https://doi.org/10.2352/EI.2024.36.2.SDA-347

Manipulating Viewing Distance and Camera Toe-in in a Stereoscopic Remote Vision System

***Eleanor O'Keefe*[1], *Eric S. Seemiller*[2], *Marc Winterbottom*[2], *and Steven Hadley*[2].**
[1]KBR, Beavercreek, OH, USA
[2]711th Human Performance Wing, Operational Based Vision Assessment Laboratory, Wright-Patterson AFB, OH, USA

Abstract

Stereoscopic 3D remote vision system (sRVS) design can be challenging. The components often interact such that changing one parameter will cause unintended distortions in the perceptual image space. For example, increasing camera convergence to reduce vergence-accommodation mismatch will have the unintended effect of increasing depth compression. In this study, we investigated the trade-offs between changes in two parameters: viewing distance and camera toe-in. Participants used a simulated telerobotic arm to complete a precision depth matching task in an sRVS environment. Both a comfort questionnaire (subjective) and eye-tracking metrics (objective) were used as indicators of visual stress. The closer viewing distance increased both depth matching performance and objective measures of visual stress, demonstrating the inherent trade-offs associated with many sRVS design variables. The camera toe-in had no effect on either user performance or comfort. While these results suggest that small amounts of camera toe-in may be more tolerable than larger manipulations of viewing distance, the consequences of both should be carefully considered when designing an sRVS.

Background

The use of stereoscopic remote vision systems (sRVS) is increasing in industry, military, and medical applications [1-4]. Design of such systems can be complex, as distortions in perceptual image space will be introduced unless the parameters of image capture and display exactly match the user's visual system (i.e., orthostereo) [5-7]. While sRVS designers should aim to minimize distortions, it is often impossible to eliminate them completely. Further, the effect of sRVS design variable manipulations and interactions are poorly understood. Thus, characterizing any effects of resulting distortions is critical for understanding both usability and comfort.

One modifiable element of an sRVS is the convergence angle of the stereoscopic cameras, which can be manipulated in one of two ways. The optical axes of the two cameras can be rotated inward, called camera toe-in. Alternatively, the camera sensors can be selectively cropped or physically shifted. This method, camera sensor shift, often requires additional computational processing and can reduce the camera field-of-view (FOV). On the other hand, camera toe-in distorts the resulting image space by producing both unresolvable dipvergence (vertical misalignment between the image channels) and depth plane curvature where horizontal motion in object space can appear as both motion-in-depth and horizontal motion in image space, [5,8].

The viewing distance of the display is also modifiable with trade-offs to consider. A nearer viewing distance generally increases both image space depth compression and vergence-accommodation (VA) mismatch. Increased depth compression can lead to perceptual inconsistencies including velocity misperception when an object moves in depth and incorrect size cues [5,9]. Under natural viewing conditions, the eyes both focus and converge to the depth of the object of interest. VA mismatch occurs with any sRVS where the eyes focus at the screen distance but converge to the stereo imagery. This inconsistency in depth cues can cause discomfort and lead to decreased task performance [10,11]. However, a nearer viewing distance likely increases the usability of binocular disparity, a 3D depth cue, which is of greater value for distances within 1-meter [12, c.f. 13].

A particular concern for sRVS use is user discomfort, specifically eye strain. In previous studies we have used variations of a visual comfort questionnaire based on those developed by Shibata et. al. [10] and Kennedy et. al. [14]. However, subjective stress metrics are limited in their sensitivity to internal psychological changes, and we have found that voluntary complaints participants make verbally to researchers during sRVS task performance are not reflected in their respective survey answers [15]. Another means of gaining insight into visual strain is to analyze objective measures related to the user's oculomotor behavior. Informative oculometrics include pupil size, pupil velocity, and saccade statistics [16-21].

In this study we investigated how changes in both camera toe-in and user viewing distance affected performance on a telerobotic precision depth task in an sRVS environment. We also measured changes in visual comfort, comparing eye tracking metrics and subjective survey results. Our findings exposed a tradeoff between performance and eye strain with viewing distance, while camera toe-in produced no significant differences. The resulting oculometrics provided insight into visual strain while the survey showed no significant differences. Not only do these results provide useful information for sRVS designers, but they advocate for the importance of eye tracking metrics as a tool for gaining insight into visual comfort information where subjective surveys are lacking.

Methods

Participants

The study was carried out at Wright-Patterson Air Force Base (AFB) and accomplished in accordance with approved Air Force Research Laboratory Institutional Review Board human subjects research protocols. Forty participants were recruited though two were unable to complete the experiment.

Apparatus

Temporally interlaced stereoscopic images were presented using a SONY 4k projector (VPLGTZ280) at 60 frames per second to each eye, with a standard VPLL Z7013 lens (SONY, Tokyo, Japan). Temporal interlacing is a technique wherein full frame 2D imagery is presented to one eye at a time on alternating frames. Projection images were shown on a Proscreen Proflight 1.0 Gain HCWA Tint Cast Acrylic rear projection screen (Proscreen, Inc., Oregon, USA). During testing, participants viewed the display through active shutter 3D Volfoni Edge RF glasses. The Volfoni IR emitter failed 42% of the way through the study and the remaining sessions were completed using SONY TDG-BT500A glasses. Projector contrast was adjusted so that measured luminance through the new shutter glasses (70 cd/m^2) remained the same. No difference in outcome metrics were found between the sessions completed before and after the change in glasses. Participants used a flight control stick (Saitek X56 flight controllers; Logitech, Lausanne, Switzerland) to perform the task. The experimental apparatus is shown in Figure 1.

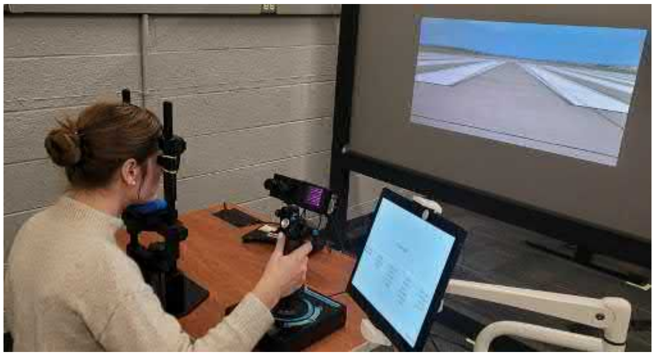


Figure 1. *Apparatus showing chin rest, eye tracker, RVS screen, and survey tablet.*

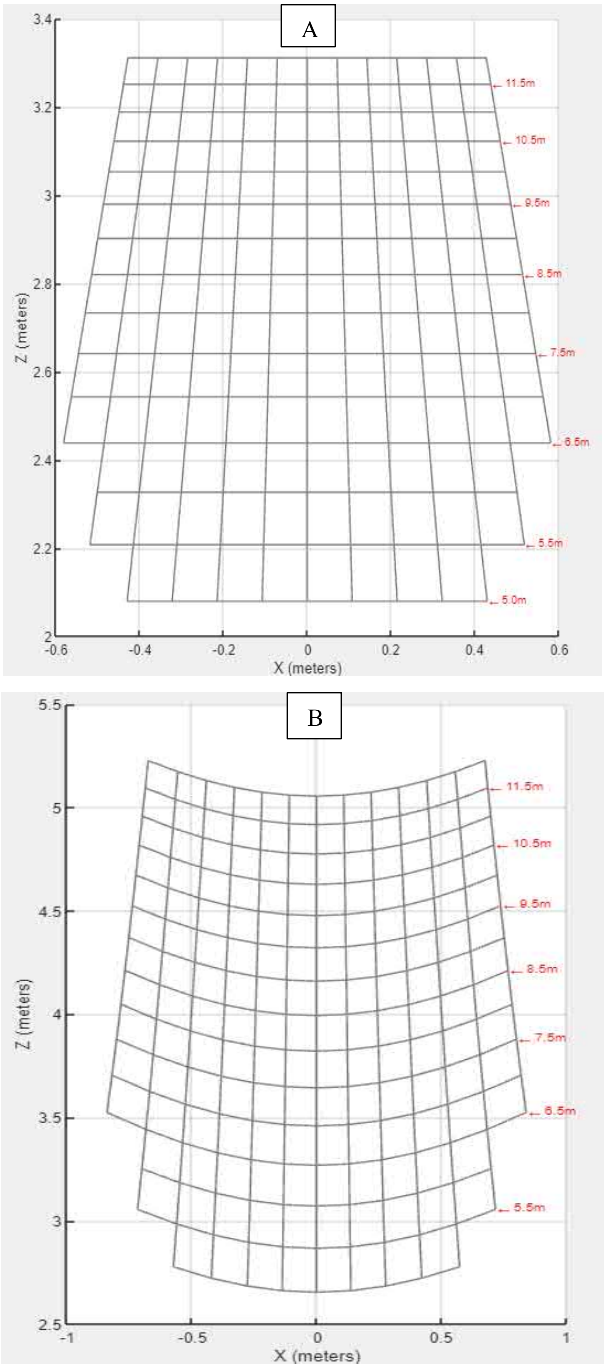
3D imagery was rendered using a multi-channel FlightSafety International Vital 1100 image generation (IG) system (FlightSafety International Visual Systems, Columbus, OH). Stereoscopic images were generated from two IG channels, each using an Nvidia Quadro P6000 video card (Nvidia, Santa Clara, CA) while frame buffers were synchronized using NVidia Quadro Sync II cards.

Two sRVS design variables, viewing distance and camera toe-in, were manipulated in a fully crossed design. Hyperstereo (i.e., camera separation) was also manipulated but those results are discussed elsewhere [22]. The complete set of viewing conditions are summarized in Table 1 along with computed depth compression and VA mismatch values. Three examples of the resulting image space distortions are shown in Figure 2. Dipvergence values across image space with camera toe-in at the near viewing distance is shown in Figure 3; note that with no camera toe-in, dipvergence values are zero. All calculations are based on analysis described by Woods et. al., 1993 [23].

Binocular eye position was recorded at 1000 Hz using an EyeLink 1000 Plus (SR Research Ltd., Ottawa, Canada) following a 13-point calibration. Participants rested their head on a chin rest throughout the experiment for stability. Ballistic saccades were detected based on their velocity profile [24] and the frequency of saccades was quantified as an estimate of visual stress [25-27]. Average pupil size and average pupil size velocity were also calculated. All three measurements were averaged at the trial level.

Table 1. Depth compression and VA mismatch values for all 3x hyperstereo conditions at x = 0 and z = 12m.

Viewing Distance (m)	Camera Toe-In	Display FOV (°)	Depth Comp. (%)	VA Mis (D)
1.70	None	29.02	22.1	0.29
1.70	0.37°	29.02	-20.0	0.39
0.96	None	49.25	116.2	0.51
0.96	0.37°	49.25	41.6	0.69



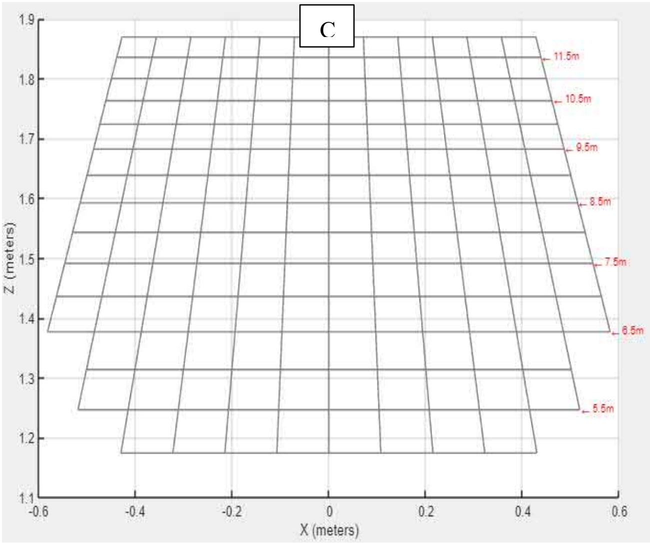


***Figure 2**. Image space distortions of 3 viewing conditions. A: Far viewing distance, no camera toe-in (parallel cameras). B: Far viewing distance, 0.37° camera toe-in. C: Near viewing distance, no camera toe-in. Note the scale change between images to represent the same object space.*

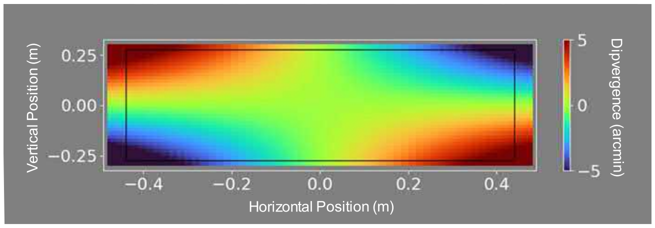


***Figure 3.** Dipvergence map for the near viewing distance with 0.37° camera toe-in.*

Tasks

Participants were instructed to drop a 6-in ball into a 12-in diameter receptacle that was situated on the ground of a runway environment (Figure 4). The ball was held by a telerobotic arm and gripper. The gripper arm hovered to prevent the use of arm geometry as a depth cue. For the same reason, there were no shadows or lighting cues in the simulation. The participant used the flight control stick to move the gripper around the image space until they felt it was directly over the receptacle, aiming as close to the center of the receptacle as possible. They pressed the joystick trigger to drop the ball and pressed it again to start the next trial. Participants were told that speed would be recorded, but accuracy was more important. On each trial the receptacle was randomly placed between 6 and 12 m downrange, and up to 1.5 m left or right of the origin. Performance was calculated in two ways: error (the distance from the ball and the center of the target), and gripper velocity (the distance between the receptacle and the starting point of the gripper divided by the length of the trial).

Participants completed the questionnaire on visual comfort after each condition of the gripper task (two times per session). This six-item questionnaire addressed issues associated with the use of stereoscopic displays, including: eye strain, eye tiredness, vision blur, headache, refocus, and eye dryness (modified from Shibata et. al. [10]). Participants responded to each question using a five-point Likert scale. The questionnaire was administered electronically.

Each session lasted 30 minutes during which the participant completed as many trials as possible. There was a mandatory 15-minute rest period between each session after they completed the visual comfort questionnaire. Each participant ran two sessions per day over the course of 4 days.

***Figure 4.** Gripper task (2D for clarity)*

Results

Gripper task performance

Overall, participants made significantly smaller errors ($F(1,37) = 10.66$, $p = 0.002$, $d = 0.24$) at the near viewing distance as compared to the far viewing distance (Figure 5). The same trend was true as measured by gripper velocity ($F(1,37) = 15.72$, $p = 0.001$, $d = 0.14$), wherein participants were faster to complete each trial in the near viewing condition (Figure 6). There was no significant difference across the camera toe-in conditions for error ($F(1,37) = 2.85$, $p = 0.10$) or gripper velocity ($F(1,37) = 0.54$, $p = 0.14$). Results are averaged across hyperstereo conditions. There were no significant interactions between viewing distance and camera toe-in for any of the performance metrics.

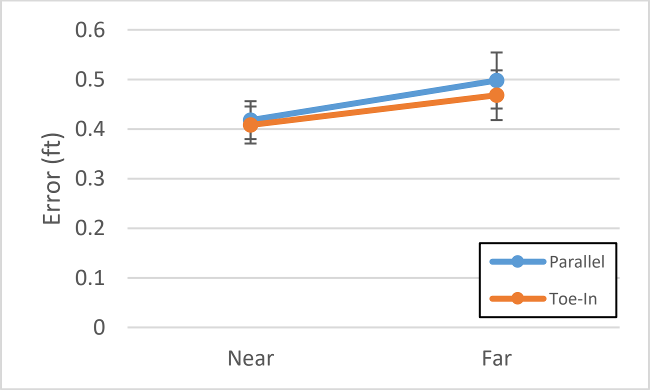


***Figure 5.** Average error for near and far viewing distances with parallel cameras and 0.37° toe-in. Error bars represent standard error.*

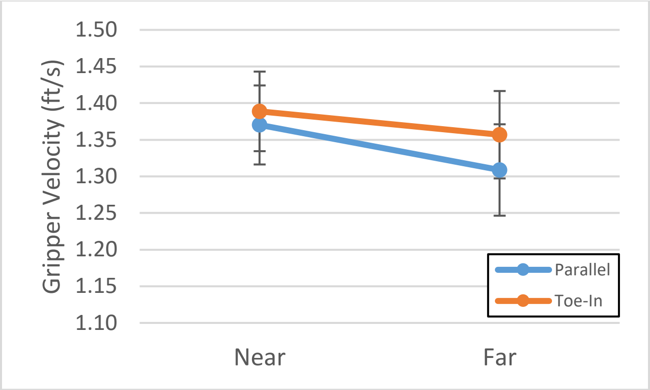


Figure 6. *Average gripper velocity for near and far viewing distances with parallel cameras and 0.37° toe-in. Error bars represent standard error.*

Eye tracking metrics

The near viewing distance led to a significant increase in saccade frequency ($F(1,24) = 114.43$, $p < 0.001$ $d = 1.46$) and significantly smaller pupil sizes ($F(1,23) = 5.31$, $p = 0.03$ $d = 0.22$) as compared to the far viewing distance (Figures 7, 8) . There was no significant difference in either saccade frequency ($F(1,23) = 2.37$, $p = 0.14$) or pupil size ($F(1,23) = 0.72$, $p = 0.40$) across the camera toe-in conditions. Pupil size velocity showed no significant differences across either viewing distance ($F(1,23) = 0.80$, $p = 0.38$) or camera toe-in ($F(1,23) = 3.00$, $p = 0.10$; see Figure 9). There were no significant interactions between distance and camera toe-in conditions for any of the oculomotor metrics. Results are averaged across hyperstereo conditions.

Subjective visual comfort scores were not statistically significantly different between conditions, using the sign test (see Table 3).

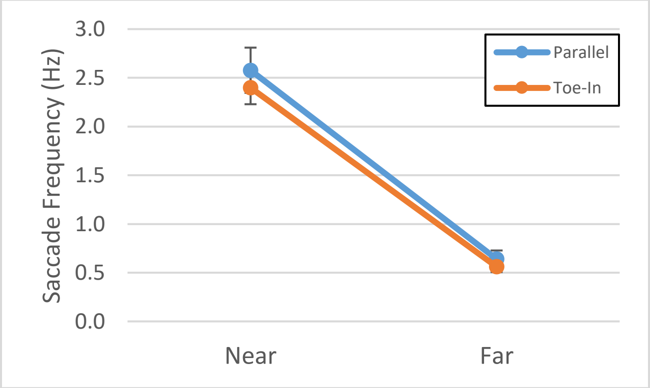


Figure 7. *Top: Average saccade frequency for near and far viewing distances with parallel cameras and 0.37° toe-in. Error bars represent standard error.*

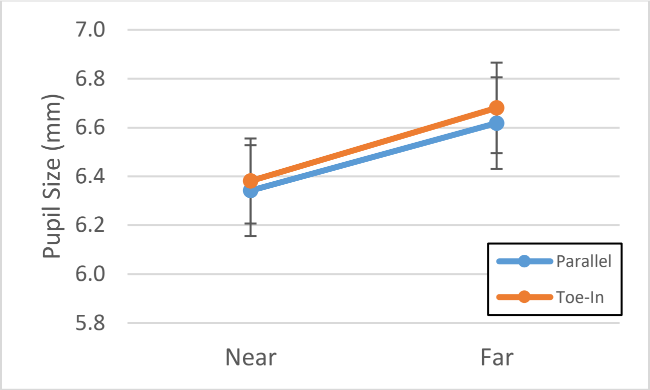


Figure 8. *Average pupil size for near and far viewing distances with parallel cameras and 0.37° toe-in. Error bars represent standard error.*

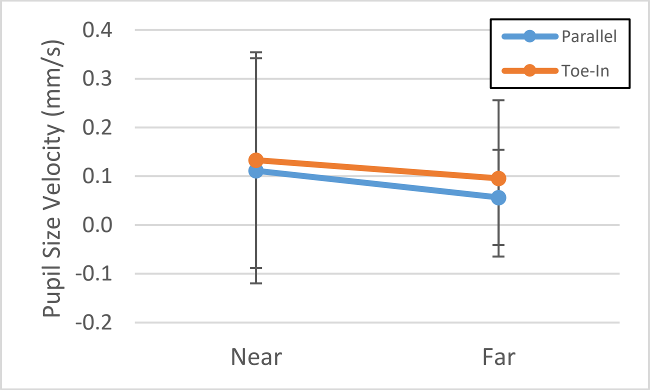


Figure 9. *Average pupil size velocity for near and far viewing distances with parallel cameras and 0.37° toe-in. Error bars represent standard error.*

Table 3. Visual comfort questionnaire results.

Survey Question	*F* approximation	*p*-value
Eyestrain	0.04	0.96
Eye tiredness	0.10	0.90
Vision blur	0.10	0.90
Headache	0.42	0.66
Fatigue	0.07	0.93
Nausea	0.24	0.79

Discussion

This study investigated the effect of two sRVS design variables, viewing distance and camera toe-in, on task performance and visual fatigue. Overall, results showed a trade-off between task performance and visual comfort with viewing distance, but no effects of a small degree of camera toe-in.

The near viewing distance condition led to an increase in task performance as quantified by both smaller error and faster task completion. This suggests that the benefit of the near stereopsis cue outweighed any effect of the larger amounts of depth compression and VA mismatch (as compared to the far viewing condition; see Table 1). However, the increase in saccade frequency, an indicator of visual stress, may have been related to those image space distortions [5-9].

The near condition also led to smaller measured pupil sizes – another potential indicator of eye strain [16, 18]. However, if this decrease in pupil size was due to visual strain, we would expect to see an increase in pupil size velocity in the same conditions, which we did not. Therefore, it is more likely that this result is related to the physiological effort of the near oculomotor response. During this response, a small pupil increases the depth of focus and reduces accommodative demand, potentially mitigating some discomfort from VA mismatch [28]. This trade-off between increased performance and increased eye strain should be considered by sRVS designers as the strain may become a more significant problem with longer viewing periods, particularly for tasks where rest periods may not be practical for the sRVS operator (e.g., air refueling, telesurgery, ordinance disposal, etc.).

Surprisingly, camera toe-in did not lead to any observable changes in task performance or visual comfort. In previous research using a simulated air refueling task, we have found that camera toe-in contributes to impaired performance and visual stress [15, 29]. However, simulated air refueling was a much more dynamic task with motion spanning the users' FOV, whereas the gripper task in the present study was at a relatively slow speed and generally confined to the center of the screen. In addition, the sRVS design in the air refueling simulation had a larger camera toe-in which resulted in a large amount of dipvergence. The dipvergence for the toed-in condition in the present study was within acceptable tolerances (< 5 arcmin) throughout the center of the display and only exceeded tolerances in the far corners (Figure 3) [30]. Larger degrees of toe-in could not be tested here due to design limitations. Therefore, a small amount of toe-in, similar to the degree tested here, may be acceptable in sRVS design in some applications. However, any amount of toe-in will create depth plane curvature and dipvergence which may be problematic.

The null survey results found here are consistent with our previous research [15] and may indicate that the physiological stressors hadn't yet met the threshold for awareness and reporting. It is likely that oculometric assessments of visual stress are generally more sensitive than subjective surveys. Extended sRVS usage may still lead to significant subjective reports, but the sensitivity of oculometric variables is likely to reduce overall human factors study time in addition to providing more reliable metrics for determining the state of internal physiological activity (e.g., eye strain) [30,31].

Conclusion

This work adds to the increasing evidence that there is no one-size-fits-all ideal design for any stereoscopic remote vision system. We demonstrated that viewing distance changes result in a trade-off between performance and comfort. And, while the results of this study showed no effect of camera toe-in on performance or visual stress, the consequences of toe-in should generally be avoided in sRVS displays.

Disclosures

The views expressed are those of the authors and do not reflect the official views of the United States Air Force, nor the Department of Defense. Mention of trade names, commercial products, or organizations do not imply endorsement by the U.S. Government. This research was supported by USAF contract FA8650-21-C-6277 to KBR. Cleared for public release (AFRL-2023-1172).

References

[1] P. J. Choi, R. J. Oskouian, & R. S. Tubbs, "Telesurgery: Past, present, and future," Cureus, vol. 10, no. 5, 2018.

[2] S. Livatino, G. Muscato, & F. Privitera, "Stereo viewing and virtual reality technologies in mobile robot teleguide," IEEE Transactions on Robotics, vol 25, no. 6, pp. 1343-1355, 2009.

[3] P. Kauff & O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," in Proc. of the 4th international conference on Collaborative virtual environments, 2002.

[4] E. O'Keefe, M. Ankrom, E. S. Seemiller, T. Bullock, M. Winterbottom, J. Knapp, & S. Hadley, "The relationship between vision and simulated remote vision system air refueling performance," Electronic Imaging, vol. 34, pp.1-6, 2022.

[5] A. Woods, T. Docherty, & R. Koch, "Image distortions in stereoscopic video systems," Stereoscopic Displays and Applications IV, vol. 1915, 1993.

[6] R. T. Held & M. S. Banks, "Misperceptions in stereoscopic displays: A vision science perspective," in Proc. of the 5th symposium on Applied perception in graphics and visualization, 2008.

[7] Z. Gao, G. Zhai, & X. Yang, "Stereoscopic 3D geometric distortions analyzed from the viewer's point of view," PloS one, vol. 15, no. 10, 2020.

[8] R. S. Allison, "The camera convergence problem revisited," Stereoscopic Displays and Applications XI, vol. 5291, 2004.

[9] D. Diner, "Danger of collisions for tele-operated navigation due to erroneous perceived depth accelerations in 3-D television," in Annual Meeting of the American Nuclear Society, 1991.

[10] T. Shibata, J. Kim, D. M. Hoffman, & M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," Jour. of Vision, vol. 11, no. 11, pp. 1-29, 2011.

[11] D. M. Hoffman, A. R. Girshick, K Akeley, & M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," Jour of Vision, vol. 8, no.33, pp. 1-30, 2008.

[12] J. E. Cutting & P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," Handbook of Perception and Cognition, vol. 5, pp. 69-117, 1996.

[13] R. S. Allison, B. J. Gillam, & E. Vecellio, "Binocular depth discrimination and estimation beyond interaction space," Jour of Vision, vol. 9, no. 1, pp. 1-14, 2009

[14] R.S. Kennedy, N. E. Lane, K.S. Berbaum, & M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," The International Jour. of Aviation Psych., vol. 3, no. 3, pp. 203-220, 1993.

[15] E. O'Keefe, K. Moffit, E. S. Seemiller, M. Winterbottom, J. Knapp, & S. Hadley, "Investigating the use of spectacle lenses to alleviate vergence-accommodation mismatch in a stereoscopic remove vision system," Electronic Imaging, vol. 35, pp. 1-6, 2023.

[16] S. Saito, M. Sotoyama, S. Saito, & S. Taptagaporn, "Physiological indiced of visual fatigue due to VDT operation: Pupillary reflexes and accommodative processes," Industrial Health, vol. 32, pp. 57-66, 1994.

[17] C. F. Chi & F. T. Lin, “A comparison of seven visual fatigue assessment techniques in three data-acquisition VDT tasks,” Human Factors, vol. 40, no. 4, pp. 577-590, 1998.

[18] A. J. Zele & P. D. Gamlin, “The pupil: behavior, anatomy, physiology and clinical biomarkers,” Frontiers in Neurology, vol. 11, 211, 2020.

[19] M. Nakayama, K. Takahashi, & Y. Shimizu, “The act of task difficulty and eye-movement frequency for the ‘Oculo-motor indices’,” in Proc. of the 2002 symposium on Eye tracking research & applications, 2002.

[20] A. T. Duchowski, “Gaze-based interaction: A 30 year retrospective,” Computers & Graphics, no. 73, pp. 59-69, 2018.

[21] K. F. Van Orden, W. Limbert, S. Makeig, & T. P. Jung, “Eye activity correlates of workload during a visuospatial memory task,” Human factors, vol. 43, no.1, pp.111-121, 2001.

[22] E.S. Seemiller, E. O’Keefe, M. Winterbottom, & S. Hadley, “Usability of hyperstereo in stereoscopic remote vision systems,” In Review.

[23] A. Woods, T. Docherty, and R. Koch, R., “Image distortions in stereoscopic video systems,” SPIE 1915, Stereoscopic Disp. and Applications.

[24] R. Engbert & K. Mergenthaler, “Microsaccades are triggered by low retinal image slip,” in Proc. of the National Academy of Sciences, 2006.

[25] R. Schleicher, N., Galley, S., Briest, & L. Galley, “Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired?” Ergonomics, vol. 51, no.7, pp. 982–1010, 2008.

[26] S. Umemoto & Y. Hirata, “Emerging order of anomalous eye movements with progressive drowsiness,” Jour. of Vision, vol. 23, no. 1, pp. 17, 2023.

[27] H. Wakui & Y. Hirata, “Detection of reduced arousal by saccadic eye movement,” Japanese Soc. Med. Bio. Eng., vol. 51, pp. 328-341, 2014.

[28] P. J. May, S. Warren, P. D. R. Gamlin, and I. Billig, “An Anatomic Characterization of the Midbrain Near Response Neurons in the Macaque Monkey,” Invest. Ophthalmology Vis. Sci., vol. 59, no. 3, pp. 1486–1502, 2018.

[29] M. Winterbottom, "Individual Differences in the Use of Remote Vision Stereoscopic Displays," Wright State University CORE Scholar, Theses and Dissertations, 1310, 2015.

[30] K. Moffitt, “Designing HMDs for viewing comfort,” in Head mounted displays: Designing for the User. New York: McGraw-Hill, pp. 117-45, 1997.

[31] M. Emoto, T. Niida, & F. Okano, “Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television,” Jour. of Display Tech., Vol 1, no. 2, pp.328-340, 2005.

[32] I. Iatsun, M. C. Larabe, & C. Fernandez-Maloigne “Investigation of visual fatigue/discomfort generated by S3D video using eyetracking data,” in Proc. SPIE 8648, Stereoscopic Disp. and Applications XXIV, 2013.

Author Biography

Eleanor O’Keefe received her BS in biopsychology from the University of California, Santa Barbara (2010) and her PhD in experimental psychology from the University of Louisville (2017). Since then, she has worked in the Operational Based Vision Assessment Lab at Wright-Patterson Air Force Base in Ohio. Her work has focused on visual perception and human factors concerning advanced 3D technology design.