



An empirical study on the user's context in mobile videoconferencing devices

Ignacio Calvo, Tomas Dorta, Jean-Marc Robert

► To cite this version:

Ignacio Calvo, Tomas Dorta, Jean-Marc Robert. An empirical study on the user's context in mobile videoconferencing devices. *Journal d'Interaction Personne-Système*, 2015, Volume 3, Issue 2, Special Issue: Best of IHM'2013 (2), pp. 1-24. 10.46298/jips.663 . hal-01166903

HAL Id: hal-01166903

<https://hal.science/hal-01166903>

Submitted on 23 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An empirical study on the user's context in mobile videoconferencing devices

IGNACIO CALVO

HYBRIDLAB Université de Montréal

TOMÁS DORTA

HYBRIDLAB Université de Montréal

JEAN-MARC ROBERT

Polytechnique de Montréal

Résumé : Ce document présente une étude empirique sur l'utilisation de la vidéoconférence mobile et le contexte de l'utilisateur afin de proposer des lignes directrices pour la conception des interfaces des dispositifs de communication vidéo mobile. Grâce à un échange riche d'informations, ce type de communication peut amener un sentiment de présence fort, mais les interfaces actuelles manquent de flexibilité pour permettre aux usagers d'être créatifs et pour avoir des échanges plus significatifs dans une vidéoconférence.

Nous avons mené des observations avec 16 participants dans trois activités où leurs conversations, les réactions et les comportements ont été observés. Deux groupes de discussion ont aussi servi pour identifier les habitudes formées à partir d'une utilisation régulière. Les résultats suggèrent une différence importante entre l'utilisation de la caméra de face ou de dos, le type de comportement, et la nécessité de fournir des outils qui offrent plus de contrôle sur l'échange dans la conversation. D'après les résultats, l'étude propose plusieurs lignes directrices de conception pour les interfaces de communication vidéo mobiles, concernant la construction du contexte mobile de l'utilisateur.

Mots clés : nouvelles technologies de communication, communication vidéo mobile, contexte, mobilité, présence, Interface.

Abstract: This paper presents an exploratory empirical study on the user's context in mobile videoconferencing in order to improve the user interface of mobile video devices. Through the rich exchange of information, mobile video communication can provide a better sense of presence than other means of communication. Yet the current mobile interfaces lack the flexibility required to be creative and more meaningful in a videoconference exchange. We conducted observations with 16 participants in three activities where their conversations, reactions and behaviours were observed. Two focus groups were used to identify habits formed from regular use. Results suggest an important difference between the use of the front-facing or back-facing camera and the importance of offering tools that provide more control over the video exchange. From these results, the study proposes several design recommendations for mobile video communication interfaces in order to support the construction of the user's mobile context.

Key words: New communication technologies; Mobile interface design; Video communication interfaces; Context; Mobility; Interfaces.

Adresse des auteurs : Ignacio Calvo, (ignacio.calvo@umontreal.ca), Hybridlab, Université de Montréal, 5155 Decelles Ave., JA Bombardier Building, 4029. H3T 2B1, Montreal, Canada. Tomás Dorta, (tomas.dorta@umontreal.ca), Hybridlab, Université de Montréal, 5155 Decelles Ave., JA Bombardier Building, 4027, H3T 2B1, Montreal, Canada. Jean-Marc Robert, (jmrobert@courriel.polymtl.ca), Polytechnique Montréal, Dep. of mathematics and industrial engineering, PO Box 6079, H3C 3A7, Montreal, Canada,

Les articles de JIPS sont publiés sous licence Creative Commons Paternité 2.0 Générique.

1. INTRODUCTION

In recent years, with the emergence of smartphones, faster networks as well as standardized development platforms, such as Android™, iOS™ and Windows Phone™, more and more mobile communication and media devices have adopted video calling as one of their communication modes. While there have been devices capable of mobile video communication for several years, only recently has access to affordable mobile bandwidth and sufficient interoperability between devices been present, thanks to software such as Skype™. This could mean that more users are likely to have the capacity to host a videoconference through a mobile device, and they are able to connect to more devices. This marks a potential change for video communication from being used primarily in a business environment to everyday use.

However, much like the problems the telephone encountered when it was first transposed to a mobile environment, this transition to mobile videoconferencing is bound to face both old and new problems [O'hara et al., 2009]. Indeed, there are still unknown problems that will arise due to the changing habits of users, as they are constantly finding new ways to use their tools and to communicate with each other. That will provide great opportunities to reconsider how we think about and design mobility and communication. How do users communicate within their contexts through videoconferencing and, more specifically, what must be taken into account when we design a mobile videoconferencing interface capable of adapting to its context?

Through this study, we aim to gain a better understanding of the practices and behaviours during mobile videoconferencing, in order to grasp how interfaces are adapted to their users when they are mobile, what kind of experience they provide in mobility and how well they serve their users. The knowledge gained from this study will then be synthesized as potential design guidelines.

We observed the use of videoconferencing in a mobile context, the problems encountered during its use and what should be considered for the design of mobile videoconferencing interfaces. Sixteen users were observed in a series of conversational activities, both moving and stationary, followed by an interview about their experiences. The exchanges were studied to identify both their use of context and the information being exchanged between participants. Further information was gathered in the form of two focus groups of power users to identify habits and preferences of use.

We find a marked difference between using the front camera and the back camera. The front camera, used mostly to talk to someone, has a much bigger impact on mobility, while the rear camera provides more flexibility, but lacks image control. When users talk to each other, the video exchange seems to overwhelm the context, almost immobilizing them, while users concentrate on each other's image. When on the move and discussing their surroundings, their attention was divided between their environment and the communication, as they tried to either curate how their environment was seen or to replicate their perception of the space. Mobility seems to have an impact on the users' experience (*flow*) [Csikszentmihalyi, 1990] and workload during a video call; these appear to vary with the activity and the content being exchanged. The interviews with habitual users provide some insights on how users have adapted to these issues and suggest that videoconferencing requires a certain familiarity with the person they are talking to, but they value the rich experience it provides.

Finally, we offer some possible design guidelines for future videoconferencing interfaces.

2. MOBILE VIDEOCONFERENCING

Videoconferencing, also known as video communication and video calling, was foreseen since the birth of the telephone, when people dreamed that new devices would allow them to see the person they were talking to. Over time, several technological implementations were proposed and since the 1970s, videoconferencing has steadily evolved, going from analog to digital and gaining acceptance in the corporate world and for personal use.

Mobile videoconferencing, a recent technology on the market since the launch of 3G networks has gained increased support by service providers and handset manufacturers. In more recent years, the popularization of smartphones and operating system development platforms, such as Android™, iOS™ and Windows Phone™, has facilitated interoperability. This type of communication is now more accessible than it's ever been, making it a part of everyday life.

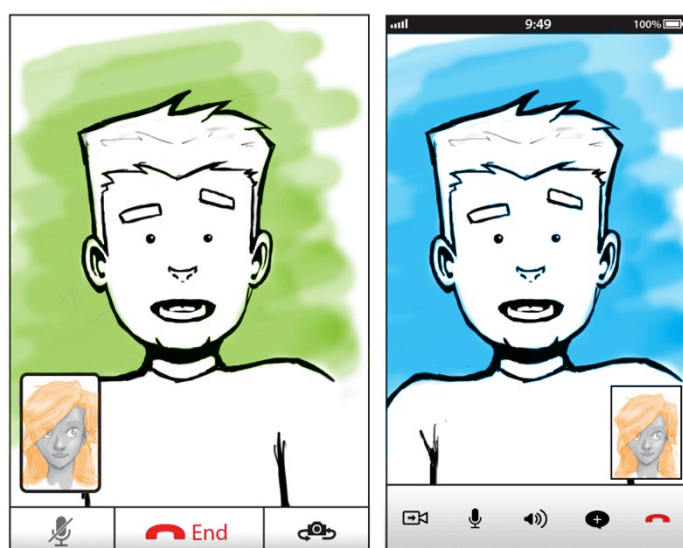


Figure 1. Interfaces of the two popular mobile videoconferencing applications (Facetime™ and Skype™)

However, while the technology has greatly evolved, the user interface has changed very little over time. Already, in 1936, the German postal service had a videoconferencing system available to the public through its postal system, in which "The head and shoulder image of a person is clearly produced" [anonymous, 1936]. By comparison, the interface offered in most popular modern systems (Figure 1) offers similar visual information, providing a view of the users' distant caller and a "mirror image" of the user. Current mobile videoconferencing interfaces have then changed very little from the ones found in personal computers. Just as in previous iterations of the technology, the user is shown both the distant and his own image, can control the volume, mute, hold or end the current call. Some features of their new platforms were adopted, such as switching to the back-facing camera to provide a view of what is in front of the user, but these features have had very little impact on the interface of these tools. Mobility also imposes some limitations, notably the need for a stable and fast connection, which often means that a Wi-Fi network is often required or, at least,

preferred to ensure call quality. Though all systems use a list of contacts in order to manage possible callers, the information provided about each contact by each system varies, with some systems communicating accessibility or status.

According to O'Hara *et al.* [2006], there are both social and technological barriers to the use of videoconferencing that explain this situation. While it was already established that it is a very formal and rigid type of communication [Reponen, Huuskonen and Mihalic, 2008], they noted that this technology requires more management from the user in order to compensate for possible poor sound, poor light or connectivity conditions and to manage the framing of the camera. Socially, it is a type of communication that is considered fairly intrusive, somewhat chaotic [O'hara *et al.*, 2006] and often reserved to communicate only with people users hold dear [Ames, Go, Kaye and Spasojevic, 2010]. Moreover, there are several social barriers between caller and recipient, caller and co-presents, as well as co-presents and others. The performance aspect of this type of interaction was also explored through other multimedia studies [Dalsgaard, Dalsgaard and Koefoed, 2008], suggesting a division between operating a system, performing for others and being spectators of their reactions, a process which shapes their interaction.

Others studies have also investigated ethnographic aspects of the use of both mobile phones [Morel, 2002] and mobile videoconferencing [Morel and Licoppe, 2009][Licoppe and Morel, 2012]. They describe how users create know-how, tricks and particular practices through the use of mobile phones, which are sure to continue with the next generation of phones. Moreover, they observed that when videoconferencing in a public setting, users adopt certain practices, such as interacting through a "talking heads" visual format [Licoppe and Morel, 2012], where they try to frame themselves in a thumbnail image (Figure 1), and how users self-organize their interactions. However, both studies concentrated on face-to-face communication and provide little information on the use of a back camera. Current phones also offer a different form factor than those used in previous research, providing a larger image and touch screen interfaces.

3. MOBILITY

Design requires understanding the context in which the devices are used. This is particularly challenging when designing mobile tools since mobility is an ill-defined term. Merriam Webster [2013] indicates it as the capacity of an objet to move or to be moved, thus suggesting a mobile context would be constantly changing. However, while it is easy to identify something that is mobile, it is much harder to understand why something is mobile.

The term "mobility" holds many meanings, depending on where it is being used. Cresswell [2006] organizes these definitions along the following axes:

Mobility as an observable fact: this definition approaches mobility more as "movement" and thus easily observable and measurable, like the definition mentioned previously. The "military" definition of mobility is an example of this [Cresswell, 2006].

Mobility as a representational strategy: the act of movement is seen as a metaphor for something else, such as "social mobility" [Cresswell, 2006].

Mobility as a way of being: mobility as an embodied experience, through which meaning is given to mobility [Cresswell, 2006]. For example, persons on a flight could be considered mobile because of their context rather than the notions that they themselves are in movement. This is in line with other research in

sociology, which approaches mobility as a construct of the user [Weilenmann, 2003].

In the field of computing, theories on mobility [Dey, 2001] are divided between quantitative approaches and more qualitative approaches, which consider the context in mobility as a representational problem. This is the approach taken by Weilenmann [2003], whose research analysed workers under constant mobility and with high demands for communication (airport and ski resort workers). Her research addresses the use of technologies during activities where people are on the move. She approaches mobility as a construct built and maintained by the user. During communication, the context is negotiated between participants in order to share a mutual understanding of the context of both parties. This mutual understanding of the context is consistent with the telepresence described for videoconferencing, which Reponen *et al.* [2008] explain as seeking to break down the physical barrier of places.

This technology then has the potential to help in the negotiation of context, by providing rich information to this exchange. Moreover, current mobile video calls might be able to compensate for problems identified in fixed videoconferencing systems, such as image control [Dourish *et al.*, 1996] and the sharing of virtual spaces [Kraut, Gergle and Fussell, 2002].

4. CONTEXT

A mobile context, however, is a very demanding environment. Through a combination of social, physical and technological demands, the cognitive resources of users are reduced to very short spans [Oulasvirta, Tamminen, Roto, and Kuorelahti, 2005]. Therefore, while mobile videoconferencing can provide a rich and useful manner to communicate, it is important to approach it in a way that will take into account the negotiation created by the users and to understand what can be done to support this process.

To understand the mobile context, we must first be able to provide a definition of context. Many authors have provided such definitions [Dourish, 2004][Dey, 2001][Reponen, Huuskonen and Mihalic, 2008], using a series of dimensions, such as *time*, to create a structure to analyse context. However, these models often describe individual contexts or fail to explain how contexts can interact with each other.

Zimmermann *et al.* [2007] propose that context is any information that can be used to characterize the situation of an entity. Elements for the description of this context information fall into five categories: individuality, activity, location, time, and relations.

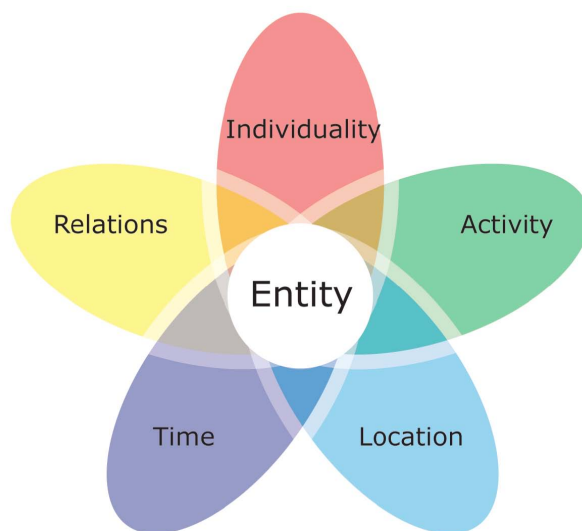


Figure 2. Interpretation of Zimmerman's context model

Individuality relates to any property that describes what the entity is and everything that can describe its current state (e.g., a plant, a human, a book or a group of people). Activity defines what the entity wants to achieve, what tasks and goals the entity will be or is involved in. It also provides insight on the need of the entity. Location is position-related information that may include quantitative locations, as well as IP addresses or other means of locating the entity. Time related information includes time zones, concepts of time, events, and intervals. The Relations aspect applies to any link the user has established to other entities. This encompasses social relations, functional concerns between two entities and compositional relations between the entity and its parts [Zimmerman *et al.*, 2007].

Zimmerman's model (Figure 2) thoroughly articulates the interaction between users and their contexts over other definitions and provides a basis of how two contexts connect and interact. This process relies on two entities sharing information of their five aspects, eventually settling on a common context. It provides a framework to analyze what is being exchanged during mobile video communications, and thus understand how context is communicated through to the other person and what aspects are important to the communication.

5. METHODOLOGY

In order to explore the potential of this type of communication, we looked at everyday use rather than work-related activities, as technologies are providing new opportunities in this space. To understand usage in everyday situations, we chose to look at how it is used during specific tasks and to meet with regular users about their habits. While this does not provide a direct view of everyday usage, it should provide clues as to how users would approach these situations in real-life and what they have learnt from their use.

5.1 Participants

For this research, participants were selected through 20-30 year-olds, as they approximate the average age of mobile phone users [Wolfram alpha, 2013]. All participants had at least some experience with video conferencing and with mobile devices. To ensure that participants would be discussing with a known interlocutor, they were chosen among university students from the same program, in order to allow exchanges to happen between people who either knew each other or shared the same academic context and thus could have more convivial exchanges.

During the course of this exploration, we conducted observations of three ad-hoc videoconferencing activities and of two focus groups.

5.2 Videoconferencing Tools

To observe how users behave during mobile videoconferencing, three different activities were devised. This segment of the research was conducted using a modified Apple iPod Touch™, equipped with an application allowing the capture of its screen in a separate computer, Display Recorder. This was chosen because it provided a near real-time view of the device's screen and helped provide a lighter test tool by taking advantage of the device's capacities. This was inspired by Morel's work in the field [Licoppe and Morel, 2012] and it allowed participants to use the device unhindered by any research equipment, such as cameras or backpacks. It could also be used during the video call with little impact on lag times, as our initial tests showed less than a second in added lag. This allowed us to remain closer to normal use of the device, albeit with a severely reduced battery time. Participants were also provided headphones, since the device's speakers were too weak to use in a public space.

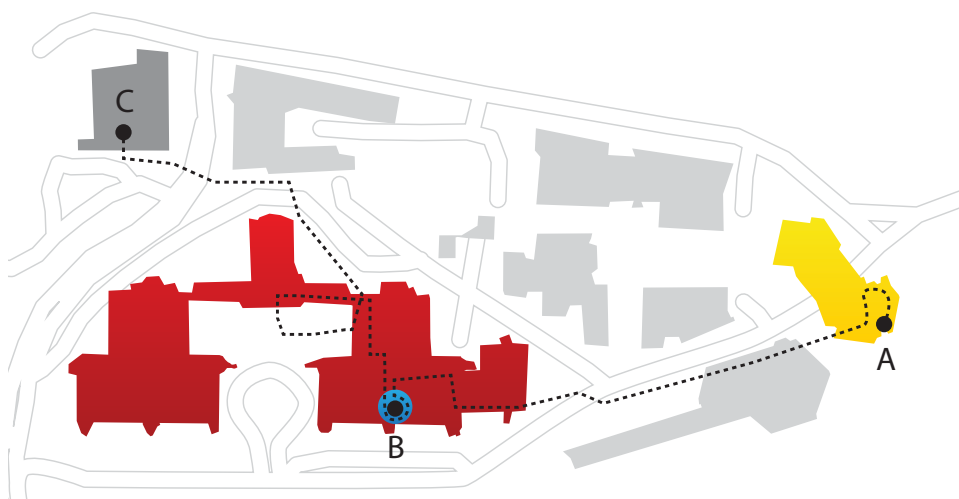


Figure 3. Map of testing activities

The research was conducted on the university campus (Figure 3), which provided large enough Wi-Fi network coverage to allow for a large array of spaces where mobile videoconferencing was possible. Because of network coverage irregularities found during test preparation, the research was conducted with a single mobile participant

discussing with a research assistant as the distant caller, sitting at a desktop computer. This limited network strains on the connexion to one side of the exchange.

5.3 Videoconferencing activities

The three communication activities were chosen according to the type of relationship with the context and with mobility they presented. While this may not represent the most popular uses of videoconferencing, they provide clear modes of interaction to observe the participants in three clear situations: a baseline situation, holding a conversation in a public space, a situation where they had control over the context of the activity, by presenting a space, and an activity where they were dependent on their caller to understand their context, by navigating a space.

During these activities, the research assistant provided the instructions and was asked to maintain a conversation during the tests, relevant to the situation at hand.

Conversation: Participants were placed in a public place, one of the main university cafeterias, where they were asked to call the research assistant and carry a conversation for 10 to 15 minutes, on a subject of their choice. Participants were free to choose where to hold the conversation. This situation serves as a comparative baseline for the other two, to observe behaviour described in other research, such as behaviour regarding usage in a public space.

Presenting a space: Participants were placed in a small exhibition hall, in the rare books section of one of the university libraries. Here, the participants had to visit the exhibit accompanied by the distant research assistant. Participants were in a situation where they had to move in a large room in order to explore the space: they were in control of their context, as they chose where to go. During this activity the assistant would inquire about what the participant was presenting, the room they were in and their impression of the exhibit. In this position, participants will have to interact with their immediate context, but will have control over the interaction, which will help us observe how they will position themselves and how they will use their context.

Navigating (from point B to point C): In this instance, participants were given instructions by the distant research assistant through a video call to guide them through various environments, such as an outdoor section. This activity placed users in movement and gave them little control over the context, as they had to rely on the instructions from the distant caller. This provides a situation where their surroundings will have different meaning depending on external elements and that will allow us to observe how that space is mediated between callers.

Activities were organized so that they would happen along a path where the video calls could take place. The participants began at a specific point on the campus and crossed a part of it, completing activities along the way. They had to present a space, have a conversation in a public space and then navigate into a building. The navigation activity was broken down into two parts to make the most out of the chosen path. Participants experienced a simpler version of the navigation activity before the conversation and a more complex version afterwards.

5.4 Assessment

During each activity, the following data was recorded:

Videoconference: the video conversation was recorded for both participant and distant caller, providing a perspective of the whole conversation. The content of these conversations was transcribed verbatim and was analysed to indicate references to any of the five aspects of context presented earlier, which could then be compared proportionally.

Shadowing [McDonald, 2005]: two observers followed the participant, taking notes on behaviour, the relationship to the observable context and actions not seen by the user's camera. The observers would follow from a distance, without interacting with the participant (Figure 4). This approach was chosen because it allowed us to observe the participants in action, but also because recording the activities in a public setting would have raised ethical concerns.

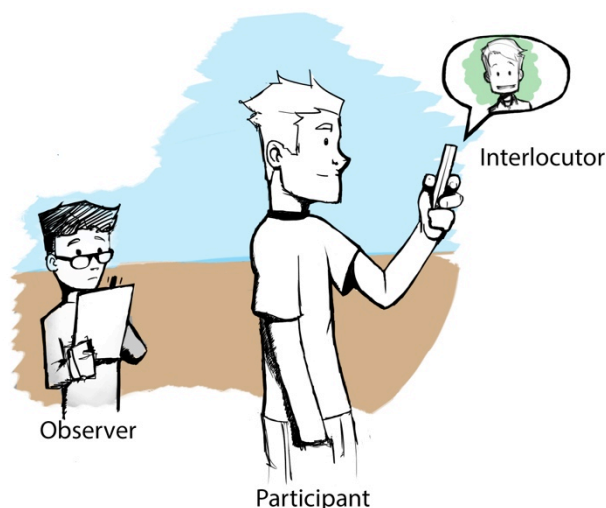


Figure 4. Mobile videoconferencing activities assessment

Workload: following each activity, participants filled a NASA TLX workload survey [Hart and Staveland, 1988] providing cognitive information on the tools used, both what was imposed on the user and the user's interactions with the tool toward the task. This information is crucial to qualify the nature of activity being undertaken and to establish a comparative element between activities.

Optimal experience [Csikszentmihalyi, 1990]: following each activity, users filled out a form inquiring about their different psychological states at the beginning, middle and end of the activity, providing information on their experience. Flow [4] is a complex psychological state that is described as a perceived optimal experience. It is described as high involvement in a complex activity that requires concentration, brings enjoyment and intrinsic motivation. The balance between challenges and skills determines the flow state, giving rise to eight possible dimensions: apathy, worry, anxiety, arousal, flow, control, boredom, and relaxation.

Assessment requires participants to identify which aspect described their best psychological state. Due to the exploratory nature of this study, this is useful in order to identify the quality of the participant's experience with the tools. They would be asked to describe their experience during the test, using a chart describing the possible states described in Flow theory (Figure 5).

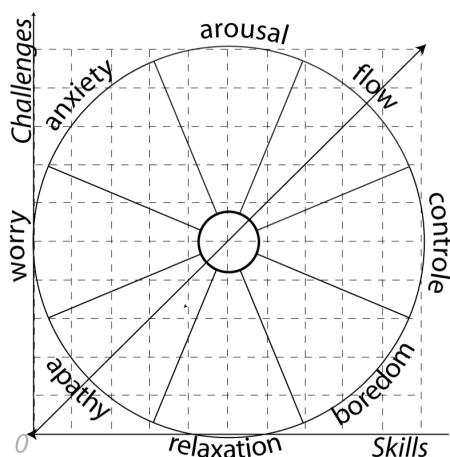


Figure 5. Flow assessment

Exit interview: at the end of the three activities, a 15-minute interview regarding the participants' experience, awareness of their context and usage during the tests was conducted to collect their comments and validate observers' opinions.

Because this was an exploratory study, we opted for a number of measurements allowing us to triangulate our observations with more precision. For instance, the experience assessment through optimal experience helped us to understand the workload better; and the observation and end interview, to better contextualise the results of both the experience and the workload assessment.

5.5 The habits of users

To investigate daily habits we formed two focus groups organized amongst a section of participants who had regular experiences with mobile videoconferencing. This approach was chosen as it would allow for more flexible discussion and a better comprehension of the participant's answers.

The first group were family members between the ages of 20 to 30, composed of three young women and two men. As part of the same family unit, they would use mobile videoconferencing tools almost daily, generally to talk with other family members in other countries, going as far as sharing some meals through video calling.

The second group was comprised of three men aged 25 to 30. They would usually use videoconferencing with their friends and family.

Both groups filled out a survey on their use of mobile videoconferencing tools. It covered general usage questions, such as the type of device and frequency of use. The participants were questioned on their habits and outlooks on videoconferencing.

6. RESULTS

Results are presented according to the observations made during the three proposed activities and the information gathered from focus group sessions.

6.1. Videoconferencing activities

Mobile videoconferencing activities were successfully completed over the course of three weeks. Through recruitment, 16 participants were chosen for the activities, rather than the 20 originally planned, due to problems of availability and difficulty in finding participants with sufficient experience with mobile video conferencing. It should be noted that tests were completed using headphones. During pre-tests the device proved to have insufficient audio levels for a public space. Attempts to correct were unsuccessful, mostly due to the extra strain of a speaker on the device's battery. It was finally decided to use headphones, which precluded exploring problem of audio in these devices beyond the problems already identified in other research [Morel and Licoppe, 2009].

Conversation

This activity provided an interesting basis, as it is a basic use case. It was completed by sixteen (16) participants in an average of thirteen (13) minutes without any major technical problems.

When presented with a choice of locations to hold the conversation, all participants looked for an area with less people. During this activity, participants sat and held the same position (Figure 6) for almost the entire length of the conversation, often trying to rest the device in front of them or fashioning temporary supports. After a few minutes, some would show signs of discomfort, but did not change their position. None of the participants chose to stand during the activity, later mentioning that it was important for them to be comfortable during this time.



Figure 6. Observed position during conversation tests

As the activity went on, the environment seemed to be of little importance, as most participants concentrated on talking to the other person and would only look elsewhere or mention their environment if something was perceived as annoying, such as a very hot room, or if someone walked by close enough for them to feel threatened.

Participants identified audio/video synchronicity as being very important for the conversation, because they felt it became difficult to understand the other person when sound and video were transmitted at different speeds. Sound was identified as being more important than video; participants felt they were properly connected if they had sound, they could deal with less-than-real-time video situations. This was something already mentioned in previous research (Tang And Isaacs, 1992), when video was limited by technology, but remains valid. The “mirror image” presented to users was also seen as fairly distracting; while some admitted to spending part of the conversation looking at their image rather than the other callers’, it was described by a participant, “as talking to someone while having a mirror next to their face”. Others felt pressured to stare at the distant caller. They mentioned feeling that the other person was constantly watching them and felt compelled to try to keep looking at them. They stated that in a normal conversation they would spend more time looking around.

When questioned about their experience, participant noted a clear tendency toward *flow*, followed by control and, at the end of the activity, boredom. When comparing the task load index (24,83) to the other activities, we find this activity to be the lowest scoring of the three. The load is described as being mostly tied to effort, followed by mental demand (Figure 8). Considering this was an activity all participants had already experienced, this last step is consistent with tools that they quickly mastered and that would not provide new challenges.

In most of the conversations, the focus of the participants was on the activity (Figure 7); while there was little mention of other aspects of context, save for a few mentions of past events. We also observed that the end of every videoconference was very structured, as users would often greet the distant caller twice before shutting down the connection.

Presenting a space

This activity introduced mobility in the conversation and left the participants in control of how their environment was communicated. Most tests were completed without any problems, but some experienced short losses of signal, which caused the connection to drop. On average, this activity was completed in 14 minutes.

In contrast to what was observed during the conversation, participants experimented with different ways of showing their environment or answering the requests of the distant research assistant.

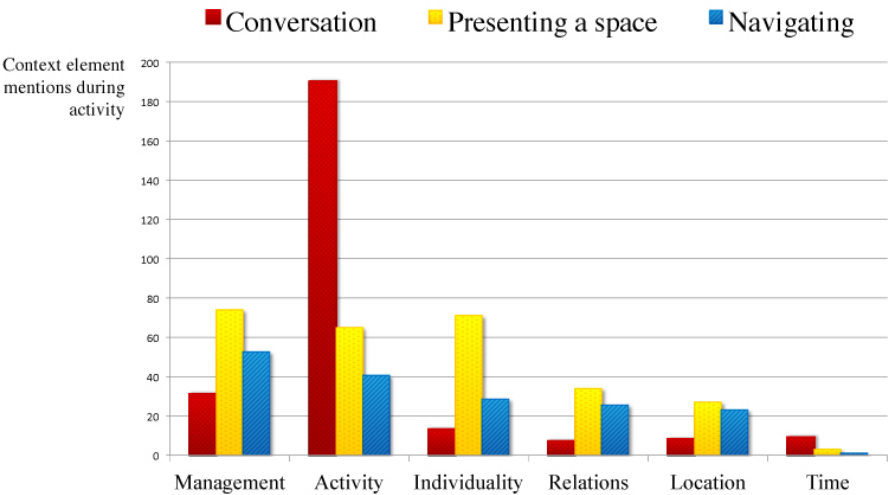


Figure 7. Proportion of context elements mentions per activity

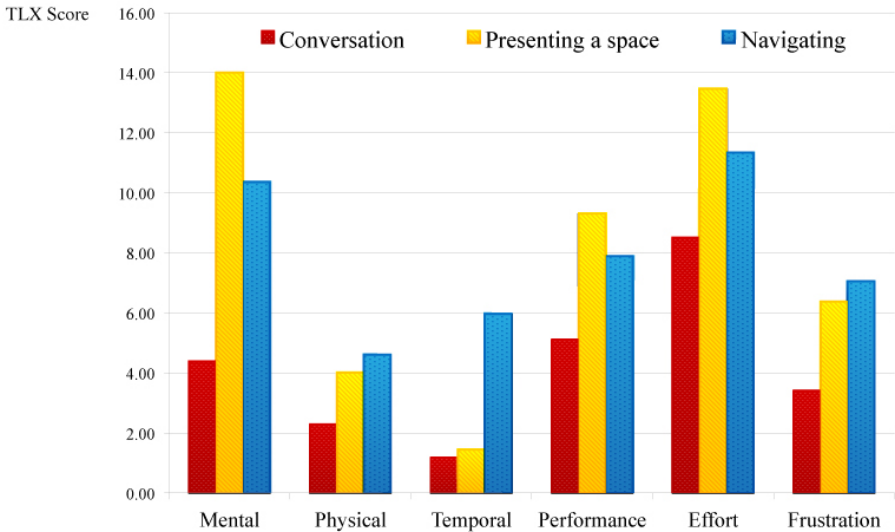


Figure 8. Workload scale per activity

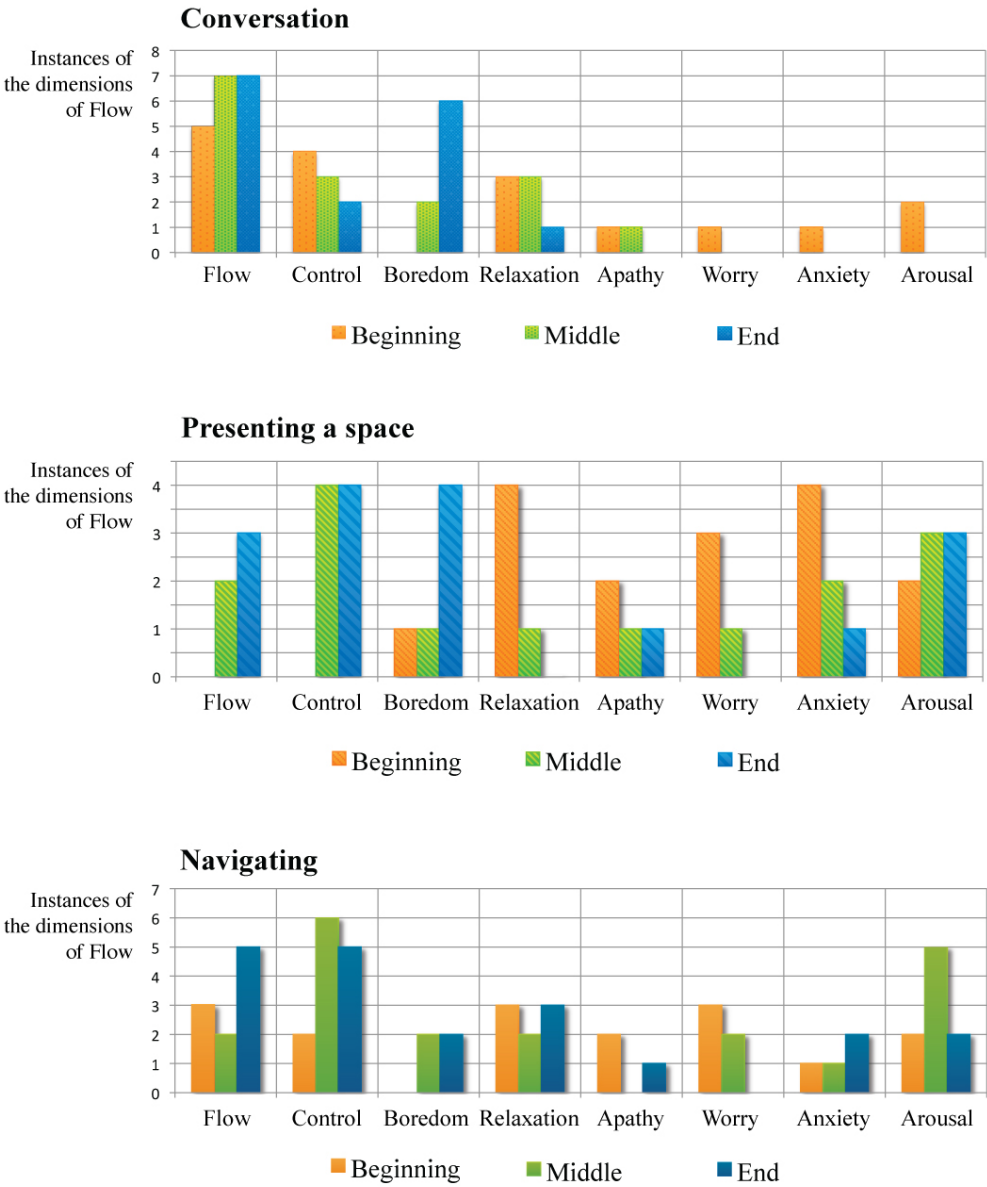


Figure 9. Number of instances of the dimensions of Flow as noted by participants per period of activity.

Near the beginning of the activity, participants would spend a few moments talking with the interlocutor. It is during this period that the participants would receive the instructions for the activity, but it was also an opportunity to notice the position the participants would take. In this situation, most participants would remain in place, holding the phone at face level (Figure 10).



Figure 10. Observed position early in the tests

When presenting a space, participants would attempt to frame elements of the exhibit and maintain a stable image, while looking around to figure out what they would show next.

We also observed that users would experiment in ways to indicate elements or provide more precision within the image that they were sending, often trying to point towards something on the image by pointing with their finger in front of the camera or maintaining the camera in position while they read out loud the object's history. During this activity, participants dissociated audio exchanges from the visual dimension. While they maintained a verbal conversation with the research assistant, they would be scanning their environment, looking for further things to discuss. They treated the device like a camera, panning, zooming and trying to provide stable images. When asked about the relationship they felt to the device at that point, nearly all of them still viewed the device as a means of real-time communication, rather than a camera.

While the majority of participants followed this pattern, a participant that did not change the camera orientation proceeded to verbally relate the experience of the space. The participant would show the general environment by turning the phone towards the environment and then returning to a face-to-face relationship to verbally describe the details of the environment.

When looking at the participant's experience, we find that this activity progresses from *relaxation* and *anxiety* towards *control*, *flow* and *arousal*, a progression similar to the one described by Csikszentmihalyi [Csikszentmihalyi, 1990].

The discussion analysis shows that participants are mainly linking the spaces and objects they are presenting (Figure 5), then identifying their context and maintaining a focus on the activity aspect of their context. Their experience during this test is divided equally between arousal, control and anxiety (Figure 6). Workload data show that this was the activity with the highest workload, mostly characterized by high mental demand, effort and performance (Figure 7).

Navigating

The navigation activity was certainly the most challenging one, since it required the participants to navigate in a changing environment while communicating. Network quality was an issue during this test, as some sections of the pavilion tend to block Wi-Fi coverage, cutting the video call and requiring a few seconds to reconnect. Even if precautions were taken to try to keep connection losses to a minimum, all participants had at least one incident during the test. On average, this activity was completed in 8:50 minutes, with a standard deviation of 1:59 minutes. It should be noted that since participants could get lost or miss instructions, their path was slightly different in each case.

In comparison to behaviours seen while presenting a space, where there was some experimentation with different means of showing the space, participants' efforts were focused on sharing their experience of the environment. From the beginning of the activity all participants chose to use the back-facing camera, keeping it in front of them and using the image provided to see their space, almost like a window. Participants would then try to maintain this position, breaking away from it only when they found themselves in front of complex obstacles like stairs.

Furthermore, we observed decisions being taken purely based on the shared video feed (as opposed to the real context surrounding them), as participants would choose which way to go by showing the corridors to the distant research assistant, who would identify which one to choose.

When describing their behaviour, participants said they tried to produce a stable image, avoiding sudden movements, considering that what they saw was what the other person was seeing. This notion was also at play when trying to explain to the assistant their position on a map, moving a hand in front of the camera, pointing out their location (Figure 11).

Once the participants understood that the communication could fail, they often checked with the distant research assistant to make sure the connection was still fast and lag-free.



Figure 11. Observed pointing behaviour

An analysis of their exchange reveals that they mostly exchanged about their activity, their individuality and their relations (Figure 7). Indeed, many decisions during the test were taken by referencing what was being shown on the camera. Participants mostly experienced this activity as being in control, followed by flow and arousal (Figure 9). Workload for this activity was higher (more demanding) than holding a conversation, ranking effort, mental demand and performance as an issue, but also noting a high temporal demand, which none of the other activities identified (Figure 8). It should be noted that technical problems during this activity were higher than in the other two activities, which some participants identified as an elevated source of frustration.

In this case, videoconferencing appears to create a limited context shared by both users, rather than an exchange of contexts. In this space, framed by what's being shown through the device, many decisions are taken with information relative to both users, for example defining a location to go to as "there".

During exit interviews some commented on the interface's behaviour during this phase. A few participants noted a difficulty in switching cameras. This process was seen as slow, which is why they chose to use only one camera, the back camera, as soon as possible. This may be more specific to Skype's mobile interface, as it hides the button to change cameras after a few seconds.

6.2. Focus groups

The purpose of the focus groups was to understand how videoconferencing was used in an everyday context and to gather information about how they related to

videoconferencing. The participants of two groups of habitual users had begun using mobile videoconferencing one to two years prior to this research.

Participants from both focus groups saw mobile video-conferencing as a way to primarily contact people they hold dear. The issue of privacy was raised during a session, as the participants agreed that a mobile video call requires the users to be in an environment that they feel is private enough. The first group, forming a family group, would arrange video calls with a member of their family and would put someone in charge of maintaining visual contact with the distant caller. This way, they managed to exchange during other activities, such as meal preparation and mealtime. The second group included three young males, aged between 25 and 30 years old. They clearly identified videoconferencing as a privileged communication medium, as they communicate with people they already have a relationship with and would frown on the usage of the technology for more commercial purposes, such as being contacted by their mobile phone service provider.

A majority of them owned an Apple iPhone™, used Facetime™ as their main videoconferencing application and held at least one 5 to 10-minute conversation every week. They generally do not prepare in any particular way before making a video call, but they do make sure they are not in a noisy room or somewhere with low light.

7. ANALYSIS

While current videoconferencing tools seem to properly support face-to-face communication, they do so in a fairly restrictive way, reducing user mobility and asking that they maintain a position the devices do not support. Moreover, there is a marked difference between what was identified when users were presenting themselves and when they were presenting a space.

When users were using the front facing camera to talk to each other, their mobility was reduced, as they struggled to maintain a position where their image would be judged acceptable and while keeping track of their context. Indeed, during any of the three activities, a switch to the front facing camera would cause participants to slow down and eventually stop moving. This behaviour supports the need to replicate an “acceptable” picture, a view of the face and shoulders of both participants, which was identified in previous research [Licoppe and Morel, 2012].

But there was a marked difference in how participants behaved when using the back camera of the device visible in the workload and optimal experience data. During the navigation and exploration activities, the video image captured by the device became a way for participants to explore the space, almost as if through a window. In this mode, real-time interaction is not as important as the capacity to convey what participants understand of their context, as it tends to act as common ground between the two callers. Context is used in a broader way; the users experiences are less optimal and are under a higher workload. Participants also tended to separate more the spectating and the directing activities mentioned by Morel and Licoppe [2012], also described in other multimedia installations [Dalsgaard *et al.*, 2008], which raised different communicational needs, such as pointing an element on the shared image.

It should be noted that observations from the tasks do not quite represent the routine usage, but that it provides insight into how users would approach these situations. Furthermore, the behaviours observed only cover one side of the conversation, as participants were talking to a research assistant. We tried to reduce biases that this could introduce by having someone the participants knew as the research assistant.

8. PROPOSED DESIGN GUIDELINES

Our approach is to take a stance as designers and explore solutions to the problems found in this study through a design process, in line with approaches proposed by Findeli and Coste [2007] for project-grounded research in design. Following our analysis of the different observations concerning the context, the psychological states and the workload, we propose some design guidelines for future mobile videoconferencing interfaces. We hope they will begin to answer the needs and problems identified during our observations, so that we can provide better ways for users to express themselves and ways for designers to rethink video communication in a mobile age.

Interfaces for meeting, interfaces for showing

Comparing our observations, data from optimal experiences and workload, we find a marked difference between using the front camera and the back camera. Those measurements suggest that the activities that would involve the latter require tools or features that could help users convey their context.

Mirror image made invisible

While our observations suggest that the videoconferencing tools available for face-to-face communication provide a pleasant experience, they still present some very interesting design challenges. Chief among them is the need to maintain a visual contact with the other person, which was seen as demanding by some participants. For example, the mirror image is necessary during the first few minutes of a video call, but it could be made invisible to oneself, as long as the device is kept stable or as long as the face of the user can be seen in the picture. This could reduce the distractions identified during our interviews, while keeping the image available when needed.

Rethinking interactions

Taking full advantage of devices currently available, videoconferencing interfaces could be designed to reduce reliance on buttons, opting for more direct ways of interacting. For example, a touch screen could also serve as a way to create a *pointer*, transmitted to the remote user to create gestures, simple sketches or directions (Figure 12). This would support behaviours observed when exploring or navigating spaces. Gesture controls could help provide more spontaneous controls, as long as care is provided to avoid gestures that could make the device harder to hold. For example, holding a finger on screen could freeze the image, letting users explore what is being shown.



Figure 12. A possible pointer system, used by moving a finger in the screen

The same goes for the accelerometer usually found in mobile platforms. For most of our tests, participants would hold the device vertically to talk to the distant research assistant, but put it sideways once they wanted to show something. This could be a very simple way to trigger the switch from the front to the back camera as well as showing the image that is most adapted to the screen orientation (Figure 13).

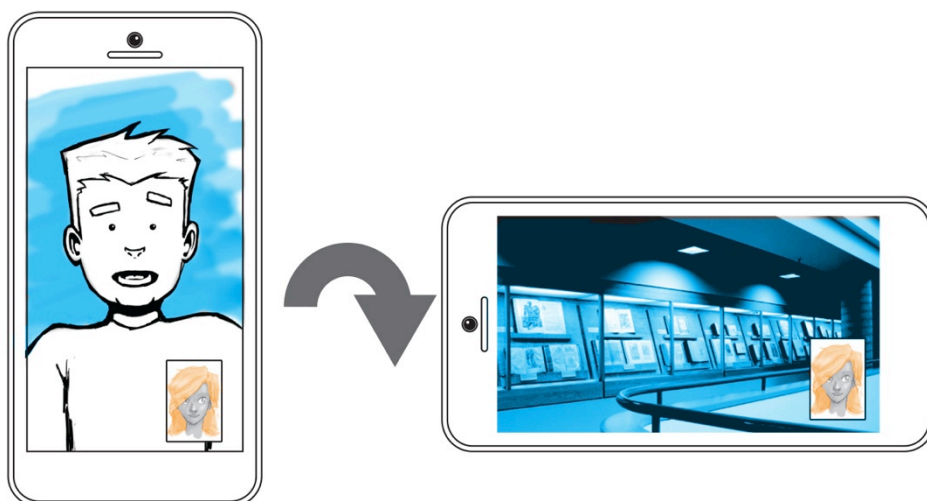


Figure 13. Switching cameras on orientation change

Stopping a conversation with the use of a simple button puts users in a position to inadvertently turn off the communication, as witnessed during one test. Designers should take this into account as some patterns already in use, such as swiping a button to confirm, would render the choice to end a conversation a conscious move and make sure the method to disconnect cannot be triggered by mistake.

Creating meaning

Creating meaning by conveying a message or adding meaning to what is being shown was a challenge for participants as it was often done against the real-time nature of the videoconference. In every one of our tests we saw participants try to go around the camera to show something in the image, or to try to keep a shot steady so they could be sure the other person could see them well. Much meaning relies on what is shown, and therefore controlling the framing of the image becomes a lynch pin of videoconferencing. Users were observed compensating for the lack of stability in the image as well as for its lack of precision, by pointing out details by moving one hand in front of the camera. Offering a way to create points of reference in the video could offer a way to provide the precision needed. Adding image stabilization, something that should be technologically feasible in the foreseeable future, might help users to require less work to maintain a stable image, but it is also important to allow them to concentrate on a single image. This could be done in a variety of ways, such as pressing and holding on the video, but it would mostly free the users to move about if needed, without having to sacrifice the possibility to discuss a point of interest.

Expanding on this idea opens possibilities to purposefully stop the video, rewind it for a few seconds, or to replace it by a picture or a video captured earlier, allowing users to review or to maintain an image without being physically tied to their position. During our tests, sound was identified as more important than video in order for the subject to feel the presence of the other caller, which could support a disjunction between video and sound. This could mean a major change for current tools, going against part of its real-time nature, but it could provide some very interesting possibilities. By maintaining a constant sound presence and providing control over the video, we create the possibility of using video as a support to the communication, as well as providing a measure of image control to the users without sacrificing the connexion they have established.

9. CONCLUSIONS

Overall, our observations suggests that the videoconferencing interface analysed here worked well for the purpose of holding a conversation, but still presented a challenge to maintain this exchange over time in a comfortable way. The constant presence of voice communication was identified as helping maintain the feeling of presence, creating at times an optimal experience for its users. Certainly a challenge lies with returning some mobility to the user who is almost immobilized during a call by the real-time nature of video communication. Users feel responsible both for the quality of the image and, especially when showing the environment, the quality of the content. If earlier research inferred that users would benefit from more image control [Dourish *et al.*, 1996], our observations suggest that controlling the image in real-time is actually very demanding for the user who feels pressured by the immediacy of the exchange and that there is a need for a different approach to image control that would provide more flexibility to the callers.

This requires not only a revision of current interfaces, but also to rethink how this type of communication is presented to the users. As such, our data seems to corroborate the findings of previous research [Licoppe and Morel, 2012][Morel And Licoppe, 2009], so it would seem that users might still be adapting to the new technology. However, there is a notable difference in how the users behaved and used videoconferencing when “sharing” their context with the back-facing camera. The content of their discussion was different; the tools received higher workload and were less conducive to an optimal experience. The behaviours observed suggest that the current tools available are lacking

in many ways and require an increased flexibility in image controls to fulfil the user's need to convey and create more meaning through what is being shown.

In fact, our observations have led us to believe that current mobile videoconferencing applications behave as old technology in new systems, and that they do not really provide the flexibility needed in a stationary conversation, when moving or showing something. The mobile videoconference interface needs to evolve towards the user experience, by enhancing, augmenting it.

The proposed design recommendations have yet to be tested, but we believe the contributions raised by this study are relevant and could be prototyped and analysed in further work. The challenge facing the design of mobile video-conferencing interfaces is dual: Keeping the relationship with the distant caller and the creation of a context shared by both callers. The introduction of videoconferencing to a mobile context is then not only a source of new problems, but a chance to create better, more meaningful communication tools.

ACKNOWLEDGMENTS

The authors would like to thank Annemarie Lesage, Edgar Pérez, Olivier Plante, Alberto Restrepo and Isabelle Deslauriers for their support. We would also like to thank the MATI-Montreal, The Faculty of Post-Doctoral Studies of the Université de Montréal, CITÉ and the Bell chair for new technologies for the financial support.

RÉFÉRENCES

- AMES, M., GO, J., KAYE, J. AND SPASOJEVIC, M. Making love in the network closet: The benefits and work of home videochat. *CSCW'10: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, (2010), 145-154.
- ANONYMOUS. Public Television in Germany. *NATURE magazine* 137, p. 391 (07 March 1936), Retrieved May 05, 2013 from <http://www.nature.com/nature/journal/v137/n3462/abs/137391a0.html>.
- CRESSWELL, T. *On the Move*. Routledge, London, UK, (2006), 3-4.
- CSIKSZENTMIHALYI, M. *Flow: The Psychology of Optimal Experience*. Harper and Row, New York, USA, (1990).
- DALSGAARD, P., DALSGAARD, H. AND KOEFOED, L. Performing perception - staging aesthetics of interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, (2008), 1-33.
- DEY, A., ABOWD, G. AND SALBER, D.A. Conceptual Framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, 16 (2), (2001), 97-166.
- DEY, A., Understanding and using context. *Personal and Ubiquitous Computing*, 5 (1), (2001), 4-7.
- DOURISH, P., What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8 (1), (2004), 19-30.
- DOURISH, P., ADLER, A., BELLOTTI, V. AND HENDERSON, A. Your place or mine? Learning from long-term use of audio-video communication. *Computer Supported Cooperative Work*, 5 (1), (1996), 33-62.
- FINDELI A., COSTE A. De la recherche-cr  ation    la recherche-projet: un cadre th  orique et m  thodologique pour la recherche architecturale. *Lieux communs*, 10, (2007), 139-161.

- HART, S.G. AND STAVELAND L.E. Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Hancock, P.S. and Meshkati, N. (eds), *Human Mental Workload*. Amsterdam, North-Holland, (1988), 139-183.
- KRAUT, R.E., GERGLE D., AND FUSSELL, R.S. The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. *Proc. 2002 ACM conference on computer supported cooperative work*. ACM Press, (2002), 31-40.
- MCDONALD, S. Studying actions in context: a qualitative shadowing method for organizational research. *Qualitative Research*, November 5 (2005), 455-73.
- MOBILE [DEF 1]. (N.D.). retrieved march 28, 2014, from <http://www.merriam-webster.com/dictionary/mobile> 2013.
- MOREL, J. AND LICOPPE, C. La vidéocommunication sur téléphone mobile. Quelle mobilité pour quels cadres? *Réseaux*, 4 (156), (2009), 165-201.
- LICOPPE, C. AND MOREL, J. Video-in-Interaction: "Talking Heads" and the Multimodal Organization of Mobile and Skype Video Calls. *Research in Language and Social Interaction*, 45 (4), (2012), 399-429.
- MOREL, J. Une ethnographie de la téléphonie mobile dans les lieux publics. *Réseaux*, 2, (112-113), (2002), 50-77.
- O'HARA, K., BLACK, A. AND LIPSON, M. Media spaces and mobile video telephony. *Media space 20+ years of mediated life*. Springer-Verlag, London, UK, (2009), 303-323.
- O'HARA, K., BLACK, A. AND LIPSON, M. Everyday Practices with Mobile Video Telephony. *Proc. CHI '06*. ACM Press (2006), 871-880.
- OULASVIRTA, A., TAMMINEN, S., ROTO, V., AND KUORELAHTI, J. (2005, April). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 919-928). ACM.
- REPONEN, E., HUUSKONEN, P. AND MIHALIC, K. Primary and secondary context in mobile video communication. *Personal and Ubiquitous Computing*, 12 (4), (2008), 281-288.
- SELLEN, A.J. Remote conversation: the Effects of Mediating Talk with Technology. *Human Computer Interaction*, 10, (1995), 401-444.
- TANG, JOHN C., AND ELLEN ISAACS. "Why do users like video?." *Computer Supported Cooperative Work (CSCW) 1.3* (1992): 163-196.
- URRY, J., *Mobilities*. Polity Press, Malden, UK. (2008).
- WEILENMANN, A. *Doing mobility*. Doctoral dissertation. Gothenburg Studies in Informatics. Göteborg University. Sweden (2003).
- WOLFRAM|ALPHA (2012). Retrieved May 14, 2013, from <http://www.wolframalpha.com/> with query [average age of cell phone users].
- ZIMMERMANN, A., LORENZ, A. AND OPPERMAN, R. An operational definition of context. *CONTEXT'07 Proc. 6th international and interdisciplinary conference on Modeling and using context*. Springer-Verlag, Berlin (2007), 558-571.



Ignacio Calvo holds a masters (M. Sc. A.) in Design and Complexity. He is a research assistant at the Hybridlab and UX designer for Yu Centrik. He also teaches advanced computer graphics and animation at the school of industrial design of the Université de Montréal.



Tomás Dorta has a background as a practitioner architect and designer. His research interests include the design process and co-design using new technologies and the development of new techniques and devices of design in the virtual realm. He obtained his Ph.D. (2001), studying the impact of virtual reality as a visualization tool into the design process. His research has been extensively funded by Canadian's federal and provincial research grant institutes and published and presented in several international scientific conferences as well as scientific journals. As a design educator, Tomás Dorta joined the School of Design of Université de Montréal in 1997 where he is now associate professor. He teaches design research and computer graphics courses and industrial design studios related to new technologies, design collaboration and bionics. Tomás Dorta is the director of the Design research laboratory Hybridlab.



Jean-Marc Robert est professeur à l'École Polytechnique de Montréal où il enseigne l'ergonomie cognitive appliquée aux interactions humain-machine (IHM). Il détient un baccalauréat et une maîtrise en psychologie de l'Université de Montréal et un doctorat en psychologie appliquée (ergonomie cognitive) de l'Université Paris V (La Sorbonne, 1982). Il a été chercheur à Xerox-PARC (Californie, 1983) et a fait des études postdoctorales au NASA-Ames Research Center (Californie, 1983-1985) dans le domaine des IHM. Ses recherches portent sur l'expérience utilisateur, l'utilisabilité des IHM, la conception et l'évaluation ergonomique des IHM, et l'ergonomie prospective pour la création de futurs produits et services.