2015

A Coalition Study of Warfighter Acceptance of Wearable Physiological Sensors

Lauren E. Menke

Christopher Best

Gregory J. Funke

Adam J. Strang

A COALITION STUDY OF WARFIGHTER ACCEPTANCE OF WEARABLE PHYSIOLOGICAL SENSORS

Lauren E. Menke[1], Christopher Best[2], Gregory J. Funke[3], Adam J. Strang[3]

[1]Ball Aerospace & Technologies Corporation, Fairborn, Ohio
[2]Defence Science and Technology Organisation, Melbourne, Australia
[3]Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio

Combat operations are often high tempo, resulting in undesirable levels of operator workload and stress. Adaptive automation has been suggested as a solution to these issues. However, this augmentation approach is predicated on operator consent to monitoring. Acceptance of such systems may be influenced by concerns regarding the use of monitor data and mistrust of automation technology. The purpose of the current investigation was to examine operator acceptance of physiological monitoring and future augmentation strategies after limited exposure to one device. During a simulated exercise, eleven command and control operators were equipped with a physiological monitor prior to each mission. Following the exercise, operators were surveyed regarding their acceptance of monitoring and several potential augmentation strategies. The results of the survey suggested that the operators were generally open to both monitoring and augmentation, but that they may also be insensitive to the limitations of current augmentation technology.

Military teams face increasingly difficult situations, characterized by high tempo operations, distributed team environments, long shift durations, high information throughput, and decision making under uncertainty (Chappelle et al., 2013). Concurrently, technological advances (e.g., for surveillance and monitoring, and cyber defense) are increasingly providing capabilities that will require rapid data processing and decision speeds that exceed human capabilities (e.g., Dahm, 2010). It has been suggested that factors such as these may result in human operators becoming a “bottleneck” in future military operations (e.g., Dorneich, Whitlow, Ververs, & Rogers, 2003).

In response to this challenge, military strategic guidance and planning documents (e.g., Dahm, 2010) suggest that human augmentation solutions need to be developed. A potential solution that has been suggested is adaptive automation (e.g., Dorneich et al., 2003). Adaptive automation is predicated on activation of assistive functions based on cues derived from operator behavior or physiology. Of particular interest is automation that is part of a data-driven feedback loop, wherein monitoring technologies track and assess physio-behavioral changes indicative of negative operator states (e.g., mental workload and fatigue; Galster & Johnson, 2013). This information can then be shared (e.g., with operators, mission commanders, other automated systems, etc.) as part of an augmentation strategy, perhaps resulting in dynamic task reallocation. However, for this approach to be viable, operators will need to be monitored during task performance. If risk factors such as stress and fatigue are to be considered (Caldwell, Caldwell, & Schmidt, 2008), extensive monitoring may be required, possibly including off-duty hours. These monitoring approaches will certainly require operator consent. This need resonates with the concept of *ubiquitous monitoring*, suggested by Moran and colleagues (2013), wherein behavioral, and potentially physiological, data are collected continuously from individuals for the purpose of monitoring and targeted intervention. Therefore, the success of adaptive automation as an augmentation strategy is contingent on operators’ acceptance of both monitoring and automation.

With regard to monitoring, research suggests that operator acceptance is likely to fall on a continuum of responses. At the “low” end of acceptance, participants may feel that monitoring is intrusive and reduces privacy (similar concerns have been raised regarding telemedicine, e.g., Beckwith, 2003; Rackett, 1997). For example, operators may fear the unwanted disclosure of health related information as a result of monitoring, particularly amid ongoing concerns regarding data privacy (e.g., Ahamed, Talukder, & Kameas, 2007). In addition, operators (e.g., aircrews) may fear that their duty status could be negatively affected by the discovery of ill-health.

A further concern may be feelings of discomfort or anxiety associated with perceptions of the presence of an evaluative “other,” such as a superior, colleague, or the monitoring system itself (e.g., Zeidner & Matthews, 2005). Research on evaluation anxiety (e.g., Zeidner & Matthews, 2005) suggests that under some circumstances, operator worries about evaluation may result in sufficient distraction to negatively impact task performance. The behavioral or physiological symptoms of such worries could result in the activation of the augmentation system, which in turn could reinforce and amplify their initial worries – creating an ongoing cycle of distraction and poor task performance.

At the “high” end of the acceptance continuum, operators may respond positively to continuous monitoring, particularly if they perceive that the benefits of the technology outweigh the risks (Moran et al., 2013). This may well be the case for military operators, considering that they are likely to a) be aware of emerging military doctrine concerning current and future reliance on automated systems, and b) have been affected by the difficult circumstances of current combat operations described previously. Operators may also endorse monitoring technologies if they are offered the opportunity to utilize the recorded physiological information for their own purposes, such as fitness or health management (e.g., Heron & Smyth, 2010).

An additional influencer of operator attitudes may be past experience with automated systems. For some operators, negative experiences with automation reliability (e.g., Parasuraman & Riley, 1997) and automation surprise (e.g., Sarter, Woods, & Billings, 1997) may elicit a general distrust in automation. There is also evidence that, under some circumstances, automation may actually increase operator workload, potentially resulting in operator underuse or disuse of augmentation technologies (Parasuraman & Riley, 1997). Finally, operators may have little understanding of the state of current automation technologies, and therefore have unrealistic expectations concerning system capabilities. Informed by popular media coverage, movies, and television, operators may believe that contemporary monitoring and automated augmentation technologies are more robust and advanced than they actually are. Similar beliefs have been expressed regarding perceptions of the capabilities of modern robots (e.g., Adams & Skubic, 2005).

Given these concerns, the purpose of the current experiment was to gauge operator opinions regarding their acceptance of monitoring and endorsement of several potential augmentation approaches. Participants in this study were a small group of Air Battle Manager (ABM) operators from the Royal Australian Air Force (RAAF) selected to take part in Exercise Black Skies (EBS; Best, Jia, & Simpkin, 2013). As part of the exercise, operators consented to physiological monitoring, providing them (limited) experience with monitoring upon which to base their ratings. We expected that operators would express general agreement to monitoring while performing their duties, and more limited approval of several augmentation strategies. Furthermore, we expected that support for specific strategies would be moderated by operational environment, with the highest endorsement during training, and reduced acceptance in more “real world” settings, such as combat missions. This would indicate a general openness of operators to emerging technologies, tempered by veridical assessment of current monitoring and augmentation capabilities and limitations.

Methods

Overview of Exercise Black Skies 2014 (EBS14)

Exercise Black Skies is a 5-day simulation training research exercise hosted by the Defence Science and Technology Organisation (DSTO) at their Air Operations Simulation Centre in Melbourne, Australia. While the specific training audience and scenarios are unique for each biannual instantiation of EBS, the broader objectives remain the same, which are to: 1) provide high-fidelity training to prepare ABM operators for a subsequent multinational, live training exercise (Exercise Pitch Black), and 2) serve as a test-bed for the development and evaluation of emerging technologies that might benefit current and future ABM operations and training.

The training audience for EBS14 included two sub-teams of ABM operators: a ground-based ABM unit (specifically, an Air Defence Ground Environment, or ADGE, unit) and an airborne unit (a mission crew from the E-7A “Wedgetail” airborne early warning and control aircraft). Participants in the exercise were 10 men and 1 woman. Their average age was 29.64 years (SD = 6.40; M_{ADGE} = 27.50, SD_{ADGE} = 6.47; $M_{Wedgetail}$ = 32.20, $SD_{Wedgetail}$ = 5.89). The ADGE team was composed of an Air Battle Director (ABD), a Tactical Director (TD), two Fighter Controllers (FCs), and two Picture Managers (PICMAN). The Wedgetail team was composed of a Mission Commander (MC), a Senior Surveillance and Control Officer (SSCO), and three Surveillance and Control Officers (SCOs).

Within these teams, the ABD, TD, MC and SSCO roles were leadership/supervisory roles, with the ABD and MC roles filled by the most experienced members (with 4,500 and 2,000 hours of controlling experience, respectively). The TD and SSCO roles were filled by the next most experienced operators (with 837 and 700 hours of controlling experience, respectively). The FC and SCO roles were tasked with tactical control of the aircraft within the team’s assigned airspace. Operators in these roles had less experience (averaging approximately 250 hours of controlling experience). The operators filling the PICMAN roles reported an average of approximately 4,000 hours experience.

While the functions and mission objectives of the two teams were mostly similar, there were several notable differences in their working environments. First, because different Command and Control interface systems are used by the RAAF in real ground-based and airborne environments, these systems were also different for the two EBS14 teams. Second, the physical configuration of the simulation facilities reflected those of each team’s typical work environment; the ADGE team sat in a semi-circular “weapons pit” arrangement of two rows (with team leaders seated behind members responsible for tactical control) while the Wedgetail team sat side-by-side (as is typical of the seating arrangement on the aircraft).

During EBS14, other command and control elements, as well as friendly and adversary airborne assets (e.g., fighter aircraft, air-lift aircraft, tankers), were simulated by an exercise “White Force” consisting of RAAF personnel and ex-military contractors. An important characteristic of EBS14 was that the mission scenarios used during the exercise were designed to simulate, in terms of airspace structure, airfield, target and sensor locations, friendly and adversary order of battle, mission types and unit roles, those that the operators would encounter several weeks later during the live exercise Pitch Black. This is noteworthy since Pitch Black is the RAAF’s largest and most complex air-combat exercise, making EBS14 a large, complex, and realistic simulation training event.

Physiological Monitoring System

During EBS14, operators consented to physiological monitoring of their responses to events in the simulation. They were told the information would be used to shape future simulation exercises and to develop augmentation technologies. It should be noted that although operators were provided an explanation for the physiological monitoring they experienced during EBS14, they were not provided information or feedback about their or their teammates’ particular physiological responses during the exercise, nor were they provided information about specific future augmentation technologies that might rely on such data.

Each operator wore a Zephyr BioHarness 3 (model BH3) during the exercise. The BioHarness is a lightweight physiological sensor designed to be worn against the wearer’s chest by means of a flexible synthetic strap (see Figure 1 for an illustration). The device was applied in accord with Zephyr’s instructions, i.e., the chest strap was aligned with the bottom of the operator’s sternum, and the recording module was located on the left side of the body in line with the operators’ armpit or slightly rotated to the back for comfort. The BioHarness records electrocardiographic (ECG), respiration, and accelerometry data (at 250, 100, and 25 Hz, respectively) and provides summary statistics once per second. Raw and summary data were recorded throughout each session to the onboard memory of the recording module. At the end of each session, data were downloaded from each operator’s module to a central database.

Figure 1. Zephyr BioHarness. The left image portrays the recording module (circular disk) and harness. The right image depicts proper placement of the harness.

Device Comfort Questionnaire (DCQ)

Following the final trial of EBS14, participants completed a novel measure, the *Device Comfort Questionnaire* (DCQ; see Appendix A). The DCQ is comprised of 19 items, representing 5 related subscales. Items of the first subscale, *device ergonomics*, relate to fit factors, such as simplicity of application and interference with task performance. The second subscale, *acceptance of physiological monitoring*, includes items related to operators’ perceptions of discomfort and intrusiveness associated with being monitored. Items of the final 3 subscales, *endorsement of use during simulation training exercises*, *live training exercises*, and *real operations*, ask operators to rate their degree of predicted acceptance of a future augmentation technology designed to utilize physiological monitoring data for a variety of purposes, including automatic adjustment of task difficulty, performance assessment, and workload monitoring. Items on the DCQ are rated on a scale of 1 (“Completely Disagree”) to 10 (“Completely Agree”). After reverse scoring relevant items (see Appendix A), subscale scores on the DCQ are computed by averaging across the pertinent item ratings.

Results

Mean operator ratings on the DCQ are presented below in Table 1, which depicts ratings aggregated (based on team role) into the categories of *lead* (ABD and TD, MC and SSCO) and *tactical* (FC and PICMAN, SCO) for the ADGE and Wedgetail teams, respectively.

Table 1.
Mean DCQ ratings by sub-team and team role.

	Team						
	ADGE			Wedgetail			
Mean Subscale and Item Responses	Lead	Tactical	**$\text{Mean}_{\text{ADGE}}$**	Lead	Tactical	**$\text{Mean}_{\text{Wedge}}$**	**Grand Mean**
Device ergonomics (Mean)	**8.67**	**7.92**	**8.29**	**7.17**	**7.44**	**7.31**	**7.80**
1. Not hindered performing duties	9.00	7.75	**8.35**	9.50	8.67	**9.08**	**8.73**
2. Device did not cause discomfort	9.00	8.75	**8.88**	5.00	8.00	**6.50**	**7.69**
5. Easy to put on and take off	8.00	7.25	**7.63**	7.00	5.67	**6.33**	**6.98**
Acceptance of physiological monitoring (Mean)	**10.00**	**9.63**	**9.81**	**9.00**	**9.00**	**9.00**	**9.41**
3. Device was not intrusive	10.00	9.75	**9.88**	8.50	8.67	8.58	**9.23**
4. Comfortable being monitored	10.00	9.50	**9.75**	9.50	9.33	9.42	**9.58**
Simulation training (Mean)	**8.17**	**9.50**	**8.83**	**6.17**	**7.44**	**6.81**	**7.82**
6. White force sets training difficulty	9.00	9.50	**9.25**	8.50	7.67	**8.08**	**8.67**
7. Identify debrief points	10.00	9.50	**9.75**	5.50	7.67	**6.58**	**8.17**
8. Automatically set training difficulty	6.00	9.50	**7.75**	5.50	7.67	**6.58**	**7.17**
9. Assessors make judgments	9.50	9.50	**9.50**	6.50	7.00	**6.75**	**8.13**
10. Inform lead about workload	9.00	9.50	**9.25**	6.50	7.67	**7.08**	**8.17**
Field exercise (Mean)	**8.00**	**9.50**	**8.75**	**6.00**	**7.44**	**6.72**	**7.74**
11. White force sets training difficulty	6.00	9.50	**7.75**	7.50	7.67	**7.58**	**7.67**
12. Identify debrief points	10.00	9.50	**9.75**	4.50	7.67	**6.08**	**7.92**
13. Automatically set training difficulty	6.00	9.50	**7.75**	5.00	7.67	**6.33**	**7.04**
14. Assessors make judgments	9.00	9.50	**9.25**	6.50	7.00	**6.75**	**8.00**
15. Inform lead about workload	9.00	9.50	**9.25**	6.50	7.67	**7.08**	**8.17**
Live operation (Mean)	**8.67**	**9.50**	**9.08**	**6.50**	**7.44**	**6.97**	**8.03**
16. Identify debrief points	9.50	9.50	**9.50**	8.00	7.67	**7.83**	**8.67**
17. Assessors make judgments	9.00	9.50	**9.25**	5.00	7.00	**6.00**	**7.63**
18. Inform lead about workload	8.50	9.50	**9.00**	6.50	7.67	**7.08**	**8.04**
19. Inform lead about fatigue	8.50	9.50	**9.00**	8.00	7.67	**7.83**	**8.42**
Grand Mean	**8.68**	**9.26**	**8.97**	**6.82**	**7.67**	**7.24**	

Perusal of Table 1 reveals several interesting effects. First, operators' ratings of the Zephyr BioHarness's *device ergonomics* were relatively high. Second, the ABMs indicated they were overwhelmingly accepting of the physiological monitoring they experienced during EBS. Third, when operators were asked to speculate about the future uses of physiological monitoring for adaptive aiding, they expressed high positive endorsement for the monitoring irrespective of the purpose or operational setting that the monitoring would be employed.

To further examine the data in Table 1 for differences in ABM operator ratings based on team and role across subscales, a 2 (team) × 2 (team role) × 5 (subscale) mixed analysis of variance (ANOVA) was computed. Results indicated a statistically significant main effect of team, $F(1, 7) = 31.06$, $p < .05$, $\eta_p^2 = .816$. No other effects in the analysis were statistically significant (all $p > .05$). Members of the ADGE team consistently provided higher agreement ratings on DCQ items than members of the Wedgetail team.

Discussion

The purpose of the current study was to provide an initial examination of operator response to physiological monitoring and potential future performance augmentation strategies. We expected that operators would express general agreement to monitoring while performing their duties, and more limited approval of the augmentation strategies. Further, we expected support for specific strategies would be moderated by operational environment. Our results suggest that operators were generally accepting of monitoring and endorsed the prospective augmentation strategies uniformly across operational environments. We also found that operator acceptance and endorsement was moderated by team; ADGE operators indicated higher agreement across items than did Wedgetail operators.

Contrary to initial predictions, the ABM operators were relatively accepting of physiological monitoring and agreed to usage of that data for all of the purposes and environments proposed. This may indicate that the perceived benefits of the proposed technology outweighed the perceived risks. Alternatively, it could suggest that operators may be unfamiliar with the capabilities and limitations of current (and near future) automated augmentation technologies. Whatever the underlying drivers may be, one consequence is relatively clear: operators are positive about future developments in monitoring and augmentation. It therefore behooves those of us working in the area to ensure that their expectations are appropriately calibrated against the actual capabilities of the systems we develop. Failure to do so is likely to result in violated expectations, mistrust, and disuse of future augmentation solutions.

Surprisingly, we found that Wedgetail operators expressed less acceptance of monitoring and augmentation than ADGE team operators. Though explanation of this effect is speculative, it could be due to disparities in operator experience across teams (ADGE team operators were generally more experienced than their Wedgetail counterparts), or other structural differences between the two groups. For example, the Wedgetail is a relatively new platform in the RAAF, and consequently those operators' attitudes may have been influenced by other factors, such as evolving organizational structure, mission requirements, and the relatively negative history associated with development of the aircraft (see e.g., Bergmann, 2013, for a brief history).

Alternatively, the observed differences in ratings may reflect differences in attitudes regarding deployment of electronic equipment. Wedgetail operators' ratings may be due to worries concerning electronic interference or safety considerations around wearing equipment in flight (e.g., it could hinder movement in the event of an emergency) – these are concerns that ADGE team operators would not necessarily share because of the ground-based nature of the unit. Yet, this explanation does not fully explain Wedgetail operator attitudes, as their ratings on the simulation training subscale of the DCQ were also lower than ADGE operators, even though simulation training exercises are not conducted on an aircraft. This may indicate that Wedgetail aircrew training has broadly sensitized them to issues of electronic interference and/or safety regardless of operational setting. This explanation has implications for other work environments. For example, if the Wedgetail operators' less-positive responses were driven by concerns about restricted movement during a crash or emergency egress, other aircrew may have similar concerns (e.g., fighter aircraft with ejection seat).

Overall, the operators surveyed in this experiment expressed high positive regard for future monitoring and augmentation approaches. Though substantial work is required to mature those technologies, it appears that operators are generally ready to accept them. In developing these devices, care must be taken to ensure that the capabilities and limitations of any such systems are communicated to operators, thereby appropriately calibrating their trust in and expectations of those devices.

References

Adams, J. A., & Skubic, M. (2005). Introduction to the special issue on human-robot interaction. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 35*, 433-437.

Ahamed, S.I., Talukder, N., & Kameas, A.D. (2007, September). *Towards privacy protection in pervasive healthcare*. Paper presented at the 3rd IET International Conference on Intelligent Environments (IE 07), Ulm, Germany. Paper retrieved from https://www.cs.purdue.edu/homes/ntalukde/papers/IE07.pdf

Beckwith, R. (2003). Designing for ubiquity: The perception of privacy. *Pervasive Computing, 3*, 40-46.

Bergmann, K. (2013, April 8). Wedgetail. Asia-Pacific Defense Reporter. Retrieved from http://www.asiapacificdefencereporter.com/articles/298/Wedgetail

Best, C., Jia, D., & Simpkin, G. (2013, October). *Air force synthetic training effectiveness research in the Australian context*. Proceedings of the NATO STO MSG 111 Multi-Workshop, Sydney, Australia.

Caldwell, J. A., Caldwell, J. L., & Schmidt, R. M. (2008). Alertness management strategies for operational contexts. *Sleep Medicine Reviews, 12*, 257-273.

Chappelle, W., McDonald, K., Christensen, J., Prince, L., Goodman, T., Thompson, W., & Hayes, W. (2013). Sources of occupational stress and prevalence of burnout and clinical distress among U.S. Air Force Cyber Warfare Operators (Report No. AFRL-SA-WP-TR-2013-0006). Wright-Patterson Air Force Base: Air Force Research Laboratory, Human Effectiveness Directorate.

Dahm, W.J.A. (2010). *Technology horizons: A vision for air force science & technology during 2010-2030* (Report No. AF/ST-TR-10-01). USAF HQ, Arlington, VA.

Dorneich, M.C., Whitlow, S.D., Ververs, P.M., & Rogers, W.H. (2003, October). *Mitigating cognitive bottlenecks via an augmented cognition adaptive system*. Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics, Washington, DC.

Galster, S.M., & Johnson, E.M. (2013). *Sense-assess-augment: A taxonomy for human effectiveness* (Report No. AFRL-RH-WP-TM-2013-0002). Wright-Patterson Air Force Base: Air Force Research Laboratory, Human Effectiveness Directorate.

Heron, K.E., & Smyth, J.M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behavior treatments. *British Journal of Health Psychology, 15*, 1-39.
Moran, S., Jaeger, N., Schnadelbach, H., & Glover, K. (2013, June). *Using adaptive architecture to probe attitudes towards ubiquitous monitoring*. Proceedings of the 2013 IEEE International Symposium on Technology and Society (ISTAS), Toronto, Canada.
Parasuraman, R., & Riley, R. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*, 230-253.
Rackett, C.M. (1997). Telemedicine today and tomorrow: Why "virtual" privacy is not enough. *Fordham Urban Law Journal, 25*, 167-191.
Sarter NB, Woods DD, Billings CE. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926-1943). New York, NY: Wiley.
Zeidner, M., & Matthews, M. (2005). Evaluation anxiety: Current theory and research. In A.J. Elliot & C.S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 141-163). New York, NY: The Guilford Press.

Appendix A: Device Comfort Questionnaire (DCQ)

Instructions to participants:
You wore a physiological monitoring device (i.e., a Zephyr BioHarness 3) during each VUL [trial] in EBS14. The purpose of these devices was to help us monitor how hard you were working during each VUL, with the idea that we could use that information to help shape future Black Skies exercises. Given that, in the following questions we are interested in your level of comfort wearing the system.

Please rate the following statements about the device on a scale from:
1 = "Completely Disagree" to **10 = "Completely Agree"**

1. I was not hindered by the device while performing my duties.
2. I felt that wearing the device caused discomfort.*
3. I felt that wearing the device was intrusive.*
4. I felt uncomfortable being monitored.*
5. I felt that the device was easy to put on and take off.

I would feel comfortable having physiological data, such as that collected during EBS14, used in a future **simulation training exercise** (e.g. Black Skies) to:

6. Help white force set or change training difficulty.
7. Help identify debrief points for after action review.
8. Automatically set or change training difficulty.
9. Help expert assessors make judgments about my performance.
10. Help inform my team leader about my workload during a mission.

I would feel comfortable having the physiological data used in a **live training exercise** (e.g. Pitch Black) to:

11. Help white force set or change training difficulty.
12. Help identify debrief points for after action review.
13. Automatically set or change training difficulty.
14. Help expert assessors make judgments about my performance.
15. Help inform my team leader about my workload during a mission.

I would feel comfortable having the physiological data used during **real operations** to:

16. Help identify debrief points for after action review.
17. Help expert assessors make judgments about my performance.
18. Help inform my team leader about my workload during a mission.
19. Help inform my team leader about my level of fatigue during a mission.

Note. Items marked with an asterisk (*) are reverse scored.
Scoring the DCQ
The DCQ includes five dimensions: *device ergonomics* (mean rating of items 1, 2, & 5); *acceptance of physiological monitoring* (mean rating of items 3 & 4); *endorsement of use during simulation training exercises* (mean rating of items 6-10); *endorsement of use during live training exercises* (mean rating of items 11-15); *endorsement of use during real operations* (mean rating of items 12-19).