

## MPEG-4 Version 1 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@ntttdt.hil.ntt.co.jp](mailto:hiroshi@ntttdt.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、東京会合 (98/3/16-20) の 2ヶ月後から FCD (Final Committee Draft) 投票期間に入っている。その間も Dublin 会合 (98/7/6-10) が開催され、FCD に対して指摘された問題点について議論が交わされ、FDIS (Final Draft International Standard) に反映される予定になっている。FDIS に成長すべき文書は Study 文書と名付けられ、将来の標準となるべくチェックが重ねられている。本稿では、98/10 に FCD となる Version 1 を中心に MPEG-4 の動向について述べる。

## Trends on MPEG-4 Version 1

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@ntttdt.hil.ntt.co.jp](mailto:hiroshi@ntttdt.hil.ntt.co.jp)

The version 1 of MPEG-4, which is an object base audio-visual coding standard, is progressing to the FCD (Final Committee Draft) balloting stage after two month later of the Tokyo meeting (98/3/16-20). During this balloting period, raised issues based on the FCD was been discussed at the Dublin meeting (98/7/6-10), so that discussion results will be reflected to FDIS (Final Draft International Standard). Documents which progress to FDIS are called "Study" and are carefully investigated for excluding bugs as International Standard. In this report, trends on MPEG-4 are described mainly focused on the version 1 which will be FCD in Oct. 1998.

## 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符号化し、シーン記述情報に基づいて合成する標準である。MPEG-4のタイトルは、従来“Very-low bitrate audio-visual coding”であったが、高いビットレートでも適用できることと、オブジェクトベースであることが最大の特徴であることを反映させ、“Coding of audio-visual objects”に変更された。低ビットレートでのアプリケーションだけでなく、HDTVの番組製作への適用が検討項目となっている。

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。現在、システム、ビジュアル、オーディオのプロファイルがほぼ固まった状態である。ビジュアルプロファイルのうち、Simpleはモバイルマルチメディアへのアプリケーションに使用される見通しである。

本稿では、98/10にFCD (Final Committee Draft) となる MPEG-4 Version 1 のプロファイルとアプリケーションおよびその動向について述べる。

## 2 MPEG-4 Version 1 のプロファイル

### 2.1 プロファイルの全体像

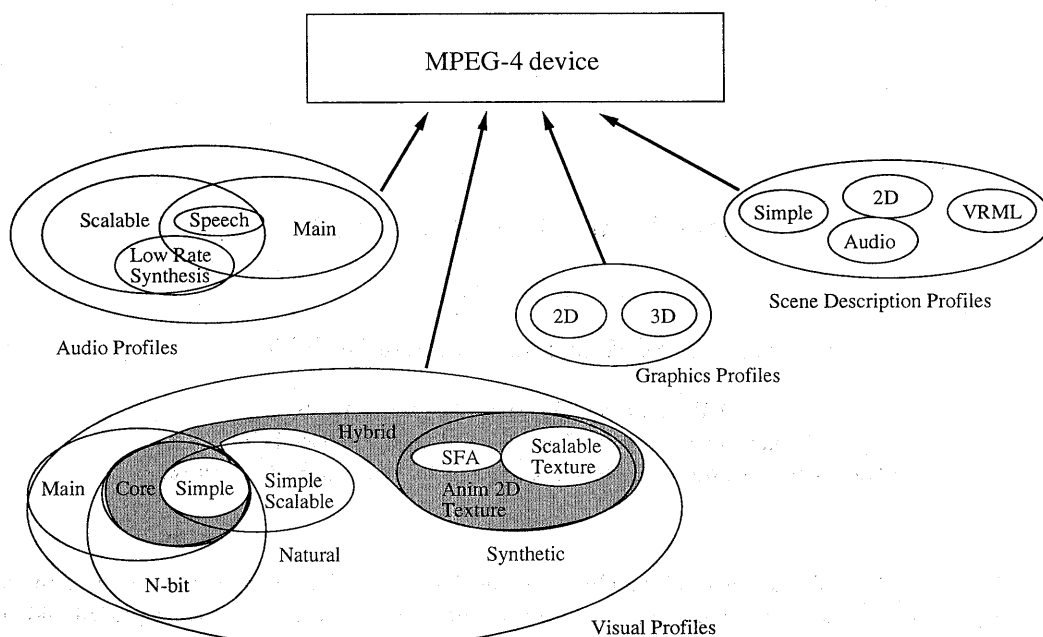


図 1: MPEG-4 Version 1 プロファイルの構成

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。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ッ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 (Scence Description Profile) の二つが含まれる。

東京会合 (98/3/16-20) までは、コンビネーションプロファイル (Combination Profile) とオブジェクトプロファイル (Object Profile) と呼ばれる2段階のプロファイルが定義されていた。しかし、ダブリン会合 (98/7/6-10) より、コンビネーションプロファイルは単にプロファイル (Profiles) に、オブジェクトプロファイルはオブジェクト (Objects) と呼ばれることになった(図2参照)。新しい定義に従うと、ビットストリーム互換の基準点に相当するプロファイルは複数のオブジェクトの組合せからなる。

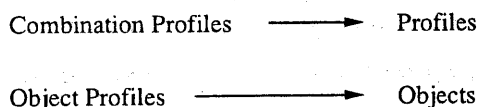


図 2: プロファイルの呼び方の変更

## 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。従来、“12-bit”と呼ばれていたオブジェクトはダブリン会合で“N-bit”に改められた。

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 Texute |        |      |      |                 |       |                  | ✓                         | ✓                | ✓           |
| 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 Pofile)  
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ダ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)**

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)、アルファブレンの機能を持ち、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。AAC, Twin VQ は楽音の符号化技術, 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。AAC も Twin VQ も時間信号の周波数成分への変換符号化である TF コーダ (Time Frequency Mapping Coder) の一種であり楽音の符号化に用いられる。AAC は MPEG-2 オーディオ (13818-7) と互換性を有している。

- **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

- **低レートシンセシスオーディオプロファイル (Low Rate 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)**

スピーチオーディオ、低レートシンセシスオーディオプロファイルを含み、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。

- **メインオーディオプロフィール (Main Audio Profile)**  
自然音響と合成音響、音声などを符号化する機能をもつプロフィール。

表 2: Audio Profiles

| Profile                   | Hierarchy                                       | Audio Object Type                                                                                                                                                                |
|---------------------------|-------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Main</b>               | Contains Scalable, Speech<br>Low Rate Synthetic | AAC Main, LC, SSR, Main/LC/SSR+PNS<br>T/F, T/F Main Scalable, T/F LC Scalable<br>Twin VQ Core<br>CELP<br>HVXC<br>HILN<br>Parametric<br>Main Synthetic<br>TTSI<br>Audio Generator |
| <b>Scalable</b>           | Contains Speech                                 | T/F LC Scalable<br>AAC-LC or/and T/F<br>CELP<br>HVXC<br>Twin VQ Core<br>HILN<br>Parametric<br>Wavetable Synthesis<br>TTSI<br>Audio Generator                                     |
| <b>Speech</b>             |                                                 | CELP<br>HVXC<br>TTSI<br>Audio Generator                                                                                                                                          |
| <b>Low Rate Synthesis</b> |                                                 | Wavetable Synthesis<br>TTSI<br>Audio Generator                                                                                                                                   |

## 2.4 システム

### 2.4.1 グラフィクスプロフィール

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つのプロフィールがある。グラフィクスプロフィー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ィール (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ィール (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 シーン記述プロフィール

シーン記述のプロフィールは、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次元グラフィッ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。

それらを表3に示す。シーン記述はBIFS(Binary Format for Scenes)により行われる。ノードの構造の一部にVRMLのサブセットを含む。2DやアニメーションやオーディオへのノードはMPEG-4固有のものである。

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<br>Sound2D, AudioSource<br>VideoObject2D |        |                | ✓            |
| 2D                        | Shared, 2D, Layer2D<br>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のプロファイルとアプリケーションおよびその動向について述べた。プロファイルと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。従来、ビデオのレベルはハードウェアによるデコーダの実現を意識して、メモリサイズや単位時間あたりの処理画素数などで規定されてきた。ところが今回、ソフトウェアによりデコーダを実現した場合の処理量の算出法が考案され、それに基づいたレベルの規定方法が考案された。このレベル規定は草案が作成された後、各国の投票によりメンバーに受け入れられれば、現在のレベル規定から変更される予定である。