

Creating an Appealing User Experience by Applying Media System Technology

Satoshi Takahashi, Yushi Aono, Shiro Ozawa, Hidenori Okuda, Atsushi Sagata, and Ryuichi Tanida

Abstract

This article introduces NTT efforts aimed at creating appealing user experiences by applying the wide variety of media system technology under development at NTT, including technologies related to speech, language, audio, still images, and video. The work described here focuses on two areas: a personal agent that is intimately close to the user and provides personalized services that stimulate human knowledge and behavior, and high-sense-of-presence media services that enable the user to enjoy extremely natural viewing experiences.

Keywords: personal agent, high sense of presence, media system processing technology

1. Creation of personalized services that stimulate human knowledge and behavior

1.1 Virtual agents

In recent years, virtual agents that can satisfy the various needs of individual users have been attracting attention. In particular, virtual agents that respond to voice input and can retrieve weather reports and answer simple questions have been implemented in smartphones and other personal devices, so such agents are now more readily available. However, these kinds of services are positioned as an input option for web search functions, and they simply present the results of searches performed using the given keywords. This form of use involves a single question and a single response, and is therefore limited to executing just one function of the original roles of the virtual agent. Currently, such virtual agents are mostly confined to virtual worlds such as the web, and the information they can retrieve is limited to that domain. They cannot use information from the user's real-world situation to interact with and influence the user. We consider this to be a major obstacle to be overcome in developing future virtual

agents.

1.2 NTT's concept of a virtual agent

In view of the situation described above, NTT has shed the idea of a virtual agent that is confined to a virtual world and interacts with users in the form of a single response to a single question in favor of a personal agent that understands the user intimately and exists together with the user in the real world (**Fig. 1**). NTT is now moving forward with research and development (R&D) to realize such an agent. The NTT concept of a personal agent involves three important elements. One is that the agent can understand the user's situation and intentions in the context of the real world through technology that senses and processes various kinds of media. Another is that it can actively influence the user based on its understanding of the user's situation and intentions. The third element is that it grows together with the user by understanding the user's situation and influencing the user accordingly. We believe that implementing these elements requires technology for understanding real-world situations and organizing and structuring that information (real-world structuring technology), and

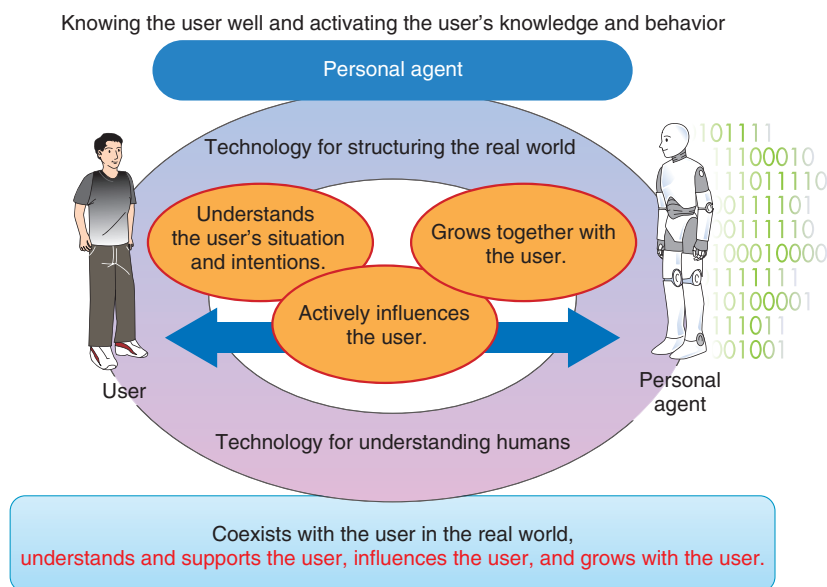


Fig. 1. Important elements of a personal agent.

technology for understanding both the explicit and latent intentions of the user (technology for understanding humans). We are developing various types of media system technology to support the required technology.

1.3 Evolution of personal agents

NTT has set two milestones for developing the personal agent and is moving forward with a policy for achieving the ultimate goal of this technology (Fig. 2). The first step is to create a user profile by collecting information on the user's interests and preferences from dialogs with the user and to function as a kind of servant or butler by providing appropriate support based on the profile. The second step is to sense the user's present situation, the ambient mood, and the user's expressions and state, and adaptively support the user accordingly, like a friend. The ultimate goal is to go beyond simply presenting a short-term optimum solution. Rather, the personal agent will anticipate future situations and influence the user with care and understanding on that basis, like a family member.

For example, consider a user who is trying to lose weight. Rather than recommending a nearby fast food restaurant, the agent would respond with proper concern for the user's goals and recommend a restaurant that has a health-oriented menu, even if it involves a circuitous route. We believe the personal agent towards which NTT is working will be capable of

providing a new and appealing user experience unlike any virtual agent that now exists.

These Feature Articles describe technology that supports the NTT concept of a personal agent and presents specific examples. The article, "Media Processing Technology for Achieving Hospitality while on the Go," describes a service that guides the user around a city, statistical machine translation technology for presenting guide information, and robust media search technology for recognizing objects in an image [1]. "Media Processing Technology for Achieving Hospitality in Information Search" describes a service for assisting users in their daily activities and subject identification technology for searching the Internet for information related to an image captured by a camera, natural language processing technology for understanding the user's intention and responding in a natural way, and user-designed speech synthesis technology for generating synthesized speech for various speakers and speaking styles [2]. "Media Processing Technology for Business Task Support" introduces technology that holds promise for applications extending beyond these service scenarios to business scenarios [3].

2. Appealing sense-of-presence media services

2.1 High-definition video

In Japan, digital broadcasting via communication satellite began in 1996. That was followed by digital

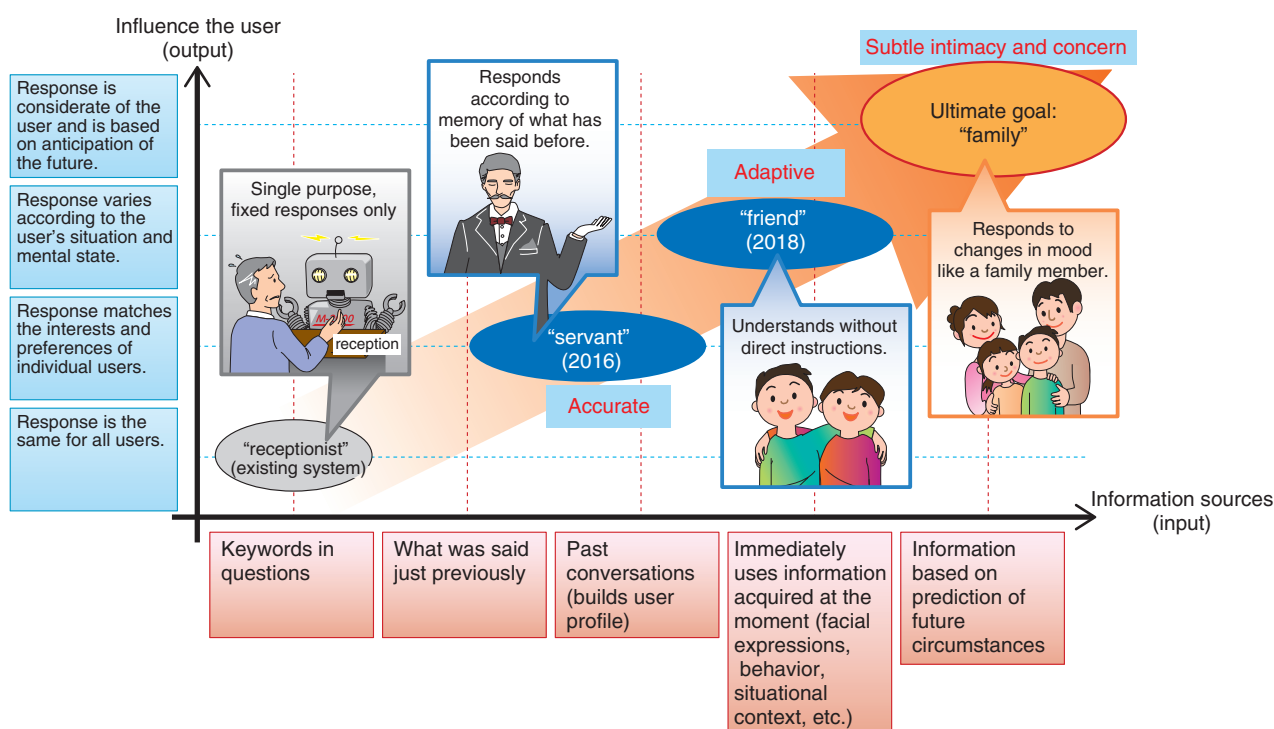


Fig. 2. Evolution of the personal agent.

broadcasting via broadcast satellite in 2000 and by terrestrial digital broadcasting in 2003. The current high-definition television (HDTV) video format^{*1} provides a remarkable improvement in image quality compared with analog broadcasting and is now used for almost all programs that are broadcast.

The next generation of high-definition video media is said to be the 4K and 8K formats,^{*2} which provide an overwhelmingly superior feeling of detail and representation of reality compared to HDTV and can be used to provide services that create a high sense of presence. The 4K video format was first introduced in movie theaters in 2007, and since then, the number of screens has been increasing. As a result, various types of practical 4K equipment have been developed, including projectors and cameras for professional use. Furthermore, the Next Generation Television and Broadcasting Promotion Forum (NexTV-F) began conducting test broadcasts in the 4K format in June 2014. Consumer-use 4K-resolution displays are also appearing on the market, and home use of 4K TV is becoming more popular as well.

2.2 Trends in Japan and NTT related to high-definition video

Countries around the world are putting more effort into achieving 4K and 8K broadcasting, and a world-leading roadmap for commercialization of these formats has been formulated by a study group of Japan's Ministry of Internal Affairs and Communications. Furthermore, an interim report from a follow-up meeting recommended acceleration of the roadmap to promote 4K and 8K broadcasting (Fig. 3) [4].

For the implementation of 4K and 8K telecom and broadcasting services, broadcasters, telecom carriers, and equipment manufacturers have established the NexTV-F as an organization for cooperation. NTT is a proponent of the organization and is therefore collaborating with various enterprises to push forward with the implementation of the world's most advanced 4K and 8K services.

*1 HDTV: 1920 × 1080 pixels, also referred to as 'Hi-Vision' in Japan.

*2 4K/8K: A high-definition video format that has twice the horizontal resolution and four times the vertical resolution of HDTV. 4K and 8K together are also referred to as ultra-high definition (UHD).

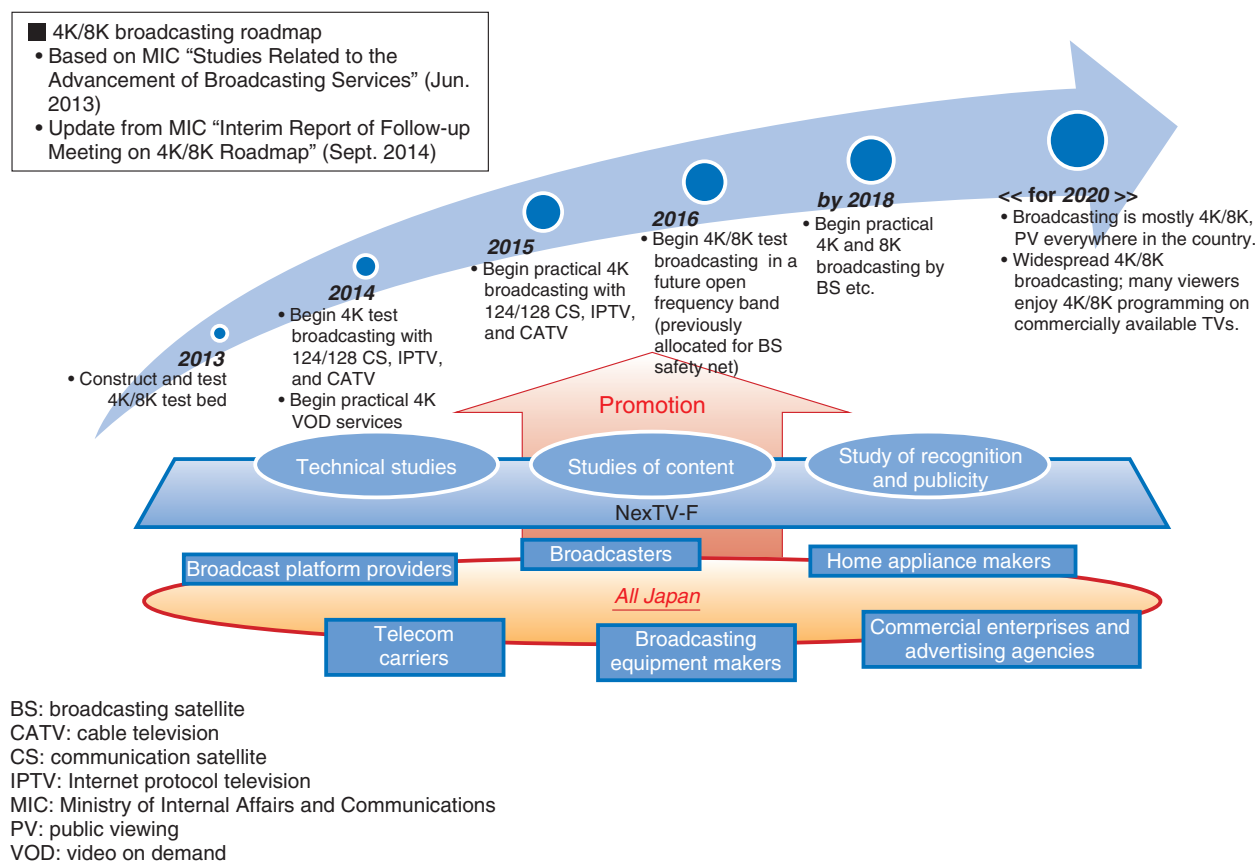


Fig. 3. 4K/8K broadcasting roadmap for Japan.

2.3 Toward implementing high-sense-of-presence media services

The NTT vision for future high-sense-of-presence media services is to realize rich life environments by providing user experiences that combine high-definition video, high-definition audio, and high sense of presence.

The subjects for ongoing R&D for telecom and broadcasting services that use high-definition media include HEVC (High Efficiency Video Coding) encoding technology, MMT (MPEG Media Transport) transmission technology, FireFort®-LDGM FEC (Low-Density Generator Matrix Forward Error Correction) technology, and other elemental technologies that are essential to service implementation [5–7].

However, implementing high-sense-of-presence media services that go beyond high definition requires more than simply improving resolution, compression quality, and transmission quality. What is needed is innovative R&D that can produce technology for

reproducing the sensation of being in a certain place or the feeling of being able to understand even more than one could understand by being in that place.

The article “Audio-visual Technology for Enhancing Sense of Presence in Watching Sports Events” describes five areas of innovative technology that NTT is working on to implement high-sense-of-presence media services [8]:

- Interactive distribution technology for omnidirectional video that enables users to view any region within a 360° video image as they choose
- Lossless audio encoding technology for compressing high-resolution, high-quality audio
- Distribution and encoding for arbitrary point-of-view video for composing video from points of view where cameras cannot be placed, such as the line of sight of the goalkeeper or the ball
- A zoom microphone system for extracting voice signals from a remote source to give the user a sensation of being on the playing field

- Reverberation removal and control technology for separating a music signal into direct sound and sound that arises indirectly from reflection from walls, etc., to reproduce sound that creates a sense of presence

3. Future development

We believe that the media system technology NTT is developing will spread throughout the world and bring about new, appealing user experiences via a variety of services. For the personal agent, we will continue investigating forms of service and technology that make it possible to respond to user needs, and we will develop services in collaboration with partners in various fields. For the high-sense-of-presence media services, too, we will continue to push for the implementation of 4K/8K services and promote innovative media technology for the evolution from *high definition to high sense of presence* in cooperation with broadcasters, video distributors, and other partners.

References

- [1] M. Horii, K. Arai, M. Nagata, K. Kashino, K. Hiramatsu, A. Fukayama, and H. Yamaguchi, "Media Processing Technology for Achieving Hospitality while on the Go," NTT Technical Review, Vol. 13, No. 4, 2015.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa4.html>
- [2] K. Sadamitsu, J. Shimamura, G. Irie, S. Tarashima, T. Yoshida, R. Higashinaka, H. Nishikawa, N. Miyazaki, Y. Ijima, and Y. Nakamura, "Media Processing Technology for Achieving Hospitality in Information Search," NTT Technical Review, Vol. 13, No. 4, 2015.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa5.html>
- [3] T. Oba, K. Kobayashi, H. Uematsu, T. Asami, K. Niwa, N. Kamado, T. Kawase, and T. Hori, "Media Processing Technology for Business Task Support," NTT Technical Review, Vol. 13, No. 4, 2015.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa6.html>
- [4] The Ministry of Internal Affairs and Communications, "Announcement of Interim Report of Follow-up Meeting on 4K/8K Roadmap," press release issued on September 2014 (in Japanese).
http://www.soumu.go.jp/menu_news/s-news/01ryutsu11_02000039.html
- [5] T. Onishi, T. Sano, K. Yokohashi, J. Su, M. Ikeda, A. Sagata, H. Iwasaki, and A. Shimizu, "HEVC Hardware Encoder Technology," NTT Technical Review, Vol. 12 No. 5, 2014.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa3.html>
- [6] K. Iso, M. Kitahara, R. Tanida, T. Mitasaki, N. Ono, and A. Shimizu, "World's Highest-performance HEVC Software Coding Engine," NTT Technical Review, Vol. 12, No. 5, 2014.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa4.html>
- [7] T. Nakachi, T. Yamaguchi, Y. Tonomura and T. Fujii: "Next-generation Media Transport MMT for 4K/8K Video Transmission," NTT Technical Review, Vol. 12, No. 5, 2014.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa6.html>
- [8] D. Mikami, Y. Kunita, Y. Kamamoto, S. Shimizu, K. Niwa, and K. Kinoshita, "Audio-visual Technology for Enhancing Sense of Presence in Watching Sports Events," NTT Technical Review, Vol. 13, No. 4, 2015.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa7.html>



Satoshi Takahashi

Executive Manager, Executive Research Engineer, Supervisor, Audio, Speech and Language Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. in information science from Waseda University, Tokyo, in 1987, 1989, and 2002, respectively. Since joining NTT in 1989, he has been engaged in research on speech recognition, spoken dialog systems, and pattern recognition. He is a member of the Acoustical Society of Japan and the Institute of Electronics, Information and Communication Engineers (IEICE).



Yushi Aono

Senior Research Engineer, Supervisor, Producer, Media Component & System Produce, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.E. in biotechnology and the M.E. and Ph.D. in control engineering from Osaka University in 1994, 1996, and 1999, respectively. Since joining NTT in 1999, he has been engaged in research on speech synthesis, speech recognition, and spoken dialog systems. He is a member of IEICE and the Information Processing Society of Japan.



Shiro Ozawa

Senior Research Engineer, Producer, Media Component & System Produce, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. from Tokyo University of Mercantile Marine (now, Tokyo University of Marine Science and Technology) in 1997 and 1999. After joining NTT as a researcher, he worked for NTT Cyber Space Laboratories. In 2005, he was transferred to NTT Comware Corporation as an engineer. He returned to NTT Cyber Space Laboratories in 2010 as a research engineer. His fields of interest are computer vision, video communication, user interfaces, and 3D displays. He received the Young Author's Award from ISPRS (International Society for Photogrammetry and Remote Sensing) in 1998 and the Encouragement Award from the Institute of Image Electronics Engineers of Japan (IEEEJ) in 2004. He is a member of the Association for Computing Machinery, the Society for Information Display, IEICE, the Institute of Image Information and Television Engineers (ITE), and IEEEJ.



Hidenori Okuda

Director of Research, Executive Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. in precision machinery engineering from the University of Tokyo in 1986 and 1988, respectively, and the M.S. in computer science from Stanford University, CA, USA, in 1994. In 1988, he joined NTT Human Interface Laboratories and was mainly involved in R&D of visual communication systems. From 1997 to 2003, he was engaged in multimedia business development in the business sections of NTT and NTT-ME Corporation. He is currently interested in creating ultra-realistic remote viewing systems. He is a member of IEICE and ITE.



Atsushi Sagata

Senior Research Engineer, Supervisor, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E. in electronic engineering from the University of Tokyo in 1994. He joined NTT in 1994 and has been engaged in R&D of video coding systems at various NTT laboratories. During 2001–2005, he was engaged in the commercial development of a video transmission system for terrestrial digital TV broadcasting at NTT Communications. He then worked on the development of an audio-visual transcoder system for terrestrial digital TV broadcasting IP retransmission services at NTT Cyber Space Laboratories until 2011. His current research interests include digital image processing, real-time hardware/software UHD-video coding systems, and UHD-video transmission systems. He is a member of IEICE.



Ryuichi Tanida

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

He received the B.E. and M.E. in electrical engineering from Kyoto University in 2001 and 2003, respectively. He joined NTT to study video coding in order to develop a real-time H.264 software encoder application. He moved to NTT Advanced Technology Corporation in 2009 to manage various software development projects related to video compression, distribution, and TV phones. He has been at NTT Media Intelligence Laboratories since 2012 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.