Executive Summary
DVB Study Mission on Virtual Reality
October 2016 (CM1706)

1. Introduction
David Wood, EBU

- The aim this study mission report is to determine whether VR video is likely to be commercially successful, and consider the role DVB can play.
- The primary interest is entertainment, information, and educational content.
- There are a wide variety of Head Mounted Display (HMD) systems, and a lack of standard file formats, as well as uncertainty about motion sickness.
- The type of VR delivery formats most commonly used will depend on quality of experience needed, type of content provided, and the display system used.
- Both tethered and untethered VR HMDs will be considered, but augmented Reality and Mixed Reality are not part of this study.

2. Glossary of basic terms
David Wood, EBU

- Virtual Reality: *A technology that replicates an environment, real or imagined, and simulates a user's physical presence and environment to allow for user interaction. Virtual realities artificially create sensory experience, which in principle can include sight, touch, hearing, and smell.*
- Augmented Reality: *The addition of multimedia to a reality which is be being experienced via an HMD or a "plain screen".*
- Mixed Reality: *The combination of several realities in one experienced reality.*

3. The definition of VR and the potential range of quality of experience
Kieren Mayers, Sony Interactive Entertainment

- A wide variety of VR experiences are being developed and technology is evolving rapidly
- Important parameters impacting quality of experience include frame rate, field of view, visual acuity, stereoscopic or flat image, degrees of immersion (panorama and degrees of freedom), head tracking latency, and information overlay
- Content can be classified into three types: 2D fixed viewpoint (3 degrees of rotational freedom), 3D fixed viewpoint (3DoF), and 3D free viewpoint (6 degrees of rotational and translational freedom)

- Content must support sufficient frame-rate, sub-pixel per degree, and latency for a comfortable viewing experience. In addition, a common approach to providing graphical overlay content may be needed
- As technology is moving quickly, this assessment only provides a 'current' snapshot

4. Challenges and prospects for VR programme production
Ralf Schaefer, Fraunhofer HHI

- Content must be stitched together during post production, to avoid parallax errors, which is extremely challenging for broadcasting live events.
- Cameras configurations used for 2D content
    - Mounted in a ring capture horizontal panoramas
    - Mounted in a ball to capture spherical video
    - Mirror-based camera systems can reduce parallax down to objects within 1 m
- Camera configurations for 3D content:
    - Mirror-based systems can capture panoramic 3D images using camera pairs
    - Stereo camera pairs arranged in a star can capture 3D panorama, but have major parallax problems
    - 'Stereo by extreme overlap' creates stereoscopic content from overlapping images captured by fish-eye or wide-angle lens, or clusters of cameras
    - Lightfield approaches can be used to generate 3D images, but requires complex computing and is error-prone.
    - Other approaches: include hybrid video composition (eg Sony) uses static 3D panorama combined with 3D video content, and insertion of 3D models (eg 8i, Microsoft, HHI).
- Camera systems can be classified into direct stitching by mirror-based systems, depth-aware stitching, and depth enabled light-field rendering, each approach with different advantages and disadvantages. Overall, cost versus quality is a key consideration in selecting VR camera systems.

5. Predictions for sales and usage of VR display equipment
Thierry Fautier, Harmonic

- We first look at the market segmentation between the tethered devices (Oculus Rift, HTC Vive), game platforms (Sony PS VR) and untethered devices (Gear VR, Consumer HMD, Cardboard). We predict that untethered devices will be 10x the volume of tethered ones, that will appeal more to gamers 'community.

- We then look at the use cases for VR Video: Bonus for 2D Movies, Documentary, Concerts , News, TV Shows, Short-Form Movies , Live Sports , Sports Highlights, Mesh Video/3D Gfx.
- We assess the size of the market on the device side considering different market researches available on a 2020 horizon. A medium scenario shows $20B revenue in 2020. This is followed by a market sizing of the VR Video services on a 2020 horizon.
- We estimate that by 2020, VR will generate between $1.0B and $1.4B revenue, the largest application being Live sports. VR Theme Parks & VR arcade games will be a lucrative business for both games and video and will, just as GPS was democratized with car rental, help evangelize VR.
- We estimate that by 2020, this will represent a revenue larger than $6B, with arcade games already widely deployed in China and VR theme parks being opened progressively.

6. Trials for VR

Simon Gauntlet, DTG

- This section lists a selection of trials and content currently available for VR, including live events, live capture – post produced, post-produced, and interactive content.

7. VR content production

Simon Gower,

- This section looks at the results of production trials by BT Sport and Sky VR, including:
    - BT broadcast and production trials:
        - Basket ball game at the O2 Arena in London in January 2016): 2D single video stream (6 users)
        - FA Cup semi-final football match at Wembley Stadium, London, in April 2016: 2D multiple video views, with 3D virtual screen & data overlay(105 users)
        - Trial of a ‘near live’ studio conducted in May 2016: using different cameras. – 2D 360 video player app for Facebook and Youtube in HD / UHD.
    - Sky have produced short 360° video segments used to support main content, such as pre-season F1 testing in Barcelona.
- There are two approaches to producing VR video: editorially (simple single video stream) or viewer driven (more complex multiple video streams selected by the viewer using a special VR media app).
- Key production insights include**:**

- *Bandwidth of 50 mbps upstream; & 10-12 mbps downstream needed*
- *HMD issues include resolution, overheating, battery life, & social interaction*
- *Most VR content is currently watched on tablets, smartphones, & PCs*
- *Standardisation can help reduce range of media players broadcasters must use*
- *VoD VR appears to offer the most compelling experience vs. VR live broadcast*
- *The length of time users can comfortably view VR content may vary*
- *Switching between multi-camera video requires fast synchronisation.*
- *Synchronised viewing capabilities are needed for social viewing*
- *User journey for immersive sports events should be considered*
- *Feedback on user's gaze orientation is useful for production & marketing*
- *Multi-viewpoint VR can be used to create interactive stories*
- *Spacialised audio is important in guiding viewer's attention within a VR scene.*
- *It is difficult to monitor panoramic video to ensure compliance*
- *XAVC300 suggested archive format fault high definition panoramic video*

8. Standardisation (MPEG, W3C, Open VR…)
David Wood, EBU & Thomas Stockhammer, Qualcomm

- Principal bodies involved in VR standardisation include ISO, IEC JTG MPEG, JPEG, and DASH IF, and possibly ITU-T and ITU-F in future. It is not clear how their activities overlap which may become the dominant standards for VR.
- MPEG are developing an Omni-directional Media Applications Format (OMAF) standard, as well as a Media Orchestration (MORE) interface for video stitching and encoding, and are considering Tiling mechanisms for region of interest encoding (using a dual layer SHVC approach).
- JPEG are developing various file formats including: JPEG XT (omni-directional photographs), JPEG XS (low-latency compression format for VR), and JPEG PLENO (lightfield video format).
- 3GGP are looking at VR standardisation for wireless mobile services, considering delivery of VR video content through current as well as 5G systems.
- DASH-IF are planning test and trials of VR delivery using DASH technology

- A VR Industry Forum is currently established to promote VR: which may develop guidelines, encourage use of common formats, and share experiences with VR.

9. Audio

Thomas Stockhammer, Qualcomm

- Sound carries an essential amount of the information that allows human beings to experience and interact with their environment. Only the three pillars, high visual quality, high sound quality and intuitive low latency interactions provide the viewer with a fully immersive virtual reality experience.
- Michael Abrash, chief scientist at Oculus, is a big believer in the power of 3D sound, he said: 'It's not an addition — it's a multiplier.' It's a force multiplier when you have 3D positional sound in VR.
- DVB understands the relevance of sound for fully immersive VR experiences, and has dedicated a specific report for audio in Virtual Reality applications. The detailed report provides an overview of the fundamental principles of immersive audio for virtual reality, dedicated use cases, an overview of state-of-the-art technologies and ongoing commercially available services and technologies.
- Based on the information collected in this report, it is obvious that a significant amount of technologies already exist for VR audio, however some are still in an exploratory phase. Experiments with VR are being conducted by many parties globally, different technologies are being investigated, and standardization for VR is in its early stages.
- There is still currently no full understanding of the impact of certain factors on the sound quality for VR audio. For example,. the relative impact of capture techniques, coding and reproduction of sound still needs investigation. The group is currently not aware of any activity on formalized quality evaluation of VR audio. Object and scene-based audio technologies are big steps towards solutions needed for sound in immersive Virtual Reality.
- There already exists a basis for VR audio in Next Generation Audio Systems as defined in the next version of DVB Audio/Video specifications. Those DVB specifications are expected to be available by the end of 2016.
- Additional functionality may need to be added in DVB transport systems to consistently signal dedicated properties of the audio, such as interactivity or other VR/3D sound related metadata. However, generally Next Generation Audio is imperative for an immersive VR experience. A very relevant aspect for the reproduction of VR audio is interactive rendering. Binaural audio (headphone) rendering is expected to be the primary configuration for consumption of VR audio. The bit rates needed for VR/3D audio (if NGA codecs are used) will probably be similar to bit rates used for immersive audio (based on NGA).
- Based on the summary the internal report provides a set of recommendations for potential work for DVB, namely to initiate a commercially driven approach to potential DVB standardization work on

VR audio. Already defined audio coding formats, primarily in the context of Next Generation Audio, may be extended for supporting VR applications. VR audio should be enabled in different delivery scenarios including broadcast and Internet streaming. Harmonization in the industry is encouraged such as coordination with other organizations such as MPEG or 3GPP.

- In the short term, metadata describing Scene-based audio streams already being used by VR hardware manufacturers requires standardization, before 'Next Generation Audio' codec`s become widely available."

10. Psycho-physical aspects of VR

Ludovic Noblet, b<>com

- The discomfort that people can feel when immersed in VR is known as 'kinetosis' or 'cybersickness', however, it is often referred to as 'motion sickness'.
- Symptoms include general discomfort, ocular problems, disorientation, and nausea, which may last for several hours, but, in the vast majority of cases, is mild & temporary.
- Causes can be due to sensory conflict, where visual inputs conflict with sensory memories of spatial movements, and also postural instability, where unfamiliar movements lead to inability to maintain postural control.
- It is theorised that nausea is either a natural defence against ingested toxins, as the brain assumes sensory conflicts are caused by hallucination, or is caused by stimulation of the vagus nerve due to eye movements.
- Both extrinsic factors (including flicker, refresh rate, display width, system like, duration of exposure) and intrinsic factors (including health, genetic background, six, page, meet an anxiety, postural stability, previous experience) can lead to motion sickness. Repeated exposure, however, can desensitise users to motion sickness.
- Strategies to prevent motion sickness include: avoiding high-risk situations (avoiding exposure of young children) limiting exposure time, and limiting system features (e.g. minimising vertical parallax, avoiding screen edge effects, avoiding chromatic differences between shots, limiting retinal disparity).
- User's sensitivity to VR can be assessed in order to implement individual strategies to avoid motion sickness (through questionnaires, use of sensors, or mathematical modelling). Measuring user experience in real-time and adapting VR systems to avoid motion sickness is likely to important in future.

11. Potential VR delivery mechanisms and systems

Thierry Fautier, Harmonic

- We describe the different relation between content capture (2D, 3D), the type of device (tethered, un-tethered), and the type of consumption (HMD, 2D experience)

12. Delivery formats for VR

Thierry Fautier, Harmonic

- VR video content creation includes capture, preparation, delivery, processing, rendering, and feedback. The method of rendering and displaying VR content can influence the transmission method what is needed that each step.
    - Camera systems can provide either a concave view (outward-facing panoramic) or a convex view (inward facing towards a single scene). Panoramic cameras typically capture 180 or 360 degree video, in 2D or 3D, and operate at p30 -p60, using rectilinear format (with 3-D video using split top bottom format).
    - Encoding can be done using legacy equirectangular format, equirectangular tiles on-demand transcoding, or polygon mapping. The equirectangular tiled system is expected to be available 2016-2017, and offers the highest resolution, whereas on-demand transcoding is the most bandwidth efficient method. Hybrid approaches might be possible.
    - VR video can be delivered either in broadcast unicast mode, although broadcast content would be limited to equirectangular formats, and polygon mapping depends on proprietary transmission methods e.g. Facebook. Otherwise unicast transmission would use DASH ISO BMFF. ABR multicast is likely to be necessary to scale service at peak usage
    - Processing: The decoder can be used to pick region of interest if required for line of sight methods.

13. The prospective VR hype cycle

David Wood, EBU & Thierry Fautier, Harmonic

- Using the Gartner ‘Hype’ cycle, VR is listed as a technology moving from disillusionment to enlightenment, although other analysis positions technology at an earlier ‘inflated expectations’ stage.
- As VR has not yet been fully deployed commercially, we expect the VR video market will be more mature and ‘enlightened’ around 2020.

14. Conclusions

- It is likely the main commercial driver for tethered VR will come from gaming, whereas the main driver for untethered VR will come from immersive video for sports and music events. The demand for content will depend on its availability and quality of experience.
- DVB should cooperate with standards bodies working in VR, as members will need to adopt common specifications for stream delivery of VR content. Requirements are needed for the minimum technical quality of VR video and audio, particularly to reduce cybersickness. Requirements should be completed within two years (mid-2018)
- In terms of quality of service, consideration must be given to the desired frame rate, field of view, visual acuity, degree of visual and audio immersion, head tracking latency, and visual overlays
- VR audio will need additional support, both for broadcast and broadband transmission.
- In the short term support is needed to avoid a multiplicity of groups and proprietary panoramic 3 degrees of freedom VR video systems, and considering requirements key parameters such as frame rate, resolution, use with tablets etc. For example, Sky's provisionally specifies the following VR formats: video: 2-4K resolution, H.264, 25-50 FPS, 20-60 mbps bitrate, audio: stereo or ambisonic.
- For the longer term it is recommended to continue the study mission to follow developments such as panoramic 6 degrees of freedom VR, augmented reality, and mixed reality.
- Commercial requirements group would begin their work the questionnaire to DVB members. In addition, the group may consider developing a DVB VR garage, where VR technologies could be neutrally badged under DVB.