Electronic Preprint for Journal of Information Processing Vol.25

Invited Paper

Virtual Reality Technologies in Telecommunication Services

Hidenobu Nagata[1,a) Dan Mikami[2,b) Hiromu Miyashita[1,c)
Keigo Wakayama[1,2,d) Hideaki Takada[1,2,e)

Received: July 5, 2016, Accepted: December 13, 2016

Abstract: Telecommunication service has been growing and progressing from telephone to high reality communication systems that are based on evolution of network and media technologies. Recognizing virtual reality (VR) as a communication tool, we provide a review of communication services and the directions they are moving in, as well as related VR technologies. The Immersive Telepresence System "Kirari!" is also introduced as the latest development example for a new telecommunication service.

Keywords: virtual reality, communication service, telepresence, experience

1. Introduction

Many times in the past, a certain year has been designated "the first year" of something or other, and 2016 has been called the first year of virtual reality (VR). The latest excitement in the VR field comes from market growth centered around low priced head mounted displays (HMDs) such as Oculus Rift [1], HTC Vive [2], PlayStation VR [3] and GearVR [4], which enable users to enjoy high reality contents. Recent years have also seen the development of theme parks that provide visitors with VR experiences. Today, not only games with HMDs but also VR technology as a whole are gaining increasing attention.

Virtual reality (VR) generally refers to computer technologies that can generate perceptual feelings as well as actual feelings from virtually created cyberspace, i.e., space that does not physically exist. It is also considered as a tool for the three C's (creation, control, and communication) and the three E's (elucidation, education, and entertainment) often cited in relation to human activities [5], so it covers a wide area of applications. It can be considered that one of the most important "final destinations" of VR is to provide multi-purpose communication tools [6], [7]. Technical development of the latest communication systems is obviously one of the VR research activities for providing users with a higher sense of realism.

Providing a communication tool, a way of enabling users remote from each other to communicate by transmitting audio and visual data via network systems, is the most fundamental role in telecommunication service. There are mainly three functional levels, 1) sensing environment data and capturing them as audio and visual signals at the origin point, 2) transmitting them to target locations, and 3) reproducing them so that they can be recognized. In the past most conventional telecommunication services have clearly been putting the highest priority on providing accurate information, i.e., reproducing and expressing words correctly. Today, high resolution cameras or displays such as 4K/8K for video and multi-channel for audio are being used to provide greater accuracy in telecommunication systems.

Recent explosive progress in network transmission capacity and computer performances is bringing more potential for realistic telecommunication services. This progress has led to devices that create a high sense of unity between users by providing multi-sensory information that includes things such as tactile feelings and even smells. Moreover, along with these technology developments toward accurate information transmission in telecommunication service, new research work involving converse approaches have been recently started. With these approaches, captured information is distorted or processed on the basis of its purpose, finally enabling it to be more appropriately conveyed. A typical example of this is found in VR utilization in sports motion training, and some studies reported that through its use athletes had improved their training efficiency by using the reduced amount of (rather than all) captured information. The concept in these new approaches has common with the Immersive Telepresence "Kirari!" concept [8], such that the perceptual feelings, understandings and even sense of reality of users can be controlled and maximized by artificially post-processing the original captured data.

In this paper, we provide a survey of VR technologies related to the expansion of communication service areas. The rest of the paper is organized as follows. First, as the most fundamental telecommunication service, interpersonal telecommunication

1 NTT Service Evolution Laboratories, Nippon Telegraph and Telephone Corporation, Yokosuka, Kanagawa 239–0847, Japan
2 NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Yokosuka, Kanagawa 239–0847, Japan
a) nagata.hidenobu@lab.ntt.co.jp
b) mikami.dan@lab.ntt.co.jp
c) miyashita.hiromu@lab.ntt.co.jp
d) wakayama.keigo@lab.ntt.co.jp
e) takada.hideaki@lab.ntt.co.jp

is described in Section 2 along with a mention of related VR research. Section 3 introduces a specific target-oriented communication service for education or knowledge transferring, along with an application example in sports motion training using VR. In Section 4, we describe a new approach in communication for maximizing users' internal sense of presence, along with an actual case example. Section 5 concludes the paper with a summary and a discussion of further VR applications in telecommunication services.

2. Interpersonal Visual Communication

Interpersonal communication between users at remote locations is the most fundamental role in telecommunication service.

The evolution of network, media, and information compression technologies and multimedia equipment such as cameras and high resolution displays has enabled rich communication systems that transmit extremely rich sensory data. These systems are able to provide not only gestures but also more emotional and atmospherical forms of expression on the part of users, even including their breathing. Recently, users have been able to feel a higher sense of unity through the use of telepresence products provided with tables and effective light equipment along with high-end cameras and displays.

In this section we focus on interpersonal communication as the most fundamental role in telecommunication service. VR related technologies are reviewed along with requirements for providing basic interpersonal communication and also higher reality. We put emphasis on three key points, which are important either for achieving higher reality, awareness, immersion, and spatial perspective or feelings of space.

2.1 Awareness

Mutual understanding of intentions and situations is necessary to make appropriate conversation. Awareness, which is one of the most important factors in personal communication, is defined as understanding the activities of others that affect one's own activity [9], [10]. It is also said that sharing participants' activities and situations is important for establishing communication.

Awareness is essential for *CSCW* (*Computer-Supported Cooperative Work*), which means computer systems for supporting collaboration among multiple users, because deep mutual understanding among users is required for effective work.

While user and workspace images were simply displayed on two GUI windows in the early stages of research, many new concepts aiming to achieve more seamless awareness have been generated and systems such as real-time workspace overlay displays have been developed [11], [12].

As the saying *"The eye is the window of the mind"* goes, gaze is considered quite an important factor that provides much valuable information for establishing communication. Thus gaze patterns in face-to-face communication were studied and applied to a gaze awareness based communication system [13].

Ishii et al. has designed a shared drawing medium as a glass board metaphor that displays the partner's face and workspace reflected on a half mirror [14]. This system allows the user to detect eye contact and monitor the partner's direction of gaze.

Face-to-face meeting systems with a surrounding table have also been presented. In these systems, gaze and visual attention from one person to another can be recognized through the use of life-size user images with relatively accurate position [15], [16].

A study aimed to activate informal communication has also been presented [17]. It reported that audio and visual data were continuously interexchanged via life-size displays in the workplace. Participants were always aware of each other and thus were able to communicate with shared feelings of being in the same place.

For another example, flexible moving displays that show a user's face to support eye contact and gaze awareness have been presented [18], [19]. Screen pose and position flexibly move automatically by mirroring the remote user's head motions, which enables the transfer of more correct information including users' behaviors such as head nodding.

2.2 Immersion

Many techniques have been researched for enabling multiple participants to feel that they have been immersed into a shared virtual space. Instead of using a flat display, IPT (immersive projection technology) and HMD (head mounted display) techniques are widely used in display systems that can create more immersive virtual space.

The CAVE, which is known as one of the most representative IPT display systems, has four surrounding screens. It was based on perspectives acquired via a viewer-centered headtracking system, and uses LCD stereo shutter glasses [20], [21], [22]. The GAVA was developed as a CAVE based system for remote communication service with the concept of space sharing, including sharing the atmosphere. The avatar is controlled by motion tracking and an ambient sound system using stereophonic sound field reproduction technology [23], [24].

In a VR project called Multimedia Virtual Laboratory, five screen CABIN and six screen COSMOS systems in Japan were connected via a gigabit network. To share users' behavior and emotional facial expressions correctly, users' images were provided as 2.5 dimensional video avatars created on the basis of depth information [25].

Surround systems are known as audio systems that are very suitable for IPT. They use multiple speakers sparsely located around users for providing a sense of direction and a listener environment. Dolby Atmos, which is a surround system with an object based rendering system, has an original GUI for flexibly projecting sound images onto any direction [26].

The 22.2ch multi-channel audio system, which has the largest number of channels among surround systems, brings higher effectiveness in acoustic perception than a surround system such as the above-described Dolby Atmos [27].

Portable spherical microphones, which have sound receiving elements inside that are isolated by reflection panels, have been provided to capture ambients with their directional components [28].

Using higher order ambisonics (HOA) [29], [30], [31], [32] makes it possible to reproduce a sound field recorded with a spherical or circular microphone array [33], [34], [35]. This enables users to perceive rich, high-level information of individual subjective impressions such as presence, verisimilitude, realism and naturalness at microphone array positions in remote spaces [36]. With the aim of walking in a remote space while keeping this perceived information, we can also reproduce a sound field simulated with source signals and sound source positions that are also obtained using various types of microphones [37]. Even enhanced monaural speech signals generated by a beamforming technique can be used as source signals [38], [39].

Wider fields of view can be achieved for VR systems with HMDs, by controlling displaying content depending on tracked head movement. In image creation, a high sense of immersion can be provided by using a remote omni-directional camera [40], [41]. Recently, HMD systems can be easily created by using the latest smartphones that have sensors [42].

Binaural reproduction is known as an audio presentation method that uses a headphone environment and is very suitable for HMD-based VR [43]. The driving signals of headphones are generated by the driving signals of virtual loudspeakers and head-related transfer functions (HRTFs) [44].

Using binaural reproduction via headphones, sound images can be set on any place around a user and also a realistic ambient can be presented. Therefore, a sense of distance of multiple sound images and immersion in audio can be perceived by the user. Even higher immersion can be provided by using binaural reproduction with videos via HMDs [45].

Another example is a stereoscopic type immersive display, which uses rotating display units consisting of LED arrays and a barrier around a viewer. Gaze awareness, which is an important factor in interpersonal communication, is not lost in this system because a user's face can be captured from outside the display [46].

Focusing on the feeling of being in the same room, Hirata et al. studied the symmetrical reproduction of communication rooms [47]. They set life size displays and cameras so that they would compose a cylinder space. Users' images were captured in front of a display and shown on a corresponding display at a remote place.

2.3 Depth Expressions for Perspective

Along with immersion, awareness of three-dimensional space around a user is also important [48]. This leads to feelings of space existence that highly affect users' perceptual feelings of each other's existence. In this section, we focus on techniques for depth expression either in audio or video.

To provide communication with natural perspective feelings in an autostereoscopic manner, binocular disparity, motion parallax, and continuous space expressions are used. Some studies have made efforts to construct three-dimensional face presentation as the most important point. In one study, for example, the face of a remote user was displayed as a three-dimensional object using a spinning display surface with a two-sided tent shape [49], which enabled eye contact to be established between users [50].

For depth presentation, depth-fused display (DFD) technology has been presented. A DFD consists of two layered screens and depth can be presented by differences in luminance between the same object on two screens [51], [52]. Relative face and eye positions can be changed and followed by an observer, enabling users to intuitively recognize the displayed character's gaze direction [53].

Transaural systems, which enable binaural audio reproduction using multiple speakers instead of having users wear headphones, was presented as a technique that is very suitable for DFD displays [54], [55]. High presence communication environments are enabled using a transaural system and robots in a space, because the sound source can be located virtually on an actual space without the need for wearing any devices.

Motion parallax is used in many communication devices that are based on life size 3D user images. Using image synthesis makes it possible to naturally reproduce relative user and background positions, depending on where the positions are [56]. A viewpoint interpolation technique using DFD has also been presented [57].

Object-based wave field synthesis (object-based WFS), which uses multiple surrounding speakers, has also been presented [58], [59]. Various implementations in wave synthesis have been studied, based on research in audio perception [60]. This technique can be used to accurately reproduce audio waves in physical space and since it can cover multiple users, it enables the creation of communication space with a higher sense of existence than transaural systems can create.

2.4 Space Sharing

For reproducing the sound field of a remote location in a space by using many microphones and loudspeakers, many researchers have proposed methods based on the boundary surface control (BoSC) principle [61], [62], methods based on a wave field reconstruction (WFR) filter [63], [64], and other methods.

Systems based on the BoSC principle are implemented by mounting loudspeakers on a wall inside a system called a "Sound Cask" [61], [62]. By getting audio signals with a microphone mounted inside "Sound Cask" and reproducing a virtual sound source inside the system, users can get the sense that other people are playing a musical instrument right next to them, thus enabling an ensemble to be created by two people in remote locations.

Systems based on a WFR filter are implemented by using a linear microphone array and a linear loudspeaker array. Because the driving signals of the loudspeaker array are calculated on the basis of an analytically derived filter, the systems can transmit sound fields to remote locations with a low delay [63], [64]. Sound images can be represented in front of loudspeakers by applying an inverse wave propagator [65], and then an original space and a destination space can be overlapped. A duplex communication system can be developed by using such sound field transmission systems and implementing multichannel acoustic echo cancellation [66], [67]. Such a system would enable users to perceive the depth of sound images and get the sense that they themselves were at the place.

Systems based on a WFR filter are also implemented by using a circular microphone array and a circular loudspeaker array. These systems can reproduce sound fields arriving from all directions in a horizontal ear-plane by limiting the target area [68]. From this a destination space can be overlapped with a part of an original space, thus enabling users to join a conference in a remote location with the sense that they were actually there.

To achieve realistic remote interpersonal communication, the perceptual reality experiences in VR must be similar to real-world face-to-face communication experiences. Studies have been made on awareness, immersion and perspectives, and many systems have been developed with the aim of achieving high sense of presence in interpersonal communication. Recently, remote space sharing has been presented as a new concept and sound space sharing systems based on sound field reproduction techniques have been presented to enable this concept. In the future, adopting light field displays to these space sharing systems will make it possible to achieve more natural interpersonal VR communication. Evolution in display resolution and device miniaturization are also greatly contributing to realistic remote communication.

Along with these evolutions and directions of VR technologies for achieving interpersonal realistic communication, new requirements for transferring information have been pointed out. New and challenging research work, which we describe in Section 3, has started with the aim of achieving broader types of telecommunication applications.

3. Target Oriented Communications

So far, we have overviewed VR systems for providing communication services, mainly focusing on telecom services such as teleconferencing. The key point for these systems is the way in which they transport auditory and visually correct situational data from one place to another. The systems may find it difficult to transfer everything, but they strive to transfer as much information as they can.

On the other hand, there is some information that is difficult to transfer even when using face-to-face communications. A typical example is "motion." Thus, this section overviews VR systems for transferring and/or acquiring motion under an extended definition of "communication." Although most of the research work done in this area has not focused on the telecommunication aspect, we believe that it has led to many important suggestions on how to further develop target-oriented communication systems. In the subsections below we will start by giving the reason VR systems are required for sports training. We will follow this up by showing examples of the information (including feedback) the systems present to users and how the information affects their motor learning.

3.1 VR Systems for Motor Learning

To date numerous VR systems for learning and acquiring motion have been proposed and their target motions have had a wide variety, ranging from dancing [69], [70], [71] and sports [72] to rehabilitation [73], [74]. A lot of advantages are indicated for such VR-based training systems. The first one is flexibility. During motor learning, trainees need to acquire multiple skills simultaneously. VR systems can easily control the target issue of the current training and make changes in other settings. In addition, they can easily create target situations requiring multiple participants. This makes it possible for users to take part without the participation of others, i.e., by using virtual participants.

Here, we would like to emphasize that information that is important for trainees is sometimes different from that provided by telecommunication services.

3.2 Providing First Person Vision for Training

Some systems have attempted to provide an athlete's first person vision under a VR environment. For example, baseball [75], handball [76], table tennis [77], and rugby football [78] were employed as the target sports. In such ball games, a trainee needs to respond to an oncoming ball. Thus, enabling the users to feel the ball trajectory stereoscopically is quite important, and these systems use devices that can display 3D information, such as CAVE and HMD.

One study examined how the information provided was applicable to the training for a handball keeper [76]. This study compared the differences in how a keeper responds to an actually thrown ball and how the keeper responds to a ball thrown by an avatar. The obtained results revealed that the keeper's motions do not change significantly for either case. These results lead to the conclusion that a VR system can be used to improve a handball keeper's performance.

In a recent study, VR was also applied to assist baseball batters in their training [75]. In this study, the batters wore a HMD that enabled them to see the ball trajectory stereoscopically. In the virtual 3D space, a 2D-based representation was used for the pitcher which means that pitching movie is shown on the flat panel placed on the pitching mound area in 3D virtual space (**Fig. 1**). After the batters had undergone the training, they were asked to fill in a questionnaire. Their responses indicated that the 2D representation had had a very insignificant effect on their quality of experience. They requested that the VR content be constantly updated.

These systems, which attempt to provide trainees with experiences they will confront in actual game situations, try to make trainees perceive the same things they will perceive in real life situations. On the other hand, interesting findings have been obtained with other systems. In observing and analyzing motions, observers felt the use of 3D CG-based representation speeded up the opponent's motions relative to what they would have been with simplified stick figure models [79] (**Fig. 2**). The authors believe that this finding may lead to more effective training under VR environments.

Fig. 1 Virtual reality (VR) baseball training system.

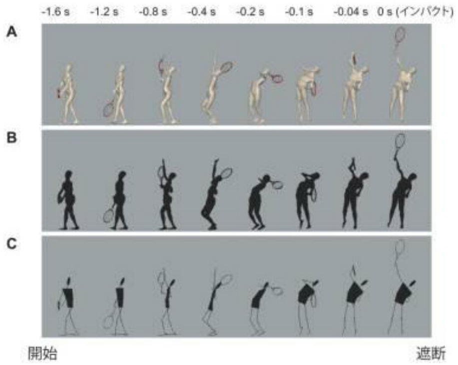

Fig. 2 Changes in cognition by presentation simplification.

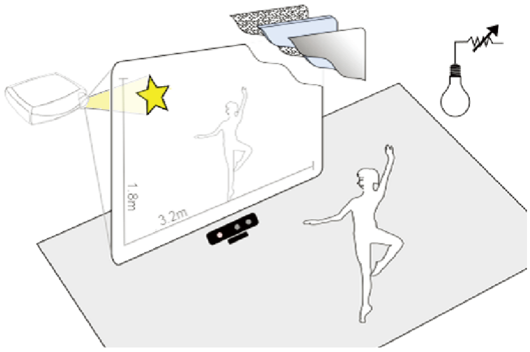

Fig. 3 YouMove.

3.3 Information Feedback to Trainees

One of the most important research topics in sport/rehabilitation supporting systems is the way of providing information feedback. This includes analyzing efficacy by using various feedback methods and sensing the feedback. Recently, it has been expected that real-time feedback can be obtained during the time trainees are performing motions and many studies have developed real-time sensing and feedback systems on the basis of this point. From this point on, we will discuss both the architecture of such systems and the representation effect they provide.

For example, a motor learning system called YouMove [69] has been proposed. As shown in **Fig. 3**, it displays feedback information onto a layered screen. During the motor learning process, the amount of feedback information decreases. Chua et al. proposed a system for training persons on how to use "tai chi", a Chinese martial art [80]. In their system, the body postures of trainees and trainers are displayed on HMDs in a real-time manner. They analyzed the effects of motor learning changes from the viewpoint and relative positions of trainee and trainer. The study results they obtained unexpectedly indicated that overlay representation of trainee and trainer motions is ineffective for motions requiring full body coordination.

Here, we will show another example, one for which intuition is unsuitable: motor learning of cyclic motions such as dribbling a basketball. This example indicates that excessive feedback deteriorates the efficacy of motor learning. It shows that it is more efficient to provide visual feedback every four or five times the motion is performed than to provide it every time the motion is performed [81]. We believe this gives an important insight into the further development of representation methods in VR environments.

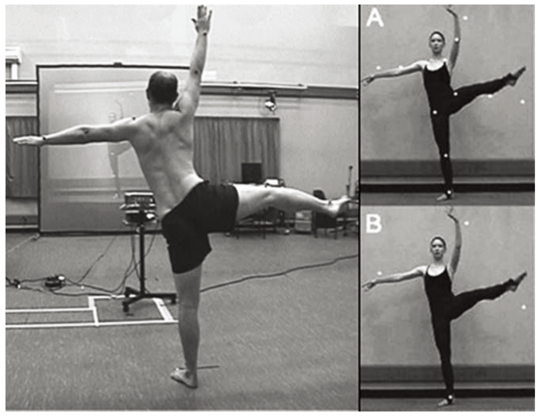

Fig. 4 Full-FB vs. Reduced-FB.

The above study can be considered as one that provides temporally reduced feedback. On the other hand, spatially reduced feedback, which gives only a limited portion of an object, has been proposed and its efficacy suggested [71]. This study employed ballet as a training target; it compared the efficacy of three feedback conditions: full, reduced, and none. **Figure 4** shows a trainee's motion (left) measured and overlaid to a reference motion; (A) shows full feedback and (B) shows reduced feedback. It shows the best result was obtained for the latter.

Most of the studies introduced in this section were examined under a laboratory setting and sometimes in actual practice the efficacy was not verified under consecutive motor learning conditions. We think it is quite interesting that some studies aim at representing all information correctly and that they verified the method's effectiveness. At the same time, it is interesting to note that reduced feedback improves motor learning because trainees tend to lapse into information overload. Even for telecommunication services, reducing the amount of information provided will become one good option for concentrating on the purpose of communication.

4. Communication for New Experience

So far, we have provided a review of VR technologies for basic interpersonal communication tools and purpose-oriented communication tools designed for transferring information that is difficult to transfer even by face-to-face communication.

In this section, we will introduce the latest VR technologies we have developed to achieve high perceptual presence and provide new experiences. These technologies are designed to provide a wide variety of telecommunication applications, including entertainment, in the future.

4.1 High Presence

Thinking about future directions in VR research, here we would like to once more consider high presence in telecommunication services.

Higher presence leads to realistic feelings, expressed lexically as "being there feelings." In conventional telecommunication services, information about external environments has been transmitted as accurately as possible to derive high presence. This means that high presence in telecommunication services is mainly considered as providing correctness in presenting the external world, which is sensed by users' senses of vision and hearing.

In contrast, sense of presence derived by psychological factors based on users' past experience or knowledge is also known as another important factor for presence. Thus, sense of presence is derived by both physical and psychological factors [82].

In many VR studies in basic telecommunication services, physical factors such as size of object, space, or sound level might well be considered in evaluations for presence level. Even though these evaluations were done on the basis of physical factors, they were able to provide correct results to some extent in presence measurements because psychological factors have a high correlation with physical factors.

However, it is known that physical accuracy is actually not the top priority in human cognitive functions [83] and there is a strong tendency to accept information as true, if the incoming information is sensory and is responded to promptly [84]. This tendency of acceptance is obviously seen especially when users are not aware of the physical correctness of each information item presented. It is also seen in the field of entertainment, where users have little interest in physical accuracy. Moreover, if the audio and visual information provided fulfills users' expectations, they may get feelings of high presence even if they are shown the information which was purely artificially created.

Also, effectiveness for controlling information based on specific purposes provides higher effectiveness in understanding as described in Section 3. The examples shown in that section also indicated that physical accuracy is not always critical in various communication services. In this section, we provide a review of new developments aiming to achieve high presence by focusing on psychological factors.

4.2 Presence Expression Using Psychological Factors

Providing or adding artificial information can often fulfill users' expectations, and this is known as one of the ways to achieve a higher sense of reality. We found many examples of this in content authoring work performed in entertainment services. For example, sound effects and color control are done in creating TV and movie scenes on an almost routine work basis.

Sound effects are known as the most representative process for creating illusions. Enhanced or artificially created sound is inserted or added to fulfill users expectations or to express the extent of space, and finally a scene providing a high sense of reality is created. Another example can be found in 3D animation creation. Using the virtual viewpoint control technique makes it possible to flexibly and dynamically express the depth perspective often used in drawing-based 2D animation creation in 3D animation creation [85]. These examples indicated that rather than physical correctness, users' impressions and the perceptual reality that follows can be controlled via presented audio and video data, which is artificially created to fulfill the expectations that users develop on the basis of their experiences or knowledge. Moreover, it can be considered that users' perceptual reality derived from psychological factors can be controlled directly.

In the next subsection we will introduce our newly developed Immersive Telepresence "Kirari!" system, which aims to provide high reality VR communication service on the basis of proactive artificial effects. It is based on multiple telecommunication technologies, including media processing, transmission, and presentation.

4.3 Immersive Telepresence "Kirari!" System

Transporting sports participants or actors from remote places to a position right in front of the audience in real time - this is the basic concept of "Kirari!" [8]. The authors expect that this concept will be able to provide users with a unique and high presence experience. Even though the concept is physically impossible, it may be able to bring about totally new experiences, thus leading to new ways of expressing reality expressions in telecommunication service.

By way of background, this type of presentation has recently become more popular and widely accepted by audiences in the entertainment field. In new types of stage production methods, digital based artificial effects are frequently used. In this way actors, CG based digital characters, and even real objects appear to be naturally interacting and existing in the same place [86]. While this is clearly physically impossible, there is a type of high presence with surprise feelings. It can be considered that the sense of presence in this case is derived by natural and prompt interactions among actors and stage objects, all of which are presented in life size. Moreover, it is considered that when there are inconsistencies in one's sensory elements or there is situational duality such that "there is something (somebody) there that shouldn't be," they could also be a factor in providing a sort of presence [82].

Extracting the essence from entertainment cases and human cognitive studies, we think that it is possible to provide a new type of presentation in our telecommunication services, one that will provide a higher sense of presence in communication as well as entertainment. In the following subsection we will describe the Kirari! system, which was developed with the idea of providing new experiences in telecommunication service, from the overall to the individual technologies it uses.

4.3.1 System Overview

Here, we will describe an overview of the "Kirari!" system. As shown in **Fig. 5**, it was designed by using several multimedia technologies to enable new experiences in live demonstration in various use cases. Audio images, images of targets (e.g., athletes, actors, or reporters to be displayed as target images to be transmitted), and background images, are captured and processed separately. Through the use of a protocol for synchronizing they are transferred via a network and displayed as dimensionally affected images at a remote place. In the research reported here, we

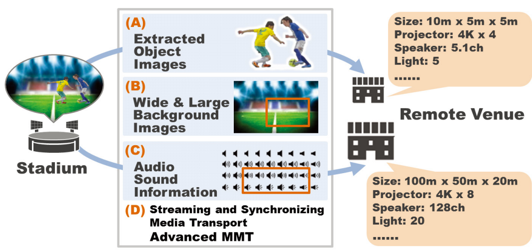

Fig. 5 Concept overview of the "Kirari!."

selected a specific display that makes it possible to present target images as realistic objects based on a luminance gap and an optical illusion called "Pepper's Ghost visual effect." The display system has two 2D image planes with different depths.

4.3.2 Realtime Object Extraction

In this area we have been researching an image processing technique for real-time, precise image extraction that is robust against background conditions. Today objects can be extracted in real time without using a green screen, if the background texture is relatively simple and luminance and color gap between object and background are appropriately created in the shooting environment. The ultimate goal is to establish a way to extract objects in real time for any background, and various types of sensors including stereo cameras have been tested with the aim of meeting this goal.

4.3.3 Image Stitching and Dynamic Shadow Processing

For this we have been researching real-time distortion correction and stitching technology for video captured by multiple 4K cameras. The goal is stitching multiple movies taken with multiple 4K cameras to generate a super high-definition movie with more than 8K resolution images in real time.

Another important post effect for background images is creating dynamic shadows that will move along with the images floating in the foreground. The shapes of the shadows are given by projectively transforming the foreground images on the basis of the lighting position of the display environment.

4.3.4 Sound Field Recording and Reproduction

The topic we have been investigating for this is sound field reproduction technologies that localize sound images on any image position. The use of virtual loudspeaker technology enables sound to be generated from a mouse or a subject on a 2D image plane. The goal is achieving 3D space composition using a wavefield reconstruction technique.

4.3.5 Multiple Media Transport with Advanced Synchronization Protocol

In this area we have been researching the synchronized media transmission technology "Advanced MPEG Media Transport (MMT)," an expansion of MMT technology, and applied H.265/HEVC and MPEG-4 ALS technologies to it. The goal is transmitting all spatial information such as target object size and position, positional relationship, and object data, as well as audio/video over IP networks.

4.4 Feasibility Experiments and Evaluations

A life size "Kirari!" system was constructed and demonstration experiments were conducted with it to cover several service cases such as live sports reporting, remote presentation, and remote communication with interaction. **Figure 6** is a photo taken at the demonstration using "Kirari!" in Japan [87]. The presenter in Fig. 6 was actually stood on the stage in the U.S. and he appeared on the stage in Japan as if he stands on right in front of audiences [88].

In this demonstration, a video transmission network is constructed by linking Japan and the USA, which consisted of GEMnet2, which is a network for research and development owned by NTT Laboratories [89], and Internet2 [90].

Fig. 6 Communication by "Kirari!" system.

A subjective experiment we conducted to evaluate reality and experience made it clear that for 2D virtual images it is better to see a large object from a long distance and this indicated that "Kirari!" is suitable for large-scale public screen service. We also found that the spatial audio technique the system features provides a sense of reality rather than the experience of enjoyment that had been enhanced by dimensional visual presentations [91]. Thus it can be considered that the sense of reality obtained from psychological factors was successfully raised by using the "Kirari!" system.

In the future, various type of displays such as glass-free 3D displays or displays using augmented reality glasses will be adopted to "Kirari!," instead of the multi-layered 2D display it now uses. More feasibility tests will also be done, including large scale online demonstrative experiments. These tests will feature not only remote communication instances but also presentations of speeches, interactive arts, and even stage entertainment. Note that the "Kirari!" system includes multiple technologies and leads the way toward new, high presence experiences in telecommunication service. Thus, in future feasibility tests in various fields, we would like to fine tune and select technologies in a very flexible manner so that they can be fitted to each service requirement. Accordingly, appropriate media technologies might be added to "Kirari!" if necessary, otherwise we may cut down current specs instead. Looking foward to further experiments, we may also conduct studies on artificial post effect processes to provide more impactful, emotional, and dimensional expressions. We believe that further research activities of this type will lead to the development of broader and more versatile telecommunication services.

5. Concluding Remarks

This paper has described research work done in the visual reality (VR) field that is closely related to telecommunication services.

In remote communication, not only the physical quality of audio and visual data but also the quality of communication between users is quite important. This is because both are means to enable precise information to be exchanged and clear mutual understanding to be obtained, including intentions and connotations. In remote communication, the perceptions and sense of engagement on the part of participants have an even greater effect on communication quality than the amount of provided information as rich audio or visual data. In this sense, high reality and high quality remote communication must be derived by the vivid feelings of real existence experienced by each participant. In this paper we described gaze awareness in communications as one of the most fundamental factors for understanding the activities of others and showed the results obtained in related research work. Immersion and feelings of space were also described as important factors for obtaining realistic perceptual feelings in VR communications, as well as immersive displays and perspective awareness technologies and audio technologies that ultimately lead to the sense of space sharing.

We also described specific-purpose-oriented communication methods. The scope of these methods is totally different from that of our remote communication service, which puts the highest priority on providing accurate information by heightening communication quality in natural conversation environments. In this service, information is distorted based on its purpose and finally target information is appropriately provided to users. The aim of this service is to provide information that is difficult to provide accurately, even via rich remote communication systems or actual face-to-face meetings. Motion in sports is introduced as an example of target data in this service, and related VR sports training systems were introduced. While there are still few remote communication systems based on this technology, the type of communication it provides is surely a very important part of telecom applications, because of the growing number of videos that transfer knowledge and learning online today.

Finally, the concept of the Immersive Telepresence Kirari! system was presented. This system represents the latest communication research that has been done with the aim of achieving high reality and new user experiences in this field. In the research work on the system that has been done, the focus was on achieving a sense of realism by using internal factors to maximize users' perceptual reality. Through the use of post production signal processes, real world physical data captured as audio visual signals was enhanced and some artificial effects were even added in the presentation process. We also introduced an example of the latest experiments that have been conducted in the entertainment field. We believe that these new developments will enable telecommunication applications to be widely extended, and it also can provide users new experiences as well as entertainment.

Applications in telecommunication service have been growing so that they are now expanding from simply providing rich communication to covering wide and diverse fields ranging from education to entertainment. Even though the use of head mounted displays (HMDs) has enabled a greater number of people to experience the world of visual reality (VR), in most cases it is necessary to provide immersive contents created for VR applications. The use of HMDs certainly brings new immersive, omnidirectional viewing experiences to users, but it can be considered that current HMD based VR applications are merely highlights that show only a quite narrow scope of VR. It is becoming increasingly clear that VR has much potential as a promising tool for enhancing human experiences in communication, one that may bring about richer and more substantial communication by transmitting multisensory information, including realistic sensations, that even high quality video conference systems have not been able to provide. This clearly indicates that the areas in the telecommunication industry where VR can be applied are growing and that the service applications that can be considered are broadening. In the future it will be necessary to stimulate users' sense of internal realism by using technology in specific-purpose-oriented communication on a widespread scale. Authors believe that VR must be positioned in the center of telecommunication applications and it will enhance communication experiences and finally enrich people's daily lives.

References

[1] Oculus Rift, 2016 Oculus VR, LLC (online), available from ⟨https://www3.oculus.com/en-us/rift/⟩ (accessed 2016-11-15).

[2] HTC Vive, HTC Corporation (online), available from ⟨https://www.vive.com/us/product/⟩ (accessed 2016-11-15).

[3] Playstation VR, Sony Interactive Entertainment LLC (online), available from ⟨https://www.playstation.com/⟩ (accessed 2016-11-15).

[4] Gear VR, Samsung (online), available from ⟨http://www.samsung.com/us/mobile/virtual-reality/⟩ (accessed 2016-11-15).

[5] Tachi, S.: *Telexistence*, World Scientific Publishing Co. Pte.Ltd., Singapore (2009).

[6] Biocca, F. and Levy, M.R.: *Communication in the Age of Virtual Reality*, Lawrence Erlbaum Associates, Inc., New Jersey (1995).

[7] Riva, G.: Virtual Reality As Communication Tool: A Sociocognitive Analysis, *Presence: Teleoper. Virtual Environ.*, Vol.8, No.4, pp.462–468 (online), DOI: 10.1162/105474699566341 (1999).

[8] Imoto, M., Uchida, S., Ashikaga, E., Wagatsuma, M. and Hidaka, K.: A Concept of Immersive Telepresence 'Kirari!', *Proc. 23rd International Display Workshops*, pp.1287–1290 (2015).

[9] Dourish, P. and Bellotti, V.: Awareness and Coordination in Shared Workspaces, *Proc. 1992 ACM Conference on Computer-supported Cooperative Work*, *CSCW '92*, pp.107–114, ACM (online), DOI: 10.1145/143457.143468 (1992).

[10] Dourish, P. and Bly, S.: Portholes: Supporting Awareness in a Distributed Work Group, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, *CHI '92*, pp.541–547, ACM (online), DOI: 10.1145/142750.142982 (1992).

[11] Tang, J.C. and Minneman, S.L.: VideoDraw: A Video Interface for Collaborative Drawing, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, *CHI '90*, pp.313–320, ACM (online), DOI: 10.1145/97243.97302 (1990).

[12] Ishii, H.: TeamWorkStation: Towards a Seamless Shared Workspace, *Proc. 1990 ACM Conference on Computer-supported Cooperative Work*, *CSCW '90*, pp.13–26, ACM (online), DOI: 10.1145/99332.99337 (1990).

[13] Vertegaal, R., Slagter, R., van der Veer, G. and Nijholt, A.: Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, *CHI '01*, pp.301–308, ACM (online), DOI: 10.1145/365024.365119 (2001).

[14] Ishii, H., Kobayashi, M. and Grudin, J.: Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments, *ACM Trans. Inf. Syst.*, Vol.11, No.4, pp.349–375 (online), DOI: 10.1145/159764.159762 (1993).

[15] Kuzuoka, H., Yamashita, J., Yamazaki, K. and Yamazaki, A.: Agora: A Remote Collaboration System That Enables Mutual Monitoring, *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '99*, pp.190–191, ACM (online), DOI: 10.1145/632716.632836 (1999).

[16] Okada, K., Maeda, F., Ichikawaa, Y. and Matsushita, Y.: Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design, *Proc. 1994 ACM Conference on Computer Supported Cooperative Work*, *CSCW '94*, pp.385–393, ACM (online), DOI: 10.1145/192844.193054 (1994).

[17] Fish, R.S., Kraut, R.E. and Chalfonte, B.L.: The VideoWindow System in Informal Communication, *Proc. 1990 ACM Conference on Computer-supported Cooperative Work*, *CSCW '90*, pp.1–11, ACM (online), DOI: 10.1145/99332.99335 (1990).

[18] Yankelovich, N., Simpson, N., Kaplan, J. and Provino, J.: Porta-person: Telepresence for the Connected Conference Room, *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '07*, pp.2789–2794, ACM (online), DOI: 10.1145/1240866.1241080 (2007).

[19] Otsuka, K.: MMSpace: Kinetically-augmented telepresence for small group-to-group conversations, *2016 IEEE Virtual Reality* (*VR*), pp.19–28 (online), DOI: 10.1109/VR.2016.7504684 (2016).

[20] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V. and Hart, J.C.: The CAVE: Audio Visual Experience Automatic Virtual Environment, *Comm. ACM*, Vol.35, No.6, pp.64–72 (online), DOI: 10.1145/129888.129892 (1992).

[21] Cruz-Neira, C., Sandin, D.J. and DeFanti, T.A.: Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE, *Proc. 20th Annual Conference on Computer Graphics and Interactive Techniques*, *SIGGRAPH '93*, pp.135–142, ACM (online), DOI: 10.1145/166117.166134 (1993).

[22] Muhanna, M.A.: Virtual Reality and the CAVE, *J. King Saud Univ. Comput. Inf. Sci.*, Vol.27, No.3, pp.344–361 (online), DOI: 10.1016/j.jksuci.2014.03.023 (2015).

[23] Kida, K., Ihara, M., Shiwa, S. and Ishibashi, S.: Motion tracking method for the CAVE^{TM} system, *Proc. 5th International Conference on Signal Processing*, *WCCC-ICSP 2000*, Vol.2, pp.859–862 (online), DOI: 10.1109/ICOSP.2000.891647 (2000).

[24] Iso, K., Yagi, T., Kobayashi, M., Iwaki, S. and Ishibashi, S.: On Natural Living Room Communication with "ComAdapter": Adapting to the Differences in Room Structure, *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, *CHI EA '03*, pp.716–717, ACM (online), DOI: 10.1145/765891.765947 (2003).

[25] Ogi, T., Yamada, T., Tamagawa, K., Kano, M. and Hirose, M.: Immersive telecommunication using stereo video avatar, *Proc. IEEE Virtual Reality*, pp.45–51 (online), DOI: 10.1109/VR.2001.913769 (2001).

[26] Altman, M., Krauss, K., Susal, J. and Tsingos, N.: Immersive Audio for VR, *Proc. 2016 AES Int. Conf. Audio for Virtual and Augmented Reality*, CA, USA (2016).

[27] Nishiguchi, T., Nakayama, Y., Okumura, R., Sugimoto, T., Imai, A., Iwaki, M., Hamasaki, K., Ando, A., Kitajima, S., Otsuka, Y. and Shimaoka, S.: Production and Live Transmission of 22.2 Multichannel Sound with Ultrahigh-definition TV, *Proc. 122nd AES Int. Conv.*, Vienna, Austria (2007).

[28] Ono, K., Nishiguchi, T., Matsui, K. and Hamasaki, K.: Portable spherical microphone for Super Hi-Vision 22.2 multichannel audio, *Proc. 135th AES Int. Conv.*, NY, USA (2013).

[29] Poletti, M.A.: Three-dimensional surround sound systems based on spherical harmonics, *J. Audio Eng. Soc.*, Vol.53, No.11, pp.1004–1025 (2005).

[30] Fazi, F., Nelson, P., Christensen, J.E. and Seo, J.: Surround System Based on Three-Dimensional Sound Field Reconstruction, *Proc. 125th AES Int. Conv.*, Paris, France (2008).

[31] Poletti, M.A., Betlehem, T. and Abhayapala, T.D.: Higher-order loudspeakers and active compensation for improved 2D sound field reproduction in rooms, *J. Audio Eng. Soc.*, Vol.63, No.1/2, pp.31–45 (2015).

[32] Wakayama, K. and Takada, H.: Spatial multi-zone sound field reproduction Using higher-order loudspeakers in reverberant rooms, *Proc. 140th AES Int. Conv.*, Paris, France (2016).

[33] Sakamoto, S., Hongo, S., Okamoto, T., Iwaya, Y. and Suzuki, Y.: Sound-space recording and binaural presentation system based on a 252-channel microphone array, *Acoustical Science and Technology*, Vol.36, No.6, pp.516–526 (2015).

[34] Bates, E., Gorzel, M., Ferguson, L., Dwyer, H. and Boland, F.M.: Comparing Ambisonic Microphones: Part 1, *Proc. 2016 AES Int. Conf. Sound Field Control*, Guildford, UK (2016).

[35] Fan, C., Salehin, S.M.A. and Abhayapala, T.D.: Practical Implementation and Analysis of Spatial Soundfield Capture by Higher Order Microphones, *Proc. Signal and Information Processing Association Annual Summit and Conference* (*APSIPA*) (2014).

[36] Suzuki, Y., Okamoto, T., Trevino, J., Cui, Z.L., Iwaya, Y., Sakamoto, S. and Otani, M.: 3D Spatial Sound Systems Compatible with Human's Active Listening to Realize Rich High-Level kansei Information, *Interdisciplinary Information Sciences*, Vol.18, No.2, pp.71–82 (2012).

[37] Noisternig, M., Katz, B.F.G., Siltanen, S. and Savioja, L.: Framework for Real-Time Auralization in Architectural Acoustics, *Acta Acustica United with Acustica*, Vol.94, No.6, pp.1000–1015 (2008).

[38] Yan, S., Sun, H., Svensson, U.P., Ma, X. and Hovem, J.M.: Optimal Modal Beamforming for Spherical Microphone Arrays, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.19, No.2, pp.361–371 (2011).

[39] Niwa, K., Hioka, Y., Furuya, K. and Haneda, Y.: Diffused sensing for sharp directive beamforming, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.21, No.11, pp.2346–2355 (2013).

[40] Hirose, M., Yokoyama, K. and Sato, S.I.: Transmission of realistic sensation: Development of a virtual dome, *Proc. IEEE Virtual Reality Annual International Symposium*, pp.125–131 (online), DOI: 10.1109/VRAIS.1993.380788 (1993).

[41] Hirose, M.: Image-Based Virtual World Generation, *IEEE MultiMedia*, Vol.4, No.1, pp.27–33 (online), DOI: 10.1109/93.580393 (1997).

[42] Sharma, P.: Challenges with Virtual Reality on Mobile Devices, *ACM SIGGRAPH 2015 Talks*, *SIGGRAPH '15*, p.57:1, ACM (online), DOI: 10.1145/2775280.2792597 (2015).

[43] Shivappa, S., Morrell, M., Sen, D., Peters, N. and Salehin, S.M.: Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics, *Proc. 2016 AES Int. Conf. Audio for Virtual and Augmented Reality*, CA, USA (2016).

[44] Zhang, W., Zhang, M., Kennedy, R.A. and Abhayapala, T.D.: On High-Resolution Head-Related Transfer Function Measurements, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.20, No.2, pp.575–583 (2012).

[45] Pike, C., Taylor, R., Parnell, T. and Melchior, F.: Object-Based 3D Audio Production for Virtual Reality Using the Audio Definition Model, *Proc. 2016 AES Int. Conf. Audio for Virtual and Augmented Reality*, CA, USA (2016).

[46] Kunita, Y., Ogawa, N., Sakuma, A., Inami, M., Maeda, T. and Tachi, S.: Immersive autostereoscopic display for mutual telexistence: TWISTER I (Telexistence wide-angle immersive STEReoscope model I), *Proc. IEEE Virtual Reality*, pp.31–36 (online), DOI: 10.1109/VR.2001.913767 (2001).

[47] Hirata, K., Harada, Y., Takada, T., Aoyagi, S., Shirai, Y., Yamashita, N., Kaji, K., Yamato, J. and Nakazawa, K.: t-Room: Next Generation Video Communication System, *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pp.1–4 (online), DOI: 10.1109/GLOCOM.2008.ECP.1058 (2008).

[48] Heath, C. and Luff, P.: Disembodied Conduct: Communication Through Video in a Multi-media Office Environment, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, *CHI '91*, pp.99–103, ACM (online), DOI: 10.1145/108844.108859 (1991).

[49] Jones, A., McDowall, I., Yamada, H., Bolas, M. and Debevec, P.: Rendering for an Interactive 360° Light Field Display, *ACM SIGGRAPH 2007 Papers*, *SIGGRAPH '07*, ACM (online), DOI: 10.1145/1275808.1276427 (2007).

[50] Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M. and Debevec, P.: Achieving Eye Contact in a One-to-many 3D Video Teleconferencing System, *ACM SIGGRAPH 2009 Papers*, *SIGGRAPH '09*, pp.64:1–64:8, ACM (online), DOI: 10.1145/1576246.1531370 (2009).

[51] Suyama, S., Ohtsuka, S., Takada, H., Uehira, K. and Sakai, S.: Apparent 3-D image perceived from luminance-modulated two 2-D images displayed at different depths, *Vision Research*, Vol.44, No.8, pp.785–793 (online), DOI: 10.1016/j.visres.2003.10.023 (2004).

[52] Takada, H., Suyama, S., Hiruma, K. and Nakazawa, K.: A Compact Depth-Fused 3-D LCD, *SID '03 Digest of Technical Papers*, Vol.58, No.2, pp.1526–1529 (2003).

[53] Iso, K., Ozawa, S., Date, M., Takada, H., Andoh, Y. and Matsuura, N.: Video Conference 3D Display That Fuses Images to Replicate Gaze Direction, *Journal of Display Technology*, Vol.8, No.9, pp.511–520 (online), DOI: 10.1109/JDT.2012.2193662 (2012).

[54] Matsui, K. and Ando, A.: Binaural Reproduction of 22.2 Multichannel Sound Over Loudspeakers, *Proc. 129th AES Int. Conv.*, CA, USA (2013).

[55] Kurabayashi, H., Otani, M., Hashimoto, M. and Kayama, M.: Sound Image Localization Using Dynamic Transaural Reproduction with Non-contact Head Tracking, *IEICE Trans. Fundamental Electron. Commun. Comput. Sci.*, Vol.97, No.9, pp.1849–1858 (2014).

[56] Ishii, R., Ozawa, S., Kawamura, H. and Kojima, A.: MoPaCo: High telepresence video communication system using motion parallax with monocular camera, *2011 IEEE International Conference on Computer Vision Workshops* (*ICCV Workshops*), pp.463–464 (online), DOI: 10.1109/ICCVW.2011.6130278 (2011).

[57] Date, M., Takada, H., Honda, Y., Ozawa, S., Mieda, S. and Kojima, A.: Highly Realistic 3D Display System for Space Composition Telecommunication, *Journal of Display Technology*, Vol.11, No.2, pp.121–128 (online), DOI: 10.1109/JDT.2014.2338858 (2015).

[58] Spors, S., Rabenstein, R. and Ahrens, J.: The Theory of Wave Field Synthesis Revisited, *Proc. 124th Conv. AES*, Amsterdam, The Netherlands (2008).

[59] Ranjan, R. and Gan, W.S.: Wave field synthesis: The future of spatial audio, *IEEE Potentials*, Vol.32, No.2, pp.17–23 (2013).

[60] Blauert, J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press (1996).

[61] Ise, S.: A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems, *Acta Acustica United with Acustica*, Vol.85, No.1, pp.78–87 (1999).

[62] Omoto, A., Ise, S., Ikeda, Y., Ueno, K., Enomoto, S. and Kobayashi, M.: Sound field reproduction and sharing system based on the boundary surface control principle, *Acoustical Science and Technology*, Vol.36, No.1, pp.1–11 (2015).

[63] Koyama, S., Furuya, K., Hiwasaki, Y. and Haneda, Y.: Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.21, No.4, pp.685–696 (2013).

[64] Koyama, S., Furuya, K., Uematsu, H., Hiwasaki, Y. and Haneda, Y.:

Real-Time Sound Field Transmission System by Using Wave Field Reconstruction Filter and Its Evaluation, *IEICE Trans. Fundamental Electron. Commun. Comput. Sci.*, Vol.97, No.9, pp.1840–1848 (2014).
[65] Koyama, S., Furuya, K., Hiwasaki, Y. and Haneda, Y.: Reproducing Virtual Sound Sources in Front of a Loudspeaker Array Using Inverse Wave Propagator, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.20, No.6, pp.1746–1758 (2012).
[66] Buchner, H., Spors, S. and Kellermann, W.: Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field synthesis, *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.* (*ICASSP*), pp.iv-117–iv-120 (2004).
[67] Emura, S. and Kurihara, S.: Echo Canceler for Real-Time Audio Communication with Wave Field Reconstruction, *Proc. 139th AES Int. Conv.*, NY, USA (2015).
[68] Koyama, S., Furuya, K., Wakayama, K., Shimauchi, S. and Saruwatari, H.: Analytical approach to transforming filter design for sound field recording and reproduction using circular arrays with a spherical baffle, *J. Acoust. Soc. Am.*, Vol.139, No.3, pp.1024–1036 (2016).
[69] Anderson, F., Grossman, T., Matejka, J. and Fitzmaurice, G.: YouMove: Enhancing movement training with an augmented reality mirror, *Proc. 26th Annual ACM Symposium on User Interface Software and Technology*, pp.311–320, ACM (2013).
[70] Charbonneau, E., Miller, A. and LaViola Jr, J.J.: Teach me to dance: Exploring player experience and performance in full body dance games, *Proc. 8th International Conference on Advances in Computer Entertainment Technology*, p.43, ACM (2011).
[71] Eaves, D.L., Breslin, G. and Van Schaik, P.: The short-term effects of real-time virtual reality feedback on motor learning in dance, *Presence*, Vol.20, No.1, pp.62–77 (2011).
[72] Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F. and Craig, C.: Using virtual reality to analyze sports performance, *IEEE Computer Graphics and Applications*, Vol.30, No.2, pp.14–21 (2010).
[73] Holden, M.K. and Todorov, E.: Use of virtual environments in motor learning and rehabilitation, *Department of Brain and Cognitive Sciences, Handbook of Virtual Environments: Design, Implementation, and Applications*, pp.999–1026 (2002).
[74] Sveistrup, H.: Motor rehabilitation using virtual reality, *Journal of Neuroengineering and Rehabilitation*, Vol.1, No.1, p.1 (2004).
[75] Ochi, D., Kameda, A., Takahashi, K., Makiguchi, M. and Takeuchi, K.: VR technologies for rich sports experience, *ACM SIGGRAPH 2016 Emerging Technologies*, p.21, ACM (2016).
[76] Bideau, B., Kulpa, R., Ménardais, S., Fradet, L., Multon, F., Delamarche, P. and Arnaldi, B.: Real handball goalkeeper vs. virtual handball thrower, *Presence*, Vol.12, No.4, pp.411–421 (2003).
[77] Brunnett, G., Rusdorf, S. and Lorenz, M.: V-Pong: An immersive table tennis simulation, *IEEE Computer Graphics and Applications*, Vol.26, No.4, pp.10–13 (2006).
[78] Miles, H., Musembi, N., Pop, S.R. and John, N.: Virtual environment for rugby skills training, *Proc. Joint VR Conference of EuroVR and EGVE* (*JVRC '11*), Nottingham, UK. Vuorimiehentie, Finland: VTT (2011).
[79] Ida, H., Fukuhara, K. and Ishii, M.: Recognition of tennis serve performed by a digital player: Comparison among polygon, shadow, and stick-figure models, *PloS one*, Vol.7, No.3, p.e33879 (2012).
[80] Chua, P.T., Crivella, R., Daly, B., Hu, N., Schaaf, R., Ventura, D., Camill, T., Hodgins, J. and Pausch, R.: Training for physical tasks in virtual environments: Tai Chi, *Proc. IEEE Virtual Reality*, pp.87–94, IEEE (2003).
[81] Ikegami, T., Hirashima, M., Osu, R. and Nozaki, D.: Intermittent visual feedback can boost motor learning of rhythmic movements: Evidence for error feedback beyond cycles, *The Journal of Neuroscience*, Vol.32, No.2, pp.653–657 (2012).
[82] Ando, H., Akiko, C., Norberto, E.N., Nishino, Y., Juan, L., Wada, A. and Sakano, Y.: Perceptual and cognitive mechanisms of presence and its evaluation technology, *NICT Journal*, Vol.57, No.1-2, pp.159–168 (2010).
[83] Marini, D., Folgieri, R., Gadia, D. and Rizzi, A.: Virtual Reality As a Communication Process, *Virtual Real.*, Vol.16, No.3, pp.233–241 (online), DOI: 10.1007/s10055-011-0200-3 (2012).
[84] Mantovani, C.: Virtual reality as a communication environment: Consensual hallucination, fiction and possible selves, *Human Relations*, Vol.48, No.6, pp.669–683 (1995).
[85] Utsugi, K., Naemura, T., Koike, T. and Oikawa, M.: E-IMPACT: Exaggerated Illustrations Using Multi-perspective Animation Control Tree Structure, *Proc. 8th International Conference on Advances in Computer Entertainment Technology*, *ACE '11*, pp.63:1–63:8, ACM (online), DOI: 10.1145/2071423.2071502 (2011).
[86] Eyeliner, studio TED (online), available from ⟨http://studioted.jp/⟩ (accessed 2016-11-01).
[87] Ultra-High-Presence Video Live Viewing of KABUKI LION SHI-SHI-O Las Vegas Performance, NTT (online), available from ⟨http://www.ntt.co.jp/topics_e/vegas_kirari/index.html⟩ (accessed 2016-05-07).
[88] SHOCHIKU channel on YouTube, SHOCHIKU Co., Ltd. (online), available from ⟨https://www.youtube.com/watch?v=E0D67sBsCY0⟩ (accessed 2016-05-07).
[89] Uose, H.: GEMnet2: NTT's new network testbed for global R&D, IEEE, (online), DOI: 10.1109/TRIDNT.2005.19 (2005).
[90] Nakagawa, J.: Collaborative Trials of Internet2 and NTT (2016), 2016 Internet2 Technology Exchange Conference, available from ⟨http://meetings.internet2.edu/2016-technology-exchange/speakers/6153/⟩.
[91] Akutsu, A., Ono, A., Takada, H., Tonomura, Y. and Imoto, M.: 2020 Public Viewing - Kirari! Immersive Telepresence Technology, *NTT Technical Review*, Vol.14, No.12, pp.1–6 (2016).

Hidenobu Nagata received his M.E. degree in Systems and Information Engineering from Faculty of Engineering, Hokkaido University, Hokkaido, Japan. He joined NTT from 2001, and studied video handling technologies including video indexing, automatic summarization, and its interface. From 2008 to 2014, he worked at NTT Electronics Corporation, and developed transcoders and embedded audio IP. He moved to NTT in 2014, and currently he has been researching ultra-realistic communication technology including Immersive Telepresence System.

Dan Mikami received his B.E. and M.E. degrees from Keio University, Kanagawa, Japan in 2000 and 2002, respectively. He has been working for Nippon Telegraph and Telephone Corporation from 2002. He received his Ph.D. from Tsukuba University in 2012. His current research activities are mainly focused on computer vision and virtual reality. He was awarded the Meeting on Image Recognition and Understanding 2009 Excellent Paper Award 2009, the IEICE Best Paper Award 2010, the IEICE KIYASU Zen'iti Award 2010, the IPSJ SIGCDS Excellent Paper Award 2013, and the IEICE Human Communication Award 2015.

Hiromu Miyashita received his B.E. and M.E. degrees from Keio University, Kanagawa, Japan, in 2008 and 2010, respectively. In 2010, he joined NTT Service Evolution Laboratories, Kanagawa, Japan, where he worked on human computer interaction, image and video processing, and ultra-high presence telecommunication service.

Keigo Wakayama received his B.E. and M.E. degrees from Waseda University, Tokyo, Japan, in 2010 and 2012, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT) in 2012. He is now a Researcher at NTT Service Evolution Laboratories, Kanagawa, Japan. His research interests include acoustic signal processing, sound field recording and reproduction, and machine learning.

Hideaki Takada received his M.E. degree in Information Systems from the University of Electro-Communications, Tokyo, Japan, in 1997. He received his D.S. degree in Global Information and Telecommunication Studies from Waseda University, Tokyo, Japan, in 2007. He joined NTT Integrated Information & Energy Systems Laboratories, NTT Corporation, Tokyo, Japan, in 1997. He is currently a Senior Research Engineer, Supervisor at NTT Service Evolution Laboratories and NTT Media Intelligence Laboratories. He has been engaged in the research of 3D visual perception and 3D display technology. He has received the Achievement Award from IEICE in 2003, Commendation by the Minister of Education, Culture, Sports, Science and Technology - Japan in 2006, and Technical Committee Prize Paper Award from IEEE IAS in 2014 and 2016.