# Multimodal Interface Agent for Enhancing Knowledge Sharing

Yoichi Takebayashi, Yasuko Nakayama, Hiroshi Kanazawa, Yasuyuki Kono and Shuichi Tsujimoto
Research and Development Center, Toshiba Corporation

## INTRODUCTION

The flood of information swirling through the Internet and pouring out of storage has risen well beyond our perceptual abilities. Studies in intelligent information retrieval and knowledge navigation [1,2,3] have sought to retrieve knowledge truly useful to a user from a vast amount of information on the Internet. However, because this kind of information is extremely diverse, it is almost impossible to define every concept needed to fully understand the information. In contrast, domain-specific information, such as "stock" information in an organization, can be easily retrieved at the semantic level.

Our interest lies in how the perceptual user interface can support knowledge/information sharing among people as well as intergroup communication. The main problem here is a lack of a practical means to store and retrieve databases of individual business knowhow. Our new multimodal interface agent system realizes this by providing knowledge/information to many people in an organization through a network.

## MULTIMODAL INTERFACE AGENT SYSTEM

We have newly developed a multimodal interface agent system "Multimodal Secretary Agent" . Figure 1 shows a user interacting with this multimodal interface agent in speech and by finger pointing. In this system, when a user enters a query in speech or natural language, it understands the user's request and retrieves the relevant knowledge/information from its knowledge databases, and then presents the appropriate answer to the user. As shown in Figure 2, the agent system consists of three components: (1) management of multimodal human-computer interaction, (2) structurization of multimedia information and (3) knowledge databases, in which knowledge bases (KBs) are linked to knowhow bases and multi databases. The first two components effectively use HI-ware (Common HI service environment), which is a set of media conversion/understanding functions [4], including speech recognition, document understanding and machine translation [5,6].

## MANAGEMENT OF MULTIMODAL HUMAN-COMPUTER INTERACTION

When a user enters a query in multimodal input, the agent understands the user's intention and situation from his/her utterances and gestures, and then retrieves the appropriate knowledge (knowhow) or information stored in knowledge databases.

In order to achieve a user-friendly perceptual user interface, the system utilizes an efficient multimodal input understanding method based on ATMS (Assumption-based Truth Maintenance System) [7], which deals with integration of delayed-arrival data, as well as elimination of ambiguities in recognizing the results of each modality and referring to previous problem solving data.

Likewise, the system's output has several user-friendly features. The present system generates a variety of responses in the form of animated facial expressions, lip and eye movements, synthesized speech, written texts, and other visual data.

## STRUCTURIZATION OF MULTIMEDIA INFORMATION

To create a user-centered multimodal interface, it is vitally important to structurize a vast amount of multimedia information using the afore-mentioned HI-ware so that it could be retrieved. For this purpose this HI-ware has the following two features: a standardized API (Application Programming Interface) whose consistency among media conversion functions enables them to be easily incorporated into various kinds of applications; and a common dictionary shared among media conversion functions, so that a newly registered word will be available to all media conversion.

It is equally important to note that this structurization is automated by the following steps. It first searches for the keywords as estimated from the titles and text provided by users by referring to office knowledge databases. Then it registers the processed information into the knowhow bases.

## KNOWLEDGE DATABASE: Linking Knowledge Bases to Knowhow Bases and Multi Databases

Large-scale KBs, powerful enough for various intelligent systems, have been built into a number of systems like CYC, but they all lack practicality for most applications in terms of cost-effectiveness of knowledge bases.

Our approach to knowledge sharing is entirely different in that its knowledge databases, which consist of KBs, knowhow bases and multi-databases, store "stock" information, as shown in Figure 1. That is, our knowledge databases are two-tiered. At the base is the established systematized data on inhouse activities; office procedures, job flows and documentation; and relations among persons and business groups. This data is reinforced by KBs derived from structured "knowhow" knowledge gathered from researchers, including a diverse range of useful information, such as advice on writing good technical

papers, computer operation, patent-related problems; and information on office security. It also links to other databases, such as archives of technical papers and employees lists.

Armed with these two types of databases, the system can recognize meaningful relations among the databases, allowing more flexible and accurate searches for business information than any other database systems and accelerate further knowledge sharing.

Figure 3 shows an example of information retrieval from knowledge databases. If a user asks "Which documents are needed to purchase equipment?", the system searches the KB and refers to the pair of the object "equipment purchase" and the property "has-document", then answers "Equipment plans, estimates and order forms." When the system is asked to give the user advice on writing equipment plans, but cannot find the properties of relations corresponding to "advice", then it searches the knowhow base and finds the information structured as: object: "equipment plan"; actions: "writing"; form: "advice".

### KIDS (Knowledge/Information on Demand System)

Using the same concept as the multimodal interface agent, we are developing a practical keyboard-based system called RDC-KIDS for everyday knowledge sharing at Toshiba Research and Development Center (RDC). Its screen layout is given in Figure 4. This Web system is capable of understanding users' request and provides such services as information retrieval, advice and suggestions, and whatever help they need, while directing a natural dialogue with them by means of keyboards. Moreover, the system has several convenient options, including registering new knowhow, making comments and evaluations on the knowhow used, and responding to unanswered questions directly from the administrator.

## CONCLUSION

We have proposed a new framework for multimodal interface agent and described two systems which use substantial knowledge databases. In order to stock knowledge and enhance its use and share, we linked deep-structured office KBs with surface-structured knowhow bases, while structurizing various types of databases by means of media conversion functions using the HI-ware. Through our experiment on about 1,000 employees at Toshiba RDC using the keyboard-based system RDC-KIDS, we have found that our system is highly efficient for enhancing knowledge sharing. We also succeeded in conducting a natural multimodal human-computer dialogue by natural language, speech, and gestures.

In the future, we envision a multimodal secretary agent system in Figure 5, which supports knowledge sharing and intergroup communication. For this end, we are expanding the existing KBs and developing more helpful functions on the basis of the user's estimated intention and situation.

## REFERENCES


1. Hammond, K.J., et al., "FAQ Finder: A Case-Based Approach to Knowledge Navigation", Proc.IJCAI-95, pp.2071-2072, 1995.
2. Miller, G.A.: "WordNet", A Lexical Database for English, CACM, 38, pp.39-41, 1995.
3. Iwazume, M., et al.,: "Ontology-Based Information Capturing from the Internet", Proc. 4th Int. Conf. On the Society of Knowledge Organization, pp.261-272, 1996.
4. Takebayashi, Y.: "Spontaneous Speech Dialogue System TOSBURG II -The User-Centered Multimodal Interface-", Systems and Computers in Japan, Vol.26, pp.77-91, 1996.
5. Takebayashi, Y., et al.: "A Multimodal Human Computer Interaction Research at Toshiba Research and Development Center", Proc.CHI97, pp.79-80, 1997.
6. Takebayashi, Y.: "Integration of Understanding and Synthesis Functions for Multimedia Interfaces", In Multimedia Interface Design, Blattner ed., pp.233-256, ACM Press Book, 1992
7. Kono, K., et al. : "A Multimodal Interpretation Method Based on Truth Maintenance", Interaction'97, pp.33-40, 1997.


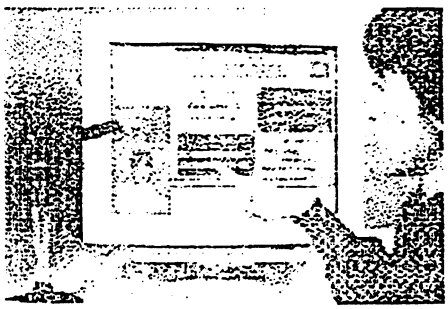

Fig.1: Interaction in speech and by finger pointing

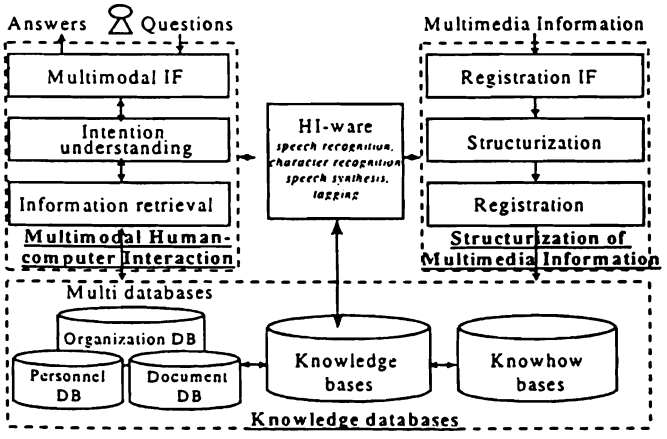


Fig.2: The system configuration

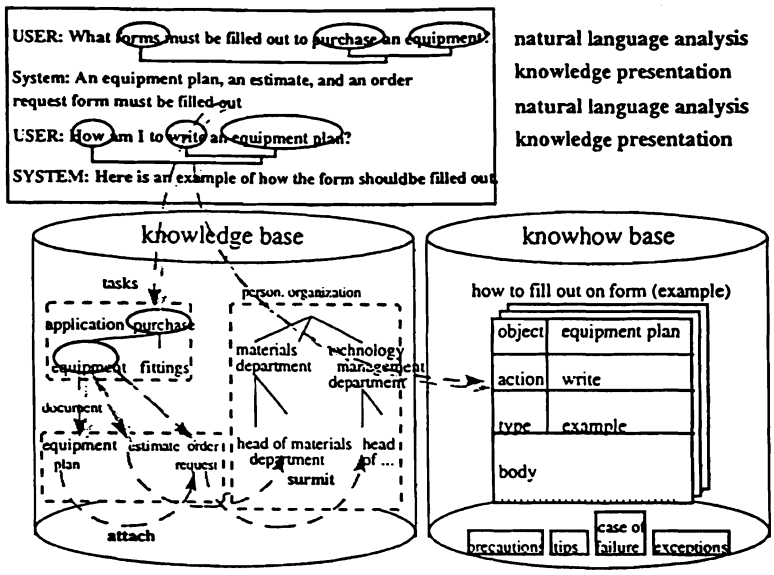


Fig.3: Content retrieval of knowledge databases

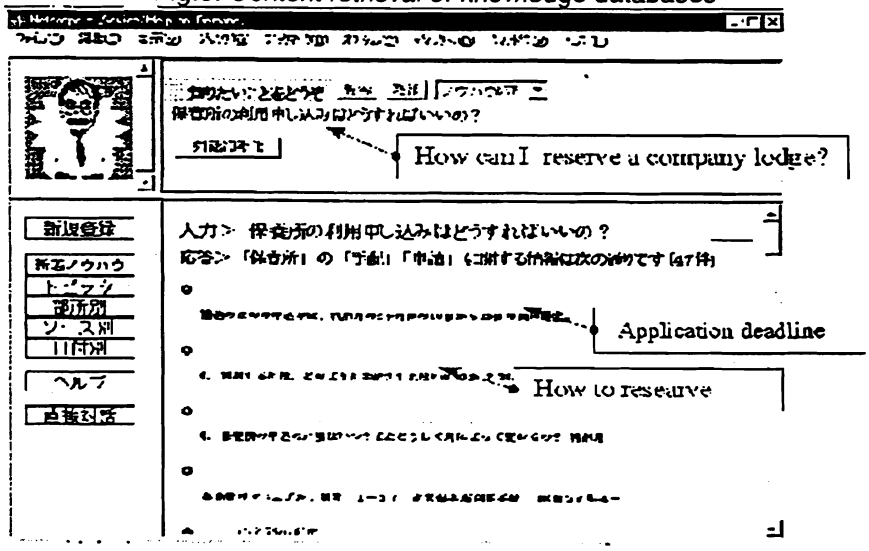


Fig.4: RDC-KID's screen image

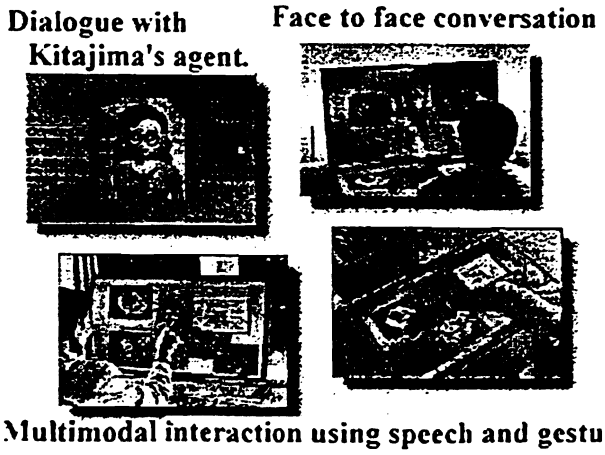


Fig:5: Multimodal Secretary Agent in 200X

# Workshop on
# Perceptual User Interfaces

# PUI '97

**Banff, Alberta, Canada**
**October 19-21, 1997**