Revolutionizing Mobile Communications through New Interface Technologies

Hirotaka Nakano, Toshiaki Sugimura†, Shoji Kurakake, and Atsushi Takeshita

Abstract

NTT DoCoMo Multimedia Laboratories are researching user interfaces that will enable you to use devices, objects, and information naturally and with ease as if they were part of your body. This article describes associated technical issues and introduces the concept of Human Centered Communication and Computing (HC^3) based on the technologies described.

1. What is HC^3?

Reflecting its corporate philosophy to "create new communication services that will revolutionize the way we communicate and establish a new global communications culture," NTT DoCoMo proposes the concept of Human Centered Communication and Computing (HC^3). The company is expending much research effort toward practical implementation of this concept.

The aim of HC^3 is to enrich communication and enhance human capabilities through interfaces that can convey a user's ideas, sensations, sensitivity, and intent. These interfaces are not about operating equipment with clear intent of accomplishing some task. They are rather technologies that will enable people to enjoy the benefits of information technology in a free and natural way, that will facilitate body-oriented communication, and that will represent and support people both physically and intellectually.

Specifically, HC^3 means creating new information-communication technologies taking into account the feasibility and means of achieving a wide variety of requirements. These include the abilities to "use something in the way one wishes," "share information and feelings as desired," "support and complement a user by inferring his or her thoughts or context," and "use things with ease in a natural, human-oriented manner."

This introductory article surveys the needs of society and the role of technology, presents a future image of communications, and describes issues and specific research themes that must be addressed and pursued to satisfy technical requirements. The other related articles in this issue describe specific research activities and their current state of progress.

2. Needs of society and role of technology

The 21st century has come to be called the "age of people", the "age of environmental management", and the "information age" [1]. Here, we discuss the needs of this diversified age in terms of what future mobile communications should aim for and describe technical requirements and the role of technology to this end.

In the "age of people", a human-centered system is essential if people are to live a truly human life, and a communication and computing system that is especially oriented to human needs should play an extremely important role here.

In this regard, there is much anticipation for true multimode communication that will enable people to reach even higher levels of mutual understanding and empathy. In addition to conventional forms of communication such as text, speech, and video, multimode communication will be able to convey information to all five senses and to exhibit embodiment char-

† NTT DoCoMo Multimedia Laboratories
Yokosuka-shi, 239-8536 Japan
E-mail: sugi@mml.yrp.nttdocomo.co.jp

acteristics (by using a physical entity to convey gestures and facial expressions more expressively).

It is also important here to provide high-reality forms of communication to make a deep impression on users and hold their attention and to provide content-based services that are both emotionally and physically impressive. Interfaces are also of great importance. For example, there is a need for an interface that can perform operations and give instructions by inputting and conveying what the user is thinking about or what the user wants to do, i.e., user intent, without requiring the user to learn how to operate it. The need is also felt for a user interface that can manage devices naturally and at will just as if they were part of one's body through simple operations and instructions using ordinary physical actions.

In addition, to enhance a person's information-processing ability and power of movement and to enrich that person's activities, there is a need for an environment that can support and complement that person in accordance with surrounding people and objects. Furthermore, in support of the aging society, there is a need for technologies that can support and complement internal organs, motor functions, and the like to preserve a healthy mind and body and promote a sense of fulfillment and satisfaction.

The "age of environmental management" demands technologies that can ease the load on the global environment. These include technologies for making environmental monitoring functions habitual and omnipresent through the five senses as a useful means of protecting and preserving the environment, and technologies for reducing traffic by providing alternatives to traveling such as videoconferencing, by optimizing traffic routes, etc.

Much attention is now being focused on a ubiquitous information-communications environment toward an "information age" in which the information world and real world that have traditionally been separate will merge to form a "super environment" [2]. Specifically, it will be possible to merge the information world within communication networks and the real world of human activities by making it easy to acquire real-world information and to use that information to communicate, and by replacing some endeavors in the real world with telecommunications. Here, within a world having a huge volume of information and a great number of devices in many forms, it is important that we have technologies that can support people or act for them in locating and handling optimal information and devices. This super environment will therefore be convenient and comfortable, but it will also raise concerns about invasion of privacy and increase in crime, for example. Under these circumstances, technologies for handling and supporting private information in public spaces, for detecting crime-related information and events, and for implementing encryption and security protocol will take on added importance.

3. Classification of technical issues by human characteristics

Table 1 lists technical requirements originating from social needs and technical issues that must be addressed to satisfy them. Horizontally, the table is divided into three sections according to human characteristics in relation to the input, processing, and output of information. These are "sensory organ system" in relation to information input; "brain and intelligent-processing system" in relation to information processing centered about the brain; and "motor and behavior system" in relation to the expression and presentation of information, emotions, and intent. Each of these sections is further divided into innate and acquired functions [3]. Vertically, the table is divided into social needs and technical requirements. Each table entry describes technical issues that must be addressed to satisfy the requirements for that row within the system for that column.

The sensory organ system consists of innate functions that govern the five senses, and assisting and extending these sensory functions are technical issues. The brain and intelligent-processing system consists of both innate and acquired functions, with the former consisting of memory, recognition, inference, and learning and the latter consisting of preferences, value judgments, comfort, and contentment. For both types of functions, researching and elucidating cognitive characteristics and extracting information-processing mechanisms as intelligent technologies are common issues with respect to the various technical requirements listed.

The motor and behavior system is where the results of internal human activities such as emotions and thoughts come to be manifested. This system involves a variety of technical issues. First, there is a need for explanations and models of the mechanisms behind information-based behavior and movement in humans. Next, there is a need for technologies for conveying and sharing emotions and intent through various forms of expressive behavior, gestures, and hand movements in accordance with human norms of communication. Finally, technologies are needed for

Table 1. Main technical

Needs of society → Technical requirements		Human characteristics: Input (sensory organ system)	
		Innate functions	
		Five senses: hearing, sight, touch, smell, taste	
Age of people	• Enable communication conducive to mutual understanding and empathy	• Raise magnitude of media quality → Convey ultra reality	• Input, convey, reproduce as new dimensions in media → Expand range of conveying and sharing sensations and sensitivities (communication and content based on the five senses)
	• Provide impressive and moving content		
	• Provide an easy-to-use human interface	• Input, convey, and reproduce directly what your body feels as sensation information → Direct five-sense interface • Input and reproduce by sensation-based devices and enable media conversion between senses → Indirect five-sense interface (by either an interface attached to your body or one attached to a remote robot) • Sense and convey internal physical functions and motor functions by fulltime-wear devices → Interface for monitoring your physical condition	
	• Enhance human capabilities (including health support)		
Age of environmental management	• Collect and use information useful for environmental management	• Achieve sensors (compact, low-power, zero-installation, cooperative operation) corresponding to the five senses and install, operate, and collect data according to objectives → Information platform using a ubiquitous sensor network (using sensors embedded in buildings, roads, vehicles, people, etc.)	
	• Reduce load on the environment		
Information age	• Improve information selectivity taking individual characteristics and context into account	• Sense and utilize information on your five senses as a means of extracting your individuality and as a means of determining your current context → Understanding individuality and context via the five senses • Sense and utilize five-sense information about your environment using wearable devices and those located nearby as a means of determining surrounding conditions → Understanding surrounding conditions via the five senses	
	• Improve information adaptability taking individual characteristics and context into account		
	• Improve environmental adaptability for enjoyable use of services		

issues for achieving HC[3].

Processing (brain and intelligent-processing system)		Output (motor and behavior system)	
Innate functions	Acquired functions	Innate functions	Acquired functions
Memory, recognition, inference, learning	Preferences, value judgments, comfort and contentment	Physical form/functions	Speaking, expressing intent, interactive behavior
• Improve translation functions → Overcome language barrier 〈Outside the scope of this research〉	• Search out agreeable conversation partners → Support and provide enhanced selectivity for individual use • Search out desirable content → Support and provide enhanced selectivity for individual use	• Create embodied expressions (gestures, hand movements, facial expressions, etc.) → Integrate emotions, intent, and atmosphere (form a collaborative space where context can be shared)	
• Deal and respond flexibly with searching/adjusting requirements and conditions mainly based on language and images (cope with a variety of expressions; adapt to insufficient information with appropriate questions, etc.) → Intelligent agents • Record one' s own behavior automatically, and select and use desired information at will → Intelligent memory support • Automatically record data obtained by monitoring behavior and physical conditions, detect changes, and issue alerts → Intelligent health monitoring	• Directly input, convey, and reproduce what one is thinking, what one wants to do (intent) without having to manipulate devices → Ultimate thought interface	• Satisfy "immediacy" in which functions can be used just as soon as one thinks about them; "wearability" in which devices are always worn on one's body in a comfortable manner, and "usability" in which input can be performed instantly without impeding the thought process. → Fulltime-wear Interface ready to use • Manipulate information by manipulating physical entities → Intuitive interface through real entities • Use ordinary physical actions to manipulate and instruct devices naturally and easily → Use part of the human body as an interface mechanism • Sense physical functions and motor functions and physically support and complement those functions → Physical function support • Project physical sensations and emotions and manipulate at will a remotely located robot sensitive to your five senses → Pseudo-teleportation (for security, nursing, remote operations, exploring undeveloped space, etc.)	
• Efficiently store large amounts of collected data → Ultralarge-scale, real-world database based on hierarchically distributed and centralized role-sharing management • Filter vast amounts of data to obtain required information and perform ultralarge-scale data mining to discover new relationships → Support and provide ultralarge-scale analysis of real-world information		• Use results of human behavioral prediction and natural phenomena prediction to select optimal and alternative means of human activity → Reduce and control energy consumption optimally by using information on a world-wide scale	
• Store five-sense information and behavioral information collected on the individual and learn and infer individual characteristics (preferences, comfort and contentment) → Support and perform learning and inference of individual characteristics • Infer current context from five-sense information and behavioral information collected on the individual → Support and perform inference of individual context • Filter out optimal information or process information adaptively for the individual based on learned individual characteristics and on five-sense information, behavioral information, and surrounding conditions currently being collected → Support and perform adaptive information selection taking individual characteristics and current context into account		• Control methods of expressing, uttering, and presenting information based on individual characteristics and context and current surrounding conditions → Support methods of expressing and presenting information taking individual characteristics and context into account • Construct an optimal service-usage environment for the individual by selecting and combining resources from the peripheral resource groups using individual characteristics and context and current surrounding conditions → Support and perform optimization of service-usage environments taking individual characteristics and context into account	

supporting and enhancing human information-processing ability and power of movement by enabling devices to be operated as an integral part of one's body and by enabling the body itself to be used as an interface.

4. Future image of HC^3

Figure 1 shows the future image of HC^3 after the above technical problems have been solved. Here, we have the birth of a "new-century man" (i-borg) with enhanced senses and no limit to the amount of knowledge that he can acquire. This is made possible through fulltime-wear information accessories and a 24-hour network connection that enables the new-century man to interact with a wide range of information-communication devices located throughout the world. In other words, HC^3 is communication and computing through interfaces that make it feel as if information, devices, and the network are all an integral part of one's body and that enable such resources to be used naturally and with ease. With HC^3, a person can operate fulltime-wear devices or nearby devices naturally just as if they were part of one's body without having to be particularly conscious of them. In this way, a person can convey and deeply share at will not only information but also sensations and emotions, and will be able to access the vast amount of information on the network as desired for use as personal knowledge. In addition, HC^3 will feature personal agents that become familiar with one's preferences, way of thinking, and other characteristics to support and even represent a person in some intellectual activities. Finally, HC^3 will enable a person to project his or her physical movements, sensations, emotions, and intent to a remotely located robot and interact with and manipulate that robot as one's alter ego. This will, in effect, extend one's physical presence to somewhere else in the world.

5. Main research themes

The main research themes in HC^3 are described below.

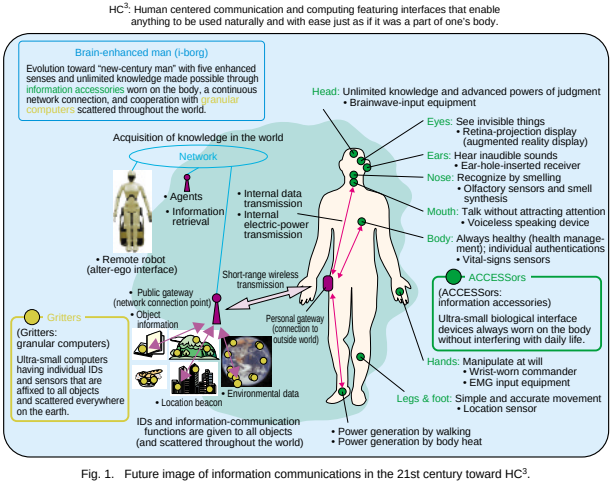

Fig. 1. Future image of information communications in the 21st century toward HC^3.

5.1 Mechanisms of human biological information

Interface technology that can directly input, transfer, and reproduce biological information should make it possible to use touch, smell, and other senses as new forms of media and bring about radical changes in the way people communicate. An essential element of this approach is research that attempts to clarify the mechanisms of human biological information. To this end, we have adopted the superconducting quantum interference device (SQUID) as a biological magnetic-measurement system. This system, which is capable of extremely high temporal and spatial resolutions, can make non-contact measurements of the small magnetic fields generated by electric currents that accompany brain, nerve, and muscle activities. The system can also be used as an interface. At present, research and development of a technique for analyzing the accuracy of SQUID output results are proceeding in parallel with research on the mechanisms behind human biological information.

5.2 Biological interface technology

Interfaces that measure, analyze, and use biological information are being researched with the aim of creating new media that can apply the five senses and operation-by-movement to communication. The first step in this research is to explain the basic mechanisms of the peripheral nervous system with the aim of developing application technologies. We are therefore targeting technologies for sensing, analyzing and using biological information related to muscle movement and technologies for sensing, analyzing, reproducing, and using surface sensations related to the sense of touch.

5.3 Fulltime-wear interface technology

In the society of the near future where high-speed wireless networks are commonplace, there will be no need to attach computers that store information onto each and every human body. The minimum requirement for equipment that a human must carry will instead be interfaces (devices) that convey the intent of the user to the network side and, conversely, that present the user with various kinds of information obtained from the network. The aim here is to achieve a “fulltime-wear interface” that satisfies three key conditions simultaneously. These are “immediacy”, or immediate access to desired information, “wearability”, or continuous wear to provide this immediacy without hindering daily life, and “usability”, or rapid input and output of information without hindering one’s flow of thought.

5.4 Alter-ego interface technology

The target here is a “real interface” that can convey one’s posture, actions, and behavior by projecting one’s self via a robot entity (alter ego), a function not possible with conventional audio and video media. Such an interface will play an important role in the evolution of the cellular telephone from “word-based communication to emotion-based communication”.

5.5 Ultra-real 3D speech/audio communications

This research theme focuses on technology that can generate a 3D sound field sensitive to space in a mobile environment using, for example, sensors that detect body movements. Included in this research is augmented reality in which a user can experience even more than what exists in reality. Another direction of this research is the generation of artificial 3D sound fields such as in three-way calling, where it would appear that the voice of one speaker is coming from the left and that of another from the right.

5.6 Ultra-real 3D video communications

This research aims to develop 3D presentation technology that can reduce the restrictions associated with wearable devices and enable a user to enjoy genuine 3D video in a mobile environment. This technology will feature both full and motion parallax and accommodation (focal adjustment).

5.7 Service selectivity in a ubiquitous environment

The technology targeted here will employ the direct use of objects (such as tools, appliances, vehicles, books, buildings, and places) in the environment that a person inhabits as triggers for finding related services without hindering that person’s activities in the real world. Also, with the aim of acquiring personal preferences and circumstances associated with the selection of an optimal service from among several candidate services found in this way, research here will also aim for technology that can continuously store information on all kinds of daily activities in a form conducive to automated processing so that behavioral patterns can be extracted.

5.8 Adaptive construction of service-usage environment

This research theme aims to develop technology that can determine what information-communication resource groups are available for a person moving in a ubiquitous environment given personal situations and policies as well as current conditions in the information-communications environment. Using the

resources so determined, this technology will be able to construct an optimal information-communications environment for the user at all times. A user will therefore be able to use services under optimal conditions without interruption no matter where he or she may be.

In a ubiquitous society providing many and diversified services, it is essential that desired services be found efficiently according to current conditions. Studies will therefore be performed on state-dependent and context-dependent service discovery and service-customizing methods to enhance this process even further.

6. HC^3 is our vision

We have described, from a human-centered point of view, the technical issues and research surrounding user interfaces that enable information, devices, and just about anything to be used freely and with ease just as if they were part of one's body. We have proposed a future image of human centered communication and computing (HC^3) based on these interfaces, and have introduced the issues that must be addressed to achieve HC^3 and approaches to solving them.

References

[1] "Information Communications Technology in 2015," NTT Publishing Co., Ltd., 2001 (in Japanese).
[2] K. Sakamura, "The Ubiquitous Computer Revolution: The World Standard of the Next Generation Society," Kadokawa Shoten Publishing, 2003 (in Japanese).
[3] "Next-generation Telecom: Current Possibilities," NTT Publishing Co., Ltd., 1992 (in Japanese).
[4] G. Salomon, "Distributed Cognitions," Cambridge University Press, 1993.

Hirotaka Nakano
Executive Manager, NTT DoCoMo Multimedia Laboratories.
He received the B.S, M.S, and Ph. D. degrees in electric and electronic engineering from the University of Tokyo, Tokyo in 1972, 1974 and 1977, respectively. Since joining NTT in 1977, he has been researching image/video retrieval systems (CAPTAIN, VOD, etc.) and multimedia communication systems. Currently, he is engaged in research on mobile multimedia. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Information and Television Engineers, and the Institute of Image Electronics Engineers of Japan.

Toshiaki Sugimura
Senior Executive Research Engineer, NTT DoCoMo Multimedia Laboratories.
He received the B.S. and M.S. degrees in computer science from Tokyo Institute of Technology, Tokyo in 1978 and 1980, respectively. In 1980, he joined Nippon Telegraph and Telephone Public Corporation (now NTT). He moved to NTT DoCoMo in 1999. He is engaged in research on natural language processing, intelligent processors, mobile multimedia, ubiquitous information-communications environments, and other subjects. He is a member of IEICE, the Information Processing Society of Japan (IPSJ), and the Association for Computing Machinery (ACM).

Shoji Kurakake
Executive Research Engineer, NTT DoCoMo Multimedia Laboratories.
He received the B.S. and M.S. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo in 1983 and 1985, respectively. Since joining NTT in 1985, he has been researching character recognition, image recognition, media handling, mobile agents, and ubiquitous services. He is a member of IEICE, IEEE, and ACM.

Atsushi Takeshita
Executive Research Engineer, NTT DoCoMo Multimedia Laboratories.
He received the B.S. and M.S. degrees in information science from Kyoto University, Kyoto in 1986 and 1988, respectively. He joined NTT in 1988 and is currently researching multimedia information searching and mobile internets. He is a member of ACM and IPSJ.