105th MPEG Vösendorf, Austria, 29 July - 2 August 2013, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1. White Paper on the MPEG-7 Audio Visual Description Profile (AVDP)

Summary

The intention of MPEG-7 AudioVisual Description Profile (AVDP) is to facilitate the introduction of automatic information extraction tools in media production, e.g. as web services in Service Oriented Architectures (SOA), by providing a common format for the exchange of the metadata they generate. AVDP is a profile (i.e., subset) of the MPEG-7 Multimedia Description Interface standard, targeting applications in media production and archiving, and includes all tools provided by Part 3 and Part 4 of MPEG-7 [4][6].

As a result of an extended requirements analysis mainly conducted inside EBU (European Broadcasting Union), the description tools in this profile can be used to describe the results of various kinds of media analysis such as shot/scene detection, face recognition/tracking, speech recognition, copy detection and summarization, etc. in a way that these data can be usefully integrated in media production processes. The AVDP profile supports temporal and spatial analysis of audiovisual material, including low-level audio and video descriptions. The profile defines a set of semantic constraints in order to facilitate interoperability.

For more details on MPEG-7, please refer to the MPEG-7 White Paper at http://www.chiariglione.org/mpeg.

Scope and functionality

The main scope of AVDP is describing the results of automatic media analysis with low-, mid- and high-level features for audiovisual content. Thus, the profile includes functionalities for representing results of several – (semi-) automatic or manual – feature extraction processes.

AVDP provides tools for describing:

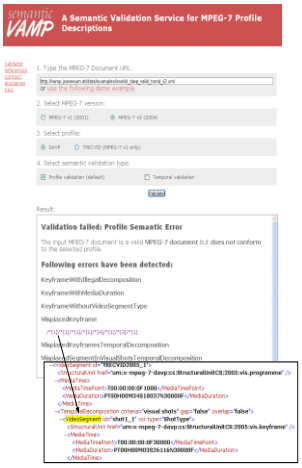
- feature extraction tool, version , contributors, and the date/time the tool applied
- several results in multiple timelines
- several results with confidence level
- various results of multimedia analysis, such as segmentation, text recognition, face detection, person identification, format detection, genre detection, keyword extraction, and speech recognition.
- results of multimedia analysis having several related descriptions of audio-visual contents, such as copy detection and summarization

Benefits

In summary, the use of AVDP provides the following benefits:

- Subset of MPEG-7 tailored to the needs of applications in media production and archiving
- Reduced complexity and improved interoperability compared to using the complete MPEG-7 standard by
    - clear separation of metadata extracted from different modalities, abstraction level and structural elements and their representation elements (e.g. key frames)
    - constraints on description tools and clear definition of their semantics (allowing formalisation and automated validation of semantic constraints)
    - mandatory use of identifiers for structural elements (decompositions, segments) and provision of classification schemes for these identifiers
- Full support for video and audio descriptors

Validation service

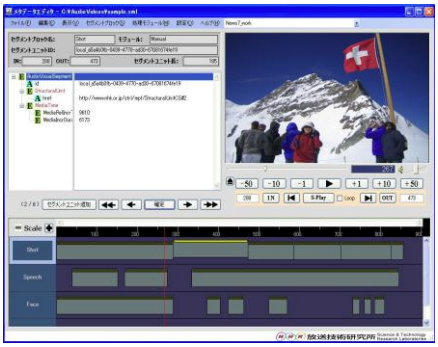


Validation of MPEG-7 documents is an important issue whenever documents are produced, exported or imported. On a syntactic level, standard tools are available for this problem, most notably XML schema validators. However, the semantic expressivity of XML schema is limited, and thus validations on higher levels cannot be done with standard tools, but need specific application logic.

VAMP (http://vamp.joanneum.at) is a service that formalises the semantic constraints in AVDP in order to automatically validate documents w.r.t. the profile definition, beyond schema validation.

VAMP is a web based application with a graphical user interface deployed at http://vamp.joanneum.at. A command line client application can be downloaded to (batch) process local files.

NHK Metadata Production Framework (MPF)



MPF is a specification proposed by NHK from 2006 that provides a common environment for the effective generation of content-based metadata for video. The MPF provides the mechanism of combining various multimedia analyses as modules for generating the desired metadata. The MPF adopts AVDP as its metadata model from version 3 on, and specifies two interface types for module control and metadata operation. Metadata Editor is part of the reference software by which user can test the basic functionality of MPF. Therefore if you follow the specified interface and develop the module with your own information extraction algorithm, you can test it on the Metadata Editor easily. The Metadata Editor and related materials are downloadable from the following site (http://www.nhk.or.jp/strl/mpf/). The Metadata Editor user interface is shown on the right. The metadata generated by the editor can be export/import as a MPEG-7 data which is compliant to AVDP specification.

Output documents

N13869 - Text of white paper on MPEG-7 Audio Visual Description Profile (AVDP)

2. Report on 3D Audio Call for Proposals

The Call for Proposals for 3D Audio (Call) [**Error! Reference source not found.**] issued at the 103rd MPEG meeting held in Geneva, CH in January 2013. Submissions to the Call were evaluated at the 105th MPEG meeting to be held in Vienna, AT, July 2013. This summary reports on the Call responses, the subjective listening test results, and the technology selected as Reference Model 0 (RM0).

Submission of processed material

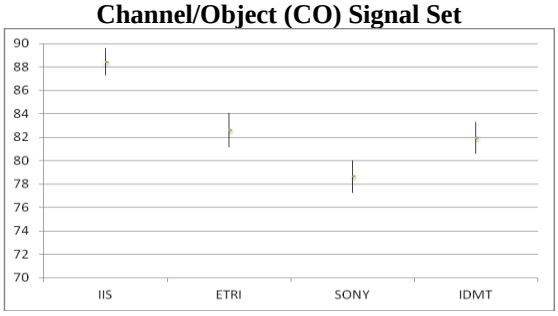
The following table lists the companies that responded to the Call. Responses could be for technology to process the Channel/Object (CO) signal set or the Higher Order Ambisonics (HOA) signal set.

Company	ID	CO	HOA
ETRI	ETRI	X	
FhG-IDMT	IDMT	X	
FhG-IIS	IIS	X	
Sony	SONY	X	
Orange	ORL		X
Qualcomm	QUAL		X
Technicolor	TECH		X

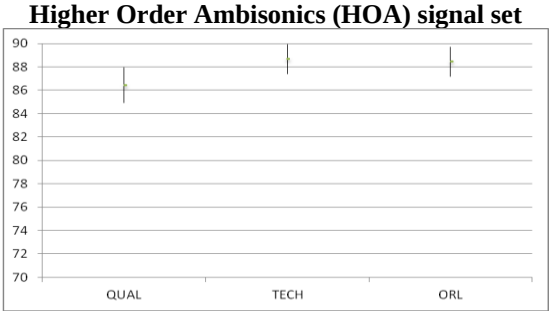
Selection of technology

The Figure of Merit was calculated for CfP responses using the CO and HOA data sets, and results are shown here:

Channel/Object (CO) Signal Set



Higher Order Ambisonics (HOA) signal set



For the Channel/Object (CO) signal set, the FhG-IIS submission (Sys1), was selected as RM0-CO.

For the Higher Order Ambisonics (HOA) signal set, the Technicolor submission (Sys2) was selected as RM0-HOA. The Technicolor submission is joint work with Orange that shares a common decoder, and could be referred to as the Technicolor/Orange submission. Thus, this Sys2 joint candidate technology from Technicolor and Orange was selected as RM0-HOA. However, it was recognized that Sys2, Qualcomm's candidate, was a very close contender, scoring within the bounds of statistical uncertainty. Based on this observation, a work plan, N13857, has been created for a fast-track core-experiment where Qualcomm will incorporate parts of their technology into the selected RM0 technology and demonstrate better performance at the 106th MPEG meeting. Based on subjective testing results, to be provided at the meeting, a decision will be made by the Audio subgroup on whether to accept the new technology as RM1 technology.

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
4. The R Project for Statistical Computing (http://www.r-project.org)

Output documents

N13854 - Report on MPEG-H 3D Audio Call for Proposals

3. Call for Proposals on Dynamic Range Control Technology

WG11 has an interest in new tools for dynamic range control (DRC) and Program Level Control (PLC). While WG11 has DRC capability integrated with the MPEG-4 Advanced Audio Coding profile coders [1], as in the AAC IMDCT-based multi-band DRC, it is interested in pursuing technology that could be more broadly applicable, be an advance on the AAC technology, and take into account recent developments in the field, including regulatory developments.

Requirements for new DRC technology are:

- Supports DRC configurations (as defined in Coding Independent Code Points 23001-8 AMD1 and MPEG-4 File Format 14496-12 PDAM3), e.g.
    - multiple DRC sets
    - different DRC for channel groups (e.g. put center in a separate group in a 5.0 channel presentation). Output loudspeaker configurations and loudspeaker index are defined in [2].
    - index of encoder DRC characteristics
- Support DRC configuration for a downmix, to be applied to the signal after downmix
- Is universally applicable (codec agnostic, codec can be a PCM format)
- Can apply DRC gain to the time domain audio signal after audio decoding is complete
- DRC processing in decoder imposed negligible additional algorithmic audio delay for single band compression
- Supports time resolution of 1ms or better for fast gain changes to meet professional requirements
- Supports gain resolution of 0.25dB or better
- Efficient gain encoding for minimizing overall bit rate increase
- If payload is embedded in a host codec's bitstream, it should maintain backwards compatibility
- Supports gain modifications, imposed by additional meta-data in bitstream: at a minimum changes in the compress and boost factors.
- Provides a loudness normalization option using loudness metadata defined in [2]
- Takes advantage of peak level signaling so as to prevent clipping
- Has low complexity at decoder side

Optional additional system functionality:

- Supports extension to multi-band with flexible number of bands and band crossover frequencies

References

1. 14496-3 MPEG-4 Audio
2. ISO/IEC 23001-8 (MPEG-B: Coding independent media description code points)

Output documents

N13858 - Call for Proposals on Dynamic Range Control Technology

N13864 - Response to USNB on Dynamic Range Control

4. Call for Proposals on MPEG User Description (MPEG-UD) - Overview and Scope

Nowadays, big data, leading to a myriad of choices, surround us. To be able to make a good and easy decision in a reasonably short time, we do need recommendations. A recommendation system can satisfy such user needs. Among all the choices that some given services offer, a recommendation system provides a set of recommendations, taking into account the user and context information.

The aim of MPEG User Description, further referred to as MPEG-UD, is to ensure interoperability among recommendation services, which take into account the user and its context when generating recommendations for the user.

MPEG-UD indents to standardize the following three data formats, as presented in Figure 1: User Description (UD), Context Description (CD), Service Description (SD), and Recommendation Description (RD).

- UD (User Description): a set of descriptions which may contain static and dynamic information about the user, including some other data like the history of the user's interactions, preferences, security settings regarding these information, etc.

- CD (Context Description): a set of descriptions that describe the environmental situation in which the user operates or is located, e.g., user's device in use, physical position, environmental variables (temperature, humidity, sound level, etc.), security settings regarding these information, etc.

- SD (Service Description): a set of descriptions containing pertinent information (including security settings) about the service (or a set of sub-services), that is offered to the end-user application, e.g. video on demand, maps, etc.

- RD (Recommendation Description) is a set of Recommended Information elements provided to the applications, in a structured, efficient, and compact form, when a customer requests a service in a certain environment. RD may include information extracted from UD, CD, SD; additional logical relations among UD/CD/SD (or their subsets) and metadata may be also included into RD. The operation of producing the RD, performed by Recommendation Engines, may have various ranges of complexity and performance, depending on the Recommendation Engine. The RD may have a general format, independent of the application.

In Figure 1, red elements indicate the formats to be specified by MPEG-UD.

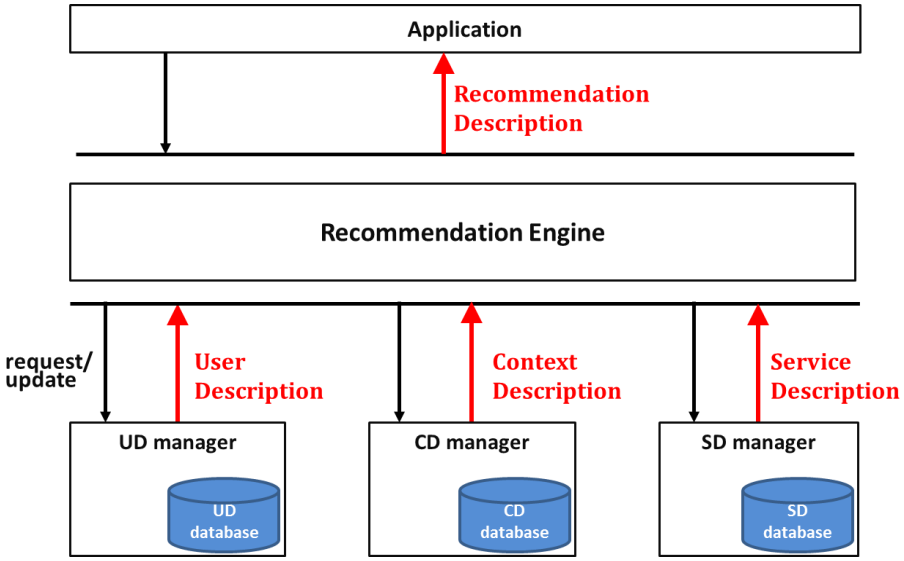


Figure 1: Conceptual Model of MPEG-UD (red elements UD/CD/SD/RD indicate the formats to be specified by MPEG-UD).

It has to be noticed that while some of the standard components, namely UD, CD, and SD refer to the description of specific entities or artefacts, respectively services, context and users, the RD is of a different nature. To better explain the need of having RD as a standard component in MPEG-UD, let us use the following example from the domain of Smart TV services.

Recent introduction of Smart TVs gives an uncountable number of possible choices of programs for a consumer, through the internet connection as well as through the broadcast channels. In such an environment, TV manufacturers consider including recommendation functionalities within the TV, to help their customers selecting the program (SD) according to their preferences (UD) and context (CD).

The recommendation solution can be acquired from other business entities with expertise in this field. With the "standardized" RD, a business entity, which has a solution for the recommendation engine, can provide an interoperable recommendation service. Then, TV manufacturers only need to provide a light application which takes the recommendation expressed in RD and provides customized recommendations to the users. Assuming that several recommendation service providers are available, the user may be asked to select one of these services. In such an environment, a standard format for RD ensures the following three benefits. *First, TV manufacturers are not constrained to select a specific recommendation engine nor to build one by themselves. Second, consumers have choices of selecting a recommendation service tailored to their preferences and context. Finally, recommendation service providers will compete to provide better services, without being tied to a specific TV manufacturer.*

From the above explanations, it can be concluded that the RD is a way to make the interface between recommender systems and Applications a standard one, so that Applications can manage to use different recommender systems, depending e.g. on licensing, performance or application domain constraints.

The connection between Applications and user, context and service descriptions realized through RD (and Recommendation Engines) ensures interoperability at this level, avoiding that each application will have to implement its own recommendation technology.

In this call, we ask for a data representation (e.g., XML schema, RDF schema or Ontologies to be used in combination) for each of the following entities: User Description, Context Description, Service Description and Recommendation Description as defined in Section 3 of N13881 and fulfilling the requirements in Section 4.2, 4.3, 4.4, and 4.5 of N13881.

Output documents

N13879 - Call for Proposals on MPEG User Description (MPEG-UD)

N13880 - Use Cases for MPEG User Description

N13881 - Requirements on MPEG User Description