A user study of augmented reality in an urban design studio

HARTMUT SEICHTER
University of Canterbury, New Zealand

Abstract. This paper outlines the design, execution and analysis of a user evaluation experiment using Augmented Reality (AR) in an urban design studio. The aim of the experiment was to gauge the differences between two interfaces in regard to their impact on the design process. The two conditions of the experiment were a direct manipulating tangible user interface and a pen-like 3D manipulation interface. Findings include differences in perceived object presence, performance measures, perceived performance and variations in the communication patterns. These findings have implications for the integration of this technology in a praxis relevant design work flow.

1. Introduction

Architectural design activity is, in large part, a joint effort of various parties. Supporting technologies have been subject of research in relation to distant collaborative work. Collaboration in the design process can, for instance, help to negotiate a consensus with the stakeholders involved. Collaboration can be conducted in various forms, through different media. However, some initial stages in the design process do not require a discussion with remote involvement.

Design collaboration in the early stages is similar to a brainstorming process, thus it requires free elaboration on a local platform for the purpose of presentation and manipulation. Usually the discussion takes place in a collocated setting. That means, the involved parties (e.g. designers, customers and consultants) gather in a local setting, allowing them to share information. To date this kind of communication in combination with digital design tools has not been investigated in detail. Emphasis in this research is given to the use of a localised table-top Augmented Reality (AR) setting as a discussion platform for spatial content for design. This study evaluates a multilayered communication system that tries to enhance collocated design
through an AR environment with Tangible User Interfaces (TUIs). Objective is to foster these technologies for an externalised thought process for spatial content. This parallels conventional non-digital design tools, like the one in architectural design, which are functional as an external entity to provide a meaningful reference within the process of work and consideration.

AR provides new possibilities of three-dimensional input for the access of spatial content. But due to a lack of established methodologies no evaluation has been attempted in this area. This study will outline important issues of using augmented spatial representation for urban design.

**Figure 1.** Left (a) The BenchWorks table showing the virtual scenario with a massing model, initial developer proposal and a student design. Right (b): An earlier prototype with a pen-like 3D input device.

Due to its status as an enabling technology AR can be applied in a diversity of user domain tasks. Naturally, existing studies are specifically tailored to task. Architectural design spans across various domains and inherently deals with spatial content. This overlaps with AR, where virtual and real spatial content converge. User evaluations in AR can utilise this breadth in order to outline a methodological framework. This study uses subjective and objective metrics to provide an insight in the practical usage of immersive AR technologies in the design process.

### 2. Context and Previous Work

This study is embedded in a wider body of work and extends the notion of other research projects. This section will shortly introduce the direct context of the research conducted and covers relevant earlier work.

Architectural design and the design process are influenced by the choice of tools and their application for the purpose of creating an architectural proposal. A series of studies in the Department of Architecture at the University of Hong Kong (Bradford, Cheng and Kvan 1994; Gao and Kvan 2004; Kuan and Kvan 2005; Kvan, Yip and Vera 1999; Schnabel and Kvan...
AUGMENTED REALITY AND TANGIBILITY IN URBAN DESIGN

2002) investigated the influence of digital media on the design process. The connecting theme of these research projects is to illuminate the impact of digital media, interfaces or settings on the design process and its outcome. Bradford et al. (1994) coined the term of the Virtual Design Studio and identified means of communication as a key factor for design within new media. Hirschberg et al. (1999) looked at the utilisation of time zones for a virtual design process and investigated the impact of the time shifts on the design process. Schnabel and Kvan (2002) extended this work by observing the quality of design and analysed communication across different settings, like Immersive Virtual Environment (IVE) and Desktop Virtual Environment (DVE). Gao and Kvan (2004) measured the influence of a communication medium in dyadic settings based on protocol analysis. Another study by Kuan and Kvan (2005) investigated the influence of voxel-based 3D modelling between two different technical implementations.

This study extends the aforementioned body of work into the domain of Augmented Reality (AR) and Tangible User Interface (TUI). This research is using a user evaluation to investigate TUI in a design aiding AR setting. It contributes new knowledge regarding the impact of different affordances of creation interfaces within the design process.

Only few studies brought forward methodologies for usability evaluation in AR by providing guidelines and a framework for simulation environments. A prominent body of research is the work at Virginia Tech (Gabbard, Swartz, Richey et al. 1999; Hinckley, Pausch, Proffitt et al. 1997) which provides taxonomies for usability evaluation of VEs. As Bowman et al. (2002) point out; these frameworks originate and are targeted at specific user groups with a ‘problem’ to solve. Therefore, each usability evaluation is tailored to a specific task.

For instance Tang et al. (2003) presented an AR system with the aim of assessing the feasibility of augmented assembly. However, it did not take into account the actual communication issues, which are important in a design setting where designers and their peers collaborate. Billinghurst et al. (2003) introduced a methodology for this problem covering a wider angle of aspects, like display technology and visual fidelity in regard to AR and communication. However, their experiment did not assess presence, which is an important metric for immersive systems (Sheridan 1992).

Designing with and within AR was investigated by few research projects. Aspects of collaboration between co-located parties have been covered by Regenbrecht et al. (2004) with a focus on social components of the collaboration process in AR. Klinker et al. (2002) identified visual fidelity and communication as important usability issues for AR through a system for design review and interactive collaboration. They highlighted the problem of feasibility in the visual representation within an immersive
augmented display. ARTHUR (Broll, Lindt, Ohlenburg et al. 2004) is a system for collaborative urban design using AR for presentation. It allows designers a direct access to geometrical data and provides a tabletop immersive environment. Additionally it uses more advanced technology like see-through HMD and utilises a computer vision based tracking system for view and input registration. However, it attempts to incorporate manipulation techniques of CAD directly through spatialised 2D menus and does not provide an evaluation of this approach. Despite this, ARTHUR is in various ways similar to the system used here in this study. Tabletop AR has been demonstrated as a viable technique to share information like plans, models and analytical data (Underkoffler and Ishii 1999).

This study looks into the usage and impact of TUIs for direct spatial manipulations in AR in a design environment. Earlier research implemented various ways to gain advantage from AR for tasks but evaluations were limited and did not cover variances in input devices for spatial content.

3. User Study

A usability evaluation allows an insight in how users perceive and interact with a system in a praxis relevant setting. It permits to create a hypothesis about the relation of interface, communication and the design task and the impact of human factors.

The aim of this study was to investigate TUIs, which lend themselves as a common ground for discussion in comparison to digitizer pens, which are inherently single user interfaces and have been adapted from 2D for the use in 3D. An urban design project in a Masters Degree course in architectural design was used as the context for the experimental setup. It provided a viable scenario for research of a collaborative AR system as it usually involves large virtual models, which need to be accessed by multiple users. Physical urban design models can be large in size and therefore difficult to handle easily. These physical properties also limit the ability of the model to present morphological information within its spatial context. For an assessment of the communication pattern the chosen scenario is valuable because urban design models are shared with several parties in order to discuss, analyse and re-represent design. These actions require the ability to change parts in-situ and visualise and discuss their impact. The experiment utilised a formal investigation task in order to evaluate the user input devices. This was preferential over using urban morphology methods, because methodological preferences would have masked the actual impact of the user interface.
3.1. INTERFACE CONDITIONS

The experiment design comprised of an observation phase and a creation phase. The observation phase utilised direct manipulated interfaces (i.e. AR Toolkit Markers) to view design proposals (see Figure 3). In the creative phase, users were asked to extend the design proposal. The independent variable in this experiment was the input interfaces for the creation phase of the design task. One condition implemented this through directly manipulated cubes (i.e. AR superimposed styro-foam cubes) (see Figure 2(b)), the other through a tool-object metaphor with a 3D pen based interface (i.e. Polhemus Isotrak with a ‘Stylus’) (see Figure 2(a)). In both conditions the users built an abstract massing model representing the extension of the existing proposal, which consisted of a fixed amount of cubes (see also Figure 1(a)).

![Figure 2. Left (a): pen-shape interface for manipulation. Right (b): tangible cube interface.](image)

3.2. PARTICIPANTS

Overall 28 students and staff from The University of Hong Kong (HKU) participated in the experiment. The age ranged between 22 and 40 years. The sample consisted of 9 female and 19 male participants. Overall, fourteen users used the pen setting, five female and nine male. The cube setting had the same amount of users, with four female and ten male. 21 were of Chinese origin, only 7 were of Caucasian origin. None of the users had encountered the system before. Four users identified themselves as practitioners, the rest as students. Six used Immersive Technology before, three of whom had more than five years of experience.

3.3. EXPERIMENT DESIGN

The experiment used a questionnaire for user survey and a video account of the trials with a subsequent coding analysis.
In order to get subjective data from the users a presence questionnaire developed by Gerhard et al. (2001) was chosen. Sheridan (1992) elaborates that presence is an essential metric for simulation environments and direct influences the performance ‘within’.

The original design for the questionnaire anchored in a study about presence with an avatar aided VE. It was chosen as the most appropriate questionnaire for this experiment as it included aspects of direct collaboration, the presence of another person, the involvement in the task and satisfaction with the system. Original items related to avatars were replaced by questions about the other participant. Other minor adjustments were made in order to address AR specific issues and to reflect the architectural task in the brief. Conole and Dyke (2004) argue that an object providing an appropriate affordance, perceived or actual, contributes to the sense of presence. The questionnaire by Gerhard et al. (2001) originated from the ‘Presence Questionnaire’ (PQ) by Witmer and Singer (1998) which measures items like sensorial factors, distraction and realism. These are relevant in this experiment and were kept in the questionnaire. Presence questionnaires are a viable source of measuring the involvement with a simulation system. Gerhard et al. argue that awareness and involvement are essential for a sense of presence. This questionnaire also allows subcategories like ‘spatial presence’ important to understand for collaborative work with large datasets in collocated settings like the one in urban design.

Post-rationalisation of communication related items in a questionnaire is an apparent problem. A more reliable communication analysis can be achieved through post-experimental coding methods like the one proposed by Kraut et al. (2002). This coding schema analyses the videos taken during the experiment for instances where users refer to objects, name their position or agree. These three items are further divided in sub-items specifying the context. Thus, one can distinguish the frequency of references to an object directly or through the spatial context. Similarly, utterances regarding the position or the position context, agreement on general ideas or agreements on actions (i.e. behaviours) are coded.

3.4. DESIGN BRIEF

The design brief was tailored to the design studio the students were enrolled in at the time of the experiment. For the practitioners attending the experiment, the design problem was known as well. The task revolved around a publicly discussed large scale development project in Hong Kong. In a high density area, called Mong Kok, a development plan proposed to redevelop an area of low-rise buildings with high-rise buildings. Part of the research in the design studio was to prepare a design proposal for the area to
be used for internal and public discussion. Hence, a key factor in the design brief was discussion. In the first phase both users had to discuss and choose between two design proposals on purely aesthetic grounds. These proposals were introduced to the users as suggestions by a fictitious developer. Both users were given AR ToolKit markers showing the virtual proposals as massing models; one L-Shape and a U-Shape high-rise building of the same height (see Figure 3). This phase of the experiment served as the initial encounter for the users with a tangible user interface. As result of their decision the users were asked to remove the rejected variant from the virtual building site.

![Figure 3. Left (a): L-shape proposal. Right (b): U-Shape proposal.](image)

In the second phase, the users were asked to create an extension to the chosen proposal. The brief introduced some constraints regarding the extension. The users were made aware that the two phases were consecutive. Two spatial attributes within the brief for the experiment design were adjusted in order to make the task ambiguous. Ambiguity is needed to force the users to reinterpret the actual problem. The intention was to induce a discussion through an artificial wicked problem (Rittel and Weber 1984). This decision originated from pilot tests where it became clear that users went straight ahead to solve the task without any discussion of the actual task given. As the experimental set-up was ideal to review spatial content, factors related to the size of the massing models were made ambiguous. The virtual models of the proposals were skewed in their actual size.

3.5. APPARATUS

To capture data for this experiment a consistent experimental set-up was needed to let the users have an undisturbed experience. A software and hardware platform for this experiment called ‘Benchworks’ (Seichter 2004) was used.
‘BenchWorks’ utilised standard PCs for compatibility and maintainability reasons. The client systems consisted of two identical Dell Optiplex GX260 systems with 2.8GHz Intel Pentium 4 HT CPU, 512 MB of RAM, Western Digital 80 GB hard drives, Intel Pro/1000 MT Adapter and nVIDIA Geforce FX 5200 graphics adapter with 128 MB of DDR2 VRAM with AGP 8x enabled. The video capturing was supported through two identical Philips ToUCam Pro II with USB 2.0 HighSpeed interface. The server for connecting the client systems and handling the start-up sequence was a generic 1.5GHz Pentium 4 system with 512 MB RAM and an Intel Pro 10/100 Ethernet Adapter. All systems were connected with the local LAN at The University of Hong Kong through a separate SMC 4-port switch in order to reduce interference with network traffic. The HMDs used in this experiment were of different types. One was an Olympus Eye-Trek FMD-700, the other was an i-glasses SVGA Pro. The Olympus Eyetrak was connected through the S-Video interface with an AVERmedia Video splitter, the i-glasses HMD was connected directly through a DVI-VGA adapter.

4. Results
This section will provide an overview of the experimental results.

4.1. TIME MEASUREMENT
The users were asked to finish both phases of the brief in 20 min. They were allowed stop earlier if they fulfilled the brief. Expectantly the time to completion did not vary between the pen and the cube conditions. The users spent a mean time of 16 min. The difference between the two conditions was insignificant. However the time taken for the design phase was statistically different (t-value: 2.27, p-value: 0.04) (see Table 1).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pen (Design)</td>
<td>5.8 (5.1)</td>
<td>4.1 (1.8)</td>
</tr>
<tr>
<td>Cube (Design)</td>
<td>10.1 (9.5)</td>
<td>3.9 (2.4)</td>
</tr>
<tr>
<td>Pen (Complete)</td>
<td>16.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Cube (Complete)</td>
<td>16.2</td>
<td>3.6</td>
</tr>
</tbody>
</table>

4.2. PERCEIVED PERFORMANCE
Two items in the questionnaire assessed the impact of the tool on their perceived performance. The users judged the overall performance of the tool in regard to finalising the task. The data (see Table 2) show that these items were quite reliably reported. For the creation phase the data reported a
significant difference between the conditions. The cube interface performed slightly better with statistical significant difference (t-value:-1.712, p-value: 0.01).

### TABLE 2: Perceived performance for interfaces.
Range for means from 1–5 (1=agree, 5=disagree).

<table>
<thead>
<tr>
<th>Relation</th>
<th>Mean</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pen (Observation Interface)</td>
<td>2.8</td>
<td>0.95</td>
</tr>
<tr>
<td>Cube (Observation Interface)</td>
<td>2.5</td>
<td>1.40</td>
</tr>
<tr>
<td>Pen (Creation Interface)</td>
<td>4</td>
<td>0.78</td>
</tr>
<tr>
<td>Cube (Creation Interface)</td>
<td>3.2</td>
<td>1.53</td>
</tr>
<tr>
<td>Observation Interface (both)</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

### 4.3. PRESENCE
Presence is a key measurement factor for simulation environments. It measures the perceived level of the users being immersed within a system. Presence goes beyond the concept of immersion which is purely technical. There are various categories of presence. Spatial presence describes the ‘being with’ of a user in regard of the space of the simulated environment. Social presence reports in regard of a user ‘being with’ avatars, synthetic entities or real persons. Object presence is the state of ‘being with’ an object in the same space and shared with others (Witmer and Singer, 1998). Presence is important for this experiment as it is closely coupled with the concept of affordance and the performance within the simulated environment (Sheridan 1992). One item explicitly addressing the presence in regard to virtual objects within the same environment showed a statistically significant difference toward a preference for the cubes (see also Table 3).

### TABLE 3: Comparing Means and Standard Deviation, t-test for all items

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (σ)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The feeling of presence inside the simulation was compelling.</td>
<td>2.7 (0.83) 2.4 (1.09)</td>
<td>0.44</td>
</tr>
<tr>
<td>The other participant was naturally within my world.</td>
<td>2.4 (1.01) 2.7 (1.33)</td>
<td>0.43</td>
</tr>
<tr>
<td>I was so involved into communication and the assigned task that I lost track of time.</td>
<td>3.1 (1.03) 2.7 (1.33)</td>
<td>0.35</td>
</tr>
<tr>
<td>Events occurring outside of the simulation distracted me.</td>
<td>4.1 (1.10) 3.7 (1.27)</td>
<td>0.34</td>
</tr>
<tr>
<td>My senses were completely engaged during the experience.</td>
<td>2.6 (0.84) 2.6 (1.15)</td>
<td>1.00</td>
</tr>
<tr>
<td>Objects on the augmented urban design table seem to be part of my world.</td>
<td>3 (1.04) 2.1 (1.07)</td>
<td>0.03</td>
</tr>
</tbody>
</table>
4.4. COMMUNICATION PATTERN

There was no significant difference in the amount of utterances between the pen group (545 or 2.4 utterances/min per person) and the cube group (521 or 2.3 utterances/min per person) (see also Table 4). Nevertheless the communication showed different properties through the frequencies of utterances (see Figure 4). To investigate the communication patterns in more detail the sub items were measured regarding their density.

The absolute references (i.e. directly addressing objects in the simulation) showed a distinctive pattern for both settings with two peaks and a trough in between. The first peak accounts for the groups deciding for one of the proposals. The trough denotes the users changing the interface. And the second peak the discussion about the extension. Therefore the pen users used considerably more bandwidth in their conversation to discuss the design proposal.

TABLE 4: Utterances Data. Mean bandwidth in utterances/min per person

<table>
<thead>
<tr>
<th>Category</th>
<th>Pen</th>
<th>Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>References (R)</td>
<td>0.39</td>
<td>0.61</td>
</tr>
<tr>
<td>References in Context (CR)</td>
<td>0.40</td>
<td>0.32</td>
</tr>
<tr>
<td>Position (P)</td>
<td>0.53</td>
<td>0.34</td>
</tr>
<tr>
<td>Position in Context (PC)</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>Agreement for Understanding (AU)</td>
<td>0.64</td>
<td>0.59</td>
</tr>
<tr>
<td>Agreement for Behaviour (AB)</td>
<td>0.21</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Figure 4. Left (a): Density diagram for references made directly. Right (b): Density diagram for agreement behaviour. (Dashed lines denote cube-users, solid line pen users.)
5. Discussion

This user study used an explorative approach in order to gauge tendencies towards improvement or degradation between two interfaces. Furthermore, a user assessment regarding a design creation tool is not free from contextual influences. Therefore the data collected in this study might be only valid within the user population the experiment was conducted in. As reported there were a high number of ‘early adopters’ which is due to the specific curriculum at the undergraduate level at the HKU where students get exposure to various Virtual Reality systems and been taught to use them for design presentation.

In both conditions the participants finished earlier, however few groups took significantly longer to complete. The time frame for the design phase differed significantly. The cube users were using about double the time for design than the pen users. From the videos one can observe that the users in the cube condition spent more time on quickly discussing different (see Figure 5(b)) approaches, whereas the pen users were cautiously placing and moving cubes in the site (see Figure 5(a)). This difference can partly be accounted to the different affordances of the tools. The users with the pen needed to change to a different device. This change resulted in an adaptation to a different affordance of the tool in order to create objects. For the cube condition the users could perform the creation phase of the brief with the same learned knowledge of the first phase.

Figure 5. Left (a): Users in the pen condition planning ahead for the next move. Right (b): cube condition users working in parallel on a solution.
Furthermore, the cubes as spatially multiplexed input devices (Fitzmaurice and Buxton 1997) were easy to share. Users with the pen needed to take turns with the input device. Only one out of the seven pen groups was taking turns with the pen, the others were deciding for a “pen” operator. For collaboration this has an interesting effect. The users of the pen condition were relying on the communication and also showed that they were considerably more exchanges (or utterances) regarding the references which clarified the site for the extension. This observation therefore confirms an observation made by Vera et al. (1998) earlier that low bandwidth tools can enhance the communication and extends this notion into the domain of interfaces.

From an observation of the video it becomes clear that the pen users spent more time with communication on details. Additionally pen users were holding off the creation phase of the task (i.e. extension building) to verbally discuss the actual problem in more detail. The cube users were discussing aspects of the design in parallel. That means they went for a “trial-and-error” strategy. Therefore, throughout the experiment, cube users were considering a multitude of design decisions in breadth and in shorter time. It also shows a distinctive pattern for the agreement on behaviour. The pen condition relied heavily on one users talking to the “pen operator” about his intention and confirming actions in the virtual scene. Cube worked independent and therefore could leave out constant agreements. The structure of the user communication was affected by the paradigm shifts for the user interface and by the interface change per se.

In both cases the video account shows clearly that the users did not face problems with the system rather than exploited the possibilities to a full extent. These observations are an indication for the novelty of the interface, as it allows a naive but explorative approach to design. The pen interface supports working anywhere in space whereas the TUI is constraint to the physical world. The experiment also confirmed earlier research by Regenbrecht and Schubert (2002), which demonstrated that low-fi media can actually produce higher presence as long as their interaction paradigm matches the task at hand. The addition of extensions through virtual cubes placed through real cubes seems to be more ‘natural’ than through a point-and-click interface.

6. Conclusion and Future Work
Technologically, AR is still in its infancy. It is at this early stage that architectural design needs to embrace it as an opportunity. For architectural design the potential of AR goes beyond a visualisation tool. The combination of real and virtual elements makes an ideal context for a design team to examine spatial problems in a collaborative setting. Boundaries
between real and virtual will be pushed further than we can predict today but it is necessary to remind the designer that the equipment should be utilised as a utility and not a push-button solution. Creativity needs to be paramount and in the control for the designer. Interfaces need to provide the designer with a maximum of freedom and a minimum of pre-programmed logic in order to maintain rather than restrict creativity during the design process.

This paper demonstrated a unique approach for an experimental study into the nature of AR for urban design. It allowed an insight into the effect of user interfaces onto communication structures. Future research needs to address the factors of communication more deeply to create highly responsive feature rich work environments to aid the need for designers to comprehend spatial content.

Acknowledgements

This research was part of the authors’ dissertation under the supervision of Prof. Thomas Kvan and Dr Justyna Karakiewicz who gave invaluable input. Additionally the author would like to thank Dr Marc Aurel Schnabel and Dr Mark Billinghurst for their critical comments. This research was made possible through a PGRS grant of The University of Hong Kong.

References


