Use of multiscale communication to improve video-mediated communication[*]

Mariusz Wisniewski
University of Fribourg
Av. Europe 20
1700 Fribourg
mariusz.wisniewski@unifr.ch

ABSTRACT

Teleconferencing systems and services are the main set of technologies developed thus far to support group work. Video-conferencing is often thought of a good way to enhance communication between parties. However, while it has been commercially available for many years, it has failed to achieve the success that was expected, and is still underused. This paper, mainly based on the work of Nicolas Roussel of the university of Paris-Sud 11, explains what multiscale communication is and how it can improve the comfort of the users of video-mediated communication. We will give some reasons as why today's videoconferencing is still seen as uninteresting, explain multiscale communication and describe one system, Pêle-mêle, which is using multiscale communication as an example. There is another system called MirrorSpace which uses multiscale communication and which would have been interesting to describe, however there will not be enough space in this paper to do so.

General Terms

Beyond being there.

Keywords

Video-mediated communication, videoconferencing, multi-scale communication, variable degree of engagement.

1. INTRODUCTION

Nowadays, 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 interactions, like 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 distant communication. [1]

CSCW researchers have investigated the reasons for the failure of traditional videoconferencing. The main problem with video-mediated communication today is that it does not take into account non-verbal communication and social aspects (gestures, eye gaze, distance, etc.). Users don't have the feeling of face-to-face communication because the systems actually available lack these aspects and make them feel uncomfortable when they use them. [3]

For example, physical proximity between individuals is a form of non-verbal communication that we all use everyday without even being aware of it. In face-to-face communication, both parties adapt their distance depending on how well they know each other. Strangers will communicate in a different way and keep a greater distance than people who already know each other. We constantly use space and distance to define and negotiate the interface between private and public matter, particularly during the moments leading up to contact. [4]

In 1963, the American E.T. Hall used the term of *proxemics* when he investigated man's appreciation and use of personal space. Hall's model lists four distances that North Americans use in the structuring of personal dynamic space: intimate (less than 30cm), personal (between 45cm and 1,2m), social (between 1,2 and 3,65m), and public (more than 3,65m). For each communication situation, we have a distance within these four categories that we find appropriate. If the perceived distance is inappropriate, we become uncomfortable and usually adjust it by physically moving closer or farther away, or even simply turning our head or looking in another direction. [2]

There are many social aspects in face-to-face communication, however, video-mediated communication systems barely take any of them into account, which make the users feel unconsciously uncomfortable. This is why nowadays video-mediated communication is still rarely used and oral or text-based communications remain more popular. Yet, these non-verbal aspects of communication should be one of the advantages of video over audio.

Nicolas Roussel proposes the concept of multiscale communication as an alternative approach for motivating video-mediated communication research. [1] The paper is organized as follows. In the first part of this article, I will present the concept of multiscale communication, and then I will

---

present and describe one system, Pêle-mêle, which is using this concept as an example and try to show how it can improve video-mediated communication.

2. MULTISCALE COMMUNICATION

The concept of multiscale communication is 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. [1]

In a quite influential paper from 1992, J. Hollan and S. Stornetta argued that rather than trying to imitate physical proximity, telecommunication research should develop tools that go *beyond being there*. *Being there* is an expression that refers to the concept of *presence*. It means 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. In their paper, 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*. [1][7] 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 (...)".

This analysis has been quite popular and has inspired a number of systems.

Although we now understand the notion of *beyond being there*, we must still find ways to achieve this feeling of *presence*. In his papers, Nicolas Roussel says he believes the notion of multiple degrees of engagement is an important one that can help structure the design space of communication systems. He says this notion could help users better choose the right communication service for a particular context. [1]

I believe Nicolas Roussel is right. A variable degree of engagement with smooth transitions between them could give a feeling of distance to the users, making a video-mediated system more appealing. According to Nicolas Roussel, the following problems, in particular should be explored:

- 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?

- How can we move from one degree to another? Can we avoid dialogue boxes? How do we perceive a remote person's degree? How can we negotiate degrees with a remote person?

- How can we combine video with other media? (e.g. email, telephone, instant messaging systems, Web)

To implement the different degrees of engagement using video, an interesting solution could be using filters to change the level of details of an image stream. For the users, the more the images are big, colourful and sharp, the more they will attract attention and expose the users, giving them a feeling of proximity. Besides, using filters will make the transition between the levels of engagement easy to make. As illustrated by Figure 1, spatial filtering techniques can be used to degrade images while temporal compositions can provide awareness of past activity. These simple techniques give additional meaning to an image stream to the users in an intuitive way. [1][2]
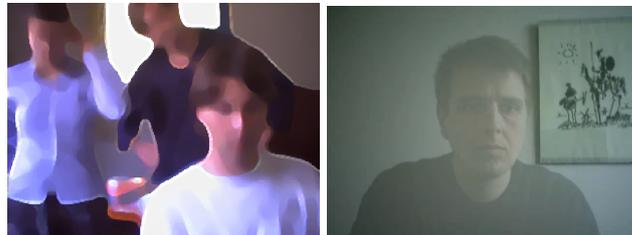

Figure 1: Degrading (left) or enriching (right) a video stream.

Like I said before, when you have different degrees of engagement, you need to make transition between them. The system must detect the actions that will trigger the transition, but this poses two kinds of problems. The degrees of engagement and the transitions should not be initiated by the users but should be detected and done by the system. For this, the camera is probably the most interesting input device, but other sensors might also be useful. The interactions must be intuitive and simple for the users to understand. To do this, feedback mechanisms such as animations can make the users understand when a transition instigated by the remote partner occurs. The second problem is that depending on the level of engagement the user is in, different tools might be needed. For example, audio is only required during direct verbal communication and an audio input device should only be activated when necessary. The system should manage these tools automatically depending on the present situation. [1]

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.

In his paper ([1]), Nicolas Roussel proposes 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. Enriching or degrading a video stream can change both its meaning and level of detail. I think that this approach is interesting and can be effective in the use of video-mediated communication. If a system can implement all these aspects and manage them automatically it could really improve communication through video.

3. PÊLE-MÊLE

Pêle-mêle is a video system designed for group interaction. It consists in a screen equipped with a video camera and connected to a small computer. Every other place equipped with the system is shown on the screen, with a detailed view of the users actually using the system. This automatically gives the user information about who is present and who is not. As we can see, the principle of the variable degree of engagement, as well as the smooth transition between them, is a key point of the system.

Pêle-mêle constantly monitors the activity of local users and classifies it according to a three-level scale: *away*, *available* and *engaged*.[5] The activity observed at each place determines the nature of its on-screen representation, which potentially combines live images and pre-recorded ones that are filtered, delayed or displayed as-is:

away : The place is represented by video clips showing past activity and a filtered view of the last image it transmitted.

available : The place is represented by video clips showing past activity and a delayed live stream.

engaged : The place is represented by video clips showing past activity and a live stream which is recorded for later use.

This system clearly follows the principles of multiscale communication.

Users engaged in using the Pêle-mêle are shown in the middle of the screen, their live-images are overlaid so that every one of them is visible. Available users are shown on the periphery using delayed images (Figure 2, left). Auditory feedback and smooth animated transitions between these two representations ease perception and understanding of the state changes. Images showing past activity are displayed on the periphery along a perspective timeline: they slowly shrink and drift toward the center of the screen over time (Figure 2, right). This helps the users to easily understand who is present and who is not and what they are doing or for how long they have left. [5][6]
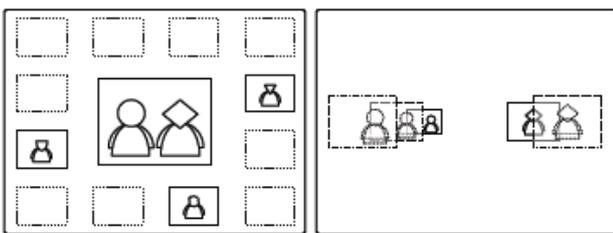

Figure 2: Focus-plus-context view of live streams and perspective timeline effect used for recorded images.

3.1 Away

Away corresponds to the situation where no face is detected and the difference with a reference image stays below a certain threshold. This just means that there's nothing happening in front of the camera. At this level, no image is

transmitted to the other users. The place is represented by the last image transmitted at the *available* level and clips recorded at the *engaged* one. These images are displayed in a small size on the periphery of the screen and slowly drift in perspective over time. [5] This is a simple and intuitive way for the users to be aware of who is connected but not really available and for how long he has been absent.

These clips are normally represented by a single image on the periphery of the screen, but if no users are in the *engaged* level, they get promoted to the focus area in the middle of the screen from time to time. The last *available* image degrades over time to make it clear it is not live and to preserve the privacy of the user. As illustrated by Figure 3, the filter produces an oil painting effect that rapidly removes details without suppressing all visible information. [5] This way a user can see, for example, that his colleague's last action, for example, was to put on a jacket, which could mean that he won't be back soon, while not revealing to many details and thus preserving his privacy. The small size of the images
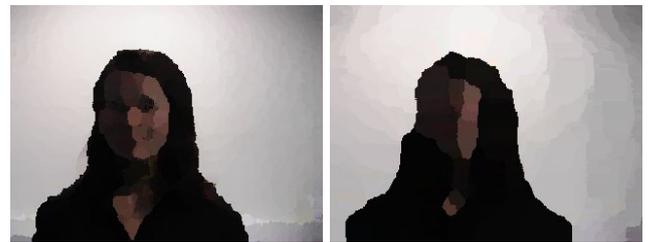

Figure 3: Image degradation over time (one minute, two minutes).

displayed at this level invites users who want to see them to move close to the display. If they come close enough, their own Pêle-mêle will switch to the *engaged* mode and start recording them.[5][6]

This gives the users the possibility to focus on the system only when needed, which can be compared to face-to-face non-verbal communication.

3.2 Available

When a face or an important change between the current scene and the reference image is detected, the Pêle-mêle system switches from *away* to *available*. The assumption made is that such a change is probably caused by incoming people or previously undetected ones. An auditory feedback is generated to warn the user and the size of the image representing the place on the periphery is slowly increased to a medium size (Figure 4). [5] As we see, the system does everything and the user can engage a transition in a way similar to face-to-face communication.

If a face is found close enough or if significant motion of the user is detected, the Pêle-mêle switches further to the *engaged* level. As we see, when the user wants to get engaged in a communication, he just has to get closer like in face-to-face communication. If no face is found and the scene does not change anymore, it switches back to *away*.

As explained above, the last image transmitted will also be rapidly degraded when used at the *away* level to provide some information without unnecessarily exposing privacy.

The delayed video stream is temporally composed to provide awareness of recent activity for the users to have a feeling

Figure 4: Image size growth during the transition from *away* to *available*.

Figure 6: Two then three users engaged together.

of presence. Selected past images are alpha-blended with the current one before it is displayed. The alpha value of each image is inversely proportional to its age, which makes it easy to perceive their temporal order (Figure 5).[5] This allows users to follow each other's activity without having to directly to communicate and to check if the other is eventually available for conversation.

Figure 5: Examples of time composed pictures.

3.3 Engaged

When in the *engaged* level, the size of the video stream slowly increases while it moves toward the center of the screen. Auditory feedback accompanies the transition and the delay is progressively suppressed. This smooth transition allows users to engage in a communication in an intuitive and non-brutal way. Just like in real face-to-face communication, the users have the time to see each other and prepare before starting the talk. They can also easily go back to the *available* level by withdrawing themselves from the camera. This offers the user the possibility for effective non-verbal communication using distance.

When the transition is finished, live images are displayed in a big size in the focus area and recorded for later use at the *away* and *available* levels. The images of all the places at the *engaged* level are actually alpha-blended together (Figure 6). This is really a good idea since it gives users a feeling of proximity that can help the communication between them. Although the combined display of local and remote participants is known to improve the co-presence feeling, the blending of multiple video sources can be quite confusing, e.g. making it difficult to associate faces and backgrounds. To minimize this problem, Pêle-Mêle uses a lower alpha value for local images. [5][6]

The transition back from *engaged* to *available* triggers when a close face and a significant motion are not detected anymore. During this transition, the size of the video stream decreases while it moves back to the periphery. The stream also temporarily decelerates to introduce the few seconds delay mentioned previously. [5]
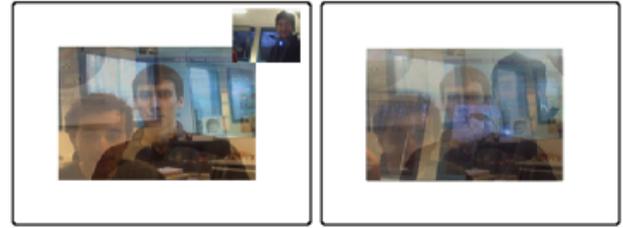
This transition is a smooth way to end a conversation between parties, I think that the absence of such smooth transitions is one of the main reasons why other systems are so unpleasant to use.

4. CONCLUSION

As I said before, video-mediated communication is far from being the success expected a few years ago.

In this paper, I presented the concept of multiscale communication as an interesting approach for video-mediated communication research. This concept should allow to realize new video-communication systems with a variable degree of engagement and smooth transition between degrees, that give the users greater comfort using non-verbal communication. Pêle-mêle is such a system. It presents a new way to approach video-mediated communication, using different types of algorithm and filters to manage the degree of engagement of the users. It could be a first step towards the right way of implementing video-communication systems in the future. Systems that, unlike most of those that exist, are easy, comfortable and pleasant to use.

5. REFERENCES

1. Roussel, N. Towards multiscale communication systems. Technical report 1439, LRI, 2006.

2. Roussel, N., and Gueddana, S. Beyond "Beyond Being There" : Towards Multiscale Communication Systems. In *Proceedings of ACM Multimedia 2007*, pages 89-98. ACM, 2007.

3. Egido, C. Videoconferencing as a Technology to Support Group Work: A review of its Failure. ACM, 1988.

4. Roussel, N., Evans, H., and Hansen, H. Utilisation de la distance comme interface à un système de communication vidéo. In *Proceedings of IHM 2003*. ACM, 2003.

5. Roussel, N., and Gueddana, S. Pêle-Mêle, a Video Communication System Supporting a Variable Degree of Engagement. Technical report 1437, LRI, 2006.

6. Roussel, N., and Gueddana, S. Pêle-Mêle, une étude de la communication multi-échelles. In *Proceedings of IHM 2009*. ACM, 2009.

7. Hollan, J., and Stornetta, S. Beyond Being There. In *Proceedings of CHI'92*. ACM, 1992.