
IMMERSIVE AUDIO WITH MPEG-H 3D AUDIO

STATUS AND OUTLOOK

Max Neuendorf, Jan Plogsties, Stefan Meltzer

Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany

NAB Booth: SU 6117

Robert Bleidt

Fraunhofer USA Digital Media Technologies, San Jose, California

NAB Show 2014, Broadcast Engineering Conference

Advanced Technologies for Television (Part 2)

Sunday, April 6, 2014

1

MPEG-H 3D AUDIO

- Challenges of a changing TV Landscape
 - Everywhere, Personalization, Immersiveness
- MPEG-H 3D Audio System
 - Codec
 - Renderers
 - Postprocessing
 - Transport
- Timeline

Challenges of Future TV

- Users have changing expectations from the TV sound system
 - Watching films on tablets and mobile devices, not just home TV
 - Interactive aspect expected also from films and TV
 - In the living room: Convenient unobtrusive (audio) products, sometimes at the cost of quality. E.g. soundbars.

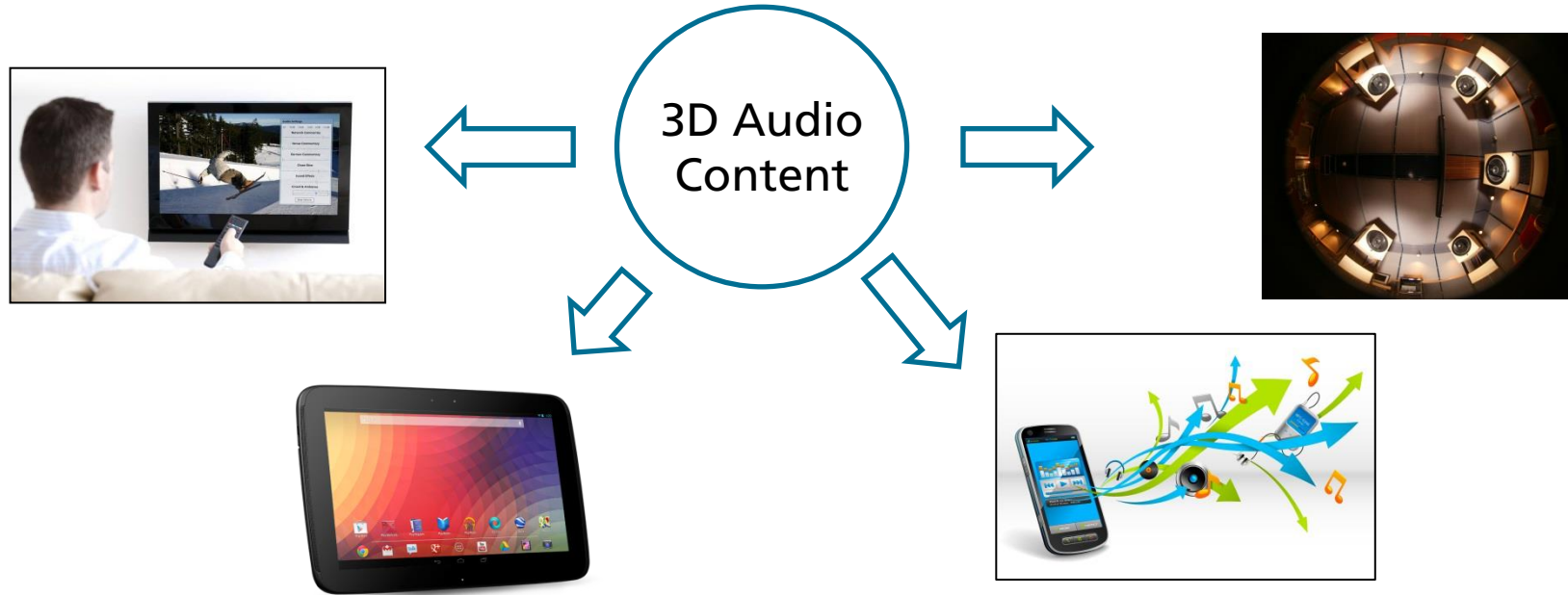
- Success of future TV will depend on how well it can compete with new media

Challenges of Future TV

- What value can a future TV audio system add to future broadcast programs?
 - **Everywhere**: Play on any device
 - Mobile over headphones, tablets, to home theatre with cinematic sound
 - **Personalization** and Interactivity
 - System that adapts to user's situation
 - Flexible sound rendering, dynamic range control, adjust and modify certain audio and sound elements
 - **Immersiveness**
 - Compelling sound experience through sound that comes from all directions

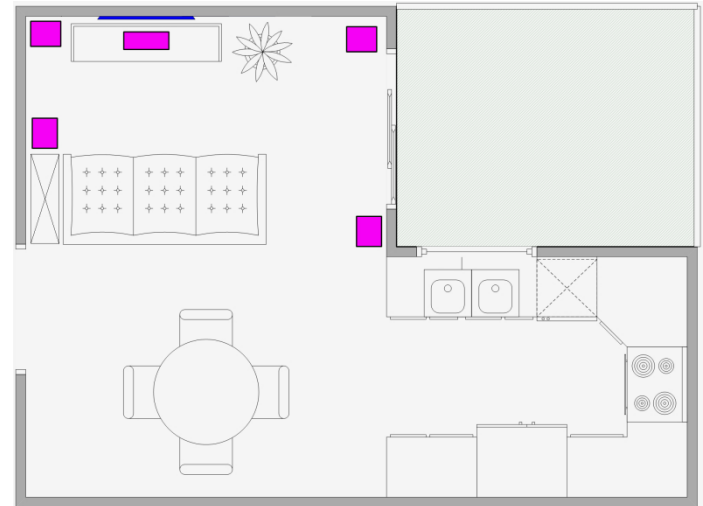
Changing TV Landscape - Everywhere

- One content plays on all devices with delivering best possible sound experience



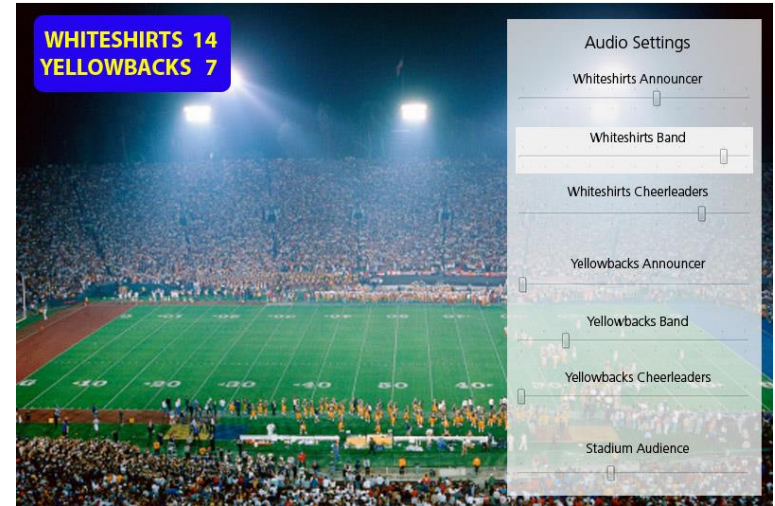
Changing TV Landscape - Personalization

- Adapt to user's situation
- Adjustment of the loudness and dynamic range based on the listening situation (home vs. mobile)
- Adjustment to the user's loudspeaker configuration through flexible rendering
 - Misplaced speakers
 - More channels than loudspeakers etc.



Changing TV Landscape - Personalization

- Set the audio mix based on your personal preference
 - Adjustment of dialogue level
 - Adjustment of voice over level
 - Changing the sound stage (home – away team commentary)

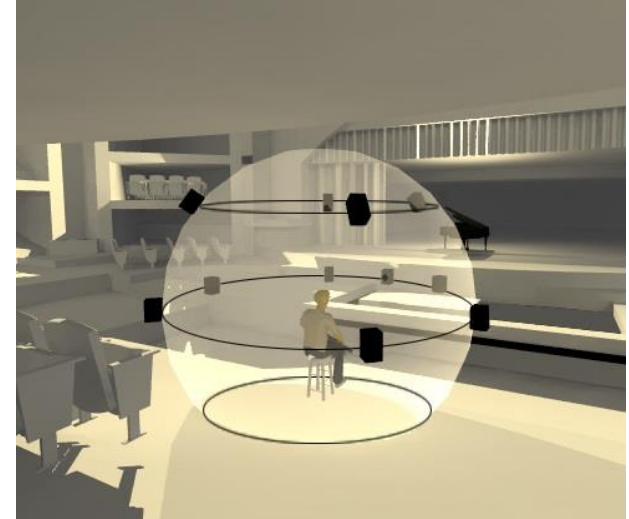
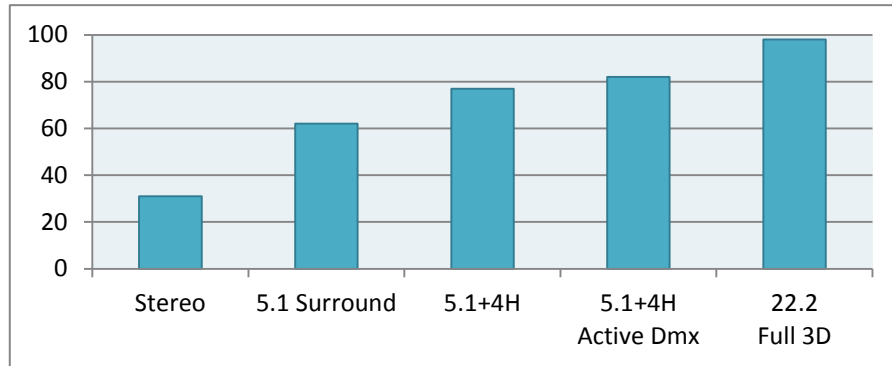


Changing TV Landscape - Personalization

- Support of multiple languages
- Selection between original and translation during a simultaneous translation
- Additional optional audio elements
(Select the team radio of your preferred driver during a car race)

Changing TV Landscape - Immersiveness

- Additional height component
 - Substantially improved sound experience



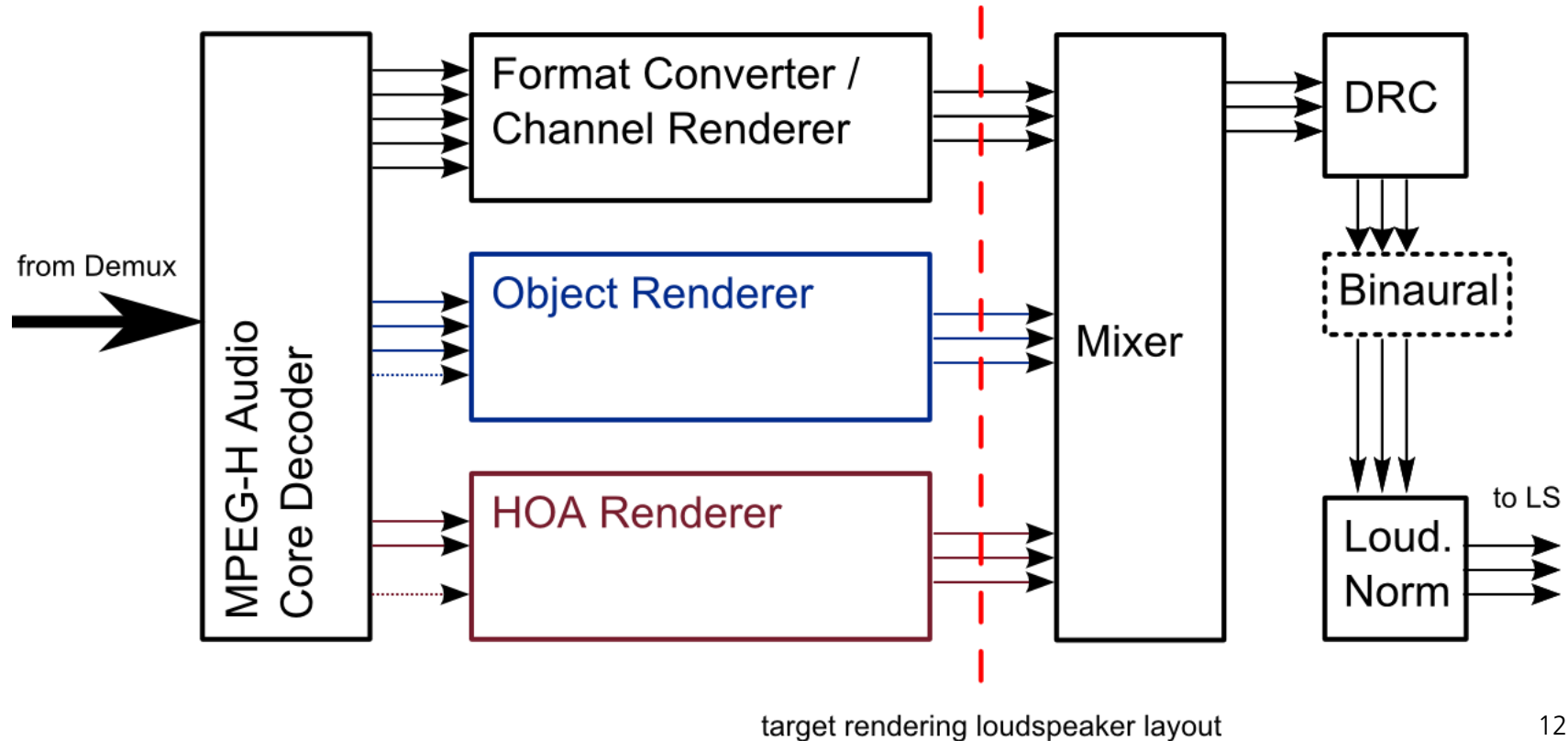
- 3D sound even on headphones through "binaural" rendering
 - Make user's perception believe she is listening to a 3D sound field

Expectations to a Future TV Audio Systems

- Play on all target devices
 - Flexible rendering from headphones to home cinema
- Allow user interactivity to tailor experience to user's liking
 - Audio objects allow user control over selection, loudness, panning of individual audio elements
- Offer compelling audio experience to the user
 - New immersive sound experience with height loudspeakers

THE MPEG-H 3D AUDIO SYSTEM

How Does MPEG-H Address these Requirements?



MPEG-H Audio Core Codec

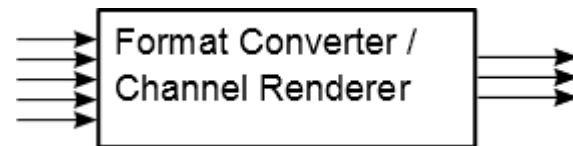
- Unsurpassed coding efficiency based on AAC evolution plus new technologies
 - Better speech quality at low bitrates
 - Improved stereo and multichannel imaging
 - Improved coding efficiency
- Additional channels or objects needed for 3D sound can be transmitted at bit rates similar to those used today for 5.1 surround broadcasts.

Bitrates in kb/s for:	Good	Recommended	Transparent
22.2 Channels	256	512	1200
7.1 + 4 Height Channels + 4 Objects	200	384	800
5.1 Channels	96	160	256
2.0 Channels	32	56	160

13

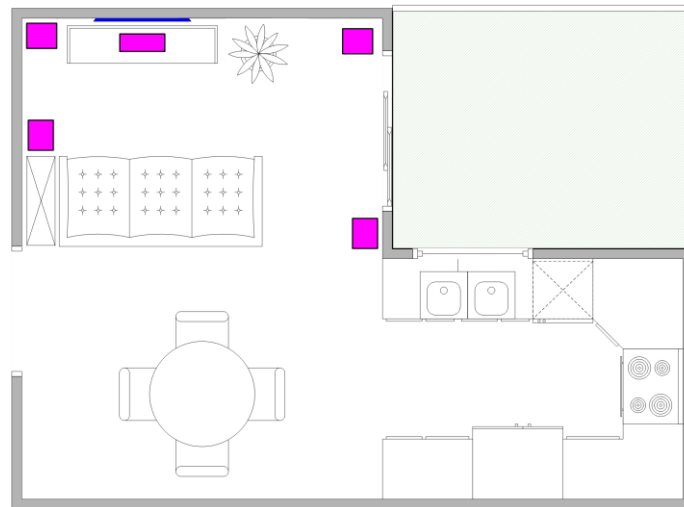
Format Converter / Channel Rendering

- Handles loudspeaker channel signals
 - traditional channel based content
 - “channel-bed” of 3D scenes
 - Flexible Rendering to target loudspeaker layout
- Retains **immersiveness** as best as possible with the available loudspeakers
- Acts like an intelligent downmix
 - Compensates timbre coloration / signal cancellation effects
- Rendering may include psychoacoustic processing to render the impression of height without height speakers, for legacy 5.1 or 2.0 consumers
- Downmix can be controlled by broadcasters if desired
- Bypassed if target loudspeaker layout is same as content layout



MPEG-H is Speaker-Foolproof

- Flexible Rendering is capable of rendering from any source layout to any target layout
 - Compensates misplaced loudspeaker positions
- Best quality playback on any configuration:
 - Stereo
 - 5.1
 - Headphones
 - Sound Frame
 - Full home theater 5.1 up to 22.2

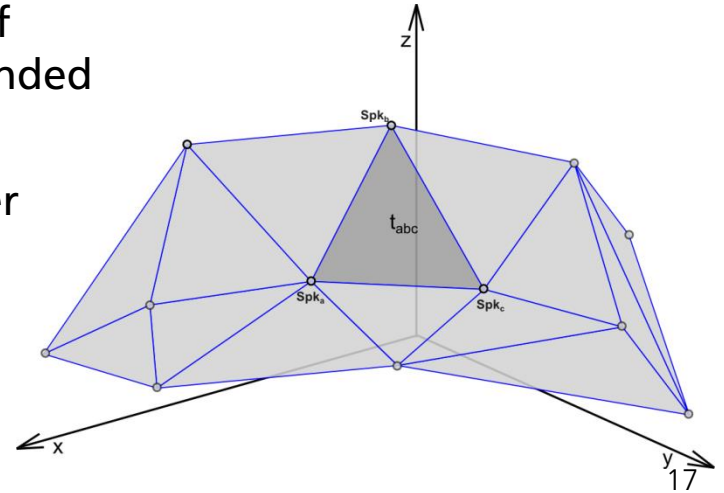


Object Renderer: What *is* an object?

- Audio object is basically a channel that can be positioned or panned during playback
- Think of transmitting mixer channels or stems instead of final outputs
 - And sending pan-pot settings as well
- Objects can be moved during playback to track action in the video
- Rendering under the viewer's control:
 - Adjust volume of objects, or turn on/off, or perhaps adjust position
- Objects can be combined with traditional channels

Object Renderer

- Renders objects to best match the loudspeaker configuration available in your **personal** listening situation
- Object channel is played over a combination of those loudspeakers that are closest to the intended object position
- Like a dynamically positioned phantom speaker
- Retains the intended perceived direction of an object



First Practical Use of Objects: Dialog Enhancement

Allows viewers to set their own dialog level

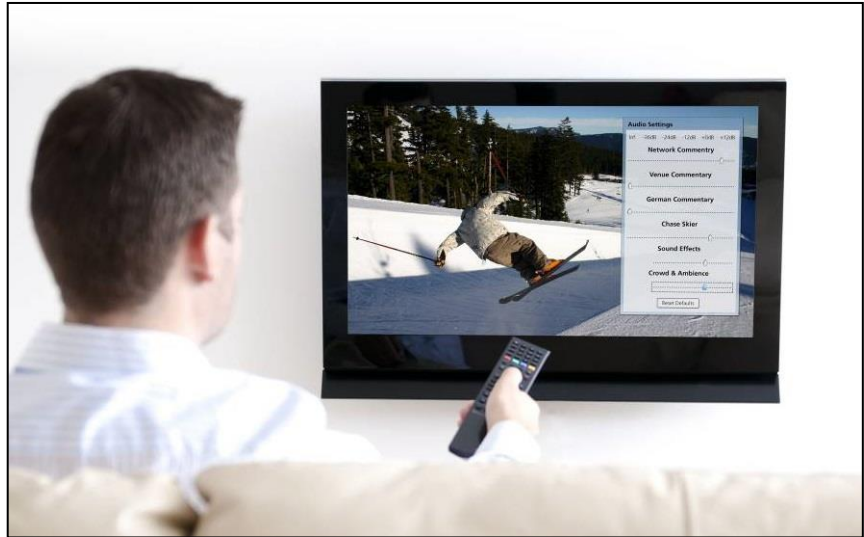
■ User Personalization

- Dialog/Announcer can be adjusted up or down
 - Helps hearing impaired understand dialog
 - Helps for listening in noisy environment
-
- Public test by BBC during the Tennis Grand Slam Championships 2011 in Wimbledon
-
- Now undergoing standardization in DVB



Other Use Cases for Objects

- Dialog in multiple languages
 - objects may be switched on or off depending on desired languages
- Audio from particular players, actors, or points of view
- Descriptive audio for visually impaired viewers
- Home team announcer vs. away team announcer at sporting event



Future Example of Objects – Football Game

The image displays a football stadium at night, filled with spectators. A blue score overlay in the top left corner reads: **WHITESHIRTS 14** and **YELLOWBACKS 7**. On the right side, a semi-transparent audio settings menu is overlaid, listing various audio sources with sliders:

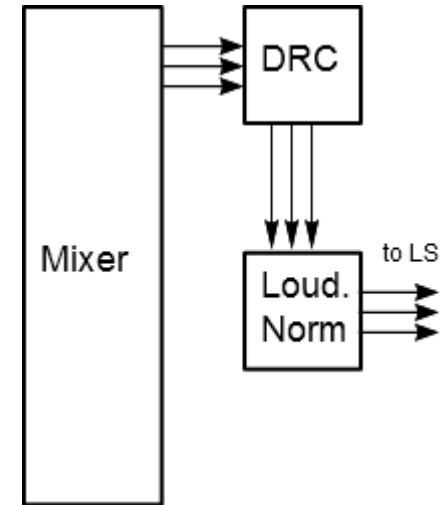
- Audio Settings
 - Whiteshirts Announcer
 - Whiteshirts Band
 - Whiteshirts Cheerleaders
 - Yellowbacks Announcer
 - Yellowbacks Band
 - Yellowbacks Cheerleaders
 - Stadium Audience

Higher Order Ambisonics (HOA) Renderer

- Loudspeaker-independent, universal, representation of a sound field
- Particularly useful for live recordings
- Captures full sound scene by using multi-capsule microphone array
- Or: synthesize HOA signals from channels/object based material
- Can be directly rendered to target loudspeaker layout to achieve optimal **Immersiveness**

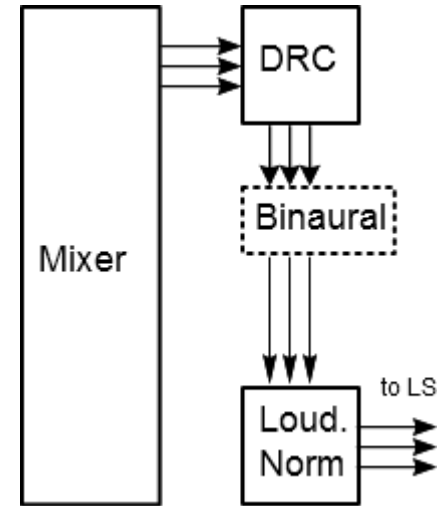
Dynamic Range Control, Loudness Normalization

- DRC allows adaptation to user's listening situation
 - Home cinema → high dynamic range
 - Mobile / airplane / late-night → low dynamic range
- Can be multi-band
- Additional DRC processing blocks in each rendering path (omitted in overview graph)
- Loudness normalization to comply to CALM
- Optional peak limiter follows loudness normalization



Binaural Rendering

- Optional post-processing step for listening on headphones
- Causes "externalization" of sound – no in-head sensation
- Applies "Binaural Room Impulse Responses" (BRIR) to Mixer output signals
- Gives sound sensation "as if" listening to loudspeakers in a certain room
- Great for getting 3D sound **immersion** from mobile devices



FURTHER ASPECTS

Aspects of Delivery / Transmission

- DASH support for stutter-free streaming
- New “audio I-frame”:
 - easy DASH bit stream switching between any configurations
 - easy splicing for Ad insertion
- Object Priority
 - Dialog maintained as object
 - Other objects mixed into bed or grouped object for mobile bandwidth reduction

Aspects of Delivery / Transmission

- Common codec allows object transmission over disparate networks with combination in MPEG-H decoder
 - Internet
 - ATSC 3.0 fixed
 - ATSC 3.0 Mobile
 - 4G and 5G wireless

MPEG Timeline

- ISO "Committee Draft" issued two days ago, available for review next Friday
- "Draft International Standard" in July → freeze status
- "International Standard" expected in February 2015

Fraunhofer's Interactive 3D Audio System for TV BC

- Come to booth SU6117
- Experience a real 3D Audio system with technology preview of a 3D Audio soundbar
- Discuss
 - Audio for Next Generation Broadcasting tomorrow, Monday, 3pm S227
- Visit
 - www.iis.fraunhofer.de/tvaudio



