RAPPORT DE RECHERCHE

L R I

TOWARDS MULTISCALE COMMUNICATIONS SYSTEMS

ROUSSEL N

Unité Mixte de Recherche 8623
CNRS-Université Paris Sud – LRI

03/2006

Rapport de Recherche N° 1439

CNRS – Université de Paris Sud
Centre d'Orsay
LABORATOIRE DE RECHERCHE EN INFORMATIQUE
Bâtiment 490
91405 ORSAY Cedex (France)

Towards multiscale communication systems

Nicolas Roussel
Projet In Situ, LRI & INRIA Futurs
Bâtiment 490, Université Paris-Sud
91405 Orsay Cedex, France
roussel@lri.fr

ABSTRACT
Forty years after AT&T's Picturephone, video is still mainly considered as a way to enhance audio communication in an attempt to reproduce face-to-face conditions. In a 1992 paper, Hollan and Stornetta argued that we should develop communication tools that go *beyond being there*. In this paper, I discuss two different interpretations of their analysis. I then propose the concept of *multiscale communication system* as an alternative approach for motivating further video-mediated communication research. The paper ends with a description of three systems that illustrate the concept.

Categories and Subject Descriptors
H.1.2 [Models & Principles]: User/Machine Systems – *Human factors*, H.4.3 [Communications Applications]: Computer conferencing, teleconferencing, and videoconferencing, H.5.2 [User Interfaces]: User-centered design, H.5.3 [Group and Organization Interfaces]: Collaborative computing.

General Terms
Design, Human Factors.

Keywords
Video-mediated communication, computer-mediated communication, multi-scale communication, coordination, communication, collaboration.

1. INTRODUCTION
Forty years after AT&T's Picturephone [24], video is still mainly considered as a way to enhance audio communication in an attempt to reproduce face-to-face conditions. Despite what futurologists predicted, videoconferencing has not replaced physical business travel. And although videoconferencing applications are available for free on the most popular software platforms (Microsoft Windows, Linux and Apple Mac OS X), few people actually use them on a regular basis. Oral and text-based communications, like email or instant messaging, remain by far the most popular solutions for asynchronous or distant communication.

CSCW researchers have investigated the reasons for the failure of traditional videoconferencing and proposed innovative uses of video for mediated communication (e.g. [9], [24], [26], [20]). This research somehow culminated in 1997 with the book *Video-mediated communication* edited by Finn, Sellen and Wilbur [10]. But strangely enough, the interest for innovative uses of video dropped off just as digital media and fast large area networks were becoming ubiquitous. As partly prophesied by Karam [34], the information superhighways killed most of the existing projects, based on analog media, like the US Interstate system killed Route 66:

> "People were not so likely to seek their fortune on the edge of a doomed road, and of those who were already there, fewer and fewer saw any value in upgrading or expanding or - sometimes - doing basic maintenance. After 1956, Route 66 remained important, but its importance was slowly moving away from the concrete toward the glorification of what the highway had been." (S.C. Kelly in *Route 66 - The highway and its people*, cited in [34])

Advances in media and networking technologies have made the implementation of video communication systems considerably easier. DSL technology brings to every home the bandwidth equivalent of a T-2 line, which AT&T used in the early 1970's to carry Picturephone signals. The new H.264 video codec promises "ultra-efficient, unprecedented video quality" [1]. But, as far as video-mediated communication (VMC) is concerned, these technologies are only used to create ultra-efficient Picturephones.

The original Picturephone was largely built on the assumption that the addition of sight to sound was both desirable and inevitable [24]. Although this assumption proved to be at least partly incorrect, few people question the motivations of current VMC research: what are we trying to achieve, why are we using video and how does this relate to other communication systems? In a quite influential paper from 1992, Hollan and Stornetta argued that rather than trying to imitate physical proximity, telecommunication research should develop tools that go *beyond being there* [16]. In this paper, I too question the goal of VMC research.

The paper is organized as follows. The next section discusses two different interpretations of Hollan and Stornetta's analysis. I then propose the concept of *multiscale communication system* as an alternative approach for motivating video-mediated communication research. Finally, I present three systems that illustrate this concept.

2. *BEYOND BEING THERE*

Being there is of course literally impossible. The expression refers to the concept of *presence*, which Lombard and Ditton define as "the perceptual illusion of nonmediation" [25]. *Being there* also refers to what has long been the main goal of VMC research: "achieving the level of information richness that we currently have in face-to-face interactions" to "interact with others that are far away just as we do with those that are near " [16].

The sense of presence, as defined by Lombard and Ditton, varies according to the media used. *Social presence* [39] and *media richness* [6] theories have been proposed and refined to characterize media, compare them and help people find the ones that maximizes efficiency or satisfaction for a particular task. Much of the research derived from these theories builds on the assumption that increased richness is linked to increased social presence [7]. As an example, the ability to support visual cues such as face expressions, eye contact, gestures or proximity is often said to increase the perceived sense of presence [39], i.e. to decrease the sense of mediation.

In their CHI 1992 paper [16], Hollan and Stornetta question the fundamental goal of telecommunication research. They suggest that instead of trying to imitate face-to-face communication, we should design tools that go *beyond being there*. The conclusion of their paper says:

> "we must develop tools that people would prefer to use even when they have the option of interacting in physical proximity (…) To create such tools, we suggest framing the problem in terms of needs, media, and mechanisms. The goal then becomes identifying needs which are not ideally met in the medium of physical proximity, and evolving mechanisms which leverage the strengths of the new medium to meet those needs".

This analysis has been quite popular and has inspired a number of systems. However, a broad look at these systems shows two very different interpretations, corresponding to different meanings of the word *beyond*: *greater than* and *other than*.

2.1 *Beyond* as *greater than*, *over*

Hollan and Stornetta ask the following question: "what would happen if we were to develop communication tools with a higher information richness than face-to-face?". Some people, notably from the Multimedia research community, take this as an invitation to pursue the prevailing technocentric approach to improve existing systems without questioning them. From this perspective, technical limitations still explain the relative failure of video-mediated communication, and further technical developments will help solve the remaining issues:

> "Why have current alternatives to physical travel such as video conferencing technology not replaced even more business travel? One hypothesis is that it is because such technology is not *immersive*." [22]

> "New sensors (e.g., touch, smell, taste, motion, etc.) and output devices (e.g., large immersive displays and personal displays integrated with eye glasses) offer the opportunity for more intimate and sensitive interaction with a remote environment. And, continued development of semiconductor technology will bring real-time three-dimensional virtual environments to every computing and communication platform. As one participant said, *interacting with a remote environment should be better than being there*." [37]

This approach focuses on *immersive*, *experiential* and *effective* telepresence[1], the proclaimed goal being to make the communication more *natural*, more *intuitive* and more *realistic*. Recent publications have indeed demonstrated impressive progress toward multiple viewpoints systems and immersive displays (e.g. blue-c [14], Twister [42], BiReality [22], Coliseum [3], MultiView [31]). But this approach has several problems.

First, in order to "beat the physical proximity", it pursues the same immediate goal of imitating it. The mark is just set higher than before, high fidelity sight and sound being considered as minimum requirements to be complemented with new technologies. Second, these new technologies often create their own problems, resulting in an endless quest for performance and fidelity:

> "Probably the biggest negative comment from users concerns the latency of the current system. One-way latency of the video is almost 700ms, so it is very noticeable. (…)We hope that the next generation of video compression cards will have reduced latency." [22]

> "Realistically, there are numerous developments that remain before this could be considered a viable alternative to travel for collaborative remote conferencing. Obvious improvements include increasing the frame rate, reducing latency, raising the quality at which people are displayed, and reconfiguring computation to enable more advanced features (such as head tracking)." [3]

The technocentric interpretation of *beyond being there* will hopefully lead to efficient high-fidelity conferencing systems. These systems might even provide services that remain valuable in the case of physical proximity, such as the ability to simultaneously manipulate shared artifacts. But their focus on synchronous face-to-face communication, combined with complex hardware and software requirements, will limit their use to formal, planed and highly engaged interactions.

2.2 *Beyond* as *other than*, *besides*

Formal interactions account for only part of typical group activity. Various studies have demonstrated the importance of more spontaneous, opportunistic, informal interactions [19]. Studies of face-to-face interactions have also shown the crucial role of visual information in monitoring and tracking availability among coworkers [43], which makes video an interesting technology for asynchronous or remote collaboration. Indeed, tracking the availability of other people for unscheduled communication is a typical need not ideally met in the physical world: how many visits to a colleague's office do you need to make before you find him or her available for discussion?

---

[1] These three terms were used for a series of workshops associated to the ACM Multimedia conference in 2002, 2003 and 2004.

Mediaspace studies [26] have investigated the potential uses of video to support collaborative activities ranging from casual awareness and informal talks – side-by-side interactions – to formal focused face-to-face communication. A variety of new services have been proposed. As an example, in addition to traditional videoconferencing, the RAVE mediaspace [12] made the following ones available: *background* (a view of a public area, used as the default connection), *glance* (a short one-way video connection), *sweep* (a series of glances), *office share* (a long-term audio and video link). These synchronous analog services were also complemented by the Portholes system [8] that presented regularly updated digitized images on the workstation screen.

While the technocentric approach focuses on the foreground activity made possible by physical proximity, most mediaspace studies were interested in the background and possibly unconscious forms of communication that go with it. One interesting finding, for example, is that in order to use it for background communication, one might need to reduce the information transmitted on a particular channel: Riesenbach [33] explains how lowering the resolution and frame rate of the permanent video connections of the Ontario Telepresence Project made them more socially acceptable by reducing the attention of the recipient and preserving the privacy of the sender.

A number of other techniques have been proposed to help mediaspace users find the appropriate tradeoff between awareness and privacy, including notification and control mechanisms [12], image and sound filtering [44], [40] and synthetic presentation of presence information [17]. Researchers later explored even more abstract, subtle and implicit forms of communication through lights, haptics and scent by taking advantage of a particular context[2], such as the intimate relation between two people [4][5][41].

But the most interesting aspect of mediaspace studies, I believe, is that they promoted the idea that a gradual engagement in communication is desirable and demonstrated that it is possible. In the next section, I will explain how this notion can be expanded to move on towards a new generation of communication systems.

3. TOWARDS MULTISCALE COMMUNICATION SYSTEMS

Although everyone seems to agree that we should develop systems that go *beyond being there*, not everyone seems to agree where to go. A technocentric interpretation of Hollan and Stornetta's analysis drives a number of researchers to a potentially endless quest for improving existing conferencing systems without questioning their goal. A more social approach, exemplified by mediaspace studies, reconsiders the problem of video-mediated communication and proposes a number of alternative services to traditional conferencing. But how do these services relate one to another? How do they relate to the many communication systems we already use? How can we structure their design space?

Gaver et al. proposed the *degree of engagement* and the *amount of planning* as two dimensions to analyze collaborative work [12]. The RAVE services (*background*, *sweep*, *glance*, *office share* and *vphone*) reflected this idea of having multiple degrees of engagement. Although less interested in the amount of planning, I believe the notion of selective engagement is an important one that can help structure the design space of communication systems. I also believe this notion could help users better choose the right communication service for a particular context.

Gaver et al. had a quite simple definition of the degree of engagement: "the extent to which a shared focus is involved". But other researchers have developed similar – although more refined – concepts. Fish et al. [11], for example, talked about the necessary balance between *accessibility* (access to others), *privacy* (control over the available information about oneself) and *solitude* (control over others' intrusion in one's space and time). Greenhalgh and Benford [13] proposed the notions of *nimbus* (one's manifestation or observability) and *focus* (one's allocation of attention) that users could explicitly manipulate, an idea recently applied to video communication in the Community Bar awareness system [27].

The Community Bar presence item proposes six levels of details (i.e. degrees of engagement) based on combinations of the following attributes: a two-state color activity indicator, the user name, a status message, a static picture, a webcam snapshot and a fast frame rate video connection. Sliders make it possible to control one's focus on each of the other users. A nimbus slider also makes it possible to specify a level of detail which others can only see up to, but not beyond (using their focus slider). Although this system makes use of video, this use is quite limited. One reason for this is probably that the Community Bar presence item, as the name suggests, is a tool for presence awareness, not something that aims at supporting the full range of collaborative activities.

Previous research on video-mediated communication has demonstrated that video, through its different forms, can be used to support a wide range of activities. Mediaspaces are probably the closest attempt at creating a single system to support the full range of these activities. I believe this should be the goal of future VMC research and development. This goal is not new. It was one of RAVE designers' for example. But it seems to have been abandoned on the way. The following problems, in particular, should be explored:

- Beyond[3] snapshots and full-rate: How can we use video to implement degrees of engagement other than static pictures and high-quality streams? How many degrees can we create? Can we create a continuum of degrees?
- Beyond buttons, sliders and labels: How can we move from one degree to another? Can we avoid dialog boxes? How do we perceive a remote person's degree? How can we negotiate degrees with a remote person?
- Beyond video: How can we combine video with other media? (e.g. email, the telephone, instant messaging systems, the Web)

As illustrated by the case of the permanent connections of the Ontario Telepresence Project, the level of detail of an image

[2] The idea that taking a particular context into account can help reduce a message while preserving its general meaning is not new. According to [30], Victor Hugo was on vacation when his book Les Misérables was published. Curious to know how it was doing, he sent a telegram to his publisher, reading simply "?". The publisher replied in an equally short way: "!".

[3] Use of the word beyond is not coincidental. As we have seen, it leaves some space for reader interpretation…

stream is probably related to the associated degree of engagement: the bigger, the more colorful, the sharper and the more frequent the images are, the more they expose the person they show and will probably attract the attention of the person that sees them. In addition to these attributes, other characteristics of an image stream could be manipulated to alter the associated engagement degree. As illustrated by Figure 1, spatial filtering techniques, for example, can be used to degrade images [44] while temporal compositions can provide awareness of past activity [17][15]. One could certainly imagine other spatial techniques to enrich the video as well as temporal techniques to degrade it (e.g. by introducing a controlled delay).

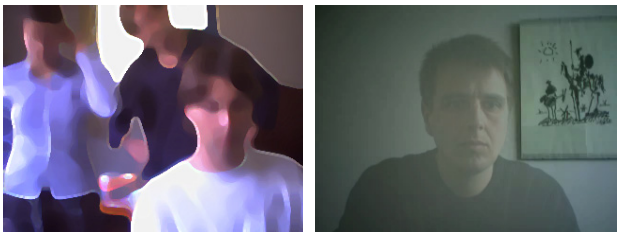

Figure 1. Degrading (left) or enriching (right) a video stream.

Transitions between engagement degrees pose two kinds of problems. First, new interaction techniques will be required to specify a desired degree. Interaction will need to be as direct and concise as possible since managing one's communications should not become a primary activity itself. The camera, in this context, is probably an interesting input device and other sensors might also be useful. In addition, feedback mechanisms (e.g. animations) will be required to make the user aware of the transitions initiated by remote partners. Combining the video system with other communication tools will again require the design of appropriate interaction techniques and feedback mechanisms. As an example, one might want to temporarily use a mobile phone as an additional audio channel to an existing video communication. Combining synchronous and asynchronous communication might also pose some interesting problems.

To summarize:

1. We should develop video communication systems that support a variable degree of engagement.
2. These systems should support smooth transitions between degrees.
3. They should also support smooth integration with other media.

The notion of *multiscale world* is defined by Jul and Furnas as a world "in which information can exist at multiple levels of detail" [23]. The degree of engagement, as I see it, somehow corresponds to the level of detail of the communication. Therefore, I propose to use the term *multiscale communication system* to designate a communication system that supports a variable degree of engagement. Smooth transitions between degrees of engagement correspond to smooth variations of the level of detail. In Zoomable User Interface terms [32], we might call them *continuous zooming*. Enriching or degrading a video stream can change both its meaning and level of detail and might thus be considered as the equivalent of a *semantic zoom*.

4. EXAMPLES

I will now present three video systems that partly illustrate the concept of multiscale communication system I just introduced. This section will complement previously published descriptions of the systems [35][18][36] by emphasizing their relation to the concepts of variable degree of engagement, smooth transitions between degrees and integration with other media.

4.1 VideoServer

VideoServer [35] was designed as a tool to support the creation of a highly tailorable Web-based media space. It is a personal HTTP server that allows a user to make live or pre-recorded images and video streams accessible to other users through simple URLs (Figure 2).

```
http://server/grab/video
http://server/push/video?framerate=5&size=QSIF
http://server/push/video?framerate=25&size=SIF
```

Figure 2. VideoServer URLs requesting a single image, a low frame rate 160x120 video and a high frame rate 320x240 video (all images are captured in real-time).

In addition to other specific protocols, videoServer is able to transmit video data to client applications on the HTTP connection itself. In this case, single images are sent as JPEG-compressed data, which can be displayed by any HTML rendering engine in place of an ordinary JPEG image, without any plug-in. Video streams are sent as a server-pushed series of JPEG-compressed images that some HTML renderers can also display in place of an ordinary image[4].

By using URLs such as those of Figure 2, users can easily integrate live images and video streams into email messages (Figure 3, left) and existing or new HTML documents (Figure 3, right). An interesting use of this feature is to include a live snapshot of one's office in one's email signature or in a Web page that shows your contact information so that people who want to reply to one of your emails or to call you can see if you're available for discussion. This ability to provide access to synchronous video services from Web publishing and email, two rather low paced asynchronous media, is a good example of the cross-media integration mentioned in the previous section.

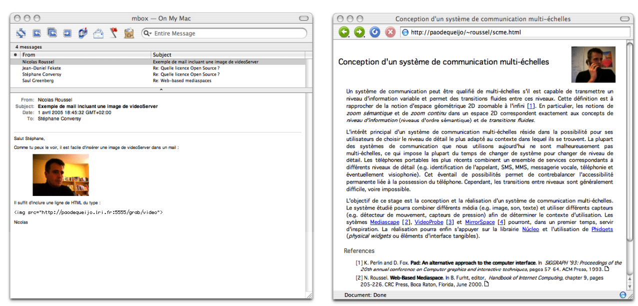

Figure 3. Live videoServer images displayed in Apple's Mail application and the Camino Web browser. Images are captured and transmitted every time the HTML message or document is rendered by the application.

Awareness views similar to Portholes are easily created by including the images from several servers in a single document and using a timer to reload it at regular intervals. Basic image and

[4] Gecko, the Mozilla HTML rendering engine is one of them. Mozilla applications such as Camino and Firefox (two Web browsers) or Thunderbird (an email client) can thus display videoServer streams without any plug-in.

video services can also be combined to support more complex interactions. A few lines of JavaScript, for example, can turn a static picture into a medium frame-rate video (e.g. 15 fps) when the mouse moves over it and pop up a new window displaying a high frame-rate and resizable stream when one clicks on it (Figure 4). While a previous study suggested that people have difficulty extracting information from snapshots unless the resolution is at least 128x128 pixels [21], experience with this three-scale focus control indicates that snapshot resolution can be reduced up to 80x60 as the ability to turn them into video streams helps resolve ambiguities.

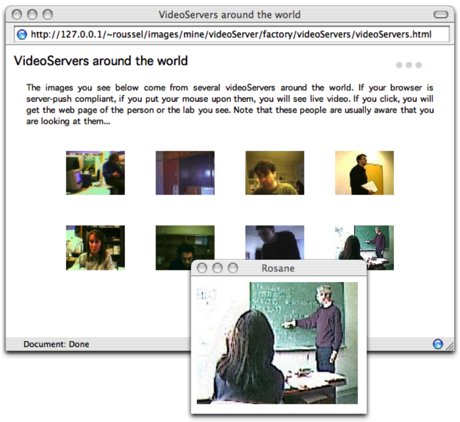

Figure 4. Focus control: from a low resolution snapshot in a Portholes-like awareness view to a high frame rate independent video that the user can freely move and resize.

As most mediaspaces and unlike webcam software, videoServer provides users with notification and access control mechanisms. For every request it receives, it executes an external program (e.g. a Python script) with arguments indicating the name of the remote machine, possibly the remote user's login name, the resource that led to the server (the HTTP referrer) and a description of the requested service. The external program uses this contextual information to generate auditory or on-screen notifications (Figure 5) and sends back to the server a description of the service to be executed. This description can be inferred from a set of pre-defined rules or negotiated with the user through some interactive dialog.

Figure 5. Sample on-screen notification.

An important feature of videoServer's control mechanism is that the external program is not limited to a binary accept/refuse choice but can freely redefine the service to be executed. It can for example request that a spatial filter be applied on the images, which the remote person will probably notice (Figure 6). It can redirect the client to another server. But it can also substitute a pre-recorded image or sequence to the live stream. This feature proved particularly useful as it supports the creation of ambiguities and stories [2]. Seeing the image 3 of Figure 6, for example, one might assume that the remote person is absent. Yet seeing this particular image too often might indicate that he simply doesn't want us to know if he's there. Seeing the image 4 might indicate that he'll be away for some time.

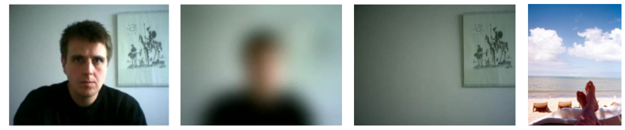

Figure 6. Nimbus control: image captured by the camera, filtered image, ambiguous pre-recorded image and explicit absence indicator.

As we have seen, videoServer makes it possible to combine synchronous video services with asynchronous communication via email or Web pages. It also provides users with flexible and powerful scripting mechanisms to control their focus and nimbus. Mastering these mechanisms, however, requires some programming knowledge. I will now describe two other systems that explored more direct and intuitive ways of varying one's degree of engagement.

4.2 VideoProbe

VideoProbe is one of the technology probes created for interLiving [18], a multi-disciplinary research project focused on the design of new technologies to support communication among family members located in different households. VideoProbe allows a group of people to share their daily lives by exchanging pictures. It physically consists in a box containing a screen, two speakers and a camera connected to a separate computer, itself connected to the Internet (Figure 6). A specific software running analyzes the images captured by the camera in real-time and decides when a picture should be taken and transmitted to similar probes in other households (only pictures are exchanged, not video streams).

Figure 7. VideoProbe.

As long as the scene observed by the camera doesn't change, the screen stays blank (Figure 8, image 1). If a change is detected, the software gradually displays the captured images, turning the screen into a mirror (Figure 8, images 2 and 3). If the same observed change persists more than three seconds, a picture is automatically transmitted to the other videoProbes. A growing translucent rectangle indicates the remaining time (Figure 8, images 4 and 5): when the rectangle reaches the full size of the video frame, an auditory cue is played, the picture is taken, displayed bigger and correctly oriented for three seconds (Figure 8, image 6) and then transmitted to the other videoProbes. If the scene doesn't change anymore, the screen gradually returns to its blank state. Otherwise, new pictures can be taken and transmitted as just described.

Figure 8. Transitions between the *sleep mode* (1), the *mirror mode* (2 to 5) and the *picture transmission mode* (6).

A remote control allows to switch the system into a browsing mode that shows the pictures taken by all the connected videoProbes. Within this mode, users can delete selected pictures or save them in a persistent album. Pictures not saved in the album gradually loose their colors and contrast and eventually disappear from the browsing interface after a few days (Figure 9).

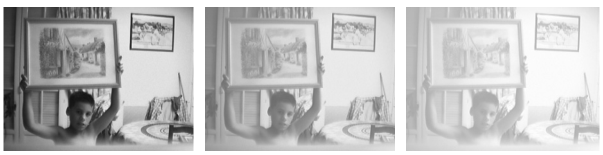

Figure 9. Picture aging in the *browsing mode* (the actual process takes about five days).

The smooth transitions between the different operation modes of videoProbe play an essential part in making the interaction simple, quick and easy. The combined use of movement detection and delayed picture taking allows to quickly switch the device from an idle state to one where it is ready to communicate while still offering an easy way to back off, as continuous move prevents the system from taking pictures. This was quickly understood by users without formal training and even turned into a little game which goal was to take a picture of an empty room, i.e. move outside the field of view of the camera at the exact moment when the picture was taken (which is in fact particularly hard to achieve).

VideoProbe supports both *explicit* and *implicit* forms of communication. The explicit form takes place when the user is consciously using the system to transmit a particular picture (Figure 10, left). The implicit form typically takes place when someone enters the room and stays there for some reason but does not pay attention to the device (Figure 10, right). In that case, the persistent scene change triggers the taking of a picture and its transmission but the user usually becomes aware of it only when he or she hears the auditory notification.

The implicit form of communication proved very useful for maintaining group awareness as it usually produces pictures that users would not or could not take themselves. At the same time, because of its motion-based control, videoProbe was perceived as less intrusive and more flexible than a purely time-based approach that would have taken pictures at regular intervals. User motion indirectly determines the rate at which the system transmits images. And although the maximum rate is quite limited (about 10 to 15 frames per second), the system was sometimes used while discussing over the phone as an acceptable replacement for a videoconferencing service. This particular example again illustrates how a single video communication system can support a variable degree of engagement ranging from asynchronous communication to synchronous one.

Figure 10. Explicit and implicit uses of videoProbe.

The process of picture taking is a slow one during which the presentation of the images captured by the camera is gradually transformed until they reach the state where one will be taken and transmitted: images first fade in and are then gradually covered by the translucent rectangle indicating the remaining time. The gradual degradation of the pictures that have been received follows the same approach: pictures don't disappear suddenly but fade away. As users had the opportunity of canceling the picture taking process, they also have the opportunity to literally save the taken pictures. This shows how the notion of variable engagement can even be used in the case of purely asynchronous communication

Our next example explores further the notion of gradual and intuitive engagement in synchronous communication.

4.3 MirrorSpace

MirrorSpace [36] is another video communication system designed for the interLiving project. Whereas existing video systems usually create a shared space corresponding to a particular interpersonal distance, the goal of MirrorSpace was instead to create a continuum of space that would allow a variety of interpersonal relationships to be expressed. As the name suggests, MirrorSpace relies on a mirror metaphor (Figure 11).

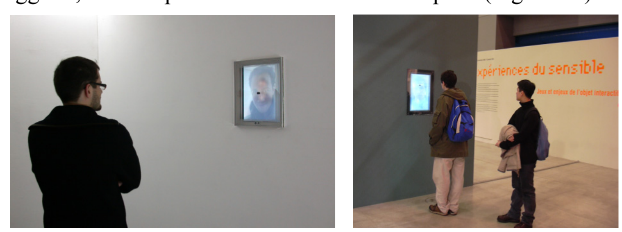

Figure 11. MirrorSpace.

Live video streams from all places connected through the system are superimposed on a single display on each site so that people see their own reflection combined with the ones of remote persons. In order to support intimate forms of communication where people might want to look into each other's eyes, the camera has been placed right in the middle of the screen. This setup allows users to come very close to the camera while still being able to see the remote people and interact with them. MirrorSpace also includes an ultrasonic proximity sensor that measures the distance to the closest object or person in front of it. A blur filter is applied on the images displayed to visually express a distance computed from the local and remote sensor values. Blurring distant objects and people allows one to perceive their movement or passing with a minimum involvement. It also offers a simple way of initiating or avoiding a change to a more engaged form of communication by simply moving closer (Figure 12) or further away.

Figure 12. Reducing the blur effect by moving closer.

MirrorSpace has been presented to the public in several art exhibitions. In one exhibition, two prototypes were placed inside a 3x3m cubicle that enabled people to directly see and hear each other. In another exhibition, they were completely isolated from each other. In other cases, they were set up in a way that people could hear without being able to see each other directly (e.g. separated by a thin wall or placed back to back). Several hours of video were shot during the exhibitions and later analyzed. Although the context of an art exhibition is somewhat particular, several interesting observations were made that are probably inherent to the system.

Proximity sensing combined with blur filtration helps creating an intimate relationship between users and the system. People seem to like the idea that the system is reacting to them and not just taking images from them, that they are in control and not only the subject. When they see another person appearing next to them on the screen, many people turn over, looking for that person behind them. As previously reported by other studies (e.g. [28]), this shows that the superposition of images creates a strong sense of shared space. The particular placement of the camera, which allows people to come really close to it, turns this shared space into an intimate one. Many people get surprised and even disturbed by this intimacy when a stranger appears to close to them on the screen. But proximity sensing and blur filtration allows them to simply step back to disengage and alter the display.

A recent study showed that blur filtration fails at providing an obfuscation level that could balance privacy and awareness for home situations [29]. Yet, I strongly believe that this type of filtering is still valuable. Not because of what it tries to remove, but because of what it adds: the filter shows the remote people that we don't want them to observe. Of course, there's no guarantee that they won't, but we know that they know they're not supposed to do so. The stronger the filter, the stronger we insist on the fact that it is socially unacceptable for them to observe. Blur filtration can be seen as a way to enrich the video communication to indicate the desire for a lesser-engaged form of communication. The fact that it does not necessarily enforce this lighter form of communication leaves room for negotiation between people.

In MirrorSpace, the strength of the blur effect applied on an image is computed from the proximity sensor values of all the connected devices. In the simplest case, the strength is the result of a transfer function applied to the local sensor value. The transfer function makes it possible to adapt the system to the particular geometry of the room where it has been installed. A more interesting case is when the blur effect applied on the image of a remote person is computed from both the local and remote sensor values. Using the sum of these values, for example, makes it possible for two people, Chris and Steve for example, to negotiate a common degree of engagement:

1. If Chris moves closer to the device, the image of Steve on his screen and his own image on Steve's screen will get sharper
2. Steven will then be able to accept the new engagement degree, to increase it further by also moving closer to the device or to go back to the previous state by stepping back

This example shows that it is possible to create communication systems that uses at least part of the physical body language to negotiate a common engagement degree in a way similar to what had been proposed by Greenhalgh and Benford for virtual environments [13].

5. CONCLUSION

In this paper, I have introduced the concept of multiscale communication system as an alternative approach for motivating video-mediated communication research. This approach aims at creating systems that support a variable degree of engagement, smooth transitions between degrees and integration with other media.

The multiscale approach to communication presented in this paper is not limited to video. I also believe that more parallels could be found between Computer-Mediated Communication and Information Visualization. As an example, Shneiderman's visual information seeking mantra [38] seems particularly relevant to the way we usually engage in a communication with another person:

Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand

After all, isn't communication the process of exchanging information?

6. REFERENCES

[1] Apple. QuickTime and MPEG-4: Now Featuring H.264. Technology Brief, Apple, April 2005.

[2] P. M. Aoki and A. Woodruff. Making space for stories: ambiguity in the design of personal communication systems. In *Proceedings of CHI '05*, pages 181–190, 2005. ACM Press.

[3] H. H. Baker, N. Bhatti, D. Tanguay, I. Sobel, D. Gelb, M. E. Goss, W. B. Culbertson and T. Malzbender. Understanding performance in coliseum, an immersive videoconferencing system. *ACM Transactions on Multimedia Computing, Communications and Applications*, 1(2):190–210, 2005.

[4] S. Brave, H. Ishii and A. Dahley. Tangible interfaces for remote collaboration and communication. In *Proceedings of CSCW '98*, pages 169–178, 1998. ACM Press.

[5] A. Chang, B. Resner, B. Koerner, X. Wang and H. Ishii. LumiTouch: an emotional communication device. In *CHI'2001 extended abstracts*, pages 313-314, 2001. ACM Press.

[6] R. Daft and R. Lengel. Information richness: a new approach to managerial behavior and organizational design. In L.

Cummings and B. Staw, editors, Research in organizational behavior 6, pages 191-233, 1984. JAI Press.

[7] A. R. Dennis and J. S. Valacich. Rethinking Media Richness: Towards a Theory of Media Synchronicity. In *Proceedings of HICSS '99*, pages 1017, 1999. IEEE Computer Society.

[8] P. Dourish and S. Bly. Portholes: Supporting Awareness in a Distributed Work Group. In *Proceedings of CHI'92*, pages 541-547, 1992. ACM Press.

[9] C. Egido. Videoconferencing as a Technology to Support Group Work: A Review of its Failure. In *Proceedings of CSCW'88*, pages 13-24, September 1988. ACM Press.

[10] K. Finn, A. Sellen and S. Wilbur, editors. Video-Mediated Communication, Lawrence Erlbaum, April 1997. 584 pages.

[11] R. S. Fish, R. E. Kraut, R. W. Root and R. E. Rice. Video as a technology for informal communication. *Communications of the ACM*, 36(1):48–61, 1993.

[12] W. Gaver, T. Moran, A. MacLean, L. Lövstrand, P. Dourish, K. Carter and W. Buxton. Realizing a Video Environment: EuroPARC's RAVE System. In *Proceedings of CHI'92*, pages 27-35, 1992. ACM Press.

[13] C. Greenhalgh and S. Benford. MASSIVE: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction*, 2(3):239–261, 1995.

[14] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. Vande Moere and O. Staadt. blue-c: A spatially immersive display and 3D video portal for telepresence. In *Proceedings of SIGGRAPH 2000*, pages 819–827, 2003.

[15] C. Gutwin. Traces: Visualizing the Immediate Past to Support Group Interaction. In *Proceedings of Graphics Interface 2002*, pages 43–50, May 2002.

[16] J. Hollan and S. Stornetta. Beyond being there. In *Proceedings of CHI'92*, pages 119-125, 1992. ACM Press.

[17] S. E. Hudson and I. Smith. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In *Proceedings of CSCW'96*, pages 248-257, November 1996. ACM Press.

[18] H. Hutchinson, W. Mackay, B. Westerlund, B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, B. Eiderbäck, S. Lindquist and Y. Sundblad. Technology Probes: Inspiring Design for and with Families. In *Proceedings of CHI 2003*, pages 17-24, April 2003. ACM Press.

[19] E. A. Isaacs, S. Whittaker, D. Frohlich and B. O'Conaill. Informal communication re-examined: New functions for video in supporting opportunistic encounters. In K. Finn, A. Sellen and S. Wilbur, editors, *Video-mediated communication*, Lawrence Erlbaum Associates, 1997.

[20] H. Ishii. Integration of Shared Workspace and Interpersonnal Space for Remote Collaboration. In M. Beaudouin-Lafon, editor, *Computer-Supported Co-operative Work*, Trends in Software Series, John Wiley &Sons Ltd, 1999.

[21] B. Johnson and S. Greenberg. Judging People's Availability for Interaction from Video Snapshots. In *Proceedings of HICSS'99*, Distributed Group Support Systems Minitrack, January 1999. IEEE Press.

[22] N. Jouppi, S. Iyer, S. Thomas and A. Slayden. BiReality: mutually-immersive telepresence. In roceedings of Multimedia'04, pages 860–867, 2004. ACM Press.

[23] S. Jul and G. W. Furnas. Critical zones in desert fog: aids to multiscale navigation. In Proceedings of UIST '98, pages 97–106, 1998. ACM Press.

[24] K. Lipartito. PicturePhone and the Information Age: The Social Meaning of Failure. *Technology and Culture*, 44(1):50-81, January 2003.

[25] M. Lombard and T. Ditton. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, 3(2), September 1997.

[26] W. Mackay. Media Spaces: Environments for Informal Multimedia Interaction. In M. Beaudouin-Lafon, editor, *Computer-Supported Co-operative Work*, Trends in Software Series, John Wiley &Sons Ltd, 1999.

[27] G. McEwan and S. Greenberg. Supporting social worlds with the community bar. In *Proceedings of GROUP '05*, pages 21–30, 2005. ACM Press.

[28] O. Morikawa and T. Maesako. HyperMirror: toward pleasant-to-use video mediated communication system. In *Proceeding of CSCW'98*, pages 149-158, 1998. ACM Press.

[29] C. Neustaedter, S. Greenberg and M. Boyle. Blur Filtration Fails to Preserve Privacy for Home-Based Video Conferencing. *ACM Transactions on Computer-Human Interaction*, in press.

[30] T. Nørretranders. The user illusion: cutting consciousness down to size, Penguin books, 1991.

[31] D. Nguyen and J. Canny. MultiView: spatially faithful group video conferencing. In *Proceedings of CHI '05*, pages 799–808, 2005. ACM Press.

[32] K. Perlin and D. Fox. Pad: An alternative approach to the computer interface. In Proceedings of SIGGRAPH '93, pages 57-64, 1993. ACM Press.

[33] R. Riesenbach. Less is more (more or less...). White Paper, Telepresence Systems, 1996. http://www.telepres.com/less_is_more.shtml.

[34] R. Riesenbach, W. Buxton, G. Karam and G. Moore. Ontario Telepresence Project. Final report, Information technology research centre, Telecommunications research institute of Ontario, March 1995.

[35] N. Roussel. Mediascape: a Web-based Mediaspace. *IEEE Multimedia*, 6(2):64-74, April-June 1999.

[36] N. Roussel, H. Evans and H. Hansen. Proximity as an interface for video communication. *IEEE Multimedia*, 11(3):12-16, July-September 2004.

[37] L. A. Rowe and R. Jain. ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications and Applications*, 1(1):3–13, 2005.

[38] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of VL '96*, pages 336-343, 1996. IEEE Computer Society.

[39] J. Short, E. Williams and B. Christie. The Social Psychology of Telecommunications, Wiley, New York, 1976.

[40] I. Smith and S. Hudson. Low disturbance audio for awareness and privacy in media space applications. In Proceedings of Multimedia'95, pages 91–97, 1995. ACM Press.

[41] R. Strong and W. Gaver. Feather, scent and shaker: supporting simple intimacy. In *Proceedings of CSCW'96*, pages 29-30, November 1996. ACM Press.

[42] K. Tanaka, J. Hayashi, M. Inami and S. Tachi. TWISTER: An Immersive Autostereoscopic Display. In *Proceedings of VR '04*, pages 59-66, 2004. IEEE Computer Society.

[43] S. Whittaker. Video as a technology for interpersonal communications: a new perspective. In A. A. Rodriguez and J. Maitan, editors, *Proceedings of Multimedia Computing and Networking 1995*, volume 2417, pages 294-304, 1995. SPIE.

[44] Q. Zhao and J. Stasko. Evaluating Image Filtering Based Techniques in Media Space Applications. In *Proceeding of CSCW'98*, pages 11-18, 1998. ACM Press.