Is “Presence” Important in Mobile Map Interaction?

Figure 1. Real world scene and a view on a mobile device to a 3D city model (top). Below, the phone (Nokia N93) used in this study.

Antti Oulasvirta
School of Information
University of California, Berkeley

Helsinki Institute for Information
Technology HIIT

Sara Estlander
Helsinki Institute for Information
Technology HIIT

Giulio Jacucci
Helsinki Institute for Information
Technology HIIT

CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract

Presence, the sense of “being there instead of here”, has become the driving design goal and topic in virtual reality research. The key idea of *mixed reality interaction* (MRI), by definition, is interaction “in” the blending of virtual environment (VE) and “real” or physical environment (PE). Mobile devices complicate MRI, because digital information cannot be directly projected onto the physical world. We ask if the notion of presence is important in such circumstances. The use of 3D mobile maps for identifying correspondences between buildings in VE and PE was examined. No signs were found of users exhibiting presence in VE. Instead, 1) they constantly *switch* between the two viewports, 2) they are aware of PE contents and 3) they interact in a way that tries to align the two spaces. Mobile MRI is better understood in terms of a *dyadic process* constituted by the user’s agency and cognitive capacities being “split” between VE and PE. The question is how actions in the two environments can *cohere* in the pursuit of a *single* task goal.

Keywords

Presence, mixed reality, 3D interaction, mobile maps

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

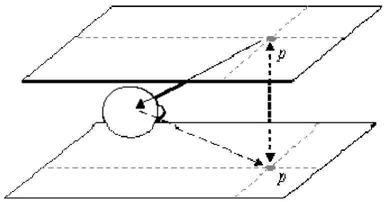

Figure 2. The MR mapping problem: how to understand the referential relationship between two spaces. In mobile MR, the user has to mentally construct the referential relationship between the virtual and the physical, because viewports to the two spaces cannot be projected onto each other.

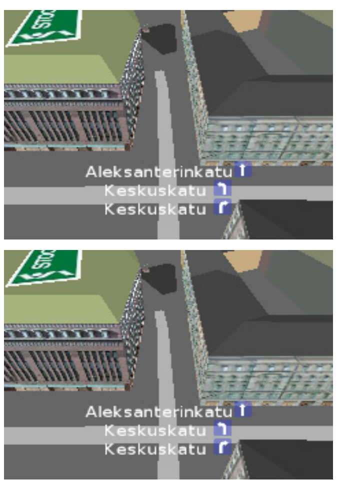

Figure 3. A roof-top view (top) and a track-based maneuvering scheme (street following, bottom). The default view was the street-level view of Figure 1.

Introduction

Presence has been defined as the sense of being physically transported to a remote workspace and as a "perceptual illusion of non-mediation" [1]. When "present", "perceived self-location and realization of action possibilities are connected to a mediated spatial environment; mental capacities ... bound by the mediated environment instead of reality" [5]. Presence has become the prime approach of psychology to the design of virtual environments and basically to all media where the goal is to surpass the limits of human sensory capacities.

One may ask what the role of this notion is in newer forms of technology that have been deliberately designed to leverage the *physical* presence of the user, or, conversely, to *blur* the boundary of the real and the virtual. Milgram and Kishino [2] defined *mixed reality* as "the merging of real and virtual worlds somewhere along the 'virtuality continuum' which connects completely real environments to completely virtual ones." Mixed reality systems either augment the virtual world with physical features or augment the physical with virtual ones. *Mixed reality interaction* (MRI) is a broader concept that involves tasks in which actions and processing of information takes place in both the PE and the VE. Thus, by definition, there is always an aspect of objects being somehow "divided" "mapped" or "shared" between two planes.

Mobile devices complicate MRI in the sense that they technically cannot be based on the projection of digital information onto physical world. Moreover, it is physiologically impossible to (foveally) process a small mobile device display *and* surrounding scene simultaneously. We call this the "mapping problem".

Figure 2 describes this view in a case where the two planes can be conceived as Euclidian spaces, as in the case of mobile maps.

There are two major *a priori* categories of how users can be "present" in a mixed reality task:

1. "Unal presence", presence in *one* world at a time:
    a. Physical environment (PE)
    b. Virtual environment (VE).
2. "Dual presence", presence in two environments:
    a. Simultaneously in PE and VE
    b. Sequentially in either PE or VE at a time.

All four options are possible *a priori*, although they do presume different cognitive mechanisms underlying the feat of achieving presence.

Regardless of which type of presence there is, the effects H1-H5 mentioned in the sidebar should be observed if presence is important in this task at all. As a source for these inferences, we refer to Lombard and Ditton's [1] well-known review.

Case: Localizing Objects with a Mobile Map

As a case we look at data from a field experiment [4] on how users localize PE buildings with 3D and 2D mobile maps. This is a MR task that explicitly requires mapping of two planes. Technical details of the study are given in the original article. Of the system and Method, it is worthwhile to know that:

- The design of the 3D map instantiates several known principles of good virtual reality design, including photorealism and maneuvering assistance.

IMPLICATIONS FROM PRESENCE RESEARCH

(H1) Realism in VE is beneficial: Realism of the virtual model should increase the sense of presence.

(H2) Consistency of switching behavior: As a behavioral indicator of presence, users should not exhibit abrupt switches between VE and PE, but either switch consistently between the two or stay concentrated on one environment for a longer time while disregarding stimuli in the other.

(H3) Mimicking: Users achieving presence should generally prefer maneuvering in 3D in ways that mimic real world movement.

(H4) Presence is useful: An increased sense of presence should generally improve performance.

(H5) Suspension of disbelief: Presence should be associated with negligence of the physical world.

Figure 4. The recording equipment. The user carries one camera on his chest (1) and two attached to the mobile device (2 and 3). The moderator follows one step behind, shooting with a wide angle lens (4).

- The default mode of movement and camera in 3D is based on a street-level view. Other views and maneuvering schemes are optional (e.g., see Figure 3).
- There is a 2D map as a comparison condition that does not have any "presence-enhancing" features. The map is a simple cartographic street map.
- The task is a pointing task: a target is indicated in the VE (map, 2D or 3D), and the user is asked to point in the direction of the target in the PE. The real target in PE may or may not be visible. The user can move her body freely during task.
- 16 subjects performed 24 tasks, spending about 1.5 h on the field with an experimenter.
- Verbal protocols, multi-video data, interaction logs, workload measures, performance measures, and various background information are collected. See Figure 4 for an illustration of the multi-camera setup.

Cue type	3D	2D
Known landmarks	Very often	Often
Building shapes	Often	-
Façade details	Often	-
Façades (whole)	Often	-
Relative directions	Often	Sometimes
Street names	Often	Very often
Street crossings	Sometimes	Often
Blocks, or part of blocks	Rarely	Sometimes
Parallelism of streets	Rarely	Rarely
Cardinal directions	Very rarely	Sometimes
Store/office names	-	Rarely
Street number	-	Rarely

Table 1. A catalogue of cues used and their frequencies. Based on analysis of think aloud data of 75 tasks.

Assessment of Evidence

The evidence speaks against the idea that presence is important in mobile MRI. The data is reported in its entirety in the original article; here, we concentrate on observations related to the hypotheses.

- *First, users are perpetually conscious of aspects of PE while interacting with VE, and vice versa.* The predominant strategy for solving the task, in both 2D and 3D maps, was finding a point that maps a location in the VE to one in the PE (hereafter: *a reference point*). One can infer the target's direction by estimating the angle difference Δ between the target and the reference point, and the target can then be pointed at Δ degrees to left or right from the reference point, assuming that one knows one's own position in relation to the reference point. One can achieve this also without locating oneself on the map, by "sandwiching" the target with two external reference points. (This did not always work: In verbal protocols we found frustration for searched-for PE visual features not found in VE and vice versa.) Now, importantly, this tactic, by definition, entails *keeping in mind a description of the target which represents the other environment.* Thus, we believe that "suspension of disbelief" cannot be in play here. In contrast, users are constantly evaluating similarities between the worlds.
- *Second, photorealism of the model does not help, it distracts.* Table 1 below shows some of the cues users relied on for finding a reference point. An important observation we made is that in 3D use, the search proceeded mostly by scanning buildings that surround the target in the VE, for example to spot a yellow building in the midst of gray ones. We observed that users *cannot* operate with a photorealistic model

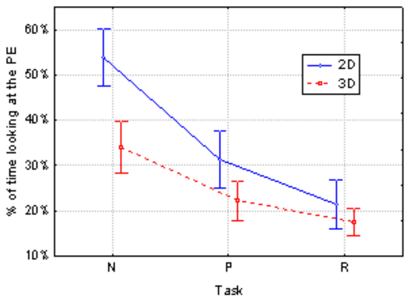

Figure 5 Time spent looking at the PE was higher in 2D than in 3D. Conversely, 3D users were more immersed in VE than 2D users. Taken together with the fact that 3D users were slower, we conclude that preferring interaction with the VE hampered performance. Good strategies allow users to concentrate on the boundaries and cross-overs.

efficiently because their attention is constantly guided to cues that may be salient in the VE but are ineffective and undiagnostic in the PE. In this respect, 2D worked much more efficiently and led to faster task completion times. Thus, contrary to findings of presence research, realism is not needed. 2D maps as representations rely on centuries of experience on effective symbols and layouts of a map.

- *Third, movement in 3D is not "realistic".* In principle, if the users so wanted, they could have immersed themselves in the VE to a view that corresponds to their real street-level view and offers rich visual detail. This was the default view. However, to mitigate the "photorealism trap", many 3D users would turn on the Tracks feature see the street names to use those as cues. Some also learned to ascend to the rooftop view and rotated there in search of statues, parks, recognizable buildings and rooftop logos of companies. If this did not work, they "flew" around in the area above the target. Thus, realistic movement did not seem important either.

- *Fourth, more "immersion" in VE means worse performance.* Time spent looking at the PE was higher in 2D than in 3D use (Figure 5). Conversely, 3D users were more immersed in VE than 2D users. During that time, they traveled the same distance in VE as 2D users. Taken together with the fact that 3D users were slower, we conclude that preferring interaction with the VE actually hampered performance. Good strategies in MR tasks allow users to concentrate on the boundaries and cross-overs of VE and PE.

- *Fifth, navigating in a 3D VE increases working memory load.* We divided subjects into two groups, according to a median split for their Corsi score (a test for visuo-spatial working memory span). The high span group was 16% quicker and used Switch perspective 26% more in 3D. They gazed at the environment 11.8% less per task, although the two groups' frequency of gazing at the environment was at the same level. Similar differences were *not* found in 2D use. These results suggest that 3D performance, but not 2D performance, may be dependent on visuo-spatial working memory span. Due to more fragmented viewing of VE, and less help from bodily strategies, 3D users have to keep in memory more locations of interest in order to solve the problem.

- *Sixth, bodily strategies in PE help performance.* We were surprised by the general tone of the results concerning bodily conduct: 2D rather than 3D tasks involved more efficient use of body and gaze. 2D users turned their upper bodies more in search of cues like street names and crossings. They deployed gaze significantly more effectively to find cues like parallel streets. They tilted their heads and rotated their devices in their hands more than 3D users. Their walking was significantly more efficient. 2D users did ego-centric alignment more, even when the target was remote and not visible in PE, by relying on the zoom out function to see the current position *and* the target POI's position in the same view. The crux of these strategies may be that they allow 2D users to avoid mental manipulation and rely more on perception in solving the task. We see a parallel to experienced Tetris players' tactic of rotating a piece in order to *see*, rather than mentally simulate, its fit. This means that it is more important to align bodily engagement in two environments, PE *and* VE, rather than be immersed in acting in only one environment, the VE.

SUMMARY OF EVIDENCE AGAINST PRESENCE

(1) Users have to be conscious of aspects of PE while interacting with VE. Being aware of the other environment's contents runs counter to the very idea of presence as "sense of non-mediation".

(2) Photorealism of the model does not help but distracts and misleads. Here, realism does *not* increase performance. This runs counter to one finding of VR research.

(3) Movement in 3D is not "realistic". Users prefer 2D-like strategies that are not realistic but that allow faster and more efficient finding of cues.

(4) More "immersion" in VE means worse performance. Contrary to main claim of presence research, even if users were present, it would not help their performance but make it worse. Efficient strategies rely on awareness of both VE and PE and thus are associated with switching behavior and focus on PE which is more information rich.

(5) Navigating in the VE in a 3D street-level view increases working memory load in this MR task as it increases the discrepancy between the VE and PE. Users rather go for direct perceptual matching or ego-centric alignment to avoid this.

(6) Bodily strategies in PE help performance. It is more important to align bodily engagement in two environments, PE *and* VE, rather than be immersed in acting only in only one environment, the VE.

Conclusions

Work on scientific concepts proceeds in a cycle of two endeavours: charting the boundaries of existing concepts and proposing new concepts. This paper has proposed that the notion of presence may not be useful in all areas of mixed reality. Our examples indicate that when operating at the augmented reality end of the spectrum, presence might not be a key issue, neither for design nor as a phenomenon describing interaction.

The results point out that comments like the following on 3D mobile maps must be overly optimistic and detached from the requirements of action arising interacting with mobile maps: "the most positive feature was found to be the possibility to recognize the features in the surrounding environment, which provides a link between the real and virtual worlds. This removes the need to map-read, which is required when attempting to link your position in the real world with a 2D map, hence the VR interface offers an effective way to gauge your initial position and orientation" [3]. Our findings lend support to a less Cartesian and more constructivist conception of human perception, where the person is seen as an active, intentional actor in an environment that offers different resources for actions.

We acknowledge the tentativeness of these conclusions. One can state that presence could not even emerge in mobile VE interaction, because of the small display, limited and slow interaction, poor visual quality, and other factors. This may be the case. The desiredatum for testing our claim would pair standard measures of presence to those of performance. Our hypothesis is directly operationalizable and testable: higher presence should be associated with poorer performance. More attention to boundaries and cross-overs of the two environments should be associated with improved performance. However, our analytical observations on the task and cognitive requirements in MR tasks hint that a sense of presence is unlikely; the distracting elements are too numerous.

We leave it as a challenge for future research to come up with a more descriptive concept. Importantly, it should capture the synergistic relationship between actions in and perceptions of the two environments.

Acknowledgements

This work has been supported by EU 6th Framework projects IPCity and PASION and by the Academy of Finland project ContextCues. The m-LOMA project was funded by EU InterregIIIA.

References

[1] Lombard, M. and Ditton, T. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication 3*, 2 (1997).

[2] Milgram, P. and Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems 77*, (1994), 1321-1329.

[3] Mountain, D., Liarokapis, F. Mixed reality (MR) interfaces for mobile information systems. *Aslib Proceedings, Special Issue: UK Library & Information Schools, Emerald Press 59*, 4&5 (2007), 422-436.

[4] Oulasvirta, A., Estlander, S. and Nurminen, A. Embodied Interaction with a 3D versus 2D Mobile Map. Accepted for publication in a forthcoming *Special Issue on Mobile Spatial Interaction* for the journal *Personal and Ubiquitous Computing*.

[5] Wirth et al. A process model of the formation of spatial presence experiences. *Media Psychology 9*, 3 (2007), 493-525.