110th MPEG Strasbourg, France, 20 - 24 October 2014, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

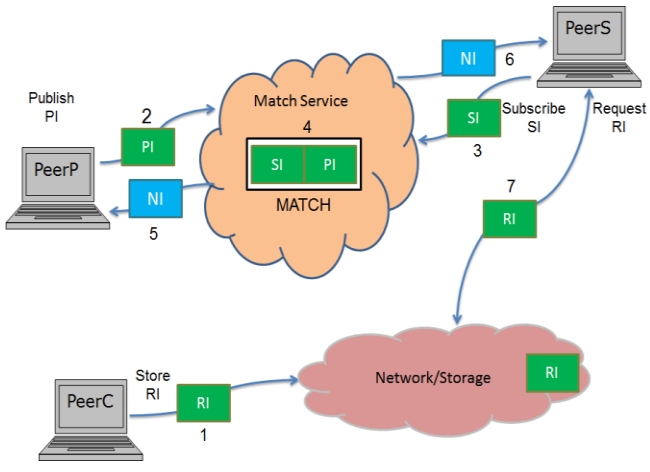
Contents

1 Publish/Subscribe Application Format (PSAF)

PSAF (ISO/IEC 23000-16) remained at Working Draft (WD) status at 110th MPEG meeting albeit progress took place in collaboration with MCO/CEL & MPEG-UD groups.

Publish/Subscrib is an established communication paradigm where senders do not communicate information directly to intended receivers but rely instead on a service that mediates the relationship between senders and receivers. In Publish/Subscribe model senders (called Publishers) post information on and receivers (called Subscribers) declare their interest in a certain type of information – before or after a publication – to a service. The Publish/Subscribe communication paradigm, adapted to the needs of media creation, publication and retrieval, can be represented by the workflow and the figure below (note that the order in which creation, publication and subscription is not relevant).

- Creator uses PeerC to
  - Store Resource
- Publisher uses PeerP to
  - Create RI
  - Store RI
  - Create PI
  - Publish PI
- Subscriber uses PeerS to
  - Create SI
  - Subscribe SI
- Match Service Provider
  - Finds Match between SI and PI
  - Notifies Subscriber and Publisher
- Subscriber uses PeerS to
  - Retrieve RI
  - Retrieve Resource



Terminology: Resource Information (RI), Publication Information (PI), Subscription Information (SI) and Notification Information (NI).

The RI, PI and SI data sets contain information that is relevant to the media context, namely:

Resource Information (RI) contains

- Resource Locator or Name the RI refers to
- Metadata describing the Resource
- RI creator's ID
- Contract between Publisher and eventual Resource consumer that includes
    - IDs of parties
    - Locator/Name of Resource the contract refers to
    - Rights to use
    - Conditions to use
- ID of users to be notified that Resource has been used
- RI ID
- Signature of User who creates RI

Publication Information (PI) contains

- RI ID the PI refers to
- Metadata for the purpose of Resource publication
- Metadata describing Publisher and Context (e.g. expressed in MPEG-UD)
- Publisher ID
- Contract between Publisher and Match Service Provider that includes
    - Permission to
        - Match PI with SIs
        - Use Query, Metadata describing Publisher and Context etc.
    - Condition on Subscriber ID
    - Duration of right
    - Match Service ID(s)
- ID of users who shall be notified that a match has been found
- PI ID
- Publisher Signature

Subscription Information (SI) contains

- Query
- Subscriber ID
- Metadata describing Subscriber and Context (e.g. expressed in MPEG-UD)
- Contract between Subscriber and Match Service Provider that includes
    - Permission to
        - Match SI with PIs
        - Use Query, Metadata describing Subscriber and Context etc.
    - Condition on Subscriber ID
    - Duration of right
    - Match Service ID(s)
- ID of users who shall be notified that a match has been found
- SI ID
- Subscriber Signature

Notification Information (NI) contains

- RI ID
- PI ID
- SI ID
- MSP ID
- NI ID
- Match Service Provider (MSP) Signature

Figure 1 depicts an exemplary use of PSAF.

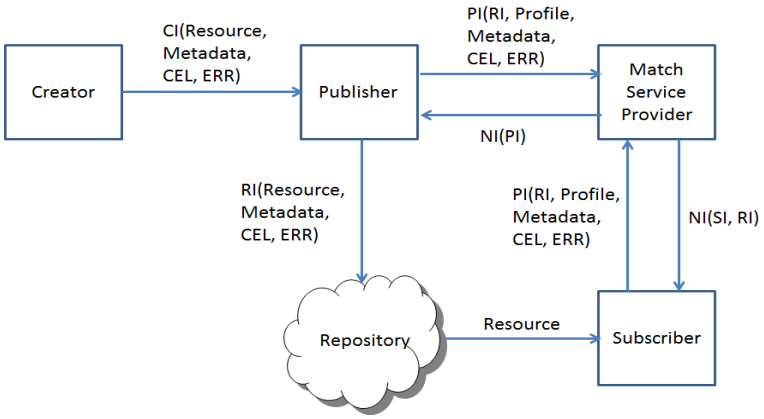


Figure 1. Relationships among PSAF entities.

The RI, PI and SI data sets and their components are expressed using MPEG-21 technologies such as:

1. Digital Item Definition Language (DIDL)
2. Digital Item Identification (DII)
3. Contract Expression Language (CEL)
4. Event Reporting (ER)

PI, RI and SI can be digitally represented as Digital Items using DIDL. The 3 Digital Items have a single ITEM that coincides with one of the 3 principal elements of PSAF (PI, RI and SI). Each ITEM has an identifier that is an element dii:identifier or a descriptor. Each ITEM is enriched with metadata of type PSAFMetadata that are specific of the 3 types of element, in particular they are metadata describing the resource (those of PI and RI) and a MPEG Query Format query (those of SI). PI and RI must contain a resource: in the former case the resource is a RI referenced by using a RI Identifier, while in the latter the resource is the resource that is the target of an eventual consumption, e.g. the URL of a media item. Digital Items and their elements are indentified using DII.

CEL is used to digitally represent contracts between Publisher and MSP, as well as, between Subscriber and MSP. ER is used to digitally represent requests to MSP to issue notifications to users. NI is based on Event Report and may benefit from Recommendation Information of MPEG-UD.

In particular, with respect to PSAF "Match" requirement CEL/MCO model amended as shown in Figure 2.

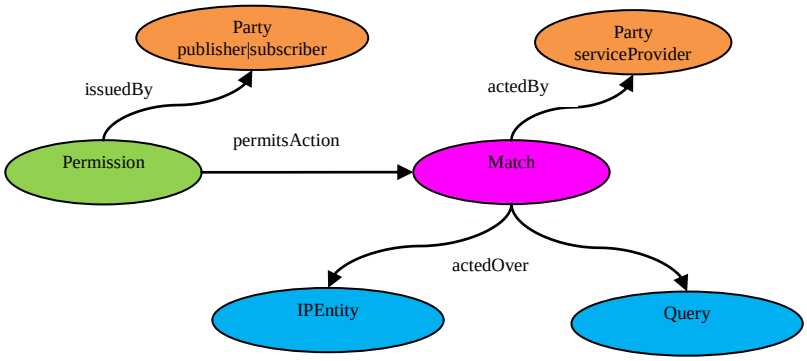


Figure 2. Extended CEL/MCO model with 'Match" act.

This is implemented in MCO by adding new classes, as:

— Match – as subclass of mvco:Action, models the action of matching queries and IP-Entities, performed for providing content indexing and/or replies to user queries in a context in which a service provider holds content material, made available by and among parties;

— Query – as subclass of mvco:IPEntity, is the expression of the desired properties of some content, that may be submitted in order to have indications about content matching such properties and at which extent.

Equivalent elements should be defined for CEL.

Furthermore, the MPEG-UD group has been contacted a study on its use cases with respect to a reference methodology adopted for consistency and effectiveness of recommendations and related schemas harmonisation. PSAF has been discussed in a joint meeting with MPEG-UD group and it has been added among its use cases for further study.

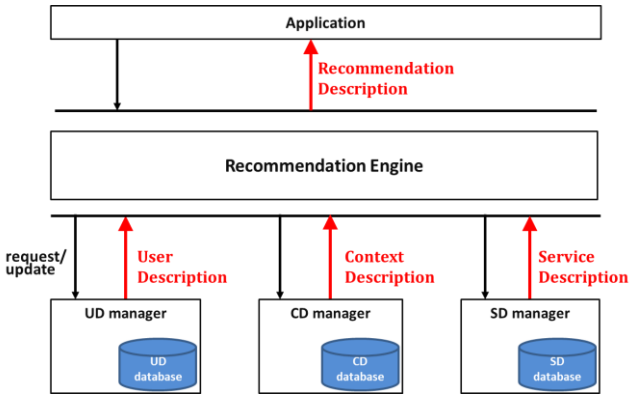


Figure 3. Conceptual Model of MPEG-UD

MPEG-UD is aimed at describing tailored "recommendations" concerning specific actions within the domain of a given application. It is believed that a complete and compact description of these recommendations allows for facilitating the application by matching the user needs. However, the logic behind how the recommendations are produced by recommendation engines is out of scope of the standard. With regards to the aforementioned study, by following a *bottom-up* approach, the starting point is given by considering a specific use case. According to it, the core "event" is identified thus classifying, in a such a way, the use case. Examples of use cases are the "fruition" of multimedia contents, the "sale" of generic products and the "purchase" of books where "fruition", "sale" and "purchase" can be recognized, respectively, as events. The recognition of a core event driving the process is fundamental because with respect to it the "recommendation" is made. The next step is represented by making all relevant information explicit, namely the "entities" and their relative "attributes". Attributes represent relevant information concerning the involved entities. The amount and the quality (for instance in terms of granularity) of attributes can strongly impact on the effectiveness of the recommendation.

Output documents
N14848 - WD of ISO/IEC 23000-16 Publish/Subscribe Application Format
N14843 - WD of ISO/IEC 21000-20 Contract Expression Language
N14844 - WD of ISO/IEC 21000-21 Media Contract Ontology
N14888 - WD of MPEG User Description
N14866 - Draft implementation guidelines for MPEG-UD

2 Media Linking Application Format (MLAF)

The notion of *bridget*

A *bridget* is a link from the programme being watched to external interactive media elements such as web pages, images, audio clips, different types of video (2D, multi-view, with depth information, free viewpoint) and synthetic 3D models. A *bridget* can be just a link from a portion of a "source" programme to a single media item but also a series of links from a collection of source programme components (images and video clips) to a set of destination media. More generally a *bridget* can be a collection of links from a set of source media items to a set of destination media items, e.g.

- An image points to an image or a set of images
- An object in an image points to an object in an image or to an image or to a set of images
- A slide part of a slide show points to audio clip
- An audio clip in an audio points to the corresponding score sheet
- Different images drawn from a programme point to different web pages
- A video clip from a video points to a set of related videos

A *bridget*, however, is not simply a URL, but contains two data structures: one related to the source media item and the other related to the destination media item. It may also contain information on how the *bridget* itself should be presented to the user.

Typical *bridget* workflow

Bridgets are links which exist because of some inherent semantic relationship between content items. As such, they can be products of an editorial decision, taken by someone as the result of the inspection (which can be manual or automatic) of content items, and can be objects of a workflow which involves different roles taking care of finding, organising and finally crafting the data that constitute them. The nature of a *bridget* is however quite different than traditional linearity of media content, and as such it induces a different, more "distributed" workflow. In fact, whether a piece of media content is a candidate source or destination for a *bridget* can be the result of an editorial decision taken at any moment and by quite different kind of users.

What is foreseeable is a sort of "layered" approach at producing *bridgets* in which actors with different roles defines *bridgets* under different perspectives and possibly concurring at the same time. Authors of programmes will define *bridget* end points (i.e. sources and destination content items) following criteria matching with the editorial intention, main distribution channel or target audience of the programme. At the same time marketing and commercial operators (e.g., advertisement agents) will define such end points following their own mind-setting, which may be independent from the authorial perspective. Last, but definitely not least, final users can define their own ways for *bridgets* through social media interaction. All the above approaches can include not only the generation of the linking information but also of information related to how referenced content have to be presented graphically or should interact with the user.

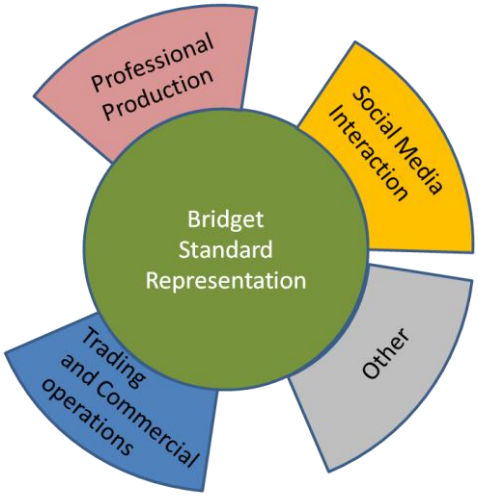


Figure 4. Bridget creation workflow.

The result is that the way in which *bridget* information would be created is quite different from traditional linear approaches to media production, and more following a daisy chain paradigm (see Figure 4).

Following this vision, as a natural consequence we see the prominent need to have a standard format for representing and exchanging bridget-related information in order to integrate all the systems that, in the different and disparate aforementioned domains, will have a role in generating *bridget* information.

Requirements
An initial set of requirements, that such a media linking application format implementing the *bridget* concept shall satisfy have been identified. These are subdivided in six main categories:

1. Source production data
    - Identification metadata
    - Technical metadata
2. Source content data
    - Access metadata
    - Descriptive metadata
3. Link related data
    - Contextual metadata (source-destination relationship)
    - Bridget production metadata
    - Segment-based linking
4. Destination content data
    - Access metadata
    - Descriptive metadata
5. Presentation data
    - Scene description
    - Media
6. Identify and analyse real signals
    - Local processing
    - Remote processing

Enabling Technologies
- MPEG-21 Digital Item Declaration
- MPEG-7 Audio Visual Description Profile (AVDP)
- EBU Core Metadata Set
- MPEG-4 Augmented Reality Application Format (ARAF)

Output documents
N15049 - Draft Requirements for Media Linking Application Format (MLAF)
N15061 - WD of ISO/IEC 23000-18 Media Linking Application Format

3 Multiple Sensorial Media Application Format

Conventional 2D/3D Media services are based on audiovisual contents that are presented by display devices and speakers. As users want more realistic experiences from media contents, they could experience stereoscopic video, virtual reality, 3-dimensional TV, multi-channel audio, etc. These media services are limited in ‘visual’ and ‘audio’ aspects. From a rich media perspective, realistic media coupled/assisted with their target devices are very beneficial to users because their experiences on media consumption can be greatly enhanced. Business markets can be enriched by coupling media and devices. Multiple Sensorial Media services use multiple devices to enhance realistic experiences, which can simulate the effect of wind, flash, tickle, and more and even adjust the intensity according to users’ preferences.

Authoring Multiple Sensorial Media is not easy. It takes a lot of time to add real sensory effects and produce a Multiple Sensorial Media content. Without a common container, each operator needs to do extra work to display Multiple Sensorial Media in synchronization with the sensory effects. With this Multiple Sensorial Media Application Format (AF), service providers can develop various real sensory effect media services and vitalize popularization of Multiple Sensorial Media. Users can also access various Multiple Sensorial Media contents easily and experience realistic media.

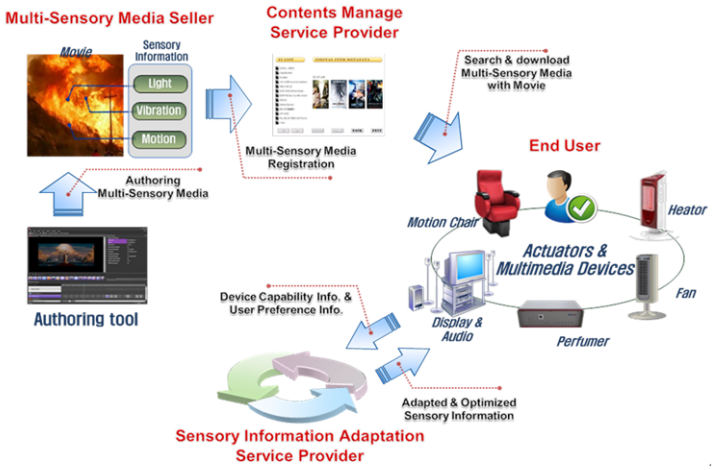


Figure 5. Multiple Sensorial Media Service Environment

Specifying a media file format (Multiple Sensorial Media Application Format) through which audio-visual contents are tightly coupled with sensory information would greatly promote the wide usage of MPEG-V standard as well as the Multiple Sensorial Media service based on the MPEG-V standard.

The Multiple Sensorial Media AF will be based on the following technologies:

- ISO Base Media File Format
- Video: MPEG Video Formats (MPEG-2, MPEG-4 AVC/H.264), MJPEG
- Audio: MPEG Audio Formats (MPEG-1 Layer-3, MPEG-2 AAC+, MPEG-4 ER-BSAC, MPEG-4 HE-AAC)
- Sensory effect metadata: MPEG-V
- Content description metadata: MPEG-7, TV-Anytime etc.
- XML Streaming Instruction: MPEG-21 XSI
- Protection and Governance information: MPEG IPMP, REL, etc.

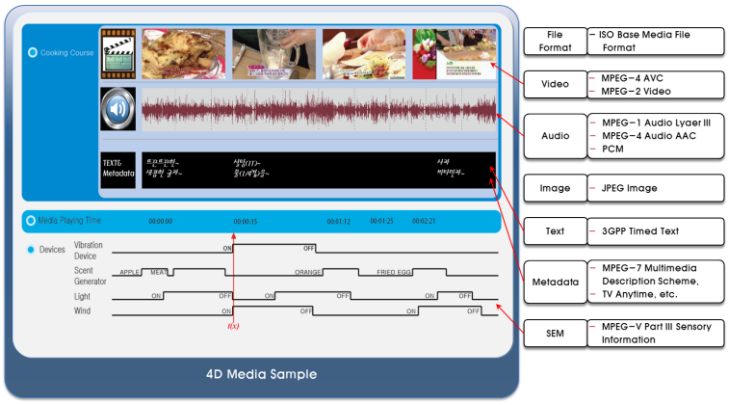


Figure 6. Base technology for the Multiple Sensorial Media AF

Output documents

N15004 WD 2.0 of Multi Sensorial Media Application Format

4 Exploration - Media-centric Internet of Things (IoT)

There is an increasing interest from the industry in the Internet of Things (IoT) technologies. There are active standardization activities to define network protocols for the Internet of Things (e.g., how to connect things). The variety and heterogeneity of "Things" make difficult to standardize descriptions, data formats, APIs in a global manner, however, when the environment is well established, this can be done. *Therefore, MPEG is exploring representations of media things as part of complex distributed systems implying interaction between things and between humans and things.*

Since 2008, MPEG has been working on a standard called *MPEG-V* (ISO/IEC 23005) which aims to define the data exchange interface between real and virtual worlds. In particular it defines the XML schema for the information about sensor/actuator capabilities, user preferences, sensed information and data types used by actuator commands. These data type elements are corresponding to descriptions of devices and messages for "talking to" and "adapting to" either devices or services in the Internet of Things.

The Media-centric Internet of Things is the collection of interfaces, protocols and associated media-related information representations that enable advanced services and applications based on human to device and device to device interaction in physical and virtual environments. Information refers to data sensed and processed by a device, and/or communicated to a human or another device.

Media-centric IoT and relationship to MPEG-V

Figure 7 shows the overall data transition in a home automation scenario using MPEG-V. This scenario was chosen because home automation is one of the major targets of the Internet of Things applications. Sensors in a home (e.g., temperature, light, humidity, proximity, gas, and camera sensors) can detect context information (i.e., sensed information) in the home in real time and send this information to a user's mobile phone or a control box with intelligent software. Each sensor identity and its capabilities can be specified by sensor capability descriptions. The sensed data can be handled and analyzed either automatically by an intelligent engine or semi-automatically by users with a mobile phone. Users can determine the actuation level of actuators at home using pre-defined actuation preferences. They can also filter or control the level of sensed data using pre-defined sensor adaptation preferences. The identity and capability of each actuator at home can be defined by actuator capability descriptions. The intelligent software in the control box, the mobile phone or even in the sensor can generate actuator commands specifying how to actuate the devices at home by combining all of the input information (i.e., sensed information, sensor capabilities, actuation preferences, sensor adaptation preferences, and actuator capabilities).

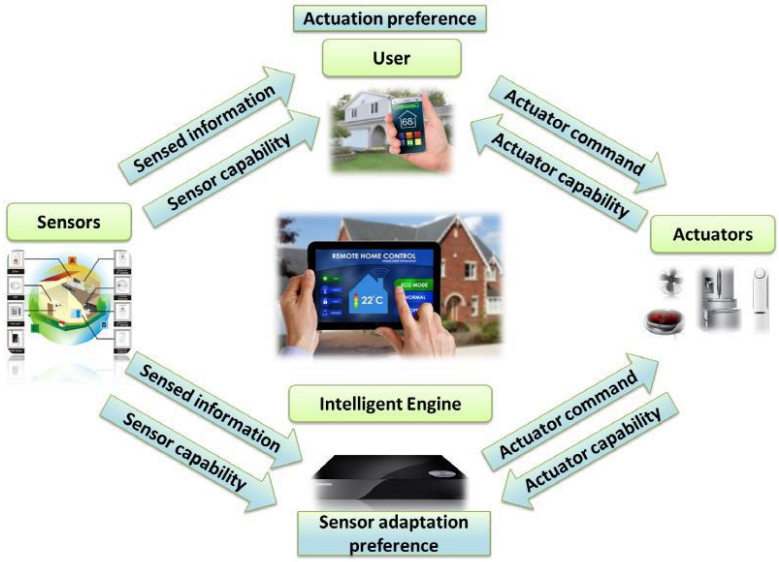


Figure 7. Data transition using MPEG-V standards in a home automation scenario.

The specification of “Things” (e.g., sensors and actuators) can be defined with the sensor capabilities and the actuator capabilities in MPEG-V. Interaction between “Things” can be achieved using the specifications of sensed information and actuator commands. Auxiliary information such as actuation preferences and sensor adaptation preferences in MPEG-V enable more intelligent Internet of Things services. Finally, the VR adaptation engine correspond to the “Intelligence” of the Internet of Things.

Output documents
N15030 - Exploration on Media-centric Internet of Things

5 Exploration - MPEG Wearable Application Format

Wearable technology is related to both the field of ubiquitous computing and the history and development of wearable computers. With ubiquitous computing, wearable technology share the vision of interweaving technology into the everyday life, of making technology pervasive and interaction friction less. Through the history and development of wearable computing, this vision has been both contrasted and affirmed. Affirmed through the multiple projects directed at either enhancing or extending functionality of clothing, and as contrast, most notably through concept of surveillance. The history of wearable technology is influenced by both of these responses to the vision of ubiquitous computing. A number of wearable technology use cases presented:

- Multimedia communication of Wearable device (smart glasses)
- Gesture recognition of Wearable device (smart glasses)
- Speech translation of Wearable device (smart watch)
- Natural language communication of Wearable device (smart watch)
- Accessibility/protection functionality of Wearable device (smart watch)
- Wearable device in cloud
- Visual Communication via Wearable Device

A possible MPEG Wearable Application Format could be based on the following technologies:

- ISO Base Media File Format
- Video : MPEG4 Simple profile based stereoscopic codec
- Video: H.264/AVC based stereoscopic codec
- Video: HEVC
- Still Image : JPEG based stereoscopic still image
- Audio: MPEG-1 Layer-2, MPEG-2 AAC+, MPEG-4 ER-BSAC MPEG-4 HE-AAC
- System: IOD/OD, BIFS, MPEG-4 File format, AVC File format,MMT.
- User metadata: MPEG-7, TV-Anytime, etc
- Stereoscopic VAF
- MPEG-UD

Output documents
N15027 - AHG on wearable MPEG

6 Workshop on media synchronisation for hybrid delivery

MPEG is currently exploring the use cases and requirements for new standards for the advanced synchronization of media, which MPEG calls Timeline Alignment. To better understand the industry development and their additional needs, MPEG has hosted a workshop including demonstrations. There were six talks from MPEG experts and external guests including:

- An overview of MPEG systems technologies providing advanced media synchronization (Samsung)
- Hybrid Broadcast - Overview of DVB TM-Companion Screens and Streams specification (TNO)
- Hybrid Broadcast-Broadband distribution for new video services : a use cases perspective (Thomson)
- HEVC and Layered HEVC for UHD deployments (Qualcomm)
- A fingerprinting-based audio synchronization technology (Sony)
- Media Orchestration from Capture to Consumption (TNO)

In addition, participants showcased six demonstrations whose titles were:

- Live augmented broadcast with user-generated content (TNO)
- Synchronised DVB-S stream with UltraHD tiled streaming on second screen (TNO)
- Hybrid Broadcast-Broadband: Alternate Camera (Telecom ParisTech)
- Hybrid Broadcast-Broadband: Timeshifting (Telecom ParisTech)
- Hybrid SHVC Delivery (INSA Rennes)
- Optimization of satellite and terrestrial broadcast resources (INSA Rennes)

MPEG has concluded to further study on the use cases and open questions presented and to continue exploration in the AhG on Timeline Alignment.

Output documents
N14881 - Presentation materials from Seminar on Media Synchronisation for Hybrid Delivery

7 Panel discussion on future video coding

MPEG together with ITU-T SG 16's VCEG hosted a panel discussion at the 110th MPEG meeting to explore use cases, requirements, and potential timelines for the development of future video coding standards. Guest speakers included:

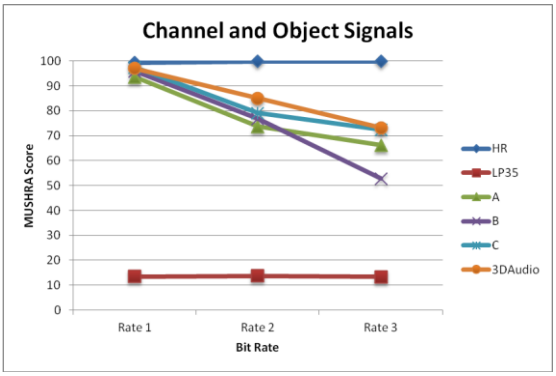
- Roger Bolton (Ericsson)
- Harald Alvestrand (Google)
- Zhong Luo (Huawei)
- Anne Aaron of (Netflix)
- Stéphane Pateux of (Orange)
- Paul Torres (Qualcomm) and
- JeongHoon Park (Samsung)

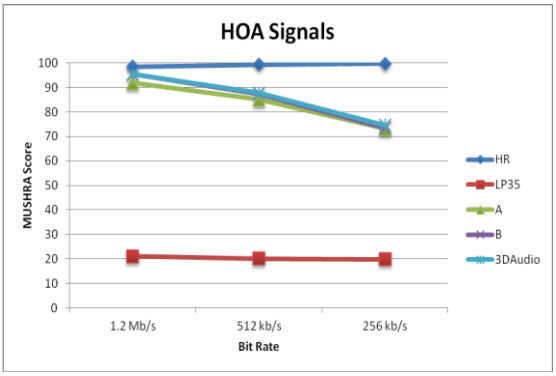
In this brainstorming session, several opinions were expressed by the panelists and other participants from a variety of industry, academia, and research institutions. Aside from confirming that maximizing compression efficiency remains a fundamental need, the expressed views were very diverse and it is evident that the needs vary among applications and industry segments. MPEG has recognized a need to further study future application requirements, and the availability of technology developments to fulfil these requirements. Toward establishing a roadmap for future video coding standardization, MPEG has therefore established two ad hoc groups to conduct this study. MPEG welcomes the input of all experts to exchange views, help map the video compression needs of the industry to future standardization activities, and discuss potential technologies that could be employed in such video coding standards.

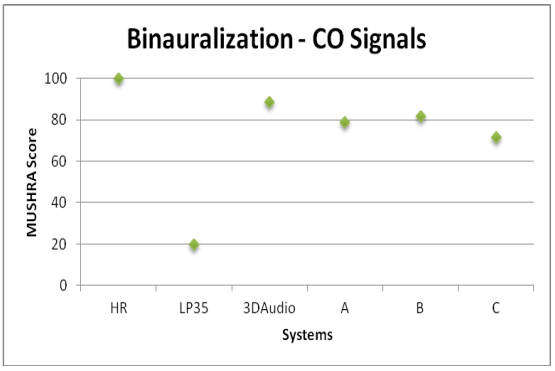
Output documents
N15050 - Presentations of the Brainstorming Session of the Future of Video Coding Standardisation

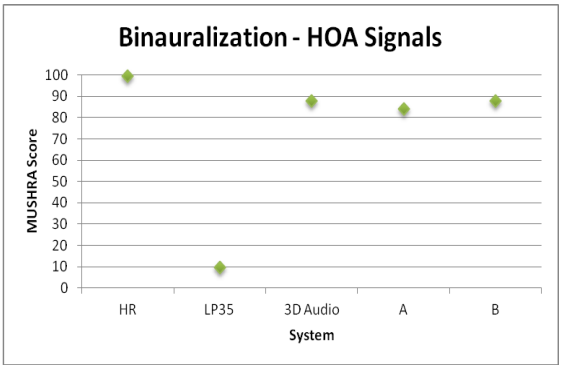
8 MPEG-H 3D Audio Performance

The Call for Proposals for 3D Audio (N13411 & N12412) was issued at the 103rd MPEG meeting held in Geneva, CH in January 2013. Submissions to the Call were evaluated at the 105th MPEG meeting held in Vienna, AT, July 2013. From the 106th to the 109th MPEG meetings, the MPEG-H 3D Audio technology was developed in a collaborative process. In July 2014 at 109th meeting, the standard progressed to DIS, at which point it was technically frozen. It is expected to progress to IS early in 2015. However, most of the collaborative work pertained to development of metadata structures, so the following results of the subjective evaluation tests (presented as mean subjective score, as averages over all test items) associated with the Call are a fair evaluation of the performance of the issued standard.









Output documents
N14928 - WD0 - 3D Audio Phase II
N14930 - Workplan on 3D Audio
N14931 - MPEG-H 3D Audio Performance Report