Turning pervasive computing into mediated spaces

by W. Mark

With pervasive computing, we envision a future in which computation becomes part of the environment. The computer forms (workstation, personal computer, personal digital assistant, game player) through which we now relate to computation will occupy only a small niche in this new computational world. Our relationship to pervasive computing will differ radically from our current relationship they have with computers. When computation becomes part of the environment, most human-computer interaction will be implicit, and it will have to take account of physical space. Physical space rarely matters in current human-computer interaction; but as computational devices become part of furniture, walls, and clothing, physical space becomes a necessary consideration. First, more than one person can occupy a space. Second, individuals within the space are doing things other than interacting with the computer: coming and going, and perhaps most strikingly, interacting with each other—not just with the computer. Finally, physical space provides a sense of place: individuals associate places with events and recurrent activities.

The emerging relationship between people and pervasive computation is sometimes idealized as a "smart space": the seamless integration of people, computation, and physical reality. This paper focuses on a particular kind of smart space, the "mediated space," in which the space understands and participates in multiperson interaction. Mediated spaces will expand human capability by providing information management within a context associated with that space. The context will be created by recording interaction within the space and by importing information from the outside. Individuals will interact with the space explicitly in order to retrieve and analyze the information it contains, and implicitly by adding to the context through their speech and gesture. Achieving the vision of mediated spaces will require progress in both behind-the-scenes technology (how devices coordinate their activities) and at-the-interface technology (how the space presents itself to people, and how the space deals with multiperson interaction). This paper explores the research challenges in both of these areas, examining the behind-the-scenes requirements of device or manifestation description and context maintenance, as well as the interface problems of metaphor and understanding natural human-to-human spoken interaction.

The pervasive computing revolution will surely occur: computation will be embodied in things, not computers. We can already put computation almost anywhere. Embedded computation controls braking and acceleration in our cars, defines the capability of medical instruments, and runs virtually all machinery. Hand-held devices (especially cell phones and pagers) are commonplace; serious computational wristwatches and other wearables are becoming practical; computational furniture and rooms are demonstrable. Relentless progress in semiconductor technology, low-power design, and wireless technology will make embedded computation less and less obtrusive. Computation is ready to disappear into the environment.

But what will it all mean? The nature of our relationship to computation in its pervasive form will nec-

0018-8670/99/$5.00 © 1999 IBM

essarily be different from our relationship to computation in its current form. The first key difference is the explicitness of the computational task. Presently people think in terms of performing explicit tasks "on the computer"—creating documents, sending e-mail, and so on. When computation is part of the environment, this comfortable explicitness will disappear. Individuals will do whatever they normally do: move around, use objects, see and talk to each other. The computation in the environment may be able to facilitate these actions, and individuals may come to expect certain services, but they will usually not be doing things "on the computer."

Implicit computation will be available everywhere; we need to figure out how to interact with it.

We see the beginnings of this form of interaction with existing embedded computers. For example, an automatic braking system engages when the driver performs the normal action of pushing the brake pedal. The "automatic" is significant: the computation is implicit—braking simply works better (most of the time) and we do not care how. Currently this form of interaction is extremely limited. We allow it only when our intent is unambiguous and when the computer can clearly do the job better than we can. In order to take advantage of pervasive computing, we must be able to greatly expand this form of interaction. Implicit computation will be available everywhere; we need to figure out how to interact with it.

A second key difference in the pervasive computing world is the importance of physical space. Current computers obviously occupy physical space, but this is usually irrelevant. Apart from dealing with limitations of "screen real estate" and ergonomic considerations of head and hand positioning, most computer interface design has nothing to do with physical space. With very rare exceptions, conventional computer interfaces are unaware of the presence, much less the identity, of human beings.

When computation is part of the environment, it will be part of everyday physical space. This single shift radically changes the relationship between humans and computation—from a fairly static single-user location-independent world to a dynamic multiperson situated environment. First, pervasive computation environments are necessarily dynamic with respect to their human users. Individuals move around in space, changing position and visual focus, coming and going. Second, more than one person can occupy a space. When more than one person is in a space, they tend to interact with each other. Finally, the physical location of the computation—or more precisely the interface to the computation—becomes relevant. Computer users are currently encouraged to disassociate computation from location: information is available from any tap; "the network is the computer." While this is a valuable viewpoint that will certainly continue in the pervasive computing world, it is based on the separation between computers and real things. A computer is an artificial entity; it does not matter very much where it is, especially in a networked world. This is very different from a computational desk or conference room table, where the interface is part of a specific spatial environment that has other attributes and associations. Individuals associate places with events ("you were sitting right there when I told you that") and recurrent activities (the conference room, my office, my favorite store for children's clothing).

Smart spaces

The relationship between people and pervasive computation that *ought* to come into being is a seamless integration of people, computation, and physical reality: a "smart space." The concept of a smart space has a long history in computer science. In the early 1960s Doug Engelbart at Stanford Research Institute (now SRI International) was exploring the concepts of human-computer systems that could augment human capability, especially humans working in groups.[1] Although this work is most famous for the mouse interface, its primary contribution is probably the "smart space" vision that still informs the research community. The "Media Room" project developed by the MIT (Massachusetts Institute of Technology) Architecture Machine Group in the mid-1970s[2] explored the concept of users interacting with room-sized computational environments. The result was a new human-computer interface based on the combination of speech and gesture input, and text and graphics output.

A decade later, Mark Weiser and his colleagues at Xerox PARC (Palo Alto Research Center)[3] were investigating a different paradigm in which spaces consist of invisibly computational objects, objects that embody computational extensions of their originals (smart Post-it** notes, badges, pads, etc.). People perform tasks primarily through interaction with these smart devices as they move through their day. Unlike the Media Room concept, explicit interaction with the computer is meant to be minimal to nonexistent. In the "ubiquitous computing" world envisioned by Weiser, people interact with computational entities pretty much the way they interact with physical entities, not the way they interact with other people.

As the technology of pervasive computing has improved, research in this area has flourished, producing significant interactive environments based on these earlier concepts. For example, Michael Coen's "Intelligent Room"[4] and more recent "Hal"[5] are intellectual descendants of the Media Room; both are conference rooms that track the people in them and understand commands as combinations of speech and gesture input. The goal is to expand the boundaries of human-computer interaction, moving toward human-human or even human-superhuman interaction patterns between people and computers. The MIT Media Lab's "Things That Think" project[6] shares much of the heritage of ubiquitous computing. The "Smart Rooms" of Sandy Pentland and colleagues[7] have a similar point of departure, but additionally create complex three-dimensional information environments that make use of human spatial reasoning capability. Smart Rooms, and Pentland and Liu's "Smart Car,"[8] also focus on inferring human intentions through their actions in order to provide enhanced interaction. The "Tangible Bits" approach of Hiroshi Ishii et al.[9] pushes still further on the boundaries of human-computer interaction, following the ubiquitous computing paradigm into physical and ambient interfaces.

These smart-space approaches are about humans dealing directly (even if implicitly) with computers to accomplish tasks. They mostly ignore the interpersonal interactions of people in the space, e.g., "in general, the room ignores spoken utterances from the lapel microphones not specifically directed to it."[4] This is an important simplifying assumption that makes implementation tractable, but it also defines the smartness of the space in terms of human-computer interaction: the capability of the space to understand what people are trying to tell it.

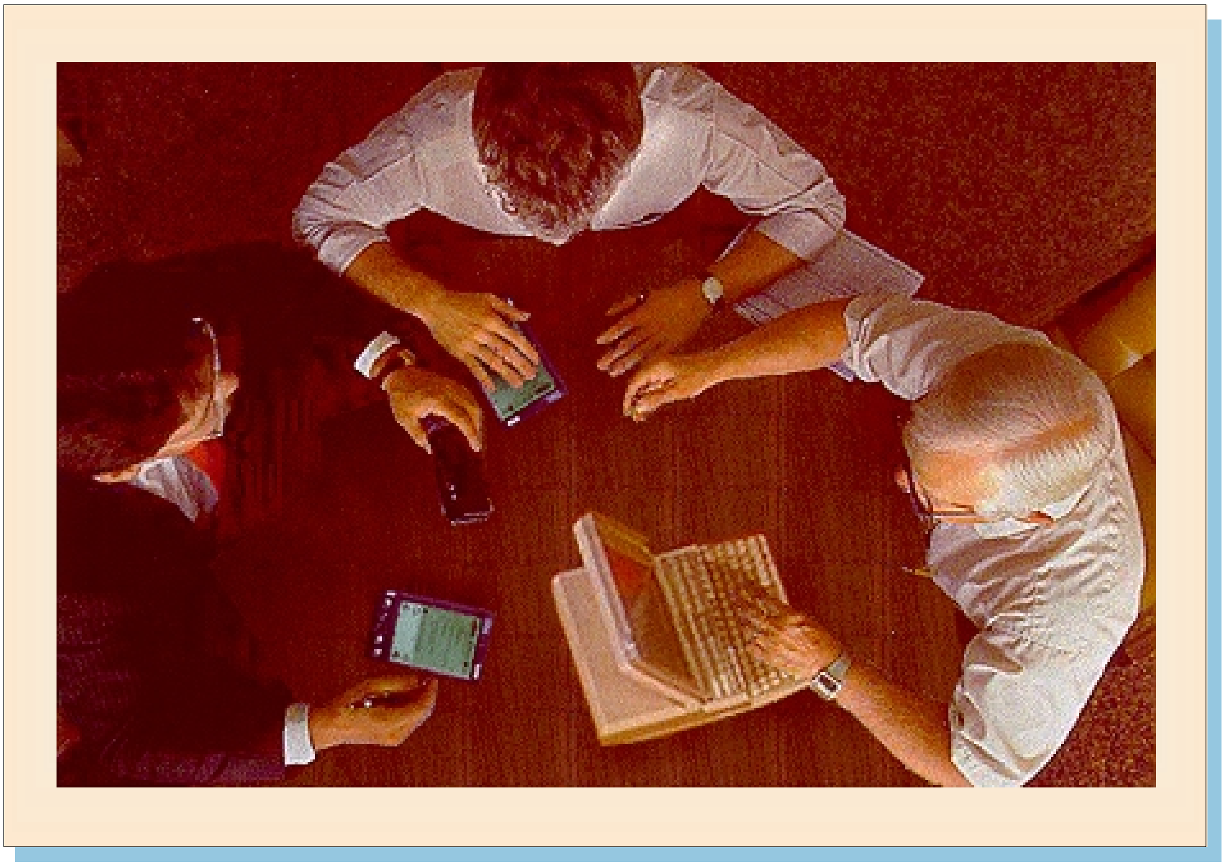
The focus here is different. Individuals in smart spaces will interact with *each other*, not just with, or even primarily with, the space.[10] I use the term "mediated space" to refer to smart spaces in which interpersonal interaction is the primary activity of the people in the space.

The mediated spaces vision

Mediated spaces will enhance human activities by enhancing the interaction of people in the space. For example, almost all complex artifacts are designed by groups, and group design requires group interaction. A design session is a meeting of a team to work on a design. Usually it is an informal interaction in which design problems and solutions are raised and discussed, often resulting in a decision. The goal is to extend or correct the group's shared representation of the design. This shared representation is partly on paper, partly in the minds of the participants, and, increasingly, partly in computer software. But even with digital desks and smart offices, even with the great advances in computer-aided design software and the computer form factors that allow it to play a role in design sessions, much of the information in a design session is lost. The interpersonal interaction that takes place during the session does not make it into the updated design representation. This information, including the rationale for decisions and the alternatives that were discussed, is only incompletely and briefly remembered by the participants and is not available at all to nonparticipants.

In a mediated design session, the computational environment incorporates the information contained in the speech and gestures of the participants into the shared design representation. The mediated space in this case is a smart conference room like the Intelligent Room. The room's paraphernalia might include a smart whiteboard, people tracking cameras and microphones, and so on.[3,4] Designers may enter the room with personal notepads or other computational devices that contain information about the design (see mock-up in Figure 1). As designers discuss and argue about the design, they point to images on paper or embodied in the computer-aided design (CAD) software of their notepads, workstations, or other displays. The difference is that the mediated space is focused on the participants' interaction with one another, not just with the devices in the space.

Figure 1 Design session in a mediated space



The role of the mediated space is to incorporate interpersonal interaction into the design representation. The goal is not to simply record everything that was said,[11] but to represent relevant information in relevant places in the design representation. The space takes a proactive role, suggesting relevant information from outside the session, including other designs it knows, along with ancillary information from the company's intranet and the Web. It proactively detects inconsistencies with both the current design ("that conflicts with an earlier decision") and the other designs it knows ("these similar designs had a higher power budget").[12,13] The result is an enhanced design experience, guided by the space's understanding of the current and previous sessions, and an enhanced design representation that incorporates rationale and alternatives.

Another mediated space example is the classroom-and-home educational environment (Figure 2). Classrooms have an interesting characteristic: generations of students learn much the same material in them year after year. Year after year teachers find ways to address the highly individual insights and learning gaps of their students. Mediated spaces offer an opportunity to enhance the learning experience by expanding access to the interaction of students with each other and with teachers. Again, this must go beyond recording and playback. The space needs to be proactive in suggesting relevant information and pedagogical approaches based on previous experiences in this and other classrooms, on information from the Web, etc.[14,15] A particular student interaction pattern cues the space to guide the teacher to pedagogical scenarios that worked in similar situations. Students can "look ahead" or review learning interactions from their class or others.

When school is over, the students go home with their individual notepads, turning the dinner table (for at

Figure 2 Mediated classroom-and-home educational environment

least part of the time) into an extended mediated classroom. The notepad can interact with other devices in the home, or simply use its own audiovisual capabilities. The usual dinner table discussion of "what we learned today" is enhanced by the context of the student's interactions during the day, perhaps annotated by the teacher during class. Parents' and siblings' comments are integrated into the lesson plan and brought back to school to be shared with others.

Mediated spaces, like any smart spaces, can consist of multiple physical environments. They can be preplanned mediation rooms or impromptu settings, like the mealtime table. The main point is that the space must deal with interpersonal interaction in order to provide benefit. People may interact directly with the space, but usually they are sharing information with and learning from each other. The space can enhance the experience, but interacting with it is not the goal of the participants.

This is a lofty vision. Mediated spaces must understand enough of speech, gesture, and personal device intercommunication to update representations and provide useful proactive information and consistency checking. The space must identify and track speakers so that it will know who is saying what to whom, identify references to objects and images in the room, and detect and follow topics in an ongoing conversation.

The technology required to implement mediated spaces can be divided into two categories:

- *Behind the scenes* (how the devices in the space coordinate their activities to support mediated interaction). A mediated space is a collection of computational devices—a dynamic collection due to the movement of people and things. Putting computational devices into a space does not make it a mediated space. Mediation is mostly about communication and coordination. To achieve the mediated space vision, these devices must do a great deal of work behind the scenes when the space is created and continuously during operation of the space.
- *At the interface* (how the space presents itself to people, and how the space understands human interaction). Participants require a way to think about and form expectations of a smart space: an interaction metaphor. Within the metaphor, people need to use the devices in the space without being explicitly aware of them. The space, in turn, must have the ability to understand and use enough of the multiperson interaction, particularly the spoken dialog, to enhance the participants' experience.

This paper explores some of the challenges in both of these categories.

Behind the scenes

Behind-the-scenes technology for mediated spaces addresses the issues of how the computational aspects of the space (the devices and the computations they perform) collectively support multiperson interaction. First, mediated spaces must maintain a description of the devices and their manifestations. The interaction of individuals in the space to some extent involves these devices and their input/output manifestations, e.g., a diagram displayed on a board by CAD software. Second, as their primary output, mediated spaces must build and continuously update a context that integrates what individuals are talking about and doing in the space with whatever computational representations exist in the devices of the space.

Device and manifestation description. Mediated spaces are *situated* collections of devices, which means that the role of any device depends not only on its own characteristics, but also on its situation. The situation of a device, in turn, depends on the individuals and the other devices in the space at the time, and the tasks the individuals are performing. The set of individuals and devices may change, either because devices enter or leave the space, or because devices within the space go in and out of operational readiness. Even the concept of space is situated, i.e., the boundaries of the mediated space may change because it is appropriate to use more time or power to interact with a larger set of devices in some situations.

The space must understand its devices and their situation, in particular the computations they are performing, in order to understand what individuals in the space are saying. One of the most powerful features of human communication is the ability to refer to objects and events without explicitly naming them. A mediated space attempting to understand human interaction must be able to interpret referring expressions, including pronouns and definite descriptions (e.g., "the hot junction"). This is a much studied topic[16] that goes far beyond mediated spaces. But mediated spaces do introduce new forms of reference and must provide the infrastructure for interpreting them. Devices and computations have to be described as real-world physical objects and concepts in order to support references.

One important class of references is based on location and the physicality of space. Mediated spaces must have descriptions of devices and their input/output manifestations that support such physical references. The space needs to know which devices are inside or outside the space and which displayed information is in the direction the speaker is pointing to, or to the left or right. References of the form "every display in this space" or "everything in use in the kitchen" need to be resolved into the identifiers for a particular set of devices.

Currently, the only location description of most computational devices is their network address. The network-address reference scheme is hierarchical, based on abstract concepts like "domain." In this scheme, computers can be addressed as groups, but these groups may or may not be associated with physical spaces. The latest version of the Internet Protocol[17] provides the capability for an almost unlimited number of unique network addresses, but having a name space is not the same thing as having a reference scheme. The Internet community is also working on reference schemes to handle some aspects of changing location.[18,19] Again, these are changes in virtual location. Internet-style addressing is entirely appropriate—even advantageous—for the virtual world of networks. From a mediated space perspective, these protocols address the issues that arise when moving from one mediated space to another, e.g., the need to disconnect from one network host and reconnect to another. They do not address the issues of changes in actual physical location that arise when individuals move within a space, change the direction of their gaze, or come and go with their personal devices.

References go beyond location and dimension, requiring the space to have other knowledge of the properties of physical objects. Some references depend on characteristics like physical possession, as in "Tom's notepad." Other references depend on the device and its manifestations, e.g., "the board that is showing the map." Still others will be references to just the manifestations. For example, a person in a mediated design session may circle part of a drawing on his or her notepad, gesture toward a wall display and say "this part of the design is critical to the heat dissipation at this point." The speech recognizer needs to resolve the referent of the phrase "this part of the design" by looking for a design object that is salient for the speaker.[20] Something in the mediated space must know that salience can be indicated by notebook gestures. There must also be knowledge that the circled drawing represents a design object. Resolution of "at this point" requires similar knowledge.

This kind of interaction can work only if the space has appropriate descriptions of devices and their manifestations. Mediated spaces must represent devices as real-world objects that come and go, move around, and change possession. Device manifestations must be represented in terms of real-world images, sounds, or touch. The beginnings of what is necessary for a descriptive framework can be found in efforts like the Open Agent Architecture (OAA) work at SRI.[21,22] Each device is represented by an agent (or set of agents) in the system. The agent's capabilities, including the manifestations it can produce, are described in Inter-agent Communication Language (ICL).[21] User interactions are interpreted by the cooperation of multiple, autonomous agents.

OAA has been used in several flexible multimodal, multidevice systems, for example in a multimodal map application in which users interact with maps via speech, pen strokes, and keyboard.[22] The capability of each device and system software component is described as an agent in ICL and registered with the system. When the user speaks an utterance like "show photo of the hotel," the ICL descriptions are used to (simultaneously) trigger relevant agents to handle the user's interaction. A natural-language agent provides a list of the most recently mentioned hotels; device agents report relevant pen gestures (circling or pointing to a hotel icon); a user-interface agent reports which hotels are currently being displayed. Higher-level meta-agents then adjudicate the response (e.g., preferring a circled icon as the referent of "the hotel").

An intriguing mediated space device/ manifestation description problem is the combination of new kinds of physical manipulation with verbal reference. Smart spaces introduce the capability of touching, moving, or otherwise manipulating smart objects in order to express intent—indeed this is the thrust of much of the previously cited human-computer interface work by Pentland[7] and especially Ishii and Ullmer.[9] This is quite different from gesture, a human-human interaction that is simply being recognized in the space. Here people manipulate objects in computationally meaningful ways. For example, moving a *phicon* (physical icon) on a map in the Tangible Geospace system[9] leads to changes in the information displayed. This kind of manipulation is a new kind of manifestation that must be described to the mediated space behind the scenes so that it can become part of the discourse, just like an utterance or gesture.

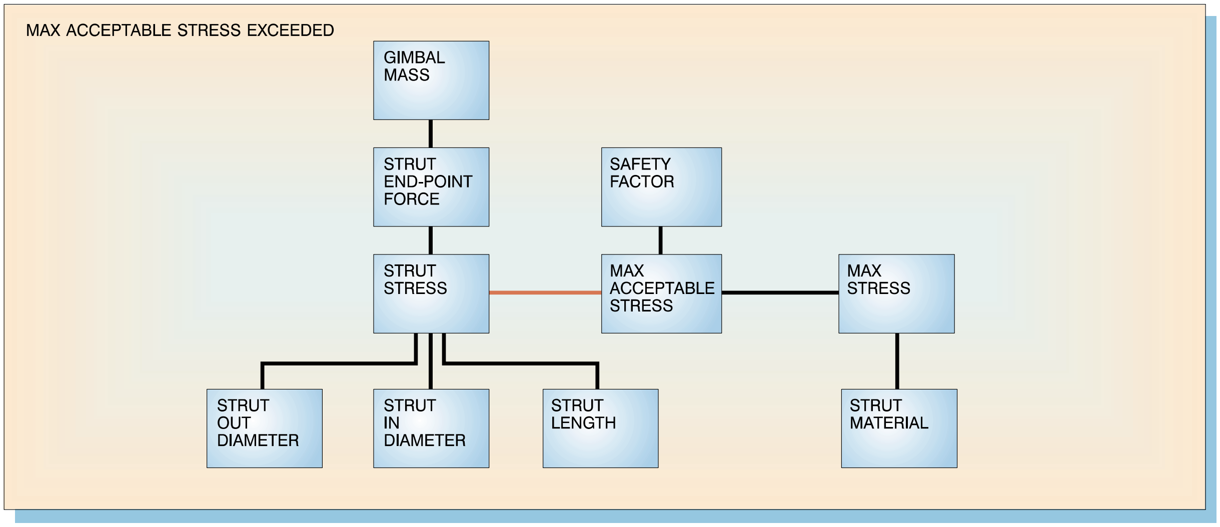
Context creation and maintenance. Context is the representation of the information that is relevant to the individuals and devices within the space. The context must be a *composition* of relevant information: a mere collection of information is of lower value. One reason is ease of access. A context that represents just the recorded interaction in the space is of lower value than one that represents indexed recorded interaction, which is in turn of lower value than one that represents summarized interaction organized into a structure. In a design session, the interaction of the participants contains a great deal of useful information, but watching and listening to it after the fact would be tedious indeed—designs can go on for months or years. Better would be the ability to go directly to specific segments of recorded material. The more flexible the access mechanism (e.g., by topic rather than by specific words) the better.[11] Better still would be the ability to access a summary of the interaction: a coherent representation of the design as a structure of decisions and their rationale.

Another reason that the context must be a composition of relevant information is that it must guide the computational understanding of further interaction in the space. Like humans, computational reasoning systems cannot understand things out of context. They need a representation of what has gone before in order to interpret utterances and gesture. The representation must encode information in a way that supports inferential reasoning: describing the meaning of the words and gestures in symbolic form, and statistical reasoning: describing the information in terms of features that can be used to discover and predict patterns.[23]

Finally, the context must be a composite of the relevant information, so that it can be used to guide mediated-space participants in performing their tasks. In the mediated design session vision described earlier, the space helps designers by finding relevant information outside of the space, looking for conflicts with earlier decisions, and so on. This requires a representation of the interaction in the space that can be related to the representation in the CAD software (or in the case of a classroom, to the lesson plan).

Examples of this kind of contextual representation can be found in systems like Cosmos[12] and PACT,[13] which provided design support based on reasoning about symbolic representations linked to CAD software. These systems supported multidesigner interaction by informing design team members of the im-

Figure 3 A Cosmos context visualization. Nodes show design elements represented in CAD tools. Lines between nodes show constraints. The red line shows a constraint violation. The diagram is a visualization of the design factors that have bearing on the maximum acceptable stress for the strut.



pact of decisions being made by other team members. Determining the impact required analysis by multiple underlying CAD systems, reasoning in terms of known constraints, and propagation of information depending on known responsibilities of the participants. For example, in the mechanical design domain of Cosmos, a designer might introduce a change in the design of a device to give it a faster reaction speed. This change would be accomplished in terms of the designer's CAD system (i.e., the designer would alter a diagram, change parameters, etc.). This change would trigger qualitative reasoning using Cosmos's model of the design domain, in this case a model of how forces affect objects of different shapes and materials. Based on the results of this reasoning, Cosmos would provide feedback to the original designer, and to other designers whose work would be affected. A key feature of the Cosmos system is that the information fed back to the original designer and shared with others is a visual representation of the context that shows the impact of the proposed decision on the design—not just a statement of the decision (see Figure 3).

Neither Cosmos nor PACT supported direct multiperson interaction in the sense of a mediated space (they were both based on designers interacting only through their CAD systems). But these systems do illustrate the kind of contextual framework that is required for mediated spaces. It is a composite representation that encodes the current and relevant past states of the design in a form that enables the system to communicate with designers in terms of the actual design. Most important for us here, it is a framework that can organize the space's interpretation of human-human interaction within the space. For example, the space can reason about how information from a dialog about maximum stress relates to other information about the design.

This contextual framework clearly depends on the application domain of the space. Again, like humans, mediated spaces will not be able to understand very much of a conversation unless they start from a prebuilt contextual framework. This is not to say that each space will have or can have its own idiosyncratic conceptual framework. It is imperative that mediated-space frameworks be based on common ontologies.[24] If every mediated space and every task within the space comes with its own *ad hoc* content framework, mediated interaction would be nearly impossible within a space, much less among different spaces.

Mediated spaces must therefore be based on context representations that support reasoning about the human-human interaction in the space. These representations must also support the continuous composition of information gained from interpreting new interactions. Finally, the representations must encompass relevant information in the underlying computational environment (e.g., CAD or educational software). Creating these representations is certainly a significant research challenge, but the foundations are present in systems that explicitly represent composite interaction information and in the growing set of core ontologies.

At the interface

The mediated spaces vision changes the whole meaning of the human-computer interface. Currently the interface means the special physical setup we use to deal with the computational world: screen, keyboard, microphone, gloves, headgear. The mediated space is a complete contrast: there is no interface in this sense. Participants do what they normally do—speak and gesture—and the mediated space simply understands them. At least that is the vision.

Turning the vision into reality will require significant research progress. Speech clearly becomes a key interface. Recent advances in speech technology have made some over-the-telephone and dictation applications practical. But understanding multiperson casual (or "natural") speech remains a difficult research endeavor. Understanding casual speech requires significant use of the contextual knowledge described in the previous section. Gesture recognition becomes important because gesture is an integral part of natural speech. The multiperson environment also raises the requirement for speaker tracking—who is saying what to whom.

But before turning to the issues of multiperson interaction, we need to examine the issue of metaphor. If participants use the devices in the mediated space without being aware of an interface, what are they aware of? How do they think about the mediated space?

Metaphor and metaphor evolution. In *Metaphors We Live By*, Lakoff and Johnson describe the role of metaphor in "the coherent structuring of experience"[25] (see Chapter 15). The importance of metaphor for structuring the experience of the human-computer interface has long been recognized, but determining metaphors for mediated spaces brings up a new set of issues.

Interface metaphors can be thought of in terms of three broad categories—with the understanding that there is some overlap among them:

- *Inherited* (e.g., brake pedal)—a computational enhancement of a physical object that presents in the same (or nearly the same) form as that object. Users think of the enhanced physical object in much the same way they think of the preenhanced original.
- *Projected* (e.g., the "desktop" interface)—an explicitly computational environment based on some aspects of physical reality. This is what most people mean by an interface metaphor: users know that they are interacting in a computational (or "virtual") world; the metaphor helps them to learn and understand the workings of that virtual world. The desktop interface (i.e., a computer screen with certain icons governed by certain rules of behavior) would never be confused with a physical desktop. Nonetheless, people can use the physical desktop as a metaphor for thinking about how to transfer activities (arranging documents, filing, etc.) from their physical office world to the virtual desktop world. Liddle points out[26] that "the critically important role of these metaphors [spreadsheets or desktops] was as abstractions that users could then relate to their jobs."
- *Created* (e.g., Tangible Bits, some video games)—Lakoff and Johnson also discuss[25] (see Chapter 21) the role of "metaphors that are outside our conventional conceptual system, metaphors that are imaginative and creative. Such metaphors are capable of giving us a new understanding of our experience." Computation can be used to create worlds that are outside our conventional conceptual system, but that are understood by users in terms of a new synthesis of known concepts. For example, the Tangible Geospace[9] creates a new metaphor in which users navigate electronic map displays by manipulating physical models of recognizable landmarks on top of the display.

Smart *objects* rely on inherited metaphors. The interface metaphor for an automatic braking system is the usual brake pedal. A smart shopping list on a refrigerator door might be able to update itself based on information about what is in (and not in) the refrigerator. But it will look like a shopping list, and most important, be thought of in terms of the shopping list metaphor. Its functional enhancement over a standard shopping list is a straightforward extrapolation of physical reality, well within the shopping list metaphor: instead of a person updating the list, the list updates itself. Notepad computers (e.g., PalmPilot**) may use projected metaphors to implement their explicit computational functions (e.g., calendar, address book), but their primary metaphor is clearly inherited: they are meant to look like and be thought of as physical notepads.

Augmented reality[27–30] blends inherited and projected metaphors. The DigitalDesk "makes the desk more like a workstation."[29] The desk becomes a smart object by incorporating features of the desktop metaphor into the physical reality of using a desk. DigitalDesk users can "drag and drop" images from paper by physically selecting them on the paper, moving them to other areas of the paper, and finally placing them—all mediated electronically. The actions make perfect sense to individuals who understand them in terms of the drag-and-drop desktop metaphor. Another example of augmented reality is using projected metaphors from CAD software to help repair physical objects, in this case literally projecting CAD diagrams onto physical pieces of equipment. The repairman's smart visor creates a new interface to the piece of equipment, based on the metaphors of CAD drawing.

Mediated spaces are physical spaces, and their focus is on the interpersonal interaction of the participants within them.

Smart *virtual environments* rely on projected metaphors. For example, a smart desktop could be understood through a "personal assistant" metaphor, a projected metaphor that is well established for software agents.[31] This works well as a projection: the computational environment can create a representation of the assistant, perhaps an avatar that has an assistant-like appearance, to reify the metaphor. The avatar can show its reaction to an interaction (a quizzical look, nod of the head, or verbal response) to indicate that it cannot understand, has understood and can perform the task, and so on.

But mediated spaces are real physical spaces, and their focus is on the interpersonal interaction of the participants within them. Real physical spaces do not *do* anything when individuals interact in them. What would be a metaphor for a space that performs functions? A mediated office space might perform personal assistant functions, but it is not clear that "personal assistant" is the appropriate metaphor. Any metaphor emphasizes some characteristics and deemphasizes others. The personal assistant metaphor gives users an idea of the functions they can expect, and provides an interaction model (the personal assistant can be given requests, will acknowledge and respond to requests, may anticipate requests and perform them without being asked, etc.). However, this metaphor captures nothing of the physicality and dynamics of a smart space. The space is smart because of the interaction of the notepads, desks, and displays interacting with the people who are interacting in the space. People are talking to each other, not to an avatar. Unlike the virtual world, in which the metaphor can be projected onto an avatar or even generically onto "the computer," the mediated space is not a virtual world to be populated with avatars, and there is no computer in evidence.

An interesting contrast can be seen in the work of Nagao and Takeuchi at Sony Corporation's Computer Science Laboratory.[10] They focus on making the computer a participant in a multiperson conversation. This research clearly shares with mediated spaces the fundamental challenge of understanding multiperson interaction. But the interaction metaphor is very different. In the work at Sony, the computer is an explicit conversational participant, metaphorically another person in the room. The avatar is a very natural representation of this metaphor, especially when the avatar closely models the reactions of the human face, as in their work. There is no notion of contextually enhancing the multiperson interaction, and no notion of interacting with a space.

(Sane) individuals do not talk to real physical spaces and expect them to respond. How does the space show its reaction to an interaction or notify users that it has performed a task? How does the space know that someone is talking to it? Science fiction authors have imagined a future in which humans interact explicitly with an anthropomorphic computer that controls a space. This *genius loci* may have a name and personality (like the infamous HAL 9000 of the movie *2001: A Space Odyssey*[32]) or have a more neutral presence (like the ship-wide computational environment addressed as "Computer" in the movie *Star Trek*[33]). The space has superhuman interaction skills. Persons in the space assume that all of their interactions with each other are understood, and that all of their interactions with the computer are properly interpreted and dealt with.

The near-term reality of mediated spaces will be much different (see next section), but the science-fiction portrayal of mediated spaces is so well known that it can itself be a viable metaphor. Indeed, the Intelligent Room refers to and uses the Star Trek metaphor explicitly: users say "Computer" when they wish to address the room.[4] Of course, this taps only a small part of the metaphor. The rest of it, in which the room understands what individuals are saying to each other and helps them by displaying information and talking, is yet to be realized (by anyone). Nonetheless, one possibility is that individuals will become comfortable with godlike metaphors for mediated spaces.

Another possibility is that a mediated space metaphor will evolve as an extension of smart object metaphors—from a smart shopping list that manages itself, to a smart refrigerator that manages the larder, to a mediated kitchen that manages the family's meals. The "horseless carriage" metaphor could evolve into the driverless or "smart" car. A dimension of this evolution is the ceding of responsibility. The smart shopping list starts out as a straightforward inherited metaphor from the real shopping list. The smart refrigerator is a conceptually easy next step. From there, the "momless" kitchen seems comprehensible. As more responsibility is ceded, the inheritance expands to encompass more and more physical reality—perhaps the entire space. In short, while there are no metaphors now for active spaces, they may evolve.

There is a story of a British lord who was staying for the weekend at a country house. On his first morning he complained mildly to his host that his "toothbrush did not foam." It seems that the lord did not realize that his usual valet had sprinkled tooth powder on his brush every day of his life. From our point of view, the apocryphal lord had ceded responsibility to his valet to the extent that he was unaware of the "apply tooth powder" task. The lord, of course, was not even aware of that; from his point of view, toothbrushes simply foamed.

In the modern world, the British lord might have been right in the first place. "Valetless" toothbrushes that automatically dispense dentifrice are quite conceivable. Similarly, from a future historical viewpoint, humans will be seen as having ceded more and more responsibility to mediated spaces. But from the point of view of the participants interacting in future mediated spaces, things simply work that way. Talking to rooms or nodding to refrigerators will seem entirely natural. The metaphors of the past will seem quaint: whoever thinks of a car as a horseless carriage?

The problem for interface researchers is to develop the appropriate metaphors to guide people along the way. Mediated space capability will be very limited for many years to come. The metaphors need to help individuals create and maintain the appropriate notion of the partial competence of the space, especially when it comes to dealing with multiperson speech and gesture.

Multiperson speech and gesture. When individuals come together in the same physical space, they talk to each other. They also gesture, look at their interlocutors, and change their body positions—usually as an essential part of their speech communication. In some mediated spaces, participants will also interact directly with computational devices (touch screens, phicons, etc.). But speech will be the main form of interaction for most individuals and most spaces, and direct device interaction will have to be integrated with speech. Understanding and incorporating multiperson speech and gesture is what is fundamentally new about mediated space interaction.

Speech is "in." The last several years have seen the mainstreaming of speech recognition technology, for both over-the-telephone dialog and dictation applications.[34] This speech technology represents a stunning achievement based on years of research—but it is very limited. Much of it is restricted to (single) human-to-computer interaction. There are many differences between human-computer and human-human interaction: when individuals talk to each other in all but the most formal settings, they have conversations. They use "casual speech," with all of its ellipses, disfluencies, and topic changes. They talk at the same time, especially if there are more than three individuals in the space.

The challenges are daunting—but this is not to say that we cannot progress until we achieve the entire solution. Key tenets of an approach to making evolutionary progress are:

- Incorporate available background information
- Use partial understanding technologies
- Use all of the interaction information in the space (spoken and visual)

We commonly observe that it is extremely difficult to understand what someone is saying unless we know what he or she is talking about. Our understanding of speech relies heavily on our knowledge of context. As mentioned earlier, contexts in mediated spaces play a critical role in informing the interaction-understanding system of what to expect and how to map utterances into the evolving representation. For example, a Cosmos-like context of design knowledge can be used as the basis for determining word meaning, modeling dialog, and resolving references.

The problem, as researchers in artificial intelligence have long known, is that it requires a great deal of detailed knowledge to *really* understand what speakers are saying. Decades of work have gone into understanding the structure of dialog and its relation to the task being performed by the speakers. The result is a considerable body of theory and experimental testing (see Grosz et al.[16] and Cole et al.,[24] Chapter 6). The long-term future of fully cognizant mediated spaces will undoubtedly be based on this work.

In the meantime, an interesting approach to incorporating contextual knowledge is to set the goal at only partial understanding of the information. Information extraction aims to understand key elements of—not necessarily all of—verbal information.[35] It has been applied to dialog in the MIMI system.[36] MIMI has been tested on conference room reservation tasks, for dialogs between a person who wants to reserve a room and the person who keeps the reservations for that room. The goal is not to completely understand the dialog, but to extract the "key" information: that the room has been booked (or canceled) for a particular day and time.

To give an idea of the performance of such systems, MIMI is reasonably successful in extracting the key information from restricted dialogs involving a single room reservation (recall 82.5 percent, precision 90 percent). The dialogs are "restricted" in that the utterance units are separated by a clear pause or by a change in speakers. When MIMI is used on unrestricted dialogs in which the speaker can make multiple reservations and in which disfluencies can be present, performance falls (recall 56.4 percent, precision 62.6 percent).

A reasonable near-term goal (that still requires significant further research) is a mediated space that uses information extraction techniques to map spoken interaction into an externally provided context like that in Cosmos. The space will not understand everything that is being said, but it will greatly enhance the interaction of the individuals in the space by representing the information they generate in a composite framework that can be queried and displayed, and that can provide guidance for future interaction.

In addition to explicit context, another source of background information is the spoken utterances themselves. Besides being used to create and tune speech recognizers, these kinds of data have been used in statistical language modeling for dialog understanding, e.g., to predict the next element of a dialog.[37–40] This approach has also been applied to handling casual speech effects, for example, in modeling and repairing disfluencies.[41–44] Using these techniques, recognition and language models will be constantly improving as more and more spoken data are collected in mediated spaces.

Understanding spoken interaction in mediated spaces will require the combination of all of these mechanisms for exploiting background information. In addition, mediated spaces offer the opportunity to add new sources of "foreground" information: the information that can be gleaned not only from microphones, but also from cameras and other sensors in the space. Speakers use prosody to inform others that they are about to finish speaking, but they also rely on gesture to give these cues. Gesture (lip movement, nodding, eye contact, change in body position) is also used to indicate desire to speak. Finally, what speakers say and how they say it is influenced by their visual perception of others. Gaze direction, eye contact, and head movement are strong indicators of attention and agreement or disagreement.

The combination of this kind of information with speech information will facilitate the understanding of interaction in mediated spaces. For example, there is fundamental work in using speech recognition technology to separate multiple speakers in the same environment;[45,46] the capability remains quite rudimentary. But there is progress in tasks like face location (finding the position and scale of faces in a complex image), head direction, and gaze tracking[47–51] that can be applied to the problem of speaker separation. Individuals certainly use the information they gain from observing the lips of other speakers (note how disconcerting it is to watch videos in which actors' lips are out of sync). Lip tracking,[52] the foundation for building this computational capability, is an active research topic.

Another example is topic tracking. Following the threads of a conversation is essential to understanding multiperson speech. Conversations move among different topics, even within a focused task. Conclusions about a topic may follow after several intervening topics have been discussed. One or more speakers may not "get their say" on a particular topic, which is often an important fact for mediation and decision making. Enumerating all of the topics may be important for later information extraction ("what did we talk about in that meeting?"), and so on. Topic tracking involves the integration of human-human dialog models (how language is used to signal topic changes), speech recognition (how prosody is used to signal topic changes), and visual recognition (how gesture is used to signal topic changes). It is unlikely that adequate solutions to the topic-tracking problem can be built without integrating all of these sources.

Finally, there is speaker tracking—following who is talking. Again this requires a combination of signal-level technologies, in this case speaker separation, speaker identification, and vision-based tracking of individuals.[53] An overall mediated space speech understanding environment must integrate topic tracking and speaker tracking to determine who is saying what to whom.

We can expect mediated spaces to start with topic-based information extraction and move along a path toward more complete understanding. As more and more interpersonal interaction data become available, statistical language modeling will contribute to the creation of robust and adaptable language and dialog models for mediated space interaction. Combining foreground information and using it in conjunction with background information is a research problem in itself—not simply a matter of agglomerating the technologies. Nonetheless, speech recognition techniques are currently being combined with the visual and haptic information available in mediated spaces. In summary, there is an evolutionary path toward multiperson speech and gesture understanding for mediated spaces.

Conclusions

This is a paper about point of view and hard problems. The point of view is that mediated space requires a computational environment that deals with the realities of physical space and multiperson interaction. The hard problems arise directly from that point of view. What is needed are:

- Descriptive schemes that treat devices and their manifestations as real-world objects, images, sounds, and sensations
- Representation of enough contextual knowledge to allow mediated spaces to at least partially understand and organize human-human interaction
- Creation of metaphors that help people interact with computation invisibly embedded in space
- Understanding of the way people communicate with each other in terms of speech and gesture

The solutions to these problems will begin to emerge over the next decade or two. A key question for researchers is: what will mediated spaces be like during the evolution of these solutions?

Mediated space evolution must be managed. Spaces must provide useful, easy-to-understand capabilities, and expectations must be aligned with those capabilities. Early mediated spaces will probably emphasize capabilities that require minimal multiperson interaction. The mediated kitchen could evolve from improving collaboration of smart objects, as described earlier, with relatively low levels of human interaction required. The metaphor must evolve with the capability. Once individuals start ceding responsibility to the space, their expectations must be managed carefully.

Even in spaces that are based wholly on multiperson interaction, partial understanding techniques can still be useful for many tasks. As discussed, information extraction can be used to place the key elements of an interaction into a composite structure that can guide future interaction. The critical added value is in organizing the information. This is far from the full-blown vision of mediated space interaction, but it is nonetheless a compelling capability. Less than full-blown mediated space environments are already impressive, e.g., Classroom 2000[14] and Virtual Meeting Rooms.[11]

A key evolutionary parameter is the explicitness of the computational task. In most working smart spaces,[4,5] even though the computation is pervasive, most of the interaction is explicit. Individuals' interaction with each other is not part of the computational task. Only their interaction directly with computers (explicitly writing on pads, pointing at displays, talking to computers) is computational. Until the technology for understanding natural multiperson interaction becomes practical, mediated spaces will also emphasize explicit computation. Mediated design sessions will focus on interaction with the CAD software; mediated classrooms will focus on interaction with representations of the lesson.[14] As the technology evolves, more of the person-to-person interaction will become part of the computational task. The CAD software will start to interact with the actual design discussion; the lesson software will augment student-teacher and student-student interaction; and the computer will start to disappear.

Mediated spaces emphasize certain aspects of interface technology. An issue for speech understanding in general, but one that is especially pressing in mediated spaces, is performance: real-time or near real-time capability is a requirement. In any speech system it is annoying if the system determines after some interval that it has not understood an utterance. In a multiperson conversational environment, this would be a disaster—the conversation would have gone on, and it would be extremely onerous or even impossible for the participants to recreate it.

Similarly, speech synthesis is a general issue in speech systems, but it is more pressing in mediated spaces. Anyone who has dealt with synthesized speech is aware of the problem: current technology produces stilted, difficult-to-understand speech. The computer cannot hold up its end of the conversation. In any mediated space metaphor that involves interaction with the persons in the space, the space must have a voice.

Finally, an interesting and potentially far-reaching aspect of mediated spaces involves the dimension of time. Given the large variety of devices and mediated spaces, it will become incumbent on devices to keep track of their own history—and to pass it on to their successors before they "die." This is important for authentication (knowing the pedigree or provenance of an artifact is an important authentication technique). But it is even more important for providing stability from the human point of view in a stressfully dynamic environment of different devices and spaces. (It would be nice if your notepad had "tribal memory" of how to coordinate with other devices to give an icon salience on a display in a particular space, and if your notepad remembered how you finally got electronic cash last time you were in Ulan Bator . . . .)

Implementing the mediated-space vision requires interdisciplinary research and an integrated approach. The research problems outlined in this paper are not independent; they define a nexus of interacting research approaches that must be realized through collaboration.

In *Things That Make Us Smart*,[54] Don Norman points out that the physical devices that last are those that can be instantly (or at least very quickly) understood by their human users. This is not an accident; it is the result of design, and usually a great deal of trial and error. Mediated spaces are no different. Their capability and metaphor must be designed to be understandable by and useful to their human users. On the very same page (page 103), Norman describes what humans are good at:

> We communicate and work well in small groups, sharing and cooperating to accomplish tasks beyond the capability of the individual. The cooperation is aided through the communicative powers of language and body: spoken and written words, gestures, eye contact, and facial expressions.

The vision of mediated spaces is to augment these most human capabilities.

Acknowledgments

The ideas about multiperson interaction were developed in conjunction with many researchers at SRI International, particularly David Beymer, Adam Cheyer, John Dowding, Horacio Franco, Luc Julia, Ray Perrault, Patti Price, Liz Shriberg, Andreas Stolcke, and Fuliang Weng. I would also like to acknowledge the insights and suggestions of the referees.

**Trademark or registered trademark of Minnesota Mining and Manufacturing Company (3M) or 3Com Corporation.

Cited references and notes

1. D. Englebart, *Augmenting Human Intellect: A Conceptual Framework*, AFOSR-3233 (October 1962), available at http://sloan.stanford.edu/mousesite/EngelbartPapers/B5_F18_ConceptFrameworkInd.html.
2. R. A. Bolt, "Put-That-There: Voice and Gesture at the Graphics Interface," *Computer Graphics* **14**, No. 3, 262–270 (1980).
3. M. Weiser, "The Computer for the 21st Century," *Scientific American* **265**, No. 3, 94–104 (September 1991).
4. M. A. Coen, "A Prototype Intelligent Environment," *Cooperative Buildings—Integrating Information, Organization, and Architecture*, N. Streitz, S. Konomi, and H.-J. Burkhardt, Editors, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg (1998).
5. M. A. Coen, "Design Principles for Intelligent Environments," *Proceedings of AAAI 1998 Spring Symposium on Intelligent Environments*, Palo Alto, CA (March 23–25, 1998), pp. 36–43.
6. See http://tangible.www.media.mit.edu/ttt/.

7. A. Pentland, "Smart Rooms," *Scientific American* **274**, No. 4, 68–76 (April 1996).
8. A. Pentland and A. Liu, "Modeling and Prediction of Human Behavior," *Neural Computation* **11**, 229–242 (1999).
9. H. Ishii and B. Ullmer, "Tangible Bits: Towards Seamless Interfaces Between People, Bits, and Atoms," *Proceedings of Conference on Human Factors in Computing Systems (CHI '97)*, Atlanta, GA (March 22–27, 1997), pp. 234–241.
10. K. Nagao and A. Takeuchi, "Social Interaction: Multimodal Conversation with Social Agents," *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, Seattle, WA (August 1–4, 1994), Vol. 1, pp. 22–28.
11. A. Ginsberg and S. Ahuja, "Automating Envisionment of Virtual Meeting Room Histories," *Proceedings of ACM Multimedia 95*, San Francisco, CA (November 5–9, 1995).
12. W. Mark and J. Dukes-Schlossberg, "Cosmos: A System for Supporting Engineering Negotiation," *Concurrent Engineering Research and Applications* **2**, No. 3, 173–182 (July 1994).
13. M. Cuskosky, R. Englemore, R. Fikes, M. Genesereth, T. Gruber, W. Mark, J. Tenenbaum, and J. Weber, "PACT: an Experiment in Integrating Concurrent Engineering Systems," *IEEE Computer* **26**, No. 1, 28–38 (January 1993).
14. G. Abowd, C. Atkeson, J. Brotherton, T. Enqvist, P. Gulley, and J. LeMon, "Investigating the Capture, Integration and Access Problem of Ubiquitous Computing in an Educational Setting," *Proceedings of CHI '98*, Los Angeles, CA (April 18–23, 1998), pp. 440–447.
15. G. D. Abowd, "Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment," *IBM Systems Journal* **38**, No. 4, 508–530 (1999, this issue).
16. B. Grosz, M. Pollack, and C. Sidner, "Discourse," *Foundations of Cognitive Science*, M. Posner, Editor, MIT Press, Cambridge, MA (1989), pp. 65–75.
17. IPv6 Specification, see http://www.ietf.org/internet-drafts/draft-ietf-ipngwg-ipv6-spec-v2-02.txt.
18. IETF IP Routing for Wireless/Mobile Hosts (Mobile IP) working group, see http://www.ietf.org/html.charters/mobileip-charter.html.
19. D. B. Johnson and D. A. Maltz, "Protocols for Adaptive Wireless and Mobile Networking," *IEEE Personal Communications Magazine* **3**, No. 1, 34–41 (February 1996).
20. A. Kehler, J.-C. Martin, A. Cheyer, L. Julia, J. Hobbs, and J. Bear, "On Representing Salience and Reference in Multimodal Human-Computer Interaction," *AAAI Workshop on Representations for Multimodal Human-Computer Interaction*, Madison, WI (July 26–27, 1998), pp. 33–39.
21. D. Martin, "The Open Agent Architecture: A Framework for Building Distributed Software Systems," *Applied Artificial Intelligence: An International Journal* **13**, No. 1–2 (January–March 1999).
22. A. Cheyer, L. Julia, and J.-C. Martin, "A Unified Framework for Constructing Multimodal Applications," in *Proceedings of the Second International Conference on Cooperative Multimodal Communication (CCMS '98)*, San Francisco (January 1998), pp. 63–69.
23. See Survey of the State of the Art in Human Language Technology, NSF/EC at http://cslu.cse.ogi.edu/HLTsurvey/.
24. See http://ontolingua.stanford.edu.
25. G. Lakoff and M. Johnson, *Metaphors We Live By*, University of Chicago Press, Chicago, IL (1980).
26. D. E. Liddle, "Design of the Conceptual Model," *Bringing Design to Software*, T. Winograd, Editor, Addison-Wesley Publishing Co., Reading, MA (1996), pp. 17–31.
27. P. Wellner, W. Mackay, and R. Gold, "Computer Augmented Environments: Back to the Real World," *Communications of the ACM* **36**, No. 7, 24–26 (July 1993).
28. L. Stifelman, "Augmenting Real-World Objects: A Paper-Based Audio Notebook," *Proceedings of CHI '96*, Vancouver, BC (April 13–18, 1996).
29. P. Wellner, "Interacting with Paper on the Digital Desk," *Communications of the ACM* **36**, No. 7, 87–96 (July 1993).
30. MIT Media Lab, Vision and Modeling Group's Smart Desk Project, see http://vismod.www.media.mit.edu/vismod/demos/smartdesk/.
31. National Information Infrastructure Report, "The Role of Intelligent Systems in the National Information Infrastructure," D. Weld, Editor, *AI Magazine* **16**, No. 3, 45–64 (Fall 1995).
32. *2001: A Space Odyssey* was released in 1968 by Paramount Pictures. It was produced and directed by Stanley Kubrik, who also wrote the screen play along with Arthur C. Clarke.
33. *Star Trek: The Motion Picture* was released in 1979 by Metro-Goldwyn-Mayer. It was produced by Gene Roddenberry and directed by Robert Wise. Harold Livingston wrote the screenplay.
34. See http://dir.yahoo.com/Business_and_Economy/Companies/Computers/Software/Voice_Recognition/.
35. D. Appelt, J. Hobbs, J. Bear, D. J. Israel, and M. Tyson, "FASTUS: A Finite-State Processor for Information Extraction from Real-World Text," *Proceedings of International Joint Conference on AI (IJCAI-93)*, Chambéry, France (August 28–September 3, 1993), pp. 1172–1178.
36. M. Kameyama and I. Arima, "Coping with Aboutness Complexity in Information Extraction from Spoken Dialogues," *Proceedings of the International Conference on Spoken Language Processing (ICSLP-94)*, Yokohama, Japan (September 21, 1994).
37. M. Nagata and T. Morimoto, "First Steps Toward Statistical Modeling of Dialogue to Predict the Speech Act Type of the Next Utterance," *Speech Communication*, No. 15, 193–203 (1994).
38. N. Reithinger, R. Engel, M. Kipp, and M. Klesen, "Predicting Dialogue Acts for a Speech-to-Speech Translation System," *Proceedings ICSLP-96*, Vol. 2, Philadelphia, PA (October 4, 1996), pp. 654–657.
39. A. Stolcke, E. Shriberg, R. Bates, R. Coccaro, D. Jurafsky, R. Martin, M. Meteer, K. Riss, P. Taylor, and C. Van Ess-Dykema, "Dialog Act Modeling for Conversational Speech," *Applying Machine Learning to Discourse Processing: Papers from the 1998 AAAI Spring Symposium*, Technical Report SS-98-01, J. Chu-Carroll and N. Green, Editors, AAAI Press, Menlo Park, CA (March 1998), pp. 98–105.
40. E. Shriberg, R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Riis, N. Coccaro, P. Martin, M. Meteer, and C. Van Ess-Dykema, "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?," *Language and Speech*, 1998.
41. D. O'Shaughnessy, "Correcting Complex False Starts in Spontaneous Speech," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, Vol. 1, Adelaide, Australia (April 19–22, 1994), pp. 349–352.
42. P. Heeman and J. Allen, "Detecting and Correcting Speech Repairs," in *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, NM (June 27–30, 1994), pp. 295–302.
43. C. Nakatani and J. Hirschberg, "A Corpus-Based Study of Repair Cues in Spontaneous Speech," *Journal of the Acoustical Society of America* **95**, No. 3, 1603–1616 (1994).
44. E. Shriberg, R. Bates, and A. Stolcke, "A Prosody-Only Decision Tree Model for Disfluency Detection," *Proceedings EU-*

ROSPEECH '97, Vol. 5, Rhodes, Greece (September 22–25, 1997), pp. 2383–2386.

45. D. Benincasa and M. Savic, "Co-Channel Speaker Separation Using Constrained Nonlinear Optimization," *Proceedings of the 1997 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*, Vol. 2, Munich, Germany (April 21–24, 1997), pp. 1195–1198.
46. M. Savic, H. Gao, and J. Sorensen, "Co-Channel Speaker Separation Based on Maximum-Likelihood Deconvolution," *Proceedings of the 1994 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-94)*, Vol. 1, Adelaide, Australia (April 19–22, 1994), pp. 25–28.
47. T. Jebara and A. Pentland, "Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico (June 17–19, 1997), pp. 144–150.
48. H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA (June 23–25, 1996), pp. 203–208.
49. K.-K. Sung and T. Poggio, *Example-Based Learning for View-Based Human Face Detection*, *AI Memo #1521*, MIT, Cambridge, MA (December 1994).
50. B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection," *Proceedings IEEE Conference on Computer Vision*, Cambridge, MA (June 20–23, 1995), pp. 786–793.
51. N. Oliver, B. Rosario, and A. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," *Proceedings of the International Conference on Vision Systems (ICVS 99)*, Gran Canaria, Spain (January 13–15, 1999), pp. 255–272.
52. R. Kaucic and A. Blake, "Accurate, Real-time, Unadorned Lip Tracking," *Proceedings of the 6th International Conference on Computer Vision*, Bombay, India (January 4–8, 1998), pp. 370–375.
53. S. Rizvi, J. Phillips, and H. Moon, "A Verification Protocol and Statistical Performance Analysis for Face Recognition Algorithms," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA (June 23–25, 1998), pp. 833–838.
54. D. Norman, *Things That Make Us Smart*, Perseus Books, Reading, MA (1993).

Accepted for publication May 12, 1999.

William Mark *SRI International, 333 Ravenswood Avenue, Menlo Park, California 94025 (electronic mail: bill.mark@sri.com).* Dr. Mark is vice president of the Information and Computing Sciences Division of SRI International. This division creates new technology in information security, system design, speech and natural language, vision and perception, planning and reasoning, and formal methods. The group performs leading-edge research, with a strong interest in intellectual property creation and commercialization. Prior to joining SRI, Dr. Mark headed the System Technology Group at National Semiconductor. He was formerly Director of Information and Computing Sciences at the Lockheed Martin Palo Alto Research Laboratories, and was a cofounder of Savoir, a company developing software tools for flexible manufacturing. Previously, he held positions at the University of Southern California Information Sciences Institute and the General Motors Research Laboratories. He holds a Ph.D. degree in computer science from MIT. His personal research interests include system design and smart spaces.

Reprint Order No. G321-5712.