International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BT.2420-2

(03/2021)

Collection of usage scenarios of advanced immersive sensory media systems

BT Series

Broadcasting service (television)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

Note*: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2021

REPORT ITU-R BT.2420-2

Collection of usage scenarios of advanced immersive[1] sensory media systems

(Question ITU-R 143/6)

(2018-2020-2021)

TABLE OF CONTENTS

Page

[1] The term “immersive” in the context of this Report is deemed to include any format or medium or platform that offers or engages an audience by employing sensory based technologies such as audio, video, or haptic) and enables any form of interaction or control of the content presentation.

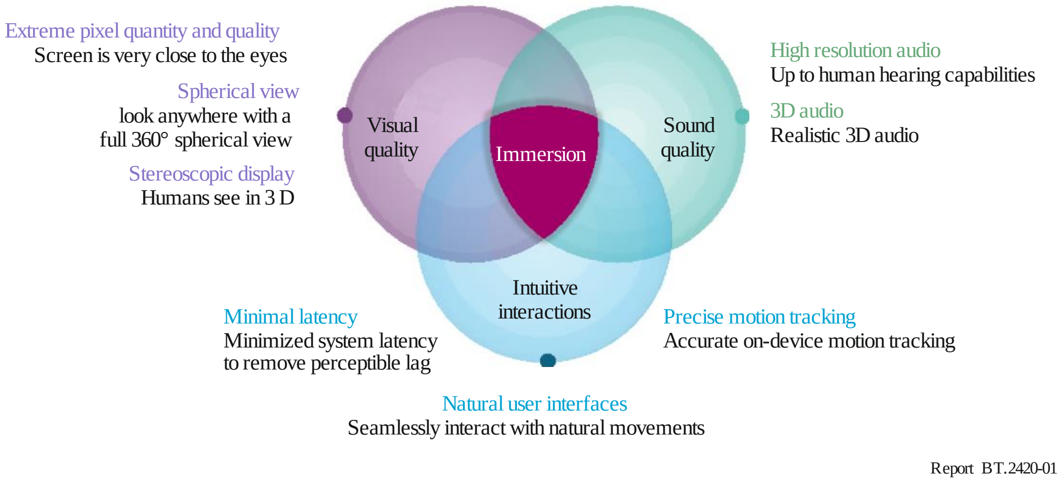
Page

1 Introduction

Advanced immersive sensory media (AISM) systems allow a user to have immersive experiences with an unprecedented degree of presence including the advanced immersive audio-visual systems. By tricking the perceptual systems of the user's brain, AISM systems can make the user believe to be somewhere else and/or somebody else. This is achieved by (re)creating audio-visual realities and allowing the user to naturally interact with these virtual environments. Figure 1 depicts the three primary quality aspects in AISM systems that contribute to immersion and presence. The sense of immersion breaks down if the information presented to these modalities does not work properly together. In some instances, users may even experience sensory sickness (see below).

FIGURE 1

The three primary quality aspects in AISM systems



This ITU-R Report is intended to describe a brief technical background and important definitions used for AISM systems, use cases for broadcasting of AISM programme material, and other challenges that have emerged through those production trials.

2 Technical background and glossary

2.1 Overview

This section provides a brief overview of technical terms and concepts. For more detailed information, the interested reader is invited to study the following guides and primers [1, 2, 3, 4].

2.2 Forms of AISM Systems

Virtual Reality (VR): A technology that replicates an environment, real or imagined, and simulates a user's physical presence and environment to allow for user interaction. Virtual reality artificially creates a sensory experience, which in principle can include sight, touch, hearing, and smell. The current VR devices primarily present content to the visual and auditory systems. On occasion, haptics information is also included.

Augmented Reality (AR): The addition of images or enhanced digital content overlaying the physical world. This can be introduced in the visual field or another sense such as audition and sound. More developed applications lead to a fusion of the physical and virtual worlds into one reality which can be experienced via an HMD as defined in § 2.3. Augmented Reality can be experienced via an

HMD or a plain screen. Examples are Microsoft's Microsoft's HoloLens1 and 2, Magic Leap 1, Bose Frames, Google Glass, Pokemon Go, and Yelp Monocle.

2.3 Presentation Modes

HMD: A head mounted display (HMD) is a display worn on the body that fits over a user's head. It has small display optics in front of the eyes and is usually equipped with additional sensors to track the viewer's head motions such as coordinate positions, pitch, roll, and yaw. In some instances, the position of the user's gaze is also captured. HMDs for AR allow the user to passively view the contextual physical world (e.g. Microsoft HoloLens), whereas HMDs for VR occlude perception of the contextual physical environment. For VR applications, some HMDs enable event-driven user integration of mobile devices to act as system displays and/or processors for the device.

Magic Window: This presentation mode enables exploration of 360° video content to be accessible without use of an HMD on: mobile devices, desktop computers, or televisions.

Using a mouse pointer or finger gesture, the user drags and rotates the 360° image on the screen to see a portion of the 360° video scene. Depending on the resolution of the 360° video, some applications may also allow the user to zoom into the scene. Further, the motion and position sensors on mobile devices allow a user to steer the mobile device in a desired direction to see a select region of the 360° video through the screen. In all cases, the accompanying sounds and acoustic scene should adapt accordingly. This mode of content presentation does not provide full immersion, but it does enable an extended mode of content interaction and consumption that has low risk for sensory sickness.

Second Screen: An AISM second screen could offer specific VR vantage points that accompany the regular television programme. These experiences do not necessarily have to replace the current television viewing paradigm, but, rather, may complement TV programmes by offering second screen services synchronized to the TV broadcasting.

Haptic interface: A haptic interface is a device that allows a user to interact with a system by receiving tactile feedback. Through the haptic device, the user receives feedback in the form of haptic stimuli, which include vibrations and changes in pressure and temperature. Haptic devices are expected to serve as interfaces for enhancing viewer immersion in broadcast programmes by presenting viewers with haptic stimuli linked with the broadcast content.

2.4 Modes of Interactivity

Three Degrees of Freedom (3DoF): Programme material in which the user can freely look around in any direction (yaw, pitch, and roll). A typical use case is a user sitting in a chair looking at 3D VR/360° content on an HMD as shown in Fig. 2.

Three Degrees of Freedom Plus (3DoF+): Programme material in which the user is free to look in any direction (yaw, pitch, and roll), plus limited translation movements due to the head movements not being centred on the optical and acoustical centre. This provides support for perceptual effects such as motion parallax which strengthen the sense of immersion. A typical use case is a user sitting in a chair looking at 3D VR/360° content on an HMD with the capability to move his head slightly up/down, left/right, and forward/backward as shown in Fig. 2.

Multiple Vantage Points: While the definitions for 3DoF and 3DoF+ are based around a single point of observation, these concepts can be extended to a scenario where users may experience a scene from multiple discrete vantage points.

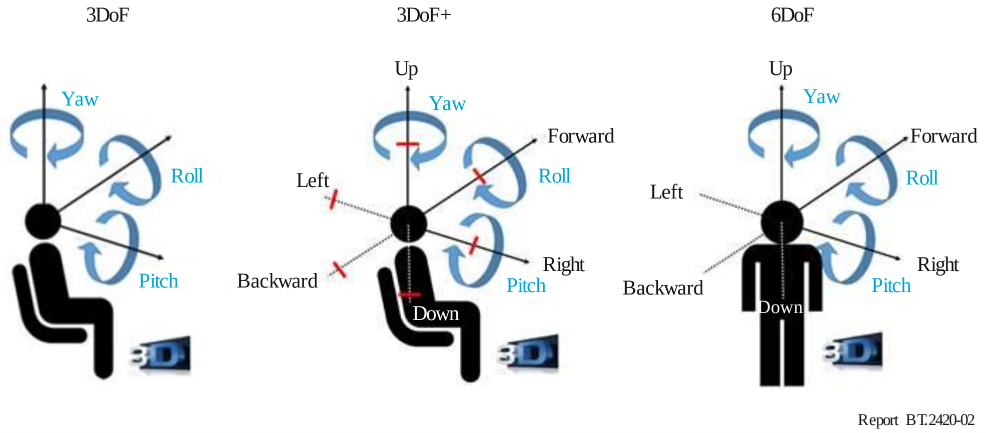
Six Degrees of Freedom (6DoF): Programme material in which the user can freely navigate in a physical space. The self-motion is captured by sensors or an input controller. Both rotation (yaw, pitch, and roll) and translation (x, y, z translation) interactions are possible. A typical use case is a user freely walking through 3D VR/360° content (physically or via dedicated user input means) displayed on an HMD as shown in Fig. 2.

It is currently expected that the majority of VR experiences to be deployed in the near term are going to be 3DoF. Mass market consumer devices and services supporting 6DoF can be expected to be widely available by 2020 (see [5, 6]).

FIGURE 2

Freedom of movement in a three-dimensional space

(This Figure is taken from ISO/IEC JTC1/SC29/WG11 N17264 − Working Draft 0.4 of Technical Report on Architectures for Immersive Media)



Diegetic and Non-diegetic Elements: A diegetic audio or video element is a stable element in the virtual scene. It is spatially rendered to be dependent on the position and movements of the user's head as recorded by the HMD sensors. The perceived position a diegetic element is unaffected by head motion. In contrast, a non-diegetic audio or video element is spatially rendered independent of the virtual scene and changes with movements of the user's head. The perceived position of the element is not constant and updates as the HMD receives user-driven information updates from the device sensors. An example of a diegetic audio element could be a person talking in the background that is not present in the virtual scene. An example of a non-diegetic video element could be a graphical overlay, such as end credits.

Linear Narrative vs. Non-Linear Interactive Content: A Linear narrative program moves sequentially through time and does not allow a user to modify how the content sequence is presented in time. Linear program material is the primary content type used in broadcasting. In contrast, Non-linear narrative programs allow a user to interact with the content. This allows a user to modify how the content sequence appears in time. Non-linear program material is common in the gaming community. Example non-linear AISM programme experiences may enable a user to walk anywhere within the scene. The user's motion and behavioural interactions with the content will directly influence how the programme material appears in time. Because it is not clear how such non-linear interactive content could be delivered via broadcast, this report will primarily focus on linear narrative content.

3 ITU-R related use cases

3.1 Overview

This section lists a number of use cases that are related to broadcast systems. Use cases less relevant to broadcasting (i.e. unicast) are studied elsewhere, e.g. at 3GPP (see [3, section 5]).

3.2 The Infinite Seat VR Broadcast

Infinite Seat VR Broadcast is a method for capture of live-VR content. At the event capture site, omnidirectional camera and microphone rigs can be placed at certain seating/viewing locations. Each audio and video capture rig delivers a distinct experience, corresponding to the unique seating locations at the event site. The infinite seating experience can be further augmented with additional audio elements (e.g. commentator voice) or visual elements (e.g. player statistics) and with techniques such as action replay, or cuts from different vantage points.

The viewer might be able to select between different seats individually, while the orientation at each seat is defined by the user's rotation (yaw, pitch, and roll).

3.3 Linear Narrative Cinematic VR Broadcast

During Linear Narrative Cinematic VR Broadcast, a consumer watches a VR movie from a fixed point in the scene (3DoF or 3DoF+). The consumer can freely turn and move his head to observe details in the scene and may follow the story by listening and watching the actors.

3.4 Free Viewpoint Television

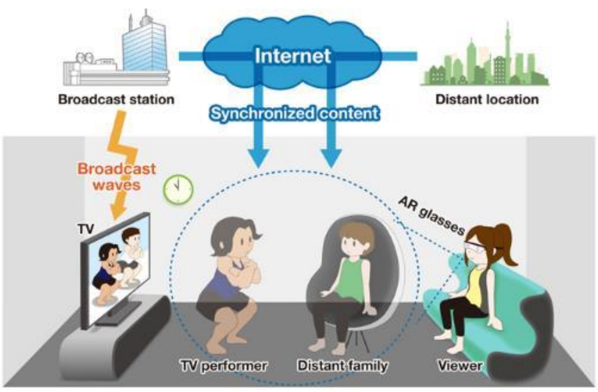
Free Viewpoint Television is a visual media that allows users to view a three-dimensionally recorded scene by freely changing their point of observation. By changing the vantage point of the broadcasted scene, the reproduced sound scene and audio elements will adapt accordingly. Free viewpoint images may also be presented simultaneously with a normal television programme by delivering additional 3D objects linked to the TV programme in real-time through broadband Internet.

3.5 Integration of TV and AR

AR has the potential to enhance the experience of TV viewing. One new viewing style is "virtual space sharing", where viewers simultaneously experience AR of six degrees of freedom (6DoF) on their handheld devices or glasses equipped with AR technology while watching a TV programme on television [7]. Figure 3 illustrates such a concept of the integration of TV and AR, where three-dimensional objects of TV performers or family members and friends in different locations are combined and displayed through the AR device. Viewers can feel as though they are sharing the same space as TV performers, family, and friends while watching TV. With AR glasses, performers or persons are displayed in their actual size to provide an increased sense of reality.

FIGURE 3

Concept of the integration of TV and AR



3.6 Haptic AR Broadcast

The term "haptic" refers to the sense of touch. While broadcasting is traditionally a medium for conveying mainly visual and/or audio information to the audience, the addition of haptic information would enable the provision of broadcast services with new sensory experiences that could greatly enhance qualitative experiences. In a sporting event, for example, vibrating a haptic device when a ball is in motion or bouncing would provide viewers with an enhanced immersive experience as if they were participating in the event rather than simply spectating. A haptic interface could also help visually or hearing-impaired members to intuitively understand the broadcast content.

3.7 HMD-based Content Consumption

3.7.1 Private VR Television

Private VR Television presents conventional 2D television programme material to a viewer in a virtual environment. The 2D television programme material is rendered directly in front of the user on a virtual rectilinear screen within the virtual environment. The virtual scene may adapt the audio and video to the user's head motion. In a different mode, the virtual screen may be head-locked, meaning that the screen is always displayed in front of the user independent of their head motions.

3.7.2 Enriched VR Television

During an Enriched VR television viewing experience, the television program is rendered on a virtual 2D-rectilinear screen inside the 3D-virtual environment. A 360° scene covers the background of the spherical environment. The user can activate the contextual menu for displaying additional information, e.g. sport statistics related to the current game, additional movie information, the electronic program guide, and a selection of different viewing angles. Also, a stereo camera can be attached to the HMD that captures the scene and depth around the viewer enabling reproduction of a mixture of real and virtual images on the HMD. In this experience, the real scene near the viewer (including the viewer's hands and body) is displayed to the user along with presentation of the virtual world (virtually represented environment and elements) seamlessly according to depth.

3.8 Supporting Visually or Hearing-Impaired Audience Members

VR/AR technology could enable an improved broadcast content experience for visually or hearing-impaired audience members. A recent article [8] demonstrated that VR/AR glasses may be used to help the visually impaired recognize objects. Haptic information is also useful for visually or hearing-impaired users to understand the content. In this way, it may be possible to use VR/AR devices to provide visually or hearing-impaired consumers with specific enhancements to broadcast content that enables a markedly improved programmed experience.

4 Broadcaster VR Productions and Trials

4.1 Overview

There is significant interest in using both advanced audio and video technologies for VR production and programme applications from programme creators, broadcasters, and media consumers. Many TV broadcasters are undertaking production trials to get familiar with the production workflow and to evaluate the feasibility of VR programme production and delivery. A representation of these broadcasters includes: BBC, Sky, ZDF, Arte, Canal+, RAI, NBC, CBS, DirecTV, Telemundo, Turner Sports, Eurosport, NHK, Nippon TV, TV Asahi, Fuji TV, and TBS-TV. The DVB authored a report referenced in § 6, that includes an overview of these engagements with technical details of 32 production trials [2, section 6]. Both experienced Video FX companies, as well as new specialized startups with proprietary VR equipment and production tools, are collaborating with broadcasters in VR productions (e.g. NextVR, Jaunt, Within, Felix & Paul). Other agencies are specializing in creation of advertisements for VR (e.g. OmniVirt, Advrtas, VirtualSKY). A recent case study on 360° advertisement [9] suggests that the 360° ad formats can trigger 85% engagement on mobile and 33% engagement on desktop respectively.

For distribution, many broadcasters have developed their own mobile device VR applications. Typically, these apps developed by broadcaster do not support content streaming. Consequently, it is necessary for users to download the VR content in its entirety prior to viewing. Recently, some broadcasters began to provide their own on-demand VR distribution channels (e.g. ABC News VR, Discovery VR). These VR distribution channels are hosted on the content platforms of HMD providers (Oculus, Viveport, Samsung VR, PlayStation VR), at streaming services (YouTube 360, Facebook 360, Twitter, dailymotion, Vimeo), or accessible via websites using WebVR, Flash, or other HTML5 extensions.

4.2 Content Types in VR trials

4.2.1 Examples content type of VR trials

This section provides a selection of resources of typical broadcast content produced for VR consumption. Most of these examples have a duration of less than 15 minutes.

In addition, media types outside the broadcast sector, such as: newspapers, comic books, radio, and musicals are using VR and AR to engage user beyond the initial content consumption endpoint or production to provide users with additional footage (such as behind-the-scene reports), bonus material, or marketing content (e.g. The New York Times, The New Yorker, The Guardian, USA Today, Huffington Post, National Geographic, Madefire Comics, CBC Radio, Dali Museum, School of Rock musical).

4.2.2 Sport and Sport Highlights

– **Rio Summer Olympics**

NBC: http://www.nbcolympics.com/news/experience-rio-olympics-virtual-reality

BBC: http://www.bbc.com/sport/36883859

– **Basketball**

http://www.recode.net/2016/10/20/13341408/nba-virtual-reality-games-nextvr

http://www.nba.com/2015/news/10/27/virtual-reality-available-for-golden-state-warriors-new-orleans-pelicans-season-opener/

– **American Football**

BTN: http://btn.com/2016/11/10/btn-to-become-first-college-sports-network-to-produce-live-football-game-in-virtual-reality

https://www.cnet.com/news/nfl-nextvr-highlights-virtual-reality-super-bowl/

– **Soccer/Football**

https://www.engadget.com/2016/12/07/this-weekends-mls-championship-will-be-broadcast-in-vr

https://www.nextvr.com/bundesliga

– **Baseball**

https://samsungvr.com/channels/58091b25c94f9a001998e72f

– **Hockey**

http://venturebeat.com/2015/02/26/nhl-streams-a-hockey-game-in-360-degree-virtual-reality

https://www.nhl.com/news/nhl-introduces-virtual-reality-experiences/c-279085566

– **Boxing**

Fox: http://fortune.com/2016/01/21/fox-sports-nextvr-team-on-boxing

DirecTV: http://variety.com/2015/digital/news/directvs-first-virtual-reality-app-takes-boxing-fans-ringside-1201613503

– **Golf**

Fox: http://www.sportsvideo.org/2016/06/14/fox-sports-nextvr-drive-virtual-reality-experience-at-u-s-open

– **Tennis**

France Television: http://advanced-television.com/2016/05/20/france-televisions-airs-roland-garros-in-4k-vr

https://www.prolificnorth.co.uk/2016/07/laduma-films-wimbledon-in-360-degrees

– **Racing**

NBC, Horse Racing: http://www.sportsvideo.org/2016/05/06/live-virtual-reality-hits-the-track-with-nbcs-first-ever-kentucky-derby-vr-production/

Fox, Car Racing: http://fortune.com/2016/02/18/fox-sports-daytona-500-virtual-reality

– **Extreme Sport**

Red Bull Media: http://www.redbull.com/us/en/video-series/1331829865492/vr-hub

– **E-Sport**

https://www.pastemagazine.com/articles/2016/11/e-sports-in-vr.html

http://www.sliver.tv/events

4.2.3 News

ABC News VR: http://abcnews.go.com/US/fullpage/abc-news-vr-virtual-reality-news-stories-33768357

The Big Picture – News In VR: https://www.youtube.com/watch?v=C5qbR5SQleY

The Economist: http://visualise.com/case-study/economist-vr-app

Associated Press: www.bigstory.ap.org/VR360

4.2.4 Documentaries

BBC: http://www.bbc.co.uk/taster/projects/invisible-italy

Doctors Without Borders: http://visualise.com/case-study/msf-doctors-without-borders-forced-home

4.2.5 Television Shows

NBC, Saturday Night Live: https://www.youtube.com/watch?v=6HS9h4xFRww

ABC, Dancing with the stars: http://abc.go.com/shows/dancing-with-the-stars/news/updates/vr-05022016

Competitive Cooking shows: https://www.youtube.com/watch?v=JpAdLz3iDPE

4.2.6 TV series, Episodic

http://www.hollywoodreporter.com/behind-screen/virtual-reality-tests-episodic-story-940425

http://variety.com/2016/digital/news/hulu-ryot-virtual-reality-news-comedy-show-1201866110

4.2.7 Animation

Spotlight Stories: http://www.polygon.com/2017/1/24/14370892/virtual-reality-first-oscar-nominated-short-film-pearl

BBC: http://www.bbc.co.uk/taster/projects/turning-forest

4.2.8 Music Videos

Reeps One: https://www.youtube.com/watch?v=OMLgliKYqaI

Muse: https://www.youtube.com/watch?v=91fQTXrSRZE

The Who: http://www.billboard.com/articles/6312181/the-who-new-app-greatest-hits-virtual-reality

4.2.9 Concert Experiences

Paul McCartney: https://www.jauntvr.com/title/PaulMcCartneyAtCandlestick-LiveAndLetDie.v12

LA Philharmonic Orchestra: http://www.laphil.com/vanbeethoven

Kasabian: http://visualise.com/case-study/kasabian-o2

4.2.10 Special Event Content

NBC, US presidential debates: http://fortune.com/2016/09/21/presidential-debate-virtual-reality

BBC, London New Year's Eve Fireworks: http://www.bbc.co.uk/taster/projects/new-years-eve-fireworks-360

4.2.11 Features Films or Promo Teaser

BBC, Planet Earth II: http://www.bbc.co.uk/programmes/articles/365zWpz7HypS4MxYmd0sS36/planet-earth-ii-in-360

ZDF, TEMPEL: http://visualise.com/case-study/360-trailer-zdfs-tv-series-tempel

HBO, Game of Thrones interactive VR experience: https://www.framestore.com/work/defend-wall

20th Century Fox, Wild: http://www.roadtovr.com/ces-2015-fox-debut-wild-vr-360-movie-experience-starring-reese-witherspoon

Sony, Ghostbusters: http://www.theverge.com/2016/6/29/12060066/ghostbusters-dimension-the-void-times-square-madame-tussauds-vr

4.3 VR/AR trials

4.3.1 VR trials

4.3.1.1 NHK

4.3.1.1.1 VR in collaboration with TV programmes and events

NHK conducted VR and AR projects including "NHK VR NEWS", "Panorama Tour", "8K VR Theatre", "Augmented TV", and others in collaboration with existing TV programmes and events.

Shooting systems currently available at NHK include Ricoh's Theta S, Samsung Gear 360, Kodak's SP360 4K, and Canon's EOS Kiss. Cameras are sometimes mounted on a stabilizer with angular sensors to record camera directions. VR content is distributed on the NHK website.

Links to examples:

NHK VR News: http://www.nhk.or.jp/d-navi/vr/

Document 72 hours: http://www.nhk.or.jp/d-navi/vr/72h/

Sports VR: https://pid.nhk.or.jp/event/nspo/detail/event19.html

Rio Paralympics athlete's view in VR: http://www.nhk.or.jp/sg-project/seed2/pa-2016/vr/

In June 2017, NHK broadcast a programme titled "BS1 Special – Real Trump World: The World That Created the New President Explored with a 360° Camera". It was filmed using a 360° camera, and broadcast as a programme for a normal-sized screen. The 360° video was delivered simultaneously over the Internet in sync with the broadcast. It was the world's first experiment allowing the viewer to freely look at images "outside" the frame that were not visible on the television screen by moving their mobile device up or down and to the left or right. The question arose as to whether viewers would find 360° video delivered in sync with television broadcasts appealing.

Public broadcaster efforts to engage with VR like the "Trump World" broadcast have only just begun and many issues remain. There is an open question regarding what public broadcasters should do when it comes to basic telecommunications services looking toward the 2020 Tokyo Olympics, Paralympics, and beyond. VR will offer an important perspective in keeping with the aspirations to become a public service media. The research report entitled "The Meaning of VR: Delivering 360 Degree Videos by Public Service Broadcasters Towards 2020 and Beyond" is available at http://www.nhk.or.jp/bunken/english/reports/pdf/report_17121201.pdf.

4.3.1.1.2 360° VR image system for comparative views

4.3.1.1.2.1 Overview

Situations in disaster areas are reported by distributing 360° images from the scene via network services such as NHK VR News. It is informative for viewers to be able to see not only how the sites are damaged but also how they are being reconstructed. NHK has developed a VR system that can show comparative views captured right after a disaster and captured several months after the disaster at the same position by displaying both images side-by-side [10]. The user can watch any direction of the 360° scene and change the border of the two images horizontally by a user interaction.

4.3.1.1.2.2 Capturing 360° images

In 2018, 360° images of areas severely damaged by a typhoon were captured at the quasi-same position in three cities in western Japan at two different times, initially captured immediately after the disaster and a second captured several months later (Fig. 4). Insta360 Pro was used to capture 360° images with the resolution of 3 840 × 1 920 pixels for each eye.

FIGURE 4

Capturing 360° images at a disaster site

4.3.1.1.2.3 Displaying comparative views

The display system for comparative views consists of a workstation, an HMD, and a controller, as shown in Fig. 5. Captured images are rendered by the workstation and fed to an HMD. The centre of a sphere is set to the position of the camera in a virtual space to render the scenes. As shown in Fig. 6, two 360° images, one captured right after the disaster and the other several months later, were mapped to the sphere. A moveable white border that the user can control between the two images indicates where the image transitions between the past and more current scenes. Figure 7 shows images of comparative views displayed on the HMD. The user is free to watch any direction of the 360° scene.

FIGURE 5

Hardware configuration

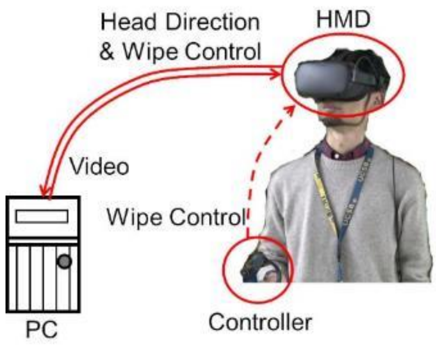


FIGURE 6

Rendering a scene

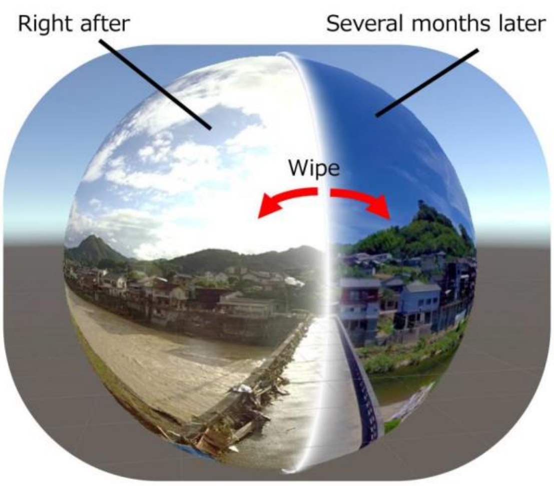


FIGURE 7

Images of comparative views displayed on an HMD

4.3.1.1.3 360° VR image system with 8K display

4.3.1.1.3.1 Overview

There are two major quality-related problems to address in 360° VR imaging to make viewing high-quality VR images comfortable; 1) the resolution of 360° images, and 2) the resolution of head mounted displays (HMDs). To prevent users from perceiving the pixel structure of a display when viewing part of a 360° image, future HMD's must include a wider field of view than currently available as well as having significantly higher spatial resolution. To support this, it is necessary for 360° images to have a much higher spatial resolution the typically captured. To cope with these problems, a prototype HMD using an OLED panel with a spatial resolution of 8K × 4K was developed, and a 360° image with a spatial resolution of 30K × 15K was captured.

4.3.1.1.3.2 HMD for VR with 8K resolution

An OLED panel with a spatial resolution of 8K × 4K was used for the HMD. Table 1 shows the specifications of the display.

TABLE 1

Specifications of 8K OLED panel used for HMD

Screen size	8.33-inch diagonal
Spatial resolution	7 680 × 4 320 for R, G, and B each
Pixel pitch	1 058 ppi (24 μm)
Frame frequency	60 Hz
Developer	Semiconductor Energy Laboratory Co., Ltd.

The HMD consisted of the OLED panel, optical components, a motion sensor unit, and image processing circuits. The motion sensor unit consisted of a 3-degree acceleration sensor, a 3-degree angular speed sensor, and a 3-degree geomagnetic sensor, and could detect the viewing direction of a user three-dimensionally in real time.

The dimensions of the panel were 103.68 × 184.32 mm, and the size of both the left and right images was 103.68 × 92.16 mm. Designing optical components so that the field of view is about 100 degrees is ideal with an 8K display, and this is achieved with a focal length of 38.67 mm. Lenses with a focal length of 50 mm were used due to easy availability. This resulted in a decreased field of view of 85.33 degrees.

It is ideal that the distance between the centres of the right and left lenses correspond to the average pupillary distance, which is about 65 mm. However, the size of the OLED panel was a little bit larger. To best use the 8K panel, the optics were designed by using an optical beam shift.

Moreover, as shown in Fig. 8, the prototyped HMD was not a goggles-type one but a hand-held one mounted to a support arm.

FIGURE 8

HMD for 8K VR

Report BT.2420-03

4.3.1.1.3.3 Capturing 360° spherical images

360° spherical images with a significantly high resolution were obtained from multiple sub-images captured by a still camera with a 5 472 × 3 648 resolution by using a robotic camera mount for automated panorama shooting. A total of 144 sub-images (12 × 12) was captured for a 360° spherical image except for the area at the foot of the camera mount. The 144 images were then stitched into a rectangular spherical image of 55 184 × 21 524 (55K × 22K) pixels by using the equirectangular projection (ERP)format. The ERP format is likely to be adopted for the MPEG-I Part 2: Omnidirectional Media Format (OMAF) for representing 360-degree video on a 2D plane.

The 55K × 22K image was scaled down to 30 720×15 360 (30K × 15K) pixels, which is sufficient to represent the 360° sphere, and a black bar was inserted at the bottom as shown in Fig. 9.

FIGURE 9

Stitched spherical image with 30 720 × 15 360 pixels

Captured at Tsukuba Space Center, Japan Aerospace Exploration Agency (JAXA)

Report BT.2420-04

4.3.1.1.3.4 Presenting 8K × 4K VR images

To present 8K × 4K VR images on the HMD, the rectangular spherical image was first re-mapped to a dome-shaped spherical image. The direction in which the user is facing was detected every 10 ms by the motion sensor attached to the HMD. In accordance with the direction, the corresponding area, which was at a size of 3 840 × 4 320 pixels (4K × 4K), was clipped from the re-mapped spherical image. The clipped image was then corrected to compensate for lens distortion and displayed side-by-side on the HMD.

4.3.1.1.4 Capture/display system for VR images with resolution beyond 8K

4.3.1.1.4.1 Overview

VR with image resolutions beyond 8K is expected to provide highly immersive experience with a sense of presence and reality. NHK set up a display system that projects images of over-8K resolution to a large cylindrical screen that provides a horizontal field of view of approximately 180° by using eight 4K projectors [11].

4.3.1.1.4.2 Capture system

A camera array consisting of three 8K cameras was set up to capture VR images covering a 180° field of view, as shown in Fig. 10 and Table 2. The three 8K cameras were aligned radially to capture synchronized images. VR images with the equirectangular format were produced through post-production including stitching and grading, as shown in Fig. 11.

FIGURE 10

Camera array consisting of three 8K cameras

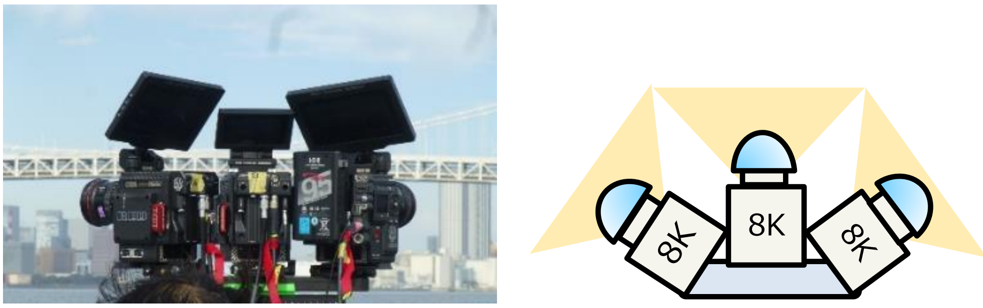


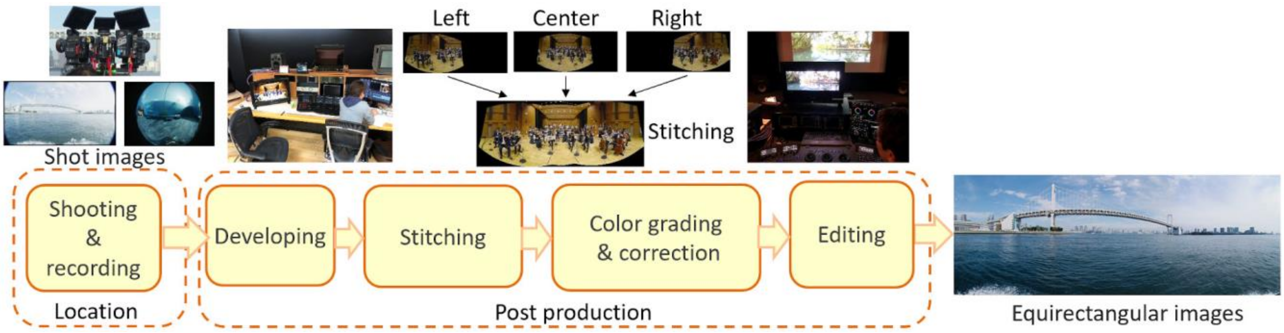
TABLE 2

Specifications of the camera array

Camera	RED / 8K MONSTRO × 3 units
Spatial resolution	8 192 × 4 320 (each camera)
Frame frequency	59.94 Hz

FIGURE 11

Workflow of high-resolution VR



4.3.1.1.4.3 Display system

A projector array consisting of eight 4K laser projectors was constructed to display the over-8K resolution images on a cylindrical screen, as shown in Fig. 12 and Table 3. Each projector was fixed in a portrait orientation. The array was placed 4 m above the floor to minimize the shadow of viewers on the screen. The diameter and the height of the screen were 11 m and 4 m, respectively. The bottom part of the screen was a round shape to widen the vertical field and minimize image distortion due to changes of a viewing position. The playout system geometrically converted the equirectangular images for each projector using the 3D shape model of the screen to display cylindrical views.

FIGURE 12

Display system

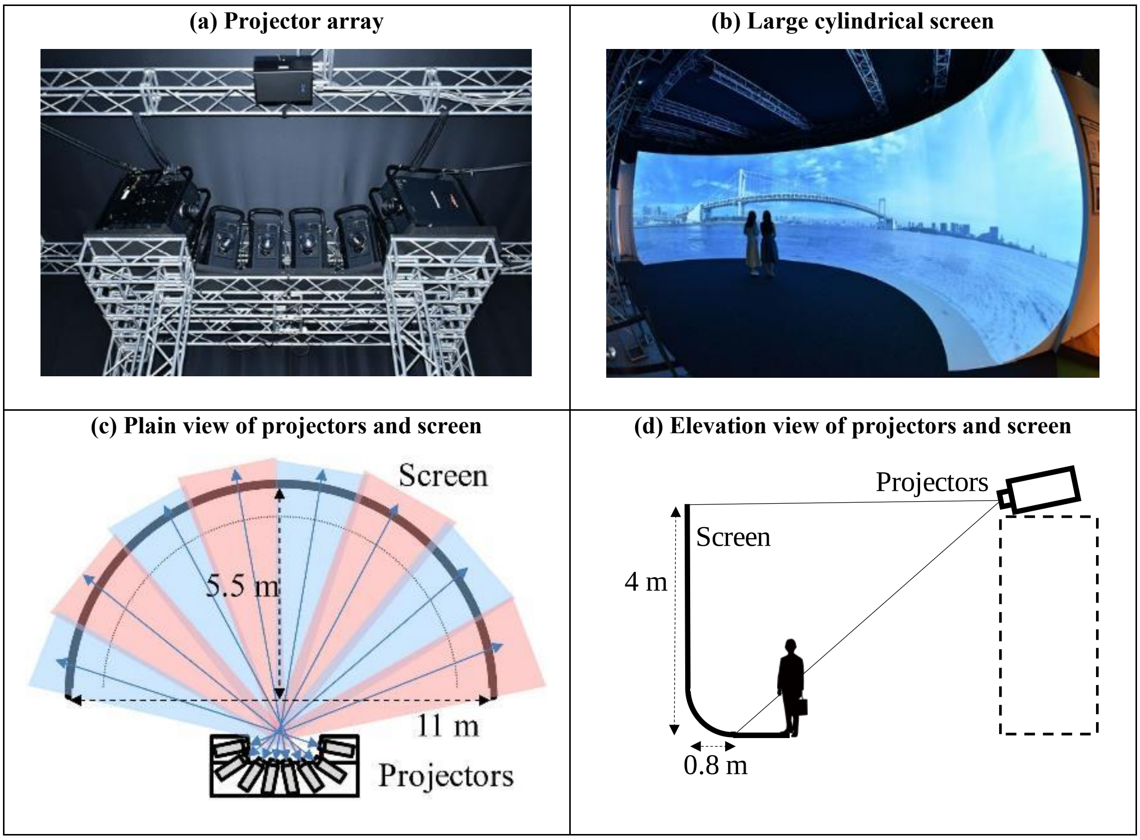


TABLE 3

Specifications of the display system

Projector array	4K laser projectors × 8 (Panasonic / PT-RQ13KJ)
Spatial resolution	3 840 × 2 160 (each projector)
	About 12K × 4K (displayed images)
Frame frequency	59.94 Hz
Screen	About 180° horizontal field of view Cylindrical (diameter: 11 m, height: 4 m) 19-face polyhedron

4.3.1.1.5 Immersive sound content for VR images

4.3.1.1.5.1 Overview

NHK has produced sound content for 3DoF VR images using audio-related metadata based on the Audio Definition Model (ADM) [19]. VR images were provided as animation of computer graphics data rendered by Unity. A user can walk through a jungle and encounter animals. VR images rendered into 1 920 × 1 080 or 3 840 × 2 160 are displayed on a projector, an LCD monitor or a head mount display (HMD) (see Fig. 13). The sound content consists of background sound, narrations, audio description and near-field sound as shown in Table 4. The main sound content rendered into 24 sound signals or their binaural stereo signals are reproduced using 24 loudspeakers of sound system H or headphones.

FIGURE 13

Reproduction system

(a) Projector and loudspeakers	(b) HMD and headphones	(c) Second sound device (wearable neck loudspeaker system)

TABLE 4

Specifications of sound content

Audio programmes (1)	60 programmes (3 viewpoints × 10 languages × 2 reproduction systems (main and second devices))
Audio objects (2)	55 audio objects
Audio channels (3)	128 channels
Background sound	72 channels (9+10+3 (sound system H [18]) × 3 viewpoints), 3 audio objects (3 viewpoints)
Music	36 channels (4+7+0 (sound system J [18]) × 3 viewpoints), 3 audio objects of 7.1.4 (3 viewpoints)
Narrations	10 channels (mono × 10 languages), 30 audio objects of mono (3 viewpoints × 10 languages)
Audio description	4 channels (2 channels × 2 languages), 16 audio objects of mono (4 objects per signal × 2 channels × 2 languages)
Near-field sound	6 channels (2 channels (stereo)) × 3 viewpoints), 3 audio objects of stereo (3 viewpoints)

(1) An audio programme contains all audio contents including narration and background music to make the complete mix. The audio programme has a single set of parameters such as language. This is specified as audioProgramme element in Recommendation ITU-R BS.2076.

(2) An audio content refers to an audio object such as narration, background music and sound effects. The audio object contains a set of the actual audio signals with the format including a loudspeaker layout or reproduced positions. This is specified as audioObject element the in Recommendation ITU-R BS.2076.

(3) An audio channel is an actual PCM audio signal. This is specified as audioTrackUID and audioChannelFormat elements in Recommendation ITU-R BS.2076.

4.3.1.1.5.2 Background sound for different viewpoints

Background sounds including music and roars of animals from three viewpoints were produced as channel-based sound signals of the sound systems H (9+10+3) and J (4+7+0) in Recommendation ITU-R BS.2051. Users can change a viewpoint by selecting an audio object for background sound. Three viewpoints at different heights of middle (human's-eye view), lower (mouse's-eye view) and upper (bird's-eye view) were provided (see Fig. 14). The area of visibility was limited by audio-related metadata to match visual contents with narration and audio description.

FIGURE 14

Viewpoints at different heights

(a) Middle **(b) Lower** **(c) Upper**

4.3.1.1.5.3 Narration and audio description

The main narration was provided as a static audio object of a monophonic sound signal located in front of the viewer. Multilingual narration was provided by audio objects for a narration (Japanese and English). A user can adjust the level balance between narration and background sound. The range of level adjustment was limited by the audio-related metadata. The user can also turn off the narration.

Audio description for individual animals was provided as static audio objects of a monophonic sound signal. The sound signal temporally recorded in the renderer is reproduced upon the user's request. A single sound signal conveys multiple audio description objects along the timeline.

4.3.1.1.5.4 Near-field sound or second sound device

The near-field sound of stereo signals including roars of animals was provided for a second sound device, a wearable neck loudspeaker. The audio programme for near-field sound is automatically switched in conjunction with the audio programme of background sound according to the viewpoint. A second device is connected via Bluetooth. The wearable neck loudspeaker system equipped with a vibrator makes the user experience sound signals as haptic stimuli.

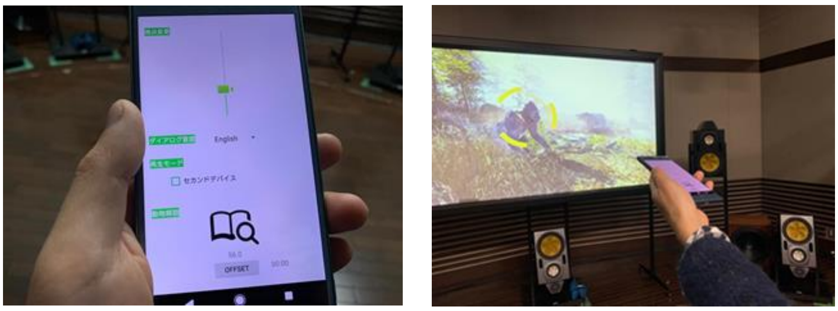
4.3.1.1.5.5 User interface

The Open Sound Control (OSC) protocol [39] was used for the user interface to control audio and video renderers. The OSC message including IDs of the audio programme and audio object for the user's action and position was conveyed via the User Datagram Protocol (UDP). A user interface was developed as an application of mobile phone and a user can select reproduction conditions and point to an audio description object using it (see Fig. 15).

FIGURE 15

User interface

(a) Application installed in mobile phone **(b) Pointing out audio description object**



4.3.1.2 Nippon TV

Nippon TV conducted VR content distribution trials twice on internet streaming services in 2016. This content was captured with a 360 degrees spherical camera and uploaded to a streaming server in a spherical image format. The content was provided as a viewing experience in a cropped rectangular image format for smartphones and HMDs. This content is aimed at gathering attention for subsequent broadcast TV programmes.

Technical specifications:

Image format: equirectangular

Picture aspect ratio: 2:1

Pixel count (horizontal × vertical): 1 920 × 960

Pixel aspect ratio: 1:1 (square pixels)

Frame frequency (Hz): 60/1.001

Scan mode: interlace

4.3.1.3 TV Asahi

TV Asahi provides VR attractions to the public at its annual Summer Festival. These attractions are related to TV programmes, including both elements of live action and computer graphics.

In 2016, two GoPro cameras were used in combination to make 360 degrees images, and HMDs such as Oculus Rift DK2 and Samsung Gear VR were used to present them. The elements of computer graphics were developed with the game engine Unity 5.0.

In 2017, the latest content was filmed by NOKIA's OZO (camera). Through the use of this camera, we were able to provide the content with the high-quality images and high realistic sensation. Since HMDs have age limit, TV Asahi developed an application compatible to monocular VR viewer for children.

Aside from movies, TV Asahi also provides other experiments that have a dramatic impact on users by using motion sensors like Leap Motion, air blowers, and vibration devices.

4.3.1.4 Fuji TV

Fuji TV provided VR content on VOD as one of its internet services, "FOD VR", in November 2016. VR content is downloaded with a dedicated app, and users are provided with a 360° viewing experience. The direction of the view follows a user's motion with a sensor when using a smartphone or HMD. In 2017, Fuji TV also provided music programmes, drama, variety shows, and promotion contents of terrestrial broadcasting programmes.

Technical specifications:

Panorama format: equirectangular (360° × 180°)

Pixels (horizontal × vertical): 4 000 × 2 000 max

Frame rate: 30-60 fps
Video recording systems: Samsung Gear 360, Kodak SP360 4K, six GoPro cameras, and NOKIA OZO.

Link to example: http://www.fujitv.co.jp/fujivr/

4.3.1.5 TBS-TV

Tokyo Broadcasting System Television (TBS-TV) has produced VR content in a variety of categories such as music, sports, drama, informational programme and sport entertainment competition programme (as known as Ninja Warrior, or SASUKE), etc. This VR content was provided to the public at its summer event “DELI-SACAS 2017” for about 40 days. Live VR content was distributed to the public there through internet links from a TV studio of a live TV show. VR viewers enjoyed the feeling as if they were audiences at this TV studio of a live TV show. This VR content was shot by 3D 360° VR camera NOKIA OZO and/or Kodak Digital Cameras SP360 4K.

Technical specifications:

Image format: equirectangular

Picture aspect ratio: 2:1

Pixel count (horizontal × vertical): 3 840 × 1 920

Pixel aspect ratio: 1:1 (square pixels)

Frame rate: 30-60 fps

Video recording systems: Samsung Gear 360, Kodak SP360 4K, and NOKIA OZO

4.3.2 AR trials

4.3.2.1 NHK

4.3.2.1.1 Augmented TV

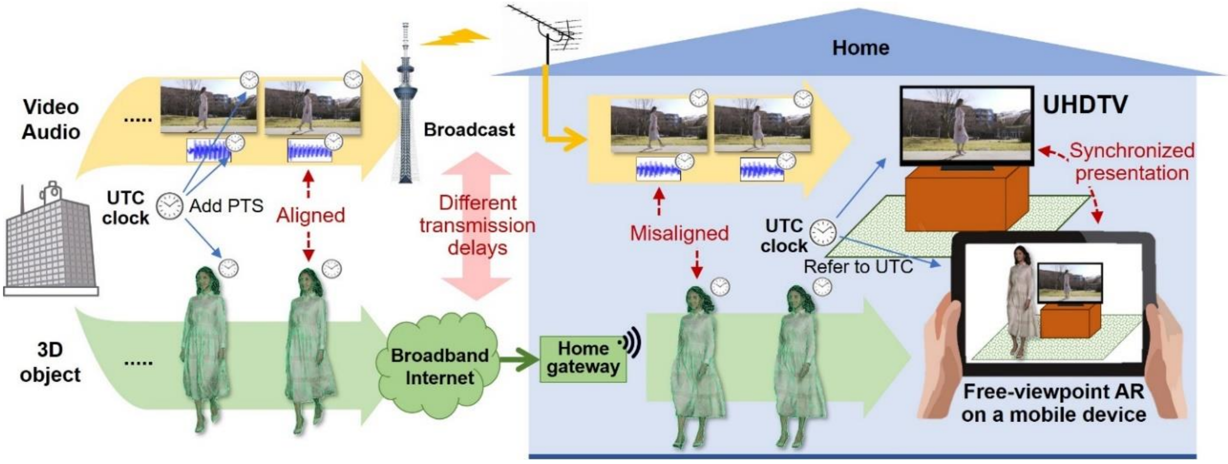
NHK developed an augmented reality system called “Augmented TV”. A viewer can see objects and characters in a TV come outside of the TV when holding a tablet up and looking through it. 2D images in the TV and 3D data in the tablet are synchronized, and 3D data is used to render images from the viewpoint calculated from the captured image and angular sensors in the tablet.

4.3.2.1.2 Free-viewpoint AR presentation synchronized with a TV programme

NHK developed a prototype for a content and distribution system based on the concept of integrated TV and AR (see § 3.5). Figure 16 shows a distribution model of a TV programme synchronized with AR content [12]. The TV programme and AR content share a common timeline and storyline. While TV video and audio are delivered over normal TV broadcasting, 3D objects linked to the TV programme are delivered in real-time through broadband Internet and rendered with free-viewpoint according to the manipulation of the AR device by the viewer. To enable synchronized presentation of AR content with a TV programme, presentation timestamp (PTS) based on Coordinated Universal Time (UTC) may be used, which is supported by MPEG media transport (MMT).

FIGURE 16

Distribution model of synchronized TV programme and AR content



Prototype content has been created in a volumetric capture studio by simultaneous shooting of UHDTV and a volumetric object. Sequential volumetric data at a fixed frame rate are generated from the images captured by multiple cameras surrounding a performer. Wavefront object files (.obj) [13] are used to represent the volumetric data consisting of polygonal geometry and texture, as shown in Fig. 17. Figure 18 shows a reception scene where the performer in the TV programme displayed on a TV appears in front of the TV on the tablet. The user can see the performer on the tablet from any perspective by changing its position and direction.

FIGURE 17

Volumetric data

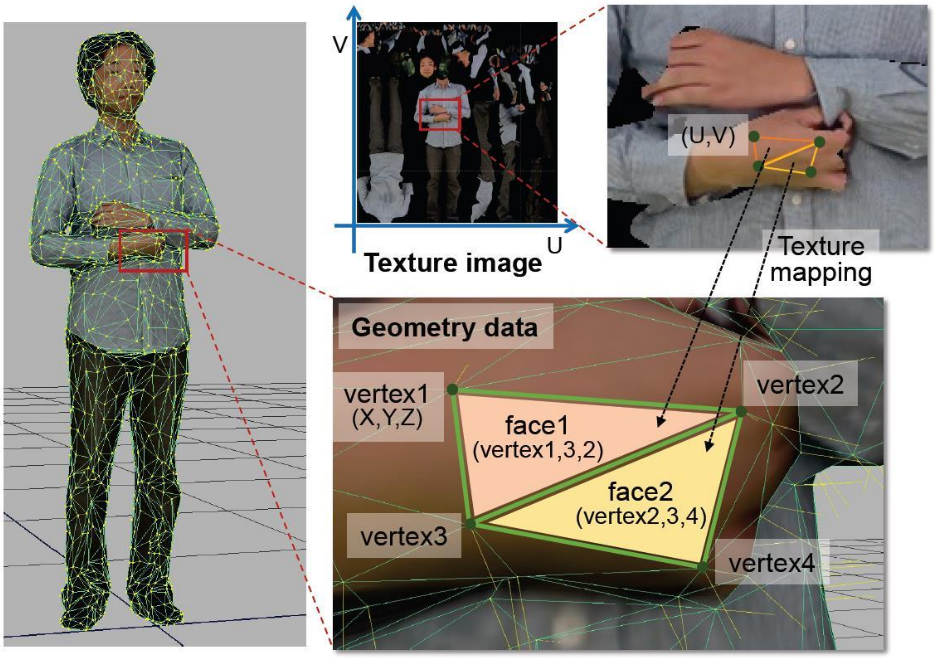
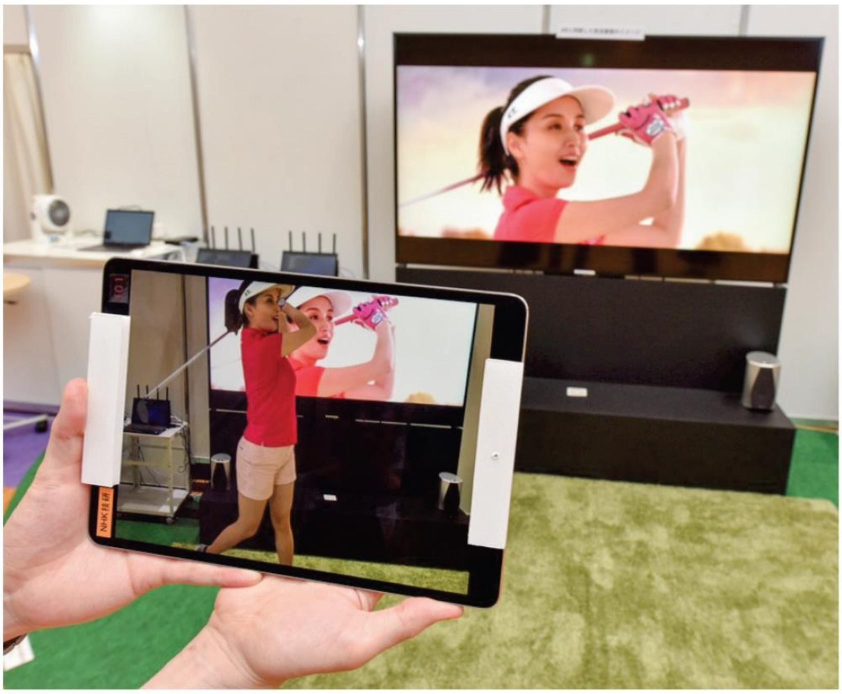


FIGURE 18

Prototype AR content synchronized with TV programme



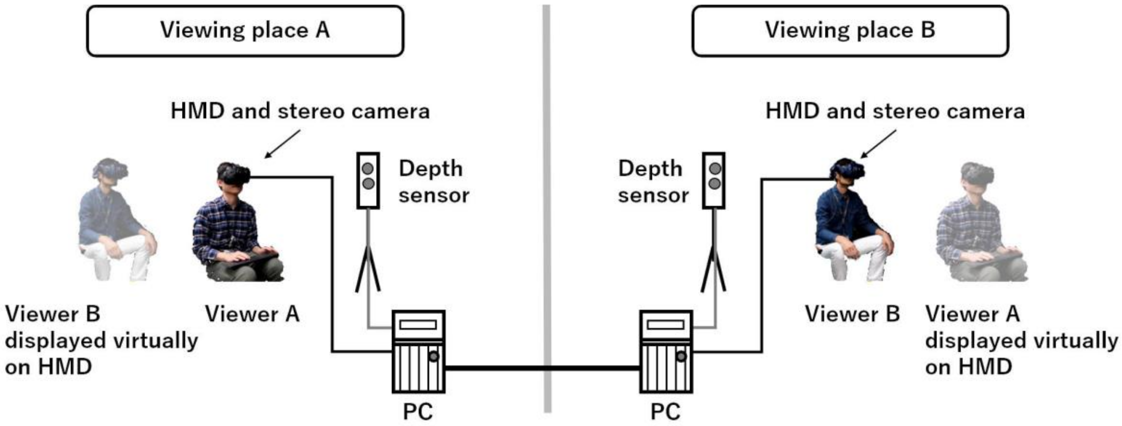
Future study for standardization will include a proposal for a common format for volumetric content and a synchronized delivery method.

4.3.2.1.3 Watching VR/AR content with remote viewers

NHK developed an experimental system for “virtual space sharing” (see § 3.5) in which three-dimensional objects of TV performers or family members and friends in different locations are combined and displayed in their actual size to the viewer through the HMD [14]. Figure 19 shows the outline of an experiment where users at two different locations simultaneously watch VR/AR content by wearing an HMD. The three-dimensional position of the HMD is captured by sensors. A microphone attached to the HMD captures the voice of the viewer and a depth-sensor captures the volumetric images of the viewer. Information elements captured from one viewing location are transmitted in real-time to the second viewing location, and vice versa. The virtual images of the person at the remote location are reproduced on the HMD along with his/her voice, thus enabling aural communications as if both viewers are watching the content together from the same location.

FIGURE 19

Experimental system for "virtual space sharing"



A stereo camera attached to the HMD captures the scene and depth around the viewer and enables reproduction of a mixture of real and virtual images on the HMD. In this experience, the real scene near the viewer (including the viewer's hands and body) is displayed to the user along with presentation of the VR content and virtual image of the remote user, as shown in Fig. 20.

FIGURE 20

Mixing real space and VR space according to distance

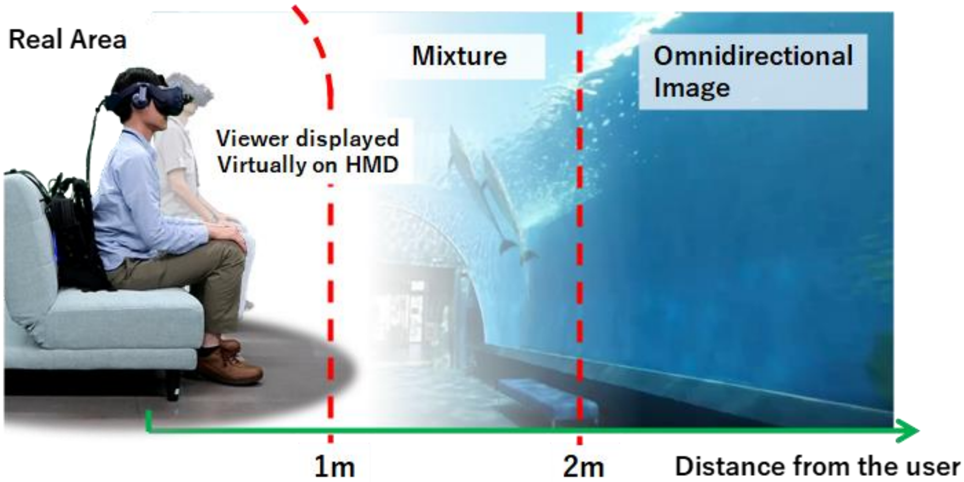


Figure 21 shows three prototype content types for watching on an HMD with remote viewers. The viewers can share the immersive visual experience in a 360° VR environment. "Aquarium" lets users enjoy the 360° space with family and friends. "Aerial TV" shows a 2D TV frame floating in the air with an arbitrary image size. "AR instructor" is a life-sized volumetric AR element, where a performer is displayed with his/her feet on the ground.

FIGURE 21

Images of prototype contents

(a) Aquarium (b) Aerial TV (c) AR instructor

4.3.2.1.4 Haptic interfaces

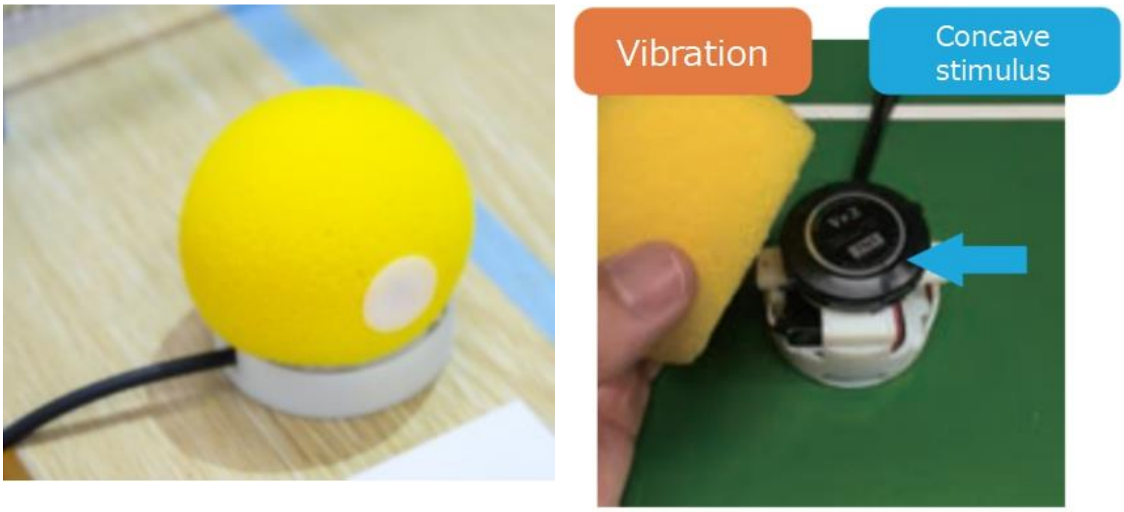
A variety of haptic devices can be envisioned depending on the genre and content of a video program. Haptic devices for which prototypes have been developed by NHK thus far are shown in Figs 22 and 23.

A ball-type haptic device envisioned for use mainly in sports programmes is shown in Fig. 22 [15]. This device is equipped with a vibrator and two servomotors, as shown in Fig. 22(b). It can convey the magnitude of an impact through vibration and the direction of a ball's movement by physically pulling in the surface of the device. This ball-type haptic device enables the viewer to experience the impact applied to a ball or athlete and their movements as haptic stimuli.

A cube-type haptic device that can be used for diverse genres of content in addition to sports is shown in Fig. 23 [16]. This device is equipped with a vibrator on each internal face of the cube, as shown on the left, right, top, and bottom face in Fig. 23(b), so that each of the four faces can be vibrated independently. This cube-type haptic device can richly convey the three-dimensional position and movement of the subject within the represented space depicted in the video. This device is also effective for showing subjects' salient motions to viewers of educational and animation programs. These haptic devices are expected to serve as interfaces for enhancing viewer immersion in broadcast content.

FIGURE 22

Ball-type haptic device



(a) External appearance (b) Internal structure

FIGURE 23

Cube-type haptic device

(a) External appearance (b) Internal structure

4.3.2.2 Nippon TV

Nippon TV developed and provided two augmented reality applications for smartphones from 2011 through 2013. These applications provided real-time on-screen movies of celebrities or 3D CG characters superimposed over images captured through smartphone cameras.

Technical specifications:

- Production format for movies
  - Picture aspect ratio: 16:9
  - Pixel count (horizontal × vertical): 1 440 × 1 080
  - Frame frequency (Hz): 60/1.001
  - Scan mode: interlace
- Presentation format
  - Picture aspect ratio: 2:3
  - Pixel count (horizontal × vertical): 320 × 480
  - Frame frequency (Hz): 15
  - Scan mode: progressive
  - File format: PNG
- Production format for 3D CG
  - Polygons: 7 000 – 10 000

Nippon TV has also developed computer graphics content that uses Augmented Reality technology referred as "Nippon TV Mixed Reality" system that enables multiple viewers in a room to wear transparent smart glasses that display various data and computer graphics outside of the television screen. Popping right out of the screen, characters and programme guests move and speak realistically in front of viewer's very eyes. During live sports broadcasts, viewers can bring athletes and useful data straight into their living room. Music show fans can also take the fun to a whole new dimension by "welcoming" the artists into their home to perform.

https://www.facebook.com/ntvmr/

4.3.2.3 TV Asahi

TV Asahi provides AR attractions at public events. In conventional technology, image recognition has been used as an AR marker. However, we have managed to get away with the complicated process of image recognition by incorporating the marker information within visual light.

TV Asahi is also investigating Augmented Reality head-mounted displays. Some applications using Microsoft HoloLens have currently been developed to study the application possibilities of adding information to television viewing.

4.3.3 Feedback of VR/AR trials

4.3.3.1 Feedback from producers and users

– Users experienced a high sense of presence (immersion) and overall positive reaction to content in the trials. This was associated with a reported positive experience from being able to freely choose where they directed their interaction with the content. However, some voiced a desire for higher image quality and image resolution. In addition, it is necessary to study and address the effects of VR/AR viewing on motion sickness as well as younger viewers.

– Production: In terms of lessening motion sickness and improving immersion, it is necessary to consider changes to the production process. This include modification in camera placement and camera resolution. Since it may become necessary for HMDs to have an age limit, it will, in turn, become necessary to develop alternative immersive experiences and content that even children can enjoy. The distance between the cameras and the subject is a critical variable that influences the production success. More study and control of the impact of this variable will help improve the experience and consistency of effective video production. Currently, super-wide-angle lenses are used during video capture. This has required a high resolution content capture to facilitate angle changes during zooming.

– Receiver: Resources are insufficient for using current receivers to develop and display 360° VR video internally. Both CPUs and memory need to be considerably upgraded. The level of specification needed is content dependent. Improved receivers are needed that enable separation of the monitor and processing parts for continued expansion.

4.3.3.2 Applicability to program genres

Applications of VR/AR are being considered that take advantage of the respective strengths of news, drama, sports, documentaries, and music. Each of these requires selection of an appropriate camera and recording system. Example relevant experiences benefitting from targeted selection include:

– Disaster reporting, where it is necessary to be able to communicate to the viewer the current surrounding situation.

– Multiple perspectives, such as to allow for multiple angles during live broadcasts of sports and entertainers and for participating in programs virtually.

– Sports content when the distance from the camera is short. This close distance provides a sense of presence but means small changes in resolution make it harder to understand what is going on. This has limited usage to less detailed content and scenes.

– A form of a second screen for hybrid broadcasting and online distribution.

4.3.3.3 Is "live" important?

– This depends on the service content. Nonetheless, content is being captured and developed that takes full advantage of the unique characteristics of "live" to provide a high sense of presence.

4.3.3.4 How might VR technologies impact storytelling?

– During live VR experiences, within a production a user is able to select their individual viewpoint. They are also often able to magnify the content in a particular location. These experiences are in contrast to conventional viewing methods where content creators and producers could rely on having control over viewpoint changes. Historically these perspective modifications happened during identified optimal moments in the content. Allowing user flexibility requires notable changes to standard practices for content recording and production.

4.3.3.5 Future activities

– VR/AR are already gaining importance for events, and it can be expected that the speed of their popularization will increase as they catch the attention of large companies.

– Current AR/VR systems impose notable inconvenience and isolation to the user. This may inspire marked changes in the design of HMDs.

– Main stream adoption requires improved development and cost savings in high-performance receivers.

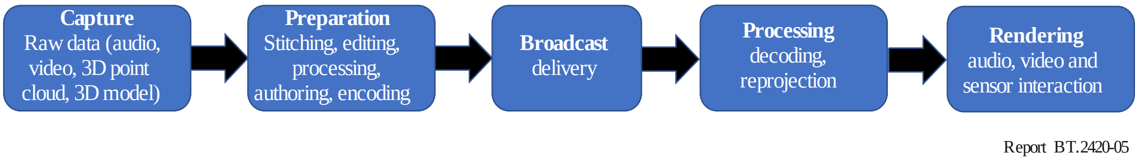
5 Challenges

5.1 Possibilities of AISM

The new possibilities of advanced immersive media created new challenges. Solving these challenges is necessary to enable a fully immersive experience over a long period of time. Figure 24 (DVB Report Virtual reality – prospects for DVB delivery, Nov. 2016) depicts the processing stages of production, broadcast delivery, and consumption of AISM programme material. Each stage brings its own challenges.

FIGURE 24

Processing flow of AISM programme material in a broadcast scenario



5.2 Production challenges

5.2.1 Format for programme exchange of AISM programme material

Recommendation ITU-R BT.2123 provides video parameter values for 360° images in 3DoF applications such as projection mapping type, image resolution, frame frequency, and colorimetry by extending the parameter values for UHDTV and HDR-TV [17]. A standard format for volumetric content is yet to be developed.

Recommendation ITU-R BS.2051 provides sound system parameters for advanced sound systems to support channel-based, object-based, or scene-based input signals or their combination with metadata [18]. The structure of a metadata model that describes the format and content of audio files, called the Audio Definition Model (ADM), is specified in Recommendation ITU-R BS.2076 [19]. The Broadcast Wave 64Bit (BW64) audio file format specified in Recommendation ITU-R BS.2088 can carry multichannel files and metadata [20]. The advanced sound systems supported by ADM and BW64 are considered the most prominent immersive audio solutions in VR applications (e.g. [6]).

However, there may be parameters specific to advanced immersive audio content that have not been addressed yet. It should be investigated how the existing specifications can be used or potentially be extended to become the programme production format for linear narrative (and maybe non-linear/interactive) advanced immersive audio content.

5.2.2 Evaluation of quality of AISM experience for broadcast applications

A recommendation on how to evaluate the quality of the AISM experience would be beneficial for the production and distribution of AISM content. For the evaluation of video coding quality of omnidirectional media, a few subjective methods have been proposed (e.g. [21, 22]). Because advanced immersive systems create multisensory experiences through the combination of audio, video, interactivity, and haptics, the QoE of AISM systems might require new QoE assessment methods. Such methods may be based on existing psychophysical methods and possibly also on psychophysiological assessment methods (e.g. [23]).

5.2.3 Production guidelines

VR production tends to be more complex compared to conventional programme production for a number of technical and aesthetic reasons: In a 360° video recording, there is nowhere to hide the equipment and production crew from the viewer. Creative solutions during production (e.g. hide crew behind props) and/or postproduction (e.g. stitching of two 180° shots into one 360° shot) are necessary. Further, to prevent sensory sickness (see § 5.4.2), quick changes in the video, often used as a cinematic storytelling technique (fast camera movements, fast changes of vantage points, scene switches), can only be used with care. The desire to tell a story under these constraints makes it necessary to find alternative techniques to direct the viewer's attention. Some broadcasters have reported on their learning experience in producing advanced immersive media (e.g. [24, 25, 26]). It has been said that VR storytelling has more in common with stage plays than with cinematic storytelling. Jaunt published a best practices guide for VR recording addressing many of these challenges [27]. In technical terms, Sky VR has published their current production guidelines, provisionally specifying their VR format[2] [28]. The VR guidelines produced by VR Industry Forum, aiming at best practices for VR services, addresses production guidelines for VR audio and video content.

5.2.4 Projection mappings

There are various types of ways to squeeze a 360° video image into a format applicable for current video encoders. The most prominent projection map is the ERP map, but there are various other methods (e.g. pyramid mapping, cube mapping). It is unlikely that consumer devices (e.g. HMDs) can support every projection mapping concept which creates the risk that content has to be produced or transcoded for different distribution channels and consumer devices.

5.2.5 Production workflow and authoring tools

For professional content creation, an automated production workflow is necessary for a timely delivery of high-quality content within budget. A studio workflow for linear narrative VR content may consist of stages such as footage ingestion, conversion, stitching, asset creation, audio production, tracking, layout, rendering, reviewing, delivering, and archival. While this workflow may be similar to a HDR video production, the amount of data that is being processed is usually significantly larger. For a non-linear narrative VR content production (i.e. where the content unfolds based on some user interaction), the VR workflow will include concepts and tools know from game development. VR and AR content production also benefits from new authoring tools tailored towards

---

[2] Video: 2-4K resolution, ITU-T H.264|MPEG-4 AVC, 25-50 FPS, 20-60 Mbps bitrate; audio: stereo or scene-based audio/ambisonics.

AISM content; for instance, sound design authoring tools to pan and visualize sound sources while watching the video content via HMDs or in 360˚ video player. It is unclear to which extent the AISM content production workflow can be integrated into the broadcast production workflow and infrastructure. Technical standards for interoperability that simplifies content exchange of raw and mastered VR content could help.

5.3 Delivery challenges

5.3.1 High transmission rates

The delivery of AISM content is the most significant challenge for current distribution systems. The transmission rates for VR content far exceed the data rates of current high-quality broadcast programmes. This stresses the network throughput and the processing capabilities of many consumer devices. Data buffering is often necessary to guarantee a consistent immersive experience.

To deliver high-quality VR content using unicast streaming, various methods are under development. These methods are based on the dynamic switching of view-dependent video streams that are compressed so that the active field of view receives high-quality video and the periphery receives reduced quality video (see e.g. [29, 30]). For preventing the perception of low-quality video content during head rotation, the seamless switch from one viewport video stream to another is crucial. For broadcasting of AISM content, those viewport-based solutions may not be feasible and the entire video feed is broadcasted at once. A proof of concept for the broadcasting of AISM content was demonstrated in 2016 at the IBC: using an ASTRA satellite connection as a distribution path the transmission of a 10K × 2K panoramic video signal to multiple devices was showcased (https://www.vrfocus.com/2016/09/ses-and-fraunhofer-hhi-to-bring-vr-via-satellite-to-ibc).

5.3.2 VR distribution format

There is a standardized distribution format to deliver high-quality immersive linear narrative AISM content. MPEG has standardized the first version of Omnidirectional Media Format (OMAF) as ISO/IEC 23090-2 which defines not only a media format of VR/360° content but also IP-based delivery methods of OMAF content using MMT and MPEG-DASH (see more in § 6). Recommendation ITU-R BT.2133 provides guidance on using ISO/IEC 23090-2 to transport AISM content in IP-based broadcasting systems [31].

5.4 Consumption challenges

5.4.1 General

Many of the technical problems during production and transmission are driven by the needs to deliver high-quality content on the consumer side that (in the best case) exceeds the requirements of current TV broadcasting. For instance, many experts suggest that a minimum of 60 frames per second and bi-ocular 4K video images are necessary to enable enjoyable high-quality VR experiences over HMDs (see also Sky's production guidelines [28]). The acceptable motion-to-sound and motion-to-photon latencies are other important technical parameters that contribute to both the realism and prevention of sensory sickness. Pointer devices or hand gesture trackers are also important for interactivity within the virtual world. These must be frictionless, intuitive, and reliable human-computer interface devices. On top of these demanding technical requirements, consumer studies show that there can be aesthetic concern with the design of current HMDs [32, 33]. As stated in [34] current VR/AR content is inconsistently rendered across HMD devices leading to different VR/AR experiences. New standards to guarantee a quality of experience may be needed.

5.4.2 Sensory sickness (kinetosis)

Currently created VR content has mostly a duration of less than 15 minutes. This can be linked to the desire to avoid triggering sensory sickness (i.e. kinetosis) that may occur when using HMDs for a longer duration, especially with high motion content. Despite years of research, preventing sensory sickness is a hard problem for VR consumption. It is known to occur as a function of several technical visual rendering parameters (motion-to-photon latency, frames per second, flicker of the displayed view, display width) and human aspects (duration of exposure, personal sensitivity, motion control, general health, genetic background, gender, age, mood, anxiety, postural stability). A detailed discussion on sensory sickness can be found in [2, section 10].

5.4.3 Content presentation

Compared to conventional television and cinema content, there is currently only a small amount of professional VR content available. Consequently, on-demand VR services aim to enhance their content offerings by adding 2D content that can be watched in VR. For instance, Netflix, Hulu and HBO are offering VR environments in which the user is placed in a virtual home cinema where the streamed 2D content is presented on a large virtual 2D-rectilinear. The content is adapted to the user's head motion. The audio content over headphones can be the original stereo mix or the binauralization of the original spatial sound mix.

On the flip side, content creators may want to repurpose VR content and make it available on traditional television. This scenario requires a workflow where a view of the VR video is rendered to traditional TV formats prior to broadcasting and presented with the correctly spatially aligned immersive audio. The German broadcaster rbb tested such solution for regular SmartTVs (https://www.fokus.fraunhofer.de/en/fame/news/FAME_BandCamp).

5.4.4 VR consumer platforms

A recent user experience report of 10 popular VR applications was published by VR testing company Fishbowl VR [35] and concluded that many VR services are not capable to deliver high-quality VR experiences. Users were asked to test and rate VR applications over HMDs in the categories UI, content variety and quality, picture quality, virtual environment, and loading times. Additionally, they were asked to rate their likelihood of recommending the app to a friend. The testers found immersive video to be enjoyable in powerful and unique new ways, however current VR delivery platforms still need improvements: Only for two tested apps, users consistently said they have the quality to eventually displace regular TV, laptop or mobile viewing. These two apps were offering both original 360° content but also traditional 2D programme material via a large 2D-rectilinear screen in the virtual world. Technical issues such as poor video quality and the lack of interactivity as well as social engagement degraded the overall VR experience. Users mostly desire a way to participate in VR video experiences together with friends and family. Content buffering, which occurs frequently in the majority of applications, led to very poor user experiences.

6 Work of ITU-T on Virtual Reality

6.1 ITU-T Study Group 16

In June 2016, ITU-T Study Group 16 established a new Question 8/16 regarding Immersive Live Experience. In January 2017, ITU-T SG 16 hosted its second workshop on Immersive Live Experience (ILE). The objective was for participants to exchange information related to immersive services and technologies between several organizations and to identify standardization gaps. The workshop featured various presentations on VR, AR, and related technologies:

http://www.itu.int/en/ITU-T/studygroups/2017-2020/16/Pages/ws/201701_ILE.aspx

6.2 ITU-T Study Group 12

At the ITU-T Study Group 12 meeting in January 2017, new work items “QoE for Virtual Reality (G.QoE-VR)” and “Subjective test methodologies for 360 degree video on HMD (G.VR-360)” were created under Question 13. These work items are intended to lead to several recommendations regarding QoE factors, QoE/QoS requirements, subjective test methodology, and objective quality estimation models for virtual reality (VR) services [36].

According to the baseline text for G.QoE-VR, the scope of the new Recommendation is as follows:

Virtual Reality (VR) is a new type of media different from the traditional video and audio media. It generates realistic images, sounds and other sensations that replicate a real environment, and simulates a user's physical presence in this environment, by enabling the user to interact with this space and any objects depicted therein using specialized display screens or projectors and other devices. The multi-sensory experiences, which can include sight, touch, hearing, and, less commonly, smell, are well coordinated and synchronized through the user’s interaction and feedback. A person using virtual reality equipment is typically able to “look around” the artificial world, move about in it and interact with features or items that are depicted on a screen or in goggles as in the real world.

In order to understand whether QoE or user-perceived performance of the VR service is good or not, benchmarking is critical. This allows measurement of user-perceived performance or QoE in that environment. Compared with traditional video and audio, the multi-sensory experience in VR imposes a new set of requirements to QoE assessment. The challenge is to characterize VR’s real-life immersive video, spatial-audio, and interactivity. Before benchmarking the QoE, it is important to address the requirements and basic factors assessing the VR quality for different VR services.

This draft Recommendation identifies different VR services and their respective requirements for Quality of Experience (QoE). This document also summarizes the key factors affecting user-perceived experience of a VR service, which can help to identify the methodologies for assessing the VR quality.

The scope of this Recommendation includes:

– VR service categories.

– QoE requirements for VR services.

– Categorization of influence factors.

7 Activities of other SDOs and VR groups

7.1 Activities of other SDOs

Many SDOs and industry groups have already started working or exploring subareas of AISM systems. The following list gives a best-effort overview of ongoing activities:

MPEG: The Motion Picture Expert Group (MPEG) is developing video and audio codecs, and transport and systems aspects with regards to VR and AR applications.

An important activity for AISM content exchange and delivery is the Omnidirectional Media Format (OMAF). OMAF is the first standard from MPEG providing formats for the support of immersive media. It specifies the application format for coding, storage, delivery, and rendering of omnidirectional images and video and the associated audio. The Final Draft International Standard (FDIS) of OMAF was produced in October 2017 as ISO/IEC 23090-2 (Part 2 of MPEG-I).

MPEG-H 3D Audio (ISO/IEC 23008-3:2017) is a next generation audio codec featuring highly efficient compression of channel-based audio, object-based, and scene-based immersive audio. For AISM applications, it specifies a 3DoF decoder-side sensor interface and normative specification for diegetic and non-diegetic audio processing.

http://mpeg.chiariglione.org/standards/mpeg-h/3d-audio

In Q3 2016, MPEG conducted an informal survey to better understand the needs for standardization in support for VR applications and services. They received 185 responses and summarized their results and conclusions [6]. Based on this survey, MPEG concluded to launch a new project on immersive media (MPEG-I). A first set of specifications that defines up to 3DoF 360° VR was finalized as OMAF to support market launches of products and services in 2018. Based on a common belief that major market launch of VR 360 services will happen in 2020, a next set of specifications including a richer feature set (such as 6DoF) may be ready in 2019.

http://mpeg.chiariglione.org/tags/virtual-reality

The Joint Video Exploration Team of MPEG and ITU (JVET) is working on the future H.266 video codec for which they (among other things) study the effect on compression when different warping methods are applied to the input 360° video before compression. On this basis, JVET has defined common test conditions, test sequence formats, and evaluation criteria for such content (http://www.content-technology.com/standards/?p=740).

JPEG: The Joint Photographic Experts Group (JPEG) is developing JPEG XT, an image format which features coding of omnidirectional (360-degrees) images; JPEG XS, a low latency compression formats for VR videos; and JPEG PLENO, a video format for point cloud, light field, and holographic images.

https://jpeg.org

DVB: In 2016 DVB carried out a study mission to determine the likelihood VR video will be commercially successful, and to find out how DVB can be involved. The study mission produced a detailed report [2]. An executive summary of this report is freely available [5].

The DVB study mission concluded that at least in the near term for broadcast use cases, untethered devices supporting 3DoF (such as slide-on HMDs) are more likely to be commercially successful than tethered devices. The dominant success factors of these VR devices and services are quality of experience (QoE), lack of sensory sickness, comfort and ease of use, cost and availability of equipment, cost and availability of content, and content desirability.

Further, it was concluded that DVB should cooperate with standards bodies working in VR, as members will need to adopt common specifications for delivery of VR content. Requirements are needed for the minimum technical quality of VR video and audio. Requirements should be completed by mid-2018. A questionnaire on VR broadcasting services has been circulated to further inform the process of defining commercial requirements. DVB is informing various standardization groups about their findings and their planned activities (e.g. MPEG, ITU-R, ITU-T). The DVB study mission continues to address topics such as AR, MR, and 6DoF VR.

https://www.dvb.org/groups/CM-VR

3GPP: 3GPP's subgroup SA4 conducted a feasibility study on virtual reality media services over 3GPP. The technical report can be accessed online [3]. Further, the subgroup SA1 specifies service requirements for the 5G system which includes aspects related to support various VR and AR use cases.

SMPTE: Specifications on production for VR content are in the scope of SMPTE. The 2016 SMPTE conference had presentations in this area (e.g. [34], [37]).

http://www.tvtechnology.com/events/0025/smpte-how-can-you-ensure-an-effective-vrar-experience/279732

W3C: The W3C community is working toward standardizing WebVR, a JavaScript-based API that provides access to VR devices, sensors, and head- mounted displays through web browsers. At the time of writing this report, it was unclear if and when the W3C standards track adopts WebVR.

https://w3c.github.io/webvr

https://www.w3.org/2016/06/vr-workshop

IEEE: Under the umbrella of the Virtual Reality and Augmented Reality Working Group (VRAR) the IEEE launched standardization activities in the area of taxonomy and definitions for VR, AR devices; quality metrics for immersive video; and VR file and streaming formats. The IEEE is also organizing the IEEE VR conference series, one of the oldest academic consortia dedicated to the study of virtual reality from a science and engineering perspective.

https://standards.ieee.org/develop/wg/VRAR.html

IEEE P2048.2

IEEE P2048.2™, Standard for Virtual Reality and Augmented Reality: Immersive Video Taxonomy and Quality Metrics. This standard specifies the taxonomy and quality metrics for immersive video.

End-to-end interoperability

IEEE P2048.3™, Standard for Virtual Reality and Augmented Reality: Immersive Video File and Stream Formats. This standard specifies the formats of immersive video files and streams, and the functions and interactions enabled by the formats.

CTA: In November 2016, the Consumer Technology Association (CTA) released a survey report [33] entitled "Augmented Reality and Virtual Reality: Consumer Sentiments". The study aimed to understand consumer's awareness and opinions regarding AR and VR technologies and gathered consumer feedback about envisioned VR/AR use cases. A summary of this report was presented at the CES 2017 and provided three recommendations. First, consumers need more education about the difference of VR vs. AR; VR device choices; and content delivery options. Second, to expedite consumer excitement and to avoid stereotyping VR technology, more diverse VR demos need to be produced and presented to the mass market. Third, to correct miscues or misuse of devices, the industry should encourage customer feedback of VR solutions.

Others: The **IETF** (Internet Engineering Task Force) explores mechanisms to carry high bandwidth media such as VR with low latency over the Internet, **Qualinet** has a task force for VR and immersive experiences (IMEx), and **Cable Labs** studies the requirement for support of VR in cable networks. **ARIB**, the Association of Radio Industries and Businesses published anticipated quality requirements and transmission rates for VR application over 5G networks [38].

7.2 Activities of VR industry groups

Multiple industry groups have been formed to promote the development and growth of the VR/AR industry.

VRIF: The VR Industry Forum (VRIF) is a not-for-profit company with the purpose to further the widespread availability of high-quality audio-visual VR experiences, for the benefit of consumers. Founded on January 5th, 2017, the goals of VRIF include:

– Advocating voluntary industry consensus around common technical standards for the end-to-end VR ecosystem, from creation to delivery and consumption.
– Advocating the creation and adoption of interoperable standards; promoting the use of common profiles across the industry, and promoting and demonstrating interoperability.
– Developing voluntary guidelines that describe best practices, to ensure high-quality VR experiences.

Describing and promoting the use of VR services and applications.

VRIF has also established a lexicon of VR terminology to encourage common usage of term and avoid misuse of terms causing confusion.

http://www.vr-if.org

VR Society: The VR Society was officially launched on July 13th 2016. Its purpose is to accelerate the transformation, innovation, and profitability of the Virtual Reality Content, Distribution and Technology Business. Further, the society promotes and fosters a marketplace of ideas based on thought leadership, sharing of best practices, marketing, consumer research, industry analytics and advocacy. Together with the Advanced Image Society, the VR Society bestows the Lumiere Awards (the 2017 edition includes expansive categories for VR content).

http://thevrsociety.com

The immersive Technology Alliance: Founded in 2009, the Immersive Technology Alliance contributes to the viability, advancement, and adoption of virtual reality, augmented reality, mixed reality, stereoscopic 3D, and related future innovations. The Alliance facilities collaborative problem solving and accelerated opportunity development at an industry-wide level.

https://ita3d.com

VRARA: The VR/AR Association (VRARA) is a global industry association founded in 2015 that aims to offer a connected local and global community of members through its initiatives. The VRARA is focused on creating a member community of VR/AR solution providers, content creators, and customers. The objective is to accelerate networking and sharing of knowledge through case studies. There are different industry committees such as VR Stories & Audience, Mobile VR & 360 Video, Entertainment, and Advertisement that strive to establish best practices, guidelines, and call-to-actions (e.g. recommendations for standards) in their areas of expertise.

http://www.thevrara.com

GVRA: The Global Virtual Reality Association (GVRA) is a non-profit organization of international headset manufacturers. It was founded on December 7th 2016 and promotes responsible development and adoption of VR globally with best practices, dialogue across stakeholders, and research.

https://www.gvra.com

Khronos: On December 6th 2016 Khronos, an open consortium of hardware and software companies, started an initiative to define a cross-vendor, royalty-free, open standard for access to modern virtual reality (VR) devices. Key components of the new standard will include APIs for tracking of headsets, controllers and other objects, and for easily integrating devices into a VR runtime. This will enable applications to be portable across VR systems, enhancing the end-user experience, and driving more choice of content to spur further growth in the VR market.

https://www.khronos.org/vr

SVA: The Streaming Video Alliance (SVA) is a consortium of organizations spanning the streaming video value chain. In November 2016, it announced a study group for VR and 360-degree video. In this group, participating members and selected outside parties seek to understand and document the

VR industry in an effort to identify opportunities for developing best practices. The group's objective is to:

– Understand the VR market and how it is impacting traditional video experiences.

– Capture the state of VR technologies, the players, and use-cases.

– Catalogue existing standards efforts.

A report on its findings is expected to be issued in Q2 2017.

https://www.streamingvideoalliance.org/technical-work/virtual-reality360-degree-video

Digital Senses Alliance: The Digital Senses Alliance is an Industry Connections program about to be formed by the IEEE Digital Senses Initiative which aims for cross-industry and cross-disciplinary collaborations to identify gaps in technologies and standards in the area of VR, AR, and Human Augmentation (HA). Further, it plans to provide training facilities and learning resources for how to create VR and AR content.

http://digitalsenses.ieee.org

OSVR: The Open Source VR group (OSVR) promotes a universal open source VR ecosystem for technologies across different brands and companies to avoid hardware fragmentation of HMDs and controllers and other challenges in the industry.

http://www.osvr.org

DASH-IF: The Industry Forum for dynamic adaptive streaming over HTTP (DASH-IF) hosted a workshop on "Streaming Virtual Reality with DASH" in May 2016.

http://dashif.org

AES: The Audio Engineering Society has just formed a new technical group to advance the science and application of Audio for New Realities (AR/VR/MR). The initial objectives will include collating the state of the art in audio for new realities across recording, composition, sound design, spatial audio, environmental analysis and auditory scene synthesis, in order to develop technical workflows that are practical and relevant to the industry and creative practitioners in the field. The AES also organized one of the first conferences purely on Audio for VR/AR:

http://www.aes.org/conferences/2016/avar.

Bibliography

[1] P. Lelyveld. (2015, July) *Virtual Reality Primer with an emphasis on camera-captured VR*. [Online]. Available at: http://www.etcenter.org/wp-content/uploads/2015/07/ETC-VR-Primer-July-2015o.pdf

[2] DVB. (2016, October) *DVB Virtual Reality – prospects for DVB delivery*, report of the DVB CM study mission on virtual reality study.

[3] 3GPP SA4. (2018, January) *Virtual Reality (VR) media services over 3GPP*. [Online]. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3053

[4] H. Nagata, D. Mikami, H. Miyashita, K. Wakayama, and H. Takada, "Virtual reality technologies in telecommunication services", Journal of Information Processing, vol. 25, pp. 142-152, 2017.

[5] DVB. (2016, November) *Executive summary DVB study mission on virtual reality*. [Online]. Available at: https://www.dvb.org/resources/public/whitepapers/dvb_vr_study_mission_report_summary.pdf

[6] MPEG. (2016, October) *N16542* – summary of survey on virtual reality. Chengdu, China. [Online]. Available at: http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/W16542%20MPEG%20VR%20Questionnaire%20Results%20Summary.pdf

[7] H. Kawakita, K. Yoshino, D. Koide, K. Hisatomi, Y. Kawamura, and K. Imamura, "AR/VR for various viewing styles in the future broadcasting", IBC2019.

[8] F. Williams. (2017, February) *OxSight uses augmented reality to aid the visually impaired*. [Online]. Available at: https://techcrunch.com/2017/02/16/oxsight-uses-augmented-reality-to-aide-the-visually-impaired

[9] Outlyer Technologies. (2017) *360 / VR test ad campaign*. [Online]. Available at: http://offers.advrtas.com/360-vr-ad-case-study-mobile-app

[10] H. Kawakita, T. Niida and K. Hisatomi, "Development of a 360 Video Player for Head-Mounted Display to Enable Comparison Between Before and After Incident", Forum on Information Technology 2020 (FIT2020), Part III, K-010, pp. 361-363. (2020).

[11] D. Koide, H. Kawakita, K. Yoshino, K. Ono, K. Hisatomi, "Development of High-Resolution Virtual Reality System by Projecting to Large Cylindrical Screen", IEEE International Conference on Consumer Electronics (ICCE), 1.16 AVS (4)-1, 2020.

[12] Y. Kawamura, T. Kusunoki, Y. Yamakami, H. Nagata and K. Imamura, "Toward tele-experience: Enhanced viewing experience by synchronized UHDTV and free-viewpoint AR", IBC2020.

[13] Wavefront Technologies, Appendix B1. Object Files (.obj), Advanced Visualizer Manual.

[14] K. Yoshino, H. Kawakita, T. Handa and K. Hisatomi,"Viewing Style of Augmented Reality/Virtual Reality Broadcast Contents while Sharing a Virtual Experience", 26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), Article 76, 2020.

[15] T. Handa, M. Azuma, T. Shimizu, S. Kondo, M. Fujiwara, Y. Makino and H. Shinoda: "Ball-type Haptic Interface to Present Impact Points with Vibrations for Televised Ball-based Sporting Event", IEEE World Haptics Conference (WHCs), TP1A.14, pp. 85-90, 2019.

[16] M. Azuma, T. Handa, T. Shimizu, and S. Kondo: "Development of Vibration Cube to Convey Information by Haptic Stimuli", Proceedings of the 24th International Display Workshops, Vol. 24, pp. 128-130, 2017.

[17] Recommendation ITU-R BT.2123-0 (01/2019) *Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting*. [Online]. Available at: https://www.itu.int/rec/R-REC-BT.2123-0-201901-I/en

[18] Recommendation ITU-R BS. 2051-2 (07/2018) *Advanced sound system for programme production*. [Online]. Available at: https://www.itu.int/rec/R-REC-BS.2051-2-201807-I/en

[19] Recommendation ITU-R BS.2076-2 (10/2019) *Audio Definition Model*. [Online]. Available at: https://www.itu.int/rec/R-REC-BS.2076-2-201910-I/en

[20] Recommendation ITU-R BS.2088-1 (10/2019) *Long-form file format for the international exchange of audio programme materials with metadata*. [Online]. Available at: https://www.itu.int/rec/R-REC-BS.2088-1-201910-I/en

[21] E. Upenik, M. Rerabek, and T. Ebrahimi, "A testbed for subjective evaluation of omnidirectional visual content", in 32nd Picture Coding Symposium, no. EPFL-CONF-221560, 2016. [Online]. Available at: https://infoscience.epfl.ch/record/221560/files/Testbed_for_omnidirectional_content_camready.pdf

[22] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes", in Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on. IEEE, 2015, pp. 31-36. [Online]. Available at: http://msw3.stanford.edu/~hlakshman/pdfs/2015_ISMAR.pdf

[23] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J. N. Antons, K. Y. Chan, N. Ramzan, and K. Brunnström, "Psychophysiology-based QoE assessment: A survey", IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 6-21, Feb 2017.

[24] J. Bloch. (2016, Oct) *5 secrets to making a virtual reality film*. [Online]. Available at: http://www.cbc.ca/news/canada/british-columbia/highway-of-tears-vr-doc-1.3800490

[25] M. Burns. (2017, March) *An exploration of VR and AR for sports: part i – the challenge facing immersive sports*. [Online]. Available at: http://www.svgeurope.org/blog/headlines/an-exploration-of-vr-and-ar-for-sports-part-i-the-challenge-facing-immersive-sports

[26] Streamshark.io. (2016, April) Ca*se study: Live streaming of the Australian Open in 360*˚. [Online]. Available at: https://streamshark.io/blog/aus-open-360

[27] Jaunt, Inc. (2017, January) *The cinematic VR field guide – a guide to best practices for shooting 360º*. [Online]. Available at: https://www.jauntvr.com/cdn/uploads/jaunt-vr-field-guide.pdf

[28] Sky. (2016, December) *Launch technical guidelines for 360 video content*. [Online]. Available at: https://corporate.sky.com/documents/about-sky/commissioning-and-ideas-submission/launch-technical-guidelines-360-video-content.pdf

[29] E. Kuzyakov. (2017, January) *End-to-end optimizations for dynamic streaming*. [Online]. Available at: https://code.facebook.com/posts/637561796428084/end-to-end-optimizations-for-dynamic-streaming

[30] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC- compliant tile-based streaming of panoramic video for virtual reality applications", in Proceedings of the 2016 ACM on Multimedia Conference, ser. MM '16. New York, NY, USA: ACM, 2016, pp. 601-605. [Online]. Available at: http://doi.acm.org/10.1145/2964284.2967292

[31] Recommendation ITU-R BT.2133-0 (10/2019) *Transport of advanced immersive audio visual content in IP-based broadcasting systems*. [Online]. Available at: https://www.itu.int/rec/R-REC-BT.2133-0-201910-I/en

[32] Perkins Coie LLP and Upload. (2016, September) *2016 augmented and virtual reality survey report*. [Online]. Available at: https://dpntax5jbd3l.cloudfront.net/images/content/1/5/v2/158662/2016-VR-AR-Survey.pdf

[33] Consumer Technology Association. (2016, October) *Augmented reality and virtual reality: Consumer sentiments*. [Online]. Available at: https://www.cta.tech/Research-Standards/Reports-Studies/Studies/2016/Augmented-Reality-and-Virtual-Reality-Consumer-Se.aspx

[34] P. Routhier, "Virtually perfect: Factors affecting the quality of a VR experience and the need for a VR content quality standard", in SMPTE 2016 Annual Technical Conference and Exhibition, Oct 2016, pp. 1-20. [Online]. Available at: http://www.sportsvideo.org/2016/10/28/smpte-2016-qa-digital-troublemakers-pierre-routhiers-13-rules-for-quality-vr

[35] Fishbowl VR. (2016, December) *The state of VR video – user experience report*. [Online]. Available at: https://www.fishbowlvr.com/resources/experience-reports/state-of-vr-video

[36] Recommendation ITU-T G.1035 (05/2020) *Influencing factors on quality of experience for virtual reality services*. [Online]. Available at: https://www.itu.int/rec/T-REC-G.1035-202005-P

[37] M. L. Champel and S. Lasserre, "The special challenges of offering high quality experience for VR video", in SMPTE 2016 Annual Technical Conference and Exhibition, Oct 2016, pp. 1-10.

[38] Association of Radio Industries and Businesses. (2014, October) *Mobile communications systems for 2020 and beyond*. [Online]. Available at: http://www.arib.or.jp/english/20bah-wp-100.pdf

[39] Adrian Freed and Andy Schmeder, "Features and Future of Open Sound Control version 1.1. for NIME", In Proc. NIME 2009, pp. 116-120.