Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement

Robert Bleidt – Fraunhofer

Based on the new MPEG-H Audio Open International Standard: ISO/IEC23008-3
Introducing the MPEG-H Audio Alliance’s New Interactive and Immersive TV Audio System
MPEG-H Audio improves the listening experience

• The newest audio codec from MPEG expands the features of today’s codecs:

• Personalize the sound to what you want to hear
• Make surround sound more realistic
• Tailor reproduction to your listening environment and hardware

• And the technical stuff:
  – One stream to every device
  – Combine multiple streams for hybrid delivery
  – Unsurpassed coding efficiency
MPEG-H Audio improves the listening experience

Personalize the sound to what you want to hear
MPEG-H Audio improves the listening experience
Personalize the sound to what you want to hear

• In addition to traditional channel-based audio, MPEG-H includes audio objects, sound elements that are transmitted separately to the decoder
• This allows the listener to adjust the elements to his liking
• Examples:
  – Simple case: Adjust the dialogue or commentary level in relation to the other audio elements such as ambient sound, music, or sound effects
    • Precursor 2011 Wimbledon “Grunt Controller” trials with BBC showed most listeners preferred either a higher or lower dialog level
    • Helpful for listeners with hearing loss or non-primary language or dialect
    • Additional language support becomes a simple matter of an additional 20-40 kb/s object
First practical use of objects: Dialogue Enhancement
Allows viewers to set their own Dialogue levels

- User benefit
  - Dialogue/Announcer can be adjusted up or down
  - Helps hearing impaired understand Dialogue
  - Helps for listening in noisy environment
- Broadcaster benefit
  - One signal broadcast to all viewers
  - Default mix played by existing receivers
- Public test by BBC during the Tennis Grand Slam Championships 2011 in Wimbledon
- Now undergoing standardization in DVB
MPEG-H Audio improves the listening experience
Personalize the sound to what you want to hear

- More complex examples:
  - 2014 Field Test with leading U.S. sports network at winter sports event
    - Pick your announcer and language
    - Want to hear more sound effects? More audience sounds?
  - Demo content from NASCAR race
    - Pick your announcer and language
    - Mix in driver’s radio
    - Hear car sounds/transducers
- Real-time decoding and rendering
- User selects audio mix from on-screen display sliders or **preset buttons**
MPEG-H Audio improves the listening experience
Personalize the sound to what you want to hear

• Summary: Personalization
• Audio objects allow listeners to:
  – Adjust elements in the sound mix to their liking
  – Boost dialogue for hearing-impaired or second language listeners
  – Listen to alternate audio content such as:
    • Home vs away team commentary
    • Driver to crew race audio
    • Alternate languages, simultaneous translation, or audio description
    • Many other future creative and revenue opportunities
MPEG-H Audio improves the listening experience
Contrasting Objects: Cinema vs Broadcast

• Cinema Theater:
• Broad listening area
• Listening primarily in reverberant field
• Objects used for spatial accuracy and dynamic motion overhead
• Standardized “house curve” and loudspeaker arrangement
• Good practices to study for TV sound in the 5.1 era

• Broadcast to Consumer’s Living Room:
• Sofa/recliner in middle of room
• Closer to near-field at most frequencies
• Objects used for personalisation / interactivity
• Dynamic objects may be used, but fewer due to transmission bandwidth
• Wide range of receiving devices: home theater, 2.1 sound bar, TV speakers, tablet speakers, earbuds
• Production concepts and sound design diverge in the 21st Century
MPEG-H Audio improves the listening experience

Making surround sound more realistic
MPEG-H Audio improves the listening experience
Making surround sound more realistic

- 3D audio systems add height information (more channels or audio objects) to the sound field to improve realism
- It becomes difficult for the listener to perceive he is hearing a recording
- Our studies show adding **four height speakers** provides a substantial improvement without the 22 speakers of classical 3D systems
MPEG-H Audio improves the listening experience  
Making surround sound more realistic

- MPEG-H currently has defined speaker configurations from 1.0 (mono) to 22.2  
  - Bitstream support for up to 128 channels and 128 objects
- **Our recommendation for immersive sound:**  
  - 7.1 Blu-ray surround configuration  
  - 4 Height Speakers above corner speakers as shown
- Other configurations are possible, for example:  
  - 5.1 + 4H  
  - Using front wide 7.1 configuration or 9.1
MPEG-H Audio improves the listening experience
Making surround sound more realistic

• MPEG-H includes sound field capture technology for alternative means of capturing and reproducing 3D sound
• Based on Higher-Order Ambisonics techniques
• Transmits predominant sounds and ambience in separate tracks, independent of speaker format
• Allows one-microphone capture of sound field or ambience
• As with channel-based system, may be combined with audio objects for interactivity or personalization
Recommended Vertical Viewing Angles Increase with 4K

<table>
<thead>
<tr>
<th>System</th>
<th>1920 x 1080</th>
<th>3840 x 2160</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing distance (relative to picture height)</td>
<td>3.2</td>
<td>1.6</td>
</tr>
<tr>
<td>Viewing angle - horizontal (degrees)</td>
<td>31</td>
<td>58</td>
</tr>
<tr>
<td>Viewing angle - vertical (degrees)</td>
<td>18</td>
<td><strong>35</strong></td>
</tr>
</tbody>
</table>

- 1.6 PH viewing is recommended for 4K (BT.2022)
- (though consumers may be able to initially afford/accept only 2 PH)
- 2 PH ≈ screen diagonal, 65” away from 65” TV
- Sound/picture divergence above 15º becomes noticeable/objectionable?
“The average consumer will never hear 3D, they listen to stereo today”

• It is true today’s consumers listen mainly on stereo in-TV speakers or with simple 2.0 or 2.1 soundbars

• 5.1 or 7.1 AVR + speaker systems aren’t seen as the logical match for that new sleek, slim flat-panel TV

• So why do 3D audio if most of the audience won’t hear it?
  – The same reason we do surround today – the high-value viewers will have 3D in their home theaters
  – Right?
Fraunhofer demonstrates future 3D listening for mainstream consumers with prototype “3D Sound Bar”

- Concept demonstration of our vision of mass consumer delivery of 3D audio
- Décor-friendly – no external wires or speakers
- No confusing AVR setup menus – un-box, hang on wall, enjoy
- Could be integrated into TV or offered as a separate frame
- Dramatic improvement over today’s stereo soundbars
- Home theater enthusiasts will still want 11.1 speakers (or more) for ultimate sound quality
- Shown at the Fraunhofer IIS booth at NAB 2014, IBC 2014, AES 2014
MPEG-H Audio improves the listening experience
Making surround sound more realistic

- Summary: Immersive Audio
- Immersive audio provides a substantial increase in sound quality beyond surround sound
- Four height speakers provide a good tradeoff between cost and sound quality
- “3D sound bar” technology could be built into TVs or soundbars to enable a dramatic improvement in sound quality and immersive sound for the mainstream consumer
- Higher Order Ambisonics offers an alternative method of capturing and transmitting 3D sound
MPEG-H Audio improves the listening experience

The Technical Stuff
A brief introduction to some selected MPEG-H concepts and features
MPEG-H Audio improves the listening experience

Tailor reproduction to your listening environment and hardware

- Multi-platform loudness control
  - Loudness measurement built into the encoder
    - Or use external values
  - Uses existing AC-3 metadata in legacy mode
  - Adjustment of dynamic range to match listening environment and device capabilities
    - Examples:
      - Home listening on good speakers
      - Listening on earbuds at airport on mobile phone
      - Listening at home on tablet
MPEG-H Audio improves the listening experience
Tailor reproduction to your listening environment and hardware

• Speaker-foolproof rendering for legacy speakers
  – May be able to correct for misplaced speakers in consumer’s home
    • HDMI 2.0 will allow communication of consumer’s installed speaker locations
  – May offer a limited impression of height on existing 5.1 or 2.0 consumer speakers
• Binaural rendering for headphone reproduction
  – Listener perceives sound stage in front of him, instead of between his ears
  – May be extended to tablet speakers with Fraunhofer Cingo technology
MPEG-H offers the best coding efficiency available today

- Includes a highly efficient new audio codec to enable true 3D Audio transmission in broadcasting and streaming:

<table>
<thead>
<tr>
<th>Bitrates in kb/s for:</th>
<th>Good</th>
<th>Recommended</th>
<th>Transparent</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.2 Channels</td>
<td>256</td>
<td>512</td>
<td>1200</td>
</tr>
<tr>
<td>7.1 + 4 Height Channels + 4 Objects</td>
<td>200</td>
<td>384</td>
<td>800</td>
</tr>
<tr>
<td>5.1 Channels</td>
<td>96</td>
<td>160</td>
<td>256</td>
</tr>
<tr>
<td>2.0 Channels</td>
<td>32</td>
<td>56</td>
<td>160</td>
</tr>
</tbody>
</table>

- Includes the unified Fraunhofer/VoiceAge improved speech codec and new spatial coding techniques
  - Better speech quality at low bitrates
  - Improved stereo and multichannel imaging
  - Improved coding efficiency
Combining streams in the decoder

• In MPEG-H, objects may be delivered over disparate networks and supplied to the decoder separately
• Enables broadcast TV delivery of main content, with alternative or auxiliary content delivery over the Internet, for example
• Effectively, late binding in the MPEG-H decoder
Implementing MPEG-H in TV Broadcasting

Easy four-stage process as individual broadcasters desire
Implementing MPEG-H in TV Broadcasting
Developing a TV Audio System Based on MPEG-H

• The MPEG-H Audio Alliance is developing a practical TV audio system based on the MPEG-H standard

• Includes not just audio coding, but a complete end-to-end system for content production, transmission and consumer delivery

• We propose a four-stage process for implementation by broadcasters:
  – From the start, consumer decoders will support all features of our system
  – Broadcasters can choose what features to use and when to implement them
Implementing MPEG-H in TV Broadcasting
Stage 1 – Today’s Surround With Current Practices

• Replace AC-3 encoders with MPEG-H encoders
• (likely simultaneous with replacing MPEG-2 video encoders with HEVC or SHVC
• Benefits:
  – Same audio quality with half the bitrate of AC-3
  – Similar operational practices using traditional loudness dialnorm and compressor profiles of AC-3
  – Automatic loudness profiles generated for mobile or tablet delivery
Implementing MPEG-H in TV Broadcasting
Stage 2 – Add Objects For Consumer Choice Of Audio Elements

- Primary language dialogue (for Dialogue Enhancement – consumer adjustment of dialogue level)
- Second or third language dialogue
- Audio Description / Visually Impaired
- Home vs. Away team announcer
- Driver – Pit Crew radio
- Athlete’s playlist?
- Sound Effects
- Mono objects require at 20-40 kb/s each
- Objects can be programmed individually or in presets or groups
- Requires additional channels in TV plant for objects
Implementing MPEG-H in TV Broadcasting
Stage 3 – Adding Immersive Sound

- Add practical immersive audio – perhaps 5.1 or 7.1 + 4 height channels
- Increasing the realism of the audio image to create the “you’re in the scene or at the event” experience.
- May use Channels-based or HOA-based immersive sound bed + audio objects
- Requires four additional audio channels in TV plant for height channels
- System supports audio up to 128 channels, configurations to 22.2 defined in standard.
Implementing MPEG-H in TV Broadcasting
Stage 4 – Adding Dynamic Objects

• Dynamic Objects – change position over time as specified by control data in bitstream
• Interesting not only for feature film or episodic drama, but for effects in sports broadcasting and other genres
• Requires transmission of control data through broadcast plant to final emission encoder
  – We’ll have a future paper on that…
Implementing MPEG-H in TV Broadcasting
Testing MPEG-H in the field

- Winter extreme sports competition (skiing, snowboarding, snowmobile racing) carried on major cable sports network
- Summer extreme sports competition (skateboarding, motorcycle racing) carried on major cable sports network
- NASCAR race (with pit crew radios) using material from NASCAR
- DTM (European race series) auto race carried on major European sports channels
Implementing MPEG-H in TV Broadcasting
Signal and work flow in field tests

- Winter games and NASCAR Mixing at Fraunhofer in Channels + Objects
- Summer games mixing at Technicolor @ Paramount in HOA + Objects
Implementing MPEG-H in TV Broadcasting
Real-time hardware ready for tests and trials

- Real-time encoder and decoder prototype
- SDI with 16 audio channels or objects
- Audio contribution encoding at 1 Mb/s
- Audio emission encoding at 384 kb/s
- M-JPEG, AVC, or HEVC video encoding
- Each encoder or decoder is 2 RU computer, including video encoding/decoding
- Audio monitoring on standard Wolher SDI monitor panel
- Decoder has IR interface to remote control for consumer interface emulation
- Demonstrated at IBC, AES, and here in booth 108
Different Approaches to 3D Microphones for Ambience

Homemade microphone from remote truck junk box: eight generic cardoids of $150 class on stands and side arms
Eight XLRs to recorder or stage box.

Precision machined mike rails, Schoeps cardoid capsules, Rycote windscreens.
Eight XLRs to recorder or stage box.

Eigenmike® from MH Acoustics.
One Ethernet cable to PC.
Post-production Using Nuendo

- Spatial Audio Designer
Monitoring MPEG-H

- Stage 1, 2: Use existing speakers
- Stage 3, 4:
  - Add four height speakers to existing 5.1 or 7.1 speakers
  - In remote trucks:
    - Small speakers for height, high power bandwidth not needed, no “spaceship flyover”
    - Mix surround / immersive mix in remote studio (as for FIFA World Cup 2014)
    - Use immersive headphones
      - Real rendering with head tracking, individual HRTF
  - For QC or confidence monitoring: MPEG-H binaural rendering
Progress with MPEG-H continues

- MPEG Standard at DIS stage
  - Final International Standard expected in January
- Being considered for upcoming ATSC 3.0 and DVB TV standards
- Pioneering experiments in sports broadcasting and other media to test production with MPEG-H's new features
- Visit the Fraunhofer’s SMPTE demonstration on the show floor to learn more: Booth 108
- Or go to www.iis.fraunhofer.de/tvaudio or www.mpeghaa.com
Fraunhofer, Qualcomm, and Technicolor
Bring MPEG-H to the world’s media and devices

• The MPEG-H Audio Alliance’s New Immersive and Interactive TV Audio System

• Our TV audio system, based on the MPEG-H Audio standard, is being jointly developed by Fraunhofer, Qualcomm and Technicolor, who have formed the MPEG-H Audio Alliance.

• In our listening rooms at industry events, we have shown:
  – Real-time encoding hardware and the 3D sound bar prototype from Fraunhofer
  – Decoding on a set-top box from Technicolor
  – Decoding on tablet computers using Qualcomm technology.

• More information on our system and future developments may be obtained at www.iis.fraunhofer.de/tvaudio and at www.mpeghaa.com.