ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T H.430.3 (V2) (09/2023)

SERIES H: Audiovisual and multimedia systems

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

Service scenario for immersive live experience (ILE)

ITU-T H-SERIES RECOMMENDATIONS

Audiovisual and multimedia systems

CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100-H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	H.200-H.499
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	H.500-H.549
VEHICULAR GATEWAYS AND INTELLIGENT TRANSPORTATION SYSTEMS (ITS)	H.550-H.599
BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES	H.600-H.699
IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV	H.700-H.799
E-HEALTH MULTIMEDIA SYSTEMS, SERVICES AND APPLICATIONS	H.800-H.899

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

Recommendation ITU-T H.430.3 (V2)

Service scenario for immersive live experience (ILE)

Summary

Recommendation ITU-T H.430.3 identifies service scenarios by analysing several use cases of immersive live experience (ILE) services, in order to classify them and to establish a reference model for ILE. The 2023 edition of Recommendation ITU-T H.430.3 appends service scenarios and use cases of interactive immersive services as part of ILE. This Recommendation also summarizes several use cases and identifies candidate technologies for implementing ILE, including related standards gap analysis.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T H.430.3	2018-08-29	16	11.1002/1000/13668
2.0	ITU-T H.430.3 (V2)	2023-09-13	16	11.1002/1000/15652

Keywords

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

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

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/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Page

Introduction

In recent times, many major sporting and music events have not only been broadcast, but also delivered to remote sites for public or live viewing with the aim of sharing the emotional experience with spectators there as if they themselves were also in the event venue. At the same time, ultra-high definition (UHD) broadcasting known as 4K (resolution of approximately 4 000 pixels) and 8K (resolution of approximately 8 000 pixels) broadcasting is becoming increasingly popular. It is therefore expected that many people all over the world 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 highly 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 that are transmitted from several locations. By using ILE, real objects, in this case musicians at the remote sites, can collaborate with three dimensional images of remote musicians that are projected on to the screen. ILE enables virtual concerts with highly realistic sensations.

In order to share the enthusiasm of event venues with large audiences even if they are remote 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 for ILE to identify key requirements and architectural frameworks to implement its services and provides several use cases around the world. The 2023 edition of this Recommendation appends service scenarios and use cases of interactive immersive services as part of ILE. This Recommendation also provides candidate technologies for implementing ILE and standardization gap analysis of ILE-related technologies.

Recommendation ITU-T H.430.3 (V2)

Service scenario for immersive live experience (ILE)

1 Scope

This Recommendation analyses several use cases for immersive live experience (ILE), specifies service scenarios, and identifies key requirements and architectural considerations.

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 terms 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.

3.1.2 haptics [b-ISO 9241-910]: Sensory and/or motor activity based in the skin, muscles, joints, and tendons.

NOTE – Haptics consists of two parts: touch and kinaesthesis.

3.1.3 touch [b-ISO 9241-910]: Sense based on receptors in the skin.

3.1.4 cutaneous [b-ISO 9241-910]: Belonging to the skin.

3.1.5 tactile [b-ISO 9241-910]: Appertaining to touch.

3.1.6 kinaesthesis [b-ISO 9241-910]: Sense and motor activity based in the muscles, joints and tendons.

3.1.7 kinaesthetic [b-ISO 9241-910]: Appertaining to kinaesthesis.

3.1.8 vibrotactile [b-ISO 9241-910]: Vibration-based stimulation of the skin.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

3D	three Dimensional
4D	four Dimensional
6DoF	six Degrees of Freedom
AR	Augmented Reality
CI	Composition Information
codec	coder-decoder
DJ	Disc Jockey
EMG	Electromyograph
HMD	Head-Mounted Device
IIS	Interactive Immersive Service
ILE	Immersive Live Experience
IPTV	Internet Protocol Television
MMT	MPEG media transport
mux	multiplexer
OMAF	Omnidirectional Media Format
QoS	Quality of Service
UHD	Ultra-High Definition
VR	Virtual Reality

5 Conventions

Service scenarios in this Recommendation are described using the following template, in relation to 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
Stakeholder (actor)/domains	Roles of related stakeholders and domains
Main requirements	Service and technical requirements
Source (references)	Use cases in Appendix I or reference document, websites

6 High-level service category

This clause briefly describes the high-level service categories shown in Figure 1. ILE services or use cases can be classified into several categories of scenario, such as live sports, entertainment and telepresence.

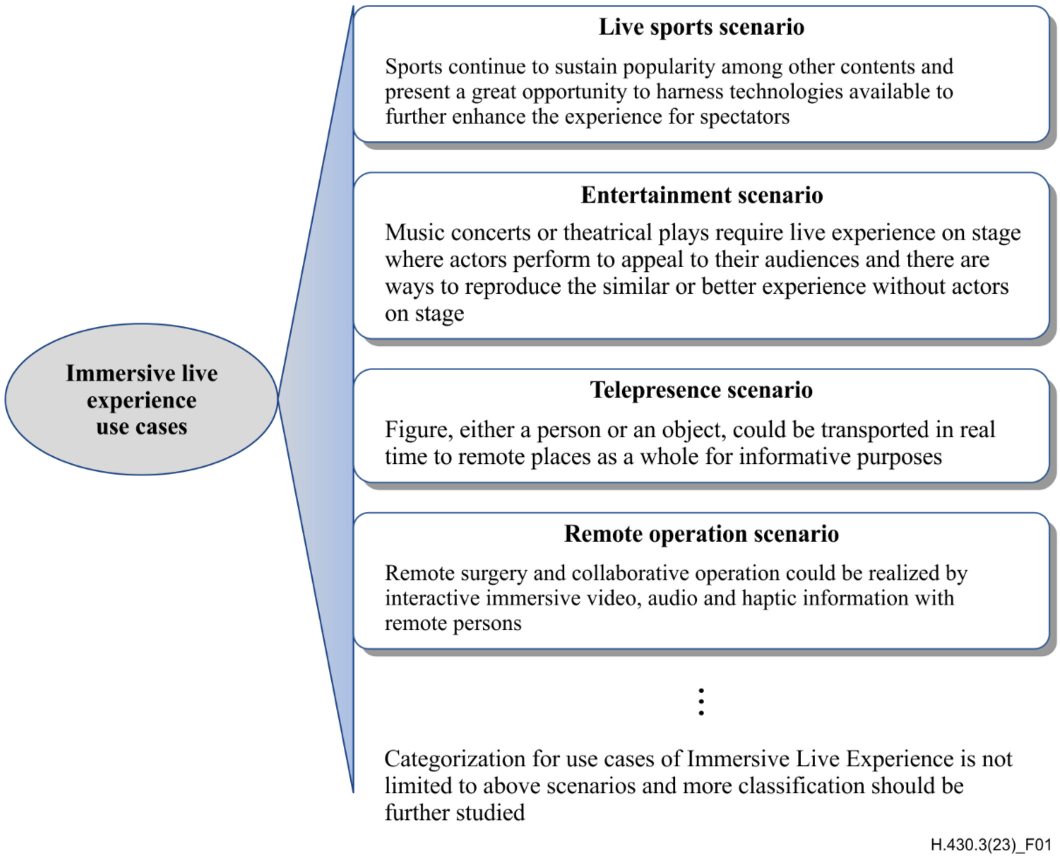


Figure 1 – High-level service categories of ILE services

Major categories for use cases of ILE include, but are not limited to, the following.

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

– Entertainment scenario: Concerts or stage plays require live experience where actors perform to their audiences; there are ways to reproduce a similar or better experience without live actors.

– Telepresence scenario: An image of a figure, either a person or an object, can be transported in real time to remote places as a whole for informative purposes.

– Remote operation scenario: Remote surgery and collaborative operation can be effectively realized by interactive immersive video, audio and haptic information with remote persons.

7 Categorized service scenarios

7.1 Live sports scenario

The following tables describe live sports scenarios.

Live sports	Live broadcast service of first-person synchronous view
Description	The service is provided by transmission by high-definition media of a first person viewpoint of an athlete in real time via a camera installed in sports equipment.

Live sports	Live broadcast service of first-person synchronous view
Service scenario	A synchronous camera is installed on the front of a fast-moving vehicle, such as a racing car or bobsleigh, whose perspective is fixed to face straight ahead along the racecourse. In other words, the camera provides a first person synchronous view as if audience members were riding in the athlete's vehicle. The camera is equipped with network capability of sufficiently high bandwidth to seamlessly upload its live feed in ultra-high definition (UHD) to a backend system that is then broadcast to user devices. The view is provided via either a big screen or mobile device screen. The additional information related to the race, such as speed, acceleration, current ranking and remaining distance to the finish line, can enhance the excitement of ongoing competition. 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, producers with broadcasting rights, network providers.
Main requirements	Extremely stringent delay, ultra-high-definition coder-decoder (codec), reliable and broadband uplink.
Source (references)	KT use case, see clause I.2.7.

Live sports	Live 360° virtual reality service of single point view
Description	The service provides an immersive experience of attending a live event at a sports venue by constructing a 360° panoramic view in real time via multiple camera feeds from the site.
Service scenario	A camera capable of filming live panoramic views or several high-definition cameras is or are installed at the venue such as a gymnasium, an ice rink (for figure skating events) and stadia (for ball games, athletic sports and opening or closing ceremonies). The live feeds from cameras at multiple angles are transmitted via a high-bandwidth uplink to a backend system for distribution with minimum latency. In order to view several media, including broadcasting, in an appropriate form, a user might need to wear a special device like virtual reality (VR) goggles or a head-mounted device (HMD) that constantly process and piece together multiple images to project the real world. For this purpose, it can be used in pseudo-three dimensional (pseudo-3D) displays without special wearable devices such as VR goggles or an HMD. The purpose of the service is to enable viewers to observe every event live around the venue so that they feel as if they are present at the remote location.
Stakeholders (actors)	Institutes managing facilities at sports venues, participating athletes, producers with broadcasting rights, network providers, VR manufacturers.
Main requirements	Extremely stringent delay, UHD codec, reliable and broadband connection, massive connectivity, VR media processing (e.g., image stitching).
Source (references)	NTT use case (Kirari!), see clauses I,2.1, I.2.8, I.2.9 and I.3. KT use case, see clause I.2.7.

Live sports	Live broadcast service of selected multi objective view
Description	The broadcasting service provides a 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 is installed at strategic locations along the racecourse where live action of athletes can be best captured simultaneously. 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 is tracked in real time and their corresponding locations continuously indicated on the map. The service provides a perspective of the racecourse from all angles to better track favourite athletes in action.
Stakeholders (actors)	Institutes managing facilities at sports venues, participating athletes, producers with broadcasting rights, network providers, outdoor positioning information providers.
Main requirements	Extremely stringent delay, UHD codec, reliable and broadband uplink, massive connectivity, location-based services (e.g., 3D map, precise positioning).
Source (references)	KT use case [b-Sohn], see also clause I.2.7.

Live sports	Becoming-player sports viewing
Description	The service provides a becoming-player experience by using haptic sense in addition to video and audio during a sports match. A combination of visual-auditory-haptic senses aims to reproduce becoming-player sensations for real players at the event site.
Service scenario	At the viewing site, people are in front of a monitor and move as if they are players in the stadium at the event site. Sound from the event site is transmitted to vibration platforms on which people stand, equipped with vibration devices. While watching the game on the monitor, each person moves their devices as if they are one of the players at the event site. As a result, people feel as if they are standing in the stadium at the event site due to the vibrations of their feet, and at the same time, they feel as if vibrations are occurring in their bodies in response to their movements. They are able to feel the physical movements of players and reproduce them through their own bodies. The purpose of the service is to provide audiences with becoming-player sports viewing by interactive ILE with haptic information, which can be further enhanced by addition of special effects, such as lighting, temperature and humidity.
Stakeholders (actors)	Interested people: producers with broadcasting rights, promoters of a sports event, video, audio and haptic equipment providers, data analytics providers, network providers.
Main requirements	High-definition codec, high definition audio, haptic sensing, haptic transmission, haptic reproduction, video/audio/haptic multiplexer (mux), converter, reliable and broadband connectivity.
Source (references)	NTT West use case, see clause I.2.16.

7.2 Entertainment scenarios

The following tables describe entertainment scenarios.

Entertainment	On-stage holographic performance show service
Description	The service provides holographic media of concerts that are live-streamed or pre-recorded for audiences at special theatres with capability of reconstructing the live environment.

Entertainment	On-stage holographic performance show service
Service scenario	A live concert performance is captured or pre-recorded by several UHD cameras at multiple angles and using high-quality microphone systems. Then, high-quality UHD video and audio is processed separately at the initial stage. Processed media streams. including video, audio and other sensory data, should be transmitted synchronously from the event to remote sites. A theatre similar to one set up for the original concert hall is reconstructed at a remote site for highly realistic effects where the processed media stream is regenerated to simulate the live event. Audiences at such a remote site can experience almost the same visual and sound quality as the original on holographic or pseudo-3D projection. The purpose of the service is to evoke the sense of a concert-like environment that is on a par with a real concert but without actual performers on stage.
Stakeholders (actors)	Concert performers, concert producers, network providers, hologram equipment providers.
Main requirements	UHD codec, hologram conversion, on stage hologram or pseudo-3D projection, video/audio mux.
Source (references)	NTT use case (Kirari!), see clauses I.2.1, I.2.8, I.2.9 and I.3. KT use case, see clause I.2.7.

Entertainment	Live interactive service
Description	The service provides virtual scenes in collaboration with real people and transmitted images in real time. The images are extracted from real people in different places and transmitted by real-time high-definition media. The transmitted images and sounds are reconstructed and interact with real people at a remote site. The service, which might be considered as a variation of live interactive service, also provides virtual scenes in collaboration with real people and transmitted images in real time. Haptic information is interactively communicated within remote places and event sites.
Service scenario	Live video images of artists at 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 to multiple transparent screens that are spatially set up. As an optional feature, the vibration of artists or audiences can be captured at each site and transmitted to other sites including studios without audience, in order to share amount of excitement and to play music synchronously even in noisy situations. A real person, such as a disc jockey (DJ) or master of ceremonies, at the remote site can collaborate and interact with artists in event venues by transmitted images. The purpose of this service is to provide the audience at a remote site with a new immersive experience as if they are watching the live stage of a DJ interacting with artists in the event venue who are teleported to remote sites beyond time and space. By adding haptic communication, this service may provide new entertainment feature which realize location free concerts, and the audience in event venues and remote sites with a new immersive experience as if they are watching one complete live stage in the event venue.
Stakeholders (actors)	Institutes managing facilities at event venue, participating artists, producers with broadcasting rights, network providers.
Main requirements	Extremely stringent delay, UHD codec, reliable and broadband uplink and downlink, haptic communication.
Source (references)	NTT use case, see clauses I.2.8, I.2.9, I.2.10 and I.2.11.

Entertainment	Multi-angle viewing service
Description	The service projects virtual pseudo-3D images on to a special arena, which can be viewed from any angle, and reconstructs sound direction. The images are captured by several cameras at the event site and transmitted with audio and spatial information synchronously to remote sites, at which the images and sound are reconstructed in a special arena. The users can simultaneously watch the images from their standpoints.
Service scenario	The images of artists or sport players at an event site are extracted by using several cameras and their sounds captured by using several microphones. In addition, 3D-sensing and object-tracking information is captured to fix spatial locations. These video and audio streams are synchronously transmitted to the viewing site with spatial information via high-speed networks. At the viewing site, these images and sounds are reconstructed in a special arena, which has multiple screens and an array of loudspeakers. The audiences at viewing sites can see images from any angle outside the special arena and synchronously hear sounds. In order to realize image or auditory localization, pseudo-3D image displaying and wave field synthesis might be used in a special arena. The purpose of this service is to provide experience of real events that are reconstructed images and sounds in 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, producers with broadcasting rights, network providers.
Main requirements	Extremely stringent delay, UHD codec, reliable and broadband uplink and downlink, pseudo-3D display, wave field synthesis.
Source (references)	NTT use case, see clauses I.2.8, I.2.9, I.2.10 and I.2.11.

7.3 Telepresence scenarios

The following tables describe telepresence scenarios.

Telepresence	Live ultra-high-definition presence video view service
Description	The service provides a real-time holographic presentation of a person of interest in UHD with augmented information or special effects around the figure for audiences at viewing sites.
Service scenario	A single person of interest such as someone being interviewed or highlighted is captured in real time by a special holographic camera. Once high-definition video and digital audio are encoded, they are synchronously transmitted via an ultra-high-speed network to a viewing site where the holographic video and audio are decoded and reproduced in real time. Various special effects like animation or information in many formats can be projected on to an original object or person for amusement or marketing. For a better feeling of reality, projection of the person or object at their actual size is preferred. The purpose of the service is to provide ILE of telepresence of a person or object at their actual size that can be further enhanced by addition of special effects.
Stakeholders (actors)	Interested people: producers with broadcasting rights, hologram equipment providers, data analytics providers, network providers.
Main requirements	High-definition codec, hologram conversion, hologram projection, high-definition audio, video/audio mux, converter, reliable and broadband connectivity.

Telepresence	Live ultra-high-definition presence video view service
Source (references)	KT use case [b-KT PC], see also clause I.2.7. NTT use case [b-NTT JKF], see also clauses I.2.8, I.2.9, I.2.10 and I.2.11.

Telepresence	Immersive live meeting
Description	This service provides teleconferencing where participants from remote sites are represented as highly realistic images to each other in order to induce an immersive experience of being present in the same meeting room.
Service scenario	Each participant image is captured by cameras, then high-resolution video and audio are transmitted over extremely low latency and high bandwidth networks. A highly realistic image of each participant is distributed to all others in real time. Users can view the scene at the reconstruction end and feel themselves to be in a fully immersive meeting environment, where they can interact with actual figures of people. The purpose of this service is to provide a portrait capture and display service for video conferences, board games, etc.
Stakeholders (actors)	Business discussants, image capture equipment providers, network providers, game producers.
Main requirements	Image capture equipment, UHD codec, reliable and broadband uplink and downlink, virtual or holographic display device.
Source (references)	OKI use case, see clause I.2.4.

Telepresence	Interactive presence service with haptic sense
Description	The service provides haptic sense in addition to video and audio of a person during face-to-face conversation. A combination of visual-auditory-haptic senses aims to reproduce the atmosphere of a real conversation, by using a communication system with other people who share a similar desk at a remote site.
Service scenario	A user is seated in front of a communication system, whose display shows the upper body of another user at a remote site. The area between the display and user is a shared space for haptic interaction. The area shows the hands of another user and vibrators are attached behind the screen to reproduce vibrotactile information at the remote site. Vibrotactile information from the remote user is captured by sensing devices and transmitted from the remote to the other site. Haptic information is reproduced by the vibration devices synchronously with visual and auditory information from the remote user. The purpose of the service is to provide interactive ILE of telepresence of a person or object at actual size with haptic information, which can be further enhanced by addition of special effects, such as lighting, temperature and humidity.
Stakeholders (actors)	Interested people: producers with broadcasting rights, hologram equipment providers, data analytics providers, network providers.
Main requirements	High-definition codec, hologram conversion, hologram projection, high-definition audio, haptic sensing, haptic transmission, haptic reproduction, video/audio/haptic mux, converter, reliable and broadband connectivity.
Source (references)	NTT use case [b-NTT ICC].

7.4 Remote operation scenarios

The following tables describe remote operation scenarios.

Remote operation	Interactive live object control
Description	This service provides interactive control of actual objects from remote sites in real time so that users can experience additional senses of touch and control, providing an increased degree of reality as a result.
Service scenario	Certain spatial dimensions are set up at a real-world location, which users can physically pilot from remote sites. The boundaries of the space are marked by sensors installed on the corners. Selected physical object(s) is (are) placed inside the setting delimited by the boundaries. Such objects are scanned and digitally encoded in advance or in real time to create an exact replica for display and control of the virtual object. With this or these physical object(s), a user can navigate and manipulate the virtual objects with a high degree of ease and comfort. This virtual object is coordinated in a central location, distributed over a network of high bandwidth and low latency and fed into an HMD or 3D display in real time for multiple users at different places where they interact with each other in similar settings to the first as if they are all in the same venue. Especially with such items under control, users interact with virtual objects in a far more immersive and live manner. The purpose of the service is to provide users with a great degree of manoeuvrability via a virtual world combined with physical object(s), further enhancing immersive and live aspects of user interaction experience.
Stakeholders (actors)	VR producers, government and travel agencies for scenery rights, sports associations, HMD and sensor equipment providers, network providers.
Main requirements	VR production, object encoding and coordination, UHD codec, reliable and broadband downlink, extremely stringent delay, HMD and VR sensors.
Source (references)	Tektonspace use case [b-Tektonspace].

Remote operation	Remote object operation
Description	The service provides interactive remote control capabilities. Users can perform operations on remote objects via their virtual images.
Service scenario	A high-resolution image of remote object(s), with the help of capture equipment, is generated by a combination of XR, holographic 3D image, surrounding audio, haptic information and other multimedia technologies. The image is transmitted to remote users over extremely low latency and high bandwidth networks in real time. Users can perform a series of operations on remote object(s) by interacting with this image. Information about operations on the image is captured and converted into control signals, transmitted to remote auxiliary facilities through extremely low latency and reliable networks. The facilities perform operations on remote object(s) in accordance with the control signals received. Feedback information about remote object(s) is reflected in the image in real time. The purpose of this service is to provide interactive remote capabilities for industrial control, remote maintenance, remote surgery, prototype design and test, etc.
Stakeholders (actors)	Manufacturers, medical institutions, image equipment providers, sensor equipment providers, auxiliary equipment providers, network providers.
Main requirements	Image capture equipment, sensor equipment, UHD codec, reliable and broadband uplink and downlink, ultra-low latency, extremely stringent delay, multiple stream synchronization, hologram conversion and projection, holographic display device.

Remote operation	Remote object operation
Source (references)	Remote surgery use case, see clause I.2.12.

Remote operation	Real-time multi-user collaboration
Description	This service provides multi-user manipulation on 3D or holographic images of real objects or digital models, which can be simultaneously viewed from any angle or position and changed by users.
Service scenario	The images of real object(s) captured by high-definition cameras or laser scanners and virtual object(s) generated by computer as an option are transmitted to multiple users after media processing (e.g., media selection, rendering). Multiple users can synchronously view the virtual object(s) on display devices from different angles, and manipulate the virtual object(s) via peripheral equipment(s) (e.g., camera, data glove). During the collaboration, remote users can see a virtual scene of the event site and all real-time changes in the virtual object(s), and communicate and collaborate with each user at the event and other remote sites as if they work in the same place. The purpose of this service is to provide users with synchronous interaction on real and virtual object(s) by immersive communication with each other user for remote teaching, remote training, remote collaborative design, etc.
Stakeholders (actors)	Capture equipment providers, network providers, display equipment providers, peripheral equipment providers, owners of the physical object, simulated data providers, collaborative participants.
Main requirements	Capture device, UHD codec, reliable and broadband uplink and downlink, hologram or 3D display, real-time media processing.
Source (references)	Multi-user collaborative design use case, see clause I.2.13.

7.5 Emergency scenarios

The following tables describe emergency scenarios.

Emergency	Live streaming service in inaccessible areas
Description	The service provides real time reconstruction of scenes in inaccessible areas that cannot be reached by humans immediately, such as post-disaster environments, polar regions or inaccessible mountains.
Service scenario	This scenario provides an immersive experience for inaccessible areas. By using robots and unmanned aerial vehicles carrying microphones, holographic cameras, environmental sensors in position or by using pre-installed sensors, UHD video and audio are captured and processed. In close proximity to these areas, in a location that people can reach, high-bandwidth and low-latency wireless network coverage is needed to transmit high-quality holographic data. 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 testing by manufacturers under extreme conditions, such as a challenging temperature and humidity simulator for human tolerance.
Stakeholders (actors)	Government, travel producers, hologram equipment providers, network providers.
Main requirements	Capture device, environment analysis and position assign system, UHD codec, reliable and broadband uplink, hologram projection.
Source (references)	NA.

7.6 Remote assistance scenarios

The following tables describe remote assistance scenarios.

Remote assistance	**Remote assistance**
Description	This service provides remote assistance to users by professionals. Users can experience highly realistic guidance by using this service.
Service scenario	This service contains one guiding site as a remote site and guided sites as event sites. An original image is captured from one guided site and transmitted to another to project the virtual image. Professionals can demonstrate on the image for guidance. The demonstration (voice, action, signs, etc.) is captured and transmitted to guided sites. Then representatives at guided sites demonstrate on the real object. Users at guided sites can do the same operation on a real object according to the guidance presented. Changes in real object and demonstration are synchronously reflected in the images at both sites.
Stakeholders (actors)	Government, rescue centres, education or technical support centres, hologram equipment providers, augmented reality (AR) equipment providers, network providers.
Main requirements	Capture device, reliable and broadband uplink, AR or holographic display device.
Source (references)	Emergency treatment guidance use case, see clause I.2.14. Remote maintenance guidance use case, see clause I.2.15.

Appendix I

Use cases for immersive live experience services

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

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

I.1 Introduction

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

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 or live viewing sites for sharing emotions with audiences. Watching styles are becoming more diversified. It is expected that sharing emotions with audiences will be increased by highly realistic television and public viewing sites using 4K and 8K technologies.

In order to provide UHD content and highly realistic sensations for audiences in remote sites, NTT has been working on research and development of immersive telepresence technology called "Kirari!", which consists of real time and synchronous media transmission, UHD 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 highly realistic sensations as if they were present in 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 watch their favourite stars with their own eyes.

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 can 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 reconstruction enables a 3D image of a real person or object to be projected on to a computer-generated background and provides an immersive experience of being present in the virtual surroundings.

Another sophisticated video streaming technology can be employed, in which many tile-based images are encoded using high efficiency video coding and then stitched together to render a cubic virtual environment.

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

With advanced technologies using six degrees of freedom (6DoF) VR reconstruction of the surrounding environment, sparse camera arrays, and a light field of representation and compression and point cloud, a 3D interactive service delivers a virtual environment in which a single or multiple users can interact with each other or computer-generated 3D objects in an ultra-realistic way. This service provides 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 in which people can collaborate as if they were right next to each other.

The general public is able to watch big sporting events such as the Olympic or Paralympic Games, which are broadcast in extremely high definition of 4K or 8K. This service provides viewers with much crisper and more accurate images of sports events and produces an ILE 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 adjusts the angle of a scene accordingly in real time, providing some degree of immersive experience. The video is produced and rendered on a backend system from signals received by remote control of movement.

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

Live video of 8K quality is broadcast over satellite with multiple videos of high quality up to 4K supplemented over broadband in order to provide a high degree of excitement about live action.

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

5G technologies of high bandwidth and low latency provide robust infrastructure on which ILE is featured for a batch of 5G services.

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

The 360° VR/TimeSlice service lets audiences with HMDs 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 teleports a real size image of a person or object from remote sites and creates a virtual environment in which an interview can be conducted live. This multi-site service utilizes hologram technologies to carefully integrate two live images from different venues into one place to coordinate or choreograph multi-party activities as if everyone were present at the same spot.

I.2.8 Use case – Interaction with images (NTT)

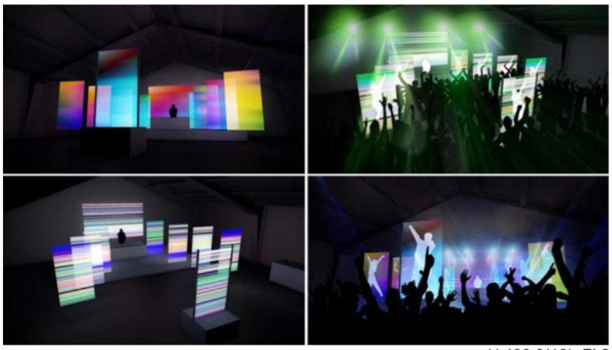
On 2017-03-13, a music live showcase was presented as a remote interactive performance between Tokyo and Austin, Texas.

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* [b-NTT Krause], in which the audience experienced an immersive synchronized performance of a DJ in Austin and the artists "teleported" from Tokyo.

The live video images of the artists in Tokyo were extracted in real time and transported in various media formats synchronously over the network, then dynamically projected on to multiple double-sided transparent screens that were spatially set up in the Austin live space (see Figure I.2) by utilizing NTT's Kirari! technology. The latency is about 400 ms, so the interaction was 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), the 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 interplay live on stage of the DJ with the artists in Tokyo who were teleported to Austin.



H.430.3(18)_FI.2

Figure I.2 – Computer-generated images of the stage at *Cyber teleportation Tokyo at SXSW*

I.2.9 Use case – Multi-angle viewing styles (NTT)

Several cameras at an event site and transmitting images all over the world, enables the atmosphere of the event site to be captured and reconstructed at viewing sites. NTT has developed a new viewing style for audiences who watch a reconstructed event together at a remote site, namely *Kirari! for Arena*, which reconstructs the atmosphere of the event site allowing viewing from several angles by audiences at viewing sites, by capturing object images with their location information. Figure I.3 shows an image constructed by "*Kirari! for Arena*" of a judo event.

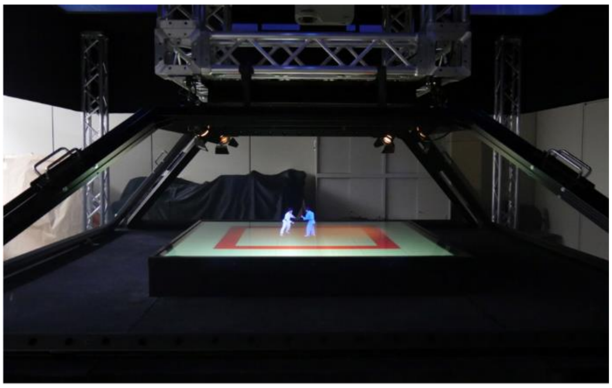
H.430.3(18)_FI.3

Figure I.3 – Image constructed by *Kirari! for Arena* of 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, capture of 3D position information of objects and object tracking are required. The combination of object recognition using deep learning, object tracking using particulate filters and depth sensors using lasers enables rough tracking of object location. 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 and displays natural motion of objects even in the depth direction. This also enables many audiences to observe a displayed object from any angle. Pseudo-3D image localization can reconstruct event venues with moving objects at viewing sites, so that dynamic display achieves spatial position relations as if players were moving around 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. Displayed at viewing sites on the horizontal floor plane is a projected pseudo-3D image reflected by an angled half mirror, which transmits some incident light and reflects the rest. In this case, it is difficult to move images in the depth direction since the pseudo-3D images are fixed at the same location. In order to display moving objects in the depth direction, utilization of visual effects is required for audience perception.



H.430.3(18)_FI.4

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

I.2.10 Use case – Remote ensemble concert using haptic communication (NTT)

Remote or virtual concerts, which are realized by using telecommunication network, are challenges addressed recently. At some remote concerts, musicians are located in several different places, like Tokyo and Paris, and they play instruments or sing songs synchronously. This type of remote concerts is called "remote ensemble concert".

Remote ensemble concerts need communication networks with very low latency, because players who are in different places have to play synchronously. Most music players recognize the baton of conductor, sound of drumstick and some vibrations when they start playing, so haptic communication may be needed for remote ensemble concerts.

As an example, Figure I.5 shows an image of a remote ensemble concert.

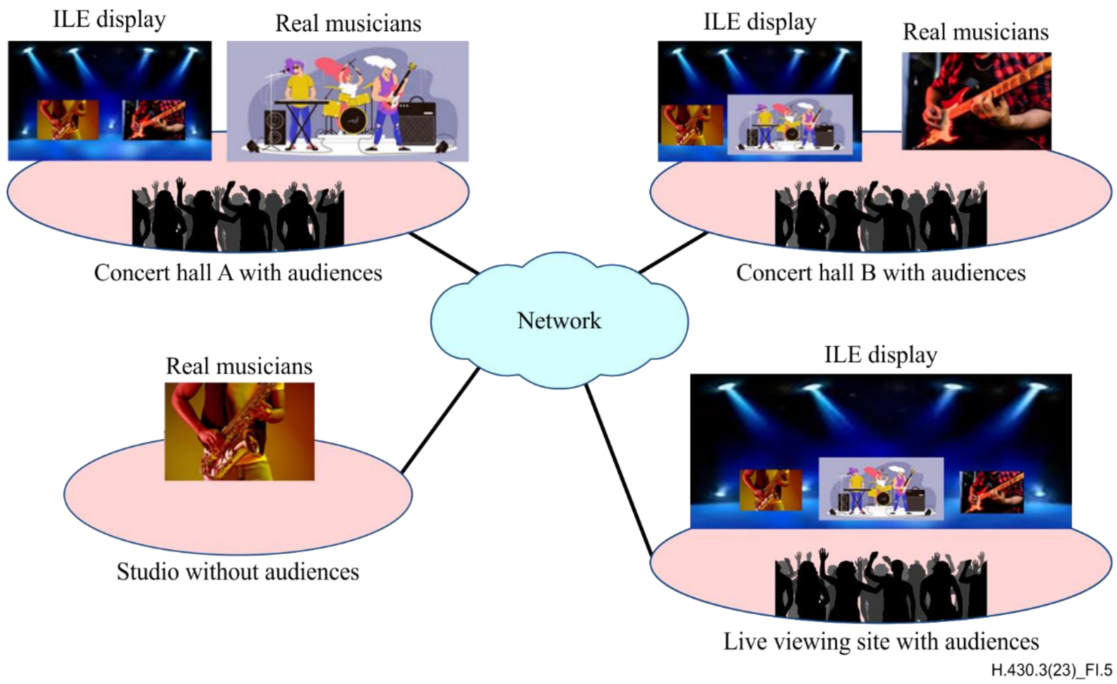


Figure I.5 – Images of a remote ensemble concert

This example is a live concert event involving four locations. Real musicians play instruments in three places, and audiences are located at three places. In concert hall A, an ILE display is installed to show real musicians at remote places, allowing audiences to see them as well as local real musicians. At a live viewing site, there are ILE systems that reconstruct a virtual event venue with combination of other places, and audiences can enjoy virtual concert. All musicians have to play synchronously, so timing recognition is critical when they start playing music. In concert halls A and B, musicians may need vibration signals as cues because of the noisy situation. Vibration signals might be sent from hall A to hall B and studio allowing synchronous playing of music. In addition, real musicians in the studio, where there is no audience, may feel the atmosphere in halls A and B and live viewing site by video, sound and some vibration information.

The benefits of haptic communication such as vibration transmission in a remote ensemble concert are considered at the start of a music performance, especially to adjust timing, as follows:

– vibration might not be affected by a noisy environment;

– for instrumental playing, vibration might not hinder music performance;

– vibration can be used even in dark concert hall such as rock concert with lighting effects or stage effects, and stage monitor display might hinder stage effects;

– players can pick up cue from the conductor cue by vibration, thus players can see audiences in the concert halls and in the remote sites through display.

When a rhythm is constant, haptic signals might be generated by a beat manager such as a conductor, sequencer or metronome. For a remote ensemble concert, a haptic signal might be generated by information from a remote site. An electromyograph (EMG) signal is obtained on starting to play musical instruments because humans move muscle to make sound. For example, an EMG signal is generated when a drummer raises a drumstick, since this involves arm movement. Thus, an EMG signal can be detected before the stick hits the drum. If information from the EMG is transmitted earlier than that of sound, other players like guitarists at another site can feel the vibration triggered by the EMG information before the drum is heard from the remote site.

In this example, it might be needed to implement interactive haptic communication, in addition to 3D videos, audios, lighting, and stage effects, within these four sites for a successful remote ensemble concert.

I.2.11 Use case – Public booth for vibrotactile communication (NTT)

A public booth for vibrotactile communication has been introduced as an interactive immersive service (IIS) using vibrotactile communication. This service utilizes an interactive audio-visual-haptic media that allows users to feel vibrations of other people at remote site who share feelings at the same desk. The system configuration is shown in Figure I.6.

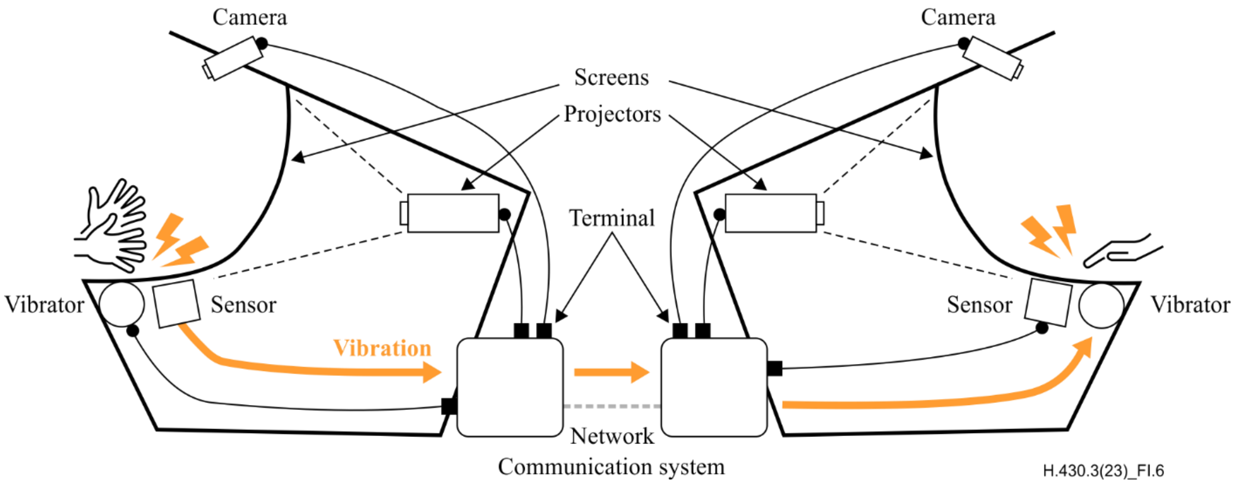


Figure I.6 – Configuration of the public booth for vibrotactile communication

Figure I.6 shows a pair of booths that connects via a communication system (terminals and network). Each booth has a camera, one round screen, a projector, a vibration sensor and a vibrator. Behind the round screen, a vibrator is attached to provide vibrotactile sense to the user and a vibration sensor to capture vibration generated by the user. In order to avoid vibration howling, a sound equalizer and a howling compressor might be needed. The sound equalizer removes high-frequency vibration signals, where humans cannot feel vibration, and signals of specific frequency bands that cause strong howling by resonant frequency of booth and screen. A compressor controls howling in real time.

NTT Labs demonstrated the public booth for vibrotactile communication in exhibition spaces in Tokyo in 2019. Figure I.7 is an image of the demonstration.

H.430.3(23)_FI.7

Figure I.7 – Demonstration of the public booth for vibrotactile communication

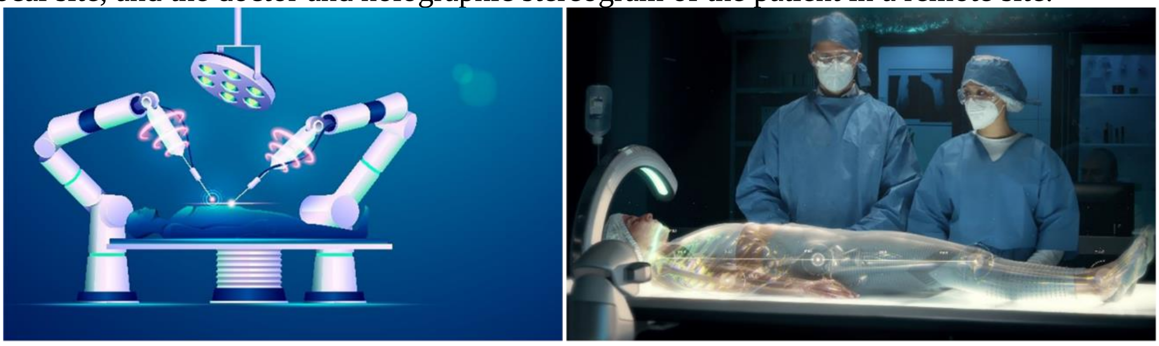
This demonstration used high-definition video and audio communication system and the Internet. Vibrotactile information was carried by one of two audio channels, which were used for stereo sound transmission, and was fully synchronized with video images. The video signals were carried by a video codec such as one compliant with [b-ITU-T H.264], and the audio and vibrotactile signals were carried by audio codec such as MPEG-4 AAC-LD. The audio and vibrotactile signals can also be carried by MPEG-4 ALS.

This demonstration provided several experiences using vibrotactile information between two users. When seated in front of a booth, user A can see the upper half of the body including the hands of user B, who is seated in front of another booth at a remote site, at actual size on the round screen. In the appropriate areas, both users can see each other's hands and feel vibrations. For example, if user A hits the desk with the hand, user B can hear the sound and feel the vibration simultaneously. This kind of interactive experience makes users feel as if they are seated at the same desk.

In this demonstration, users can also feel vibrations of rolling balls, drawing pictures, and touching and scuffing the screen from the other side.

I.2.12 Use case – Remote surgery

Based on extremely stringent delay and reliable connectivity, remote surgery is one critical application of telemedicine. As shown in Figure I.8, it involves the patient and auxiliary facility in the local site, and the doctor and holographic stereogram of the patient in a remote site.



H.430.3(23)_FI.8

a) **b)**

Figure I.8 – Remote surgery: a) local site; b) remote site

First, the holographic stereogram of the patient in the local site is generated by holographic technology and captured with the help of cameras and sensors; the result is displayed at the remote site. The remote doctor can view the real-time status (such as pulse count, blood pressure measurements and blood vessel status on the surface of the skin) of the local patient.

Second, the remote doctor performs a series of operations on the holographic stereogram. Haptic information about operations (such as touch and kinaesthesis) is captured by cameras and sensors, converted into control signals and transmitted to local auxiliary facilities (such as mechanical arms) through networks.

Finally, local auxiliary facilities perform represented operations on the local patient and the feedback interactive information is reflected on the remote holographic stereogram in real time. The remote doctor can judge the result and continue to the next.

I.2.13 Use case – Multi-user collaborative design

With advanced 3D display technologies, a design prototype or physical model is represented by a high-resolution 3D image at actual size at different sites, and can be watched or modified by users. Highly realistic image(s) of the user(s) at the event site is or are transferred to the remote site and presented at life size and in real time, and *vice versa*. Figure I.9 is an image of an example of multi-user collaboration for vehicle design. Colleagues who work in the office (event site) can communicate and collaborate with a colleague who works from home (remote site) to interact with the virtual vehicle model.

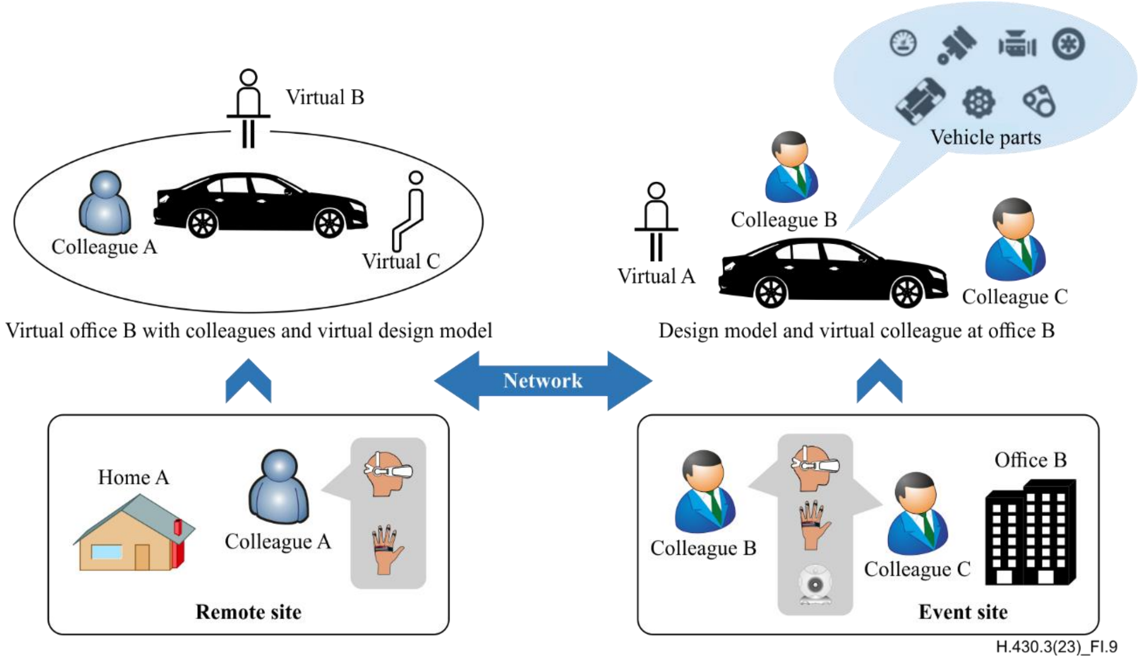


Figure I.9 –Multi-user collaboration

In the collaborative design process, each user's ideas and operations on the virtual object are tracked and captured via cameras, wearable devices or other accessories, and delivered to other users working at different sites in real-time. Users at the event site can design the virtual object with natural experience, e.g., by moving it, modifying components or attributes, labelling instructions and testing new configurations. The users at the event and remote sites can watch the same virtual object from an arbitrary viewing angle, discuss the design plan and design the virtual object simultaneously. This service provides 3D visualization of the design process, immersive interaction with virtual objects and immersive communication for users in real-time.

I.2.14 Use case – Emergency treatment guidance

Many injured people die in emergency situations (drowning, heart disease, stroke, etc.) due to lack of appropriate treatment in the first minutes after the event. Guidance is in high demand due to the deficiency of emergency treatment knowledge.

Remote rescue centre professionals can guide on-site people to treat emergencies appropriately in a timely fashion. The operation on the replica of the patient performed by surgeons is captured and presented to people on site, who can then perform the same treatment on a real patient. In addition, if a natural disaster such as an earthquake occurs, trapped people can be guided to treat themselves.

I.2.15 Use case – Remote maintenance guidance

If faults occur, a customer calls a technical support centre for assistance and is guided by a specialist for maintenance.

An image of the real device is captured at the customer site and transmitted to the support centre. The specialist demonstrates maintenance on the virtual image using voice, action, signs, etc. The demonstration is captured and transmitted to the customer site and represented as if demonstrating on the real device. The customer conducts the same operation on the real device, and any changes of device are synchronously reflected in virtual image. The device is then fixed with the guidance of the specialist.

I.2.16 Use case – Becoming-player sports viewing (NTT West use case)

In sports viewing, audiences cheer with enthusiasm because they have empathy with players, and player performance can improve through the enthusiastic cheering. "Becoming-player sports viewing" is a new experience, in which audience emotions at viewing sites can be evoked using haptic information from the event site.

One use case is the fencing project that was demonstrated in Numazu-city in Shizuoka, Japan in 2021. In the stadium (event site) for a fencing match at the Roppongi Hills Arena in Tokyo, Japan, six microphones were installed at the side of the piste (on which the match must occur), mainly to capture the sound of a fencer's foot stepping on the piste and the sound of swords colliding with each other. The captured sounds were mixed into two channel audio, and transmitted to Numazu by telecommunication network. This use case involves a sabre match (one of three disciplines in fencing), in which the weapon is used to slash rather than stab, unlike the other two. Thus, the sabres make big sounds and intense vibrations.

Figure I.10 is an image of this use case. At the viewing site in Numazu, two children with experience of sabre stood facing each other in front of the monitor, as if they were in the stadium in Tokyo. The video image captured by the camera in Tokyo was displayed on the monitor at Numazu. The sound from Tokyo was transmitted to the speakers, including output from the vibration platform that the two children stood on and the sword-type vibration device held in their hands. The sound signal from the left channel mix of output from the platform and the device on the left side of the viewing site was sent to one speaker, and the sound signal from the right channel mix was sent to another for the right side. The children were instructed to move their hands while synchronizing to each of the players they were becoming. The video images and sounds from Numazu were sent to the monitor and speakers in Tokyo.

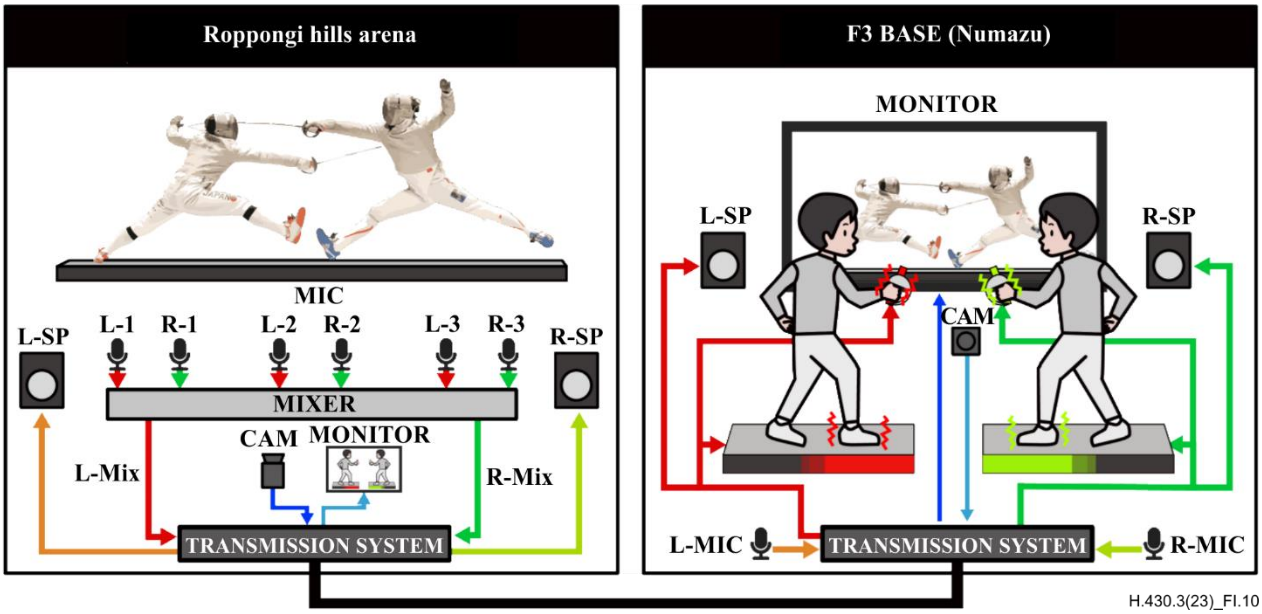


Figure I.10 – Becoming-player sports viewing

As a result, children felt as if they were standing in the stadium in Tokyo due to the vibrations of their feet, and at the same time, they felt as if vibrations were occurring in their hands in response to their movements. They were able to feel the physical movements of the players, such as footwork and swordplay, and reproduce them through their own bodies.

After the demonstration, the children's impressions were, "I was happy to win", and, "I was disappointed to lose", so the children talked about winning and losing as if they had become players themselves.

I.3 Configuration of ILE

If transmitting a soccer game to gymnasia for example, there is a big difference in the sizes of venues as well as the sound and lighting equipment. Figure I.11 shows a configuration of Kirari! as an example of an ILE service, in order to reconstruct the atmosphere of a soccer stadium in a remote site. The following approaches can achieve highly realistic sensations as if audiences in a gymnasium had entered a 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 in the gymnasium, information about the object is transmitted in real time.

2) Video, audio and lighting information, except for objects previously described, is also transmitted to gymnasia 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 gymnasium environmental data to reconstruct the atmosphere of the stadium at remote sites. For example, the following are the actual parameters: i) (*x*, *y*, *z*) coordinates of objects; ii) range of captured environmental and spatial information except for objects; iii) region of sound space.

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

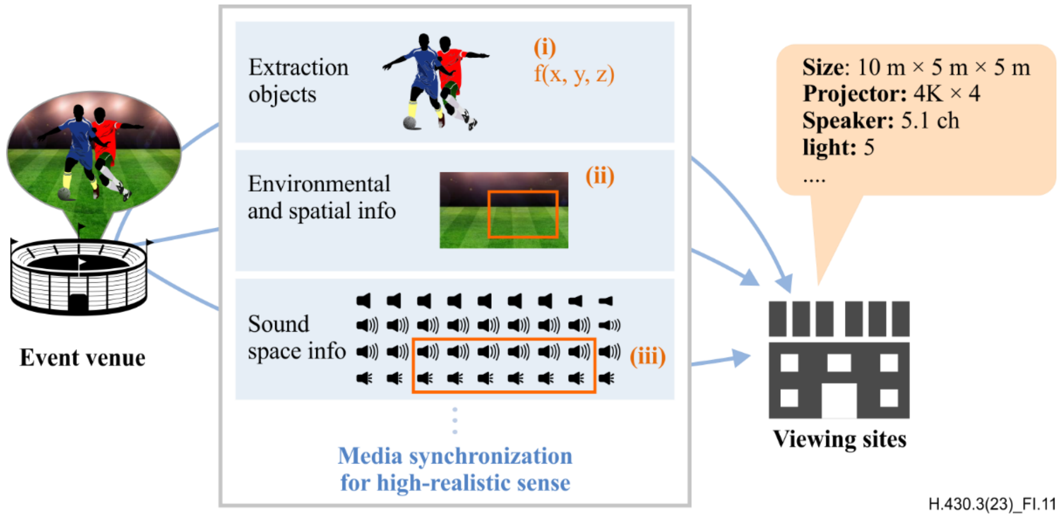


Figure I.11 – Configuration of Kirari!

As shown in Figure I.12, audiences in remote sites can feel ultra-highly 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.

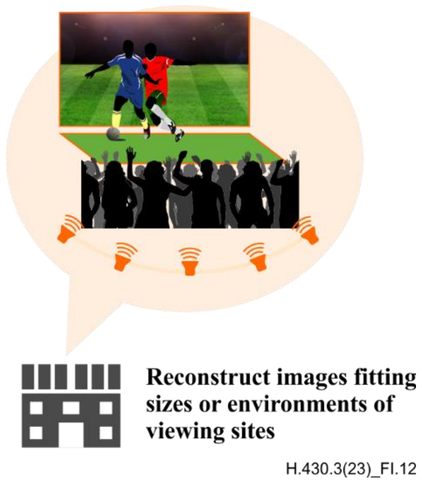


Figure I.12 – Reconstruction atmosphere of event venue at viewing site

I.4 Candidate technologies used for ILE

Candidate technologies used in ILE that aim to realize pseudo 3D presentations in real time are described in this clause and include highly realistic media synchronization, real-time object image extraction, highly realistic auditory lateralization and extra-wide video stitching technologies.

(1) Highly realistic media synchronization

The MPEG media transport (MMT) technology has features that 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. A 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 highly realistically at the viewing site, depending on its specifications.

For IISs, interactive media require multiple concurrent streams to be delivered. Coordination and synchronization of multiple concurrent streams is a key technology to control their establishment, synchronization and reconstruction according to specific scenarios and service policies.

(2) Real-time object image extraction

Real-time object image extraction is a key technology to realize live experience, and extracted objects are displayed as pseudo-3D images at the 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 at high speed and in 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) Highly realistic auditory lateralization

Highly realistic auditory lateralization aims to reconstruct a virtual acoustic field over a wide viewing area. This enables sound direction from actual-sized objects projected on to wide screens to be reproduced, thus audiences can feel voice and sound as if they come from the projected objects.

(4) Extra-wide video stitching

Video-stitching technology pieces together processed video images of multiple high-definition cameras, adjusts the images and enables highly realistic field images to be reconstructed in real time for major sports venues, and track and field events such as athletics.

(5) Intelligent distribution technology of massive multimedia data

Intelligent distribution technology of massive multimedia data aims to select appropriate media processing servers in accordance with the requirements of codec, compression and rendering.

(6) Network status awareness and quality of service scheduling

Network status awareness and quality of service (QoS) scheduling aim to guarantee the quality of experience of IISs. Network status awareness is the capability to collect QoS and network resource information. QoS scheduling is the capability to update policies based on the collected information.

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], establishes 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 (MPEG) work

ISO/IEC JTC1 SC29, MPEG, has been working with ITU-T SG16 on video codecs for many years. MPEG has developed media transport technologies called MMT. Media coding technology called visual volumetric video-based coding (V3C) has been developed by MPEG as well.

ISO/IEC 23008-1

[b-ISO/IEC 23008-1], which was developed in MPEG, gives an MMT overview, and 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 [b-ISO/IEC 23008-1]. In other words, MMT cannot resolve synchronous media transportation.

ISO/IEC 23008-11

MPEG has also developed [b-ISO/IEC 23008-11], which is a composition information (CI) specification of terminal devices. Display of a real-sized object needs to be considered as CI; however, [b-ISO/IEC 23008-11] does not cover spatial environments such as bezel size of display devices and spatial position information. Therefore, CI is inefficient for displaying actual-sized objects.

MPEG-I Project

From the systems perspective, the MPEG-I project has laid out a basic foundation of coding, signalling, encapsulation and delivery formats for both mono and stereoscopic 360° video and audio in omnidirectional media format (OMAF) in 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.

ISO/IEC 23090-2

[b-ISO/IEC 23090-2], which was developed in MPEG, specifies the OMAF for coding, storage, delivery and rendering of omnidirectional media, including video, images, audio and timed text. An omnidirectional image or video can contain graphics elements generated by computer graphics but encoded as image or video. Multiple viewpoints, each corresponding to an omnidirectional camera, are supported. [b-ISO/IEC 23090-2] also specifies storage and delivery of overlay images or video intended for rendering over an omnidirectional background image or video.

ISO/IEC 23090-5

[b-ISO/IEC 23090-5], which was developed in MPEG, specifies the syntax, semantics and decoding for visual volumetric media using video-based coding methods. [b-ISO/IEC 23090-5] also specifies processes that can be needed for reconstruction of visual volumetric media, which can also include additional processes such as post-decoding, pre-reconstruction, post-reconstruction and adaptation. This enables the coding and processing for holographic images.

I.5.3 ISO work

ISO 9241-910

[b-ISO 9241-910] provides a framework for understanding and communicating various aspects of tactile or haptic interaction. [b-ISO 9241-910] defines terms, describes structures and models, and gives explanations related to the other parts of the ISO 9241-9XX subseries. [b-ISO 9241-910] also provides guidance on how various forms of interaction can be applied to a variety of user tasks. [b-ISO 9241-910] applies to all types of interactive systems making use of tactile/haptic devices and interactions. [b-ISO 9241-910] does not address purely kinaesthetic interactions such as gestures.

I.5.4 Institute of Electrical and Electronics Engineers work

In March 2016, the Institute of Electrical and Electronics Engineers (IEEE) Standards Association approved the creation of the IEEE P1918.1 standards working group. The core activities of the working group can be classified into three distinct areas: 1) definitions and use-cases; 2) haptic codecs; and 3) reference architectural framework.

Ongoing IEEE activity on haptic codecs for the tactile internet [b-IEEE P1918.1.1] is attempting to standardize the first kinaesthetic codec. [b-IEEE P1918.1.1] focuses on standardization of a corresponding tactile codec. [b-IEEE P1918.1] and [b-IEEE P1918.1.1] draft standards were planned fore completion in 2019. However, there is no codec standard published at the time of publication.

I.5.5 Work in broadcasting area

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

Association of Radio Industries and Business

[b-ARIB STD-B60] is a media transport standard based on MMT and [b-ISO/IEC 23008-1]. [b-ARIB STD-B60] aims to enhance terrestrial broadcasting services. However, as shown in clause I.6.2, [b-ARIB STD-B60] does not specify synchronous media transportation.

ITU-R

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

I.5.6 Work in cinema area

Today, four dimensional (4D) cinemas that provide highly realistic sensation for users have been developed and deployed all over the world. 4D means that the 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 that are equipped with special devices, with special programmes that are designed for 4D. Moreover, a 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 cannot be covered by this area.

I.5.7 Standardization gaps

Standardization of terminal devices decreases the cost of highly realistic IPTV services and standardization of the transmission issue provides interoperability between providers of services and content.

As previously mentioned, standardization of all key features of ILE services has not progressed, so it is necessary to start standardization work in ITU-T.

Bibliography

[b-ITU-T H.222.0] Recommendation ITU-T H.222.0 (2023), *Information technology – Generic coding of moving pictures and associated audio information: Systems: Various corrections and updates*.

[b-ITU-T H.264] Recommendation ITU-T H.264 (2021), *Advanced video coding for generic audiovisual service.*

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

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

[b-ISO 9241-910] ISO 9241-910:2011, *Ergonomics of human-system interaction – Part 910: Framework for tactile and haptic interaction*.

[b-ISO/IEC 23008-1] International Standard ISO/IEC 23008-1:2023, *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport (MMT).*

[b-ISO/IEC 23008-11] International Standard ISO/IEC 23008-11:2015, *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 11: MPEG media transport composition information*.

[b-ISO/IEC 23090-2] International Standard ISO/IEC 23090-2:2023, *Information technology – Coded representation of immersive media – Part 2: Omnidirectional media format*.

[b-ISO/IEC 23090-5] International Standard ISO/IEC 23090-5:2023, *Information technology – Coded representation of immersive media – Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC).*

[b-IEEE P1918.1] IEEE P1918.1, *IEEE Draft Standard for tactile Internet: Application scenarios, definitions and terminology, architecture, functions, and technical assumptions*.

[b-IEEE P1918.1.1] IEEE P1918.1.1, *IEEE Draft Standard for haptic codecs for the tactile Internet*.

[b-ARIB STD-B60] Association of Radio Industries and Businesses Standard ARIB STD-B60 version 1.14 (2019), *MMT-based media transport scheme in digital broadcasting systems*. Available [viewed 2023-11-20] at: https://www.arib.or.jp/english/html/overview/doc/6-STD-B60_v1_14-E1.pdf

[b-KT PC] Chosunbiz (2016). *Kt기자간담회.. 5G 홀로그램라이브* [KT press conference. 5G hologram live]. San Bruno, CA: Youtube. Available [viewed 2023-11-20] at: https://www.youtube.com/watch?v=cJgo-b4aEjE

[b-NTT JKF] SHOCHIKUch (2016). *『KABUKI LION 獅子王』遠隔舞台挨拶in Las Vegas*. [Japan Kabuki Festival in Las Vegas 2016]. San Bruno, CA: Youtube. Available [viewed 2023-11-20] at: https://www.youtube.com/watch?v=E0D67sBsCY0 .

[b-NTT Krause] Krause, J. (2017). *Cyber teleportation Tokyo at SXSW*. Austin, TX: SXSW. Available [viewed 2023-11-21] at: https://www.sxsw.com/news/2017/cyber-teleportation-tokyo-at-sxsw/

[b-NTT ICC] Intercommunication Center (2023). *Public booth for vibrotactile communication*. Tokyo: Intercommunication Center. Available [viewed 2023-11-20] at: https://www.ntticc.or.jp/en/archive/works/public-booth-for-vibrotactile-communication/

[b-Sohn] Sohn Ji-young (2018). Hard to follow cross-country skiing? 5G is here to help. *The Korea Herald*. Seoul: Herald Corporation. Available [viewed 2023-11-20] at: http://www.koreaherald.com/view.php?ud=20180213001003

[b-Tektonspace] Tektonspace (2017). *VR walk through*. San Bruno, CA: Youtube. Available [viewed 2023-11-20] at: https://www.youtube.com/watch?v=EUDTTLNLfDo

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

Published in Switzerland
Geneva, 2023