Summary of the MPEG-4 Codec

Michael Murray

January 26, 2004

Imagine you download a video by your favorite band in MPEG-4 format, but instead of sitting idly by, listening and watching the video you have more options. You can switch between two different versions of the video by clicking one button, or perhaps you would like to use the foreground of one version with the background of the other. You could do the same with the audio track. Actually you want every word the lead singer says to come straight from her mouth as she is singing. After awhile you realize you would rather the lead singer be a man, so you change their face for the remainder of the video to that of Dan Rather. This all might seem far fetched, but it is this kind of interactivity, integration and control that MPEG-4 is designed to handle.

1 Introduction

According to the International Standards Organization (ISO) and the Motion Picture Experts Group MPEG-4 "builds on the proven success of digital television, interactive graphics and interactive multimedia" [4]. The specification of the standard calls for a complete integration of all media that a user can access or influence over the world wide web, which can be divided into two types of elements, visual elements and audio elements . This summary reveals the advancements made to both of these elements under the MPEG-4 specification.

It is important to visualize an MPEG-4 file as madeup of many different streams, some holding visual elements, some holding audio elements and some holding instructions on how to manipulate the end user software and the elements mentioned above. The MPEG-4 file structure was based on Apple's QuickTime movie file structure.

2 Visual Elements

In MPEG-4 the nature of visual elements can be summarized by the following specifications and features:

Bit-rates : Since MPEG-4 files are designed for streaming use as well as native use, they include information on the bit-rate necessary for realtime streaming. Compression algorithms allow MPEG-4 to be stream-able in bit rates from 5kb/sec -¿ 1gb/sec and beyond.

Formats : Both video and graphics are similar to other current pixel based formats(eg. jpeg) and many known pixel based formats can be imported as a stream in an MPEG-4 file. Video can be either of the progressive or interleaved variety.

Resolutions : Both graphics and video can be represented in resolutions of sub-QCIF to studio level resolutions (4000 x 4000 pixels).

Compression : Video and Graphic compression is said to range from " "acceptable" for very high compression ratios up to "near loss-less"" [4].

Control : User has random access to any point in a video or animation (2D or 3D) stream as well as all standard controls: play, rewind, fast forward, stop, pause.

Coding : There are many visual coding features which apply to different visual elements within an MPEG-4 file:

- Content-Based Coding: this means that each individual object that makes up a scene can be accessed separately over time. This feature also allows each object to change over time, including any kind of size or shape morphing or layer-ability.
- Shape Coding: Each of the above objects does not have to be a standard rectangle shape but can be of any outlined shape.
- 'Gray Scale' Coding: Refers to the ability to assign and realize different levels of transparency for any object.
- 2D and 3D Coding: All visual objects will be represented by 2D or 3D meshes which will take on the characteristics outlined in the other coding methods.

Scalability : Because of coding methods MPEG-4 visual objects are scalable in a number of ways.

- Complexity Scalability: Each visual element in an MPEG-4 file can be encoded into various complexities (reolution/compression etc.) and decoded from these same levels.
- Spatial Scalability: Each visual item can be scaled in size within the scene over time.
- Temporal Scalability: Every video or animation can be slowed down or sped up to one of three relative levels.

Animation : Both 2D and 3D animation of any mesh is possible through the variance of shape, location and layer. There is specific tools available in MPEG-4 for face and body animation.

Recognition : There are tools available in the MPEG-4 specification for shape and movement recognition, especially for face and body objects.

All animation and recognition features are currently being upgraded through the Animated Framework Extension. This one of many extensions still being worked on for the specification.

3 Audio Elements

MPEG-4 Audio is a very versatile tool which combines existing compression encoding (eg. MPEG-2 AAC) with MPEG-4 specific tools (eg. Error Correction tool (ERtool)). We will review all the basic features of MPEG-4 Audio and take a closer look at MPEG-2 AAC compression and how it has changed in the MPEG-4 environment.

Bitrates : MPEG-4 Audio can be transferred in bitrates of 2kbits/sec to 64kbits/sec and up for each channel of audio, all classified on different streams.

Channels: Up to 48 channels.

Sampling Rates: 2kHz to 48kHz.

Compression: MPEG-4 Audio streams can each be encoded using any number of different compression codecs. Usually speech is encoded with different compression at lower bitrates then general audio.

The speech codecs include:

- HVXC (Harmonic Vector Exchange Coding), for very low bitrates from 2kbits/sec to 4kbits/sec. Proven to be functional using VBR(Variable Bit Rate) at 1.2kbits/sec.
- CELP (Code Excited Linear Predictive), for bit rates from 6kbits/sec to 18kbits/sec.

The general audio codecs include:

- TwinVQ
- MPEG-2 AAC: This is the main compression technique used in the MPEG-4 audio specification which is a perceptual audio codec based on the popular MPEG-1/layer 3(A.K.A. mp3). Some improvements have been made to AAC audio under the MPEG-4 specification and besides improvements to the AAC codec itself, tools have been developed under the MPEG-4 specification to enhance the use of AAC audio. These are some of the enhancements and tools applicable to MPEG-2 AAC audio under the MPEG-4 specification:
  - TNS(Temporal Noise Shaping): This is a technique to more effectively mask quantization noise created by compression. Jurgen Herre from Fraunhofer summarizes the technique: "If such predictive coding is applied to spectral data over frequency, the temporal shape of the quantization error signal will appear adapted to the temporal shape of the input signal at the output of the decoder." [3]
  - Low-Delay Transmission: MPEG-4 uses a version of AAC using little or no process buffer and different windowing techniques to minimize the delay time inherit perceptual coding. This low-delay allows for real-time application of the codec.
  - Prediction: "A technique commonly established in the area of speech coding systems. It benefits from the fact that certain types of audio signals are easy to predict." [1] This allows the MPEG-4 format to more effectively prevent pre-transient artifacts found in perceptual coding.

  When the EBU(European Broadcasting Union) tested MPEG-4 audio it said there was an imperceptible difference in the original and the compressed audio under the following conditions:
  - 5 x 64kbits/sec per channel in five channel reproduction. Total 320kbits/sec.

  - 2 x 128kbits/sec per channel in stereo reproduction. Total 256kbits/sec.

Scalability:

- Bit-rate: Bit-rate scalability refers to MPEG-4's ability to have streams chopped up and reassembled either in transmission or decoding. This allows many chunks of one stream to be sent over many different transmission lines simultaneously, allowing greater download speeds.
- Complexity: Complexity scalability allows different streams to be of different bitrates and different compression types. This allows for example the base stream to be a 2kbit/sec HVXC speech stream enhanced by a 24kbit/sec AAC stream carrying optional reverberant information.

Synthesized Sound:

- Structured Audio: The MPEG-4 specification includes a unique way of using synthesized audio called Structured Audio. This part of the codec was designed by researchers in MIT's MAchine listening Group at MIT Media Laboratory. Structured AUdio uses instructions on part and score to determine not only what an end user's synthesizer should play(notes) but also exactly the type of sound it should be played with(beyond midi instruments). The MIT Machine Listening Group webpage describes the first implementation of Structured Audio to be included in MPEG-4, "we decided to use the opportunity(*of a redesign phase encouraged by MPEG*) to revisit some of the synthesis-language issues represented in Csound, and dive deeper into the Structured Audio concepts than NetSound did. The result was the language SAOL, which we designed over the winter of 1996-1997 and submitted to MPEG soon after, meeting with general enthusiasm." [2]
- TTS: Text To Speech: From Rob Koenen's description of the MPEG-4 specification, "Text To Speech. TTS coders bit-rates range from 200 bit/s to 1.2 Kbit/s, which allows a text or a text with prosodic parameters (pitch contour, phoneme duration, and so on) as its inputs to generate intelligible synthetic speech. It supports the generation of parameters that can be used to allow synchronization to associated face animation, international

languages for text and international symbols for phonemes. Additional markups are used to convey control information within texts, which is forwarded to other components in synchronization with the synthesized text. Note that MPEG-4 provides a standardized interface for the operation of a Text To Speech coder (TTSI = Text To Speech Interface), but not a normative TTS synthesizer itself." [4]

4 Conclusion

This summary only partly reveals some of the many features designed under the MPEG-4 specification, which is obviously a very dynamic and powerful format but one that is also which is a relatively new. The first complete version of the codec was released in 1998 and extensions are still being worked on today. Many of the features in the codec do not even have practical implementations as of yet, but it is exciting to see where this format could bring us in the future. To keep an eye on the codec itself navigate your web browser to www.m4if.org, the MPEG-4 user's forum.

References

[1] Mpeg-2 advanced audio coding: Data compression for the 21st century. *Fraunhofer Institut Integrierte Shaltungen (website).*

[2] Mpeg-4 structured audio: Saol, sasbf, sasl, audio bifs, and more. *Machine Group: MIT Media Laboratory (website).*

[3] Jurgen Herre. Temporal noise shaping, quantization and coding methods in perceptual audio coding. *Proceedings of the AES 17th International conference on High Quality Audio Coding*, 1999.

[4] Rob Koenen. Overview of the mpeg-4 standard. *ISO/IEC*, WG11, 2002.