107^{th} MPEG San Jose (CA), USA, 13-17 Jan. 2014, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 MPEG wins Emmy for MPEG-2 Transport Stream

At the Consumer Electronics Show held earlier this month in Las Vegas, *MPEG received its fourth Emmy award*, this time for the ubiquitous MPEG-2 Transport Stream (TS) technology which is part of the Systems specification formally known as Rec. ITU-T H.222.0 | ISO/IEC 13818-1. The Transport Stream, which was approved in November, 1994, is widely deployed across a very broad range of application environments, including cable TV, satellite broadcast, terrestrial broadcast, IPTV, and Blu-ray Disc. Its use is not just historical, but rather, MPEG-2 TS is also used in conjunction with the latest state-of-the-art video and audio compression specifications. The number of devices using this standard is still increasing. MPEG-2 TS is projected to sustain approximately 650 million units per year.

2 Unified Architecture for MPEG-H 3D-Audio CO and HOA

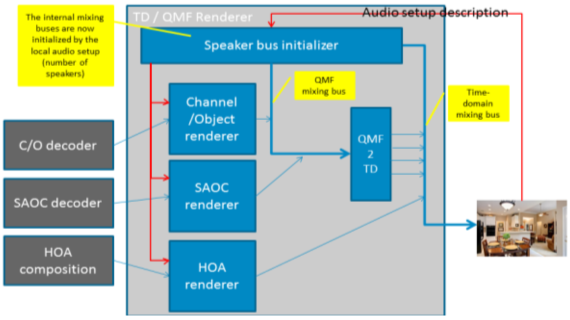
At the 107^{th} MPEG meeting two working draft documents were issued capturing the current development progress of the two reference models (RM) for MPEG-H 3D Audio Channels & Objects (CO) and MPEG-H 3D Audio Higher Order Ambisonics (HOA) (N14263, N14264).

Contribution M32244 gave convincing evidence that the HOA system can operate with the same core channel coder as the CO RM.

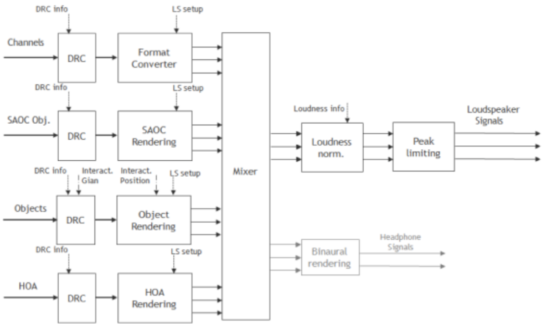
At the same meeting several proposals were brought forward which propose a reorganization of functional blocks of RM1 of MPEG-H in order to accommodate the two different coding 3D Audio coding approaches, Channels & Objects (CO), and Higher Order Ambisonics (HOA) in one algorithmic architecture (M32241, M32245).

Envisioned Architectures as seen by various proponents

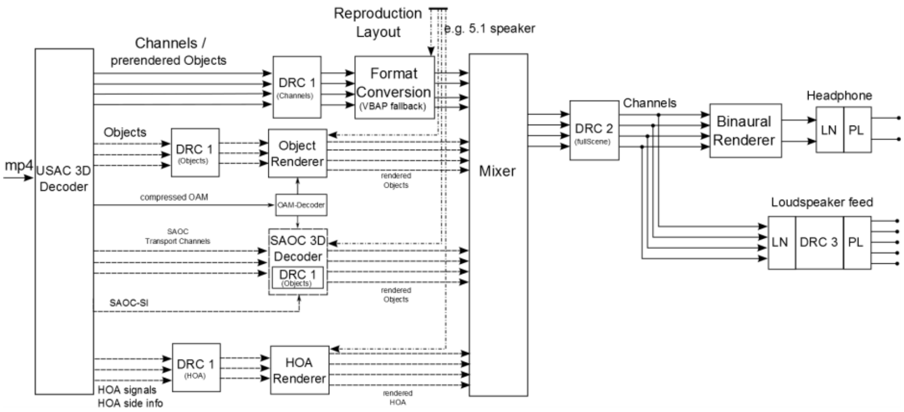
Philips / Orange / BlueRippleSound



Technicolor



Fraunhofer IIS



MPEG-H 3D Audio Time Line

Meeting	Date	Action
~~105 Vienna~~	~~August 2013~~	~~Selection of Reference Model 0 technology~~
~~106 Geneva~~	~~October 2013~~	~~Submission of Reference Model 0 Working Draft text and Reference Software.~~
107 San Jose	January 2014	Produce WD1 text of CO Produce WD1 text of HOA
AhG period		Work on Unified architecture for MPEG-H 3DA
108 Valencia	March 2014	Produce CD text. Initiate reduced 2 month CD ballot.
109 Sapporo	July 2014	Produce DIS text. Initiate regular 5 month DIS ballot.
110 Strasbourg	October 2014	
AhG period		DIS ballot approval with all "Yes" votes
111 Geneva	February 2015	IS

References

1. M32244 "Progress Report on Unification of MPEG-H 3D-Audio CO and HOA"
2. M32241 "Unified decoder and renderer architecture and API"
3. M32245 "Vision of a MPEG-H 3D Audio unified architecture"

Output documents

N14263 - WD1-CO Text of MPEG-H 3D Audio

N14264 - WD1-HOA Text of MPEG-H 3D Audio

3 3D-Audio Decoder Interfaces

Multiple input contributions to the 107th MPEG meeting have addressed the topic of communicating properties of the playback setup and Binaural Room Impulse Response (BRIR) and object control data to the decoder. Syntax proposals for the BRIR data [M31427], external loudspeaker configuration data [M32241, M31427] and object control data [M32184], [M32241] are available as contributions to the 106th and 107th meetings.

Objectives of this work are to define an interface for reading non-audio data from applications and connected devices, namely:

- Playback-side channel configuration for connected loudspeakers or
- BRIR data to the binaural renderer
- Object control data to modify object gains.

In addition also data written from the decoder to an external application or device should be considered, using the same or similar interface. Specific aspects to be taken into account are:

- Potential alignment of speaker configuration data from the user with reference speaker configuration data from the content [M32182],
- The type of BRIR interface, i.e. generic (FIR) and/or low-level parameters, where low level parameters are defined by the binaural rendering of MPEG-H 3D audio
- Using a suitable data representation format for data exchange

Interested parties are encouraged to share their ideas and contributions regarding the interface via the mpeg-audio-call reflector. Input contributions for the next meeting should propose complete syntax and functionality of the decoder to read such data.

Timeline

Activity	Completion Date	Responsible Party
Define data decoder interface data format	Input to 108th meeting	MPEG Audio Subgroup
Describe data formatting inside decoder interface	Input to 108th meeting	MPEG Audio Subgroup
Prepare text and software for decoder interface	Input to108th meeting	MPEG Audio Subgroup

References

1. M31427 "Thoughts on Binaural decoder parameterization"
2. M32341 "Unified decoder and renderer architecture and API"
3. M32182 "Flexible Signalling of 3D Loudspeaker Configurations for MPEG-H 3D Audio"
4. M32184 "Thoughts on an Interface to Device-specific Rendering"

4 Dynamic Range Control

Dynamic Range Control (DRC) technology for audio has been selected at the 106th MPEG meeting. The corresponding WD on DRC [M32271] was submitted to the current 107th MPEG meeting.

Consumer audio systems and devices are used in a large variety of configurations and acoustical environments. For many of these scenarios, the audio reproduction quality can be improved by appropriate control of content dynamics and loudness.

This part of ISO/IEC 23003 provides a universal dynamic range control tool that supports loudness normalization. The DRC tool offers a bitrate efficient representation of dynamically compressed versions of an audio signal. This is achieved by adding a low-bitrate DRC metadata stream to the audio signal. The DRC tool includes dedicated sections for clipping prevention and for generating fade-in and fade-out to supplement the main dynamic range compression functionality. The DRC effects available at the DRC decoder are generated at the DRC encoder side. At the DRC decoder side, the audio signal may be played back without applying the DRC tool, or an appropriate DRC tool effect is selected and applied based on the given playback scenario.

The DRC tool provides efficient control of dynamic range, loudness, and clipping based on metadata generated at the encoder. The decoder can choose to selectively apply the metadata to the audio signal to achieve a desired result. Metadata for dynamic range compression consists of encoded time-varying gain values that can be applied to the audio signal. Hence, the main blocks of the DRC tool include a DRC gain encoder, a DRC gain decoder, a DRC gain modification block, and a DRC gain application block. These blocks are exercised on a frame-by-frame basis during audio processing. Various applicable DRC configurations are encoded separate from the DRC gain codes. Based on the applicable DRC configurations, the DRC set selection block decides based on the playback scenario which DRC gains to apply to the audio signal. Moreover, the DRC tool supports loudness normalization based on loudness metadata.

A typical system for loudness and dynamic range control in the time domain is shown in Figure **1**. The decoder part of the DRC tool is driven by metadata that efficiently represents the DRC gain samples and parameters for interpolation. The gain samples can be updated as fast as necessary to accurately represent gain changes down to at least 1 ms update intervals. In the following the decoder part of the DRC tool is referred to as “DRC decoder”, which includes everything except the audio decoder and associated bitstream de-multiplexing.

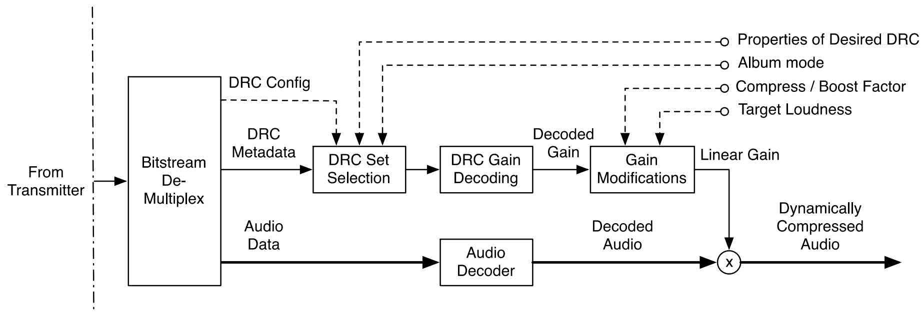


Figure 1 - Block diagram of a typical system with audio decoder and DRC tool modules to achieve loudness normalization and dynamic range control.

References

1. M32271 "Working Draft 0 on DRC"
2. M32397 "Proposed Updates to MPEG-4 Audio"

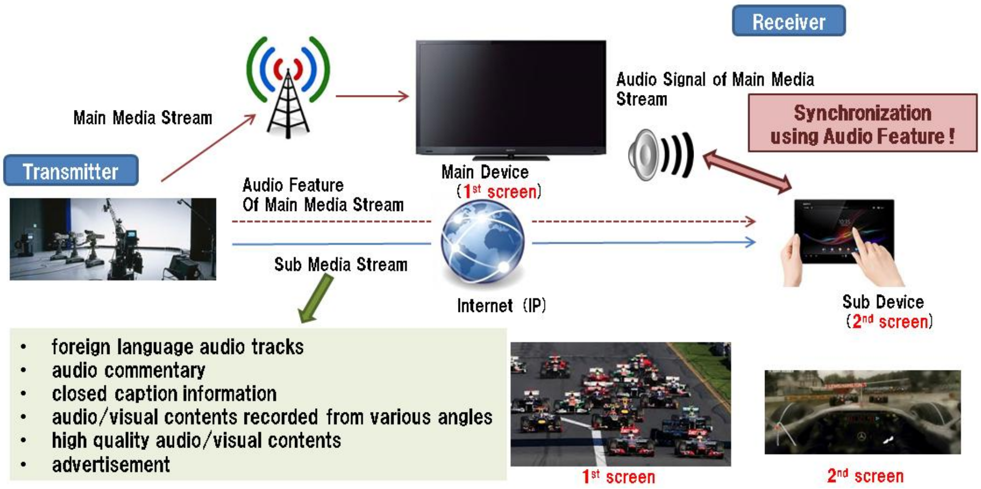
Output documents

N14261 - WD1 Text of Dynamic Range Control

N14262 - RM1 Software for Dynamic Range Control

5 Exploration - Audio Synchronization based on Fingerprinting

The audio synchronization technology allows “second screen” applications where the 2^{nd} screen content is synchronized to the 1^{st} screen content. In this scenario, no common clock covering the 1^{st} and 2^{nd} screen devices is required, nor way to exchange time-stamps between the devices. Synchronization of the contents between the devices is done by using audio features (fingerprint) extracted from the 1^{st} screen content.



In this example, 1st screen content is distributed over existing broadcast system, and 2nd screen content is distributed over IP network, where 1st screen content is a regular TV program of car racing, and 2nd screen content is a drivers view of the car racing. Feature of the 1st screen audio content is used for the synchronization. The audio feature (fingerprint) stream of the 1st screen content is sent to 2nd screen together with 2nd screen audio/video content over the IP network. In the 2nd screen device, audio of the 1st screen content is captured by microphone and its feature is extracted. The extracted feature from the microphone input and received feature from IP network is compared and time difference is computed. This time difference is used to align 2nd screen audio/video content to the 1st screen content.

References

1. M31368 "Proposed scheme for synchronization of multiple audio streams"
2. M32189 "Cross-check report of audio synchronization"
3. M32236 "Cross Check Report of Audio Synchronization Experiment"

Output documents

N14069 - Work plan for audio synchronization

6 Evaluation Report of the responses to the CfP on MPEG User Description

At the 105th MPEG meeting, the Requirements group issued a Call for Proposals (CfP) on MPEG User Description [1], with Requirements on MPEG-UD [2] and Use Cases on MPEG-UD [3]. In the 106th and 107th meetings, five parties (namely ETRI, RAI, KETI, TSP, InSignal) have submitted their technologies as responses to the call for proposals. In their contributions, these five parties proposed data representations using XML schema and RDF format for User Description (UD), Context Description (CD), Service Description (SD), and Recommendation Description (RD) responding to the call. To evaluate the CfP responses, the AHG followed the procedure specified in the CfP document [1]. In the following tables a summary of the evaluation results is presented, w.r.t to the CfP requirements, towards a working draft on MPEG User Description (MPEG-UD).

Legend

O	Requirement completely satisfied
x	Requirement partially satisfied
-	No inputs for this Requirement
	Requirement not satisfied by this party

Table 1. Conformance to General requirements

	General	ETRI	RAI	KETI	TSP	INSIGNAL	Proposals to harmonize	Satisfied?	Based on
1	**Structure**	o	o	o	o	O	ETRI, KETI, TSP, INSIGNAL, RAI	Y	
2	**Extensibility**	o	o	o	o	O	ETRI, KETI, TSP, INSIGNAL, RAI	Y	
3	**Security & Privacy & Traceability**	-	-	-	-	-		N	
4	**Access, Modularity, and Adaptivity**	-	-	-	-	-		N	
5	**Licensing**	-	-	-	-	-		N	
6	**Priorities**			O	o		KETI + TSP	Y	
7	**Compatibility**	x	x	O	x	x	KETI	Y	

Table 2. Conformance to User Description (UD) requirements

	UD	ETRI	RAI	KETI	TSP	INSIGNAL	Proposals to harmonize	Satisfied?	Based on
1	**User Identification**	o		o	o	x	KETI + ETRI + TSP	Y	MPEG-7 and others
2	**Birthdate and Birthplace**	x		o	x	x	KETI	Y	birthplace from MPEG 7
3	**Gender**	o		o	o	o	ETRI, KETI, TSP, INSIGNAL	Y	MPEG-7 AMD 4
4	**Profession**			o	x		KETI	Y	MPEG-7
5	**User's Specialties**			o			KETI	Y	
6	**Additional User Information**	x		o			KETI	Y	
7	**Interactions with Other Users (Social Networks)**	o		o			KETI + ETRI	Y	MPEG-7
8	**Interaction with Multimedia**	o	o	o			KETI + RAI + ETRI	Y	MPEG-7
9	**Data Usage**		x	o			RAI, KETI	Y	MPEG-7
10	**Actions with Data**			o			RAI, KETI	Y	MPEG-7
11	**Classification of User Experience when Interacting with Application**		o	x			RAI	Y	
12	**Multi-Domain Identification of Users**			x			KETI	Partially	
13	**Proxies and Authorizations**	-	-	-	-	-		N	
14	**User Language Description**	o		o		x	KETI + ETRI	Y	MPEG-7
15	**Preferences on Languages and Speaking Manners**	o		o	x		KETI + ETRI	Y	
16	**Accessibility**			o			KETI	Y	MPEG-21
17	**Service Usage Preferences**			o	o		KETI + TSP	Y	
18	**Preference on Presentation Style**			o			KETI	Y	
19	**Preference on Service Provider**			o			KETI	Y	
20	**Personalization Profile**			o			KETI	Y	
21	**Interested Topics**			o			KETI	Y	MPEG-7
22	**Interested Media**			o		x	KETI	Y	MPEG-7
23	**Intention**	o		o	o		KETI + ETRI + TSP	Y	
24	**Emotion**			o			KETI	Y	
25	**User's Age-related Functional Limitation**			o			KETI	Y	
26	**Different Abilities**			o			KETI	Y	
27	**History of Service Usage**		o		o	x	RAI + TSP	Y	
28	**Capturing Context and Service Information**	-	-	-	-	-		N	
29	**Score of Satisfaction**			o			KETI	Y	
30	**Additional User Information**	x		o			KETI	Y	

Table 3. Conformance to Context Description (CD) requirements

	CD	ETRI	RAI	KETI	TSP	INSIGNAL	Proposals to harmonize	Satisfied?	Based on
1	**Static Information of current device in use**	o		o	o	x	KETI + ETRI + TSP	Y	MPEG-21
2	**Dynamic Information of Current Device in Use**			o	o	x	KETI + TSP	Y	MPEG-21
3	**Available Device**	o		o		x	KETI + ETRI	Y	
4	**Time of the Day**	o		o		x	KETI + ETRI	Y	MPEG-7 and others
5	**Season**	o		o			KETI + ETRI	Y	
6	**Temperature**			o			KETI	Y	MPEG-V
7	**Weather**	x		o	x		KETI	Y	MPEG-V
8	**Geographic Location Information**	o		o	o		KETI + ETRI + TSP	Y	MPEG-7
9	**Semantic Location Information**	o		o	o		KETI + ETRI + TSP	Y	MPEG-7
10	**Additional Viewing Environmental Information**	x		o			KETI	Y	MPEG-21
11	**Aggregation of Context Information**	x		o			KETI	Y	
12	**Sharing/Synchronizing**	x		o			KETI	Y	

Table 4. Conformance to Service Description (SD) requirements

	SD	ETRI	RAI	KETI	TSP	INSIGNAL	Proposals to harmonize	Satisfied?	Based on
1	**Social Networks and Communities**	o				x	ETRI	Y	
2	**Social Activities**	o					ETRI	Y	
3	**Usage Permissions**	x		o	x	x	KETI	Y	MPEG-21
4	**Supported Multimedia Objects**	o		o	o	x	KETI + ETRI + TSP	Y	MPEG-M
5	**Supported Representation Formats**	o		o	x	x	KETI + ETRI	Y	MPEG-M
6	**Limit of Usage**	x		o		x	KETI	Y	MPEG-21
7	**Service Products Category**	x		o			KETI	Y	
8	**Service Products Information**	x		o	o		KETI + TSP	Y	
9	**Service Level Agreement**	o		o		x	KETI + ETRI	Y	MPEG-M
10	**Aggregated Services**			o			KETI	Y	
11	**Baseline Information**			o			KETI	Y	
12	**Service Priorities**			o	o		KETI + TSP	Y	
13	**Service Messages**	-	-	-	-	-		N	
14	**Additional Service Information**			o			KETI	Y	

Table 5. Conformance to Recommendation Description (RD) requirements

	RD	ETRI	RAI	KETI	TSP	INSIGNAL	Proposals to harmonize	Satisfied?	Based on
1	**Representation of Usage Roles of UD, CD and SD elements**		x	o	o		KETI + TSP	Y	
2	**Objects Classification Representation**	o	o	o	x		RAI + ETRI + KETI	Y	
3	**History of Acceptance**		o		o		RAI + TSP	Y	
4	**Representation of Objects Equivalence**		o				RAI	Y	
5	**Process Representation**				o		TSP	Y	
6	**Representation of objects relationships**	o	o				RAI + ETRI	Y	

Table 6. Summary of the evaluation

	General	UD	CD	SD	RD	Total
Satisfied	**2**	**27**	**12**	**13**	**6**	**62**
Partially satisfied	**0**	**1**	**0**	**0**	**0**	**1**
Not satisfied	**3**	**2**	**0**	**1**	**0**	**6**

References

1. N13879 “Call for Proposals on MPEG-UD”, Vienna, AT, August, 2013
2. N13881 “Requirements on MPEG-UD”, Vienna, AT, August, 2013
3. N13880 “Use Cases for MPEG-UD”, Vienna, AT, August 2013

Output documents

N14287 - Evaluation Report of the responses to the CfP on MPEG User Description

N14185 - WD of MPEG User Description

N14186 - Work Plan of MPEG User Description