

Supporting Technologies for Hikari Living

Akira Nakagawa, Mushin Nakamura, Suehiro Shimauchi, Kiyoshi Nakahama, Takeshi Matsumoto, Ryuichi Tanida, Kazunori Kobayashi, and Koichi Sugiura

Abstract

Hikari Living is not only a means for visual communication but also a tool that includes a function for sharing content during visual communication. This tool enables family members separated by distance to visually share content from their living rooms. In this paper, we explain supporting technologies that enable *Hikari Living* to provide comfortable visual communication.

Keywords: content sharing, media tuning, echo canceller

1. Introduction

The *Hikari Living* concept includes the following features:

- New type of *Hikari* visual communication via a large television (TV) screen in the living room
- Easy and relaxed communication
- Shared atmosphere providing a feeling of togetherness

These features were achieved by using the four technologies below, which are explained in detail in the subsequent sections.

- (1) Content sharing technology on the *data connect* service provided in the Next Generation Network (NGN) *Hikari* telephone
- (2) Speech delay reduction technology to minimize speech delay, which is a requirement for fixed telephones
- (3) Video quality tuning technology to stream high definition video via the NGN *Hikari* telephone (about 1.8 Mbit/s)
- (4) Acoustic echo cancellation technology to stably reduce echo in the living environment of general households

2. Content sharing technology

Hikari Living enables users to share digital content such as pictures taken with a digital camera during

audio communication and visual communication. Steps for sharing pictures while carrying out visual communication are as follows (**Fig. 1**):

Step 1: The users establish visual communication.

Step 2: The sender chooses one photo from a selection of photos and sends the data to the receiver.

Step 3: When all of the data has been captured by the receiver, the screens of both the sender and receiver are automatically divided into right and left panes. The left side displays the visual communication image and the right side displays the shared data.

As described above, you can share digital content with the person you are speaking to. To realize this function, we used our new *Content Sharing During Communication Protocol*.

This protocol works are shown in **Fig. 2**. The protocol enables either the caller or callee (receiver) to send digital content. Only two requirements are set for the devices used.

Requirement 1: Both devices must be able to support incoming/outgoing communications (audio or audio/visual) and be capable of sharing digital content from either side.

Requirement 2: Both devices must be capable of determining the types of content the other side can process, for example, by previewing it.

To confirm that these requirements can be met, this

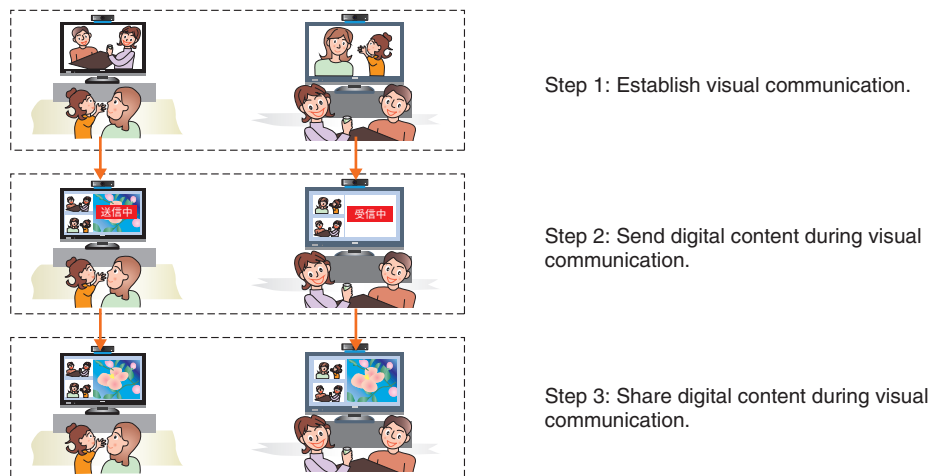


Fig. 1. Function to share pictures during visual communication.

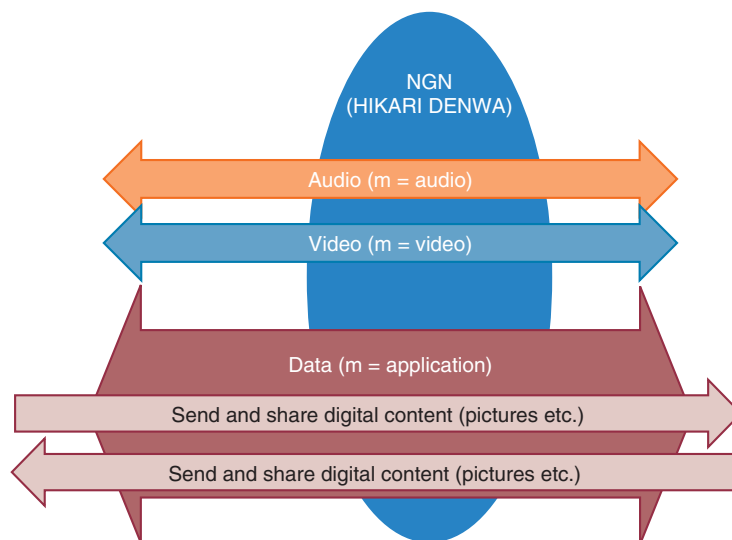


Fig. 2. Content sharing during communication protocol.

protocol enforces the following rules:

Rule 1: Establish transmission path in a TCP (transmission control protocol) session in order to share content multiple times with two-way commands (Fig. 3).

Rule 2: Exchange device capabilities such as what kinds of previewable content will be allowed after the path is established.

When we were designing the Content Sharing During Communication Protocol, the protocol design for normal state operation went smoothly, but of course, we also had to consider all abnormal states.

Specifically, we considered all situations (especially the collision of processes and collision between manipulation commands from the sender and receiver), in designing the protocol so that it could recover from all eventualities.

Consequently, designing the Content Sharing during Communication Protocol was rather a challenge given the complexity and frequency of abnormal states, as opposed to designing completely new services such as Hikari Living.

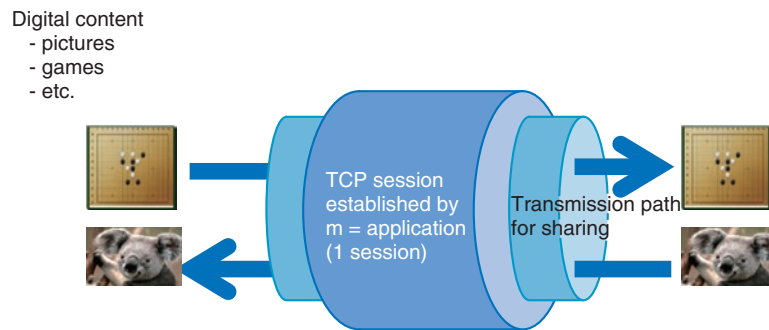


Fig. 3. Establishment of transmission path for digital content.

3. Speech delay reduction technology

In Hikari Living, the target for one-way speech delay is less than 150 ms. The assumed maximum network delay is 30 ms, so the upper limit of delay for the Hikari Living terminal is 120 ms. Furthermore, because the packet size of the Hikari telephone is 20 ms, 60 ms is required for the transmission of voice packets. This 60 ms comprises 20 ms each for capture delay, playback delay, and synchronization delay. The processing delay in the speech codec and echo canceller is also 20 ms. The remaining delay is 40 ms (= 120 ms – 60 ms – 20 ms). Therefore, it was necessary to implement a jitter buffer and synchronization buffer with delay of less than 40 ms. To achieve a small delay, it is necessary to reduce the buffer size. However, a shorter buffer size makes it easy for sound interruptions to occur as a result of external disturbances, e.g., if the CPU (central processing unit) load is too high. Accordingly, we attempted to determine the minimum buffer size that would not generate sound interruptions under various conditions. By controlling the read and write buffer sizes, we were able to control the accumulated amount of audio data to the minimum buffer size. With this control, the target for the one-way speech delay (less than 150 ms) was achieved.

4. Video quality tuning technology

We adjusted the configuration of the camera, encoder, and decoder units to improve video quality, while satisfying not only related standards but also the limitations for both bandwidth and delay time (Fig. 4).

First, the camera and its preprocessing section were tuned in order to capture clear and easily recogniz-

able images under low bitrate conditions. The light-receiving sensitivity of an image sensor in the camera was increased so as to obtain an easily viewable picture under dark conditions. In addition, a function to correct peripheral darkening, which is a typical characteristic of wide lenses, was added. It works by brightening the edges of problematic pictures. Furthermore, we developed a new adaptive noise reduction filter that maintains edge information in order to get a well-modulated picture.

Second, we focused on the encoder. For practical use of the limited bandwidth, we changed the rate of bit allocation for each picture because it greatly affects video quality. However, we checked the peak rate and delay time simultaneously because these factors are also affected by it.

The last adjustment involved the decoder. Because there is a tradeoff between latency and playout smoothness, we adjusted the buffer size in order to smooth out the network jitter and bit-rate fluctuations. Bit-rate fluctuations are the result of biased bit allocation by the encoder.

These tests were repeated under various illumination conditions to ensure stable operation.

5. Acoustic echo cancellation technology

When visual communication (or audio communication) is initiated on Hikari Living, you can hear the other person's voice from the TV's built-in loudspeakers connected to the Hikari Living terminal, as if you were listening to the sound of a TV program. You can talk to the other person without being aware of the microphone, since the small microphone is built into the Hikari Living camera device set on the TV. Natural hands-free visual communication is achieved even in such audio setup conditions through

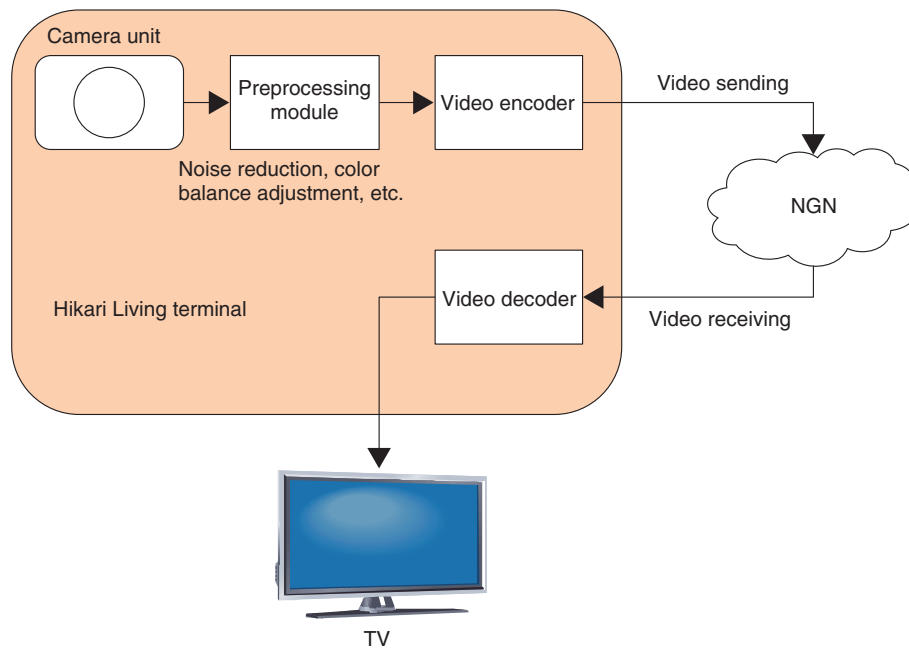


Fig. 4. Configuration of video transmission.

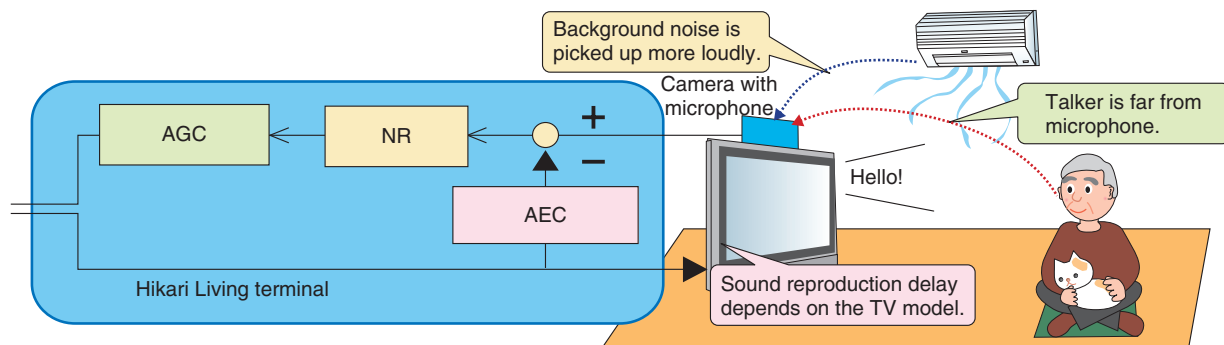


Fig. 5. Acoustic echo cancellation technology used with Hikari Living.

the implementation of newly developed audio processing functions, i.e., acoustic echo cancellation (AEC), noise reduction (NR) and automatic gain control (AGC), in the Hikari Living terminal (Fig. 5).

Acoustic echo caused by acoustic coupling between the loudspeaker and the microphone seriously disturbs visual communication. Therefore, AEC is necessary to cancel the annoying echo; it works by predicting the echo signal and subtracting the predicted signal from the microphone signal. Digital TVs usually reproduce sound with a relatively long delay after inputting the sound signal to the TV due to the syn-

chronization of the video processing. The length of the delay varies depending on the TV model or manufacturer. The AEC can automatically adapt to the delay differences of most TV models and can subtract the predicted echo in the proper timing so that a TV can be used as-is for Hikari Living.

Another problem can occur when the talker is too distant from the microphone on the TV. When the distance between the microphone and the talker increases, the talker's voice that is picked up is quieter and the background noise is picked up louder. We have addressed this issue by incorporating NR to

reduce the background noise and AGC to amplify the talker's voice with the AEC function.

6. Conclusion

Hikari Living utilizes many technologies and much

knowhow to provide convenience and ease of use to users. Moreover, tuning operations were performed repeatedly to improve convenience and service quality. We believe that Hikari Living has many benefits and will provide a lot of enjoyment to users.



Akira Nakagawa

Senior Research Engineer, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. degrees from Kyushu Institute of Technology, Fukuoka, in 1992 and 1994, respectively. Since joining NTT in 1994, he has been engaged in research on acoustic signal processing for echo cancellation. He received a paper award from the Acoustical Society of Japan (ASJ) in 2002. He is a member of ASJ.



Takeshi Matsumoto

Engineer and Researcher, Networked Appliance Project, NTT Service Evolution Laboratories.

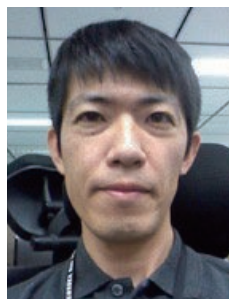
He received the B.E. and M.E. degrees in engineering from Waseda University, Tokyo, in 2006 and 2008, respectively. He joined NTT Service Evolution Laboratories (formerly NTT Cyber Solutions Laboratories) in 2008 and has been engaged in designing protocols of NGN appliances. He designed the NGN content transfer protocol, which enables digital content to be transferred by telephone like a fax.



Mushin Nakamura

Research Engineer, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.S. degree in mathematics and the M.E. degree in computer science from the University of Tokyo in 2001 and 2003, respectively. He joined NTT in 2003. He has been engaged in R&D of NGN appliances.



Ryuichi Tanida

Research Engineer, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. degrees in electrical engineering from Kyoto University in 2001 and 2003, respectively. He joined NTT to study video coding for developing a real-time H.264 software encoder application. He moved to NTT Advanced Technology Corporation (NTT-AT) in 2009 to manage various software development projects related to video compression, distribution, and TV phones. He has been at NTT laboratories since 2012, where he works as a research engineer in the visual media coding group. His current research focuses on the development of HEVC software encoders. He is a member of IEICE.



Suehiro Shimauchi

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 1991, 1993, and 2007, respectively. Since joining NTT in 1993, he has been engaged in research on acoustic signal processing for acoustic echo cancellers. He is a member of ASJ, the Institute of Electronics, Information and Communication Engineers (IEICE), and IEEE.



Kazunori Kobayashi

Senior Research Engineer, Audio, Speech, and Language Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. degrees in electrical and electronic system engineering from Nagaoka University of Technology, Niigata, in 1997, 1999, and 2003, respectively. Since joining NTT in 1999, he has been engaged in research on microphone arrays, acoustic echo cancellers, and hands-free systems. He is a member of ASJ and IEICE.



Kiyoshi Nakahama

Senior Research Engineer, Supervisor, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.E. degree in engineering from Muroran Institute of Technology, Hokkaido, in 1990. He joined NTT in 1990 and has been engaged in R&D of an IC card public telephone, shared use personal computers (Shared PC), NGN appliances, and an NGN development support environment. He is a member of IEICE.



Koichi Sugiura

Senior Manager, Terminal Equipment Technology Center, Application Solutions Business Headquarters, NTT-AT.

He received the B.A. degree from Tokyo University of Foreign Studies in 1995. Since joining NTT-AT in 2000, he has been engaged in software development, with a focus on lossless audio codec and a song synthesis engine.