The World of Mushrooms: Human-Computer Interaction Prototype Systems for Ambient Intelligence

Yasuhiro Minami, Minako Sawaki, Kohji Dohsaka, Ryuichiro Higashinaka, Kentaro Ishizuka, Hideki Isozaki, Tatsushi Matsubayashi, Masato Miyoshi, Atsushi Nakamura, Takanobu Oba, Hiroshi Sawada, Takeshi Yamada, and Eisaku Maeda
NTT Communication Science Laboratories
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, JAPAN
+81 774 93 5323
minami@cslab.kecl.ntt.co.jp

ABSTRACT

Our new research project called "ambient intelligence" concentrates on the creation of new lifestyles through research on communication science and intelligence integration. It is premised on the creation of such virtual communication partners as fairies and goblins that can be constantly at our side. We call these virtual communication partners mushrooms.

To show the essence of ambient intelligence, we developed two multimodal prototype systems: mushrooms that watch, listen, and answer questions and a Quizmaster Mushroom. These two systems work in real time using speech, sound, dialogue, and vision technologies.

We performed preliminary experiments with the Quizmaster Mushroom. The results showed that the system can transmit knowledge to users while they are playing the quizzes.

Furthermore, through the two mushrooms, we found policies for design effects in multimodal interface and integration.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – prototyping.

General Terms

Design, Human Factors, Experimentation, Measurement, Languages

Keywords

dialog multimodal interfaces, visual-auditory feedback.

1. INTRODUCTION

In October 2004, our laboratories initiated a research project whose theme was ambient intelligence [3][14][16][20]. It had two main purposes: 1) creating new lifestyles through the research and disciplines and strategically developing transdisciplinary areas, as symbolized by the term "intelligence integration."

Terms such as ubiquitous or pervasive approach a description of the concept. However, since such R&D studies tend to focus on computers and sensors, they are geared toward hardware or devices.

Ambient intelligence, which suggests intelligence that surrounds us, is also used for research being pursued in the EU and the USA [4][21][22]. These studies aim to design a future information society by exploiting computers and sensors embedded in the environment to create more natural and intelligent interfaces.

Our aim is to discover what kind of communication we really need and to achieve it. We plan to achieve ambient intelligence using information and communication science technologies. That is, we will exploit our areas of expertise—speech, sound, language, dialogue, vision, data retrieval, and networking—to create a new style of ambient intelligence that places human intelligence and intellect at the forefront. This will lead to proposals about future lifestyles and clarify future specific issues [15]. Here mental stability and satisfaction rather than physical convenience have become new targets. Furthermore, ambient intelligence may drive a robot brain or be embedded in the environment of ubiquitous computers and networks [5][24].

We discussed the possible lifestyles that could be realized by ambient intelligence and suggested specific concrete issues to be tackled to achieve ambient intelligence in [15]. To achieve an embodied world of ambient intelligence, making prototype systems using state-of-the-art technologies is important. Evaluating prototype systems allows us to find weaknesses in these technologies and the missing technologies as well as providing a chance to enhance them and create new research fields. For this purpose, we made two prototype systems. One shows the concept of mushrooms that watch, listen, and answer questions, and another is a Quizmaster Mushroom.

In this paper, we briefly review the concept of ambient intelligence and then explain the state-of-the-art technologies utilized in the prototype systems. Finally, we explain the scenario, the main flows of the two prototype systems, and a preliminary experiment.

2. BASIC CONCEPT OF MUSHROOMS

One aspect of ambient intelligence is based on the creation of virtual communication partners that are constantly at our side. To

ICMI'07, November 12–15, 2007, Nagoya, Aichi, Japan.

develop this concept, we considered appropriate fundamental features. For example, such entities need a personality and the ability to evolve, grow, and express feelings. There are many entities; they lurk nearby, unobtrusively, even discreetly; when called, they answer, and otherwise, they are hidden, as shown in Figure 1.

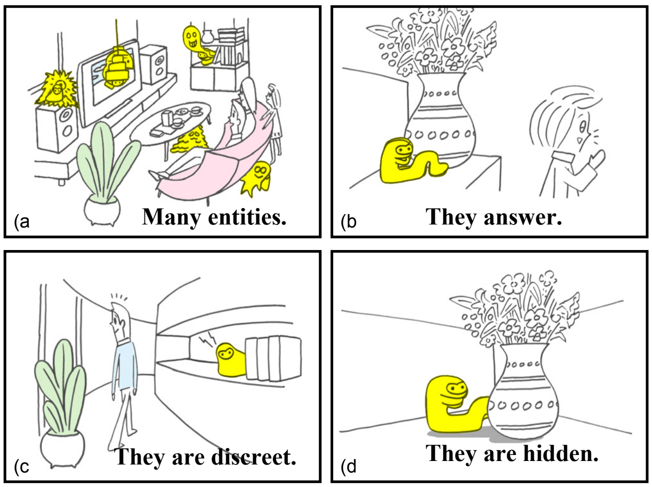

Figure 1. Ecology of mushrooms.

The world of ambient intelligence resembles the world of fairies and goblins that is familiar in both the East and the West. Therefore, they are a satisfactory concept on which to base a common understanding of ambient intelligence. Ambient intelligence also provides a means for communication that addresses the heart, as fairies and goblins used to do.

2.1 Mushrooms and Their Design

Mushrooms are what we call our fairies and goblins. We gave them shapes and design coherence. Each mushroom has one or two specialized abilities to which their names and shapes correspond. Three examples of mushrooms are shown in Figure 2. Keekie has the ability to listen, Chacha can speak, and Sheeshie can provide knowledge acquired from newspaper articles. Their shapes change as they evolve or as their abilities improve. For example, they may become more complicated in shape, lose the dark circles under their eyes, grow longer hair, or acquire more rings around their body as their knowledge increases. We think that above abilities are kinds of individualities that mushrooms have. If each ability is assigned to a mushroom that has an original shape, it is easy for us to understand what kind of thing the mushroom can do by looking at its shape.

Considerations related to these design issues may be of secondary importance in terms of research and development. Nevertheless, inspired by such designs or shapes, new ideas may emerge.

The evolution of mushrooms is as shown in Figure 3. The process and future perspectives of mushroom evolution represent the progress of R&D itself and a road map for the future of ambient intelligence.

Figure 2. Appearance of mushrooms.

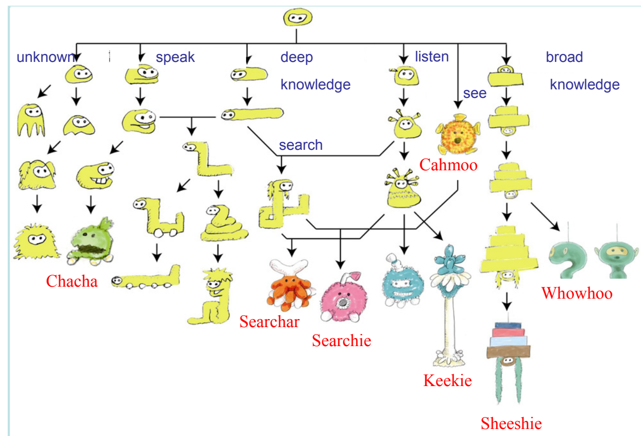

Figure 3. Evolution of mushrooms.

3. KEY TECHNOLOGIES IN PROTOTYPE SYSTEMS

This section explains the latest technologies developed in our laboratory, many of which must be combined to achieve the potential of the World of Mushrooms system. We integrate and evaluate state-of-the-art technologies to find weaknesses in these technologies and to open new research fields made possible by intelligence integration.

3.1 Face Detection

Face detection locates a user's facial area and determines its size in an image continuously captured by the mushroom's USB camera. This is done to check whether the user is approaching the mushroom. Face detection speed is accelerated by two-step processing. First, most of the non-face areas in the captured image are rejected using a method based on [23]. Here, the sum of the pixel values of sub-images is evaluated to quickly reject obvious non-face areas. In the second step, the remaining sub-images are then classified in detail as face or non-face using SVM [12]. The SVM employs a polynomial kernel (d=2).

3.2 Blind Audio Source Separation

Blind audio source separation is applied on a wide range of systems including noise robust speech recognition and hands free telecommunication. To focus on the target speaker's voice, an array of microphones is used. By only using the signals captured by these microphones, array directivity can be controlled to discriminate between target speech and undesired sounds and to suppress the undesired sounds. Such processing can be performed in a blind manner, i.e., without any *a priori* information about the source direction [2]. Several kinds of blind audio source separation methods have been proposed. In the prototypes, we use maximum signal to noise ratio (MaxSNR) beamformers [1] among these methods to combine face detection technology. The MaxSNR beamformer maximizes the ratio between signal power and interference and noise power.

3.3 Extremely Large Vocabulary Speech Recognition

A Speech recognizer with OutLook On the Next generation (SOLON) speech recognizer [19] features an algorithm based on fast on-the-fly composition for weighted finite-state transducer (WFST) type advanced computational models. Due to its accuracy and efficiency, the algorithm enables, for the first time in the world, real-time speech recognition with a vocabulary of several million words [9]. This technology makes it possible to handle a huge variety of expressions not only including common words and neologisms but also personal, organization, and place names.

We made WFSTs to prepare a basic recognizer for the two prototypes. The WFSTs used here were created from two models: phoneme Hidden Markov Models (HMMs) and language models. The language models are basically trained using two databases: newspaper articles covering the last 12 years and about 24,000 question answering (QA) task sentences. Named entities were extracted for all the training data using named entry recognition described in 3.5. Consequently the vocabulary contains 1.8 million words. Phoneme models were trained using 67,244 sentences uttered by about 360 speakers.

3.4 Open Domain Question Answering (QA)

Our laboratories have been working on speech dialogue systems and QA systems for years [6], [10], and this prototype 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. The system looks for answers in 14 years worth of computerized newspaper articles from 1991 to 2004.

3.5 Named Entity Recognition (NER)

We developed NER systems based on state-of-the-art machine learning methods. We used 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 [11]. To extract names from speech, we implemented a more robust NER system because automatic speech recognition (ASR) systems sometimes misrecognize words [26]. 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 [28].

3.6 Speech Synthesis

We used a corpus-based text-to-speech technology called “Cralinet” developed by NTT Cyber Space Laboratories [18]. Since the original corpus was designed for a voice used in public announcements, we prepared a new corpus that consists of a voice database of expressions uttered in a childish way by a voice actress. This approach was taken so that the synthesized voice would match the playful mood of the mushrooms.

4. THE WORLD OF MUSHROOMS

We developed two prototype systems to embody the world of mushrooms. Several mushrooms as prototypes were created based on the basic design concept.

However, even if we use the latest technologies, it is difficult to completely embody the concept. Therefore as the first step, we achieved mushrooms that have the basic functions described in Figure 3: seeing, speaking, having broad knowledge, and listening. Prototypes of mushrooms that watch, listen, and answer questions were implemented based on this concept.

In the world of mushrooms, mental stability and satisfaction rather than physical convenience are important. For this concept, we made a prototype Quizmaster Mushroom that transmits knowledge from the system to users while they are playing a quiz with it.

4.1 Prototype of Mushrooms that Watch, Listen, and Answer Questions

The goal of prototype mushrooms that watch, listen, and answer questions is to capture user questions under different noisy environments and answer questions from a large amount of newspaper articles. Figure 4 shows the system diagram of mushrooms that watch, listen, and answer questions. This prototype uses five main intelligent modules: blind audio source separation, extremely large vocabulary speech recognition, Japanese question answering, face detection, and speech synthesis. We designed a control module to organize all intelligent modules. The characters in the figure are mushrooms assigned to the input and output devices and designed to appear collaboratively. Keekie and Chamoo serve speech and image inputs, respectively. Sheeshie serves speech and display outputs. Figure 5 shows an actual picture of this prototype.

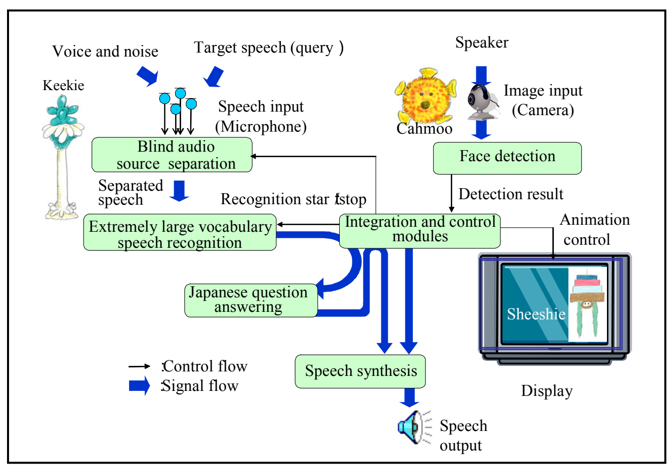

Figure 4. System diagram of mushrooms that watch, listen, and answer questions.

4.1.1 Scenario

The system is based on the following scenario. Cahmoo is always watching. When a person approaches close enough to Cahmoo and faces its camera, Cahmoo assumes that the person wants to talk and starts a conversation with a greeting, such as “Hello” or “What’s up?” Sheeshie starts a conversation the moment Cahmoo notices that a person is facing him. When the person asks Keekie a question such as "Where did the 2006 World Cup take place?" Sheeshie answers, "Germany."

Figure 5. Mushrooms that watch, listen, and answer questions.

4.1.2 Main Flow of Mushrooms that Watch, Listen, and Answer Questions

Figure 6 shows the main flow of the prototype mushrooms that watch, listen, and answer questions. This flow is obtained by rewriting the signal flows (thick arrows) in Figure 4 in order of time. The system uses four microphones located at vertexes of regular tetrahedron frames. They record utterances with environmental noise and voices coming from other speakers. Blind audio source separation suppresses the environmental noise and the voices in the recorded sounds to emphasize utterances from the interlocutor. To do this, we introduce a maximum Signal to Noise Ratio (SNR) beamformer described in [1]. Then the SOLON speech decoder recognizes the emphasized speech using WFSTs converted from phoneme HMMs and the language model described in Section 3. The recognized speech is analyzed by QA to check the question type. Based on the question type, QA seeks the answer from the newspaper database. Finally, the answer is synthesized by child voice speech synthesis.

In addition to the main flow, we introduce control flow using face detection (thin arrows in Figure 4) to achieve higher audio discrimination performance and reduce recognition errors. Figure 7 shows this flow. When no face is detected, the captured signal is judged to consist only of undesired sounds whose directions are estimated by minimizing their gains. If a face is detected, the probability of hearing speech increases, and target speech is extracted by removing the undesired sounds using the previously obtained information (their directions). This allows the system to clearly hear the speech even in a noisy environment such as an exhibition hall.

Moreover, while answering, Cahmoo continues to watch the interlocutor to guess whether he/she wants to pose another query. This new capability is achieved by combining speech recognition and face detection technologies. The system performs speech recognition when faces are detected to allow the system to listen to speech only when users face the system. When the system detects a user's face by the face detection method described in 3.1, it starts the speech recognition function. On the other hand, when the system does not detect a face within a certain time threshold, it stops the speech recognition.

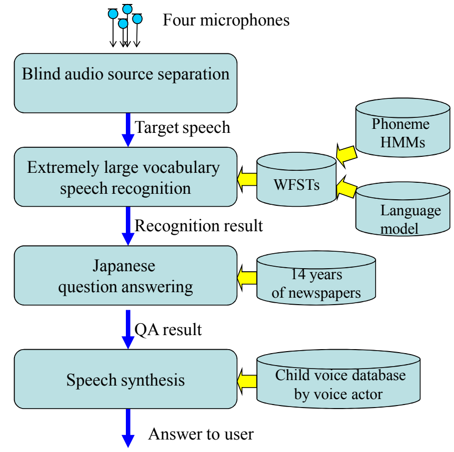

Figure 6. Main flow of mushrooms that watch, listen, and answer questions.

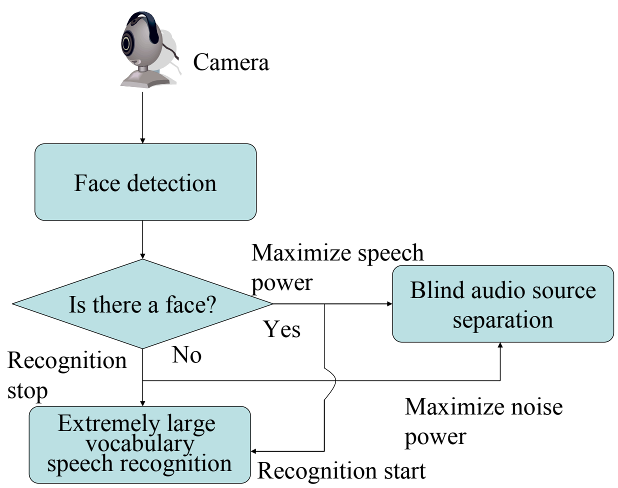

Figure 7. Procedure using face detection in mushrooms that watch, listen, and answer questions.

4.2 Prototype: Quizmaster Mushroom

The general aim of interactive systems is to shorten the interaction time and efficiently transmit information from the users to the system or from the system to the users. On the contrary, we aim to achieve a system that users actually want to use. For this purpose, we believe that it is important to transmit information for a long time while the users enjoy the system. Developing a system that can be used for a long time compels us to find a way to maintain mental satisfaction, since a large volume of information would be transferred to the users.

The purpose of the prototype Quizmaster Mushroom is to transmit knowledge from the system to users while they are playing a quiz game with the prototype system. The prototype system can give a quiz on a particular person whose name is selected from the

Internet. Figure 8 shows a diagram of the Quizmaster Mushroom, which employs five intelligent modules: automatic quiz creation, quiz ranking, network mining, extremely-large-vocabulary speech recognition, and speech synthesis. We also implement a control module to organize all of the intelligent modules. For this prototype, we created the main mushroom personality, named "Whowhoo," who functions as an emcee. The only system input is speech. System outputs are speech and display. Figure 9 shows a photograph of this prototype.

4.2.1 Dialogue Example

Using this prototype, we learn new things through casual conversations with the mushrooms. Whowhoo engages the user in repartee with a guessing game. For example:

Whowhoo: Hi! Let's play a quiz. Who is this?
First hint: He is related to computers.
Person: *Mavin Minsky*?
Whowhoo: No, not an information engineer.
Second hint: He studied chemical engineering at Berlin University and the Swiss Federal Institute of Technology, and mathematics at Budapest University.
Person: *Kurt Gödel*?
Whowhoo: Nice Try! But He lived before *Kurt Gödel*.
Third hint: He is one of the pioneers of dynamic meteorology, together with *Jule Gregory Charney* and *Fjortof*.
Person: *Albert Einstein*?
Whowhoo: Close! But not a German physicist.
Fourth hint: Born into a noble and wealthy banking family, he showed a remarkable talent and received an early education.
Person: *Alan Turing*?
Whowhoo: No, but a good guess in terms of a computer-related person.
Fifth hint: He could calculate eight-digit numbers mentally.
Person: John von Neumann
Whowhoo: You're right! Well done! Excellent!

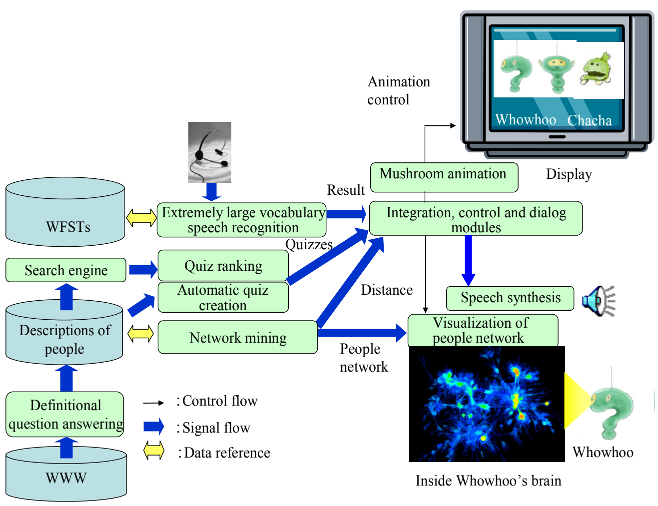

Figure 8. System diagram of Quizmaster Mushroom.

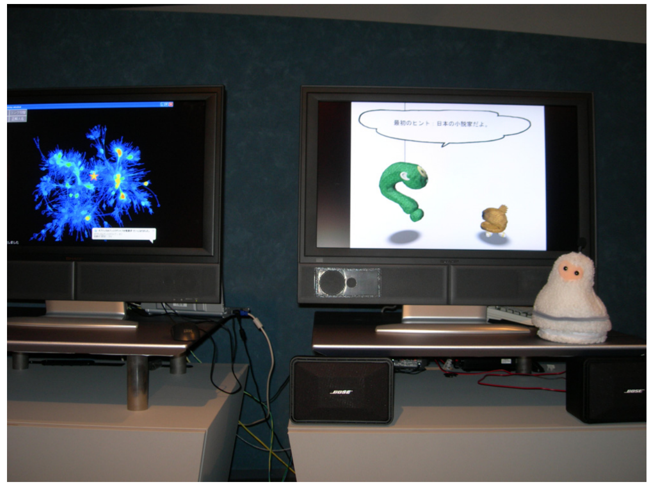

Figure 9. Photograph of Quizmaster Mushroom prototype.

4.2.2 Main Flow of Quizmaster Mushroom

Figure 10 shows the main flow of the Quizmaster Mushroom. This flow is obtained by rewriting the signal flows (thick arrows) in Figure 8 in order of time.

In advance, about 30,000 famous people were selected from the Internet to generate a name list. "Who is this?" quizzes are created by ordering N descriptive statements about person X based on the difficulty of naming X. First, descriptive sentences about X are collected from the web by a pattern-based approach [27] or from Wikipedia by using all sentences in the entry of *X*. Then, by hand-crafted rules, we filter ambiguous and answer-revealing sentences that are ordered based on their content word co-occurrences with *X* measured by pointwise mutual information (PMI) [7].

The name list is also used to make a language model for speech recognition (see preliminary experiment).

After starting a quiz, the system gives the top quiz from the ordered sentences using speech synthesis. The user responds to the quiz through a close-talking microphone; the system recognizes the utterance with the SOLON speech decoder.

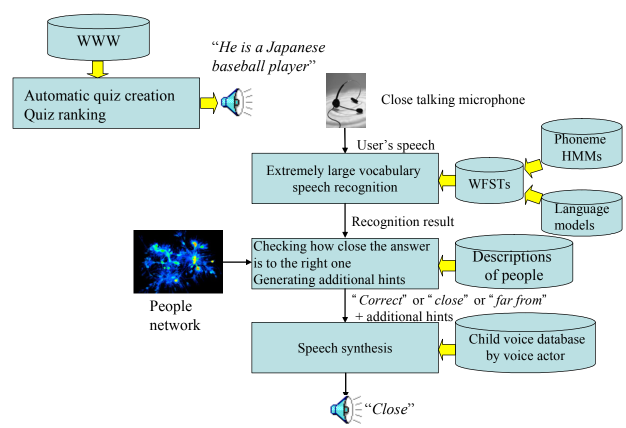

Figure 10. Main flow of Quizmaster Mushroom prototype.

The system then indicates how close the answer is to the right one. It judges inter-person distance using network mining techniques that analyze inter-person relationships and visualizes the result as a network defined in two-dimensional space on a monitor screen (see left picture in Figure 10) [13][25]. A person network is generated from information on about 10,000 people from the Internet based on the co-occurrences of two arbitrary people. The network representation provides a structure of a person's relations to us by examining an embedded layout in a low-dimensional Euclidean space. We use a fast method, inspired by the hierarchical individual time-step method used in astrophysical dynamics, to generate this network [17]. The system simultaneously generates additional hints with user answers.

If the user's answer is incorrect, the system selects the next quiz question from the person's ordered sentences. Then the system iterates the above procedure until the user answers correctly, quizzes for the person are exhausted, or the user quits the quiz.

4.2.3 Preliminary Experiment

To check the performance of the prototype quizmaster, a preliminary experiment was performed in a sound booth. Twenty-four speakers (12 males and 12 females) were used in this experiment. We selected 50 person names from the name list described in 4.1 for the quiz. Each speaker used the system in eight sessions. Each session lasted fifteen minutes. We evaluated this system for speech recognition and the users' correct answer rates in the quiz.

We modified the language model described in 3.3. Table 1 shows the sentences that were used to make a lexicon and a language model.

Table 1. Sentences used for training trigram language model.

	# of sentences	Weight
News paper sentences	3255408	1
Name sentences	312780	30
Interrogative sentences	23772	120
Various expressions	1284	120

The original newspaper database has a large number of sentences that do not contain person names. To eliminate these sentences, we used only the newspaper sentences that had at least one person name. We also prepared the sentences with a person name in such a way to improve the recognition performance for quiz answers. These sentences were made by inserting pauses and fixed phrases before and after a person name in the name list. We also used interrogative sentences, which were made by hand for the QA task. Although such sentences may not be optimally effective for the quiz task, we took into account the possible extension by combining the prototype systems in 4.1 and 4.2; therefore, we used the sentences for training the language model so that the system could recognize a large variety of expressions. Considering the situation when users would have a question about the system's hint and would ask the question to the system, the language model would enrich the dialog between the users and the system. Finally, we prepared various expressions such as greetings (e.g. "hello Keekie,") and some control commands (e.g. "give up" and "next hint").

We made the language model using the above sentences. Each set of sentences was multiplied by the weight in count frequency (see Table 1) to make the language model; the vocabulary contained 1.2 million words.

Table 2. Recognition conditions and recognition results.

	# of subjects	# of utt.	Word percent correct	
			With "next hint"	Without "next hint"
Male	12	4234	88.2%	65.3%
Female	12	3710	91.3%	68.0%

Although the system can recognize various sentences, the speakers were requested to answer with a single person name after hearing the system's hints. The speakers were also advised that they could use the phrase "next hint" to pass the hint.

Table 2 shows recognition conditions and recognition results. Total recorded utterances are 4234 for males and 3710 for females; word percent corrects were 88.2% and 91.3%, respectively. However, since hints were sometimes too cryptic to help the user answer the person name, the users used many "next hint" phrases. We removed utterances of "next hint" from the evaluation data, and the last column shows the corresponding recognition results. Word percent corrects were 65.3% and 68.0%, respectively. The experiment's results indicate that the system works very well for the users.

The total number of quizzes was 1092, and the percentage of correct answers for the quizzes was 60.4%. The users answered 258 quizzes twice, where the correct answer rate for the first time was 63.6% and that for the second time was 84.5%. There was one week interval between two experiments. They seemed to learn about the person from the system. This means that the system can transmit knowledge to users while they are playing a quiz game with it. In fact, our previous work [8], which showed user understanding of a quiz system using text input, had significantly better results than the read-out system.

Since the evaluation in [8] was obtained with an interface using text, we will perform a more detailed evaluation using this prototype system.

5. Discussion

The two prototypes can work in real time. Their machine specifications are shown in Tables 3 and 4, respectively.

Table 3. Machine specifications for mushrooms that watch, listen, and answer questions.

Process	CPU	Clock	Memory
Main control	Xeon * 2	3.6 GHz	4 GB
Blind audio source separation	Athron 64	2.4 GHz	3 GB
WWW server	Athron 64	2.4 GHz	3 GB
Speech recognition	Pentium 4	3.4 GHz	3 GB
Speech recognition (back up)	Pentium 4	3.4 GHz	3 GB
Face detection & display control	Pentium 4	3.4 GHz	3 GB
Speech synthesis	Pentium 4	3.6 GHz	2 GB
QA server	Pentium 4	3.4 GHz	3 GB

Table 4. Machine specifications for Quizmaster Mushroom.

Process	CPU	Clock	Memory
Speech recognition	Pentium 4	3.4 GHz	3 GB
Quiz control	Pentium 4	3.4 GHz	3 GB
Speech synthesis Display control	Pentium 4	3.4 GHz	3 GB
Speech synthesis Information visualization	Pentium M	1.3 GHz	1 GB

Although one prototype used eight computers while the other used four computers, neither used special computers. When researchers used the prototype systems, communication between the systems and the researchers was very natural.
We demonstrated the two prototypes in our lab's 2006 Open House event. During demonstrations, many visitors mentioned that they felt a good impression of our prototypes. Female visitors in particular seemed to treat the mushrooms with great interest due to their appearance and shape. From our subjective observations of both prototypes' demonstrations, we found the following considerations.

(1) The appearance of the mushrooms must play a role in the multimodal interface. However, there is no good methodology for evaluating these roles, since the effect of appearance on users cannot be objectively observed.
(2) It is also important that mushrooms have individuality, such as amusing characteristics and earnestness. That is, when two or three mushrooms are talking to each other, such individuality would play a big role in transferring the information to the users.
(3) Integrating face detection, speech recognition, and audio source separation, as well as their real-time performance, is crucial. Since face detection generally involves error, such error sometimes causes fatal mistraining of the blind source separation parameters, which in turn causes fatal error in communication. Another problem is that a delay in any process disturbs natural communication between users and the system. To overcome these problems, we must continue to improve each technology and closely integrate the technologies to reduce error and shorten processing time.

Finally, opening new research paths is vital to the world of mushrooms.

6. CONCLUSIONS

Our new research project called "ambient intelligence" concentrates on the creation of new lifestyles through research on communication science and intelligence integration. To show the essence of ambient intelligence, we developed two multimodal prototype systems: mushrooms that watch, listen, and answer questions and a Quizmaster Mushroom. These two systems work in real time using sound, speech, dialogue, and vision technologies. This paper first described these fundamental technologies and then described the two prototype systems. Prototype mushrooms that watch, listen, and answer questions heed user questions under various noisy environments and answer questions from a large volume of newspaper information. The prototype Quizmaster Mushroom quizzes users. Preliminary experiments with the Quizmaster Mushroom show that the system can transmit knowledge from the system to users while playing a quiz game.

By demonstrating our two prototypes, we discovered new research areas: designing effects in multimodal interfaces and integrating them with face detection, blind audio source separation, and speech recognition. Consequently, we can demonstrate the effectiveness of our approach in realizing future lifestyles with current technologies. And from there we can find new ideas.

Although we found that some people enjoyed the system and received some knowledge while playing with it, the current status of our technology is still too immature to discuss the mental effects described in section 2.1. Before discussing methods to attain appropriate mental impressions, we have to step toward new research areas such as recognition and understanding of human emotions and actions, real-time intelligence-integration processing, and planning and execution of actions that cause changes in human emotions and intelligence.

7. ACKNOWLEDGMENTS

We would like to thank K. Shimomura, Y. Kashihara, and K. Maeda for supporting the system implementations. We would like to express our appreciation to the Minoh Lab. at Kyoto University for help with the experiments.

8. REFERENCES

[1] Araki, S., Sawada, H, and Makino, S. Blind Speech Separation in a Meeting Situation with Maximum SNR beamformers, In Proc of ICASSP 2007, vol.1, pages 41-44, 2007.

[2] Benesty, J., Makino, S., and Chen, J. Speech Enhancement, Springer, The Netherlands, pages 299-327, 2005.

[3] Dohsaka, K., Minami, Y., Mori, A., and Kondo, T. Report on "Ambient Intelligence Symposium 2006—The Future: A Tapestry Woven from Threads of Intelligence. NTT Technical Review, Vol. 4, No. 12, pages 64-69, 2006.

[4] Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., and Burgelman, J-C. Scenarios for Ambient Intelligence in 2010 Final Report. ISTAG, 2001.

[5] Fujita, M. On Activating Human Communications with Pet-type Robot AIBO. In Proc. of IEEE, vol. 92, no.11, pages1804-1813, Nov. 2004.

[6] Higashinaka, R., Prasad, R. and Walker, M. Learning to Generate Naturalistic Utterances Using Reviews in Spoken Dialogue Systems. In Proc. of COLING-ACL-2006, pages 265-272, 2006.

[7] Higashinaka, R., Dohsaka, K., Isozaki, H. Learning to Rank Definitions to Generate Quizzes for Interactive Information Presentation, In Proc. of ACL 2007 (Poster Presentation) pages 117-120, 2007.

[8] Higashinaka, R., Dohsaka, K., Amano, S., Isozaki, H., Effects of Quiz-style Information Presentation on User Understanding, In Proc. of Interspeech 2007, pages 2725-2728, 2007.

[9] Hori, T., Hori, C., and Minami, Y. Fast on-the-fly Composition for Weighted Finite-State Transducers in 1.8 Million-Word Vocabulary Continuous Speech Recognition. Proc. of Interspeech 2004, pages 289-292, 2004.

[10] Isozaki, H. An Analysis of a High Performance Japanese Question Answering System. ACM Transaction on Asian Language Information Processing, Vol. 4, No. 3, pages 263-279, 2005.

[11] Isozaki H. and Kazawa, H. Efficient Support Vector Classifiers for Named Entity Recognition. In Proc. of COLING-2002, pages 390-396, 2002.

[12] Joachims, J. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Scholkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[13] Kazama, K., Saito, K., Yamada, T. and Matsubayashi, T. Analysis of Reader Network on a Blog Service. JWEIN06, 2006 (in Japanese).

[14] Maeda, E. and Minami, Y. Steps towards Ambient Intelligence. NTT Technical Review, Vol. 4, No. 1, pages 50-55, 2006.

[15] Maeda, E., Minami, Y., and Dohsaka, K. Resurgence of Fairies and Goblins—A Proposal for the New Vision of Ambient Intelligence. IPSJ Magazine, Vol. 47, No. 6, pages 624-640, 2006 (in Japanese).

[16] Maeda, E., Minami, Y., Miyoshi, M., Sawaki, M., Sawada, H., Nakamura, A., Yamato, J., Yamada, T., and Higashinaka, R. The World of Mushrooms---Transdisciplinary Approach to Human-Computer Interaction with Ambient Intelligence. NTT Technical Review, Vol. 4, No.12, pages17-25, 2006.

[17] Matsubayashi, T., and Yamada, T. A Force-directed Graph Drawing based on the Hierarchical Individual Timestep Method, In Proc. of IJECS, pages 116-121, 2007.

[18] Mizuno, H., Asano, H., Isogai, M., Hasebe, M. and Abe, M. Text-to-Speech Synthesis Technology Using Corpus-Based Approach, NTT Technical Review, Vol. 2, No. 3, pages 70-75, Mar. 2004.

[19] Nakamura, A., Oba, T., Watanabe, S., Ishizuka, K., Hori, T., Schuster, M., McDermott, E., and Minami, Y. A Benchmark Evaluation of Speech Recognizer SOLON using the Corpus of Spontaneous Japanese. In Proc. of Spring ASJ meeting, pages 1185-1186, 2006 (in Japanese).

[20] NTT Communication Science Laboratories (ed.) The World of Mushrooms. Private edition, http://www.brl.ntt.co.jp/cs/ai/index-j.html (Contact us for a copy of the printed booklet available in Japanese and English).

[21] Remagnino, P., Foresti, G. L., and Ellis, T. (eds.). Ambient Intelligence: A Novel Paradigm. Springer, 2004.

[22] Riva, G., Vatalaro, F., Davide, F., and Alcaniz, M. (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.

[23] Viola, P. and Jones, M. Rapid Object Detection using a Boosted Cascade of Simple Features. In Proc. of CVPR, Vol. 1, pages 511-518, 2001.

[24] Sato, M., Sugiyama, A., and Ohnaka, S. Auditory System in a Personal Robot, PaPeRo. In Proc. of ICCE ’06, pages19-20.

[25] Saito, K., Yamada, T., and Kazama, K. Extracting Communities from Complex Networks by the k-dense Method. 2nd International Workshop on Mining Complex Data, 2006.

[26] Sudoh, K., Tsukada, H., and Isozaki, H. Incorporating Speech Recognition Confidence into Discriminative Named Entity Recognition of Speech Data. In Proc. of COLING-ACL-2006, pages 617-624, 2006.

[27] Sun, R., Jiang, J., Tan, Y. F., Cui, H., Chua, T.-S., and Min-Yen Kan, M.-Y. Using Syntactic and Semantic Relation Analysis in Question Answering. In Proc. of TREC, 2005.

[28] Suzuki, J., McDermott, E., and Isozaki, H. Training Conditional Random Fields with Multivariate Evaluation Measures. In Proc. of COLING-ACL-2006, pages 217-224, 2006.