Mpeg 4

general overview
9.4
MPEG History

- MPEG is a subcommittee of the ISO
  - Moving Pictures Experts Group
- MPEG-1: 1993 (ISO 11172)
  Digital audio/video coding (MP3)
- MPEG-2: 1994-7 (ISO 13818)
  Digital coding for broadcast
- MPEG-4: 1998 (ISO 14496)
  - Low-bitrate coding
  - Interactive, synthetic, distributed, object-based coding
Goals of MPEG-4

- Represent scenes in terms of objects
- Enable reusability for authoring
- Facilitate interactive content
- Access scalable transport (bitrate, QoS)
- Allow sophisticated IPR management
- Organize many technologies into unified framework
Mpeg 4

Standardized ways to support:
- Coding
- Composition
- Multiplex
- Interaction

The parts of MPEG-4
- Systems (streaming and compositing)
- Video (including synthetic objects)
- Audio (including sound synthesis)
- Conformance
- Reference software
- DMIF (delivery API)
MPEG-4: Object-based coding

MPEG-4 integrates...
- Audio and video
- Basic media and media compositing
- Streaming and downloaded content
- Natural and synthetic video
- Video with text and graphics
- 2D and 3D coding
- Still images and videos
- Recorded sound and synthetic sound

... in an open-standards framework
MPEG-4: The big picture
A note on encoding

- **Traditional coding philosophy**
  - Encoding is a **hands-off**, automated process that is a black box to the author

- **MPEG-4 philosophy**
  - By **involving** the content author in the encoding process, more **sophisticated** encoding is possible

- **Implication**
  - MPEG-4 aware content-creation tools are **crucial**!
Application Scenarios of MPEG-4

- Live Content
- On-demand Content
- Stored Content

Media Encoder

License Server

Media Services Server

Download & Play Streaming

Wired & Wireless

Media Player

PC, Hand-held, STB

Compression → Access → Interaction
Overview of MPEG-4

- The coded representation of the combination of streamed elementary audiovisual information
- 1) Compression, 2) content-based interactivity, 3) universal access
- To provide a bridge between the www and conventional AV media
- To delivery streaming AV media on the Internet and wireless networks

```
<table>
<thead>
<tr>
<th>Audiovisual Scene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded Representation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natural and Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Information</td>
</tr>
<tr>
<td>Coded Representation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natural and Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Information</td>
</tr>
<tr>
<td>Coded Representation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synchronization</th>
</tr>
</thead>
<tbody>
<tr>
<td>of Audiovisual Information</td>
</tr>
</tbody>
</table>
```
Integration of Natural and Synthetic Contents
Object-Based Coding

- Object-Based Coding = Content-Based Coding
  - increases compression efficiency
  - allows the user to access arbitrarily-shaped objects in a coded scene
  - enables high interaction with scene content

- Manipulation of scene content on bit stream level
Objects in Audio-Visual Scene

- Speech
- Video Object
- 2D Background
- AV Presentation
- 3D Furniture
BIFS – Binary Format for Scene

- Person
  - Speech
  - Video
- 2D Background
- Furniture
  - Globe
  - Table
- Audio-visual Presentation
Object-Based Coding

Each Video Object in a Scene is Coded and Transmitted Separately
Scene Reconstruction
Example of Video Decoding

Bitstream

DEMULTIPLEXE

Shape Decoding

Shape Information

Motion Decoding

Motion Compensation

Texture Decoding

VOP Memory

Reconstructed VOP

Compositor

Video Out

Compositing Script in BIFS
Composition

- Composition information
- Spatial relationships
  - Each elementary object has its coordinate system
- Temporal relationships
  - time base associated to BIFS;
  - time stamps associated to media streams
  - time value: duration or instant in time (relative to time stamps)
Multiplex layers

- Synchronization layer:
  - Specific synchronization information
- Flexible multiplex layer:
  - multiplexing streams with different characteristics (e.g., bit rate)
- Transport multiplex layer:
  - Adapting to particular network characteristics
Video

Functionalities

- Content based interactivity
  - Access and manipulation tools
  - Hybrid natural and synthetic data

- Compression
  - Images and video
  - Textures
  - Implicit 2D meshes
  - Time varying geometry that animate meshes

- Universal access:
  - Robustness
  - Scalability
    - Content based scalability
    - Spatial, temporal, quality scalability
Possible Applications of MPEG-4

- IVS  Internet Video Streaming
- VA   Video Archive
- VCD Video Content Distribution
- IMM Internet Multimedia
- IVG Interactive Video Games
- IPC Interpersonal Communications (videoconferencing, videophone, etc.)
- ISM Interactive Storage Media (optical disks, etc.)
- MMM: Multimedia Mailing
- NDB: Networked Database Services (via ATM, etc.)
- WMM: Wireless Multimedia
MPEG 4

Audio
Objects
- Natural audio
- Synthetic audio
- Control

Operations on objects
- Synchronize
- Decode
- Compose into compound objects
- Present
- Interact
Advantages of Object Framework

- Each signal coded with most efficient coding system
  - Natural
  - Synthetic
- Composition of objects into audio scene
  - Rate conversion
  - Mix and Equalize
  - Effects
- Final mix is done in the terminal
Audio Object Functionalities

- Signal compression
- Scalability
  - bit rate
  - signal bandwidth
  - presentation rate
  - encoder or decoder complexity
- Extraction and re-use
- Robustness to channel errors
System overview
MPEG4 audio components

- Contains a set of different coders for different signal classes and bit rate ranges
  - Parametric Speech and audio coders
  - CELP speech coders
  - General audio (G/A) coders
  - Combined scalable audio coders
- Methods for synthetic material
  - Structured Audio
  - Text To Speech interface
- Sound effect and composition
  - Speech change tool
  - Audio compositor
  - Sound effects
Scalability

Satellite  Cellular phone  Internet  ISDN
Secure com.

24 bit-rate (kbps)  32  48  64

Scalable Coder

TTS

Speech coding

General audio coding

4 kHz  8 kHz  Typical Audio bandwidth  20 kHz
MPEG-4 Audio

- **High-quality** sound
  - Based on MPEG-AAC algorithm: twice as good as MP3

- **Low-bitrate** sound
  - For WWW and cellular: speech/music as low as 4 kbps

- **Synthetic** sound
  - Interface to Text-to-Speech (TTS) synthesizers
  - High-quality audio synthesis with Structured Audio

- **AudioBIFS**
  - Mix and postproduce multi-track sound streams
Application Domains: Profiles

- Speech
  - low rate speech coders and TTS

- Synthesis
  - wavetable synthesis
  - score driven synthesis
  - TTS

- Scalable
  - speech coders
  - general audio coders
  - all coders in scalable configuration

- Main
  - all of the above
Mpeg 4: synthetic audio

- Mpeg-4 audio
  - Structured Audio (SA)
  - Text-To-Speech Interface (TTSI)
- Mpeg-4 systems
  - Audio scene description (AudioBIFS)
  - 3D models of sound source (VRML)
Structured audio in Mpeg-4

- Structured Audio = sound coding by structural description
- decoding: synthesis of music and audio effects
- **SAOL:** Structured Audio Orchestra Language
  - description of synthesis algorithm
- **SASBF:** Structured Audio Sample Bank Format
  - efficient description of wavetable banks
  - for terminal of limited functionality (e.g. karaoke)
- **SASL:** Structured Audio Score Language
  - description of control parameters
    - interface with SAOL
    - MIDI
Text-To-Speech in MPEG-4

- phonetic syntax (independent from language)
- prosody, timing
- language, dialect, gender, age
- automatic synchronization with face movement (FBA)
- specified only the interface, not the method TTS
Audio scene description in Mpeg-4

- different parts of audio track are coded in different way
- production, mixing, spatialization are done at the terminal
- **AudioBIFS**: description of mixing, effects and postproduction
  - mix and synchronize multichannel sounds
  - uses SAOL for effects description (AudioFX)
  - position sounds in virtual 3D space
- sound reflection and transmission through objects
  - parameters (reflection and transmission) associated to polygonal (visual) surfaces
  - detailed model of reflections echoes, delays and directionality when they arrive to listener
  - allows inclusion of absorbing objects in BIFS scene