The Prospects of Sensory Transmission and UX in the 5G Era

Yuko Ohgishi
General Manager
Corporate Technology Strategy Division
Sony Corporation

Kazuyuki Yamamoto
General Manager
Application Tech. Dev. Field Tokyo Laboratory 01
R&D Center
Sony Corporation

1. Introduction

Fifth generation (5G) mobile communication systems are now being introduced in various countries around the world. Radio communication began in the early 1900s. It was initially used for military and special industrial applications, and was later developed as an important means of civilian communication. By the time third and fourth generation (3G, 4G) communication systems arrived, it had become a popular way of connecting people and had transformed our communication lifestyle.

In the 2000s, 5G is set to change the industrial world by providing a platform that supports objects-to-object communication and enables the transfer of large quantities of information, such as 4K, 8K and immersive VR video. This is expected to evolve not only as a simple communication platform, but also as an infrastructure to support everyday living and social activity. Compared to the time when radio communication was used for special industrial applications, the transmission cost per bit has become about a trillion times smaller. It is thus expected that radio communication in the so-called "Beyond 5G" era will be widely used as a system that supports society with services such as telepresence and telexistence.

Sony has conducted research and development on various 5G applications, and we have published the results of various demonstration experiments based on our R&D of key technologies and their applications with regard to "remote" operations, primarily in our main business areas of entertainment but also in various other sectors. We believe that we can contribute to the solution of diverse social issues by further developing the remote technology that has been realized in the pursuit of ultimate reality and real-time performance for the Beyond 5G era.

There are four major categories of solutions to social issues that can benefit directly from remote technology.

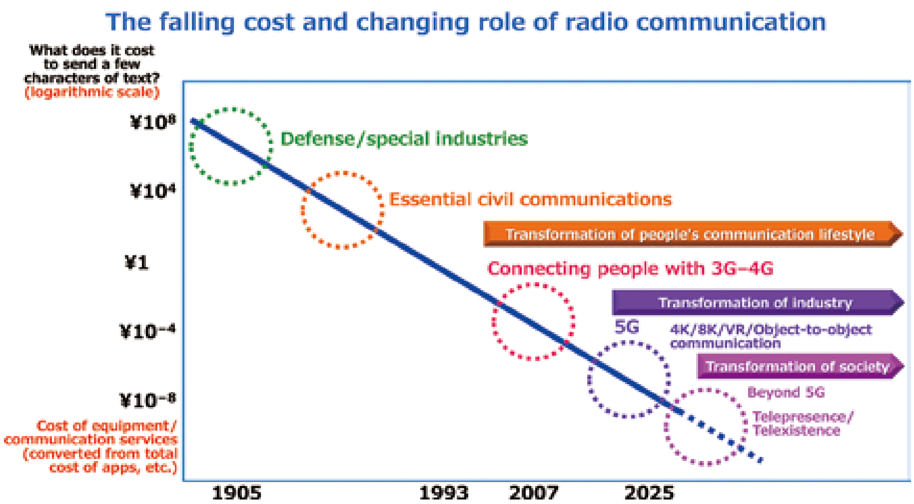
The first is energy consumption. It is considered that the resources, time and physical effort expended in moving people from one place to another can be reduced by developing remote technology, and that the realization of a society where it is unnecessary for people to move around will help to address issues such as the concentration of population in urban areas.

The second is climate change. Remote technology allows people to continue working in a comfortable indoor setting without being exposed to an outdoor environment of extreme high or low temperatures, heavy rain and strong winds all over the world.

The third is mitigation of the effects of natural disasters to ensure the safety of society. According to recent data, the numbers of natural disasters and disaster victims more than tripled between the 1970s and 2000s. It is expected that this will not only contribute to teleoperation activities in dangerous locations and the reduction of unnecessary and non-urgent excursions, but also lead to the realization of virtual experiences in extreme environments where people cannot go.

And the fourth category relates to the management of epidemics such as the one in which we currently find ourselves. A reduction in the movement of people is expected to help alleviate

Figure 1: The falling cost of radio communication



the everyday difficulties faced by people during the spread of an infectious disease.

In this article, we discuss UX proposals and verification tests that assume the availability of 5G/Beyond 5G wireless communication technology, and we discuss the key technologies that will be needed to realize these proposals.

2. Supporting remote technology

The realization of remote technologies such as telepresence and telexistence requires processes that are capable of sensing environments and information, performing signal processing on this data, transmitting it via communication networks, and displaying/presenting it to people and/or objects in visual, auditory, tactile or other forms. Figure 3 summarizes the key technologies required at each stage. In sensing, image sensors and other sensor devices are used to acquire information of various types, including not only images and positional data but also acoustic data from multiple microphones and tactile

Figure 2: From 5G to Beyond 5G

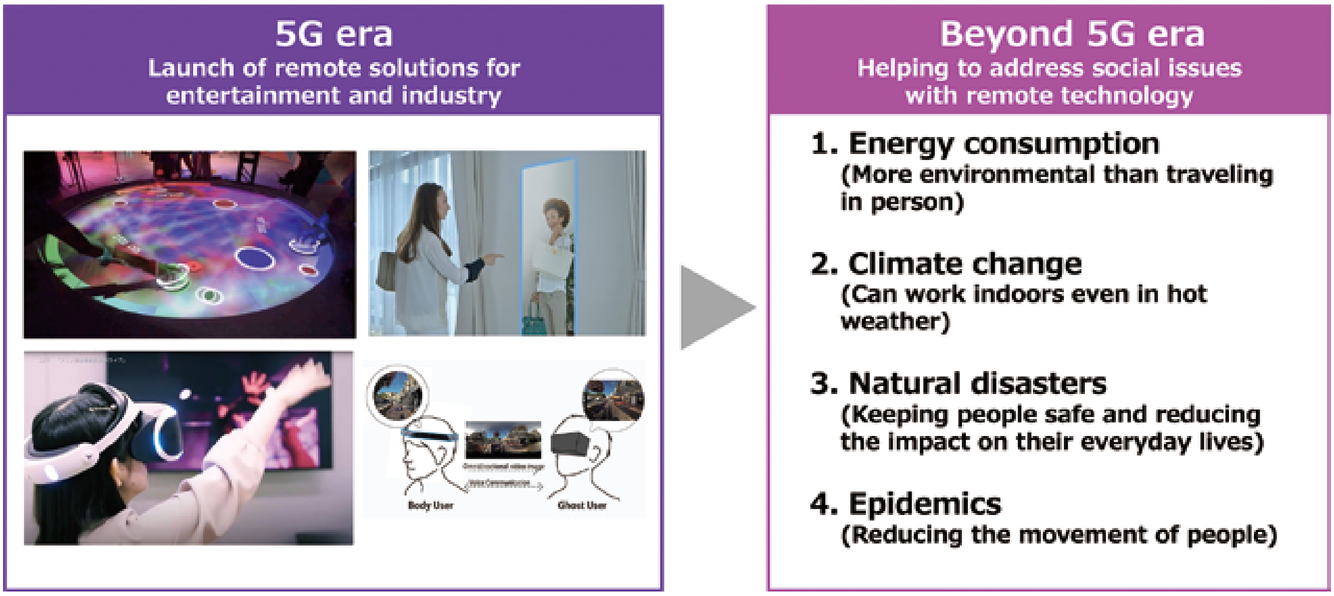
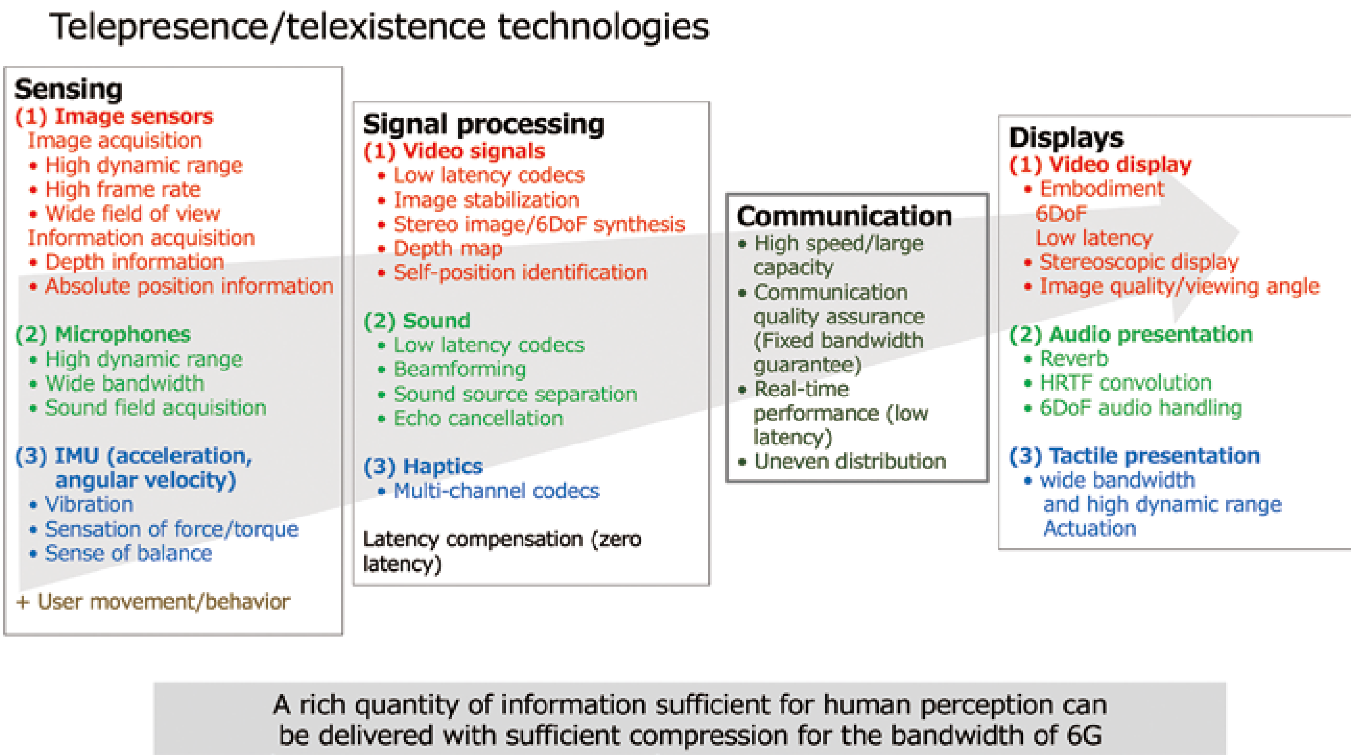


Figure 3: Telepresence/telexistence technologies



sense data from IMUs (inertial measurement units). The data collected in this way is subjected not only to appropriate signal processing, but also to processes such as compression to facilitate efficient transmission through the communication transmission environment at the next step. At the communication stage, due to the development of technologies such as 5G, it is possible to transmit large amounts of data at high speed with high quality and high reliability. This data is then displayed and presented via visual, auditory and tactile means in order to deliver a sufficient quantity of information needed for perception and intuitive understanding by humans.

Here, we introduce the elemental technologies that Sony is developing to help users achieve expanded perception and intuitive understanding.

2.1 Volumetric 3D video capture

First, we will introduce some examples of how images can be captured and displayed.

People intuitively understand their environment by observing it from different angles with six degrees of freedom (6DoF). However, much of this information is lost when images are presented on current 2D display devices.

We have therefore developed a technique whereby an object is simultaneously captured by multiple cameras from multiple different directions, and then image processing technology is used to combine these images into a 3D spatial image that can be transmitted and displayed as 3D data.

By incorporating this spatial three-dimensional information, we can produce video works that would have been impossible to implement with a single camera due to physical constraints and the difficulty of camera work. In this way, we can reproduce the 6DoF environment expected by users. Since this technique can reproduce features including the smooth surfaces of clothing with high quality, it allows data to be presented to users without compromising on the required amount of video information[1].

Figure 4: Volumetric capture of a dancing couple

2.2 Spatial Sound technology: 360 Reality Audio

For this sort of spatial display to provide users with a sense of presence, it is essential to support the intuitive recognition abilities of the user by using not only visual means but also spatial sound technology, which is necessary for the auditory sense.

At Sony, we have developed signal processing technology based on Object Audio technology that can record and play back sounds to reproduce a realistic sound field. We have also defined a music distribution format conforming to MPEG-H 3D Audio specifications in order to provide a new "360 Reality Audio" music experience that surrounds listeners with an immersive spherical sound field.

With this technology, content creators can arrange multiple sound sources at any orientation in this sound field, and can faithfully reproduce these sources during playback so as to provide the user with an acoustic experience that feels just like listening to the real thing[2]. We also plan to develop next-generation 360 Reality Audio content that delivers an even more realistic experience by expanding it to include video and VR content.

2.3 Wide bandwidth and high dynamic range haptic presentation

The sense of touch is perhaps the third most important of the human senses.

Tactile sensations can provide a user with information about the surrounding environment without disturbing the user's visual and auditory awareness, and can even enhance audiovisual information by interacting with the visual and auditory senses (cross-modal effect).

We have developed highly realistic haptic technology that uses audio technology to present tactile feedback having a wide bandwidth and high dynamic range in multiple channels.

This technology can be used to implement haptic presentations of various forms, ranging from palm-of-the-hand presentations to whole-body presentations. In this article, we will focus on the latter. Here, we introduce a technique for reproducing a rich

Figure 5: 360 Reality Audio

360 REALITY AUDIO

virtual haptic experience by using multiple VCMs (voice coil motors) in a vest worn by the user to present vibrations with a wide bandwidth and high dynamic range in synchronization with video and/or audio content. This results in a more immersive experience than can be achieved with audio and video alone.

Technologies that contribute to the sensing and reproduction of the three senses that make the largest contribution to feelings of immersion and reality (sight, hearing and touch) have already been implemented in practice[3].

■ Figure 6: Users taking part in an immersive haptic experience

2.4 Presenting information on output devices with 6 DoF

To reproduce realistic sensations, it is important to consider how the acquired information can be correctly presented to stimulate these three senses simultaneously.

Here, we introduce two types of video device that can display 3D video information (including the abovementioned volumetric video) with a high degree of realism.

The first is an eye-sensing light field display with eye position recognition that allows the user to view images in 3D without having to wear special glasses. This device reproduces a realistic 3D virtual image inside a box, with the viewpoint of this image constantly adjusted to match the viewpoint corresponding to the position of the user's eyes[4].

The second is an optical see-through augmented reality (AR) HMD (Head Mounted Display) [5].

An AR display overlays virtual objects on real-world places or objects, thereby presenting users with additional information in a way that can easily be intuitively understood. When the user's viewpoint moves, even slight misalignment between real objects and overlaid virtual objects can cause them to lose their sense of presence and prevent users from understanding them intuitively. The main causes of this misalignment are errors in the viewpoint pose estimation, and the processing latency from sensing the user's eye position/gaze direction to drawing virtual objects in the AR visor.

To mitigate this problem, the AR visor performs pose estimation with a combination of an image sensor and an IMU to achieve both highly accuracy and low latency, and the rendered virtual image to be overlaid on the real world is modified immediately before display, depending on how the pose of user's viewpoint changes while rendering. As a result, it is possible to create overlaid images without any apparent delay. This technique is called latency compensation.

Experimental measurements of how the displayed positions of real and virtual objects change over time are shown in the graph below. Without latency compensation, the position of the virtual object lags behind the position of the real object by an amount corresponding to the processing delay, but with latency compensation it can be seen that the positions of the real object and the virtual object are almost identical. This makes it possible to line up the positions of real and virtual objects in a way that appears much more natural[6].

3. UX demonstration experiments

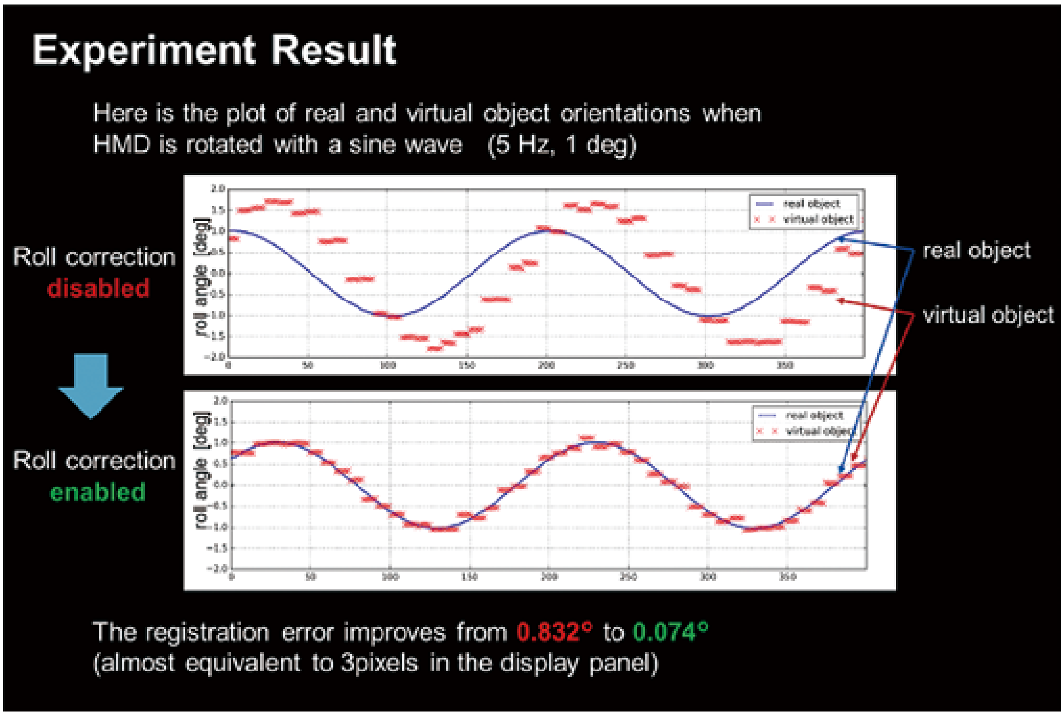
As an application that integrates the key technologies described above, we conducted demonstration experiments on several application cases.

In our research and development of these sorts of key technologies we have repeatedly built prototypes to evaluate what kind of user experience can be obtained by applications that

■ Figure 7: Eye-sensing Light Field Display

■ Figure 8: Optical see-through AR visor

■ Figure 9: Results of latency compensation experiment



combine these elements, and we have performed experiments to determine the necessary requirements and specifications. Some examples are presented here.

Virtual reality (XR) remote live experience

To develop this system, we teamed up with developers of various technologies, experts and artists in the entertainment field such as music and games to build and demonstrate XR entertainment prototypes that allow users to experience live music from a remote location.

Through prototyping, we are able to not only check the progress of XR technology under development, but also to accumulate know-how related to content production. For example, we found that a five-piece band and a solo singer require different shooting methods and expressions, and that the addition of game-like elements allows users to participate in live events with a greater sense of accomplishment. We have also been able to gain insights such as discoveries relating to the perceived value of new

■ **Figure 10: A virtual reality (XR) remote live experience**

■ **Figure 11: JackIn Head**

XR experiences[7]. For example, music fans do not feel particularly involved in a concert if they are situated too far away from the performing artists, but conversely feel more involved when they share a virtual environment that adapts to the direction in which they are looking.

JackIn Head

We are currently developing a technology called JackIn Head, which offers a way of sending omnidirectional video images from a "body user" (equipped with a wearable 360° omnidirectional camera) to a remote "ghost user" (who can watch this video on a screen or head-mounted display). The body user and ghost user can also communicate via a two-way voice connection. This system should enable the provision of services ranging from synchronous travel experiences and remote assistants to the real-time delivery of special experiences, sports-based entertainment, and professional training. It could also be used in situations where specialized knowledge is required such as medical cooperation efforts, or in disaster-affected areas[8].

Telepresence system

Next, we will introduce some examples of the development and demonstration of a telepresence system aimed at facilitating natural communication between people as if they were all present in the same space. This system uses a 4K vertical display to display images from remote locations in real time. By making use of sound quality enhancement and echo cancellation technology, it allows users to converse with people in remote locations just as if they were in the same room together.

In various demonstration experiments, we found that it is important to match the user's line of sight with the camera axis in order to achieve a realistic sense of communication with the remote images. We also found that hand gestures play an important role in communication, and that it is possible to provide a sense of unity by sharing information with no explicit purpose, such as background video and ambient sounds[9].

■ **Figure 12: Telepresence window**

AR applications

Next, we will introduce two examples of AR applications based on projection.

Both of them use system configuration and prediction algorithms that minimize the latency between the user's movements and the projection images in the same way as the AR visor described above, so that there is virtually zero lag between the user's movements and the displayed motion.

The Doodle Pen AR application allows users to draw virtual letters and pictures anywhere in the environment with a digital pen. By eliminating as far as possible the latency and misalignment between the movements of the user holding the pen and the characters and images drawn using this pen, this system allows the user to concentrate on content creation instead of having to keep thinking about how the system behaves[10].

We also built a prototype A(i)R Hockey application that combines haptic technology with zero-latency projection to give players the impression that they are really hitting a virtual puck. As the game progresses, the players tend to start losing track of whether the pucks they are hitting are real or virtual[11].

In projection examples, including the AR visor mentioned above, it is possible to reduce latency and improve quality by making the position sensing, information processing and prediction algorithms run as fast as possible.

Although we were able to improve the user experience by

■ **Figure 13: Doodle Pen**

■ **Figure 14: A(i)R Hockey**

minimizing latency in the examples shown here, we have also repeated these experiments in order to understand the network conditions that are required for these applications.

With the AR visor mentioned above, we are conducting a demonstration experiment in which multiple users share the same superimposed content while constantly communicating and interacting with one other, not only indoors but also outdoors. When used indoors, the devices are covered at the Wi-Fi level, but when users take them outside, the amount of communication increases greatly, even within the controlled space, due to the significantly increased reception of data such as background image information. This traffic requires 5G connectivity. If the devices are used in arbitrary locations that are not confined to a controlled space, then this exchange of information becomes extremely large, and would require communication with an even wider bandwidth.

4. Conclusion

5G has evolved from a conventional communication infrastructure to part of the social fabric, and its successor in the Beyond 5G era is expected to play an even more central role in supporting society's infrastructure. The digitization and low-latency transmission of information directly related to human senses are key requirements. In particular, in systems that integrate cyberspace with the real (physical) world and obtain feedback through communication based on sensing of people and things, we expect to be able to adapt to the diverse needs of people and industries by exploiting AI technology through advances in communication such as high speed/capacity, ultra-low latency, and the ability to handle large numbers of simultaneous connections. In the future, we will continue to work on core technologies for remote applications, and on field trials of applied technologies. We will also pursue the ultimate level of reality and real-time performance in applications combining images and sounds, and even tactile sensations. In this way, we hope to deliver experiences that exceed the expectations of users. Also, for the forthcoming Beyond 5G era, we hope to develop remote technology that works more closely with people to help solve various social issues.

[Reference URLs]
[1] https://www.sony.co.jp/SonyInfo/technology/activities/SonyTechnologyDay2019_demo2/
[2] https://www.sony.com/electronics/360-reality-audio
[3] https://www.youtube.com/watch?v=DMY4LB-4-24
[4] https://www.youtube.com/watch?v=VIPEckQ9wnk&feature=youtu.be
[5] https://www.sony.net/SonyInfo/design/stories/ghostbusters/?ui-ux
[6] https://doi.org/10.1002/sdtp.12923
[7] https://www.youtube.com/watch?v=Dmo5Hq4ROBk
[8] https://www.sonycsl.co.jp/project/2373/
[9] https://www.sony.co.jp/SonyInfo/News/Press/201810/18-085/
[10] https://www.youtube.com/watch?v=qE6tAoaPhyo&feature=youtu.be&list=PL7tWrqXC5RzO4yWWjd814dC7si4-rgaih&t=239
[11] https://www.sony.net/SonyInfo/design/stories/AiRhockey/?ui-ux