Technical Section

Using mobile group dynamics and virtual time to improve teamwork in large-scale collaborative virtual environments

Trevor J. Dodds *, Roy A. Ruddle

Visualization and Virtual Reality Research Group, School of Computing, University of Leeds, UK

ABSTRACT

Mobile group dynamics (MGDs) assist synchronous working in collaborative virtual environments (CVEs), and virtual time (VT) extends the benefits to asynchronous working. The present paper describes the implementation of MGDs (teleporting, awareness and multiple views) and VT (the utterances of 23 previous users were embedded in a CVE as conversation tags), and their evaluation using an urban planning task. Compared with previous research using the same scenario, the new MGD techniques produced substantial increases in the amount that, and distance over which, participants communicated. With VT participants chose to listen to a quarter of the conversations of their predecessors while performing the task. The embedded VT conversations led to a reduction in the rate at which participants traveled around, but an increase in live communication that took place. Taken together, the studies show how CVE interfaces can be improved for synchronous and asynchronous collaborations, and highlight possibilities for future research.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Collaborative applications in general may be classified in terms of time (synchronous vs. asynchronous) and space (co-located vs. remote) [1]. For example, applications using shared tables and shared wall displays provide for co-located and *synchronous* interaction. Leaving post-it notes in a shared space (or using software which provides the digital equivalent on a single shared display) is an example of co-located and *asynchronous* interaction. Collaborative virtual environments (CVEs) are one way of enabling remote collaboration. They allow virtual co-location of people who are physically remote, by providing a 3D virtual spatial world for people to co-exist in.

Historically, users have had difficulty in understanding the actions of others in CVEs [2,3], and the problems mushroom in a large-scale environment (e.g., a virtual building or city) because of the extra challenges of navigating and locating the whereabouts of one's collaborators. To help with this we have developed techniques called mobile group dynamics (MGDs), which helped groups of people work together while they traveled around large-scale CVEs [4].

This paper: (a) addresses shortcomings in MGDs, which centered on the time it took users to regroup in a place to discuss or see what each other was interested in, and (b) implements the

concept we call virtual time (VT) that allows (to a certain extent) virtual synchronization of people who are physically separated in time. Taken together, our techniques allow both synchronous and asynchronous collaborations in large-scale CVEs. The following sections describe the background and implementation of both suites of techniques (our updated version of MGDs and VT), and then experiments evaluate both. The MGDs work was previously reported in [5], but the VT research is entirely new. Our hypothesis was that the teleporting, awareness and multiple views functionality would improve teamwork. To analyze teamwork, we looked for improvements in two specific areas. First, we wanted to tackle problems of participants spending time collocating to communicate (or waiting until they are collocated before they talk to each other). Second, we wanted to help people work as a team by providing an awareness of the actions and perspectives of others (multiple views tackling problems 1 and 2). These were analyzed using the quantitative data provided by the server's log of activity, and a conversation transcript.

2. Methods for real-time collaboration

Previous research showed how even a basic set of MGDs techniques helped users communicate while they traveled around a virtual urban development and reviewed its design [4]. However, two major areas for improvement were also identified. First, participants tended to spatially regroup to discuss their findings, even though MGDs allowed communication over an infinite distance (there was no distance attenuation for audio

* Corresponding author. Tel.: +44 7752239057.
  E-mail addresses: trev@comp.leeds.ac.uk (T.J. Dodds),
royr@comp.leeds.ac.uk (R.A. Ruddle).

communication between group members). This meant that unnecessary amounts of time were spent traveling to meeting places. Second, if participants wanted to see what others were looking at (e.g., a point of interest that was being discussed) then they had to 'walk' to the appropriate location.

These shortcomings in real-time (i.e., synchronous) collaboration were tackled by adding new functionality to MGDs, taking advantage of the fact that CVEs do not need to be bound by real world constraints [6]. This new functionality: (1) used visual feedback to provide 'awareness' about who was receiving audio at a given moment in time and who was speaking, (2) supplemented a participant's own (main) view by small viewports that showed the views of fellow group members, and (3) implemented teleporting so participants could move directly to any point in the environment by clicking on it ('walking' is time consuming).

The basic MGDs techniques incorporated an explicit hierarchical grouping system, represented using a 'group graph' metaphor, and methods to assist movement as a group. The awareness functionality (see Fig. 1) used a head-up display (HUD) to display the faces of all participants who were within hearing range of you at a given moment in time (this included all participants in one's own group because there was no distance attenuation for intragroup audio communication). These faces were photographs of the participants (extracted from their photographic avatars), so they could be easily recognized. This was designed to make the participant aware that they could be heard by all the participants shown on their HUD, even if some of them were fellow group members whose avatars were a considerable distance away. When another person was talking, their face was highlighted on the

HUD, with a speech icon next to it. This gave participants additional information as to who was speaking, which was particularly useful if the associated avatar was out of sight.

In VEs, users experience two kinds of problems understanding the actions of others. (1) 'Fragmented views', where another participant refers to an object or point of interest in the environment, but their avatar and the point of interest are not simultaneously visible in the viewport [2]. (2) What you see is *not* what I see, which makes it difficult to understand another's perspective. A combination of these two problems occurs if two users wish to meet at a point of interest. This is a 'Come here! Look at this' scenario (see [7, p. 136]), where the respondent needs to know the location of the user who is talking (they are unlikely to be within the viewport, see problem 1), and what they are referring to (problem 2).

To overcome these problems, Wössner et al. [8] provided a 'what you see is what I see' (WYSIWIS) view in their CVE, which would eradicate problem 2. They designed two CVE interfaces, one of which provided a master/slave style view (where one participant had complete control), and the other which provided a more flexible approach where participants still had some independence (they could change orientation). However, it was found that users preferred the independent viewpoint, so they did not interfere with the other participant. Sonnenwald et al. [9] found that users saw a benefit in both independent views and shared perspectives—users liked to be able to figure things out on their own and then discuss them collaboratively. Therefore, we provided each participant with a main window (their own view of the world) and thumbnails showing the view of each of their fellow group members (see Fig. 1).

The teleporting was implemented as rapid but visually continuous movement, rather than a sudden 'jump' to the new location. This was to help prevent disorientation associated with an instantaneous change of location [10]. The teleporting algorithm took its inspiration from [11], with the addition of gradual acceleration as well as deceleration, and to avoid problems caused by traveling through walls and hedges, raised a participant to a birds-eye view so they could clearly see where they were being taken. Teleporting was achieved either by clicking on a particular place in the VE scene, or on a fellow group member's thumbnail view (this teleported you to be next to that person). Our hypothesis was that the teleporting, awareness and multiple views functionality would improve teamwork. To analyze teamwork, we looked for improvements in two specific areas. First, we wanted to tackle problems of participants spending time collocating to communicate (or waiting until they are collocated before they talk to each other). Second, we wanted to help people work as a team by providing an awareness of the actions and perspectives of others (multiple views tackling problems 1 and 2). These were analyzed using the quantitative data provided by the server's log of activity, and a conversation transcript.

3. Methods for VT collaboration

Traditional CVEs bring together people who are physically remote, and adding VT makes it easier for people to collaborate even if they are not in the CVE at the same time. In other words, combining VT with a CVE allows asynchronous, remote collaboration. There are few examples of VT being implemented in CVEs, but exceptions are 'temporal links' to playback recorded content (e.g., 3D flashbacks to tell a story), which in some cases was activated by a production crew working behind the scenes [12], and in a second example the links were represented as virtual objects that a user could interact with to playback a recording or send messages to other users [13].
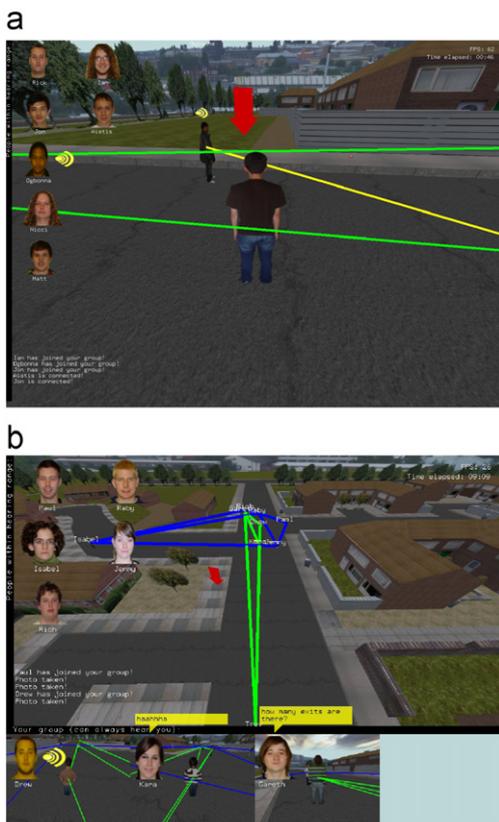
Fig. 1. Screenshots of the environment in the two conditions: teleport and multiple views. The graph metaphor, speech icon, teleporting arrow, and participants within hearing range can be seen in both figures. The views of fellow group members can be seen along the bottom of the screen in (b). (a) Teleport condition, shown using an over-the-shoulder view. (b) Multiple views condition, shown using a bird's-eye view.

We consider a spoken or written utterance to be the basic unit of collaboration, a basic VT system would just contain what was said in that CVE, but nothing about who said them, what they were talking about, where they were in the CVE, or when. At the other extreme, a sophisticated VT system would allow you to travel through a virtual world, walking with people who had been there in the past, chipping in to their conversations as if they were still there, to the extent that an observer who came along later still would be unable to determine who were the original inhabitants vs. who was the impostor who has been added later?

Analysis of these examples highlights a rich complexity of possible functionality for VT. Therefore, the following sections present a framework of VT, and then describe the practicalities of implementation.

3.1. A framework for VT

Given that utterances are the basic building blocks of collaboration and communication in virtual worlds, then a key challenge for VT is determining how those utterances should be organized and associated. The primary methods for doing this are in terms of: (a) people, (b) time, (c) space and (d) topic. Each method has several levels (see Table 1), which can provide context to help us understand the meaning of what was said, influence where we choose to go next in the CVE, and help users control the number of utterances that are visible/audible at any given time so the VT system is scalable.

Adding people's identity to the utterances in a VT CVE allows users to discriminate everything that was said by a particular person, for example, someone who provided particularly insightful comments. Allowing people to choose their virtual appearance will have other effects on whose utterances a given user chooses to listen to, as found in real-time collaborative worlds [14].

Statistics terminology is adopted for the levels of time. At the nominal level, a future user would have no clue as to when, or in what order, different utterances were spoken. Ordinal information would allow utterances to be listened to in the sequence that they originally occurred, and the time interval (either absolute or rebased to when the speaker entered the environment) would allow sets of utterances that took place in quick succession to be distinguished from those that were separated by a lengthy delay.

Indicating the point in space where each utterance was spoken would help a future listener understand what was being talked about, and reduce the need for users in CVEs to devote much more effort to making the 'implicit explicit' than is the norm in real life [2,3]. Linking utterances by the path the speaker had taken would provide the listener with even more information about the things the speaker had seen and which led them to a particular conclusion.

Organizing and associating utterances can be done according to certain topics (or subtopics), for example defining whether greetings were due to users meeting or departing, whether idle chat was humorous or not, or which part of a task a given conversation was based on. Natural language processing (NLP)

algorithms could be used to process utterances into topics, to which user-supplied quality ratings could be added by borrowing techniques from recommender systems and search engines.

Finally, there are many possible combinations of the above methods. For example, combining point (space) and topic (content) would allow the main items of interest in a given area to be quickly determined, adding interval (time) to identity (people) would allow the conversations of a group of people to be followed, and adding path (space) to interval/identity would help a future listener comprehend the bigger picture of a conversation that took place after a group of people had split up to explore an area and then regrouped to discuss their findings.

3.2. Implementing VT

For the evaluation described in Experiment 2, we implemented VT using level 3 utterance association for people (appearance) and level 2 association for time (ordinal), space (point) and content (topic) (see Table 1). Details of the implementation are as follows.

First, all of the utterances from two previous studies of synchronous teamwork in CVEs [4,5] were divided into blocks of communication and categorized as either task-specific or not. The latter were discarded to avoid cluttering the CVE with irrelevant utterances (e.g., idle chat). The task-specific blocks were classified using keywords from the 13 questions on the urban planning report that participants were asked to complete (see Section 4) and, although this was performed manually in the present study, it could have been done using NLP techniques.

The classification used a two level hierarchy, with the 13 questions (subtopics) grouped according to three topics (permeability, character, and safety and security) that were used on the urban planning report. The topics were rendered with different hues (yellow, cyan and magenta), using a different lightness for each subtopic. In addition, tags flashed when they were being played, and were visually caged in black stripes when they had been viewed. A color-coded checkbox was provided for users to choose which utterance subtopic(s) were displayed (see Fig. 2), allowing related comments to be identified even if they are separated in space and time.

The point where each utterance commenced was represented with a hemispherical visual object known as a tag, which put the conversations into context by showing where they took place. To reduce clutter, all utterances in a given block that were within line

Table 1
Four methods for organizing/associating utterances.

Method	Level 1	Level 2	Level 3
(a) People	Anonymous	Identity	Appearance
(b) Time	Nominal	Ordinal	Interval
(c) Space	Amorphous	Point	Path
(d) Content	Undefined	Topic	Quality

Level 1 corresponds to a basic VT system, with Levels 2 and 3 providing ever richer possibilities for virtual time.

Fig. 2. All of the utterance tags used in the VT study (Experiment 2). The 'Audio tag filters' was a list of checkboxes, shown in the top right hand corner of the screen, that allowed the utterances associated with each of the 13 questions to be toggled on/off. Utterances that a pair of participants had already listened to were visually caged in black stripes (see 'Exit points' on left-hand side of the image).

of sight of each other were represented by a single tag that was at the mean position of the individual utterance tags.

The system was tested using a pilot study and refined in response to participants' feedback. The main improvements were:

- Making tag selection explicit (instead of walking into a tag to play it, the users wanted to be able to select a tag with the mouse).
- Allowing users to pause/resume/stop tags, instead of always playing the whole of a tag.
- Providing more time for the task than was allowed in Experiment 1, because there were a lot of recorded utterances that the users wanted to watch and listen to.

Due to the exploratory nature of this work, it was difficult to generate meaningful hypotheses for the way participants would use the VT system and the changes it would make to their behavior. Experiment 2 was a first step in analyzing asynchronous collaborations in CVEs, and the results will help us make more informed design decisions and predictions in future work.

4. Experiment 1: real-time collaboration

The experiment used the context of urban planning, with participants asked to use a CVE to review the design of a new housing estate. Participants were run in two batches. In the first of these (the teleporting condition), participants had all the basic MGD functionality from [4], and new MGD functionality to provide awareness of who was talking, who was within hearing range and teleporting. Participants in the second batch were provided with multiple views (the multiple views condition), in addition to all the MGD functionality that was provided to the other batch of participants.

4.1. Method

The experiment took place in an undergraduate computing laboratory. Each participant was provided with a headset, and they were spread out across the laboratory so they could only communicate using audio and text communication from within the environment. Participants used two adjacent computers, one for the CVE and the other for the urban planning report write-up. The CVE application, environment and experimental procedure were the same as in [4].

4.1.1. Participants

All participants were undergraduate students from the School of Computing, who had not taken part in the previous study. Eight participants were recruited for each run, but one participant in the teleporting condition was unavailable on the day of the experiment. The remaining seven participants in the teleporting condition (6 men and 1 woman) had a mean age of 21.7 ($SD = 5.2$). The eight participants in the multiple views condition (5 men and 3 women) had a mean age of 21.8 ($SD = 4.1$).

All the participants volunteered for the experiment, gave informed consent and were paid an honorarium for their participation.

4.1.2. CVE application

The software application and 3D sound model are described in the previous study [4].

Distance attenuation was turned off for communications between members of the same group. This was clarified by displaying photographs of the faces of participants who would
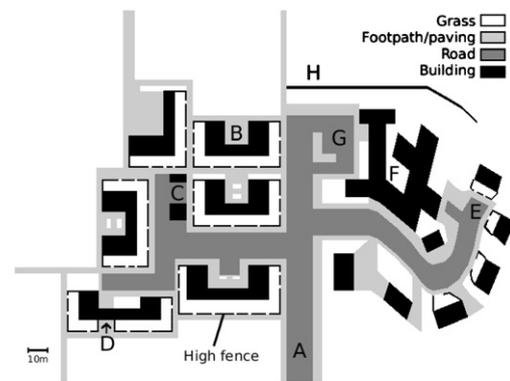
Fig. 3. A map of the estate. The estate had an entrance road in the middle (point A), which acted as a dividing line between two styles of building. On the left-hand side of the entrance road, there were brown-bricked terraced houses, which were mostly horse-shoe shapes creating partially enclosed private space (e.g. point B). The front gardens were bordered by high fences, and there were six garages in the road (C). There was an archway under one of the terraces (D). On the right-hand side of the entrance road there were red-bricked bungalows (single story buildings) along the edge of the curved road, with gardens bordered by low brick walls (e.g. E). There was a single-story care home for elderly people (F), with a car park to the left with space for six cars (G), and a hedge-row above it partly separating private land around the care home from public parkland (H).

receive any transmitted audio. These faces were displayed on the HUD, and were added and removed appropriately as participants changed their position in the environment and switched groups. In addition, an icon was placed above a participant's avatar, and by the side of their face on the HUD, when they were talking.

4.1.3. Environment

The environment was a residential estate that was based on a real estate in Leeds. An annotated map of the estate is shown in Fig. 3.

All participants were represented in the environment with a photographic avatar (using four photos: front, back, left and right). Participants were given an over-the-shoulder perspective, with the option of switching to and from a bird's-eye view. An over-the-shoulder perspective meant that participants could see each other relative to their avatar, and be more aware of how others perceived them [15].

4.1.4. User interface

The participants used desktop workstations, and a two-handed control method, with one hand on the keyboard and the other hand on a 3-button mouse. By holding down appropriate arrow keys a participant could move forward/backward/left/right at 6 m/s, and heading and pitch could be changed by moving the mouse. This is a common gaming control method (e.g. [16]).

The 'Insert' key was used to take screenshots, the 'Home' key to toggle between over-the-shoulder and bird's-eye views, and holding down the 'Page Down' key allowed the participant to use voice communication.

Text communication was achieved by simply typing letters or numbers, which were transmitted the moment each was typed, appearing in a speech bubble above the participant's avatar. The text expired after approximately 10 seconds from the moment the enter key was pressed. Each participant was provided with a stereo headset for audio communication. The default recording and playback volumes were automatically set using a shell script.

The basic MGD functionality used three mouse buttons, and the 'Delete' key to move up one level in the group hierarchy. The display had a crosshair in the middle used for selection. Selecting an avatar with the left mouse button formed/joined a group. Selecting the avatar of a fellow group member with the right

mouse button rapidly moved to their location and automatically followed them. Pressing the middle mouse button anywhere moved to the mean location of the group.

Holding down the numpad zero key released the mouse from controlling heading and pitch, and allowed it to control the position of the red teleporting arrow. Once the arrow was positioned in the desired location, a left mouse click teleported the participant there.

Participants in the multiple views condition could position the teleporting arrow over one of their group members views, and clicking the left mouse button would teleport them to that group member's location. By default, the participant's subsequent movements were tethered to that group member (the automatic following functionality in basic MGDs) but the participant could 'free' themselves simply by pressing a movement key.

The multiple views took up the bottom quarter of the screen. A limit was imposed of three views, each taking up a quarter of the horizontal space, with the remaining quarter reserved for displaying the faces of any other group members. These could be selected using the numpad zero key to release the mouse pointer. Selecting them showed their view in one of the existing viewports, swapping out the member whose view had been replaced.

4.1.5. Procedure

A 10 minute meeting was held with participants a few days before the experiment. They received a verbal explanation of the experiment, a single-sided A4 sheet containing extracts from UK urban planning guidelines and a consent form. They also had photos taken for their avatar during this time.

The experiment itself lasted one hour. At the start participants were provided with another copy of the urban planning guidelines sheet, an instruction sheet for using the CVE, an experiment schedule, and an electronic copy of an urban planning report which they had to complete during the experiment. The report contained the following questions, which participants were asked to illustrate using screenshots:

- *Question* 1, *Permeability*: (a) How many entrance and exit points are there around the estate? What are these for (i.e. cars or pedestrians)? (b) What reduces the speed/volume of traffic? (c) Are there suitable pedestrian routes around the environment? (d) Are the blocks small enough or do you have to walk too far before you reach a choice of direction?
- *Question* 2, *Character*: (a) Which parts of the environment follow the same pattern/building structure? (b) Find a part of the environment that is not consistent with the layout of the estate. (c) Is this acceptable or should it be changed? (d) Does the estate have character?
- *Question* 3, *Safety and security*: (a) Comment on the safety and security of the estate based on your own thoughts, the information in the guidelines and your discussion with other participants. (b) Find examples of where public and private space is clearly distinguished and where it is not. (c) Discuss which part(s) of the estate you think are least safe. (d) Can you find any blank walls that you think should be overlooked to improve the feeling of safety and help prevent graffiti? (e) Try to suggest some improvements with regard to the safety and security of the estate.

5. Results

There were two types of work that took place in the experiment: taskwork and teamwork [17]. Taskwork refers to the answers given in participants' reports, whereas data about teamwork were provided by the server's log of the movements, communication, and groups that participants formed.

The urban planning reports were marked like an exam. Participants names were on the reports, marking was not blind. An independent samples *t*-test showed no significant difference between the teleporting and multiple views conditions, $t(13) = 1.49, p = 0.16$. Participants in the teleport condition had a mean mark of 18.7 ($SD = 3.3$) out of 24, and 16.3 ($SD = 3.1$) in the multiple views condition. Our focus, however, was on how participants went about doing the task (i.e. the teamwork), and how different MGD functionality affected participants' behavior.

For each batch of participants, the spoken and text communication was transcribed and analyzed using a communication coding approach [18] to classify each utterance as one of the following:

(a) *Greetings* (e.g. 'Hey *M*!', 'Hi *G*!').
(b) *Functionality*—communication regarding the system and the groups (e.g. 'Think we need smaller groups than all of us!', 'You do realize that if you just press 'Home' you get a bird's-eye view and it's a lot easier to see!').
(c) *Environment*—discussion about the 3D world, but not in relation to the task (e.g. 'I swear you should be able to see uni from here.', 'I kind of might have figured out where the pictures were taken of, you know the Leeds skyscrapers ones.').
(d) *Task related* (e.g. 'Which part's the least safe?', 'I'd say where we're stood now, *J*.').
(e) *Idle chat* (e.g. 'Party at my flat. Come on, let's go!').

These data were analyzed in terms of the quantity of communication that took place, and where participants were relative to each other when they communicated. For comparison, data are provided from a previous study [4] when other participants had performed the same urban planning task either in a conventional CVE ('control' in Fig. 4) or with basic MGDs functionality (see Figs. 4 and 5). Note that the average group size in the basic MGDs, teleport and multiple view conditions was 3.5, 2.5, and 3.0, respectively.

The total number of utterances made by participants in the basic MGDs (data from [4]), teleport and multiple views conditions (data from the present study) was analyzed using a univariate analysis of variance (ANOVA). This showed that there
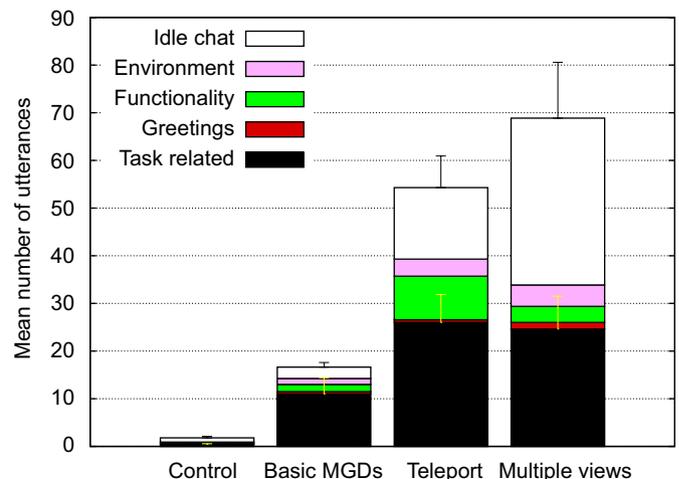
Fig. 4. Mean number of utterances made by the participants in each condition. The control and basic MGDs conditions are from [4]. The error bars are shown for task related utterances and idle chat.
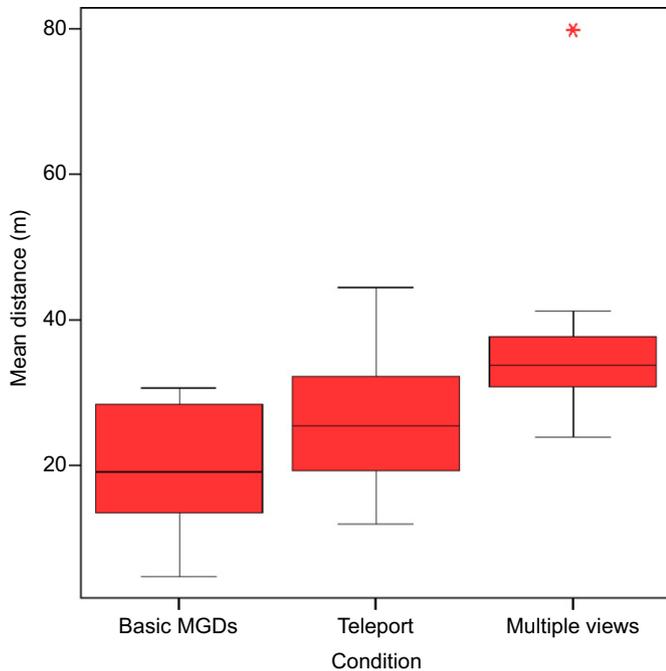
Fig. 5. Mean distance to the nearest group member at the time of each participant's utterances. The basic MGDs condition was from [4].

was a significant difference between the conditions, $F(2, 20) = 3.91, p = 0.04$. Tukey HSD posthoc tests showed that the difference between basic MGDs and multiple views was significant ($p = 0.03$) but the other pairwise comparisons were not. The mean amount of communication increased by 226% from the basic MGDs to the teleport condition, and by another 27% from the teleport to the multiple views condition. Within this, task related communication increased by a factor of 2 from basic MGDs to the teleport and multiple views conditions, but this was not significant. Idle chat more than doubled from the teleport to the multiple views condition (see Fig. 4).

One of the limitations identified in our previous research was that participants tended to assemble in one place in the CVE before communicating, even though this was unnecessary with the basic MGDs functionality that was provided (see Introduction). To determine whether the new functionality provided in the present study overcame this limitation, each time a participant made an utterance the distance to their nearest group member was calculated, and the mean for each participant in the basic MGDs, teleport and multiple views conditions was analyzed using a univariate ANOVA. The two participants who did not speak at all during the experiment were excluded from the analysis, one was from the basic MGDs condition and the other was from the multiple views condition. The ANOVA showed that there was a significant difference between the conditions, $F(2, 18) = 3.56$, $p = 0.05$. Tukey HSD posthoc tests showed that the difference between basic MGDs and multiple views was significant ($p = 0.04$) but the other pairwise comparisons were not (see Fig. 5).

6. Discussion

It is well known that in conventional CVEs users often have difficulty understanding the context of what each other is talking about (see problems 1 and 2 in Section 2). Our previous research into MGDs [4] showed how a group graph metaphor could help users find each other, since the graph 'tracked' participants and the nodes corresponded to avatars, with edges denoting group

membership (see Fig. 1). One could find a group member by following a line from their avatar until they reached a node. In Experiment 1 of the present study, the teleporting and multiple views functionality took this a step further. It allowed participants to teleport directly to a group member of their choice by selecting the appropriate viewport, and providing participants with multiple views specifically tackled the problem of understanding another's perspective.

One of the places where the original MGDs techniques fell short of their goals was in facilitating communication when participants were spatially separated within the environment [4]. The fact that participants tended to collocate to communicate in the basic MGDs condition was a sign of inefficient group-work—participants were either taking time to collocate when they wanted to communicate, or they were waiting until they were coincidentally collocated before they said anything.

The present study indicates that by providing feedback to the participants, they became more aware of how the system works and communicated across greater distances than in the basic MGDs condition (see Section 5). The interesting thing about this feedback from the system is it is not specifically new functionality in the sense of a new tool at the users' disposal, like teleporting and multiple views are (in the previous research it was possible for participants to communicate at a distance, because distance attenuation was turned off for within-group audio communication). Instead it boosts *awareness* of existing functionality. As Schroeder et al. reflect, do we improve usability 'by means of improving the systems and features of the environment, or by improving the users' awareness of their activities and settings?' [19, p. 666].

Finally, in previous research, participants communicated a great deal to overcome the lack of sensory information that CVEs provided [2,3]. By contrast, in the present study substantially more communication took place when extra sensory information was provided (e.g. awareness of who could hear you and who was speaking, and multiple views providing an 'extra pair of eyes'). This increase in communication was indicative of more teamwork taking place.

7. Experiment 2: VT collaboration

The second experiment used the same urban planning context and the same environment as Experiment 1. Participants were run in pairs, and had access to all the task-related conversations of 23 people who had previously done the same task in the environment (the 15 participants of Experiment 1, and the 8 MGDs participants from [4]). These previous conversations were embedded in the environment using the tags described in Section 3.2.

7.1. Method

The method was similar to that of Experiment 1. A total of 10 participants (5 pairs) took part. There were 7 men and 3 women, with a mean age of 22.2 ($SD = 3.3$). They had not taken part in any of the previous studies. Each pair communicated with each other and had access to the task-related conversations of 23 participants who had done the task in previous real-time experiments.

After pilot testing, the total time for Experiment 2 was increased to 90 minutes from the 60 minutes used in Experiment 1. The time in the training environment was extended from 15 to 30 minutes, and the time in the residential environment from 30 to 45 minutes. The final 15 minutes were allocated for a semi-structured interview.

The grouping interface and functionality were identical to that of the multiple views condition in Experiment 1, however, each pair was placed into a group together and could not join the groups of participants from the past. Participants could select a conversation tag by positioning the crosshair with the mouse and pressing the left mouse button. Participants could stop conversation tags by pressing the 'Escape' key, and pause/resume the playback with the 'F1' key.

7.2. Results

As in Experiment 1, the urban planning reports were marked like an exam, and participants had a mean mark of 17.9 out of 24 ($SD = 3.9$). An independent samples t-test showed no significant difference between the teleporting and multiple views conditions, $t(16) = 0.978, p = 0.343$. However, our main interest lay in how participants used VT and the effect it had on their behavior in the CVE. To investigate this, participants' communication, movement and tag usage were analyzed. Statistical comparisons were made with the multiple views condition from Experiment 1, whose interface was the same as the one used in Experiment 2 except for the VT functionality. In considering the findings, readers should bear in mind obvious differences between the experiments (especially the number of live participants at any given time), which could have affected the results.

Each participant's rate of communication was calculated by dividing the number of utterances they made by the time they spent working in the CVE (the difference between the time they first and last moved). This took account of the extra time allowed for Experiment 2 as a whole (45 vs. 30 minutes) and the fact that some participants remained 'in' the CVE (but not moving) while they finished writing their report.

Fig. 6 shows the mean rate of communication for task-related and non-task-related utterances for the multiple views (Experiment 1) and VT (Experiment 2) groups. An independent samples t-test was carried out on the task-related communication, and showed that the difference between the groups was significant, $t(16) = 3.258, p = 0.005$.

The distances that participants covered as they walked and teleported around the environment were calculated from the server log. The rate of travel was calculated by dividing the distance each participant traveled by the time they spent working in the CVE (calculated as above). Independent samples t-tests showed that the multiple views group walked significantly further in unit time than the VT group, $t(16) = 2.790$, $p = 0.013$ (Fig. 7).
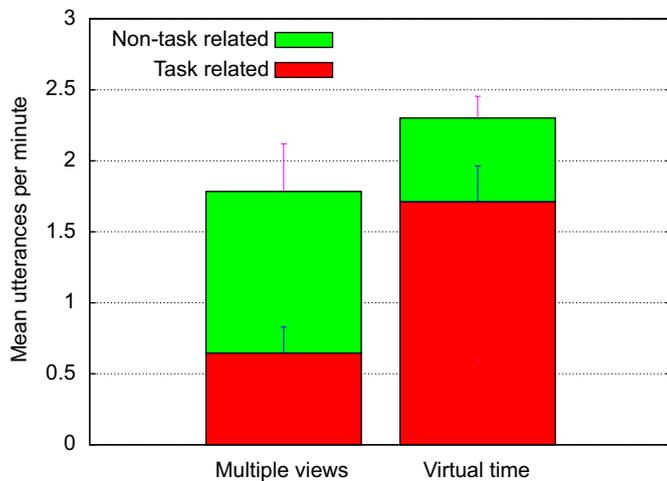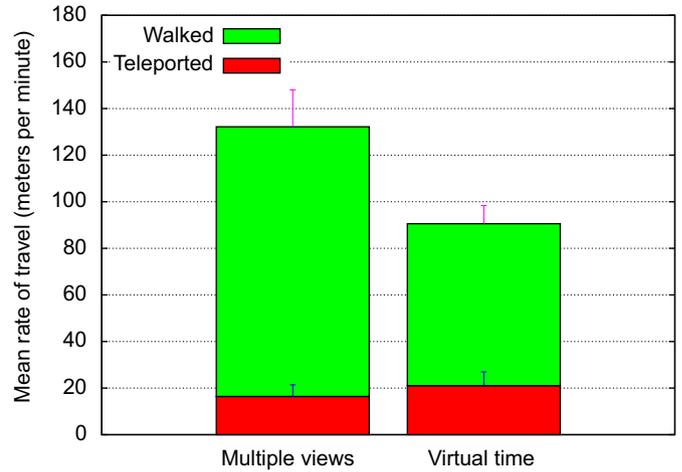

Fig. 7. The mean rate of travel for the multiple views and virtual time conditions.
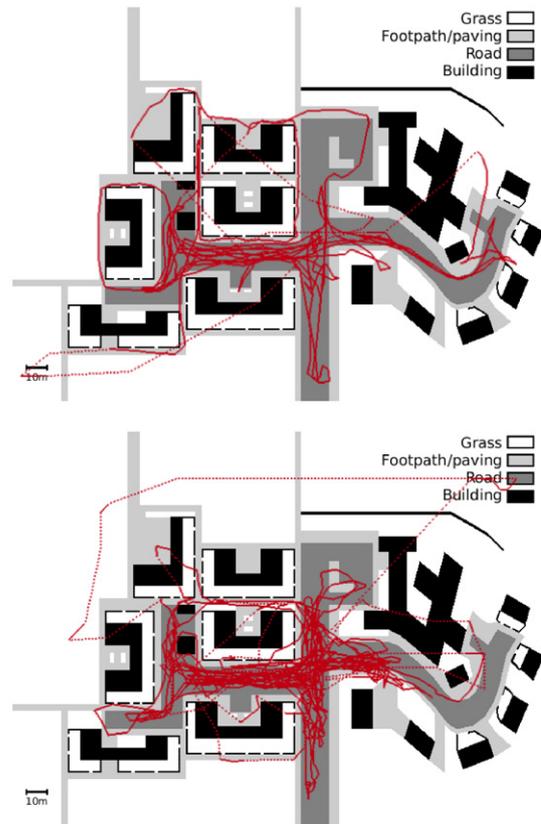

Fig. 8. The paths of two participants who had the median percentage distance teleported. The solid lines represent walking and the dotted lines represent teleporting.

The difference in rate of travel for teleporting was not significant, $t(16) = 0.578$, $p = 0.571$.

The paths of two participants were plotted as a qualitative analysis of the usage of teleporting. The two participants chosen had the median percentage distance teleported. The paths are shown in Figs. 8(a) and (b). The paths show examples of how teleporting was used: to cover large distances and to 'jump' over buildings.

There were a total of 67 tags containing 'VT' communication. All of participants' usage of the tags, and the tag filter menu, was recorded in the server log. Analysis of this log showed that pairs of participants typically selected one subtopic in the menu at a time,

Fig. 6. The mean utterances per minute for the multiple views and virtual time conditions.

Table 2
The mean number and SD of utterance tags in each subtopic that were played by each pair of participants.

Topic	Subtopic	Total tags	Mean tags played	SD	%
Permeability	Exit points	18	4.80	3.49	26.7
	Traffic	10	4.20	2.39	42.0
	Pedestrians	0	0.00	0.00	0.0
	Block size	4	2.00	1.58	50.0
Character	Buildings	3	1.40	0.89	46.7
	Inconsistent	5	0.80	0.84	16.0
	Acceptable?	0	0	0	0.0
	Character	4	2.00	1.41	50.0
Safety and security	Security	10	1.40	1.52	14.0
	Public/private	3	1.00	1.22	33.3
	Least safe	2	0.40	0.55	20.0
	Graffiti	4	0.80	1.30	20.0
	Improvements	4	1.20	1.30	30.0

Fig. 9. The environment with the utterance tags. The tags are colored based on how many pairs of participants played them: white = never played, blue = played by the minority of pairs, and red = played by the majority (3+).

allowing participants to focus on VT utterances that were relevant to the question being answered at a given time, and went through the tags in a logical order (subtopic by subtopic, matching the order in which the questions appeared in the urban planning report).

The mean number of tags played by each pair was 20.0 ($SD = 6.6$), and the breakdown by (sub)topic is presented in Table 2 (note: the playback was shared across the network, so both participants in a pair heard the same utterances). The distribution of the utterance tags and the frequency with which each was played is represented in Fig. 9.

7.3. Discussion

In this study, VT was implemented via a system of conversation tags (Sections 3.1 and 3.2) so participants could take advantage of the comments their predecessors had made. Participants used the environment in pairs, so each had one other real-time collaborator to communicate with and the conversations of 23 previous inhabitants to listen to. The results showed that the VT system led to a significant increase in task related communication between the 'live' pair when compared to the

same interface without the VT (the multiple views condition). Furthermore, the results showed significantly less travel around the environment.

The reduced travel suggests that the points of interest were found by watching and listening to the conversations from the past: the usage results showed that on average a quarter of the conversation tags available for each subtopic were played. There is one drawback to this, however, as identified from the participant's comments in the semi-structured interviews. They were concerned about the quality of content of the tags, since they were using the information from past participants to perform the task. Assuming the past work was thorough and correct, VT provides a large pool of ideas to be shared from one group of participants to the next, thus focusing their conversation on the task. This highlights the importance of a measure of quality of content, level 3 in the VT framework, Section 3.1.

8. Conclusions

Our goal was to develop techniques to support synchronous and asynchronous collaborations in large-scale CVEs. For synchronous collaboration we identified problems in conventional real-time CVEs and built upon our existing MGDs functionality by adding teleporting and multiple view conditions. For asynchronous collaboration we presented a framework for VT and used the data from the real-time studies to implement VT, so live participants could benefit from their predecessors' comments.

The results from the real-time experiment (Experiment 1) showed that the awareness MGDs functionality produced a significant increase in communication, and an increase in the distance over which participants communicated, making participants behavior more efficient. In the VT experiment, participants listened to an average of 27% of their predecessors' conversations, spent more time talking about the task themselves and traveled significantly less. In other words, VT stimulated communication between the live participants.

Finally, future developments of VT systems could take many directions within the various levels of associations identified in Table 1. In particular, research is needed into the issue of quality of content of the conversation tags (level 3 of content). For example, tags could be user-generated, so that participants could explicitly request that particular conversations be recorded and tagged, a user-rating system would allow participants playing back the conversations to contribute to the quality control, and methods from recommender systems (e.g., Amazon's 'people who bought X also bought Y') could be incorporated.

References

[1] Johansen R. GroupWare: computer support for business teams. New York, NY, USA: The Free Press; 1988.
[2] Hindmarsh J, Fraser M, Heath C, Benford S, Greenhalgh C. Object-focused interaction in collaborative virtual environments. ACM Transactions on Computer–Human Interaction 2000;7(4):477–509.
[3] Ruddle RA, Savage JCD, Jones DM. Verbal communication during cooperative object manipulation. In: Proceedings of the 4th international conference on collaborative virtual environments (CVE '02), Bonn, Germany, September 30–October 2 2002. ACM. p. 120–7.
[4] Dodds TJ, Ruddle RA. Mobile group dynamics in large-scale collaborative virtual environments. In: Proceedings of IEEE virtual reality (VR '08), Reno, Nevada, USA, March 8–12 2008. p. 59–66.

[5] Dodds TJ, Ruddle RA. Using teleporting, awareness and multiple views to improve teamwork in collaborative virtual environments. In: Proceedings of the 14th eurographics symposium on virtual environments (EGVE '08), Eindhoven, The Netherlands, May 29–30 2008. p. 81–8.

[6] Pekkola S. Critical approach to 3D virtual realities for group work. In: Proceedings of the second nordic conference on human–computer interaction (NordCHI '02), Denmark, October 2002. ACM. p. 129–38.

[7] Yang H, Olson GM. Exploring collaborative navigation: the effect of perspectives on group performance. In: Proceedings of the 4th international conference on collaborative virtual environments (CVE '02), Bonn, Germany, September 30–October 2 2002. ACM. p. 135–42.

[8] Wössner U, Schulze JP, Walz SP, Lang U. Evaluation of a collaborative volume rendering application in a distributed virtual environment. In: Proceedings of the workshop on virtual environments 2002 (EGVE '02), Barcelona, Spain, May 30–31 2002. The Eurographics Association. p. 113–221.

[9] Sonnenwald DH, Whitton MC, Maglaughlin KL. Evaluating a scientific collaboratory: results of a controlled experiment. ACM Transactions on Computer–Human Interaction 2003;10(2):150–76.

[10] Ruddle RA, Howes A, Payne SJ, Jones DM. The effects of hyperlinks on navigation in virtual environments. International Journal of Human–Computer Studies 2000;53(4):551–81.

[11] Mackinlay JD, Card SK, Robertson GG. Rapid controlled movement through a virtual 3D workspace. Computer Graphics 1990;24(4):171–6.

[12] Benford S, Fraser M, Reynard G, Koleva B, Drozd A. Staging and evaluating public performances as an approach to CVE research. In: Proceedings of the 4th international conference on collaborative virtual environments (CVE '02), Bonn, Germany, September 30–October 2 2002. ACM. p. 80–7.

[13] Greenhalgh C, Flintham M, Purbrick J, Benford S. Applications of temporal links: recording and replaying virtual environments. In: Proceedings of IEEE virtual reality (VR '02), Orlando, FL, USA, March 24–28 2002. p. 101–8.

[14] Schroeder R. Social interaction in virtual environments: key issues, common themes, and a framework for research. In: Schroeder R, editor. The social life of avatars: presence and interaction in shared virtual environments. London: Springer; 2002. p. 1–18.

[15] Cheng L, Farnham S, Stone L. Lessons learned: building and deploying shared virtual environments. In: Schroeder R, editor. The social life of avatars: presence and interaction in shared virtual environments. London: Springer; 2002. p. 90–111.

[16] Brown B, Bell M. CSCW at play: 'There' as a collaborative virtual environment. In: Proceedings of the 2004 ACM conference on computer supported cooperative work (CSCW '04), Chicago, Illinois, USA, November 2004. ACM. p. 350–9.

[17] Baker K, Greenberg S, Gutwin C. Empirical development of a heuristic evaluation methodology for shared workspace groupware. In: Proceedings of the 2002 ACM conference on computer supported cooperative work (CSCW '02), New Orleans, Louisiana, USA, November 16–20 2002. ACM. p. 96–105.

[18] Bowers CA, Jentsch F, Salas E, Braun CC. Analyzing communication sequences for team training needs assessment. Human Factors 1998;40(4):672–9.

[19] Schroeder R, Heldal I, Tromp J. The usability of collaborative virtual environments and methods for the analysis of interaction. In: Presence: teleoperators and virtual environments, vol. 15 (6). Cambridge, MA: MIT Press; 2006. p. 655–67.