City: A Mixture of Old and New Media

Matthew Chalmers
Computing Science, University of Glasgow, UK

1. Introduction

The majority of the chapters in this book describe what might nowadays be called 'traditional' inhabited information spaces: collaborative virtual environments (CVEs) or virtual worlds. Although not all CVEs centre on computer–rendered 3D graphics, the paradigmatic CVE does so. Shared 3D virtual environments are emblematic of CVE research, but have only gained public acceptance in the form of computer games. The focused engagement in such games is designed to fit with the closed world of the virtual environment. A player can become immersed in a game — closed off from the 'real' world — by attention as much as by apparatus. A PC at home can be as engaging as the head–mounted displays and immersive projection technologies in research labs. However, even a single–player non–networked game may be a resource for social interaction e.g. played by one person while friends and family shout advice from the sofa, order pizza by phone and slip into the kitchen to get more drinks. A computer game is a resource for far more social interaction than the software's architecture may suggest. In general, the wider context of use is hardly modelled or represented in the system. Games' internal data structures are designed to be decoupled, i.e. closed off from the other media people use in everyday life, and this decontextualised design approach has worked well in this domain.

Although many households, schools and workplaces have computers that could support 2D or 3D virtual environments, such CVEs are rarely used as a medium of family interaction, education or work. One reason for this may be the decoupling from the more traditional interactions of family members, the overall educational activities of the school and the business of the workplace. The information within the CVE would have to correspond with each user's wider context, and this would require sensing and tracking of users' activity in media beyond that of the computer. To paraphrase part of this book's introduction, a large amount of our activity relies on the knowledge of what other people do, and what people do in the home, street, school and workplace involves many non–computational media. However, CVEs are designed on the basis of narrowly focused engagement, decoupled from their users' wider context of family, friends, learning and work. Other than in games, CVEs' decontextualised design approach has not led to popularity or widespread use.

As the introduction to the book also points out, a number of researchers have begun to work on inhabited information spaces that are more 'out in the world' than traditional CVEs. A rhetorical example the editors give is "a system that enables co–located groups to co–operatively work with information by using a display projected onto physical artifacts". An IIS may include tangible artefacts in more traditional media, such as urban models and interaction devices made from wood, wire and plastic (Underkoffler and Ishii 1999), or tiles and pages made from toner and paper in augmented reality systems such as MagicBook (Billinghurst et al. 2001).

The work discussed in this chapter is intended to go further in this direction. As part of the Equator interdisciplinary research collaboration (www.equator.ac.uk), the City project explores digital spaces that are peers with others rather than a digital space that is the primary focus or locus of activity. For example, in our system, one person's use of a 3D VR model of a museum exhibition is coupled with another person's use of hypermedia describing the

exhibition as well as a third person's use of the traditional 'bricks and mortar' exhibition. No one of the three is primary; each is part of the context of the other two. We combined the media of traditional exhibitions, mobile computers, hypermedia and virtual environments in one design, and support interaction between people using different subsets of this heterogeneous set of media.

The project was initially theory–led. A number of theoretical issues were outlined in a discussion document, and then exemplified them by scenarios of technology use. As the project grew and developed, our theoretical issues, design scenarios, system development, evaluation of pilot trials, and observational studies all affected each other. Although I sometimes campaign for the rule of theory, no one of these areas is primary. Instead, each is part of the context of the others. The project has always aimed to get out into the streets of the city, but we decided to begin our work in a more controlled setting: the Mackintosh Interpretation Centre, a permanent exhibition devoted to the life and work of Charles Rennie Mackintosh (1868–1928). Mackintosh was a Glasgow architect, designer and artist, and several of his buildings and other institutions related to his work stand in the city. Often simply called the 'Mack Room', the Centre is comprised of textual and graphical displays with some original artifacts, as well as over 20 screens presenting video and interactive digital material. The Mack Room is in The Lighthouse, Scotland's Centre for Design, Architecture and the City (www.thelighthouse.co.uk).

More generally, we are exploring the way that digital information is just another part of an interwoven set of media, the use of which constitutes inhabiting the city. Unlike traditional CVEs, we aim for systems coupled with and contextualised in everyday activity, and hence in accord with contemporary theory of the use of language and space. A basic theoretical premise is that we can only use digital media because of such interweaving and interdependence, and we are looking for ways to increase and take advantage of the interdependence of traditional and new media. This theoretical standpoint is set out in the next section of this chapter. The subsequent section outlines a system design based on this standpoint, followed by a discussion of user experience in trials of an implemented research prototype. Some of the details of our ongoing and future work are then outlined before a concluding section that offers more general reflections on the project.

2. Theory

This section focuses on conceptions of space and the media associated with work, and how we often conceive of space as a medium that stands above or apart from others. It is this usually implicit assumption that lets us talk of information spaces as being 'inhabited'. I'd like to present an opposing view that treats information spaces as merely one medium among the many used in everyday life. My approach is based on experience with information visualisation and virtual environments, as well as some borrowing of structuralist semiotics (Saussure 1983; Nöth 1995) and philosophical hermeneutics (Grondin 1994). Part of this latter work was set out in a recent J. CSCW paper (Chalmers 2003).

When discussing work, and designing systems for remote collaboration, we all too often concentrate on emulating the spatial aspects of the workplace, e.g. modelling spatial forms and supporting remote communication that appears to be like face–to–face interaction. All design has to be biased in some way, and the bias towards space in CVE research may be due to it being technologically led more than sociologically, semiologically or philosophically led. The arrival of cheap graphics hardware and the eye–catching novelty of 3D images gave rise to a good deal of work that focused rather narrowly on the construction of rather

decoupled and decontextualised information spaces. This is true in some of my own work over the past twelve years, ranging from (Chalmers 1991) to (Morrison et al. 2002), and there are strengths, weaknesses and alternatives to such a bias (Chalmers 2001).

Newer technologically driven research is weakening or revealing CVE research's implicit assumptions of space's primacy and independence. Many of the characteristic design principles and assumptions were established before the current fashion for mobile computers and ubiquitous computing. Nowadays, it possible to obtain tolerable frame rates for 3D graphics on a wirelessly net–connected handheld computer. A person can thus be walking down a city street with a friend, chatting about a museum they intend to visit, while simultaneously watching the avatar of another friend moving through a CVE — with that 'remote' friend also taking part in the conversation. In this case, it would seem difficult to claim that the person 'inhabits' the information space. One might ask whether the person is in digital or virtual space, or in real or physical space, but the question is based on two false dichotomies: digital media are no more or less real than older media, and computers are just as physical as buildings and books.

The workplace has always been affected by communication with people in other locations. Many traditional, everyday and non–digital media support remote interaction, for example letters, books, maps and the land–line telephone. There are already digital media in the contemporary workplace that support remote interaction, such as email and mobile telephones. Nowadays, why do we not speak of 'entering cyberspace' when we use email, as people did a decade ago? Why do we not inhabit telephone space, or speech space, or MacDonalds' employee name badge space? I suggest that a principle from philosophical hermeneutics is useful here: we don't talk about these technologies in such marked ways because we have appropriated them into our everyday life and language. We no longer 'enter cyberspace' because email is so interwoven in our everyday life and familiar in our experience that we don't need to mark it out in such a way. We don't inhabit telephone space because we understand telephones, in particular how to present ourselves through them and how to present ourselves to 'spectators' nearby who can perceive our use of them. We only 'inhabit' virtual worlds because their designs are so new and decoupled from other media. Experience and understanding of such coupling lets us focus on the task of communication, not on the tool for communication, just as a carpenter engaged in his work focuses on hammering and not the hammer.

We continually mix and couple media in our everyday communication: walking, gesturing and pointing while talking, and referring to places and what people did in them as one writes. Space is an essential part of this mix. It has its unique characteristics that differentiate it from other media, but it has no privileged position above or apart from them. It does not stand alone as a paragon for computational media to emulate. More generally, a medium cannot be fully used or understood in an isolated or 'singular' way. People's activity continually combines and cuts across different media, interweaving those media and building up the patterns of association and use that constitute experience and understanding. A person's work or activity may be influenced by the configuration of space around them and the interactions that space affords, but also by books, telephones, hypermedia, 3D computer graphics and so forth. People act and work through the full range of media they have ready to hand. A narrow emphasis on space as the paramount resource for activity underrates the influence of other media. Recent technological developments, such as mobile phones and email, heighten or highlight a phenomenon already familiar in the use of older media such as written text, maps and cinema, and well–explored in older disciplines than computer science.

For example, a city's meaning is not just in its bricks and mortar, but also in our understanding and use of the information about it. At any time, one is likely to have symbols in a number of heterogeneous media available for interpretation and use. As I walk through a train station towards a city square, the map in my hand, the voice of a friend on my mobile phone, the signs informing me of exit routes and the posters advertising exciting shopping opportunities are all open for my interpretation and action. Temporally, symbols in an even broader range of media influence me, as my activity is influenced by my past experience and my expectations of the future. Past experience may include my previous visits to that city, my browsing of a web site with good maps to print out, and my experience of magazines, books and films about urban life, and so forth. My language and culture, spanning media old and new, affect me as much as the immediate perception of spatial form. Since Heidegger and Saussure, a fundamental tenet of philosophy and linguistics has been that language is constituted by all the symbols and all the media one uses, with each symbol interpreted through immediate individual perception as well as past social experience. Contemporary neurophysiology is in strong accord with this view (Churchland and Churchland 1998; Edelman and Tononi 2000), as is the field most obviously related to the design of space, architecture and urban design (Leach 1997).

The differences between media are usually very obvious. We can characterise media and treat each one as if it were an isolated individuated entity because of the senses we use in perceiving each one, and also because of our understanding of how to relate and to distinguish examples of each one. For example, it is easy to distinguish the spoken word "red" from the written word *red* because of the senses one uses in each case. Despite having the same letters, it is easy to distinguish *tar* from *rat* by looking at the order of letters within each written word. Simple rules about what one can immediately see, hear, etc. within a word begin to strain and then break when one considers, for example, how we distinguish homonyms such as *rose*. The written word *rose* can mean many things, including a flower and having risen. When spoken, the same syllables can also mean linear structures (rows), about or belonging to fish eggs (roe's), moving in a boat (rows), small deer (roes) and multiple occurrences of the Greek letter ρ (rhos). The word's usage is understood through its context—one's understanding of the other symbols co–occurring with its use—rather than perception of the word's pattern of syllables or letters.

Context becomes progressively more important as we turn from thinking about the differences between media, and distinguishing symbols, to considering the similarities of media and the relatedness of symbols. For example, the spoken word "red" and the written word *red* are related because we can use either of them in the context of rose blooms, fresh blood, the former USSR and so forth. We understand, relate and differentiate symbols through experience of contexts of use within a culture. As shown in the early 20th century by Saussure, this understanding is not solely dependent on the form or medium of each symbol, but also on how we use each symbol in the context of other symbols — and this context includes symbols in other media.

Taking fuller account of the interdependence of media enriches our understanding of space and of work. Consideration of how to make systems that are consistent with this standpoint opens up new possibilities for technology design and for computer–mediated social interaction. More particularly, it opens up a wealth of approaches based on coupling and contextualisation. For example, a museum exhibition might be associated with a set of web pages so that walking into a room on a particular architect triggers the display of text describing the life and work of that architect. Similarly, reading the text might trigger display of a map or visualisation of the room, affording access to a structured collection of blueprints, design sketches and building models. The space of the room would be coupled with the text

of the page, with each becoming part of the context of the other. In terms of social interaction, a person walking into the room might be made aware of a friend's reading of the web page, and hence open to conversation about the exhibition despite the two people being geographically remote from each other.

Our intention is to support social interaction, as is familiar in traditional museums where co–visitors use awareness of each other's interaction with exhibits as a resource for their interaction with each other, and use interaction with each other in interpreting the exhibits (Falk and Dierking 1992; Galani and Chalmers 2002). The City project explores the coupling of new and traditional media, weaving them together to form resources for social interaction and interpretation.

In particular, our 2002 experiment explored social interaction between people in different locations and contexts where, by definition, they have different resources at hand. As they discuss and refer to contextual information, heterogeneity is inevitable: one person can use the non–digital resources of his or her location while others have only digital representations of that location. A case that is more easily handled is audio: each person will hear his or her own voice and sounds from other nearby sources differently to others, because of the digitisation and transmission of audio, but we have become relatively accustomed to handling this. A much more challenging heterogeneity is that of people's position, orientation and gesture within rooms and buildings. For example, the Mack Room presents much greater visual and tactile richness than the room's digital representations e.g. maps and VR models. Unlike most earlier CSCW research, the City project addresses this inevitable heterogeneity by coupling media together, tracking activity in each medium and representing it in others, and so letting participants interweave these media in their social interaction.

3. System

This section outlines our prototype system, beginning with its infrastructure: the EQUIP platform, the Auld Leaky contextual link server, the VR Juggler framework, and the Bristol ultrasonic positioning system. More detail of this system can be found in (MacColl et al. 2002).

The EQUIP platform is being developed within Equator to support information sharing between heterogeneous devices. It provides a run-time infrastructure to support interoperation between Java and C++, and supports extensibility e.g. dynamic loading of C++ classes. The University of Nottingham is leading development of EQUIP, with contributions from the various Equator projects. City uses it as a blackboard architecture through which VR Juggler, Auld Leaky and the Bristol ultrasonics interoperate. Data items representing user context, an underlying spatial model and context–dependent content are stored for manipulation by City clients and services. Additional EQUIP facilities support real-time 3-dimensional graphics and mathematical operations, and abstract, renderable scene graph nodes. In addition, interfaces between EQUIP and a number of other systems have been developed, including the University of Iowa's VR Juggler.

VR Juggler is used as the renderer for 3-dimensional graphics. It is described as a virtual platform for virtual reality (VR) application development. It is a high-level, cross-platform architecture supporting immersive and non-immersive presentations. Both UCL and Nottingham have immersive projection facilities, and the UCL facility has been used for development and pilot trials in the City project. The 3-dimensional graphics rendering is used to provide an analogue to the traditional exhibition space visited by traditional visitors. For

World Wide Web visitors, the space is represented as a 2-dimensional map. We also require a presentation of the information in the exhibition displays, and this is provided by Auld Leaky.

Auld Leaky is a lightweight contextual link server being developed within Equator to store and serve hypermedia structures, using context to filter query results. The model used to define the structures is capable of representing a variety of hypermedia domains: link–based, spatial and taxonomic. Auld Leaky is being developed by the University of Southampton, and is written in Perl although it has a Java API. Information is encoded as an XML linkbase, loaded into Auld Leaky and queried using HTTP. The text and images of the hypermedia were taken from the Mack Room's catalogue. Contextual queries are used to generate location– and device–specific content to be delivered by the Apache World Wide Web server and servlet engine. For 2– or 3–dimensional renderings, a location can easily be derived from the position of a visitor's avatar or map marker. For the mobile computer, an ultrasonic system is used to provide position information.

The University of Bristol is developing a low cost indoor positioning system using a combination of radio frequency (RF) and ultrasonics (Muller and Randell 2001) as part of its contribution to Equator. The system uses a single RF transmitter for synchronisation, with ultrasonic transmitters for positioning. The ultrasonics transmit at known intervals after the RF, and are received by a handheld or wearable. Each second, the variations in flight time of the ultrasonic transmissions are used to calculate the spatial position of the receiver. The receiver incorporates a magnetic compass to provide orientation information. The City project installation involves eight ultrasonic transmitters covering the approximately 10m by 20m area of the Mack Room. The room is a challenging environment for ultrasonics, as it is split into two large areas by a partial 'time line' wall and has some areas set up as cubicles within which ultrasonic reception is virtually impossible. For aesthetic and coverage reasons, the transmitters are set on top of walls, displays and cubicles so that ultrasonic transmissions are reflected off the ceiling.

The system supports a shared visiting experience, with one visitor using a handheld or wearable computer in the Mack Room, a second visitor using the World Wide Web on a laptop or PC in another room, and a third using 3–dimensional graphics on a similar machine in a third room. These computers communicate through 802.11 wireless Ethernet. A separate audio subsystem, that we will not detail here, handles visitors' speaking to and hearing each other. In discussion and development scenarios we name the visitors Vee, Dub and Ana respectively, and these will also be used in the remainder of this chapter. The names do have roots: Vee is for 'visitor', and Vee was the user in our first design scenario; Dub is from the first syllable of "double-U", as in WWW; and Ana stems from 'analogue', playing with the way that the digital space of the VR is an analogue of the Mack Room. Some people have suggested that Vee should be the 'virtual visitor' but we decided to keep the name on so as to irritate those who think that Vee's experience is not strongly influenced by digital media.

Spatial awareness is supported by tracking activity in each of the 2D, 3D and handheld systems, sending position and orientation information for each one into EQUIP, and then rendering the information about all visitors to each individual visitor. The components of the prototype system operate similarly for each visitor, broadly as follows:

1. store spatial position and orientation in EQUIP;
2. retrieve and render positions of other visitors;
3. store named location in EQUIP in response to position change;
4. store content from Auld Leaky in EQUIP in response to location change;
5. format content for presentation and advise client program of availability in response to content change.

For Ana, position and orientation information is automatically published in the EQUIP data space by the VR Juggler client, and the positions of all visitors stored in the data space are automatically rendered as 3-dimensional avatars. Figure 1 shows a non-immersive spatial awareness display for Ana, with avatars representing Vee and Dub (displaying only heads rather than complete avatars).

Figure 1. A non-immersive VR display of the Mack Room for Ana, with avatars representing Vee and Dub (displaying only heads rather than complete avatars).

Vee uses a Hewlett–Packard Jornada that polls position and orientation sensors, and sends the results via a proxy into EQUIP. The proxy is also responsible retrieving the information about the other visitors, and the positions and orientations of all visitors are presented to Vee on a 2-dimensional map. Figure 2 shows a visitor the Mack Room, with a handheld and an ultrasonics receiver. The figure also shows a close-up view of the handheld as an inset.

Dub interacts with a Java applet in a World Wide Web browser frame. The applet communicates via a proxy that converts mouse clicks on a 2-dimensional map of the Mack Room to position and orientation information. The applet also displays representations of all visitors. An example of Dub's map is shown in Figure 3, corresponding to Ana's 3-dimensional display in Figure 1. The red boxes on Dub's map are trigger zones, discussed in the next section. Vee's map is similar to, but simpler than, Dub's.

Figure 2. A visitor in the Mack Room, 'Vee', with a handheld computer and ultrasonics receiver. The figure also shows a close-up view of the handheld as an inset.
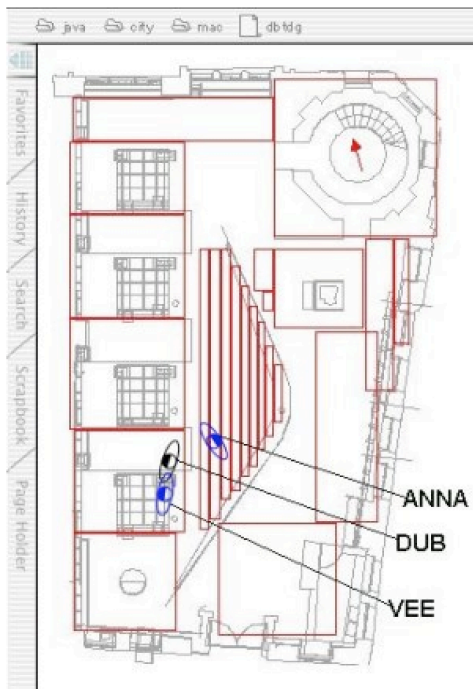

Figure 3. An example of Dub's map. For paper publication, we have annotated the image with the names of the visitors.

Shared visiting requires a sense of shared context, and hence some comparability of the information available to each visitor, but we also wanted to maintain and explore a degree of the heterogeneity that is inevitable in remote collaboration. Vee has the rich environment of the Mack Room's displays and artefacts, as shown in Figure 2. Providing hypermedia to Dub and Ana involves converting positions to named locations, querying Auld Leaky with the visitor's device and location in order to generate informational content, and then formatting and presenting the content. In initial trials of our prototype system we did not deliver hypermedia content to Vee, encouraging her to use the rich content of the existing room when interacting with her friends. Dub and Ana have rich access to the web that Vee lacks, and they can move and jump between Mack Room locations in ways that Vee cannot. Ana's 3D view of the Mack Room VRML model has greater visual richness than the 2D maps of Dub and Vee, but she also has visual occlusions to deal with.

Positions are converted to locations by an EQUIP–connected service. These locations represent semantically significant volumes or extents within the spatial model (shown as red outlines on Dub's map, as in Figure 3). Also, for each visitor a target is inserted into the EQUIP data space, currently equivalent to a 10cm cube held in the hand. Detection of a collision between a target and a sensor invokes code that inserts a new *(user, location)* item into the data space. Adding or updating a user-location item in the data space triggers a query to Auld Leaky and the results — a set of hypermedia fragments — are stored in the EQUIP data space. Adding or updating such a set of hypermedia fragments in the data space triggers formatting for delivery to the visitor. The content fragments are retrieved and combined into an HTML page. Dub's applet displays the HTML page in a separate browser frame set aside for this purpose. Ana also runs a browser and applet that, when advised, displays HTML pages i.e. Ana has no map display but does have a textual display.

The overall effect, then, is that all three visitors have location information about their co–visitors. Dub and Ana each have spatial and textual information about the exhibition, with the text updated as they use the 2D map and the 3D graphics. Vee has the traditional information of the exhibition room. Many of the artefacts and exhibits have corresponding representations for each of the three visitors. We sometimes refer to such artefacts and exhibits as 'hybrid objects' because of the visitors' tight interaction around corresponding heterogeneous representations. Space, text and audio afforded sufficient interaction, context and reference to support a shared visiting experience, as the next section discusses.

4. User Experience

In the summer of 2002, the City project carried out a set of system trials in the Mackintosh Interpretation Centre. Rather than trying to make the best possible system for the Mack Room in particular, our focus was on general lessons we could learn for the design of systems involving heterogeneous representations and interactive media. We aimed to increase our experience and understanding of how these systems serve as constraints and resources for users' interaction. We had already studied the use of a number of cultural institutions, including the Mack Room, without our technology (Galani and Chalmers 2002), and were interested in comparing the Mack Room with and without our technological intervention. Fuller discussion of the trial can be found in (Brown et al. 2003).

The trials involved 34 participants: ten groups of three and two groups of two. The groups of three consisted of a Dub (using the web), an Ana (using VR) and a Vee (using a mobile computer). Dub was in the Interpretation Centre, while Ana and Dub were in separate rooms on a different floor of the Lighthouse. The pairs explored different combinations: Dub and

Ana visiting without the physical visitor (one trial), and Vee and Ana (one trial). For the first half of the trial, participants were asked to explore the Centre together, to familiarise themselves with the technology and how they could co-visit. Since we were specifically interested in how the system supported social interaction, we introduced an artificial task for the second half of the trial. Each participant was given three questions, and the group was asked to answer these questions together. Some questions were designed to provoke open–ended discussion and interaction between the participants. For example, participants were asked "What is the group's favourite Mackintosh painting?" and "What contribution has Mackintosh made to Glasgow?" as well as more factual questions such as "What was Mackintosh's birthday?" This combination of open and task-centred behaviour allowed us to study activity that was typical of a museum visit, such as finding exhibits, and to observe how the system supported the shared aspects of visiting a museum. During the trial, use of the system was heavily logged and each Dub was video taped. After the trial, the participants were interviewed as a group in a recorded semi–structured debriefing.

For analysis, we combined the map view used by participants, with the video and audio recordings. We analysed transcripts of the post-trial debriefings, and the logs of the visitors' use of the system. We paid close attention to the details of how users interact with each other and with technology, especially through video analysis. With particular interest in the use of location and of exhibits, we looked for 'critical moments' where the system was used in a way that would let us reveal design lessons, consistencies and inconsistencies with theory, and comparisons with earlier studies.

The participants engaged in rich social interaction around the hybrid or coupled exhibits. When participants found that objects corresponded, in this way they quickly were able to move on to using them in their shared tasks and activities, e.g. discussing the qualities of the object and comparing it to the other exhibits. In the following extract the participants discuss a set of Mackintosh pictures to decide which one they like the most. Square brackets (e.g. [pet]unias) show overlapping talk and underlining shows a speaker's emphasis.

Vee: Petunias is errm better for me than Rosemaries
Ana: Ok [pet]unias
Dub: [hhh] Petunias it is
Vee: Early work
Ana: Hey guys see: this other one it's really nice. It's called Fort Mailly hhh Fort Mailly in nineteen twenty seven
Vee: Nineteen twenty seven
Ana: Yeah, it's got the light
Vee: Yeah I know but I like [Roses]
Ana: [Can you] see it?
Dub: Fort Mailly?
Ana: Hmmmm
Dub: Yeah that's quite nice
Vee: I still prefer Roses

The visitors do not focus on the system or the media involved, such as the differences between the digital and the printed reproductions of the Fort Mailly painting. Instead they focus on its aesthetic qualities and the task of deciding which picture they like most. They don't focus on the tool but on the task.

However, interacting around these hybrid exhibits was not without its problems. In ordinary face–to–face interaction, we assume a degree of commonality in the objects that we can see, hear, touch and so forth. Each user of our system had to build an understanding of the perspectives of the other users had of the Centre. They could often see similar things, or

related things, but not the same things from the same perspective.

Dub and Vee frequently guided Ana to specific exhibits verbally. We designed the system so that Ana did not have an overview map, and therefore might find that occlusion of objects in the exhibit was a problem. In turn, since Dub and Ana could shift attention between artifacts placed far apart in the room much more quickly than Vee, they frequently found information and then guided Vee to the corresponding location. Again this was often done verbally, but more spatial or graphical guiding is discussed later.

Participants also made use of shared location and orientation, using icons on an outline map for Vee and Dub, and avatars in Ana's 3D display. Shared awareness of location also allowed users to quickly move to their friends, and to quickly find or confirm the exhibits being discussed i.e. to quickly find what their friends were looking at and then move so as to look at the same or a closely related thing. They developed simple means to gesture. In one case, Ana moved her avatar back and forth in a 'wiggle' so as to confirm which icon represented her and to show her location to Dub. This gesture was something like the wave of a hand used by someone to show he or she is in a crowd. Global location could be seen 'at a glance' on the map, without the need for the visitors to use talk, and but examples such as Ana's wiggle show participants' awareness of how they would be perceived by each other in different media.

Participants learned about each other's perspectives through questions and observations, building understanding of what they shared and what they did not, and thus how to more smoothly interact with each other through the resources at hand. Indeed, participants put considerable effort into designing their interactions to take into account the characteristics and limitations of their varying views of the Centre. For example, the above extract shows Vee emphasizing and confirming the year in which 'Fort Mailly' was painted ("nineteen twenty-seven"). One of the largest exhibits in the Mack Room is a long wall of panels, with each of the chronologically ordered panels showing a number of images associated with a year of Mackintosh's life. Mentioning a year to Vee would let her move quickly to a corresponding part of the exhibit. The information presented to Ana and Dub about this wall was broken up into a page for each year, so any one of the visitors could help guide the others to a particular image in that exhibit by specifying the year.

The use of our system involved more talk, and louder talk, between the co-visitors than we observed in conventional museum settings. For example, during a post–visit discussion two participants were asked:

Q: Is it different to a museum visit?
A: Yeah, it's really talkative.
B: You kind of go "Mmmm, that's nice" […] If you find something interesting, you go "Look", and "That's over here"

Another commented on being able to talk without disturbing others:

I quite enjoyed the social engagement […] being able to talk about everything more and not feeling that you are disturbing. Not thinking about other users in the gallery, you know it's kind of liberating

I offer here a few of the many possible reasons for this, stemming from the use of multiple rooms, the audio hardware used and, perhaps more interestingly, the set of media available for interaction. Two of the participants were in rooms other than the traditional exhibition room, and so were less influenced by the normal hushed reverence adopted inside museums. All three participants were engaged in a trial of an unfamiliar technology, without established social norms for speech levels. Vee often talked more loudly and paid less attention to other members of the public in the Mack Room. Much like a telephone, the loudness of speech

needed for audibility through the audio hardware may not be the same as in face–to–face interaction. Lastly, the participants often used shared audio because the coarse–grained representations of position and orientation did not afford familiar use of gesture and posture. However, participants were not solely using talk in order to circumvent the tools given; they were often talking while engaged in the tasks. Also, they developed gestures for the media at hand, for example Ana's wiggle: a gesture made in a 3D VR to be seen on a 2D map.

In everyday face-to-face situations, interaction between two people can be impeded for a number of reasons such as one or both of the participants choosing to interact with others (e.g. on the telephone) or to interact with objects, or by participants being forced to interact with other people or objects because of interruption, breakdown, occlusion and so forth. In the trial, interaction would often pause when participants found a difference between the visitors' representations of the Centre. For example, the room's interactive video displays were only available to Vee. When a visitor started to use and talk about a display or exhibit that was not shared, the other participants would refrain from interacting and move on to other exhibits.

Similarly, Dub and Ana generally used movement on the map or VR in order to access information, in preference to the more conventional hyperlink navigation. This seems to have been partly due to the fact that such spatial movement was more a part of the shared experience, in terms of conversational references to locations but also as a way of avoiding future confusion: following links to a web page about a new exhibit did not move the participant's icon or avatar to the corresponding new position. (This capability was implemented during the trial period but, to maintain consistency, was not deployed.) This meant that 'web movement' could leave a visitor's icon in a potentially confusing place.

While there were occasional interactional breakdowns, they were not fatal for the sense of a shared visit or for interaction. Overall, participants showed skill in finding ways to handle the differences between the different representations, and exploiting corresponding and coupled features. The system successfully supported a shared experience by enabling users to talk about and interact around the exhibition, offering a socially engaging experience beyond that available to a conventional web site visitor.

5. Ongoing and Future Work

We continue to explore remote collaboration in cultural information and cultural institutions, in particular collaboration involving heterogeneous media. We support social context as a resource for the interpretation of information, and contextual information as a resource for social interaction. We are extending our system to be used in more of the city than the Lighthouse, adding GPS, dead reckoning, GPRS, and further 802.11 aerials. We are using the publicly available VRML model of central Glasgow from the University of Strathclyde, and 2D maps from services such as EDINA (www.edina.ac.uk). We have been undertaking field studies of visitors in a range of locations in Glasgow, seeing how their visits include far more than traditional cultural institutions, and how they use resources such as tourist information centres, maps, guidebooks and signage. We aim to run another field trial that involves a number of participants visiting the city, each of whom has a wearable computer on which tightly coupled 3D VR, 2D maps and hypermedia can be used. Each participant's activity will be available to the others synchronously, much as in the Mack Room, but also asynchronously.

One reason for this latter development is to move beyond the traditional objectifying or 'scientific' systems of classification and retrieval that too often are the only means of access

to digital information. Influenced by the theoretical standpoint outlined in Section 2 and the Recer system (Chalmers et al. 1998), we have built a central information resource, connected to EQUIP and thus accessible via a variety of media and devices, and which stores a growing and evolving body of individuals' paths or narratives through a range of symbols: our own images and fragments of hypertext, annotations made by users, locations in the city, and other locations and web pages worldwide. The system uses this resource to make contextually specific recommendations of people, places and things by comparing each person's recent activity with similar sections of the past activity of selected others. Paths can also be more directly shown on maps, in VR models and woven into web pages. We will allow this body of information to grow as people use it, making new associations between symbols and adding in new ones. Information access based on evolving inter-subjective patterns of contextual association and use will complement access based on more static and objective interpretation. We are also experimenting with bridging between the two, using pre–written 'official' explanations of the connections between symbols as a means to enrich the dynamically created recommendations.

Remote collaboration brings abstraction and approximation as a system monitors and senses activity such as a person walking across a city. Issues such as sampling, resolution, delay, disconnection and uncertainty have to be faced as one decides how to represent the activity inside a system, even before one considers how to represent the activity to a remote collaborator. There is no getting away from the way that activity is going to be interpreted through sensors and transducers such as cameras, GPS and ultrasonic systems, and any digital medium has characteristic losses and uncertainties. While we want to make new resources for interaction available to people, combining new media with old in a perfectly seamless way is not going to happen. Uncertainties and inaccuracies ('seams') are an inherent part of any communicative medium, and people often learn to use these characteristics for their own ends. For example, mobile phones can be set to display the current cell, if the service provider permits, and some people choose to enable this facility. This is an elegant ambient or peripheral presentation of potentially useful information: users can choose what use to make of it, e.g. seeking a stronger signal by moving to a location that forces handover to another cell. Cell boundaries and signal strengths are interactional resources of the medium. Similarly, long–term use of video–mediated communication was reported by Dourish and Adler (1994) to lead to "complex patterns of behaviour built up around the interactional details of the video medium […] When the medium changes, the mechanisms change too; but the communicative achievements remain."

Recalling a term used by Mark Weiser (1994), we see *seamfulness* as an important design goal for our future work. We plan to design in explicit representations of the errors and uncertainties in our systems, letting people take account of the characteristic heterogeneity, errors and limitations of the systems we offer them. For example, we are starting to develop explicit presentations to accommodate uncertainty due to ultrasonic and GPS–based positioning, showing a person's sensed position as a spatial extent, rather than as a point, and showing estimates of sensing accuracy and communications bandwidth on our city maps and models.

6. Conclusion

The City project emphasises the interdependence of media, such as computer graphics and audio, with others. We have explored the combination of CVE technology with hypermedia and mobile computers, and also with the architecture and exhibits of the Mackintosh Interpretation Centre. Supporting broad social context in remote collaboration involves

heterogeneity, and our project aimed to address this through coupling and correspondences between media. Users of a mixed collection of interactive media were able to enjoy a shared visit experience, engaging in collaboration through awareness of each other's activity and through more focused talk and interaction around 'hybrid' objects.

By presenting the theoretical issues underlying much of this work, as well as system design and experiences of use, this chapter may serve to complement many other chapters of this book. Rather than considering that users inhabit our information space, we see people as inhabiting cities and towns, and using new technologies and older media to interact with friends, relatives and colleagues. Looking to the near future, CVE technologies will be widely available via phones and mobile computers. I suggest that taking fuller account of their use among a wider set of technologies and media, and designing for contextuality, heterogeneity and seamfulness, will greatly enrich our work.

7. Acknowledgements

References

1. Billinghurst M, Karo H, Poupyrev I (2001) The MagicBook: A Transitional AR Interface. Computer Graphics, 25:745-753.
2. Brown B, MacColl I, Chalmers I, Galani A, Randell C, Steed A (2003) Lessons from the Lighthouse: Collaboration in a Shared Mixed Reality System. To appear in Proc. ACM Computer–Human Interaction (CHI03), Fort Lauderdale.
3. Chalmers M (1991) Seeing the World Through Word-Coloured Glasses. Proc. 2nd Intl. Conf. on Cyberspace, U. California Santa Cruz.
4. Chalmers M (1999) Comparing Information Access Approaches. J. ASIS 50th Anniversary Issue, 50(12):1108-1118.
5. Chalmers M (2003) Awareness, Representation and Interpretation, J. CSCW 11:389–409.
6. Chalmers M, Chitson P (1992) Bead: Explorations in Information Visualisation. Proc. ACM Conf. on Information Retrieval (SIGIR'92), Copenhagen. Published as a special issue of SIGIR Forum, June 1992, ACM Press, pp. 330–337,
7. Chalmers M, Rodden K, Brodbeck D (1998) The Order of Things: Activity–Centred Information Access. In: Proc. World Wide Web (WWW98), Brisbane. Published as Computer Networks and ISDN Systems, 30:359-367.
8. Churchland PM, Churchland PS (1998) On the Contrary: Critical Essays 1987–1997. MIT Press.
9. Dourish P, Adler A (1996) Your Place or Mine? Learning from Long-Term Use of Audio-Video Communication. J. CSCW, 5:33-62.
10. Edelman G, Tononi G (2000) Consciousness: How Matter Becomes Imagination. Allen Lane Penguin Press.
11. Falk H, Dierking L (1992) The Museum Experience. Whalesback Books, Washington.
12. Galani A, Chalmers M (2002) Can You See Me? Exploring Co-Visiting between Physical and Virtual visitors. Proc. Museums and the Web. Archives & Museum Informatics, Boston.
13. Grondin J. (1994) Introduction to Philosophical Hermeneutics. Trans. J. Weinsheimer. Yale U. Press.
14. Leach N (1997) (ed) Rethinking Architecture: A Reader in Cultural Theory. Routledge.
15. Morrison A, Ross G, Chalmers M (2002) A Hybrid Layout Algorithm for Sub–Quadratic Multidimensional Scaling. In: Proc. IEEE Information Visualisation, Boston, pp. 152-160.

16. Nöth W (1995) Handbook of Semiotics. Indiana University Press.
17. Randell C, Muller H (2001) Low Cost Indoor Positioning System. Proc. UbiComp 2001: Ubiquitous Computing, pp. 42–48, Springer.
18. de Saussure, F (1983) Course in General Linguistics. Trans. Wade Baskin. McGraw–Hill. (Original published in 1906.)
19. Underkoffler J, Ishii H (1999) Urp: A Luminous-Tangible Workbench for Urban Planning and Design. Proc. ACM Conf. Computer–Human Interaction (CHI99), pp. 386–393.
20. Weiser M (1994) Creating the Invisible Interface (abstract). Proc. ACM User Interface Software and Technology, p. 1.
21. Wittgenstein L (1958) Philosophical Investigations. 3rd edn, trans. G.E.M. Anscombe, Oxford University Press.