**ABSTRACT**

Goal-oriented visual search in augmented reality can be facilitated by using visual cues to call attention to a target. However, traditional use of explicit cues can degrade visual search performance due to scene distortion, occlusion and addition of visual clutter. In contrast, Subtle Cueing has been previously proposed as an alternative to explicit cueing, but little is known about how well it works for head-tracked head worn displays (HWDs).

We investigated the effect of Subtle Cueing for head-tracked head worn displays, using visual search research methods in simulated augmented reality environments. Our user study found that Subtle Cueing improves visual search performance, and serves as a feasible cueing mechanism for AR environments using HWDs.

**Keywords:** Attention, Subtle visual cueing, Visual search.

**Index Terms:** H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.1.2 [Models and Principles]: User/Machine Systems—Human factors

1 INTRODUCTION AND RELATED WORK

Goal-oriented visual search is an action performed whenever a person searches for a known target in the visual environment [9]. In video–see-through Augmented Reality (AR), AR visual cues can be used to facilitate rapid visual search. Explicit AR cues have traditionally been used for this purpose [1,8], but such methods of cueing may introduce distortion, occlusion and clutter [2]. In turn, these problems may degrade visual search performance [6], and require additional steps to mitigate them [4].

To address these problems, Subtle Cueing [3] has been proposed as an alternative to explicit cueing. Figure 1 shows an illustration of a subtle cue as compared to an explicit cue. By using Subtle Cueing, it appears possible to redirect the user's attentional spotlight [7] without introducing the problems mentioned above. Previous research into Subtle Cueing has focused on desktop monitor-based, video–see-through, static camera setups [3]. Therefore, how Subtle Cueing performs in mobile AR, using head-tracked Head Worn Displays (HWDs), is still unknown.

To help answer this question, we investigate the effects of Subtle Cueing on visual search performance for head-tracked HWDs. We make the following contribution: Through simulated video–see-through AR user studies, we show that Subtle Cueing improves visual search response time and error rate in scenarios typical of mobile HWD AR.

2 RESEARCH QUESTION AND METHODOLOGY

We adapted the experimental method of Lu et al. [3] to address the research question:

"Can Subtle Cueing be used as a feasible cueing mechanism for video–see-through HWD-based AR in head-tracked scenarios?"

To achieve this, we first formulated an empirical user study with a within-subjects design. We developed this as a laboratory experiment to ensure repeatable and reliable test conditions. The apparatus consisted of the following components (Figure 2):

1. i-visor DH-4400VPD HWD
2. Emotiv Epoc wireless motion sensor headset
3. Fujitsu Lifebook Windows Laptop PC
4. Generic trackball mouse trigger

The Emotiv Epoc headset allowed us to track head orientation using its built-in dual-axis (yaw and pitch) gyroscope. The HWD and Emotiv Epoc were combined into a single unit using a head harness, weighing 450g in total, which allowed the standardization of component positions when worn by different subjects, thereby accommodating many head sizes and shapes. This allowed users to look around freely in a simulated video–see-through AR environment, with the view in the HWD being updated based on the orientation of the user's head.

The simulated AR environment was created using a 4096×2048 pixel panoramic image depicting an outdoor scene. At each point in time, a user could view through each eyepiece of the HWD a 640×480 pixel portion of the image (centered within the 800×600 panel), subtending a 25.2º diagonal viewing angle. The specific portion visible through the HWD was determined by head yaw and pitch, simulating an AR environment viewed from a fixed position, as illustrated in Figure 3.

Within this environment, a single target was embedded in the background: a black cross “+” (1.5º diagonal viewing angle), similar to that used in Lu et al. [3] and more than the minimum size recommendation for legibility [5]. The subject’s task was to find the target as quickly as possible and press the trackball mouse trigger the moment s/he saw the target. A subtle cue in the form of a white square was inserted in between the target and the scene (as shown in Figure 3), and the opacity of the subtle cue was varied as

![Figure 1: Cue (shaded square) and target (“+”) in an outdoor scene. Notice how less obvious (almost invisible) the subtle cue is compared to the explicit cue, even though the subtle cue still produces significant cueing effects [3].](image-url)
To further reduce order and learning effects, five different locations of the target was randomized to prevent learning effects. The starting view window location was always set to be at the center of the image (coinciding with the neutral head orientation that the subject would adopt). There was never a target in this initial starting region, thereby requiring the user to rotate their head (and therefore the camera view) and explore the rest of the image in order to find the target. When the user pressed the trigger, the system would determine whether or not the target was within the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.

To further reduce order and learning effects, five different scenes were used. The Feature Congestion (FC) value was used in the selection of the images, as FC has been used as an indicator of scenes were used. The Feature Congestion (FC) value was used in

The starting view window location was always set to be at the center of the image (coinciding with the neutral head orientation that the subject would adopt). There was never a target in this initial starting region, thereby requiring the user to rotate their head (and therefore the camera view) and explore the rest of the image in order to find the target. When the user pressed the trigger, the system would determine whether or not the target was within the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.

To further reduce order and learning effects, five different scenes were used. The Feature Congestion (FC) value was used in the selection of the images, as FC has been used as an indicator of the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.

To further reduce order and learning effects, five different scenes were used. The Feature Congestion (FC) value was used in the selection of the images, as FC has been used as an indicator of the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.

To further reduce order and learning effects, five different scenes were used. The Feature Congestion (FC) value was used in the selection of the images, as FC has been used as an indicator of the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.

To further reduce order and learning effects, five different scenes were used. The Feature Congestion (FC) value was used in the selection of the images, as FC has been used as an indicator of the current view window. If the target was within the view window when the user pressed the trigger, the user was deemed to have found the target successfully. Otherwise, the user was deemed to have failed that attempt. This was to ensure that equal attention was paid to all regions of the image within the intended search area, thereby reducing biases due to image location. The location of the target was randomized to prevent learning effects.