Past and Future of Video Communication for Working from Anywhere

Shohei KAMAMURA

Abstract

As a measure to respond to recent social conditions, video communication, such as web conferences, is attracting more attention and spreading rapidly. In this paper, I would like to look back on the history of video communication that evolved and supported by the widening bandwidth of networks and the development of video coding technologies in the first place. Then, I will take the use of a device called a multipoint control unit (MCU) as an example and explain the basic mechanism with which video and audio are shared and communication is realized among participants. Last, I will briefly introduce the future outlook of video communication.

Keywords: Video communication, web conference, videoconference, work-style reform

1. Introduction

Video communication is a generic name for a means to realize web conferences for people teleworking from home and remote classes for students. Although it has been used for a long time, its use began to spread rapidly as measures to avoid the infection of world-shaking COVID-19 that started at the beginning of 2020.

I teleworked from home almost every day from February to May 2020. I was freed from round-trip commutes of more than two hours on packed trains and, as a result, was able to use my time effectively. I feel that I am more productive now.

Telework is defined as a flexible work style without the restriction of time and place. Although there are many tools to realize telework, video communication is indispensable for white-collar workers who need to attend meetings very often during work. Originally, communication meant conveying perceptions, emotions, and thoughts between people who lived a social life. Visual communication is defined as having two-way communication with the aid of audio and video using telecommunication technologies. By conveying nonverbal information, including facial expressions and body language, unlike communication using only voice, such as the one via telephone, participants can feel a sense of reality as if they were meeting face-to-face even though they were apart from each other.

Shohei KAMAMURA (Regular member: Senior Member)
NTT Network Service Systems Laboratories, Nippon Telegraph and Telephone Corporation,
Musashino-shi, Tokyo 180-8585 Japan.
E-mail: shohei.kamamura@ieee.org
The Journal of IEICE Vol.103, No. 12, pp. 1261-1265, December 2020

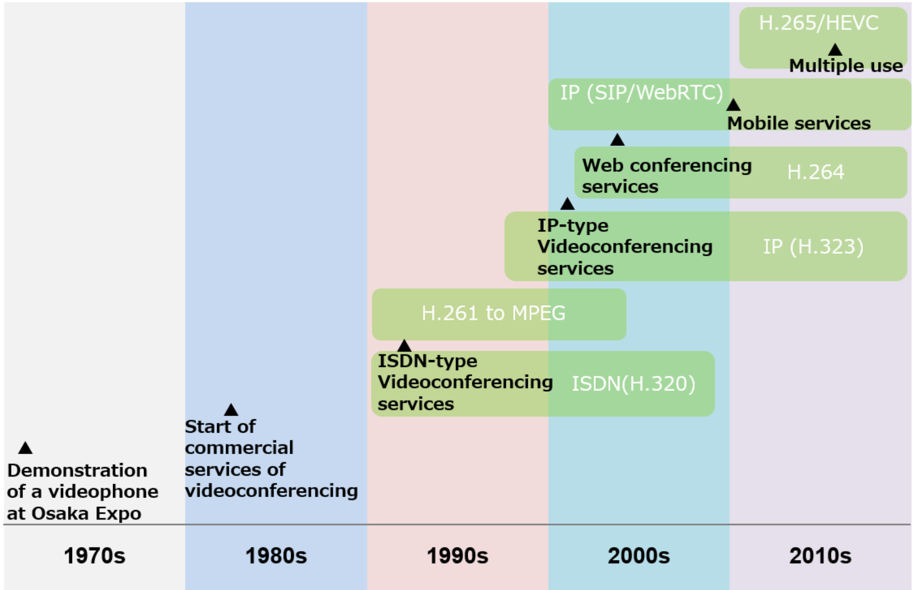

Figure 1: History of Video Communication

The history of video communication started with the demonstration of a videophone in 1970 and was developed into substantial video and web conferencing, supported by the development of network technologies and video coding technologies. Video communication has been used on various devices and in various use cases so that it meets the needs of the times.

In this paper, I would like to explain the history of video communication and the basic mechanism in which video and audio are shared among participants. I hope it will help readers, who are expected to play their roles in society, fully utilize video communication by deepening their understanding of it.

2. History of Video Communication

The history of video communication is shown in Figure 1. Although a videophone was demonstrated and made its debut at the Osaka Expo in 1970, substantial commercial service was started by the then Nippon Telegraph and Telephone Public Corporation in the 1980s. In the 1990s, the digital communication network named Integrated Services Digital Network (ISDN), as well as videoconferencing using a protocol for ISDN called H.320, came into widespread use. Furthermore, in the late 1990s, the H.323[(terminology)1] protocol for internet protocol (IP) networks was recommended, and when we entered the 2000s, higher quality videoconferencing came to be widely used. For example, the image quality of videoconferencing used to be the standard definition (SD) of 720 x 480 pixels, which was the same quality as DVDs, but was upgraded to the high definition (HD) of 1,280 x 720 pixels.

In the middle of the 2000s, in addition to videoconferencing using dedicated terminals, web conferencing services using PCs via the internet started. Later, not only web conferencing services using dedicated applications but also ones that run on the web browser named WebRTC[2,3] came onto the market, which will be explained later. Smartphones started to diffuse explosively in the late 2000s, and as

Terminology

H.323: It is recommended at ITU-T for real-time audio and video communications over IP networks. The H-series of ITU-T are protocols related to audiovisual and multimedia systems.

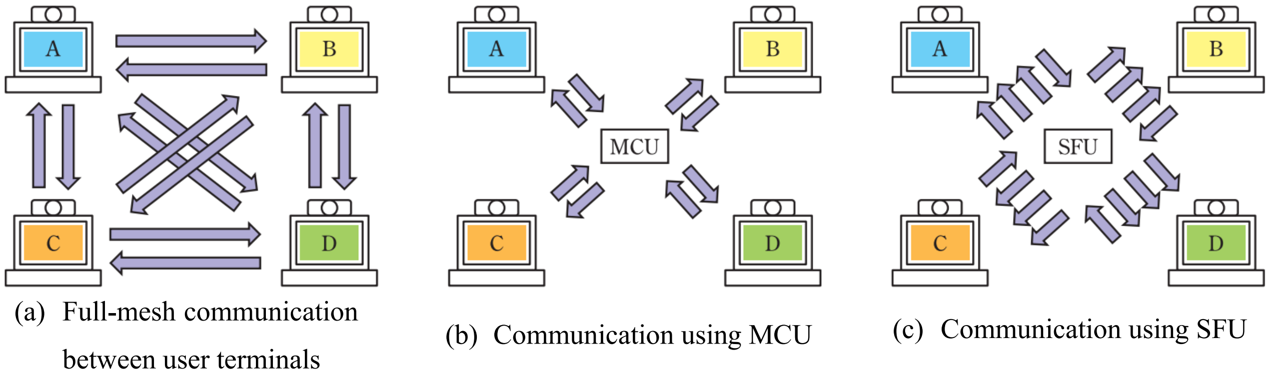

Figure 2: Mechanism of Multipoint Control In method (a), images and sounds are communicated between all user terminals over a full-mesh network. In method (b), a multipoint control unit (MCU) composes images sent from user terminals that then sends the composite images back to the user terminals. In method (c), a unit called the SFU distributes images and sounds sent from user terminals to other user terminals in place of a real sender.

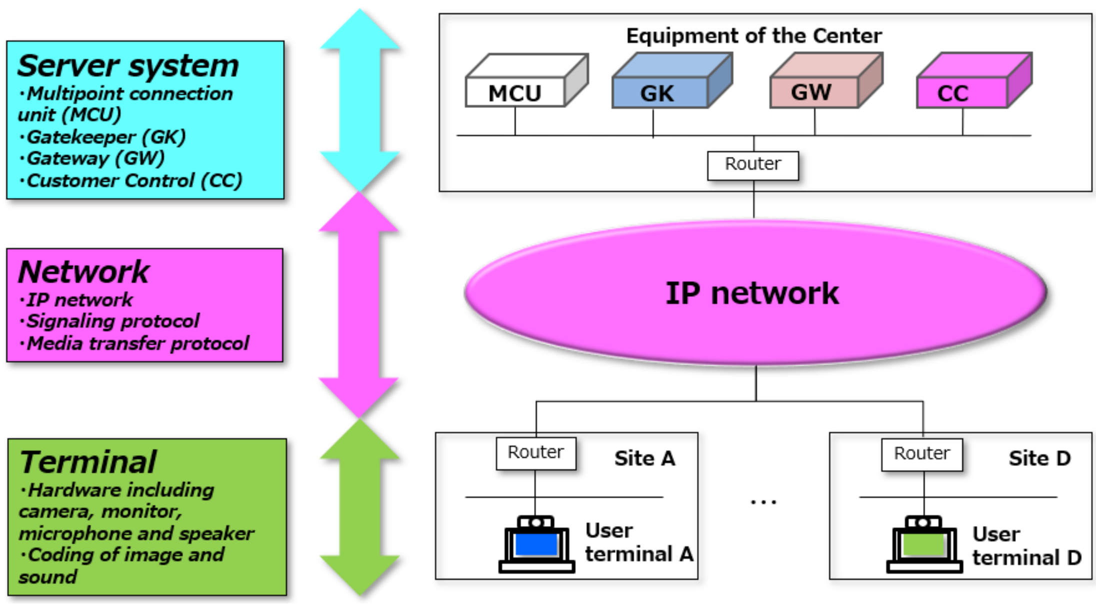

Figure 3: Consisting Technologies When a Multipoint Control Unit is Used
Video communication using a multipoint control unit (MCU) consists of technologies roughly classified into server systems, networks, and terminals.

a result, mobile services like LINE also started to become hot in the 2010s. Bandwidth of IP networks has been further widening, and full HD video communication of 1,920 x 1,080 pixels, which is as high quality as Blu-ray, is possible today. Since the mid-2010s, improvement of video quality and diversification of devices, as mentioned above, have changed the idea that video communication is solely for meetings. Expectations for use for other purposes than meetings are gradually increasing. Examples include online conclusion of a real estate contract, which is called an explanation of important information using IT, online medical treatment in medically underserved areas, and use for education, such as online English conversation and other classes.

The development of video communication has been supported not only by the increase in network capacity represented by IP but also by the progress of video and audio coding technologies. Coding technologies mean, simply put, technologies to reduce the data size while maintaining viewing and listening quality as high as possible. If we focus on the area related to video communication, H.261, developed in 1990, was used for ISDN videoconferencing services. Next, H.264/AVC[4] established in 2003, was used for IP video communication services and is still a mainstream video coding technology. H.265/HEVC[5] was proposed in 2013 and is expected to be the next-generation standard that realizes video communication of higher definition.

In this way, video communication has been used on various devices and in various use cases, supported by the development of network technologies and video coding technologies so that it meets the needs of the times. In section 3, I would like to explain the basic mechanism of video communication.

3. Technologies Supporting Video Communication

3.1 Mechanism of Multipoint Control

Video communication requires a mechanism where participants connect to each other at the same time to communicate instead of one-to-one communication like a telephone call. The simplest connection method may be a full-mesh network where each participant is directly connected to every other participant (Figure 2 [a]). However, as this connection method needs to accommodate video traffic between all users, the load on user terminals becomes heavy, and the use of network bandwidth increases. To solve this problem, two other methods are proposed (Figure 2 [b] and [c]), where servers are placed in between and communicate in place of a real sender. Figure 2 (b) shows the mechanism where participants communicate via an MCU. The MCU composes images and sounds sent from user terminals and then sends the composite images back to the user terminals. For this reason, high processing capability is required for the server where the MCU is installed, but, as a result, the load on user terminals is reduced, and the use of network bandwidth can also be reduced. Figure 2 (c) shows the mechanism where a Selective Forwarding Unit (SFU) distributes images, which have been sent from user terminals, to other user terminals in place of a real sender. User terminals receive images from all terminals but send them only to the SFU. This mechanism makes it possible to reduce the processing load on the server compared with the case that uses MCU and to reduce the load on user terminals and the use of network bandwidth compared with the case that uses a full-mesh network.

3.2 Consisting Technologies When a Multipoint Control Unit (MCU) is Used

Next, I will take the case in which a representative method, the multipoint control unit (MCU), is used as an example and explain the technologies that constitute video communication. This method is composed of a wide variety of element technologies, which are roughly classified into server systems, networks, and terminals as shown in Figure 3.

A server system is composed of an MCU, a gatekeeper, a gateway and a customer control as shown at the top of Figure 3. As explained above, an MCU is a key technology of video communication, which composes images and sounds sent from user terminals and then sends them back to the user terminals. A gatekeeper, as its name clearly suggests, provides a reception control function. For example, when only user terminals A and D in Figure 3 are registered on the gatekeeper, other user terminals cannot access the MCU and therefore cannot participate in the meeting. Thus, the security of communication is ensured. In addition to this control function, a gatekeeper provides, in general, an address resolution function that connects video-communication-specific IDs and IP addresses as well as a bandwidth control function that adjusts the bandwidth for connection of user terminals. Next, I would like to explain a gateway. It is a technology that realizes interconnections between various user terminals and applications by providing a protocol conversion function. A protocol means a common procedure for establishing communication. Therefore, in a normal situation, applications having different protocols cannot communicate with each other. To solve this problem, a gateway comes in between and makes the communication appear as one having a common protocol, enabling interconnection between different terminals and applications. Last, I would like to explain customer control. It is a function to provide control of a meeting for customers or users. For example, it provides control functions, such as opening and closing a meeting, controlling screen layout and muting, and automatically opening a meeting with a reservation function, on its user-friendly web-based interface.

Next, let me move on to the explanation of a network. User terminals and a server system are connected by an IP network, in principle. I would like to omit the explanation of an IP network here. Especially in video communication, signaling protocols and media transfer protocols, which become active on the upper IP layers, play an important role. The role of a signaling protocol is to establish a connection, called a session, between terminals on which communication is conducted. Representative protocols include H.323, which is used for videoconferencing, and the session initiation protocol (SIP),[6] which is for IP telephone calls and web meetings. After the connection is established between terminals by the signaling protocol, sounds and images are transferred in real time using a media transfer protocol. For media transfer, international standards, such as the real-time transfer protocol (RTP) and real-time transfer control protocol (RTCP)[7] are used in most cases. RTP is a protocol that delivers data streams, such as sounds and images, in real time and has a characteristic of simple processing and rareness in delay. For this reason, it is possible to make supplementary use of RTCP, if you want to confirm the arrival of data or control the flow when delivering data.

Next, I would like to explain the technologies used for user terminals. Users may use not only the terminals dedicated to videoconferencing but also general-purpose computers that include PCs, smartphones, and tablets. It is necessary, of course, for such computers to be equipped with hardware for video communication, such as cameras, monitors, microphones, and speakers. Internally, however, video coding technologies, explained in section 2, assume a significant role. Coding is a technology for compressing or decompressing sounds and images. The process to compress the volume before transferring media to a network is called encoding, and reversely, the process to decompress the compressed media information to the state which people can listen to and view is called decoding. Especially for images, technologies to compress data efficiently while maintaining good quality have become an important subject of research. For example, the technologies of intra-frame prediction, which pays attention to the similarity of adjacent pixels on the same plane when sending images, and inter-frame prediction, which predicts temporal motion vector, enable data compression through sending only the differences instead of sending the entire data.

Last, I will introduce the closely related technology called WebRTC[2,3] by reviewing the technologies I have introduced thus far.

WebRTC is a technology that realizes video communication on a web browser, and its standardization is under way by W3C and IETF. It may help your understanding if you think this way: WebRTC is not a single independent communication technology but various technologies combined to realize WebRTC. For this reason, the method to implement WebRTC is highly flexible. For example, while standard implementation of a WebRTC application is full-mesh communication as shown in Figure 2 (a), implementation using MCU in Figure 2 (b) and SFU in Figure 2 (c) is also possible. If you use the above-mentioned gateway, interconnection with videoconferences and web meetings with different protocols are also possible. As a signaling protocol in WebRTC, the technology called WebSocket is often used but SIP, as explained above, can also be used without any problem. For media transfer, the secure real-time transport protocol (SRTP),[8] which is a more secure version of RTP, is used. In WebRTC, a method to transfer text and other data is also defined, in addition to the rules on media transfer. For example, video communication and chat can be combined using this method. Although the choice of video coding engine varies depending on the implementation of a web browser, H.264/AVC, VP8, and VP9 are used in the main. It can be specified when establishing a session, if the browser supports it. In this way, you will notice that WebRTC, for example, is composed of the elementary technologies shown in Figure 3 in principle. For many of the web conferencing services you usually use on your browser, in fact, WebRTC is used. [9]

4. Conclusion

In this paper, I have explained the history and basic technologies of video communication. The use of video communication is rapidly expanding because of the improvement of the sense of reality and immersion realized by higher definition images, which has been achieved by advances in network and video coding technologies, as well as the recent social situations. On the other hand, many of you may think that it is easier to convey your feelings face-to-face than online. Last, what I would like you to know is that video communication is not a simple substitute for face-to-face communication. Although I could not explain in detail here because of the restrictions of space, by linking with such technologies as speech recognition and machine translation for example, it becomes possible to acquire the content of speech at a meeting and to take minutes automatically, or to translate speech in English into Japanese, which are beginning to be offered as services. In this way, by accumulating and utilizing the audio and video data as big data, I believe video communication will continue to evolve to deliver new value that has not been realized in real spaces.

Last, it is my hope that readers of this paper will be able to realize a flexible work style by utilizing video communication more than ever and, as a result, improve their work-life balance and quality of life (QoL).

Reference

(1) Recommendation ITU-T H. 323, "Packet-based multimedia communications systems, Infrastructure of audiovisual services-Systems and terminal equipment for audiovisual services," 2003.
(2) IETF, "Real-time communication in WEB-browsers (rtcweb)," http://datatracker.ietf.org/wg/rtcweb/
(3) W3C, "WebRTC 1.0 : Real-time communication between browsers," https://www.w3.org/TR/webrtc/
(4) ISO/IEC 14496-10, "Coding of audio-visual objects-Part 10 : Advanced video coding,", Recommendation ITU-T H. 264, "Advanced video coding," 2003.
(5) ISO/IEC 23008-2, "High efficiency coding and media delivery in heterogeneous environments-Part 2 : High efficiency video coding,",
Recommendation ITU-T H. 265, "High efficiency video coding," 2013.
(6) IETF RFC 3261, "SIP : Session initiation protocol," 2002, https://tools.ietf.org/html/rfc3261
(7) IETF RFC 3550, "RTP : A transport protocol for real-time applications," 2003, https://tools.ietf.org/html/rfc3550,
(8) IETF RFC 3711, "The secure real-time transport protocol (SRTP)," 2004, https://tools.ietf.org/html/rfc3711
(9) "SMART communication & collaboration cloud," https://www.nttbiz.com/solution/vcs/service/smart_cc/

(Accepted on May 20, 2020; Finally accepted on June 3, 2020)

Shohei KAMAMURA
(Regular member: Senior member)

Received B.E., M.E. and Ph.D. degrees from Waseda University, Tokyo in 2004, 2006 and 2013, respectively. Joined Nippon Telegraph and Telephone Corporation in 2006. Since then, has been engaged in the research and development of IP and optical transport networks. Senior Researcher at NTT Network Service Systems Laboratories. PhD (Global Information and Telecommunication)