Cognitive Performance Impact of Augmented Reality for Network Operations Tasks

Steve Beitzel, Josiah Dykstra, Sean Huver, Michael Kaplan, Michael Loushine and Jason Youzwak

Abstract We examine whether the use of Augmented Reality (AR) may aid the performance of network security operators (NSOs) performing time critical tasks. While attempting to achieve some primary goal, NSOs must continuously monitor and respond to a wide range of ancillary events which may impact mission performance. Responding to these additional events in a time sensitive manner introduces significant stress and impacts primary task performance. We conducted experiments with the Epson Moverio and Vuzix m100 head-mounted displays. Test subjects performed a simulated NSO exercise with and without AR glasses while simultaneously responding to randomly generated ancillary events. Test subjects using AR reported less overall mental demand, performed the primary task more quickly, and more often successfully responded to ancillary events within a required time. We find that AR devices are a promising aid for maintaining focus in primary network operations tasks while reducing overall stress load.

Keywords Augmented reality · Network security · Cognitive workload · CTF · Situational awareness

S. Beitzel · S. Huver · M. Kaplan · M. Loushine · J. Youzwak (✉)
Applied Communication Sciences, Basking Ridge, NJ, USA
e-mail: jyouzwak@appcomsci.com

S. Beitzel
e-mail: sbeitzel@appcomsci.com

S. Huver
e-mail: shuver@appcomsci.com

M. Kaplan
e-mail: mkaplan@appcomsci.com

M. Loushine
e-mail: mloushine@appcomsci.com

J. Dykstra
Laboratory for Telecommunication Sciences, College Park, MD, USA
e-mail: jdykstra@LTSnet.net

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 501,
DOI 10.1007/978-3-319-41932-9_12

1 Introduction

Computer network operations occur at the intersection of technology and human operators. While much research and development have been focused on the tools for these tasks, insufficient attention has been given to the usability, utility, and satisfaction with these tools, and specifically to overcoming the fatigue and cognitive workload associated with the tasks.

In determining whether a technology aids a user in performing a task, industrial psychologists typically use a combination of objective and subjective means to form a conclusion. Objective testing includes task-specific quantitative testing performance, and sometimes also physiological data, such as heart rate, perspiration, eye movement, and even electroencephalogram (EEG) testing to monitor electrical activity in the brain. Subjective testing is often evaluated with the NASA Task Load Index (TLX) survey [1], which is a numerically ranked self-assessment of a user's mental demand, physical demand, temporal demand, effort, performance, and frustration during a task. The following study uses a simulated network operations mission to investigate whether or not augmented reality glasses help to improve operator performance on time-sensitive tasks while experiencing a demanding cognitive load. We show (albeit with a small sample size of seven users) a reasonable possibility that augmented reality (AR) devices may reduce the cognitive load of operators, while also decreasing operator reaction time for time-sensitive events.

Our simulated operation measures the total time to completion, the number of time-sensitive events of interest (EOI) successfully mitigated within the time limit, as well as the average response time to an EOI. The operation is performed both with and without augmented reality glasses, and after a non-timed walkthrough to gain familiarity with the operational environment.

The NASA TLX is the most commonly used subjective ranking method for mental workload, and has been cited in over 4400 task performance studies. We note that our control group may have some cultural biases as they were homogenous in demographics (Caucasian males, aged 34–60).

Augmented reality is a combination of virtual reality and the actual world. It uses components such as camera, video and microphone to enhance the user's interaction with an existing environment. Users are able to interact with both virtual and

Fig. 1 A test subject using augmented reality (AR) glasses to perform a network "capture the flag" scenario. The AR glasses aid in helping to identify and classify ancillary events that are of interest to the operator

actual contents, and can distinguish between the two. By contrast, Virtual Reality (VR) creates a complete world in which the user can interact (Fig. 1).

Our experiment was limited by commercially-available AR hardware. There are several models of AR glasses coming to the market in the next 6–12 months that appear to have better capabilities than the two AR devices used in this study (Vuzix m100 and Epson Moverio). The Microsoft Hololens and the Osterhout Design Group R7 glasses both promise impressive hardware and more sophisticated SDKs than what is currently available.

2 Problem

NSOs often need to perform a primary mission while monitoring and reacting to simultaneously occurring events that may or may not impact their task at hand. During these so-called "vigilance" tasks, operators routinely monitor 10 or more terminal windows; all displaying potentially relevant information to the task at hand. The time and mental focus required to respond to these terminal events routinely interrupts the focus of the operator, while also introducing additional stress which often negatively impacts performance and the time required to accomplish a primary task [2]. Previous research shows that humans naturally have difficulty with this type of task, and that the average person can only keep 3 to 5 chunks of information in working memory [3].

Optimizing vigilance task performance is an active area of research, with numerous studies recently appearing that examine responsiveness as a function of fatigue, cognitive load, eye gaze location, and task repetitiveness [4–6].

Recently, studies have also appeared which examine the ability of AR headsets to improve task performance. These include areas such as surgery, and driving performance [7, 8].

For network security, any technology which helps to streamline the flow of information from numerous terminals could not only potentially reduce operator stress, but also help improve and maintain focus in achieving a primary task. Augmented reality devices have the potential to achieve both of these goals. AR devices, with their displays encompassing a user's field of view, can give important notifications without gaze disruption, as well as provide audio cues that can assist in alerts.

3 Methodology

We designed a simulated network penetration test in the form of a capture-the-flag scenario (described in Sect. 3.3). Test subjects completed the scenario multiple times, both with and without the use of AR devices. Metrics were recorded as the subjects performed, and at the end of each test run the test subject filled out a subjective survey.

This is the first known application of using AR devices in a network penetration testing environment. Given this, a small sample size of seven subjects was deliberately chosen to ascertain whether the approach is reasonable and to provide a measurable indication of the utility of future development and evaluation.

3.1 Test Subjects

Test subjects were selected from a pool of Caucasian males age (34–60) working in telecommunications/research industry. Using actual NSOs would be ideal, but were not availability during the testing period. Most test subjects had little to no hands-on network penetration experience, but were well-versed on the Linux command line. Given the lack of experience in the network security domain, a simulation was designed to cause the test subject to perform similar mental activities.

The test subjects volunteered their time and were not financially compensated.

3.2 Augmented Reality (AR) Devices

Two Android based devices were selected for use in this study: Epson Moverio BT-200 [9] and Vuzix M100 [10]. Both devices connected and communicated to the test environment through the use of the android debugging bridge (adb) over a USB connection.

Epson Moverio BT-200. The Epson glasses (Fig. 2) project a 960-by-540 resolution over each eye producing a stereographic effect similar to viewing am 80 in. screen from 16 ft away. The screen is directly displayed the user's field of vision in both lenses. The Epson glasses were not practical for users that wore corrective lenses.

Fig. 2 Moverio BT-200 augmented reality (AR) headset. The BT-200 has a 1.2 GHz dual core processor, 1 GB RAM, and 8 GB of onboard memory. The display over both eyes can be distracting if an application is not specifically designed to display information

Fig. 3 Vuzix m100 AR headset. The M100 has an onboard 1.2 GHz processor with 1 GB of RAM running Android as its operating system. Its display covers one eye, and has a visibility equivalent to a 4 in. mobile screen held 14 in. away. There is onboard audio for one ear

Vuzix m100. The Vuzix glasses (Fig. 3) showed the display in a 428-by-240 screen in the user's peripheral vision. Users who wear corrective lenses can use these glasses attached to a headset.

3.3 Capture the Flag

Capture the flag [11] scenarios serve as educational exercises of offensive and defensive network security. Often participants are required to use a variety of security skills such as network sniffing, protocol analysis, and cryptanalysis.

Our capture the flag exercise required the user to locate a set of encrypted text files scattered randomly across as a series of hosts on a network. The contents of one file were used to decrypt the contents of another file, which in turn decrypted the contents of another file, and so on until a final success message was obtained.

We developed a python program (puzzlecreator.py) to create the series of encrypted messages using the following procedure.

Let E(k,m) be a symmetric encryption cipher:

(1) Let secret_msg = "You finished the test".
(2) Let n = the total number of flags
(3) Let word[n] = a random dictionary word
(4) Let flag[n] = E(word[n], secret_msg)
(5) Let i = n − 1
(6) While i >= 0:

    (a) Let word[i] = a random dictionary word
    (b) Flag[i] = E(flag[i + 1], word[i])
    (c) i = i − 1

For encryption, we implemented a simple Vigenère cipher (coder.py). The output of each flag was written to a text file.

An example of creating the flags (where N = 4) and decrypting is shown in Fig. 4.

```
$ ./puzzlecreator.py
[flag0, flag1, flag2, flag3 created.]

$ cat flag0.txt
maintainability

$ cat flag1.txt
09bW0-c=

$ ./coder.py maintainability 09bW0-c=
fumes

$ cat flag2.txt
2OTizNvU2tDQ5g==

$ ./coder.py fumes 2OTizNvU2tDQ5g==
roughnecks

$ cat flag3.txt
xtTo24jR1NDb39fj2g==

$ ./coder.py roughnecks xtTo24jR1NDb39fj2g==
Test complete
```

Fig. 4 Decrypting flags using keys found in previously discovered flag files

For each experiment, a set of seven flags was used. Login credentials and flag target locations were given to the participants at the beginning of the scenario in order to normalize performance as a function of network security skill level. While not removing UNIX and network security skill levels entirely from the test, it helped to make the scenario more a function of focus and managing information.

3.4 Primary Task

The primary task was to find a set of files that were randomly hidden in directories across several connected virtual machine systems. Files were named such as “flag1.txt”, “flag2.txt”, and so on. The first flag contained a plaintext dictionary word which would later be used as the cipher to decrypt the next flag. Each subsequent flag would then act as the cipher for the next, until all three were decrypted and the scenario was over.

3.5 Events of Interest

Test subjects were required to monitor and respond to randomly generated events of interest (EOIs) within a limited time window of 45 s. These network EOIs were

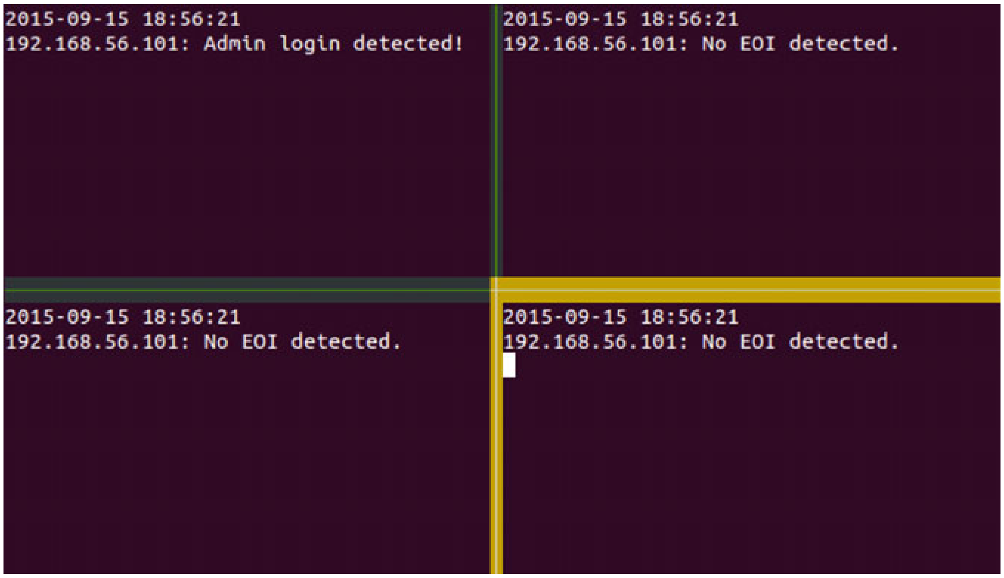

Fig. 5 Four terminal windows with randomly generated events of interest (EOI) to which test subjects must respond

simulated, but correspond to actual events an NSO would typically deal with during an operation; such as responding to an adversarial network administrator logging into the same network, or performing a network capture.

Monitoring Via Terminal Window. To monitor EOIs without an AR device, twelve terminal windows were created which ran a simulated monitoring script. If an event were randomly generated, a notice would appear in the terminal window (Fig. 5).

Monitoring Via AR Device. To monitor for EOIs with an AR device, an Android application was created which would create audible and visual alerts for the NSO as shown in Fig. 6.

Responding to Events. In both cases the subject would respond to an EOI by running a shell script command which would simulate a counter-action to the specific EOI generated. In addition to the script command, the subject was also tasked with solving a randomly generated math problem to further induce stress and distractions. If the subject responded to and successfully countered the EOI as well as solved the math problem, the event was considered to be successfully handled.

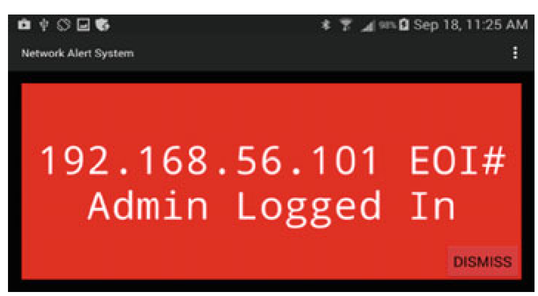

Fig. 6 Event of interest (EOI) notification seen by a test subject in the augmented reality device. The test subjects were required to keep track and respond to randomly generated EOIs while achieving their primary capture the flag objective

4 Results

During each test execution, the following metrics were recorded:

- *Total Elapsed Time*: time between start of experiment and decryption of final flag
- *Average Response Time*: time between the EOI generation and successful countermeasure
- *Success EOIs*: number of times countermeasures where successfully applied
- *Failed EOIs*: number of times where the test subject did not apply a countermeasure for an EOI within the given response window (roughly 60 s)
- *Countermeasure Failure Rate*: (Failed EOIs)/(Success EOIs + Failed EOIs)

In addition, each test subject completed the TLX assessment to indicate their experience (scaled 1–20):

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

4.1 Total Elapsed Time

In general, the Total Elapsed Time decreased when using the glasses. As shown in Fig. 7, test subjects finished the task 6 min faster on average when wearing the glasses (with the exception of one tester).

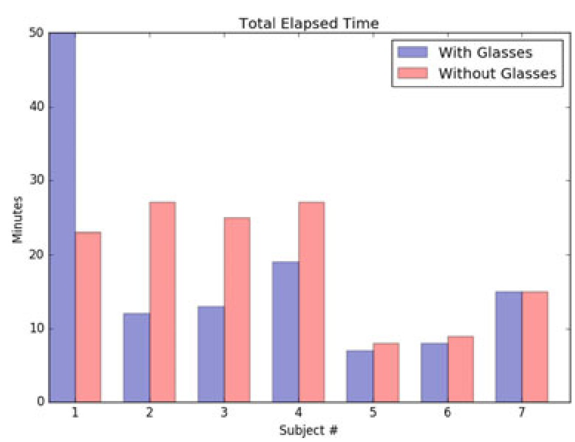

Fig. 7 Total elapsed time to perform the experiment for each subject, with, and without glasses

4.2 Average Response Time

The Average Response Time to EOIs decreased for all test subjects when using the glasses. As shown in Fig. 8, on average, test subjects responded 7 s faster when wearing the glasses.

4.3 Countermeasure Failure Rate

The failure rate (number of failed countermeasures/number of total EOIs) decreased for test subjects when using the glasses, with a couple of exceptions. As shown in Fig. 9, on average, test subjects experienced 10 % fewer failures when wearing the glasses.

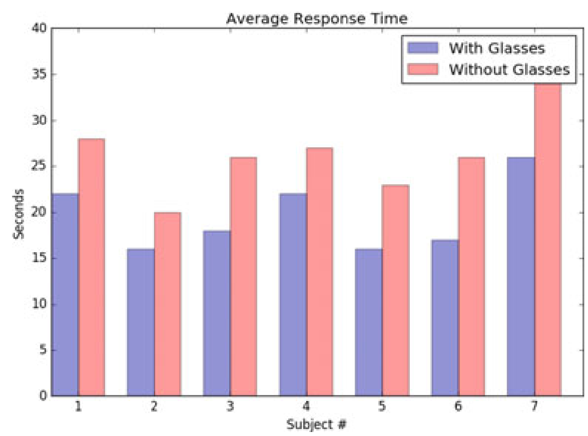

Fig. 8 Average time taken to respond to randomly generated EOIs

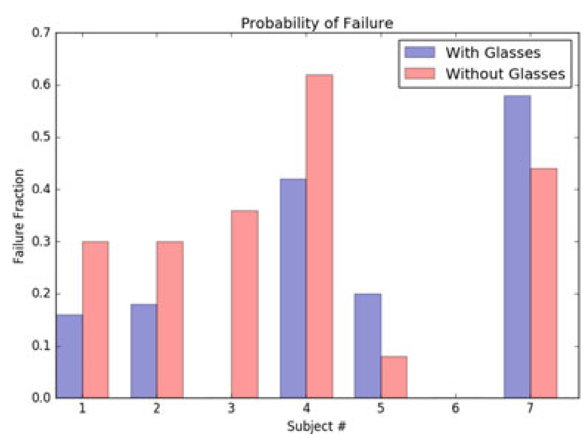

Fig. 9 Countermeasure failure rate in responding to randomly generated EOIs while working on primary task

4.4 User Experience

Most test subjects commented that they had a positive experience when using the glasses. Many test subjects said that they:

- were less distracted
- could focus better on the main task
- found it easier to recognize the EOI alerts

About half the subjects tested used the Vuzix glasses (those that wore corrective lenses), while the other used the Epson glasses. Some subjects that wore the Epson glasses (that displayed the alert directly in the field of vision) commented that the bright red alert was annoying, and thus further encouraged them to respond with a countermeasure quicker.

A few test subjects had difficulty with the glasses and commented that they:

- found the glasses distracting and uncomfortable
- moved their focus to often
- found the alert too annoying (on the Epson glasses)
- had trouble reading the alerts
- actually saw more alerts but found it more stressful

The TLX self-assessment results were quite varied, with the meaningful metric being the difference reported between the two cases.

We first give an example of the overall self-reported scores for Mental Demand in Fig. 10. Five of the seven subjects who completed the TLX reports claimed the glasses lowered their mental demand, while one claimed it raised it, and a sixth saying it made no difference.

As shown in Fig. 11, the average user results show overall improvements in five of the six TLX categories, with only physical demand being reported as getting worse (overall scores were sums of all six categories). The greatest improvement seems to have been in the temporal demand of the task, as users felt less rushed knowing they didn't have to continuously search for network EOIs.

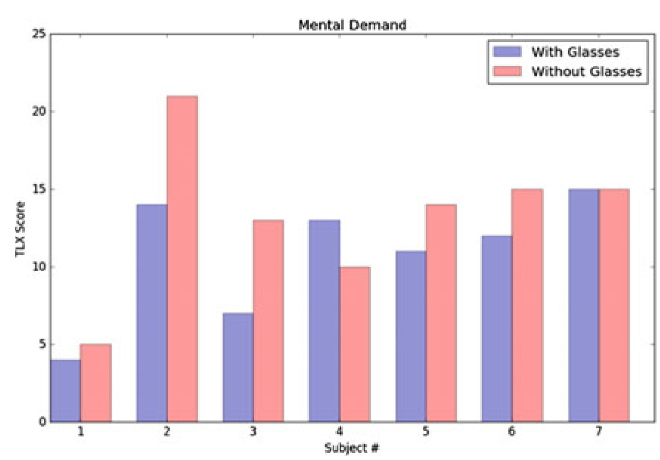

Fig. 10 Self-reported mental demand of carrying out the capture the flag scenario both with and without the use of AR glasses

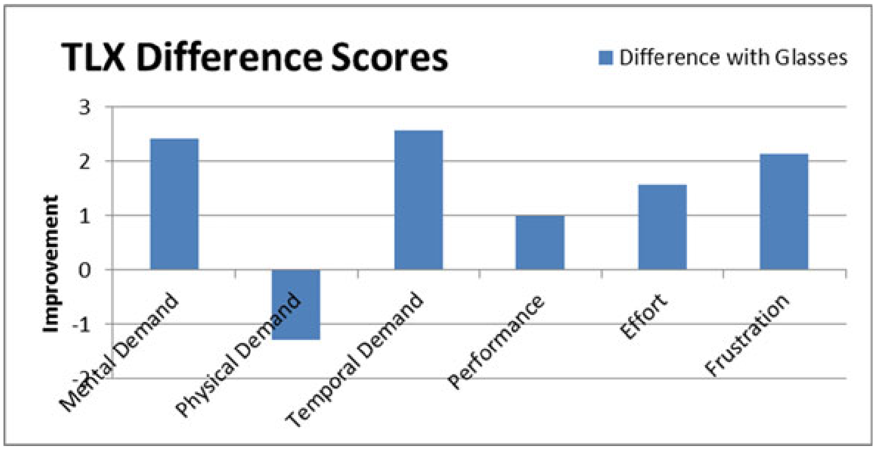

Fig. 11 Self-reported differences in NASA TLX scores between using the glasses and not using the glasses during the capture the flag scenario

5 Discussion

In summary, when using the glasses, test subjects performed better for our particular test than when not wearing the glasses. The quantitative measure of performance was judged on:

- Total time elapsed
- Average response time for applying countermeasures, and
- Failure rate

The improvement appears to be correlated to the overall less distraction and greater ability to focus when using the glasses, as stated by the test subjects. Looking at this closer, consider the activities a test subject performs during the scenario without the glasses:

- Focusing on the main goal (finding the flags)
- Actively maintaining a discipline to periodic shift their attention to the alert screen
- Visually shifting their focus several inches away from the main task

In software design, the periodic shifting of attention is similar to a "pull" request. In other words, the user needs to consciously perform periodic mental interrupts to see if an alert has occurred. This process is not efficient, since sometimes the user interrupts his or her workflow to check the notifications window when no alerts exist. On the contrary, if the user forgets to check the window within a given time frame, he or she may miss applying the countermeasure.

Contrast this to the user's activities when performing the scenario with glasses:

- Focusing on the main goal (finding the flags)
- Passively receiving alert notifications audibly and visually when an event occurs
- Visually shifting attention and visual focus slightly when an alert occurs

This is analogous to a "push" request. The user need not "remember" to check periodically for alerts. Rather, when an alert occurs, it will be "pushed" to the user. The alert itself interrupts the user from their current task, only when necessary. Thus the user will be interrupted less overall, and will be more efficient in using his or her time.

Also, spending less time context switching should cause less mental stress to the user.

Push notifications could be integrated into the desktop screen, rather than in the AR device. However our belief is that success requires maintaining a gaze on the primary task. An AR device helps maintain this concentration by offering the ability to maintain task focus without adverting the user's gaze.

Some test subjects essentially ignored the alerts when not using the glasses and had a higher failure rate. When using the glasses, it was more difficult to ignore the alerts and thus the test subjects responded more often. Some test subjects commented that wearing the glasses was more distracting, but that could be due to perceiving more alerts than they did when not wearing them. However, the results indicate that the overall performance of users was greater when wearing the glasses compared to when not.

In summary, using the glasses freed the test subject from switching focus when unnecessary, and reduced the amount of overall time the test subject took handling EOI detection. It also reduced the amount of mental burden on the test subject by shifting less often.

6 Conclusion

These results suggest that AR devices may be beneficial for network security operations scenarios where multiple streams of ancillary information must be monitored in order to ensure optimal performance. Additionally, we see that AR in some cases reduces the mental demand and stress of these types of time sensitive operations.

As AR hardware and software continue to mature, there may be additional opportunities to create solutions that aid the performance of high stress tasks. Some of these possible features may include teaming, or the sharing of information in a virtual environment, as is currently possible with the Microsoft Hololens. Other possibilities include making use of computer vision techniques which could allow the AR device to see and make recommendations about information it sees through a front facing camera, rather than through an application specific android app as it currently does.

AR is still in its infancy, and has struggled to find a widespread use case in the consumer market. We believe it may have a role to play in niche applications such as network security operations where management of information flow is critical to success.

Acknowledgements The research described in this paper was sponsored by the U.S. Department of Defense (DOD) through Applied Communication Sciences (ACS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

References

1. Hart, S., Staveland, L.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) Human Mental Workload. North Holland Press, Amsterdam (1988)
2. Dykstra, J., Paul, C.: Stress and the cyber warrior: cognitive workload in a computer operations center. J. Sensitive Cyber Res. Eng. **3**(1), 1–23 (2015)
3. Cowan, N.: The magical mystery four: how is working memory capacity limited, and why? Curr. Dir. Psychol. Sci. **19**(1), 51–57 (2010)
4. Sawyer, B., Finomore, V., Funke, G., Mancuso, V., Funke, M., Matthews, G., Warm, J.: Cyber vigilance: effects of signal probability and event rate. Proc. Hum. Factors Ergonomics Soc. Ann. Meet. **58**(1), 1771–1775 (2014)
5. Amico, A., Whitley, K., Tesona, D., O'Brien, B., Roth, E.: Cyber defense situational awareness: a cognitive task analysis of information assurance analysts. Proc. Hum. Factors Ergonomics Soc. Ann. Meet. **49**(3), 229–233 (2005)
6. Popola, A.: The effects of eye gaze based control on operator performance in monitoring multiple displays. Masters Thesis, Embry-Riddle Aeronautical University (2009)
7. Dixon, B., Daly, M., Chan, H., Vescan, A., Witterick, I., Irish, J.: Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. Surg. Endosc. **27**(2), 454–461 (2013)
8. Rusch, M., Schall Jr., M., Gavin, P., Lee, J., Dawson, J., Vecera, S., Rizzo, M.: Directing driver attention with augmented reality cues. Transp. Res. Part F Traffic Psychol. Behav. **16**, 127–137 (2013)
9. Epson Moverio. http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H560020
10. Vuzix m100. https://www.vuzix.com/Products/M100-Smart-Glasses
11. Capture the Flag. https://ctftime.org/ctf-wtf/