Using Visual and Auditory Cues to Locate Out-of-View Objects in Head-Mounted Augmented Reality

Nicola Binetti*

UCL Interaction Centre, University College London, London, United Kingdom

Luyan Wu

UCL Interaction Centre, University College London, London, United Kingdom

Shiping Chen

Department of Industrial Engineering, Tsinghua University, Beijing, China

Ernst Kruijff

University of Applied Sciences Bonn-Rhein-Sieg, Sankt Augustin, Germany

Simon Julier

UCL Interaction Centre, University College London, London, United Kingdom

Duncan P Brumby

UCL Interaction Centre, University College London, London, United Kingdom

Abstract

When looking for an object in a complex visual scene, Augmented Reality (AR) can assist search with visual cues persistently pointing in the target's direction. The effectiveness of these visual cues can be reduced if they are placed at a different visual depth plane to the target they are indicating. To overcome this visual-depth problem, we test the effectiveness of adding simultaneous spatialized auditory cues that are fixed at the target's location. In an experiment we manipulated which cue(s) were available (visual-only vs. visual+auditory), and which disparity plane relative to the target the visual cue was displayed on. Results show that participants were slower at finding targets when the vi-

*Corresponding author
Email address: nicolabinetti@gmail.com (Nicola Binetti)

sual cue was placed on a different disparity plane to the target. However, this slowdown in search performance could be substantially reduced with auditory cueing. These results demonstrate the importance of AR cross-modal cueing under conditions of visual uncertainty and show that designers should consider augmenting visual cues with auditory ones.

1. Introduction

Many potential applications of AR, such as maintenance, repair, instructions and training, involve *visual search* [1, 2]. Visual search tasks require people to look for a target within the surrounding environment [3, 4]. For example, a user might be tasked with finding and interacting with real-world or virtual targets located in the scene.

Visual search in AR can be assisted using a variety of visual cueing techniques. Head-fixed virtual pointers, such as arrows, wedges or halos [5, 6, 7, 8, 9, 10], can be intuitive and effective ways of visually guiding users towards task relevant information. However, the effectiveness of visual cueing in binocular AR Head-Mounted Displays (HMDs) relies on cues being displayed at the same visual depth as the target they are pointing to. While this is generally ensured in devices that leverage spatial mapping and ray casting, there can be instances in which cue / target (CT) depth discrepancy can arise.

Factors that contribute to CT depth discrepancy include: modelling errors (i.e. incorrect *a priori* knowledge of the depth at which targets might be located), measurement errors (i.e. uncertainty with the measurement of real-world depths), calibration errors (i.e. optics that have to be properly aligned or calibrated), and/or depth perception errors due to vergence / accommodation conflicts. Importantly, all current consumer based AR HMDs such as the Microsoft HoloLens 1 and 2, and the Magic Leap One are susceptible to CT depth discrepancy. Another major factor is the actual view management method used for placing augmented information, in which label placement is often optimized to reduce clutter and improve visibility and legibility [11, 12]. For example, this often results in labels or other cues that are not directly co-located with the object they refer to.

Previous research has demonstrated that depth discrepancies in AR systems can result in context switching, in which a user has to alternate between the virtual and real-world content positioned at different depths. This can be a source of visual discomfort [13] and fatigue [14], and can negatively impact visual search performance [15]. However, to date, depth discrepancy is an overlooked issue in relation to the design of AR visual search cues.

Cross-modal approaches have been employed as solutions for aiding search tasks under conditions of sensory uncertainty [16, 17, 18]. Auditory cueing is effective at directing behaviour [19, 20, 21], producing strong attentional orienting responses [22], and can aid visual search [13, 16] without cluttering the visual scene or impacting on visuo-attentional processing [23].

In this paper, we characterized the impact of depth discrepancy on the effectiveness of a visual cue in guiding users towards a visual target. We also explored the effectiveness of simultaneous spatialized auditory cueing in aiding search performance under conditions of visual depth discrepancy.

1.1. Problem Space

To illustrate the problem, consider the scenario shown in Figure 1. An AR user must locate and discriminate an out of view visual target which is set amongst potential distracting stimuli. The AR display uses a virtual arrow (cue) that continually points towards the target [7, 8]. We assume that, while target and distractors are at the same depth relative to the observer, the arrow cue is displayed closer to the observer (Figure 1a). Given the difference in cue / target (CT) disparity planes, fixating on either stimulus will require a different amount of left / right eye vergence [24]. As shown in Figure 1b, the vergence is smaller when the target is further away than the cue. If the user fixates on the cue, the target will appear diplopic (i.e. 'double'), conversely if the user fixates on the target, the cue will appear diplopic as well [15, 25]. In order to integrate cue and target position information requires users to switch back and

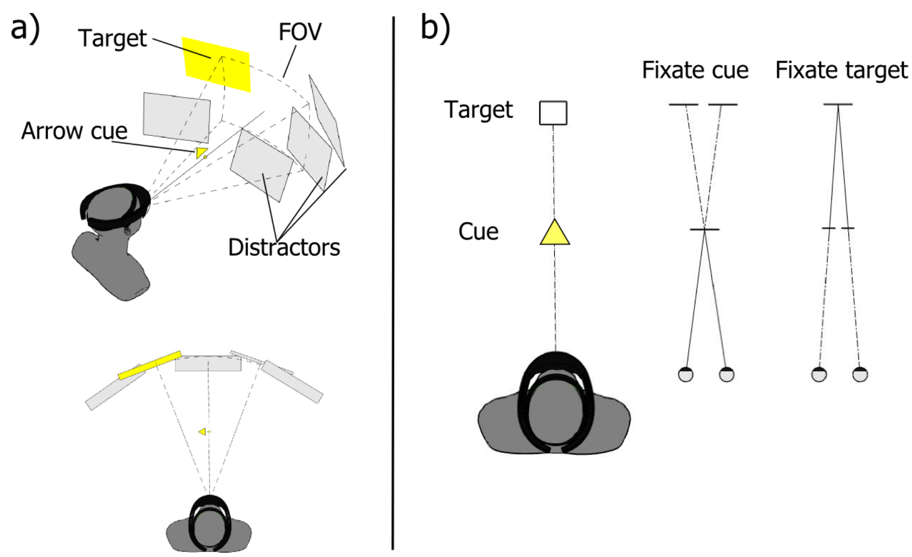

Figure 1: Cue target depth plane discrepancy. a) the AR user must locate and identify a visual target (yellow) set amongst potential distractors. A virtual arrow (cue) facilitates the task by persistently pointing in the target's direction. While target and distractors are located at the same depth relative to the observer, the arrow cue is displayed on a closer depth plane. b) Fixating on a cue or target at different depths require different degrees of eye vergence. Fixating on the cue will impair visibility of the target (vergence overshoot), whereas fixating on the target will impair visibility of the cue (vergence undershoot).

forth between disparity planes [26, 27].

The scenario illustrated in Figure 1 raises two questions. First, to what extent does depth discrepancy reduce the cue's effectiveness in guiding the user towards the target? Second, what measures can be put in place to counteract or minimize any potential disruption to the effectiveness of this cue?

1.2. Contributions

The study reported in this paper makes two contributions. First, we provide the first quantification of the impact of CT depth discrepancy on AR visual search performance. We do this by measuring the effectiveness of the cue in assisting users to find and discriminate a visual target that has been placed amongst task-irrelevant distractor stimuli. Visual depth discrepancy is known to induce fatigue and visual discomfort [25]. We find that depth discrepancy significantly slows down AR cued visual search performance. This clearly demonstrates that CT depth discrepancy can significantly limit the effectiveness of AR visual cueing interfaces.

The second contribution is to test a combined (cross-modal) visual-auditory cueing approach as a way to mitigate the costs incurred by CT depth discrepancy. Spatialized auditory cues that are fixed at the target's location are effective at directing user attention [19, 20]. Most importantly, auditory cues are likely unaffected by costs in visual context switching. These properties make simultaneous spatialized auditory stimuli a candidate solution to alleviate costs in visual search performance determined by visual CT depth discrepancy. We find that the slowing in visual search could be substantially reduced with auditory cueing, approaching performance observed with no depth discrepancy. This finding encourages the use of cross-modal AR cueing interfaces to bypass limitations caused by visual depth discrepancy.

2. Related Work

In this section, we describe previous work and concepts relevant to the current study. These are: visual search tasks and performance, visual search and visual cueing in AR, and the constraints introduced by depth discrepancy in AR on visual search performance.

2.1. Visual Search Tasks

Visual search is the process by which an observer has to locate a target embedded in the surrounding environment. It has been extensively investigated in vision and attention research [28], as well as in the context of interface design [29]. The need for visual search is a consequence of the physiological and computational constraints of visual, attentional and working memory systems. Humans have a vertical FOV of about 150 degrees of visual angle, and a horizontal binocular FOV of about 210 degrees of visual angle [30]. Furthermore, foveated (high visual acuity) vision is limited to a narrow 2 degrees of visual angle [31]. Despite the possibility of pre-attentively processing peripheral visual information [32], the narrow cone of foveated vision limits the observer's ability to simultaneously process most of the information in the visual field. These limitations in the visual system mean that search strategies must be used. The performance of this search strategy depends upon the number and type of distractors. When the features for the target and the distractors are very different, observers are capable of rapidly spotting the target, which 'pops-out' from the surrounding environment [4, 33]. In these instances, search time is not modulated by number of stimuli, revealing a form of parallel visual processing [34]. However, when the target and distractor features are not distinct from one another, search times tend to be longer because there is competetion for observer attention [35].

2.2. Visual Search in AR

When performing visual search in AR, HMDs introduce further constraints on search performance. Commercially available see-through AR headsets typically provide a horizontal FOV of 30-43 degrees of visual angle [36, 37]. Narrow augmented fields of view reduce the observer's awareness of surrounding virtual information, limiting visual search performance [2, 38, 39]. In contrast, widening the FOV to very wide (100 degree+) has been shown to be advantageous for search [38], while also the design and motion parameters of labels may contribute to the awareness of information in the periphery [40]. AR tasks involving visual search therefore require interfaces that have been designed to work around these limitations. Some approaches are aimed at capturing user attention via non-visual cues [41] or annotations and labels with features aimed at leveraging pop-out effects and parallel visual processing [42, 43]. While effective, these forms of cueing can potentially lead to increase in visual clutter.

Alternative approaches have explored more subtle forms of cueing, enhancing the target through less obtrusive lightness and colour contrast saliency modulation techniques [1, 44]. Yet more approaches have been designed to actively guide the user towards task relevant information. These include head-fixed visual arrows, wedges, halos [5, 6, 7, 8, 9, 10], 3D 'tunnelling' interfaces [45], or peripheral visual cues [46], that persistently point the user in the target's direction. Search tasks can also be guided through other sensory modalities. For example, 3D auditory cueing (i.e. originating from the target's location in 3D space) is effective at directing behaviour [19, 20, 21], producing strong reflexive attentional orienting responses [22], without cluttering the visual scene or impacting on visuo-attentional processing [23]. Vibrotactile cues have also been successfully exploited in navigation and visual search tasks [47, 48]. However, non-visual forms of cueing also prove less effective for precisely discriminating a target location amongst distractors in cluttered visual spaces [49, 50]. Finally, cross-modal approaches, combining audio-tactile [37, 51, 52] and audio-visual [8] signals can provide effective forms of behavioural cueing, by combining directional information across modalities [13]. While these studies focus on the general advantage of combining different sensory cues, our study more specifically deals with the question of weather a spatialized auditory cue can bypass limitations in visual cueing.

2.3. Depth Perception, Vergence, and Depth Discrepancy

There are many visual cues that contribute to depth perception in AR, including occlusion, relative size, motion paralax, accomodation and binocular visual disparity [53]. In our setup, the depth of cue, target and distractors was mainly conveyed through binocular disparity. Binocular disparity arises from the fact that, when an observer views a nearby stimulus, this projects horizontally shifted left and right retinal images. Compensatory eye rotations (vergence) are performed to properly fixate on the stimulus and fuse the images [24]. Accommodation arises from the fact that the eye can vary its convexity, changing the depth at which objects will appear in focus. Some methods to resolve accomodation-vergence problems in head-mounted displays have been proposed before. An overview of these methods can be found in [54]. Examples of techniques can be hardware-based, such as multi-focal displays that use deformable membranes, tunable lenses, parallax barriers or pinlight displays (e.g., [55]), or software-based, such as gaze-contingent blurring [56]. These methods to a certain extent simulate or support the perception of different focal distances (disparity planes). However, multi-focal displays that cover the full depth range (instead of two focal planes like in the Magic Leap) are still not commercially available as they are mostly lab prototypes. Furthermore, while gaze-contingent blurring provides additional depth cues, it does not overcome the accomodation-vergence conflict when used with non multi-focal displays.

If tasks require integrating information carried by stimuli located at different depths (e.g. applying text instructions contained in a label to a separate target stimulus, or integrating cue and target spatial information), users will have to repeatedly switch across depth planes [26]. These switches will in turn introduce access costs (time to rotate eyes, and accommodate to different planes when alternating between real-world and virtual stimuli), with measurable impacts on AR usability [1, 15, 20, 27].

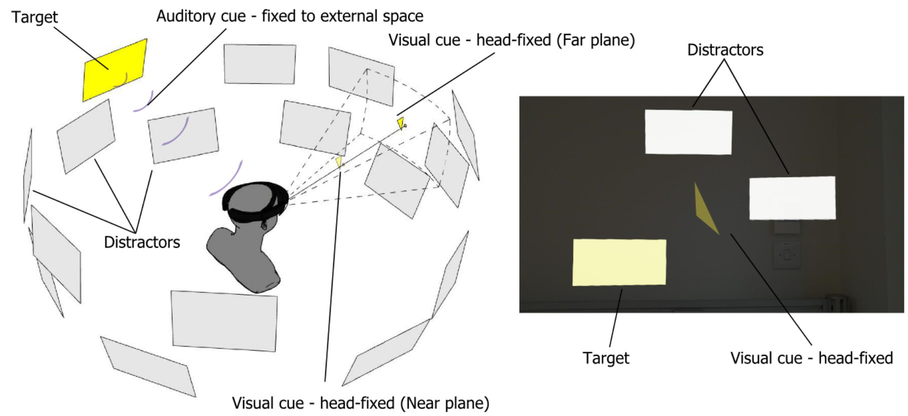

Figure 2: Experimental setup. Participants had to identify and interact via an air-tap gesture with an out of view virtual target (yellow box) amongst distractor stimuli (white boxes). Target and distractors were radially arranged around the participant, positioned at a constant distance of 200 cm. Participants performed the task with visual cueing, with visual and auditory cueing, or with no cueing. We also manipulated the plane the visual cue was displayed on relative to the target and distractor stimuli: on the same plane (200 cm - Far; opaque arrow), or closer plane relative to the target (40 cm - Near; semi-transparent arrow). The right pane shows cue, target and distractors from the user view perspective.

3. Effectiveness of Spatialized Auditory cues in Visual Search with Cue-Target depth discrepancy

AR cueing interfaces typically require integrating virtual (cue direction) and real-world target information and, as stated above, can be negatively impacted by CT depth discrepancy. We currently lack a quantification of the impact of CT depth discrepancy on AR cueing effectiveness, and a description of measures that can be put in place to counteract or minimize this disruption.

To address these points, in the present study we quantified the impact of CT depth discrepancy on search performance. We also measured the effectiveness of simultaneous spatialized auditory cueing in aiding search performance under conditions of visual depth discrepancy. Specifically, we investigated whether auditory directional cues can offset the performance costs that are incurred when head-fixed directional cues (i.e. augmented arrows) and target stimuli positioned at external fixed locations are displayed at different depths. As detailed in the previous section, discrepancies in cue and target disparity planes can require alternating vergence behaviours (eye rotation): i.e. users may need to continuously switch their vergence (and shift visual focus of attention) between disparity planes to properly integrate information across stimuli [27]. The fact, however, that CT information cannot be clearly visualized simultaneously is likely to negatively impact visual search performance. If auditory positional information centred on the target stimulus were also available (which is not bound to these constraints), then this should override these limitations. We chose spatialized auditory cueing, opposed to other forms of sensory cueing (e.g. vibrotactile), as it can be easily implement on current AR-HMDs such as the HoloLens and the Magic Leap, requiring no additional hardware and minimum setup.

Our study was guided by the following two hypotheses:

- H1: Visual search performance is negatively impacted when a head-fixed visual directional cue (an arrow at the centre of the display) and a virtual target stimulus (a box displayed at an external fixed location) occupy different disparity planes.
- H2: Costs in visual search performance caused by discrepant depth information between cue and target will be mitigated by additional auditory positional cueing (auditory signal originating from the target stimulus).

3.1. Participants

We recruited 32 participants (20 female) with mean age 26.8 ± 7.4 years (range: $20 - 51$). All participants had normal or corrected to normal vision and hearing. Informed consent was obtained from all participants before starting the experiment. Only three participants had prior experience using a Microsoft HoloLens. Participants were compensated £10 (GBP) for their participation.

This study was approved by the institutional Research Ethics Committee and was in agreement with the local research guidelines and regulations.

3.2. Design

Throughout five experimental blocks, participants had to find and select a virtual visual target. The target was cued by a head-fixed visual arrow lying amongst distractor stimuli. We adopted a within-subject design to evaluate the effects of visual depth discrepancy between cue and target stimuli, and the effects of auditory cueing on visual search performance.

3.2.1. Independent Variables

Across separate blocks we manipulated the plane on which a head-fixed arrow cue was displayed, half way between the participant and the target or on the same plane occupied by the target (Cue Plane: Near vs. Far), and the presence of an additional auditory cue centred on the target (Auditory Cueing: On vs. Off).

3.2.2. Dependent Variables

We measured participants' response time (RT - amount of time required to air-tap on the cued target) and response accuracy (ACC - whether the participant interacted with the target or a distractor stimulus) at the visual search task. We additionally collected subjective measures of task load (physical demand, mental demand, temporal demand, effort and frustration) using a modified NASA-TLX questionnaire [57]. These task load subjective measures complemented objective performance measures (RT/ACC). We focused on objective performance measures, opposed to subjective performance measures, as the latter often tend to be biased [58].

3.3. Materials

The Experiment was conducted in a controlled testing environment, with artificial overhead lighting. Participants wore a Microsoft HoloLens AR-HMD, with a horizontal field of view of about 30°, while sitting on a swivel chair in the centre of the testing area. Participants viewed a three dimensional array of 16 white virtual box stimuli (encompassing 11.4 x 5.7 degrees of visual angle, each), radially arranged around the participant at a constant distance of 200cm (Figure 2), arranged at different heights relative to the participant's head. This stimulus arrangement required participants to rotate their head/body both leftwards/rightwards as well as upwards / downwards in order to fully scan the virtual test area. The right pane of Figure 2 depicts a user view perspective of the stimuli contained in the augmented visual field of view. At most three distractors could be simultaneously present in the FOV. All boxes were non-overlapping. While the study was aimed at behavioural cuieng in AR, we decided to exclusively adopt virtual stimuli in order to precisely control target / distractor positioning, that varied randomly across trials.

A small hollow circular cursor was presented at the centre of the display as a proxy of user gaze. Participants had to align the gaze cursor with target stimuli while performing an 'air-tap gesture' (pinching gesture with the index and thumb in front of the device) in order to interact with any given target. A yellow central visual arrow cue was also presented, offset by 1cm relative to the gaze cursor in the direction of the target it pointed to (the arrow rotated about the gaze axis in the target's direction). If the target was behind the subject, the arrow cued participants to rotate leftwards / rightwards, based on the shortest cue/target trajectory. The cue disappeared once it overlapped with the target. The shape and positioning of the arrow cue was not based on ideal cueing interface design principles, but simply provided a test bed for measuring effects of CT depth discrepancy and cross-modal integration on search performance, all other conditions being equal.

Across trials, the arrow pointed constantly at one of the 16 boxes (the current target box, randomly selected) by rotating on a single axis parallel to the participant's coronal plane. Across different blocks, the cue was either displayed 40cm in front of the participant (near), or on the same disparity plane at 200cm occupied by the box stimuli (far). Given that HoloLens displays are generally fixed at an optical distance of 200cm, design guidelines suggest displaying stimuli at this distance to ensure optimal visibility. However, AR devices, including the HoloLens, are also aimed at assisting tasks within typing/reaching distance ($30 \sim 50$cm), such as assembly work on a workbench. We therefore displayed near stimuli at 40cm in order to deliberately introduce differences in vergence angles across disparity planes containing task relevant stimuli. The arrow size was scaled across disparity planes to encompass an equivalent area on the retina (2.8 degrees of visual angle). In a subset of blocks, a sound cue was presented simultaneously to the visual cue. The sound cue consisted of a sound clip of a ringing alarm clock (1 second period), centred on the target box. The sound cue was spatially rendered with the Unity head-related transfer function (HRTF). We opted for spatialized rendered sounds (opposed to external physical sounds) for a matter of convenience in the experimental setup. This allowed for virtual stimulus positioning to be independent from the layout of the physical environment, which would be otherwise challenging considering that target positioning was randomized across trials. Stimulus presentation, data logging and experiment logic were implemented in Unity 2018.1.

Throughout the experiment participants also performed an online modified NASA-TLX questionnaire [59] by inputting responses in a laptop online form, rating: (i) physical demand,(ii) mental demand, (iii) temporal demand, (iv) effort invested and, (v) level of frustration associated with the task. Every question was scored on a rating scale ranging from 0 (Very Low) to 100 (Very High). Participants performed the modified NASA-TLX questionnaire between blocks.

3.4. Procedure

Throughout trials participants were required to interact (gaze and 'air-tap') with a target box stimulus. At the beginning of each trial one of the 16 boxes was randomly selected as a target stimulus, changing colour from white (RGB: 255, 255, 255) to light yellow (RGB: 255, 230, 160). The remaining 15 boxes acted as distractor stimuli. This transition was relatively subtle, in order for the target to not immediately stand out at a glance amongst distractors. As soon as the target changed colour, the head-fixed arrow cue was presented on the side of the gaze cursor, persistently pointing in the direction of the target. In trials with additional auditory cueing, a ringing sound played back simultaneously with the presentation of the arrow cue: the spatially localized ringing sound persisted until the participant selected the target stimulus.

Participants had to search for and interact with the target as quickly as possible. As soon as an air-tap was successfully detected on a box stimulus (target or distractor), its colour changed to a darker shade of yellow (RGB: 255, 230, 100) for 1 second, providing user feedback of a successfully registered air-tap. If the stimulus was incorrect (not the current target), the response was flagged as 'incorrect' and the participant was required to continue the search for the target (incorrect trials were later removed from the analysis of RT data). Trials were labelled as failed when exceeding a duration of 15 seconds. If the box stimulus was the current target, the RT was recorded, the stimulus colour was reset and the visual cue stimulus disappeared (and auditory signal, when present, was interrupted). This was immediately followed by another trial, where a new target box was selected and cued by the arrow. A constraint was applied to the selection of next target box: it could not be currently present within the field of view, thus requiring the participant to explore the surrounding environment through head and body rotations.

Each participant performed the visual search task across five blocks with 10 trials each. Four blocks were based on combinations of visual cue disparity plane (Plane Near / Far) and the presence of the Sound cue (Sound On / Off). We collected an additional block with no visual or auditory cue, in theory providing a baseline measure of visual search performance. Block order was counterbalanced across participants. The entire testing session for each participant lasted approximately 45 minutes. Prior to conducting the experiment, we had participants perform a practice run to familiarize themselves with the search task and pinch gesture. This involved correctly identifying targets in five practice trials. Participants on average required 10 attempts to achieve the desired performance. This ensured that all participants were capable of adequately performing the visual search and target identification task via an air tap gesture.

4. Results

We report findings related to visual search response times (RTs) and response accuracy (ACC), and mental / physical fatigue scores based on the NASA-TLX questionnaire inventory.

4.1. Visual Search Task Performance

Participants ideally contributed 10 RT and ACC data points per condition. Outlier data points, lying 1.5 standard deviation from the group mean, were excluded from analysis. The majority of participants failed at performing the task when lacking both visual and auditory cueing, requiring the experimenter's intervention. Thus, RT and ACC scores were evaluated in terms of relative performance differences as a function of visual cue disparity plane (near or far plane), and auditory cueing (presence or absence of sound).

Figure 3a shows participants' mean response time in seconds across each condition for trials in which the target was correctly selected on the first attempt. Data were analysed with a 2x2 Repeated Measures ANOVA, with factors Visual Cue Plane (Near vs. Far), and Auditory Cueing (On vs. Off). The analysis revealed a significant main effect of Cue Plane, $F(1,31) = 25.38$, $p < .001$, $\eta^2 = .45$, where RTs were significantly longer when the visual cue was displayed on the near plane opposed to the far plane that was also occupied by the target and distractor stimuli. Consistent with the first hypothesis, search times were slower when the directional cue was presented on a closure disparity plane than the target.

A significant main effect of Auditory Cueing, $F(1,31) = 11.61$, $p = .002$, $\eta^2 = .27$, and a significant Cue Plane x Auditory Cueing interaction, $F(1,31) = 5.44$, $p = .026$, $\eta^2 = .149$, evidenced improved RTs when visual cueing was accompanied by additional auditory information. t-test post-hoc comparisons were run to explore the significant interaction, revealing that RTs were significantly impaired in trials with near visual cues when there was no auditory cueing (Cue Plane-near / Auditory Cueing-off), relative to all other conditions: Cue Plane-far / Auditory Cueing-on, $t(62) = 3.72$, $p < .001$, $d = .92$, Cue Plane-near / Auditory Cueing-on, $t(62) = 1.96$, $p = .027$, $d = .56$, Cue Plane-far / Auditory Cueing-off, $t(62) = 3.19$, $p = .001$, $d = .79$. Crucially, the above comparisons show a significant reduction in RTs when near visual cues were assisted by auditory cueing, with respect to instances in which the near visual cue was not accompanied by an auditory cue, approaching performance observed with far visual cues. Therefore, in keeping with the second hypothesis, the presence of additional auditory cueing was effective at offsetting costs in RT caused by visual depth mismatch.

Figure 3b shows participants' target selection accuracy (proportion of correct trials in a block: N correct / total) across each condition in the experiment. These ACC values were submitted to a 2x2 Repeated Measures ANOVA. This analysis only revealed a significant main effect of Cue Plane, $F(1, 31) = 17.2$, $p < .001$, $\eta^2 = .36$, where ACC was significantly impaired in trials with a near visual cue with respect to a far visual cue, irrespective of the presence or absence of auditory information. There was neither a significant main effect of Auditory Cueing, $F(1, 31) = 2.81$, $p = .1$, $\eta^2 = .08$, nor a significant Cue Plane x Auditory Cueing interaction, $F(1, 31) = 1.27$, $p = .3$, $\eta^2 = .04$. Sound was therefore ineffective at reducing costs in accuracy associated with near visual cueing.

4.2. Questionnaire Scores

Figure 3c shows modified NASA-TLX responses given by participants for each experimental condition. Participants provided a 0-100 score for each of the five dimensions of the NASA-TLX inventory at the end of each testing block. These data were submitted to a non-parametric Friedman test. This analysis revealed that mental load varied across conditions, $\chi^2(3) = 27.35, p < 0.001$. Bonferroni-corrected Wilcoxon signed-rank tests revealed that participants re-

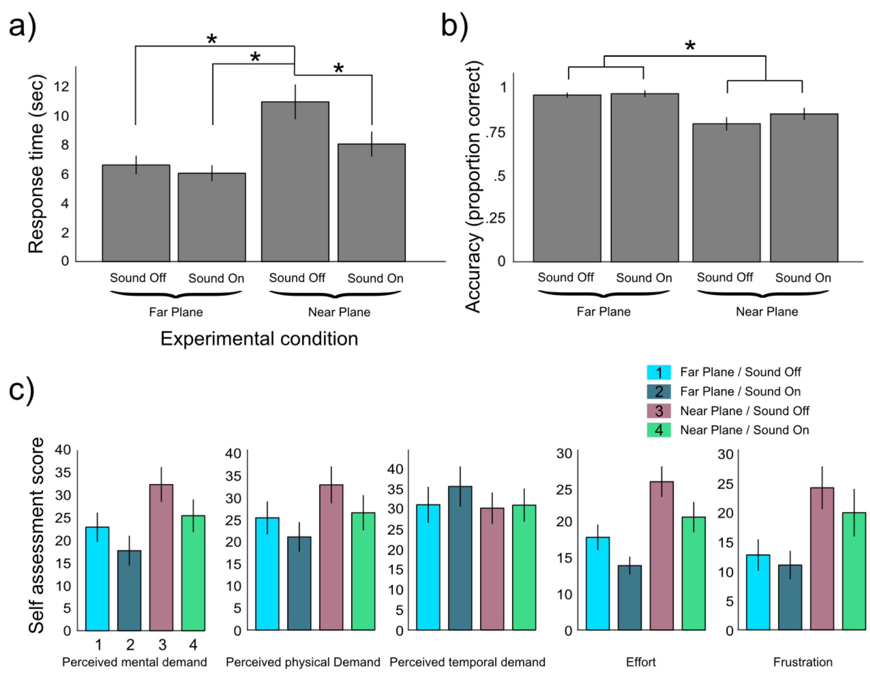

Figure 3: Visual search task performance measures for response time (a), response accuracy (b), and NASA-TLX self-assessment scores (c), across experimental conditions. Error bars represent standard error of the mean.

ported a significantly higher mental demand when auditory information was absent, opposed to instances in which it was available: far visual cues ($Z = -3.019, p = .003$), near visual cues ($Z = -2.8, p = .02$). Similarly, physical demand was modulated by task conditions, $\chi^2(3) = 12.74, p < 0.005$, revealing that trials with far cues and without sound were as physically more demanding: far visual cues ($Z = -2.69, p = .04$), for near visual cues ($Z = -2.8, p = .12$). No significant differences in temporal demand were observed across experimental conditions. Effort varied across conditions, $\chi^2(3) = 26.37, p < 0.001$, revealing that sound reduced effort for far cues: far cues ($Z = -3.07, p = .008$), near cues ($Z = -1.93, p = .2$). Finally, frustration levels varied across conditions, $\chi^2(3) = 20.93, p < 0.001$, revealing that participants perceived trials with near visual cueing as more frustrating: with sound ($Z = -3.36, p = .004$), without sound ($Z = -2.89, p = .016$).

5. Discussion

We predicted that visual search and target identification amongst distractors would be negatively impacted when the visual cue and target occupy different disparity planes. It would force participants to continuously switch their vergence, and shift visual focus of attention, between disparity planes to compare positional information across stimuli. We also predicted that additional auditory cueing, which is not bound to these constraints on visual attention, should mitigate costs in visual search caused by discrepant depth information.

5.1. H1: CT Depth Discrepancy Impacts Visual Search

We demonstrate that depth discrepancy indeed hinders the ease of use of AR visual cueing interfaces, introducing substantial costs on visual search time and accuracy. RT and ACC results showed that response times and accuracy were significantly impaired when visual cues were displayed on the near plane, likely forcing participants to alternate fixations to integrate CT information provided at different depths, with a measurable impact on visual search performance.

This was consistent with the idea of greater information access costs incurred by eye movement and/or shifts of visual attention [23, 60], in response to disparity plane discrepancy. Stimuli situated at different depths, requiring different amounts of eye vergence, cannot be clearly visualized simultaneously. Fixating on one stimulus would cause the other stimulus appearing diplopic ('double'). In order to clearly visualize and integrate information carried by both stimuli (directional and positional information for cue and target, respectively), participants were likely forced to repeatedly alternate fixations between cue and target planes, yielding costs in visual search performance.

Depth discrepancy can easily occur in AR systems, given that real-world and virtual stimuli are embedded in 3D space and frequently located at different visual depths. By extension, depth discrepancy can also arise in the context of AR cueing interfaces, when visual cues, targets, and other potentially relevant stimuli are located at different depths. While this can be considered a design flaw, it can nonetheless emerge in AR cueing interface design, due to lack of *a priori* knowledge of the depth at which targets might be located, or to uncertainty associated to measurements of real-world depths.

Several factors can potentially contribute to CT depth discrepancy. for example, CT could be introduced by measurement errors, especially in the case of rapidly moving targets, where precise real-time depth determination of target depth can prove unreliable in devices with a big eye box such as the Hololens 1 and 2 and the Magic Leap One. CT depth discrepancy can also be engendered by calibration errors, where optics are not properly aligned to ensure consistency in the perceived visual depth of virtual cues and real-world target stimuli. This can be further compounded by vergence / accommodation conflicts in devices with fixed focal length [61]. While closer / further cues trigger different degrees of eye rotation (vergence), these are not accompanied by changes in accommodation given that in devices such as the HoloLens, displays are fixed at an optical distance of 2m. This mismatch in vergence-accomodation cues can in turn introduce biases in the perceived depth of virtual stimuli [62]. All or subsets of these factors might engender depth discrepancies between virtual and world stimuli, and in turn impair the effectiveness of visual cueing interfaces relaying on virtual cues pointing towards real-world targets. Furthermore, there can be instances in which text based labels associated with real-world targets, are intentionally placed at fixed distances (or not co-located with objects) to minimize occlusion and improve readability [11]. CT discrepancy might affect performance if these labels also serve as guidance cues for localizing targets.

5.2. H2: Sound Cueing Mitigates Depth Discrepancy Costs

When directional auditory information was presented alongside a near visual cue, response times neared those observed in trials with far visual cues. Consistent with the second hypothesis, auditory cueing was successful at reducing costs associated with arrow / target disparity plane discrepancy. This finding is consistent with prior research showing that 3D audio can effectively bypass drawbacks associated with visual search (i.e. limited FOV and visual clutter) [16, 17, 18]. In keeping with performance data, participants reported that trials with auditory cues were perceived as significantly less demanding and less effortful, suggesting an abatement of attentional and/or cognitive demands in the presence of additional auditory information. Response accuracy results painted a slightly different picture with respect to what was observed with response times. As was the case for response times, response accuracy was significantly lower in trials with a near visual cue. However, the presence or absence of auditory information had no modulatory effect on response accuracy. This would suggest that sound was ineffective at reducing the costs in accuracy associated with CT depth discrepancy.

The findings from our study suggest that CT depth mismatches should be explicitly accounted for in AR cueing interface design. Best performance is achieved when cue and target are at the same depth. However, this cannot always be feasible or convenient, due to lack of prior knowledge of target positioning, cue implementation (the distance at which the cue is rendered), and depth measurement / optical calibration errors. We demonstrate that mixed audio-visual AR cueing is highly effective at dealing with CT visual depth discrepancy (even substantial mismatches that were deliberately introduced in the study). Thus, unless depth is fully accounted for, AR cuing design greatly benefits from cross-modal approaches that provide effective redundancy against sources of CT depth mismatch.

5.3. Multimodal Cueing and AR-Based Visual Search Stages

Multimodal view management is still an open venue of research in AR interface design. Combinations of sensory cues can be aimed at reducing perceptual load within single sensory channels, by distributing workload across modalities [63]. Multimodal approaches have been specifically investigated in the context of cueing interfaces. Several studies have shown improved visual search performance when pooling directional information across different sensory channels [8, 51, 52]. However, cross-modal interference effects have also been reported [64], suggesting that effects of multimodal cueing on visual search performance are sensible to cue and task constraints. Our results suggest that users adopted different strategies, exploiting additional auditory information as a function of the reliability of visual information. When cue and targets were positioned on the same disparity plane, participants relied exclusively on visual information to guide behaviour, with no aggregate benefit of additional auditory cues. Whereas, when visual / attentional access costs were introduced by discrepant disparity plane information between the visual cue and targets / distractors, participants shifted strategy by relying on sound. Furthermore, the dissociation observed between response time and accuracy suggests a two-stage AR visual search process: 1) an initial orienting of attention in the general direction of the target, migrating task relevant information into the FOV, 2) followed by the discrimination of the target amongst distractors. Sound benefits the first stage, whereas it has no beneficial impact on target identification accuracy, given reduced positional determination with sound, or, the fact that participants preferentially rely on visually degraded information.

5.4. Limitations and Open Questions

A limitation in the design of the study was that we had no 'sound only' condition. This implies that we could not compare the auditory cue performance against that with the visual cue. Consequently, we could not determine, when both were available, whether participants were integrating visual and auditory information or if they were elusively relying on the auditory signal under conditions of visual uncertainty. Another limitation is that, based on the current dataset, we cannot determine whether performance degradation in the near condition was due to sequentially alternating fixations between cue and target planes or due to participants exclusively fixating on one plane while simultaneously utilizing visually degraded information on the other plane. While the increases in RTs suggest the first possibility, and is consistent with prior literature [27], we cannot exclude the adoption of other strategies. Insights on task strategies, and how these relate to performance, could be obtained by running an equivalent study with eye and head tracking. These measures could clarify under conditions of CT depth discrepancy whether participants sequentially fixate between depth planes or default preferentially to one plane, whether they integrate visual and auditory cues or disregard the visually degraded cue in favour of the auditory information, and whether the effectiveness of the auditory cue is modulated by target eccentricity.

6. Conclusion

This paper gives insights into the effectiveness of visual and auditory cues for helping users locate out-of-view objects when using a head-mounted AR system. Results show that visual search performance is negatively impacted (i.e. response times increase and accuracy decreases), when a visual cue is not located on the same depth plane as the target it refers to. Reduced search performance is likely caused by participants needing to switch continuously between different disparity planes to identify the target. Critically, we found that auditory cueing, which was not bound to these visual constraints, can be used to effectively overcome these performance costs in response time, but not those in accuracy. These results demonstrate the importance of AR cross-modal cueing under conditions of mismatching depth cues and show that designers should consider augmenting visual cues with auditory ones.

These findings are of relevance to visual cueing interface design in binocular HMDs, where due to potential combinations of factors, including modelling/measurement/calibration/perceptual errors and/or view management constraints, the co-location of virtual cues / real-world targets at a given visual depth can not always be ensured. Future studies could investigate visual search as a function of different distances between cue and target on the depth axis, to characterize the relationship (linear/non-linear) between the magnitude of depth discrepancy and the corresponding cost in search performance. Furthermore, cue / target depth discrepancy could also be explored when the cue lies at a further distance relative to the target depth plane.

Acknowledgments

This study was supported by the HuMan project under the European Union's Horizon 2020 research and innovation programme, grant agreement no. 723737. http://humanmanufacturing.eu/. This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) under grant KR 4521/2-1.

References

[1] W. Lu, B.-L. H. Duh, S. Feiner, Subtle cueing for visual search in augmented reality, in: 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2012, pp. 161–166. doi:10.1109/ISMAR.2012.6402553.

[2] C. Trepkowski, J. Maiero, E. Kruijff, D. Eibich, A. Marquardt, S. Feiner, The effect of narrow field of view and information density on visual search performance in augmented reality, in: 2019 IEEE Conference on Virtual

Reality and 3D User Interfaces (VR), 2019, pp. 575–584. doi:10.1109/VR.2019.8798312.

[3] J. Duncan, G. W. Humphreys, Visual search and stimulus similarity, Psychological Review 96 (3) (1989) 433–458. doi:10.1037/0033-295X.96.3.433.

[4] J. M. Wolfe, Guided search 2.0 a revised model of visual search, Psychonomic Bulletin Review 1 (2) (1994) 202–238. doi:10.3758/BF03200774.

[5] P. Baudisch, R. Rosenholtz, Halo: A technique for visualizing off-screen objects, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03, Association for Computing Machinery, New York, NY, USA, 2003, p. 481–488. doi:10.1145/642611.642695.

[6] S. Burigat, L. Chittaro, S. Gabrielli, Visualizing locations of off-screen objects on mobile devices: A comparative evaluation of three approaches, in: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '06, Association for Computing Machinery, New York, NY, USA, 2006, p. 239–246. doi:10.1145/1152215.1152266.

[7] U. Gruenefeld, A. E. Ali, W. Heuten, S. Boll, Visualizing out-of-view objects in head-mounted augmented reality, in: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17, Association for Computing Machinery, New York, NY, USA, 2017. doi:10.1145/3098279.3122124.

[8] U. Gruenefeld, D. Lange, L. Hammer, S. Boll, W. Heuten, Flyingarrow: Pointing towards out-of-view objects on augmented reality devices, in: Proceedings of the 7th ACM International Symposium on Pervasive Displays, PerDis '18, Association for Computing Machinery, New York, NY, USA, 2018. doi:10.1145/3205873.3205881.

[9] S. Gustafson, P. Baudisch, C. Gutwin, P. Irani, Wedge: Clutter-free visualization of off-screen locations, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, Association for Computing Machinery, New York, NY, USA, 2008, p. 787–796. doi:10.1145/1357054.1357179.

[10] N. Henze, B. Poppinga, S. Boll, Experiments in the wild: Public evaluation of off-screen visualizations in the android market, in: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, NordiCHI '10, Association for Computing Machinery, New York, NY, USA, 2010, p. 675–678. doi:10.1145/1868914.1869002.

[11] M. Tatzgern, D. Kalkofen, R. Grasset, D. Schmalstieg, Hedgehog labeling: View management techniques for external labels in 3d space, in: 2014 IEEE Virtual Reality (VR), 2014, pp. 27–32. doi:10.1109/VR.2014.6802046.

[12] K. Ichihashi, K. Fujinami, Estimating visibility of annotations for view management in spatial augmented reality based on machine-learning techniques, Sensors 19 (4) (2019). doi:10.3390/s19040939.

[13] P. A. Hancock, J. E. Mercado, J. Merlo, J. B. F. V. Erp, Improving target detection in visual search through the augmenting multi-sensory cues, Ergonomics 56 (5) (2013) 729–738. doi:10.1080/00140139.2013.771219.

[14] D. Gupta, An empirical study of the effects of context-switch, object distance, and focus depth on human performance in augmented reality, Ph.D. thesis, Virginia Tech, Blacksburg, VA, USA (2004).
URL https://vtechworks.lib.vt.edu/handle/10919/33507

[15] G. K. Edgar, J. C. Pope, I. R. Craig, Visual accomodation problems with head-up and helmet-mounted displays?, Displays 15 (2) (1994) 68–75. doi:https://doi.org/10.1016/0141-9382(94)90059-0.

[16] T. Arce, H. Fuchs, K. McMullen, The effects of 3d audio on hologram localization in augmented reality environments, in: Proceedings of the Hu-

man Factors and Ergonomics Society Annual Meeting, Vol. 61, 2017, pp. 2115–2119. doi:10.1177/1541931213602010.

[17] G. Luo, E. Peli, Use of an augmented-vision device for visual search by patients with tunnel vision, Investigative Ophthalmology Visual Science 47 (9) (2006) 4152–4159. doi:10.1167/iovs.05-1672.

[18] B. D. Simpson, D. S. Brungart, R. H. Gilkey, J. L. Cowgill, R. C. Dallman, R. F. Green, K. L. Youngblood, T. J. Moore, 3d audio cueing for target identification in a simulated flight task, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 48 (16) (2004) 1836–1840. doi:10.1177/154193120404801610.

[19] P. Flanagan, K. I. McAnally, R. L. Martin, J. W. Meehan, S. R. Oldfield, Aurally and visually guided visual search in a virtual environment, Human Factors 40 (3) (1998) 461–468. doi:10.1518/001872098779591331.

[20] J. P. McIntire, P. R. Havig, S. N. J. Watamaniuk, R. H. Gilkey, Visual search performance with 3-d auditory cues: Effects of motion, target location, and practice, Human Factors 52 (1) (2010) 41–53. doi:10.1177/0018720810368806.

[21] Z. Zhou, A. D. Cheok, Y. Qiu, X. Yang, The role of 3-d sound in human reaction and performance in augmented reality environments, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 37 (2) (2007) 262–272. doi:10.1109/TSMCA.2006.886376.

[22] I. SanMiguel, D. Linden, C. Escera, Attention capture by novel sounds: Distraction versus facilitation, European Journal of Cognitive Psychology 22 (4) (2010) 481–515. doi:10.1080/09541440902930994.

[23] C. D. Wickens, Multiple resources and performance prediction, Theoretical Issues in Ergonomics Science 3 (2) (2002) 159–177. doi:10.1080/14639220210123806.

[24] V. V. Krishnan, L. Stark, A heuristic model for the human vergence eye movement system, IEEE Transactions on Biomedical Engineering BME-24 (1) (1977) 44–49. doi:10.1109/TBME.1977.326207.

[25] M. Lambooij, M. Fortuin, I. Heynderickx, W. IJsselsteijn, Visual discomfort and visual fatigue of stereoscopic displays: A review, Journal of Imaging Science and Technology 53 (3) 30201–1–30201–14. doi:10.2352/J.ImagingSci.Technol.2009.53.3.030201.

[26] J. L. Gabbard, D. G. Mehra, J. E. Swan, Effects of ar display context switching and focal distance switching on human performance, IEEE Transactions on Visualization and Computer Graphics 25 (6) (2019) 2228–2241. doi:10.1109/TVCG.2018.2832633.

[27] E. Kruijff, J. E. Swan, S. Feiner, Perceptual issues in augmented reality revisited, in: 2010 IEEE International Symposium on Mixed and Augmented Reality, 2010, pp. 3–12. doi:10.1109/ISMAR.2010.5643530.

[28] J. M. Wolfe, What can 1 million trials tell us about visual search?, Psychological Science 9 (1) (1998) 33–39. doi:10.1111/1467-9280.00006.

[29] D. Scott, Visual search in modern human-computer interfaces, Behaviour & Information Technology 12 (3) (1993) 174–189. doi:10.1080/01449299308924378.

[30] J. W. Kalat, Biological psychology, 7th ed, Biological Psychology, Wadsworth/Thomson Learning, Belmont, CA, USA, 2001.

[31] M. Jacob, S. Hochstein, Graded recognition as a function of the number of target fixations, Vision Research 50 (1) (2010) 107–117. doi:10.1016/j.visres.2009.10.019.

[32] J. M. Wolfe, Guidance of visual search by preattentive information, in: L. Itti, G. Rees, J. K. Tsotsos (Eds.), Neurobiology of Attention, Academic Press, Burlington, 2005, pp. 101–104. doi:10.1016/B978-012375731-9/50021-5.

[33] P. T. Quinlan, Visual feature integration theory: Past, present, and future, Psychological Bulletin 129 (5) (2003) 643–673. doi:10.1037/0033-2909.129.5.643.

[34] M. Arguin, P. Cavanagh, Parallel processing of two disjunctive targets, Perception & Psychophysics 44 (1) (1988) 22–30. doi:10.3758/BF03207471.

[35] S. Forster, N. Lavie, Failures to ignore entirely irrelevant distractors: The role of load, Journal of Experimental Psychology: Applied 14 (1) (2008) 73–83. doi:10.1037/1076-898X.14.1.73.

[36] O. Cakmakci, J. Rolland, Head-worn displays: a review, Journal of Display Technology 2 (3) (2006) 199–216. doi:10.1109/JDT.2006.879846.

[37] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, Non-visual cues for view management in narrow field of view augmented reality displays, in: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2019, pp. 190–201. doi:10.1109/ISMAR.2019.000-3.

[38] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, E. Kruijff, Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks, in: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2014, pp. 177–186. doi:10.1109/ISMAR.2014.6948425.

[39] A. Sigitov, E. Kruijff, C. Trepkowski, O. Staadt, A. Hinkenjann, The effect of visual distractors in peripheral vision on user performance in large display wall systems, in: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, ISS '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 241–249. doi:10.1145/2992154.2992165.

[40] E. Kruijff, J. Orlosky, N. Kishishita, C. Trepkowski, K. Kiyokawa, The influence of label design on search performance and noticeability in wide

field of view augmented reality displays, IEEE Transactions on Visualization and Computer Graphics 25 (9) (2019) 2821–2837. doi:10.1109/TVCG.2018.2854737.

[41] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, J. Schöning, Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays, IEEE Transactions on Visualization and Computer Graphics 26 (12) (2020) 3389–3401. doi:10.1109/TVCG.2020.3023605.

[42] L. Bonanni, C.-H. Lee, T. Selker, Attention-based design of augmented reality interfaces, in: CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05, Association for Computing Machinery, New York, NY, USA, 2005, p. 1228–1231. doi:10.1145/1056808.1056883.

[43] J. Wither, S. DiVerdi, T. Höllerer, Annotation in outdoor augmented reality, Computers Graphics 33 (6) (2009) 679–689. doi:10.1016/j.cag.2009.06.001.

[44] E. E. Veas, E. Mendez, S. K. Feiner, D. Schmalstieg, Directing attention and influencing memory with visual saliency modulation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 1471–1480. doi:10.1145/1978942.1979158.

[45] F. Biocca, A. Tang, C. Owen, F. Xiao, Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06, Association for Computing Machinery, New York, NY, USA, 2006, p. 1115–1122. doi:10.1145/1124772.1124939.

[46] U. Gruenefeld, T. C. Stratmann, A. E. Ali, S. Boll, W. Heuten, Radiallight: Exploring radial peripheral leds for directional cues in head-mounted displays, in: Proceedings of the 20th International Conference

on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA, 2018. doi:10.1145/3229434.3229437.

[47] V. Lehtinen, A. Oulasvirta, A. Salovaara, P. Nurmi, Dynamic tactile guidance for visual search tasks, in: Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 445–452. doi:10.1145/2380116.2380173.

[48] R. W. Lindeman, R. Page, Y. Yanagida, J. L. Sibert, Towards full-body haptic feedback: The design and deployment of a spatialized vibrotactile feedback system, in: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST '04, Association for Computing Machinery, New York, NY, USA, 2004, p. 146–149. doi:10.1145/1077534.1077562.

[49] F. Biocca, C. Owen, A. Tang, C. Bohil, Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality, Journal of Management Information Systems 23 (4) (2007) 163–184. doi:10.2753/MIS0742-1222230408.

[50] J. C. Middlebrooks, D. M. Green, Sound localization by human listeners, Annual Review of Psychology 42 (1) (1991) 135–159. doi:10.1146/annurev.ps.42.020191.001031.

[51] K. Hopkins, S. J. Kass, L. D. Blalock, J. C. Brill, Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario, Ergonomics 60 (5) (2017) 692–700. doi:10.1080/00140139.2016.1198495.

[52] M. K. Ngo, C. Spence, Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli, Attention, Perception, Psychophysics 72 (6) (2010) 1654–1665. doi:10.3758/APP.72.6.1654.

[53] J. E. Cutting, P. M. Vishton, Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth, in: W. Epstein, S. Rogers (Eds.), Perception of Space and Motion, Handbook of Perception and Cognition, Academic Press, San Diego, 1995, pp. 69–117. doi:10.1016/B978-012240530-3/50005-5.

[54] G. Kramida, Resolving the vergence-accommodation conflict in head-mounted displays, IEEE transactions on visualization and computer graphics 22 (2015) 1912 – 1931. doi:10.1109/TVCG.2015.2473855.

[55] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, H. Fuchs, Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources, ACM Transactions on Graphics 33 (4) (Jul. 2014). doi:10.1145/2601097.2601141.

[56] A. T. Duchowski, D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, B. Bazyluk, Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field, SAP '14, Association for Computing Machinery, New York, NY, USA, 2014. doi:10.1145/2628257.2628259.

[57] S. G. Hart, L. E. Staveland, Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: P. A. Hancock, N. Meshkati (Eds.), Advances in psychology, 52 Human mental workload, North-Holland, Oxford, UK, 1988, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9.

[58] D. Tempelaar, B. Rienties, Q. Nguyen, Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application, PLOS ONE 15 (2020) 1–29. doi:10.1371/journal.pone.0233977.

[59] S. G. Hart, Nasa-task load index (nasa-tlx); 20 years later, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 50, 2006, pp. 904–908. doi:10.1177/154193120605000909.

[60] C. D. Wickens, C. M. Carswell, The proximity compatibility principle: Its psychological foundation and relevance to display design, Human Factors 37 (3) (1995) 473–494. doi:10.1518/001872095779049408.

[61] Y. Lu, B. Deng, Y. Yan, Z. Qie, J. Li, J. Gui, Vergence-accommodation conflict potential solutions in augmented reality head mounted displays, in: AIP Conference Proceedings, Vol. 2185, 2019, p. 020001. doi:10.1063/1.5137845.

[62] M. Mon-Williams, J. R. Tresilian, Ordinal depth information from accommodation?, Ergonomics 43 (3) (2000) 391–404. doi:10.1080/001401300184486.

[63] P.-A. Oskarsson, L. Eriksson, O. Carlander, Enhanced perception and performance by multimodal threat cueing in simulated combat vehicle, Human Factors 54 (1) (2012) 122–137. doi:10.1177/0018720811424895.

[64] T. Koelewijn, A. Bronkhorst, J. Theeuwes, Competition between auditory and visual spatial cues during visual task performance, Experimental Brain Research 195 (4) (2009) 593–602. doi:10.1007/s00221-009-1829-y.