MPEG-H Audio System for Broadcasting

ITU-R Workshop “Topics on the Future of Audio in Broadcasting”

Jan Plogsties
Challenges of a Changing Landscape

**Immersion**
Compelling sound experience through sound that comes from all directions produced in different formats

**Personalization and interactivity**
Adaptation to the user’s liking and listening situation

**Everywhere**
Play on any device with delivering best possible sound experience in terms of spatial impression and dynamic range
What is MPEG?

- Moving Picture Experts Group (MPEG) is a working group of ISO/IEC

- MPEG has a 25 year history of delivering open standards for broadcast
  - like MPEG-2 Video and transport,
  - MPEG-4 AAC familiy of audio codecs,
  - MPEG-4 AVC or MPEG-H HEVC

- In 2013 a Call for Proposal was issues for technology for coding of immersive and interactive audio
  - Extensive testing and refinement over several meeting periods
- MPEG-H 3D Audio is International Standard since 2015
MPEG-H Audio in the broadcasting chain

Audio from Post-Production
With Control Data for Objects and Loudness

Audio from Live Console
No Control Data for Objects and Loudness

Advanced sound file format

A/V Playout Server
SDI (A/V)

Immersive Audio Monitoring & Authoring
Joy Stick
Monitor Mode
Integrated Loudness
Object Setting (Interactivity, Loudness, ...)

E.g. 7.1+4 Speakers

MPEG-H Audio Emission Encoder
Transport Stream
Adaptive Streaming Segments
Web Server

MPEG-H Audio Decoder incl. Rendering
User control
IP

Tablet
2.0, 5.1, 7.1+4 Loudspeakers, Headphones

TV
Audio Content Types

- **Channel-based**
  Mixes produced for a defined target loudspeaker locations

- **Object-based**
  Sound of a single element and related metadata
  - **Positional object** - Target locations or arbitrary movements specified by metadata
  - **Interactive object** – Object that can be controlled by the listener

- **Higher Order Ambisonics (HOA)**
  Corresponds to a spherical expansion of the sound field in a point

**Speaker layout agnostic; Rendering to target loudspeaker setup**
MPEG-H Audio - Basic Decoder Architecture
MPEG-H Audio Core Codec

- Based on MPEG Unified Speech and Audio Coding (USAC)
- Extensions for use in the context of 3D audio
- Improved coding efficiency by parametric tools
  - Improved stereo and multichannel imaging
  - Instantaneous bit rate and stream switching
  - Signaling for 3D content/loudspeaker layouts

<table>
<thead>
<tr>
<th>Bed</th>
<th>Objects</th>
<th>Total # of chan. + obj.</th>
<th>Total Bitrate (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1+4</td>
<td>3</td>
<td>15</td>
<td>504</td>
</tr>
<tr>
<td>5.1</td>
<td>3</td>
<td>9</td>
<td>288</td>
</tr>
<tr>
<td>2.0</td>
<td>3</td>
<td>5</td>
<td>180</td>
</tr>
</tbody>
</table>
Format Converter / Channel Rendering

- Flexible Rendering to target loudspeaker layout
- Retains immersion as good as possible with available speakers; including height perception

- Intelligent downmix
  - Compensation of timbre coloration / signal cancellation effects
  - Automatic generation of optimized downmix matrices broadcast-controlled downmix
Object Renderer – Positional Objects

- Object channel is played over a combination of available loudspeakers that are closest to the intended object position
- Dynamic Metadata is used to control object position
- Uses virtual loudspeakers in case too few loudspeakers are available
Object Renderer – Interactive Objects

- Object = audio essence + metadata
- Interaction during playback:
  - Selection on/off
  - Level adjustment
  - Position

  fully controlled and restricted by metadata

- Presets are a combination of enabled and disabled groups, and a specific mix

![Diagram showing Object Renderer, Object Meta-data, and User Interactivity connections.](image)
Higher Order Ambisonics Renderer

- Dedicated coding scheme and renderer for HOA content
- Loudspeaker-independent, universal, representation of a sound field
- Multi-capsule microphone array
- Synthesize HOA signals
- Directly rendered to target loudspeaker layout
Dynamic Range Control

Source: EBU tech 3343

Receiver type: AV Receiver, TV Set, Tablet

Watching TV late at night

0 dB FS
Loudness Normalization

For different programs

- No normalization
  - Commercial: 7.1+4
  - Sports: 7.1+4
  - Film: 2.0
- With normalization
  - Commercial: 7.1+4
  - Sports: 2.0
  - Film: 2.0

For different playback

- No normalization
  - TV Mix: 7.1+4
  - Dialog+: 2.0
- With normalization
  - TV Mix: 7.1+4
  - Dialog+: 2.0

For different presets

- No normalization
  - TV mix: 7.1+4
  - Dialog+: 2.0
- With normalization
  - TV mix: 7.1+4
  - Dialog+: 2.0
Highly Efficient Core Audio Codec

- Object based option with up to 128 objects
- Sound-field-based
- Channel based option with up to 128 channels

Enhanced Loudness and Dynamic Range Control
Flexible Rendering
Synchro-nization-ready for second screen audio

Delivery over Broadcast, Internet, and Mobile Networks

MPEG-H 3D Audio

- Fully Open Standard
- Reference software
- Multi-vendor support
MPEG-H AUDIO ALLIANCE
The Next-Generation System for Interactive and Immersive Sound

Introducing MPEG-H Audio:

With consumers viewing - and hearing - content on more devices and in more places than ever before, broadcasters are increasingly pressured to meet demands for new and better experiences in a cost-effective manner. Consequently, broadcasters are assessing the capabilities of existing infrastructures and determining how new developments in audio and video technology will affect their ability to deliver enhanced services to a broad array of end-user technologies - from high-end home theaters, to tablets and smart phones. The new MPEG-H Audio standard allows broadcasters to meet these challenges beyond 5.1 surround sound while incorporating new groundbreaking interactive features in a rapid and cost-effective manner.
MPEG-H Audio in the broadcasting chain

**Production**
- Audio from Post-Production
  - With Control Data for Objects and Loudness

**Live Production**
- Audio from Live Console
  - No Control Data for Objects and Loudness

**Broadcast Control**
- Immersive Audio Monitoring & Authoring
  - Joy Stick
  - Monitor Mode
  - Integrated Loudness
  - Object Setting (Interactivity, Loudness, ...)

**Encoding**
- MPEG-H Audio Emission Encoder
  - Transport Stream
  - Adaptive Streaming Segments
  - Web Server

**Consumer**
- MPEG-H Audio Emission Decoder incl. Rendering
  - User control
  - IP
  - TV
  - 2.0, 5.1, or 7.1+4 Speakers

---

**ITU-R BS.2051-0**
Advanced sound system for programme production

**ITU-R BS.2076**
Audio Def. Model

**ITU-R BS.1352**
BWF extension?
(Profiles for typical use)

**ITU-R BS.1770**
Loudness (for immersive/objects content?)

---

© 2015 Fraunhofer IIS
Thank you.

Jan.Plogsties@iis.fraunhofer.de