Implicit and Explicit Interactions in Video Mediated Collaboration

Lyndsey Fisk, Marcus Carter, Behnaz Rostami Yeganeh, Frank Vetere, and Bernd Ploderer
Interaction Design Lab and Microsoft Centre for Social NUI
The University of Melbourne
Parkville, Victoria 3010, Australia
research@lyndseyfisk.com, {marcus.carter, rostamib, f.vetere, ploderer}@unimelb.edu.au

ABSTRACT
In this paper we report the results of a study comparing implicit-only and explicit-only interactions in a collaborative, video-mediated task with shared content. Expanding on earlier work which has typically only evaluated how implicit interaction can augment primarily explicit systems, we report issues surrounding control, anxiousness and negotiation in the context of video mediated collaboration. We conclude that implicit interaction has the potential to improve collaborative work, but that there are a multitude of issues that must first be negotiated.

Author Keywords
Explicit, Implicit, Interaction, Video Mediated Collaboration

ACM Classification Keywords
H5.2. User Interfaces: Interaction Styles.

INTRODUCTION
With the rise of "smart devices", *implicit interactions* are now a significant aspect of the human-computer experience. Systems that support implicit interactions serve the ideal of making technology invisible, seamlessly addressing the interests of the user without explicit instruction or supervision (e.g. a light that automatically turns on when you enter the room). These systems aim to operate proactively, outside the attentional foreground of the user (e.g. new email alerts from priority senders, or auto saving documents).

In this paper we explore the role of implicit interactions in video mediated (VM) collaboration. VM collaboration is an area of growing interest both socially and commercially, with everyday consumer technology possessing the capacity to facilitate video communication (eg. Skype). VM collaboration has not only become widely familiar, but it is evolving, in particular with complimentary emerging technologies that allow us to share content simultaneously with a video feed. We report the results of a Wizard of Oz (Kelley, 1984; Weiss et al, 2009) study exploring implicit-only and explicit-only interactive systems in the context of video-mediated collaborative tasks with shared content.

Our paper begins with a brief overview of the concepts underpinning implicit and explicit interactions. We then discuss how the use of non-traditional modalities, such as gesture and voice, are particularly useful in interrogating implicit and explicit interactions in VM collaborative tasks, before reporting the results of our study.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

OzCHI '14 , December 02 - 05 2014, Sydney, NSW, Australia
Copyright 2014 ACM 978-1-4503-0653-9/14/12…$15.00
http://dx.doi.org/10.1145/2686612.2686650

IMPLICIT AND EXPLICIT INTERACTIONS
Albrecht Schmidt (2000), one of the first authors to use the term *implicit interaction* in HCI, contextualised his definition within the example of when information is exchanged between humans 'implicitly', such as by body language, gestures, tone, and context which serves to help disambiguate communication. He defines implicit interaction as *"an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input"* (2000, p. 192). Schmidt is primarily concerned with how to use sensors and other technologies to make systems more aware of their context, and argues implicit interaction will mostly be used to augment explicit interaction.

Schmidt's "*simple example*" of an implicit interaction is a garbage bin that scans barcodes and reproduces a list of thrown away items as suggestions for a shopping list. He notes the action (throwing an item away) is not changed by the presence of the computing system in the garbage bin, and thus the system "*makes use of the action performed by the user"* but that "*the user does not explicitly interact with the computer"* to generate a shopping list (2000, p. 192), and indeed does not even need to be aware of (or consent to) the interaction for it to be 'successful'. The implicitness of this interaction is based on perception and interpretation, and the system's ability to "*act proactively"* by making use of an action already made. Thus, we see in Schmidt's context-aware approach, implicit interactions lying across dichotomies of intent, control, seamlessness and the proactivity of systems.

Another body of research explores implicit interactions in terms of activity monitoring (Atterer et al., 2006; Weakliam et al., 2005; Wilson et al., 2005; Blažica et al., 2013; George et al. 2000). These studies attempt to make use of "*unconsciously provided*" (Atterer et al., 2006) information, such as how long a user looks at a

photograph during photowork (Blažica et al., 2013) or how fast a user scrolls while reading a webpage (Atterer et al., 2006). This approach defines implicit interaction based on (a lack of) awareness and intent. While these studies use implicit interaction in a normatively positive way (improving usability, for example), they extend Schmidt's (2000) definition (which regarded implicit actions as 'primarily aimed' at technology) to include 'wholly unintentional', and perhaps sometimes 'intentionally withheld' actions. A system involving this type of implicit interaction suggests the potential for ethical concerns, moving from 'context-aware computing' to 'lacking informed consent computing'.

Wendy Ju (2008) and colleagues (Ju & Leifer, 2008; Ju, Lee & Klemmer 2008) have proposed a design-based approach to understanding implicit interactions in HCI. Ju and Leifer (2008) present an implicit interaction framework drawing on Buxton's (1995) division between background ("*peripheral*") and foreground ("*intentional*") attentional demand, further applying the perpendicular axes of *initiative* (p. 75), between reactive interactions initiated by the user, and proactive interactions initiated by the computer system. In this context, Ju understands explicit interactions as those where "*the user issues commands ... and receives overt feedback*" (p. 75), and implicit interactions as those that occur outside the attentional foreground of the user (such as auto-saving files) or if it is initiated by the system (such as a new email alert). In their work (Ju, Lee & Klemmer, 2008), have illustrated how implicit systems can work to support (co-located) collaboration between users on an electronic whiteboard, however there is no related work exploring implicit interactions in video-mediated collaboration.

VIDEO MEDIATED COLLABORATION

Explicit or implicit interactions that are based on communication modalities such as gesture or voice open up rich ways of interacting with technologies (Tse et al., 2006). These types of interactions do not impose significant spatial and gestural constraints. In fact, they act to ensure "*the interface will no longer be a barrier to users*", instead "*the interface will be* [the user] *and their gestures* [or other similar input modalities]" (O'Hara et al., 2013). This form of interactivity enables interactions in scenarios where traditional human interface devices are not appropriate (Wilson & Oliver, 2005).

A domain that largely suffers from limitations of constrained degree of freedom caused by traditional interaction methods is video mediated (VM) collaboration. The studies conducted by O'Hara et al. (2011), Greenberg et al. (2011), Hauber et al. (2006) show that the shared content in distributed collaborations requires users to interact through traditional interaction methods such as mice, keyboards and controllers typically on WIMP *(Windows, Icons, Menus, Pointer)* systems. The effect of these interaction methods in the context of collaboration is a restriction on mobility when interacting with their collaborators, both gesturally and spatially. Hence, integrating non-traditional ways of interactions with VM collaboration technologies has been suggested as a new direction for further research (O'Hara et al., 2011; Greenberg et al., 2011). Moreover, recent commercial attempts such as XBox Kinect, HP Halo and Cisco Telepresence already show the potential of such ways of interaction to create new enhanced experiences.

The suitability of gesture- or voice-based implicit or explicit interactions in the context of VM collaboration is heavily dependent on the nature of task. The findings from studies conducted by Olson & Olson (2000), Dourish et al. (1996) and Buxton (2009) show lack of context, common grounds, awareness, and interfacing capabilities often reduce the effectiveness of VM collaboration environments. However, not all tasks suffer from those deficiencies. Instead, there are some that gain beneficial values in distributed collaborations through video. Thus, introduction of non-traditional interactions could have the potential to enhance those VM collaboration experiences.

Remote negotiation has been recognised as one of the tasks that take advantage of VM interactions. For social tasks primarily involving negotiation and bargaining the visual channel is of significant importance, however "the video channel is of little benefit" for tasks focusing on collaborative problem solving (Anderson et al., 1996). Similarly, Veinott, Olson, Olson, & Fu (1999) find VM negotiation of common grounds as an exceptional task to which quality of work improves by adding a video channel. Therefore, VM collaboration is most effective in tasks with a "high common ground and loosely coupled work" (Olson & Olson, 2000) such as a negotiation task using shared content.

Although the research reveals the potential benefits of adding non-traditional interaction methods (such as gesture and voice) to remote negotiation tasks with shared content, it could also raises new complications. Harper & Mentis (2013) discuss that human behaviour changes when they articulate their actions towards interactive devices as opposed to the time they interact with other humans. They call the former *technomorphic* (in contrast with anthropomoric) concluding that in order to understand natural behaviour we need to ensure that it is referenced to the space users engage in, where both the machine and user can see each other (Harper & Mentis, 2013). Here, the role of implicit or explicit non-traditional interactions and the consequent frustrations that they could create in the mixed environments of human-system and human-human interactions becomes highlighted. Currently, there is a limited understanding of human behaviour in implicit and explicit interactions with gesture and voice, especially in the context of VM collaborations.

As a result of the potential advantages in this area, our work aims to provide knowledge about implicit and explicit interaction, using gesture and voice, in VM collaboration environments with shared content.

METHOD

We designed a testing environment and interactive system called CONIS (Collaboration Orientated Natural Interfacing System). In this section we overview the research method through the setup of CONIS, the

experiment structure and design, employed task scenario, interaction control, participant recruitment and data collection, and analysis.

CONIS Setup

The experiment took place in a laboratory consisting of three rooms; two for participants engaged in VM collaboration and one for the research controller. Due to the complexity of the interaction methods, particularly the difficult nature of modelling implicit motions and their intrinsically sophisticated parameters we choose to conduct the experiment through a WOZ (Wizard of Oz) approach (see Kelley, 1984; Weiss et al, 2009), which is particularly suitable for understanding new forms of interaction techniques (e.g. Salber, 1993). The WOZ approach has also been used in other studies of implicit interaction (Ju, 2008).

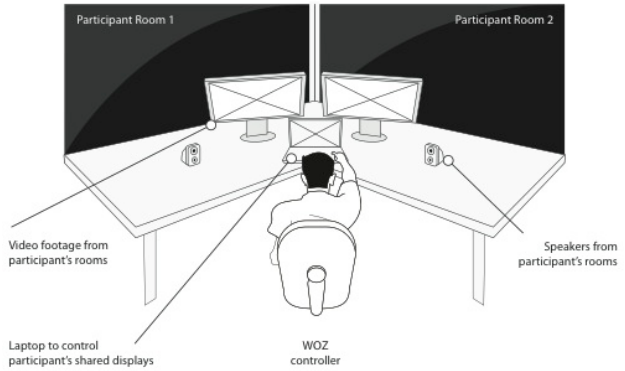

Figure 1. CONIS Setup for WOZ

To maintain consistency across data samples, we elected to have a single 'Wizard' who was highly familiar with the implicit and explicit interaction parameters of this study. The Wizard would determine user interactions and manually control the content on the participant's interactive displays. This was conducted by mirroring the participant's displays with a laptop operated by the Wizard in the control room. Both participants' implicit and explicit interactions would be observed by the Wizard through one-sided glass, and over a video monitor with speakers in the control room. The video monitor and speakers were connected to cameras and microphones in the each of the participant rooms.

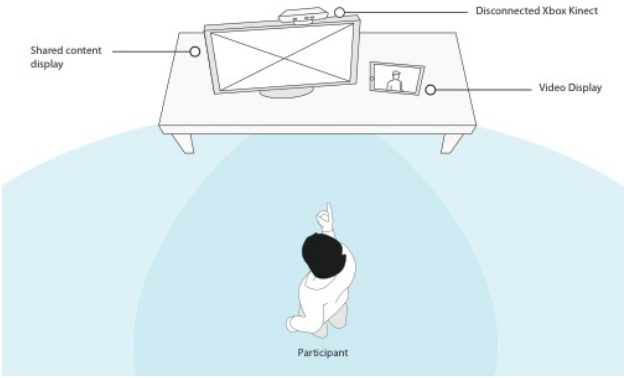

Figure 2. Individual participant's system setup and environment

Each participant's room consisted of a shared content display, a Skype video feed to their partner's room (on a tablet oriented to face the participant), and a disconnected Xbox Kinect which was intended to give an impression of gestural and voice capability (see Figure 2). The shared content display was a television that housed an interactive user interface (ostensibly) controlled through speech and gesture. If a user interaction changed the display on their screen, it would also change their partners' display; they would always be viewing the same collaborative content. The video feed was presented on a tablet and set at a distance to replicate the near full body display view that an Xbox Kinect would normally capture. The purpose of setup was to replicate the requirements of a full body camera that can detect and capture interaction through broad body gestures.

The disconnected Xbox Kinect was placed in the room to not only convince the participants that it was the gesture and speech input mechanism, it also simulated an environment where the user was present with the technology that would be required for this system to function outside of a WOZ method. No participants suspected this low-level deception during the study, allowing us to conclude that the interactions we identified were technomorphic behaviours (rather than anthromorphic, see Harper & Mentis, 2013) and evaluate the impact of implicit or explicit interaction on collaboration in a way relevant for the design of 'real' implicit systems.

Experiment Structure & Design

The structure for the experiment involved three sets of interview questions and two separate tasks: An 'implicit interaction only' task and an 'explicit interaction only' task (described under 'task scenario'). The decision to separate the implicit and explicit tasks (rather than having implicit interaction complement explicit interaction) was made to explore the impact of both interaction types on VM Collaboration in greater depth, respectively identifying each interaction type's strengths and weaknesses.

The tasks were designed to require "high common ground and loosely coupled work" (Olson & Olson, 2000). Hence both tasks were constructed so that they could only be achieved by two participants working on a shared goal. To ensure loose coupling as a collaborative construct the task was a simple exercise, demanding minimal and infrequent interactions and achievable without an ongoing reference to the shared content. To ensure high common ground, we required participants to know each other prior to participating in the research and we chose questions that elicited a common goal; negotiating a selection of activities to undertake together. Moreover we chose content that would contain topics familiar to all participants.

Task Scenario

The scenario of our experiment was the same for both the implicit and explicit tasks. The collaborative pairs were given a hypothetical set of adventure park passes and provided information about four parks (figure 3). In the first task participants were asked to negotiate the parks that they'd like to attend for a full day and a half day holiday. In the second task they were asked to choose

from three to five activities from each park that they would like to do for varying durations. Both tasks required the participants to come to a collaborative answer with their combined preferences after a four-minute period. The adventure parks and activities within the implicit and explicit tasks changed theme from winter to summer once the first task had been completed and the second task was underway, ensuring the second task required further negotiation.

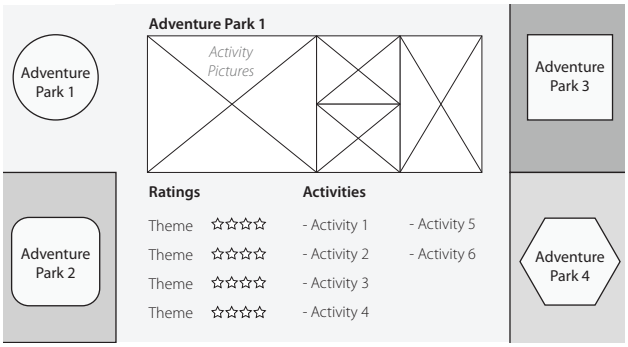

Figure 3. Wireframe for the shared content interface used in both implicit and explicit tasks.

The scenario of an adventure park and its activities was selected because it made use of a diverse range of general entertainment events such as roller coasters, skiing, a zoo, and comedy shows. This was done to evoke a high degree of personal interest for participants and to create a collaborative environment. These characteristics are important in effective VM collaboration (Anderson et al., 1996).

Interaction Parameters

Participants were informed about the interaction parameters before each task. For the implicit task they were told they would not have direct control over the content, and that the system would "implicitly" follow the context of the conversation and gestures directed toward their partner. For the explicit task they were told that the content would only change upon direct interaction through speech and body gesture at the shared content display, however they were not specifically told what type of gestures or verbal commands to use in order to interact with CONIS. This was a deliberate decision to align with Poddar et al. (1998); exploring natural user habits without imposing a predefined interaction method on the user. This type of experiment design is also highly suitable for WOZ studies (e.g. Höysniemi et al., 2004).

The researcher manually operating the content would act upon the following interactions, in accordance with previous definitions and examples of implicit/explicit gesture and voice interactions. The implicit parameters included verbal cues and hand gestures that referenced or represented the discussion around any form of the content presented throughout the task. These interactions involved voice, tone and indirect expressive gestures. The explicit parameters included verbal commands stating the park name and direct pointing, swiping or any form of directional swing at the park icon on the display. The majority of the interactions were unambiguous and worked well with the WOZ set up.

An Example of the Parameters

As an example of an implicit interaction, if User B was to interrupt their partner (User A) and say to "hold on, what about the roller coasters!", whilst gesturing their excitement towards User A, the system controller (i.e. the Wizard) would recognise User B's excitement as well as the context of conversation and change the shared display from the zoo to represent the referenced content containing a rollercoaster.

However, in the explicit task, if User B interrupted and said "hold on, what about the roller coasters!", whilst gesturing their excitement towards User A, the content would not change. For the content to change in the explicit task User B's gestures or voice would have to be "*primarily aimed"* (adhering to Schmidt's (2000) definition) at CONIS. This was determined by evidence of participants directly focusing on the shared display and signalling a direct command to change the content. To make this direct command, User B might say 'Let's look at the rollercoaster park' whilst pointing to the corner of the screen in which the parks icon resides.

In accordance with evaluating interaction types independently, explicit interactions were ignored by the Wizard during the implicit task, and vice versa. This meant that if User B was attempting to control the shared content explicitly in the implicit task by (for example) yelling at the screen to change ("Zoo, ZOO!, ZOO!!, ZOO!!!!") or forcibly pointing to the screen whilst disconnected from their partner, the system would not change. This is consistent with earlier work on implicit only systems where the user is not in control (Dietz et al, 2004). Content change in this example would only occur at the next implicit cue not directed at the system.

Participant Recruitment

For this experiment we used 10 participants, aged between 24 and 35. All participant spoke English fluently and all had some form of tertiary education. All participants were at least aware of both voice and gestural control systems. 70% of participants had used a voice control system such as Apple's Siri however no participant had used a gesture control system such as Xbox Kinect (and thus did not have preconceived notions regarding what it could or could not do).

Data Gathering and Analysis

To gather data we used observations and interview questions. Observations were conducted during the experiment. The interview questions took place over three stages; before the tasks, conducting general demographic and system familiarity questions and after both the implicit and explicit task.

Our data analysis took place through several key stages. To initiate the analysis process, the recordings of participant interviews and the observations of participants in their tasks were transcribed in their entirety. Open coding was then applied to generate the theoretical properties of data categories.

We used a member checking (Miles & Huberman, 1994) to ensure the collected interpretations of the data was

correct. This involved following up participants a few days after their session, discussing the researchers' summary of their experience to ensure the direction of the collected data was accurate.

After themes were established, results were taken to round table discussion of academic peers to ascertain clarity and coherence (Miles & Huberman, 1994). All peers had observed the experiment during the data collection phase. A form of axial coding was then used to group code according to conceptual categories that identified commonalities (Fielding & Lee, as cited by Miles and Huberman 1994). In addition to this the video footage of the participants tasks were reviewed in context to the categorical codes providing further detail into the events that took place and suitable quotations were extracted.

RESULTS

We observed a range of findings that demonstrate how implicit and explicit interactions significantly change the dynamics of VM collaboration.

Across both implicit and explicit tasks our clearest observation was that participants elected to disengage with their collaborative partner visually, instead focusing their attention on the interactive display and the shared content. This occurred even though the task was designed to be accomplishable without a constant reference to the shared content.

Participants became reliant on the shared content display to negotiate their outcomes, wheras the Skype video feed was used primarily for its audio link. Typical responses in our post-session interviews included: *"I wasn't looking [at my partner] when I was speaking; when I was speaking I was always concentrated on the content"*. Some believed they *"could have done it on the phone"* and others would have *"a quick look"* at their partner but stated it "*wasn't needed to get the task done*". This is in contrast to Anderson et al's (1996) investigations that found a video screen is of benefit in tasks with negotiation, bargaining or conflict. We found the addition of a shared interface in this environment meant users would manage their conversation through viewing and changing the interactive content, reducing the value of the video screen to a minimal benefit. They would allow the presented information to lead communication cues, and in return it would keep the conversation going onto the next topic; "*Skype became basically irrelevant, because you're talking to the point when you were directing [the shared content]"*

Upon questioning the participants about their satisfaction and experience of the task all 10 participants questioned the necessity of the video in the exercise. However it should be noted the video channel was not completely worthless. Through our observations we saw on occasion people would either focus on or go up to the video display in moments of high emotional discussion, to better understand what their partner was trying to say. In the post-session interviews a few participants acknowledged the indirect benefit of the video channel, a benefit that we describe as maintaining connectedness; it allowed participants to feel "*closer*" to their partner, as Jeremy stated: "*I could feel* [my partners] *presence in the room*".

Implicit Task

Collaboration over the shared content in the implicit-only task was marred by frustrations and difficulties. As a result of not being engaged with their partners via the video-link, there was a lack of implicit body action for the Wizard to interpret. The lack of direct control over the system caused significant frustration, and the tangential nature of collaborative discussions saw the shared content change frequently, consequently diminishing the conversation structure. Finally, issues around surveillance, comfort and 'naturalness' further reduced the quality of collaboration in the implicit-only task.

Implicit Body Movement and Gestures

The observed body language in the implicit tasks saw participants acting somewhat reserved and slightly ridged. Typically they would have their arms crossed, stand in a thinking pose with one hand on their chin or have their hands in their pockets. They would occasionally use their hands to count out activities or point at things on the shared content display as a reference to their conversation, this happened despite the inability for their partners to see directly what they were pointing at (without significant interpretation). Participants also tended to display limited movement around the room, generally they were fidgeting, rocking back and forth on the spot, leaning, standing still or sitting down. There was undoubtedly some unease in the lab-based environment. The majority of spatial movement came when participants would occasionally step towards the shared content to take a closer look.

The users' focus was on the shared content rather than the video display of their partner, and because of this, participants were gesturing directly to the shared content instead of their partner on the video display. The result of these actions meant there was a limited amount of implicit gestures for the Wizard (acting as the system) to extract; implicit control tended to come from the context of conversation and indirect vocal commands as opposed to indirect gestures.

Conversation Structure and Collaboration

The conversation between participants would generally flow from one topic to another, sometimes meaning topics were often left behind or skipped, affecting the structure (and success) of the collaboration. These tangential conversational cues caused the system to implicitly change the shared content to match the context of conversation on a regular basis, sometimes exacerbating the extent to which some facets of the discussion were forgotten. Further, when it became noticeable to participants that they were not effectively negotiating, the flow of conversation and constant change in topic became a clear disruption.

Frustration and Control

In one example, Tom, a participant, raised a new idea as sidenote to the current discussion. In doing so the content display changed from the current topic they were still

discussing to his new topic. In response to this event James (Tom's partner) raised his voice and said "*Tom* [in frustration], *we've got to focus here, we've got to apply a framework … let's stop jumping around and determine where we want to go*". The result was frustrating for both participants as James was obviously annoyed at Tom for changing the topic of conversation and the content on the shared display, similarly Tom was frustrated as he hadn't intended to change the content, he simply wanted to briefly discuss something else, a topic that the conversation flow evoked.

As each task progressed some participants began to change their vocal approach, directing their words at the interface. Participants were attempting to take direct control over the shared content even though they knew the interactive display would only change to represent the context of the conversation via indirect body gestures and verbal cues, not through direct actions at the shared content. They would often try to make the system change with explicit pointing and direct vocal commands.

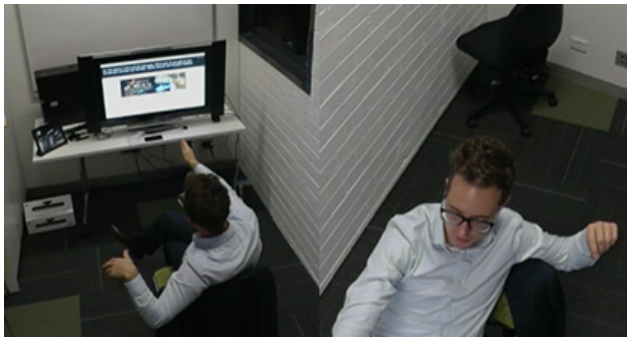

Figure 4. A participant as his interaction began to turn explicit through explicit pointing within the implicit task.

In another example, Adam would occasionally disconnect the conversation with his partner by focusing on the system, his voice changed in both pitch and volume in an attempt to help the system understand his interactive intent, he would raise his voice and say "*adventure island... ADVENTURE ISLAND*". This action halted the collaborative conversation until the slide changed based on some form of implicit cue, such as reengaging with their partner on another topic or the participant moving on to discuss an activity associated with adventure island.

The Naturalness of Implicit Interaction

Participants initially reported an improved conversational flow, however, upon further questioning we found that collaboration was difficult in an implicit only task. There was an 'unnatural' lack of control and a feeling of discomfort in the interactive environment.

Participants initially responded well to the implicit interaction methods, they found it a beneficial because it didn't get in the way of their conversation and it simulated a collaborative environment that they considered natural. Tom explained his feelings in the implicit post-session.

"I felt it was natural that we could control at the same time so it was a collaborative effort, you're talking about something and you both have an effect on it. Because usually, you've got control of your space and he's got control of his space. So I thought… it was a natural way of collaborating if you were with someone and we'd both say something, it would have an effect on the content."

Tom was comparing his experience to that of a video mediated environment without shared interactive content. He was illustrating that the implicit interaction was not changing the dynamic of previous communication experiences he'd had over video channels. Instead he felt it enhanced the communication because it naturally displayed content to show the topic of discussion.

These views were consistent across most participants, Jane described the conversation as "*natural* [because] *it flowed*", Sarah also commented on the benefit of the relevant content coming up as they conversed in their normal conversational manner, and Rod felt it was "*pretty responsive; whatever* [he] *said, it flashed up on the screen*". Although initial communication over the system worked well, the collaboration was not optimal. Jane stated that she felt it "*took* [them] *a long time to come to the answer*", Tom said the collaborative aspect of the task was a negative experience as he wasn't able to easily effect who had a greater "*share of the power*" without raising his voice or cutting out his partner, he was referring to the displayed content that their collaboration became dependent on. In response James thought that if you had raise your voice to take control of the content and the topic it would be bad user experience as "*people would change the way they communicate with each other*", stating "*he'd just call* [his partner] *on the phone where he couldn't cut him off*".

As stated in the observations, participants began to try and control the system explicitly, by doing this it became less responsive to their interactions and in turn "*a bit unnatural*" for users to collaborate over. Adam explained his feelings during one of these occasions; *"I felt like I was waiting for it, it wasn't always seamless…. so sometimes I felt like I just wanted to go up and touch it"*.

Surveillance and Comfort

Another topic raised was about the level of comfort in the environment as multiple users described a certain level of uneasiness. Jane felt as though simply standing there was a bit unnatural and being watched via the video display made her feel as though she had to stand up straight. Steve also discussed his uneasiness in the environment. He said that he would have liked to move things around such as the TV, the iPad and the Kinect, saying "*I don't know whether that's a function perspective or an ownership perspective, like a musician that moves his microphone stand up and down before performing*".

Steve's feelings were matched with the type of body language identified in our observations; he changed his position, leaned on the wall and realised he was leaning so he stood up straight again. He seemed somewhat unsure of his environment and uneasy about the lack of control due to not knowing how the system functioned.

Explicit Task

Overall, we found that conversation structure in the explicit-only task was much more efficient, leading to better negotiations and successful collaborations. As a

result of an increased interest in their partner, participants were more engaged in the task, and an increased level of content control improved participant's comfort whilst reducing the feelings of anxiousness reported in the implicit-only task.

Conversation Structure and Collaboration

It was apparent that the freedom of direct content control allowed participants to come to an agreement on discussion points, and it improved the ability for conversation turn-taking. We quickly saw the shared content act as a mediator for each person to take turns conveying their thoughts. In several cases the turn-taking was not just a process that collaborators fell into, it was a process which explicit control evoked. An individual participant would acknowledge they'd taken control over the system and as a result they gained ownership of the conversation. If they had been controlling the content for an extended duration they would tend to unambiguously pass control of the system to their partner, this pass of control often resulted in the receiver taking the communicative lead for the next topic of conversation.

An example was when James had taken control in a conversation early, and as he continued to discuss his points of interest he seemingly became aware that he had interacted with the shared content more than Tom had. James simply took a step back and said to Tom "*I'm happy to do whatever you like, you just do the pointing*". The result of this was an immediate switch of conversation control that saw Tom's interests better represented in the negotiation.

This was also the case if the partner didn't recognise the other person wanted to contribute. In the case of Steve and Jeremy there had been an ongoing conversation led by Jeremy's preferences. Jeremy's conversation had taken him onto another topic. Steve cut Jeremy short by changing the content and taking the conversation back to one of the first topics Jeremy had mentioned, as the content changed Jeremy immediately stopped his conversation and Steve took over. This example was a strong juxtaposition to Tom and James's frustrating experience in the implicit task where Tom's sidenote changed the content and Tom had no method of changing the content back without socially overriding Tom in conversation, which he did by verbally breaking out and disrupting the existing dynamic of the collaboration.

It was also evident that through explicit control participants had the opportunity to take a more balanced approach to the structure of the collaboration. Adam and Jane used it to facilitate their conversation, taking turns to control and observe the shared content. The collaborative structure they employed involved one person changing the shared content, they'd then both observe the screen, the lead content controller would talk and the partner would reply. After a small conversation the participant in control would change to another adventure park option and use the newly presented information to refresh the conversation topic. This conversational structure was a consistent pattern in the explicit collaborative exercise; the method appeared to be a lot less ad-hoc than their corresponding task with implicit control.

Explicit Body Movement and Gestures

Additionally we observed an increased level of bodily movement and gestures. James' interaction in the explicit task was a good example of this finding. He would try new gestures to change the shared content; from single-handed pointing to double handed pointing, behind the back and even trying to initiate system change by head-butting in the direction of the content he wanted to appear. He was very active and heavily involved with the environment. This was a similar case for Roger who increasingly exaggerated his movements as he tried to explore interaction methods whilst collaborating. He would overtly gesture mountain signs to change to ski related content or jump in the air to represent an icon of one of the topics he wanted to view. They would eventually determine that the majority of these interaction methods would remain unresponsive and switch to using a direct point or directional swipe with a voice command, however their level of engagement and bodily action within the room did not diminish.

Figure 5. A participant becoming heavily engaged during the explicit task.

Lastly, we did notice that the explicit task created an environment where some people would look at their partner in the video channel a little bit more regularly than the implicit task, although the difference was minimal.

Increased Engagement

The findings from interviews indicated users felt more engaged and comfortable in the explicit tasks. The general feeling from the participants was that the method of conversation structure for the explicit task was more successful than that of the implicit task. Furthermore throughout the interviews we had an unexpected finding that explicit control decreased feelings of anxiety for some participants.

Several participants commented on an increased level of engagement with both the system and with their partners. The system engagement elevated as users felt the desire or requirement to increase bodily action and spatial movement around the room. James described his experience in the post-session.

"We were up, running around the room gesturing… it changed the way we conducted ourselves… I was more engaged and interested in the exercise because I was doing things".

Participant engagement increased not only because they were more active and enthusiastic but also as there was a greater interest in what their partners were doing, Jessica identified her feelings of a greater level of engagement with her partner Roger. She described herself as becoming somewhat more interested in the video channel; as Roger gestured she wanted to watch. Similarly James found it was funny to observe his partner, however, he raised an important question that perhaps some of the desire to watch each other could come from the novelty of a new experience, relating it to games such as Singstar which are enjoyable to spectate (ala work by Downs et al, 2014; 2014)

When participants were asked how much they felt they engaged with their partner, most estimated they looked at their partner 10-20% more than they did in the implicit task. However whilst our observation identifies that some people did look at their partner a little bit more, it was not to the degree in which they felt they did in the interviews, suggesting that greater control over the shared content increased the perception of stronger engagement with partners.

Comfort and Control

Due to the heightened engagement between users and an increased freedom to move around without implicit interactions changing the content, the level of comfort increased with explicit interactions. Jane explained her feelings in a post-session question.

"I was more comfortable around my control of the material", "I'd look at [my partner] *more in a conversation way because* [the shared content on the screen] *wasn't going to change, it seemed to be a lot more natural".*

Adam found it comfortable because he "*could stand there in different poses, for example a thinking pose and do* [his] *normal kind of actions without the screen changing*", furthermore Steve related this experience to feeling "*normal*" through explicit control; "*I want to do things with my hands, it's natural… like being in a bar you want a drink in your hands*". Furthermore Steve explained the impact of explicit control in the style of his conversation.

"It almost facilitated an agenda or a structured meeting, I don't know if that's natural but it helped the experience".

"It's kind of like saying I want to look at 'that,' I want to look at 'this'. You're kind of saying that I am now owning the next 10-15 seconds of this conversation, so when my partner would say they wanted to look at Park 2, I knew he had something to say about Park 2".

What Steve was alluding to was that direct control inherently allowed users to come to an agreement on points to be discussed. James' response furthered this explanation as he explained explicit control made both his partner and himself more reliant on collaboration to decide who should control the system and in turn the conversation. This type of collaborative control was indicative of technomorphic behaviour outlined by Harper & Mentis (2013), users often adjusted their body language and method of communication to meet the expectations of the interactive devices present in the room.

Lastly some participants described a reduced level of anxiousness in their explicit experience. During the post-session interviews of the explicit task Steve came to somewhat of a realisation about the described unease in his previous implicit task; "*I felt this time I didn't want to move the objects* [in the room] *as much because I felt that the sense of control had been provided to me on the screen*", subsequently the sense of control he felt to move the room around "*dissipated*". "*I felt like I wanted to do something in the* [implicit] *task and this task I got to do something*".

DISCUSSION

Technologies with the capacity to make use of implicit interactions are increasingly becoming available. The recent Samsung Galaxy S5, for example, uses the front-facing camera to determine if the user is looking at the screen, consequently not turning off to save power if they are looking at it. As our devices and systems become increasingly context-aware and proactive, further research is needed to understand the impact of implicit interactions on user experience. In this study we evaluated the impact of implicit interaction on a relatively simple video-mediated collaborative task, contrasting user experience with an explicit-only task to reveal the constraints and affordances of the different interaction paradigms.

Overall, we found when using a shared-screen collaborative task, participants did not seem to engage with the video-link to their partner. While some participants noted that the video-link enhanced the sense of presence, engagement primarily revolved around the interactive display and the shared content. This is in contrast to the identified benefits of video communication in prior studies, where video usage was beneficial for tasks involving negotiation, bargaining or conflict (Anderson et al. 1996). This is also despite our task being purposefully loosely coupled and with high common ground (Olson & Olson, 2000) to ensure users did not have to be dependent on the shared content to complete the task.

The collaborative dynamic of the task meant that implicit interactions often did not support the intent of the user, as conflict emerged due to the complex negotiation of human-human and human-system interactions. As has been identified, users typically adjust their behaviour to become technomorphic in the presence of interactive hardware, as opposed to behaving in an anthromorphic manner with their video-mediated collaborator (Harper & Mentis 2013). The introduction of an implicit system, presented to supplement the interaction between two collaborators, was no different; user's behaviour was altered. Because the participants adopted behaviours that did not match the input methods for implicit interaction (indirect gestural cues), the system became less able to reflect their intent and support the collaborative task.

It is likely that behaviours would adjust as users become more familiar with implicit interactions. However, as the

majority of user interactions were based directly around the interactive content display as opposed to the video channel, human-to-human interactions were evidently diminished. Collaborative tasks that require or encourage VM interaction and negotiation (such as tasks requiring high emotional awareness) may be more suitable to implicit tasks than our findings have been able to identify.

Despite the experiment being purposefully designed to be simple, and thus not requiring the users rely on the shared screen to complete the task, control over the display was an issue for the success and positive user experience in the implicit-only task. We speculate that this was not specifically due to a lack of control, but rather the way in which unintended or unexpected changes to the shared content interrupted the flow and structure of collaborative discussion.

Control is one of the more complex and undefined concepts at the heart of implicit interactions, as it is still not clear within the literature if implicit systems inherently lie within or external to user control. We observed collaboration improving during the explicit task when users were able to delegate control, or share control through turn-taking, and consequently structure their negotiation. Prior work has frequently highlighted issues around control in implicit only systems (Hinckley, 2000; Dietz, 2004), which the introduction of explicit overrides (e.g., Ju et al., 2008) was able to significantly reduce. We therefore echo Schmidt (2000) in concluding that it is unlikely implicit interaction will (or should) be used exclusively. This conclusion also supports Ju and Leifer's (2008) framework that implicit interactions should be based on attentional foreground and background, rather than independently proactive systems. Control sharing (such as turn taking) is a natural collaborative interaction for users, i.e., control sharing actions are *"the actions they perform with technology 'apposite', 'appropriate', or 'fitting' to the particular social setting*" (O'Hara et al., 2013). We therefore note future work could explore the ability to explicitly delegate control with implicit systems, or provide turn-taking (perhaps through implicit interactions such as proxemics, ala Ju et al., 2008) in order to supplement this constraint.

In contrasting the two tasks, we found that explicit tasks saw a reduction in anxiety and improved collaboration. This is perhaps a combination of the affordances of the interaction paradigm, but also potentially due to all our users' lack of familiarity with implicit systems. Familiarity with explicit interaction, and the clarity around what was considered an 'interaction' and what was not, decreased feelings about control and their ownership (and comfort within) the environment. We noted in our review of the complex concepts involved in defining implicit interaction questions regarding surveillance, intent and consent, which these findings suggest are valid concerns that users have when interacting with an implicit system.

CONCLUSION

In this research we have conducted an investigation into the impact of implicit and explicit use for VM collaboration with shared content. We have shown that explicit and implicit interactions with shared content are a natural fit into a VM collaborative environment and through our findings we have supported existing knowledge that distance collaboration can be successful when the task is loosely coupled and maintains high common grounds.

A detailed exploration of the strengths and weaknesses of implicit and explicit interaction has been discussed across a number of themes integral to definitions of implicit interaction; control, naturalness, intent and surveillance. Through a discussion of these themes we have found that implicit interactions are generally not preferable primary interactions for VM collaborative environments. We have offered an explanation as to why we believe this to be the case. Furthermore we have defined the strengths that explicit control can contribute to not only systems interaction but also towards enhancing the user experience of collaborating individuals. Additionally we have found that explicit interactions can reduce unease and anxiety within similar interactive environments.

Extending upon these finding there is significant opportunity for further studies to work both in parallel with this research as well as build upon it. Firstly, there is an opportunity to investigate a similar study across different cultures, demographics and experience with implicit systems to exploring the variance of the results. Secondly, there is room to further examine the impact of other implicit and explicit controls that come under the natural interaction umbrella such as eye gaze, facial recognition and spatial navigation. Finally, by using this research there is not only an opportunity to enhance combined natural interaction and VM collaborative systems, there is a potential to use these findings as groundwork in defining and developing a framework for both explicit and implicit controls appropriate for this type of combined technology.

ACKNOWLEDGMENTS

We would like to thank Bernard de la Coeur, Kenton O'Hara, Jeni Paay, Jesper Kjeldskov and all participants for their support in this study.

REFERENCES

Anderson, A.H., Newlands, A., Mullin, J., Marie Fleming, A., Doherty-Sneddon, G. and Van der Velden, J. Impact of video-mediated communication on simulated service encounters. Interacting with Computers 8, 2 (1996), 193-206.

Atterer, R., Wnuk, M. and Schmidt, A. Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In Proc. WWW'06, ACM Press (2006), 203-212.

Blažica, B., Vladusic, D. and Mladenic, D. A personal perspective on photowork: implicit human-computer interaction for photo collection management. Personal and Ubiquitous Computing 17, 8 (2013), 1787-1795.

Buxton, W. Integrating the periphery and context: A new taxonomy of telematics. Proc. Graphics Interface '95 (1995), 239-246.

Buxton, W. Mediaspace-meaningspace-meetingspace. In S. Harrison (ed.) Media space: 20+ Years of Mediated Life (2009), Springer: London, 217-231.

Dietz, P., Raskar, R., Booth, S., Van Baar, J., Wittenburg, K. and Knep, B. Multi-projectors and implicit interaction in persuasive public displays. In Proc. AVI'04, ACM Press (2004), 209-217.

Dourish, P., Adler, A., Belotti, V. and Henderson, A. Your place or mine? Learning from long-term use of audio-video communication. Computer Supported Collaborative Work 5, 1 (1996), 33-62.

Downs, J., Vetere, F., Howard, S., and Loughnan, S. Measuring audience experience in social videogaming. In Proc. OzCHI 2013, ACM Press (2013), 217-220.

Downs, J. Vetere, F., Howard, S., Loughnan, S. and Smith, W. Audience experience in social videogaming: effects of turn expectation and game physicality. In Proc. CHI 2014, ACM Press (2014), 3473-3482.

George, L. and Lécuyer, A. An overview of research on 'passive' brain-computer interfaces for implicit human-computer interaction. In Proc. ICABB'10 (2010).

Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R. and Wang, M. 2011. Proxemic interactions: The new ubicomp?. Interactions 18, 1 (2011), 42–50.

Harper, R. and Mentis, H.M. The mocking gaze: the social organization of kinect use. In Proc. CSCW 2013, ACM Press (2013), 167-180.

Hauber, J., Regenbrecht, H., Billinghurst, M. and Cockburn, A. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In Proc. CSCW 2006, ACM Press (2006), 413-422.

Hinckley, K., Pierce, J., Sinclair, M. and Horvitz, E. Sensing techniques for mobile interaction. In Proc. UIST'00, ACM Press (2000), 91-100.

Höysniemi, J., Hämäläinen, P. and Turkki, L. Wizard of prototyping of computer vision based action games for children. In Proc. IDC'04, ACM Press (2004), 27-34.

Ju, W. The Design of Implicit Interactions. PhD Dissertation, Stanford University (2008).

Ju, W., Lee, B.A., Klemmer, S.R. Range: Exploring implicit interaction through electronic whiteboard design. In Proc. CSCW'08, ACM Press (2008), 17-26.

Ju, W. and Leifer, J. The design of implicit interactions: Making interactive systems less obnoxious. Design Issues 24, 3 (2008), 72-84.

Kelley, J. F. An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Office Information Systems 2, 1 (1984), 26-41.

Miles, M. B. and Huberman, A. M. Qualitative data analysis: an expanded sourcebook. SAGE Publications, Thousand Oaks, CA. (1994).

O'Hara, K., Harper, R., Helena, M., Abigal, S., Taylor, A. On the naturalness of touchless: Putting the "interaction" back into NUI. ACM Transactions on Computer-Human Interaction 20, 1 (2013), 1-25.

O'Hara K., Kjeldskov, J., Paay J. Blended interaction spaces for distributed team collaboration. ACM Transactions on Computer-Human Interaction 18, 1 (2011), 1-28.

Olson, G.M. and Olson, J.S. Distance matters. Human Computer Interaction 15, 2 (2000). 139-178.

Poddar, I., Sethi, Y., Ozyildiz, E. and Sharma, R. Toward natural gesture/speech HCI: A case study of weather narration. In Proc. 2nd Workshop on Perceptual User Interface (1998), 1-6.

Salber, D. and Coutaz, J. Applying the wizard of oz technique to the study of multimodal systems. In Proc. CHI'93, ACM Press (1993), 219-230.

Schmidt, A. Implicit human-computer interaction through context. Personal Technologies 4, 2-3 (2000), 191-199.

Tse, E., Shen, C., Greenberg, S. and Forlines, C. Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In Proc. AVI'06, ACM Press (2006), 336-343.

Venoitt, E. S., Olson, J., Olson, G. M. and Fu, X. Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In Proc. CHI 1999, ACM Press (1999), 302-399.

Weakliam, J., Bertolotto, M. and Wilson, D. Implicit interaction profiling for recommending spatial content. In Proc. GIS'05, ACM Press (2005), 285-294.

Weiss, A., Bernhaupt, R., Schwaiger, D., Altmaninger, M., Buchner, R. and Tscheligi, M. User experience evaluation with a wizard of oz approach: Technical and methodological considerations. In Proc. Humanoid Robots'09, IEEE (2009), 303-308.

Wilson, A. and Oliver, N. Multimodal sensing for explicit and implicit interaction. In Proc. HCII'05, ACM Press (2005).