2005

Analyzing the Physical and Vestibular Effects of Varying Levels of Immersive Displays for Controlling Unmanned Aerial Vehicles from an Aircraft Platform

John Bruyere

Justin Gripp

Christopher Nagy

Terence Andre

ANALYZING THE PHYSICAL AND VESTIBULAR EFFECTS OF VARYING LEVELS OF IMMERSIVE DISPLAYS FOR CONTROLLING UNMANNED AERIAL VEHICLES FROM AN AIRCRAFT PLATFORM

John Bruyere, **Justin Gripp**, **Christopher Nagy**, and **Terence Andre**
USAF Academy
Colorado Springs, Colorado

This study attempted to further the base of knowledge concerning effects on watching video taken from an Unmanned Aerial Vehicle (UAV). Sixteen participants from the U.S. Air Force Academy were involved in watching UAV video under 2 conditions of motion (with and without) and 2 conditions of video presentation (laptop computer screen and a head-mounted display). Each video was about 5 minutes long and following each condition the subject filled out a questionnaire which judged their sickness level based on many different factors. Our results did not show any significant difference in sickness levels between the 4 conditions, and further research will have to be performed to fully investigate the effects of watching UAV camera video in an immersive environment.

Introduction

The military has significantly increased its focus on Unmanned Aerial Vehicles (UAVs) in recent years and as this focus and need increases, increased applications of UAVs will continue to arise. While most UAVs have been controlled from a ground station so far, the near future will likely present the need to control UAVs from mobile systems on the ground and military aircraft deployed to an area of interest. Controlling UAVs from other airborne vehicles presents some unique challenges. In particular, operators will have to deal with the potential of sensory conflict between the display from the UAV workstation and the sensory input from the motion of the aircraft. In anticipation of this future need, understanding the unique demands that this will place on the operator(s) of the UAV is crucial. Furthermore, determining which type of display mode will be important when trying to control UAVs from airborne platforms. Consideration should be given to portable 3-D immersive displays (e.g., Head-Mounted Display, HMD) as well as a 2-D laptop computer screen (LCS) in presenting UAV information.

Directly applicable research in this new and focused area is scarce, but there has been a fair amount of research on the slightly broader areas of motion sickness and effects of virtual reality environments. The main theory behind the origins of motion sickness in different environments is the sensory-conflict theory (Yardley, 1992). "Sensory-conflict theory proposes that symptoms occur as a result of conflict between signals received by the three major spatial senses: the visual system, the vestibular system, and nonvestibular proprioception" (Cobb, Nichols, Ramsey, & Wilson, 1999, p. 170). In their study, Cobb et al. (1999) analyzed nine different experiments examining after-effects from different virtual reality (VR) systems, virtual environment (VE) designs, and task requirements, resulting in a total sample pool of 148 participants. A variety of measures, from surveys to physiological indicators, were used to measure different effects, sickness being the item we are most interested in (Cobb et al., 1999). Their results from the self-report data indicated that symptoms of sickness usually occurred within 15 minutes of being immersed. They also found that symptom levels were highest on the first immersion trial and negligible on the third, leading to the conclusion that the participants habituated to the environment after two trials. This information would be helpful in designing our experiment.

As any person experienced with flight simulators knows, UAV pilots today and those of the future will likely have to deal with the phenomenon of vection. Vection "refers to the powerful illusory sensation of self-motion induced in viewing optical flow patterns" (Hettinger, Berbaum, Kennedy, Dunlap, & Margaret, 1990, p. 172). Hettinger et al. (1990) performed a study in which subjects sat and watched a 15-minute flight simulation video which included turns, banks, and altitude changes. The subjects had to watch the display and indicate how much vection was experienced. The researchers hypothesized that those who had more experiences of vection would be more likely to experience sickness. Hettinger et al. (1990) found that those who experienced vection got sick a significantly higher percentage of the time than did those with very limited or no experiences of vection, showing that symptoms of motion sickness can arise by just viewing a screen, screens that usually involve a large field-of-view, even when there is no physical movement. This finding is very relevant to our condition using the HMD because it will likely have a large field-of-view while the individual is tracking an object, compared to our LCS condition, helping us to determine which viewing method is best.

Hettinger et al. (1990) also explained that very few subjects reported symptoms following the initial 15-minute display because motion sickness is a cumulative phenomenon. An important research question in our study is whether a small change in presentation mode (LCS vs. HMD) can impact the feeling of sickness in just a short amount of time.

Studying more about the effects of virtual environments (VE), Stanney, Kingdon, Graeber, and Kennedy (2002) performed a study in which individuals were exposed to a 3-D VE and required to perform certain tasks, such as locomotion, manipulation, turning, etc. The researchers found that the more movement control VE users had, the more presence they would experience, although complete control would make them sicker (Stanney, et al., 2002). These results indicate that there is something of a tradeoff between full movement control in an environment, leading to a higher sense of presence along with a greater level of sickness, and less control, correlated with less sickness. This could be important for how much immersion and how much control (versus possibly more automation) is presented to pilots controlling UAVs.

Another study that is directly applicable to our experiment was performed by Kennedy, Lane, Berbaum, and Lilienthal (1993) which consisted of creating a new survey from which to judge simulator sickness (SS). The Simulator Sickness Questionnaire (SSQ) was developed to replace the Pensacola Motion Sickness Questionnaire (MSQ) which was not created for SS and thus less applicable to the special circumstances of SS. The SSQ was derived partly from the MSQ using a series of factor analyses to come up with the appropriate format and substance. Our study used a modification of the SSQ in order to compare the levels of sickness experienced by our subjects.

In our experiment, we introduced two independent variables (IVs) with two conditions each. The first IV is whether the students viewed camera video with an immersive HMD or with a non-immersive LCS. The second IV is whether the subject experienced motion in the chair they were sitting in during the trial. Motion is defined as random turns, from 90 degrees to 180 degrees, that the subject experienced. This motion is meant to simulate the motion that might be felt onboard an aircraft, though it is severely limited in that it can only move in two degrees of motion and cannot simulate turbulence. The dependent variable (DV) is simply what level of sickness they feel, defined by their results from the SSQ.

Our null hypothesis is that there will be no significant difference between the HMD and LCS conditions on the level of reported nausea in the participants. Our alternative hypothesis is that the HMD environment will increase the feeling of motion sickness, and that when coupled with movement, the increase will be even greater. We feel that when fully immersed, an individual will feel more sickness than when not immersed, and that those feelings will be intensified with even slight motion.

Method

Participants

For our study, we utilized 16 male and female cadets enrolled in an introductory psychology class at the United States Air Force Academy. These cadets consisted of both freshman and sophomore cadets who voluntarily signed up and received extra credit from their teachers for participating.

Apparatus

In order to expose the participants to a UAV environment, we used a fully immersive HMD virtual reality system to recreate the pilot's view. The HMD was a Virtual Research V8 with 800x600 resolution. To recreate the view from a flat screen display to compare against the HMD, we used an IBM ThinkPad Laptop computer for the non-immersive Laptop Computer Screen (LCS). A Dell 3.2 GHz, Pentium 4 processor desktop computer was used to display the video for the HMD. For the video, we used UAV flight video flown over recognizable areas of the United States Air Force Academy which lasted 4 minutes and 40 seconds in length. To simulate the movements for the participants, we used a non-motorized Barany Chair. A modified SSQ was given to each participant in-between each condition. A picture of the setup can be seen in Figure 1.

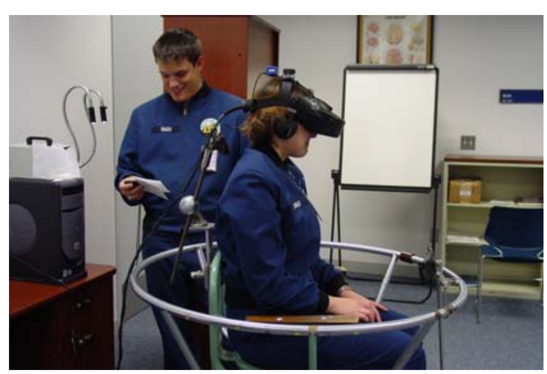

Figure 1. *Experiment setup of HMD condition.*

Procedure

Upon arriving to the testing room, the participants read and completed the consent form. After they were told that they may withdraw at anytime, we explained the procedures and began the experiment. Participants watched the UAV camera video in 4 different conditions. Those conditions were the HMD with motion, the HMD with no motion, the LCS with motion, and the LCS with no motion. Our experimental design includes a balanced within-subjects design in which each participant watched the video under all four conditions. The design is balanced so that each possible order of the four conditions is performed in order that learning effects do not affect the outcome of the data. For the movement conditions, the participant began with the chair facing south. The participants were exposed to lateral turns in the Barany chair at exact times and rotation degree values predetermined by the researchers as shown in Table 1.

Table 1. *Times and directions for chair movements during experiment.*

Time	Direction	Degrees
0:15	L	90
0:45	R	90
1:00	R	90
1:20	L	180
1:25	R	180
2:00	L	90
2:10	R	90
2:40	L	180
3:00	R	180
3:10	L	90
3:15	L	90
3:45	R	180
4:00	L	180
4:10	R	90
4:30	L	90
4:35	R	90

For the "no movement" conditions, the participants sat in the Barany chair during the video presentation. For the "movement" conditions, the participants wore the HMD; during the LCS conditions, participants held the laptop in their lap for the duration of the video. For all conditions, the lights were turned off in the room in order to prevent any distractions and to simulate a dark aircraft cabin. All the participants were given tasks to complete during the video under all conditions. They were told to count and keep a running total in their heads of any moving vehicles they saw, count and keep a running total in their heads of any street intersections, and provide heading information (North, South, East, or West) pertaining to the UAV flight path. After each condition, the participant completed a modified version of the SSQ and was given 1 minute to walk around and rest before the next condition began. The modified SSQ we used did not include 3 components that did not apply to our very basic simulator experience or might be confusing based on our pilot study (i.e., stomach awareness, burping, and fullness of head). We replaced two of these terms with more general items (dizziness and sickness feeling) hoping to make the questionnaire slightly more sensitive to the very slight differences that might exist between our conditions.

Results

The data generated by the SSQ was entered into SPSS v 11.0 and analyzed. The data from each condition was matched with each participant based upon the order in which the participants signed up to participate in the study. Throughout the analysis the alpha was selected as $\alpha = .05$.

Participant scores for each item of the SSQ were summed according to a weighted scale subscribed by the SSQ as illustrated in Table 2. Simulator sickness has three subcomponents which make the whole construct (nausea, oculomotor, and disorientation). Certain items such as difficulty focusing and nausea counted towards two of the three subcomponents. Other items, like fatigue, counted toward just one subcomponent. No item counted towards all three. We made an assumption that the new items we replaced on the SSQ (dizziness and sickness feeling) counted in a single category as did the original items. We multiplied any item answer that counted towards two subcomponents by a factor of two to weight it properly and any answer that counted toward just one subcomponent was multiplied by 1. We then summed all of the scores for each condition in order to obtain a total sickness score for each participant in all four conditions.

Table 2. *Weighting scale showing example of weights for two symptoms.*

	Weight		
SSQ Symptom	*Nausea*	*Oculo-motor*	*Disorientation*
Fatigue	0	1	0
Nausea	1	0	1

We analyzed our data using a repeated measures analysis of variance. Table 3 shows the mean and standard deviations for each of the four conditions. For the display presentation type, we found no significant results, $F(1,15) = 3.615$, $p = 0.077$. For motion we again found no significant results, $F(1,15) = 1.90$, $p = 0.188$. There was also no interaction between display type and motion, $F(1,15) = 2.241$, $p = 0.155$. Figure 2 shows a graph of the changes in sickness means for both of our conditions (motion and display).

Table 3. *Mean and standard deviation SSQ scores for all four conditions.*

	Display Type	
	HMD	LCS
Motion Type	Mean (SD)	Mean (SD)
Motion	7.125 (6.20)	3.4375 (3.65)
No Motion	4.8750 (4.60)	3.3750 (3.42)

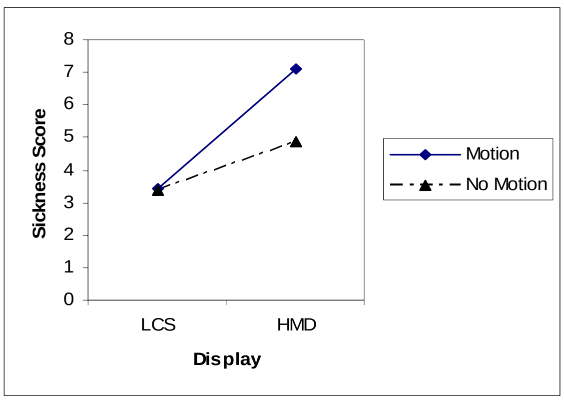

Figure 2. *A graph showing the marginal means for all four conditions.*

Discussion

The mean values for the HMD conditions showed a slight increase in the level of sickness reported, although the difference was not statistically significant using an alpha = .05. There are many probable reasons for why we were unable to show any significant results. The first and most salient reason is the number of participants. Under optimal conditions, we would need approximately 30 participants in each cell to show reliable results. In our case we were limited to 16 based on class time constraints and scope of project. Perhaps with a larger participant pool we would be able to show the results we had expected. A second mitigating factor was our inability to utilize motion on any axis outside of the z-axis. In a real life condition, a UAV pilot operating from an AWACS would be subject to motion on all three axes, not just one. Another real-life condition we did not have the means to replicate was the flying of a UAV. We subjected our participants to videos of UAV flight, but this condition could not fully represent the attention that would be given to the screen if the participant was actually piloting a UAV, even despite our attempts to alleviate this factor by having the participant attend to heading, moving vehicles, and intersections. Another factor that influenced our results was the amount of time the participants were subjected to the video. Due to time constraints, we were only able to show a 4 minute and 40 seconds video, where Cobb et al. (1999), found that 15 minutes of immersion is necessary to generate the sickness we were looking for.

For future research, we would recommend that the above issues be addressed by meeting a few important requirements (assuming optimal conditions and appropriate assets). First the sample size should be increased to include at least 30 participants to help validate the findings. Secondly, a full-motion and fully immersive simulator should be utilized, in which the participants would be required to actually operate a simulated UAV. Finally, participants should be subjected to each condition for at least fifteen minutes before the SSQ is administered.

References

Cobb, S. V., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments, 8*(2), 169-187.

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

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

Stanney, K. M., Kelly, S., Graeber, D., & Kennedy, R. S. (2002). Human performance in immersive virtual environments. *Human Performance, 15*(4), 339-367.

Yardley, L. (1992). Motion sickness and perception: a reappraisal of the sensory conflict approach. *British Journal of Psychology, 83*(4) 449-472.

EFFECTS OF WORKLOAD AND LIKELIHOOD INFORMATION ON HUMAN RESPONSE TO ALARM SIGNALS

Ernesto A. Bustamante
James P. Bliss
Old Dominion University
Norfolk, VA

The purpose of this study was to examine how workload and likelihood information would affect participants' responses to alarm signals while they performed a battery of tasks. As expected, participants' overall response rates and false alarm response rates were significantly lower, and true alarm response rates were significantly higher when they used a likelihood alarm system. These results were particularly noticeable under high workload conditions. Results from this study suggest that although people may respond less often to alarm signals when they are provided with likelihood information, they will more likely respond to true signals rather than false alarms. Therefore, designers should incorporate likelihood information in alarm systems to maximize people's ability to differentiate between true and false alarms and respond appropriately.

Introduction

Technological advances have made the use of automated alarm systems a common practice in aviation (Bliss, 2003). Such systems serve a crucial function in the cockpit by alerting pilots of potential or imminent dangerous conditions. Nevertheless, even the most sophisticated alarm systems emit a high number of false alarms, increasing pilots' level of workload and jeopardizing their flight performance (Getty, Swets, Pickett, & Gonthier, 1995; Gilson & Phillips, 1996).

A possible solution to this problem is to provide pilots with additional information regarding the positive predictive value (PPV) of alarm signals through the use of a likelihood display. The PPV of a signal, which is also commonly referred to as its "alarm reliability," is defined as the conditional probability that given an alarm, a problem actually exists. Researchers have shown that people adjust their responsiveness based on the outputs given by alarm systems (Meyer & Ballas, 1997; Robinson & Sorkin, 1985). More specifically, people's responsiveness to alarm signals is dependent on the PPV of such signals (Bliss & Dunn, 2000; Bliss, Gilson, & Deaton, 1995; Getty et al., 1995). The purpose for using a likelihood alarm display is to provide people with information about the PPV of different signals so that they can respond more often to high-likelihood signals and less often to low-likelihood signals.

However, researchers have questioned the usefulness of such displays by pointing out that they may actually decrease pilots' responsiveness, thereby jeopardizing flight safety (Sorkin, Kantowitz, & Kantowitz, 1988). Nonetheless, providing pilots with likelihood information may enhance their decision-making strategies such that they might respond more often to signals that signify actual problems and disregard false alarms. However, few researchers have examined how operators of complex tasks react when faced with signals generated by a likelihood alarm system. Similarly, there is little awareness of how other task variables might interact with likelihood information to influence alarm reaction patterns or primary task performance. The purpose of this study was to examine how workload and likelihood information would affect people' responses to alarm signals.

Participants performed the tracking and resource-management tasks from the Multi-Attribute Task (MAT) Battery (Comstock & Arnegard, 1992) and an engine-monitoring task that the experimenters designed. We manipulated workload level by automating the tracking task and by increasing the difficulty of the resource-management task. While performing their tasks, participants reacted to alarms generated by either a binary alarm system (BAS) or a likelihood-alarm system (LAS).

We assessed participants' response rates to false alarms and true signals. We expected participants to respond more often to false alarms when they interacted with the BAS, particularly during low workload (Sorkin et al., 1988). This hypothesis was consistent with previous research, which suggests that people are generally more likely to respond to alarm signals under low workload conditions (Meyer, 2002). However, we hypothesized that participants would respond more often to true signals when they interacted with the LAS compared to the BAS, and that this difference would be greater under high workload conditions. The reason for this was that we expected the LAS would improve participants' ability to detect alarms that were more likely to be true signals. Such an expectation is reflected by Selcon, Taylor, and Shadrake (1991), who demonstrated the benefits of redundant information on pilot reactions to displays in the cockpit.

Method

Experimental Design

We used a full within-subjects design. Preliminary analyses consisted of descriptive statistics to ensure that we did not violate any statistical assumptions. We set statistical significance for all inferential tests *a priori* at $\alpha = .05$.

Participants

An *a priori* power analysis revealed that approximately 30 participants would be necessary to obtain a power of .80, assuming a medium effect size ($f = .25$) at an alpha level of .05 (Cohen, 1988). Therefore, we used convenience sampling to select 30 (18 females, 12 males) undergraduate and graduate students from Old Dominion University to participate in this study. Participants ranged from 18 to 38 years of age ($M = 22.70$, $SD = 4.54$). All participants had normal or corrected-to-normal vision and hearing. To motivate participants, we provided them with three research credit points to apply to their class grades, and awarded a $10 prize to the person who performed best.

Materials and Apparatus

To increase the realism of the experimental design, participants performed a set of complex primary tasks at the same time they performed the secondary task. The primary tasks consisted of a compensatory-tracking task and a resource-management task, both taken from the MAT (Comstock & Arnegard, 1992). We loaded the MAT on an IBM-compatible computer and displayed it to participants using a 17-inch monitor. Participants performed the MAT using a standard mouse and a QWERTY keyboard.

While performing the MAT tasks, participants also performed an engine-monitoring task that the experimenters designed. We presented this task to participants on a separate 17-inch monitor, located at 90º to the right of the primary task. This engine-monitoring task required participants to respond to a series of alarms that indicated a potential problem with two engines. As they performed the MAT, participants encountered different alarms and had to decide whether to ignore them or respond to them by searching for critical system-status information. To search for this information, participants had to divert their attention from the primary task and press the space bar on the keyboard located in front of the computer hosting the secondary task. Once they did this, the screen presented them with the system-status information regarding the current oil temperature and pressure of the two engines. Participants then assimilated this information and decided whether they needed to correct the problem by pressing the space bar again, or cancel the information by pressing the escape key and returning to the primary task. To keep participants motivated, they received a score on the engine-monitoring task, which was updated after each alarm depending on their response.

Participants received one point for searching for further information when an alarm was true and for ignoring false alarms. They lost one point for searching for further information when an alarm was false, but they lost three points for ignoring a true alarm. If they checked the status of the two engines, they received two points for correctly resetting actual problems and one point for canceling the information when there was no problem. They also lost one point for resetting the system when there was no problem, but they lost three points for canceling the information when a problem actually existed. The rationale for using this point system was to more closely simulate the payoff associated with responding to and ignoring alarm signals in a complex task situation, such as flying an airplane, where adequately responding to true alarms is crucial for flight safety.

Alarm Systems

Binary Alarm System We modeled the performance of the binary alarm system based on prior research (Bustamante, Anderson, & Bliss, 2004). The probability of a problem was .01. The system had a high sensitivity (d'=3.98) and a low threshold (β=.23). Based on these parameters, the system was able to detect the presence of a problem 99% of the time, while issuing a false alarm rate of 5%. The system had a sampling rate of 1s. Each experimental session lasted 30 minutes, and a problem could arise at any given second throughout each session. Based on the prior probability of the problem, a total of 18 engine malfunctions occurred throughout each session. The system was able to detect the presence of all the problems, thereby generating a total of 18 true alarms throughout each session. However, because of the low base rate of the problem and the system's low threshold, it generated a total of 82 false alarms, resulting in an overall system reliability of 18%. The true and false alarms generated by the system looked and sounded exactly alike, to reflect real-world situations where the operator must search for additional information to ascertain alarm validity. The visual component of the alarm signal consisted of a yellow circle accompanied by the word "WARNING" written underneath it. The auditory component of the alarm signal was a simple sine wave at a frequency of 500 Hz, presented at 65 dB(A) through a set of flat-panel speakers. The ambient sound pressure level was approximately 45dB(A).

Likelihood Alarm System. The overall performance of the likelihood alarm system was the same as the binary system. However, this system generated two types of alarms depending on the likelihood that they would be true. To determine the likelihood of each alarm, the system had two simulated thresholds instead of one. We set the lowest threshold of this system at the same value as the binary system, and the highest threshold at β=88.40. Based on these two thresholds, the system generated a total of 84 low-likelihood alarms, 4 of which were true and 80 of which were false. As a result, these alarms had a 5% likelihood of being true. This system generated a total of 16 high-likelihood alarms, 14 of which were true and 2 of which were false. As a result, these alarms had an 88% likelihood of being true. The low-likelihood alarm signals consisted of the same stimuli used for the binary system. The visual component of the high-likelihood alarms consisted of a red circle accompanied by the word "DANGER" written underneath it. The auditory component of these alarms was a simple sine wave at a frequency of 2500 Hz, also presented at 65dB(A).

The rationale for using this particular design for the likelihood alarm system was to use peripheral cues such as color, signal word, and sound frequency to enable participants to easily differentiate between low- and high- likelihood alarms. Although these cues may affect the perceived urgency of such signals, prior research suggests that the effect of the PPV of alarms overshadows any effect that could be attributed to perceived urgency (Burt, Bartolome-Rull, Burdette, & Comstock, 1999).

Procedure

As part of this study, participants completed two experimental sessions during which they interacted with an alarm system and an automatic pilot. During one of these sessions, participants used a binary alarm system, and for the other session, they used a likelihood alarm system. We fully counterbalanced the order in which participants used these systems.

Participants came to the laboratory individually. When they entered the laboratory, they first read and signed an informed consent form and then completed a background information form. The purpose of the background information form was to collect information relevant to the exclusionary criteria for the experiment, such as participants' age and whether they had any visual or auditory problems. Once participants completed this form, we provided them with the instructions about how to perform the MAT tasks. Next, participants performed a 5-min practice session.

Once participants completed this practice session, the experimenter provided them with the instructions about how to complete the engine-monitoring task. Participants then went through another 5-min practice session, performing all tasks at the same time. Next, the experimenter informed participants of the overall reliability of the system and the likelihood of each type of alarm. Then, participants performed the two experimental sessions, taking a 5-min break between them. Before participants began the second session, we provided them with information about the other alarm system. Then, participants went through another 5-min practice session, using the other alarm system. After this practice session was over, participants performed the second experimental session using the other alarm system.

Each experimental session lasted 30 min. During the first and last 7.5 min, participants performed the tracking task manually, and they experienced a series of random pump malfunctions in the resource-management task. At other times, the autopilot performed the tracking task, and participants did not experience any pump malfunctions in the resource-management task. The rationale for doing this was to more closely simulate the distribution of workload levels found in applied settings, such as in aviation, where the take-off and landing phases of flight are associated with higher levels of workload than the cruising phase.

Dependent Measures

We assessed participants' overall response rates (ORR), which was the proportion of alarms that participants responded to in a given session. We also assessed participants' false alarm response rates (FARR), which was the proportion of false alarms that participants responded to in a given session. Last, we assessed participants' true alarm response rate (TARR), which was the proportion of true alarms that participants responded to in a given session.

Results

We conducted three 2 x 2 repeated-measures ANOVAS. We used workload (Low, High) and system (BAS, LAS) as independent variables. We used ORR, FARR, and TARR as dependent measures. Results from the first ANOVA showed a statistically significant main effect of workload on ORR, $F(1,29) = 46.25$, $p < .001$, partial $\eta^2 = .62$. Participants' ORR was significantly higher during low workload ($M = .51$, $SD = .24$) than during high workload ($M = .40$, $SD = .23$). Results from this first analysis also showed a statistically significant main effect of system on ORR, $F(1,29) = 28.04$, $p < .001$, partial $\eta^2 = .49$. Participants' ORR was significantly higher when they interacted with the BAS ($M = .54$, $SD = .26$) than when they interacted with the LAS ($M = .37$, $SD = .19$). These results are shown in Figure 1.

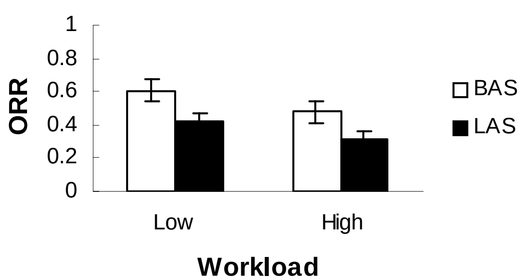

Figure 1. *Overall response rate as a function of workload and system.*

Results from the second ANOVA showed a statistically significant main effect of workload on FARR, $F(1,29)=35.67$, $p<.001$, partial $\eta^2=.55$. Participants' FARR was significantly higher during low workload ($M = .46$, $SD = .27$) than during high workload ($M = .34$, $SD = .26$). Results from this second analysis also showed a statistically significant main effect of system on FARR, $F(1,29)=57.93$, $p<.001$, partial $\eta^2=.67$. Participants' FARR was significantly higher when they interacted with the BAS ($M = .54$, $SD = .25$) than when they interacted with the LAS ($M = .27$, $SD = .22$). These results are shown in Figure 2.

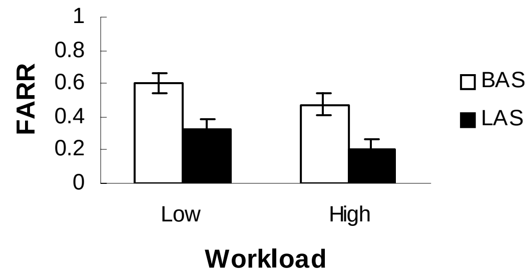

Figure 2. *False alarm response rate as a function of workload and system.*

Last, results from the third ANOVA showed a statistically significant workload by system interaction effect, $F(1,29)=7.20$, $p<.05$, partial $\eta^2=.20$, and statistically significant main effects of workload, $F(1,29)=14.10$, $p<.01$, partial $\eta^2=.33$, and system, $F(1,29)=30.22$, $p<.001$, partial $\eta^2=.51$, on TARR. Participants' TARR was significantly higher when they interacted with the LAS ($M = .80$, $SD = .13$) than when they interacted with the BAS ($M = .56$, $SD = .31$), but this difference was greater during high workload. These results are shown in Figure 3.

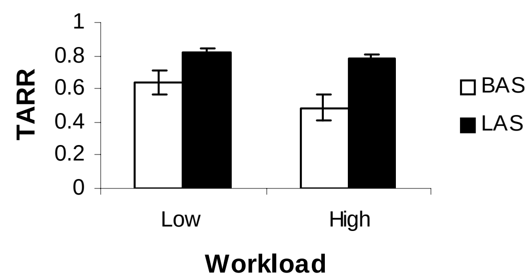

Figure 3. *True alarm response rate as a function of workload and system.*

Discussion

Results supported our hypotheses. As expected, participants responded significantly more often to false alarms when they interacted with the BAS, particularly under low-workload conditions. However, participants responded significantly more often to true signals when they interacted with the LAS, especially during high-workload conditions.

In general, the results of this experiment support the use of redundant information to signify alarm validity, or lack thereof. As noted by Selcon, et al. (1991), the presence of such information can improve pilot reactions to displayed information in the cockpit. Bliss, Jeans, and Prioux (1996) showed similar results; when participants were faced with an unreliable alarm system, they benefited most from the presence of additional information upon which to base their judgments of individual alarm validity.

Results from this study have potential applications for designing alarm systems in the field of aviation. These results suggest that although pilots may respond less often to alarm signals when they are provided with likelihood information, they are more likely to respond to true signals rather than false alarms. Therefore, designers should incorporate likelihood information in alarm systems to maximize pilots' ability to differentiate between true and false alarms and respond appropriately. This, in turn, may increase safety by directing pilots' attention to actual problems without jeopardizing flight performance by minimizing responsiveness to false alarms.

References

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics, 43*(9), 1283-1300.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Bliss, J. P., Jeans, S. M., & Prioux, H. J. (1996). Dual-task performance as a function of individual alarm validity and alarm system reliability information. Proceedings of the IEA 1996/HFES 1996 Congress, Santa Monica, CA: International Ergonomics Association/Human Factors and Ergonomics Society, 1237-1241.

Burt, J. L., Bartolome-Rull, D. S., Burdette, D. W., & Comstock, J. R. (1999). A psychological evaluation of the perceived urgency of auditory warning signals. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 151-169). Aldershot, Hants: Ashgate Publishing Limited.

Bustamante, E. A., Anderson, B. L., & Bliss, J. P. (2004). Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 1948-1952.

Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research* (NASA Technical Memorandum No. 104174). Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*, 19-33.

Gilson, R.D., & Phillips, M.J. (1996). Alert or sound alarm? *Aerospace Engineering*, 21-23.

Meyer, J. (2002). Task demands and responses to warnings. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 308-312.

Meyer, J. & Ballas, E. (1997). A two-detector signal detection analysis of learning to use alarms. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 186-189.

Robinson, D. E. & Sorkin, R. D. (1985). A contingent criterion model of computer assisted detection. In R. E. Eberts & C. G. Eberts (Eds.), *Trends in ergonomics/human factors II* (pp. 75-82). Amsterdam, North-Holland: Elsevier.

Selcon, S.J., Taylor, R.M., & Shadrake, R.A. (1991). Giving the pilot two sources of information: help or hindrance? In E. Farmer (Ed.), *Human Resource Management in Aviation* (p. 139-148). Aldershot: Avebury Technical.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.