ARL-TR-10484 ● Nov 2026

Characterizing Dimmable Augmented Reality for Highlighted Target Detection

by Roman A. Stemp and Chou P. Hung

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-10484 ● NOV 2026





Characterizing Dimmable Augmented Reality for Highlighted Target Detection

Roman A. Stemp
Army Educational Outreach Program Student

Chou P. Hung
DEVCOM Army Research Office

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
November 2026	Technical Report	June 2025	October 2026

4. TITLE AND SUBTITLE
Characterizing Dimmable Augmented Reality for Highlighted Target Detection

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
W911SR-26-2-A003		0601102A
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Roman A. Stemp and Chou P. Hung

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLR-EH
Adelphi, MD 20783-1138

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-10484

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Roman A. Stemp, 0009-0003-1936-6731; Chou P. Hung, 0000-0003-1447-9828

14. ABSTRACT
Augmented reality can enhance visual search performance, but the luminance contrast of target highlighting can be overly distracting, impairing search performance in real-world environments. Real-world scenes, particularly with combinations of shadow and strong illumination, can vary in luminance up to 100,000:1 in a single view, making it difficult to control the highlighting contrast to avoid attentional capture, impairing the ability to detect unhighlighted targets. We characterized how target detection performance is affected by a novel target highlighting technique called divisive display augmented reality (ddAR), in which a region is highlighted by dimming the transparency versus adding luminance. Across a choice of six possible target locations, we tested varying levels of target and highlight contrast and different numbers of targets. We also tested conditions where the highlighting is missing or misleading, simulating different types of aided target recognition malfunction. We report that 20% dimming significantly increased target detection, with low cost to the detection rate of a second unhighlighted target. Lower highlight contrasts (2%, 5%, and 10% dimming) had little-to-negligible effect on detection compared with baseline (0%), whereas higher contrasts (40%, 60%, and 100% dimming) became too visually salient, producing an attentional capture effect distracting from unhighlighted targets.

15. SUBJECT TERMS
Humans in Complex Systems, augmented reality, visual search, attention, aided target recognition, heads-up display, cockpit

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	31

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Chou P. Hung	(520) 942-2081

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures v

List of Tables v

Acknowledgments vi

Declaration of Generative AI and AI-Assisted Technologies vi

1. Introduction 1

1.1 Target Highlighting for AI-Augmented Visual Search 1

1.2 Specific Aims 4

2. Methods 5

2.1 Subjects 5

2.2 Experimental Setup 5

2.3 Visual Search Task 6

2.4 Analysis 8

3. Results 8

3.1 Detection Performance for 1-Target and 2-Target Conditions 8

3.2 Analysis of Malfunctioning Highlighting Conditions 10

3.3 Cost–Benefit Analysis 12

4. Discussion and Conclusion 13

6. References 15

Appendix. Supplemental Material 17

A-1. Hallucinations 18

A-2. Radial Bias 18

A-3. Participant Bias 19

A-4. Reaction Time Analysis 20

List of Symbols, Abbreviations, and Acronyms **22**

Distribution List **23**

List of Figures

Figure 1. Example of attentional capture in additive vs. divisive display highlighting for AR. 4

Figure 2. Dimmable AR test bed, stimulus display, and timecourse. 6

Figure 3. Highlight contrast vs. detection performance for 1-Target trials. 9

Figure 4. Highlight contrast vs. detection performance for 2-Target trials. 10

Figure 5. Highlight contrast vs. correct rejection for 0-Target trials. 11

Figure 6. Highlight contrast vs. false positive reports for misleading trials in which one target was present but the highlight was on a nontarget pedestal. 12

Figure 7. Costs vs. benefits of each contrast level. 13

Figure A-1. Detection accuracy across target locations. 19

Figure A-2. Mean reaction time vs. highlight contrast. 21

List of Tables

Table 1. Trial types. 7

Acknowledgments

We thank Chloe Callahan-Flintoft (ARL), Patrick Cox (Lehigh University), Stephen Mitroff (George Washington University), and Nathan Schwalm (DEVCOM Army Research Office) for helpful discussions, and further thank Chloe Callahan-Flintoft and Nathan Schwalm for participating in the pilot testing. R.S. was supported by the U.S. Army Educational Outreach Program.

Declaration of Generative AI and AI-Assisted Technologies

The authors affirm that Cursor AI was used to generate Python code using matplotlib to create the figures in this report. Cursor was also consulted to implement the Weibull fits and help draft a preliminary analysis.

1. Introduction

The Undersecretary of Defense for Research and Engineering highlighted 14 Critical Technologies Areas as part of the 2023 National Defense Science and Technology Strategy (DoD 2023). Two of these critical technologies were 1) Human–Machine Interfaces and 2) Trusted AI and Autonomy. An important gap to be addressed for these technologies is how augmented and mixed reality devices should be designed to enable the user to leverage new AI capabilities and autonomous platforms. Specifically, they must allow the Warfighter to sense, understand, decide, and act on the overwhelming speed and volume of information in Multi-Domain Operations, particularly in drone and autonomous systems warfare. Although AI and autonomous platforms can develop a common operating picture and filter for important, timely, and relevant information to present to the user, the principles of human visual psychophysics for interfacing and interacting with that information are not well understood. Academic research in neuroscience and psychophysics has historically focused on understanding natural biological vision. In this domain, evolution has resulted in precise circuits and computations for perceiving, anticipating, and surviving with natural scene statistics and dynamics, typically in the near field with local actors. With augmented reality (AR) and AI, new nonevolutionary principles are needed to deal with more abstract, high-dimensional, and multiscale information beyond our immediate vicinity. At the intersection of the abstract and physical worlds, future AR systems must work within the constraints of evolutionary biology while creating new capabilities for new problem spaces never before encountered in hundreds of millions of years of evolution.

1.1 Target Highlighting for AI-Augmented Visual Search

The potential of AR displays to improve visual search performance is an emerging topic of research, building upon a large body of work in aided target recognition (AiTR) and visual psychophysics. The science of allocating limited visual attentional resources to detect, track, and act upon targets is being transformed by modern warfare where sensors are pervasive. A wide variety of use cases exist including mission command, crew stations, cockpits, night-vision goggles, and rifle scopes. The result is a tsunami of AI-augmented information (Lance et al. 2020; Su et al. 2021; Cox et al. 2022a, 2022b; Kase et al. 2022).

The human brain has limited capacity to sift through this information; it is evolutionarily adapted to simplify its task by actively trying to predict what information will be important at the next moment, to reallocate its sensory and attentional resources. Indeed, a variety of brain areas—including the frontal and

supplementary eye fields, as well as "top-down" processes from frontal cortex to lower visual areas—are engaged in preparing the eyes to capture the next useful bit of information. The top-down processes anticipate incoming visual patterns, simplifying the task of detecting mismatches between the retinal input and what we expect, whether the differences are in luminance contrast, feature contrast, or higher-order information. Thus, as we shift our gaze to search the environment, we are not only visually attracted to salient aspects of the environment, but also to features that share similarities to targets of interest and relevant to the task at hand, or that differ markedly from the background patterns we expect. Framed in this context, the attentional attractiveness of the information provided by AiTR does not have a static value, nor is it simple to calculate—its strength depends on both background and task context, as well as the user's moment-to-moment expectations.

This squishy, inexact nature of how AiTR highlighting can interact with attentional reallocation remains more art than science. Several types of highlighting have been considered: dark crosshairs, bounding boxes, circles, arrows, and disks. For acuity-sensitive tasks, crosshairs are necessary for precise alignment. However, when the target is in the visual periphery and hidden among the background clutter, the reduced spatial sensitivity and increased temporal and luminance contrast sensitivity of peripheral vision (at both retinal and cortical levels) favors highlighting that is lower spatial frequency with some temporal energy. In this sense, flickering Gaussian blobs or disks are an obvious choice for attracting attention beyond 5° eccentricity (Larkin et al. 2019, 2020a, 2020b).

A physical problem arises for AR systems, besides the obvious questions of the size and temporal dynamics of the highlighting: how exactly should this highlight be displayed? For many years, head-mounted AR devices were designed for indoor environments and focused on displaying interactive objects and avatars. To prevent "ghosting" (the appearance of a semitransparent virtual object floating in front of the background), it was necessary to mask out the background based on the object's outline. Thus, in commercial off-the-shelf AR systems, visual objects were displayed with exact luminances regardless of (or anchored to) the background scene. This is fine for indoor use where scenes are well-lit, and the goal is simply to have the virtual object appear lifelike. However, AiTR visual search performance depends on whether the highlighting achieves three things:

1. It must help to find unseen AI-identified targets (minimizing errors of missing an AI-identified target; called a "type 2" error).
2. Simultaneously, it should avoid harming the baseline ability to find targets without the aid of highlighting (minimizing errors of missing a target that AI also missed; also "type 2").

3. It must also avoid overallocation of attentional resources to highlighting of AI-misidentified targets (a "type 1" error) or to only a subset of true highlighted targets (attentional capture).

Designing and titrating the strength of the highlighting thus becomes a critical challenge to AR search performance. The challenge is exacerbated by the movement of the head and eyes, which can often result in 100-fold changes in luminance in the retinal image (high dynamic range [HDR] visual search). Although such huge luminance changes would seem to render moot the idea of fine titration of attention, we previously showed that visual acuity remains virtually unaffected even after such large luminance changes and even for low-contrast letters (Hung et al. 2021). Based on the early visual pathway's resilience to large, abrupt changes in luminance, we previously suggested that divisive display augmented reality (ddAR) (i.e., manipulating highlight transparency rather than highlight luminance) may be an effective and economical approach to titrate attention in AR-based AiTR (Hung et al. 2021). The reasoning behind manipulating transparency was that it would enable finer and computationally simpler control of highlight contrast, unlike luminance-defined highlights that would be prone to accumulated errors from eye and head movements, system delays, and other limitations.

Figure 1 shows examples of these phenomena for additive highlighting (where luminance is added, top image) and ddAR (bottom image). Additive highlighting over the right target (in clearing) is effective for drawing attention to a true target, but its distracting luminance contrast may lead to attentional capture, harming the ability to see an unhighlighted target (left target, under trees). With ddAR (bottom image), luminance is reduced by titrating transparency, providing finer control.

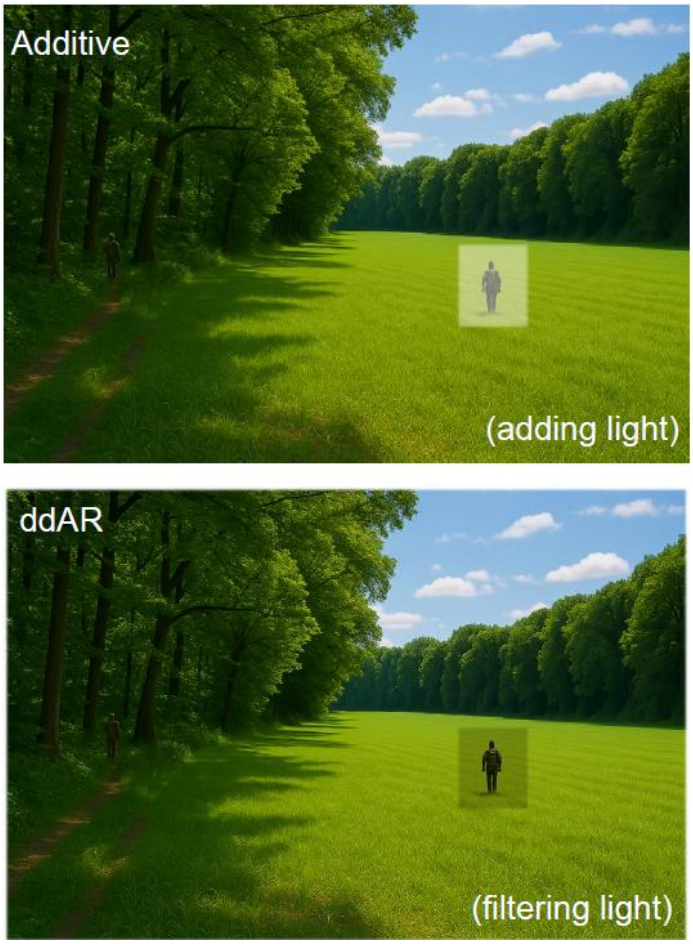


Figure 1. Example of attentional capture in additive vs. divisive display highlighting for AR.

1.2 Specific Aims

We have developed a new ddAR test bed and characterized how target-highlighting parameters affect visual search performance across both accurate and misleading highlighting conditions. We hypothesized that as highlight contrast increases, it also produces attentional capture, but that there may be a "sweet spot" of highlight contrast for visual search performance. To test this, we varied target-highlighting properties and compared the resulting benefits and costs on target-detection performance. To be consistent with the expected use case of a head-mounted display to aid in searching for a distant target, we designed a search task in which zero, one, or two small targets may appear across six possible locations (up to one target per location). Across trials, a highlight was presented at one correct target, one of the nontarget locations ("misleading"), or no highlight was presented. To mitigate against the possibility of the participant simply reporting the highlight location, the highlight was only helpful on half the trials, and each trial ended with auditory feedback about the correct number of targets.

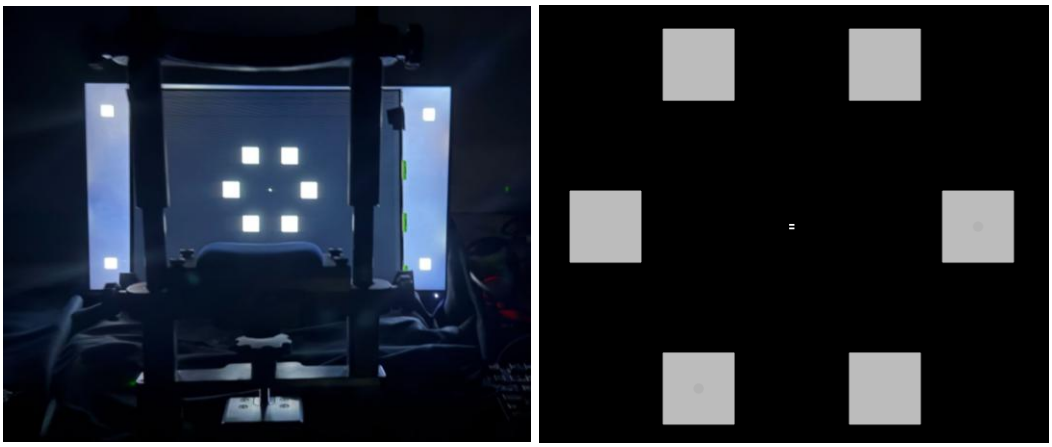
2. Methods

2.1 Subjects

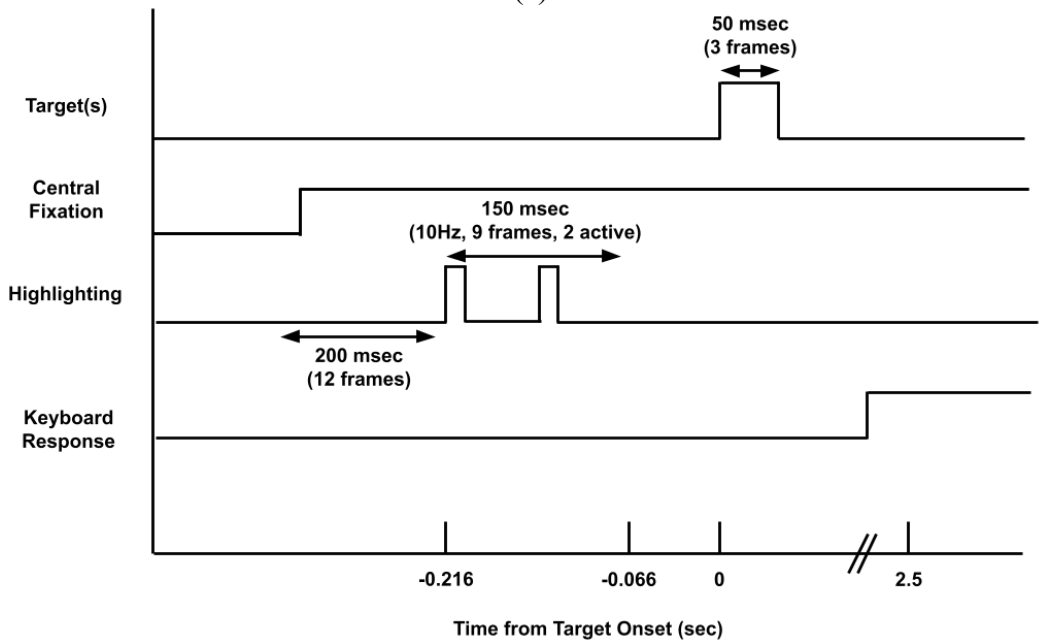
All tests were conducted according to protocol ARL 26-044 as approved by the Army Human Research Protections Office (Hung 2026). All three participants have been anonymized via random assignment to two-letter strings (KB, AU, and OY). Participants were male, right-handed, and ranged in age from 17 to 52. Each had normal or corrected-to-normal vision.

2.2 Experimental Setup

The experiment ran on a Dell D11S computer with Intel HD 530 Graphics displayed on a custom-made ddAR test bed (Figure 2a, left) combining an unmodified liquid crystal display (LCD) monitor (Dell U2417H, referred to as "target display") with a backlight-stripped LCD screen (LG Philips LM190E08-TLL1, referred to as "ddAR display"). The target display was 1920 × 1080 pixels (54.1 × 30.5 cm), corresponding to 78.6° × 49.6° at a 33-cm chin-rest–stabilized viewing distance. The ddAR display, positioned 2.5 cm in front of the target display, was 1280 × 1080 pixels (36.0 × 30.5 cm), corresponding to 61.1° × 53.1° at a 30.5-cm viewing distance. Both displays were 60 Hz (16.7 ms/1 frame) with 8-bit color and were controlled using PsychoPy software (ver. 2025.1.1) (Peirce et al. 2019). Target display luminance and gamma across all 256 luminance levels at uniform red, green, and blue nit guns were measured via Konica Minolta NT-1/3P luminance meter, and a similar process was used to measure transparency levels for the ddAR display. At full target-display luminance and full ddAR-display transparency, the maximum target display luminance was 11.6 cd/m^2 due to the ddAR display's 5% maximum transparency. The experiment was conducted in a windowless office with ambient luminance less than 0.04 cd/m^2 with four 30 cd/m^2 1.65° square luminance anchors centered at ±34.8° h, ±18.4° v (±850 h, ±400 v pixels) (Gilchrist et al. 1999).



(a)



(b)

Figure 2. Dimmable AR test bed, stimulus display, and timecourse.

2.3 Visual Search Task

Participants were tasked with detecting dimming target(s) on a set of six square 3.8° (80-pixel wide) "pedestals" arranged in hexagon formation, 8.7° eccentricity, 1 cd/m^2 while maintaining their gaze on a central fixation (Figure 2a, right: example shows targets at 3 and 7 O'clock positions). Targets were dimming disks at the centers of 1 or 2 pedestals and were 0.41° × 0.46° (10 × 10 pixels) radius, dimmed for 1 video frame (16.7 ms). In cases of two targets, the targets were dimmed simultaneously. At the start of the experiment, target luminance dimming was titrated near 10% for each participant until a baseline target detection rate of approximately 35%–50% was achieved without highlighting.

Each block consisted of 960 trials presented in pseudorandom order, comprising 4 possible conditions ("1-Target," 1 correct highlight; "2-Targets," 1 correct highlight; "0-Target," 1 unhelpful highlight; and "Misleading," 1 target, 1 misleading highlight) × 60 trials with variations of target and highlight positions × 4 variations of highlight contrasts (Block A: 0%, 5%, 10%, and 20%; Block B: 20%, 40%, 60%, and 100%) (Table 1). Each block required approximately 90–120 min and was broken into groups of 60 trials, 4–6 s per trial, with typically 1–2 min of rest between groups of trials and with bio breaks allowed. Participants were shown Block A before Block B. The two blocks were conducted on either the same day or two different days.

Table 1. Trial types.

Block	Highlight contrast (%)	Target and highlight conditions			
		"1-Target," 1 correct highlight	**"2-Targets," 1 correct highlight**	**"0-Target," 1 unhelpful highlight**	**"Misleading," 1 Target, 1 misleading highlight**
A	0	60 trials	60 trials	60 trials	60 trials
	5	60 trials	60 trials	60 trials	60 trials
	10	60 trials	60 trials	60 trials	60 trials
	20	60 trials	60 trials	60 trials	60 trials
B	20	60 trials	60 trials	60 trials	60 trials
	40	60 trials	60 trials	60 trials	60 trials
	60	60 trials	60 trials	60 trials	60 trials
	100	60 trials	60 trials	60 trials	60 trials

The timecourse of each trial began with a pair of central fixation bars at 2 cd/m^2 (6 × 6 pixels, 0.3° long, 0.3° separation), which rotated from vertical to horizontal after a variable delay of 1.5 to 3.5 s (Figure 2b). For each trial, 0, 1, or 2 targets were presented, with the highlighting appearing on the ddAR screen 217 ms (13 frames) before target onset. The highlight was presented as a 10-Hz (i.e., 1 frame on, 5 frames off, 1 frame on) flickering darkening disk on the ddAR display with 6.2° (120 pixel) diameter, obscuring the entire pedestal. Highlight contrast level was independent of target number or position (pseudorandom; equally distributed across all trials), but highlight position depended on trial type. In the 1-Target case, highlighting was correctly over a target, and in the 2-Targets case, highlighting was correctly over one of the two targets. The 0-Target and Misleading conditions had incorrect highlighting over pedestals without targets for contrasts above 0%; responses recording detection of targets under these false highlights are called hallucinations. Participants were told beforehand of these conditions and the percentage of trials they comprised (25% each, evenly distributed).

To record their perceived targets, participants responded by pressing one of six keys surrounding the G key (H, Y, T, F, V, and B) corresponding to each respective

pedestal; the G key itself was inactive. Participants could record two targets at most or indicate no targets (or no additional targets after already pressing a letter key) by pressing the space bar to advance to the next trial. Response latency was recorded as the time from target onset to the first keypress. At the end of each trial, an audio tone indicated how many targets were present, providing instant feedback (high tone = 2 targets, mid tone = 1 target, low tone = 0 targets).

2.4 Analysis

Responses and the benefit function were fit via a Weibull function to capture asymmetries in the tails of the distributions (Britten et al. 1992). The fitted psychometric function ψ was

$$\psi(c) = \gamma + (1 - \gamma - \lambda)\,[1 - e^{(-(c/\alpha)^{\beta})}]$$

where c is highlight contrast (%), α is a scale parameter representing the contrast threshold (in % units) at which a certain level of performance is expected, β controls steepness, γ is the floor performance at $c = 0$ (chance/unaided detection), and λ is an upper asymptote. The Weibull function can rise more steeply on one side of threshold and flatten asymmetrically, which better matches how hit rates rise steeply with contrast to reach saturation.

3. Results

3.1 Detection Performance for 1-Target and 2-Target Conditions

We examined target detection performance as a function of highlight contrasts ranging from 0% to 100% across the 1-Target and 2-Target conditions with helpful highlighting at 1 target location (Figures 3 and 4). Overall, increasing contrast improved highlighted target detection up to 20% highlight contrast where correct performance saturates. Costs have a less predictable trend, being low at approximately 20% contrast but worsening significantly at 40% contrast (Figure 3). Individual subject detection rates, benefits, and costs, are expressed as mean ± standard error of the mean. Dots indicate the score for each participant.

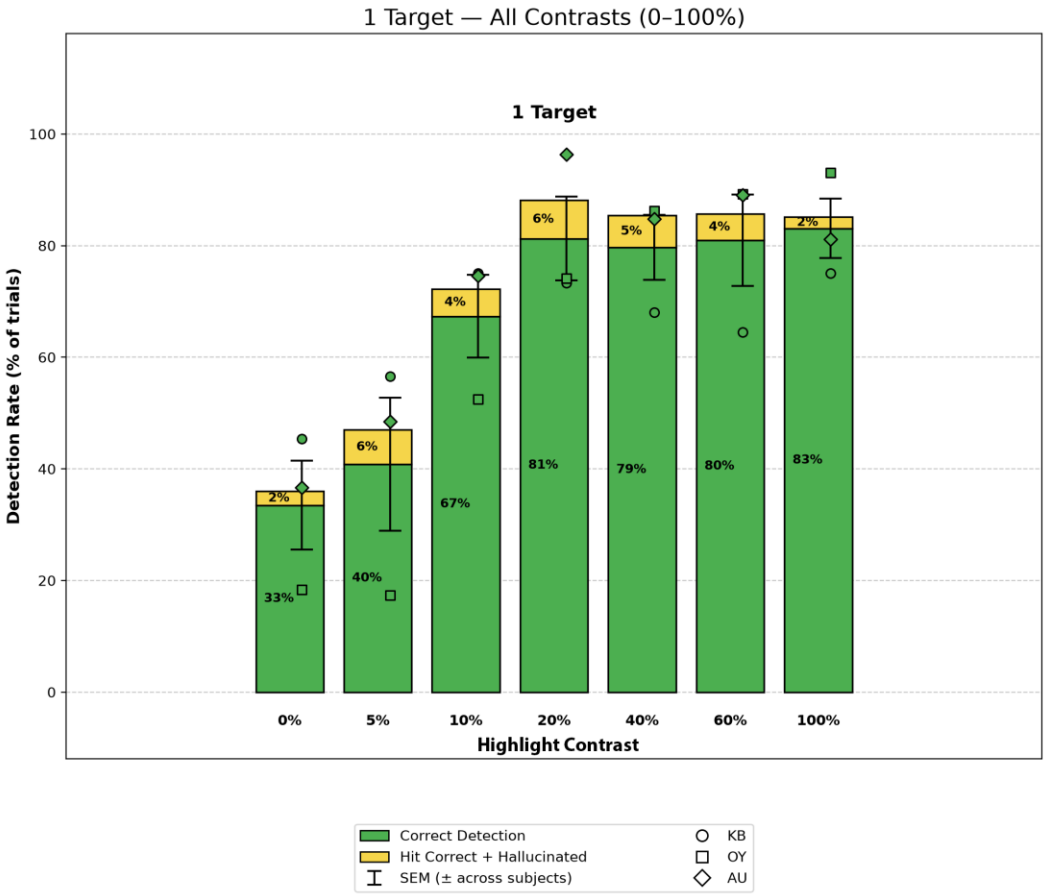


Figure 3. Highlight contrast vs. detection performance for 1-Target trials.

We observed an improvement in 1-Target correct performance (shown in Figure 3's green bars) across all participants between 0% and 20% contrast, with accurate detection rate increasing from 33% to 81% detection. Within each participant, the performance improvement was weaker between adjacent contrast levels (between 0% and 5%, between 10% and 20%), with overlapping SEMs and subject-specific location biases (see Appendix). Above 20% contrast, 1-Target performance remained saturated at 79%–83% (Figure 3). Hallucinations (responses at nontarget locations) were present across all contrast levels. A Weibull fit to Correct Detection + Hallucinated (green + yellow) curves yielded a threshold near 9% highlight contrast ($\alpha \approx 9\%$), with performance saturating at 20% contrast.

In 2-Target trials (Figure 4), the pattern is similar with an increase in highlighted target detection (orange + green bars) that saturated beyond 20% contrast, increasing from 30% (green + 0.5 × lime) to 80%–86%. For detection of any target (cream + orange + green), detection performance also saturated beyond 20% highlighting. Attentional capture, defined here as trials in which only the highlighted target was reported but the unhighlighted target was missed, emerged

with as little as 5% highlight contrast. At 5% highlight contrast, attentional capture (27% of trials) was more likely than detection of just the unhighlighted target (14%), but the overall effect of the 5% highlighting was a weak, nonsignificant boost to target detection (51%, green + orange + cream) compared with no highlighting (48%, green + lime). At 10% and 20% highlight contrasts, attentional capture increasingly cannibalized from nonhighlighted target detection, but the overall target detection rate nearly doubled versus baseline (88% detection rate for 20% highlight contrast, versus 48% detection rate at baseline). A Weibull fit to 2-Target highlighted target detection (orange + green) yielded a threshold near 10% highlight contrast ($\alpha \approx 10\%$), with performance also saturating at 20% highlight contrast.

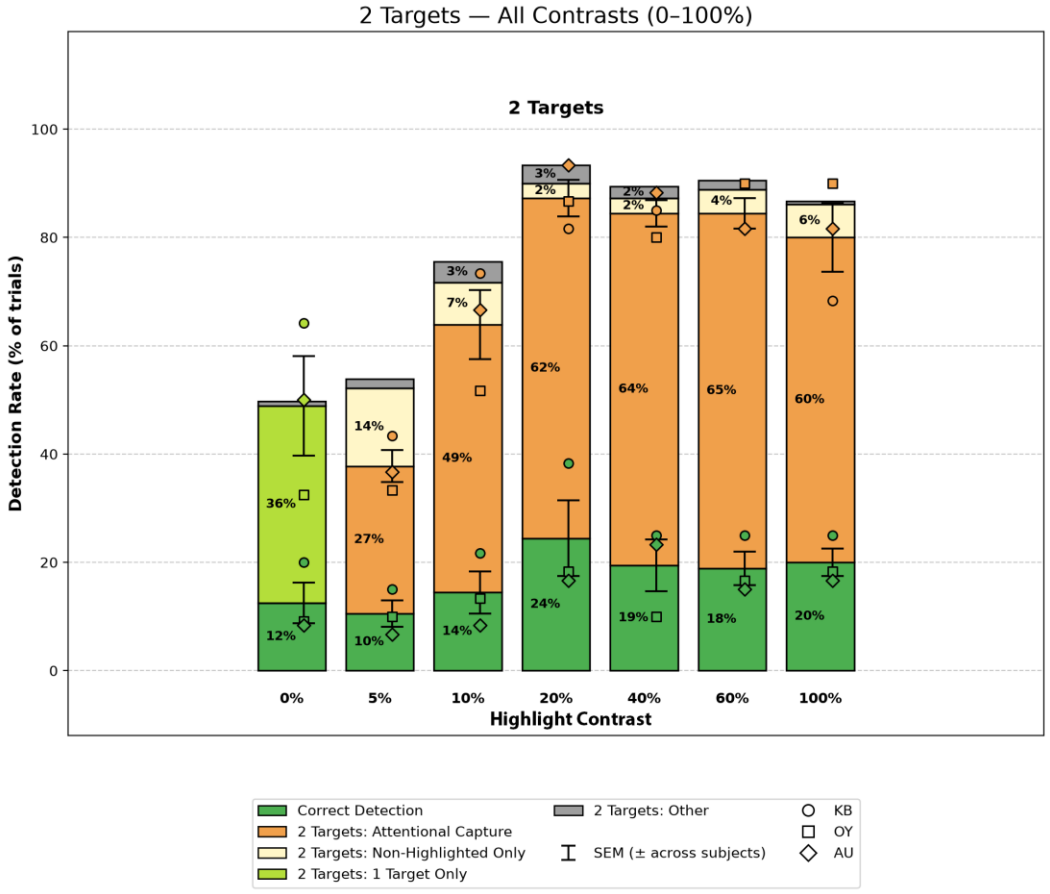


Figure 4. Highlight contrast vs. detection performance for 2-Target trials.

3.2 Analysis of Malfunctioning Highlighting Conditions

Randomly interspersed among the 1-Target and 2-Target conditions, we also tested for correct rejection in a 0-Target condition (Figure 5) and false positive reporting of nontarget(s) in a misleading highlight condition (Figure 6). Both the 0-Target condition and the misleading highlight condition emulate examples of possible

AiTR malfunction, where highlighting does not coincide with an active target location. Without a true target present (0-Target, Figure 5), correct rejection rates ("Detection Rate," i.e., one minus the false positive rate) decrease steeply as misleading highlighting contrast increases, resulting from an increase in hallucinations (see the Appendix), falling from 96% correct at 0% highlight contrast (i.e., no highlight) down to approximately 28%–37% correct at ≥20% highlight contrast, driven by incorrect false reports. Figure 6 illustrates a similar trend with an increase in hallucination rates (yellow + blue) going from approximately 2% at 0% highlight contrast to approximately 61%–69% trials having a hallucinated target at ≥20% highlight contrast.

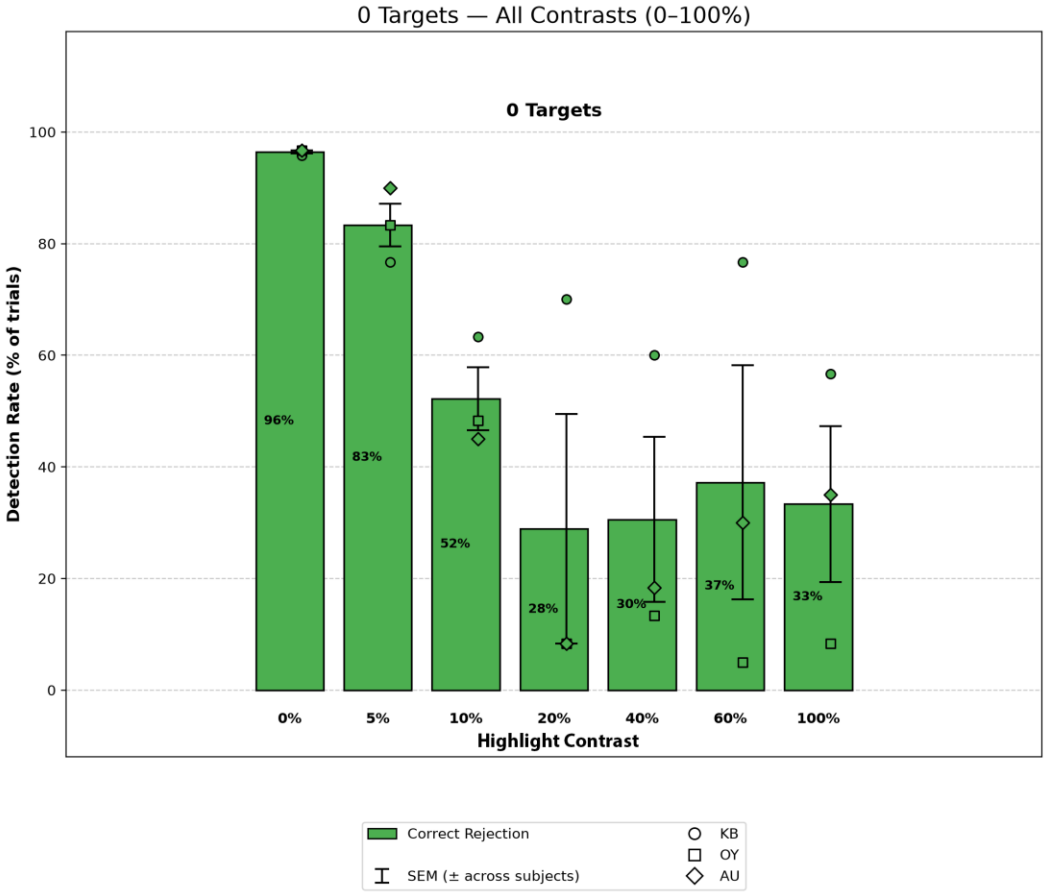


Figure 5. Highlight contrast vs. correct rejection for 0-Target trials.

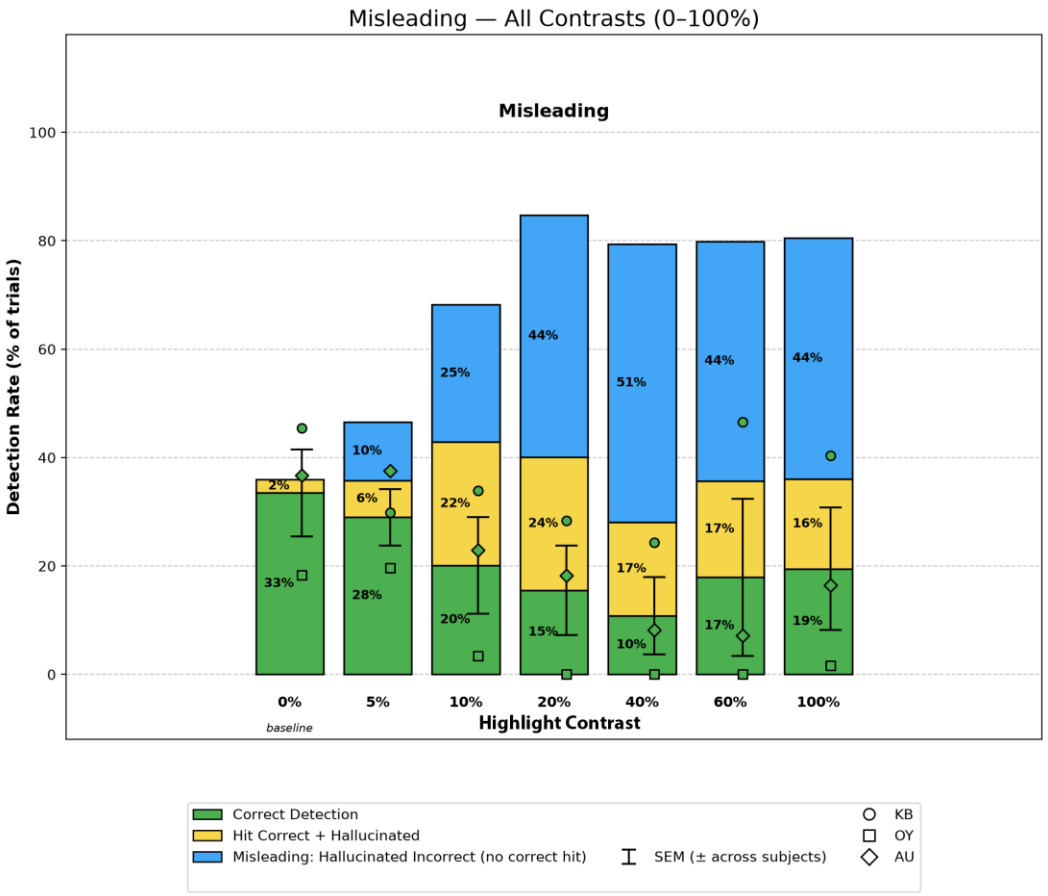


Figure 6. Highlight contrast vs. false positive reports for misleading trials in which one target was present but the highlight was on a nontarget pedestal.

When highlighting correctly helps (Figures 3 and 4), hallucination rates are present across all contrasts (yellow in 1 Target and gray in 2 Target; excludes attentional capture, which is a missed second target rather than a false report) without a strong relationship connecting them. When one target is present but the highlight is at a nontarget pedestal (Misleading condition), hallucination rates (blue) also increase; correct nonhighlighted target identification (green + yellow) fluctuates without clear pattern across participants.

3.3 Cost–Benefit Analysis

We generated cost and benefit functions each with two components to calculate best performance. Highlighted target detection across the 1-Target (green + yellow) and 2-Target (green + orange) conditions were averaged to produce a mean benefit curve via a Weibull fit (Figure 7, green). We also calculated a cost function with two parameters: unhighlighted target detection rate reduction in the 1) 2-Target and 2) Misleading conditions (Figure 7, red). A tradeoff score was generated per contrast level: tradeoff(c) = benefit(c) – benefit_at_0% - cost(c), which peaked at

20% highlight contrast. The highest cost benefit difference is at 20% highlight contrast (Figure 7, dotted vertical line). Across the three participants, the individual tradeoff comparisons also peaked at 20% highlight contrast for two participants but were at 60% for participant KB, due to surprisingly low cost at the strongest highlight contrasts. Figure 7 shows combined cost for 20% contrast is near zero and significantly larger at 40%, and the benefit curve begins to asymptote near 20% contrast. Subtracting cost from benefit across all contrast levels revealed 20% contrast as the best.

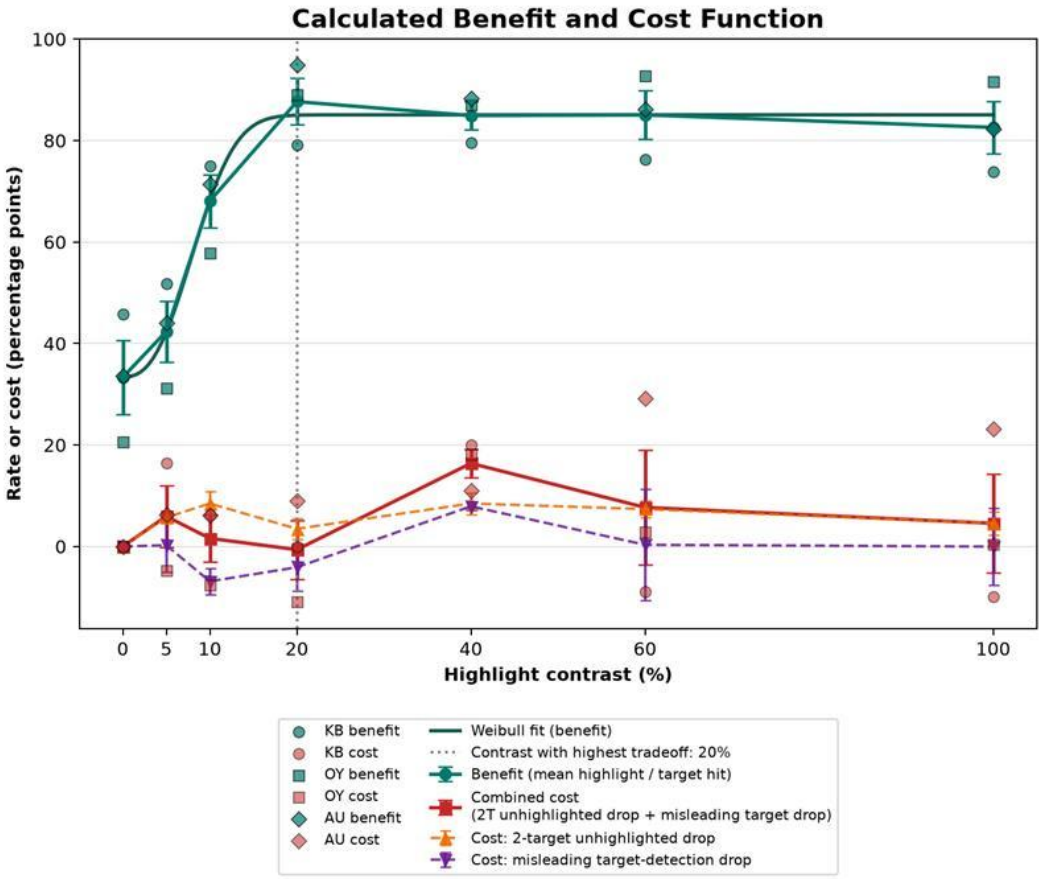


Figure 7. Costs vs. benefits of each contrast level.

4. Discussion and Conclusion

A 20% highlight contrast produced the clearest benefit for visual search performance, based on the tradeoff between the benefit of highlighting for target-detection accuracy versus the cost of false alarms and attentional capture. This 20% threshold was true for both 1-Target and 2-Target search and was consistent across two-thirds of the participants, with a third participant having a higher threshold due to reduced costs at strong highlight contrasts. A 20% highlight contrast should be

easy to produce via ddAR transparency, as today's AR systems already include dynamic dimming to enable real-world outdoor performance.

Besides the main conditions reported here, we also pilot-tested other variations of highlight size (3.6° diameter instead of 6.2°) and temporal frequency (20 Hz instead of 10 Hz). Because they did not produce substantial changes in the results, we chose not to explore them further.

The generalizability of this result will likely depend on the circumstances of the search conditions. In our case, the participants were told that the highlighting would be inaccurate in 50% of trials. We chose to balance trials across accurate versus inaccurate highlighting and to provide audible end-of-trial feedback, so that participants would not be biased toward simply reporting the highlight location or ignoring the highlighting. Future AiTR systems with higher or lower accuracy would likely shift this ideal contrast threshold upward or downward. Also, our target and highlight locations were in the near-periphery (between 5° to 10° eccentricity). Due to differences in photoreceptor sensitivity, cortical representation, and limits of the field of view, other conditions may require different highlighting contrasts, sizes, or temporal energy, or even multisensory integration for targets outside the field of view. Additionally, the eyes remained fixated on a central point in our study, whereas the background and task contexts may exert a greater influence in free viewing. Furthermore, strong abrupt luminance changes can result in grouping effects that may interact with the effectiveness of highlighting (Hung et al. 2020a, 2020b). Finally, detection performance saturated at 20% highlight contrast under our stimulus conditions. With more high-contrast or energetic backgrounds (e.g., foliage, waves, rocks, urban scenes), it is possible that performance would continue to improve with stronger highlight energy, shifting the cost–benefit curve rightward.

5. References

Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci. 1992;12(12):4745–4765.

Cox KR et al. A summary of the effects of technology aids and crew size on a next generation combat vehicle platoon section. DEVCOM Army Research Laboratory (US); 2022a Mar. Report No.: ARL-TN-1109.

Cox KR et al. Effects of technology aids and crew size on a next generation combat vehicle platoon section. DEVCOM Army Research Laboratory (US); 2022b Feb. Report No.: ARL-TR-9403.

[DoD] Department of Defense. National defense science & technology strategy. Department of Defense (US); 2023 May 9. https://media.defense.gov/2023/May/09/2003218877/-1/-1/0/NDSTS-FINAL-WEB-VERSION.PDF

Gilchrist A et al. An anchoring theory of lightness perception. Psychol Rev. 1999;106(4):795–834.

Hung CP. Visual attention with dynamic diming augmented reality. Exemption Determination. DEVCOM Army Research Laboratory Human Research Protection Program (US); 2026 Aug. Report No:. ARL 26-044.

Hung CP et al. A 100,000-to-1 high dynamic range (HDR) luminance display for investigating visual perception under real-world luminance dynamics. J Neurosci Methods. 2020a;338: 108684. https://doi.org/10.1016/j.jneumeth.2020.108684

Hung CP et al. Abrupt darkening under high dynamic range (HDR) luminance invokes facilitation for high-contrast targets and grouping by luminance similarity. J Vision. 2020b;7:9. https://doi.org/10.1167/jov.20.7.9

Hung CP et al. Low contrast acuity under strong luminance dynamics and potential benefits of divisive display augmented reality (ddAR). J Percept Imaging. 2021;4(1):010501-1. https://doi.org/10.2352/J.Percept.Imaging.2021.4.1.010501

Kase SE et al. The future of collaborative human-AI decision-making for mission planning. Front Psychol. 2022;13:850628. https://doi.org/10.3389/fpsyg.2022.850628

Lance BJ et al. Minimizing data requirements for soldier-interactive AI/ML applications through opportunistic sensing. SPIE Defense + Commercial Sensing 2020: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II. 2020;11413:1141306.

Larkin GB et al. Augmented target recognition display recommendations. DSIAC J. 2020a;7(1):28–34.

Larkin GB et al. Emerging recommendations for initial automated target recognition (ATR) display considerations: implications for concept development and evaluation. PsyArXiv; 2019 Aug 15. https://doi.org/10.31234/osf.io/w9qge

Larkin GB et al. Human performance with complex technology: How visual cognition is critical to enhanced performance with aided target recognition (AiTR). In: Ayaz H, Asgher U, editors. Advances in Neuroergonomics and Cognitive Engineering. Virtual; 2020 July 16–20. Springer. 2020b;1201:134–140.

Peirce JW et al. PsychoPy2: experiments in behavior made easy. Behav Res Methods. 2019;51:195–203.

Su S et al. Mixed reality visualization of friendly vs hostile decision dynamics. Proceedings of the Virtual, Augmented and Mixed Reality: 13th International Conference, VAMR 2021, held as part of the 23rd HCI International Conference, HCII 2021. Virtual; 2021 July 24–29. Association for Computing Machinery; 2021.

Appendix. Supplemental Material

A-1. Hallucinations

Under the Misleading and 0-Target conditions, target highlighting can induce errors in the form of false alarms (e.g., response to nontarget pedestals) and misses of true targets (i.e., type 2 error). Cluster-robust logistic regression showed a significant positive relationship between hallucinations and Misleading highlight contrast ($p < 0.001$) and in 0-Target false-alarm trials ($p = 0.018$). Hallucination rates on helpful trials did not show a comparable increase. A contrast × misleading interaction on 1-Target trials confirmed that the contrast effect was significantly steeper when the highlight was misleading than when it was helpful ($p = 0.016$), indicating that elevated hallucination rates at higher contrasts are selective to misleading (and no-target) conditions rather than a nonspecific effect of highlight intensity.

A-2. Radial Bias

Through pseudorandomization, radial target location and highlight location were evenly distributed across all trials. A chi-squared test of hit versus miss counts (pooled across all contrast levels in 1-Target trials) by position yielded a significant relationship ($\chi^2(5) = 55.07$, $p = 1.3\times10^{-10}$). Thus, a bias towards Positions 0 and 3 (0 and π radians, counterclockwise from rightmost pedestal) can be observed in Figures A-1a and A-1b with pooled detection rates being 81% and 76%, respectively, while others range from 57% to 62% (bars show means ± standard errors of the mean [SEM] across all contrasts). This is most clearly prevalent in the low-contrast cases (0%, 5%), dropping from 65% and 55% in Positions 0 and 3 respectively down to just 15% in Position 5. Interestingly, the mean detection rate at 100% contrast fell below the mid-contrast peak (often 20%) at several positions, though SEMs frequently overlap. Average SEM magnitude was also noticeably larger for Position 3 compared with other positions, driven by larger participant spread (symbols).

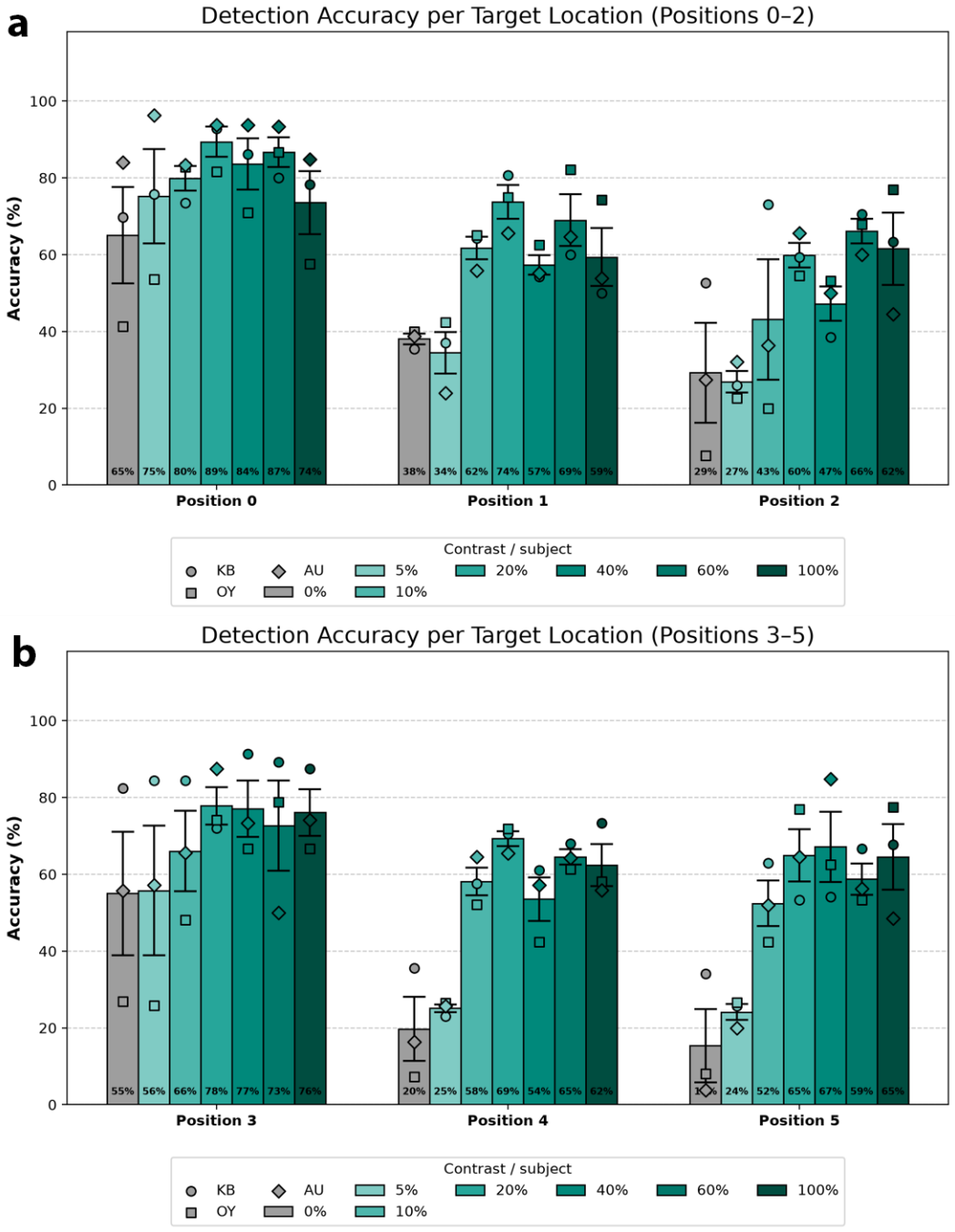


Figure A-1. Detection accuracy across target locations.

A-3. Participant Bias

Overall detection/correct-rejection accuracy differed across participants in each trial type. On 1-target helpful trials, subject × correct/incorrect was significant ($\chi^2(2) = 18.53$, $p = 9.5\times10^{-5}$), with pooled full-correct rates of 67.7% (AU), 62.9% (KB), and 54.4% (OY). On 2-Target helpful trials, the subject effect was also

significant ($\chi^2(2) = 26.62$, $p = 1.7\times10^{-6}$), with KB showing a higher pooled full-correct rate (23.8%) than OY (13.1%) and AU (12.9%). Zero-target correct-rejection performance also showed large between-subject spread (KB often most conservative at mid-to-high contrast; OY often least). Critically, rank order across contrast levels was not stable on 1-Target trials (Kendall's coefficient of concordance $W \approx 0.02$, $p = 0.87$[1]), so these differences are best described as individual variability that widens SEM rather than a single consistently superior participant.

A-4. Reaction Time Analysis

Reaction times were measured as time from target onset to first keypress. Trials with reaction times greater than 2.5 s were excluded as likely trials where the participant missed target onset until the trial pass reminder occurred at 2.5 s. For each participant, mean reaction time was computed within each condition (1-Target, 2-Targets, 0-Target, and Misleading) and contrast level; group summaries are means ± SEM across the three subject means (Figure A-2). Mean reaction time generally decreased as highlight contrast increased, most clearly on 2-Target trials (from 0.94 ± 0.09 s at 0% to 0.73 ± 0.07 s at 60%). Trials in which reaction time exceeded 2.5 s were excluded.

We fit cluster-robust ordinary least squares regressions of reaction time on contrast (clustered by subject); only the 2-Target condition was significant (slope = −0.16 s per 100% contrast, p = 0.037). The other cases showed similar downward trends that did not reach significance (p's ≥ 0.063). Overall, stronger highlight contrast reduced response latency, mainly when two targets competed for attention.

[1] Siegel S, Castellen Jr NJ. Nonparametric statistics for the behavioral sciences. 2nd ed. McGraw-Hill Book Co; 1988.

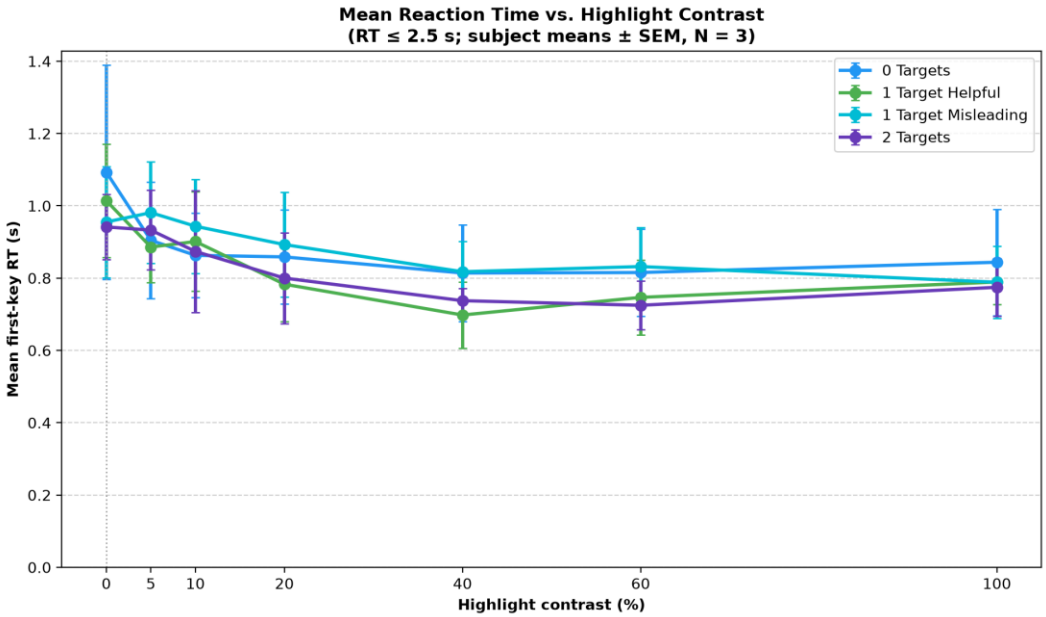


Figure A-2. Mean reaction time vs. highlight contrast.

List of Symbols, Abbreviations, and Acronyms

AI	Artificial Intelligence
AiTR	Aided Target Recognition
AR	Augmented Reality
ddAR	Divisive Display Augmented Reality
DoD	Department of Defense
HDR	High Dynamic Range
LCD	Liquid Crystal Display
SEM	Standard Error of the Mean

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CB TECH LIB