# Exploring Augmented Reality for Situated Analytics with Many Movable Physical Referents

Aimée Sousa Calepso*
University of Stuttgart
Stuttgart, Germany
Aimee.SousaCalepso@visus.uni-stuttgart.de

Philipp Fleck*
Graz University of Technology
Graz, Austria
philipp.fleck@icg.tugraz.at

Michael Sedlmair
University of Stuttgart
Stuttgart, Germany
Michael.Sedlmair@visus.uni-stuttgart.de

Dieter Schmalstieg
Graz University of Technology
Graz, Austria
schmalstieg@tugraz.at
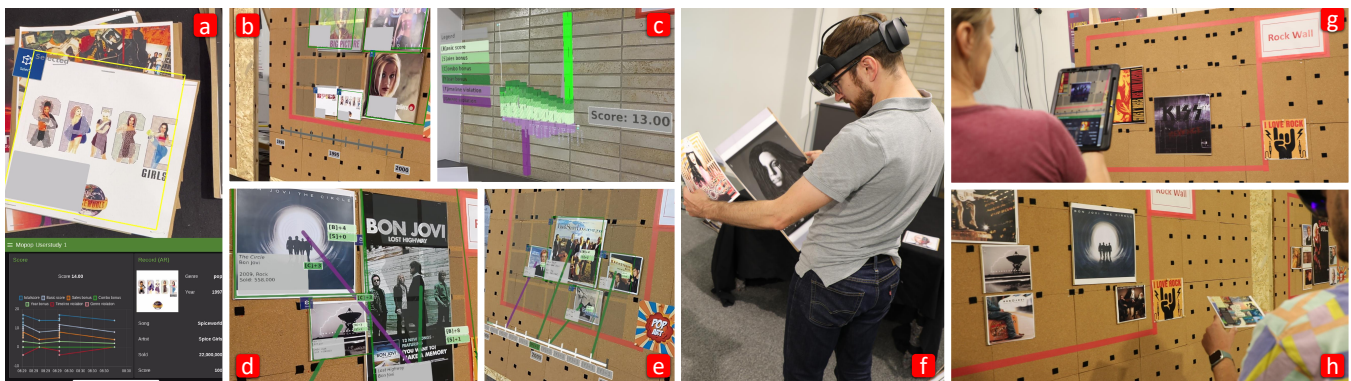
Figure 1: Images taken from our second study on SitA. We compared (a) a 2D mobile analytics interface on an iPad with (b-e) a SitA interface on a HoloLens 2. (b, e) The SitA interface enriches exhibits placed on an exhibition wall directly where they are located. Additional information included, for instance, (d) leader lines depicting the chronological order of the exhibits on a (e) timeline, and the evolution of the (c) participant's score as a barchart visualization. The images on the right show the participants using our application (f, h) on the HoloLens 2 and (g) on the iPad, while (f) scanning or (g, h) placing exhibits on the exhibition wall.

## ABSTRACT

Situated analytics (SitA) uses visualization in the context of physical referents, typically by using augmented reality (AR). We want to pave the way toward studying SitA in more suitable and realistic settings. Toward this goal, we contribute a testbed to evaluate SitA based on a scenario in which participants play the role of a museum curator and need to organize an exhibition of music artifacts. We conducted two experiments: First, we evaluated an AR headset interface and the testbed itself in an exploratory manner. Second, we compared the AR headset to a tablet interface. We summarize the lessons learned as guidance for designing and evaluating SitA.

*A. S. Calepso and P. Fleck contributed equally as leading authors to this work.

## CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Visual analytics**; **Empirical studies in visualization**; **Visualization design and evaluation methods**.

## KEYWORDS

Immersive analytics, Situated analytics, Augmented Reality

## 1 INTRODUCTION

Sensemaking using visual analytics (VA) tools on desktop computers is now a mature discipline with practical relevance. In contrast, many fundamental aspects of *immersive analytics* (IA) have not

been systematically studied [42, 53]. Yet, IA has arguably demonstrated benefits for both spatial data [24] and abstract data [34], which warrants a deeper investigation.

*Situated analytics* (SitA) is a close relative of IA, focusing on sensemaking tasks related to the physical environment of the user [56]. While IA makes the inclusion of physicality optional, SitA makes it an essential aspect of the sensemaking process. While situated analytics can be implemented using a variety of display technologies, e.g., small embedded displays [7], this paper focuses on **augmented reality** (AR) [51] **as an enabling technology for situated analytics**. We feel that the connection of the virtual and physical world – an essential aspect of AR according to Azuma's classic definition [2] – has not yet been sufficiently explored in the larger VA context.

Arguably, the sensemaking requirements of users working on physical tasks are different from those working on purely abstract tasks. Desktop computer users need to spend negligible effort on coordinating physical actions. Even more importantly, the desktop user does not need to alternate between epistemic and pragmatic actions [33], i.e., between interpreting the data and making changes to the physical world.

In contrast, the defining property of SitA is the user's engagement with physical *referents* [65], i.e., relevant objects in the real environment. Hence, the interleaving of physical and analytical actions is not only possible but even mandatory. Such a combination is intrinsic to many everyday activities. Even mundane tools are subject to Norman's gulfs of evaluation and execution [47]: For example, *a cook must constantly switch attention between stove and cookbook*. Activities situated in the physical environment are never purely analytical and, consequently, can be improved by narrowing the gap between digital information and physical reality.

The potential of situated analytics for such scenarios has been proclaimed many times [42, 64, 65]. At the moment, however, we have little empirical evidence that grounds these promises in actual studies, a deficiency that has recently been described as one of the "grand challenges of immersive analytics" by Ens et al. [17]. One reason for the insufficient evidence is certainly owed to technical limitations of existing AR technology. To deal with these limitations, realistic scenarios need to be considerably simplified in order to implement and study them. This simplification often happens on the *physical* side. Typically, only a very small number of physical referents is used in SitA interfaces [19]. Another simplification is that referents have little semantic meaning for the analytical task, such as when projecting a visualization onto a piece of paper [30].

We want to push these boundaries by allowing the study of SitA under more realistic conditions. We address this goal in a testbed that allows **study situated analytics with many movable referents**, leading to an increase in the complexity of the physical task. To this end, we introduce a scenario that fosters enough physical and analytical depth. We chose a museum curation task (Figure 1) and conducted two experiments in this scenario. First, we developed and tested a SitA application using an AR headset in an exploratory study. Second, we improved the SitA application and conducted another experiment, comparing the headset AR interface to a mobile tablet interface. From both experiments, we report lessons learned to inform researchers seeking to design and evaluate situated analytics. Our contributions are as follows:

- an evaluation scenario for situated analytics (museum curation),
- two interface designs to support this scenario (a situated interface with a headset, a mobile interface with a tablet),
- two user experiments, in which we study these two interfaces in the given scenario, and
- lessons learned for designing and evaluating situated analytics.

## 2 BACKGROUND AND RELATED WORK

We briefly review the state of the art in immersive and situated interfaces for visualization and analytics.

### 2.1 Immersive, spatial, and mobile visualization

The introduction of inexpensive VR headsets has sparked a new wave of research into immersive analytics (IA). Bowen and McMahan [5] list several advantages which VR visualization can have over conventional desktop visualization: Increased spatial understanding can be facilitated by additional depth cues. Virtual displays provide plenty of space, reducing information clutter and supporting peripheral awareness. Together with enhanced support for spatial memory, VR displays can drive the user's perception with higher information bandwidth. Positive effects of IA have not only been demonstrated for scientific visualization [24], where the use of 1:1 spatial encoding of data is prevalent, but also for abstract data, *e.g.*, node-link diagrams [34]. Especially tangible manipulation can be instrumental in building multi-dimensional views [3, 14].

We note that spatial interfaces in visualization are not exclusive to VR, but have, for instance, also been implemented via large wall-sized displays [1]. These spatial interfaces differ from VR in that they restrict the placement of visualizations in free space [50], but in return do not require the user to be equipped with a headset [38]. In contrast, researchers have also started to explore how mobile and HMD technologies can be combined for a better experience in data analysis [11, 31, 36]. The combination of mobile devices and AR allows for extending 2D displays into 3D and offers a tangible prop for interaction.

Both desktop and VR visualizations lack one important property – mobility. Even a "naive" mobile visualization, i.e., a visualization displayed on a device with a mobile form factor, such as a smartphone or tablet, already unlocks enormous opportunities. After all, the community of smartphone users is one order of magnitude larger than that of desktop users, and these mobile users can employ visualization at any location of their choosing. The goal of our work is to further unlock the potential of more mobile visualizations, by offering a new testbed for SitA.

### 2.2 Situated visualization

Compared to the aforementioned visualization interface styles – immersive, spatial, mobile –, AR offers the unique opportunity of making visualizations spatial and mobile at the same time. Virtual views presented in an AR display can literally be situated anywhere in the environment, including in mid-air [18, 20]. Not only does this freedom make AR display more versatile, and, ultimately, cheaper than physically embedding displays in our environment, it also paves the way for *situated visualization*. White [62] introduces

the term *situated visualization* to describe a visualization that is intrinsically related to its physical environment. He describes key characteristics that a situated visualization must have:

(1) Data in visualization is related to physical context
(2) Visualization is based on relevance of data to physical context
(3) Display and presentation of visualization lies in physical context

The relationship to physical context (or, reality) is what sets situated visualization apart from immersive visualization. Immersive visualizations (and, by extension, immersive analytics) consider "the use of engaging, embodied analysis tools to support data understanding and decision making" [42], but typically do so in a purely virtual environment.

In contrast, situated visualization techniques can use precise spatial information to *embed* [65] visualizations into the perception of referents. If data comes with intrinsic spatial characteristics, we can directly overlay its visualization onto a physical referent. Such visualizations were called embedded visualizations [65] and can reveal the correspondences between real and virtual dimensions, *e.g.*, for temperature [26], Wi-Fi signal strength and seating accessibility [25], pollution levels [63], water levels [57], viticulture [32], geological formation [37], corrosion [59], construction progress [66], or CAD models [29]. Others have also tried to leverage situated visualization for abstract data without an inherent spatial dimension, *e.g.*, tourist maps [9], charts [10], nutritional information [16], citation data in academic articles [39], free-form annotations on post-its [54], or everyday activities [6]. Some studies explore how the setup influences the way that people interact with immersive content [41].

## 2.3 Situated analytics

SitA depicts analytical settings in which visualizations are perceived in close temporal and spatial proximity to the respective physical referents [56]. This definition implies that we know the referents' spatiotemporal location, but it also assumes analytic purposes. Without getting into a debate about the exact requirements for a use case to qualify as "analytics", we observe that desktop, VR, and mobile interfaces must derive their sensemaking complexity exclusively from the digital data. In contrast, a situated visualization/analytics application may very well have lightweight use of visualization, but high complexity in the user's sensemaking engagement with referents in the real world – or, epistemic actions in the language of Kirsh [33]. This consideration implies an important change of perspective, possibly shifting not only the focus of the technology but redefining its entire scope of application.

Many AR researchers have considered visual instructions that support physical activities, such as assembly and maintenance [60]. Such instructions should definitely be considered as situated visualizations. Various authors have compared situated visualizations with visualizations delivered on paper [55, 61], via stationary screens [28, 55], video [40] or side by side with the referents (as opposed to embedding the visualization) [4]. However, all these works focus on supporting the user in performing pragmatic actions and include little or no support for epistemic actions. In fact, making it easy to blindly follow instructions may encourage a user to adopt a mechanical workstyle while investing as little cognitive effort as possible. Consequently, the evidence provided by these

works concerning the advantages of situated visualizations over conventional visualizations may not extend to scenarios where SitA helps users in cognitively demanding sensemaking. Few such scenarios exist in the literature, which is why we set out to further explore the area.

## 3 MOPOP: AN EXPERIMENTAL TESTBED

Our main goal was to *create a testbed to study situated analytics in AR with high physical complexity, involving many movable referents*. With such a testbed, we can design and implement applications and investigate various aspects in user experiments. In this section, we first elaborate on the requirements that a suitable scenario must have. Then we describe the scenario and the associated task objectives that we created to address the requirements mentioned above. In section 4, we describe the initial AR interface design. In section 5 and section 6, we describe two experiments conducted in our testbed. Throughout this process, our main intention was to collect *lessons learned* on how to design and evaluate SitA interfaces, which we summarize in section 7. Due to the inherent technical limitations of current AR devices [21, 49] as well as the novelty of the field, we expected many challenges could be captured.

## 3.1 Requirements

We identified four requirements for a scenario suitable for SitA:

*Combined physical and analytical task.* We wanted to explore a scenario that truly places the referents at the center of attention: Not only did we want the user to deal with a large number of referents, but we also wanted a task that requires the user to physically move the referents around. Especially the latter aspect makes such a scenario distinct from existing immersive analytics scenarios. IA often works on a digital twin rather than using any real objects as physical referents. The emphasis on referents implies that the user's attention during the sensemaking is split between physical and virtual (i.e., analytical) actions. We are interested in understanding whether the potential adversary effects of this split can be mitigated by SitA.

*Spatially situated data.* Traditional visual analytics and, by extension, immersive analytics often rely on datasets that have no direct relation to the spatial surroundings (for instance, newspaper articles or tabular sales data). In contrast, the scenario we seek must incorporate the physical world in a meaningful way. The essential requirement here is that the data needs to be situated in the physical environment of the user. The data itself can be either primarily abstract, *e.g.*, table data attached to IoT sensors [22], or spatial, *e.g.*, water levels [58].

*Balance between control and realism.* Ideally, the study benefits from the increased ecological validity of a real-world task location, such as an industrial shop-floor [29] or a building maintenance area [48]. Alas, AR studies conducted outside of the lab are subject to many confounding and hard-to-control factors, such as noise, poor lighting, environments hostile to 3D tracking [44], or – in case of cooking – untidy or even dangerous side effects. Consequently, we seek a motivating, complex scenario that can be operated within a lab [45].

*Gamification.* While we desire a scenario with sufficient analytic depth, it must remain accessible to ordinary audiences (*e.g.*, the local student population), from whom we want to recruit experimental subjects. Recruiting genuine domain experts rather than subjects from the general population makes it easier to select a scenario affording analytic depth, but this strategy severely restricts the achievable sample size and the kind of experiment that can be conducted [52]. Choosing a thematic area that appeals to a broad population lets us adopt a fictional scenario that casts test subjects in an expert "role" and draws motivation from gamified tasks. For example, the widely adopted VAST challenges on text analytics [12] essentially cast subjects as forensic investigators. Gamification allows us to collect participants' performance as a quantitative measure.

## 3.2 Scenario

Based on these requirements, we chose a scenario that involves widespread and familiar referents: musical recordings. We drew inspiration from the Museum of Pop Culture, *MoPop*, in Seattle (USA). MoPop exhibits detailed collections of promotional artifacts, such as vinyl record sleeves, compact disc (CD) covers, and tour posters. Based on this scenario, we simulate the work of a curator. Specifically, we postulate that the curator's job is to design an exhibition that showcases the evolution of two musical genres (*Pop*, *Rock*) throughout a certain period (1990–2000).

We designed the tasks to involve both analytical and physical aspects. To select which exhibits should be placed on the wall, the user needs to go over the collection and analyze the information represented therein. Participants need to analyze the current state of the wall, too. This analysis includes taking into consideration the information about the year, number of sales, and artists associated with the exhibits that have already been placed. When a change is made (adding or removing an exhibit from one of the walls), the information on the score chart is updated accordingly.

The exhibition wall is subdivided into a regular grid to ease placement. Free online databases, *e.g.*, MusicBrainz, let us assemble a rich collection of artworks without design cost. We have access to information related to release data, cover art, and genre for each of the exhibits. The acquired data points are used in visualizations that inform the curator in the design task.

## 3.3 Task objectives

The historical or aesthetic value of an exhibition hardly has a clear optimum that lends itself to quantitative benchmarking. To introduce a more measurable objective, we formulated a set of rules for the exhibition design which could be expressed in a numerical score. Subjects were encouraged to aim for a maximum score.

The rules are intended to convey to the curator a set of objectives, such as how to best use the available space, how to determine the most iconic works of a period, and how to establish a sense of the timeline. The curator must consult meta-data about exhibits (number of sales, release year, genre), making it necessary to interleave analytic thinking with the physical task of arranging exhibits. Instructions to the curator impose constraints on the exhibition design that propagate through the whole exhibition, since one picked

**Table 1: The score $S$ is computed as a sum of all the positive and negative aspects derived from a curator's design. Only exhibits which do not violate the rules contribute positively to the score.**

| Element | Category | Description | $S$ |
|---------|----------|-------------|-----|
| CD | Basic score | Correct wall | +1 |
| Vinyl | Basic score | Correct wall | +4 |
| Poster | Basic score | Correct wall | +8 |
| Any | Sales bonus | > 1M copies sold | +1 |
| exhibit | Sales bonus | > 10M copies sold | +2 |
| Adjacent | Combo bonus | Same artist | +3 |
| exhibits | Combo bonus | Same format | +1 |
| Timeline | Year bonus | Every year covered | +2 |
| In column | Timeline violation | Year mismatch | -1 |
| Columns | Timeline violation | Year mismatch | -1 |
| Any exhibit | Genre violation | Wrong wall | -50 |

artifact will influence the selection of consecutive ones. We define that as orthogonal constraints, and take care to keep them simple and measurable so that we can evaluate user performance accordingly. Specifically, we imposed the following rules:

- Two genre walls, one for *Pop* and one for *Rock*, show the art of the chosen period sorted by year from left to right.
- Each wall is subdivided into a grid with five rows and five columns. A CD cover is $1 \times 1$ grid cells wide, a vinyl sleeve, $2 \times 2$, and a poster, $2 \times 4$. Hence, a CD partially occupies one row, a vinyl partially occupies two rows, and a poster occupies two columns.
- The musical collection consists of 100 exhibits, 50 per genre, which are split into 25 CD, 20 vinyl and five poster exhibits.
- A basic score is awarded for every exhibit placed on the correct genre wall. Larger exhibits yield a better score.
- Bonus points are awarded for exhibits with high selling numbers.
- A further "combo" bonus is granted for exhibits which are immediately adjacent and feature music by the same artist or have the same format (i.e., CD, vinyl, poster).
- A bonus is awarded for every year covered by at least one exhibit.
- The exhibits must be sorted by year in ascending order. In one column, only a single year may be present, and it must be the same or larger than the year used in the preceding column. Exhibits covering two columns (vinyl, poster) force the same year across both columns. Hence, the 11 different years (1990–2000) could not be represented in the five available columns, compelling the curator to make a certain trade-off.
- Violations of the timeline incur a penalty.
- Placing an exhibit on the wrong wall incurs a large penalty.

Based on these rules, participants are then asked to optimize their score in a given timeframe, *e.g.*, in 1 hour in our first study. The score computation is summarized in Table 1.
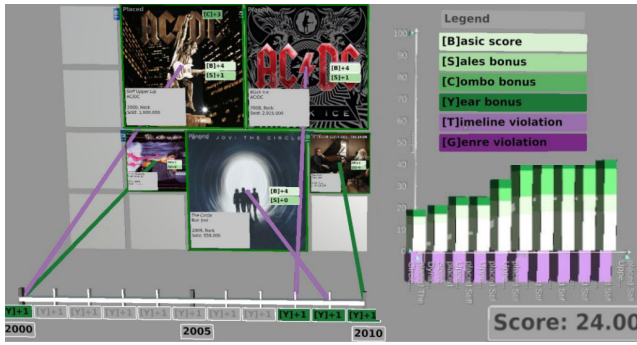
**Figure 2: Visualizations in the curator's interface (shown without a video background for better clarity): (left)** *Rock* **exhibition wall with overlays, including** *leader-lines* **to indicate the year and boxes (green) to indicate bonuses, (top right) legend explaining the color code used for the score, (bottom right); stacked chart visualizing the score over time.**

## 4 HEADSET INTERFACE

To explore the testbed, we designed a SitA application for an AR headset. The headset interface presents several interactive visualizations that inform about the exhibition currently being designed. Figure 2 shows the main visualizations of the final headset interface (without a video background for better clarity), while Figure 3 shows the first version of the headset interface through a Microsoft HoloLens 2. Such an optical see-through headset demands that visualizations be robust against visual clutter. Guaranteeing such robustness is challenging and makes the design differ considerably from visualizations intended for desktop computers or mobile devices. Labels need to be much larger, colors chosen carefully to not interfere with the background, and the overall design needs to be substantially simplified [35].

The interface components and visualizations were designed to support the primary tasks of our scenario. They should help communicate the influence of the curator's decisions on the score (subsection 3.3). As a first goal, a curator will explore the physical exhibits. When an exhibit is gazed at, the interface instantly shows the metadata as a *preview* overlay in AR, conveying important information on the artist, title, year, sales, and reputation (Figure 5a).

Next, the curator places exhibits on the wall. To support this step, we show an occupancy grid (Figure 2, left, and Figure 3) that marks empty and occupied elements. Below each wall, a virtual *timeline* is shown to help keep the exhibits in chronological order. *Leader lines* connect the exhibits to the timeline and establish a sense of order. Crossings of leader lines indicate timeline violations, which should be avoided (subsection 3.3).

Next to the exhibition walls, a score chart is placed (Figure 2, right, and Figure 3), which shows the evolution of the score over time. Here, the score is shown cumulatively. In the final version (Figure 2), we broke it down into categories, which are visualized as stacked bar charts. Every placement or removal of an exhibit is graphed as an event changing the score and annotated with a summary of how the score changed as a result of the curator's action.

When an exhibit is changed (placed or removed), the action is picked up by image recognition after a brief dwelling period. When a change action is detected, a confirmation sound is played, and the color of the corresponding grid position changes accordingly. Moreover, the score and all visualizations are updated.

We implemented the first version of the headset interface on a Microsoft HoloLens 2, using the RagRug toolkit [23]. The AR visualizations are presented using an extended version of IATK [13] in Unity 3D, which runs on current AR devices. To detect and track exhibits, we used Vuforia for image recognition. The maximum size of the Vuforia recognition library is 100 images, which was deemed sufficient for our scenario.

## 5 EXPLORATORY USER STUDY

Using the headset interface, we conducted a first study in the form of an exploratory experiment. By observing users and by gathering qualitative and quantitative feedback, the primary goal of our first study was formative: What works well? And what not? What strategies are used in situated analytics? What challenges are still ahead of us?

### 5.1 Methods

*Setup.* We performed the study in two locations, one in Venue B and the other one in Venue A. In each location, we set up the scenario as illustrated in Figure 3. As the goal of the study was exploratory, we decided to split the locations in order to get more diverse feedback from participants with different backgrounds. At one location, people had more expertise with computer graphics-related AR topics. The other location had participants coming from institutes more focused on visualization and simulation topics. We thought that splitting sites would be more effective in diversifying participants instead of gathering many people in a single location. In addition, it would be a good opportunity to see how replicable the study setup and results would be since AR strongly depends on environmental factors such as lightning, tracking, calibration, *etc.* We made sure that the settings were as similar as possible, so that we could analyze the results as a single experiment without major confounding factors.

*Design, procedure, tasks.* After collecting consent and demographic data, each participant was introduced to the system with a video tutorial and instructions about the task and visualizations. They were given 1 hour to complete the task.

*Collected measures and data.* During the experiment, we collected the time needed to complete the task, the score generated in the end, the evolution of the score over time, and the changes to the wall that led to the final score. We also encouraged participants to comment on their experience while working. We recorded audio and additionally took notes of the subjects' comments. At the end of the experiment, we asked them to complete standard SUS [8] and NASA TLX [27] questionnaires to rate the overall user experience. . We added 3 questions targeting the four specific visualizations: 1) I used this visualization frequently, 2) I found the visualization unnecessarily complex and 3) I felt very confident using the visualization.
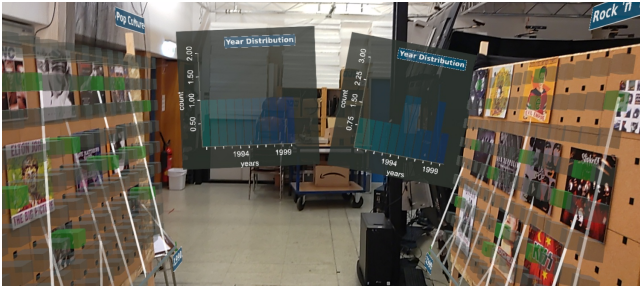
**Figure 3: Results of the exploratory user experiment using the initial version of headset interface, showing the histogram visualization of a finished exhibition.**

*Participants.* Overall, we tested 16 participants in the first study (14m/2f). As the primary goal of the study is formative, we opted for a small sample size, following the typical practice for such studies in HCI [46]. Two of the participants (1m/1f) stopped the study prematurely, and we excluded them from the analysis of the results, so 14 participants remained, 6 in Venue B, and 8 in Venue A.

## 5.2 Results

*Scores.* We collected quantitative data to better explain and judge the outcome. The best participant scored 325, while the worst only scored 16.8 points. On average, the subjects reached a score of 176.4 ($std = 114.4$, $median = 188.825$). None of the participants stopped early; all used the full hour. On average, the participants picked up exhibits 172 times. The three participants with the highest score had pick-up numbers close to or below average (172, 82, 85 times). In contrast, the participants with the lowest scores picked most often (219 and 258 times). These participants extensively applied physical sorting, apparently without consulting the score visualization or otherwise following a clear strategy to include the available virtual information.

*Usability.* Most participants judged the visualizations easy to understand: For the wall (11/14), score (10/14) and cover (14/14) visualizations, the majority *disagreed* or *strongly disagreed* with the statement "I found the visualization unnecessarily complex". After a closer look at the data, we detected an interesting bimodal distribution of user experience feedback. While the feedback to the SUS questionnaire from the participants in Venue A was consistently enthusiastic (Figure 4, right), the SUS feedback from the participants in Venue B was mixed and much more overall negative (Figure 4, left). Due to the exploratory nature of the study, we can only speculate on the reasons. These results likely stem from a larger number of technical (tracking) issues in Venue B. It is a known challenge in the AR community that controlled studies are very sensitive to even seemingly small imperfections of AR technology [44]. It is important to keep these sensitivities in mind when designing and interpreting future studies on situated analytics.

*Strategies for problem-solving.* We were also interested in identifying strategies for problem solving that are used in our scenario. As expected, the score and year histograms are the core elements to complete the task (Figure 3). The highest-scoring participant

aimed for a uniform year histogram and selected the highest-selling records as exhibits. After becoming familiar with the interface, most participants (11/14) quickly adopted a routine of mixing physical actions with virtual reactions and observations: (1) select wall; (2) look at empty spots, identify missing year; (3) search for fitting exhibits; (4) place exhibit and verify score; (5) repeat until wall is completed. We conclude that our requirement for a mixed routine was met (see subsection 3.1 and that our subjects were able to quickly explore the space of interactions available to them, and the headset interface was essential for any successful strategy.

## 5.3 Limitations of first version of headset interface

As both the field of research (SitA) and the study testbed (MoPop) were novel, we expected to encounter various challenges in this first study. We specifically identified the following design and testbed issues that impaired the user experience in performing the tasks.

*Tracking limitations.* In general, the tracking in the AR headset is satisfactory. Yet, when tracking over larger distances (several meters) and longer periods of intense interaction, limitations become evident. Registration with the real world may be affected by subtle drift, and features such as hand detection may not always work well under fast motions. We observed that users are very familiar with manipulating physical referents (*e.g.*, quickly leafing through a pile of vinyl covers) and get confused when the AR system has difficulty catching up with fast motions.

*Interaction mechanics.* In our first implementation, changes on the exhibition wall (i.e., placement and removal of exhibits) were purely detected by the proximity of an exhibit to the wall. This frequently led to missed detections, where the user performed an action too fast without the system registering the change.

*Visualization design.* A critical point we identified with regard to visualization was displaying the development of the score over time. We realized that the score was composed of multiple components, but the original visualization showed only the cumulative score, not the individual components. This omission made it difficult for participants to understand how their actions affected the overall score.

*Information density.* We also noticed that we were overly ambitious with regard to the amount of information displayed for the participants and the amount of interaction required to process this information. Compared to what is on display in a traditional record store, a selection of 100 musical recordings seemed a lower limit for a realistic scenario. However, we observed that participants struggled to browse all the exhibits and felt overwhelmed by the amount of work presented.

*Score system.* We observed that our attempt to gamify the curation task by scoring the user's choices led to some confusion. Participants were able to understand the basic rules of sorting by genre and year, but were uncertain about the more advanced rules, such as the bonus system.
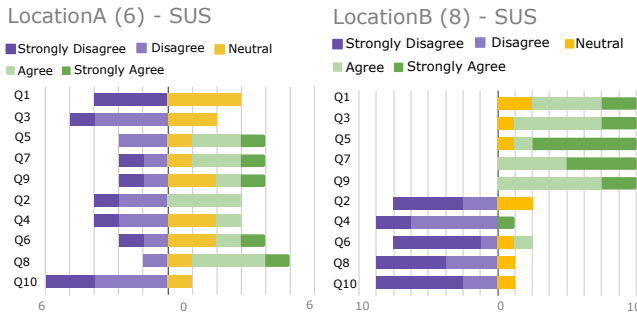
**Figure 4: Feedback from SUS for both locations regarding the AR interface. Questions are ordered according to SUS-intrinsic direction (rows 1-5: higher=better, rows 6-10: lower=better).**

## 6 COMPARATIVE USER STUDY

Based on the positive feedback from the participants of the first study, we conducted a second user experiment. Our goals were twofold. On the one hand, we wanted to fix the design and testbed issues that surfaced in the first study and to evaluate the effectiveness of these changes. On the other hand, we wanted to move from a primary focus on usability to a more utility-based, comparative focus.

Eventually, we decided on a comparison of the AR headset interface with a more traditional tablet interface. We first considered implementing an alternative interface on a stationary screen placed between the exhibition walls. This setup would be very straightforward and present the information in a traditional way. However, making the system aware of the changes to the wall would be difficult. Attempts to place stationary surveillance cameras to oversee the walls and detect any changes are technically not feasible with the Vuforia library, because image targets need to be seen up close for successful detection. We also ruled out a Wizard of Oz solution [15] that employs a human for event detection, since it would be too labor-intensive and error-prone.

As a feasible alternative, we settled on designing a tablet interface. Carrying a tablet computer occupies one hand of the user, which can affect the user's ability to manipulate referents. This drawback is partially compensated for by the freedom to look at the visualization anytime without having to return to a stationary display.

### 6.1 Improved headset interface and scenario

We set out to improve the headset prototype design and the scenario in order to address the limitations identified in the first study (subsection 5.3). In terms of the visual interface, we specifically worked on improving the visual quality compared to the prototype used in the first study. The redesign introduced the following main differences, resulting in the visualizations presented in Figure 2 and Figure 5:

*Tracking limitations.* We enhanced tracking performance by optimizing the application code so that it used fewer resources of the rather overloaded HoloLens 2 hardware. This not only increased the responsiveness of exhibit detection, but also reduced automatic

thermal throttling of the hardware after extended use. Moreover, we modified the working area by adding additional light sources to ensure a uniformly lit, bright environment, and we removed any strongly reflective or featureless surfaces to help the HoloLens 2 self-tracking.

*Interaction mechanics.* To minimize problems with missed detection of changes to the exhibition wall, we added a virtual "intent button" overlaid on every exhibit. By pressing the button, the user signals the "intent" to place the exhibit on the wall, or remove it from there. The change only becomes effective after the user *actually* adds or removes the exhibit; tending to another exhibit instead resets the intent button. This strategy imposes a small amount of extra work (pressing the intent button) on the user, but almost completely resolves the issue of missed events.

*Visualization design.* We changed the score chart from a simple line chart showing only the cumulative score to a stacked bar chart showing each component of the score separately. We did so to provide users with a more transparent way of understanding the changes in scores over time. This change was accompanied by the introduction of a color code for the score components, with the bonuses using shades of green and the penalties using shades of purple. This color code was used throughout all visualizations, *e.g.*, on the timeline to color the leader lines and the respective axis labels.

*Information density.* The scenario was altered to be less overwhelming for the participants, introducing a smaller exhibition wall and a shorter curation time. We also reduced features that were not considered essential. Furthermore, we selected 100 additional records (2000-2010) for one set of records per condition.

*Score system.* We simplified the scoring system. The original score was based on multiplying certain bonus factors and behaved in a non-linear manner that was hard for participants to understand. The new scoring system introduced a fixed positive score per exhibit or bonus and a fixed negative score per penalty. The final score is just the sum of these individual scores.

### 6.2 Tablet interface

From a technical point of view, the tablet computer provides a straightforward solution to the issue of detecting changes on the wall. Using the tablet computer's camera, we run software for self-tracking and for recognizing exhibits in the same way as on the HoloLens 2. The user is instructed to scan every exhibit with the tablet camera when picked up, which is confirmed with a sound.

However, we redesigned the tablet interface compared to the headset interface to be more conservative. The tablet interface, while providing a video see-through mode, makes only minimal use of embedded visualizations. The main visualizations on the screen, which show the score (Figure 5h, l) and the meta-data of the scanned exhibit (Figure 5i) are purely 2D. Video see-through AR is only used to show an outline of the exhibit being scanned (Figure 5g), gray squares for empty spots on the wall (Figure 5j) to aid users in proper placement, and the basic timeline without bonus highlights (Figure 5k). These elements are essentially static

**Figure 5: Visualizations for the Head condition (HoloLens 2) on the left and the Tab condition (iPad Pro) on the right. Situated information is not available in the Tab condition; only visual cues which are necessary to perform the task are kept in 3D. For example, the tracked border of the exhibit is highlighted to provide feedback on the successful recognition of objects. (a) A tracked exhibit in Head shows detail information on the bottom left and the "select" button on the top left, while (g) the tracked exhibit in Tab does not present detail information; (i) the detail information in Tab is shown as a 2D panel. (b) Head shows the color-coded score chart, with the legend to the left and the current score to the right, on a nearby wall, while (h) Tab shows it as part of the panel mentioned above. (c) Head introduces the leader lines from the exhibits to the corresponding year, while the leader lines are missing in Tab. (d) Head shows bonus scores for an exhibit placed on the wall, while (i) Tab shows only the border without additional information. (e) The timeline in Head indicates the year bonus (green "Y+1"), (k) which is missing in Tab. (f, m) The "recalibration" markers for the pop and rock wall. Similarly, the (j) grid is also the same under both conditions.**

visualizations necessary to navigate the exhibition space and could have been printed permanently on the wall.

As a consequence of this design decision, our two conditions, while technically equivalent, appear very different to the user: One emphasizes spatial visualization and interaction, while the other one relies on traditional space-agnostic visualization and interaction. We chose this design to elicit reactions and comments specifically on the utility of embedded visualizations.

## 6.3  Methods

*Setup.* Our setup was the same as in study 1, with the improvements described above. Study 2 was carried out in a single location.

*Design, procedure, tasks.* We designed this study as a within-subjects experiment, with interface type – headset condition (Head) or tablet condition (Tab) – as the sole independent variable. In order to reduce  learning effects, we counterbalanced the starting condition.

Each participant started with a brief introduction to the scenario and an explanation of the tasks to perform. We also explained the score mechanism and how it is influenced by moving exhibits. To minimize the overwhelming effect that the high information density had on the participants in the first study, we started each condition with an in-game tutorial. The walls were initially occupied with a few exhibits, which yielded a starting score ≠ 0. We intentionally made the preset exhibits cause rule violations, in order to educate the participants on the scoring system. We then gave the participants four distinct tasks:

$T_1$  Fix genre violation: A pop exhibit was placed on the rock wall. We asked participants to identify it and remove it from the wall.
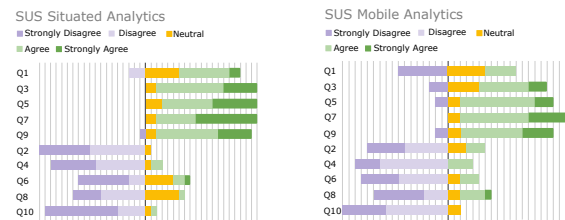


**Figure 6: SUS results for both conditions of the study**

$T_2$  Fix timeline violation: An exhibit was placed five years from where it should be. We asked participants to find a better spot.

$T_3$  Group placement: The combo bonus was introduced by placing exhibits of the same artist on the wall. From the collection of exhibits in storage, we offered three exhibits and asked participants to pick one to be placed to obtain a bonus.

$T_4$  Free curation: Following the introductory tasks, we asked the users to continue populating the wall with exhibits of their choice, with the overall goal of maximizing their score.

These tasks allow us to perform a quantitative and qualitative comparison of how users perform. Users were instructed to work on tasks $T_1$ to $T_3$, followed by $T_4$ after completion. Support for questions and technical problems was offered throughout the session. The session ended when the user was satisfied with the exhibition, or 15 min after starting $T_4$. Questionnaires regarding usability of interface and visualizations, along with written feedback, were examined afterward. The procedure was repeated for the other condition. Overall, the experiment took around 50 minutes.

*Collected measures and data.* We collected the final score, as a proxy to measure how successful the participants were in following the instructions. The maximum achievable score based on the accumulation of all points and bonuses was 100 for each condition. After completing each condition, participants completed questionnaires for SUS and NASA TLX. . After concluding both conditions, we conducted an open interview to obtain qualitative feedback from users.

*Participants.* A group of 20 participants (aged 15 to 42, 5f/15m) from the campus student population participated in the experiment. All used a computer daily, with six having at least some experience with AR headsets, and 17 having at least some experience with tablets. During the evaluation, we had a single case of severe tracking failure on the HoloLens 2. We needed to remove this participant from the analysis of the score performance data.

## 6.4 Results

*Scores.* Head users obtained a score with $M = 49.7$, $SD = 19.3$, while Tab users obtained a score with $M = 36.8$, $SD = 17.3$. We performed a statistical analysis using a repeated-measures t-test with $\alpha = 0.05$. The results indicated significantly better performance with Head ($t(18) = 2.44, p = 0.025$). In addition, 5/19 participants performed better in Tab than in Head. Of these five, only one performed Tab first, which may indicate that the other four may have benefited from the learning effect. Figure 8 illustrates the score differences between conditions visually.

*Usability.* On average, we collected SUS ratings in the "acceptable" range for both Head ($M = 78$, $SD = 13.56$) and Tab $M = 70$, $SD = 14.73$), following the same trend as the score values. In comparison, the usability preferences of the participants were significantly higher with Head ($t(19) = 2.19, p = 0.040$). Figure 6 illustrates the SUS answers. Q8 *"I found the application very cumbersome to use"* differed the most between the two conditions , suggesting that the non-hands-free operation of Tab was perceived as an issue, as could be expected.

*Task load.* NASA TLX lead to similar replies for both conditions. We observed a slightly higher physical demand in Tab ($M = 6.0$) as in Head ($M = 4.65$) and an impression of better performance in Tab (M=8.95) over Head (M=10). The score was visible for the duration of the task, but it did not influence the subjective perception of performance. A higher frustration in Tab ($M = 8$) over Head ($M = 6.1$) was reported. Overall, we found that our efforts toward providing a comparable environment for both conditions were reasonably successful. Figure 7 shows the results per category.

*Qualitative comments from participants.* Apart from complaints about tracking issues, we were surprised to find how much interaction with the devices differed from participant to participant, especially for Tab. Some of the common comments were *"I'm definitely enjoying this [Head] more"*, *"The immersion was also practical in the sense of less concentration needed"* and *"I was a bit confused that I had to click on the screen, I felt like I wanted to click in the real world"*, the latter being a feature only available in Head.
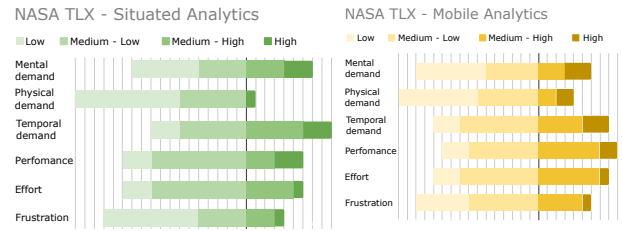


**Figure 7: NASA TLX questionnaire comparing both conditions**

## 7  LESSONS LEARNED

Combining all our experiences in implementing a testbed and running user evaluations, we distilled several lessons learned:

*Perceptual requirements of SitA are more complex than in desktop computing.* Transposing visualizations that were initially designed for 2D screens into a SitA scenario is usually not straightforward. Many factors in SitA influence the perception of a visualization, including display (headset, tablet), the location of viewing (*e.g.*, background clutter, lighting), the size and distance of the visualization (scale *vs.* distance) and the placement (referent, world, heads-up display). Many choices that are considered common sense in traditional visualizations may not be effective in SitA.

*Physical interactions with referents trump abstract interactions.* One of the most distinctive aspects of MoPop is that the user's interaction is dominated by the manipulation of physical referents. Pure physical interactions yield a physical result (*e.g.*, change of space), but also a virtual result (*e.g.*, triggering an artist-combo bonus by placing the same artist next to each other). Referents can also involve purely virtual interactions, but these are still presented in close proximity to the referent (*e.g.*, virtual buttons on an exhibit can only be pressed when the exhibit is within arm's length). Users do not respond well to ergonomically challenging tasks (*e.g.*, non-hands-free referent manipulation with a tablet), irrespective of whether they involve physical or virtual interactions. Overall, the design of the interaction in SitA always needs to consider the physical aspects of the interaction, and is hardly ever purely virtual.

*Less is more.* One important improvement from the first to the second interface version was a reduction in the amount of information displayed to users. In SitA, one can easily underestimate how overwhelming the cognitive processing of an augmented environment can be. Overlaying a lot of virtual content tends to increase clutter and may not provide the expected benefits. A possibly fruitful strategy relies as much as possible on physical referents and their properties, and includes as virtual objects (especially, visualizations) only those elements that cannot easily be deduced from observing the real world. We also noted that the tablet interface could be used sporadically to look at visualizations and otherwise ignored, while the headset interface unconditionally showed embedded visualizations. The reception in the headset interface improved when we redesigned it to include fewer overlays. Users may also benefit from additional configuration options that give them the ability to filter visualizations or marks on the fly.
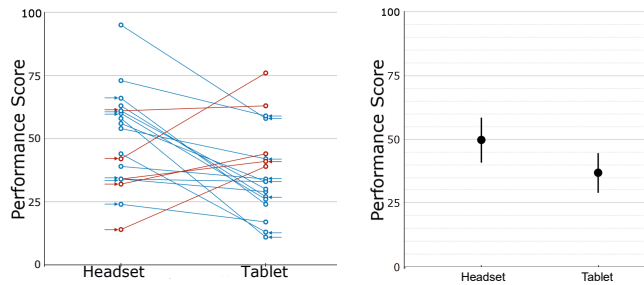
**Figure 8: Score difference between the conditions. Left: actual scores (0–100) of the 19 participants; lines connect the two scores of the same participant; arrows indicate which condition a participant started with; 14 participants performed better with SitA (blue) and 5 with mobile analytics (red). Right: Mean values with 95% confidence intervals; SitA performed better; average of 49.7 (95% CI 41 to 58.4) against 36.8 (95% CI 29 to 44.6).**

*Technical challenges in AR mask possible benefits of SitA.* The complexity of engineering SitA applications is high, with many more aspects to consider and control than in desktop applications. Alas, both development tools for SitA and experience with SitA development are in early stages. Users tend to expect mature interfaces, which can be hard to achieve if technical challenges cannot be reliably eliminated or worked around. This characteristic makes it even harder to replicate the same technical setup reliably at different locations and thus might add confounds to studies, as we have seen in our first study. The most obvious technical challenges still come from tracking. Even though commercial tracking is mature, drift can accumulate after tens of minutes. We compensated with on-demand drift reset, which was acceptable to users (but not ideal). We also changed the size of all real and virtual objects to be safe above 2× the size of the expected tracking error, so that the interaction was robust enough. There is also an inherent trade-off between optical tracking, which requires "feature-rich" environments, and visual design, which favors a "clean" background with little visual clutter.

*Repeatability and cost.* Most of the effort to create the testbed went into designing and implementing the software. Creating the physical setup (exhibits and sticky walls) was comparably light-weight after all required materials had been identified, acquired and tested. In this sense, creating a duplicate of the testbed of Venue A at Venue B for Study 1 was not too difficult. Since the technical equipment was already available in the Venue B lab, the additional monetary cost of physically creating the testbed was rather negligible. However, ensuring an optimal experience in Venue B turned out to be significantly harder. Even when running tried-and-tested software, the physical characteristics of a new testbed location require some amount of "debugging". The steps that require iteration include ensuring good illumination, sufficient wireless network reception, calibration of tracking coordinate systems, and several other tasks.

*Creating comparable conditions is non-trivial.* Another challenge we faced was the creation of comparable conditions for different

flavors of situated analytics interfaces. Trade-offs need to be made in designing meaningful physical tasks, while, at the same time, keeping the tasks measurable and understandable for the users. When the interface condition is altered (*e.g.*, when switching from a headset to a tablet), not one, but all factors that characterize the experience are altered. This entanglement of factors makes it difficult to isolate a single objective, while still keeping the overall experience motivating and non-artificial. For the tablet, the lack of hands-free operation seemed to be the strongest factor. This was to be expected since our focus on many movable referents creates a condition in favor of a headset. Nevertheless, the results indicated that other factors played a role as well, for example, the advantage of embedded over side-by-side visualizations.

## 8 CONCLUSIONS

This paper explores how situated visualizations can benefit analytic tasks in physical environments. We proposed a testbed that could be used to study the SitA interfaces. We also implemented two interfaces, headset and tablet, and ran two studies, exploratory and comparative. Throughout the process, we collected a first set of lessons learned. Our results are meant to inspire others in their own endeavors to design and evaluate SitA applications. Upon acceptance, we will also make all code freely available.

*Limitations.* As with all design and empirical work, our approach has limitations. Our work is exploratory; at this point, we focus on a specific scenario to design and test concrete instances of SitA interfaces. This approach naturally limits generalizability. Studies with other scenarios, tasks, interfaces, and visualizations are needed to gradually build up an understanding of SitA.

*Future work.* This work is just a starting point. Many aspects of SitA are still largely unexplored. For example, an important area to investigate is the placement of the visualizations relative to the physical referent. We also want to study how guidance can be best incorporated in an AR interface, for example, by including suggestions or navigation hints to the best exhibits in the visualization. Also, factors such as color, shape, and light conditions of the surroundings have an important influence on the performance of SitA interfaces. Systematically studying these factors will necessitate studies under different conditions. More work will also be needed to measure the amount of information that can be displayed without overloading the user. Another interesting direction is studying scenarios in which the gamified objective is replaced by a real-world objective. Such a change would require moving to a non-laboratory location like a real museum. At the same time, we also need studies in more strictly controlled environments that can reveal precise cause-and-effect relationships, as well as studies of different scenarios to foster generalizability [43].

# REFERENCES

[1] Christopher Andrews, Alex Endert, and Chris North. 2010. Space to Think: Large High-resolution Displays for Sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. New York, NY, USA, 55–64. https://doi.org/10.1145/1753326.1753336

[2] Ronald Azuma. 1997. A Survey of Augmented Reality. *Presence* 6, 4 (1997), 355–385.

[3] Andrea Batch, Andrew Cunningham, Maxime Cordeil, Niklas Elmqvist, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. 2020. There Is No Spoon: Evaluating Performance, Space Use, and Presence with Expert Domain Users in Immersive Analytics. *IEEE Transactions on Visualization and Computer Graphics* 26 (1 2020), 536–546. Issue 1. https://doi.org/10.1109/TVCG.2019.2934803

[4] Jonas Blattgerste, Patrick Renner, Benjamin Strenge, and Thies Pfeiffer. 2018. In-situ instructions exceed side-by-side instructions in augmented reality assisted assembly. In *Proceedings of the 11th PErvasive technologies related to assistive environments conference*. 133–140. https://doi.org/10.1145/3197768.3197778

[5] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual reality: How much immersion is enough? *Computer* 40, 7 (jul 2007), 36–43. https://doi.org/10.1109/MC.2007.257

[6] Nathalie Bressa, Jo Vermeulen, and Wesley Willett. 2022. Data Every Day: Designing and Living with Personal Situated Visualizations. In *CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3491102.3517737

[7] Nathalie Bressa, Kendra Wannamaker, Henrik Korsgaard, Wesley Willett, and Jo Vermeulen. 2019. Sketching and Ideation Activities for Situated Visualization Design. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 173–185. https://doi.org/10.1145/3322276.3322326

[8] John Brooke et al. 1996. SUS: A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[9] Z. Chen, Y. Su, Y. Wang, Q. Wang, H. Qu, and Y. Wu. 2019. MARVisT: Authoring Glyph-Based Visualization in Mobile Augmented Reality. *IEEE TVCG* 26, 8 (2019), 2645–2658.

[10] Zhutian Chen, Wai Tong, Qianwen Wang, Benjamin Bach, and Huamin Qu. 2020. Augmenting Static Visualizations with PapARVis Designer. *ACM CHI* (2020). https://doi.org/10.1145/3313831.3376436

[11] Neil Chulpongsatorn, Wesley Willett, and Ryo Suzuki. 2023. HoloTouch: Interacting with Mixed Reality Visualizations Through Smartphone Proxies. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–8.

[12] Kristin Cook, Georges Grinstein, and Mark Whiting. 2014. The VAST challenge: History, scope, and outcomes: An introduction to the special issue. , 301–312 pages.

[13] Maxime Cordeil, Benjamin Bach, Andrew Cunningham, Bastian Montoya, Ross Smith, Bruce Thomas, and Tim Dwyer. 2020. Embodied Axes: Tangible, Actuated Interaction for 3D Augmented Reality Data Spaces. In *ACM CHI 2020*. https://doi.org/10.1145/3313831.3376613

[14] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. 2017. ImAxes: Immersive Axes as Embodied Affordances for Interactive Multivariate Data Visualisation. In *ACM UIST*. 71–83. https://doi.org/10.1145/3126594.3126613

[15] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*. 193–200.

[16] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. 2015. Situated Analytics. *Big Data Visual Analytics* (2015). https://doi.org/10.1109/BDVA.2015.7314302

[17] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim Dwyer, Jens Grubert, Jason Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matthew Whitlock, and Yalong Yang. 2021. Grand Challenges in Immersive Analytics. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*. Yokohama and Japan. https://doi.org/10.1145/3411764.3446866

[18] Barrett Ens, Rory Finnegan, and Pourang Irani. 2014. The personal cockpit: A spatial interface for effective task switching on head-worn displays. In *Conference on Human Factors in Computing Systems - Proceedings*. 3171–3180. https://doi.org/10.1145/2556288.2557058

[19] Barrett Ens, Sarah Goodwin, Arnaud Prouzeau, Fraser Anderson, Florence Y. Wang, Samuel Gratzl, Zac Lucarelli, Brendan Moyle, Jim Smiley, and Tim Dwyer. 2020. Uplift: A Tangible and Immersive Tabletop System for Casual Collaborative Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–1. https://doi.org/10.1109/TVCG.2020.3030334

[20] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 2014. Ethereal Planes: A Design Framework for 2D Information Space in 3D Mixed Reality Environments. *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction* (2014), 2–12. https://doi.org/10.1145/2659766.2659769

[21] Åsa Fast-Berglund, Liang Gong, and Dan Li. 2018. Testing and validating Extended Reality (xR) technologies in manufacturing. *Procedia Manufacturing* 25 (2018), 31–38.

[22] Philipp Fleck. 2022. RagRug Github page. https://github.com/philfleck/ragrug.

[23] Philipp Fleck, Aimee Sousa Calepso, Sebastian Hubenschmid, Michael Sedlmair, and Dieter Schmalstieg. 2022. RagRug: A Toolkit for Situated Analytics. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–1. https://doi.org/10.1109/TVCG.2022.3157058

[24] Adrien Fonnet and Yannick Prí. 2019. Survey of Immersive Analytics. (2019). https://doi.org/10.1109/TVCG.2019.2929033

[25] Renan Guarese, Joao Becker, Henrique Fensterseifer, Aimee Calepso, Marcelo Walter, Carla Freitas, Luciana Nedel, and Anderson Maciel. 2020. A usability assessment of augmented situated visualization. In *2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*. IEEE, 636–637. https://doi.org/10.1109/VRW50115.2020.00169

[26] Ann Sofie Gunnarsson, Malinda Rauhala, Anders Henrysson, and Anders Ynnerman. 2007. Visualization of sensor data using mobile phone augmented reality. In *IEEE ISMAR*. 233–234. https://doi.org/10.1109/ISMAR.2006.297820

[27] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* 52 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[28] Steven J Henderson and Steven K Feiner. 2011. Augmented reality in the psychomotor phase of a procedural task. In *IEEE ISMAR*. 191–200. https://doi.org/.Augmentedrealityinthepsychomotor

[29] Dominik Herr, Jan Reinhardt, Robert Krueger, Guido Reina, and Thomas Ertl. 2017. Immersive Visual Analytics for Modular Factory Layout Planning. In *IEEE Immersive Analytics Workshop*. https://doi.org/10.1016/j.procir.2018.03.200

[30] David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. 2005. Paper windows: interaction techniques for digital paper. ACM, 591–599. https://doi.org/10.1145/1054972.1055054

[31] Jinbin Huang, Shuang Liang, Qi Xiong, Yu Gao, Chao Mei, Yi Xu, and Chris Bryan. 2022. SPARVIS: Combining Smartphone and Augmented Reality for Visual Data Analytics. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 111–117.

[32] Gary R. King, Wayne Piekarski, and Bruce H. Thomas. 2005. ARVino - Outdoor augmented reality visualisation of viticulture GIS data. In *IEEE ISMAR*. 52–55. https://doi.org/10.1109/ISMAR.2005.14

[33] David Kirsh and Paul Maglio. 1994. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science* 18, 4 (1994), 513–549. https://doi.org/10.1207/s15516709cog1804/_1

[34] Matthias Kraus, Johannes Fuchs, Björn Sommer, Karsten Klein, Ulrich Engelke, Daniel Keim, and Falk Schreiber. 2022. Immersive Analytics with Abstract 3D Visualizations: A Survey. *Computer Graphics Forum* 41, 1 (2022), 201–229. https://doi.org/10.1111/cgf.14430

[35] Ernst Kruijff, J. Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, ISMAR 2010 - Proceedings*. 3–12. https://doi.org/10.1109/ISMAR.2010.5643530

[36] Ricardo Langner, Marc Satkowski, Wolfgang Büschel, and Raimund Dachselt. 2021. Marvis: Combining mobile devices and augmented reality for visual data analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[37] Juan I Larregui, M Luján Ganuza, Ernesto A Bjerg, Silvia Castro, Nicoí As, F Gazcón, Gazc´ Gazcón, Juan M Trippel Nagel, María L Ganuza, and Silvia M Castro. 2018. *Immersive Analytics for Geology: Field Sketch-Like Visualization to assist geological structure analysis during fieldwork*. Technical Report. https://www.researchgate.net/publication/323424944

[38] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer. 2021. Shared Surfaces and Spaces: Collaborative Data Visualisation in a Co-located Immersive Environment. *IEEE TVCG* 27, 2 (2021), 1171–1181. https://doi.org/10.1109/TVCG.2020.3030450

[39] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. 2019. Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *ACM CHI*. 1–14. https://doi.org/10.1145/3290605.3300917

[40] Frieder Loch, Fabian Quint, and Iuliia Brishtel. 2016. Comparing Video and Augmented Reality Assistance in Manual Assembly. In *2016 12th International Conference on Intelligent Environments (IE)*. 147–150. https://doi.org/10.1109/IE.2016.31

[41] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. 2022. Where Should We Put It? Layout and Placement Strategies of Documents in Augmented Reality for Collaborative Sensemaking. In *CHI Conference on Human Factors in Computing Systems*. 1–16. https://doi.org/10.1145/3491102.3501946

[42] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H. Thomas. 2018. *Immersive Analytics*. Springer LNCS 11190.

[43] Joseph McGrath. 1995. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction*. Elsevier, 152–169. https://doi.org/10.1016/B978-0-08-051574-8.50019-4

[44] Leonel Merino, Magdalena Schwarzl, Matthias Kraus, Michael Sedlmair, Dieter Schmalstieg, and Daniel Weiskopf. 2020. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). In *IEEE ISMAR*. 438–451.

[45] Alessandro Mulloni, Jens Grubert, Hartmut Seichter, Tobias Langlotz, Raphael Grasset, Gerhard Reitmayr, and Dieter Schmalstieg. 2012. Experiences with the Impact of Tracking Technology in Mobile Augmented Reality Evaluations. In *Proc. MobileHCI Workshop on Mobile Vision and HCI (MobiVis) 2012*.

[46] Jakob Nielsen. 2000. Why you only need to test with 5 users. *Nielsen Norman Group, Nielsen* (2000).

[47] Donald A. Norman. 1988. *The design of everyday things*. Basic Books.

[48] Arnaud Prouzeau, Antoine Lhuillier, Barrett Ens, Daniel Weiskopf, and Tim Dwyer. 2019. Visual link routing in immersive visualisation. In *ACM International Conference on Interactive Surfaces and Spaces (ISS)*. https://doi.org/10.1145/3343055.3359709

[49] Ihsan Rabbi and Sehat Ullah. 2013. A survey on augmented reality challenges and tracking. *Acta graphica: znanstveni časopis za tiskarstvo i grafičke komunikacije* 24, 1-2 (2013), 29–46.

[50] P. Reipschläger, T. Flemisch, and R. Dachselt. 2021. Personal Augmented Reality for Information Visualization on Large Interactive Displays. *IEEE TVCG* 27, 2 (2021), 1182–1192. https://doi.org/10.1109/TVCG.2020.3030460

[51] Dieter Schmalstieg and Tobias Höllerer. 2016. *Augmented Reality - Principles and Practice*. Addison-Wesley Professional. http://www.augmentedrealitybook.org

[52] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG* 18, 12 (2012), 2431–2440. https://doi.org/10.1109/TVCG.2012.213

[53] Richard Skarbez, Nicholas F. Polys, J. Todd Ogle, Chris North, and Doug A. Bowman. 2019. Immersive Analytics: Theory and Research Agenda. *Frontiers in Robotics and AI* 6 (9 2019). https://doi.org/10.3389/frobt.2019.00082

[54] Hariharan Subramonyam, Steven M. Drucker, and Eytan Adar. 2019. Affinity lens data-assisted affinity diagramming with augmented reality. In *ACM CHI*. 1–13. https://doi.org/10.1145/3290605.3300628

[55] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/642611.642626

[56] Bruce H. Thomas, Gregory F. Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, Neven A. M. ElSayed, Ross T. Smith, and Wesley Willett. 2018. *Situated Analytics*. Springer International Publishing, 185–220. https://doi.org/10.1007/978-3-030-01388-2_7

[57] Eduardo Veas, Raphael Grasset, Ioan Ferencik, Thomas Gruenewald, and Dieter Schmalstieg. 2013. Mobile Augmented Reality for Environmental Monitoring. *Personal and Ubiquitous Computing* 17, 7 (2013). https://doi.org/10.1007/s00779-012-0597-z

[58] Eduardo Veas, Raphael Grasset, Ernst Kruijff, and Dieter Schmalstieg. 2012. Extended Overview Techniques for Outdoor Augmented Reality. *IEEE TVCG* 18 (April 2012), 565–572. http://arbook.icg.tugraz.at/schmalstieg/Schmalstieg_226.pdf

[59] James A. Walsh and Bruce H. Thomas. 2011. Visualising Environmental Corrosion in Outdoor Augmented Reality. In *Australasian User Interface Conference*. 39–46.

[60] X. Wang, S. K. Ong, and A. Y.C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4 (3 2016), 1–22. Issue 1. https://doi.org/10.1007/S40436-015-0131-4/FIGURES/9

[61] Stefan Werrlich, Austino Daniel, Alexandra Ginger, Phuc-Anh Nguyen, and Gunther Notni. 2018. Comparing HMD-based and paper-based training. In *IEEE ISMAR*. 134–142. https://doi.org/10.1109/ISMAR.2018.00046

[62] Sean White and Steven Feiner. 2009. SiteLens: situated visualization techniques for urban site visits. In *Proc. of the international conference on human factors in computing systems (CHI)*. 1117–1120. https://doi.org/10.1145/1518701.1518871

[63] Sean White and Steven Feiner. 2009. SiteLens: Situated Visualization Techniques for Urban Site Visits. In *ACM CHI*.

[64] Wesley Willett, Bon Adriel Aseniero, Sheelagh Carpendale, Pierre Dragicevic, Yvonne Jansen, Lora Oehlberg, and Petra Isenberg. 2022. Perception! Immersion! Empowerment! Superpowers as Inspiration for Visualization. *IEEE TVCG* 28, 1 (2022), 22–32. https://doi.org/10.1109/TVCG.2021.3114844

[65] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. 2017. Embedded Data Representations. *IEEE TVCG* 23, 1 (2017), 461–470. https://doi.org/10.1109/TVCG.2016.2598608

[66] S. Zollmann, D. Kalkofen, C. Hoppe, S. Kluckner, H. Bischof, and G. Reitmayr. 2012. Interactive 4D overview and detail visualization in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 167–176.