Invited Paper

Digital to Natural - Innovation for Smart World

Susumu Yamamoto^{*}, Akira Nakayama^{*}, and Katsuhiko Kawazoe^{*}

*NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Japan {susumu.yamamoto.vk, akira.nakayama.vr, k.kawazoe}@hco.ntt.co.jp

Abstract - With the development of information and communication technologies, it is hoped that a world in which all people can live bountiful and happy lives can be achieved using innovative technologies. In other words, a Smart World. "Digital to Natural" is a transformation that is crucial to turning the concept of a Smart World into a reality. It means not only pursuing the ultimate digital vision of high-speed, high-capacity, high-definition performance, but also creating new value that can be achieved by naturally capturing and making the best use of a variety of information that previously could not be captured by humans. This will allow people to naturally and unconsciously benefit from technology. This paper describes what should be considered in order for technology to evolve into a more natural form, and shows technologies that support it such as AI, visual media, and ICT infrastructures - IOWN. This paper also presents concepts for several services that this technology can enable.

Keywords: natural, generous AI, Kirari, IOWN, Point of Atmosphere

1 INTRODUCTION

With the development of information and communication technologies (ICT), the society of the future will be the fifth society to evolve, following on from the previous four stages of human social evolution. Society 5.0 can be regarded as a super smart society that takes the fullest advantage of digital innovation[1]. The world that will be created through society 5.0 is a Smart World, in which everyone's lives and society itself will be totally changed for the better in every way.

How can we create this new Smart World? In creating new value for society, it is of course important to consider "technology innovation" that seeks to improve performance, such as communication capacity, but it is also important to consider "value innovation" that creates new social values at the same time.

ICT has evolved as the frontier of "digital." But should we continue along this path? In other words, what should we aim for next at this time when "digital" has reached its peak? We believe that what we should aim for is embodied by the keyword "natural." That is "Digital to Natural." The next step in digital's evolution is towards natural.

This paper describes what should be considered so that technology can evolve into a more natural form, and shows the technology that supports it and the concepts of the services that it enables. The rest of this paper is organized as follows; Section 2 presents some examples of what we think of as natural. Section 3 shows the direction in which the technologies supporting natural should go, particularly for AI, visual media, and ICT infrastructures. Section 4 presents the concepts of services that blend naturally into people's lives, and Section 5 concludes this paper.

2 DIGITAL TO NATURAL

"Digital to Natural" is the next paradigm change for technologies. This section describes two viewpoints: why and what we think of as natural for the evolution of technologies.

2.1 "Natural" for Diverse Value Judgement

In a diverse world, inhabited by people of all nationalities, ages and backgrounds, it is full of different value judgements. In such a world, "natural" is important if individuals and businesses, regardless of their characteristics, are to benefit from technology. Sometimes it keeps a caring eye on the lives of people without them being aware of it; sometimes it helps people do things more efficiently; and sometimes it appeals to their emotions. It also helps to provide a comfortable environment that is friendly to people and the global environment. We think the future of ICT should be seen in this way.

For technology to make this vision a reality, it is not enough to develop high-definition and high-sensitivity sensors and obtain more information. We will need innovative information processing that understands the senses of others and the subjectivity of each individual, and to look beyond conventional wisdom toward the diversity of life and systems. Where the result of such technology can be enjoyed by humans without feeling any stress or discomfort, we call this congenial state "natural" and this state is the objective which we pursue.

2.2 Creating Innovation through Learning from Living Things

The diversity of values is not limited to the human world. For example, imagine beautiful yellow flowers. Human beings perceive colors in a certain way that allows them to appreciate and enjoy the beauty of flowers (Figure 1(a)). The way in which honeybees perceive flowers, on the other hand, is very different as they can see ultraviolet rays, and the center of the flower is emphasized (Figure 1(b)). For honeybees, a beautiful-looking flower has no intrinsic value. However, it is important for them to know where the nectar and pollen of flowers is in order to live, and that is the "value" for honeybees.

Let's take another example; imagine mantis shrimps in the sea. They have receptors that can discriminate as many as 12 different colors [2]. Humans discriminate intermediate colors based on three primary color (red, blue, and green) receptors and information processing in the brain. In contrast, mantis shrimps sense things extremely rapidly with their 12-color receptors and minimal information-processing (Figure 2). What's of value to them is a high-speed response in order to capture their prey moving through the water. In the case of honeybees, what's remarkable is that they can see what humans cannot. What is notable about mantis shrimps is that they have a mechanism for directly processing information.

For a long time, humans have created innovations through learning from other living things, especially by mimicking them. The designs of jet planes and trains are examples of using such a mimicry approach. We want to take it a step further. Every animal species has a unique sensory world in which it lives. Jakob von Uexküll, a German biologist, called this phenomenon "umwelt" to signify a self-centered world, using the German term for "environment." [3] It means that how things look varies depending on the viewer. Accordingly, the type of information a species transmits and the way it processes information vary depending on what is of value to that species. We believe that if we regard the world as a field that holds diverse values, and if we can use "natural" information as it exists in the world, hitherto unused due to the digitization of current sensors being limited, such as digital sampling and quantization, we will be able to create new values that we could have never previously im-



Photo by Bjørn Rørslett / Science Photo Library (Aflo) Figure 1: Flower as viewed by humans and honeybees.



Figure 2: Comparison between the ways in which humans and mantis shrimps see something.

agined.

3 TECHNOLOGIES FOR NATURAL

This section provides some examples of technologies that can lead to natural value, and how these technologies should evolve hereafter, particularly for AI, visual media, and ICT infrastructure.

3.1 AI

(1) Natural conversation with robots

The first one is interaction with robots. Even now, there are examples of robots being used in shops to guide customers. But they are not able to function at a particularly high level. This is because that, currently, most robots can only converse within the context of a predetermined scenario, or are capable of asking only simple questions and providing simple answers. However, human conversations frequently involve unscripted small talk and discussions.

Higashinaka et al. [4][5] have developed robots which can enter into a more realistic discussion. This research led to the creation of a discussion structure that allows opinions about certain topics to be expressed, and they developed a mechanism to extract opinions that either support or disagree with a certain proposition depending on what the other party says. Through the evolution of this kind of technology, human activities are supported more naturally by making people and AI understand each other better through authentic conversations.

(2) Crossmodal voice / face conversion

Another example is how AI now has the ability to discern sound in a more human way. Humans can conjure an image of what a person's face looks like from the impression created by his/her voice, or what someone's voice may sound like from the impression created by his/her face. Kameoka et al. [6] showed how AI might be able to do this. They developed a cross-modal voice conversion model using deep generative models, which can convert speech into a voice that matches an input face image and generate a face image that matches the voice of the input speech by leveraging the correlation between faces and voices. Humans can recognize things by using different senses, such as vision and hearing. We believe it is important to create natural AI that is more responsive to people's feelings by providing AI with these unique human abilities.

(3) Seeking more natural AI

The performance of AIs when listening, speaking, and viewing things is improving. The time will come when it is more than adequate. We believe that at some stage in the future, thinking AI that supports human thought processes will become more important than ever.

The future of AI should be one that incorporates various values and helps people to consider complicated problems for which there is no single answer. For example, the best route to travel from one place to another may depend on a person's values, personality, circumstances, and habits, such as wanting to go quickly, have fun, or go safely. We believe it is important for AI to understand these differences and handle variations in values in order to find the best solution for each individual. We call such AI "generous" AI.

3.2 Ultra-realistic Viewing

(1) Kirari

High image definition technologies such as 4K and 8K will create a greater than ever sense of excitement. However, it is questionable whether it is effective to continue increasing the resolution in order to further raise the level of excitement. People feel a sense of presence through complex perceptual processing, and it is important to utilize multiple types of sensory information beyond the monitor frame. The objective of the "Kirari!" project [7][8] is to provide more natural video services.

The concept of Kirari was announced in 2015 and used a table tennis match as an example (Figure 3). The table tennis tables were real, but the players and the balls were not; they were actually displayed in virtual 3D. This demonstrated that by combining real and virtual, you could experience sports with a very realistic feeling, just as if you were present in the stadium. The most important aspect of the Kirari concept is that it creates a natural and highly realistic space by decomposing a scene into its constituent elements such as video streams, audio, lighting and other sensory information, transmitting these elements separately, and then recomposing them in a way that is best suited to the conditions at the viewing site.

(2) Ultra-wide viewing

Another technical feature of Kirari is ultra-wide viewing. For example, windsurfing competitions are difficult to follow as they are taking place far away from the beach, and spectators cannot understand what is happening in a contest if they can only see a specific part of the course. To understand what is taking place, even a 4K screen alone is not



Figure 3: Kirari.



Figure 4: Ultra-wide viewing.

enough. Kirari's ultra-wide viewing technology stitches together multiple 4K screens in real time to create a very wide screen at a remote venue (Figure 4). The spectators will actually feel as if they are present at the venue and watching what is taking place. This can be applied to many different sports including soccer and baseball [9][10]. It also makes it possible for spectators who cannot easily visit a distant venue, such as those confined to a wheelchair, to enjoy sports.

(3) Use case of Kirari

Kirari has also led to innovations in the field of entertainment, for example, new Kabuki productions using ICT. In the show "Cho Kabuki," a virtual idol called Miku Hatsune, and a real Kabuki actor, Shido Nakamura, were able to perform together using ICT [11]. Many people, especially young people who are not familiar with Kabuki, watched Cho Kabuki and discovered how entertaining Kabuki really is. It has really contributed to the creation of new values.

Kirari will continue to evolve. Its aim is not only to convey images and sound as they are, but also to activate people's sensitivity and psychology based on stories, such as those from history, and knowledge. This is truly Natural, as it conveys emotions to the human mind.

3.3 ICT Infrastructure

(1) IOWN

The conventional motivations for development of information and communication processing have been to increase speed, capacity and efficiency using digital signal processing. For example, the Internet has proven useful for providing many services and supporting business activities through adoption of common protocols to make an inexpensive network available based on the "best-effort" principle. Today, the volume of Internet traffic is increasing exponentially. Cisco estimates that global IP traffic is growing at a rate of 26% per year, and will total nearly 400 exabytes of traffic per month in 2022 [12]. However, there are limits. As traffic and data processing volumes grow, the power consumption of IT equipment also continues to increase. To date, integrated circuit performance has continued to increase exponentially, but we are approaching its limits.

If we are to evolve technology to simultaneously capture the diverse values that exist in the world, we need to penetrate new technical domains. Both the umwelt of honeybees and that of mantis shrimps actually exist in this world. But, these worlds are missing from the current world of IP or the digital world. We want to convey sufficient information and process it appropriately based on diverse values, and to provide benefits in a natural way.

To achieve this innovation, NTT, Intel and Sony have announced a new infrastructure concept called IOWN, Innovative Optical and Wireless Network [13][14]. IOWN is aimed at providing an innovative information processing platform that supports processing on a massive, unprecedented scale, and brings about changes that surpass the limitations of conventional technologies, such as power consumption barriers.

Figure 5 shows the three elements that comprise IOWN:

• All-photonics network is designed to dramatically enhance the potential of the information processing base.



Figure 6: Development of technologies towards Photonics-electronics convergence

- Digital twin computing is designed to create a new environment for services and applications.
- The goal of the Cognitive Foundation is optimal harmonization of all ICT resources.
- (2) All-photonics network

Optical technology, which was originally used for longdistance transmission such as intercontinental networks, has come to be used for short-distance transmission like FTTH and intra-data-center networks. IOWN's all-photonics network expands the applications of optical technology, applying it not only to networks but also to inside semiconductors in terminals and servers. The foundation for achieving an all-photonics network is photonics-electronics convergence. This makes it possible to process light and electricity on a single chip (Figure 6). For example, Nozaki et al. [16] developed a femtofarad-scale optical transistor using a photonic-crystal platform, which reduces power consumption by about two orders of magnitude in comparison to conventional devices (Figure 7). All-photonics does not necessarily



Figure 7: Ultra-low-energy and high-speed optical transistor.

mean that all electrical components will be replaced with optical parts but, rather, to use optical technology everywhere.

The all-photonics network has three technical advantages. The first is low power consumption. We solve the heat problem by incorporating photonics technology into everything from networks to terminals. The second is transmission capacity. Multi-core, multi-wavelength communication technology is increasing the capacity of 1 optical fiber up to 1 peta bps. Yet another advantage is low latency. Some services will be able to greatly reduce processing delay by directly transmitting raw information in a wavelength band, which is currently delayed by such factors as IP packet waiting and data compression.

(3) Digital twin computing

Today, digital twin technology is attracting attention across a wide range of industrial fields. A digital twin is an accurate cyberspace representation of a real-world object such as a production machine in a factory, aircraft engine or automobile (Figure 8(a)). However, most of the current approaches are focused on creating a single umwelt in cyberspace, which is simply a copy of the real world. But, as can be seen in the examples of the umwelts of bees and mantis shrimps, our world is full of various kinds of information that humans cannot perceive.

Digital twin computing [17] is a new computing paradigm



(a) Conventional Digital Twin

(b) Digital Twin Computing

Figure 8: Digital Twin Computing.

that will allow us to perform operations and interactions between many digital twins in cyberspace such as cars, robots, and people (Figure 8(b)). For example, digital twin computing can simulate the future of a digitized city in cyberspace. This means the future can be changed by simulating the development and decline of a city and feeding back the results to actual city planning. Thanks to technical innovation, humankind will be able to go beyond its own umwelt and access all types of available information. We believe digital twin computing will capture everything in the real world and by re-presenting it in cyberspace it will enable the creation of new values, previously inconceivable to humans.

Although today's cloud and other computing platforms are underpowered to perform complex calculations between many of the individually focused digital twins, the lowpower-consumption, high-capacity information processing enabled by all-photonics network will be the answer.

(4) Cognitive Foundation

The Cognitive Foundation is a concept that flexibly controls and harmonizes all ICT resources. Its key points are self-evolution and optimization. There have been many attempts to detect failures based on logs issued by telecommunications infrastructure devices and to deal with them autonomously using AI. The time has come to move well beyond that approach. The Cognitive Foundation incorporates a wide variety of information that cannot be monitored by the network, such as weather prediction information about the strength and path of an approaching typhoon and information about planned events. It also incorporates information about the various umwelts referred to earlier in this paper. Based on the collected information, the system will optimize the network autonomously. For example, it will plan and execute measures against a disaster before it occurs. It will make predictions and evolve itself accordingly, in other words, self-evolving service lifecycle management.

(5) IOWN Global Forum

Realization of IOWN requires not only telecommunications and computer technologies, but also the insight of researchers and experts from various fields, such as the social sciences and the humanities. The IOWN Global Forum was established in 2020. This is a new industrial forum to facilitate cooperation among global partners with the objective of driving forward research and development for IOWN [18]. The forum accelerates the adoption of a new communication infrastructure to meet our future data and computing requirements through the development of new technologies, frameworks, specifications and reference designs in a number of areas.

4 FUTURE SERVICES CONCEPTS

This section describes the concepts of future services that blend naturally into people's lives as a result of evolving technologies described in the previous section.



Figure 9: Point of Atmosphere



Figure 10: Heart-warming elderly care.

4.1 Point of Atmosphere

"Point of Atmosphere" is a concept for future natural devices (Figure 9). Imagine someone living in a world where there are no terminals anymore. A Point of Atmosphere is a connecting point for all devices to the environment itself. For example, when this person wakes up, various things in the immediate environment can be used to ascertain the person's state of health, traffic and other conditions, and then naturally give this person relevant advice. When this person is about to go out, the bag may appear to be bringing attention to itself by blinking using new projection technology, thereby letting the person know that something has been forgotten. As soon as the closet is opened, the down jacket hanging there advises the person to wear it because it's cold outside. Or when going into a wine bar, the wine bottle itself suggests that this is the wine to be tried. There will be a world where various things around people watch and support their lives. That's natural.

4.2 Heart-warming Elderly Care

The final example of a service concept is one that supports dementia sufferers so that they are able to lead rewarding lives [19]. Imagine a mother who suffers from dementia and she has forgotten the name of her son. To help jog her memory, her digital twin may display videos of the son's childhood (Figure 10). Moreover, the digital twin may sense her emotions and thoughts and find a piece of music that can help awaken her memories. We believe this helps her to recapture memories of her son, and bring about a future in which everyone can be connected to his or her memories, be connected to other people, and enjoy a contented life supported by mental and physical health.

5 CONCLUSION

This paper described a world where many people from diverse backgrounds and cultures can naturally benefit from technology. Although ICT has evolved as the frontier of "digital," the next step in digital's evolution is towards "natural." Many technologies such as AI, ultra-realistic communication, networks, and information processing are evolving towards creating natural value. We showed how these technologies could continue to evolve in the future and how they will be able to assimilate and respond to various values, how they can convey excitement, and how they can support people's lives in a natural way. This paper also showed the concepts for future services that blend naturally into people's lives as a result of evolving technologies. We hope such a smart world will become a reality in the near future.

REFERENCES

- Government of Japan, The 5th Science and Technology Basic Plan, https://www8.cao.go.jp/cstp/english/ basic/5thbasicplan.pdf (2016).
- [2] H. Thoen, M. How, T. Chiou, and J. Marshall, "A Different Form of Color Vision in Mantis Shrimp", Science, Vol.343, Issue 6169, pp.411-413 (2014).
- [3] J. von Uexkül, J. D. O'Neil (trans.), "A Foray into the Worlds of Animals and Humans: with a Theory of Meaning", University of Minnesota Press (2010).
- [4] R. Higashinaka, K. Sakai, H. Sugiyama, et al., "Argumentative dialogue system based on argumentation structures", Proc. of the 21st Workshop on the Semantics and Pragmatics of Dialogue (SemDial2017), pp.154-155 (2017).
- [5] K. Sakai, R. Higashinaka, Y. Yoshikawa, et al., "Hierarchical Argumentation Structure for Persuasive Argumentative Dialogue Generation", IEICE Trans. on Information and Systems, Vol.E103-D, No.02, pp.424-434 (2020).
- [6] H. Kameoka, K. Tanaka, A. Puche, et al., "Crossmodal Voice Conversion", arXiv:1904.04540 [cs.SD] (2019).
- [7] H. Nagata, D. Mikami, H. Miyashita, et al., "Virtual Reality Technologies in Telecommunication Services", Journal of Information Processing, Vol.25, pp.142-152 (2017).
- [8] A. Akutsu, K. Minami, and K. Hidaka, "Kirari! Ultrarealistic Communication Technology: Beyond 2020", NTT Technical Review, Vol.16, No.12, pp.6-11 (2018).
- [9] Press release issued by NTT, NTT Delivers Successful Ultra Reality Viewing of Live MLB Postseason Game, https://www.ntt.co.jp/news2019/1910e/191008a.html (2019).
- [10] T. Sato, K. Namba, M. Ono, et al., "Surround Video Stitching and Synchronous Transmission Technology for Immersive Live Broadcasting of Entire Sports

Venues", NTT Technical Review, Vol.15, No.12 (2017).

- [11] S. Usui, K. Kimura, S. Kinoshita, and K. Minami, "A Kabuki and Information Communication Technology Collaboration: Kabuki × ICT", NTT Technical Review, Vol.17, No.1, pp.20-27 (2019).
- [12] Cisco Visual Networking Index: Forecast and Trends, 2017-2022 White Paper, https://www.cisco.com/c/en/ us/solutions/collateral/service-provider/visualnetworking-index-vni/white-paper-c11-741490.html (2019).
- [13] M. Ii and K. Kawazoe, "The IOWN Innovative Optical and Wireless Network", NTT Publishing (2020).
- [14] Press release issued by NTT, NTT, Intel and Sony Establish New Global Forum Dedicated to Realizing the Communications of the Future, https://www.ntt.co. jp/news2019/1910e/191031a.html (2019).
- [15] QSFP-DD MSA, QSFP-DD Hardware Specification for QSFP DOUBLE DENSITY 8X PLUGGABLE TRANSCEIVER Rev 5.0, July 9, 2019, http://www. qsfp-dd.com/wp-content/uploads/2019/07/QSFP-DD-Hardware-rev5p0.pdf
- [16] K. Nozaki, S. Matsuo, T. Fujii, et al., "Femtofarad optoelectronic integration demonstrating energy-saving signal conversion and nonlinear functions", Nature Photonics, Vol. 13, pp. 454-459 (2019).
- [17] NTT, Digital Twin Computing White Paper, Ver. 2.0.0, https://www.ntt.co.jp/svlab/e/DTC/DTC_Whitepaper_

en_2_0_0.pdf

- [18] IOWN Global Forum, https://iowngf.org/
- [19] NTT, Natural ICT supports aging society (concept video), https://youtu.be/rHzgX_kgtHo (2019).

(Received March 16, 2020)



Susumu Yamamoto is a senior research engineer of NTT Media Intelligence Laboratories. He received his B.E. degree in electrical engineering and M.E. degree in computer science from Keio Univ., Kanagawa, in 1997 and 1999, respectively. After his joining NTT in 1999, he has

been engaged in image and video processing such as digital watermarking and virtual reality system. He is a member of IPSJ and IIEEJ.



Akira Nakayama is a manager of R&D vision group, Research and Development of NTT holdings. He received his M.E. and Ph.D. degrees in computer science from Nara Institute of Science and Technology in 1999 and 2007, respectively. After his joining NTT in 1999, he has

been engaged in robotics, CSCW, recommendation and people flow analysis. He is a member of IPSJ, ACM and RSJ.



Katsuhiko Kawazoe is a senior vice president and the head of Research and Development of NTT holdings. He received his B.E. and M.E. degrees in engineering from Waseda Univ., Tokyo, in 1985 and 1987, respectively. He also received his Ph. D. in Information engineering

from Kyoto Univ., Kyoto in 2009. After his joining NTT in 1987, he had been engaged in radio and satellite communication systems, the personal handyphone system (PHS) and broadcast-broadband convergence services. He became the head of NTT Service Evolution Laboratories in 2014, and head of NTT Service Innovation Laboratory Group in 2016. He assumed his present position in June 2018. He had served as President of ITE from 2018 to 2019.