

# Beyond “Beyond Being There”: Towards Multiscale Communication Systems

Nicolas Roussel  
LRI (Univ. Paris-Sud – CNRS), INRIA  
Bâtiment 490, Université Paris-Sud  
91405 Orsay Cedex, France  
roussel@lri.fr

Sofiane Gueddana  
LRI (Univ. Paris-Sud – CNRS), INRIA  
Bâtiment 490, Université Paris-Sud  
91405 Orsay Cedex, France  
gueddana@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, we discuss two different interpretations of their analysis. We then propose the concept of *multiscale communication* as an alternative approach for motivating telecommunication research, an approach that aims at creating systems that support a variable degree of engagement, smooth transitions between degrees and integration with other media. Finally, we present three video systems from which the multiscale communication concept emerged and that partly illustrate it.

## Categories and Subject Descriptors

H.4.3 [Communications Applications]: Computer conferencing, teleconferencing, and videoconferencing; H.1.2 [Models & Principles]: User/Machine Systems - Human factors; 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, multiscale communication, coordination, communication, collaboration

## 1. INTRODUCTION

Forty years after AT&T’s Picturephone [29], 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.

Historians and CSCW researchers have investigated the reasons for the failure of traditional videoconferencing (e.g. [29, 12]) and proposed innovative uses of video for mediated communication (e.g. [4, 24]). This research somehow culminated in 1997 with the book *Video-mediated communication* edited by Finn, Sellen and Wilbur [13]. 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 [40], 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 [40])

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. New video codecs such as H.264 promise “ultra-efficient, unprecedented video quality” [2]. 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 [29]. 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* [20]. In this paper, we too question the goal of video-mediated communication and telecommunication research in general.

The paper is organized as follows. The next section discusses two different interpretations of Hollan and Stornetta’s analysis. We then propose the concept of *multiscale communication* as an alternative approach for motivating telecommunication research, an approach that aims at creating systems that support a variable de-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM’07, September 23–28, 2007, Augsburg, Bavaria, Germany.  
Copyright 2007 ACM 978-1-59593-701-8/07/0009 ...\$5.00.

gree of engagement, smooth transitions between degrees and integration with other media. Finally, we present three video systems from which the multiscale communication concept emerged and that partly illustrate it.

## 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” [30]. *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” [20].

The sense of presence, as defined by Lombard and Ditton, varies according to the media used. *Social presence* [45] and *media richness* [9] 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 [10]. 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 [45], i.e. to decrease the sense of mediation.

In their CHI 1992 paper [20], 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:

“If we ever hope to solve the telecommunication problem, we must develop tools that people would prefer to use even when they have the option of interacting in physical proximity as they have heretofore. To do that requires tools that go *beyond being there*. 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”: the ultra-high fidelity approach

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 technology-driven 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:

“Systems rarely support more than two participating sites, and specially equipped rooms are often required. Frame rates and image quality lag expectations, and the resulting experience is of blurry television watching rather than personal interchange. Our intention in

Coliseum has been to push the envelope in all dimensions of this technology – display frame rate and resolution, response latency, communication sensitivity, supported modalities, etc.” [3]

“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*.” [26]

“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*.” [43]

This approach focuses on *immersive*, *experiential* and *effective* telepresence<sup>1</sup>, 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 [17], Twister [48], BiReality [26], Coliseum [3], MultiView [35]). But this approach has several problems. First, it focuses on media and mechanisms but often neglects user needs. Second, 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. Lastly, these new technologies often create their own problems, resulting in an endless quest for performance and fidelity:

“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- ing).” [3]

“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.” [26]

The ultra-high fidelity approach will hopefully lead to efficient lifelike 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, planned and highly engaged interactions.

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

## 2.2 “Beyond” as “other than”: the high-diversity approach

Formal interactions account for only part of typical group activity. Various studies have demonstrated the importance of more spontaneous, opportunistic, informal interactions [23]. Studies of co-located interactions have also shown the crucial role of visual information in monitoring and tracking availability among coworkers [49], 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?

Mediaspace studies [4, 31] 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 communications. A variety of new services have been proposed. As an example, in addition to traditional video-conferencing, the RAVE mediaspace [15] 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 [11] that presented regularly updated digitized images on the workstation screen.

While the ultra-high fidelity 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 [39] 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 trade-off between awareness and privacy, including notification and control mechanisms [15], image and sound filtering [46, 50] and synthetic presentation of presence information [21]. Researchers later explored even more abstract, subtle and implicit forms of communication through lights, haptics and scent by taking advantage of a particular context<sup>2</sup>, such as the intimate relation between two people [47, 5, 6]. But the most interesting aspect of mediaspace studies, we 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, we 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. An ultra-high fidelity interpretation of Hollan and Stornetta’s analysis drives a number of researchers to a potentially endless

<sup>2</sup>The idea that taking a particular context into account can help reduce a message while preserving its general meaning is not new. According to [36], 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: “!”.

quest for improving existing conferencing services without questioning their goal. A more social approach, exemplified by mediaspace studies, reconsiders the problem of video-mediated communication and proposes an increasing 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, like email, instant messaging or the telephone? Can we structure their design space in a way that includes both high fidelity systems for face-to-face interactions as well as subtle, implicit and abstract forms of communications?

Gaver et al. proposed the *degree of engagement* and the *amount of planning* as two dimensions to analyze collaborative work [15]. 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, we believe the notion of selective engagement is an important one that can help structure the design space of communication systems. We 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 for the degree of engagement: “the extent to which a shared focus is involved”. Other researchers have developed similar – although more refined – concepts. Fish et al. [14], 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 [16] also suggested that users should be able to separately control their *nimbus* (one’s manifestation or observability) and *focus* (one’s allocation of attention). This idea was notably applied to video communication inside a Collaborative Virtual Environment [38] and more recently in the Community Bar awareness system [32]. Based on these different concepts, our own definition for the degree of engagement is the following: “the extent to which users are ready to expose themselves and open to others”.

The Community Bar presence item proposes six 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.

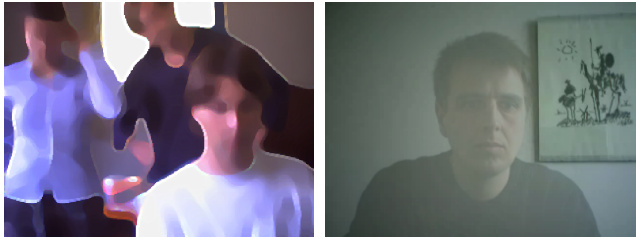
Previous research on video-mediated communication has demonstrated that video, through its different forms, can be used to support a wide range of activities. Mediaspaces are probably the closest attempt at creating a single system to support the full range of these activities. We believe this should 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<sup>3</sup> 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?

<sup>3</sup>Use of the word beyond is not coincidental. As we have seen, it leaves some space for reader interpretation. . .

- Beyond buttons, sliders and labels: How can we move from one degree to another? How can we perceive a remote person's degree? How can we negotiate degrees with remote people? Can we avoid explicit dialog boxes and support more intuitive interactions?
- Beyond video: How can we combine video with other media? (e.g. email, the telephone, instant messaging, the Web)

As illustrated by the case of the permanent connections of the Ontario Telepresence Project, the level of detail of an image 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, filtering techniques can be used to degrade images [50] while temporal compositions can provide awareness of past activity [21, 19]. One could certainly imagine other image filtering techniques to enrich the video as well as temporal techniques to degrade it (e.g. by introducing a controlled delay). More subtle filters could also eliminate some details while enhancing others [28, 7].



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

Transitions between engagement degrees pose two kinds of problems. First, new interaction techniques are required to specify the desired degree. These techniques 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. Feedback mechanisms such as animations can probably help make the user aware of the transitions initiated by remote partners. Combining the video system with other communication tools again requires 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 also poses some interesting problems.

To summarize:

- We believe our goal should be to develop new communication systems that support a variable degree of engagement.
- These systems should support smooth transitions between degrees. They should also support smooth integration with other media or communication systems.
- Video is a good starting point, as it has already been shown to support a wide range of collaborative activities and can also be used as an input channel for Human-Computer interaction.

A *multiscale world* is defined by Jul and Furnas as a world “in which information can exist at multiple levels of detail” [27]. The degree of engagement, as we see it, somehow corresponds to the level of detail of the communication. Therefore, we 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 [37], 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

We will now present three video systems that partly illustrate the concept of multiscale communication we just introduced. This section will complement previously-published descriptions of these systems by emphasizing aspects of their design and use that are related to the concepts of variable degree of engagement, smooth transitions between degrees and integration with other media.

The first system, VideoServer, shows how focus and nimbus control mechanisms can be used to combine synchronous video services, and how these services can be integrated with asynchronous text-based communication to support lightweight coordination. The second system, VideoProbe, shows how activity sensing techniques can be used to support both implicit (i.e. peripheral) and explicit (i.e. highly engaged) interactions, and transitions from asynchronous to almost synchronous communication. The last system, MirrorSpace, further illustrates the use of sensing techniques to support the implicit control of the degree of engagement in a synchronous communication through the usual body language.

It is important to understand that these systems were not designed according to the multiscale communication approach, but that the concept emerged from them. In other words, these examples are not here to “validate” the concept but rather to explain its genesis.

### 4.1 VideoServer

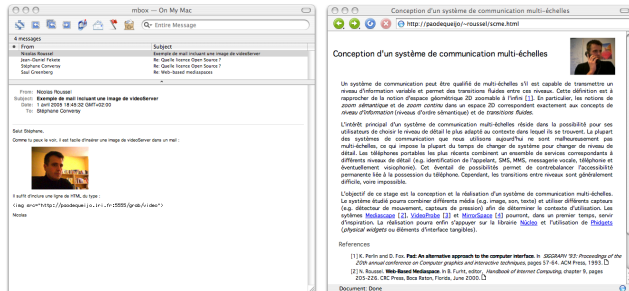
VideoServer [41] was designed as a tool to support the creation of a highly tailorable Web-based mediaspace. It is a personal HTTP server that allows a user to make live or pre-recorded images and video streams accessible to other users through simple URLs (Figure 2). In addition to other specific protocols, 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<sup>4</sup>.

```
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).**

<sup>4</sup>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.

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 that users developed 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 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.

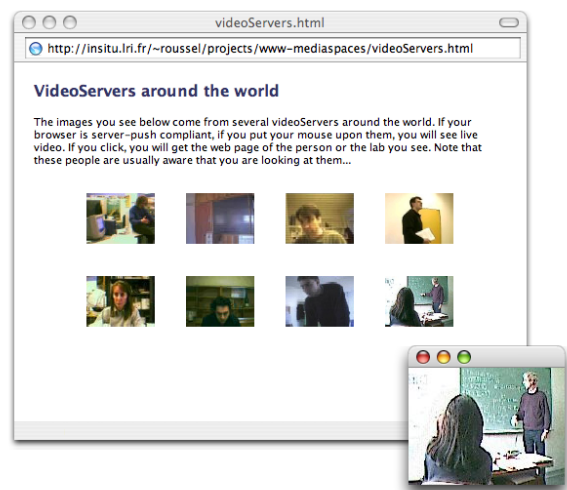


**Figure 3: Live VideoServer images displayed in Apple's Mail application and the Camino Web browser. Images are captured and transmitted every time the HTML message or document is rendered by the application.**

Awareness views similar to Portholes are easily created by users themselves, by including images from several servers in a single HTML document and using a timer to reload it at regular intervals. Basic image and 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 [25], 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.

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 (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.

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, image 2). 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 [1]. Seeing the third image of Figure 6, for example, one might assume that the remote person is absent. Yet seeing this par-

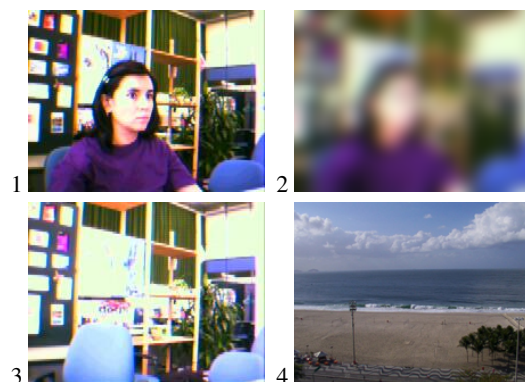


**Figure 4: Focus control: from a low resolution snapshot in a Portholes-like awareness view to a high frame rate independent video that the user can freely move and resize.**



**Figure 5: Sample on-screen notification showing a description of the requested service, the document to contain the requested images and the remote user's address.**

ticular image too often might indicate that she simply doesn't want us to know if she is there. Seeing the fourth image might indicate that she will be away for some time.



**Figure 6: Nimbus control: image captured by the camera (1), filtered image (2), ambiguous pre-recorded image (3) and explicit absence indicator (4).**

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. We will now describe two other systems that illustrate



more direct and intuitive ways of varying one's degree of engagement.

## 4.2 VideoProbe

VideoProbe [22] was created as part of INTERLIVING<sup>5</sup>, a multidisciplinary european 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 7). A specific software analyzes the images captured by the camera in real-time and decides when a picture should be taken and transmitted to similar devices installed 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 lose their colors and contrast and eventually disappear from the browsing interface after a few days (Figure 9).

<sup>5</sup><http://interliving.kth.se/>



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

As confirmed by long-term user studies [8], 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 video-conferencing 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.

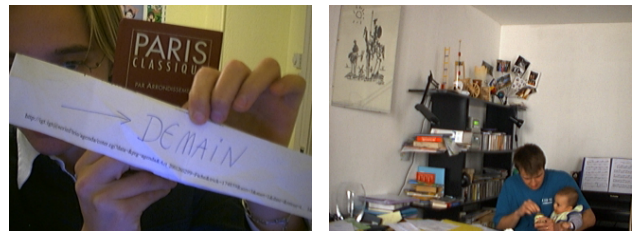


Figure 10: Explicit ("I'll be in Paris tomorrow") and implicit uses of VideoProbe.

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 grad-

ual 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 illustrates further the notion of gradual and intuitive engagement in synchronous communication.

### 4.3 MirrorSpace

MirrorSpace [42] 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.

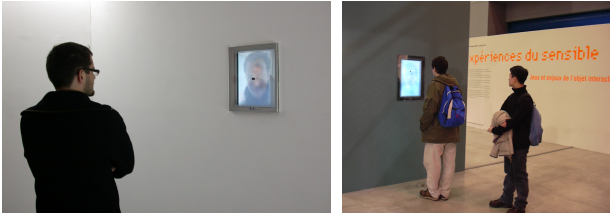


Figure 11: MirrorSpace.

MirrorSpace relies on a mirror metaphor (Figure 11). 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 provides a *social catalyst* [28] to support and encourage distributed interaction. It 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 and blur filtration help creating an intimate relationship between users and the system. People 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. [33]), 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 too close to them on the screen, but proximity sensing and blur filtration allow 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 [34]. Yet, we 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:

- 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
- 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 [16].

## 5. CONCLUSION

In this paper, we have introduced the concept of *multiscale communication* as an alternative approach for motivating video-mediated communication research, and telecommunication research in general. This approach aims at creating systems that support a variable degree of engagement, smooth transitions between degrees and integration with other media. We have also presented the three video systems from which this concept originated, each

of them illustrating one or more aspects of it (e.g. integration with other media, transitions between asynchronous and synchronous communication, intuitive control of the engagement degree).

We are currently working on a series of new communication systems to further explore the design space offered by the multiscale approach. The first one, Pêle-Mêle [18], is another multiparty video system that combines computer vision techniques, spatial and temporal filtering and an original layout to support both asynchronous and synchronous communication, three degrees of engagement and the transitions between them. But the multiscale approach to communication is not limited to video. We are particularly interested, for example, in the potential transitions between various forms of communication based on text (e.g. email and instant messaging), audio, video and shared artifacts.

We hope the multiscale approach will stimulate other researchers interested in multimedia communication. We are particularly curious about the other parallels that might be found between Computer-Mediated Communication and Information Visualization. As an example, Shneiderman's visual information seeking mantra seems particularly relevant to the way we usually engage in a communication with another person, and summarizes pretty well the idea of gradual engagement:

“Overview first, zoom and filter, then details-on-demand” [44]

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

## 6. ACKNOWLEDGEMENTS

This work has been supported by France Télécom R&D as part of the DISCODOM projects (2005-2008). Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen and Wendy Mackay all contributed to the design and implementation of the three systems we described. Thanks to Rosane Ushirobira, Stéphane Chatty, Stéphane Sire and Angela Sutan for their helpful comments on an earlier version of this paper.

## 7. REFERENCES

- [1] P. M. Aoki and A. Woodruff. Making space for stories: ambiguity in the design of personal communication systems. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 181–190, New York, NY, USA, 2005. ACM Press.
- [2] Apple. QuickTime and MPEG-4: Now Featuring H.264. Technology brief, Apple, Apr. 2005.
- [3] 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, May 2005.
- [4] S. Bly, S. Harrison, and S. Irwin. Mediaspaces: Bringing people together in a video, audio and computing environment. *Commun. ACM*, 36(1):28–47, Jan. 1993.
- [5] S. Brave, H. Ishii, and A. Dahley. Tangible interfaces for remote collaboration and communication. In *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 169–178. ACM Press, 1998.
- [6] A. Chang, B. Resner, B. Koerner, X. Wang, and H. Ishii. Lumitouch: an emotional communication device. In *CHI 2001 extended abstracts on Human factors in computer systems*, pages 313–314. ACM Press, 2001.
- [7] D. J. Chatting, J. S. Galpin, and J. S. Donath. Presence and portrayal: video for casual home dialogues. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 395–401, New York, NY, USA, 2006. ACM Press.
- [8] S. Conversy, W. Mackay, M. Beaudouin-Lafon, and N. Roussel. VideoProbe: Sharing Pictures of Everyday Life. Rapport de Recherche 1409, LRI, Université Paris-Sud, France, Apr. 2005. 8 pages.
- [9] 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. JAI Press, 1984.
- [10] A. R. Dennis and J. S. Valacich. Rethinking media richness: Towards a theory of media synchronicity. In *HICSS '99: Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences*, page 1017. IEEE Computer Society, 1999.
- [11] P. Dourish and S. Bly. Portholes: Supporting Awareness in a Distributed Work Group. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems*, pages 541–547. ACM Press, 1992.
- [12] C. Egidio. Videoconferencing as a Technology to Support Group Work: A Review of its Failure. In *Proceedings of ACM CSCW'88 Conference on Computer-Supported Cooperative Work*, pages 13–24. ACM Press, Sept. 1988.
- [13] K. Finn, A. Sellen, and S. Wilbur, editors. *Video-Mediated Communication*. Lawrence Erlbaum, Apr. 1997. 584 pages.
- [14] 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.
- [15] W. Gaver, T. Moran, A. MacLean, L. Löfstrand, P. Dourish, K. Carter, and W. Buxton. Realizing a Video Environment: EuroPARC's RAVE System. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems*, pages 27–35. ACM Press, 1992.
- [16] C. Greenhalgh and S. Benford. Massive: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction*, 2(3):239–261, 1995.
- [17] 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. *ACM Transactions on Graphics, Proceedings of SIGGRAPH 2000*, 22(3):819–827, 2003.
- [18] S. Gueddada and N. Roussel. Pêle-mêle, a video communication system supporting a variable degree of engagement. In *Proceedings of ACM CSCW'06 Conference on Computer-Supported Cooperative Work*, pages 423–426. ACM Press, Nov. 2006.
- [19] C. Gutwin. Traces: Visualizing the Immediate Past to Support Group Interaction. In *Proceedings of Graphics Interface*, pages 43–50, May 2002.
- [20] J. Hollan and S. Stornetta. Beyond being there. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems*, pages 119–125. ACM Press, 1992.
- [21] S. E. Hudson and I. Smith. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In *Proceedings of ACM CSCW'96 Conference on Computer-Supported Cooperative Work, Boston, Mass.*, pages 248–257. ACM Press, Nov. 1996.



- [22] 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 ACM CHI 2003 Conference on Human Factors in Computing Systems*, volume 5(1) of *CHI Letters*, pages 17–24. ACM Press, Apr. 2003.
- [23] E. 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.
- [24] 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.
- [25] B. Johnson and S. Greenberg. Judging People’s Availability for Interaction from Video Snapshots. In *Proceedings of the Hawaii International Conference on System Sciences, Distributed Group Support Systems Minitrack*. IEEE Press, Jan. 1999.
- [26] N. Jouppi, S. Iyer, S. Thomas, and A. Slayden. Bireality: mutually-immersive telepresence. In *MULTIMEDIA ’04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 860–867. ACM Press, 2004.
- [27] S. Jul and G. W. Furnas. Critical zones in desert fog: aids to multiscale navigation. In *UIST ’98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 97–106, New York, NY, USA, 1998. ACM Press.
- [28] K. Karahalios and J. Donath. Telemurals: linking remote spaces with social catalysts. In *CHI ’04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 615–622, New York, NY, USA, 2004. ACM Press.
- [29] K. Lipartito. PicturePhone and the Information Age: The Social Meaning of Failure. *Technology and Culture*, 44(1):50–81, Jan. 2003.
- [30] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), Sept. 1997.
- [31] 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.
- [32] G. McEwan and S. Greenberg. Supporting social worlds with the community bar. In *GROUP ’05: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 21–30, New York, NY, USA, 2005. ACM Press.
- [33] O. Morikawa and T. Maesako. HyperMirror: toward pleasant-to-use video mediated communication system. In *Proceeding of ACM CSCW’98 Conference on Computer-Supported Cooperative Work, Seattle, Mass.*, pages 149–158. ACM Press, 1998.
- [34] C. Neustaedter, S. Greenberg, and M. Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction*, 13(1):1–36, 2006.
- [35] D. Nguyen and J. Canny. Multiview: spatially faithful group video conferencing. In *CHI ’05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 799–808, New York, NY, USA, 2005. ACM Press.
- [36] T. Nørretranders. *The user illusion: cutting consciousness down to size*. Penguin books, 1991.
- [37] K. Perlin and D. Fox. Pad: An alternative approach to the computer interface. In *SIGGRAPH ’93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 57–64. ACM Press, 1993.
- [38] G. Reynard, S. Benford, C. Greenhalgh, and C. Heath. Awareness driven video quality of service in collaborative virtual environments. In *CHI ’98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 464–471, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.
- [39] R. Riesenbach. Less is more (more or less...). White Paper, Telepresence Systems, 1996.
- [40] R. Riesenbach, W. Buxton, G. Karam, and G. Moore. Ontario Telepresence Project. Final report, Information technology research centre, Telecommunications research institute of Ontario, Mar. 1995.
- [41] N. Roussel. Mediascape: a Web-based Mediaspace. *IEEE Multimedia*, 6(2):64–74, April-June 1999.
- [42] N. Roussel, H. Evans, and H. Hansen. Proximity as an interface for video communication. *IEEE Multimedia*, 11(3):12–16, July-September 2004.
- [43] 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.
- [44] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *VL ’96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society, 1996.
- [45] J. Short, E. Williams, and B. Christie. *The Social Psychology of Telecommunications*. Wiley, New York, 1976.
- [46] I. Smith and S. Hudson. Low disturbance audio for awareness and privacy in media space applications. In *MULTIMEDIA ’95: Proceedings of the third ACM international conference on Multimedia*, pages 91–97. ACM Press, 1995.
- [47] R. Strong and B. Gaver. Feather, scent and shaker: supporting simple intimacy. In *Proceedings of ACM CSCW’96 Conference on Computer-Supported Cooperative Work*, pages 29–30. ACM Press, Nov. 1996.
- [48] K. Tanaka, J. Hayashi, M. Inami, and S. Tachi. Twister: An immersive autostereoscopic display. In *VR ’04: Proceedings of the IEEE Virtual Reality 2004 (VR’04)*, pages 59–66, Washington, DC, USA, 2004. IEEE Computer Society.
- [49] S. Whittaker. Video as a technology for interpersonal communications: a new perspective. In A. Rodriguez and J. Maitan, editors, *Proceedings of Multimedia Computing and Networking 1995*, volume 2417, pages 294–304. SPIE, 1995.
- [50] Q. Zhao and J. Stasko. Evaluating Image Filtering Based Techniques in Media Space Applications. In *Proceeding of ACM CSCW’98 Conference on Computer-Supported Cooperative Work, Seattle, Mass.*, pages 11–18. ACM Press, 1998.