The Affordances of Media Spaces for Collaboration

William W. Gaver

Rank Xerox Cambridge EuroPARC
61 Regent Street
Cambridge CB2 1AB, U.K.
+44 223 341 527
gaver@europarc.xerox.com

ABSTRACT
In this paper, I discuss the affordances offered by media spaces for collaboration, contrasting their properties with those of the everyday medium and exploring the implications for perception and interaction. Collaboration is situated in a physical environment which supports or constrains the various forms social interactions might take. An analysis of the affordances of the environment – the properties that offer actions and interactions to those within it – thus complements analyses which emphasize social and cultural factors. Examining the "physics" of media space systems is helpful both in understanding how people use them to collaborate and in suggesting possibilities for design.

KEYWORDS
Video, mediaspaces, affordances, ecological approaches

INTRODUCTION
Recently there has been increasing interest in "media spaces" – computer-controllable networks of audio and video equipment used to support synchronous collaboration (e.g., 7, 16, 3, 17, 19, 21). Because they allow simultaneous, two-way transmission of visual and auditory information, these systems are often thought to simulate the everyday medium within which we collaborate. In fact, a number of analogies to the everyday world have arisen to describe the experience with such systems. The term "media space" itself indicates an analogy to the everyday space; they are said to support "virtual copresence," to act as "tailorable office-spaces," "meeting rooms," or "hallways."

In *Proceedings of CSCW'92* (Toronto)

These metaphors are useful in indicating some of the intended functions of media spaces, but they may well be misleading in suggesting that media spaces are like the everyday medium for interaction. In the everyday world, collaboration is situated within a shared, encompassing space, one which is rich with perceptual information about objects and events that can be explored and manipulated. The "space" created by current audio-video technologies, on the other hand, has very different properties for perception and interaction. By comparing these properties with those of the everyday medium, we can better understand the nature of these systems and the challenges and opportunities for their design.

An Ecological Approach To Understanding Media
In this paper, I discuss the *affordances* offered by media space systems for perception and interaction. Affordances are properties of the environment that offer actions to appropriate organisms (8). They are defined with respect to both the environment and the interacting organism, and thus provide physical analyses that can complement sociological perspectives on situated action. In addition, because they focus on relevant physical properties of the environment, analyses of affordances can directly suggest implications for design (6).

An examination of the affordances of media spaces for collaboration is a useful complement to analyses which emphasize social and cultural influences. Social activities are *situated* in their environment (22): if collaboration depends on complex, subtle social relations, it also depends on a medium in which these relations can work. From this point of view, an important aspect of understanding collaboration is understanding the environment in which it takes place. Because the concept of affordances focuses directly on the relations between properties of the environment and possibilities for (inter)action, it is a particularly valuable tool for succinctly describing situational factors relevant for collaboration.

In considering the affordances of audio-video technologies, I am inspired by Gibson's (8) discussion of the affordances of the everyday medium. According to this approach, the everyday medium is characterized by a number of properties and corresponding affordances. Gibson lists six. In his account, the medium:

- transmits light, affording vision,
- transmits vibrations, affording hearing,
- offers little resistance, affording motion,
- allows chemical diffusion, affording smell,
- contains oxygen, affording breathing,
- has an intrinsic up and down, affording orientation.

Clearly this analysis is too coarsely grained to serve as a basis for a comparison with media spaces. But it is useful in suggesting that the properties of the everyday medium can be directly related to the actions it affords – the fact that it transmits light, for instance, allows vision, the fact that it offers little resistance allows motion, and so on. This is the essence of affordances: they point to the relationship between properties of the environment and the possibilities for action it allows.

THE AFFORDANCES OF MEDIA SPACES

Typical media spaces have different properties from those of the everyday medium and thus provide different possibilities for action. They use computer-controlled links to transmit perceptual information to and from nodes incorporating one or more cameras, microphones, monitors and speakers (see Figure 1). Thus people are not surrounded by information in media spaces, they cannot move within them, and communication within them is constrained by the limited bandwidth they offer. In fact, the "space" created by these technologies is both discontinuous and arbitrary. These properties of the audio-video

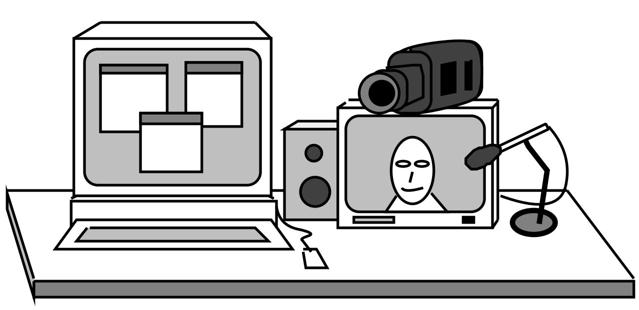

Figure 1. A typical media space node: a video monitor with a nearby camera, microphone, and speaker.

medium shape how we collaborate using it just as the everyday medium shapes our interactions within it. In further exploring these properties, then, we can better understand how people interact in media spaces.

Moreover, because an analysis of media spaces' affordances focuses on the physical attributes of the medium, it highlights possibilities for design. Neither media spaces nor their component technologies – video, audio, or computer – are stable artifacts. With changes in the technologies come changes in their affordances; creating or emphasizing affordances is, in fact, a useful way to characterize the purpose of design (8).

In order to stress possibilities for design, this account focuses on properties of media spaces which constrain various forms of perception and interaction. I take this stance not to damn media spaces, however, but to understand them. In the end, some of the constraints of media spaces can be put to good use. Many of the truly limiting aspects can be improved. And some offer entirely new possibilities: Above all, audio-video technologies have an important affordance that the everyday medium does not:

> *Media spaces convey visual and auditory information between arbitrary points, and thus afford remote collaboration.*

In the following discussion, I explore several of the properties of audio-video systems that shape their affordances for perception and interaction. I start with an account of vision, audition, and movement in media spaces as a way of describing the environment within which action and collaboration is situated. Using this as a foundation, I turn from affordances for obtaining perceptual information to an explicit consideration of collaboration, and consider the ways that the audio-video medium shapes our interaction within it.

AFFORDANCES FOR VISION

Simple awareness of one's colleagues is a strong predictor of their collaboration (19). Because media spaces provide visual information about remote sites, we might expect to maintain awareness of events in media space much as we do local events. For example, Figure 2 shows an image taken from the video camera in a colleague's office. This view makes visual information available that supports the video equivalent of face-to-face interaction, and is more generally useful for maintaining awareness of activities in her office. However, awareness of the scene is significantly constrained by the video technologies employed. For instance, is she looking at her computer, at the camera, or at my image on her monitor? Is there somebody else in her office, perhaps standing just to the left of the view? What's that paper on the table? Clearly, the properties of current video technologies shape their affordances for vision.

Field of View and Resolution

Video provides a restricted field of view on remote sites. This limits peripheral vision and constrains perceptual exploration.

Unless we are standing next to an obstacle, we usually have access to a spherical field of optical information. Even if the head and eyes are kept motionless, our natural range of vision is almost hemispherical. Video cameras, in contrast, have a much narrower field of view; typically well less than a quarter and often closer to an eighth of their surroundings. This is compounded by the fact that video monitors are often relatively small, taking up only a fraction of one's normal field of view. When we look at video images, then, we only have access to a constrained view of the remote site. In addition, peripheral awareness of activities in such scenes is limited, both in the sense that remote information is not available to the entire visual field and in the sense that remote events often go unnoticed when the monitor is on the periphery of the visual field.

The resolution of video limits detailed inspection.

In the everyday medium, constraints on visual acuity come from our visual systems rather than from the nature of the medium itself. The level of detail that is made available by video images, in contrast, is always fixed at pixel size. In addition, to the extent that pixels are sharply defined, spurious high-frequency visual information is added which further masks details (1). The result is that close scrutiny of video images reveals the structure of the medium itself rather than details of a remote scene. We can't read remote partners' computer screens, nor share a detailed drawing with them. Our ability to inspect remote scenes is limited by the technology, not our visual systems.

Design Implications

Because of the limited resolution of current video technologies, there is a tradeoff between the field of view offered and the ability to focus on details. The effects of this tradeoff can be seen in the image shown in Figure 2: the camera is positioned against a corner of the room to view an appreciable amount of the scene, and this limits our access to distant objects. Wide-angle lenses can make more of the ambient information at a distant site available, but imply a loss of detail (especially at the edges of displays). Wall-sized video displays or eye-covering displays can make video information more encompassing, but again as the image grows visually the lack of resolution becomes more apparent. This tradeoff can only be overcome by the development of video with resolution greater than our eyes can register, by the integration of multiple video images, or – as I will discuss below – by allowing movement in remote scenes.

Information for Three Dimensions

Video conveys a limited amount of information about the three-dimensional structure of remote scenes, and thus limits exploration, inspection, and peripheral awareness.

Light provides a great deal of visual information about the three-dimensional structure of the everyday world. Much of this information is lost using current video technologies. Clearly, stereopsis and convergence are impossible given the single image presented by current video displays. Less obviously, but equally important, both motion parallax – changes in the scene caused by movement independent from the observer – and especially movement parallax – changes in the scene produced by self-induced movement around a fixation point – are also lost. Only occlusion, texture gradients, and movement in the scene itself provide information about a space within media spaces – and this is countered by information about the flatness of the monitor screen itself. This interferes with our ability to discriminate and attend to people and things on the basis of their distance. It also means that awareness of colleagues and events on the periphery is often disrupted: movements in the remote space are largely two-dimensional and do not seem to impinge on the local scene. The “space” created by video is hardly a space at all, and is certainly not one that extends continuously from the local environment.

Figure 2: A framegrabbed image from a colleague's video node. Is she looking at her computer, at the camera, or at my image on her monitor? Is there somebody else in her office, perhaps standing just to the left of the view? What's that paper on the table? This is a reduced and reproduced image, but the same limitations hold for full-sized video.

Design Implications

Information for three-dimensional structure can be increased in several ways. Stereopsis and convergence can be made available by various technologies that allow different images to be presented to each eye (see, e.g., 18, 11). The best-known systems require that viewers wear equipment over their eyes: either small monitors such as used in virtual reality research, polarized or red-green glasses, or spectacles which use rapidly-switched liquid crystal lenses to occlude the view to alternating eyes. Such systems create a convincing experience of depth, but are intrusive both in requiring viewers to wear special equipment and in interfering with the view of the local scene. More interesting are passive displays, which present different views to each eye without requiring spectacles. These include systems which collimate the light emitted by a conventional screen into narrow parallel beams which are swept through the azimuth (24), lenticular lens systems, and Visidep systems (15) which rapidly alternate the view from two vertically-mounted cameras. As yet, such systems have not been employed in media-space research; the incorporation of stereoscopic information in such systems is an interesting possibility for further exploration.

AFFORDANCES FOR LISTENING

Just as the everyday medium surrounds us with light, so we are enveloped in a rich, reverberating field of structured sound, one created by the coupling between vibrating surfaces and air. In media spaces, the situation is very different: sound is transmitted in limited ways by constrained channels. This has several implications for listening in such environments.

> *Audio equipment is usually monophonic, impeding localization and conveying a biased sample of remote audio information.*

Because we are surrounded by the everyday medium, sounds reach us from every direction, allowing us to localize their sources. Most media spaces, in contrast, use only a single microphone to capture sounds, drastically reducing cues for localization. Not only does this interfere with orientation to individual speakers, but also with attention to a single “stream” of sound in a noisy environment (2). Localization helps us *ignore* undesirable sounds – we can hold a conversation in a crowded room, for instance, by attending to a single stream of sound. When localization is reduced as it is in video systems, unwanted noises intrude on desirable signals.

> *Sounds are altered by the electronic medium, biasing the auditory information made available.*

Most sounds are produced by a coupling between sources and the air, and thus are all conveyed with roughly the same efficiency. In most media spaces, on the other hand, microphones are usually attached to some solid object which itself may vibrate. The effect is similar to placing ones head against a vibrating surface: the sounds of the object overwhelm those conveyed by air. In addition, microphones pick up a disproportionate amount of airborne mechanical noise from locations near where they are mounted – noise from nearby machines, footsteps and so forth. Finally, microphones, cables, amplifiers and speakers all shape audio signals as well. The result is that remote colleagues may be difficult to understand, and all are likely to be readily distinguished from those in the local environment.

Design Implications

These observations have obvious implications for design. Localization can be aided by using stereo audio equipment. Microphones can be hung freely or mounted on damped supports, just as our ears are supported by a highly-damped material, to reduce biases caused by mechanical coupling. Finally, attention to the fidelity of audio equipment is of paramount importance if remote events are to sound like local ones, or to be interpretable at all.

AFFORDANCES FOR MOVEMENT

So far this discussion has concerned static perception of remote scenes – the differences between viewing a remote scene via a video image and being present but motionless, sitting in a chair with an unmoving head. However, perception is seldom static. Instead, we explore our environment to discover new perceptual information. In media space, we cannot do this: the effect is as if we were paralyzed at the remote location, unable to move or even shift our eyes.

> *Cameras and microphones are stationary or only moved remotely, preventing perceptual exploration.*

As Gibson (8, 9) stressed, the visual system consists of movable eyes in a movable head on a movable body. Movement allows us to achieve panoramic vision from a single station point by bending over, turning around, and so forth. We can view a shared workplace from different angles by moving around it, we can inspect the details of a shared work object by moving closer, or achieve a more encompassing field of view by moving further away. Moreover, movement allows us to obtain a great deal of information about the three-dimensional layout of a scene; as we move, the visual flow fields corresponding to different depths shift at different rates.

Similarly, movement allows listeners to orient to each other, to avoid locations which are particularly noisy and so forth. We can lean towards an interesting conversation, or away from an obtrusive one, or walk over to an entirely different group of people. In general, movement is crucial for perception and collaboration.

But exploratory movement is not supported by typical media space systems. Cameras are usually mounted at stationary vantage points meant to provide satisfactory views and are often left at a wide angle of focus to give maximum contextual information. Microphones are usually placed at a location near cameras, and are usually omnidirectional. Cameras and microphones are moved or camera foci changed only by people at the distant location – producing an experience more like watching a home movie than like that of self-generated perceptual exploration.

Design Implications

In general, providing remote controlled movement in media spaces has the potential to compensate for many of the limitations of still video cameras and monitors. For instance, using multiple cameras at each remote location to provide access from various angles and distances provides a partial solution to the tradeoff between field of view and resolution. This approach might be extended, for instance by the development of a synchronous version of video mosaic techniques (12) which allow the progressive unfolding of detail in a video image to approximate the ability to focus more or less narrowly on a scene. Finally, remotely controlled pan-tilt-zoom units, already used in many commercial teleconferencing systems, can provide an even richer degree of self initiated movement.

Similarly, user controlled mixing of the signals from multiple remote microphones would not only allow a more representative sample of ambient audio information, but produce an effect similar to moving between the microphones' remote locations. In this case, microphones should be well spaced and speakers separated, since the purpose is to capture a more representative sample of the remote audio information rather than to create a stereo image.

But systems might also be designed which allow true exploratory movement in remote spaces. Possibilities include mounting cameras and microphones on remotely-controllable robot arms or vehicles. For instance, at EuroPARC we have experimented with a camera mounted on a radio-controlled model car; though impractical for everyday usage, the device allows true exploration of remote spaces and is remarkably easy to control.

The development of such systems is promising, but emphasizes issues concerning the control of movement. Successful systems must *afford* movement – more than making it possible, they must make it easy to initiate and control. Pan-tilt-zoom units used in teleconferencing systems are often very slow and awkward to use; cameras mounted on radio-controlled cars are fun, but impractical. Unless the cost of gaining additional information is low enough, it will not seem worth the additional effort.

But systems can be designed that rely on natural exploratory movement. For instance, the intuitive experience of looking through a window into a remote space, with the ability to move with respect to the scene, has been simulated by using a head-tracking device to control remote camera movement (18). This system follows head movement quickly and accurately enough to create true movement parallax, and thus three-dimensional exploration of remote scenes is possible. With controls such as these, systems allowing remote movement present exciting possibilities for future media space research.

AFFORDANCES FOR INTERACTIVE MOVEMENT

Understanding the effects of the medium offered by current media-space technologies on perception and action provides a good foundation for understanding their effects on collaboration. Collaboration is situated in a medium for perception and action; what we can see and do in this medium has profound effects on how we can communicate and interact. If the properties of audio-video technologies have implications for perception of remote spaces, they combine to produce even more pronounced effects on interaction.

For instance, Heath and Luff (10) have observed that certain kinds of communicative gestures seem to be relatively ineffective over video. Trying to capture a remote person's attention, for example, one may begin by making relatively subtle gestures. These are often overlooked, and increasingly large movements result, sometimes ending in a humorous display of waving, face-making, etc. The effect is that we sometimes seem not only to be paralyzed in media space, but invisible as well.

> *Media spaces are often discontinuous, reducing the ability to create communicative gestures and to control conversational access.*

In the everyday medium, we can move with respect to one another in a variety of ways. We can reach into other peoples' views, almost literally grasping their attention and directing it to oneself or to a relevant direction. We can manipulate other people's access to our conversations through control of the audibility of speech, the visibility of facial expressions, and so forth. Movement affords a smooth transition between privacy and accessibility: people may subtly gesture their willingness to converse, bend and talk with their neighbors, or straighten and address an entire assembly.

In current media spaces, these abilities are often limited. One can't gesture within a shared space because of the barrier presented by monitor screens. Peripheral vision is limited by constrained camera views and screen sizes, and also because video emits light rather than reflecting it, so that people and objects in a video scene appear more similar to each other than to those in the surrounding local environment. The paucity of information about the three-dimensionality of remote scenes produces further discontinuities between local and remote spaces. The result is that gesturing is limited both because it is not always noticed and because there is not a mutual space for gesturing but instead two separated territories.

Control over conversational access is also constrained by current audio-video systems (4). Media space systems usually allow only one video and audio signal to be broadcast from a given location to all others; this does away with any possibility of speaking privately with a subset of members. Even when addressing one person in a group, the loss of visual and auditory cues produces confusion. In general, because media spaces are not shared, we cannot move within them, but only gesture and talk from one side of a barrier to all other spaces that may be connected to it.

Design Implications

Gesturing can be facilitated simply by using larger monitors, thus making movement more salient (10). But the perceived distance of a person in the remote scene often conflicts with that of the monitor itself. This leads to the problem that increasing monitor size enough to make others noticeable can make the monitor itself overwhelming. A more demanding but potentially more effective solution is to increase the sense of shared space in media spaces, providing greater information for three-dimensionality either through stereopsis or movement. This would allow gestures to move towards and perhaps even "into" the space surrounding remote colleagues. Media space would thus become a much more encompassing and fluid medium, capable of supporting more natural interaction.

Conversational access can be re-established by using multiple cameras and microphones (20), or by using voice-tracking to determine which of several possible speakers is displayed on a screen (4). This sort of approach is relatively new, but extremely promising in overcoming limitations in conversational control.

AFFORDANCES FOR PREDICTABLE INTERACTION

Considering the affordances of the everyday medium is helpful in highlighting the affordances of the audio-video medium. But considering the affordances of media spaces can also highlight those of air, including affordances so taken for granted as to be easily overlooked. For instance, *isotropism* is a term used in physics to refer to a material which has characteristics that are the same when measured along any axis.

Media spaces are often anisotropic, making prediction difficult.

Air is isotropic with respect to light and – unless it is moving – with respect to sound as well. This means that air affords reciprocal communication, that people can predict what their partners will see and hear by what they themselves see and hear. Many of the social conventions which surround interaction rely on this property of the medium; without it, they are problematic.

Media spaces, in contrast, are often functionally anisotropic. Because of the independence and separation of cameras and monitors, microphones and speakers, the effect is that light and sound travel differently between different points (see Figure 3). This interferes with the foundation of predictability that lies behind many social conventions of interaction. For instance, if I can see your eyes in the everyday medium, then you can see mine (unless I'm wearing sunglasses). This is often not the case for video.

The anisotropism of media spaces is as crucial for social interaction as immobility is for perception. Here I discuss just two implications of anisotropism; there are many more which might be considered.

Video is anisotropic, interfering with the design of communicative gesture and with gaze awareness.

One of the reasons communicative gesturing is often ineffective in media spaces is that there is no shared space between conversational participants. But the anisotropism of video as a medium further interferes with gesturing

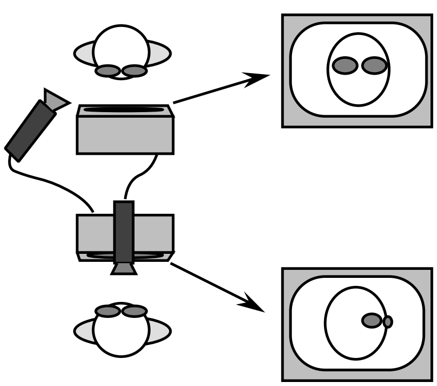

Figure 3. Video connections are often anisotropic: views travel in different ways between different nodes.

because the affordance of predictability is weakened. Gesturing not only requires a shared space in which to move, but the ability to design a gesture within that space for the desired impact. The loss of predictability in media spaces undermines affordances for communicative gesture (cf. 10).

Gaze awareness also relies on a predictable shared space to be effective. Gaze is used a great deal in conversations to facilitate turn-taking, indicate interest, and reflect social relations (16). The anisotropic nature of video disrupts gaze just as it does gesture. Not only does it prevent true eye contact, but it constrains the perception of gaze direction. For instance, because monitors are usually smaller than the scenes they portray, looking from one side of a video scene to another results in relatively small eye movements.

Design Implications

Two broad classes of solution to this problem have emerged: one technical, the other social. On the one hand, the loss of eye-contact has prompted the development of “video tunnels” which employ half-silvered mirrors to allow cameras to view a scene as if mounted directly in front of a monitor (3). Similarly, the ClearBoard system (14) merges video tunnel technology with that used to create VideoDraw (23) to produce the illusion of drawing on, and looking through, a pane of glass. Not only does this allow eye contact, but participants can see what their partners are looking at. Both these systems represent technological solutions to the problem of gaze awareness, though at the expense of affordances such as movability of the equipment itself.

On the other hand, a social solution has also evolved to compensate for the loss of reciprocity in looking. Experienced inhabitants of EuroPARC's media space sometimes choose to look at their cameras instead of their monitors while engaged in conversations – giving up visual access to their partner in order to create the illusion of access for communicative purposes. This is an interesting example of a social custom being reshaped to fit the physical properties of the medium in which it is situated. More generally, this example highlights the complementary nature of ecological and social explanations of interaction. On the one hand, the physics of the environment, understood at an ecological level, constrains possibilities for social patterns. On the other hand, society and culture shape our experience of the environment and guide our actions within it. These sorts of issues become particularly clear when considering technologies such as media spaces, which involve the design of physical artifacts that affect social relations.

> *Media space is anisotropic, affording "spying," but also unobtrusive "glances" and "office shares."*

Because air is an isotropic medium, finding an observation point which allows access to a scene without allowing visual access to oneself is difficult. Video, in contrast, affords one-way viewing and listening to a far greater degree. In the everyday medium, to obtain visual information is usually to make information available; in media spaces, making information available is an independent act from obtaining it.

There are both positive and negative aspects to the affordance for asymmetrical access offered by media spaces. These technologies afford spying: we don't need to let people know that we're watching or listening to them. But they also afford unobtrusive awareness: we don't have to interrupt people to find out if they're around or busy.

Design Implications

Attempts to change media spaces' affordances for asymmetrical access must find a balance in the tension between privacy and intrusion. One might reduce the ability to hide by enforcing "symmetry," so that if I can see or hear you then you can see or hear me. But symmetry is introduced at the expense of unobtrusive awareness. Because cameras and monitors are usually kept in fixed locations, there is no way to differentiate an appearance on them that is meant to be unobtrusive from one meant to capture attention: symmetry implies interruption.

A different approach to the problem is to provide control over who can access one's camera, allowing privacy to be maintained without interruption (5). In addition, another channel of feedback (e.g., notifications on the workstation, nonspeech audio cues) may be provided to indicate when someone is tapping the signal from one's camera. The combination of these two approaches seems to recreate the mixture of unobtrusive but knowable access found in the everyday world (7). But these sorts of solutions are post hoc additions to media space systems; asymmetrical access seems an enduring feature of media space technologies.

CONCLUSIONS

Although I have only touched on a few of the properties of media spaces and their implications for perception and interaction, it is evident from this analysis that the "space" created by audio-video technologies is significantly different from spaces as found in hallways, offices or meeting rooms. Compared to the everyday medium, the audio-video medium conveys a limited subset of visual and auditory information, prevents movement and exploration, and is often arbitrary and discontinuous. These properties shape the possibilities media spaces offer for collaboration.

Of course, saying that media spaces have different affordances than the everyday medium does not imply that collaboration is always "worse" in media spaces, merely that it is different. For instance, the ability to make one-way video connections has been useful in allowing unobtrusive "glances" into remote spaces. Similarly, the fact that audio from remote sites is shaped by the technology and thus easily distinguishable can be useful in allowing remote office mates to ignore conversations that do not concern them. Comparisons to the everyday world are useful in emphasizing the potential richness of interactions in media spaces. But an appreciation of the everyday should not interfere with an understanding of the new affordances offered by audio-video technologies. For as applications such as ClearBoard (14) demonstrate, media space technologies can potentially allow us to go "beyond being there" (13).

From this point of view, comparing media spaces to the everyday medium represents a first step towards understanding the unique affordances of media spaces. But it is a step that is useful in understanding collaboration using these technologies. Such an analysis is valuable in complementing those which stress social and cultural influences on interaction over media spaces. Because it emphasizes the relations between the "physics" of media spaces and the interactions they support, it highlights the fact that these social and cultural effects are situated, that they take place using a group of technologies that can be designed and changed, and points directly to relevant changes that might be made. For as I have hoped to make clear in this discussion, there are not only a number of differences between media spaces and the everyday medium, but a wealth of possibilities for the design of systems allowing richer, more intuitive forms of interaction and collaboration.

ACKNOWLEDGEMENTS
My thanks to Wendy Mackay, Allan Maclean, Don Norman, and Anne Schlottmann for their many helpful comments on this paper. I also thank Victoria Bellotti, Sara Bly, Paul Dourish, Steve Harrison, and Christian Heath for fruitful discussions about these ideas.

REFERENCES

1. Berry, D., Baron, R., Kean, K., and Misovich, S. (1991). Meaning-in-motion: Acquiring social knowledge from dynamic events. In Studies in Perception and Action: Posters presented at the 6th International Conference on Event Perception and Action. Beek, P., Bootsma, R., and van Wieringen, P. (eds). Rodopi, Amsterdam.

2. Bregman, A. (1990). Auditory scene analysis: The perceptual organization of sound. MIT Press.

3. Buxton, W., and Moran, T. (1990). EuroPARC's integrated interactive intermedia facility (iiif): Early experiences. In Proceedings of the IFIP WG8.4 Conference on Multi-User Interfaces and Applications (Herakleion, Crete, September 1990).

4. Buxton, W. & Sellen, A. (1992). Interfaces for multiparty videoconferencing. Technical Report No. EPC-92-129, Rank Xerox EuroPARC, Cambridge U.K.

5. Dourish, P. (1991). Godard: A flexible architecture for AV services in a media space. Technical Report No. EPC-91-134, Rank Xerox EuroPARC, Cambridge U.K.

6. Gaver, W. (1991). Technology affordances. In proceedings of CHI, 1991 (New Orleans, Louisiana, April 28 - May 2, 1991) ACM, New York, pp. 79 - 84.

7. Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K., and Buxton, W. (1992). Realizing a video environment: EuroPARC's RAVE system. In proceedings of CHI, 1992 (Monterey, California, May 3 - 7, 1992) ACM, New York, pp. 27 - 35.

8. Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin, New York.

9. Gibson, J. J. (1966). The senses considered as perceptual systems. Allen and Unwin, Ltd., London.

10. Heath, C. and Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. In proceedings of CHI, 1991 (New Orleans, Louisiana, April 28 - May 2, 1991) ACM, New York, 1991. pp. 99 - 103.

11. Hodges, L., and McAllister, D. (1987). True three-dimensional CRT-based displays. Information Display.

12. Hodges, M. (1986, May). The visual database project – Navigation. Internal Technical Report, Educational Services, Digital Equipment Corporation.

13. Hollan, J., & Stornetta, S. (1992). Beyond being there. In proceedings of CHI, 1992 (Monterey, California, May 3 - 7, 1992) ACM, New York, pp. 119 - 125.

14. Ishii, H., & Kobayashi, M. (1992). Clearboard: A seamless medium for shared drawing and conversation with eye contact. In proceedings of CHI, 1992 (Monterey, California, May 3 - 7, 1992) ACM, New York, pp. 525 - 532.

15. Jones, E., McLaurin, A., and Cathey, L. (1984). Visidep: Visual image depth enhancement by parallax induction. SPIE Advances in Technology, 457, 16 - 19.

16. Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T., and Wellman, B. (1991). Experiences in the use of a media space. In proceedings of CHI, 1991 (New Orleans, Louisiana, April 28 - May 2, 1991) ACM, New York, 1991. pp. 203 - 208.

17. Moran, T. P. and Anderson, R. J. (1990). The workaday world as a paradigm for CSCW design. In Proceedings of CSCW'90 (Los Angeles, California, October 1990). ACM, New York.

18. Overbeeke, C., and Stratmann, M. (1988). Space through movement. Non-published doctoral thesis, TU Delft, The Netherlands.

19. Root, R. W. (1988). Design of a multi-media vehicle for social browsing. In proceedings of CSCW'88 (Portland, Oregon, September 1988) ACM, New York. pp. 25-38.

20. Sellen, A., Buxton, W. & Arnotte, J. (1992). Using spatial cues to improve desktop video conferencing. 8 minute videotape presented at CHI'92 (Monterey, CA., May 2 - 7, 1992).

21. Stults, R. (1986). Media space. Xerox PARC technical report.

22. Suchman, L. (1987). *Plans and situated actions: The problem of human machine communication*. Cambridge University Press, Cambridge, U.K.

23. Tang, J., & Minneman, S. (1990). VideoDraw: A video interface for collaborative drawing. In proceedings of CHI, 1990 (Seattle, Washington, April 1 - 5, 1990) ACM, New York, pp. 313 - 320.

24. Travis, A., & Lang, S. (1991). The design and evaluation of a crt-based autostereoscopic 3-D display. In Proceedings of SID, Vol. 32/4.