Workload-Adaptive Training Scenarios for Synthetic Training Environments

Nathan D. Smith[1], Ezekiel D. Gunnink[1], Thomas Schnell[1], Christopher Reuter[1], Jason D. Moss[2]
Operator Performance Laboratory (OPL)[1], U.S. Army Research Laboratory (ARL)[2]

INTRODUCTION

The Army is working toward enhancements in soldier training systems using synthetic training environments (STEs) and mission rehearsal capabilities. This capability will augment live training on the range and in many cases will be self-guided, meaning that the trainee will not require a tutor or instructor to administer the training. The STE enables single user training and training of teams in small local groups and across operational networks involving large groups. Adapting the training scenarios to the capabilities and training needs of individual trainees is a proven way to enhance individual training effectiveness.

In distributed training evolutions, proper adaptation of the training scenarios takes on an even more important role, as such exercises involve many trainees at various stages of their training maturity and skill. Problems arise when less experienced, or lower-skilled, trainees are exposed to training scenarios that are too advanced, or complex, for their level of experience. This can easily happen if the STE does not consider the capabilities and limitations of the individual trainees. Such unprepared trainees are more likely to fail their training mission and thereby reduce the benefits of training, further exacerbating frustration in the trainee. In addition, the failure may jeopardize the success of other trainees who depended on a reasonably successful outcome of a mission task element in the scenario. Failure of a single trainee to accomplish his/her mission may result in a chain reaction of adverse events in the training evolution that may reduce the value of the training exercise or increase cost. Conversely, trainees exposed to missions that are not sufficiently challenging may experience boredom, or even apathy, resulting in a negative training benefit. Adaptive scenario administration is needed in STEs to avoid such breakdowns and to enhance individual training effectiveness.

There are many STEs and tools available such as VBS 3 (Virtual Battle Space 3). These tools often allow the creation and storing of scenarios that contain the starting conditions of a training module but the scenarios themselves are usually administered on a brute-force lesson plan. The structure of the simulation tools actually encourage such lesson based administration as it is very easy to create and save static scenarios. What is needed, however, is a mechanism to continually adapt the scenarios to match them to student abilities at their respective stages in the training program. Additionally, students need rich feedback on their performance and guidance on ways to modify behaviors to increase performance, if it is not at or above expectation.

In our work, we have created an adaptive training framework from three separate systems, (1). the Generalized Intelligent Framework for Tutoring (GIFT), (2) the VBS 3 simulation framework, and (3) the Cognitive Assessment Tool Set (CATS) workload quantification library. We developed a generic method to incorporate the GIFT performance grading scheme into VBS 3. This allows for on-the-fly configuration of adaptive VBS 3 training scenarios. Additionally, through CATS, this script can take into consideration the workload exhibited by the trainee and adapt the scenario to avoid over or underload conditions. This adaptive training framework is governed by student performance, workload, and task difficulty. Performance and workload were incorporated as aggregated scores. Workload is assessed using the CATS workload library that is attached to GIFT. Both performance and workload drive the selection of upcoming training scenarios by modulating task difficulty such that the trainee is challenged at an optimum level. A script in VBS 3 uses a decision tree on the basis of performance (below, at, above expectation) to manipulate the level of task difficulty to maximize training effectiveness.

In the current development cycle, we will perform a human factors study to assess the efficacy of adaptive training using two distinct adaptation schemes, one based on performance only, and one based on a combination of performance and workload. The results of the study will be used to determine if workload-adaptive training scenarios are more effective than training scenarios that only consider performance.

BACKGROUND

The value of adaptive training and its positive effect on training effectiveness has been well documented. The idea is of course not new (Lintern G. & Gopher D., 1978). The underlying principle is based on two hypotheses, (1) the learning of a complex task is best accomplished using less difficult versions of the task and increasing levels of difficulty until the whole task can be mastered, (2) learning of a task is better when transition from one level of difficulty to the next is guided by the performance of the student rather than brute-force administration of a rigid training regimen.

In one-on-one training settings, expert instructors use this principle almost instinctively to keep students motivated throughout the building of critical skills. For example, flight instructors may teach the difficult skill of auto-rotating a helicopter using increasing levels of difficulty by gradually increasing the complexity of the maneuver. In the example of autorotation, adaptation is not only representative of good training didactics, it is essential for survival of both the instructor and the student, as poor performance can lead to mechanical damage to an expensive helicopter, such as through over speeding the rotor system, or it could lead to a fatal crash such as allowing the rotor RPM to drop below an allowable minimum or initiating the landing flare too late. Control of this task requires manipulation of four inter-dependent controls (collective, lateral cyclic, longitudinal cyclic, and tail rotor pedals) as well as at least four inter-dependent performance parameters (airspeed, flight path, rotor RPM, aircraft attitude). To an uninitiated person, this maneuver is extremely scary and cognitive workload will be very high. It makes no sense to scare a student on each and every repetition of that maneuver as this will only increase the possibility that the student will never master it and be unable to use it as a needed emergency skill.

Expert instructors will ease their students into autorotations through adaptive training principles by giving the student only one control axis at a time (e.g. the collective) or through adjustment of the flight path (straight in path instead of curved). As the student gains confidence in his/her ability to master this skill at a given difficulty level, performance will improve and workload will go down. As is typical in the acquisition of many critically important skills, the decrease in workload is highly indicative of autonomous mastery. In the early stages, students may be able to master the skill at an acceptable technical level but only with the highest levels of cognitive workload expenditure. This is usually sufficient for passing a checkride or to graduate with a certificate but it is hardly a proper level of training for critical skills in warfighters. Instructors and instructional systems owe it to the warfighter to train them to a higher standard. High levels of cognitive demand causes significant draw on limited attentional resources (see Figure 1) which adversely affects the performance of perception, memory, decision making, and response execution. Trainees who master the skill to a point of automaticity will expend less cognitive workload and thus retain more attention resource capacity. This will afford them to devote those resources to mission critical task elements, which is essential in the projection of military power and for the self-protection of the warfighter.

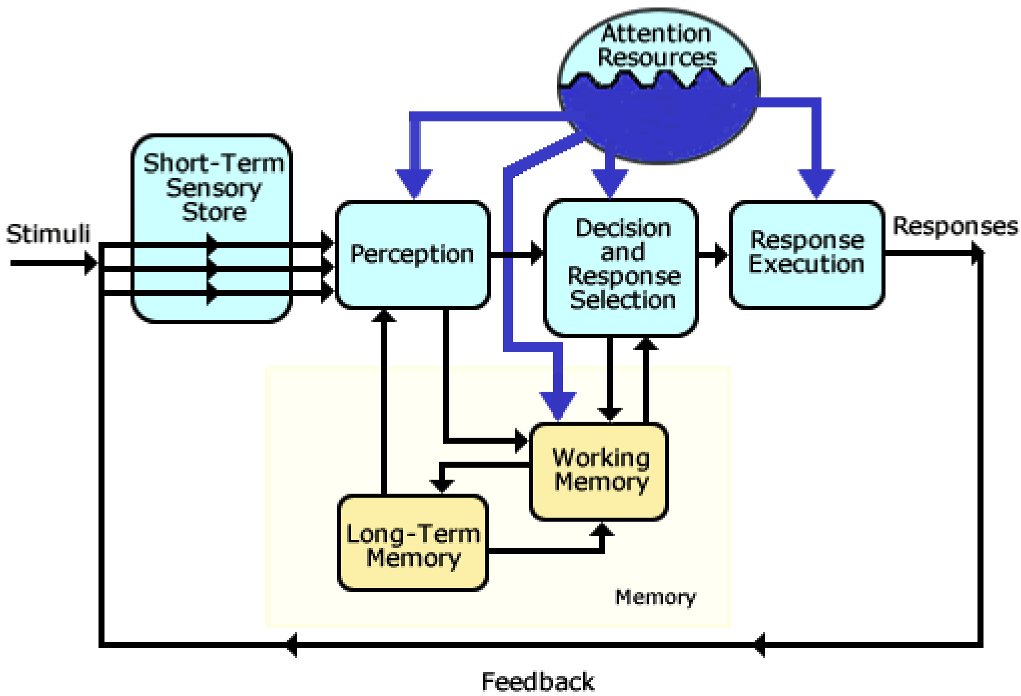

Figure 1. Wickens Information Processing Model (Wickens C. D., 1992, 2008)

In our research, we are working on building the adaptive expertise that good instructors apply almost instinctively into automated Synthetic Training Environments (STEs).

Even as far back as the Seventies, adaptive training was conceived of as a closed-loop controller system (Lintern G. & Gopher D., 1978). Such a system depends on measurement of task performance. Unfortunately, automatically generating performance measures is not always easy and in many training tasks has eluded us to this day. Additionally, the most optimal way to make the training scenario adaptive is not always easily evident. Much research has been devoted to the question of how to make training tasks adaptive. Part-whole training is an adaptation scheme where essential subtasks are learned as building blocks to enable mastery of the whole task. Part

Task (PT) training was found to lead to significantly faster convergence of a tactical skill in a video game when compared to Full Task (FT) training (Mané, Adams, & Donchin, 1989). An interesting observation of their work is that the part tasks were not fully representative as fractions of the whole task but when learned in sequence lead to better performance than if the full task is learned at once. An additional observation is that the PT training took longer than FT training. However, the skill transfer rates from the PT were 100% and the overall performance of demonstrating the full task was much better. Thus, while PT may not yield net time savings, the fact that better performance is achieved may mean that less remedial training will be needed.

Mane and Wickens (Mane A. & Wickens C., 1986) studied the effects of task difficulty and workload on training. They noted that training systems should adapt to maintain high levels of workload as otherwise, trainees will learn short-term resource preserving strategies that are counterproductive toward mastering of the long-term skill. Rigid (i.e. non-adaptive) training methods allow such maladaptive resource preservation strategies to take hold. In our work, we use real-time measures of cognitive workload to quickly close that short-term loophole for the trainee by adjusting training difficulty to maintain high levels of workload while at the same time preventing overload or defeat of the student through scenarios that are too difficult (e.g. autorotations).

Gerjets et al. (Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014) describe the relationship between cognitive load theory (CLT) and training outcomes through optimal loading of working memory load (WML). The main challenge is a continuous classification of cognitive workload to allow adaptation of the training scenario to modulate WML. They describe methods such as subjective probes or secondary tasks measures. Both methods of workload estimation are disruptive and hinder training effectiveness. They used EEG as a means to estimate workload with some success. The use of EEG signals for classification of workload is well represented in the literature, two additional examples of which are presented here. Wilson and Russell (Wilson & Russell, 2003) attempted to classify workload using a combination of sensors, including six channels of brain electrical activity, eye, heart, and respiration measures. Those authors were able to achieve classification accuracies around 82%. However, their tasks consisted of only two variants of the same test. Additionally, the high number of sensors used to collect the data, is sub-optimal for many scenarios including in flight measurements. Matthews *et al* (Matthews et al., 2008) used a wireless EEG sensor helmet to classify workload in real-time. Those authors achieve classification accuracies on an average of 80.5%. In well over a decade of workload estimation research at the University of Iowa Operator Performance Laboratory (OPL), we have come to the conclusion that the technical readiness level and diagnostic capabilities of EEG based workload probes is very low and unsuitable for a real-world training environment outside a highly controlled laboratory.

A much simpler sensor montage is possible through a three-lead electrocardiogram (ECG). At OPL, we have used discrete deterministic nonlinear models of the full ECG waveform to obtain reliable and highly diagnostic real-time measure of cognitive workload. It is important to note that our method of ECG based workload estimation is NOT a heart rate based method or a time-series based analysis. Rather, we continually transform the entire ECG signal into an embedded phase space and classify workload on the basis of the dynamic representation of the heart though an ergodicity map of the electrical heart signal. We start with the realization that the heart is a chaotic system that is under control of the nervous system. 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 multi-dimensional embedded phase space (Richter & Schreiber, 1998). In our method, the ECG time series data is transformed into phase space using the CATS software tool (OPL, 2014). This step established the Ergodicity Transition Matrices (ETMs) (Engler & Schnell, 2013) that represented the dynamics of the ECG signal in phase space for the different workload conditions. To generate a real-time workload estimation, we can either use the ETMs directly through lookup of model ETMs using nearest neighbor classifiers or through models of statistical transitions within the ETM called the Transition Probability Variance (TPV). TPV calculates the variance of the probabilities of transition from one cell to another different cell of the course-grained ETM. The TPV therefore captures the variability in the dynamics of the ECG signal as the trainee undergoes different levels of cognitive loading. TPV varies inversely to the degree of workload with higher TPV numbers seen under low workload conditions and low TPV numbers seen under high workload conditions. The benefit of the direct ETM based discrete classifier is its very high accuracy level (near 100%). The downside of this method is that model ETMs need to be established for each participant and each desired level of workload. The TPV method is less accurate (around 85-90% classification accuracy) but it does not require a model. The TPV method provides a continuous measure of workload no more than three heartbeats after the ECG system has been turned on. The TPV system has excellent cross-person and cross-task validity and is easily deployed in complex real-world environments (Schnell T et al., 2017; Schnell T., Hoke J., & Romeas T., 2017; Schnell T., Reichlen C., & C., 2017).

Another trainee specific dimension that may be applied in the context of adaptive training systems is that of trainee affect and engagement. As with performance based adaptation, expert instructors generate a motivating and interesting training experience and they have the ability to detect affectual cues from the trainee such as frustration, fear, boredom, or anger. Effective instructors can interpret affectual cues as levers that affect learning. The affective domain of training provides a framework for instruction that includes student awareness, response, value perception, organization, and integration (FAA, 2008). A trainee has to be aware of the material being taught. It is the responsibility of the instructor or instructional system to raise the awareness level in the trainee through immersive and interesting content. The student responds through active participation, decides on the value of the training, organizes the training into his/her belief system, and finally, internalizes it. Motivation and enthusiasm are important enabling components of the affect domain.

Ocumpaugh et al. (Ocumpaugh et al., 2017) provide a thorough review of the role of emotions in training. A quantitative understanding of affect dynamics allows not only for an understanding of a learner's current affective state but also enables prediction of future affective states. Ocumpaugh et al. leverage data of the trainee's affect dynamics toward making better adaptive training transitions. A proposed approach for incorporating affective state assessment into the GIFT training system draws from the observed model of affect dynamics (**Error! Reference source not found.**) presented by D'Mello and Graesser (D'Mello & Graesser, 2012).

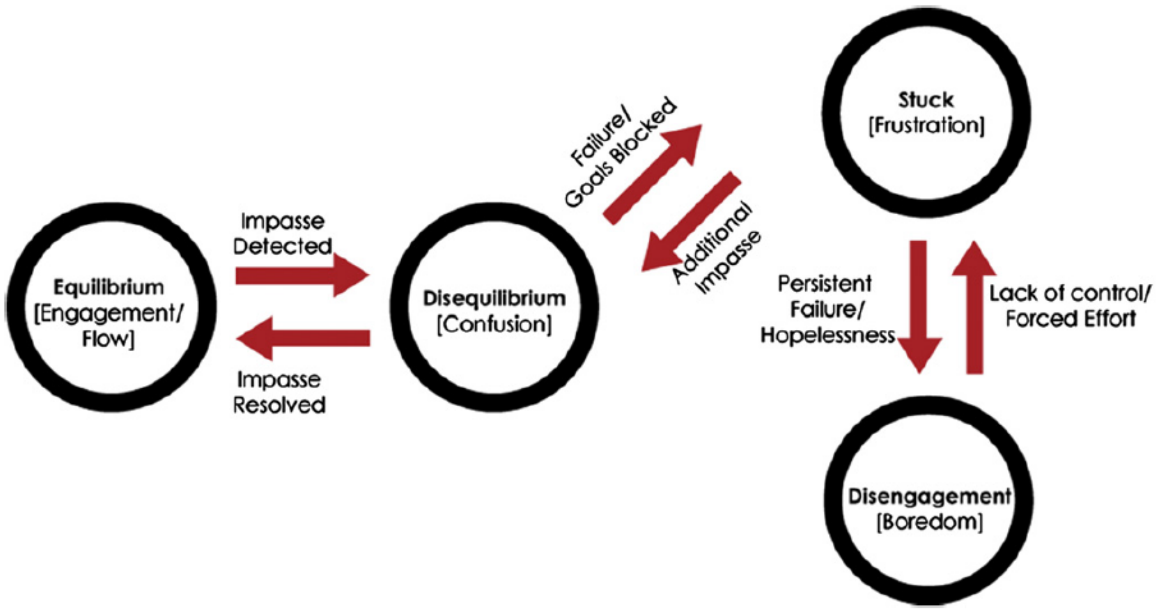

Figure 2. D'Mello & Graesser Model of Affect Dynamics

We identify the following states and definitions from the referenced work

- s_1 – Engagement/Flow: A state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete. This is of course a desired state in a training system.
- s_2 – Confusion: A state experienced while encountering the "cognitive disequilibrium" that occurs when confronted with obstacles to goals, interruptions of organized action sequences, impasses, contradictions, anomalous events, dissonance, incongruities, unexpected feedback, uncertainty, deviations from norms, and novelty. In the context of a training system, this is a "productive" form of confusion as the resolution of the impasse provides a sense of accomplishment.
- s_3 – Frustration: A state experienced while encountering the "hopeless confusion" that occurs when an impasse cannot be resolved, the learner gets stuck, there is no available plan, and important goals are blocked
- s_4 – Boredom: A state experienced when a learner disengages from the learning process

The model depicts six primary state transitions, but the design focuses on four (4) transitions that have pedagogical implications to an adaptive training system

- $s_1 \rightarrow s_2$: Caused when an impasse is detected and the learner engages is effortful problem solving
- $s_2 \rightarrow s_1$: Caused when an impasse is resolved. Additional positive affective states, such as delight, may occur as a result of achieving goals or receiving positive feedback
- $s_2 \rightarrow s_3$: Caused when an impasse cannot be resolved, the learner is stuck, or important goals are blocked

- $s_3 \rightarrow s_4$: Caused when persistent frustration prompts the learner to disengage from the learning process

While it is not documented in this model, a direct $s_1 \rightarrow s_4$ transition may also occur if the learner is under-tasked, or when concentration or attention is broken. Proposed training adaptations are presented in two specific contexts: affective state alone and affective state coupled with physiological workload. If the trainee is in a prolonged state of equilibrium, scenario complexity should be increased to trigger the $s_1 \rightarrow s_2$ transition and cause the learner to engage in effortful problem solving. Sustained equilibrium should be managed to prevent an $s_1 \rightarrow s_4$ transition. The $s_2 \rightarrow s_1$ transition back into equilibrium does not require immediate intervention, as it indicates problem solving has been applied to successfully achieve a goal or resolve an impasse. However, the transition should trigger the system to monitor for a prolonged state of equilibrium. The $s_2 \rightarrow s_3$ transition into frustration does not require an immediate intervention; however, it should trigger the system to monitor for a prolonged state of frustration. Sustained frustration should be managed by reducing scenario complexity to prevent the $s_3 \rightarrow s_4$ transition. If the $s_3 \rightarrow s_4$ transition occurs, the scenario complexity should be reduced to present the learner with a more simplified problem, but the complexity of the problem must also increase the learner's interest in re-engaging with the training session. If an $s_1 \rightarrow s_4$ transition occurs, the scenario complexity should be increased to present the learner with a more complex problem that also increases the learner's interest in re-engaging with the training session.

Physiological workload assessment techniques can reinforce, or modify, the adaptations based solely on affective state. For brevity, the differences to the list above are included here. Stable or decreasing workload reinforces the adaptation that increases scenario complexity and triggers the $s_1 \rightarrow s_2$ transition during a prolonged state of equilibrium. A decreasing workload trend should immediately trigger the adaptation to prevent an $s_1 \rightarrow s_4$ transition. Workload provides an added dimension to the $s_2 \rightarrow s_3$ transition into frustration. The transition to frustration, paired with a stable or moderate increase in workload, does not require an immediate intervention, but should trigger the system to monitor for prolonged frustration. The $s_2 \rightarrow s_3$ transition accompanied with a dramatic increase in workload should result in a reduction in scenario complexity to prevent a rapid $s_3 \rightarrow s_4$ transition. Ideally, the coupling of affective and physiological state should allow for early detection and prevent the $s_3 \rightarrow s_4$ transition from occurring.

STUDY: PHYSIOLOGICAL BASED ADAPTIVE TRAINING

This paper describes a study that we are preparing to conduct over the next few months. Unfortunately, we cannot present any results art this time. However, we feel that there is value in conveying our test plan to the scientific community.

The present study is intended to assess the value of adaptive training systems that use measures of subject workload. We intend to test the hypothesis that adaptation using performance and workload (P+WL) will lead to better training outcomes than adaptations using performance only (P). Stated as a testable hypothesis EH_1:

- H_0: performance only based adaptive training score = performance with workload adaptive training score

- H_1: performance only based adaptive training score < performance with workload adaptive training score

In this experiment, both groups (A and B) will receive task training using their respective P+WL or P only adaption scheme. The effectiveness of that training will then be assessed in a graded capstone checkride. Throughout the training, we will periodically administer subjective workload probes to allow us an independent validation of the accuracy of the OPL workload algorithm.

Each subject will wear a NeXus 4 channel wireless ECG system that collects raw data used by the UPCAT system to assess workload of the participants. Performance metrics from within the virtual environment along with workload are used to adapt the scenario. Figure 3 shows the system architecture used to collect and assess subjects' performance and workload. All audio and video from the HMI, as well as audio and video of the subject is recorded and synchronized. Figure 4 shows the system architecture used to collect and synchronize audio and video data.

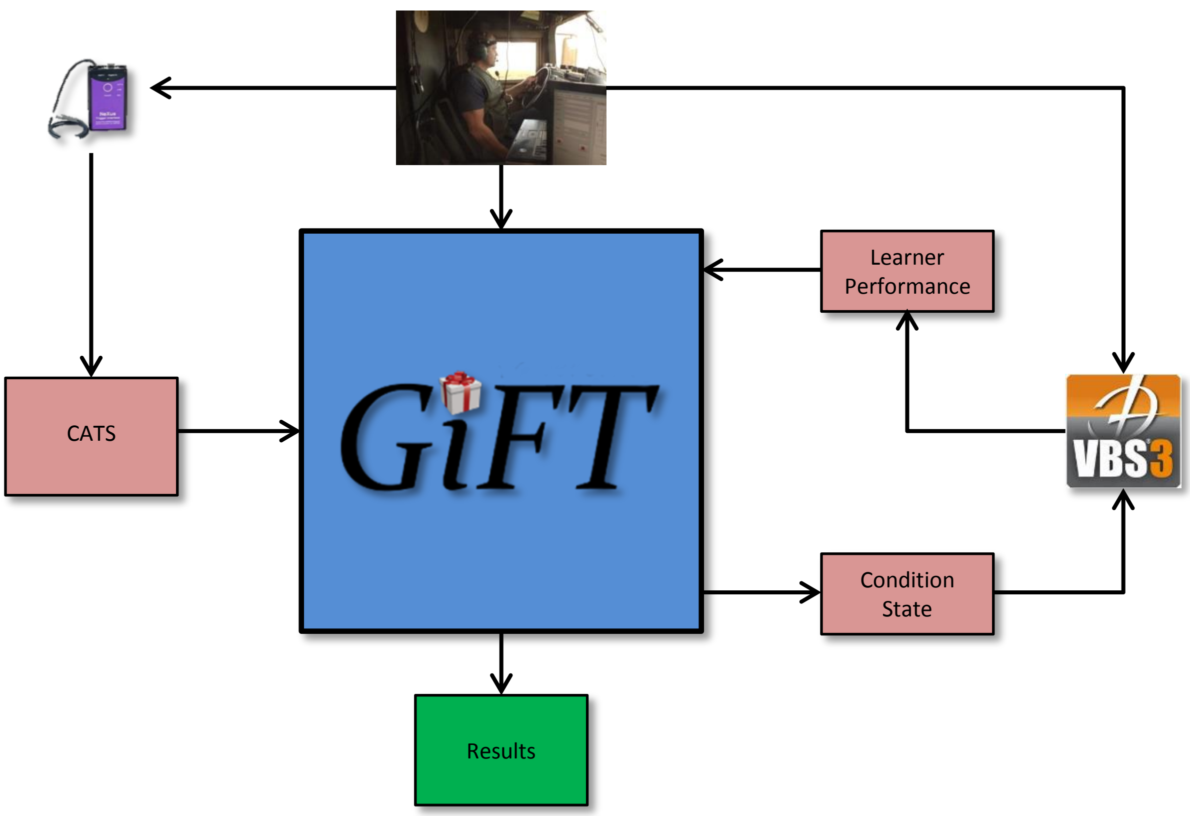

Figure 3. UPCAT System Architecture

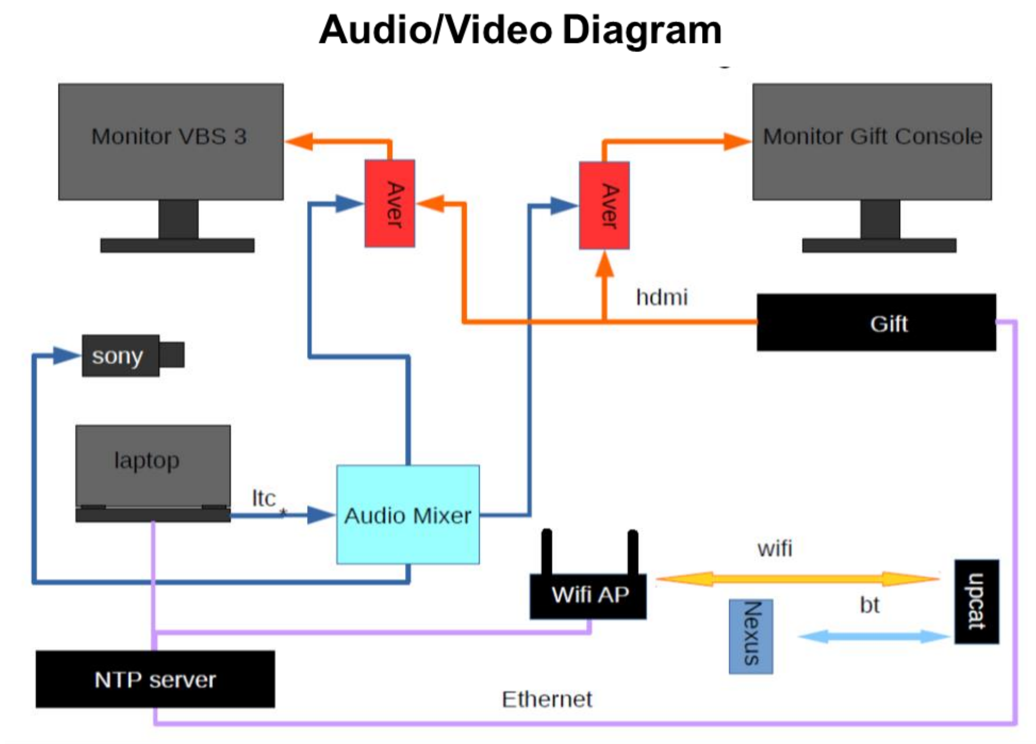

Figure 4. System Architecture for Video and Workload Data Capture

Each participant will complete a GIFT based training course in accordance with the group assigned adaption scheme (P+WL), (P). Within this course, each participant will complete a number of tasks. With the exception of the non-adaptive introduction (warmup), each task has three levels of difficulty (i.e. Easy, Medium and Hard). Participants will return approximately 14 days after their initial course to complete the capstone checkride. This general GIFT course flow can be seen in

Figure 5.

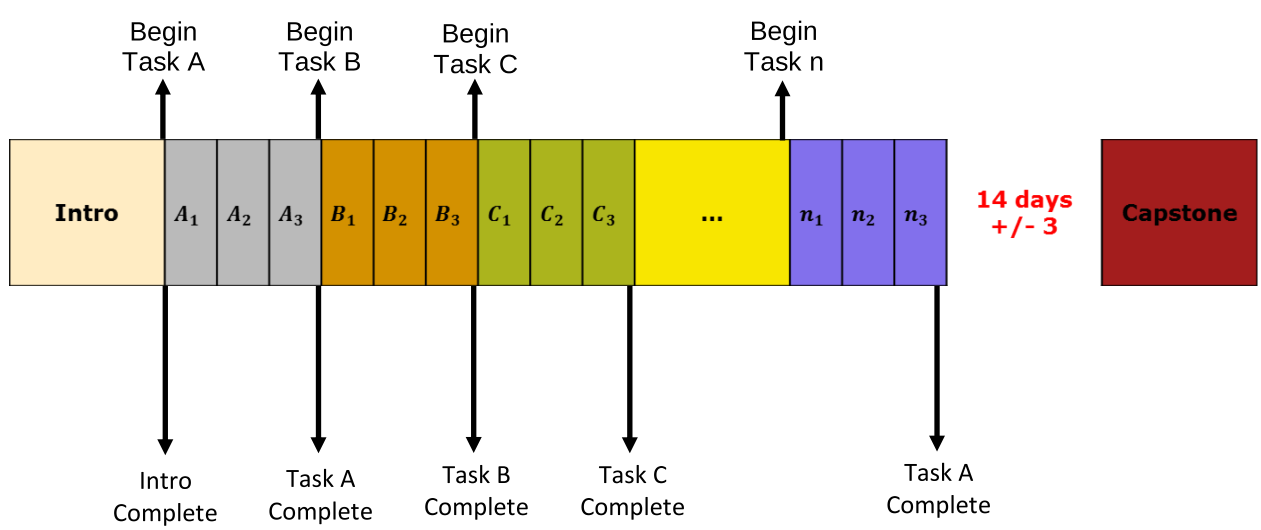

Figure 5. Training Course Flow

Each participant will first attempt each task at the medium difficulty level. Participant performance and workload are assessed throughout the training task and summarized for the adaptation decision at the end of each attempt. For each separate level of difficulty, participant outcomes are classified into one of three groups based on their performance score (green bubbles) as being below expectation, at expectation, or above expectation. The transition to the ensuing task level follows the decision tree shown in Figure 6 and Figure 7 for (P) and (P_+WL) groups, respectively. These adaptation decision trees were adapted from (Mark et al., 2018).-Figure 4

.

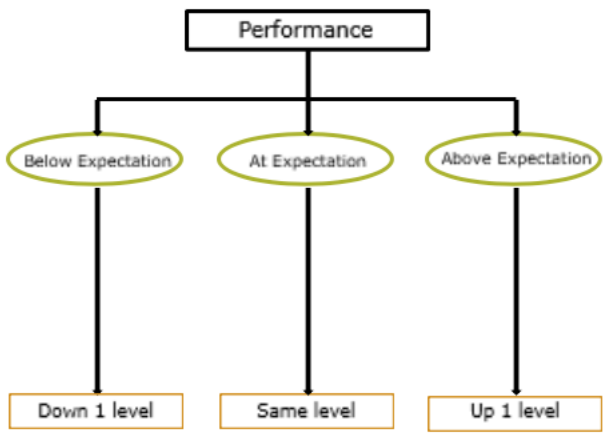

Figure 6. Adaptation Flow for Performance Only Adaptation (P)

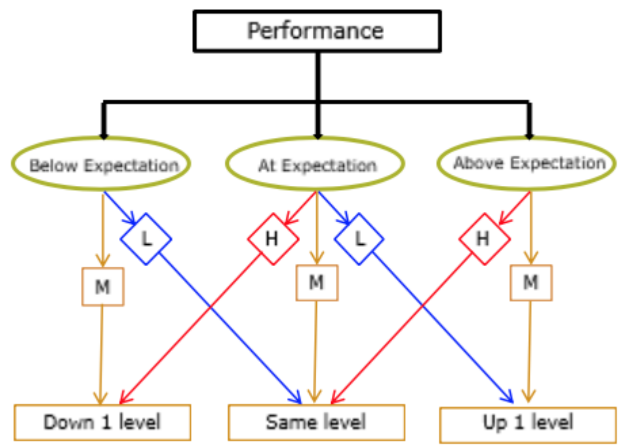

Figure 7. Adaptation Flow for Performance with Workload Adaptation (P+WL)

The capstone checkride consists of one ever increasingly difficult task that encompasses all task elements from all previous part tasks. Participants continue through this increasingly difficult capstone checkride until they fall below performance thresholds. The point in the checkride where they fail is the dependent measure of training effectiveness with a later failure being better than an early one. We chose this method of testing to avoid ceiling or floor effects where many or all participants pass or fail a checkride of a selected level of difficulty.

Throughout the experimental GIFT driving course we evaluate four conditions. They include the GIFT Corridor Boundary, OPL Workload, Maintain Speed, and Collision Avoidance. GIFT evaluates both Corridor Boundary and Workload Classifier conditions while VBS 3 evaluates Maintain Speed and Collision Avoidance conditions. VBS 3 maintains a state variable for each Corridor Boundary, Workload Classifier and Maintain Speed conditions. Each GIFT condition has three state transition strategies: one for each of the below, at or above expectation evaluations (increasing, decreasing and maintaining for workload), in accordance with the flow graphs shown in Figure 6 and Figure 7.

We added six new Environmental Control Enums; one for each condition at each evaluation which are used in GIFT state transition strategies. Using the *sendCommand()* function from GIFT's VBS 3 Plugin Interface we are able to send any valid VBS 3 script command. For example, assume that the subject has trouble with tracking the vehicle in the middle of the driving lane. Therefore, the Corridor Boundary condition will evaluate to a value of *below expectation*. GIFT executes its corridor boundary *from anything to below expectation* state transition strategy which sends the VBS 3 command *["BELOW"] call setCorridorState*, and the Corridor boundary state variable maintained by VBS 3 is updated to *BELOW*. The same happens for all evaluations and accompanying state transition strategies for both the Corridor Boundary and Workload Classifier conditions.

Currently we have hard-coded the commands through the use of the Environmental Control Enum. This is restrictive as VBS 3 allows for thousands of commands. We experimented with the *sendCommand()* function, and were able to send multiple commands separated by a semi-colon with a single call to *sendCommand()*. We believe the ability to create custom commands within the state transition strategies instead of the restrictive hard coded example we are using to be an appropriate addition to GIFT. We could add a single CUSTOM_COMMAND enum to the list of GIFT Environmental Control Enums. The command(s) could then be written into, and read from, the course .dkf file when GIFT calls the state transition strategy implementing that command.

As mentioned before, both Maintain Speed and Collision Avoidance Conditions are handled by VBS 3. Both are called inside of an event handler attached to the subject object which fires every time the subject object moves. The event handler includes a timer that only calls the evaluation functions for both conditions for every evaluation interval (currently every 1 second while the vehicle is moving). For both the Corridor Boundary and Maintain Speed conditions, VBS 3 maintains a timer for each of the below, at or above expectation evaluations. At every evaluation interval, VBS 3 checks the current state of the two conditions and adds the elapsed time from the previous evaluation to its corresponding timer. The final evaluation for each of the Corridor Boundary and Maintain Speed conditions is assigned based on what percentage of the total time was spent in each state based on Table 1 (note that actual logic accounts for ranges and not set values).

Below/Total	At/Total	Above/Total	Evaluation
0%	0%	100%	ABOVE
0%	25%	75%	ABOVE
0%	50%	50%	ABOVE
0%	75%	25%	AT
0%	100%	0%	AT
25%	0%	75%	ABOVE
25%	25%	50%	AT
25%	50%	25%	AT
25%	75%	0%	AT
50%	0%	50%	AT
50%	25%	25%	BELOW
50%	50%	0%	BELOW
75%	0%	25%	BELOW
75%	25%	0%	BELOW
100%	0%	0%	BELOW

Table 1. Evaluation Assignment for Corridor Boundary and Maintain Speed Conditions

Maintain Speed condition is graded through the use of a target speed and a speed window. If the subject is outside the speed window, they are evaluated to below expectation. If the subject is inside the center one-third of the speed window, then they are evaluated to above expectation. If the subject is between inner one-third and outside of the speed window, then they are evaluated to at expectation. Let the target speed be 35 km/h, and the speed window be 6km/h. If the subject's speed is more than 41 km/h or less than 29 km/h, then they are outside the speed window and are evaluated to below expectation. If the subject's speed is between 37 and 41 km/h or between 29 and 33 km/h, then they are evaluated to above expectation. If the subject's speed is between 33 and 37 km/h, then they are evaluated to at expectation.

Collision Avoidance is graded through the use of upper and lower bounds. If, at the end of an attempt, the subject has had fewer collisions than the lower bound they are evaluated to above expectation. If the subject has had more collisions than the upper bound, then they are evaluated to below expectation. Anything in between receives an evaluation of at expectation.

The performance evaluation used for adaptation is an aggregate of these three condition evaluations. Each task weights the evaluation of the three conditions differently, and some are not even used at all for some tasks. Let Task one be driving in reduced visibility, where the subject is evaluated on maintaining speed and corridor boundary while driving through a sandstorm. It is important to maintain their speed, but it is more important to stay on the road. So a fair weighting of the singular evaluations to determine the aggregate performance evaluation could be set to Equation 1. The aggregate performance and workload evaluations are then used to decide on the scenario adaptation based on the adaptation trees from Figure 6 and Figure 7.

$$aggPer = 0.40 \times speedEvaluation + 0.60 \times corridorEvaluation$$

Equation 1. Aggregate Performance Grade for Task 1

The aforementioned evaluation and adaptation logic is controlled by various scripts and event handlers. VBS 3 *init.sqf* script (called at the start of the scenario) compiles multiple scripts that set-up the global variables; the VBS 3 waypoints, create the files used for data collection; task, time, grading and GIFT message related functions; event handlers; and scripts that set-up the evaluation of conditions. The scenario adaptations needed for each level of difficulty for each task are also contained within their own scripts.

The current GIFT Corridor Boundary condition did not allow an evaluation of above expectation, and we were concerned about fairness in the evaluations of the two groups (A & B). For example: the ability of subjects from group A to reach an evaluation of above expectation and an adaptation of up 1 level compared to subjects from group B's ability to reach the same adaptation through an evaluation of at expectation with a decreasing workload as shown in the adaptation trees in Figure 6 and Figure 7. We saw a potential for a biased evaluation and made changes to allow GIFT's Corridor Boundary condition to evaluate to above expectation.

It works in much the same way as the Maintain Speed condition. If the subject is outside the corridor, they are evaluated to below expectation just as before. The change we made affected the way the subject is graded while inside the corridor. If the subject is inside the center one-half of the corridor, then they are evaluated to above expectation. If the subject is between inner one-half and outside the corridor, then they are evaluated to at expectation. Let the corridor be 10 meters wide. If the subject is more than 5 feet away from the center of the corridor, then they are outside the corridor and are evaluated to below expectation. If the subject is less than 2.5 meters from the center of the corridor, then they are evaluated to above expectation. If the subject is less than 5 meters but more than 2.5 meters away from the center of the corridor, then they are evaluated to at expectation.

For purposes of our study, we write all data related to decision making with respect to the evaluation of the different conditions, aggregate scoring and adaptations throughout the course to .csv files. Each data point is timestamped with the system time (the exact time and date according to the computer the subject is using).

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

We invested considerable effort in the establishment of an architecture that tightly integrates the capabilities of the GIFT framework with VBS 3 as a representative of an Army Synthetic Training Environment (STE). This architecture provides a robust control interaction capability between the two systems. Additionally, this architecture includes tight integration of a continuous workload assessment system (CATS) using a deterministic nonlinear workload classifier that analyses the ECG waveform in embedded phase space. This apparatus is capable of assessing learner state in real-time, in this case using a driving task, and applying performance and workload assessments to automatically configure scenario transitions for adaptive training.

Additionally, we invested significant effort integrating an existing GIFT learner affect classifier library (Ocumpaugh et al., 2017) into this framework. This classifier uses a Kinect sensor to track features on the learner's face to classify states of emotion. Even though we spent a tremendous amount of effort in an attempt to integrate this library, we were, to date, not yet able to gain a reliable classification from it. Therefore, in the upcoming validation study using this apparatus, we decided not to use learner affect as a state variable to invoke scenario transitions. If we manage to get the affect state library to work, we will collect data from its affect state classifier for separate and off-line analysis.

REFERENCES

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

Engler, J., & Schnell, T. (2013). *Deterministically Nonlinear Dynamical Classification of Cognitive Workload.* Paper presented at the I/ITSEC 2013, Orlando, FL.

FAA. (2008). *Pilot's Handbook of Aeronautical Knowledge* (FAA-H-8083-25A). Retrieved from

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience, 8*(385). doi:10.3389/fnins.2014.00385

Lintern G., & Gopher D. (1978). Adaptive training of perceptual-motor skills: issues, results, and future directions. *International Journal of Man-Machine Studies, 10*(5), 521-551.

Mane A., & Wickens C. (1986). *The Effects of task Difficulty and Workload on Training.* Paper presented at the 30th Annual Meeting of the Human Factors and Ergonomics Society.

Mané, A. M., Adams, J. A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, 71*(1), 179-196. doi:https://doi.org/10.1016/0001-6918(89)90008-5

Mark, J., Thomas, N., Kraft, A., Casebeer, W. D., Ziegler, M., & Ayaz, H. (2018). *Neurofeedback for Personalized Adaptive Training*, Cham.

Matthews, R., Turner, P., McDonald, N., Ermolaev, K., McManus, T., Shelby, R., & Steindorf, M. (2008, Aug. 20-24, 208). *Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes.* Paper presented at the The 30th Annual Intl Conf of the IEEE on Engineering in Medicine and Biology, Vancouver, BC.

Ocumpaugh, J., Andres, J. M., Baker, R., DeFalco, J., Paquette, L., Rowe, J., . . . Sottilare, R. (2017). *Affect Dynamics in Military Trainees Using vMedic: From Engaged Concentration to Boredom to Confusion*, Cham.

OPL. (2014). *Cognitive Assessment Tool Set (CATS) User Manual* Retrieved from Iowa City, Iowa:

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, Reichlen C., Geiselman E, Knox J., Williams H., & Ercoline W. (2017, May 8 - May 11, 2017). *A Comparison of Helmet-Mounted Display Symbologies During Live Flight Operational Tasks.* Paper presented at the 19th International Symposium on Aviation Psychology, Dayton, Ohio, USA, May 8 - May 11, 2017.

Schnell T., Hoke J., & Romeas T. (2017, Mar 7-9 2017). *Achieving the Third Offset: Maximizing Human-Machine Symbiosis.* Paper presented at the 2017 NDIA Human Systems Conference,, Waterford at Springfield.

Schnell T., Reichlen C., & C., R. (2017). *Spatial Disorientation Threat Characterization for F-35 Representative Helmet-Mounted Display Use in the Flight Environment* Retrieved from Dayton, OH:

Wickens C. D. (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins.

Wickens C. D. (2008). Multiple resources and mental workload. *The Journal of the Human Factors and Ergonomics Society, 50*(3), 449-455.

Wilson, G., & Russell, C. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors, 45*(4), 635-644.

ABOUT THE AUTHORS

Mr. Nathan D. Smith *Nathan joined the Operator Performance Laboratory's (OPL's) team in September of 2017. He currently works as a Research Assistant and is also a student at the University of Iowa College of Engineering. Nathan will graduate in the fall of 2018 BSE in Electrical and Computer Engineering, with a focus in Software Design and a minors in both Business and Computer Science.*

Mr. Ezekiel D. Gunnink *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.*

Mr. Chris Reuter *joined the Operator Performance Laboratory in October 2012. He has a bachelor's degree in mathematics from the University of Iowa and is currently pursuing a PhD in Industrial Engineering. Prior to joining OPL, Chris worked in the financial industry and brings to the team a wide range of management and operational experience. Chris's primary focus at OPL is working with human factors studies and data analysis.*

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 an 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.*

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.*