The World of Mushrooms—A Transdisciplinary Approach to Human-Computer Interaction with Ambient Intelligence

Eisaku Maeda†*, Yasuhiro Minami, Masato Miyoshi, Minako Sawaki, Hiroshi Sawada, Atsushi Nakamura, Junji Yamato, Takeshi Yamada, and Ryuichiro Higashinaka*

Abstract

This paper provides an update to the January 2006 letter on "ambient intelligence", describing this year's achievements and the latest demonstration systems. The Ambient Intelligence Project (also known by the code name project Mushroom) aims to bridge the boundaries between technological fields and thus cover the entire field of communication science.

1. Introduction

NTT Communication Science Laboratories (NTT CS Labs.) initiated a research project with the theme of "ambient intelligence" (*kankyo chinou* in Japanese) in October 2004. It is being primarily undertaken by the Ambient Intelligence Research Group and the Intelligence Integration Open Laboratory of NTT CS Labs. *Kankyo chinou* has now been officially adopted as the main vision of the Labs. It has two main purposes: the first is to create new lifestyles through the research and development of communication science and technologies based on the concept of ambient intelligence, and the second is to break down the barriers surrounding existing research disciplines and to strategically develop the transdisciplinary area, as symbolized by the term "intelligence integration".

Last year we presented the key concepts of ambient intelligence and demonstrated a prototype at Open House 2005 of our laboratories [1]. A letter in the January 2006 issue of NTT Technical Review introduced these concepts and achievements to a wider, international audience [2]. We also discussed the lifestyles that could be made possible by ambient intelligence and suggested the specific issues to be tackled toward the achievement of ambient intelligence in a published paper [3].

In this paper, we review the background of the Ambient Intelligence Project in detail for the convenience of new readers who missed the previous letter [2] because the concepts are important but probably unfamiliar. We then go on to outline the project's achievements to date.

2. Ambient intelligence by information and communication science technology

Information science was born at the beginning of the 20th century and has grown remarkably. The communication science technologies of Japan have advanced significantly together with the Japanese economy, bringing significant benefits to our lives today. However, it is difficult to expect a similar rate of growth in the coming 50 years. Japan's low birthrate and longevity are changing the nation's demographic structure and causing a labor shortage. This will increase the burden of social welfare. Furthermore, global issues such as global warming, energy shortfalls, and waste disposal problems will only become more serious. In these circumstances, our society is searching for major qualitative changes.

In these social circumstances, the roles of communication science in enriching our lives have changed.

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
Email: maeda@cslab.kecl.ntt.co.jp

Mental stability and satisfaction rather than physical convenience have become new targets. While we are accustomed to depending on doctors and religion for relief in times of difficulty, we believe that the majority of people are looking for a more immediate and casual environment to ease some of their anxieties and enhance the vitality of daily life. Our long term goal is to apply communication science technologies to create such an environment.

In achieving the environment we envisage, one key word is communication because it is nothing less than the exchange of human kindness. Today, the Internet is increasingly popular as a medium for exchanging information between individuals, and new communities are created every day. The Internet enables an enormous amount of communication beyond the physical limits imposed in daily life. Yet no matter how much Internet communication is developed, people cannot dispel the sense of isolation unless they can receive the same mental support that they get from their family and friends, who share in their memories and knowledge. The popularity of communication over the Internet gives us a renewed recognition of the importance of direct communication in shared spaces without using networks.

For example, the casual remark or subtle smile common in direct communication can enrich our lives. We may not be conscious of this in most situations, but we often find out later that even small events are essential to our wellbeing. This support can also be achieved by communication with entities outside our everyday experience. We believe in harnessing the power of what we call “fairies and goblins”, as explained in the following section. We anticipate that communication science technologies will usher in the world of fairies and goblins within the next 50 years.

“Ambient intelligence” implies intelligence that surrounds us. The term ambient intelligence is also used for research being pursued in the EU and the USA [4]-[6]. Those studies aim to produce a design for a future information society by making use of computers and sensors embedded in the environment so as to create more natural and intelligent interfaces. Terms such as “ubiquitous” or “pervasive” come close to describing the concept. However, such R&D studies tend to focus on computers and sensors, and as such are geared toward hardware or devices.

On the other hand, our Ambient Intelligence Project is focused more on such things as hearing, vision, language, and knowledge, which are all related to human intelligence. The aim is to discover what kind of communication we really need and to achieve it. This is where our “ambient intelligence” differs from such concepts as ubiquitous computing or conventional ambient intelligence. We plan to achieve ambient intelligence by using information and communication science technologies. That is, we will use NTT’s areas of expertise—speech, sound, language, dialogue, vision, data retrieval, networking, and so on—to create a new style of ambient intelligence that places human intelligence and intellect at the forefront. That will lead to a proposal about a future lifestyle and clarify specific issues to be tackled towards it [3]. Furthermore, “ambient intelligence” may drive a robot brain or be embedded in the environment of ubiquitous computers and networks.

3. Fairies and goblins

“Ambient intelligence” is premised on the creation of communication partners that we want constantly at our side. To develop this concept, we started by considering appropriate fundamental features. For example, such entities must have different personalities and be able to evolve, grow, and express their feelings. They might seem to have the same level of intelligence as precocious children. They hide nearby and are unobtrusive, unassertive, and even a little mischievous. They respond when called and are visible to some people but not to others.

This world of “ambient intelligence” is rather similar to the world of fairies and goblins, which is familiar in both the East and the West. This makes “fairies and goblins” an adequate concept on which to base a common understanding of ambient intelligence. Ambient intelligence also provides a means for communication that speaks to the heart, as fairies and goblins used to do. Since information technologies have advanced so much, the time is ripe for thinking about new ways of using them in our daily lives as tools for refreshing the heart and mind. This approach will, we believe, make information technologies desirable to the general public.

We gave the name “mushrooms” to the fairies and goblins that are our idealized vision of ambient intelligence. These mushrooms live in a world called “Mush-Room”. Another important issue in terms of concept design is the evolution of the mushrooms. The process and the future perspective of mushroom evolution (**Fig. 1**) represent the progress of the R&D itself and a road map for the future of “ambient intelligence”. As mushrooms evolve, they will approach human intelligence or intellect. The world of “ambi-

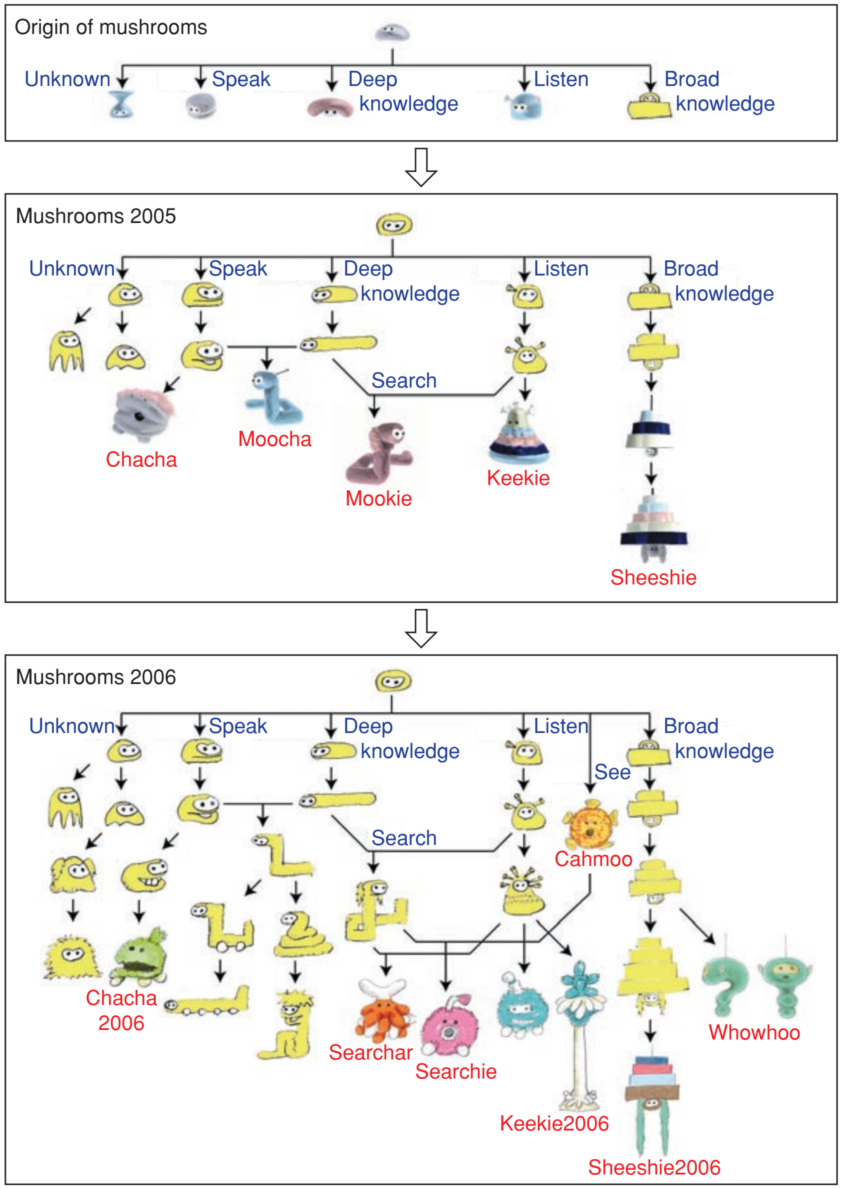

Fig. 1. Evolution of mushrooms.

ent intelligence" cannot be produced simply by advances in this or that technology. That is why we require intelligence integration that covers the whole field of communication science.

4. The world of mushrooms 2006

To achieve the ideal world of ambient intelligence, it is important to provide missing technologies as well as to enhance the current information processing technologies. This year's Open House (officially called Mirai-Soron 2006) offered three demonstrations (**Fig. 2**) in line with three fundamental concepts of ambient intelligence: watching over, keep looking, and hiding and other actions. They well illustrate our research achievements of this year [7]. These demonstration systems were implemented and are being maintained mainly by K. Shimomura, T. Matsuda, Y. Kashihara, and K. Maeda.

4.1 Demonstration A: mushrooms that watch, notice, listen, and answer questions

The keywords for this demonstration are watching over, noticing, and listening and answering questions.

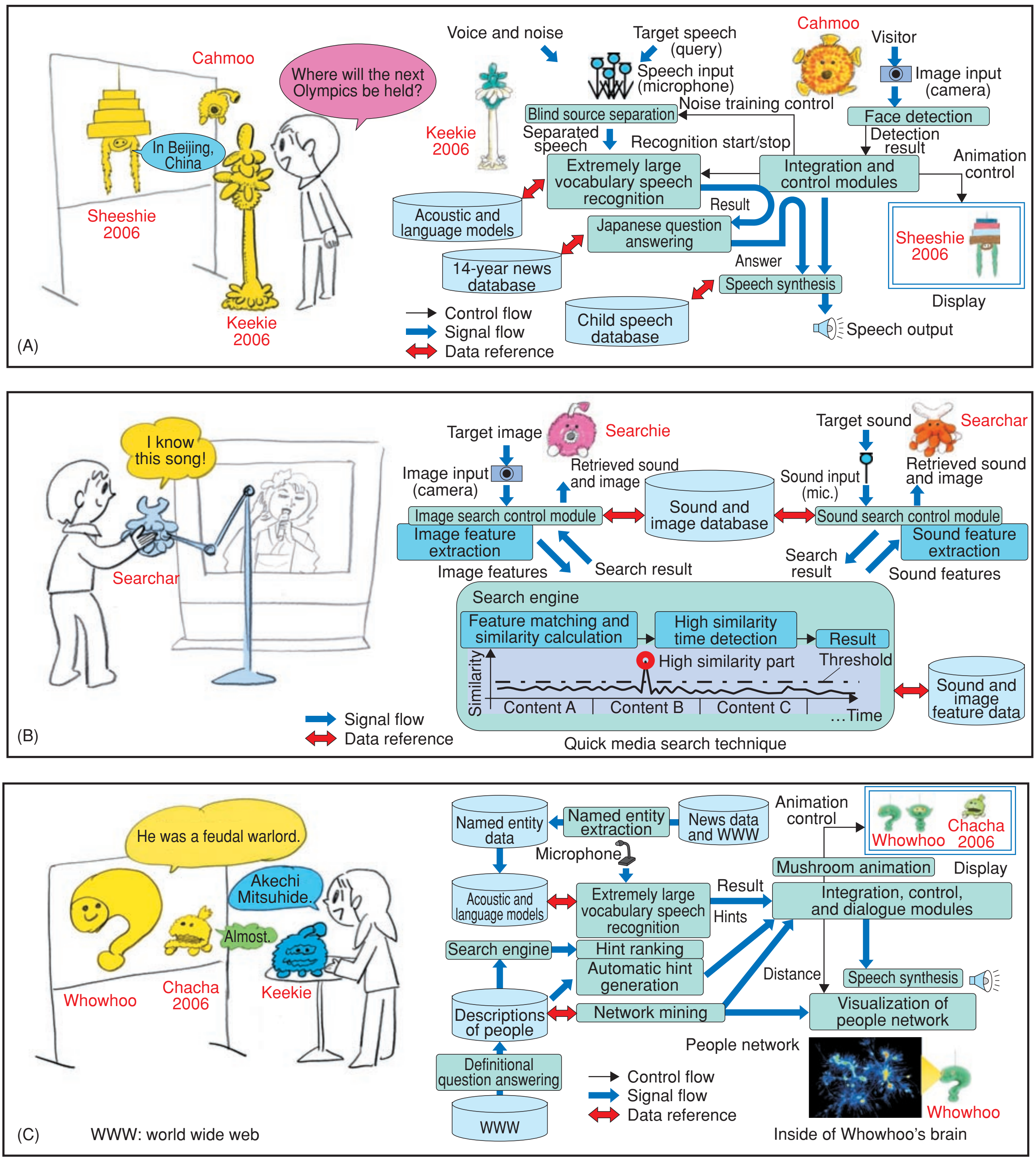

Fig. 2. Overviews of three demonstrations presented in Laboratory Open House 2006 (Mirai-Soron 2006).
The illustrations depict the demonstrations (left) and schematic block diagrams of their systems (right).

This demonstration was based on the following scenario. Mushrooms are watching and waiting to help people sharing a common space. A mushroom notices that person A would like to chat with person B (process I). It interacts with A through speech to confirm the situation (process II) and helps overcome the problems that person A faces in communicating with B.

This demonstration uses several mushrooms. For process I, the mushroom called Cahmoo is equipped with a video camera and face detection technology[*1]. When a person gets close to Cahmoo and faces its camera, Cahmoo assumes that the person wants to chat and tries to start a conversation by saying "Hello" or "What's up?" in Japanese. Simultaneously, the mushroom called Keekie, which uses four microphones as its ears, suppresses environmental noise and the voices coming from other speakers in order to emphasize the speech from the interlocutor. This preparation for the next stage consisting of question answering is based on a combination of blind audio source separation[*2] and face detection technologies.

For process II, Keekie recognizes the interlocutor's speech and transmits the query to the mushroom called Sheeshie. Sheeshie looks for answers in 14 years' worth of computerized newspaper articles (1991 to 2004) and finally answers the query by using speech synthesis technology. Such a speech-driven question answering system uses extremely large vocabulary speech recognition[*3] and question answering[*4] technology. Keekie can quickly and accurately recognize queries expressed in various ways. Moreover, while answering the query, Cahmoo keeps watching the interlocutor to guess whether he/she wants to pose another query. This new faculty

*1 **Face detection technology:** Cahmoo has a camera that captures images and sends them to a computer. The computer detects the position and size of a human face in the images. The face detection speed is accelerated by using two-step processing. First, through a comparison of averaged pixel values in neighboring areas, most candidate areas in the images that clearly do not correspond to a face are removed [5]. The remaining areas are then analyzed in detail by an assessment of their pixel values. In Demonstration A, the results of face detection were used to improve the performance of blind audio source separation[*2] and speech recognition[*3].

*2 **Blind audio source separation technology:** Keekie has an array of four microphones. By using only the signals captured by these microphones, it controls the array directivity to discriminate between target speech and undesired sounds, and then suppresses the undesired sounds. Such processing can be performed in a blind manner, i.e., without any *a priori* information about the source direction [9]. Keekie uses additional information provided by Cahmoo to achieve higher audio discrimination performance. When no face is detected, the captured signal is considered to consist of just undesired sounds and the directions of those sounds are estimated. If a face is detected, the probability of hearing some speech becomes high, and the target speech is extracted by removing the undesired sounds using the previously obtained information (their directions). This allows Keekie to hear the speech clearly even in a noisy environment such as an exhibition hall.

*3 **Extremely large vocabulary speech recognition technology:** Keekie has a SOLON speech recognizer [10], which features an algorithm based on fast on-the-fly composition for weighted finite-state transducer type advanced computational models. Due to its accuracy and efficiency, the algorithm enables, for the first time in the world, realtime speech recognition with a vocabulary of several million words [8]. This technology makes it possible to handle a very wide variety of expressions including not only common vocabulary words but also personal names, organization names, place names, and newly coined words.

Moreover, using information provided by the face detection technology, Keekie performs speech recognition only when Cahmoo detects a face. When we speak with someone else, we use not only the information contained in speech utterances, but also other information provided by gesture or face direction. For example, by watching the direction in which the other person is looking, we may recognize that he/she is talking to himself/herself. In such cases, we usually do not answer the interlocutor. In a similar way, Keekie uses the face direction information extracted by Cahmoo. We believe that with this architecture we have made a significant step toward more natural human-machine speech communication.

*4 **Open domain question answering (QA):** One of the most important features of ambient intelligence is its ability to conduct natural speech dialogues with humans. Our laboratories have been working on speech dialogue systems and QA systems for years [12], [13], and this demonstration uses our QA technology for "definitional questions" such as "Who is Claude Monet?" Our QA system answers this question by locating various bits of information about the person from numerous documents. Whowhoo uses this technology not to answer questions but to generate hints.

is achieved by combining speech recognition and face detection technologies. The combination of face detection, blind audio source separation, and speech recognition technologies makes these mushrooms a bit smarter.

4.2 Demonstration B: mushrooms that watch, listen, and search

Here, we introduce the mushrooms Searchie and Searchar. They watch and listen to things around them all the time and then find what someone wants from a tiny clue. These functions are achieved by media search technologies*5 developed over the last ten years at NTT CS Labs. In last year's "The World of Mushrooms" [1], we introduced Mookie and Moocha, which search for music based on short clips. They use audio search technology and a music database to reply with the name of a song that is being played. This year, we unveiled Searchie and Searchar, which can instantly locate the original image or audio file in a huge database based on small image segments or sound clips. They use video and audio search technologies and databases of music, music video clips, and TV programs. They provide the user with information such as the program title and its release date.

Searchie and Searchar use small video cameras as eyes and microphones as ears. Nowadays, it is not difficult to capture images from video cameras as digital information and sound from microphones. Modern large-capacity hard disk drives can now store several years' worth of data. The challenge is to find the information you want from among the huge number of sound and image files. Stored information cannot be used just by watching and listening without using a search method, i.e., a media search technology. Mushrooms watch TV and they instantly search huge databases for a variety of information about current TV programs, such as the title of music being played or information about whether or not the scene was used before. Even if images are distorted or intermittent or if the sound is noisy or interrupted, accurate searches remain possible.

4.3 Demonstration C: quiz-master mushroom

In this demonstration, we can learn new things subconsciously through a casual conversation with the mushrooms. The Whowhoo mushrooms engage in repartee in the form of a guessing game. They usually hide, but can occasionally pop up and quiz each

*5 **Media search technology:** In researching media search technology, we have concentrated on three key factors: rapid, robust, and smart. Rapid is the most basic characteristic of this technology [14], [15]. Indeed, our search technology can locate specific query matches in huge video and sound clip databases at extremely high speeds; for example, a database containing several hundred hours of material can be searched in one second. This key feature enables us to scan a huge database of hundreds of thousands of music pieces to find the name of background music or to scan a TV program database to find the specific scene in a drama.

Robust is another basic characteristic. Most media data is not cleanly observable in the real world. We have to catch the music in the presence of noise or recognize it in the presence of the distortion created by microphones/speakers and codecs used for transmission or archiving. We have to recognize video with occlusion, geometrical distortion, or color shifts. Robustness against noise and distortion is very important if media search technology is to lead to practical applications in the real world. To date, we have implemented a music search with mobile phone transmission [16] and a background music search in the presence of loud speech or other sounds [17]. We have also made a video search that is robust against partial occlusion and color shifts caused by camera/display characteristics.

The third characteristic, smart, is a little tricky. As explained above, robust targets changes caused by the external environment. Here, we would like to handle the changes in the target itself. For example, we would like to find the same object, not the same scene, even if the object was captured from a different angle. Taking a music search as an example, we would like to find the same musical score even though it was performed by a different person using a different instrument. To achieve these types of media search, we need a technique for modeling objects or music that parameterizes the possible changes, for example, in viewing angle or tone color of sounds. This is crucial for mushrooms, which will live with us for a long time and record all the happenings in our lives and then retrieve relevant information at the appropriate time and in the appropriate manner to help us become happy. Research on modeling in media searches is an important theme and we have already made significant progress.

other. For example:

Whowhoo-1: Hi! Let's start a quiz. Who is this? First hint: He's is a Japanese baseball player.
Whowhoo-2: Shigeo Nagashima?
Whowhoo-1: No, but Shigeo Nagashima is also a Japanese baseball player. Second hint: His wife is five years older than he is.
Whowhoo-2: Hiromitsu Ochiai?
Whowhoo-1: No, but Hiromitsu Ochiai is close to the correct answer. In fact, this person was also a member of Chunichi Dragons. Third hint: His team was defeated by Yanai High School in the finals at Koshien high school baseball championship.
Whowhoo-2: Ludwig van Beethoven?
Whowhoo-1: No, Ludwig van Beethoven is far from the correct answer. This person was born much later than Ludwig van Beethoven. Fourth hint: His name begins with B.

To make the quiz challenging enough, these hints are presented to the answerer in the order from the most ambiguous and thus most difficult one to the most definitive and thus easiest one based on the question answering technology*4.

Just overhearing this conversation allows us to extend our knowledge about the domain. You can join and answer the quiz yourself if you want. Our extremely large vocabulary speech recognition technology*3 [10] enables the recognition of a large vocabulary of two million words including many personal names. These entries were extracted from a large collection of documents by using named entity recognition (NER)*6.

Whowhoo indicates how close the answer was to the right one. It judges inter-person distances by using network mining techniques [18], [19], which analyze inter-person relationships and visualize the result as a network on a monitor screen. It is just like looking into the brain of Whowhoo.

5. Concluding remarks

This paper introduced demonstration systems that partially achieve our concept of ambient intelligence and described the fundamental technologies that we have developed so far. These technologies for ambient intelligence could transparently permeate our everyday lives. To create more ideal lifestyles and enhanced personal satisfaction through the use of ambient intelligence, a transdisciplinary approach that links related technical fields is essential. While some of these technologies have already been used in commercial products, their practical utility must be further enhanced and new application scenarios must be found. New ideas leading to future business should emerge from the lifestyles created by ambient intelligence.

Ambient Intelligence (*kankyo chinou*) Symposium 2006—The Future: A Tapestry Woven from Threads of Intelligence

To raise issues about the possibilities, future visions, and problems of information and communication technology (ICT) with a focus on ambient intelligence and to assert the importance of NTT's basic research related to ICT, we held "Ambient Intelligence (*kankyo chinou*) Symposium 2006—The Future: A Tapestry Woven from Threads of Intelligence" on September 22, 2006. The reports are summarized in another article in this issue [23]. A video of the symposium is available [24] for those interested in the details of the discussions at the symposium.

*6 **Named entity recognition (NER):** We developed NER systems based on state-of-the-art machine learning methods. Whowhoo uses a Japanese NER system based on a machine learning method called support vector machines (SVM). Since standard SVM algorithms are too inefficient to process gigabytes of documents, we devised an efficient algorithm optimized for NER [20]. To extract names from speech, we implemented a more robust NER system because automatic speech recognition (ASR) systems misrecognize some words [21]. To make the NER system more accurate, we are also working on another machine learning method called conditional random fields (CRF). To get better results in an arbitrary evaluation measure, we devised a general "consistent" framework for training CRF [22].

References

[1] NTT Communication Science Laboratories (ed.), "The World of Mushrooms," private edition, http://www.brl.ntt.co.jp/cs/kikang/index.html (Please contact us if you would like to receive a copy of the printed booklet (available in Japanese and English versions)).
[2] E. Maeda and Y. Minami, "Steps towards Ambient Intelligence," NTT Technical Review, Vol. 4, No. 1, pp. 50-55, 2006.
[3] E. Maeda, Y. Minami, and K. Dohsaka, "Resurgence of Fairies and Goblins—A Proposal for the New Vision of Ambient Intelligence," IPSJ Magazine, Vol. 47, No. 6 , pp. 624-640, 2006 (in Japanese).
[4] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J-C. Burgelman, "Scenarios for Ambient Intelligence in 2010 Final Report," ISTAG, 2001.
[5] P. Remagnino, G. L. Foresti, and T. Ellis (eds.), "Ambient Intelligence: A Novel Paradigm," Springer, 2004.
[6] G. Riva, F. Vatalaro, F. Davide, and M. Alcaniz (eds.), "Ambient Intelligence: The Evolution of Technology, Communication and Cognition Towards the Future of Human-computer Interaction," IOS Press, 2005.
http://www.ambientintelligence.org (2005)
[7] NTT Communication Science Laboratories (ed.), "The World of Mushrooms 2006," private edition (in Japanese). (Please contact us if you would like to receive a copy of this booklet).
[8] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc. of CVPR, Vol. 1, pp. 511-518, 2001.
[9] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind Extraction of Dominant Target Sources Using ICA and Time-Frequency Masking," IEEE Trans. Audio, Speech, and Language Processing, Vol. 14, No. 6, pp. 2165-2173, Nov. 2006.
[10] A. Nakamura, T. Oba, S. Watanabe, K. Ishizuka, T. Hori, M. Schuster, E. McDermott, and Y. Minami, "A benchmark evaluation of speech recognizer SOLON using the Corpus of Spontaneous Japanese," Proc. of Spring ASJ meeting (in Japanese), pp. 1185-1186, 2006.
[11] T. Hori, C. Hori, and Y. Minami, "Fast on-the-fly composition for weighted finite-state transducers in 1.8 million-word vocabulary continuous speech recognition," Proc. of Interspeech 2004, pp. 289-292, 2004.
[12] H. Isozaki, "An Analysis of a High Performance Japanese Question Answering System," ACM Transaction on Asian Language Information Processing, Vol. 4, No. 3, pp. 263-279, 2005.
[13] R. Higashinaka, R. Prasad, and M. Walker, "Learning to Generate Naturalistic Utterances Using Reviews in Spoken Dialogue Systems," COLING-ACL-2006, pp. 265-272, 2006.
[14] K. Kashino, T. Kurozumi, and H. Murase, "A Quick Search Method for Audio and Video Signals Based on Histogram Pruning," IEEE Trans. Multimedia, Vol. 5, No. 3, pp. 348-357, 2003.
[15] K. Kashino, A. Kimura, and T. Kurozumi, "A Quick Video Search Method Based on Local and Global Feature Clustering," Proc. of International Conference on Pattern Recognition (ICPR), Vol. 3, pp. 894-897, 2004.
[16] T. Kurozumi, K. Kashino and H. Murase, "A Robust Audio Searching Method for Cellular-Phone-Based Music Information Retrieval," Proc. of ICPR2002, Vol. 3, pp. 991-994, Aug. 2002.
[17] H. Nagano, K. Kashino and H. Murase, "A Fast Search Algorithm for Background Music Signals Based on the Search for Numerous Small Signal Components," Proc. of the 2003 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Vol. V, pp. 796-799, Apr. 2003, also in Proc. of 2003 IEEE International Conference on Multimedia and Expo (ICME 2003), Vol. I, pp. 165-168, July 2003.
[18] K. Saito, T. Yamada, and K. Kazama, "Extracting Communities from Complex Networks by the k-dense Method," 2nd International Workshop on Mining Complex Data (in press), 2006.
[19] K. Kazama, K. Saito, T. Yamada, and T. Matsubayashi, "Analysis of Reader Network on a Blog Service," JWEIN06, 2006 (in Japanese).
[20] H. Isozaki and H. Kazawa, "Efficient Support Vector Classifiers for Named Entity Recognition," Proc. of COLING-2002, pp. 390-396, 2002.
[21] K. Sudoh, H. Tsukada, and H. Isozaki, "Incorporating speech recognition confidence into discriminative named entity recognition of speech data," Proc. of COLING-ACL-2006, pp. 617-624.
[22] J. Suzuki, E. McDermott, and H. Isozaki, "Training Conditional Random Fields with Multivariate Evaluation Measures," Proc. of COLING-ACL-2006, pp. 217-224.
[23] K. Dohsaka, Y. Minami, A. Mori, and T. Kondo, "Report on "Ambient Intelligence Symposium 2006—The Future: A Tapestry Woven from Threads of Intelligence," NTT Technical Review, Vol. 4, No. 12, pp. 64-69, 2006 (this issue).
[24] http://www.kecl.ntt.co.jp/KCS2006/ (in Japanese).

Eisaku Maeda

Senior Research Scientist, Supervisor, Group Leader of Ambient Intelligence Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E. and M.E degrees in biological science and the Ph.D. degree in mathematical engineering from the University of Tokyo, Tokyo, in 1984, 1986, and 1993, respectively. He joined NTT in 1986. He was a guest researcher at the University of Cambridge, UK, in 1996-1997. His research interests are in statistical machine learning, intelligence integration, and bioinformatics. He is a senior member of IEEE and a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, the Information Processing Society of Japan (IPSJ), and the Japanese Society of Bioinformatics.

Yasuhiro Minami

Senior Research Scientist, Ambient Intelligence Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Tokyo, in 1986, 1988, and 1991, respectively. He joined NTT in 1991 and worked on robust speech recognition. He was a visiting researcher at MIT from 1999 to 2000. Since February 2000, he has been with NTT Communication Science Laboratories. He is interested in modeling for speech recognition. He is a member of IEEE, IEICE, IPSJ, and the Acoustical Society of Japan (ASJ).

Masato Miyoshi

Senior Research Scientist, Supervisor, Group Leader of Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Ph.D. degrees in electrical engineering from Doshisha University, Kyoto, in 1981, 1983, and 1991, respectively. Since joining Nippon Telegraph and Telephone Public Corporation (now NTT) as a researcher in 1983, he has continued to study signal processing theory and its application to acoustic technologies. In 1991-1992, he was a visiting researcher at Univ. Bochum, Germany. He is a senior member of IEEE and a member of IEICE, ASJ, and the Audio Engineering Society.

Minako Sawaki

Senior Research Scientist, Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

She received the B.E. degree in electrical engineering from Keio University, Tokyo, in 1989. She joined NTT in 1989. She is interested in pattern recognition and computer vision. She is a member of IEICE and the Institute of Image Information and Television Engineers.

Hiroshi Sawada

Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, in 1991, 1993, and 2001, respectively. He has been engaged in research on signal processing, microphone array, and blind source separation. He is a senior member of IEEE and a member of IEICE and ASJ.

Atsushi Nakamura

Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Dr.Eng. degrees in communication engineering from Kyushu University, Fukuoka, in 1985, 1987, and 2001, respectively. He joined NTT in 1987. His research interests include speech recognition and learning theory and its application to spoken language processing. He is a member of IEEE, IEICE, and ASJ.

Junji Yamato

Senior Research Scientist, Supervisor, Group Leader of Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E, M.E., and Ph.D. degrees in precision machinery engineering from the University of Tokyo, Tokyo, in 1988, 1990, and 2001, respectively, and the S.M. degree in electrical engineering and computer science from MIT, Massachusetts, in 1998. He joined NTT in 1990. His research interests are in computer vision, human-computer interaction, and robotics. He is a member of IEEE, the Association for Computing Machinery (ACM), IEICE, and the Robotics Society of Japan.

Takeshi Yamada

Senior Research Scientist, Supervisor, Group Leader of Emergent Learning and Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the B.S. degree in mathematics from the University of Tokyo, Tokyo, in 1988 and the Ph.D. degree in informatics from Kyoto University, Kyoto, in 2003. He joined NTT in 1988. His research interests are in machine learning, data mining, and combinatorial optimization. He is a member of ACM, IEEE, IEICE, IPSJ, and the Scheduling Society of Japan.

Ryuichiro Higashinaka

Research Scientist, Knowledge Processing Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received the B.A. degree in environmental information and the Master of Media and Governance degree from Keio University, Tokyo, in 1999 and 2001, respectively. He joined NTT in 2001. During 2004-2006, he was a visiting researcher at the University of Sheffield, UK. His research interests are in question answering systems and utterance understanding and generation in spoken dialogue systems. He is a member of the Association for Computational Linguistics and IPSJ.