



Launch Technical Guidelines For 360 Video Content

Contents

1	Overview	3
1.1	User Experience	3
1.2	Hardware	3
1.3	Content Creation Guidelines.....	4
1.4	Programme Compliance.....	4
2	SKY VR and Spherical 360 Programme Delivery Specification.....	6
2.1	Master Format.....	6
2.2	Master Format Picture Resolution.....	6
2.3	Frame Rates	6
2.4	Master File Format.....	6
2.5	Audio for VR	7
2.5.1	Master Audio Format.....	6
3.0	Interim Specification for Final Encode 3 rd Party Content.....	8
4	Appendix	9
5	Relevant Research and Standards Papers.....	9
6	VR Terminology.....	10
7	Version Control	11

1.0 Overview

This document will detail the user experience, hardware, applications for VR and Immersive content. It will provide a framework for approaches and best practice for content creation, and an outline of a technical standards framework.

This document details the Production requirements and provides Post Production Guidelines for Commissions, as well as the expectations of Acquisitions, Acquired content used within Commissions, Programme Compliance and the Delivery formats.

This document is intended to supplement the full Technical Delivery Requirements detailed in the DPP technical specifications.

All 360 and VR content must satisfy a Quality Control process specified by Sky. Failure to adhere to these guidelines could result in content being rejected or not qualifying for Content delivery.

The advancements in technology around 360 and VR will necessitate amendments. Please check with Sky to ensure you are using the most up-to-date version of this document.

1.1 User Experience

360 content displayed in a Head-mounted-Display (HMD) allows the viewer to engage with content in a unique, immersive manner. The viewer is separated from visual cues from the real environment and placed in a virtual environment or experience. This disconnect can, in some instances, lead to disorientation and unease: care must be taken to minimise these effects.

1.2 Hardware

This can be categorised as follows:

a) Mobile phone-based displays

May be either hand held or inserted/integrated into a Head-Mounted-Display (HMD). The phone must have a gyro and accelerometer and suitable software to display spherical content.

Audio is fed from the phone to earphones.

A Bluetooth hand controller may be integrated with the phone.

Examples include Oculus Gear VR, Google Cardboard, Google Daydream

b) Tethered HMD devices

Units currently rely on a separate processor to render content, thereby enabling higher quality output and greater functionality. The system comprises an HMD with orientation sensors and on- or off-board positional, spacial and gesture sensors. Hand-held controllers are generally integrated into the system. Audio is fed to earphones.

Current display resolutions are in the range of 1000x1000 to 2000x2000 per eye with refresh rates at up to 120fps

Examples include Oculus Rift, HTC Vive, Sony Playstation VR

Future developments indicate that a wireless link will replace the cable connection.

1.3 Content creation guidelines

It is essential that immersive content is comfortable to view and does not induce feelings of disorientation or unease in the viewer.

The following topics will be covered in the guidelines

- Field of view
- Motion
- Orientation
- Perceived viewer height
- Proximity of objects in the scene
- Duration of content
- 3D Stereo content

1.3.1 Field of view

Head-mounted Displays (HMD) vary in the field of view that they can reproduce, but generally have a 60° to 90° per eye range. For content designed to be viewed in a seated position, it is best to limit the action in the scene so that the viewer is only required to move their head through an arc of 90° to 120°; this will give an overall viewable arc of 180° to 210°, with significant action within a 90° range of head movement.

For content designed to be viewed in a standing position a slightly wider arc may be considered.

1.3.2 Motion

There are several considerations of the effect of motion within content.

1 - Image capture rate and motion capture fidelity.

Content should be shot or created in a manner that reduces motion blur or motion stepping to comfortable levels, as flickering images can lead to rapid fatigue or disorientation.

Image capture at high frame rates and designing content to work within the display capabilities will reduce discomfort. When capturing moderately fast-paced action such as dance, sport and vehicles a minimum capture frame rate of 50fps should be considered. If shooting at higher frame rates consideration of the target display(s) characteristics should be made. Current maximum display refresh rates are 120fps (Playstation VR)

2 - Perceived motion of the viewer in the experience

If the viewer is stationary and the scene moves independently from the viewer there is a perceived disconnect between the viewer's visual system and motion sensing system. This can lead to disorientation and in some cases nausea.

If motion is to be used as a creative intention then care should be taken. The following techniques can be useful: Suspension of disbelief by placing the viewer in a recognisable situation ie: inside a vehicle or a flying bird's eye view.

If the motion is at ground level, avoid:

- Accelerating too rapidly
- Turns that are not relevant to the action
- Yaw and pitch of the horizon
- Vertical motion - low frequency - as in replication of human walking PoV

If the action follows a relevant person or object this can reduce motion effects as the viewer again suspends disbelief.

1.3.3 Orientation

In 360 content, it is best practice to keep the horizon level at all times unless angled for a specific purpose. The viewer will always tend to match their head inclination to the perceived horizon - this can lead to discomfort and vertigo.

If the horizon oscillates this can prove to be very uncomfortable.

Panning and tilting the camera can also be very uncomfortable.

When transitioning between shots, care should be taken to orientate the action or intended direction of view of the subsequent shot to the outgoing shot.

1.3.4 Perceived Viewer eye height

Camera to scene height should always be relevant to the scene.

A small discrepancy in height can result in surprisingly off-putting perceptions of the relative sizes of subjects. The wide-angle distortion of some camera formats can exacerbate this, a low camera height can make subjects appear unusually tall - similarly a high camera height can make subjects appear unusually small.

1.3.5 Proximity of objects in the scene

Objects close to camera have a disproportionate influence on the viewer and can unnecessarily dominate the scene. This effect can be very disquieting to the viewer and can cause reaction in attempt to move away from the object. The wide-angle perspective distortion of some camera formats can also make objects approaching the camera appear to accelerate as they draw nearer, causing disquiet.

1.3.6 Duration of Content

Some research (see Appendix 4.3) has been conducted and has shown that content durations of up to 20 minutes are acceptable if the content is not challenging. Content with fast-paced action or motion will be more comfortable if shorter.

1.3.7 3D Stereo Content

3D stereo content can significantly enhance the viewing experience, creating a more natural perception of the environment. Care must be taken not to create too great a positive or negative disparity which might cause eye-strain or defensive reactions in the viewer.

1.4 Programme Compliance

All programme submission to SKY must comply with the Editorial Compliance Policy and Guidance for all SKY Entertainment Content.

A copy of this document can be found at:

<https://corporate.sky.com/documents/about-sky/commissioning-and-ideas-submission/editorial-compliance-pack---feb-2015.pdf>

2.0 SKY VR and Spherical 360 Programme Delivery Specification

Before delivery please consult the commissioner to detail the precise delivery requirements

2.1 Master Format

SKY requires the delivery of a Master VR Video file from which transcodes for the various end-user platforms can be made. In certain circumstances SKY may permit the Programme maker to deliver end-user transcoded files.

The Master Video presentation format is **Equirectangular Lat/Long image video**, 2:1 Aspect Ratio with the attributes set out below:

- No Zenith or Nadir Blind-spots are permitted.
- The video may be generated from multiple camera arrays or composited imagery.
- The image should display no apparent stitch lines and occluded or missing picture information.
- Frame motion should not display motion artefacts such as blur or step motion.

2.2 Master Format Picture resolution

Monoscopic: 4096 H x 2048 V or 3840 H x 1920 V

Stereoscopic: 4096 H x 4096 V or 3840 H x 3840 V

Stereoscopic format is: Over-and-Under frame packed, Left Eye above

2.3 Frame rates

The following frame rates are acceptable fps: 25, 30, 50, 60, 75, 90, 100, 120 fps, progressive
Other frame rates may be permitted by agreement.

2.4 Master Format File formats

Bit Depth	10-bit
Colour Sampling	4:2:2
Colour space	Rec.709 (gamut levels within the threshold defined by EBU R103)
Scan	Progressive Frame
Delivery format Option 1	MXF Programme Contribution (AMWA AS-11 X1 as per DPP Specification)
Delivery format Option 2	IMF Application 2e / JPEG 2000 minimum data rate 150Mb/s
Delivery format Option 3	DnxHR HQ
Delivery format Option 4	ProRes 422 HQ

In certain instances SKY may acquire Third-party content encoded for final delivery rather than as a Master Format file

2.5 Audio For VR

Care must be taken to ensure that the audio design for content enhances the experience.

Audio is primarily experienced with headphones or in-ear buds and must be mixed to comply with the EBU R128.

Spatial audio works very effectively and sound design should preferably reflect this.

There are a number of spacial recording microphones available on the market, however the use of these microphones alone, located at the camera position is not always effective, especially if dialogue is to be recorded.

Best practice is generally to record stems and remix in Post-Production.

2.5.1 Master Audio Format

The minimum requirement is for Stereo audio, preferably spatially encoded.

A 5.1 Audio mix may be supplied by agreement.

A B-format Ambisonics First Order WXYZ (*.amb) file may be submitted by agreement.

Delivery in discrete 8 track format as follows:

Tracks 1 and 2, Stereo Mix

Tracks 3 to 6, 5.1 Audio

Track configuration:

1 - Stereo Left

2 - Stereo Right

3 - Centre

4 - Low Frequency Effect

5 - Left Surround

6 - Right Surround

7 - No Audio

8 - No Audio

The orientation of B-format and 5.1 mixes should be Centre to picture reference zero

INTERIM SPECIFICATION – MAY BE SUBJECT TO IMMEDIATE CHANGE

3.0 Delivery Guidelines for Final Encoded 3rd Party Content

Video

Format HD

Resolution	1920 x 960
Encoding	H.264 (Constrained Baseline Profile 4)
Framerate	25 or 30 FPS
Bitrate	20Mbps
Throughput	20M
Distortion	Equirectangular
Zenith and Nadir	Blind Spot Top, Bottom, None
Encode Options	movflags faststart

Format 2K

Resolution	2048 x 1024
Encoding	H.264 (Baseline Profile 4.1 or 4.2 for 50fps)
Framerate	25 or 50 FPS
Bitrate	25 or 35Mbps
Throughput	35M
Distortion	Equirectangular
Zenith and Nadir	Blind Spot Top, Bottom, None
Encode Options	movflags faststart

Format 4K

Resolution	4096 x 2048 or 3840 x 1920
Encoding	H.264 (Baseline Profile 5 or 5.2 for 50fps)
Framerate	25 or 50 FPS
Bitrate	45 or 60Mbps
Throughput	60M
Distortion	Equirectangular
Zenith and Nadir	Blind Spot Top, Bottom, None
Encode Options	movflags faststart

Audio for all picture formats

1 – Stereo

Encoding	MPEG-4 AAC
Channels	Stereo
Sample Rate	48000 HZ
Bitrate	128 Kbp/s (kilobits per second)

2 – Spatial Audio – Ambisonics

Encoding	MPEG-4 AAC-LC, Layout 4.0
Channels	4 WXYZ
Sample Rate	48000 Hz
Bitrate	512 Kbps (kilobits per second)

4.0 Appendix

Set out below is an abstract from DPP File-based delivery specification:

4.1 File Formats

The programme must be delivered as one or more files that conform to the AMWA AS-11 X1 Specification ("MXF Program Contribution – DPP UHD").

Note: In summary, the key file-format properties of a file that conforms to AMWA AS-11 X1 are:

- MXF OP1a.
- AVC / H.264 coded video:
 - 10-bits per sample
 - LongGOP (inter coded) or I-Frame (intra coded).
 - Multi-channel Sound Tracks containing 48kHz 24-bit PCM audio.
 - Each Sound Track contains labelling to identify its audio configuration, language, etc.
 - An embedded XML document containing descriptive metadata.
 - A Segmentation Track to identify the in-point and out-point of each programme segment.
 - A continuous Timecode Track.

4.2 Additional Video Constraints

The choice of AVC / H.264 coding style (LongGOP (inter coded) or I-Frame (intra coded)) must be discussed in advance with the broadcaster.

The AVC / H.264 coded video in each MXF file must have a mean data rate (mean value across the entire file) of at least:

- LongGOP (inter coded):
 - 2160p/25: 150Mb/s
 - 2160p/50: 250Mb/s*
- I-Frame (intra coded):
 - 2160p/25: 250Mb/s
 - 2160p/50: 500Mb/s

5.0 Relevant Research and Standards Papers

- 1) International Standard ISO/DIS 9241-392, Ergonomics of Human System Interaction, Part 392 [Link](#)
- 2) U.S. Army Research Institute for the Behavioral and Social Sciences, Research Report 1832, Introduction to and Review of Simulator Sickness Research [Link](#)

6.0 VR Terminology

360 Content	Post produced or Live 360 Video content with limited interactivity
Cardboard	Google project to display 360 in cheap viewer using Gyro, accelerometer, compass of smart phones
Equirectangular	Projection of a spherical image in a rectilinear image frame
Haptic Feedback	Palpable feedback, usually servo-driven in human interfaces (hand controllers)
HMD	Head Mounted Display
Hot Spot	Virtual object embedded in 360 content enabling additional functionality
Lat/Long	Projection of a spherical image in a rectilinear image frame
Nadir	Lowest point of a sphere – often used to refer to dead-spot in image below camera array
Nodal point	The point at which the optical centres of the camera array is centered
Simulator Sickness	Uneasiness or disorientation when wearing a Head-Mounted Display
Spatial Audio	Audio which is perceived to have orientation in three possible axes: L/R, Up/Down, Back/Front
Spherical Video	Video content captured or displayed simultaneously in all directions
Stitch Lines	Visible junctures of blended or overlapped images from multiple camera arrays
Unity	(Game Engine) A cross-platform engine for game and VR design and implementation
VR	Virtual Reality, historically referred to CGI -based 360 immersive experiences, now 360 video too
Zenith	Highest point of a sphere – often used to refer to dead-spot in image above camera array

7.0 Version Control

Version	Date	Section	Update
1.0	12/08/16	-	Initial publication