Head movements during visual exploration of natural images in virtual reality

Brian Hu, Ishmael Johnson-Bey, Mansi Sharma, Ernst Niebur
Zanvyl Krieger Mind/Brain Institute
Johns Hopkins University
Baltimore, Maryland 21218

Abstract—**During natural visual exploration, both head and eye movements can be used to redirect gaze to new points of interest. In order to better understand the role of head movements in this process, we recorded subjects' head orientations while they explored a set of natural images from five different categories using virtual reality head-mounted displays. While head movements are likely influenced by image content or image saliency, here we focus on their stereotyped patterns, which have a consistent relationship between the amplitude, duration, and peak velocity of movements. We find that most head movements occur along the cardinal directions, and furthermore, the head position and head velocity distributions are similar across image categories. Our results provide greater insight into the kinematics of head movements during visual exploration in virtual reality environments.**

I. Introduction

Humans operating in natural or virtual environments need to gather information about their surroundings in order to achieve optimal or, at least, acceptable performance in whatever task they are working on. Vision is arguably the most important sensing modality but high-resolution visual input is only available from a very small part of the environment, from the projection onto the fovea with an extent of a few square degrees. Therefore, during natural exploration, humans move their center of gaze constantly, making several saccades per second (in addition to other types of eye movements). Many studies have been devoted to characterizing and understanding eye movement behavior in a large number of tasks and environments, but humans also move their heads (and other body parts) to change their center of gaze. Much less is known about head movements compared to eye movements. This deficiency in our knowledge is perhaps acceptable when an observer looks at a stationary screen (like a computer monitor, television *etc.*) that fills a relatively small part of his or her visual environment. Understanding head movements becomes imperative, however, in virtual environments where the usable visual environment is much larger, going beyond the visual field of a human observer. In this study, we therefore study human head movements in a Virtual Reality (VR) environment.

Speed and direction of head movements are limited by the biomechanics of the neck. Humans are far more likely to move their heads horizontally than vertically, with oblique movements occuring even less frequently. Gaze shifts often include multiple eye movements, such that the eyes can fixate on multiple targets while the head is moving [1]. Non-human primates asked to redirect their gaze to a new location make head movements that are well correlated with their eye movements [2], [3]. These studies have also found that initial head and eye position can have substantive effects on gaze shifts.

Einhäuser et al [4] studied the coordination of head and eye movements of humans under natural viewing conditions. Subjects were fitted with a wearable eye tracker with integrated cameras ("EyeSeeCam") and allowed to explore different natural environments (forest, train station, apartment) while their eye and head movements were simultaneously recorded. These authors found that the majority of co-occurring head movements and eye movements point in opposite directions, consistent with a role of eye movements in stabilizing gaze while the head was moving. A smaller proportion of head and eye movements pointed in the same direction, allowing for synergistic interactions, possibly to avoid excessively large saccades. Visual exploration also occurs when searching for an object, *e.g.* while picking out a product on a supermarket shelf. In this case, selective attention can guide visual search by allowing organisms to direct their necessarily limited information processing capabilities to the most relevant sensory inputs gathered in a complex world. Nakashima and Shioiri [5], [6] studied the influence of head and eye position on the allocation of visual attention in a search task. When searching for a simple target among distractors, they found that a relative difference in the initial head and eye positions can interfere with visual search.

The studies discussed above are examples of two different experimental paradigms for studying the role of head and eye movements. While Einhäuser et al [4] allowed for unconstrained exploration in real-world environments, making their study potentially applicable to realistic scenarios, their results are difficult to reproduce and to interpret due to the large number of uncontrolled factors. The number of subjects used in the experiment was also low ($N = 4$), so it is unclear whether their results are representative of larger populations. On the other hand, stimuli used by Nakashima and Shioiri [5], [6] and other eye/head movement studies [2], [3] were very well controlled but also simplified and quite abstract. While their results provide insight into the influence of head and eye position on visual perception and attention, it is not always clear how they translate to natural scenes where the features and objects are more complex.

Here, we leverage recent advances in consumer-grade VR technology to create a novel experimental setup that allows us to record head movements while subjects view natural images in a VR setting. VR environments can render realistic natural images, and, at the same time, give the experimenter tight control over all details of the scene [7]. This is crucial for being able to reproduce the experimental setting between subjects and studies. Our approach thus complements and serves as a balance between the discussed experimental paradigms, allowing us to probe head movements in a well-controlled environment but using complex, naturalistic scenes.

II. Methods

All experimental procedures were approved by the Institutional Review Board of Johns Hopkins University. 27 subjects (15 male; mean age = 20.1 years, SD = 2.7) participated in the experiment. Before the experiment, all subjects gave written informed consent. Participants with neurological disorders were excluded from the study. Subjects received a pair of Google Cardboard VR glasses as compensation for their participation in the experiment. In the first part of the study, subjects filled out a participant data sheet that recorded demographic information as well as previous experience with video games and/or VR systems, see Table I. Each subject was then fitted with a pair of Google Cardboard VR glasses. Subjects wearing glasses had to remove them. A special compartment in the Google Cardboard housed a Samsung Galaxy S5 smartphone, which was used to display the VR environment and collect head movement data with a custom-designed script using the Google Cardboard SDK (available at https://developers.google.com/vr/unity/) and the Unity 5.0.2 game engine. While wearing the Google Cardboard, subjects first performed a nine-point calibration where they had to accurately redirect their head to targets within a fixed grid. Throughout the experiment, subjects could also "recenter" the VR environment's coordinate system to correct for drift.

Each subject was then asked to view in the VR environment a total of 70 different scenes (13 images from each of five image categories, with five repeat images, one from each category). All images had a resolution of 640 × 480 pixels. Four of the categories (buildings, fractals, "old" home interiors and landscapes) were introduced by Parkhurst et al [8] and we added an additional category ("new" home interiors) for reasons explained below. Image categories were chosen to provide subjects with a variety of scenes that are ecologically important. The scenes also differ in semantic content (*e.g.* fractals are devoid of meaning), which likely leads to differences in the allocation of top-down visual attention. We chose these images because previous studies have extensively characterized eye movements (fixations) that humans make in these scenes [8], [9]. Furthermore, a large study with hundreds of subjects determined which portions of these scenes were considered subjectively. The fifth, additional set of ("new") home interior images was collected from the internet and used in our experiment because the original home interior images were digitized from photographs and, as a consequence, were often perceived as blurry when rendered in the VR environment.

TABLE I

Results of pre- and post-experiment surveys, averaged over all participants (means and standard deviations). **Pre-experiment questions:** *"Game experience:"* Self-declared participant's experience with computer games (1=Least, 5=Most). *"VR experience:"* Percentage (fraction) of participants having had any previous experience in a VR environment. **Post-experiment questions:** *"Nausea/dizziness:"* Experience of any nausea or dizziness during the experiments (1=Least, 5=Most). *"Ease of use:"* Was the VR system easy to use (1=Very easy, 5=Very difficult). *"Small image used:"* Percentage (fraction) of participants who found that the small image was useful to guide image exploration.

Survey questions answered by participants	
Game experience (1-5)	M = 3.81, SD = 1.0
VR experience (yes)	25.9% (7/27)
Nausea/dizziness (1-5)	M = 1.48, SD = .94
Ease of use (1-5)	M = 2.37, SD = 1.6
Small image used (yes)	74.1% (20/27)

Subjects were seated on a stationary (non-swiveling) chair in a quiet room to minimize the influence of body movements and noise disturbances. Each image presentation began with a 1-second, small view of the image to give subjects an overview (gist) of the whole image. This then immediately zoomed into a large-scale, immersive image. The small image subtended approximately 30 degrees in the horizontal direction and 23 degrees in the vertical direction, while the full-size image subtended approximately 116 degrees in the horizontal direction and 100 degrees in the vertical direction. Subjects viewed the images through the field of view of the VR glasses, a square aperture with side length 74 degrees. As a result, the full size image was larger than the visible portion of the VR environment by more than a factor of two in surface area, and the head had to be moved to see parts of the image outside the center. Subjects were instructed to visually explore the images and were told that they would be asked about image contents. No explicit mention of head movements was made. After each image viewing, subjects were asked to describe the scene concisely in one sentence, and their verbal description was recorded using a voice recorder. Viewing time was set to 10 seconds for each image, with unlimited time for image description. No analysis of the recorded audio data is reported in this study. After completing the experiment, subjects were asked post-study survey questions about their experience, see Table I. Each experiment lasted approximately one hour.

We excluded all trials where the recorded head movements went outside the image area, *i.e.* when the vector indicating the head's forward direction intersected with the image plane outside the projected image. This was the case in $<2\%$ of the data. We resampled and interpolated the raw position data at 50 Hz and converted raw image pixel values to visual degrees. Head movement velocity was obtained by convolving each position trace with a derivative of Gaussian filter with a width $\sigma = 100ms$ [10]. We separated the head movements into "head fixations" (referred to below sometimes simply as

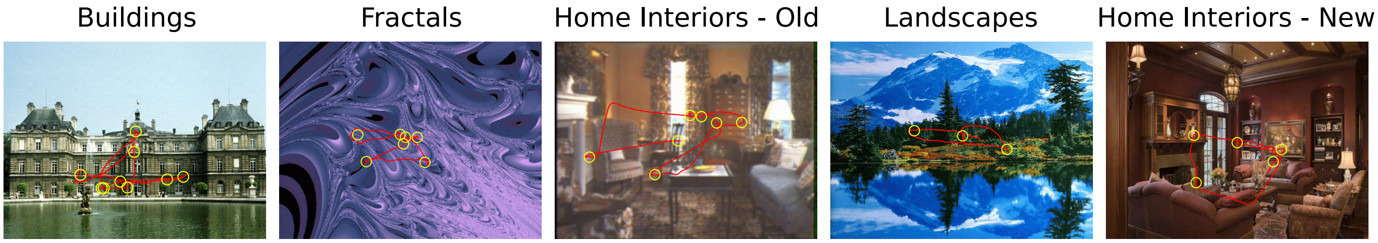
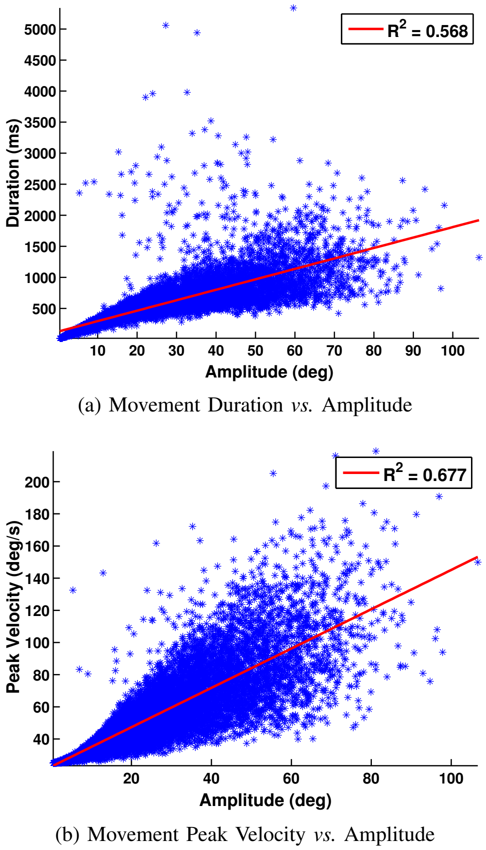


Fig. 1. Example head movement patterns for one subject on sample images from each category. Red lines show the movement trajectory and yellow circles indicate the locations of head "fixations," illustrating the diversity in head movements used for image exploration.



(a) Movement Duration *vs.* Amplitude

(b) Movement Peak Velocity *vs.* Amplitude

Fig. 2. Head movement main sequence. Head movements follow a stereotyped pattern, where both movement duration (a) and peak velocity (b) increase linearly with amplitude. The data is from all head movements over all subjects and images, and the red lines are least-squares fits to the data.

"fixations") and non-fixations by using a velocity threshold of 25 degrees/second. Although this threshold value is somewhat arbitrary, other studies have reported using similar thresholds between 15-25 degrees/second [1]–[3], [11]. Movements below the threshold were classified as fixations, and the centroid of recorded movements during fixation periods was used as the fixation center.

All data and code associated with this paper can be found online at: https://github.com/brianhhu/VR_HeadMovements.

III. Results

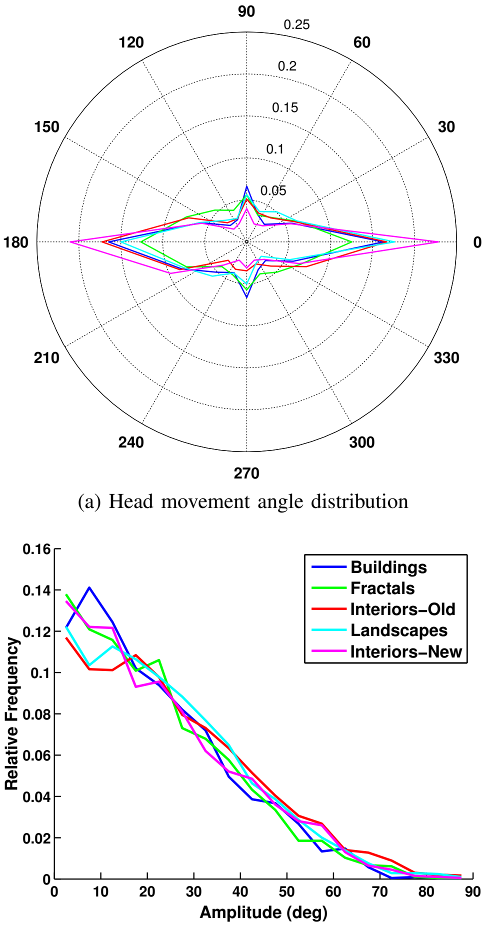
Subjects made a variety of different head movements in response to the natural scenes that they viewed in VR. Figure 1 shows example head movements for one subject viewing one image from each category. While subjects could have used the same approach to explore all images (*e.g.* a rectangular scanning pattern similar to that for the "Home Interiors – New" image in Figure 1), visual inspection of movement trajectories across subjects and images confirmed that there was a diversity of patterns in head movements. This suggests that head movements may be influenced by image content or image saliency, and that differences in head movements may exist between the different image categories. In the following analyses, we focus on understanding the kinematics of the underlying head movements that make up the diverse movement trajectories we observed. As a result, we ignore the temporal and location-specific aspects of head movements and leave this as a future avenue of research.

Eye movements are known to be relatively stereotyped, and the fixed relationship between their amplitude, duration, and peak velocity is known as the "main sequence" [12]. We found a similar rule governing head movements, with a predictable relationship between head movement amplitude and duration (Figure 2a), as well as between head movement amplitude and peak velocity (Figure 2b). Both movement duration and peak velocity increased as a noisy linear function of movement amplitude ($R^2 = 0.568$ and $R^2 = 0.677$, respectively). Our results are in agreement with, and extend the range of, previous experimental findings. Several studies [2], [3] reported the existence of a head movement main sequence. We confirm this result, demonstrating that the movement kinematics that govern head movements during visual exploration of natural images are similar. In addition, we show that this applies to a large class of head movements, while previous studies mainly tested lateral head movements [2], [3]. Furthermore, to the best of our knowledge, our results are the first demonstration of a main sequence for head movements recorded in a VR setting. Our findings can likely be explained, at least in part, by constraints given by the biomechanics of the neck musculature, which constrain the speed and magnitude of head movements. Because our setup did not allow for recording of eye movements, we cannot draw conclusions about the contribution of head movements to overall gaze shifts, which include the contribution of both head and eye movements. Others have found a dependence of head movement amplitude on the overall size of the gaze shift [1]–[3].

We then analyzed the distribution of head orientation angles and amplitudes. Figure 3a shows that the majority of head movements occurred along the cardinal axes, with horizontal movements being more prevalent than vertical movements. Movements along oblique directions were relatively rare. This head movement pattern is again consistent with the biomechanics of the neck, and represents a strategy in which subjects largely explored images using horizontal and vertical scanning movements. We found this strong bias towards horizontal head movements for all image categories, even though in the scenes, points of interest could lie anywhere in the image. Even though our images extended more in the horizontal than in the vertical direction, their aspect ratio was only 4:3, much smaller than the horizontal to vertical ratio of head movement amplitudes. Our results suggest that head movements may have re-oriented the direction of gaze, but eye movements (not recorded) may have been responsible for shifting gaze to different salient image locations. A recent study found a strong equator bias when viewing 360-degree panoramic images in virtual reality [11]; however, the images used in that study all contain a strong horizon line, which may have biased their results.

Figure 3b shows that movement amplitudes (as defined by our metrics) were typically small, with about 80% of movements falling below 40 degrees. The largest recorded head movements were about 100 degrees, while the lowest head movements were below 5 degrees (our bin size). We also observed that the distribution of head movement amplitudes decreased very close to linearly, and this was true for all image categories. We do not have a clear explanation for this phenomenon; understanding it is an area for further research. Our results are consistent with previous experimental findings [4], which showed similar patterns in the distribution of head movement angles and amplitudes when viewing natural scenes (although in that study, velocity rather than position was reported). Different from our study, Einhauser et al [4] did not use a velocity threshold to separate head movements into fixations and non-fixations. As a result, their results pool data from both types of head movements, while our results in Figure 3 are exclusively for periods when the head was moving (non-fixations).

To more directly compare our results with ref [4], we also looked at the distribution of head velocities without separating head fixations from non-fixations, shown in Figure 4. The distribution of head velocity angles shows again a strong anisotropy for horizontal velocities, but the difference at other orientations was less pronounced than for the head movement amplitudes, Fig. 4a. Figure 4b shows that the head velocity magnitude distributions also differed from the head amplitude distributions, Fig. 3b. While the head amplitude distributions were largely linear (relative frequency decreasing with amplitude), the velocity magnitude distributions show a decelerating decrease, with a larger fraction of low-velocity movements compared to high-velocity movements. The tail of the distribution was very long, indicating that high velocity movements were also recorded in our data.

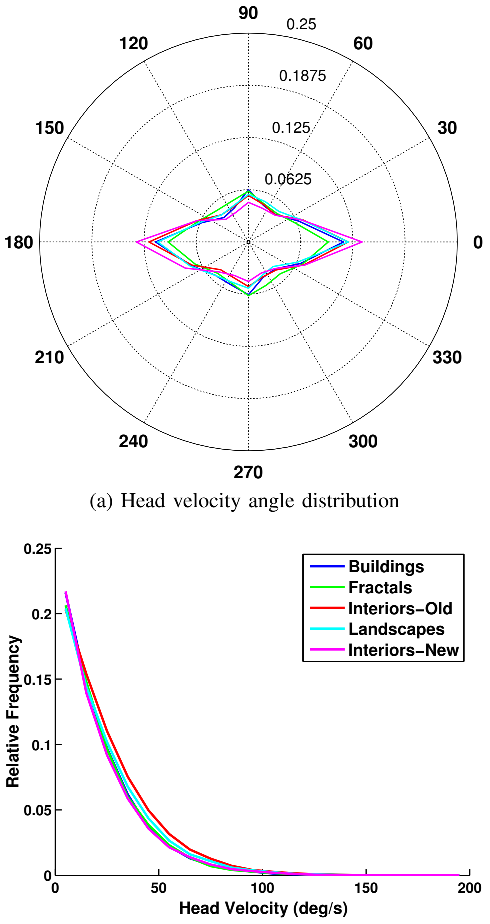


(a) Head movement angle distribution

(b) Head movement amplitude distribution

Fig. 3. Head position angle and amplitude distributions during periods when the head was actively moving. (a) Histograms of head movement angles (binned at eight orientations) are normalized to unit integral for comparison. Colors denote different image categories, with the legend shown in (b). Directions in the panel (left-right, up-down) correspond to direction of head movement. (b) Histograms of head movement amplitudes, with a bin width of 5 degrees.

Einhäuser et al [4] found a similar trend in their data, with the majority of head movements having low velocities. Overall, however, we found much higher head velocities in our experiment. While only a small proportion ($< 10\%$) of head movements exceeded 30 degrees/second in their set up, our head velocity magnitude distribution has a much longer tail, with many more head movements that exceeded this threshold. We do not know the exact cause of this difference in head movement velocities between the experiments, although we note that the behavioral task was different in the two cases. For our experiments, subjects had to view images and remember them in order to recall what they saw in the scene. We also placed a time limit of ten seconds for image viewing, which meant subjects had to use head movements efficiently in order to explore the image. In contrast, in the set of experiments in ref. [4], subjects moved freely in real-world environments and did not have any specific task constraint placed on them. Many of the recorded eye and head movements may have been



(a) Head velocity angle distribution

(b) Head velocity magnitude distribution

Fig. 4. Head velocity angle and magnitude distributions for all movements (head fixations and non-fixations). (a) Histograms of head velocity angles (binned at eight orientations) are normalized to unit integral for comparison. Colors denote different image categories, with the legend shown below in (b). Directions in the panel (left-right, up-down) correspond to direction of head movement. (b) Histograms of head velocity magnitudes, with a bin width of 10 degrees/second.

reflexive, *e.g.* employed while navigating through different types of terrain.

Another reason for the slower eye movements in the Einhüser *et al.* study [4] may have been the weight and mechanical inertia of the equipment their participants had to wear. Although no specifics on the weight and mechanical characteristics of the EyeSeeCam system are given in that report, the system of multiple cameras, mirrors, servo motors, power supply *etc.* may have prevented participants to make the fast natural head movements that were possible for our participants who were fitted with comparatively light-weight equipment. Furthermore, while wearing the EyeSeeCam, the field of view of the participants in the Einhüser *et al.* study was not limited by the equipment, which allowed subjects to make larger amplitude eye movements in order to explore the environment. In contrast, in our experiment, by virtue of the constraints we put on image presentation in the virtual reality environment, the image aperture required that head movements be used in order to fully explore an image. As a result, eye movements were less useful in our experiment, as they could only explore parts of the image that had been uncovered by head movements. As a result, the head movement velocities may be increased in our experiment as a consequence of a different exploration strategy.

Our head velocity magnitude distribution also matches recent results showing a similar distribution in longitudinal head velocities [11]. In that study, the head velocity distributions were separated based on whether the eyes were fixating or not fixating. Interestingly, these authors found a non-linear decrease in head velocities only when the eyes were fixating. When the eyes were not fixating, the distribution of head movement velocities became linear. They proposed that the different velocity distributions correspond to two distinct modes: exploration, when the head is relatively still and the eyes are fixating, and re-orienting, where the head is moving to new image regions. While we did not record eye movements during our experiment, the non-linear distribution of head movement amplitudes that we observed suggest that the eyes were mainly fixating while the head was moving.

Finally, we did not find substantial differences in average head movement kinematics across image categories. This result suggests that the underlying head movements subjects made when exploring different image categories were largely the same. However, this analysis ignores the temporal and location-specific aspects of head movements, which may serve as the basis for the diversity in movement trajectories that we observed. While we found that head movements themselves were quite stereotyped, when subjects choose to move their heads compared to when they choose to fixate, as well as where they choose to direct their heads within images, may be critical for understanding how subjects perceive and attend to different parts of the image. One area of future study will be to understand where subjects oriented their head when viewing images, and whether these locations correlate to previous findings on eye fixations [13] or interest points [14] as well as to predictions of gaze control by computational models of attentional selection [15]–[18].

IV. Conclusion

VR systems represent an emerging technology that could change how our society interacts with visual content. While VR allows for the creation of new and immersive experiences, many questions still remain. How do people explore content in VR scenes? What kinds of head movements do users make? Where do people look in VR scenes, and what items draw people's attention? We developed a novel experimental setup that allowed us to record head movements of human participants as they viewed natural images in a VR environment. Our results give insight into head movement kinematics during natural visual exploration. Our behavioral data may also be useful in informing future designs of VR systems and user interfaces. An interesting extension of our work would be simultaneous recording of eye and head movements.

Acknowledgment

This work is supported by the Office of Naval Research under Grants N000141010278 and N00014-15-1-2256 and the National Institutes of Health under Grant R01EY027544. We thank Dr. Kathleen Cullen for helpful discussions.

References

[1] Y. Fang, R. Nakashima, K. Matsumiya, I. Kuriki, and S. Shioiri, "Eye-head coordination for visual cognitive processing," *PloS one*, vol. 10, no. 3, p. e0121035, 2015.

[2] E. G. Freedman and D. L. Sparks, "Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys," *Journal of neurophysiology*, vol. 77, no. 5, pp. 2328–2348, 1997.

[3] M. K. McCluskey and K. E. Cullen, "Eye, head, and body coordination during large gaze shifts in rhesus monkeys: movement kinematics and the influence of posture," *Journal of neurophysiology*, vol. 97, no. 4, pp. 2976–2991, 2007.

[4] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König, "Human eye-head co-ordination in natural exploration," *Network: Computation in Neural Systems*, vol. 18, no. 3, pp. 267–297, 2007.

[5] R. Nakashima and S. Shioiri, "Why do we move our head to look at an object in our peripheral region? Lateral viewing interferes with attentive search," *PloS One*, vol. 9, no. 3, p. e92284, 2014.

[6] ——, "Facilitation of visual perception in head direction: Visual attention modulation based on head direction," *PloS one*, vol. 10, no. 4, p. e0124367, 2015.

[7] C. J. Wilson and A. Soranzo, "The use of virtual reality in psychology: A case study in visual perception," *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[8] D. Parkhurst, K. Law, and E. Niebur, "Modelling the role of salience in the allocation of visual selective attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, 2002.

[9] D. Parkhurst and E. Niebur, "Scene Content Selected by Active Vision," *Spatial Vision*, vol. 16, no. 2, pp. 125–54, 2003.

[10] N. C. Anderson, F. Anderson, A. Kingstone, and W. F. Bischof, "A comparison of scanpath comparison methods," *Behavior research methods*, vol. 47, no. 4, pp. 1377–1392, 2015.

[11] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, and G. Wetzstein, "Saliency in VR: How do people explore virtual environments?" *ArXiv e-prints*, Dec. 2016.

[12] A. T. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements," *Mathematical Biosciences*, vol. 24, no. 3-4, pp. 191–204, 1975.

[13] D. Parkhurst and E. Niebur, "Variable resolution displays: a theoretical, practical and behavioral evaluation," *Human Factors*, vol. 44, no. 4, pp. 611–29, 2002.

[14] C. Masciocchi, S. Mihalas, D. Parkhurst, and E. Niebur, "Everyone knows what is interesting: Salient locations which should be fixated," *Journal of Vision*, vol. 9, no. 11, pp. 1–22, October 2009, pMC 2915572.

[15] E. Niebur and C. Koch, "Control of Selective Visual Attention: Modeling the "Where" Pathway," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, vol. 8, pp. 802–808.

[16] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based fast visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[17] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Neuroscience*, vol. 2, pp. 194–203, 2001.

[18] A. F. Russell, S. Mihalas, R. von der Heydt, E. Niebur, and R. Etienne-Cummings, "A model of proto-object based saliency," *Vision Research*, vol. 94, pp. 1–15, 2014.