INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2012/M29227
April 2013, Incheon (KR)

Status **For consideration at the 104nd MPEG meeting**
Title **Timeline Delivery for TS enhancements**
Authors J. Le Feuvre, C. Concolato, **Telecom ParisTech**
R. Monnier, A-L. Mevel, P. Gendron, **Thomson Video Networks**
M. Mattavelli, C. Alberti, D. Renzi, **Ecole Polytechnique de Lausanne**
F. Pescador, N. García, **Universidad Politécnica de Madrid**
M. Raulet, **IETR/INSA**

1 Introduction

In m26903 presented during the 102nd MPEG meeting and its follow-up m28136 presented during the 103rd MPEG meeting, a solution was presented, enabling carriage of DASH timeline values in MPEG-2 TS in order to unambiguously recover MPD media time in the current period, regardless of time discontinuities introduced at the MPEG-2 TS network level. This contribution presents detailed use cases for time synchronization of MPEG-2 TS content with external content, such as MPEG-DASH or MPEG-MMT.

2 Use cases for MPEG-2 Transport Stream Timeline

2.1 Generic concerns

The work presented in this contribution focuses on enhancing broadcast MPEG-2 Transport streams with broadband media. Various use cases for possible enhancements exist, and can be classified as follows:

- Enhancement with no synchronization or loose synchronization: in these cases, existing tools from already deployed TSs are sufficient to estimate the current time of the broadcast service with little accuracy, typically up to a few seconds or more;
- Enhancement requiring subjective synchronization: in these use cases, timing has to be computed with the same precision as regular inter-media synchronization (A/V sync, Text/Video sync …), which is usually a few tens of milliseconds up to a few hundreds;
- Enhancement requiring frame accurate: in these cases, timing has to be perfectly reconstructed (synchronization error tolerance is zero), otherwise either the decoding or the presentation of one media fails.

In this contribution, we mainly focus on the last type of use cases, as solutions covering this type also cover the use cases requiring subjective synchronization. However, examples of both subjective synchronization and frame accurate synchronization use cases are given in order to better understand the differences and constraints.

The following use cases descriptions are provided by the H2B2VS[1] project. This description is not exhaustive but reflects the preliminary work done in the very first months of the project. This contribution is fully supported by the Industrial Partners of the H2B2VS project which are listed in the table below.

Company	Contact	Email address
Thomson Video Networks	Raoul Monnier	raoul.monnier@thomson-networks.com
NagraVision	Minh-Son Tran	minh-son.tran@nagra.com
Neusoft Mobile Solutions	Lauri Lehti	lauri.lehti@fi.neusoft.com
SmartJog	Aurélien Violet	aurelien.violet@smartjog.com
TDF	Jean-François Travers	jean-francois.travers@tdf.fr
Teleste	Ilkka Ritakallio	ilkka.ritakallio@teleste.com
Turk Telekom Argela	Burak Gorkemli	burak.gorkemli@argela.com.tr
Vestel Elektronik	Kagan Bakanoglu	kagan.bakanoglu@vestel.com.tr

2.2 Use cases requiring enhancements with subjective synchronization

2.2.1 Audio switch from/to enhanced audio use case

This Use Case describes how the ability to switch audio quality can provide a richer and more compelling end-user experience when consuming content across different devices. The support of several audio channels configuration can be a feature to be sold with premium content that can be an additional revenue stream for content distributors.
In the context of an audio-visual or audio content consumption, an end user may need to move from a high-end audio device (e.g. home theatre in the living room) to a less performing device (e.g. simple TV set or mobile device). In this case, he should be able to switch the audio settings from 5.1 to stereo. The opposite is also true when switching from stereo to an enhanced audio channels configuration.
The advantages of supporting such use case are at least twofold. On one side the switch to a simpler audio channel configuration (e.g. from 5.1 to stereo) helps in saving bandwidth especially with the coming rich configuration related to 3D and augmented reality content where the number of channels can grow up to 24; on the other side, it supports the creation of a wider offer for the end user whereby basic content can be distributed by the broadcast network with a fixed audio configuration (with the limitations thereof) while premium content provides via the broadband network more flexibility and an enhanced experience.

2.2.2 Audio switch to any possible language

This use case describes how the ability to switch audio to any possible language can give the user a richer and more compelling experience and, at the same time, permit broadcaster to seek wider audiences for their content.

From the end user perspective, in the context of a single audio-visual or audio content consumption, the user may want to move from a language (e.g. dubbing track in native language)

[1] H2B2VS is a Celtic+ project which aims at investigating the hybrid Broadcast Broadband distribution of TV programs and services, using HEVC as a compression technology. It is coordinated by Thomson Video Networks. For more information, visit http://www.celtic-initiative.org/Projects/Celtic-Plus-Projects/2012/H2B2VS/h2b2vs-default.asp

to another one (e.g. track in original language). In this case he/she should be able to switch the audio from one language to the other and vice-versa.
From the broadcaster perspective, the support of multiple languages is a feature that is already deployed in many broadcast services and that can be enhanced furthermore by hybrid broadcast/broadband delivery, e.g., by using broadband servers to provide as many languages as many linguistically different regions have to be served. With the support of several languages, broadcasters can also promote greater viewer loyalty and increase advertising revenues through campaigns targeted to wider audience. Broadband delivery of multiple audio/tracks can simplify the automatic deployment of audio tracks out of different regions, by taking advantage of the IP address geo-location feature and moving the audio tracks routing intelligence to the client.

The advantages of supporting such use case are manifold. At first, the switch from e.g., a dubbing track in native language to the original language track, might allow the user to avoid poor dubbing quality. Second, the user may want to switch to a more universally understood language, like English, if, e.g., the audio content has to be consumed by an audience made by people speaking different native languages (e.g., the audience of an international event). Also, the switch may be led by learning purposes (e.g., for practising English by listening movies played by native English actors and eventually back if native English is too difficult to understand). Finally, from the broadcaster perspective, using the broadband for transmitting audio tracks in several language solves the problem of audio shuffling during playout: instead of requiring routing systems or master control switchers to shuffle audio tracks, the system can use the IP address geo-location feature typical of broadband systems to ensure the correct audio tracks are played out on the correct channels in the different regions.

2.2.3 Augmented user experience through Picture-In-Picture

This use case describes how to take advantage of the hybrid transmission to improve the user experience and create additional services for users.

The user experience can be increased by adding information into the main screen (Picture in Picture -PIP-). In this context, the user can select the information to be integrated: a second view (several views could be available), some additional information about the contents or other TV channels.

Additionally, the hybrid distribution will allow developing services oriented to the integration of groups of citizens such as handicapped people (e.g. sign language video during news programmes). The broadband network will be used to transmit the information for these groups of users because it is specific to these groups and the broadcast network cannot address so many different groups.

All these services need a subjective synchronization between both streams but no frame accurate synchronization as the information presented over the main picture have not a strict time dependency with the primary view.

2.3 Use cases requiring enhancements with frame-accurate synchronization

2.3.1 3D-Hybrid HEVC distribution use case

This Use Case defines how 3D applications can take benefit from HEVC new standard and hybrid distribution. In the few past years, 3DTV has drawn much attention as a new broadcasting service. In spite of technical advances in 3DTV broadcasting services, there are two barriers that hinder these new services from being launched for terrestrial broadcasting: One is the lack of available bandwidth for transmission of additional-view video via terrestrial channel and the other one is the necessary backward compatibility with the legacy 2D HDTV services. As a

solution, a hybrid 3D video distribution scheme is proposed for a 3DTV service where the traditional 2D programme is transmitted through the legacy broadcasting systems and the additional information to obtain a 3D rendering is delivered via the Broadband network to which the TV terminal is connected.
The rationale for this use case is that the Terrestrial Broadcast network is suffering from its bandwidth limitation. The 1st Digital Dividend already reduced the spectrum allocated to terrestrial TV. The 2nd Digital Dividend will force Terrestrial operators to release one third of their spectrum for mobile applications. There is no way to envisage 3D TV broadcasting on pure Terrestrial Broadcast networks. Cooperation of Terrestrial Broadcast networks with Broadband networks through the use of hybrid distribution technologies is a solution to allow terrestrial operators to offer 3D services.

2.3.2 UltraHD TV Hybrid HEVC distribution use case

This use case is quite similar to the3D-Hybrid HEVC distribution use case.

UltraHD seems to attract networks operators as a number of pre-commercial field tests are in progress or announced. However, for the same reason as for 3D (spectrum scarcity), it is very unlikely that Terrestrial Broadcast operators will be in a position to offer UltraHD services. This is why the same approach as for 3D can be used. The Broadcast network will carry the legacy 2D HDTV services while the Broadband network will bring the additional data allowing UltraHD decoding in the terminal.

The UltraHD TV Hybrid HEVC distribution will take advantage of the scalable version of the HEVC standard, which is under discussion. The broadband layer will carry the UltraHD enhancement layer, which will be used in the terminal together with the legacy 2D HDTV layer to display UltraHD pictures.

3 Media Timeline Delivery over MPEG-2 Transport Stream

3.1 Proposal Description

The proposal described below is design to enable the use-case presented in the previous section, but does no intend to provide all technical tools required by these use cases. More specifically, the proposal does not describe how to combine the primary media and the enhancement media, as the proponents believe this is a key differentiation factor for future products.
The proposal is designed on the carriage of a single stream sent in PES packets. By matching timing information carried in the stream and associated DTS/PTS carried in the PES header, a unique, accurate timing of each media access unit can be recomputed, regardless of PCR discontinuities that may happen in the broadcast.
Meta-data such as enhancement location, prefetch or reload directives are sent in the timeline stream. The format is designed to be small, so that no more than one TS packet is usually needed to carry the information. The format is defined to be DTV friendly, and avoids as much as possible usage of PMT updates.

3.2 Technical Proposal

Based on the previous section and contributions already presented at MPEG, we propose to define a media timeline format suited for MPEG-2 Transport Stream, but usable over other protocols.
It is based on a single payload format carrying al l the information identified above.

Syntax	No. of bits	Format
MediaTimelineTime_PES_Payload {		
has_mime	1	
has_location	1	
force_location_reload	1	
is announcement	1	
reserved=0	4	
		uimsbf
if (has_mime) {		
mime_length	8	
for (i=0;i<mime_length;i++) {		uimsbf
mime_type	8	
}		uimsbf
}		
if (has_location) {		
url_length	8	uimsbf
for (i=0;i< url_length;i++) {		
addon_url	8	uimsbf
}		
}		
timescale		
if (is_announcement) {		
time_before_activation	32	uimsbf
} else {		
media_time_anchor	64	uimsbf
}		
	64	uimsbf
}		

Semantics:

has_mime: 1-bit flag indicating that a mime type is included in this message.
has_location: 1-bit flag indicating that add-on location is included in this message.
force_location_reload: 1-bit flag indicating that add-on description shall be reloaded before attempting to map media times or locate media components.
is_announcement: 1-bit flag indicating that this message describes a media addon that is not yet active.
mime_length: 8-bit field indicating the mime_type length in bytes
mime_type: indicates the mime type of the add-on available at the indicated location. If no mime type is set, it is the responsibility of the terminal to determine the mime type. An implementation can decide to fetch or not the add-on based on this mime type indication.
url_length: indicates the length in bytes of the addon location.
addon_url: indicates the location of the addon. For MPEG-DASH, this points to the URL of the MPD. If a DASH period has to be identified, fragment URIs shall be used. if this string is empty, this implies that no addon is currently associated with the current program; if a DASH client detects a change of period, it shall switch to the indicated period regardless of the current payback state of other broadband media.

`timescale`: indicates the timescale used to express time fields (time_before_activation or media_time_anchor) in this message.
`time_before_activation`: indicates the time in timescale units until the timeline identified by timeline_id becomes active. An implementation may use this information to start prefetching content.
`media_time_anchor`: indicates the media time in timescale units corresponding to the PES PTS value of this packet for the timeline identified by timeline_id.
For MPEG-DASH addons, this indicates the MPD time of the active period identified by the given URL id as follows: let PTS_0 be the PTS of the packet carrying media time anchor MTA_0; until a new media timeline packet is received, the PTS of subsequent PES packets of other PIDs in this program have a Media Presentation Time, relative to the period start (cf 7.2.1 of MPEG-DASH spec), of:

$$MPT = (PTS - PTS_0) / 90000 + MTA_0/timescale - @presentationTimeOffset/SegmentBase.timescale$$

Each `MediaTimelineTime_PES` packet shall carry a PTS in its PES packet header. NOTE: in order to avoid interpolation issues when frame-accurate synchronization is required, the indicated PTS should be the same as the PTS of the associated video or audio stream for which frame accurate sync is needed.

4 Conclusion

In this contribution, we have reviewed use cases for enhancing MPEG-2 TS content with broadband add-ons; we have proposed a generic solution covering all hybrid delivery use cases, allowing time synchronization of add-on media such as MPEG-DASH or MPEG-MMT. While mainly designed for MPEG-2 TS environments, the solution is generic and could be used as a MediaTimeLine stream for other media transport protocols and formats (RTP, ISOBMF, WebM, …). We ask WG11 to start an amendment to the MPEG-2 Transport Stream format to include the proposed solution.