Human Information Science: Opening up Communication Possibilities for the Future

Tatsuya Hirahara† *and Hiroshi Murase*

Abstract

NTT Communication Science Laboratories conducts scientific research aimed at understanding human information-processing mechanisms and technological research aimed at realizing human-friendly interfaces in computer environments. The ultimate goal of our work is to enrich communications among people as well as between people and computers. This special feature presents the leading edge of our research activities.

1. Introduction

Many obstacles must be overcome before NTT can make its HIKARI vision a reality. One of the barriers is the hardware required in order to create platforms that will enable anyone to access high-speed networks regardless of time or location. In keeping with the developments in computer and network technologies, the capacity to accumulate, process, and transmit information has increased dramatically compared with 20 years ago. However, these advancements fall short of what is needed for HIKARI. There are strong expectations that in the near future, as optical technologies become more refined, we will see optical memory on the scale of several terabytes (tera: 10^{12}) and optical networks with speeds reaching several exabits per second (exa: 10^{18}), which would meet the requirements. Another obstacle is the software required to make full use of these high-speed networks and the computers connected to them. When we talk about software, we are not referring only to computer programs and content. Software also includes the man-machine interface technologies that make telecommunications equipment easier to use and that are necessary to effectively convey "information." This "information," in turn, refers not only to that expressed in letters and numbers, but also sounds and images, as well as touch, smell, and taste in other words, all of multi-modal stimuli received by the five senses.

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: hirahara@idea.brl.ntt.co.jp

2. Human information science

Throughout the long history of evolution, human beings have learned strategies for extracting and using information from the real world, which is subject to physical laws. These strategies have been imprinted into our brain. In recent years, however, the world we come in contact with has been expanding at a rapid pace. As computers become more powerful, it is becoming much easier to modify or even create multi-modal information. With broadband networks being expanded on a global scale, it is becoming much easier to transmit and receive multi-modal information regardless of time and space. As a result, we have more opportunities to be exposed to various types of information and the brain has to deal with them. We face a set of circumstances that our ancestors never had to tackle.

First, increasing the complexity of high-tech communication tools could make it difficult for us to grasp their internal model. The internal model is the cause-and-effect relationship governing operations of equipment, tools, or the mind. When we cannot understand something's internal model, we hesitate to use it: we do not like to use things unless we know how they work or we fear unexpected surprises arising from incorrect use. So we have to learn how to use a new high-tech tool when we get one. The use of high-tech tools, however, is often unnatural because they are designed for the convenience of the tools not users. The process of learning to use new tools relies on the plasticity of our brain, but it has some limitations. It is thus important to design high-tech tools so that they fit our internal model smoothly.

Second, the increasing frequency of encounters with highly exaggerated "information" can potentially have a very strong effect on people and could even be life threatening. The "Pokemon incident" is one such example. On December 16, 1997, over 600 people across Japan (mostly children) suffered from convulsive seizures and other symptoms while watching the animated TV program "Pocket Monsters." A number of people were hospitalized. The cause is believed to be photosensitivity to the strong flashing lights used for the video effects in the animation. The influence of "information" on our brain may be only temporary or it may be long term. A temporary influence is called adaptation and a long-term influence is called learning. Thus, it is becoming necessary for us to make responsible choices when selecting information from the flood of information available. At the same time, the parties that convey and distribute information will also be required to act responsibly, evaluating in advance its effects to ensure that no unforeseen incidents occur.

Third, the increasing amount of time spent in virtual worlds could bring about social disruption. The virtual worlds created on computers encompass many different levels, from simplified imitations of the real world to purely imaginary artificial worlds. Among the latter are the worlds of on-line network games, in which more than one hundred thousand people are participating at any given time. For some people who spend many hours in such virtual worlds, the boundary between the real world and virtual ones becomes indistinct. As a consequence, we are seeing social withdrawal, increasing dependency, and decreasing communication skills. A disregard for the value of "life" comes from the expectation of being able to press the "reset" button anytime. There is also a tendency for some to bring the stakes in the virtual world into the real world. It is thus becoming more important to contemplate the compatibility of these virtual worlds with people and societies in the real world.

The only way to overcome these software technology barriers is to master ways to get along with computers and ways to get along with information. Our relationship with computers is still comparatively new, going back just a few decades. And our relationship with the virtual worlds created through computers and networks has only just begun. In order to reduce constraints and create more attractive communication environments, which should vitalize social and economic activities, it is important to gain a comprehensive understanding of human beings, the creators and the recipients of information, and explore ways of symbiosis with computers.

NTT Communication Science Laboratories performs scientific research aimed at clarifying the mechanisms of human information processing and technological research aimed at using engineering approaches to develop communication methods that are adapted to the unique characteristics of human information processing. In this special feature, we will introduce the leading edge of our scientific and technological research in this area.

3. Scientific research

Among the five senses, by which the brain receives information from the external world, the sensory systems for sight and sound (*i.e.*, vision and hearing) are highly developed. Normally, we live day to day without being conscious of the system's operations. However, if we lose these functions, or receive only limited information due to restrictions on the communication equipment, or try to make a computer mimic human functions, we are reminded of their complexity and their capability to solve very difficult problems instantaneously. In "Human Visual Mechanism" and "Human Auditory Mechanisms," we show some results that have deepened our understanding of how the visual and auditory systems work. These results also provide us with useful insights that will contribute to the development of new technologies that will expand our senses.

We use speech every day without devoting much thought to the speech production process. How do we generate speech sounds by moving articulators like the glottis, the lips, and tongue? Articulatory motions are not only physical motions of vocal organs but are also movements peculiar to human beings, which generate acoustic signals involving linguistic information processed by their auditory system. A better understanding of speech production mechanisms will allow us to create a sophisticated speech synthesizer or a talking robot that can produce various and more natural sounding voices. In "Human Speech Production Mechanisms", we review speech production mechanisms as revealed by our decades of research, then present an outline of a speech production model that enables a computer to mimic human voices by simulating the movement of vocal organs.

Motions of the mouth, eyes, hands, and body are essential to human communication, since they are closely related to perception. In “Human Motor Control Mechanism”, we outline the information processing mechanisms required to generate motions and introduce a computational model of arm motion control, and our latest findings concerning interactions with the environment during arm motion and control mechanisms for compensatory articulation during speech production.

4. Technological research

NTT Communication Science Laboratories has been conducting extensive research aimed at reproducing the human functions of “seeing,” “hearing,” and “speaking” on computers. Here, we introduce our latest research related to smooth speech dialog, easy-to-use gaze measurement devices, and timbre/music synthesis technologies.

It might appear at first glance that only speech recognition and speech synthesis are required to achieve speech dialog between humans and computers, but these alone offer only a minimal level of usability. To achieve smoother, more natural dialogue, the system must incorporate an interruption function that enables the user to ask a different question while the computer is still speaking as well as a dialog control function that allows the computer to quickly understand what the speaker wants. With these functions, a computer dialogue system can respond flexibly in short questions. In “A Mechanism for Smooth Conversation,” we introduce sophisticated technologies that enable a computer to communicate with people smoothly.

When executing speech dialog, more naturalness can be achieved by using non-verbal information such as gestures, which is separate from language information. We know that among the types of non-verbal information, eye gaze conveys important information for guessing what the speaker is thinking. In “Gaze Tracking System for Gaze-Based Human-Computer Interaction,” we introduce a gaze detection system called “Eye Camera” that can measure gaze quickly and accurately without the need for any special equipment worn on the face.

In addition to language and other logical information, information related to emotion plays an important role in enriching communication. It is thus important to create content that is capable of generating an emotional response. In “Timbre Synthesis Technology that Serves Communication,” we present research on the synthesis and editing of timbre, music, and related content.

5. Conclusion

To support safe and pleasant information distribution services in the 21st century, we need technologies that help us handle information and technologies that handle information safely and properly on computers and networks. Developing such technologies requires a deeper and comprehensive understanding of how the human brain processes information. Human information science research will open up wondrous communication possibilities, in which the dream of abundant communication between people as well as between people and computers comes true.

Tatsuya Hirahara
Executive Manager, Human and Information Science Laboratory, NTT Communication Science Laboratories.
He received the B. Eng., M. Eng., and Dr. Eng. degrees in electronic engineering from Hokkaido University, in 1978, 1980, and 1983, respectively. In 1983, he joined the Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT), Yokosuka, Japan. In 1986, he temporary transferred to ATR Auditory and Visual Perception Laboratories, Kyoto, and in 1991, he returned to NTT Basic Research Laboratories, Musashino, Japan. In 1997, he temporary transferred to the Corporate Planning Division of ATR, Kyoto, then in 2000, he moved to NTT Communication Science Laboratories, Atsugi. He received the NTT President’s Award in 1997 and the Awaya Kiyoshi award from the Acoustical Society of Japan in 1988. Dr. Hirahara is a member of the Acoustical Society of Japan, the Institute of Electronics, Information and Communication Engineers, the Virtual Reality Society of Japan and the Acoustical Society of America.

Hiroshi Murase
Former Executive Manager, Media Information Laboratory, NTT Communication Science Laboratories.
He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Nagoya University, in 1978, 1980, and 1987, respectively. In 1980, he joined the Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT), Tokyo, Japan. From 1992 to 1993 he was a visiting research scientist at Columbia University, New York. He was awarded the Telecom System Award in 1992, the IEEE CVPR Best Paper Award in 1994, the IEEE ICRA Best Video Award in 1996, the Takayanagi Memorial Award in 2001, and the IEICEJ Achievement Award in 2002. Dr. Murase is a member of the IEEE, the Institute of Electronics, Information and Communication Engineers, and the Information Processing Society of Japan.