On Facilitating Multimodal Multimedia Communications in Networked Environment
Ferguson, Innes; Davlouros, J.D.

National Research Council Canada   Conseil national de recherches Canada

Canada

On Facilitating Multimodal Multimedia Communications in Networked Environments

Innes A. Ferguson

Interactive Information Group, Institute for Information Technology,
National Research Council, Ottawa, ON, K1A 0R6

James D. Davlouros

Computing and Intelligent Systems, Canadian Space Agency,
6767 route de l'Aeroport, St-Hubert, PQ, J3Y 8Y9

Abstract

The PeopleFinder is an application of technology which aims to support a sense of social proximity among geographically distributed users. The PeopleFinder employs agent- and knowledge-based design techniques to assist users in determining the whereabouts of other staff located in an office or network environment. The tool makes use of several modes of input and output, as well as employing a number of interface and communications media with which to present information and interconnect both local and remote system users.

Introduction

The PeopleFinder is a knowledge-based tool to assist users in determining the whereabouts of other staff located in a departmental or office environment [Ferguson and Davlouros, 1995]. The tool makes use of several modes of input and output, as well as employing a number of interface and communications media with which to present information and interconnect geographically distributed system users. The PeopleFinder employs agent-based design techniques [Ferguson, 1992] as a way of integrating a variety of heterogeneous applications running on a number of different computing platforms and operating systems. Specifically, the PeopleFinder's functionality is carried out by a collection of coordinated robotic software agents which can:

(i) assume user-delegated tasks (e.g. interpret spoken commands, invoke email and voice recording tools, dial out on a telephone);

(ii) perform a number of application level tasks (e.g. update the graphical user interface, apply knowledge-based heuristics to decide whether and/or how system users want to be contacted); and

(iii) carry out a number of lower level operating system level activities, the majority of which will be beyond the level of expertise of most casual system users (e.g. translate between different speech encoding formats, establish which users are logged on to the various multi-platform computer networks throughout the building).

The remainder of this paper will provide a brief description of the main design features and functional capabilities of the PeopleFinder. In addition to relating the PeopleFinder to some existing multimodal and/or multiagent communications applications, some comments on preliminary evaluations of the tool as well as plans for future work on the tool will be given.

The PeopleFinder

The PeopleFinder is a system that aims to provide support for general awareness in a distributed work group. By collecting data from a number of on-line sources (e.g. login and active process statistics, electronic calendar data, telephony status information) and applying a series of heuristic rules that are designed to encapsulate typical work patterns of fellow users, the PeopleFinder attempts to inform users, as quickly and as accurately as possible, of the activities and whereabouts of other system users.

Some of the key design features of the PeopleFinder include:

- **Agent-based design**. The system is based on a number of coordinated autonomous agents, each specialized in one or more system tasks, a number of which have been mentioned above.

- **Multimodal interaction**. The system makes use of various input and output modalities for human-computer interaction: in particular, the keyboard, mouse, and speech for input; and audio and (visual) screen-based feedback for output. Figures 1 and 2, for example, illustrate some of the feed-
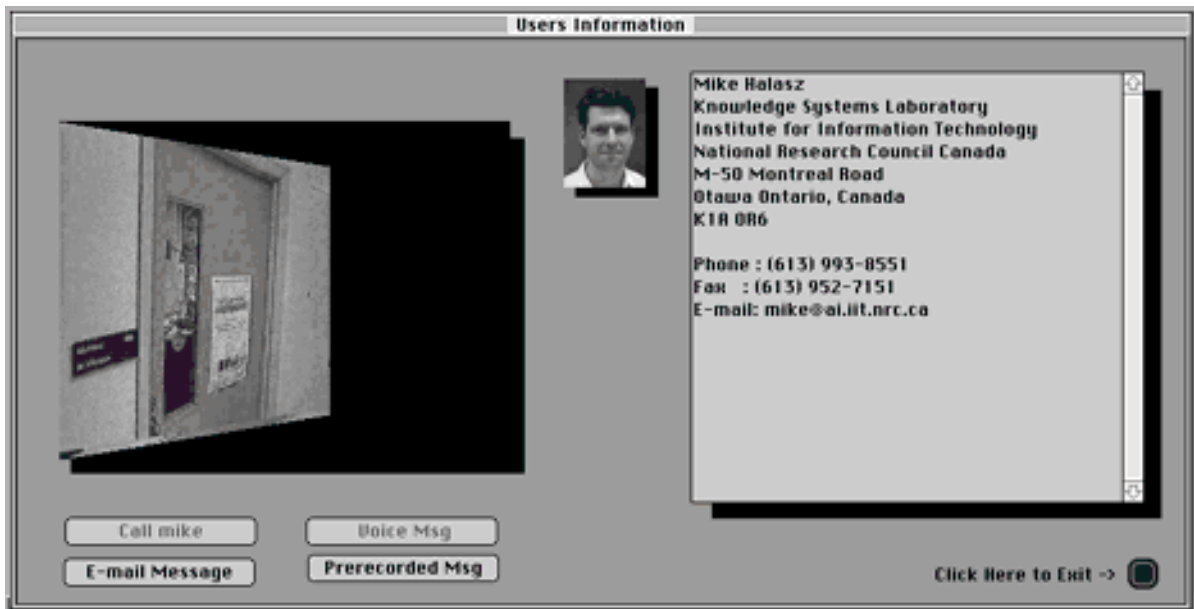
FIGURE 1. Example Users Information windows showing one user to be "idle" or willing to be contacted (a) and one user who is presumed to be out of the building (b). Appropriate means of communicating with the sought user are suggested in the bottom left-hand corner of each window. General contact information (e.g. information typically found in users' ".plan" files) is listed in the scrollable text field on the right-hand side of the window.
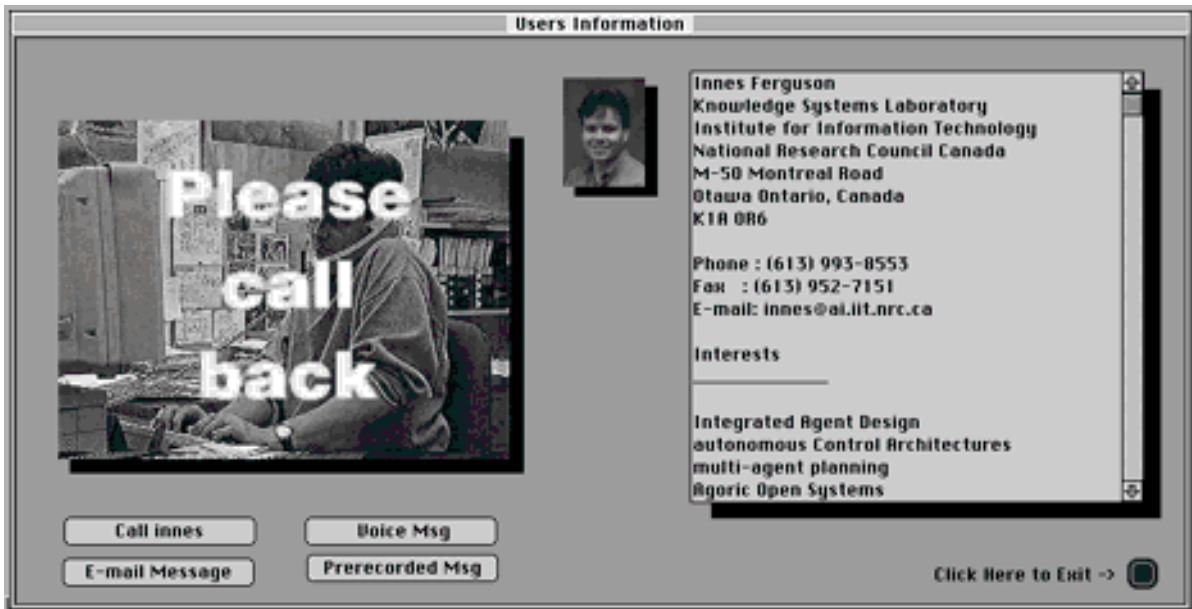
(a)
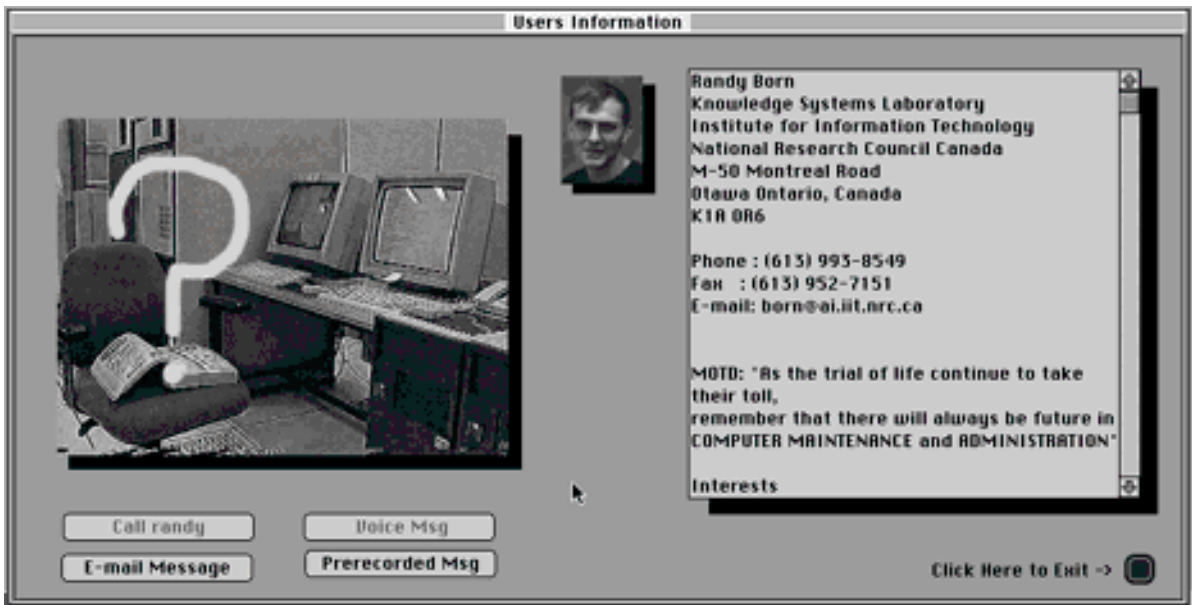
(b)

FIGURE 2. More Users Information windows showing one user to be busy on the phone (a) and yet another user to be out of his office, but presumably still somewhere in the building (b). In the former case, the remote user's status is determined by placing a telephone call and monitoring the local modem for a received busy signal; in the latter case, the remote user's status is determined heuristically on the basis of elapsed idle computing time.

FIGURE 3. PeopleFinder: System Level View.

FIGURE 4. CALVIN: Control Framework.

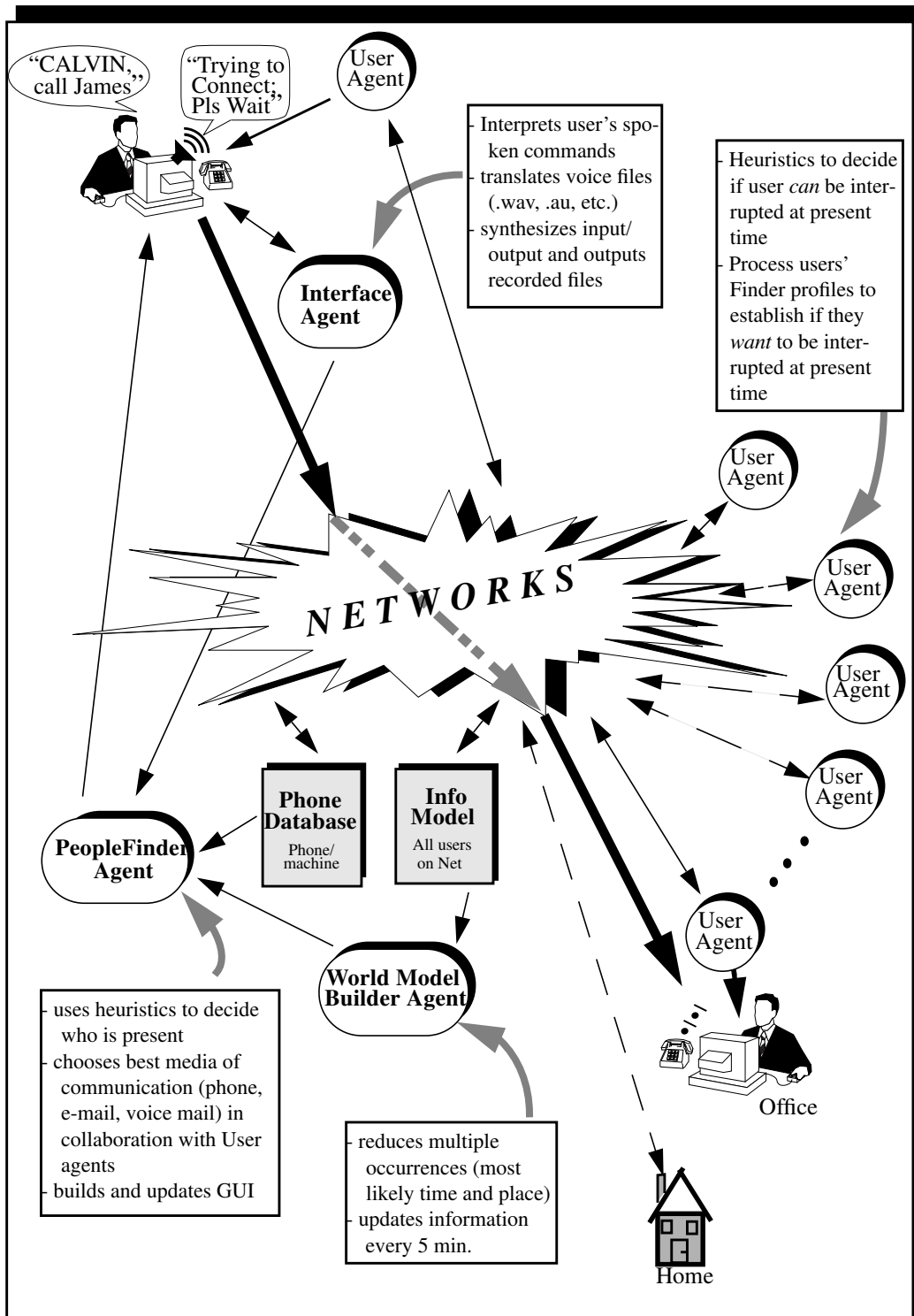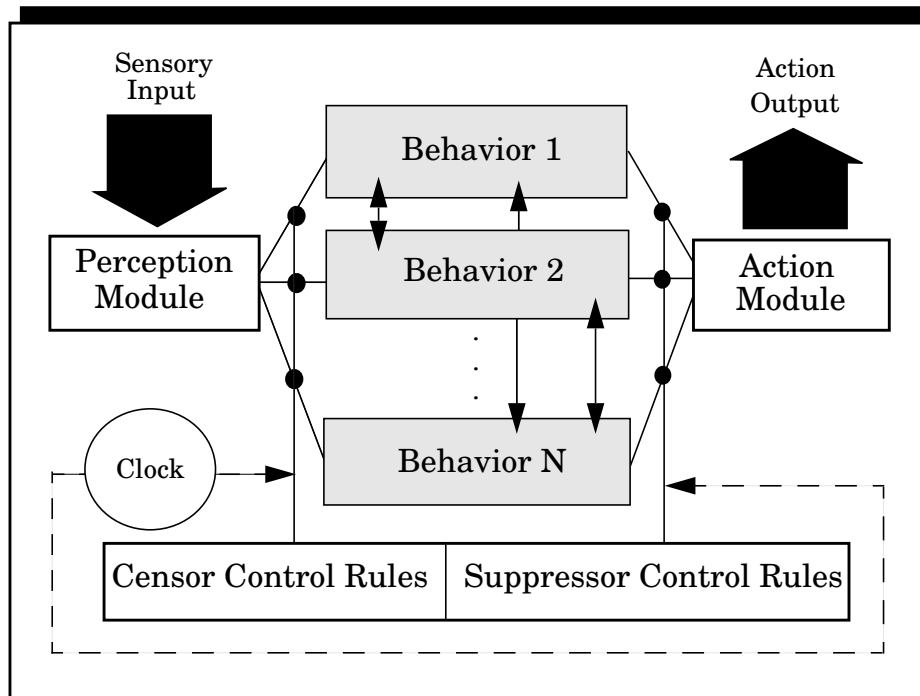which aims to support a sense of social proximity among geographically distributed users. Besides supporting presentation of diverse sources of multimedia information, the PeopleFinder emphasizes (and exploits) the use of different human sensory modalities and channels of communication in order to maximize its potential usefulness. A primary goal in designing and deploying the tool is to provide a computational system which closely matches the skills and work practices of its intended users. Early evaluations of the tool suggest the present work is heading in the right direction.

Future work already underway includes porting the graphical user interface portion of the PeopleFinder to run on both PC and Unix platforms (in the interest of extending the tool's audience and ensuring a more thorough testing and empirical evaluation phase of the project[††]); integrating a number of software applications such as teleconferencing, voice dictation, and video camera-based face recognition; extending agents' capabilities for autonomously resolving runtime conflicts resulting from shared access to the different presentation and communications resources used by the system (see Werkman's BN negotiation-based conflict resolution work for related issues [Wer94]); adding privacy features so that individual users may control how much information regarding their activities and whereabouts may be accessed through the PeopleFinder; and formalizing the various rules used by the PeopleFinder to combine multiple media with multiple modalities for both human-computer interaction and user-user communication (much along the lines of the work of Arens *et al.* [1993] on allocating multiple media).

References

Arens, Y., Hovy, E. and Vossers, M., [1993] On the Knowledge Underlying Multimedia Presentations. In Mark T. Maybury, editor, *Intelligent Multimedia Interfaces*. AAAI Press: Menlo Park, CA.

Buxton, W., [1994] The Three Mirrors of Interaction: A Holistic Approach to User Interfaces. In L.W. MacDonald and J. Vince, editors, *Interacting with Virtual Environments*. New York: Wiley.

Cohen, P.R., Cheyer, A., Wang, M. and Baeg, S.C. [1994] An Open Agent Architecture. In *Working Notes of the AAAI-94 Spring Symposium on Software Agents*, Palo Alto, CA, pp. 1–8.

Faure, C. and Julia, L. [1994] An Agent-Based Architecture for a Multimodal Interface. In *Working Notes of the AAAI-94 Spring Symposium on*

[††]In fact, we have now made the PeopleFinder available on the WWW. In such an environment, users are able to regard other remote users as inhabiting a (very large!) virtual hallway or office spanning several physical offices and organizations.

Intelligent Multi-Media Multi-Modal Systems, Palo Alto, CA, pp. 82–86.

Ferguson, I.A. [1992] TouringMachines: An Architecture for Dynamic, Rational, Mobile Agents. Ph.D. diss., Computer Laboratory, University of Cambridge, Cambridge UK.

Ferguson, I.A. [1995] Integrating Models and Behaviors in Autonomous Agents: Some Lessons Learned on Action Control. In *Working Notes of the AAAI-95 Spring Symposium on Lessons Learned from Implemented Software Architectures for Physical Agents*, Palo Alto, CA, March 27–29.

Ferguson, I.A. and Davlouros, J.D. [1995] PeopleFinder: a Multimodal Multimedia Communications Tool for Interconnecting Office Staff. In *Proceedings International Joint Conference on Artificial Intelligence*, Montreal, PQ, August 20-25. pp. 2059-2060.

Finin, T., McKay, D. and Fritzson, R. [1992] An Overview of KQML: A Knowledge Query and Manipulation Language. Available through the Stanford University Computer Science Department, Palo Alto, CA, March.

Giachino, L. [1993] Activity Sensing Through Portholes Images: A Bridge between Passive Awareness and Active Awareness, Technical Report, OTP-93-08, Ontario Telepresence Project, University of Toronto, Toronto, ON, August.

Want, R., Hopper, A., Falcao, V. and Gibbons, J. [1992] The Active Badge Location System, *ACM Transactions on Information Systems*, 10(1):91–102.

Weiser, M. [1993] Some Computer Science Issues in Ubiquitous Computing, *Communications of the ACM*, 36(7):75–83.

Werkman, K.J. [1994] A DAI Architecture for Coordinating Multimedia Applications. In *Working Notes of the AAAI-94 Spring Symposium on Intelligent Multi-Media Multi-Modal Systems*, Palo Alto, CA, pp. 93–97.

Acknowledgments

back users receive when attempting to contact other network users.

- **Multimedia presentation**. The system also makes use of a number of different media for human-computer interaction; in particular text, graphics, animation, and prerecorded video (the latter acting as a "cheap", but nevertheless very effective, substitute for in-office video cameras and/or ubiquitous computing devices such as active badges[*]).

- **Communications-oriented capabilities**. The system enables transparent communication across different computer platforms (Macintosh, PC, Unix) and facilitates the interconnection of system users via telephone, email, and voice messaging (see Figure 3).

The robotic software agents used in the PeopleFinder application are based on the *CALVIN* open agent framework (see Figure 4) which in turn is an extension of the TouringMachine architecture [Ferguson, 1992]. This framework provides application developers with a powerful set of agent programming tools including libraries of intra- and inter-agent protocols (e.g. KQML[†]), sensory and effectory apparatus, internal behavior APIs, persistent storage management, and (currently under consideration) CORBA[‡] compliance.

The tool is implemented using a variety of different scripting languages (AppleScript, Quickeys, and C-shell) and runs on a Macintosh Quadra 840 AV. The tool also makes use of Quicktime movies for displaying prerecorded video as well as the Macintosh's Apple Phone tool and Geoport Telecom Adapter for performing its various computer-telephony integration tasks.

Evaluation

A number of features of the *CALVIN*[**] architecture have proven useful for developing multimodal applications that integrate a number of distributed

---

[*]Such as those used in various in-house applications at Xerox's Palo Alto Research Center [Want *et al.*, 1992; Weiser, 1993].

[†]The Knowledge Query and Manipulation Language (KQML) is a protocol intended to support interoperability among intelligent agents in distributed applications [Finin e*t al.*, 1992; Werkman, 1994].

[‡]The Object Management Group's Common Object Request Broker Architecture (CORBA).

[**]The *C*ommunicating *A*gents *L*iving *V*icariously *I*n *N*etworks (*CALVIN*) architecture is an agent framework developed at the National Research Council's Knowledge Systems Laboratory (see WWW page http://ai.iit.nrc.ca/software.html for more details).

media resources. In particular, rapid responses to users' commands are facilitated through integration of appropriate reactive behaviors in the system's Interface and User agents (see Figure 3); in addition, blending of complementary input modalities is facilitated through the execution of multiple concurrent agents (which in turn are able to execute multiple concurrent, task-specific behaviors).

Early evaluations of the PeopleFinder application are generally quite positive. Feedback from users of the tool within our own laboratory (which comprises between 25 to 30 staff) has enabled us to identify a list of the most popular capabilities of the tool. These include the ability to integrate a number of commonly used tools and devices (e.g. electronic mail, telephone, and "finger"-like OS-level commands); the aesthetic use of multiple media (in particular, the prerecoded video); and the ability to offer multi-sensory and multi-channel means of interaction between human and computer. In terms of negative user feedback, this was mostly directed at the less than satisfactory quality of the Macintosh Quadra 840 AV speech processing (automatic recognition and text-to-speech generation) capabilities.

Related Work

Besides permitting the relatively seamless integration of both non-deliberative (e.g. reactive) and deliberative (e.g. planned) behaviors, as well as providing an added level of flexibility and robustness to the overall system performance [Ferguson, 1995], the existence of multiple autonomous — and, therefore, concurrently operating — agents also facilitates the application of multiple channels of interaction between user and system [Faure and Julia, 1994]. A number of similar benefits and issues have been identified in the application of the Open Agent Architecture to the management of email; in particular, the roles of such agent skills as *delegation* ("the ability to receive a task to be performed without the user's having to state all the details"), *data-directed execution* ("the ability to monitor local or remote events, such as database updates, OS, or network activities"), and *communication* ("the ability to enlist other agents... in order to accomplish a task") [Cohen *et al.*, 1994, page 1]. In addition, the PeopleFinder can be seen to offer a number of desirable features which are characteristic of technology for *telepresence*; that is, technology which provides, despite geographical or temporal distance, a sense of social proximity — a kind of *social prosthesis* for overcoming gaps and weaknesses with an organizational structure [Giachino, 1993; Buxton, 1994].

Summary

The PeopleFinder is an application of technology