2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 979-8-3503-2891-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR-Adjunct60411.2023.00093

Beyond the Wizard of Oz: Using Imperfect Machine Learning to Examine the Impact of Reliability of Augmented Reality Cues on Visual Search Performance

Aditya Raikwar* Colorado State University
Domenick Mifsud† Colorado State University
Christopher D. Wickens‡ Colorado State University
Brendan Kelly§ Colorado State University
Amelia C. Warden¶ Colorado State University
Benjamin A. Clegg|| Colorado State University
Francisco R. Ortega** Colorado State University

Abstract

Searching for an object in their visual field, even when the person knows exactly what the object looks like, is a complex task. This is time-consuming and can be error-prone. We present findings of automation bias when imperfect cues are presented using real-time machine learning, presented in the Augmented Reality head-mounted display using the YOLOv5 machine learning model. We validated that Wizard-of-Oz studies produce results that are equivalent to using a real-time machine-learning system for visual search aided by cues.

Index Terms:

Human-centered computing—Empirical studies in HCI—

1 Introduction

Visual Search tasks with Augmented Reality (AR) Head-Mounted Displays (HMDs) allow users to find an object in the real world using different techniques, compared to a regular display (e.g., monitor and tablet). In some types of visual search tasks, the user is provided a target ahead of time and asked to find it. These may consist of simple virtual geometries like spheres [1] or rectangles [2] or real-world objects such as office items [3]. Even when the user knows what the target looks like (i.e., they are looking for a specific object/person in the search field), there is a performance increase when aid is provided. For example, Warden et al. showed that the arrow cue performed significantly better than the no-cue condition [5]. Yet, when controlling the accuracy (83% using Wizard of Oz (WoZ)) users showed a clear decrease in performance in visual search when the object was not cued correctly. Our approach continues from Warden et al.'s [5] work by using the best cue found (i.e., the arrow) while adding a real-time Machine Learning (ML) providing a non-constant accuracy, with an average of 88.9%. Automation bias [4], defined as the tendency to automatically follow the automation's recommendation, was also observed as it had been in WoZ studies before.

*e-mail: adirar@colostate.edu
†e-mail: dmifsud@rams.colostate.edu
‡e-mail: chris.wickens@colostate.edu
§e-mail: brendan.kelley@colostate.edu
¶e-mail: acwarden@colostate.edu
||e-mail: benjamin.clegg@colostate.edu
**e-mail: fortega@colostate.edu

2 Methodology

Our hypothesis (H_1) was based on the large database of research on automation bias: *Imperfect cueing automation would negatively impact the overall search performance due to high erroneous searches by humans on those imperfect trials*.

Participants: The total number of participants was 53. Each participant was assigned either of 2 conditions (perfect or imperfect cueing). There were 25 participants (17 male, 7 female, and 1 non-binary) with imperfect cueing and 28 (16 male, 12 female) with perfect cueing. This study was approved by the university IRB.

Materials: Participants completed the experiment using the Varjo XR-3 video pass-through AR HMD. To track the objects, an OAK-1 camera was mounted to the front of the Varjo XR-3. OAK-1 has dedicated hardware for running neural networks on the device. The neural network used was YOLOv5-Nano which was trained from scratch on a synthetic dataset of 500,000 images. The synthetic dataset was created in Unity 2020.3.27 using 3D models of the 40 target objects + 3 practice objects. The study software was developed on Intel core i9-9900K, 64GB RAM, and RTX 2080Ti graphics card.

Task: Participants completed a visual search task in 3D space in which they were asked to locate real-world objects (Mega Bloks), with and without the aid of a target cue presented via the Varjo XR-3 AR HMD video-pass through. Participants were told in the instructions that the system is not 100% reliable and may fail. They were instructed to confirm the suggestion presented by the system. Each trial consisted of three steps: (1) Target image shown. (2) Searching for the designated object. (3) Selecting object (In case of cue condition, the participants could select or reject the object suggested by the cue, and search for the target on their own. The target cue condition consisted of a 3D arrow pointing to the target object's location in 3D space.)

This was a mixed-subject design where the groups were split based on different levels of cue reliability (between subjects), and each group had two conditions, cue versus no cue (within-subjects, repeated measures). The two levels of cue reliability were: 1) the locations of all objects were known, and the cue always pointed at the correct object, and 2) the objects were located using ML and had an average accuracy of 88.9%. Participants did not receive a cue to help find the target object for the no-cue condition (i.e., the control condition). The search field of the room incorporated 180° surrounding the participant when looking forward from the chair in which they were seated. A total of 40 potential target objects + 3 practice objects were uniformly distributed across the 180° search field in the horizontal direction and approximately $\pm 15°$ in the vertical direction relative to the chair. Objects were placed at three different height levels: ground, table, and shelf level (separation between levels was 28″).

2771-1110/23/$31.00 ©2023 IEEE
DOI 10.1109/ISMAR-Adjunct60411.2023.00093

3 Results

The Speed-Accuracy Trade-off Function: The combined effects of imperfect cueing on accuracy and response time – that is, on overall performance as stated in H_1 - are represented in the speed-accuracy trade-off space in Figure 1. Within the figure, high performance (high accuracy, fast responses) can be seen in the upper left region, while poorer performance is in the lower right corner. Figure 1 first illustrates with high prominence, the general equivalence in both measures of the uncued condition, the two circled points. Thus, any concerns about fundamental differences in search performance abilities between the two populations are eliminated. The differential impact of cueing on the two groups is illustrated by the two arrows. When the cues are perfect, their enormous benefits to both dimensions of performance are pronounced, as illustrated by the long dashed orange arrow pointing to the upper left. But for the imperfect ML cueing group, with this relatively small drop in cueing reliability, from 100% to 88.9%, the cues' benefit to accuracy is eliminated, and its benefit to processing speed, while not eliminated, is greatly reduced, as illustrated by the short dashed blue arrow.

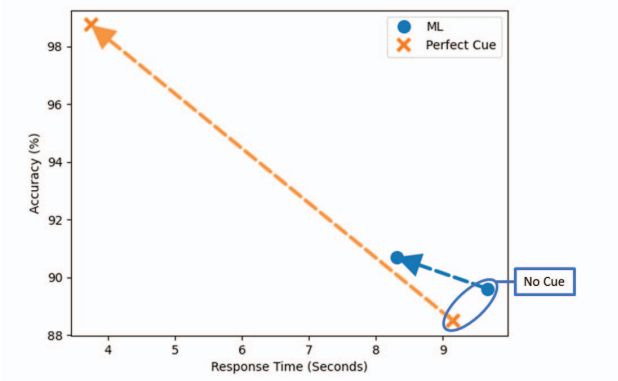

Figure 1: Speed-accuracy trade-off: the combined impacts of imperfect cueing.

In Figure 2, we have used the same speed-accuracy space as in Figure 1, to depict the automation bias observed in the current data for the imperfect automation group. In the upper left is depicted the excellent performance when the cue was correct. In the lower right is the performance when the cue was wrong. The figure reveals the large and statistically significant loss in performance accuracy from 88.3% to 33.1% ($t(23) = 8.15$, $p < .001$), Cohen's $d = 1.66$), as well as the large increase in response time, from 7.3s to 16.32s ($t(25.7) = 5.04$; $p < .01$, Cohen's $d = 1.63$). Collectively, both of these effects further confirm H_1. Not only does performance suffer on all search trials with imperfect automation cueing as shown in Figure 1, but that performance is particularly bad when automation errs and the wrong object is cued.

The finding of the response time delay, when automation is wrong, illuminates the cognitive strategies employed by the participants. This delay indicates participants did not always "blindly follow" the automation recommendation. Rather, when they saw a cue they thought was wrong, they sometimes tried to figure out the correct target (based on their imperfect memory of the target image) but failed and selected the wrong target, a process which took an added 11.4 seconds.

The two strategies to be followed when the cue is wrong would seemingly produce a different pattern of effects of response time and accuracy. "Blindly following" (the automation bias) would keep response time at its very short value typical of the control condition (8.3 sec) or the cue-correct condition (7.3 sec); and produce a near 100% error rate. In contrast, noticing the error and triggering a "re-search" for the correct target would produce this prolonged delay. Accuracy would be improved but be far from perfect because, by this time, the participant's memory of what the true target looked like would be degraded. We cannot estimate the proportion of trials on which each of these two strategies was deployed, nor the extent to which different participants or different trials for the same participant contributed to this difference in strategy choice, as this study was not designed with this purpose.

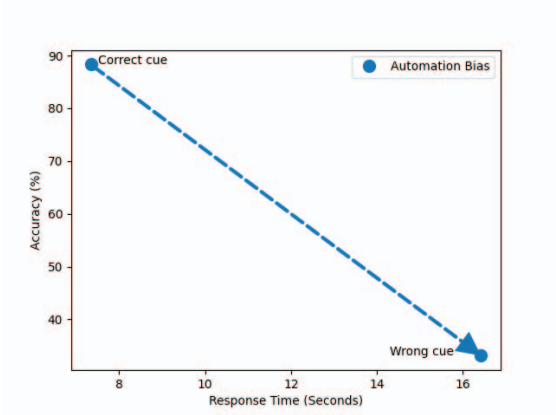

Figure 2: In the speed-accuracy trade-off space, automation bias.

In conclusion, showing the speed-accuracy trade-off function of our study, we validated that visual search suffers when machine learning is imperfect, as previous WoZ studies did. In future work, the study may be expanded to look at different objects (real-world objects that contextually belong to the environment) and add visual search on the move (e.g., walking).

Acknowledgments

ONR N00014-21-1-2949, ONR N00014-23-1-2298; Dr Peter Squire was the scientific/technical monitor.

References

[1] S. Chung, H. Joh, E. Lee, and U. Oh. Panocue: An efficient visual cue with a omnidirectional panoramic view for finding a target in 3d space. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 218–223, 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00052

[2] R. J. Cunio, D. Dommett, and J. Houpt. Spatial auditory cueing for a dynamic three-dimensional virtual reality visual search task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):1766–1770, 2019. doi: 10.1177/1071181319631045

[3] W. Lu, H. B.-L. Duh, S. Feiner, and Q. Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):404–412, 2014. doi: 10.1109/TVCG.2013.241

[4] R. Parasuraman and D. H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010. PMID: 21077562. doi: 10.1177/0018720810376055

[5] A. C. Warden, C. D. Wickens, D. Rehberg, F. R. Ortega, and B. A. Clegg. Fast, accurate, but sometimes too-compelling support: The impact of imperfectly automated cues in an augmented reality head-mounted display on visual search performance, 2023.