



Bring2Me: Bringing Virtual Widgets Back to the User's Field of View in Mixed Reality

Charles Bailly, François Leitner, Laurence Nigay

► To cite this version:

Charles Bailly, François Leitner, Laurence Nigay. Bring2Me: Bringing Virtual Widgets Back to the User's Field of View in Mixed Reality. AVI '20: International Conference on Advanced Visual Interfaces, Sep 2020, Ischia Island, Italy. pp.1-9, <10.1145/3399715.3399842>. <hal-02960599>

HAL Id: hal-02960599

<https://hal.science/hal-02960599v1>

Submitted on 7 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Bring2Me: Bringing Virtual Widgets Back to the User's Field of View in Mixed Reality

Charles Bailly^{1,2}

¹Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
F-38000 Grenoble, France
charles.bailly@univ-grenoble-
alpes.fr

François Leitner²

²Aesculap SAS
F-38130 Echirolles, France
francois.leitner@bbraun.com

Laurence Nigay¹

¹Univ. Grenoble Alpes, CNRS,
Grenoble INP, LIG
F-38000 Grenoble, France
laurence.nigay@univ-grenoble-
alpes.fr

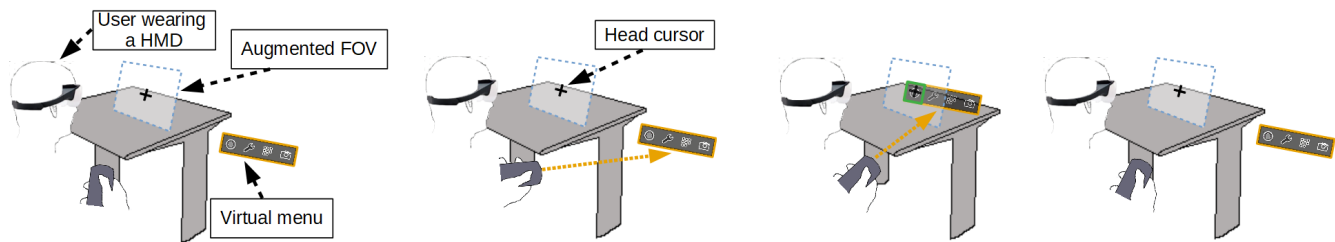


Figure 1: Bringing back a virtual menu to the FOV without turning the head. The user: 1) points to the off-screen menu using raycasting, 2) moves the selected menu with a drag and drop gesture and 3) aligns an item with the head cursor to select it.

ABSTRACT

Current Mixed Reality (MR) Head-Mounted Displays (HMDs) offer a limited Field Of View (FOV) of the mixed environment. Turning the head is thus necessary to visually perceive the virtual objects that are placed within the real world. However, turning the head also means losing the initial visual context. This limitation is critical in contexts like augmented surgery where surgeons need to visually focus on the operative field. To address this limitation we propose to bring virtual objects/widgets back to the users' FOV instead of forcing the users to turn their head. We carry an initial investigation to demonstrate the approach by designing and evaluating three new menu techniques to first bring the menu back to the users' FOV before selecting an item. Results show that our three menu techniques are 1.5s faster on average than the baseline head-motion menu technique and are largely preferred by participants.

CCS CONCEPTS

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

KEYWORDS

Mixed Reality, HMD, Pointing task

ACM Reference Format:

Charles Bailly, François Leitner and Laurence Nigay. 2020. Bring2Me: Bringing Virtual Widgets Back to the User's Field of View in Mixed Reality. In

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

AVI '20, September 28-October 2, 2020, Salerno, Italy

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7535-1/20/09...\$15.00

<https://doi.org/10.1145/3399715.3399842>

International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy. ACM, New York, NY, USA, 9 pages.
<https://doi.org/10.1145/3399715.3399842>

1 INTRODUCTION

Current Head-Mounted Displays (HMD) for Mixed Reality (MR) still offer a limited Field Of View (FOV) far from covering the human eyes capabilities [8, 13, 24]. For instance, the recent HoloLens2 [21] has a FOV of 43° horizontally and 29° vertically, which corresponds to only a part of the mid-peripheral vision [27]. Patterson et al. suggest that such limited FOVs impact spatial orientation and immersion [24]. While several interaction techniques have been proposed to select distant visible objects in MR [18, 23], a size-limited FOV implies that most of the mixed environment is not visible. Turning the head to move the augmented FOV is thus necessary to make targeted objects visible.

Such a scenario can be found in many contexts, including augmented surgery. As explained by Bailly et al., surgeons are used to interact with menus during an operation [5]. By wearing a HMD, surgeons no longer rely on a single physical screen in the operating room: virtual menus can be located across the whole operating theatre. However, turning the head to look at these menus implies losing the initial visual context. This is a strong issue when users need to focus their visual attention on their main task. Surgeons looking away from the patient have to interrupt their surgical task, which adds cognitive load to already busy-minded users [5]. Alternative solutions to interact with virtual widgets that preserve users' visual context are required. In this work, we focus on a commonly used type of widget: menus.

On the one hand, eyes-free pointing techniques allow users to directly interact with virtual items without looking at them [20, 28]. Such techniques rely on spatial memory and proprioception, but do

not consider bringing the menu back to the FOV to make it visible. On the other hand, other studies propose gesture-based interaction techniques to invoke a menu and make it appear in the FOV [10, 16]. Such an approach requires learning the set of gestures and poses associated with the different available menus. In this case, spatial memory cannot be used to facilitate the interaction.

Our approach is a combination of eyes-free pointing and menu invocation. We take advantage of the spatial memory to point at menus in an eyes-free manner in order to make these menus become visible in users' FOV. Moreover, we study menu interaction by focusing on how to bring the menu back to the FOV and then select a menu item.

The contribution of this work is the exploration of a new approach to compensate the limited FOV of current HMDs. We present the design and evaluation of 3 menu techniques. These techniques preserve the visual context of the user, benefit from the spatial memory and make the targeted menu temporarily appear in the users' FOV for the time of the interaction. We evaluate these 3 techniques with an experimental study where they are compared to a baseline head-motion technique. In this work, we take augmented surgery as an application context, but the lab study was conducted without ecological conditions to be able to generalise the results. After discussing experimental results, we conclude by describing future pathways for interacting with virtual widgets in MR with a size-limited FOV.

2 RELATED WORK

2.1 Selecting visible targets in MR

There are two main approaches to select visible targets in a 3D space: raycasting and virtual hand [3]. Raycasting techniques have been deeply studied in the literature [13, 17], and previous works found it was efficient for spatial object selection [23]. Contrary to virtual hand, raycasting techniques allow the selection of targets beyond the area of reach of the user's arms. In the case of distant targets, several strategies like target expansion can be applied to facilitate the selection [15].

Using a hand-held tool as the origin of the ray is fairly common [6, 13]. Since MR with a HMD implies that the augmented FOV is directly linked to the position of the head, several works also explored head-based raycasting techniques [5, 18]. In these studies, the position of the cursor is directly linked to the position of the head. Özacar et al. compared this approach with the GyroWand technique, where the ray origin is fixed on the user's chin and its direction is controlled using the IMU of a hand-held device [23]. The authors found that the head cursor outperformed GyroWand in terms of speed. Lee et al. suggests that hand-based raycasting outperforms head-based raycasting in terms of task completion time and accuracy, but the latter has the advantage of not requiring users' hands at the moment of the interaction. Nonetheless, both head and hand raycasting technique are eyes-engaged and require one to see the target.

2.2 Selecting off-screen targets in MR

The limited size of the FOV on current HMDs implies that most virtual targets located across the environment are not visible at a

given time. In their work on direct off-screen pointing with a mobile device, Ens et al. reported three strategies to deal with off-screen targets [12]:

- **S1:** Scaling the workspace to make it encloses the targets visible on screen (zooming)
- **S2:** Shifting the workspace to make targets appear in the viewport (scrolling)
- **S3:** Moving the viewport to the targets (peephole pointing).

However, adapting these strategies to MR with a HMD raises several issues as the viewport (the augmented FOV) is egocentric and targets are located across the whole 3D environment. Scaling the workspace (S1) would distort the 3D augmented environment and make targets too small: it is thus not adapted. While S2 is usable in MR, shifting the entire virtual workspace would move all the virtual objects present in the environment, which is not suitable. Finally, moving the viewport to the targets (S3) can be done in MR with a HMD by turning the head to the target locations. Nonetheless, this approach implies losing the initial visual context.

2.3 Preserving the visual context

Users executing visually demanding tasks cannot afford to turn the head to look at virtual targets. To address this issue of initial visual context, previous work mainly studied two approaches: eyes-free pointing and menu invocation.

2.3.1 Eyes-free pointing. A few studies focused on eyes-free pointing in MR. Li et al. proposed a system to trigger shortcuts in a eyes-free manner by orienting a spatially-aware mobile device towards virtual application icons placed around the user [20]. Selections rely on kinesthetic and spatial memory, and their system outperformed the classical interface of a Nokia device for common phone tasks.

More recently, Yan et al. studied eyes-free target acquisition around the body in VR [28]. The authors conducted several studies to evaluate the acceptance of target location, the accuracy of control of eye-free target acquisition and finally compared eye-free selection with eye-engaged selection. Eyes-free selections were found to be faster and more error prone than eyes-engaged selections while still reaching a satisfying level of accuracy [28]. Moreover, most of their participants preferred the eyes-free technique. Such results are encouraging for eyes-free interaction techniques, especially in the case of MR where the FOV is reduced (around 50° diagonally at best) compared to current VR FOVs (between 90° and 110° on currently available VR HMDs). Besides, eyes-free pointing benefits from the spatial memory since a virtual target has a physical position in the environment.

2.3.2 Invoked menus. Another possibility to compensate a reduced FOV is to rely on *invoked menus*. These menus do not have permanent physical positions in the environment: they appear directly in users' FOV when invoked, and disappear after the selection of an item or when users no longer need to interact with them. Contrary to *persistent* menus that have a continuous position in the environment, no spatial memory is involved when using temporary menus. For instance, Datcu et al. proposed a system where a menu appears at the non-dominant hand location, letting users select items by pointing them with their dominant hand [10]. In their work, the authors considered only one menu, but different hand

gestures could be used to make the different available menus appear. The main downside of this approach is that users must learn the gestures and the mapping of gestures with menus. Such interaction cannot take advantage of the spatial memory as the menus do not have a specific position in the environment.

2.4 Positioning of our work

Our work aims at combining the features and benefits of eyes-free pointing and menu invocation. We explore an approach relying on the spatial memory and eyes-free pointing to bring a virtual menu back to the users' FOV. This approach is closely related to the "Shifting the workspace" strategy (S2) identified by Ens et al. to manage off-screen targets with a mobile device [13]. In our case, we detach temporarily the menu from its initial location for the time of the interaction. Only the selected menu is affected instead of the whole workspace. Besides, we also study different techniques to select an item once the menu became visible. In the following section, we detail this approach and the design of the corresponding menu techniques.

3 INTERACTION TECHNIQUES

Instead of forcing users to turn the head to see the menu they want to interact with, we explore bringing the menu to users' FOV. In that case, the initial visual context is preserved: only the target menu is temporarily moved for the interaction duration (pseudo-mode) before going back to its initial location. To explore this approach, we propose three new *Bring2Me* interaction techniques to interact with menus in MR. We consider the entire interaction workflow, which includes 1) selecting a particular menu 2) moving the selected menu to users' FOV and 3) selecting an item of the selected menu that is now visible. We base the design of our techniques on 3 design parameters:

- (1) The technique to select a given menu among the menus placed in the environment
- (2) The technique to move the menu
- (3) The technique to finally select the menu item.

3.1 Design parameters

3.1.1 Initial menu selection. In order to allow users to bring a given menu to their FOV, we rely on spatial memory while preserving the users' visual context. Inspired by previous work on bimanual interaction using a stylus [9] and eyes-free pointing [28], we introduce another input modality based on a hand-held tool. An example of such a tool can be found in augmented surgery, where surgeons use hand-held tools tracked by a camera during operations [1]. In our case, we use the hand-held pad of the HMD as a ray-casting device: users can point at the location of a virtual menu without turning the head (and thus, without seeing the menu). The menu selection is validated using the pad hand-held button. An illustration of this initial menu selection is available at Figure 1.

3.1.2 Technique to control the menu position. We distinguish two types of manipulation over the menu position: *discrete* control and *continuous* control.

With discrete control, the selected menu will instantly jump to users' FOV. Previous work suggest that the loss of control during

the interaction can impact user performance and comfort by creating disorientation issues [26]. For instance, cursor movements predictions systems like the Delphian desktop [4] can sometimes be inaccurate or wrong, perturbing the initial trajectory planned by users. However, in our case the behavior of the menu is deterministic as it will always appear at the top of the FOV. We hypothesize that such a behavior would reduce the disorientation of users.

Continuous control of the menu location implies users freely controlling the position of the menu over time once the menu is selected. In this case, users have to move the menu to their FOV (mid-air dragging).

3.1.3 Item selection. Finally, our last design parameter concerns the technique to select the desired item when the menu is visible in the FOV. We focus on three techniques: 1) head only using a head cursor 2) hand only using ray-casting with the hand-held pad and 3) a mixed approach (head + hand).

3.2 Resulting interaction techniques

The three *Bring2Me* techniques we propose are based on a combination of the design parameters. The three designed techniques share the same first design parameter: only the two other design parameters vary. The three *Bring2Me* techniques are depicted in Figure 2.

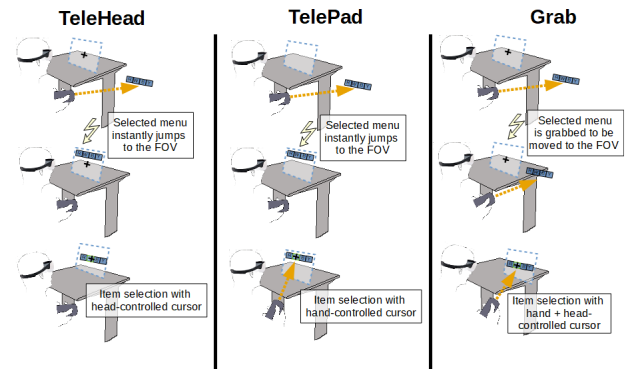


Figure 2: The three *Bring2Me* techniques.

The *TeleHead* technique is based on discrete control of the selected menu and on a head cursor to select the desired menu item. Its name comes from this combination of parameters: the menu *teleportation* (instant menu jump) and the item selection technique (*head* cursor). Similarly, the *TelePad* technique also relies on discrete control of the menu position, but it uses ray-casting with the HMD hand-held pad to select the menu item. Therefore, all the selection is done using only the hand in this case.

The *Grab* technique is a mixed approach based on a drag-and-drop gesture. Like the *TeleHead* and *TelePad* techniques, users must first select the desired menu by doing off-screen ray-casting. However, in this case they must keep the pad button pressed instead of simply clicking on it. While the button is pressed, the menu is linked to the pad ray-cursor and users freely control its position (continuous control). Once the menu is displayed in the FOV and

Technique	Design parameters			Properties	
	Menu selection	Menu position control	Item selection	Visual context	Hands-free
TeleHead	Pad raycasting	Discrete (jump to FOV)	Head cursor	Preserved	No
TelePad	Pad raycasting	Discrete (jump to FOV)	Pad ray cursor	Preserved	No
Grab	Pad raycasting	Continuous	Head cursor + hand	Preserved	No
HeadPointing	None	None	Head cursor	Lost	Yes

Table 1: Comparison of the 3 designed interaction techniques compared to the baseline HeadPointing technique.

the head cursor is hovering the desired menu item, the pad button can be released to confirm the selection.

We eliminated the technique based on continuous control of the menu position and a head-controlled cursor to select the menu item. Such a change of input modality would break the single drag-and-drop gesture metaphor we wanted to study. Contrary to *TeleHead* and *TelePad* based on two distinct pointing steps (selecting the menu, then the item), *Grab* is therefore based on a continuous gesture to select and move the menu as well as select the item. Besides, the two modalities (hand and head movement) can be used in parallel with this technique.

For the experimental comparison of the three techniques, we first considered a standard hand raycasting technique as a baseline. Nevertheless, such hand raycasting approach is not eyes-free and requires users to have a hand available to interact. We considered that if a hand is available at the moment of the interaction, it should be used to perform eyes-free pointing and preserve the visual context. Therefore, we chose a standard head-controlled cursor technique [5, 18, 23] which is eyes-engaged and does not require gestures with the hand. A cross-shaped cursor is displayed at a fixed position at the center of the HMD display, and has to be aligned with the target by turning the head. When correctly placed, the target item is highlighted and the selection can be confirmed. The four compared techniques are described in Table 1.

Initially, we planned to use a two-pedal footswitch for the confirmation of the selection, which is an input device used by surgeons. However, as part of a pilot study we observed that some participants were disturbed by the latency when confirming the selection with the pedals (around 1s to obtain a visual feedback in worst case scenarios). We thus removed the pedal for this experiment and let participants confirm the selection of targets by pressing a button on the HMD pad held by the dominant hand. Therefore, the validity of the *HeadPointing* technique is restricted to pointing tasks (thus excluding the confirmation mechanism).

4 DESIGN AND IMPLEMENTATION

We used the Epson Moverio BT-300, running on Android 5.1. This HMD has a 23 degree vertical FOV and a resolution of 1280*720 pixels, and has the advantage of being light-weight. Three passive markers were fixed on it in order to track its position with a set of five Flex3 cameras. These cameras were fixed to the ceiling and monitored by the *Tracking Tools* software (version 2.3.1) running on a distant laptop. A Python script sends the computed head position to the HMD through Wifi. Then, our custom software running on the HMD uses the position of the head (location and orientation) to display the corresponding view. This software is based on OpenGL

ES 2.0 to display virtual content with side-by-side 3D. Overall, the implemented system is able to maintain 30 fps. An average of 24ms was required to obtain the head pose estimate: 10ms from cameras, 4ms for pose computations and 10ms to send the pose to the HMD through Wifi.

For menus, we considered only 1-level menus with 4 items. This design choice was motivated by the similarity with menus in existing surgical navigation systems like the OrthoPilot [1]. Virtual menus are 12cm long, 3cm high and their items are labelled with letters.

5 EXPERIMENTAL STUDY

5.1 Setup

5.1.1 Workspace. As shown in Figure 3a, participants stood behind a grey squared table (81cm long, 73cm high) representing a working area. A white panel (102cm long, 125cm high) was placed vertically behind the table and elevated from the floor (+35cm). Near the top of this white panel, a 30cm long green square called *Visual Context* was displayed in the HMD (Figure 3b). A physical object (here, a plastic bone) materialises the center of this area used as the main visual focus during the experiment. Our goal when positioning the *Visual Context* was to allow participants to look in front of them instead of keeping the head turned down as this could create additional neck tiredness.

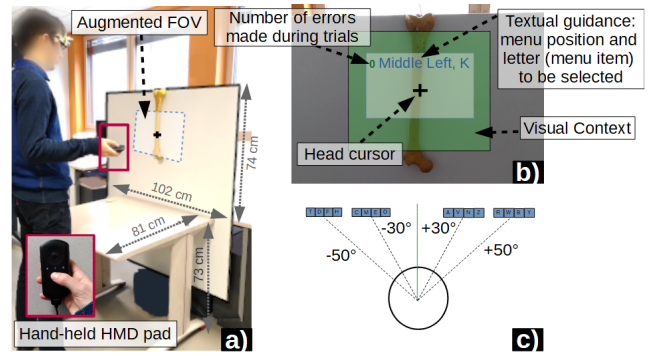


Figure 3: Experimental setup. a) A participant standing in front of the experimental workspace. b) Illustration of the FOV at the beginning of a pointing task c) Top view of the angular distances (horizontal angles) between the initial FOV center and the 4 virtual menus.

5.1.2 Menu positions. Two criteria guided the positioning of the virtual menus. First, we wanted to force users to completely change their initial visual context when turning their head with the *Head-Pointing* technique. However, we also wanted to stay in a range of angular distances with acceptable eyes-free pointing accuracy, as studied by Yan et al. [28]. They observed that the best accuracy results were concentrated in a range of horizontal angles from -60 to +60 degrees. We thus chose 30 and 50 degrees for the two possible angular distances between the center of the Visual Context and the menus. These angular distances are included in the mid-peripheral vision. Moreover, between 30 and 60 degrees (horizontally) the vision is still binocular and colors can be efficiently detected [27].

We thus obtained four horizontal positions for menus: -50, -30, +30 and +50 degrees, as shown in Figure 3c. The menus were all vertically aligned above the white panel at a angular distance close to 20 degrees from the center of the Visual Context. The four menus were designated to participants by their location: Left, Middle Left, Middle right and Right respectively. To facilitate the off-screen pointing phase, the menu targets were expanded using invisible 80cm long hitbox areas. The Visual Context was the only virtual area visible on the central white panel.

5.2 Task

To start a trial, participants had to look and select the Visual Context at the center of the setup using ray-casting with the pad. The head cursor and the pad cursor were both visible and had to be inside the limits of the Visual Context before clicking on the pad button. Our goal was to make sure participants started in neutral head and hand positions. For *Bring2Me* techniques, participants were also told to keep looking at the Visual Context area.

Inspired by previous work on pointing task in MR [5], a textual guidance indicated the current menu and item to be selected at the beginning of a trial (e.g. "Middle Left, K" in Figure 3b). This visual textual feedback is displayed at the top of the HMD display.

Participants had to turn their head or bring the targeted menu to their FOV depending on the current interaction technique. The targeted menu item was highlighted in green when hovered by the head cursor or the pad cursor. After a successful item selection, participants had to go back to the initial position at the center of the Visual Context before starting another trial. In case of error, the trial was interrupted and rescheduled at the end of the block. The current number of errors made was always displayed at the top left corner of the display. Green at the beginning (no error made), this error number became orange then red when too many errors were done (red for more than 3 errors). Participants were instructed to be as quick as possible while avoiding errors and to adjust their speed if the error number became red.

5.3 Study design

The experiment was divided into 4 sections, one section per interaction technique: *HeadPointing*, *TeleHead*, *TelePad* and *Grab*. Each section was divided in two blocks of 16 trials, leading to 32 pointing tasks per technique. During these 32 pointing tasks, each of the 16 menu items was selected twice, as shown in Figure 4. Overall, we had a total of 128 pointing tasks per participant during the experiment. The order of sections was balanced using a Latin-square design

while the order of menu items was randomized. After each section, participants were invited to take a short break if they wanted, and then had to complete a Raw-TLX questionnaire to collect their feedback about the technique they just used.

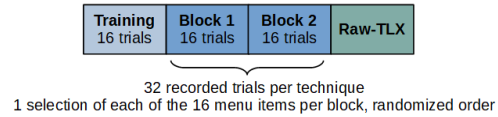


Figure 4: Experimental workflow for each of the 4 interaction techniques.

The experiment began with the explanations of the instructor about the pointing tasks and the position of virtual objects in the environment illustrated through pictures. After adjusting the HMD and the stereo effect if necessary, participants were invited to look at each virtual area (the Visual Context and the 4 menus) and asked to memorise their positions. Participants were told that the locations of menus would remain the same during the entire experiment. They were free to explore the mixed environment until they felt confident with it. Our goal was to be comparable to a peephole pointing scenario with prior knowledge of target locations [17]. Before each technique session, a training block of 16 trials was imposed. The goal of this long training phase was to make sure participants were used to the interaction technique and the menu positions.

5.4 Participants

12 volunteer participants (5 female, 7 male) from 23 to 43 years old ($sd = 5.3$) were recruited for this study. As reported in our preliminary questionnaire, they were all novices at interacting in MR and only 2 of them had already tried MR before. Participants could wear glasses or contact lenses with the HMD.

5.5 Measures

We recorded the time of each phase of the pointing task. For the *HeadPointing* technique, we recorded the time to turn the head to the menu until it became visible (head search time) and the time to select the menu item (item time). For the other techniques, we recorded the time to select the menu (menu ray-casting time) and the time to select the item. Three types of errors were also recorded: 1) focus error (looking away from the Visual Context instead of focusing on it); 2) menu selection error (incorrect off-screen menu selection) and 3) incorrect menu item selection. Head and hand trajectories were also recorded every 100ms to have a precise understanding of participants behaviour during pointing tasks.

For the qualitative analysis, we used Raw-TLX questionnaires based on a 5-items Likert scale. Subjective preferences about the interaction techniques and their ranking were discussed during semi-structured interviews.

5.6 Results

To conduct statistical analysis, we used Repeated-measures ANOVA (with $\alpha = 0.05$). Post hoc tests were conducted using pairwise t-tests

and Bonferroni corrections. When using ANOVA was not appropriate, we employed a Friedman test instead followed by a Wilcoxon post hoc test with Bonferroni corrections. The results of these statistical tests are reported in the following. However, as suggested by Dragicevic [11], we put more emphasis on confidence intervals and effect sizes. In all figures, error bars are 95% Confidence Intervals (CI). In case of positive skewness in time measurements, we applied a log transform to the data before analysis [11].

5.6.1 Selection times. We observed a clear difference of total selection times between the 4 techniques. Turning the head with the baseline *HeadPointing* techniques seems 1.5s longer on average compared to *TeleHead* ($p = 2.6e-05$), *TelePad* ($p = 0.0015$) and *Grab* ($p = 0.0007$). On the contrary, we did not find any evidence of difference between the three techniques we proposed as total selection times were close to 4.3s for each.

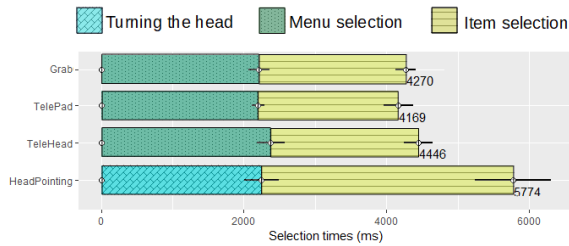


Figure 5: Mean selection times per interaction technique. Error bars are 95% CI.

As illustrated in Figure 5, the times to reach the menu position by turning the head or using off-screen pointing were also similar. However, we found good evidence of difference for the item selection time between *HeadPointing* and the other techniques. Overall, we did not find any evidence of an influence of menu positions over selection times ($p = 0.98$).

In order to detect any learning effect in the data, we compared the two blocks of trials for each technique (16 first trials versus the 16 last ones). Overall, we only found a good evidence of difference for *TeleHead* ($p = 0.006$). When focusing on this particular technique, we observed a mild tendency suggesting the difference was mostly due to menu item selection times (comparison of the two trials blocks: $p = 0.035$).

5.6.2 Errors. We observed a large discrepancy between participants about the number of errors made during the experiment, as illustrated in Figure 6. Overall, it was difficult to draw any strong conclusion. This was supported by our statistical analysis, with which we found no clear evidence of difference between techniques ($\chi^2 = 5.74$, $p = 0.13$). However, when comparing only errors related to menu item selections, we observed that *HeadPointing* led to more error than *TeleHead* ($p = 0.025$), *TelePad* ($p = 0.02$) and *Grab* ($p = 0.012$). In particular, no error was made during the menu item selection phase with *Grab*.

Unsurprisingly, the number of errors during the menu selection with off-screen ray-casting seems to be quite similar between the techniques (all techniques but the baseline). This result was supported by our statistical analysis as no evidence of difference was

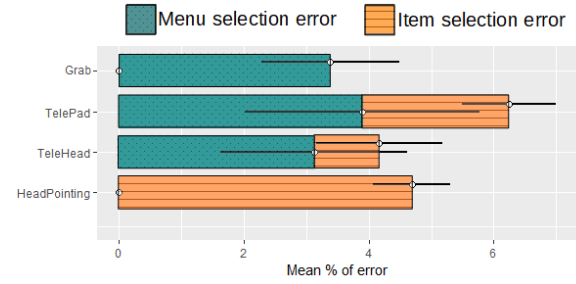


Figure 6: Mean percentage of error per interaction technique. Error bars are 95% CI.

observed ($p = 1$). Only two focus errors were made during the entire experiment, making it negligible.

5.6.3 Head and hand trajectories. With the *Grab* technique, two interaction modalities were available at the same time during the item selection phase (once the menu was grabbed). The analysis of head trajectories suggest that only 2 of the 12 participants used their head during the item selection with the *Grab* technique. For the other 10 participants, no evidence of parallel usage of the two available modalities was found. The time to start the menu item selection when the menu became visible in the FOV (*Grab*, *TelePad* and *TeleHead* techniques) was close to 0.5s.

5.6.4 Qualitative results. A clear tendency emerged from both qualitative questionnaires and semi-structured interviews: turning the head was the most disliked approach for participants, as shown in Figure 7. *HeadPointing* was judged more physically tiring and frustrating. Interestingly, participants P10 and P11 declared that they found this technique more "simple and intuitive" at first glance during the explanations with the instructor. These two participants thus expected to prefer *HeadPointing*, but reported that they changed their mind after a few training trials. All participants appreciated being able to bring a menu to their FOV: 6 rated it *useful* and the other 6 *really useful*.

When asked to compare the *Bring2Me* techniques during the semi-structured interviews, several tendencies were reported by participants. First, the instant menu jump occurring with *TeleHead* and *TelePad* was always judged better (less frustrating and less tiring) than turning the head. Only P9 reported that "it was not always obvious to know where the menu would appear exactly". Interestingly, P2 and P7 reported that the teleportation "created a break" compared to the continuous movement with *Grab*. 5 participants selected *Grab* as their favourite technique, explaining it felt easier, more intuitive and more precise. P12 also reported that he appreciated using both the head and the hand to select menu items with this technique. However, 3 participants also reported that the continuous gesture was more physically tiring than the jumping menu, especially for the most distant targets. P1, P4 and P6 also explained that their hand was more stressed keeping the pad button pressed, which was not comfortable.

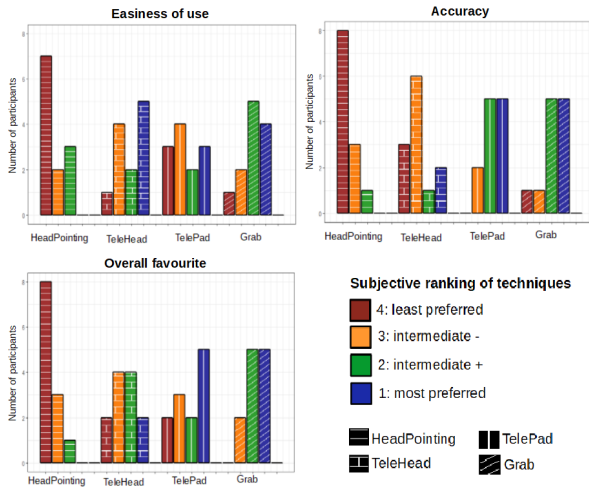


Figure 7: Results from the final qualitative questionnaire. Participants were asked to rank the 4 techniques according to criteria including easiness of use, accuracy and overall preference.

6 DISCUSSION

6.1 Turning the head

Experimental results clearly suggest that turning the head to see the menu and interact with it is the worst approach among the interaction techniques we evaluated. *HeadPointing* was the slowest and most disliked technique, even if it was the most straightforward to explain to participants. However, the times to make the menu visible (by turning the head to see it or by making it jump to the FOV) were similar. Two reasons may explain this similarity. First, participants perfectly knew the locations of the four virtual menus thanks to the long training phases and thus did not have to explore the environment to discover targets. Secondly, menu selection with Bring2Me techniques was facilitated using target expansion and textual guidance.

The difference of performance between techniques mostly relies on item selection times. When using *Headpointing*, participants do not have any control on the virtual menu location and must point at a distant target using the head cursor. On the contrary, *Bring2Me* techniques alter the location of menus and can temporarily bring it as close to the participant as possible. Selections can thus be facilitated since targets are bigger. This is a non-negligible advantage of *Bring2Me* techniques over the baseline, even if in our setup the difference between the target sizes was limited.

Beyond the difference of performance, the negative qualitative feedback about turning the head must be emphasized. Even if our experimental protocol did not involve a distraction task requiring visual attention from participants, the loss of the initial virtual context when turning the head to interact was judged tiring, frustrating and disturbing. Such results highlight the potential of *Bring2Me* techniques.

6.2 Bring2Me techniques

Overall, the three *Bring2Me* techniques have similar performance, but *TeleHead* was less appreciated than *TelePad* and *Grab*. Reasons evoked by participants are mainly the lack of stability and the lack of familiarity of head pointing compared to hand pointing, the latter being closer to using a mouse. This users preference tendency confirms previous work [18, 19]. In their work about head and eye-gaze pointing, Kyto et al. found that participants preferred using a hand-held device to precisely refine the selection than using head pointing [18]. Lee et al. compared different raycasting selection techniques, and found that hand-based raycasting outperformed head-based raycasting in terms of task completion time and error rate [19].

6.2.1 The mixed approach: *Grab*. Noteworthy, *Grab* was perceived as one of the most accurate technique in our setup. Errors made during the menu selection phase were similar between the three *Bring2Me* techniques, but no item selection error was made with *Grab* during the item selection phase. Continuous control over the menu position may thus be a significant design parameter in contexts like augmented surgery where precision and the feeling of control are key factors [7]. Continuous control allows more flexibility by letting users choose the position of the menu for the item selection phase, which can for instance limit disturbing visual occlusion compared to *TeleHead* and *TelePad*.

Moreover, *Grab* is also the only technique allowing to use two input modalities at the same time to select menu items. While the menu position is controlled through the HMD pad, the head can also be used to move the head cursor. However, our experimental protocol prevented participants to look away from the Visual Context. This design choice limited the amplitude of possible head movements. The analysis of trajectories suggest that only 2 participants over 12 took benefit from the possibility of coordinating movements of the head and the pad. More research is thus required to evaluate the impact of the parallel usage of the two input modalities. In particular, studying further the balance between 1) letting users to freely turn their head to reach the grabbed menu faster and 2) not losing the visual context would further determine the potential of the *Grab* technique.

6.2.2 Menu instant jumps. Contrary to *Grab*, *TeleHead* and *TelePad* techniques offer only a discrete control over the menu location. When selected, the menu makes an instant jump to its new location and then stays fixed at this position until an item is selected. This mechanism creates visual occlusion, but the menu behavior is deterministic. The ability to predict where the menu will appear may have limited any disorientation effect in our study: no effect was observed and only one participant reported not being sure of where the menu would appear. This hypothesis is coherent with previous work like the Jumping Menu from Ahlstrom et al. [2]. With the Jumping Menu, the cursor jumps to the first item of the sub-menu when a mouse click is detected inside a parent item. The authors observed that if some participants found this menu system somewhat irritating, most of participants judged it better than a standard menu and had better performances with it once familiar with the system [2].

Design parameters	Techniques	Benefits	Limitations
Turning the head	<i>HeadPointing</i>	Hands not required to interact Easy to explain to participants	Initial visual context not preserved Slowest and least preferred technique
Bringing back the menu to the FOV	<i>Bring2Me</i> techniques: <i>TeleHead</i> , <i>TelePad</i> and <i>Grab</i>	Initial visual context preserved Preferred by all participants Menu closer to user: bigger visual size	1 hand required to interact
Discrete control of menu position	<i>TeleHead</i> , <i>TelePad</i>	Instant jump to FOV: no disorientation effect observed Limit hand tiredness compared to continuous control	2-steps pointing: first the menu, then the item Cannot control the menu position after the instant jump
Continuous control of menu position	<i>Grab</i>	Can place the menu anywhere: avoid or limit visual occlusion	Additional hand tiredness
Head-based item selection	<i>TeleHead</i>	Easy to perform once the menu is in the FOV	Modality change between menu and item selection More time required to reach optimal performance
Hand-based item selection	<i>TelePad</i>	Same modality for menu and item selection	Additional hand tiredness
Mixed (head+hand) item selection	<i>Grab</i>	Possible parallel use of the two modalities Most accurate technique for item selection	2 participants only used head+hand in parallel




 Making the menu visible
  Controlling the menu position
  Selecting the item

Figure 8: Benefits and limitations of the techniques evaluated in our experimental study.

In addition, *TeleHead* and *TelePad* only teleport a menu to users' FOV. Selecting the wrong menu has limited consequences as it only requires to close the menu or to perform another selection to switch the unwanted menu with the targeted one. Therefore, erroneous menu selections with *TeleHead* and *TelePad* techniques cause less issues than approaches based on directly selecting an item [20, 28].

6.3 Perspectives

6.3.1 Location of targets. In our experimental setup, the menu selection phase for *TeleHead* and *TelePad* can be achieved without moving the arm as using wrist rotations only is enough to reach the menus. However, this may not be the case in a more realistic environment with targets all around users. Most of the participants in our experimental study were right-handed and 2 of them reported that the rightmost menu (+50 degrees horizontally) was more difficult to select. These participants explained that they were close to the limit of their wrist rotation amplitude. Beyond this limit, an arm movement may thus become necessary to point at further menu targets. More research is therefore required to study *Bring2Me* techniques in more complex environments. For instance, it would be interesting to greater amplitude of position for targets like -180°/+180° horizontally and -60°/+60° vertically in [28]. Greater amplitudes could also be used to further study the parallel use of two input modalities (head+hand) for *Grab* during the item selection phase.

6.3.2 Generalizing the results. In this work, augmented surgery is a promising and concrete application scenario as MR techniques preserving the visual context are needed by surgeons. However, we chose to conduct a lab study without ecological conditions in order to obtain generalizable results. Other domains with highly demanding visual tasks such as manufacturing [22], machine assembly and maintenance [29] and airline cockpits [14, 25] could also benefit from our work. Moreover, results suggest that even without a visual demanding task, all participants preferred bringing back the menu to their FOV than turning the head. Therefore, *Bring2Me* techniques could also be used in more general scenarios of MR interaction with widgets.

The *Bring2Me* techniques need to be studied in other scenarios with more complex menu designs and positions. Moreover, they

need to be compared with existing eyes-engaged techniques requiring hands to interact [6, 18]. These techniques may be faster and easier to use than the *Bring2Me* techniques, but at the cost of losing the visual context. We believe that verifying one of the two properties (preserving the visual context and being hands-free) necessarily impacts the interaction performance. The design choice about these two properties will depend on the application domain, and Figure 8 summarises the benefits and limitations of our techniques to help designers in their choices.

7 CONCLUSION

This paper contributes to addressing a significant issue in MR interaction with a HMD: the loss of the visual context occurring when users turn their head to see virtual widgets. This phenomenon is common as the size of the augmented FOV is limited with current HMDs. In our study, we focus on techniques preserving this visual context in the case of interaction with menus. The *Bring2Me* techniques we propose allow users to bring virtual menus to their FOV. These techniques take benefit from the spatial memory, are based on off-screen pointing by raycasting and consider the entire menu interaction, from menu selection to item selection.

Results suggest that head pointing with a head-controlled cursor is 1.5s longer on average than *Bring2Me* techniques. Head pointing should thus be used only when hands are not available. Otherwise, when: 1) at least one hand can be used at the moment of the interaction and 2) users' visual context must be preserved, *Bring2Me* techniques define a promising approach. These techniques were preferred by all participants, and we propose a set of recommendations by discussing their benefits and limitations.

Thanks to an industrial partnership, we are implementing the interaction techniques in a professional surgical system called the OrthoPilot [1]. Our goal is now to evaluate the *Bring2Me* techniques with surgeons.

REFERENCES

- [1] Aesculap. 2020. OrthoPilot® navigation system. Retrieved January 9th, 2019 from <https://www.bbraun.com/en/products-and-therapies/orthopaedic-joint-replacement/orthopilot.html/orthopilot.html>
- [2] David Ahlstrom, Rainer Alexandrowicz, and Martin Hitz. 2006. Improving Menu Interaction: A Comparison of Standard, Force Enhanced and Jumping Menus. In

- Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1067–1076.
- [3] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.
 - [4] Takeshi Asano, Ehud Sharlin, Yoshifumi Kitamura, Kazuki Takashima, and Fumio Kishino. 2005. Predictive interaction using the delphian desktop. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. Association for Computing Machinery, New York, NY, USA, 133–141.
 - [5] Charles Bailly, François Leitner, and Laurence Nigay. 2019. Head-Controlled Menu in Mixed Reality with a HMD. In *IFIP Conference on Human-Computer Interaction*. Springer, 395–415.
 - [6] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: a 3D Pointing Facilitation Technique based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 101.
 - [7] Joan Cassell. 1987. On control, certitude, and the paranoia of surgeons. *Culture, medicine and psychiatry* 11, 2 (1987), 229–249.
 - [8] Isha Chaturvedi, Farshid Hassani Bijarbooneh, Tristan Braud, and Pan Hui. 2019. Peripheral vision: a new killer app for smart glasses. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 625–636.
 - [9] Lawrence D Cutler, Bernd Frohlich, and Pat Hanrahan. 1997. Two-handed direct manipulation on the responsive workbench. In *Proceedings of the 1997 symposium on Interactive 3D graphics*.
 - [10] Dragos Datcu and Stephan Lukosch. 2013. Free-hands interaction in augmented reality. In *Proceedings of the 1st symposium on Spatial user interaction*. ACM, New York, NY, USA, 33–40.
 - [11] Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330.
 - [12] Barrett Ens, David Ahlström, Andy Cockburn, and Pourang Irani. 2011. Characterizing user performance with assisted direct off-screen pointing. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 485–494.
 - [13] Barrett Ens, David Ahlström, and Pourang Irani. 2016. Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View Limitations. In *Proceedings of the 2016 Symposium on Spatial User Interaction*. ACM, New York, NY, USA, 107–110.
 - [14] David C Foyle, Anthony D Andre, and Becky L Hooey. 2005. Situation awareness in an augmented reality cockpit: Design, viewpoints and cognitive glue. In *Proceedings of the 11th International Conference on Human Computer Interaction*, Vol. 1. 3–9.
 - [15] Maxime Guillon, François Leitner, and Laurence Nigay. 2014. Static Voronoi-based Target Expansion Technique for Distant Pointing. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 41–48.
 - [16] Zhenyi He and Xubo Yang. 2014. Hand-based interaction for object manipulation with augmented reality glasses. In *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*. ACM, New York, NY, USA, 227–230.
 - [17] Bonifaz Kaufmann and David Ahlström. 2012. Revisiting peephole pointing: a study of target acquisition with a handheld projector. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, New York, NY, USA, 211–220.
 - [18] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages.
 - [19] Sangyoon Lee, Jinseok Seo, Gerard Jounghyun Kim, and Chan-Mo Park. 2003. Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, Vol. 4756. International Society for Optics and Photonics, 38–44.
 - [20] Frank Chun Yat Li, David Dearman, and Khai N Truong. 2009. Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 125–128.
 - [21] Microsoft. 2020. HoloLens 2 HMD by Microsoft. Retrieved January 2020 from https://www.microsoft.com/en-us/hololens?icid=SSM_AS_Promo_Devices_HoloLens2
 - [22] SK Ong, ML Yuan, and AYC Nee. 2008. Augmented reality applications in manufacturing: a survey. *International journal of production research* 46, 10 (2008), 2707–2742.
 - [23] Kasım Özacar, Juan David Hincapié-Ramos, Kazuki Takashima, and Yoshifumi Kitamura. 2016. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. *Interacting with Computers* 29, 4 (2016), 579–591.
 - [24] Robert Patterson, Marc D Winterbottom, and Byron J Pierce. 2006. Perceptual issues in the use of head-mounted visual displays. *Human factors* 48, 3 (2006), 555–573.
 - [25] Sylvain Pauchet, Catherine Letondal, Jean-Luc Vinot, Mickaël Causse, Mathieu Cousy, Valentin Bequet, and Guillaume Crouzet. 2018. GazeForm: Dynamic Gaze-Adaptive Touch Surface for Eyes-Free Interaction in Airliner Cockpits. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 1193–1205.
 - [26] Sebastien Pelurson and Laurence Nigay. 2015. Multimodal interaction with a bifocal view on mobile devices. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 191–198.
 - [27] Viktoria Witka. 2016. Perceptual Issues in Multi-Display Environments. *Mobile Multi-Display Environments* (2016), 1.
 - [28] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 42.
 - [29] ML Yuan, SK Ong, and AYC Nee. 2008. Augmented reality for assembly guidance using a virtual interactive tool. *International journal of production research* 46, 7 (2008), 1745–1767.