Tabletop AR with HMD and Tablet: A Comparative Study for 3D Selection

Carole Plasson
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
F-38000 Grenoble France
carole.plasson@univ-grenoble-alpes.fr

Yann Laurillau
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
F-38000 Grenoble France
yann.laurillau@univ-grenoble-alpes.fr

Dominique Cunin
Ecole Superieure d'Art et Design Grenoble-Valence
Unite de recherche
F-26000 Valence, France
dominique.cunin@esad-gv.fr

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

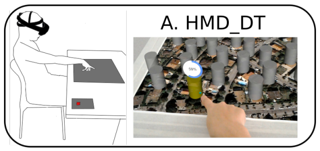
ISS '19, November 10–13, 2019, Deajon, Republic of Korea.

ACM ISBN 978-1-4503-6891-9/19/11.
http://dx.doi.org/10.1145/3343055.3360760

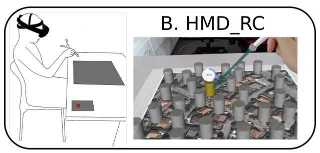
Abstract
We experimentally compare the performance and usability of tablet-based and see-through head-mounted display (HMD)-based interaction techniques for selecting 3D virtual objects projected on a table. This study is a first step toward a better understanding of the advantages and limitations of these interaction techniques, with the perspective of improving interaction with augmented maps. To this end, we evaluate the performance of 3 interaction techniques in selecting 3D virtual objects in sparse and dense environments: (1) the direct touch interaction with a HMD; (2) the ray-casting interaction with a HMD; and (3) the touch interaction on a tablet. Our results show that the two techniques using a HMD are faster, less physically tiring and preferred by the participants over the tablet. The HMD-based interaction techniques perform equally well but the direct touch technique seems to be less impacted by small targets and occlusion.

Author Keywords
Augmented Reality; Augmented Maps; HMD; Tablet; Tabletop; Touch; Comparative study

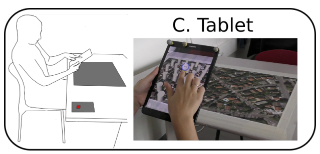
CCS Concepts
•Human-centered computing → Human computer interaction (HCI); Mixed / augmented reality;



(a) HMD direct touch (close interaction)



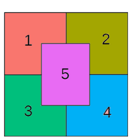
(b) HMD ray-casting (far interaction)



(c) Tablet touch (far interaction)

Figure 1: Comparison of three techniques

(a) Terrain example



(b) Zones

Figure 2: Layout of terrains

Introduction: Goal and Motivation

Tabletop Augmented Reality (AR) systems and especially augmented maps combine a tangible physical surface and virtual objects. Handheld devices (tablets/smartphones) or HMDs are used to implement such systems. For instance, handheld devices are used to explore interior decoration over a floor plan [6] or to visualize 3D buildings over a 2D map ([7], Ullmer & Ishii's *Metadesk*) in urban planning systems. In addition to exploration tasks, ARTopos [9] enables annotation of the 3D augmentation of the 2D map by touching the screen of the smartphone. Other systems use a HMD to augment a 2D map with 3D geographic information. To interact with the augmented content, users directly touch it using freehand [4] or by holding a physical object as a pointing device (e.g., a stick, a pen or a paddle [2, 3]).
Our goal is to answer the question: Which device/technique (i.e. tablet or HMD) is better and preferred to the other for these systems. But very few comparative studies have been conducted. Recently Bach et al. [1] compared visualisation tasks on the desktop or on two AR devices (a tablet and an HoloLens HMD) based on tangible markers interaction. The results showed that desktop performed well for all studied tasks, but the AR technique using the HMD showed good performance for "tasks that can be solved through spatial perception and interactions with a high degree of freedom". For tabletop AR, Looser et al. [5] compared 3 HMD-based techniques for 3D selections: a virtual hand, a virtual pointer, and a magic lens technique. The results showed that the lens technique was faster and preferred over other techniques. However, the authors explained that the virtual hand and virtual pointer techniques were impacted by the lack of stereoscopic vision, biasing the experiment towards the magic lens technique.
Considering the currently available devices and platforms for AR, the question of the best and preferred device/technique remains.

Three techniques to compare for 3D selection

Towards our goal, we conducted two comparative studies to investigate the performance and usability of HMDs and tablets by comparing 3 interaction techniques. From the literature review, we indeed identify 3 commonly used interaction techniques (see Figure 1) that we classify along two axes: the device and the proximity (in motor space) of the user from virtual objects. These techniques are:
- HMD_DT. HMD with direct touch (close interaction). An augmented sphere follows users' fingertip and selection occurs when the sphere collides with a cylinder;
- HMD_RC. HMD with ray-casting (far interaction). Users hold a stick extended by a virtual laser (its width is equal to the sphere's diameter of HMD_DT) and selection occurs when it collides with a cylinder. We choose the stick instead of the finger for stability reasons.
- Tablet. A tablet is used to visualize the 3D scene. Users directly touch the screen to select a cylinder (far interaction).
For the two devices, we use a Microsoft HoloLens 1 HMD and a Samsung Galaxy Tab S3 tablet (9.7" display with a resolution of 2048x1536 and weights 429g). For implementing the 3 techniques, we use Unity to create and manage the 3D scene and Vuforia to track the image of the residential area stuck to the table. The Optitrack technology (high-speed infrared cameras) and reflective markers are used to track the stick and the user's fingertip.
As the majority of prior work is about map exploration or selection of points of interest, we focus on these fundamental tasks of search and selection of 3D virtual objects in sparse and dense environments in the two following experiments.

First Experiment: Search and Selection

The experiment is designed to distinguish 2 phases of target selection: the search and the selection phases. Indeed, the 2 devices differ on users' perception of a 3D scene; this

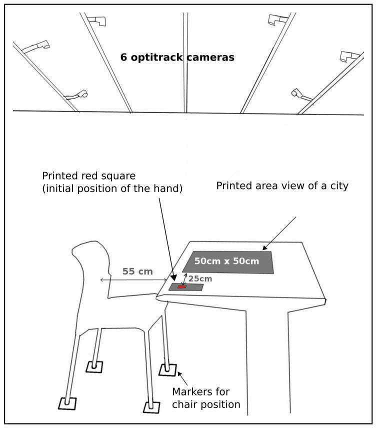
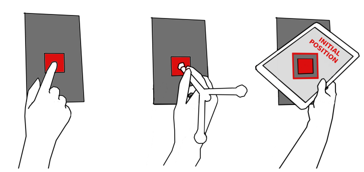


Figure 3: Setup of the experiment



(a) HMD_DT **(b)** HMD_RC **(c)** Tablet

Figure 4: Initial positions

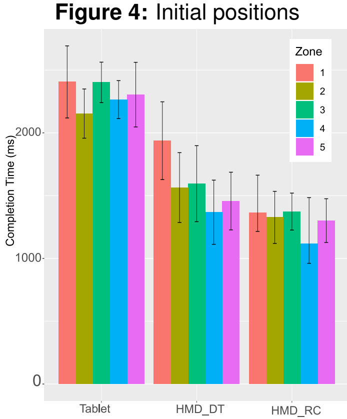


Figure 5: Completion time: zoning effect for selection tasks

difference may have an impact on the time to visually locate a target in the scene. The goals of this experiment are to compare the techniques for each of the 2 phases and compare our results with those of the similar study [5]. We recruited 12 volunteers from our lab, aged 24 to 40 years old ($mean = 31.4$, $sd = 5.78$). All participants had a previous experience in AR, especially with a HMD, but none of them were experts.

Terrain Generation and Setup

Thirty cylinders (1 red target and 29 gray distractors) are displayed on the table over a printed image (Figure 2a). They measure 5cm in height with a diameter of 3.5cm. We chose cylinders as targets because their widths are always the same for all lines of approach. As distance impacts movement time, we divide the terrain into 5 (invisible) zones (Figure 2b). The cylinders are pseudo-randomly placed to avoid occlusion and to obtain 6 cylinders in each zone. We generated 15 terrains allowing 3 selection tasks inside each of the 5 zones. Figure 3 shows the setup of the experiment. The position of the chair allows participants to reach all the zones at hand but they have to lean forward to reach the farthest zones 1 & 2.

Experimental Design and Procedure

The experiment is designed as a within-subject user study with 2 independent variables (*Zone*, *Technique*). The techniques order is counterbalanced with a latin square and the terrains are presented in a random order. The experiment lasts approximately 40 minutes.

We start with a short introduction of the experiment. As the execution task is separated into 2 phases, participants have to return to the initial position (see Figure 4) before each phase. First, they have to find the target as quickly as possible and to say "okay". Then, they select the same target. They perform these phases for 15 terrains plus 3 training terrains for each technique. At the end of each technique, participants fill a questionnaire inspired by NASA-TLX (same questionnaire of [5] to compare our results with theirs), with additional questions about participants' level of fatigue (4-point Likert scale) for several body areas. We conclude with a final questionnaire to compare the techniques (preference and fatigue) and an interview.

Results

As data are not normally distributed (Shapiro-Wilk test), we use a non-parametric Friedman χ^2 test and a Wilcoxon signed rank test with Bonferroni adjustment for pairwise comparisons. In the graphs, we represent the mean values and the 95% confident intervals computed with Bca bootstrap method (mean is the bootstrap function).

We find a statistically significant difference between the completion times of the two devices for search [$\chi^2 = 18$, $df = 2$, $p = 0.00012$] and selection [$\chi^2 = 20.667$, $df = 2$, $p < 0.00001$] tasks. Pairwise comparisons show that the HMD-based techniques are faster than the tablet for both phases ($p = 0.0015$).

For the selection, see Figure 5, we find that *Technique* has a significant effect on the completion times (for all zones, Friedman $p < 0.0005$). For all zones, HMD_DT and HMD_RC are faster than the tablet ($p < 0.0029$). The two HMD-based techniques perform equally well except for the farthest zone (HMD_RC is faster in this zone, $p = 0.0146$). Contrary to the distant interaction techniques, we find that HMD_DT seems to be impacted by the zones [$\chi^2 = 23.2$, $df = 4$, $p = 0.00012$].

From our qualitative evaluation, we find that HMD-based techniques (and especially HMD_RC) are preferred and less tiring than tablet interaction. They rank the interaction techniques as follows: 5/12 participants preferred HMD_RC, 5/12 HMD_DT, and 2/12 Tablet. According to participants,

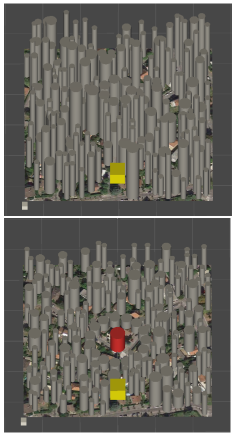


Figure 6: Example of a terrain with a visible/occluded target from the starting point of view

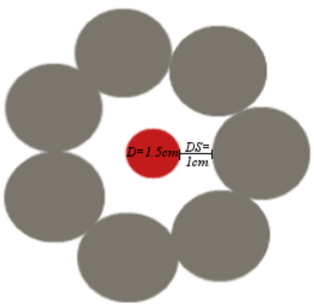


Figure 7: Distractors around a target; with $DS = 1cm$, $S = small$

holding a tablet is more physically demanding than performing large movements with HMD_DT and wearing a HMD.

Second Experiment: Occlusion

When we use the augmented map of a dense city, several buildings can be hidden behind others and thus, it can be difficult to select them. This experiment studies the impact of dense environments and occluded targets on the performance of a selection task (no search task). Our protocol is based on the protocol of Vanacken et al. [8].
We recruited the 12 volunteers from our previous study. They had already used the three interaction techniques which reduced the WoW effect and the learning phase. We use the same setup as in the first experiment, but we remove the chair and place a line on the ground at 70cm from the table indicating the starting position for the task. At this position, the occluded targets are not visible and participants have to move forward to see them.

Terrain Generation
A terrain is composed of 200 distractors and 2 targets: the initial target is a yellow cube and the second target is a red cylinder (see Figure 6). The red cylinder is always placed at a distance of 17cm from the cube (target's position already studied in the first experiment). The cube is always visible and at the same location on the terrain. We consider three independent variables:
- *Visibility Condition (V):* The red target is visible or occluded for a participant standing at 70cm from the table (starting point), see Figure 6.
- *Target size (S):* The diameter of the target (6cm high) is either 1.5cm (small target) or 4cm (large target).
- *Density Spacing (DS):* In a terrain, several distractors are positioned circularly around the red target (Figure 7), at a distance of 1cm (close to) or 4cm (far from) to the target. For occluded targets, these distractors measure 18cm high. Otherwise, they measure 3cm ($target_{Height}/2$). The other distractors are randomly placed.
For this experiment, we have 2(V) $\times$ 2(S) $\times$ 2(DS) $\times$ 3(repetitions) = 24 terrains per technique.

Experimental Design and Procedure
The experiment is designed as a within-subject user study, with the 4 independent variables (*V, S, DS, Technique*). The techniques order is counterbalanced with a latin square and the terrains are presented in a random order. The experiment lasts approximately 50 minutes.

We start with a short introduction of the experiment. After a training phase, participants perform 24 selections per technique. They execute the task as follow: (1) they stand behind the line and in front of the table; (2) during 3.5s, they only look at the 3D scene containing the yellow cube and the red target; (3) during 4s, the distractors grow up until they reach their final size; (4) participants can select the cube when it turns from yellow to green; (5) they then select the red cylinder. During the steps 2 & 3, participants can easily find out targets' positions. These steps allow us to eliminate the search phase (already studied in the first experiment). Participants are free to move after step 2. They fill the same questionnaire of the first experiment for both visibility conditions after each technique and fill a comparative final questionnaire at the end of the experiment.

Results
For the analysis, we report the results for 11 participants as we removed an outlier (with major differences). As they were free to move after step 2 (some of them moved the table forward during step 3), we measured the time between step 2 & 5 (selection of the red target). As data sets are modeled by a normal distribution, we used ANOVAs and t-tests with Bonferroni adjustment for pairwise comparisons.

We find a main effect of *Technique* [$F_{2,20} = 19.96$, $p < 0.0001$],

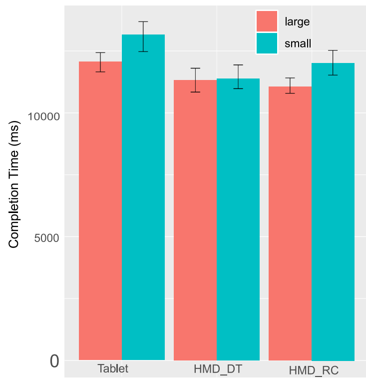


Figure 8: [Completion Time] Effect of the target size

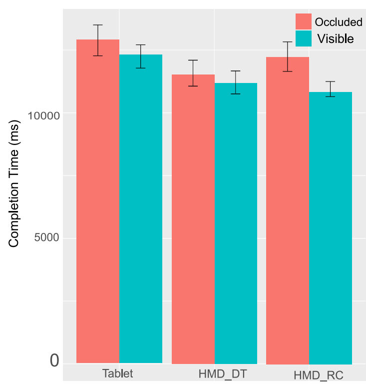


Figure 9: [Completion Time] Effect of the visibility condition

the target size *S* [$F_{1,10} = 80.1, p < 0.0001$] and the visibility condition *V* [$F_{1,10} = 39.65, p < 0.0001$]. Overall, HMD is faster than the tablet ($p < 0.0156$), regardless of V and S. The completion times seem not to be affected by the density spacing *DS*.

We find that *Technique* has significant interaction effect with *S* [$F_{2,50} = 6.1326, p = 0.0042$] (data preprocessed using an ART), see Figure 8. HMD_DT seems to be not affected by the target size *S*, contrary to the distant interaction techniques ($p < 0.011$). This may be one advantage of HMD_DT over HMD_RC.

We find that *Technique* has a significant interaction effect with *V* [$F_{2,50} = 5.7921, p = 0.005464$] (data preprocessed using an ART), Fig 9. We find a statistically significant effect of *V* only for HMD_RC (Post-hoc analysis $p = 0.00098$). We find no significant effect of *V* for the others.

For qualitative results, participants find it easy to select the target with all the techniques, but the HMD is better perceived than Tablet for both visibility conditions. We note that the three techniques have a lower ranking for the occlusion condition than with visible condition. HMD_DT is the most preferred, easiest, fastest and accurate technique.

Discussion: Key results

HMD faster than and preferred over Tablet. With our first study, we find the search and selection tasks are achieved faster with the HMD-based interaction than with the tablet. Indeed, HMDs facilitate uncoupled and synchronized head/arms movements, and users only need head movements for search tasks. Participants also largely prefer the HMD and explain that wearing it is less tiring than holding a tablet. Contrary to the results of [5], both HMD-based techniques have good scores. They seem better perceived than the techniques of 10 years ago. These new results open up new prospects for the HMD in tabletop AR interfaces.

	Benefits	*Limitations*
Tablet	• Perceived accuracy, easiness, and rapidity • Low occlusion effect on completion time • Low zone effect on completion time	• Slowest interaction • Most tiring • Least preferred • Perceived as least easiest, fastest, and accurate • Size effect on completion time
HMD	• Fastest interaction • Perceived accuracy, easiness, and rapidity; Less tiring • Most preferred, contrary to [5]	• Zone effect for search task (close objects on left side)
HMD_DT	• Perceived as most easiest and fastest (hidden target) • Perceived as most accurate • Low occlusion and size effects on completion time • Elicit interaction strategies	• Slower than HMD_RC for selecting distant virtual objects (left side) • Zone effect for selection task (distant vs. central and close objects)
HMD_RC	• Faster for distant virtual objects (left side) • Low zone effect	• Occlusion and size effects on completion time (see [8]) • Perceived as least accurate

Table 1: Benefits and limitations of the techniques

HMD_DT most preferred, perceived as easiest and fastest. In our second study, the density spacing variable has no effect on completion time, which confirms the observation of [5]. Not studied in [5], we also considered the visibility condition (i.e. target visible or occluded). Unlike other techniques, HMD_RC is significantly affected by the occlusion of the target. This is confirmed by the study conducted in [8] in order to improve the ray-casting technique for the case of occlusion. This suggests that HMD_RC should be replaced by another type of ray-casting technique as proposed by [8] or by a technique like HMD_DT which is perceived to be direct, natural and more robust to occlusion and target size than HMD_RC.

Strategies with obstacles. By analyzing the movements of users' fingertip (HMD_DT condition) in dense environments in the second experiment, three different strategies emerge for selecting the red target hidden behind high obstacles:
- Going through obstacles to directly reach the red target: 9/11 participants used this strategy;
- Moving above obstacles to reach the red target: We observe a jump movement between the starting target and the end target for occlusion conditions. The height of the jump is more than 20cm which corresponds to obstacles' height (18cm). 1/11 participant used this strategy;
- 1/11 participant both used the two previous strategies.

Mixed mid-air/touch strategies. We additionally observe two users' strategies to achieve the selection task with HMD_DT in the first experiment. We were expecting participants to place their finger on the table and not in mid-air in order to manage only 2 of the 3 dimensions. However, only 4/12 participants always touched the table during this task while 7/12 participants always selected the target in mid-air (1 participant used both strategies). This suggests that tabletop AR interfaces based on direct touch with a HMD should be designed to be flexible enough to take into account these strategies.

Acknowledgements

We gratefully acknowledge the support of the AP2 project ANR-15-CE23-0001.

REFERENCES

[1] Benjamin Bach et al. 2018. The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *IEEE TVCG* (Jan 2018), 457–467. DOI: http://dx.doi.org/10.1109/TVCG.2017.2745941

[2] Joachim Bobrich. 2003. An Immersive Environment Based on Paper Maps. In *Proc. ICC*.

[3] Nicholas R. Hedley et al. 2002. Explorations in the Use of Augmented Reality for Geographic Visualization. *Presence: Teleop. VE* 11, 2 (April 2002), 119–133. DOI:http://dx.doi.org/10.1162/1054746021470577

[4] Seungjun Kim et al. 2006. A Reliable New 2-Stage Distributed Interactive TGS System Based on GIS Database and Augmented Reality. *IEICE Trans. Inf. Syst.* E89-D, 1 (Jan. 2006), 98–105. DOI: http://dx.doi.org/10.1093/ietisy/e89-d.1.98

[5] Julian Looser et al. 2007. An Evaluation of Virtual Lenses for Object Selection in Augmented Reality *(GRAPHITE '07)*. ACM, New York, NY, USA, 203–210. DOI:http://dx.doi.org/10.1145/1321261.1321297

[6] Alex Olwal and Anders Henrysson. 2007. Lumar: A hybrid spatial display system for 2d and 3d handheld augmented reality. In *ICAT*. IEEE, 63–70.

[7] Julian Seifert et al. 2014. Hover Pad: Interacting with Autonomous and Self-actuated Displays in Space *(UIST '14)*. ACM, NY, USA, 139–147. DOI: http://dx.doi.org/10.1145/2642918.2647385

[8] Lode Vanacken et al. 2009. Multimodal Selection Techniques for Dense and Occluded 3D Virtual Environments. *Int. J. Hum.-Comput. Stud.* 67, 3 (March 2009), 237–255. DOI: http://dx.doi.org/10.1016/j.ijhcs.2008.09.001

[9] Frederik Wiehr et al. 2017. ARTopos: Augmented Reality Terrain Map Visualization for Collaborative Route Planning *(UbiComp '17)*. ACM, NY, USA, 1047–1050. DOI: http://dx.doi.org/10.1145/3123024.3124446