Ubiquitous Information Interfaces

By Keiji YAMADA,* Shuji SENDA,† Toshiyuki ASAHI,† Toshihiko HIROAKI,†
Hiroaki HATTORI* and Akitoshi OKUMURA*

ABSTRACT This paper describes ubiquitous information interface, which consists of an input function, an output function, and an interactive function. These functions have to be designed under the criteria such as real time, adaptability, and personalization in addition to the basic user interface design rules. As examples of ubiquitous information interface, it describes a camera-typing interface, a speech interface, a semantic web browser, a fast document image browser, and an automatic spoken language translator. These examples demonstrate that the ubiquitous information interfaces can overcome the problems which may occur in the ubiquitous society and make ubiquitous systems be easy to use and efficient.

KEYWORDS **Ubiquitous, Interface, Image processing, Internet, Browser, Speech processing, Translator**

1. INTRODUCTION

At the end of the 1980's, Mark Weiser proposed the concept of ubiquitous computing[1] and Norman put forward the idea of invisible computing[2]. They predicted that computers would be miniaturized and embedded in various objects, and that in future, computers would be used very differently from the way they were currently used.

Until now, computers have usually consisted of a box with a keyboard, mouse, and screen. To use a computer, the user goes to the computer and uses the keyboard (**Fig. 1**). With ubiquitous computing, however, users do not have to visit a computer to use it. They can access information services every time they need them wherever they are. This is not merely "anywhere and anytime" computing, but the ability to use information services naturally without any special effort.

As a step towards the ubiquitous information society, a number of computers can be connected to one another via a ubiquitous network that combines several network systems. These sorts of systems provide access to almost unlimited information. To utilize them efficiently, however, requires the impractical task of mastering numerous types of terminal devices with different kinds of access networks. Conventional GUIs (Graphical User Interfaces) are incapable of solving this problem. Instead, we need a new kind of user interface, which we call a ubiquitous information interface. This paper describes the concept of a ubiquitous information interface and gives some examples along with their advantages.

2. PROBLEMS OF UBIQUITOUS SYSTEMS

Ubiquitous systems include the following three types (see Fig. 1).

(1) Access Networks

Several computer systems may be integrated via a ubiquitous network, which is a seamless combination of numerous networks such as a Gigabit Ethernet, W-CDMA network, W-LAN, and UWB (Ultra Wide Band) network.

(2) Terminal Devices

Different types of terminal devices, including mobile phones, PDAs, POS (Point-Of-Sale) terminals,

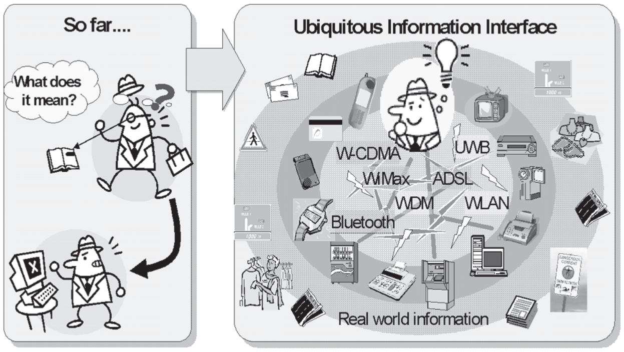


Fig. 1 From conventional interfaces to ubiquitous information interfaces.

*Media and Information Research Laboratories
†Internet Systems Research Laboratories

next-generation IC cards, and terminals attached to vending machines, are connected to one another and exchange information such as users' attributes and logs through a ubiquitous network.

(3) Content Presentation

Information may be presented in various multimedia formats such as text, images, video, speech, three-dimensional images, or some sort of physical force.

It might seem that these various systems provide users with access to information services whenever they want them. However, they do not offer users the benefits of a ubiquitous system; rather, they may actually prevent users from experiencing these benefits. Every time users want to access an Internet service or content, they have to select the optimal access network and terminal device depending on the circumstances. They therefore need to know how to use all sorts of different terminal devices so they can successfully use the one in front of them. This chaotic situation makes it much more difficult to use ubiquitous systems.

To address this problem and to provide a truly accessible ubiquitous system, we need a ubiquitous information interface that enables everybody to easily access information services using any kind of terminal device via any kind of network.

3. CONCEPT OF A UBIQUITOUS INFORMATION INTERFACE

An information interface is defined as an interface between the real world and a digital information source.

In the real world, information is available in the form of printed documents, notebooks, speech, signs, and objects. Information is also available about our surroundings, such as the temperature, humidity, and location of acquaintances.

This information is observed by sensors and is input as digital information. In addition, digital information is produced by computers. Its examples are computer graphics.

An information interface has three functions (see **Fig. 2**), input, output, and interactive functions. Each of these needs to be developed to overcome the problems described in the previous section.

(1) Input Function

The input function observes information on a real object to generate digital information with minimal effort on the part of the user. The digital information produced is not only stored in a computer, but also used as a hyperlink from the real-world object to other digital information related to that object. For example, if users want to know the meaning of a word printed on the page of a book, all they have to do is point their finger to it.

(2) Output Function

The output function presents digital information to the real world via display devices such as screens, printers, and speakers. This function not only outputs pieces of digital information to display devices, it also transforms them into a suitable form for the display device in front of the user. For example, when images and videos are presented on a small display device, essential information such as a headline is extracted. These parts are displayed at a higher resolution and other parts are shrunk and presented in a reduced form. Thus, users can acquire important information at a glance.

(3) Interactive Function

The interactive function enables dialogue between machines and users or facilitates communication between users. For example, an electronic commerce system discovers users' requirements through dialogue with them and then provides information about products that best meet their preferences. The system then guides them to complete a purchase according to their individual levels of computer literacy.

Another example is man-machine-man communication. An interactive function inputs a message from an individual and transforms it so that a partner can easily understand the message. If two people are using different terminal devices, transformation techniques, such as video transcoders, web-page-layout changes, or text-to-speech synthesis, can be used to

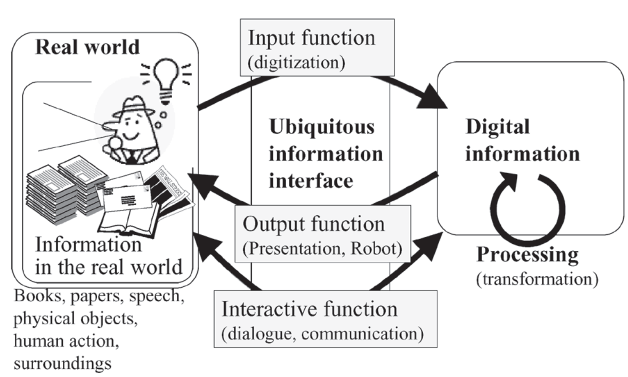


Fig. 2 Ubiquitous information interface.

change the media format. In addition, spoken Japanese text can be translated into other languages such as English and Chinese, facilitating international communication.

4. DESIGN OF A UBIQUITOUS INFORMATION INTERFACE

A ubiquitous information interface effectively utilizes both real-world and digital types of information, making ubiquitous systems easier to use. Each of these information types has advantages and disadvantages, as shown in **Table I**.

1) Intuitive usability: The most important feature of real-world information is that pieces of information are presented via familiar media such as newspapers and sign boards and the information is intuitively recognized at a glance. There is no need to learn how to use it.

2) Retrieval: Specific digital information can be found using information retrieval systems, which is a very helpful feature.

3) Amount of information: Unfortunately, not every piece of real-world information is digitized. Some effort is therefore required to digitize pieces of real world information into a useful form.

4) Transformation: Even if a piece of digital information is not represented in the required form, it can be transformed into that form.

A ubiquitous information interface is like a natural extension of current user interface techniques and should be designed according to Shneiderman's eight golden rules for the design of user interfaces[3]. As a consequence of the previous considerations, the following three criteria should be added to these rules (**Fig. 3**).

(1) Real Time

Input and interactive functions should be designed to respond quickly in situations where there is a limited amount of digital information and to enhance the intuitive usability of real-world information. Even if the input function makes an error when a speech recognition technique is being used, the operator will try again as long as the system produces a quick response. But if the response is both slow and incorrect, the operator will give up.

(2) Adaptability

The output function should be capable of transforming image content depending on the size of the display and network bandwidth. For improved usability, not only changes in image size, but also semantic transformation should be implemented based on image analysis, speech recognition, and automatic translation technologies.

(3) Personalization

To enable anybody to use ubiquitous information interfaces easily, they should be adaptable to individual skills and preferences. That is, an individual should be able to use different terminal devices in the same way.

Table I Advantages and disadvantages of real world information and digital information.

Type	Examples	Advantages	Disadvantages
Real world information	Books, newspapers, sign boards, speech, music, facial expressions, behavior	Intuitive usability	Difficult to retrieve Transformation not possible
Digital information	Digital documents, document images, pictures, music, Web pages	Retrievable Easily transformed	Limited amount of information

Shneiderman's eight golden rules

1 Strive for consistency.
2 Enable frequent users to use shortcuts.
3 Offer informative feedback.
4 Design dialog to yield closure.
5 Offer simple error handling.
6 Permit easy reversal of actions.
7 Support internal locus of control.
8 Reduce short-term memory load.

Additional criteria for ubiquitous information interfaces

9 Real time
10 Adaptability
11 Personalization

Fig. 3 Design criteria for ubiquitous information interfaces.

If ubiquitous information interfaces are designed according to the eleven criteria described above, they will overcome the chaotic state produced by the three various systems that exist at present and enable general use of ubiquitous systems. In the following section, five examples are described.

5. EXAMPLES OF UBIQUITOUS INFORMATION INTERFACES

5.1 Text Input Interface Using a Camera with a Mobile Phone

In Japan, more than 90% of the mobile phones shipped in the fourth quarter of the 2003 fiscal year included cameras. These cameras are mostly used for visual communication, but can also provide an input function for ubiquitous information interfaces.

We have developed a new technology called a 'camera-typing interface' that provides a way of inputting characters printed on magazines or business cards, such as URLs, using a low-resolution camera[4]. For example, if someone finds a URL within an interesting magazine article, they can immediately access the web page by shooting it using the camera on their mobile phone. Similarly, if they see an interesting sign board, they can access digital information relating to it by pushing a few buttons on their mobile phone.

Our technology has two main advantages in addition to the capability of recognizing low-quality characters with small computational resources; one is automatic concatenation of sequential shots, and the other is automatic error correction by re-shooting. For example, to input a long piece of text, the user shoots it bit by bit. The system concatenates the recognition results for the partial images by detecting overlaps between them (**Fig. 4**). If users discover recognition errors, they can retake the image around the misrecognized character. Then, the system changes the relevant recognition result to repair the error automatically (**Fig. 5**).

In tests, after automatic correction was performed once, the recognition rate increased significantly from 97.1% to 99.0%. After repeated corrections, the recognition rate increased to 99.6%.

5.2 Text Input Interface Based on Speech Recognition

A speech interface is an effective input function for ubiquitous information interfaces.

To demonstrate the effectiveness of speech interfaces, we developed an on-line manual retrieval system (**Fig. 6**) for cellular phones with Web-browsing capability[5]. The system recognizes user's naturally spoken queries via a telephone speech recognizer and searches an on-line manual with a retrieval module on a server. The results are then displayed on the screen via a Web interface. **Figure 7** also shows a traffic guide system with a speech interface based on VoIP (Voice over Internet Protocol), which transmits high-quality (wide band) speech, thus improving recognition. These systems are based on integrating speech interfaces and Web access technology and have been shown to be effective in evaluation experiments.

We plan to improve the usability and performance of the systems through field trials, and to explore technologies that will enable speech understanding and synthesis technologies to be applied to ubiquitous information interfaces.

5.3 Semantic Zoom Browser and JPEG2000 Image Browser for Adaptive Output Function

Currently, web pages are designed for specific terminal devices, mostly for PCs. When users access a

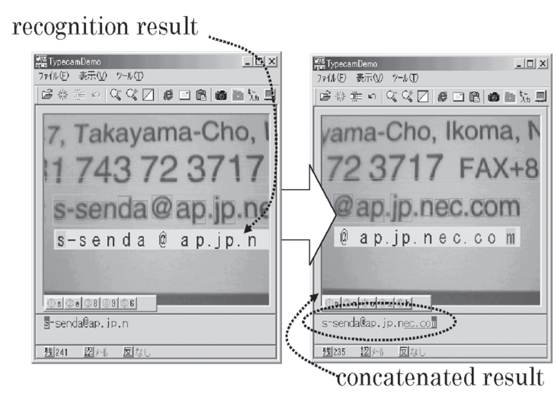


Fig. 4 Concatenation of recognition results.

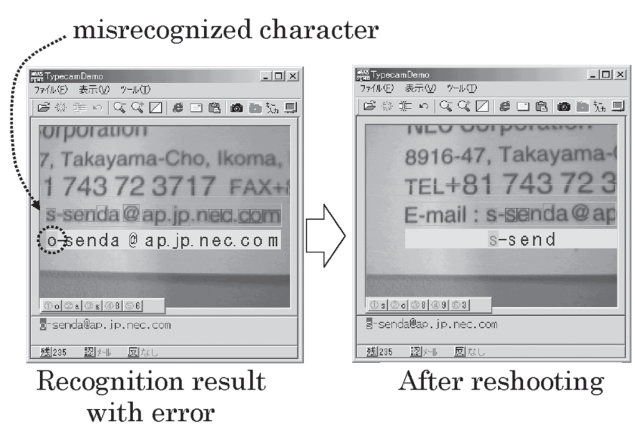


Fig. 5 Correcting misrecognized results by reshooting.

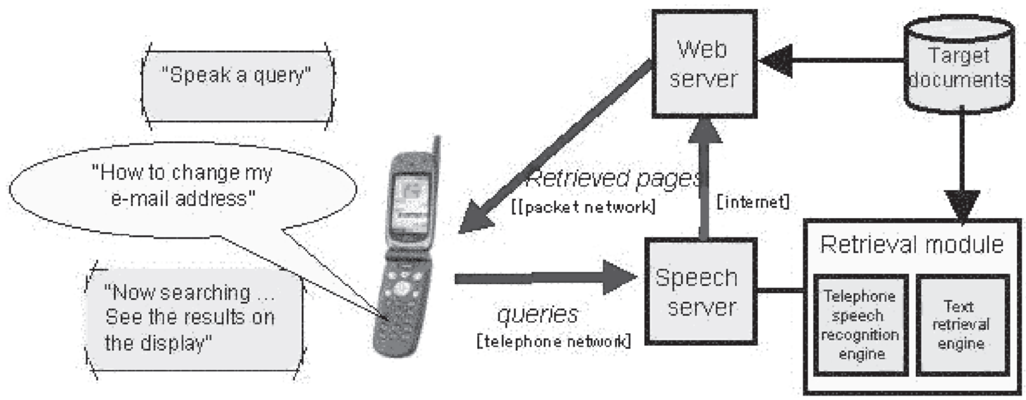


Fig. 6 A manual retrieval system.

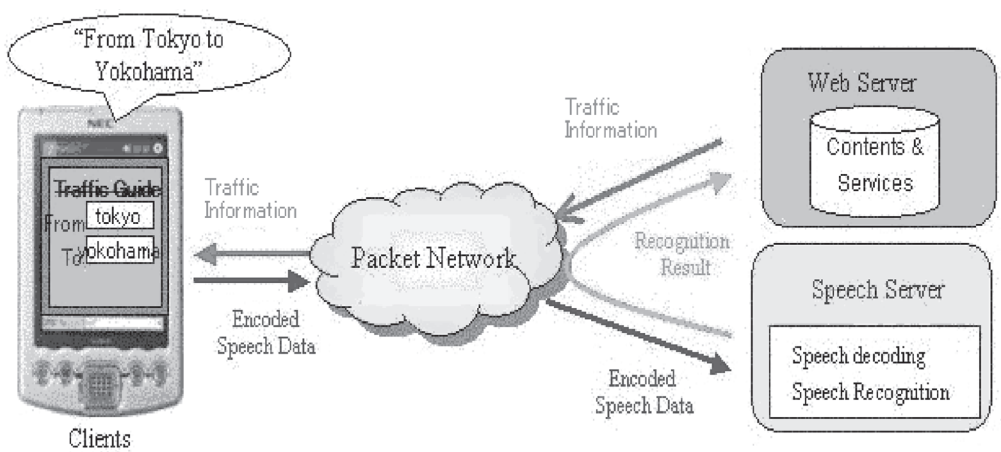


Fig. 7 Traffic guide service on VoIP.

web page via a mobile phone, they cannot actually view it. One reason for this is that the description language is unsuitable for the terminal. This problem can be solved by translating description languages such as HTML and C-HTML. Another major reason is that the assumed page size is different from the terminal display size. In this case, users cannot find the information they need because they can only see a very small area of the whole page. Even more annoying, although the display shows one small part of the page, it takes a long time to download a large page. We developed two technologies to address these problems: (1) a Semantic Zoom browser for Internet browsing using different types of terminals, and (2) a fast document image browser based on JPEG2000 image formatting.

(1) Semantic Zoom Browser

To solve the former problem, we developed a Semantic Zoom function that controls the degree of reduction/enlargement locally as well as the characteristics of pointer movement based on the semantic attributes of the content displayed on the screen[6]. This function is used in addition to the regular zoom function, which simply reduces and enlarges the images on the screen.

Using the regular zoom function, it would be possible to display the entire content of a page on a single screen, but the characters and other information would be impossible to read. Using the extended zoom function, however, the entire screen can be reduced, with titles and other important information maintained at a larger, more readable size (**Fig. 8**). The user can thus view the content as a whole, even on TVs and other screens with low resolution, and at the same time can understand the details by reading the headlines. This function uses XML descriptions of Web content layout information (inclusive relations, etc.) to judge the importance of titles and other display elements, enabling control of the degree of zoom for each individual display element.

In the past, if users wanted to view a segment of the content that fell outside of the main screen, they had to move the pointer to a small area on the screen and scroll to control the display position. This operation was difficult, however, when using a device with limited input capabilities, such as a remote control. With this newly developed function, the display position can be controlled by moving the pointer on the screen directly to the desired position. At the same time, by adjusting the pointer movement units in stages using "semantic units," for example, large, medium or small titles, the user can display the desired segment quickly and easily. This function uses the layout information described above to control the units of pointer movement.

(2) Fast Document Image Browser

A full-color A4 document image of 600 dpi (dots per inch) consists of about 100 megabytes of data. When users receive an image like this, it takes about ten seconds to see the page even using a 100Mbps network. This means that they cannot leaf through the pages of a magazine. We therefore developed a fast document image browser using the JPEG2000 image format, which provides hierarchical image descriptions using seven levels of resolution from low to high[7] (**Fig. 9**). Users at a terminal with a QVGA-size display do not receive an image with 2,000 × 3,000 dots; instead, the system sends the image to the terminal at the lowest resolution. Then it sends images at progressively higher resolutions. Users can go to the next page as soon as they want to skip the current page. If they want to magnify an image, the system will send that part at a higher resolution so that it fits the display. For users at terminals with different sizes of display, the system operates adaptively to the display size. This system enables users to view 10 full-color pages of a magazine per second via 384kbps communication channels.

We also developed image analysis technology that recognizes the content of document images. It extracts information about document layout, such as text lines, headlines, figures, and photos and also estimates the reading order of the text. The system sends images of text lines according to the reading order, enabling users to read them as quickly as possible. It can also send only photos to reduce the transmission time if users are looking for photos.

5.4 Automatic Speech Translation for Interactive Functions

We developed an automatic speech translation system for PDAs (Personal Digital Assistants) to assist oral communication between Japanese and English speakers in various situations while traveling (**Fig. 10**)[8]. This compact system requires only a small memory to immediately translate speech input in Japanese into English or vice versa and to read aloud the result.

As shown in **Fig. 11**, the system consists of speech recognition, translation and speech synthesis modules. The speech recognition module has a large-scale vocabulary of tens of thousands of words and

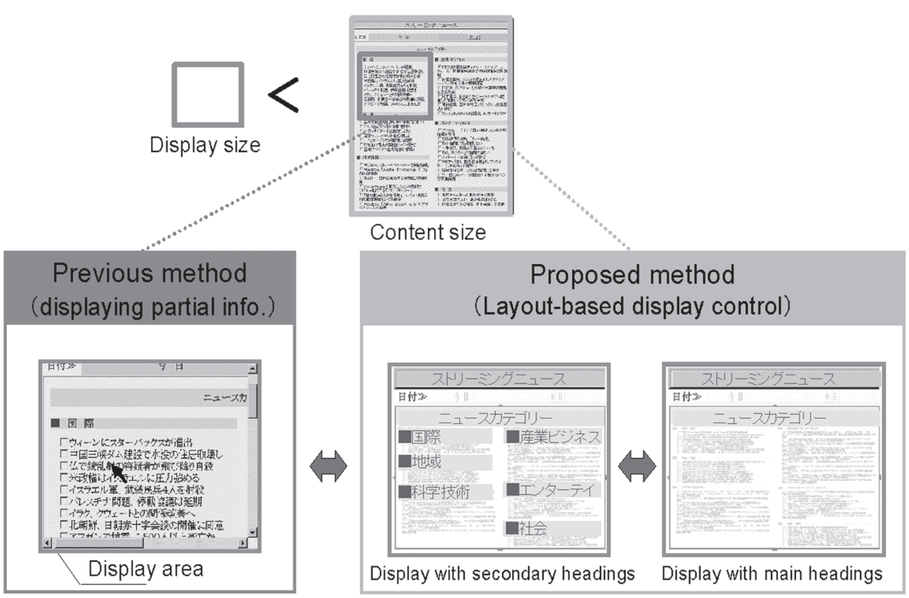
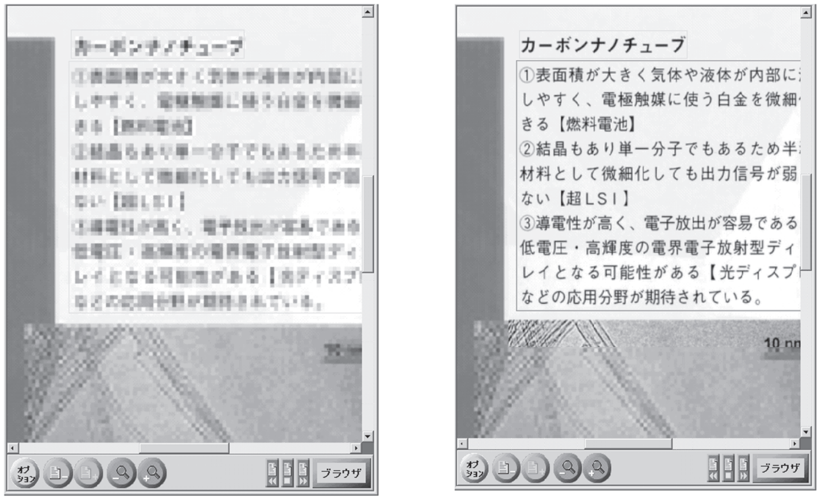


Fig. 8 Semantic zoom browser.



Low-resolution image　　High-resolution image

Fig. 9 Progressive transmission of document images.

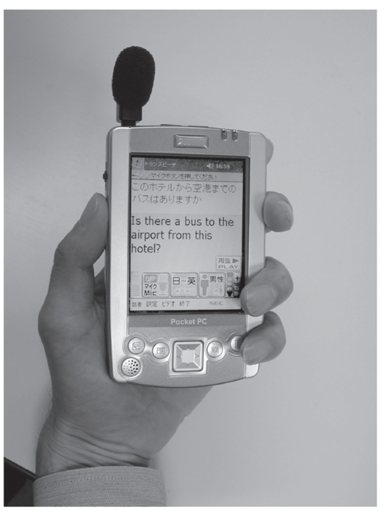


Fig. 10 Speech translation system on a PDA.

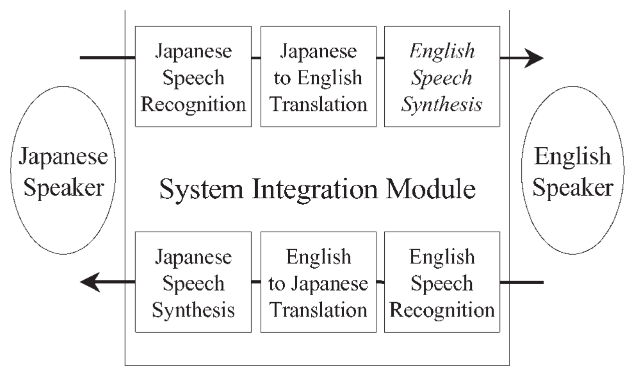


Fig. 11 Overview of speech translation system.

recognizes a variety of conversational expressions spoken by travelers. The translation module has a large-scale grammar dictionary and translates various travel-related expressions including colloquial idioms and polite expressions.

The performance of the new system has been aided by the development of a number of functions including a high-speed large-vocabulary continuous speech recognition engine that adjusts its operation in accordance with the processor and memory capacity used.

We are planning to continue developing automatic speech translation technology with improved performance and enhanced functions to provide users with the ideal translation support solution. We also plan to attempt to increase the number of languages that can be translated.

6. CONCLUSION

In this paper, we described the design of a ubiquitous information interface with properties of real time, adaptability, and personalization. To date, we have developed component technologies suitable for practical application. For input functions, we have developed interfaces using character and speech recognition, which provide hyperlinks between the real world and digital information. Users can easily access rich information relating to their interests. For output functions, adaptability is essential because different types of terminals have various-sized displays.

The system adjusts the content to suit the terminal. Media transcoding such as text-to-speech and text-to-animation will become more important. We also described an automatic interpreter which provides an interactive function. In future, we will develop a universal interpreter and a robot to engage in natural dialogue with people.

REFERENCES

[1] M. Weiser, "Some Computer Science Problems in Ubiquitous Computing," *Communications of the ACM*, July 1993.

[2] D. A. Norman, "Invisible Computer -Why good products can ail, the personal computer is so complex, and information appliances are the solution," The MIT Press, 1998.

[3] B. Shneiderman, "Designing the User Interface," Addison Wesley, Third Edition, 1998.

[4] S. Senda, et al., "Camera-Typing Interface for Ubiquitous Information Services," *Proc. of IEEE PerCom 2004*, pp.366-370, 2004.

[5] S. Ishikawa, et al., "Speech-Activated Text Retrieval System For Multimodal Cellular Phones," *Proc. of ICASSP 2004*, **1**, pp.453-456.

[6] Y. Tatsumi, H. Noda and T. Asahi, "Pointer Zooming: Pointing Method under Limited Input and Output Environment," *The 10th Workshop on Interactive Systems and Software (WISS2002)*, pp.67-72, 2002.

[7] S. Wang, et al., "Adaptive Data Transmission on Browsing of Scanned Documents Using JPEG2000," *Proc. of IEEE International Workshop on Knowledge Media Networking (KMN'02)*, 2002.

[8] R. Isotani, et al., "An Automatic Speech Translation System on PDAs for Travel Conversation," *Proc. ICMI'02*, pp.211-216, Oct. 2002.

Received July 26, 2004

* * * * * * * * * * * * * * *

Keiji YAMADA received his B.E., M.E. and Dr.E. degrees in Information Science from Kyoto University in 1982, 1984 and 1987, respectively. He joined NEC Corporation in 1987 and is currently the general Manager of Media and Information Research Laboratories.

Dr. Yamada is a member of the IEEE, the IEICE (Institute of Electronics, Information and Communication Engineer), and IPSJ (Information Processing Society of Japan). He received Sakai memorial Awards from IPSJ and some other awards.

Shuji SENDA received his B.E., M.E. and D.E. degrees in information science from Kyoto University in 1991, 1993 and 1996, respectively. He joined NEC Corporation in 1996, and is now Assistant Manager of Ubiquitous System Technology Group, Internet Systems Research Laboratories. He is engaged in the research and development of character recognition and its application.

Dr. Senda is a member of the Institute of Electronics, Information and Communication Engineers, and the Information Processing Society of Japan.

Toshiyuki ASAHI is a principal researcher at Internet Systems Research Laboratories, NEC Corporation. His research interests include usability engineering, information visualization and adaptive user interface. He received the M.E. degree from Osaka University in 1984 and the Ph.D. degree from Nara Institute of Science and Technology in 1997.

Dr. Asahi is a member of the Human Interface Society, Information Processing Society of Japan, and Usability Professionals Association.

Toshihiko HIROAKI received his M.S. degree in material engineering from Tsukuba University in 1990. He joined NEC Corporation in 1990, and is now Principal Researcher of the Internet Systems Research Laboratories. He is engaged in the research and development of ubiquitous multimedia information systems and user interface.

Mr. Hiroaki is a member of the IPSJ.

Hiroaki HATTORI received his B.E. and M.E. degrees from Hokkaido University, in 1983 and 1985, respectively. He joined NEC Corporation in 1985, and is now senior manager of Speech Interface Technology Group, Media and Information Research Laboratories. He received Dr. Eng. degree from Hokkaido University in 1994.

Dr. Hattori is a member of the Institute of Electronics, Information and Communications Engineers, and the Acoustical Society of Japan.

Akitoshi OKUMURA received his B.E. and M.E. degrees from Kyoto University, Kyoto, Japan, in 1984 and 1986, respectively. He joined NEC Corporation in 1986, and is currently researching human language technology at the Media and Information Research Laboratories as a Senior Manager. He received his Ph.D. from Department of Computer Science, Tokyo Institute of Technology in 1999. His current research interests include speech translation, translingual information management, information extraction and summarization.

* * * * * * * * * * * * * * *