MPEG Immersive Media

Thomas Stockhammer (Qualcomm Incorporated)

On behalf of MPEG

MPEG

- Organized under ISO/IEC
- Some joint work with ITU-T, e.g. HEVC
- Participants are accredited by their national organization, i.e. country
- Development of specifications follows a due process structure; voting conducted by country
- Usually meets 3-4 times per year; roughly 400 experts attend each meeting
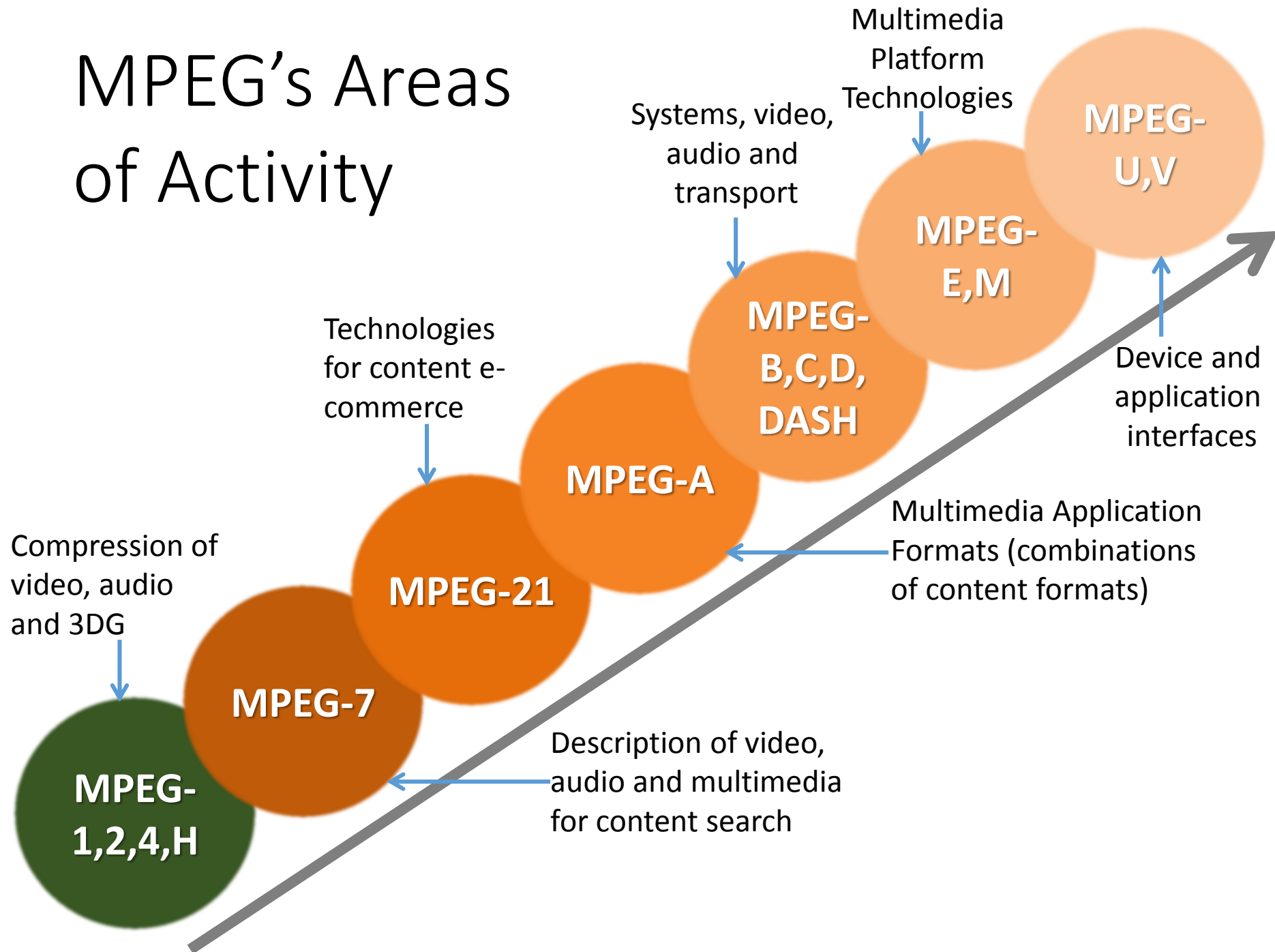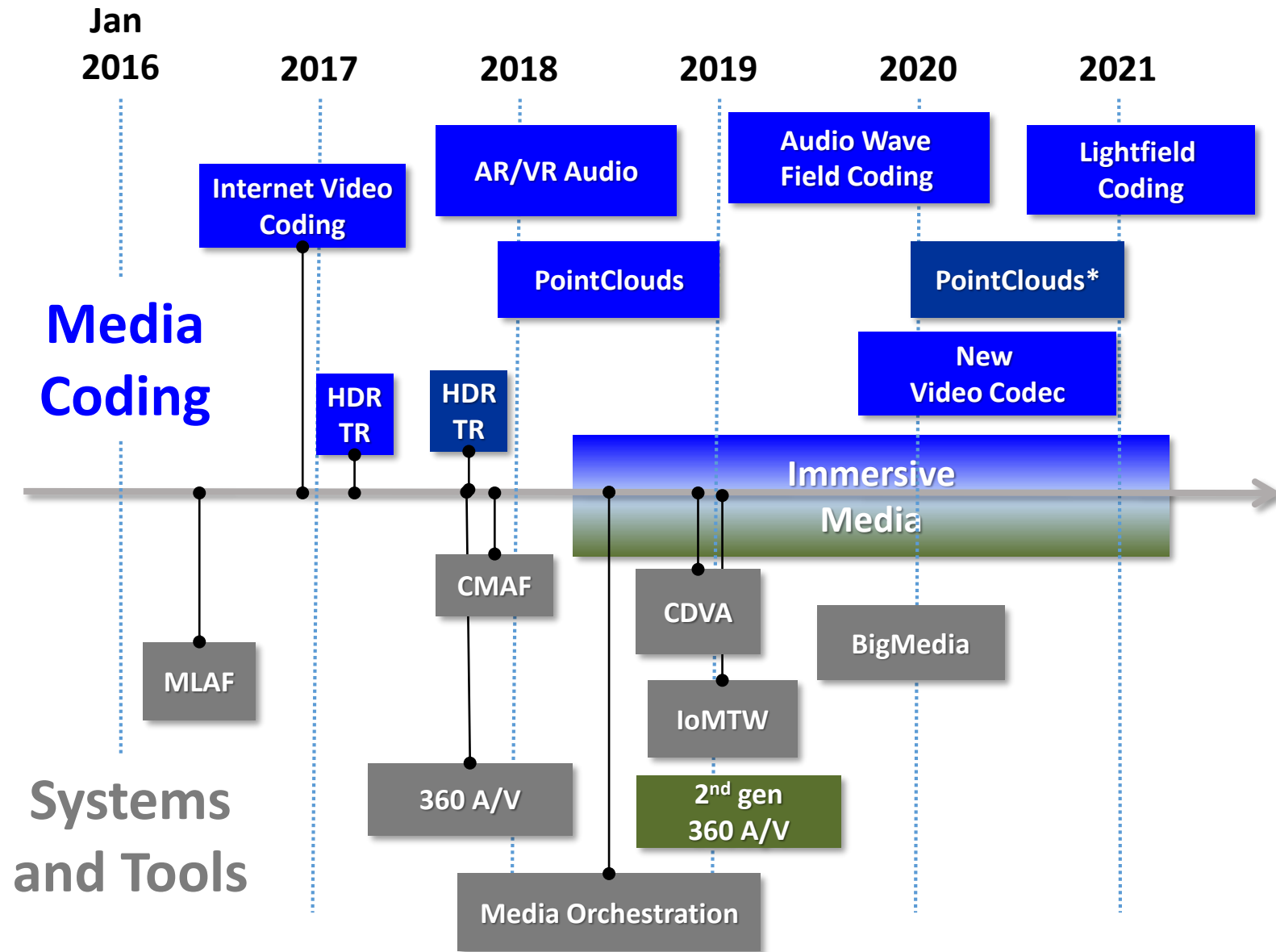
MP20 Standardisation Roadmap
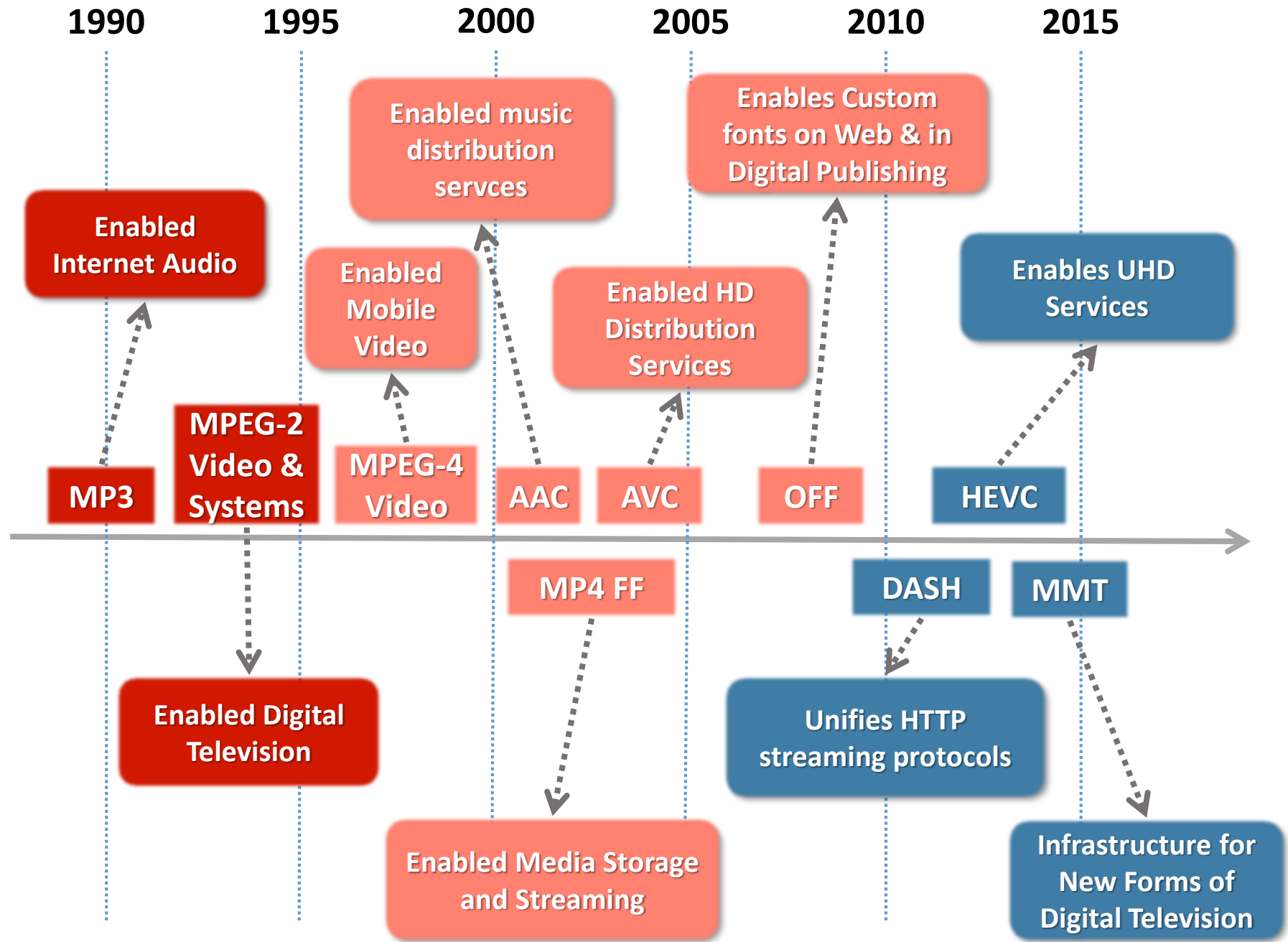
Questions to MPEG's Customers

- Which **needs** do you see **for media standardisation**, between now and years out?

- What MPEG standardisation **roadmap** would best meet your needs?

- To accommodate your use cases, what should **MPEG's priorities** be for the delivery of specific standards? For example, do you urgently need something that may enable basic functionality now, or can you wait for a more optimal solution to be released later?

Program of Workshop Jan 18th, 2017

16:00	**Opening Address**	Leonardo Chiariglione, MPEG
16:10	**MP20 Roadmap**	Rob Koenen; José Roberto Alvarez, MPEG
16:20	**DVB VR Study Mission Report**	David Wood, EBU
16:40	**Video formats for VR: A new opportunity to increase the content value… But what is missing today?**	Gilles Teniou, Orange
17:00	**Snapshot on VR services**	Ralf Schaefer, Technicolor
17:20	*Break*	
17:35	**Today's and future challenges with new forms of content like 360°, AR and VR**	Stefan Lederer, Bitmovin
18:55	**The Immersive Media Experience Age**	Massimo Bertolotti, Sky Italia
18:15	**Discussion**	All Speakers
18:40	**Final Remarks, Conclusion**	Chairs
18:45	End	

What is virtual reality?

Virtual Reality is a rendered environment (visual and acoustic, pre-dominantly real-world) providing an immersive experience to a user who can interact with it in a seemingly real or physical way using special electronic equipment (e.g. display, audio rendering and sensors/actuators)*

*MPEG's definition

Elements of VR from MPEG point of view

- Single view
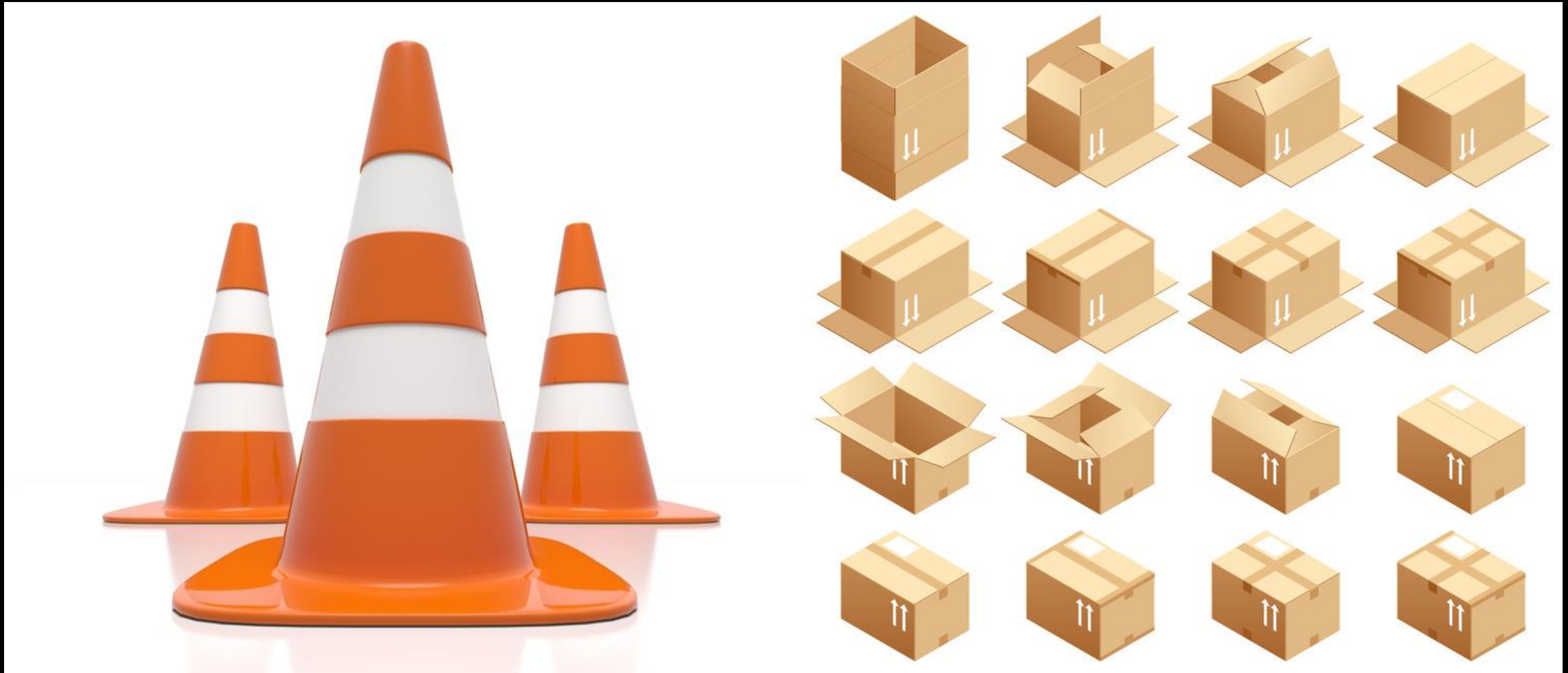- Multiple view 3 DoF
- Multiple view with continuous parallax + 6 DoF

Immersive audio

- Projection of audio waveforms in more natural way

- Listener receives audio signal coherent with his/her position

Summary of MPEG VR Questionnaire Results

Introduction

- In August and September 2016, MPEG conducted an informal Survey to better understand the needs for standardisation in support for VR applications and services.
- This document summarises the results of the Survey
- The Summary does not list individual comments; these have been analysed by MPEG and are reflected in the Conclusions, which are also included in this Survey
- This result summary can be distributed to interested parties. It has also been sent to the Respondents.

Instructions given to respondents

ISO/IEC SC29/WG11, also known as the Moving Picture Experts Group (MPEG), is aware of the immense interest of several industry segments in content, services and products around Virtual Reality (VR). In order to address market needs, MPEG has create the following survey.

In order to provide some context, consider the following definition for Virtual Reality: "Virtual Reality is a rendered environment (visual and acoustic, pre-dominantly real-world) providing an immersive experience to a user who can interact with it in a seemingly real or physical way using special electronic equipment (e.g. display, audio rendering and sensors/actuators)."

MPEG believes that VR is a complex ecosystem and that already deployed technologies can begin to fulfill the very high commercial expectations on VR services and applications, but standards-based interoperability for certain aspects around VR is required. Therefore MPEG is in the process of identifying those technologies that are relevant to market success in order to define an appropriate standardization roadmap. The technologies considered include, but are not restricted to, video and audio coding and compression, metadata, storage formats and delivery mechanisms.

This questionnaire has been developed with the goal of obtaining feedback from the industry on the technologies whose standardisation may have a positive impact on VR adoption by the market. The questionnaire will be closed on 23rd September 2016. Should this deadline not be manageable for you, please contact the organizers and we will attempt to accommodate your request for a possible extension.
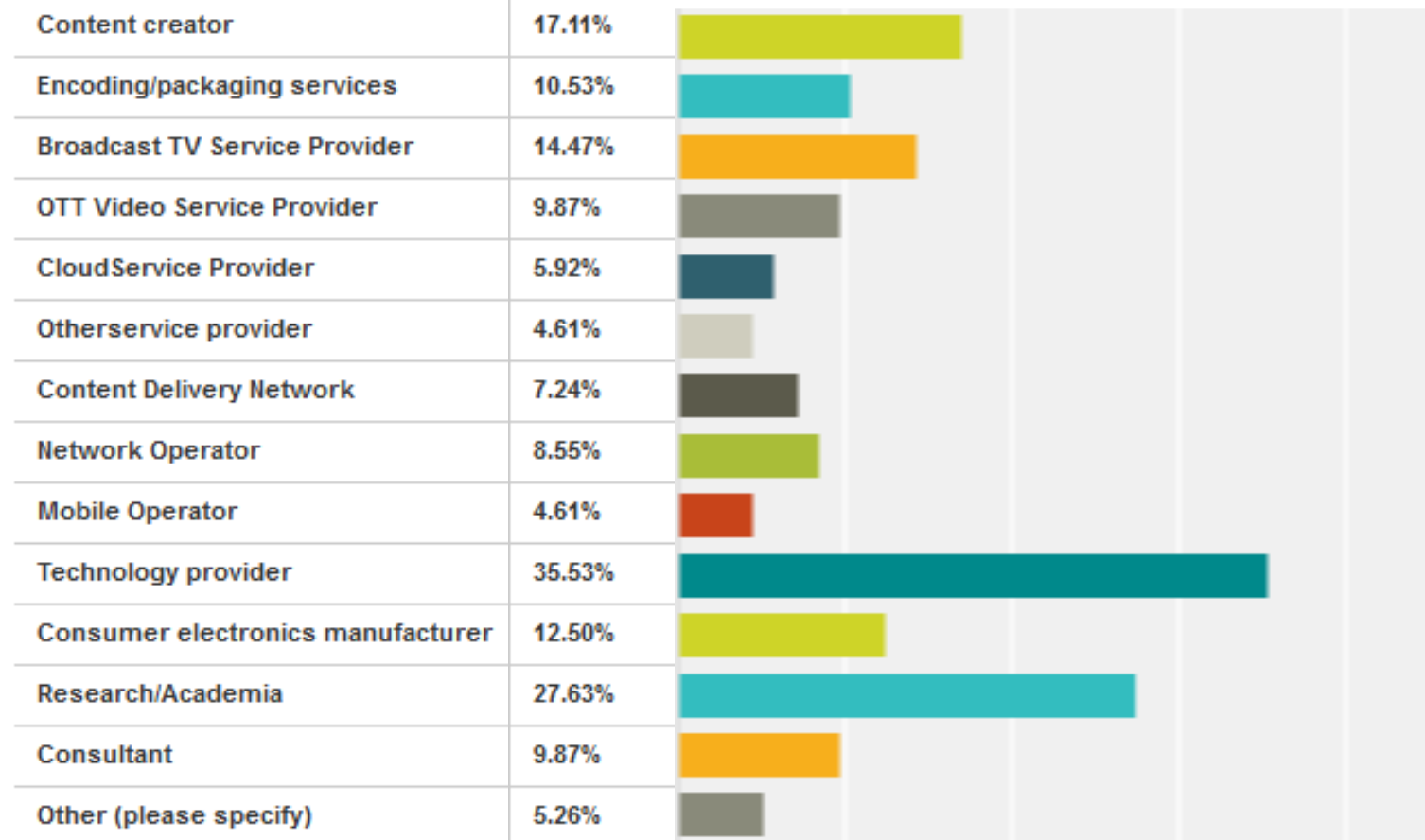
Please attempt rate/answer all items in the questions. However, while we seek complete answers, we are also interested in receiving partially filled out questionnaires.

Please use the comment box below a question if you wish to make a comments or suggestions. You may also use the comment box at the very end of the questionnaire for general comments. The use of comment boxes is encouraged because they help us disambiguate your answers to our questions.
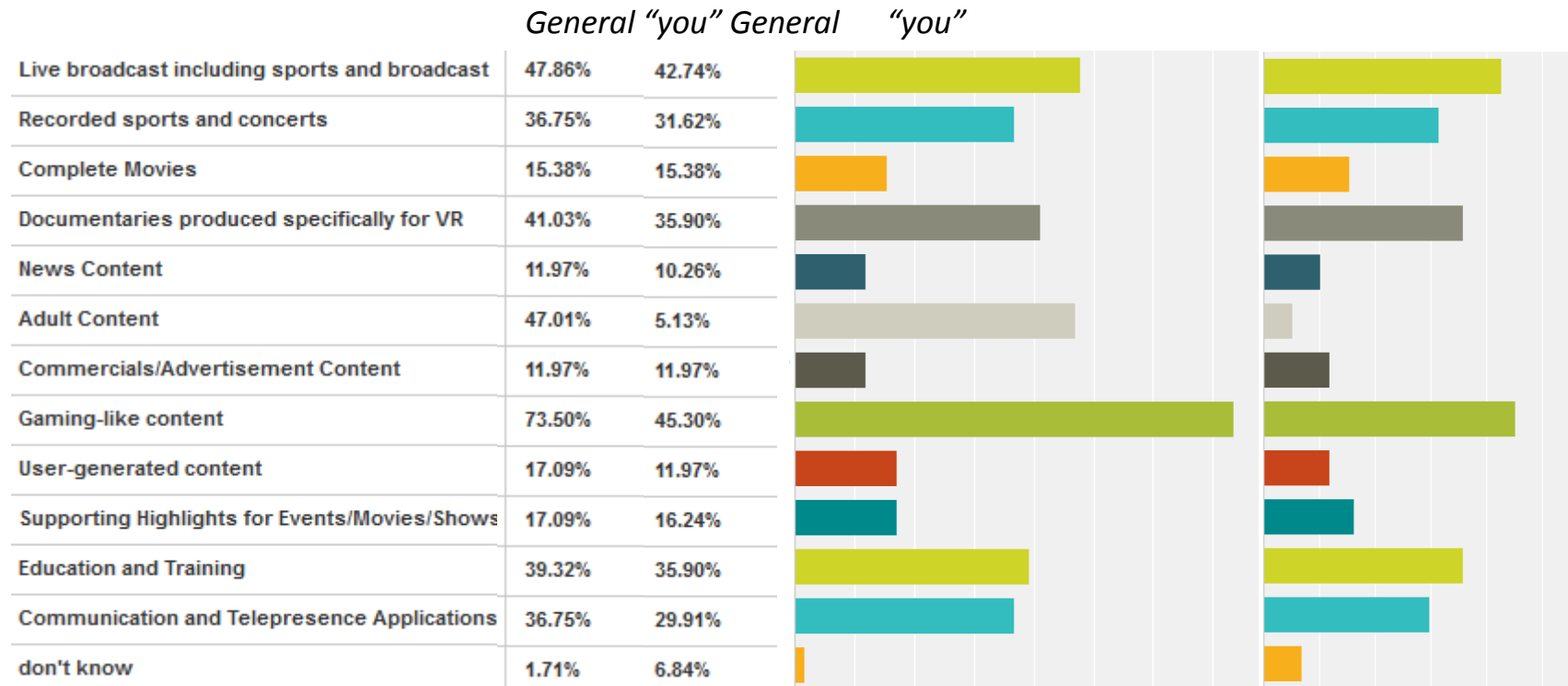
Thank you for participating in our survey. Your feedback is important.

The Chairs of the MPEG Virtual Reality Ad-hoc Group

What Business are you in?

Content creator	17.11%
Encoding/packaging services	10.53%
Broadcast TV Service Provider	14.47%
OTT Video Service Provider	9.87%
CloudService Provider	5.92%
Otherservice provider	4.61%
Content Delivery Network	7.24%
Network Operator	8.55%
Mobile Operator	4.61%
Technology provider	35.53%
Consumer electronics manufacturer	12.50%
Research/Academia	27.63%
Consultant	9.87%
Other (please specify)	5.26%

Use of VR is for …

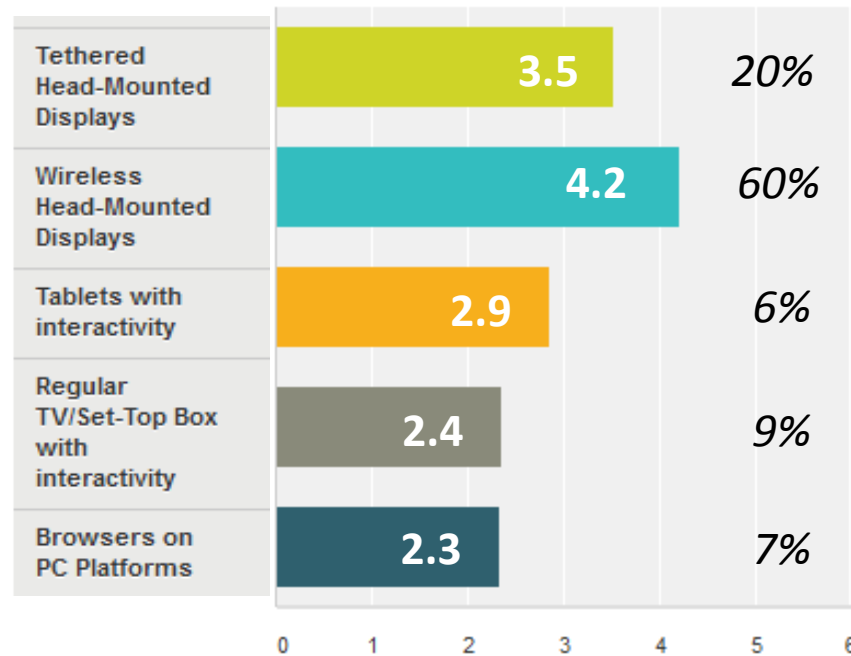|	General "you"	"you"
Live broadcast including sports and broadcast	47.86%	42.74%
Recorded sports and concerts	36.75%	31.62%
Complete Movies	15.38%	15.38%
Documentaries produced specifically for VR	41.03%	35.90%
News Content	11.97%	10.26%
Adult Content	47.01%	5.13%
Commercials/Advertisement Content	11.97%	11.97%
Gaming-like content	73.50%	45.30%
User-generated content	17.09%	11.97%
Supporting Highlights for Events/Movies/Shows	17.09%	16.24%
Education and Training	39.32%	35.90%
Communication and Telepresence Applications	36.75%	29.91%
don't know	1.71%	6.84%

Other mentions:
- *Episodic content (<30 min)*
- *Professional support tools.*
- *Remote sign language interpreting*
- *Too little options, believe for almost all of them*
- *Note, like many in the industry, VR does not represent 360 degree video solutions and can only be for rendered content.*
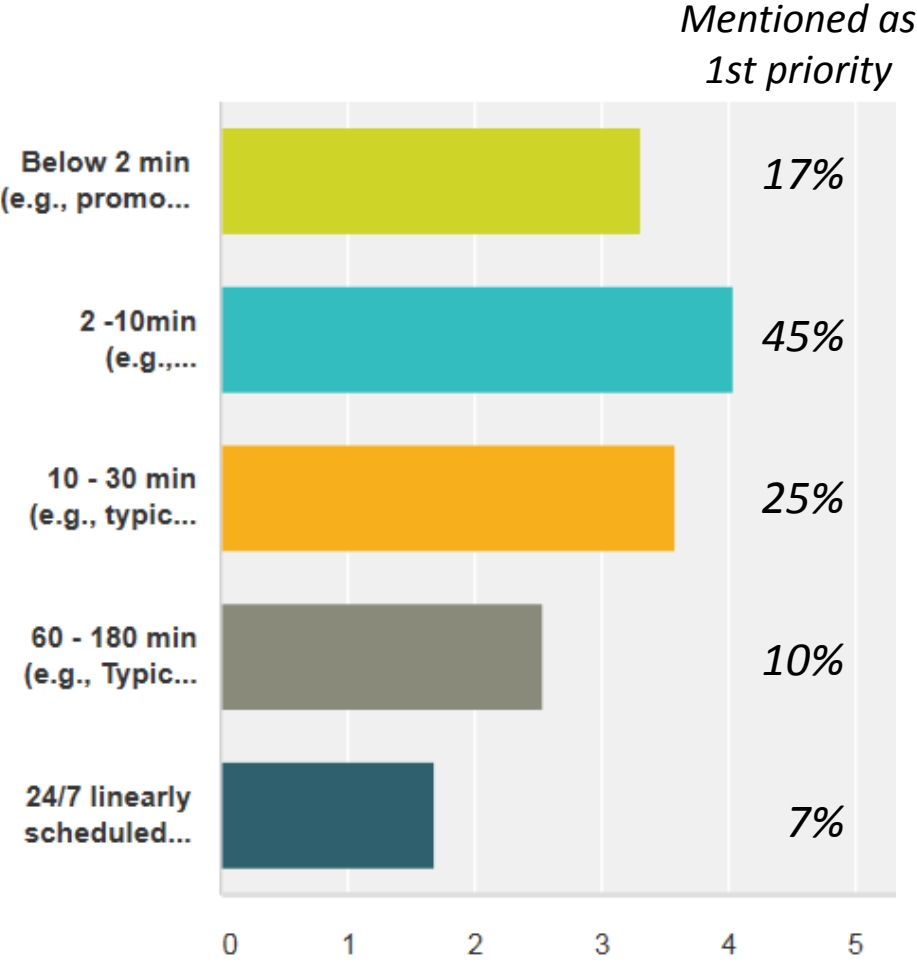- *Eventually VR will be used everywhere and will replace existing services*

Most relevant devices?

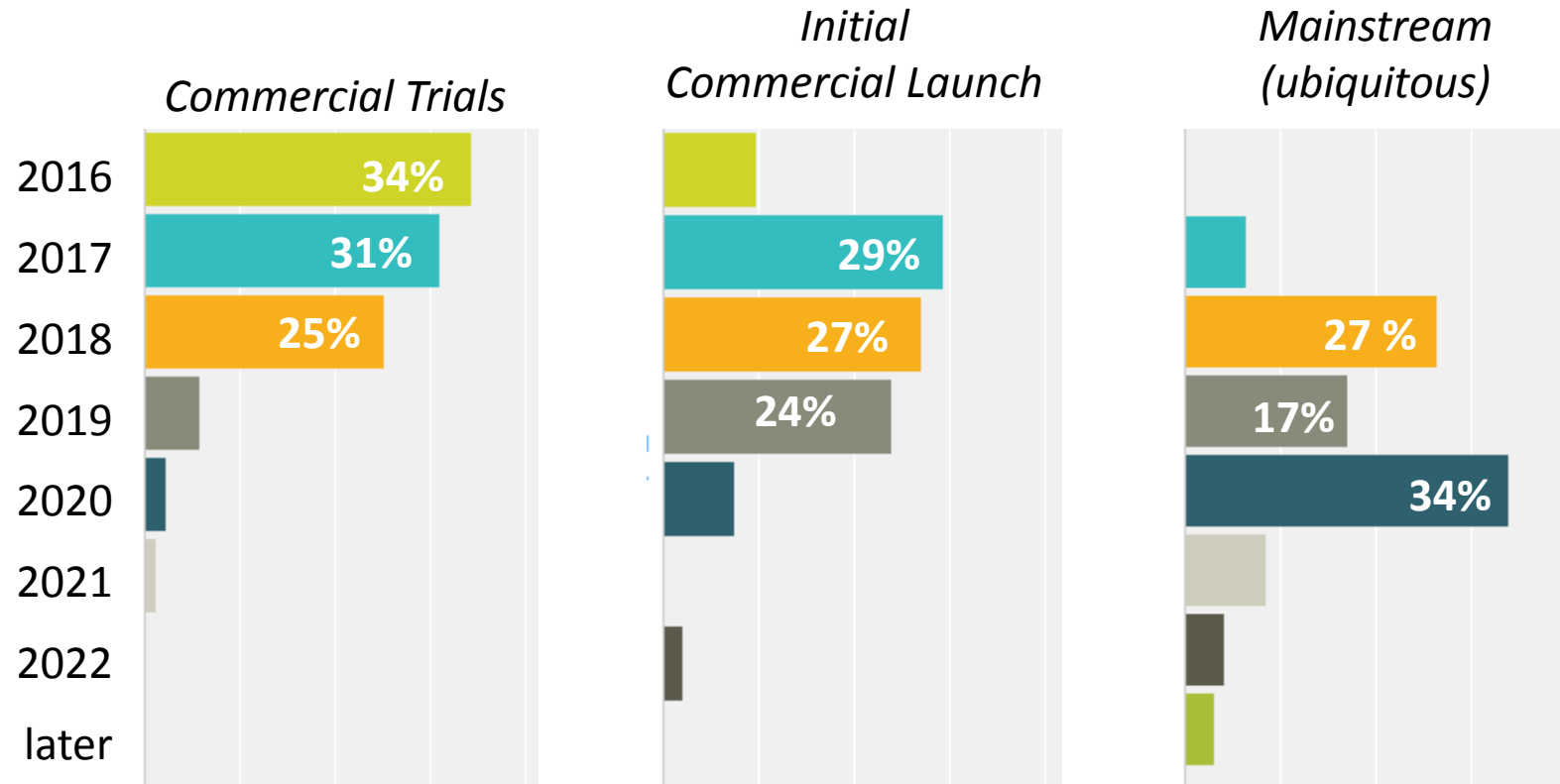|		
Tethered Head-Mounted Displays	**3.5**	*20%*
Wireless Head-Mounted Displays	**4.2**	*60%*
Tablets with interactivity	**2.9**	*6%*
Regular TV/Set-Top Box with interactivity	**2.4**	*9%*
Browsers on PC Platforms	**2.3**	*7%*

Conclusion: VR will be used on all these devices, while HMDs are considered the most important, especially wireless ones.
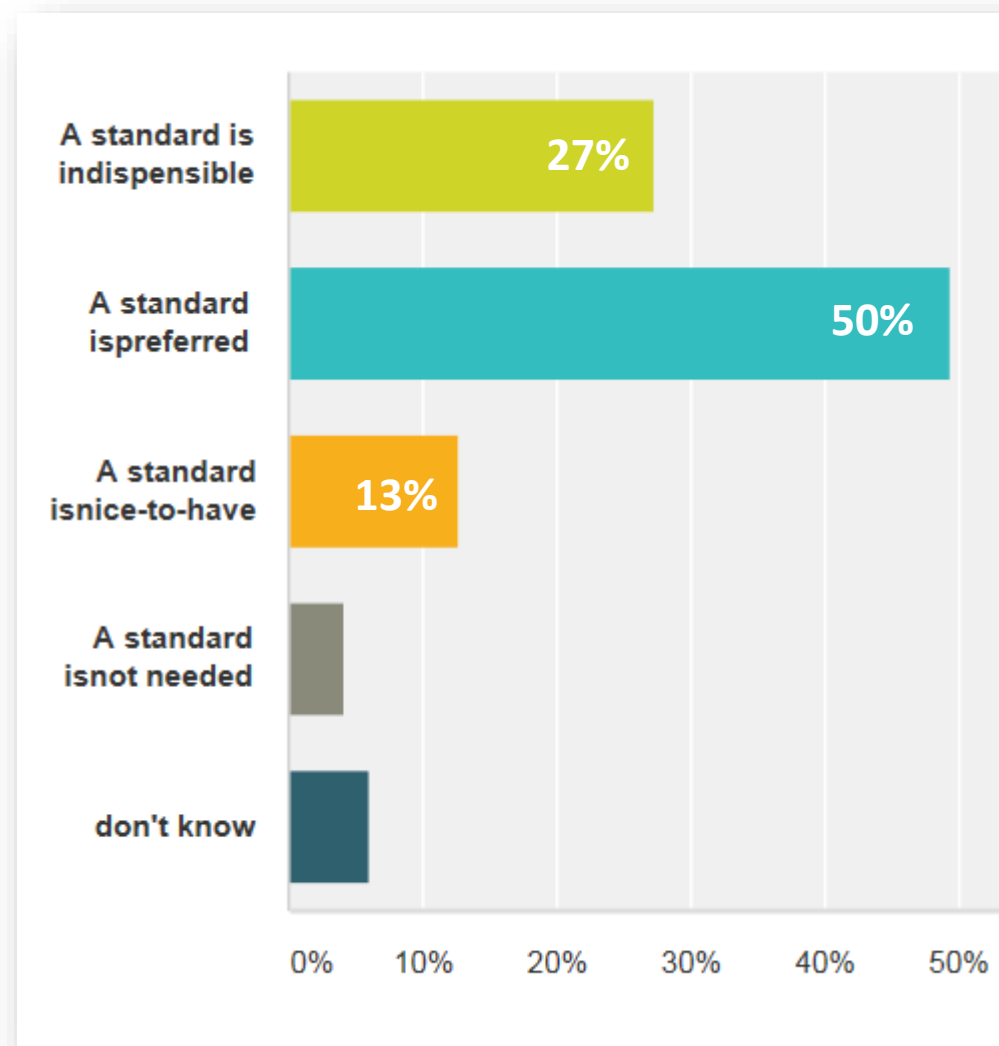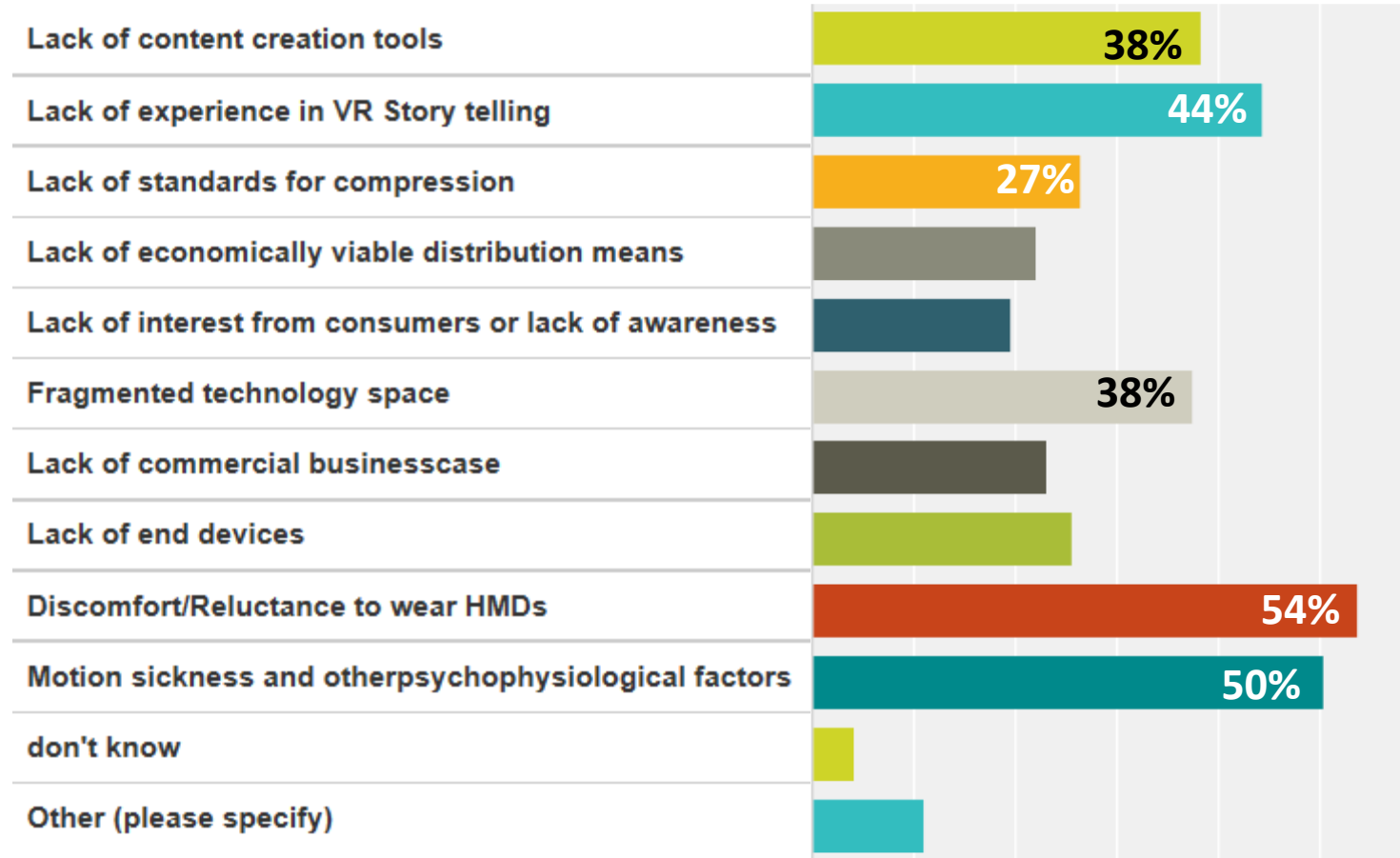
Typical Content Duration?

Mentioned as 1st priority

Duration	1st priority
Below 2 min (e.g., promo...	17%
2 -10min (e.g.,...	45%
10 - 30 min (e.g., typic...	25%
60 - 180 min (e.g., Typic...	10%
24/7 linearly scheduled...	7%

Deployment timelines?

Commercial Trials

Year	
2016	34%
2017	31%
2018	25%
2019	
2020	
2021	
2022	
later	

Initial Commercial Launch

Year	
2016	
2017	29%
2018	27%
2019	24%
2020	
2021	
2022	
later	

Mainstream (ubiquitous)

Year	
2016	
2017	
2018	27 %
2019	17%
2020	34%
2021	
2022	
later	

- Commercial trials this and next year.
- Launches starting seriously next year
- Mainstream in 2-4 year timeframe

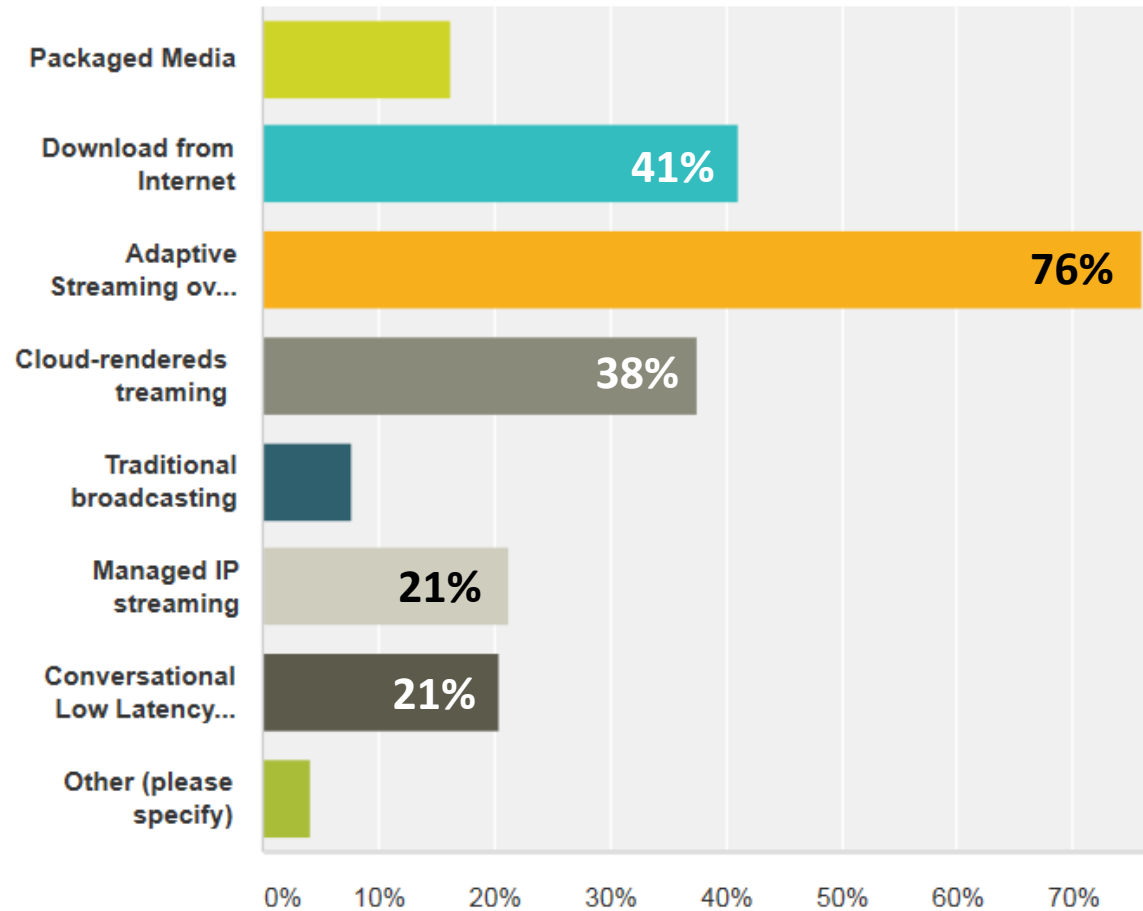# How about standards?

What Hurdles/Obstacles?

Obstacle	Percentage
Lack of content creation tools	38%
Lack of experience in VR Story telling	44%
Lack of standards for compression	27%
Lack of economically viable distribution means	
Lack of interest from consumers or lack of awareness	
Fragmented technology space	38%
Lack of commercial businesscase	
Lack of end devices	
Discomfort/Reluctance to wear HMDs	54%
Motion sickness and otherpsychophysiological factors	50%
don't know	
Other (please specify)	

Major Cost Factors?

Who Selects Technology?

The content producers, such as movie studios	**41%**
The Web content providers	
The Internet community, such asbrowser vendors	
The game development companies	41%
The HMD platform companies	38%
The mobile device companies	
Service providers, such as sports leagues or pay TVservice providers	26%
OTT Service providers	
Standardization organizations	
don't know	
Other (please specify)	

Content is King, also here (Game Development is also Content creation)
Consumers were mentioned quite a few times under "Other"

Most Important Delivery Means?

Category	Value
Packaged Media	
Download from Internet	41%
Adaptive Streaming ov...	76%
Cloud-rendereds treaming	38%
Traditional broadcasting	
Managed IP streaming	21%
Conversational Low Latency...	21%
Other (please specify)	

What Motion-Photon Latency vs. Bitrate is required?

- A very elaborate question that not everyone completed. A rough summary is as follows:
  - At least 10 -20 Mbit/s required at 5 msec
  - 20 - 40 Mbit/s sec required at 10 msec
  - No consensus at 20 msec (100 Mbit?)
  - Never good enough at 50 msec or higher

Video Production Formats in the near Future?

Format	Percentage
less than 360 degree spherical video with 3 degrees of freedom (head stays in but user can look around)	37%
full 360 degree spherical video with 3 degrees of freedom	65%
spherical video with 6 degrees of freedom (user can also move around)	44%
monoscopic 3D video	
stereoscopic 3D video	35%
Light fields	
Point clouds	
do not know	
Others (please specify)	
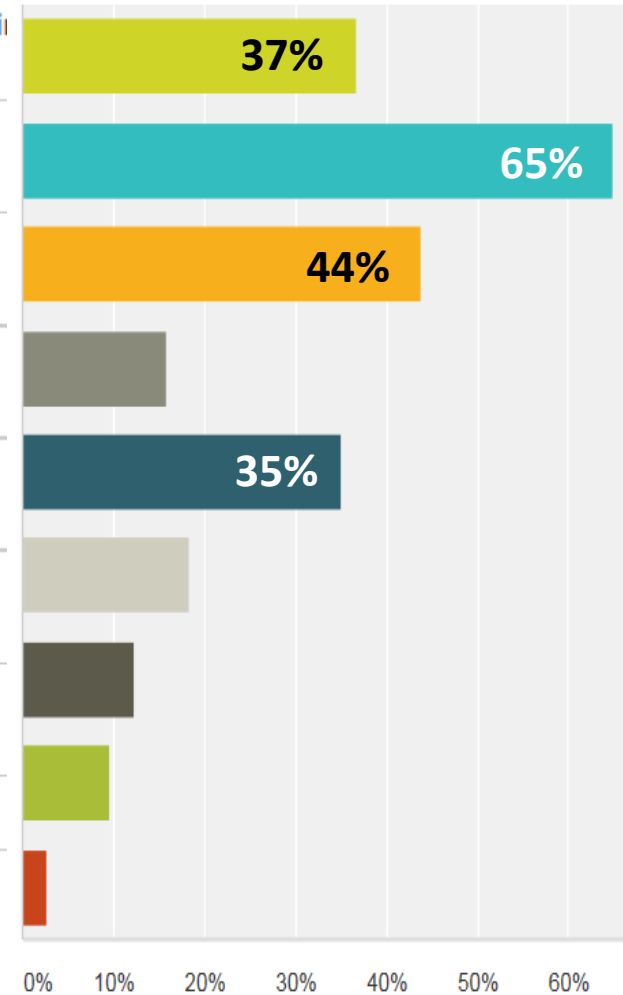
HEVC (including extensions) sufficient?

Yes, single decoder is sufficient	24%
Yes, but multiple parallel decoders are necessary	33%
Yes, with some adjustments (please specify shorto	
No (please provide comments in Comment box)	
Don't know	23%

0%   10%   20%   30%   40%
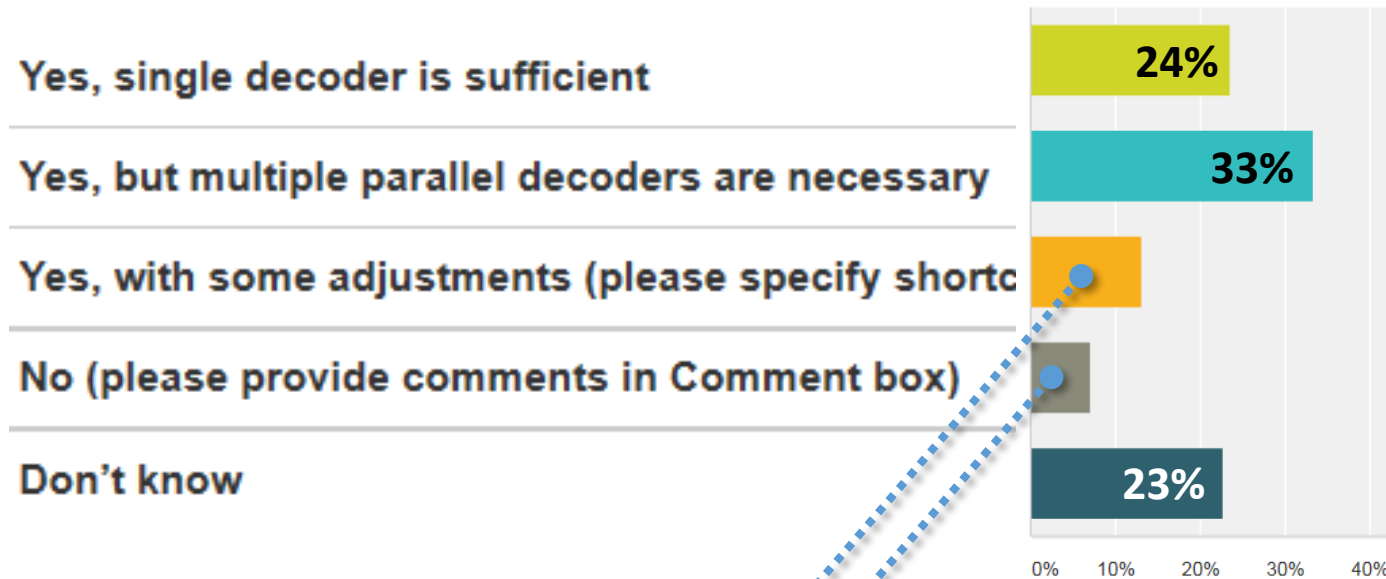
MPEG has analyses "write-in" comments

Quality Issues with 360 / 3 Degrees of Freedom?

Quality Issue	Percentage
Screen door effect	
Motion-to-photon latency	33%
Geometry artifacts due to stitching	40%
Color artifacts due to stitching	
Incorrect focus due to usage of multiple cameras	
Compression artefacts	
incorrect parallax (Incorrect positioning in space)	33%
Absence of self-presence (user embodiment)	
don't know	
Other (please specify)	
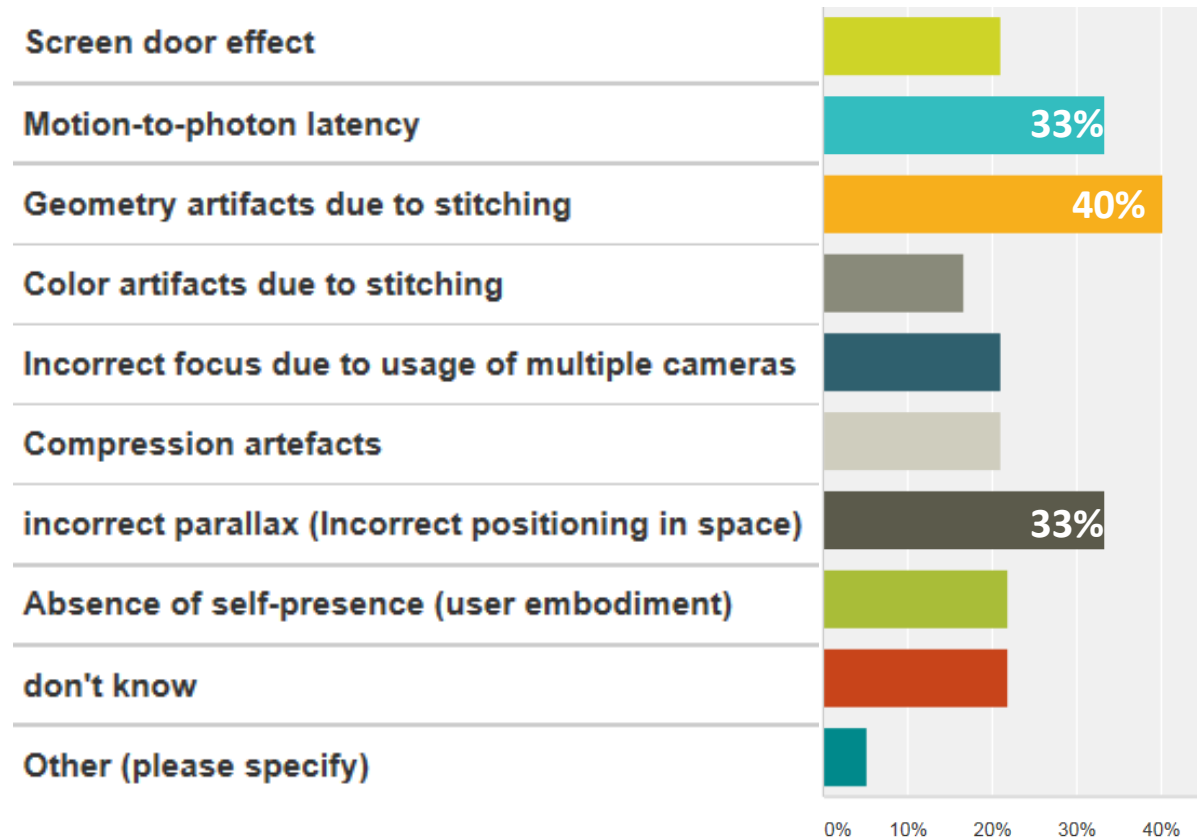
Low resolution mentioned multiple times in comments

Minimum Ingest Format Reqs?

- A very elaborate question that not everyone completed. A rough summary:
  - 30 fps inadequate, although maybe at 6k and up …
  - 60 fps acceptable at 4K and up
  - 90 fps and higher perhaps doable at HD; good enough at 4k+

Minimum Reqs per Eye?

- A very elaborate question that not everyone completed. A rough summary:
  - 30 fps never good enough, well, maybe at 8k and up?
  - 60 fps usable at 4k and up
  - 90 fps and higher clearly usable at 4k and up, but not at HD

3D Audio for VR in near future?

(max 3 answers)
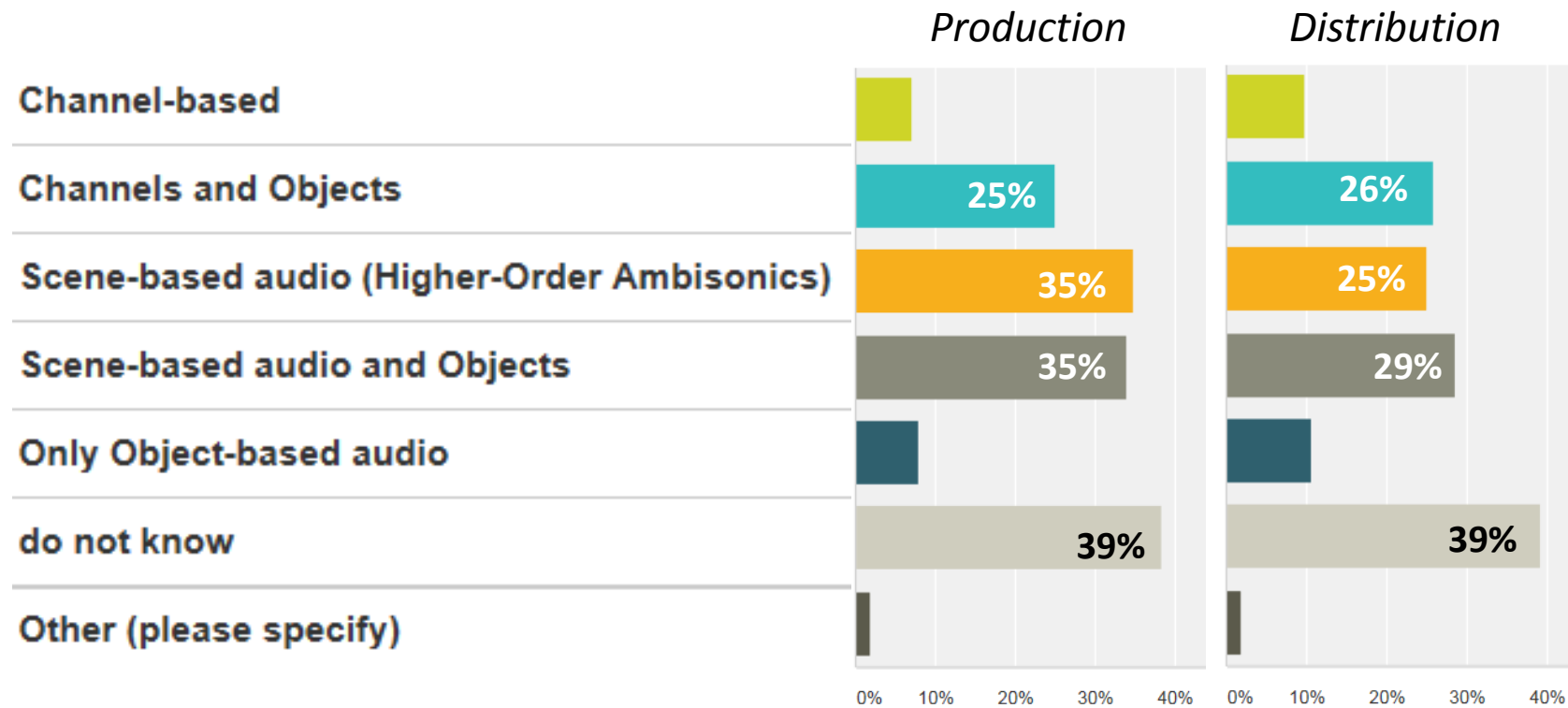
MPEG-H 3D Audio Sufficient for Initial Deployments?

MPEG will has analysed write-in answers
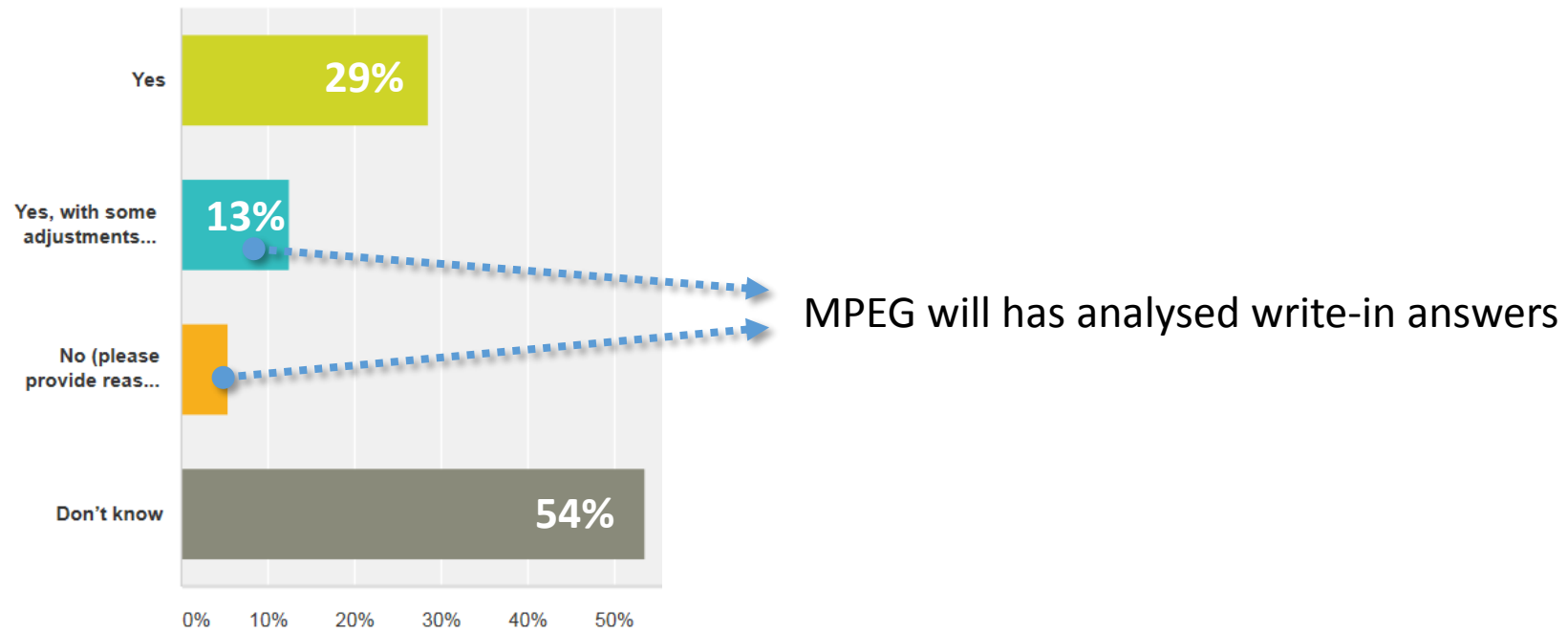
Which Quality Issues with 3D Audio?
(max 4 answers)

Audio coding artefacts	
Lack of immersiveaudio rendering (stereo-rendering only)	24%
Low spatial accuracy	18%
Motion-to-sound latency	24%
Lack of personalized experience (interaction, adaptation to listener'sears)	16%
Lack of loudness adjustment and dynamic range control	
don't know	51%
Other (please specify)	
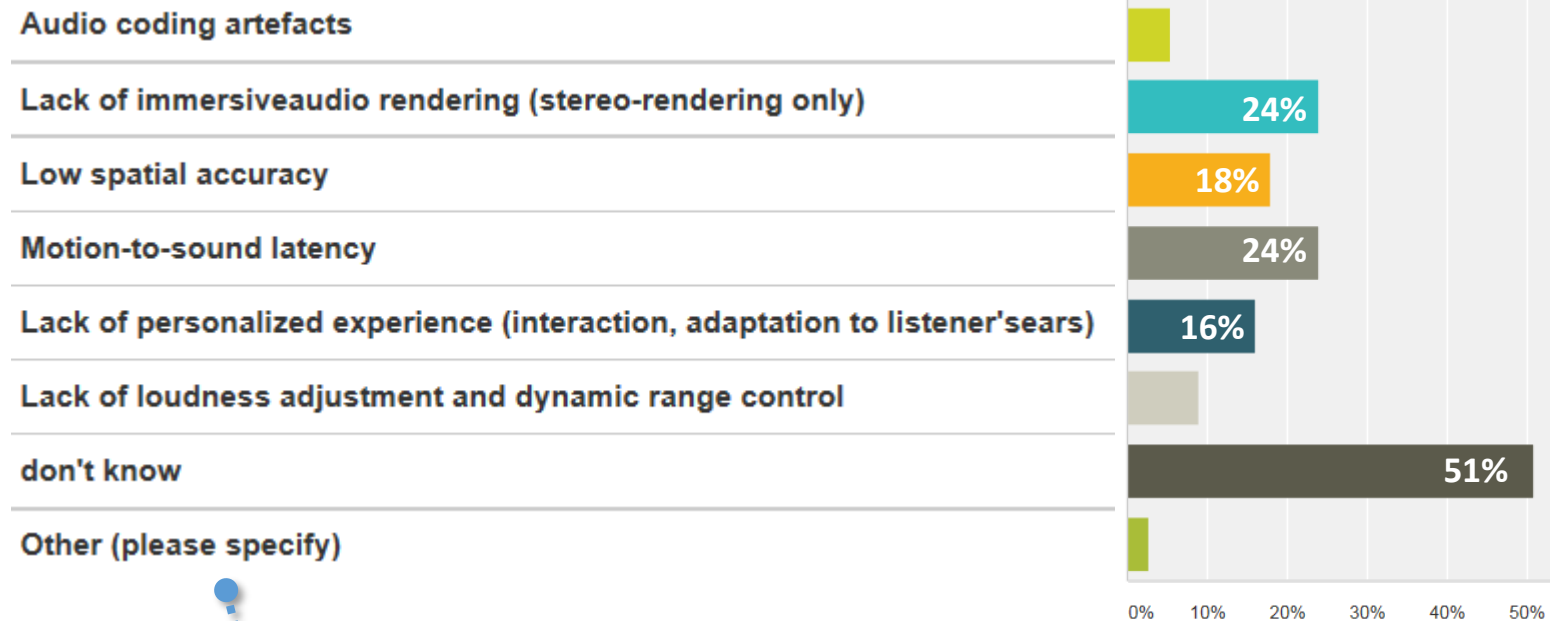
MPEG will analysed write-in answers

What Specs should MPEG Create?

Spec	Percentage
A complete end to end VR (360 multimedia) system	48%
Enabler for 3DOF spherical video compression	34%
Enabler for 6DOF spherical video compression	33%
Extensions to 2D existing video codecs (Projection Maps, etc.)	25%
Enablers for 3D audio compression	19%
Enabler for storage and file based distribution of VR Content	25%
Enabler for broadcast distribution	17%
Enabler for Internet distribution	35%
Others (please specify)	

Conclusions

There is a significant interest in having standards

- An analysis learns that there is no significant difference between MPEG participants and non-participants
- Question 14 teaches that MPEG deliver compression tools, make for a less fragmented technology space; and should support short motion-to-photon delay

Application space:

- The focus is now on 360 Video with 3 Degrees of Freedom (monoscopic or stereoscopic)
- There is a clear interest in 6 Degrees of Freedom.

Business Models

- Broadcast is considered an interesting business model by a significant amount of respondents.
- This raises the question if broadcasting brings specific requirements, and whether broadcast as a service also implies broadcast as a distribution model. Most respondents seem convinced that adaptive streaming is the best way to distribute VR content.

Transport

- Adaptive streaming is considered very important
- There is also an understanding that it needs to get better, i.e. more adaptive to viewing direction (in terms of motion to photon delay)

Video

- Most respondents believe that HEVC is useful, but a significant amount believe that extensions may be desired or required, e.g. in tiling support, or the use of

multiple decoders.

- No clear picture emerges on quality requirements for video, although it is clear that very high resolutions are desired. Current VR quality is not yet enough for a good experience, and MPEG should provide tools that enable higher quality.
- Respondents also indicate that MPEG-defined projection methods are desirable.
- Coding technologies will be required to support experiences with 6 degrees of freedom

Audio

- Many respondents did not have an opinion on Audio, but those that did, think that the required tools are available.

General

- There is a need to look at the interaction between projection mapping and video coding, and to find optimal solutions.
- Question 16 shows us that requirements from those who create the content are important, as content creators are seen as an important factor in determining what tools are used.

Timing:

- The survey gives a fairly uniform picture when it comes to deployment timelines:
- Commercial Trials: 2016 and 2017, then levelling off
- Initial Commercial Launch: 2017/2018
- Mainstream: 2018 to 2020

MPEG-i Project

Immersive Media in MPEG

Phase 1a

- Timing is what guides this phase
- Goal: to deliver a standard for 3DoF 360 VR in the given timeframe (end 2017 or maybe early 2018)
- Aim for a complete distribution system
- Based on OMAF activity; using OMAF timelines;
- Audio: a 3D Audio profile of MPEG-H geared to a 360 Audiovisual experience with 3 DoF,
- Transport: Basic 360 streaming, and if possible optimizations (e.g., Tiled Streaming)
- Video: Adequate tiling support in HEVC (may already exist) and projection, monoscopic and stereoscopic
- MPEG should be careful not to call this MPEG VR, as the quality that can be delivered in the given timeframe may not be enough.

Phase 1b

- Mainly motivated by desire by a significant part of respondents to launch commercial services in 2020
- deploy in 2020; spec ready in 2019, (which may match 5G deployments)
- Extension of 1a; focus very likely still on VR 360 with 3 DoF (again monoscopic and stereoscopic)
- If there are elements that could not be included in phase 1a, improving quality – it is not a foregone conclusion that there will be a phase 1b, and if there is such a phase, it is to be further defined what this would comprise
  - E.g., optimization in projection mapping
  - E.g., further motion-to-photon delay reductions
  - Optimizations for person-to-person communications
- Phase 1b should have some quality definition and verification

Phase 2

- A specification that is ready in 2021 or maybe 2022
- This would be a "native" VR spec ("MPEG VR")
- Goal is support for 6 DoF
- Most important element probably new video codec with support for 6 DoF; to be decided by Video Group what tools are most suitable
- Audio support for 6 degrees of freedom
- Systems elements perhaps required too in support of 6 DoF, as well as 3D graphics.

Phase 1 Baseline Technology

Omni-directional Media Application Format and others

Basic framework of VR

There can different **media formats (codec and metadata), and different signaling and transmission protocols**

There can be different camera settings, with different optical parameters

What is rendered needs to be **immersive** to the user:

- Visual: high pixel quantity and quality, broad FOV, stereoscopic display
- Sound: high resolution audio, 3D surround sound
- Intuitive interactions: minimal latency, natural UI, precise motion tracking

Image stitching and equi-rectangular mapping

Video encode

Network

Video decode

Video rendering on a sphere
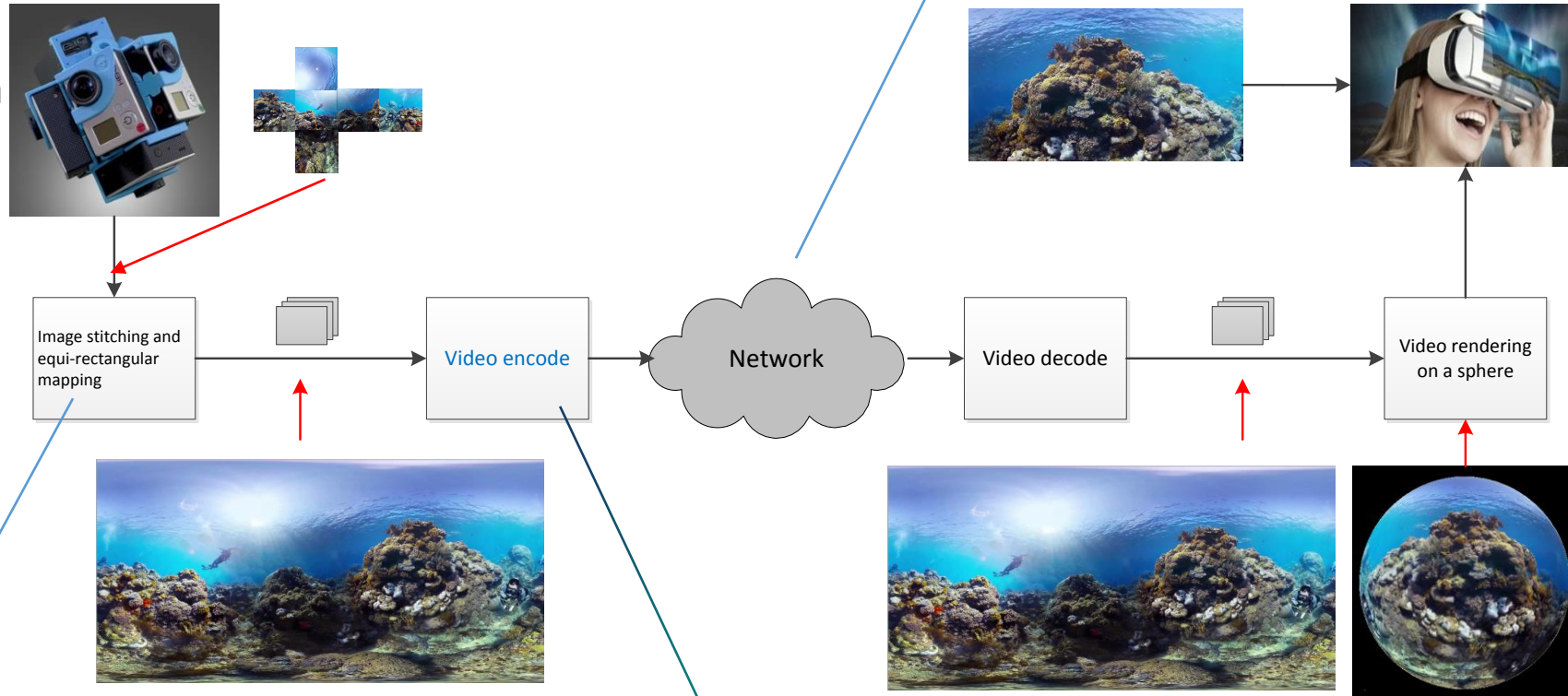
There can be different stitching and **projection mapping algorithms**

There can be **different video codecs and different encoding schemes**

Work already underway or completed

- Omni-directional Media Format
- DASH extensions for streaming VR and signaling ROI
- HEVC enhanced for flexible tiling
- Experiments for 360° stereo + 3 DoF video
- Audio completed for 3 DoF
- Experiments for many formats of projection mappings and necessary signaling

OMAF – when

- Will probably be the first industry standard on virtual reality (VR)
- MPEG started looking at VR and started the OMAF project in Oct. 2015
- Technical proposals started in Feb. 2016
- First working draft (WD) as an output of the Jun. 2016 MPEG meeting
- CD after Jan 2017 MPEG meeting, FDIS late 2017

OMAF – what

- The focus of the first version of OMAF would be 360° video and associated audio

- The scope of OMAF

  - Projection mappings

  - File format encapsulation and metadata signalling

    - Extensions to ISO base media file format (ISOBMFF) needed

  - DASH encapsulation and metadata signalling

    - Extensions to DASH needed

  - Codec and coding configurations

  - Guidelines of viewport dependent VR media processing, based on either of the following

    - Viewport dependent video encoding and decoding

    - Viewport dependent projection mapping

OMAF architecture

A: Real-world scene
B: Multiple-sensors-captured video or audio
C: Sphere video
D/D': Projected video
E/E': Coded video or audio bitstream
F/F': ISOBMFF file
G/G': DASH media presentation
M1: Projection mapping etc. metadata
M2: Metadata info of current FOV

Formats to be standardized in OMAF: E/F/G, incl. signalling of M1 in F/G.

For audio, the entire stitching process is not needed, and D is the same as B.

Sphere – equi-rectangular
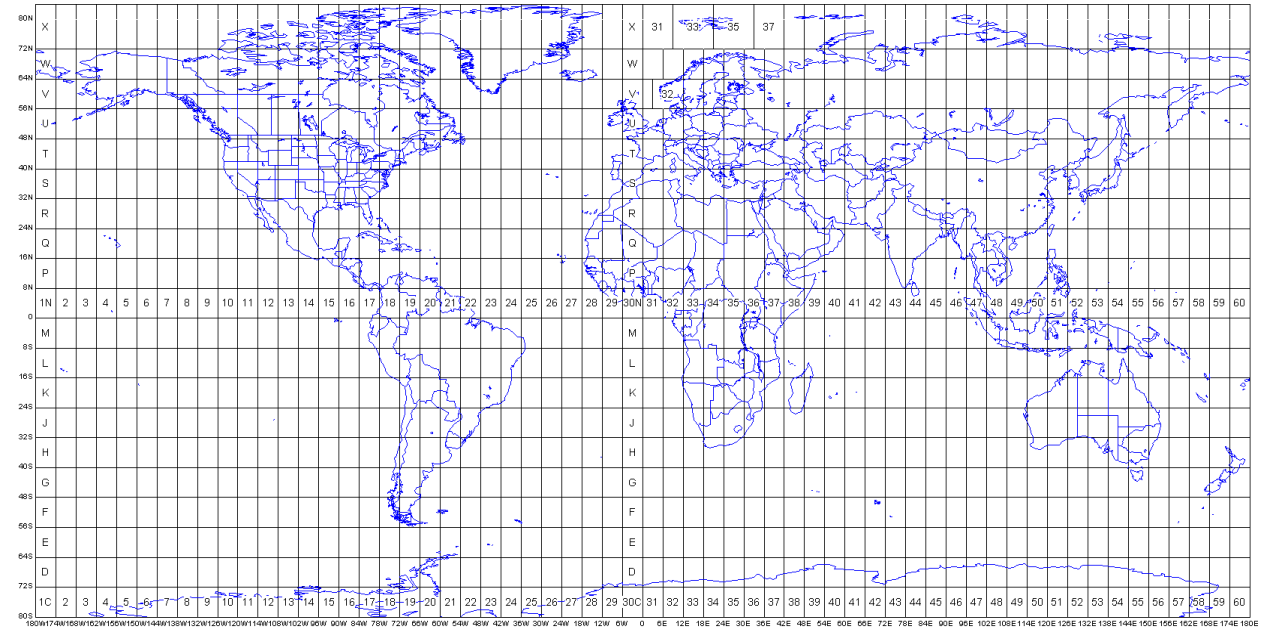
Cylinder
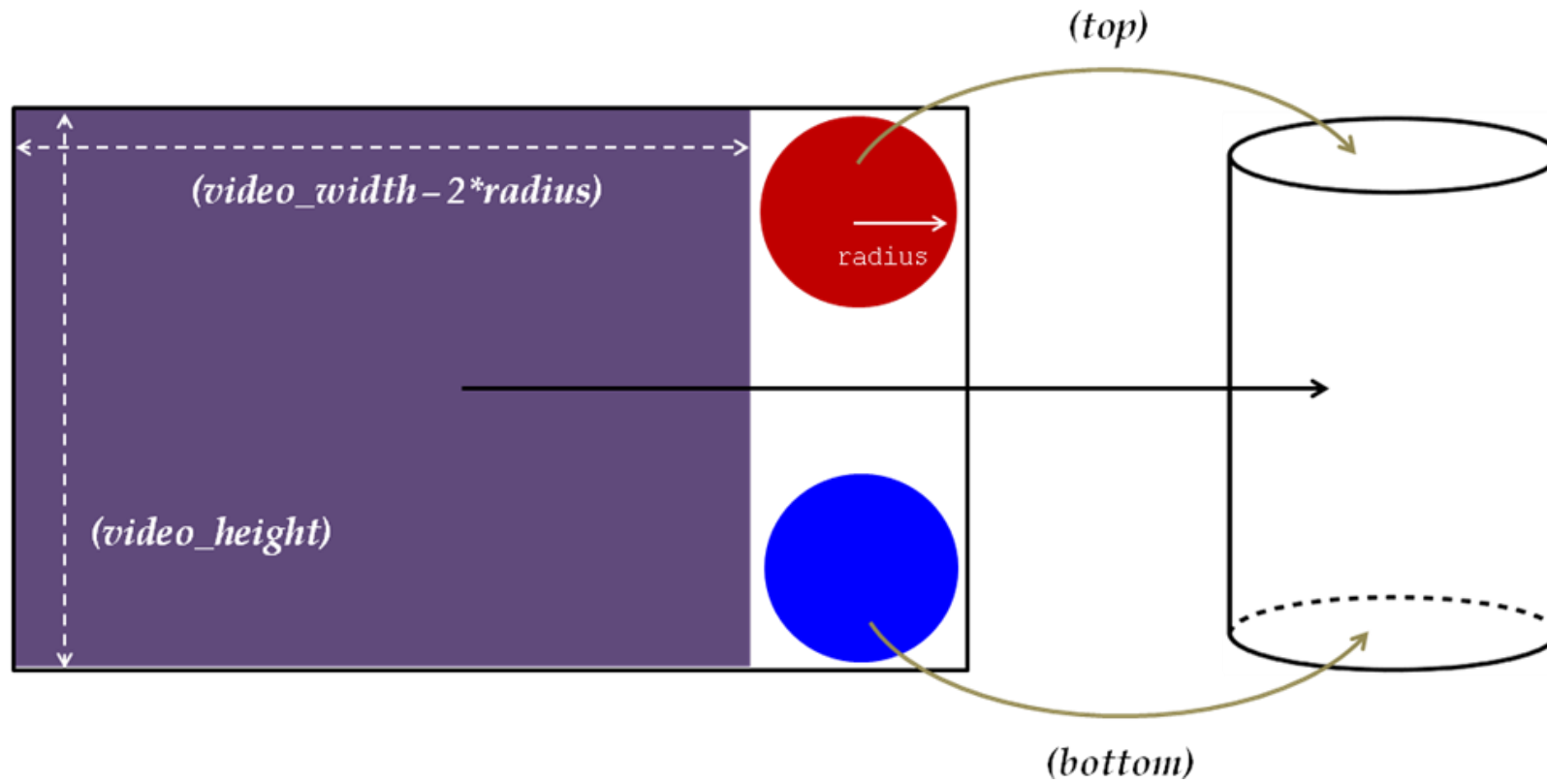
Cube map
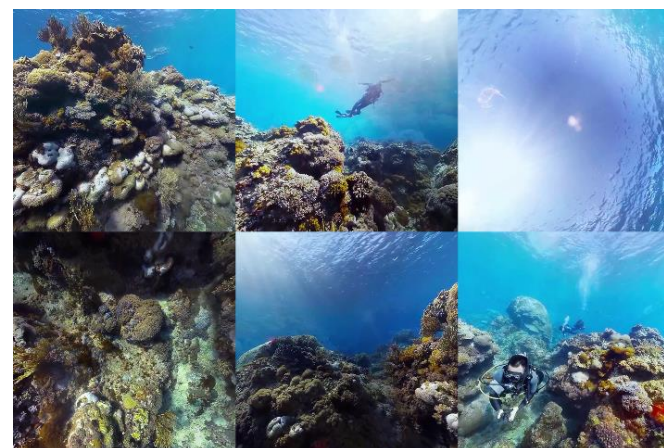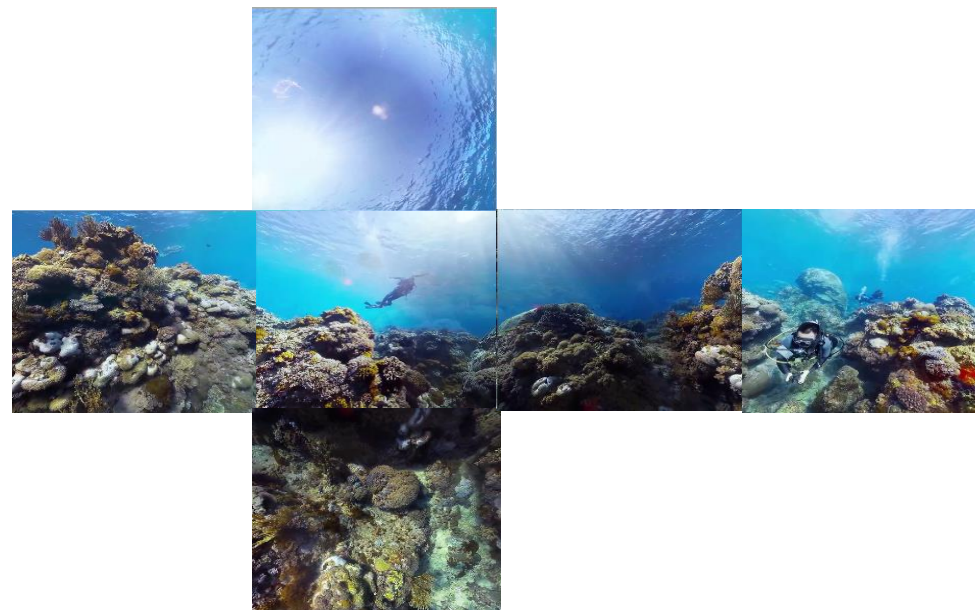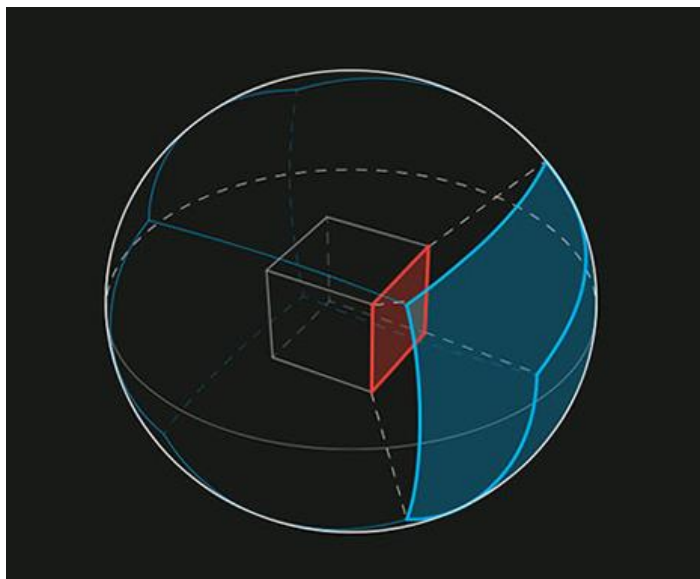
Left	Front	Up
Down	Right	Bottom

Icosahedron

Truncated pyramid

Image stitching and equi-rectangular mapping

Video encode

Network

Video decode

Video rendering on a sphere
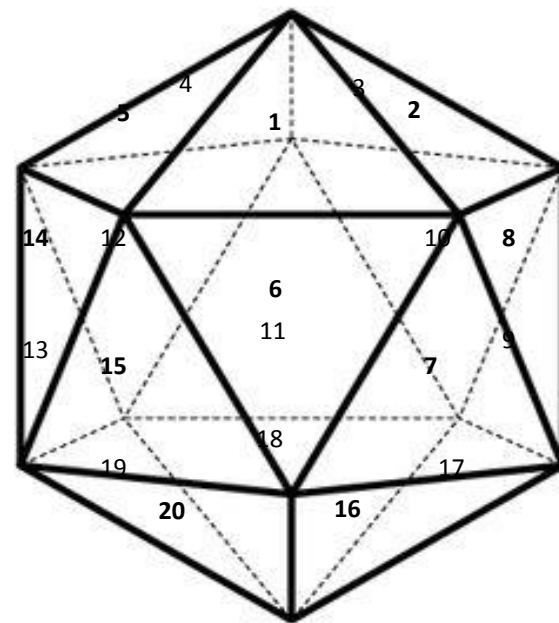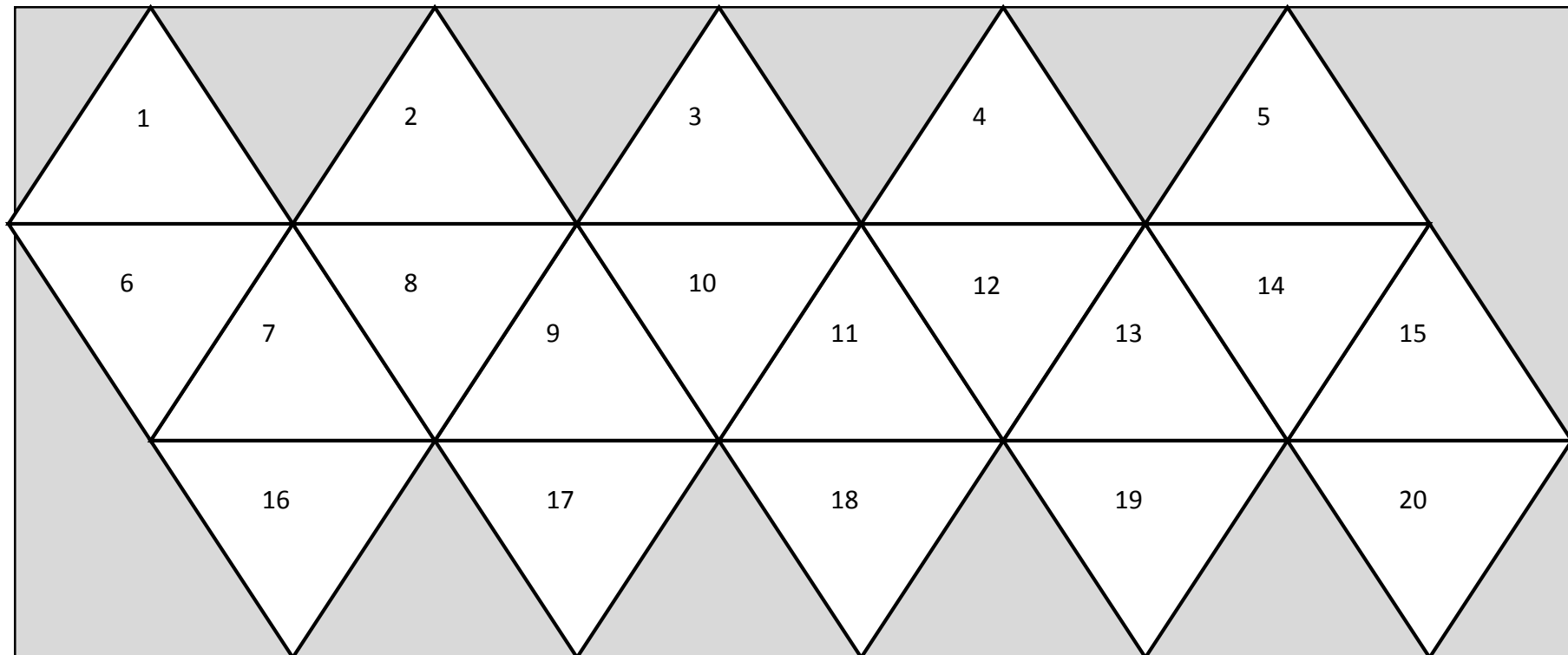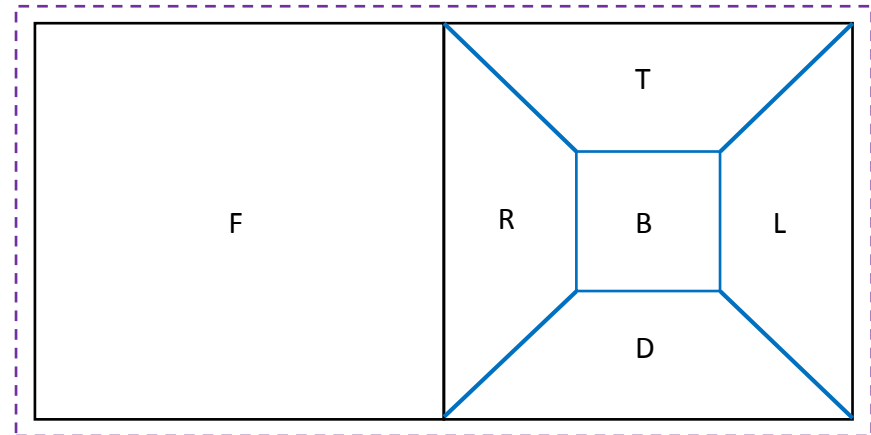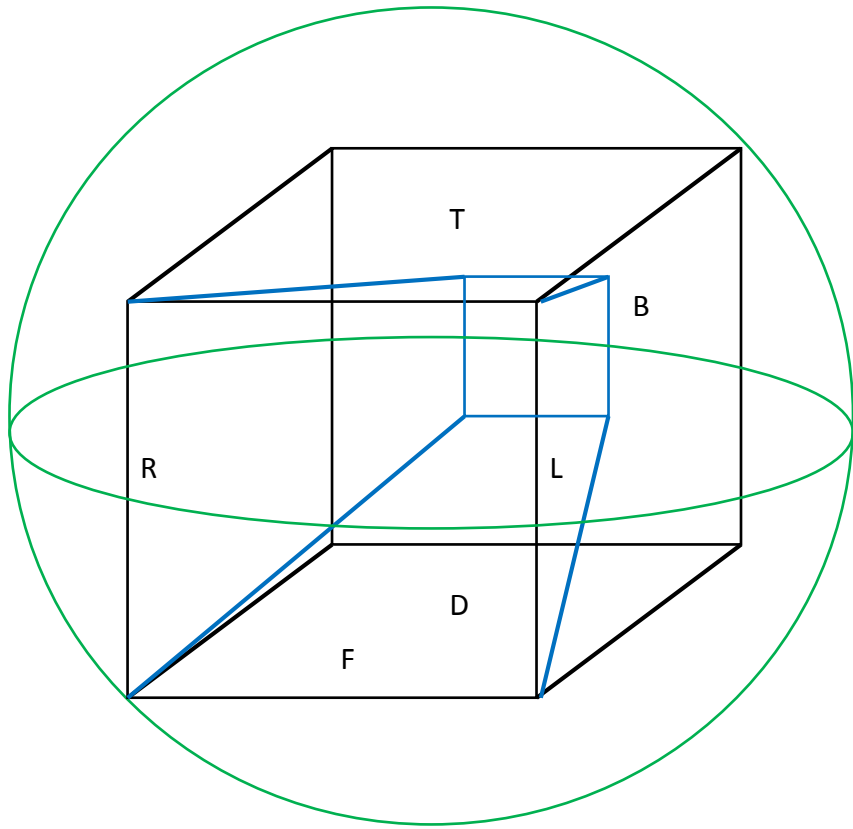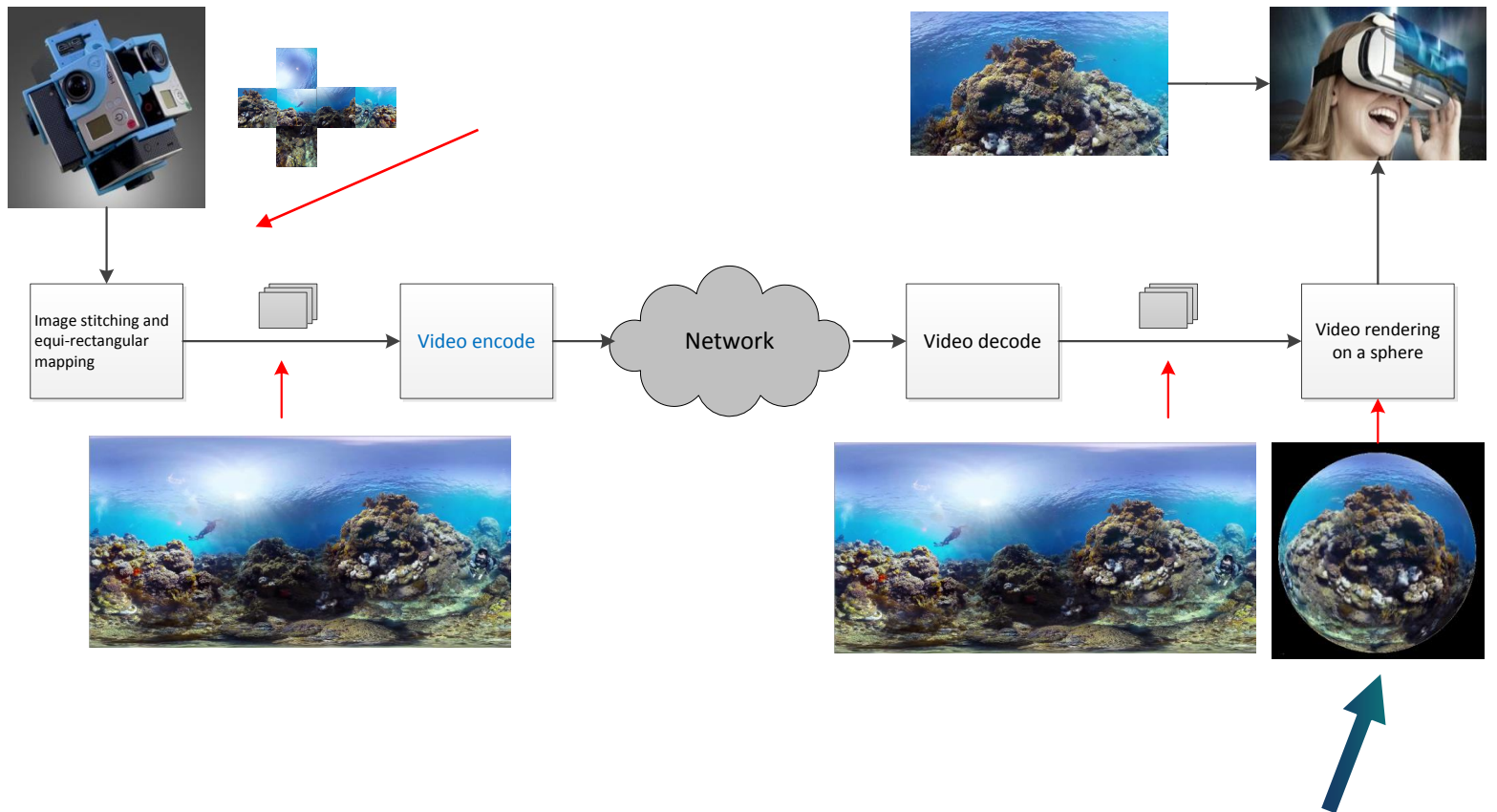
For VR to be really immersive, bandwidth and processing complexity remain two of the biggest changes

At any moment, only a portion of the entire coded sphere video is rendered

Viewport dependent video processing

- To tackle the bandwidth and processing complexity challenges
- To utilize the fact that only a part of entire encoded sphere video is rendered at any moment

- Two areas can be played with to reduce needed bandwidth and to reduce video decoding complexity
  - **To play with (viewport dependent) video encoding, transmission, and decoding**
  - To play with a viewport dependent projection mapping scheme

VR/360 video encoding and decoding – conventional
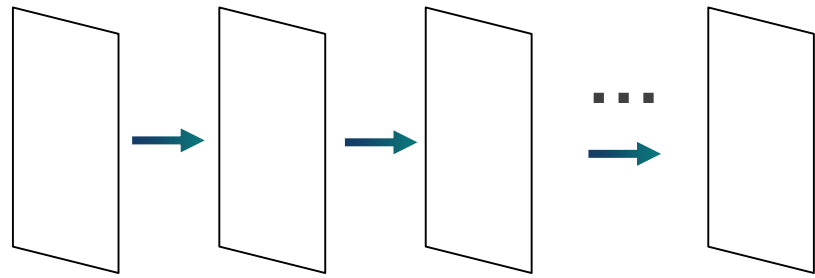
Temporal inter prediction
(TIP)

- Each picture is in a format after a particular type of projection mapping, e.g., equi-rectangular or cube-map
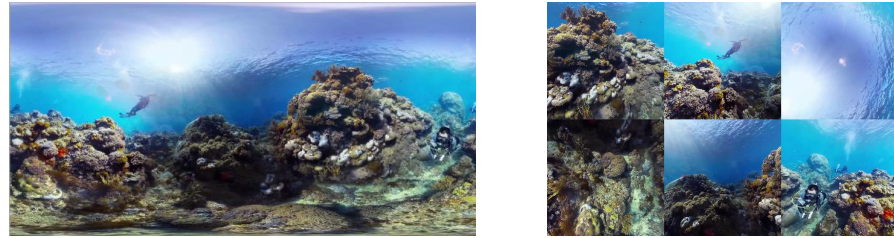
- The video sequence is coded as a single-layer bitstream, with TIP used
- The entire bitstream is transmitted, decoded, and rendered

Simple Tiles based Partial Decoding (STPD)

Encoding and storage

Transmission and decoding

→ Temporal inter prediction (TIP)
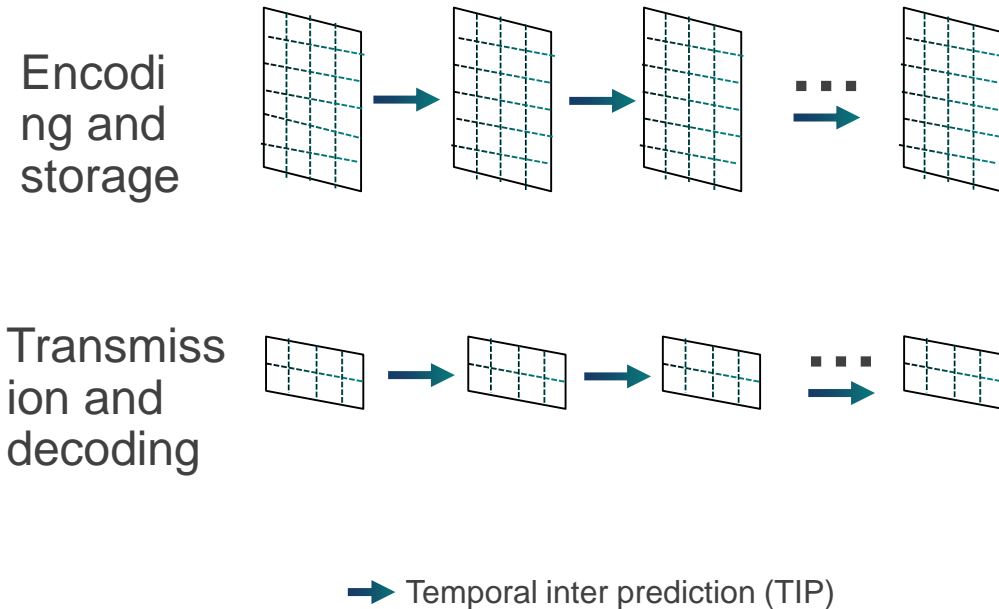
- The video sequence is coded as a single-layer bitstream, with TIP **and motion-constrained tiles** used

- Only a part of bitstream (the minimum set of motion-constrained tiles covering the current viewport/FOV) is transmitted, decoded, and rendered

- Compared to the conventional scheme:
    - Lower decoding complexity (or higher resolution under the same decoding complexity)
    - Lower transmission bandwidth
    - Same encoding and storage costs
    - The latency between user head turning and user seeing the new viewport is really problematic today

Preliminarily bandwidth comparison of SLPD#1 vs Conventional

BD-rate reduction

Bethlehem		Scuba	
Equi-rectangular	Cube map	Equi-rectangular	Cube map
- 19.7%	-23.9%	-27.9%	-31.8%

Scuba (equi-rectangular)

Bethlehem (equi-rectangular)

Bethlehem (cube map)

Scuba (cube map)
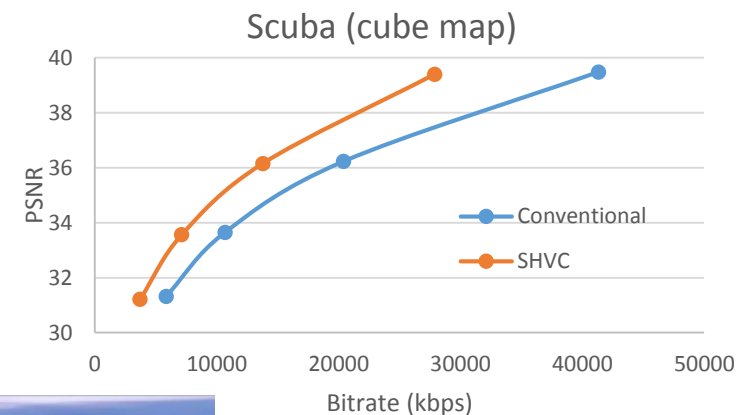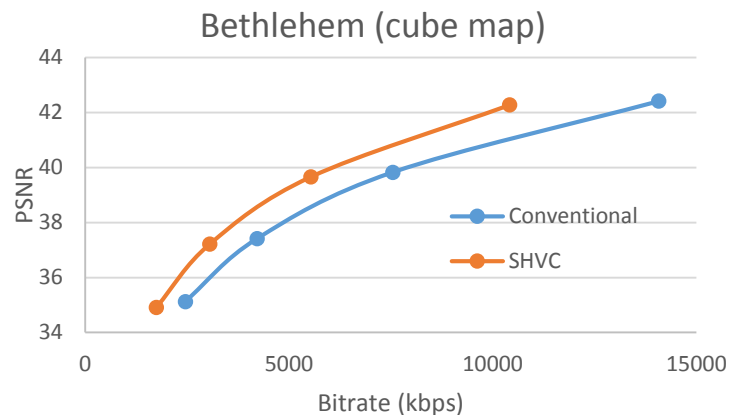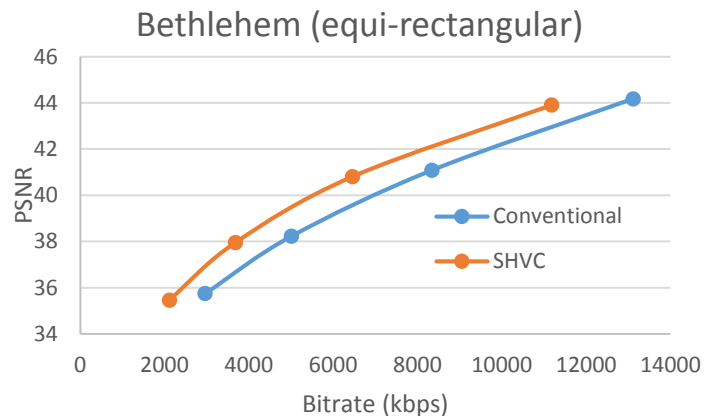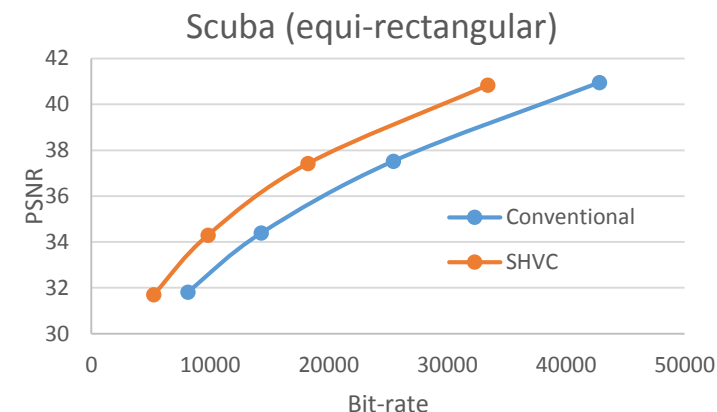
Bethlehem

Scuba

Elements of Phase 2

Studies, Exploration, etc.

Six degrees-of-freedom

- Part of MPEG's vision for *native VR*

- What technology?
  - Light Fields?
  - Point Clouds?
  - Could depend on use case

- May require entirely new video codec (TBD)
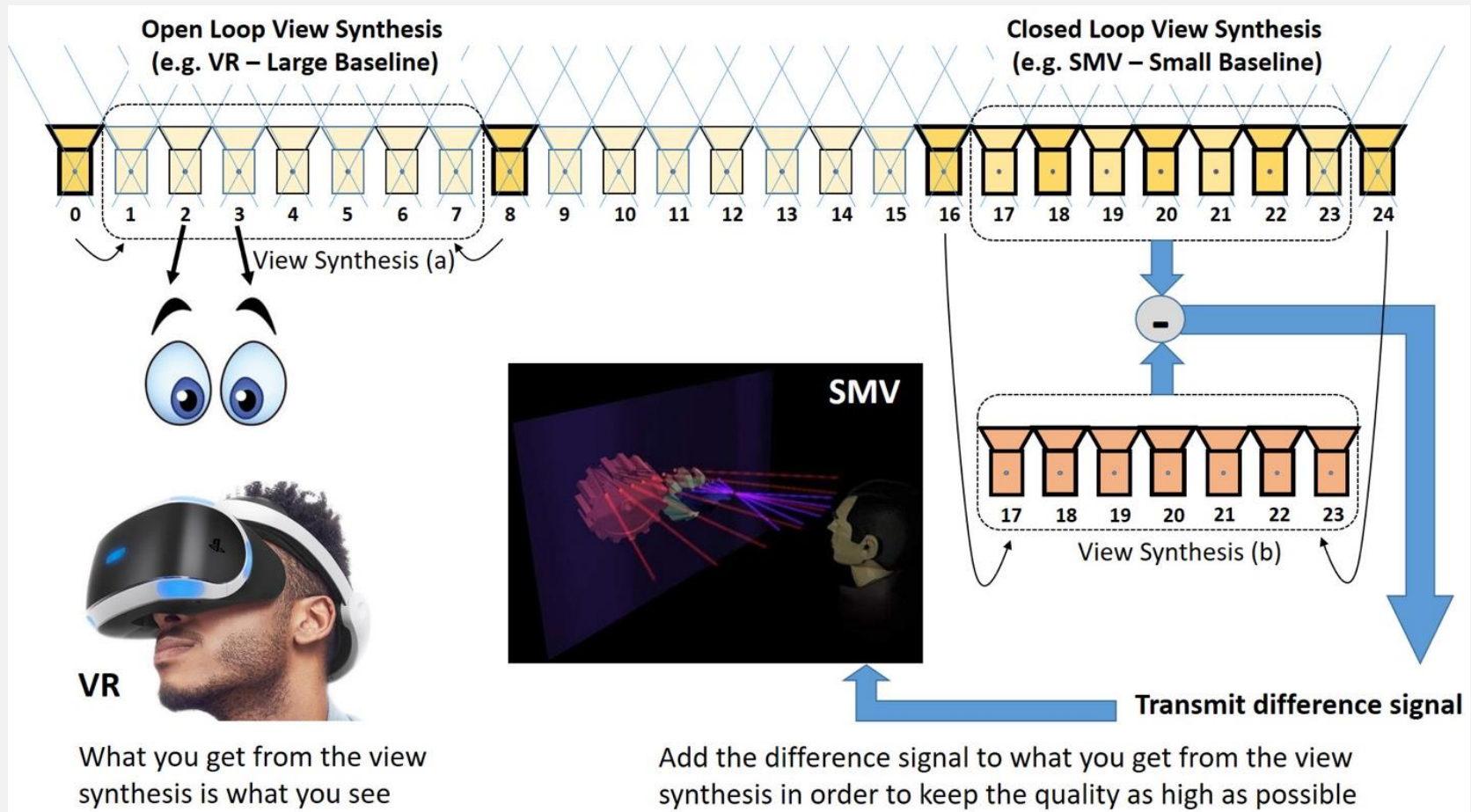
- Point cloud activity already underway
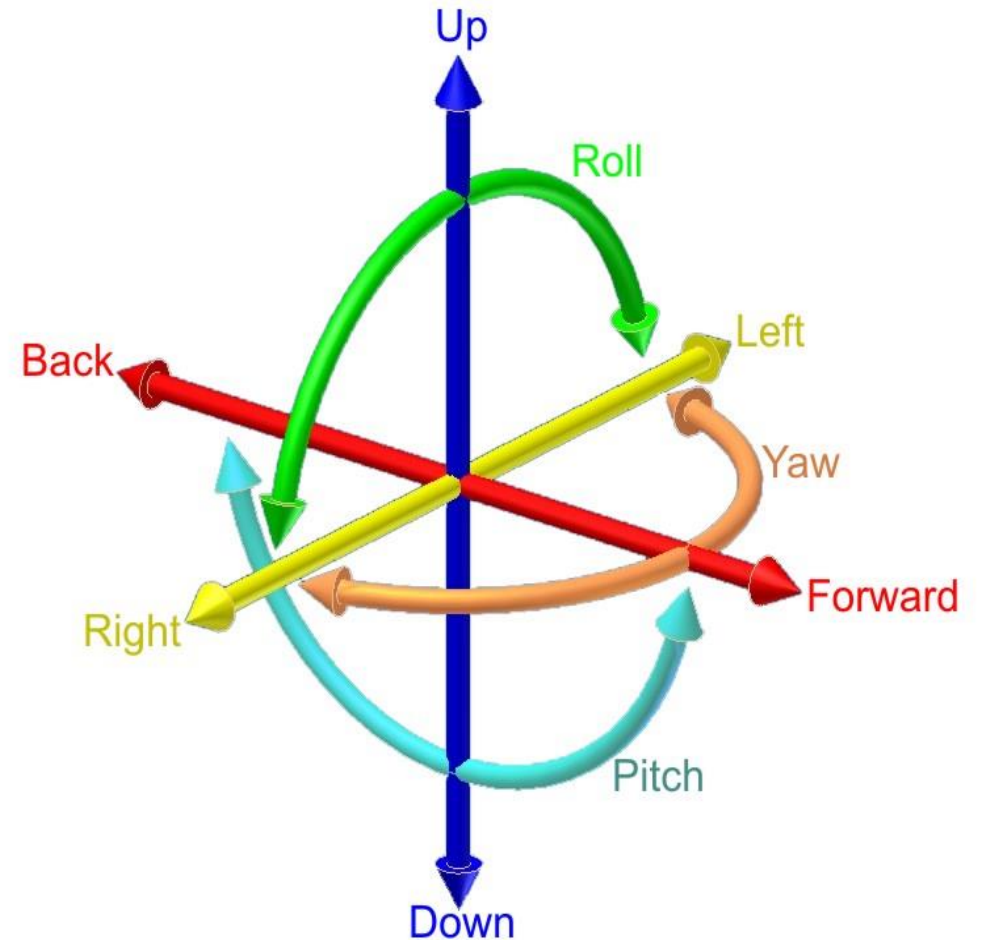
Point clouds
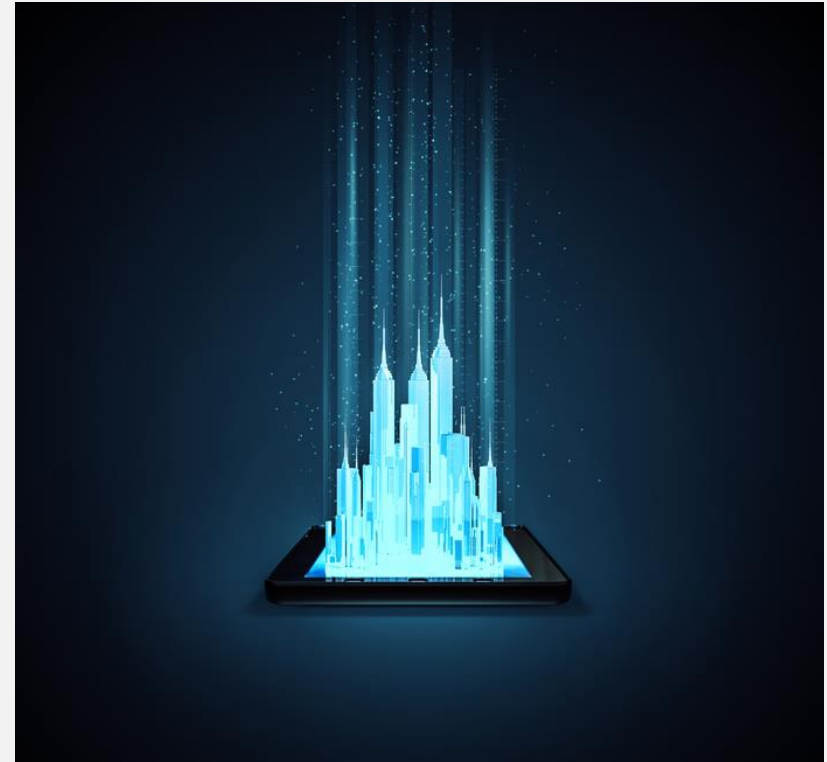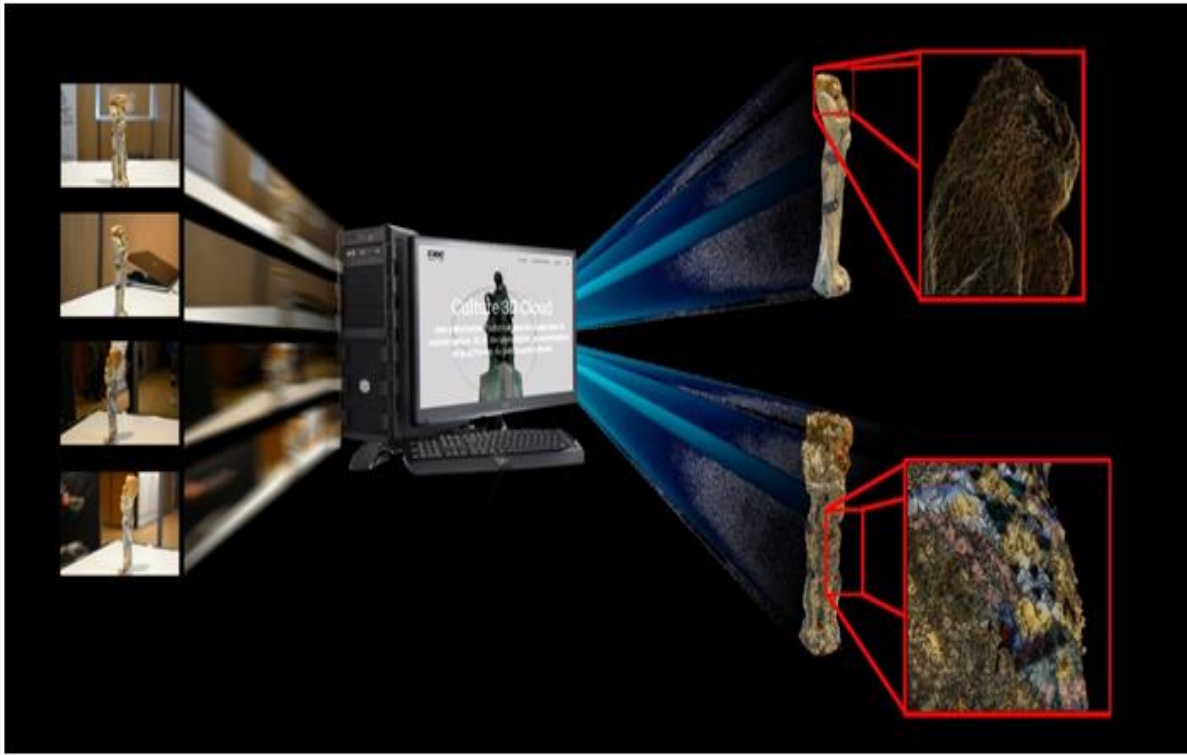
- Natural and computer generated content, 3D meshes

- Efficient compression for storage, streaming, and download
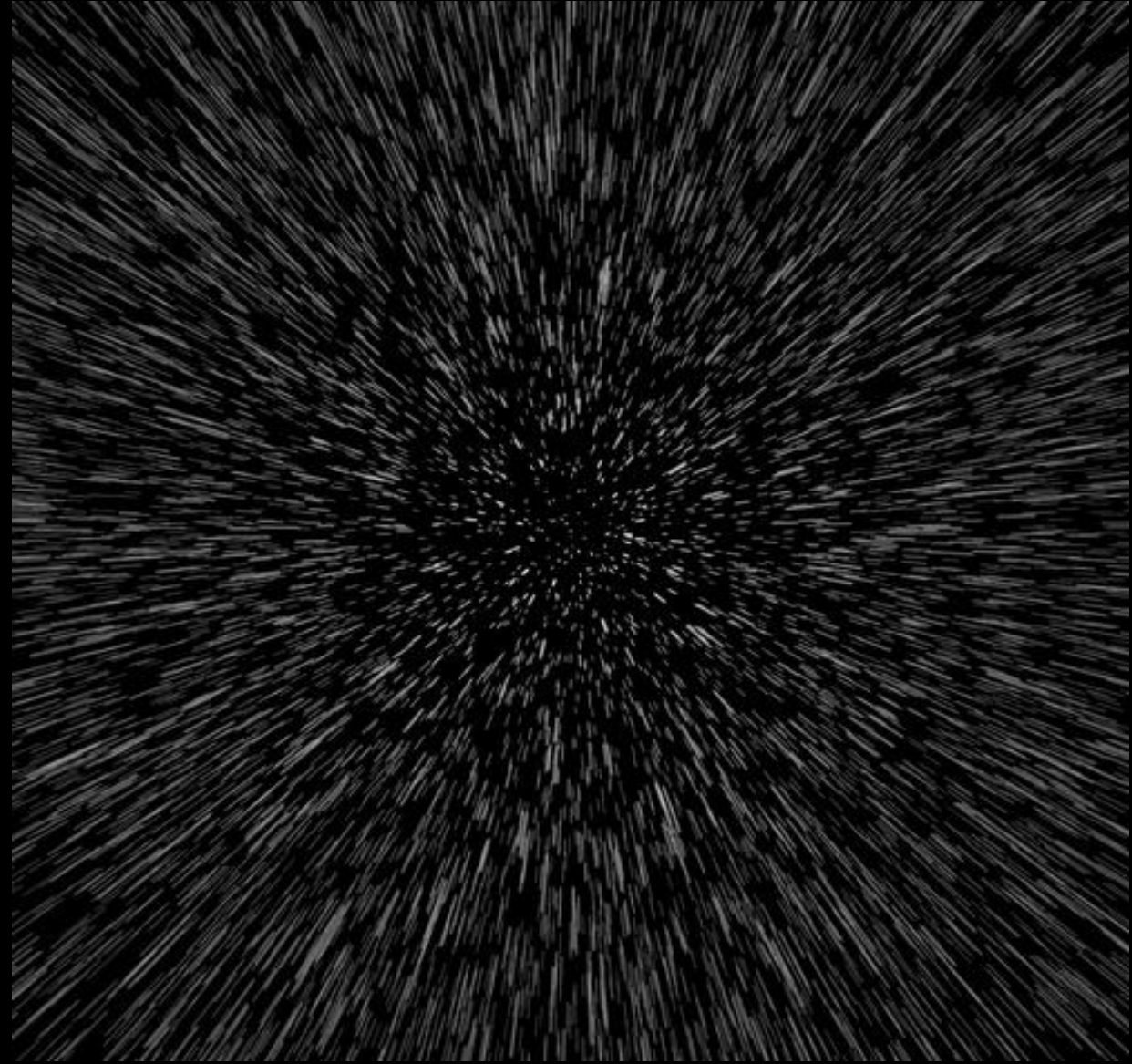
- CfP to be issued Jan 2017

Free navigation

- Capture of converging or diverging views from camera arrays
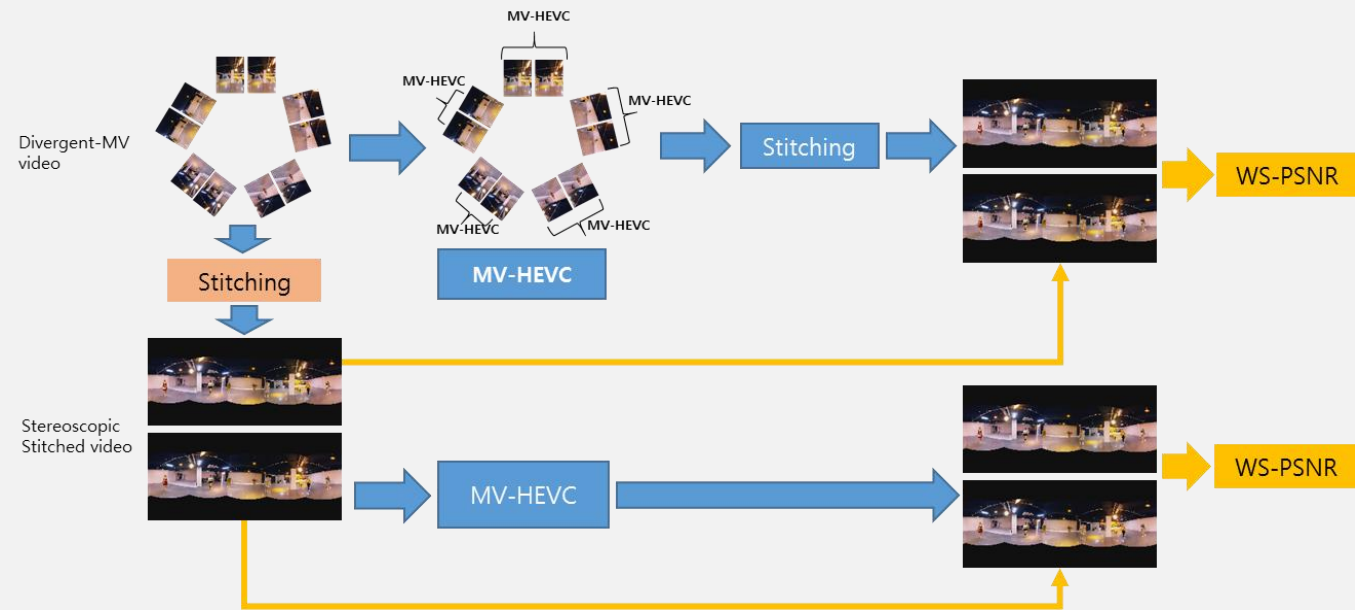- Viewer can freely choose the desired view

Goals and devices for light fields

- Future light field head mounted displays

- Free navigation, e.g. sporting events

- Full coherent parallax

- Super multi-view display or light field display

Next steps for light field video

- Free navigation experiments with various camera array configurations

- Testing with plenoptic video and highly dense camera array video test material

- One solution may be to extend JPEG Pleno's support of static light field images

- Consider viewer fatigue, motion sickness, eye strain, coherent sensory fusion

Thank you