音声音響符号化技術の進展 —線形予測分析の貢献

Progress in Speech and Audio Coding
—Role of Linear Prediction

守谷健弘 Takehiro MORIYA
原田 登 Noboru HARADA

鎌本 優 Yutaka KAMAMOTO
杉浦亮介 Ryosuke SUGIURA

アブストラクト 多様な信号処理の全般で重要な役割を果たしている線形予測分析技術の利用の観点で、音声音響符号化技術の進展を紹介する。開発当初から線形予測分析による合成フィルタは音声生成の声道モデルと親和性が高く、音声合成や電話音声に特化した音声音響符号化で広く使われてきた。一方、現在普及している典型的な音声音響符号化には線形予測符号化技術が使われていないが、低ビットレート化、音声音響統合符号化の要請に合わせて、スペクトル包絡を効率的に表現する手法として広く使われるようになった。これらの経緯を説明し、あわせて最近実用段階にあるロスレス音声音響符号化 MPEG-4 ALS、携帯電話用の符号化 3GPP EVS について紹介する。

キーワード 音声音響符号化、線形予測、情報圧縮、携帯電話、デジタル放送

Abstract Progress in speech and audio coding is presented, focusing on the technology of linear predictive coding (LPC), which has played important roles in various processing schemes for speech and audio signals in general. From the first, LPC has been used for speech synthesis and telephone bandwidth speech coding, since it was found to be well suited to represent the characteristics of the vocal tract. In contrast, typical audio coding schemes have not used LPC. However, because of the needs of a low bitrate and the unification of speech and audio coding, LPC has come to be used for representing spectral information in audio coding as well. In line with this trend, two new standard coding schemes are introduced: audio lossless coding, MPEG-4 ALS, for digital broadcasting, and 3GPP EVS for recent cellular telephones.

Key words speech and audio coding, linear prediction, compression of information, cellular phone, digital broadcasting

1. ま え が き

放送や携帯電話は貴重な電波帯域を多くの利用者で共有するため、品質を維持しつつ信号を情報圧縮して伝送する符号化技術は極めて重要である。音声音響信号の圧縮技術の進展は、1990年代以降世界中で幅広く使われている経済的なデジタル通信サービスの実現に多大な貢献をしてきている。

符号化技術には情報圧縮を目的とする情報源符号化と伝送路誤りを防ぐことを目的とする伝送路符号化があるが、本稿では、50年にわたる情報源符号化としての音声音響符号化の研究開発と実用化の経緯を振り返る。特にこれらを支える多くの技術の中で、最も重要で共通に使われていると考えられる線形予測符

号化技術を周波数領域の符号化への展開も含めて紹介する。なお本稿では線形予測の分析、合成、量子化などの符号化で用いる関連技術を含める意味で線形予測符号化技術、または LPC (Linear Predictive Coding) という言葉を使わせて頂く。

これらの技術の実用例として、最近話題になっている二つの国際標準化技術を紹介する。その一つは衛星放送、IPTV、音楽配信などの用途に期待されているひずみを許さないロスレス符号化^(注1)の MPEG (Moving Picture Experts Group) -4 ALS (Audio Lossless Coding) 方式である。もう一つは携帯電話の高品質通話サービスに利用され始めている低ビット低遅延音声音響符号化の 3GPP (3rd Generation Partnership Project) EVS (Enhanced Voice Services) 方式である。

2. 技術とサービスの進展

2.1 歴 史

図1は音声音響符号化に関する技術と応用の分類を試みたものである。この図では主にフレーム単位の処理で高能率の圧縮を目指すものを採り上げ、双方向通信用の狭帯域低遅延音声音

守谷健弘 正員：フェロー NTT コミュニケーション科学基礎研究所
E-mail t.moriya@m.ieice.org
鎌本 優 正員 NTT コミュニケーション科学基礎研究所
E-mail kamamoto.yutaka@lab.ntt.co.jp
原田 登 正員 NTT コミュニケーション科学基礎研究所
E-mail harada.noboru@lab.ntt.co.jp
杉浦亮介 NTT コミュニケーション科学基礎研究所
E-mail sugiura.ryosuke@lab.ntt.co.jp
Takehiro MORIYA, Fellow, Yutaka KAMAMOTO, Noboru HARADA, Members, and Ryosuke SUGIURA, Nonmember (NTT Communication Science Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Atsugi-shi, 243-0198 Japan).

電子情報通信学会 基礎・境界ソサイエティ
Fundamentals Review Vol.10 No.4 pp.246-256 2017 年 4 月
©電子情報通信学会 2017

(注1)：本誌が準拠している本会編の改訂電子情報通信用語辞典では「無損失符号化」となるが、本編では、著者の希望により一般的に用いられている「ロスレス符号化」を用いている。

図1 音声音響符号化のカテゴリ⁽¹⁾

図2 音声音響符号化の進展の歴史 (横軸は年代、縦軸はおおまかなビットレートで、矢印は基本的な技術の影響を示している。) ⁽¹⁾

号符号化、片方向通信用の高品質音響信号符号化 (オーディオ信号と同義で以下音響信号とさせていただきます) ひずみを許さないロスレス符号化の3種に分類している。ロスレス符号化は、無損失符号化、無ひずみ符号化あるいは可逆符号化とも呼ばれ、ひずみがない条件で、できるだけ情報量を節約することが目標である。ロスレス符号化以外は与えられた情報量 (ビットレート) の範囲内で、できるだけひずみを少なくして品質を維持することが目標である。

これらの符号化の要求条件や利用形態は多様であるが、それらの発展や実用化には、デジタル信号処理技術や情報理論の進展、聴覚や音声生成などの人間科学の知見の利用、プロセッサやメモリなどのハードウェア技術の進展が不可欠であった。

図1を大まかな開発の年代に展開したものが図2である。これらの図を使って技術やサービスの進展を大雑把に振り返る。

2.2 双方向通信：音声符号化

双方向通信の主な用途は電話、特に携帯電話で、低遅延かつ低ビットレートが要求される。入力 は 8 kHz サンプリングのモノラル信号である。1980 年代初頭までの音声符号化は分析合成方式 (ボコーダ) が主流であった。分析合成は音声を生成モデルを使って音源と声道の特性に分離してそれらのパラメータをそれぞれ低ビットレートで伝送して、デコーダで再構成する。この方式では再生される音声波形のスペクトル形状とピッチ周期をできるだけ原音と近づけることを目的としている。

分析合成の構成はアナログフィルタの時代から試みられてい

たが、デジタル処理として線形予測符号化技術⁽²⁾⁻⁽⁵⁾が試みられて大きな発展を遂げている。声道の特性の表現や人間の聴覚が対数軸でスペクトルの谷の部分より山の部分に敏感であることが極めて有利であった。また線形予測係数の等価パラメータである PARCOR (PARTIAL autoCORrelation)⁽⁷⁾や LSP (Line Spectrum Pair)⁽⁸⁾⁽⁹⁾ など実用に適した効率的表現が考案され、更にその量子化についても熱心に研究された。ボコーダによって圧縮率が高く、肉声に近い合成音の実現できたが、電話のような有料の公衆通信用途までの品質を実現することは困難であった。

1985 年頃からボコーダのスペクトルモデルとサンプルごとの波形符号化を組み合わせた 8 kbit/s 付近でのハイブリッド符号化の開発が進んだ。線形予測に基づく合成フィルタを用い、そのフィルタの入力となる励振波形を多数のベクトルとして用意し、合成波形と入力との誤差が最小となるベクトルを閉ループで決定する CELP (Code Excited Linear Prediction)⁽¹¹⁾で自然性の高い低ビットレート符号化が実現できた。

CELP を基に励振波形の数多くの簡素化、パラメータの符号化、高品質化の要素技術の研究が進み、線形予測係数の表現としてはボコーダの研究成果が生かされ、ほぼ全てで LSP のベクトル量子化が利用されるに至った⁽¹⁰⁾。1990 年代には携帯電話用符号化方式が標準化され、励振波形を構造化した VSELP (Vector Sum Excitation LP) 符号化⁽¹²⁾や PSI-CELP (Pitch Synchronous Innovation CELP) 符号化⁽¹³⁾が商用化された。その後、励振波形は更に簡素化され、単一パルス群を励振波形とする ACELP (Algebraic CELP)⁽¹⁴⁾が世界中のほぼ全部の 3G 携帯電話の規格である 3GPP AMR (Adaptive Multi-Rate)⁽¹⁵⁾や 3GPP2 EVRC (Enhanced Variable Rate Coding)⁽¹⁶⁾更に IP 電話の規格となった ITU-T G.729⁽¹⁷⁾⁽¹⁸⁾に採用され、現在に至っている⁽¹⁹⁾。

2.3 片方向通信：音響符号化

片方向通信はほぼ音響符号化を指し、入力 は 48 kHz サンプリングのステレオが主体である。主要な用途は放送や記録媒体への録音である。多くの符号化はサブバンド領域または MDCT (Modified Discrete Cosine Transform)⁽²⁰⁾による周波数領域の係数の適応的量子化とパワースペクトルを示す補助情報で符号化されている。符号化の遅延の要求条件は緩く、ビットレートもステレオで 100 kbit/s 以上でよいが、ほとんど劣化が感じられない程度の高い品質が要求される。このため、情報理論的圧縮に加えて、周波数領域の聴覚特性を丁寧にモデル化して聴覚量子化雑音を最小化する技術が進展した。1990 年代初頭からの MPEG の標準規格が大きな役割を果たしている。MP3 (MPEG-1/2 Layer III)⁽²¹⁾や AAC (Advanced Audio Coding)⁽²²⁾、更に帯域拡張で低ビット化した HE-AAC (High-Efficiency AAC)⁽²³⁾⁽²⁴⁾など世界中で幅広く使われている音響符号化は MPEG 標準である。ディスクへの録音やインターネット配信などではデファクトとなった独自方式も幾つか存在するが、技術的には MPEG 標準方式と共通するものが多い⁽²⁵⁾。

更に MPEG ではマルチチャネル音源やオブジェクト単位で

の音源の符号化 MPEG SAOC (Spatial Audio Object Coding)⁽²⁶⁾, 立体音響処理 MPEG-H 3D audio⁽²⁷⁾へと発展している。またこれに先駆け、音声と音響符号化の統合を目指した MPEG-D USAC (Unified Speech and Audio Coding)⁽²⁸⁾⁻⁽³⁰⁾や 3GPP AMR-WB+ (extended Adaptive Multi-Rate Wide-Band)⁽³¹⁾⁻⁽³²⁾といった国際標準が制定されている。本稿の最後で改めて紹介する 3GPP EVS は更に遅延の少ない条件下での双方向通話用の統合符号化と位置付けられる。

2.4 ひずみを許さない符号化

図 2 の最上段に分類されているのはひずみを許さないという拘束の下にできるだけ情報量を削減することを目指すロスレス符号化である。この点で情報量が与えられてできるだけひずみを抑制することを目指す前節までで紹介した二つの符号化の分類とは異なる。

フレームごとに可変ビットレートとなり、原音の情報量からの圧縮効率は相対的に小さいという制約があるが音源の長期保存、高品質素材伝送、高品質放送などで有効である。独自方式やフリーソフトも各種使われているが⁽³³⁾⁻⁽³⁵⁾、国際標準としては、後で紹介する音響用の MPEG-4 ALS⁽³⁶⁾⁻⁽⁴⁰⁾や対数圧伸 PCM 用の ITU-T G.711.0⁽⁴¹⁾⁻⁽⁴³⁾が制定されている。

なおテキストなどを圧縮する可逆圧縮技術として文字列や単語の辞書を自動生成して圧縮効率を高める ZIP など多くの符号化法が知られている。ただし、それらの方式は音響信号の圧縮効率が低く、また長期的なデータの依存性があるために指定した時間の途中から再生できる構造を持っていないので、音響信号への適用には適さない。

3. 情報圧縮

3.1 圧縮符号化の基本技術

音声音響信号の圧縮には、①サンプル値の分布の偏り、②サンプル値間の相関、③人間の聴覚特性、を使うことができる。サンプル値の分布が一様でないことを利用することは情報圧縮の基本で、具体的には可変長符号、最適量子化、更にそれぞれを多次元化するなどの手段が使われる。これらの効果はサンプル値間の相関とは独立の圧縮効果を持つが、多くの信号の場合、効果は限定的である。また符号誤りがある場合のような通信には使いにくいことが多い。

2 番目のサンプル値間の相関については音声音響信号の圧縮では有効に利用できる。通常有意義な情報はサンプル値間の相関が大きく、すなわち周波数軸上でのパワースペクトルの値の強弱の偏りが大きく、しかも時間ごとに変化する。これを利用する有力な手段が線形予測符号化技術である。

図 2 での矢印はそれぞれのカテゴリーをまたいだ技術の影響を大雑把に示したものである。下からの矢印はほぼ線形予測符号化技術に関連するもので、上からの矢印は可変長符号化に関連するものを想定している。このように 1970 年代からの線形予測符号化技術の基盤はボコーダの開発から出発し、音声符号化、

図 3 ロスレス符号化の圧縮と高圧縮符号化の圧縮

音響符号化、ロスレス符号化にも応用されている。

本稿では線形予測符号化技術によるサンプル値間の相関を利用する技術を中心に紹介させて頂く。サンプル値の分布が一様でないことを利用する圧縮技術は、画像や映像及びテキストも含む多岐にわたる符号化の研究と関連する。また聴覚特性を利用した品質の改善技術については音声符号化でも⁽⁴⁴⁾、音響符号化でも⁽⁴⁵⁾古くから研究が進められてきている。いずれも整理して紹介するのは難しく、本稿では割愛させて頂く。

なお、音響符号化は聴覚特性を利用した圧縮効果をうたうので、別名聴覚符号化と呼ばれる場合が多い。実際の音響符号化は図 3 に表現するように時間領域の符号化やロスレス符号化と同様にサンプル値間の相関(パワースペクトルの偏り)やサンプルの分布の偏りによる圧縮効果を利用している。その上で音声符号化でも音響符号化でも聴覚特性を利用して更に情報削減による聴覚的劣化を防いでいる。音響符号化の圧縮効果は聴覚特性だけの効果ではないことに注意したい。

3.2 サンプル値間の相関の利用

図 4 は異なるカテゴリーの符号化でもサンプル値間の相関を利用して、線形予測分析と合成の枠組みで統一的に扱えることを示したものである。いずれも線形予測分析に基づく予測フィルタとその逆の合成フィルタが基本になり、振幅が小さくパワースペクトルがほぼ平坦になっている予測誤差信号の表現モデルが異なることを示している。整数で表現した予測残差をひずみなく符号化するものがロスレス符号化、予測残差を波形としてベクトルなどで低ビットに圧縮するのが高圧縮(ハイブリッド)符号化、予測残差をピッチ周期のパルスと雑音でモデル化するものが分析合成符号化(ボコーダ)である。予測残差の表現ビットレートの違いで、ロスレス符号化は元の PCM 信号の平均半分程度までのビットレートの圧縮が普通であるが、高圧縮符号化では 1/10 程度までに圧縮した形で実用的符号化に使われている。分析合成では予測波形そのものは符号化しないので、予測パラメータとピッチ周期だけの符号化として例えば 1/30 程度のビットレートまでに圧縮できる。

図 4 には周波数領域の符号化が出てこないが、後述するように線形予測の予測と合成による情報圧縮効果は周波数領域での

図4 3種のLPC合成の比較 ロスレス符号化、高圧縮波形符号化、分析合成（ボコーダ）⁴⁶⁾

スペクトル包絡の形を使った符号化で同様の圧縮効果がある。当初の音響符号化は線形予測符号化技術とは無関係であった。線形予測分析は人間の声道の伝達特性との対応が良いので、音声の合成や音声符号化にとって極めて親和性が高い。このため、一般の音響信号に対する符号化では線形予測符号化技術の利点はないとも誤解されてきた。1990年代後半からは音響符号化の高圧縮波形符号化にも線形予測符号化技術が使われ始め、最近ではそれが広く普及している。

3.3 線形予測の役割

線形予測分析技術は上述のように音声音響符号化に欠かせないものであることは明らかであるが、音声合成、音声認識、音声強調など広い範囲の音声音響処理技術に重要である。その発端は今からほぼ50年前の1966年の国際会議に始まる。AtalとSchroeder⁴⁾はフレーム単位の線形予測分析と時間領域の予測符号化の原点となるフィルタを提案した。ItakuraとSaito⁵⁾は線形予測分析と周波数領域のパワースペクトルの対応を明らかにした。これは後の音声認識や周波数領域での符号化の原点となった。

その後、符号化の観点でPARCOR係数による量子化、更に効率の良いLSPパラメータでの量子化が考案された。更にLSPパラメータはスペクトル包絡の形状との変換の対応が素直なために、LSPパラメータ領域での値をユークリッド距離を使った予測、補間、ベクトル化などで処理することで更に効率を高めることができた。

符号化で通常使われる手法は下記のように全極形または自己回帰形であり、自己相関法で予測誤差を最小化するための正規方程式をLevinson-Durbinアルゴリズムで解く。これにより安定な予測係数が高速に求められる⁶⁾。

予測係数 $\{a_n\}_{n=1}^p$ とする式(1)の

$$H(z) = \frac{\sigma}{1 + \sum_{n=1}^p a_n z^{-n}} \quad (1)$$

σ^2 は正規化予測誤差エネルギーであり、入力波形のエネルギーを1としたとき、予測誤差のエネルギーが σ^2 となり、下記の式でも表すことができる。

$$\sigma^2 = \prod_{n=1}^p (1 - k_n^2), \quad (2)$$

図5 線形予測によるスペクトル包絡の例

ここで k_n は n 次の PARCOR 係数である。

$$G = \frac{1}{\sigma^2} = \frac{\text{入力信号のエネルギー}}{\text{予測残差信号のエネルギー}} \quad (3)$$

G は予測利得で予測によって波形のエネルギーを削減できた効果であり、情報圧縮の源泉となる。

周波数領域で量子化ひずみを最小化する場合には図5の例のようなスペクトル包絡を $\log(|H(z)|^2)$ で求める。この包絡は低ビットで表現でき、この形状を利用して適応的ビット配分の量子化あるいは適応的重み付け距離探索のベクトル量子化が可能となる。周波数領域のサンプル値の偏りによるひずみ削減効果（変換利得） G_f は

$$G_f = \frac{\text{パワースペクトル包絡の相加平均}}{\text{パワースペクトル包絡の相乗平均}} \quad (4)$$

で表される。線形予測符号化によるパワースペクトル包絡を使えば、 G_f は理想的には時間領域の利得 G と同一と近似できる¹⁹⁾ ⁴⁷⁾。すなわち、周波数領域でも線形予測符号化技術で時間領域と同等のひずみ削減効果（ビットレート圧縮効果）を実現できる。

4. 周波数領域での線形予測

4.1 時間領域と周波数領域の対比

2.で紹介してきたとおり、電話用途の音声符号化と放送用途の音響符号化はいろいろの観点で異なっているが、これを改め

	音声符号化	音響符号化
サンプリング	8 kHz	48 kHz
処理	時間領域	周波数領域
圧縮	予測符号化	変換符号化
用途	双方向通信	片方向通信
チャネル	モノラル 単音	ステレオ 複合音
フレーム長	20 ms	80 ms
主な学会	Interspeech 情報・システム(本会) 音声B(日本音響学会)	Audio Engineering Society 基礎・境界(本会) 電気音響(日本音響学会)

図6 典型的な音声符号化と音響符号化の特徴

て対比したものが図 6 である。

主な用途が異なり、それによって帯域、遅延、ビットレートが大きく異なる。主流の符号化が周波数領域と時間領域になり、また標準化団体やメインの学会も異なっていた。この結果、研究者や技術の交流も限定的であった。

一方で、2000 年以降、音声と音響信号を共通に符号化する音声音響統合符号化の開発が要請された。音声符号化は CELP の枠組みの性能が高く、周波数領域ではまだ追いつかない。一方、一般の音響信号には CELP は性能が低いので統一の枠組みは難しい。ただし時系列の符号化の観点からは共通点も多く、音声符号化の道具立てが音響符号化にも影響を及ぼした例として線形予測符号化技術が挙げられる。

4.2 周波数領域での対比

線形予測符号化技術の周波数領域の音響符号化への展開を意識して符号化の基本的構造の分類と代表的方式を図 7 に示す。

(A) のカテゴリー図 7 は MP3 や AAC に代表される周波数領域の典型的な音響符号化である。周波数領域での符号化はパワースペクトルの振幅が周波数方向で平坦でないことで情報圧縮を実現するが、補助情報としてスペクトルの概形の情報が必要である。この目的で周波数帯域を不均一な帯域（高周波ほど幅が広いバーク尺度）ごとのパワースペクトル値を量子化して符号化する。このパワースペクトルの形状と聴覚特性に基づいてスケールファクタ帯域ごとに適応的に MDCT 係数の正規化利得と量子化ビット数が決定される。量子化値はサンプルごとの可変長符号として最適な複数のハフマン符号化で圧縮される。USAC や EVS でもビットレートが高い動作モードでこの

図 7 符号化の基本的構造の分類と代表的方式

(A) 音響符号化のエンコーダ [上段]/デコーダ [下段], (B) LPC 包絡に基づく周波数領域の音響符号化, (C) LPC 合成に基づく周波数領域の音響符号化 (狭い意味での TCX と呼ばれる), (D) 時間領域の LPC (音声) 符号化⁽⁴⁶⁾

カテゴリの符号化が使われているが、可変長符号としては適応的な算術符号化が使われている。

(B) のカテゴリは周波数領域の符号化に必要なパワースペクトル形状の推定に線形予測符号化に基づくパワースペクトル包絡、すなわち LPC 包絡を使うものである。古くは TwinVQ (Transform domain weighted interleave Vector Quantization)^{(48) (49)} MPEG-4 general audio TwinVQ⁽⁵⁰⁾, 低遅延の符号化⁽⁵¹⁾, MPEG-D USAC での低ビットレートモード、3GPP EVS の低ビットレート寄りの動作モードなど、比較的新しい符号化標準で使われている。量子化ひずみの制御原理は基本的に (A) の場合と同じである。

(C) のカテゴリは LPC 予測残差を周波数領域で符号化し、再生信号を LPC 合成フィルタで再生するものである。(A), (B) のカテゴリではフレーム境界雑音を避けつつ、周波数領域の符号化サンプル数が時間領域のサンプル数と同じという利点を有する MDCT が使われているが、(C) の構成では LPC 合成フィルタがフレーム境界雑音を抑えることができるので、オーバーラップのない DCT (Discrete Cosine Transform) が使われる。TCWVQ (Transform Coding with Weighted Vector Quantization)⁽⁴⁷⁾ や 3GPP AMR-WB+ は狭い意味での TCX (Transform Coded Excitation)^{(52) (53)} に基づいており、このカテゴリに属する。なお広い意味で“TCX”という用語はカテゴリ (B) と (C) とともに使われる場合が多いので注意する。このカテゴリの符号化でも周波数領域での量子化には (A) や (B) と同様のビット割当てやひずみ尺度の制御を行うことで、変換利得に対応する量子化ひずみを削減できる。

カテゴリ (D) は LPC 合成フィルタを用いる典型的な時間領域の CELP である。デコーダでの合成フィルタの入力となる励振波形は当初、乱数や学習されたベクトルであったが、その後ほとんど代数符号帳による ACELP^{(14) (18)} となった。ACELP の励振波形は位置の異なる等しい振幅の単一パルスから成る複数の系列の和で表される。励振波形は簡素で符号帳を必要としないが、符号器は聴覚補正付き LPC 合成フィルタ後の波形ひずみで時間領域の励振波形を探すため、演算量削減のための工夫がされており、演算量を削減する目的で探索時にのみ周波数領域で行うこともある^{(54) (55)}。

4.3 今後の展開

カテゴリ (A) と (B) はパワースペクトルの形状の表現方法が違うだけであるが、定性的にはビットレートが高い場合には (A)、低い場合には (B) が主流である。(B) の場合は LSP パラメータの効率良い符号化やパワースペクトルの周波数方向の変化が大きくても効率が落ちない点で、LPC 包絡を使う方が効率が良いと思われる。この点での定量的な考察や最適化は今後の課題である。

分析合成のための信号処理を周波数領域の符号化に応用する検討は既に 1970 年代に試みられている^{(56) (57)}。当時はフレーム境界のひずみを除く効果的方法がなかったために進展は途絶えていたようであるが、現在は MDCT の利用で解決している。

今後低遅延，低ビットの周波数領域の符号化では (B) のカテゴリーが主流となると思われ，多面的に発展してきた信号処理を使って更に発展する可能性がある．例えば，LPC スペクトル包絡の表現の改善として，最適化した正值の行列をパワースペクトルに適用した周波数分解能の伸縮^{(58) (59)}や周波数領域の係数の振幅分布推定に基づく最適包絡推定である PAPSE (Powered All-Pole Spectrum Estimation)^{(46) (60) (61)}はこの例である．このように (B) のカテゴリーの符号化の考察は進行中である．

5. 実用符号化

5.1 音響ロスレス符号化 MPEG-4 ALS

5.1.1 経緯

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 は ISO/IEC JTC1 SC29 の標準化グループである．これまでに実用的な符号化方式が制定され，放送や光ディスク媒体などで広く使われている．2005 年には NTT 主導で国際協力の下に音響ロスレス符号化方式 MPEG-4 ALS の標準規格を制定した⁽³⁶⁾．この規格は入力として線形 PCM の様々なフォーマットや条件に幅広く対応し，浮動小数点形式対応などの多くの機能や処理を選択できる柔軟性を持っているが，実用的な配信や放送での要求に合わせて経済的なデコードを定義する必要があった．サービスの要求を満たす範囲で演算量やメモリ消費量を削減するために機能や処理構成を限定した ALS のシンプルプロファイルを制定し，入力のサンプリングレートとチャンネル数に合わせた 4 種のレベルを制定した⁽⁶²⁾．

更に 2008 年以降，ALS は日本の衛星放送などに使われる可能性が出てきたので，日本の電波利用システムの実用化及びその普及を促進する団体である ARIB (Association of Radio Industries and Businesses) 標準規格 ARIB STD-B32 において，超高精細度テレビジョン放送 (4K/8K 放送) で使われる音声符号化方式の一つとして規定され，衛星放送の運用規定にも制定された．

更にデジタルオーディオインタフェースの国際標準規格の IEC 61937-10 が NTT と国内メーカーの協力で ALS の伝送を可能とするよう制定された．更に ARIB STD-B32 の規定に合わせ LATM/LOAS (Low-overhead Audio Transport Multiplex/Low Overhead Audio Stream) 形式の ALS ストリームを伝送する拡張は 2017 年中の完成が見込まれている．これにより ALS ストリームをテレビの受信機で復号せずに，外部の AV アンプなどにそのまま伝送し，AV アンプで復号するという構成も可能となる．

5.1.2 特徴と技術

2.4 でも紹介したように ALS はロスレス符号化で，原音と同一の波形の再生を保証する一方，入力の信号の音量や特徴に依存してビットレートが変動する．振幅が大きく，サンプル値間予測が難しい白色雑音では原音の線形 PCM から情報量をほとんど削減できない．一方，振幅が小さく，時間領域か周波数領

図 8 AAC と原音 (ALS) のデコーダでの再生音声のスペクトル比較 (CD 帯域音楽)⁽¹⁾

域でエネルギーが偏っている，言い換えればサンプル値間の予測ができれば半分以上の情報まで圧縮できる．

ALS は典型的な全極型の線形予測符号化に基づいており，過去のサンプルからの予測波形を整数化し，入力波形との差分の整数値をサンプルごとに可変長符号化する．予測は最大 1023 次までで，予測係数の符号化には PARCOR 係数の可変長符号化が使われる⁽⁶³⁾．フレーム単位で符号化が完結し，任意のフレームの先頭から波形を再現できるようにフレームの先頭から予測次数を漸増させる技術や，演算量の増加を抑えつつ圧縮率を高めるために，フレームごとの適応的長期予測，適応的チャンネル間予測，適応的サブフレーム分割，予測次数の適応化などの多様な要素技術が組み込まれている^{(64)~(69)}．ALS との対比を見るために広く普及している代表的な聴覚符号化の AAC とパワースペクトルを原音の場合と比較した例が図 8 である．入力是一般的な CD (44.1 kHz サンプル振幅 16 ビット) のステレオ音楽で左チャンネルのみのスペクトルを表示している．AAC は 144 kbit/s で符号化されているが，聴覚の感度が低い 17 kHz 以上の成分や，原音のスペクトルの谷間の成分の情報が伝わっていないことが分かる．AAC の復号音の音質はほとんど原音と区別できないが，復号音を更にミクシングしたりサラウンド処理をする場合には想定外の音質となることが懸念される．

5.1.3 今後の展開

日本の衛星放送，IPTV などの高品質モードで ALS が普及することが期待される．これにより，スタジオで音楽家や音楽編集の専門家が作成した音響信号と全く同じデジタル信号を一般家庭でも安心感を持って鑑賞することができるようになる．

オーディオ業界ではサンプリング周波数の高いハイレゾ (ハイレゾリューション) 音楽の関心が高まってきている．JEITA (電子情報技術産業協会) は CD 音源よりもサンプリング周波数または量子化精度が高い音源を指すと定義している．なお人間の耳は 20 kHz までの信号しか検知できないことが分かっているが，20 kHz 以上の信号は脳内の活動に影響を与えていることも判明している⁽⁷⁰⁾．このような効果はまだ十分解明されているとは言い難いが，ハイレゾの符号化にはひずみのない圧縮が

必須である。また、将来の研究の進展に期待して、音源の長期保存などには最低でも 96 kHz サンプルの信号のひずみのない符号化が使われる。この点でも ALS の利用が期待されている。あわせて音響信号以外のデータも含めてパッケージとして長期保存をするためのフォーマットの標準規格も MPEG-A PA-AF (Professional Archival Application Format) として制定されている^{(71)~(73)}。

ハイレゾ音楽のロスレス圧縮には FLAC⁽⁷⁴⁾などのフリーソフトや独自方式も使われている。これらの方式はビデオや多重化の標準が制定されておらず、長期の保守や維持管理は保証されていないという問題がある。この点で MPEG 標準である ALS の利用が期待されている。また ALS は柔軟な符号化であるので、音音楽だけでなく生体信号⁽⁷⁵⁾や環境信号などの圧縮符号化にも期待されている。

5.2 携帯電話向け音声符号化 3GPP EVS

5.2.1 経緯

移動通信システムの国際標準化団体である 3GPP では移動通信システム全体が回線交換から IP ベースの LTE (Long Term Evolution) に更改されるタイミングに合わせて EVS (Enhanced Voice Services) 向けの音声符号化の標準化を開始した。要求条件の確定や、予備試験、膨大な品質評価試験を経て、EVS 符号化方式として (以降方式名としても EVS とする) 音声符号化標準規格を 2014 年に制定した。この規格には Ericsson, Fraunhofer IIS, HUAWEI, NOKIA, ORANGE, Qualcomm, Samsung, VoiceAge, ZTE のほか、日本からもパナソニック, NTT ドコモ, NTT による共同開発技術が採用されている。

これまでの携帯電話向け音声符号化は CELP 方式が用いられ、人間の声を低いビットレートにもかかわらず高品質で伝送してきている。しかしながら、音楽など一般の音響信号に対して符号化ひずみが極端に目立つことや、2000 年以降進展した帯域拡張技術や、VoLTE のように IP ベースの伝送でビット単位での符号誤りを考慮しなくてよい技術が生かされていなかった。これらの課題の解決が標準化の要求条件となった。

5.2.2 技術

EVS では CELP に加えて新たに開発された低遅延の音楽向けの音響符号化方式を入力に合わせてフレーム単位で切り換える。更に後述のように、高圧縮、低遅延、広帯域、高音質の両立を目指して多くの技術が結集された^{(76)~(82)}。図 9 はこれまでに標準化された代表的な符号化方式では同時に達成できなかった高圧縮、低遅延、広帯域、高音質を EVS が実現していることを示している。なお放送用符号化は音響符号化であるが、商用のビットレートでは音声品質も良好で、音声品質はこれまでのいずれの符号化でも大きな問題はなかった。

一般の LTE で利用できるビットレートは増大しているが、利用者の増大にも対応できる優先的帯域保障を実現する VoLTE (Voice over LTE) のビットレートは従来の回線交換方式の携帯

	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	
ビットレート [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	

図 9 EVS による機能の達成 (高圧縮, 低遅延, 広帯域, 高音質。背景が斜線になっている枠が好ましい性能を表現しており、これまでの標準方式では各用途での性能が優先で、全ての条件で達成されたものはなかった。)⁽¹⁾

電話より大きく増やすことができないという制約がある。このため依然として品質や高圧縮性能の改善は必須である。この目的で例えばこれまで永く固定的に使われてきた線形予測分析用のラグ窓の係数の適応化⁽⁸³⁾技術等、数多くの要素技術が考案され採用されている。一方 LTE によるパケット単位の伝送ではシステムとしてパケット全体が管理されるので、ペイロード中のビット誤りを考慮する必要がなくなる。このため情報圧縮の基本技術である可変長符号化やモード切替が多角的に使われている。これまでの 3G (第三世代) 携帯電話での回線交換ではビット列の誤りが避けられないので、1 ビットの誤りによってフレーム全体の情報が正しく復号できなくなる可能性を持った可変長符号化やモード切替は使われなかった。

EVS はこれまで固定電話や 3G 携帯電話で使われてきたサンプリングレートが 8 kHz の音だけではなく、VoLTE で使われている 16 kHz (AM ラジオ相当) や 32 kHz (FM ラジオ相当)、更に 48 kHz (TV 放送相当) の音までの幅広い周波数帯域をサポートしている。サンプリングレートが例えば 4 倍になると波形の単純な符号化では 4 倍の情報量が必要になる。このため低ビットを維持したままで、広い周波数帯域を符号化するには帯域拡張技術が不可欠である。帯域拡張技術は時間領域の RELP (Residual Excited LP)⁽⁸⁴⁾や HE-AAC⁽²³⁾などに使われているサブバンド領域での SBR (Spectral Band Replication) が考案されているが、EVS では更に発展させた技術も使われている。いずれも低周波成分 (例えば 4 kHz 以下) は通常の波形ひずみを最小化する符号化を適用し、高帯域側は波形でなくパワースペクトルの大雑把な形状だけをパラメータとして低ビットで符号化する。この手法は高域の位相成分を無視することになり、聴覚の特性を反映させたものになる。

従来の携帯電話や IP 電話の音声符号化は低遅延 (20 ms フレーム) で時間領域の符号化が使われていたが、音楽などの音声以外の信号に対して品質が大幅に劣化するという問題があった。一方従来の音響符号化はフレーム長や分析窓を長くとらざるを得なかった。EVS ではフレーム長や窓長を要求条件範囲内に短縮した周波数領域での音響符号化を用意し、入力によってフレームごとに音声符号化と切り換える。音響符号化ではフレーム長や窓長の短縮で特に継続的な調波構造を持つ音楽などの品

図 10 携帯電話音声の主観品質評価（日本語入力，日本人評価）¹⁾

質劣化が最も深刻であった。これに対応して調波構造のモデルを利用する量子化などで品質の改善を達成している。

VoLTE はビット誤り耐性は不要であるが，パケット全体が失われる可能性はある。パケット消失のあったフレームの音声をそれよりも前に送られたフレームの情報を使って消失したフレームの波形を合成して違和感を小さくしている。更に，ビットレートも 5.9 から 128 kbit/s まで対応し，入力周波数帯域とビットレートにフレームごとに切り換えることができる。また既に 2014 年から普及している VoLTE で使われている AMR-WB (adaptive multi-rate wideband)⁸⁵⁾と相互接続できるモードを備えているので，サービスの円滑な切替が可能である。

5.2.3 評価と今後の展開

標準化の過程において，様々な条件，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⁸⁶⁾や個別に報告されている⁸⁷⁾。これによれば EVS は総合的に従来の音声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る。

図 10 に日本語による携帯電話音声の主観品質評価結果を示す。入力には音声，背景雑音音声，音楽が含まれる。参照音（超広帯域）と比較した評価値である。従来の 3G 携帯では AMR 方式による狭帯域，VoLTE (HD) では AMR-WB 方式による広帯域，最新の VoLTE (HD+) では EVS 方式による超広帯域の音声が出力される。約 13 kbit/s の情報量で比較すると，格段の品質改善が確認される。EVS は 2016 年 5 月から NTT ドコモの VoLTE (HD+) としてサービスが始められた。相前後してアメリカ，ヨーロッパでもサービスが始まっており，今後順次世界的な普及が期待されている。これまでは通信会社，あるいは国際間の通話のほとんどは何度も圧縮符号化，復号を繰り返されざるを得なかった。今後，近い将来，真に統一規格としての EVS の国際的普及で世界の通信会社間をまたがった通話でも一度符号化された圧縮符号列が IP のペイロードとして伝達されることで，高品質，低遅延，広帯域の通話が実現されるものと期待される。

音声音響符号化技術の進展と線形予測符号化技術がほぼ全ての種類の音声音響符号化技術で使われるようになったを紹介した。更に最近の実用例としてロスレス符号化 MPEG-4 ALS と携帯電話用 3GPP EVS を紹介した。音声音響符号化サービスは世界的に広く普及し，低価格化，高品質化の進展を続けている。今後とも臨場感の伝送も含む広い意味での高品質化や便利な機能が追加される方向で技術やサービスが進展するものと思われる。

電話も放送も互換性が最も重要で，これまで国際標準規格の制定に多大な労力が払われ，量産化や品質の保証が達成されてきた。今後のサービスの相互接続性はソフトウェアのダウンロードで確保される可能性が高まってきている。それを支える技術開発はサービスの多様性や開発スピードの利点で，国際標準を経ないでオープンソースなどで開発される可能性がある。これまでは，一旦標準化された技術は改良の意義が小さく，研究開発を抑止する傾向もあったが，今後は随時多様な技術やアイデアが柔軟にサービスに適用されることになるかもしれない。

また一方，センサやデバイスの進展により，様々な分野で大容量のこれまでにない情報源を利用することが予想されている。これまでの音声音響符号化の技術はこのような情報源を人間が活用するための分析，符号化，処理などの技術基盤として幅広く展開されていくことが期待される。

文 献

- (1) 守谷健弘，“音声音響符号化技術の最近の話題，” 情処学音楽情報科学研報 vol. 2015-MUS-107, No. 8, 2015.
- (2) F. Itakura and S. Saito, “Analysis synthesis telephony based upon the maximum likelihood method,” Proc. 6th International Cong. Acoust. (ICA'68), no. C-5-5, pp. C17-20, 1968.
- (3) B. S. Atal and M. R. Schroeder, “Predictive coding of speech signals,” Proc. 6th International Cong. Acoust. (ICA'68), no. C-5-4, pp. C13-16, 1968.
- (4) B. S. Atal, “The history of linear prediction,” IEEE Signal Process. Mag., vol. 23, no. 2, pp. 154 - 161, 2006.
- (5) 板倉文忠，“音声信号に関する統計的・数理的手法 最ゆうスペクトル法から PARCOR を経て LSP へ，” Fundamentals Review, vol. 3, no. 3, pp. 3-25-3-30, Jan. 2009.
- (6) 中田和男，改訂 音声，日本音響学会（編），1995.
- (7) F. Itakura and S. Saito, “On the optimum quantization of feature parameters in the PARCOR speech synthesizer,” Proc. 1972 Conf. Speech Commun. Process. pp. 434 - 437, 1972.
- (8) F. Itakura, “Line spectrum representation of linear predictive coefficients of speech signals,” J. Acoust. Soc. Am., vol. 57, no. 533(A), 1975.
- (9) F. Itakura, T. Kobayashi, and M. Honda, “A hardware implementation of a new narrow and medium band speech coding,” Proc. ICASSP, vol. 7, pp. 1964 - 1967, 1982.
- (10) 守谷健弘 “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.
- (11) M. R. Schroeder and B. S. Atal “Code-excited linear prediction(CELP): High-quality speech at very low bit rates,” Proc. ICASSP, vol. 10, pp. 937 - 940, 1985.
- (12) I. A. Gerson and M. A. Jasiuk, “Vector sum excited

- linear prediction (VSELP) speech coding at 8 kbps," Proc. ICASSP, vol. 1, pp. 461 - 464, 1990.
- (13) 三樹 聡, 守谷健弘, 間野一則, 大室 伸, "ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP)," 信学論 (A), vol. J77-A, no. 3, pp. 314-324, April 1994.
 - (14) J.-P. Adoul, P. Mabilieu, M. Delprat and S. Morissette "Fast CELP coding based on algebraic codes," Proc. ICASSP, vol. 12, pp. 1957 - 1960, 1987.
 - (15) 3GPP TS 26.071, "Mandatory speech CODEC speech processing functions; AMR speech codec; General description," 1999.
 - (16) 3GPP2. C.S0014-0 Version 1.0, "Enhanced variable rate codec (EVRC)," pp.4-4 - 4-12, 1999.
 - (17) ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)," 1996.
 - (18) R. Salami, C. Laflamme, J.P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: A toll quality 8kb/s speech coder," IEEE Trans. Speech Audio Process. vol. 6, no. 2, pp. 116 - 130, 1998.
 - (19) 守谷健弘, 音声符号化, 電子情報通信学会, 1988.
 - (20) J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," IEEE Trans. Acoust. Speech Signal Process., vol. 34, no. 5 pp. 1153 - 1161, 1986.
 - (21) ISO/IEC 11172-3, "Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio," 1993.
 - (22) ISO/IEC 13818-7, "Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced audio coding," 1997.
 - (23) ISO/IEC 14496-3, "Information technology - Coding of audio-visual objects - Part 3: Audio," 2009.
 - (24) J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding [Standards in a Nutshell]," IEEE Signal Process. Mag., vol. 25, no. 3, pp. 137 - 142, 2008.
 - (25) 杉山昭彦, "オーディオ符号化とその応用 これまでの 25 年と今後の展望," Fundamentals Review, vol. 6. no. 1, pp. 51-60, July 2012.
 - (26) ISO/IEC 23003-2, "Information technology - MPEG audio technologies - Part 2: Spatial audio object coding (SAOC)," 2010.
 - (27) ISO/IEC 23008-3, "Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio," 2015.
 - (28) ISO/IEC 23003-3, "Information technology - MPEG audio technologies - Part 3: Unified speech and audio coding," 2012.
 - (29) S. Quackenbush, "MPEG unified speech and audio coding," IEEE MultiMedia, vol. 20, no. 2, pp. 72-78, June 2013.
 - (30) M. Neuendorf, M. Multrus, N. Rettelbach, G. Fuchs, J. Robilliard, J. Lecomte, S. Wilde, S. Bayer, S. Disch, C. Helmrich, R. Lefebvre, P. Gournay, B. Bessette, J. Lapiere, K. Kjolring, H. Purnhagen, L. Villemoes, W. Oomen, E. Schuijers, K. Kikuri, T. Chinen, T. Norimatsu, C. K. Seng, E. Oh, M. Kim, S. Quackenbush, and B. Grill, "MPEG unified speech and audio coding - The ISO/MPEG standard for high-efficiency audio coding of all content types," Proc. AES 132nd Convention Paper, no. 8654, April 2012.
 - (31) 3GPP TS 26.290, "Audio codec processing functions; Extended adaptive multi-rate wideband (AMR-WB+) codec; Transcoding functions," Release 11, 2012.
 - (32) R. Salami, R. Lefebvre, A. Lakaniemi, K. Kontola, S. Bruhn, and A. Taleb, "Extended AMR-WB for high-quality audio on mobile devices," IEEE Commun. Mag., vol. 44, no. 5, pp. 90-97, May 2006.
 - (33) T. Robinson, "SHORTEN: Simple lossless and near-lossless waveform compression," Cambridge Univ. Eng. Dept., Tech. Rep. no. 156, Cambridge, UK, 1994.
 - (34) M. Hans and R. W. Schafer, "Lossless compression of digital audio," IEEE Signal Process. Mag., vol. 18, no. 4, pp. 21-32, July 2001.
 - (35) S. Salomon and G. Motta, Handbook of data compression, pp. 279-280, pp. 1018-1030, Springer, 2010.
 - (36) ISO/IEC 14496-3, "Information technology - Coding of audio-visual objects - Part 3: Audio subpart 11, audio lossless coding," 2005.
 - (37) T. Liebchen, Y. Reznik, T. Moriya, and D. T. Yang, "MPEG-4 audio lossless coding," Proc. AES 116th Convention, no. 6047, May 2004.
 - (38) T. Liebchen and Y. Reznik, "MPEG-4ALS: an emerging standard for lossless audio coding," Proc. Data Compression Conference 2004, pp. 439-448, March 2004.
 - (39) Y. Reznik, "Coding of prediction residual in MPEG-4 standard for lossless audio coding (MPEG-4 ALS)," Proc. ICASSP, vol. 3, pp. 1024-1027, 2004.
 - (40) T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, and Y. A. Reznik, "The MPEG-4 audio lossless coding (ALS) standard - technology and applications," Proc. AES 119th Convention, no. 6589, Oct. 2005.
 - (41) ITU-T G.711.0, "Lossless compression of G.711 pulse code modulation," 2009.
 - (42) N. Harada, Y. Kamamoto, T. Moriya, Y. Hiwasaki, M. A. Ramalho, L. Netsch, J. Stachurski, L. Miao, H. Taddei, and F. Qi, "Emerging ITU-T standard G.711.0 - lossless compression of G.711 pulse code modulation," Proc. ICASSP, pp. 4658 - 4661, 2010.
 - (43) N. Harada, Y. Kamamoto, and T. Moriya, "Lossless compression of mapped domain linear prediction residual for ITU-T recommendation G.711.0," Proc. Data Compression Conference 2010, p. 532, March 2010.
 - (44) B. Atal and M. Schroeder, "Predictive coding of speech signals and subjective error criteria," IEEE Trans. Acoust. Speech Signal Process., vol. 27, no. 3, pp. 247 - 254, 1979.
 - (45) J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. Sel. Areas Commun., vol. 6, no. 2, pp. 314 - 323, 1988.
 - (46) T. Moriya, R. Sugiura, Y. Kamamoto, H. Kameoka, and N. Harada, "Progress in LPC-based frequency-domain audio coding," APSIPA Trans. on Signal and Information Processing, vol. 5, 2016.
 - (47) T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," IEEE J. Sel. Areas Commun., vol. 6, no. 2, pp. 425-431, 1988.
 - (48) N. Iwakami, T. Moriya, and S. Miki, "High-quality audio-coding at less than 64 kbit/s by using transform-domain weighted interleaved vector quantization (TWINVQ)," Proc. ICASSP, vol. 5. pp. 3095-3098, 1995.
 - (49) T. Moriya, N. Iwakami, K. Ikeda, and S. Miki, "Extension and complexity reduction of TwinVQ audio coder," Proc. ICASSP, vol. 2, pp. 1029-1032, 1996.
 - (50) ISO/IEC 14496-3, "Information technology - Coding of audio-visual objects - Part 3: Audio subpart 4, general audio," 2009.
 - (51) G. D. T. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre and post-filters and lossless compression," IEEE Trans. Speech Audio Process., vol. 10, no. 6, pp. 379 - 390, 2002.
 - (52) R. Lefebvre, R. Salami, C. Laflamme, and J.-P. Adoul, "8 kHz/s coding of speech with 6 ms frame-length," Proc. ICASSP, vol.2, pp. 612-615, 1993.
 - (53) R. Lefebvre, R. Salami, C. Laflamme, Adoul, J.-P., "High quality coding of wideband audio signals using transform coded excitation (TCX)," Proc. ICASSP, vol.1, pp. I/193 - I/196, 1994.
 - (54) I. Trancoso and B. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders," Proc.

- ICASSP, vo. 11, pp. 2375 - 2378, 1986.
- (55) I. Trancoso and B. Atal, "Efficient search procedures for selecting the optimum innovation in stochastic coders," IEEE Trans. Acoust. Speech Signal Process., vol. 38, no. 3, pp. 385 - 396, 1990.
- (56) J.M. Tribolet and R. E. Crochiere, "Frequency domain coding of speech," IEEE Trans. Acoust. Speech Signal Process., vol. 27, no. 5, pp. 512 - 530, 1979.
- (57) P. Zelinski and P. Noll, P "Adaptive transform coding of speech signals," IEEE Trans. Acoust. Speech Signal Process., vol. 25, no. 4, pp. 299 - 309, 1977.
- (58) R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka, and T. Moriya, "Representation of spectral envelope with warped frequency resolution for audio coder," Proc. EUSIPCO, vol. TU-L03-1, pp. 51 - 55, Sept. 2014.
- (59) R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka, and T. Moriya, "Resolution warped spectral representation for low-delay and low-bit-rate audio coder," IEEE/ACM Trans. Audio, Speech, and Language Processing, vol. 23, no. 2, pp. 288-299, Feb., 2015.
- (60) R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka, and T. Moriya, "Golomb-Rice coding optimized via LP for frequency domain audio coder," Proc. GlobalSIP, pp. 1024 - 1028, 2014.
- (61) R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka, and T. Moriya, "Optimal coding of generalized-gaussian-distributed frequency spectra for low-delay audio coder with powered all-pole spectrum estimation," IEEE/ACM Trans. Audio, Speech, and Language Processing, vol. 23, no. 8, pp. 1309 - 1321, Aug. 2015.
- (62) ISO/IEC 14496-3:2009/Amd 2:2010, "ALS simple profile and transport of SAOC," 2009, 2010.
- (63) Y. Kamamoto, T. Moriya, N. Harada, and Y. Hiwasaki, "Low-complexity PARCOR coefficient quantization and prediction order estimation designed for entropy coding of prediction residuals," Acoust. Sci. Technol., vol. 34, no. 2, pp. 105-112, 2013.
- (64) 鎌本 優, 守谷健弘, 西本卓也, 嵯峨山茂樹, "チャンネル相関を用いた多チャンネル信号の可逆圧縮符号化," 情処学論, vol. 46, no. 5, pp. 1118 - 1128, 2005.
- (65) 鎌本 優, 守谷健弘, 原田 登, 西本卓也, 嵯峨山茂樹 "ISO/IEC MPEG-4 audio lossless coding (ALS) におけるチャンネル内とチャンネル間の長期予測," 信学論(B), vol. J89-B, no.2, pp. 214 - 222, Feb. 2006.
- (66) 守谷健弘, 原田 登, 鎌本 優, 関川 浩, 白柳 潔, "MPEG-4 ALS -歪みを許さない「ロスレス・オーディオ符号化」の国際標準," NTT 技術ジャーナル, vol. 18, no. 6, pp. 42 - 44, 2006.
- (67) Y. Kamamoto, N. Harada, and T. Moriya, "Multi-channel linear prediction method compliant with the MPEG-4 ALS," IEICE Trans. Fundamentals, vol. E91-A, no. 3, pp. 756 - 762, March 2008.
- (68) 鎌本 優, 原田 登, 守谷健弘, Csaba 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.
- (69) 鎌本 優, 守谷健弘, 原田 登, "携帯電話や放送をきれいな音で伝えるための音声符号化技術," NTT 技術ジャーナル, vol. 28, no. 9, pp. 9 -12, 2016.
- (70) 大橋 力, 音と文明, 岩波書店, 2003.
- (71) ISO/IEC 23000-6, "Information technology - Multimedia application format (MPEG-A) - Part 6: Professional archival application format," 2012.
- (72) N. Harada, Y. Kamamoto, T. Moriya, Hendry, H. Sabirin, and M. Kim, "Archive and preservation of media contents using MPEG-A," IEEE Multimedia, vol. 17, no. 4, pp. 94 - 99, Oct.-Dec. 2010.
- (73) 原田 登, 鎌本 優, 守谷健弘, "MPEG-4 Audio Lossless Coding (ALS) を用いた楽曲マスター音源アーカイブシステム," 情処学論, vol.57, no.5, pp.1355-1364, May 2016.
- (74) <https://xiph.org/flac/>
- (75) Y. Kamamoto, N. Harada, and T. Moriya, "Interchannel dependency analysis of biomedical signals for efficient lossless compression by MPEG-4 ALS," Proc. ICASSP, pp. 569 - 572, 2008.
- (76) G. Fuchs, M. Multrus, M. Neuendorf, and R. Geiger, "MDCT-based coder for highly adaptive speech and audio coding," Proc. EUSIPCO, pp. 1264-1268, 2009.
- (77) 3GPP TS 26.441, "Codec for enhanced voice services (EVS); General overview," Release 12, 2014.
- (78) 3GPP TS 26.445, "Codec for enhanced voice services (EVS); Detailed algorithmic description," Release 12, 2014.
- (79) M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, H. Pobloth, 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. ICASSP, pp. 5698 - 5772, 2015.
- (80) G. Fuchs, C. Helmrich, G. Marković, M. Neusinger, E. Ravellil, and T. Moriya, "Low delay LPC and MDCT-based audio coding in EVS," Proc. ICASSP, pp. 5723 - 5727, 2015.
- (81) T. Bäckström and C. Helmrich, "Arithmetic coding of speech and audio spectra using TCX based on linear predictive spectral envelopes," Proc. ICASSP, pp. 5127 - 5131, 2015.
- (82) T. Moriya, Y. Kamamoto, N. Harada, T. Baekstroem, C. Helmrich, and G. Fuchs. "Harmonic model for MDCT based audio coding with LPC envelope," Proc. EUSIPCO 2015, no. AASP-L1.2, pp. 794 - 798, 2015.
- (83) Y. Kamamoto, T. Moriya, and N. Harada "Adaptive selection of lag-window shape for linear predictive analysis in the 3GPP EVS codec," Proc. IEEE GlobalSIP, pp. 493 - 497, 2015.
- (84) Chong Un and D. Magill, "The residual-excited linear prediction vocoder with transmission rate below 9.6 kbits/s," IEEE Trans. Commun., vol. 23, no. 12, pp. 1466 - 1474, 1975.
- (85) 3GPP TS 26.171, "Speech codec speech processing functions; Adaptive multi-rate - Wideband (AMR-WB) speech codec; General description," Release 5, 2001.
- (86) 3GPP TR 26.952, "Codec for enhanced voice services (EVS); Performance characterization," Release 12, 2014.
- (87) 鎌本 優, 守谷健弘, 原田 登, "3GPP 標準 EVS コーデックの日本語音声品質評価," 春季音響論集, no. 3-2-9, pp. 285 - 286, 2016.

(EA 研究会提案, 平成 28 年 12 月 13 日受付)

守谷健弘 (正員:フェロー)

1980 東大大学院工学系研究科計数修士課程了。同年日本電信電話公社(現 NTT)入社。以来, 音声符号化の研究開発, 標準化に従事。1989 工博。同年 AT&T ベル研究所客員研究員。現在, 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, IEEE (フェロー), 情報処理学会, 日本音響学会各会員。

鎌本 優 (正員)

2003 慶大・理工・物理情報卒。2005 東大大学院情報理工学系研究科システム情報修士課程了。同年日本電信電話株式会社入社。以来, 信号処理及び情報理論に関する研究に従事。2012 博士(情報理工学)。現在, NTT コミュニケーション科学基礎研究所勤務。情報処理学会, 日本音響学会, IEEE 各会員。

原田 登 (正員)

1995 九工大・情報工卒。1997 同大学院情報システム修士課程了。同年日本電信電話株式会社入社。以来、音声・楽音の高能率符号化、音声情報処理、VoIP 通信、マルチメディアコンテンツアーカイブ / 長期保存の研究に従事。2017 博士 (工学)。現在、NTT コミュニケーション科学基礎研究所勤務。情報処理学会、日本音響学会、AES、IEEE 各会員

杉浦亮介

2015 東大大学院情報理工学系研究科システム情報修士課程了。同年日本電信電話株式会社入社。以来、信号処理、情報理論、及び音声音響符号化に関する研究に従事。現在、NTT コミュニケーション科学基礎研究所勤務。日本音響学会、IEEE 各会員。