

MPEG-4 Version 1 FDIS の動向

渡辺 裕

SC29 専門委員会幹事, NTT

239-0847 横須賀市光の丘 1-1

NTT ヒューマンインタフェース研究所 画像通信研究部 922A

phone: 0468-59-2823, fax: 0468-59-2829, e-mail: hiroshi@nttvdt.hil.ntt.co.jp

オブジェクトベースのオーディオビジュアル符号化標準 MPEG-4 Version 1 は、第 45 回 MPEG Atlantic City 会合 (98/10/12-16) において FDIS (Final Draft International Standard) となった。ただし、テキストは編集作業の都合から 11/15 に FTP により MPEG サイトにアップロードされることになっている。作成された FDIS に対して、今後各国は Yes/No の投票を行い、IS (International Standard) となる。したがって、実質的な仕様は FDIS の時点で固められたことになり、以後変更は不可能である。MPEG-4 Version 1 の IS は、ISO/IEC 14496 シリーズとして発行される予定である。今回は、14496-1 (Systems), 14496-2 (Visual), 14496-3 (Audio), 14496-6 (DMIF) が FDIS となった。14496-4 (Reference Software) は 1999 年 3 月に FDIS 化される予定である。また、14496-5 (Conformance Testing) は 1 年後に FDIS 化されるスケジュールである。

Trends on MPEG-4 Version 1 FDIS

Hiroshi Watanabe

Japanese SC29 committee Secretariat, NTT

1-1 Hikarinooka, Yokosuka, 239-0847 JAPAN

Visual Communication Lab., NTT Human Interface Labs.

phone: 0468-59-2823, fax: 0468-59-2829, e-mail: hiroshi@nttvdt.hil.ntt.co.jp

The version 1 of MPEG-4, which is an object base audio-visual coding standard, is promoted to FDIS (Final Draft International Standard) at the 45th MPEG Atlantic City Meeting (98/10/12-16). However, the texts will be uploaded to MPEG FTP site on 11/15 due to the editing process. Each national body will vote "Yes/No" by ballot to the edited FDIS, and FDIS will be promoted to IS (International Standard). Thus, an actual specification of MPEG-4 is fixed at the FDIS stage so that it will not be modified. MPEG-4 version 1 will be formally published as ISO/IEC 14496 series. This time, 14496-1 (Systems), 14496-2 (Visual), 14496-3 (Audio) and 14496-6 (DMIF) are promoted to FDIS. 14496-4 (Reference Software) will be promoted to FDIS in March 1999. 14496-5 (Conformance Testing) will be promoted to FDIS a year later.

1 まえがき

MPEG-4 は、通常のビデオ、CG によるアニメーション、テキストや図形などのグラフィクス、オーディオ、音声、合成音などの情報をオブジェクトとして圧縮し、シーン記述情報に基づいてそれらを合成して出力する符号化標準である。

MPEG-4 では、MPEG-1 や MPEG-2 などと同様に、ビットストリームのシンタクスと復号の論理のみが規定される。様々なオブジェクトを組合せるために、多くの符号化ツールが用意されている。そのため、エンコーダのバリエーションは多岐にわたる。ビットストリームの互換点を確保し、デコーダのバリエーションをいくつかに集約するために、プロファイルとレベルという概念が使用されている。

本稿では、99/3 に IS (International Standard) となる MPEG-4 Version 1 の FDIS (Final Draft International Standard) 段階での動向について述べる。FDIS に対する投票は、"Yes/No" であり、仕様変更は受け付けられない。そのため実質的な最終仕様が出揃ったことになる。ここでは特に FCD (Final Committee Draft) からの変更点及びプロファイルについて述べる。

2 MPEG-4 Version 1 のプロファイル

2.1 プロファイルの全体像

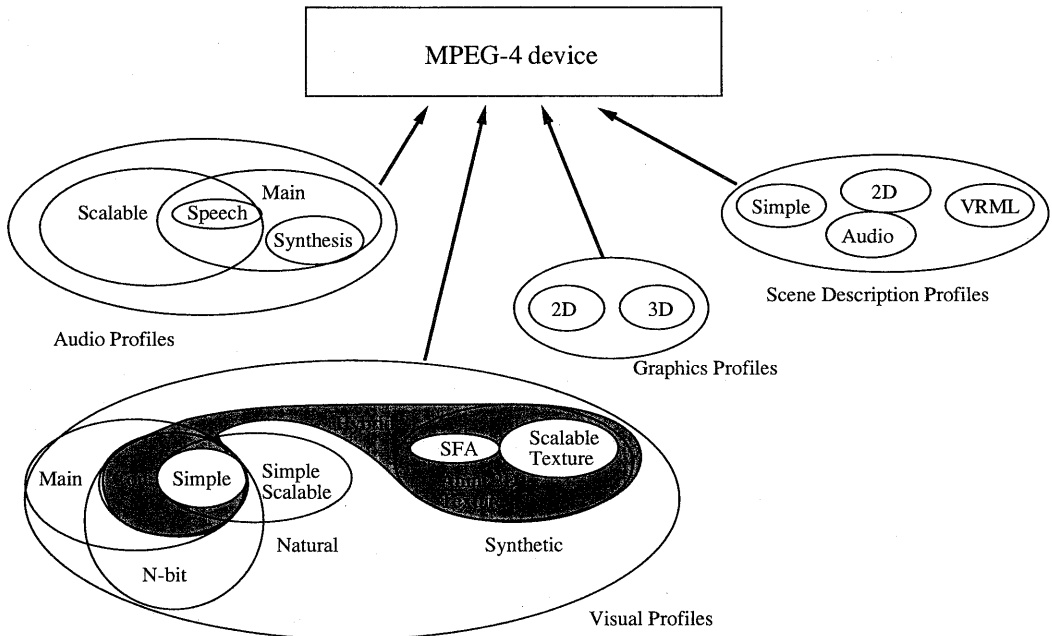


図 1: MPEG-4 Version 1 プロファイルの構成

Version-1 は、システム (Systems)、ビジュアル (Visual)、オーディオ (Audio)、適合性試験 (Conformance Testing)、規範ソフトウェア (Reference Software)、マルチメディア制御 (DMIF) の 6 部構成からなる。今回は、14496-1 (Systems)、14496-2 (Visual)、14496-3 (Audio)、14496-6 (DMIF) が FDIS となった。14496-4 (Reference Software) はソフトウェア実装の遅れから、1999 年 3 月に FDIS 化されることになった。また、14496-5 (Conformance Testing) はスケジュール通り、1 年後に FDIS 化される予定である。システム、ビジュアル、オーディオでは、複数のツール (技術) を集めてある機能を満たすことのできるプロファイル (仕様) とレ

レベル(規模を示すパラメータ)が定義される。これらはシステム、ビジュアル、オーディオでそれぞれに規定されるものであり、3つをまたがった端末レベルのプロファイルというものは存在しない。これらの選択は市場原理に任されている。プロファイルはビットストリームのシンタクスのサブセットを規定する。これらをまとめたものを図1に示す。システムにはグラフィクスプロファイル(Graphics Profile)とシーン記述プロファイル(Scene Description Profile)の二つが含まれる。

2.2 ビジュアルプロファイル

ビジュアルプロファイルがどのようなオブジェクトから構成されているかを表1に示す。レベルをどのように決定するかは現在まだ議論が行われている。MPEG-2では、あるレベルを規定するのに、パラメータは最大画素数或は単位時間あたりのマクロブロック数などが用いられている。MPEG-4でも Simple Profile ではその規定を踏襲している。しかし、Animationを含む Main Profile では、ソフトウェアでの実装を考慮して幾つかの要因の線形和で表した複雑さ指標により、レベルを規定しようとしている。ただし、動作チェックが完了していないため、現在では仕様への盛り込みを躊躇している状態である。

表 1: Visual Profiles

Profile	Object								
	Simple	Core	Main	Simple Scalable	N-bit	Animated 2D Mesh	Basic Animated 2D Texture	Scalable Texture	Simple Face
Simple	✓								
Simple Scalable	✓			✓					
Core	✓	✓							
Main	✓	✓	✓					✓	
N-bit	✓	✓			✓				
Hybrid	✓	✓				✓	✓	✓	✓
Basic Animated 2D Texture							✓	✓	✓
Scalable Texture								✓	
Simple FA									✓

自然画像のプロファイルには以下のものがある。

- シンプルビジュアルプロファイル (Simple Visual Profile)
エラー耐性のある矩形のビデオの符号化が可能であり、PCS や IMT2000 のような移動体通信のアプリケーションに適している。
- シンプルスケーラブルビジュアルプロファイル (Simple Scalable Visual Profile)
シンプルビジュアルプロファイルに時間 / 空間スケーラビリティを加えたもので複数の品質レベルを供給できるためインターネットやソフトウェアデコードのアプリケーション向きである。
- コアビジュアルプロファイル (Core Visual Profile)
シンプルビジュアルプロファイルに対して、任意形状を扱つことができ、また時間スケーラビリティを持つことから、簡単なオブジェクト単位のインタラクティブ性を活かしたマルチメディアアプリケーションに適する。
- メインビジュアルプロファイル (Main Visual Profile)
システムコアビジュアルプロファイルに加えて、インタレース画像処理、スプライト (背景画像)、アルファ

プレーンの機能を持ち、放送やDVDレベルのインタラクティブマルチメディアのアプリケーション用である。

- **N-ビットビジュアルプロファイル (N-bit Visual Profile)**

画素が4ビットから12ビットの精度の信号を処理し、遠隔監視のアプリケーションに用いられる。

合成画像あるいは自然画像と合成画像の混ざったハイブリッド画像のプロファイルには、以下のものがある。

- **シンプルフェイシャルアニメーションプロファイル (Simple Facial Animation Profile)**

顔のモデルを簡単に動作させる機能を持ち、耳の不自由な人へのAVプレゼンテーションなどのアプリケーションに適する。

- **スケーラブルテクスチャビジュアルプロファイル (Scalable Texture Visual Profile)**

多くのレベルからなる静止画の空間スケーラビリティを持ち、ゲームのアプリケーションなど物体にテクスチャを張り付ける場合に適する。

- **基本アニメ2次元テクスチャビジュアルプロファイル (Basic Animated 2D Texture Visual Profile)**

空間/画品質スケーラビリティと静止画が張り付けられた物体のメッシュを使ったアニメーションが可能であり、また顔画像のアニメーションも含まれている。

- **ハイブリッドビジュアルプロファイル (Hybrid Visual Profile)**

コアプロファイルまでの自然画像とアニメーション関係の全てのオブジェクトを持つ多機能なプロファイル。

2.3 オーディオプロファイル

オーディオプロファイルを表2に示す。FDISではプロファイルの名称が多少変更になり、オブジェクトのいくつかが除去された。FCDには含まれていたHLIN及びBSAC (Bit Slice Audio Coding) が除かれ、Version 2の検討課題となった。従来のAACはT/F Coderと合わせてGA (Generic Audio) と呼ばれることになった。CELPとHVXCは音声、HVLNはパラメトリック楽音、TTSIはテキスト-音声変換へのインタフェースである。

- **スピーチオーディオプロファイル (Speech Audio Profile)**

CELPとテキスト-音声変換 (Text-to-Speech) へのインタフェースによる超低レートの音声エンコーダ

- **シンセシスオーディオプロファイル (Synthesis Audio Profile)**

波形テーブル (Wavetable) とテキスト-音声変換 (Text-to-Speech) へのインタフェースを用い、低レートでの合成音声、合成サウンド用である。

- **スケーラブルオーディオプロファイル (Scalable Audio Profile)**

スピーチとオーディオを含み、スケーラビリティを持たせたものであり、インターネットや狭帯域デジタルオーディオ放送 (Narrow Band Audio Digital Broadcasting, NADIB) などのアプリケーションに適する。FCD段階ではシンセシスが含まれていたが、FDISでは除かれた。

- **メインオーディオプロファイル (Main Audio Profile)**

自然音響と合成音響、音声など全てを符号化する機能をもつプロファイル。

表 2: Audio Profiles

Profile	Hierarchy	Audio Object Type
Main	Contains Scalable, Speech Synthetic	GA Main, LC, SSR, Main/LC/SSR+PNS Twin VQ CELP HVXC Wavetable Synthetic TTSI
Scalable	Contains Speech	GA LC Scalable CELP HVXC Twin VQ
Speech		CELP HVXC TTSI
Synthesis		Wavetable Synthesis TTSI

2.4 システム

2.4.1 グラフィクスプロファイル

システムにはグラフィクスに関して2つ、シーン記述に関して5つのプロファイルがある。グラフィクスプロファイルはシーンの中で使われるグラフィクスやテキストを規定する。

- **2D グラフィクスプロファイル (2D Graphics Profile)**
様々な2次元グラフィクスを扱い、簡単な処理のみをカバーする。
- **コンプリートグラフィクスプロファイル (Complete Graphics Profile)**
2次元と3次元のグラフィクスをサポートする。

2.4.2 シーン記述プロファイル

FDISの段階で大きく変化した。VRMLに関しては、Proto/External Protoのノードを除く全てのVRML97のノードをサポートすることになった。また、IPMP (Intellectual Property Management and Protection)のためのインタフェースが追加された。従来のシステムの参照モデルであるIM1デコーダに加えIPMPデコーダが必要となった。そのためにStream Flow ControllerがIM1用に製作されている。FDISでは、Script node, Bitmap nodeが追加され、FCDに含まれていたVideoObject2D, Image2D, Composition Mapが除去された。またTTSからFAへのインタフェースが作成された。

シーン記述のプロファイルは、オブジェクトをノードとして関連付けて取り扱うMPEG-4において重要な機能を持っている。これらには以下のものがある。

- **シンプルプロファイル (Simple Profile)**
ビデオとオーディオが複数扱え、旧来のテレビ放送などのアプリケーションに適する。
- **2D プロファイル (2D Profile)**
2次元の変換やアルファ合成など簡単な2次元グラフィクスを実装するのに適する。
- **VRML プロファイル (VRML Profile)**
MPEG-4で取り扱うVRMLノードをシーン記述に用い、ゲームなど疑似的な3次元空間内の表示を使うアプリケーションに適する。

- **Audio プロファイル (Audio Profile)**
オーディオを使用するために用いる。
- **コンプリートプロファイル (Complete Profile)**
上記の全てを含むプロファイル。

FCD 段階でのプロファイルを元に修正したものを表 3 に示す。ただし、これは FDIS の編集段階では変更されている可能性があり、本原稿の執筆段階では正確ではない。

表 3: シーン記述プロファイル

Scene Description Profile	Content			
	Nodes	ROUTEs	BIFS Animation	BIFS Updates
Simple	Layer2D, Transform2D Sound2D, AudioSource			✓
2D	Shared, 2D, Layer2D Composite2DTexture	✓	✓	✓
VRML	Common VRML/MPEG-4 nodes	✓		✓
Audio	Audio, Sound	✓		✓
Complete	All	✓	✓	✓

2.4.3 DMIF

DMIF(Delivery Multimedia Integration Framework) はインタラクティブに MPEG-4 オブジェクトを操作する場合に必要なプロトコルなどを規定するものであり、プロファイルを持たない。

3 むすび

本稿では、MPEG-4 Version 1 の FDIS 段階でのプロファイルと動向について述べた。プロファイルを共にビットストリームの互換のポイントを規定する重要な項目にレベルがあるが、本原稿の執筆時点では、Visual の Main Profile での規定の仕方がまだ決定されていないという問題点がある。