

Head-controlled Menu in Mixed Reality with a HMD

Charles Bailly^{1,2}, François Leitner², and Laurence Nigay¹

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

² Aesculap SAS, F-38130 Echirolles, France
{charles.bailly, francois.leitner}@bbraun.com

Abstract. We present a design-space and three new techniques for head-based interaction with menus in Mixed Reality (MR) with a Head-Mounted Display (HMD). Usual input modalities such as hand gestures and voice commands are not suitable in noisy MR contexts where the users have both hands occupied as in augmented surgery and machine maintenance. To address the two issues of noisy MR contexts and hand-free interaction, we systematically explore the design space of head-controlled menu interaction by considering two design factors: 1) head-controlled menu versus head-controlled cursor 2) virtual targets versus mixed targets anchored on physical objects. Based on the design space, we present three novel menu techniques that we compared with a baseline head-controlled cursor technique. Experimental results suggest that head-controlled menu and head-controlled cursor techniques offer similar performance. In addition, the study found that mixed targets do not impact ultimate user performance when users are trained enough, but improve the learning phase. When using virtual targets, users still progressed after the training phase by reducing their mean selection time by 0.84s. When using mixed targets, the improvement was limited to 0.3s.

Keywords: Mixed Reality · HMD · Menu technique

1 Introduction

Mixed Reality (MR) using a Head-Mounted Display (HMD) has a strong potential in several domains [6, 38], including medicine, manufacturing and repair. Superimposing virtual content on a working environment allow users to visualize relevant data directly in place instead of relying on a distant screen. For instance, MR can replace manuals by superimposing instructions directly on the working environment to help technicians [36, 52]. MR is also studied as a training tool for assembly tasks [41].

For augmented surgery, a HMD can be used: 1) to display medical information directly on the patient body [3, 27, 32] and/or 2) to display the Graphical User Interface (GUI) of the surgical system including medical images, menus and palettes. Surgical navigation systems like the OrthoPilot®[1] are interactive systems supporting surgeons during the entire workflow of the surgery (hence the

term *navigation* systems). Issuing commands with such systems is usually done through 1-level menus currently displayed on a distant screen in the operating room. As a first step towards replacing the distant screen by using a HMD, this paper focuses on pointing to a menu in MR.

Existing interaction techniques for MR using a HMD are not always adapted to specific contexts of use as in augmented surgery and machine maintenance. Surgeons must follow strict sterility rules and often have both hands occupied by surgical tasks. The same applies for mechanical tasks: technicians need both hands to hold tools and perform their repair task while still needing to interact with the system. In addition, operation rooms and machine areas are both noisy and congested environments. Finally, surgeons and military mechanics both have a significant cognitive load because of the complexity and critical aspects of their tasks [11, 19]. Such constraints limit the possible techniques for interacting with the system. New hand-free interaction techniques adapted to these specific contexts of use are needed. This work focuses on menu techniques for HMD-based MR that are adapted for these contexts of use.

The contribution of this work is the design and the evaluation of 3 head-based techniques for interacting with a menu in MR. These techniques are based on two design factors: (1) Head-controlled menu versus Head-controlled cursor and (2) virtual targets versus mixed targets. In the following, the augmented surgery scenario is taken as a running example. However, the techniques we designed target any context where users work with both hands occupied in a noisy and congested environment. In the rest of this paper we discuss background work, present the design space of head-based menu techniques in MR and describe the three designed techniques we developed. We then present the results of two conducted studies to compare the performance of the designed menu techniques. We conclude by discussing design implications for the menu techniques we have studied, and suggest potential future pathways for developing menu techniques in MR using a HMD.

2 Related work

2.1 Interaction techniques in MR

Gesture-based interaction: Gesture-based interaction is commonly used for HMD-based MR. Some approaches are based on specific gloves to allow direct hand manipulation of virtual objects as [10], while more recent approaches use bare hand gesture recognition systems [12, 13, 18]. However, adapting the gestural recognition system to specific environments like operating rooms requires significant effort [37]. Operation rooms are very bright environments and gestures from other surgical team members must not be interpreted as commands. Besides, gesture-based interaction requires one to use one or two hands. This is an issue for surgeons and mechanics, who need both hands to perform their technical tasks.

Direct touch interaction: For the same reasons as before, interaction based on tangibles or body direct touch is not adapted to augmented surgery or maintenance. In the case of surgery, strict sterility constraints prevent direct touch interaction on the body [5]. Furthermore, all objects in the operating field must be sterilized before the surgery. Introducing new tangibles to interact with menus as in [48] would increase this already long preparatory phase. As explained in section 1, the lack of available space is also an issue. For instance, Henderson and Feiner [19] describe a prototype MR application to help military mechanics perform maintenance inside an armored vehicle turret, where available space is really limited.

Voice control: When hands are not available for interacting, voice commands offer an alternative interaction technique [2, 34]. However, ambient noise is a issue for voice recognition software [21, 37]. The operating theatre is a noisy environment due to the medical equipment and oral communication between surgical team members. This also holds for machine rooms which often have noisy equipment. Besides, as pointed out by Mitrasinovic et al. [35], speaking slowly and distinctly is a luxury during a surgery since time is a critical factor. Therefore, voice commands are not suitable for our targeted application contexts.

2.2 Head-based interaction

Head-based gestures have been studied in the literature as a technique to interact with virtual content displayed on a HMD [33, 50, 51]. However, using a set of previously defined head-gestures may not be suitable in our application contexts. In [33], users have to perform different head gestures depending on the task. Having to learn and remember which gesture must be done for each type of command may increase the cognitive load of an already highly-focused user with busy hands (e.g. a surgeon, a technician). Moreover, head gestures also raise the issue of tiredness, especially for the case of surgeons performing more than one hour long operations. Finally, as for voice commands, head gestures must also be distinguished from unintentional inputs [50].

In addition to head gestures, head motion has also been explored as a technique for pointing tasks. Head motion is used to control a virtual cursor, in Virtual Reality (VR) [4, 40] and in MR [29, 39]. Jagacinski and Monk showed in 1985 that Fitt's Law is a good predictor of a head pointing task performance [25]. When compared to eye-gaze input, head-controlled cursor was found more comfortable to use and more accurate [26, 29]. Comfort and accuracy are key factors, in particular for surgeons: comfort because operations can last more than one hour and a half; accuracy to avoid errors and feel in precise control. Cassell highlighted that the feeling of certitude and control are crucial in the operating room because surgeons have to take life-and death decisions [11]. Anything threatening the confidence surgeons have in the mastery of their tools and environment must be avoided [15].

Using the head as a technique for interacting with graphical widgets displayed on

the HMD is therefore a promising approach that matches the requirements of our targeted contexts. As mentioned in section 1, this paper focuses on head-based interaction with a menu.

3 Design space for Head-based menu techniques in MR

We used two factors to define the basis of a design space for head-based menu techniques. Our design rationale is based on the specificities of MR with a HMD. In particular, the augmented Field Of View (FOV) is limited but directly controlled by the head.

3.1 Interaction element controlled by head movements

The first design factor is the element controlled by users when moving their head: a cursor or the menu itself. When using a HMD, the augmented FOV is directly linked to the head position. Therefore, any graphical element displayed at a fixed position on the HMD will move in the physical world frame of reference according to the head movements. The graphical element is fixed in the screen frame of reference and movable in the physical world frame of reference. When interacting with a menu, this principle leads to two solutions: moving the cursor or moving the menu. With a Head-Controlled Cursor (HCC), users aim at targets (i.e. menu items) that are fixed in the physical world frame of reference by orienting their head and aligning a cursor on the targets (see Figure 1a and section 2.2). HCC minimizes the visual intrusion on the HMD: the cross displayed at the center of the screen. This is a significant advantage since the augmented FOV is limited on current HMDs. Besides, HCC reproduces the same metaphor as in GUI using the mouse, and is therefore familiar to MR novice users.

Instead of moving a cursor to a menu item, the menu can be directly controlled with the head. When using a Head-Controlled Menu (HCM), the targeted menu item must be brought to a fixed location in the physical world frame of reference (i.e. a cursor position) in order to be selected (see Figure 1b). To some extent, this approach is similar to the Toolglass technique [9]. The Toolglass is a transparent sheet with a set of tools (*i.e.* a palette) which can be controlled with a trackball. This palette can thus be superimposed onto the object of interest with one hand while the other hand control the mouse, allowing users to select both the command and its target at the same time. HCM and Toolglass share one idea: bringing what Beaudouin-Lafon [8] calls meta-instruments to a selection area (i.e. the cursor for a HCM and the object of interest for a Toolglass). Of course, a Head-Controlled Menu implies a significant visual intrusion by displaying the menu at the center of the HMD, especially as compared to a Head-Controlled Cursor. During the critical phases of their work, surgeons and mechanics only focus on the complex gestures to be performed, and any visual intrusion may threaten the process. However, when interacting with the navigation system, these users are not performing such critical phases. Therefore,

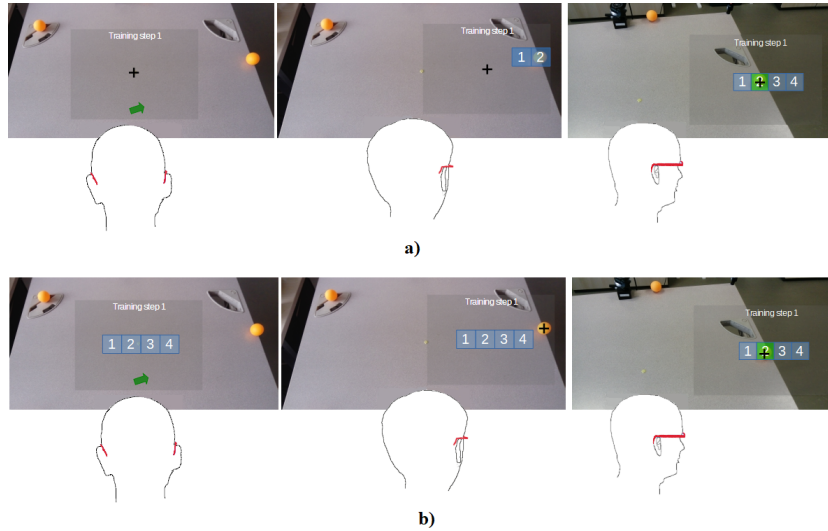


Fig. 1: a) Head-Controlled Cursor: head-based movement of a cursor towards a menu. b) Head-Controlled Menu: head-based movement of a menu towards a cursor

visual intrusion at these moments is less of an issue. Visual intrusion can also be mitigated by the fact that menus do not require permanent screen space. Menus are transient: they are closed immediately after the selection of an item and they appear on demand [7]. For instance, the system could display the menu only when users move their head away from the main area of interest in a given direction, or after a footswitch pedal press. These actions could be a trigger to indicate that the current goal is to interact with the system. One benefit of HCM is that right at the beginning of the selection users perceive the menu and are therefore aware of the currently available commands.

Head-Controlled Cursor (HCC) and Head-Controlled Menu (HCM) define the first axis of our design space, namely Controlled Interaction Element (see Figure 2a).

3.2 Tangibility of the target

Contrary to Virtual Reality where every visible element is virtual, MR opens more possibilities along the Reality-Virtuality Continuum. Interaction elements can either be physical, mixed or virtual. Physical and virtual elements are straightforward: physical objects are visible even without the HMD while virtual elements are only visible through the HMD. The tangible menu proposed by Ullmer et al. [47] is an example of a physical interaction element. Mixed elements are different since they have both a physical and a virtual component. These physical and virtual components can share the same position (in the case

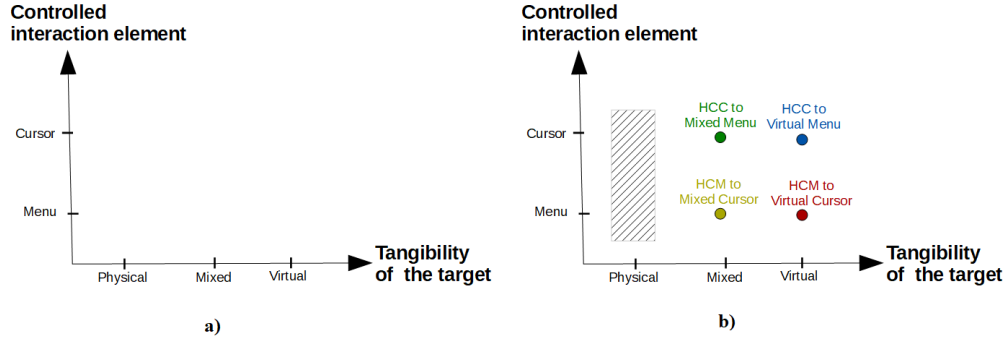


Fig. 2: (a) Design space for head-based interaction with a menu in MR with HMD. (b) The 4 studied interaction techniques located in the design space

of augmented objects for instance), but can also be spatially disjoint. Nonetheless, even when outside of the limited augmented FOV, mixed elements are still partially visible thanks to their physical component.

On the one hand, since we focus on head-based interaction with a menu, the interaction element controlled by head movements must be virtual. Attaching an object to the head just to be able to interact would be much more cumbersome. Thus, by moving the head, users control either a virtual cursor (Head-Controlled Cursor) or a virtual menu (Head-Controlled Menu). On the other hand, the target can belong to any position along the Reality-Virtuality Continuum. This aspect defines the second axis of our design space, which describes the tangibility of the target (see Figure 2a). Our goal was to determine if materializing the graphical element brings any benefit to the user. We study if embodying the location of targets to make them visible in the peripheral vision even outside the augmented FOV is suitable or not. Using physical objects as anchors for virtual content is not new: Henze et al. mentioned this approach as future work in their study of visualization of off-screen objects in augmented maps [20]. However, this approach needs to be studied and evaluated. In particular, we are interested in studying if such physical anchors have an impact on learning the location of targets (in our case, menu items).

As detailed in section 2, introducing new objects as anchors for virtual content in operation rooms is not suitable. However, some already present objects could be used as anchors. For instance, surgeons fix *rigid bodies* on the patient's body to allow a navigation system like the OrthoPilot® to get anatomical points of reference. Rigid bodies are metallic objects with infrared markers and are already part of the sterilization process. Therefore, they could be used as physical anchors for virtual targets. The same approach is applicable to others contexts like manufacturing and maintenance. Introducing new physical objects can be difficult, but objects already present in the environment could be detected and used.

4 Design and implementation of Head-based menu techniques

We extracted four menu techniques from the design space. Three of these techniques are new, and the fourth one is used as a baseline since it already exists in the literature.

4.1 Design of the menu techniques

The two-axes design space of Figure 2a defines 6 types of menu techniques. However, we choose to study only 4 menu techniques by excluding the 2 techniques that involve a physical target (see Figure 2 b). We motivate the decision by the targeted contexts of use:

1. Pointing to a tangible menu is a very unlikely scenario in surgery or maintenance since it would mean introducing a new tangible in these specific environments. We thus eliminated the *Head-Controlled Cursor to Physical Menu* technique.
2. The other menu technique with a physical target, *Head-Controlled Menu to Physical Cursor*, would be easier to design and implement than a complete tangible menu. However, since we exclude the first type of menu technique involving a physical target, we do not consider *Head-Controlled Menu to Physical Cursor* in this study. We thus keep the entire Physical target category (hatched area in Figure 2b) for a later study.

Among the 4 remaining menu techniques, we particularly focus on the 2 techniques with a Head-Controlled Menu (HCM) as there are the most innovative ones. The *Head-Controlled Cursor to Virtual Menu* already exists in the literature [25, 30] and in some VR headsets like the first version of the Oculus Rift. This technique is used as a baseline. However, the influence of pointing to a *Mixed* menu instead of a *Virtual* menu still needs to be studied (*HCC to Mixed Menu* technique).

As a first exploration of the design space, we consider a simple one-level menu. The menu techniques can be used in several contexts, like surgery, machine maintenance and repair. We thus did not adapt the menu to a specific group of users. Environmental conditions can also be different from one context to another: surgery rooms are very bright environments, which is not the case for the mechanics working on vehicles in [19] for instance. Therefore, since we do not aim to recreate specific ecological conditions, we choose to consider a simple menu design as shown in Figure 1. A one-level horizontal menu is used and each menu item is labeled with a number.

4.2 Implementation of the menu techniques

The 4 menu techniques were developed with the Epson Moverio BT-300, running on Android 5.1. We chose this HMD because of its light weight (70g for the

headset). Machine maintenance and surgery are physically and mentally tiring tasks, so minimizing the weight of the HMD was a determinant factor of choice. To be able to track the position of the HMD, we fixed three passive markers on it. The markers were tracked by a set of four Flex3 cameras monitored by the *Tracking Tools* software (version 2.3.1). This software was executing on a nearby laptop. The position of the MR glasses (and thus the participant’s head orientation) was sent to the HMD through Wifi thanks to a custom Python script. The screen of the HMD was then updated according to where the participant was facing.

To validate a selection, dwell time is a classical approach to validate the selection of a target, and was often used for eye-gaze selection [24, 44]. However, it was also criticized by several authors because this method slows down the interaction in order to avoid errors and it requires significant concentration from the user [46]. Since both of these limitations would have a significant impact for the case of our targeted scenarios, we chose another approach using a pedal like in [17]. To validate a target, a 2-pedals Steute orthopedic footswitch was used as this device is commonly used by surgeons for validation in the operating room. This technique corresponds to the first foot interaction style identified by Rovers et al.: a simple toggle action [42]. In the following section, we describe the experimental studies we conducted to compare the 4 implemented menu techniques.

5 Study 1

This first experimental study had two goals. Our first goal was to compare the Head-Controlled Cursor (HCC) and Head-Controlled Menu (HCM) approaches. Since HCC is more similar to well-known WIMP interaction than HCM, we expected participants to prefer and perform better with HCC (**H1**). Our second goal was to evaluate the benefits of physically anchoring virtual targets using physical objects. The resulting *Mixed targets* have a physical component (the anchor) visible even outside of the augmented FOV. We hypothesized that *Mixed targets* would have a positive impact on the menu interaction for two reasons: 1) it may facilitate learning target locations; 2) it could change how participants are planning their head movements, leading to better performance (**H2**).

5.1 Participants

Twelve participants (3 female, 9 male) from 22 to 39 years old ($sd = 5.3$) were recruited for this study. As reported in our preliminary questionnaire, they were all novices in Mixed Reality. Participants had normal or corrected-to normal vision, and could use glasses in addition to the HMD if necessary.

5.2 Setup

The goal of our setup was to recreate the conditions of a general scenario when users have both hands occupied. To do so, we were inspired by a generic aug-

mented surgery scenario. However, we did not try to recreate ecological conditions as it is beyond the scope of this study. Participants stood behind a grey squared table (81cm long, 73 cm high) representing the working area. They were asked to hold two light plastic bones to keep both hands occupied. Physical anchors for the 2 *Mixed targets* conditions were represented by orange ping-pong balls to be sure they were clearly visible on the grey table in terms of size (diameter: 3.7cm) and color (see Figure 3).

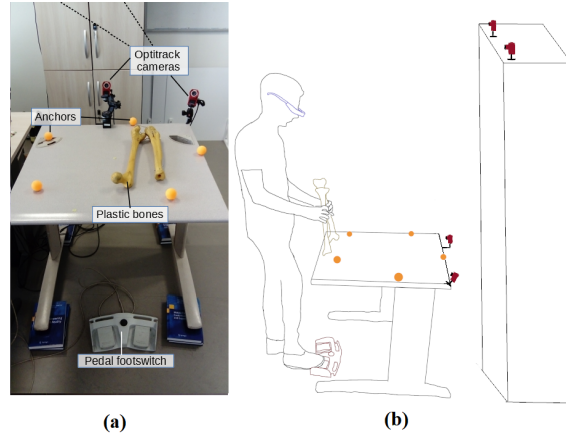


Fig. 3: (a) Annotated experimental setup. (b) A participant's pose during the experiment

5.3 Task

Participants were told they had to quickly and accurately select an item from a virtual menu using their head. They had to hold the plastic bones with both hands and keep them in a given position. A supervisor payed attention to be sure they maintained a correct pose during the entire experiment. This was done to ensure they could not use their hands to interact. A one-level horizontal menu with four 100*100 pixel items was used (see Figure 1). Such a one-level menu with four items corresponds to the main menu of the OrthoPilot® navigation system [1]. This system from the Aesculap company guides surgeons during orthopaedic operations of the knee or ankle.

Similarly to the study performed by Louis and Berard [31], we chose to keep a small number of five targets to allow participants to quickly memorize the different target locations. We wanted to be sure that participants had prior knowledge of target locations to avoid exploration phases during trials. This was motivated by the fact that surgeons and mechanics cannot afford wasting time searching for targets while performing their main tasks.

Target positions: Inspired by the approach adopted by Guillon et al. [16], we applied a modified version of the ISO 9241-9 pointing task [23]. As stated before, five targets are placed on the operating table and form a circular shape. We asked participants to go back to the center of this circle (which was also the center of the table) to start a new pointing task. To be sure they were correctly placed, they were asked to align a small circle displayed on the HMD with a small circle at the center of the table. When correctly oriented the circle displayed on the HMD became green, and participants could validate in order to get the next target.

The order of targets was randomized. To identify the current target, a green arrow pointing to the target location was displayed at the bottom of the HMD (see Figure 1). This guidance arrow disappeared when the target menu (in the HCC case) or the target cursor (in the HCM case) became visible on the HMD. All targets were outside of the initial augmented FOV. With *Mixed targets* conditions, physical anchors were visible in the peripheral vision. Besides, we chose to set the Index of Difficulty (ID) of all targets to 3.2 to have tasks with a comparable difficulty to the tasks in the study by Guillon et al. [15]. The authors studied target expansion techniques for distant pointing in the context of augmented surgery with the OrthoPilot®[1]. The ID was computed using Wingrave’s and Bowman’s formulation of Fitts’ law [49]. An ID of 3.2 is relatively easy according to the range of ID values between 2 and 8 suggested by Soukoreff and Mackenzie [45].

Selection: When the mobile part of the interaction was correctly positioned, the targeted menu item was highlighted with a green color (see Figure 1). To validate the selection, participants used a footswitch similar to what surgeons currently use in the operating room as a validation device. If the selection was correct, the target disappeared and the participant was asked to go back to the initial position. When correctly positioned, participants had to click again to start a new trial. If an error was made during a trial, the participant had to continue with the current target until he finally reached it.

5.4 Study design

The experiment was divided into 4 blocks of 30 trials each, one block per Technique. The order of blocks was counterbalanced using a Latin-square design. The menu item which had to be selected (item number 2) was the same for the entire experiment to keep the same index of difficulty for each target. Therefore, we had 4 menu techniques (2 *Controlled interaction element* x 2 *Tangibility of target* design options) x 5 target positions x 6 repetitions, leading to 120 trials per participant.

Before each block, a corresponding training session was conducted. Training trials were identical to real trials, but nothing was recorded except the total training time. At the beginning of the experiment, we explained to participants that they

had to learn the position of the five targets around the operating table, and that these targets would remain the same during the entire experiment. Participants were asked to train with the current menu technique until (1) they knew the five target locations and (2) they felt efficient and comfortable with it. A minimum of 12 training trials was imposed.

We observed during our pilot study that some participants experienced eye tiredness after doing more than 60 consecutive trials. Therefore, between each block of 30 trials, participants were encouraged to take a break. After the break, participants were given a questionnaire to collect their feedback about the menu technique they just experienced. A final questionnaire was also given at the end of the experiment to obtain feedback about the comparison of all techniques.

5.5 Measures

For each trial, the time to perform the valid selection and the number of errors made by participants were recorded. When a mistake was made, the distance between the current position and the target center was recorded as well. Hereafter, this distance will be called *error distance*. Moreover the head trajectory was also recorded to be able to compare it to the ideal one. To do so, the head position was saved every 100ms during trials. The goal was to better understand the participant’s performance and discriminate movement phases. Qualitative feedback was divided in two kinds of questionnaire. First, participants were given a questionnaire based on a 7-degrees Likert scale after each block. The questions were about subjective performance (speed, accuracy) and subjective comfort (easy to understand, physical and cognitive loads). Then, at the end of the experiment, a global questionnaire asked participants to rank the 4 menu techniques according to the same criteria (subjective performance and comfort). Tiredness, perceived utility of mixed targets and preferences were also discussed during semi-structured interviews.

5.6 Results

For analysis, we used a repeated-measures ANOVA (with $\alpha = 0.05$). Post hoc tests were conducted with pairwise t-tests and Bonferroni corrections. Following the guidelines by Dragicevic [14], a log transform was applied to the data when it was relevant (i.e. for the case of positive skewness in time measurements). When using ANOVA was not appropriate, we used a Friedman test instead followed by a Wilcoxon post-hoc test with Bonferroni corrections. 13 trials (0.9% of the 1440 total trials) were removed from the analysis because of lag due to tracking system (*e.g.* hair partially masking the tracked markers of the HMD) or because participants had to adjust the HMD on their head in the middle of a trial.

Target selection time: Considering the entire experiment, results are largely inconclusive concerning the differences between the four techniques in terms of target selection time ($F(3, 33) = 0.11, p = 0.95$), as shown in Figure 4a. The mean

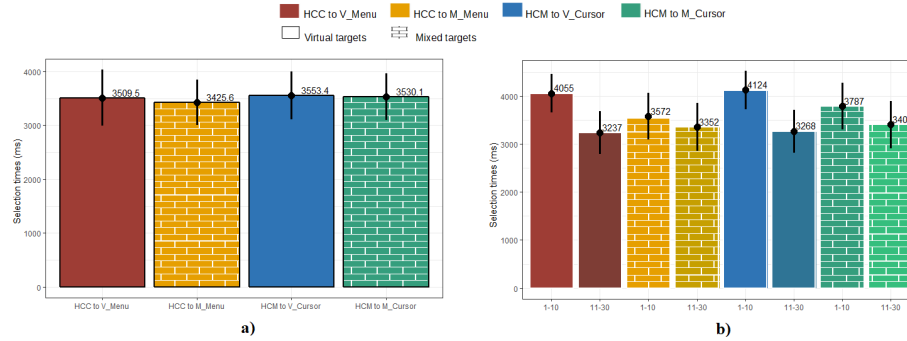


Fig. 4: (a): Study1: Target selection times for the four menu techniques. (b): Study1: Comparing selection times within each block: the 10 first trials versus the 20 last trials. Error bars are 95% CI

selection time was close to 3.5 seconds for all menu techniques. No difference was found between the HCC and HCM techniques. However, we observed a different tendency when comparing the first 10 trials of each block with the 20 remaining trials (Figure 4b). In this case, we found good evidence of differences between techniques with virtual targets and techniques with mixed targets. The evidence was strong for *HCM to Virtual Cursor* ($p = 0.01$) and *HCC, to Virtual Menu* ($p = 0.02$). In comparison, we did not find evidence for *HCM to Mixed Cursor* ($p = 0.66$) and no evidence at all for *HCC to Mixed Menu* ($p = 1$). When using virtual targets, participants still progressed after the training phase by reducing their mean selection time by 0.84s. When using mixed targets, the improvement was limited to 0.3s. Besides, participants' training times were similar for all techniques (no differences found, $p = 0.48$). Participants thus still progressed significantly during trials when using the two techniques with virtual targets. In order to counter-balance this learning effect, we also checked if the performances were the same once the learning curve was stabilized for all techniques. To do so, we analyzed only the 20 last trials. Once again, we found no evidence of difference between techniques ($F(3,33)=0.2$, $p = 0.91$).

Errors: Similarly to selection time, results are inconclusive about the 4 techniques in terms of error. More precisely, the mean number of errors ($\chi^2 = 2.9$, $p = 0.41$) and mean error distances ($F(3,33) = 1.73$, $p = 0.18$) did not highlight any evidence of difference between techniques (Figure 5). In both cases, the discrepancies between participants also prevented us from drawing any conclusion. We observed similar results when comparing the beginning and the end of each trial block. No difference was found between the 10 first trials and the 20 last trials for both the number of errors ($\chi^2 = 1.9$, $p = 0.2$) and error distances ($F(7,77) = 0.9$, $p = 0.51$).

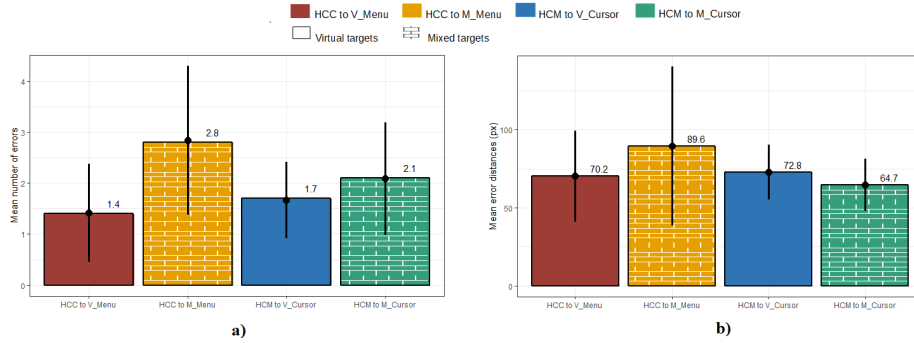


Fig. 5: (a): Study1: Mean error number for the four menu techniques. (b): Study1: Mean error distance for the four menu techniques. Error bars are 95% CI

Qualitative feedback: Based on the global questionnaire at the end of the experiment, *HCM to Virtual Cursor* was ranked first (globally preferred). Most participants preferred using the Head-Controlled Menu rather than using the Head-Controlled Cursor. 8 participants over 12 reported preferring Head-controlled menu techniques because they felt it was easier, quicker and/or more precise. Figure 6 presents an overview of the main qualitative results for each technique.

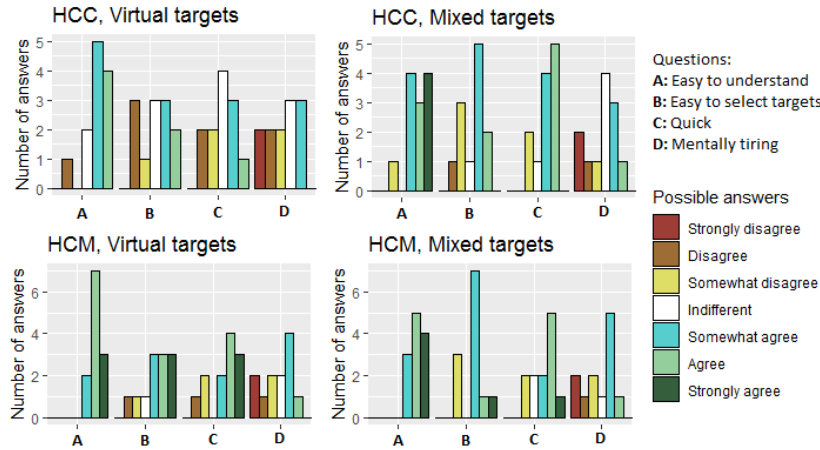


Fig. 6: Study1: Qualitative feedback collected from participants

Interestingly, Participant 12 explicitly mentioned the benefit of seeing the menu from the start: *"I preferred moving the menu with the head (HCM) because you already see it and just search the cross. The two others (HCC) are like twice*

the effort because first you search the menu, then the item". Participant 11 also pointed out that there were less moving elements on screen with HCM (since the menu was fixed at the center of the screen), thus he found HCM less confusing. The 4 participants who preferred HCC techniques gave the same type of reasons: they found HCC easier to use, quicker or felt more in control. Participant 3 also reported being more familiar with the HCC condition.

Five participants did not perceive the difference between HCC and HCM. They said during the training that it was exactly the same (for instance, doing *HCC to Virtual Menu* after *HCM to Virtual Cursor*), and needed several training steps to understand the difference. For the other 7 participants, the difference was obvious from the first training step. At the end of the trials, 2 participants were still unsure of the difference between HCC and HCM techniques. They finally captured the difference during the final questionnaire thanks to the reminder drawings representing each technique that were given with the questionnaire. Opinions about mixed targets and their physical anchors were more diverse. On the one hand, 4 participants declared that they were really useful. On the other hand, 2 participants disliked them. Three other participants reported that they used anchors at the beginning to learn target locations, but not afterwards since they already knew where the targets were positioned. Interestingly, the last 3 participants reported that they found anchors useful, but that the superposition of anchor and targets created visual discomfort.

6 Discussion

All the menu techniques lead to similar performances in terms of target selection time, number of errors and error distances. This result is in contradiction with H1 since we thought HCC techniques would be faster than their HCM counterparts. Moreover, the HCM condition was preferred by most participants. These results highlight the potential of HCM when using simple menus. We purposely chose a simple one-level menu design, with only 4 items to be close to the existing main menu of the OrthoPilot® navigation system [1]. As commented by Participant 10, our current results may not be valid with more complex menus with multiple levels.

Partially supporting H2, marking the location of virtual targets with physical objects impacted the learning process. Participants reached optimal performance faster with mixed targets than with virtual targets. Since participants had to confirm they felt ready before each trial block and that training times were not statistically different, it seems that more repetitions were necessary with virtual targets to reach optimal performance. Further study is required to determine the influence of mixed targets on the different phases of the learning process. Scarr et al. proposed in 2011 a framework of interface expertise [43]. This framework characterizes the evolution of user performance from novice to expert mode, and can be applied to menus [7]. A valuable research question would then be to evaluate to which extent mixed targets impact the *initial performance* and *extended learnability* phases of the user performance [43].

However, it is important to note that once participants were trained enough, physical anchors did not improve performance, which is in contradiction with H2. This performance may be explained by the fact that participants perfectly knew the target locations. Thus, anchors were not useful anymore. This explanation is supported by the feedback of 3 participants.

The visual discomfort reported by some other participants may also have an influence on the performance when using mixed targets. This visual discomfort origin may be due to a vergence-accommodation conflict, a well-known issue with stereo displays [28]. This conflict occurs when the eyes converge on an object (i.e. the physical anchor), but the eyes' lenses accommodate at a different depth (i.e. the depth of the HMD screen where the virtual target is displayed). Vergence-accommodation conflicts have a negative impact on performance and increase visual fatigue [22]. The 3 participants who reported visual discomfort all declared feeling visual fatigue at the end of the experiment. We conducted a second experiment to further study this issue.

7 Study 2

The goal of this second study is to focus on the visual discomfort that was reported by some participants in Study 1. To do so, we designed and implemented alternative versions of the two menu techniques with mixed targets that suppress this visual discomfort. We then compared the 2 new menu techniques with the two menu techniques of Study 1 with virtual targets. The goal was to evaluate if these new menu techniques with mixed targets allow participants to outperform the menu techniques with virtual targets because the visual discomfort was suppressed. The experimental protocol of this second study was similar to the one of the first study. Only the differences between the two protocols are reported in this section. 9 participants (5 female, 4 male) from 20 to 30 years old ($sd = 3.3$) were recruited for this second study.

7.1 Alternative menu techniques: Design rational

When using mixed targets, the users have to focus first on the virtual green arrow to know which target has to be selected, then on the physical anchor, and finally back on virtual elements when the menu appeared. This forces additional eyes accommodation and potentially creates vergence-accommodation conflicts. Moreover, since both the menu and the cursor are partially transparent, it is still possible to perceive the real world behind them, which could disturb the users. When no physical anchor is used, the users can only focus on virtual elements and ignore the real world features.

To address this issue, we propose to progressively mask the real world. The closer the users are to the target, the more opaque the background color of the augmented FOV will become, as illustrated in Figure 7. The opacity change starts as soon as the target appears in the augmented FOV. Thus, when the users have reached the virtual target, they can no longer perceive any real world feature in

the augmented FOV and can fully focus on virtual elements. The goal of this design is to eliminate both the required eye focus transitions and the disturbance from real world features. Conceptually, it corresponds to a progressive transition from a Mixed Reality (MR) mode to a Virtual Reality (VR) mode. However, users are still conscious of their physical environment since the real world remains visible in their peripheral vision. The separation with the physical environment is thus limited in time (only when close to the target) and space (only in the augmented FOV). The concept of these new techniques can also be

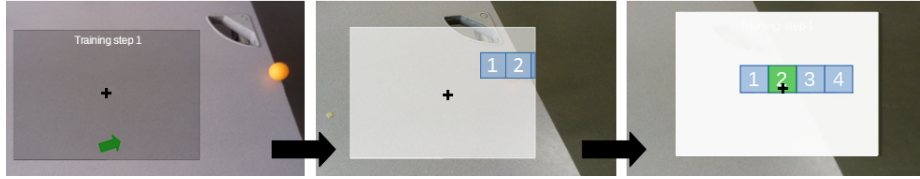


Fig. 7: New menu techniques: Progressive change in the background opacity illustrated with the *HCC to Mixed Menu* technique

seen as a shift in the design space from a mixed target to a virtual one (Figure 2a). A particularly interesting aspect of this shift is that it happens during the interaction. Users begin with a mixed target to be able to see the target location in the peripheral vision. Then, when they no longer need the guidance of the physical anchor, they switch to a virtual target to limit vergence-accommodation conflicts.

We hypothesized the following: masking the real world at the end of the pointing task would help participants by minimizing vergence-accommodation conflicts. Thus, menu techniques with change in opacity will outperform menu techniques without physical anchor (**H3**).

7.2 Results

Selection time: As shown in Figure 8, results for target selection times for the 4 techniques were inconclusive ($F(3,24) = 0.35$, $p = 0.81$). Applying the same analysis process as for Study 1, we checked if participants reached optimal performance after the training phase. Like in Study 1, we compared each pair of the 8 conditions (4 techniques divided in two parts each, 10 first trials and 20 last trials). However, we only found a weak evidence of difference ($F(7,56) = 1.8$, $p = 0.11$). When running a post-hoc test, the sole identified tendency was between <the 20 last trials of *HCC with Mixed targets*> and <the 10 first trials of *HCM with Virtual targets*>, which is not relevant for comparison. However, the tendency for *HCM to Virtual Cursor* seems different from Study 1, where user improvement during trials was clearly observed. Besides, target selection times for the two HCM techniques seem a bit higher than for Study 1 (around 3.9s instead of 3.5s).

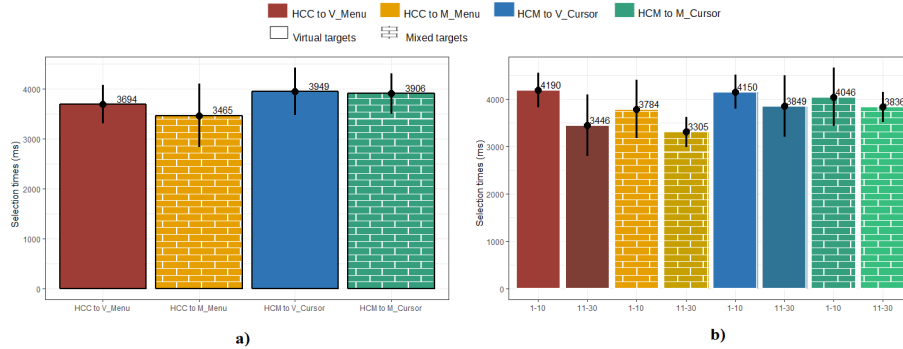


Fig. 8: (a): Study2: Target selection times for the four menu techniques. (b): Study2: Comparing target selection times within each block: the 10 first trials versus the 20 last trials. Error bars are 95% CI

Errors: Similarly to Study 1, we did not find any evidence of difference between techniques when focusing on errors. Both the mean number of errors ($\chi^2 = 9.7$, $p = 0.2$) and mean error distances ($F(3,24) = 0.46$, $p = 0.72$) were similar for all techniques.

Qualitative feedback: 6 of the 9 participants perceived the change in background opacity as an additional feedback for pointing. They reported that this change helped them to adjust their speed when they got close to the target. Interestingly, this was the only benefit reported by the participants: none of them mentioned the change in the nature of the target from a mixed to a purely virtual target. The 3 other participants were more skeptical about the impact of the change in opacity and felt neutral about it. Mixed targets were judged less useful than the change in opacity. 5 participants reported that physical anchors could be useful at the beginning to learn the location of targets. They reported that after a while they did not use the physical anchors anymore. Only 2 participants found anchors really helpful during the whole experiment, and the last 2 participants disliked them. However, no participant reported any visual discomfort.

H3 does not seem to be supported by the results. By masking the real world and therefore minimizing the visual discomfort reported in Study 1, we wanted to improve the performance of the menu techniques with mixed targets. However we did not find evidence of such an improvement. A possible explanation is that participants perfectly knew the positions of the virtual targets and did not consider physical anchors anymore. Vergence-accommodation conflicts then disappeared since participants did not focus on the physical anchors and results suggest equivalent performance between virtual and mixed targets.

8 Lessons learned and future work

This paper contributes to interaction with a menu displayed on a HMD without using hands or voice control. We establish a design space by identifying two factors: the element controlled by the user (Head-Controlled Cursor or Head-Controlled Menu) and the presence (or absence) of physical anchors at the location of the virtual targets (i.e. mixed targets). Based on the design space, we designed and evaluated three head-based techniques. In the following, we first discuss the lessons learned from the two experimental studies and then the future research axes.

In the first study, Head-Controlled Cursor (HCC) and Head-Controlled Menu (HCM) have similar performances, but HCM was preferred by most participants. While causing a significant visual intrusion, HCM allows users to directly see the menu items. This was possible thanks to a simple menu design where all menu items were visible on screen from the beginning of the task. This opens new perspectives on head-based techniques beyond moving a cursor displayed at the center of the HMD.

The first study also highlighted that mixed targets did not improve performance. To evaluate if the visual discomfort reported by some participants was related to the simultaneous visual perception of the virtual and physical components of mixed targets (i.e. vergence-accommodation conflicts), we conducted a second study with alternate menu techniques. However, temporarily masking the real-world did not impact performance. A possible explanation of the absence of effect of physical anchors is that after enough repetitions participants perfectly knew the virtual target positions. In this case, participants did not need to focus on physical anchors anymore, thus eliminating the vergence-accommodation conflicts and their impact on performance. However, we do not conclude that physically anchoring virtual targets is not useful. Our first study provided good evidence that mixed targets allowed participants to reach optimal performance faster than virtual targets. Such a result raises several follow-up research questions about the choice of physical anchors, their design and their precise impact on the different phases of the learning process [7]. Furthermore, we also plan to conduct a study in which the users will not know in advance the positions of the targets. The goal is to experimentally study the extension of the limited FOV of the HMD by relying on physical anchors.

The menu techniques were designed to be used in several possible contexts, and we purposely evaluated the techniques in a non-ecological lab setting. This is an unavoidable step to first explore the design choices. Our on-going work focuses on the augmented surgery scenario. So far, we have implemented the three techniques into the classical surgery workflow of the OrthoPilot®[1] system as part of an established partnership with the Aesculap company. We collected promising informal feedback from Aesculap engineers who have a deep knowledge of the operation workflow and of surgeons needs. Our next step is to conduct experimental studies with surgeons in pseudo ecological conditions, namely the demonstration room of Aesculap.

References

1. Aesculap: Orthopilot® navigation system, <https://www.bbraun.com/en/products-and-therapies/orthopaedic-joint-replacement/orthopilot.html/orthopilot.html>
2. Allaf, M.E., Jackman, S.V., Schulam, P.G., Cadeddu, J.A., Lee, B.R., Moore, R.G., Kavoussi, L.R.: Laparoscopic visual field. *Surgical Endoscopy* **12**(12), 1415–1418 (1998)
3. Andersen, D., Popescu, V., Cabrera, M.E., Shanghavi, A., Gomez, G., Marley, S., Mullis, B., Wachs, J.: Virtual annotations of the surgical field through an augmented reality transparent display. *The Visual Computer* **32**(11), 1481–1498 (2016)
4. Atienza, R., Blonna, R., Saludares, M.I., Casimiro, J., Fuentes, V.: Interaction techniques using head gaze for virtual reality. In: 2016 IEEE Region 10 Symposium (TENSYMP). pp. 110–114 (May 2016)
5. Azai, T., Ogawa, S., Otsuki, M., Shibata, F., Kimura, A.: Selection and manipulation methods for a menu widget on the human forearm. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 357–360. CHI EA '17, ACM, New York, NY, USA (2017)
6. Azuma, R.T.: A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* **6**(4), 355–385 (1997)
7. Bailly, G., Lecolinet, E., Nigay, L.: Visual menu techniques. *ACM Computing Surveys (CSUR)* **49**(4), 60 (2017)
8. Beaudouin-Lafon, M.: Instrumental interaction: An interaction model for designing post-wimp user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 446–453. CHI '00, ACM, New York, NY, USA (2000)
9. Bier, E.A., Stone, M.C., Pier, K., Fishkin, K., Baudel, T., Conway, M., Buxton, W., DeRose, T.: Toolglass and magic lenses: The see-through interface. In: Conference Companion on Human Factors in Computing Systems. pp. 445–446. CHI '94, ACM, New York, NY, USA (1994)
10. Buchmann, V., Violich, S., Billingham, M., Cockburn, A.: Fingertips: Gesture based direct manipulation in augmented reality. In: Proceedings of the 2Nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia. pp. 212–221. GRAPHITE '04, ACM, New York, NY, USA (2004)
11. Cassell, J.: On control, certitude, and the paranoia of surgeons. *Culture, medicine and psychiatry* **11**(2), 229–249 (1987)
12. Colaço, A., Kirmani, A., Yang, H.S., Gong, N.W., Schmandt, C., Goyal, V.K.: Mime: Compact, low power 3d gesture sensing for interaction with head mounted displays. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology. pp. 227–236. UIST '13, ACM, New York, NY, USA (2013)
13. Dacu, D., Lukosch, S.: Free-hands interaction in augmented reality. In: Proceedings of the 1st Symposium on Spatial User Interaction. pp. 33–40. SUI '13, ACM, New York, NY, USA (2013)
14. Dragicevic, P.: Fair statistical communication in hci. In: *Modern Statistical Methods for HCI*, pp. 291–330. Springer (2016)
15. Guillon, M., Leitner, F., Nigay, L.: Static voronoi-based target expansion technique for distant pointing. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces. pp. 41–48. AVI '14, ACM, New York, NY, USA (2014)

16. Guillon, M., Leitner, F., Nigay, L.: Target expansion lens: It is not the more visual feedback the better! In: Proceedings of the International Working Conference on Advanced Visual Interfaces. pp. 52–59. AVI '16, ACM, New York, NY, USA (2016)
17. Hatscher, B., Luz, M., Nacke, L.E., Elkmann, N., Müller, V., Hansen, C.: Gaze-tap: towards hands-free interaction in the operating room. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction. pp. 243–251. ACM (2017)
18. He, Z., Yang, X.: 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. pp. 227–230. VRCAI '14, ACM, New York, NY, USA (2014)
19. Henderson, S.J., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret (2009)
20. Henze, N., Boll, S.: Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. pp. 191–194. MobileHCI '10, ACM, New York, NY, USA (2010)
21. Hirsch, H.G., Pearce, D.: The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW) (2000)
22. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* **8**(3), 33 (2008). <https://doi.org/10.1167/8.3.33>
23. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9. Standard, International Organization for Standardization (2000)
24. Jacob, R.J.: What you look at is what you get: eye movement-based interaction techniques. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 11–18. ACM (1990)
25. Jagacinski, R.J., Monk, D.L.: Fitts law in two dimensions with hand and head movements. *Journal of motor behavior* **17**(1), 77–95 (1985)
26. Jalaliniya, S., Mardanbeigi, D., Pederson, T., Hansen, D.W.: Head and eye movement as pointing modalities for eyewear computers. In: Wearable and Implantable Body Sensor Networks Workshops (BSN Workshops), 2014 11th International Conference on. pp. 50–53. IEEE (2014)
27. Katic, D., Wekerle, A.L., Gortler, J., Spengler, P., Bodenstedt, S., Rohl, S., Suwelack, S., Kenngott, H.G., Wagner, M., Muller-Stich, B.P., Dillmann, R., Speidel, S.: Context-aware augmented reality in laparoscopic surgery. *Computerized Medical Imaging and Graphics* **37**(2), 174 – 182 (2013)
28. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 2010 IEEE International Symposium on Mixed and Augmented Reality. pp. 3–12 (Oct 2010)
29. Kytö, M., Ens, B., Piumsomboon, T., Lee, G.A., Billinghamurst, M.: Pinpointing: Precise head- and eye-based target selection for augmented reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 81:1–81:14. CHI '18, ACM, New York, NY, USA (2018)
30. Lin, M.L., Radwin, R.G., Vanderheiden, G.C.: Gain effects on performance using a head-controlled computer input device. *Ergonomics* **35**(2), 159–175 (1992)
31. Louis, T., Berard, F.: Superiority of a handheld perspective-coupled display in isomorphic docking performances. In: Proceedings of the 2017 ACM International

- Conference on Interactive Surfaces and Spaces. pp. 72–81. ACM, New York, NY, USA (2017)
32. Mahmoud, N., Grasa, Ó.G., Nicolau, S.A., Doignon, C., Soler, L., Marescaux, J., Montiel, J.M.M.: On-patient see-through augmented reality based on visual slam. *International Journal of Computer Assisted Radiology and Surgery* **12**(1), 1–11 (2017)
 33. Mardanbegi, D., Hansen, D.W., Pederson, T.: Eye-based head gestures. In: *Proceedings of the symposium on eye tracking research and applications*. pp. 139–146. ACM (2012)
 34. Mentis, H.M., O’Hara, K., Gonzalez, G., Sellen, A., Corish, R., Criminisi, A., Trivedi, R., Theodore, P.: Voice or gesture in the operating room. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. pp. 773–780. CHI EA ’15, ACM, New York, NY, USA (2015)
 35. Mitrasinovic, S., Camacho, E., Trivedi, N., Logan, J., Campbell, C., Zilinyi, R., Lieber, B., Bruce, E., Taylor, B., Martineau, D., et al.: Clinical and surgical applications of smart glasses. *Technology and Health Care* **23**(4), 381–401 (2015)
 36. Neumann, U., Majoros, A.: Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In: *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*. pp. 4–11. IEEE (1998)
 37. O’Hara, K., Gonzalez, G., Sellen, A., Penney, G., Varnavas, A., Mentis, H., Criminisi, A., Corish, R., Rouncefield, M., Dastur, N., Carrell, T.: Touchless interaction in surgery. *Commun. ACM* **57**(1), 70–77 (Jan 2014)
 38. Ong, S., Yuan, M., Nee, A.: Augmented reality applications in manufacturing: a survey. *International journal of production research* **46**(10), 2707–2742 (2008)
 39. Özacar, K., Hincapié-Ramos, J.D., Takashima, K., Kitamura, Y.: 3d selection techniques for mobile augmented reality head-mounted displays. *Interacting with Computers* **29**(4), 579–591 (2016)
 40. Qian, Y.Y., Teather, R.J.: The eyes don’t have it: An empirical comparison of head-based and eye-based selection in virtual reality. In: *Proceedings of the 5th Symposium on Spatial User Interaction*. pp. 91–98. SUI ’17, ACM, New York, NY, USA (2017)
 41. Reiners, D., Stricker, D., Klinker, G., Müller, S.: Augmented reality for construction tasks: Doorlock assembly. *Proc. IEEE and ACM IWAR* **98**(1), 31–46 (1998)
 42. Rovers, A., Van Essen, H.: Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles. *Virtual Reality* **9**(2-3), 177–191 (2006)
 43. Scarr, J., Cockburn, A., Gutwin, C., Quinn, P.: Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In: *Proceedings of the sigchi conference on human factors in computing systems*. pp. 2741–2750. ACM (2011)
 44. Sibert, L.E., Jacob, R.J.: Evaluation of eye gaze interaction. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. pp. 281–288. ACM (2000)
 45. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of fitts law research in hci. *International Journal of Human-Computer Studies* **61**(6), 751 – 789 (2004), fitts’ law 50 years later: applications and contributions from human-computer interaction
 46. Špakov, O., Majaranta, P.: Enhanced gaze interaction using simple head gestures. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. pp. 705–710. ACM (2012)

47. Ullmer, B., Sankaran, R., Jandhyala, S., Tregre, B., Toole, C., Kallakuri, K., Laan, C., Hess, M., Harhad, F., Wiggins, U., et al.: Tangible menus and interaction trays: core tangibles for common physical/digital activities. In: Proceedings of the 2nd international conference on Tangible and embedded interaction. pp. 209–212. TEI '08, ACM (2008)
48. White, S., Feng, D., Feiner, S.: Interaction and presentation techniques for shake menus in tangible augmented reality. In: Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality. pp. 39–48. ISMAR '09, IEEE Computer Society, Washington, DC, USA (2009)
49. Wingrave, C., Bowman, D.: Baseline factors for raycasting selection. In: Proceedings of HCI International. Citeseer (2005)
50. Yan, Y., Yu, C., Yi, X., Shi, Y.: Headgesture: Hands-free input approach leveraging head movements for hmd devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **2**(4), 198:1–198:23 (Dec 2018)
51. Yi, S., Qin, Z., Novak, E., Yin, Y., Li, Q.: Glassgesture: Exploring head gesture interface of smart glasses. In: Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on. pp. 1–9. IEEE (2016)
52. Yuan, M., Ong, S., Nee, A.: Augmented reality for assembly guidance using a virtual interactive tool. International Journal of Production Research **46**(7), 1745–1767 (2008)