USING PAIR-WISE RANKINGS IN THE ASSESSMENT OF ADAPTIVE AIDING

Christina Gruenwald
Oak Ridge Institute for Science and Education
Belcamp, Maryland
Matt Middendorf
Middendorf Scientific Services
Medway, Ohio
Chelsey Credlebaugh
Ball Aerospace and Technologies Corp.
Fairborn, Ohio
Jonathan Mead
Oak Ridge Institute for Science and Education
Belcamp, Maryland
Scott Galster
Air Force Research Laboratory
Wright-Patterson AFB, Ohio

In remotely piloted aircraft (RPA) operations, operator cognitive workload is an important concern. High workload could result in performance decrements and operational mishaps. In research, physiological data can be used by models to assess the operator's cognitive state. When a model detects the onset of cognitive overload, assistance could then be provided to the operator to help mitigate the overload in some form of augmentation. However, it is imperative that the assessment is accurate and completed in a timely manner. The accuracy of a workload assessment model and augmentation application can be evaluated using a psychometrically determined scale of man/machine conditions. Both the operator and machine can be in various conditions at any point in time. In three prior studies, eighteen participants were asked to perform pair-wise rankings of sixteen conditions to generate the rankings. These rankings will be used to evaluate the accuracy of the workload assessment model in future research.

Operator cognitive workload is an important concern in remotely piloted aircraft (RPA) operations. RPA use is increasing for missions in hostile environments, or those considered too dangerous for manned aircraft (U.S. Department of Defense, 2011). This places more importance on monitoring the cognitive state of the RPA operators. When task demands are high and cognitive overload occurs, operator performance can suffer and lead to mission failure (Young & Stanton, 2002).

The Sense-Assess-Augment (SAA) paradigm was developed for use in research studying operator cognitive state (Galster & Johnson, 2013). In general, the paradigm serves to *sense* physiological measures, *assess* the state of the operator, and if necessary, provide tools to *augment* the operator's performance.

The most complex aspect of the SAA paradigm is within the assessment stage. Models must be able to accurately assess cognitive state using physiological measures. If the assessment is inaccurate, the augmentation strategy applied may not be beneficial. In the current research, the SAA paradigm is applied to simulated RPA operations in order to prevent detrimental errors that could occur during a mission as a result of cognitive overload.

Adaptive augmentation is best suited for RPA operations to mitigate high workload. The Yerkes-Dodson Law states that workload must fall within a middle range (not too low and not too high) in order to achieve optimal performance (Cohen, 2011). Augmentation must be adaptive for two reasons. First, if

DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited. 88ABW Cleared 02/09/2017; 88ABW-2017-0541.

augmentation is not necessary for the operator's current cognitive state, the resources necessary for the augmentation would be unnecessarily tied up. Secondly, if the operator's cognitive state is already in the middle range of workload, we do not want to provide an augmentation that will allow the operator to fall into a cognitive underload state. A low workload state has been shown to be just as detrimental to performance as having high task demands (Desmond & Hoyes, 1996).

In past research, artificial neural network (ANN) models have been used in an attempt to model cognitive workload (Hoepf et al., 2016). These models can be used to trigger augmentation and close the SAA loop. This report focuses on the development of a methodology that can be applied to studies that use an assessment model to determine if augmentation is being triggered at the correct time. Further, augmentation should only be provided when it is truly needed and not when the workload level is already manageable. Previous research has used control groups such as yoked or random augmentation to determine if augmentation is being applied correctly (e.g., Bailey, Scerbo, Freeman, Mikulka, & Scott, 2006). However, in field operations, having experimental trials is not possible and there will still be a need for ensuring accurate cognitive state assessment. The development of the pair-wise rankings methodology and how it can be applied to future research is reported.

Method

Participants

A total of 18 individuals from three previous studies participated in the pair-wise rankings evaluation after completion of the studies. Eleven of the participants were male and seven were female. The average age of the participants was 21.3. All participants read and signed an informed consent document prior to participating. Study procedures were reviewed and approved by the Air Force Research Laboratory (AFRL) Institutional Review Board.

Task

Prior to the pair-wise rankings, each participant completed one of three studies that each consisted of surveillance, tracking, and communication tasks. The main differences between the three studies were the physiological measures being collected and modeling techniques used to estimate workload. Therefore, the experimental tasks and manipulations were consistent across these studies. All studies consisted of two 2 x 2 within-subjects designs.

For the surveillance task, participants were instructed to search a market place to find four high value targets (HVTs). The HVTs carried a sniper rifle whereas all other pedestrians in the market place were distractors: not holding anything, holding a handgun, or holding a shovel. Once the HVT was found, participants pressed the F key and tracked the target (i.e., kept the target on screen) until he went under a tent. Participants would then proceed to search and find the next HVT. Only one HVT was present in the scenario at a time. Once an HVT was found, participants started accumulating points. Therefore, the sooner the participant found the HVT, the higher performance score they received.

There were two experimental manipulations, each consisting of two levels for the surveillance task: distractors and fuzz. Distractors could be low or high, meaning there were either 16 or 48 other entities walking around the market place during the trial. The fuzz manipulation could either be off or on. When fuzz was off, the camera feed was clear and it was fairly easy to identify the HVT. However, when fuzz was on, the camera feed was degraded.

For the tracking task, participants were instructed to follow HVTs that were traveling by motorcycle. The two manipulations for the tracking task were number of HVTs and route the HVT was traveling. Participants would either track one or two HVTs. Two HVTs required continuous clicking back and forth on two camera feeds. The HVT(s) would either travel a country route, which consisted of a paved, straight road, or a city route in which the HVT(s) turned often and could be occluded by buildings. Performance for the tracking task was based on keeping the HVT(s) in the camera feed. In addition, more points were awarded for keeping the HVT(s) closer to the center of the feed, compared to the edges.

In conjunction with the surveillance and tracking tasks, participants had a secondary task to perform. Evenly distributed throughout the trial, mental math questions were asked over the headset. These questions consisted of altitude change, travel time, and speed change questions relative to the RPA. For example, if the RPA was traveling at 40 knots, a question might ask how long it would take to arrive at a location 100 nautical miles away with a headwind of 15 knots.

Man/Machine Conditions

Throughout the experiment, participants experienced different man/machine conditions. Both the man and machine could be in various conditions depending on the experimental manipulations and the specific point of time within a trial. There are 16 possible combinations of various man and machine conditions for both the surveillance and tracking tasks as seen in Table 1 and Table 2.

For surveillance, the machine could be in a condition of fuzz on or off and distractors high or low (experimental manipulations). The man could be in a condition of either looking for an HVT or tracking the HVT until it goes under a tent. The man could also be in a condition of answering a communication question or not having a question to answer. Although the machine condition was consistent throughout each trial, the man condition could change within a trial.

Similarly, for tracking, the machine could be in a condition of tracking one or two HVTs, along a country or city route (experimental manipulations). The man could be in a condition of either successfully tracking the HVT or searching for a lost HVT. Also, the man could be answering or not answering a communication question. Identical to the surveillance task, the machine condition was constant throughout each trial, but the man condition varied.

Table 1.
Possible Man/Machine Conditions for the Surveillance Task.

Condition	HVT	Question	Fuzz	Distractors
A	Looking	Yes	On	High
B	Looking	Yes	On	Low
C	Looking	Yes	Off	High
D	Looking	Yes	Off	Low
E	Looking	No	On	High
F	Looking	No	On	Low
G	Looking	No	Off	High
H	Looking	No	Off	Low
I	Tracking	Yes	On	High
J	Tracking	Yes	On	Low
K	Tracking	Yes	Off	High
L	Tracking	Yes	Off	Low
M	Tracking	No	On	High
N	Tracking	No	On	Low
O	Tracking	No	Off	High
P	Tracking	No	Off	Low

Table 2.
Possible Man/Machine Conditions for the Tracking Task.

Condition	Tracking	Question	Route	Targets
A	Lost	Yes	City	2
B	Lost	Yes	City	1
C	Lost	Yes	Country	2
D	Lost	Yes	Country	1
E	Lost	No	City	2
F	Lost	No	City	1
G	Lost	No	Country	2
H	Lost	No	Country	1
I	Tracking	Yes	City	2
J	Tracking	Yes	City	1
K	Tracking	Yes	Country	2
L	Tracking	Yes	Country	1
M	Tracking	No	City	2
N	Tracking	No	City	1
O	Tracking	No	Country	2
P	Tracking	No	Country	1

Procedure

After completing each experiment, participants performed the pair-wise comparison survey. Each of the 16 man/machine conditions were compared to each other. Participants simply selected which of two conditions had higher workload. There were 120 comparisons for each task.

The number of times a condition was selected as the higher workload condition was summed for each participant. This number was then averaged across participants producing an overall ranking of the man/machine conditions.

Results

The results from the pair-wise comparisons are reported in Table 3 for the surveillance task. It is clear that the largest factor to affect workload was whether the participant was looking for the HVT or had already found the HVT and was tracking it. This is evident by “Looking” for the HVT being in the top seven rankings. The lowest workload man/machine condition was tracking the HVT, with no question being asked, fuzz off, and distractors low, as anticipated. In contrast, the highest workload man/machine condition was looking for the HVT, answering a question, fuzz on, and distractors high.

Table 4 shows the results from the pair-wise comparisons for the tracking task. It was rated that when a target was lost, a question was being asked, and they were tracking two targets in the city (Condition A) was the highest level of workload. Conversely, Condition P had the lowest level of workload.

Table 3.
Surveillance Man/Machine Rankings.

Condition	Rank	Average	HVT	Question	Fuzz	Distractors
A	16	14.9	Looking	Yes	On	High
E	15	13.3	Looking	No	On	High
B	14	12.3	Looking	Yes	On	Low
C	13	11.7	Looking	Yes	Off	High
F	12	10.0	Looking	No	On	Low
G	11	9.7	Looking	No	Off	High
D	10	8.3	Looking	Yes	Off	Low
I	9	8.3	Tracking	Yes	On	High
H	8	6.5	Looking	No	Off	Low
J	7	5.9	Tracking	Yes	On	Low
K	6	5.7	Tracking	Yes	Off	High
M	5	5.5	Tracking	No	On	High
L	4	2.8	Tracking	Yes	Off	Low
N	3	2.6	Tracking	No	On	Low
O	2	2.3	Tracking	No	Off	High
P	1	0.1	Tracking	No	Off	Low

Table 4.
Tracking Man/Machine Rankings.

Condition	Rank	Average	Tracking	Question	Route	Targets
A	16	14.8	Lost	Yes	City	2
E	15	13.4	Lost	No	City	2
C	14	12.3	Lost	Yes	Country	2
I	13	10.9	Tracking	Yes	City	2
G	12	10.3	Lost	No	Country	2
M	11	9.2	Tracking	No	City	2
B	10	9.0	Lost	Yes	City	1
K	9	8.1	Tracking	Yes	Country	2
F	8	7.6	Lost	No	City	1
D	7	6.3	Lost	Yes	Country	1
O	6	5.6	Tracking	No	Country	2
J	5	4.4	Tracking	Yes	City	1
H	4	4.1	Lost	No	Country	1
N	3	2.3	Tracking	No	City	1
L	2	1.7	Tracking	Yes	Country	1
P	1	0.1	Tracking	No	Country	1

Discussion

The pair-wise ranking methodology gives insight into how the man/machine conditions compare to each other. For the surveillance task, it was determined that looking for the HVT usually ranked as higher workload than tracking the HVT after it was already found. Similarly, for the tracking task, tracking two targets was harder than tracking one target in most conditions.

This methodology can be applied to future experiments to evaluate the accuracy of the workload assessment model and the activation of augmentation. This can be accomplished by taking the average of the man/machine condition when the augmentation was triggered and comparing it to when performance decrements occurred or if the man/machine ranking was in the upper portion of the rankings. This information can then be used to evaluate if ANN models are activating the augmentation at the correct times. The augmentation should be provided when the man/machine condition was in a condition that was rated as having higher workload, although the point in which augmentation is needed can vary due to individual differences in performance. In conclusion, the pair-wise ranking methodology will be applied to ongoing and future studies, in addition to other statistical analysis to ensure modeling and adaptive augmentation is working properly.

Acknowledgements

The authors would like to thank Paul Middendorf, Kevin Durkee, Avi Hiriyanna, and Noah Depriest for assistance in data collection and software support.

This research was supported in part by an appointment to the Student Research Participation Program at the U.S. Air Force Research Laboratory (USAFRL) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USAFRL. The views expressed in this report are solely those of the authors and do not necessarily reflect the views of the employers or granting organizations.

References

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors*, *48*, 693-709. doi: 10.1518/001872006779166280

Cohen, R. A. (2011). Yerkes–Dodson Law. In *Encyclopedia of clinical neuropsychology* (pp. 2737-2738). Springer New York. doi: 10.1007/978-0-387-79948-3_1340

Desmond, P. A., & Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task: evidence for compensatory effects. *Safety Science, 22*(1), 87-101. doi: 10.1016/0925-7535(96)00008-2

Hoepf, M., Middendorf, M., & Mead, J., Gruenwald, C., Credlebaugh, C., Middendorf, P. & Galster, S. (2016). Evaluation of Physiologically-Based Artificial Neural Network Models to Detect Operator Workload in Remotely Piloted Aircraft Operations (Report No. AFRL-RH-WP-TR-2016-0075) Wright-Patterson Air Force Base: Air Force Research Laboratory, Human Effectiveness Directorate.

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.

U.S. Department of Defense. (2011). Unmanned systems integrated roadmap FY2011-2036 (Reference No. 11-S-3613). Washington, DC: Department of Defense.

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science, 3,* 178-194. doi: 10.1080/14639220210123789