Visual Softphone: New Ways to Communicate

Koichiro Kanaya, Hironori Ohata, Dai Ando, Machio Moriuchi, Hiroshi Jinzenji, and Jotaro Ikedo†

Abstract

NTT is providing a visual softphone for use on Windows personal computers as a tool for expanding the means of communication. This article provides an overview of it and describes various communication formats that it can provide.

1. Softphone with OAB-J number capability

The visual softphone is a software-based telephone terminal that can run on Windows personal computers (PCs). It is being provided by NTT EAST and NTT WEST free of charge to HIKARI DENWA*1 users of FLET'S HIKARI NEXT broadband service under the name HIKARI Softphone. The initial version of the visual softphone was provided in February 2009 and a content transfer function supporting data connections (enabling data files to be sent and received by a PC) was added in July 2010.

Although Skype and other softphone products are available on the market, achieving a short end-to-end delay has been a difficult problem to solve. At present, only a few softphones (including NTT's visual softphone) can achieve a delay of 150 ms, which is specified as a requirement for using 0AB-J numbers (10-digit numbers starting with 0).

2. Architecture

In general, software on a PC operates by receiving instructions (input) from the user, performing various types of processes, and returning results to the user. In contrast, a softphone operates not only by receiving instructions from the user but also by recognizing and processing an incoming call. A softphone must also support a variety of commercially available external devices such as headsets and cameras. In addition, a softphone by its very nature achieves most of its functions by software, which means that it supports the addition of diverse functions and modification of its operations in a relatively easy manner. To exploit this feature to the maximum, we need to use an architecture that can simplify function addition and specification modification as much as possible without sacrificing overall performance.

With these characteristics taken into account, the visual softphone can be broadly divided into utility, control logic, and graphical user interface (GUI) elements, as shown in **Fig. 1**.

The utility element enables the visual softphone to support various external devices by concealing differences in PC hardware components, and it uniformly supports incoming calls by performing SIP (session initiation protocol) processing and general-purpose processing such as media RTP (Real-time Transport Protocol). The logic element controls all softphone operations and controls input/output according to the state of the software. The GUI element handles input/output with respect to the user.

3. Main functions

The main functions of the visual softphone are listed in **Table 1**. Its functions for videoconferencing, broadband telephony, and content transfer greatly

† NTT Cyber Space Laboratories
Yokosuka-shi, 239-0847 Japan

*1 Hikari is the Japanese word for light; denwa is the Japanese word for telephone.

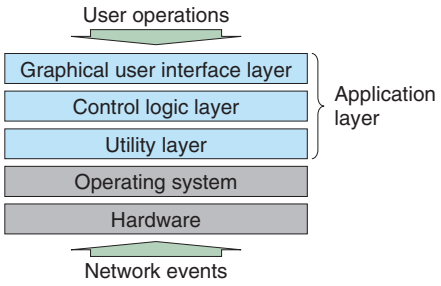


Fig. 1. Architecture overview.

Table 1. Main functions.

Video	VGA (640x480) @ 30 fps: MEPG-4
	QVGA (320x240) @ 15 fps: MPEG-4
	QCIF (176x144) @ 15 fps: MPEG-4
Audio	Wideband (7-kHz) voice: G.711.1 & UEMCLIP
	Narrowband (3.4-kHz) voice: G.711
Digital contents	Simultaneous contents transfer and voice call
	Simultaneous contents transfer and video call
	Individual contents transfer

fps: frames per second

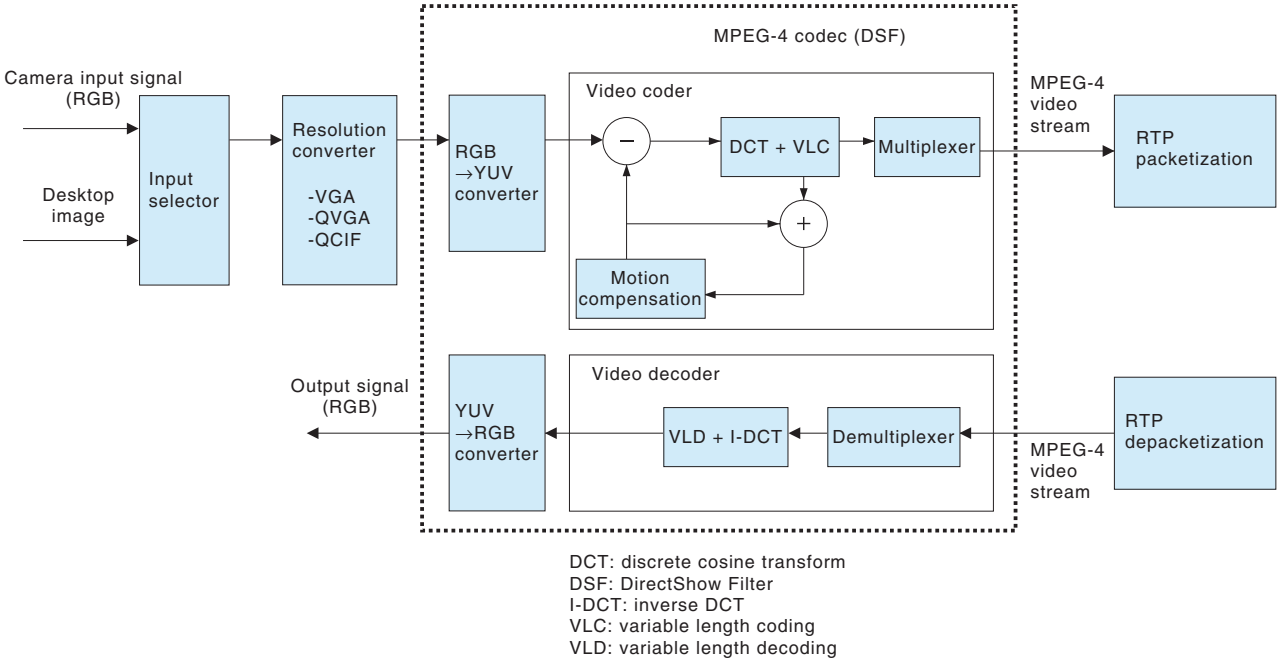


Fig. 2. Visual softphone's video processing module.

exceed conventional telephone functions. These functions can be optimally allocated to suit user needs.

3.1 Video communications function

The configuration of the visual softphone's video communications processing module is shown in **Fig. 2**. This module takes a video signal input from a camera and converts it to a previously set resolution in the resolution-conversion section. It then passes the resulting signal to the video coding section where it undergoes RGB-YUV[*2] conversion and compression/coding. The coding system uses an MPEG-4-based codec developed by NTT Cyberspace Laboratories and achieves high-quality video while suppressing the processing load. Supported video sizes are VGA (640 pixels × 480 lines), QVGA (320 pixels × 240 lines), and QCIF (176 pixels × 144 lines). In general, high-quality videophone communications is provided using VGA video, but QVGA video, which generates a somewhat smaller processing load for coding, is also supported to enable the use of relatively low-performance PCs such as netbook

*2 RGB: red, green, and blue; YUV: color space defined in terms of one luma and two chrominance components.

Table 2. Terminals that can connect to the visual softphone.

	Network	Terminals
Voice call	NGN	HIKARI Softphone
		HIKARI DENWA terminal
	PSTN	Analog terminal
		Mobile phone
		INS voice terminal
		050 number IP phone
		PHS (personal handy-phone system)

INS: Information Network System
NGN: Next Generation Network

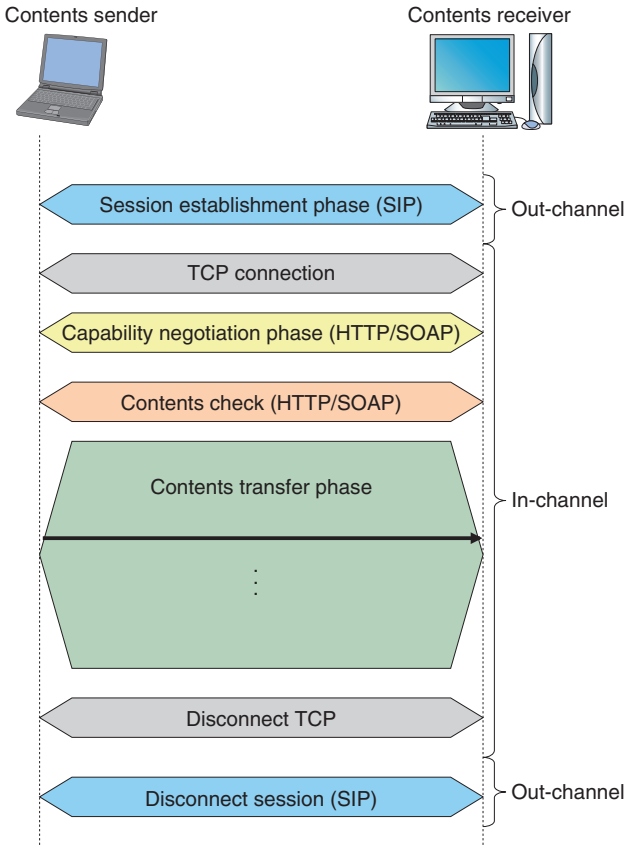


Fig. 3. Example of the contents transfer sequence.

computers. The visual softphone also supports videophone communications with FOMA mobile phones in the same way as the PC Communicator [1], for which QCIF video is used to match the video on the FOMA side.

Although the usual format is for a video signal to be input from a camera and then compressed and sent to the other party's terminal, a desktop sharing function can also be used. In this case, the input-switching section of the video communications processing module switches the input signal from a camera-fed video signal to a rectangular portion of the PC's desktop screen selected by the user. This signal is then compressed and sent to the other party. This configuration enables a screen image to be sent to the other party's terminal during videophone communications without the need to switch applications, which leads to smooth screen sharing.

3.2 Voice communications function

Call control on the visual softphone is performed by SIP, and calls can be made to a variety of PSTN (public switched telephone network) telephone terminals via a home gateway (**Table 2**).

In addition to the G.711 coding system used by the conventional telephone network for speech coding, the visual softphone is also equipped with the UEMCLIP system developed by NTT Cyberspace Laboratories and the G711.1 wideband speech coding system [2] to provide high-quality voice calls with little delay.

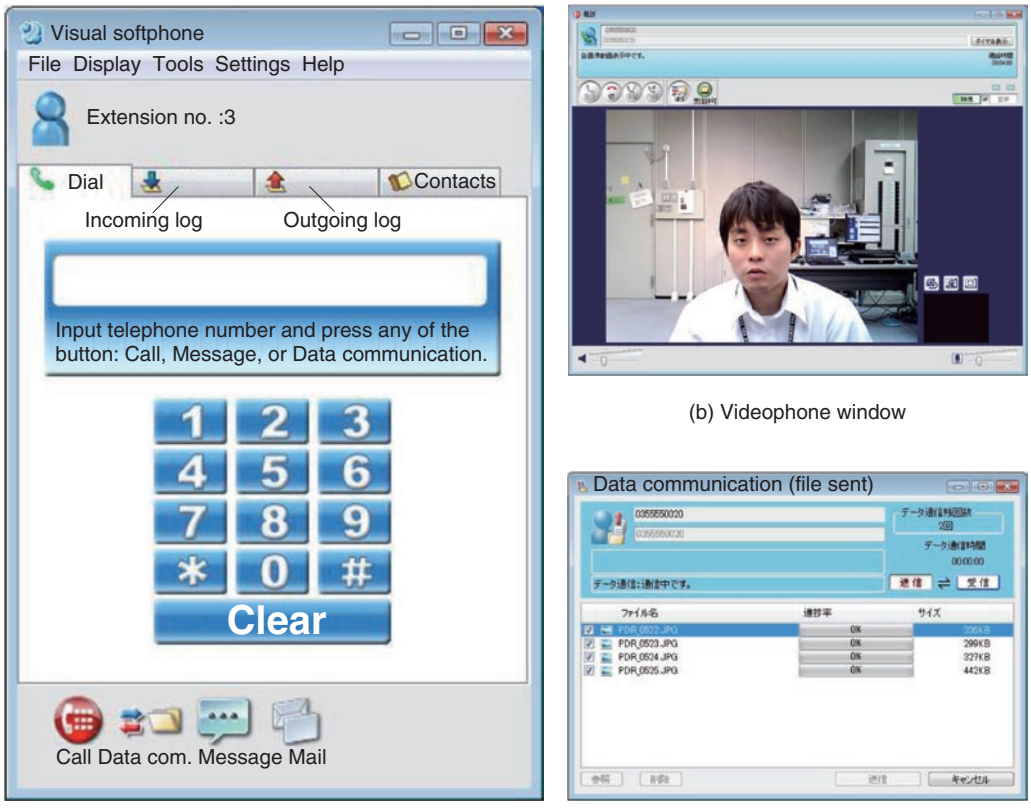
Switching between these speech encoding systems is performed automatically by a negotiation process at the beginning of a call. The user does not need to know what kind of terminal the other party is using.

It is common to incorporate a fixed-size buffer for voice communications that uses Internet protocol (IP) to deal with packet transmission fluctuations and packet loss. However, this buffer processing is directly related to increases in voice delay. In the visual softphone, this problem is dealt with through various measures such as dynamically controlling the buffer size and performing buffer processing frequently with the aim of preventing buffering-related delay as much as possible. The visual softphone also prioritizes the allocation of CPU (central processing unit) resources for voice packet processing to prevent unwanted effects from other applications.

The visual softphone also incorporates a simple echo canceller [3]. This enables the user to connect a microphone and speakers to the PC and perform hands-free calling in an environment where the distance between the microphone and speakers is no greater than 2 m.

3.3 Content transfer function

In addition to video and voice communications,



(a) Main window

(b) Videophone window

(c) Contents transfer window

Fig. 4. Examples of PC windows.

the visual softphone provides a content transfer service that supports the sending and receiving of digital data such as images and documents using the "m=application" media type in SIP/SDP (SDP: session description protocol). This service enables users to send and receive all sorts of digital data in the manner of fax transmission using 0AB-J numbers and even to send data that must be carefully handled such as personal information by utilizing the secure network characteristic of the Next Generation Network (NGN). In short, the content transfer service is expected to stimulate the creation of completely new usage formats.

The content transfer service uses technical specifications for content transfer established by NTT Cyber Solutions Laboratories. The visual softphone can connect to and transfer content to other devices that use the protocol specified in these technical specifications. In particular, the specifications prescribe a method for describing call settings on the basis of SIP/SDP and an in-channel communication protocol using HTTP/SOAP (HTTP: hypertext transfer protocol; SOAP: simple object access protocol) for use after the call has been established.

The content transfer service can be used in both a non-calling state and a calling state. An example of content transfer from a non-calling state is shown in **Fig. 3**. The session is started by a SIP INVITE command specifying m=application. This causes in-channel communications to begin. The softphone on the sending side connects to the softphone on the receiving side by TCP (transmission control protocol). It sends a list of digital content specified by the user (maximum 100 files), transmits the actual digital content, and closes the TCP connection and terminates in-channel communications. The sending softphone sends a SIP BYE command to terminate the session. For content transfer while a call is in progress, the softphone sends a SIP reINVITE command with m=application added to begin in-channel communications. After in-channel communications has been performed, media deletion by reINVITE(port0) is

performed to return the communications state to voice/video.

The NGN currently has an upper limit for the number of times that media addition can be performed during a voice/video call, and when m=application media addition is performed at the time of content transfer, a limit can be set for the number of content items that can be transmitted at one time. We therefore considered how to exceed such limits in content transfer by enabling the sending-side softphone to select whether content transfer will continue after one transmission. If it does continue, the user can specify and send new content in the same session by continuing the m=application session.

4. User interface

The GUI is divided into a main window, telephone window, and content transfer window (**Fig. 4**). Information needed at non-calling times, such as contacts and calling history, is all contained in the main window to minimize the display area occupied during non-calling times. At the time of a call, the GUI automatically switches to the telephone window; the procedure for accepting an incoming call was carefully designed to be uncomplicated. The content transfer window enables files for transfer to be selected by drag-and-drop to facilitate intuitive use similar to the operation of other Windows applications.

5. Conclusion

This article outlined NTT's visual softphone, which enables people to communicate using various types of new media on the NGN. Looking beyond the visual softphone's obvious use as a videophone, we intend to propose new ways of communicating using diverse media.

References

[1] H. Takeda, D. Ando, S. Sakauchi, S. Onishi, and H. Jozawa, "Functions of PC Communicator: PC-to-FOMA IP Videophone Technology," NTT Technical Review, Vol. 3, No. 10, pp. 18–24, 2005. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200510018.pdf

[2] S. Sasaki, T. Mori, Y. Hiwasaki, and H. Ohmuro, "Global Standard for Wideband Speech Coding (G.711 wideband extension)," NTT Technical Review, Vol. 6, No. 8, 2008. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200808le1.html

[3] S. Sakauchi, Y. Haneda, M. Okamoto, J. Sasaki, and A. Kataoka, "Echo Canceller with Noise Reduction Provides Comfortable Hands-free Telecommunication in Noisy Environments," NTT Technical Review, Vol. 2, No. 3, pp. 59–63, 2004. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200403059.pdf

Koichiro Kanaya
Senior Research Engineer, Promotion Project 1, NTT Cyber Space Laboratories.
He received the B.E. degree in mechanical engineering from Waseda University, Tokyo, in 1990. In 1990, he joined NTT Video and Record Communications Division and has mainly been engaged in developing video communication systems.

Hironori Ohata
Senior Research Engineer, Network Appliance and Services Project, NTT Cyber Solution Laboratories.
He received the B.E. degree in communication engineering from Tokai University, Kanagawa, in 1989. In 1989, he joined NTT Telecommunications Equipment Division and has mainly been engaged in the development of telecommunications equipment. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Dai Ando
Research Engineer, Promotion Project 1, NTT Cyber Space Laboratories.
He received the B.E. degree in information technology from Tohoku University, Miyagi, in 1989. In 1989, he joined NTT Human Interface Laboratories and has mainly been engaged in developing video communication systems. He is a member of IEICE.

Machio Moriuchi
Senior Research Engineer, Promotion Project 1, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in electrical engineering from Tokyo Science University in 1987 and 1989, respectively. He joined NTT Electrical Communication Laboratories, Kanagawa, in 1989. He has been engaged in R&D of videoconferencing systems.

Hiroshi Jinzenji
Senior Manager, Network Appliance and Services Project, NTT Cyber Solution Laboratories.
He received the B.E. and M.E. degrees in communication engineering from Osaka University in 1989 and 1991, respectively. In 1991, he joined NTT Human Interface Laboratories and has mainly been engaged in developing customer premises equipment. He is a member of IEICE.

Jotaro Ikedo
Senior Research Engineer, Promotion Project 1, NTT Cyber Space Laboratories.
He received the B.E. degree in electronic engineering and the M.E. degree in electrical engineering from Kogakuin University, Tokyo, in 1989 and 1991, respectively. Since joining NTT in 1991, he has been engaged in R&D of low-bit-rate speech coding and wireless transmission. He contributed to establishing the ARIB STD-27 and ITU-T G.729 standards. He is now developing VoIP systems. He is a member of IEEE, IEICE, and the Acoustical Society of Japan.