121st MPEG Gwangju, Rep. of Korea, 22 - 26 January 2018, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 Network-Based Media Processing

MPEG has developed various technologies for multimedia coding and transport, such as AVC/HEVC, 3D audio, MPEG-2 TS, ISOBMFF, DASH, and MMT. These technologies have been widely adopted and are heavily used by various industries in various applications, such as digital broadcasting, audio, and video streaming over the Internet, in mobile terminals, etc.

In order to develop standardized and efficient solutions for network-based media processing (NBMP), especially given the recent increase in demand for distribution of MPEG media in next generation network environments such as 5G, MPEG evaluates and addresses the current limitations of available standards in the MPEG media distribution area including taking considerations of processing units in networks and challenges in emerging network environments into account.

1. Interfaces for media processing functions in networks/cloud
2. Supplementary information for media processing
3. A format for standardization of chaining and composition of network based media processing functions
4. NBMP will also provide API to media processing
5. These specifications will together compose the NBMP framework

To enable interoperable distributed media processing in these networks and cloud environments, NBMP defines interfaces between media processing functions. These interfaces will allow configuration of the media processing functions and efficient exchange of media data. Second, these interfaces will facilitate a better tradeoff between resources (e.g., bandwidth, compute, storage) in this environment by having supplementary information useful for the media processing functions. Further, NBMP will define formats for chaining and composing such media processing functions and provide an API for the media processing. These functionalities together compose the NBMP framework for interoperable network-based media processing.

NBMP is a framework that allows service providers and end users to describe media processing operations that are to be performed by the network. NBMP describes the composition of network-based media processing services out of a set of network-based media processing functions and makes these network-based media processing services accessible through Application Programming Interfaces (APIs).

An NBMP media processing entity performs media processing tasks on the input media data and the related metadata. NBMP also provides Control Functions that are used to compose and configure the media processing. In addition, NBMP provides transport methods for communication between media source and media processing entities.

Figure 1 depicts the NBMP architecture that will be used as a reference architecture to scope the NBMP work.

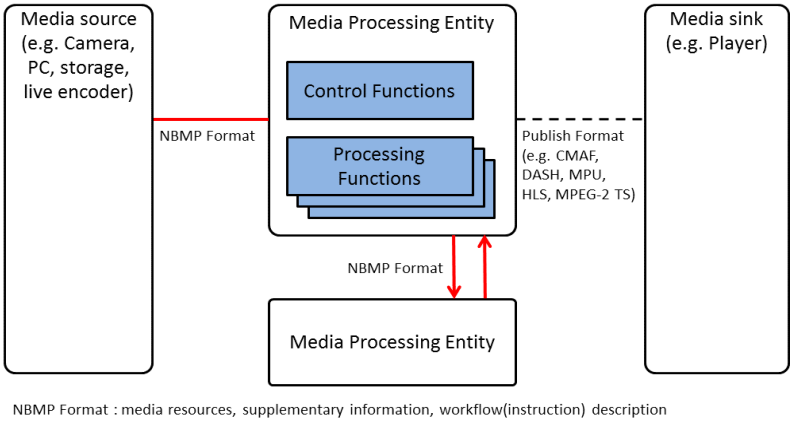


Figure 1 - Potential framework for Network-Based Media Processing system

The definition of NBMP format such as media resources, supplementary information, workflow(instruction) description will be described with respect to the following:

1. **Media Resources (M)**
   - ✓ Audio/ Visual (raw data) bit-stream, compressed stream, packaged stream
   - ✓ Encoded data, packaged data
2. **Supplementary information (S)**
   - ✓ Metadata
   - ✓ auxiliary information (side information)
3. **Workflow description (describes the ingestion, processing steps, distribution/ end-to-end procedures, wish list from content provider) (W)**
   - ✓ Java script
   - ✓ server manifest
   - ✓ Instruction for media processing
   - ✓ QoS requirement descriptor
   - ✓ QoS/QoE feedback management

The NBMP framework consists of two generic functions inside media processing entities: Processing Functions, and Control Functions.

1. **Processing Functions:** provide functionalities for media processing and analysis given control instructions from the control functions. These functions are considered to generate processing output information, which includes the following functions:
   - ✓ **Media processing function:** the core media processing function which performs processing of the input media that can generate output media or metadata. Examples of media processing are; content encoding, decoding, content encryption, content conversion to HDR, content trans-multiplexing of the container format, streaming manifest generation, frame-rate or aspect ratio conversion, and content stitching, etc.
   - ✓ **Analytics function:** provides functionalities for analysis of the logged information and makes the analysis report available via request.
2. **Control Functions:** provide functionalities for controlling and management of the media processing tasks and workflows, and how input media is processed into output media published into the media sink. A key control function is the "service/workflow manager" function.
   - ✓ **Service/Workflow function:** provides functionality for the composition of media processing workflows by chaining a set of media tasks. This includes matching the output of each media processing entity to the input of the succeeding media processing entity.
   - ✓ **Monitoring Function:** provides functionalities for monitoring the processing pipeline, and to guarantee media processing task execution, correctness, or to detect failures during media processing.
   - ✓ **Logging Function:** logs information about media processing and/or services
   - ✓ **Pub/Sub Function:** provides functionality for messaging and information exchange, including content retrieval and publishing. This may be used to trigger processing to start at one media processing entity after processing ends in a previous media processing entity.
   - ✓ **Security functions:** provides functionalities for ensuring user's and content security, including user authentication, content encryption, and other functionalities, as necessary.

Use cases

1. Use case: Network-assisted media quality enhancement
   - Network-assisted VR stitching
   - Network-assisted video up-scaling
   - Mobile edge encoding for the adaptive streaming
   - Content-aware cloud transcoding
   - Cloud-based 360 VR stitching
   - Cloud-based multiple capture device for VR streaming
2. Use case: Network-assisted media distribution
   - Live Media Ingestion
   - Online transcoding for media-aware caching
   - Coded caching for media distribution
   - Targeting caching for media distribution
   - Large size caching
   - Multicast ABR streaming
3. Use case: Network-assisted media composition
   - Augmented Video streaming
   - Intelligent user-centric broadcasting
   - Interactive media services
   - Customized media composition
4. Use case: Immersive media handling
   - Network aggregated point cloud media
   - Network pre-rendering
5. Additional use cases
   - Media processing for vehicles

Output Documents
N17327 - Use cases and requirements for NBMP (v3)
N17328 - Draft Call for Proposals on Network-Based Media Processing

2 Workplan on MPEG-I Audio Evaluation Platform

The development of evaluation platform addressed in [N17177] is work in progress. Joint collaboration is needed within the Audio Subgroup to define and create the proper testing platform and choose the most suitable evaluation concept supporting all needs of the MPEG-I 6DoF Audio development and evaluation process.

The following table summarizes the set of tasks planed for the period between the 121st and 122nd MPEG meetings. The tasks formulated in the MPEG120 workplan [N17254] are extended by the set of new ones resulted from the inputs and discussions during the MPEG121 meeting. The set of tasks categorized into the following sub-sections:

- Evaluation platform software development tasks
- Evaluation concept development tasks

The first category includes the tasks related to development of the audio testing platform software. The second one contains the tasks related to justification of the testing methodology, configuration of the evaluation platform and other aspects.

Timetable

#	Description	Due
	MPEG120	
	Evaluation platform SW development tasks	
1.1	Access to evaluation platform (+ licensing statements)	Ongoing
1.2	Assessment of evaluation platform	Done
1.3	Switching between renderers	Done
1.4	Multiple VST plugins	Done
1.5	A/V Synchronization	Done
1.6	MUSHRA-VR GUI	Done
1.7	Loop functionality	Ongoing
1.8	Recording user actions and playback functions	2018-03-16
1.9	Recording function for the VR presentation	2018-03-16
	Evaluation concept development tasks	

2.1	VR treadmill exploration	Done
2.2	“Off-line” evaluation option	Done
2.3	ADM format specification	Done
MPEG121		
	Evaluation platform SW development tasks	
3.1	Setup and access to the Git repository	2018-03-02
3.2	Licensing statement	2018-03-02
3.3	Support of VST3 plugins	MPEG122
3.4	Latency of uniOSC	MPEG122
3.5	General improvement of the SW and documentation	-
	Evaluation concept development tasks	
3.6	Test using the MUSHRA-VR 6DoF Evaluation Platform and using the regular MUSHRA approach	MPEG122
4	Exploration of testing concepts, the following can optionally be studied and contributions are welcomed: • "limited body motion" • "audio-only" • "AR" • "multi-dimensional" MUSHRA • "off-line" • etc.	MPEG122

MPEG-I 6DoF Audio Test Material
In order to support the MPEG-I 6DoF Audio effort, there is a need for test material suitable for assessing AR/VR coding and rendering technology. A first use of such material is to check that proposed MPEG-I 6DoF evaluation procedures and envisioned tests permit effective discrimination between various technologies under test. Such study may be a means to determine a final content format.

Re-use of MPEG-H 3D Audio Test Material
A first step to the creation of object-based VR content could be expressed as follows:
Rather than starting from scratch, it would be possible to build on top of the high-quality object-based content that was used for MPEG-H 3DA verification testing. These are object signals and static or dynamic rendering metadata which construct these scenes.

At the time of MPEG-H 3DA there was no notion of object distance or the user moving around in the scene. Instead, the distance of an object can be considered implicit as being equal to the radius of the loudspeaker sphere (and the size of the reproduction room) that was used for reproduction (say, r=1). In MPEG-I 6DoF, distance needs to be made explicit in order to position the objects within the reproduction scene in which the user is able to move around. Thus, the original content describes VR scenes where all objects sit at the default position, i.e. on the surface of the sphere which sounds the same as in original MPEG-H presentation when the listener is in the original (middle) listening position. This may be sufficient to properly represent sounds at the nominal distance. As a next step, the object positions could be adjusted by setting their radius to something within the whole sphere (i.e. r<1) such that the user could walk around them and possibly create some visual counterpart at this location. This is getting one step further towards 6DoF.

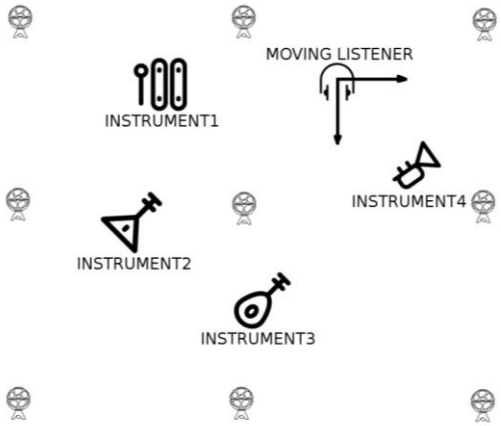
Beyond that, there are further refinements that should follow, like integrating radiation directivity or to include environmental acoustic properties to realize the full potential of 6DoF AR/VR.

HOA content for 3D audio scene
The goal is to capture the sound field across the fairly large recording scene. Zylia will use Zylia ZM-1 spherical microphone arrays, which are capable of recording up to 3rd order HOA. Potential issues to be reconsidered before the recording session are listed below:

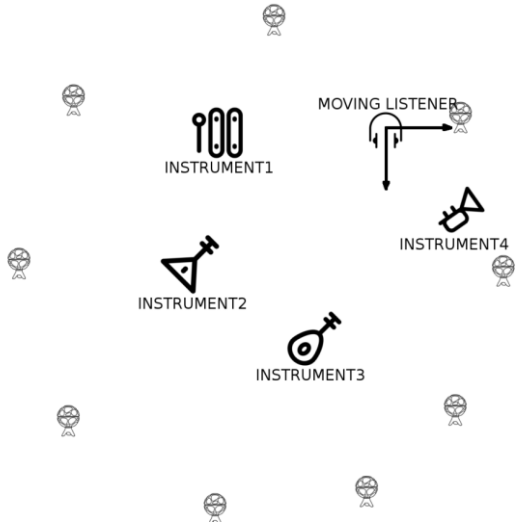
- Optimal situation would be to distribute microphones as dense as possible. This would allow to obtain better sound field approximation between microphones and to reduce the spatial aliasing effect.
- Installation – precise positioning of microphones is crucial. The microphones would be placed on the equal height of approx. 1.5 m.
- Planar layout of the microphones may result in less precise sound image across the Z axis.

Zylia proposes to use nine ZM-1 microphones distributed in two different layouts. The recording material will be short 3-4 minute performances with 4 instruments of different kinds: string, wind etc. All recordings will be synchronized in time. The proposed microphone layouts are as follows:

- **Layout 1** - equal distance planar microphone distribution at ~1.5 m height. Distance between consecutive microphones would be up to 2 m to cover performance space - square of area equal up to 16 m^2.



- **Layout 2** - circular planar microphone distribution at ~1.5 m height with the same square area of 16 m^2.



The recorded material will be provided in two audio formats:

- 9 x 19-channel raw WAVE, 24 bit, 48000 Hz.
- 9 x 16-channel WAVE (HOA 3rd order), 24 bit, 48000 Hz.

Audio Objects

Audio Research Labs (ARL) will work with New York University (NYU) Steinhart School of Music to record a number of individual instruments and voices. Recording will be conducted in the NYU recording studio using the studio's microphones and Digital Audio Workstation.

It is proposed to record the following audio objects:

- String instrument, e.g. violin
- Brass instrument, e.g. trumpet
- Speaking voice

Each audio object will be recorded in 8 channels, where microphones are placed at equidistant angles in the horizontal plane at a height so as to achieve the best capture of the sound. In other words, microphones are places every 45 degrees and aimed toward the audio object.

It is anticipated that the 8 channels will capture sufficient details of the radiation pattern of the audio object to permit determination of a parametric model of the object's radiation pattern. In this way one of the microphone signals plus the model of directivity could be used as test material for MPEG-I Audio work.

Output Documents

N17443 - Workplan on MPEG-I Audio Evaluation Platform
N17444 - Workplan on MPEG-I Audio Test Material
N17445 - Thoughts on MPEG-I Audio Requirements

3 MVCO Extensions on Time-Segments and Multi-Track Audio

MPEG-21 'Media value Chain Ontology (MVCO) Extensions on Time-Segments and Multi-Track Audio' in terms of specification (ISO/IEC 21000-19:2010 FDAM 1) has reached Final Draft Amendment (FDAM) stage at the 120th MPEG meeting, Macau, CN, 23 - 27 Oct. 2017. Whilst its associated reference software (ISO/IEC 21000-8:2008 AMD 4) reached Amendment (AMD) stage at the 121st MPEG meeting, Gwangju, KR, 22 - 26 Jan. 2018. Both are expected to be published as ISO/IEC international standards in the next few months.

MVCO facilitates rights tracking for fair and transparent royalties payment by capturing user roles and their permissible actions on a particular IP entity. However, widespread adoption of interactive music services (remixing, karaoke and collaborative music creation) - thanks to MPEG-A: Interactive Music Application Format (ISO/IEC 23000-12) - raises the issue of rights monitoring when reuse of audio IP entities is involved, such as, tracks or even segments of them in new derivative works. This amendment addresses this issue by extending MVCO functionality related to description of composite IP entities in the audio domain, whereby the components of a given IP entity can be located in time, and for the case of multi-track audio, associated with specific tracks. The introduction of an additional 'reuse' action enables querying and granting permissions for the reuse of existing IP entities in order to create new derivative composite IP entities.

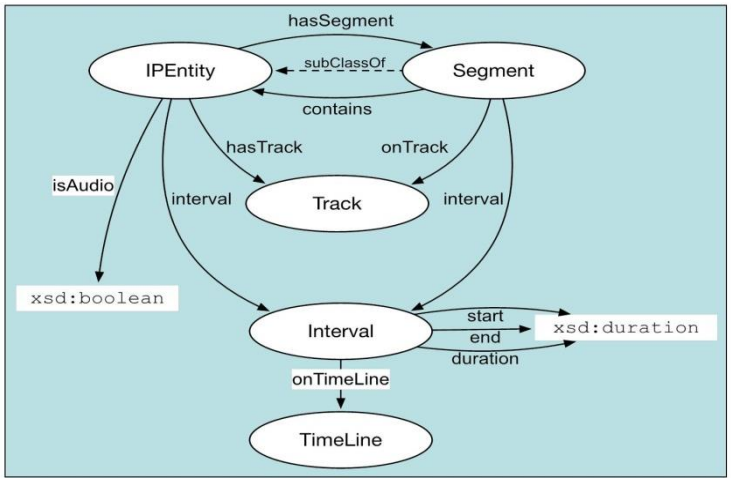


Figure 2 - Classes and relationships for the representation of IPEntity that contain other existing IP Entities. Segments may also be associated with individual Tracks of a Multi-track IP Entity.

With qMedia::C4DM proposed MVCO Extensions on Time-Segments and Multi-Track Audio, it is now possible to query for information about user collectives and the components of the composite IP Entities.

List members of a user collective:

```
$ java -jar rvdac.jar -r -lcu Performers
RVD Administration Console
Guitarist
Vocalist
```

List components of a composite IP Entity (including locations specified by segment and tracks where applicable):

```
$ java -jar rvdac.jar -r -lic MusicInstance
RVD Administration Console
LyricsInstance | segment: 30s to 150s | track: 2
GuitarInstance | track: 1
```

Further information on this particular use case can be found in the MVCO Extensions on Time-Segments and Multi-Track Audio Guidelines Document.

Output documents

121st MPEG meeting, Gwangju, KR, 22 - 26 Jan. 2018
N17403 - DoC on ISO/IEC 21000-8:2008/DAM 4 MVCO Extensions on time-segments and multi-track audio
N17404 - Text of ISO/IEC 21000-8:2008/AMD 4 MVCO Extensions on time-segments and multi-track audio

120th MPEG meeting, Macau, CN, 23 - 27 Oct. 2017
N17169 - DoC on ISO/IEC 21000-19:2010/DAM 1.2 Extensions on time-segments and multi-track audio
N17170 - Text of ISO/IEC 21000-19:2010/FDAM 1 Extensions on time-segments and multi-track audio