモでは2016年の夏モデルからVoLTE (HD+) としてEVSコーデックを用いたサービスが始められています⁽⁸⁾。また、海外でも使われ始めており^{(9),(10)}、EVSコーデックは世界中で使われることが期待できます。今後、皆様が携帯電話・スマートフォンを新しいものに変えると、より自然な通話ができるようになり、雰囲気も伝わりやすくなることでしょう。

ハイレゾ放送向け音響ロスレス符号化ALS

動画や音声などのデータの符号化方式を検討する委員会であるMPEG (Moving Picture Experts Group) では、これまでに実用的な符号化方式を策定してきており、私たちの生活の中で使われています。MPEGでは音を対象に、圧縮しても元に戻る音響ロスレス符号化方式としてMPEG-4 ALSを標準化し、NTTもそれに貢献してきました^{(11),(12)}。圧縮率は入力信号(音源)に依存しますが、おおよそ30%から70%の大きさのビットレートへ圧縮することができます^{(13),(14)}。近年、ハイレゾオーディオ*¹が普及しつつあり、高精細にデジタル化された

*1 ハイレゾオーディオ：JEITAはCDよりもサンプリング周波数や量子化精度が高ければハイレゾリューションオーディオと呼ぶことを周知しました⁽¹⁵⁾。

音を楽しめるようになってきています。せっかくのハイレゾ音源も歪みのある圧縮をしてしまっただけでは性能が活かしきれずもったいないため、ALSのような音響ロスレス符号化が注目されています。

例えば、TV放送の音声は、放送局内では48 kHz、24 bitのいわゆるハイレゾ品質で制作されています。ところが、放送コンテンツを家庭まで届けるために使われる電波は有限な資源であり無駄遣いはできないため、人間の耳にはほとんど気にならない程度の劣化に抑えつつ、大幅な情報圧縮して、皆様の元へ届けられていました。これか

ら始まる超高精細度テレビジョン放送（4K/8K放送）では、映像で使われる情報量が大きくなるため、それに伴い、音声に割り当てられる情報量も大きくなるのが期待されます。通信・放送分野における電波利用システムの実用化およびその普及を促進する団体であるARIB（Association of Radio Industries and Businesses）では放送分野の標準規格ARIB STD-B32において、4K/8K放送で使われる音声符号化方式の1つとしてALSを利用できるように規定しています⁽¹⁶⁾。ALSは圧縮してもデジタル信号として必ず元に戻る方式なので放送局のスタジオ

で制作された音声をそのまま家庭まで届けることができるようになります。音声に十分な伝送帯域を割り当てることができれば、ハイレゾ放送を家庭で楽しむことができるようになります（図2）。電波による放送に先駆けて、通信回線を使うIPTV（Internet Protocol TeleVision）での放送コンテンツ伝送においても、ALSが使われる可能性が高まっています。

ALSの普及が見込まれることから、NTTでは周辺分野の標準化にも取り組んできました。演算量が小さく、使いやすいツールに限定したALS Simple Profileを策定し、これを使う

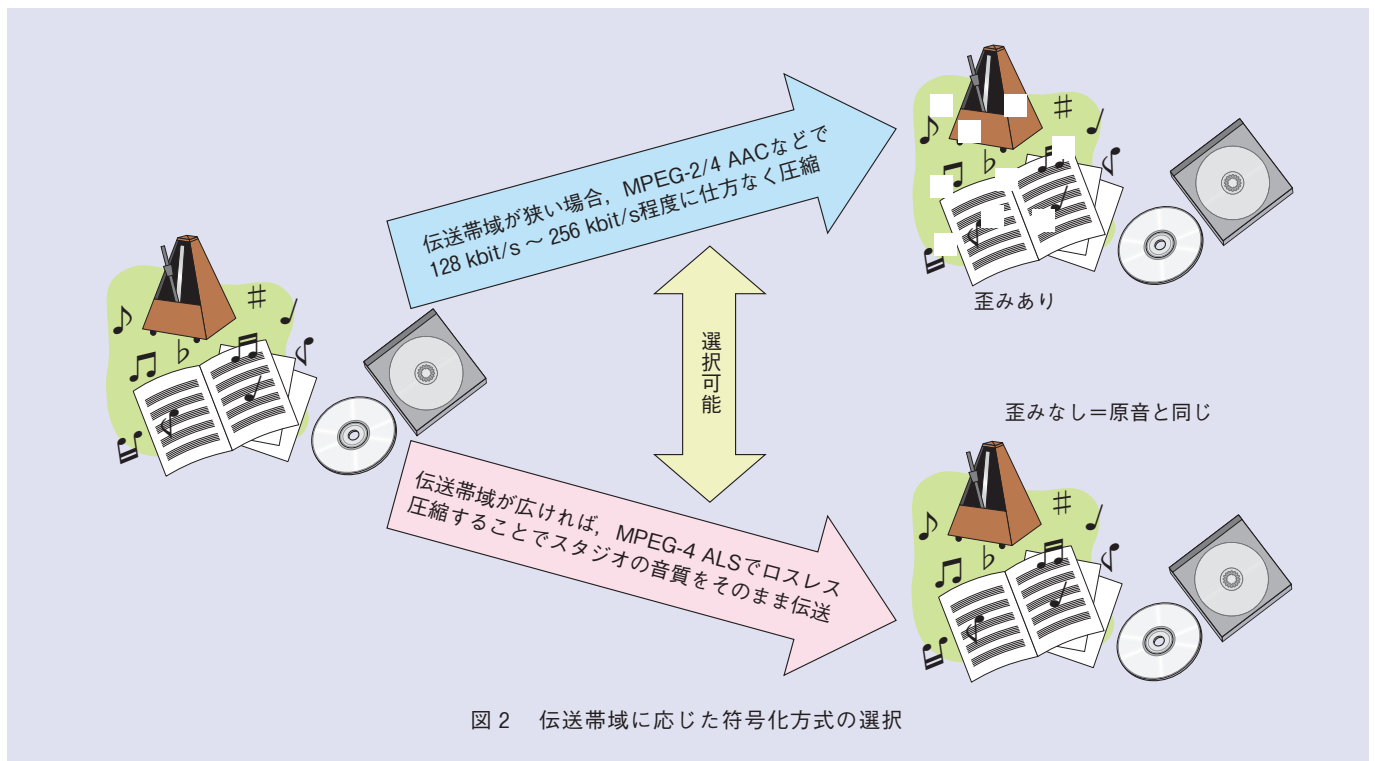


図2 伝送帯域に応じた符号化方式の選択

ことがARIB STD-B32にも記載されています。また、デジタルオーディオインタフェースの国際標準規格であるIEC 61937シリーズでALSを伝送することができるIEC 61937-10の標準化や⁽¹⁷⁾、ARIB STD-B32の規定に合わせてLATM/LOAS (Low-overhead Audio Transport Multiplex/Low Overhead Audio Stream) 形式^{*2}のALSストリームをそのまま外部アンプなどへ出力するためにIEC 61937-10のEdition 2への改訂作業も行っています。今後、ALSが普及することにより、皆様を意識せずとも放送の音質が向上することが期待できます。

今後の展開

EVSコーデックにより低いビットレートかつ低遅延にもかかわらず、音声と音楽の両方に対してより良い通話品質を得られるようになりました。また、ALSによりある程度の大きさのビットレートを確保できれば原音のままの音質を伝送できるようになりました。良い音を送ることができるのは当たり前になりつつあるので、今後は臨場感を向上させるためには良い音以外に何が 필요한のかを検討していきたいと思っています。例えば、ネットワークを利用したリアルタイムコンテンツ配信

サービスにおいて、配信元と配信先での相乗効果による臨場感の向上が期待できます。そのためにネットワークを利用するメリットである双方向性を活かした付加価値の高い通信システムの実現に向けて理論基盤の構築を進めたと思います。今後も、通信手段の要求条件に合わせた音声音響符号化技術をタイミングよく提供できるように、基礎研究を継続していきたいと思っています。

参考文献

- (1) 3GPP TS26.441 : “Codec for Enhanced Voice Services (EVS); General overview,” 2015.
- (2) 3GPP TS26.445 : “Codec for Enhanced Voice Services (EVS); Detailed algorithmic description,” 2015.
- (3) M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, H. Poblath, L. Miao, Z. Wang, L. Laaksonen, A. Vasilache, Y. Kamamoto, K. Kikuri, S. Ragot, J. Faure, H. Ehara, V. Rajendran, V. Atti, H. Sung, E. Oh, H. Yuan, and C. Zhu : “Overview of the EVS codec architecture,” Proc. of ICASSP2015, pp.5698-5702, Brisbane, Australia, April 2015.
- (4) 守谷・鎌本・原田・菊入・堤・仲・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, SP2014-139, pp.25-30, 2015.
- (5) from NTTドコモ : “VoLTEのさらなる高音質化と音楽の活用を実現する3GPP標準EVSコーデック,” NTT技術ジャーナル, Vol.27, No.7, pp.32-35, 2015.
- (6) 3GPP TR26.952 : “Codec for Enhanced Voice Services (EVS) ; Performance characterization,” 2015.
- (7) 鎌本・守谷・原田 : “3GPP標準EVSコーデックの日本語音声品質評価,” 2016年春季日本音響学会研究発表会講演論文集, 3-2-9, 2016.
- (8) <https://www.nttdocomo.co.jp/support/area/volte/>
- (9) https://www.ericsson.com/news/160614-evolved-hd0-voice-for-volte_244039855_c
- (10) <https://mitarbeiterapp.vodafone.de/article/31310460>
- (11) ISO/IEC 14496-3 : “Information technology — Coding of audio-visual objects — Part3:

Audio,” fourth edition, 2009.

- (12) T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, Y.A. Reznik : “The MPEG-4 Audio Lossless Coding (ALS) Standard-Technology and Applications,” Proc. of AES119th Convention, New York, U.S.A., Oct. 2005.
- (13) 鎌本・守谷・原田・Kos : “ロスレス・オーディオ符号化MPEG-4 ALSの高性能化,” NTT技術ジャーナル, Vol.20, No.2, pp.11-18, 2008.
- (14) 原田・守谷・鎌本 : “MPEG-4 ALSの性能・応用と関連する標準化活動,” NTT技術ジャーナル, Vol.20, No.2, pp.19-25, 2008.
- (15) http://home.jeita.or.jp/page_file/20140328095728_rhsiN0Pz8x.pdf
- (16) ARIB STD-B32 : “デジタル放送における映像符号化, 音声符号化及び多重化方式,” 2011.
- (17) IEC 61937-10 : “Digital audio- Interface for non-linear PCM encoded audio bitstreams applying IEC60958- Part10: Non-linear PCM bitstreams according to the MPEG-4 audio lossless coding (ALS) format,” 2011.



(左から) 守谷 健弘/ 鎌本 優/
原田 登

皆様を意識せずに、より自然な音や制作者の意図した音を聞くことができるように、圧縮符号化の研究開発を進めていきます。IoTやM2Mなどの信号にも対応していきたいと思っています。

◆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 046-240-3593
FAX 046-240-3145
E-mail MPEG4ALS@lab.ntt.co.jp

*2 LATM/LOAS形式 : MPEG-4 Audioで規定されている、MPEG-4 Audioのヘッダやストリームを伝送するときに使われる方式の1つ。