# Eye vs. Head: Comparing Gaze Methods for Interaction in Augmented Reality

Nelusa Pathmanathan
st141156@stud.uni-stuttgart.de
University of Stuttgart
70569 Stuttgart, Germany

Michael Becher
Nils Rodrigues
Guido Reina
{first}.{last}@visus.uni-stuttgart.de
University of Stuttgart
70569 Stuttgart, Germany

Thomas Ertl
Daniel Weiskopf
Michael Sedlmair
{first}.{last}@visus.uni-stuttgart.de
University of Stuttgart
70569 Stuttgart, Germany

## ABSTRACT

Visualization in virtual 3D environments can provide a natural way for users to explore data. Often, arm and short head movements are required for interaction in augmented reality, which can be tiring and strenuous though. In an effort toward more user-friendly interaction, we developed a prototype that allows users to manipulate virtual objects using a combination of eye gaze and an external clicker device. Using this prototype, we performed a user study comparing four different input methods of which head gaze plus clicker was preferred by most participants.

## CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **HCI design and evaluation method →** *User studies*; • **Interaction paradigms →** *Mixed/augmented reality*.

## KEYWORDS

Immersive analytics, visualization, augmented reality, eye tracking, interaction

**ACM Reference Format:**
Nelusa Pathmanathan, Michael Becher, Nils Rodrigues, Guido Reina, Thomas Ertl, Daniel Weiskopf, and Michael Sedlmair. 2020. Eye vs. Head: Comparing Gaze Methods for Interaction in Augmented Reality. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3379156.3391829

## 1 INTRODUCTION

Immersive analytics comes with a novel design space that leverages virtual reality (VR) and augmented reality (AR) for visualizing data [Dwyer et al. 2018]. This trend raises the question of how interaction can be realized in such environments. A natural choice is to adopt existing VR/AR interaction approaches. For instance, Microsoft HoloLens—an optical see-through and head-mounted display (HMD) for AR—is designed to be used via head gaze. For that, a cursor located at the center of the display is linked directly to head movements. Hand gestures or an external clicker device allow the user then to select virtual objects that the cursor is pointed at. While simple head movements are usually not very strenuous, the constant need to move the head in order to point at virtual objects can lead to fatigue. The same can hold true for the use of hand gestures over longer periods of time. Our work explores the use of eye tracking for interaction in AR. Previous work has suggested that leveraging eye movements might be a comfortable and natural alternative in such situations [Sibert and Jacob 2000]. Since the eyes move anyways during interaction with a computer, interaction using the gaze usually does not pose a great challenge for users [Drewes 2010].

Our contribution is two-fold. First, we developed a prototype software for an optical see-through HMD (Microsoft HoloLens) that allows for object manipulation with eye gaze and an external clicker device. Second, we performed a user study to evaluate the feasibility of object manipulation in combination with pointing by eye gaze and head gaze as well as interaction by gesture and clicker device.

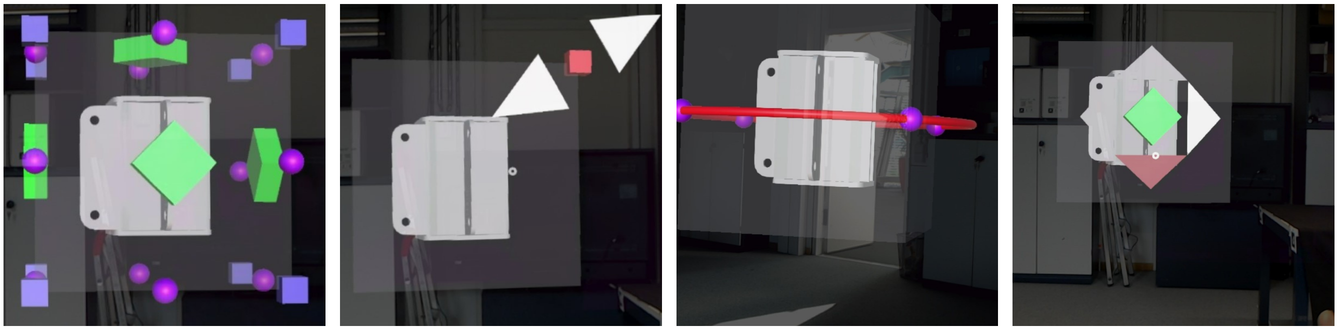
## 2 RELATED WORK

Visual analytics is often used in conjunction with 2D applications. Flat visualizations have also been the target of research that aims at using eye tracking as a method for interaction. For instance, gaze can be used to select targets, which can then manipulated by touch on a display [Pfeuffer et al. 2014]. It then does not matter on what part of the display the touch input occurred, as it is the gaze that defines the target. Selection is also possible by matching the movement path of objects on a display to the gaze trajectory [Vidal et al. 2013]. Assuming that the eyes follow the object of interest, trajectories of different objects in the field of view are compared with the gaze trajectory and if a match is found, the object is recognized as being of interest to the user.

Immersive analytics encompasses analytical reasoning, decision making, and collaborative work [Dwyer et al. 2018] and might require interaction in 3D spaces. It is possible for users to interact with an AR environment solely with their eyes by integrating an eye tracker in an AR system and using dwell-time [Lee et al. 2011]. By designing a 3D user interface around gaze interaction and using a VR HMD with integrated eye tracking, such as the FOVE, different spatial manipulations of virtual objects can be supported [Groß et al. 2019]. While inspired by Groß et al.'s approach, our design improves on several aspects: UI elements (e.g., for translation) are more accessible from arbitrary user positions, usage of different


*ETRA '20 Short Papers, June 2–5, 2020, Stuttgart, Germany*

ACM ISBN 978-1-4503-7134-6/20/06…$15.00
https://doi.org/10.1145/3379156.3391829



(a) Interaction handles default state (b) Scaling (c) Rotation (d) Translation

**Figure 1: Tool handles placed on the bounding box and the three transformation modes with their individual sub-handles.**

transformations is more consistent, and a stricter separation between tool activation and object manipulation prevents accidental manipulations. Also, we re-evaluate the approach for AR with an aftermarket eye tracking setup for HoloLens 1 that is ergonomically more challenging and error-prone than the permanently installed eye tracker of the FOVE.

The FOVE was also used in a study to compare the performance of head gaze and eye gaze for selection tasks [Qian and Teather 2017]. Contrary to their expectations, the results showed poor performance in terms of error rate, selection time, and throughput for eye-gaze-based selection. However, it was also mentioned that the poor eye tracking quality is a possible cause for this result. Another study investigating interaction with head gaze and eye gaze in AR and VR shows that eye gaze can surpass head gaze in speed, workload, required head movement, and user preference [Blattgerste et al. 2018]. In addition, the benefits for eye-gaze-based interaction increased with the field of view. For that study, the HTC Vive was used to simulate an AR display with configurable field of view (FoV). Eye tracking was performed using the SMI HTC Vive Integration Scientific Premium eye tracker. In our work, instead of relying on a video see-through HMD, we use an optical see-through HMD (Microsoft HoloLens) to explore the use of eye tracking in AR. Moreover, our prototype does not only investigate selection but also object manipulations that require several consecutive interactions using eye gaze.

The use of VR and AR in a broader range of applications, especially combined with user interaction based on eye tracking, raises the question of eye fatigue caused by the use of HMDs. Eye tracking itself can be used to measure the eye fatigue introduced by HMDs [Wang et al. 2019].

## 3 PROTOTYPE SYSTEM

To explore whether eye tracking is a feasible input method for AR, we developed a prototype system that features basic 3D object manipulation via eye gaze for an optical see-through AR device. We chose to implement tools for well-known spatial transformations (translate, rotate, scale) that can be used to interact with arbitrary objects in a 3D environment, e.g., interacting with parts from a CAD model within an immersive analytics setting.

Our assumption is that the interaction with eye gaze provides a convenient method of interaction. It is possible to interact with different objects and visualizations for a longer time without straining the user. The main challenge in providing an eye-gaze-based interaction is given by the Midas touch problem, that is, the challenge to differentiate natural gaze from the intention to interact [Jacob 1990]. To circumvent this problem, gaze is solely used for aiming purposes and a clicker device for the selection of the targeted objects.

To manipulate objects in a virtual scene, we equip them with handles on their bounding box (see Figure 1a): rhombuses on each side for translation, spheres on each edge for rotation, and cubes on each vertex for scaling. These handles have different states, indicated by different colors for visual feedback. Every handle starts in its *default* state, which does not have a specific color. When the user's gaze lands on a handle, it changes into the *hover* state and turns yellow. Pressing the HoloLens clicker *activates* the hovered handle (it turns red) and enables manipulation of the associated object. At the same time, all other handles are *disabled*, turning them invisible. To return to the initial state of the tool, as shown in Figure 1a, handles must be deactivated (hover+click) again.

An active handle will be surrounded by more specific sub-handles that are unique to each transformation (see Figures 1b–1d). These sub-handles are used for the actual manipulation of the associated object and are not persistent across state changes. To avoid the Midas touch problem, we designed all sub-handles to go through a two-step activation process before any object manipulations can be performed: Sub-handles must first be activated and then gazed at and clicked simultaneously. Figures 1b–1d show the design of the transformation tool for each of the following manipulation modes:

*Scaling.* Two arrows placed around the active scaling handle (pointing toward and away from the object) allow the user to decrease or increase the object's scale by a fixed amount. We chose adjusting scale in discrete steps instead of a continuous manipulation following recommendations of previous work [Groß et al. 2019].

*Rotation.* Rotation uses ring shaped sub-handles oriented orthogonal to one of the three local main axes. Depending on which rotation handle is active, only one of three rings is shown. The user needs to click two points on the ring to rotate the object around the axis orthogonal to the ring by the angle between the two points.

*Translation.* Four arrows placed around the active translation handle allow the user to move the object along two of the local main axes. As the axes change depending on which handle is active, it is possible to move an object along all three main axes.

In addition to the basic transformation tools, our prototype provides alternative rotation and translation modes. *Free rotation* is modeled after the equally named feature by Groß et al. [2019]. It allows users to rotate the object so that the surface point they gaze at is rotated toward the center of the object. The speed of rotation is proportional to the distance of the gaze point from the center. Consequently, if the user follows the direction of rotation with their eyes, i.e., keeps a specific surface point in focus, the rotation slows down. Once the gaze falls within a small *dead zone* around the center, which is visually communicated to the user by a small circle, the rotation stops completely.

*Free translation* allows users to move the object by larger distances with a single interaction. The three local main coordinate axes are represented by auxiliary lines placed through the center of the object. If the user gazes at a point on a line and then clicks, the object is immediately moved to that location.

## 4 USER STUDY

We conducted a study testing the usability of the different functionalities of the prototype. The aim of this study was to compare the following four input methods using the different manipulation functionalities available within the prototype:

- **Head-Gesture:** Head gaze and the air tap gesture (default HoloLens input modality)
- **Head-Clicker:** Head gaze and the dedicated clicker device
- **Eye-Cursor:** Eye gaze with the clicker device and a visible cursor
- **Eye-NoCursor:** Eye gaze with the clicker device but no visible cursor

We propose the following hypotheses:

- **H1:** Eye tracking is the most convenient way to control objects.
- **H2:** The clicker device is the most convenient way to select objects.
- **H3:** The visual feedback from virtual objects is easy to understand.
- **H4:** Users prefer *free rotation* to the rotation with the transformation tool.

*Study Design.* We conducted the study with 12 (3 female) participants. The age of the participants ranged from 18 to 23 years. Eleven participants already had prior experience with AR and VR, but only two were working with immersive technologies on a weekly basis. We chose a within-subject study design in which the participants had to use the four different input methods in four consecutive trials. The order of the trials was randomized for each participant to counter-balance learning effects. The different input methods, as well as the different tool functions, are our independent variables. The dependent variable is the user experience of the participants. Moreover, we consider the poor quality of eye tracking to be a potential confounding variable.

*Setup.* The study was conducted with a Microsoft HoloLens 1 device, and a modified Pupil Labs binocular eye tracking add-on. The right eye camera broke before the start of the study, and was replaced with a compatible part from a Pupil Core eye tracker. At the beginning of the trial, we asked participants to calibrate the HoloLens device. For the input methods using eye gaze, a calibration of the eye tracker was required as well. During the study, we observed that even with a good initial calibration of the eye tracker the reliability and precision of the eye tracking deteriorated over time. We suspect that this happens because the HoloLens position shifts over time, even when fitted very tightly to the user's head. In our study, the tool bounding box was 0.5 m in size and only a single object was shown at the same time. The object was placed at a distance of 1 m from the users, which allowed for all handles to remain visible in HoloLens' limited FoV (ca. $30° \times 17.5°$). The interaction handles can be selected with some tolerance, i.e., the area covered by the invisible colliders for gaze intersection is larger than the visible handles. During the study, the colliders covered around 5° of the FoV, while the eye tracker is accurate to about 1°.

For each input method, the participants performed the following tasks: Loading a single object into the scene (using a standard VR/AR floating menu window), carrying out *free rotation* on an object in the scene, and carrying out the different manipulations with the transformation tool. After each trial, the participants filled out a questionnaire in which they were asked to rate the different tasks on a Likert scale ranging from 1 to 5. After all trials were performed, the participants were asked to fill out another questionnaire in which they should rate the different input methods as well as the different functions of the prototype in general. Again, the rating was based on a Likert scale ranging from 1 to 5.

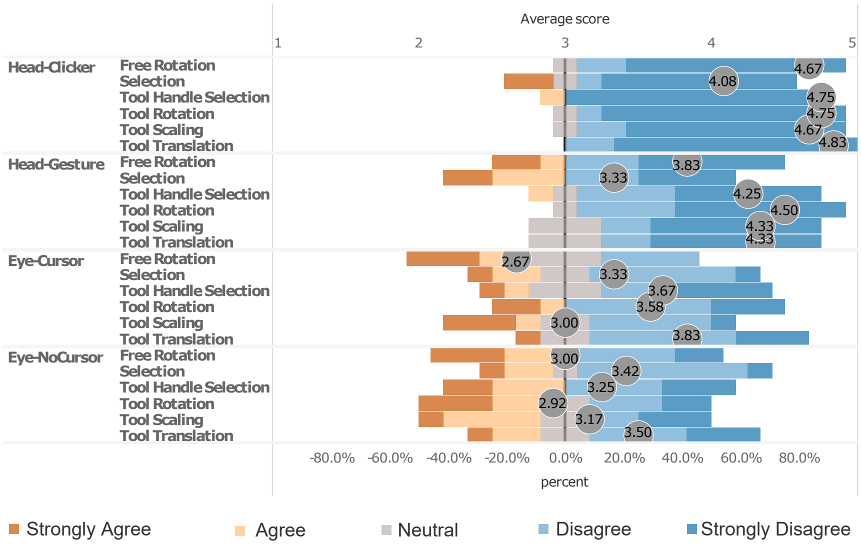
## 5 RESULTS AND DISCUSSION

The hypotheses are examined by evaluating the questionnaires. Furthermore, a non-parametric Friedman test was executed for statistical analysis. Due to the poor eye tracking quality two participants were not able to complete one part of the trials using eye tracking. For these two participants, the questionnaires on these particular parts were rated completely negatively, i.e., given the worst possible rating.

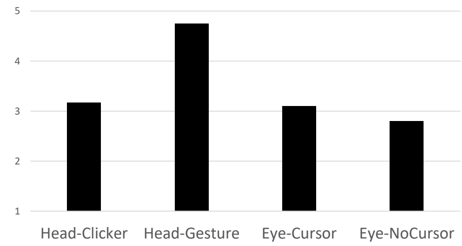
The diverging stacked bar chart in Figure 2a shows the Likert Scale ratings of the different functions of the prototype, for each interaction method. It is clearly visible that the interaction method *Head-Clicker* tends to have a better rating for each of the functions in contrast to the other interaction methods. Also, the eye-gaze-based interaction methods are generally rated more negatively then the head-gaze-based approaches. Since the gaze data from the eye tracker was not very reliable, the input methods with head gaze were preferred by the majority of the participants.

The Friedman tests show significant results regarding the rating of the interaction methods for each function ($10<\chi^2(3)<17$, $p<0.018$), except for the selection functionality ($\chi^2(3)=4.57$, $p=0.2$). Post hoc pairwise comparisons with the Wilcoxon signed-rank test show that the differences between *Head-Clicker* and *Eye-Cursor* ($2.3<Z<2.9$, $p<0.02$, $0.48<r<0.59$) as well as *Head-Clicker* and *Eye-NoCursor* ($2.2<Z<3.1$, $p<0.027$, $0.46<r<0.64$) are significant.

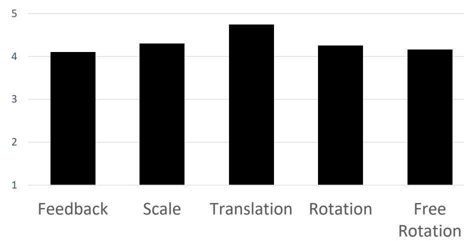
Figure 2b shows how the participants rated each interaction method. The Friedman test indicates significant differences ($\chi^2(3)=11.382$, $p<0.01$). Post hoc comparisons with the Wilcoxon test shows




**(a) A diverging stacked bar chart [Heiberger and Robbins 2014] showing the Likert Scale ratings of the different user tasks for each input method respectively. Bars diverge from the center score '3'; the lengths of the stacked bars encode the percentage in the respective rating category. Gray circles show the average rating score.**




**(b) The subjective rating of the input methods.**




**(c) The rating of the prototype's functions.**

**Figure 2: Evaluation of the different questionnaires as diagrams. High values correspond to a better rating.**

again that there are significant differences between the *Head-Clicker* method and the eye-gaze-based interaction methods (*Eye-Cursor* ($Z$=-2.865, $p$<0.004, $r$=0.584), *Eye-NoCursor* ($Z$=-2.5673, $p$<0.009, $r$=0.524)). In general, interaction with hand gestures is straining, especially if the hand gestures are not recognized well enough by the HoloLens tracker, which we observed quite often during the study. Figure 2b shows that the *Head-Gesture* method has nearly the same rating as *Eye-Cursor*, which is also significantly different from Head-Clicker ($Z$=-2.74, $p$<0.008, $r$=0.558). Participants who were not used to interacting with hand gestures had difficulties in getting the system to recognize their movements as valid input gestures. Figure 2a, shows that *Head-Gesture* generally performs well, except for the selection task, for which it is in fact the worst performing method across all four. For the selection, the recognition of the hand gestures plays a key role, so this could be a reason for the bad performance of head gaze with the air tap gesture.

Overall, the results show that the interaction with head gaze was much more pleasant for the participants and thus contradicts H1. By comparing *Head-Clicker* and *Head-Gesture*, it can be observed that the clicker is preferred to hand gestures and therefore H2 can be confirmed. Figure 2c shows that all of the prototype's functionalities get good ratings (a value of 4 or better in the Likert scale), including the visual feedback provided by the tools, which confirms H3. We cannot confirm H4 because the transformation tool and *free rotation* have very similar ratings in Figure 2c. Also, Figure 2a shows similar ratings between the two functionalities, when taking a look at the *Head-Clicker* method.

## 6 CONCLUSION

We studied how different interaction modes for spatial manipulation of virtual objects in a 3D scene are adaptable to an eye tracking interface. A working prototype was implemented for Microsoft HoloLens with a Pupil Labs binocular eye tracker add-on. To evaluate the feasibility of eye tracking for user interaction in AR with hardware that is currently readily available, we compared different input methods. For the comparison, a user study was conducted in which the participants rated the different input methods and the functionalities of the prototype.

The evaluation of the user study shows that the majority of participants prefer the use of a head gaze for targeting. However, the user experience during the study was affected by the poor accuracy of the gaze data. The study should be repeated with more stable and reliable eye tracking to get clearer results. The built-in gaze tracking of Microsoft HoloLens 2 might be a possible solution to tackle this problem in the future. Apart from the accuracy of the eye tracking system, our method is also limited by some of the tool's design decision: Transformations are often executed in fixed step sizes or at a fixed constant speed. This can be exhausting, if manipulations are carried out over a long period of time. Nevertheless, participants of our study were able to successfully use eye tracking for pointing tasks and interaction with virtual scene elements.

## ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project ID 251654672 – TRR 161 (projects A08, B01) and – project ID 279064222 – SFB 1244 (project B05).

## REFERENCES


Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-gaze over Head-gaze-based Selection in Virtual and Augmented Reality Under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN '18)*. ACM, 1–9.

Heiko Drewes. 2010. Eye Gaze Tracking for Human Computer Interaction. (March 2010).

Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H. Thomas. 2018. *Immersive Analytics: An Introduction*. Springer International Publishing, Cham, 1–23.

Anja Groß, Michael Becher, Guido Reina, Thomas Ertl, and Michael Krone. 2019. A User Interaction Design for Object Manipulation via Eye Tracking in Virtual Reality. *Workshop on Novel Input Devices and Interaction Techniques (NIDIT)* (2019).

Richard M Heiberger and Naomi B Robbins. 2014. Design of diverging stacked bar charts for Likert scales and other applications. *Journal of Statistical Software* 57, 5 (2014), 1–32.

Robert J. K. Jacob. 1990. 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 (CHI '90)*. ACM, 11–18.

Jae-Young Lee, Hyung-Min Park, Seok-Han Lee, Tae-Eun Kim, and Jong-Soo Choi. 2011. Design and Implementation of an Augmented Reality System Using Gaze Interaction. In *2011 International Conference on Information Science and Applications*. 1–8.

Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, 509–518.

Yuan Yuan Qian and Robert J. Teather. 2017. 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 (SUI '17)*. ACM, 91–98.

Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, 281–288.

Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans W. Gellersen. 2013. Pursuits: Eye-based Interaction with Moving Targets. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, 3147–3150.

Yan Wang, Guangtao Zhai, Sichao Chen, Xiongkuo Min, Zhongpai Gao, and Xuefei Song. 2019. Assessment of eye fatigue caused by head-mounted displays using eye-tracking. *BioMedical Engineering OnLine* 18, 111 (2019).