The Effect of Increased Body Motion in Virtual Reality on a Placement-Retrieval Task

Thibault Friedrich
thibault.friedrich@gmail.com
École de technologie supérieure
Montreal, Canada

Arnaud Prouzeau
arnaud.prouzeau@monash.edu
Inria & LaBRI
Bordeaux, France

Michael J. McGuffin
michael.mcguffin@etsmtl.ca
École de technologie supérieure
Montreal, Canada

ABSTRACT

Previous work has shown that increased effort and use of one's body can improve memory. When positioning windows inside a virtual reality, does the use of a larger volume, and using one's legs to move around, improve ability to later find the windows? The results of our experiment indicate there can be a modest benefit for spatial memory and retrieval time, but at the cost of increased time spent initially positioning the windows.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *Empirical studies in HCI*; Graphical user interfaces.

KEYWORDS

Virtual reality, locomotion, controlled experiment

ACM Reference Format:
Thibault Friedrich, Arnaud Prouzeau, and Michael J. McGuffin. 2021. The Effect of Increased Body Motion in Virtual Reality on a Placement-Retrieval Task. In *27th ACM Symposium on Virtual Reality Software and Technology (VRST '21), December 8–10, 2021, Osaka, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3489849.3489888

1 INTRODUCTION

Immersive headsets, for both virtual reality (VR) and augmented reality (AR), are increasingly common, with more than 5,000 games for VR now available on the Steam video game platform. Office and knowledge workers may soon be performing more work with immersive headsets, for teleconferencing (e.g., https://spatial.io/), design reviews, training and simulations, or simply for the benefits of having information cover a larger field-of-view (FoV) than possible with one, two, or three physical desktop monitors. An often-proposed scenario is to have virtual windows or documents floating in front of and around the user [8, 10, 20, 22, 27].

A second, related trend is that office workers are increasingly using standing desks or desks that "convert" between sitting and standing modes [28], in part for health benefits. Immersive headsets could not only supplement or replace physical monitors, but also allow a user to work in standing or sitting modes, as desired. Standing and using a headset could also be more natural during (remote or collocated) collaborative work where users interact with a (virtual or physical) whiteboard or with a 3D model of a building, automobile, or other virtual content.

On one hand, such rich, immersive work environments can help users view and manage larger collections of information than with physical monitors. Users may even wish to switch between multiple collections of windows and documents, similar to how users of desktop computers sometimes switch between multiple virtual desktops [25]. On the other hand, surrounding users with content could also make users more reliant on their *spatial* memory to find documents, rather than other forms of memory (like the path of subfolders leading to a document), especially if the user is switching between multiple personal layouts and between collaborative and solo work spaces. A better understanding of factors affecting spatial memory could inform the design of appropriate user interfaces.

Many previous studies have measured the effects of different factors on spatial memory [29], such as display size [12, 17, 30], use of landmarks [9, 31], 2D vs 3D [3, 5, 26], and 'locomotion' (i.e., walking instead of standing or sitting still) [11, 12, 17]. Spatial memory can benefit from increased effort and time [2]. Performing actions with one's body can affect memory [4, 14, 18] and more particularly spatial memory [13, 24]. However, studies of spatial memory involving immersive headsets have been more limited: none of the studies we found [9, 15, 32] allowed a user to freely position content in 3D for later recall, which limits their implications for window management in VR and AR.

The current work involves spatial memory in VR, and focuses on a placement-retrieval task, as this is highly relevant to window management. The task is similar to that in previous studies [3, 5, 12, 26] which were not conducted in VR. In particular, we present the first study of the effect of locomotion on this placement-retrieval task in VR: how does standing in one spot to position and recall windows compare with taking small steps within a larger volume to perform the same task? We present evidence suggesting that locomotion yields a modest benefit in retrieval time and in spatial memory, but at the cost of increased time positioning windows.

2 BACKGROUND

Scarr et al. [29] survey both the scientific theory and practical implications of spatial memory, proposing over a dozen UI design guidelines supported by previous literature, as well as several "methodological cautions" and research questions. Their discussion of immersive environments is, however, limited, and they do not cover the effect of gestures or body motions on spatial memory.

Our work focuses on a "placement-retrieval task", where users are asked to freely position several items in a (2D or 3D) space before being asked to retrieve certain items previously positioned.

VRST '21, December 8–10, 2021, Osaka, Japan

ACM ISBN 978-1-4503-9092-7/21/12…$15.00
https://doi.org/10.1145/3489849.3489888

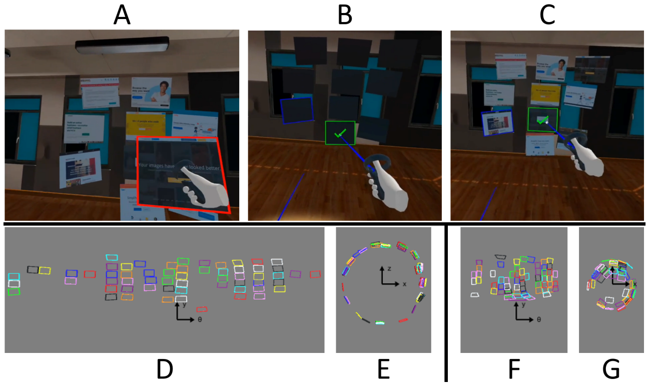

Figure 1: A: Placement phase in VR, in locomotion condition. The user has already positioned several windows, and is positioning one more window with their right hand controller, and the red border indicates that the window must be moved further away before being released to have a valid position. B: Blind recall phase, where the user has selected the correct window with the blue raycast, and success is shown with the green checkmark. C: Visible recall phase, where the content of all windows is visible. D and E: example placement of windows in the locomotion condition, seen in a cylindrical projection (y+ up, θ+ rotation right) and top projection (x+ right, z+ forward). F and G: in the stationary condition, the windows are placed closer to the origin.

Table 1 summarizes previous studies involving this task. In most of these studies, users positioned each item one-by-one and were not allowed to *re*position any items before moving on to the recall phase where items had to be retrieved. In the recall phase, items might still be visible, e.g., in the form of thumbnail images, which we refer to as *visible recall*. Visible recall makes it difficult for researchers to separate the performance of pure spatial memory versus visual search (i.e., simply recognizing the thumbnail image). This issue is discussed further in [29, section 3.1].

Table 1: Previous studies with a placement-retrieval task.

Study	[26]	[5]	[3]	[12]
Year	1998	1999	2002	2019
Number of users	32	9	69	80
Number of items positioned	100	n/a	99	100
Repositioning allowed?	yes	no	no	no
Number of items recalled	100	100	30	25
Type of recall	visible	mixed	visible	blind

In contrast, if the content of the items is blanked out (but the locations of items are still visible), we call this *blind recall*, which was used in recent work [12]. Visible recall is more realistic, whereas blind recall is more sensitive to effects on spatial memory. Unlike most previous studies, our study uses *both* visible and blind recall, to benefit from both advantages. Also, in the interest of having a realistic task, our study allowed users to reposition items before the recall phases, unlike most previous studies. Finally, none of the previous studies in Table 1 were done in VR.

3 STUDY

Our study was a controlled experiment in VR where users had to manually position dozens of windows in 3D space, one-by-one and with 6 degrees-of-freedom (DoF) for each window, and then recall where they had positioned certain windows. There were two main conditions, requiring users to either (1) stand in a single spot (*Stationary*) while positioning windows, limiting the user to occupy a smaller volume of space, or (2) take small steps with their legs (*Locomotion*) and fill a larger volume with the windows.

3.1 Equipment

The VR headset used for the experiment was the (first generation) Oculus Quest. This headset is untethered and performs inside-out tracking. The user held two "Oculus Touch" controllers in their hands, which are tracked with 6 DoF. Each controller has the same buttons available on it, of which we made use of the A, B, and Trigger buttons, as we mention below.

3.2 Tasks and Phases

For each of the two main conditions (*Stationary* and *Locomotion*), users went through a Placement phase where users positioned windows, followed by two Recall phases (Visible Recall and Blind Recall) where users had to remember and identify certain windows.

For each Placement phase, the system would generate 60 windows, which were rectangles textured with 60 different randomly-chosen screenshots of websites. These windows were initially invisible. The first window would be shown to the user at a standard initial position. The user would then need to use a controller (left or right, whichever they wanted) to drag-and-drop the window to a new position (and orientation), with full freedom over the 6 DoF of the window. Once the user was satisfied with the new position of this window, they would press the A button (on the right controller) to confirm, causing the second window to appear at the same initial position. The user would then need to drag-and-drop that 2nd window to a new position, and could also (optionally) drag-and-drop the first window to a new position. After pressing the A button to confirm, the 3rd window would appear, etc., until all 60 windows were positioned. The user was always free to re-position any of the windows by dragging them, for example, to create room for a new window or to change groupings or layout. When the user was satisfied with the layout of all 60 windows, they pressed the B button to confirm and finish the Placement phase.

During the Placement phase, to drag-and-drop any window, the user had to first grab a window by positioning either controller within 10 cm of the window's plane and within 30 cm of the window's center. The user then pressed and held the Trigger button, and moved the controller to drag the window (whose position and orientation followed the controller's with a 1:1 mapping), and completed the operation by releasing the Trigger button. Drag-and-drop operations on different windows could be done with both hands simultaneously if the user desired.

The end of the Placement phase would start one of the two Recall phases (either Visible Recall or Blind Recall). For each Recall phase, the system randomly chose 20 windows from the set of 60 that had been positioned by the user. We will refer to these 20 windows as $w_1, w_2, \ldots, w_{20}$. The system would then display a copy of w_1 at the same initial position, and the user had to remember and/or find where they had positioned the original w_1. The user would then use raycast selection to indicate the original w_1. Next, the system would display a copy of w_2 at the initial position, and the user had to raycast-select the original w_2, etc., until the user finally raycast-selected w_{20}. The system would then choose another random 20 windows for the 2nd and final Recall phase, which required the same raycast-selection of windows, one-by-one.

During both Recall phases, raycast selection could be performed with either controller and was done with the Trigger button. The controllers were shown projecting rays (like laser beams) to make it clear which window(s) were being intersected by the rays and would be selected before the user pressed Trigger.

The difference between the two Recall phases was that, in Blind Recall, the content of the 60 windows that had been positioned was hidden (i.e., they were shown as blackened rectangles). The only window whose content was visible was that of the *copy* of w_i at the initial position. This forced the user to *use their memory* to identify where they had positioned the original w_i. On the other hand, for Visible Recall, the content of all windows was visible, which is more realistic as a task, and also allowed the user to *use visual search and recognition* to find w_i rather than relying purely on their memory.

In the Blind Recall phase, the user had only one chance to identify each of the 20 windows, and in the case of an error, the system recorded the EuclideanErrorDistance between the centers of the selected window and the correct window. We also computed, for each window w_i, the maximum distance between w_i and all other windows, and then computed the average of these maxima over all the windows to obtain an overall measure of the distance between all windows. For each window that was blind-recalled, we computed a NormalizedErrorDistance equal to the EuclideanErrorDistance divided by that overall measure.

In the Visible Recall phase, if the user selected a wrong window, the user was required to try again until they successfully selected the correct window. (In practice, users never made more than 1 such mistake for each w_i.) This yielded an ErrorCount. The system also recorded the RecallTime to select the correct window.

The subsets of 20 windows chosen by the system for each Recall phase were guaranteed to be disjoint, so the user would not be asked to identify the same window in both Recall phases.

Throughout all phases, textual instructions prompted the user on what to do, step-by-step.

3.3 The Two Main Conditions

The independent variable Movement has two levels (i.e., two conditions): *Stationary*, and *Locomotion*. In the *Stationary* condition, users had to stand on a single spot during the Placement phase. This was ensured by requiring that the projected position of the user's head remain within a virtual circle (30 cm radius) displayed on the floor. Moving outside this circle prevented the user from dragging-and-dropping windows The volume available for placing windows was therefore limited by the length of the user's arms.

In the *Locomotion* condition, there was a larger virtual circle (of 140 cm radius) displayed on the floor, and each window had to be positioned *outside* of this circle, i.e., such that the projected center of the window was outside. During the Placement phase, the initial position for each new window was inside the circle, and the border of the window appeared red. This border would remain red (Figure 1) until the user had dragged-and-dropped the window outside the circle, causing the red border to disappear, indicating that the position was now acceptable. The 140 cm radius meant that users often took one or two steps with their legs to position each window, and made use of a larger volume overall.

3.4 Content and Size of Windows

To generate realistic content for the windows, we took a list of the 500 most popular websites [21] and wrote a script to download, render, and save a screenshot of each website. We manually inspected the resulting screenshots, eliminating any that had not rendered properly, leaving 341 distinct screenshots at 900×720 pixels.

For each of the two main conditions, a random subset of 60 screenshots was chosen by the system, and these were different for each user. Windows were displayed in VR with a diagonal size of approximately 15 inches, roughly the size of a laptop.

3.5 Hypothesis

Our hypothesis was that the Locomotion condition, which forced users to move their body more and occupy more volume with the windows, would benefit the spatial memory of users, yielding improved RecallTime or ErrorCount in the Visible Recall phase, and an improvement in EuclideanErrorDistance or NormalizedErrorDistance in the Blind Recall phase. This hypothesis is motivated by previous work that has found gestures and body motion benefiting memory or spatial memory [4, 13, 14, 18, 24].

3.6 Experiment

The Placement phase was followed by both Blind Recall and Visible Recall phases. The ordering of Movement conditions was counterbalanced, with half of the users experiencing *Stationary* first, and the other half experiencing *Locomotion* first. In addition, an orthogonal half of the users experienced the Blind Recall phase first, while the other half experienced Visible Recall first.

Precautions due to the covid-19 pandemic included disinfecting surfaces, equipment, and hands; masks and visors; and social distancing. Each user had their depth acuity measured with a Titmus stereo acuity test, filled out pre- and post-questionnaires, and performed tasks with the headset adjusted to their inter-pupillary distance (IPD). Prior to donning the headset, a video demonstrated the tasks to the user, to make instructions consistent. A break was taken between the two conditions. Each session with one user took ≈60-90 minutes, not including time to disinfect.

We recruited 18 users for our university, 16 male, 2 female; aged 20-36 years (mean 25.6); 16 right-handed, 2 left-handed; none self-reported color-blind; IPD ranging from 58 to 70mm (within the normal range [6]); stereo acuity measured with Titmus test ranging from 0 to 10 out of 10 (mean 7.8). Of these 18 users, one (user #9) was removed from our analysis because their post-questionnaire was incompletely filled out, and another (user #18) was removed so that the remaining 16 had a perfectly counter-balanced ordering of the Movement variable.

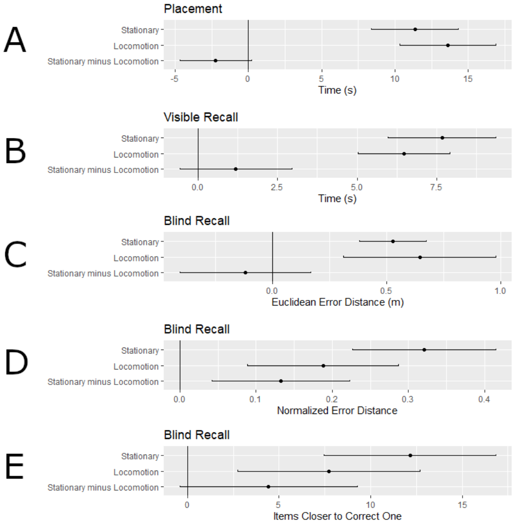

Figure 2: 95% confidence intervals.

The data that remained for analysis comprised 16 participants × 2 conditions (*Stationary*, *Locomotion*) × 60 windows = 1920 windows positioned, and 16 participants × 2 conditions (*Stationary*, *Locomotion*) × 40 windows recalled (20 blind + 20 visible) = 1280 windows recalled.

3.7 Results

We adopt several pieces of advice from Dragicevic [7] that are informed by recent trends in statistics and that have been discussed extensively in the statistical literature. In particular, we eschew null hypothesis significance testing (NHST) and p-values as these support misleading, dichotomous thinking ("Tip 25" in [7]); we present effect sizes visually and with confidence intervals (Tips 15, 16), where the confidence intervals are computed using only one (averaged) value for each (user, condition) pair (Tip 9); we clearly distinguish between our pre-experiment hypotheses and post-hoc exploratory data analysis, because not doing so encourages HARKing (Hypothesising After the Results are Known) and p-hacking; and we are deliberately vague in our interpretation of results (Tip 26) to allow readers to draw their own conclusion.

In Figure 2, A and B, the position of the "zero" with respect to the difference CIs indicates that users spend more time positioning windows in the *Locomotion* condition (unsurprisingly, since they must move their legs), but seem to spend less time retrieving them in the *Locomotion* condition (during Visible Recall). This is despite the fact that raycast pointing at windows in the *Locomotion* condition requires pointing at targets that are farther away, and therefore that cover a smaller angle, and hence should require more time according to Fitts' law [23]. Thus, Figure 2, B provides some evidence that *Locomotion* can improve performance.

The ErrorCount in the Visible Recall task was always below 3%, with never more than one wrong selection per trial.

Figure 2, C and D, indicate that *Locomotion* results in a larger variance in EuclideanErrorDistance and in an improvement in NormalizedErrorDistance. We later realized that neither of these is the best way to quantify the error in Blind Recall condition. The windows were the same size in both Movement conditions, and we observed that users often placed neighboring windows close to each other, regardless of the Movement condition (compare Figure 1, D and F). Thus, if a user committed an error where they were "off by one", selecting the neighbor of a target rather than the target, this could plausibly result in the same EuclideanErrorDistance, and therefore a smaller normalized distance in the *Locomotion* condition. In the next section, we present a new way to quantify error distance.

Subjective likert ratings suggest users found *Locomotion* required more physical effort (3.6/7 for *Locomotion* vs 2.8/7 for *Stationary*), but *Stationary* required more mental effort (3.5 vs 4.8) and was more frustrating (2.1 vs 3.6). Subjective results are further detailed in the supplemental document.

3.8 Post-hoc Exploratory Data Analysis

Due to the problems inherent in EuclideanErrorDistance and NormalizedErrorDistance, we computed (post-hoc) a new dependent variable for the Blind Recall trials, the ItemsCloserToCorrectOne, i.e., the number of windows closer to the correct target window than the selected window. A selection that is "off by one" is quantified by this metric the same way regardless of Movement condition, making it a better measure of small errors. Figure 2, E provides good evidence that *Locomotion* resulted in smaller errors. This may be due to the locomotion engaging spatial memory, and/or the larger volume allowing for more spacious layouts that better trigger memory (compare Figure 1, D and F).

We also separated our data into two halves. The performance of users during the 2nd half of each session may be more reflective of realistic performance, since by the 2nd half, users have better understood the task. We found that average RecallTime and average error distances are all smaller in the 2nd half of the experiment. We also found that RecallTime in the 2nd half is 6.9-5.4=1.5 seconds (or 22%) lower for *Locomotion* than for *Stationary*. However, in the 2nd half of the experiment, users spent 14.1-10.7=3.4 seconds more in initial placement of windows in *Locomotion* than in *Stationary*. In a real scenario, as the user performs repeated retrievals to navigate between windows, spatial memory should consolidate and recall time in both conditions should converge. The **implication for designers** is that the benefit to spatial memory from increased locomotion is modest. Hence, if users want to position windows within a large volume, it seems likely that doing so through indirect input (e.g., [16]), with little locomotion (and without any need for a wide open physical space for walking), will not substantially harm later retrieval performance.

4 CONCLUSION AND FUTURE WORK

We found some evidence (Figure 2, B and E) for a modest benefit to retrieval performance when users had to use locomotion to place windows. Future work could include helping users manage multiple, large collections of windows or documents with 3D analogs of piles [1, 19] or fences (https://www.stardock.com/products/fences/). Such features would help users organize and find content, and agree with some of the UI design guidelines in [29].

REFERENCES

[1] Anand Agarawala and Ravin Balakrishnan. 2006. Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In *ACM CHI*.

[2] Andy Cockburn, Per Ola Kristensson, Jason Alexander, and Shumin Zhai. 2007. Hard lessons: effort-inducing interfaces benefit spatial learning. In *ACM CHI*. 1571–1580.

[3] Andy Cockburn and Bruce McKenzie. 2002. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *ACM CHI*. 203–210.

[4] Susan Wagner Cook, Terina KuangYi Yip, and Susan Goldin-Meadow. 2010. Gesturing makes memories that last. *J. Memory and Language* 63, 4 (2010), 465–475.

[5] Mary Czerwinski, Maarten Van Dantzich, George G Robertson, and Hunter G Hoffman. 1999. The Contribution of Thumbnail Image, Mouse-over Text and Spatial Location Memory to Web Page Retrieval in 3D. In *INTERACT*.

[6] Neil A Dodgson. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic displays and virtual reality systems XI*, Vol. 5291. International Society for Optics and Photonics, 36–46.

[7] Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330.

[8] Barrett M Ens, Rory Finnegan, and Pourang P Irani. 2014. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *ACM CHI*.

[9] 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 *IEEE BigComp*.

[10] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. 2018. The Office of the Future: Virtual, Portable, and Global. *IEEE CG&A* (2018).

[11] Mikkel R Jakobsen and Kasper Hornbæk. 2015. Is moving improving?: Some effects of locomotion in wall-display interaction. In *ACM CHI*.

[12] Yvonne Jansen, Jonas Schjerlund, and Kasper Hornbæk. 2019. Effects of Locomotion and Visual Overview on Spatial Memory when Interacting with Wall Displays. In *ACM CHI*. 1–12.

[13] 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 *AVI*. 83–90.

[14] Nathaniel B Klooster, Susan W Cook, Ergun Y Uc, and Melissa C Duff. 2015. Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Frontiers in Human Neuroscience* 8 (2015).

[15] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. 2019. Virtual memory palaces: immersion aids recall. *Virtual Reality* 23 (2019), 1–15.

[16] Joon Hyub Lee, Sang-Gyun An, Yongkwan Kim, and Seok-Hyung Bae. 2018. Projective Windows: Bringing Windows in Space to the Fingertip. In *ACM CHI*.

[17] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E Mackay. 2014. Effects of display size and navigation type on a classification task. In *ACM CHI*. 4147–4156.

[18] Christopher R Madan and Anthony Singhal. 2012. Using actions to enhance memory: effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology* 3 (2012), 507.

[19] Richard Mander, Gitta Salomon, and Yin Yin Wong. 1992. A 'pile' metaphor for supporting casual organization of information. In *ACM CHI*.

[20] Mark Mcgill, Aidan Kehoe, Euan Freeman, and Stephen Brewster. 2020. Expanding the bounds of seated virtual workspaces. *ACM TOCHI* (2020).

[21] Moz, Inc. 2020. The Moz Top 500 Websites. https://moz.com/top500.

[22] Oculus. 2020. Infinite Office. https://youtu.be/5_bVkbG1ZCo.

[23] Julian Petford, Miguel A Nacenta, and Carl Gutwin. 2018. Pointing all around you: selection performance of mouse and ray-cast pointing in full-coverage displays. In *ACM CHI*. 1–14.

[24] 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 *Proc. ACM ITS*.

[25] Meredith Ringel. 2003. When one isn't enough: an analysis of virtual desktop usage strategies and their implications for design. In *Extended Abstracts of ACM CHI*.

[26] George Robertson, Mary Czerwinski, Kevin Larson, Daniel C Robbins, David Thiel, and Maarten Van Dantzich. 1998. Data mountain: using spatial memory for document management. In *ACM UIST*. 153–162.

[27] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. 2020. Maps around me: 3D multiview layouts in immersive spaces. *Proc. ACM on HCI* (2020).

[28] William Saunders and Daniel Vogel. 2016. Tap-kick-click: Foot interaction for a standing desk. In *ACM DIS*.

[29] 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.

[30] Desney S Tan, Darren Gergle, Peter Scupelli, and Randy Pausch. 2006. Physically large displays improve performance on spatial tasks. *ACM TOCHI* (2006).

[31] Md Sami Uddin, Carl Gutwin, and Andy Cockburn. 2017. The effects of artificial landmarks on learning and performance in spatial-memory interfaces. In *ACM CHI*. 3843–3855.

[32] Fumeng Yang, Jing Qian, Johannes Novotny, David Badre, Cullen Jackson, and David Laidlaw. 2020. A Virtual Reality Memory Palace Variant Aids Knowledge Retrieval from Scholarly Articles. *IEEE TVCG* (2020).