Physiological Based Adaptive Training

Thomas Schnell¹, Cory J. Reuter¹, Ezekiel D. Gunnink¹, Bradley M. Parker¹, Carlton H. Richey¹, Jaclyn A.

Hoke², Jason D. Moss³

Operator Performance Laboratory (OPL)¹, Rockwell Collins², U.S. Army Research Laboratory (ARL)³

INTRODUCTION

The Army faces an emerging adversary environment that is very competitive, dangerous, and cognitively intense. To address this challenge, Army soldiers must out-learn and out-train their adversaries and this training challenge must be met in a climate of austere or shrinking training budgets (Army, 2011a, 2011b). The gold standard in training is one-to-one human tutoring, which has been shown to be significantly more effective than the one-many method of instruction such as the traditional classroom setting or self-study using static training materials such as manuals and books (VanLehn K, 2011). The proliferation of computer based games including massively multiplayer online games (MMOG), low-cost simulations, and exciting virtual immersion technologies opens new doors in the training domain. Additionally, considerable progress has been made in areas that include training pedagogy, methods of instruction/feedback, artificial intelligence, virtual humans, and trainee state assessment. Through a well-crafted learning concept roadmap, the Army plans to leverage those technological game changers to create systems that will allow self-paced, adaptive training capabilities that will enhance training effectiveness while at the same time be very cost effective. To address this challenge, the Army Research Lab has developed the Generalized Intelligent Framework for Tutoring, which is known as GIFT (Sottilare R, Sinatra A, & Boyce M, 2015). In our project, we are adding a component to GIFT that uses the Cognitive Assessment Tool Set (CATS) as a system to acquire the real-time operator state (Ellis K.E, 2014; T. Schnell, 2012; T. Schnell & Engler, 2013; T. Schnell, Melzer, & Robbins, 2009). This includes task technical performance, cognition (workload, engagement), and attention (degree of focus), so that the training content can be adapted through GIFT to maximize training effectiveness. We selected a demonstration use case centered on self-study driving instruction training for military vehicles, particularly for the High Mobility Multipurpose Wheeled Vehicle (HMMWV). Initially, we are using the Virtual Battle Space (VBS3) as the driving simulation tool. In Year 2 of this project, we will migrate the GIFT Framework into our instrumented Model 997 HMMWV Off-Road Testbed for testing in a real-world off-road environment.





a. VBS 3 HMMWV Simulator

b. OPL's Instrumented HMMWV/Simulator

Figure 1. Physiological Based Adaptive Training using GIFT Framework for HMMWV

The training scenarios in our adaptive training concept progress in difficulty from simple driving tasks on a flat and level tarmac to complex urban navigation and off road maneuvering assignments. This specific

use case was developed after various training domains and application domains have been reviewed, alternatives were developed, and a down-selection was performed to arrive at the particular driver training use case which will form the basis of the testbed in this project.

THE PROBLEM

Current training tools do not usually have an ability to acquire trainee data beyond simple performance data (e.g. right and wrong answers). Therefore, current training systems are generally not able to associate trainee state to specific elements of instruction. In one-to-one human tutoring settings, the instructor observes the student's performance and exterior performance indicators such as body language, facial expressions, head position, and hand movements to make a determination if the trainee is on the right path to acquiring the skill. For example, an Instructor Pilot (IP) may carefully and unobtrusively observe his/her flight students during landings to see if they are referencing the correct instruments and perform the correct manual movements for this phase of flight. Through such exterior performance observations, the IP can assess trainee state in real time and take corrective action, if necessary. Such actions could include physical interventions (e.g. to prevent a crash), the provision of explanations, a decision to repeat the task, or a decision to abandon the task and allow the student to rest. Unfortunately, there are circumstances where it is impossible for an instructor to discriminate with external indicators alone, if a student has frozen up or if he/she is cool and in control but is not currently moving around. Additionally, instructors are a limited resource and it is not feasible to have one-on-one tutoring in all training settings.

Therefore, the Army is looking for a data driven approach that will automatically and unobtrusively assimilate trainee information and then reliably and automatically classify trainee state including performance, cognition (workload, engagement), attention (degree of focus), and affect (joy, confusion, frustration, boredom, surprise, and anger) so that the training content can be adapted to maximize training effectiveness.

PHYSIOLOGICAL BASED TRAINEE STATE MODULE

In the project described in this paper, we are adding a component to GIFT that uses the Cognitive Assessment Tool Set (CATS) (OPL, 2014) as a system to acquire the real-time cognitive workload of the trainee to close the loop, through GIFT, with the training application. This means that the workload experienced by the trainee affects the progression of the training application. We selected a demonstration use case centered on self-study driving instruction training for military vehicles, particularly for the HMMWV. We are using the Virtual Battle Space (VBS3) as the driving simulation tool. We call this combination of GIFT, CATS, and a training application, in the specific use case VBS3, the Unobtrusive Physiological Classification and Adaptive Training (UPCAT) system. This specific use case was developed after various training domains and application domains have been reviewed, alternatives were developed, and a down-selection was performed to arrive at the particular use case which will form the basis of the testbed in this project. In Year 1 of this project (current year), we are integrating CATS and GIFT with the Virtual Battle Space (VBS 3) simulation tool. This constellation will allow us to test the adaptive capabilities of GIFT instruction on the basis of a simulated driving task. In the following project year (Year 2), we will migrate the framework into our instrumented Model 997 HMMWV (see Figure 1). This vehicle can be used as an Automobile-In-Loop (AIL) simulator and it can also be driven on and off-road as a human factors driving research testbed.

Cognitive Assessment Tool Set (CATS)

Understanding and monitoring the changes in the cognitive workload of trainees can offer critical quantitative information about their progression and performance. Unfortunately, accurate real-time objective quantification of cognitive workload using physiological signals has, thus far, proven elusive and is often neglected in favor of subjective self-reports. In well over a decade of physiological based assessment work, we investigated many sensors and came away with the conclusion that the electrocardiogram (ECG) waveform is by far the best signal for workload assessment. Based on our extensive real-world data collection experience, we discourage the use of invasive sensors such as electroencephalogram (EEG) for operational training contexts. The test-retest validity of these EEG appliances is usually very poor, approaching chance probability of prediction. For our ECG based workload assessment, we are using a deterministically nonlinear dynamical classifier to assess cognitive workload with great success (Engler & Schnell, 2013; T Schnell & Engler, 2014). The research community has known for a number of years that human physiological signals in general, and ECG specifically, are deterministically nonlinear (also known as chaotic) systems (Govindan, Narayanan, & Gopinathan, 1998; Kozma, 2002; Owis, Abou-Zied, Youssef, & Kadah, 2002). Chaotic systems are often not well represented via the normal scalar time series. Instead, the dynamics of the system are obfuscated in the single dimension whereas they become apparent when a transform of the data is made. This transform moves the data from the single dimensional scalar space into a multidimensional embedded phase space (Richter & Schreiber, 1998). The transformation to phase space using the mutual information and false nearest neighbor techniques can be illustrated nicely with an ECG signal. The panel on the left of Figure 2 depicts a portion of an ECG signal from a subject in a recent study. After calculating the parameters as described above, the phase space can be generated with time delay $\tau = 8$ and embedding dimension d = 3. The panel in the middle of Figure 2 shows the phase space that is generated from the signal using the methods described above. The image of the phase space does not necessarily illicit new knowledge about the ECG signal in and of itself. However, the phase space can be coarse-grained (right panel in Figure 1) into a numerical array that represents a quantitative signature of operator state in ECG phase space and thus offers the possibility for accurate operator state characterization. We refer to this as the Chaotic Physiological Classifier (CPC) method.



Figure 2. Example of a Scalar ECG (left) Transformed into Embedding Phase Space (middle) and then Coarse-Grained for Numerical Classification (right).

The CATS backbone is a relational database which forms the repository of all data collected during a study. Data is collected via providers within CATS which form communications links between the relational database and the sensing hardware. All data that is collected in CATS is time stamped at the source of the data with a globally synchronized time stamp. The relational database is then structured such that the time stamps form a candidate key for each table, thereby inherently synchronizing the data as it is recorded. Multiple tables exist within the relational database representing multiple signal sources such as vehicle state, environmental (simulator) state, eye tracking, and ECG. Each of these tables is linked through foreign keys indicating which subject, vehicle, and task to which each record is linked. This form of candidate and foreign keys forms a robust, indexable data backbone for the operator state classification effort. In addition to recording data, the CATS system calculates certain metrics in real-time. These metrics are then available to be shared, in real-time, with research partners through the previously mentioned communications portals. CATS uses a CPC model to produce the workload classification based upon the real-time input of ECG data.

UPCAT System Architecture

Figure 3 shows the architecture of UPCAT and its connectivity to the GIFT framework. As shown in this (greatly simplified) diagram, CATS receives data from the ECG sensor (Nexus 4 made by MindMedia) through its standard sensor provider that makes it manufacturer independent. Inside of CATS, the Chaotic Physiological Classification (CPC) (OPL, 2014; T Schnell & Engler, 2014) embeds the time-series ECG data in phase space and applies the ergodicity classification to it. In this context, it is easiest to think of CATS as a processor that translates full ECG waveforms to cognitive workload numbers. The real-time workload number is passed to a processor in CATS which aggregates the rapidly fluctuating number into a relatively stable score that indicates the degree of trainee engagement in the task. This score is transmitted to the Workload Condition in GIFT. The workload measured indicates a type of effort expenditure of the trainee. This expenditure yields a level of driving performance as a function of experience level. A novice driver may expend a significant effort to achieve a relatively low level of driving performance. As training iterations are performed at a certain difficulty level (as driven by the scenario), the effort expenditure should decrease and the driving performance should increase. At some point, both metrics may plateau and if driving performance and workload expenditures are considered satisfactory, the trainee is advanced to the next level of difficulty, either within the scenario or by switching to a more difficult scenario in VBS3. Driving performance is assessed through quantitative metrics that relate to automatically measurable outcomes such as speed maintenance, lane control, steering wheel rotation entropy, etc.



Figure 3. UPCAT System Architecture

The VBS 3 training scenarios in UPCAT progress in increasing levels of difficulty, much like advancing through chapters of a book. At strategic points in the scenarios, the driving events are stopped (frozen) and the trainees fill out surveys provided by GIFT to indicate the level of self-rated performance and workload expenditure. After the surveys are completed, the responses are combined with the physiological based data and the driving performance results to make decisions on how the scenario should proceed. In GIFT, the following steps are needed to accomplish this: 1). Create the workload scoring filter, 2). Create the workload scoring condition, 3). Create the VBS3 scoring filter, 4). Create the VBS3 scoring condition, 5). Create the surveys per scenario, 6). Create the survey scoring condition, 7).

interface (VBS 3 scenarios), 8). Combine the workload and VBS3 filters into a Domain Module, 9). Transmit the information via a Learner Module into a Pedagogical Module so that we can act on these conditions as a group, 10). Create a Domain Knowledge File (DKF, a set of rules for performance) and author Triggers to change scenarios and trigger difficulty changes. GIFT has a VBS3 interface that allows transmission of commands to a running VBS3 instance. These commands include calling a script remotely and reading out the results of its execution. In our architecture, this facility is used to extract driving performance score calculations out of VBS3.

The GIFT system is implemented as constellation of state machines in different states. The GIFT Intermission Stage (shown in Figure 4 a) handles the transitions between scenarios by analyzing the results of previous scenarios. This allows us to adaptively increase or decrease the difficulty of the scenarios as a whole unit. The GIFT Run Stage in (shown in Figure 4 b) adjusts the difficulty within the current scenario in real time as it is being run. These adaptive changes in difficulty will be smaller incremental changes than when the scenario difficulty is changed as a whole unit. Scripts inside of VBS 3 respond to changes in scenarios and difficulty according to the script state machine (shown in Figure 4 c) in response to the other GIFT stages.



Figure 4. State Machine Diagrams for GIFT Implementation in UPCAT

UPCAT CONCEPT OF OPERATIONS (CONOPS)

The following is the CONOPS narrative that describes what the UPCAT system, once completed and when applied for the HMMWV driving use case, should be able to do. This CONOPS has driven our architecture design and will help us to complete the UPCAT system in accordance with set requirements. The CONOPS is a narrative that describes how we want the finished UPCAT system to work. While the CONOPS is relatively detailed and thorough, we will focus the effort in this project on the physiological based adaptive training capability and only "rough in" some of the training environment capabilities described in this CONOPS. We developed a detailed visual storyboard that describes the graphical content of the UPCAT screens that the trainee would see. Some selected images from that storyboard are represented as figures hereinafter.

The expected trainee is an army recruit who has a valid US driver license and about 3-4 years of on-road driving experience on normal US highways under 4 season day and night driving conditions. As a baseline,

we assume that the trainee has no prior off-road driving experience and no driving experience in foreign countries.

The trainee is assigned to an UPCAT workstation where he/she logs in and starts the enrollment process using an interactive screen to fill in information. The trainee enters pertinent information about his/her person to facilitate tracking of course credit. Additionally, the trainee enters information related to his/her driving experience such as number of years driven, area where driving was performed, urban vs rural driving, day vs nighttime, driving on snow, type of vehicle, etc. This is done to establish a baseline database of driving exposure. The UPCAT system then provides the trainee with an instructional video that illustrates how the UPCAT ECG sensor is to be applied. This shows attachment of the electrodes using a schematic view of a person's torso to be sure the electrodes and leads are attached correctly. The video then stops to give the trainee a chance to set up the electrodes and go to the next screen. UPCAT then tells the trainee how to start CATS and verify ECG signal accuracy (Figure 5). Next, the trainee goes through a set of slides that introduce the HMMWV. This is basically a Computer Based Training (CBT) user manual review introducing the HMMWV controls. Once the trainee completes the basic CBT, a quiz will be administered (Figure 6) to ensure the trainee is ready to progress to the first driving simulator module.



Figure 5. UPCAT CBT and Sensor Application



Figure 6. UPCAT Vehicle Familiarization

The first driving simulator module (Level 1) is a simple drive on a very large parking lot or tarmac without obstacles. Using graphical interactive content (Figure 7, left) the trainee is told to drive around the parking lot perimeter in a clockwise direction, one car width away from the apron edge, at an appropriate speed, not to exceed 20 MPH. Performance are measured by UPCAT to ensure that the trainee has maintained the speed and positional assignment. CATS is used to assess workload to ensure that sufficient replications of the drive around the tarmac have been completed. The trainee is considered ready for the next level when workload has levelled off and driving-technical performance is within boundaries. A point score is then calculated from the performance metrics as credit similar to the score in a video game (Figure 7, right).



Figure 7. UPCAT Simple Driving Task (Level 1)

Upon completion of the first level, the point score and other statistics are shown. Feedback is provided by GIFT using trigger points such as a) great job, all is well, b) watch your speed, c) watch your lane control etc. These feedback points are illustrated with performance graphs and verbal narratives from canned AVIs playing an instructor (Gunny) chastising or praising the student (Figure 7, right). The trainee then proceeds to the next level and the process repeats for as many levels as needed by the particular use case. In our project, the progression of scenarios may look as follows:

- Level 2: Parking lot with obstacles placed to drive around.
- Level 3: Driving on an open, mostly straight highway in a foreign country with appropriate visuals and a simple navigational assignment.
- Level 4: Addition of curves and reasonable up and down grades.
- Level 5: Addition of urban areas.
- Level 6: Addition of roadway threats to avoid, requiring severe braking and swerving.
- Level 7: Addition of off-road, straight up and down grade.
- Level 8: Addition of off-road, along grade (slant), left and right.
- Level 9: Addition of driving at nighttime and in degraded visibility conditions.
- Level 10: Addition of IED detection and avoidance.
- Level 11: Addition of ambush event with backup retreat.
- Level 12: Capstone driving event that is assembled from all the parts that the trainee did not do well on.

Provisions should be made so that a driving session can be interrupted and taken up again. We still will need to determine how feasible this is with regard to physiological based workload assessment without baseline. It may be necessary to repeat the level upon resuming after a long break (e.g. days).

Once all levels on the driving simulator have been completed, the GIFT training record is forwarded to the driving instructor for review and scheduling of the first live driving lesson. The idea of the live drive is that UPCAT rides along, permitting the use of a safety observer who is not a qualified driving instructor but rated in the vehicle only. Therefore, staffing is easier because the safety observer does not need to have the qualities of an instructor as that job will be done by UPCAT. Instead, the safety observer simply monitors the drive with regard to safety. In the live drive, UPCAT receives vehicle state not from VBS3 but from the vehicle inertial system.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The GIFT architecture facilitates the integration of external tools such as VBS 3 and CATS in a very effective fashion. CATS is an operational workload and performance assessment system that has been used by OPL in real-world driving and flight contexts for a number of years. CATS has been used many times to assess the performance of fighter pilots in OPL's instrumented jet aircraft or in flight simulators at OPL and numerous government research facilities. In this project, we are using the workload assessment capability of CATS and integrate it with the GIFT framework using a Direct Link Library (DLL) methodology.

At the time of writing this paper we are about 5 months into the first program year. We have finished the architecture design and implemented an initial prototype in accordance with Figure 3. In the remainder of Year 1, we will complete the initial UPCAT prototype and demonstrate the physiological based adaptive training scenario capability.

For the Year 2 effort, we are planning to test and evaluate the UPCAT system using N=12 participants undergoing a full-mission training evolution as described in the CONOPS. There are several research questions that this experiment will seek to answer.

- 1. Is the UPCAT workload assessment accurate (absolute) and precise (narrow distribution) when compared to a known or self-assessed baseline workload scale?
- 2. Is the UPCAT affect assessment accurate (absolute) and precise (narrow distribution) when compared to a self-assessed baseline affect scale (e.g. joy, confusion, frustration, boredom, surprise, and anger)?
- 3. Is the adaptive portion of the training more effective than its non-adaptive counterpart?
- 4. Is the base program UPCATS/GIFT system acceptable for actual training in an Army context?

Research questions 1-3 will be answered though experimental hypotheses resulting from a full factorial experiment with assessments performed using appropriate statistical tests. Research question 4 will be answered through analysis of debriefing interviews and with the use of subjective preference rating questionnaires. The experimental hypotheses will be structured along the following lines in accordance with the research questions:

- 1. Experimental hypothesis EH₁:
 - a. H₀: workload assessment error is less than 10% of baseline
 - b. H₁: workload assessment is higher than 10% of baseline
 - Independent variable: Workload driver (rest, low, medium, high, very high)
- 2. Experimental hypothesis EH₂:
 - a. H_0 : affect assessment error per emotion is less than 10% of baseline
 - b. H₁: affect assessment error per emotion is higher than 10% of baseline

Independent variable: Affect driver (story) at levels of joy, confusion, frustration, boredom, surprise, and anger

- 3. Experimental hypothesis EH₃:
 - a. H_0 : adaptive training performance = non-adaptive training performance
 - b. H_1 : adaptive training performance > non-adaptive training performance
 - Independent variable: without and with adaptive training

In program Year 3 (Option, if funded) we intend to expand the use of UPCAT from individual performance assessment to team performance assessment. However, while there may be some simple cases of team training, we readily acknowledge the complexity of teamwork in general. Well-functioning teams must have a balanced workload with roles that mutually support each other. Unbalanced workload levels may be indicators of dysfunctionality between the team members. In crew resource management research, for example, crew members should have balanced levels of workload and they should both function within their assigned roles rather than having to reach into the other team-member's role. Ellis (Ellis K.E, 2014) measured workload in flight crews and then subjectively assessed each pilot's view of the other's workload. Discrepancies in actual workload and perception of the "other guy's" workload were found to be strong indicators of a team dysfunction. With that in mind, a machine learning feature extraction engine could be added to automatically watch out for such discrepancies.

REFERENCES

- Army. (2011a). *The US Army Learning Concept for 2015*. Retrieved from http://www.tradoc.army.mil/tpubs/pams/tp525-8-2.pdf
- Army. (2011b). *The US Army Learning Concept, 2012-2020*. Retrieved from http://www.tradoc.army.mil/tpubs/pams/tp525-8-2.pdf
- Ellis K.E. (2014). Fight Deck Crew Coordination Indices of Workload and Situatoin Awareness in Termninal Operations. (PhD), University of Iowa, Iowa City, Iowa.
- Engler, J., & Schnell, T. (2013). *Deterministically Nonlinear Dynamical Classification of Cognitive Workload*. Paper presented at the I/ITSEC 2013, Orlando, FL.
- Govindan, R., Narayanan, K., & Gopinathan, M. (1998). On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis. *Chaos*, 8(2), 495-502.
- Kozma, R. a. F., W. (2002). Classification of EEG Patterns using Nonlinear Dynamics and Identifying Chaotic Phase Transitions. *Neurocomputing*, 44-46, 1107-1112.
- OPL. (2014). Cognitive Assessment Tool Set (CATS) User Manual Retrieved from Iowa City, Iowa:
- Owis, M., Abou-Zied, A., Youssef, A., & Kadah, Y. (2002). Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification *IEEE Transactions on Biomedical Engineering*, 49(7), 733-736.
- Richter, M., & Schreiber, T. (1998). Phase Space Embedding of Electrocardiograms. *Physical Review E*, 58(5), 6392-6398. doi:10.1103/PhysRevE.58.6392
- Schnell, T. (2012). Workload Classification Algorithm Development and Validation for Human Factors Test Vehicle for DIAM. Retrieved from
- Schnell, T., & Engler, J. (2013). *Operator State Sensor Investigations and Operator State Classification* Retrieved from Hampton, VA:
- Schnell, T., & Engler, J. (2014). Entropic skill assessment of unmanned aerial systems (UAS) operators. *Journal of Unmanned Vehicle Systems*, 2(02), 53-68.
- Schnell, T., Melzer, J. E., & Robbins, S. J. (2009). The cognitive pilot helmet: enabling pilot-aware smart avionics. Paper presented at the Head- and Helmet-Mounted Displays XIV: Design and Applications, Orlando, FL, United states.
- Sottilare R, Sinatra A, & Boyce M. (2015). Domain Modeling for Adaptive Training and Education in Support of the US Army Learning Model—Research Outline Retrieved from
- VanLehn K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197-221.

ABOUT THE AUTHORS

Dr. Tom Schnell is a Professor in the Department of Industrial and Mechanical Engineering at the University of Iowa. He is also the Director of the Operator Performance Laboratory (OPL) where he has been the principal investigator on around 230 research projects. Tom has a undergraduate degree in Electrical Engineering and a MS and PhD in Industrial Engineering. He is a commercial pilot, test pilot, and flight instructor for fixed wing airplanes and rotorcraft.

<u>Mr. Cory J Reuter</u> is a full-time Research Assistant at the Operator Performance Lab (OPL). He is currently finishing his undergraduate degree in Mechanical Engineering at the University of Iowa. Before working at OPL, Cory spent

several years in the construction industry and has built everything from timber peg homes to designing and building furniture and custom built cabinets. At OPL, he assists in the design and manufacturing of the hardware needed for our human factors research and aircraft maintenance.

<u>Mr. Ezekiel D. Gunnink</u> is a full-time Research Assistant at the Operator Performance Lab (OPL). He is currently completing his computer & electrical engineering degree, with a specialization in software engineering. He is an expert in computer vision, graphics, and advancement in sensor technology. He is one of the lead developers and software architects of the CATS software

<u>Mr. Bradley M. Parker</u> is a full-time Engineering Coordinator at the Operator Performance Lab (OPL). He is a graduate of Idaho State University where he studied Graphic Arts Technology. He worked in the graphic arts industry for 15 years supporting computer graphics and high resolution imaging systems. Prior to joining OPL, he was an IT support specialist at the University Hospital and Clinics.

<u>Ms. Jaclyn Hoke</u> is a Principal Engineer in the Rockwell Collins Advanced Technology Center (ATC) with 10 years of experience in research, technology development, and technical project management. She is the research lead for Rockwell Collins initiatives in State Assessment and Training Effectiveness. Jaclyn has a B.S. in Applied Mathematics, an M.S. in Computer Engineering, and is currently pursuing her Ph.D.

Dr. Jason D. Moss is a Research Psychologist for the U.S. Army Research Laboratory (ARL)– Human Research and Engineering Directorate (HRED), Advanced Training and Simulation Division (ATSD) in Orlando, Fl. He has over 14 years of experimental research experience in the areas of military psychology, training, perception, human factors psychology, simulator sickness, and virtual environments. He has a Ph.D. in Human Factors Psychology from Clemson University.