Evaluation of Mixed-Space Collaboration

Raphaël Grasset, Philip Lamb, Mark Billinghurst
HIT Lab NZ
University of Canterbury, New Zealand
{*Raphael.Grasset*|*Philip.Lamb*|*Mark.Billinghurst*}*@hitlabnz.org*

Abstract

Recently Augmented Reality (AR) technology has been used to develop the next generation collaborative interfaces. First results have shown the value of using AR for co-located tasks based on exocentric viewpoints. In contrast, Virtual Reality (VR) seems to offer interesting advantages for immersive collaborative experiences with egocentric viewpoints.

In this paper we focus on a new area: a mixed collaboration between AR and VR environments. We present a new conceptual model of transitional interfaces that allow users to move between AR and VR viewpoints. We then describe the results of a quantitative evaluation with an AR exocentric viewpoint and a VR egocentric viewpoint for a navigational task. We also conducted a second experiment on the impact of the relationship between the interaction and visualization space in mixed collaboration. Results of these studies can provide a better understanding of how to design interfaces for multispace and transitional collaboration.

1. Introduction

Recently, The Magic Book project [4] introduced the concept of a *transitional augmented reality interface*. Based on a physical book, the interface can be used to seamlessly move between the real world, the virtual world or an AR world (along the continuum defined by Milgram [19]).

In the MagicBook interface a dedicated hand-held device provides an augmented view and also supports navigation in the virtual environment (VE). The MagicBook also explores a new metaphor for multi-viewpoint collaboration. For example, two scientists can use the MagicBook interface to share an AR view of a virtual 3D molecular simulation superimposed on a real table between them. In this case they both have an exocentric viewpoint. At any time, either person can switch the context, and can immerse themselves inside the virtual molecule for observing more precisely its internal structure. In this way the MagicBook can aid collaboration by providing multiple viewpoints to improve the understanding of a task ([28]).

In informal studies users have appreciated the transitional concept, but a more quantitative evaluation still needs to be done to evaluate the collaborative potential of this technology. Therefore, in our work we are interested in studying the advantages of a transitional collaboration. More precisely, in this paper we evaluate a collaborative application where users are in different spaces; a *mixed-space collaboration*.

The next section will present related research, followed by a formal description of the transitional interface concept. Then we present the first experiment followed by the second experiment. Finally we will finish by a discussion of the results and the conclusion.

2. Related Work

Providing representation and viewpoint specification to allow efficient explorations of data is an old problem in computer human interaction research ([8]). For example, Ainsworth studied the function of multiple representations and proposed a taxonomy in [1]. Complementary roles, constrained interpretation and construction of deeper understanding are the main categories he defined. In the case of 3D applications, the user view is generally related to the spatial notion of *frame of reference*, i.e. choosing an appropriate 3D/6D viewpoint for manipulating the data ([5]).

In the context of virtual reality, Darken et al. [7] demonstrated the value of using a body-user oriented navigation map for a wayfinding task. In the World In Miniature work (WIM) Stoakley [25] extended the limited egocentric viewpoint with a complementary exocentric view. The exocentric view is represented a scale model attached to the hand, and can be manipulated to change the users position in the virtual environment. As Lamb [15] shows, in a teleoperation application, there is a high degree of interaction between spatial properties of the task and the optimal interface conditions. Results demonstrated that an egocentric control frame of reference offers superior performance and an egocentric viewpoint seems to offer a better spatial representation (but is less efficient for task completion time).

For collaborative situations, Salzman [23] conducted a study on the impact of the frame of reference (FOR) in the context of scientific learning. He shows that using complimentary frames of reference enhances the process of understanding abstract information. A group of student demonstrated generally good learning outcomes using a cooperative bicentric viewpoint (exo and egocentric) although there were individual differences between subjects. Leigh et al. [16] have also been interested in ego/exocentric multi-scale collaboration. They introduced the concept of a *mortal*'s view (egocentric) and a *deity*'s view (exocentric) with a dedicated avatar representation in an immersive virtual environment. Yang [28] conducted an experiment with a navigation task using a collaboration between a VR egocentric viewpoint for one user and different VR conditions for the other user. Contrary to expectations, the results show that collaboration between two VR egocentric viewpoints are the best in terms of search and traveling time. Yang felt that this was due to the difficulty of performing mental rotation of the user position during the communication.

To enhance the interaction range, Zhang proposed [29] the notion of a multi-scale Collaborative Virtual Environment (CVE), where a user can control the scale of interaction and by implication the level of collaboration. Based on a searching and traveling task, he demonstrated [30] that multiscale collaboration can improve efficiency. In fact, users can access multiple scales of representation of the task and associated navigation speed.

Few rigorous experiments have been conduct to explore collaborative manipulation. Pinho [21] introduced a collaborative manipulation in a multi-scale environment. Schafer et al. [24] demonstrated that cooperation between exocentric and egocentric views seems efficient, but both users having egocentric viewpoint provides a better feeling of presence for two users with different roles (on a desktop VR setup).

Recently, Kiyokawa [12] has demonstrated that AR is better than VR for a co-located collaborative targeting task. His evaluation was restricted to the case where both users had a similar virtual world representation. He also hypothesizes that VR is probably more efficient in an immersive or egocentric scenario. Billinghurst et al.[3] analyzed the influence of the coupling between the communication and task space on a face to face collaboration task. The authors noticed that even with a reduced field of view due to the technology used, communication behavior patterns were very similar to those used in natural unmediated face to face collaboration.

These results show that the use of different viewpoints and working with a collaborator generally improves efficiency of manipulative or navigational tasks in VR. However, the application needs high-level control and good design of the communication layer ([28]).

Previous works have been largely limited to use of the same spaces for the collaboration, although some more recent works have explored the use of different space for each user. For example, Kiyokawa [13] introduced one of the first collaborative transitional interfaces in his VLEGO system. He used a virtual slider-like tool to smoothly switch from an AR world into a VR world. His interface did not support transition along the entire Mixed-Reality continuum, just between AR and immersive VR. In contrast, Brown et al. [6] describe spatially mixing an AR view (virtual scene on a real table) with a VR view (immersed view on a curved surround screen). Finally, the 3D Live system [22] can support collaboration between a user in AR world seeing a live body capture of the other user (immersed in a VR world) integrated in the virtual scene. However, in none of these projects were a comprehensive user evaluation carried out, nor analysis of the collaboration processes.

One exception is the work of Brown [6], a collaborative experiment where users visited a museum in different environments (immersive VR, a standard desktop interface or a pervasive computing enhanced real location). A qualitative study showed that data coherence, good location awareness and a comprehensive range of different referential controls are essential. On a similar scale, Koleva [14] introduced the concept of a traversable interface, crossing the gap between the real and virtual world. They demonstrate a physical prototype but have not conducted formal studies on users responses to their prototype. Finally, Nakanishi et al. [20] explore the notion of *transcendent communication*, a collaboration between a user in the VR environment with a visitor in the real public space. Based on the idea that a bird's-eye view can help people, two unevaluated prototypes have been demonstrated.

The past research work has implied that using multiple collaboration spaces of different types can improve the collaboration efficiency for a specific task. However this depends on the frame of reference or the scale of the user, and also the type of the space (and by extension the user's representation). We are interested in exploring whether the properties of the space and the transition between different spaces can provide an improvement for the realization of different collaborative tasks.

In this paper, we are interested in a *mixed collaborative setup*, involving collaboration between AR and VR spaces. Our work is the first that provides a rigorous user study that compares collaboration between a user in an AR space and a second user in VR space. This experiment serves two purposes:

- A good understanding of the implications of a collaboration between different spaces is a prerequisite to evaluating and using transitional interfaces. A rigorous study of the employed mechanisms of collaboration and the impact of the chosen space on performance need to be analyzed.
- A large number of applications can directly use the results of this experiment, because they are based on a mixed-space collaboration. For example, firefighters immersed in a stressful situation in a VR training

simulation can benefit from the collaboration with a guided user in another environment. Based on results of Kiyokawa [12], an AR setup can provide a useful element for co-located fire-officers, discussing action options for the immersed firefighers by viewing the model of a 3D building directly on a table.

However, before we describe our experiment further, in the next section we will discuss more precisely the concept of a transitional interface and the conceptual model that we based our experiments on.

3. Conceptual Model

In [14], a schematic representation of traversable interface is proposed. Except for this work, few related models have been introduced for representing a VR/AR mixed or a 'transcendental' collaboration. In this paper, we extend this schema to a more conceptual model for transitional collaboration (Figure 1). We also redefine some terms generally used in papers studying this new kind of collaboration. This model not only furnishes support to describe or classify previous works, but also offers support for the design of transitional collaborative scenarios.

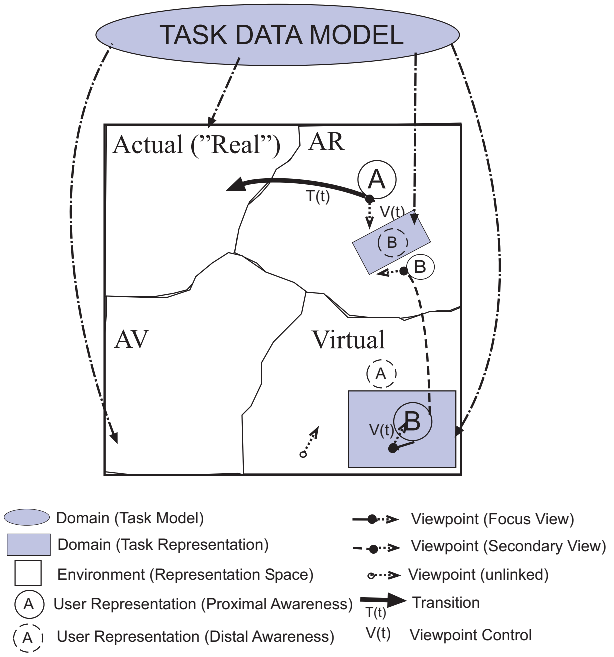

Figure 1. The transitional collaborative model.

The model of a task (*task model*) can be presented in different *environments* along the actual-virtual continuum[1] [19], with a possible dedicated representation (*task representation*) on each one. A *viewpoint* defines a shared or subjective view inside one environment. The relative position in the task representation provides the notion of egocentric or exocentric viewpoint (e.g. Figure 1 user *A* has an exocentric viewpoint). A user can have multiple viewpoints spatially-multiplexed (screen-aligned, WIM, etc.), like user *B* in Figure 1. Each of these viewpoints can dynamically change, controlled by a motion function (*viewpoint control*). In our schema, user *B* has a tethered control mode in the VR environment, and also a static view in the AR environment. We define the *focus view* as the view to which the user is devoting the most attention at a given instant in time. In each environment a user can have a *user representation* (embodiment), used for proprioperception or the body awareness of other users sharing this environment (in our example, user *B* has two user representations since he has two views). A *distal embodiment* is also defined for the awareness support during a collaboration between different spaces (shown as the dashed circle for user *A* for in user *B*'s view in VR environment, and for the one of user *B* for user *A* in AR environment).

A user can have also time-multiplexed viewpoints, the *transition* between them can be defined by a function $T(t)$ (uni-or bidirectional). A transition can be generalized to any movement inside (or between) the environment(s). A transition can be between two existing viewpoints (e.g. unlinked, or another user's viewpoint) or from an existing to a new one (e.g. automatically assigned, user defined). The notion of a *transitional interface* then defines the interface used to support this transition.

In this generic model, we have identified five main collaboration types:

- *Standard collaboration*: Users can cooperate in an environment without transitions, such as a standard VR immersive collaboration in a virtual building.
- *Multi-perspective Collaboration*: The user has access to multiple viewpoints in a collaborative environment (spatially or time-multiplexed).
- *Multi-space collaboration*: Users are in the same environment, and transition simultaneously and together to another environment. For example, navigation as a group between an AR exocentric viewpoint to a VR immersive viewpoint of a building.
- *Mixed-space collaboration*: Each user can work separately in a specific environment. For example, an architect with an AR exocentric viewpoint could be manipulating parameters of a virtual house while their client is immersed in a VR environment inside the house.
- *Transitional collaboration*: this is the generalized case, users can collaborate in the same environment but also switch independently to other environment. This last case is illustrated by the concept used in the MagicBook demonstration.

---

[1] We prefer to reserve the term 'reality' for the ontological level of exploration and so here we will replace Milgram's use of 'real' with 'actual' (c.f. 'virtual').

Notice that the model doesn't define any constraint on the control space or physical setup (device choices, co-located/remote cooperation) or the spatial and semantic properties of the space (e.g. non-linear projective coordinate system). The communication cues provided (audio, visual) and ways of supporting awareness 'feedthrough' of a collaboration are also not represented ([10]). As an example of how this framework can be applied, Figure 2 illustrates our model applied to the MagicBook concept.

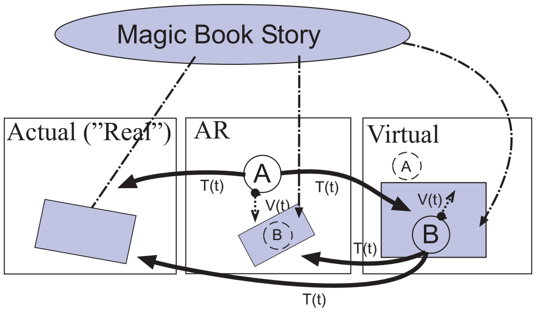

Figure 2. The Magic Book Model

In the MagicBook interface each user has one non-static viewpoint, constrained to be exocentric in the AR environment and egocentric in the VR view. In the actual environment the task representation is defined by the illustration of the real book.

Comparing every different possible type of collaboration in this framework remains out of the scope of this paper. Therefore, we present in the next section a first study domain, mixed-space collaborations.

3.1. Restricted Case: Mixed-Space Collaboration

The case chosen is a collaboration between an AR environment and a VR environment. We are interested to study the efficiency of working together when two users collaborate between AR and VR environments during a specific task. We focus on these major questions:

1. Does a mixed-space collaboration affect the performance on a task that can be completed by a user alone? Does a mixed-space collaboration provide more efficient performance during a task than in a standard collaboration?
2. Does a mixed-space collaboration evoke similar communication processes to a standard collaboration?
3. Is a mixed-space collaboration affected by the type of the environment, the viewpoint used, the control space or the display affordances?
4. Which kind of collaborative applications are best suited to a mixed-space collaboration ?

We expect that with an efficient collaboration process and design ([6]), a mixed-space collaborative task can be completed faster than a standalone one, based on the support for a simultaneous combination of multiple viewpoints (one for each user). In regards to combination of different environments, like Yang [28], we expect that sharing the mental model of the task for the users positioned in different environments offers an efficient way to keep semantic information. But with dedicated representational properties of this environment (AR and VR), we can improve performance on a dedicated task. As demonstrated in [18], [26] and [17], natural proprioperception or haptic feedback can improve a task, elements naturally supported in AR (reproduced virtually in VR). In contrast, as hypothesised by Kiyokawa [12], immersive VR can be interesting for egocentric viewpoint interaction as seeing the surrounding actual environment can be disturbing when navigating inside a virtual world. Thus, complemented by the results of [24], we hope that a combination of an AR exocentric viewpoint and a VR egocentric viewpoint (Figure 3) can be an efficient way to collaborate for a specific task.

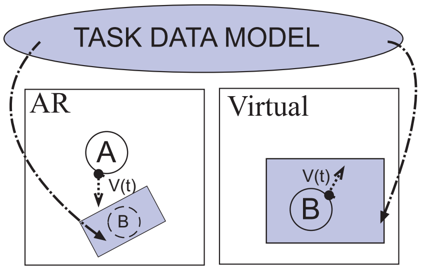

Figure 3. The chosen scenario. The user A with an AR exocentric viewpoint collaborates with user B with a VR egocentric viewpoint.

We hypothesise also that the relationship between interaction and visualization space will impact on the task performance, and the collaboration, as demonstrated on a co-located case in [3]. In summary, we formulate three hypotheses:

Hypothesis 1*: A mixed-space collaboration is more efficient than working individually on a navigation task.*

Hypothesis 2*: The mixed-space scenario of an AR exocentric viewpoint + VR egocentric viewpoint collaboration is more efficient than a standard VR exocentric viewpoint + VR egocentric viewpoint collaboration.*

Hypothesis 3*: The relationship between the visual and control space of the AR exocentric user influences the efficiency of the collaboration.*

Related to this last point, we chose to evaluate the efficiency as proposed in [3] by decomposing on three main categories of measures: performances measures (provided by the task outcome), subjective measures (user opinion, appreciations of the success of the collaboration) and process measures (identified characteristics of the coordination/communication, compared to real situation).

We conducted two experiments, the first, to evaluate *Hyp 1.* and *Hyp 2.*, the second, to evaluate *Hyp 3.*. We present each of these experiments in turn.

4. Experiment 1: Impact of an AR space in a mixed-collaboration

4.1. Task

For our experiment, we chose a navigation context based on a search and wayfinding task. In contrast to [28], we focus on a 2D task but with a 3D spatial representation. The context is a maze, where one user is immersed in the virtual maze and needs to reach a specific location assisted and guided vocally by the other user. To overcome the problem of the cost of negotiation during communication ([30]), we gave roles to each of the users, where the AR user is a director and the VR user is an actor, referenced in the terms proposed by [24]. This dominant position is created by a monodirectional body awareness and a high level control on the task.

4.2. Conditions

We used three conditions for the experiments:

1. *No collaboration*: the immersed user performs the task alone, in a VR egocentric viewpoint.
2. *VR collaboration*: the two users collaborate on the wayfinding task. The first user (in a VR egocentric viewpoint, named *VR Ego User*) is helped by a second user with a VR exocentric viewpoint on the shared virtual world (named *VR Exo User*).
3. *Mixed-space collaboration*: the two users collaborate using different spaces. The first user has a VR egocentric view of the maze, and is helped by a second user with an AR exocentric viewpoint on the scene (named *AR Exo User*).

The *AR Exo User* sees the maze displayed above a physical object (fiducial marker fixed on a table), a reference for the virtual content. The *VR Exo User* sees the maze floating in front of him with a view of a simplified surroundings. These users see an avatar representation of the other user and his or her gaze direction (represented by a red line).

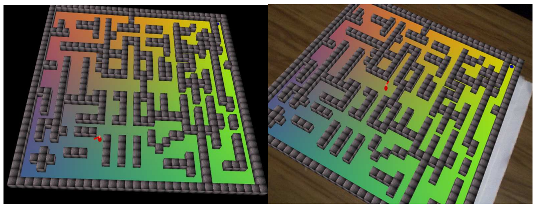

Figure 4. The view of the exocentric user: condition VR (left) and condition AR (right).

The *VR Ego User* in the maze can freely move his or her head, and can advance in the direction of gaze by clicking a handheld input device (mouse). Movement is the discrete steps to one of the eight adjacent spaces. There is no embodiement of the exocentric user in the VR environment. A sample view for the exocentric user is shown in Figure 4 while one for the egocentric user is shown in Figure 5.

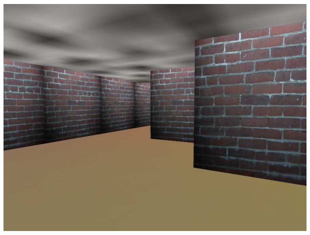

Figure 5. The view of the immersed user.

In designing the maze task, we deliberately chose to limit the visual input for the immersed user to landmark knowledge information ([7]). There were no top-down map or other spatial cues that could be used. The main reason is that use of a map or any survey knowledge information (that might help the immersed user in condition 1) would bias our study about using only one view for the egocentric user. We hope nevertheless to perform comparative evaluations with multiple views in future experiments.

4.3. Metrics

We used objectives performance metrics, subjective metrics and process metrics. The performance measures gauged the effectiveness of the interaction and we used time of completion, length of path traveled by the user, path error[2] and participants head movement/velocity of subjects. For the qualitative subjective evaluation, we used a survey questionnaire (similar to [24], see Appendix 1) during the experience to evaluate the awareness, ease of communication and the task realization. We also recorded video and audio for the subjects working together and used collaborative process cues [3] and identified high level communication processes by video-analysing the data. We focused on a coarse estimation of speech dialogue protocol and head/hand movement of subjects between the different conditions.

4.4. Apparatus

The setup is shown in Figure 6. The two subjects stood in a room separated by a curtain that blocked out their view of each other but allowed audio communication. The *VR Exo User* wore an Olympus Eye-Trek HMD, the *AR Exo User* wore a similar display but with a video camera attached to allow it to be used for video see-through AR. The VR Exo user also had a black sheet attached around the edges of the HMD to block peripheral views of their environment and

---

[2] A path error is defined by passing the same position twice, i.e. cycling path)

provide a more immersive experience. For tracking the head movement of the *Exo User* we used an Ascension Bird magnetic tracker for condition 2 and ARToolKit [11] computer vision based tracking for condition 3. For the *Ego User*, we used Intersense ICube3 to obtain head orientation[3]. The navigation interaction was performed with a wireless Gyration mouse. Both HMDs, camera and tracking devices were plugged into a single computer (Intel P4 3Ghz, NVIDIA GeForce 5900).

The graphics code is a custom OpenGL application with a sufficient number of cues provided to allow the depth perception. The ground is defined by a color gradient (4 colors) giving sufficient information for *Ego User* about his relative position in the maze. The virtual representation of the scene is identical for the two users.

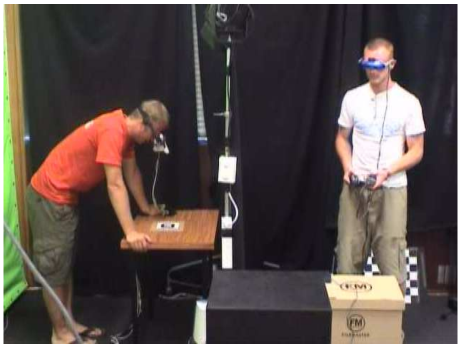

Figure 6. The setup with condition 3: the *Ego User* **(left), the** *AR Exo User* **(right).**

4.5. Procedure

After randomly assigning pairs of participants to the ego and exo roles, we briefly described the task and equipped each user. We demonstrated individually to each user how he or she could visualize and navigate in the maze for each condition. After that, users had one training trial before completing the task. The order that the conditions were performed in were counterbalanced during the experiment.

At the end of each condition, a questionnaire was completed by the subjects and an informal interview was conducted at the end of the experiment. An earlier pilot study indicated that participants generally took no more than 5 minutes to complete the task.

Eighteen people participated to the experiment (9 pairs of 2, 8 male pairs and 1 female pair). Participants ranged in age from 22 to 29. Sixteen were right handed.

4.6. Results

The results of this first experiment support our first hypothesis but do not support the second.

Performances Measures

In each condition we measured the time it took to complete the maze. There was a significant effect of condition on time to completion, as indicated by a one-way ANOVA with repeated measures ($F_{(2,8)} = 64.58$, $p < 0.001$). As can be seen in Figure 7, there was a large difference between the Alone condition and the other conditions. Post-hoc application of Tukey's HSD test indicated that this was the only significant difference. There were 7 participants who failed to complete the Alone condition in the five minutes allowed.

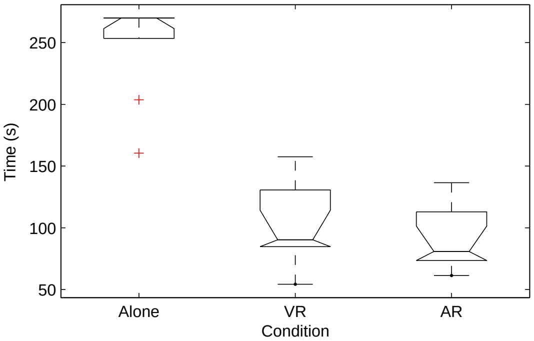

Figure 7. Task Completion Time.

There was also a significant difference in path error between conditions (ANOVA $F_{(2,8)} = 27.21$, $p < 0.001$) (see Figure 8). There was a significant difference between the Alone condition and the two other conditions ($F = 26.24$, $p < 0.001$, $F = 27.75$, $p < 0.001$), but no difference between the AR and VR conditions.

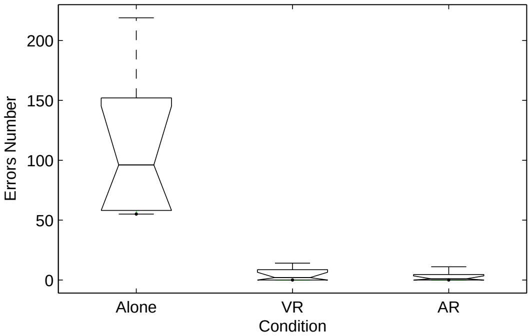

Figure 8. Path Error.

To understand how the egocentric user moved through the maze we measured their position and head orientation. We observed a significant difference in the velocity of the immersed user (ANOVA $F_{(2,8)} = 5.473$, $p < 0.05$). Users moved 1.5 times as fast in the VR and AR conditions as in the Alone condition (mean = 1.3, std = 0.22, vs mean = 0.94, std = 0.29). The egocentric user's head movement angular velocity was also higher in the Alone condition than the other conditions ($F = 14.72$, $p < 0.001$). This might suggest

---

[3] We did no extra correction of the relative position between the user's eyes and the tracker as these were relatively close in our case

more effective motion, except that in the path analysis we observed that in the Alone condition some users chose an effective naive-search approach, while others users moved without pattern.

The head motion of the exocentric user was also interesting. In both the AR and VR cases, the user generally converged to an average static orientation, viewing the maze from a viewpoint perpendicular to the ground plane (also observed from the video analysis). This orientation makes the 3D maze model look more like a traditional 2D map, allowing the exocentric user to most easily observe the solution path.

Subjectives and Process Measures

The performance measures were matched by the results from the subjective survey (see Appendix 1) where the egocentric user reported that the task was much easier in the AR and VR conditions (Q4: $F = 15.13$, $p < 0.001$, Q5: $F = 8.375$, $p < 0.001$, Q6: $F = 4.973$, $p < 0.05$). Apart from this there were few differences observed in response to questions about the VR and AR conditions. The high values on the collaboration questions (mean = 6.5) indicates that participants felt that it was easy to collaborate, and indeed no communication problems or misunderstandings were observed or reported.

The exocentric users didn't report any difference in their phenomenal experience or situational awareness between the AR and VR viewpoint conditions. They reported that it was easier to understand their partner's ideas than to communicate their own ideas.

4.7. Discussion

Contrary to our expectations, the AR condition didn't provide a significant performance improvement over the VR condition. Contrary to [9], [18] the proprioperception and visual background provided by the actual environment in the AR condition did not improve the efficiency of the task. However, the task for the exocentric user was not a typical 3D task and didn't require manipulation where natural hand perception could play a role. Also we expected that a physical FOR would improve the feeling of the natural head movement afforded by the HMD (6DOF) in a known environment. But since users reduced the 3D maze problem to a 2D one, they didn't require large movements of the head. In further investigations we propose to study other types of tasks with a higher DOF in the task of the immersed user (e.g. astronomy).

Nevertheless, the results show that a mixed-space collaboration doesn't disrupt the efficiency of the task. A future step will to repeat the same experience by adding a exocentric view to the immersed user and comparate at new the communication cost ([28], [29]). Furthermore, similarity in results between the AR and VR conditions demonstrates that an AR interface can be as useful as an immersive interface for supporting collaboration. Seeing the real environment doesn't appear to interfere with successful cooperation or communication processes.

The awareness provided by the avatar and gaze direction indicator seems sufficient support for collaboration in the maze navigation task.

5. Experiment 2: Impact of the relation between interaction and visualization space

5.1. Task

In the second experiment we wanted to explore the effect of display affordances on the AR condition. We used a modified version of the previous task that would increase the demands for 3D viewing of the maze by the exocentric user.

In the new task, the exocentric user was required to guide the egocentric user successively to four specific locations inside the maze before going to the exit. At each location there was a white billboard with a code written on one face. The billboard was oriented vertically and the code could only be read by viewing the maze from side-on (i.e. in an elevation view rather than plan view). The content of the code was different for the *Exo* and *Ego User*. Each of the users needed to read aloud the code he read from his own viewpoint in order to proceed.

We chose the following four conditions for the experiment (see Figure 9):

1. *Monitor AR solution with colocated FOR*: The interaction and visualization space are co-located and registered. The user can manipulate the physical marker behind the screen.
2. *Monitor AR solution with decoupled FOR*: Same as condition 1 except that the visual and interaction space are separated and unaligned (in position and orientation).
3. *HMD AR solution with colocated FOR*: the AR User can naturally turn his view around the maze, by simple head movement. In this case, the interaction and visualization space are also in the same position.
4. *Hand-Held AR solution with colocated FOR*: We replace the HMD solution with a hand-held tablet PC display.

We used the same procedure as in experiment 1, with an additional measure of when visual tracking was lost in the AR condition. We used also a more complete post-experiment interview. The monitor used in conditions 1 and 2 was a ViewSonic LCD panel and the tablet PC was a Compaq TC1100. Each display had a resolution of 800x600 pixels and the camera video input was 320x240 pixels.

Ten pairs of students were recruited for this experiment. They ranged in age from 18 to 27 years.

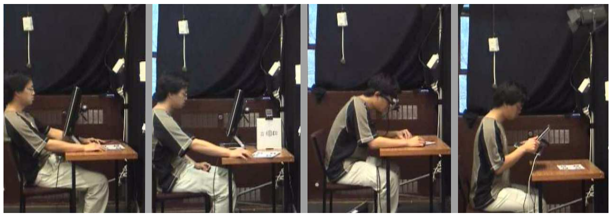

Figure 9. Setup: The four different conditions for the Exo User.

5.2. Results

Performances Measures

There was a significant effect of condition on time to completion ($F_{(3,9)}$ = 0.1093, $p < 0.001$). As can be seen in Figure 10, the slowest performance was in condition HandHeld with the tablet PC (mean = 220.12, std = 43.18), while condition Monitor Aligned was fastest (mean = 139.38, std = 22.9). These results were also mirrored in the measures of the egocentric users' head movement angular velocities (F = 7.303, $p < 0.001$).

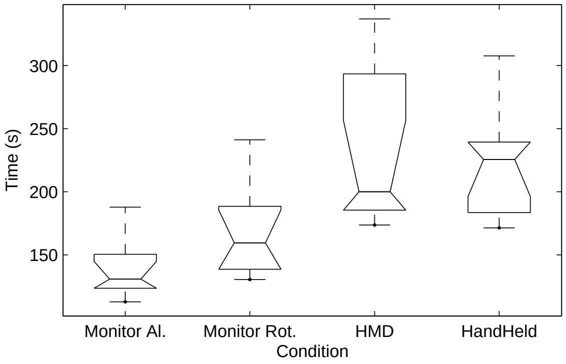

Figure 10. Task Completion Time.

We have also observed a significant difference in the tracking lost error between conditions (F = 2.68 $p < 0.1$). The HMD had the largest error, explained by the general problem observed during the reading of the code: the users tended to move their head very close to the marker to read the code and in doing so occluded the border of the fiducial tracking marker. The main reason is the low field of view provided by the HMD (associated with higher head movement angular velocities) in comparison to the larger field of view provided by other configurations (and associated with lower head movement angular velocities).

In regards to the tracker movement of the exocentric user (related also to the head movement), a significant difference was observed on the angular velocity (F = 3.443, $p < 0.05$) and position velocity (F = 13.32, $p < 0.001$). The rankings of the means are the same for the two types of movement, with the HMD condition highest velocity, followed by the tabletPC and monitor conditions with low values.

Subjectives Measures

As in experiment 1, results from the questionnaire contained few significant differences in support for collaboration or the task between the conditions. We observed though, that the tablet PC had a significant difference in the reported overall collaborative session satisfaction (Q1: F = 3.402, $p < 0.05$).

In contrast, the interview provided a lot of information. Participants generally found the task easy to do and the communication efficient. A large number of users commented on the one-way nature of the communication processes and the limited role for the immersed user. The egocentric user noticed some differences between the success of the task, and these reports were generally a good match with the objective performance results (in 66% of cases). The preferred devices for the exocentric user were the HMD (40%) and the aligned screen (40%). The least preferred devices were the tablet PC (40%), and the non-aligned screen (30%). A lot of users complained about the weight of the tablet PC (condition 4) and the difficult mental mapping required with the non-aligned screen (condition 2). Some people also raised the difficulty in reading with the HMD the label on the billboard. Generally, users appreciated having the feedback from being able to visually observe their hand and found it useful (60%).

Process Measures

More interesting are observations of the hand and head movement of the exocentric user. The manipulation of the marker was generally dominated by small rotations (mainly rolling) and small translational movement between different observational periods (an 'orbiting the scene' metaphor). Users generally manipulated the marker based on intuition, with two hands in condition 1 and 3. During the billboard label identification task, where some users choose to orient the marker to have the label at the top, some users guided the egocentric user first to the label and once the user was there, rotated the marker to read the label themselves.

In regards to head manipulation, even though the HMD is preferred by users for its "naturalness" for visualizing scenes from different angles, people generally preferred transferring the choice of the viewpoint by manipulating the marker with small movement of the hand.

We observed also that in contrast to experiment 1, users didn't choose a static 2D view throughout the experiment. This is explained by the requirement to identify and read the label. The tablet PC condition generally showed users interleaving periods of two hand manipulation of the tablet for changing the viewpoint with a grasping with one hand while the other hand moved the marker. We also observed around half of the users standing up during condition 4 (bad visibility), or moving their head close to the marker in condition 3.

5.3. Discussion

The results show that the choice of the interface significantly influence the realization of the collaborative task. These results are visible not only on the performance but also on the subjective and process measure.

The performance measures demonstrate that the Monitor approach seem to offer the most efficient solution to collaborate, user reacting and interacting in natural way during the collaboration. The results for the HMD are more mitigated. Users feel the HMD comfortable and natural but the actual performances impact on the efficiency of the task. Like [3] we think until there are large improvements of the quality in the HMD, this solution remains limited. In [3], users interviews show a large interest in hand-held displays during a co-located collaboration. In our case, the results are poorer mainly because the technology choice was different (tabletPC vs Palmtop) and the co-presence requirement absent fom the task. The weight of the HandHeld and the constraint of two handed grasping limited its usage in collaborative scenario.

We also notice that even users moving more fast in HMD or tabletPC condition (better velocity), it's not improving the efficiency of the task. It seems the stability and reduced DOF offer by the Monitor conditions encourage the user to changing efficiently his viewpoint.

Also, like [27], we show that a rotation and translation transformation between the interaction and visualization space reduces performances by complicating the FOR mental transformation to the egocentric user. It seems interesting to notice the ability of subjects to manipulate the marker to provide themselves an efficient viewpoint (oriented or not) of the avatar and the task focus.

6. Design Lessons and Implications

Even these experiments furnish only initial results on mixed-space collaboration study, we present some lessons and implications on their design.

The type and dimension of the task influences considerably the choice of collaborative environment to use. Without large requirements on 3D spatial navigation or manipulation, a standard VR environment with desktop interface can be a good solution for its demonstrated efficiency. On that, we think that a mixed-space collaboration can be an interesting choice in the case of multiple users co-located in a real environment interacting with other users immersed on VR environment or for keeping a seamless transition with real content (books, maps, mockup, etc.). But interesting results during our experiment of the manipulation of the marker to changing its viewpoint (and previous one of [2]) shows a potential usage of a AR for natural navigation/viewpoint control.

In terms of awareness, a standard VR avatar representation has been sufficient to support the mixed-collaboration and a lot of research results on CVE can be a good basis for the design of the collaborative support. As proposed by some subjects, displaying the view of the ego user on the screen of the exocentric user can improve the coordination during the communication. Definitively, a deeper exploration about the user representation remains to be done on mixed-collaboration. For example, the physical space on the AR environment can provide the opportunity to use more tangible interface to “externalize” the immersed user or interact with him (i.e. using plastic dolls or motor-controlled tangible avatar).

For the egocentric user, the audio communication serves largely to coordinate with the other user but can be largely optimized. For example, in the firefighter scenario introduced previously, users suggested providing telepointer or path trajectory for the immersed user (that can be controlled by tangible interface). In this way, the audio support would be reduced to more efficient and directive information like “follow this way”, “go to the door I point to you”.

A monitor see-through solution seems nowadays a simple and efficient solution for supporting the mixed-space collaboration but can't be the most useful solutions for co-located users interacting with an immersed user ([3]). Also, applying a simple mapping between visualization and interaction space seems also indispensable in video see-through AR context.

7. Conclusions

We conducted one of the first quantitative and qualitative experiments on a mixed-space collaboration between AR and VR environment, with an exocentric-egocentric collaboration. Firstly, we have presented a new conceptual framework that can be used for designing transitional, mixed-space or multi-perspective applications. Secondly, the results of this first experiment show that a collaboration between the two environments doesn't affect the efficiency of the task and a mixed-space approach offers similar results to a VR standard collaboration. Finally, the second experiment demonstrated that a monitor see-through approach for the exocentric user offers an efficient way to realize the collaborative task.

The conceptual model introduced in this paper can be largely refined. For that, we are converging by enhancing the cardinality and the restricted representation of Milgram's continuum with a more ecological approach providing a new base for our model. Also, other combinations of mixed-space collaboration need to be studied like AR egocentric with VR exocentric. In the future, we will concentrate on adding the transition component to the mixed-space collaboration, introducing a large number of new interesting research foci like the transitional awareness aspect.

References

[1] S. Ainsworth. A functional taxonomy of multiple representations. *Computers and Education,*, 33(2):131–152, 1999.

[2] D. Belcher, M. Billinghurst, S. Hayes, and R. Stiles. Using augmented reality for visualizing complex graphs in three dimensions. In *ISMAR'03.*, 2003.

[3] M. Billinghurst, D. Belcher, A. Gupta, and K. Kiyokawa. Communication behaviors in co-located collaborative ar interfaces. *Lawrence Erlbaum Associates International Journal of Human Computer Interaction*, 2002.

[4] M. Billinghurst, H. Kato, and I. Poupyrev. The magicbook - moving seamlessly between reality and virtuality. In *IEEE Computer Graphics and Applications*, volume 21, pages 6–8. IEEE Computer Society Press, 2001.

[5] D. A. Bowman. *Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation, and Application*. PhD thesis, Virginia Polytechnic State University, 1999.

[6] B. Brown, I. MacColl, M. Chalmers, A. Galani, C. Randell, and A. Steed. Lessons from the lighthouse: collaboration in a shared mixed reality system. In *CHI'03*, pages 577–584. ACM Press, 2003.

[7] R. P. Darken and J. L. Sibert. Wayfinding strategies and behaviors in large virtual worlds. In *CHI'96*, pages 142–149. ACM Press, 1996.

[8] G. W. Furnas. Generalized fisheye views. In *CHI'86*, 1986.

[9] E. D. Graham and C. L. MacKenzie. Physical versus virtual pointing. In *CHI'96*, pages 292–299, 1996.

[10] C. Gutwin and S. Greenberg. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *CSCW'98*, pages 207–216. ACM Press, 1998.

[11] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *IWAR'99*, 1999.

[12] K. Kiyokawa, M. Billinghurst, S. E. Hayes, A. Gupta, Y. Sannohe, and H. Kato. Communication behaviors of co-located users in collaborative ar interfaces. In *ISMAR'02)*, pages 139–149, 2002.

[13] K. Kiyokawa, H. Takemura, and N. Yokoya. Seamlessdesign for 3d object creation. *IEEE MultiMedia*, 7(1):22–33, 2000.

[14] B. Koleva, H. Schnodelbach, S. Benford, and C. Greenhalgh. Traversable interfaces between real and virtual worlds. In *CHI'00*, pages 233–240. ACM Press, 2000.

[15] P. Lamb. Human performance in space telerobotic manipulation. Master's thesis, University of Canterbury, Christchurch, New Zealand, 2002.

[16] J. Leigh and A. E. Johnson. Supporting transcontinental collaborative work in persistent virtual environments. *IEEE Computer Graphics and Applications*, 16(4), 1996.

[17] B. Lok, S. Naik, M. Whitton, and J. Frederick P. Brooks. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoper. Virtual Environ.*, 12(6):615–628, 2003.

[18] A. H. Mason, M. A. Walji, E. J. Lee, and C. L. MacKenzie. Reaching movements to augmented and graphic objects in virtual environments. In *CHI '01*, pages 426–433, 2001.

[19] P. Milgram and F. Kishin. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(9):1321–1329, 1994.

[20] H. Nakanishi, S. Koizumi, T. Ishida, and H. Ito. Transcendent communication: location-based guidance for large-scale public spaces. In *CHI'04*, pages 655–662. ACM Press, 2004.

[21] M. S. Pinho, D. A. Bowman, and C. M. Freitas. Cooperative object manipulation in immersive virtual environments: framework and techniques. In *VRST'02*, pages 171–178. ACM Press, 2002.

[22] S. Prince, A. D. Cheok, F. F. andTodd Williamson, N. Johnson, M. Billinghurst, and H. Kato. 3d live: Real time captured content for mixed reality. In *ISMAR'02)*, 2002.

[23] M. C. Salzman, C. Dede, and R. B. Loftin. Vrs frames of reference: A visualization technique for mastering abstract information spaces. In *CHI'99*, pages 489–495. ACM Press, 1999.

[24] W. A. Schafer and D. Bowman. Evaluating the effects of frame of reference on spatial collaboration using desktop collaborative virtual environments. *Springer-Verlag Virtual Reality*, 7, 2004.

[25] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *CHI'95*, pages 265–272. ACM Press/Addison-Wesley Publishing Co., 1995.

[26] Y. Wang and C. L. MacKenzie. The role of contextual haptic and visual constraints on object manipulation in virtual environments. In *CHI'00*, pages 532–539, New York, NY, USA, 2000. ACM Press.

[27] C. Ware and J. Rose. Rotating virtual objects with real handles. *ACM Trans. Comput.-Hum. Interact.*, 6(2):162–180, 1999.

[28] H. Yang and G. M. Olson. Exploring collaborative navigation: the effect of perspectives on group performance. In *CVE'02*, pages 135–142. ACM Press, 2002.

[29] X. Zhang and G. W. Furnas. Social interaction in multiscale cves. In *CVE'02*, 2002.

[30] X. Zhang and G. W. Furnas. The effectiveness of multiscale collaboration in virtual environments. In *CHI '03 extended abstracts*, pages 790–791. ACM Press, 2003.

A Appendix 1

The survey from [24] is based on a likert scale with 3 main parts: collaboration, task and awareness. In function of the conditions, the user or the experiments some were removing.

1. I felt the collaboration session overall went great. We had no problems and did not struggle to complete the task.
2. I could easily understand my partner's ideas.
3. I could easily communicate my ideas.
4. The task was easy to complete.
5. The task required little effort.
6. I did not have to concentrate very hard to do the task.
7. How often did you know where your partner was located?
8. During the trial, how often did you know what your partner could see?
9. During the trial, how often did you know what your partner was directly looking at?