Remedying Video Call Paralysis: An Exploration of IoT-based Support for Telepresence in Meetings

Salu Ylirisku
University of Southern Denmark
Universitetsparken 1, 6000 Kolding
ylirisku@mci.sdu.dk

Peter Heinzl
Ondrej Henek
University of Southern Denmark
Universitetsparken 1, 6000 Kolding

Abstract
In the paper we address the problems that many people face when participating in a meeting through a video call, such as Skype or Google Hangouts. Compared to a person, who is embodied in a situation, the person who participates through a video call is severely handicapped. This influences not only to how well a person can participate in the ongoing interaction, but also on how the others who are co-located treat the person calling. We present a design concept named Point'n'Turn, which uses IoT to alleviate some of the limitations in embodied interaction of the caller.

Author Keywords
Telepresence, video call paralysis, user interfaces, IoT

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

What is Video Call Paralysis

Symptoms: You cannot move your head, you cannot see nor hear clearly, nor touch of point at things when you are calling into a meeting. People cannot see where you are looking at. Sometimes you think that what you say becomes heard, but after a moment you get notified that some time ago your voice was cut. You come to wonder from the others: “What exactly you didn’t hear?”

Responses: People treat you as handicapped and slow, which causes them to invest extra energy to take care of your special needs.

Effects: Video call paralysis lowers the intensity of interaction and the flow of discussion and, especially, of ideation.

Introduction

It has become a commonplace that one or more of the participants in a meeting cannot be physically present in the location. An often-used solution is to use internet-based video calls, such as Skype or Google Hangouts to establish the connection to the meeting. This solution, however, often leads to the calling participant to be limited in significant ways. Especially this is visible on moments when the co-present participants begin to present, discuss, and use visual and textual materials in their interaction. The person participating through skype cannot turn the camera, cannot see clearly due to pixelation, cannot hear well due background noise reduction, has limited capabilites to give subtle hints if one wants to say something without talking over the others, and one cannot point at or touch things. These deficits not only limit the person calling in one’s ability to fully participate, but also results in the others, the co-located people, to treat the person on the video call as a particularly disabled one.

Discovering Video Call Paralysis

As part of an MA Apprentice course in the ITPD study progamme at the University of Southern Denmark, we conducted a project to explore possibilities to support meetings where one of more of the participants would be joining remotely. Our intent was to understand how IoT could provide support for the apparent deficiencies of participation that the caller has. We began our study by going through video recordings of our previous projects. We studied a project meeting in Cambrigde UK, where one of the participants was calling in through Skype. We also studied two project meetings in real projects at SDU design.

Below we explicate how people respond to a person participating in a meeting through skype.

Turning the video call laptop

Figure 1. Others in the meeting flip the computer with video call to different directions.

In meetings it is often important to pay attention to surfaces around the space in which the meeting takes place. We observed that people are actively turning the laptop in order to keep the person on the video call able to participate in discussions about particular things on walls.

Holding the video call computer

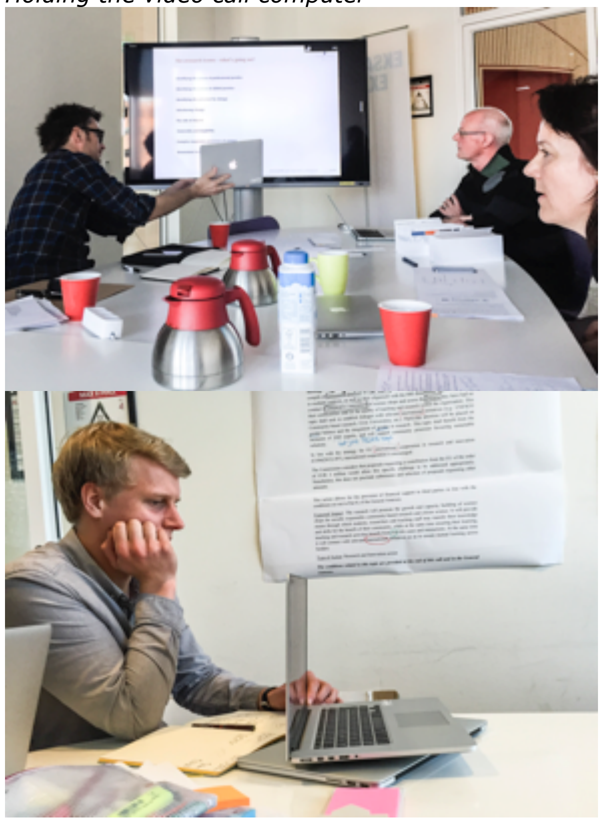


Figure 2. Others enacting as the body of the caller

People enact as the body for the person on video call. They hold the laptop and help the person on the call to see what is being discussed.

Talking like to a severerly impaired

Figure 3. Explaining to the skype paralysed makes interactions more worksome.

This involves exaggerated facial and bodily gesturing and slow and clear verbal articulating. The display of artefacts is done patiently to get the image still and focussed.

The handling of things and pointing at these is done by the co-present people

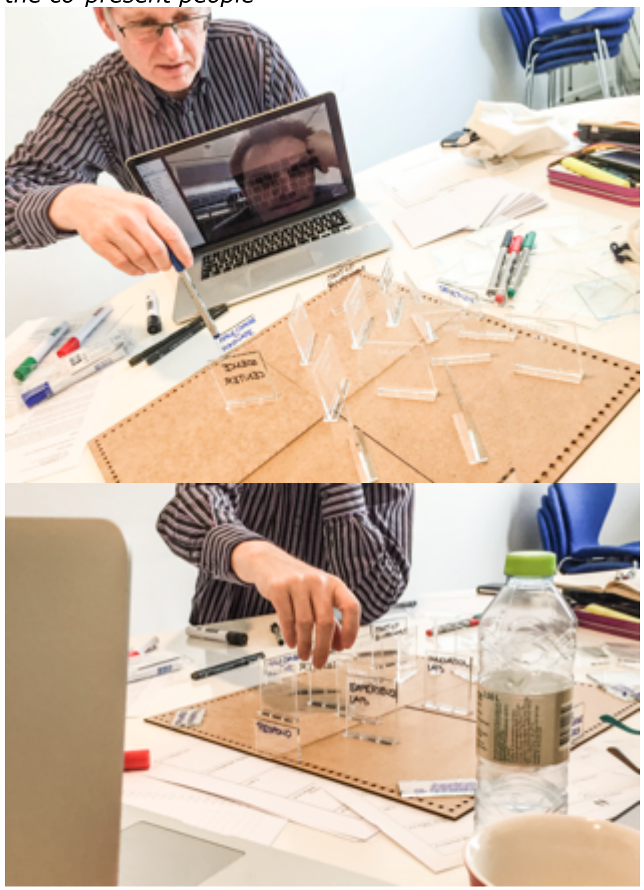


Figure 4. Participating directly in tactile interactions was impossible.

We observed how discussions of physical things happened by the co-located pointing at things and thus orhcestrating the discussion. The person in video call cannot point or touch things.

Missing the peripheral discussions

Figure 5. The person on video call become at many times an onlooker as the others in the meeting discussed with each other informally.

The ability of co-present people to face each other and to engage in subtle face-to-face discussions in parallel does not happen in video calls when a caller is participating in a meeting with multiple co-present people. There are parallel conversations that are missed by the person on video call. In these parallel conversations people are resourcing each other’s eye contact, physical orientation, and facial expressions in discussing.

We also observed that the person on video call had to use one’s voice to intervene discussions. In co-located situations people have various ways to indicate to others that they would like to take the next turn, e.g. by raising a hand, or by turning one’s face at someone.

We also observed that people are investing extra work to make the meetings more fluent. One of the strategies is to create alternative channels to mediate the communications, such as intranet sites, Google docs, and chats. Shared digital resources enable alternative paths of communication.

Research as design

In our research we followed the conceptual design approach approach as outlined by Ylirisku et al. [16]. We organised our research process as a design process, which was aimed at developing and presenting novel conceptual understandings of unexpected opportunities. The process included an initial field study phase, where we collected experiences from meetings, where one participant was participating through video call. We also used our own project as an example, since one of the team members was travelling a lot, and participated in the project meetings often through a video call. Hence, we also set up altenative channels to support the discussions. These included Trello, Slack, DropBox, GitHub, personal web servers, and e-mail.

We also experimented how difficult it is to do collaborative sensemaking with post-its when one of the participants is not in the same room, and we ultimately decided to postpone our conclusions of the best ideas until a moment when we were all co-present. It was a crucial step in the process to group the findings and choose the key ideas to focus on. As the sun was shining from the window behind our white board, we got new ideas about pointing, see Figure 6.

Figure 6. Pointing with shadows.

The design concept

Based on our exploration and ideation, we developed a design concept. The purpose of the concept was to support people participating through video calls in getting better involvement in on-going action. We outlined three key principles for the way in which we thought the design object should be created.

1. Building on what users already have, e.g. laptops, tablets or smartphones with skype or similar telepresence software
2. Not having to install anything
3. Using standard web technologies

With these principles it was intended to make the resulting design object affordable and better accessible to people than the expensive existing solutions (see related work section below).

We ideated different components to address different areas of the video call paralysis. We envisioned

1. a turntable to enable turning one’s face,
2. a laser pointer to be able to point at things,
3. a signal light to indicate that one would like to take a turn,
4. a zoomable camera to enable seeing small texts clearly,
5. a directional or throwable microphone to be able to attend one’s hearing better and to overcome the way in which video calls cut off the lower sounds,
6. an indicator to show that the call-loop is broken, i.e. that the people in the meeting room do not hear what you say, to avoid the “we did not hear you” problem

We had also ideas to support grabbing objects, moving the camera in space, and having a 360° view to use to give a peripheral vision. We, however, choose to focus only on a handful of our ideas to make an actual working prototype. We decided to focus on two key features: to be able to point and turn.

The Point’n’Turn design prototype

Based on searching of alternative physical designs to enable turning a laptop around, we decided to go with an approach that we call the ‘Lazy Susan’. It is a revolving stand or tray on table. Based on our design principles we decided our turntable to be such that it would have its own battery. We also envisioned that if the prototype would be developed into a commercial product, it could have a 4G-network connectivity to avoid problems with connecting to the local wifi, which often asks for a password. In our design we used the ESP8266 IoT platform (embedded into the Adafruit HUZZAH board), which needs to be given the new password for each different wifi. And this introduces complications for designing the interactions for starting the use of the device. In our final design prototype we decided to require people to set WiFi credentials and a unique device ID into the program code. It was the simplest option, but not very user friendly. So, the wifi password needs to be entered into the code that is compiled and uploaded to the ESP. However, one can make things simpler by using, e.g. one’s own smart phone to set up a wifi hotspot to be used by the ESP to connect to the Internet. This is a sensible alternative as the communication of the ESP uses rather limited amount of data through a WebSocket connection.

The physical design was first mocked up with a book on the table and old LP record on top of it to simulate the behavior of the platform. After putting a laptop on it with a video call on, we simulated the rotation, and the resulting experience was encouraging. The final design of the turntable is shown in Figure 7. It is battery driven and can be put underneath a laptop and it is ready to work. It can also be powered through USB.

Figure 7. The final design of the turntable.

The laser pointer feature was designed to sit on the shoulder of the laptop. It would enable the person in video call to put on the light and move it at a desired location within the video view. The prototype version of the laser pointer is shown in Figure 1Figure 8.

Figure 8. A prototype of the laser pointer functionality.

The design of the laser pointer was intended to be implemented with tiny stepper motors, but these were too unreliable for the purpose. The current implementation uses two servos instead, although, causing undesired noise when turning, and being too inaccurate for the purpose.

Software design

Software design could be broken into 3 parts: The ESP8266, the controller and communication between them.

ESP8266

Before the sketch is uploaded to the ESP8266, it's required to set WiFi credentials and a unique device ID - it's the simplest option, but not very user friendly. For the platform's stepper motor we use Accel Stepper library[1] as it is nonblocking because we want the device to be responsive all the time. For controlling laser pointer we use the default Servo library and it's command moveMilliseconds for better movement precision.

Communication is based on WebSocket standard because it is open and well supported in modern browsers. To enable WebSocket in ESP8266 we chose the library Links2004/arduinoWebSockets[2] as the only developed and supported solution. Because the ESP8266 could be connected from any WiFi, it cannot function as a server and browser neither. Therefore we utilized a free WebSocket server achex.ca[3], which works with lightweight JSON protocol. This is a native language for web browsers but ESP8266 requires additional library[4].

The controller

The controller is built as a responsive web application in HTML and JavaScript. The graphical user interface layer utilizes Velocity library and jQuery library. It is split into two parts, one controls the laser pointer and the other the platform. At the beginning user is prompted to input device ID which has been set in the remote

[1] http://www.airspayce.com/mikem/arduino/AccelStepper/

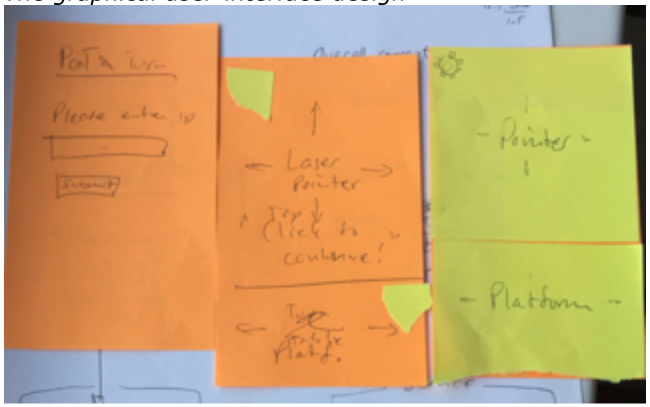
[2] https://github.com/Links2004/arduinoWebSockets

[3] http://achex.ca

[4] https://github.com/bblanchon/ArduinoJson

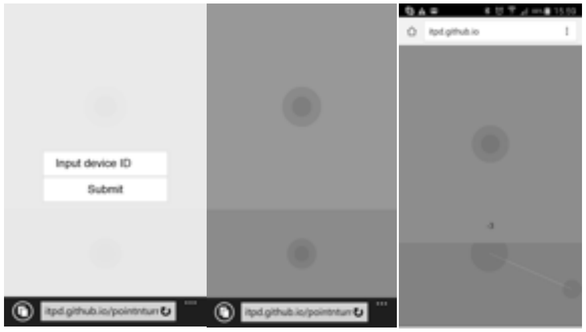
ESP8266. Then the user can interact with the interface by dragging control circles in desired direction and feedback is observed real-time through the video conference software. The controller utilizes our own JavaScript Point'N'Turn API[5] which is well documented and developed in a way that it can be easily extended or embedded into other projects.

The graphical user-interface design



We first created a paper prototype of the GUI and then implemented it with JavaScript. The idea of the interactivity was to use the smart phone of the person who calls as the control device.

The interface works simply by opening a particular URL on the web browser of the smart phone. It recognises when the user puts finger on the screen and then calculates how much the finger is moved and sends commands through the WebSocket to the EPS that controls the turntable as well as the laser pointer. The further the finger is moved from the original point, the greater speed values are transmitted to the ESPs. The upper part of the user interface controls the laser pointer and the lower part controls the turning table. The interface works also on computer's web browser.



Open source project

We decided to do this as an open source project. Therefore we based the production on laser cutting and the Arduino platform, which is widely used in the maker community. We used materials MDF and acrylic that is available in your local handyman center. The whole project is available through GitHub for anyone to use, build and develop further.

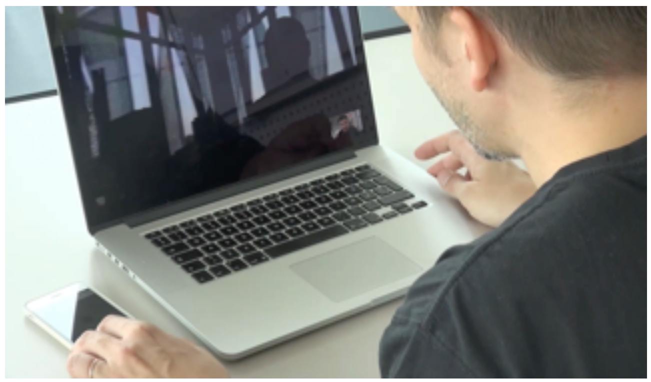
User study

We tested the prototype with two users. With the first user we observed an unexpected delay when accelerating and decelerating. Our test participant was using the GUI on smart phone by swiping it repeatedly, and as the delay was so long, this caused frustration. We used the 3 Volts from the ESP board to power the motor and the speed with this power was considered too slow. We omitted acceleration and deceleration phases of the motor to make the system more

[5] https://github.com/itpd/pointnturn

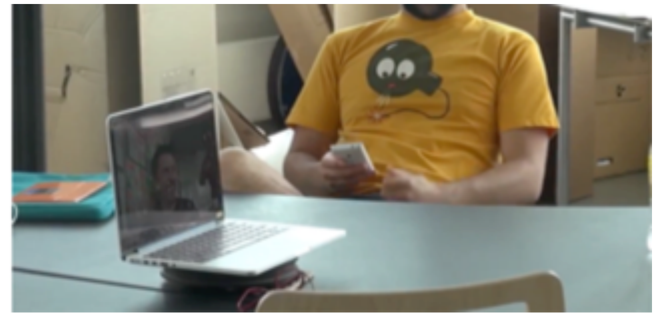
responsive. There will always remain some delay due to network connections and data communications.

In our second test we gave the functioning prototype with the turntable to a person to use for the first time. He was impressed of the ability to ‘drive’ the video from distance. It became possible to look around in the space.



(Starts turning the laptop through the interface.)
- “Wow! This is actually.. like driving.”
(Another person appears on the screen)
- “Hello Jack! How are you?”

The design prototype clearly gave the caller the possibility to establish a personal connection with one of the participants of a meeting by embodied interaction, i.e. by facing a person. The person calling now has a ‘body’ that is oriented towards another person, establishing a novel contextual configuration in the meeting room.



Technical considerations
While internally testing the laser pointer we were force to switch between several motors and servos. The main issue there is the size of one step which has to be low enough to offer precise control. Some servos were also not reliable in precision when we tested pointing repeatedly in two different locations. When we introduced our prototype to a wider audience we've also heard concerns about health relate to the laser beam. We tried to put boundaries to where the laser can point but this would also limit the versatility of the device in regards to the location where it is placed.

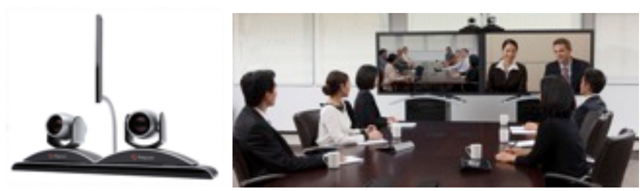
We were also observing users controlling the laser and platform with our web application on their mobile device. The graphical user interface doesn't offer a lot of affordance and cues for the user. Therefore they had to be instructed to move the circles in a desired way. We noticed interesting behavior when users were not holding the circle at desired direction for continuous movement but they were swiping rapidly over the screen to send more commands faster. We decided to put a brief introduction to the app and a "help" overlay which gives the user basic understanding of the control interface.

Related work

Supporting telepresence in meetings is a widely studied topic, and there also exist numerous products in the market to address challenges in telepresence. We first review existing products and then related research.

Existing products

Polycom RealPresence Centro[6]. The system captures an entire meeting room with a 360-degree camera and enables the person calling to see everybody simultaneously. The problem with the system is the lack of embodied signalling of where the person calling is attending. Also the system incorporates expensive custom components, and the price is only given through official enquiries.



Polycom EagleEye Director[7]. This camera locates automatically the person who speaks, zooms in and frames the active speaker. Voice triangulation and face-finding technologies move directly to the speaker. The price of the system is around $13 000, which reduces the interest of small teams to purchase it.

Logitech Conference Cam[8]. The camera has 180-degree rotation for panning. It also can tilt and zoom. The price is around $250, which makes it much more affordable than the Polycom systems. The Logitech Conference Cam, however, cannot be controlled by the caller, but needs to be controlled by the people in the meeting room.

RambleBot[9] is an inexpensive telepresence robot ($199), which can be controlled remotely online with an ordinary browser. There exists also a tall version of the RableBot ($299). The robot uses a mobile phone as the means to establish connection with the robot and with the incoming call.

---

[6] http://www.polycom.com/hd-video-conferencing/telepresence-room-systems/realpresence-centro.html?playVideo=Introducing_Polycom_RealPresence_Centro

[7] http://www.polycom.com/content/dam/polycom/common/documents/brochures/polycom-solutions-product-portfolio-br-enus.pdf

[8] http://www.logitech.com/en-us/product/conferencecam-bcc950

[9] http://www.ramblebot.com/

There exists currently a variety of different kind of remote presence robots, but we only included RambleBot for its affordability and that it shared many of our aspirations, e.g. supporting makers.

Related research

BiReality [7] is one of the first telepresence robots to act as a surrogate for a person, who participates in a meeting from distance. It features a large stroller with three displays showing the person who participates, and at the remote end there is a large screen that provides a wide-angle view of the meeting room.

There is an increasing recognition for better support of telepresence in meetings [8,9]. Giusti et al. [3] discovered that people invest significant effort in setting the stage for video calls when they need to work together. People use different kinds of ‘stage-setups’ for different kinds of tasks, such as presenting, sharing documents, face-to-face discussions, and comparing things [3]. There are different kinds of solutions to virtualise physical interactions to be done on computers [13], see also LiveStroke [4], .

Some of the research has focussed on exploring the design for the local end. Sellen and Buxton [12] explored the signifance of spatial cues in videoconferencing with the system called Hydra. Another study continuing on this was a study on multiple views to support collaboration [2]. Wallace and Scott studied how the co-located work can be supported with collaborative tables [14]. Tangibility is important in co-located work [6].

Another stream of research relates to enabling people to share resources across distance, see [5,15]. Studies of co-located vs. distributed teams are also relevant [1,10,11].

Conclusions

We presented a study that explored how the problems that we conceptualised in terms of Video Call Paralysis could be alleviated by the use of Internet of Things. We presented an inexpensive solution with several IoT peripherals (smart phone, turntable, laser pointer) that could be used together. A functioning interactive prototype was developed, tested, and distributed as an opensource project.

References

1. Andrew L. Cohen, Debra Cash, Michael J. Muller, and Curtis Culberson. 1999. Writing apart and designing together. ACM Press, 198. http://doi.org/10.1145/632716.632840
2. William W. Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One is not enough: multiple views in a media space. ACM Press, 335–341. http://doi.org/10.1145/169059.169268
3. Leonardo Giusti, Kotval Xerxes, Amelia Schladow, Nicholas Wallen, Francis Zane, and Federico Casalegno. 2012. Workspace configurations: setting the stage for remote collaboration on physical tasks. ACM Press, 351. http://doi.org/10.1145/2399016.2399071
4. Michael J. Gormish and John Barrus. 2014. Connecting content and annotations with livestroke. ACM Press, 131–134. http://doi.org/10.1145/2644866.2644884
5. Amar Gupta, Elisa Mattarelli, Satwik Seshasai, and Joseph Broschak. 2009. Use of collaborative technologies and knowledge sharing in co-located and distributed teams: Towards the 24-h knowledge factory. *The Journal of Strategic Information Systems* 18, 3: 147–161. http://doi.org/10.1016/j.jsis.2009.07.001
6. Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. ACM Press, 437. http://doi.org/10.1145/1124772.1124838
7. Norman P. Jouppi. 2002. First steps towards mutually-immersive mobile telepresence. ACM Press, 354. http://doi.org/10.1145/587078.587128
8. Eunyee Koh. 2010. Conferencing room for telepresence with remote participants. ACM Press, 309. http://doi.org/10.1145/1880071.1880124
9. Osamu Morikawa and Takanori Maesako. 1997. HyperMirror: a video-mediated communication system. ACM Press, 317. http://doi.org/10.1145/1120212.1120412
10. Toni Robertson, Jane Li, Kenton O'Hara, and Susan Hansen. 2010. Collaboration Within Different Settings: A Study of Co-located and Distributed Multidisciplinary Medical Team Meetings. *Computer Supported Cooperative Work (CSCW)* 19, 5: 483–513. http://doi.org/10.1007/s10606-010-9124-9
11. Jonathan Sapsed, David Gann, Nick Marshall, and Ammon Salter. 2005. From here to eternity?: The practice of knowledge transfer in dispersed and

co-located project organizations. *European Planning Studies* 13, 6: 831–851. http://doi.org/10.1080/09654310500187938

12. Abigail Sellen, Bill Buxton, and John Arnott. 1992. Using spatial cues to improve videoconferencing. ACM Press, 651–652. http://doi.org/10.1145/142750.143070
13. Sudi Sharifi and Kulwant S. Pawar. 2002. Virtually co-located product design teams: Sharing teaming experiences after the event? *International Journal of Operations & Production Management* 22, 6: 656–679. http://doi.org/10.1108/01443570210427677
14. James R. Wallace and Stacey D. Scott. 2008. Contextual Design Considerations for Co-located, Collaborative Tables. In *2008 IEEE International Workshop on Horizontal Interactive Human-Computer System: Amsterdam, Netherlands, 1 - 3 October 2008*. IEEE, Piscataway, NJ, 57–64.
15. Maria C Yang and Yan Jin. 2008. An examination of team effectiveness in distributed and co-located engineering teams. *International Journal of Engineering Education* 24, 2: 400.
16. Salu Ylirisku, Giulio Jacucci, Abigail Sellen, and Richard Harper. 2015. Design Research as Conceptual Designing: The Manhattan Design Concept. *Interacting with Computers*. http://doi.org/doi:10.1093/iwc/iwv040