ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory

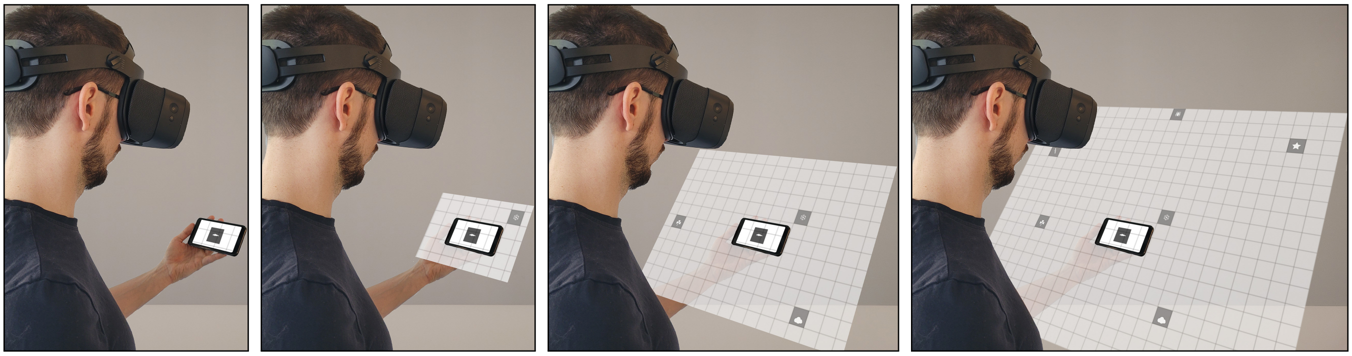
Sebastian Hubenschmid
Sebastian.Hubenschmid@uni.kn
University of Konstanz

Johannes Zagermann
Johannes.Zagermann@uni.kn
University of Konstanz

Daniel Leicht
Daniel.Leicht@uni.kn
University of Konstanz

Harald Reiterer
Harald.Reiterer@uni.kn
University of Konstanz

Tiare Feuchtner
Tiare.Feuchtner@uni.kn
University of Konstanz
Aarhus University



(a) NO-AR (b) SMALL-AR (c) MEDIUM-AR (d) LARGE-AR

Figure 1: Exploring virtually-extended displays using a video see-through augmented reality head-worn display, we compare four different display sizes: (a) the NO-AR condition is used as baseline; (b) the SMALL-AR condition is comparable to a tablet; (c) MEDIUM-AR is similar to a desktop monitor; and (d) the LARGE-AR condition is equivalent to a television monitor.

ABSTRACT

Smartphones conveniently place large information spaces in the palms of our hands. While research has shown that larger screens positively affect spatial memory, workload, and user experience, smartphones remain fairly compact for the sake of device ergonomics and portability. Thus, we investigate the use of hybrid user interfaces to virtually increase the available display size by complementing the smartphone with an augmented reality head-worn display. We thereby combine the benefits of familiar touch interaction with the near-infinite visual display space afforded by augmented reality. To better understand the potential of virtually-extended displays and the possible issues of splitting the user's visual attention between two screens (real and virtual), we conducted a within-subjects experiment with 24 participants completing navigation tasks using different virtually-augmented display sizes. Our findings reveal that a desktop monitor size represents a "sweet spot" for extending smartphones with augmented reality, informing the design of hybrid user interfaces.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; *User studies.*

KEYWORDS

spatial memory; augmented reality; hybrid user interfaces

ACM Reference Format:
Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. 2023. ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581438

1 INTRODUCTION

Through smartphones, vast information landscapes are conveniently available at our fingertips. Their relatively small form factor makes them easily portable, while their ergonomics are (mostly [4, 44]) well-suited for one-handed interaction. Yet, this design for portability and ergonomics comes at the cost of a limited display size: prior work has demonstrated that a bigger screen can improve

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581438

spatial memory, workload, and user experience [60], especially when using touch interaction [78].

In contrast, augmented reality (AR) head-worn displays (HWDs) offer a near-infinite visual space. While AR hardware can be expected to further improve and become commonplace in everyday life (e.g., as sunglasses), the ergonomics [3, 6, 32] and accuracy [7, 14] of mid-air interaction may present a critical challenge for wide adoption. One promising possibility to address this issue are hybrid user interfaces [21, 76]. For example, by combining AR HWDs with smartphones in a hybrid system, we complement the familiar touch interaction with the near-infinite visual display space of an AR HWD. We can thereby virtually increase the smartphone's display size to benefit from better spatial memory and usability. In particular, virtually-extended screen-aligned displays (VESADs) [54] have been found effective for seamlessly extending physical smartphone screens, presenting them as one (possibly infinitely) large screen. But how much (virtual) display real-estate is actually necessary or even beneficial?

To better understand the design space of hybrid user interfaces, we investigate the impact of the size of virtually-extended screens through the complementary use of an AR HWD and a smartphone. Although we expect to find many similarities to prior studies on spatial memory (e.g., [60, 78]), hybrid user interfaces also face unique challenges that must be considered. For example: (1) By separating the display into a real and virtual display, users may have to split their visual attention [61], counteracting potential benefits gained from virtually increasing the display size [30]; (2) larger screen sizes may require more head-movement, which may affect ergonomics [12]; and (3) larger screen sizes cannot be fully kept in view or visually processed effectively, which could affect cognitive load. All of these may prove detrimental to spatial memory. To investigate this in detail, we conducted a controlled laboratory experiment with 24 participants testing the effect of different sizes of virtually-extended displays on spatial memory, workload (e.g., ergonomics, cognitive load), and user experience. Our findings support that larger virtually-extended screen can consistently improve spatial memory, workload, and user experience, but also indicate that these benefits decrease again once the virtually-extended screen becomes too big. Our findings also reveal that using a small virtually-extended screen can indeed perform worse than providing no virtually-extended screen at all.

As part of our experiment, we developed a hybrid user interface that extends smartphone display space through AR and contribute: (1) *key findings* from an evaluation thereof with 24 participants, comparing four sizes of virtually-extended displays; (2) we thereby investigate *spatial memory*, *workload*, and *user experience*, based on which we (3) present *design and research implications* for virtually-extended displays.

In the following sections, we review findings from *related work*, explain our *experiment* and the *results* thereof, provide a *discussion* of the results in relation to our research questions, discuss *limitations and future work*, and present *insights for design and research.*

2 RELATED WORK

In this section, we review prior findings in the context of *spatial memory* and *hybrid user interfaces.*

2.1 Spatial Memory

Spatial memory is an important aspect of human cognition that has been well-studied in relation to HCI [66], especially in the context of reducing cognitive effort for navigation and search tasks [1, 58]. In this context, prior work has investigated different input (e.g., peephole navigation [40, 50, 56], body movement [22, 41, 59], mouse and touch [39, 73]) and output modalities (e.g., audio cues [24], tab interfaces [27]), display sizes [60, 78], visualization and memorization techniques (e.g., fisheye lenses [69], providing an overview [34, 38], focus+context [11, 55], storytelling [23]), as well as the use of landmarks (e.g., gridlines [46], body parts [4], anchors and background images [74], in graphical user interfaces [65] and 3D environments [26, 52]).

Specifically, prior work indicates that – compared to indirect mouse input – direct touch interaction can improve memorization accuracy [73], spatial memory, and navigation performance [39]. Zagermann et al. [78] showed that embodied interaction can increase spatial memory when compared to indirect touch (e.g., via trackpad) and direct touch interaction, but at the cost of user experience and efficiency. This is especially relevant in the context of peephole navigation [49], as smartphones can be used for both static (i.e., touch) and dynamic (i.e., spatial movement, e.g., [56]) peephole navigation. In terms of peephole size, a study by Rädle et al. [60] show that an increased peephole size can positively affect learning speed, navigation speed, and task load – albeit with diminishing returns. For AR, we also need to consider the virtual field of view. In this regard, a study by Caluya et al. [12] shows that a smaller virtual field of view has no significant impact on spatial memory, but can increase head movement.

In summary, prior work highlights the relation between peephole size and spatial memory. We expect that bigger peepholes (i.e., larger screens) perform better in terms of spatial memory, workload, and user experience (cf. [60, 78]), but may also negatively impact workload (i.e., head movement) as screen sizes exceed the HWD's field of view (cf. [12]). Here, a 2D environment in AR with direct touch interaction (e.g., via smartphone), gridlines, and visual anchors can strike a good balance in terms of clutter [15], efficiency [15, 78], and user experience [78].

2.2 Hybrid User Interfaces

Hybrid user interfaces [21] combine complementary devices [76] such as AR HWDs and smartphones to offset the disadvantages of each device. Recent research has demonstrated their applicability and relevance in a variety of use cases, combining mixed reality HWDs with a wide range of interactive devices, such as smartwatches [30], smartphones [42, 45, 54, 75, 79], tablets [2, 17, 18, 36, 43, 48, 68, 72], interactive surfaces [5, 10, 63, 70], display walls [64, 71], and desktop computers [35, 37]. Initial studies show that hybrid user interfaces can improve navigation performance [9] and user experience [54, 79]. One common use case for hybrid user interfaces, especially across smaller handheld devices, is the extension of screen real-estate. While prior work has explored different technologies for screen extensions (e.g., projectors [13, 29, 31], using multiple devices for cross-device interaction [8, 77]), AR HWDs allow for a truly seamless screen extension. Normand and McGuffin [54] describe such virtually-extended screen-aligned displays

(VESADs) as virtual screens that are *"centered on, and co-planar with, a smartphone"*. Prior work has explored the design space of such VESADs for annotations [17], immersive analytics [43], and general interaction concepts [79]. Grubert et al. [30] show how extending the display of a smartwatch can improve task completion times at the cost of higher workload.

Although an increased screen real-estate can be beneficial (e.g., for spatial memory, see Section 2.1), prior work in multi-display environments also hints at several potential issues that may counter the benefits. One issue is the split attention effect [28], as users have to split their visual attention between multiple displays, resulting in overall worse performance [61]. In this regard, Rashid et al. [62] provide an overview of different factors influencing attention switches in multi-display environments, such as display contiguity and angular coverage. In addition, a study by Nacenta et al. [53] shows that a physical gap between displays can significantly reduce performance. Yet, unlike prior methods of expanding screens, VESADs leverage AR to seamlessly extend the smartphone screen, thus eliminating any "displayless space" [53] and potentially avoiding the attention split between two displays. Still, Grubert et al. [30] and Eiberger et al. [20] observed significant overhead when switching between display output of an AR HWD and smartphone due to different focal planes, while Normand and McGuffin [54] observed no such overhead with video see-through HWDs.

In summary, hybrid user interfaces are increasingly used to, for example, seamlessly extend the screen of handheld devices with AR. Although the combination of handheld device with AR HWD presents many benefits, prior work in multi-monitor environments also indicates unique challenges, such as splitting the user's visual attention between two screens, which can have a negative impact on performance. To the best of our knowledge, there is no prior research investigating these challenges for VESADs and their size.

3 EXPERIMENT

Building on prior research, we aim to investigate whether reported findings on spatial memory can be transferred to the use case of hybrid user interfaces. On the one hand, larger display screens have been shown to be beneficial for spatial memory [60], especially when using touch interaction [78]. On the other hand, the split attention effect may negatively impact performance [13, 62]. We conducted a controlled laboratory experiment using different virtual screen sizes to investigate the impact of virtually-extended display sizes on spatial memory, workload, and user experience. In addition, we also investigated possible differences in the split attention effect between different virtual screen sizes.

3.1 Research Questions

Our experimental setup is guided by three main research questions.

RQ1: **Spatial Memory.** How does screen size affect users' navigation behavior and performance in a *spatial memory* task?

RQ2: **Workload.** Does a virtual screen extension cause an increased *workload* in terms of cognitive load and ergonomics due to increased context switching between the real and virtual display?

RQ3: **User Experience.** In what way is *user experience* influenced by screen size?

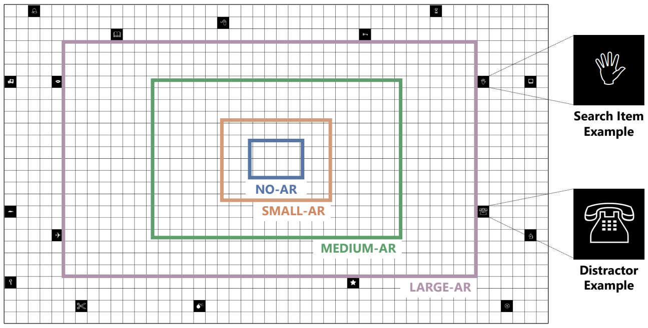


Figure 2: Exemplary 2D grid map used during the *navigation phase* containing search icons and distractor icons. A space in the middle of the map with the size of LARGE-AR was intentionally left blank to hide any icons in the starting position of each condition. Colored borders indicate the size and starting positions of each condition.

The effect on *spatial memory* (RQ1) is evaluated by analyzing the navigation path, task completion time, navigation speed, and accuracy of object location recall. Further, the *workload* (RQ2) is operationalized via the pupil size as indicator for cognitive load [19], total amount of gaze movement, total degree of head movement during navigation, and the NASA task load index [33]. Finally, we evaluate *user experience* (RQ3) with a user experience questionnaire [67] and subjective preference ratings. Overall, we expect that display size positively affects spatial memory, workload, and user experience (cf. [60, 78]). We therefore also expect to see diminishing returns, as participants no longer profit from the increased screen size beyond a certain threshold (e.g., tablet-sized [60]).

3.2 Conditions

We differentiate between three extension sizes (similar to Rädle et al. [60]) to mimic existing devices. In addition, we added a condition without any virtually-extended display as baseline condition. Lastly, participants wore an AR HWD in all conditions to guarantee for a better comparability of conditions. In order of smallest to largest display size, we compared the following display sizes (see Figures 1 & 2):

- **NO-AR:** This condition serves as baseline with no virtual extension, using a display size of 5.5″.
- **SMALL-AR:** This condition mimics the size of a tablet, using a total display size of 11″.
- **MEDIUM-AR:** This condition is similar to current desktop monitors with a display size of 23″.
- **LARGE-AR:** The largest condition has approximately the size of a television monitor at 43″.

3.3 Tasks

To keep our results comparable to prior studies on spatial memory, we employed an established task (see [39, 41, 46, 50–52, 59, 78]) which consists of a *navigation phase* and an *object location recall phase.* The task makes use of a 2D grid map (46 columns × 27 rows, see Figure 2) with an approximate real world size of 124 cm × 73 cm.

While the visible area differed for each condition, the map size stayed consistent. All conditions and maps had a visible aspect ratio of 16:9.

Navigation phase. For the *navigation phase*, participants had to put on an AR HWD and were provided with a smartphone that was extended with a VESAD (depending on the condition, see Figure 1). Participants had to search for and navigate to a symbol on the grid map using touch panning gestures on the smartphone to move the map (i.e., using static peephole navigation [49]). To increase ecological validity, the navigation behaved similar to off-the-shelf map applications by emulating physical inertia and stopping when the participant touched the phone again. For each search trial, the application started in its default position in the middle of the grid map (see Figure 2). The current search icon was shown as a semi-transparent symbol, which remained statically in the middle of the smartphone screen. The task was automatically completed once participants navigated to the icon and placed it approximately in the middle of the smartphone (i.e., once the map symbol's center touched the semi-transparent search symbol). In addition, we included a total of five item sets: an item set for the training task, showing letters of the alphabet and four distinct item sets (see Figure 2) to avoid learning effects between conditions. The locations of all icons were randomized on each map (i.e., in every condition) to prevent learning effects across conditions. Yet, we ensured that the length of navigation paths remained comparable across conditions and that there were no differences with regard to the complexity or theme of the icons (cf. [78]).

Participants were tasked to find a series of 6 different icons (with 4 repetitions each), which were distributed on the 2D grid map that also included similarly-looking symbols that served as distractors (cf. [78], see Figure 2). As shown in Figure 2, no icons were visible in the starting positions of all conditions and icons were not placed within an area equivalent to the size of LARGE-AR. With this, we ensured that, for example, participants do not visually scan the map in larger conditions before actually interacting with it, which might impair the comparability of conditions (e.g., navigation path lengths and task completion times). Each condition therefore consisted of 24 search trials, resulting in 96 trials per participant and a total of 2304 trials over all participants.

Object location recall phase. For the *object location recall phase*, participants sat in front of a desktop PC with a mouse, reducing potential influences of motoric or kinesthetic memory to increase the internal validity of the spatial memory measurement, as a common practice for studying spatial memory (cf. [39, 78]). Here, participants were first presented with an empty grid map. In this phase, the entire map was visible on the screen and no navigation was possible. Instead, participants had to place the icons from the *navigation phase* in their prior location by clicking on the corresponding position on the map using the mouse. The current icon was shown at the top of the screen. The icon order was based on the search order from the *navigation phase.*

3.4 Measurements

We employed different quantitative and qualitative metrics to address our research questions.

Spatial Memory. To measure the impact on spatial memory, we measure the path length, task completion time, and navigation speed during the *navigation phase* as well as the icon placement recall accuracy during the *object location recall phase.* For better comparability across conditions, we use the normalized path length, which is calculated as a ratio between the participant's actual path and the shortest possible path using Euclidean distance. Here, we omitted the first repetition from our analysis due to the initial randomness during the first navigation [39]. The task completion time was logged as duration in seconds between the start of a repetition until the icon was found and placed in the middle of the smartphone. Since navigation speed can be derived from the path length and task completion time, we complemented this measurement by recording the maximum navigation speed to investigate whether larger display screen sizes allow participants to flick through the map more quickly. Lastly, the icon placement accuracy was measured in the Euclidean distance in pixels between the icon's actual position and the position where participants placed the icon during the *object location recall phase.*

Workload. To evaluate the objective workload, we measure the pupil size [19], the total amount of gaze movement, total degree of head movement during navigation, and the subjective task load. The pupil size was measured using a built-in eye-tracker which logged a relative value between 0–1 according to the pupil size range detected by the AR HWD. Here, a larger pupil size is seen as an indicator for increased task load [19]. The total amount of gaze movement and total degree of head movement were both calculated in the quaternion distance (i.e., angles) between each data point and divided by the duration until the next data point. We also used the raw NASA TLX [33] which allows us to measure subjectively perceived task load. Lastly, we traced the participant's gaze onto the virtually-extended screen to measure how much time participants spent looking at which screen (i.e., smartphone or virtual screen).

User Experience. We used the user experience questionnaire (UEQ) [67] to gain more insights into the attractiveness, hedonic qualities, and pragmatic qualities of each condition. We complemented these results with a semi-structured interview at the end of the study session to gather qualitative insights into participant's preferences.

3.5 Apparatus

For all navigation tasks, we employed a Varjo XR3 as video see-through AR HWD due to its high digital field of view (155° horizontal field of view; 90 Hz refresh rate; 12 megapixel video pass-through per eye, 100 Hz eye-tracker) attached to a state-of-the-art computer (Intel i9 9900K, Nvidia RTX 3090). We intentionally decided against an optical see-through HWD to avoid potential issues with different focal planes [20, 30, 54]. The information landscape was overlayed on top of a Google Pixel XL (5.5″, 2560 × 1440 pixel, Android 10). The smartphone was cut out from the digital overlay, allowing participants to still fully see their hands and the smartphone's display and its content. We also reduced the transparency of the entire map to 40 % so that participants were still able to make out their physical surroundings. The AR HWD was tracked using four Valve Base Stations placed in every corner of the room, while the smartphone was tracked with a fiducial marker mounted to the smartphone and tracked via the HWD's front-facing cameras. We used a stabilization algorithm to avoid inaccurate smartphone tracking when participants were not looking at the smartphone (e.g., for LARGE-AR). The object location recall task was performed using a desktop PC on a 4K 27″ monitor.

The software for all devices was implemented in Unity 2021.1 and is available as open source project[1]. The applications communicated through a client-server architecture using TCP. We connected all devices via 5 GHz Wifi or ethernet to reduce latency – ensuring that there was no perceivable latency between smartphone display and AR overlay.

3.6 Participants

We recruited 24 participants (10 female, 14 male) aged 21–36 (M = 24.4, SD = 3.1) from the local university. Participants were recruited through flyers that advertised an AR study about memory games. We recruited participants that were fluent in the local language to avoid potential differences in the linguistic meaning of different icons (cf. [39, 46, 78]). 22 participants were undergraduate students from different fields (e.g., computer science, social studies, history, biology, life science, law), 1 participant was a PhD student, and 1 participant was administrative staff. Although participants were mostly experienced in the use of smartphones (M = 4.375, SD = .824, on a Likert scale from 1 (inexperienced) to 5 (experienced)) and all participants owned a smartphone (n = 24), experience with AR applications was mixed (M = 2.75, SD = 1.327, on a Likert scale from 1 (inexperienced) to 5 (experienced)). All participants had normal (n = 12) or corrected to normal (n = 12) vision.

3.7 Procedure

Participants first signed a consent form, completed a demographic questionnaire, and received an introductory presentation about the task and guidelines for wearing the HWD to ensure correct eye-tracking calibration. We assigned each participant a different order of conditions using full counterbalancing to avoid any learning effects. In each condition, participants started by putting on the AR HWD and received a smartphone. During all *navigation phases*, participants remained seated at a table and held the smartphone in landscape orientation. Participants started in the *navigation phase* where they first solved a training task until they felt comfortable with the system. Next, participants solved 4 repetitions of finding 6 different icons. After the *navigation phase*, participants took off the AR HWD to use a desktop system, where they completed the *object location recall phase* using mouse input. At the end of each condition, participants filled out a raw NASA TLX [33] and a UEQ [67]. We concluded each session with a semi-structured interview about topics such as memorization strategies, subjective preferences, and preferred display sizes. The study duration ranged between 40–70 minutes, and all participants received monetary compensation for their time. We also awarded an additional monetary reward to the fastest participant to further encourage participants to perform the tasks as quickly as possible. We followed all necessary ethical and sanitary guidelines provided by the local university.

4 RESULTS

In this section, we present our results based on our three research questions of *spatial memory*, *workload*, and *user experience*. Since a Shapiro-Wilk test revealed that our data did not follow a normal distribution, we analyzed the data with a non-parametric approach. We used a Friedman test followed by a pairwise Wilcoxon test with Bonferroni correction as post-hoc analysis to test for statistical significance, where appropriate. We indicate the medians (Mdn) and standard deviations (SD) using subscripts $_{NO}$ for NO-AR, $_{S}$ for SMALL-AR, $_{M}$ for MEDIUM-AR, and $_{L}$ for LARGE-AR to improve readability. We assume $\alpha = .05$ for statistical significance. For pairwise comparisons, we adjusted significance values by the Bonferroni correction for multiple tests. The user study data is available in a data repository[2]. To improve readability, we report results of statistical analyses as tables in the appendix of this paper.

4.1 Spatial Memory

We measure spatial memory based on the *navigation path length*, *task completion time*, and *recall accuracy*. As a related measure, we also investigated the *navigation speed* during the *navigation phase*. Our findings are summarized in Figure 3.

4.1.1 Navigation Path Length. We found statistically significant differences in each repetition when comparing normalized navigation path length across conditions. Pairwise post-hoc comparisons showed that LARGE-AR (Mdn_L = 2.629, SD_L = .365) and MEDIUM-AR (Mdn_M = 2.532, SD_M = .339) had consistently shorter navigation path lengths than NO-AR (Mdn_{NO} = 4.168, SD_{NO} = 1.317) and SMALL-AR (Mdn_S = 3.92, SD_S = 1.274) throughout all repetitions (see Appendix A: Tables 1 & 2 and Figure 3 (A)); pairwise comparisons between NO-AR and SMALL-AR or MEDIUM-AR and LARGE-AR did not show any statistically significant differences. Comparing normalized navigation path lengths between repetitions within each condition showed a decrease in path lengths: Here, the overall test showed statistically significant differences for all conditions with individual differences for each condition, when comparing the first with following repetitions (see Tables 3 & 4). Only MEDIUM-AR showed a significant difference between Repetition 1 and 2.

4.1.2 Task Completion Time. We found statistically significant differences ($\chi^2(3) = 48.05, p < .001$) when comparing task completion times across conditions (see Figure 3 (C)). A pairwise post-hoc comparison reveals that NO-AR ($Mdn_{NO} = 72.883$, $SD_{NO} = 41.538$) is significantly longer than both MEDIUM-AR ($Mdn_M = 44.01$, $SD_M = 16.569$, $z = 1.875$, $p < .001$) and LARGE-AR ($Mdn_L = 40.782$, $SD_L = 13.11$, $z = 2.333$, $p < .001$). Similarly, SMALL-AR ($Mdn_S = 56.884$, $SD_S = 23.817$) is also significantly longer than MEDIUM-AR ($z = 1.542$, $p < .001$) and LARGE-AR ($z = 1.083$, $p = .022$). No statistically significant differences were found between NO-AR and SMALL-AR, or MEDIUM-AR and LARGE-AR.

4.1.3 Recall Accuracy. We found statistically significant differences in the object recall accuracy during the *object location recall phase* ($\chi^2(3) = 14.121$, $p = .003$, see Figure 3 (B)). A pairwise post-hoc comparison shows that NO-AR ($Mdn_{NO} = 3.248$, $SD_{NO} = 1.026$,

---

[1] https://github.com/hcigroupkonstanz/ARound-the-Smartphone

[2] DOI: 10.18419/darus-3326

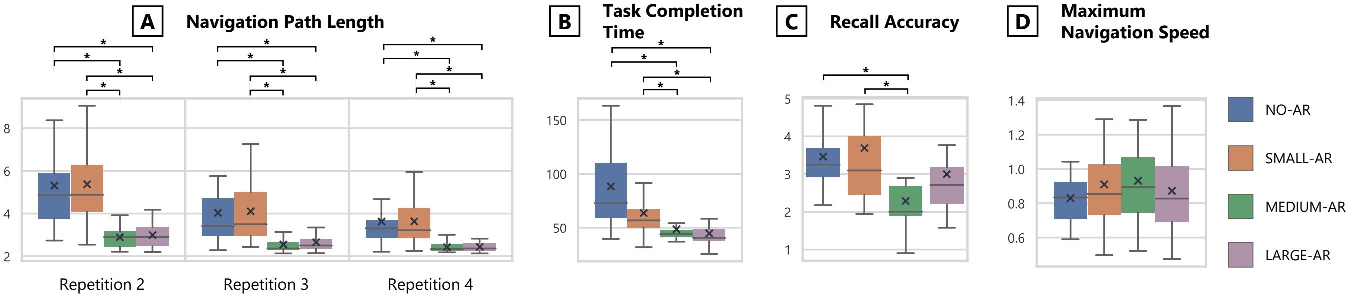


Figure 3: Summary of measures for *spatial memory.* (A) Boxplots of pairwise comparison of normalized navigation path length during the *navigation phase* (optimal normalized path length is 1). We omitted Repetition 1 due to the randomness of initial navigation trials [39]. (B) Boxplots showing average task completion time in seconds during the *navigation phase.* (C) Boxplots showing participants' accuracy of the *object location recall phase* in average Euclidean distance in blocks between the actual and the recalled position of the sign. (D) Maximum navigation speed in cm/s.

$z = 1.12$, $p = .015$) and SMALL-AR ($Mdn_S = 3.095$, $SD_S = 1.804$, $z = 1.188$, $p = .009$) are significantly less accurate than MEDIUM-AR ($Mdn_M = 2.003$, $SD_M = .873$). While MEDIUM-AR performed best on average, no significant differences could be found between LARGE-AR ($Mdn_L = 2.715$, $SD_L = 1.199$) and MEDIUM-AR.

4.1.4 Navigation Speed. For navigation time, we divided *normalized navigation path length* by the *task completion time*. Although both measures already showed statistically significant differences between conditions, we still include the results here to substantiate our findings. Since we found statistically significant differences in navigation speed ($\chi^2(3) = 49.35$, $p < .001$), we performed a pairwise post-hoc comparison. The comparison shows that NO-AR ($Mdn_{NO} = .196$, $SD_{NO} = .039$) had a significantly slower navigation speed than MEDIUM-AR ($Mdn_M = .263$, $SD_M = .06$, $z = -2.417$, $p < .001$) and LARGE-AR ($Mdn_L = .24$, $SD_L = .051$, $z = -1.625$, $p < .001$). Likewise, SMALL-AR ($Mdn_S = .212$, $SD_S = .038$) also was significantly slower than MEDIUM-AR ($z = -1.792$, $p < .001$) and LARGE-AR ($z = -1$, $p = .044$). To further complement these findings, we also measured the maximum navigation speed during each condition (see Figure 3 (D)). Although MEDIUM-AR performed best on average, we found no statistically significant differences ($\chi^2(3) = 3.052$, $p = .384$).

4.2 Workload

To better understand the effects of display sizes on workload-related measures such as ergonomics and cognitive load, we measured the amount of *head movement*, the participant's *pupil size*, the amount of *eye-gaze movement* during all conditions, the participant's *visual attention*, and the *subjective task load* (see Figure 4 and Figure 5 (A)). For eye-tracking measures, we omitted data from 7 participants due to insufficient tracking quality. For movement-based data, we omitted data from 3 participants due to technical issues with our prototype.

4.2.1 Head Movement. We measured the amount of head movement using the rotational data of the AR HWD (see Figure 4 (B)). Here, we found statistically significant differences in the amount of head rotation between conditions ($\chi^2(3) = 35.682, p < .001$). A pairwise post-hoc comparison reveal that LARGE-AR ($Mdn_L = 4.932$, $SD_L = 5.47$) is significantly higher than both NO-AR ($Mdn_{NO} = .176$, $SD_{NO} = 1.889$, $z = -2.353$, $p < .001$) and SMALL-AR ($Mdn_S = .278$, $SD_S = 2.388$, $z = -2.118$, $p < .001$), while MEDIUM-AR ($Mdn_M = 1.812$, $SD_M = 3.85$) is significantly higher than NO-AR ($z = -1.294$, $p = .021$).

4.2.2 Pupil Size. We also analyzed participants' pupil size as an objective indicator for mental demand [19]. We compared the relative pupil size between conditions which revealed statistically significant differences ($\chi^2(3) = 15.568$, $p = .001$). A pairwise post-hoc comparison shows participants had a significantly larger pupil size in NO-AR ($Mdn_{NO} = .072$, $SD_{NO} = .012$) than MEDIUM-AR ($Mdn_M = .069$, $SD_M = .009$, $z = 1.211$, $p = .023$) and LARGE-AR ($Mdn_L = .069$, $SD_{NO} = .01$, $z = 1.579$, $p = .001$). No statistically significant differences were found for SMALL-AR compared to all other conditions ($Mdn_S = .07$, $SD_S = .01$).

4.2.3 Eye-Gaze Movement. We found statistically significant differences in the amount of eye-gaze movement between conditions ($\chi^2(3) = 31.518$, $p < .001$, see Figure 4 (C)). A pairwise post-hoc analysis shows that SMALL-AR ($Mdn_S = .398$, $SD_S = .596$, $z = -1.353$, $p = .013$), MEDIUM-AR ($Mdn_M = .391$, $SD_M = .551$, $z = -1.471$, $p = .005$), and LARGE-AR ($Mdn_L = .686$, $SD_L = .553$, $z = -2.471$, $p < .001$) have a larger amount of eye-gaze movement when compared against NO-AR ($Mdn_{NO} = .041$, $SD_{NO} = .0595$). No significant differences between the AR conditions were found.

4.2.4 Visual Attention. To better analyze the visual attention during conditions, we investigated the amount of time participants focused on the AR extension (see Figure 4 (D)) and visualize gaze behavior in a heatmap for each screen size (see Figure 4 (E)). We thereby omit the NO-AR condition as participants were fully focused on the smartphone. For duration of visual focus, we found significant differences in how much time participants spent looking at the smartphone screen or the AR extension depending on the screen size ($\chi^2(2) = 20.235$, $p < .001$). A pairwise post-hoc analysis shows that participants looked significantly less at the AR extension during SMALL-AR ($Mdn_S = .33$, $SD_S = .185$) when compared to MEDIUM-AR ($Mdn_M = .453$, $SD_M = .205$, $z = -.941$, $p = .018$) and LARGE-AR ($Mdn_L = .697$, $SD_L = .196$, $z = -1.529$, $p < .001$).

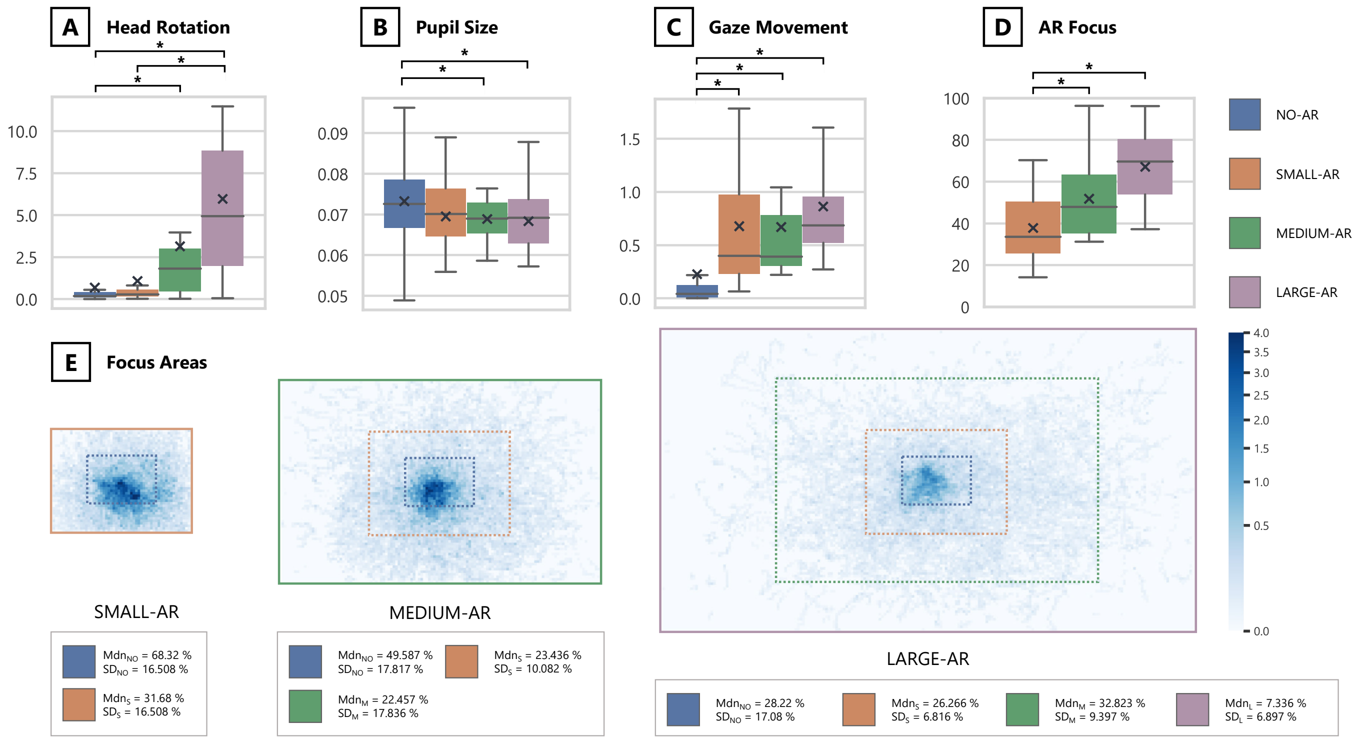


Figure 4: Summary of objective measures for *workload*. (A) Boxplots showing the average head rotation in °/s. (B) Boxplots showing the average pupil size during all conditions. Here, pupil sizes ranges are a relative value between 0 and 1 as calculated by the AR HWD. (C) Boxplots showing the average gaze movement in °/10ms. (D) Boxplots showing the average time in percent that participants spent looking at the AR screen extension. (E) Heatmaps showing how long participants focused on what part of the VESAD. Heatmaps indicate fixation duration per position related to total gaze time, whereby results were scaled with a power-law function ($\gamma = 0.5$) to increase the visibility of lower values. We marked corresponding virtual extensions sizes in each heatmap to better contextualize the values. We also added statistics below each heatmap indicating how much percent of a participant’s focus each display size received (excluding inner conditions).

This gaze behavior also becomes apparent from the heatmap visualization (see Figure 4 (E)), where clusters of darker color indicate that especially for SMALL-AR participants looked mostly at lower central portion of the smartphone screen, with occasional glances at the virtual extension around the smartphone. Even in MEDIUM-AR and LARGE-AR, where participants overall spent more time looking at the virtual extension, the participants’ gaze was mostly centered on or just around the smartphone. For LARGE-AR, participants’ gaze appears to traveled across the extended display up to the MEDIUM-AR size (percent of total gaze time in MEDIUM-AR area, excluding inner sizes: $Mdn = 32.828\%$, $SD = 9.397\%$), but quickly falls off towards the edge (percent of total gaze time in LARGE-AR area, excluding inner sizes: $Mdn = 7.336\%$, $SD = 6.897\%$). Similarly, gazes in MEDIUM-AR also fall off towards the edges, but the overall virtual display area of MEDIUM-AR appears more evenly used.

4.2.5 Subjective Task Load. We used the NASA TLX [33] to measure task load after each *navigation phase* (see Figure 5 (A) and Appendix A: Tables 5 & 6). We found statistical significant differences in all subscales: *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration* as well as the *overall scores*. In a pairwise post-hoc analysis, we found a statistically significant improvement of SMALL-AR compared to NO-AR in *effort*. A comparison of NO-AR and MEDIUM-AR shows statistically significant improvements for MEDIUM-AR across all subscales. When comparing NO-AR to LARGE-AR, we found significant improvements for LARGE-AR in *mental demand*, *temporal demand*, and *frustration*. We also found statistically significant improvements from SMALL-AR to MEDIUM-AR in *mental demand* and *frustration*. Repeating this pairwise comparison for the overall score, results show that MEDIUM-AR and LARGE-AR have a significantly lower workload than NO-AR, while MEDIUM-AR also has a significantly lower workload than SMALL-AR. On average, MEDIUM-AR performed better than LARGE-AR in all subscales, but no statistically significant differences were found.

4.3 User Experience

We measure user experience based on the *user experience questionnaire* and *subjective preferences* gained from a semi-structured interview. Figure 4 (B) shows an overview of our statistical findings.

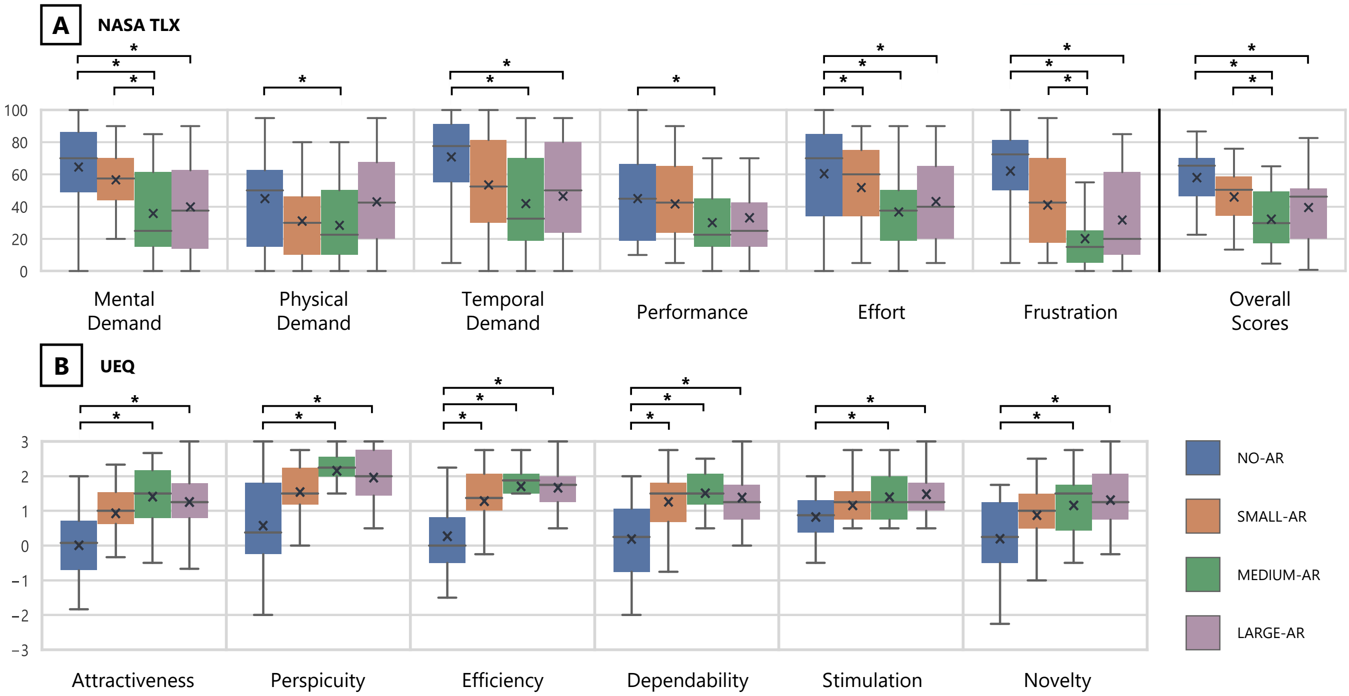


Figure 5: Measures for *workload* (A) and *user experience* (B). (A) Boxplots showing the results of the NASA TLX questionnaire for all conditions. (B) Boxplots of the results from the UEQ for all conditions.

4.3.1 User Experience Questionnaire.

We employed a UEQ [67] after each *navigation phase* (see Figure 5 (B) and Appendix A: Tables 7 & 8). We discovered statistically significant differences in all scales: *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, and *novelty*. A pairwise post-hoc analysis reveals that SMALL-AR was ranked better than NO-AR in *efficiency* and *dependability*. MEDIUM-AR was ranked better than NO-AR in all scales; similarly, LARGE-AR was ranked better than NO-AR in all scales. Although MEDIUM-AR ranked, on average, best in all scales except for *stimulation* and *novelty*, no statistically significant differences compared to LARGE-AR could be found.

4.3.2 Subjective Preferences.

During a final semi-structured interview, we asked participants about their most and least favorite condition (multiple choices were allowed) and the reasoning behind this choice.

Regarding the most favored condition, participants were split between the MEDIUM-AR (n = 14) and the LARGE-AR extension (n = 11). Furthermore, 4 participants chose the SMALL-AR extension as their favorite. While some participants (n = 8) stated they liked the LARGE-AR condition for the extensive display size that provided a better overview, other participants found the LARGE-AR size too big (n = 5), too overloaded (n = 1), and disliked the head movement associated with the LARGE-AR condition (n = 3). For LARGE-AR, participants noted that *"I liked the television size [. . .] because when I scan the whole map to see where each symbol is, then the biggest size helps the most"*—[P19] and that *"you could really take advantage of the headset by really looking around to see where the different items are"*—[P13]. In contrast, other participants argued that *"the TV monitor was almost too big, it was hard to keep everything in sight"*—[P11]. Participants (n = 2) also felt that they were most familiar with the MEDIUM-AR condition, as it resembled a typical desktop monitor in size. Lastly, one participant indicated some possible motion sickness issues due to increased amount of eye gaze movement in the LARGE-AR condition and therefore rated the SMALL-AR condition higher in comfort: "[. . .] *because I had to move my eyes less.* [. . .] *Once I have to move my eyes too much, I get motion sick."*—[P2].

Regarding the least favored condition, participants almost unanimously (n = 23) chose the NO-AR condition. One participant only disliked the SMALL-AR condition as it provided no real advantage to NO-AR. Another participant disliked both the NO-AR condition and the LARGE-AR condition because they felt that the LARGE condition made it harder to remember the icons, as a result of seeing the icons almost instantly. Participants expressed that they felt lost on the map (n = 13) due to seeing to little of the map (n = 10), especially when no other icon was visible. Participants also felt that the small display size forced them to search too much (n = 5) and see little to no relations with other objects (n = 3).

5 DISCUSSION

In this section, we discuss the results of our laboratory experiment in the context of findings from prior work. We structure our discussion based on our research questions concerning *spatial memory*, *workload*, and *user experience*.

5.1 Spatial Memory

In terms of spatial memory, MEDIUM-AR and LARGE-AR significantly improved navigation path length and task completion time, thereby also improving overall navigation speed. While MEDIUM-AR also clearly resulted in a significantly higher learning effect and recall accuracy, we did not find any similar significant effects for LARGE-AR. Thus, our findings suggest that – similar to findings from Rädle et al. [60] and Zagermann et al. [78] – there is a "sweet spot" for display size in terms of spatial memory and that a larger available display size improves navigation performance. Prior findings by Gao et al. [25] and Uddin et al. [74] indicate that additional landmarks (e.g., seeing more icons in larger conditions) can facilitate the formation of spatial memory, which is in line with our qualitative findings. Despite this effect and unlike Rädle et al. [60], however, participants actually performed slightly worse again beyond the MEDIUM-AR size. Surprisingly, SMALL-AR also slightly decreased spatial memory despite being bigger than NO-AR – further deviating from prior findings [26, 60, 74]. This implies that AR extensions have an implicit cost associated with splitting the display into a real and virtual screen.

Contrary to our expectations, maximum navigation speed did not see a significant difference between the conditions. We expected the maximum navigation to increase with display size, as participants could be more confident in quickly flicking across the information space (as they are "scrolling into the unknown") and scanning the available area for the symbol. Instead, our results show a relatively consistent maximum navigation speed, indicating that the time required to visually scan the map and the time required to navigate to a new segment were roughly consistent across all conditions.

In summary, a larger display size generally contributes to a better task completion time. Spatial memory, however, can actually decrease for small virtually-extended screens. Here, MEDIUM-AR presents a "sweet-spot" for spatial memory, after which spatial memory starts to slightly degrade again.

5.2 Workload

In line with findings by Caluya et al. [12], our results show that head rotation increases with larger display sizes (i.e., insufficient virtual field of view for the given content is compensated by increased head movement). The LARGE-AR size comes at the cost of significantly more head rotation than both NO-AR and SMALL AR. Although MEDIUM-AR fits comfortably within the participant's field, our data also shows a significant increase in head movement and rotation compared to SMALL-AR and NO-AR. Conversely, the cognitive load significantly decreased for the MEDIUM-AR and LARGE-AR conditions, which indicates a trade-off between ergonomics in terms of head rotation and cognitive load and is in line with prior findings by Rädle et al. [60].

Similarly, our data also shows that virtually-extended display size correlates with how much participants spent looking at the virtual screen. However, our analysis reveals that most of the participants' gaze is still focused on or around the smartphone. While the increased display space is well-used in the SMALL-AR condition, adding more display space beyond the MEDIUM-AR sees barely any use. Given the significant increase in head rotation for LARGE-AR, the extra space only caused additional physical load with little to no additional benefit.

To our surprise, gaze movement did not significantly increase with display size, but rather whether or not an virtually-extended screen was used. Since we ensured that both (real and virtual) screens were on the same focal plane (i.e., using a video see-through HWD, cf. [20, 30, 54]) and that there was no visible gap between the displays (cf. [53]), we expected gaze movement to correlate with display size. Here, further research is necessary to explore the actual underlying causes. For example, the smartphone bezel might provide a physical frame of reference, thereby further splitting the screen into two distinct displays, contributing to an increase in context switching (cf. [28, 62]). Alternatively, the smartphone may provide a "sweet spot" in terms of angular coverage [62], thus fitting well within the fovea-wide field of view. Another reason might be due to our participants' familiarity with a smartphone's physical affordances: By introducing a virtual screen, we added an unfamiliar affordance, thus leading to a higher cognitive load without much added benefit for small extensions: *"[During SMALL-AR], I was still focused on the smartphone. It took me a while until I looked at the [VESAD] again, it took me a while to convince myself that I can peek across the border"*—[P18].

In summary, although larger virtually-extended displays are worse in terms of ergonomics, the increased display space was well-used until MEDIUM-AR. However, there is no benefit in increasing the size beyond MEDIUM-AR (cf. [60]). In contrast, a small virtually-extended display causes a disproportionately high cognitive workload.

5.3 User Experience

Our results show a clear subjective preference for any AR display extension over the NO-AR condition across attractiveness, pragmatic qualities, and hedonic qualities. Here, both MEDIUM-AR and LARGE-AR were consistently rated significantly higher than the NO-AR condition, which was confirmed in our semi-structured interviews: *"It's not too much to overwhelm you with information, but it's also not too small so that you have to search too much.* [...] *It's like, if your monitor is too big you start to lose track of your cursor."*—[P9].

Although MEDIUM-AR is ranked slightly better than LARGE-AR (cf. [78]), the difference might be due to a legacy bias [57]: Most of our participants are likely used to working on a notebook or desktop monitor, thus explaining their preference towards MEDIUM-AR over LARGE-AR. In contrast, other participants might use a television-sized monitor or multi-monitor environment in their everyday life, thus preferring LARGE-AR over MEDIUM-AR.

In summary, participants consistently favored larger AR extensions such as MEDIUM-AR and LARGE-AR. The subjective size preference between MEDIUM-AR and LARGE-AR might depends on the use case and participants' day-to-day experiences.

6 LIMITATIONS & FUTURE WORK

Due to our narrow focus on comparing virtually-extended display sizes against a smartphone baseline condition, our study has the following limitations.

We intentionally limited the overall map size to accommodate both NO-AR and LARGE-AR. While LARGE-AR may see further improvements with a larger information space (e.g., with regard to task completion time), this would negatively affect the NO-AR condition. Future studies could exclude the NO-AR condition to better study the effects of larger virtually-extended display. In addition, we assigned each condition dedicated icons and icon locations to achieve full counterbalancing of our conditions. Although we ensured that icons were equally placed between different maps and our results are mostly in line with prior work (e.g., [60, 78]), our results may be influenced by the layout of each map. We also intentionally switched to a desktop interface during the *object location recall phase*. While this reduced the ecological validity, it allowed us to better compare our results with prior studies (e.g., [60, 78]) and isolate spatial memory from other confounding influences (e.g., muscle memory).

Another limitation may be given by the use of a video see-through AR HWD. We intentionally decided against an optical see-through AR HWD to avoid confounding factors with respect to different focal planes (cf. [20, 30, 54]). However, video see-through HWDs are more cumbersome than optical see-through HWDs and have a reduced real-world field of view, which may negatively impact larger display sizes. In this regard, prior research already indicated that a restricted virtual field of view does not negatively impact spatial memory [12]. As AR HWDs continue to improve, further studies are necessary to investigate the impact of device ergonomics on larger virtually-extended display sizes. In line with this, future research could investigate, whether a fully virtually simulated AR environment (e.g., combining a virtual reality HWD and a physical prop) could be used to increase the internal validity of the measurements – albeit at the cost of ecological validity.

Since our study compared different virtually-extended screen sizes with a fixed input modality as a static smartphone-sized peephole [49], there are many aspects left unexplored. For instance, future work could explore the effects of different physical screen sizes (e.g., smartwatches, tablets) and their relation to virtually-extended screens. Furthermore, future work could also investigate dynamic peephole navigation as input modality by tracking the smartphone in space (e.g., see [56]): Replacing touch interaction with physically moving the handheld device to explore the information space could further improve spatial memory (cf. [78]).

7 DESIGN & RESEARCH IMPLICATIONS

In this section, we synthesize our findings from our laboratory experiment (see Sections 3 and 4) and our discussion (see Section 5) to provide design implications (D1–D4) for the design of VESADs and research implications (I1–4) to inspire future work. We summarize our key implications in call-out boxes at the end of this section.

While our results generally show that any virtually-extended display size can have significant benefits in terms of task completion time and user experience **(D1)**, our results consistently favor MEDIUM-AR over LARGE-AR. Looking at our workload results, we conclude that MEDIUM-AR presents the best trade-off in terms of performance (e.g., task completion time, navigation path length) and increased task load (e.g., subjective and objective physical demand) **(D2)**. However, a much finer granularity in the comparison of display sizes is necessary to find the "tipping point" that presents the ideal virtually-extended screen size **(I1)**.

For LARGE-AR, the advantages gained from increasing the display sizes begin to diminish (cf. [60]), while the workload (e.g., ergonomics) starts to outweigh the benefits **(D3)**. In this regard, we could fully realize the potential of AR to bend the information space around the user (i.e., similar to off-the-shelf ultra-wide monitors or CAVE [16] systems). Here, we could investigate whether bending the virtually-extended screen shows any benefits for different virtually-extended display sizes (e.g., comparing CAVE-like system, ultra-wide monitor, and straight display) **(I2)**.

For SMALL-AR, our objective results contradict the subjective user preferences: While users appreciate even the smallest virtually-extended display size (e.g., subjectively increased efficiency), our objective results indicate adverse effects in terms of spatial memory (e.g., navigation path length, recall accuracy). In terms of spatial memory and workload, we therefore conclude that a small virtually-extended screen is worse than providing no virtually-extended screen **(D4)**. Although our results indicate clues (e.g., increased eye gaze movement) why SMALL-AR performed worse, further research is necessary to find the underlying cause **(I3)**. However, prior works (e.g., [17, 43, 54, 63, 79]) already show several promising scenarios of utilizing the small space next to a display (e.g., for pushing UI elements out of the screen). As we did not study the interaction with the AR content *per se* (e.g., via mid-air input [54]), we suggest to systematically investigate use cases and scenarios for interaction with out-of-screen UI elements (comparable to the size of SMALL-AR) regarding the effect on a user's workload as the interaction might alleviate shortcomings related to spatial memory **(I4)**.

Design Implications

- **D1** Any virtually-extended screen size is beneficial in terms of user experience and task completion time for navigation tasks.
- **D2** Virtually extending a smartphone to the size of a desktop monitor presents a "sweet spot" in terms of spatial memory, workload, and user experience.
- **D3** Virtual extensions that are larger than a desktop monitor are detrimental to spatial memory and ergonomics with little additional benefit.
- **D4** Small display extension negatively impact spatial memory and workload.

Research Implications

- **I1** Find "tipping point" of virtually-extended display size that represents optimal trade-off between ergonomics and performance.
- **I2** Compare different levels of bending the virtually-extended display around the user (cf. [25, 26, 47]).
- **I3** Investigate effects of context switch between real and virtual displays (cf. [62]).
- **I4** Investigate if interaction with out-of-screen UI elements can alleviate shortcomings of small display extensions (cf. [54]).

8 CONCLUSION

In this work, we investigate the effects of virtually-extended display sizes on spatial memory, workload, and user experience. For this, we combine an augmented reality head-worn display with a smartphone to seamlessly extend the smartphone with a virtually-extended screen-aligned display. We conducted a controlled laboratory experiment with 24 participants using a within-subject design to compare a baseline condition (smartphone with no virtual screen extension) with three commonly found display sizes (smartphone extended with virtual screen of tablet size, desktop monitor size, or television size). Our experiment used a well-established task to measure spatial memory, which consists of a navigation phase and an object location recall phase. Our findings confirm results from prior work that bigger (virtually-extended) screens contribute to better task completion times, but with diminishing returns. However, our results also show that spatial memory benefits from a "sweet spot" of virtually-extended display sizes: If the virtual display extension is too small, the disadvantages of splitting the screen into a real and virtual screen outweigh the benefits of an increased screen size; if the extension is too large, device ergonomics start to supersede any benefit gained from extending the screen size. We found that virtually extending a smartphone to the size of a desktop monitor provides the best trade-off, consistently leading to a significantly improved spatial memory, decreased workload, and better user experience. Based on our results, we synthesize design and research implications for virtually-extended screen-aligned displays. Our work contributes towards a better understanding of virtually extending a physical screen using hybrid user interfaces.

ACKNOWLEDGMENTS

This research was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

REFERENCES

[1] Gary L. Allen (Ed.). 2004. *Human Spatial Memory* (first ed.). Psychology Press. https://doi.org/10.4324/9781410609984

[2] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Montreal QC, Canada, 1–15. https://doi.org/10.1145/3173574.3173759

[3] Myroslav Bachynskyi, Gregorio Palmas, Antti Oulasvirta, Jürgen Steimle, and Tino Weinkauf. 2015. Performance and Ergonomics of Touch Surfaces: A Comparative Study Using Biomechanical Simulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 1817–1826. https://doi.org/10.1145/2702123.2702607

[4] Joanna Bergstrom-Lehtovirta, Sebastian Boring, and Kasper Hornbæk. 2017. Placing and Recalling Virtual Items on the Skin. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 1497–1507. https://doi.org/10.1145/3025453.3026030

[5] Verena Biener, Daniel Schneider, Travis Gesslein, Alexander Otte, Bastian Kuth, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. 2020. Breaking the Screen: Interaction Across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers. *arXiv:2008.04559 [cs]* 26, 12 (dec 2020), 3490–3502. https://doi.org/10.1109/TVCG.2020.3023567 arXiv:2008.04559 [cs]

[6] Sebastian Boring, Marko Jurmu, and Andreas Butz. 2009. Scroll, Tilt or Move It: Using Mobile Phones to Continuously Control Pointers on Large Public Displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. Association for Computing Machinery, New York, NY, USA, 161–168. https://doi.org/10.1145/1738826.1738853

[7] Gerd Bruder, Frank Steinicke, and Wolfgang Stürzlinger. 2013. Effects of Visual Conflicts on 3D Selection Task Performance in Stereoscopic Display Environments. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 115–118. https://doi.org/10.1109/3DUI.2013.6550207

[8] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. 2019. Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–28. https://doi.org/10.1145/3290605.3300792

[9] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating Smartphone-based Pan and Zoom in 3D Data Spaces in Augmented Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '19*. ACM Press, Taipei, Taiwan, 1–13. https://doi.org/10.1145/3338286.3340113

[10] Simon Butscher, Sebastian Hubenschmid, Jens Müller, Johannes Fuchs, and Harald Reiterer. 2018. Clusters, Trends, and Outliers: How Immersive Technologies Can Facilitate the Collaborative Analysis of Multidimensional Data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–12. https://doi.org/10.1145/3173574.3173664

[11] Simon Butscher and Harald Reiterer. 2016. Applying Guidelines for the Design of Distortions on Focus+Context Interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, Bari Italy, 244–247. https://doi.org/10.1145/2909132.2909284

[12] Nicko R. Caluya, Alexander Plopski, Christian Sandor, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. 2022. Does Overlay Field of View in Head-Mounted Displays Affect Spatial Memorization? *Computers & Graphics* 102 (Feb. 2022), 554–565. https://doi.org/10.1016/j.cag.2021.09.004

[13] Jessica R. Cauchard, Markus Löchtefeld, Pourang Irani, Johannes Schoening, Antonio Krüger, Mike Fraser, and Sriram Subramanian. 2011. Visual Separation in Mobile Multi-Display Environments. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11*. ACM Press, Santa Barbara, California, USA, 451. https://doi.org/10.1145/2047196.2047256

[14] Li-Wei Chan, Hui-Shan Kao, Mike Y. Chen, Ming-Sui Lee, Jane Hsu, and Yi-Ping Hung. 2010. Touching the Void: Direct-Touch Interaction for Intangible Displays, In Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2625–2634. https://doi.org/10.1145/1753326.1753725

[15] Andy Cockburn and Bruce McKenzie. 2002. Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. Association for Computing Machinery, Minneapolis, Minnesota, USA, 203–210. https://doi.org/10.1145/503376.503413

[16] Carolina Cruz-Neira, Dj Daniel J Sandin, and Ta Thomas A DeFanti. 1993. Surround-Screen Projection-Based Virtual Reality : The Design and Implementation of the CAVE. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '93)*. ACM Press, New York, NY, USA, 135–142. https://doi.org/10.1145/166117.166134

[17] Francesco Riccardo Di Gioia, Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. 2022. Investigating the Use of AR Glasses for Content Annotation on Mobile Devices. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (Nov. 2022), 430–447. https://doi.org/10.1145/3567727

[18] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376628

[19] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load *Vis-à-Vis* Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173856

[20] Anna Eiberger, Per Ola Kristensson, Susanne Mayr, Matthias Kranz, and Jens Grubert. 2019. Effects of Depth Layer Switching between an Optical See-Through Head-Mounted Display and a Body-Proximate Display. In *Symposium on Spatial User Interaction*. ACM, New Orleans LA USA, 1–9. https://doi.org/10.1145/3357251.3357588

[21] Steven Feiner and Ari Shamash. 1991. Hybrid User Interfaces: Breeding Virtually Bigger Interfaces for Physically Smaller Computers. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology (UIST '91)*. Association for Computing Machinery, Hilton Head, South Carolina, USA, 9–17. https://doi.org/10.1145/120782.120783

[22] Thibault Friedrich, Arnaud Prouzeau, and Michael McGuffin. 2021. The Effect of Increased Body Motion in Virtual Reality on a Placement-Retrieval Task. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. ACM, Osaka Japan, 1–5. https://doi.org/10.1145/3489849.3489888

[23] Bruno Fruchard, Eric Lecolinet, and Olivier Chapuis. 2018. Impact of Semantic Aids on Command Memorization for On-Body Interaction and Directional Gestures. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. ACM, Castiglione della Pescaia Grosseto Italy, 1–9. https://doi.org/10.1145/3206505.3206524

[24] BoYu Gao, Zhiguo Chen, Xian Chen, Huawei Tu, and Feiran Huang. 2021. The Effects of Audiovisual Landmarks on Spatial Learning and Recalling for Image Browsing Interface in Virtual Environments. *Journal of Systems Architecture* 117 (Aug. 2021), 102096. https://doi.org/10.1016/j.sysarc.2021.102096

[25] BoYu Gao, Byungmoon Kim, Jee-In Kim, and HyungSeok Kim. 2019. Amphitheater Layout with Egocentric Distance-Based Item Sizing and Landmarks for Browsing in Virtual Reality. *International Journal of Human–Computer Interaction* 35, 10 (June 2019), 831–845. https://doi.org/10.1080/10447318.2018.1498654

[26] BoYu Gao, HyungSeok Kim, Byungmoon Kim, and Jee-In Kim. 2018. Artificial Landmarks to Facilitate Spatial Learning and Recalling for Curved Visual Wall Layout in Virtual Reality. In *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, Shanghai, 475–482. https://doi.org/10.1109/BigComp.2018.00076

[27] Varun Gaur, Md. Sami Uddin, and Carl Gutwin. 2018. Multiplexing Spatial Memory: Increasing the Capacity of FastTap Menus with Multiple Tabs. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Barcelona Spain, 1–13. https://doi.org/10.1145/3229434.3229482

[28] Paul Ginns. 2006. Integrating Information: A Meta-Analysis of the Spatial Contiguity and Temporal Contiguity Effects. *Learning and Instruction* 16, 6 (Dec. 2006), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

[29] Andrew Greaves and Enrico Rukzio. 2008. Evaluation of Picture Browsing Using a Projector Phone. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '08*. ACM Press, Amsterdam, The Netherlands, 351. https://doi.org/10.1145/1409240.1409286

[30] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. MultiFi: Multi Fidelity Interaction with Displays On and Around the Body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, Seoul, Republic of Korea, 3933–3942. https://doi.org/10.1145/2702123.2702331

[31] Alina Hang, Enrico Rukzio, and Andrew Greaves. 2008. Projector Phone: A Study of Using Mobile Phones with Integrated Projector for Interaction with Maps. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '08*. ACM Press, Amsterdam, The Netherlands, 207. https://doi.org/10.1145/1409240.1409263

[32] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-Body Interaction: Armed and Dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction (TEI '12)*. Association for Computing Machinery, New York, NY, USA, 69–76. https://doi.org/10.1145/2148131.2148148

[33] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[34] Kasper Hornbæk, Benjamin B. Bederson, and Catherine Plaisant. 2002. Navigation Patterns and Usability of Zoomable User Interfaces with and without an Overview. *ACM Transactions on Computer-Human Interaction* 9, 4 (Dec. 2002), 362–389. https://doi.org/10.1145/586081.586086

[35] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmqvist, and Harald Reiterer. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–20. https://doi.org/10.1145/3491102.3517550

[36] Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. 2021. STREAM: Exploring the Combination of Spatially-Aware Tablets with Augmented Reality Head-Mounted Displays for Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445298

[37] Sebastian Hubenschmid, Johannes Zagermann, Daniel Fink, Jonathan Wieland, Tiare Feuchtner, and Harald Reiterer. 2021. Towards Asynchronous Hybrid User Interfaces for Cross-Reality Interaction. In *ISS'21 Workshop Proceedings: "Transitional Interfaces in Mixed and Cross-Reality: A New Frontier?"*, Hans-Christian Jetter, Jan-Henrik Schröder, Jan Gugenheimer, Mark Billinghurst, Christoph Anthes, Mohamed Khamis, and Tiare Feuchtner (Eds.). https://doi.org/10.18148/kops/352-2-84mm0sggczq02

[38] Yvonne Jansen, Jonas Schjerlund, and Kasper Hornbæk. 2019. Effects of Locomotion and Visual Overview on Spatial Memory When Interacting with Wall Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300521

[39] Hans-Christian Jetter, Svenja Leifert, Jens Gerken, Sören Schubert, and Harald Reiterer. 2012. Does (Multi-)Touch Aid Users' Spatial Memory and Navigation in 'panning' and in 'Zooming & Panning' UIs?. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*. ACM Press, Capri Island, Italy, 83. https://doi.org/10.1145/2254556.2254575

[40] Bonifaz Kaufmann and David Ahlström. 2013. Studying Spatial Memory and Map Navigation Performance on Projector Phones with Peephole Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 3173–3176. https://doi.org/10.1145/2470654.2466434

[41] Daniel Klinkhammer, Jan Oke Tennie, Paula Erdoes, and Harald Reiterer. 2013. Body Panning: A Movement-Based Navigation Technique for Large Interactive Surfaces. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*. ACM, St. Andrews Scotland, United Kingdom, 37–40. https://doi.org/10.1145/2512349.2512822

[42] Pascal Knierim, Dimitri Hein, Albrecht Schmidt, and Thomas Kosch. 2021. The SmARtphone Controller: Leveraging Smartphones as Input and Output Modality for Improved Interaction within Mobile Augmented Reality Environments. *i-com* 20, 1 (April 2021), 49–61. https://doi.org/10.1515/icom-2021-0003

[43] 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 (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445593

[44] Huy Viet Le, Sven Mayer, Patrick Bader, and Niels Henze. 2018. Fingers' Range and Comfortable Area for One-Handed Smartphone Interaction Beyond the Touchscreen. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173605

[45] Chi-Jung Lee and Hung-Kuo Chu. 2018. Dual-MR: Interaction with Mixed Reality Using Smartphones. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, Tokyo, Japan, 1–2. https://doi.org/10.1145/3281505.3281618

[46] Svenja Leifert. 2011. The Influence of Grids on Spatial and Content Memory. In *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '11*. ACM Press, Vancouver, BC, Canada, 941. https://doi.org/10.1145/1979742.1979522

[47] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. 2022. Effects of Display Layout on Spatial Memory for Immersive Environments. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (Nov. 2022), 468–488. https://doi.org/10.1145/3567729

[48] Weizhou Luo, Eva Goebel, Patrick Reipschlager, Mats Ole Ellenberg, and Raimund Dachselt. 2021. Exploring and Slicing Volumetric Medical Data in Augmented Reality Using a Spatially-Aware Mobile Device. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, Bari, Italy, 334–339. https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00076

[49] Sumit Mehra, Peter Werkhoven, and Marcel Worring. 2006. Navigating on Handheld Displays: Dynamic versus Static Peephole Navigation. *ACM Transactions on Computer-Human Interaction* 13, 4 (Dec. 2006), 448–457. https://doi.org/10.1145/1188816.1188818

[50] Jens Müller, Roman Rädle, Hans-Christian Jetter, and Harald Reiterer. 2015. An Experimental Comparison of Vertical and Horizontal Dynamic Peephole Navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 1523–1526. https://doi.org/10.1145/2702123.2702227

[51] Jens Müller, Roman Rädle, and Harald Reiterer. 2016. Virtual Objects as Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1245–1249. https://doi.org/10.1145/2858036.2858043

[52] Jens Müller, Johannes Zagermann, Jonathan Wieland, Ulrike Pfeil, and Harald Reiterer. 2019. A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices. In *Proceedings of Mensch Und Computer 2019*. ACM, Hamburg Germany, 399–410. https://doi.org/10.1145/3340764.3340773

[53] Miguel A. Nacenta, Regan L. Mandryk, and Carl Gutwin. 2008. Targeting across Displayless Space. In *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*. ACM Press, Florence, Italy, 777. https://doi.org/10.1145/1357054.1357178

[54] Erwan Normand and Michael J. McGuffin. 2018. Enlarging a Smartphone with AR to Create a Handheld VESAD (Virtually Extended Screen-Aligned Display). In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Munich, Germany, 123–133. https://doi.org/10.1109/ISMAR.2018.00043

[55] Kenton O'Hara, Abigail Sellen, and Richard Bentley. 1999. Supporting Memory for Spatial Location While Reading from Small Displays. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems - CHI '99*. ACM Press, Pittsburgh, Pennsylvania, 220. https://doi.org/10.1145/632716.632853

[56] Michel Pahud, Ken Hinckley, Shamsi Iqbal, Abigail Sellen, and Bill Buxton. 2013. Toward Compound Navigation Tasks on Mobiles via Spatial Manipulation. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services - MobileHCI '13*. ACM Press, Munich, Germany, 113. https://doi.org/10.1145/2493190.2493210

[57] Thomas Plank, Hans-Christian Jetter, Roman Rädle, Clemens N Klokmose, Thomas Luger, and Harald Reiterer. 2017. Is Two Enough?! Studying Benefits, Barriers, and Biases of Multi-Tablet Use for Collaborative Visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, New York, New York, USA, 4548–4560. https://doi.org/10.1145/3025453.3025537

[58] Kathrin Probst. 2016. Peripheral Interaction in Desktop Computing: Why It's Worth Stepping Beyond Traditional Mouse and Keyboard. In *Peripheral Interaction*, Saskia Bakker, Doris Hausen, and Ted Selker (Eds.). Springer International Publishing, Cham, 183–205. https://doi.org/10.1007/978-3-319-29523-7_9

[59] Roman Rädle, Hans-Christian Jetter, Simon Butscher, and Harald Reiterer. 2013. The Effect of Egocentric Body Movements on Users' Navigation Performance and Spatial Memory in Zoomable User Interfaces. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*. ACM, St. Andrews Scotland, United Kingdom, 23–32. https://doi.org/10.1145/2512349.2512811

[60] Roman Rädle, Hans-Christian Jetter, Jens Müller, and Harald Reiterer. 2014. Bigger Is Not Always Better: Display Size, Performance, and Task Load during Peephole Map Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 4127–4136. https://doi.org/10.1145/2556288.2557071

[61] Umar Rashid, Miguel A. Nacenta, and Aaron Quigley. 2012. The Cost of Display Switching: A Comparison of Mobile, Large Display and Hybrid UI Configurations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. Association for Computing Machinery, Capri Island, Italy, 99–106. https://doi.org/10.1145/2254556.2254577

[62] Umar Rashid, Miguel A. Nacenta, and Aaron Quigley. 2012. Factors Influencing Visual Attention Switch in Multi-Display User Interfaces: A Survey. In *Proceedings of the 2012 International Symposium on Pervasive Displays - PerDis '12*. ACM Press, Porto, Portugal, 1–6. https://doi.org/10.1145/2307798.2307799

[63] Patrick Reipschläger and Raimund Dachselt. 2019. DesignAR: Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces - ISS '19*. ACM Press, Daejeon, Republic of Korea, 29–41. https://doi.org/10.1145/3343055.3359718

[64] Patrick Reipschläger, Tamara Flemisch, and Raimund Dachselt. 2021. Personal Augmented Reality for Information Visualization on Large Interactive Displays. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (feb 2021), 1182–1192. https://doi.org/10.1109/TVCG.2020.3030460 arXiv:2009.03237

[65] Md. Sami Uddin and Carl Gutwin. 2021. The Image of the Interface: How People Use Landmarks to Develop Spatial Memory of Commands in Graphical Interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445050

[66] Joey Scarr, Andy Cockburn, and Carl Gutwin. 2013. Supporting and Exploiting Spatial Memory in User Interfaces. *Foundations and Trends® in Human–Computer Interaction* 6, 1 (2013), 1–84. https://doi.org/10.1561/1100000046

[67] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40. https://doi.org/10.9781/ijimai.2017.445

[68] Mickael Sereno, Stéphane Gosset, Lonni Besançon, and Tobias Isenberg. 2022. Hybrid Touch/Tangible Spatial Selection in Augmented Reality. *Computer Graphics Forum* 41, 3 (June 2022), 403. https://doi.org/10.1111/cgf.14550

[69] Amy Skopik and Carl Gutwin. 2005. Improving Revisitation in Fisheye Views with Visit Wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Portland Oregon USA, 771–780. https://doi.org/10.1145/1054972.1055079

[70] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. 2017. VRRRRoom: Virtual Reality for Radiologists in the Reading Room. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM Press, New York, NY, USA, 4057–4062. https://doi.org/10.1145/3025453.3025566

[71] Tianchen Sun, Yucong Ye, Issei Fujishiro, and Kwan-Liu Ma. 2019. Collaborative Visual Analysis with Multi-level Information Sharing Using a Wall-Size Display and See-Through HMDs. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 11–20. https://doi.org/10.1109/PacificVis.2019.00010

[72] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300243

[73] Desney S. Tan, Randy Pausch, Jeanine K. Stefanucci, and Dennis R. Proffitt. 2002. Kinesthetic Cues Aid Spatial Memory. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems - CHI '02*. ACM Press, Minneapolis, Minnesota, USA, 806. https://doi.org/10.1145/506443.506607

[74] Md. Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017. The Effects of Artificial Landmarks on Learning and Performance in Spatial-Memory Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 3843–3855. https://doi.org/10.1145/3025453.3025497

[75] Katja Vock, Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. 2021. IDIAR: Augmented Reality Dashboards to Supervise Mobile Intervention Studies. In *Mensch Und Computer 2021 (MuC '21)*. ACM, New York, NY. https://doi.org/10.1145/3473856.3473876

[76] Johannes Zagermann, Sebastian Hubenschmid, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. *it - Information Technology* 0, 0 (Sept. 2022). https://doi.org/10.1515/itit-2022-0031

[77] Johannes Zagermann, Ulrike Pfeil, Carmela Acevedo, and Harald Reiterer. 2017. Studying the Benefits and Challenges of Spatial Distribution and Physical Affordances in a Multi-Device Workspace. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. ACM, Stuttgart Germany, 249–259. https://doi.org/10.1145/3152832.3152855

[78] Johannes Zagermann, Ulrike Pfeil, Daniel Fink, Philipp von Bauer, and Harald Reiterer. 2017. Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 1899–1910. https://doi.org/10.1145/3025453.3026001

[79] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376233

A STATISTICAL TESTS

This appendix reports on the statistical findings presented in Section 4. Our data is also available on our project page[3]. We omitted all pairwise comparisons that did not show any statistically significant differences.

---

[3] DOI: 10.18419/darus-3326

Table 1: Results of the Friedman's test for navigation path lengths between repetitions, as visualized in Figure 3 (A). Statistically significant entries are marked with a star*.

Repetition	Result
Repetition 2	$\chi^2(3) = 44.10, p < .001^*$
Repetition 3	$\chi^2(3) = 44.50, p < .001^*$
Repetition 4	$\chi^2(3) = 53.55, p < .001^*$

Table 2: Results of the pairwise comparison of navigation path lengths between repetitions, as visualized in Figure 3 (A). Statistically significant entries are marked with a star*.

Comparison	Repetition 2	Repetition 3	Repetition 4
NO-AR ↔ MEDIUM-AR	$z = 1.750, p < .001^*$	$z = 1.750, p < .001^*$	$z = 1.958, p < .001^*$
NO-AR ↔ LARGE-AR	$z = 1.750, p < .001^*$	$z = 1.583, p < .001^*$	$z = 1.708, p < .001^*$
SMALL-AR ↔ MEDIUM-AR	$z = 1.750, p < .001^*$	$z = 1.917, p < .001^*$	$z = 2.125, p < .001^*$
SMALL-AR ↔ LARGE-AR	$z = 1.750, p < .001^*$	$z = 1.750, p < .001^*$	$z = 1.875, p < .001^*$

Table 3: Results of the Friedman's test for learning effects between repetitions for each condition. Statistically significant entries are marked with a star*.

Condition	Result
NO-AR	$\chi^2(3) = 39.35, p < .001^*$
SMALL-AR	$\chi^2(3) = 40.35, p < .001^*$
MEDIUM-AR	$\chi^2(3) = 42.65, p < .001^*$
LARGE-AR	$\chi^2(3) = 42.95, p < .001^*$

Table 4: Results of the pairwise comparison of learning effects between repetitions for each condition. Statistically significant entries are marked with a star*.

Condition	NO-AR	SMALL-AR	MEDIUM-AR	LARGE-AR
Repetition 1–2	$z = 0.708, p = .344$	$z = 0.500, p = 1$	$z = 1.167, p = .010^*$	$z = 0.958, p = .061$
Repetition 1–3	$z = 1.667, p < .001^*$	$z = 1.542, p < .001^*$	$z = 1.958, p < .001^*$	$z = 1.750, p < .001^*$

Table 5: Results of the Friedman's test for the raw NASA TLX, as visualized in Figure 5 (A). Statistically significant entries are marked with a star*.

Scale	Result
Mental Demand	$\chi^2(3) = 21.618, p < .001^*$
Physical Demand	$\chi^2(3) = 14.397, p = .002^*$
Temporal Demand	$\chi^2(3) = 25.07, p < .001^*$
Performance	$\chi^2(3) = 10.235, p = .017^*$
Effort	$\chi^2(3) = 20.127, p < .001^*$
Frustration	$\chi^2(3) = 35.226, p < .001^*$
Overall	$\chi^2(3) = 38.949, p < .001^*$

Table 6: Results of the pairwise comparison of the raw NASA TLX, as visualized in Figure 5 (A). Statistically significant entries are marked with a star*.

Scale	NO-AR ↔ SMALL-AR	NO-AR ↔ MEDIUM-AR	SMALL-AR ↔ MEDIUM-AR	NO-AR ↔ LARGE-AR
Mental Demand	$z = 0.333, p = 1$	$z = 1.521, p < .001^*$	$z = 1.188, p = .009^*$	$z = 1.063, p = .026^*$
Physical Demand	$z = 0.833, p = .152$	$z = 1.229, p = .006^*$	$z = 0.396, p = 1$	$z = 0.354, p = 1$
Temporal Demand	$z = 0.958, p = .061$	$z = 1.479, p < .001^*$	$z = 0.521, p = .974$	$z = 1.646, p < .001^*$
Performance	$z = 0.313, p = 1$	$z = 1.042, p = .031^*$	$z = 0.729, p = .302$	$z = 0.813, p = .175$
Effort	$z = 1.000, p = .044^*$	$z = 1.542, p < .001^*$	$z = 0.542, p = .877$	$z = 1.125, p = .015^*$
Frustration	$z = 0.688, p = .390$	$z = 2.000, p < .001^*$	$z = 1.313, p = .003^*$	$z = 1.396, p = .001^*$
Overall	$z = 0.750, p = .265$	$z = 2.229, p < .001^*$	$z = 1.479, p < .001^*$	$z = 1.354, p = .002^*$

Table 7: Results of the Friedman's test of the user experience questionnaire, as visualized in Figure 5 (B). Statistically significant entries are marked with a star*.

Scale	Friedman's Test
Attractiveness	$\chi^2(3) = 25.146, p < .001^*$
Perspicuity	$\chi^2(3) = 24.614, p < .001^*$
Efficiency	$\chi^2(3) = 36.500, p < .001^*$
Dependability	$\chi^2(3) = 30.434, p < .001^*$
Stimulation	$\chi^2(3) = 18.067, p < .001^*$
Novelty	$\chi^2(3) = 21.790, p < .001^*$

Table 8: Results of the pairwise comparison of the user experience questionnaire, as visualized in Figure 5 (B). Statistically significant entries are marked with a star*.

Scale	NO-AR ↔ SMALL-AR	NO-AR ↔ MEDIUM-AR	NO-AR ↔ LARGE-AR
Attractiveness	$z = -0.854, p = .131$	$z = -1.667, p < .001^*$	$z = -1.479, p < .001^*$
Perspicuity	$z = -0.896, p = .097$	$z = -1.646, p = .001^*$	$z = -1.375, p < .001^*$
Efficiency	$z = -1.250, p = .005^*$	$z = -1.896, p < .001^*$	$z = -1.854, p < .001^*$
Dependability	$z = -1.500, p < .001^*$	$z = -1.812, p < .001^*$	$z = -1.521, p < .001^*$
Stimulation	$z = -0.646, p = .499$	$z = -1.062, p = .026^*$	$z = -1.458, p = .001^*$
Novelty	$z = -0.667, p = .442$	$z = -1.146, p = .013^*$	$z = -1.604, p < .001^*$