



European Broadcasting Union | Seminar



Production Technology Seminar 2009

organised with the EBU Production Management Committee (PMC)

EBU Headquarters, Geneva

27 - 29 January 2009

Report

Written by Jean-Noël Gouyet, EBU TRAINING

Revised and proof-read by the speakers

Opening speech	4
Seminar quick reference guide	5
1 2008 HD events – user reviews	6
1.1 Euro Cup 2008 – Launching HDTV at ORF	6
1.2 Hear it like Beckham! A new audio recording technique for sports	7
1.3 Simulcasting Eurosport HD & SD	8
1.4 Acquisition formats for HDTV production	9
1.5 New studio codec tests & concatenation issues	11
1.6 HDTV Distribution Encoder Tests Results	13
1.7 Loudness Group – 1 st results of work	14
1.8 HE-AACv2 listening tests for DAB+	17
1.9 Handling surround audio in 50p production & contribution broadcast infrastructures	18
2 IT-based production and archives	21
2.1 File-based production: problem solved?	21
2.2 Asset Management & SOA @ EBU	23
2.3 SOA Media Enablement - a media specific SOA framework. From ESB to abstract service description	26
2.4 Medianet technology – The missing link from SOA to file-based production	29
2.5 EBU-SMPTE Task Force: The (almost) final report	30
2.6 Request for Technology & first agreements	31
2.7 The Time-related Labelling (TRL)	32
2.8 How can you possibly synchronise a TV plant using Ethernet?	33
2.9 What replaces shelves: solutions for long-term storage of broadcast files	35
2.10 Living in a Digital World - PrestoSpace & PrestoPRIME	37
2.11 Metadata for radio archives & AES	39
2.12 Video Active – Providing Access to TV Heritage	41
3 The future in production	43
3.1 SMPTE Task Force on 3D to the Home	43
3.2 3D TV – Market Overview	44
3.3 High Frame Rate (HFR) Television	49
3.4 Future Television Production – Proof of Concept	51
3.5 LIVE extends the interactive television experience – The 2008 Beijing Olympic Games	52
List of abbreviations and acronyms	54

Foreword

This report is intended to serve as a reminder of the presentations for those who came to the 2009 seminar, or as an introduction for those unable to be there. So, please feel free to **forward this report to your colleagues within your broadcasting organisation**, but make sure to protect the password access to the presentations!

It may be a detailed summary of the presentation or sometimes a quasi-transcription of the lecture, for some more tutorial-like presentations (e.g. on Audio loudness, SOA, IEEE 1588, EBU Core, 3D...) or for some comprehensive test results or experience reports.

For more details, the reader of this report should refer to the PDF version of the **speakers' presentations**, which are available on the following site via browser: <http://tech.ebu.ch/production09>. You may also contact: Nathalie Cordonnier, Project manager - Tel: +41 22 717 21 48 - e-mail: cordonnier@ebu.ch

The **slides number** [in brackets] refer to illustration slides of the corresponding presentation in the .PDF version.

To help "decode" the (too) numerous¹ **abbreviations and acronyms** used in the presentations' slides or in this report, a list is provided at the end of this report. Short explanation of some terms may complete the definition.

Web links are provided in the report for further reading.

Many **thanks** to all the speakers and session chairmen who revised the report draft. Special thanks to Bob Edge (Thomson GV), Colin Smith (ITV PLC) with Ami Dror (XpanD) and Ethan Schur (TDVision), who provided a very nicely edited and comprehensive version of their presentations. "Merci !" to Nathalie Cordonnier and Corinne Sancosme who made the final editing.

The **reports** of the Production Technology **2006**, **2007** and **2008** seminars are still available on the EBU site:

http://www.ebu.ch/CMSimages/en/PMC08_Report-FINAL_tcm6-64142.pdf

http://www.ebu.ch/CMSimages/en/EBU-2007ProdTechnoSeminar-Report_FINAL_tcm6-50142.pdf

http://www.ebu.ch/CMSimages/en/EBU-2006ProdTechnoSeminar-Report_FINAL_tcm6-43103.pdf

¹ About 265!

Opening speech

Lieven Vermaele, EBU Technical Director, Switzerland

EBU is a union of 75 Active members from 56 countries, and 45 Associate members around the world, representing a big force towards the industry and Standards organisations. The EBU Technical Department, in order to become your reference in media technology and innovation, developed this last year a **strategic plan**, in three steps:

- 'Have your say' with questionnaires and interviews of the members, to analyse the situation.
- Redefine key missions, vision, values and objectives with the management committees.
- Define a clear activity plan in the different domains.

Our **vision and mission** are clear, expressed in the following strategic **objectives**:

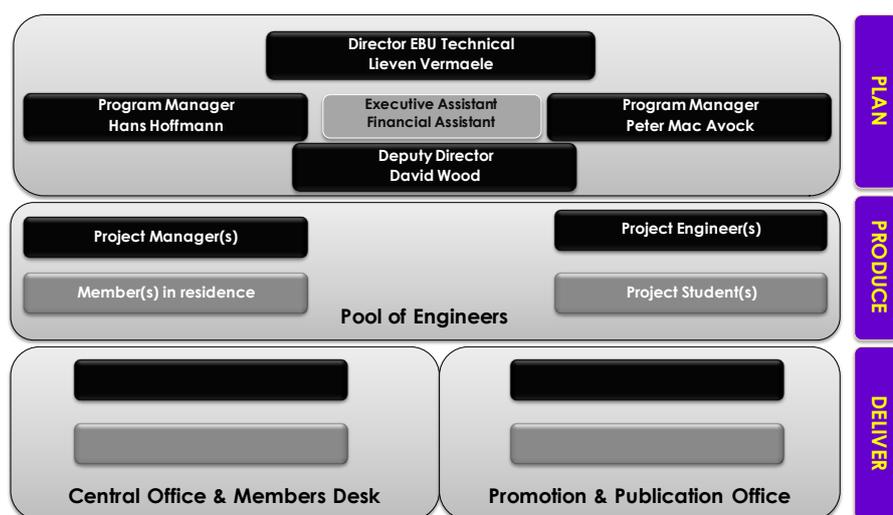
- We connect and share, bringing people together, sharing experience (seminars, on-line...).
- We develop and guide.
- We promote (open standards, interoperability, network neutrality, maximised user access, spectrum requirements...) and represent EBU members in industrial bodies, international organisations, regulation bodies.
- We drive (pushing for innovation where and when necessary, e.g. time labelling...) and harmonise (e.g. Digital Radio) for cost efficiency.

We will concentrate our activities in three **technological domains**, with corresponding programmes:

- Content creation, contribution and production technology, through PMC and NMC and the following programmes: HDTV and beyond – File-based production systems and infrastructure, and Networks - Archive and Storage.
- Media delivery technology, through DMC and SMC and the following programmes: Broadcast technology – Broadband fixed and wireless technology (spectrum and applications).
- Consumer equipment and applications technology, through DMC, and the following programmes: Display technologies – Applications and Interactivity – Security, Rights and Metadata.

For each programme we defined an activity matrix, matching the strategic objectives (Connect & Share...).

In order to achieve our goals we changed the **organisation** to plan – produce – deliver the work.



We also developed the website, which is becoming:

- One place where the information is put together (news, publications, presentations, seminars and reports, webinars...)
- The central place of the project groups, with on-line meeting tools such as the 'EBU Network' (finding a colleague).

Seminar quick reference guide

Per BOEHLER, NRK, Norway, Chairman of the Production Management Committee (PMC)

'Production Technology 2009' offers a very comprehensive programme.

The first day dedicated to **High Definition** presents first **users' experiences**: launching a HDTV channel in 8 months with 2 different production and emission formats (§ 1.1), a new audio recording technique to capture the kicking of the ball in a football match (§ 1.2), the infrastructure and the formatting for Eurosports HD & SD simulcast (§ 1.3).

The second session is focusing on the **tests** results of some **HD equipment**: camera systems for drama production (§ 1.4), studio and acquisition codecs with concatenation issues (§ 1.5), and distribution encoder (§ 1.6).

Audio loudness jumps control (§ 1.7), low-bit rate encoding HE-AACv2 test results (§ 1.8) and Dolby E in a 50p production environment (1.9) are the subjects of the last session.

The second day central topic is **IT-based production**, starting with the presentation of a new EBU working group on Networked Production (§ 2.1) and then trying to decrypt the buzz word **SOA** (*Service Oriented Architecture*) through a tutorial and the positioning of the EBU work (§ 2.2), two vendors' presentations of a real SOA framework system and modules (§ 2.3) and of a 'linked' network technology (§2.4).

The 2nd session reports on the final work to define a new **Synchronisation and Time-related labelling** (§ 2.5 - 2.7) with a zoom on a way to synchronise a digital TV plant using Ethernet and IEEE 1588 (§ 2.8).

The storage, costs and control of loss issues in **Digital archives** are analysed in the last session (§ 2.9) as well as the migration strategies through the projects PrestoSpace and PrestoPRIME (§ 2.10). The use of the EBU Core is illustrated for radio archives metadata (§ 2.11) as well as the access to TV archives at the European level (§ 2.12).

The third day brings us into the **future**, the "beyond-HD". **3D** is there, SMPTE is defining a 3D Home Master format for distribution (§ 3.1), the technologies... and the market are ready (§ 3.2). The future may also be **high frame rate** television (§ 3.3), **computer-assisted search and production** (§ 3.4) and **interactive** television (§ 3.5).

1 2008 HD events – user reviews

Chairperson: Reinhard KNOER, IRT, Germany

2008 has seen a fair amount of big HD productions – produced and distributed to the homes. The sessions of the first day highlight the experiences and challenges, including audio, encountered by broadcasters in the production and contribution of large sporting events in HD and the issues in the exchanges between the facilities.

1.1 Euro Cup 2008 – Launching HDTV at ORF + Corr1 + Corr3

Manfred Lielacher, Head TV Production management, ORF, Austria

In September 2007 the ORF Management took the decision to start HDTV with Euro 2008. Eight months later, on the 2nd of June 2008 ORF1 HD, simulcast with the ORF1 SD channel, was on air, just 5 days before the 1st match!

The **HD production** takes place in **1080i 25 Hz** format, in accordance with EBU Tech 3299E², because:

- 1080i 25Hz equipment was available from several manufacturers, and the short timeline did not allow for any delay (which would have been caused by introducing the 720p50 format).
- This format is in wide use internationally (e.g. EBU-Contribution, OB-vans on the rental market, etc.).
- The native1080i25 camera(IKEGAMI HDK-79EXIII) was available, and the Sony older HDCAM tape was already in use in ORF as exchange medium which did not support 720p/50 at this time.
- All ORF Productions so far (since 2004) were in1080i25 (e.g. New Years Concert, Operas, etc.).

The **HD emission** parameters are:

720p 50 Hz format, in accordance with EBU TR112 - 2004³;

Compression format: MPEG-4 AVC HP@L4 (H.264) at 14 Mbit/s;

Audio: 2x PCM, Dolby AC3 Multichannel Audio;

Distribution via DVB-S (ASTRA TP57) and DVB-C;

Conditional access system: Cryptoworks;

DRM: geographic copy protection must be possible.

The HD content broadcast should consist of: live sports events, movies & series, with a minimum of one HD transmission each day.

On the functional block diagrams [9] + [10], the colours indicate the necessary use of different codecs in the signal chain. We tried to keep the signal in the same parameters as long as possible and to reduce transcoding at the intermediary steps as much as possible. Slides [13] to [21] detail the HD facilities and equipment used.

What were the **lessons learned** from this experience in the following domains?

- **HD equipment:** the tight schedule and budget made it necessary to trust the manufacturers, who helped to solve problems with the new products (Alchemist Converter of Snell & Wilcox⁴, AV-HDXMUX-13T of Flashlink⁵, NV5000XP-HD Router of NVision⁶, HDC6800+ of Leitch⁷).

² EBU Tech 3299 – High Definition (HD) Image Formats for Television Production. December 2004

<http://tech.ebu.ch/docs/tech/tech3299.pdf>

³ EBU Technical Recommendation R112 – 2004. EBU statement on HDTV standards. 10/2004

<http://tech.ebu.ch/docs/tr/r112.pdf>

⁴ http://www.snellwilcox.com/products/conversion_restoration/

⁵ <http://www.network-electronics.com/flashlink>

⁶ <http://www.nvision.tv/>

⁷ http://www.broadcast.harris.com/product_portfolio/product_details.asp?sku=HDC_6800

▪ **Formats and Codecs:**

- some consumers think that the emission format 720p is “small HD” compared to 1080i, and this must be balanced by a good picture quality;
- one must be careful in down-converting from 1080i25 to 576i25 (SD) - mostly it works perfectly, but some visual elements, e.g. the vertical lines of a football field, created visual distortion;
- unhomogeneity in the codec line makes transcoding necessary bringing with it conversion artefacts;
- VITC is not well supported in the HD domain (RP188/ATC), with the need to add some TC inserters.

▪ **Alignment**

- the transition from the CRT monitor to the TFT panels forced the camera control operator to get the right aperture adjustment;
- large TV screens challenge frame stability, i.e. take care of steady shots from portable cameras;
- take care of lip sync⁸ along the chain, that means more necessary delays have to be inserted;
- take care of handling the Dolby E signal through the different stages: there may some critical guard interval alignment, the Ingest AVID Air Speed server only handles 24-bit, the Telestream FlipFactory cannot pass through such a Dolby signal.

▪ **HD Material**

- Work hand in hand with the programme and marketing departments to create awareness for producing original HD content and mark the HD programmes on the screen with a HD-channel logo.
- Be flexible in the budget planning: rental companies are charging extra for HD copies.
- Be careful by using SD material up-converted into HD programmes: the quality difference is visible.

ORF wants to increase the number of HD programmes per day: major sports events (Olympic Games, Champions-League, National top events, etc.), blockbuster movies as much as available, TV-Series, Spring Event (Dancing Stars, in March), in parallel with the continuous expansion of HD-enabled equipment, cabling and facilities (edit suites, production studios, on-air recording, etc.).

1.2 Hear it like Beckham! A new audio recording technique for sports

Gerhard Stoll, Senior Engineer, Audio System, IRT, Germany

Sports productions are a great platform for HDTV and multichannel audio. The combination of "best picture and best sound" provides a coherent and emotional coherent impression.

The basic audio elements of sports production are:

- Atmosphere, through a mix of crowd/ambience (front and surround fields).
- 'Sports sound effects' (e.g. the sound of the kicking of the ball) through mono and stereo game microphones.
- Other sounds (coach talking) through camera microphones of hand-held cameras (typically provided with stereo shotgun microphones).
- Stereo elements / music from videotape playback.
- Contributions (short replay from another game) from external links.

There are two kinds of audio environments in sports: 'static' (ambience scene of the stadium/arena/sports hall) and the 'dynamic' sports sound effects – you want to hear what you see on the ground. The reproduction of the 'sports sound effects' is quite difficult, especially in soccer: high level of the ambient noise, quite close to mikes, and low level of the ball, quite far from mikes.

The solution mostly used today to capture the 'ball sound' is to install directional mikes (typically 8 to 12) around the soccer field [8] in a height of about 60 cm above the ground. This needs a constant and dynamic adjustment of the mikes' levels, and however this is quite often not done by the sound engineer.

⁸ Cf Production Technology 2007 seminar report § 3.5

A new technique to capture the sports sound effects is based on a **system** [16] including:

- The automatic tracking of highly directional microphones, installed on 6-7,5 m mast or poles ideally positioned 7 m behind each goal [11]-[13], with the help of highly dynamic and silent remote heads.
- The identification of the ball position either by an automatic image ball-tracking system [19]-[22], eventually coupled with the camera parameters, or manually by an operator following the ball with a mouse on a screen. For automatic tracking, either a stereoscopic camera system (Tracab of Sweden) or a simple live tracking system (of Signum Bildtechnik - Munich) including 2 fixed cameras with wide-angle lenses, can be used.
- The control and the following processing of the microphone signals by a central controlling computer. The controller software for remote heads [21] proceeds to:
 - the automatic levelling of the remote heads mikes depending on the position and distance of the ball;
 - the automatic compensation of the time delay of audio related to each mike;
 - the automatic extrapolation of the latency of the system versus shots speed, by motion estimation.
- The outcome of the system is a processed audio signal of the ball that could be used and mixed with the ambiance into the 5.1 as a "field noise" component. The noise of spectators [18] is maximum between 500 Hz and 1 kHz, where there is not so much ball sound [15], so the ground noise can be further attenuated in the control unit.

Note the very convenient use on the field of a single optical link (Unilinks), only one cable through all the stadium, for audio (analog and digital), video (SDI, HD-SDI) and control data (RS232, RS485).

In addition to the question of a well-balanced mix of atmosphere and sports sound effects, some **other questions** (for the next year seminar!?) are put forward:

- How to deal with the commentary, i.e. well-balanced mix of commentary and atmosphere? Must be the commentary always in the center-channel only?
- Does the center channel carry basic sound events (e.g. sound of ball, coach's voice)? Or use only phantom center?
- What is with the use of audio compression? How are we dealing with dynamics when down-mixing to stereo?
- Stereo clips, commercial spots within a surround event: up-mix to surround or play them as they are?
- How to deal with the LFE? Extra mikes for LFE signals or simple bass management?
- Etc.

1.3 Simulcasting Eurosport HD & SD

Pascal Crochemore, CTO Distribution & Vincent Gerard-Hirne, Technical Director, Eurosport France

Eurosport [3]-[7] started to consider HD 2-3 years ago. At the end of 2007 the 'go ahead' for the project was given for a SD - HD simulcast, and on the 24th of May 2008, HD was launched with the Roland Garros championship.

Eurosport HD is now a simulcast of the Eurosport channel to 28 countries, offering over 140 sporting disciplines, amounting to more than 6500 hours of sports (over 3100 hours live). The HD video is transmitted in MPEG-4 AVC compression format at 12 Mbit/s, and the sound in stereo channels in 15 (soon +2) major languages with a rapid migration to 100% 5.1 surround sound (presently only English, German and French). The total bit rate is about 15 Mbit/s.

Inside the central Eurosport facility (south of Paris), where the main production takes place and with the transmission uplinks to 19 local facilities [16], the playout suite is 100% HD and 5.1.

The SD inputs are up-converted, if necessary, with a Snell & Wilcox up-converter (for live or delayed programmes) or with the K2 Thomson Grass Valley server for recorded programmes (magazines). For

SD distribution, the HD output is down-converted to SD and cropped to 4:3. The picture aspect ratios of the sources, 4/3 SD (still 25% of incoming feed for all channels) or 16/9 SD (67%) or HD 16/9 (10%), are handled in different ways at the HD and SD output [19]-[21]. The commercials are in a "14/9" box to manage in the same way the HD and the SD output. The switch to the 16/9 picture aspect ratio will be complete mid 2010.

Tape formats are only used for ingest: Digital Betacam for SD and HDCAM for HD. For the recording and editing, the SD compression format has been upgraded on the server from 24 Mbit/s to IMX 50 Mbit/s and the HD compression format corresponds to XDCAM MPEG HD 422 50 Mbit/s [25].

In order to get higher HD programme quality, what is needed?

- Productions by the host broadcasters in SD 16:9 at least.
- More productions by the host broadcasters in native HD with Dolby E multi-channel audio.
- High bit rate on SD or HD contribution links from the event.
- An efficient workflow to limit up/down conversion and aspect ratio conversion.
- Good encoders (the Thomson Grass Valley is the only one to handle 12 audios).

Test reports – HD User equipment

1.4 Acquisition formats for HDTV production

Walter Demonte, Camera and Sound department, WDR, Germany

The key points in HDTV drama production are: formats, camera systems and workflows.

Formats

For HDTV drama production the 16mm film is suitable only with optimal operation and under use of the best technology. The grain noise and resulting sharpness of the film material is always critical. This brings no real HDTV impression with existing production workflows. The way out would be to use 35 mm film, but this is impossible because of the budget.

Full digital HD-production avoids disadvantages of analogue film and is future proof. First the direct transfer to the postproduction workflow offers possible savings on production costs (no film stock, no film processing, no telecine). Second, during the last two years the current camera systems has been improved so far, that they can achieve the look and quality of film, with enhanced dynamic range and optimized colour space being the major improvements. In comparison to 16mm film the resolution of digital cinema cameras is now much better.

Camera systems

Currently there are 3 camera systems on the market which are suitable for drama production: Sony F23/F35, Arri D-21, RED ONE. WDR conducted tests [6] on all systems in case of drama production with the following criteria: picture quality (on 50" display), operational and handling aspects, postproduction workflow. Sony F-23 has even been tested in comparison to 16mm film and the HDCAM-Camcorder HDW-750⁹ with „Digi-Primes“ and Pro 35. The test results are presented in the table hereafter

⁹ Shows recognisable weakness in dynamic range and picture quality.

	Sony F-23 compared to Super 16 Kodak Vision 3 + ARRI 416	Arri D-21	RED ONE
Camera system		Basically D-21 is a 35mm camera system. Up to the CMOS-target, the camera technique is similar to ARRI 535. Target size is the same than full 35mm 3:2. It can be used as full 35mm or 16:9 target with 1080x1920 pixels. Camera uses mirror shutter and optical viewfinder.	RED ONE is a completely new designed camera. CMOS target has the same size than 35mm film. The basic resolution of the sensor is 4k. It works also in 1080x1920 mode with nearly the same target size than 16mm. Modular camera concept with individual design and accessory.
Frame/s		Camera runs up to 60 frame/s.	Frame rates up to 120 frame/s are possible (for slow motion).
Lenses		All 35mm lenses on the market can be used.	All 35mm film lenses can be used.
Grain noise	No grain noise. 16mm film shows different strong grain noise. Less grain noise improves three-dimensionality of picture.	Almost no grain noise.	
Dynamic range	Dynamic range achieve up to 11 f-stops. This is comparable to film.	Dynamic range achieve up to 11 f-stops. This is comparable to film.	
Quality features	Sony F-23 can provide nearly the same picture quality than 16mm film. Picture impression is much better than film. Picture looks much more detailed. Colour space and dynamic range is as expected high. In some cases there are slightly defocused areas visible. Possible to detect operational failures on focusing work. Working space for colour correction is quiet fair. But there is no electronic sharpness. On low light settings, same picture quality than film can be achieved. Even though the sensitivity of F-23 is less than Vision 3 film stock.	ARRI D-21 can provide "state of the art" picture quality with excellent sharpness, fine details and well balanced colour correction. The depth of field is absolutely the same than 35mm. WDR did all test in HDCAM-SR and ARRI-RAW. In any case the results are excellent. Nearly the same functionality than with 16mm or 35mm camera can be achieved. The full range of ARRI accessories can be used.	Picture quality of the camera can be referred to as excellent. RedCode file format, which is based on JPEG 2000 (very easy to handle on Final Cut but not on Avid based postproduction). Further on, special software tool provide a wide range of manipulation options. Setting up "look up tables" for accurate picture control, even in the field, is possible. Unfortunately RedOne is still in beta-status. Therefore the camera does not work absolutely stable. The RedCode is not fully supported by the major postproduction systems (i.e. Avid, Quantel). Case depending it is necessary to do 6 to 8 work steps to transfer the material to the postproduction system. In some cases loss of Time Code is possible. Rendering time needs 15 to 30 times of real-time
Recording	Recording format is HDCAM-SR. Recorder can be attached to camera or be connected via HD-SDI dual link. HDCAM-SR is basically a tape workflow. Therefore integration to existing workflows in WDR is easily possible	Recording device is not integrated to the camera. Choice of HDCAM-SR or ARRI RAW-File recording. Several manufactures provide recording devices for tape-based or rather file-based recording. Flashpacks, attachable to the camera, are also available.	RedOne uses CF flash-packs and hard disks as recording devices. Therefore the camera is very lightweight and flexible.
Dimensions Weight	Dimensions of camera and recorder does not really support handheld and steadycam operation.	Dimensions and weight supports handheld and steadycam operation.	
Price		(complete) ~500 k€	~150 k€

The test results show that all 3 camera systems are suitable for drama production. But colour space and the colour reproduction (esp. skin colour) are limited on the recording format of HDCAM. Therefore,

WDR decided to use **ARRI D-21**, which offers maximum picture quality, no disadvantages in comparison to film, and a quality comparable to 35mm film. The cost saving induced by the absence of the processing of negative film allows the higher cost of the digital camera equipment rental.

Workflows

In the workflow with the camera D-21 directly connected to the HDCAM SR recording unit [14], the video is first down-converted to XDCAM SD for off-line editing. One can use the recorder internal look-up table, to transfer basic colour-corrected material to the postproduction suite. Audio is transferred on hard disc drives to postproduction.

In the workflow with the Venom Flashpack (unfortunately 4:2:2 only, not 4:4:4) attached to the camera [15], the material data are transferred on the set to avoid to have too many Flashpacks.

The postproduction workflow [16] consists of the final on-line video editing with Quantel eQ and of the audio editing on Avid Media Composer.

1.5 New studio codec tests & concatenation issues

Massimo Visca, Centro Centro di Produzione TV di Torino, RAI, Italy

The new **P/HDTV** group (successor of the PMC project P/HDTP, which conducted the initial studies on HDTV in the production environments) has defined the following tasks:

- 1 Share the experience between EBU Members (lead by EBU).
- 2 Investigate studio compression codecs (lead by IRT).
- 3 Analyse the performances of the lenses for HDTV cameras (lead by NRK).
- 4 HDTV cameras and camcorders (lead by BBC).

Concerning the HD production codecs, the final goal is to provide guidance and neutral information to EBU members, to help them to take their own decisions. The preliminary tasks were to:

- define a test plan¹⁰ [8] of stand alone chains (cascading of the same encoders [12]) and production chains (cascading of the same encoders [24]+[25]);
- define the corresponding test conditions [13] + [26] and select reference displays [14].

Concerning the requirements for HDTV codecs, it was noted that it is necessary to test them to the 7th generation, with a quality headroom mandatory in any case, because large display acts as magnifier of artefacts and archives must be usage future-proof.

Picture quality is only one of the parameters to be considered in the comparison of different solutions on the market. Other key parameters are: storage requirements, network requirements, physical media bearer, error resilience, quality and cost related to use.

The activity started 2 years ago [10] with the test of legacy algorithms (HDCAM, HDCAM SR, DVCPRO HD, XDCAM HD) and of 4 new algorithms (Sony XDCAM HD 422, Panasonic AVC-I, Avid DNxHD, Thomson GV JPEG2000)¹¹.

In November 2008 the expert viewing of the **Apple ProRes422** codec was performed. The results for **standalone chains** of this codec are presented in the table hereafter.

¹⁰ EBU BPN 076-079 Supplement, December 2007 – New HDTV Studio and Acquisition Compression System Analysis

¹¹ Cf. Production Technology seminar 2008 report - § 2.4

Algorithm		Apple ProRes422 proprietary codec, intended for High-Quality NLE
Frame rate		1080i/25, 1080p/25, 720p/50
Bit rate		122 Mbit/s ProRes422, 184 Mbit/s ProRes422HQ (codec's target bit rate, i.e. VBR codec)
Subsampling		NO
Chroma format		4:2:2
Bit resolution		10 bits
1st Gene (3H)	[16]	The source and coded pictures were rated as <i>identical</i> for all bit rates (122 & 184 Mbit/s) and for all the formats (1080i/25, 1080p/25, 720p/50)
4th Gene (3H)	1080i 720p [17]	For 122 Mbit/s <i>just perceptible increase of noise</i> was noted for some sequences and a <i>perceptible increase of noise</i> was noted for the most critical sequences. For 184 Mbit/s pictures were rated as <i>identical</i> for non critical sequences and <i>nearly identical</i> for critical sequences, where a just perceptible increase in noise was noted in sub-areas.
	1080p/25 [18]	For 122 Mbit/s <i>just perceptible increase of noise</i> was noted in sub areas of the most critical sequences. For 184 Mbit/s pictures were rated as <i>identical</i>
7th Gene (3H)	1080i 720p [19]	For 122 Mbit/s : a <i>clearly perceptible</i> increase in noise was noted for almost all sequences at 122 Mbit/s compared to those codec at 184 Mbit/s, for both 1080i and 720p formats. For 184 Mbit/s : - for 1080i, a <i>just perceptible</i> increase of noise was noted for some sequences (or perceptible for the most-critical sequences); - for 720p, pictures were rated as <i>nearly identical</i> for most sequences. Just perceptible increase of noise was noted in sub areas of the most critical sequences.
	1080p/25	For 122 Mbit/s . A <i>perceptible</i> increase in noise was noted for almost all sequences at 122 Mbit/s compared to those coded at 184 Mbit/s. For 184 Mbit/s . Pictures were rated as <i>identical</i> for most sequences and <i>just perceptible</i> increase in noise was noted in sub-areas of the most critical sequences.
Compared at 184 Mbit/s with legacy HDCAM SR (3H)	1080i 720p	At 4th generation . Pictures were rated as <i>identical</i> for non critical sequences and <i>nearly identical</i> for critical ones, where just perceptible increase of noise was noted in sub-areas for ProRes422. At 7th generation . For ProRes422 a <i>just perceptible</i> increase in noise was noted for some sequences and a perceptible increase in noise was noted for the most critical sequences.
Compared at 122 Mbit/s with legacy DVCPRO HD (3H)	1080i 720p	At fourth generation . In general, ProRes has <i>perceptibly higher resolution</i> but also a <i>just perceptible increase in noise</i> in most sequences.

The full test results of all compression systems are available to EBU members in the **BPN reports series 076 to 080**.

One further BPN report will deal on simulation of selected **concatenated production chains**, with the following conclusions:

Without NLE (interconcatenation of the acquisition formats XDCAM-HD 422 50 and AVC-I 100)		[27] In general, loss of resolution and noise are perceptible but, on the average, performance is slightly better than expected. Whole production chain compared with fourth generation of a single algorithm provides similar results.
With NLE (interconcatenation of the acquisition formats XDCAM-HD 422 50 and AVC-I 100 with the algorithms used in NLE systems DNxHD and ProRes422)	At lower bit rate (~ 120 Mbit/s)	[28] In general, a whole production chain at ~ 120 Mbit/s, introduces artefacts in the range between <i>just perceptible</i> and <i>perceptible</i> . The comparison between the whole production chain at ~ 120 Mbit/s, and the 4th generation of a single compression algorithm, provides <i>comparable</i> performance. WARNING: An inter-concatenated production chain based on ~ 120 Mbit/s NLE seems to be able to provide only a limited amount of <i>headroom quality</i> .
	At higher bit rate (~ 185 Mbit/s)	[29] In general, a whole production chain at about 185 Mbit/s, provides, for non critical sequences, a picture quality <i>identical</i> to the picture quality available in acquisition. For critical pictures, artefacts are <i>just perceptible</i> . [30] The comparison between the whole production chain at about 185 Mbit/s, and the fourth generation of a single compression algorithm, provides <i>nearly identical</i> performance.

The EBU Recommendation R 124¹², accessible to all publics, provides guidelines for the 'Choice of HDTV Compression Algorithm and Bit rate for Acquisition, Production and Distribution'.

1.6 HDTV Distribution Encoder Tests Results

Rainer Schaefer, Head of Production Systems TV, IRT, Germany

The previous report of the P/HDC group work [4] was presented in January 2008¹³.

For the tests, two sets of sequences were assembled from transparent sources and from the former P/HDTV group [6]-[7]+[20]. The parameters used for evaluation were:

- Formats: 1920 x 1080i/25, 1440 x 1080i/25, 1280 x 720p/50
- Bit-rates: 6 to 20 Mbit/s in steps of 2 Mbit/s
- GOP structures (with an I-frame distance ~ 0,64 s): N16M3 for 1080i/25, N32M3 for 720p/50, and dynamic GOP enabled, if supported.

The **H.264 (MPEG-4 AVC)** encoders used for evaluations, during two experts viewing sessions, were: Ateame Kyrion (status Q3/07), Harmonic Electra 7000 (status Q3/07), Scientific Atlanta D9054 (status Q3/07), Tandberg EN8090 (status Q3/07, and status Q2/08 optimised), GVG/Thomson Vibes EM3000 (status end Q4/08). The state-of-the-art HD MPEG-2 encoder Sciati D9050 was used as an anchor and reference. Three different tests were undertaken, with the following results:

Test	Tasks	Test results
Step 1 Distribution only [12]+[17]	Identify critical sequences of the whole set of sequences and record general observations on H.264 encoder	H.264 (MPEG-4 AVC) codecs are generally better than an MPEG-2 encoder operating at twice the bit-rate Generally less coding artefacts for AVC except for grass and diva On average MPEG-2 (at doubled bit-rate) & H.264 are comparable for critical scenes; just perceptible loss of resolution for some scenes in H.264 depending on the codec optimisation Visible artefacts for sequences.... (Diva, Oly Flags...) at low bit-rates such as 6..8 Mbit/s H.264 Sometimes perceptible loss of sharpness for H.264 (one encoder, certain sequences...), depending on optimization of encoder Good sharpness in general... GOP pumping perceptible to very perceptible in some sequences (...) GOP pumping in sub-areas of certain sequences ...
Step 2 Distribution only [13]+[18]	Find the bit-rate of device under test to match an "upper anchor" (24 Mbit/s of the reference MPEG-2 encoder).	According to the formats, the following average bit-rates of the 5 codecs tested were needed to match: 1920 x 1080i/25: 12,833 Mbit/s (8 - 16 Mbit/s interval) 1440 x 1080i/25: 12,133 Mbit/s (10 - 14 Mbit/s) 1280 x 720p/50: 10,533 Mbit/s (8 - 14 Mbit/s)
Cascaded 4th generation Production + Distribution [14] + [20]	Identify whether production encoder stresses distribution encoder at low bit-rates and whether production encoder limits distribution quality at high bit-rates	At high bit-rates: No significant impairments of production encoder and type of production encoder visible In some cases: It may be the case that the picture quality has improved for 720p/50 in such a way that differences now become more visible At low bit-rates: Just perceptible loss of resolution at 3H viewing distance Just perceptible increase of coding artefacts at 3H GOP pumping in certain sub-areas of critical sequences Perceptible increase of noise for certain sequences In summary: no significant dependency on any production encoder!

¹² <http://tech.ebu.ch/docs/r/r124.pdf>

¹³ Cf. Production Technology seminar 2008 report - § 2.5

The distribution encoders were also tested on Lipsync and latency [21].

The full test results of all compression systems are available to EBU members in **the BPN reports series 085 to 087** with one supplement and two more reports to come.

Encoders

Encoders behaved different in terms of trade-off between resolution (sharpness) and coding artefacts for critical sequences (some encoders are optimised for low bit rates and perform pre-filtering, others for high bit rates).

Encoders showed different behaviour in terms of buffer control (GOP pumping).

Differences between encoders have become significantly smaller over the last 2 years. Differences in “emergency strategies” with demanding content are still visible. Bugs have been reported and solved (one encoder varied the resolution with demanding sequences, but did not recover from stress sequences and maintained on low resolution mode forever)

Sampling formats

1280 x 720p/50 shows advantages over 1920/1440 x 1080i/25 for typical screen sizes in terms of bit-rate savings (about 20%) and in terms of processing in the display.

Bit-rates

H.264 performs up to/about 50% better than MPEG-2 and even better for certain sequences.

Some experts felt strongly that even with the best encoder that 8 Mbit/s (1920x1080i / 25) is insufficient for HD broadcast of critical material. All experts felt strongly that with the best encoder, 6Mbit/s (1280x720p / 50) is insufficient for HDTV broadcast.

Recommended minimum bit-rates “for critical material but not unduly so”:

- **10,5 Mbit/s** minimum CBR for 1280 x 720p/50
- **12,1 Mbit/s** minimum CBR for 1440 x 1080i/25
- **12,8 Mbit/s** minimum CBR for 1920 x 1080i/25(MPEG-2 24 Mbit/s reference)

Quality of the encoders has reached a mature level for various vendors and less drastic improvements in terms of picture quality are expected in the future. Other parameters may prevail: differences in statistical multiplexing, optimisation to sharpness or minimum coding noise, other features such as supported audio formats and integration aspects.

As further work, cascading with converter is investigated in N/SC.

Audio developments

1.7 Loudness Group – 1st results of work

Florian Camerer, 'Tonmeister' & Trainer, ORF, Austria

The fact is that we broadcast a range of programmes with very different levels and very different dynamic ranges. How to prevent or to get rid of 'loudness jumps'?

The EBU group **P/LOUD** - of more than 70 members - has been launched with the following objectives and work areas:

- Change the levelling paradigm from peak to loudness.
- Define a new true maximum peak level. The Recommendations we had so far still accommodate the analog days with the -9 dB FS maximum permitted level, measured with a QPPM (Quasi-Peak Programme Meter).
- Look at the dynamic range of programmes directly related to the loudness issue.

Some **extreme examples** [13]: movies with very low level of average loudness (e.g. -28 dBFS) and a big difference between the average loudness and the maximum peak level – and on the other hand commercials with very high average loudness level (e.g. -13 dBFS) and very small dynamic range. And we cannot transmit audio unaltered with this 15 dB difference - that would be unacceptable for listeners.

What did we do up to now? We normalized to the **peaks** with PPM (Peak Programme Meter) [14] (to 'quasi-peaks', taking into account the 10 ms reaction time meters), making the situation of average loudness even worse with an even larger difference between movies and commercials! So to broadcast it, we compress the audio signal [15]. We still have the same peaks, but we push up the low-level details. Of course we sacrifice the dynamic range. So, this solution is already a compromise.

The ideal solution would be to **normalise to loudness** instead of peaks [16]. The -31 dB figure in the "**Line mode**" stems from the Dolby system (it is the lowest possible value of their loudness metadata parameter 'DIALNORM'). Everything, from the line level signal to the decoder of the set-top box, is aligned to the same -31 dBFS level, loudness normalised, and we have a totally varying peak value with the constant loudness value. That would already be a fantastic solution, the consumer at home not being forced anymore to adjust the volume with his/her remote control. An interesting area is this huge amount of headroom for the programmes that used to be highly compressed, especially the commercials that could be produced again in a transparent dynamic way. Compression would only be used for artistic reasons and not for the only purpose of sounding louder and louder! A problem might be that the dynamic range may be too big for the living room.

But we live in a non-homogeneous coding world (PCM, Dolby, MPEG...). If we switch to a channel with MPEG-1 Audio Layer II, usually the loudness is in the range of -20 dB [17], and so there is a gap of 11 dB between programmes normalised at -31 dB loudness and programmes transmitted as they are. Therefore there is a second mode in the system called "**RF mode**" normalising the loudness to -20 dB, more comparable to the legacy MPEG Audio programmes found in other channels. We have again loudness normalisation, but we have nevertheless to apply some compression to avoid overshooting, for the action movies, for example.

Therefore we are looking at the **Recommendation** developed by the **ITU Working group 6G**, normalising everything to -23 dB, suggesting a new maximum true peak level of -2 dB [18]. The ultimate goal is that the ITU Rec. will be the same as the EBU Rec.

Another recommendation from the Utrecht school of Music Technology [19] suggests -21 dB as the target level (but it is only 2 dB apart ITU Rec., even less in fact) and a maximum peak level of -5 dB, the reason for this value being the concern for the analog re-broadcasters.

As far as **measurement** is concerned the ITU Group issued Recommendation **BS.1770**, which is now the basis for the implementation of most loudness meters. It is a very simple measurement, easy to implement. Starting from the well-known weighting curves [21]: A (very low level signals) ... D (for noise measurement). The revised low-frequency B-curve is a very easy to implement high pass filter, and the B-curve has been modified for surround sound to include a high-frequency weighting filter [22], and this is the basis for the ITU measurement. This 2nd Revised Low Frequency B-curve is named **R2LB**, or **K**-weighting. If you measure Loudness, you then speak of LKFS (**L**oudness, **K**-Weighting and **F**ull **S**cale), for example "-23 LKFS" (no need to say "dB LKFS"...), and if you substitute R2LB to K it becomes LR2LBFS [26]!

One of the issue is on which **signal type** do we base our measurement? There are strong contenders for voice (Dolby) but there are others for music (concerts), sound effects (commercials - there are very short pieces where an algorithm for detecting dialog and speech has not enough time). The ultimate goal is to find a basis as broad as possible, and we certainly will recognise and recommend all three types of signals [28].

Gating is also very important. For example, during a golf sport transmission, nothing is happening in audio for a long time (very little atmosphere, presenter saying something, then half a minute silence...). You don't want the measurement to be just too low because most of your transmission is very low level. So we are thinking about a threshold level below which the measurement is paused. This must be the

matter of investigation and extensive testing.

Time constants are also a very important issue, especially for short-term measurements, because in the end we want to switch to loudness measurement in live production. You then need then a meter which gives you appropriate feedback, fast enough, so you can react, but not too fast, like a Peak Programme Meter. So we are looking into which time constants are appropriate for short- and medium-term measurements.

Does the inclusion of the **LFE** in the measurement make any difference? 5.0 compared to 5.1, does it make a big difference? It of course depends on the level of the LFE...In the ITU measurement the LFE is currently discarded.

Looking at the **new maximum digital peak level**, it is not going to be zero dBFS, because you need headroom for the encoder, but probably -2, -3.

ITU-R has already in it the recommendation to use **oversampling true peak meters**, not only counting samples. If you, for example, use a regular counting samples digital meter for the newest release of Metallica heavy metal band, you will get a constantly LED 'Over' lit – that would not help at all. In that kind of production, there are samples peaks that go almost 2 dB ABOVE 0 dBFS and they distort the digital-to-analog converter. Metallica sounds then more heavily distorted than it is already! Metallica is the new world champion in perceived loudness, because their CD has a loudness level of -3.8 LKFS, which is almost 5 dB louder than pink noise at Full Scale! The loudness race comes to a catastrophic situation.

We have to learn to produce now with **loudness meters** compared to peak meters. There are quite a lot of companies offering loudness meters based on the new ITU standard. They have still differences, because the time constant, etc. are not fixed yet - therefore the behaviour of their meters is slightly different. Some snapshots: T.C. Electronic LM5 [36] with the radar display. RTW [37] with the blue bars for loudness and the two adjacent bars for peak levels, and with short-term and long-term loudness (on the right side). We will probably put in our recommendation some basic requirements how loudness meters should look like, without specifying too many details... but the simultaneous display of loudness levels and peak levels (you do not want to distort your signal chain) is a good thing. With the DK Audio meter [38], we are used to levelling to zero and want to normalise to this magical number. If we come to recommend a target level of -22 LKFS, then there is of course the possibility for the meter manufacturer to interpolate that into a so called zero Loudness Unit (LU). Behind that stands -22 LKFS. So for people who are not aware of the standards (editors...), it is probably easier to level in a way that at the end the bar hits zero! The software meter from Dolby [39] with the option of letting an algorithm try to distinguish between dialog and non-dialog. All these meters integrate the ITU algorithm already and they will adapt if we have ongoing modifications.

As far as the **target level** is concerned, this is one of the most important goals. The ideal would be that we find one single target level, one figure where everything is normalised to. Looking at the multichannel programmes, it might be the case that we need a range of possible loudness target levels, since the preferred listening level of multichannel is highly dependent on the production itself: movie compared to a rock concert or to a nature documentary...with different mixing styles, dynamic range, etc. Again, this needs testing.

In conclusion, follow the ongoing development in P/LOUD and anticipate it your own company, because it means equipment, training, looking at your own programme flow, your own current practices. And if you don't come to P/LOUD, P/LOUD will come to you! We will effectively set up a roadshow after the Recommendation is finalised, and visit all the main broadcasters.

1.8 HE-AACv2 listening tests for DAB+

Mathias Coinchon, EBU Technical Department, Switzerland

HE-AAC is a low bit rate audio codec (also called AAC+), which may use two tools [3], Spectral Bandwidth Replication (SBR) and Parametric Stereo (PS), inducing different bit rate ranges:

- Plain AAC: generally for bit rates >96kbps (stereo)
- AAC+SBR (v1): generally for bit rates <96kbps (stereo)
- AAC+SBR+PS (v2): for bit rates <56kbps (stereo)

It is standardised in MPEG-4 Audio ISO/IEC 14496-3:2005 Amd.2 and specified in many applications, for Digital Radio: DAB+ and Digital Radio Mondiale (DRM) / Digital TV: DVB-S, DVB-H... / Mobile TV: DVB-H, T-DMB (EU profile) / Mobile Phones: specified in 3GPP / Multimedia players, Internet streaming.

Two possible transform lengths can be used: AAC 960 in DAB+ and DRM, and AAC1024 in all other applications.

Another version, MPEG Surround, is using Spatial Audio Coding (SAC) MPEG-D Part 1 (ISO 53003-1).

DAB+ (ETSI TS 102 563) is an enhancement of the Digital Audio Broadcasting standard. When traditional DAB is using MPEG-2 Layer II (24 or 48kHz sampling), DAB+ is using HE-AACv2 (with 960 transform length, for 32kHz or 48kHz sampling) or MPEG Surround.

One of the tasks of the EBU D/DABA project group is to evaluate the DAB+ Audio Quality. A Phase 1 consists of listening tests (error free channel) and a Phase 2 of evaluating the performance in radio channels (errored).

For the listening test the chosen parameters were: 48kHz sampling, sub-channel bit rates (= audio bit rate + short X-PAD associated data bit rate) [9] of 32, 40, 48 kbits/s for AAC+SBR+PS - 48, 64, 96 kbits/s for AAC+SBR - 96, 128 kbits/s for plain AAC - 112, 128, 192 kbits/s for MPEG Layer II.

The listening test procedure [10]+[11] was the one of the MUSHRA test (Multi Stimulus test with Hidden Reference and Anchors), according to ITU-R BS.1534.

The test equipment (headphones, amplifier, equaliser) was validated by IRT [12].

Some of the test results in high end listening conditions of critical extracts with selected listeners [13] are commented in the table hereafter.

Results / audio extract [Slide number]	Comments
Average all items & 95% confidence interval [15]	There is quite a difference between expert and non-expert listeners, but the tendency is the same. The original (01 bar) should be at 100%. The better encoder remains the MPEG Layer II at 192 kbit/s (15). For the experts DAB Layer II at 128 kbit/s (14) is equivalent to AAC+SBR at 64 kbit/s (08). The software encoders (0-12) perform slightly better (cf plain AAC at 96 kbit/s) than the hardware encoder (102-04). There is a real gain with SBR at 96 kbit/s (07); there is a real gain with PS at 48 kbit/s but only for the experts (04) (10).
Electro pop [17]	One of the most critical (high processing at the studio: heavy clipping, leveled before coding). Look (listen) at the difference between plain AAC at 128 kbit/s (05) and AAC+SBR+PS at 32 kbit/s (12) rated under 20!
Female speech swedish [18]	When you listen on the headphones (after it has been encoded) you hear a sort of 'ghost sound' coming through the left channel – this is why the AAC has been graded low. The PS version which takes to AAC mono and then applies to stereos much better.
Drums – Jazz [19]	Quite powerful extract with a lot of high frequency components – the removal of HF (17) is quite critical for people.
Jingle English [20]	Typical of radio broadcast: high loudness, no dynamic, and with probably lot of sounds coming from DJs to assemble them, probably with cascading and so on... with a terrible panning noise
Brass, tympany and castanets [22]	Castanets are very difficult for most of the encoders, even Layer II
Pipe organ Slowly [25]	Experts could not hear the difference between the original (01) and plain AAC 96 kbit/s (02) of the hardware encoder.

This study is here to provide elements for decision, Broadcasters remaining free to choose bit rates depending on their objectives.

Be very careful under 64kbits/s! And be careful on the production side (coding formats, processing).

There are still some open questions: Performance in a cascading environment (tandem coding)? What future optimizations on HE-AACv2 encoders? What can be done for pre-processing? What are the differences with raw HE-AACv2 (with less framing constraints)? And not yet tested: mono with 32 kHz sampling.

1.9 Handling surround audio in 50p production & contribution broadcast infrastructures

Jason Power, Director Broadcast Systems & Will Kerr, Applications Engineer, Dolby, USA & UK

What is Dolby E?

A professional (never reaching the home) cascaded coded audio format enabling the convenient distribution of surround 5.1 audio, through a single AES3 channel in the production and contribution infrastructures, prior to transmission. It's a mature solution, with over 24 000 encoders/decoders shipped (Dolby and partner products) and many infrastructure products compatible with Dolby E. It carries up to 8 audio channels plus sets of metadata specific to each program (to adapt the control of audio in the home receiver and to create a stereo version or a mono version down-mixed) – essential for HD. Dolby E frames must be aligned with video frames so it can be switched and edited without creating clicks or pops. To facilitate these operations, there is a guard band of null data between Dolby E frames, centered on the video switch point. Because the guard bands occur at a 25 Hz rate, switching at a 50 Hz rate risks “cutting a Dolby E frame in half”, and causing a click or 40 ms mute in the decoded audio.

Could we create a 50 Hz Dolby E?

This would essentially be a new format, e.g. Dolby ‘X’, with a set of compromises which are not acceptable:

- 1) In order to keep the same guard interval length, the data payload should be reduced, that means dropping the number of cascades, or carrying fewer channels.
- 2) Moving from 40 ms blocks to 20 ms blocks changes the behaviour of the transform function on which the audio coder is based and that will lower the coding margin.
- 3) Existing hardware assumes 25 or 29.97 Hz. This new solution would require purchasing new devices.
- 4) It would be difficult to 'down-convert' from Dolby ‘X’ 50 Hz to Dolby E 25Hz and to derive where the correct timing of the frame should be.

How to best handle Dolby E at 50p?

By taking care in the design of the system and by intelligent handling (e.g. switching) of Dolby E in broadcast infrastructure products. A Dolby team has recently worked on a set of guidelines for manufacturers suggesting how to enhance infrastructure features in order to handle Dolby E in a 50 Hz environment. A basic Broadcast system [8] with good practices should ensure that all Dolby E sources have the same alignment (when switched from one source to another, no change in alignment) and that the alignment is correct (the guard band is located around the video switch point).

System Considerations.

Concerning progressive video there are several existing references [8]: tri-level sync, Time Code (LTC or VITC) for automation, 25 fps black burst signal (normally used to lock Dolby E equipment). This can all help in the quest of reducing the 50% chance (to switch 50 Hz video in the middle of a 25 Hz Dolby E frame) to a much better value. It must also be noted that:

- A Dolby E decoder, although it expects to receive a stream aligned to a 25 Hz reference, can accept any amount of misalignment on the input stream without producing any audio artefact on the output audio.

- If there is any corruption on the Dolby E input stream, then it is quite difficult to determine how the decoded audio will sound, because the exact location of the error in the bit stream determines how the decoder behaviour results on the output audio – So in the worst case this may be a small glitch, in the best case it may be a 40 ms mute, but the decoder will do its best to try to conceal this error.

One appropriate place to use Dolby E is in the **contribution** system for live/sports events [9]. Some consistent set-ups should ensure:

- That the encoder is clocked to a synchronous reference.
- The locking of the IRD to the incoming MPEG-2 Transport Stream making sure that the Programme Clock Reference is used as a basis by the IRD to decode.
- The mapping of AES data into MPEG-2 TS¹⁴. SMPTE 302M specifies that each audio PES packet should last the same duration as one video PES packet. This is the case for interlaced 25 Hz (40 ms for video and audio) and for progressive 50Hz (20ms) and in some cases that does not cause a problem, but some IRDs try to make some realignment of the PES packets to time them to some local reference.
- The encoder is clocked to input or synchronous reference. We suggest requesting that in the Dolby E contribution mode, the audio PES packets last 40 ms so they encapsulate complete Dolby E frames [9].

After an ingest point or an IRD in a broadcast plant, **frame synchronising** can improve the robustness, by always dropping 2 video frames along with 1 Dolby E frame, and re-aligning E frames to the 25 Hz house reference signal [10].

In the context of the video **switching** router, switch on 25 Hz frame boundaries or parse the Dolby E input to find the guard bands [11-left]. For **editing**: use 25 Hz rate, decode and encode via plug-ins, or use separate A/V edit points [11-right].

If Dolby E is not practical: use discrete audio (e.g. embedded in HD-SDI), with a separate metadata channel; ensure metadata is carried throughout all equipment. Real time and file-based audio processors both require metadata. To get it, SMPTE RDD-6 describes how to transmit Dolby metadata on a real time serial protocol (via e.g. 9-pin RS-485) and SMPTE 2020 specifies the embedding of RDD-6 into HD-SDI VANC. In the file world, the 'dbmd chunk' allows to encapsulate Dolby metadata in a section of a .WAV header.

Equipment for embedding and disembedding audio metadata (per SMPTE RDD6) in the VANC data space (per SMPTE S2020) is available¹⁵. Concerning **SMPTE 2020**, ensure that:

- The Audio (discrete or embedded) / Video timing is preserved [14-right-up].
- The metadata is timed correctly to the Audio it is describing. For example, in the case of a channel configuration change between a 5.1 service and a stereo service, you want to ensure that the home cinema loudspeakers will turn 'on' or 'off' along with the audio changes.
- Channel allocation remains not undefined if metadata is erased [14-right-bottom]. What happens when the SMPTE 2020 embedder loses its serial metadata input – does it switch to an internal metadata preset?

How the samples timing accuracy between **discrete audio** channels may affect audio? You may have to split 6 audio channels over 2 embedded HD-SDI groups, 4 in group 1 and 2 in group 2. What happens if these 2 groups are misaligned? [15]. If there is a similar audio content on all the channels (music/drama) all we get is one signal and a delayed version. And the downstream stereo down-mixes could sound “phasey” with a comb-filtering effect [15-right-bottom].

For **file-based** applications the 'dbmd chunk' to encapsulate Dolby metadata in any .WAV is already implemented in some vendors' equipment software. It can be then re-encapsulated into MXF (SMPTE 382M) via WAV. In the future, XML schemas may be used in automation systems. Possible applications are: postproduction editing, into file-based processors; Dolby E file-based processors relying on dbmd chunk; interchange and delivery of Dolby metadata in files.

¹⁴ Linear PCM or other audio/data (SMPTE 337M – Format for non-PCM Audio and Data in AES3 Serial Digital Audio Interface)

¹⁵ Miranda, Evertz...

All these emerging methods should allow the handling of surround audio plus metadata in 5.0p systems. Additionally, further effort is being made to ensure that the techniques discussed in this presentation are standardised into SMPTE documentation.

2 IT-based production and archives

Chairperson: Wieslaw Lodzikowski, TVP, Poland

Because broadcasters need to deliver richer content across a large number of delivery platforms, production needs to meet new business requirements. Sharing resources and combining best of breed market solutions is key. Will file-based production and new architectures fulfil their promises?

Service-oriented Architecture

2.1 File-based production: problem solved?

Giorgio Dimino, RAI Research Centre, Italy

The concept was introduced 10-12 years ago. The "EBU-SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bit-streams" started to think of the future TV infrastructure based on computer technology. It was a fundamental think tank, which gave birth to most of the concepts and standards that we are using today: compression formats in TV production, file wrappers (AAF, MXF), metadata (SMPTE dictionary, UMID), exchange of content as file (file transfer, streaming)... It formulated the need for a level of "system management services" and for a "Reference Object Model for System Management in order to insure interoperability in the longer term". Up to then, broadcast infrastructure was based on audio/video interfaces and cabling. Now, with IT-based technologies, interfacing is much more complex, you need more intelligence, formats and models. But at the time there was not enough knowledge of the processes we factorised in the view of IT technology to be able to build a sound model. The follow-up of this work was undertaken by several EBU projects [4], providing clear advances.

Many broadcasters have implemented IT-based production islands (self-contained), but very few have been able to integrate all of these islands into a coherent production system. The system organisation is still video-centric in most cases, and sometimes it is even faster than trying to integrate the islands, because of the lack of common interfaces. And even when the system integration is implemented, it is based on proprietary solutions. That means that the technology migration, the extension of the system and of the workflow is a challenge and is expensive. This is because, each time, you have to redo a part of the system integration work. Especially when several manufacturers update their product independently from each other and then apply the upgrade on one part, you have also to upgrade all the others and perhaps rework the interfaces.

As an example, a very simplified scheme with different production islands [6], with different equipment of different manufacturers designed at different times. When you want to interconnect them, you have to define a custom interface at both ends. That may be not very efficient in many cases, because it was not meant from the beginning, and sometimes simply because the formats do not match. So, it is difficult to run the workflow around it. In some case, you want to use some resources that have been installed for another similar facility, you want to put them together... and again you need a specific interface. And when you get rid of one of the components, you probably have to rework the interface. So this is adding cost and you never know where and when this story is going to end. We have also to keep in mind that in the IT world nothing lasts more than a few years. It is not economical to keep running a system that is older, because the maintenance costs are in many cases higher than rebuilding the system with updated technologies...

Our **vision** is to:

- redefine integration as a pool of production resources interconnected over a network via standardized interfaces
- implement workflows via a resource orchestrator that just calls resources and chain them, these workflows being supported by management services

To make this vision become reality, the technology which today seems the more promising is the **SOA** (Service Oriented Architecture) which is becoming popular in many IT domains. Its advantages:

- It uses widespread technologies: a network layer using HTTP, passing XML messages from one service to another.
- It provides a loose coupling of resources. You simply wrap the existing interface of any object in such a way that you can pass a message over a network. You do not need to enter in the internal of the object or to rework the object itself.
- It is platform independent and can be very well deployed over the infrastructure, with no problem with, for example, firewalls – that was instead a problem with previous technologies like CORBA.

All this led the PMC to give another chance to the standardisation of a model, or at least to the definition of a model, which could be the basis for the standardisation of future systems. Therefore, a new EBU project was launched called **P/NP** (Networked Production)¹⁶, with the following main **goals**:

- to analyse the shortcomings of current IT based TV production system integration,
- to collect the missing user requirements,
- to investigate new relevant technologies and architectures in co-ordination with the industry.

The challenge is to design an IT-based production system based on "**data manipulation through loosely coupled network services**" offering interoperability, scalability and evolution. The available enabling technologies comprise: essence formats and container formats associated with metadata models, services with their description, service invocation and discovery protocols, and an Enterprise Service Bus, ESB, which is the basic infrastructure on which all this will run.

The following **tasks** are undertaken to reach this goal:

Task 1 - Strategy for future TV Production systems, providing a kind of Executive summary to disseminate the findings of the project.

Task 2 - Handling of file formats.

Task 3 - Handling of files and streams (exchange) in IT-based networks.

Task 4 - Business process management (P/CP).

Task 5 - Service based system integration.

File formats. A prerequisite for system integration is file interoperability between services. A file format is given by the combination of file wrapper, coding and metadata schemes. Since the industry cannot support all the variants on the market, the users must clarify their requirements and provide a minimal set of preferred file formats to reduce the transcoding need. Even if the standardisation is based on MXF, in practice there are many variants that cannot talk to each other, and probably too many variants.

Networking. One hour of HD video can require 45 GB (at 100Mbit/s) of data or more, depending on the coding scheme used. When moving video content as files from one service to the other, we have to be very careful. If for one reason or the other (e.g. transfer not complete) we have to redo it, this increases the burden on the network and create bottlenecks. Critical operations are to be considered like: file transfer, integrity check, transcoding (can we reduce the number?), security...

Guidelines are needed to properly interconnect systems (in cooperation with NMC), as well as the guidelines on Time Code and Synchronization (SMPTE/EBU Task Force).

Process modelling. P/CP is advancing in the modelling of production processes in news and drama environments. From this analysis the basic building blocks will be derived and described (e.g. capture, playback, transcoder, storage unit, video processor).

Services. A number of functionalities are common to any service, including service discovery, resource locking, status polling, error logging, etc...After having collected user requirements concerning common service behaviour and core services, the goal is to show if the concept works and to provide the 'skeleton' of an open model (vendor independent), which can be enriched in cooperation with the industry, to become a real system.

¹⁶ <http://tech.ebu.ch/groups/pnp>

2.2 Asset Management & SOA @ EBU

Jean-Pierre Evain, EBU TECHNICAL, Switzerland

The EBU and several members have met key players at **IBC 2007 and 2008**: Asset Management providers and manufacturers (Adobe, Ardeno, Avid, Blue Order, Cisco, Dalet, IBM, S4M, Silex Media, etc.). Several questions were identified. From the broadcasters: How could MAM (Media Asset Management) be characterized? What are the key selection criteria and features? To the industry: could the EBU help in defining best practice workflows for News, for drama...? For all: what role will Service Oriented Architecture (SOA) play in the future?

In May 2008, EBU organised the "Latest trends in digital TV production" seminar, which helped to define the business and technical challenges.

For broadcasters, the audio-visual landscape is changing. They have to deal with more delivery platforms (broadcast, mobile, IPTV), more competition. So, they have to maximise the use of all the resources in the production environment. Moreover the consumption habits and viewer expectations are evolving. So, broadcasters have to adapt and keep within range of their audience.

- The **business challenge** include the needs to rationalise and be present on a variety of platforms, to adapt content to the specific needs (usability, availability, etc.), to control production costs ("produce once, publish many?") and to share resources.
- EBU members have to face **the technical challenge** including the needs to:
 - Adapt to business needs and rationalise platform independent production.
 - Combine the best of breed of available tools from different providers (e.g. MAM products are good at managing assets, but very often they are specialised into a particular tool).
 - Maximise reuse of well defined common resources by similar 'roles' having similar 'needs' across different production units;
- Support modularity, scalability, evolution capacity to allow, maintenance, upgrade and customisation (e.g. an MAM provider develops customised 'patches' for a broadcaster – but what happens if the vendor comes to the next MAM generation. So, if you have a more modular architecture, like SOA, you have then a more clever solution to deal with this sort of problem).
- Modularise functions for more 'agile' workflow orchestration.
- "Start small, think big!"¹⁷.

Some broadcasters are already working with SOA, but to a certain extent using proprietary solutions. So we want to investigate now how far we can go to have real interoperability when using the SOA concept, by:

- sharing knowledge on Asset Management and SOA (since may 2008);
- starting EBU project on file-based production and SOA-like architectures (now!);
- establishing a network between broadcasters and the industry (to be continued)

The **SOA proposal** is to provide:

- An environment within which you can combine heterogeneous functional tools (legacy and new equipment, tools from different manufacturers, software platforms, asset management tools, in-house developments)
- A better management of metadata collected through well defined interfaces and contributing to each broadcaster's data model.
- Modularity and scalability, a box of tools exposed as 'services'.
- Flexible workflow management through 'service' invocation. Since all the different functions are available, different workflows (which can correspond to the different production units) can be easily reorganized.
- Easier maintenance and higher ability to upgrade the production system.

¹⁷ E-L. Green, SVT

SOA makes sense in a file-based production environment.

SOA has the potential of a standard if it is implemented according to common rules. But 'what is' and 'what means' **SOA compliance?**

Step 1 - In order to define **the process**, we take the OASIS reference model [5], presented as "an architecture paradigm for organising and utilising distributed capabilities that may be under the control of different ownership domains..."

The EBU work is compatible with this model:

- "The 'ownership domains' mean, for our broadcast environment, different tools from different providers or in-house development.
- At the input [5-left], concerning the 'Requirements', the EBU is collecting members' requirements (What do you need? What would you like the system to do?)
- Concerning the 'Patterns' [5-Center], if we speak about the business patterns we can refer to the work of the P/CP group, analysing common processes (almost finished for News, in progress for Drama). Concerning 'Related Models', e.g. Metadata, EBU is still working on metadata models and processes analysis (P/CP & P/MAG).
- The related work around the 'Protocols', 'Profiles', 'Specifications', 'Standards' [5-right], this is also obviously what EBU does.

Step 2 – Defining **business patterns**

Starting from a simplified overall broadcasting production model [6], EBU is producing detailed business patterns for News and Drama. See, for example, the more detailed analysis of the Ingest process for News [7]. The difficulty is to decide how far to go and where to stop. For the time being we have a quite complete set and we are even working on the metadata flow through the different interfaces for the different functions.

Step 3 – **Web services** [8]

The next step is what SOA is all about: exchanging messages, exchanging information, activity, functionalities. The definitions of the Web services are the core of SOA.

- This is important to know which Web services are available - the visibility of the Web services is an important criteria. Then how you can reach these Web services – where are they located? How you can activate them – through which interface¹⁸? What you can expect from them – referring to the description part.
- The more technical part of the description of the Web service concerns the actual activation of the functionalities, with 2 levels of description:
 - The behaviour model is a representation of the functionality: what you can expect? What is going to be the real-life effect of this particular Web service (activating a particular system or sub-system)?
 - The information model concerns metadata and system parameters – which information do you need to send to this Web service to activate it and which information do you expect from this Web service?
- The real world effect is the actual process and expected results.

Compliance will require the agreement of common web service description rules and formats!

Web service definition: "a mechanism to enable access via internet protocols to processes via an interface described using predefined rules and procedures".

As a typical example of a function eligible as 'Web service' take the '**Ingest**' [9]

The device is the camera from which you wish to ingest the content. Then, the Web Service Interface (WSI) with 3 levels of description: the binding protocol to interact with this service (SOAP over e.g. HTTP or FTP), the behavioural model – what you expect from the service (wrap and packetize content), and the information model (technical audio-video parameters, and other metadata e.g. automatically generated). And this is how you can make this functionality available on the network as a Web service.

¹⁸ The service interface is the communication element through which services will be activated (with or without parameters) and through which information (metadata and states) will be returned.

Altogether the **EBU scope** is the following one. We started to discuss on asset management. At IBC 2008, some vendors told us “we are already working on SOA, and our Web services are publicly described”. Other said “This is all the know-how of our company and we do not want to disclose the way we manage the different functionalities when we pull up one of the Web services”. One of the company is developing its Web Service Interface as a big bag and you activate only the part of the functionalities that you need according to the Interface on which you are working. So, the only way is to have a very high-level abstract Web service description language. The diagram [10] with the Enterprise Service Bus (ESB) in the middle, the 2 layers of the abstract Web Service Description Language (WSDL) represents a similar approach as the one from IBM (§ 2.3), and the lower layer is to be managed by different people to be connected to this SOA layer (cf. Cisco - § **Error! Reference source not found.**).

Starting from this complete picture, what do we want to do? First, some of the MAM providers are tempted to take over the layer of the abstract description language. We do not want them to do that. We think it is not beneficial to the industry and we want it to be open. Second, what can we do to describe this directory of services? This is where you should know which services are available, if you want to refine some workflows. And finally, because we have all these problems of interoperability in MXF, we also have to deal with the lower layers. In order **to reach a ‘plug and play’ service description, discovery and use**, we propose to:

- Investigate possible solutions for a common abstract WDSL
 - Recommend a preferred protocol for Web Service access (<binding> definition and SOAP parameters).
 - Recommend a common approach to describe the operations / functions available through the web service (<portType>).
 - Recommend common rules and formats for message exchange (<message>) and common datatypes (<types>).
 - Harmonise service localisation and associated network definitions.
 - Support mapping to publicly defined or more abstract WS interfaces from different MAM providers or manufacturers.
- Register services in a common directory (adapting and restricting the UDDI concepts to production)
 - Provide harmonised WS description about functionalities, requested parameters and expected effects.
 - Provide localisation information.
 - Support additional profiling (contextualisation) and access information.

An unexpected potential bonus: **a metadata logical reference model** [12]

We have been working with the identification of metadata at the different interfaces. Considering the different steps in production, you get technical metadata on the video format, then some editorial information, some Edit list, and finally publication data. So, all these metadata that you may collect now through the Web services is going to contribute to your overall data model in your Broadcasting facilities. What we have done at the start in EBU was to develop metadata specifications that look at different models and tried to make THE metadata model. But we are a little stepping back from this position (although we have more EBU members using P/META that we are still supporting and maintaining). But on the other hand, because of the impact of Web services on metadata, we now want to have a much higher Common Logical Data Model (CLDM). It is not a question of structuring this data, but of understanding what is your amount of data that participates to the logical data model. By doing this we still benefit from the experience we have gathered embedding the metadata specifications and from the experience of the EBU members. But this logical data model could become a common reference to allow Broadcasters to discuss between themselves or with 3rd parties like manufacturers or MAM providers. For instance, if a broadcaster has a data model, he would map it to this CLDM. A MAM provider mapping its data model to this CLDM, can then compare its data model to all broadcasters’ data models, because he has one common reference to which each broadcaster has mapped its own data model.

Conclusions

File-based tapeless production is becoming a reality, but issues still need to be addressed through additional rules and guidelines, and EBU can help.

Tapeless production is a trigger to develop new architectures and improve asset and workflow

management, giving more control to broadcasters:

- You have the know-how, manage the production your way!
- Get what you need (even if you have not the necessary R&D capacities) and not only what is 'available' (which is not necessarily doing everything what you would like to do in the way you would like to do it)! SOA is one chance to give you more flexibility, to get control again on the development of your systems. Take the best from the different providers!
- Give your metadata its strategic dimension!

Will Service Based production fulfil its promises? Watch this space, P/NP will challenge the concepts (such as 'claimed' flexibility)!

The goal: To re-adapt in the production domain the concepts of 'plug and play' and 'content and service discovery', that we have today in the distribution domain.

2.3 SOA Media Enablement - a media specific SOA framework. From ESB to abstract service description

Dieter Haas, IT Architect Media & Telco, Industry Technical Leader Media, IBM, Germany
Frank Schaffa, IBM Research, Mgr. Multimedia Communications Systems, US

In the media business environment, the integration of new resources and applications and the automation of processes [3] face rigid architectures. This makes it difficult to adapt new technologies and achieve a level of flexibility to meet today's challenges. Maintenance of those grown infrastructures where resources are connected in a point-to-point approach is another issue and demands a conceptual change - in a real case, there were 32 MAM applications with 205 inter-applications connections [4]! Therefore, the objective here is to explain the additional capabilities and benefits of SOA when applied to media processing especially in the sense of reusing resources versus dealing with fixed and hardened production flows.

Our approach is based on the OASIS definition for SOA as "a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations". It is based on following **principles**:

Loose coupling	Services maintain a relationship that minimizes dependencies - one can use the applications and wrap them with the appropriate adapters and Web service interfaces to run them.
Autonomy	Services control the logic they encapsulate (self sufficient) – the adapter itself keeps the logics encapsulated.
Contract	Services adhere to communications agreement (service interface) – the service interfaces need to be really stable so that everybody can rely on this content.
Abstraction	Services internally behave as black boxes with high granularity.
Reusability	Services to be architected for reusability (contract, abstraction). If one wants, for example, to use a transcoder as a service, it is not just in one situation, ideally it is in as many processes as possible.
Composition	Assembly and sequencing of services to form composite services - one wants to be able to combine process steps to a composite service.
Discoverability	Services have to be able to be discovered (description, registration) - if one wants a proposed or exposed service, it has to be discovered, otherwise it is hidden and hardly anybody will be able to use it.

SOA today is well established and works fine with many business processes. SOA is understanding and handling messaging exchange, calling the services, etc. but SOA today does not understand anything about associated media objects and their processing or transport. It is this kind of '**media awareness**' that we want to bring into the SOA business and into the entire complex metadata. And we wanted this media awareness to enhance the SOA layers, starting with the Enterprise Service Bus (ESB), to understand about media. So, if we come back to some aspects of SOA benefits, what does that mean in the media context?

SOA Benefit	Media-Aware Benefit
Loose coupling of applications	Media applications produce/consume both content and metadata (messages). A media application may have a 1 GB video file that can't be understood / handled in a standard SOA SOAP message. A media-aware ESB synchronizes the capture and delivery of both between services. We obviously do not want to move large media gigabytes files through an ESB bus – we have to find a way to synchronise the services with the content that has to be processed at that time.
Service abstraction and mediation	A media-aware ESB has the ability to "inspect" the media through its metadata and leverage the appropriate mediation services with dynamic runtime service selection (i.e., dynamically invoke the most appropriate service/route for the media) – it knows what has to be done with the media.
Workflow persistence	A media-aware ESB manages both the transaction flow and the media essence transparently between services. SOA provides dynamic routing capabilities: the messages are routed through the infrastructure to the appropriate service - it should as well ensure the delivery of content to the right place, but this requires an extension.
Transformation and mediation	A media-aware ESB transforms both the message (metadata) and the media essence to meet the requirements of a service. When a message is routed through the infrastructure, it gets transformed by this architecture to meet the format of the target service - media content has also to be implicitly converted from one format to another. This has to deal with adapters, with the infrastructure requesting, responding appropriately to partner applications.

In order to realise these benefits, we enhanced our standard Websphere SOA framework with appropriate Media extensions to achieve a Media Industry SOA Solution Framework - called **Media Hub** [11], which links business and content processes to support end-to-end workflows for media and other enterprises.

The media extension could be considered as 2 conceptual enhancements:

- Media Awareness
- Abstract Service Definitions

'**Media awareness**' means that the entire components that are relevant in this process (like registry, like ESB, like services) need to understand what the media is about, what it is, what format, what type it is, what the size, etc. Therefore, we need some additional information describing the content characteristic that is provided with the message running through the infrastructure and that is provided in the registry that identifies and describes the service. To describe the media content we use **MPEG-21**¹⁹ [12]. It is an open standard, it is applicable and used by other industries. It provides the capability to describe the media in XML-like format. It can be used separated from the essence itself, so that we can use this description in the ESB, in a message format running through the infrastructure – the essence being moved in a separate one. The MPEG-21 DIDL (*Digital Item Declaration Language*) structure is used in this context and it might be very complex [13]+[14]. This structure contains all the information (metadata) about a media object (essence).

'**Abstract Service Definition**' (ASD) is about combining same category of services into one class. For example, we want to deal with a general interface for a transcoder, regardless of the specific implementation. The benefit is to easily exchange one transcoder with another one without necessarily touching the workflow. It is just within an adapter which has the abstract interface and beneath the specific interface. So we focus here on the function, not on the proprietary interfaces, and that help us to manage the resources. How does it look like? The ASD comprises 2 major components [16]:

- The 1st is the service class design, which is the specific class (e.g. transcoder / watermark / data mover...);
- The 2nd one is the specific mapping from the class to the specific service provider API.

In term of operation we need an 'Adapter' between the 'Orchestration & Monitoring' and the application itself [17]. This adapter has the 'Abstract WSDL' (Web Service Description Language) and has the 'Adapter Logic' inside which matches to the entire application at that point.

To recapitulate [18]:

- we started from an ESB with a mediation flow, the message models and the communication

¹⁹ MPEG-21 Multimedia Framework (ISO/IEC TR 21000-1)

protocols;

- we extended with the media enhancements, which is based on MPEG-21 Digital Item Declaration, a metadata registry for the semantic representation of the services, and the abstraction of the process orchestration of these services;
- and, on top of that, we have the abstract service definition for the media for various service classes (transcoder, etc.).

This makes it easier to exchange a specific service instance if we for example want to introduce a new transcoder. Of course, we need to look for the new transcoder application, we may need to write the adapter and publish the service to the registry... the rest remains. At the end we have a Media Hub, a media-enabled SOA infrastructure and a solution framework which is flexible enough with the media extensions to support media enterprise business. For more information, an IBM **Redpaper** 'Abstract Service Definition for Media Services' is available²⁰.

Let's look on the benefits of this solution with some **examples of a media-aware abstract service selection**.

- If we for example model a process sequencing [20-left/left], starting from a service A with a content object to the next service B (e.g. watermark). At that point we only need to model the abstract service 'watermark'. In the infrastructure there might be several instances 'watermark' like a watermark for audio, a watermark for video, etc. Due to the information that is carried inside a message about the actual media, the infrastructure is capable to catch this information, look into the registry where the various instances are described, pick out the appropriate instance that matches for this format and pass it to the instance [20-left/right]. So, the physical sequence looks different and more complex while the model [20 – left] is simple and at an abstract level that a business person can handle.
- As mentioned before we need support for transcoding media, implicitly in a similar manner as it is done by the infrastructure with the messages (usually by XML style sheets transformation) [21]. We do not want to build a transcoder into the media extensions – there are transcoders which are good at doing the job (Telestream FlipFactory, Rhozet...). So, we enabled the infrastructure to use this service implicitly.
- Starting again from service A to service B, which is 'Playout'. Let's assume in A we have a media in the compression format MPEG-2 and I want to playout it in MPEG-4. The infrastructure recognises (due to the MPEG-21 information) the content characteristics in the message that we are dealing with a MPEG-2 media object, but that the next service requires a content characteristics MPEG-4. It recognises a format mismatch and the need for a transformation. It also recognises that the essence is in the wrong place and needs to be moved. So it looks for a service capable to do a data movement. Here again, the idea is not to implicitly integrate this in the infrastructure but to integrate existing application services (Aspera, FileCatalyst, Signiant, FTP adapter, etc.). So, the infrastructure recognised the format and the location mismatch and supplies implicit additional processes for data movement, for transcoding to arrive finally at the playout. That might look complex in this way [21-right]. We modelled [21-left] the same process going from a repository to a publisher playout – and again with a format mismatch and a location mismatch, that the infrastructure resolves. It simplifies the entire process. Of course, we could also model the right hand side with Web services, but if we want to modify the publishing format we need to look for a new transcoder, to write a new adapter, to publish it into the registry, to change the workflow, to display the new infrastructure, and to re-test. It is a bit simpler with the left hand side model. Of course, we need to look for the new transcoder, to write the adapter, to publish to the registry... but the rest remains.

To conclude, we presented Media Hub, a media-enabled SOA infrastructure and solution framework based on Abstract Service Definition which is flexible enough with the media extensions to support your business.

²⁰ IBM Redpaper - Abstract Service Definition for Media Services
<http://www.redbooks.ibm.com/Redbooks.nsf/RedpieceAbstracts/redp4464.html>

2.4 Medianet technology – The missing link from SOA to file-based production

Dimitris Papavassiliou, Head of Digital Workflows Solutions, Media & Broadcasters, European Markets, Cisco

SOA as a concept enables the dynamic and collaborative workflows [5]. However when we are talking about SOA Web Services, we are talking about applications capabilities, we are talking about the way for applications to communicate, but we are not talking about the actual communications. Communications are not guided over the ESB, never meant to, communications are guided over a network, and a network in a SOA architecture is essentially a SERVICE, a bunch of services: connectivity service, virtualization service, security service... And this has always been the case in the IT where the actual movement of data has been guided through a network service element.

However, when we are moving to a media space this communication is becoming more complex. It is not just because video loads the network. it changes the network; The network has to react differently on video, because it has more stringent requirements than any other traffic type and application so far. So we are addressing this challenge with a **medianet**, a "**Media-aware network**". Our initiatives for medianets are not just around production, not just around the media industry, it is around all industries and also the home, it is our overall drive behind our video strategy.

Why medianet? On one side [6] users have more demand for video, for more video applications, for more video devices, and on the other side networks providers, media providers, service providers have to optimise the quality of experience, to reduce complexity and to accelerate the deployment of services. So it is important to introduce a different sort of network that is able to handle this. Medianets optimise networks for the dominant traffic type: the CISCO VNI projects that video will amount to 90% of the network traffic by 2012, and during the last Summer Olympics Games over 3600 hours of content were produced by NBC (more than the total coverage previously accumulated) and most of this was over the Internet [7]. Consumers are driving requirements, as they are looking for a more visual, social, personal and interactive experience, new requirements on all networks. These requirements are feeding back on how we are building networks in the service provider side, in the media side, and how application vendors have to think about networks. So this is a medianet [8] – a network has not just to be network-aware, it has to be media-aware, it has to be end point aware. Medianets are created by implementing new technologies to converged IP Networks able to support rich media services.

We are encompassing four over-arching pillars for medianet technology[9]:

- Transforming video experience, to differentiate end user experience
- Media aware IP NGN, to ensure end user experience
- Virtualization, to manage complexity and scale;
- Monetization, for new revenue streams

Focusing on virtualization [10], we are speaking about benefits in production, contribution, distribution and user experience. As an example of virtualization in the file-based production is the **Unified Fabric** innovation we are bringing in our Data Center 3.0 technology [11], a lose-less Ethernet Fabric, able to carry both Fibre Channel, Ethernet traffic and even inter-process communications over a 10 Gbit/s Ethernet interface. Unified Fabric can minimise cabling for a more efficient, simpler, greener operation, reducing the total cost of ownership - just one cable supporting all communications types. Unified fabric is based on a series of standard-based technologies.

There is a technology evolution in building medianets. First a '**Converged Media Ready Network**' [13] is about a well designed, integrated and verified end-to-end solution. It is a network architecture to support the media workflow applications in their operations – how actually, media essences, Web services, signalling, metadata, flow within the network. This is the Media Workflow platform architecture, a validated design to support multiple digital workflow applications over a common converged topology. It is based on Data Center 3.0 innovations (virtualization, Unified Fabric, application acceleration) and it consists of architecture blueprints for end-to-end solutions. One use case is the blueprint for the Avid application suite [14].

The next step is the **Medianet Service Interface**. Its objectives are, to:

- Enhance the media application development, deployment and use by highly integrating media applications with a medianet infrastructure.
- Provide a comprehensive and consistent service interface to access the network services.

Basically there are the application domain and the network domain, but they are not aware of each other. So, the application assumes there is a network and the network knows it expects some request directly from the application, but in a sense they do not know each other. The aim here is to create this set of application interfaces with the middleware, and particularly a software tools stack provided with the application, in order for the application to explicitly invoke network services [16].

From a service definition point of view, we are looking at Video Network Services [17] like 'Quality of experience' services (QoS, etc.), 'Security' service (Identity, etc.), 'Session control' services (Scheduling, etc.). The workflow is dynamic, not static. Supposing, you want to do something at a certain point of time and you want to notify in advance the network about this activity and the network should reserve the required resource to perform this activity. In this context an application will be able to explicitly require services from the network and the network will provide the service to the application. And of course for a legacy applications, the network provides the same services implicitly, based on policies - policies defined in advance [18].

The objective of the **Adaptive Media Aware Network** is to enhance the support for media applications by reacting, by providing advanced media-aware functionality for key network services like admission control, routing, monitoring, and resiliency mechanisms. So, that media-aware services can adapt to real-time usage and requirements, and can optimize infrastructure support or provide options to applications and users.

As **use case**, let's look at a telepresence application [20] with HD 1080p video 6Mbit/s streams providing lively user interaction. The network has to provide the resources to support a telepresence session. Let's look how an adaptive media aware network can react here.

(1) It recognises the type of traffic.

(2) Then the end point provides some information, for example about the active 'face' screens It is up to the network to disregard the streams for inactive screens.

(3) And probably depending on the usage of the network, as to adapt to the changing conditions, a decision has to be made of dropping packets. So, the network has to understand what part of the content will have minimal impact here to the video quality, or decide to discard other network traffic.

(4) The network has notified the end point, the other side of the communication, that there is an invitation to fall down on SD, because there is not enough bandwidth.

For these kinds of interaction, the network has an understanding of what kind of traffic is carried over and adapt on the existing conditions.

with the medianet technology we are looking at improving efficiencies by optimising CAPEX (capital expenditures) and OPEX (operational expenditures) as well as looking at new functionalities to enable new services [21].

EBU/SMPTE Time labelling and synchronization

2.5 EBU-SMPTE Task Force: The (almost) final report

Hans Hoffmann, EBU Technical Department & Peter Symes, SMPTE, TF co-chairmen

Both organisations, SMPTE and EBU, recognised the need to address the issue of sync and Time Code. There are huge difficulties in synchronising facilities, particularly in the multi-standard environment (e.g. HD with 3-level sync, black burst problems...) and with the trend of moving to IT infrastructure. They decided to bring their forces together and set up a **EBU-SMPTE joint activity**, similar to the Task Force initiatives of the past (Rec.601 and harmonised bit-streams), to achieve results much faster. This Task Force works clearly on a next generation system that will come in place not tomorrow, but will provide the foundation for interoperable sync and time in about 2-5 years

Why did we undertake the work?

- The current reference signals are about 30 years old and are based on colour black. They rely on zero crossings of 3.579545454 MHz and 4.43361875 MHz and require a dedicated infrastructure
- This solution does not support multi-TV standards (e.g. 1080p 50Hz running in the infrastructure with a sampling frequency of 148 MHz!) and it is not easy to sync Audio and Video. The future digital, networked and multi-standard media creation and production environments definitely require a new form of synchronisation signal.
- The current Time Code signal is also about 30 years old and has been many times modified (version 20!) and tweaked,. At the beginning it was designed for linear audio tracks and not for the video. It does not support frame rates greater than 30 Hz (imagine a system running at 50/60 KHz and even higher in the future...). It has found many "interpretations" in the market, being implemented by certain manufacturers in very "individual" ways. The future digital, networked and multi-standard media creation and production environments also require a new form of time labelling.

At the start of the **Task Force project** [6], over 100 people subscribed to it, but it came down to a core of 20-30 active parties from broadcast, cable, telcos and users. The TF defined 'User Requirements' (UR), published then in the form of Request for Technology. It got 6 responses from industry, including IPR (Intellectual Property Rights) declarations. After mapping them against the UR, a first proof of concept using IEEE1588 (standard for moving a synchronised session over Ethernet - § 2.8) was presented. The work should be finished in March 2009 and handed over to the SMPTE for standardisation.

2.6 Request for Technology & first agreements

Friedrich Gierlinger, Production Systems Television, IRT, Germany

A Request for Technology was formulated and published (March 2008). The table hereafter lists examples of User Requirements.

General user requirements	
Intellectual property rights (*)	Respondents to this RFT must declare any patents known or believed to be essential to the Implementation.
Software platform	Users shall be free to make their own software implementations of the standards without dependence on a particular operating system or hardware platform
Transition to use of the new standards	The transition from current to new standards should be achieved in broadcast production plant with infrastructure based on current standards.
Continued availability	The proposed technology shall have a high likelihood of continued availability, or availability of backward-compatible technology, for the foreseeable future.
Basic Value and Economy	The proposal should offer significant additional value when compared to the existing colour black system.
Universal Format Support	The synchronization signal must convey sufficient information to generate any appropriately specified video or audio standard
Deterministic Phasing between multiple systems	The system must provide deterministic phasing of all current video and audio standards. It must be able to accommodate potential future standards (e.g. based on arbitrary frequencies) without change to the synchronization signal.
Frequency reference (*)	...A global frequency reference if possible
External lock	The proposal must provide for master generators that lock to an external reference frequency, and specifically it shall be possible to lock to a global time/frequency reference, such as GPS.
Frequency accuracy and stability (*)	The proposal must support frequency accuracy (at least) sufficient to meet the most stringent requirements: currently this is the PAL system, requiring accuracy of <1 Hz at subcarrier frequency, or approximately 0.225 ppm... at the moment we don't know how precise, accurate the new system should be.
Time reference	
Time of day	The synchronization signal shall convey sufficient information to provide a "time of day" clock with date information to the slave. In addition, the synchronization method must convey sufficient information to convey local timezone offset from UTC as well as a Daylight Savings Time (DST) flag which would be used in conjunction with UTC to determine the actual time-of-day in a facility.

Leap second and DST management (*)	The proposal must provide for appropriate management of leap seconds and Daylight Savings time, which is specific to geographic / political region.
Extensibility	It is likely that over the required lifetime of this standard there will be the need to transport additional data specific to the extension of the capabilities of the system. The system shall provide a mechanism for extensibility of the transported data to accommodate future requirements
Compatibility with legacy systems (*)	It shall be possible for a slave system to generate legacy synchronization signals such as colour black that meet all existing standards.
Synchronization signal transport considerations (*)	The synchronizing system should not necessarily require its own infrastructure dedicated to the distribution of the synchronization signal. This preference could be met by using an infrastructure that is already in place in an existing plant, or that would have to be provided for other reasons in a new plant.

Responders to the RFT were: Harris, Skotel-Edlmax, Symmetricom, Sony, Thomson Grass Valley. Out of these responses should a common solution be developed and the responses have been intensively discussed and evaluated.

The proponent X offered two solutions. The basic idea was a 3-layer system [7] with a 3-level sync or a more developed black burst signal driver for the first solution ('StreamSync') and a 1588 network interface for the second version ('NetworkSync') [8]. The proponent Y provided a solution [9], via IEEE 1588 network which is more precise. This proposal can be synchronised with GPS or with an analog black burst. The proponent Z designed a layered model [10] containing counters synchronised with different possibilities (GPS, PCR...) and a transport via a network.

The 1st common solution [11] was divided in 3 sections: a 'Master generator' synchronised (with GPS or the old black burst), the 'Network' (IEEE 1588, or streaming via coaxial cable) and the 'Client'. The 'Client' should be able to generate timing signals as well as **Time-related Labelling (TRL)** signals out of the signals coming from the networks.

The agreed **Common Synchronization Interface** divided in the sections 'Master'/'Network'/'Slave'. The network is either a streaming network on coaxial cable or a IEEE 1588 network. The Transport layer provides the necessary network drivers. The layer above is the Session layer, where all signals needed at the client side must be inside the 'Common Synchronization Interface'. It contains data coming from the 'Cyclic counter', 'Time count' and additional 'Control data' of the Presentation layer. In the Application layer, a possibility to synchronise the system with GPS or with the black burst signals must be available. The 'Client' side must be able to generate all needed sync signals and TRL signal out of the CSI-signal which comes over the network.

A plug fest will be organised with the different manufacturers to see whether the systems works together before the standardisation starts.

2.7 The Time-related Labelling (TRL)

John Fletcher, BBC R&D, UK

The SMPTE 12M Time Code looks like a time (hours/minutes/seconds/frames). Actually it is a count of "frames since midnight". It has severe limitations: limited support for higher frame rates (≤ 30 Hz, possibly 50-60 Hz, not really supported beyond), the labels are only unique in a 24-hour period (and in many applications you may want to record longer than that), there is no indication of frame rate, and it has limited support for multiple labels (e.g. acquisition time, time along the tape, film edge code, etc.).

The two main uses for time labels are:

- Synchronising independent recordings, like multiple camera recording [6], separate recording of audio & video) by labelling the recordings with the capture time (i.e. the "time of day" Time Code).
- Identifying temporal position into material: log sheet of events happened during recording, edit decision list (EDL to identify particular frames for edit points), at what time the subtitle appears in the programme...

Is the label to be based on **time or frame count**?

Frames (or other media unit)	Time
(+) Obvious way to index material, like pages in a book	(+) Same for all material types. Very good e.g. for the multi-camera capture – the label will match regardless of the different frame rates or type of application.
(-) But different for different material types: audio and video media units different... and establishing the correspondence between labels may work not so well	(-) But numbers don't increment simply – the frame rate may not be exactly locked to your Time Code.

You may think it does not matter to choose time or frame count, because you can convert one to the other world, but it is not so straightforward as one may think.

The phase of the essence signal which has been labelled makes a difference. If your decision boundary for whether this time matches to one frame count or to the next is close to the actual frame boundary, there can be difficulties.

And you must rely on a constant frame rate exactly related to the time.

To address these questions of frame versus time, it was decided to basically include both types of labelling, depending on the application, in 2 proposals.

TRL Type 1

Includes: a timestamp to high precision fraction of a second (960 Hz resolution = ~1.0416 ms), a sufficient size to count up to AD 2117, information about time zone, leap seconds etc. It also includes the media unit rate (nominal rate).

Use: e.g. acquisition time, for variable rate, over/undercranking or all speed cameras.

TRL Type 2

It includes: a media unit number which is basically an incrementing count (e.g; frames), the media unit rate (can only be the nominal rate), plus the time stamp of the 1st unit you have labelled or phase datum (allowing to calculate the current timestamp, assuming your media unit is locked to time and constant).

Use: e.g. postproduction EDL.

Binding

There is no use defining a label if we can't store it or carry it throughout the system without it being lost. It is not too difficult to include one additional field, or to add a bit of extra information to files formats, to packet data. It just become more difficult with synchronous streams (such as SDI video, or AES3 audio, etc.) with a constrained data size and with plenty of devices which will strip off the information.

2.8 How can you possibly synchronise a TV plant using Ethernet?

Bob Edge, Manger, Standards and technology, Thomson Grass Valley

The computer industry has been working on IP network time synchronisation for decades. Most IP network solutions (protocols) have accuracies measured in milliseconds, digital TV plants can operate with 50 nanoseconds or more of jitter. Why is IP network jitter so difficult to control? How does IEEE 1588²¹ solve the problems?

In an ISO layered network [4] with a master clock on one device, timing protocol messages start at the application layer, then are then sent through the layered software network stack, they are then transferred on a physical network as a packet which is received by the layered software network stack on the receiver side, and finally delivered to the application layer in the receiver. The network timing protocol manages these packets at the application layer on each device. In an ideal world application

²¹ IEEE standard for a Precision Clock Synchronisation Protocol for Networked Measurement and Control. 2002 / July 2008

layer packets would arrive with constant delay times [5]. If networks are not heavily loaded, there is a distribution of the times that it takes to the packet to get from the 'application' layer on one computer to the 'application' layer on the other [6]. There are several things which result in transport times being unpredictable. If the network is overloaded, this time is going to be even larger [7]. If the network is saturated, these times can be seconds to transport a packet [8]. In fact, the packet can even be lost. There is a significant variation in the end-to-end transport times on different networks.

Most of a network stack is implemented in software [9]-[10]. Engineers and managers cannot figure out how long it takes to write software and we cannot accurately estimate how much time it takes for software to run in a loaded computer. When a packet is transported on a fibre or a copper cable, it moves at the speed of light on the specific media. So the delivery time across the fibre or the copper is a constant. As soon as the packet is received, we are back in the software world where non-deterministic timing occurs... As the packet moves down to the software there are unpredictable timing, constant time on the fibre, and unpredictable times in the receiving computer's software network stacks. [11] (the pink colour on the slides showing what is predictable and what is not [12].)

NTP (Network Time Protocol) and other network timing protocols, use a special packet which is constructed by the application layer, goes down to the network protocol stack and across to the network to the receiver [13].

This process is equivalent to taking a good Swiss watch and trying to synchronise it with an international time standard using the Post! You do not acquire accurate time or have good jitter management. You might get that watch synchronized to the right day, but it might take a month for that to happen...

So why is **IEEE 1588** different? A packet starts at the application layer [14] and the packet is moved from a memory buffer onto the network and at a fixed place in the packet a high precision clock is inserted. This is like time stamping a truck as it leaves a warehouse. When the packet gets to the receiver time is extracted (the sender's time stamp plus the transmission time is also recorded).

You can implement IEEE 1588 in software by placing parts of the protocol at the network driver layer [15]. This eliminates some of the timing jitter. So, a software IEEE 1588 implementation is better than NTP but it is not as good as hardware IEEE 1588.

With IEEE 1588v1, the time stamps are inserted and extracted at the network hardware interface. The unpredictable software run times do not impact the transport times. In addition the high-level protocols use these accurate time stamps to lock the receivers' clock rate to the master clock, to calculate the physical network "round trip" times, and this information can also be used to lock the receiver's clock value to the master clock.

What happens on a large Switched IP Network [17]? Network switches add more timing jitter. When a packet is received by a switch, that packet is stored in the switch. These are switching decisions for IP routing that add unpredictable delay times. Furthermore, the packet is held in the switch until the outbound port is available [19]. IEEE 1588v2 offers a solution for unpredictable packet routing times [20].

For example; when the packet leaves the application layer it starts at an uncertain time. As it is moved onto the network, you start the stopwatch. When the packet leaves the network and is captured in an IP router, you pause the stopwatch. When the packet leaves the switch, you start the stopwatch again, and at the receiving device you stop the stopwatch again. Now you have all the transport times (time the packet spent on the fiber or on the wire) and this other time is taken out by the high-level protocols.

In summary:

- IEEE 1588 can be used on IP networks with other traffic; IEEE 1588v2 can work in large switched IP networks with normal network loads
- The self-discovered network round-trip times can be used to back-time a facility.
- IEEE 1588 improves timing precision from milliseconds to nanoseconds by using time stamps recorded by the network interface hardware as packets go on and off the wire.
- IEEE 1588 is being used by other industries (factories with robots, instrumentation managed through

Ethernet, power companies for power grid management...)

- A few broadcast equipment vendors have built proof-of-concept systems.

Using IEEE 1588 is a good path forward to synchronise digital TV plants.

Digital Archives

2.9 What replaces shelves: solutions for long-term storage of broadcast files

Richard Wright, BBC R&D, UK

The PrestoSpace project was about the 'Preservation factory' concept [5]. It had many areas of work: Digitisation, restoration, metadata, storage... [6]. But first of all, the content on our archives shelves is very much at risk: about 70% of the material is concerned with obsolescence, or decay, or fragile. 30 million hours of content were specifically identified by PrestoSpace and the European project TAPE²². UNESCO extrapolated and estimated 200 million hours worldwide in audiovisual collections. This is why digital storage for the preservation of this material²³ (except may be some film) becomes a critical issue. So, **how much digital storage** would we need? In the table hereafter are the BBC weekly requirements, beside its legacy archives (650 khours video + 350 khours audio + 2M stills) [8]-[10].

Summary of Storage – now (BBC production, archiving, preservation)		Summary of Storage – soon	
Standard Definition	Storage Requirements	High Definition (~ SD x 4)	Storage Requirements
Raw Material, 10 khours (30 hrs/1hr drama series)	1000 TB/week	Raw Material	4000 TB/week
Completed Material 1 khour/week	100 TB/week	Completed Material	400 TB/week
Archiving 300 hrs (Legacy) Digitisation 800 hours	30 TB/week 80 TB/week	Archiving Digitisation (Digitising old material is still in SD)	120 TB/week 80 TB/week
But – only Archiving and Digitisation require permanent storage	= 110 TB/week	Requirement for permanent storage	200 TB/week

The storage requirements for audiovisual preservation are huge - in Europe: 50 million hours (20M video, 20M audio, 10M film) – worldwide: 200 million hours. Assuming following digitisation parameters: video at 200 Mbit/sec (“Rec.601”), audio at 1.4 MB/sec (CD quality), film 2k (1.5 Gbit/sec), saving 1/3 of this material brings to a total of 600 PB + 4.2 PB + 2400 PB!

What is happening to storage systems?

- Storage capacity goes up according to the Moore's law) [12].
- Media/Device cost (e.g. cost per gigabyte) goes down: the cost reduction for storage has been faster than Moore's Law since mid 1990's.
- The usage goes up.
- The risk (= number of devices x capacity of the device) goes up by the square! Device reliability has increased, but the number of devices in use has greatly increased.

What Archives want from storage²⁴ is not first storage media but they want a functionality: “everything necessary to maintain access”. We want to keep things (= *persistence*) and that we can use them immediately in current formats (= *currency*).

²² Training for Audiovisual Preservation in Europe

<http://www.tape-online.net/>

²³ <http://wiki.prestospace.org/>

²⁴ Richard WRIGHT – “What Archives Want – the requirements for digital technology”

http://tech.ebu.ch/docs/techreview/trev_308-archives.pdf

Persistence

We do not necessarily want to keep everything forever (every item going into BBC Archives has a 'review date'). We certainly keep what we've already selected and what we're about to select. Price, risk, errors, loss have to be balanced. If the transfer of 2" and 1" tapes was "excellent", for U-matic it was about 97%, over approximately 20 years. A suggestion is that 99% could be much more cost effective than striving for 100%.

Currency

Because of the high rate of change of technologies (formats, encodings, carriers, file management systems, operating systems, networks), most of these elements have a life expectancy of less than 10 years (before something changes). The implications for digital archives (especially for audiovisual archives where you must have an encoding usable by current productions): we are on something like a **5-year** (possibly 7-/-9) **verification-migration cycle**. This is based on the obsolescence of all the previous listed elements, plus the fact that data tapes formats (with a lot of material will be saved on) have a 3-year cycle (LTO). So the verification-migration cycle is between 3-9 years.

Cost of Ownership

The TCO breakdown includes:

- the maintenance cost: for keeping more or less 'the same' technology running properly; it applies to all forms of storage (shelves, robots, servers) and all media (tape, disc, optical, magneto-optical);
- the migration cost for coping with obsolescence.

Thereafter is an estimate of the cost of keeping digital archives either in managed high-end servers farms, compared to shelves, and compared to un-managed cheap raw discs storage. In this last case, the price becomes cheaper than shelves. But the problem is that it is associated with high risks. Managing raw media like LTO tapes and cheap hard disc drives without losing material is now the issue. We cannot afford high-end servers for everything. Is adding management to these raw discs a cost-effective way?

Year	Managed Servers Cost per gigabyte (media+management)	Managed Shelves Cost per gigabyte	Un-managed (raw) Discs Cost per gigabyte	Managed raw disc/tape
2002	\$15=8+7	\$0.10	\$4	Cost ?
2006	\$9 =2+7	\$0.11	\$1	
2010	\$7.5=.5+7	\$0.12	\$0.25	
2020	\$7 =0+7	\$0.15	\$0.02	
		Managed shelves are really cheap BUT: migration cost >> shelf cost		Key issue: cost of managing offline storage

For the **management of storage**, there are many approaches:

- Systems/storage managers software/hardware (cf SUN Honeycomb).
- Hierarchical Storage Management (HSM), Life Cycle Management...
- Content/Asset management (DAM, Digital Asset Management, MAM).
- Digital Libraries (OAIS and related processes, standards).
- Digital Preservation (UK: Avatar, encoding for storage; EC: PrestoPRIME for audiovisual digital preservation).

The **management by migration** 'solves' all obsolescence issues. If you transfer every 5 years, then you are coping with all the problems of currency – you can have current file systems, current encoding formats. Datatape copying every 5 years is much cheaper than an analogue migration every 20 years. BBC transferred in 6 months, with one person, at a cost of £30k 40khours of audio files from DVD to HDD and data tapes - this is 1% of the £3 million original cost of digitisation which took 4 years of work.

The **risk of loss of data** is proportional to the number of devices, to the size of the devices (because each holds more data), to the complexity of the storage management (the more servers farms, the more

servers managers, the more fingers in the pie!) - unless somehow complexity can be used to reduce risk - and to the reliability of individual devices. Besides the loss of storage devices there are many more risks: format obsolescence, IT infrastructure obsolescence, file corruption, system corruption, human errors and other human actions. They all increase in significance (impact) in proportion to the amount of storage in use.

First conclusion: as storage gets *really* cheap... it gets *really* risky.

The control of loss includes:

- The **prevention of loss**. This is where most of the attention (and research) is directed: reducing MTBF for devices, making copies (!), using storage management layer(s), introducing virtual storage layer(s), using Digital Library technology (OAIS 'packages' and preservation metadata).
- The **mitigation of loss**. When it's gone, it's gone? [30]. But this doesn't have to be the case. For example, if you have a BMP raster scan image file, with 160 errors in 40k [31-right], the bit errors are completely local and only affect the byte they occur on. If you have a GIF compressed file with 3 errors in 10k [30-center], the bit errors affect the equations which re-create the image from an encoded form and that propagate the errors. That is another reason to have uncompressed files in the archives. Fortunately, there are files that can be read despite errors - anything with a sequence (lines, pages, images with raster scan (or any sequential ordering of the data), audio, video. The structure 'unit of loss' needs then to be identifiable: particular pixels, samples, lines (text or video!), pages/frames. A structure of independent units is also contributing to the mitigation of loss: files with independent units (pages, lines, bytes) - so that the loss of one element does not affect any others. Unfortunately, most files have lost this property because compression removed redundancy - using the similarities between units ties them together - and whole conglomerations of data are affected by a single byte error.

In summary, a **survival strategy**:

- Understand costs and risks. But is very difficult to get from the storage industry substantial information (besides the MTBF of their hardware devices); e.g. how much it will cost to have a 1% error rate versus a 0.1 error rate.
- Keep (uncompressed) master material off-line (for now).
- Use only expensive managed on-line servers where usage (production/public access) justifies the cost.
- The cost-benefits equation of robots needs re-analysis against very cheap hard drives, for low-volume access.
- Uncompressed files have lowest risk.
- Migrate every five years.

... If delegating storage: maintain control!

2.10 Living in a Digital World - PrestoSpace & PrestoPRIME

Daniel Teruggi, INA Recherche, France

1) Getting your old analogue assets into the Digital World

The European **PrestoSpace**²⁵ project (2004 - 2008) involved 35 Partners collaborating in the project, 180 Archives users in 52 countries, 144 service providers in 26 countries.

The project [4] was about from making new machines for audio, video, and film handling, acquisition and digitisation [5], restoration [6], storage and archive management, metadata extraction [7] to a Turnkey System for hosting Digital Audiovisual Archives. The main objective was to make preservation faster, better, and cheaper! A main concept was the *Preservation Factory*²⁶, to take the industrial model versus the mainly applied artisanal approach.

²⁵ <http://prestospace.org/>

²⁶ This concept had not been protected and was 'taken over' and trademarked by Sony! Since the new name of 'PrestoSpace Factory'

During the project we changed our view. Instead of aiming towards a factory, we realised first that the audiovisual archives were in such a bad condition that you have to spend a lot of time handling one object, bringing down any industrial perspective. And, secondly, what was mainly needed was 'guidance' provided by a reference instance, that would help to get the knowledge, the orientations, the expertise, the tools, the methodology for preservation²⁷. And also business guidance, how to calculate the costs, where to find the money, how to "sell" the project to your top management... This met the European Commission wishes to have distributed Competence Centres in different domains of activity (→ PrestoPRIME).

2) Staying in the Digital World

Migration

We, Archives, are highly accustomed to '**conservation**': keeping objects (contents) from the past to make them available in the future. This is not a good position when you have media objects. At the BBC, there is a policy to review the status of contents every year. But INA (the French National Institute for Audiovisual archives) has no time frame – the law is "keep it forever!" On the other side these contents are highly demanded, they are used by producers, broadcasters and now made available to a wide public. So we live in a divided responsibility.

We have used the word '**preservation**' for 'communication with the future'. We have something today, and we have to carry it on and to convey it to somebody in a future time which we cannot measure.

In the communication field, you have a sender and a receiver, and between both you have a common vector (language, writing, technology...) [12]. You have to be sure that between the present and the future this link stays alive [13]. There are three ways of sending digital contents to the future, three **migration strategies**:

1) **Change now, recover later!** This is the **migration on ingest** [15]. You change now your contents from all the different historical formats to a chosen unique format (BBC: D3, INA: Digital Betacam 14 years ago). So, we have a homogeneous collection of contents, and it should be easier to conceive a unique transfer in a certain number of years and it should cost less, since the initial cost was so high. This postpones the migration but does not solve the problem. This approach is related to the concepts of: **UPF** (Universal Preservation Format), uncompressed, that would guarantee accessibility a very long amount of time – and **UVC** (Universal Virtual Computer) making the media data accessible anytime (in the future), anywhere. This is an efficient solution for repositories.

2) **Change continuously!** This is the **batch migration** [16]. You change when necessary, mainly when you have media or format obsolescence. When Archives work for the production world, their contents have to be in the format that is immediately accessible by any producer. So we have to follow the production trends, but taking into account the reality of the market (e.g. MPEG-4 not as spread in production as predicted). The associated concepts are: refreshing (transfer the same data on a new carrier), integrity check (is the data the same that it is supposed to be?), transcoding or conversion (changing formats). This is an efficient solution for continuously used contents.

3) **Don't change, it will be solved later!** This is the **migration on access** [17]. You postpone the migration until you need it. If I have my data on a floppy disc recorded in 1982, and I want to access the media data today, it does not make sense. But it has if you do not need to access regularly to the data. The only condition is that you have a very high-level original quality, plus a detailed information of what you have and how to access it, and of the structure of the record (OASIS deals with that), and a description of the preservation environment. This is intended for **emulation**, for replicating the functionality of an obsolete system, in order to 'replay' the original media²⁸. This generates the need to very precisely describe the original environment and to archive format converters or develop access software.

²⁷ PrestoSpace Preservation Guide

<http://wiki.prestospace.org/>

²⁸ Verdegem R. - 'Back to the future': Dioscuri, emulation in practice. FIAT/IFTA Digital Archives seminar 2008
http://www.ebu.ch/CMSImages/fr/FIAT-Archives-SeminarReport-FINAL_tcm7-59431.pdf

E.g. the National Archives of Australia have the obligation to keep the digital content in its original format.

Migration strategies have to be evaluated, planned and applied regularly. The big change is that it becomes an active preservation. In any case **replication** (creating duplicate copies) has to be continuously applied. Very important related issues are **authenticity** and **versioning**. Each domain has its own preservation constraints and strategies. The audiovisual domain, due to the huge volumes, represents a very particular and complex case for migration.

3) A new project for Digital contents: PrestoPRIME

PrestoPRIME²⁹ is an European project of 42 months, which started 1/1/2009, with following partners: INA, BBC, RAI, Joanneum Research Forschungsgesellschaft, Beeld & Geluid, ORF, ExLibris, Eurix, Doremi Technologies, Technicolor, IT Innovation, Vrije Universiteit Amsterdam, Universität Innsbruck, European Digital Library Foundation.

It is a R&D project for the long-term preservation of digital audiovisual objects, programmes and collections. It aims to increase the access by integrating the media archives with European on-line digital portals in a digital preservation framework. [21]. PrestoPRIME is the continuation of the philosophy of PrestoSpace with the common objective of fostering Audiovisual Digital Libraries. The challenge is: we have contents in the Digital world; but once you got there, how to stay there! We started in 2000 with PrestoSpace, creating the conditions for opening preservation factories, PrestoPRIME will bring solutions to manage digital contents (migration, protection, search, access).

Some of the **actions** PrestoPRIME is working on:

- Models and Metadata for Audiovisual long-term preservation
- Storage strategies and rule sets for preservation
- Processing and workflows for Audiovisual migration
- (Original and after migration) content quality appraisal and risk management
- Multivalent approaches to long-term AV media preservation
- Infrastructures for AV content storage and processing
- Metadata interoperability for access
- User-generated and contextualised metadata
- Content provenance and tracking
- Audiovisual rights modelling at European level
- Integration of Archives, Libraries and user generated content
- and...

To reach the objectives and conduct these actions, PrestoPRIME is setting up and managing a networked Competence Centre [23].

2.11 Metadata for radio archives & AES

Tormod Vaervagen, System Architect & Gunnar Dahl, System Administrator, NRK, Norway

In the beginning there was the tape. The know-how was from the librarians with the paper card indexes. It was more a library than a Radio Archive [3].

Years later, we started to use computer technology and index cards were typed into the computer as they were [4]. This was forming a digital island. We did not change any of the workflows. There was no common data model and no relation between planning-, production- and archive systems [5]-[7]. The connection here, were the people working at the Broadcaster's facility.

Then we started to change the planning [7], the tape recorders were also replaced [8] and the playout was computer-assisted [9]. But we kept the same workflow architecture, combining the worst of the old days with the worst of the new days.

²⁹ <http://wiki.prestospace.org/pmwiki.php?n=Main.PrestoPRIME>

What we are now doing at NRK is to take these 4 classes of systems [10] and to change to 'virtualization' and 'standardisation' around the architecture. This system is wrapped in common standards for data exchange [12]. The data structure is used between the systems, not necessarily inside the archives. This facilitates development and replacement of systems, so we can easily change one system, because it will act as the previous one, and the overall architecture remains unchanged [13]. This the 1st step towards a Service Oriented Architecture (SOA).

The archive is now a part of the production cycle, not the end point. It is not about storage, not about shelves, it is about preparing metadata for the retrieval of the content. Archiving, publishing, reporting, work as one system, but even if they are still separate units, a common metadata standard ensures a strong yet flexible integration.

Let us have a closer look on such a metadata standard. The **EBU Core Metadata Set** (Tech 3293 – 2008)³⁰ was finalised just before Christmas 2008. The main building blocks are based on Dublin Core, DC [15]. And this is actually a good thing, because when we are going to other libraries, archives or industries it is a well known and used metadata standard. But Dublin Core in its original definition is not very strict, so we had to define refinements on each DC element. In addition to that we defined how the core may be extended, as an XML-framework. This work has been a joint effort of the EBU Technical Department and of NRK, with contributions from several other EBU broadcasters.

The timeline behind this work [18]: around 2000, audiovisual scandinavian groups based and inspired by Dublin Core, set up a metadata standard (SAM) that became the start point for EBU Tech 3293 (2001). Inspired by this, in NRK we started to make an XML version of this 'AXML' that was implemented in the middleware scheme ('gluon'³¹) of the content production system, that we had already started to use, and which has been running in NRK for the last 7 years. At the same time, EBU following the Digital Strategy Group requirements also started to make an XML implementation of Tech 3293. The 2 partners came together, and the EBU Core 2008 standard will be used in our Archives for the new 'gluon'

This is an 'object oriented approach'. The standard defines different attributes. These attributes are a standard set attach to each object. If the object we are describing is a title, the title field will form the 'programme title', if it is an item it will be of course the 'item title' [19]. The type element denotes the kind of object.

An **example with XML** [20]. On top the title 'The Vikings are coming', the name given to the object, with the alternative title, e.g. the series title 'Norwegian Diplomacy' – it could also be a working title, an original title... We can re-use data – we have here more than one alternative title.

XML is an eXtensible Mark-up Language, with for example the title element mark-ups [21] showing the title field with data elements, and the content in the middle 'The Vikings are coming'. This is an ordinary text document that can be written and read in several applications.

These forms are a part of a XML framework, which is extensible. For example with the 'Title Type' [24] we have: the plain DC element where we store the content, then we have 3 attribute sets – one giving the title history (date), the next one giving the title status (e.g. working title) and another one giving the title type (e.g. series title, main title). The 2 last attributes are forming the **extension framework**.

The core is extended by reference to separate 'contracts'. For example, the 3 attributes of the 'titleType' points to a contract [25]. This contract is actually a data dictionary defining the terms you are using. A contract can be either between 2 partners, or more, or can be a standardised data dictionary you have agreed upon. So, you have the 'typeLabel' pointing to a certain defined term, the 'typeDefinition' is of course which data dictionary is used, and the 'typeLink' pointing to a resource which can give you the data dictionary. If you need to change your data application, you can easily just change the data dictionary and that will change how the title will be read and understood. So this is a way to simplify, extend and build standards.

³⁰ <http://tech.ebu.ch/docs/tech/tech3293-2008.pdf>

³¹ <http://gluon.nrk.no/>

If we have two partners wanting to exchange data [26], they select one common industry XML-based standard, the EBU Core for example [27], they agree upon common terms through a data dictionary [28], in order to be able to read and understand the data exchanged [29].

In NRK we have used this system for 7 years now. Because of its extensibility, we do not use it only for programme data but also for programme guides and archive reports, internal and external - for any kind of programme related metadata, music, news – for traffic situation information – for sports events and results information – for news and other content for new media. All are defined by this scheme

As we worked with this standard, AES needed an XML standard with descriptive metadata for audio programmes. **AES** discussed during the 124th and the 125th AES Conventions its **X098A** proposal with the EBU, and the 2 partners saw quite early that it could be modelled as a subset of the EBU Core Tech 3293-2008 [31] [32]. The two share the same origin; a similar structure and have a similar purpose. A formal documentation defining the subset is work in progress. You add the dictionaries you need, in this case the 'roles' (role list or the different Publisher types) and also 'format' (of the duration field) definitions [33].

2.12 Video Active – Providing Access to TV Heritage

Johan Oomen, R&D Department Manager, Netherlands Institute for Sound and Vision & Siem Vaessen, Noterik B.V., Netherlands

Video Active is a 36-month European project of the eContentplus programme, which started in September 2006. Its primary aim was to provide on-line access to a well balanced collection (10.000 video items by 2009) of Audiovisual heritage coming from AV archives, providing also contextual data (i.e. stills, programme guides, articles written by academics). The Web site³² is accessible in 10 languages, not only providing different language schemes but also a multilingual thesaurus. 14 members from 10 countries and 11 content providers in 10 languages are involved [3].

There are a lot of collections, very heterogeneous (e.g. TV Catalonia is just 15 years established, BBC on the other hand with a very long tradition). How do we select 10 000 items to represent the European Broadcast community? Our academics partners came up with a content selection policy based on the History of Television in Europe, and the European History on Television, exploring and showing the cultural and historical differences and similarities.

Let's have a look on the project results by watching two clips:

- Video Active clip³³
- BBC 08/11:1987 'Money programme' on 'digital phones' at the Telecom 87 exhibition in Geneva³⁴

As you can see this is presented in a Flash environment. At the very beginning we had a lot of discussions with the content providers that we needed to have one single format for playout. At the birth of the Web TV we had RealMedia (not really an option now), and Windows Media, which is an option, and is included in this project by some of our partners. But the majority of content is streamed using Flash with H.263 codec (old Flash) and now the H.264 codec. From 2009 onwards we need to offer some highest quality footage.

In the portal there are 5 different 'European television History' access pathways: Technology – Institutions – Events – Watching.

The European History is accessed through 34 topics. For example 'Terrorism' brings 40 items divided in different sets of Genres (e.g; news/documentary) / Languages / Owner / Colour/ Type (A/V) / Transmission period, allowing the user to filter this offer.

The Video Active architecture [6] comprises various modules, all using Web technologies. It has not a single streaming playout platform. This is due to IPR restrictions from many of the broadcasters AV

³² www.videoactive.eu

³³ On top of the portal: 'Video Active' → then click on the 'videoactive.eu' key frame picture

³⁴ Use the 'Advanced Search' to access it

archives, which are not allowed to stream from our facility based in Amsterdam. So, we had to find a workflow that did enable a consistent streaming playout for all partners from our location in Amsterdam or from Belgium, Italy, Greece...

The annotation process starts with a legacy database [7]. Each Archive has its own method of annotating and its own metadata workflow in place. So, we had to align all these processes together into a single Video Active scheme, in order to introduce proper searching and ranking. In the 'Web Annotation Tool' it is possible to use the Web interface of the Video Active backend [8], or (at the very beginning at least) to upload Excel sheets with archive material. The data available in the Web Annotation Tool can be edited, modified, and we automatically have 'RDF Triples' which gets us a semantic metadata model. Everything is actually stored in a 'Semantic Store'.

Every partner, who is a content provider, has its own overview of items [8] and can filter on genres, on topics, create new items, import metadata (from Excel sheets or batch import), and can add contextual information.

One can access to the thesaurus, used for the multilingual purpose: If I want to search for 'war' using the English term, or the Dutch term, all items related... from Greece, Germany...will also be retrieved.

Concerning the item creation [9], there are options for information production, adding something on significance, adding more classification, adding video files. In this case there are 2 video formats available: Flash and Windows Media. If one chooses Flash, just upload the video and it will be automatically transferred by the server transcoder. If it is going to be Windows Media files, a simple link to this file on a streaming server is sufficient. There is an option for setting your selected key picture, instead of our automated extracted key frame.

Once logged-in the user has access to his/her 'User Workspace' with registration details, favourites and settings. Beyond the simple search there is an 'Advanced Search'³⁵ to look for specific partners, items in specific languages. There is a new 'Timeline' [10]-[11]³⁶, similar to the one developed by the MIT (USA)³⁷.

The European Commission published a new call in the eContentPlus programme, and we were granted a funding for a new project called **EUScreen**³⁸ which will be "exploring Europe's television heritage in changing contexts".

Since we had the technology developed for Video Active, it was time to involve more archives. In the meantime the Europeana portal³⁹ had been launched in November 2008. 'Europeana' is the 'marketing name' of what was formerly called European Digital Library. At the moment there is primarily material from national archives and libraries rather than from audiovisual archives [13]-[15]. So the new drive was to provide Europeana with an A/V impulse.

EUScreen brings together 26 partners from 19 countries [16]. Its objectives are listed hereafter:

O1: To develop technical solutions to provide harmonized and highly interoperable audiovisual collections, using for example EBU Core (§ 2.11).

O2: To provide the necessary technical solutions that the Europeana portal needs to be able to support audiovisual content.

O3: To create demand and user-led access to television content from broadcasters and archives across the whole of Europe.

O4: To develop and evaluate a number of scenarios amongst a range of users, including the research learning and leisure sectors (this implies new functionalities at the front end).

O5: To build a community (network) of content providers, standardisation bodies and users, and to build and share knowledge among these on the key issues and challenges relevant to the audiovisual heritage domain and beyond.

The EUScreen project starts in October 2009 – Please, join the initiative!

³⁵ <http://www.videoactive.eu/VideoActive/search/AdvancedSearch.do>

³⁶ <http://videoactive.wordpress.com/2009/01/20/video-active-presents-new-search-feature/>

³⁷ <http://simile.mit.edu/timeline/>

³⁸ http://ec.europa.eu/information_society/events/cf/document.cfm?doc_id=9107

³⁹ <http://www.europeana.eu/portal/>

3 The future in production

Chairperson: Roberto Cecatto, RAI, Italy

Television, Radio, Web, mobile. Yes indeed we would have enough on our research and development plates to last us for decades. But should we not also look beyond the scope of strictly broadcasting developments? 3D-TV for instance, from which the broadcasters have a lot to learn. And what about the future plans in broadcasting...

3D

3.1 SMPTE Task Force on 3D to the Home

Bill Zou, DTS, Standards & Business Development; Task Force Chairman, USA

The Task Force (TF) was formed primarily under the request of the technology vendors. The driving force of movies studios produced a lot of 3D content over the last couple of years, and the success at the box office made them wonder why they just show it in theatres and if they can push this content all the way to the home. There is also the driving force of consumer electronics. Now, more and more homes already have HD and what is next? 3D is perhaps the next killer application. If there are driving forces at both ends, something is missing in the middle. There is no way to move the content from content owner to home. And without standards you cannot launch a successful business.

Therefore, at the Task Force kick-off meeting, 19 August 2008, 200 people attended at the Entertainment Technology Center in Los Angeles.

The 3D Task Force mission is first to answer the questions "What standards are needed?" for rapid adoption of stereoscopic content, from mastering to consumption in the home on a fixed home display via multiple types of distribution channels (broadcast, package media, Internet) – "What standards should be written by SMPTE?" in liaison with other bodies to ensure other needed standards are written.

In the 3D End-to-End Value Chain [3], the yellow box relates to what the Task Force is focusing on: the '3D Home Master' format requirements with consideration from content creation, distribution and display.

The corresponding **specific tasks** [5] are divided between 4 drafting teams in charge of:

- Defining the **issues and challenges** related to 3D distribution for the home market, by:
 - describing end-to-end distribution chain, to precise the demarcation of the Task Force scope;
 - creating use cases (with inputs from cable, DTH, Studios, broadcasters) and prioritize;
 - creating standard terms/definitions (3D terminology to ensure that discussions and documents within the SMPTE 3D Task Force remain coherent);
 - identifying unique challenges in determining solutions.
- Defining **minimum requirements** needed to overcome the issues and challenges, by
 - defining functional requirements for each distribution channel;
 - defining performance requirements;
 - consolidating and prioritizing.
- Defining **evaluation criteria** for content creation, content formatting, distribution channels, display - including both 3D quality and 2D compatibility/quality.
- Defining and recommending a **minimum set of standards** that would need to be written to provide sufficient interoperability.

We are close to completing a task-force report. This is a document to include use cases, end-to-end system diagram, terminology and minimum requirements for a single 3D Home Master that can be used for various downstream distribution platforms. The 3D Home Master will be an uncompressed and unencrypted image format or file package derived from a 3D Source Master and intended to be used in the creation of 3D distribution data.

The report of the Task Force should be complete by Q1 2009.

3.2 3D TV – Market Overview

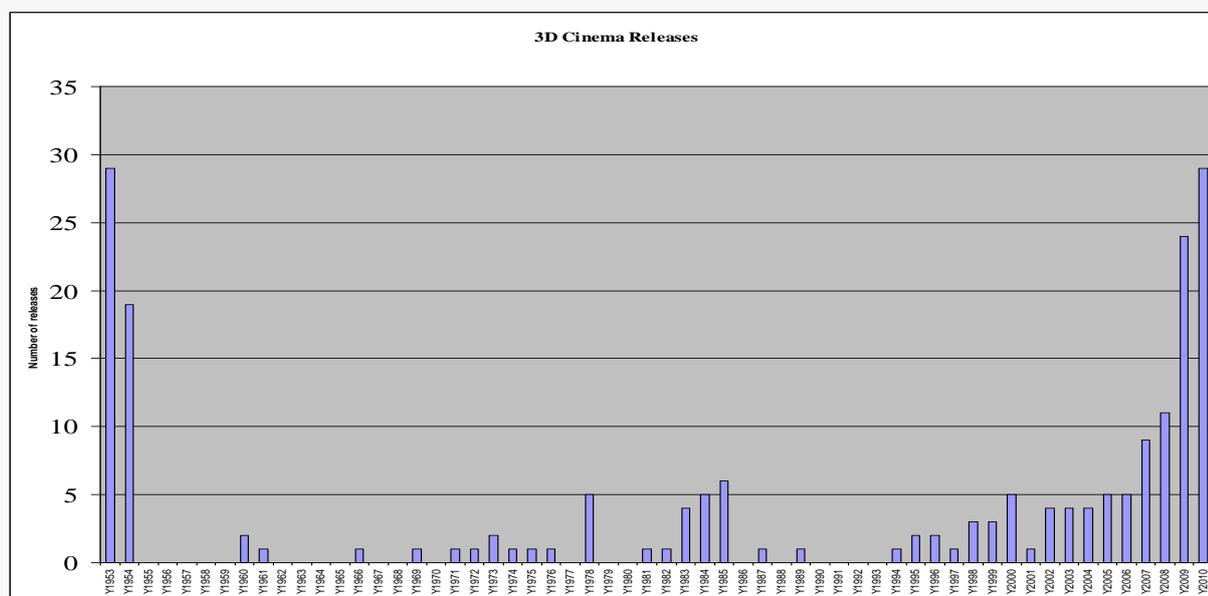
Ami Dror, Xpand TV, Ethan Schur, Tdvision, Colin Smith, ITV

3.2.1 Context of 3D. Cinema and 3DTV

Stereoscopic 3D is finally here to stay after decades of sporadic peaks of interest. Anyone that has seen the latest full colour digital 3D content (in cinema and/or on one of the new generation of 3DTV's) can testify that the experience is considerably better than previous generations. There are around 20 3D movies coming to cinemas in 2009. There are almost 70 movies right now under production. This is a 5 billion Euro investment. One of the reasons that 3D is here to stay is the heightened experience of feeling connected and immersed in the content. Many of the world's leading film producers have termed this 'the greatest evolution in large screen and television entertainment since the advent of colour'⁴⁰.

The growth forecast for 3DTV is significant [4] and highlights 3DTV is not a niche product. Manufacturers predict mass production levels of stereoscopic enabled displays by 2010. Most major Consumer Electronics (CE) manufacturers have demonstrated commercial and prototype models. In certain markets these have already been launched (Japan: Hyundai). By 2010 a continued adaptation of products will occur and there is a good chance (depending on early uptake levels) most television displays will be "3D Ready" or "Full 3D Ready" just as we have today with HDTV and HD Ready.

3D is simply, yet powerfully, adding another dimension, another way of feeling 'inside' the movie, at the match, at the concert or at the event. It is far removed from the previous 3D film generations where the director/producer was trying to have 3D content coming out of the screen to 'poke you' in your eyes/face. The cinema industry is settling into a new form of 3D language. We have seen a gradual reduction in the use of negative parallax (content coming out of the screen) which from a story telling perspective has limited appeal especially after someone has seen the effect a few times. This is possibly connected to the perception that 3D was previously seen as a "gimmick" effect.



This table shows the history of 3D film releases - confirmed or currently in production. The growth has built up in the past few years and is quite different from any previous 3D cinema release window. Notably

⁴⁰ <http://www.today3d.com/>
<http://www.linkedin.com/groupInvitation?gid=3671&sharedKey=0135C6665B53>

from a broadcasting perspective, the first 3D boom in 1953/54 (requiring colour film for the anaglyph filtering) has been commented on as a response to early colour broadcasting in the USA. This latest 3D dynamic is based on the viewer's experience and the proven uplift in 3D vs. 2D release.

3.2.2 Comparison between leap from SD to HD and from HD to 3DHD

Perhaps the most significant consideration for 3DTV is to approach/understand it from a non technical perspective. Many technologies have been introduced to the public yet the nature and balance between the consumer experiences vs. the business model has limited appeal. Many broadcasters are still trying to master HD and may consider 3DTV as something medium/long term. It could be argued this misses out the fundamental motivation of 3DTV production/broadcast. To a user, the leap from SD to HD is minor compared with the HD to 3DTV. An advertiser conveying a message in HD is a minor improvement from SD (message recall, etc) compared with the leap from B&W to colour. Work is underway to align the advertising community with the 3DTV proposition so it may result in a common view that a premium is justified. If not, 3DTV might be purely for pay television operators but beyond the full reach of FTA broadcasters. The table below compares the transition from SD-HD & from HD-3HDTV.

	SD to HD	HD to 3DHD
A	Improvement in picture quality noticed by engineers – often <i>not</i> by the consumer.	Dramatically and immediately noticeable by consumers ("wow" factor). It's completely different to the first time they would have seen HD.
B	Business model to produce or broadcast HD is challenging and perhaps still presents problems. It is difficult to apply a premium to that content.	This depends on the standards process. If you can control access to the 3D element, so you can apply more flexible business models (e.g. you get programmes and/or adverts in 2D unless consumer/advertisers have paid a premium).
C	New cameras	Mostly same HD cameras (just using special rigs). Can use specialist cameras for certain shots.
D	New infrastructure (HD-SDI)	Same infrastructure as in HD (HD-SDI) with some dual path or mezzanine compression for single path - depending on infrastructure.
E	New displays. You either had a SD television or an HD television.	This depends on the standards process. New displays (but stepping stone via HD anaglyph, consuming 3D content on an HD display). Perhaps a start with CGI generated programming once a week etc.
F	New editing software	New editing software or plug in's
G	New sets/make up etc (potentially)	Same sets/make up as HD.
H	400%+ more bandwidth than SD	This depends on the standards process. One option is a full resolution per eye model. Depending on content, if live or from playout it needs between 30-50% extra bandwidth than 2DHD. With the 3D model where HD is inside a dedicated channel it's possible to have half resolution per eye and use the same bandwidth as HD. Thus, that method of 3DTV would be 100% extra.
I	No control over access (HD by default) so long as they could access the channel.	This depends on the standards process. Ability to apply business rules at point of transmission or in the STB. For example, a 3D STB one-off licence activation to view content fee for a non advertising PSB for certain channels.
J	Not backwards compatible with SD	Backwards compatible with 2DHD. Providing the viewer experience to watch content in 2D (without wearing glasses).
K	No real alternative market for 2DHD content in cinemas - if you can watch it at home	Cinemas looking for alternative 3D content (can help cover production costs) – as the production itself can open a new revenue stream by showing it in the cinemas in 3D.
L	By-product of HD can be SD but SD content is plentiful and that is not a real value to have an additional SD feed	3DHD "by-product" is 2DHD. Yet this has a strong value as HD content still commands a premium. Plus again to (2DSD) if the edit compromise was deemed acceptable.
M	HD-STB baseline too quick for 1080p50/60 (format gap) – so is this format going to happen to the consumers?	The 3DHD STB baseline (drawing board) still open – it is possible to include, or to migrate to 1080p50/60. This is an ideal junction in time.
N	Not that different to shoot HD compared to SD	Shooting good 3D is a new skill, takes time to master. This really is a new way of involving a viewer and that skill will take time. Recommend to start learning process at this stage – not when the standards have been finalised & displays in the shops. 3D will not go away from the cinema and the pressures for consumer options will increase. In a way it's a far greater leap from SD to HD.
O	HD, for many broadcasters, had no premium felt of value by advertisers. It was just a bit better colour TV. Accordingly for FTA broadcasters, HD is often not easy to monetize.	This depends on the standards process. Proven uplift with 3D cinema. Sets precedent (up to 200%). Research on message recall of stereo 3D information many help brands justify a small increase in advertising to have the option to advertise in 3D – thus providing a viable business model for advertising based FTA broadcasters. Similar to B&W to colour.

3.2.3 Display options for 3DTV

The following is a basic summary of the 3 types of 3DTV seen recently at various trade shows and press demonstrations. It doesn't attempt to review products that will be based on further technology research - currently underway for more advanced types of 3DTV. The active and passive examples could be thought of as first generation 3DTV.

Active 3D	Passive 3D (Circular polarization)	Auto-stereoscopic Glasses Free
Panasonic, Samsung, Mitsubishi, LG, Viewsonic (so far).	Hyundai / JVC (so far).	Philips, LG (so far)
Using synchronized shutter glasses	Each line is polarized alternatively	Lenticular lens with multiple optimal viewing positions.
The advantages		
Excellent 3D quality. Capable of full 1080p per eye.	Low Cost Glasses (\$1) – glasses are starting to look better. No need for batteries or re-charging.	NO Glasses!
Not viewing angle sensitive	Very good 3D Quality	
Does not increase the display's cost	LCD Based	
Perfect Quality in 2D & 3D		
Optimal with OLED		
Good solution for home based projected 3DTV -as only a single low cost projector is required		
When viewing 2D content without glasses 0% quality reduction in light loss etc.		
The disadvantages		
Need wireless synchronized shutter glasses	Need polarized glasses	Expensive display (at present)
Expensive Glasses (\$30-\$200) depending on volume and requirements.	Increase the cost of the TV (depends on volume/profit uplift (upto 50% increase)).	Low quality 3D experience (but improving!) – still a very long way to go the home. Esp. multi-view.
DLP / PDP Based (LCD soon)	Minor decrease the quality of 2D Viewing (small light loss).	For multiview 3D potentially a 4K panel is required - every viewing angle requires additional information/pixels.
Glasses need batteries (last upto 300 hours of viewing) or recharging.	Capable only of half resolution per eye (at present).	Requires viewers head to be in correct location. Due to nature of lenticular lens. This may change in time.
	Vertical viewing angle sensitivity (getting better though).	Suboptimal viewing of 2D content due to lenticular lens.
	Degree of ghosting present depending on screen size and viewing position.	Alternative auto-stereo that uses a barrier method can be 2D switchable & half resolution per eye but requires exact head position (not practical for home consumption).

One 3DTV challenge is to provide a 3DTV format to work for all display types:

- Stereoscopic systems (active and polarised)
- Single view auto-stereoscopic (2D+depth, etc)
- Be open/friendly to multi-view auto-stereoscopic system developing forwards - potentially
- Different screen sizes / viewing distances (if your seat is on the 1st row of the cinema, the amount of 3D effect that you will perceive will be very hard to process for sustained viewing).
- Legacy HD TV support via anaglyph (potentially)

3.2.4 Distribution format to the home

The consumer who purchases a television, whether it is '3D ready' or not, should be able to watch the programmes clearly in 2D or in 3D at the highest resolution possible per-display. Users deserve to be given the choice. Certain content may be deemed acceptable in anaglyph – thus providing an entry experience. Perhaps CGI production formats.

Content can be generated by stereoscopic HD cameras, computer generated virtual stereo rigs, or 2D to 3D conversion done offline. Certain technical methods permit a sampling of the left and right frames so that when the 50% is removed from each view and the content placed in a format known as 'side by side' (SBS) [11]. When viewed on a 46" display, the result can still appear impressive. Issues manifest when SBS content is played on existing 2D televisions; users will not be able to view the content in 2D. SBS cuts the resolution per-eye in half, and when transcoding this frame for specific displays such as row interleaved LCD, there is an additional loss of another 50% resulting in a stereoscopic image that has only 25% of its original pixels [12] and the rest interpolated. Interpolation techniques that may work moderately well in 2D have more acute consequences when applied to 3D motion picture images. This is because interpolating pixels that are geometrically neighbours but dimensionally not neighbours leads to incorrect depth cues.

In a similar way with 720P vs. 1080i the issue often goes beyond technology. 720P was perhaps more technically suited to all non cinema content and yet 1080i was easier to sell to the public. Panasonic has already opened the debate with its "Twin Full HD 3D" message. Taking aside 3DTV for a moment the most significant issue is the new channel model vs. evolution of 2DHD channel issue. If you have 3D content as an extension of HD it permits a gradual increase according to the market/budget/skill set and format suitability. In a similar way that colour broadcasting was mostly consumed in black and white (not everyone had a colour television when broadcasting started and nor did the quantity of colour content) 3D would fit this evolved format better as a broadcasting proposition than as a new channel model. This does not stop the eventual migration to full 3DTV channels but the content gap makes this look challenging to say the least over the medium term (3 to 5 years).

If any format of 3DTV is broadcast that puts both left & right eye views in a single HD frame (1920x1080) it will need line processing to generate a full 2DHD frame. This line processing is not resident in any native sets so existing users would not be able to watch the content on their 2D sets. Whilst line processing may appear a minimal issue it would still, no matter how good it was, be a compromise in image quality. Attempting to market this to current HD consumers would present issues.

3.2.5 2D Backwards compatibility - key to permit user freedom and gradual introduction

From a FTA broadcasting perspective, for 3DTV to take off, 2D backward compatibility is essential [13]. The simplest way is to use one of the views (if you close one of your eyes, you see in 2D!). Many consumers will not want to be forced to wear glasses to consume content – even if they had a new 3DTV. Home consumption has many use cases. You might be watching content, eating dinner or cooking whilst chatting to friends, etc. This presents challenges to all environments where 3DTV might not be viewed in 3D. For example, a home that has gone for active glasses based 3DTV might only have 4 glasses and in certain times of the year would have more than 4 people viewing the content. So until auto-stereo 3D finds a quality experience similar to the first generation of glasses based 3DTV we will never reach the majority of our potential audience base. This is why a gradual migration is needed from the 2D to 3D environment. Supporting 3D, at this stage, means the gradual building up of skills and content to provide the justification bases to purchase an auto-stereo display. It can also re-address any issue over quality reduction in 2D consumption due to the lenticular lenses until sufficient 3D content would be available.

TDVision Systems Inc. has proposed a standard solution based on '**2D+Delta**'. This is an advanced matching correlation of all the pixels and colour information of the left view and of the right view. You discard what is the same and you end up with the 'Delta' - the difference information [15]. You run a DCT on this secondary or stereoscopic information, make a modified stereoscopic B-frame (inter-view frame) and place it in the transport stream. The legacy HD STBs discard the 3D data and simply playback in 2D. You can also use the delta to reconstruct the full resolution Left, the full resolution Right and prepare the picture for any type of display whether it is 2D, anaglyph [14], DLP, LCD or dual/single projector. The latter extending from home cinema all the way to live 3D broadcasting in cinemas/custom screenings.

This abstraction of the broadcast signal from the end consumer device is the only way to permit various CE vendors to continue to support their preferred technology type with the least bill of material cost in the

display to cope with various technologies. This can provide early manufacturing confidence on what can be sold as “3D Ready” and start to build up capacity for mass production when displays are “Full 3D Ready”. That way the consumer is clearly aware of what the display is capable of at the time of purchase. The alternative is to take a gamble that a certain type of display format will be the only one on the market and deliver in a format that the end display supports. This would limit evolution as full resolution per eye displays will introduced by at least one manufacturer. Knowing the standard permits multiple quality points to the consumer helps facilitate other business models such as live broadcast to cinema or brand funded screenings. This larger audience projector-based screening benefits from full resolution per eye. Passive and active glasses based 3DTV both have their market and consumers will base purchasing decisions on many factors.

3.2.6 What to consider now – independently of 3D broadcasting/production decisions

Many issues still need more consideration. For example, in a cinema you have a deliberate decision to watch a film in 3D. Pay per view is similar. DVD or Blu-ray is similar. For broadcasting it eventually would involve an increase in content types. This will include adverts (commercially significant) and other content types such as promotions and interstitials. Consideration is required for seamless extended 3DTV consumption. This will take time to understand fully and can be implemented using best practice methods over time. This suits gradual 3DTV introduction - the learning process is likely to develop/improve when the content is produced and consumed.

A discussion is needed now on the likely method of 3DTV introduction. This affects what option might be best to consider from a standards perspective. Should 3DTV be gradual evolution from 2DHD or a new channel? Who should pay the cost premium in 3D content production (brand, advertiser, public funded PSB, etc)? To what level is some degree of access control to 3D viewing required? Should parents be able to control the number of hours a child might view 3D content? These are just a few points to consider in standardisation. Either directly to the display without any real standards consideration, or as part of a new base line for a “Full 3D Broadcast Ready” STB/TV.

A broadcaster may (or may not) have plans to instigate 3DTV broadcasting yet thought is needed now on 3DTV standardisation as input/contribution. If no view is expressed developments may move forward closing any door on gradual introduction of stereo 3D to the consumer – with an increase in the shift to pay channel consumption (i.e. as completely new channels might be required). 3DTV as a full channel proposition, from day one, rather than as a gradual introduction - may be far worse than ignoring 3DTV and letting the standards process develop forwards without EBU member input.

3.2.6 Conclusions

3DTV is a controversial subject. Primarily because the consumer may not want to view 3D in the home requiring glasses. Of course consumption of 3DTV without glasses is preferable in the medium term but it will take a while for auto stereo 3DTV to reach the same quality point as ‘glasses based’ solutions. Various tests have been performed that prove the “glasses on” 3DTV showing the experience leap is beyond the negative perception of wearing glasses. In other words, after they have seen 3D even with glasses on, their opinion may change. JVC’s 3D glasses have far more style than the types we have seen in the cinema. This will help change perception as it’s often not the wearing glasses that is an issue – but the wearing of “funny 3D glasses”. The press are doing a good job of showing the worst possible glasses which would never be used as typical home 3D eye wear.

What is dramatically different for 3D, compared to the introduction of HD, is its support and interest from the cinema industry. Avatar is James Cameron’s first feature release after Titanic. This is in 3D and it will raise the creative perception of 3D forever. Quantel’s leadership in post production raised awareness of new business models with the release of “Hannah Montana” to great commercial and industry acclaim. They currently lead high end 3D post production opening many new applications of 3D such as ‘catch up’ screenings at music festivals etc. The majority of people in the 3D industry have a view that 3D will reach the home in a short to medium term time frame. Certain consumers will wait for ‘without glasses’ 3DTV and, in time that will occur. This means a significant viewer base would potentially still value 3DTV in its initial introduction. Perhaps more so than HD as the experience leap from SD to HD was/is far lower.

Every consumer/family/home will have its own personal preferences or display type suitability. The television used to be a platform just to receive broadcast television. Now it used for a variety of other purposes including gaming, DVD’s/Blu-ray, etc. The gaming industry has started its stereoscopic 3D offering and it is likely to continue the need to place importance on supporting a heterogeneous display type and an abstraction of the 3D broadcast in the STB level (initially).

An analogy may be given. Imagine if HD could have first been broadcast from day one in 1080P50 and

the consumer could first select if they wanted to view in 720P or in 1080i. Naturally if that happened 1080P50 would have reached the consumer already as the transport stream would have supported it. This matter of considering the short, medium and long term evolution to 3DTV can be given thought today *before* defacto or knee jerk standards enter the market place and shift consumption away from FTA channels. Effectively this kind of option is possible with 3D providing a good level of backwards and forwards compatibility and the model to evolve (to the pace felt comfortable) with 3D broadcasting.

HDTV required both new displays *and* set top boxes. This was issue because of the considerable change from standard definition components and the attendant costs. The model put forward by TDVision will still require either a box firmware change or a new STB. However, many of the existing silicon available today is compatible with this 3D broadcast model. It's not such a great leap up technically in comparison to SD to HDTV. In addition, many displays that support 120Hz can work with active glasses and give a "Full 3D Ready" display at the same cost as a 2DTV. The only extra cost is active glasses. The passive polarised displays do permit a greater flexibility of glass design and are suited to when you might need to support many viewers at the same time. To apply the polarization plane to the LCD does require additional cost but at mass volumes this is not considerable. To compromise 3DTV by limiting consideration to options that only go inside a video frame limits quality, evolution potential, user experience, the option of 2D/3D single EPG channel and denies the highest quality 3DTV to the consumer.

Finally a point to consider - if we are to have a new generation of "Full 3D Ready" set top boxes should this be used as a way of future proofing to cope with 1080P50? If not at this rare junction in time how can 1080P broadcasting ever occur?

Future technologies

3.3 High Frame Rate (HFR) Television

Richard Salmon, Sam Davies, Mike Armstrong, Steve Jolly, BBC R&D, UK

Over 70 years ago, the TV frame/field rates were chosen, to: exceed the threshold for apparent motion, avoid visible flicker (on small screens on these days), avoid interaction with the mains frequency and provide a way of showing cinema film on TV systems. The current 50/60Hz TV was a match to standard definition pictures and smaller CRT displays. But it is not a good match for larger displays, increased picture resolution and sample-and-hold display technology (such as LCD).

Because the camera has a **shutter** that is open for 1/50th of the second, you get a loss of detail on moving objects. So, the static objects in the background are nice and sharp, the moving object is blurred out by the camera integration [6]. If you have a long shutter, then you have motion blur. If you have a short shutter you loose of the smoothness of the motion and introduce temporal aliasing (leading to jerky motion and spoked wheels running backwards). The short shutter will sharpen the picture, but you reduce the amount of light coming in, hence causing a loss of sensitivity.

Consider a ball moving across the screen [8]. We want the ball to remain looking like a ball as it moves. If you capture it with a short shutter, you get nice sharp images (but juddery motion) – with a 50% shutter, the images are still somewhat blurred out – but with a 50% shutter at double rate you get much sharper images and much smoother motion as well, without using excessive camera shuttering.

Another way of looking at the problem is, say for example you are following the action in a football match and you have upgraded your cameras and broadcast system **from SD to HD**, the action still happens at the same speed. If you follow this action by panning at the same speed as in SD, since you have got 3 times more horizontal pixels it is 3 times more blurred as it was in SD [11]. You lost all the advantages of HD. So if you increase the resolution still further, you have to slow down the rate of panning. The problem is that the dynamic resolution of HDTV is actually no better than the dynamic resolution of SDTV [11].

Historically, about 20 years ago, the BBC proposed that we should have a 80fps for HDTV, in part because we managed to get a CRT to work at 80 H

What is the impact on the viewer? If there is a large difference between the static resolution and the dynamic resolution of (moving objects) in a picture, this can lead to a feeling of nausea. Therefore the higher the *static* resolution, the higher the *dynamic* resolution must be for comfortable and lifelike images. The solution in the case of the football was to reduce the shuttering slightly to give a sharper image, and also to reduce the aperture correction in the camera.

Up-converting displays. 100/120Hz LCD TVs are available, and 180/200/240/480Hz models are now being exhibited. But there are still fed with 50/60 Hz signals, with the frame rate interpolated up in the set. That solves the problems of large areas flicker and display smearing. Motion prediction which is used to create intermediate pictures is never perfect - and to get 480 Hz they are inserting black fields, which shortens the display aperture and helps sharpen up motion. But that is entirely to mitigate the problems of sample-and-hold displays (LCDs). Of course it cannot reduce the motion blur captured in the camera and it cannot predict complex motion (for example, these displays have a problem with rotating motion).

Therefore, to make motion rendition more lifelike we need higher frame rates in the camera, for distribution and in the display. We would suggest that: if SD is acceptable at 50Hz then full HDTV needs 150Hz and as resolution increases, we probably want at least 300Hz - a multiple of 50 Hz and 60 Hz - is easy to convert to 50 or 60Hz and is compatible with mains frequencies. Or may be we can go to 600 Hz to incorporate 24 Hz as well!

The potential **HFR issues** – it cannot be an all win! Clearly, higher frame rates require:

- Increased storage and increased bandwidth. The good news is that HFR video should be easier to compress, because of: smaller changes between each picture, each frame is sharper making the motion easier to predict, less temporal aliasing, and video compression could make use of three-dimensional transforms. You also can have longer GOP length (still half a second while having 6 times as many frames within that GOP).
- Shorter exposure for each frame leading to higher noise levels. But if each image is cleaner in term of motion blur, you should be able to do better motion prediction and hence also better noise removal, and at higher display rates random noise is far less visible to the human eye (cf. DLP and plasma displays)
- Interaction with AC lighting. It will lead to variations, fluctuations in illumination between pictures. This may make compression more difficult, but will not be noticeable when displayed. Multiple of mains frequency (such as 300 Hz!) should be used to avoid beating. It is also simpler to filter out temporal lighting problems and photographic flash.
- Loss of “film-look”? With a higher frame rate shoot you can change the temporal characteristics of the video in postproduction, for example: add film-look later, add film-look to only part of the picture (and if part of the picture has a problem you can average fewer or more frames) and develop a new range of motion characteristics (with shaped temporal filters).

HFR Production

High Frame rate production also gives the possibility of creating higher quality standard rate productions. We can convert 300Hz production equally well for 50 and 60Hz TV. Better temporal down-sampling can be used to minimise aliasing and give improved motion portrayal. A greater range of motion FX can be applied.

Demonstrations of High Frame Rate TV that took place at IBC 2008 [27]+[28], with video shot 1920x1080 at 300fps and down converted to display at 1400 x 788 100fps, showed that against the smeary picture of 50 fps, the 300 fps picture is captured absolutely sharp and that the eye can track the object moving across the screen, with significant improvement even at only 100 Hz.

Further work still remains to be conducted:

To understand how well HFR video compresses.

To understand the trade-off, as the frame rate increases and the visibility of noise decreases vs. increase in noise with loss of sensitivity.

To understand what bit depth is required as frame rate increases (bit depth comes down?).

Meet a compromise, choosing the optimum frame rate for a given resolution, which is also a compromise between data capacity and the visual effect.

HFR Conclusion

Increasing the static resolution without improving the frame rate makes the TV system less and less suitable for moving pictures (NHK is now considering what it should do about frame rate for Super HiVision). We assert that increasing the frame rate for capture and display of television pictures produces a very significant improvement in video quality, and increases production flexibility, especially for sports material (e.g. covering tennis from the side as well as from the end of the court becomes possible).

3.4 Future Television Production – Proof of Concept

Maarten Verwaest, VRT R&D Department, Belgium

PISA - Production, Indexing and Search of Audiovisual Material is a 30 man-year research project of the VRT-Medialab⁴¹ with IBBT (Interdisciplinary Research Institute)⁴², on video search technology, unsupervised feature extraction and computer-assisted production.

Context of the project. In previous years we did some extensive research on file-based production in terms of networked storage and of a single file-based 'production machine' [3]. On top of that we have investigated and installed a lot of software applications that have virtualised and integrated our production process, including Ingest, Editing, Playout [4]... However, if we look at the particular context of drama production, we actually see that the metadata flow, which sits on the top of the system and that associates the management information for all the media assets, is usually combined by documents [5]. News are a little more advanced – and journalists will use an editorial application – but it is still unstructured text and it is difficult to manage and to automate the information flow that actually controls the production 'machine'. However we think it is crucial, that when all the material we produce is going out in digital form (via DVB-S, cable or telecoms lines) to capitalise on that metadata to be able to bring it out in a structured way, to take care of our EPG, etc. We should find the means to harvest the different sources of metadata and offer that in a nice and attractive form to the consumer [7].

In order to solve that we bought a MAM system [6] and did a lot of 'plumbing' and then we could end up with an information portal on top of our News material⁴³ [8]. We took a look on what BBC did with iPlayer⁴⁴ and had this broadband experience of News. It is like TV experience on the first sight but it combines different ways of sorting out your news items, mark up 'My personal channels' and includes search functionalities. We have in Flanders normally 1 million news consumers per day, between 100 000 and 200 000 people using the News Web site, that is 10% of our audience, and a significant part is attracted by the 'VideoZone' [8] introduced mid January 2009.

Users want something more than just redistribution of a single cast on different distribution channels. They expect "**configurable**" **content**, that means: scalable content served by multiple distribution channels, hybrid formats complementary using multiple distribution channels, value added applications (EPG, Favourites, MyChannel,...) relying on various metadata sources. On-line offering is characterised by personalisation.

Our definition of a **future television production** is based on the following **assessments**:

- Concurrent engineering (collaborative production) increases productivity.
- Modular production apparatus enables a configurable product.
- A Digital Supply Chain Manager should ensure overall consistency and performance. Individual Media Asset Management systems don't match this requirement and ad hoc plumbing ("best of breed") compromises stability and quality of the product.

⁴¹ <http://medialab.vrt.be/pisa>

⁴² <http://projects.ibbt.be/pisa>

⁴³ <http://www.deredactie.be/cm/de.redactie>

⁴⁴ <http://www.bbc.co.uk/iplayer/>

- **More, better and structured metadata** will be the driver to manage the evolution from bare application integration to information integration and supply chain optimisation.

A value-added application like the Search engine for video [16] is a bit of a problem, because as long as we can index video files by crawling the hypertext surrounding it is searchable, but as soon as we put every piece of video on-line before an editor has added some text we have a major issue because it does not exist any search engine to index video as such – So that was considered a focal area in our R&D.

A regular procedure would be to include logging activities during capturing or annotation activities in the Archives process. We think it is not very scalable and we need to accelerate this type of process [15]. We did a lot of research on **computer-assisted archiving** and particularly **features extraction**.

Shot segmentation is an easy one but a more difficult one is **scene recognition** [17], when you want to offer a time-coded index of a logical unit of work, which corresponds to the unit of the editor or of a director.

Another work is **video copy detection** [18] to identify duplicates, to group related copies or search results, for Intellectual Property Protection and for computer-assisted analysis.

The work on **face detection** [19] was not based on the pixels but on morphing 3D models.

All these features extractors would run in the background during ingest. Instead of offering the archivist an annotation client with which it should start from scratch, we have proposed an annotation client that collects the preparatory work [20] including all the scripts that have been collected from the editorial department, if available, including subtitles – that are available from another department. The most difficult final task of our research project has to integrate these different sources, those different aspects of the same item, to perform the 'semantic aligning' of all these dimensions.

As a proof of concept, we came up with an advanced Search Engine 'Trouvaille' (lucky find) [21]. It is optimised for video with a lot of metadata processing going underneath. We include it in our on-line News bulletin, and we are looking how to re-use the research results in our broadband Internet news player [22].

Last challenge: starting from CAD/CAM [23], we are wondering if we could apply 3D modelling technology for pre-visualisation for regular TV drama production, using a very interactive interface, a "3D set modeler"⁴⁵ [27].

3.5 LIVE extends the interactive television experience – The 2008 Beijing Olympic Games

Philipp Krebs, ORF, Austria

LIVE⁴⁶ is an integrated project partially funded under the European Union's IST 6th Framework Programme. It was launched on January 2006 and will run for 45 months. The coordinator is Fraunhofer IAIS. The LIVE project creates novel content formats with new production methods and intelligent tools for interactive digital broadcasters. TV consumers are able to influence the TV authoring of live content. Professional users are enabled to create a non-linear multi-stream video show in real-time, which changes due to the interests of the consumer (end user).

The LIVE system and new TV format were tested at ORF (Austrian public broadcaster) during the 2008 Beijing Summer Olympic Games. The LIVE trial went "on air" over the three weekends of the Olympic Games 9-24 August 2008 from 9am to 3pm on both Saturday and Sunday. The input video sources were: 12 multi-streams live from the Olympic Games sites, 1 studio mix (3 cameras), 1 live camera for backstage 'visits' and more than 120 video clips (archives) through 3 parallel channels. The output 'ORF1 Interaktiv' broadcast consisted of 5 interlinked channels [4]. A simple onscreen menu enabled viewers to

⁴⁵ http://www.vrtmedialab.be/index.php/nederlands/publications/publ_medialab_MV_20080913_ibc2008

⁴⁶ <http://www.ist-live.org> http://www.ist-live.org/promo-material/promo-material/factsheet_2008.pdf

easily zap across the channels and with the remote control respond to requests and messages from the ORF production team.

The **interactivity** took place in all levels of production and consumption [18].

- On the viewer side with: of course his/her remote control, the mini EPG [13]+[16], his/her responses to Voting [20] and to Switch [20], and by 'Skyping' with the production team.
- On the production side:
 - A 'Feedback Application' interface displaying the use of the channels [35] and the viewers' responses [36].
 - A 'Recommender tool' [40]-[41], a search & retrieval tool based on the information provided by the system.
 - The producer making informed decisions about which sports to broadcast live or which content of high interest to repeat. Patterns began to form in the voting and rating of content that indicated high interest in the behind the scenes action of sports events, this also included the behind the scenes of the production broadcast itself [15]. The result was the production of new content 'on-the-fly' such as studio guest interviews, documentaries about sports personalities and interviews with the production team [14] or visits to viewers [15].
 - Moderators on a dedicated channel for permanent, informal and spontaneous live moderation. It soon took on the form of a home channel where viewers would return either for a break from the action or for an update on the action across the 5 channels [22] or to hear what the other viewers had to say. The producer also relied on the studio channel for immediate reaction to patterns in viewer behaviour such as a major switch to the medal ceremony involving an Austrian athlete.

The core of the system was an '**Intelligent Media Framework**' using an 'Intelligent Content Model' [26] which integrated the knowledge about the content sources (live events, production archives), the staging concepts (When do we switch - In which way? - How do we communicate? – Do we treat the main topic now on the 4 channels? Do we spread our topics?) and the information coming from the supporting tools. For generating metadata the Fraunhofer Institute developed a 'Human annotation' interface [42].

Results of the field trial⁴⁷. LIVE was "on air" for a total of 33 hours per channel, in which a total of 254 onscreen interactive elements were produced. On average more than half of all viewers took regular advantage of the interactive elements to switch to dramatic events on another channel or to vote or rate the content on screen. 87% of viewers participated in the trial on at least two weekends. Overall satisfaction with the broadcast was very high. In fact 63% of viewers who participated on all three weekends were very happy with the broadcast [61]. Nearly all of the respondents in the follow-up user survey indicated that they would like to use such an interactive TV service in the future.

⁴⁷ http://www.ist-live.org/promo-material/promo-material/trialfolder_s.pdf

Annex 1: Abbreviations and acronyms

Note: Some terms may be specific to a speaker or/and his/her organization

1080i/25	High-definition interlaced TV format of 1920 x 1080 pixels at 25 frames per second, i.e. 50 fields (half frames) every second
1080p/25	High-definition progressively-scanned TV format of 1920 x 1080 pixels at 25 frames per second
1080p/50	High-definition progressively-scanned TV format of 1920 x 1080 pixels at 50 frames per second
2D	Two-dimensional
2k	Horizontal definition (number of pixels) in Digital Cinema
4k	
3D	Three-dimensional
3GPP	3rd Generation Partnership Project
3H	Three times the TV monitor/set height value
4/3, 4:3	Picture aspect ratio (width/height)
14/9, 14:9	
16/9, 16:9	
4:4:4	Ratio of sampling frequencies to digitize the Luminance / Chrominance (Cb, Cr) components
4:2:2, 422	
4:2:0	
5.0	Left front / Centre / Right front / Left Surround / Right Surround sound channels
5.1	5.0 + LFE channel (surround sound)
720p/50	High-definition progressively-scanned TV format of 1280 x 720 pixels at 50 frames per second
AAC	Advanced Audio Coding
AAC+	AACplus = HE-AAC
AAF	Advanced Authoring Format
AC	Alternative Current
AC3	Audio Coding 3, known as Dolby Digital
AD	'Anno Domini': After the Christ's birth
ADC, A/D	Analog-to-Digital Conversion/Converter
Ads	Adverts, commercials
AES	Audio Engineering Society
ANC	Ancillary (SDI, HD-SDI)
API	Application Programming Interface
ASD	Abstract Service Definition
ATM	Asynchronous Transfer Mode
ATSC	Advanced Television Systems Committee (USA)
A/V	Audio/Video
AV	Audiovisual
AVC	Advanced Video Coding (MPEG-4 Part 10 = ITU-T H.264)
AVC-I	Advanced Video Coding - Intra (Panasonic)
AXML	Audiovisual XML ??? (SAM / NRK)
B	Bidirectional coded picture (MPEG)
B2B, BtoB	Business-to-Business
B2C, BtoC	Business-to-Consumer
BB	Black Burst
BBC	British Broadcasting Corporation
BMP	BitMaP file format
BR	Bit-Rate
C	Centre (surround sound)
CAC	Call Admission Control (Cisco)
CAD	Computer-Aided Design
CAM	Computer-Aided Manufacturing
CAPEX	CAPital EXpenditures
CCAAA	Co-ordinating Council of Audiovisual Archives Associations http://www.ccaaa.org/
CD	Compact Disc

CEA	Consumer Electronics Association (USA)
CEO	Chief Executive Officer
CLDM	Common Logical Data Model
CMOS	Complementary Metal-Oxide Semiconductor
CMS	Content Management System
CRT	Cathode Ray Tube
CSI	Common Synchronization Interface
D-Cinema	Digital Cinema
D/DABA	DAB+ Audio quality evaluation (EBU Project Group)
D/HDC	Evaluation of HD codecs (EBU Project Group – Delivery Technology)
D-10	Sony's IMX VTR SMPTE standard
DAB	Digital Audio Broadcasting
DAM	Digital Asset Management
DB	Database
DC	Dublin Core http://dublincore.org/
DEA	Detection, Extraction & Annotation (LIVE project)
DIDL	Digital Item Declaration Language (MPEG-21 Part 2)
DLP	Digital Light Processing
dm	Downmix
DMAM	Digital Media Asset Management
DMF	Digital Media Factory (VRT)
DMS	Descriptive Metadata Scheme (MXF)
DNxHD	High Definition encoding (Avid) http://www.avid.com/resources/whitepapers/DNxHDWP3.pdf?featureID=882&marketID=
DRM	Digital Radio Mondiale
DRM	Digital Rights Management
DSL	Digital Subscriber Line
DST	Daylight Saving Time or Summer Time
DTH	Direct-To-Home
DVB	Digital Video Broadcasting
DVB-C/-H/-S/-T	Digital Video Broadcasting (Cable/ Handheld/ Satellite/ Terrestrial)
e.g., eg	exempli gratia, for example
E2E	End-to-End
EBU	European Broadcasting Union
EDL	Edit Decision List
ENC	Encoder / Encoding
EPG	Electronic Programme Guide
ESB	Enterprise Service Bus (IBM) http://www-306.ibm.com/software/info1/websphere/index.jsp?tab=landings/esb
f/s, fps	Frame/second
f-stop	Focal-number or focal-ratio (focal length of a camera lens divided by the "effective" aperture diameter) adjusted in discrete steps
fps	frame per second
FS	Full Scale
FTP	File Transfer Protocol (Internet)
FX	Special effects
GB	Gigabyte
Gbit/s, Gbps	Gigabit per second
GIF	Graphics Interchange Format
GOP	Group Of Pictures (MPEG-2/-4)
GPS	Global Positioning System
GUI	Graphical User Interface
GVG	Grass Valley Group
H	Horizontal
HBR	High Bit-Rate
HCI	Human-Computer Interface
HD(TV)	High Definition (Television)
HDD	Hard Disk Drive
HD-SDI	High Definition SDI (1,5 Gbit/s)

HE-AAC	High Efficiency - AAC http://tech.ebu.ch/docs/techreview/trev_305-moser.pdf
HFR	High Frame Rate
HP	High Profile (MPEG)
HTTP	HyperText Transfer Protocol
HSM	Hierarchical Storage Management
HW, H/W	Hardware
I	Intra coded picture (MPEG)
i	Interlaced
IBBT	Interdisciplinair instituut voor BreedBand Technologie http://www.ibbt.be/index.php?node=293&table=LEVEL0&id=1&ibbtlang=en
IBC	International Broadcasting Convention (Amsterdam)
ID	Identifier, identification
i.e.	id est, that is to say
IMF	Intelligent Media Framework (LIVE project)
INA	Institut National de l'Audiovisuel (France)
IP	Internet Protocol (OSI Network layer)
IPR	Intellectual Property Rights
IPTV	Internet Protocol Television
IRD	Integrated Receiver/Decoder
IRT	<i>Institut für Rundfunktechnik GmbH</i> (German broadcast technology research centre) http://www.irt.de/
IT	Information Technology (<i>Informatics</i>)
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union – radiocommunication sector
iTV	Interactive Television
ITV	Commercial television network (UK) http://www.itv.com/aboutITV/
JP2K, J2K	JPEG2000
JPEG	Joint Photographic Experts Group
KLV	Key-Length-Value coding (MXF)
L / Ls	Left / Left surround sound
LAN	Local Area Network
LBR	Low Bit-Rate
LCD	Liquid Crystal Display
LFE	Low Frequency Effects channel (Surround Sound)
LKFS	Loudness, K-weighting, Full Scale
Ln	Level n
LTO	Linear Tape Open (IBM, HP, Seagate)
LUT	Look-Up Table
M	Mega
MAC	Media Access Control
MAM	Media Asset Management
MB	Megabyte
Mbit/s, Mbps	Megabit per second
MCR	Master Control Room
Mgmt, Mgt	Management
MIT	Massachusetts Institute of Technology
MPEG	Moving Picture Experts Group
MTBF	Mean Time Between Failures
MUSHRA	Multi-Stimulus test with Hidden reference and Anchors
MUX, MX	Multiplexer
N/SC	Standard Converter (EBU Project Group)
MXF	Material eXchange Format
NewsML-G2	News Markup Language - 2 nd Generation (IPTC)
NGN	New Generation Network
NHK	Nippon Hoso Kyokai (Japan)
NLE	Non-Linear Editing
NMC	Network Management Committee (EBU)
NRCS	NewsRoom Computer System
NRK	Norsk rikskringkasting (Norway)

NYC	New Year Concert (ORF, Austria)
OAI	Open Archives Initiative http://www.openarchives.org
OASIS	Organization for the Advancement of Structured Information Standards http://www.oasis-open.org/home/index.php
OB	Outside Broadcasting
OLED	Organic Light-Emitting Device (Diode)
OPEX	Operational EXpenditures
ORF	Österreichischer Rundfunk
OSI	Open Systems Interconnection
p	Progressive
P/AGA	Advisory Group on Audio (EBU Project Group)
P/CHAIN	Television production CHAIN (EBU Project Group)
P/CP	Common Processes (EBU Project Group)
P/FTA	Future Television Archives (EBU Project Group)
P/FTP	Future Television Production (EBU Project Group)
P/HDTP	High Definition in Television Production (ex - EBU Project Group)
P/HDTV	High Definition Television (EBU Project Group)
P/LOUD	Loudness in broadcasting (EBU Project Group)
P/MAG	Metadata Advisory Group (EBU Project Group)
P/MDP	Middleware for Distribute Production (EBU Project Group)
P/META	EBU Metadata Exchange Scheme
P/NP	Networked Production (EBU Project Group)
P/TVFILE	Use of FILE formats for TeleVision production (EBU Project Group)
PCR	Programme Clock Reference (MPEG-TS)
PDP	Plasma Display Panel
Ph	Physical layer (OSI)
PH	Picture Height
PiP	Picture in Picture
PISA	Production, Indexing and Search of Audiovisual material (VRT & IBBT)
PMC	Production Management Committee (EBU Technical Department)
PPM	Peak Programme Meter
PS	Parametric Stereo (HE-AAC)
PSNR	Peak Signal-to-Noise Ratio
PTP	Precision Time Protocol (OSDI Application layer)
QC	Quality Control
QPPM	Quasi-Peak Programme Meter
R / Rs	Right / Right surround
R&D	Research & Development
R2LB	2 nd Revised Low Frequency B-curve (ITU)
RAI	Radiotelevisione Italiana
RDF	Resource Description Format (W3C)
Res.	Resolution
RF	Radio Frequency
RFT	Request For Technology (SMPTE)
S&R	Search & Retrieve
SAC	Spatial Audio Coding (MPEG Surround)
SAM	Scandinavian Audiovisual Metadata group
SAN	Storage Area Network
SBR	Spectral Bandwidth Replication (HE-AAC)
SD(TV)	Standard Definition (Television)
SDI	Serial Digital Interface (270 Mbit/s)
SDK	Software Development Kit
SIA	'Stuck in active' (Cisco)
SIS	Sports Information System (LIVE project)
SMIL	Synchronized Multimedia Integration Language
SMPTE	Society of Motion Picture and Television Engineers
SNMP	Simple Network Management Protocol
SNR	Signal-to-Noise Ratio
SOA	Service Oriented Architecture

SOAP	Simple Object Access Protocol http://www.w3.org/TR/soap/
STB	Set-top box (-> IRD)
SVT	Sveriges Television och Radio Grupp (Sweden)
SW, S/W	Software
T-DMB	Terrestrial Digital Multimedia Broadcasting
TB	Terabyte
TC	Time Code
TCO	Total Cost of Ownership
TFT	Thin-Film Transistor
TRL	Time-Related Label
TS	Transport Stream (MPEG-2)
TF	Task Force
Tx	Transmission / Transmitter
UDP	User Datagram Protocol (OSI Transport layer)
UGC	User-Generated Content
UHDTV	Ultra High Definition TV (NHK)
UMID	Unique Material Identifier (SMPTE)
UR	User Requirements
V	Vertical
VBI	Vertical Blanking Interval
VC-1	Ex - Windows Media Video Codec, now SMPTE 421M-2006
VC-2	SMPTE code for the BBC's Dirac Video Codec
VC-3	SMPTE code for the Avid's DNxHD Video Codec
VOD	Video On Demand
VRT	Vlaamse Radio en Televisie (Belgium)
vs.	versus, against, compared to, opposed to
VTR	Video Tape Recorder
WAN	Wide Area Network
WAV	WAVEform audio file format (Microsoft)
WMF, WMA, WMV	Windows Media format, Windows Media Audio, Windows Media Video
WS	Web Service
WSDL	Web Service Description Language
WSI	Web Service Interface
X-PAD	eXtended Programme-Associated Data
XHTML	eXtensible HyperText Markup Language
XML	eXtensible Markup Language