106th MPEG Geneva, Switzerland, 28 Oct. - 1 Nov. 2013, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

1. An excerpt: Website Content Audio Promotion - www.mpeg-audio.org

For more than ten years, MPEG Advanced Audio Coding (AAC) has been the most important audio codec for state-of-the-art multimedia systems. During that time it had been continuously updated and is still the state-of-the-art audio codec today.

Thanks to its versatility, MPEG AAC is natively supported in every leading operating system (including Windows and OSX), media player and browser. Additionally, MPEG AAC is part of every mobile operating system (including Android and iOS), it is used for streaming audio to billions of devices and used for audio in all mobile TV standards around the globe. Also, the classic consumer electronics world relies heavily on AAC, as it is used in all tablets and game consoles, in most DTVs, Bluray players, STBs and music players and in most cars and digital radios. At the end of 2012, MPEG AAC has been licensed for more than six billion devices worldwide, a number which is constantly growing.

The AAC codecs offer a balanced sound quality over a wide spectrum of bit rates: from highest audio quality at bit rates of 256 kbit/s, to good audio quality at bit rates as low as 16 kbit/s. At all bit rates AAC represents the state-of-the-art audio codec. This versatility allows for both, uncompromised high-end audio with 24 bit/96 KHz up to 24 bit/192 KHz as well as an impressive audio experience over channels with limited bandwidth, such as those in broadcast or mobile multimedia streaming. AAC also offers special low-delay versions, which are used in Apple's Facetime and which are today's de facto audio standard for high-quality communication systems, such as video conferencing and telepresence.

On top of being a versatile, reliable audio codec, AAC offers all the features needed for delivery of stereo and surround content to any device: Loudness metadata, binaural headphone mode, adaptive streaming and many more. As an open ISO Standard, AAC offers a freedom unmatched by any closed eco system.

To help AAC decoder manufacturers with testing their implementation, this website offers test bitstreams that go beyond MPEG audio conformance. With those files manufacturers can make sure that their implementation works across all the different applications and ecosystems (digital TV, mobile and Internet streaming, digital radio, gaming etc.).

The test bitstreams, further informational material, and FAQ section will make it easy for companies to use MPEG AAC codecs.

In another context, it is worth to be mentioned that MPEG would be the recipient by "The National Academy of Television Arts & Sciences" of its third EMMY award, this time, for the development and standardisation of the MPEG-2 Transport Stream.

2. Working Draft Text of MPEG-H 3D Audio Channels/Objects

At the 105th MPEG meeting in Vienna the audio group evaluated submitted proposals for the CfP on MPEG-H 3D Audio and selected Reference Model 0 for Channels and Objects (RM0-CO). RM0-CO is based on MPEG-D USAC and MPEG-D SAOC and extends it with additional tools.

Encoder and decoder block diagrams

The 3D Audio Codec System is based on an MPEG-D USAC Codec for coding of channel and object signals. To increase the efficiency for coding a large amount of objects, MPEG SAOC technology has been adopted. Three types of renderers perform the tasks of rendering objects to channels, rendering channels to headphones or rendering channels to a different loudspeaker setup.

When object signals are explicitly transmitted or parametrically encoded using SAOC, the corresponding Object Metadata information is compressed and multiplexed into the 3D-Audio bit stream.

Figure 1 and Figure 2 show the different algorithmic blocks of the 3D-Audio system.

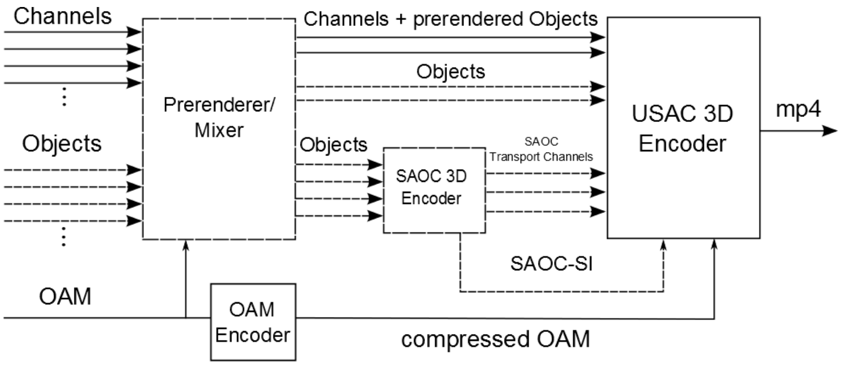

Figure 1 - Block diagram of the 3D-Audio encoder

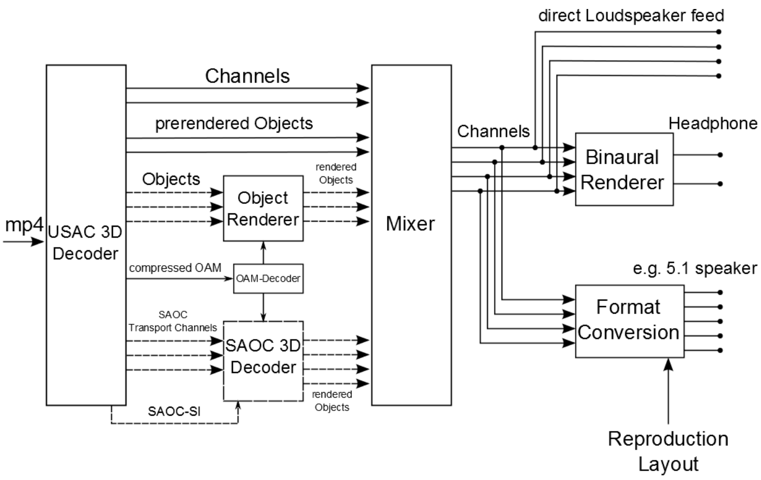

Figure 2 - Block diagram of the 3D-Audio decoder

3. Working Draft Text of MPEG-H 3D Audio Higher Order Ambisonics

At the 105th MPEG Meeting in Vienna the proposal of Technicolor and Orange has been selected as Reference Model 0 for coding of Higher Order Ambisonics content (RM0-HOA) in the envisioned MPEG-H 3D Audio standard. This proposal does not use the ISO file format, but uses a very basic format only containing the absolute inevitable parameters to allow for decoding the bit stream. For this reason and due to the planned technology merge of the RM0 for coding of channel-based and object-based content (RM0-CO) and the RM0-HOA, it is very likely that changes will be applied to the bit stream architecture.

Decoder block diagram

A block diagram of the decoder architecture is depicted in Figure 3.

First, the input bit stream is de-multiplexed into I bit streams originally created by the AAC-family mono encoders plus the parameters required to recompose the full HOA representation from these bit streams.

In the multi-channel perceptual decoding component the I bit streams are individually decoded by AAC-family mono decoders to produce I signals.

In the successive spatial decoding component, first, the actual value range of these signals is reconstructed by the inverse gain control processing. In a next step, the I signals are re-distributed to provide the M pre-dominant signals and $(I-M)$ HOA coefficient signals representing the more ambient HOA components.

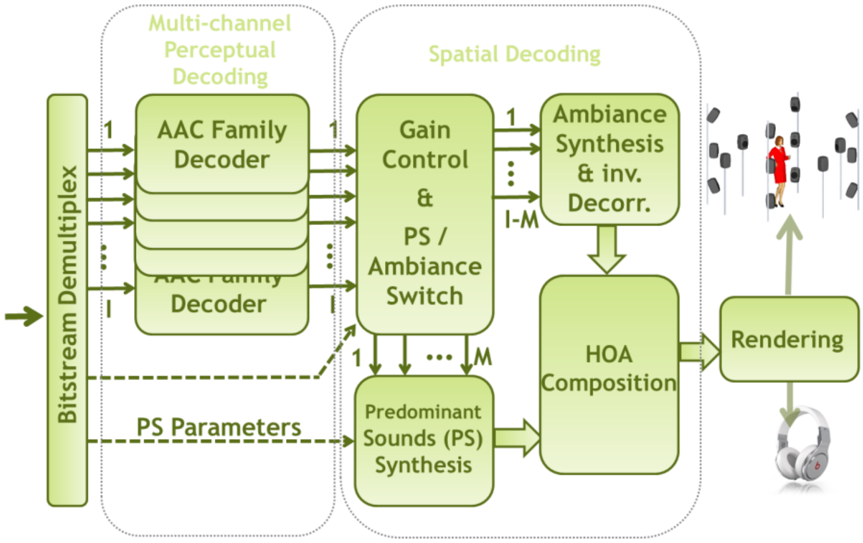

Figure 3 - Simplified block diagram of the decoder

The fixed subset of the $(I-M)$ HOA coefficient signals is re-correlated, this means the decorrelation at the HOA encoding stage is reversed.

Next, all of the $(I-M)$ HOA coefficient signals are used to create the ambient HOA components.

The predominant HOA components are synthesized from the M predominant signals and the corresponding parameters.

Finally, the predominant and the ambient HOA components are composed into the desired full HOA representation, which is then rendered to a given loudspeaker setup.

4. Submission and Evaluation Procedures for 3D Audio Phase 2

The Call for Proposals for 3D Audio (Call) issued at the 103rd MPEG meeting held in Geneva, CH in January 2013. Submissions to the Call are to be evaluated at the 109th MPEG meeting to be held in Sapporo, JP, in July 2014.

The following companies, as in Table 1, have registered their intent to participate in the Call. The rightmost columns indicate an intention to participate in Phase 2 of the Call and to submit Set 1 (CO) or Set 2 (HOA) bit streams and decoded signals.

Table 2 gives the timeline for submissions to the Call. Reference Model X (RM X) is the latest RM as of the time of the relevant timeline date or meeting date.

Table 1 - Registered Companies

	ID	Company	Set	
			1	2
1	ETRI	ETRI	x	
2	IDMT	FhG-IDMT	x	
3	IIS	FhG-IIS	x	
4	HUA	Huawei	x	x
5	IOS	Iosono	x	
6	NHK	NHK	x	
7	ORL	Orange Labs	X	x
8	PHI	Philips	x	
9	QUAL	Qualcomm	x	x
10	SONY	Sony	x	
11	SWS	Swissaudec	x	x
12	TECH	Technicolor	x	x
13	VIS	VisiSonics	x	x
		Phase 2 Totals	**13**	**6**

Table 2 – Timeline for Phase 2

Meeting / Date	Action
April 14, 2014	Proponent Contact sends email to Test Administrator declaring the intent to participate in Phase 2 testing.
April 18, 2014	Proponent Contacts that have sent email to Test Administrator will receive Call FTP site information from Test Administrator
May 16, 2014	Proponent processed test items submitted to Call FTP site.
May 23, 2014	Test items available to Listening Labs on Call FTP site
	Conduct evaluation listening tests
June 27, 2014	Listening Labs submit • Raw scores via email to Test Administrator • Contribution to 109th MPEG meeting describing their test setup.
109th meeting, July 2014	Proponent written documentation submitted as contribution to 109th MPEG meeting (see Call for Proposals for 3D Audio [1] for details on what documentation must be submitted and when it must be submitted).
109th meeting, July 2014	Selection of technology to incorporate into Reference Model X for both CO and HOA technologies.
110th meeting, October 2014	Proponent(s) must submit Reference Model X Working Draft text and Reference Software.

References

1. N13411, Call for Proposals for 3D Audio
2. N13412, Encoder Input Format for MPEG-H 3D Audio
3. N12633, Submission and Evaluation Procedures for 3D Audio

Output documents

N14060 - Working Draft Text of MPEG-H 3D Audio CO RM0

N14061 - Working Draft Text of MPEG-H 3D Audio HOA RM0

N14062 - Software for MPEG-H 3D Audio CO RM0

N14063 - Software for MPEG-H 3D Audio HOA RM0

N14064 - Submission and Evaluation Procedures for 3D Audio Phase 2

N14065 - Workplan on 3D Audio

N14066 - Workplan on Binauralization CE

5. Dynamic Range Control

At the 105th MPEG meeting in Vienna WG11 issued a Call for Proposals on Dynamic Range Control Technology [N13858], with submissions due at the 106th meeting in Geneva.

There were three responses to the call, from Apple [M31471, M31472, M31473], Fraunhofer IIS [M31384] and Sony [M31359]. These submissions were reviewed at the 106th MPEG meeting. It was decided to select the Apple technology and to investigate and extend this with elements from the Fraunhofer IIS and Sony submissions. The remainder of this brief is a workplan for this investigation.

In order to investigate the impact of the Sony proposal considering the current Apple proposal, and to perform an integration of the Fraunhofer IIS proposal into Apple's proposal, the following activities are planned.

Activity	Completion Date	Responsibility
Merge the functional blocks of the Fraunhofer IIS proposal into Apple's proposal.	Nov. 22, 2013	Apple, Fraunhofer IIS
Supply test sets of uncompressed DRC gain sequences in audio sampling rate resolution.	Nov. 22, 2013	Apple, Sony, Fraunhofer IIS
Exchange of DRC gain bit streams for cross-check. Report bitrates for Apple's current proposal and the version extended by Sony's predictive coding based on the test sets.	Dec. 17, 2013	Apple, Sony
Perform a listening test to evaluate the audio quality for Apple's spline interpolation according to [M31471] and Sony's linear interpolation proposal [M31359] using similar bit rates for the coded DRC gains.	Dec. 17, 2013	Sony, Apple, Fraunhofer IIS
Report of results	Contribution to 107th meeting	Sony, Apple, Fraunhofer IIS
Extend and adapt DRC WD to support the FhG-IIS DRC proposal (i.e. Loudness Normalization, DRC, guided Clipping Prevention, Peak Limiter) and to work within the 3D Audio architecture as outlined in doc [M31360]. Use of Peak Limiter is optional, but strongly encouraged	Contribution to 107th meeting	Sony, Apple, Fraunhofer IIS
Study how to represent, store and transmit Downmix Matrix Coding and optional equalization parameters for use prior to downmix	Contribution to 107th meeting	Apple, Fraunhofer IIS

Output documents

N14067 - Working Draft on Dynamic Range Control

N14068 - Workplan on Dynamic Range Control

6. Technologies under Consideration for MPEG User Description (MPEG-UD)

In the 106th meeting in Geneva, several technologies submitted for evaluation according to the call for proposals [1,2]. In the following, the preliminary evaluation of these technologies presented w.r.t to the CfP requirements, towards a working draft on MPEG User Description (MPEG-UD).

Legend	O	(completely satisfied)	X	(partially satisfied)		(not satisfied)

Conformance to User Description (UD) requirements

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

Conformance to Context Description (CD) requirements

CD		ETRI	RAI	KETI	TSP	INSIGNAL
1	Static Information of current device in use	X		O	X	X
2	Dynamic Information of Current Device in Use			X	X	X
3	Available Device	X		O		X
4	Time of the Day	X				X
5	Season	X				
6	Temperature			X		
7	Weather	X		X	X	
8	Geographic Location Information	X		O	X	
9	Semantic Location Information	X		O	X	
10	Additional Viewing Environmental Information	X		X		
11	Aggregation of Context Information	X		O		
12	Sharing/Synchronizing	X				

Conformance to Service Description (SD) requirements

SD		ETRI	RAI	KETI	TSP	INSIGNAL
1	Social Networks and Communities					X
2	Social Activities					
3	Usage Permissions	X			X	X
4	Supported Multimedia Objects				X	X
5	Supported Representation Formats				X	X
6	Limit of Usage	X				X
7	Service Products Category	X				
8	Service Products Information	X			X	
9	Service Level Agreement	X				X
10	Aggregated Services					
11	Baseline Information					
12	Service Priorities				X	
13	Service Messages					
14	Additional Service Information					

Conformance to Recommendation Description (RD) requirements

RD		ETRI	RAI	KETI	TSP	INSIGNAL
1	Representation of Usage Roles of UD, CD and SD elements		X		X	
2	Objects Classification Representation		O		X	
3	History of Acceptance		O		X	
4	Representation of Objects Equivalence		O			
5	Process Representation				X	
6	Representation of objects relationships		O			

Conclusion

General		ETRI	RAI	KETI	TSP	INSIGNAL	
1	Structure	O	O	O	O	O	
2	Extensibility	O	O	O	O	O	
3	Security & Privacy & Traceability						
4	Access, Modularity, and Adaptivity	X	X	X	X	X	At this stage of the development all proposals can be compatible with this requirement. However, it cannot be technically assessed.
5	Licensing						
6	Priorities				X		Just placeholders defined, no structure.
7	Compatibility	X	X	X	X	X	All proposals are compatible w.r.t. this requirement, but the level of maturity may vary.

We have found that by combining/merging multiple proposals, most of the requirements are fulfilled. The remaining requirements will be fulfilled by adding more details to the technologies under consideration, during the development of the standard.

The following table summarizes the conformance to the requirements related to UD, CD, SD and RD.

Summary	ETRI	RAI	KETI	TSP	INSIGNAL
Data formats proposed	**UD, CD, SD**	**UD, RD**	**UD, CD**	**UD, CD, SD, RD**	**UD, CD, SD**
n. of completely satisfied Requirements (o)	**1** **UD(1)**	**7** **UD(3)+RD(4)**	**19** **UD(14)+CD(5)**	**1** **UD(1)**	**1** **UD(1)**
n. of partially satisfied Requirements (x)	**20** **UD(5)+CD(10)+ SD(5)**	**2** **UD(1)+RD(1)**	**9** **UD(5)+RD(4)**	**20** **UD(6)+CD(5)+ SD(5)+RD(4)**	**15** **UD(5)+CD(4)+ SD(6)**
Total (o+x)	**21**	**9**	**28**	**21**	**16**

References

1. ISO/IEC JTC1/SC29/WG11/N13879 "Call for Proposals on MPEG-UD", Vienna, AT, August, 2013
2. ISO/IEC JTC1/SC29/WG11/N13881 "Requirements on MPEG-UD", Vienna, AT, August, 2013
3. ISO/IEC JTC1/SC29/WG11/N13880 "Use Cases for MPEG-UD", Vienna, AT, August 2013

Output documents

N13994 - Technologies under Consideration (TuC) of User Description