（4）モバイルマルチメディアにおける符号化技術の進化

メディア符号化技術に関する最新動向をAMR-WB，AAC+，H.264の音声，音響および動画符号化の最新標準として紹介する．また，USA研究所における研究成果を含め，今後の符号化技術の発展方向について述べる．

栄藤（えとう） 稔（みのる） Khosrow Lashkari（コズロー ラシュカリ）
Frank Bossen（フランク ボッセン） Wai Chu（ワイ チュー）

1. まえがき

無線ネットワークにおけるメディア符号化を考えるとき，2つの重要な法則がある．1つは有名な「ムーアの法則」で，半導体の処理能力が18～24カ月で2倍になるというものである．「ムーアの法則」は，符号化・復号化（CODEC：COder DECoder）の進化にも当てはまり，MPEG（Moving Pictures Experts Group）のMPEG-2標準化から10年間で符号化効率は飛躍的に向上した．もう1つはまだ法則としては知られておらず，筆者の1人が唱えているにすぎないが，“無線ネットワークと有線ネットワークとの間に常に1桁～2桁の伝送帯域ギャップが存在する”という“法則”である．この帯域幅ギャップのために，無線ネットワークではメディアデータを効率良くコンパクトに表現する符号化技術が要求される．伝送されるメディア品質を向上させるには無線アクセス技術だけでなく，メディア符号化技術も欠かすことができない．USA研究所では過去3年にわたりドコモのマルチメディア研究所と共同で音声，音響，および映像の圧縮技術を開発してきた．

本稿では，2001年に本誌で発表した文献[1]以降の技術の進歩と，より高度な符号化技術に向けた研究の方向性について説明する．既存のCODECは，過去のハードウェアアーキテクチャの制約に適合するように設計されてきた．これに対して，将来のCODECはムーアの法則から予測される半導体技術の進歩を考慮して，より自由な発想で開発されるべきである．computational complexity（計算複雑さ）*の増加を許容することにより，符号化効率を確実に向上させることができる．これは，CODECの進化に対するUSA研究所の基本理念である．

＊ computational complexity（計算複雑さ）：計算機学科，情報科学科で使われる用語．チューリング機械（現在のコンピュータのモデル）を用いて問題を解くのに必要な空間量と時間量のことを表す．

2. 音声および音響符号化

音声および音響符号化とは，音声・音響信号のコンパクトなディジタル表現を求めることである．その目的は信号を正確に再現することではなく，“知覚的に等価な波形再現”にある．これは，信号から冗長性（音声の場合）と非相関性（音楽の場合）を除去することで達成される．つまり，目的は同じであっても，音声と音響の符号化技術は大きく異なる．

2.1 音声符号化

音声符号化は，いわゆる発声機構特有の情報源モデルを使って，音声の音源，すなわち人間の発声器官（声門，口腔，唇）をモデル化するものである．音声符号化アルゴリズムは，音声信号に極めて特化されており，一般に音楽の符号化には適していない．これまでにも，多数の符号化標準規格が開発されてきた．文献[1]には，音声および音響符号化標準規格が年代順，標準化団体別にリストアップされている．図1にさまざまなアプリケーションのデータ転送速度と，それらに使用されているCODECの平均オピニオン値（MOS：Mean Opinion Score，音声および音響の知覚品質の尺度）を示す．一般に，音声CODECは以下の3つの主要カテゴリに分類できる．

① パルス符号変調（PCM：Pulse Code Modulation），差分PCM（DPCM：Differential Pulse Code Modulation），

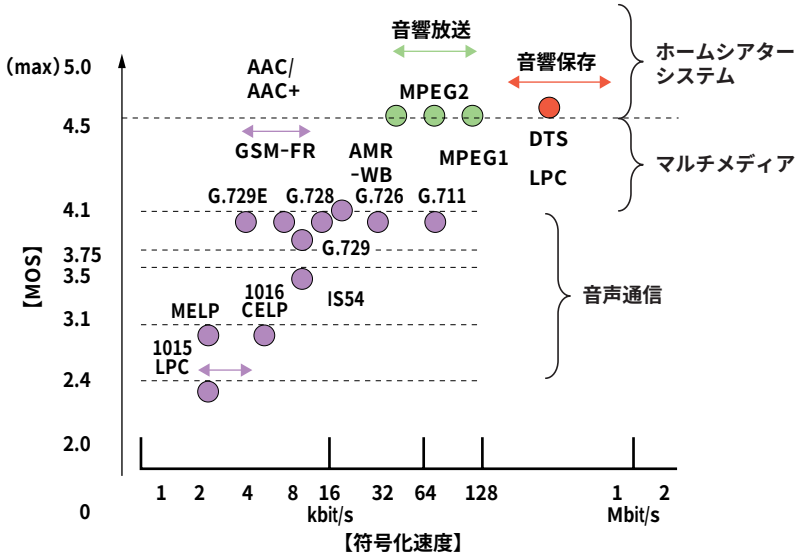

図1 用途および符号化速度によるCODECの位置付け

適応的差分PCM（ADPCM：Adaptive Differential Pulse Code Modulation）などの波形CODEC[2].

② 線形予測符号化（LPC：Linear Predictive Coding）などのボコーダ.

③ 符号励振線形予測（CELP：Code Excited Linear Prediction）[3]などや適応マルチレート（AMR：Adaptive Multi Rate）CODECなどのハイブリッドCODEC.

FS1015，FS1016，およびMELP（Mixed Excitation Linear Prediction）などの標準規格は，狭帯域無線チャネルを前提としてセキュリティの高い通信を実現するために開発されたもので，音質は高くない．一方，国際電気通信連合・電気通信標準化部門（ITU-T：International Telecommunications Union, Telecommunication standardization sector）の標準規格G.729は，携帯電話ネットワーク上で十分な音質を実現するものとして設計された．RPE-LTP（Regular Pulse Excitation-Long Term Prediction）は，GSM（Global System for Mobile communications）携帯電話での利用を目的に欧州電気通信標準化機構（ETSI：European Telecommunications Standards Institute）が開発したものである．AMR-NB（Adaptive Multi Rate-NarrowBand）は，VoIP（Voice over IP）などの帯域幅が変動するアプリケーションに適しており，3GPP（3rd Generation Partnership Project）において，広帯域符号分割多元接続方式（W-CDMA：Wideband Code Division Multiple Access）システムの必須音声CODECとして採用されている．また，AMR-WB（Adaptive Multi Rate-WideBand）は，テレビ会議などのアプリケーションで対面通話に適した品質を実現している．図2に，AMR-WB CODECの概念図を示す．AMR-WB CODECは，6.6～23.85kbit/sのビットレートを持つ9つのソースコーダで構成されている．このCODECはCELPモデルをベースにしており，サブバンドのフィルタリングメカニズムによって効率的なエンコードを可能にしている．このCODECは多くの点で前世代のCELPに類似しているが，さまざまな画期的技術を組み込むことにより，高効率と高品質を実現している．その1つが16kHzでサンプリングされた広帯域（7kHz）音声に対応するように設計されている点である．一方，USA研究所では，符号化の遅延を抑制し，信号対雑音比（SNR：Signal to Noise Ratio）を改善した標準CODECの最適化技術を開発している[4][5].

2.2 音響符号化

音響にはさまざまな楽音があり，普遍的な情報源モデルが存在しないため，音響符号化では，音源モデルではなく，人間の聴覚モデルにより符号化を行う．MPEG-1[6]音響符号化は，レイヤI，II，IIIと呼ばれる3つの符号化スキームで構成され，32kbit/s～448kbit/sのビットレートをサポートしている．MP3（MPEG-1 Audio Layer-3）は，CDに近い音響品質を特長とする．MPEG-2は，より低いサンプリング周波数（16kHz，22.05kHz，および24kHz）と5.1音響などのマルチチャンネル音響に対応するようにMPEG-1を拡張したものである．MP3の後継バージョンになり得るMPEG-2のAAC（Advanced Audio Coding）は，MPEG-1と下位互換性がないが，128kbit/sで実際の音源と区別できないほどクリアなステレオ音響品質を実現している．また，AACでは，96kHzまでのサンプリングレートをサポートしており，CDを超える音質の超高品位音響を可能にしている．aacPlus（AAC+とも表記される）は最新の音響CODECであり，低いビットレートで高品質の音響を特徴としている．ほとんどの音響CODECは，128kbit/sを境として知覚品質が低下し始めるが，AAC+はスペクトルバンドレプリケーション（SBR：Spectral Band Replication）と呼ばれる技術を用いて，48kbit/sでも優れたステレオ品質を，32kbit/sでも高品質な音響を実現している．SBRでは，全帯域の音響スペクトルが低域と補完的な高域セクションに分割されている．スペクトルの低域部分はAACコアを用いて符号化されるが，高域部分は直接には符号化されず，この帯域に関するわずかな情報が送信されることで，デコーダが全帯域の音声スペクトルを再構築できるようになっている．AAC+は次の2つの事実を活用して高品質を実現している．第1に，可聴周波数において高周波が与える聴覚心理学的影響が比較的低いということ．第2に，音響スペクトルの低周波と高周波には利用可能な高い相関性があるということである.

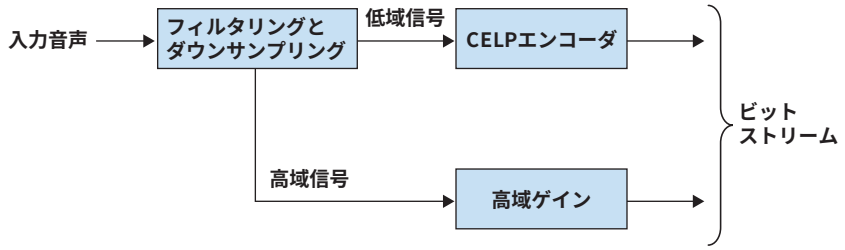

図2　AMR-WB CODECの概念図

2.3 音声と音響符号化に関する研究の方向性

現在，端末装置には，例えば，音響にAAC，音声にAMRとなるように，音声用と音響用のそれぞれに適したCODECを使用する必要がある．しかしながら，音声と音響の両方を処理できる統合CODECの実現が，極めて望ましい．そこで，USA研究所はマルチメディア研究所と共同で音声音響統合CODEC（USAC：Unified Speech and Audio Coding）の開発に着手した．このCODECはMPEG-4のHILN（Harmonics, Individual Lines and Noise）パラメータ音響符号化をベースにしている．図3にその概念図を示す．基本原理は，音源合成として入力波形をモデル化する情報源モデルに基づいている．このモデルによれば，音声・音響波形は以下の3種類の信号成分の合成から成り立っているとする．

① 基本周波数と部分音の振幅のスペクトル包絡線で表現される倍音．
② 周波数，振幅，位相によって特徴付けられる単一正弦波（「個別ライン」ともいう）．
③ 振幅とスペクトル形状によって特定されるノイズ．

音声・音響信号は時間軸上の区間，フレーム単位で上記3成分に分解されたパラメータとして符号化される．この分解をスケーラブルにして，1つの音源モデルから複数の音源モデルへ調整することにより，音声信号から音響信号まで扱えるようにした点がこのアプローチの特徴である．

USA研究所が目指すもう1つの目標は，無線機器で使用する超小型スピーカーで3D音響や高品質音響を実現するためにCODEC機能を強化することである．USA研究所では，非線形信号処理技術を活用しながら，この目標の実現に向けて取り組んでいる．

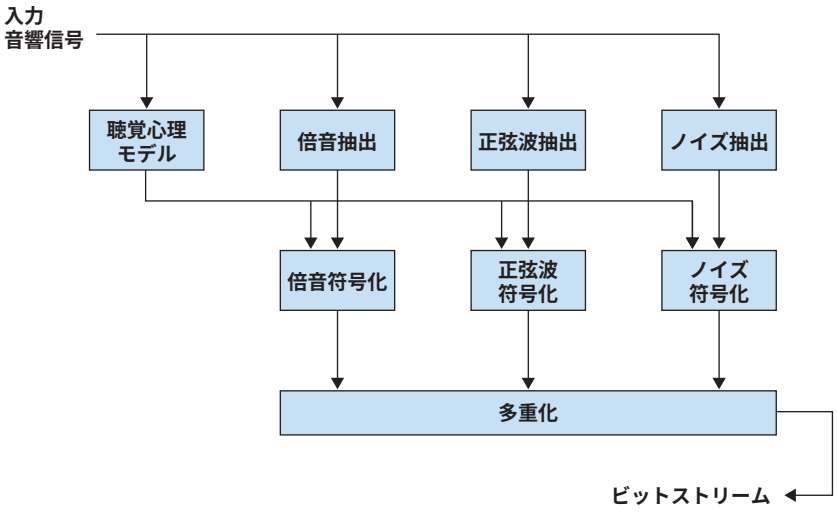

図3　音声音響統合CODECの概念図

3. 動画符号化

3.1 新しい動画CODEC

半導体技術と無線ネットワーク技術の進歩により，モバイル機器でも動画アプリケーションを使用できるようになった．利用可能なリソースを最大限に活用し，リアリティのある映像体験を提供するには，高度な動画圧縮アルゴリズムが必要になる．最近15年間で，主に2つの標準化団体，すなわち国際標準化機構（ISO：International Organization of Standardization）/国際電気標準会議（IEC：International Electrotechnical Commission）のMPEGとITU-TのVCEG（Video Coding Experts Group）によっていくつかの圧縮アルゴリズムが標準化された．これらには，H.261, H.263, MPEG-1, MPEG-2, MPEG-4などの標準規格が含まれている．H.263とMPEG-4は現在，モバイルやテレビ会議などのアプリケーションに使用されているが，その中・低速ビットレートの品質は限られたものである．MPEGとVCEGは最近協力して，JVT（Joint Video Team）を結成し，H.264/AVC（Advanced Video CODEC）[7]を開発した．2003年5月に最終承認されたこの標準規格では，符号化効率が大幅に向上する見込みである．一方でWindows Media Video 9[8]のような独自仕様のソリューションも出現し，事実上の業界標準になることを目指している．

H.264/AVCは，3つのフレームタイプ（I，P，およびB）を含む従来の標準規格（図4のブロック図を参照）で使用されていたのと同じハイブリッド動き補償・変換符号化アーキテクチャをベースにしている．しかし個々の技術では，この符号化規格は大きく異なる技術を使用している．高品質を実現する主な要素の1つは予測精度の改善である．この予測には，空間予測（フレーム内予測）と時間軸方向予測（フレーム間予測）の2つの形式がある．MPEG-4では単純なアルゴリズムを使って水平または垂直方向の係数値を予測するが，H.264/AVCはベースバンド領域で9方向の空間的予測を行う．一時的予測も，1/4画素単位（MPEG-4 Simple Profileは1/2画素単位の精度）の動き補償を使用するとともに，動作境界をより正確に表現できる4×4画素（MPEG-4 Simple Profileは8×8画素）のブロックサイズを実現することで改善されている．フレーム間予測を複数のフレームから選択して行

えることも，符号化効率向上の要因になっている．また，H.264/AVCは，離散コサイン変換（DCT：Discrete Cosine Transform）を4×4画素単位で可逆に整数変換を行うところが他のCODECと異なる点である．従来の標準規格に見られる8×8画素単位の浮動小数点による不可逆演算から脱却することで，エンコーダとデコーダ間のドリフト問題が解消され，すべてのデコーダが同一画像を確実に再構築できるようになる．H.264/AVCのもう1つの特徴は，ブロックノイズを取り除くループフィルタを備えていることである．（後処理ではなく）ループ内でこのフィルタを使用すると，動き補償された予測精度が向上する．最後に，H.264/AVCは高度なエントロピー符号化方式を用いている．これには，コンテクスト適応ハフマン符号とコンテクスト適応算術符号が含まれている．

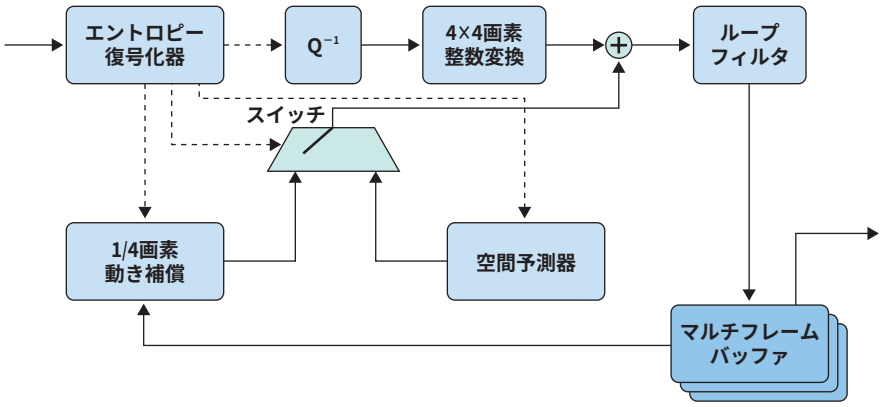

図4　H.264/AVC復号化器のブロック図

Windows Media Video 9は，H.264/AVCと多くの類似点を持つCODECであり，実際，I，P，およびBフレーム，ループフィルタ，2分の1画素単位の動き補償，整数変換を特徴としている．H.264/AVCとの相違点は，多様なサイズのブロック変換と，動き補償に適応したフィルタを備えていることである．ただし，動き補償の最小ブロックサイズが8×8画素に限られ，また算術符号にも対応していない．それでも，両CODECの類似性は顕著であり，興味深い．

3.2 動画符号化に関する研究の方向

現行世代の動画CODECは十分な品質を備えているが，将来はよりいっそうの改善が期待されている．改善にはいくつかの方向がある．

(1) JPEG（Joint Pictures Experts Group）のJPEG-2000とは異なり，H.264/AVCでは，算術符号が既存アルゴリズムの先頭に付加されている．算術符号化を前提として動画CODECを設計すると，さらなるメリットが生まれるであろう．これによって，マクロブロックを基本符号ユニットとしない新たなアーキテクチャが構築できる．

(2) 従来の8×8 DCT変換を脱却しつつあるとはいえ，H.264/AVCは依然としてDCTをベースとした変換に依存している．DCTやKLT（Karhunen-Loève Transform）などの変換は，信号のガウス分布を前提にエネルギー最小化を目的として導かれたものである．しかし，この前提は符号量最小を目的とすると最適とはいえず，この前提に当てはまらずに，圧縮性能を向上させる変換方法があると考えられる．一般化された最適な変換方法の設計においては，エントロピー符号化を視野に入れた最適化プロセスに重点が置かれる．

(3) フレーム間予測は，効率的な符号化の鍵となる．次の2つの条件がフレーム間予測の精度を向上させると思われる．1つは，画像に映る異なる動きの境界をより明確に表現すること，もう1つはフィルタを動き補償ループに適応させることである．以上により，予測符号化がより効率的なものとなる．

(4) エンコーダの最適化は通常，レート・歪み特性を最大限にまで高めることによって達成される．PSNR（Peak Signal to Noise Ratio）は便利な歪み評価尺度であるが，PSNRに比例して常に視覚品質が向上するとは限らない．他の歪み評価尺度（重み付けPSNRなど）を使用しても，知覚される画質が向上する可能性がある．

以上の方向に沿ってUSA研究所は動画CODECの研究開発を行っている．

4. あとがき

本稿ではここ数年のCODEC技術進化を鳥瞰し，AMR-WB，AAC+，H.264/AVCを最新技術として紹介した．そして将来の発展方向について述べた．USA研究所では音声・音響統合符号化，算術符号に基づくビデオCODECなどを軸に研究開発を行っており，今後ともモバイルマルチメディアのニーズに応えるCODECの進化に寄与したいと考えている．

文　献

[1] 中野，ほか：“モバイルマルチメディア信号処理技術概要，”本誌，Vol. 8，No. 4，pp.6-11，Jan. 2001.

[2] N. S. Jayant: “ Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers, ” Proc. IEEE, Vol. 62, pp. 611-632, May 1974.
[3] ITU-T, Recommendation G.729 Coding of speech at 8 kbit/s using Conjugate- Structure Algebraic-Code-Excited- Linear-Prediction (CS-ACELP), ITU-T, Geneva, Sep. 1998.
[4] K. Lashkari and T. Miki: “ Joint Optimization of Model and Excitation in Parametric Speech Coders, ” Proc. ICASSP 2002, pp. 277-280, May 2002.
[5] W. Chu and T. Miki: “ Optimization of Window and LSF Interpolation Factor for the ITU-T G.729 Speech Coding Standard, ” Proc. Eurospeech 2003, pp. 1061-1064, Sep. 2003
[6] ISO/IEC 13818-3: “ Coding of Moving Pictures and Associated Information-Part 3: Audio, ”May 1995.
[7] ITU-T Recommendation H.264 | ISO/IEC 14496-10, Geneva, May 2003.
[8] Microsoft, WMV9-an advanced video CODEC for 3GPP, Document S4(03)0613 submitted to 3GPP, Sep. 2003.

用 語 一 覧

AAC：Advanced Audio Coding
ADPCM：Adaptive Differential Pulse Code Modulation（適応的差分PCM）
AMR：Adaptive Multi Rate（適応マルチレート）
AMR-NB：Adaptive Multi Rate-NarrowBand
AMR-WB：Adaptive Multi Rate-WideBand
AVC：Advanced Video CODEC
CELP：Code Excited Linear Prediction（符号励振線形予測）
CODEC：COder DECoder（符号化・複合化）
DCT：Discrete Cosine Transform（離散コサイン変換）
DPCM：Differential Pulse Code Modulation（差分PCM）
ETSI：European Telecommunications Standards Institute（欧州電気通信標準化機構）
GSM：Global System for Mobile communications
HILN：Harmonics, Individual Lines and Noise
IEC：International Electrotechnical Commission（国際電気標準会議）
ISO：International Organization of Standardization（国際標準化機構）
ITU-T：International Telecommunications Union, Telecommunication standardization sector（国際電気通信連合・電気通信標準化部門）
JPEG：Joint Pictures Experts Group
JVT：Joint Video Team
KLT：Karhunen-Loève Transform
LPC：Linear Predictive Coding（線形予測符号化）
MELP：Mixed Excitation Linear Prediction
MOS：Mean Opinion Score（平均オピニオン値）
MP3：MPEG-1 Audio Layer-3
MPEG：Moving Pictures Experts Group
PCM：Pulse Code Modulation（パルス符号変調）
PSNR：Peak Signal to Noise Ratio
RPE-LTP：Regular Pulse Excitation-Long Term Prediction
SBR：Spectral Band Replication（スペクトルバンドレプリケーション）
SNR：Signal to Noise Ratio（信号対雑音比）
USAC：Unified Speech and Audio Coding（音声音響統合CODEC）
VCEG：Video Coding Experts Group
VoIP：Voice over IP
W-CDMA：Wideband Code Division Multiple Access（広帯域符号分割多元接続方式）
3GPP：3rd Generation Partnership Project