NAMRU-D-19-13 Nov-2018

NAVAL MEDICAL RESEARCH UNIT DAYTON

Adapting Virtual Reality and Augmented Reality Systems for Naval Aviation Training

by LT Adam T. Biggs, MSC, USN[1],Daniel J. Geyer[1,2] ,Valarie M. Schroeder[1,2],F. Eric Robinson[1], and John L. Bradley[3]

[1] Naval Medical Research Unit Dayton
[2] Henry M. Jackson Foundation for the Advancement of Military Medicine
[3] University of Pikeville Kentucky, College of Optometry

v.1 Oct2018

NOTICES

Disclaimer

The views expressed in this article reflect the results of research conducted by the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government.

Source of Support

This work was supported by Office of Naval Research and funded by work unit number z9PU.

Human Research/Institutional Review Board (IRB) statement

The study protocol was approved by the Wright-Patterson Air Force Base Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects.

Copyright

Adapting Virtual Reality and Augmented Reality Systems for Naval Aviation Training

by

LT Adam T. Biggs, MSC, USN, Daniel J. Geyer, Valarie M. Schroeder, F. Eric Robinson,
& LCDR John L. Bradley, MSC, USN

Foreword

This research was funded by U.S. Naval Air Systems Command under work unit number H1604. The authors have no financial or non-financial competing interests in this manuscript. The views expressed in this article reflect the results of research conducted by the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government. I am a military service member or federal/contracted employee of the United States government. This work was prepared as part of my official duties. Title 17 U.S.C. 105 provides that `copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. 101 defines a U.S. Government work as work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

Table of Contents

Summary Overview

Naval aviators face substantial training requirements because their operational responsibilities involve some of the most complex and challenging cognitive tasks of any military activity. Specific training activities vary, but pilots must be able to perform and adapt to extreme scenarios and high-risk maneuvers—which in turn requires high-risk training. New initiatives and technology create emerging opportunities to improve training so that all naval aviation training, from the most mundane in-flight activities to the most extreme high-risk maneuvers, will more effectively prepare naval aviators to perform their duties and meet mission requirements. Among these new training platforms, virtual reality (VR) and augmented reality (AR) stand out as some of the most promising initiatives. Both VR and AR technologies create simulated environments designed to immerse the user in a world or activity that does not actually exist. Commercial entities are exploring these platforms for gaming and general entertainment, but these simulations offer intriguing possibilities for naval aviation as well. For example, virtual platforms provide practice opportunities for high-end tactics and procedures in forward operating environments where cumbersome training infrastructure cannot be transported for either cost efficiency reasons or operations security issues. A virtual approach provides vital experience to the aviator in a safe, cost-effective environment that allows repetition and control over key aspects of the flight. In some cases, virtual simulation is the only safe way to create the training scenario because the scenario being simulated involves life-or-death decisions.

VR/AR platforms can create these training scenarios for our aviators, although it is important to understand the relative differences of each platform. VR systems are immersive multimedia that simulate a physical presence in an entirely virtual environment. AR systems blend computer-generated elements with the real-world environment to create a hybrid of virtual elements and reality. The end result is a thoroughly immersive environment that still allows for real-world interactions with other stimuli or other people—virtual or real in both cases.

Recently, substantial improvements have taken VR/AR systems from blocky, artificial constructions to increasingly immersive experiences. More realistic and immersive experiences increase the potential range of capabilities to be trained, which makes both technologies increasingly appealing as aviation training platforms. The problems arise when determining how to implement these new systems into training programs. For example, VR is generally a known-but-improving quantity, whereas realistic AR systems are so new that very little is known about them or how they may differ from VR. There are also many unanswered aeromedical and human factors questions about the impact of these new technologies on pilot safety, training effectiveness, and practical implementation.

The current review examines four potential issues involved with integrating VR/AR systems into naval aviation training: simulator sickness, spatial disorientation countermeasures, fatigue/eye strain, and vergence-accommodation conflict issues. The review will begin with an overview of all four areas, including a brief summary of the full literature review, several takeaway questions/answers the authors found highly informative, and finally a subset of suggested further reading which the authors found most useful during the review process. The literature reviews for each topic are then presented in full at the end of the document. We hope these discussions will provide the reader with a better understanding of each VR/AR challenge as it pertains to naval aviation training as well as greater insight into the existing knowledge/application gaps for future research.

Simulator Sickness

Summary

Simulator sickness has been a problem for military aviation since the first simulators were introduced. It is believed to originate from conflicting sensory cues, such as the individual perceiving motion visually without analogous vestibular sensations. Simulator sickness differs from classical motion sickness in that oculomotor symptoms tend to be more prevalent than nausea, and vomiting is very rare. The problem has persisted despite the increasing fidelity and sense of immersion offered by new generations of simulators. No validated objective measurement has been identified due to the widespread individual variability in susceptibility and symptom development. Instead, symptoms are recorded subjectively via the Simulator Sickness Questionnaire (SSQ). Causal factors that have been identified include the user's available field-of-view (FOV), latency, fidelity, vection (the visually-induced sense of motion when there is none), length of time in simulator, and tasks performed in the simulator.

There are multiple types of simulators used in aviation training. Cave automatic virtual environments (referenced as CAVEs) are immersive virtual reality devices which use multiple projectors to direct images on walls and create an artificial environment. The F/A-18 Tactical Operational Flight Trainer (TOFT) would be an example of a CAVE. Alternatively, helmet- or head-mounted displays (HMDs) are another method to simulate virtual environments, which often attach directly to the display or the head to provide an immersive simulation. CAVEs represent the more standard simulation technique, whereas HMDs are a rarer, but developing method in aviation training. HMDs tend to average higher levels of sickness than CAVE simulators, with HMD users experiencing significantly more disorientation symptoms. Though HMDs tend to have higher SSQ scores, HMD studies typically use college student populations and CAVE studies typically use military aviators. This difference in symptomology has led to an ongoing debate about whether simulator sickness induced by CAVE-style simulators is different than the “cybersickness” elicited from HMDs. Researchers retain the term simulator sickness to reference all symptoms derived from all simulators. Commercial entities and media tend to use the term “cybersickness;” however, the term is rarely synonymous with the full array of simulator sickness symptoms, and is typically used only to reference nausea and vomiting. There is no concrete evidence that simulator sickness negatively affects training, although there is no research examining training retention beyond several hours. In several studies, subjects were trained on either HMDs or desktop computer displays, with the HMD condition inducing significant levels of sickness. Training retention and effectiveness were similar in both conditions, though it is possible that, sans sickness, HMDs would provide more efficacious training.

Actual incidence of simulator sickness has yet to be determined due to the wide array of simulator types and technologies. Likewise, the time course of symptom development and the duration of symptoms post-exposure are unknown, as no recent or reliable studies have followed simulator users beyond one-hour post exposure. This omission is especially problematic concerning any aftereffects. No research has reliably investigated the long-term results of simulator sickness post-exposure in the past 20 years, although several studies have identified symptoms such as ataxia (e.g., loss of muscle control and/or balance) existing for at least one hour post-exposure. AR devices are especially under-studied, and some researchers theorize AR will prompt even more sickness due to the merging of virtual and real environments.

The typical motion sickness countermeasures, such as pharmaceuticals or temporary occlusions in visual display, have not been fully tested for efficacy in simulators, and their ability to counter oculomotor symptoms are unknown. Commercial countermeasures in HMDs are not ideal for naval aviation training because the primary commercial solution (e.g., temporary occlusions based on head movements) obscures peripheral vision when making rapid head movements. Aviation communities will be understandably reluctant to accept a solution that directly and intentionally limits an aviator's visual field. The most successful countermeasure is adapting to the simulator through continued use. However, there is a fear that adaptation will lead to a rearrangement of sensory perception—an adaptation ultimately detrimental to the pilot once inside an actual aircraft. A uniform time course and severity of adaptation has not been identified in research due to the difference in technologies and populations used. It is possible that any sensory rearrangement due to adaptation can be countered with requiring simulator users to undertake re-adaptation tasks immediately after simulator use (e.g., hand-eye coordination tasks, balance tasks, etc.).

Takeaways

- What type of simulator causes simulator sickness?
    - Due to the wide variety of technologies involved, there can be no assumption of incidence made across simulator types. Generally, HMDs have higher sickness levels than any other formats. Among non-HMD military flight simulators, rotary-wing simulators have higher sickness rates than fixed-wing simulators.
- Does simulator sickness affect training capabilities?
    - There is no evidence that simulator sickness negates training effectiveness. However, there is no research looking at the relationship between long-term (more than one hour) retention and simulator sickness.
- Is simulated motion necessary for simulator sickness?

  - No, there is no evidence that the absence or presence of simulated motion is directly related to causing simulator sickness. However, it is theorized that if a simulator can exactly match the motions of the real craft being simulated, sickness will only occur during maneuvers that would prompt sickness in the real craft.
- What are potential causal factors of simulator sickness?
  - Individual susceptibility, the user's available field-of-view (FOV), latency, fidelity, vection (the visually-induced sense of motion when there is none), length of time in simulator, and tasks performed in the simulator
- What factors are, without question, known to be prerequisites of simulator sickness?
  - A sense of vection and unobstructed peripheral vision. Neither are causes in isolation, but both are required before simulator sickness can occur.
- What causal factors are the most likely to yield new countermeasures or insight into motion sickness from future research?
  - FOV and latency.
- What is the most effective means to prevent simulator sickness?
  - Adaptation through repeated use is the best means for countering simulator sickness. However, adaptation may cause a realignment of sensory cues, which in turn may be problematic for pilots in real aircraft. No research has thoroughly examined the effects of adaptation, although there is some evidence that re-adaptation strategies can mitigate negative adaptation factors, such as users undertaking hand-eye coordination tasks immediately after simulator use.
- What is the biggest gap in simulator research?

  - The biggest gap is in the time course and long-term aftereffects of simulator sickness and adaptation. No study in the past twenty years has examined simulator sickness symptoms beyond one hour, which is especially troubling due to the preponderance of oculomotor problems induced by simulator sickness.

Suggested Further Reading

- Champney, R. K., Stanney, K. M., Hash, P. A. K., Malone, L. C., Kennedy, R. S., & Compton, D. E. (2007). Recovery from virtual environment exposure: expected time course of symptoms and potential readaptation strategies. *Human Factors, 49*(3), 491-506.
- Fernandes, A. S., & Feiner, S. K. (2016, March 19-20). Combating VR sickness through subtle dynamic field-of-view modification. Paper presented at the 2016 IEEE Symposium on 3D User Interfaces (3DUI), Greenville, SC (201-210).
- Johnson, D. M. (2005). *Introduction to and review of simulator sickness research*. (ARI Report No. 1832). U.S. Army Research Institute Field Unit, Fort Rucker, AL.
- Kennedy, R. S., & Fowlkes, J. E. (1992). Simulator sickness is polygenic and polysymptomatic: Implications for research. *International Journal of Aviation Psychology, 2*(1), 23-38.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.
- Kinsella, A., Mattfeld, R., Muth, E., & Hoover, A. (2016). Frequency, not amplitude, of latency affects subjective sickness in a head-mounted display. *Aerospace Medicine and Human Performance, 87*(7), 604-609.

- Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors, 53*(308), 308-319.

Spatial Disorientation Countermeasures

Summary

Spatial disorientation (SD) remains one of the leading causes of flight mishaps despite decades of effort to address the problem. This report describes the potential for virtual reality (VR; flight simulation) and augmented reality (AR) to positively impact the effectiveness of SD training. SD occurs when a pilot fails to correctly perceive the relative position, motion, or attitude of the aircraft relative to the earth. Such misperceptions are most often due to inaccurate vestibular cues, or the alteration of visual spatial strategies due to loss of visual cues. Current SD training seeks to familiarize pilots with known vestibular and visual illusions via classroom training and practical demonstration, but this approach has not been successful in reducing the rate of SD mishaps. VR systems offer tremendous potential to deliver effective SD training, but we must first reexamine current training methods and evaluate how best to utilize available simulation before realizing the potential benefits of such technology.

First, we must consider how training is conducted currently. Current training focuses on vestibular illusions based on the known physiological mechanisms of SD. However, research indicates that visual processes (i.e., cognitive rather than physiological) may be more prominent in SD during flight. We should refocus training efforts on the types of SD that, based on mishap reports, are most likely to cause an accident. Further, current training centers on SD recognition. More research on the spatial strategies pilots use and the cognitive contribution to SD will help shift the focus of training from recognition to avoidance. Finally, training is most often evaluated by subjective feedback or the use of objective measures administered immediately after training. We must conduct longitudinal research to validate the long-term impact of SD training.

After identifying the most critical types of SD and gaining a better understanding of the cognitive strategies relevant for avoiding such illusions, we will be able to determine how best to utilize simulation to accomplish SD training. Simulator acquisition and use should be driven by learning requirements rather than new technological features, as the content of scenarios is more important than the fidelity of the simulator. Motion-based and high fidelity simulators offer potential value, but neither might be necessary. More research is needed to determine whether fixed-base simulators can address the range of required SD illusions for training. Finally, we should continue to evaluate the use of emerging technologies such as head-mounted displays and augmented reality. Such displays offer the potential to enhance training by facilitating access to simulations aboard ship or by helping instructors highlight key information, but may also have unintended consequences on training in the form of increased pilot head movements and altered spatial strategies. These possibilities must be evaluated prior to the adoption of such technology.

Takeaways

- How effective are current training methods?
    - Trainee and instructor feedback along with performance immediately post-training suggest that training is effective. However, no long-term studies of training effectiveness have been conducted and rates of SD mishaps have not changed despite falling mishap rates overall. This indicates that current training methods are not mitigating the most serious consequences of SD.
- What is the biggest impediment to effective SD training?

    - We see three main impediments to effective training based on current practices. First, evidence indicates that current training may overemphasize the role of vestibular illusions in SD mishaps and neglect visual and cognitive processes. Second, training focuses on recognizing SD rather than avoiding it. Finally, too much attention has been paid to physical simulator fidelity at the expense of how well simulators and scenarios actually support learning.
- What are some ways to better capitalize on current technologies?
    - Focusing on how well simulators support learning rather than on how real they "feel" will help improve training outcomes. Scenario fidelity in the simulator is more important than the fidelity of the simulator itself. Scenarios will better support training if they emphasize spatial strategies based on critical cues, whereas the simulator itself will better support training if it facilitates instructor manipulation and feedback.
- What new VR technology is needed to improve SD training?
    - Current technology appears capable of recreating nearly any type of SD with high visual and physical fidelity. Rather than promote new technical advancements, we should start investigating which capabilities are necessary to generate SD sensations as well as which capabilities genuinely add value to training and which do not. We should also investigate how best to employ those capabilities.
- What is the most prudent non-technology area for future investment?
    - We recommend investigating the cognitive contributions to SD, such as how pilots seek information from the world and prioritize/integrate multiple sources of information (e.g., visual, vestibular, or instruments) to form a sense of orientation.

    Understanding SD from this perspective—in addition to vestibular or visual illusions—will facilitate the identification of active steps that pilots can take to change their spatial strategies and reduce the risk of becoming disoriented. These strategies can then be incorporated into training.

- How can augmented reality be used to supplement SD training?
    - AR technologies can help highlight relevant spatial cues and demonstrate the interaction between them to promote more effective spatial strategies among pilots.
- What are the biggest risks of using augmented reality in SD training?
    - AR used in training may conflict with HUDs or other displays in aircraft simulators, potentially leading to issues with confusion or display clutter. Pilots may also become too dependent on the AR when seeking spatial cues.
- What are the biggest risks for using head-mounted displays in simulators?
    - Depending on the field of view, HMDs may affect how pilots move their heads and how they utilize visual cues. Such changes may alter how pilots experience SD and lead to ineffective or negative training.

Suggested Further Reading

- Cheung, B. (2013). Spatial disorientation: More than just illusion. *Aviation, Space, and Environmental Medicine, 84,* 1211-1214.
- Grierson, L. (2014). Information processing, specificity of practice, and the transfer of learning: Considerations for reconsidering fidelity. *Advances in Health Science Education, 19,* 281-289.

- Patterson, F., Arnold, R., & Williams, H. (2013). Visual perspective illusions as aviation mishap causal factors. *Paper presented at the International Symposium on Aviation Psychology, May 6-9, Dayton, OH.*
- Patterson, F., & Muth, E. (2009). Cybersickness onset with reflexive head movements during land and shipboard head-mounted display flight simulation. NAMRL report #10-43.
- Patterson, F., Williams, H., Folga, R., & Arnold, R. (2014). Identification of pilot spatial strategies for mitigation of disorientation: A new look at an old problem. *Paper presented at the annual meeting of the Aerospace Medicine Association.*
- Patterson, F., Williams, H., Folga, R., & Arnold, R. (2015). Effects of spatial disorientation countermeasures training on black hole illusion (BHI) onset distance and recovery altitude. *Paper presented at the annual meeting of the Aerospace Medicine Association.*
- Salas, E., Bowers, C., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8,* 197-208.

Fatigue/Eye Strain

Summary

Fatigue and eye strain can be distracting, increase cognitive load, and reduce the amount of time spent working—all of which can reduce training effectiveness. Research documenting the association between the use of VR and AR systems with fatigue and eye strain is limited. Despite promising preliminary studies, more information is needed to determine the impact these systems have with regards to fatigue and eye strain prior to implementation within shipboard training programs.

The fatigue literature is especially sparse. The majority of information gained from cognitive fatigue research is based upon viewing 3D displays and older Defence Evaluation and Research Agency (DERA) AR research. Results from these studies indicated 3D viewing increases cognitive load, and hence cognitive fatigue, whereas DERA research demonstrated AR systems could be equally as effective as standard PowerPoint training programs. However, two separate studies examining VR and AR systems suggest that cognitive load is lowered with more interactive uses of these technologies. Research examining physical fatigue, or fatigue due to sleep loss, and its association with VR and AR systems has mostly been conducted within the medical field. The majority of such research suggests no effects of fatigue on learning new surgical procedures using VR simulators. One study demonstrated a benefit to learning a procedural task using an AR platform. Such research is promising, but it needs to be explored further.

No studies assessing eye strain unrelated to accommodation and convergence issues with AR systems could be located. An especially informative study examined eye strain across various VR platforms. This study found that all but one platform produced a significantly increasing amount of eye strain and visual discomfort for the duration of the study. Additional research comparing 3D to 2D displays further attest to a problem of eye strain when using 3D versus 2D displays. The major cause of eye strain resulting is believed to be binocular disparity when using these displays, with spatial distortions, poor filters, and stereoscopic conflicts impacting eye strain the most.

Based on this limited information, it appears that incorporating interactive components and programs within VR and AR systems could counter the negative effects of both cognitive and physical fatigue. These active components would require the participant to engage or respond to certain situations or prompts. Such interactive programs likely allow for greater engagement within the program, resulting in increased arousal and greater motivation to perform well. Similarly, research is currently underway to combat the effects of binocular disparity within VR/AR displays in an effort to reduce eye strain. Studies examining the effectiveness of the application of filters, image adjustments, and calibrations have been proposed, but these proposed studies are still in their infancy.

Additional research is recommended on the viability of these systems prior to shipboard utilization as it is unclear if the studies identified within this report can generalize to success among Navy shipboard trainers. The following is a short list of issues that should be addressed:

1. Given that interactive components within VR and AR programs lessen the impact of fatigue on performance, would shipboard trainers be able to implement these components into their training programs?
2. Shorter time-on-task is associated with greater task performance likely due to greater engagement. However, what is the minimum and maximum time necessary for optimal training while reducing fatigue and eye strain?
3. Are there short- and long-term effects on the visual system due to repeated exposure?
4. Filters, image adjustments, and calibrations have been proposed as methods to reduce visual discomfort during VR/AR exposure. Can such measures, individually and combined, effectively reduce such symptoms?

Takeaways

- What aspects of VR/AR technology make it prone to producing fatigue?

  - The level of human-machine interaction of the system appears to be of importance when examining the impact of both cognitive and physical fatigue on performance. Greater mental overload, decreased arousal due to habituation of a repetitive task, and sleep loss all influence levels of fatigue.
- What techniques can diminish fatigue attributed to VR/AR technology?
  - Although in its infancy, VR/AR research related to fatigue suggests that shorter duration tasks and greater interaction within the VR/AR program appears to result in greater performance, likely due to greater engagement and motivation of the task. Physical fatigue may further be diminished by avoiding testing and/or training during circadian troughs. Still, this approach would represent the classic approach to reducing fatigue, which simply requires less prolonged work or exposure.
- What aspects of VR/AR technology make it prone to producing eyestrain and visual fatigue?
  - Research has suggested that imperfect binocular disparity could explain a significant amount of eyestrain complaints. This problem occurs when there is a perceivable difference between right and left eye images.
- What techniques can diminish eyestrain and visual fatigue attributed to VR/AR technology?
  - Within VR/AR technology, it is suggested that image manipulations can diminish imperfect binocular disparity related to visual fatigue. These manipulations can include filters, blurs, image adjustments, and calibrations, although validation is needed for these procedures.
- Are the costs and benefits of VR/AR technology similar for fatigue and eye strain?

  - Fatigue research indicates greater interaction within VR/AR systems is associated with greater performance due to decreased fatigue effects. Greater interaction may translate to greater feelings of immersion, which in turn may prove detrimental with regard to eye strain. As such, there may be trade-offs for which any determination of the most effective shipboard training system must account.

Suggested Further Reading

- Baumeister, J., Dorrian, J., Banks, S., Chatburn, A., Smith, R. T., Carskadon, M. A., et al. (2016). Augmented reality as a countermeasure for sleep deprivation. *IEEE Transactions on Visualization and Computer Graphics, 22*(4), 1396-1405.
- Dan, A., & Reiner, M. (2016). EEG-based cognitive load of processing event sin 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology*, *INTPSY-11157*, 1-10.
- Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays, 25*, 99-108.
- Reedy, G. B. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing, 11*, 355-360.
- Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, *29*, 58-69.
- Tomasko, J. M., Pauli, E. M., Kunselman, A. R., & Haluck, R. S. (2012). Sleep deprivation increases cognitive workload during simulated surgical tasks. *The American Journal of Surgery*, *203*, 37-43.

- van Krevelen, D. W., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *The International Journal of Virtual Reality, 9*, 1-20.

Vergence-Accommodation Conflict

Summary

Some training problems will occur in almost any scenario with or without VR/AR technology, whereas other problems are directly attributable to the VR/AR system itself. Vergence-accommodation conflict is a particular type of problem almost unique to virtual environments. In normal binocular vision, the eyes work together to achieve accurate depth perception by using both vergence and accommodation. Vergence describes the process by which the eyes move separately to aim at the same location. For near objects, the eyes must converge upon a nearby point as the left eye moves right and the right eye moves left. For far objects, the eyes must diverge upon a distant point as the left eye moves left and the right eye moves right. These general eye movements work in conjunction with accommodation under normal circumstances. Accommodation describes the process by which ciliary muscles contract or relax to round or lengthen the lens within the eye. A more rounded lens allows for focus upon nearby objects, whereas a less rounded allows for focus of more distant objects. These two processes are naturally coupled and work in unison during normal vision.

Vergence-accommodation conflict occurs when an observer is trying to view a simulated depth upon a nearby screen. An image may attempt to recreate a three-dimensional display with depth, and the eyes will attempt to converge or diverge as appropriate for said depth. However, the lenses are constantly trying to accommodate the nearby screen whether the eyes or converging or diverging, which means the lenses and accommodative processes are no longer functioning in concert with the vergence process. This conflict can cause symptoms of visual fatigue and visual discomfort. Visual fatigue represents an objective decline in human visual operations, whereas visual discomfort represents the subjective experience of unpleasantness or pain felt by the individual following prolonged vergence-accommodation conflict issues.

There are currently mixed options to solve the problem with different emphasis on more short-term or long-term solutions. Short-term solutions involve changes to the scenario design based upon the existing technology and vision science literature. For example, conflict symptoms can be reduced by increasing the distance between the eyes and the screen when possible, or by limiting any rapid depth perception changes in the virtual scenario. These solutions are design-based and therefore immediately applicable to current systems. The ultimate advantage is reduced symptomology, which allows for longer training session durations.

Long-term solutions involve changes to the existing technology. The current systems depend largely upon stereoscopic viewing displays to simulate depth, although the technology may prove to be a dead end at solving vergence-accommodation conflict. Symptoms can be minimized through countermeasures (e.g., pharmacological) and image resolution enhanced; even so, it remains unclear whether stereoscopic images will ever fully eliminate vergence-accommodation conflict because they still require a screen placed too close to the eyes. New technology offers promising solutions and potential elimination of vergence-accommodation conflict altogether in virtual scenarios. Several such technologies are discussed, including parallax barriers, pinlight arrays, and light fields. However, an important concern becomes whether the Navy should directly invest in these technologies. Commercial investment into new products and technological advancement already amounts to billions within the past few years. This existing investment already will pursue any viable solutions without further naval investment. The Department of the

Navy could benefit greatly by adapting from the commercial advancements rather than investing further billions into solving the same problem. As such, it is likely within the Navy's best interest to utilize available resources in exploring and adapting off-the-shelf technology for naval training purposes. This approach would require, at most, minor or targeted investments through Small Business Innovation Research (SBIR) or Small Business Technology Transfer (STTR).

Takeaways

- Should conflict solutions be operator-based or technology-based?
    - Vergence-accommodation conflict occurs due to typical functioning of the human visual system when a virtual element is incorporated to replace real-world elements. Despite some interesting vision training options to enhance vergence or visual processing, the conflict solutions will need to come from new technology.
- How severe are vergence-accommodation conflict symptoms?
    - The primary symptoms are visual fatigue and visual discomfort—the objective reduction in visual system performance, and the subjective experience, respectfully. The symptoms accumulate with prolonged exposure, and this accumulation can eventually result in fusion difficulties. At that point, training scenarios should cease as any further training benefits would be severely limited by the aviator's inability to properly perceive visual information.
- Does the conflict equally impact virtual reality and augmented reality systems?
    - It is likely that there will be fewer vergence-accommodation issues with AR systems than with VR systems. This suspected difference is due to an AR system incorporating real-world elements, which would not be prone to the vergence-

    accommodation problems. However, an AR system could continue to cause this issue for any computer-generated images. This issue should be explored further.

- What solutions appear to be the most promising?
    - There are numerous promising technologies that could address vergence-accommodation conflict issues. Short-term solutions can minimize the symptomology through optimal design, but new technology innovations (e.g., parallax barriers, pinlight arrays, light fields) could eventually eliminate any symptoms. Unfortunately, it is difficult to lay out a true time table for the development of these new technologies, thereby making design solutions the more immediate option.
- Should the Navy try to reduce conflict symptoms through VR/AR development?
    - There are already substantial commercial investments aimed at improving VR/AR technology and solving issues with vergence-accommodation conflict. Unless the Navy is willing to contribute billions to the research, available funding could be better used exploring and adapting off-the-shelf technology for naval purposes.

Suggested Further Reading

- Banks, M. S., Kim, J., & Shibata, T. (2013, May). Insight into vergence/accommodation mismatch. In *SPIE Defense, Security, and Sensing* (pp. 873509-873509). International Society for Optics and Photonics.
- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, *8*(3), 33: 1-30.

- Kim, J., Kane, D., & Banks, M. S. (2014). The rate of change of vergence–accommodation conflict affects visual discomfort. *Vision research*, *105*, 159-165.
- Kramida, G. (2016). Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE transactions on visualization and computer graphics*, *22*(7), 1912-1931.
- Lambooij, M., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, *53*(3), 30201-1.
- Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011, February). Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict. In *IS&T/SPIE Electronic Imaging* (pp. 78630P-78630P). International Society for Optics and Photonics.
- Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, *29*(2), 106-116.

The Persistent Issue of Simulator Sickness in Naval Aviation Training:

History, Known Challenges, and Future Directions

by

Daniel J. Geyer & LT Adam T. Biggs, MSC, USN

Introduction

Modern virtual reality (VR) and augmented reality (AR) systems have nearly unlimited potential for naval aviation training. The newer head- and helmet-mounted displays (HMDs) are lightweight, portable, affordable, and capable of producing high fidelity displays previously available only through massive screen-and-projector setups. These technologies could allow trainees to simulate any number of scenarios they might encounter in the air—only, they would face these encounters on the ground, in complete safety, at minimal cost, and perhaps most important of all, these trainers could be deployed anywhere at any time. One important complication and potential drawback involves the role of simulator sickness and its resultant symptoms, including dizziness, pallor, cold sweating, increased salivation, stomach awareness, headache, fatigue, apathy, nausea, and the traditional end point of vomiting. These substantial and varied symptoms might nullify any benefits of exposure to this environment under normal conditions, and the symptoms might be further complicated by operational requirements such as integrating these virtual systems aboard an aircraft carrier. Given the immense potential and aeromedical risks, it is important to fully understand the advantages and disadvantages of the technology before the Navy decides to fully pursue these capabilities with significant financial investment.

Our goal here is to discuss simulator sickness as it pertains to naval aviation training. We will cover several topics, including: the prevailing theories on why symptoms develop, methods of measurement, incidence rates, contributing factors, effects on training, effects when used shipboard, aftereffects, and countermeasures. Based on the review of this information, we will conclude with a discussion about the most and least promising avenues for future research involving virtual simulations in an aviation training environment.

Definitions and Distinctions

In the approximately fifty years of research literature for motion sickness, terms are routinely used interchangeably, often without explicit definition. For the purpose of this paper, the term “motion sickness” will be used in the generic sense, to indicate the well-recognized symptoms occurring when a susceptible individual is exposed to provocative motion. Additionally, the term will serve as an umbrella category, covering all subcategories in the abstract. Visually-induced motion sickness will refer specifically to motion sickness derived from exposure to a provocative environment lacking physical motion (e.g., optokinetic rotating drums, video games, Oculus Rift). The terms simulator sickness and cybersickness are often used interchangeably in the academic literature. There have been sporadic efforts to quantitatively differentiate the two as separate subcategories of motion sickness, with overlapping but distinct symptomology; these attempts will be addressed in the section on measurement (Stanney, Kennedy, & Drexler, 1997). Outside of academic research, when used by commercial or media outlets, the term ‘cybersickness’ is typically referencing only the classical endpoint of nausea and vomiting induced by exposure to a computer-generated image; other symptoms are usually ignored. As this paper’s focus is on military flight simulators, it will retain the term simulator sickness to reference all motion sickness generated by a virtual environment (VE).

VE refers to any computer-generated environment – or aspects thereof– with which a user can explore and interact. VR is an entirely artificial VE within which the user is fully immersed. Immersion, referenced as presence in the literature, is the perception of a user being physically present in a VR. The less physical reality that intrudes into the VR, the greater the sense of immersion. AR, in contrast to VR, introduces synthetic elements to a user’s visual display while allowing continuous visualization of the real environment; it creates a VE “mixed” with reality. High immersion in AR would constitute a seamless mix of the virtual with the real. HMD AR can be implemented as either an optical- or video-based display. Optical-based AR, as with the Google Glass or F-35 helmet, superimposes computer-generated objects on a user’s real-world view. Video-based AR uses external cameras to relay the real-world environment to the user’s display and combined with computer-generated images. Theoretically, simulator sickness is a lesser concern with optical-based AR, in contrast to video-based, as the latter requires higher resolution and processing powers in addition to challenges arising from improper placement of the external cameras. From a technological perspective, video-based AR is identical to VR.

There are two primary simulator technologies used in military aviation training: HMDs and the cave automatic virtual environments (hereafter referred to as CAVE). Though configurations will vary, CAVEs are essentially large rooms with projectors directed to display 3D images across most, if not all, surfaces, and a motion capture system recording the real-time position of the user. For the purpose of this review, military flight simulators using screens and projectors for an out-the-window (OTW) display will be viewed as essentially CAVE constructs outfitted with high systems fidelity cockpits. Examples include CAE USA’s six degree-of-freedom MH-60R and L-3 Link’s F/A-18E/F tactical operational flight trainers (TOFT). Due to their lack of affordability and transportability – and that only a handful of research universities are equipped with CAVEs – academic research on simulator sickness has focused almost exclusively on HMDs. Military research has been broader by design.

Theories on Motion Sickness

Researchers have identified multiple contributing factors to the development of motion and simulator sickness, yet there is no singular identifiable explanation for its existence or the pathway of symptom development. The most widely accepted theory is the Sensory Conflict Theory (SCT). This theory proposes that when motion detected by the vestibular, visual, and proprioceptor systems conflicts with expected (or previously learned) motion, a centralized signal gradually builds until reaching an individually variant threshold. If this threshold is surpassed, symptoms endemic to motion sickness begin to occur (Reason and Brand, 1975). These sensory conflicts can be either inter-sensory (e.g., between the visual and vestibular system) or intra-sensory, within the vestibular system between angular (semicircular canals) and linear motion (otoliths) detection (Bertolini & Straumann, 2016). A modification of the SCT hypothesizes that the only sensory conflict necessary is between the expected (learned) and sensed direction of gravity, referenced as the subjective vertical (Bles, Bos, de Graaf, Groen, & Wertheim, 1998; Bos & Bles, 1998). However, the SCT does not predict motion sickness or explain individual susceptibility.

The prime alternative to SCT is the postural instability theory (PIT). Rooted in ecological theory, Riccio and Stoffregen (1991) noted that sensory information is conflicting even in everyday tasks, and they theorized motion sickness develops when an individual is placed in a novel environment in which the correct balance – or posture – had not yet been learned. Therefore, postural instability should always precede motion sickness (Stoffregen & Smart, 1998). Other theories have been proposed, though none have received the attention of SCT and PIT, and these alternatives are intended to explain the ‘why’ rather than the ‘how’ (cf., Treisman’s Evolution Theory, 1977). Unless otherwise stated, this paper follows the trend of most research and presumes SCT. Using the SCT paradigm, simulator sickness can be differentiated from classical motion sickness as being caused by the inability to accurately replicate motion with which the individual is accustomed, rather than provocative motion in and of itself (Kennedy et al., 1988; Kolasinski, 1995; Pausch, Crea, & Conway, 1992). A prime example is when experienced pilots trend higher in simulator sickness incidence during flight simulator training than student pilots—the latter not yet having “learned” the real motion of aircraft (Kennedy et al., 1988).

Physiologically, research has yet to identify a clear relation between the level of sensory conflict and/or postural instability and the time course, incidence, or symptomology of motion sickness (Warwick-Evans, Symons, Fitch, & Burrows, 1998). A functioning vestibular system is a prerequisite of all motion sickness. Previous evidence has indicated that persons with bilateral labyrinthectomies have proven immune, whereas blind persons remain susceptible (Cheung, Howard, & Money, 1991; Reason, 1978). Any illusory sense of motion when none occurs is vection, and this illusory motion is a natural outcome of achieving a high degree of realism in simulators. Vection is not analogous to motion sickness, but it is a prerequisite as sickness does not develop without it (Flanagan, May, & Dobie, 2002; Hettinger, Berbaum, Kennedy, Dunlap, & Nolan, 1990; Keshavarz, Riecke, Hettinger, & Campos, 2015; Stern, Hu, Vasey, & Koch, 1989).

Measurement

To date, no validated objective measurement of either motion sickness or simulator sickness has been identified. In research studies, physiological variables are routinely measured pre- and post-exposure, but the magnitude and direction of change remain inconsistent between studies, regardless of physiological indices used. Postural equilibrium tests (PETs) provide a measure of ataxia (e.g., loss of muscle control and/or balance) and/or sway, but not all persons experiencing simulator sickness develop measureable postural instabilities. Dennison and

D’Zmura (2017) exposed 15 subjects to a rotating virtual tunnel in an HMD at six different speeds in both sitting and standing conditions. The authors concluded that the lack of correlation between postural sway and simulator sickness excluded it as an objective measurement of sickness levels. Lubeck, Bos and Stins (2015) exposed 15 stationary subjects to both moving and still images in separate sessions on a projection screen. Simulator sickness occurred only in the motion condition, whereas both conditions saw increased postural instability as measured by increased sway.

This inability to identify an objectively valid measurement has led to a reliance on self-reported symptoms collected via questionnaires. For classical motion sickness, the Pensacola Diagnostic Index (PDI) [Graybiel, Wood, Miller, & Cramer, 1968] and the Pensacola Motion Sickness Questionnaire (MSQ) [Kellogg, Kennedy, & Graybiel, 1965; Kennedy, Tolhurst, & Graybiel, 1965] are the predominate favorites. However, these questionnaires are designed to assess classical motion sickness symptomology as a univariate symptom, existing along a single continuum with variance extant only in the level of severity (Gianaros, Muth, Mordkoff, Levine, & Stern, 2001). The Simulator Sickness Questionnaire (SSQ) [Kennedy & Fowlkes, 1992; Kennedy, Lane, Berbaum, & Lilienthal, 1993] was developed from more than 1200 pre- and post-exposure MSQs from 10 different Navy simulators. Any MSQ symptom that did not reach statistical significance in the simulator survey was eliminated, resulting in the SSQ containing 16 symptoms to be scored on a four point scale (0-3). The symptoms are divided into three subscales: oculomotor (i.e., eye strain, difficulty focusing, blurred vision, headache), nausea (i.e., stomach awareness, nausea, salivation, burping), and disorientation (i.e., dizziness, vertigo). Of all the symptoms collected from the Navy simulators MSQs, oculomotor symptoms trended highest and nausea lowest. Vomiting was not included in the SSQ as it occurred in less than two percent of all cases (Johnson, 2005; Kennedy et al., 1988; Kennedy et al., 1992; Kennedy et al., 1993; Lilienthal, Kennedy, Berbaum, Dunlap, & Mulligan, 1987; Muth & Lawson, 2003; Uliano, Lambert, Kennedy, & Sheppard, 1986). Each SSQ subcategory is weighted and summed together for a maximum total score of approximately 300. Simulators are defined as “problem” simulators once central tendency scores reach 20 or higher. While not included in the SSQ, additional research found increased postural instability positively correlated with the disorientation subscale (Kennedy, Berbaum, & Lilienthal, 1997). Throughout this paper, all sickness score references have been derived from the SSQ.

Kennedy, Drexler, Compton, Stanney and Harm (1997) noted that in a review of simulation technology, CAVE flight simulators – both rotary- and fixed-wing – tended to rate from 8 to 20, with most scoring under 10, whereas HMD total scores ranged from 19 to 55. These differences led Stanney et al. (1997) to argue for cybersickness as an entity distinct from simulator sickness, noting that, compared to the CAVE flight simulators, sickness arising from HMDs trended higher on disorientation, lower on oculomotor, and had average total scores three times higher.[1] In 2005, Ames et al. proposed an alternative to the SSQ: the Virtual Reality Sickness Questionnaire (VRSQ). The VSRQ was designed to be taken in under one minute, contained only 13 symptoms, and removed symptoms related to nausea and stomach awareness. In contrast to the approximately 1200 questionnaire results used to validate the SSQ during design, the VRSQ was internally validated using a sample of 16 immobile college students wearing an HMD while sitting or standing. It has yet to see widespread usage in VE research, but is not without its supporters (Buker,

---

[1] However, we should note that this conclusion is based upon the state of technology twenty years ago. A theoretical difference between cybersickness and simulator sickness would need to be replicated with current technology standards before fully accepting the two as theoretically distinct entities.

Vincenzi, & Deaton, 2016). However, for military populations using the more dynamic CAVE flight simulators, the SSQ is preferred.

Incidence

Incidence of simulator sickness has always been difficult to determine for several reasons, including: the experiences vary widely between individuals, the tremendous array of technologies being used, the environments with which individuals and technology are combined, and the tasks being performed. For population incidence, two studies are commonly used throughout the literature. Miller and Goodson (1958, 1960) identified 12 percent of flight students and 60 percent of instructors as experiencing some degree of sickness as a result of exposure to a military rotary wing simulator, though their sample consisted of only 28 subjects and pre-dated computer-generated graphics. A larger post-exposure questionnaire study involving approximately 700 U.S. Navy and Army pilots found 45 percent reporting at least one incident of post-exposure effects, though again, the surveyed technology is now more than 30 years old (Baltzley, Kennedy, Berbaum, Lilienthal, & Gower, 1989). Yet these incidence rates continue to be repeated in the academic literature as population totals, likely due to the difficulty in determining incidence otherwise.

As simulator sickness incidence varies so widely between simulators, there is no reliable measurement to encompass all existing simulators. However, aggregate incidence by simulator type can be estimated (Kennedy et al., 1992; Kennedy et al., 1993). Several reviews have identified rotary-wing simulators as the most problematic, with incidence as high as 60% in some simulators (Baltzley et al., 1989; Kennedy, Lim, Berbaum, Baltzely, & McCauley, 1989; Braithwaite & Braithwaite, 1990; Johnson, 2005). Johnson (2005) suggested the higher sickness scores were likely due to increased perception of visual flow from the combination of lower altitude and a more detailed terrain than seen in the higher altitude fixed-wing simulators. Drexler (2006) conducted a meta-analysis with approximately 2100 subjects from 21 different simulator studies and 16 different VR studies, totaling over 4000 SSQs. She identified the types of simulators in order of decreasing sickness: driving, rotary-wing, and fixed-wing. She also noted that in comparing five HMD studies to one CAVE study, total and nausea subscale scores were higher in HMDs than CAVEs, with CAVEs scoring higher on oculomotor symptoms. However, samples used in rotary- and fixed-wing simulator studies are almost exclusively military population, whereas car/driving studies and HMD studies relied almost exclusively on college student populations.

The incidence and duration of negative aftereffects post-simulator use is, despite being of prime importance for military pilots, mostly theoretical. Only two large studies were found that followed post-exposure symptomology beyond an hour. Both relied entirely on retrospective questionnaires and are approximately 30 years old. Baltzley et al.'s (1989) post-exposure survey of 742 pilots using 11 different Navy and Army military flight simulators found 45 percent reported some after-effects. Thirty-four percent said it lasted less than an hour, six percent lasting longer than four hours, and four percent lasting longer than six hours. Four pilots (0.5% of the group) reported spontaneously occurring flashbacks of vection post-exposure. Ungs (1989) surveyed U.S. Coast Guard pilots undergoing training via three rotary- and one fixed-wing simulator. Of the 196 respondents, 4.6% reported adverse symptoms lasting 24 hours or longer. Listed symptoms included visual flashbacks, ataxia, and decreased hand-eye coordination. Three pilots (1.3%) reported subsequent difficulties flying aircraft. However, each of these respondents noted on the survey that they would not hesitate to use the simulators again.

Factors Linked to Simulator Sickness

Research has identified multiple factors that can contribute to simulator sickness in VR and AR. These factors fall into one of three categories: individual, technological, and usability factors (Ames, Wolffsohn, & McBrien, 2005; Kolasinski, 1995).

Individual

Individual susceptibility is often based upon subjective differences in symptomology through prior experience, although some factors are linked with higher rates of motion sickness. Very young children and older adults are relatively immune (Reason et al., 1975; Golding, 2006). Females tend to be more susceptible than males and trend higher in oculomotor and disorientation symptoms (Reason et al., 1975; Stanney & Kennedy, 2008). There is some evidence of a genetic predisposition, especially for those of Chinese ancestry (Finley et al., 2004; Stern, Hu, LeBlanc, & Koch, 1993). However, the most validated individual predictor of developing any variant of motion sickness is a prior history of motion sickness (Braithwaite et al., 1990; Stanney et al., 2008).

Technological

Researchers have found it difficult to isolate the technological aspects responsible for simulator sickness due to the wide array of technologies and devices available. Four technological aspects common to all devices make up the bulk of the research: the visual field-of-view (FOV), the accommodation-vergence conflict, latency, and the frequencies of simulated motion and/or visual displays.

Unobstructed, real-world FOV is 180 degrees or more. When discussing HMDs, two FOVs must be considered: the display FOV (DFOV) and the geometric FOV (GFOV). The DFOV is the

FOV allowed by the physical dimensions of the device, and GFOV the simulated FOV. If DFOV is larger than GFOV, the image must be magnified to fill the physical dimensions, and vice versa. Most HMDs GFOV ranges between 110 and 120 degrees. There is a general consensus that a wide FOV is a prerequisite for simulator sickness because a wider the FOV induces greater vection—another prerequisite of simulator sickness (Biocca, 1992; Keshavarz et al., 2015; Mooij, 1988; Young, 2003). Wider FOVs can also increase the likelihood of flicker perception, where the display's frame rate has fallen below threshold. The eye is then able to distinguish light modulations as the frames change, which can contribute to oculomotor disturbances (Kolanski, 1995). Image scale, which is the ratio of DFOV to GFOV, is another potential contributor to sickness scores. Bos, deVries, van Emmerik, and Groen (2010) found that altering the image scale factor in HMDs in either direction from a 1:1 ratio decreased simulator sickness incidence. However, Moss and Muth (2011) found no effect in HMDs in either direction, using 2.0 and 0.88 ratios, whereas Draper, Viirre, Furness & Gawron (2001) found sickness levels significantly greater in both directions, using 2.0 and 0.5 image scale factors in HMDs.

The accommodation-vergence conflict is a major source of eye strain and fatigue. When viewing an object, the eyes have two primary actions: converging dually upon an object so that both eyes are directed at the point of interest (vergence); and changing the shape of the eyes' lenses to sharpen the retinal image (accommodation, or focal distance). These two actions are coupled as dual parallel feedback control systems, and in natural viewing the distance for both are equal. Yet in stereoscopic AR and VR, they are artificially decoupled, as the stereoscopic display causes the vergence distance to vary depending on image contents (i.e. where a 3D image is projected to be in space), whereas the focal distance remains the same, being the distance between the eyes and the display screen (Lambooij, Ijsselsteijn, Fortuin, & Heyndericks, 2009; Shibata, Kim, Hoffman,

& Banks, 2011). This topic is related to simulator sickness for its problematic oculomotor disturbances, but it is considered a separate entity with an achievable engineering solution.

Latency, or lag, is the time delay between an HMD user's head movement and the display's ability to update. Sixty milliseconds is frequently cited as the maximum allowable to remain undetected by the user, though some researchers have argued for rates below 20 milliseconds (LaValle, Yershova, Katsev, & Antonov, 2014). DiZio & Lackner (1997) and Cobb (1999) found the longer the latency, the higher the simulator sickness scores. In contrast, three separate studies of VR HMDs found no effect of latencies ranging between 40 and 250 milliseconds on sickness scores (Nelson, Roe, Bolia, & Morley, 2001; Draper et al., 2001; Moss et al., 2011). As with most VR studies, this contradiction raises the issue of differences in technology or specific platforms driving the effect. However, in all of the aforementioned studies, latency remained constant throughout the experiment. More recent work has examined the relation between latency and amplitude over time. A VR HMD study found that sinusoidally varying latency amplitude by 20-100 milliseconds at a 0.2 Hz frequency caused significantly higher sickness scores than chronic-rate latency (St. Pierre, Banerjee, Hoover, & Muth, 2015). A second VR HMD study found that varying latency amplitude with a frequency of 0.2 Hz had significantly higher sickness levels than 0.1 Hz (Kinsella, Mattfeld, Muth & Hoover, 2016).

Latency may be more problematic in AR than VR HMDs. In the 1990s, AR HMDs routinely suffered from significant visual delays and the "swimming" movement of virtual objects and text, causing sickness scores to rise (Azuma et al., 1994). Allison, Harris, Jenkin, Jasiobedzka and Zacher (2001) suggested that latency in AR HMDs is more nauseogenic than in VR HMDs, as humans are more sensitive to relative than absolute motion. Buker et al. (2012) suggested that in AR HMDs, there are two levels of sensory conflict due to latency: inter-sensory, between head movements and the AR, and intra-sensory, between the AR and OTW displays. However, we were unable to identify any research explicitly comparing sickness levels between AR and VR HMDs.

There is a general understanding that the development of motion sickness depends on motion frequency. Prior research identified the most problematic frequencies as between 0.1 – 0.3 Hz, with the greatest incidence centered at 0.2 Hz (Golding, Mueller, & Gresty, 2001; Golding & Gresty, 2005; Shupak & Gordon, 2006). The proposed reason behind the nauseogenecity of this frequency is that the otoliths' 'break' frequency between tilt and oscillation perception is approximately 0.2 Hz, causing vestibular indecision regarding movement – an inter-sensory conflict per the SCT (Shupak et al., 2006). This would explain the increased sickness levels in St. Pierre et al.'s (2015) study with a latency amplitude at 0.2 Hz compared to the 0.1 Hz condition. In Kennedy et al.'s (1997) technical report to NASA regarding simulator sickness, a prime recommendation was to ensure any simulator motion frequency occurred below 0.01 Hz or above 0.8 Hz. There is evidence that this frequency range is problematic not just as a direct motion frequency, but in *differences* between motion frequencies. Groen and Bos (2008) looked at the frequency of the mismatch signal between simulator platform motion and the motion of an actual vehicle. The simulator was a full-sized car on a six DOF platform, with the OTW display generated by screen and projectors. They analyzed 58 SSQs from two different experiments, and found significantly higher sickness scores in the experiment with the main frequency component of the mismatch signal at 0.08 Hz, compared to the other experiment's main frequency at 0.46 Hz.

Usability

Usability factors correlated with increased sickness scores include sense of presence, sense of vection, peripheral vision, length of time in the simulator, and the user's actions within the simulator (Stanney et al., 1997). As with technological factors, these factors are inextricably linked, and likely work in tandem for an additive effect of increased sickness scores.

It was previously believed that as a simulator's fidelity increased to the point of accurately replicating reality, sickness levels would drop (Witmer & Singer, 1998). More precisely, the greater the presence, the lower the potential for sickness. While this appears to be true for CAVEs, heightened visual detail in HMDs has been proving more nauseogenic (Stanney et al., 2008). In an experiment with the Oculus Rift, seated subjects were shown two different 14 minute roller coaster scenarios with either high or low fidelity. FOV, latency, and all other factors were identical between scenarios. Researchers found that the high fidelity track not only had significantly higher sickness scores, but that significantly more users were unable to finish the simulation compared to the low fidelity track (Davis, Nesbitt, & Nalivaiko, 2015; Nalivaiko, Davis, Blackmore, Vakulin, & Nesbitt, 2015). An interesting issue could then arise if there are differences in fidelity between certain factors. For example, high fidelity in resolution, but low fidelity in tracker latency. This fidelity mismatch scenario could induce even higher simulator sickness levels, albeit that interpretation is speculative and requires further research.

As aforementioned, research suggests that vection is a prerequisite for simulator and cybersickness, as previous research has shown that persons unable to sense vection are unlikely to develop symptoms (Hettinger et al., 1990). Bonato, Bubka, Palmisano, Phillip & Moreno (2008) found that not all vection experiences are similar, and that subjects viewing an optic flow pattern on a computer screen were significantly more likely to develop simulator sickness when the pattern, and hence the direction of vection, changed, as compared to a steady-state flow.

Physiologically, vection – and, by extrapolation, simulator sickness – is dependent on peripheral vision (Webb & Griffin, 2003). Moss et al. (2011) found that after adjusting for latency and image scale factor, simulator sickness was highest when peripheral vision was wholly ensconced by VR, suggesting that allowing some aspect of peripheral vision to remain open to the external environment may reduce sickness levels.

The length of time in the simulator has been correlated with increased sickness in several studies (Johnson, 2005; Kolanski, 1995; Stanney, Lanham, Kennedy, & Breaux, 2000). However, time to the development of sickness is invariably linked with type of simulator, type of simulation, tasks being performed, etc., preventing the identification of a uniform time course to sickness. Commercial HMD manufacturers, such as Oculus Rift and HTC Vive recommend users should take short breaks after every 30 minutes of use. Johnson’s 2005 review of rotary-wing military flight simulators suggested a maximum time of two hours in-simulator.

Actions and tasks performed within a simulator are, as are all usability factors, inextricably linked with myriad other factors. Increased head movements have been linked to increased sickness scores in several studies of HMDs (Kolanski, 1995; Moss & Muth, 2008). However, as with the length of time in simulator, head movements are likely intertwined to other potential causes. Researchers have noted that as sickness increases, users will minimize head movements, adopting a “move-and-wait” strategy (Buker et al., 2012; Draper et al., 2001), though whether this is due to head movements being the nauseogenic cause, or another property such as latency, remains to be determined (Walker, Muth, Switzer, & Hoover, 2010). There is also evidence that sickness increases when users are passive observers compared to when users are allowed to control movement in the VE. Jaeger and Mourant (2001) had subjects traverse a virtual hallway while wearing an HMD using either a mouse while sitting or via a treadmill-operated system, with sickness levels significantly higher in the passive condition.

Simulator Sickness and Training

Kennedy et al.’s 1987 report on simulator sickness among U.S. Navy simulators recommended the development of a simulation sickness program to ensure simulator-enhanced training is not compromised. The report cited two rationales: first, that simulators with a reputation for inducing sickness will be avoided and rendered useless as users will lack confidence in the simulator’s training abilities; and second, sickness-inducing simulators could lead users to avoid actions or movements that would enhance symptoms, such as pilots avoiding looking at the OTW display and relying as much as possible on mock-up instruments. Johnson (2005) notes that at the time of his review of military rotary- and fixed-wing simulators, there was virtually no evidence showing simulator sickness prevented adequate training. There have been cases of users avoiding – or at least complaining of – specific simulators due to reputation, in both civilian and military settings, though there lacks any concrete evidence of poor reputation affecting training (Hicks & Durbin, 2011). Nor is there any specific evidence that users adopt anti-nauseogenic methods in simulators that preclude acquiring the necessary training.

There is evidence that simulator sickness negatively effects performance within simulators. One study was found where simulator sickness was negatively correlated with accuracy in a target and shooting task, though the experiment did not involve flight simulators. Jerome (2007) had 64 subjects participate in a tracking-and-shooting simulation in an approximately 30 square meter urban physical mock-up. All subjects wore haptic vests and optical-based AR HMDs. Audio was presented via multiple speakers throughout the mock-up. HMDs, vests, and speakers were coordinated in real time, presenting subjects with multiple computer-generated enemies in a multi-modal VE. Simulator sickness was negatively correlated with accuracy and positively correlated with reaction time, indicating that as symptomology increased, subjects became less accurate and slower to acquire targets.

There is also evidence that not all technology is equal regarding the intersection between training and simulator sickness. Draper, Ruff, Fontejon and Napier (2002) had 10 military personnel participate in a large-area search task using either an HMD or stationary CRT display. Subjects participated in both conditions. Subjects acquired as many targets as possible using either a joystick (CRT) or by line-of-sight (HMD). Each session lasted two minutes. Sickness scores were significantly higher in the HMD condition, whereas situational awareness and accuracy were significantly higher in the CRT condition. Morphew, Shively and Casey (2004) conducted a similar experiment with eight subjects, substituting the CRT display for a conventional computer monitor. Subjects performed an Unmanned Aerial Vehicle (UAV) sensor operator target search task, again via joystick or direct line-of-sight. Subjects participated in both conditions. Again, sickness scores were significantly higher in the HMD condition, whereas targeting accuracy was significantly higher in the computer monitor condition. Taylor and Barnett (2013) examined the training effectiveness of an HMD in two separate experiments. In the first experiment, 98 subjects were assigned to one of three conditions: desktop computer, HMD, or interactive videos. In the desktop computer and HMD conditions, a trainer explained and demonstrated each procedural task in the VE, and provided feedback to subject participation, whereas the interactive video group viewed three different videos used by the U.S. Army to assist in training. Training incorporated functions related to standard military tasks, including movement, observation, target engagement, and communication. Each condition lasted approximately 20 minutes. In the second experiment, 62 subjects were split between three conditions: desktop computer, HMD, or live training. All subjects were then trained in the Army’s hostage rescue missions. Once training was complete, subjects completed four live missions under the same conditions as the live training group, but without instructions or assistance from researchers. In both experiments, there was no significant difference in training retention or training transfer between conditions. However, the HMD conditions elicited significant levels of simulator sickness, and these levels were obtained within 20 minutes. It is possible that HMD-based training would be more effective if the potential for developing simulator sickness was completely removed.

Shipboard Simulator Training

Existing evidence is limited regarding any shipboard use of simulators, and further research can and should examine the efficacy of simulators when considered for shipboard use. While no studies were found that specifically looked at training acquisition using different simulator technologies shipboard, two studies did examine flight simulator usage and sickness levels aboard U.S. Navy Yard Patrol (YP) craft. Muth et al. (2003) looked at operating a fixed-base flight simulator aboard a U.S. Navy Yard Patrol (YP) craft. Twenty-six active duty participants were exposed to three conditions over three separate days: piloting the flight simulator ashore, riding aboard the YP craft without using the simulator; and piloting the simulator aboard the YP craft. Each stimulus lasted one hour. Content was displayed on three 21-inch monitors in front of a mock-up cockpit. SSQ scores in all three conditions were minimal, with no significant differences between conditions. There was, however, a small but significant decrease in dynamic visual acuity among participants after using the simulator aboard the YP craft, approximated as equivalent to one half a line on a standardized eye chart. Though there was no time-course follow up, the authors did note this visual detriment could be the beginning of simulator sickness development, and cautioned against lengthier simulator exposures.

In a follow-up study by Patterson and Muth (2010), the fixed-base simulator was replaced with an HMD. Nine subjects used stick-and-throttle controls to navigate a one-hour flight simulation, with flight instruments digitally overlayed on a virtual heads-up display (HUD) in the HMD. Subjects were exposed to the same three conditions as Muth et al. (2003). Simulator sickness levels were five times higher post-exposure for HMD usage on both land and YP craft than pre-exposure levels, though there was no significant difference between conditions. These results suggest that shipboard usage does not have an additive effect on sickness levels. Given the importance of this issue for VR/AR use in naval training, additional research is critical.

Aftereffects

There is no direct evidence for any simulator sickness aftereffects and actual aircraft accidents. Potentially negative aftereffects were first described anecdotally. Miller and Goodson's 1958 and 1960 Naval Air Station Pensacola reports on simulator sickness described the first significant adverse reaction post-exposure: "One of these men had been so badly disoriented in the simulator that he was later forced to stop his car, get out, and walk around in order to regain his bearing enough to continue driving" (Quoted in Johnson, 2005). Despite the incident occurring prior to the advent of computers and having no recorded second occurrence, the anecdote has become a proof by repeated assertion, and it is often referenced as the cumulative aftereffect in the literature.

Research does suggest the existence of aftereffects that would be problematic during flight. Generally, these include post-exposure ataxia and disorientation (Kennedy et al., 1987; Kolanski, 1995; Riccio et al., 1991), disorienting flashbacks (Ungs, 1989; Kolanski, 1995), and adaptation to behaviors and stimuli that, when transferred to the aircraft being simulated, would be detrimental (Kolasinki, 1995; Nelson et al., 2000).

Increased ataxia does occur for some persons post-exposure, but duration and severity have been difficult to determine, likely due to the difference in technologies and tasks being performed. Most studies measure ataxia only pre- and post-immersion into a VE, and many are explicitly invested more in proving the PIT tenant that increased postural instability is directly linked to subsequent sickness scores rather than identifying any ataxia-related after effects (Stoffregen, Hettinger, Haas, Roe, & Smart, 2000). Cobbs (1999) and Cobbs and Nichols (1998) had 40 subjects play an interactive VR game via HMD while standing immobile. Navigation and shooting tasks were controlled via a hand-held input device. Each exposure lasted a maximum of 20 minutes. A mild increase in ataxia was noted post-exposure, though it was brief and not correlated to sickness scores. Murata (2004) had eight subjects play a VR game via HMD while immobile for three hours, with posture stability measured hourly. While instability was higher post-exposure, there was no correlation between increased postural instability and time in the simulator. A much larger study, however, with more than 900 subjects exposed to VR HMD from anywhere between 15 and 60 minutes found significant levels of postural instability post-exposure. These levels did not return to pre-exposure levels by 60 minutes post-exposure (Champney et al., 2007).

Evidence of flashbacks is the least documented aftereffect. Johnson's (2005) seminal review on simulator sickness in the military notes the only direct evidence of truly dangerous disorientation post-exposure comes from the afore-mentioned Goodwin-Miller anecdote, which has been repeated throughout the literature. Other than a brief reference to 'visual flashbacks' as having been written on at least one anonymous questionnaire (Ung, 1989), there is no other evidence of such drastic disorientation occurring.

Adaptation is the aftereffect with the highest potential for adverse effects. Johnson (2005) noted that most users will adapt to the simulator, with simulator sickness gradually attenuating after repeated exposures. It is believed that repeated exposure to the sickness-producing experience will modify the centralized storage of previously learned motion, until the new motion is no longer recognized as a sensory conflict. An estimated three to five percent of individuals never adapt, for unknown reasons (Reason et al., 1975). The fear is that the behaviors, postures, and sensory interactions learned by the user to prevent or adapt to simulator sickness will cause adverse aftereffects once the user is placed in real aircraft or other real environments. Essentially, users' nervous system plasticity will allow them to incorporate the sensory-motor cues provided by a VE (Champney et al., 2007). The literature often compares this potential effect to the sickness astronauts experience after returning to earth, having adapted to microgravity, though re-adapting to standard gravity after an extended stay in space is hardly equivalent to occasional one-hour hops in a simulator (Barrett, 2004). Stanney, Mourant, and Kennedy (1998, pg 34) summarized the theoretical problem of adaptation: "…those individuals who exit VE interactions feeling less affected (less ill) may actually be the ones at greatest risk." Cobb (1999) theorized that adaptation to a VE may effect postural instability, whereby users leaving the simulator would suddenly encounter simulator sickness symptoms due to a need for re-adaptation to the real environment; a hypothesis that has yet to be tested beyond the immediate post-exposure effects. There is also the possibility that as eye accommodation is controlled by the autonomic nervous system (ANS), simulator sickness' ANS-driven symptomology and subsequent adaptation could resultantly effect accommodation in heretofore unknown ways (Fowlkes, Kennedy, Hettinger, & Harm, 1993). No research linking oculomotor symptoms and adaptation was found.

Only one large-scale study was found measuring not just negative aftereffects post-exposure, but re-adaptation strategies as well. Champney et al. (2007)'s experiment with over 900 college students found that not only had postural instability not returned to baseline at sixty minutes post-dose, but neither had hand-eye coordination skills. Subjects wore an HMD while seated, the VR consisting of a series of tasks including locomotion, object manipulation, and choice reaction time tasks. Time in the VR ranged between 15 and 60 minutes. Postural stability, hand-eye coordination (via a pointing task), and an SSQ were measured before exposure, immediately post-exposure, and then every fifteen minutes for a maximum of one hour post-exposure.

All subjects were divided into three different re-adaptation groups: no stimulus (or natural decay), where subjects would sit with eyes closed between measurements; vestibular re-adaptation, where subjects walked heel-to-toe along an eight-foot metal rail flat on the floor for five minutes between measurement sessions; and hand-eye re-adaptation, where subjects filled a pegboard with 25 pegs as quickly as they could, one at a time, and then removed the pegs in the same manner, repeating the task for five minutes between measurement sessions. The hand-eye coordination re-adaptation condition showed a significant decrease in pointing errors compared to the other two groups, while natural decay and the vestibular re-adaptation conditions had no effect. However, mean hand-eye coordination and postural stability scores did not return to baseline levels by the final measurement in all three conditions. An interesting follow-up question warranting further research would be the consistency of such results after the same individual received multiple exposures across a series of days or weeks.

Countermeasures

For VR- and AR-induced simulator sickness, adaptation is the single, most effective countermeasure to simulator sickness. However, adaptation can result in an alteration of sensory perception. The negation of any such maladaptive behaviors must also be considered. Champney et al. (2007) identified two routes for re-adaptation: natural decay and active recalibration of the sensory systems. Natural decay is the standard route. All Services ascribe to natural decay when mandating a specific duration of time to occur between simulator exposure and flight. Champney et al. (2007) did find that certain tasks might assist in active recalibration of at least hand-eye coordination skills, though it is not known which tasks, of what duration, and for which simulator would be effective. This uncertainty thus raises the issue of situationally-modulated gains.

Besides adaptation, several engineering remedies have been proven to attenuate sickness scores. For HMD latency problems, predictive compensation has shown some promising results. With predictive compensation, a computer algorithm pre-determines the direction and speed of user head movements and updates the display appropriately, without apparent lag (Buker et al., 2012). Dynamic FOV is another potential remedy. A recent study found that reducing the FOV during real or virtual head movements reduced sickness scores (Fernandes & Feiner, 2016). Researchers had 24 stationary subjects view a VR HMD with joystick controls to rotate the displayed view. During head or display movement, the FOV would begin blacking out at the 120-degree mark and, depending on the speed of the movement, all the way to the 50-degree mark. Remaining still would allow the FOV to return to focus at a rate which was not discernible to most subjects. Sickness levels were significantly lower using this dynamic FOV, though the authors note more research is needed to increase FOV efficiency, and that textures and shapes may also affect sickness levels. Essentially, dynamic FOV is the occlusion of peripheral vision, thereby reducing the sense of vection, and, potentially, sickness levels.

Though pharmaceutical countermeasures, such as scopolamine or diphenhydramine, are used to combat motion sickness, there is significantly little research on their efficacy against simulator sickness. Only one study was found, examining the effect of the antihistamine cinnarizine in preventing simulator sickness in flight simulators – a medication that is not legally available in the U.S. There was no difference in SSQ scores between cinnarizine and placebo (Lucertini, Mirante, Casagrande, Trivelloni, & Lugli, 2007). It should be noted that, depending on the mechanism of action, pharmaceuticals aimed at relieving the nausea-like symptomology (i.e. antihistamines) of classical motion sickness may not resolve the oculomotor or disorientation symptoms typically seen with HMDs.

The commercial VR industry's recent forays into galvanic vestibular stimulation (GVS) present a potential means of bypassing sensory conflict by creating a false sense of acceleration when one does not exist. An electrode is placed on each mastoid, and when a small current is applied, the vestibular nerve is activated. The individual's perception of acceleration in a particular direction can be controlled by the direction of the applied current. A variation of GVS is galvanic cutaneous administration, using the same current application but placing electrodes on the neck, usually on the sternocleidomastoids. A study compared GVS and GCS in a fixed-base driving simulator. Subjects received either GVS, GCS, or no stimulus while driving around curves. Both stimuli conditions had significantly lower sickness scores than the no stimulus condition. In the GVS condition, however, subjects retained realistic driving behaviors while navigating the curves – an effect absent in the GCS condition (Reed-Jones, Reed-Jones, Trick, Toxopeus, & Vallis, 2009). Glavez-Garcia's (2015) study applied GCS to subjects in a fixed-base driving simulator known to induce simulator sickness. Subjects underwent three conditions: no stimulus, GCS delivered at 40 meters from a curve until the curve end, or GCS applied intermittently at either curves or during straight patches. Only subjects experiencing GCS at curves had significantly lower sickness scores than the other two conditions. However, at this time, GVS is not sensitive or specific enough to warrant usage in military flight simulators, as its abilities differ highly between persons, and its ability to artificially create a sense of acceleration is limited to a simultaneous perception of strong roll and slight yaw (Wiedemann, Remlinger, & Bengler, 2014).

Future Research

There are numerous opportunities to improve existing systems and thereby enhance simulators through cost-effectiveness, versatility, realism, and ease of deployment. Incidence and prevalence have, to this point, largely been simulator specific; there has been no systematic attempt to define actual rates of simulator sickness, though this omission is largely due to the continual advance of the technology. This factor, and the array of companies and technologies involved, precludes using a single simulator or simulator-type's incidence as a benchmark measurement across different simulators. Nor is there any direct knowledge of the duration of simulator sickness and the latency of any aftereffects. Research is also needed that will follow the time course of sickness and aftereffects beyond the immediate post-exposure questionnaire. These issues do not necessarily make existing simulators ineffective, but they could limit the effectiveness of any platform. Resolving these issues would therefore increase overall training efficacy.

It remains to be seen if there is an accurate methodology to predict simulator sickness susceptibility, especially for a subpopulation such as naval pilots. Whereas experienced pilots were once known to be more susceptible to simulator sickness than student pilots, there is the question as to whether this remains a relevant statement, as there is now an entire generation of experienced pilots coming-of-age with full exposure to video games, VR and AR; are pilots who have grown up with these technologies less susceptible to sickness or negative aftereffects? And will this new generation of pilots adapt to simulators in the same manner that previous generations did, or will they be able to move between simulation and real aircraft without any change in sensory perception?

Regarding technological causal factors: would allowing HMD users to retain some aspect of peripheral vision with the real environment alleviate simulator sickness, or would this eliminate both vection and a sense of presence? Is there a singular aspect of latency that can be directly linked with simulator sickness? There is evidence that a sickness-inducing frequency range's ability to induce sickness exists not just in a singular frequency, but in the difference between different components (i.e., real- and simulated-motion frequencies). And, is it possible to identify causation between specific technological aspects and specific symptoms (i.e., FOV scale directly affects accommodation?).

Additional research is needed regarding usability aspects, especially in correlating levels of sickness with task. While increased head movements in HMDs is linked with increased sickness levels, are head movements alone the cause or are there additional causal factors? High fidelity in HMDs is linked with increased sickness, but what is the exact relation between fidelity, vection, and sickness scores? And, if a high fidelity HMD is coupled with a moving cockpit base, would sickness scores be reduced?

Questions likewise remain for countermeasures. Though adaptation is recognized as the most efficacious solution to simulator sickness, additional research is needed not only to identify negative aftereffects and their duration (especially in regards to oculomotor and disorientation/vestibular disturbances), but what forms of re-adaptation processes – if any – will be effective in reducing/eliminating aftereffects. And, is it possible that a re-adaptation method could create a new set of maladaptive behaviors that would diminish a pilot's performance in real aircraft?

One alleviation method would involve pharmacological countermeasures, yet the efficacy of different medications remains unknown. Current pharmacological interventions are designed to address the symptoms of motion sickness – primarily the endpoint of nausea and vomiting – but addressing the more subtle symptoms of simulator sickness might require a different pharmacological regimen than addressing the symptoms of motion sickness. There is also the question of whether pharmacological interventions would be considered for non-mission essential simulations unless smaller doses could be used or fewer side effects induced. An in-depth study would be necessary to compare the impact of different pharmacological interventions to determine their efficacy in alleviating simulator sickness rather than motion sickness.

Conclusion

What do we know about VR and AR systems when it comes to simulator sickness? VR systems have been repeatedly investigated, although constantly evolving technology also requires constantly updating research findings to make informed decisions about state-of-the-art research. As such, we know quite a lot about simulator sickness within VR systems while simultaneously acknowledging that previous results will not always extrapolate to new systems as the technology advances. However, any conclusions about AR systems are largely informed by previous research about VR systems. There is so little research on AR that assumed parallel findings between the two platforms is the primary metric by which AR can be evaluated. The assumption is that AR systems will induce equivalent or lesser simulator sickness symptoms than VR due to changes in how the information is conveyed. This assumption is based upon theories largely derived from work with VR systems. Although video-based AR is effectively VR and the assumption likely holds, the assumption may not hold with optical-based AR. This aspect is particularly important when considering that many recent optical-based AR platforms are designed for commercial or home use and therefore are intended for situations which might minimize the potential for simulator sickness; however, military missions will expose individuals and AR systems to situations far different than those experienced during commercial or home use. As such, AR systems require additional research into simulator sickness with a particular emphasis on studies that emphasize the consequential effects of operating AR systems during a military mission. For example, AR systems need to be tested with military-grade flight simulators that can incorporate real motion into the simulation where possible.

Caution must be exerted when HMDs and other devices are marketed as free of "cybersickness." Though the term is often used interchangeably with simulator sickness, one should not assume the terms to be synonymous. Commercially, cybersickness is almost exclusively used to reference nausea and vomiting, the end points of classical motion sickness. However, simulator sickness encompasses an array of symptoms, including multiple disorientation and oculomotor disturbances – symptoms which the civilian HMD user may be able to tolerate, but are unacceptable for highly skilled military pilots flying multi-million dollar aircraft.

Several areas of additional research are warranted. Latency and FOV issues have emerged as the potential leading factors causing sickness, especially in HMDs. The commercial sector's emerging solution is to occlude peripheral vision during rapid head movements, decreasing the sense of vection and, concurrently, preventing simulator sickness. However, this is not an option for military pilots, as they require full FOV at all times.

Although the evidence does not suggest that immediate training performance is impaired due to simulator sickness symptoms, there could be lingering issues about retention quality and motivation. Likewise, there is no known time course for symptoms, let alone aftereffects. As such, there is still a significant advantage in reducing simulator sickness symptoms during training for prone individuals.

References

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001). *Tolerance of temporal delay in virtual environments*. Paper presented at the Proceedings of IEEE Virtual reality Conference 2001, New York, NY.

Ames, S. L., Wolffsohn, J. S., & McBrien, N. A. (2005). The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optometry and Vision Science, 82*(3), 168-176.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34-47.

Baltzley, D. R., Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., & Gower, D. W. (1989). The time course of postflight simulator sickness symptoms. *Aviation, Space, and Environmental Medicine, 60*(11), 1043-1048.

Barrett, J. (2004). *Side effects of virtual environments: a review of the literature* (DSTO Technical Report No. 1419). Edinburgh, Australia: Defence Science and Technology Organisation.

Bertolini, G. & Straumann, D. (2016). Moving in a moving world: A review on vestibular motion sickness. *Frontiers in Neurology*, *7*(14), 1-11.

Biocca, F. (1992). Will simulation sickness slow down the diffusion of virtual environment technology? *Presence, 1*(3), 334-343.

Bles, W., Bos, J. E., de Graaf, B., Groen, E., & Wertheim, A. H. (1998). Motion sickness - only one provocative conflict? *Brain Research Bulletin, 47*(5), 481-487.

Bonato, F., Bubka, A., Palmisano, S., Phillip, D., & Moreno, G. (2008). Vection change exacerbates simulator sickness in virtual environments. *Presence, 17*(3), 283-292.

Bos, J. E., & Bles, W. (1998). Modelling motion sickness and subjective vertical mismatch detailed for vertical motions. *Brain Research Bulletin, 47*(5), 537-542.

Bos, J. E., de Vries, S. C., van Emmerik, M. L., & Groen, E. L. (2010). The effect of internal and external fields of view on visually induced motion sickness. *Applied Ergonomics, 41*, 516-521.

Braithwaite, M. G., & Braithwaite, B. D. (1990). Simulator sickness in an Army simulator. *Journal of the Society of Occupational Medicine, 40*, 105-110.

Buker, T. J., Vincenzi, D. A., & Deaton, J. E. (2016). The effect of apparent latency on simulator sickness while using a based helmet-mounted display: reducing apparent latency with predictive compensation. *Human Factors, 54*(2), 235-249.

Champney, R. K., Stanney, K. M., Hash, P. A. K., Malone, L. C., Kennedy, R. S., & Compton, D. E. (2007). Recovery from virtual environment exposure: expected time course of symptoms and potential readaptation strategies. *Human Factors, 49*(3), 491-506.

Cheung, B. S., Howard, I. P., & Money, K. E. (1991). Visually-induced sickness in normal and bilaterally labyrinthine-defective subjects. *Aviation, Space, and Environmental Medicine, 62*, 527-531.

Cobb, S. V. G. (1999). Measurement of postural stability before and after immersion in a virtual environment. *Applied Ergonomics, 30*(1), 47-57.

Cobb, S. V. G., & Nichols, S. C. (1998). Static posture tests for the assessment of postural instability after virtual environment use. *Brain Research Bulletin, 47*(5), 459-464.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). *Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters*. Paper presented at the Proceedings of the 11th Australasian Conference on Interactive Entertainment, Sydney, Australia.

Dennison, M. S., & D'Zmura, M. (2017). Cybersickness without the wobble: experimental results speak against postural instability theory. *Applied Ergonomics, 58*, 215-223.

DiZio, P., & Lackner, J. (1997). Circumventing side effects of immersive virtual environments. In M. J. Smith, G. Salvendy & R. H. Koubek (Eds.), *Advances in Human Factors/Ergonomics: Design of Computing Systems* (Vol. 21, pp. 893-896). Amsterdam: Elsevier

Draper, M., Ruff, H. A., Fontejon, J. V., & Napier, S. (2002). *The effects of head-coupled control and a head-mounted display (HMD) on large-area search tasks* (Abstract). Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors, 43*, 129-146.

Drexler, J. M. (2006). *Identification of system design features that affect sickness in virtual environments* (Unpublished doctoral dissertation). University of Central Florida, Orlando, Florida.

Fernandes, A. S., & Feiner, S. K. (19-20 March 2016). *Combating VR sickness through subtle dynamic field-of-view modification*. Paper presented at the 2016 IEEE Symposium on 3D User Interfaces (3DUI), Greenville, SC (201-210).

Flanagan, M. B., May, J. G., & Dobie, T. G. (2002). Optokinetic nystagmus, vection, and motion sickness. *Aviation, Space, and Environmental Medicine, 73*, 1067-1073.

Fowlkes, J. E., Kennedy, R. S., Hettinger, L. J., & Harm, D. L. (1993). Changes in the dark focus of accommodation associated with simulator sickness. *Aviation, Space, and Environmental Medicine, 64*(7), 612-618.

Galvez-Garcia, G. (2015). Alleviating simulator sickness with galvanic cutaneous stimulation. *Human Factors, 57*(4), 649-657.

Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine, 72*(2), 115-119.

Golding, J. F. (2006). Motion sickness susceptibility. *Autonomic Neuroscience, 129,* 67-76.

Golding, J. F., Kadzere, P., & Gresty, M. A. (2005). Motion sickness susceptibility fluctuates through menstrual cycle. *Aviation, Space, and Environmental Medicine, 76*, 970-973.

Golding, J. F., Mueller, A. G., & Gresty, M. A. (2001). A motion sickness maximum around the 0.2 Hz frequency range of horizontal translational oscillation. *Aviation, Space, and Environmental Medicine, 72*(3), 188-192.

Graybiel, A., Wood, C. D., Miller, E. F., & Cramer, D. B. (1968). Diagnostic criteria for grading the severity of acute motion sickness. *Aerospace Medicine, 39*, 454-455.

Groen, E. L., & Bos, J. E. (2008). Simulator sickness depends on frequency of the simulator motion mismatch: An observation. *Presence, 17*(6), 584-593.

Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., & Nolan, M. D. (1990). Vection and simulator sickness. *Military Psychology, 2*(3), 171-181.

Hicks, J. S. & Durbin, D. B. (2011). *A summary of simulator sickness ratings for U.S. Army Aviation Engineering simulators* (ARL-TR-5573). U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Jaeger, B. K., & Mourant, R. R. (2001). Comparison of simulator sickness using static and dynamic walking simulators. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 45*(27), 1896-1900.

Jerome, C. J. (2007). *Effects of spatial and non-spatial multi-modal cues on orienting of visual-spatial attention in an augmented environment* (ARI-TR-1215). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Johnson, D. M. (2005). *Introduction to and review of simulator sickness research*. (ARI Report No. 1832). U.S. Army Research Institute Field Unit, Fort Rucker, AL.

Kellogg, R. S., Kennedy, R. S., & Graybiel, A. (1965). Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. *Aerospace Medicine, 36,* 315-318.

Kennedy, R. S., Berbaum, K. S., Allgood, G. O., Lane, N. E., Lilienthal, M. G., & Baltzley, D. R. (1988). Etiological significance of equipment features and pilot history in simulator sickness. In *AGARD, Motion cues in flight simulation and simulator induced sickness* (AGARD Conference Proceedings No. 433, pp. 1.1-1.22). Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development.

Kennedy, R. S., Berbaum, K. S., & Lilienthal, M. G. (1997). Disorientation and postural ataxia following flight simulation. *Aviation, Space, and Environmental Medicine, 68*(1).

Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., Dunlap, W. P., Mulligan, B. E., & Funaro, J. F. (1987). *Guidelines for alleviation of simulator sickness symptomatology* (NAVTRASYSCEN TR-87-007). Naval Air Systems Command, Washington, D.C.

Kennedy, R. S., Drexler, J. M., Stanney, K. M., & Harm, D. L. (1997). Configural scoring of the self-report of symptoms in different motion sickness environments: Normative data and comparison with other scoring systems. *Published in Abstracts of the International Workshop on Motion Sickness: Medical and Human Factors* (pp. 78-82).

Kennedy, R. S., & Fowlkes, J. E. (1992). Simulator sickness is polygenic and polysymptomatic:

Implications for research. *International Journal of Aviation Psychology, 2*(1), 23-38.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1989). Simulator sickness in U.S. Navy flight simulators. *Aviation, Space, and Environmental Medicine, 60*, 10-16.

Kennedy, R. S., Tolhurst, G. C., & Graybiel, A. (1965). *The effects of visual deprivation on adaptation to a rotating environment* (Naval School of Aviation Medicine Tech. Rep. No. 918). Pensacola, FL: Naval School of Aerospace Medicine.

Keshavarz, B., Riecke, B. E., Hettinger, L. J., & Campos, J. L. (2015). Vection and visually induced motion sickness: how are they related? *Frontiers in Psychology, 6*(472).

Kinsella, A., Mattfeld, R., Muth, E., & Hoover, A. (2016). Frequency, not amplitude, of latency affects subjective sickness in a head-mounted display. *Aerospace Medicine and Human Performance, 87*(7), 604-609.

Kolanski, E. M. (1995). *Simulator sickness in virtual environments*. (Army Research Institute Technical Report 1027). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology, 53*(3).

LaValle, S. M., Yershova, A., Katsev, M., & Antonov, M. (2014, May). Head tracking for the Oculus Rift. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 187-194).

Lilienthal, M. G., Kennedy, R. S., Berbaum, K. S., Dunlap, W. P., & Mulligan, B. E. (1987, June 23-26). Vision/motion-induced sickness in Navy flight simulators: Guidelines for its prevention. *Proceedings of the 1987 Image Conference IV*, 275-285.

Lubeck, A. J. A., Bos, J. E., & Stins, J. F. (2015). Motion in images is essential to cause motion sickness symptoms, but not to increase postural sway. *Displays, 38*, 55-61.

Lucertini, M., Mirante, N., Casagrande, M., Trivelloni, P., & Lugli, V. (2007). The effect of cinnarizine and cocculus indicus on simulator sickness. *Physiology & Behavior, 91*, 180-190.

Miller, J. W., & Goodson, J. E. (1958). *A note concerning "motion sickness" in the 2-FH-2 Hover Trainer* Project No. 1 17 01 11, Subtask 3, Report No. 1. Pensacola, FL: Naval School of Aviation Medicine.

Miller, J. W., & Goodson, J. E. (1960). Motion sickness in a helicopter simulator. *Aerospace Medicine, 31*(3), 204-212.

Morphew, M. E., Shively, J. R., & Casey, D. (2004). Helmet-mounted displays for unmanned aerial vehicle control (Abstract). *Proceedings of the SPIE, 5442*, 93-103.

Mooij, H. A. (1988). Technology involved in the simulation of motion cues: The current trend. In AGARD, *Motion cues in flight simulation and simulator induced sickness* (AGARD Conference Proceedings No. 433, pp. 2.1-2.15). Neuilly Sur Seine, France: Advisory Group for Aerospace Research and Development.

Moss, J., Scisco, J., & Muth, E. (2008). Simulator sickness during head mounted display (HMD) of real world video captured scenes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(19), 1631-1634.

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors, 53*(308), 308-319.

Murata, A. (2004). Effects of duration of immersion in a virtual reality environment on postural stability. *International Journal of Human-Computer Interaction, 17*(4), 463-477.

Muth, E. R., & Lawson, B. (2003). Using flight simulators aboard ships: Human side effects of an optimal scenario with smooth seas. *Aviation, Space, and Environmental Medicine, 74*, 497-505.

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., & Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & Behavior, 151*, 583-590.

Nelson, W. T., Roe, M. M., Bolia, R. S., & Morley, R. M. (2000). *Assessing simulator sickness in a based HMD: effects of time delay, time on task, and task complexity*. Paper presented at the IMAGE 2000, Scottsdale, Arizona.

Patterson, F. R., & Muth, E. R. (2010). *Cybersickness onset with reflexive head movements during land and shipboard head-mounted display flight simulation* (NAMRL Report No. 10-43). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Pausch, R., Crea, T., & Conway, M. (1992). A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. *Presence, 1*(3), 344-363.

Reason, J. T. (1978). Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine, 71*, 819-829.

Reason, J. T., & Brand, J. J. (1975). *Motion sickness.* London: Academic.

Reed-Jones, J. G., Reed-Jones, R. J., Trick, L. M., Toxopeus, R., & Vallis, L. A. (2009). Comparing techniques to reduce simulator adaptation syndrome and improve naturalistic behavior during simulated driving. *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design, 276-283*.

Riccio, G. E., & Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural stability. *Ecological Psychology, 3*(3), 195-240.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accomodation conflict. *Proceedings of SPIE, 7863*.

Shupak, A., & Gordon, C. R. (2006). Motion sickness: Advances in pathogenesis, prediction, prevention, and treatment. *Aviation, Space, and Environmental Medicine*, *77*(12), 1213-1223.

St. Pierre, M. E., Banerjee, S., Hoover, A. W., & Muth, E. R. (2015). The effects of 0.2 Hz varying latency with 20-100 ms varying amplitude on simulator sickness in a helmet mounted display. *Displays, 36*(1-8).

Stanney, K. M., & Kennedy, R. S. (2008). Chapter 6: Simulation Sickness. In P. A. Hancock, D. A. Vincenzi, J. A. Wise & M. Mouloua (Eds.), *Human Factors in Simulation and Training*. Boca Raton, FL: CRC Press.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, 2*, 1138-1142.

Stanney, K. M., Lanham, S., Kennedy, R. S., & Breaux, R. (2000). Designing virtual environments to enhance human performance. *Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence, 7*(4), 321-351.

Stern, R. M., Hu, S., LeBlanc, R., & Koch, K. L. (1993). Chinese hypersusceptibility to vection-induced motion sickness. *Aviation, Space, and Environmental Medicine, 64*, 827-830.

Stern, R. M., Hu, S., Vasey, M. W., & Koch, K. L. (1989). Adaptation to vection-induced symptoms of motion sickness. *Aviation, Space, and Environmental Medicine, 60*, 566-572.

Stoffregen, T. A., & Smart Jr., L. J. (1998). Postural instability precedes motion sickness. *Brain Research Bulletin, 47*(5), 437-448.

Stoffregen, T. A., Hettinger, L. J., Haas, M. W., Roe, M. M., & Smart, L. J. (2000). Postural instability and motion sickness in a fixed-base flight simulator. *Human Factors, 42*, 458-469.

Taylor, G. S., & Barnett, J. S. (2013). Evaluation of wearable simulation interface for military training. *Human Factors, 55*(3), 672-690.

Treisman, M. (1977). Motion sickness: an evolutional hypothesis. *Science, 197*, 493-495.

Uliano, K. C., Lambert, E. Y., Kennedy, R. S., & Sheppard, D. J. (1986). *The effects of asynchronous visual delays on simulator flight performance and the development of simulator sickness symptomatology* (NAVTRASYSCEN 85-D-0026-1). Orlando, FL: Naval Training Systems Center, US Navy.

Ungs, T. J. (1989). Simulator induced syndrome: Evidence for long-term aftereffects. *Aviation, Space, and Environmental Medicine, 60*, 252-255.

Walker, A. D., Muth, E. R., Switzer, F. S., & Hoover, A. (2010). Head movements and simulator sickness generated by a virtual environment. *Aviation, Space, and Environmental Medicine, 81*, 929-934.

Warwick-Evans, L. A., Symons, N., Fitch, T., & Burrows, L. (1998). Evaluating sensory conflict and postural instability. Theories of motion sickness. *Brain Research Bulletin, 47*(5), 465-469.

Webb, N. A., & Griffin, M. J. (2003). Eye movement, vection, and motion sickness with foveal and peripheral vision. *Aviation, Space, and Environmental Medicine, 74*, 622-625.

Wiedeman, M., Remlingger, W., & Bengler, K. (2014). Application of galvanic vestibular stimulation for the evaluation of vehicle settings in a fixed-base simulator. In N. Stanton, S. Landry, G. Di Bucchianico & A. Vallicelli (Eds.), *Advances in Human Aspects of Transportation Part II* (pp. 21-28): AHFE Conference 2014.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Young, L. R. (2003). Spatial orientation. In P. S. Tsang & M. A. Vidulich (Eds.)*, Principles and practice of aviation psychology* (pp. 69-113). Mahwah, NJ: Erlbaum.

The Use of Virtual Reality and Augmented Reality Technologies in Spatial Disorientation Training: Current Uses and Future Opportunities

by

F. Eric Robinson & LT Adam T. Biggs, MSC, USN

Overview

Spatial disorientation (SD) remains a stubborn and costly problem despite decades of effort to solve it. SD is the number one human factors cause of mishaps in the Navy (Gibb, Musselman, & Farley, 2012). More troubling than the rate of mishaps is the fact that SD-related mishaps tend to be particularly severe. This single issue is a contributing factor in up to 33% of mishaps and it has a higher fatality rate than other causes (Estrada, Adam, & Leduc, 2002; Gibb, Ercoline, & Scharff, 2011). SD caused approximately 20% of the class A mishaps in the Air Force between 1980 and 2010, resulting in 86 fatalities and over $2 billion worth of damage between 1993 and 2010 (Gibb, Musselman, & Farley, 2012; Wickens, Small, Andre, Bagnall, & Brenaman, 2008).

This problem is defined as a failure to correctly perceive the relative position, motion, or attitude of an aircraft relative to the earth (Cheung, 2013). When a pilot is unaware of the mismatch between perception and reality or is unable to ignore misleading sensory inputs and rely on instruments, SD can lead to a mishap (Heinle & Ercoline, 2003). The problem is also troublingly common. One estimate puts the likelihood of a pilot becoming disoriented between 90-100% at some point in their career (Newman, 2007). The risk is also not isolated to a particular platform. One survey found that 8% of fixed wing pilots experienced SD severe enough to impact flight safety compared to only 6% of pilots who never experienced any form of SD (Matthews, Previc, & Bunting, 2002). Another survey found that over 70% of rotary wing pilots report experiencing SD as well (Estrada, Adam, & Leduc, 2002). Given the high incidence of SD among pilots, it is not surprising that SD contributes to a high percentage of mishaps. The prevalence of SD highlights the severity of the problem, although another troubling factor involves the persistent nature of this challenge. Previous efforts have proven largely ineffective as SD mishap rates remained largely constant over the last 30 years. However, new advances in virtual reality (VR) or augmented reality (AR) might be able to improve training and thereby reduce the human and economic cost of SD.

This report will review issues related to SD training and how VR/AR technologies may be used to enhance such training, including a particular emphasis on the most productive focus of training and the fidelity requirements for effective SD training in simulators. We also discuss the implications of using head-mounted displays (HMDs) in simulations. The report begins with a brief discussion of SD as well as various descriptions of SD types that must be addressed in training. We next describe prior efforts to reduce SD via training, followed by a discussion of the use of VR and the fidelity required to train various types of SD. We conclude with recommendations for future research into SD and training using VR/HMDs.

SD Mechanisms

Sensory, perceptual, and cognitive processes are all involved in orientation (Cheung, 2013). Even so, these varied sensory inputs are only pieces of a whole. The sense of orientation is constructed from the integration, interpretation, and interaction of these multiple systems based on schemas established through training and experience (Cheung, 2013). Discrepancies between these systems can lead to a misperception of an aircraft's position in space. Contributing factors to such discrepancies include attentional anomalies, pilot expectations, and misinterpretation of sensory inputs (Heinle & Ercoline, 2003).

Humans are naturally terrestrial creatures – our sensory systems evolved under ground-level conditions and therefore have certain limitations that can be exceeded during flight (Bos et al., 2003; Kirkham et al., 1978). Most disorientation events are due to inadequate or unreliable sensory information. Visual, vestibular, somatosensory, and cognitive cues all contribute to human spatial orientation, with visual and vestibular cues being the most important (Bos, Bles, Hosman, & Groen, 2003; Eriksson, 2010; Kirkham, Collins, Grape, Simpson, & Wallace, 1978). Pilots rely on stable primary visual cues such as the horizon and moving secondary cues such as the glare shield. The interaction between these cues offers information regarding the attitude and motion of the aircraft (Patterson, Arnold, & Williams, 2013). Visual information often dominates vestibular inputs and removes sensory confusion, allowing the pilot to see which way is up and how the aircraft is moving (Bos et al., 2003). If the visual information is lost, such as during transition to instrument meteorological conditions (IMC), the pilot can have difficulty ignoring inaccurate vestibular information.

The vestibular system responds to both motion and gravity, but algorithms that function on the ground do not necessarily work in the air. At times, we cannot differentiate between motion and gravity in the flight environment, causing confusion or misperceptions of orientation. In the absence of visual cues, such confusion can be considered one of the main causes of SD (Bos et al., 2003). Pilots in instrument meteorological conditions (IMC) have to construct their orientation rather than perceive it, and workload increases as the pilot must rely on instruments to remain oriented (Gibb, Musselman, & Farley, 2012). The misperception/illusion itself is therefore only a small part of an SD mishap scenario. The precursors to the misperception, the environment, and past experience/training play a role as well (Cheung, 2013). Pilot cognition must be considered in addition to the visual or vestibular contributions to SD.

Types of SD

Several unique SD phenomena have been identified. Unfortunately, SD illusions can be difficult to categorize due to lack of agreement on how to define them precisely. Heinle and Ercoline (2003) describe three main categories of SD—identified based on the pilot's response:

1. *Unrecognized* SD occurs when the pilot fails to realize that the perceived attitude is not correct.
2. *Recognized* SD occurs when the pilot is able to recognize a conflict between perceived attitude and actual attitude.
3. *Incapacitating* SD occurs when controls do not appear to respond to the pilot's inputs, despite a functional aircraft.

Although perhaps useful for describing SD in the context of a mishap, this categorization system is somewhat flawed for research or training purposes because the same illusion can fall into multiple categories depending solely on the pilot's awareness of it. Patterson, Arnold, and Williams (2013) describe an alternate categorization system based on the primary cause of various SD illusions:

1. *Vestibular* illusions are caused primarily by conflicting or inaccurate interpretation of input from the vestibular system.
2. *Visual* illusions occur due to optical illusions based on the relative motion of the aircraft and the world, or when pilots inadvertently apply an incorrect spatial strategy (such as

using a coastline as a substitute for the horizon). These issues are most likely to occur when typical visual cues have been lost.

3. *Cognitive* illusions occur when the pilot misinterprets instruments or other cues. These issues are most likely to occur under task saturation or unexpected transitions between visual cues and instrument-based cues.

This categorization system is more amenable to research and training because it maintains a consistent grouping of SD illusions based on common causal mechanisms. The illusions within each category are therefore likely to have similar solutions, facilitating the identification of interventions for any given illusion. Despite similarity within categories, each illusion is likely to have different training and intervention requirements (Cheung, 2013). Even so, a complete understanding of SD can take advantage of both classification approaches by integrating pilot awareness of the situation and the causal mechanism behind the situation. Each aspect delivers something unique for different purposes, and together they provide a comprehensive understanding of SD mishaps.

SD and Virtual Reality

SD countermeasures generally fall into two categories: training and technology (in the form of enhanced displays or warning systems; Braithwaite, Ercoline, & Brown, 2004). There is no "cure" for SD. Training and understanding the nature of SD are the only ways to help mitigate the problem (Kowalczuk & Mikuliszyn, 2002). The general approach of SD training to date has been based on the recognized, unrecognized, and incapacitating taxonomy of SD. The Air Force has stated goals for SD research (along with investigating displays/technologies and causal mechanisms of SD) involving an increase in the rate of recognized as opposed to unrecognized SD under the guidance that pilots can respond to recognized SD (Heinle & Ercoline, 2003). SD training therefore typically aims to increase pilots' awareness of the dangers of SD and the ease of becoming disoriented, as well as increasing the reliance on standard recoveries from disorienting situations (Johnson, Estrada, Braithwaite, & Manning, 1999). This training rests on the assumption that if pilots are exposed to various types of SD, they can recognize and avoid it in flight (Cheung, 2013). The earliest training programs focused on vestibular processes, eventually incorporating vision and an increased emphasis on instrument use as well (Previc & Ercoline, 2004). Four broad principles guide SD training (Braithwaite, Ercoline, & Brown, 2004):

1. Minimize the likelihood of SD using frequent and systematic monitoring of flight instruments.
2. Expect to become disoriented.
3. Recognize SD when it occurs.
4. Try to make the instruments read correctly (i.e., actively control the aircraft and focus on instruments).

These principles are instilled using a combination of classroom instruction and practical demonstrations intended to increase pilots' awareness of SD and help them recognize when it occurs (Johnson, Estrada, Braithwaite, & Manning, 1999). Classroom instruction emphasizes the physiological causes of SD as well as conditions that can lead to SD such as formation flying, low visibility flight, or flying with night vision devices (Braithwaite, Ercoline, & Brown, 2004). This lecture-based instruction helps pilots identify causal or environmental factors upon entering a situation likely to induce SD, thereby raising their situational awareness. Practical demonstrations are intended to illustrate the fallibility of the vestibular system and the ways in which it can be misled. These efforts aim to make pilots rely on their instruments rather than their senses. Demonstration-based instruction most often occurs on the ground in a Barany chair (a device designed to induce vestibular illusions), although in-flight demonstrations are sometimes used (Johnson, Estrada, Braithwaite, & Manning, 1999).

Lecture-based training enhances situational awareness and helps a pilot avoid entering situations that might induce SD, whereas lectures and demonstration-based training each enhance pilot awareness of SD and emphasize the importance of relying on instruments rather than sensory inputs. Tips to avoid SD focus on avoiding the conditions that are likely to lead to specific types of illusions, such as avoiding approaches over dark terrain or not looking away from instruments during IMC flight (Braithwaite, Ercoline, & Brown, 2004). Not only are these tips impractical due to the constraints of flight, the principles behind them are too general to convert into a training syllabus. Military flight often requires flying into difficult conditions, under high workload, while having to simultaneously scan the outside world and instruments. Even routine operations such as carrier landings often require approaches over a dark surface – a direct and unavoidable violation of one tip. Simply telling a pilot what they should do is inadequate. Training must provide pilots with the necessary tools and experience to stay oriented under realistic conditions.

Issues with training. Classroom training alone is inadequate for SD training; demonstrations are the best way to help pilots be aware of the limitations of their sensory systems (Gibb, Schvaneveldt, & Gray, 2008). However, the nature of such demonstrations requires careful consideration. Simple textbook demonstrations are easy to perform on the ground, but difficult to reproduce in an aircraft. Unlike SD on the ground, SD in the aircraft occurs with the pilot's eyes open, unrestricted head movement, and manifests during high workload, poor weather, or otherwise difficult flight conditions (Johnson, Estrada, Braithwaite, & Manning, 1999; Patterson, Arnold, & Williams, 2013). There is no longitudinal evidence to suggest that exposures such as those experienced during SD training prevents SD in flight (Cheung, 2013). Passively experiencing SD illusions in a Barany chair or even in flight is insufficient because sensing and perceiving depend on interaction with the world.

Allowing the trainee to actively interact with the environment is more effective than passive experience such as that gained using a Barany chair (Cheung, 2013; Parmet & Ercoline, 2008; Powell-Dunford, Bushby, & Leland, 2016). Given the difficulty of delivering effective training on the ground, and the need for a more interactive experience to adequately reproduce the conditions of SD in flight, a VR environment holds great promise to train pilots by better replicating the circumstances under which SD tends to occur.

VR and SD training. VR in the form of flight simulators offers several advantages over traditional training. For instance, many types of SD cannot be safely demonstrated in flight, but these experiences can be replicated in a simulator (Estrada, Adam, & Leduc, 2002). Simulators also allow for complete control over workload and weather conditions—all in a safe, reproducible, cost-effective setting (Johnson, Estrada, Braithwaite, & Manning, 1999; Newman, 2016). The earliest simulator is generally considered to be Edward Link's "Pilot Maker" in 1929. The device was capable of mimicking aircraft pitch, roll, yaw, and vibration as well as an instructor's station to monitor instruments and flight progress. The simulator could mimic many actions of the aircraft such as pre-stall buffet, spinning, and overspeed in the undercarriage in addition to being useful for instrument flight training (Newman, 2016). Improvements in simulator technology since the "Pilot Maker" include a larger field of view (FOV), the addition of GPS imagery, the incorporation of G forces, and motion cuing. The planes of motion have also increased from three (pitch, roll, and yaw) to six (pitch, roll, and yaw, along with translation along the X, Y, and Z axes). Some simulators can even incorporate emergency conditions such as smoke in the cockpit (Powell-Dunford, Bushby, & Leland, 2016).

Despite the long history of flight simulation and the promise of simulation for SD training over other means such as a Barany chair, the use of VR in SD training is a relatively recent development. Simulation was not a common means of SD training through most of the 1990s and even into the 2000s. The Polish Air Force implemented simulation as part of their SD training program in 1998 (Kowalczuk & Mikuliszyn, 2002), but as of 2002, most NATO countries still did not use simulation for SD training (Estrada, Adam, & Leduc, 2002). The Royal Air Force has since established the Disorientation Trainer (DISO) in order to provide training in vestibular and visual illusions (RAF, 2016). The US Army developed SD training for rotary wing pilots based on reproducing real-world mishaps in a flight simulator, which are then flown by the student and debriefed by an instructor (Johnson, Estrada, Braithwaite, & Manning, 1999). The scenarios were well-received by the pilots, who believed them to be beneficial. However, only the National Guard implemented the training on a full-time basis, whereas the Army used the scenarios on a voluntary basis. Only 49% of surveyed pilots received any type of SD training in a simulator, and only 30% used the generated scenarios (Estrada, Adam, & Leduc, 2002).

Using Simulation Effectively. Training effectiveness tends to be measured subjectively based on trainee feedback, or objectively based on changes in performance immediately following training. Most anecdotal evidence from both students and instructors indicates that simulator-based SD training is beneficial (Powell-Dunford, Bushby, & Leland, 2016). Pilots' rating of SD training quality tends to be negatively related to SD incidents such that better ratings lead to fewer incidents

(Mathews, Previc, & Bunting, 2002). The long-term efficacy of such training is more difficult to assess given the mobile nature of military members (Powell-Dunford, Bushby, & Leland, 2016). Even so, mishap rates seem to call the long-term efficacy of SD training into question, at least in terms of preventing the most serious consequences of SD. The rate of SD-related mishaps has remained stable for the last 30-40 years, despite declining mishap rates overall (Estrada, Adam, & Leduc, 2002; Gibb, Ercoline, & Scharff, 2011; Gibb, Musselman, & Farley, 2012; Heinle & Ercoline, 2003).

Simulation by itself does not guarantee effective training. The simulation must provide the necessary instructional features to promote learning (Salas, Bowers, & Rhodenizer, 1998). Simulator development follows strict guidelines, but no similar protocols exist for the training conducted in the simulator. Instructors teach according to their own style (Salas et al., 1998). Among rotary wing pilots, proposed best practices include conducting a student debrief, utilizing simulators for refresher training, utilizing simulators for relevant research, utilizing reality-based scenarios, and utilizing advanced technologies (Powell-Dunford, Bushby, & Leland, 2016). Two of these recommendations imply that higher fidelity enhances training (use of reality-based scenarios and advanced simulation technology). We must therefore critically examine the fidelity requirements necessary to promote effective learning in a flight simulator.

VR Fidelity Requirements. The realism of a given VR training scenario is often equated with the training effectiveness of that scenario (Salas, Bowers, & Rhodenizer, 1998). Indeed, Newman (2016) illustrates this perception when he states that realism is the key component in an effective simulator. While high fidelity simulations are certainly required for certain training domains, such a requirement may not apply for all types of SD training. Several studies have found failure to transfer skills from high-fidelity simulators to actual flight, and in some cases, lower-fidelity simulations can actually provide superior training (Salas et al., 1998). Fidelity has gained such a prominent place in simulation training because simulators and training scenarios are typically evaluated based on the subjective response of the instructors and trainees. These evaluations lead to a false sense that higher fidelity is better because pilots tend to prefer higher fidelity simulations. Although pilot buy-in is an important consideration for training design, it is not a good measure of training effectiveness (Salas et al., 1998). We must redefine how we think about fidelity and the qualities we look for in a training simulator. Training features embedded in the scenario determine training quality much more than the simulation alone.

At its most basic, fidelity is a matter of how faithfully a simulation represents the real world (Grierson, 2014). Fidelity for training purposes should be judged based on whether the simulation captures the important elements of task performance rather than how well it recreates the physical space in which that task occurs (Grierson, 2014). Therefore, fidelity should be based on the cognitive and behavioral requirements to support learning of the task being trained (Salas, Bowers, & Rhodenizer, 1998). The first principle of aligning a simulation with the world is that the simulation should reflect the relevant stimuli and responses during criterion performance. The type of information presented and the way that information is processed are both important (Grierson, 2014). In some instances, such as SD training for visually-based illusions, high visual fidelity will likely be important. In others, such as training pilots to scan instruments properly, visual fidelity will be less important than physical fidelity (replicating the instrument panel). On the other hand, teaching a pilot how to process information and react in complex environments will rely on the fidelity of the scenario more than the fidelity of the simulator. In the case of SD, which rests heavily on visual and vestibular sensations, we must ask what capabilities are required of a simulator to implement effective SD training scenarios.

Fixed-base Simulators and SD. Given the prominence of vestibular sensations in many SD illusions, one might expect that fixed-base simulators could not deliver effective SD training. However, visual information accounts for approximately 80% of a pilot's orientation (Gibb, Musselman, & Farley, 2012). In fact, the most common types of SD appear to be generated from conflicting visual cues which confound pilots' spatial strategies, and the dominant role of vestibular misperception has been questioned in many other SD illusions (Patterson, Arnold, & Williams, 2013; Patterson, Williams, Folga, & Arnold, 2014). Due to the apparent dominance of visual cues for most SD illusions, fixed-base simulators may prove to be very helpful in reducing mishap rates due to SD in flight.

Fixed-base simulators have proven quite useful for some types of SD, particularly visually-based illusions. The Black Hole Illusion is one specific example of a visual illusion that can be replicated in a fixed-base simulator. This illusion occurs when a pilot flying an approach over dark featureless terrain misjudges the altitude of the aircraft and adopts a low glide slope. The Black Hole Illusion has been studied in fixed-base simulators since at least 1978 (Kraft, 1978). Later studies have successfully replicated the illusion in fixed-base simulators as well (e.g., Gibb, Schvaneveldt, & Gray, 2008; Patterson, Williams, Folga, & Arnold, 2015). For some domains such as rotary wing flight (where visual SD illusions are predominant), fixed-based simulators can be just as effective as motion-based simulators for SD training (Powell-Dunford, Bushby, & Leland, 2016).

Fixed-base simulators are also able to replicate cognitive and vestibular SD. Control reversal errors are one cognitive type of SD that have been replicated in a fixed-base simulator, where pilots misinterpret the attitude display and move the aircraft in an unintended direction (Williams, Littman, Folga, & Patterson, 2014). Vestibular types of SD can be replicated without motion due to the presence of the Optokinetic Cervical Reflex (OKCR). The OKCR is a reflex where pilots tilt their head during a bank to keep the head aligned with the vertical axis and the horizon level in the visual field. This reflex gives fixed-base simulators the potential to induce vestibular illusions such as the leans (where pilots feel a sensation of tilt even in level flight) (Patterson, Williams, Folga, & Arnold, 2016). The OKCR has been induced quite reliably in fixed-base simulators (Eriksson, 2010; Moore et al., 2008; Patterson, Cacioppo, Gallimore, Hinman, & Nalepka, 1997; Smith, Cacioppo, & Hinman, 1997).

Fixed-base simulators appear capable of providing effective training for many of the most common types of SD experienced in flight, but field of view (FOV) must also be taken into account as a major consideration. Generally speaking, simulators with a wider FOV are superior to simulators with a smaller FOV unless the simulated task is very simple (Rogers, Asbury, & Szoboszlay, 2003). Wide FOV displays help stimulate peripheral vision, which is a key source of information for orientation and locomotion (Rogers, Asbury, & Szoboszlay, 2003). Small FOV simulators are able to induce the OKCR response, but not to the same extent as simulators with a large FOV. Beyond a certain point, however, the benefits of increasing FOV level off (Williams, Littman, Folga, & Patterson, 2014). A FOV of around 130° horizontally x 60° vertically appears to offer the best balance between ability to induce SD and cost (Williams, Littman, Folga, & Patterson, 2014).

Motion-based simulators and SD. Despite the promise of fixed-base simulators for successfully replicating many types of SD, some illusions may require a motion-base simulator for pilots to be trained effectively. Pilots must be exposed to situations during training in which they cannot trust vestibular and proprioceptive cues. This approach helps pilots learn to control their behavior and utilize instruments to ignore perceptual information (Tropper, Kallus, & Boucsein, 2009). Motion-base simulators are able to recreate certain proprioceptive and vestibular cues that cannot be replicated in fixed-base simulators. For instance, the sudden accelerations present in a stall situation are absent in a fixed-base simulator. Students had difficulty anticipating stalls during slow flight training in a fixed-base simulator that lacked these cues (Anderson & Macchiarella, 2004). For these reasons, motion-base simulators offer intuitive appeal for SD training.

In line with their intuitive appeal, empirical evidence indicates that motion-base simulators can provide effective SD training. Motion-base simulators can induce unrecognized SD (Previc et al., 2007), and students demonstrated fewer crashes and reduced physiological indicators of stress during post-training test flights when trained using a simulator with six degrees of freedom (Tropper, Kallus, & Boucsein, 2009). Motion-base simulators may even offer superior training compared to fixed-base simulators in some instances, particularly for vestibular illusions. Students trained in a motion-base simulator demonstrated superior objective performance results and subjective performance ratings for combatting the pitch-up illusion, spin recovery, and inadvertent flight into IMC. Students trained in a fixed-base simulator actually demonstrated the worst performance of any training group on the pitch-up and spin recovery tasks, indicating the possibility of negative training (Kallus, Tropper, & Boucsein, 2011).[2] Despite the general effectiveness of motion-base simulators, however, certain limitations remain.

Aside from practical considerations of cost, motion-base simulators do have one main physical flaw in that they cannot eliminate the effects of gravity (Cheung, 2013). Tilting the simulator to simulate bank goes against the gravitational vertical and reduces the sense of coordinated flight (Nooij & Groen, 2011). In this way, motion-base simulators may sometimes offer *less* fidelity than fixed-base simulators by disturbing the sense of realism felt by the pilot.

---

[2] We note, however, that the training consisted primarily of exposure as opposed to guided instruction and debrief. More deliberate instruction could have potentially allowed training in the fixed-base simulator to be more effective.

Ground-based training in simulators can be improved by recreating the sensations and sensorimotor responses found in flight (Cheung, 2013).

Simulation and Gravitational Force. Although both fixed-base and motion-base simulators have proven effective for recreating and training different types of SD illusions, such simulators are inherently limited because they cannot replicate the flight envelopes and the sensations of flight (Cheung, 2013). Prior work by the Polish Air Force indicated that a Stewart-type hexapod motion system was not effective for SD research or training involving illusions due to angular or linear acceleration (Kowalczuk & Mikuliszyn, 2002). Even though a Barany chair is able to create some isolated sensations, the Barany chair is limited to a single axis and cannot recreate the other visual and task demands present in the flight environment (Nooij & Groen, 2011). Simulators that incorporate centrifugal motion provide a solution to the problem of recreating the sensations found in flight (Kowalczuk & Mikuliszyn, 2002).

Centrifuge-based simulators have successfully recreated types of SD such as the G excess illusion (Jia et al., 2001). Two of the most advanced flight simulators in the world—the Desdemona in the Netherlands and the Disorientation Research Device (DRD; a.k.a. "The Kraken") at the Naval Medical Research Unit-Dayton—incorporate centrifugal motion along with six degrees of freedom (pitch, roll, yaw, and translation along the x, y, and z axes; Powell-Dunford, Bushby, & Leland, 2016). Such centrifuge-based simulators are complex and expensive, but they allow researchers to replicate the sensations felt across nearly the entire range of the flight envelope and create a feeling of coordinated flight. For instance, the Desdemona successfully recreated the somatogyral illusion with as little as two seconds of roll (Nooij & Groen, 2011).

Simulators such as the Desdemona and DRD offer the ability to study nearly every component of SD in a controllable, replicable manner to isolate and understand the various contributing factors of SD. Such knowledge will facilitate future training design by allowing us to target the most critical elements of SD. For instance, if we discover that vestibular inputs are largely irrelevant for in-flight SD, we can focus instead on visual cues and teach pilots more effective strategies to gather spatial information from the world.

Head-mounted displays

One emerging avenue for simulation-based training is the use of virtual head-mounted displays (V-HMDs).[3] Rather than external screens displaying the flight console and outside-the-window scene as in a typical simulator, V-HMDs are worn over the eyes as a single display that alters the visual scene based on the position of the wearer's head. Such displays offer the potential to produce a visual representation of the virtual world in every direction, without the cost or space requirements of a large dome and projection system. Another issue, along with how much motion is necessary for any SD training simulator, is whether the increased head movement of a V-HMD interacts with the motion of the simulator to produce unintended vestibular sensations or increases motion sickness. However, very little research has examined how V-HMDs compare to more traditional simulations for inducing SD illusions.

What research has been conducted on V-HMDs and SD illusions has been mixed. Patterson & Muth (2009) were able to demonstrate that V-HMDs can elicit an OKCR response both on land and aboard ship. Some work has examined FOV issues in V-HMDs, but those studies did not properly replicate the pilot's sight picture with the appropriate peripheral and secondary cues found in the cockpit (Williams, Littman, Folga, & Patterson, 2014). Pilots rely on the interaction between cues such as the horizon and glare shield to maintain orientation (Patterson, Arnold, & Williams,

---

[3] In order to distinguish between head mounted displays for virtual reality and helmet-mounted displays for actual flight, I will use V-HMD to refer to virtual reality displays.

2013). Given the importance of visual cues and their impact on the spatial strategies used by pilots, any SD simulation utilizing V-HMDs must recreate the pilot’s sight picture accurately. Failure to do so is likely to impact the pilot’s experience of SD and/or lead to ineffective or negative training.

Not only must the primary and secondary cues be accurately represented, but simulations must preserve the pilot’s ability to scan between the outside world and the cockpit. Traditional simulators allow pilots to switch between the cockpit instruments and the outside scene with a glance, as they would in an aircraft. If the vertical FOV of a V-HMD is too short, the pilot would have to artificially adjust his or her behavior to look down using head movements rather than glances. The resulting slower transitions to instruments and the increased head motions may alter the effects of SD, particularly for control reversal errors or vestibular illusions. This potential issue may be partially mitigated in aircraft using head-up displays (HUDs) or helmet-mounted displays (HMDs) as the primary flight display.

SD and Augmented Reality

HUDs and HMDs represent augmented reality (AR), which overlays information from a display over the real world. HUDs are projected on a small fixed screen in the front of the cockpit, whereas HMDs are incorporated into the helmet and projected onto a visor or monocle. Both are intended to reduce scanning behavior by allowing the pilot to maintain their gaze outside the cockpit while still retaining access to important display information (Lorenz, Tobben, & Schmerwitz, 2005). AR offers several benefits over traditional cockpit instruments, but also has certain tradeoffs that must be accounted for. We begin this section with an overview of the benefits of AR technology to general flight, followed by a discussion of the potential drawbacks of AR technology and the importance of symbolism to the display’s effectiveness. Finally, we incorporate these findings into a discussion of AR in the context of SD and SD training.

AR Benefits. AR reduces the need to visually scan between the outside-the-window scene and the cockpit instruments because relevant flight information can be superimposed on the outside world (Fadden, Ververs, & Wickens, 2001). Pilots can therefore spend more time looking at the outside world because visual transitions are kept to a minimum (Weinstein & Ercoline, 1992). To maximize this benefit, AR flight displays are most often collimated to optical infinity, reducing the burden of re-accommodation between instruments and the outside world while also smoothing the transition between the two types of information (Fadden et al., 2001; Lorenz, Tobben, & Schmerwitz, 2005). AR displays can also link information to the real world, such as an artificial horizon overlaid on the real horizon. Such conformal symbolism helps to integrate different sources of information and better link the display with the world (Fadden et al., 2001).

AR displays have proven to be very useful since their introduction, largely because of the reduced need for visual scanning (Martin-Emerson & Wickens, 1997). For instance, carrier waveoffs were reduced and precision approaches/weapons delivery became more accurate after the introduction of HUDs (Ercoline, 2000). Research has also demonstrated that HUDs generally lead to improved performance on flight tracking tasks and facilitate the detection of expected events (e.g., acquiring the runway on approach; Fadden, Ververs, & Wickens, 1998; Martin-Emerson & Wickens, 1997). Although AR displays offer several advantages for flight in general, some disadvantages have been identified as well.

AR Disadvantages. The primary disadvantages of AR displays center around issues of attention allocation and clutter. Despite reduced scanning costs, people must still divide attention between the display and the world. Evidence suggests that HUDs may lead to attentional focus (or "tunneling") on the display at the expense of paying attention to events in the world (Francis & Rash, 2010; Jarmasz, Herdman, & Johannsdottir, 2005). HUDs can impair detection of unexpected events such as runway incursions compared to traditional head-down cockpit instruments (Ercoline, 2000; Fadden et al., 1998). Object-based models of attention suggest that people cannot process multiple sources of information simultaneously unless they are part of the same perceptual object (Foyle, McCann, Sanford, & Schwirzke, 1993). Despite overlapping images, people appear to treat the information in the HUD as separate from the world, most likely because the images on the HUD move together against the background of the world to form a unique perceptual object (Foyle, McCann, Sanford, & Schwirzke, 1993; Jarmasz, Herdman, & Johannsdottir, 2005; Levy, Foyle, & McCann, 1998).

Although attentional tunneling hinders one's ability to detect changes in the outside world by capturing attention, item clutter can make AR displays difficult to use by making the outside world more difficult to see or making information more difficult to find (Fadden, Ververs, & Wickens, 2001; Francis & Rash, 2010). HUDs in particular have a limited viewing area (Newman & Haworth, 2004; Parmet & Ercoline, 2008), making display space valuable real estate that must be used carefully to avoid filling the screen. The lack of color in HUDs compounds the issue because rather than using color to differentiate between different system states, additional symbols must be used (Ercoline, 2000).

The Importance of Symbolism. Even more than attentional tunneling or clutter issues, the symbolism of the AR display has a dramatic effect on how useful the display can be. AR is not inherently useful; display design can greatly impact performance. AR displays with different symbolism have demonstrated markedly different performance from one another on tasks such as approach and landing (Cheung et al., 2015) and unusual attitude recovery (Ercoline, Self, &

Matthews, 2002; Weinstein & Ercoline, 1992; Wickens, Small, Andre, Bagnall, & Brenaman, 2008). Good design can even outweigh display location, such that flight performance can be equivalent regardless of whether the display is in a head-down or head-up location (i.e., whether the display overlays the outside world or not; Fadden et al., 2001).

The availability of information is no guarantee of its use (Francis & Rash, 2010). There is an important distinction between displays that simply make data available and displays that actively support information extraction. The former can burden memory and hinder problem solving, whereas the latter can help the pilot integrate raw data to support problem solving and reduce workload (Bennett & Flach, 1992). In particular, conformal or scene-linked symbolism on an AR display can help alleviate some problems with attention and clutter (Fadden, Ververs, & Wickens, 2001; Martin-Emerson & Wickens, 1997; Ververs & Wickens, 1998). Conformal symbols correspond to something in the outside world to allow the formation of a single object, such as an artificial horizon overlaid on top of the true horizon (Martin-Emerson & Wickens, 1997). Scene-linked symbols are displayed such that they appear as an object out in the world (Levy, Foyle, & McCann, 1998). Both conformal and scene-linked symbols often prove superior to traditional instrument symbolism, particularly for tracking tasks (Fadden, Ververs, & Wickens, 1998; Levy, Foyle, & McCann, 1998; Martin-Emerson & Wickens, 1997). Conformal and scene-linked symbolism likely benefit from being able to form a perceptual object with the world to facilitate shared attention or more complete integration between the world and the display (Jarmasz, Herdman, & Johannsdottir, 2005; Levy, Foyle, & McCann, 1998).

AR and SD. Just as AR displays have benefits and drawbacks for routine flight, they have advantages and disadvantages when it comes to the potential for SD. Early HUDs were actually reported to increase the rate of SD, particularly during situations known to promote SD occurrence.

As training and design improved, the reported increase in SD was mitigated (Newman & Haworth, 2004). Certain inherent design limitations still apply, however, making pilots potentially more vulnerable to SD while using a HUD. The small FOV on a HUD can make the display difficult to read in turbulent conditions. The relatively small display area also constrains the spacing of pitch lines, causing small changes in pitch to appear quite large and large changes in pitch to be difficult to follow (Ercoline, 2000). The limited viewing area can also make the horizon hard to use, particularly when the “0” line disappears under certain maneuvers (Parmet & Ercoline, 2008). Other limitations of HUDs are related to the symbolism. The necessary orientation information can sometimes be difficult to find, and the orientation ladder can be difficult to interpret or look similar in different attitudes (Ercoline, 2000; Gawron, 2004). Some studies have indicated that the HUD can be less effective than the traditional attitude display for unusual attitude recovery (Newman & Haworth, 2004), though others have found no difference (Huber, 2006). Many of the issues related to symbolism could be corrected with a new design. For instance, Cheung et al. (2015) found that the rate of SD was five times greater for one HUD display than another during a comparison of rotary wing landing displays. Finally, attentional tunneling may contribute to SD if the pilot focuses on a single part of the display at the expense of spatial information or the outside world.

On the other hand, AR displays have the potential to actually reduce the rate of SD. AR displays can preserve the relationship between the primary and secondary visual cues that pilots use as their frame of reference, whereas head-down displays reverse the relationship and may cause confusion (Rogers, Asbury, & Szoboszlay, 2003). HUDs also reduce the need for the pilot to move their head, which may help avoid certain vestibular illusions (Ercoline, 2000; Moore et al., 2008). In fact, HMDs with a wide FOV could potentially serve as a strong enough visual stimulus to override many vestibular illusions completely. For instance, Eriksson (2010) presented visual motion cues to participants who were standing still on a fixed platform. The participants altered their posture in response to the cues with as little as 108° FOV, indicating that they perceived movement even though they were standing still and received no vestibular input. HMDs that superimpose a horizon on the world can provide relief for people with motion sickness or reduce the number of therapy sessions required for people with vestibular conditions, offering further evidence that visual stimuli can influence vestibular sensations (Krueger, 2011). Despite the potential of an HMD with a wide FOV, HMDs have their own considerations that must be taken into account.

HMDs offer a much wider FOV than HUDs (Lueken, Doehler, & Schmerwitz, 2016) and allow the pilot to access instrument information from any head position, but the increased FOV and access to information must be managed carefully. HMDs allow the pilot to maintain access to display information while looking off-axis, but orientation information may conflict with the visual scene if the display is not designed carefully (Ercoline, Self, & Matthews, 2002). This issue is particularly important for rotary wing pilots, who spend considerable time looking off axis and flying at low altitudes (Rogers, Asbury, & Szoboszlay, 2003). Because HMD displays are always in the pilot’s line of sight, particular care must be taken to provide usable information without cluttering the pilot’s visual field (Cheung, 2013). Frames of reference, head tracking, and head orientation are all considerations when planning how best to utilize a HMD (Newman & Haworth, 2004). Increased capability can also lead to unanticipated consequences. For instance, the F-35 HMD famously allows pilots to “see through” the aircraft. Depending on how this is implemented, pilots may lose secondary visual cues and face an increased risk of SD. On the other hand, such a capability may allow the pilot to locate the ground in nearly any attitude, potentially aiding orientation. Further research is needed to explore the various implications of such AR technology for SD and how they impact pilot orientation cues and spatial strategies. Training practices may need to be altered to account for changes in orientation brought about by AR technology.

AR and SD Training. Very little research has examined the impact of V-HMDs on SD simulation—even less has examined whether AR can be incorporated into V-HMD training. Despite the lack of empirical research on the subject of whether AR and V-HMD can be integrated simultaneously, multiple questions come to mind that would be of great research interest. First, there is the basic question of whether AR provides any benefit to standard SD training. For instance, AR could be used to highlight various visual cues and demonstrate how they interact in real time. This approach could potentially help pilots identify key spatial cues and adopt more effective spatial strategies, but it could also lead to dependence on the AR and lead to negative effects when the pilots try to fly without it. We should examine how effective such training could be and how best to implement it without negative consequences. We must also assess how AR for SD training may interact with the AR already present in the simulation (e.g., the HUD in a simulated fighter aircraft). The additional spatial AR information could potentially conflict with or clutter the indigenous HUD AR. If so, pilot performance could suffer artificially or negative training could result. To combat this, we should also examine how aircraft-specific SD training needs to be. We should examine whether the cues and strategies learned in one aircraft transfer to another. If so, perhaps a generic training platform with only spatial AR can be used to teach pilots basic orientation tools. More advanced practice could then be done on that pilot's specific platform with their typical HUDs or HMDs.

Recommendations for the Future

We recommend research in several areas to better utilize existing technologies and ensure effective, cost-efficient training in the future. The results of the proposed research will inform both training design and simulator acquisition to promote the development of robust training systems. The stubbornly high rate of SD mishaps over the previous decades indicates that current SD interventions are largely failing to prevent the most serious consequences of SD, despite the introduction of advanced displays and training simulators. Simulation offers promise for improving SD training and reducing mishaps, but technology advancement has outpaced our existing knowledge regarding the application of these new capabilities. HMDs for simulators in particular represent an unknown quantity, as many interpretations about the efficacy of HMDs for practical use and training are really just assumptions based on what we know about VR technology. Broadly speaking, we need a better understanding of SD from a cognitive perspective in order to design more effective training interventions that best utilize the available technology. We recommend the following to create a more effective SD training program by better capitalizing on existing technology and maximizing the training benefit of future technologies:

1. Validate current training methods.
    a. Ensure that training targets the correct illusions. Much of the current training focuses on vestibular illusions, but empirical research has questioned how prevalent such illusions are in the actual flight environment. We must conduct research based on mishap reports to identify the most common causes of SD-related mishaps in the real world so that training interventions can be targeted to the most common/serious illusions.

   b. Ensure that training is effective. Most training evaluations are based on subjective reports of trainee satisfaction or instructor ratings of improvement. Objective measures of training effectiveness are most often collected immediately after training. In order to get a better sense of the true effectiveness of training, we recommend conducting longitudinal (or even archival) research into the rate of SD-related mishaps among pilots who have completed training, what type of training was received by pilots with and without mishaps, and time between training and a mishap.

2. Shift the focus of training away from recognition in favor of avoidance.
   a. We should stop thinking about SD training in terms of the mechanical causes. Current demonstration-based training is based on the question "what is SD?" Researchers have identified the vestibular and visual causes of disorientation, and worked to recreate these sensations during training so that they can be recognized by pilots. Teaching pilots to recognize SD is difficult because SD is subtle; illusions are difficult to become aware of by their very nature. Further, recognition does nothing to prevent SD from occurring in the first place. Also, even if pilots identify a conflict between their perception and instrument readings, as with recognized SD, there is no guarantee the pilot will recover from the SD-induced flight problems.
   b. We should start thinking about SD in terms of the cognitive strategies pilots use to stay oriented. Rather than focusing on "what is SD?" a more productive line of research may be "how do we avoid SD?" SD is not just a sensation. SD involves action or inaction based on a pilot's perception, decision making, and seeking spatial cues. We must turn away from a strictly vestibular or visual perspective on

SD and incorporate a cognitive component to account for how pilots perceive, integrate, and prioritize multiple types of inputs (visual, vestibular, and instruments) to form a coherent sense of orientation. By reframing SD research to identify the strategies pilots use to remain oriented and the limitations of those strategies, we can teach pilots active ways to avoid SD rather than try to make them recognize SD once it has already occurred. Simulators would be ideal for such training because they afford multiple practice attempts for students and the ability for instructors to manipulate different cues for emphasis or to highlight/promote various strategies. They also afford researchers the ability to manipulate different features of scenarios to better identify relevant stimuli and spatial strategies that pilots use to judge orientation. Such an approach has already demonstrated promise. A study by Patterson and his collaborators (2015) demonstrated that training based on the visual strategies pilots use to judge glide path can help reduce the incidence of black hole approaches. Rather than simply telling pilots that black hole illusions exist and pilots should trust the instruments, Patterson et al. explained the visual cues that regulate landing and how the use of those cues can change at night to cause a low approach. They emphasized the spatial strategy rather than the visual cue itself.

3. Emphasize learning when acquiring simulators
    a. In many cases the training instruction and scenario are greater determinants of training effectiveness than the physical capabilities of the simulator. After we gain a better understanding of what types of SD are most important for preventing mishaps and the key contributors to those illusions, we will be able to design more

effective training scenarios and provide more targeted instruction to promote pilot orientation. Simulators must be designed to accommodate such training by providing performance measures for instructors and supporting a variety of scenarios that can be manipulated and adjusted as necessary to support instruction. Simulator design and acquisition must account for how people will learn in the simulator (Salas, Bowers, & Rhodenizer, 1998). Simulator designers must be able to explain how the simulator and proposed training scenarios support learning from a cognitive perspective and how the simulator captures the relevant aspects of any given illusion.

4. Identify key simulator requirements
    a. Visual processes appear dominant in SD, raising the issue of whether fixed-base simulators may be adequate for nearly all SD training. Just as Williams et al. (2014) examined the FOV requirements for various visual illusions, we should examine the motion requirements (if any) for reproducing other types of illusions. Such work could be accomplished in the DRD. More “bells and whistles” does not necessarily equate to better training. By determining the minimum requirements for a simulator to effectively reproduce the necessary visual, vestibular, and cognitive elements of SD, we can deliver effective training at lower cost.
    b. An indirect requirement with the simulators involves the manpower needed to run them. Comprehensive flight operations require coordinating across many different platforms and many different mission spaces. Any such all-inclusive approach will require significant manpower to design and operate, which further complicates the simulator requires above and beyond display or fidelity issues.

5. Explore emerging technology
    a. Finally, we should evaluate the utility of newer technologies such as HMDs. Among the most critical issues with current technology are FOV requirements and how well the display is able to recreate the pilot's primary and secondary visual cues. If the display forces pilots to alter their behavior via increased head movement or does not allow for the normal use of spatial cues, the effectiveness of any SD training will be called into question. Depending on how much motion is required to successfully train SD in a simulator, the interaction of such motion and increased head movements from the HMD may contribute to simulator sickness and reduce students' ability to train. At a minimum, HMDs must be compared to traditional simulators to confirm that they offer equivalent training.
    b. We should explore the potential for AR technologies to enhance SD training. AR may facilitate training by highlighting relevant spatial cues and demonstrating effective orientation strategies. However, we must remain mindful that AR spatial cues may clutter or conflict with HUD cues already present in some simulators. We should examine the effectiveness of various combinations of AR technologies to optimize SD training.

Conclusion

The emergence of flight simulators as a viable SD training tool offers the potential to finally begin to reduce stubbornly high SD mishap rates. However, this technology is not utilized in a vacuum. The simulator is a tool within a larger training system; how it is used will have a far larger impact on the effectiveness of training than the features of the simulator itself. In order to capitalize on the training potential of simulators, we must focus not only on the capabilities of the simulator, but also on the scenarios used, the types of illusions emphasized, and even the nature of how we think about SD. We must work to shift our research emphasis to better understand the strategies used by pilots and determine the key elements of pilot orientation in flight. Simulator design should emphasize learning rather than simple fidelity and we should critically examine the overall effectiveness of current simulator-based training practices. Finally, we must identify the features of a simulator (visual fidelity, motion, etc.) required to faithfully reproduce the critical components of SD and determine whether HMDs meet these requirements and how AR technology can enhance learning in a simulator.

References

Anderson, R., & Macchiarella, N. (2004). Nonmotion flight simulation development of slow flight and stall tasks for use in ab initio flight training. *Human Performance, Situation Awareness, and Automation: Current Research and Trends,* 180-184.

Bennett, K.B., & Flach, J.M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34(5),* 513-533.

Bos, J., Bles, W., Hosman, R., & Groen, E. (2003). The cause of spatial disorientation. *Paper presented at the RTO HFM Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences, and Cures". La Coruna, Spain, 15-17 April.*

Cheung, B. (2013). Spatial disorientation: More than just illusion. *Aviation, Space, and Environmental Medicine, 84,* 1211-1214.

Cheung, B., Craig, G., Steels, B., Sceviour, R., Cosman, V., Jennings, S., & Holst, P. (2015). In-flight study of helmet-mounted symbolism system concepts in degraded visual environments. *Aerospace Medicine and Human Performance, 86,* 714-722.

Patterson, F., & Muth, E. (2009). Cybersickness onset with reflexive head movements during land and shipboard head-mounted display flight simulation. NAMRL report #10-43.

Ercoline, W. (2000). The good, the bad, and the ugly of head-up displays. *IEEE Engineering in Medicine and Biology,* 66-70.

Ercoline, W., Self, B., & Matthews, R. (2002). Effects of three helmet-mounted display symbologies on unusual attitude recognition and recovery. *Aviation, Space, and Environmental Medicine, 73,* 1053-1058.

Eriksson, L. (2010). Toward a visual flow integrated display format to combat pilot spatial disorientation. *International Journal of Aviation Psychology, 20,* 1-24.

Estrada, A., Adam, G., & Leduc, P. (2002). Use of simulator spatial disorientation awareness training scenarios by the U.S. Army and National Guard. *Paper presented at the RTO HFM Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences, and Cures", La Coruna, Spain, 15-17 April 2002.*

Fadden, S., Ververs, P., & Wickens, C. (1998). Costs and benefits of head-up display use: A meta-analytic approach. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 16-20.

Fadden, S., Ververs, P., & Wickens, C. (2001). Pathway HUDs: Are they viable? *Human Factors, 43*, 173-193.

Foyle, D., McCann, R., Sanford, B., & Schwirzke, M. (1993). Attentional effects with superimposed symbolism: Implications for head-up displays (HUD). *Proceedings of the Human Factors and Ergonomics Society 37th Meeting*, 1340-1344.

Francis, G., & Rash, C. (2010). Cognitive considerations for helmet-mounted display design. In P. Marasco and P. Havig (eds.), *Proceedings of SPIE vol. 7688, Head- and Helmet-Mounted Displays XV: Design and Applications.*

Gibb, R., Ercoline, B., & Scharff, L. (2011). Spatial disorientation: Decades of pilot fatalities. *Aviation, Space, and Environmental Medicine*, *82,* 717-724.

Gibb, R., Musselman, B., Farley, R.; "Spatial disorientation … still a problem in 2012". USAF journal *The Combat Edge*, July/August 2012. http://www.acc.af.mil/librar y/accsafety.asp

Gibb, R., Schvaneveldt, R., & Gray, R. (2008). Visual misperception in aviation: Glide path performance in a black hole environment. *Human Factors, 50,* 699-711.

Grierson, L. (2014). Information processing, specificity of practice, and the transfer of learning: Considerations for reconsidering fidelity. *Advances in Health Science Education, 19*, 281-289.

Heinle, T., & Ercoline, W. (2003). Spatial disorientation: Causes, consequences, and countermeasures for the USAF. *Paper presented at the RTO HFM Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences, and Cures". La Coruna, Spain, 15-17 April.*

Huber, S. (2006). Recovery from unusual attitudes: HUD vs. back-up display in a static F/A-18 simulator. *Aviation, Space, and Environmental Medicine, 77,* 444-448.

Jarmasz, J., Herdman, C., & Johannsdottir, K. (2005). Object-based attention and cognitive tunneling. *Journal of Experimental Psychology: Applied, 11,* 3-12.

Jia, H., Yu, L., Bi, H., Wang, K., Liu, Z., & Xie, S. (2001). Ground simulation of the G-excess illusion. *Space Medicine and Medical Engineering, 14,* 88-91.

Johnson, P., Estrada, A., Braithwaite, M., & Manning, J. (1999). Assessment of simulated spatial disorientation scenarios in training U.S. Army aviators. USAARL report #2000-06.

Kallus, K., Tropper, K., & Boucsein, W. (2011). The importance of motion cues in spatial disorientation training for VFR-pilots. *The International Journal of Aviation Psychology, 21,* 135-152.

Kirkham, W., Collins, W., Grape, P., Simpson, J., & Wallace, T. (1978). Spatial disorientation in general aviation accidents. Report prepared for the Federal Aviation Administration (FAA; report # FAA-AM-78-13). Oklahoma City, OK.

Kowalczuk, K., & Mikuliszyn, R. (2002). Spatial disorientation in military vehicles: Causes, consequences, and cures. *Paper presented at the RTO HFM Symposium, La Coruna, Spain, April 15-17, 2002.*

Kraft, C. (1978). A psychophysical contribution to air safety: Simulator studies of visual illusions in night visual approaches. In H.L. Pick et al. (eds.), *Psychology: From Research to Practice*, pp. 363-385. New York: Plenum Press.

Krueger, W. (2011). Controlling motion sickness and spatial disorientation and enhancing vestibular rehabilitation with a user-worn see-through display. *The Laryngoscope*, 121, S17-S35.

Levy, J., Foyle, D., & McCann, R. (1998). Performance benefits with scene-linked HUD symbolism: An attentional phenomenon? *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 11-15.

Lorenz, B., Tobben, H., & Schmerwitz, S. (2005). Human performance evaluation of a pathway HMD. In J. Verly (ed.), *Enhanced and Synthetic Vision 2005, Proceedings of SPIE Vol. 5802*, 166-176.

Lueken, T., Doehler, H., & Schmerwitz, S. (2016). Helmet mounted display supporting helicopter missions during en route flight and landing. In J. Sanders-Reed and J Arthur III (eds.), *Degraded visual environments: Enhanced, synthetic and external vision solutions. SPIE.*

Martin-Emerson, R., & Wickens, C. (1997). Superimposition, symbolism, visual attention, and the head-up display. *Human Factors, 39*, 581-601.

Matthews RS, Previc F, & Bunting A. (2002). "USAF spatial disorientation survey". Paper presented at the Research and Technology Organization/Human Factors and Medicine

Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences, and cures. April; La Coruna, Spain. HFM-085, RTO-mp-086.

Moore, S., MacDougall, H., Lesceu, X., Speyer, J., Wuyts, F., & Clark, J. (2008). Head-eye coordination during simulated orbiter landing. *Aviation, Space, and Environmental Medicine, 79,* 888-898.

Newman, D. (2007). An overview of spatial disorientation as a factor in aviation accidents and incidents. *Australian Transport Safety Bureau report #B2007/0063.*

Nooij, S., Groen, E. (2011). Rolling into spatial disorientation: Simulator demonstration of the post-roll (Gillingham) illusion. *Aviation, Space, and Environmental Medicine, 82,* 505-512.

Parmet, A., & Ercoline, W. (2008). Spatial orientation in flight. In J. Davis, R. Johnson, J. Stepanek, & J. Fogarty (Eds.), *Fundamentals of Aerospace Medicine, 4th Ed., pp. 142-205.* Philadelphia: Lippencott Williams & Wilkins.

Patterson, F., Arnold, R., & Williams, H. (2013). Visual perspective illusions as aviation mishap causal factors. *Paper presented at the International Symposium on Aviation Psychology, May 6-9, Dayton, OH.*

Patterson, F., Cacioppo, A., Gallimore, J., Hinman, G., & Nalepka, J. (1997). Aviation spatial orientation in relationship to head position and attitude interpretation. *Aviation, Space, and Environmental Medicine, 68,* 463-471.

Patterson, F., Williams, H., Folga, R., & Arnold, R. (2014). Identification of pilot spatial strategies for mitigation of disorientation: A new look at an old problem. *Paper presented at the annual meeting of the Aerospace Medicine Association.*

Patterson, F., Williams, H., Folga, R., & Arnold, R. (2015). Effects of spatial disorientation countermeasures training on black hole illusion (BHI) onset distance and recovery altitude. *Paper presented at the annual meeting of the Aerospace Medicine Association.*

Powell-Dunford, N., Bushby, A., & Leland, R. (2016). Spatial disorientation training in the rotor wing flight simulator. *Aerospace Medicine and Human Performance, 87,* 890-893.

Previc, F., Ercoline, W., Evans, R., Dillon, N., Lopez, N., Daluz, C., & Workman, A. (2007). Simulator-induced spatial disorientation: Effects of age, sleep deprivation, and type of conflict. *Aviation, Space, and Environmental Medicine, 78,* 470-477.

RAF Centre of Aviation Medicine (2016). Training Memorandum #35: Standard operating procedures for the conduct of practical training in spatial disorientation for British military aircrew.

Rogers, S., Asbury, C., & Szoboszlay, Z. (2003). Enhahced flight symbolism for wide field-of-view helmet-mounted displays. In C. Rash and C. Reese (eds.), *Helmet- and Head-Mounted Displays VIII: Technologies and Applications. SPIE.*

Salas, E., Bowers, C., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8,* 197-208.

Smith, D., Cacioppo, A., & Hinman, G. (1997). Aviation spatial orientation in relationship to head position, altitude interpretation, and control. *Aviation, Space, and Environmental Medicine, 68,* 472-478.

Tropper, K., Kallus, K.W., & Boucsein, W. (2009). Psychophysiological evaluation of an antidisorientation training for visual flight rules pilots in a moving base simulator. *The International Journal of Aviation Psychology, 19,* 270-286.

Ververs, P., & Wickens, C. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology, 8,* 377-403.

Newman, J. (April, 2016). Virtual becoming reality: The next training evolution. *Naval Aviation News.*

Weinstein, L., & Ercoline, W. (1992). The utility of a ghost horizon and climb/dive ladder tapering on a head-up display. *Proceedings of the Human Factors Society 36th Annual Meeting,* 48-51.

Wickens, C., Small, R., Andre, T., Bagnall, T., & Brenaman, C. (2008). Multisensory enhancement of command displays for unusual attitude recovery. *The International Journal of Aviation Psychology, 18,* 255-267.

Williams, H., Littman, E., Folga, R., & Patterson, F. (2014). Field of view evaluation for flight simulators used in spatial disorientation training. *NAMRU-D report #14-41*.

Fatigue and Eye Strain Associated with Virtual Reality and Augmented Reality Systems

by

Valarie M. Schroeder, LT Adam T. Biggs, MSC, USN, & LCDR John L. Bradley, MSC, USN

Introduction

Simulation technologies (e.g., VR/AR, HUD, HMD, etc.) all require some form of display monitor, and prolonged training exercises expose personnel to these monitors for many hours at a time. This combination creates two related issues: eye strain and general cognitive/physical fatigue. Eye strain represents discomfort resulting from increased effort by the visual system, and general cognitive/physical fatigue represents the detriments in performance someone might experience from prolonged time-on-task. These two issues present a major challenge for VR/AR training scenarios due to the cognitive and perceptual processes required by such systems. Despite a recent resurgence in the development of these technological systems, studies have documented the need for improvement in some areas in order for these systems to demonstrate effective training outcomes.

The purpose of this report is to present the relevant research regarding VR and AR systems as related to fatigue and eye strain for the purpose of implementing such systems into shipboard training programs. Implications and recommendations for integration of these systems into shipboard training programs will be discussed following a review of the relevant literature. Although the processes utilized by the visual system to view various displays associated with VR and AR is relevant to eye strain and fatigue, that research has been reported elsewhere; this review will focus on the associations between such displays and eye strain and fatigue.

Fatigue.

Fatigue is an ill-defined term, characterized most broadly as a feeling of tiredness or exhaustion, and is typically further defined based upon the cause of the fatigue. This review will cover cognitive and physical fatigue as these forms of fatigue are most likely to impact VR/AR training systems for military shipboard operations.

Cognitive fatigue. Cognitive, or mental, fatigue can further be defined as exhaustion relating to objective performance effects on tasks of cognitive abilities (Ackerman, 2011), and is associated with physical fatigue. Cognitive fatigue results in decreased performance across a variety of factors relevant to occupational performance, to include tasks of vigilance (Smit, Eling, & Coenen, 2004), task engagement (van der Linden, Frese, & Sonnentag, 2003), task-switching flexibility (Plukkard, Huizinga, Krabbendam, & Jolles, 2015), attention (Hopstaken, van der Linden, Bakker, Kompier, & Leung, 2016), and emotion regulation (Grillon, Quispe-Escudero, Mathur, & Ernst, 2015). Research suggests that such fatigue results from a combination of factors, to include the limited availability of resources necessary for information processing (i.e., mental overload), as well as decreased arousal due to habituation of a repetitive task (i.e., time-on-task), and can be further influenced based upon motivation levels (Ackerman, 2011; Eysenck, 1982; Mackworth, 1969; Smit, Eling, & Coenen, 2004).

Cognitive load theory is pertinent to VR/AR training paradigms as it considers working memory capacity (WMC) when determining instructional efficacy (Sweller, van Merrienboer, & Paas, 1998), especially with simulation-based programs (Reedy, 2015). Information must be processed by working memory in order for the information to be retained, or to enter into long term memory (Baddeley & Hitch, 1974). Because WMC is limited, information cannot be processed for retention if that capacity is exceeded, resulting in cognitive fatigue. Specifically, task performance will ultimately be hindered as more resources are required to complete the task.

Moreover, repetitiveness of tasks over time or low-arousing tasks can result in disengagement due to a lack of stimulation and/or effort required to complete the task (Eysenck, 1982, Mackworth, 1969). Task performance efficacy is lost in both instances. This has been demonstrated with increased task reaction times (RTs) for a high- vs. low-demanding task (Smit et al., 2004), for cognitively fatigued vs. non-fatigued individuals (Grillon et al., 2015; Plukkard et al., 2015; van der Linden et al., 2003), and for longer time on task (TOT; Hopstaken et al., 2016). Reduced performance has been demonstrated further from electroencephalogram (EEG) recordings, with decrements in cognitive performance for high-demand, but not low-demand, novelty detection tasks evidenced by reduced event-related potential (ERP) amplitudes (Massar, Wester, Volkerts, & Kenemans, 2010).

Cognitive fatigue has additionally been associated with physiological outcomes such as pupil diameter and dilation, as well as physical fatigue. Increased cognitive fatigue results in decreased pupil diameter (Hopstaken, van der Linden, Bakker, & Kompier, 2015a, 2015b), as well as increased physical fatigue, defined as a decrease in time to exhaustion as measured by cycling performance (Marcora, Staiano, & Manning, 2009). However, motivational rewards provided to cognitively fatigued subjects resulted in a restoration of pupil diameter (Hopstaken et al., 2015b). Similar findings have been reported elsewhere, in that motivation may counteract the effects of cognitive fatigue. Indeed, Boksem and Tops (2008) remarked how despite highly demanding cognitive tasks, or workload, fatigue is not imminent; only when motivation associated with work is low will cognitive fatigue be evidenced.

In addition to the use of pupilometry in the measurement of cognitive fatigue (Hopstaken et al., 2015a, 2015b), Pimenta, Carneiro, Neves, and Novais (2016) suggest cognitive fatigue can be assessed based on human-computer interactions. Simple measurements of mouse velocity and accelerations, as well as time between mouse clicks, length of key presses, and mouse location, over prolonged periods of time are associated with varying levels of fatigue. Such measurements can be used for real-time monitoring of cognitive fatigue during extensive human-computer interfacing.

Technology and cognitive fatigue. Cognitive fatigue resulting from depletion of resources is notably applicable to the implementation of AR/VR technology for training due to the novelty of such technological advances and the potential for this novelty to disrupt cognitive processes. These advances could result in an increase in the complexity of cognitive demands associated with AR/VR usage, to include demands on visual resources. It is important to determine that information processing requirements do not detract from operational performance by leading to cognitive overload, and hence fatigue (Huang, Alem, & Livingston, 2013; Livingston, Ai, Karsch, & Gibson, 2011).

Fatigue resulting from 3D exposure has been evaluated using typical cognitive performance metrics (e.g., RTs, performance accuracy), EEG, as well as heartbeat evoked potential (HEP). Research has found that exposure to one hour of a 3D video resulted in subsequent reduced performance accuracy and increased RTs, increased ERP latency, and increased HEP activity on a stimulus detection task compared to 2D video exposure (Park, Won, Lee, Mun, & Whang, 2015). The authors suggest that 3D exposure involves greater workload by the brain simply to process the complex visual information, and leads to increased brain-heart synchronization. This influx of information and communication throughout the body (i.e., brain, heart, vision) is what likely results in cognitive fatigue (Park et al., 2015). Similarly, ERPs and steady-state visually evoked potentials (SSVEPs) from EEG recordings during a stimulus detection task demonstrated that P600 ERP latency increased (suggesting delayed processing), while amplitude decreased (suggesting decreased attention) after one hour of 3D viewing (Mun, Park, Park, & Whang, 2012).

However, applied research suggests that 3D displays are associated with lower levels of cognitive load than 2D displays. Dan and Reiner (2016) recently compared training capabilities utilizing 3D vs. 2D displays in the art of origami on an objective physiological measure of cognitive load. They found that individuals who viewed the 5-minute training videos in 3D had reduced cognitive load as compared to those viewing the instructions in 2D. The 3D viewing proved especially advantageous for individuals who scored lower on a spatial visualization task.

There is little research specifically examining VR systems and cognitive fatigue. Lee and Wong (2014) examined performance of an anatomy lesson using an interactive VR program with high school seniors. They found improved performance for those trained with the VR system as compared to students who received instruction using PowerPoint slides. They also found that students with low spatial abilities benefited the most.

Upon examining cognitive fatigue with AR, lower cognitive workload was indicated in a navigational task with the use of Google Glass as compared to the use of a hand held device (McKendrick et al., 2016), where Glass users demonstrated longer latencies to reach a cognitively fatigued state. Glass-wearing subjects also had fewer errors on an auditory working memory task while navigating a route compared to the subjects utilizing the hand held device. The UK's Defence Evaluation and Research Agency (DERA) Centre for Human Sciences undertook the initiative to examine the use of AR for military embedded training systems, especially for shipboard training for the Royal Navy (Kalawsky, Hill, Stedmon, Cook, & Young, 2000). A series of studies evaluated differences between an AR training format, which involved the use of see-through virtual I-O glasses, and the conventional training format of a PowerPoint presentation utilized on standard PC software. Performance on tests of short term memory, comprehension, and retention of information (Stedmon, Hill, Kalawsky, & Cook, 1999), as well as on an operational task examining system status monitoring (Kalawsky et al., 2000), was equal regardless of the display, suggesting that AR platforms are equally effective as standard PC training platforms with no additional cognitive load. However, cognitive workload was subjectively rated as lower utilizing the AR display compared to the standard PC display (Kalawsky et al., 2000).

Such discrepant findings may be explained by way of the lack of stimulation and/or effort required to complete the task as mentioned earlier. Specifically, the studies finding evidence of greater cognitive load resulting from 3D viewing (Mun et al., 2012; Park et al., 2015) lasted for a duration of one hour and were not interactive or instructional by nature. Conversely, those studies finding evidence of lower cognitive load resulting from 3D viewing or VR/AR technologies (Dan & Reiner, 2016; Lee & Wong, 2014; McKendrick et al., 2016) were interactive training programs, and of shorter duration. That the DERA projects found no difference in cognitive load between the AR and standard PC displays may be an interaction effect. Although the tasks used for the study were shorter in duration, they were not interactive learning paradigms, resulting in a similar cognitive load rather than a degradation or improvement in performance. Another explanation for the conflicting results may be due to the system platforms, and may indicate differences between 3D displays, VR, and AR systems. Unfortunately, such a comparison is premature due to limited availability of research examining each of these platforms.

Physical fatigue. The effects of physical fatigue due to sleep loss, as a result of chronic sleep restriction and total sleep deprivation, are well documented and extremely relevant to military operations given the need for variable operational tempos. Sleep loss is associated with decreased attentional vigilance (Miller, Matsangas, & Shattuck, 2007), memory (Walker, 2008), decision making (Harrison & Horne, 2000; Whitney, Hinson, Jackson, & Van Dongen, 2015), moral reasoning (Olsen, Pallesen, & Eid, 2010), visual encoding (May & Kline, 1987), and visual search (Pomplun et al., 2012), tasks necessary for effective occupational performance. Sleep loss has further been implicated in the Three Mile Island, Chernobyl, and Space Shuttle Challenger disasters (Mitler et al., 1998), and is considered equally as dangerous as driving while intoxicated (Williamson & Feyer, 2000).

In terms of total sleep deprivation, Angus and Heslegrave (1985) found one night of sleep deprivation produced a 30% reduction in cognitive performance, while two nights of sleep loss resulted in a 60% reduction of performance. A meta-analysis by Wickens, Hutchins, Laux, and Sebok (2015) further demonstrated that simple cognitive tasks (e.g., simple attention) are more severely disrupted by sleep restriction and deprivation as compared to complex cognitive tasks (e.g., crystallized intelligence, reasoning, etc.). Once sleep loss occurs, a sleep debt is incurred, defined as the accumulation of sleep loss in which performance will not be restored until a specified amount of sleep is achieved to counteract the effects of the sleep loss. An important aspect of sleep is the circadian rhythm, or the predictable physiological and biological changes that occur in humans in a rhythmic pattern (Stampi, 1992). These rhythms explain why the tendency for sleep is greatest in the early afternoon and early morning hours.

Technology and fatigue due to sleep loss. Research examining the effects of fatigue on performance within VR/AR systems is limited. Fatigue research demonstrates that simple tasks are associated with less arousal as compared to complex tasks, resulting in poorer performance; greater arousal associated with complex tasks tends to compensate for the effects of sleep loss by increasing task engagement, and thus improving performance (Wickens et al., 2015). These results are similar to those found in the cognitive fatigue literature, and may translate to VR/AR systems by determining the level of VR/AR system complexity required by the visual system. If the level of engagement provided by the VR/AR training program is high, performance may not be as compromised as with low levels of engagement despite sleep loss. However, it should be noted that sleep loss and restriction significantly influenced performance on both complex and simple cognitive tasks (Wickens et al., 2015), and some performance degradation may be expected within these systems regardless of complexity.

There has been some research examining fatigue and sleep loss and simulator performance. Fatigue due to sleep loss has resulted in degraded pilot performance in flight and locomotive simulators. Using a moving-base flight simulator, Morris and Miller (1996) found that errors on altitude, airspeed, heading, and velocity significantly increased as pilots became more fatigued. Caldwell, Caldwell, Brown, and Smith (2004) found that flight performance, as assessed with a stationary F-117A Weapon System Trainer (WST) simulator began to decline after 17 hours of continuous wakefulness, with the greatest decline in performance occurring after 27 hours of sleep deprivation. Degraded performance was demonstrated further with fatigued engineers using a locomotive simulator, in that fatigue was associated with greater extreme speed violations and use of air brakes (Roach, Dorrian, Fletcher, & Dawson, 2001). Additional research in the medical domain has examined fatigue specifically in virtual reality simulators.

The majority of medical research examining the use of VR trainers for surgical procedures is promising. Laparoscopic surgery performance was examined under fatigued conditions using a VR trainer with medical students (Tomasko, Pauli, Kunselman, & Haluck, 2012). Students reporting less than two hours of sleep the previous night performed equally well on the VR trainer as those reporting six or more hours of sleep, despite subjective reports of greater sleepiness and greater workload by the sleep deprived group. A similar study examined performance among night and day shift medical students on learning how to perform a renal artery lesion using a vascular VR simulation trainer (Naughton et al., 2011). Both day and night shift students were able to effectively learn the surgical procedure and retain the knowledge a week after VR simulated training. However, the night shift students took one day longer to learn the procedure. Other studies using similar VR trainers have found different results.

Taffinder, McManus, Gul, Russel, and Darzi (1998) found that dexterity performance was degraded for fatigued surgeons using a VR laparoscopic simulator, with fatigued surgeons producing 20% more errors and taking 14% longer to complete the task as compared to rested surgeons. Similarly, fatigue resulted in degraded surgical skills on a VR surgical simulator following a work schedule change to a night shift in which fatigue is greatest due to circadian rhythms. Performance improved, however, over consecutive nights, likely due to an adaptation to the night shift schedule (Leff et al., 2008).

Only one study to date has examined the use of an AR system under fatigued conditions. Baumeister and colleagues (2016) examined the use of a spatial AR projection system compared with a 3D monitor display on the presentation of instructions for completion of a procedural task. The spatial AR system consisted of the projection of procedural instructions onto an object; in the monitor condition, participants received the instructions on a 3D monitor display. Both groups then had to manipulate the object to carry out the instructions. After one night of continuous wakefulness, AR user performance remained stable whereas monitor users demonstrated increased reaction times and errors. These results suggest a protective factor by the AR system, in that despite the effects of sleep loss, performance remained stable through the use of the spatial AR system (Baumeister et al., 2016).

In summary, the effects of cognitive and physical fatigue on VR/AR systems are extremely limited. Preliminary studies suggest that the use of interactive VR and AR systems resulted in equal or improved performance over non-VR and AR comparisons, likely due to decreased cognitive fatigue. Longer duration of time spent viewing 3D compared to 2D displays resulted in degraded performance, likely due to increased cognitive fatigue. Physical fatigue due to sleep loss does degrade flight and locomotive performance as determined with simulators. However, the effects of fatigue on performance using VR medical trainers was mixed, while a spatial AR program was shown to benefit performance under fatigued conditions. One explanation for the discrepancies of the reports pertaining to VR systems may relate to training. Specifically, those studies reporting no effects of fatigue used a training paradigm for a novel task, in which medical students were fairly naïve to the surgical procedure prior to the study. Conversely, those studies showing effects of fatigue, to include the flight and locomotive simulators, trained participants on the systems prior to the study. It appears, therefore, that fatigue may impact performance more greatly for already achieved skills versus those skills sought after in training programs. This could be considered an engagement effect, in that participants were more engaged for the novel tasks, and so effects of fatigue were overridden. An additional explanation, also related to level of engagement, could be due to time on task. Shorter tasks allow for greater arousal, whereas longer tasks, or tasks assessed repetitively over a longer period of time, may result in disengagement.

Eye Strain

The terms eye strain, asthenopia, and visual fatigue are often used synonymously throughout the literature. As such, the present report regards eye strain as an umbrella construct that includes symptoms of tiredness, headaches, and general eye soreness or discomfort resulting from increased effort by the visual system. In our technologically advanced society, additional concepts have been coined, such as computer vision syndrome (CVS) or digital eye strain (DES; Rosenfield, 2016). CVS/DES involves any form of eye or vision problem directly related to the use of digital technology. A survey found that eye strain was associated with less than two hours of computer use daily in 25% of individuals, but in 53% of individuals who use a computer for more than six hours per day (Agarwal, Goel, & Sharma, 2013). As such, an investigation regarding the use of VR/AR shipboard training systems on eye strain is warranted.

Technology and Eye Strain. There are specific components of digital technology in general that may result in increased eye strain, such as reduced blink rate, inappropriate gaze angle, glare, and refractive errors (Rosenfield, 2016). For instance, more cognitively demanding tasks have been shown to result in reduced blink rate (Rosenfield, Jahan, Nunex, & Chan, 2015; Nielsen, Sogaard, Skotte, & Wolkoff, 2008), which then leads to dry eyes and subsequently greater eye strain. However, increased blink rate is also an indication of eye strain, in that the viewer is increasing their rate of blinking in an attempt to minimize dry eyes and maximize focus (Heo, Lee, Shin, & Park, 2014). Gaze angle has also been debated, with a lower visual gaze angle recommended (Fostervold, 2003; Mon-Williams, Plooy, Burgess-Limerick, & Wann, 1998; Nielsen et al., 2008). Lower gaze angle allows for reduced ocular surface area exposure, and thus reduced opportunity for moisture loss (Nielsen et al., 2008), less visual strain (Fostervold, 2003), and increased eye-hand coordination (Leoni, Molle, Scavino, & Dichmann, 1994) which could prove important for VR/AR users. Glare can influence visual perception through screen display washout, resulting in visual fatigue and subsequent increases in performance time (Livingston, 2013; Wimalasundera, 2006), and prove to be a concern to AR systems and displays. Convergence-accommodation conflicts are additionally highly relevant to VR/AR trainer usage; however, they are discussed elsewhere in this report and so will not be discussed here.

For VR/AR technology to be considered effective for training platforms, eye strain and fatigue must be reduced to allow users to interact with such displays (Urvoy, Barkowsky, & Le

Callet, 2013). Some research has examined eye strain by comparing movie viewing on stereoscopic 3D versus 2D displays. A survey found that movie goers watching a 3D movie were 15 times more likely to report symptoms of eye strain than individuals who watched a 2D movie (Solimini, 2013). Yang et al. (2012) report that blurred vision, seeing double images, eye soreness and pain, burning sensation, pulling sensations within the eye, and motion sickness symptoms (e.g., dizziness, nausea, disorientation) were more likely to occur upon 90 minutes of viewing a movie on a 3D as compared to a 2D display, with younger individuals between the ages of 24 and 34 reporting greater visual symptoms than adults over the age of 45.

Subjectively, Kuze and Ukai (2008) report that 85 minutes of viewing a 3D movie resulted in significantly greater eye strain and difficulty focusing as compared to 2D viewing. Iatsun, Larabi, and Fernandez-Maloigne (2013) examined visual fatigue on exposure to six 10-minute 3D and 2D video clips, and found that subjects reported increased fatigue over shorter time periods of 3D exposure compared to 2D exposure. They additionally examined objective eye fatigue via eye tracking, and found that video content significantly contributed to fatigue for both 3D and 2D viewing, but video exposure time contributed to fatigue for 3D, but not 2D, viewing.

In more interactive analyses of 3D displays, twenty minutes of playing a 3D video game resulted in a decrease in blink rate and quality of tear production, as well as a significantly greater area of dryness compared to playing a 2D game (Cardona, Garcia, Seres, Vilaseca, & Gispets, 2011). These results suggest that 3D displays are more greatly associated with visual fatigue and discomfort via dry eye symptoms and reduced blink rate as compared to 2D displays.

Mon-Williams, Wann, and Rushton (1993) examined VR/AR systems specifically by immersing subjects for 10 minutes into an exploratory VR system that utilized a head mounted devise (HMD). They found that 20% of subjects experienced reduced distance vision, and 70%

experienced shifts in distance heterophoria (a deviation of the visual axes from parallelism). However, no subject subjectively reported specific feelings of eye strain, and only 2 subjects reported either sore or tired eyes.

Eye strain resulting specifically from 3D displays are believed to be caused by binocular asymmetry, of which there are three groups that can increase eye strain: spatial distortions, poor filters, and stereoscopic conflicts (Heo et al., 2014; Kooi & Toet 2004; Shibata, Kim, Hoffman, & Banks, 2011). Spatial distortions result from geometric differences in the left and right images (shifts, magnifications, rotations, or blurring of images), poor filters result from photometric differences in the images (luminance, color, blur, contrast, accommodation, or crosstalk), and stereoscopic conflicts occur when depth planes are not optimal, and may involve accommodation-convergence and motion parallax-convergence mismatches (Kooi & Toet, 2004). These authors found that in a passive experiment examining these types of binocular asymmetry using HMD stereoscopic displays, vertical disparity, crosstalk, and blur were the three most influential factors in producing visual discomfort. These symptoms likely arise from the combination of depth cues and motion parallax presented on the display that tax the perceptual system (Reichelt, Häussler, Fütterer, & Leister, 2010), but that can be minimized by not exceeding a screen disparity value of one degree (Lambooij, IJsselsteijn, Fortuin, & Heynderickx, 2009).

Studies have supported such causes of visual discomfort and eyestrain. Heo and colleagues (2014) found that the degree of change in the stereoscopic disparity of a 3D movie was the leading cause of eye strain as measured by blink rate. Lambooij, IJsselsteijn, and Heynderickx (2011) further attest that stereoscopic disparity, screen disparity range, disparity offset, and lateral motion contribute to visual discomfort during a 24-minute viewing of a 3D movie. Kim, Yoo, and Seo (2013) examined subjective discomfort of viewers while watching 10-18 minute 3D movies. They found that animated and dynamic content produced greater visual discomfort compared to real and static content, as did fast camera movement, unpredictable scene changes, moving objects with large disparities at the focal point, high luminance, and large and sudden changes in color and/or luminance. However, an exception was found in terms of motion content. Static content is liable to produce more visual discomfort than motion content when the distance between the foreground and background is greater, as greater demands are placed on the visual system when the viewer must switch their attention between these areas (Li, Barkowsky, & Le Callet, 2014).

Another study examined subjective virtual reality induced symptoms and effects (VRISE) before and after 30 minutes of exploring passive and active VR factory environments across four different VR platforms: an HMD, desktop VR system, reality theatre, and VR projected onto a screen (Sharples, Cobb, Moody, & Wilson, 2008). They found that oculomotor symptoms consisting of eye strain, difficulty focusing, blurred vision, and headache as reported on the Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993) significantly increased from pre- to post-exposure for the HMD, desktop, and projection VR platforms, but not for the reality theatre platform (Sharples et al., 2008). They further found that oculomotor symptoms increased from pre- to post-exposure for light, but not dark conditions, and that symptoms were greater for passive as compared to active post-exposure conditions.

Very little research is available regarding AR systems. Eye strain is still a concern, however. Livingston, Gabbard, Swan, Sibley, and Barrow (2013) suggest that vertical disparity associated with the see-through displays of AR systems are likely to produce eye strain due to extended periods of use. They contend that stereo displays are not calibrated properly, and that stereoacuity needs to be properly corrected for such displays.

Some research has been conducted in an attempt to relieve eye strain and fatigue caused by 3D viewing. Carnegie and Rhee (2015) found that depth-of-field (DoF) blurring can decrease eye strain, blurred vision, and fatigue during a 15-minute VR location exploration task on a desktop computer using an Oculus Rift HMD. They state that visual discomfort is alleviated due to a reduction in 1) the amount of focusing a user must perform, and 2) the range of depths a user must encounter, as less screen area is in focus due to the blurring effect. Blink rate can also be influenced by stimulating a blink with digitally presented visual stimuli via eye monitoring, additionally allowing for an alleviation of eye strain (Crnovrsanin, Wang, & Ma, 2014). Individual differences may contribute to eye strain and visual discomfort susceptibility during 3D stereoscopic viewing for a limited number of people, who can be detected with relatively simple measures in advance (Lambooij, Fortuin, IJsselsteijn, Evans, & Heynderickx, 2011). Such countermeasures suggest the feasibility of the inclusion of VR/AR systems for shipboard training while limiting eye strain and discomfort.

Implications for Naval aviation training

The U.S. Navy has already started the process of implementing VR/AR systems into its shipboard trainers. The Conning Officer Virtual Environment (COVE) is a system that uses PC-based simulators to provide navigation, seamanship, ship handling, piloting, and tactical force protection training for shipboard personnel, while the Virtual & Constructive Representation on Live Avionics Displays (VCR-LAD) initiative plans to incorporate VR/AR programs into aircrew training avionics (NAWCTSD, 2016). Greunke (2015) further demonstrated the feasibility of using commercial off-the-shelf (COTS) VR technology for shipboard Landing Signal Officer (LSO) training. While such programs are designed to optimize occupational performance at a reduced cost, VR/AR programs may present issues that should be addressed prior to immersion into the Naval training system.

Cognitive Fatigue. Little information is available regarding cognitive fatigue associated with the use of either VR or AR training systems. What is known based on the current research is 3D displays in interactive instructional designs provided improved outcome performance compared to 2D displays. An interactive VR training platform was associated with improved performance as compared to a PowerPoint presentation. Lastly, AR systems were associated with equal or better performance due to decreased cognitive fatigue compared to non-AR handheld digital devices and PowerPoint instructional programs. Non-interactive viewing of 3D displays at longer durations, however, is associated with poorer subsequent performance. Therefore, higher levels of interaction within VR and AR deployable trainer systems could prove advantageous to enhance performance outcomes, as has been demonstrated previously (Bailey & Witmer, 1994).

Stanney, Kingdon, Graeber, and Kennedy (2002) refine this finding, however, by reporting that greater interaction or control enhanced performance for locomotion tasks, but limited interaction or control enhanced performance for manipulation tasks. Therefore, level of interaction within VR/AR systems appears to be dependent upon the learning task. It should be noted Stanney and colleagues (2002) further found that greater interaction or control during VR sessions is also associated with more symptoms of sickness.

Unfortunately, this implication is based on limited information; even less is known regarding training durations. Although hour-long 3D viewing is detrimental to performance, it does not appear to be known how long interactive VR/AR training sessions should be before performance is negatively impacted. Moreover, performance of VR/AR trainers has not been examined based on types of training programs. Specifically, performance improvement from the integration of VR in a high school biology class may not translate to improved high-end tactical performance from the same integration of VR.

When considering effective methods of learning in VR/AR deployable trainers, one may wish to examine various aspects of the learning environment, such as design of the program and learner capabilities (Reedy, 2015). For example, Reedy recommends that clinical simulated trainers should 1) allow for goal-free learning, where learners are allowed to explore various responses to a problem rather than finding the one true answer, which only increases intrinsic cognitive load; 2) consider the prevention of "shock and surprise" situations or scenarios by preparing learners in advance, as such scenarios increase extraneous cognitive load that detracts from the training; 3) gradually increase task difficulty over time; and 4) gradually move from lower to higher fidelity simulators so as not to overwhelm learners with extraneous load. Such recommendations may or may not be applicable to military deployable trainers.

Spatial ability of learners also plays a significant role in the effectiveness of VR/AR trainers. Individuals with low spatial abilities benefit the most from such simulated trainers (Dan & Reiner, 2016; Höffler & Leutner, 2011; Lee & Wong, 2014). It is likely that these simulated programs reduced the cognitive load that would typically be needed to mentally produce a 3D image in order to process the displayed image. With a lower cognitive load as compared to 2D displays, the learner now has more cognitive resources available for the processing of the actual information in the training program. No perceived degradation of performance was indicated for high spatial ability learners, indicating that VR/AR trainers are a viable method for instruction at no cost in learning effectiveness.

Physical Fatigue. Research examining the influence of physical fatigue on VR and AR use is also severely limited. Use of simulators under fatigued operating conditions demonstrates that flight and locomotive operating performance is degraded as sleep loss increases. These simulators allow for greater generalization to actual operating performance, but may lack the technological advances seen in today’s VR and AR systems and displays. Moreover, they are not feasible for use for shipboard training programs due to space constraints.

The majority of research examining sleep loss and the use of VR systems has been conducted in the medical field, as such programs allow for training of high-risk procedures without jeopardizing patient safety. However, results from such studies are conflicting. Whereas some studies demonstrated no impaired performance due to fatigue, others found an increase in errors and procedural completion times. Such differences may result from the actual procedures and level of engagement provided by each procedure. As has been demonstrated, greater task engagement may override the effects of sleep loss (Wickens et al., 2015). Task engagement may be lost due to repetitiveness of the tasks being asked to complete, and hence result in poorer performance. Moreover, it is unclear if these findings would translate to shipboard trainers. Surgical procedures are high in dexterity and fine motor skill; such tasks may not transfer to aviation training.

Only one study could be located examining the influences of fatigue on AR performance. Specifically, Baumeister and colleagues (2016) found that an AR system protected performance under fatigued conditions. As compared to instructions provided on a 3D display, performance due to instructions received by the AR display was maintained across the fatigued condition, providing an effective countermeasure to fatigue. Although such results are promising, additional research is required to validate the study, as well as to determine the effects of fatigue on visual AR systems in addition to the spatial AR system that was examined. Moreover, the implementation of a spatial AR system into a shipboard training program is likely not feasible, again due to space constraints required by the projectors.

Of particular importance to aviation trainers is engagement, which may vary within training sessions. Specific to pilots, for example, is that flight maneuvers may prove complex enough to provide adequate levels of engagement allowing for appropriate flight performance by counteracting the effects of fatigue. Flying straight and level, however, may not allow enough engagement to override such influences, as such operations produce low levels of arousal (Matthews & Desmond, 2002; Wickens et al., 2015). Additionally, task duration is sure to influence level of engagement while physically fatigued, as seen previously in the cognitive fatigue literature.

Another critical piece that may need to be considered relates to perceived egocentric distances within VR environments. Research suggests that distance judgements of objects from one's self tend to be inaccurate within VR environments, in that individuals incorrectly perceive distances as bring shorter with the use of VR displays compared to actual distance (Renner, Velichkovsky, & Helmert, 2013). While not directly related to the fatigue research, this may prove a safety issue for fatigued individuals within the shipboard environment. Individuals commencing VR training may be fatigued, and such visual misperceptions may carry over into real environments, posing problematic navigational issues within the ship (van Krevelen & Poelman, 2010). Additional information is needed to determine if such effects are relevant to the VR and/or AR systems of interest, if they do in fact carry over into non-VR environments, and the duration of these effects.

A comparison of VR and AR systems as related to fatigue is not viable at this point due to the limited research available. Specifically, only one study examined cognitive fatigue using a VR trainer, and only a handful of studies have examined cognitive fatigue with AR systems, some of which are over 15 years old. Given the rapid transition and development of new technology, such platforms are likely outdated. Moreover, VR research on physical fatigue is mostly limited to the medical field, and only one study could be located assessing fatigued performance within an AR system. Although promising, additional research is needed.

With the exception of training simulators and the spatial AR platform, the vast majority of the VR and AR systems mentioned in this review could likely be incorporated into shipboard trainers. Most utilize desktop computers and monitor displays, with some taking up even far less space (i.e., HMDs). Durability of such systems are comparable to typical desktop computers and monitors. Those systems utilizing larger wall-mounted or projection screens have greater space constraints.

Eye Strain. There has been considerable research on eye strain and visual discomfort associated with stereoscopic 3D displays that is highly relevant to VR/AR applications. However, there are concerns still to be addressed in an effort to reduce these effects. There are numerous 3D configurations and platforms available for use (Reichelt et al., 2010), which can prove beneficial as these programs can be adapted to specific training programs. Unfortunately, it is difficult to determine the effects of these systems on eye strain without evaluating each specific configuration and platform, as they each present their own unique challenges (Livingston, 2013). Such an in-depth evaluation has yet to be performed. As such, findings reported here should be used with caution.

Imperfect binocular disparity appears to be the greatest concern pertaining to eye strain and fatigue during VR/AR exposure (Kooi & Toet, 2004). Binocular disparity occurs when a difference is created between right and left eye images. When two differing images are presented to the right and left eye, binocular rivalry occurs and details from the competing images are suppressed (IJsselsteijn, Seuntiens, & Meesters, 2006). IJsselsteijn et al. suggest two competing images will result in the brain choosing a dominant stimulus while ignoring the alternate stimulus resulting in a singular image on the visual field (2006). This could result in the absence of crucial information when training scenarios remain unsynced.

Some efforts are underway to reduce these effects, such as the application of filters, image adjustments, and calibrations (Lin & Woldegiorgis, 2015). However, it appears that such efforts are in their infancy. Additional research is needed to determine the potential for such countermeasures. Validation of these efforts would further be recommended prior to implementation into VR/AR systems.

It appears that viewing duration is not as great of importance for eye strain and discomfort as it is for fatigue, as even minimal exposure durations of 3D displays can produce increased visual discomfort (Cardona et al., 2011; Mon-Williams et al., 1993; Sharples et al., 2008). What is not well understood is whether such visual discomfort has the ability to override performance effectiveness. For example, despite increased visual symptomology associated with VR/AR-relevant displays, relatively few, if any, individuals across the above studies are reported as having withdrawn specifically due to eye strain or visual discomfort. Therefore, while symptoms do exist, it is unknown if they are severe enough to warrant an interference with training completion and/or performance.

Another aspect of VR/AR training programs that should be considered is short- and long-term effects of exposure on eye strain and fatigue, as well as overall vision. Research examining exposure duration on eye strain and visual discomfort is conflicting. It has been demonstrated that individuals gradually adapt to visual discomfort during a session (Lambooij et al., 2011; Li et al., 2014), but also that increased exposure duration was associated with increased oculomotor problems (Stanney et al., 2002). Less is known if such repeated exposure has any harmful effects on the visual system over time (Lambooij et al., 2009; Nichols & Patel, 2002). Mon-Williams and colleagues (1993) expressed concern for the potential of permanent degradation of binocular function. In terms of practical considerations, shipboard lighting may need to be considered if introducing VR/AR systems in terms of brightness, contrast, resolution, and field of view (van Krevelen & Poelman, 2010). Lighting can cause glare and hence eye strain and visual fatigue. Combined with VR/AR lighting requirements, one may wish to determine if ships can accommodate such lighting requirements, and if such requirements will impact eye strain and visual fatigue.

Present available information does not reveal whether VR or AR systems produce less detrimental visual effects than the other. The literature of VR and AR systems are closely linked, and is lacking on separate experiences to differentiate user visual experiences. Both reality systems could have similar visual side effects arising from differing depth cues and motion parallax-convergence depth cues (Kooi & Toet, 2004; Reichelt et al., 2010). Unfortunately, more research is needed to confirm or deny such effects. Moreover, the fatigue research indicates greater interaction is associated with greater performance due to decreased fatigue effects. Greater interaction may translate to greater feelings of immersion, which may prove detrimental in regards to eye strain. As such, there may be trade-offs that must be determined between these two lines of research in order to arrive at the most effective system for shipboard trainers.

Categorical recommendations

Based on the available research pertaining to fatigue and eye strain, the following questions should be considered when determining whether to integrate VR/AR systems into shipboard training programs:

- Would such programs be able to implement interactive components into the training? Training efficacy appears to increase by reducing cognitive fatigue through implementation of interactive VR/AR systems (Bailey & Witmer, 1994; Stanney et al., 2002). Assuming adequate shipboard space and relevancy to each specific training program (e.g., LSO vs. CSO training), implementing interactive components could prove advantageous.
- Can the results obtained from the research in this review be generalized to navy-specific VR/AR programs? McIntire, Havig, and Geiselman (2014) found that only 36% of studies demonstrated a beneficial effect of using 3D displays for training purposes, but 52% of spatial understanding and 67% of spatial manipulation studies found 3D systems beneficial. More information is needed on the relevancy and efficacy of using VR/AR systems specifically for shipboard training, especially high-end tactical navy-specific programs.
- What is the minimum and/or maximum time allowed for a training session in order to achieve optimal performance with minimal fatigue and eye strain? Shorter time on task appears to be associated with greater task performance in the fatigue research, and may minimize eye strain.
- What short- and long-term effects pertaining to the visual system can be expected, if any, of repeated exposure to such systems? Current research is conflicting; more information is needed to determine any residual effects of repeated exposure of both VR and AR systems.
- Filters, image adjustments, and calibrations have been proposed as methods to reduce visual discomfort during VR/AR exposure. Can such measures, individually and

combined, effectively reduce such symptoms? Currently, this question remains unanswered due to lack of research and validation of such measures.

Conclusion

It appears the most critical technological aspects of reducing cognitive fatigue with VR/AR systems relates to the level of human-machine interaction of the system. Greater levels of interaction appear to reduce cognitive fatigue. Similarly, level of engagement appears to significantly influence the degree to which physical fatigue impacts performance in these systems. With respect to eye strain, binocular disparity is the single greatest problem found in VR/AR systems. The ability to determine which system, VR or AR, is “better” for integration into shipboard training is unrealistic at this point, mainly due to the paucity of research on both systems.

Little information is available regarding fatigue as related to VR/AR exposure, but what is available suggests few problematic areas, especially if content is interactive. Such systems appear to produce more detrimental effects on eye strain and visual discomfort. According to Reichelt and colleagues (2010), “Inherent visual conflicts of stereo 3D to natural viewing are always going to limit capabilities and features” (pg. 11). However, an evaluation is needed to determine if such conflicts can be controlled or countered to a degree that allows effective training performance.

References

Ackerman, P. L. (Ed.). (2010). *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*. Washington, DC: American Psychological Association.

Agarwal, S., Goel, D., & Sharma, A. (2013). Evaluation of the factors which contribute to the ocular complaints in computer users. *Journal of Clinical and Diagnostic Research, 7*(2), 331-335.

Angus, R. G., & Heslegrave, R. J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulation. *Behavior Research Methods, Instruments, & Computers, 17*(1), 55-67.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 47-89). New York: Academic.

Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a Virtual Environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(18), 1158-162.

Baumeister, J., Dorrian, J., Banks, S., Chatburn, A., Smith, R. T., Carskadon, M. A., et al. (2016). Augmented reality as a countermeasure for sleep deprivation. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1396-1405.

Boksem, M. A., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*, 125-139.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology, 16*(3), 163-181.

Cardona, G., Garcia, C., Seres, C., Vilaseca, M., & Gispets, J. (2011). Blink rate, blink amplitude, and tear film integrity during dynamic visual display terminal tasks. *Current Eye Research, 36*(3), 190-197.

Carnegie, K., & Rhee, T. (2015). Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE Computer Graphics and Applications, 35*, 34-51.

Crnovrsanin, T., Wang, Y., & Ma, K. (2014, April). Stimulating a blink: Reduction of eye fatigue with visual stimulus. Proceedings of the Special Interest Group on Computer-Human Interaction Conference, 2055-2064, Human Factors in Computing Systems.

Dan, A., & Reiner, M. (2016). EEG-based cognitive load of processing event sin 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology, INTPSY-11157*, 1-10.

Eysenck, M. W. (1982). *Attention and arousal*. Berlin, New York: Springer.

Fostervold, K. I. (2003). VDU work with downward gaze: The emperor's new clothes or scientifically sound? *International Journal of Industrial Ergonomics, 31*(3), 161-167.

Greunke, L. C. (2015). "Charlie," development of a light-weight, virtual reality trainer for the LSO community: Time to make the leap toward immersive VR. *Calhoun: The Naval Postgraduate School Institutional Archive*. Thesis Collection.

Grillon, C., Quispe-Escudero, D., Mathur, A., & Ernst, M. (2015). Mental fatigue impairs emotion regulation. *Emotion, 15*(3), 383-389.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology, 6*(3), 236-249.

Heo, H., Lee, W. O., Shin, K. Y., & Park, K. R. (2014). Quantitative measurement of eyestrain on 3D stereoscopic display considering the eye foveation model and edge information. *Sensors, 14*, 8577-8604.

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations – evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior, 27*, 209-216.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. P*sychophysiology, 52*(3), 305-315.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. (2015b). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology, 110*, 100-106.

Hopstaken, J. F., van der Linden, D., Bakker, A. B., Kompier, M. A., & Leung, Y. K. (2016). Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology*, *42*(6), 878-889.

Huang, W., Alem, L., & Livingston, M. A. (2013). *Human factors in augmented reality environments*. New York: Springer.

Iatsun, I., Larabi, M., & Fernandez-Maloigne, C. (2013, February). Investigation of visual fatigue/discomfort generated by S3D video using eye-tracking data. Proceedings from the International Society for Optical Engineering (8648), Stereoscopic Displays and Applications XXIV.

IJsselsteijn, W. A., Seuntiens, P. J., & Meesters, L. M. (2006). Human factors of 3D displays. In O. Schreer, P. Kauff, & T. Sikora (Eds.), *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication*. West Sussex, England: John Wiley & Sons, Ltd.

Kalawsky, R. S., Hill, K., Stedmon, A. W., & Cook, C. A. (2000). Experimental research into human cognitive processing in an augmented reality environment for embedded training systems. *Virtual Reality, 5,* 39-46.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Kim, D., Yoo, J., & Seo, Y. (2013). Qualitative analysis of individual and composite content factors of stereoscopic 3D video causing visual discomfort. *Displays, 34*, 223-240.

Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays, 25*, 99-108.

Kuze, J., & Ukai, K. (2008). Subjective evaluation of visual fatigue caused by motion images. *Displays, 29*, 159-166.

Lambooij, M. T., Fortuin, M., IJsselsteijn, W. A., Evans, B. J., & Heynderickx, I. (2011). Susceptibility to visual discomfort of 3-D displays by visual performance measures. *IEEE Transactions on Circuits and Systems for Video Technology, 21*(12), 1913-1923.

Lambooij, M. T., IJsselsteijn, W. A., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology, 43*, 1-14.

Lambooij, M. T., IJsselsteijn, W. A., & Heynderickx, I. (2011). Visual discomfort of 3D TV: Assessment methods and modeling. *Displays, 32*, 209-218.

Lee, E. A., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers and Education, 79*, 49-58.

Leff, D. R., Aggarwal, R., Rana, M., Nakhjavani, B., Purkayastha, S., Khullar, V., et al. (2008). Laparoscopic skills suffer on the first shift of sequential night shifts: Program directors beware and residents prepare. *Annals of Surgery, 247*(3), 530-539.

Leoni, F. M., Molle, F., Scavino, G., & Dichmann, A. (1994). Identification of the preferential gaze position through evaluation of visual fatigue in a selected group of VDU operators. *Documenta Ophthalmologica, 87*(2), 189-197.

Li, J., Barkowsky, M., & Le Callet, P. (2014). Visual discomfort of stereoscopic 3D videos: Influence of 3D motion. *Displays, 35*, 49-57.

Lin, C. J., & Woldegiorgis, B. H. (2015). Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display, 23*(7), 319-332.

Livingston, M. A. (2013). Issues in human factors evaluations of augmented reality systems. In W. Huang, L. Alem, & M. A. Livingston (Eds.), *Human Factors in Augments Reality Environments* (pp. 35-66). New York: Springer.

Livingston, M. A., Ai, Z., Karsch, K., & Gibson, G. O. (2011). User interface design for military AR applications. *Virtual Reality, 15*, 175-184.

Livingston, M. A., Gabbard, J. L., Swan, J. E., Sibley, C. M., & Barrow, J. H. (2013). Basic perception in head-worn augmented reality displays. In W. Huang, L. Alem, & M. A. Livingston (Eds.), *Human Factors in Augments Reality Environments* (pp. 3-10). New York: Springer.

Mackworth, J. F. (1969). *Vigilance and habituation: a neuropsychological approach*. Harmondsworth: Penguin Books.

Marcora, S. M., Staiano, W., & Manning, V. (2009). Mental fatigue impairs physical performance in humans. *Journal of Applied Physiology, 106*(3), 857-864.

Massar, S. A., Wester, A. E., Volkerts, E. R., & Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: Evidence from novelty processing and simulated driving. *Psychophysiology, 47*(6), 1119-1126.

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology, 55A*(2), 659-686.

May, J., & Kline, P. (1987). Measuring the effects upon cognitive abilities of sleep loss during continuous operations. *British Journal of Psychology, 78*, 443-455.

McIntire, J. P., Havig, P. R., & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance: A comprehensive review. *Displays, 35*, 18-26.

McKendrick, R., Parasuraman, R., Murtza, R., Formwalk, A., Baccus, W., Paczynski, M., & Ayaz, H. (2016). Into the wild: Neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Frontiers in Human Neuroscience, 10*, 1-15.

Miller, N. L., Matsangas, P., & Shattuck, L. G. (2007). Fatigue and its effect on performance in military environments. In P. A. Hancock & J. L. Szalma (Eds.), *Performance Under Stress*, 231-250. Burlington, VT: Ashgate Publishing Company.

Mitler, M. M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1998). Catastrophes, sleep, and public policy: Consensus report. *Sleep, 11*(1), 100-109.

Mon-Williams, M., Plooy, A., Burgess-Limerick, R., & Wann, J. (1998). Gaze angle: A possible mechanism of visual stress in virtual reality headsets. *Ergonomics, 41*, 280-285.

Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Opthalmic and Physiological Optics, 13*, 387-391.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology, 42*, 343-360.

Mun, S., Park, M., Park, S., & Whang, M. (2012). SSVEP and ERP measurement of cognitive fatigue caused by stereoscopic 3D. *Neuroscience Letters, 525*, 89-94.

Naughton, P. A., Aggarwal, R., Wang, T. T., Van Herzeele, I., Keeling, A. N., Darzi, A. W., et al. (2011). Skills training after night shift work enables acquisition of endovascular technical skills on a virtual reality simulator. *Journal of Vascular Surgery, 53*(3), 858-866.

Naval Air Warfare Center Training Systems Division (2016). *Programs: I/IT SEC 2016 Fact Sheets*. Retrieved from http://www.navair.navy.mil/nawctsd/Programs.

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: A review of empirical evidence. *Applied Ergonomics, 33,* 251-271.

Nielsen, P. K., Sogaard, K., Skotte, J., & Wolkoff, P. (2008). Ocular surface area and human eye blink frequency during VDU work: The effect of monitor position and task. *European Journal of Applied Physiology, 103*, 1-7.

Olsen, O. K., Pallesen, S., & Eid, J. (2010). The impact of partial sleep deprivation on moral reasoning in military officers. *Sleep, 33*(8), 1086-1090.

Park, S., Won, M. J., Lee, E. C., Mun, S., & Whang, M. (2015). Evaluation of 3D cognitive fatigue using heart-brain synchronization. *International Journal of Psychophysiology, 97*, 120-130.

Pimenta, A., Carneiro, C., Neves, J., & Novais, P. (2016). A neural network to classify fatigue from human-computer interaction. *Neurocomputing, 172*, 413-426.

Plukkard, S., Huizinga, M., Krabbendam, L., & Jolles, J. (2015). Cognitive flexibility in healthy students is affected by fatigue: An experimental study. *Learning and Individual Differences, 38*, 18-25.

Pomplun, M., Silva, E. J., Ronda, J. M., Cain, S. W., Munch, M. Y., Czeisler, C. A., et al. (2012). The effects of circadian phase, time awake, and imposed sleep restriction on performing complex visual tasks: Evidence from comparative visual search. *Journal of Vision, 12*, 1-19.

Reedy, G. B. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing, 11*, 355-360.

Reichelt, S., Häussler, R., Fütterer, G., & Leister, N. (2010). Depth cues in human visual perception and their realization in 3D displays. Presented at Proc. SPIE 7690, Three Dimensional Imaging, Visualization, and Display.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments – A review. *ACM Computing Surveys, 46*(2), 23:1-23:40.

Roach, G. D., Dorrian, J., Fletcher, A., & Dawson, D. (2001). Comparing the effects of fatigue and alcohol consumption on locomotive engineers' performance in a rail simulator. *Journal of Human Ergology, 30*, 125-130.

Rosenfield, M. (2016). Computer vision syndrome (a.k.a. digit eye strain). *Optometry in Practice, 17*, 1-10.

Rosenfield, M., Jahan, S., Nunex, K., & Chan, K. (2015). Cognitive demand, digital screens and blink rate. *Computers in Human Behavior, 51,* 403-406.

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays, 29,* 58-69.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision, 11*, 1-19.

Smit, A. S., Eling, P. A., & Coenen, A. M. (2004b). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica, 115,* 35-42.

Solimini, A. G. (2013). Are there side effects to watching 3D movies? A prospective crossover observational study on visually induced motion sickness. *PLOS One, 8*(2), e56160.

Stampi, C. (1992). *Why we nap: Evolution, chronobiology, and functions of polyphasic and ultrashort sleep*. Boston, MA: Birkhauser.

Stanney, K. M., Kingdon, K. S., Graeber, D., & Kennedy, R. S. (2002). Human performance in immersive virtual environments: Effects of exposure duration, user control, and scene complexity. *Human Performance, 15,* 339-366.

Stedmon, A. W., Hill, K., Kalawsky, R. S., & Cook, C. A. (1999). Old theories, new technologies: Comprehension and retention issues in augmented reality systems. Proceedings of the 43rd Annual Human Factors and Ergonomics Society.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Taffinder, N. J., McManus, I. C., Gul, Y., Russel, R. C., & Darzi, A. (1998). Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *The Lancet, 352*, 1191.

Tomasko, J. M., Pauli, E. M., Kunselman, A. R., & Haluck, R. S. (2012). Sleep deprivation increases cognitive workload during simulated surgical tasks. *The American Journal of Surgery, 203*, 37-43.

Urvoy, M., Barkowsky, M., & Le Callet, P. (2013). How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors. *Annals of Telecommunications, 68*, 641-655.

van der Linden, D., Frese, M., & Sonnentag, S. (2003). The impact of mental fatigue on exploration in a complex computer task: Rigidity and loss of systematic strategies. *Human Factors, 45*(3), 483-495.

van Krevelen, D. W., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *The International Journal of Virtual Reality, 9*, 1-20.

Walker, M. P. (2008). Cognitive consequences of sleep and sleep loss. *Sleep Medicine, 9*, S1-S34.

Whitney, P., Hinson, J. M., Jackson, M. L., & Van Dongen, H. P. (2015). Feedback blunting: Total sleep deprivation impairs decision making that requires updating based on feedback. *Sleep, 38*(5), 745-754.

Wickens, C. D., Hutchins, S. D., Laux, L., & Sebok, A. (2015). The impact of sleep disruption on complex cognitive tasks: A meta-analysis. *Human Factors, 57*(6), 930-946.

Williamson, A. M., & Feyer, A. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine, 57*(10), 649-655.

Wimalasundera, S. (2006). Computer vision syndrome. *Galle Medical Journal, 11*(1), 25-29.

Yang, S., Schlieski, T., Selminsm B., Cooper, S. C., Doherty, R. A., Corriveau, P. J., & Sheedy, J. E. (2012). Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and Vision Science, 89*(7), 1068-1080).

Vergence-Accommodation Conflict and the Challenge of Adapting Virtual Reality and Augmented Reality Systems to Normal Binocular Functioning

by

LT Adam T. Biggs, MSC, USN, Daniel J. Geyer, & LCDR John L. Bradley, MSC, USN

Introduction

Many virtual reality (VR) and augmented reality (AR) issues arise as a result of common elements involved in nearly every form of training. For example, fatigue effects will transpire in due course no matter what training platform is used. However, some VR/AR problems are directly caused by the virtual platform. The vergence-accommodation conflict is a prime case of the platform itself causing problems. This conflict is difficult to solve because the problem involves the physiological factors of binocular vision rather than the design or comfort of the mixed reality environment. As such, the conflict underlies and contributes to other problems (e.g., fatigue and eye strain), thereby making it an important issue to understand—and, where possible, resolve through technique or technology.

The challenge begins with the basic functionality of binocular vision. Both eyes must simultaneously aim towards and acquire a particular target (von Noorden, 1996; von Noorden & Campos, 2002). Human perception depends upon successfully integrating the visual information acquired from both eyes. The result is stereopsis (Furukawa, 2010; Howard & Rogers, 1995; Levi, Knill, & Bavelier, 2015; Vishwanath, 2014), wherein the separate yet overlapping images from the two eyes are combined into a single representation. This process is almost unnoticeable between saccades, the rapid eye movements between fixation points which occur on a scale of milliseconds (Deubel & Schneider, 1996; Findlay & Walker, 1999; Munoz & Everling, 2004; Schall & Hanes, 1993). In fact, the entire procedure is so seamless, that humans are actually blind while the eyes are moving. This is saccadic suppression or saccadic masking, the process that blocks the blurred images produced when the eyes move rapidly between fixation points, exemplified by an individual looking into a mirror and shifting their gaze from one eye to the other without ever detecting actual eye movement (Breitmeyer & Ganz, 1976; Bridgeman, Hendry, & Stark, 1975; Burr, Morrone, & Ross, 1994; Tabernero & Artal, 2014). This prospect is particularly impressive given that, while reading, a human will make approximately four fixations per second (Henderson & Reisberg, 2013), and each saccade requires approximately 20 milliseconds to 100 milliseconds based upon the length of the eye movement (Purves et al., 2001). So, a reader is actually blind during about 10% of the time spent reading without ever noticing the disruption—thanks to the seamless integration of stereopsis in binocular vision and human perception.

VR/AR platforms can disrupt this otherwise seamless process due to the vergence-accommodation conflict. Vergence describes the process by which your eyes either converge on an item nearby, or diverge on an item far away (Masson, Busettini, & Miles, 1997; Mays, 1984; Purves et al., 2001; Toates, 1974). This process allows for depth perception and differs significantly from other types of eye movements. For example, trying to see something to your left requires both eyes to move left—a conjugate eye movement. Vergence movements require the left eye to move right and the right eye to move left when focusing upon something near. The reverse (left eye moving left and right eye moving right) is true for diverging movements. These eye movements align the fovea of each eye with the corresponding target and aid with depth perception. Accommodation describes the process by which the lens adjusts to create clear vision for either a far or near point (Helmhotlz, 1909; Leigh & Zee, 2015; Toates, 1972). Acute depth perception require both components working together, although they are distinct physiological processes as one involves the physical eye movement (vergence) and one involves a reshaping of the lens within the eye (accommodation).

Under normal circumstances, vergence and accommodation are naturally coupled to create the best possible depth perception. The eyes converge to focus upon something close while tightening muscles to allow the pliable lens to become more rounded. The eyes diverge to focus upon something far away while relaxing the ciliary muscles to keep the lens less rounded. These procedures represent the natural approach, or the A-A/B-B coupling that occurs during normal vision. In a VR/AR viewing environment, however, the artificial environment forcibly decouples these processes. A virtual scenario creates images with perceivable depth based on the relative sizes of each item in the display, but these images are still being displayed on screens that often are mere inches away from the physical eye. Essentially, vergence distance varies according to the needs of the virtual scenario, whereas accommodation distance remains uniform throughout. This disruption changes a natural A-A/B-B coupling to a less natural A-A/B-A coupling, where the lens is always trying to remain rounded rather than adapting to distance along with the normally occurring eye movements.

Unfortunately, most VR or AR platforms use head-mounted displays (HMDs) to create the virtual world, which imposes the vergence-accommodation problem described here onto almost all VR or AR scenarios. This prevalence is important because the vergence-accommodation conflict can cause or exacerbate other simulator sickness symptoms as well as visual fatigue. Moreover, this conflict creates an all-round unpleasant experience for the user. If these systems are going to be used for training simulations, then it is necessary to understand the risks involved, the severity of user symptoms, and the technological avenues to solving these issues. This review will consider the larger issue of vergence-accommodation conflict with particular attentiveness towards these issues as they pertain to training issues in naval aviation. We will begin with an evaluation of the symptoms which arise from prolonged vergence-accommodation conflict.

Symptoms Occurring due to the Vergence-Accommodation Conflict

Vergence-accommodation conflict symptoms are similar in several ways to simulator sickness symptoms (cf. Champney, Stanney, Hash, Malone, Kennedy, & Compton, 2007; Johnson, 2005; Kennedy & Fowlkes, 1992; Moss & Muth, 2011), and indeed the conflict plays a role in the presence and development of several symptoms, though the full extent is yet to be identified. Still, although some aspects are more indicative of simulator sickness in general, some aspects are more directly attributable to vergence-accommodation conflict. For example, visual fatigue may develop from prolonged use of HMDs (Bando, Iijima, & Yano, 2012; Hoffman, Girshick, Akeley, & Banks, 2008; Lambooij, Fortuin, Heynderickx, & IJsselsteijn, 2009; Reichelt, Häussler, Fütterer, & Leister, 2010). Visual fatigue is conceptually distinct from other forms of fatigue covered earlier in the larger literature review, including cognitive fatigue (cf. Grier, Warm, Dember, Matthews, Galinsky, Szalma, et al., 2003; Hockey, 2011; Langner & Eickhoff, 2013; Pattyn, Neyt, Henderickx, & Soetens, 2008), or sleep deprivation (Baldwin & Daughtery, 2004; Pilcher & Huffcutt, 1996; Samkoff & Jacques, 1991). Visual fatigue refers to an objective decrease in performance of the visual system, whereas visual discomfort—occasionally also referenced as visual fatigue in the literature—represents subjective discomfort involving the eyes (Lambooij et al., 2009). Both objective and subjective measures of visual fatigue and discomfort are prominent following prolonged vergence-accommodation conflict, and both measures are likely to have a negative impact on any task-relevant performance metrics.

Some degree of visual fatigue and visual discomfort is highly likely due to the vergence-accommodation conflict, but these symptoms are mitigated or exacerbated by the given circumstances. This evidence comes from research into both virtual systems and normal vision.

For example, the primary stimulus for accommodation is retinal blur, whereas the primary stimulus for vergence is binocular disparity (Alpern & Ellen, 1956; Maxwell, Tong, & Schor, 2010). This functional differentiation allows researchers to isolate various aspects of the conflict by adapting experimental stimuli to specific visual functions. Granted, the vergence and accommodation functions are inherently related, though the vergence system does appear able to adapt to prolonged exposure (Lee, Granger-Donetti, Chang, & Alvarez, 2009; Kim, Vicci, Han, & Alvarez, 2011; Schor, 1979; Vienne, Sorin, Blondé, Huynh-Thu, & Mamassian, 2014). Certain visual anomalies, such as convergence insufficiency (CI), can lead to reduced vergence adaptation with prolonged convergence accommodation (Sreenivasan & Bobier, 2014). Although developmental visual anomalies are uncommon (CI has a prevalence of 5% in school aged children; Cooper, Schulman, & Jamal, 2013; Scheiman & Wick, 2002), scientific research can compare such anomalies to normal visual functions. This approach allows for a better understanding of the physiology involved.

One important factor involves the rate of change in the vergence-accommodation conflict. Specifically, stimuli can be paired to simulate natural viewing—that is, changes in the stimulus cause vergence and accommodation to change together—or stimuli can be paired to simulate a virtual environment—that is, changes in the stimulus cause a change in vergence, but not accommodation as any image remains a fixed distance away from the eye. Results have indicated that visual discomfort is more extreme as the changes are more extreme (Kim, Kane, & Banks, 2014). This effect can be worse for a simulated environment, although the difference again applies more to the extreme scenarios. Disparity and distance are additional factors that can impact visual fatigue and discomfort symptoms. Specifically, greater symptoms will occur for longer distances with uncrossed disparities or shorter distances with crossed disparities

(Shibata, Kim, Hoffman, & Banks, 2011). Visual fatigue and discomfort also increase for longer viewing distances (Banks, Kim, & Shibata, 2013). A prime underlying factor thus appears to be the extent of change involved with switching focus from one stimulus to another. Put another way, rapid changes exacerbate vergence-accommodation conflict, and the aggravated symptoms are only further exacerbated by a virtual environment.

Exercises, Risk Factors, and Individual Differences

Vergence-accommodation is a common problem when adapting human visual functioning to virtual systems, yet there are numerous other factors that can impact the extent of the problem for better or worse. One example involves eye exercises, which have shown promise in improving vergence and/or accommodation. CI symptoms have been reduced using orthoptic exercises (Horwood; Toor, & Riddell, 2014; Serna, Rogers, McGregor, Golden, Bremer, & Rogers, 2011). The vergence improvements have been significant in experiments, although the beneficial effects for accommodation have been far fewer. Still, vision training offers an interesting take on the adaptability of oculomotor functions.

Another study demonstrated that binocular function could be improved even among university athletes (Zwierko, Puchalska-Niedbał, Krzepota, Markiewicz, Woźniak, & Lubiński, 2015), which suggests that vision training may be successful among healthy participants. An important note here is that vision training could refer to either oculomotor training to enhance binocular vision or stroboscopic training to enhance attentional and visual processing. Stroboscopic training involves specialized eyewear that limits the visual field by turning the lenses from clear to opaque and back at an adjustable interval. This alternative training has been shown to be successful in numerous healthy athletic participants, including college baseball players (Clark, Ellis, Bench, Khoury, & Graman, 2012), college softball players, (Appelbaum, Lu, Khanna, & Detwiler, 2016), and professional athletes from the National Hockey League (Mitroff, Friesen, Bennett, Yoo, & Reichow, 2013). Stroboscopic training appears to enhance anticipatory timing (Smith & Mitroff, 2012) and visual cognition (Appelbaum, Schroeder, Cain, & Mitroff, 2011) rather than binocular vision or oculomotor functions. As such, there is an important distinction to draw in that vision training could apply either to the movement of the eyes or the processing of any accumulated visual information. Vergence issues could be more directly addressed by vision training as it pertains to eye movement, although an interesting but untested idea would involve how various forms of vision training might interact.

Symptoms and related issues of visual discomfort/fatigue are well-known, and there does appear to be some avenues for adaptation or training to overcome or mitigate the problems. However, a more important operational point comes in defining a quitting threshold—when should an operator stop training in a virtual environment? Session duration is a significant concern because visual fatigue becomes greater over time and has a larger impact on binocular vision as session duration increases (Suzuki, Onda, Katada, Ino, & Ifukube, 2004). Individual session length will depend upon the type of activity and training involved. Once fusion difficulties are experienced though, the viewer is encouraged to cease any further action with the virtual scenario (Ukai & Howarth, 2008). Still, this suggestion may prove too simplistic for training requirements. Multiple unknown factors could impact proscribed session length, including individual differences and long-term effects from repeated use. These possible factors represent significant avenues for future research, although more basic science research into visual fatigue mechanisms could likewise prove useful in delineating safe and repeatable session durations.

Age is another notable factor in considering stopping rules. Convergence function does not decrease with age, whereas accommodative functioning does (Schaeffel, Wilhelm, & Zrenner, 1993; Mordi & Ciuffreda, 1998; 2004; Ukai & Howarth, 2008). Older individuals are much more likely to have a reduced accommodative range, and many vision studies will exclude participants on the basis of presbyopia. It is an interesting consideration that increased age might actually reduce vergence-accommodation conflict because accommodation functions are reduced enough to avoid extreme levels of conflict.

Vergence-accommodation conflict issues may benefit significantly from future research into individual differences and basic science research into the mechanisms of human vision. Even so, it is unclear precisely what benefits or tangible impact these future research initiatives might have upon improving training scenarios for naval aviation. More immediate initiatives could involve improvements in technology or the scenarios themselves. This approach could reduce visual fatigue through design rather than individual differences, training, or far-future technology pursuits. For example, studies have shown that there are no differences between certain e-readers and paper with ink when it comes to visual fatigue (Benedetto, Drai-Zerbib, Pedrotti, Tissier, & Baccino, 2013). Certain existing technology may also cause reading behaviors to be very similar to paper/ink (Siegenthaler, Wurtz, Bergamin, & Groner, 2011). With these promising possibilities in mind, the discussion will move to how scenario design and existing technology could improve vergence-accommodation issues in virtual environments.

Technological Solutions to Vergence-Accommodation Conflict: Short-term

At the moment, experimental design and technological improvement are the most promising avenues to reducing vergence-accommodation conflict symptoms while using VR/AR devices. The vision science literature is well-aware of this problem, and although it is difficult to provide definite rules for widespread application, there are several guidelines to consider when designing VR/AR scenarios. These approaches will not eliminate symptoms arising from vergence-accommodation conflict. However, they can reduce the severity of symptoms and allow individuals to engage in prolonged session durations. Moreover, these changes can be implemented right now into the existing VR/AR systems as they alter scenario development rather than require new technology. Here are several guidelines for reducing vergence-accommodation conflict through scenario design:

- Limit extreme changes in viewed object distance because vision-related symptoms become more severe with extreme changes (Kim et al., 2014). Allow distance changes to be slower, and the symptoms will be reduced.
- When large disparities are required, there are conflicting suggestions. Content creators can either interweave sections with smaller disparities into the larger disparities (Mendiburu, 2009), or the large disparities as slowly as possible (Kim et al., 2014). It is currently unclear which approach is better, although either approach could alleviate any vision-related symptoms.
- Use long viewing distances between the eye and the screen when possible because there should be little vergence-accommodation conflict beyond 1m (Hoffman et al., 2008; Watt, Akeley, Ernst, & Banks, 2008)
- Maximize the reliability of other depth cues in the stimulus beyond focus cue (Hoffman et al., 2008). Focus cues will impact vergence-accommodation conflict issues based upon their relative reliability.

- Make existing conflicts less salient (Hoffman et al., 2008). Stacking objects at varying depths so that they overlap can exacerbate issues. So, symptoms can be reduced by minimizing salient depth differences by minimizing object overlap.
- Do not fully depend upon adaptation to the scenario. Although the visual discomfort from optical correction is temporary (Henson & North, 1980; North & Henson, 1985), stereoscopic displays and VR/AR scenarios prevent successful adaptation in a sufficient time frame.
- End the VR/AR scenario, or at least allow for breaks, when the viewer experiences fusion difficulties (Ukai & Howarth, 2008).

For AR platforms, there are similarly low-tech solutions to reducing vergence-accommodation conflict. The Microsoft HoloLens is easily one of the most advanced AR platforms at the moment. Their developer resources provide information to help improve hologram stability, and these developer notes specifically reference convergence and accommodation (Microsoft, 2016). According to these resources, their product users will accommodate to a distance of 2.00m to maintain a clear image. Their straightforward recommendation is therefore to place hologram content as close to the 2.00m distance as possible so as to avoid any vergence-accommodation conflict. When this distance is not possible, they recommend an optimal placement distance of 1.25m and 5.00m to avoid any conflict-related issues. Even in AR systems, this approach can reduce vergence-accommodation conflict and the associated symptoms of visual fatigue simply by adapting the limitations of the technology and user tendencies to the desired scenario.

Unfortunately, these limitations may prevent certain uses for naval aviation training. For example, aviation training scenarios may not be able to fully utilize current HoloLens system for a wide variety of training scenarios if the hologram items need to sit between 1.25m and 5.00m away. Cockpit controls will obviously need to be within 1m of the aviator. This limitation does not preclude its use, although it may limit session duration times to prevent extreme levels of fatigue and discomfort on the trainee.

Conversely, AR systems might be precisely the optimal current solution to vergence-accommodation conflict—depending on how the AR is integrated into the scenario. Lambooij et al. (2009) laid out numerous factors that impact visual fatigue and visual discomfort. One particular factor included that retinal disparities beyond 1° were assumed to cause discomfort. In short, discomfort can be minimized if eye focus remains at or close to infinity. Precise methodologies have already been developed in how to alter stereo content to achieve this goal for certain scenarios (Didyk, Ritschel, Eisemannn, Myszkowski, Seidel, & Matusik, 2012; Lang, Hornung, Wang, Poulakos, Smolic, & Gross, 2010; Liu, Huang, Chang, Lee, Liang, & Chuang, 2011; Shamir & Sorkine, 2009). For commercial VR/AR purposes, dynamic and interactive content relies upon nearby interactions, which violates the constraints and can cause problems (Celikcan, Cimen, Kevinc, & Capin, 2013; Rolland, Krueger, & Goon, 1999; Sherstyuk & State, 2010). For aviation training, distant focal points are inherent when examining the skies around the aircraft. The proposed solution would require a realistic cockpit for interactions, although the surrounding skies—and their content—could be simulated with an AR environment presented via HMD. The trainee would experience an immersive and realistic simulation with the vergence-accommodation conflict likely to arise while interacting with stimuli very close to him or her. These symptoms could be mitigated for some of the more distant content. An AR solution to training issues would then be rather direct—make the cockpit real, and simulate the sky.

Technological Solutions to Vergence-Accommodation Conflict: Long-term

Thus far, the proposed solutions represent the short-term or more immediate options. VR/AR technology is currently evolving, and there are some future technological developments in the works which may provide more complete and satisfactory answers to vergence-accommodation conflict. VR systems currently depend upon stereoscopic displays to create 3-D images from mostly flat screens. Significant effort has gone into improving screen resolution and enhancing image quality so that the virtual environment appears more realistic. Although these efforts have taken virtual reality from the blocky images of the mid-1990's to the flurry of realistic products now available, stereoscopic displays are not a long-term solution to the vergence-accommodation conflict that underlies many symptoms of visual fatigue and visual discomfort. The precise solution remains one of large interest, particularly commercial interest, as new products gain increased demand. Here we will discuss three promising technological approaches with varying degrees of readiness: parallax barriers, pinlight arrays, and light fields.

Parallax describes the effect when the apparent position (or direction) of an object differs when viewed from different lines of sight. A parallax barrier is a device that allows a stereoscopic image to be displayed without 3-D glasses by placing the device in front of the image source. These screens or stacks can modulate light rays spatially and angularly as the rays pass through (Lanman, Hirsch, Kim, & Raskar, 2010a; 2010b). Light rays can then be integrated as a sum of perceived light rays rather than individual rays. This procedure can synthesize perceived light rays at particular locations on the pupil, thereby allowing the eyes to accommodate naturally to the desired depth of perceived virtual objects. Accommodation issues are reduced, although there are additional problems introduced. First, computational requirements are significant as some designs require several minutes to render a single scene. Second, the design requires no relative motion between the display and the pupil, which could be a significant issue for an aviation training scenario. The problem can be circumvented with eyetracking, although that solution prompts further computational and hardware requirements.

Pinlight arrays represent a similar design to a parallax barrier, albeit with a slightly different operating principle (Maimone, Lanman, Rathinavel, Keller, Luebke, & Fuchs, 2014). A pinlight HMD projects through a barrier onto the eye through a dense array of point light sources. These various lights operate with minimal overlap to form a dense grid. While the projected result is out of focus on the pupil, a sharp image forms at the back of the retina. Light rays are not truly converging then, but rather counting upon further refraction by the pupil to form the desired image. The design is innovative, though there are still complications. Multiple projectors are required for a sufficient field of view, and like parallax barrier, the design assumes no motion. Further design adaptations are thus required. Eye tracking and creativity in the manner of pinlight projection (or pixel sharing) could address some issues, but the design must be adapted for increased motion to be useful for naval aviation training.

One particularly exciting possibility involves light fields. Light field rendering employs radiance as a function of position and direction in a space free of occluders (Levoy & Hanrahan, 1996). Creating a light field requires many difference 2-D images to create slices of a 4-D light field, and the eventual representation samples from these different images to create the field itself. The result is a full light field that allows the observer to naturally focus upon any point in the visual display. Stereoscopic displays project an image from a flat screen and utilize design tricks to minimize the visual discomfort. Viewers are still trying to interpret light coming from a screen close to their eyes. Light fields project light directly onto the eye in a way much closer to how the human visual system processes light in a natural setting. This approach requires recreating a 4-D space by adding a temporal component to length, width, and depth, which then allows for calculations of vector and speed. Generating the field then becomes a mathematics calculation and a computer processing challenge as much as anything else. A light field is then a complicated reconstruction of many 2-D images that creates a new environment by incorporating depth and time. Put another way, the 4-D field could be described as such: 1-D is a line (length), 2-D is a circle (add width), 3-D is a sphere (add depth), and 4-D is a rolling sphere (add time to allow for changes).

A light field solution can address vergence-accommodation conflict because the viewer can focus naturally. Other image aspects can blur just as portions of the real world blur when the viewer is focused upon something. The current challenge is that this approach cannot be fully utilized with the existing technology. Recent advancements have utilized light fields to create a wearable VR display with high image resolution and focus cues (Huang, Chen, & Wetzstein, 2015). Similar advances offer potential methods to incorporate light fields into augmented reality (Lee, Jang, Moon, Cho, & Lee, 2016). Still, light fields offer an intriguing possible solution to vergence-accommodation conflict, whereas stereoscopic displays may be a dead end—technique and technology might minimize the conflict without being able to fully eliminate it. Either possibility is speculation at the moment. Each approach requires new technology and new research for further investigations, which raises the real issue about who, when, and where funding investments should be made into these technologies.

Commercial investments into these new technologies are already quite substantial. A single startup, called Magic Leap, received a $542,000,000 investment from Google to pursue new approaches to creating 3-D imagery (Metz, 2015). The funding would reach $794,000,000 total investment as it entered the pilot production phase of its augmented reality device (Tilley, 2016). It is also worth noting that this investment involves one company—not the only company doing so. Microsoft is also heavily invested into new mixed reality products such as its HoloLens. By comparison, the entire Department of the Navy Science and Technology (S&T) budget requested approximately $2 billion in fiscal year 2016 (Winter, 2015).

Scope is thus a significant consideration. Although the technology would be commercial, nearly half of the annual Navy S&T budget would be necessary to match the commercial investment. Moreover, the commercial systems could be adapted for naval training purposes. The greatest disadvantage is then the uncertain timeline for these new technologies as commercial companies control when new products are released. Still, this burden is hardly a difficult one to bear as the companies will try to release new and better technologies as quickly as possible to see returns on investment. Another notable disadvantage could be our control and access in designing new scenarios specifically for aviation. This drawback seems significant, although the platforms are increasingly developed for many content creators rather than a single source; and, it is hardly uncommon for the military to work with commercial entities when designing hardware or software for military purposes. As such, it may be in the Navy’s best interest to use—rather than sponsor—commercial development of new technology to reduce any vergence-accommodation conflict issues.

Conclusion

Vergence-accommodation conflict is a significant issue for VR/AR systems. The conflict can cause visual fatigue and visual discomfort from prolonged use, which contributes to significant symptomology that eventually impair any benefits associated with using this technology. With standard HMDs for VR, the problem is virtually unavoidable because the screen sits too close to the eyes. However, there are technological tweaks and scenario development approaches that help circumvent some issues. The existing tweaks are short-term solutions, although long-term commercial development and investment in this area is already substantial. Despite the lack of a definite timetable for solving any vergence-accommodation conflict issues in VR/AR systems, it would be difficult to justify matching the current investments from the commercial section. Naval research funding could be better utilized by exploring training applications with off-the-shelf technology that continues to improve.

References

Alpern, M., & Ellen, P. (1956). A quantitative analysis of the horizontal movements of the eyes in the experiment of Johannes Mueller: I. Method and results. *American Journal of Ophthalmology*, *42*(4), 289-296.

Appelbaum, L. G., Lu, Y., Khanna, R., & Detwiler, K. R. (2016). The Effects of Sports Vision Training on Sensorimotor Abilities in Collegiate Softball Athletes. *Athletic Training and Sports Health Care*, *8*, 154-163.

Appelbaum, L. G., Schroeder, J. E., Cain, M. S., & Mitroff, S. R. (2011). Improved visual cognition through stroboscopic training. *Frontiers in Psychology*, *2*, 276.

Baldwin Jr, D. C., & Daugherty, S. R. (2004). Sleep deprivation and fatigue in residency training: results of a national survey of first-and second-year residents. *Sleep*, *27*(2), 217-223.

Bando, T., Iijima, A., & Yano, S. (2012). Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review. *Displays*, *33*(2), 76-83.

Banks, M. S., Kim, J., & Shibata, T. (2013, May). Insight into vergence/accommodation mismatch. In *SPIE Defense, Security, and Sensing* (pp. 873509-873509). International Society for Optics and Photonics.

Benedetto, S., Drai-Zerbib, V., Pedrotti, M., Tissier, G., & Baccino, T. (2013). E-readers and visual fatigue. *PloS one*, *8*(12), e83676.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, *83*(1), 1-36.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719-722.

Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, *371*(6497), 511-513.

Celikcan, U., Cimen, G., Kevinc, E. B., & Capin, T. (2013). Attention-aware disparity control in interactive environments. *The Visual Computer*, *29*(6-8), 685-694.

Champney, R. K., Stanney, K. M., Hash, P. A. K., Malone, L. C., Kennedy, R. S., & Compton, D. E. (2007). Recovery from virtual environment exposure: expected time course of symptoms and potential readaptation strategies. *Human Factors*, *49*(3), 491-506.

Clark, J. F., Ellis, J. K., Bench, J., Khoury, J., & Graman, P. (2012). High-performance vision training improves batting statistics for University of Cincinnati baseball players. *PloS One*, *7*(1), e29109.

Cooper, J., Schulman, E., & Jamal, N., 2013. Convergence insufficiency – A major review. Optometry online. American Optometric Association. 16-1-2013.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827-1837.

Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., Seidel, H. P., & Matusik, W. (2012). A luminance-contrast-aware disparity model and applications. *ACM Transactions on Graphics (TOG)*, *31*(6), 184.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(04), 661-674.

Furukawa, Y., & Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, *32*(8), 1362-1376.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 349-359.

Henderson, J. M., & Reisberg, D. (2013). Eye movements. *The Oxford handbook of cognitive psychology*, 69-82.

Henson, D. B., & North, R. (1980). Adaptation to prism-induced heterophoria. American Journal of Optometry and Physiological Optics, 57(3), 129–137.

Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*, 167-187.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33: 1-30.

Horwood, A. M., Toor, S. S., & Riddell, P. M. (2014). Change in convergence and accommodation after two weeks of eye exercises in typical young adults. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *18*(2), 162-168.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford University Press, USA.

Huang, F. C., Chen, K., & Wetzstein, G. (2015). The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Transactions on Graphics (TOG)*, *34*(4), 60.

Johnson, D. M. (2005). *Introduction to and review of simulator sickness research*. (ARI Report No. 1832). U.S. Army Research Institute Field Unit, Fort Rucker, AL.

Kennedy, R. S., & Fowlkes, J. E. (1992). Simulator sickness is polygenic and polysymptomatic: Implications for research. *International Journal of Aviation Psychology*, *2*(1), 23-38.

Kim, J., Kane, D., & Banks, M. S. (2014). The rate of change of vergence–accommodation conflict affects visual discomfort. *Vision Research*, *105*, 159-165.

Kim, E. H., Vicci, V. R., Han, S. J., & Alvarez, T. L. (2011). Sustained fixation induced changes in phoria and convergence peak velocity. *PloS One*, *6(6)*, e20883.

Kramida, G. (2016). Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE transactions on visualization and computer graphics*, *22*(7), 1912-1931.

Lambooij, M., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, *53*(3), 30201-1.

Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., & Gross, M. (2010). Nonlinear disparity mapping for stereoscopic 3D. *ACM Transactions on Graphics (TOG)*, *29*(4), 75.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870-900.

Lanman, D., Hirsch, M., Kim, Y., & Raskar, R. (2010a, July). Content-adaptive parallax barriers for automultiscopic 3D display. In *ACM SIGGRAPH 2010 Posters* (p. 59). ACM.

Lanman, D., Hirsch, M., Kim, Y., & Raskar, R. (2010b). Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Transactions on Graphics (TOG)*, *29*(6), 163.

Lee, Y. Y., Granger-Donetti, B., Chang, C., & Alvarez, T. L. (2009). Sustained convergence induced changes in phoria and divergence dynamics. *Vision Research, 49(24)*, 2960–2972.

Lee, S., Jang, C., Moon, S., Cho, J., & Lee, B. (2016). Additive light field displays: realization of augmented reality with holographic optical elements. *ACM Transactions on Graphics (TOG)*, *35*(4), 60.

Leigh, R. J., & Zee, D. S. (2015). *The neurology of eye movements* (Vol. 90). Oxford University Press, USA.

Levi, D. M., Knill, D. C., & Bavelier, D. (2015). Stereopsis and amblyopia: a mini-review. *Vision Research*, *114*, 17-30.

Levoy, M., & Hanrahan, P. (1996, August). Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 31-42). ACM.

Liu, C. W., Huang, T. H., Chang, M. H., Lee, K. Y., Liang, C. K., & Chuang, Y. Y. (2011, November). 3d cinematography principles and their applications to stereoscopic media processing. In *Proceedings of the 19th ACM international conference on Multimedia* (pp. 253-262). ACM.

Maimone, A., Lanman, D., Rathinavel, K., Keller, K., Luebke, D., & Fuchs, H. (2014, July). Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies* (p. 20). ACM.

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*(6648), 283-286.

Maxwell, J., Tong, J., & Schor, C. M. (2010). The first and second order dynamics of accommodative convergence and disparity convergence. *Vision Research*, *50*(17), 1728-1739.

Mays, L. E. (1984). Neural control of vergence eye movements: convergence and divergence neurons in midbrain. *Journal of Neurophysiology, 51*(5), 1091-1108.

Mendiburu, B. (2009). 3D movie making: stereoscopic digital cinema from script to screen. Oxford, UK: Focal Press, Elsevier.

Metz, R. (2015, March/April). Magic Leap: A startup is betting more than half a billion dollars that it will dazzle you with its approach to creating 3-D imagery. *MIT Technology Review*.

Microsoft. (2016, December 28). Hologram stability. Retrieved from: https://developer.microsoft.com/en-us/windows/holographic/hologram_stability

Mitroff, S. R., Friesen, P., Bennett, D., Yoo, H., & Reichow, A. W. (2013). Enhancing ice hockey skills through stroboscopic visual training: a pilot study. *Athletic Training and Sports Health Care*, *5*(6), 261-264.

Mordi, J. A., & Ciuffreda, K. J. (1998). Static aspects of accommodation: age and presbyopia. *Vision Research*, *38*(11), 1643-1653.

Mordi, J. A., & Ciuffreda, K. J. (2004). Dynamic aspects of accommodation: age and presbyopia. *Vision Research*, *44*(6), 591-601.

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors, 53*(308), 308-319.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience, 5*(3), 218-228.

North, R., & Henson, D. B. (1985). Adaptation to lens-induced heterophorias. *American Journal of Optometry and Physiological Optics, 62(11)*, 774–780.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue?. *Physiology & Behavior, 93*(1), 369-378.

Pilcher, J. J., & Huffcutt, A. J. (1996). Effects of sleep deprivation on performance: a meta-analysis. *Sleep: Journal of Sleep Research & Sleep Medicine*.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A. S., McNamara, J. O., & Williams, S. M. (2001). Types of eye movements and their functions. Accessed from: http://www.ncbi.nlm.nih.gov/books/NBK10991/.

Reichelt, S., Häussler, R., Fütterer, G., & Leister, N. (2010, April). Depth cues in human visual perception and their realization in 3D displays. In *SPIE Defense, Security, and Sensing* (pp. 76900B-76900B). International Society for Optics and Photonics.

Rolland, J. P., Krueger, M. W., & Goon, A. A. (1999, May). Dynamic focusing in head-mounted displays. In *Electronic Imaging'99* (pp. 463-470). International Society for Optics and Photonics.

Samkoff, J. S., & Jacques, C. H. (1991). A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Academic Medicine*, *66*(11), 687-93.

Schaeffel, F., Wilhelm, H., & Zrenner, E. (1993). Inter-individual variability in the dynamics of natural accommodation in humans: relation to age and refractive errors. *The Journal of Physiology*, *461*, 301.

Schall, J. D., & Hanes, D. P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, *366*(6454), 467-469.

Scheiman, M., & Wick, B. (2002). *Clinical management of binocular vision: Heterophoric, accommodative and eye movement disorders.* Philadelphia: Lippincott Williams & Wilkins.

Schor, C. M. (1979). The relationship between fusional vergence eye movements and fixation disparity. *Vision Research, 19(12),* 1359–1367.

Serna, A., Rogers, D. L., McGregor, M. L., Golden, R. P., Bremer, D. L., & Rogers, G. L. (2011). Treatment of symptomatic convergence insufficiency with a home-based computer orthoptic exercise program. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *15*(2), 140-143.

Shamir, A., & Sorkine, O. (2009, December). Visual media retargeting. In *ACM SIGGRAPH ASIA 2009 Courses* (p. 11). ACM.

Sherstyuk, A., & State, A. (2010, November). Dynamic eye convergence for head-mounted displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (pp. 43-46). ACM.

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011, February). Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict. In *IS&T/SPIE Electronic Imaging* (pp. 78630P-78630P). International Society for Optics and Photonics.

Siegenthaler, E., Wurtz, P., Bergamin, P., & Groner, R. (2011). Comparing reading processes on e-ink displays and print. *Displays*, *32*(5), 268-273.

Smith, T. Q., & Mitroff, S. R. (2012). Stroboscopic training enhances anticipatory timing. *International Journal of Exercise Science*, *5*(4), 344-353.

Sreenivasan, V., & Bobier, W. R. (2014). Reduced vergence adaptation is associated with a prolonged output of convergence accommodation in convergence insufficiency. *Vision Research*, *100*, 99-104.

Suzuki, Y., Onda, Y., Katada, S., Ino, S., & Ifukube, T. (2004). Effects of an eyeglass-free 3-D display on the human visual system. *Japanese Journal of Ophthalmology*, *48*(1), 1-6.

Tabernero, J., & Artal, P. (2014). Lens oscillations in the human eye. Implications for post-saccadic suppression of vision. *PloS One*, *9*(4), e95764.

Tilley, A. (2016, February 2). Augmented reality startup Magic Leap closes $794 million investment as it enters pilot production phase. Retrieved from: http://www.forbes.com/sites/aarontilley/2016/02/02/augmented-reality-startup-magic-leap-closes-794-million-investment-as-it-enters-pilot-production-phase/#524ecd6076e1

Structure: Last, F. M. (Year, Month Date Published). Article title. Retrieved from URL

Toates, F. M. (1972). Accommodation function of the human eye. *Physiological Reviews*, *52*(4), 828-863.

Toates, F. M. (1974). Vergence eye movements. *Documenta Ophthalmologica*, *37*(1), 153-214.

Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, *29*(2), 106-116.

Vienne, C., Sorin, L., Blondé, L., Huynh-Thu, Q., & Mamassian, P. (2014). Effect of the accommodation-vergence conflict on vergence eye movements. *Vision Research*, *100*, 124-133.

Vishwanath, D. (2014). Toward a new theory of stereopsis. *Psychological Review*, *121*(2), 151-178.

von Helmholtz, H. H. (1909). Handbuch der Physiologishen Optik. In Southall J. P. C. (Translator), Helmholtz's treatise on physiological optics (pp. 143-172) 1962. New York: Dover.

von Noorden, G.K. (1996) *Binocular vision and ocular motility: Theory and management of strabismus.* 5th ed. C.V. Mosby: St. Louis.

von Noorden, G. K., & Campos, E. C. (2002). *Binocular vision and ocular motility*. Mosby.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5(10):7*, 834–862,

Winter, M. W. (2015, March 26). Statement of Rear Admirial Mathias W. Winter, United States Navy Chief of Naval Research before the Emerging Threats and Capabilities Subcommittee of the House Armed Services Committee on the Fiscal Year 2016 Budget Request.

Zwierko, T., Puchalska-Niedbał, L., Krzepota, J., Markiewicz, M., Woźniak, J., & Lubiński, W. (2015). The Effects of Sports Vision Training on Binocular Vision Function in Female University Athletes. *Journal of human kinetics, 49*(1), 287-296.