

5

オリンピックと
次世代 MPEG 符号化

—高能率から 3D, フリーナビゲーションへ—



妹尾孝憲 (情報通信研究機構)

金子 格 (東京工芸大学)

安藤彰男 (富山大学)

MPEG とオリンピック

本稿では、MPEG (Moving Picture Expert Group) とオリンピックの関係をたどりつつ、今後の課題の拡大について紹介する。それは高能率化から、3D やフリーナビゲーションへ、いわば圧縮率という量的目標から、質的目標への拡大である。

MPEG は ISO (国際標準化機構) の傘下で動画像と音声の符号化標準を策定してきたが、その標準化はオリンピックの開催に大きな刺激を受けて発達してきた。今回特集のために、MPEG とオリンピックの歴史を振り返り図-1 にまとめた。MPEG 標準化は、オリンピックで初めてハイビジョン放送が実現したソウルオリンピックが開催された 1988 年に始まった¹⁾。

当時 1950 年代から続いた NTSC (カラー TV 放送規格) 方式など従来のテレビ放送方式が改良の限界に達し、日本のアナログハイビジョン方式を始め多くの次世代テレビ技術が提案されていた。その中でデジタル方式である MPEG の標準化も開始されたのである。

その当時デジタルカメラも実用化途上であり、テレビのデジタル化は非常に技術的難易度が高く、20 世紀中には不可能と考えられていた。しかし規格を標準化し、研究開発を集中することで、飛躍的に開発が速まると期待された。MPEG 標準化の開始により、一転して世界中の研究機関が標準化に向けてしのぎを削る状況になった。

その後の何回かのオリンピックは、MPEG 標準が一候補から次世代放送方式の本命へと認知される重要なステップになった。日本はバルセロナ、アトランタの 2 大会で、唯一ハイビジョン方式をアナロ

グ方式で実施し当時先行していた日本のハイビジョン技術の優越性をアピールした。アトランタ大会では米国はグランドアライアンス (大連合) を組み初のデジタルハイビジョン放送を目指したが試験放送は実現できなかった。しかしグランドアライアンスは大きな成果を挙げ、アナログ技術で先行していた日本の勢力もデジタル放送へ大きく舵を切った。欧州も同調し、わずか 4 年後のシドニーオリンピックで、国際標準の MPEG 符号化技術を用いたデジタル放送が実現した。そして現在、世界中ほぼすべてのデジタル放送で MPEG 符号化方式が利用されるようになった。

オリンピック憲章に示されたオリンピックの目標は、人類の調和と平和な社会であるが、デジタル技術の分野でもオリンピックは調和と平和をもたらした、ともいえそうだ。その後も MPEG 符号化技術の発展は続き、HDTV (High Definition Tele-Vision) 映像のブルーレイディスクへの録画やストリーミング配信が実用化し、昨年、8K × 4K 画素のスーパーハイビジョン映像を効率良く符号化する HEVC (High Efficiency Video Coding) を完成させた。

では、今後 2020 年に向けて、さらにどのような次世代標準が計画されているのであろうか？ 現在 10 以上の次世代標準化が進んでいるが、以下ではその中から、オリンピック観戦の臨場感を飛躍的に高め、オリンピックでの実現が期待される 2 つの規格を紹介する。1 つは FTV (自由な視点からの映像を見られる映像符号化)、もう 1 つは 3D オーディオ (上下、前後定位を含めた、その場の臨場感を再現するオーディオ) である。

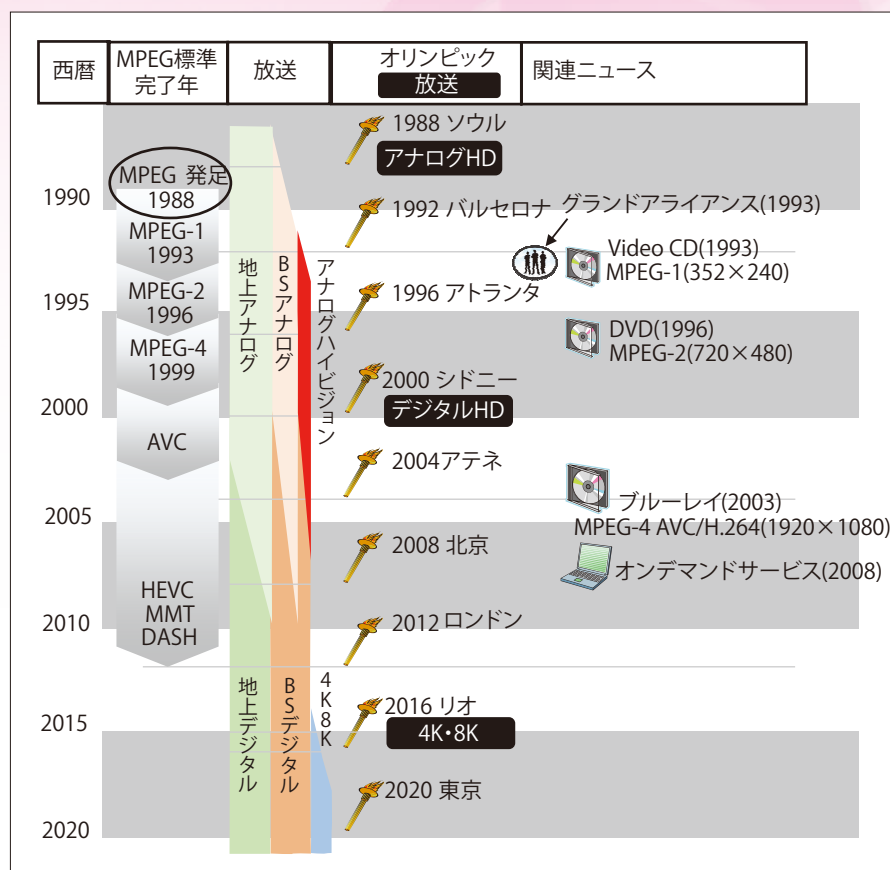


図-1 MPEG とオリンピックの歴史

べて異なり、従来のブロックベースの動き補償予測の適用のみでは、約1.5倍の圧縮率改善にとどまらざるを得なかった。図-2に示すように、2007年から始まった第2フェーズFTVでは、カメラから被写体までの距離が分かれば、コンピュータビジョン技術を用いて、任意視点映像を合成可能であるので、被写体の距離情報を、視点映像間の視差を表す奥行マップとして追加する代わりに、符号化すべき視点映像を間引く方式（MVC+Depth, 3D-AVC, 3D-HEVC）が検討され、約2倍の符号化効率が達成された。

奥行マップは、多視点映像間で対応点探索を行うことや、距離カメラで被写体を照明し

FTV（多視点映像符号化とフリーナビゲーション）

MPEGは、2001年に立体映像符号化方式の標準化検討を開始した。当初は、欧州を中心としたコンピュータグラフィクス技術に基づく立体モデルベースの符号化と、日本を中心とした多視点映像符号化の両方式を検討したが、立体ディスプレイと相性の良い多視点映像符号化方式に絞り、標準化を2005年に開始し、2009年にMVC（Multi-view Video Coding）として完成し、AVC（Advanced Video Coding）規格書に含めた。この標準は、ブルーレイディスクで2眼立体映像を記録・再生するために使われている。この期間は、多視点映像からの自由視点映像合成を主張したグループによって第1フェーズFTV（Free-viewpoint Tele-Vision）と呼ばれる。しかし、静止したシーンであっても、視点位置が異なると、各視点映像内の画素位置がす

た光が反射して戻って来るまでの時間や角度を計測することで得られる。しかし、多視点映像のマッチングでは、前景の被写体の陰に隠れた背景部分の奥行をいかにして得るかというオクルージョン問題や、類似のテクスチャによる偽マッチングの問題があり、距離カメラでは、計測できる奥行範囲が狭い問題や、黒色等の被写体の色の影響等が課題である。

最近、超多視点3Dディスプレイやインテグラル3Dディスプレイに加えて、360°どの方向からでも見ることのできる3Dディスプレイ等が出現したのをきっかけに、図-3に示すように、2013年からは、超多視点映像の符号化と、3D映像空間を自由に動き回れるフリーナビゲーションの実現を目指す第3フェーズFTVの検討が始まった²⁾。これらのディスプレイは、100視点以上の映像を密に配置することによって、それまで実現が困難であった広い視野角の実現と、2眼立体映像の課題であった、被写体に注目すると目の焦点がスクリーン面から離れて

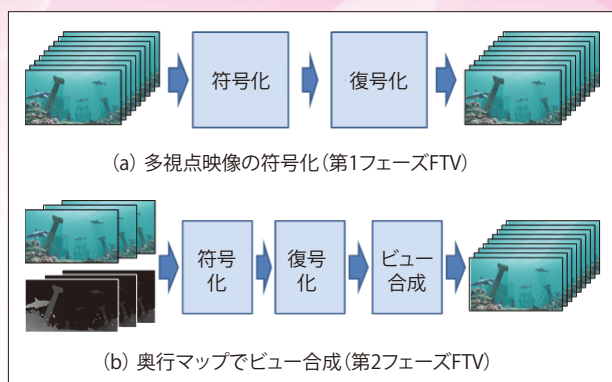


図-2 立体映像符号化



図-3 超多視点映像とフリーナビゲーション (第3フェーズFTV)

像がぼける「輻輳と調整の乖離」の問題を解決している。

第3フェーズFTVでは、図-4に示すように、第2フェーズで開発された奥行探索ソフトウェア DERS (Depth Estimation Reference Software) や、3D-HEVC (Three-Dimensional-High Efficiency Video Coding) コーデック、任意視点映像合成ソフトウェア VSRS (View Synthesis Reference Software) を用いた探索実験を行っているが、非常に多くの視点映像を符号化する必要があるため、符号化効率をさらに上げるとともに、簡単で高速なコーデックが望まれている。また、自由な視点映像の合成

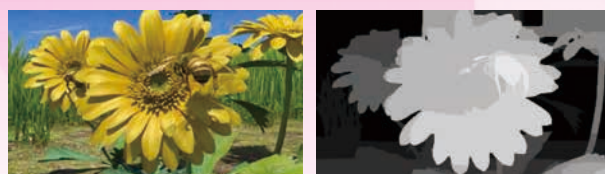


図-4 185視点の中心映像と推定された奥行マップの例



図-5 フリーナビゲーション映像の合成例⁷⁾

FTV	略称	特徴
第1フェーズ	MVC	多視点映像をそのまま圧縮 (低圧縮率)
第2フェーズ	3D V	奥行マップでビュー合成 (高圧縮率)
第3フェーズ	SMV+FN	超多視点映像とフリーナビゲーション

図-6 FTVのフェーズと特徴

を可能にするために、従来の8ビット精度の奥行マップを16ビット精度に拡張することや、オクルージョン問題を解決するために、多数の視点映像から奥行マップを推定する方法等が検討されている。

フリーナビゲーションの検討では、任意のシーンからの自由視点映像の合成は困難であるが、図-5に示すように、シーンがサッカー場等と既知の形状である場合、完全な奥行マップを求めなくても、コート白線等の特徴点を3Dモデルに当てはめることで、実写映像の3Dモデルが得られ、それにコンピュータビジョン技術を適用すれば、任意の視点から見た映像を合成できる。そこに実写シーンから切り出した選手映像を貼り付けた平板であるビルボードを、角度を合わせながら表示することで、フリーナビゲーションを実現する方法等が検討されている。

図-6に、FTVの各フェーズをまとめた。

3D オーディオ符号化

MPEGでは、2013年に、新しい規格MPEG-Hの中に第3部3Dオーディオを設け、3次元音響符号化方式の標準化を開始した。3次元音響は、上方にもチャンネルを配置することにより、音の上下感も表現できる方式であり、5.1チャンネルなどの従来システムよりも高い臨場感を実現できる。3Dオーデ

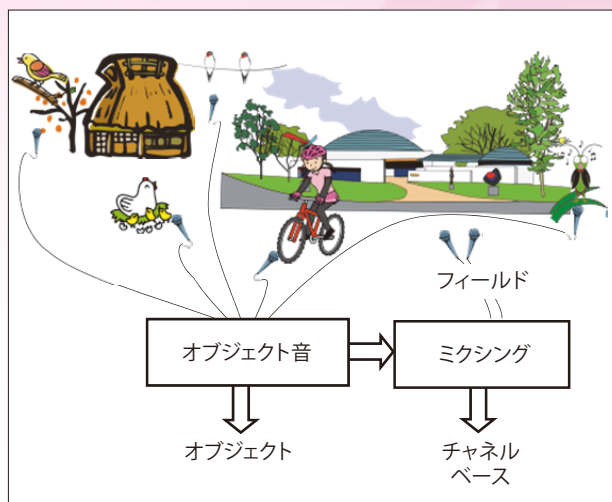


図-7 チャンネルベースとオブジェクトベース

ィオ符号化は、このような音の立体的な空間性も符号化する方式である。MPEG-H 3D オーディオでは、従来のチャンネルベースシステム、オブジェクトベースシステムのほか、将来の音響システムとして研究開発されている高次アンビソニックスシステムも扱うことが定められた。また、ヘッドホンによる簡易な受聴方式としてバイノーラル方式も検討されることになった。

チャンネルベースとは、2チャンネルステレオを拡張したマルチチャンネル音響信号の各チャンネル信号を扱う方式であり、そのチャンネル数は常に一定である。一方、オブジェクトベースとは、オブジェクト音、すなわち明確な対象物が発する音を独立して記録して伝送する方式であり、送られる信号数は常に一定とは限らない。その違いは、オブジェクト音とフィールド音（風が渡る音のような明確な方向性を持たない音）からなる3次元音響の情報を、一定のチャンネル数にミクシングして音場情報を送るか、それともオブジェクト音のみを独立に送るかという点にある。また、チャンネルベースを基本として、ある特定のオブジェクト信号を必要に応じて送ることも検討されている。特に、コメント音声のみをオブジェクト信号として伝送し、デコーダ側で必要に応じてコメント音声の信号レベルを増幅するダイアログ強調が知られている。

高次アンビソニックス信号は、音の方向性情報を

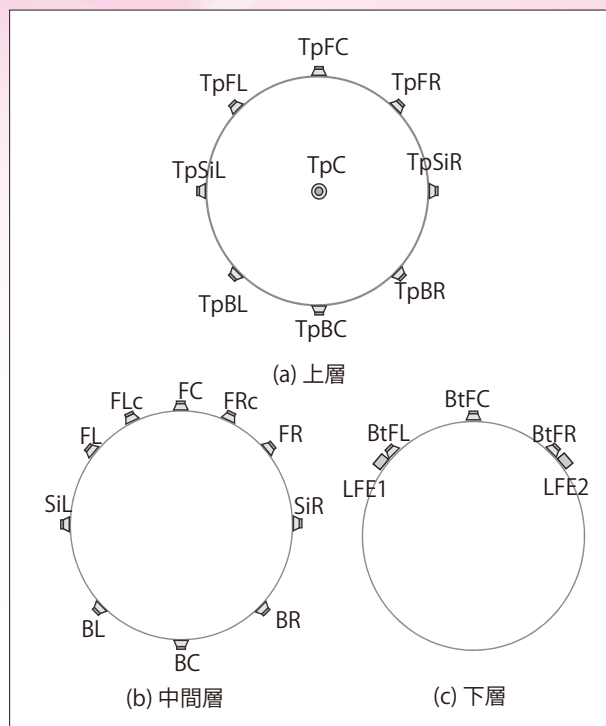


図-8 22.2チャンネルにおけるチャンネル配置

直交関数展開して、その展開係数を符号化／伝送し、デコーダにおいて、送られてきた展開係数に基づいて、元の音場を合成する方式である。高次アンビソニックスは、フィールド音の記録に適した方法であるが、コンテンツを制作する道具立てが不十分なことから、オリンピックでも利用される見通しは立っていない。現状では、臨場感の高いコンテンツを制作可能なチャンネルベース方式か、図-7に示すように、チャンネルベースにオブジェクトベースを併用する方式が有望である。

チャンネルベースでは、高さ方向に3層化したチャンネル配置を持つ22.2チャンネルが、3次元音響の標準的なシステムとして採用されている。そのチャンネル配置を図-8に示す。図-8のチャンネルラベルは、それぞれ、FL: Front left, FR: Front right, FC: Front center, FLc: Front left center, FRc: Front right center, BL: Back left, BR: Back right, BC: Back center, SiL: Side left, SiR: Side right, TpFL: Top front left, TpFR: Top front right, TpFC: Top front center, TpBL: Top back left, TpBR: Top back right, TpBC: Top back center, TpSiL

: Top side left, TpSiR : Top side right, TpC : Top center, BtFL : Bottom Front Left, BtFR : Bottom Front Right, BtFC : Bottom Front Center, LFE1, 2 : Low Frequency Effect Channels である。さまざまな実験的検討の結果, 3次元音響のコンテンツを制作する上で, このようなチャンネル配置が必要であることが知られている³⁾。

MPEG-H 3D オーディオで特に注目されている技術は, 家庭におけるダウンミックスなどの音響レンダリング技術である。たとえば, 5.1 チャンネル信号を 2 チャンネルステレオで聴取する場合には, 元の信号を線形演算によって 2 チャンネル信号に変換するダウンミックス技術が利用される。このようなレンダリング技術では, レンダリング前の信号が有する臨場感を保ちながら, そのチャンネル数を減らす技術が重要である。また, 各チャンネル信号に同相成分が多く含まれると, ダウンミックスの際に, 信号がクリップするという現象が起こる。そこで, オブジェクト信号を効果的に利用しながら, 少ないスピーカ数への音響レンダリングを行う技術とともに, 音のラウドネスを保ちながらクリップを防ぐダイナミックレンジ制御技術の標準化作業が行われている。

ヘッドホン再生は, 3次元音響の簡易な再生方式として知られている。3次元音響の場合でも人間は 2つの耳で聞いているのであるから, 両耳の入り口での音が再現できれば, 立体的な音が再現できるはずである。このため, MPEG-H 3D オーディオでは, バイノーラル室内インパルス応答 (BRIR), すなわち一定の残響がある部屋において, さまざまな到来方向からダミーヘッドの両耳位置までのインパルス応答の利用も検討している。

このほか, 2013 年には, MPEG-4 AAC (Advanced Audio Coding) において 22.2 チャンネルの符号化方式が定められた⁴⁾。これを受けて, 2016 年からの超高精細度テレビジョン放送実用化試験放送では, 22.2 チャンネルも含めた音声の符号化方式として, 基本サービス用に MPEG-4 AAC が採用され, ロスレスサービス用に MPEG-4 ALS (Audio Lossless Coding) が採用される予定である。

Green Metadata	グリーン (省エネルギー) 符号化
Uniform Signaling for Timeline Alignment	スマホなどの外部機器との再生タイミング同期
Mixed and Augmented Reality Reference Model	拡張現実のための符号化, システム標準
Compact Descriptors for Video Analysis	動画像の分析のためのコンパクトな特徴記述の符号化
Multimedia Preservation Application Format	マルチメディア長期保存フォーマット
MPEG-UD	ユーザ記述の符号化

表-1 現在進行中の MPEG 標準
(<http://mpeg.chiariglione.org> を参考とした)

2020 年に向けて

2020 年のオリンピック観戦は, 次世代 MPEG 符号化でどう変わるだろう。MPEG-H を採用した 4K/8K スーパーハイビジョンの実験放送は 2015 年に開始される。2016 年のリオ・オリンピックは, 8K 映像と, 22.2 チャンネルのオーディオで観戦したい。本稿で紹介した第 3 世代 FTV や 3D オーディオも, 2020 年の東京オリンピックには技術的には実現可能と期待できる。

すでに標準化が完了した ISO/IEC 23008-1:MMT と ISO/IEC 23009-1:DASH の普及によって, 放送とネットワークの融合は格段に進むだろう。これらは会誌 2014 年 9 月号⁵⁾ と 10 月号⁶⁾ に, 詳しい記事が掲載された。

また, 現在進行中または今後開始が検討されている標準化項目を表-1 に示す。スマホ, タブレット, ウェアラブルなど新しい端末では, Green Metadata (省エネ技術) や Uniform Signaling for Timeline Alignment (同期技術) が応用できる。また, HMD (Head Mount Display), 3D ディスプレイ, 3D プリンタなど, 3D や拡張現実の普及が進んでいるが, Mixed and Augmented Reality Reference Mode (複合現実, 拡張現実の参照モデル) が応用できる。さらに Compact Description for Video Analysis (動画像の分析のためのコンパクトな特徴符号化), Multimedia Preservation Application Format (マルチメディア長期保存フォーマット)



はクラウド化やビッグデータでの利用に備えるものだ。MPEG では常に新しい応用や技術の登場に合わせて、あるいはその将来を見越して、新しい標準の準備を進めている。

2020 年のオリンピックまでにはこれらのうちのいくつかの標準も実現しているかもしれない。オリンピック競技の記録と同様、次世代 MPEG 符号化の開発も終わることがない。

参考文献

- 1) Sweet, W. : Chiariglione and the Birth of MPEG, IEEE Spectrum, Vol.34, No.9 (1997).
- 2) Shimizu, S., et al. : Use Cases and Requirements on Free-Viewpoint Television (FTV), ISO/IEC JTC 1/SC 29/WG 11 N14178 (2014).
- 3) Hamasaki, K., et al. : A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV), SMPTE Motion Imaging Journal, pp.40-49 (Apr. 2008).
- 4) ISO/IEC 14496-3:2009/Amd4:2013 New Levels for AAC Profiles (2013).
- 5) 青木秀一：MMT により実現される 4K/8K 放送システムメディアトランスポート技術の 20 年ぶりの大改定一，情報処理，

Vol.55, No.9, pp.996-1002 (Sep. 2014).

- 6) 平林光浩：次世代 HTTP ストリーミング標準 DASH，情報処理，Vol.55, No.10, pp.1138-1146 (Oct. 2014).
- 7) Kawamura, K., et al. : [FTV AHG] EE5 related, Preliminary Test Preparation with a Vide Sequence of a Single Camera for Free Navigation, ISO/IEC JTC 1/SC 29/WG 11 MPEG 2014/M33365 (2014).

(2014 年 6 月 16 日受付)

妹尾孝憲 senoh@nict.go.jp

情報通信研究機構，ユニバーサルコミュニケーション研究所，超臨場感映像研究室主任研究員，博士（工学），立体映像と電子ホログラフィの研究に従事。

金子 格（正会員） Itaru-k@acm.org

東京工芸大学工学部コンピュータ応用学科准教授，博士（工学），コンピュータの応用システム全般，ISO/IEC MPEG 標準化，デジタル音声処理。

安藤彰男 andio@eng.u-toyama.ac.jp

富山大学理工学研究部（工学）／工学部電気電子システム工学科教授，博士（工学），音響伝送・再生に関する研究に従事，2006 年度本会業績賞等受賞。