Gaze-based Augmented Reality Interfaces to Support Human-Robot Teams

Christina Petlowany
cpetlowany@utexas.edu
The University of Texas at Austin
Austin, Texas, USA

Nathan Hahn
DEVCOM Army Research Laboratory
Adelphi, Maryland, USA
nathan.p.hahn.civ@army.mil

ABSTRACT

Supervision and teaming with automated systems increase in difficulty as the number of automated systems increases, which is becoming more and more common with improvements in artificial intelligence and growing accessibility of these devices. It is difficult for people to manage multiple systems and most current interfaces do not scale well to larger teams. Augmented Reality (AR) can place information on top of the real world allowing users to track the environment and the robot at the same time–perhaps improving the scaling of these devices. To further investigate the user interface scaling in AR, we implement two modes of a gaze-based interface: active and passive. Gaze is a powerful indicator of attention and may create more reactive systems that can reduce the cognitive burden of a user. In the active mode, the user looks at a menu and presses a button to request additional information. The passive mode does not require active effort from the user, instead switching the menu when the user's gaze dwells on the menu for a set amount of time. We implement a user study where participants perform a visual search task and provide feedback on their user preferences. Results show that the passive and active interfaces provide better scaling as the number of robots increases. However, users slightly prefer the passive interface for its low mental demand, effort, and frustration.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; Graphical user interfaces; **User studies**; **HCI theory, concepts and models**; Empirical studies in HCI; Scenario-based design.

KEYWORDS

mixed/augmented reality, eye-tracking, user interfaces, human-robot teaming

ACM Reference Format:
Christina Petlowany and Nathan Hahn. 2023. Gaze-based Augmented Reality Interfaces to Support Human-Robot Teams. In *Proceedings of The ACM Symposium on User Interface Software and Technology (UIST '23).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

UIST '23, Oct. 29–Nov. 01, 2023, Bay Area, CA
© 2023 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Technology is becoming ever present around us - no longer is computing limited to specific devices, but is now in items like coffee makers and even planters. As these devices begin to incorporate computing – their functions and features become more advanced. We now have robots that can vacuum and mop autonomously, and drones that can follow you and record a video of you skateboarding. While this connectivity and computation make them more capable than ever, it also changes the nature and means of interaction and feedback from these devices. A vacuum is no longer a simple on and off device that you drive around - instead it can be scheduled and must navigate around obstacles in a room, and notify you when it has trouble doing so. For human interaction with robotics, this leads to an "automation conundrum" [5] - where a human's knowledge of a systems status tends to decrease as system automation increases.

Novel means of interaction, such as Augmented Reality (AR), offer a way to interrogate and understand the status of these systems at a glance. By overlaying digital information on the physical world, AR augments a user's gaze with additional information. However, as the number of devices within a user's area increases, they can quickly be overloaded with information and find the digital world overtaking the physical. This can not only create emotional stress, but also overload an individual's visual processing system, making tasks such as visual search take longer.

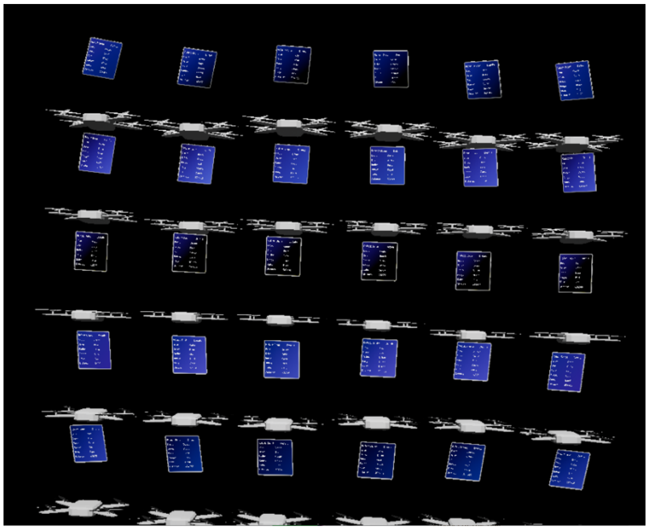


Figure 1: Simulated drones floating with the dense menu above them.

Eye-gaze offers a potential way to mediate the glut of information. By using eye-gaze as a proxy for attention and interest, the amount of information shown to an individual can be increased where their attention is focused, and limited otherwise. Compared to a traditional gesture or selection technique, this offers a much faster response to a change in the user's attention. Simultaneously, it potentially allows users to find information faster by limiting the amount of data they visually must sort through, especially in situations where the number of objects in their view is large.

In this paper, we focus on how an eye-gaze based information mediation interaction can reduce task time and cognitive burden for end users in augmented reality. We developed two different interactions - a passive and active interaction. In the case of the passive interaction, information is limited to the name of a device, and only after a certain dwell time, is additional information about that device shown. In the active condition, a user looks at a device and then clicks a button to show additional information about that device. We then constructed a visual search task using a Microsoft HoloLens 2 that compares these two interactions versus a situation where all the information about a set of devices is shown. We performed 21# trials per participant that increases the number of devices in their field of view where they must find a specific piece of information from a particular device.

Both the active and passive interactions improved performance time over a traditional interface, shown in Figure 1, as the number of robots increased; the gaze interfaces scale better with larger robot teams. Participants prefer both of the gaze-based interactions over the traditional method with a preference towards the passive interaction, even though it did not improve performance quite as much as the active interaction. Users expressed a wish for customizing the passive dwell time setting.

2 RELATED WORK

Gaze is appealing as an interaction technique because it is fast, hands-free, and is naturally employed during perception. Researchers have widely investigated gaze in typical, two-dimensional user interfaces as discussed below, but investigations of gaze-based systems with robotics and AR are burgeoning and primed for design guidelines.

For this effort, we discuss related works in human-computer interfaces, gaze-based or scalable robotic interfaces, and AR robotic interfaces.

2.1 HCI

Most two-dimensional computer interfaces incorporate navigation and selection as the main control methodologies [8]. Gaze movement naturally aligns with navigation, but selection remains difficult.

Human-computer interaction researchers have long been interested in gaze for interfaces as gaze is hands-free and swift. Still, gaze is tricky to use because, as Zhai et al. note, "it is unnatural to overload a perceptual channel such as vision with a motor control task" [18]. For instance, a user may accidentally trigger a dwell selection while they are resting their eyes and not looking at anything in particular or may be trying to trigger a selection when their eyes are drawn to periphery movement. These scenarios are part of the "Midas Touch" problem of user interfaces–a user interface needs to balance being too easy and too difficult to trigger or users may choose not to adopt the system [10].

To circumscribe these issues, Zhai et al. created a dual-modality gaze and button clicking system to [18] employ gaze for navigation but not selection. The Manual and Gaze Input Cascaded (MAGIC) pointing system follows the user's gaze direction for navigation but reverts to manual device input for selection. One version of the gaze-based system allowed participants to click on appearing targets faster than a traditional point-and-click interface.

In later works, Zhai again recommended the use of gaze in attentive, or implicit, interfaces that did not require the user's active gaze effort [17]. Wang, Zhai, and Su created an eye-typing interface that combined gaze direction with a space-bar press to select a Chinese character from a suggested list created typing speeds comparable to a traditional system [15].

2.2 Robotic Interfaces

Humphrey et al. employ a "halo" display to help users manage teams with increasing numbers of robots [9]. The display shows the feed for one specified robot along with a halo of arrows that point to the locations of the other robots with respect to the current one. The results did not confirm if the system helped with scalability or not as there was no comparison interface.

Yu et al. try gaze-based drone teleoperation with a laptop interface [16]. Users could control the robot either by dwelling on six specified areas of the screen to move the drone in the corresponding direction (left, right, up, down, forward, backward) or could implement gaze gestures, or sequential movements between locations, to select a direction. Neither gaze method outperformed keyboard or joystick control. Participants struggled to remember the gaze gestures and the dwell methods resulted in many false positives.

2.3 AR/MR Robotic Interfaces

The Augmented Robot Environment (AugRE) [12] functions as a status monitor for human-robot teams. The system provides localization and communication between robot teams and users wearing Microsoft HoloLens 2 headsets. In theory AugRE is more scalable over a traditional interface because it reduces context-switching, but this has not yet been explored.

Ruffaldi et al. create a system that reduced teaming inefficiency between a participant and a flying drone by communicating the drone's intent to the user so the user could adjust their plan accordingly [13]. They did not consider scalability and tested the system with only one robot.

Hedayati et al. seek to improve drone teleoperation with AR visual overlays that show either the drone's field-of-view or camera feed [7]. These overlays improved performance time and accuracy over existing tablet software, though researchers did note that "participants preferred designs that moderately improved performance over the best-performing design."

Other researchers implement AR cues to improve teleoperation of a co-located robot manipulator [1]. Teleoperation by itself is inherently not scalable, although teleoperation takeovers of autonomous systems could have scaling capabilities.

3 INTERFACE DESIGN

In this work, we explore how gaze can be leveraged to reduce information overload in a scenario where AR is used for interacting with a large number of autonomous systems. For the design and implementation of the interface and study, we utilized a Microsoft HoloLens 2 as the AR platform. In order to simulate interaction with robotic assets in a space, we leverage the AugRE platform [12] to manage the visualization and display of robotic information. We manipulate the menu showing the robot data, creating three interface designs for the study:

- Simple
- Passive
- Active

3.1 Simple Interface

The Simple Interface acts as the control for this study. It shows all the information about a robot with no additional eye-based information mediation. This menu can be seen in the right side of Figure 3. The top left of the menu displays the robot's name and there are six categories of information below on the left with their details on the right. The Distance category includes dynamic information to more closely represent a changing robot status.

3.2 Passive Interface

The Passive Interface incorporates lessons learned from the related works in an implicit design. As shown in Figure 3, the menu first shows only the robot name: the "Sparse Menu." After a user's gaze dwells on the menu for 1.5 seconds, the menu switches to the "Dense Menu" with all of the information. The Sparse Menu returns when the user is no longer looking at the Dense Menu (no intersection of gaze with the menu for 0.5 s). The text for the robot name is the same size on both the Sparse and Dense menus to eliminate effects from text size on the performance time compared to the Simple interface.

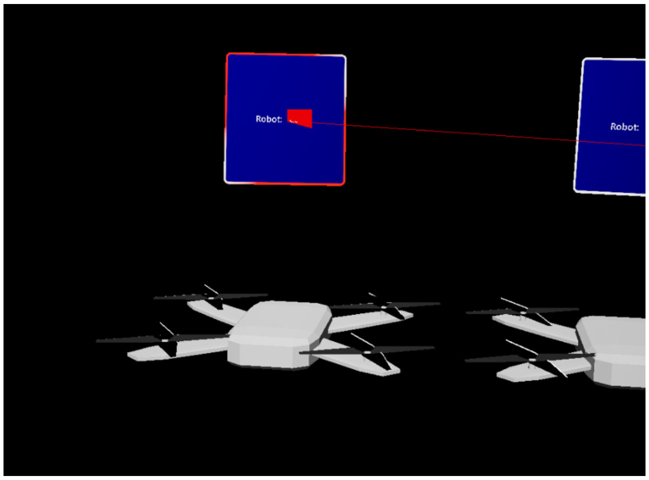


Figure 2: The sparse menu and its intersection with the user's gaze vector (red, shown for visualization purposes).

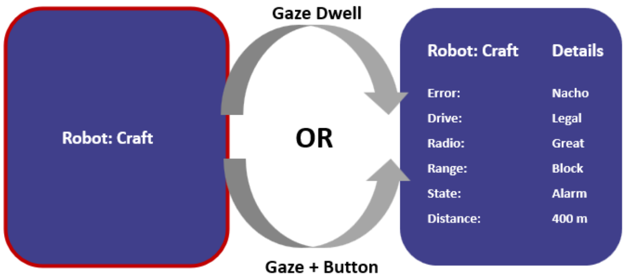


Figure 3: Depiction of the menus and the conditions that prompt a transition.

The 1.5-second-dwell was informed by the literature and initial testing. Bafna et al. led an eye-typing study and found typing speed to increase with an increased dwell time, while errors increased [2]. The best performance occurred with dwell times of 350 - 450 ms (tested range of 250 ms to 1150 ms). We provided additional time for users to read the two words on the initial menu. Research shows the average reading rate is around 238 words per minute [3], or about four words per second. We doubled this time to account for the moving menus and increased difficulty while reading from an AR device.

To provide the user with indicator of where they are currently looking, as long as the progress of the interaction, the outline of the menu begins to darken as the user dwells on the Sparse Menu, shown in Figure 2.

3.3 Active Interface

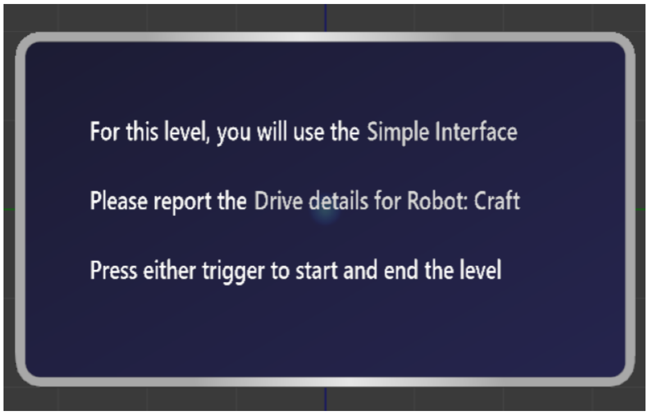
The Active Interface demonstrates an explicit UI design; the user must actively press a button to receive more information; in this case, the button is the red "B" button on the front of an Xbox controller. Otherwise, this interface functions similarly to the Passive Interface, except the trigger condition is a user looking at a menu and simultaneously pressing a button on a controller, shown in Figure 3, and the interface switched back immediately once the user averted their gaze from the menu. The menu outline turns red when the user is looking at a menu to give visual feedback.

4 EXPERIMENT DESIGN

This research seeks to understand 1) the impact of the interface on users seeking information, 2) the effect of the interface on the scalability of robot teams, and 3) user preferences between the different interfaces.

Regarding the first goal, we utilize a standard visual search task similar to [14] or [11]. Participants are instructed to seek out specific details from the interface, e.g. the Drive details (in Figure 3, Legal). The system measures the time for each trial to see if there is an impact of the interface on the user's ability to quickly find the relevant information. To minimize the effect of reading times on the results, the words are chosen from a word bank of five-letter words, listed in Appendix A.

The trial is designed to stay within a person's working memory limits [4]. The users must remember a robot name and an information category in addition to the menu interaction.



For this level, you will use the Simple Interface

Please report the Drive details for Robot: Craft

Press either trigger to start and end the level

Figure 4: The trial start menu.

Participants first view a starting menu, in Figure 4, that lists the interface and the requested information for the trial. To start and stop the trial, the users press a trigger on an Xbox controller. Participants report the information to the study lead after they end the trial.

These interfaces appear as a menu floating above a robot, much like as in [12]. The robots in the trials are simulated drones, shown in Figure 5. The drones drift slowly around the scene as they move from one random nearby point to the next.

Regarding the second goal, participants complete multiple trials for each interface with sets of 1, 5, 10, 15, 20, 25, and 30 robots and sought the details from one specific robot (in Figure 3, Robot:Craft). The drones start in random sections of a 5x6 vertical and horizontal grid 5 meters in front of the participant as in Figure 6. Initial results with the robots circling the user showed that the robot starting position dominated the performance time and hid effects from the interface. The randomized robot location aims to minimize further effects from robot starting position. The 30-robot maximum was intended to capture a reasonable maximum number of robotic teammates while being large enough to invoke cognitive overload in participants.

Finally, participants answered surveys to reveal their preferences between the interfaces. Participants fill out four surveys: three surveys based on the NASA-TLX scale for each of the three different interfaces [6] and one final survey with Likert-scale responses comparing the different interface modalities and long-form answer questions. B

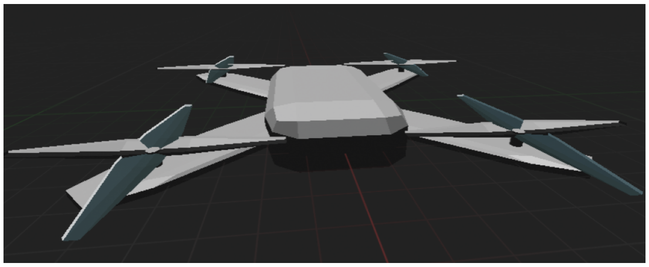


Figure 5: The simulated drone design.

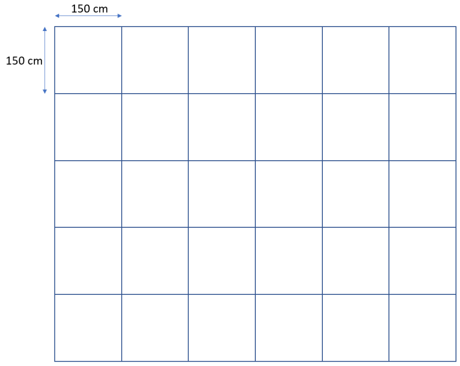


Figure 6: The grid for robot placement.

To account for learning effects, the order of interfaces was counterbalanced across all participants. Within the set of trials for each interface, the participants first completed a practice trial to further reduce learning effects and then completed the rest of the trials in that interface in a randomized order. Participants filled out the NASA-TLX survey for each interface directly after each set of trials while the experience with that interface was recent.

5 RESULTS

We recruited 16 participants (12M | 4F) with ages ranging from 19-36 (25.75 +- 4.75 years). All participants save one had previous AR experience, with phone-based games such as Pokémon Go or even a Microsoft HoloLens 2. This study was approved by the ANONYMIZED FOR REVIEW IRB under Study Number STUDY NUMBER.

We applied a Mixed Linear Model to the dependent variable of trial time with the interface order, the number of robots, the type of interface, and the number of robots plus the interface type as independent variables. This analysis excluded five outlier trials with times greater than three standard deviations from the average trial time. Excluding these outliers, the average trial times were: Simple interface - 10.6 s, Passive interface - 9.0 s, and Active interface 8.7 s. The Active interface resulted in 1.9-s improvement over the Simple interface and the Passive has a 1.6-s improvement.

For all of the analysis, we set the significance level to $p = 0.05$. There was no significant dependence between the interface order or type of interface and trial time. When the previous variables are removed from the model, the number of robots was significant ($p = 0.000$) and so was the Active interface compared to the Simple interface including number of robots ($p = 0.002$) and the Passive interface compared to the Simple interface including the number of robots ($p = 0.010$). The average trial time for each interface with increasing numbers of robots are shown in Figure 7.

Trials with the incorrect answer were excluded from the results. There were 22 trials with incorrect answers, all noted in Table 1: six from the Simple interface, eight from the Passive interface, and eight from the Active interface. Some participants noted that it was difficult to remember which details they needed to report; participants gave the details for the correct robot but wrong details category for 15 of the 22 incorrectly answered trials. During the

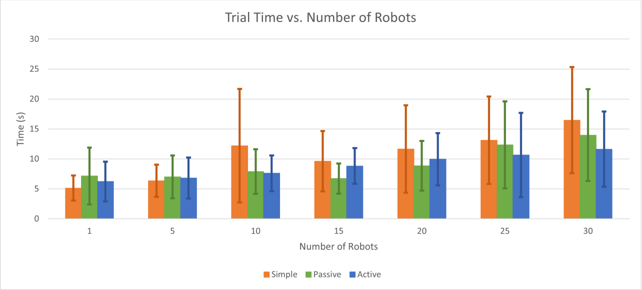


Figure 7: Average trial times versus number of robots for each interface type.

Table 1: Trials with Errors

Interface	Simple	Passive	Active
1 robot	2	1	1
5 robots	0	1	0
10 robots	0	1	1
15 robots	0	1	0
20 robots	3	1	1
25 robots	1	2	4
30 robots	0	1	1

Table 2: False Positives

Interface	Simple	Passive	Active
1 robot	0	0	0
5 robots	0	0	0
10 robots	3	0	0
15 robots	1	0	0
20 robots	2	0	0
25 robots	2	2	0
30 robots	7	3	1

Table 3: NASA-TLX Survey Results

Interface	Simple	Passive	Active
Mental Demand	5.625	4.1875*	4.4375*
Physical Demand	3.5	2.9375	3.5
Temporal Demand	5.3125	5.5625	4.9375
Performance	8.4375	8.4375	8.875
Effort	5.8125	4.75*	5.125
Frustration	4.125	3.0625*	3.375

trials, a few participants made verbal comments that the words were similar and hard to tell apart.

Additionally, we explored the time to view the correct menu once the trial started. We first removed outliers from the data. The average time to view the correct menu is heavily dependent on the number of robots and the number of robots plus the type of interface (p = 0.000 for all). It took participants longer to find the correct robot as the number of robots increased and the Active and Passive interfaces decreased the time to find the correct robot as the number of robots increased. The average trial times to find the correct menu were: Simple interface - 7.8 s, Passive interface - 5.3 s, and Active interface 5.1 s. This shows that at least part of the performance improvements with the Passive and Active systems resulted from them enabling the users to quickly identify the correct robot's menu.

We defined false positives as when a participant viewed the correct menu but looked away for longer than two seconds. We detected 21 false positives over the trials: 15 for the Simple interface, 5 for the Passive interface, and 1 for the Active interface. False positives occurred more frequently with more robots and with the Simple interface; this is summarized in Table 2.

Overall, the surveys revealed participants preferred both the Active and Passive interface over the Simple interface, with a proclivity for the Passive interface in particular.

Paired t-tests on the NASA-TLX survey data indicate that participant responses were not statistically significantly different ($p < 0.05$) between the three interfaces for Physical Demand, Temporal Demand, or Performance. Participants, however, did rate the Passive interface as less mentally demanding ($p = 0.0010$), requiring less effort ($p = 0.012$), and less frustrating ($p = 0.045$) than the Simple interface. Participants found the Active interface to also be less mentally demanding than the Simple interface ($p = 0.10$). Survey responses did not show a significant difference between the Active and Passive interfaces on these measures. These results are summarized in Table 3 where the categories in which an interface outperformed the Simple interface are starred.

The final survey presented to the participants further supports this interpretation. Out of the participants, 31.3% disagreed or strongly disagreed that the Simple interface was intuitive. Meanwhile, all participants at least neutrally agreed that the Passive interface was intuitive as summarized in Figure 8. This result was statistically significant for Passive vs. Simple ($p = 0.0063$), but not for the other combinations of interfaces.

Almost half of participants (43.8%) said they disagreed that the Simple interface was easy to use, while none did for the Passive interface. Only 12.5% of participants thought the Active interface was not easy to use, shown in Figure 9. The Passive interface was easier to use than the Simple interface ($p = 0.0014$), same with the Active interface ($p = 0.043$).

We asked participants if they felt the interface made them feel like they and the robots were part of the same "team". More people said they felt like a member of the team with the Active interface ($p = 0.017$) and Passive interface ($p = 0.030$) compared to the Simple Interface, likely because the system responded to their input.

Only two participants said they preferred the Simple interface over the Passive interface, but four said they preferred the Simple interface over the Active interface. The results were split for Active vs. Passive, with seven preferring the former and nine preferring the latter, shown in Figure 10.

Participants found the Active and Passive systems responsive, with none disagreeing with the statements: "The passive gaze system correctly identified when I looked at a label" and "The active gaze system responded when I asked to see more information."

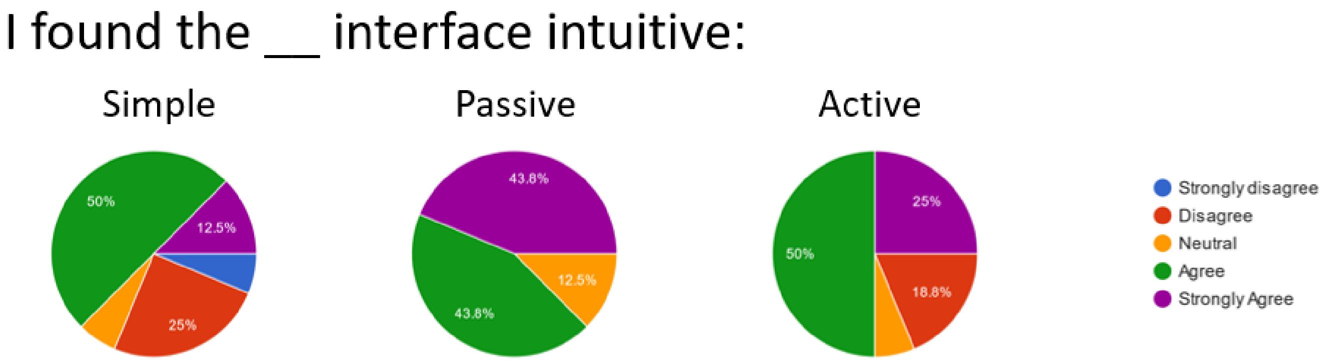


Figure 8: Responses to the survey intuitiveness questions.

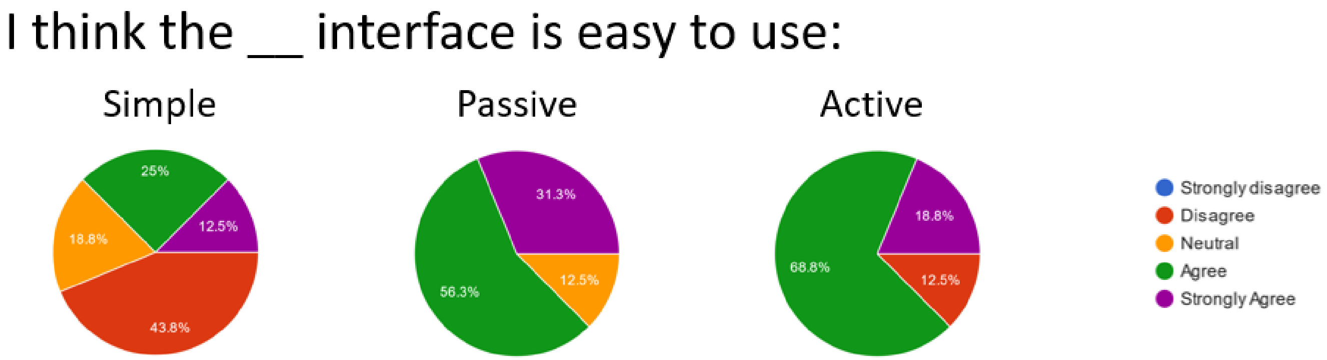


Figure 9: Responses to the survey ease of use questions.

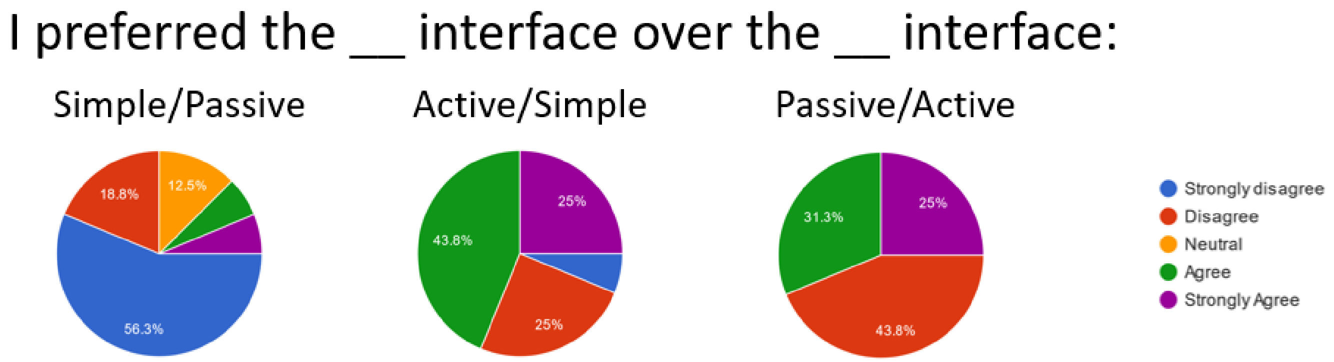


Figure 10: Responses to the survey interface preference questions.

Participants preferred the Passive interface over the Active interface because: " I had to do the least amount of work both in looking for the robot (higher for simple) and in button presses (higher for active)" and the Passive interface's "simplicity." Opponents of the Passive system disliked when its timing wasn't quite right. "[T]he passive gaze could sometimes give extra, distracting information if my sight lingered for too long," one participant stated. Another said, "I liked the active gaze the best... The passive gaze was also good, though it was difficult to get the timing right. I.e. sometimes it responded too quick before I could figure out the name of the robot (especially for those further away)." Even advocates for the Passive interface noted this as a downside, even if it wasn't necessarily a deal breaker, as one participant noted, "The passive gaze sometimes switched to the details screen earlier than I expected to. That wasn't necessarily bad, just surprising the first time it happened." Another stated, "[I liked the] Passive interface, but it should not take too long to show the details."

One participant verbally noted a tendency to avert their gaze while pressing the Active interface button, which prevented the switch to the more detailed menu and could further explain some of the preferences for the Passive interface. The Active interface could have a delay on the switch back to the sparse menu, similar to the delay on the Passive interface.

6 DISCUSSION

Overall, the results indicate that eye-gaze based interactions can help to improve visual search performance in an augmented reality environment, and help to reduce the strain of information overload. The value of the interaction to end users is shown through the NASA TLX workload surveys, task performance time, and self-report user preference, especially in environments with significant augmented reality information overlayed. As augmented reality becomes more ubiquitous, and more items seek to interact through it, eye-gaze serves as a natural way to mediate overload. In the future, we see an eye-gaze based information mediation interactions as a fundamental piece of the augmented reality development toolkit. Toolkit fundamentals, such as titles, menus, and text fields could have a "low information" and a "high information" state, with an eye-gaze interaction allowing users to switch between those two modes. If this was implemented at a platform level, it would allow the device to naturally reduce information clutter, and prevent overload strain on the end user.

While participants indicated they like the passive interface the most, we saw a faster task time for the active interface. We have a number of theories for why this might occur. The first is that the dwell time required during the passive condition could have negated any time saving effect from the reduced information load, resulting in a non-significant performance improvement. We had a few participants indicate to us that the dwell time was too long, and they would have preferred for the information to appear quicker. Instead of dwell time being fixed across all users, this could instead be a preference that is either set by an end-user, or determined automatically through some type of calibration with the end-user. Alternatively, the preference of the passive interface over the active interface could be related the form factor of the input device users utilized to trigger the information shown. Having to hold an additional device in your hands would be burdensome compared to a hands-free interaction. A different confirmation interaction in the "Active Condition", such as a hand gesture or a tap on the device could provide the necessary performance, without the burden of having to hold a controller or additional device.

We designed these interfaces to reduce a user's cognitive load, but did not measure their cognitive load levels. Future studies could record data that correlates to cognitive load, such as pupil dilation.

This study incorporated simulated drones and future studies could investigate the effects of these interfaces when there are physical drones performing more realistic tasks. This type of visual search task may not be directly applicable to a real-life human-robot teaming scenario. For instance, the gaze-based interfaces withhold information until a user indicates a desire for more and it is possible that user may miss important status updates.

Future work should investigate applications of these interactions to less menu-based systems. One possible avenue is hiding the menus until a user views the robot. We are actively investigating gaze as feedback for robot task execution and gaze in a robotic system to help users learn tasks and collaborate with the robot.

7 CONCLUSIONS

This paper presents investigations into gaze-based user interfaces in Augmented Reality to allow users to supervise increasing numbers of robotic systems. We directed 16 participants to complete information search tasks with a passive, gaze dwell-based method and an active, gaze with button-press method compared to a simple interface requiring no input from the user. Participants performed better with both of the gaze-based interfaces as the number of simulated autonomous drones increased and preferred these interfaces over the traditional, simple system. This investigation demonstrates that these types of gaze-based AR systems enable users to control information flow and find information more quickly, perhaps by reducing their cognitive burden from when the information is presented all at once.

ACKNOWLEDGMENTS

This study was approved by the ANONYMIZED FOR REVIEW IRB under Study Number STUDY NUMBER

REFERENCES

[1] Stephanie Arevalo Arboleda, Franziska Rücker, Tim Dierks, and Jens Gerken. 2021. Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.

[2] Tanya Bafna, John Paulin Paulin Hansen, and Per Baekgaard. 2020. Cognitive load during eye-typing. In *ACM symposium on eye tracking research and applications*. 1–8.

[3] Marc Brysbaert. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of memory and language* 109 (2019), 104047.

[4] Nelson Cowan. 2010. The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science* 19, 1 (2010), 51–57.

[5] Mica R Endsley. 2017. From here to autonomy: lessons learned from human–automation research. *Human factors* 59, 1 (2017), 5–27.

[6] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[7] Hooman Hedayati, Michael Walker, and Daniel Szafir. 2018. Improving collocated robot teleoperation with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 78–86.

[8] Anke Huckauf and Mario H Urbina. 2008. On object selection in gaze controlled environments. *Journal of Eye Movement Research* 2, 4 (2008).

[9] Curtis M Humphrey, Christopher Henk, George Sewell, Brian W Williams, and Julie A Adams. 2007. Assessing the scalability of a multiple robot interface. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 239–246.

[10] Robert JK Jacob. 1993. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction* 4 (1993), 151–190.

[11] Jonathan Ling and Paul Van Schaik. 2002. The effect of text and background colour on visual search of Web pages. *Displays* 23, 5 (2002), 223–230.

[12] Frank Regal, Christina Petlowany, Can Pehlivanturk, Corrie Van Sice, Chris Suarez, Blake Anderson, and Mitch Pryor. 2022. AugRE: Augmented Robot Environment to Facilitate Human-Robot Teaming and Communication. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 800–805.

[13] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. 2016. Third point of view augmented reality for robot intentions visualization. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 471–478.

[14] Missie Smith, Jillian Streeter, Gary Burnett, and Joseph L Gabbard. 2015. Visual search tasks: the effects of head-up displays on driving and task performance. In *Proceedings of the 7th international conference on Automotive User Interfaces and Interactive Vehicular Applications.* 80–87.

[15] Jingtao Wang, Shumin Zhai, and Hui Su. 2001. Chinese input with keyboard and eye-tracking: an anatomical study. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 349–356.

[16] Mingxin Yu, Yingzi Lin, David Schmidt, Xiangzhou Wang, and Yu Wang. 2014. Human-robot interaction based on gaze gestures for the drone teleoperation. *Journal of Eye Movement Research* 7, 4 (2014), 1–14.

[17] Shumin Zhai. 2003. What's in the Eyes for Attentive Input. *Commun. ACM* 46, 3 (2003), 34–39.

[18] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 246–253.

A WORD BANK

Table 4: Word Bank

Adult	Alarm	Among	Anger
Audio	Baker	Bench	Block
Bread	Cabin	Cause	Chair
Coach	Craft	Crime	Crown
Dance	Doubt	Dozen	Dream
Eagle	Entry	Equal	Extra
Fable	Final	Fruit	Grass
Great	Guide	Habit	Heart
Horse	Index	Jacks	Juice
Known	Legal	Limit	Magic
Money	Nurse	Oasis	Ocean
Party	Pilot	River	Royal
Sharp	Shelf	Smoke	Study
Tired	Truck	Union	Urban
Voice	Water	World	Yield

B SURVEYS

NASA-TLX Survey

- Mental Demand - How mentally demanding was the task?
- Physical Demand - How physically demanding was the task?
- Temporal Demand - How hurried or rushed was the pace of the task?
- Performance - How successful were you in accomplishing what you were asked to do?
- Effort - How hard did you have to work to accomplish your level of performance?
- Frustration - How insecure, discouraged, irritated, stressed, and annoyed were you?

Final Survey

- I found the simple interface intuitive
- I found the passive gaze interface intuitive
- I found the active gaze interface intuitive
- I think the simple interface is easy to use
- I think the passive gaze interface is easy to use
- I think the active gaze interface is easy to use
- I felt like a member of the team with the simple interface
- I felt like a member of the team with the passive gaze interface
- I felt like a member of the team with the active gaze interface
- I preferred the simple interface over the passive gaze interface
- I preferred the active gaze interface over the simple interface
- I preferred the passive gaze interface over the active gaze interface
- The passive gaze system correctly identified when I looked at a label
- I did not have to look to long at a label for it to show more information in the passive gaze system
- The active gaze system responded when I asked to see more information
- What did you like the best? (long form)
- Any questions about working with robots or AR? (long form)
- Any further comments? (long form)

Received 5 April 2023