International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.430.3

(08/2018)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Telepresence, immersive environments, virtual and extended reality

# Service scenario of immersive live experience (ILE)

Recommendation ITU-T H.430.3

ITU-T H-SERIES RECOMMENDATIONS

**AUDIOVISUAL AND MULTIMEDIA SYSTEMS**

| | |
|---|---|
| CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS | H.100–H.199 |
| INFRASTRUCTURE OF AUDIOVISUAL SERVICES | |
| General | H.200–H.219 |
| Transmission multiplexing and synchronization | H.220–H.229 |
| Systems aspects | H.230–H.239 |
| Communication procedures | H.240–H.259 |
| Coding of moving video | H.260–H.279 |
| Related systems aspects | H.280–H.299 |
| Systems and terminal equipment for audiovisual services | H.300–H.349 |
| Directory services architecture for audiovisual and multimedia services | H.350–H.359 |
| Quality of service architecture for audiovisual and multimedia services | H.360–H.369 |
| **Telepresence, immersive environments, virtual and extended reality** | **H.420–H.439** |
| Supplementary services for multimedia | H.450–H.499 |
| MOBILITY AND COLLABORATION PROCEDURES | |
| Overview of Mobility and Collaboration, definitions, protocols and procedures | H.500–H.509 |
| Mobility for H-Series multimedia systems and services | H.510–H.519 |
| Mobile multimedia collaboration applications and services | H.520–H.529 |
| Security for mobile multimedia systems and services | H.530–H.539 |
| Security for mobile multimedia collaboration applications and services | H.540–H.549 |
| VEHICULAR GATEWAYS AND INTELLIGENT TRANSPORTATION SYSTEMS (ITS) | |
| Architecture for vehicular gateways | H.550–H.559 |
| Vehicular gateway interfaces | H.560–H.569 |
| BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES | |
| Broadband multimedia services over VDSL | H.610–H.619 |
| Advanced multimedia services and applications | H.620–H.629 |
| Ubiquitous sensor network applications and Internet of Things | H.640–H.649 |
| IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV | |
| General aspects | H.700–H.719 |
| IPTV terminal devices | H.720–H.729 |
| IPTV middleware | H.730–H.739 |
| IPTV application event handling | H.740–H.749 |
| IPTV metadata | H.750–H.759 |
| IPTV multimedia application frameworks | H.760–H.769 |
| IPTV service discovery up to consumption | H.770–H.779 |
| Digital Signage | H.780–H.789 |
| E-HEALTH MULTIMEDIA SYSTEMS, SERVICES AND APPLICATIONS | |
| Personal health systems | H.810–H.819 |
| Interoperability compliance testing of personal health systems (HRN, PAN, LAN, TAN and WAN) | H.820–H.859 |
| Multimedia e-health data exchange services | H.860–H.869 |
| Safe listening | H.870–H.879 |

*For further details, please refer to the list of ITU-T Recommendations.*

# Recommendation ITU-T H.430.3

## Service scenario of immersive live experience (ILE)


### Summary

Recommendation ITU-T H.430.3 identifies service scenarios by analysing several use cases on immersive live experience (ILE) services, in order to classify ILE services and to clarify a reference model of ILE. This Recommendation also summarises several use cases and identifies candidate technologies for implementing ILE, including standards gap analysis related to ILE technologies.



### History

| Edition | Recommendation | Approval | Study Group | Unique ID* |
|---|---|---|---|---|
| 1.0 | ITU-T H.430.3 | 2018-08-29 | 16 | 11.1002/1000/13668 |

### Keywords

Immersive live experience (ILE), service scenario, standards gap analysis, use cases.


* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

## NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

## Table of Contents

## Introduction

In recent times, a number of major sporting and music events have not only been broadcast, but have also been delivered to remote sites for public or live viewing with the aim of sharing the emotional experience with spectators in remote sites as if they themselves were also in main event venue. At the same time ultra high definition (UHD) broadcasting known as 4K (resolution of approximately 4000 pixels) and 8K (resolution of approximately 8000 pixels) broadcasting is becoming increasingly popular. It is therefore to be expected that many people in all over the world will wish to share in the excitement of viewing sporting or musical events on UHD TV or public viewing sites in the near future. However, realistic experience with flat displays has limitations.

In order to provide high-realistic sensations to audiences at remote sites, immersive live experience (ILE) needs to be implemented to reconstruct event sites virtually with presentation of real-sized objects and sound direction by transmitting environmental information together with audio and video streams. For example using ILE musicians could harmonize with player images which are transmitted from several locations. By using ILE, the real objects such as in this case the musicians at the remote sites can collaborate with 3D images of remote musicians which are projected onto the screen. ILE enables virtual concerts with high-realistic sensations.

In order to share the enthusiasm of event venues with large audiences even if they are in remote sites, far from the event venue, implementing ILE services based on standardized designs is desirable. ILE can allow audiences anywhere in the world cheer their favourite sporting teams or musical artists at remote sites even if they are not themselves present in the event venue and they may feel a sense of togetherness and even become passionate just as though they were actually present in the event venue.

This Recommendation specifies several service scenarios of ILE to identify key requirements and architectural frameworks for implementing ILE services and also provides several ILE use cases around the world. This Recommendation also provides candidate technologies for implementing ILE and standardization gap analysis of ILE related technologies.

# Recommendation ITU-T H.430.3

## Service scenario of immersive live experience (ILE)

### 1 Scope

The objective of this Recommendation is to analyse several use cases for immersive live experience (ILE), specify service scenarios and help to identify key requirements and architectural considerations.

The scope of this Recommendation includes:

– Service scenarios

– Use cases of ILE services

– Candidate technologies for implementing ILE

– Standards gap analysis of ILE related technologies

### 2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None.

### 3 Definitions

#### 3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

**3.1.1 immersive live experience (ILE)** [b-ITU-T H.430.1]: A shared viewing experience that stimulates emotions within audiences at both the event site and the remote sites, as if the audience at the remote sites had wandered into substantial event venue and had actually watched the events taking place in front of them. This impression is due to high-realistic sensations provided by a combination of multimedia technologies such as sensorial information acquisition, media processing, media transport, media synchronization and media presentation.

#### 3.2 Terms defined in this Recommendation

None.

### 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

6DOF 6 Degrees Of Freedom

AR Augmented Reality

CI Composition Information

HEVC High Efficiency Video Coding

HMD Head Mounted Device

IPTV Internet Protocol Television

ILE Immersive Live Experience

OMAF Omnidirectional Media Format

UAV Unmanned Aerial Vehicle

UHD Ultra High Definition

VR Virtual Reality

## 5 Conventions

Service scenarios in this Recommendation are described using the following template, in relation with other Recommendations; use cases, requirements, etc.

| **Title-X: high-level service scenario title** | **Sub-title: Specific title of service scenario related to the high-level service scenario** |
|---|---|
| Description | General description for service scenario of ILE |
| Service scenario | General explanation of how a service works |
| Stakeholders (Actors)/Domains | Roles of related stakeholders and domains |
| Main requirements | Service and technical requirements |
| Source (References) | Use cases in Appendix I and/or reference document, websites |

## 6 High-level service category

This clause provides a brief description for the high-level service categories shown in Figure 1. ILE services or use cases can be classified into several categories, such as live sports scenarios, entertainment scenarios, telepresence scenarios, etc.

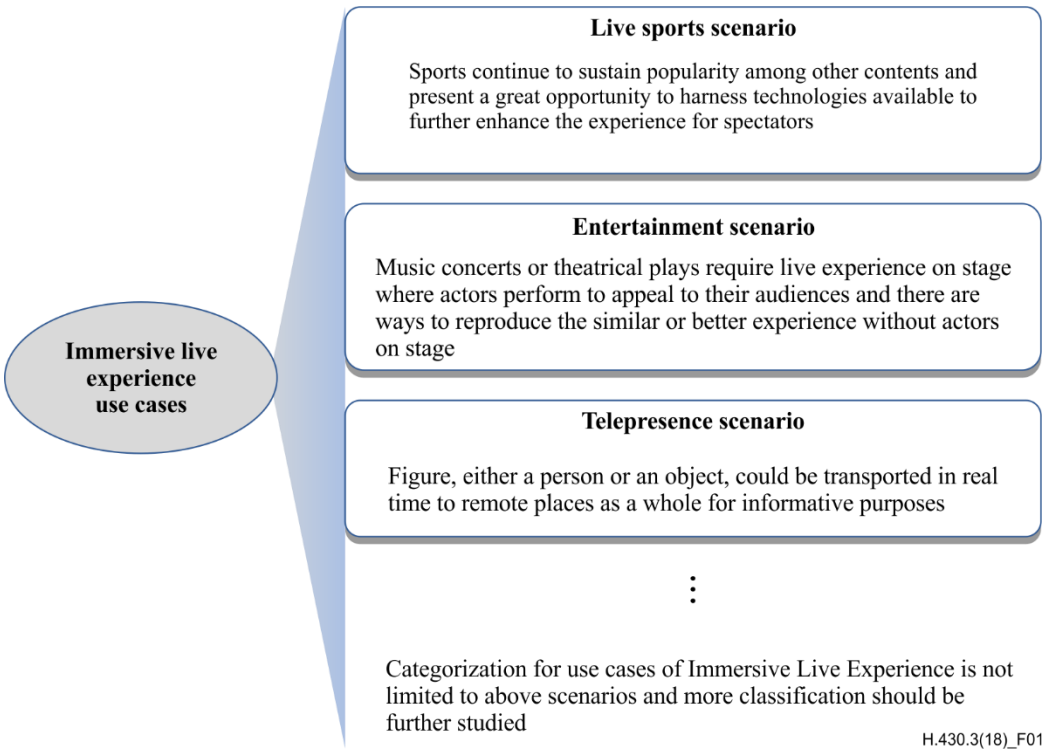



**Figure 1 – High-level service categories of ILE services**

Major categories for use cases of ILE include, but are not limited to, those listed below:

– Live sports scenario: Sport continues to sustain popularity among other contents and presents a great opportunity to harness technologies available to further enhance the experience for spectators

– Entertainment scenario: Music concerts or theatrical plays require live experience on stage where actors perform to appeal to their audiences and there are ways to reproduce a similar or better experience without actors on stage

– Telepresence scenario: A figure, either a person or an object, could be transported in real time to remote places as a whole for informative purposes

## 7 Categorized service scenarios

### 7.1 Live sports scenario

| Live sports | Live broadcast service of first-person synchronous view |
|---|---|
| Description | The service will be provided by transmission of high definition media of athlete's first person viewpoint in real time via a camera installed in sports equipment. |
| Service scenario | A synchronous camera will be installed on the very front of a fast-moving vehicle such as a racing car or bobsleigh where its view will be fixed to face straight ahead on the race course. In other words, the camera will provide the first person synchronous view as if audience is riding the athlete's vehicle.<br>The camera will be equipped with network capability of high bandwidth enough to seamlessly upload its live feed in ultra-high definition to a backend system which will then broadcast to user's devices.<br>The view will be provided via either a big screen or mobile device screen. The additional information related to the race such as speed, acceleration, current |

| Live sports | Live broadcast service of first-person synchronous view |
|---|---|
| | ranking and remaining distance to the finish line and so on can enhance the excitement of ongoing competition.<br>The purpose of the service is to simulate the game-like experience of becoming an athlete on the track in real time. |
| Stakeholders (Actors) | Institutes managing facilities at sports venues, Participating athletes, Producer with broadcasting rights, Network provider |
| Main requirements | Extremely stringent delay, Ultra-high definition codec, Reliable and broadband uplink |
| Source (References) | KT use case in Appendix I |

| Live sports | Live 360° virtual reality service of single point view |
|---|---|
| Description | The service will provide an immersive experience of attending the live event at a sports venue by constructing 360° panoramic view in real time via multiple camera feeds from the site. |
| Service scenario | A camera capable of filming live panoramic views, and/or several high-definition cameras will be installed at the venue such as a gymnastic hall, an ice rink for figure skating events and stadiums for ball games, athletic sports and opening/closing ceremony.<br>The live feed of cameras in multiple angles will be transmitted via an uplink of high bandwidth to a backend system for distribution with minimum latency.<br>In order to view several media, including broadcasting, in proper form, a user might need to put on a special wearable device like virtual reality (VR) goggles or a head mounted device (HMD) which constantly processes and pieces together multiple images to project the real world. For this purpose, it could be used in pseudo 3D displays without special wearable devices such as VR goggles or HMD.<br>The purpose of the service is to enable viewers to observe every event live around the spot so that they feel as if they are present at the actual remote location. |
| Stakeholders (Actors) | Institutes managing facilities at sports venues, Participating athletes, Producer with broadcasting rights, Network provider, VR manufacturer |
| Main requirements | Extremely stringent delay, Ultra high definition codec, Reliable and broadband connection, Massive connectivity, VR media processing (e.g., image stitching) |
| Source (References) | NTT use case (Kirari!) in Appendix I<br>KT use case in Appendix I |

| Live sports | Live broadcast service of selected multi objective view |
|---|---|
| Description | The broadcasting service will provide user selected view of live game action in high definition from among other preinstalled cameras at the venue with the aid of position information of athletes on a 3D map. |
| Service scenario | A number of cameras will be installed at strategic locations along the race course where live action of athletes can be best captured simultaneously.<br>With the aid of a 3D map interface, viewers can select any point of interest on the map at any given time for better angles of an action. Every competing athlete will be tracked in real time and their corresponding locations will be indicated on the map all the time. |

| Live sports | Live broadcast service of selected multi objective view |
|---|---|
| | The purpose of the service is to provide omniscient viewpoint on all corners of the race course to better track favourite athletes in action. |
| Stakeholders (Actors) | Institutes managing facilities at sports venues, Participating athletes, Producer with broadcasting rights, Network provider, Outdoor positioning information provider |
| Main requirements | Extremely stringent delay, Ultra high definition codec, Reliable and broadband uplink, Massive connectivity, Location based services (e.g., 3D Map, Precise positioning) |
| Source (References) | KT use case (http://www.koreaherald.com/view.php?ud=20180213001003) |

## 7.2 Entertainment scenarios

| Entertainment | On-stage holographic performance show service |
|---|---|
| Description | The service will provide holographic media of music concerts that are live-streamed or pre-recorded for audiences at special theatres with capability of reconstructing the live concert environment. |
| Service scenario | A live concert performance is captured or pre-recorded by several ultra-high definition cameras in multiple angles and using high quality microphone systems. Then, ultra-high definition video and audio of high quality will be processed separately at the initial stage. Processed media streams including video, audio and other sensor data should be transmitted synchronously from the event site to remote sites.<br>A theatre similar to one set up for original concert hall event site will be reconstructed at remote site for high realistic effects where the processed media stream will be regenerated to simulate the live event. Audience at such remote site can experience almost the same visual and sound quality as the original on holographic or pseudo-3D projection.<br>The purpose of the service is to evoke the sense of concert like environment which is on a par with real concert but without actual performers on stage. |
| Stakeholders (Actors) | Concert performers, Concert producer, Network provider, Hologram equipment provider |
| Main requirements | Ultra-high definition codec, Hologram conversion, On stage hologram or pseudo-3D projection, Video/audio mux |
| Source (References) | NTT use case (Kirari!) in Appendix I<br>KT use case in Appendix I |

| Entertainment | Live interactive service |
|---|---|
| Description | The service will provide virtual scenes collaboration with real person and transmitted images in real time. The image will be extracted from real person in different places and will be transmitted in real time high definition media. The transmitted images and sounds will be reconstructed and interact with real person in remote site. |
| Service scenario | The live video images of the artists in event venues are extracted in real time and transmitted in various media formats synchronously over high-speed networks with low latency. In viewing sites, transmitted images and sounds are dynamically projected and mapped on multiple transparent screens which are spatially set up.<br>A real person, such as a disc jockey (DJ) or master of ceremony (MC), in the |

| Entertainment | Live interactive service |
|---|---|
| | remote site can collaborate and interact with artists in event venues by transmitted images.<br>The purpose of this service is to provide the audience in a remote site with a new immersive experience as if they are watching the live stage of a DJ interplaying with the artists in the event venue who are teleported to remote sites beyond time and space. |
| Stakeholders (Actors) | Institutes managing facilities at event venue, Participating artists, Producer with broadcasting rights, Network provider |
| Main requirements | Extremely stringent delay, Ultra-high definition codec, Reliable and broadband uplink and downlink |
| Source (References) | NTT use case in Appendix I |

| Entertainment | Multi-angle viewing service |
|---|---|
| Description | The service will provide virtual psudo-3D images onto a special arena, which can be viewed from any angles and reconstruct sound direction. The images are captured by several cameras at the event site and are transmitted with audio and spatial information synchronously to remote sites. At the remote sites, the images and sound are reconstructed onto a special arena. The users can simultaneously watch the images from their angles. |
| Service scenario | The images of artists or sport players at an event site are extracted by using several cameras and the sounds of them are also captured by using several microphones. In addition, 3D sensing information and object tracking information are also captured for fixing spatial locations. These video and audio streams are synchronously transmitted to the viewing site with spatial information via high-speed networks.<br>At the viewing site, these images and sounds are reconstructed onto a special arena, which has multi-screens and arrayed speakers.<br>The audiences at viewing sites can see images in any angles outside the special arena and can hear sounds from images. In order to realize image or auditory localization, pseudo-3D image displaying and wave field synthesis might be used in a special arena.<br>The purpose of this service is to provide experiences of real events which are reconstructed images and sounds onto a special arena in real time. By using this service, several audiences can watch the event at viewing sites such as sport bars. |
| Stakeholders (Actors) | Institutes managing facilities at event venue, Participating artists, Producer with broadcasting rights, Network provider |
| Main requirements | Extremely stringent delay, Ultra-high definition codec, Reliable and broadband uplink and downlink, Psudo-3D display, Wave field synthesis |
| Source (References) | NTT use case in Appendix I |

## 7.3 Telepresence scenarios

| Telepresence | Live ultra high definition presence video view service |
|---|---|
| Description | The service will provide real time holographic presentation of a person of interest in ultra high definition with augmented information or special effects around the figure for audiences at viewing sites. |

| Telepresence | Live ultra high definition presence video view service |
|---|---|
| Service scenario | A single person of interest such as someone being interviewed or highlighted will be captured in real time by special holographic camera.<br>Once high definition video and digital audio are encoded, they will be synchronously transmitted via ultra high speed network to a viewing site where the holographic video and audio are decoded and reproduced in real time.<br>Various special effects like animation or information of many formats could be augmented onto an original object or person for amusement or marketing/promotion. For a better feel of reality, projection of the person/object in real size is preferred.<br>The purpose of the service is to provide ILE of telepresence of person/object in real size which could be further enhanced by addition of special effects for good measure. |
| Stakeholders (Actors) | A person of interest, Producer with broadcasting rights, Hologram equipment provider, Data analytics provider, Network provider |
| Main requirements | High definition codec, Hologram conversion, Hologram projection, high definition audio, Video/audio mux, Converter, Reliable and broadband connectivity |
| Source (References) | KT use case (https://www.youtube.com/watch?v=cJgo-b4aEjE )<br>NTT use case (https://www.youtube.com/watch?v=E0D67sBsCY0) |

## 7.4 Controller scenarios

| Controller | Live object controller interaction |
|---|---|
| Description | The service provides control of real objects in virtual reality so that users can experience additional senses of touch and control, as a result providing an increased degree of realism. |
| Service scenario | Certain dimension of space is set up at a real world location where users can physically pilot through virtual reality. The boundaries of the space are marked by sensors installed on the corners. Inside the setting defined within the boundaries, selected physical object(s) is (are) placed. Such objects are scanned and digitally encoded in advance or in real time to create an exact replica for the virtual reality. With this physical object(s), user can navigate and manipulate the virtual world with high degree of ease and comfort.<br>This virtual reality is coordinated in a central location, distributed over a network of high bandwidth and low latency and fed into HMD in real time for multiple users at different places where they interact with each other as if they are all in the same venue. Especially with such items under control, users will interact with virtual reality in far more immersive and live manner.<br>The purpose of the service is to provide users with a great degree of manoeuvrability in a virtual world via physical object(s), further enhancing immersive and live aspects of user experience. |
| Stakeholders (Actors) | VR producer, Government and travel agency for scenery rights, Sports association, HMD and sensor equipment provider, Network provider |
| Main requirements | VR production, Object encoding and coordination, Ultra high definition codec, Reliable and broadband downlink, Extremely stringent delay, HMD and VR sensors |
| Source (References) | TEKTONSPACE use case (https://www.youtube.com/watch?v=EUDTTLNLfDo) |

## 7.5 Meeting scenarios

| Meeting | Immersive live meeting |
|---|---|
| Description | The service provides teleconferencing where participants from remote sites will be represented in high-realistic images to each other in order to induce an immersive experience of being present in the same meeting room. |
| Service scenario | Each participant's image will be captured by cameras, then high resolution video and audio will be transmitted over extremely low latency and high bandwidth networks. Each participant's high-realistic image will be distributed to all others in real time.<br>Users can view the scene at the reconstruction end and feel themselves in a fully immersive meeting environment, where they can interact with people's real figures.<br>The purpose of this service is to provide a portrait capture and display service for video conference, board games, etc. |
| Stakeholders (Actors) | Business discussant, Image capture equipment provider, Network provider, Game producer. |
| Main requirements | Image capture equipment, Ultra high definition codec, Reliable and broadband uplink and downlink, Virtual or holographic display device |
| Source (References) | OKI use case in Appendix I |

## 7.6 Emergency scenarios

| Emergency | Live streaming service in inaccessible areas |
|---|---|
| Description | The service will provide real time reconstruction of scenes in inaccessible areas which cannot be reached by humans immediately such as post-disaster environments, polar regions or inaccessible mountains. |
| Service scenario | The purpose of this scenario is to provide immersive experience for inaccessible areas. By using robots and unmanned aerial vehicles (UAVs) carrying microphones, holographic cameras, environmental sensors in position or by using pre-installed sensors, ultra high definition video and audio will be processed and captured.<br>In close proximity of these areas somewhere that people can reach, a high-bandwidth and low-latency wireless network coverage needs to be provided, to transmit high quality holographic data.<br>These scenes are reconstructed in real time at the viewing site (possibly the disaster relief command centre or the polar experience museum). In addition to holographic video and high quality audio projection abilities, viewing sites also require extreme conditions manufacturing capabilities, such as a challenging temperature and humidity simulator for human tolerance |
| Stakeholders (Actors) | Government, Travel producer, Hologram equipment provider, Network provider. |
| Main requirements | Capture device, Environment analysis and position assign system, Ultra high definition codec, Reliable and broadband uplink, hologram projection |
| Source (References) | NA |

# Appendix I

## Use cases for immersive live experience (ILE) services

(This appendix does not form an integral part of this Recommendation.)

This appendix provides several use cases for immersive live experience (ILE) services and identifies candidate technologies for implementing ILE. In addition, this appendix summarizes standards gap analysis related to ILE technologies.

### I.1 Introduction

This appendix provides several use cases for immersive live experience services. ILE is a completely new service, so it is important to create common understanding through identifying ILE use cases. In addition, this appendix identifies standardization gaps within ITU-T and outside ITU-T, since several key technologies may already be standardized in ITU-T and/or in other SDOs.

This appendix includes:

– Use cases of ILE services
– Configuration of ILE
– Candidate technologies for implementing ILE
– Standards gap analysis of ILE related technologies

### I.2 Use cases of ILE services

#### I.2.1 Use case – Kirari! (NTT)

Nowadays, big events such as worldwide sports events like the soccer world cup are not only broadcast on live television but can also be watched on public viewing or live viewing sites for sharing emotions with audiences. The watching styles become more diversified. It is expected that sharing emotions with audiences will be increased by high-realistic televisions and public viewing sites using 4K and 8K technologies.

In order to provide ultra high definition content and high realistic sensations for audiences in remote sites, NTT has been working on research and development of immersive telepresence technology called "Kirari!" Kirari! consists of real time and synchronous media transmission, ultra high definition media codecs, spatial and environmental information transmission, etc. As shown in Figure I.1, Kirari! enables audiences who can be anywhere in the world to feel high-realistic sensations as if they had wandered into a major event venue. At the event site, multiple cameras and sensors capture information about objects and the environment and the information is sent to remote sites with some media transformation in real time. At the remote sites, virtual images reconstruct the atmosphere of the event site using time synchronized media such as 3D projection, surrounding audio and lighting. Audiences at remote sites can share emotions with spectators at the event site, and they can see their favourite stars with their own eyes.

H.430.3(18)_FI.1

**Figure I.1 – Concept of Kirari!**

### I.2.2 Use case – High quality video for VR streaming (Fraunhofer HHI)

One of the critical elements to produce ILE effectively is to capture the 360° surrounding environment of a source venue. Several camera systems could be used to deliver such VR video content such as direct stitching by using mirrors, depth-aware stitching and depth-enabled light-field rendering.

Multi-view 3D reconstruction enables a 3D image of a real person or object to be projected onto computer generated background and provides immersive experience of being present in the virtual surrounding.

Another sophisticated technology, tile-based video streaming, could be employed where many tile-based images are encoded using high efficiency video coding (HEVC) and then stitched together to render a cubic-shaped virtual environment.

### I.2.3 Use case – 3D interactive experience (B<>Com)

With advanced technologies using 6 degrees of freedom (6DOF) VR reconstruction of the surrounding environment, sparse camera arrays, light-field of representation and compression and point cloud, a 3D interactive service delivers a virtual environment where single or multiple users can interact with each other or computer generated 3D objects in an ultrarealistic way. This service will provide users with the ultimate natural feeling of virtual presence.

### I.2.4 Use case – Ultra-realistic telework and 4K/8K broadcast (OKI)

High resolution displays along with high definition sound systems transform separate working places into one integrated immersive office where people can collaborate as if they are right next to each other.

The general public is able to watch big sport events such as the Olympic/Paralympic games which will be broadcasted in extremely high definition of 4K/8K. This service will provide viewers with much crisper and more accurate sports events and produce immersive live experience even in a 2D setting.

### I.2.5 Use case – High quality cloud-based 360° rendering and streaming (Fraunhofer FOCUS)

Cloud-based 360° video playout enables high quality 360° experience on low capability devices such as hybrid TVs and mobile devices. By moving a simple remote control, video will adjust the angle of a scene accordingly in real time, providing some degree of immersive experience. The video will be produced and rendered on a backend system upon signals received by remote control movement.

### I.2.6 Use case – Multi-channel 4K/8K broadcast service (NHK)

Live video of 8K quality will be distributed over satellite broadcast with multiple videos of high quality up to 4K supplemented over broadband in order to provide a high degree of excitement on live actions.

### I.2.7 Use case – Immersive live experience over 5G technologies (KT)

5G technologies of high bandwidth and low latency will provide the robust infrastructure on which immersive live experience will be featured for a batch of 5G services.

A high speed bobsleigh equipped with a high resolution camera will provide first person synchronous view in real time over 5G radio technologies augmented with the athlete's various physical status information points such as heart rate to give an immersive feel of what the athlete sees and feels.

360° VR/TimeSlice service lets audiences with HMD absorb a sense of immersive presence with access to all angle views around the venues and through mobile devices, also a 360° peripheral view of any athlete selected by the viewer allowing a close up monitoring of events such as ice hockey tournaments or figure skating.

Hologram Live will teleport a real size image of a person or object from remote sites setting up a virtual environment where one conducts a live interview in person. Multi-site hologram service utilizes hologram technologies to carefully integrate two live hologram images from different venues into one place to coordinate or choreograph multi-party activities as if everyone is present at the same spot.

### I.2.8 Use case- Interaction with images (NTT)

On March 13, 2017, Music Live Showcase "CYBER TELEPORTATION TOKYO at SXSW" was presented as a remote interactive live performance between Tokyo and Austin, Texas.

Featuring a Japanese electronic music producer and DJ in Austin, and Japanese artists in Tokyo, NTT showed a live performance experience "CYBER TELEPORTATION TOKYO at SXSW" (https://www.sxsw.com/news/2017/cyber-teleportation-tokyo-at-sxsw/), in which the audience experienced an immersive synchronized performance of DJ in Austin and the artists "teleported" from Tokyo.

The live video images of the artists in Tokyo are extracted in real time and transported in various media formats synchronously over the network, then dynamically projected on multiple double-sided transparent screens which are spatially set up in the Austin live space (see Figure I.2) by utilizing NTT's Kirari! technologies. The latency is about 400 ms, so the interaction could be sufficient.

The network infrastructure used for this showcase was established using NTT's network testbed GEMnet2, in cooperation with Internet2 (a consortium of US universities providing an advanced networking environment), LEARN (Lonestar Education and Research Network) and the University of Texas System and Greater Austin Area Telecommunication Network.

These technologies provided the audience in Austin with a new immersive experience as if they were seeing the live stage of the DJ interplaying with the artists in Tokyo who are teleported to Austin beyond time and space.

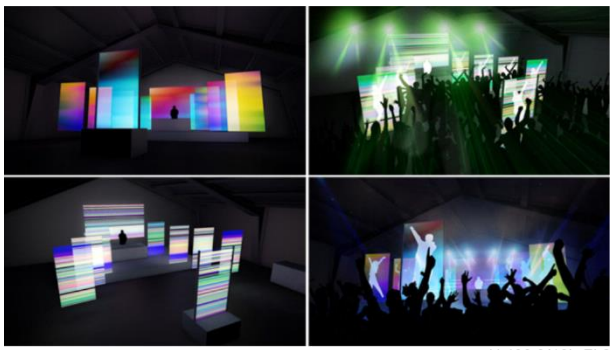



**Figure I.2 – CG Images of the stage at Japan Factory in SXSW 2017**

### I.2.9 Use case- multi-angle viewing styles (NTT)

Capturing atmosphere by several cameras at an event site and transmitting images to all over the world, enables the atmosphere of the event site to be reconstructed at viewing sites. NTT developed a new viewing style that audiences who encompass a reconstructed event site at viewing sites can watch the same event all together. The developed new viewing style was called "Kirari! for Arena", which realized the reconstruction of the atmosphere of the event site for watching the event from several angles by many audiences all together at the viewing sites, by capturing object images with location information at event site. Figure I.3 shows the image of "Kirari! for Arena" at a Judo event.

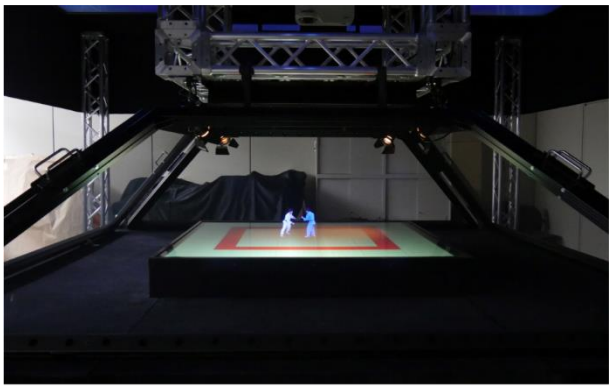

**Figure I.3 – Images of Kirari! for Arena on Judo**

3D sensing and object tracking technologies are key functionalities for realizing "Kirari! for Arena". In order to reconstruct the atmosphere of an event site at viewing sites, it is required to capture the 3D position information of objects and to track objects. The combination of object recognition using deep learning, object tracking using particulate filters and depth sensors using lasers enables roughly tracking the location of objects. The location information enables reproduction of natural motion in depth in addition to horizontal direction at viewing sites.

Another key functionality is pseudo-3D image localization which reconstructs the atmosphere of an event site at viewing sites is the displaying of natural motion of objects even in depth direction. This also enables many audiences to observe a displayed object from any angles in same time. Pseudo-3D image localization can reconstruct event venues with moving objects at viewing sites, so that, dynamic displaying achieves spatial position relations as if players move around the fields in front of audiences.

By using a simple display device, as shown in Figure I.4, audiences can watch objects displayed as pseudo-3D images without recognition of the display devices. The image on the horizontal plane is a projected pseudo-3D image on the displayed floor at viewing sites by reflecting images at an

angled half mirror, which transmits some incident light and also reflects the rest. In this case, it is difficult to move images to depth direction since the pseudo-3D images are fixed on the same location. In order to display for moving objects to depth direction, it is required to utilize visual effects and psychological implication.



H.430.3(18)_FI.4

**Figure I.4 – Simple display device using four half mirrors**

## I.3 Configuration of ILE

In the case of transmitting a soccer game to gymnasium halls for example, it is clear that there is a big difference in the sizes of venues as well as the sound and lighting equipment. Figure I.5 shows the configuration of Kirari! as an example of an ILE service, in order to reconstruct the atmosphere of the soccer stadium to a remote site. The following approaches can achieve high-realistic sensations as if audiences in the gymnasium were entering the soccer stadium:

1) Objects such as soccer players are measured in X-Y-Z coordinates and are extracted images in real-time at the stadium. In order to reconstruct the extracted objects at the gymnasium hall, the information of the object is transmitted in real-time.

2) Video, audio and lighting information except for objects described above is also transmitted to gymnasium halls after transforming information based on the data of the hall environment such as the width and height of screens, the number of projectors, the types of audio equipment and characteristics of the lighting equipment. This transforming function calculates parameters from the hall environmental data to reconstruct the atmosphere of the stadium at the remote sites. For example, the following are the actual parameters: i) X-Y-Z coordinate of objects, ii) range of captured environmental and spatial information except for objects, iii) region of sound space.

3) At the remote sites, environmental information is reconstructed by projection mapping technology. Then another projector shows objects in 3D shapes and numerous speakers reconstruct the audio space and auditory lateralization.

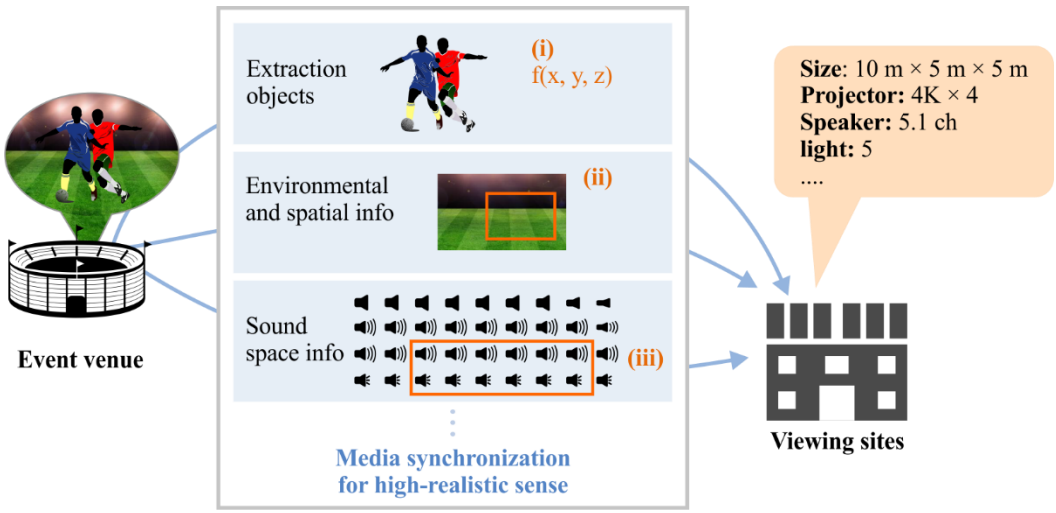



**Figure I.5 – Configuration of Kirari!**

As shown in Figure I.6, audiences in the remote sites can feel ultra-high-realistic sensations by pseudo-3D projection and auditory lateralization of Kirari! Audiences' experience emotions as though they are watching actual events, exceeding those provided by 2D displays. Kirari! provides new experiences for audiences.

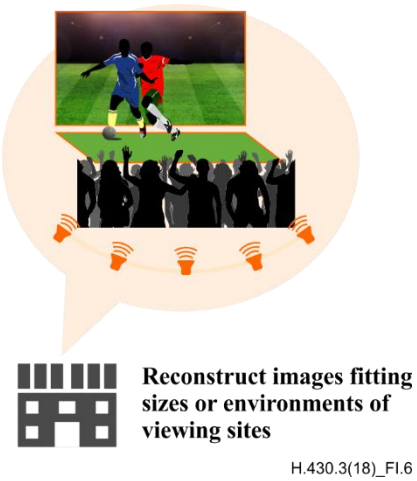



**Figure I.6 – Reconstruction atmosphere of event venue at viewing site**

## I.4 Candidate technologies used for ILE

Candidate technologies used in ILE which aim to realize pseudo 3D presentations in real time are described below and include high-realistic media synchronization, real time object image extraction, high-realistic auditory lateralization and extra-wide video stitching technologies:

(1) High-realistic media synchronization

The MPEG media transport (MMT) technology has features which synchronize video and audio streams by using absolute time management. However, it is necessary to synchronize several elements such as lighting and spatial information to reconstruct a virtual field of the event venue at remote sites. Signalling descriptor of MMT is required to model spatial information such as size of objects, positional relations and direction of sounds.

By using this technology, it is possible to handle asset information related to physical spatial position information such as the size of screen, resolution of displays and spatial position. Using the transmitted spatial size of images and auditory lateralization information together with multiple audio and video streams constituting the event venue, a

virtual field of the event venue can be reconstructed with high-reality at the viewing site, depending on the specifications of each viewing site.

(2) Real-time object image extraction

Real-time object image extraction is a key technology to realize live experience, and extracted objects are displayed pseudo-3D images at viewing site. This may be achieved by a combination of contour detection technology using sensor information such as distance and temperature and image processing technology which identifies boundaries of target objects in high-speed and high-resolution. Real-time image extraction can clip purely the target object to be presented as a pseudo-3D image at the viewing site.

(3) High-realistic auditory lateralization

High-realistic auditory lateralization aims to reconstruct a virtual acoustic field onto a wide viewing area. This enables sound direction from real-sized objects projected onto wide screens to be reproduced, thus audiences can feel voice and sound as if the projected objects emitted a sound.

(4) Extra-wide video stitching

The video stitching technology which stitches processed video images of multiple high definition cameras, and adjusting images, enables high-realistic field images to be reconstructed in real time for major sports venues and wide fields such as athletic sports.

## I.5 Standards gap analysis of ILE related technologies

### I.5.1 ITU-T work

There exist several Recommendations related to Internet protocol television (IPTV) services. For example, IPTV architecture, [b-ITU-T Y.1910], defines the reference model and function blocks. However, there are no specifications about immersive services in those IPTV related Recommendations.

### I.5.2 ISO/IEC JTC1 SC29 WG11 (MPEG) work

ISO/IEC JTC1 SC29 WG11, MPEG, has been working with ITU-T SG16 in the video codecs for many years. MPEG has developed media transport technologies called MPEG media transport (MMT).

**ISO/IEC 28008-1**

[b-ISO/IEC 28008-1], which was developed in MPEG, defines an MMT overview. This standard extends existing media transport technologies such as MPEG2-TS [b-ITU-T H.222.0] and specifies multi-source content delivery. This enables synchronous media transmission between video and sound, but transporting special environment information is not considered in this standard. In other words, MMT cannot resolve synchronous media transportation.

**ISO/IEC 28008-11**

MPEG has also developed [b-ISO/IEC 28008-11], which is a composition information (CI) specification of terminal devices. Displaying real-sized object needs to be considered as CI, however, this standard does not cover spatial environment such as bezel size of display devices and spatial position information. Therefore, CI is inefficient for displaying real-sized objects.

**MPEG-I Project**

From the systems perspective, MPEG-I project has laid out a basic foundation of coding, signalling, encapsulation and delivery formats for both mono and stereoscopic 360-degree video and audio in omnidirectional media format (OMAF) at its initial phase. However, a rich set of features in immersive media including 6DOF, point clouds, light fields that could enhance new use cases is set to be further explored in the next phase.

### I.5.3 Work in broadcasting area

In the broadcasting standardization area, standardization of media transport technologies has been started.

**ARIB**

The Association of Radio Industries and Business (ARIB) created the media transport standard, B-60 [b-ARIB STD-B60], based on MMT, [b-ISO/IEC 28008-1]. This standard aims to enhance terrestrial broadcasting services. However, as shown in clause I.6.2, this standard does not specify synchronous media transportation.

**ITU-R**

ITU-R has started considering new media transport technologies based on ARIB B-60. This work has not been finalized yet.

### I.5.4 Work in cinema area

Today, 4D cinemas which provide high-realistic sensation for users have been developed and deployed all over the world. The 4D means that user can sense several feelings such as vibration of seats, lighting and smell in addition to 3D video and surround sound. 4D cinemas can only be provided at specific theatres which are equipped according to special devices, with special programs that are designed for 4D. Moreover, 4D cinema does not have live content transmission capability.

On the other hand, there exist several 4D cinema technologies such as 4DX and MX4D, but there are no standards so far.

Therefore, ILE services could not be covered by this area.

### I.5.5 Standardization gaps

Standardization of the terminal issues provides decreasing terminal device cost of high-realistic IPTV services and standardization of the transmission issue provides interoperability between service providers and content providers.

As mentioned above, all key features of ILE services have not been progressed on any standardization work, so it is necessary to start the standardization work in ITU-T.

# Bibliography


[b-ITU-T H.222.0] Recommendation ITU-T H.222.0 (2006), *Information technology – Generic coding of moving pictures and associated audio information: Systems*.

[b-ITU-T H.430.1] Recommendation ITU-T H.430.1 (2018), *Requirements for immersive live experience (ILE) services*.

[b-ITU-T H.430.2] Recommendation ITU-T H.430.2 (2018), *Architectural framework for immersive live experience (ILE) services*.

[b-ITU-T Y.1910] Recommendation ITU-T Y.1910 (2008), *IPTV functional architecture*.

[b-ISO/IEC 28008-1] ISO/IEC 28008-1 (2014), *Information technology — High efficiency coding and media delivery in heterogeneous environments —Part 1: MPEG media transport (MMT)*.

[b-ISO/IEC 28008-11] ISO/IEC 28008-11 (2014), *Information technology — High efficiency coding and media delivery in heterogeneous environments —Part 11: MPEG Composition Information*.

[b-ARIB STD-B60] ARIB STD-B60 (2015), *MMT-Based Media Transport Scheme in Digital Broadcasting Systems* (in Japanese).

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | Tariff and accounting principles and international telecommunication/ICT economic and policy issues |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| Series G | Transmission systems and media, digital systems and networks |
| **Series H** | **Audiovisual and multimedia systems** |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Telephone transmission quality, telephone installations, local line networks |
| Series Q | Switching and signalling, and associated measurements and tests |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities |
| Series Z | Languages and general software aspects for telecommunication systems |

Printed in Switzerland
Geneva, 2019