Telethrone: a situated display using retro-reflection based multi-view toward remote collaboration in small dynamic groups

John O'Hare

UNIVERSITY OF SALFORD

SCHOOL OF COMPUTING, SCIENCE AND ENGINEERING

DECEMBER 13, 2018

This document is toward the degree of Doctor of Philosophy

Contents

List of Figures

Acknowledgement

Claire O'Hare, David Roberts, Rob Bendall, Allen Fairchild, Bex Tennenhaus, Toby Duckworth, and Margaret O'Hare have provided enormous patience, support and technical wisdom.

0.1 Abstract

This research identifies a gap in the tele-communication technology. Several novel technology demonstrators are tested experimentally throughout the research. The presented final system allows a remote participant in a conversation to unambiguously address individual members of a group of 5 people using non-verbal cues. The capability to link less formal groups through technology is the primary contribution.

Technology-mediated communication is first reviewed, with attention to different supported styles of meetings. A gap is identified for small informal groups.

Small dynamic groups which are convened on demand for the solution of specific problems may be called "ad-hoc". In these meetings it is possible to 'pull up a chair'. This is poorly supported by current tele-communication tools, that is, it is difficult for one or more members to join such a meeting from a remote location. It is also difficult for physically located parties to reorient themselves in the meeting as goals evolve.

As the major contribution toward addressing this the 'Telethrone' is introduced. Telethrone projects a remote user onto a chair, bringing them into your space. The chair seems to act as a situated display, which can support multi party head gaze, eye gaze, and body torque. Each observer knows where the projected user is looking. It is simpler to implement and cheaper than current comparable systems.

The underpinning approach is technology and systems development, with regard to HCI and psychology throughout. Prototypes, refinements, and novel engineered systems are presented. Two experiments to test these systems are peer-reviewed, and further design & experimentation undertaken based on the positive results. The final paper is pending.

An initial version of the new technology approach combined retro-reflective material with aligned pairs of cameras, and projectors, connected by IP video. A counterbalanced repeated measures experiment to analyse gaze interactions was undertaken. Results suggest that the remote user is not excluded from triadic poker game-play. Analysis of the multi-view aspect of the system was inconclusive as to whether it shows advantage over a set-up which does not support multi-view.

User impressions from the questionnaires suggest that the current implementation still gives the impression of being a display despite its situated nature, although participants did feel the remote user was in the space with them.

A refinement of the system using models generated by visual hull reconstruction can better connect eye gaze. An exploration is made of it's ability to allow chairs to be moved around the meeting, and what this might enable for the participants of the meeting. The ability to move furniture was earlier identified as an aid to natural interaction, but may also affect highly correlated subgroups in an ad-hoc meeting. This is unsupported by current technologies. Repositioning of several onlooking chairs seems to support 'fault lines'. Performance constraints of the current system are explored.

An experiment tests whether it is possible to judge remote participant eye gaze as the viewer changes location, attempting to address concerns raised by the first experiment in which the physical offsets of the IP cameras lenses from the projected eyes of the remote participants (in both directions), may have influenced perception of attention.

A third experiment shows that five participants viewing a remote recording, presented through the Telethrone, can judge the attention of the remote participant accurately when the viewpoint is correctly rendered for their location in the room. This is compared to a control in which spatial discrimination is impossible. A figure for how many optically seperate retro-reflected segments is obtained through spatial anlysis and testing. It is possible to render the optical maximum of 5 independent viewpoints supporting an 'ideal' meeting of 6 people. The tested system uses one computer at the meeting side of the exchange making it potentially deployable from a small flight case.

The thesis presents and tests the utility of elements toward a system, and finds that remote users are in the conversation, spatially segmented with a view for each onlooker, that eye gaze can be reconnected through the system using 3D video, and that performance supports scalability up to the theoretical maximum for the material and an ideal meeting size.

0.2 Declaration

0.2.1 Contributing research

For nearly a decade, research at The University of Salford built better tools for communication across a distance. Roberts and Steed worked together in supporting gaze in Telepresence [266].
Oliver Grau (BBC R&D) introduced research into visual hull reconstruction [122], which informed the design of the Octave multi-modal research platform with Roberts, Aspin, and I (all University of Salford), which was used in subsequent research toward connecting people better over a distance [270].
Toby Duckworth [82, 80, 83, 81], Carl Moore [210, 211] (Salford PhDs) and I built the Octave's 3D capture and reconstruction systems under the direction of Prof Roberts. I am a named author on two of these papers [269] [273], with the system later being used successfully in the Telethrone research.
The BBC's research and development department did some pilot work, trying to connect informal spaces with Skype. This approach was found inadequate for their problem and informed their research proposal to the University of Salford.
Roberts had proposed connecting eye gaze through furniture, in an unpublished 2009 paper called 'Talking with the Furniture'[267]. Roberts and Graham Thomas (BBC R&D) discussed spatial segmentation using retro-reflective cloth more normally found in TV production, with Thomas suggesting that it be applied to a chair, per Roberts' idea. This idea of projecting onto a chair was tested without retro-reflection in 2010, an echo of the Microsoft Room2Room research which would come much later. This can be seen inset left in Figure 1.

Paul Sermon (University of Salford) suggested that it was important that chairs could be 'pulled up' into the conversation. This was based on his telepresence projection research supporting ad-hoc social engagements[289, 291, 31].

Duckworth and Roberts tested rear projection of multiple video streams onto a semi-transparent film, at the suggestion of Thomas, but this material did not provide spatial segmentation. Duckworth and I assembled a three projector trial onto Chromatte retro-reflective cloth in 2010, simply projecting three different photographs of a head onto a flat surface. Three views extracted from a video which was shot at the time can be seen in Figure 1.

This seemed to work well, with Duckworth suggesting that vertical lenticular lenses should be added to improve the effect. Once again this echoed research which would come much later in Telehuman 2 [121]. Application of a lenticular sheet without a barrier was briefly attempted, but the materials available provided poor results.

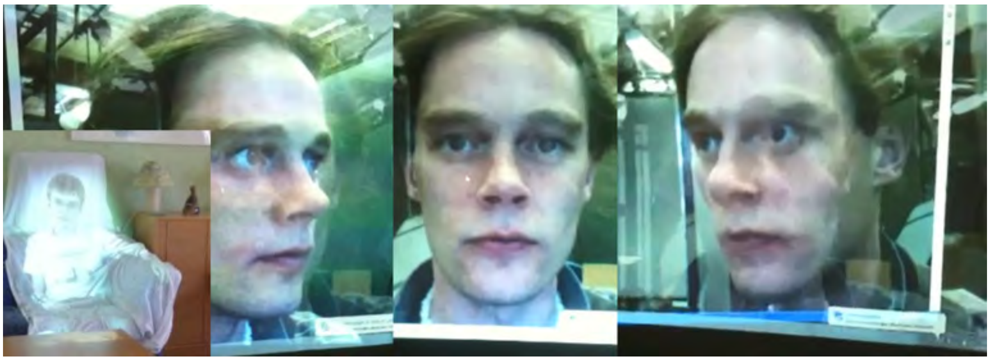


Figure 1: Early discussion and some testing across the group suggested that a chair (inset left) might be combined with retro-reflective material (seen demonstrating three isolated viewpoints of the same person). Rear projection of a micro-lensed material and lenticular sheets were trialled at this time with poor results.

Roberts, Bruce Weir (BBC R&D), and Steve Bowden (University of Salford) performed a pilot experiment; applying the cloth to furniture, first on a sofa, then a version of Telethrone.

In support of the research presented in this thesis John Rae (Roehampton)[256, 255, 254] provided statistical support and insight for the first experiment. R. C. A. Bendall (University of Salford) was instrumental in supporting the first experiment and paper [228].

Since the capture system was built work in the group has continued around representing 3D scanned avatars within immersive collaborative virtual environments, notably in the EU CROSSDRIVE project with contribution from Wolff, Fairchild, Campion, and Roberts [93, 94]. CROSSDRIVE has been adapted by Fairchild to specification from the Telethrone research to provide reconstruction for the Telethrone system. This collaborative effort is detailed in a Appendix 1 and will form a journal paper submission in the future.

Concurrently with our research Pan, Steed, and Steptoe (UCL) furthered situated displays, with some crossover with our group.

0.2.2 Novel Components

The presented research contains novel parts which intersect with the research of others in the group. Some are derivations or developments of ideas presented before the commencement of the research, and are attributed in the Venn diagram Figure 2.
Critically for this researcher the systems presented and tested are themselves technically novel. Any references to support for evolving conversational fault lines, and to subgroups within the ad-hoc meeting - supported by the Telethrone - are specific to the PhD. This ad-hoc and dynamic element is the author's novelty and therefore central to the Venn.

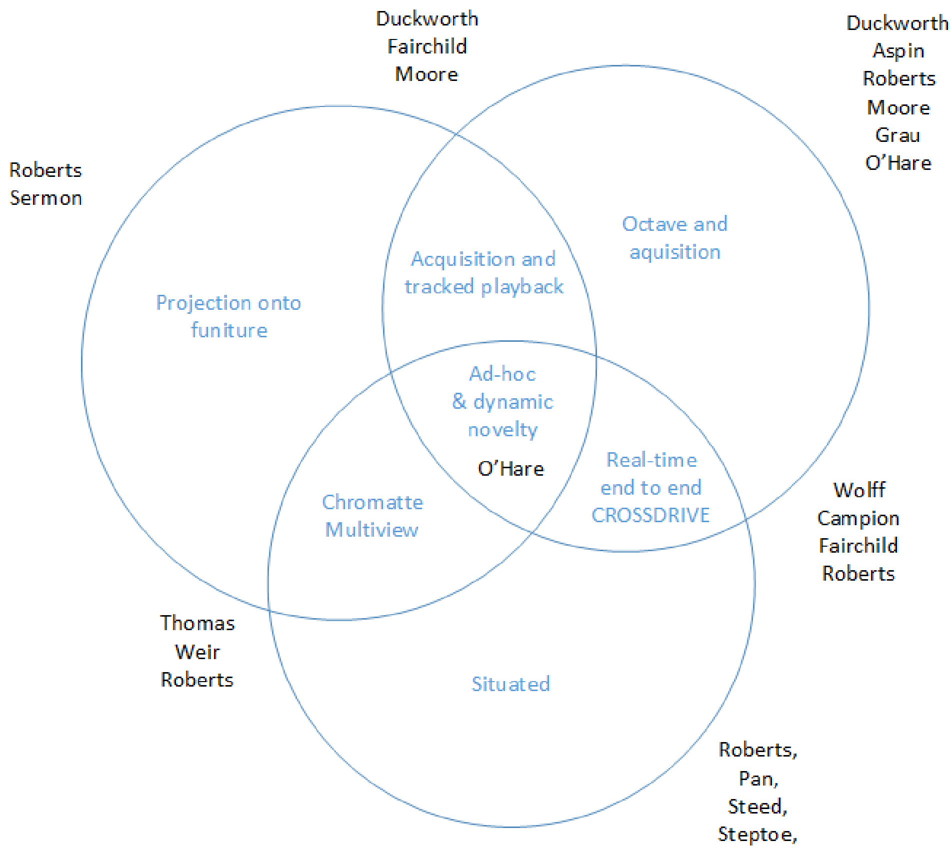


Figure 2: Group and wider contributors to the system. The fields of study are in blue, while the notable researchers are in black. The literature review reveals many more key players from these and other connected fields, but this diagram captures the intersections of colleagues who were directly involved. The author contributed novelty across the Venn to the previous work of others, but personal novelty is the centre of the Venn.

0.3 Abbreviations and Definitions

- Ad-hoc : For the purpose of this document ad-hoc is a meeting convened on demand and at short informal notice. The implication (from the problem brought to us by the BBC) is that it is technical teams solving a problem with a physical separation between elements of the group.
- Affordance : This is Gibson's definition taken as "an action possibility available in the environment to an individual, independent of the individual's

ability to perceive this possibility” [204].

- AR : Augmented Reality : Addition of computer stimuli over and into the normal sensory channels.
- Balun : A balun is an electrical transformer used to connect an unbalanced circuit to a balanced one. In the use cases stated it is always from a balanced video signal to a fibre signal designed to extend range and vice versa.
- CAT6 : Twisted pair copper Ethernet cable.
- Chomatte : Proprietary retro-reflective cloth used in video and TV production.
- Cluster : More than one computer or discrete compute element working on a common task nominally at the same time.
- DVI : Digital Visual Interface is the most common video signal standard in this research.
- Dyadic : Interaction between a pair of individuals.
- Embodiment : A contested word in this research domain but generally here taken to mean the full representation of an individual externally to themselves.
- Faultlines : In the context of this document always used to describe pragmatic functional and temporary junctions between subgroups within a larger group context.
- Frustum : A geometric description of a solid truncated by two parallel planes. In this research the reference is to the ‘projection frustum’ which is the totality of the light path between the solid state image chip and the display surface.
- HD : High definition always taken to be 1920x1080 pixels of video but not necessarily frame rate or bit depth.

- HMD : Head Mounted Display such as Oculus Rift which place visual input on the users head.
- ICVE : Immersive Collaborative Virtual Environment is a specific suite of technologies which seek to connect remote users for collaborative purposes.
- Immersive : Generally taken to be a spectrum of engagement for a user of a computer simulation where the posited 'full immersion' is indistinguishable from real life. A contested term.
- IP : The principle communication protocol employed by the internet, and ubiquitous in computer networking.
- Latency : Temporal offset, delay, normally network related, here normally referring to perceived delay which causes cognitive dissonance when using a technology.
- Mindmap : Information structuring methodology.
- Mixed Reality : Integration of computer generated content with the real world with a view to blurring the distinction.
- Multi-modal : Here taken to mean that more than one sense is engaged.
- Multiview : Set of technologies which variously present one or more users with different views when they simultaneously apprehend the same display.
- Octave : A specific multi-modal research platform used in the research.
- Point to Point : Communication between only two users without regard for distance.
- Presence : The subjective feeling of being in a place or context. A complex and contested research term.
- Projection mapping : Mathematically or algorithmically adjusting a 2D image which is projected onto a complex known 3D surface to add perceptive or aesthetic features.

- RAW : For simplicity the least processed version of an image or image stream with respect to the capture device.

- Reconstructed : Generation of 3D data from a synthesis of 2D/3D data by various means.

- RPT : Retro-reflective projection technology is a new field which capitalises on the properties of some materials to reflect digitally projected structured light back along the source light path.

- Segmenting : A step in the reconstruction pipeline where 2D images are separated into pertinent and useless pixels.

- Shader lamp : Projection mapping from a projector or projectors onto surfaces by distorting and/or manipulating the image stream.

- Shape from silhouette : Segmenting plus some basic encoding to describe 2D structures.

- Situated : A relatively new and somewhat ambiguous term for a display which is suitable for its surrounding environment. Not incongruous. In place.

- Structured light : Another strategy for resolving 3D shape data for transmission of physical information. Uses multiple invisible known infrared patterns to reverse engineer shape.

- TCP/IP : Structured network protocol which invariably underpins digital communication. UDP is less relevant but similar.

- Telepresence : A somewhat contested definition which here is taken to mean the transmission of sufficient 'self' via technology to facilitate communication to a degree which warrants serious analysis.

- Time of flight : Another strategy for resolving 3D shape data for transmission of physical information. Uses the round trip time of lasers.

- Triadic : Interaction between three individuals.

- UML : Universal Markup Language is a standard for functional diagrams
- Uncanny : In this case always describes the cognitive dissonance when perceiving a less than faithful representation of a human with the result that the observer instinctively rejects the image, sometimes exhibiting revulsion.
- VC : Video Conferencing are technologies which seek to add a visual component to audio communication and is now established enough to cover a multitude of similar strategies.
- Vicon : Optical near-infrared spatial tracking system vendor. Such systems provide 3D data about the location of groups of markers in a volume.
- Visual Hull : The end product of whatever various technologies are used to generate 3D video from multiple viewpoints. See reconstruction, shape from silhouette and time of flight.
- VR : Virtual Reality, also VE, Virtual Environments.
- VRPN : Virtual Reality Peripheral Network is a long established communication system which simplifies development between vendors of different VR hardware.
- WLAN : Wireless local area network

Chapter 1

Introduction

With global pressure to justify travel for ecological reasons, there is an increasing need for a step change in technologies which can unite small groups over a distance. The goal is for "low overhead communication that naturally occurs when people are located in close proximity"[16]. To achieve this the technology should duplicate as many features of normal small group interaction, as transparently as possible.

Current ubiquitous systems support chiefly audio channels on a one to one basis (telephony). The telephone is indeed so ubiquitous that there are now more mobile phones than working toilets [331]. Just having an audio channel leads to proven emergent issues such as the actor-observer effect [159] where divergent expectations about context lead to complications in conversational flow. In their 1976 book "The Social Psychology of Telecommunication", Short, Williams, and Christie assert that face-to-face is the gold standard of interpersonal communication [293], with video mediated conversation less effective, and 'audio only' less effective still [233]. Erickson & Kellogg provide an empirical study of a large (n=123) cohort working at IBM [89]. This study ranked physical face-to-face meetings as preferable, with interesting results across various metrics for video conferencing, SecondLife, and web teleconferencing. Slater et al. found increased "group accord"

in the real as compared to the virtual [297].
Much research effort and commercial investment has been expended in providing additional modality for remote communication, trying to close this gap between real life and technology-mediated communication. This work has concentrated primarily on providing a visual representation of the distal participant. Single point to point connection of two users is well supported through technologies such as Skype, though significant problems still remain, with the reconnecting mutual eye gaze chief among these. This is explored in depth.
There is already established industry support for well structured meetings through systems such as ‘Cisco Telepresence’ rooms, but these too have shortcomings which are explored. There is seemingly no provision for groups which employ less structured and/or more dynamic meeting styles, with the exception of some attempts in SecondLife and public telematic installations [290].
The literature survey examines the historic body of research around interpersonal conversation, and seeks to capture the current state of the art around communicating across a distance using technology. Attempts to understand the transmission of attention across technology are explored, and where appropriate studied in more detail. Limitations in current technologies are identified and/or highlighted to expose remaining research challenges.
The experimental research aims to better understand the connection of small groups through a novel approach which addresses some of the current shortcomings.

1.1 Research Problem

The initial Telethrone concept was a response to a problem BBC R&D encountered through split site working. They found they could no longer ‘rub shoulders’ in informal or ad-hoc meetings. BBC R&D suggested that 10 unstructured meetings were optimal for leveraging the best outcomes from a single formal meeting. In a single site environment, such as their original London headquarters, this kind of interaction was supported through communal areas and the eponymous ‘water cooler meetings’. With the advent of split site working in London and Salford it became apparent that these clarifying meetings were being compromised. The ideal solution, therefore, would be a technology which could be economically integrated into a social space and could be ‘always-on’. This would potentially allow both serendipitous and ad-hoc meetings. Serendipitous meetings through a screen on a wall had been attempted by BBC R&D but did not seem to work, even if the system was always-on. One limitation may be the difficulty in correctly grabbing the attention of passing colleagues due to the spatial misalignment inherent in the Mona Lisa effect [4] and the limitation of the flat screens ‘containing’ another space.

1.1.1 Research Questions

The research questions evolved over the course of the research and make more sense in the context of the lessons learned throughout. They are as follows:

- **Q0:** Do the experimental set-up pre-requisites exist? i.e. Does light reflected from Chromatte cloth fall off in intensity as the angle subtended from the axis of projection increases. Is this in-line with the manufacturers datasheet when the material is used outside of their guidelines as in the proposed Telethrone. Are there additional subjective concerns?
- **Q1:** Does the Telethrone operate as an effective situated display (judged by questionnaire), which can show multiple views of a remote participant without excluding them from the attention (judged empirically) of the participant during a three-way conversation? Structured poker play designed to minimise but not preclude eye contact should be supported between three parties with the Telethrone. This will be examined statistically by the number of times the participant looks at the co-located player against the remote

player.

- **Q2:** In the course of investigating the social presence capabilities of the system in the first experiment a new line of inquiry was explored. What are the theoretical limits of the system, especially when moving chairs around dynamically in the meeting?

- **Q3:** Addressing shortcomings in the capability of the system to connect mutual eye gaze required 3D capture of the remote participant. To what extent does the new system, which allows the observer to move around the space, allow detection of eye gaze? Can it enable a group of five people to reliably detect non-verbal cues transmitted by the remote participant, in a simulated ideal sized meeting?

1.2 Scope and Limitations

Prof Roberts envisaged a multi-site system with Telethrones at each location, always-on, and always up-to-date with one another's relative positions such that they are spatially faithful. Roberts recommended that the Telethrone be tracked such that moving it in the space could maintain faithful relationships with the onlooking chairs. Such functionality is beyond the scope of this PhD. The presented research explores a single Telethrone deployment addressing several moving chairs, with asymmetry between the sites (the remote participant doesn't employ Telethrone technology).

The research set out to create as natural a setting as possible for the experimentation, but the available equipment and spaces mean that elements of the technology remain exposed and this limits the naturalness of the environment.

Initial investigation of Chromatte cloth, a commercially available retro-reflective surface, found it is potentially useful for separating multiple video channels to different observers, essentially a multi-view technology allowing for mutually spatial faithful installations [221, 54, 111]. Some additional exploration materials which might add value were made and discounted and the focus was applied to Chromatte. The extent to which this testing of retro-reflective cloth adds value or novelty (given that this is itself an established field), is discussed at the end of the thesis, along with other identified shortcomings.

The time it took to perform each run of the exclusion experiment [228], and the requirement that participants be able to play poker, limited the number of participants in the first experiment. This experiment is the most detailed and firmly establishes the system as supportive of social interaction.

Space and budget constraints limited the time available for deployment of the tracked system which underpinned the second experiment [229], and at the time there were no available research confederates. This made experimentation between members of a small group extremely difficult. For this reason, it was a more pragmatic choice to address a testable element based around the reconstruction - that of reconnected eye gaze. Reconstructed eye gaze was examined between a static frame from a recorded model and a participant. This limitation brought a compensation of complete repeatability to the experiment.

The final system tests a small group, and end to end reconstruction onto the Telethrone. Although all chairs are tracked, and the viewpoints generated are as described in the proposal for evolving fault lines, budget restrictions meant that it was impossible to move stations while supporting testing on subjects at the same time because of the available equipment. The system was tested within the parameters available. This work was prepared for submission but time constraints mean that the work is unpublished at this time.

1.3 Contributions

Primary Contributions which further the design and development of Telethrone are enumerated here, and correspond to the research questions seen above. The thesis will revisit this format of 0 through 3 throughout for simplicity of reference, and the chapters are laid out conformal to the questions, with the hypotheses enumerated within the same schema. The contributions, though listed here, are better framed and discussed in the final chapter:

- **C0:** Performed testing of physical characteristics of the retro-reflective cloth suggested by Roberts and Graham Thomas (BBC), to ensure that it was suitable for the intended purpose in principle. This performance data for the cloth formed the basis of a later discovery that the optical characteristics of this specific brand of retro-reflective material degrade over time. This is in itself a contribution. The early testing and prototyping considered and disregarded some image processing approaches which might have allowed better fitting of the projected image of the remote collaborator onto the chair. These approaches are useful markers for when the performance of the components improves sufficiently to match human perceptual considerations.
- **C1:** Performed rigorous set-up of a system capable of transmitting bi-directional video streams between two physically separated locations. This supported a three-way conversation with the display. An experiment tried to get a feel for how situated the display was, and a rigorous analysis of attention finds that a remote collaborator is not excluded from structured play by some characteristic of the technology. This contribution builds on others research into both situated displays and multiple spatially segmented views onto a remote collaborator.
- **C2:** Discussion of how the elements of the system support novel interaction within small groups, which are partially mediated by a tele-presence boundary. Analysis of how the system could and should be scaled explores the likely limits. Some theoretical consideration is applied to the current system implemented in the research, detailing the constraints and challenges.
- **C3:** Integration with spatial tracking, and playback of a recorded 3D video session in Octave enabled flexible proxemics with a view to potentially supporting dynamic meetings. An experiment tests the extent to which participants could resolve the eye gaze of the polygonal hull reconstruction (previously developed by others) in this new more demanding context. A

second experiment explores the degree to which a maximal system supports non-verbal cues in a group context. This builds directly on work done in the wider research group into ability to resolve eye gaze at social distance and makes it applicable to the Telethrone. The experimental set-up proves for the first time that it is possible to move around the Telethrone with multiple spatially accurate viewpoints onto the remote collaborator, as envisaged by Roberts, Thomas, and Sermon.

Chapter 2

Literature Review & Theoretical Framework

2.1 Chapter Overview

A good literature review is central to new research and an undertaking such as this thesis [47]. An attempt is made to capture the current available knowledge surrounding subsets of research into interpersonal communication.

Interaction between co-located parties and groupings, which are further mediated by technology over a distance, are considered. This builds knowledge to summarise any effects pertaining to human perception of others through media, and to explore the current state of the art in human communication across a distance.

It should be noted that the literature review was conducted in the first two years of a 7 year part time investigation (around 2013). The methodology mandates revisiting the literature iteratively but these later additions are described in the closing chapter rather than this one. Consequently, the span of time may make some of the review seem out of date given the pace of change in this research area.

The theoretical framework outlines and/or defines key concepts around the research problem. Pertinent theories are identified and referenced, and where they are critical to the investigation they are justified. Relationships between the concepts are identified and explained. The specific area of research of the thesis should sit at the centre of these interrelated fields.

2.2 Introduction

It is evident that point to point connection of multiple users, with IP network encoding of voice and/or facial camera, is well supported through technologies such as telephony & Skype, with Skype alone accounting for 280 million connections per month [215]. Multi person teleconferencing and visually enhanced teleconferencing such as IBM's IEAC or AccessGrid as seen in Figure 2.1 provide a shared channel for group interaction [76].

Established technologies extend traditional telephony to provide important multi-modal (multiple sense) cues through non-verbal communication [12, 349]. However, these technologies demonstrate shortfalls compared to a live face-to-face meeting, which is generally agreed to be optimal for human-human interaction [349, 335].

The research aims to add knowledge to the efforts to best create comfortable and pervasive telecommunication. It will be shown that there is advantage to accurate connection of the gaze between conversational partners and there is a large body of evidence that physical communicational channels extend beyond the face and include both micro and macro movement of the upper body.

An understanding of baseline human-human interaction underpins any investigation into communication which is obfuscated by a technology boundary.

There are established fields of study within social sciences which break this problem down, these are explored next.

2.3 Human - Human interaction: Communication theory

Oxford English Dictionary 2010 defines this as the "branch of knowledge dealing with the principles and methods by which information is conveyed". This is a broad

Figure 2.1: An access grid system. Such systems form a loose specification of 'quality' for aspects of the connection including latency, resolution, echo cancellation, etc. They are difficult to set-up and maintain and increasingly rare with Google searches for the term peaking in 2004 at 20 times the current level.

church with a huge scope. The thesis concentrates on verbal and non-verbal cues most important to the wider research into technology-mediated communication.

2.3.1 Vocal

A good way to begin an investigation into interaction is to use delineations provided by the social sciences; in his book 'Bodily Communication' [10] Michael Argyle divides vocal signals into the following categories:

1. Verbal

2. Non-Verbal
    (a) Linked to Speech
        i. Prosodic
        ii. Synchronising
        iii. Speech Disturbances
    (b) Independent of Speech
        i. Emotional Noises
        ii. Paralinguistic (Emotion and Interpersonal attitudes)
        iii. Personal Voice and Quality of Accent

Additional to the semantic content of verbal communication in the real world there is a rich layer of meaning in pauses, gaps, and overlaps [139] which help to mediate who is speaking and who is listening (turn passing), in multiparty conversation. Mediation of turn passing to facilitate flow is by no means a given and is highly dependent on context and other factors [168]. Interruptions are also a major factor in turn passing.
This extra-verbal content [324] extends into physical cues, so-called 'nonverbal' cues, and there are utterances which link the verbal and non-verbal [236]. This will be discussed later, but to an extent, it is impossible to discuss verbal communication without regard to the implicit support which exists around the words themselves.
In the context of all technology-mediated conversation the extra-verbal is easily compromised if technology used to support communication over a distance does not convey the information or conveys it badly. This can introduce additional complexity [236].
When examining just verbal / audio communication technology it can be assumed that the physical non-verbal cues are lost, though not necessarily unused. In the absence of non-verbal cue it falls to timely vocal signals to take up the slack when framing and organising the turn passing. For the synchronising of vocal signals between the parties to be effective the systemic delays must remain small. System latency, the inherent delays added by the communication technology can allow slips or a complete breakdown of 'flow' [163].

It is clear then that transmission of verbal / audio is the most critical element for interpersonal communication since the most essential meaning is encoded semantically. There is a debate about ratios of how much information is conveyed through the various human channels [192], but it is reasonable to infer from its ubiquity that support for audio is essential for meaningful communication over a distance. We have seen that it must be timely, to prevent a breakdown of framing, and preferably have sufficient fidelity to convey sub-vocal utterances.
The Telethrone system designed through the research based on this review should support sound with sufficient directionality to convey the impression of where the talking participant is positioned.

2.3.2 Nonverbal

We have already seen that verbal exchanges take place in a wider context of sub vocal and physical cues. In addition, the spatial relationship between the parties, their focus of attention, their gestures and actions, and the wider context of their environment all play a part in communication [120]. These are summarised well by Gillies and Slater [117] in their paper on virtual agents.

Gaze

Of particular importance is judgement of eye gaze which is fast, accurate and automatic, operating at multiple levels of cognition through multiple cues [10, 9, 11, 12, 13, 166, 209].
Gaze in particular aids with smooth turn passing [136] [226] and lack of support for eye gaze has been found to decrease the efficiency of turn passing by 25% [340].
There are clear patterns to eye gaze in groups, with the person talking, or being talked to, probably also being looked at [339] [184]. To facilitate this groups will tend to position themselves to maximally enable observation of the gaze of the

other parties [166]. This intersects with proxemics which will be discussed shortly. In general people look most when they are listening, with short glances of 3-10 seconds [11]. Colburn et al. suggest that gaze direction and the perception of the gaze of others directly impacts social cognition [60] and this has been supported in a follow up study [196].

The importance of gaze is clearly so significant in evolutionary terms that human acuity for eye direction is considered high at 30 sec arc [314] with straight binocular gaze judged more accurately than straight monocular gaze [171], when using stereo vision.

Regarding the judgement of the gaze of others Symons et al. suggesting that "people are remarkably sensitive to shifts in a person's eye gaze" in triadic conversation [314]. Wilson et al. found that subjects can "discriminate gaze focused on adjacent faces up to [3.5m]" [348]

This perception of the gaze of others operates at a low level and is automatic. Langton et al. cite research stating that the gaze of others is "able to trigger reflexive shifts of an observer's visual attention" and further discuss the deep biological underpinnings of gaze processing [183].

When discussing technology-mediated systems Vertegaal & Ding suggested that understanding the effects of gaze on triadic conversation is "crucial for the design of teleconferencing systems and collaborative virtual environments" [338], and further found correlation between the amount of gaze, and amount of speech. Vertegaal & Slagter suggest that "gaze function(s) as an indicator of conversational attention in multiparty conversations" [339]. Vertegaal provides an "attentive state model" [336] which may be of use for analysis in Telethrone experiments.

There is research investigating this sensitivity when the gaze is mediated by a technology, finding that "disparity between the optical axis of the camera and the looking direction of a looker should be at most 1.2°in the horizontal direction, and 1.7°in vertical direction to support eye contact" [334, 42]. It seems however that humans assume that they are being looked at unless they are sure that they are not, and this may give some additional lattitude in the design of the Telethrone [57].

Vertegaal et al. found that task performace was 46% better when gaze was synchronised in their telepresence scenario. As they point out, gaze synchonisation (temporal and spatial) is 'commendable' in all such group situations, but the precise utility will depend upon the task [338].

There has been some success in the automatic detection of the focus of attention of participants in multi party meetings [311, 309]. More recently, advanced eye tracking technologies allow the recording and replaying of accurate eye gaze information [308] alongside information about pupil dilation toward determination of honesty and social presence [305]. Such application of technology would likely fall outside the scope of this research with more traditional encoding techniques such as those applied by Argyle and Cook [9] alongside video recording being more realistic.

In summary, gaze awareness does not just mediate verbal communication but rather is a complex channel of communication in its own right. Importantly for the research, gaze has a controlling impact on those who are involved in the communication at any one time, including and excluding even beyond the current participants.

The Telethrone system should support directionality of gaze with sufficient accuracy such that participants know when they are being looked at, and perhaps enough flexibility to support gaze cues into a wider environment.

Mutual Gaze

Aygyle and Cook established a great deal of science around gaze and mutual gaze, with their seminal book of the same title [9], additionally detailing confounding factors around limitations and inaccuracies in observance of gaze and how this varies with distance [13] [10] [61].

Mutual gaze is considered to be the most sophisticated form of gaze awareness with significant impact on dyadic conversation especially [61, 168, 92]. The effects seem more profound than just helping to mediate flow and attention, with mutual eye gaze aiding in memory recall and the formation of impressions [44].

While reconnection of mutual eye gaze through a technology boundary does not seem completely necessary it is certainly important, with impact on subtle elements of one to one communication, and therefore discrimination of eye gaze direction should be bi-directional if possible, and if possible have sufficient accuracy to judge direct eye contact. In their review Bohannon et al. said that the issue of rejoining eye contact must be addressed in order to fully realise the richness of simulating face to face encounters [44].

Mutual gaze is a challenging affordance for Telethrone as bi-directional connection of gaze is not a trivial problem since the camera or virtual camera must be coincident with the eyes of the participants. This problem is described in a later section.

Head Orientation

Orientation of the head (judged by the breaking of bilateral symmetry and alightment of nose) is a key factor when judging attention. Perception of head orientation can be judged to within a couple of degrees [348].

It has been established that head gaze can be detected all the way out to the extremis of peripheral vision, with accurate eye gaze assessment only achievable in central vision [191]. Features of illumination can alter the apparent orientation of the head [326].

Head motion over head orientation is a more nuanced propostion and can be considered a micro gesture [45]. Regarding video mediated teleconferencing; Vertegaal et al. [340] note that the "larger the distance of head to screen, or the smaller the projected images, the more head movement of users is tolerable without impairing the conveyance of gaze at the eyes". While users of video conferencing equipment tend to position themselves correctly in front of the camera, movement speed of the head within the frame during use is typically in the region 2-3 meters per second. Framerate of the camera should support this [43].

It is possible that 3D displays are better suited to perception of head gaze since it is suggested that they are more suitable for "shape understanding tasks" [155] and a possible further development to Telethrone could involve multiple narrow baseline stereo pairs similar to the system outlined by Cooke [63]. This would allow stereoscopic projection of IP video streams.

Bailenson, Baell, and Blascovich found that giving avatars rendered head movements in a shared virtual environment decreased the amount of talking, possibly as the extra channel of head gaze was opened up. They also reported that subjectively, communication was enhanced [18].

Clearly head orientation is an important indicator of the direction of attention of members of a group and can be discerned even in peripheral vision. This allows the focus of several parties to be followed simultaneously and is an important affordance to replicate on any multi-party communication system. To support head gaze (i.e. the attention vector implied by the reletive angle of the head) it is necessary to have a continuum or spectrum of views onto a tele-present participant, or at a minimum multiple independent viewpoints as in the Telethrone design presented.

Combined Head and Eye Gaze

Rienks et al. found that head orientation alone does not provide a reliable cue for identification of the speaker in a multiparty setting [262]. Stiefelhagen & Zhu found “that head orientation contributes 68.9% to the overall gaze direction on average” [310], though head and eye gaze seem to be judged interdependently [171]. Langton noted that head and eye gaze are “mutually influential in the analysis of social attention” [182], and it is clear that transmission of ‘head gaze’ by any mediating system enhances rather than replaces timely detection of subtle cues. Combined head and eye gaze give the best of both worlds and extend the lateral field of view in which attention can be reliably conveyed to others.

Other Upper Body: Overview

While it is well evidenced that there are advantages to accurate connection of the gaze between conversational partners [13, 168], there is also a body of evidence that physical communication channels extend beyond the face [168, 222] and include both micro (shrugs, hands and arms), and macro movement of the upper body [87]. Goldin-Meadow suggests that gesturing aids conversational flow by resolving mismatches and aiding cognition [119].

In their technology-mediated experiment which compared face to upper body and face on a flat screen Nguyen and Canny found that “upper-body framing improves empathy measures and gives results not significantly different from face-to-face under several empathy measures” [222]. An image from their experiment is seen in Figure 2.2.

Figure 2.2: Head and upper body framing experiment by Nguyen and Canny. They used physical masks to block out the body shown on the large screen.

Other Upper Body: Facial Much emotional context can be described by facial expression (display) alone [87, 58], with a cognitive preference for high temporal resolution over image resolution [280]. Some aspects of conversational flow appear to be mediated in part by facial expression [230]. There are gender differences in the perception of facial affect [142].

Other Upper Body: Gesturing Gesturing (such as pointing at objects) paves the way for more complex channels of human communication and is a basic and ubiquitous channel [152]. Conversational hand gestures provide a powerful additional augmentation to verbal content [174].

Other Upper Body: Posture Some emotions can be conveyed through upper body configurations alone. Argyle details some of these [10] and makes reference to the posture of the body and the arrangement of the arms (i.e. folded across the chest). These are clearly important cues. Kleinsmith and Bianchi-Berthouze assert that ”some affective expressions may be better communicated by the body than the face” [169].

Other Upper Body: Body Torque In multi-party conversation, body torque, that is the rotation of the trunk from front facing, can convey aspects of attention and focus [279].

In summary, visual cues which manifest on the upper body and face can convey meaning, mediate conversation, direct attention, and augment verbal utterances. The Telethrone system should support display of the upper body and face with directionality.

Effect of Shared Objects on Gaze

Ou et al. detail shared task eye gaze behaviour “in which helpers seek visual evidence for workers’ understanding when they lack confidence of that understanding, either from a shared, or common vocabulary” [239].
Murray et al. meanwhile found that in virtual environments, eye gaze is crucial for discerning what a subject is looking at [217]. This work is shown in Figure 2.3.
It is established that conversation around a shared object or task, especially a complex one, mitigates gaze between parties [12] and this suggests the choice of task is an important factor when examining eye gaze in experimental groups.
In the shared poker task (which is used to assess the Telethrone design), turn taking, and conflict resolution, sit alongside attempts to ‘read’ the other players. This seemed a good balance, though came with it’s own issues explored later.

Figure 2.3: Eye tracked eye gaze awareness in VR. Murray et al. used immersive and semi immersive systems alongside eye trackers to examine the ability of two avatars to detect the gaze awareness of a similarly immersed collaborator in the same virtual space when they took turns to look at a grid of objects in the air between them.

2.4 Human - Human Interaction: Interpersonal Psychology

2.4.1 Proxemics

Proxemics is the formal study of the regions of interpersonal space begun in the late 50's by Hall and Sommers and building toward The Hidden Dimension [130], which details bands of space that are implicitly and instinctively created by humans and which have a direct bearing on communication. These are seen in Figure 2.4. Distance between conversational partners, and affiliation, also have a bearing on the level of eye contact [11] with a natural distance equilibrium being established and developed throughout through both eye contact and a variety of subtle factors.

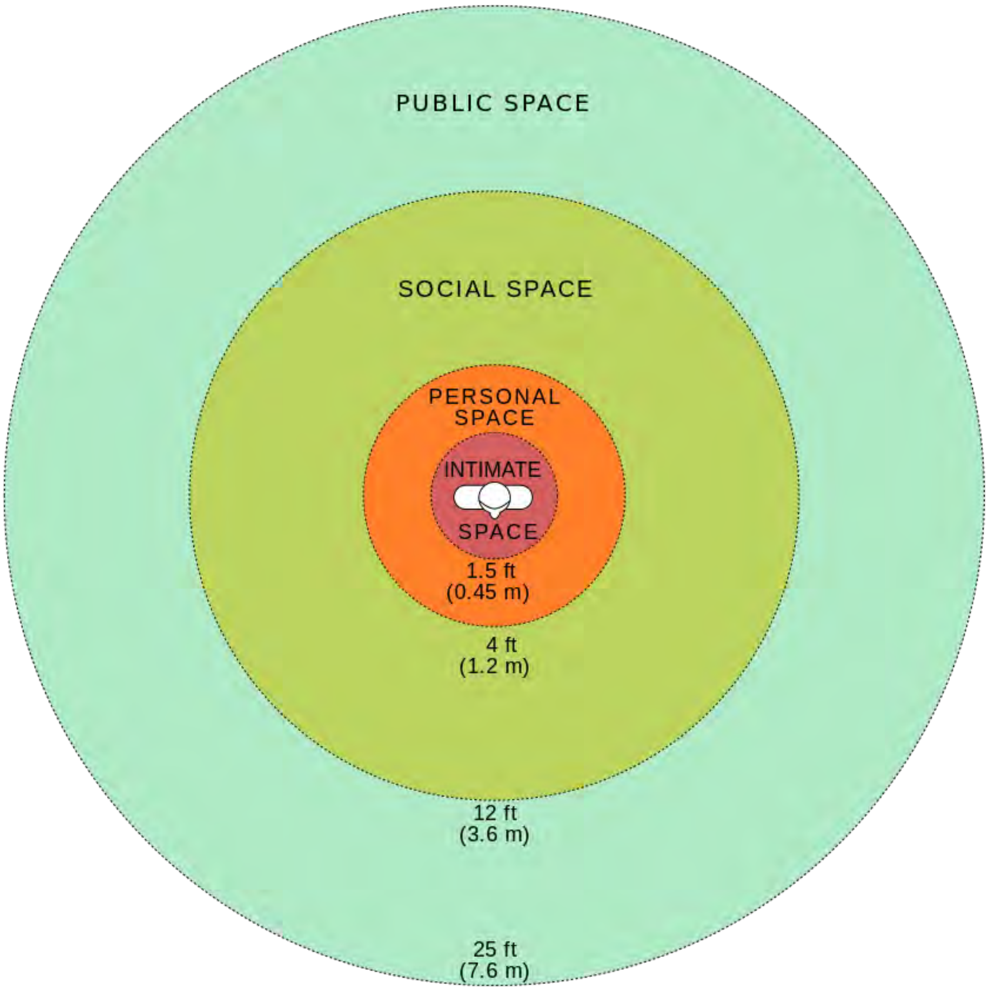


Figure 2.4: Hall's proxemics zones show where people respond differently to one another depending on the distance between them. Telethrone should be in the social space.

Argyle & Ingham provide levels of expected gaze and mutual gaze against distance [13]. These boundaries are altered by ethnicity [346, 10] and somewhat by gender [49], and age [299, 142].

Even with significant abstraction by communication systems (such as SecondLife) social norms around personal space persist [352, 17, 19]. Bailenson & Blascovich found that even in Immersive Collaborative Virtual Environments (ICVE's) "participants respected personal space of the humanoid representation"[17] implying

that this is a deeply held 'low-level' psychophysical reaction [41].
It is highly applicable to experimental design that the noted 'social space' be considered throughout.
Maeda et al. [197] found that seating position impacts the level of engagement in teleconferencing. Taken together with the potential for reconfiguration within the group as well as perhaps signaling for the attention of participants outside of the confines of the group Telethrone design should be mindful of proxemics.
When considering the attention of engaging with people outside the confines of a meeting Hager et al. found that gross expressions can be resolved by humans at tremendous distances [129, 10]. It seems that social interaction begins around 7.5m in the so-called 'public space' [130].

2.4.2 Attention

The study of attention is a discrete branch of psychology. It is the study of cognitive selection toward a subjective or objective sub focus, to the relative exclusion of other stimulae. It has been defined as "a range of neural operations that selectively enhance processing of information" [55]. In the context of interpersonal communication it can be refined to apply to selectively favouring a conversational agent or object or task above other stimuli in the contextual frame.
Humans can readily determine the focus of attention of others in their space [311] and preservation of the spatial cues which support this are important for technology-mediated communication [286] [309].
Symons et al. suggested that "people are remarkably sensitive to shifts in a person's eye gaze" in triadic conversation [314]. Wilson et al. found that subjects can "discriminate gaze focused on adjacent faces up to [3.5m]" [348]
The interplay between conversational partners of attention and the reciprocal noticing of attention is dubbed the perceptual crossing [72, 116].
This is a complex field of study with gender, age, and ethnicity all impacting the behaviour of interpersonal attention [33, 299, 10, 142, 241]. Vertegaal has done a great deal of work on awareness and attention in technology-mediated

situations and the work of his group is cited throughout the research[336]. As an example it is still such a challenge to "get" attention through mediated channels of communication, that some research [97, 286] and many commercial systems such as 'blackboard collaborate', use tell tale signals (such as a microphone icon) to indicate when a participant is actively contributing. Some are automatic, but many are still manual, requiring that a user effectively hold up a virtual hand to signal their wish to communicate.

Langton et al. cite research stating that the gaze of others is "able to trigger reflexive shifts of an observer's visual attention". Novick et al. performed analysis on task hand-off gaze patterns which is useful for extension into shared task experiment design [226].

Attention and the sensing of attention rely on the reconnection of non-verbal cues across the whole of the space in which communication occurs. In the case of a dynamic informal meeting this area of communicational attention may extend beyond the bounds of the meeting.

Functioning at a more complex level of cognition than simple attention is interpersonal behaviour. Some of this is derived from information provided by attention and that of others, but behaviour at this level can be considered to apply 'modifiers' which reflect more complex drivers.

Regarding the attention of others, Fagal et el demonstrated that eye visibility impacts collaborative task performance when considering a shared task [92].

The implication from this is that the whole of the space addressed by the Telethrone participant must be spatially faithful from the viewpoints of the onlookers in order to support their ability to discriminate focus of attention.

2.4.3 Behaviour

Hedge et al. suggested that gaze interactions between strangers and friends may be different which could have an impact on research recruitment [136]. Voida et al. elaborate that prior relationships can cause "internal fault lines" in group working [343]. This concept of fault lines became increasingly pertinent as the research evolved and is explored in more detail later.

When new relationships are formed the primary concern is one of uncertainty reduction or increasing predictability about the behavior of both themselves and others in the interaction [34]. This concept of smoothness in the conversation is a recurring theme, with better engineered systems introducing less extraneous artifacts into the communication, and so disturbing the flow less. In a similar vein the actor-observer effect describes the mismatch between expections which can creep into conversation. Conversations mediated by technology can be especially prone to diverging perceptions of the causes of behaviour [159]. This may be especially prevelant in shared immersive environments where different subjective views of the same objective shared dataset exist [301].

Interacting subjects progress conversation through so-called ‘perception-action’ loops which are open to predictive modeling through discrete hidden Markov models [206].

It may be that the perception-behaviour link where unconscious mirroring of posture bolsters empathy between conversational partners, especially when working collaboratively [56], and the extent to which posture is represented through a communication medium may be important.

Landsberger posited the Hawthorne effect [246]. Put simply this is a short term increase in productivity that may occur as a result of being watched or appreciated. This has broader implications for experimentation where tasks are observed or are perceived to be observed. The impression of being watched changes gaze patterns during experimentation, with even implied observation through an eye tracker modifying behaviour [263].

There is much historic work describing “the anatomy of cooperation” [173] and this may be useful in analysing experimental data around common collaborative tasks.

Blascovich proposes a model “within technology-mediated and/or determined contexts” [41]. This work suggests that some catagories of behaviour are more important than others [345] and deserve different degrees of support, though again this is context dependent.

Johnson discusses the importance to cognition of represtations from a psychology perspective and presents guidelines for digital representations [156].

Cuddihy and Walters defined an early model for assessing desktop interaction mechanisms for social virtual environments [68].

Perception Of Honesty

Hancock et al. state that we are most likely to lie, and to be lied to, on the telephone [131]. Technology used for communication impacts interpersonal honesty. It seems that at some level humans know this; lack of eye contact leads to feelings of deception, impacting trust [143]. This has a major impact on video conferencing which often does not support mutual gaze. Trust is crucial for business interactions. Further there are universal expressions, microexpressions, and blink rate which can betray hidden emotions [251], though the effects are subtle and there is a general lack of awareness by humans of their abilities in this regard [143]. Absence of support for such instinctive cues inhibits trust [273]. Support for these rapid and transient facial features demands high resolution reproduction in both resolution and time domains. There is detectable difference in a participant's ability to detect deception when between video conference mediated communication and that mediated by avatars [306]. Systems should aim for maximally faithful reproduction. This testing would be an interesting experiment for Telethrone and could be addressed in future research.

2.5 Technology-Mediated Interaction

2.5.1 Dyadic

Point to Point Telephony (Audio only)

The ubiquitous technology to mediate conversation is, of course, the telephone. The 2015 Ericsson mobility report [1] states that there are a staggering 7.4 billion mobile subscriptions. More people have access to mobile phones than to working toilets [331].

Joupii and Pan designed a system which focused attention on spatially correct high definition audio. They found “significant improvement over traditional audio conferencing technology, primarily due to the increased dynamic range and directionality. [160]. Aoki et al. also describe an audio only system with support for spatial cues [8].

Point to Point Video Conferencing

O’Malley et al. showed that face-to-face and video mediated employed visual cues for mutual understanding, and that addition of video to the audio channel aided confidence and mutual understanding. However, video mediated did not provide the clear cues of being co-located [233].

Dourish et al. make a case for not using face to face as a baseline for comparison, but rather that analysis of the efficacy of remote tele-collaboration tools should be made in a wider context of connected multimedia tools and ‘emergent communicative practises’ [79]. While this is an interesting viewpoint it does not necessarily map well to a recreation of the ad-hoc meeting.

There is established literature on human sensitivity to eye contact in both 2D and 3D VC [337, 334], with an accepted minimum of 5-10°before observers can reliably sense they are not being looked at [57]. Roberts et al. suggested that at the limit of social gaze distance (4m) the maximum angular separation between people standing shoulder to shoulder in the real world would be around 4°[272].

Sellen found limited impact on turn passing when adding a visual channel to audio between two people when using Hydra, an early system which provided multiple video conference displays in an intuitive spatial distribution[287]. She did however find that the design of the video system affected the ability to hold multi-party conversations [288].

Monk and Gale describe in detail experiments which they used for examining gaze awareness in mediated and unmediated communication. They found that gaze awareness increased message understanding [209].

Both Kuster et al. and Gemmel et al. have successfuly demonstrated software systems which can adjust eye gaze to correct for off axis capture in real time video

systems[113, 180].
Shahid et al. conducted a study on pairs of children playing games with and without video mediation and concluded that the availability of mutual gaze affordance enriched social presence and fun, while its absence dramatically affects the quality of the interaction. They used the 'Networked Minds', a social presence questionnaire. Social presence will be discussed in detail after the following section, but this work first pointed to the potential usefulness of this kind of survey for evaluating the quality of interaction in poker play on Telethrone [292].

2.5.2 Triadic and Small Group

Early enthusiasm in the 1970's for video conferencing as a medium for small group interaction quickly turned to disillusionment. It was agreed after a flurry of initial research that the systems at the time offered no particular advantage over audio only communication, and at considerable cost [347].
Something in the breakdown of normal visual cues seems to impact the ability of the technology to support flowing group interaction. Nonetheless, some non-verbal communication is supported in VC with limited success.
Additional screens and cameras can partially overcome the limitation of no multi-party support (that of addressing a room full of people on a single screen) by making available more bidirectional channels. For instance, every remote user can be a head on a screen with a corresponding camera. The positioning of the screens must then necessarily match the physical organization of the remote room.
Egido provides an early review of the failure of VC for group activity, with the "misrepresentation of the technology as a substitute for face to face" still being valid today [85].
Commercial systems such as Cisco Telepresence Rooms cluster their cameras above the centre screen of three for meetings using their telecollaboration product, while admitting that this only works well for the central seat of the three screens. They also group multiple people on a single screen in what Workhoven et al. dub a "non-isotropic" configuration [248]. They maintain that this is a suitable trade off as the focus of the meeting is more generally toward the important contributor

in the central seat [315]. This does not necessarily follow for less formal meeting paradigms.

In small groups, it is more difficult to align non-verbal cues between all parties, and at the same time, it is more important because the hand-offs between parties are more numerous and important in groups. A breakdown in conversational flow in such circumstances is harder to solve, a perception of the next person to talk must be resolved for all parties and agreed upon to some extent.

However, most of the conventional single camera, and expensive multi camera VC systems, suffer a fundamental limitation in that the offset between the camera sight lines and the lines of actual sight introduce incongruities that the brain must compensate for [349].

For experimental design Bailenson found the game '20 questions' to be effective in analysis of triadic attention, specifically watching the head gaze [18]; this may be useful for the Telethrone experiments.

2.5.3 Presence, Co-presence, and Social Presence

Presence is a heavily cited historic indicator of engagement in virtual reality, though the precise meaning has been interpreted differently by different specialisms [28, 284]. It is generally agreed to be the 'sense of being' in a virtual environment [295]. Slater extends this to include the "extent to which the VE becomes dominant".

Beck et al. reviewed 108 articles and synthesised an ontology of presence [28] which at its simplest is as follows:

1. Sentient presence
    (a) Physical interaction
    (b) Mental interaction
2. Non-sentient
    (a) Physical immersion
    (b) Mental immersion = psychological state

When presence is applied to interaction it may be split into Telepresence, and Co/Social presence [137, 36]. Co-presence and/or social presence is the sense of “being there with another”, and describes the automatic responses to complex social cues [53, 103, 135]. Social presence (and co-presence) refers in this research context to social presence which is mediated by technology (even extending to text based chat [128]), and has its foundations in psychological mechanisms which engender mutualism in the ‘real’. This is analysed in depth by Nowak [227]. An examination of telepresence, co-presence and social presence necessarily revisits some of the knowledge already elaborated.

The boundaries between the three are blurred in research with conflicting results presented [50]. Biocca et al. attempted to enumerate the different levels and interpretations surrounding these vague words [37], and to distill them into a more robust theory which better lends itself to measurement. They suggest a solid understanding of the surrounding psychological requirements which need support in a mediated setting, and then a scope that is detailed and limited to the mediated situation.

Since ‘social presence’ has been subject to varied definitions [37] it is useful here to consider a single definition from the literature which defines it as “the ability of participants in the community of inquiry to project their personal characteristics into the community, thereby presenting themselves to the other participants as real people.” [107, 28]. Similarly to specifically define co-presence for this research it is taken to be the degree to which participants in a virtual environment are “accesible, available, and subject to one another” [37].

Social presence has received much attention and there are established questionnaires used in the field for measurement of the levels of perceived social presence yet the definitions here also remain broad, with some confusion about what is being measured [38].

Telepresence meanwhile is interaction with a different (usually remote) environment which may or may not be virtual, and may or may not contain a separate social/co-presence component.

Even in simple videoconferencing Bondareva and Bouwhuis stated (as part of an experimental design) that the following determinants are important to create social presence [46, 160]. In principle the Telethrone system can support all of these.

1. Direct eye contact is preserved
2. Wide visual field
3. Both remote participants appear life size
4. Possibility for participants to see the upper body of the interlocutor
5. High quality image and correct colour reproduction
6. Audio with high S/N ratio
7. Directional sound field
8. Minimization of the video and audio signal asynchrony
9. Availability of a shared working space.

Bondareva et al. went on to describe a person to person telepresence system with a semi silvered mirror to reconnect eye gaze, which they claimed increased social presence indicators. Interestingly they chose a checklist of interpersonal interactions which they used against recordings of conversations through the system [46]. Their examination of eye contact frequency and duration is a good quantitative method with little ambiguity and is used in the Telethrone experiment.

The idea of social presence as an indicator of the efficacy of the system suggests the use of social presence questionnaires in the evaluation of the system [37]. Subjective questionnaires are however troublesome in measuring effectiveness of virtual agents and embodiments, with even nonsensical questions producing seemingly valid results [296]. Usoh et al. found that 'the real' produced only marginally higher presence results than the virtual [332].

Nowak states that "A satisfactory level of co-presence with another mind can be achieved with conscious awareness that the interaction is mediated" and asserts that while the mediation may influence the degree of co-presence it is not a prohibiting factor [227].

Baren and IJsselsteijn [333, 132] list 20 useful presence questionnaires in 2004 of which "Networked Minds" seemed most appropriate for the research. Hauber et al.

employed the "Networked Minds" Social Presence questionnaire experimentally and found that while the measure could successfully discriminate between unmediated and mediated triadic conversation, it could not find a difference between 2D and 3D mediated interfaces [134, 128].

In summary, social presence and co-presence are important historic measures of the efficacy of a communication system. Use of the term in literature peaked between 1999 and 2006 according to Google's ngram viewer and has been slowly falling off since. The questionnaire methodology has been challenged in recent research and while more objective measurement may be appropriate, the networked minds questions seem to be able to differentiate real from virtual interactions [132].

2.5.4 Mona Lisa Type Effects

The so-called Mona Lisa effect describes the phenomenon where the apparent gaze of a portrait or 2 dimensional image always appears to look at the observer regardless of the observer's position [342, 7, 350]. This situation manifests when the painted or imaged subject is looking into the camera or at the eyes of the painter [191, 104].

Single user-to-user systems based around bidirectional video implicitly align the user's gaze by constraining the camera to roughly the same location as the display. When viewed away from this ideal axis, it creates the feeling of being looked at regardless of where this observer is [4, 342, 7, 350], or the "collapsed view effect" [221] where perception of gaze transmitted from a 2 dimensional image or video is dependent on the incidence of originating gaze to the transmission medium.

Multiple individuals using one such channel can feel as if they are being looked at simultaneously, leading to a breakdown in the normal non-verbal communication which mediates turn passing [338]. This was mentioned briefly earlier when describing the Cisco Telepresence suites which often have a single central camera.

It seems that the effect is truely limited to 2D surfaces. A 3D projection surface (a physical model of a human) designed to address this problem completely removed the Mona Lisa effect [214].

2.5.5 Mutual Gaze in Telepresence

We have seen that transmission of attention can broadly impact communication in subtle ways, impacting empathy, trust, cognition, and co-working patterns. Mutual gaze (looking into one another's eyes), is currently the high water mark for technology-mediated conversation.

Many attempts have been made to re-unite mutual eye gaze when using teleconferencing systems. In their 2015 review of approaches Regenbrecht and Langlotz found that none of the methods they examined were completely ideal [260]. They found most promise in 2D and 3D interpolation techniques, which will be discussed in detail later, but they opined that such systems were very much ongoing research and lacked sufficient optimisation.

A popular approach uses the so called 'Peppers Ghost' phenomenon [304], where a semi silvered mirror presents an image to the eye of the observer, but allows a camera to view through from behind the angled mirror surface. The earliest example of this is Rosental's two way television system in 1947 [275], though Buxton et al. 'Reciprocal Video Tunnel' from 1992 is more often cited [52]. This optical characteristic isn't supported by RPT technology, and besides requires careful control of light levels either side of the semi-silvered surface.

The early GAZE-2 system (which makes use of Pepper's ghost) is novel in that it uses an eye tracker to select the correct camera from several trained on the remote user. This ensures that the correct returned gaze (within the ability of the system) is returned to the correct user on the other end of the network [341]. Mutual gaze capability is later highlighted as an affordance supported or unsupported by key research and commercial systems. This comparison takes place in the literature review without comparison to Telethrone, then is expanded with more systems and Telethrone included in the concluding chapter. It is therefore unnecessary to examine which important systems feature this capability here.

2.5.6 Other Systems to Support Business

There have been many attempts to support group working and rich data sharing between dispersed groups in a business setting. So called 'smart spaces' allow interaction with different displays for different activities and add in some ability to communicate with remote or even mobile collaborators on shared documents [23], with additional challenges for multi-disciplinary groups who are perhaps less familiar with one or more of the technology barriers involved [3].
Early systems like clearboard [150] demonstrated the potential for smart whiteboards with a webcam component for peer to peer collaborative working. Indeed it is possible to support this modality with Skype and a smartboard system (and up to deployments such as Accessgrid). They remain relatively unpopular however.

Tabletop and Shared Task

In early telepresence research Buxton and William argued through examples that “effective telepresence depends on quality sharing of both person and task space [51].
In their triadic shared visual workspace Tang et al. found difficulty in reading shared text using a ‘round the table’ configuration, a marked preference for working collaboratively on the same side of the table. They also found additional confusion as to the identity of remote participants [319]. Tse et al. found that pairs can work well over a shared digital tabletop, successfully overcoming a single user interface to interleave tasks [327].
Tang et al. demonstrate that collaborators engage and disengage around a group activity through several distinct, recognizable mechanisms with unique characteristics [318]. They state that tabletop interfaces should offer a variety of tools to facilitate this fluidity.
Camblend is a shared workspace which is hybrid physical and digital and which maintains some spatial cues between locations [225, 224]. Participants successfully resolved such co-orientation within the system.

The t-room system implemented by Luff et al. surrounds co-located participants standing at a shared digital table with life sized body and head video representations of remote collaborators [195] but found that there were incongruities in the spatial and temporal matching between the collaborators which broke the flow of conversation. Tuddenham et al. found that co-located collaborators naturally devolved 'territory' of working when sharing a task space, and that this did not happen the same way with a tele-present collaborator [330]. Instead remote collaboration adapted to use a patchwork of ownership of a shared task. It seems obvious to say that task ownership is a function of working space, but it is interesting that the research found no measurable difference in performance when the patchwork coping strategy was employed.

The nature of a shared collaborative task and/or interface directly impacts the style of interaction between collaborators. This will have a bearing on the choice of task for experimentation [153, 154]. A shared digital table task was used in the first Telethrone experiment and it suggests there might be utility in such supporting systems around the Telethrone, but it seems that this should be left for future research.

2.5.7 Beyond 2D Screens

Displays need not be limited to 2 dimensional screens and can be enhanced in various ways.

Stereoscopy allows an illusion of depth to be added to a 2D image by exploiting the stereo depth processing characteristics of the human vision system. This technical approach is not perfect as it does not fully recreate the convergence and focus expected by the eyes and brain [294]. There are multiple approaches to separating the left and right eye images, these primarily being active (where a signal selectively blanks the input to left then right eyes in synchronicity with the display), passive, where either selective spectrum or selective polarisation of light allow different portions of a display access to different eyes, or physical arrangements which present different displays (or slices of light as in lenticular systems) to different eyes.

These barrier stereoscopy / lenticular displays use vertical light barriers built into the display to create multiple discrete channels of display which are accessed by moving horizontally with respect to the display. In this way it is possible to generate either a left/right eye image pair for 'autostereoscopic' viewing, or with the addition of head tracking and small motors. With these techniques multiple viewpoint or an adaptive realtime viewpoint update can be presented without the glasses required for active or passive stereoscopic systems. Telethrone aims to support ad-hoc and should therefore eschew glasses.

Spatially Faithful Group

Hauber et al. combined videoconferencing, tabletop, and social presence analysis and tested the addition of 3D. They found a nuanced response when comparing 2D and 3D approaches to spatiality: 3D showed improved presence over 2D (chiefly through gaze support), while 2D demonstrated improved task performance because of task focus [133].

I3DVC reconstructs participants from multiple cameras and places them isotropically (spatially faithful) [164, 165]. The system uses a large projection screen, a custom table, and carefully defined seating positions. They discussed an "extended perception space" which used identical equipment in the remote spaces in a tightly coupled collaborative 'booth'. It employed head tracking and multi camera reconstruction alongside large screens built into the booth. This system exemplified the physical restrictions which are required to limit the problems of looking into another space through the screen. Towles et al. [102] demonstrated a similar system over a wide area network but achieved only limited resolution and frame rate with the technology of the day.

University of Southern California used a technically demanding set-up with 3D face scanning and an autostereoscopic 3D display to generate multiple 'face tracked' viewpoints [157]. This had the disadvantage of displaying a disembodied head.

MAJIC is an early comparable system to support small groups with life size spatially correct video, but without multiple viewpoints onto the remote collaborators it was a one to 'some' system rather than 'some' to one like the proposed Telethrone. Additionally users were rooted to defined locations [148, 231].

Multiview In order to reconnect directional cues of all kinds it is necessary for each party in the group to have a spatially correct view of the remote user which is particular for them. This requires a multi-view display, which has applications beyond telepresence but are used extensively in research which attempts to address these issues.
Nguyen and Canny demonstrated the 'Multiview' system [221]. Multiview is a spatially segmented system, that is, it presents different views to people standing in different locations simultaneously. They found similar task performance in trust tasks to face to face meetings, while a similar approach without spatial segmentation was seen to negatively impact performance.

In addition to spatial segmentation of viewpoints it is possible to isolate viewpoints in the time domain. Different tracked users can be presented with their individual view of a virtual scene for a few milliseconds per eye, before another viewpoint is shown to another user. Up to six such viewpoints are supported in the c1x6 system [178] Similarly MM+Space offered 4 Degree-Of-Freedom Kinetic Display to recreate Multiparty Conversation Spaces [237]

Robots, Shader Lamp, and Hybrid

Virtuality human representation extends beyond simple displays into robotic embodiments (which need not be humanoid [202]), shape mapped projection dubbed "shader lamps", and hybridisations of the two.

Uncanniness When employing simulation representations of humans it may be the case that there is an element of weirdness to some of these systems, especially those that currently represent a head without a body. Mori has demonstrated The Uncanny Valley [212] effect in which imperfect representations of humans elicit revulsion in certain observers. This provides a toolkit for inspecting potentially 'weird' representations, especially if they are 'eerie' and is testable through Mori's GODSPEED questionnaire. This may include the double image seen on the Telethrone and so the Uncanny Valley effect will inform the research [112, 207, 140].

With an improved analysis of the shape of the likeability curve estimated later showing a more nuanced response from respondents where anthropomorphism of characters demonstrated increased likeability even against a human baseline [24, 26].

A mismatch in the human realism of face and voice also produces an Uncanny Valley response [207].

However, there is a possibility that Mori's hypothesis may be too simplistic for practical everyday use in CG and robotics research since anthropomorphism can be ascribed to many and interdependent features such as movement and content of interaction [25].

Bartneck et al. also performed tests which suggest that the original Uncanny Valley assertions may be incorrect, and that it may be inappropriate to map human responses to human simulacrum to such a simplistic scale. They suggest that the measure has been a convenient 'escape route' for researchers [25]. Their suggestion that the measure should not hold back the development of more realistic robots holds less bearing for the main thrust of this telepresence research which seeks to capture issues with imperfect video representation rather than test the validity of an approximation.

Interestingly Ho et al. performed tests on a variety of facial representations using images. This is far closer to the core investigation. They found that facial performance is a 'double edged sword' with realism being important to robotic representations, but there also being a significant Uncanny Valley effect around 'eerie, creepy, and strange' which can be avoided by good design [141].

More humanlike representations exhibiting higher realism produce more positive social interactions when subjective measures are used [351] but not when objective measures are used. This suggests that questionnaires may be more important when assessing potential uncanniness.

Ho et al. also identified problems with the original GODSPEED indices used to measure Uncanny Valley effects. They proposed a new set of measures which they found to be generally valid [140], though they admit they were only tested with a single set of stimuli.

A far more objective method would be to measure user responses to humans, robots, and representations with functional near-infrared spectroscopy and while this has been attempted it is early exploratory research [312], an emotional response to 'eerie' was discovered.

Embodiment through robots Robots which carry a videoconference style screen showing a head can add mobility and this extends the available cues [2, 187, 328, 247, 176]. Interestingly Desai and Uhlik maintain that the overriding modality should be high quality audio [75].

Tsui et al. asked 96 participants to rate how personal and interactive they found interfaces to be. Interestingly they rated videoconferencing as both more personal and more interactive than telepresence robots, suggesting that there is a problem with the overall representation or embodiment [329].

Kristoffersson et al. applied the Networked Minds questionnaire to judge presence of a telepresence robot for participants with little or no experience of videoconferencing. Their results were encouraging, though they identified that the acuity of the audio channel needing improvement [175].

There are a very few lifelike robots which can be used for telepresence, and even these are judged to be uncanny [277]. This is only an issue for a human likeness since anthropomorphic proxies such as robots and toys perform well [212].

Physical & Hybrid embodiment Embodiment through hybridisation of real-time video and physical animatronic mannequins has been investigated as a way to bring the remote person into the space in a more convincing way [189, 190, 258]. These include telepresence robots [187, 277, 328], head in a jar implementations such as SphereAvatar [240, 244, 242] and BiReality [161], UCL's Gaze Preserving Situated Multi-View Telepresence System [243], or screen on a stick style representations [176].

Nagendran et al. present a 3D continuum of these systems into which they suggest all such systems can be rated from artificial to real on the three axes, shape, intelligence, and appearance [218].

Itoh et al. describe a 'face robot' to convey captured human emotion over a distance. It uses an 'average face' and actuators to manipulate feature points [151]. It seems that this is an outlier method for communication of facial affect but demonstrates that there are many development paths to a more tangible human display.

Shader lamps Projection mapping is a computational augmented projection technique where consideration of the relative positions and angles of complex surfaces allows the projection from single or multiple sources to augment the physical shapes onto which they appear. It was first considered by the Disney corporation in 1969 [252] and was given prominence by Raskar and Fuchs with "office of the future" [257] and later by Raskar and other researchers [194, 258]. It has since gained considerable commercial popularity in live entertainment [325].

Shader lamps [258] is the more formal academic designation for projection mapping. It is possible to use the technique alongside reconstruction to project onto a white facial mannequin. Researchers have attempted to use the technology for remote patient diagnostic, projecting onto styrofoam heads [264].

Bandyopadhyay et al. demonstrated [22] that it is possible to track objects and projection map [69] onto them in real time. This is beyond the scope of the proposed projection onto furniture since we wish to keep the system as simple as possible, but could be useful for shared tasks in the future work.

Lincoln et al. employed animatronic avatars which they projected with shader

lamps. This combination recreated facial expression and head movement though they were limited in speed and range of control of the remote head [190].
While shader lamps are an important and useful technology, there are limitations imposed by its use. In particular if a realtime video feed or reconstruction of a subject is used then that scanned subject must either remain still enough to be correctly mapped onto geometry on the remote side (useful for some virtual patients for instance [32], or else there must be a computational adjustment made for their changing position to make them appear static, or the projection surface must move to match their movement as in Lincoln et al. .

This technique was investigated and disregarded for Telethrone as elaborated on later.

Holography and Volumetric

Blanche et al. have done a great deal of research into holographic and volumetric displays using lasers, rotating surfaces, and light field technology [39, 321]. They are actively seeking to use their technologies for telepresence and their work is very interesting, but it is too technologically complex for the application space which Telethrone seeks to occupy.
Similarly Jones et al. “HeadSPIN” is a one-to-many 3D video teleconferencing system [157] which uses a rotating display to render the holographic head of a remote party. They achieve transmissible and usable framerate using structured light scanning of a remote collaborator as they view a 2D screen which they say shows a spatially correct view of the onlooking parties.
Eldes et al. used a rotating display to present multi-view autostereoscopic projected images to users [88].
Seelinder is an interesting system which uses parallax barriers to render a head which an onlooking viewer can walk around. The system uses 360 high resolution still images which means a new spatially segmented view of the head every 1°of arc. They claim the system is capable of playback of video and this head in a jar

multi-view system clearly has merit but is comparatively small, and as yet untested for telepresence [353].

These systems do not satisfy the requirement to render upper body for the viewers and are not situated (as described soon).

Reconstructed Viewpoint

Although it is possible to manipulate a 2D view of a remote participant to bring the camera, eye, and screen into correct alignment better results can be obtained by capturing a 3D representation of the remote user, then modifying that. Several groups have investigated stereo reconstruction of face and upper body, so-called immersive videoconferencing [15].

Multiple viewpoints from multiple cameras allow either playback from different cameras on a continuous basis, so-called free viewpoint TV [354]. This multiple camera approach can be extended with the use of light field to interpolate between real or virtual cameras [238, 5, 21], or else algorithmic generation and display of a polygonal mesh through photogrammetry can be undertaken. This can now be performed fast enough to compute a continuous surface from a video stream with little latency [66]. These multi camera systems aim to create geometry for a single virtual viewpoint (such as a face in video conferencing), which is correct to align eye gaze. This is sometimes called“3D video”, which is different to stereoscopic video used in cinema and commercial TV.

A depth map, point cloud, then polygon mesh can be calculated for one side of a person from two adjacent cameras which have slightly different views of a subject. This so-called ‘narrow baseline’ technique examines sub-pixel disparities between the images [172].

Multiple ‘narrow baseline’ pairs can create stereo geometry reliably from a tightly coupled pair of cameras which are integrated with geometry from another pair with an independent viewpoint [232, 63, 64].

Alternatively there is the so-called “wide baseline” approach in which a pair of cameras with a significant distance between them capture both sides of the face or body. Sadagic et al. demonstrated a system which brought 2 camera reconstructed [276] people into a desk based collaborative space. The user addressing these two tele-present collaborators could move around in their chair while being presented with both passive stereo depth cues, and parallax cues appropriate to the desk environment. The system's main limitation was the necessity for the user who addressed the two desk screens to wear both stereo glasses and a camera on their head. This made their system unbalanced. There are also problems with geometry and texture management using a wide baseline approach since the cameras are resolving for different lighting and views of the scene.

Pan, Steptoe and Steed found similar results with their spherical display with a decrease in trust toward avatar mediated conversation when viewing 2D displays at oblique angles [244].

Microsoft Kinect sensors demonstrate effectiveness in real-time scanning with between 1 and 5 deployed in research systems [205]. Multiple Kinect v1 sensors interfere with one another, and though this problem is partially addressed with v2 hardware there are constraints with temperature drift causing frequency shifts which allow interference to creep in. There are workarounds for this issue but no completely reliable implementation.

For some time it has been possible to create geometric models of the human form using a technique called "shape from silhouette" [6, 249, 203, 302, 101, 27, 186, 122, 344, 100, 96, 62, 303]

Similar effect can also be accomplished for the human shape using depth cameras such as the Microsoft Kinect v2 [199], and indeed this seems now to be the prevailing method. Sensors in this vein continue to improve.

Structured light systems provide still another technique, back as early as the 1980's [145], and as recently as the Microsoft Kinect v1 sensor. Deformations in projected visible or infrared patterns can be compared to the known baseline and shape reverse engineered from the differences. 3D-Board is an interesting example since two standing collaborators can interact through touch with a digital whiteboard which also forms the perceived barrier between them [356].

Rendering and telepresence Cooke et al. investigate the difference between so-called narrow and wide baseline camera capture for dealing with the more complicated elements of reconstructed geometry such as hands which gesture. They maintain that multiple stereo camera pairs (each narrow baseline) arranged as a wide baseline system, with post processing to remove artifacts is of most use [65]. This is likely too complex for Telethrone.

Petit et al. adopt a wide baseline system with conventional green screen background segmentation and get good 3D video results. Because they are using a chroma based technique for their silhouette system they cannot exclude objects in the scene by 'training' the system [250].

Kuster et al. use a depth based approach and have a compelling demonstrator which uses what they call an "anisotropic transparent background back projection foil". They rear project a Kinect reconstructed head and upper body in stereoscopic 3D onto a transparent film suspended so as to appear to bring the remote person to the edge of a table. This brings the 3D avatar into a space at 15 frames per second across significant network distances but is again peer to peer[179].

Similarly Maimone and Fuchs use large tiled displays and Kinect sensors but attempt to 'merge' two spaces by scanning both the person and the remote space to create perspective correct viewing through a virtual window into the other space. Without multi-view this system is one to one, although they demonstrate multiple people in the background of the space [200].

The Fusion4D system successfully captured complex and challenging scenes such as dancers with flowing clothing for playback from arbitrary angles [78].

withyou is an experimental capture and playback system which uses the Octave multi-modal suite. This system uses wide baseline single cameras and CPU based shape from silhouette reconstruction alongside a novel network transport system to send a full 3D video polygonal hull to another rendering location [273]. Previous tests on the system suggested that the capture and playback made it possible to judge the eye gaze of the reconstructed subjects to within normal social/biological limits [272].

A particularly well developed example is the blue-c system which enables telepresence through multi-camera 3D video. The users at multiple locations are scanned and transmitted as a point cloud captured during projector blanking frames [127].

The system is flexible enough to use either CAVE style surround or a projected screen style arrangement.
ViPiD is similar to the Octave capture system used to support Telethrone. The ViPiD system interestingly uses Chromatte in its intended role to improve segmentation for their capture system [170].
Schreer at al demonstrate a capable multi camera capture system suitable for small group telepresence, though its technical overhead is high [282]. They discuss the need for multi-view capable screens only insomuch as they agree one must be developed for the system.
Maimone at al demonstrate a system which has similar effect to the Telethrone but uses a HMD system to project a reconstructed person into the correct seat at a table in augmented reality [201].

Immersive Collaborative Virtual Environments

Simple online virtual environments with an external viewpoint and many avatars interacting online generated much hype in the early days as a potential means to support group meetings, even back as far as the VRML standard [98]. SecondLife in particular generated significant interest. Erickson et al. went so far as to say that a virtual conference hosted in SecondLife was 'fairly successful' [90]. For whatever reason however these systems have fallen out of favour with the public and research communities. It might be that the external viewpoint perspective, or the clunky internal viewpoint are a barrier to communication.
In contrast 'immersive' collaborative virtual environments place the user physically into the virtual scene. These systems are less scalable than online virtual communities which can take better advantage of distributed resources [126, 31].
Goebbels et al. designed a taxonomy for what they termed simply CVE's. In their design they provide an excellent high level description of the ICVE as spaces which provide “distributed collaborative teams with a virtual space where they could meet as if face-to-face, co-exist and collaborate while sharing and manipulating virtual data”, crucially designed in a way to 'disburden' the users' senses [118] and

reduce fragmentation of the shared perceptual environment [265]. The ability to collaborate in such systems extends even to closely coupled physical tasks [271]. There are various technology systems which demonstrate heightened immersion and presence in a virtual environment while allowing interaction between parties who may or may not be physically co-located [217]. Attempts at bringing people together in VR extend back to the early days of the technology and systems such as DIVE [30].

Bailenson et al. noted that while increased realism of such avatars increases co-presence it decreases self disclosure [20]. Such systems seem to compromise social interaction even as the realism increases. Avatar representations increased in fidelity and eventually it became possible to share avatar representations of participants wearing body tracking [281] and eye tracking equipment [105, 106]. This enabled tracked, viewpoint independent interaction with reconnection of eye gaze cues within VR in a spatially faithful way [266, 216, 307, 308, 306].

It is also possible (at least in research implementations) to fully reconstruct people as 3D models, and connect more than two users together utilising life-size multi-view video [273, 122, 77, 127, 29]. These 3D video representations of users can share a virtual space in which spatial and directional information are maintained in a natural way [273, 349].

Fairchild et al. have developed a system for ICVE and desktop based on earlier work by Duckworth et al. [82] which is adapted for use in the Telethrone research [93].

Eye tracking in ICVE More recently, advanced eye tracking technologies allow the recording and replaying of accurate eye gaze information[59, 308, 307] alongside information about pupil dilation toward determination of honesty and social presence [305]. Heldal found that collaborative tasks manifested fewer disturbances due to “misunderstandings of reference or action” when using more immersive systems [138].

Situated Displays

Between the complexity of ICVEs and the more ubiquitous screen based VC technology there now exist so-called situated displays.
Conversation does not exist in isolation, but rather in the context of semiotic resources. Participants make constant reference to the surrounding assets through mutual orientation, gesture, and diverse cues [120]. So called situated displays seek to embed the represented participant within the spatial and contextual framework of the conversation such that the referential cues are better supported. This has many implications, but chief amongst these is support for a spatially faithful conversational environment supportive of gaze [243, 240, 242, 355].
Such displays place a representation of the remote user into a space, theoretically allowing all participants to physically interact with the 'contextual configurations' around them [120]. This is a relatively new field of research. These could include the aforementioned telepresence robots [187, 277, 328], head in a jar implementations such as SphereAvatar [240, 244, 242], and Gaze Preserving Situated Multi-View Telepresence System [243]. Sphereavatar [240] demonstrates that there are problems with accurate mapping, distortion, and projection, and movement of the captured actors outside of very tight bounds.
Telehuman brings the whole body of a standing remote user into a space via a cylindrical display with a single tracked observer viewpoint [167] and is a very useful comparative system for Telethrone as detailed later.

Retro-reflective Projective Technology (RPT)

Retro-reflective materials such as Chromatte(tm) cloth reflect light back along the angle of incidence. An everyday application of such material is high visibility jackets.
Inami et al. described the first use of RPT in 2000 with their visuo-haptic display [149]. This system used head mounted projectors to augment RPT shapes in the real world. Krum et al. describe the REFLCT system [177] which uses large

retro-reflective surfaces to “provide[s] users with a personal, perspective correct view of virtual elements that can be used to present social interactions with virtual humans”. They use helmet mounted projectors and describe a military training application in a large volume which allows faithful transmission of attention and gestures from the virtual to the real. They also briefly describe augmenting a facial mannequin by projecting onto RPT adhered to the surface. They point out that the optical characteristics of the material maintain polarisation and so could support passive stereoscopy. This is an important finding for future research opportunities for Telethrone but it is more important at this stage to maintain unencumbered interaction.

Tachi describes an augmented reality system where a helmet mounted projector places an image of a remote human onto a robot. This is a system they describe as Teleexistance RPT [316, 317]. This system is the closest analogy to the Telethrone except that it demonstrates a single user, wearing a head mounted projector, viewing only a head which is captured from a single viewpoint. It is however enabled for motion by means of robotics under the Chromatte cloth like the hybrid robotic systems described earlier.

Hua et al. demonstrated elements of ’SCAPE’, a tracked head mounted projection system surrounded by RPT cloth which could surround multiple physical users in a shared immersive experience [147]. Hypothetically this system could be extended to include what they term ’passive remote users’ projected into the views of the co-located and immersed users. The set-up however is quite complex, and involves wearing headgear, so is again less suitable for informal meetings.

Surman et al. discuss a potential development of their retroreflection based autostereoscopic display (which head tracks a single user to present a depth enabled stereo pair). They suggest that multiple users could address a large screen, but think that their stereopsis would break down and the system would be extremely challenging [313].

Although less pertinent to this research there are other interesting deployments using RPT. Darken et al. and later Hahn in his PhD thesis in the same group describe a novel system which mounts a chromakey LED light ring and camera on a HMD [70]. The camera takes in a mocked up physical cockpit and windows which are made of Chromatte cloth. This video image is very easy to segment to replace the green-screen windows with a simulated external view. The pilot trainee sees their own hands and the cockpit instrument panels (albeit in monoscopic video), alongside the generated external view. This augmented reality is much cheaper to deploy than a full simulator set-up.

Furniture as a Mediating Display

Paul Sermon experimented with projection onto beds in his telematic dreaming work, research which while not truly situated certainly informed the idea of Telethrone [289].
Room2room from Microsoft Labs [248] demonstrates the utility of projecting onto furniture by building upon their previous automatic projection mapping research [158]. It is seen in Microsoft's Room2Room which was discovered later in the research and is pictured in the final chapter. The system uses a Kinect 2.5D camera to capture a remote participant, and an overhead projector combined with a Kinect to projection map a viewpoint correct image of the telepresent person onto a complex surface such as a chair. However, this is a single static view-point.

Augmented Reality Collaboration

Lehment at al described what they term a "consensus reality" for head mounted AR in which they compute the correct locations for remote participants as 3D video representations overlaid on the real [188].

2.5.8 Network Issues

Bradner & Mark established that latency matters in remote telecollaboration [48], while an extensive body of work notes that system latency impacts key factors such as interruptability [16]. Gergle et al. found that remote collaborating pairs of people were tolerant of delays in visual feedback up to 939ms, after which shared task performance suffered considerably [114]. Overall it is clear that for communication systems, especially those with potential for shared or collaborative task, those aspects of the system latency which can be controlled and/or minimised should be.

Early research around COVEN and DIVE described the benefits of multicast for multiparty collaborative technology systems. While multicast still has advantages in latency critical set-ups the bandwidth/throughput has to a degree 'caught up', while the technical demands of a wide area multicast network remain the same. The DIVE system has a standalone WAN application layer network transport called DIVEBONE which attempted to address this issue [125]. Steed et al. describe a hybrid multi/uni cast system which allows local multicast with unicast on the WAN [124].

Bradner used social impact theory to suggest it is important to either say the distally located person is geographically remote to emulate real systems, or else explain they are in the next room to exclude this factor [48], while an extensive body of work notes that system latency impacts key factors such as interruptibility [16].

Gasparello et al. [108] then later Moore et al. [210] worked within the wider research group on novel network transport systems for synchronised 2D and 3D video information to support telepresence reconstruction in the later withyou system [273]. Elements of this research are used in the Telethrone system. When Lamboray extended their blue-c 3D video system (an implementation similar to Telethrone) they considered that latency of up to 200ms was acceptable so long as jitter (changes in latency) remained low [245, 181].

2.6 Comparing Key Systems and their Affordances

Affordances are properties of the environment that offer actions to appropriate organisms [115]. In this research the Gibson definition of 'an action possibility' is adopted rather than the more restrictive Norman definition of 'a perceived suggestion' [223]. McGrenere and Ho sum up the difference in definitions of affordances

Table 2.1: Affordances of some key systems and technologies. An expanded version of this table in the concluding chapter references the papers.

Systems	Joint Eye Gaze	Viewpoint independent / Reconstructed	3D stereo	Multiple independent viewpoints	Natural Setting	Weird / Uncanny	Commodity components	Able to move around
Skype Facetime	No	No*	No	No	Somewhat	Somewhat	Yes	No
Realpresence Centro	Some	No	No	Sort of**	Yes	No	No	Yes
Cisco TP Rooms	Some	No	No	No	Yes	No	No	No
Sphere Avatar	Yes	Yes	No	Yes	No	Yes	No	Yes
Telehuman	Yes	Yes	Yes	No	Some***	No	Some	Yes
Immersive Group to Group	Yes	Yes	Yes	Yes	No	No	No	Yes
Withyou & blue-c	Yes	Yes	Yes	No	No	No	No	Yes
Tripleview	Yes	Yes	Yes	Yes	Yes	No	No	No

by considering a hidden door in a paneled room [204]. In Gibson's definition the existence of the door (or parallax cues, or stereoscopy, or eye gaze for instance in this research) does not need to be explicitly known to the user, just its existence is enough. In Norman's view the door would need to have a handle to suggest the action to the user. This is a less relevant HCI definition for this research.

Triple-view is an important system for comparison in the first Telethrone experiment [54]. It affords correct spatial alignment through the use of an interpolated camera pair per screen, providing three bi-directional channels which preserve the directionality of the user's gaze. Importantly this system still uses screens which show backgrounds from the remote space resulting in the feeling of looking into another space.

The more technically demanding Immersive Group-to-Group [29] wall places the reconstructed remote users in a bland virtual space which can be set to match the surrounding walls and draws less attention to the elements which are not the remote collaborator. It has more in common with the later Telethrone experiments.

The previously mentioned TeleHuman supports 360°motion parallax as the viewer moves around a human scale cylinder which is internally projected. This is optionally stereoscopic 3D [167].

Table 2.1 shows the most important systems identified at the time of writing for the literature review, along with what affordances they support.

* New sensor technologies in mobile phones do now in principle allow facial reconstruction

** In Centro four screens face outward

*** The Telehuman cylinder would look appropriate in for instance retail settings

2.7 Support for Less Formal?

Outside of the sphere of simple webcam systems such as Microsoft Skype Google Hangouts there has been little attention investing in advanced telecommunication tools for informal settings.
Slovak et al. describe a group-to-group videoconferencing system called GColl in which they highlight mutual gaze support and modest technical requirements [300]. GColl is a monitor based system which seeks to align mutual gaze by rendering to a small window within 5°of the mounted camera. It seems to address the right problems but is a questionable development over and above a careful videoconferencing set-up. De Greef and Ijsselsteijn describe a system which provides collaborative working tools alongside video and audio conferencing for the home [71]. Judge and Neustaedter examined how 21 families used videoconferencing in a home setting and found that requirement for planning and availability for using the tools inhibited their use. They suggest an 'always on' system might better mediate availability [162]. This matches the assertion from the BBC which informed the Telethrone concept, and may be similarly important for all informal business meetings.
Interestingly Lottridge et al. identified that it is the empty reflective moments during mundane activity which might benefit most from the ability to strike up intimate conversations between separated couples [193]. Whether and how much this supposition is analogous to the so-called 'water cooler meeting' moments in business remains an open question but is perhaps worth bearing in mind.
There is, of course, Room2Room which provides a far better informal home experience, but is inherently one to one [248]. The same is true of 'Holoportation',

a reconstructed and tracked 3D video based system designed by Microsoft for head mounted display and seemingly aimed at the home market [234]. Similarly utilising Microsoft hardware, but on a tight budget, HomeProxy uses a simple combination of consumer devices to render a reconstructed remote partner to a cabinet style display on a desk [320]. Users reported that they appreciated the look and feel of the HomeProxy system which is designed to fit in sympathetically to home furnishings.

Gaver explains that ”the space created by audio-video technologies is significantly different from spaces as found in hallways, offices or meeting rooms”, conveying a limited subset of visual and auditory information, preventing movement and exploration, and is often arbitrary and discontinuous. He says that ”these properties shape the possibilities.. for collaboration.” [110].

In a work setting Fish et al. said (in 1993) that “informal communication cuts across organisational boundaries and often happens spontaneously.” They reference research which suggests that informal communication is more frequent, expressive, and interactive [99]. Fish went on to look at early telepresence systems which might support informal, but by any standard these systems are too clunky to be deemed capable of such. Although this has been recognised as an issue for decades it could nonetheless seem that less formal business meetings are of insufficient utility to attract the technology and research to better support them. Computer Supported Collaborative Working (CSCW) is a well known acronym but there seems to be no single point of entry to research into less formal systems. This could be a function of the technology being better suited to deployment in a controlled formal setting.

The paper “Ad hoc versus established groups in an electronic meeting system environment” turns out to be using the other definition of ad-hoc meetings which is meetings convened from random members with no past and no future together. The study is also dated being from 1990, but still provides insight. In the paper Dennis et al. reviewed what little research existed at the time noted that meeting outcomes varied between studies as to whether ad-hoc or more established and formal meetings were more productive [73].

Fayard and Weeks examined the affordances of informal interaction (indeed exactly the ‘water cooler meetings’ that the BBC were talking about) [95]. They identified:

proximity, privacy, and legitimacy as key affordances which should be supported, and highlighted: functional centrality, semi-enclosure, reciprocal visibility, easy access and egress, multiple shared resources as important further affordances of the environment. The affordance of legitimacy is interesting. This means that the location of the informal meeting should support the feeling that it is legitimate to be there and strike up a conversation. This seems to be something which could and should be translated to the Telethrone system and as such an analysis of the system as it stands will be made against these affordances in the discussion chapter.
A potential model example is explored in Horgan's book 'Excellence by design: Transforming workplace and work practice'. 'The LX common' was a semi communal space for informal meetings. It was semi-enclosed to give some privacy, but located on a busy thoroughfare and housed the photocopiers, printers, shared reference material, kitchen etc. People who passed felt free to listen and occasionally join in. Three rules evolved from the research; Traffic through the common was acceptable at any time, anyone was free to join any meeting in the common, anyone was free to leave any meeting in the common at any time [144].
Supporting spatial aspects such as mutual gaze or 'full gaze awareness" [209] in this way normally demands large purpose built installations which poorly support casual or ad hoc meeting paradigms [283, 349]. It seems clear that there is a justification for development of simple affordable technologies which better mediates communication over distance for less formal situations.

2.7.1 Support for Dynamic Meetings

Voida et al. discuss natural sub grouping in partially distributed teams [343]. These 'fault lines' within meetings are a natural component of more informal meeting structure with groups potentially forming and dispersing naturally over time.
Greenburg and Jerald found in group interaction with trainers that members often changed their seating position to facilitate discussion goals [123]. This may not be true of all group meetings but it is under supported in telepresence systems.
This dynamic reorganisation of meetings is potentially an affordance for ad-hoc and casual meeting styles. It demands the ability to reorganise chairs such that

they still have faithful views onto remote participants. There is no support for this found in literature.

2.8 Theoretical Framework

2.8.1 Problem Statement

It is clear that there are multiple factors which contribute to successful human-human communication. These factors remain important in telecommunication supported by technology, and are variously supported, unsupported, or modified by particular technologies.

Of particular importance is interpersonal gaze [61, 168, 92], and gaze is an excellent dependant variable for experimentation. Non-verbal cues are also important across multiple modalities of sight, sound [236], and position of interlocutors [166], extending to the whole body [168, 222].

The Mona Lisa effect demands that some novel system which allows spatial signals to be maintained must be employed [214].

While formal meeting paradigms are supported to an extent by commercially deployed systems this does not suit all meeting styles. Such systems are generally booked well in advance and so meetings tend toward a formal structure. These meetings seem to demand many smaller supporting meetings between parties or groups of parties.

The problem is a requirement for a system or work toward a system which connects informal spaces, for groups, with an always on technology, which allows dynamism, connection of natural non vocal cues, without too much encumbering technology overhead.

2.8.2 Core Assumptions

Figure 2.5 shows the interlocking relationships between baseline communication where the participants are present, and technology which attempts to support across distance.

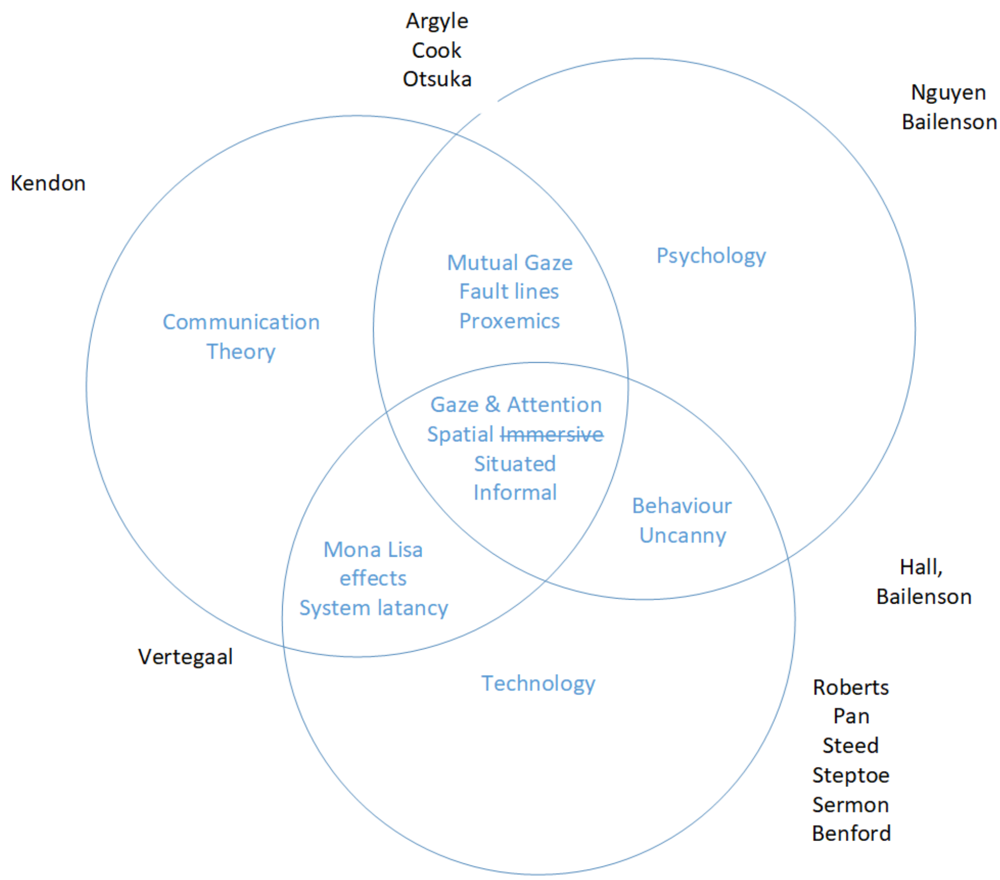


Figure 2.5: The Venn diagram shows areas of research which have been identified in blue. These interlock and overlap as shown. The most relevant identified researchers from the literature are shown in black close to the fields of study which they represent. This diagram is a view of the core assumptions for the research, with the most important fields at the centre.

Of most interest to this research is the centre of the Venn where meeting styles which are less formal, and perhaps dynamic, may occur. Looking at these items one by one gives us our core assumptions.

1. Gaze
   Gaze is broadly agreed to be highly important for mediating flow. Mutual gaze is a rich emotional channel. The research must consider gaze. All of the researchers listed around the Venn have at some point engaged with this topic.

2. Attention
   The non-verbal communication channel employed in 'attention' is assumed based upon the literature to be critical to smoothly leaving and entering a fast flowing conversation where concentration around a defined problem may be high (gesturing to a chair for instance). Again, all of the listed researchers have made reference to attention in their work.

3. Spatial (immersive)
   Support for spatiality is important in a group setting so that directional non-verbal cues can find their target. The topic of spatial relationships between interlocutors cuts across all of the researchers, but this is not true of immersion. Immersion in a shared virtuality can certainly support the underlying requirements spatial, but the technical infrastructure required is out of scope (so this is struck through on the diagram). Roberts and Steed are the main expertise referenced even though this element is not expanded in the research.

4. Situated
   Situated displays are those which are appropriate for their surrounding context, in this case the informal meeting. Roberts, Pan, Steed and Steptoe seem the most relevant researchers in these technology spaces.

5. Informal
   Based on the literature proxemics is believed to be relevant in a meeting where subgroups can be instantiated and destroyed as the meeting evolves, and those where people can be invited in from outside the physical bounds of the meeting (informal spaces). Hall is the best source for this work. If it is assumed that people may come and go, and subgroups may be convened then Sermon and Benford are the best references through their work blending

real and virtual spaces. This may be more consistent with less organised meetings such as those convened on demand (ad-hoc).

2.8.3 Peripheral Assumptions

Surrounding the centre of the Venn are additional relevant topics from social science branches of theory

From verbal communication

It is assumed that the directionality of sound is important [8], and this will be engineered into the experimental design. It is assumed that movement of the lips is an indicator and this is tied to latency and frame rate in the vision system.

From non-verbal communication

It is assumed that eye gaze is of high importance, and that this information channel is supported by head gaze and body torque to a high degree. It is further assumed that mutual eye gaze is of less relevance in a multi party meeting where there is a common focus for attention but can be significant for turn passing. It is assumed that upper body framing and support for transmission of micro and macro gesturing is important for signaling attention in the broader group, and for message passing in subgroups.

2.8.4 Revisiting the Literature Review

The iterative process of revisiting the literature throughout the methodology threw up many more important historical papers, and some new research which came out after the conclusion in presentation of this review at Internal Evaluation. This newer body of discovered work is explored in the concluding chapter in ”Revisiting the literature”.

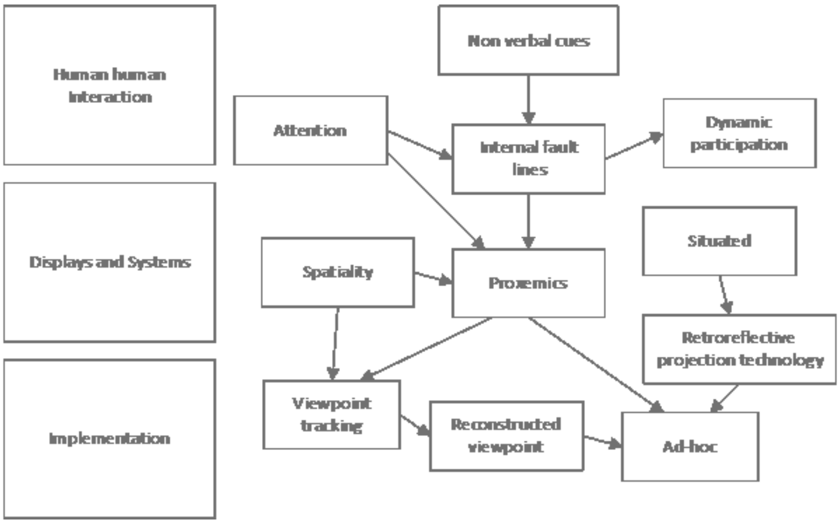


Figure 2.6: The conceptual framework shows a map of the connected concepts revealed during the literature review, and roughly how they relate to one another. The web of connected concepts on the right hand side moves roughly top left to bottom right as the attention focuses increasingly down toward the more pertinent specialisms around the research. The left hand side shows the broad flow of the focus of the research from interpersonal communication in the real (top) to the implementation of Telethrone (bottom). There is a broad horizontal relationship between the left hand side and the right hand side, though blurred distinctions mean it is not appropriate to delineate with visual boundaries. In general the diagram should be considered top left to bottom right as the research developed through the concepts.

Chapter 3

Methodology

3.1 Chapter Overview

This chapter details the aim of the research, the objectives, research questions, and hypotheses. Following this, the research methods and how they interacted are outlined. The experiments and development of the research then provide the framework for the remaining chapters.

3.2 Introduction

The research examines variants of small group interaction with a view to building support for more informal styles of meeting mediated by technology.

This is perhaps an unusual PhD thesis in that it presents a series of novel integrations in the IT domain, alongside experimental design and analysis more normally associated with the field of psychology [40]. This reflects changes in the way science is being conducted more generally in VR, and telepresence, as explored in the literature review.

At it's core, this PhD is about integrating novel components into systems and performing experimental analysis using subjects. This can be further broken down as follows:

- Building things.
  - Prototyping
  - Testing components
  - Understanding characteristics of the components
  - Refining
  - Testing systems
  - Understanding characteristics of the systems
- Evaluating systems in isolation.
  - Careful experiment design
  - Quantitative measurement
- Integrating into broader research.
  - Adapting existing systems to the approach
  - Understanding the existing system in the new context

- Evaluating them in concert with other research
  - Careful experiment design
  - Quantitative measurement

Prototyping, refining design, testing components, and experimentation was an iterative process.

3.3 Aims and Objectives

3.3.1 Aim

The intention of the presented research is to develop and examine a novel multi-view telepresence system based around Chromatte retro-reflective cloth. This computer supported collaborative working tool may better align to the informal meeting paradigm. The research is an iterative step toward support for natural and spatially flexible conversations in small groups across multiple sites, integrating this into the broader research undertaken by the group.

Experimental evaluation of triadic and small group interaction, mediated in part by the Telethrone underpins the contribution.

3.3.2 Objectives

Objectives are zero indexed since ‘0’ was a preamble to the research proper, and this is reflected throughout the research.

- **O0:** With a view to building a novel tele-presence system which can support ad-hoc meetings; test the physical properties of Chromatte cloth, cameras, and projection systems to determine suitability in experimental context for use in spatial segmentation. Trial techniques such projection mapping and adaptive software masking which might more accurately apply the projection to the complex surface of the chair. The testing must give confidence that it is possible for observers to perceive a view of a projected remote person that is correct for each of them separately.

- **O1:** Test if a remote participant brought into the space by the system is for some reason excluded from triadic poker play. Primarily seeking evidence for (statistically) significantly fewer looks and duration of looking toward the remote player, with additional examination of social engagement through questionnaire feedback.
- **O2:** Explore how the system might scale, and how it might support novel subgroups within a meeting.
- **O3:** Integrate previous research (3D video reconstruction) which can present a representation of a remote participant such that the observer can walk around the projected image, continuously seeing a viewpoint which is generated for their position. Test this experimentally. Test this reconstructed viewpoint system in the context of a small meeting experiment.

3.3.3 Research Questions

As a reminder of the research questions seen in the introduction the following were generated over the course of the research

- **Q0:** Do the experimental set-up pre-requisites exist? i.e. Does light reflected from Chromatte cloth fall off in intensity as the angle subtended from the axis of projection increases. Is this in-line with the manufacturers datasheet when the material is used outside of their guidelines as in the proposed Telethrone. Are there additional subjective concerns?
- **Q1:** Does the Telethrone operate as an effective situated display (judged by questionnaire), which can show multiple views of a remote participant

without excluding them from the attention (judged empirically) of the participant during a three-way conversation? Structured poker play designed to minimise but not preclude eye contact should be supported between three parties with the Telethrone. This will be examined statistically by the number of times the participant looks at the co-located player against the remote player.

- **Q2:** In the course of investigating the social presence capabilities of the system in the first experiment a new line of inquiry was explored. What are the theoretical limits of the system, especially when moving chairs around dynamically in the meeting?
- **Q3:** Addressing shortcomings in the capability of the system to connect mutual eye gaze required 3D capture of the remote participant. To what extent does the new system, which allows the observer to move around the space, allow detection of eye gaze? Can it enable a group of five people to reliably detect non-verbal cues transmitted by the remote participant, in a simulated ideal sized meeting?

3.4 Summary of Research Process

Initial investigation was deliberately broad and iterated down while informing the direction of investigation. Testing of the system was exclusively quantitative.

3.4.1 Common Approaches

The presented research is a synergy of social science, and system integration aspects of computer science. The social science elements are presented from an HCI (human computer interfaces) perspective, focusing attention chiefly on technology-mediated social interaction and using tools developed in that field.

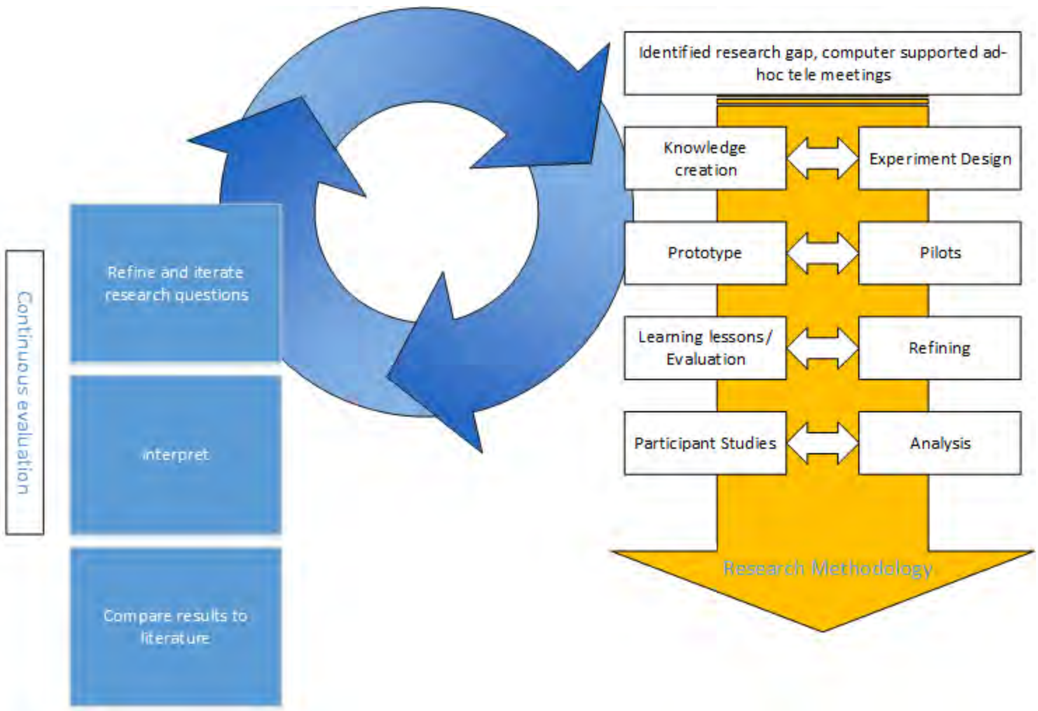


Figure 3.1: The right hand side shows the workflow for an experiment to test a hypothesis. This can be seen evaluated later in 'swim lanes' in the discussion at the end of the thesis. The method is iterated to generate results from the research questions, and this feeds back into the overall research framework on the left. The overarching research focus is checked against the framework and new literature to inform the next experiment.

Choosing a Research Approach

Investigation of the research approaches of others in HCI and telepresence, VR, augmented, and mixed reality yielded a variety of commonalities and differences. It was necessary to synthesise the available research tools into a programme which would best fit the Telethrone research. The following initial structure of tools was chosen:

1. Literature review

2. Feasibility and component testing

3. Concept design
4. Pilot
5. User experiments
6. Quantitative analysis
7. Comparison against literature

Quantitative Tools

Effective interpersonal communication is notoriously difficult to study since to gauge when 'enough' is provided by a technology for it to be deemed a success, is subjective to the individual using the system, furthermore the general consensus of synthesised opinion changes over time as the support provided by the systems improve with technology. This is not an issue peculiar to telepresence research, but the fast pace of change in interpersonal communication sometimes seems to outstrip the ability of tools to measure the improvements. Quantitative questionnaires which attempt to judge levels of connection across technology boundaries are losing favour with noted researchers in the field, yet they remain as pragmatic tools until sufficient study replaces them.
Decisions about quantitative analysis are explored in detail later but from the beginning manual encoding from video was chosen for gaze analysis, while positional analysis of participants was derived from 3D tracking data. The Networked Minds social presence questions and GODSPEED questionnaires identified earlier were chosen as the primary quantitative interrogation tools following conversations with researchers in the field, and appraisal of the literature.

3.4.2 Literature Review

Literature review is critical for grounded research, ensuring that the investigation is appropriately nested in a broader continuum of research, without merely duplicating

previous efforts [47]. Consistent investigation in the fields revealed new avenues for critical analysis, which informed the direction of the research, which in turn led to additional judgement about which literature to include.
Key researchers and teams, conferences, and journals were identified, and further material sought through forward and backward referencing. Backward referencing gave context to the state of the art at that time, while forward referencing showed further refinements.
Key outlets for literature searches included Google Scholar, ACM, Web of Science, UoS library, and Research Gate.

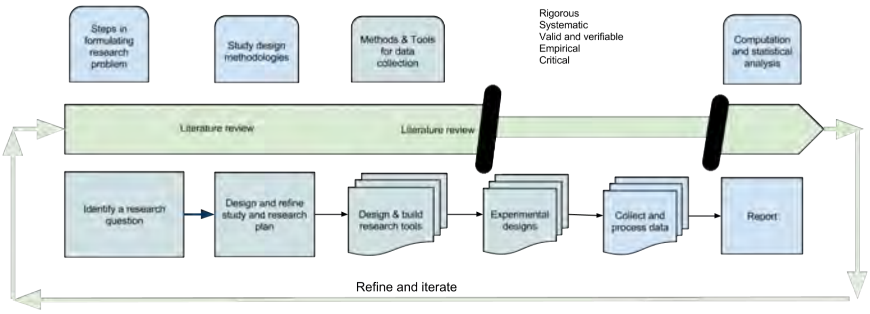


Figure 3.2: Viewed in another way the research methodology wraps around the literature review, with each element of the iterated methodology returning to the literature throughout. The green ‘spine’ of the literature constantly informs. In a way this is separate from the formal literature review at the start of the research, and is a body of work in itself. This delineation will be revisited later.

The review aimed to gain a high level view of current understanding of channels of interpersonal conversation, and a detailed view of attempts to duplicate these channels over a technology boundary. This contextualises attempts to develop a novel situated display which isn’t ‘uncanny’.
Initial focus is placed on human-human interaction with special attention placed on less formal / more dynamic conversation. It was important to understand the baseline condition of interaction between people enough to roughly delimit what the technology might be expected to recreate, but this knowledge proved to be tempered and constrained by subtle variations in behaviour as a result of the use of

technology.
Investigation draws from communication theory, and the attention and behaviour elements of psychology research. Technology-mediated communication research which relies on these social science measures is investigated and compared.
The review was then broadened to include detailed investigation of technologies which mediate conversation, and associated investigatory tools.
Building upon this base there is a critique of current telepresence research systems with reference to their ability to address the requirements of ad-hoc communication. Any gaps or shortcomings are isolated and highlighted toward refining the area of research for the PhD.
The search methodology was loosely broken into three strands which pragmatically responded to the demands of the research. These are interdependent investigations and deliberately not completely delineated. The first two were begun concurrently at the start of the research, with the third added early on during implementation.

1. Knowledge pertaining to verbal and nonverbal cues, interpersonal attention, and behaviour in human-human communication. Both face to face and technology-mediated are considered.

2. Telepresence / video conferencing research and techniques. Commercial systems are included as a reference to the state of the art. Consideration of the social science elements previously considered is reinforced but with more focus on the technology. This element is further refined into applicable research systems which address the more specific challenges such as mutual eye gaze and spatiality which are outside the purview of conventional video conferencing.

3. Concurrently with the execution of the first experiment, in response to a growing suspicion that the offset of the camera might be a problem, reconstruction systems, such as the visual hull which was finally employed were investigated. These allow manipulation of the perceived virtual camera point such that the tele-present subject is perceived to be looking directly into the camera rather than above or below, and additionally frees them from a defined position, allowing free movement around the remote subject.

Knowledge pertaining to psychology, (the study of mental functions and behaviour, interpersonal attention, and communication) formed the first strand. This is a well established field. Much of the foundation literature is available in its original form in seminal books from the 1960's and 70's with one reference going back to the 1800's.

Google Scholar searches exposed key players in the field with much of the relevant material boiling down to three volumes [9, 13, 130]. Additional and more recent pertinent literature toward interpersonal psychology was easily identified through analysis of citations in the current telepresence field. Particular attention was paid to factors which modify typical eye behaviour such as group dynamics, age, [299], sex [33, 45, 142], ethnicity [12]; which may better constrain experiments by controlling variables or adjusting independent variables appropriately.

Keywords for this element concentrated on physiometrics, non-verbal cues, gaze, mutual gaze, proxemics, attention, body, torque, arousal.

The second strand is a study of telepresence research and techniques, including commercially available systems. The focus was a rigorous examination of the well cited researchers and teams in telepresence publications. Current state of the art is heavily cross cited and further makes much reference to the underlying psychological elements covered in the first section. Keyword searching within Athens, Google Scholar, and the library's SOLAR system exposed a broad body of relevant and heavily cited literature. Consideration of relative prevalence of citation by others highlighted the best regarded groups and individuals, and a prioritisation of newer works (roughly the past ten years) further honed the selection. This was followed by an in depth reverse citation search adding context and exposing additional themes for investigation, while a forward citation search of the key papers added confidence in the accuracy of the conclusions.

Researchgate proved invaluable in obtaining surveys and material directly from researchers when it proved impossible to access them through other means.

Although some attempts were made to parse collections of proceedings such as CSCW and CHI it seems that a more catholic approach using the new meta search engines was far more productive. Grey literature such as New Scientist and online blogs provided some interesting additional steer in terms of the direction of the developing technologies and cultural drivers. Examination of the currently avail-

able commercial offerings such as CISCO, Lifesize, Tandberg, IOCOM, Google Hangouts, and Skype provided a current consumer and business level technology snapshot.
Keywords for this element included ICVE, telepresence, mutual gaze, collaborative virtual environments, teleconferencing, video conferencing, (co)presence, immersion, multi-view, RPT, situated displays, and uncanniness.
The third area of examination is a brief overview of technologies pertinent to 3D reconstruction and communication through media which use such technologies. This topic has been heavily investigated by the wider research group in recent years so obtaining a body of papers around this subject was primarily a matter of parsing these works.
Keywords included; shape by silhouette, 3D reconstruction, camera reconstruction, depth field reconstruction, visual hull, structured light 3D, lenticular, and lightfield. Figure 3.2 shows the literature review running concurrently throughout the research methodology.

3.4.3 Inclusion & Exclusion Criteria

Eligibility criteria for inclusion in the research study included renowned author expertise, high citation count, more applicable outlets, journals, and/or conferences, and outstanding or seminal work in the field(s). The disparate fields involved in the research meant that capture was necessarily reliant on keyword searching to a degree. Exclusion criteria focused mainly on low citation documents, grey stats, and conclusions drawn in popular scientific press which lacked sufficient authority, though this is again caveated in that certain supporting anecdotal evidence proved compelling.

3.4.4 Software Tools Used in the Review

Papers were mapped into Docear, a document management package which allows linking from groupings of papers to research themes. Initially everything was included under the headings:

1. Physiometrics
2. attention
3. social presence
4. uncanniness
5. mona lisa effects

Later investigation added:

1. Shader lamps
2. holography and reconstructed
3. shared workspace
4. poker
5. physical and hybrid
6. scene reconstruction
7. ad-hoc interaction
8. tabletop
9. augmented reality
10. public spaces

Some principles were then identified to better define the research direction, and visual mappings were established between relevant sub sections of the mind map and these principles. This can be seen in Figure 3.3, where principles especially are shown 'unwrapped' in yellow. It is not possible to show the links between lower level groupings of papers and the principles at this scale on paper, though the PDF version of this thesis allows zooming into the detail.
A further iteration added areas for investigation as follows:

1. 3D reconstruction
2. segmentation
3. latency
4. jitter
5. temporal misalignment

The mind map also carries metadata about citation strength and this was used alongside mapped notes about important features from papers to feed stronger papers into the literature survey. By the end of the process the mind map directly correlated to the structure of the review section. An early snapshot from this can be seen in Figure 3.4.

Tasks were continuously identified using the research methodology and mapped through the MoSCoW prioritization technique into many hundreds of Agile 'tickets' in Trello, an online tool to support Agile delivery. These tasks migrated from left to right through the columns 'Could have', 'Should have', 'Must have', 'Won't have', 'Shaping', 'Doing', and 'Done'. An example of this from part way through the PhD, from the main PhD Trello (there was also a separate board for each paper and paper attempt) can be seen in Figure 3.5.

The research was written up throughout in Latex using Miktex, and TexMaker under first OSX then later Windows. This allowed for semi-automatic migration of the references from Docear, through cleanup in Jabref, into Latex, for the creation

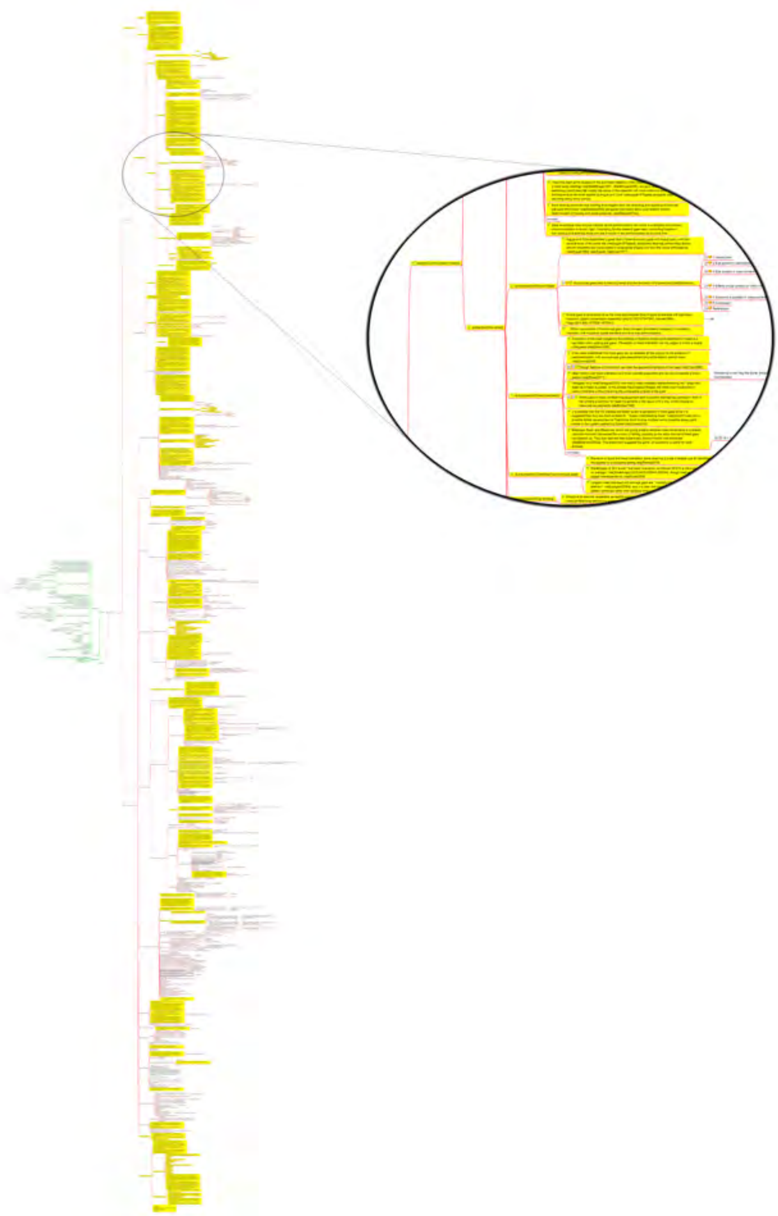


Figure 3.3: High level mind mapping was crucial to the methodology of this research, but difficult to capture visually because of the complexity and scope.

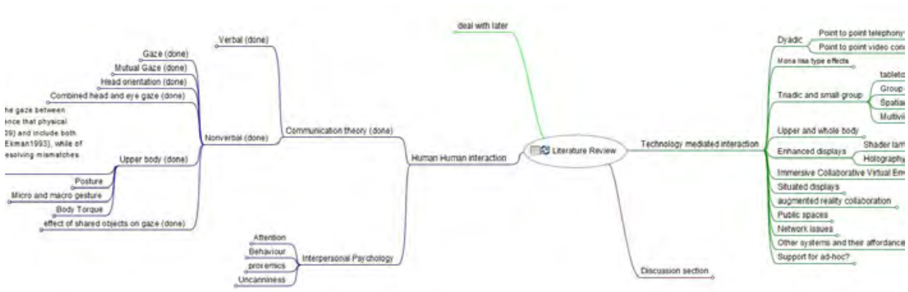


Figure 3.4: Literature survey structure is shown captured from a point in time in the survey. If the nodes at the ends were unfolded then the detail would break down through individual papers to key quotes in a short contextual sentence. These 'elements' for use in papers would include the 'citep' LaTeX code necessary to use the reference in the papers directly, and later this thesis.

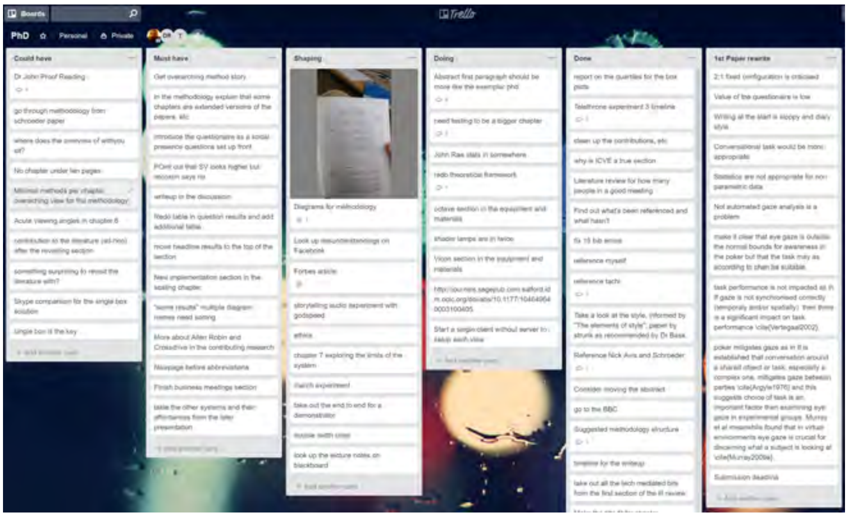


Figure 3.5: The main Trello task board with part of the MoSCow technique and Agile. Tasks migrate from left to right as they are either completed or discarded. This is a very common tool in project delivery.

of references and bibliography. This technique ensured that it was simple to refactor the papers for different outlets in peer-review.

Backup and version control throughout was handled by Google Drive.

3.4.5 Publishing Methodology

Experiments were designed and written up for appropriate outlets in the field. There were many attempted submissions. The first experiment was submitted unsuccessfully to CHI, CSCW, before being accepted for ICAT/EGVE. The second experiment was refused by Springer, and ieeeVR, and accepted by International AR/VR. The third experiment was designed for initial submission to Frontiers Telepresence Journal but there was insufficient time to submit. The papers were later reshaped and expanded into chapters, accumulating many revisions from peer-review. The first experiment was a more significant work with resubmissions over three years and is elaborated in Chapter 5. The second and third experiments made more sense as a combined chapter on reconstructed viewpoint in Chapter 7.

3.5 Overview of Methods for Experiments

It is difficult to disambiguate the Telethrone research equipment from that used by the wider group. The Telethrone chair, itself and all of the various projectors which address it throughout the experiments are exclusive to this research, though all the components have been used before for things like AccessGrid. The Vicon Bonita and Tracker integration is again particular to this research, though it has also been used elsewhere in the group in the past for different purposes. Many of the compute elements were overhauled or assembled specifically for the research, but cannot claim to be in any way unusual.

Software to support the experiments was written by the author for the research with the exception of the software capture system used in the later experiments (which is based on CROSSDRIVE). This was built collaboratively with another team for the research, and to an extent is therefore outside of the scope of this reporting. Much of the capture hardware has been used for this purpose (with slightly different software) before, and this is referenced where appropriate. Some of that capture hardware has been upgraded since the earlier work. It is most useful to simply enumerate everything below and then disambiguate in detail later during the chapters which describe the experiments in detail.

3.5.1 Common Materials

Cameras

All experiments made use of up to 10 x Basler piA1000-48gm 4:3 aspect ratio and/or 1 x piA1900-32gm 16:9 aspect ratio IP cameras. These were fitted with either 8.5mm or 12.5mm Pentax lenses depending on use case. The piA1000-48gm IP cameras capture 1004x1004 pixel images at up to 48 Hz dependent on the available light while the piA1900 captures 1920x1080 pixels at up to 32 Hz.

Lighting

A combination of fluorescent tube room lighting, incandescent Dedolight DLH4's and DS500 studio lights were used throughout the experimentation. These were aimed at white painted walls in the observation suite, or bounced off the matt foamex flooring in Octave to provide diffuse illumination from above and below the subject.

In addition, the Christie rear projected walls in Octave were either set into a white test pattern, or set to project a white jpg image, the top half of which contained an off white rendering of the text script used by the captured subject. In the capture system the three illuminated walls in front of the subject are seen by the segmentation system as empty but allow the subject to look out over the tops of the cameras in a relatively natural way while reading the monologue which is later presented on Telethrone.

Octave Capture System

Experiments 2 & 3 make use of the camera capture system developed in previous research in the group [273].
Octave is a configurable immersive display and capture space designed by the author and Roberts to support this and other research. It was conceived to study the impact of faithful appearance and attention in shared VR (making it ideal for the Telethrone study) as well as providing a multi-modal space in virtual worlds. The space is a large octagon with rear projected walls and a front projected floor. Scale can best be visualised in Figure 3.6.
The capture system is part of this Octave suite. It was initially conceived by Roberts in part to address the evergreen problem of capturing both appearance and attention to support Telepresence.
The core compute in Octave has varied over the course of the research, but the

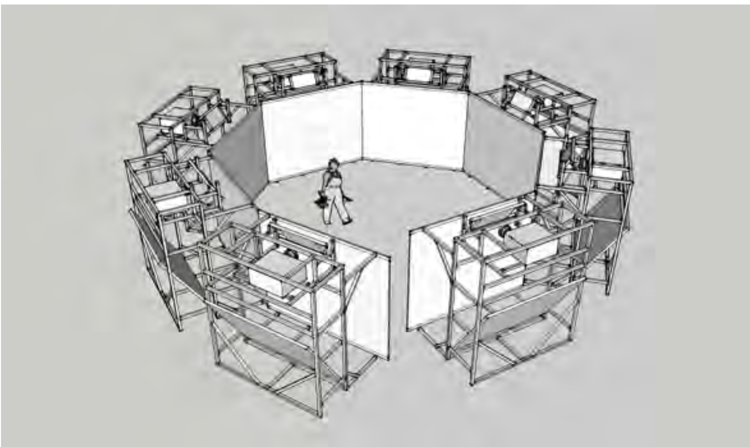


Figure 3.6: Scale diagram for the Octave capture system used in later experiments. For the purposes of the research it provided a capture space and various compute elements for the 3D avatar creation. Image credit Kyle MacDonald.

only unit applied directly to the research from the suite is a 16 XEON cores (32 hyper threads) and 12288 CUDA cores across four Nvidia M6000 GPU's. This can be seen in Figure 3.7

The Octave suite was reconfigured to provide optimised capture for the Telethrone

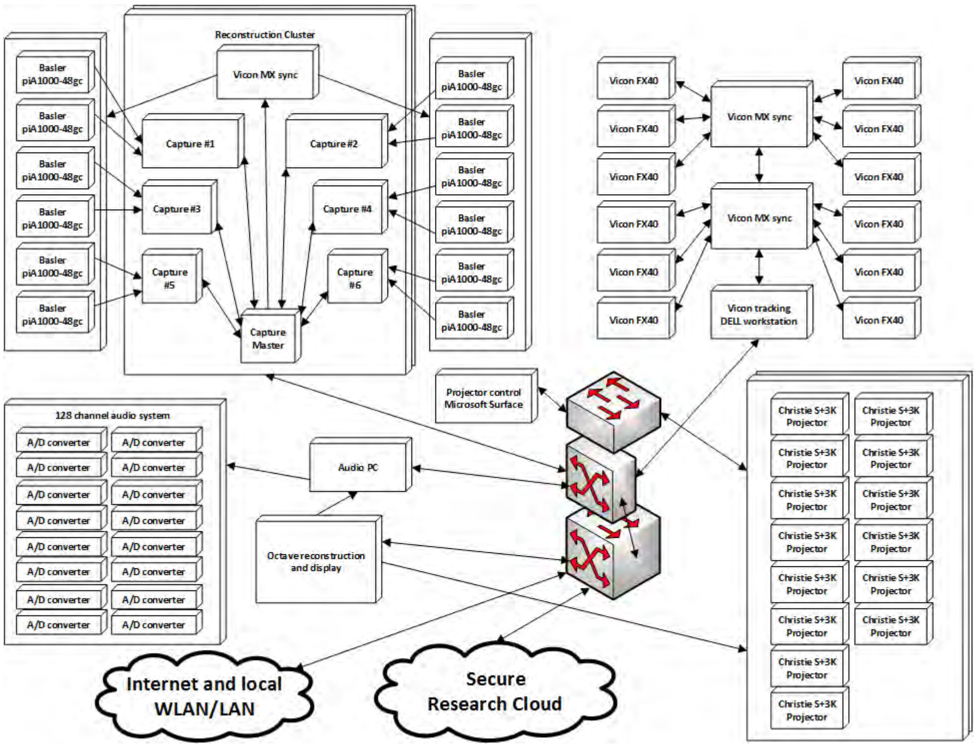


Figure 3.7: UML diagram for the full Octave. In the later experiments everything except the speaker system (bottom left) and the motion tracking (top right) was employed.

reconstruction as in previous research. This meant moving Basler cameras down to roughly eye level for the seated captured user and increasing the light available for the avatar capture. More light means a higher possible frame-rate. As the Baslers have relatively low sensitivity the available light on the target rapidly becomes the limiting factor on the possible capture rate and thereby the playback rate.

Vicon Motion Tracking

In order to analyse the position of experimental participants in the eye gaze experiment, and to generate correct viewpoint renders in the scaled system experiment up to 6 Vicon heads were used alongside Vicon Tracker 2.0 which outputs cartesian coordinates to the local network (either internal loopback or adjacent PC) as UDP packets.

Compute & Software

Hardware used throughout was as follows:

1. Hammer 32 thread XEON system with 4 x Nvidia Quadro M6000 GPUs and 64GB of memory running Windows 10
2. Scan 16 core graphics workstation with 4 x Nvidia Quadro K5000 graphics cards running Windows 10
3. Sun Ultra 40M2 workstation with Nvidia Quadro FX5600 graphics and high gain wireless 802.11n networking running Windows Server 2008R2 HPC edition
4. Various Basler IP cameras and Vicon MX power supplies to run them
5. 3 Christie LX380L 4000 Lumen 1400x1050 projectors
6. 2 Christie DS25W projectors
7. Infocus HD projector
8. 4 Christie LX500 projectors
9. 5 Saville wide angle projectors
10. Custom truck and chair arrangement.
11. 40 inch touch screen display as a touch table
12. Logitech webcam cameras
13. Speakers (powered monitor type)
14. Psychology observation suite in Allerton Building

15. 60 inch plasma TV screen
16. Various Apple Mac Minis
17. Powerful i9 system for offline reconstruction
18. Dell M6400 laptop for recording tracking data
19. Various Manfrotto tripods
20. Vicon Bonita tracking and PoE switches to facilitate their use
21. Various DSLR cameras
22. Behringer CE500A powered monitor speaker
23. Various headphones
24. Sennheiser clip on radio mic
25. Tascam DR-701D (DR701D)6-Channel audio recorder
26. Fibre Baluns for display extension

Software used throughout was as follows:

1. Basler Pylon camera control software allowed fine control of capture and network characteristics to minimise latency and maximise frame-rate for the first experiment.
2. Mplayer network media player linked to the IP cameras through DirectX and provided fast access to the video feeds for the first experiment.
3. Poker software was written for the experiment in simple flash and replicated the action of cards in poker for the touch screen shared task.
4. Google Sheets & Forms for questionnaire entry and logging in the experiments.

5. Cinelerra CV non linear video editing suite under Linux for frame by frame analysis of the video streams from the first experiment.
6. Adobe Audition for matching the audio to the video in experiment 3
7. Excel for Wilcoxxon statistical analysis and question response analysis
8. SPSS for statistical analysis of the exclusion experiment
9. R - suite for statistical analysis of the eye gaze experiment
10. Custom tracking logging software under Unity3D for the eye gaze experiment
11. Vicon tracker 2.0 for both the eye gaze experiment and the final scaled build.
12. Windows 7 / 10 / 2008 Server R2, OSX, and Linux across various systems.
13. Camera recording software provided as part of the observation suite in the psychology department allowed multi channel synchronised digital recording of the experiment for later analysis.
14. Octave capture and reconstruction system [270] was modified by Allen Fairchild through collaborative working to underpin the second and third experiments.
15. VPT projection mapping software
16. Docear (authoring)
17. Jabref (authoring)
18. Miktex (authoring)
19. TexMaker (authoring)

3.5.2 Hypotheses

If a research question addresses a testable element of an objective, then hypotheses were proposed for experimental analysis.

Feasibility Study

The feasibility study examined the suitability of Chromatte cloth for the application. Specifically the ability of the material to spatially segment through retro-reflection back to multiple projectors. This atypical use of the material was subjected to simple optical tests using Adobe Photoshop for RAW camera processing and then histogram comparison between images. Additionally, projection mapping in software called VPT (Seen later in Figure 4.6) was trialed, and while this provided good projection masking features it introduced a further latency and was deemed unnecessary. Consequently physical masks on the fronts of the projectors were employed throughout.

- **H0:** In Telethrone Chromatte cloth is draped loosely onto a chair, such that it has many deformities. The manufacturer of the material recommends that the cloth be flat and smooth for their intended use in chromakey. When an image of a person is projected onto this draped form of the cloth for use in Telethrone, it will similary retro-reflect light back along the axis of projection. Specifically the angle subtended before 5% of light remains detectable will be approximately 15°to the side of the projector (horizonally). This corresponds to the 'spatial' element from the centre of the theoretical framework as confirming this hypothesis enables spatial segementation using the material in this novel context.

Experiment 1

- **H1.1:** Natural conversational attention between static seated participants, judged by eye gaze, is not attenuated in some way by any subtlety of the system such that it can be detected statistically or else be consistently highlighted in questionnaire responses. This corresponds to the 'gaze' element of the theoretical framework.
- **H1.2:** The multi-view condition demonstrates more natural looking be-

haviour than the single-view condition as examined through eye gaze events and questionnaire. This corresponds to the 'attention' element of the theoretical framework.

- **H1.3:** The Telethrone is not reported as anomalous in some way when compared to a real person when examining social presence and uncanniness through questionnaire. This corresponds to the 'situated' element of the theoretical framework since support for this hypothesis indicates that the display is operating in context.

Theory Investigation

- **H2:** In less formal meetings dynamism and flexibility are more appropriate. Conventional formal tele-presence systems poorly support this. In contrast the Telethrone will allow five distinct spatial display segments of the Telethrone even if the onlooking seats are arranged further out, in smaller dynamic sub groupings. This corresponds to the 'informal' element of the theoretical framework.

Experiment 2

- **H3.1:** Subjects will be allowed to walk around freely (in counterpoint to the static arrangement of H1), until they are comfortable that they are meeting the simulated eye gaze of a reconstructed model of a remote person (3D video). The accuracy is not significantly worse than previous research from the group when deployed in this more challenging setting. This corresponds to the 'Mutual Gaze' and 'proxemics' elements of the theoretical framework.

Experiment 3

- **H3.2:** The Telethrone system is arranged to support up to five loosely arranged onlooking participants. The participants can identify when they are being referred to through non-verbal cues alone when the multi-view affordance of the Telethrone is used statistically more than when the multi-view affordance is not used. This corresponds to ‘informal’ from the centre of the theoretical framework.

3.5.3 Participants

Feasibility Study

Experiment 1

Sixteen participants (14 male) aged between 18 and 46 years (*M* = 31.89 years, *SD* = 8.5 years) participated in this study. Written informed consent was gained from each participant after they were given procedural information about the study. Ethical approval was obtained from the School of Computer Science and Engineering Research Ethics Panel at the University of Salford (CST 15/03). All participants received an inconvenience allowance.

Experiment 2

n = 39 participants were recruited for a between subjects experiment with two conditions from Social VR 2016 workshop at MediaCityUK in Salford. No details were taken from the participants in line with CST 15/03 ethical approval. The first 19 subjects performed the best case condition, while 21 subsequent attendees performed the worst case scenario.

Experiment 3

The n=20 participants were recruited directly in accordance with the approved ethics for the experiment (CST15-03) and had a median age of 39. There were 5 women and 15 men.

3.5.4 Variables

Feasibility Study

Independent Variable Position of sensor relative to RPT surface and projector

Dependent Variable RGB light values

Experiment 1

Independent Variable A within-participants design was used with two independent variables; the medium of communication (co-located or projected onto the Telethrone), and the support for directional view. The experiment compared gaze behaviour of the participant between a co-located researcher and a remotely located projected researcher. Two types of remote projection technique were tested. Of primary interest was multi-view, theoretically supporting directionality and mutual gaze between participant and tele-present researcher. Also investigated was single-view which employed an offset camera (centre of the TV) such that the view of the tele-present researcher would be impossible to reliably resolve (having more than 10°the horizontal offset).

In the single-view configuration the video from a camera in the centre of the TV was linked to both projectors simultaneously, and then these video images were aligned to one another on the Chromatte. In this mode both local users see the transmission from this same camera, situated directly in front of the remote person. This created a false off-axis view similar to the spatial offset problem demonstrated by commodity VC systems.

Dependent Variable The dependent variables were: The number of look events (glances and looks), and total duration in seconds of participant gaze to both the co-located and the Telethrone remote researcher.
Responses which were recorded from repeated self-report questionnaires for both the multi-view and the single view conditions.

Experiment 2

Independent Variable The independent variables are how much the captured confederate deviates their view from their body centreline, and how central to the retro-reflected light cone their deflected view is.

In the best case condition the capture is at 45°deflection of gaze, eyes and head are aligned. The reconstructed eye vector is central to the cone of reflected light from the Telethrone surface.

In the worst case condition the capture is at 66.5°deflection of (combined) head and eye gaze. The reconstructed eye vector is to the edge of the cone of reflected light from the Telethrone surface.

Dependent Variable The dependent variable is the angle offset from the correct simulated view vector which the participants settle on during the experiment.

Experiment 3

Independent Variable The primary independent variable is the projection condition, either multi-view with spatial segments for each chair, or else a single projected image of a front on view of their remote avatar recording.

The secondary independent variable is the position of the onlooking chair relative to the Telethrone.

Dependent Variable The dependent variable is the accuracy with which participants can correctly identify their role in the project as described in the monologue.

Chapter 4

Initial Exploration

4.1 Chapter Overview

To begin the research it was useful to explore some prerequisites, and pilot design ideas, before beginning the experiment design. Foremost was the suitability of the material which was available. Some techniques for real time video manipulation to better fit the projected image of the remote collaborator were also attempted. This is detailed first below.

4.2 Other Materials and Investigation Prior to the PhD

Before the start of the PhD there were trials with a trans-retro-reflective materials called MT-561 UNITIKA SPARKLITE film. This self adhesive translucent film is no longer available for purchase. Attempts were made to assess if the micro lenses applied to the non sticky side of the film would provide the same spatial segmentation capability when the projectors were mounted behind the surface. The diffusion through the adhesive backing, or else some other optical property of the interface between the lenses and the substrate means that in practise the image is

diffused far too much to be a viable platform for spatial segmentation.
Large lenticular sheets were also attempted, overlaid on the Chromatte, in a similar way to “Head tracked retroreflecting 3D display” by Surman et al. [313]. This 2015 system is different in principle in that it uses a barrier lenticular system, steered by the tracked head position of a single user, to generate a different image for each eye. It was built long after the early trial attempted here, but shows that the suggestion by Duckworth to spread the retro-reflected light from the cloth more evenly in the vertical plane has optical merit. Some combination of the materials which were used in this attempt failed to produce a viable result however, and this avenue was abandoned prior to the start of the PhD.

4.3 Image Manipulation

There will always be some warping of a projected image stream captured from any camera. The lens will likely have a distortion, especially a wide angle lens. At the time of the initial exploration various Pentax lenses were available. 6mm, 8mm, and 12.5mm CCTV lenses like these are especially subject to barrel distortion, especially toward the extremities of the frame.
The projectors were initially envisaged to address the Telethrone from arbitrary angles. This too introduces a distortion in that the projected image will cover more of the projection surface laterally when shone from a more oblique angle. In the hope of addressing these issues together, and up at the front of the pipeline, projection mapping of the image was explored to test fit the resultant output to a known good shape, effectively engineering out all distortion in one step.

4.3.1 Projection Mapping

For the image alignment an 8 point grid warp was attempted which had a side benefit of some masking capability. Figure 4.1 shows a live webcam warped and masked to fit the experiment VPT, real-time image processing.

4.3.2 Software Masks

Software masking was attempted before this research began, by Thomas, as seen in the inset element of 1. This was briefly attempted again using software called VPT, but was similarly discounted as it added a small but potentially significant delay to the video stream.

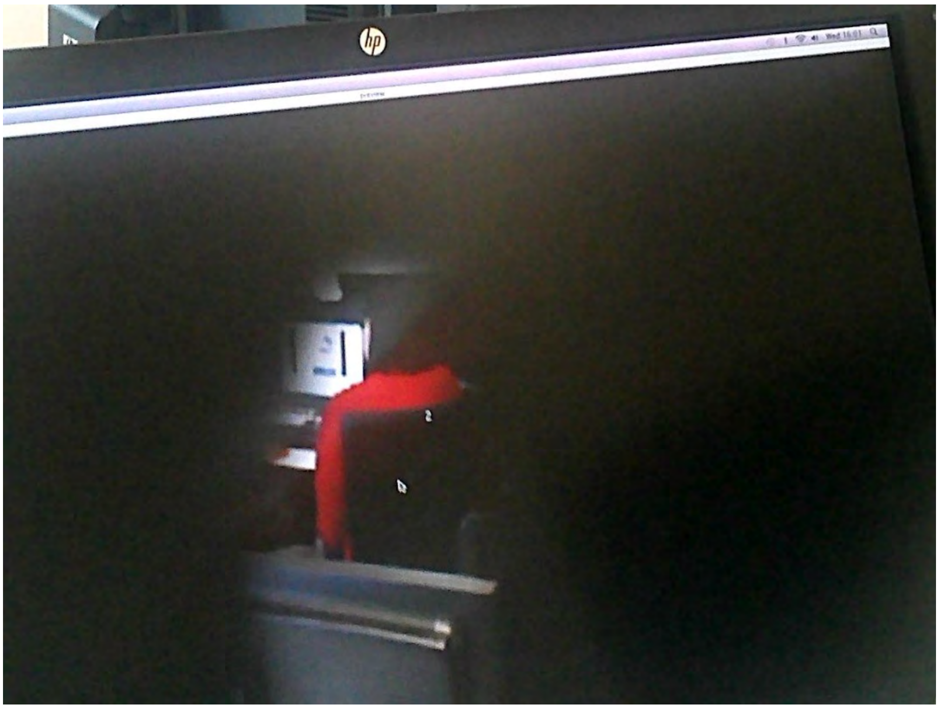


Figure 4.1: Early warping and masking test using VPT software. Points are dragged around the elements of the image need to be kept, and the software blanks the rest. A warping grid can distort the video stream to fit a new 'shape' in real-time.

4.4 Chromatte

Chromatte cloth is a stiff gray cloth coated with microscopic glass beads as shown in Figure 4.2. The optical characteristics of a glass sphere (or a lensed hemisphere as in the diagram) with a backing are such that a ray of light entering the sphere bounces internally round the structure of the glass and back out at the same (but opposing) angle that it entered. This is shown in Figure 4.3

Figure 4.2: The Chromatte surface is a matrix of flattened micro lenses with an open front and reflective back surfaces. The lines are the stitches of the cloth which acts as a substrate for the lenses, (image credit reflecmedia)

According to datasheets[259] and early testing performed by the group and BBC R&D, it should be possible to use Chromatte cloth to spatially segment images.

4.4.1 Scalability

The datasheet states that Chromatte retro reflection falls off to a minima after 15°deviation in any direction from the angle of incidence of the light. In principle this means that a 360°horizontal sweep around the surface of the cloth will offer a maximum of 12 spatially segmented regions. Half of this region is behind the Telethrone, and the two points to the extreme sides of the Telethrone are at such an oblique viewing angle as to be useless. This means that the notional maximum useful number of spatial segments which can address Telethrone with the crosstalk at minima is 5. There is more detail on this in a later chapter 6.

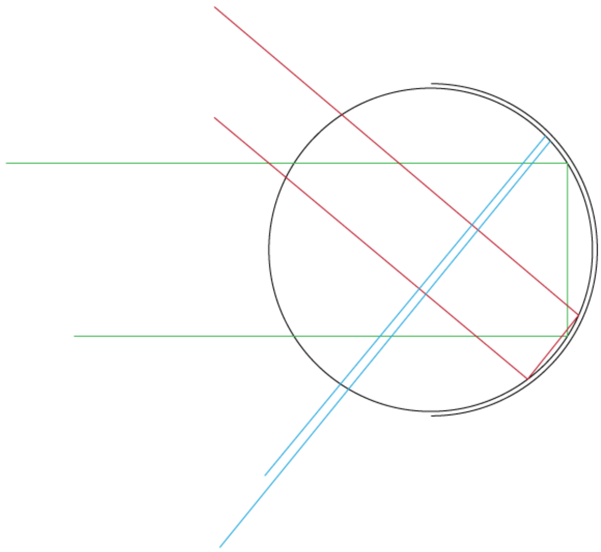


Figure 4.3: Three light rays (RGB) reflecting internally inside a glass sphere. Regardless of the point of entry the geometry of the internal reflection ensures that each light ray exits parallel to its entry. In the context of the tiny lenses involved and the limits of human vision it can be considered that the light simply returns on the same path it arrived.

4.4.2 Testing Chromatte

Early tests with Chromatte material highlighted crosstalk issues between projected images when deployed as a multi-view surface.

Initial pilot testing was performed in various controlled lighting tests at MediaCityUK. Static images of a seated subject were projected from both projectors onto a variety of seats and surfaces. Real time projection mapping software was trialed and ruled out because of latency and complexity concerns. Similarly software masking was attempted and discarded.

It was decided to obtain data for levels of crosstalk in the proposed set-up through emission matching and falloff analysis. Two Christie LX380 projectors were factory reset. They were then checked against one another for consistency of light emission by projecting BBC Test Card F onto a screen at 1.5m, and taking a manually set up RAW photograph, then averaging light across the test card element of the received images. Samples of RGB levels were taken in 5°arc (15cm) increments between the two projected images (Figures 4.4 and 4.5, by combining 3x16 bit RAW images to 25 High Dynamic Range (HDR) sets seen plotted in Figure 4.7.

Figure 4.4: Controlled test set-up in MediaCity:UK studios. The sensor was moved along the line of markers on the floor which were set every 5°from the projecting lens (seen at the top of the shot).

4.4.3 Conclusion

This demonstrates that there is a cross bleed effect between the two spatially separated images which minimizes after around 15°(commensurate with manufacturer's data-sheets), but which otherwise remains at a value of around 5%.

This means that at least in principle there can be 12 distinctly segmented images in a full 360°space. Potentially then up to 6 Telethrones could be used simultaneously

Figure 4.5: A clearer image of the light falloff set-up at MediaCity:UK with two stations. Both have an emitting projector. Multiple shots by 5°deviation the lefthand projector toward the right hand projector show the blend between the two spatial segments.

Figure 4.6: This is the projected image retro-reflected and photographed. The left hand image shows the bulk of the light from the image giving a clear view of the projected person. Very quickly each 5°move to the right demonstrates a substantial falloff in the reflected light to that location. The human eye continues to resolve the image of a person beyond the right hand image in this example.

(having discounted the segments behind the chair), making the system scalable in a location.

The worst of the cross bleed problem is seen in Figure 4.8, but to the human eye (not

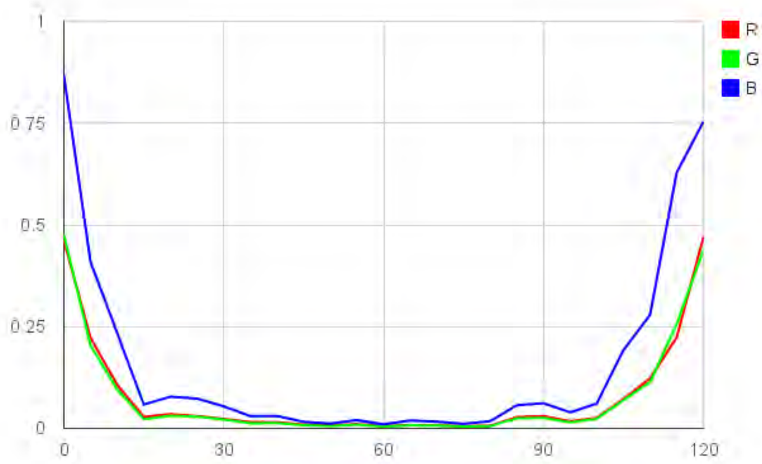


Figure 4.7: Since photographs of the reflected light seems to contain no data it was necessary to carefully control the environment and average detected light across a RAW data capture in RGB from a camera sensor. This is the plot of the averaged light reflection per angle of deviation and by colour where 0°and 120°both host a projector as seen in the other images of the set-up. While it seems that the reflected light drops to zero in the centre between the projectors the human eye can still make out the shape of a person. This is a limitation of the sensors used. Blue plots differently because the human eye is less sensitive to blue and both projectors and camera sensors account for this through increased output /sensitivity.

images which can be captured for this thesis) there remains a slight ghost image at all times. Potential effects of this characteristic were thereby a consideration throughout the investigation, i.e. would a slight double image seem weird, or more specifically be testable as uncanny?

Pilot Deployments

The system was piloted for an event at MediaCityUK as seen in Figure 4.9. This system showed IP video images from a chair in the same room. Even directly next

Figure 4.8: Mid-point cross-bleed. This image was difficult to obtain using stacked long exposures of multiple static images to two projectors in the dark. It is however highly representative of what the human eye sees. Remember this is the point on the previously seen graph where almost no reflected light can be detected, and yet there are clearly two convoluted images here. The deformation of the cloth would not normally be visible to a user of the system and is an artefact of the image optimisation necessary to demonstrate the problem.

to one another the spatial segments seemed distinct, though no data was gathered.

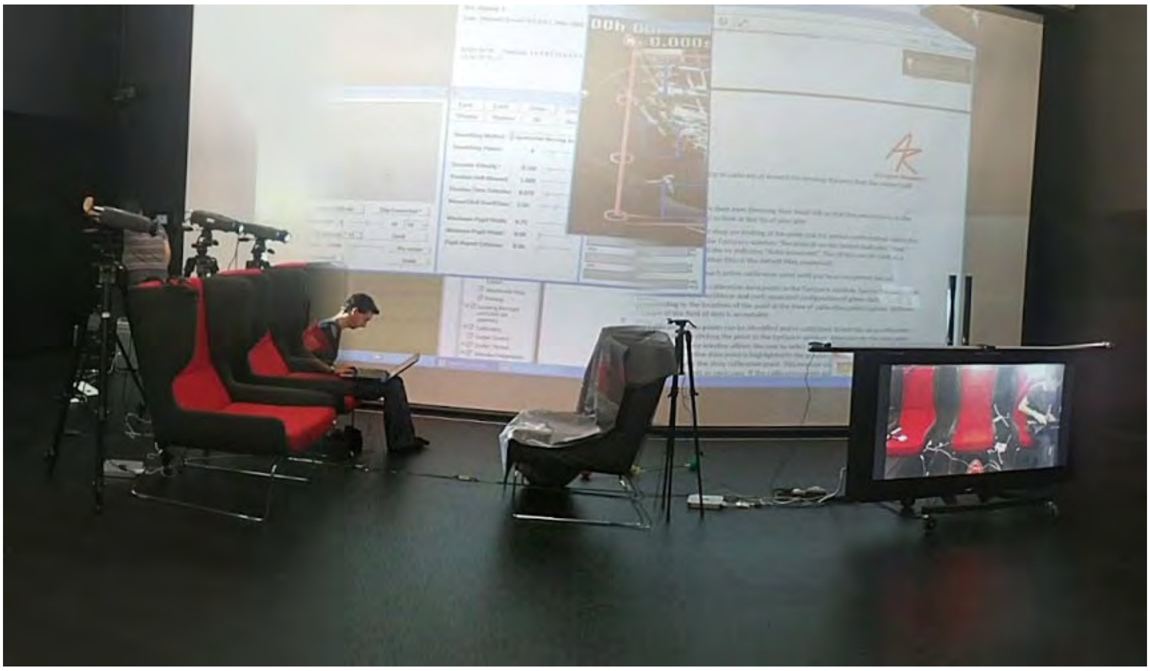


Figure 4.9: In this pilot deployment for 3 observers of one 'remote' user the system was deployed in the digital performance lab at MediaCity:UK UoS studios and connected to the large projector at the back of the shot for demonstration to visitors to an event. At the time, even with the chairs abutted next to one another the system was clear in its ability to show three viewpoints from three web cams above the TV.

Chapter 5

IP Camera System and Triadic Experiment

5.1 Chapter Overview

The following chapter is an expansion of a peer-reviewed paper with two novel components; a system which may improve current small group telecommunication, and an experiment to test the efficacy. The paper was submitted to CHI 2016, rejected and resubmitted to CSCW journal with changes reflecting feedback from reviewers. It was again rejected and changes were integrated before being accepted to ICAT/EGVR in Litte Rock USA. The paper as published can be found in Appendix 11.

Telethrone projects a remote user onto a chair, bringing them into your space. The chair acts as a situated display which can support multi party head gaze, eye gaze, and body torque such that each observer knows where the projected user is looking. It is simpler to implement and cheaper than current systems. Our primary contribution is a counterbalanced repeated measures experiment to analyse gaze interactions. Results suggest that the remote user is not excluded from three-way poker game-play. The multi-view aspect of the system is investigated to test if it demonstrates advantage over a set-up which does not support multi-view and in this results are inconclusive. User impressions from questionnaire suggest that the current implementation still gives the impression of being a display despite its situated nature, although participants did feel the remote user was in the space with them.

5.2 System Development

Wouldn't it be nice to have natural conversations with someone in another office, home, or country, without feeling that technology was somehow 'in the way'? Video conferencing (VC; 'Skype', in its best known form) has made real inroads in supporting this, but looking through a screen keeps the other party very much *in their space* as in Triple-View [54]. Unlike VC, situated displays [243] attempt to put the remote person *inside your space*, while ensuring that gaze and gesture can retain spatial context.

Telethrone is a novel situated display that places a representation of a remote participant in an actual chair within the user's space. This maintains a natural context, in contrast to other contemporary research systems as outlined in the literature review. It may be the case that visualizing the whole of the remote user may be less weird than an anthropomorphic tele-robot or 'head in a jar' style approach.
The Telethrone can provide independent spatially correct views from positioned seats, supporting multi-directional channels of gaze and body torque, and other non-verbal communication. In Figure 5.1 & 5.2) the remote participant projected onto the Telethrone is looking toward the right hand seat as they perceive them. In both of the images this is clear, demonstrating that there are two spatial views of the remote participant.
The degree to which this comparatively simple system represents the remote user is explored. The aim is to bring an impression of the whole of the remote user into the space in a natural way, without noticeably drawing in aspects of their environment with them. It is asserted here that this support for affordable, natural setting multi-view is poorly supported by current available commercial and research systems.

Figure 5.1: View from the participant station toward the Telethrone. The remote participant whose gaze is toward the right hand side of the 60 inch TV seems to be looking toward the participant.

Figure 5.2: View from the experimenter toward the Telethrone. The remote participant is still looking to the right hand side of the TV (where the participant is shown for them). In this case the view is such that the experimenter in the room can see the remote user looking to the left of the shot, where the participant is seated. Considered together, and imagined to be taken at the same instant, these images are an excellent demonstration of the technology.

5.3 Experiment: Does Telethrone Exclude The User From Interaction?

5.3.1 Materials

Video System

Video was swapped between two rooms to form a telepresence connection. Two cameras (on tripods) by the large TV on the left of Figure 5.2 were transmitted directly to two projectors mounted behind the locally situated users on the right of Figure 5.3. The retro-reflection from the projector frustum to the eyes of the observer is illustrated in 5.4

The Figure 5.3 illustrates the two adjoining rooms connected by a one way mirror which hides the cameras as recommended by Hall for reducing the observer effect [130]. The diagram also illustrates the proxemics social space in green, with personal space in orange and intimate in red.

Room & Camera Connections The remote room to the left of Figure 5.3 consistently hosted a tele-present researcher. The researcher sat facing an LG 60PF95 plasma TV whose image showed the other researcher and the participant seated in the observation room (Figure 5.5).

Two Basler piA1000-48gm 4:3 aspect cameras with 8.5mm Pentax lenses were positioned on tripods in front, and to the sides of the TV in the observation room. These capture points were slightly offset from the video ‘eyes’ of the two co-located participants, who occupied the very extremity of the image when they were leaning back (Figure 5.5). This video was IP cabled to the observation room and then played on Mac Minis running Windows 7, which in turn supplied two Christie DS300W projectors.

The live video stream from the observation room to the TV in the remote room

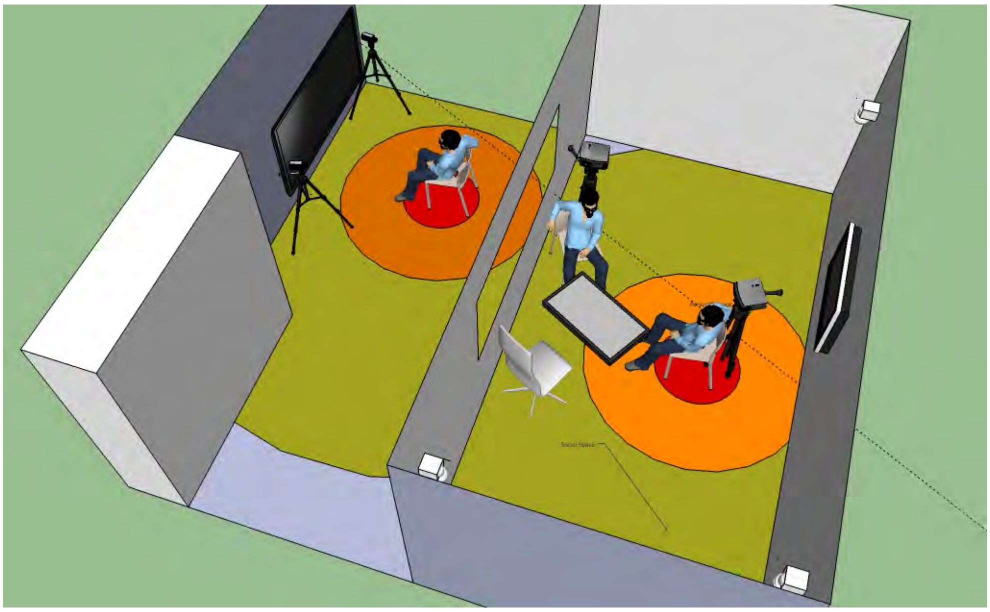


Figure 5.3: The two 'sites' which are connected by the Telethrone systems can be seen on the left and right of the image. What looks like a gap or window in the wall between the spaces is in fact a large one way mirror used for observation of experiments. The Telethrone is shown as an empty chair in this model. The intimate, personal, and social space boundaries discussed earlier are shown in red, orange and green respectively, with the Telethrone participants all engaging with one another in the social band as is most suitable for the meeting style. It can be seen here that the triangle formed by the user and edges of the TV screen in the left hand room are a spatial match for the arrangement around the touch screen table in the Telethrone room.

was captured by a Basler piA1900-32gm 16:9 aspect IP camera (matching the TV aspect), with a wide 12.5mm Pentax lens. The camera position is as close to the projected eye point of the Telethrone as possible, shown at the very top of Figure 5.1. The video stream was Bayer encoded to 8 bit, cropped vertically at capture time to maximize the available frame rate then transmitted over TCP/IP to a Mac mini.

We attempted to maximize spatial accuracy, replicating the triad in the observation room for the remote observer. This transmitted the correct angle of head rotation from the remote person when their attention switched from participant to researcher or vice versa. The offset from the center of the eyes on the TV screen to the camera

Figure 5.4: Light retro-reflecting from the two projectors back along the angle of incidence into the eyes of the associated onlookers. This is a gross over simplification of what's going on optically with the reflected frustum, but is useful in that it makes clear that there are two segments of display here.

was roughly 5°, leading to a 5°deviation from true on the returned gaze.

The low latency (˜<10ms) direct connections provided around 20fps constrained by the brightness of the lighting in the room which was in turn limited by the relative brightness of the available projectors. Faster projectors would potentially enable a higher frame-rate up to the 48fps maximum of the cameras. No attempt was made to emulate the delay of a typical telepresence set-up, as this would present a further confounding factor.

The projectors were height adjusted per experimental run (on tripods) such that their projection frustums sat just above the heads of the participant and researcher who shared the observation room. This tuning per session accommodated height differences between participants and can be seen demonstrated with the base of a projector just above the head of the researcher as in Figure 5.2.

Figure 5.5: The remote user's station with IP cameras. As discussed elsewhere it can be seen that the cameras which capture the views for the two users on the right and left of the screen are not co-incident with their eyes being roughly 5°apart.

Audio

Audio was transmitted to the remote person in stereo from two directional ClearOne desk condenser microphones. These were stationed under the central table close to the users in the room, and sent to discreet headphones visible in Figure 5.5. This stereo signal was given phantom voltage and routing using a Yamaha 01v mixing desk which also served to power a single condenser microphone. This microphone in the remote room delivered the voice of the tele-present researcher from the remote to a Behringer CE500A powered monitor speaker positioned directly behind the projected head(s) on the Telethrone, creating the impression of their voice coming from the head area of the projection.

Recording the Experiment

The observation room was set up with comparatively dimmed lighting and two Cannon Legria FS306 video cameras focused through the mirror onto the participants, with black cowlings to reduce reflection artifacts. One was positioned to overlook the shoulder of the co-located researcher and the other over the projected "shoulder" of the Telethrone. The telltale lights that video was recording were obscured to prevent the cameras being visible to the subject through the glass. This set-up minimized any effects of feeling watched which is noted by Risko et al. [263] as being a significant distraction or modifier in experiments. There were additional cameras in the ceiling for body posture analysis. All playback was synchronized in a multi-track environment (Cinelerra-CV) using a start and end event (a simple clap). These cameras are shown in Figure 5.6.

Figure 5.6: Observation suite camera control station. All of these feeds are steerable and zoomable using the controller, but are less visible to the users of the space as they are in enclosed ceiling domes. These were later synchronised and analysed for the gaze encoding alongside the two which captured through the one way mirror over the shoulders of the participants.

Lighting and Environmental Factors

It was important to control lighting levels to strike a balance which maximized the impact of the system. If the room lighting was too high then the images from the projectors would be too weak. If the light levels were too low then the cameras would be constrained to lower frame rates. A combination of natural daylight, and incandescent Dedolight DLH4's bounced off the ceiling were employed alongside standard ceiling lighting, which were selectively occluded to minimize washout from the point of view of the subject chair. Overall light level was adjusted by blackout blinds to a setting standardised via a light meter aimed at a blank wall. This provided complex mixed lighting, while remaining mindful of the contrast restrictions of the Chromatte, and the necessity to standardize from session to session. The observing space (LHS 5.3) was controlled in a similar fashion to be at a slightly dimmer ambient level, enabling the one way mirror. The remote station was further isolated by material suspended from the ceiling to shield the emissive TV from view. Additional lighting of the remote researcher was provided by a Dedolight studio spotlight aimed at a matt reflector on the floor in front of the TV and two DS500 studio lights behind the TV reflecting off the matt wall above and behind the TV.

There is a balance of lighting to be struck per experimental set-up. Projectors need to be the minimum brightness possible given the ambient lighting levels sufficient for eye gaze to be resolved when the system is in use.

Questionnaire

Participants completed telepresence questions (e.g. ''I felt that the person on my left/right was in the space with me?'') between projection conditions and at the end of the sessions. Participants answered the questions on a 7 point Likert scale ranging from 1 (not at all) to 7 (very much). To attempt to assess potential uncanny effects of the double image from the projectors the likeability and anthropomorphism sections from the GODSPEED uncanny valley questionnaire

[6] were administered using a 7 point Likert scale ranging from 1 (machinelike) to 7 (humanlike).

5.3.2 Procedure

Participants started the study after reading an information sheet and providing informed consent. The experiment separated three people by putting two together in a common room, with the third remote person in an adjacent room. The remote researcher interacted with a TV showing the two people in the common room; while those in the common room viewed the remote person through the medium of the Telethrone. Two of the users were confederate researchers. One researcher was located in the remote room whilst the other researcher was located in the observation room. The other user was a participant.

The experimental set-up reflects the 2m optimum distance identified by Hall [130]. The participant is seen on the right of 5.3 in the proxemics rings with the other players in the 'social space' banding. The three seats in the main room were distributed evenly around the table to balance the conversation spatially.

Shared common tasks are a prevalent feature of group telepresence research as this more accurately reflects professional or social group meetings. A structured task also enhances repeatability under experimental conditions. A disadvantage of a less familiar task is that it can radically reduce eye contact and especially mutual eye contact. More structured group problem solving tasks promote turn taking and thereby support gaze however they are complicated to create and analyse [340]. A card game was implemented in computer graphics to simplify the design.
The card game was designed for a touch table (Figure 5.7). Poker was chosen as the shared task as it is a familiar group activity to many. Additionally, it may be that poker bolsters observation of other players (reading a player's bluff). Olsson et al. discovered in internet poker play which lacked social cues that experienced players who used statistical analysis of the game were able to better control the game through strong play. Clearly there is a social element to poker when social cues

are included. Also, it was established by Ostrem [235] that the act of gambling for money increases emotional engagement, so a financial incentive was included in the study design. The card game was implemented on touch table in Flash/ActionScript and the code is available in Appendix 5.

The experimental design closely matched the physical layout of a real game through the use of the 42 inch touchscreen display mounted horizontally at the same height as a normal card table. Player's cards were dealt face up under a physical mask that closely matched the positioning of cards which would be face down in a real game (Figure 5.2). To look at the cards the players were forced to physically move.

Figure 5.7: A screenshot from the poker game. The cards in the bottom corners here were shrouded behind a physical mask which prevented the two users of the table from seeing one another's hand. Control buttons in the centre allowed the community cards to be dealt at random into the centre of the screen. The Telethrone cards were also physically masked on the touch table. At the other side a laptop with appropriate masks was held on the knees of the experimenter with the screen folded flat away from them. This roughly duplicated the condition on both sides and is highly analogous to the physical interactions one would expect in real play.

This movement resembles the action employed in physical play, where a player will lift the edges of their own cards and duck slightly to check their values.

Buttons available on the touchscreen display in communal view allowed bets to be increased and decreased. They also facilitated folding and the passing of the locus of control. Player chip values and community pots were shown in communal view as in a traditional game.

The distributed poker game allowed game infractions in the same way a physical card game might (i.e. cards could be dealt at the wrong time, players could make out of turn actions). Claiming chips was also a manual process rather than automated. These incidents and interactions promoted discussion during the game through observation of specific actions and additional attention to other players' behaviour.

The tele-present player interacted using a laptop on their knees with the screen pushed back to near the horizontal. This closely spatially approximated the 42" touch table. Physical masks taped to the laptop screen covered the other players' hands. The set-up was faithful to a genuine poker game, with no cheats or biases for the experimenter.

Two five-minute practice games were completed which combined instruction and practice during which conversation was encouraged (typically initiated by the researchers). After this participants completed two rounds of ten minutes on one technology (multi-view or single-view). They then completed the questionnaire comparing co-located researcher with the Telethrone projection. The projection set-up was swapped during this period, and the next two rounds were played before an additional - and final - repeat of the questions.

Chip totals were summed on paper between hands with eventual 'chip leader' initially agreed to win 20 pounds and the other two players 5 pounds. In actuality each subject was thanked and awarded the full 20 pounds at this stage.

5.3.3 Analysis and Results

Analysis

Gatica-Perez notes that there is currently no agreed single system for appraising non-verbal cues and interactions in computer science experimentation and that overall the approaches are ‘trial and error’ [109].

An attempt was made to extract facial action coding from the video streams with Nodus Facereader. The video was stripped into separate frames which were optimised for ‘levels’ with Adobe Photoshop, then roughly aligned with Hugin ‘align image stack’ which also auto cropped the images. This was necessary as FaceReader is quite particular about input images and the faces from the experiments moved around a broader frame which contained elements which were not ‘face’. This image sequence was presented to FaceReader, but the experiment was unsuccessful.

Statistical analysis was performed variously in Excel (Wilcoxon ranked signed sum), SPSS (Wilcoxon and t tests), and ’R’ (box plots).

To prepare for analysis, the 6 video streams were synchronized. Sections of game play (i.e. not briefing or interruptions for scoring or breaks) were isolated, presenting a condensed dataset for inspection.

Analysis was initially based upon continuous visual inspection of the video streams (predominantly from a single ‘over the shoulder’ view use employed in Hall [13]). Glances from the participant toward what appeared to be faces for either co-located or tele-present players were counted. This method was repeated for consistency by another researcher. Next a deeper *frame by frame* analysis was undertaken marking from start of glance to termination of glance for all glances. Anything deemed to be directed toward one or the other players’ faces was marked in Cinelerra-CV for the duration of the look event. Where the glance was ambiguous inspection of multiple camera angles was undertaken. Removing the gaps between these marks

gave a total time for 'looking' for each session of play.
The relatively small sample size of n = 16 meant that it was not possible to test if the data was parametric, so Wilcoxon signed rank tests were applied to the data.
An example of a full four session experimental run with all eye gaze to co-located and Telethrone shown as vertical bars is seen in Figure 5.8

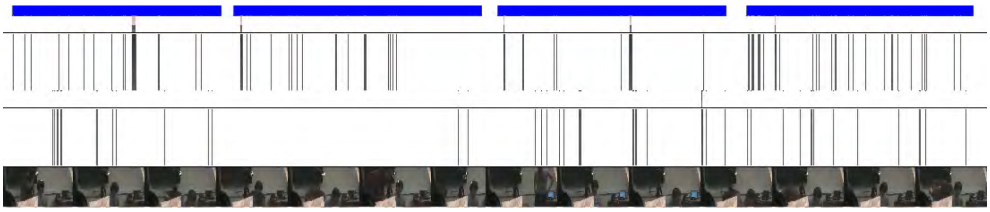


Figure 5.8: All of the look events from one session of 4 games. The video is continuously recorded for the whole session, including breaks and discussion between breaks. The four blue bars at the top of the image show the four games which were played with this participant. Confederates swapped round positions during the breaks. The multi-view vs single view condition was swapped during the central break. Start conditions were randomised. Glances to the Telethrone are the vertical black bars in the upper row, while glances to the co located experimenter are the black bars on the bottom row. The thickness of the bars is the look duration to within 1/30th of a second. The software later concatenated these vertical bars to give overall durations. Video stills from behind the one way mirror showing the right side of the face of the participant can be seen along the bottom. Zooming into this high resolution image in the PDF version of this thesis allows this to be clearly seen but there is little to be added by additional figures to show this in detail.

Eye gaze results

For both multi-view and single-view projection conditions there were more looks (gaze events) toward the Telethrone than toward the co-located researcher.

In the multi-view projection condition the number of gaze events toward the Telethrone were significantly greater (*Mdn* = 33) than the co-located researcher (*Mdn* = 22.35), T = 22, $p = .017$. In the single-view projection condition there were more gaze events toward the Telethrone (*Mdn* = 31) than the co-located researcher (Mdn = 22.35), T = 113, $p = .020$. In comparing projection conditions there is no significant difference between gaze events toward multi-view (*Mdn* = 33) vs gaze events toward single-view (*Mdn* = 31), T = 57.5, p = .587. There is no significant difference between single-view and multi-view for the number of look events when compared to one another, T = 78.5, p = .587.

In both projection conditions there was more gaze duration toward the Telethrone than the co-located researcher. In the multi-view projection condition the total duration of gaze toward the Telethrone was significantly higher (*Mdn* = 29s) than for the averaged co-located researcher (*Mdn* = 23.28), T = 23, p = .020, while for the single-view projection condition gaze duration toward the Telethrone was significantly higher (*Mdn* = 39.88) than for the averaged co-located researcher (*Mdn* = 23.28), T = 92.5, p = .012. There is no significant difference between single-view and multi-view for look duration when compared to one another, T = 67, p = .959.

Questionnaire Responses

Embodiment questions were designed to try to examine how present in the space the remote confederate was.

There was no significant difference in whether participants viewed the confederates in the space with them.

t = -2.0466, df = 26.843, p-value = 0.0506 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -2.803956697 0.003956697 sample estimates: mean of x mean of y 4.666667 6.066667

Participants did regard the Telethrone confederate as being viewed significantly more through a screen

t = 2.0852, df = 27.928, p-value = 0.04631 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 0.0245162 2.7754838 sample estimates: mean of x mean of y 3.133333 1.733333

	co-located re-searcher	**Telepresent researcher**					
		Single-view			**Multi-view**		
Question	***Mdn* -**	***Mdn***	**T**	***p***	***Mdn***	**T**	***p***
I felt that person was in the space with me	7	5	102	.002	6	105	.001
I felt I was looking at the person through a screen	1	4	2.5	.004	3	115	.003
It seemed to me the person occupied a different space to me	1	3.5	1	.002	3	136	.001
When I looked at the person it seemed I was looking into another room	1	2	2.5	.012	2.5	77	.003

Figure 5.9: Questionnaire response results designed for the research attempt to assess how similar the co located researcher is to the more multi-view condition where the Likhart scale runs from 1 (not at all) to 7 (very much). The single view condition is included but is of less interest since this is a control condition.

and to be significantly more occupying a different space to themselves

t = 3.7954, df = 15.381, p-value = 0.001691 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 0.9085365 3.2247968 sample estimates: mean of x mean of y 3.333333 1.266667

They also thought that the Telethrone participant was significantly more in another room.

t = 2.8691, df = 19.432, p-value = 0.009671 alternative hypothesis: true difference

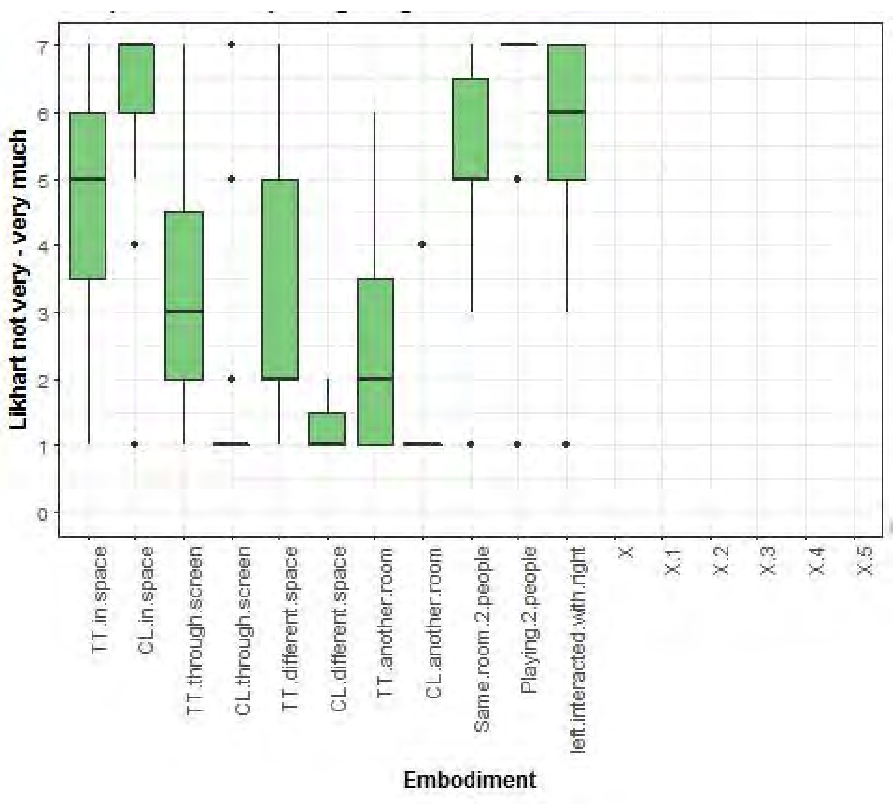


Figure 5.10: Embodiment responses compare the multi-view condition to the co-located confederate on a 1-7 Likert scale. Note that the black lines at 7 and 1 mean that all responded with 7 or 1, meaning no box is drawn. The 0 used on this graph makes these bottom lines more visible, but there was no 0 option in the scale. Working from left to right along the X axis: Co located feels more 'in the space', telepresent is more 'though a screen', Telepresent seems 'in a different space' and somewhat 'in a different room'. Being in a room with two people, playing with two people, and the Telethrone interacting naturally with the co located scored well. Even so there are no major alarming anomalies.

in means is not equal to 0 95 percent confidence interval: 0.3802446 2.4197554 sample estimates: mean of x mean of y 2.6 1.2

Participants felt somewhat strongly that they were in the room with 2 people (Mdn=5) and very strongly that they were playing with two other people (Mdn=7). They also felt that the Telethrone confederate interacted naturally with the co-located confederate (Mdn=6).

The Networked Minds questionnaire (Appendix 10) gives us tools which compare the level to which the participant feels socially engaged with the Telethrone and co-located researchers respectively.

These results can be directly compared. Only statistics which compare the multi-view projection condition with the co-located are reported since this is the desired use state for the Telethrone. In the box plots multi-view is seen in green as the left hand side of each paired bar with the single-view condition rendered in purple to the right of the pair.

The first section of the questionnaire is co-presence which Harms et al. define [132] as "the degree to which the observer believes he or she is not alone and secluded, their level of peripheral or focal awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them."

Responses can be seen plotted against indicative summary wording of the questions in Figure 5.11.

Comparing how much the participant noticed the Telethrone and the co-located confederate there is no significant difference in the relevant multi-view condition

t = 1.0364, df = 28.54, p-value = 0.3087 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.5280251 1.6113584 sample estimates: mean of x mean of y 5.875000 5.333333

Comparing how much the experimenters were perceived to notice the participant there is no significant difference in the relevant multi-view condition

t = 0.6473, df = 29.959, p-value = 0.5224 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.6735169 1.2985169 sample estimates: mean of x mean of y 5.2500 4.9375

Asking how obvious the confederates were to the participants there is no significant difference.

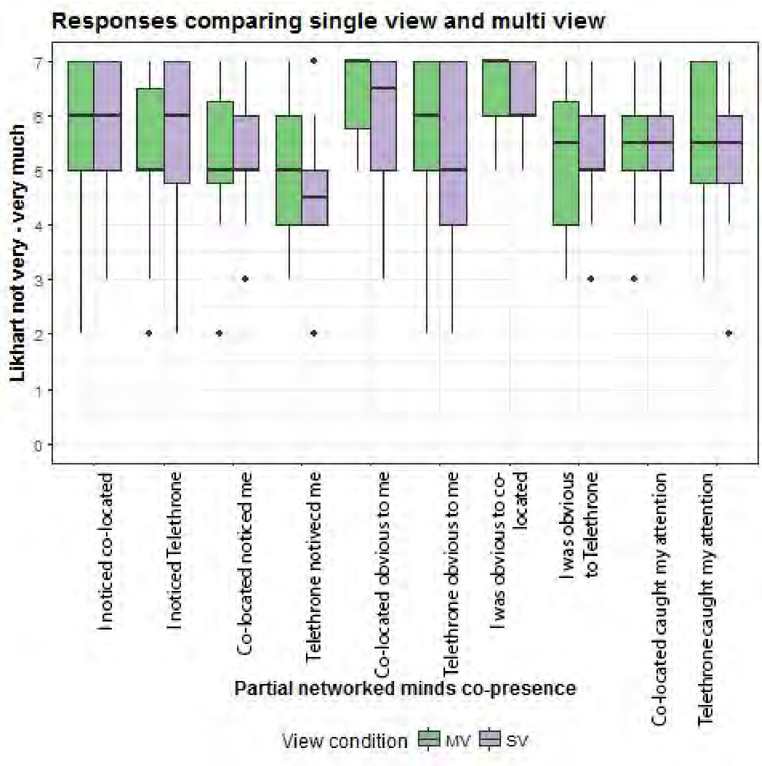


Figure 5.11: Responses comparing copresence of single and multi-view conditions. Left is Telethrone while right is colocated. SV is single-view, MV is multi-view. There is no statistical difference between any of the conditions which is a broadly positive finding for the system, though a notable poorer score for the single view condition would have been interesting.

t = 1.6007, df = 22.194, p-value = 0.1236 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.2101557 1.6351557 sample estimates: mean of x mean of y 6.3125 5.6000

While in asking how ‘obvious the participants felt they were to the confederates’ there is a significant difference

t = 2.4973, df = 24.407, p-value = 0.01964 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: 0.1778969 1.8637697 sample estimates: mean of x mean of y 6.333333 5.312500

The participants felt that both confederates caught their attention equally t = 0, df = 28.841, p-value = 1 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.9325716 0.9325716 sample estimates: mean of x mean of y 5.4375 5.4375

Attentional allocation is the amount of attention the user experiences with the other person [132]. This question of distraction can be seen plotted in Figure 5.12.

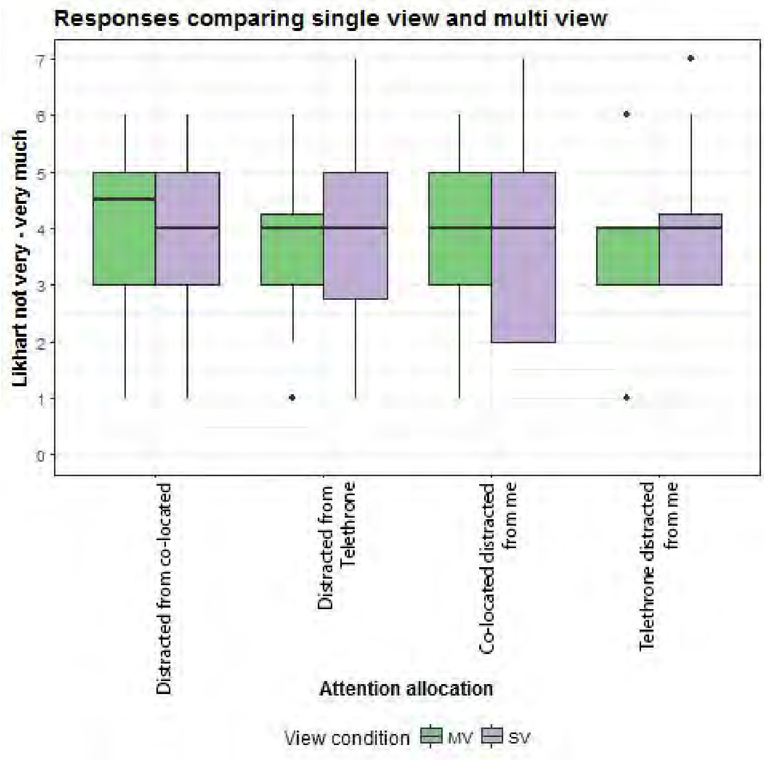


Figure 5.12: Responses comparing allocation of attention between the single-view and multi-view conditions, a measure of how much the participants distracted the other players when something was going on. There is little to glean from these results since the responses are extremely similar.

Participants felt the same level of distraction from both confederates

t = 0.85607, df = 61.973, p-value = 0.3953 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.417209 1.042209 sample estimates: mean of x mean of y 4.09375 3.78125

Nor did the participants feel that the Telepresent confederate was more distracted from them than the co-located participant.

t = 0, df = 58.799, p-value = 1 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.6707134 0.6707134 sample estimates: mean of x mean of y 3.90625 3.90625

”Perceived message understanding is the ability of the user to understand the message being received from the interactant as well as their perception of the interactants level of message understanding.” [132] Responses are plotted in Figure 5.13.

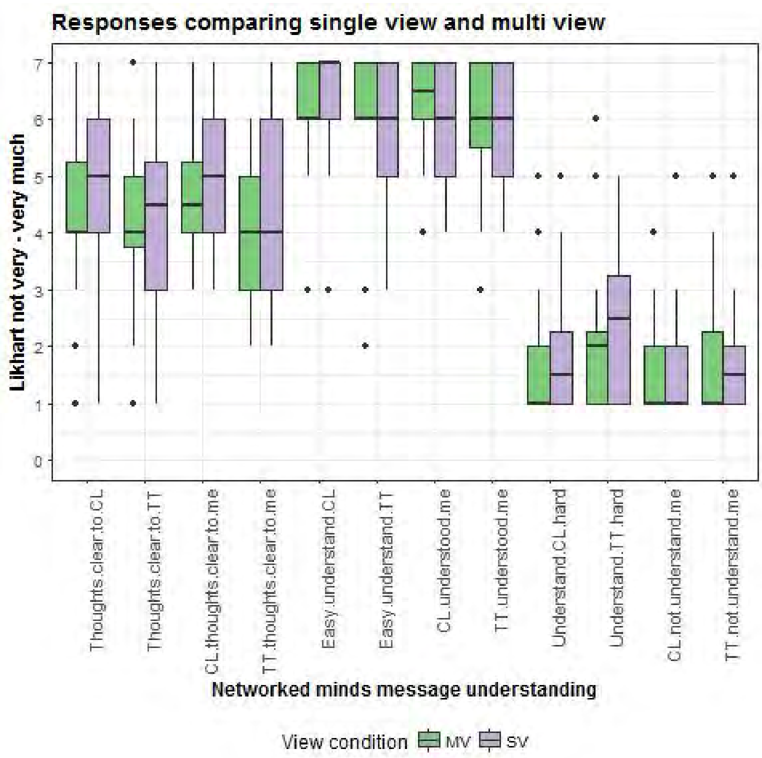


Figure 5.13: Responses comparing message understanding of single and multi-view conditions. TT is Telethrone while CL is colocated. SV is single-view, MV is multi-view. It can clearly be seen that there are no major differences here, though understanding the Telethrone user in single view condition is a full point away from understanding the co located, this is still in the noise statistically.

Participants perceived that both Telethrone and co-located confederates believed the participant's thoughts were equally clear.

t = 0.85568, df = 61.506, p-value = 0.3955 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.4594197 1.1469197 sample estimates: mean of x mean of y 4.65625 4.31250

and that the confederates thoughts were equally clear to the participants.

t = 1.7132, df = 59.027, p-value = 0.09192 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.09447845 1.21947845 sample estimates: mean of x mean of y 4.78125 4.21875

The participants found it easy to understand both confederates (Mdn=6) with no significant difference between the two.

t = 1.7132, df = 59.027, p-value = 0.09192 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.09447845 1.21947845 sample estimates: mean of x mean of y 4.78125 4.21875

They also did not find it hard to understand the confederates (Mdn=1.5) with no significant difference between the two

t = 1.2023, df = 55.671, p-value = 0.2343 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.2290826 0.9165826 sample estimates: mean of x mean of y 6.18750 5.84375

The participants further felt that both confederates could understand them (this was phrased as a counter question and thereby Mdn=1), with no significant difference between the two

t = -1.6081, df = 58.346, p-value = 0.1132 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -1.2625946 0.1375946 sample estimates: mean of x mean of y 1.8750 2.4375

“Perceived affective understanding is the users ability to understand an interactants emotional and attitudinal states as well as their perception of the interactants ability to understand the users emotional and attitudinal states.” [132].

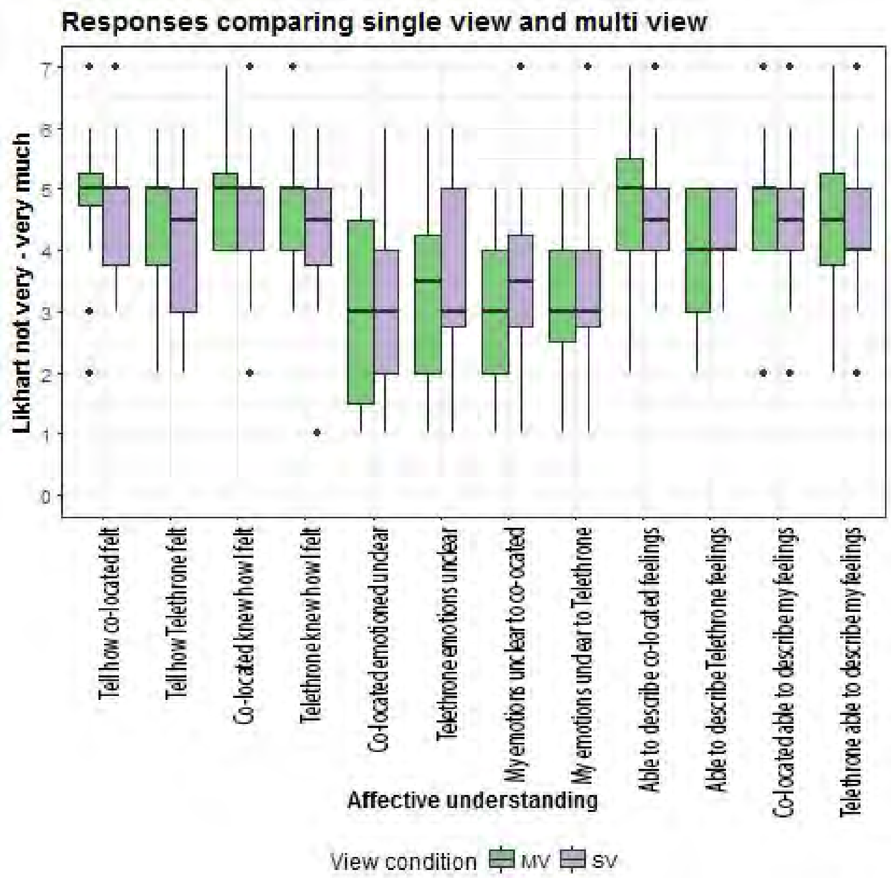


Figure 5.14: Responses comparing affective understanding of single and multi-view conditions. Left is Telethrone while right is colocated. SV is single-view, MV is multi-view. Again there is not much of interest here except to say that the ability to understand emotions and feelings seems well mediated by the system.

”Perceived affective interdependence is the extent to which the users emotional and attitudinal state affects and is affected by the emotional and attitudinal states of the interactant” [132]. These responses are plotted in Figure 5.15.

There was no significant difference in the ability of participants to tell how either confederate felt.

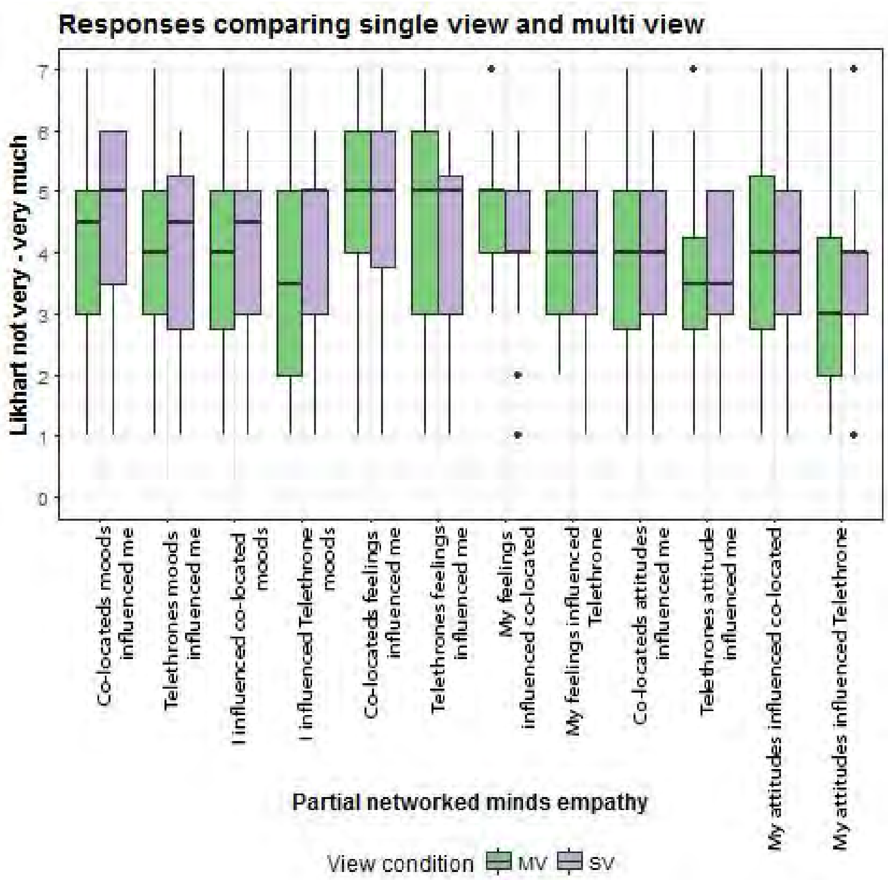


Figure 5.15: Responses comparing empathy of single and multi-view conditions. Left is Telethrone while right is colocated. SV is single-view, MV is multi-view. While not adding anything to the analysis of the Telethrone in this setting it is interesting that the empathy scores were fairly neutral throughout. This could perhaps be a function of the task.

t = 1.6064, df = 61.998, p-value = 0.1133 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.1145595 1.0520595 sample estimates: mean of x mean of y 4.75000 4.28125

nor their perception of how confederates judged the participants felt.

t = 1.2347, df = 61.85, p-value = 0.2216 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.2321307 0.9821307 sample estimates: mean of x mean of y 4.875 4.500

Participants did not believe that the tele-present confederates emotions were significantly more unclear than that of the co-located.

t = -1.0681, df = 60.963, p-value = 0.2897 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -1.1610518 0.3525841 sample estimates: mean of x mean of y 3.064516 3.468750

and their impression of how clear their emotions were to the confederates were not significantly different

t = -0.028828, df = 60.975, p-value = 0.9771 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.7093232 0.6891619 sample estimates: mean of x mean of y 3.312500 3.322581

Participants felt that they would be able to describe the feelings of both confederates equally well.

data: affectUnderstand[10] and affectUnderstand[11] t = 1.6258, df = 55.088, p-value = 0.1097 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.105760 1.015034 sample estimates: mean of x mean of y 4.548387 4.093750

and that they felt confederates would be able to describe their feelings equally well.

data: affectUnderstand[12] and affectUnderstand[13] t = 0.77608, df = 61.58, p-value = 0.4407 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.3940157 0.8940157 sample estimates: mean of x mean of y 4.65625 4.40625

Analysed in SPSS the GODSPEED anthropomorphism section found no significant differences between multi-view (*Mdn* = 5) and single-view (*Mdn* = 6) projections conditions T = 410, p = .309. The GODSPEED likeability section found no significant differences between multi-view (Mdn = 6) and single-view (*Mdn* = 6), T = 460, p = .696.

Figure 5.16 shows the data as plotted by R.

GODSPEED uncanniness responses are seen in Figure 5.16

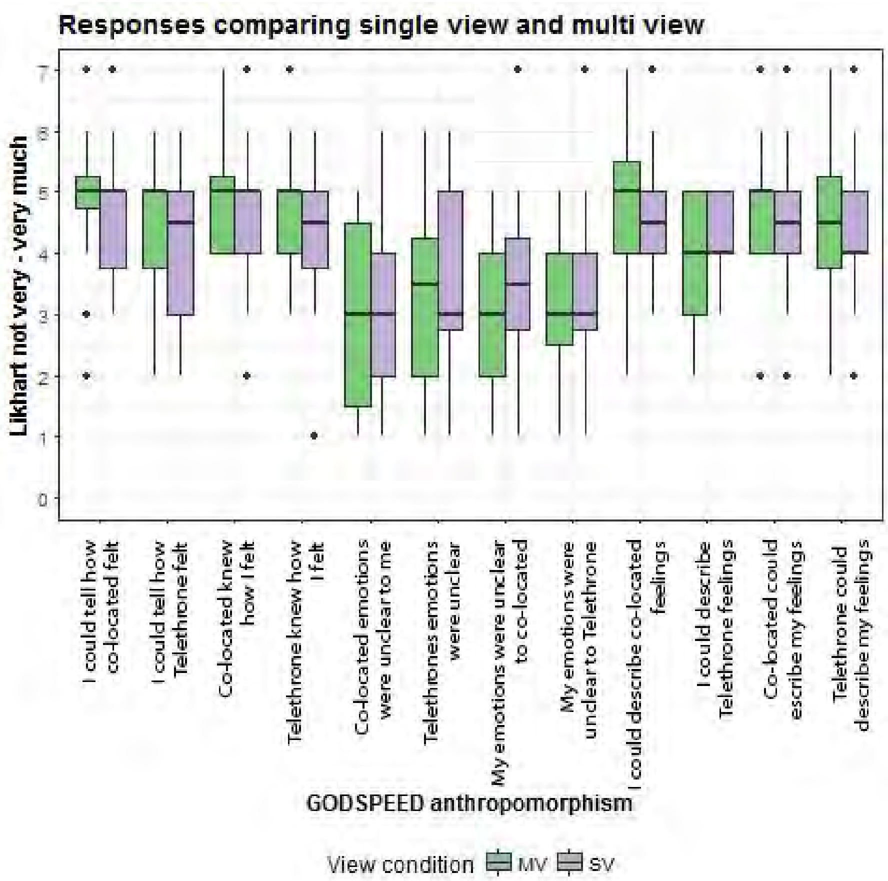


Figure 5.16: Responses comparing godspeed uncanniness of single and multi-view conditions. TT is Telethrone while CL is colocated. SV is single-view, MV is multi-view.

5.4 Discussion

The aim of study presented in this chapter was to investigate if the remote participant was excluded from the conversation in terms of looking / eye gaze events. Had analysis found that eye gaze events were similar between co-located participant and the Telethrone this would have been highly suggestive of the system successfully operating as a contextually situated, tangible, telepresence display. The experiment found that not only were they not excluded but that the Telethrone attracted more attention than the co-located participant. This was true for the number of look events in both projection modes, and duration of looking in both projection modes.

The reasons for this general bias toward the Telethrone remain unclear. It is suggested that the physical flow of play may give rise to increased glances at the Telethrone as play is always to the left, and the Telethrone always to the left of the participant. It may also be the novelty of the Telethrone. It may be inherent weirdness, or the compounding weirdness of the slight double image stimulates additional looking. It is also conceivably a function of the subjects taking more time attempting to resolve the attention of the remote player. This is potentially suggested in the data in that the deliberately poor single-view is looked at the most (*Mdn* = 39.88) with multi-view second (*Mdn* = 29s), and the averaged co-located researcher the least (*Mdn* = 23.28). This difference between multi-view and single-view conditions is not significant ($p = .587$).

We also investigated whether the Telethrone supported directionality of gaze, mutual gaze, and body torque. The difference between these conditions was unclear. Had such a difference been evident in the data it may have been possible to assert that the system also supported mutual gaze in the multi-view system, and thereby would have supported it as a system above and beyond conventional approaches. That the investigation did not expose such a difference is somewhat at odds with findings from the Triple-View system, which compared their spatially faithful system with a similarly cut-down version to the single-view mode of Telethrone. Triple-view found that gaze direction was important when they employed a collaborative task and analysis of turn taking while solving language puzzles [340].

Vertgaal et al. also found that there was a slight increase in all of their reported questionnaire metrics for mutual gaze [54].

It may then be that the contrived and competitive nature of poker, especially the 'covert' aspect of some of the observation gives rise to unnatural look patterns. It was observable in the analysis that there was very little mutual gaze, with competitors preferring observational glancing to judge game-play over conversational and/or communicational glancing and looking.

It is possible that the very fact of a shared task reduced interpersonal gaze to an extent that compromised the experiment as found in The Central European Experiment [12].

However, the most likely explanation is that both conditions do not in fact properly reconnect eye gaze. Even in the multi-view arrangement there is a $5°$vertical mis-alignment in the eye gaze vectors. This problem is addressed in the next experiment which uses reconstruction to more properly connect the gaze.

If the system had demonstrated a level of uncanniness or weirdness then this might have been exposed by the GODSPEED questionnaire, however this was not observed. The observation that there are no differences between GODSPEED responses for the Telethrone and co-located individuals suggests that neither condition is particularly "uncanny".

Chapter 6

Dynamic Meetings and Meetings Within a Meeting

6.1 Chapter Overview

This chapter presents a theoretical analysis of meetings from the perspective of how the research might add value to informal contexts which are supported by one or more Telethrones. David Roberts and Paul Sermon introduced the concept of dynamic meetings insomuch as they thought the most natural way to interact with a group of people is to spontaneously ‘pull up a chair’ into the space, defining your initial physical relationship with the group, and thereby beginning to lay out the social relationships. Experimentation during the research suggested that there was perhaps an opportunity to further leverage the tracking. This optical positioning system enabled the users to pull up a chair. It seemed on reflection that simply pulling up to a correct viewpoint need not be the end point of the tracked affordance, but a valuable support for an intrinsic human social wish to transparently redefine the meeting as needs evolve.

It must be noted that development of this chapter happened concurrently with the final experiment and the chronology of the research is thereby somewhat muddled by this chapter. The second experiment, which is presented in the next chapter seeded this idea, and the data which informs the limitations comes from the system developed for experiment three, again, presented in the next chapter. Nonetheless this is a better structure for the thesis.

Before examining this new potential, the theoretical limits of the system are examined, a technical solution is described which might allow many sites with many Telethrone to be supported.

The challenges of re-orienting the projection frustum of the moved projector with respect to the Telethrone chair are then explored for a hypothetical system which mounts a single wide angle projector per chair.

The spatial limits of a tracked chair are investigated as it is moved around in the most challenging spatial segment around 0 and/or 160°. This provides tentative worst case limits for the most oblique angle of projection.

Texture blending across the whole sweep of viewing angles onto a Telethrone is shown using a real example.

6.2 Revisiting the Literature

During the second experiment (in the next chapter), where the users were freely tracked, it seemed that the ability to move around during the meeting might add flexibility. This was over and above the suggestion by Roberts and Sermon that pulling up a chair might be useful. Revisiting the literature around this revealed very little of use. The knowledge around fault lines suggested that natural groupings could present themselves at the outset of the meeting and that they might evolve over time guided only by intrinsic biases in the group. Additionally it seemed probable that in the kind of social space meeting outlined in the research problem there would be opportunity for changing priorities, and focus of the discussion could possibly be split. This additional tranche of research was added to the literature survey here, under the sections “Business Meetings”, and “Faultlines”.

6.2.1 Baseline: The Formal Meeting

Without even recourse to literature it can intuitively be seen that a formal meeting is normally organised with seats evenly spaced around a table in such a way that each chair is very close to the table. This can be seen in Figure 6.1.
It is usual for these chairs to offer maximum spatial efficiency for the table at hand, that is to say, they are packed close. Invariably they are not wheeled, mobile, office style chairs, but rather the stiff backed square chair of the Victorian dining table. This allows the distribution to be set out once when the table is installed, with participants filling as many seats as required in an arrangement which suits them, but inevitably compromised by the fixed arrangement of the seating.

Figure 6.1: An example formal meeting layout around a table.

6.2.2 Business Meetings

As Scott et al. note in their special issue of Small Group Research [285] there is surprisingly little analysis of the conduct of business small group meetings. Instead, meetings tend to be studied to examine other things such as decision making or the quality of a technology to mediate the meetings. There is in fact far more interest in technology-mediated meetings than in the conduct of meetings themselves. Much reliance is made on a few older studies in the USA. For instance Monge et al. examined the 3M meeting study and found that the highest proportion of meetings (41%) are 6-10 people [208]. Philip Slater at Harvard in the 1950's found that the optimal meeting size is 5 [298] with 4-6 seeming to lead to least problems. Romano and Nunamaker examined historic research and found the optimal to be five to seven members, with a breakdown in group dynamic above seven (although larger groups do offer advantages in viewpoints and quality of the decisions reached) [274].

The BBC were clear that a meeting size of 6 was optimal for supporting their serendipitous meeting target in order to leverage the maximum outcomes from related formal meetings. With these figures in mind 5-6 seems a pragmatic optimal meeting size for the Telethrone technology, so it is serendipitous that this is what can be supported.

Eisenbart et al. found that scheduled meetings allowed participants time to reflect on the circulated agenda and form opinions ahead of time which could become entrenched and potentially polarising within the meeting [86]. They found that in ad-hoc meetings discussion “is likely to revolve around members’ existing expertise”. Sauer and Kauffeld suggested that less centralised meetings may foster increased participation and better outcomes [278].

Amongst other factors Napier and Gershenfeld found that for a meeting to be effective there should be no physical barriers to the free flow of verbal and non-verbal cues as a result of the environment [220, 219].

Duffy and McEuen examine how to choose when a meeting might best be conducted face to face rather than through technology and isolate ‘capturing attention’ as a feature unsupported by virtual meetings. They also identify the ability to inspire a positive emotional climate, and building networks and relationships as features critical to the face to face meeting [84]. Arvey rounds up various research which similarly points to the advantages of face to face over technology-mediated meetings with “a variety of psychological as well as business outcomes” [14]. Clearly a technology which could get closer to supporting the underlying psychology of face to face meetings could save on business travel costs.

Denstadli et al. compared use of VC and face to face meetings in Norway and found that in practice the two fulfilled different and complimentary roles. They found that VC was more used in distributed projects and management meetings, while face to face allowed more informal meeting styles and enabled new business connections [74]. This finding is very much in line with the issue which the BBC raised in that VC is seen not to be useful in the less formal moments in business.

6.2.3 Faultlines

Bezrukova et al. define faultlines as “hypothetical dividing lines that split a group into relatively homogeneous subgroups” [35]. Adapting this interpretation, the ’faultline model’ by Lau, & Murnighan [185] concentrated their attention on diversity as a driver for group interaction. More recent investigation in this area such as that by Voida et al. [343] cites research which identifies fault lines forming around location, nationality, professional or organizational affiliation, shared group identity, power, information flow and diversity, subgroup size, resource distribution, values, race, gender, and age. Most pertinent of these is the discussion of natural sub grouping in partially distributed teams [343], a feature of conventional systems which may be open to investigation under experimental conditions.
Such study is beyond the scope of this work since there are no plans to control for natural fault lines. The concept of dividing lines within a more homogeneous group is valuable however. This section is picked up again in discussion where the literature is revisited. Telethrone should be able to support evolving shapes of meeting, at least in principle.

6.3 Analogies in Fiction

The film *Kingsman* (2014) (Figure 6.2) features a fictional set-up with several perspective correct views of remote participants in a meeting. There are two real participants at the meeting, Colin Firth, seen on the right, and the character who occupies the point of view of the camera. In the film the technology in use is some unspecified and obviously fictional holographic projection system. The text in shot, and the fact that all participants are wearing glasses suggests perhaps that they envisage some kind of mixed reality system like Hololens or Magic Leap (but more discrete). Telethrone is designed without fictional, or indeed any headgear, but it can be seen that the Telethrone could be extended to provide something akin to the Kingsman system. A closer analogous system from cinema is the Jedi Council chamber from the film *Revenge of the Sith* (2005). This is shown in Figure 6.3.

In the image it can be seen that the science fiction here forgoes eyewear, pitching a true holography system of some sort. It's interesting to note that in both the Kingsman and Jedi systems the seats seem constrained to their locations.

Figure 6.2: The fictional tele presence conference from the film *Kingsman* (2014). The film's protagonist is wearing some kind of fictional AR system, using glasses, and can see both the co-located character on the right (Colin Firth) and the group of tele-present (transluscent) figures around the table. In the film spatial faithfulness is a given with eye contact made to the protagonist through the glasses, into the point of view of the cinema viewer. This all seems very natural, and this is what Telethrone would like to offer to a group.

6.4 Hypothetical Maximum

Roberts' original idea for Telethrone suggested many sites, with many Telethrones at each site. It is clear from the research presented that the maximum theoretical distribution is 6 Telethrones around a space or table, since this is how many discriminating spatial segments RPT can provide. The serendipity that research finds this number (or thereabouts) an 'ideal' size for a business meeting has been mentioned. While the thesis goes on to report investigation of 5 chairs addressing a Telethrone it is a different proposition to have 6 Telethrones addressing one another.

To accomplish 6 Telethrones each station would need to be a self contained projection system, with on-board networking and computer, and perhaps tracking.

Consider first (in counterpoint to the formal meeting previously shown) the arrangement of chairs at an informal meeting where 6 people are present. This is

Figure 6.3: The fictional telepresence conference from the film *Revenge of the Sith* (2005) is a level further into science fiction. No sops are made here to potentially feasible technology, everyone can implicitly see everyone else, even as they move around the space. This ability to move around without eyewear or headgear while engaging with holograms of remote avatars is currently the preserve of fantasy films such as this, but Telethrone seeks to engage with these issues.

shown in Figure 6.4.

In a system where all of these chairs were a Telethrone, linking 6 separate sites, each chair would have to capture the real-time mesh and texture information from one or more physically present participants, and send that information out to the other systems. That's a simple enough proposition for a single real participant in the cluster of 6, where perhaps a Kinect could perform the capture. It does not scale to collecting multiple physically present users though. To do this requires either multiple Kinects (each aimed at a different seat), which we know doesn't work at this time, or an unspecified time of flight sensor (again, nothing suggests itself from the literature survey), or else perhaps a multi-camera reconstruction system build into all of the chair stations. To date it seems that this 'image based' approach is the most possible, with both withyou and CROSSDRIVE software demonstrating the capability to capture multiple people at the same time [273, 93]. The Telethrone surface (when sat upon by a user) could conceivably provide the necessary background segmentation surface for such a system, since chromakey segmentation is, in fact, the intended use of Chromatte. When proposing this hypothetical 'seat that is also an RPT projection surface' it's worth bearing in mind

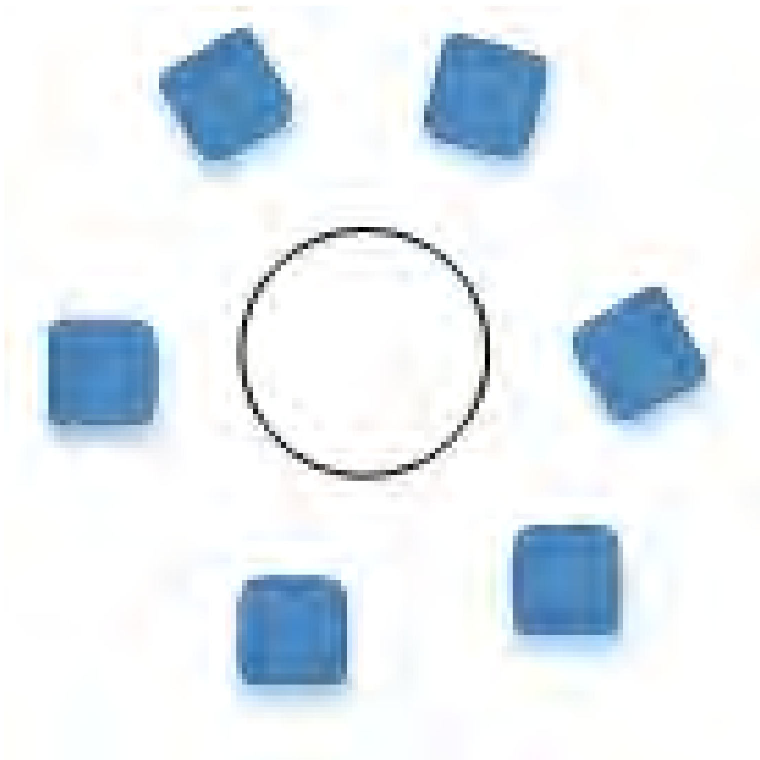


Figure 6.4: An example informal meeting with a coffee table. Chairs can obviously be moved around, removed, added. This differs from a formal meeting table.

that this research finds that the Chromatte used is easily physically damaged, or worn out. Since this is hypothetical, it may be that another, more durable fabric be used. Chromatte for instance have a flooring product with the same characteristics which is likely to be more durable. Segmentation could perhaps be done best using a separate infra-red camera and light, adjacent to the capture camera, splitting the texture and segmentation pipelines and thereby increasing the throughput of the system.

There would then have to be some way to project from a given Telethrone onto all of the other seats, because they are all Telethrones too. If a 4k projector (let's say 4096x2304 pixels, so a 19:4 projection ratio) were built into the back of each Telethrone, with a 130°ultra wide throw lens each, then the angles of projection would be as pictured in Figure 6.5. Such projectors are rare at this time, but increasingly available. Most pertinent to this research is the new breed of laser sourced 4k projectors which separate the light source (a big box which could go

under the seat), from the light engine (a tiny quiet head which could sit above the headrest on a swivel mount). These are currently very expensive units (of the order of $150,000), so 36 of them for a 'complete' Telethrone system would be something in the order of five million pounds.
Ignoring this for a moment; of the 130 degrees available there would be around 15 degrees addressable per chair. This translates to around 250-300 pixels width for the projection of each Telethrone avatar, with around 600-700 available for the height of the body. It is possible to provide more accurate calculations, but not in the context of mobile seats, so these figures are acceptable for this thought experiment.
Taking a real photograph and applying these pixel calculations across the whole frame gives a mock-up for the distribution, and resolution, which can be seen in Figure 6.6.

It can be seen in Figure 6.5 that in the more oblique projection condition onto the Telethrone surfaces, on the stations either side of the projecting Telethrone, that the angle subtended to project the image of the person is smaller, and consequently fewer pixels are available. Some additional testing of this limit is required and is added to the 'future research' options in the final chapter.
Clearly, optically, and in terms of projection technology, addressing 5 chairs is right at the limit of what is possible, but since this whole section of the thesis is discussing theoretical limits it seems appropriate to continue with this 6 seat arrangement. The six site set-up shown in Figure 6.7 is then the theoretical maximum.

6.4.1 A More Pragmatic Maximum

The first experiment was an asymmetric system, with different technologies at each end of the telepresence boundary. This can be seen in Figure 6.8.

While 36 Telethrones across six sites as described might be possible in the future, it does not seem useful to consider this at all at this stage. For the purposes of

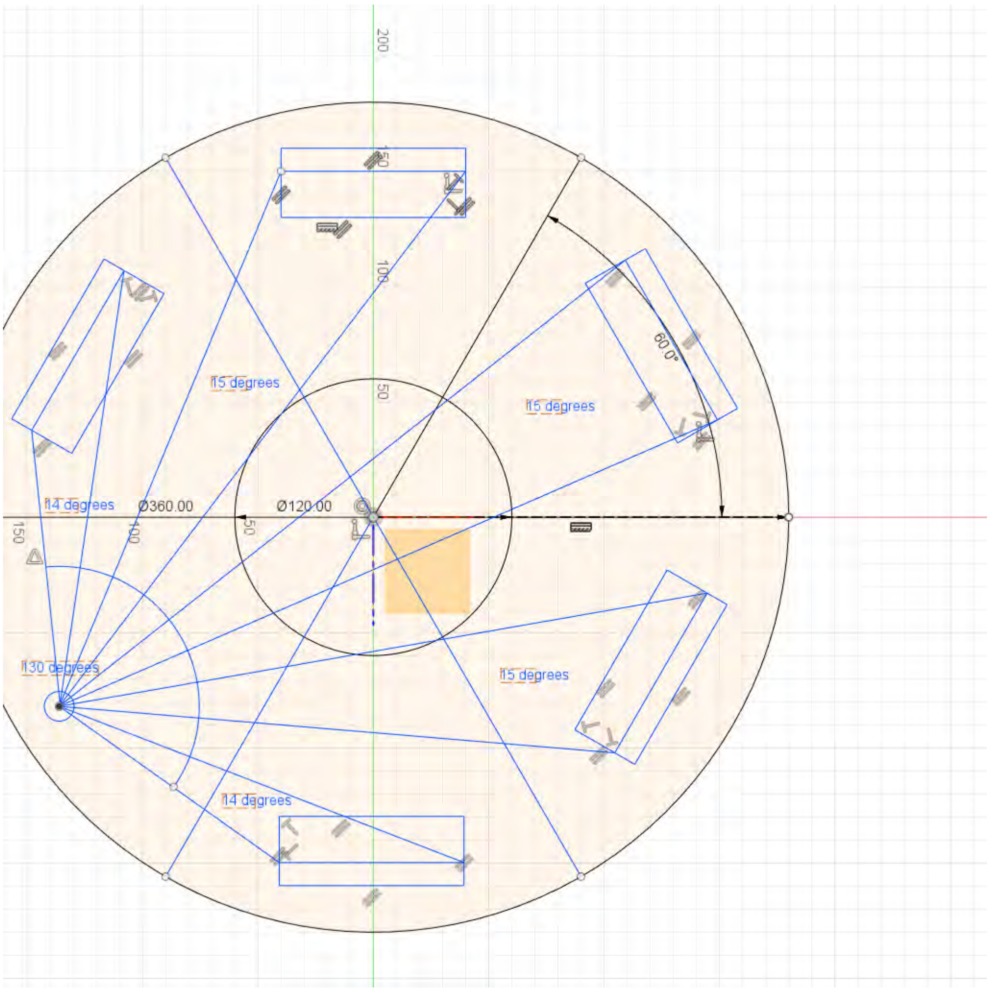


Figure 6.5: 130°projection frustum showing the distribution on the chairs. The more oblique angle of the most adjacent chairs to the projector (which is bottom left) means that slightly less angle is subtended, and so slightly fewer pixels can be used for the human form. It is only around 5% difference however.

the research, in order to propose something which could actually be built with current technology, the maximum system arrangement we should consider is a symmetrical system where a Telethrone at each of two connected sites is always-on, and extends those sites into one anothers' spaces. This is shown in Figure 6.9.

In the following section it will be considered what could be done with such a system.

Figure 6.6: 130°projection frustum showing the distribution of pixels in the postulated 4k projection window (4096x2304 red area). The edge images are reduced by 5% as these are the adjacent seats to the projector and are more challenging and oblique targets. The whites of the eyes can still be made out as can be seen in the (actual pixels) zoom at the top of the image.

6.5 Applying to an Example Meeting

As an example a scenario is suggested where a meeting is convened to discuss whether a project is likely to overrun. For consistency this description will adopt PRINCE2 project role descriptors, and these will be applied when discussing hypothetical business meetings throughout the research.

Remember that Figure 6.1 shows how a traditional meeting table might look. Instead however this meeting comes about without warning; an executive in the company is in the building on other business, and gets into a conversation in the social space with a developer. This is much more like the kind of seating arrangement seen in Figure 6.4.

The product owner who might be expected to oversee such matters is off-site. In some companies it might be possible to quickly book the formal Telepresence

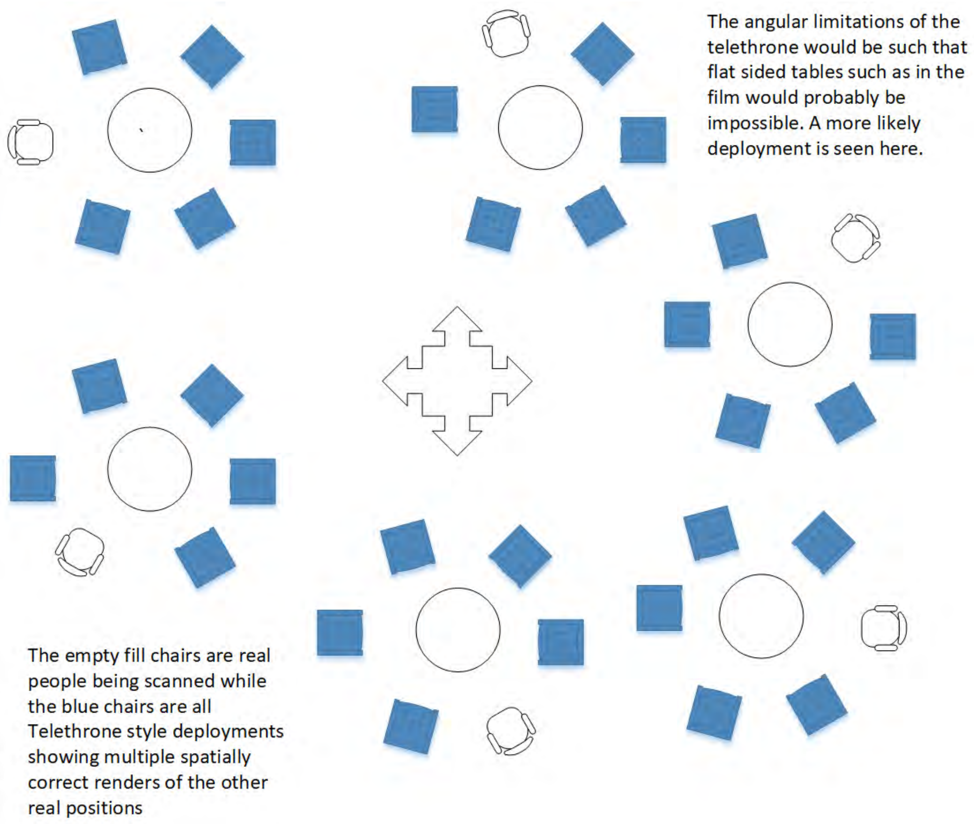


Figure 6.7: Six sites connected with Telethrones. One chair at each site is occupied with a real person. In this figure these are an outline chair in black, but would also be a working Telethrone surface when unoccupied.

room like the Cisco IX5000 seen in Figure 6.10.

Such systems are generally booked well in advance. A more likely option for this situation is something like the Tandberg VC system seen in Figure 6.11. The problems with such a single camera system were explored early on in the thesis.

In this hypothetical example however the Product Owner can be present through a single Telethrone which is installed adjacent to the hot desk work area.

The executive invites two clients to join her. As the meeting begins another developer from the product team passes and is signalled to join by the telepresent product owner. Now five people are physically present in the space, while one is telepresent.

The members of the meeting sit down and naturally cluster into two contiguous

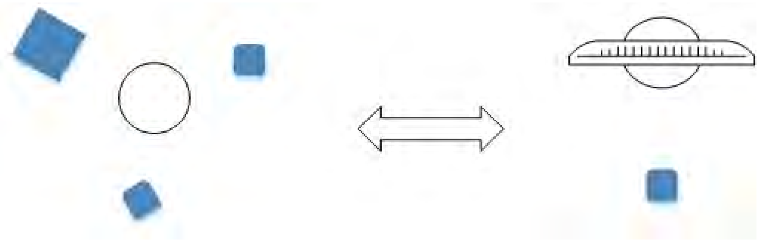


Figure 6.8: Two site (asymmetrical), Telethrone to PC with a large TV as in the first experiment. This could potentially be extended using a Kinect to offer free viewpoint.

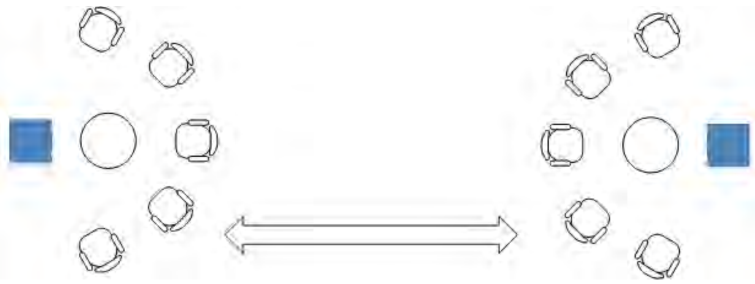


Figure 6.9: Two site (symmetrical), Telethrone to Telethrone informal meetings. This is the idea set-up for this stage of the research and would have allowed testing of an end to end connection of meeting spaces, but required significantly more equipment.

groupings: The two clients and the executive, and the three members of the product team. This distribution can be seen in Figure 6.12.

This is a poor meeting by project management standards, and is happening out of the normal project management cycle and without an agenda. In other words this is a typical ad-hoc, informal, or impromptu meeting.
As the meeting evolves it becomes necessary for the product team to clarify a technical point amongst themselves. In the posited Telethrone supported system it is possible for the two developers to form a 'huddle' with the product owner. This meeting sub group can fork the conversation, while the clients and executive continue with other matters. This can be seen illustrated in Figure 6.13. The chairs which are moved away from the Telethrone could (and should) have projectors which can be rotated to address the Telethrone surface. In this way the 5 members who are physically present maintain their visual connection to the telepresent

Figure 6.10: Two site, technology-mediated formal meeting using Cisco IX5000.

Figure 6.11: Two site, technology-mediated less formal meeting with Tandberg Video Conferencing.

product owner throughout. Their positions in the spatial segments can be seen roughly represented in Figure 6.14. The two stations which share a viewpoint at the top of the diagram would naturally negotiate who was being referenced (near, or far).

6.5.1 Analysis of Space Constraints

To test the limits of moving a single chair around when addressing a recorded remote session it was necessary to move the 'observer' into the kinds of spaces

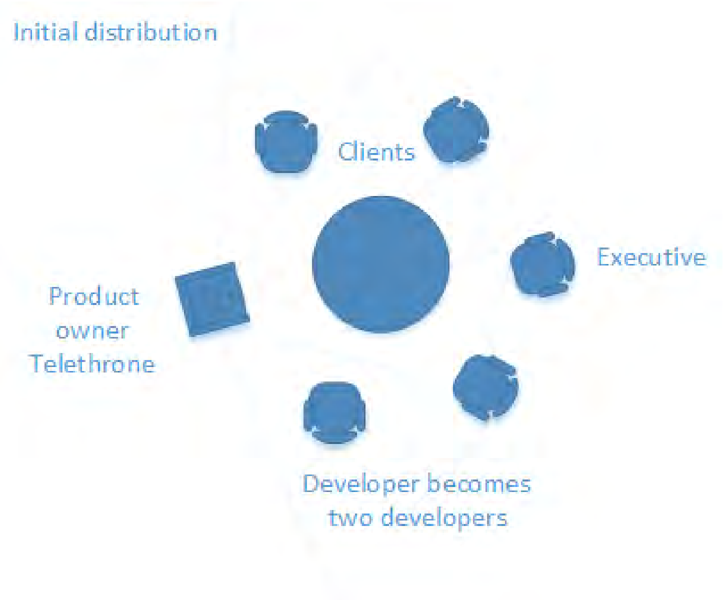


Figure 6.12: Initial seating distribution allows for a person walking past the group to be invited to join the group through verbal and NVC.

where these posited breakout huddles, subgroups, or faultlines might end up located. This was possible just by tracking a video camera around the space and continuously reconstructing for that tracked viewpoint.

As mentioned, instead of tracking the projector stations it was decided to track a hand-held video camera. An opportunity arose to capture a Channel 4 TV presenter during a broadcast they were making about this and other elements of the Octave system (Figures and 6.15 and 6.16). Aside from increasing the engagement impact for the work this was a good dataset to test with.

6.5.2 Limits on the Reconstructed Viewpoint

As in experiment 2 (described in the next chapter), the tracking data from the camera was recorded. In the end this data wasn't used as the image was clear and stable out to the limit of the space available, and the head pivot which locks the 3D video in place on the Telethrone was correct up to around 1m from the Telethrone

Figure 6.13: A temporary internal faultline causes physical redistribution of the chairs.

surface. This head pivot and the software technology which underpins reconstruction onto the Telethrone was developed in collaboration with other researchers for the Telethrone, and is consequently in Appendix 1 rather than described here. To give brief context the software is set up to effectively lock the centre of the head of the reconstructed person onto the headrest of the Telethrone. This prevents the remote user from moving off the projection surface.

When attempting to move closer than 1m to the Telethrone the body of the reconstructed hull pivoted up toward the tracked camera. This could be engineered out (by simply locking this software rotation), but it is unlikely that anyone would want to stand that close to the projection in the proposed use.

In summary the spatial limit of the system is imposed by two things; the size of the tracked volume, and the ability of the observer to resolve the white of the eyes on the projection. In the testing was possible to move 7m away from the surface and still make out the white of the eyes (by eye).

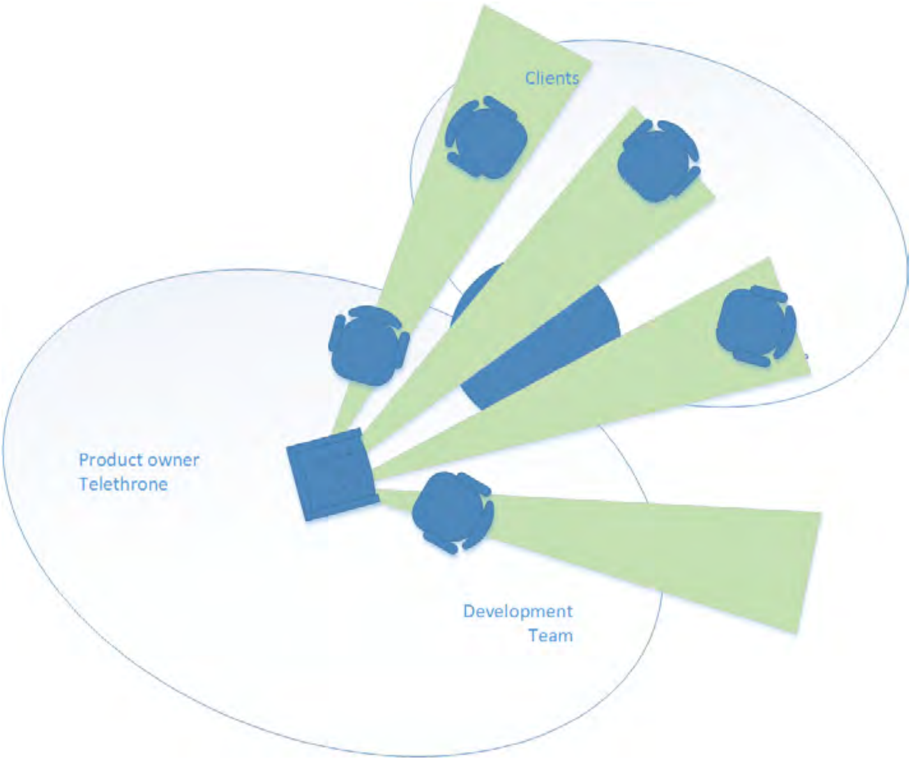


Figure 6.14: Everyone present can still be in a spatial segment and get a clear image of the Telethrone, albeit that two participants share a viewing angle.

6.6 Texture Blending (Not Ideal Capture Placement)

In the experiments the capture cameras have been positioned such that they utilise the new feature of the codebase to select a texture channel, without blending, if that channel is in line with the camera. This gave the best case texturing for the experiments, avoiding some of the problems in the previous implementations of withyou and CROSSDRIVE. It seemed that this contrivance was worthwhile to remove the confounding factors of a capture system which is to an extent providing a service to the Telethrone concept, rather than being core to it's design. To fill this gap in the knowledge of the current system, and provide a look at what a chair and user might see if they were to move to a new (less orchestrated) viewing position, it was necessary to check the blending between the projection stations. This analysis also gives some additional knowledge for the discussion in the final chapter around

Figure 6.15: A TV presenter facing away from the Telethrone capture cameras in Octave with the TV camera man filming the set-up.

the potential for people passing by, in the social space, how they might be somehow tracked, and invited to join by the Telethrone member of the meeting.
A photograph of the reconstructed face of the TV presenter was taken every 5°at a distance of 5m from the surface. This can be seen in Figure 6.17. 5m is outside of the arc of the seating, and very much a worse case presentation.

6.7 Discussion

It has been demonstrated in this chapter that Telethrone is more than the demonstrators presented in the experiments throughout this thesis. The fully scaled system is within the reach of current technology, albeit at a high economic cost.
Such a system , when always-on, and backed by a wall projection to give some broad context of the users of the other spaces could offer an incredible level of ad-hoc joined meeting spaces across a large organisation. It's important to remember here something that isn't immediately obvious when looking at the diagram of the scaled 36 seat system across six sites. Only six participants can be joined at any time, across the 36 Telethrones. There might be one physically present at each site, or there might be five physically present in one site, and just one at one of the others (as in the more pragmatic maximum discussed). Combinations

Figure 6.16: TV cameraman captures the playback onto Telethrone from a single projector while the camera is tracked to present a perspective correct viewpoint

are obviously arbitrary within the constraints that only six people can be in the system. The sites which do not have active users could show the meeting of six virtual avatars, and it is not impossible to imagine being able to hand over a seat to a newcomer at any site, much as one might in a real dynamic informal meeting. There is a lot of extra research, design, and unpacking of the social ramifications of this potential to be done, but this hypothetical maximum feels like too much of a stretch at this stage of the development to even include in the further opportunities in the concluding chapter.

As a last word on the 36 seat maximum system, it is hard to discuss what is and is not acceptable for the viewers based on the mock-up, but the pixel density seen in Figure 6.6 is a close match to that seen later on the four stations addressed by Christie projectors in experiment 3. The whites of the eye are visible (depending on the subject, and the lighting, and the frame in question). This can be seen in 6.17. The best guess is that it would probably deliver the functionality the system needs. Further improvements could (and clearly would) be made to capture, reconstruction, and perhaps even some additional lenticular lenses on the suggested 4k projectors, to minimise the lost pixels. Its seems that the technology is 'in reach',

Figure 6.17: Photographs of the reconstruction on the Telethrone every 5°, showing the texture blending of the system

and yet not in the spirit of the envisaged affordable system.

The more realistic target of two systems connected by a Telethrone at each end seems to be both in-line with what The BBC requested, and possible with current technology. Ten small projectors mounted on swivels above ten chairs seems plausible and affordable. No extreme 4k solution is required if only one Telethrone surface is being addressed at each site.

Chapter 7

Reconstruction Based Telethrone and Experiments

7.1 Chapter Overview

In this chapter the system is expanded closer to the goal of supporting less formal meetings through the integration of a 3D reconstructed model of a person with live viewpoint dependent rendering. The two demonstrators (and associated experiments) presented in this chapter are the first implementation of Roberts idea that chairs can be moved around the Telethrone concept system and still provide correct viewpoints to the observers.

In Section 7.3 the results of an n=39 between subjects experiment show the accuracy of reconnected mutual gaze mediated by a prototype system. Subjectively easier and harder situations (best and worst case) are compared [229].

The deployment is then optimised as far as possible with the available equipment to demonstrate playback (with audio) to five onlookers. In section 7.5 an experiment is presented which tests transmission of non-verbal cues when n = 20 participants watch a recorded monologue captured in the 3D video system and played back from disk. Performance figures are also presented for real-time, end-to-end, transmission of a 3D video using the new system.

7.2 System Development

The Telethrone decribed in the previous chapter is a faithful representation of the remote participant as it uses video, but like other video systems this locks the users into predefined locations. Addition of a rendered polygonal representation of the remote user to the system not only allows better eye gaze vector alignment, but also allows movement around the model, freeing the viewpoint. This is similar to Sphereavatar [242] and Gaze Preserving Situated Multi-View Telepresence System [243]. However, it may be the case that projection of a person onto furniture is less odd (or perhaps uncanny), and more familiar than seeing a head in a jar as with Sphereavatar, or a person in a tube as in Telehuman [167].
The Telethrone system described in this chapter moves toward simple support for social space deployments, and dynamic arrangement of chairs within the group as detailed in the previous chapter.

7.2.1 Previous System

The Telethrone described in chapter 5 used IP cameras to capture spatially distinct views of the remote collaborator, and a return camera above and behind the Telethrone (Figure 7.1). This arrangement demanded that both sides of the system have a fixed distribution. This made scaling the system difficult. Additionally the offset between camera and eye point resulted in ambiguous gaze and mutual gaze awareness and may have resulted in the null hypothesis 1.2.

7.2.2 Blending with 'withyou'

'withyou' is an experimental capture and playback system which uses the Octave multi-modal suite [273]. This system uses shape from silhouette recontsruction [83] alongside a novel network transport system [268] to send a full 3D video polygonal hull to another rendering location. Previous tests on the system suggested that the

Figure 7.1: It is perhaps useful to remind the reader of how limited the previous system was. The two onlookers with spatially segmented views retro-reflected from Telethrone cannot move from their positions, and neither can the remote participant.

capture and playback made it possible to judge the eye gaze of the reconstructed subjects to within limits which underpin social interaction [272]. It was developed onward for a mixed reality system with multiple sites collaborating on shared data [93].

For the purposes of this research elements of Octave are employed for capture and processing (it is more broadly used as a CAVE style display environment [67]). Developments of the UML diagram seen earlier in Chapter 3 can seen in Figure 7.2 as deployed for Experiment 2 Figure 7.3 for experiment 3.

Combined System, Telethrone Reconstructed

Further refinements of the *withyou* capture system have been undertaken for integration with Telethone. Some of these developments are specific to the demands of the Telethrone system.
Development requests to the withyou and CROSSDRIVE codebase were submitted to the programmers and are detailed in Appendix 1.
Physically closer capture and better projection gives higher effective pixel density.

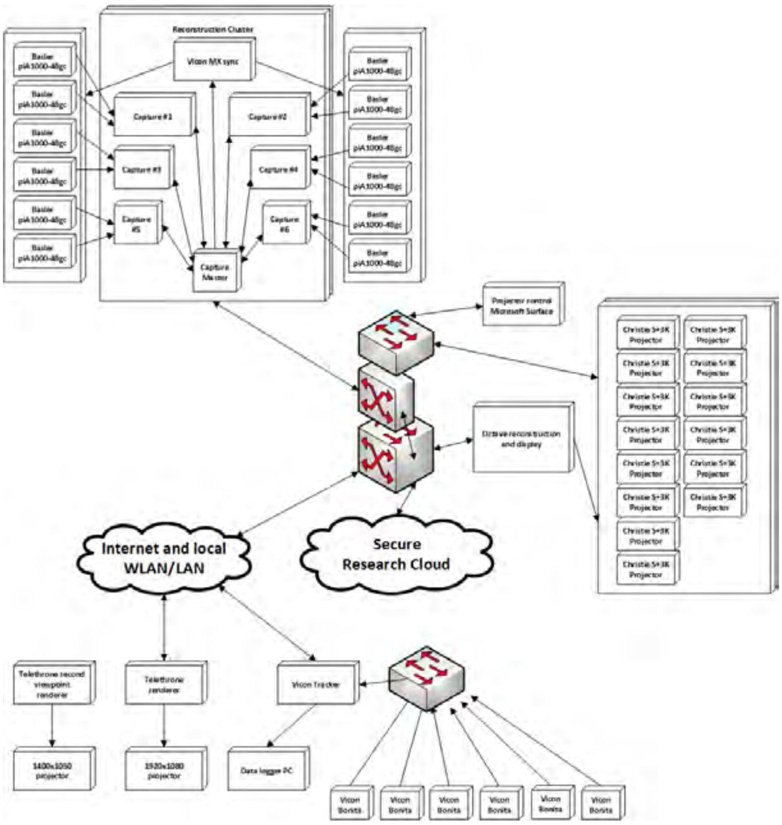


Figure 7.2: UML diagram for the Telethrone system used in experiment 2. The top half of the diagram is the capture system, while at the bottom the two projectors, two image generators, and two more systems are required for the tracking.

Vertical and horizontal resolution is around twice that previously available for the face and eyes throughout the capture and display pipeline. A new texturing technique was also employed which picks and applies the best texture for the viewpoint. Additionally the texturing camera is at eye level, which was found to provide most faithful 3D video in the *withyou* system.

Broadly speaking the following steps are taken to transmit a faithful likeness of a person from a remote site to the Telethrone:

1. 2D video is captured in ten HD streams to memory. The most pertinent camera angle is specifically set up for optimal capture of the face,

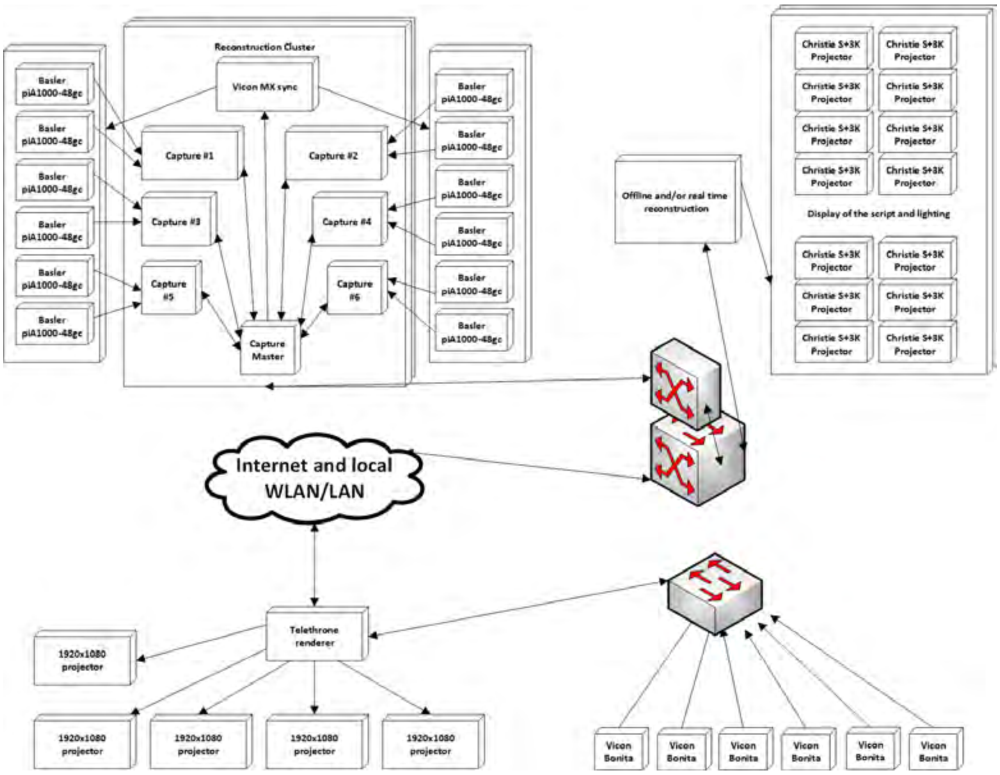


Figure 7.3: This UML diagram for the end to end capable system used in experiment 3 shows a single image generator driving all the projection and the Vicon unit. The capture system in Octave is as before but is now real-time end to end capable.

2. The video is segmented on the capture cluster resulting in separate mask images,
3. The segmented data (image plus mask) is sent across the network to the Telethrone system,
4. The Telethrone image generator reconstructs a visual hull using its CPU (not GPU) based approach,
5. Appropriate textures are selected and blended depending on the viewing angle supplied by VRPN,
6. The 3D visual hull is then displayed from the viewer's perspective (Figure 7.5).

Figure 7.4: Capture of the subject in the Octave system for experiment 2. No additional lighting was used in this set-up and only a single camera was brought out of the ring in the ceiling space. The results were satisfactory as only a still image was required for this experiment. More light is needed only when moving 3D video is to be used.

7.3 Experiment: Testing Eye Gaze Using Reconstructed Avatars

The experiment was conducted in a semi-public space with compute out of sight in a secure room. This served to further demonstrate toward the principle of a deployable system which could work in real business environments.

The captured subject in Figure 7.4 was asked to focus their attention away from their body centreline to marked points in the Octave either 45°to their right (best case condition) or 66.5°to their right (worst case condition). 66.5°is a clearly

Figure 7.5: The polygonal hull reconstruction from experiment 2 shows the quality of the facial capture when looking directly into the camera which has been brought into the space. This is nonetheless a full 3D scan, but viewed here from the best angle, which was all that was required for the experiment.

resolved location in the geometry of the capture system and was selected as the maximum one might expect to deviate from the body centreline. The reconstructed polygonal hull can be seen in Figure 7.5.

7.3.1 Materials

Scope

It was difficult with the prototype system available at the time to move the chair and associated projector around the floor on its wheels to take up a different viewpoint on the Telethrone (Figure 7.7 shows that the custom truck arrangement built to test this was unwieldy).

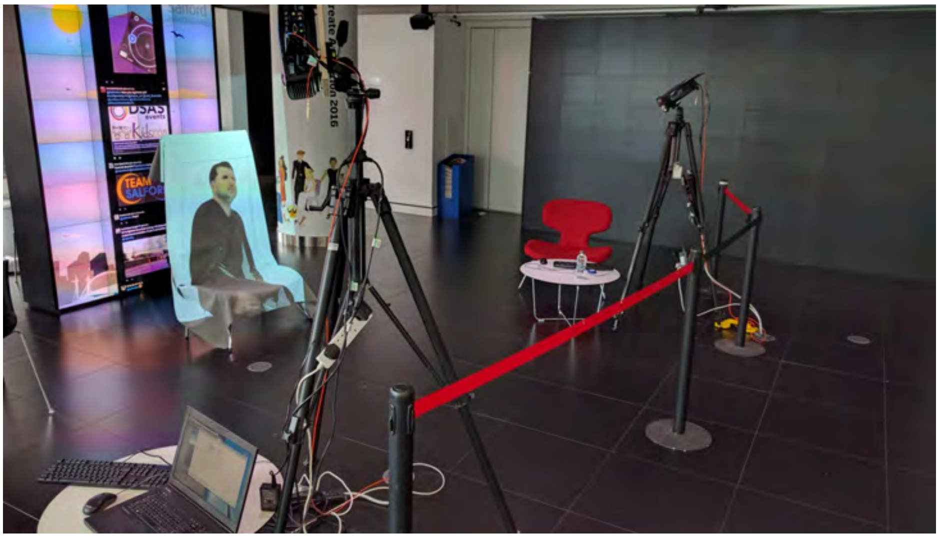


Figure 7.6: Telethrone is here deployed in the Egg suite at MediaCity:UK. The two small projectors on tripods show the same perspective correct rendering of a still avatar. This brings the blend inherent to the material into play. This set-up is though not designed to show multiple independent viewpoints. Participants walked along the red tape line from the back of the shot and across to the left of the frame to where ever they felt most comfortably in the eye gaze of the model.

Figure 7.7: The prototype observing chair and projector are tracked for the viewpoint reconstruction but was not used in the experiment due to its weight. Instead, the computer shown was removed to a systems room.

For this reason, a component was tested, in which the tracking was decoupled from the projection and chair. This allowed the participants to move unencumbered

when finding their position of mutual eye gaze. This tests the readiness of the reconstruction for the broader Telethrone system, but means that viewpoint of the onlookers is not necessarily in the centre of the light cone which is retro-reflected from the surface of Telethrone.
For the purposes of this incremental test experiment it was desirable to maximise the available frame rate at the Telethrone, and have complete repeatability. Rather than attempting a full end to end network linkup the capture was stored to disk for playback during the experiment. The 3D video was 'paused' in playback such that it displayed a single frame on the Telethrone with 60fps viewpoint manipulation.

Tracking

The Vicon Bonita system was deployed into ceiling rigging. It can be seen at the top of the frame in Figure 7.8. The VRPN from the Vicon Tracker 2 was transmitted over wireless LAN from a local systems room to the Windows laptop also seen in Figure 7.6. This laptop ran the Unity games engine with bespoke C sharp code (available in Appendix 4) which continuously logged the position of the hardhat seen in Figure 7.10.

Figure 7.8: The ‘egg’ suite ceiling infrastructure was designed by the author for the deployment of this and other systems. The Bonita cameras can be seen highlighted.

Projection

Two Christie DS25 projectors were loaned from the facility in which the experiment took place as these had better brightness than the DS300 units previously used. This was necessary as the semi public foyer has natural lighting.

Compute & Networking

The Sun Ultra 40M2 workstation seen in Figure 7.7 was moved to a systems room (Figure 7.9), and fed the display to the projectors over fibre optic baluns in the floor infrastructure.

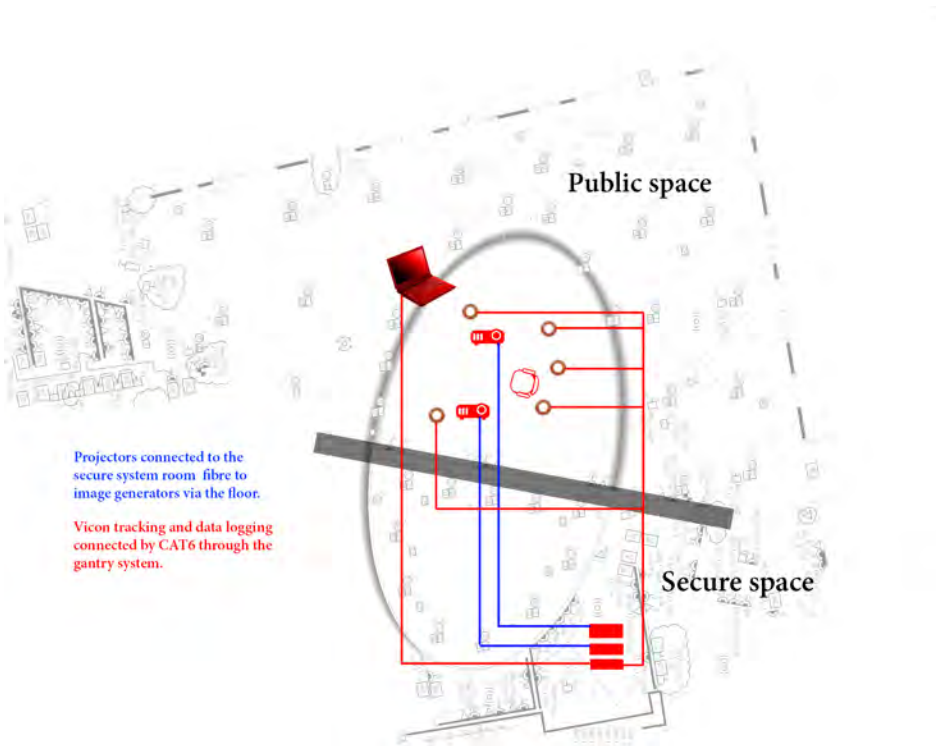


Figure 7.9: A schematic of the deployment showing the remote baluns connections. This demonstrates the suitability of the system for integration into a semi public context.

Procedure

Each participant was asked to wear a tracked hat which presented their location to the renderer and the logging system. They were asked to stand between the two projectors facing the Telethrone. In this location they were presented with a blend of two spatially distinct views onto the Chromatte cloth.

They were asked to walk slowly along a line demarked by a barrier (shown in Figure 7.6). This took them into the projection frustum of the projector displaying the reconstructed and tracked image, which was continuously rendered as perspective correct based upon the VRPN tracking data.

The effect was that the participants gradually felt that they are walking more into the head and eye gaze of the projected subject.

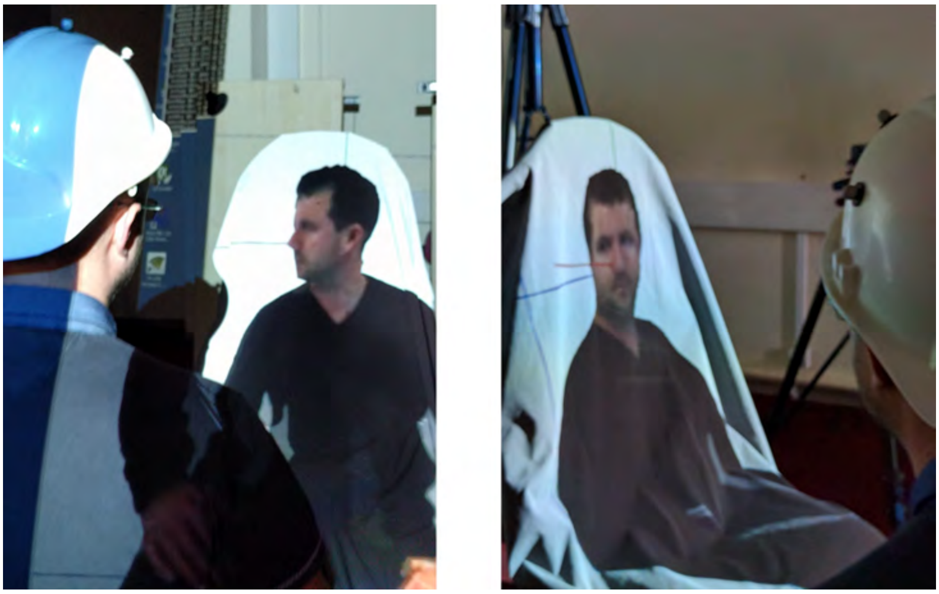


Figure 7.10: Participant wearing the tracked hat. In the left hand image they are standing directly in front of the Telethrone chair and to their mind the captured avatar on the chair is looking away to their left. As the participant moves left the model is manipulated around its central axis such that the viewer perceives they are moving into the gaze of the Telethrone avatar as seen in the right hand image.

When they were satisfied that they were in the correct position, they signaled to the experimenter who immediately stopped the data logging to record their position. The calculated offset from either 45 or 66.5°depending on condition is a measure of accuracy for gaze discrimination.

7.3.2 Results

For the best case condition n=18 the deviation from correct angle is Mdn=-0.78 M=-0.50 StDev=6.28 Q1=-4.00 Q3= 2.46. Shapiro-Wilk significance of 0.087 suggests that the data is normally distributed.
For the worst case n=21 the deviation from correct angle is Mdn=8.27 M=7.74 StDev=4.48 Q1=3.78 Q3=10.86. Shapiro-Wilk significance of 0.419 indicates that the data is normally distributed.
Wilcoxon ranked signed comparison between conditions shows significant differ-

ence between accuracy in 45 and 66.5 z=-2.765 p=0.006 while an independent sample t-test likewise shows a difference with t=-4.769 p=.000. Figure 7.11 shows box plots comparing the two conditions.

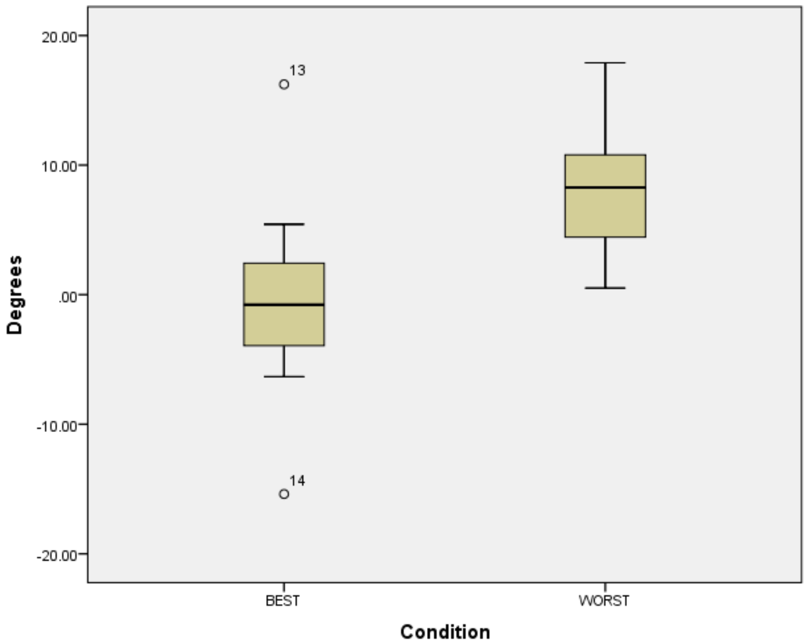
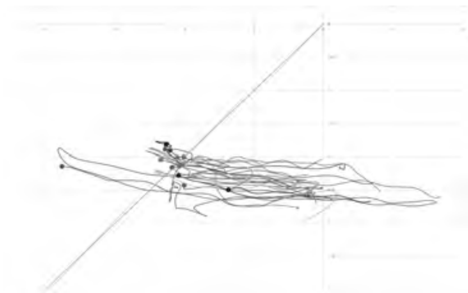
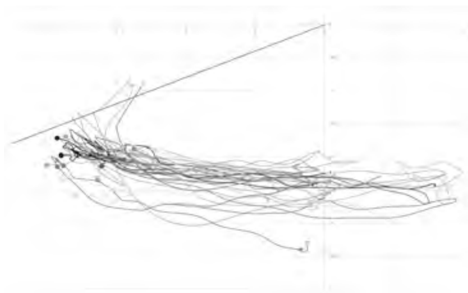


Figure 7.11: Accuracy of the two conditions compared as a box plot. In the best case condition the median accuracy is very close while the worst case condition demonstrates a consistent undershooting of the target.

Best case can be seen in Figure 7.13a while worst case can be seen in Figure 7.13b. The cross bleed and ambiguous eye direction are visible in the worst case.
Figure 7.12a shows the walking tracks (recorded position over time) for the best case condition above a photograph from the correct viewpoint in Figure 7.13a.
Tracks from the worst case condition can be seen in 7.12b above the worst case projection photograph in Figure 7.13b.
It can be observed that the final positions for the participants in the best case condition are grouped around the eye vector for the model (solid line at 45°). The positions for the worst case condition are all short of the line.



(a) Best case condition



(b) Worst case condition

Figure 7.12: The walking tracks of the 18 participants in 7.12a and the 21 participants in 7.12b trying to resolve the gaze deflected 45 and 66.5°from the vertical axis respectively. Participants started on the right hand side and stopped at the dots to the left. Axes are in meters.

(a) Best case condition

(b) Worst case condition

Figure 7.13: Best case view onto the Telethrone from 7.13a: 45°from front, head and eyes aligned, no image cross talk and 7.13b: 66.5°from front, head and eyes not aligned, cross talk from the other spatial segment is visible.

7.3.3 Discussion

All statistics were performed in SPSS comparing within the experiment and against the most similar experiments, Telehuman [167] and “Evaluating the gaze of a virtuality human” [272].
The difference in the final positions seen in Figures 7.12a and 7.12b may be down to the difficulty in resolving the additional component from the eyes of the model, or it may be as the image intensity falls as a function of the retro-reflection or cross-talk between the projected images. All subjects stopped short of the correct position.
Roberts et al. used an earlier version of the free viewpoint reconstruction system and projection into an immersive environment. When comparing to their paper ‘estimating the gaze of a virtuality human’ [272] medians and standard deviations are used create box charts from the original data. This is shown in Figure 7.14
R0L in Roberts et al. has eyes aligned forward in the head, and the head turned away from the body. This is most similar to TT Best.
R’RL in Roberts is eyes, head, and body not aligned and is most similar to TT Worst.

When comparing to research from the literature the best case condition is analogous to Telehuman in their reported ‘looking at’ scenario where participants had to decide where they were being looked at. Telethrone has higher mean accuracy in the best case condition with 0.85°compared to 5.2°. Telehuman has a far better STDEV at .89 compared 6.27. The standard deviation of Telehuman is potentially lower as many more experimental runs were performed.

The mean angle of deviation of the Telethrone system in the best case condition compares well to Telehuman with a higher mean accuracy, and a standard deviation in the same range as the offset reported in Telehuman. Without access to the detailed Telehuman data it is hard to make detailed comparison.

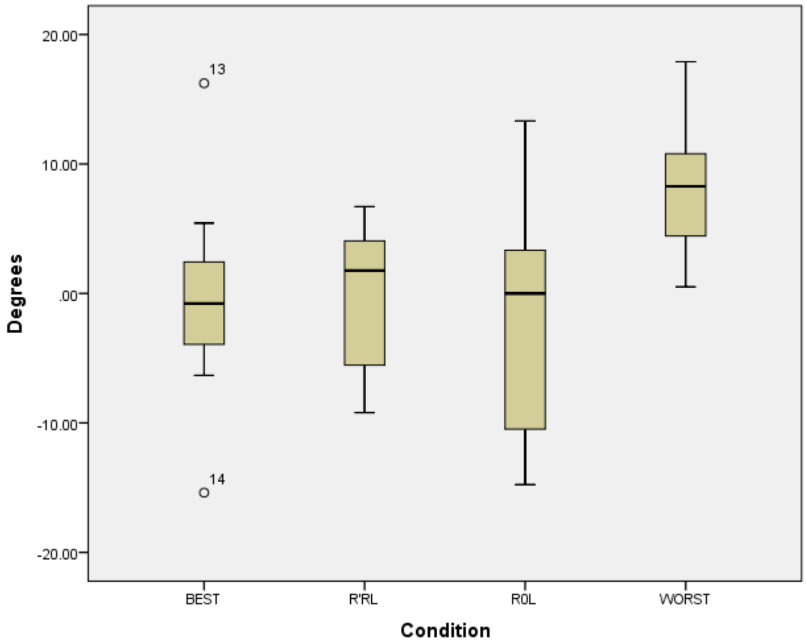


Figure 7.14: Best is analogous to R0L while worst case is most similar to R'RL and exhibits crossbleed between projectors in addition to misalignment of eyes with respect to head. Positive and negative are plotted rather than absolute values (as reported in Robert's paper) as this allowed examination of the directionality of the system.

It is notable that the maximum negative deviation for the best case condition is less than the lower quartile deviation from the analogous condition in Roberts et al's experiment. It seems to be the case that the better quartile ranges in both best and worst case condition is due to the better texturing method and higher resolution.

Subject 14 complained that the system did not work for him and he could not seem to resolve the gaze at all. His final position is an outlier seen on the far left hand side of Figure 7.12a. He made three passes through the correct eye vector. It is interesting that he was the tallest participant with the tracking data showing him to be approximately 15cm taller than the mean height of 1.55m for the group. While height might be a factor it can be seen that subject 13 (the other outlier visible as a dark spot in the centre of Figure 7.12a) was close to the mean height at 1.50m

Table 7.1: Results from comparable experiments

	TT best	TT worst	R0L	R’RL	Telehuman
Sample Size	18	21	17	17	high
Shaprio Wilks	.087	.419	.194	.091	
Median Angle	-.77	8.27	.00	1.77	
Mean Angle	.85	8.22	-2.05	-.34	5.2
STDEV	6.27	4.48	8.61	5.36	.89
Lower Quartile	-4.00	.96	-11.80	-5.90	
Upper Quartile	2.46	10.86	3.33	4.26	
Wilcoxon	z=-2.77	p=.006	z=-.734	p=.463	
TTest	t=-4.77	p=.000			

7.4 System Development

We previously established that the system does not exclude users from shared poker play but found little evidence for support of mutual gaze. An extension of the system demonstrated that it was possible for participants to correctly resolve the gaze direction of a reconstructed avatar projected onto the chair.

Telethrone is now extended to five reconstructed viewpoints. An n=20 between subjects experiment is presented which assesses the degree to which participants feels they are being individually addressed by a 3D reconstructed recorded person as that person delivers a monologue. Comparison against a control in which there is no support for directionality show that the system works as intended at this scale.

Performance figures for real-time capture and playback over the network are also presented.

The demonstrator supports an optimal group meeting size [208, 298, 274] as discussed later. It demonstrates the principle of multiple tracked onlooking chairs which can be moved around in the space, continuously updating the viewpoints onto the reconstructed remote avatar. This new feature enables the concept of meetings within a meeting, meeting fault-lines, and evolving meetings as discussed in the previous chapter. It reduces the amount of equipment required when compared to the previous iterations such that it could plausibly be stored in a flight case, a car boot, or an office cupboard.

From a quantitative performance standpoint figures are presented for rendering multiple viewpoint reconstructions from the Octave capture suite from both stored SSD and over the network (in real time). The new system demonstrates performance improvements over the previous implementations.

As the primary experimental contribution the core capability of directional awareness is tested with an n=20 experiment between subjects experiment assessed by questionnaire.

As a further contribution a technical demonstrator for an end to end system capable of rendering 5 independent views for tracked chairs it presents to. This brings the system very close to supporting all of the listed affordances. This will be discussed in the final chapter.

7.5 Experiment: Testing Perception Within a Group

We present an experiment to evaluate how five simultaneous users perceive the gaze direction, relative to themselves, of a reconstructed human projected onto the chair.
Time stamped playback synced with audio for the first time gives the best indication yet of the utility of the spatial faithfulness of the full system.

7.5.1 Materials

The Octave system was reconfigured as in the eye gaze experiment to optimise camera capture for a chosen eye level viewpoint. This tweak to the system ensured that the issues identified in previous research had minimal bearing on the current system.
A UML style system diagram can be seen in Figure 7.15.

Scope

The Telethrone was modified based on testing to be double width. This allowed better transmission of the pointing gestures made by the captured model. A return channel was impossible with the current set-up.

5 eye level cameras

With evenly distributed cameras the arrangement of chairs facing the Telethrone would each expect to receive texture maps derived from multiple cameras and blended in the software pipeline as in CROSSDRIVE. This configuration was attempted in a pilot trial and also brought down three cameras to eye level (distributed to match the spatial arc of the onlooking participants). Although the texture blended without a visible seam it was decided to align a camera to each viewpoint to give a best possible condition for the texture mapping element of the capture system. A camera was therefore positioned at eye level commensurate with each of the participant viewpoints for the experiment. In this way when the chairs are arranged evenly around the Telethrone each view could expect a reconstructed hull which is textured from a single image coincident with the generated view point. This in no way undermines the functionality of being able to move the stations since the blend feature remained enabled.
To ensure the spatial alignment of the positioned chairs, the capture co-ordinate system, and the playback co-ordinate system a test 3D capture was performed

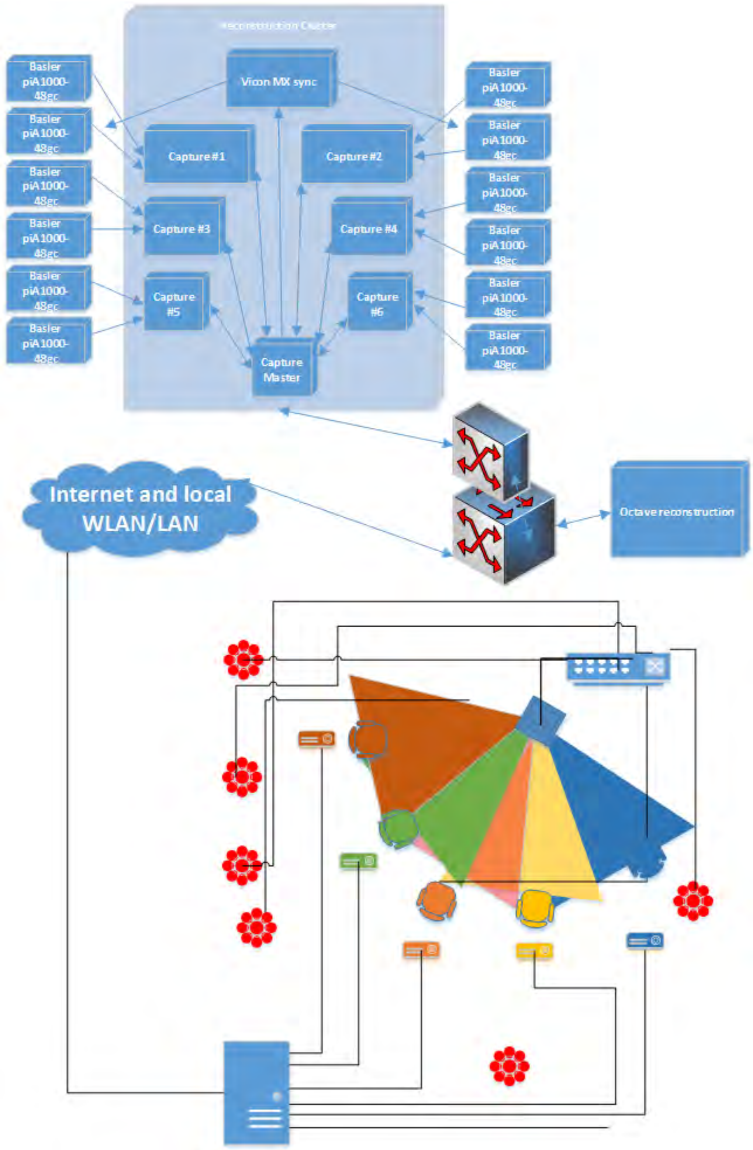


Figure 7.15: Full system diagram for the capture, tracking, and rendering elemtents. This image differs from the previous UML in that the arrangement of the room and tracking is highly indicative at the bottom of this image. The different colour segments are an attempt to show the retro-reflected light frustums from the Telethrone surface.

in which the subject held their hands up at arms length, facing each of the five cameras. The camera capture arrangement can be seen in Figure 7.16.

Figure 7.16: Five capture cameras and lighting provided significantly better framerate and texturing for experiment 3. The walls around the Octave were also illuminated. The stool can be seen in the centre of the space

Compute & Networking

Capture and segmentation The capture nodes 1 through 6 in Figure: 7.15 are Intel Core i5 Processor i5-4460, 3.20GHz, Quad Core systems with Nvidia GTX1050 GPU's and were upgraded for this research from their previous incarnation under CROSSDRIVE. They receive the network streamed raw frames into an additional Intel network card per machine. Segmentation can take place on either CPU or GPU. For the purposes of this paper CPU segmentation was employed. This is further detailed in the forthcoming paper on the modified shape from silhouette system.

Local networking The capture and segmention nodes are connected to a CISCO 4900M switch by onboard 1Gbit ethernet. This switch connects onward to the reconstruction PC by 1Gbit Cisco networking provided by the building network infrastructure.

Offline reconstruction Reconstruction for the experiment was performed on an i9 system loaned from the developer.

Realtime reconstruction, tracking & display Realtime reconstruction was performed on a 10 core i9 system (20 logical cores) with 2 x Nvidia 1080 Titan GPUs (non SLI). This system also ran the optical tracking system (Vicon Bonita). Initial testing was performed with 4 x Nvidia K5000 GPUs on an older single CPU but found that the system lacked the performance to render more than 3 simultaneous channels while staying in sync with the audio as explained. The system was upgraded for the experiments.

Projectors

Projectors sit behind five on looking chairs distributed across 160°, with around 35°between each station. This is around the limit of the capability for the Chromatte cloth and so represents a worst case scenario optically.
Four of these spatially segmented channels cannot be moved though they are still tracked, and the viewpoint is based on the tracker data rather than being hard coded. This means that in principle the current equipment could be replaced with a mobile chair with an integrated projector. Each of these four locked viewpoints currently comprises a chair in front of a tripod base with a rigidly mounted Christie LX380 projector.
Aim of the image onto the Telethrone is adjusted by differential lengthening of the tripod legs. This can be seen in Figure 7.17. The other (more pertinent) spatial segment is projected from an infocus IN8606HD. This full HD projector is mounted sideways on a tripod such that the vertical resolution available is 1920 pixels which can be zoomed to fit the entirety of the Telethrone. This is a decent demonstration of how a commodity projector costing less than 700 pounds could readily support a high resolution reconstructed image. Since the Christie projectors are heavy it was impossible to run them in portrait mode as with the Infocus unit. This means that the projection is 4:3 aspect ratio with 768 pixels available vertically for the image.

Figure 7.17: Five participants viewing the control condition. For the control the Chromatte was covered with a matt white board and only the central of 5 projectors was turned on. The set-up of the tripods in the room and the limitations on space that this incurred can be seen.

This translates in reality to about 700 vertically and around 200 pixels across for the reconstructed human. The more flexible and modern Infocus projector yields roughly 900w x 1700h pixels for the human image. This is some 10 times more pixels. For this reason, the quantitative performance tests are performed against this projector only. The five projectors were adjusted against the RGB intensity sampling technique established in Chapter 4. This primarily served to balance the Christie projectors against the Infocus unit, but also to check the Christie units against one another. The adjustment meant that the Christie units were turned down to 28 percent of their nominal 4000 lumen output, with the bulbs set to their power saving mode.

Tracking

Vicon Bonita tracking was once again deployed for the experiment. Since the space available was limited it was necessary to use five tracking heads. Experience from the previous experiment highlighted the difficulty of connecting up a separate tracking PC on the available network, with onerous authentication issues, port security, subnet connections, and similar.

Instead, the system was redesigned and optimised with the tracking provision provided by the image generator. The PoE network switch which supports the 6 Bonita optical heads was connected to a second network interface card which was appropriately configured. When forced to upgrade the system for audio sync a microusb 3.0 Ethernet dongle to provide an additional network interface, and this is useful knowledge to anyone attempting to implement the same set-up on a system with only one Nic. This new deployment takes up less space in a theoretical packaged system.

The Vicon Tracker 2.2 software was trained to recognise and track the projectors then output their location relative to the origin as Tracker0, through Tracker4 internally to the computer using a virtual network port and VRPN (Virtual Reality Peripheral Network).

The Cartesian origin of the system was set to the floor in front of the Telethrone chair as per previous experiments by removing the retro-reflective surface from the chair to avoid conflating the return signal. Tracking markers were attached to each of the projectors, with a different pattern on each. The Infocus projector which was to be moved around in the room used 6 markers instead of three to improve the reliability of the tracking when moving. This is shown in Figure 7.16.

This version of the Telethrone chair

Since it was anticipated that there might be more expansive gesturing around the 160°arc of the onlooking group the Telethrone was adapted; making it wider. This was accomplished by spreading the Chromatte cloth over two chairs instead of one. This can be seen in the multiview condition in Figure 7.19.

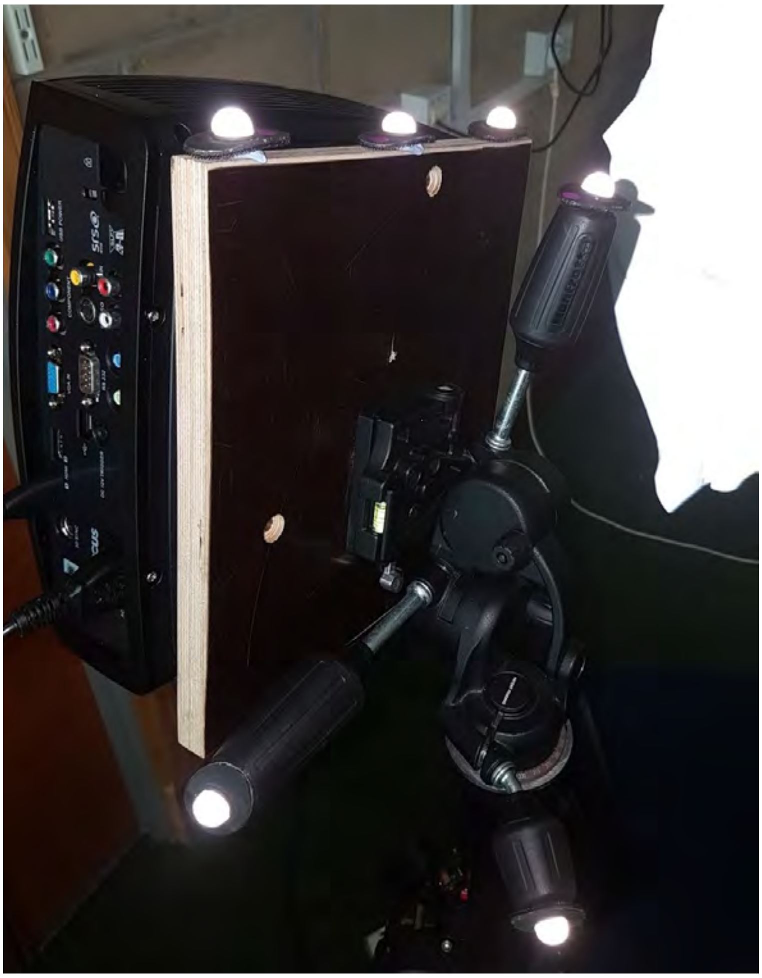


Figure 7.18: The Vicon retroreflective markers on the InFocus 1920x1080 projector show up clearly in the camera flash. This projector delivered 4x the pixel density of the other older units.

Participant Chairs

Onlooking chairs were deliberately limiting. Four legged static theatre style chairs were instead of the more mobile office chairs used in previous implementations of the Telethrone. This helped to keep the participants' heads located under the light cones of the projectors.

Figure 7.19: Five participants viewing the multiview condition. The double width Telethrone can be seen. It is notable that the captured subject appears to be learning forward. In fact they were seated on a stool in Octave and therefore assumed a body position slightly incongruous with the reclined aspect of this version of the Telethrone. In practice nobody mentioned the effect, which was worse for stations 1 and 5 at the outer edges.

Audio sync

The audio sync presented a particular challenge. Audio was recorded using a Senheisser clip on radio mic connected to a Tascam DR-701D 6-Channel Audio Recorder with playback started in Windows Media Player using a script at the same time as the reconstruction. Performance issues discussed in another paper to be published soon led to loss of synchronisation. In the end extensive upgrades were required (as mentioned above in the compute materials) to the playback system in order that dropped frames did not cumulatively lead to asynchronous audio.

Procedure

Participant briefing Participants were instructed to read a briefing, provide consent, and fill in some personal details. They were then taken into the experimental room. They were shown a poster on the wall which listed the fictional attendees at the meeting as an aid to memory. These roles can be seen in Appendix 6 alongside the script for the monologue.

Spatialised playback to group Half of the participants were shown a presentation which rendered a perspective correct viewpoint derived from the tracking data for their seat position. This is shown in 7.19

Non spatialised playback to group The Telethrone surface was swapped for a matt board which did not retro-reflect and the centre projector played back only the centre perspective stream to the whole group. This is analogous to a conventional single camera and screen VC set-up and is seen in 7.17.

Survey All participants were asked explicitly which role they represented at the meeting (building on the hypothetical business meeting roles elaborated in the previous chapter), then they were asked how much they felt they represented each of the five roles on a Likhart scale, and finally they responded to a subsection of Mori's uncanny valley questionnaire (Appendix 9).

Debrief Participants were asked if they had any comments or questions and were shown how the experiment worked.

Recording and Data Capture One of the sessions was recorded on a digital camera and this footage is available on request.
In addition, two phone cameras were synced with a clap then separated to the

Octave and the Telethrone room. These two cameras filmed live streaming of the system and were then clap synced and checked for drift in Adobe Premier. This action allowed later identification of ten movement artifacts in the scene which could act as data points for latency measurement.

7.5.2 Results: Perception of Attention

In answer to the simple question "Which person in the meeting were you?" every member of the multi-view condition group in both experimental runs (n=10) responded correctly, with the exception of station 5 in the second run, who responded affirmatively in the multiple choice to both the Corporate role associated with station 5, and the executive role associated with station 4. The reasons for this are discussed later.

In the single-view condition in which the central station projected a front view of the remote participant onto a board which did not retro-reflect, the results were more complex with a headline failure rate of 50%. The main finding is that every participant in the seats 2 and 4 incorrectly identified themselves as the centre seat.

Statistical Analysis of the Results

All participants were also asked to what extent they felt they represented each role at the meeting in a 1-5 Likhart scale. These answers were 'scored' such that correctly identifying strongly with the role associated with the station would score 5 points, while strongly refuting association with the wrong role could score 5 points for each of the other stations. This question was asked for each station. This gave a maximum of 25 points for each participant (station) in each experimental run. The matrix produced allows score for the accuracy of identification with the role for each participant as well as the level of misidentification with incorrect roles.

The power factor was calculated with GPower to be 0.948 indicating that although the sample size was small there was significant difference in the conditions. This can be seen in Figure 7.20.

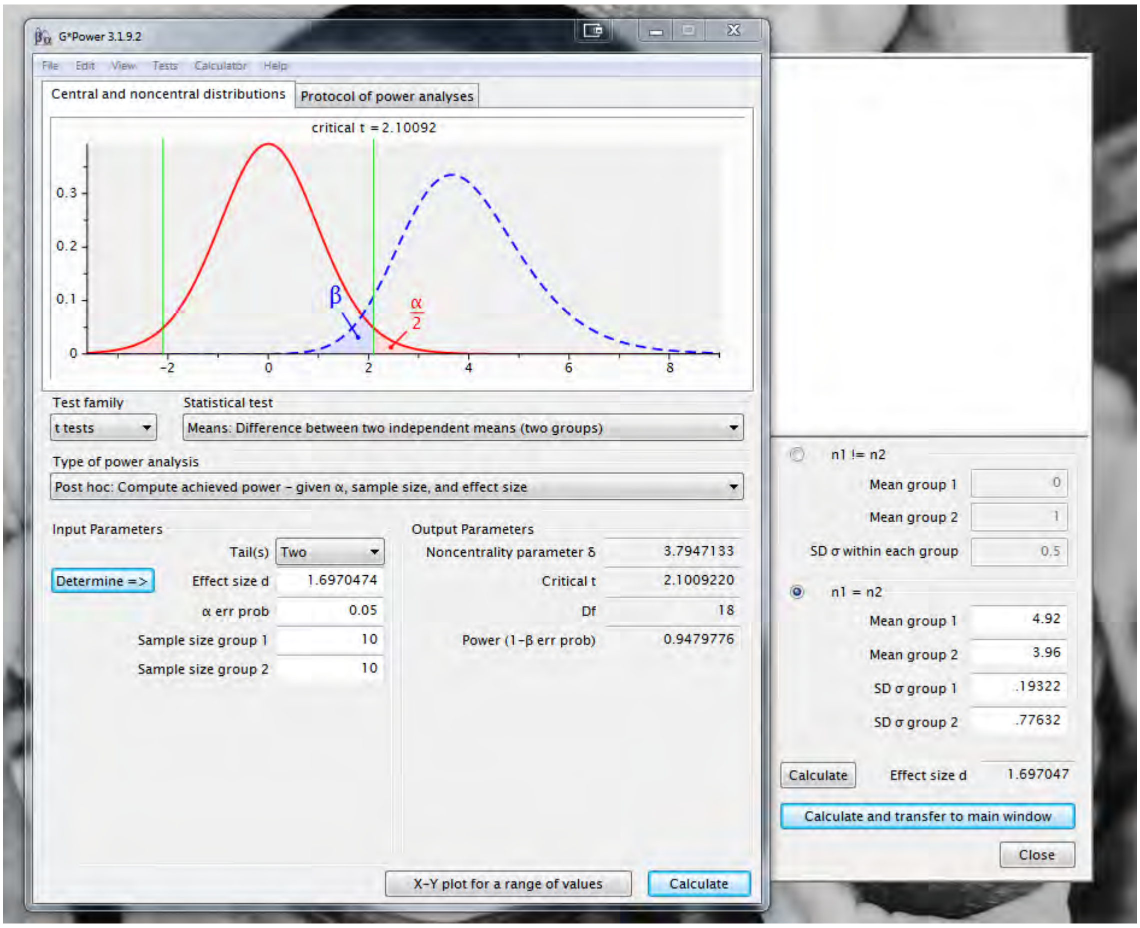


Figure 7.20: Gpower analysis for the dataset. The post-hoc power analysis shows that there is significant different in the conditions even with the small sample sizes.

A boxplot of the two conditions further illustrates the difference in the scored responses in Figure 7.21.

It can be observed that the small sample size leads to large differences in quartiles since the effects of a single choice in each station are proportionately large. Taking a geometric approach to the analysis suggests that it is appropriate to combine the results for the outer station 1 and 5, and also combine 2 and 4. This somewhat

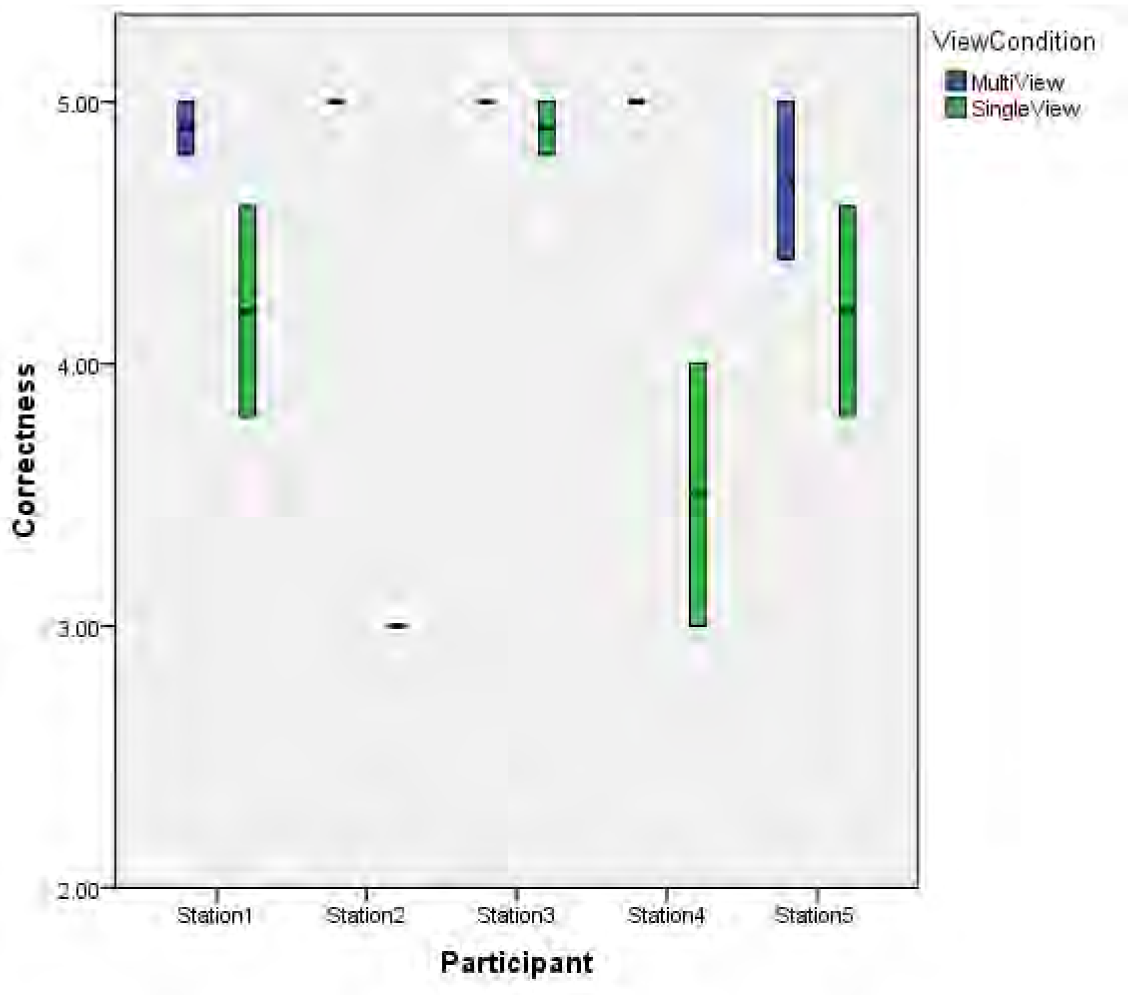


Figure 7.21: For each of the stations 1 to 5 the synthesis of all of their answers can be seen plotted here for both conditions. Stations 2,3, and 4 were in no doubt about who they were across all the questions, so their 'bars' are seen as blue lines at the top. Stations 1 and 5 had a little doubt about perhaps being the adjacent role. 5 leaned far out from her chair to the left and was most confused. Nonetheless when pressed by a single question they all answered correctly. In the green multiview condition only the central station 3 identified their role correctly, with all others having different degrees of confusion about who they were supposed to be in the meeting.

smooths the data and shows the level of correct identification from the point of view of the stations. It is important to note for instance that the centre has a very

Table 7.2: Degree to which participants correctly identified their role

Particpant performance	Multiview			Singleview		
	Outer	Inner	Centre	Outer	Inner	Centre
Average correctness	4.8	5	5	4.2	3.25	4.9
StDev	0.52	0	0	0.31	0.49	0.32

low score not because the centre participants incorrectly identified themselves in the centre role, but because other participants in the outer seats thought they were the centre role.

7.5.3 Other results: Software Performance

From Disk

Reconstruction from disk is now 30FPS thanks to improvements in the playback code as compared to the previous system. The playback is now time stamped in order to correctly synchronise to the audio playback. While this is an important development for the software system it will be reported in another forthcoming paper.

End to end

We demonstrate end to end 3D video in line with previous implementations to 5 tracked chairs. The end to end system combines all the elements which have been engineered so far within the Telethrone research and also expands on the previous related 3D video research from the wider group. This was a significant undertaking but was a collaborative effort with the other researchers in the group and will be reported elsewhere in detail.
As in previous versions of the software [94, 273] the framerate was low compared to the 20FPS which is deemed acceptable for Skype [273]. The new implementation achieves 4FPS which is lower than the 10FPS achieved using withyou.

Changes to the visual hull subsection made for CROSSDRIVE and Telethrone make the current implementation far more stable compared to withyou, but slower. Further optimisations are possible but outside of the scope of this paper and will be explained in a forthcoming publication.
Latency was Mdn 1.02s (StDev 0.06) across ten samples, which compares favourably with the 1.5s achieved in CROSS DRIVE [94].

7.5.4 Discussion

Acute Viewing Angles

The two zones at acute viewing angles to the cloth have been touched upon in the possible arrangements and in the test of 66.5°eye gaze (which is in the 30°at the edge). The experimentation so far suggests that as the viewing angle onto the surface becomes more acute it is harder to make out the attention of the projected participant. This chapter seeks to explore this in depth.
Testing projection at the oblique limits of the capability provided additional insights and suggest areas which could be improved. The projection of the legs becomes far more distorted for instance. A suggestion for what might be done to solve this is provided in the closing chapter.

7.6 Conclusion

The novel contribution of the chapter is the presentation of version 2 of the Telethrone system, with tracked 3D video augmenting the spatial segmentation in experiment 2, as well as an experiment which builds toward a deployable system in experiment 3.

In the experimental results the best case scenario in experiment 2 compares well against “Estimating the gaze of a virtuality human” in the similar R0L condition, and against Telehuman in the ‘Look at’ condition. This satisfies H3.1 and clears the way for further integration with the Telethrone prototype and a behavioural test with multiple participants.

In particular is it suspected that the better texturing approach gave higher accuracy and smaller inter quartile ranges over previous experiments.

The worst case scenario satisfies the assertion that the accuracy would be compromised with deflected eye gaze and cross talk. While this data demonstrates less accuracy it is by no means a broken system for the support of directional gaze. Results are less and 4°worse than Telehuman.

Telethrone appears to support small ad-hoc group meetings with dynamic participation and subgroups within a group in experiment 3. The simplicity of the system, it’s affordability, flexibility, and scalability seem to be appropriate for high traffic social spaces which is a less researched application for telepresence displays.

Chapter 8

Evaluation, Discussion & Conclusion

8.1 Chapter Overview

This chapter first looks at how the literature has informed the development of the Telethrone system, which aims to connect small groups over a distance in a less formal setting, and then briefly summarises how the research has impacted back on the literature.

Opening the discussion section is a summary of the arc of development. This reiterates the initial aims then the summary review relates the journey of the investigation, showing how the lessons from each iteration of the methodology informed the following steps. The evolution of the research over the seven years is best captured throughout this summary.

The methodology is then reviewed and differences between the planning and the actuation are considered.

The discussion then directly lists the experimental hypotheses against the results and objective measures and attempts to qualify the degree to which they were confirmed or otherwise.

A more wide ranging evaluation opens the scope of the analysis, highlighting the more important threads of the research, and including some conjecture and opinion.

Contributions and shortcomings, though discussed in-line with the sections of the chapter are then summarised briefly in their own section. This informs the options and opportunities for expanding the research in the future.

8.2 Literature Review

8.2.1 Impact of the Literature

It was clear from the literature review that attention is mediated in a large part by non-verbal cues. A large body of research across psychology tells us that Extra Verbal Content is important [120] for mediating attention [311, 286, 309, 72, 116, 33, 299, 10, 142, 241, 336], conversational flow, and transmission of mood and manner.

Support for these cues is limited in conventional commercial systems which convey the image of a person over a distance through technology.

Eye gaze, mutual eye gaze, and facial expression [129, 10, 230] are particularly rich information channels, but Telethrone also transmits body torque, gesture, and head gaze. Proxemics shows us that interpersonal distance is an important feature of communication and can evolve as situations change [130].

Telethrone has been designed and tested against its ability to better support connection of non-verbal cues across this telepresence boundary.

It is especially difficult to support mutual eye gaze well because of the Mona Lisa effect where a viewer of a screen showing a camera view of a remote participant will only perceive that they are being looked at when the remote user is looking directly into the camera lens which captures them. Fixing this disconnect of mutual gaze in telepresence systems is an active area in both research and commercial system. The 3D video approach used in the later versions of Telethrone reflect some of this research.

Telepresence solutions aimed at collaborative working have traditionally attempted to join remote spaces so that people in each can look into the other, seeing each other, at best, as if through a window. Immersive Collaborative Virtual Environments (ICVE) have been used to join remote spaces so that they coincide. Free-viewpoint video has been combined with immersive projection technology to allow people to move around spaces, seemingly together [273]. For these spaces to be aligned so that they coincide, removing the impression of being separated by glass, users must wear stereo glasses.

More recently, situated displays have attempted to give the impression of placing people within each other's room, but either lack the spatial faithfulness available within ICVE, or else do not render the whole of the remote collaborator.

Situated displays can place a representation of the remote user within a space. Most informative to the research at the time of the literature review was the work of Pan et al. [240, 244, 242, 243]. Paul Sermon explored situating live telepresence video onto furniture, most notably beds in his work at the University of Salford [289].

Co presence and social presence are the degree to which a user of a system feels they are 'in a place' with a remote user. This is an important feature for Telethrone since it gives an indication of how situated and natural the system is.

The bulk of commodity systems for telecommunication remain dyadic, that is, person to person. More interesting and difficult is small group interaction, so-called 'teleconferencing'. This is a difficult field, with known issues in all of the market leading systems. Mona Lisa effects, latency, jitter, and the requirement of a lot of expensive equipment to be maintained alongside a booking system for its use amongst the most problematic.

Inami pioneered the use of RPT [149] and this was built on by others, most notably Tachi [316, 317].

Tachi's projection of a face onto a robotically manipulated surface is the closest analogy in literature to the Telethrone system but is not a situated display. Krum et al's REFLCT system [177] has some similarity to the Telethrone in that observers move around to gain viewpoint independent and spatially segmented views on a retro-reflective surface, but the application is very different.

Thus none of the above methods allow people in different spaces to use a wide range of contextualised non-verbal communication (NVC) when addressing less formal groups in a workplace context. Some give a true representation of the whole face but restrict movement, while others allow movement but abstract the physicality of the remote participant (SecondLife).

Table 8.1: Telethrone offers additional affordances

Systems	Joint Eye Gaze	Viewpoint independent / Reconstructed	3D stereo	Multiple independent viewpoints	Natural Setting	Weird / Uncanny	Commodity components	Able to move around
Skype[215] Facetime	No	No*	No	No	Somewhat	Somewhat	Yes	No
Realpresence Centro & Tandberg	Some	No	No	Sort of**	Yes	No	No	Yes
Cisco TP Rooms[315]	Some	No	No	No	Yes	No	No	No
Sphere Avatar[240]	Yes	Yes	No	Yes	No	Yes	No	Yes
Telehuman[167]	Yes	Yes	Yes	No	Some***	No	Some	Yes
Immersive Group to Group[29]	Yes	Yes	Yes	Yes	No	No	No	Yes
Withyou[273] blue-c[127]	Yes	Yes	Yes	No	No	No	No	Yes
Tripleview[54]	Yes	Yes	Yes	Yes	Yes	No	No	No
Room2Room[248]	Yes	No	No	No	Yes	Some*	Yes	No
Encumbrance Free [198]	Yes	Yes	Yes	No	Some**	No	Yes	Yes
Holovizio Lightfield [323]	Yes	Yes(lightfield)	Yes	Yes	No	No	No	Yes
Telehuman2 Lightfield[121]	Maybe	Yes	Yes	Yes	Some	No	Some	Yes
Telethrone[228, 229]	Yes	Yes	Maybe	Yes	Yes	No	Some	Yes

* Viewed from oblique angles the Room2Room projection can be distorted by the furniture onto which it is projected.

** Addressing a screen at the end of a desk in similar to commercial systems and can be made to look fairly natural with enough integration

In examining current telepresence technologies various pertinent systems are detailed, and their affordances highlighted against one another best in Table 8.1. In the literature review no comparison was made to Telethrone, so the table is repeated here with Telethrone in place to show how the research adds to knowledge. Also included are systems which came to light after the completion of the initial literature survey. In this updated version of the table, the relevant papers are highlighted with references.

8.2.2 Point of Departure

The point of departure builds upon previous research knowledge in the wider team concerning attention within small groups when those groups are partly or fully mediated by different levels of technology. This is explored in more detail in the declaration at the start of the thesis but to recap in summary: The work of Paul Sermon [289] and a decade of research conducted by Prof David Roberts [217, 80, 81, 210, 211, 273, 93] and Steed at UCL [106, 124, 138, 213], provided a rich basis for the technical inquiry. Grau from the BBC introduced 3D reconstruction to the group in 2007 and Thomas from the BBC introduced the retro-reflective fabric in 2009 as a potential method for spatially separating multiple channels.

This concept had previously been explored by Tachi [316, 317] and others.
David Roberts had suggested reconnecting eye gaze through a chair and Thomas suggested the use of the Chromatte fabric on furniture. Roberts thought this might create a situated display based on his long experience with Steed. Roberts suggested a system which might build upon the ability suggested by Sermon for joining a group discussion 'ad-hoc', and suggested that Telethrone might be tracked somehow to maintain spatially faithful relationships to the group. Bowden at Salford and Weir at the BBC did some unpublished experimentation, while Roberts, Duckworth and this researcher tried some different combinations of materials.

8.2.3 Revisiting the Literature Throughout, a Summary

Suggestions from peer-review (from the many submissions to outlets), new research published through the course of the investigation, and additional discoveries, informed the work and are included here rather than being engineered into the literature survey chapter. The exception to this is the investigation into meetings and fault lines which sits better in the previous chapter.
The best single review on mutual gaze in telepresence happened after the initial literature review but is worth noting here. It is by Regenbrecht in 2015, and notes that there is still no perfect solution to the problem [261].
Just prior to Internal Evaluation Microsoft Labs announced that they had been researching projection onto furniture using their Kinect system and projectors. This Room2Room system [248] is seen in Figure 8.1. This lent weight to the assertion in this research that projection onto furniture is an exciting and important area. Their system provides spatially correct viewing through reconstruction but only on a point to point basis with a single user at each end. Telethrone still offered significant advantage through its support for multiple viewpoints.

The Holovizio real-time 3D light field transmission system[323] was raised during peer-review and is added to Table 8.1.
Maimone et al. presented their "Encumbrance free telepresence" [198], which was

Figure 8.1: Microsoft Room2Room projects a reconstructed remote user onto an arbitrary surface, in the case of this image a chair. This technology could be blended with Telethrone in future research.

not captured in the initial review but is now added to Table 8.1. It can be seen that this system provides many of the desired affordances but falls short on being able to support small groups. A blending of this system with Telethrone might be desirable but is beyond the scope of this work. During the peer-review process the "SCAPE" system [146] was highlighted. This comparatively complex system combines immersive projection from head mounted projectors (onto surrounding walls and objects) with tracking and virtual avatars to create multiple viewpoints into a shared virtuality.

Ad-hoc in SecondLife

It is a significant shortcoming of the research throughout that ad-hoc meetings in SecondLife were not captured until the end of the research. This is undoubtedly due to a lack of regard for this approach by the researcher, causing the body of work to be inappropriately excluded. There is no other reason than personal preference to this choice and this is a limitation of the research.

8.2.4 Impact on Literature

Two experiments have been published to date. [228] [229]. A third is planned discussing the support Telethrone offers to small meetings and the test of that capability through integration and modification of the CROSSDRIVE codebase.

8.3 Discussion

8.3.1 Original Aims

The aim of the research was to test the suitability of Chromatte cloth when repurposed to support telecollaboration with a small informal business group.
In order to do this it was necessary to gain some understanding of communication within group meetings when they are mediated partly by a telepresence technology boundary.
This meeting paradigm was identified as a specific challenge by BBC R&D, and they provided some initial steer as to a possible solution. They had found that with split site working they were not able to leverage the best outcomes from formal meetings within their R&D department as they no longer found themselves in serendipitous informal meetings. By their own estimates they stated that ten so-called 'water cooler meetings' were required to totally follow up points from a given formal meeting.
As the research developed the focus shifted slightly to accommodate some additional novelty, integrate new knowledge, and to fit to reality of constraints in the development. This section gives an overview of the objectives, and their ongoing impact on the investigation. This forms the story of the research, which will be evaluated in the discussions which follow.

8.3.2 Summary Review of Objectives and Questions

Initial Exploration

When the research was begun there was little to no consideration of the field of RPT as outlined throughout this thesis. Concurrently with the literature review it was necessary to understand the retroreflective material (Chomatte). There is a strong case for this being an unnecessary step, as much work had already been done in this area, and to an extent this 'zero' objective, the research question, and the hypothesis are therefore redundant. This is a shortcoming of the research approach with hindsight. Nonetheless lessons were learned and integrated, and at the time this felt like valuable exploration. The toolchains developed to measure reflected light values were used throughout the experiments to calibrate projectors against one another. Also, it is appropriate to declare the avenues which were unsuccessfully attempted.

O0: With a view to building a novel tele-presence system which can support ad-hoc meetings; test the physical properties of Chromatte cloth, cameras, and projection systems to determine suitability in experimental context for use in spatial segmentation. Trial techniques such projection mapping and adaptive software masking which might more accurately apply the projection to the complex surface of the chair. The testing must give confidence that it is possible for observers to perceive a view of a projected remote person that is correct for each of them separately.

As a prelude to the research proper it was necessary to test a few elements of the system in isolation to establish that these prerequisites warranted further investigation. Most pertinent was how the Chromatte cloth responded to projection of images when draped on a chair. Both the manufacturers data sheets and previous RPT research suggested that it was appropriate, and this was born out in the main, with some residual concerns about the 5% of light which scattered rather than retro-reflecting. In addition, software masking of the projected image, and image warping were discounted due to latency concerns.

Informing and adjusting the research After getting a feel for how the cloth responded it was discovered that the human eye still discerns the image of a person even at the reduced 5% light levels which are always scattered regardless of viewing angle. It was deemed necessary to address the potential for experimental participants being able to see a ghost of different angles overlaid on the image angle they were supposed to see. Examining the literature did not yield any exact fits for this and so a test for a corruption of the 'likeness' to a human face was sought, with Mori's uncanny valley research seemingly a good fit.

IP Camera based Telethrone

The first experiment reflects the knowledge at the time of the experimental design. It was thought that IP cameras would give less latency for an end to end system over a network, so this was attempted first.

O1: Test if a remote participant brought into the space by the system is for some reason excluded from triadic poker play. Primarily seeking evidence for (statistically) significantly fewer looks and duration of looking toward the remote player, with additional examination of social engagement through questionnaire feedback.

In the first experiment there was good indication that the telepresent participant was not excluded from the structured poker play. A careful examination of 'look events' (any looking from one participant toward another) was made, and found that in fact the Telethrone system drew more attention, though the reasons remain unclear. Questionnaire responses presented a mixed message with a detectable perceived difference between the physically present participant and the tele present participant, but encouraging social indicators summarised in more detail in the next section.

Informing and adjusting the research The first experiment suffered from angular offsets of the captured gaze in both directions. The camera next to the 60 inch TV was around 5°to the side of the eyes of both the filmed participants in the observation suite, while the camera above the Telethrone which returned the image to the TV was nearly 15°higher. There was a feeling that this lack of mutual gaze availability was contributing (alongside the choice of task employed in the experiment) to the lack of mutual gaze events, and also potentially to the additional attention toward the Telethrone participant (as participants perhaps instinctively tried to resolve the ambiguity in the gaze direction of the remote collaborant). To address this experiment with 3D reconstruction techniques was undertaken.

Novel Support for New Affordances

It is difficult to establish exactly where the idea for dynamic meetings occurred, or more properly it is difficult to define exactly where the idea changed the flow of the investigation. In attempting to better align eye gaze though reconstruction some early testing work was attempted, and this engendered a new line of inquiry.

O2: Explore how the system might scale, and how it might support novel sub-groups within a meeting.

While designing the first tracked and reconstructed system, which presented a 'static', but viewpoint correct avatar on the Telethrone, it was noticed that ability to move fluidly around the space while being tracked (and observing the Telethrone) was itself extremely interesting. Some consideration of how this might be useful suggested a novel affordance for meetings with a telepresent component. Rather than just pulling up a chair as suggested by Sermon and Roberts, what would be the impact to the meeting of moving around as the group evolved over time? Investigation of meeting science (as explained in the 'Business Meeting' section above) and a revisiting of the telepresence literature suggested that this challenge had not been investigated outside of avatar based systems like SecondLife.
While it is true that in many formal meetings the seating positions are static over time, it seemed that in the posited problem which the BBC brought to the group,

there was opportunity for members entering and leaving the meeting, and that different fault lines or subgroups might come into play as this happened. Some investigation of this potential affordance suggested that this could be supported in principle by the optics of the Chromatte cloth.

Informing and adjusting the research Integrating the capability of dynamic meetings happened in parallel with putting reconstructed viewpoints into the system. The third experiment addressing Q3 (that of five tracked chairs) contains the synthesis of this idea with the ability to correctly render a viewpoint through tracking in that all the stations are tracked in real time. A large enough space was found to investigate both problems with the same set-up.

Reconstruction Based Telethrone

Shortcomings in the results from the first experiment seemed to suggest that it was very important to reconnect gaze vectors through the system. Looking at the surrounding research with hindsight this conclusion could be seen to be inevitable. This is a shortcoming of the research plan, with concerns about latency taking precedence over suggestions from literature that the system would be flawed without proper correction of the eye gaze.

O3: Integrate previous research (3D video reconstruction) which can present a representation of a remote participant such that the observer can walk around the projected image, continuously seeing a viewpoint which is generated for their position. Test this experimentally. Test this reconstructed viewpoint system in the context of a small meeting experiment.

In order to begin to address the offset from the IP based system the decision was made to use 3D reconstruction, as in other research systems, to allow the model which represented to remote user to be manipulated such that the gaze of the remote collaborant was aimed toward a new virtual camera, aligning gaze vectors through

the system.
The second experiment used a cut down and limited version of the Octave reconstruction system to test if participants could correctly estimate the gaze of a rendering of a remote participant, in a well lit social space, with some of the cross bleed effects from the multiview in play. In this second experiment there was indication that application of the reconstruction system could enable accurate mutual eye gaze on the Telethrone system.
The third experiment presents evidence that it is possible to apply the Telethrone technology to a suitable size of group to address the research aims while using only a single graphics workstation. This is a novel contribution in that the reconstruction system developed by the group has never been scaled in this way before. The system employs tracking on the same workstation to support repositioning of the chairs, supporting the novel affordance of reconfiguring the meeting (as discussed in principle). Both playback from disk of a recorded session, and end to end capture and playback on Telethrone, in real time, were demonstrated. This last, the real time end to end was a huge body of work undertaken in collaboration with the previous CROSSDRIVE team, and it is a shame that there was insufficient time to publish and reference this work. The collaborative nature of the effort means that it is currently in the thesis in Appendix 1. Meanwhile the playback from disk experiment was extremely encouraging, with non-verbal cues successfully resolved by all participants while a control showed confusion typical to current commercial systems.

8.3.3 Review of Methodology

The degree to which the research followed the methodology can be judged in the following pages.
First let us consider the original methodology again in Figure 8.2. The right hand side of the iterating methodology corresponds to the vertical swimlanes in the review of work done in the following pages. After each run down through the right hand side of the methodology diagram the literature is reviewed to inform the onward research questions on the left hand side of the methodology diagram,

before the next cycle of knowledge creation and testing by experiment. It can be seen that the knowledge creation swim lane encapsulates all the left hand side of the methodology diagram. If the methodology were followed then the overall flow of work should move back and forth regularly across the swim lanes, iterating through the research questions. This is mainly true, though there was some concurrency in practice.

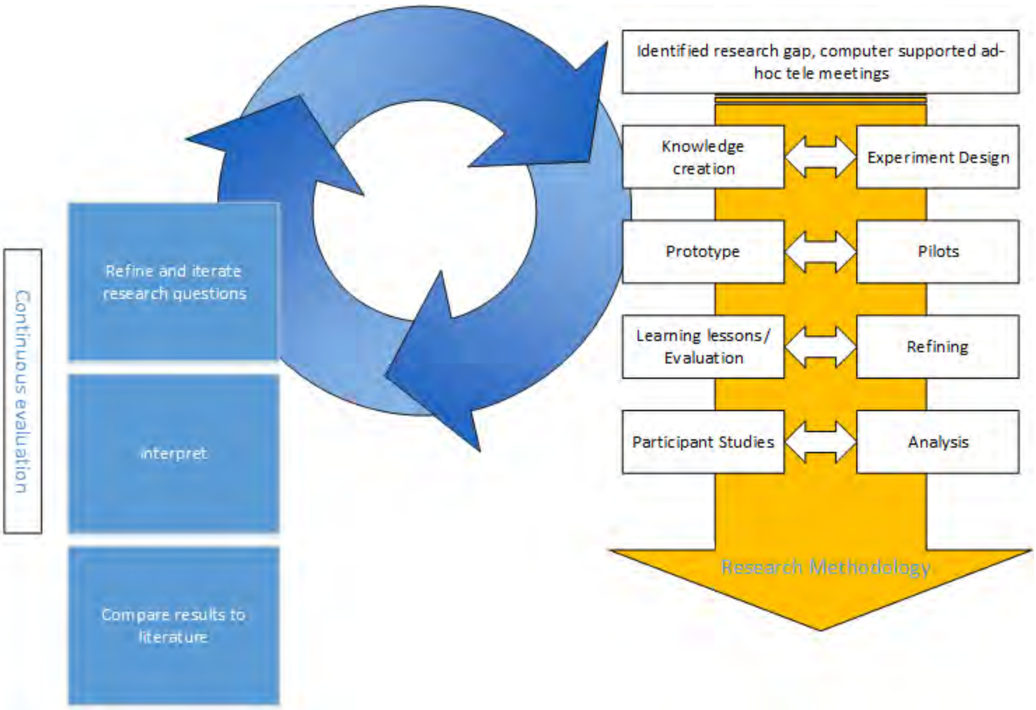
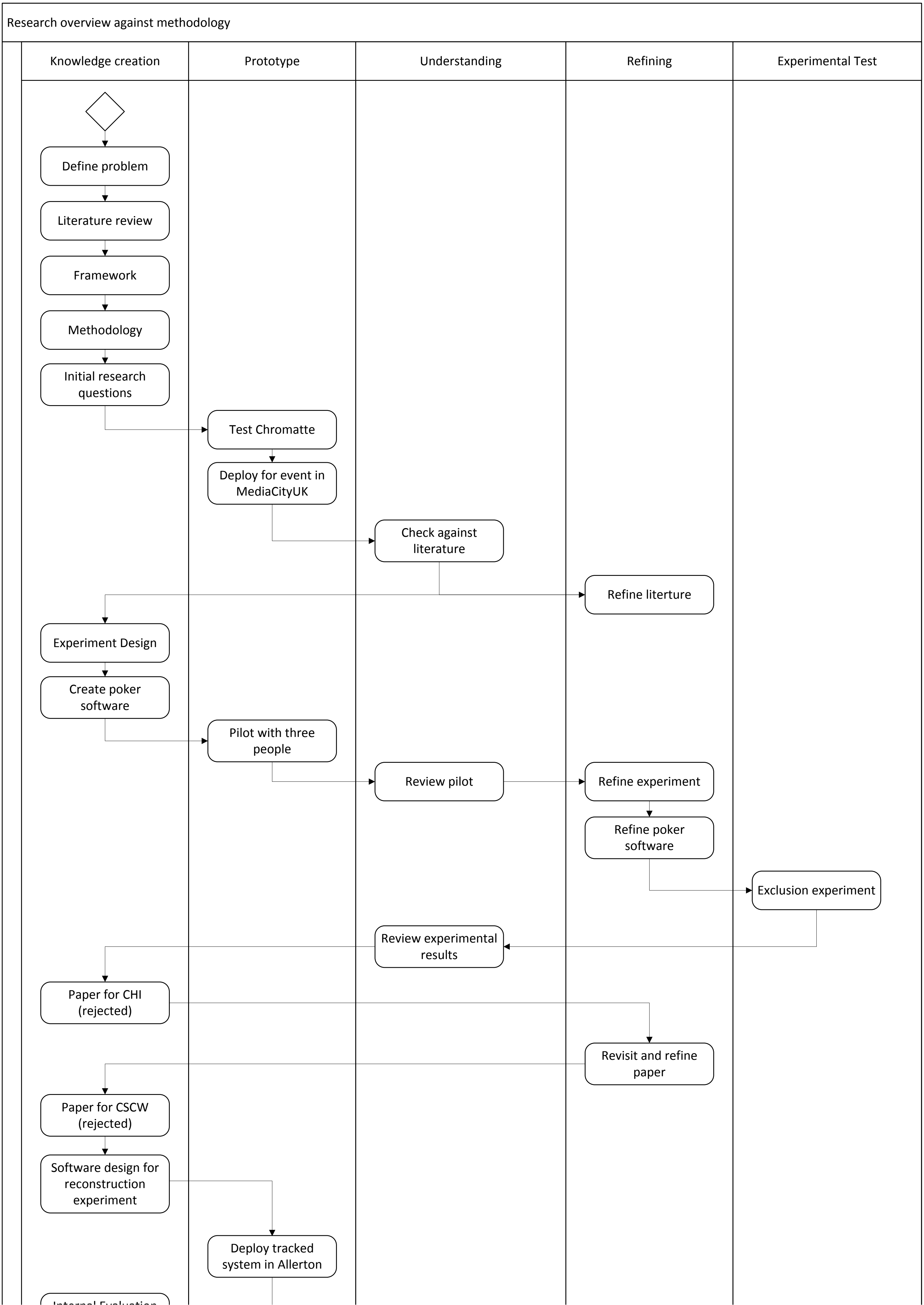
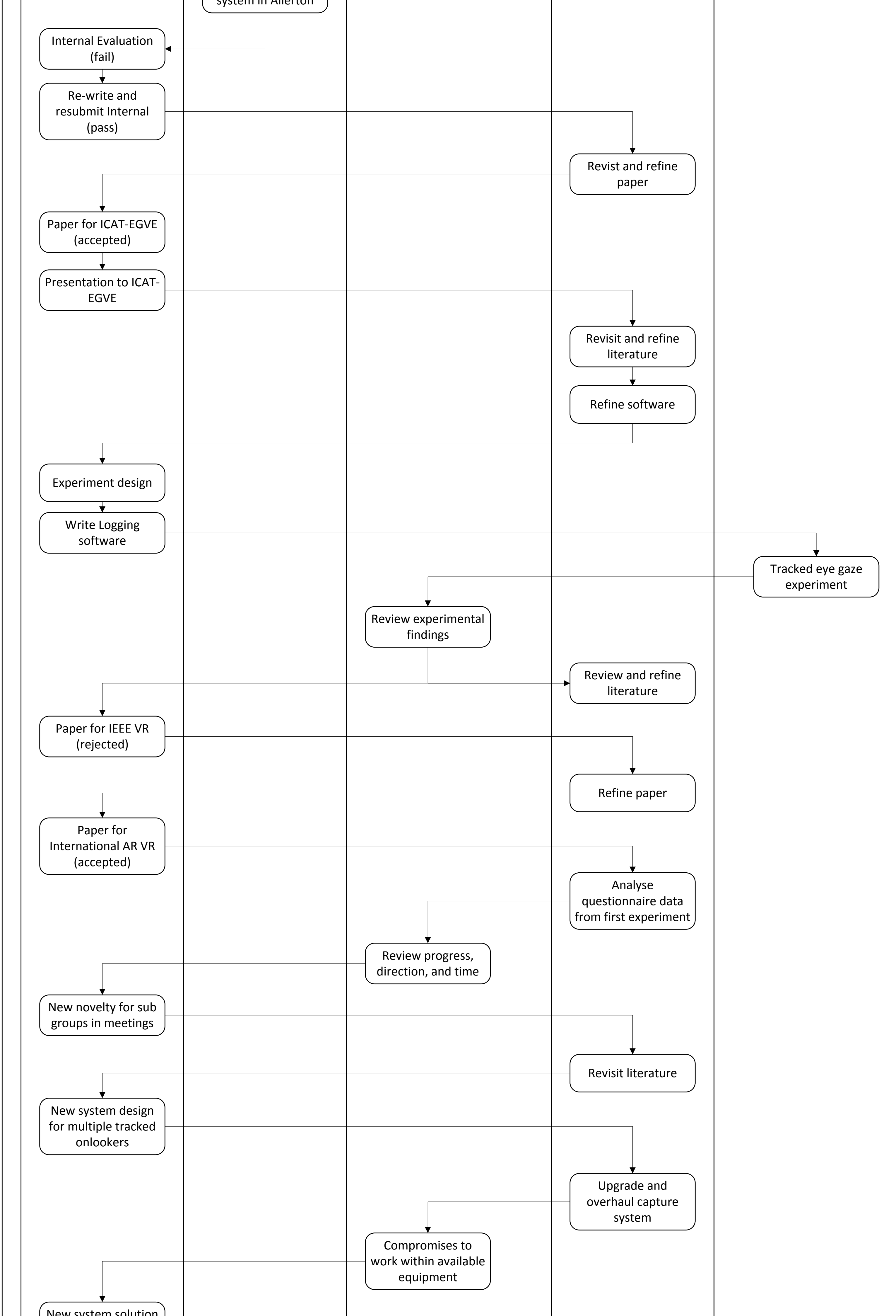
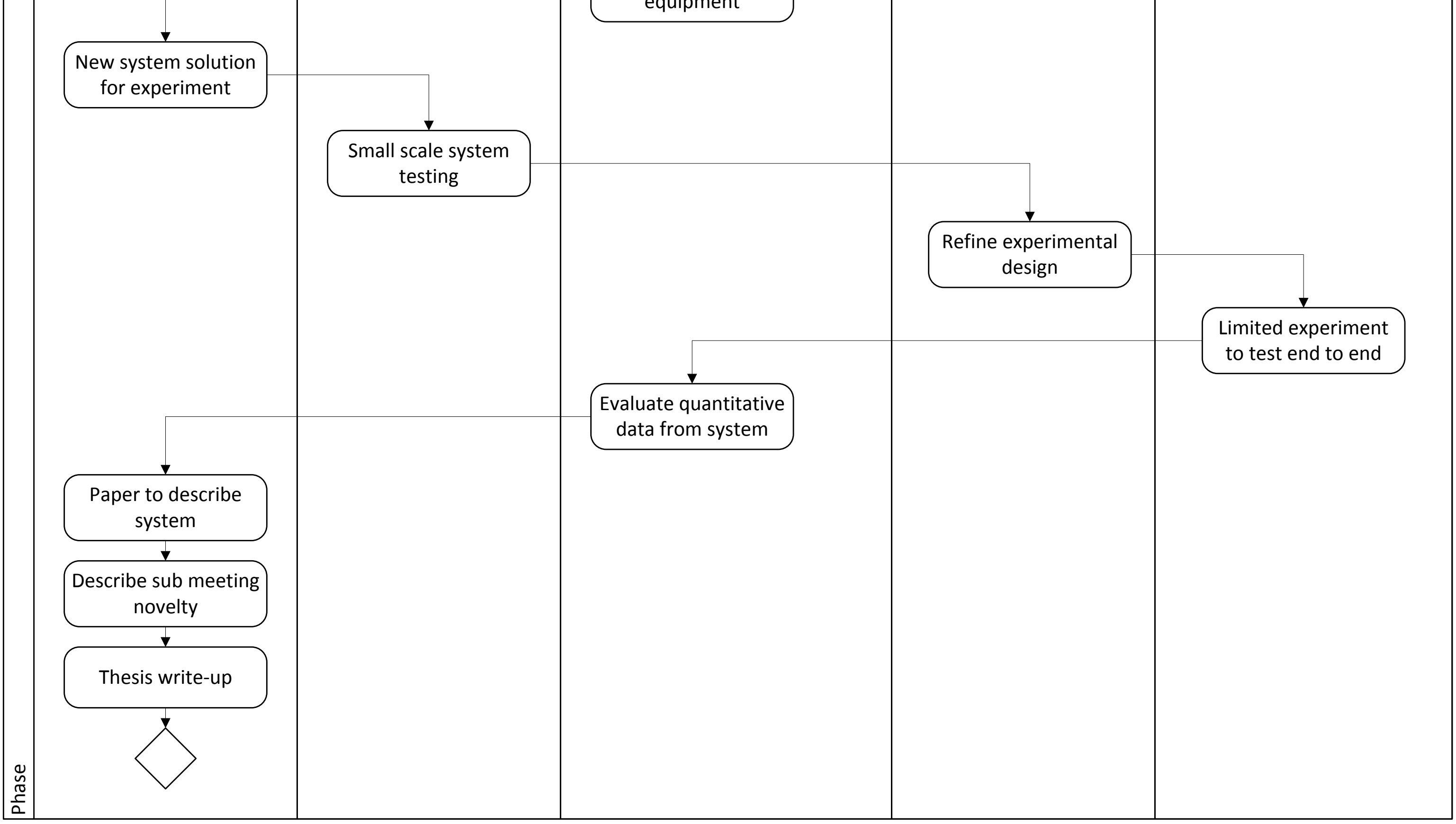


Figure 8.2: Methodology interacts with the technology side of the theoretical framework Venn diagram.


Research overview against methodology
Knowledge creation
Prototype
Understanding
Refining
Experimental Test
Define problem
Literature review
Framework
Methodology
Initial research questions
Test Chromatte
Deploy for event in MediaCityUK
Check against literature
Refine literture
Experiment Design
Create poker software
Pilot with three people
Review pilot
Refine experiment
Refine poker software
Exclusion experiment
Review experimental results
Paper for CHI (rejected)
Revisit and refine paper
Paper for CSCW (rejected)
Software design for reconstruction experiment
Deploy tracked system in Allerton


system in Allerton
Internal Evaluation (fail)
Re-write and resubmit Internal (pass)
Revist and refine paper
Paper for ICAT-EGVE (accepted)
Presentation to ICAT-EGVE
Revisit and refine literature
Refine software
Experiment design
Write Logging software
Tracked eye gaze experiment
Review experimental findings
Review and refine literature
Paper for IEEE VR (rejected)
Refine paper
Paper for International AR VR (accepted)
Analyse questionnaire data from first experiment
Review progress, direction, and time
New novelty for sub groups in meetings
Revisit literature
New system design for multiple tracked onlookers
Upgrade and overhaul capture system
Compromises to work within available equipment
New system solution


equipment
New system solution for experiment
Small scale system testing
Refine experimental design
Limited experiment to test end to end
Evaluate quantitative data from system
Paper to describe system
Describe sub meeting novelty
Thesis write-up
Phase

By far the biggest issue with the Methodology was the initial plan to iterate through the literature in tandem with the experiment designs. This caused the first experiment to be designed and started before the bulk of the literature review was assessed. This is a predictable problem looking back, the urgency to get started early on meant that time was wasted proving things which were already well established in literature. The time spent looking at the material could have been avoided. The first experiment with offsets between camera capture and the projected eyes should have been done differently. It would have been much better to spend a lot longer on trials and basic research, moving into the reconstructed tracked projection later.

8.3.4 Review of Hypotheses

The following section is a presentation and critique of the tested hypotheses. The objectives and questions previously discussed form the framing for these, with the ordinal references the same throughout the thesis, though there are sometimes multiple hypotheses generated per objective/question pair. For this reason, they will not be explicitly framed here. Evidence is summarised from the chapters where it is deemed to support or refute a hypothesis but evaluation and expansion of the discussion will be discussed against the same ordinal set in the next section.

H0: In Telethrone Chromatte cloth is draped loosely onto a chair, such that it has many deformities. The manufacturer of the material recommends that the cloth be flat and smooth for their intended use in chromakey. When an image of a person is projected onto this draped form of the cloth for use in Telethrone, it will similary retro-reflect light back along the axis of projection. Specifically the angle subtended before 5% of light remains detectable will be approximately 15°to the side of the projector (horizonally). This corresponds to the 'spatial' element from the centre of the theoretical framework as confirming this hypothesis enables spatial segementation using the material in this novel context.

Hypothesis 0 (testing) is supported quantitatively but found that the residual 5% is visible to the human eye, informing later experiments.

H1.1: Natural conversational attention between static seated participants, judged by eye gaze, is not attenuated in some way by any subtlety of the system such that it can be detected statistically or else be consistently highlighted in questionnaire responses. This corresponds to the 'gaze' element of the theoretical framework.

In the multi-view projection condition the number of gaze events toward the Telethrone were significantly greater (*Mdn* = 33) than the co-located researcher (*Mdn* = 22.35), T = 22, $p = .017$. In the single-view projection condition there were more gaze events toward the Telethrone (*Mdn* = 31) than the co-located researcher (Mdn = 22.35), T = 113, $p = .020$.

In both projection conditions there was more gaze duration toward the Telethrone than the co-located researcher. In the multi-view projection condition the total duration of gaze toward the Telethrone was significantly higher (*Mdn* = 29s) than for the averaged co-located researcher (*Mdn* = 23.28), T = 23, $p = .020$, while for the single-view projection condition gaze duration toward the Telethrone was significantly higher (*Mdn* = 39.88) than for the averaged co-located researcher (*Mdn* = 23.28), T = 92.5, $p = .012$.

Hypothesis 1.1 is strongly supported and unexpectedly revealed that Telethrone attracts significantly more attention.

H1.2: The multi-view condition demonstrates more natural looking behaviour than the single-view condition as examined through eye gaze events and questionnaire. This corresponds to the 'attention' element of the theoretical framework.

In comparing projection conditions there is no significant difference between gaze events toward multi-view (*Mdn* = 33) vs gaze events toward single-view (*Mdn* = 31), T = 57.5, $p = .587$. There is no significant difference between single-view and multi-view for the number of look events when compared to one another, T = 78.5, $p = .587$.
There is no significant difference between single-view and multi-view for look

duration when compared to one another, T = 67, p = .959.

Hypothesis 1.2 is null leading to questions about the design of the system which led to further development.

H1.3: The Telethrone is not reported as anomalous in some way when compared to a real person when examining social presence and uncanniness through questionnaire. This corresponds to the 'situated' element of the theoretical framework since support for this hypothesis indicates that the display is operating in context.

Here results are mixed but generally very positive:

On the negative side from the Networked Minds responses: Participants did regard the Telethrone confederate as being viewed significantly more through a screen, and to be significantly more occupying a different space to themselves. They also thought that the Telethrone participant was significantly more in another room. How much these fell short of 'the real' will be discussed later as it is a judgement call.
In asking how 'obvious the participants felt they were to the confederates' there is a significant difference between the Telethrone and the co-located.
Participants felt somewhat strongly that they were in the room with 2 people (Mdn=5) and very strongly that they were playing with two other people (Mdn=7). They also felt that the Telethrone confederate interacted naturally with the co-located confederate (Mdn=6).
There were many responses from the questions which further support the hypothesis: Comparing how much the participant noticed the Telethrone and the co-located confederate there is no significant difference in the relevant multi-view condition. Comparing how much the experimenters were perceived to notice the participant there is no significant difference in the relevant multi-view condition.
Asking how obvious the confederates were to the participants there is no significant difference.
The participants felt that both confederates caught their attention equally.
Participants felt the same level of distraction from both confederates..
Nor did the participants feel that the Telepresent confederate was more distracted

from them than the co-located participant.
Participants perceived that both Telethrone and co-located confederates believed the participant's thoughts were equally clear and that the confederates thoughts were equally clear to the participants.
The participants found it easy to understand both confederates (Mdn=6) with no significant difference between the two. The participants further felt that both confederates could understand them, with no significant difference between the two.
There was no significant difference in the ability of participants to tell how either confederate felt, nor their perception of how confederates judged the participants felt.
Participants did not believe that the tele-present confederates emotions were significantly more unclear than that of the co-located, and their impression of how clear their emotions were to the confederates were not significantly different.
Participants felt that they would be able to describe the feelings of both confederates equally well, and that they felt confederates would be able to describe their feelings equally well.

It would seem that the hypothesis is broadly supported, though there is some feeling that the display shows another space, and is somewhat like a screen. This informed later design in that the following implementation 'segments' out the room behind the remote participant.

H2: In less formal meetings dynamism and flexibility are more appropriate. Conventional formal tele-presence systems poorly support this. In contrast the Telethrone will allow five distinct spatial display segments of the Telethrone even if the onlooking seats are arranged further out, in smaller dynamic sub groupings. This corresponds to the 'informal' element of the theoretical framework.

Hypothesis 2 is somewhat supported as paper exercises without experimental data except that provided by the other work. It was possible to place a sub meeting in the room with both meetings having spatially correct visual access to the Telethrone participant. In practice it seems that the options for where to position chairs are limited and this informs the discussion about future research.

H3.1: Subjects will be allowed to walk around freely (in counterpoint to the static arrangement of H1), until they are comfortable that they are meeting the simulated eye gaze of a reconstructed model of a remote person (3D video). The accuracy is not significantly worse than previous research from the group when deployed in this more challenging setting. This corresponds to the ‘Mutual Gaze’ and ‘proxemics’ elements of the theoretical framework.

Hypothesis 3.1 is somewhat supported with the best case for the Telethrone projection very similar to the previous system as shown in 8.2. Even though the space was far more challenging, and the projectors significantly cheaper it was still possible for participants to judge eye gaze as well as in Roberts previous paper in the better of the two conditions. This suggested that the reconstruction system was viable for the Telethrone in the kind of spaces where it would potentially be deployed.

Table 8.2: Results against Roberts et al.

	TT best	TT worst	Old best	Old worst
Median Angle	-.77	8.27	.00	1.77
STDEV	6.27	4.48	8.61	5.36

H3.2: The Telethrone system is arranged to support up to five loosely arranged onlooking participants. The participants can identify when they are being referred to through non-verbal cues alone when the multi-view affordance of the Telethrone is used statistically more than when the multi-view affordance is not used. This corresponds to ‘informal’ from the centre of the theoretical framework.

Hypothesis 3.2 is well supported with all of the experimental participants correctly identifying themselves in the role associated with their station when asked directly. The control condition exhibited the well known problems associated with the Mona Lisa effect, with the control group making significant errors in judgement about their role in the meeting.

8.3.5 Evaluation

This section highlights results which are most pertinent to the development arc against the questions which investigated them and distillation of the results of the experimental hypotheses above, along with any additional lessons learned and some suppositions and expansions. This is a broader view of the original questions, the generated answers, and thoughts and discussion as a result. While there is some repetition here this section represents the most valuable (if selective) summary of the work.

Evaluating Chromatte

Q0: Do the experimental set-up pre-requisites exist? i.e. Does light reflected from Chromatte cloth fall off in intensity as the angle subtended from the axis of projection increases. Is this in-line with the manufacturers datasheet when the material is used outside of their guidelines as in the proposed Telethrone. Are there additional subjective concerns?

We evaluated the performance characteristics of the Chromatte cloth in the context of the Telethrone proposal. Optically (from an objective point of view) the manufacturers' data sheets were correct, but from a subjective point of view of an observer could still see the 5% of the other image(s) overlaid on the retro-reflection of the intended image. This in practice meant looking at an image of a face which exhibited some cross bleed from another projected image of a face and it was clear that there might be problems. There seemed no objective way to measure this conflating of images using objective sampling. Experimental design was adjusted to attempt to address this open question using questions about uncanniness.
In the final experiment it became clear that the material has a shelf life in even occasional use. The cloth which had been employed over the 8 years of part time research was suspected to be exhibiting much worse cross fade than in early experiments. It might have been useful to attempt to quantify this drop-off in the retro-reflective capability of the material but there was little time available,

and given that 10 year old material from the stores performed well it would seem that the issue is mechanical, and thereby pegged to how much the cloth is used, or moved. With that said, no new fabric was available to test either 10 year old material against so this remains an open question.
It is therefore suggested that the fabric be replaced as necessary or that the more resilient floor foam version of the product be used (it can be rolled up for storage).

Experiment for Q1

Q1: Does the Telethrone operate as an effective situated display (judged by questionnaire), which can show multiple views of a remote participant without excluding them from the attention (judged empirically) of the participant during a three-way conversation? Structured poker play designed to minimise but not preclude eye contact should be supported between three parties with the Telethrone. This will be examined statistically by the number of times the participant looks at the co-located player against the remote player.

It seems based on questionnaire responses that the Telethrone is somewhat situated. The participants in experiment one thought that the remote confederate in the multi-view projection condition was somewhat viewed through a screen, scoring an average of 3. This compared to 1 (not at all) for the co-located confederate on a 1-7 Likert scale. This single data point is difficult to judge in isolation. If it were a screen instead of a chair then one might expect the participants to rate this as 7, but without having a screen as a control (which is another shortcoming of the research) we cannot know for sure. Either way is seems 'better' that they thought it closer to real than they did a screen. They similarly thought that this remote confederate somewhat occupied a different space to them, again scoring 3 out of 7, compared to 1 (not at all) for the real. These are statistically significantly different scores, but are still 'good' scores considering the amount of technology mediating the conversation. It is also statistically significant that they thought the remote particiant was more in another room than them, scoring 2.5 compared to 1 (not at all). Given that they were indeed in a different room, this again seems a

decent score for the system. Conversely the participants thought very much (6 out of 7) that the remote participant was in the space with them.
As seen in the detailed results from the Networked Minds questions there were strong social indicators that the telepresent person seemed to be situated in the chair in the same way as the real confederate, and in watching the video it seemed that there was no breakdown in normal conversation caused by the technology.
Although this slightly mixed evidence that Telethrone acts as an effective situated display it is helpful to remind ourselves of Nowak's assertion that "A satisfactory level of co-presence with another mind can be achieved with conscious awareness that the interaction is mediated" [227].

The eye gaze data indicates that the Telethrone does not exclude the remote user from triadic poker. There is a lot more looking at the Telethrone in fact. It has been discussed that the flow of the poker play (for which there seems no supporting evidence), and the novelty of the screen may have led to this extra attention. The experimenter however feels instinctively that the angular offset in the gaze is the real culprit here.

There was no difference in the first experiment between the gaze patterns of the participants when comparing multi view against a control. Hypothesis 1.2 is null with the caveat that the experiment failed to sufficiently explore this. This was a frustrating finding at the time as the multiview affordance of the material felt critical to the exploration of this system, and later design of the experiments took a lot more care in this regard. There wasn't time however to solve both the technical problem of 3D video avatars on the Telethrone and the rigour of the first gaze experiment. This is a shortfall of the research.

The uncanniness questionnaire responses indicate that the multi view is successful insomuch as there is no indication that the other image bleeding through causes a problem when using two spatial segments 120°apart.

Q2: In the course of investigating the social presence capabilities of the system in the first experiment a new line of inquiry was explored. What are the theoretical limits of the system, especially when moving chairs around dynamically in the meeting?

The exercise exploring the potential for arrangement of the seats in a meeting was somewhat tested with a tracked handheld camera. The analysis somewhat supports the capability for meetings within meetings but is limited.

It seems that the system has demonstrated the necessary component parts to support a dynamic meeting where people come and go, move around in the space, potentially changing their direction of attention toward another smaller grouping.

Q3: Addressing shortcomings in the capability of the system to connect mutual eye gaze required 3D capture of the remote participant. To what extent does the new system, which allows the observer to move around the space, allow detection of eye gaze? Can it enable a group of five people to reliably detect non-verbal cues transmitted by the remote participant, in a simulated ideal sized meeting?

3D video can be applied to the chair with a novel pivot around the head. This lock of the head position was a major feature of the development but was implemented through requests to the programmers who built the CROSSDRIVE system. Locking the head position mitigated problems with co-aligning the spatial frames of the capture and playback systems and is touched upon in Appendix 1. It will be explored in detail in a future paper. Hypothesis 3.1 is supported.

In the more challenging scaled deployment there is noticeable bleed through at 35°separation (the maximum number of onlooking chairs that Telethrone supports), but participants reliably identify which spatial segment is being referenced by the reconstructed model. All of the 10 participants who saw the multi-view system correctly identified themselves as roles attributed in the monologue, while there was predictable confusion due to the Mona Lisa effect in the stations either side of

the central chair in the control condition.
There is no significant difference in the uncanny questions which were used to interrogate potential effects of the crossbleed.

We have demonstrated that it is possible to render the maximum 5 viewpoints while reconstructing a live view from the Octave system. This was performed on commodity graphics hardware (2 x Nvidia 1080). Frame rate was 4FPS and latency was around 1 second.

8.3.6 Contributions

As an overarching technical contribution this research presents and refines a novel system, loosely draping retro-reflective cloth over a chair, supporting affordances which are agreed to be important for telepresence systems. It compares well to other systems as detailed in table 8.1.
The ability to rapidly deploy a cost effective solution for normal spaces was specifically identified by BBC R&D, and while the final implementation is not deployable against their requirement, there are many considered and valuable iterative steps toward this. The contributions break down as follows, repeating the assertions made in the section 1.3 and inline with the structure presented throughout the thesis.

Contributions from Literature Review

The literature review details which affordances are agreed to be important for teleconferencing systems that aim to replicate the richness of natural meetings. It finds that some systems have demonstrated success, including objective measures such as task performance and trust. There seems agreement from Benford et al. in 1998 [31], through to Regenbrecht in 2015[261] that to date, nobody has convincingly provided an emulation of the face-to-face meeting that could be considered useful

in supporting real meetings. The technology either gets in the way, or else meets some requirements, or at worst hinders flow. Commercially available systems (in which there is considerable interest and investment) are agreed to be limited. There seems very little viable support for ad-hoc or informal systems, which are always-on, and can be trusted to link spaces with minimal disruption. These small informal group meetings are supposedly supported by available telepresence technologies, but problematic webcam solutions are the mainstay.
The literature review suggests that there is sufficient novelty, and need, to justify research into the Telethrone system. This assessment is significantly bolstered by the recent interest of Microsoft labs in this space. Crucially and as its primary differentiating characteristic the Telethrone offers multi-view support to situated furniture, without onerous technical overheads.
This assertion of a gap in provisioning across both commercial and research systems is clearly demonstrated in the review, supports the original challenge as presented by the BBC, and seems to be novel in its clarity.

Original Contributions Seeded from the Supporting Research

Application of the research methodology to the ideas proposed by Roberts, Sermon, and Thomas has generated several novel additions to science.

C0: Performed testing of physical characteristics of the retro-reflective cloth suggested by Roberts and Graham Thomas (BBC), to ensure that it was suitable for the intended purpose in principle. This performance data for the cloth formed the basis of a later discovery that the optical characteristics of this specific brand of retro-reflective material degrade over time. This is in itself a contribution. The early testing and prototyping considered and disregarded some image processing approaches which might have allowed better fitting of the projected image of the remote collaborator onto the chair. These approaches are useful markers for when the performance of the components improves sufficiently to match human perceptual considerations.

The early testing demonstrates objectively that five spatial segments (of a possible available 12) can give support for five independent viewpoints onto the Telethrone, independently of their distance from the Telethrone surface. The cross bleed effect between the different projected/reflected images was somewhat unexpected, with the human eye resolving the other faces in the early tests far more than cameras or sensors would suggest. These are immutable characteristic of the tested material. In addition, the optical performance of the Chromatte material unexpectedly degraded in the performance over time, even with very careful handling of the cloth.

C1: Performed rigorous set-up of a system capable of transmitting bi-directional video streams between two physically separated locations. This supported a three-way conversation with the display. An experiment tried to get a feel for how situated the display was, and a rigorous analysis of attention finds that a remote collaborator is not excluded from structured play by some characteristic of the technology. This contribution builds on others research into both situated displays and multiple spatially segmented views onto a remote collaborator.

The first experiment introduced the Telethrone display for the first time in publication, a technical novelty. The presented behavioural study of gaze suggested Telethrone to be a situated display [228]. It clearly supports transmission of upper body cues, which is an important channel of communication [10, 169]. Visible body torque allows transmission of attention [279]. The poker task employed in the experiment is a valuable methodological addition to such tests and could be refined and repeated in broader experimentation.

C3: Integration with spatial tracking, and playback of a recorded 3D video session in Octave enabled flexible proxemics with a view to potentially supporting dynamic meetings. An experiment tests the extent to which participants could resolve the eye gaze of the polygonal hull reconstruction (previously developed by others) in this new more demanding context. A second experiment explores the degree to which a maximal system supports non-verbal cues in a group context. This builds directly on work done in the wider research group into ability to resolve eye gaze at social distance and makes it applicable to the Telethrone. The experimental set-up proves for the first time that it is possible to move around the Telethrone with multiple spatially accurate viewpoints onto the remote collaborator, as envisaged by Roberts, Thomas, and Sermon.

The second experiment adds full body rendering from reconstruction, demonstrating the potential for reconnection of eye gaze [229], while the observer is free to move around in the space. This is the first implementation of the idea first suggested by Roberts, and suggested a futher individual contribution for continuous movement of participants in the meeting.
The third experiment integrates many adaptations to previous research software (CROSSDRIVE), to better support the Telethrone (Appendix 1). It also adds the double width Telethrone better supporting expansive spatial gestures, a 'basic and ubiquitous channel' [152]. In experiment three, the multi-view, tracked, Telethrone demonstrates the ability to address 5 onlookers with spatial discrimination.
Playing back a recorded session to a small group in order to establish (with high repeatability) the degree to which participants identify with the visual segment they occupy is a novel methodological contribution.

Individual Contribution to Knowledge

C2: Discussion of how the elements of the system support novel interaction within small groups, which are partially mediated by a tele-presence boundary. Analysis of how the system could and should be scaled explores the likely limits. Some theoretical consideration is applied to the current system implemented in the research, detailing the constraints and challenges.

The concept of supporting moving chairs around; within the flow of a meeting, to reorganise the nature and scope of the meeting, while retaining the spatial affordance of the Telethrone system, is an individual contribution, and is therefore detailed in its own chapter. The research demonstrates Telethrone with optical positional tracking, and shows that the viewpoint updates correctly in an end-to-end system. Multiple 'viewpoint correct', spatially segmented, live reconstructions of a single network stream are presented. The spatially correct viewpoints 'could' be moved on the fly thereby supporting re-organisation of the meeting.

8.3.7 Shortcomings

Although some shortcomings have been identified throughout, it is useful to list them in a chronological and structured fashion here.

Literature Survey Methodology

The writing style of the thesis is (to an extent) a function of the methodology decisions made early on in the research. The use of a mind map to author much of the work meant that paragraphs were written to be self supporting, then dragged around as the text evolved. This was a conscious decision but it is at times obvious that the flow of the text could be better, with some isolated assertions and inconsistent positioning of referencing. A loss of this mind map in final months of the PhD means that sometimes statements which deserve a reference at the end of the document were very hard to locate. In the end those that do have links toward the end were the result of lengthy searching through some 1200 archived papers using Adobe Acrobat file search. Re-assertiing links to previously referenced works seemed like a poor use of time toward the end of the research; at times there are obvious ommisssions.

Recruitment

Hedge at al [136] established that attention and eye gaze patterns are different between people who know one another well, as compared to people who have just met. Because recruitment was so difficult some of the participants in experiments 1 and 3 were known to the researcher. No analysis was done against this potentially confounding factor.

Misuse of Questionnaires

In the first experiment a mistake was made in the use of the Mori uncanny questions. They are only applied to the Telethrone image, and the scale was Likert 1-7, not the 1-5 reported in the literature. This meant there was no baseline comparator for the data which was collected.

Null Conclusion from Experiment 1

From a purely scientific perspective there is nothing wrong with the null result when comparing the multi-view and single view conditions, the result stands. The literature however strongly asserts that reconnection of eye gaze is important for systems such as Telethrone, and not finding this effect through the various measures is far more likely a shortcoming of the experimental design than a discovery that mutual eye gaze is somehow unimportant for triadic gameplay such as this. It remains an open question from an experimental point of view, but pinning this down would have consumed disproportionate resources, just to reinforce an accepted piece of knowledge from other studies. It would have been preferable to find a reassuring difference in the conditions, and it would have given confidence in the Telethrone, but instead it led to a redesign of the whole system. This is perfectly acceptable from a scientific point of view, but the poorly designed experiment remains a shortcoming.

Asymmetry of the System

Currently the system is one way because of limitations imposed by the available capture system. It is beyond the scope of this research to address that problem.
In both of the experiments where end to end transmission was tested the Telepresent confederate had to use a channel which was different to the Telethrone system. In the first experiment this was a TV, in the third there was no option of a return channel. This is a without doubt the major compromise in the research. Roberts

posited a system with a Telethrone deployment at either end, and it would have been valuable to attempt this. One of the persistent challenges with the research was availability of rooms for the set-up of the system. Working within this constraint mean designing the experiments differently.

Different Projectors in the Scaled System

It was necessary to use two types of projector in the third experiment. One of the projectors presented 4 times the pixels to the participants in that chair. The experiment was carefully designed around upper body non-verbal cues rather than relying on eye gaze which would be more vulnerable to this disparity.

Division of the End to End Processing Pipeline

In testing the end to end pipeline it was necessary to make a compromise in the distribution of the processing pipeline. The optimal system captures on 5 PC's, then segments locally to those PC's such that a texture and a mask are passed onward to be reconstructed. This onward transport of the images is optimised through jpeg compression but is still very network intensive. The Octave capture system is nominally configured so that the next stage, polygonal hull reconstruction, takes place on the local 10G network on a sufficiently capable computer. Several options were attempted, including the Octave image generator but it became clear that the best results were obtained by passing the 20 image streams of textures and masks to the Telethrone PC, which was the most modern in the pipeline.

No Audio Channel on the Live End to End

The end to end test of the visual channel was a 'nice to have' test of the reengineering of the established reconstruction system, and no experiment was designed to test this functionality because of time constraints. It was unclear at the time what a demonstration of transmission of audio would add to the system and so this was not undertaken. This was nontheless a desirable thing to demonstrate and it is a shame that it is not part of the final chapter.

Insufficient Investigation of Support for Subgrouping

It would have been valuable to explore this aspect further though more physical prototyping, but it was impractical to move the projector and tripod combinations with participants in play. This is a shortcoming of the implementation as the large and aged projectors could only be affixed to the heavy tripods via custom metal plates. This meant that the only way to incline the projectors such that they correctly addressed the Telethrone was to extend the tripod legs across the whole of the available floor space. This arrangement, and the placement of the tracking tripods, limited the physical placement of the equipment. It was unsafe to attempt allowing participants to relocate their positions in the room. For this reason, this is principally a theoretical contribution.
A compromise which allowed some bounds to be placed on where the equipment could possibly be placed was to track a video camera around the room. This allowed the limits imposed by the pivot which locks the head in place to be ascertained. This is a novel methodological contribution.

8.3.8 Future Research

Deployment - Proximity, Privacy, & Legitimacy

Checking against affordances from Fayard and Weeks [95] as discussed in the literature review: These are useful not to critique the research but to inform how the research would be best deployed.
Fayard et al. are concerned with proximity as a prerequisite for informal conversation. The Telethrone provides an almost arbitrary level of proximity - subject to

the projector used - in that an onlooking chair need not be confined to a position on the other side of a table but can rather be pulled right up to the Telethrone. The limits of this were tested as part of the scaled system experiment and found that the Infocus projector allows the whites of the eyes to be visible between 1m and 6m. The degree to which Telethrone can support privacy is not a technical issue but a deployment one. The privacy described by Fayard et al. is very much linked to their affordance of proximity, that is, a prerequisite which unlocks a certain kind of interpersonal behaviour is satisfied, so enabling informal conversation. In a similar vein to proximity there is ample flexibility in Telethrone for users to define their own natural privacy through movement of the seats closer to the Telethrone surface. Beyond this instinctive affordance the ability to control the volume of the remote participant or add bluetooth headphones could certainly support additional privacy. The social space in which a Telethrone system notionally sits could certainly support legitimacy, especially if communal business assets such as water coolers, kitchen elements, or photocopiers occupied the same general space.

Deployment - Secondary Affordances

Functional centrality implies that the system should be deployed in an appropriate social space. As it stands the available tripods and projectors are not suitable for such a deployment, but this is a matter of investment not technical impossibility. So called ‘reciprocal visibility’ is difficult for the envisaged Telethrone system. The BBC R&D pilot system for joining their spaces used an always-on screen in each space, but this system failed to join the spaces sufficiently to leverage opportunities for small group interaction. Telethrone seems to support small group interaction well, but in a hypothetical fully scaled system where Telethrones were also seats which could be occupied by real people (as described in Chapter 6, it wouldn’t be possible to transmit awareness of the remote space to trigger ad-hoc conversations. This could be ameliorated by a hybrid system with a large wall screen like that tested by the BBC to allow some awareness of the remote space, or more interestingly perhaps projection of segments of the remote space onto the Telethrone surfaces in an ’always-on’ capacity.

The spaces between the chairs mean that there is in principle ease of access, but again it could not be asserted that the current system allows access and egress. A channel was established around the tripods to allow participants to safely enter the space during the third experiment. Again, this could be solved with more modern and capable equipment.
The literature survey briefly touched on ‘smart spaces’ and the attendant supporting technologies for collaborative working across a distance. The research began to explore shared digital resources with the poker table in experiment one, and this seems to be a good fit for the space. The answer as to whether the system can support multiple shared resources is a tentative yes, but as stated in the literature survey this is outside the scope of this research.

Integrated Systems

The next opportunity for Telethrone is to move it toward a deployable system. The affordances which are not currently met for servicing small meetings demand that the projectors be small, bright, and integrated into rolling chairs, preferably with their own power and wireless networking.

Different Tracking

The tracking should be either attached to the walls in a semi-permanent installation in an appropriate space, or perhaps be fiducial based, from a single camera on the Telethrone. Such an image based system using ARtoolkit [253] could use markers above the projectors to locate the chair stations, or the newer object based room tracking [322] using SLAM (simultaneous localisation and mapping) to identify and locate the visual features on the stations using computer vision algorithms.
The most interesting potential tracking system could use a single lens on the Telethrone to automatically track and reconstruct viewpoints for people in the

scene [91]. This is a PhD in itself as zones of engagement could be generated into the depth of the scene (the workplace), with the system choosing to detect and generate viewpoints depending on the proximity of the passing users to the chairs. This would allow better connection of the non-verbal cues of the Telethrone user to space outside of the meeting. Careful research would be required since the users would pass across and between the spatial segments generated by the Telethrone system. If the software were sufficiently aware of which stations were occupied by onlookers at any given time then multiple segments could be assigned to correctly generate the same viewpoint for a passing person in order to attract them to the empty chairs.
With these upgrades it would be possible to engage with more research questions which add value to everyday use of the system, perhaps deploying in a real social space.

Structured Telethrone Surface

Testing projection at the oblique limits of the capability provided additional insights and suggest areas which could be improved. The image stretches out a little at the more oblique angles, covering more pixels, reducing the quality of the image. The projection of the legs becomes far more distorted for instance.

Different 3D capture

The Octave is not a suitable 3D capture system for the Telethrone. It was utilised because it was available. A quick look at time of flight and photogrammetry systems like Kinect ruled them out at the time the research was started. By the time that capture systems had matured (as with Room2Room), Telethrone capture through photogrammetry was already well under way, and it was infeasible to change track. This is a shortcoming of long, part time, technical PhD's. This would be the first problem to engage with in ongoing research.

Group Interaction with a Symmetric System

The best opportunity to build on the current system is to develop the proposed symmetric system discussed in the previous chapter as a ‘pragmatic maximum’. This would require different 3D capture, a different less encumbering tracking solutions, and probably a full set of small matching projectors for each side of the system. None of these things are technically impossible, or even difficult, they simply require resourcing.

A credible use case example

During the course of the research there were some initial discussions with solicitors and a barrister about a potential use in a legal context. It is considered vital to examination by interview, cross examination and the legal process that the full gamut of human communication channels is available. This imperative to see and to judge the manner and facial affect of a witness can be in tension with the need to protect witnesses from the physicality of the courtroom. The most visceral example of this is child abuse cases where video links are already used. There is a clear opportunity for the Telethrone system in this context.

8.4 Conclusion

The research set out to establish if it was possible to support group interaction through a novel application of a retro-reflective cloth, draped over a chair, in a less formal social context. In the course of the research an additional novelty was discovered, in that the system seems to support characteristics which may allow reorganising and subdividing the meetings, an area of telepresence which seems to have little consideration outside of avatar representations in outmoded systems like SecondLife.
The literature review identifies pertinent surrounding knowledge, and latterly adds

an examination of meetings, and types of meetings which seems to be novel in this technology context.
The survey identifies a gap in current knowledge for persistent teleconferencing systems, which are able to support transmission of non-verbal cues, for small, and dynamic informal groups.

An ostensibly convenient research problem, a likely solution, and supporting technologies with which to investigate, were all in place from the start. It is appropriate and important that this initial condition and context was so carefully outlined in the declaration, and explains why the sources for various ideas are so explicitly reiterated through the thesis.

The methodology is nominal for the field of telepresence / HCI, but there are new methods presented within the experiments. Use of poker to control interpersonal eye gaze seems to be novel, and there are certainly suggestions in this approach which could inform other researchers. The use of a scripted spatially keyed address, to a mock business meeting, was highly effective and seems novel. The new implementation of the software is itself part of the methodology and it seems that the head being locked as a pivot on the display, to limit motion artefacts is novel, though this was a collaborative work.
The findings from the first experiment are broadly positive, with many good indications (or rather a lack of counter indications) that the Telethrone acts to involve the remote participant as a natural participant in the game.
The second experiment demonstrated the potential for integration of 3D reconstruction with the Telethrone system, and did so in a challenging daylight context. Comparison with other systems validated the approach.
The goal at that stage remained a teleconferencing system which could join a remote user to a small group: A bi-directional system. The novel individual contribution outlined added the challenge of dynamic meetings.
In the final analysis the thesis falls short of the aim of a demonstrable telepresence system. Subsets of the problem have been rigorously tested, obtaining results

which support the premise, but it proved to be too ambitious an undertaking to examine social interaction using a two way system which supported groups.

The blending of the Telethrone with previous research in the group is a valuable addition to the work which has been done over the past decade, with a new and novel code fork which is stable and reusable. This is the first implementation of Roberts' vision for tracked informal telepresence. The collaborative effort which brought this broader system to fruition is under-reported in this thesis, but will be reported in a future publication.

The ability to move chairs around in a meeting and still have access to the affordances of the Telethrone is an important individual contribution, and is explored within the limits imposed by the equipment, and certainly warrants more investigation in future.

The results from experiment three unambiguously indicates that the concept can provide spatially correct eye gaze, and gesture to a group of 5 people. This is a solid conclusion to the experimental arc, and validates the novel deployment of Chromatte loosely over a chair surface in an informal setting.

Bibliography

[1] Ericsson mobility report, 2015. URL http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf.

[2] Sigurdur O Adalgeirsson and Cynthia Breazeal. Mebot: a robotic platform for socially embodied presence. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 15–22. IEEE Press, 2010. doi: 10.1145/1734454.1734467.

[3] Piotr D Adamczyk and Michael B Twidale. Supporting multidisciplinary collaboration: Requirements from novel hci education. In *Proceedings Of The Sigchi Conference On Human Factors In Computing Systems*, pages 1073–1076. ACM, 2007. doi: 10.1145/1240624.1240787.

[4] Samer Al Moubayed, Jens Edlund, and Jonas Beskow. Taming mona lisa: Communicating gaze faithfully in 2d and 3d facial projections. volume 1, page 25, 2012. doi: 10.1145/2070719.2070724.

[5] Asma Al-Saidi, Nick J Avis, Ian J Grimstead, and Omer F Rana. Distributed collaborative visualization using light field rendering. In *Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on*, pages 609–614. IEEE, 2009. doi: 10.1109/ccgrid.2009.79.

[6] Jeremie Allard, Jean-Sebastien Franco, Clement Menier, Edmond Boyer, and Bruno Raffin. The grimage platform: A mixed reality environment for interactions. In *IEEE International Conference on Computer Vision Systems, 2006 ICVS'06.*, pages 46–46. IEEE, 2006. doi: 10.1109/icvs.2006.59.

[7] Stuart M Anstis, John W Mayhew, and Tania Morley. The perception of where a face or television portrait is looking. volume 82, pages 474–489. JSTOR, 1969. doi: 10.2307/1420441.

[8] Paul M Aoki, Matthew Romaine, Margaret H Szymanski, James D Thornton, Daniel Wilson, and Allison Woodruff. The mad hatter's cocktail party: A social mobile audio space supporting multiple simultaneous conversations. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 425–432. ACM, 2003. doi: 10.1145/642611.642686.

[9] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, 1976. ISBN 9780521208659. doi: 10.1017/s0007125000073980. URL https://books.google.co.uk/books?id=ZBR0QgAACAAJ.

[10] Michael Argyle. *Bodily communication*. Methuen, 1988.

[11] Michael Argyle and Janet Dean. Eye-contact, distance and affiliation. pages 289–304. JSTOR, 1965. doi: 10.2307/2786027.

[12] Michael Argyle and Jean Ann Graham. The central europe experiment: Looking at persons and looking at objects. volume 1, pages 6–16, 1976. doi: 10.1007/BF01115461.

[13] Michael Argyle and Roger Ingham. *Gaze, Mutual Gaze, and Proximity*. Number 1965. 1969. doi: 10.1515/semi.1972.6.1.32.

[14] Richard D Arvey. Why face-to-face business meetings matter. *White Paper for the Hilton Group*, 2009.

[15] Nicole Atzpadin, Peter Kauff, and Oliver Schreer. Stereo analysis by hybrid recursive matching for real-time immersive video conferencing. volume 14, pages 321–334. IEEE, 2004. doi: 10.1109/tcsvt.2004.823391.

[16] Daniel Avrahami, James Fogarty, and Scott E Hudson. Biases in human estimation of interruptibility: Effects and implications for practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 50–60. ACM, 2007. doi: 10.1145/1240624.1240632.

[17] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. volume 10, pages 583–598. MIT Press, 2001. doi: 10.1162/105474601753272844.

[18] Jeremy N. Bailenson, Andrew C. Beall, and Jim Blascovich. Gaze and task performance in shared virtual environments. volume 13, pages 313–320, #dec# 2002. doi: 10.1002/vis.297.

[19] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Interpersonal distance in immersive virtual environments. volume 29, pages 819–833. Sage Publications, 2003. doi: 10.1177/0146167203029007002.

[20] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, 2006. doi: 10.1162/pres.15.4.359.

[21] Tibor Balogh and Pter Tams Kovcs. Real-time 3d light field transmission, 2010. URL http://dx.doi.org/10.1117/12.854571.

[22] D. Bandyopadhyay, R. Raskar, and H. Fuchs. Dynamic shader lamps : Painting on movable objects. pages 207–216. IEEE Comput. Soc. ISBN 0-7695-1375-1. doi: 10.1109/ISAR.2001.970539.

[23] Jakob Bardram, Sofiane Gueddana, Steven Houben, and Søren Nielsen. Reticularspaces: Activity-based computing support for physically distributed and collaborative smart spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2845–2854. ACM, 2012. doi: 10.1145/2207676.2208689.

[24] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Is the uncanny valley an uncanny cliff? In *Robot and Human interactive Communication, 2007.*, pages 368–373. IEEE, 2007. doi: 10.1109/roman.2007.4415111.

[25] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. My robotic doppelgänger-a critical look at the uncanny valley. In *Robot and Human Interactive Communication, 2009.*, pages 269–276. IEEE, 2009.

[26] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[27] Bruce G Baumgart. A polyhedron representation for computer vision. In *Proceedings of the May 19-22, 1975, national computer conference and exposition*, pages 589–596. ACM, 1975. doi: 10.1145/1499949.1500071.

[28] Dennis Beck, Paul Fishwick, Rasha Kamhawi, Amy Jo Coffey, and Julie Henderson. Synthesizing presence: A multidisciplinary review of the literature. volume 3, 2011.

[29] Stephan Beck, Andre Kunert, Alexander Kulik, and Bernd Froehlich. Immersive group-to-group telepresence. volume 19, pages 616–625. IEEE, 2013. doi: 10.1109/tvcg.2013.33.

[30] Steve Benford, John Bowers, Lennart E Fahlén, Chris Greenhalgh, and Dave Snowdon. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, Denver, #may# 1995. doi: 10.1145/223904.223935.

[31] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)*, 5 (3):185–223, 1998. doi: 10.1145/292834.292836.

[32] LOK Benjamin. Shader lamps virtual patients: the physical manifestation of virtual patients. *Medicine Meets Virtual Reality 19: NextMed*, 173:372, 2012.

[33] Gary Bente, WC Donaghy, and Dorit Suwelack. Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. volume 22, 1998.

[34] Charles R Berger and Richard J Calabrese. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human communication research*, 1(2):99–112, 1975. doi: 10.1111/j.1468-2958.1975.tb00258.x.

[35] Katerina Bezrukova, Karen A Jehn, Elaine L Zanutto, and Sherry MB Thatcher. Do workgroup faultlines help or hurt? a moderated model of faultlines, team identification, and group performance. *Organization Science*, 20(1):35–50, 2009. doi: 10.1287/orsc.1080.0379.

[36] Frank Biocca. The cyborg dilemma : Embodiment in virtual environments. pages 12–26, 1997. ISBN 0818680849.

[37] Frank Biocca, Chad Harms, and Judee Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. volume 12, pages 456–480. MIT press, 2003. doi: 10.1162/105474603322761270.

[38] Frank Biocca, Chad Harms, and Judee K Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence*, 12(5):456–480, 2003. doi: 10.1162/105474603322761270.

[39] P-a Blanche, a Bablumian, R Voorakaranam, C Christenson, W Lin, T Gu, D Flores, P Wang, W-Y Hsieh, M Kathaperumal, B Rachwal, O Siddiqui, J Thomas, R a Norwood, M Yamamoto, and N Peyghambarian. Holographic three-dimensional telepresence using large-area photorefractive polymer. volume 468, pages 80–3, #nov# 2010. doi: 10.1038/nature09521.

[40] Ann Blandford, Anna Cox, and Paul Cairns. Controlled experiments. 2008.

[41] Jim Blascovich. A theoretical model of social influence for increasing the utility of collaborative virtual environments. *Proceedings of the 4^{th}*

international conference on Collaborative virtual environments - CVE '02, pages 25–30, 2002. doi: 10.1145/571878.571883.

[42] Simon W Bock, Peter Dicke, and Peter Thier. How precise is gaze following in humans? *Vision research*, 48(7):946–957, 2008. doi: 10.1016/j.visres.2008.01.011.

[43] Martin Böcker, Werner Blohm, and Lothar Mühlbach. Anthropometric data on horizontal head movements in videocommunications. In *Conference Companion on Human Factors in Computing Systems*, pages 95–96. ACM, 1996. doi: 10.1145/257089.257175.

[44] Leanne S Bohannon, Andrew M Herbert, Jeff B Pelz, and Esa M Rantanen. Eye contact and video-mediated communication: A review. *Displays*, 34(2): 177–185, 2013. doi: 10.1016/j.displa.2012.10.009.

[45] Steven M Boker, Jeffrey F Cohn, Barry-John Theobald, Iain Matthews, Michael Mangini, Jeffrey R Spies, Zara Ambadar, and Timothy R Brick. Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3):874, 2011. doi: 10.1037/a0021928.

[46] Yevgenia Bondareva and Don Bouwhuis. Determinants of social presence in videoconferencing. In *AVI2004 Workshop on Environments for Personalized Information Access*, pages 1–9. Citeseer, 2004.

[47] David N Boote and Penny Beile. Scholars before researchers: On the centrality of the dissertation literature review in research preparation. volume 34, pages 3–15. Sage Publications, 2005. doi: 10.3102/0013189x034006003.

[48] Erin Bradner and Gloria Mark. Why distance matters: effects on cooperation, persuasion and deception. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, CSCW '02, pages 226–235, New York, NY, USA, 2002. ACM. ISBN 1-58113-560-2. doi: 10.1145/587078.587110.

[49] Nicola Bruno and Michela Muzzolini. Proxemics revisited: Similar effects of arms length on men's and women's personal distances. *Universal Journal of Psychology*, 1(2):46–52, 2013.

[50] Saniye Tugba Bulu. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58(1):154–161, 2012. doi: 10.1016/j.compedu.2011.08.024.

[51] William Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, volume 92, pages 123–129, 1992.

[52] William Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, volume 92, pages 123–129. Canadian Information Processing Society Toronto, Canada, 1992.

[53] Barrett S Caldwell, Shiaw-Tsyr Uang, and Lilas H Taha. Appropriateness of communications media use in organizations: situation requirements and media characteristics. *Behaviour & Information Technology*, 14(4):199–207, 1995. doi: 10.1080/01449299508914633.

[54] Apoorva Sachdev John Canny. Triple-view: Improving persuasion in group video conferencing through spatial faithfulness. 2011.

[55] D.E. Carlston. *The Oxford Handbook of Social Cognition*. Oxford Library of Psychology. OUP USA, 2013. ISBN 9780199730018. URL https://books.google.co.uk/books?id=nagzAAAAQBAJ.

[56] Tanya L Chartrand and John A Bargh. The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999. doi: 10.1037//0022-3514.76.6.893.

[57] Milton Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '02, pages 49–56, New York, NY, USA, 2002. ACM, ACM. ISBN 1-58113-453-3. doi: 10.1145/503376.503386.

[58] Nicole Chovil. Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction*, 25(1-4):163–194, 1991. doi: 10.1080/08351819109389361.

[59] Jan Ciger, Bruno Herbelin, and Daniel Thalmann. Evaluation of gaze tracking technology for social interaction in virtual environments. In *Proc. of the 2nd Workshop on Modeling and Motion Capture Techniques for Virtual Environments (CAPTECH'04)*. Citeseer, 2004.

[60] Alex Colburn, Michael F Cohen, and Steven Drucker. The role of eye gaze in avatar mediated conversational interfaces. 2000.

[61] Mark Cook. Gaze and mutual gaze in social encounters. *American Scientist*, 65(3):328–333, 1977. ISSN 00030996.

[62] Eddie Cooke, Oliver Schreer, Bernhard Pasewaldt, and Peter Kauff. Extension of incomplete 3d for arbitrary multi-view-synthesis. In *VMV*, pages 205–212, 2000.

[63] Eddie Cooke, Peter Kauff, and Oliver Schreer. Multiple narrow-baseline system for immersive teleconferencing. In *Video/Image Processing and Multimedia Communications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom*, pages 367–370. IEEE, 2002.

[64] Eddie Cooke, Peter Kauff, and Oliver Schreer. Image-based rendering for teleconference systems. 2002.

[65] Eddie Cooke, Peter Kauff, and Oliver Schreer. Image-based rendering for teleconference systems, 2002.

[66] Antonio Criminisi, Jamie Shotton, Andrew Blake, and Philip H S Torr. Gaze manipulation for one-to-one teleconferencing. In *Proceedings of the 9th IEEE International Conference on Conputer Vision (ICCV '03)*, pages 191–198, Nice, #oct# 2003. doi: 10.1109/iccv.2003.1238340.

[67] Carolina Cruz-Neira. *Virtual reality based on multiple projection screens: the cave and its applications to computational science and engineering*. PhD thesis, University of Illinois, Chicago, #sep# 1995.

[68] Elisabeth Cuddihy and Deborah Walters. Embodied interaction in social virtual environments. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 181–188. ACM, 2000. doi: 10.1145/351006.351039.

[69] Peter Dalsgaard and Kim Halskov. 3d projection on physical objects: design insights from five real life cases. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1041–1050. ACM, 2011. doi: 10.1145/1978942.1979097.

[70] Rudolph P Darken, Joseph A Sullivan, and Mark Lennerton. A chromakey augmented virtual environment for deployable training. 2003.

[71] Paul De Greef and Wijnand A Ijsselsteijn. Social presence in a home tele-application. *CyberPsychology & Behavior*, 4(2):307–315, 2001. doi: 10.1089/109493101300117974.

[72] Eva Deckers, Stephan Wensveen, Pierre Levy, and Rene Ahn. Designing for perceptual crossing: designing and comparing three behaviors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1901–1910. ACM, 2013. doi: 10.1145/2470654.2466251.

[73] Alan R Dennis, Annette C Easton, George K Easton, Joey F George, and JF Nunamaker. Ad hoc versus established groups in an electronic meeting system environment. In *System Sciences, 1990., Proceedings of the Twenty-Third Annual Hawaii International Conference on*, volume 3, pages 23–29. IEEE, 1990.

[74] Jon Martin Denstadli, Tom Erik Julsrud, and Randi Johanne Hjorthol. Videoconferencing as a mode of communication: A comparative study of the use of videoconferencing and face-to-face meetings. *Journal of Business and Technical Communication*, 26(1):65–91, 2012.

[75] Munjal Desai, Katherine M Tsui, Holly A Yanco, and Chris Uhlik. Essential features of telepresence robots. In *Technologies for Practical Robot Applications (TePRA), 2011 IEEE Conference on*, pages 15–20. IEEE, 2011. doi: 10.1109/tepra.2011.5753474.

[76] Xianghua Ding, Thomas Erickson, Wendy A Kellogg, Stephen Levy, James E Christensen, Jeremy Sussman, Tracee Vetting Wolf, and William E Bennett. An empirical study of the use of visually enhanced VoIP audio conferencing: the case of ieac. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1019–1028. ACM, 2007. doi: 10.1145/1240624.1240780.

[77] O Divorra, Jaume Civit, F Zuo, H Belt, I Feldmann, O Chreer, E Yellin, W Ijsselsteijn, R van Eijk, D Espinola, et al. Towards 3d-aware telepresence: Working on technologies behind the scene. 2010.

[78] Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escolano, Christoph Rhemann, David Kim, Jonathan Taylor, et al. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)*, 35 (4):114, 2016. doi: 10.1145/2897824.2925969.

[79] Paul Dourish, Annette Adler, Victoria Bellotti, and Austin Henderson. Your place or mine ? learning from long-term use of audio-video communication. *Victoria*, 1993. doi: 10.1007/bf00141935.

[80] T Duckworth and D.J Roberts. Accelerated polyhedral visual hulls using opencl. In *IEEE Virtual Reality Conference (VR '11)*, pages 203–204, Singapore, #mar# 2011. doi: 10.1109/vr.2011.5759469.

[81] Tobias Duckworth and David J Roberts. Camera image synchronisation in multiple camera real-time 3D reconstruction of moving humans. In *Distributed Simulation and Real Time Applications*, pages 138–144, Salford, #sep# 2011. IEEE Computer Society. doi: 10.1109/ds-rt.2011.15.

[82] Tobias Duckworth and David J Roberts. 3drecon, a utility for 3D reconstruction from video. *Joint Virtual Conference of ICAT - EGVE - EuroVR (JVRC '12)*, #oct# 2012.

[83] Tobias Duckworth and David J Roberts. Parallel processing for real-time 3d reconstruction from video streams. *Journal of real-time image processing*, 9 (3):427–445, 2014.

[84] Christine Duffy and Mary Beth McEuen. The future of meetings: the case for face-to-face. 2010.

[85] C Edigo. Videoconferencing as a technology to support group work: A review of its failure. In *Proceedings of the ACM conf. on Computer-Supported Cooperative Work*, 1988.

[86] Boris Eisenbart, Massimo Garbuio, Daniele Mascia, and Federica Morandi. Does scheduling matter? when unscheduled decision making results in more effective meetings. *Journal of Strategy and Management*, 9(1):15–38, 2016. doi: 10.1108/jsma-03-2014-0017.

[87] Paul Ekman. Facial expression and emotion. *American psychologist*, 48(4): 384, 1993. doi: 10.1037//0003-066x.48.4.384.

[88] Osman Eldes, Kaan Akşit, and Hakan Urey. Multi-view autostereoscopic projection display using rotating screen. *Optics express*, 21(23):29043–29054, 2013. doi: 10.1364/oe.21.029043.

[89] Thomas Erickson and WA Kellogg. Telepresence in virtual conferences: An empirical comparison of distance collaboration technologies. pages 1–6, 2010.

[90] Thomas Erickson, N Sadat Shami, Wendy A Kellogg, and David W Levine. Synchronous interaction among hundreds: an evaluation of a conference in an avatar-based virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 503–512. ACM, 2011. doi: 10.1145/1978942.1979013.

[91] A. Ess, B. Leibe, K. Schindler, and L. Van Gool. A mobile vision system for robust multi-person tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008. doi: 10.1109/CVPR.2008.4587581.

[92] Sascha Fagel, Gérard Bailly, Frédéric Elisei, and Amélie Lelong. On the importance of eye gaze in a face-to-face collaborative task. In *Proceed-*

ings of the 3rd international workshop on Affective interaction in natural environments, pages 81–86. ACM, 2010. doi: 10.1145/1877826.1877847.

[93] Allen J Fairchild, Simon P Campion, Arturo S García, Robin Wolff, Terrence Fernando, and David J Roberts. A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016. doi: 10.1109/tcsvt.2016.2580425.

[94] Allen J Fairchild, Simon P Campion, Arturo S García, Robin Wolff, Terrence Fernando, and David J Roberts. A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(4):814–827, 2017. doi: 10.1109/tcsvt.2016.2580425.

[95] Anne-Laure Fayard and John Weeks. Photocopiers and water-coolers: The affordances of informal interaction. *Organization studies*, 28(5):605–634, 2007. doi: 10.1177/0170840606068310.

[96] I Feldmann, N Atzpadin, O Schreer, J-C Pujol-Acolado, J-L Landabaso, and O Divorra Escoda. Multi-view depth estimation based on visual-hull enhanced hybrid recursive matching for 3D video conference systems. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 745–748. IEEE, 2009. doi: 10.1109/icip.2009.5414274.

[97] DI Fels and PL Weiss. Toward determining an attention-getting device for improving interaction during video-mediated communication. *Computers in human behavior*, 16(2):189–198, 2000. doi: 10.1016/s0747-5632(99)00055-2.

[98] Alois Ferscha and James Johnson. Distributed interaction in virtual spaces. In *Distributed Interactive Simulation and Real-Time Applications, 1999. Proceedings. 3rd IEEE International Workshop on*, pages 5–13. IEEE, 1999.

[99] Robert S Fish, Robert E Kraut, Robert W Root, and Ronald Rice. Video as a technology for informal communication. *ACM*, (36):48–61, 1993. doi: 10.1145/151233.151237.

[100] J Franco and Edmond Boyer. Efficient polyhedral modeling from silhouettes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(3): 414–427, 2009. doi: 10.1109/tpami.2008.104.

[101] Jean-Sébastien Franco and Edmond Boyer. Exact polyhedral visual hulls. In *British Machine Vision Conference (BMVC'03)*, volume 1, pages 329–338, 2003. doi: 10.5244/c.17.32.

[102] H Fuchs, N Kelshikar, J Mulligan, and K Daniilidis. 3d tele-collaboration over internet 2. In *International Workshop on Immersive Telepresence (ITP '02)*, Juan Les Pin, #dec# 2002.

[103] Janet Fulk, Charles W Steinfield, Joseph Schmitz, and J Gerard Power. A social information processing model of media use in organizations. *Communication research*, 14(5):529–552, 1987. doi: 10.1177/009365087014005005.

[104] Chris Fullwood and Gwyneth Doherty-Sneddon. Effect of gazing at the camera during a video link on recall. *Applied Ergonomics*, 37(2):167–175, 2006. doi: 10.1016/j.apergo.2005.05.003.

[105] Maia Garau, Mel Slater, Simon Bee, Martina Angela Sasse, Gower St, and Martlesham Heath. The impact of eye gaze on communication using humanoid avatars. In *SIGCHI*, pages 309–316, 2001. ISBN 1581133278. doi: 10.1145/365024.365121.

[106] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M Angela Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536. ACM, 2003. doi: 10.1145/642700.642703.

[107] D Randy Garrison, Terry Anderson, and Walter Archer. Critical inquiry in a text-based environment: Computer conferencing in higher education. volume 2, pages 87–105. Elsevier, 1999. doi: 10.1016/s1096-7516(00)00016-6.

[108] Paolo Simone Gasparello, Giuseppe Marino, Filippo Bannò, Franco Tecchia, and Massimo Bergamasco. Real-Time Network Streaming of Dynamic 3D Content with In-frame and Inter-frame Compression. In *IEEE/ACM 15th International Symposium on Distributed Simulation and Real Time Applications (DSRT '11)*, pages 81–87, Salford, #sep# 2011. IEEE. doi: 10.1109/ds-rt.2011.24.

[109] Daniel Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and vision computing*, 27(12):1775–1787, 2009. doi: 10.1016/j.imavis.2009.01.004.

[110] William W. Gaver. The affordances of media spaces for collaboration. *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92*, 92:17–24, 1992. doi: 10.1145/143457.371596.

[111] WW Gaver, Abigail Sellen, Christian Heath, and Paul Luff. One is not enough: Multiple views in a media space. 1993. doi: 10.1145/169059.169268.

[112] FC Gee, WN Browne, and K Kawamura. Uncanny valley revisited. In *Robot and Human Interactive Communication, 2005. IEEE International Workshop on*, pages 151–157. Ieee, 2005.

[113] Jim Gemmell, Kentaro Toyama, C Lawrence Zitnick, Thomas Kang, and Steven Seitz. Gaze awareness for video-conferencing: A software approach. *IEEE MultiMedia*, (4):26–35, 2000. doi: 10.1109/93.895152.

[114] Darren Gergle, Robert E Kraut, and Susan R Fussell. The impact of delayed visual feedback on collaborative performance. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1303–1312. ACM, 2006. doi: 10.1145/1124772.1124968.

[115] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[116] James J Gibson and Anne D Pick. Perception of another person's looking behavior. *The American journal of psychology*, 76(3):386–394, 1963. doi: 10.2307/1419779.

[117] Marco Gillies, Mel Slater, et al. Non-verbal communication for correlational characters. 2005.

[118] Gernot Goebbels, Vali Lalioti, and Martin Göbel. Design and evaluation of team work in distributed collaborative virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 231–238. ACM, 2003. doi: 10.1145/1008653.1008693.

[119] S Goldin-Meadow. The role of gesture in communication and thinking. volume 3, pages 419–429, #nov# 1999. doi: 10.1016/s1364-6613(99) 01397-2.

[120] Charles Goodwin. Action and embodiment within situated human interaction. *Journal of pragmatics*, 32(10):1489–1522, 2000. doi: 10.1016/ s0378-2166(99)00096-x.

[121] Daniel Gotsch, Xujing Zhang, Timothy Merritt, and Roel Vertegaal. Telehuman2: A cylindrical light field teleconferencing system for life-size 3d human telepresence. In *CHI*, page 522, 2018.

[122] Oliver Grau, Graham A Thomas, A Hilton, J Kilner, and J Starck. A robust free-viewpoint video system for sport scenes. In *3DTV Conference, 2007*, pages 1–4. IEEE, 2007. doi: 10.1109/3dtv.2007.4379384.

[123] Jerald Greenberg. The role of seating position in group interaction: A review, with applications for group trainers. *Group & Organization Studies*, 1(3): 310–327, 1976. doi: 10.1177/105960117600100306.

[124] Chris Greenhalgh, Adrian Bullock, Emmanuel Frécon, David Lloyd, and Anthony Steed. Making networked virtual environments work. *Presence*, 10(2):142–159, 2001. doi: 10.1162/105474601750216777.

[125] Chris Greenhalgh, Adrian Bullock, Emmanuel Frécon, David Lloyd, and Anthony Steed. Making networked virtual environments work. *Presence: Teleoperators and Virtual Environments*, 10(2):142–159, 2001. doi: 10.1162/105474601750216777.

[126] I.J. Grimstead, D.W. Walker, and N.J. Avis. Collaborative visualization: A review and taxonomy. *Ninth IEEE International Symposium on Distributed Simulation and Real-Time Applications*, pages 61–69, 2005. doi: 10.1109/DISTRA.2005.12.

[127] Markus Gross, Stephan Würmlin, Martin Naef, Edouard Lamboray, Christian Spagno, Andreas Kunz, Esther Koller-Meier, Tomas Svoboda, Luc Van Gool, Silke Lang, et al. blue-c: a spatially immersive display and 3d video portal for telepresence. 22(3):819–827, 2003. doi: 10.1145/1201775.882350.

[128] Charlotte N Gunawardena and Frank J Zittle. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education*, 11(3):8–26, 1997. doi: 10.1080/08923649709526970.

[129] Joseph C Hager and Paul Ekman. Long-distance of transmission of facial affect signals. *Ethology and Sociobiology*, 1(1):77–82, 1979. doi: 10.1016/0162-3095(79)90007-4.

[130] Edward Twitchell Hall and Edward T Hall. *The Hidden Dimension*, volume 1990. Anchor Books New York, 1969. doi: 10.2307/1572461.

[131] Jeffrey T Hancock, Jennifer Thom-Santelli, and Thompson Ritchie. Deception and design: The impact of communication technology on lying behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–134. ACM, 2004. doi: 10.1145/985692.985709.

[132] Chad Harms and Frank Biocca. Internal consistency and reliability of the networked minds measure of social presence. 2004.

[133] J Hauber, H Regenbrecht, M Billinghurst, and A Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. *Interfaces*, (November), 2006. doi: 10.1145/1180875.1180937.

[134] Jörg Hauber, Holger Regenbrecht, Aimee Hills, Andrew Cockburn, and Mark Billinghurst. Social presence in two-and three-dimensional videoconferencing. 2005.

[135] Caroline Haythornthwaite, Barry Wellman, and Marilyn Mantei. Work relationships and media use: A social network analysis. *Group Decision and Negotiation*, 4(3):193–211, 1995. doi: 10.1007/bf01384688.

[136] BJ Hedge, BS Everitt, and Christopher D Frith. The role of gaze in dialogue. volume 42, pages 453–475. Elsevier, 1978. doi: 10.1016/0001-6918(78)90033-1.

[137] Carrie Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992. doi: 10.1162/pres.1992.1.2.262.

[138] Ilona Heldal, Anthony Steed, Maria Spante, Ralph Schroeder, Sophia Bengtsson, and Marja Partanen. Successes and failures in co-present situations. *Presence: Teleoperators and Virtual Environments*, 14(5):563–579, 2005. doi: 10.1162/105474605774918679.

[139] Mattias Heldner and Jens Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, 2010. doi: 10.1016/j.wocn.2010.08.002.

[140] Chin-Chang Ho and Karl F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Comput. Hum. Behav.*, 26(6):1508–1518, #nov# 2010. ISSN 0747-5632. doi: 10.1016/j.chb.2010.05.015.

[141] Chin-Chang Ho, Karl F MacDorman, and ZA Dwi Pramono. Human emotion and the uncanny valley: a glm, mds, and isomap analysis of robot video ratings. In *Proceedings of the 3rd ACM/IEEE international conference*

on Human robot interaction, pages 169–176. ACM, 2008. doi: 10.1145/1349822.1349845.

[142] Stefan G. Hofmann, Michael Suvak, and Brett T. Litz. Sex differences in face recognition and influence of facial affect. *Personality and Individual Differences*, 40(8):1683–1690, #jun# 2006. ISSN 0191-8869. doi: 10.1016/j.paid.2005.12.014.

[143] Håkan J Holm and Toshiji Kawagoe. Face-to-face lying–an experimental study in Sweden and Japan. *Journal of Economic Psychology*, 31(3):310–321, 2010. doi: 10.1016/j.joep.2010.01.001.

[144] Turid Horgen. *Excellence by design: Transforming workplace and work practice*. John Wiley & Sons, 1999.

[145] G Hu, A K Jain, and G Stockman. Shape from light stripe texture. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '86)*, pages 412–414, Miami, #jun# 1986.

[146] H. Hua, L. D. Brown, and C. Gao. Scape: supporting stereoscopic collaboration in augmented and projective environments. *IEEE Computer Graphics and Applications*, 24(1):66–75, Jan 2004. ISSN 0272-1716. doi: 10.1109/MCG.2004.1255811.

[147] Hong Hua, Leonard D Brown, and Chunyu Gao. Scape: supporting stereoscopic collaboration in augmented and projective environments. *IEEE Computer Graphics and Applications*, 24(1):66–75, 2004. doi: 10.1109/mcg.2004.1255811.

[148] Yusuke Ichikawa, Ken-ichi Okada, and Giseok Jeong. Majic videoconferencing system: Experiments, evaluation and improvement'. *Proceedings of . . .* , pages 279–292, 1995. doi: 10.1007/978-94-011-0349-7_18.

[149] Masahiko Inami, Naoki Kawakami, Dairoku Sekiguchi, Yasuyuki Yanagida, Taro Maeda, and Susumu Tachi. Visuo-haptic display using head-mounted projector. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 233–240. IEEE, 2000.

[150] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems (TOIS)*, 11(4):349–375, 1993. doi: 10.1145/159764.159762.

[151] Kazuko Itoh, Hiroyasu Miwa, Yoshitaka Onishi, Kazutaka Imanishi, Kouki Hayashi, and Atsuo Takanishi. Development of face robot to express the individual face by optimizing the facial features. In *Humanoid Robots, 2005 5^{th} IEEE-RAS International Conference on*, pages 412–417. IEEE, 2005.

[152] Jana M Iverson and Susan Goldin-Meadow. Gesture paves the way for language development. *Psychological science*, 16(5):367–371, 2005. doi: 10.1111/j.0956-7976.2005.01542.x.

[153] Izdihar Jamil, Kenton O'Hara, Mark Perry, Abhijit Karnik, and Sriram Subramanian. The effects of interaction techniques on talk patterns in collaborative peer learning around interactive tables. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3043–3052. ACM, 2011. doi: 10.1145/1978942.1979393.

[154] Hans-Christian Jetter, Jens Gerken, Michael Zöllner, Harald Reiterer, and Natasa Milic-Frayling. Materializing the query with facet-streams: a hybrid surface for collaborative search on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3013–3022. ACM, 2011. doi: 10.1145/1978942.1979390.

[155] Mark St John, Michael B Cowen, Harvey S Smallman, and Heather M Oonk. The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1):79–98, 2001. doi: 10.1518/001872001775992534.

[156] Scott Johnson. What's in a representation, why do we care, and what does it mean? examining evidence from psychology. *Automation in Construction*, 8(1):15–24, 1998. doi: 10.1016/s0926-5805(98)00062-4.

[157] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. Headspin: a one-to-many

3d video teleconferencing system. In *ACM SIGGRAPH 2009 Emerging Technologies*, page 13. ACM, 2009. doi: 10.1145/1597956.1597969.

[158] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 637–644. ACM, 2014. doi: 10.1145/2642918.2647383.

[159] Edward Ellsworth Jones and Richard E Nisbett. The actor and the observer: Divergent perceptions of the causes of behavior. General Learning Press Morristown, NJ, 1971.

[160] Norman P Jouppi and Michael J Pan. Mutually-immersive audio telepresence. In *Audio Engineering Society Convention 113*. Audio Engineering Society, 2002.

[161] Norman P Jouppi, Subu Iyer, Stan Thomas, and April Slayden. Bireality: mutually-immersive telepresence. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 860–867. ACM, 2004. doi: 10.1145/1027527.1027725.

[162] Tejinder K Judge and Carman Neustaedter. Sharing conversation and sharing life: video conferencing in the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 655–658. ACM, 2010. doi: 10.1145/1753326.1753422.

[163] Yasuhero Kashagiri. Aiduti in japanese multi-party design conversations. 2007.

[164] Peter Kauff and Oliver Schreer. Virtual team user environments-a step from tele-cubicles towards distributed tele-collaboration in mediated workspaces. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 2, pages 9–12. IEEE, 2002.

[165] Peter Kauff and Oliver Schreer. An immersive 3D video-conferencing system using shared virtual team user environments. In *Proceedings of the 4th international conference on Collaborative virtual environments*, pages 105–112. ACM, 2002. doi: 10.1145/571878.571895.

[166] A Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26(1):22–63, 1967. ISSN 00016918. doi: 10.1016/0001-6918(67)90005-.

[167] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. Telehuman: Effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2531–2540. ACM, 2012. doi: 10.1145/2207676.2208640.

[168] Chris L Kleinke. Gaze and eye contact: a research review. *Psychological bulletin*, 100(1):78, 1986. doi: 10.1037//0033-2909.100.1.78.

[169] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33, 2013. doi: 10.1109/t-affc.2012.16.

[170] Patrick Klie, Torsten Büschenfeld, and Jörn Ostermann. Vipid-virtual 3d person models for intuitive dialog systems. In *IEEE Workshop on Content Generation and Coding for 3D-television*, volume 1, 2006.

[171] Nathan L Kluttz, Brandon R Mayes, Roger W West, and Dave S Kerby. The effect of head turn on the perception of gaze. *Vision research*, 49(15):1979–93, #jul# 2009. ISSN 1878-5646. doi: 10.1016/j.visres.2009.05.013.

[172] Daniel Knoblauch and Falko Kuester. Virtualizeme: interactive model reconstruction from stereo video streams. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 193–196. ACM, 2008. doi: 10.1145/1450579.1450620.

[173] Peter Kollock. Social dilemmas: The anatomy of cooperation. *Annual review of sociology*, pages 183–214, 1998. doi: 10.1146/annurev.soc.24.1.183.

[174] Robert M Krauss, Yihsiu Chen, and Purnima Chawla. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in experimental social psychology*, 28:389–450, 1996. doi: 10.1016/s0065-2601(08)60241-5.

[175] Annica Kristoffersson, Silvia Coradeschi, Kerstin Severinson Eklundh, and Amy Loutfi. Sense of presence in a robotic telepresence domain. In *Universal Access in Human-Computer Interaction. Users Diversity*, pages 479–487. Springer, 2011. doi: 10.1007/978-3-642-21663-3_51.

[176] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, 2013:3, 2013. doi: 10.1155/2013/902316.

[177] David M Krum, Evan A Suma, and Mark Bolas. Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing*, 16(1):17–26, 2012.

[178] Alexander Kulik, André Kunert, Stephan Beck, Roman Reichel, Roland Blach, Armin Zink, and Bernd Froehlich. C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. *ACM Transactions on Graphics (TOG)*, 30(6):188, 2011. doi: 10.1145/2024156.2024222.

[179] C Kuster, N Ranieri, A Agustina, H Zimmer, JC Bazin, C Sun, T Popa, and M Gross. Towards next generation 3D teleconferencing systems. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, pages 1–4. IEEE, 2012. doi: 10.1109/3dtv.2012.6365454.

[180] Claudia Kuster, Tiberiu Popa, Jean-Charles Bazin, Craig Gotsman, and Markus Gross. Gaze correction for home video conferencing. *ACM Transactions on Graphics (TOG)*, 31(6):174, 2012. doi: 10.1145/2366145.2366193.

[181] Edouard Lamboray, Stephan Wurmlin, and Markus Gross. Data streaming in telepresence environments. *Visualization and Computer Graphics, IEEE Transactions on*, 11(6):637–648, 2005. doi: 10.1109/tvcg.2005.98.

[182] SRH Langton. The mutual influence of gaze and head orientation in the analysis of social attention direction. Number November 2012, pages 825–845, 2000. doi: 10.1080/713755908.

[183] Stephen R. H. Langton. The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology Section A*, 53(3):825–845, 2000. doi: 10.1080/713755908. PMID: 10994231.

[184] Stephen R.H. Langton, Roger J. Watt, and Vicki Bruce. Do the eyes have it? cues to the direction of social attention. volume 4, pages 50–59, #feb# 2000. doi: 10.1016/S1364-6613(99)01436-9.

[185] Dora C Lau and J Keith Murnighan. Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, 23(2):325–340, 1998. doi: 10.2307/259377.

[186] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(2):150–162, 1994. doi: 10.1109/34.273735.

[187] Min Kyung Lee and Leila Takayama. Now, i have a body: Uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 33–42. ACM, 2011. doi: 10.1145/1978942.1978950.

[188] Nicolas H Lehment, Daniel Merget, and Gerhard Rigoll. Creating automatically aligned consensus realities for ar videoconferencing. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 201–206. IEEE, 2014. doi: 10.1109/ismar.2014.6948428.

[189] Peter Lincoln, Greg Welch, Andrew Nashel, Adrian Ilie, Andrei State, and Henry Fuchs. Animatronic shader lamps avatars. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 27–33. IEEE, 2009. doi: 10.1109/ismar.2009.5336503.

[190] Peter Lincoln, Greg Welch, Andrew Nashel, Andrei State, Adrian Ilie, and Henry Fuchs. Animatronic shader lamps avatars. *Virtual Reality*, 15(2-3): 225–238, #oct# 2010. doi: 10.1007/s10055-010-0175-5.

[191] Jack M Loomis, Jonathan W Kelly, Matthias Pusch, Jeremy N Bailenson, and Andrew C Beall. Psychophysics of perceiving eye-gaze and head direction with peripheral vision:¬†implications for the dynamics of eye-gaze behavior. volume 37, pages 1443–1457, 2008. doi: 10.1068/p5896.

[192] Jack M Loomis, Roberta L Klatzky, and Nicholas A Giudice. -sensory substitution of vision: Importance of perceptual and cognitive processing. In *Assistive technology for blindness and low vision*, pages 179–210. CRC Press, 2012.

[193] Danielle Lottridge, Nicolas Masson, and Wendy Mackay. Sharing empty moments: design for remote couples. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2329–2338. ACM, 2009. doi: 10.1145/1518701.1519058.

[194] Kok-Lim Low, Greg Welch, Anselmo Lastra, and Henry Fuchs. Life-sized projector-based dioramas. *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '01*, page 93, 2001. doi: 10.1145/505025.505026.

[195] Paul Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. Hands on hitchcock: embodied reference to a moving scene. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 43–52. ACM, 2011. doi: 10.1145/1978942.1978951.

[196] C Neil Macrae, Bruce M Hood, Alan B Milne, Angela C Rowe, and Malia F Mason. Are you looking at me? eye gaze and person perception. volume 13, pages 460–4, #sep# 2002. doi: 10.1111/1467-9280.00481.

[197] Hiroyuki Maeda, Tomohiro Tanikawa, Jun Yamashita, Koichi Hirota, and Michitaka Hirose. Real world video avatar: Transmission and presentation of human figure. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 237–238. IEEE, 2004.

[198] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 137–146, Oct 2011. doi: 10.1109/ISMAR.2011.6092379.

[199] Andrew Maimone and Henry Fuchs. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 137–146. IEEE, 2011. doi: 10.1109/ismar.2011.6092379.

[200] Andrew Maimone and Henry Fuchs. A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall. *Proceedings of ICAT 2011, to appear*, pages 4–9, 2011.

[201] Andrew Maimone, Xubo Yang, Nate Dierk, Andrei State, Mingsong Dou, and Henry Fuchs. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In *Virtual Reality (VR), 2013 IEEE*, pages 23–26. IEEE, 2013. doi: 10.1109/vr.2013.6549352.

[202] Stefan Marti and Chris Schmandt. Physical embodiments for mobile communication agents. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 231–240. ACM, 2005. doi: 10.1145/1095034.1095073.

[203] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J Gortler, and Leonard McMillan. Image-based visual hulls. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 369–374. ACM Press/Addison-Wesley Publishing Co., 2000. doi: 10.1145/344779.344951.

[204] Joanna McGrenere and Wayne Ho. Affordances: Clarifying and evolving a concept. In *Graphics interface*, volume 2000, pages 179–186, 2000.

[205] Rufael Mekuria, Michele Sanna, Stefano Asioli, Ebroul Izquierdo, Dick CA Bulterman, and Pablo Cesar. A 3D tele-immersion system based on live

captured mesh geometry. In *Proceedings of the 4th ACM Multimedia Systems Conference*, pages 24–35. ACM, 2013. doi: 10.1145/2483977.2483980.

[206] Alaeddine Mihoub, Gérard Bailly, Christian Wolf, and Frédéric Elisei. Learning multimodal behavioral models for face-to-face social interaction. *Journal on Multimodal User Interfaces*, 9(3):195–210, 2015. doi: 10.1007/s12193-015-0190-7.

[207] Wade J Mitchell, Kevin A Szerszen Sr, Amy Shirong Lu, Paul W Schermerhorn, Matthias Scheutz, and Karl F MacDorman. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1):10, 2011. doi: 10.1068/i0415.

[208] P.R. Monge, C. McSween, and J.A. Wyer. *A Profile of Meetings in Corporate America: Results of the 3M Meeting Effectiveness Study*. Annenberg School of Communications, University of Southern California, 1989. URL https://books.google.co.uk/books?id=RPjbtgAACAAJ.

[209] Andrew F Monk and Caroline Gale. A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33 (3):257–278, 2002. doi: 10.1207/s15326950dp3303_4.

[210] Carl Moore, Toby Duckworth, Rob Aspin, and David Roberts. Synchronization of Images from Multiple Cameras to Reconstruct a Moving Human. In *IEEE/ACM 14th International Symposium on Distributed Simulation and Real Time Applications (DSRT '10)*, pages 53–60, Fairfax, #oct# 2010. IEEE. doi: 10.1109/ds-rt.2010.15.

[211] Carl Moore, Tobias Duckworth, and David J Roberts. Investigating the Suitability of a Software Capture Trigger in a 3D Reconstruction System for Telepresence. *IEEE/ACM 15th International Symposium on Distributed Simulation and Real Time Applications (DSRT '11)*, pages 134–137, #sep# 2011. doi: 10.1109/ds-rt.2011.18.

[212] Masahiro Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

[213] Jesper Mortensen, V Vinayagamoorthy, Mel Slater, Anthony Steed, B Lok, and MC Whitton. Collaboration in tele-immersive environments. In *ACM International Conference Proceeding Series*, volume 23, pages 93–101. Citeseer, 2002.

[214] Samer Al Moubayed, Jens Edlund, and Jonas Beskow. Taming mona lisa. volume 1, pages 1–25, #jan# 2012. doi: 10.1145/2070719.2070724.

[215] Parri Munsell. Skype and messenger coming together: The next chapter. web blog, #feb# 2013. URL http://blogs.skype.com/2013/02/15/skype-and-messenger-coming-together-the-next-chapter/#fbid=oFh6i6Q6aMm.

[216] Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, and John Rae. An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 3(4):8, 2007. doi: 10.1145/1314303.1314311.

[217] Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, John Rae, and Robin Wolff. Eye gaze in virtual environments: evaluating the need and initial work on implementation. *Concurrency and Computation: Practice and Experience*, 21(11):1437–1449, #aug# 2009. ISSN 15320626. doi: 10.1002/cpe.1396.

[218] Arjun Nagendran, R Pillat, C Hughes, and Greg Welch. Continuum of virtual-human space: Towards improved interaction strategies for physical-virtual avatars. doi: 10.1145/2407516.2407554.

[219] Rodney Napier and Matti K Gershenfeld. *Making groups work: A guide for group leaders*. Houghton Mifflin Boston, 1983.

[220] Rodney W Napier and Matti K Gershenfeld. *Groups: Theory and experience.* Houghton Mifflin, 1973. doi: 10.2307/581729.

[221] David Nguyen and John Canny. Multiview: spatially faithful group video conferencing. In *Proceedings of the SIGCHI conference on human factors*

in computing systems, pages 799–808. ACM, 2005. doi: 10.1145/1054972.1055084.

[222] David T Nguyen and John Canny. More than face-to-face: empathy effects of video framing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 423–432. ACM, 2009. doi: 10.1145/1518701.1518770.

[223] Donald A Norman. The psychopathology of everyday things. *Foundations of cognitive psychology: core readings. MIT Press, Cambridge, MA*, pages 417–443, 2002.

[224] James Norris, Holger Schnädelbach, and Guoping Qiu. Camblend: an object focused collaboration tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 627–636. ACM, 2012. doi: 10.1145/2207676.2207765.

[225] James Norris, Holger M Schnädelbach, and Paul K Luff. Putting things in focus: establishing co-orientation through video in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1329–1338. ACM, 2013. doi: 10.1145/2470654.2466174.

[226] DG Novick, Brian Hansen, and Karen Ward. Coordinating turn-taking with gaze. 1996.

[227] Kristen Nowak. Defining and differentiating copresence, social presence and presence as transportation. In *Presence 2001 Conference, Philadelphia, PA*, pages 1–23. Citeseer, 2001.

[228] J O'Hare, RCA Bendall, J Rae, Graham Thomas, B Weir, and DJ Roberts. Is this seat taken?: behavioural analysis of the telethrone: a novel situated telepresence display. In *Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments*, pages 99–106. Eurographics Association, 2016.

[229] John OHare, Allen J Fairchild, Robin Wolff, and David J Roberts. Telethrone reconstructed; ongoing testing toward a more natural situated display. In *Augmented Reality and Virtual Reality*, pages 323–337. Springer, 2018.

[230] Kohtaro Ohba, Takehito Tsukada, Tetsuo Kotoku, and Kazuo Tanie. Facial expression space for smooth tele-communications. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 378–383. IEEE, 1998.

[231] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. Multiparty videoconferencing at virtual social distance: MAJIC design. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)*, pages 385–393, Chapel Hill, #oct# 1994. doi: 10.1145/192844.193054.

[232] Masatoshi Okutomi and Takeo Kanade. A multiple-baseline stereo. *IEEE Transactions on pattern analysis and machine intelligence*, 15(4):353–363, 1993. doi: 10.1109/34.206955.

[233] C O'Malley and Steve Langton. Comparison of face-to-face and video-mediated interaction. volume 2, pages 177–192, 1996. doi: 10.1016/0953-5438(96)01027-2.

[234] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754. ACM, 2016. doi: 10.1145/2984511.2984517.

[235] Camilla Ostrem. A behavior analytic perspective on gambling behavior. Høgskolen i Oslo og Akershus, 2014.

[236] Kazuhiro Otsuka, Yoshinao Takemae, and Junji Yamato. A probabilistic inference of multiparty-conversation structure based on markov-switching models of gaze patterns, head directions, and utterances. In *Proceedings of*

the 7th international conference on Multimodal interfaces, pages 191–198. ACM, 2005. doi: 10.1145/1088463.1088497.

[237] Kazuhiro Otsuka, Shiro Kumano, Ryo Ishii, Maja Zbogar, and Junji Yamato. Mm+ space: nx 4 degree-of-freedom kinetic display for recreating multiparty conversation spaces. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 389–396. ACM, 2013. doi: 10.1145/2522848.2522854.

[238] Maximilian Ott, John P Lewis, and Ingemar Cox. Teleconferencing eye contract using a virtual camera. In *INTERACT'93 and CHI'93 Conference Companion on Human Factors in Computing Systems*, pages 109–110. ACM, 1993. doi: 10.1145/259964.260136.

[239] Jiazhi Ou, Lui Min Oh, Jie Yang, and Susan R Fussell. Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 231–240. ACM, 2005. doi: 10.1145/1054972.1055005.

[240] Oyewole Oyekoya, William Steptoe, and Anthony Steed. Sphereavatar: A situated display to represent a remote collaborator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2551–2560, 2012. ISBN 9781450310154. doi: 10.1145/2207676.2208642.

[241] Xueni Pan, Marco Gillies, and Mel Slater. Male bodily responses during an interaction with a virtual woman. In *Intelligent Virtual Agents*, pages 89–96. Springer, 2008.

[242] Ye Pan and Anthony Steed. Preserving gaze direction in teleconferencing using a camera array and a spherical display. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, pages 1–4. IEEE, 2012. doi: 10.1109/3dtv.2012.6365433.

[243] Ye Pan and Anthony Steed. A gaze-preserving situated multiview telepresence system. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems, pages 2173–2176. ACM, 2014. doi: 10.1145/2556288.2557320.

[244] Ye Pan, William Steptoe, and Anthony Steed. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. In *Proceedings of the annual ACM conference on Human factors in computing systems*, pages 1397–1406. ACM, 2014. doi: 10.1145/2556288.2557276.

[245] Kyoung Shin Park and Robert V Kenyon. Effects of network characteristics on human performance in a collaborative virtual environment. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 104–111. IEEE, 1999.

[246] H. M. Parsons. What happened at hawthorne? *Science*, 183(4128):922–932, 1974. ISSN 0036-8075. doi: 10.1126/science.183.4128.922. URL http://science.sciencemag.org/content/183/4128/922.

[247] Eric Paulos and John Canny. Prop: personal roving presence. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 296–303. ACM Press/Addison-Wesley Publishing Co., 1998. doi: 10.1145/274644.274686.

[248] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 1716–1725. ACM, 2016. doi: 10.1145/2818048.2819965.

[249] Benjamin Petit, Jean-Denis Lesage, Jean-Sébastien Franco, Edmond Boyer, and Bruno Raffin. Grimage: 3D modeling for remote collaboration and telepresence. In *ACM Symposium on Virtual Reality Software and Technology (VRST '08)*, pages 299–300, Bordeaux, #oct# 2008. ACM. doi: 10.1145/1450579.1450662.

[250] Benjamin Petit, Jean-Denis Lesage, Clément Menier, Jérémie Allard, Jean-Sébastien Franco, Bruno Raffin, Edmond Boyer, and François Faure. Multicamera Real-Time 3D Modeling for Telepresence and Remote Collaboration.

International Journal of Digital Multimedia Broadcasting, 2010:1–12, 2010. ISSN 1687-7578. doi: 10.1155/2010/247108.

[251] Stephen Porter and Leanne ten Brinke. Reading between the lies identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5):508–514, 2008. doi: 10.1111/j.1467-9280.2008.02116.x.

[252] ProjectionMappingCentral. History of projection mapping. URL http://projection-mapping.org/the-history-of-projection-mapping/.

[253] H. Qing-xi, L. Tao, and Y. Yuan. An easy system of spatial points collection based on artoolkit. In *2009 WRI World Congress on Computer Science and Information Engineering*, volume 1, pages 582–586, March 2009. doi: 10.1109/CSIE.2009.680.

[254] Irene Rae, Leila Takayama, and Bilge Mutlu. One of the gang: supporting in-group behavior for embodied mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3091–3100. ACM, 2012. doi: 10.1145/2207676.2208723.

[255] Irene Rae, Leila Takayama, and Bilge Mutlu. In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1921–1930. ACM, 2013. doi: 10.1145/2470654.2466253.

[256] John Rae. Organizing participation in interaction: Doing participation framework. *Research on Language and Social Interaction*, 34(2):253–278, 2001. doi: 10.1207/s15327973rlsi34-2_4.

[257] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998. doi: 10.1145/280814.280861.

[258] Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*, pages 89–102. Springer, 2001. doi: 10.1007/978-3-7091-6242-2_9.

[259] Reflecmedia. Reflecmedia chromatte information. URL http://www.reflecmedia.com/broadcast/products/chromatte/index.htm.

[260] Holger Regenbrecht and Tobias Langlotz. Mutual gaze support in videoconferencing reviewed. *CAIS*, 37:45, 2015.

[261] Holger Regenbrecht and Tobias Langlotz. Mutual gaze support in videoconferencing reviewed. *CAIS*, 37:45, 2015.

[262] Rutger Rienks, Ronald Poppe, and Dirk Heylen. Differences in head orientation behavior for speakers and listeners. *ACM Transactions on Applied Perception*, 7(1):1–13, #jan# 2010. ISSN 15443558. doi: 10.1145/1658349.1658351.

[263] Evan F Risko and Alan Kingstone. Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, 73(2): 291–296, 2011.

[264] Diego Rivera-Gutierrez, Greg Welch, Peter Lincoln, Mary Whitton, Juan Cendan, David a Chesnutt, Henry Fuchs, and Benjamin Lok. Shader Lamps Virtual Patients: the physical manifestation of virtual patients. *Studies in health technology and informatics*, 173:372–8, #jan# 2012. ISSN 0926-9630.

[265] D Roberts, A Al-Liabi, R Wolff, O Otto, and A Al-Khalifah. Reducing fragmentation in telecollaboration by using IPt interfaces. In *Proceedings of the 11th Eurographics conference on Virtual Environments*, pages 211–216. Eurographics Association, 2005.

[266] Daniel M Roberts, Robin Wolff, Oliver Otto, and Anthony Steed. Constructing a gazebo: supporting teamwork in a tightly coupled, distributed

task in virtual reality. *Presence*, 12(6):644–657, 2003. doi: 10.1162/105474603322955932.

[267] David Roberts. Talking with the Furniture. 2009.

[268] David Roberts, Toby Duckworth, Carl Moore, Robin Wolff, and John O'Hare. Comparing the end to end latency of an immersive collaborative environment and a video conference. In *Proceedings of the 2009 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, pages 89–94. IEEE Computer Society, 2009. doi: 10.1109/ds-rt.2009.43.

[269] David Roberts, Toby Duckworth, Carl Moore, Robin Wolff, and John O'Hare. Comparing the end to end latency of an immersive collaborative environment and a video conference. *IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications (DSRT '09)*, pages 89–94, #oct# 2009. doi: 10.1109/ds-rt.2009.43.

[270] David Roberts, Robin Wolff, John Rae, Anthony Steed, Rob Aspin, Moira McIntyre, Adriana Pena, Oyewole Oyekoya, and William Steptoe. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *Virtual reality conference, 2009. VR 2009. IEEE*, pages 135–142. IEEE, 2009. doi: 10.1109/vr.2009.4811013.

[271] David J Roberts, Robin Wolff, and Oliver Otto. Supporting a closely coupled task between a distributed team: using immersive virtual reality technology. *Computing and Informatics*, 24(1):7–29, 2012.

[272] David J Roberts, John Rae, Tobias W Duckworth, Carl M Moore, and Rob Aspin. Estimating the gaze of a virtuality human. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):681–690, 2013. doi: 10.1109/tvcg.2013.30.

[273] David J Roberts, Allen J Fairchild, Simon P Campion, John O'Hare, Carl M Moore, Rob Aspin, Tobias Duckworth, Paolo Gasparello, and Franco Tecchia. withyou—an experimental end-to-end telepresence system using

video-based reconstruction. *Selected Topics in Signal Processing, IEEE Journal of*, 9(3):562–574, 2015. doi: 10.1109/jstsp.2015.2402635.

[274] Nicholas C Romano and Jay F Nunamaker. Meeting analysis: Findings from research and practice. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, pages 13–pp. IEEE, 2001.

[275] Adolph H Rosenthal. Two-way television communication unit, May 6 1947. US Patent 2,420,198.

[276] Amela Sadagic, Herman Towles, Loring Holden, Kostas Daniilidis, and Bob Zeleznik. Tele-immersion portal: towards an ultimate synthesis of computer graphics and computer vision systems. In *4^{th} Annual International Workshop on Presence*, pages 21–23, 2001.

[277] Daisuke Sakamoto, Takayuki Kanda, Tetsuo Ono, Hiroshi Ishiguro, and Norihiro Hagita. Android as a telecommunication medium with a human-like presence. In *Human-Robot Interaction (HRI), 2007 ACM/IEEE International Conference on*, pages 193–200. IEEE, 2007. doi: 10.1145/1228716.1228743.

[278] Nils Christian Sauer and Simone Kauffeld. The structure of interaction at meetings: A social network analysis. *Zeitschrift für Arbeits-und Organisationspsychologie A&O*, 2016. doi: 10.1026/0932-4089/a000201.

[279] Emanuel A Schegloff. Body torque. *Social Research*, pages 535–596, 1998.

[280] Diane J Schiano, Sheryl M Ehrlich, and Kyle Sheridan. Categorical imperative not: facial affect is perceived continuously. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–56. ACM, 2004. doi: 10.1145/985692.985699.

[281] Oliver Schreer, Ralf Tanger, Peter Eisert, Peter Kauff, Bernhard Kaspar, and Roman Englert. Real-time avatar animation steered by live body motion. In *Image Analysis and Processing–ICIAP 2005*, pages 147–154. Springer, 2005. doi: 10.1007/11553595_18.

[282] Oliver Schreer, Ingo Feldmann, Nicole Atzpadin, Peter Eisert, Peter Kauff, and HJW Belt. 3d presence-a system concept for multi-user and multi-party immersive 3d video conferencing. IET, 2008. doi: 10.1049/cp:20081083.

[283] Ralph Schroeder, Anthony Steed, Ann-Sofie Axelsson, Ilona Heldal, Åsa Abelin, Josef Wideström, Alexander Nilsson, and Mel Slater. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics*, 25(5):781–788, 2001. doi: 10.1016/s0097-8493(01)00120-0.

[284] Martijn J Schuemie, Peter Van Der Straaten, Merel Krijn, and Charles APG Van Der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001. doi: 10.1089/109493101300117884.

[285] Cliff W Scott, Linda Rhoades Shanock, and Steven G Rogelberg. Meetings at work: Advancing the theory and practice of meetings. *Small Group Research*, 43(2):127–129, 2012.

[286] Abigail Sellen, Bill Buxton, and John Arnott. Using spatial cues to improve videoconferencing. *Proceedings of the SIGCHI conference on . . .* , pages 651–652, 1992. doi: 10.1145/142750.143070.

[287] Abigail Sellen, Bill Buxton, and John Arnott. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 651–652. ACM, 1992. doi: 10.1145/142750.143070.

[288] Abigail J Sellen. Remote conversations: The effects of mediating talk with technology. *Human-computer interaction*, 10(4):401–444, 1995. doi: 10.1207/s15327051hci1004_2.

[289] Paul Sermon. Telematic dreaming. *Leonardo*, 33(2):90–90, 2000. doi: 10.1162/leon.2000.33.2.90.

[290] Paul Sermon. Over one hundred years of telepresence. 2016.

[291] Paul Sermon and Charlotte Gould. Occupy the screen. 2013.

[292] Suleman Shahid, Emiel Krahmer, and Marc Swerts. Video-mediated and co-present gameplay: Effects of mutual gaze on game experience, expressiveness and perceived social presence. *Interacting with Computers*, 24(4): 292–305, 2012. doi: 10.1016/j.intcom.2012.04.006.

[293] J. Short, E. Williams, and B. Christie. *The Social Psychology of Telecommunications*. Wiley, 1976. ISBN 9780471015819. doi: 10.2307/2065899. URL https://books.google.co.uk/books?id=Ze63AAAAIAAJ.

[294] Donald E. Simanek. The illusion of reality in stereoscopy. URL https://www.lhup.edu/~dsimanek/3d/stereo/reality.htm.

[295] Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, 1999. doi: 10.1162/105474699566477.

[296] Mel Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4):484–493, 2004. doi: 10.1162/1054746041944849.

[297] Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder. Small-group behavior in a virtual and real environment: A comparative study. volume 9, pages 37–51. MIT Press, 2000. doi: 10.1162/105474600566600.

[298] Philip E Slater. Contrasting correlates of group size. *Sociometry*, 21(2): 129–139, 1958. doi: 10.2307/2785897.

[299] Gillian Slessor, Louise H Phillips, and Rebecca Bull. Age-related declines in basic social perception: evidence from tasks assessing eye-gaze processing. *Psychology and aging*, 23(4):812, 2008. doi: 10.1037/a0014348.

[300] Petr Slovák, Pavel Troubil, and Petr Holub. Gcoll: a flexible videoconferencing environment for group-to-group interaction. In *Human-Computer Interaction–INTERACT 2009*, pages 165–168. Springer, 2009. doi: 10.1007/978-3-642-03658-3_22.

[301] Dave Snowdon, Chris Greenhalgh, and Steve Benford. What you see is not what i see: Subjectivity in virtual environments. *Framework for Immersive Virtual Environments (FIVE95)*, 1995.

[302] Jonathan Starck and Adrian Hilton. Model-based human shape reconstruction from multiple views. *Computer Vision and Image Understanding*, 111 (2):179–194, 2008. doi: 10.1016/j.cviu.2007.10.001.

[303] JR Starck. *Human modelling from multiple views*. PhD thesis, Citeseer, 2003.

[304] Jim Steinmeyer. *The Science Behind the Ghost!: A Brief History of Pepper's Ghost*. Hahne, 2013.

[305] W Steptoe, A Steed, A Rovira, and J Rae. Lie tracking: Social presence, truth and deception in avatar-mediated telecommunication. CHI '10, pages 1039–1048. ASSOC COMPUTING MACHINERY, 2010. doi: 10.1145/1753326.1753481.

[306] WAH Steptoe. *Eye tracking and avatar-mediated communication in immersive collaborative virtual environments*. PhD thesis, UCL (University College London), 2010.

[307] William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts, and Anthony Steed. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08*, page 197, 2008. doi: 10.1145/1460563.1460593.

[308] William Steptoe, Oyewole Oyekoya, Alessio Murgia, Robin Wolff, John Rae, Estefania Guimaraes, David Roberts, and Anthony Steed. Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 83–90. IEEE, 2009. doi: 10.1109/vr.2009.4811003.

[309] Rainer Stiefelhagen. Tracking focus of attention in meetings. In *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pages 273–280. IEEE, 2002.

[310] Rainer Stiefelhagen and Jie Zhu. Head orientation and gaze direction in meetings. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 858–859, New York, NY, USA, 2002. ACM. ISBN 1-58113-454-1. doi: 10.1145/506443.506634.

[311] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Estimating focus of attention based on gaze and sound. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–9. ACM, 2001. doi: 10.1145/971478.971505.

[312] Megan Strait and Matthias Scheutz. Measuring users' responses to humans, robots, and human-like robots with functional near infrared spectroscopy. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 1128–1133. IEEE, 2014. doi: 10.1109/roman.2014.6926403.

[313] Phil Surman, Sally Day, Xianzi Liu, Joshua Benjamin, Hakan Urey, and Kaan Aksit. Head tracked retroreflecting 3d display. *Journal of the Society for Information Display*, 23(2):56–68, 2015. ISSN 1938-3657. doi: 10.1002/jsid.295. URL http://dx.doi.org/10.1002/jsid.295. SID-01-14-0154.R2.

[314] Lawrence A Symons, Kang Lee, Caroline C Cedrone, and Mayu Nishimura. What are you looking at? acuity for triadic eye gaze. volume 131, pages 451–69, #oct# 2004.

[315] Cisco Systems. Heres looking at you: Eye contact and gaze perspectivein the two dimensional world of telepresence.

[316] Susumu Tachi. Telexistence and retro-reflective projection technology (rpt). In *Proceedings of the 5th Virtual Reality International Conference (VRIC2003) pp*, volume 69, pages 1–69, 2003.

[317] Susumu Tachi, Naoki Kawakami, Masahiko Inami, and Yoshitaka Zaitsu. Mutual telexistence system using retro-reflective projection technology. *International Journal of Humanoid Robotics*, 1(01):45–64, 2004. doi: 10.1142/s0219843604000095.

[318] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1181–1190. ACM, 2006. doi: 10.1145/1124772.1124950.

[319] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C Tang, and Bill Buxton. Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 271–280. ACM, 2010. doi: 10.1145/1718918.1718969.

[320] John C Tang, Robert Xiao, Aaron Hoff, Gina Venolia, Patrick Therien, and Asta Roseway. Homeproxy: exploring a physical proxy for video communication in the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1339–1342. ACM, 2013. doi: 10.1145/2470654.2466175.

[321] Savaş Tay, P-A Blanche, R Voorakaranam, AV Tunç, W Lin, S Rokutanda, T Gu, D Flores, P Wang, G Li, et al. An updatable holographic three-dimensional display. *Nature*, 451(7179):694, 2008. doi: 10.1038/nature06596.

[322] Lucas Teixeira, Alberto Raposo, and Marcelo Gattass. Indoor localization using slam in parallel with a natural marker detector. pages 27–33, 03 2013. doi: 10.1145/2480362.2480370.

[323] Pter Tams Kovcs Tibor Balogh. Real-time 3d light field transmission, 2010. URL https://doi.org/10.1117/12.854571.

[324] Stella Ting-Toomey and Leeva C Chung. *Understanding intercultural communication*. Oxford University Press New York, NY, 2012.

[325] Google Trends. Google trends for projection mapping. URL https://www.google.co.uk/trends/explore#q=%22projection%20mapping%22.

[326] Nikolaus F Troje and Ulrike Siebeck. Illumination-induced apparent shift in orientation of human heads. *Perception*, 27(6):671–680, 1998. doi: 10.1068/p270671.

[327] Edward Tse, Chia Shen, Saul Greenberg, and Clifton Forlines. How pairs interact over a multimodal digital table. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 215–218. ACM, 2007. doi: 10.1145/1240624.1240659.

[328] Katherine M Tsui, Munjal Desai, Holly A Yanco, and Chris Uhlik. Exploring use cases for telepresence robots. In *Human-Robot Interaction (HRI), 2011 ACM/IEEE International Conference on*, pages 11–18. IEEE, 2011. doi: 10.1145/1957656.1957664.

[329] Katherine M Tsui, Munjal Desai, and Holly A Yanco. Towards measuring the quality of interaction: Communication through telepresence robots. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, pages 101–108. ACM, 2012. doi: 10.1145/2393091.2393112.

[330] Philip Tuddenham and Peter Robinson. Territorial coordination and workspace awareness in remote tabletop collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2139–2148. ACM, 2009. doi: 10.1145/1518701.1519026.

[331] UNstats. Deputy un chief calls for urgent action to tackle global sanitation crisis. URL http://www.un.org/apps/news/story.asp?NewsID=44452&Cr=sanitation&Cr1=#.Vwtx7vkrK5P.

[332] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence*, 9(5):497–503, 2000. doi: 10.1162/105474600566989.

[333] Joy Van Baren and Wijnand IJsselsteijn. *Measuring presence: A guide to current measurement approaches*. 2004.

[334] Rob van Eijk, André Kuijsters, Klaske Dijkstra, and Wijnand A IJsselsteijn. Human sensitivity to eye contact in 2D and 3D videoconferencing. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 76–81. IEEE, 2010. doi: 10.1109/qomex.2010.5518040.

[335] Various. *Distributed Work*. MIT Press, 2002.

[336] Roel Vertegaal. Catching the eye: management of joint attention in cooperative work. Number 29, 1997. doi: 10.1145/270950.270998.

[337] Roel Vertegaal and Yaping Ding. Explaining effects of eye gaze on mediated group conversations:: amount or synchronization? In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 191–198. IEEE, 2002. doi: 10.1145/587078.587085.

[338] Roel Vertegaal and Yaping Ding. Explaining effects of eye gaze on mediated group conversations:: amount or synchronization? pages 278–285, 2002. ISBN 1581135602. doi: 10.1145/587078.587085.

[339] Roel Vertegaal and R Slagter. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. pages 301–308, 2001. ISBN 1581133278. doi: 10.1145/365024.365119.

[340] Roel Vertegaal and Gerrit van der Veer. Effects of gaze on multiparty mediated communication. In *In Proceedings of Graphics Interface*, pages 95–102. Morgan Kaufmann, 2000.

[341] Roel Vertegaal and Ivo Weevers. GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *Proceedings of the SIGCHI . . .* , (5):521–528, 2003. doi: 10.1145/642700.642702.

[342] Dhanraj Vishwanath, Ahna R Girshick, and Martin S Banks. Why pictures look right when viewed from the wrong place. volume 8, pages 1401–1410. Nature Publishing Group, 2005. doi: 10.1038/nn1553.

[343] Amy Voida, Nathan Bos, Judith Olson, Gary Olson, and Lauren Dunning. Cross-cutting faultlines of location and shared identity in the intergroup cooperation of partially distributed groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3101–3110. ACM, 2012. doi: 10.1145/2207676.2208724.

[344] Wolfgang Waizenegger, Nicole Atzpadin, Oliver Schreer, and Ingo Feldmann. Patch-sweeping with robust prior for high precision depth estimation in real-time systems. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 881–884. IEEE, 2011. doi: 10.1109/icip.2011.6116699.

[345] Joseph B Walther and Malcolm R Parks. Cues filtered out, cues filtered in. *Handbook of interpersonal communication*, 3:529–563, 2002.

[346] O. Michael Watson and Theodore D. Graves. Quantitative Research in Proxemic Behavior. *American Anthropologist*, 68(4):971–985, #aug# 1966. ISSN 0002-7294. doi: 10.1525/aa.1966.68.4.02a00070.

[347] Ederyn Williams. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5):963, 1977. doi: 10.1037//0033-2909.84.5.963.

[348] Hugh R Wilson, Frances Wilkinson, Li-Ming Lin, and Maja Castillo. Perception of head orientation. volume 40, pages 459–472. Elsevier, 2000. doi: 10.1016/s0042-6989(99)00195-9.

[349] Robin Wolff, Dave Roberts, Alessio Murgia, Norman Murray, John Rae, Will Steptoe, Anthony Steed, and Paul Sharkey. Communicating eye gaze across a distance without rooting participants to the spot. pages 111–118. Ieee, #oct# 2008. ISBN 978-0-7695-3425-1. doi: 10.1109/DS-RT.2008.28.

[350] William Hyde Wollaston. On the apparent direction of eyes in a portrait. pages 247–256. JSTOR, 1824. doi: 10.1098/rstl.1824.0016.

[351] Nick Yee, Jeremy N Bailenson, and Kathryn Rickertsen. A meta-analysis of the impact of the inclusion and realism of human-like faces on user

experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1–10. ACM, 2007. doi: 10.1145/1240624.1240626.

[352] Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. volume 10, pages 115–21, #feb# 2007. doi: 10.1089/cpb.2006.9984.

[353] Tomohiro Yendo, Toshiaki Fujii, Masayuki Tanimoto, and Mehrdad Panahpour Tehrani. The seelinder: Cylindrical 3D display viewable from 360 degrees. *Journal of visual communication and image representation*, 21(5): 586–594, 2010. doi: 10.1016/j.jvcir.2009.10.004.

[354] Cha Zhang. Multiview imaging and 3DTV. *IEEE Signal Processing Magazine*, 1053(5888/07), 2007. doi: 10.1109/msp.2007.905873.

[355] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 851–860. ACM, 2013. doi: 10.1145/2470654.2470775.

[356] Jakob Zillner, Christoph Rhemann, Shahram Izadi, and Michael Haller. 3d-board: a whole-body remote collaborative whiteboard. pages 471–479, 2014. doi: 10.1145/2642918.2647393.

Appendix A

Appendices

A.1 New Version of withyou, CROSSDRIVE

Rendering

A series of feature requests were generated for the programmers who maintain the previous codebase. The geometry streamer software was modified by Fairchild and Wolff to include new features important for this continued research.

This included:

1. A new server and client for the local system which allows synchronised playback from 5 multiple viewpoints
2. Window height, width, and screen origin can be set from the command line
3. Network integration of a vrpn server including object identifiers to allow multiple tracked projectors
4. Parameters to lock the view to a location in x, y, z space, including a field for the distance of the virtual camera from the target location

5. Controls to move the model origin within the scene (necessary to correctly align the model with the capture coordinate system)
6. Controls to adjust the tracking origin within the scene (necessary because setting the playback origin correctly with the vicon wand was difficult in the experimental space)
7. Frustum controls for the projector such as field of view (necessary because different types of projectors were used)
8. Ability to draw target axes for the lookat feature (Useful for debugging the system during development and alignment)
9. A switch between disk and network operation (between real-time and experimental conditions)
10. Sync to system time to allow correct playback against the recorded audio for the experiment

The software has been written to be source agnostic in that a local server takes input from disk alongside the tracking data, and serves to an internal 127.0.0.1 address the correct number of streams for display by multiple client windows.

For instance, the Infocus projector is opened with a window 1080 pixels wide and 1920 pixels high. It is instructed to open 2560 pixels into the desktop, which moves the window off the first screen and centres it on the Infocus projector. The software is instructed to open the pre-recorded file from disk with an offset applied to the position of the model which ensures everything is correctly positioned. The horizontal field of view is set to 30 degrees. The lookat point is passed coordinates from the command line such that it is located in the head of the recorded model. The lookat distance is locked to 2.7m to match the layout of the room and give correct scale. The command line passes the location of the Infocus projector tracking data to the software as Tracker0. The batch file passes correct details for each projection station.

A.2 Ethical approval forms

Academic Audit and Governance Committee

College of Science and Technology Research Ethics Panel (CST)

MEMORANDUM

To	**John O'Hare**
cc:	**Professor Sunil Vadera, Head of School of CSE**
From	**Nathalie Audren Howarth, College Research Support Officer**
Date	**20/04/2015**

Subject: **Approval of your Project by CST**

Project Title: **Attention in Telepresence: A systematic comparison of tele-collaboration approaches and their ability to introduce non verbal cues normally associated with face to face meetings.**

REP Reference: **CST 15/16**

Following your responses to the Panel's queries, based on the information you provided, I can confirm that they have no objections on ethical grounds to your project.

If there are any changes to the project and/or its methodology, please inform the Panel as soon as possible.

Regards,

Nathalie Audren Howarth
College Research Support Officer

For enquiries please contact:
College of Science and Technology
College Research Support Officer
The University of Salford
Maxwell building, (7th floor, room 721)
Telephone: 0161 295 5278
Email: n.audren@salford.ac.uk

Academic Audit and Governance Committee

College of Science and Technology Research Ethics Panel (CST)

University of Salford MANCHESTER

MEMORANDUM

To	**John O'Hare (and Prof Dave Roberts)**
cc:	**Prof Sunil Vadera, Head of School of CSE**
From	**Nathalie Audren Howarth**, **College Research Support Officer**
Date	**11th May 2012**

Subject: **Approval of your Project by CST**

Project Title: **Attention in Telepresence: A systematic comparison of tele-collaboration**

REP Reference: **CST 12/11**

Following your responses to the Panel's queries, based on the information you provided, I can confirm that they have no objections on ethical grounds to your project.

If there are any changes to the project and/or its methodology, please inform the Panel as soon as possible.

Regards,

Nathalie Audren Howarth
College Research Support Officer

For enquiries please contact:
College of Science and Technology
College Research Support Officer
The University of Salford
Maxwell building, (7th floor, room 721)
Telephone: 0161 295 5278
Email: n.audren@salford.ac.uk

PGR Ethics Audit

100% complete

Thank you

Completion receipt

Receipt number: 299740-299732-25969664
Submission time: 2017-10-11 15:53:50 BST

A.3 Survey for poker

Edit this form

Participant survey

By participating in today's experiment you consent to that analysis of data gathered during your time here may be used in the broader study. Please read through this information and feel free to ask the experimenters any questions.

The purpose of the study is to see if the naturalness of telecommunication can be improved. We will be applying analysis to two games of Texas Holdem poker which you will play with two other participants.

Procedures

- You will be asked to provide some background information
- You will be given £10 which you are expected to stake in the game
- You will be asked to tick a box for consent and sign a separate receipt with your name on to legally confirm this (both will be stored seperately & securely digitally
- You have the right to withdraw from the experiment and have your data deleted completely after the experiment or at any time where you notify the experimenter in writing)
- You may be instructed in the use of an eye tracker which will be started
- You will be instructed in the use of a computer running internet poker
- You will play practise hands of poker without money for ten minutes with chip leader 'winning'. One of the experimental participants will guide you in this.
- You will then play multiple hands of poker with two other participants for money over a period of 40 minutes, overall chip leader being winner and taking £20 with second and third places each taking £5
- Afterwards there will be a short de-brief with the experimenter to assess how you felt about the experience and technology.
- The whole study is expected to take about 1.5 hours.
- Please consider that all of these questions are required questions unless advised otherwise or you have a question or objection which you are welcome to raise with the experimenter.

· You are free to withdraw from the study at any time and without giving reasons for withdrawing (by leaving the room).

*Required

Section 1 - Personal details

These elements are necessary for the statistical analysis but personal data is not retained against your name, and if you do not feel comfortable answering a question then either skip it or ask for help.

Your personal information will not be made public and your name will not be stored with the data. However it may be useful for us to publish a photograph or video footage of you during the experiment in support of the study findings. Do you consent to this? *

- ○ yes
- ○ no

Sex *

[▼]

Age

[▼]

Confirm you have normal (or normalised corrected) eyesight

Glasses and contact lenses are fine.

☐ I have normal eyesight or else my corrected eyesight is normal

Occupation

[ ]

Culture/country in which you spent your formative years (early life) *

This is important as there are different non verbal responses coded in early life

[ ]

You normally write with your

◯ Right hand

◯ Left hand

Rate how familiar you are with computers on the 7 point scale

	1	2	3	4	5	6	7	
Very unfamiliar, I do not use computers at all	◯	◯	◯	◯	◯	◯	◯	Very familar, I consider myself an expert

How familiar do you consider yourself with technology which allows communication over a distance

This question attempts to identify researchers who use these systems all the time

	1	2	3	4	5	6	7	
Unfamiliar, limited to phones	◯	◯	◯	◯	◯	◯	◯	Very familiar, research on immersive telepresence systems

How often do you play poker

[▼]

If you do play then how do you normally play poker

◯ Online

◯ In person

◯ A mix of the two

Are you aware of any unusual factors today which might impede or affect your performance or attitudes such as stress or lack of sleep?

☐ Yes

☐ No

I understand all the information above and give my consent to this experimental procedure *

☐ please check

Insertion of participant number here signifies that you have signed the form *

Participant number (experimenter fills this in

[ ]

Experimenter should fill in this question, please stop here. *
Start condition

- SV JOH
- MV JOH
- SV RCB
- MV RCB

Section 2 - Fill this section in after the experiment

Please read the question carefully, each one is slightly different.

I noticed the player on my right.

	1	2	3	4	5	6	7	
Not at all								Very

I noticed the player on my left

	1	2	3	4	5	6	7	
Not at all								Very

The player on my right noticed me

	1	2	3	4	5	6	7	
Not at all								Very

The player on my left noticed me

	1	2	3	4	5	6	7	
Not at all								Very

The player on my rights presence was obvious to me

	1	2	3	4	5	6	7	
Not at all								Very

The player on my lefts presence was obvious to me

	1	2	3	4	5	6	7	
Not at all								Very

My presence was obvious to the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My presence was obvious to the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right caught my attention

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left caught my attention

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was easily distracted from the player on my right when things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was easily distracted from the player on my left when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right was easily distracted from me when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left was easily distracted from me when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My thoughts were clear to the player on my right

	1	2	3	4	5	6	7

Not at all ○ ○ ○ ○ ○ ○ ○ Very

My thoughts were clear to the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The thoughts of the player on my right were clear to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The thoughts of the player on my left were clear to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It was easy to understand the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It was easy to understand the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right found it easy to understand me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left found it easy to understand me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Understanding the player on my right was difficult

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Understanding the player on my left was difficult

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my right had difficulty understanding me

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my left had difficulty understanding me

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I could tell how the player on my left player felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I could tell how the player on my right felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my right could tell how I felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my left could tell how I felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my rights emotions were not clear to me

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my lefts emotions were not clear to me

1 2 3 4 5 6 7

Not at all	○	○	○	○	○	○	○	Very

My emotions were not clear to the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My emotions were not clear to the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I could describe the player on my rights feelings accurately

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I could describe the player on my lefts feelings accurately

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right could describe my feelings accurately

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left could describe my feelings accurately

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was sometimes influenced by the player on my rights moods

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was sometimes influenced by the player on my lefts moods

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right was sometimes influenced by my moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my left was sometimes influenced by my moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my rights feelings influenced the mood of our interaction

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my lefts feelings influenced the mood of our interaction

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My feelings influenced the mood of interactions with the player on my right

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My feelings influenced the mood of interactions with the player on my left

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my rights attitudes influenced how I felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my lefts attitudes influenced how I felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My attitudes influenced how the player on my right felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My attitudes influenced how the player on my left felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Do you have any comments about the projected player on your left?

Section 3 - Regarding the player on my left

Machinelike / Humanlike

	1	2	3	4	5	6	7	
Machinelike	○	○	○	○	○	○	○	Humanlike

Artificial / Lifelike

	1	2	3	4	5	6	7	
Artificial	○	○	○	○	○	○	○	Lifelike

Fake / Natural

	1	2	3	4	5	6	7	
Fake	○	○	○	○	○	○	○	Natural

Uncouncious / Concious

	1	2	3	4	5	6	7	
Unconcious	○	○	○	○	○	○	○	Concious

Moving rigidly / Moving elegantly

1 2 3 4 5 6 7

Moving rigidly ◯ ◯ ◯ ◯ ◯ ◯ ◯ Moving elegantly

Awful / Nice

	1	2	3	4	5	6	7	
Awful	◯	◯	◯	◯	◯	◯	◯	Nice

Unpleasant / Pleasant

	1	2	3	4	5	6	7	
Unpleasant	◯	◯	◯	◯	◯	◯	◯	Pleasant

Dislike / Like

	1	2	3	4	5	6	7	
Dislike	◯	◯	◯	◯	◯	◯	◯	Like

Unfriendly / Friendly

	1	2	3	4	5	6	7	
Unfriendly	◯	◯	◯	◯	◯	◯	◯	Friendly

Kind / Unkind

	1	2	3	4	5	6	7	
Kind	◯	◯	◯	◯	◯	◯	◯	Unkind

Section 4

I felt that the person on my left was in the space with me *

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very much

I felt that the person on my right was in the space with me

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very much

I felt I was in the same room as two other people

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

I felt that I was looking at the person on my left through a screen

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

I felt that I was looking at the person on my right through a screen

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

It seemed to me that the person on my left occupied a different space to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

It seemed to me that the person on my right occupied a different space to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

When I looked at the person on my left I felt that I was looking into another room

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

When I looked at the person on my right I felt that I was looking into another room

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

I felt that the person on my left interacted naturally with the person on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

I felt that the person on my right interacted naturally with the person on my left

	1	2	3	4	5	6	7	

Not at all ○ ○ ○ ○ ○ ○ ○ Very much

During the game I felt that I was playing with two other people

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much

Please stop filling in the questionnaire here and alert the experimenter

There will now be more poker

Section 5

I noticed the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I noticed the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right noticed me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left noticed me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my rights presence was obvious to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my lefts presence was obvious to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My presence was obvious to the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My presence was obvious to the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right caught my attention

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left caught my attention

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was easily distracted from the player on my right when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I was easily distracted from the player on my left when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right was easily distracted from me when other things were going on

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left was easily distracted from me when other things were going on

	1	2	3	4	5	6	7

Not at all ○ ○ ○ ○ ○ ○ ○ Very

My thoughts were clear to the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My thoughts were clear to the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The thoughts of the player on my right were clear to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The thoughts of the player on my left were clear to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It was easy to understand the player on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It was easy to understand the player on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right found it easy to understand me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left found it easy to understand me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Understanding the player on my right was difficult

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Understanding the player on my left was difficult

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right had difficulty understanding me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left had difficulty understanding me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I could tell how the player on my right felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I could tell how the player on my left felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my right could tell how I felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my left could tell how I felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

The player on my rights emotions were not clear to me

1 2 3 4 5 6 7

Not at all ◯ ◯ ◯ ◯ ◯ ◯ ◯ Very

The player on my lefts emotions were not clear to me

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My emotions were not clear to the player on my right

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My emotions were not clear to the player on my left

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I could describe the player on my rights feelings accurately

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I could describe the player on my lefts feelings accurately

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my right could describe my feelings accurately

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my left could describe my feelings accurately

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I was sometimes influenced by the player on my rights moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

I was sometimes influenced by the player of my lefts moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my right was sometimes influenced by my moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my left was sometimes influenced by my moods

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on the rights feelings influenced the mood of our interaction

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my lefts feelings influenced the mood of our interaction

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My feelings influenced the mood of the interactions with the player on my right

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

My feelings influenced the mood of the interactions with the player on my left

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my rights attitudes influenced how I felt

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very

The player on my lefts attitudes influenced how I felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My attitudes influenced how the player on my right felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

My attitudes influenced how the player on my left felt

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Section 6 - Regarding the player on my left

Machineline / Humanlike

	1	2	3	4	5	6	7	
Machineline	○	○	○	○	○	○	○	Humanlike

Artificial / Lifelike

	1	2	3	4	5	6	7	
Artificial	○	○	○	○	○	○	○	Lifelike

Fake / Natural

	1	2	3	4	5	6	7	
Fake	○	○	○	○	○	○	○	Natural

Unconcious / Concious

	1	2	3	4	5	6	7	
Unconcious	○	○	○	○	○	○	○	Conscious

Moving rigidly / Moving elegantly

	1	2	3	4	5	6	7	
Moving rigidly	○	○	○	○	○	○	○	Moving elegantly

Awful / Nice

	1	2	3	4	5	6	7	
Awful	○	○	○	○	○	○	○	Nice

Unpleasant / pleasant

	1	2	3	4	5	6	7	
Unpleasant	○	○	○	○	○	○	○	Pleasant

Dislike / like

	1	2	3	4	5	6	7	
Dislike	○	○	○	○	○	○	○	Like

Unfriendly / friendly

	1	2	3	4	5	6	7	
Unfriendly	○	○	○	○	○	○	○	Friendly

Kind / unkind

	1	2	3	4	5	6	7	
Kind	○	○	○	○	○	○	○	Unkind

Section 7

I felt that the person on my right was in the space with me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I felt that the person on my left was in the space with me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I felt that I was in the same room as two other people

1 2 3 4 5 6 7

Not at all	○	○	○	○	○	○	○	Very

I felt that I was looking at the person on my right through a screen

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I felt that I was looking at the person on my left through a screen

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It seemed to me that the person on my right occupied a different space to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

It seemed to me that the person on my left occupied a different space to me

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

When I looked at the person on my right I felt I was looking into a different room

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

When I looked at the person on my left I felt I was looking into a different room

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I felt that I interacted naturally with the person on my right

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

I felt that I interacted naturally with the person on my left

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

During the game I felt I was playing with two other people

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very

Do you have any comments after this second session about the projected player on your left

Section 8 - Leave this for the experimenter

Eye tracked

- Yes
- No

FNIR

- Yes
- No

This box should be filled in by the experimenter

notes

Submit

Never submit passwords through Google Forms.

A.4 Software to record participant position for Exp 2

The tracking data was received into a demo version of MiddleVR software which ran a Unity programme capable of saving the Cartesian position data for the tracked 'player' (participant).

```
using UnityEngine;
using System.Collections;
using System.Xml;
using System.Xml.Serialization;
using System.IO;
using System.Text;

public class SaveAndLoad: MonoBehaviour {

    // An example where the encoding can be found is at
    // http://www.eggheadcafe.com/articles/system.xml.xmlserialization.asp
    // We will just use the KISS method and cheat a little and use
    // the examples from the web page since they are fully described

    // This is our local private members
    Rect _Start, _Stop, _StartMSG, _StopMSG;
    bool _ShouldSave, _ShouldLoad,_SwitchSave,_SwitchLoad;
    string _FileLocation,_FileName;
    public GameObject _Player;
    UserData myData;
    string _PlayerName;
    string _data;
    string _points;
    float _nextStoreTime;
    bool _isRecording;
    GameObject _headnode;

    // When the EGO is instansiated the Start will trigger
    // so we setup our initial values for our local members
    void Start () {
        // We setup our rectangles for our messages
        _Start=new Rect(10,80,100,20);
        _Stop=new Rect(10,100,100,20);
        _StartMSG=new Rect(10,120,400,40);
        _StopMSG=new Rect(10,140,400,40);

        // Where we want to save and load to and from
        _FileLocation = Application.dataPath;
        _FileName = "HeadTrackerData.xml";

        _PlayerName = "Head Tracker Data";

        // we need soemthing to store the information into
        myData = new UserData ();

        _isRecording = false;
    }

    void Update ()
    {
        if (_isRecording)
        {
            float interval = 0.2F;
            if (Time.time > _nextStoreTime) {
                // Wait till next timer
                _nextStoreTime = Time.time + interval;

                // Store the current point in the array
                Vector3 point = _headnode.transform.position;
                myData._iUser.points += "x=" + point.x + ",y=" + point.y + ",z=" +
                point.z + ";";
            }
        }
    }

    void OnGUI()
    {

        //***************************************************
        // Loading The Player...
        // **************************************************
        if (GUI.Button(_Start,"Start")) {

```

```
            _isRecording = true;
            _headnode = GameObject.Find("HeadNode");

            /*
            GUI.Label(_StopMSG,"Loading from: "+_FileLocation);
            // Load our UserData into myData
            LoadXML();
            if(_data.ToString() != "")
            {
                // notice how I use a reference to type (UserData) here, you need this
                // so that the returned object is converted into the correct type
                myData = (UserData)DeserializeObject(_data);
                // set the players position to the data we loaded
                VPosition=new Vector3(myData._iUser.x,myData._iUser.y,myData._iUser.z);

                _Player.transform.position=VPosition;
                // just a way to show that we loaded in ok
                Debug.Log(myData._iUser.name);
            }
            */
        }

        //*************************************************
        // Saving The Player...
        // ************************************************
        if (GUI.Button(_Stop, "Stop")) {
            _isRecording = false;

            GUI.Label(_StartMSG, "Saving to: "+_FileLocation);

            /*
            for (int index = 0; index < _points.Count; index++)
            {
                Vector3 point = Vector3(_points[index]);
                myData._iUser.points += "x=" + point.x + ",y=" + point.y + ",z=" + point.z + ";";
            }
            */
            myData._iUser.name = _PlayerName;

            // Time to creat our XML!
            _data = SerializeObject(myData);

            // This is the final resulting XML from the serialization process
            CreateXML();
            Debug.Log(_data);
        }


    }

    /* The following metods came from the referenced URL */
    string UTF8ByteArrayToString(byte[] characters)
    {
        UTF8Encoding encoding = new UTF8Encoding();
        string constructedString = encoding.GetString(characters);
        return (constructedString);
    }

    byte[] StringToUTF8ByteArray(string pXmlString)
    {
        UTF8Encoding encoding = new UTF8Encoding();
        byte[] byteArray = encoding.GetBytes(pXmlString);
        return byteArray;
    }

    // Here we serialize our UserData object of myData
    string SerializeObject(object pObject)
    {
        string XmlizedString = null;
        MemoryStream memoryStream = new MemoryStream();
```

```
        XmlSerializer xs = new XmlSerializer(typeof(UserData));
        XmlTextWriter xmlTextWriter = new XmlTextWriter(memoryStream, Encoding.UTF8);
        xs.Serialize(xmlTextWriter, pObject);
        memoryStream = (MemoryStream)xmlTextWriter.BaseStream;
        XmlizedString = UTF8ByteArrayToString(memoryStream.ToArray());
        return XmlizedString;
    }

    // Here we deserialize it back into its original form
    object DeserializeObject(string pXmlizedString)
    {
        XmlSerializer xs = new XmlSerializer(typeof(UserData));
        MemoryStream memoryStream = new MemoryStream(StringToUTF8ByteArray(pXmlizedString));
        XmlTextWriter xmlTextWriter = new XmlTextWriter(memoryStream, Encoding.UTF8);
        return xs.Deserialize(memoryStream);
    }

    // Finally our save and load methods for the file itself
    void CreateXML()
    {
        StreamWriter writer;
        FileInfo t = new FileInfo(_FileLocation+"\\"+ _FileName);
        if(!t.Exists)
        {
            writer = t.CreateText();
        }
        else
        {
            t.Delete();
            writer = t.CreateText();
        }
        writer.Write(_data);
        writer.Close();
        Debug.Log("File written.");
    }

    void LoadXML()
    {
        StreamReader r = File.OpenText(_FileLocation+"\\"+ _FileName);
        string _info = r.ReadToEnd();
        r.Close();
        _data=_info;
        Debug.Log("File Read");
    }
}

// UserData is our custom class that holds our defined objects we want to store in XML format
public class UserData
{
    // We have to define a default instance of the structure
    public DemoData _iUser;
    // Default constructor doesn't really do anything at the moment
    public UserData() { }

    // Anything we want to store in the XML file, we define it here
    public struct DemoData
    {
        public string points;
        //public float x;
        //public float y;
        //public float z;
        public string name;
    }
}
```

A.5 Actionscript code for shared poker task

```
/* *********************************************************************
AS3 Class for Flash CS4+ by Doug Ensley of http://www.flashandmath.com/
Last modified: November 20, 2010
******************************************************************** */

package flashandmath.as3.cards {
    import flash.display.Sprite;
    import flash.display.Bitmap;
    import flash.display.BitmapData;
    import flash.events.Event;

    import flashandmath.as3.cards.PlayingCard;

    public class CardDeck extends Sprite {

        private var arrCards:Vector.<PlayingCard>;

        public function CardDeck(arr:Vector.<PlayingCard>) {
            arrCards = new Vector.<PlayingCard>()
            arrCards = arr.concat();
            initialLayerCards();
        }

        private function initialLayerCards():void {
            var i:int;

            for (i=0; i<arrCards.length; i++) {
                arrCards[i].x = 0;
                arrCards[i].y = 0;
                arrCards[i].z = 0;
                this.addChildAt(arrCards[i],i);
            }
        }
        // Sets the depth level to match the array order for the cards in the deck. This allows us in the public methods that follow below to manipulate just the array of cards.
        private function layerCards():void {
            var i:int;

            while (this.numChildren > 0) {
                this.removeChildAt(0);
            }

            for (i=0; i<arrCards.length; i++) {
                this.addChildAt(arrCards[i],i);
            }
        }

        // getCardArray allows the user to gain access to all PlayingCard objects in this CardDeck object.
        public function getCardArray():Vector.<PlayingCard> {
            return arrCards.concat();
        }

        // getCardAt allows the user to gain access to one particular PlayingCard in this CardDeck object.
        public function getCardAt(i:int):PlayingCard {
            return PlayingCard(arrCards[i]);
        }

        // removeCardAt allows the user to remove (and return) a PlayingCard object from this CardDeck object.
        public function removeCardAt(i:int):PlayingCard {
            var pc:PlayingCard = arrCards.splice(i,1)[0];
            layerCards();
            return pc;
        }

        // addCardAt allows the user to add a PlayingCard object from this CardDeck object.
        public function addCardAt(pc:PlayingCard, i:int):void {
            arrCards.splice(i,0,pc);
            layerCards();
```

```
        }

        public function moveCard(from:int,to:int):void {
            var thisPC:PlayingCard = removeCardAt(from);
            addCardAt(thisPC,to);
        }

        public function reverseDeck():void {
            arrCards.reverse();
            layerCards();
        }

        // getCardIndex returns the index of a particular PlayingCard object in the
        CardDeck
        public function getCardIndex(pc:PlayingCard):int {
            return arrCards.indexOf(pc);
        }

        // Return the number of cards in this CardDeck
        public function get numCards():int {
            return arrCards.length;
        }
    }
}
/* *********************************************************************
AS3 Class for Flash CS4+ by Doug Ensley of http://www.flashandmath.com/
Last modified: November 20, 2010

ImageLoader class by Barbara Kaskosz of Flash and Math
******************************************************************** */

package flashandmath.as3.cards {
    import flash.display.Sprite;
    import flash.display.Bitmap;
    import flash.display.BitmapData;
    import flash.events.Event;

    //We will use our custom class, ImageLoader to load a list of bitmaps at runtime.
    import flashandmath.as3.ImageLoader;

    public class CardLoader extends Sprite {
        private var imgLoader:ImageLoader;
        private var arrCards:Array;

        public static const CARDS_LOADED:String = "imgsLoaded";
        public static const LOAD_ERROR:String = "loadError";

        public function CardLoader(arrImages:Array, stBackFile:String) {
            arrCards = new Array();
            imgLoader = new ImageLoader();
            imgLoader.addEventListener(ImageLoader.LOAD_ERROR,errorLoading);
            imgLoader.addEventListener(ImageLoader.IMGS_LOADED,allLoaded);
            imgLoader.loadImgs(arrImages.concat([ stBackFile ]));
        }

        private function errorLoading(e:Event):void {
            dispatchEvent(new Event(CardLoader.LOAD_ERROR));
        }

        private function allLoaded(e:Event):void {
            makeCards();
        }

        private function makeCards():void {
            var arrImages:Array = imgLoader.bitmapsArray;
            var n:int = arrImages.length - 1;
            var i:int;

            for (i=0; i<n; i++) {
                arrCards[i] = new PlayingCard(arrImages[i].bitmapData,arrImages[n].
                bitmapData);
            }

```

```
            for (i=0; i<arrCards.length; i++) {
                arrCards[i].x = 15*i;
                arrCards[i].y = 0;
                this.addChildAt(arrCards[i],i);
            }

            dispatchEvent(new Event(CardLoader.CARDS_LOADED));
        }

        // getCardArray allows the user to gain access to all PlayingCard objects in this CardDeck object.
        public function getCardArray():Array {
            return arrCards;
        }

        // getCardAt allows the user to gain access to one particular PlayingCard in this CardDeck object.
        public function getCardAt(i:int):PlayingCard {
            return PlayingCard(arrCards[i]);
        }

        // getCardIndex returns the index of a particular PlayingCard object in the CardDeck
        public function getCardIndex(pc:PlayingCard):int {
            return arrCards.indexOf(pc);
        }

        // Return the number of cards in this CardDeck
        public function numCards():int {
            return arrCards.length;
        }
    }
}

/* ***********************************************************************
ActionScript 3 Tutorial by Barbara Kaskosz

http://www.flashandmath.com/

Last modified: March 11, 2008
********************************************************************** */

package flashandmath.as3 {

    import flash.display.*;

    import flash.events.*;

    import flash.net.URLRequest;

    /*
    We are extending the EventDispatcher class contained in the flash.events
    package. Any instance of a subclass of EventDispatcher is capable
    of dispatching custom events. Many of AS3 built-in classes are subclasses
    of the EventDispatcher class, for example, the Sprite class and other
    DisplayObjects.
    */

    public  class ImageLoader extends EventDispatcher {

          /*
          We are defining constants corresponding to our two custom events.
          Similarly as for built-in events, later, when we add listeners
          to instances of ImageLoader, we can refer the events by the names
          of the constants, e.g. ImageLoader.IMGS_LOADED, or by their string value
          e.g. 'imgsLoaded'.
          */

          public static const IMGS_LOADED:String = "imgsLoaded";

          public static const LOAD_ERROR:String = "loadError";

          private var loadersArray:Array;
```

```
		private var numImgs:int;

		private var numLoaded:int;

		private var isError:Boolean;

		private var _bitmapsArray:Array;

		private var loadCanRun:Boolean;

		public function ImageLoader(){

		 //The constructor of the class sets the value of 'loadCanRun' variable only.
		 //It is the method 'loadImgs' below that performs all the main tasks.

		 this.loadCanRun=true;

	}

	 /*
	 'loadImgs' method takes an array of strings as a parameter. For the method
	  to function properly, the strings should represent addresses of the image files
	  to be loaded. The method listenes for IO loading errors. (For example,
	  the server is too busy and the file appears non-existent.) The method does not listen
	  to FlashPlayer security errors. We assume that the image files are at locations
	  that do not violate the security settings of the swf file that uses ImageLoader.
	 */

	public function loadImgs(imgsFiles:Array):void {

		if(loadCanRun){

		  loadCanRun=false;

		  //The conter variable counting how many images have been loaded.

		  numLoaded=0;

		  //The variable that remembers the current error status.

		  isError=false;

		  //The number of images to be loaded.

		  numImgs=imgsFiles.length;

		  //The array of bitmaps, each representing a loaded image.

		  _bitmapsArray=[];

		  /*
		  For each image file, we will use a separete instance of the Loader class.
		  That is because we will be loading images simultaneously rather than consecutively.
		  Consecutive loading is easier to code but it causes visible delays.
		  In the loop that follows, we populate the array of Loaders and attach
		  listeners to each Loader. One listens to an image finishing loading, the other
		  to an occurance of a loading error. Then we evoke the 'load' method for each Loader
		  with the address of the corresponding image.
		  */

		  loadersArray=[];

		 for(var i:int=0;i<numImgs;i++){

		    loadersArray[i]=new Loader();

		    loadersArray[i].contentLoaderInfo.addEventListener(Event.COMPLETE,
```

```
                    imgLoaded);

                    loadersArray[i].contentLoaderInfo.addEventListener(IOErrorEvent.IO_ERROR, errorOccured);

                    loadersArray[i].load(new URLRequest(imgsFiles[i]));

                }

            }

        }

        private function imgLoaded(e:Event):void {

            //When any of the images finishes loading, the count is increased by 1
            //and the function 'chackLoadStatus' is called. The function checks if all the images
            //have been loaded succesfully.

            numLoaded+=1;

            checkLoadStatus();

        }

        /*
        If a loading error occurs with any of the images, the function 'errorOccured' runs.
        The function dispatches one of our custom events: ImageLoader.LOAD_ERROR.
        Note the syntax when dispatching a custom event.
        */

        private function errorOccured(e:IOErrorEvent):void {

            isError=true;

            dispatchEvent(new Event(ImageLoader.LOAD_ERROR));

        }

        /*
        'checkLoadStatus' function runs each time an image is completely loaded.
        If the number of images loaded is equal to the total number of images to be
        loaded, the function dispatches the custom event: ImageLoader.ALL_LOADED.
        Then the function removes all the listeners and clears Loaders which we no longer
        need as the images have been stored by the function in _bitmapsArray.
        If all the images have been loaded successfully, loadCanRun is set to 'true'
        so the 'loadImgs' method can be called again for a different set of images.
        */

        private function checkLoadStatus():void {

            var i:int;

            if(numLoaded==numImgs && isError==false){

                 for(i=0;i<numImgs;i++){

                    _bitmapsArray[i]=Bitmap(loadersArray[i].content);

                 }

              for(i=0;i<numImgs;i++){

                    loadersArray[i].contentLoaderInfo.removeEventListener(IOErrorEvent.IO_ERROR, errorOccured);

                    loadersArray[i].contentLoaderInfo.removeEventListener(Event.COMPLETE, imgLoaded);

                    loadersArray[i]=null;

```

```
            }

            loadersArray=[];

            loadCanRun=true;

            dispatchEvent(new Event(ImageLoader.IMGS_LOADED));

        }

    }

    /*
    In order for 'bitmapsArray' to act as a public, read-only property, we define the getter
    method without defining the setter.
    */

    public function get bitmapsArray():Array {

        return _bitmapsArray;

    }


   }

}
/* ***********************************************************************
AS3 Class for Flash CS4+ by Doug Ensley of http://www.flashandmath.com/
Last modified: November 20, 2010
********************************************************************** */

package flashandmath.as3.cards {

    import flash.display.Sprite;
    import flash.display.Bitmap;
    import flash.display.BitmapData;
    import flash.events.Event;
    import flash.geom.Point;
    import flash.geom.PerspectiveProjection;


    import fl.transitions.Tween;
    import fl.transitions.TweenEvent;
    import fl.transitions.easing.*;
    import flash.geom.Vector3D;

    public class PlayingCard extends Sprite {
        public static const MOTION_DONE:String = "tweenMotionDone";
        public static const MOTION:String = "tweenMotion";

        private var bdFirst:BitmapData;
        private var bdSecond:BitmapData;

        private var _isFaceUp:Boolean;
        private var _value:String;
        private var _numValue:int;
        private var _suit:String;

        private var picWidth:Number;
        private var picHeight:Number;

        private var holder:Sprite;
        private var side0:Sprite;
        private var side1:Sprite;

        private var side0Img:Bitmap;
        private var side1Img:Bitmap;

        private var ptFrom:Vector3D;
        private var ptTo:Vector3D;
        private var objTween:Object;
```

```
        private var twMove:Tween;

        private var pp:PerspectiveProjection;

        /*
        The constructor of the PlayingCard class takes two BitmapData objects
        representing the images for the front and the back of the card to be used
        by the class. This will typically be called from the CardDeck class, which
        first loads all actual image files before constructing the individual cards.
        */

        public function PlayingCard(bmdFace:BitmapData,bmdBack:BitmapData){
            ptFrom = new Vector3D(0,0,0);
            ptTo = new Vector3D(0,0,0);

            objTween = {t: 0};
            twMove = new Tween(objTween, "t", None.easeIn, 0, 1, 1, true);
            twMove.stop();
            twMove.addEventListener(TweenEvent.MOTION_CHANGE, tweenMover);
            twMove.addEventListener(TweenEvent.MOTION_FINISH, tweenDone);

            side0Img=new Bitmap(bmdFace);
            side1Img=new Bitmap(bmdBack);

            picWidth=side0Img.width;
            picHeight=side0Img.height;

            holder=new Sprite();
            this.addChild(holder);

            holder.x=picWidth/2;
            holder.y=picHeight/2;

            side0=new Sprite();
            holder.addChild(side0);

            side0Img.x=-picWidth/2;
            side0Img.y=-picHeight/2;

            side0.x=0;
            side0.y=0;

            side0.addChild(side0Img);

            side1=new Sprite();

            holder.addChild(side1);

            side1Img.x=-picWidth/2;
            side1Img.y=-picHeight/2;

            side1.x=0;
            side1.y=0;

            side1.addChild(side1Img);

            //In order to appear correctly after a flip, the back side has to be
            //rotated initially.

            side1.rotationX = 180;
            _isFaceUp = true;

            // We have easy-to-access properties for value and suit so the card can be used in a game,
            // but these values will have to be set at runtime if the programmer wants to use them.
            _value = "";
            _numValue = 0;
            _suit = "";

            //Each instance of the PlayingCard class has its own PerspectiveProjection object.
            pp=new PerspectiveProjection();
```

```
        pp.fieldOfView=60;
        pp.projectionCenter=new Point(picWidth/2,picHeight/2);
        this.transform.perspectiveProjection=pp;

        rotateView(0,"horizontal");
    }

    //End of constructor.

    private function tweenMover(twe:TweenEvent):void {
        this.x = ptFrom.x + objTween.t * (ptTo.x - ptFrom.x);
        this.y = ptFrom.y + objTween.t * (ptTo.y - ptFrom.y);
        this.z = ptFrom.z + objTween.t * (ptTo.z - ptFrom.z);

        dispatchEvent(new Event(MOTION));
    }

    private function tweenDone(twe:TweenEvent):void {
        ptFrom = new Vector3D(0,0,0);
        ptTo = new Vector3D(0,0,0);
        dispatchEvent(new Event(MOTION_DONE));
    }


    // The tweenMotion method moves the card from coordinages (sx,sy,sz) to coordinates (fx,fy,fz)
    //  over the course of sec seconds.
    public function tweenMotion(sx:Number,sy:Number,sz:Number,fx:Number,fy:Number,fz:Number,sec:Number):void {
        ptFrom.x = sx;
        ptFrom.y = sy;
        ptFrom.z = sz;

        ptTo.x = fx;
        ptTo.y = fy;
        ptTo.z = fz;

        twMove.duration = sec;
        twMove.stop();
        twMove.rewind();
        twMove.start();
    }

    public function get isFaceUp():Boolean {
        return _isFaceUp;
    }

    public function get value():String {
        return _value;
    }

    public function set value(v:String):void {
        _value = v;
    }

    public function get numValue():int {
        return _numValue;
    }

    public function set numValue(v:int):void {
        _numValue = v;
    }

    public function get suit():String {
        return _suit;
    }

    public function set suit(s:String):void {
        _suit = s;
    }

    //The method switchSideUp flips the card immediately -- it is not an animated effect!
```

```
        public function switchSideUp():void {
            if (_isFaceUp) {
                makeFaceDown();
            }
            else {
                makeFaceUp();
            }
        }

        // The following methods, used above, are public so they can be called directly for greater control of the facing of the card.
        public function makeFaceUp():void {
            rotateView(0,"horizontal");
            _isFaceUp = true;
        }

        public function makeFaceDown():void {
            rotateView(180,"horizontal");
            _isFaceUp = false;
        }

        /* The rotateView method manages the rotation of the card and the correct visibility
        settings for the two card faces. This avoids depth swapping within the card itself.
        The value of t is the number of degrees of rotation, where t=0 means the card is face up.
        The spinType (default=vertical) specifies the axis of rotation, always the center of the card. */

        public function rotateView(t:Number,spinType:String="vertical"):void {
            var goodT:Number = t - 360*(Math.floor(t/360));

            if ( (goodT < 90) || (goodT > 270) ) {
                side0.visible = true;
                side1.visible = false;
            }
            else {
                side0.visible = false;
                side1.visible = true;
            }

            if(spinType=="vertical") {
                holder.rotationX = 0;
                holder.rotationY = goodT; }
            else {
                holder.rotationY = 0;
                holder.rotationX = goodT;
            }
        }
    }
}
```

A.6 Script for experiment 3

Prince2 Roles

Corporate Senior User Senior supplier Executive Project Manager

A major company is starting a new IT project. They need to update their payroll system to comply with new legislation before the end of the current financial year.

The company owner, referred to for this presentation as Corporate (indicate station 1) meets with their deputy, the Executive (indicate station 2) of the project to outline why change is required, what needs to happen, and what the tolerances for the project might be. They agree that they are interested in the project being delivered as a custom solution using their own Agile software department, who are represented by a software architect termed the Senior Supplier (indicate station 3) for the purposes of the project. They are familiar enough with Agile to know that it can rapidly deliver a minimum product Senior User (indicate station 3) fficient to meet the legislative demands, while being flexible enough to perhaps add much needed additional benefits to the payroll department, who are represented by a manager who takes the role of the Senior User (indicate station 4) for this project. Because theyre unsure of the specifics of translating their proposal for their current scrum masters they contract an external Project Manager (indicate station 5) for the duration of the project.

The Project Manager (indicate station 5) is a practitioner of PRINCE2, which he suggests be used for the project under an Axelos PRINCE2/Agile framework. The Senior User (indicate station 3) , Executive (indicate station 2) and Project Manager (indicate station 5) sit down to set out the initial project in a phase which is formally called project startup under PRINCE2 project management.. The Project Manager (indicate station 5) begins a daily log and formally records the names and roles of the Executive (indicate station 2) , the representative for payroll, the Senior User (indicate station 3), and the lead of the project delivery, the Senior supplier (indicate station 3).

Together they prepare the outline business case, discussing what will be done, how it will be done, and why each element of the case is necessary. The Senior Supplier (indicate station 3) has previous experience of implementing a similar system from a previous project in another company and is able to provide some lessons for capture by the Project Manager (indicate station 5), who records them in a lessons log which will be used throughout the project. As an essential part of the outline business case the Corporate (indicate station 1) explains how the project will be funded, with this going into the project brief owned by the Project Manager (indicate station 5) . The initial stage planning formulated by these senior stakeholders goes forward to the Corporate (indicate station 1) who gives formal approval to the project. The business case is refined according to the PRINCE2 guidelines, and the documentation is updated.

Corporate (indicate station 1) provides details of the corporate strategic planning to the executive and the Project Manager (indicate station 5). This allows a project plan to be designed. This project planning includes detailed planning for the first formal management stage of the project. Crucially it also includes an exception plan, that details in what circumstances the project manager must alert the executive

A.7 Experiment 3 full results

A.8 Final scaled system questions

Telethrone small group experiment

By participating in today's experiment you consent to that analysis of data gathered during your time here may be used in the broader study. Please read through this information and feel free to ask the experimenters any questions.

The purpose of the study is to see if the naturalness of telecommunication can be improved. We will be applying analysis to your responses to the questions which you answer after a short presentation

Procedures

- You will be asked to provide some background information
- You will be asked to tick a box for consent
- You will be instructed in what you will see during the presentation
- You will watch a short presentation (6 minutes)
- Afterwards there will be a short de-brief with the experimenter to assess how you felt about the experience and technology.
- The whole study is expected to take about 15 minutes.
- Please consider that all of these questions are required questions unless advised otherwise or you have a question or objection which you are welcome to raise with the experimenter.
- You have the right to withdraw from the experiment at any time, and have your data deleted completely after the experiment or at any time where you notify the experimenter in writing). You do not need to provide reasons for withdrawing and may simply leave the room.

*Required

1. **Participant Number (experimenter fills this)** *

____

Section 2 - Personal Details

These elements are necessary for the statistical analysis but personal data is not retained against your name, and if you do not feel comfortable answering a question then either skip it or ask for help.

2. **Your personal information will not be made public and your name will not be stored with the data. However it may be useful for us to publish a photograph or video footage of you during the experiment in support of the study findings. Do you consent to this?** *

Mark only one oval.

- ( ) Yes
- ( ) No

3. **Sex** *

Mark only one oval.

- ( ) Female
- ( ) Male
- ( ) Prefer not to say
- ( ) Other: ____________________

4. **Age** *

____

5. **Confirm you have normal (or normalised corrected) eyesight** *

Mark only one oval.

- ( ) Yes I have normal eyesight or else my corrected eyesight is normal

6. **Occupation**

____

7. **Culture/country in which you spent your formative years (early life)**

____

8. **Rate how familiar you are with computers on the 7 point scale**

Mark only one oval.

	1	2	3	4	5	6	7	
Very unfamilar, I do not use computers at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very familiar, I consider myself an expert

9. **How familiar do you consider yourself with technology which allows communication over a distance on the 7 point scale**

Mark only one oval.

	1	2	3	4	5	6	7	
Unfamiliar, limited to phones	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very familiar, research on such systems

10. **Stop here, signal the experimenter. You will be assigned a number between 1 and 5. Find that labelled chair and sit on it when told to do so.** *

Mark only one oval.

- ⬭ Station 1
- ⬭ Station 2
- ⬭ Station 3
- ⬭ Station 4
- ⬭ Station 5

How did I feel about the representation of a person in the projected chair / screen.

Do your best to rate these answers somewhere between one word and the other word.

11. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Fake	⬭	⬭	⬭	⬭	⬭	Natural

12. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Machinelike	⬭	⬭	⬭	⬭	⬭	Humanlike

13. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Unconscious	⬭	⬭	⬭	⬭	⬭	Conscious

14. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Artificial	⬭	⬭	⬭	⬭	⬭	Lifelike

15. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Moving rigidly	◯	◯	◯	◯	◯	Moving elegantly

16. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Dead	◯	◯	◯	◯	◯	Alive

17. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Stagnant	◯	◯	◯	◯	◯	Lively

18. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Inert	◯	◯	◯	◯	◯	Interactive

19. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Apathetic	◯	◯	◯	◯	◯	Responsive

20. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Dislike	◯	◯	◯	◯	◯	Like

21. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Unfriendly	◯	◯	◯	◯	◯	Friendly

22. **I felt that the representation on the chair was** *
Mark only one oval.

	1	2	3	4	5	
Unkind	◯	◯	◯	◯	◯	Kind

23. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Unpleasant	◯	◯	◯	◯	◯	Pleasant

24. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Awful	◯	◯	◯	◯	◯	Nice

Who did I most think I was when I was pointed at and referred to.

Do the best you can to score this section so that it represents who you most think you were in the meeting. If you are unsure then you can indicate multiple roles for yourself.

25. **I felt that I represented the Corporate (head of the company) role in the room**

Mark only one oval.

- I strongly disagree, I was not corporate
- I disagree, I was not corporate
- Neutral
- Agree, I think I was corporate
- Strongly agree, I was corporate

26. **I felt that I represented the Executive (company deputy) role in the room**

Mark only one oval.

- I strongly disagree, I was not executive
- I disagree, I was not executive
- Neutral
- Agree, I think I was executive
- Strongly agree, I was executive

27. **I felt that I represented the Senior Supplier (software architect) role in the room**

Mark only one oval.

- I strongly disagree, I was not the senior supplier
- I disagree, I was not the senior supplier
- Neutral
- Agree, I think I was the senior supplier
- Strongly agree, I was the senior supplier

28. **I felt that I represented the Senior User (payroll representative) role in the room**

Mark only one oval.

- I strongly disagree, I was not the senior user
- I disagree, I was not the senior user
- Neutral
- Agree, I think I was the senior user
- Strongly agree, I was the senior user

29. **I felt that I represented the Project Manager (external consultant) role in the room**

Mark only one oval.

- I strongly disagree, I was not the project manager
- I disagree, I was not the project manager
- Neutral
- Agree, I think I was the project manager
- Strongly agree, I was the project manager

Table A.1: A table which contains all the results before they are merged

Condition / run / station	**Thought they were station #**
Multiview condition run 1 station 1	1
Multiview condition run 1 station 2	2
Multiview condition run 1 station 3	3
Multiview condition run 1 station 4	4
Multiview condition run 1 station 5	5
Multiview condition run 2 station 1	1
Multiview condition run 2 station 2	2
Multiview condition run 2 station 3	3
Multiview condition run 2 station 4	4
Multiview condition run 2 station 5	4 & 5
Singleview condition run 1 station 1	1
Singleview condition run 1 station 2	3
Singleview condition run 1 station 3	3
Singleview condition run 1 station 4	3
Singleview condition run 1 station 5	4
Singleview condition run 2 station 1	1
Singleview condition run 2 station 2	3
Singleview condition run 2 station 3	3
Singleview condition run 2 station 4	3
Singleview condition run 2 station 5	5

A.9 GODSPEED Indices

The Godspeed Questionnaire Series – Christoph Bartneck, Ph.D.

The web pages proposes a series of questionnaires to measure the users' perception of robots. This series shall be called "Godspeed" because it is intended to help creators of robots on their development journey. Below you find the English, Spanish, Dutch,Japanese, and Chinese version. Please email me your translations into other languages. The Spanish translation was contributed by Javier Ruiz-del-Solar. The Dutch and Chinese translation was contributed by Bram Vanderborght. The German translation was contributed by Mary Ellen Foster and Manuel Giuliani. The Arabic translation was contributed by Micheline Ziadee. Alexander Astaras translated the questionnaire to Greek. Wafa Johal provided the French translation. Carina Dantas provided the Portuguese translation.

This questionnaire has been discussed in detailed in this publication:

Bartneck, C., Croft, E., Kulic, D. & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1) 71-81. | DOI: 10.1007/s12369-008-0001-3

Godspeed I: Anthropomorphism, Antropomorfismo, Anthropomorphismus, Ανθρωπομορφισμός, Anthropomorphisme, Antropomorfismo

Παρακαλούμε βαθμολογείστε την εντύπωση που σχηματίσατε για το ρομπότ στις παρακάτω κλίμακες:
Bitte beurteilen sie Ihren Eindruck des Roboters auf diesen Skalen:
Please rate your impression of the robot on these scales:
Por favor de su opinión del robot en los siguientes aspectos:
Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:
以下のスケールに基づいてこのロボットの印象を評価してください。
人格化
من فضلك حدد\حددي انطباعك عن الروبوت على المقاييس التالية:
Veuillez noter vos impressions au sujet du robot sur les échelles ci-dessous:
Por favor, avalie a sua impressão sobre as caraterísticas humanas do robô nas seguintes escalas:

Ψεύτικο Unecht Fake Falso Onecht 偽物のような 虚假的 مزيف Faux Falso	1	2	3	4	5	Φυσικό Natürlich Natural Natural Natuurlijk 自然な 自然的 طبيعي Naturel Natural
Μηχανόμορφο Wie eine Maschine						Ανθρωπόμορφο Wie ein Mensch

Machinelike Con Aspecto de Máquina Lijkend op een machine 機械的 似机械的 شبيه بالالات D’aspect Machinale Com aspeto mecânico	1	2	3	4	5	Humanlike Con Aspecto Humano Lijkend op een mens 人間的 似人类的 شبيه بالانسان D’aspect Humain Com aspeto humano
Χωρίς συνείδηση Hat kein Bewusstsein Unconscious Inconsciente Onbewust 意識を持たない 无意识的 غير مدرك Inconscient Inconsciente	1	2	3	4	5	Με συνείδηση Hat ein Bewusstsein Conscious Consciente Heeft een bewustzijn 意識を持っている 有意识的 مدرك Conscient Consciente
Τεχνητό Künstlich Artificial Artificial Kunstmatig 人工的 人工的 اصطناعي Artificiel Artificial	1	2	3	4	5	Ζωντανό Realistisch Lifelike Parece Vivo Levensecht 生物的 逼真的 يبدو حقيقي Paraissant vivant Realista
Άκομψη κίνηση Bewegt sich steif Moving rigidly Se Mueve Rígidamente Houterige bewegingen ぎこちない動き 动作僵硬 حركه متصلبه Bougeant avec rigidité Move-se com rigidez	1	2	3	4	5	Φυσική κίνηση Bewegt sich flüssig Moving elegantly Se Mueve con Fluidez Vloeiende bewegingen 洗練された動き 动作流畅 حركه انيقه Bougeant avec fluidité Move-se com fluidez

Godspeed II: Animacy, Animacidad, Belebtheit, Κινητικότητα, Animation, Expressão de vida

Παρακαλούμε βαθμολογείστε την εντύπωση που σχηματίσατε για το ρομπότ στις παρακάτω κλίμακες:

Bitte beurteilen sie Ihren Eindruck des Roboters auf diesen Skalen:

Please rate your impression of the robot on these scales:

Por favor de su opinión del robot en los siguientes aspectos:

Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:

以下のスケールに基づいてこのロボットの印象を評価してください。

生命性

من فضلك حدد\حددي انطباعك عن الروبوت على المقاييس التالية:

Veuillez noter vos impressions au sujet du robot sur les échelles ci-dessous:

Por favor, avalie a sua impressão sobre a expressão de vida do robô nas seguintes escalas:

Άψυχο						Ζωντανό
Tot						Lebendig
Dead						Alive
Muerto						Con Vida
Dood	1	2	3	4	5	Levend
死んでいる						生きている
死的						活的
ميت						حي
Mort						Vivant
Morto						Com vida
Στάσιμο						Ευκίνητο
Unbewegt						Lebendig
Stagnant						Lively
Inactivo						Vivaz
Stilstaand	1	2	3	4	5	Levendig
活気のない						生き生きとした
静止的						活泼的
ساكن						حيوي
Inanimé						Animé
Parado						Enérgico
Μηχανικό						Βιολογικό
Mechanisch						Organisch
Mechanical						Organic
Mecánico						Orgánico
Mechanisch	1	2	3	4	5	Organisch
機械的な						有機的な
机械的						有机的
الي						عضوي
Mécanique						Organique
Mecânico						Orgânico
Τεχνητό						Μοιάζει ζωντανό
Künstlich						Realistisch
Artificial						Lifelike

Artificial						Parece Vivo
Kunstmatig	1	2	3	4	5	Levensecht
人工的な						生物的な
人工的						逼真的
اصطناعي						يبدو حقيقي
Artificiel						Réaliste
Artificial						Realista

Αδρανές						Διαδραστικό
Träge						Interaktiv
Inert						Interactive
Estático						Interactivo
Passief	1	2	3	4	5	Interactief
不活発な						対話的な
迟钝的						互动的
جامد						متفاعل
Inert						Interactif
Estático						Interativo

Απαθές						Ανταποκρίνεται
Apathisch						Reagierend
Apathetic						Responsive
Indiferente						Atento
Apatisch	1	2	3	4	5	Responsief
無関心な						反応のある
冷淡的						反应迅速的
لا مبالي						متجاوب
Apathique						Attentif
Apático						Participativo

Godspeed III: Likeability, Simpatía, Συμπαθητικότητα, Appréciation, Simpatia

Παρακαλούμε βαθμολογείστε την εντύπωση που σχηματίσατε για το ρομπότ στις παρακάτω κλίμακες:
Bitte beurteilen sie Ihren Eindruck des Roboters auf diesen Skalen:
Please rate your impression of the robot on these scales:
Por favor de su opinión del robot en los siguientes aspectos:
Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:
以下のスケールに基づいてこのロボットの印象を評価してください。
可爱度
من فضلك حدد\حددي انطباعك عن الروبوت على المقاييس التالية:
Veuillez noter vos impressions au sujet du robot sur les échelles ci-dessous:
Por favor, avalie a sua impressão sobre a simpatia do robô nas seguintes escalas:

Δε μου αρέσει						Μου αρέσει
Nicht mögen						Mögen
Dislike						Like

Disgusta						Gusta
Afkeer	1	2	3	4	5	Geliefd
嫌い						好き
不喜欢						喜欢
لا احبه						احبه
Déplaisant						Plaisant
Não gosto						Gosto
Μη φιλικό						Φιλικό
Unfreundlich						Freundlich
Unfriendly						Friendly
No Amigable						Amigable
Onvriendelijk	1	2	3	4	5	Vriendelijk
親しみにくい						親しみやすい
不友好的						友好的
غير ودود						ودود
Inamical						Amical
Hostil						Amigável
Αγενές						Ευγενές
Unhöflich						Höflich
Unkind						Kind
Descortés						Amable
Niet lief	1	2	3	4	5	Lief
不親切な						親切な
不亲切的						亲切的
غير طيب						طيب
Malaimable						Aimable
Antipático						Gentil
Δυσάρεστο						Ευχάριστο
Unangenehm						Angenehm
Unpleasant						Pleasant
Desagradable						Agradable
Onplezierig	1	2	3	4	5	Plezierig
不愉快な						愉快な
不愉快的						愉快的
غير ممتع						ممتع
Désagreable						Agréable
Desagradável						Agradável
Απαίσιο						Συμπαθές
Furchtbar						Nett
Awful						Nice
Feo						Lindo
Afschuwelijk	1	2	3	4	5	Mooi
ひどい						良い

恶劣的 مقيت Horrible Horrível						良好的 لطيف Gentil Simpático

Godspeed IV: Perceived Intelligence, Inteligencia Percibida, Εκλαμβανόμενη Ευφυΐα, Intelligence Perçue, Inteligência Percebida

Παρακαλούμε βαθμολογείστε την εντύπωση που σχηματίσατε για το ρομπότ στις παρακάτω κλίμακες:
Bitte beurteilen sie Ihren Eindruck des Roboters auf diesen Skalen:
Please rate your impression of the robot on these scales:
Por favor de su opinión del robot en los siguientes aspectos:
Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:
以下のスケールに基づいてこのロボットの印象を評価してください。
感知能力
من فضلك حدد\حددي انطباعك عن الروبوت على المقاييس التالية:
Veuillez noter vos impressions au sujet du robot sur les échelles ci-dessous:
Por favor, avalie a sua impressão sobre a inteligência percebida do robô nas seguintes escalas:

Ανίκανο Inkompetent Incompetent Incompetente Onbekwaam 無能な 无能力的 غير كفوء Uncompétent Incompetente	1	2	3	4	5	Ικανό Kompetent Competent Competente Bekwaam 有能な 有能力的 كفوء Compétent Competente
Αδαές Ungebildet Ignorant Ignorante Onwetend 無知な 无知识的 جاهل Ignorant Ignorante	1	2	3	4	5	Καταρτισμένο Unterrichtet Knowledgeable Culto Veel wetend 物知りな 有知识的 مطلع Cultivé Sabedor
Ανεύθυνο Verantwortungslos Irresponsible Irresponsable Onverantwoordelijk	1	2	3	4	5	Υπεύθυνο Verantwortungsbewusst Responsible Responsable Verantwoordelijk

無責任な 无责任的 غيرمسؤول Irresponsable Irresponsável						責任のある 有责任的 مسؤول Responsable Responsável
Κουτό Unintelligent Unintelligent Sin inteligencia Onintelligent 知的でない, 无智力的 غير ذكي Inintelligent Pouco inteligente	1	2	3	4	5	Έξυπνο Intelligent Intelligent Inteligente Intelligent 知的な 有智力的 ذكي Intelligent Inteligente
Ανόητο Unvernünftig Foolish Insensato Dwaas 愚かな 笨拙的 غير عاقل Insensé Insensato	1	2	3	4	5	Λογικό Vernünftig Sensible Juicioso Gevoelig 賢明な 敏感的 عاقل Sensé Sensato

Godspeed V: Perceived Safety, Seguridad Percibida, Sicherheit, Εκλαμβανόμενη Ασφάλεια, Sureté perçue, Segurança Percebida

Παρακαλούμε βαθμολογείστε τη δική σας συναισθηματική κατάσταση στις παρακάτω κλίμακες:
Bitte bewerten Sie Ihren emotionalen Zustand auf diesen Skalen:
Please rate your emotional state on these scales:
¿Cómo se sintió durante su interacción con el robot?
Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:
以下のスケールに基づいてあなたの心の状態を評価してください。
安全性
من فضلك حدد\حددي انطباعك عن الروبوت على المقاييس التالية:
Veuillez noter votre état émotionel sur les échelles ci-dessous:
Por favor, avalie o seu estado emocional sobre a segurança percebida do robô nas seguintes escalas:

Ανήσυχος Ängstlich Anxious Ansioso						Χαλαρός Entspannt Relaxed Relajado

Angstig 不安な 不安的 قلق Anxieux Ansioso	1	2	3	4	5	Ontspannen 落ち着いた 放松的 مسترخي Relaxé Descontraído
Ήρεμος Ruhig Calm Tranquilo Kalm 冷静な 冷静的 منفعل Calme Calmo	1	2	3	4	5	Ταραγμένος Aufgewühlt Agitated Agitado Opgewonden 動揺している 焦躁的 مطمئن Agité Agitado
Ήσυχος Still Quiescent No Sorprendido Rustig 平穏な 沉默的 هادئ Serein Sereno	1	2	3	4	5	Ξαφνιασμένος Überrascht Surprised Sorprendido Verrast 驚いた 诧异的 متفاجئ Surpris Surprendido

A.10 Networked Minds

Networked Minds Social Presence Inventory:
(Scales only, Version 1.2)
Measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry

Frank Biocca, Chad Harms

This document includes all the items that comprise the Networked Minds Social Presence Inventory. For more information on the scale, please consult the Guide to the Networked Minds Inventory and other related papers at http://www.mindlab.msu.edu/networkedminds/.

1. First order social presence: Co-presence

The following items form the measure of co-presence, the degree to which the users feel as if they are together in the same space.

Perception of self	Perception of the other
I often felt as if (my partner) and I were in the same (room) together.	I think (my partner) often felt as if we were in the same room together.
I was often aware of (my partner) in the (room).	(My partner) was often aware of me in the (room).
I hardly noticed (my partner) in the (room)	(My partner) didn't notice me in the (room).
I often felt as if we were in different places rather than together in same (room)	I think (my partner) often felt as if we were in different places rather than together in the same (room).

2. Second order social presence: Psycho-behavioral interaction

These items seek to measure the user perception of attention, emotional contagion, and mutual understanding with their partner or participant.

Perceived psychological engagement

Perception of self	Perception of the other
Perceived attentional engagement	
I paid close attention to (my partner).	(My partner) paid close attention to me
I was easily distracted from (my partner) when other things were going on.	(My partner) was easily distracted from me when other things were going on.
I tended to ignore (my partner).	(My partner) tended to ignore me.
Perceived emotional contagion	
I was sometimes influenced by (my partner's) moods.	(My partner) was sometimes influenced by my moods.
When I was happy, (my partner) tended to be happy.	When (my partner) was happy, I tended to be happy.
When I was feeling sad (my partner) also seemed to be down.	When (my partner) was feeling sad, (my partner) I tended to be sad.
When I was feeling nervous, (my partner) also seemed to be nervous.	When (my partner) was nervous, (my partner) I tended to be nervous.
Perceived comprehension	
I was able to communicate my intentions clearly to (my	(My partner) was able to communicate their intentions

partner.)	clearly to me.
My thoughts were clear to (my partner).	(My partner's) thoughts were clear to me.
I was able to understand what (my partner) meant.	(My partner) was able to understand what I meant.

Perceived behavioral interdependence

Perception of self	Perception of my partner
My actions were often dependent on (my partner's) actions.	(My partner's) actions were often dependent on my actions.
My behavior was often in direct response to (my partner's) behavior.	The behavior of (my partner) was often in direct response to my behavior.
What I did often affected what (my partner) did.	What (my partner) did often affected what I did.

3. Third order social presence: Subjective and Intersubjective Symmetry

Third order social presence is derived from the scales used for first order and second order social presence.

Subjective Symmetry: Analysis

Subjective symmetry is a measure of the degree to which the user perceives their level of social presence to be symmetrical or correlated with that of their partner's. It is the calculated as a correlation between the ratings of the social presence of the self ("Perception of self") and the other ("Perception of my partner"). This can be calculated for each scale or for the inventory as a whole.

Intersubjective Symmetry: Analysis

Intersubjective symmetry is a measure of the degree to which the user's rating of their social presence is symmetrical (correlated) with their partner's rating of the user's level of social presence. It is the calculated as a correlation between the ratings of the social presence of the self ("Perception of self") and the other's rating of the user (i.e., the partner's "Perception of my partner"). The intersubjective symmetry can be calculated for each self-partner pairing. This can be calculated for each scale or for the inventory as a whole.

4. Notes on the use and analysis of the scales

Contextualizing the scale with substitutions for "my partner" and "the room" terms

To prevent confusion, difficulty, and to make questions clear it is preferable that users respond to concrete references rather than abstract references. Therefore the words in brackets indicate phrases to be substituted for appropriate terms relevant to the mediated interaction.

Example substitutions for the ***agent term*** holder,
"my partner" = the other, participants, robot, player, opponent, guest, "Alicia," etc.

Example substitutions for the ***place term*** holder,
"the room" = office, city, screen, virtual environment, etc.

A.11 Presented at ICAT-EGVE in Little Rock USA

Call for Papers ————— ICAT-EGVE 2016 is the merger of the 26th International Conference on Artificial Reality and Telexistence (ICAT 2016) and the 21st Eurographics Symposium on Virtual Environments (EGVE 2016). ICAT-EGVE 2016 will be held in Little Rock, Arkansas, USA, from December 7th to 9th 2016.

We are fortunate to have been able to secure an exceptional venue for this event: the conference will be held at the Clinton Presidential Library.

This international event will be a unique opportunity for researchers, developers, and users to share their experience and knowledge of virtual reality, as well as augmented reality, mixed reality and 3D user interfaces. And, of course, it is a good time to renew old friendships, make new ones, and experience all that Little Rock has to offer.

ICAT-EGVE 2016 seeks original, high-quality research papers in all areas of virtual reality, as well as augmented reality, mixed reality and 3D user interfaces. Research papers should describe results that contribute to advancements in the following areas:

- 3D interaction for VR/AR - VR/AR systems and toolkits - Immersive projection technologies and other advanced display technologies - Presence, cognition, and embodiment in VR/AR/MR - Haptics, audio, and other non-visual modalities - User studies and evaluation - Multi-user and distributed VR, tele-immersion and tele-presence - Serious games and edutainment using VR/AR/MR - Novel devices (both input and output) for VR, AR, MR, and haptics - Applications of VR/AR/MR

Papers in other related areas are welcome, too, of course.

All accepted papers will be published in the Digital Library of the Eurographics. In addition, they will be cited and indexed in the ACM Digital Library including the DOI.

A.12 Presented at International AV VR

Call for Papers

International AR and VR Conference 2017: Empowering Human, Place and Business through AR & VR

Manchester, 23rd February 2017

Host:

Creative Augmented and Virtual Reality Hub, School of Tourism, Events & Hospitality Management Faculty of Business and Law, Manchester Metropolitan University

Augmented and Virtual Reality (AR & VR) offer exciting opportunities for human computer interaction, the enhancement of places and new business cases. The Augmented Reality and Virtual Reality Conference organisers seek original, high-quality papers in all areas related to augmented reality (AR), virtual reality (VR), mixed reality and 3D user interfaces.

With a theme Empowering Human, Place and Business through AR & VR, International AR & VR Conference 2017 will focus on exploring cutting edge Augmented and Virtual Reality concepts, applications and business models which shape our everyday life, place and businesses.

Issues to be covered at the conference include, but are not limited to, the following areas:

Augmented and virtual reality adoption behavior Augmented and virtual reality business model Augmented and virtual reality applications in tourism, events, hospitality, retail, cultural heritage, architecture, education, entertainment, health, media etc. Development and implementation of augmented and virtual reality applications Mobile and wearable technologies and applications Impacts of augmented and virtual reality on value co-creation Augmented and virtual reality for customer engagement Enhancing customer experience via augmented and virtual reality Multi-sensory experience via augmented and virtual reality Cultural differences in the use of augmented and virtual reality Legal, ethical, and regulatory issues of augmented and virtual reality Augmented and virtual reality marketing Augmented and virtual reality gamification Augmented and virtual reality in smart city & smart tourism Any other related topic

Telethrone small group experiment

By participating in today's experiment you consent to that analysis of data gathered during your time here may be used in the broader study. Please read through this information and feel free to ask the experimenters any questions.

The purpose of the study is to see if the naturalness of telecommunication can be improved. We will be applying analysis to your responses to the questions which you answer after a short presentation

Procedures

- You will be asked to provide some background information
- You will be asked to tick a box for consent
- You will be instructed in what you will see during the presentation
- You will watch a short presentation (6 minutes)
- Afterwards there will be a short de-brief with the experimenter to assess how you felt about the experience and technology.
- The whole study is expected to take about 15 minutes.
- Please consider that all of these questions are required questions unless advised otherwise or you have a question or objection which you are welcome to raise with the experimenter.
- You have the right to withdraw from the experiment at any time, and have your data deleted completely after the experiment or at any time where you notify the experimenter in writing). You do not need to provide reasons for withdrawing and may simply leave the room.

*Required

Section 2 - Personal Details

These elements are necessary for the statistical analysis but personal data is not retained against your name, and if you do not feel comfortable answering a question then either skip it or ask for help.

1. **Your personal information will not be made public and your name will not be stored with the data. However it may be useful for us to publish a photograph or video footage of you during the experiment in support of the study findings. Do you consent to this?** *

Mark only one oval.

- Yes
- No

2. **Sex** *

Mark only one oval.

- Female
- Male
- Prefer not to say
- Other: ____________________

3. **Age** *

____

4. **Confirm you have normal (or normalised corrected) eyesight** *

Mark only one oval.

- Yes I have normal eyesight or else my corrected eyesight is normal

5. **Occupation**

____

6. **Culture/country in which you spent your formative years (early life)**

____

7. **Rate how familiar you are with computers on the 7 point scale**
Mark only one oval.

	1	2	3	4	5	6	7	
Very unfamilar, I do not use computers at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very familiar, I consider myself an expert

8. **How familiar do you consider yourself with technology which allows communication over a distance on the 7 point scale**
Mark only one oval.

	1	2	3	4	5	6	7	
Unfamiliar, limited to phones	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very familiar, research on such systems

How did I feel about the representation of a person in the chair.

Do your best to rate these answers somewhere between one word and the other word.

9. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Fake	⬭	⬭	⬭	⬭	⬭	Natural

10. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Machinelike	◯	◯	◯	◯	◯	Humanlike

11. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Unconscious	◯	◯	◯	◯	◯	Conscious

12. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Artificial	◯	◯	◯	◯	◯	Lifelike

13. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Moving rigidly	◯	◯	◯	◯	◯	Moving elegantly

14. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Dead	◯	◯	◯	◯	◯	Alive

15. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Stagnant	⬭	⬭	⬭	⬭	⬭	Lively

16. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Inert	⬭	⬭	⬭	⬭	⬭	Interactive

17. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Apathetic	⬭	⬭	⬭	⬭	⬭	Responsive

18. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Dislike	⬭	⬭	⬭	⬭	⬭	Like

19. **I felt that the representation on the chair was**
Mark only one oval.

	1	2	3	4	5	
Unfriendly	⬭	⬭	⬭	⬭	⬭	Friendly

20. **I felt that the representation on the chair was** *

Mark only one oval.

	1	2	3	4	5	
Unkind	( )	( )	( )	( )	( )	Kind

21. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Unpleasant	( )	( )	( )	( )	( )	Pleasant

22. **I felt that the representation on the chair was**

Mark only one oval.

	1	2	3	4	5	
Awful	( )	( )	( )	( )	( )	Nice

Who did I most think I was when I was pointed and and referred to.

Do the best you can to score this section so that it represents who you most think you were in the meeting. If you are unsure then you can indicate multiple roles for yourself.

23. **I felt that I represented the Corporate (head of the company) role in the room**

Mark only one oval.

- ( ) Strongly disagree
- ( ) Disagree
- ( ) Neutral
- ( ) Agree
- ( ) Strongly agree

24. **I felt that I represented the Executive (company deputy) role in the room**

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

25. **I felt that I represented the Senior Supplier (software architect) role in the room**

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

26. **I felt that I represented the Senior User (payroll representative) role in the room**

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

27. **I felt that I represented the Project Manager (external consultant) role in the room**

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

A.13 R analysis

```r
library(ggplot2)
library(reshape2)

setwd("c:/rwork")

partialCoPresence <- read.csv(file="partialCoPresence.csv",stringsAsFactor=FALSE)
partialCoPresence.m <- melt(partialCoPresence, id.var = "View")
p2 <- ggplot(data = partialCoPresence.m, aes(x=variable, y=value)) +
        geom_boxplot(aes(fill=View)) +
        scale_y_continuous(name = "Likhart not very - very much",
                           breaks = seq(0, 7, 1),
                           limits=c(0, 7)) +
        scale_x_discrete(name = "Partial networked minds co-presence") +
        ggtitle("Responses comparing single view and multi view") +
        theme_bw() +
        theme(plot.title = element_text(size = 14, family = "Tahoma", face = "bold"),
              text = element_text(size = 12, family = "Tahoma"),
              axis.title = element_text(face="bold"),
              axis.text.x=element_text(size = 11, angle = 90, hjust = 1),
              legend.position = "bottom") +
        scale_fill_brewer(palette = "Accent") +
        labs(fill = "View condition")
jpeg(file="partialCoPresence.jpeg")
p2
dev.off()

partialCoPresenceMV <- subset(partialCoPresence, View=="MV")

index <- 1

for(i in seq(2, 11, 2)) {
model <- t.test(partialCoPresenceMV[i],partialCoPresenceMV[j])$p.value
partialCoPresenceMVResults[[index]] <- model
index <- index+1
}


messUnderstand <- read.csv(file="messUnderstand.csv",stringsAsFactor=FALSE)
messUnderstand.m <- melt(messUnderstand, id.var = "View")
p4 <- ggplot(data = messUnderstand.m, aes(x=variable, y=value)) +
        geom_boxplot(aes(fill=View)) +
        scale_y_continuous(name = "Likhart not very - very much",
                           breaks = seq(0, 7, 1),
                           limits=c(0, 7)) +
        scale_x_discrete(name = "Networked minds message understanding") +
        ggtitle("Responses comparing single view and multi view") +
        theme_bw() +
        theme(plot.title = element_text(size = 14, family = "Tahoma", face = "bold"),
              text = element_text(size = 12, family = "Tahoma"),
              axis.title = element_text(face="bold"),
              axis.text.x=element_text(size = 11, angle = 90, hjust = 1),
              legend.position = "bottom") +
        scale_fill_brewer(palette = "Accent") +
        labs(fill = "View condition")
jpeg(file="messUnderstand.jpeg")
p4
dev.off()


messUnderstandMV <- subset(messUnderstand, View=="MV")

index <- 1

for(i in seq(2, 11, 2)) {
  model <- t.test(messUnderstandMV[i],messUnderstandMV[j])$p.value
  messUnderstandMVResults[[index]] <- model
  index <- index+1
}


mydata <- read.csv(file="embodyInSpace.csv",stringsAsFactor=FALSE)
tests<-split(mydata,mydata$View)
multiview=tests[[1]]
singleview=tests[[2]]
mydata.m <- melt(mydata, id.var = "View")
p5 <- ggplot(data = mydata.m, aes(x=variable, y=value)) +
        geom_boxplot(aes(fill=View)) +
        scale_y_continuous(name = "Likhart not very - very much",
                           breaks = seq(0, 7, 1),
                           limits=c(0, 7)) +
        scale_x_discrete(name = "Embodiment") +
        ggtitle("Responses comparing single view and multi view") +
        theme_bw() +
        theme(plot.title = element_text(size = 14, family = "Tahoma", face = "bold"),
              text = element_text(size = 12, family = "Tahoma"),
```

```
            axis.title = element_text(face="bold"),
            axis.text.x=element_text(size = 11, angle = 90, hjust = 1),
            legend.position = "bottom") +
      scale_fill_brewer(palette = "Accent") +
      labs(fill = "View condition")
jpeg(file="saving_plot5.jpeg")
p5
dev.off()


attAlloc <- read.csv(file="attAlloc.csv",stringsAsFactor=FALSE)
attAlloc.m <- melt(attAlloc, id.var = "View")
p6 <- ggplot(data = attAlloc.m, aes(x=variable, y=value)) +
      geom_boxplot(aes(fill=View)) +
      scale_y_continuous(name = "Likhart not very - very much",
                         breaks = seq(0, 7, 1),
                         limits=c(0, 7)) +
      scale_x_discrete(name = "Attention allocation") +
      ggtitle("Responses comparing single view and multi view") +
      theme_bw() +
      theme(plot.title = element_text(size = 14, family = "Tahoma", face = "bold"),
            text = element_text(size = 12, family = "Tahoma"),
            axis.title = element_text(face="bold"),
            axis.text.x=element_text(size = 11, angle = 90, hjust = 1),
            legend.position = "bottom") +
      scale_fill_brewer(palette = "Accent") +
      labs(fill = "View condition")
jpeg(file="attAlloc.jpeg")
p6
dev.off()
attAllocMV <- subset(attAlloc, View=="MV")

index <- 1





affectUnderstand <- read.csv(file="affectiveUnderstand.csv",stringsAsFactor=FALSE)
affectUnderstand.m <- melt(affectUnderstand, id.var = "View")
p7 <- ggplot(data = affectUnderstand.m, aes(x=variable, y=value)) +
      geom_boxplot(aes(fill=View)) +
      scale_y_continuous(name = "Likhart not very - very much",
                         breaks = seq(0, 7, 1),
                         limits=c(0, 7)) +
      scale_x_discrete(name = "Affective understanding") +
      ggtitle("Responses comparing single view and multi view") +
      theme_bw() +
      theme(plot.title = element_text(size = 14, family = "Tahoma", face = "bold"),
            text = element_text(size = 12, family = "Tahoma"),
            axis.title = element_text(face="bold"),
            axis.text.x=element_text(size = 11, angle = 90, hjust = 1),
            legend.position = "bottom") +
      scale_fill_brewer(palette = "Accent") +
      labs(fill = "View condition")
jpeg(file="saving_plot7.jpeg")
p7
dev.off()
affectUnderstandMV <- subset(affectUnderstand, View=="MV")
```