SubTAG Session Abstracts
DoD HFE TAG Meeting 69

SubTAG: Modeling & Simulation

Chair: Ranjeev Mittu Co-Chair: John Rice

Tuesday, 5 May 2015
Session: I

1240 - 1330 **Past, Present, and Future of Modelling and Simulation: A Panel Discussion**

Dr. Lauren Reinerman-Jones, IST, UCF; Dr. Shawn Burke, IST, UCF

The department of defense has a long history of investing in the use of simulation to train warfighters to operate in complex, mission critical environments. While there is some evidence to suggest its effectiveness (Tannenbaum & Yukl, 1992), it is only a tool. And, as with many tools, the quality with which it is used can greatly vary. In order to be effective, simulation must be designed and utilized appropriately. Research has begun to suggest some of the basic requirements for human-in-the-loop simulation. For example, Salas and Burke (2002) argued that at a minimum the following characteristics must be in place: (1) instructional features are embedded within the simulation, (2) carefully crafted scenarios, (3) opportunities for assessing and diagnosing individual and team performance, (4) a guided learning experience, (5) simulation fidelity is matched to training requirements, and (6) a reciprocal partnership between subject matter experts and learning/training specialists exists. However, ensuring that such characteristics are embedded within human-in-the-loop simulations often presents a challenge in practice. Panel members will speak to their experience in the design and use of human-in-the-loop simulation, challenges in implementing the science, and how the fields they represent have worked to overcome such challenges. Panel members represent a range of experience, expertise, focus, and domain. We have designed our panel such that members represent a variety of time periods: past, present, and future. This look back to look forward not only highlights where the field is going (and corresponding capabilities and roadblocks), but where it has been. This capitalizes on lessons learned, as well as identifying those challenges that remain persistent.

1350 - 1405 **Simulation-Based Surgical Education and Training: Current and Future Considerations**

Dr. T. Robert Turner, Jr., American College of Surgeons

This presentation will include an overview of various simulation-based surgical education and training initiatives of the American College of Surgeons (ACS), as well as high-level discussion around the growing international network of ACS-accredited simulation centers. Within this framework, I will also present some of the key human factors objectives, challenges, and potential obstacles that lie ahead. Finally, I will discuss trends in current human factors research as they pertain to advancing the role of simulation in selecting, educating, and training future medical students, residents, and practicing surgeons. The opinions expressed and statements made in this presentation reflect the presenter's personal observations and do not imply endorsement by or official policy of the American College of Surgeons.

1405 - 1430 **Discrete Event Simulation for airborne Battle Management Command and Control crew reduction of the Joint Surveillance Target Attack Radar System: A Monte Carlo study.**

Mr. Stephen McGee, 711th Human Performance Wing: Human Systems Integration

Alion Science and Technology collaborated with The 711th Human Performance Wing (711HPW) to develop general Human Systems Integration (HSI) considerations for reducing the crew size of the Joint Surveillance Target Attack Radar system (JSTARS) airborne Battle Management Command and Control (BMC2). They analyzed potential HSI issues, such as increased cognitive workload, task queue and message response time associated with a reduced BMC2 personnel. Discrete event simulations of BMC2 on JSTARS were developed using Improved Performance Research Integration Tool (IMPRINT), which allows a currently used baseline crew of 16 to be evaluated and compared to a reduced crew of 10. Based on the IMPRINT models it appears that a reduced crew of ten would have no significant differences in workload with a lower message response time and a lower task queue with the assumed technology advances in place. This paper utilizes the familiar Monte Carlo simulation method for message response time based on operator task queue and cognitive workload for a comparison with the analytical models. Based on this information, we might make different choices when reducing or reconfiguring crew size.

Tuesday, 5 May 2015
Session: II

1615 - 1635 **VIRTUAL ENVIRONMENT FOR LAUNCH OFFICERS**

LT Lee Sciarini, Naval Postgraduate School; LCDR Jeffrey Korzatkowski, Naval Postgraduate School

Naval aircraft Launch and Recovery Officers or 'shooters' are responsible for the safe launch and recovery of aircraft from the flight deck of aircraft carriers. To become fully qualified as a shooter, trainees must complete the Aircraft Launch and Recovery Equipment Officer course. Upon completion of this training course, the officer is assigned to an aircraft carrier to complete qualification training while on-the-job. This training can only be conducted during flight operations. Consequently, training cannot be accomplished while performing non-flight operations while underway or whenever in port. Ultimately, commands that are in port or in the shipyard must coordinate training opportunities with commands that are underway and conducting flight operations in order to achieve pre-deployment training requirements. This effort will discuss the assessment, design, and development of a low-cost virtual system capable of producing an environment in which core shooter skills can be practiced. Specifically, this presentation will discuss a task analysis which identified core shooter skills, an assessment of the skills to determine those which could be effectively performed and/or practiced in a low-cost virtual environment, a review of current commercially available technology, and finally, the development of a virtual environment to perform the shooter tasks.

1615 -1635 **Robotics Operator Manager ACT-R Model and Validation**

Dr. Daniel Cassenti, U.S. Army Research Laboratory; Kristin Schaefer, U.S. Army Research Laboratory

With the U.S. military's proposed increase in activated robots in the field, and the subsequent mental workload demands associated with control of each robot, there is an opportunity to form a new role for the management and coordination of multiple robotic assets. The role of a Robotics Operator Manager

(ROM) was studied using a Model Test Model (MTM) approach during a scenario in which robots were used to detect engineered explosives during a convoy mission. This work is the final phase in the MTM approach. We 1) discuss the ROM role; 2) provide a brief discussion of the initial model; 3) explore the Test phase, which showed promise to accomplish military missions with robots in light of the negative effects of time and bandwidth limits on performance and trust in network; and 4) conclude with a modeling and simulation approach for the final model using Adaptive Control of Thought-Rational (ACT-R) modeling. An examination of the results of the final model indicated not only a close approximation of the mental processes required to fill the ROM role, but also can be used to establish a personnel selection process.

SubTAG: Standardization

Chair: Al Poston

Tuesday, 5 May 2015
Session: I

1230 - 1245 **Military Standard 1474E: Design Criteria for Noise Limits vs. Operational Effectiveness**

Mr. Bruce Amrein, Army Research Laboratory

In the spring of 2015, the U.S. Department of Defense published a significant revision to Military Standard 1474 (MIL-STD-1474): Design Criteria-- Noise Limits. This standard was last revised in 1997 to the "D" version (MIL-STD-1474D). Through the efforts of a multi-service working group every aspect of MIL-STD-1474 has been revised to improve readability; reduce conflicting guidance; and consolidate requirements common to steady-state and impulsive noise produced by weapons systems, and ground-, air- and water-borne platforms. Noise requirements in military environments differ significantly from typical industrial or occupational situations, and mission success requires offensive equipment and weapons to be more lethal and survivable than those used by the adversary. Higher muzzle velocities, heavier projectiles, and more powerful engines result in high levels of both steady-state and impulsive noise and an increased risk of hearing loss for the users. Weapons firing can expose the user to more energy in a single event than typically experienced in a working lifetime of occupational exposure. For the first time in MIL-STD-1474E, a computer-based, electro-acoustic model of the human auditory system is used to evaluate hazard from impulsive noise events typical of weapon firing. Producing material suitable for military operations requires unique design criteria often exceeding civilian national or international standards. MIL-STD-1474E provides the design tools and measurement techniques necessary to satisfy these unique and often contradicting requirements. This presentation describes the salient requirements of MIL-STD-147E necessary to produce and deploy military systems maximizing Warfighter effectiveness, while minimizing hearing damage caused by their use.

SubTAG: Human Factors Engineering/Human Systems Integration

Chair: Pamelyn Maynard

Tuesday, 5 May 2015
Session: I

1235 - 1305 **Acquisition Program Human Systems Integration (HSI) Assessment: An Air Force HSI Emerging Approach**

Ms. Julia Naga, Booz Allen Hamilton consultant of Air Force Human Systems Integration Directorate (711 HPW/HP)

OVERVIEW: Since establishment with the Air Force Research Laboratory's Human Performance Wing, the Human Systems Integration Directorate's (711 HPW/HP) has provided program office support across the Life Cycle Management Center acquisition portfolio. The HSI Baseline Assessment is a new approach to supplying programs with a human centered perspective on the system being procured. It allows the HSI Directorate and the associated AFLCMC program to summarize HSI domain areas, their derived HSI risk, and provides possible support options to the program. The progression of the various iterations, the approach development, future directions and applications will be discussed.

AUTHORS:

1) Julie Naga, HSI Analyst, Booz Allen Hamilton in support of Air Force Human Systems Integration Directorate (711 HPW/HP)

2) Tim Bush / Technical Advisor/ Air Force Human Systems Integration Directorate (711 HPW/HP)

3) Bill Kosnik / Chief, Material Branch /Air Force Human Systems Integration Directorate (711 HPW/HP)

1310 - 1350 **AF Human Systems Integration Capabilities and Requirements Assessment Tool**

Mr. Arpan Patel, Booz Allen Hamilton; Dr. William Kosnik, 711 HPW/HPI; Ms. Andrea Cooks, Booz Allen Hamilton; Ms. Sheryl Cosing, Booz Allen Hamilton; Ms. Jessica Shihady, 711 HPW/HPI

AF Human Systems Integration Capabilities and Requirements Assessment Tool Authors: Arpan Patel, William Kosnik, Andrea Cooks, Sheryl Cosing, Jess Shihady, John Plaga Establishing human-centered requirements early in the system design, development and acquisition process is key to delivering effective and useable systems to the warfighter. To achieve this goal, it would be beneficial for the HSI practitioner to have a standardized, data-driven tool to assist in generating, tracking, and documenting human-centered requirements. To provide a means for identifying human related risks and concerns, the Survivability/Vulnerability Information Analysis Center created the Human Systems Integration Capabilities and Requirements Assessment Tool (HSI – CRAT). The HSI-CRAT uses domain-based questions to analyze human related risks in the Capabilities-Based Assessment (CBA), Analysis of Alternatives (AoA), and Development Planning (DP) process. The questions were developed by leveraging DoD and AF requirements guides. HSI-CRAT is based on a risk management approach that is familiar to system engineering and program management disciplines. The practitioner responds to a set of questions with simple yes/no answers that query the status of human centered risks at any point in the requirements planning process. The practitioner is then prompted to rate the potential human performance risk using the standard DoD risk methodology. The tool also allows the practitioner to document the HSI status by entering data that supports the risk rating. Based on user responses, the HSI-CRAT provides a report documenting the human performance risks associated with the analysis being performed. A preliminary set of questions has been developed for HSI-related CBA, AoA, and DP activities and is being vetted by domain and HSI subject matter experts. A prototype tool is currently undergoing usability testing. The final validated tool is scheduled for release in June 2015.

1355 - 1420 **The Multi-Service Human Systems Integration Framework (HSIF): Activities for DoD pre-Acquisition.**

Mr. Frank Lacson, Pacific Science and Engineering Group; Dr. Matthew Risser, Pacific Science and Engineering Group; Dr. William Kosnik, USAF HSI Division, 711th Human Performance Wing

The HSI Framework (HSIF) is an interactive process diagram that visualizes and aligns the technical activities, collaborations, and products of all HSI Domains across the DoD Acquisition Life Cycle. Activities originate from relevant HSI-related guidance, standards, and best practices across DoD Services and non-DoD organizations. A timeline display references activities to Systems Engineering Technical Reviews and Acquisition milestones. The HSIF provides technical value to HSI Practitioners, Program Managers, MAJCOM/SYSCOM Technical Authorities, System Engineers, and Prime Contractors. This presentation outlines user-driven updates to the HSIF software application and introduces HSI activities for pre-Acquisition.

Currently, HSI activities, guidance, and planning have been scoped well after a decision is made to develop a system. This scoping limits the effectiveness of HSI planning and requirements development. Thus, unrealized system performance may not be realized until late in the Acquisition cycle, resulting in costly re-work and additional engineering change proposals. The advent of alternative and rapid acquisition strategies increases the need to conduct relevant and cost-effective HSI efforts early in the system life cycle.

Analysis of policy, standards, guidance, and best practices related to Capabilities Based Assessment (CBA), Development Planning (DP), and the Analysis of Alternative (AoA) Study Plan Guidance were used to generate Pre-Acquisition HSIF content. Best practices were also derived from Modeling & Simulation (M&S) across the HSI Domains, as well as lessons learned from Science & Technology (S&T) down-select efforts for systems in early-Acquisition.

For HSI Practitioners and Systems Engineers, access of early, explicit, and timely pre-Acquisition guidance improves the likelihood of documenting human-related constraints and opportunities throughout System Development. In turn, Program Managers are given the information needed to accurately assess human-related program risks, leading to relevant and scoped HSI plans. At the Service level, conducting HSI activities during pre-acquisition can also provide technical insight to the user-related components of non-materiel solutions: doctrine, organization, training, leadership and education, personnel and facilities.

Tuesday, 5 May 2015
Session: II

1505 - 1525 **Execution of the US Army Anthropometric Survey (2010-2012) and Preparation of the Database**

Mr. Joseph Parham, Natick Soldier RD&E Center; Dr. Claire Gordon, ORISE Fellow - Natick Soldier RD&E Center; Mr. Steve Paquette, Natick Soldier RD&E Center; Ms. Cynthia Blackwell, Natick Soldier RD&E Center; Mr. Joseph Venezia, Natick Soldier RD&E Center

The US Army maintains a comprehensive anthropometric database mainly used to inform the design and sizing of military clothing, protective equipment, workstations, vehicles, and digital human models. Before 2011, the last anthropometric survey of US Army personnel (ANSUR) was conducted in 1988 (Gordon et al, 1989), and that benchmark survey positively influenced countless programs of record since its publication and release. However, amid concerns that the 1988 database no longer represented the modern Army, in 2006 the US Army Natick Soldier Research, Design & Engineering Center (NSRDEC) undertook a preliminary study of Soldiers to determine whether a new anthropometric survey was needed (Paquette et al, 2009). Finding significant differences between the 1988 and 2006 data, a comprehensive follow-up was initiated, dubbed ANSUR II. Around this same time, the USMC Project Manager, Infantry Combat Equipment decided to conduct a full-scale Marine Corps anthropometric

survey (MC-ANSUR) and approached NSRDEC to conduct the study. These joint service anthropometric surveys were therefore conducted in conjunction, with many of the same personnel, and with the same data collection protocols in place. This presentation reports the methods and techniques used to insure that the ANSUR II database accurately represents the US Army, and provide valuable data for use by scientists, designers, and engineers. Methods and Results: From October 2010 to April 2012, NSRDEC oversaw and lead data collection for ANSUR II at 12 Army bases spread geographically across the country. A total of 6,866 male Soldiers and 3,721 female Soldiers were measured during ANSUR II using standardized protocols to minimize observer error and maximize comparability to the 1988 ANSUR survey database (Hotzman et al., 2011). For example, each measurement was scrutinized by statistical software during data collection, comparing it both univariately and multivariately to previously collected data. Flagged measurements could then be immediately re-measured while the Soldier was still participating in the survey. Demographic information was also carefully collected on each Soldier, to be compared to Defense Manpower Data Center (DMDC) data on the total Army. Following data collection, outliers within the anthropometric data were identified through multivariate regression modeling. They were then investigated, and a determination of their reliability was made. Because some anthropometric data were missing from participating Soldiers, these data were tested for bias. It was shown that including Soldiers with missing data in the database did not significantly change key metrics, and therefore removing Soldiers with incomplete measurement sets from the database would not introduce bias. The demographic data were likewise reviewed and decisions were made to keep, change, or remove inconsistent demographic entries. Because ANSUR II minorities were oversampled during data collection, the demographic distribution of the resultant databases did not match DMDC census data of the Army. Therefore, representative working databases of 4,082 male Soldiers and 1,986 female Soldiers were selected through stratified random sampling. The summary statistics of the ANSUR II working databases have been recently published (Gordon et al, 2014); summary statistics for the MC-ANSUR effort were published earlier (Gordon et al, 2013). Analyses of both service's anthropometric databases and their use in modeling and design efforts are ongoing. Discussion: Conducting a large-scale military anthropometric survey is a complicated and challenging endeavor under ideal conditions. During a time of war, with units deploying and redeploying, gaining access to large numbers of Marines and Soldiers is particularly difficult. High level support is critical in conducting such surveys, as is ensuring that the sample is representative of the forces and the data collected are accurate. Methods utilized and lessons learned in the execution of MC-ANSUR and ANSUR II should inform future large-scale anthropometric data collection efforts.

Keywords: Military Anthropometry; Anthropometric Survey; Data Cleaning

References:

Gordon CC, Blackwell CL, Bradtmiller B, Parham JL, Barrientos P, Paquette SP, Corner BD, Carson JM, Venezia JC, Rockwell BM, Mucher M, and Kristensen S. (2014) 2012 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics. NATICK/TR-15/007. Natick, MA: U.S. Army Natick Soldier RD&E Center.

Gordon CC, Blackwell CC, Bradtmiller B, Parham JL, Hotzman J, Paquette SP, Corner BD, and Hodge BM. (2013) 2010 Anthropometric Survey of U.S. Marine Corps Personnel: Methods and Summary Statistics. NATICK/TR-13/018. Natick, MA: U.S. Army Natick Soldier RD&E Center.

Gordon CC, Churchill T, Clauser CE, Bradtmiller B, McConville JT, Tebbetts I and Walker RA(1989). 1988 Anthropometric Survey of U.S. Army Personnel: Methods and Summary

Statistics. NATICK/TR-89/044. Natick, MA: U.S. Army Natick RD&E Center.

Hotzman J, Gordon CC, Bradtmiller B, Corner BD, Mucher M, Kristensen S, Paquette S, and Blackwell CL (2011). Measurer's Handbook: US Army and Marine Corps Anthropometric Surveys, 2010-2011. NATICK/TR-11/017. Natick, MA: U.S. Army Natick Soldier RD&E Center.

Paquette S, Gordon CC, Bradtmiller B (2009). Anthropometric Survey (ANSUR) II Pilot Study: Methods and Summary Statistics. NATICK/TR-09/014. Natick, MA: U.S. Army Natick Soldier RD&E Center.

1530 -1620 **Implementing Human System Integration Across the Department of Homeland Security Systems Engineering Lifecycle: Guidance on the Selection of Tools**

Ms. Angelia Sebok, Alion Science and Technology

Human System Integration (HSI) practitioners and managers encounter numerous challenges during the acquisition process. As HSI specialists well know, HSI impacts can be greatest and most effective when they are made early in the lifecycle (e.g., identification of human performance gaps in the needed capability, HSI inputs to the mission needs, and analysis of alternatives). The majority of tools that currently exist, however, to support HSI evaluations are more relevant for much later stages (e.g., preliminary and detailed design) with a focus on user interface design and test and evaluation. This paper describes an effort to support the Department of Homeland Security (DHS) HSI personnel throughout the entire acquisition lifecycle by identifying relevant tools that can be used at each stage in the process. In this effort, we have compiled a set of nearly 200 HSI tools and described them in terms of the capabilities they offer, the HSI domains they support, and the phases and stages of the acquisition lifecycle for which they can be used. We have also assigned rankings to the tools based on their breadth of applicability (the number of the HSI domains and lifecycle stages they support), the depth of analysis they allow (comparisons only or qualitative assessments, quantitative assessments, or “what if” scenario analysis), the availability of validation evidence, and the cost associated with obtaining the tool.

These tools are included within an Excel workbook that includes pages (or tabs) that identify the specific lifecycle stages. Each tab lists the tools that are relevant for that particular stage. The tools are sorted according to their rankings, but users can filter the data in a variety of ways. The workbook also includes “deep dive” reviews of a select sample of the tools. These tools are also included within the spreadsheets, and reviewed according to the same criteria as other tools. This “catalog” of tools will support DHS HSI practitioners and managers in selecting the appropriate techniques for conducting evaluations or gathering data throughout the acquisition lifecycle.

In the past decade, there have been numerous efforts to compile human factors tools. This current project differs from many of those previous efforts by the interaction allowed in the Excel workbook. Users of the DHS HSI spreadsheets can identify specific tools that are relevant for a particular phase in the acquisition lifecycle. They can easily identify which tools can be used for each of the HSI domains. They can get recommendations for particular tools by observing the rankings in the spreadsheet. For a limited set of tools, those examined in the deep dive analyses, the HSI catalog users can review more detailed information on how the end user would work with the tool to perform an analysis, gather data, and interpret the data.

The eventual goal for this tool is to provide a repository of tools that HSI practitioners and managers can access, and to provide guidance to support those personnel in the selection of the most appropriate tools for their evaluations.

1605 - 1630 **Methods & Metrics Metadata Construct (M3C): Standardized Expression of Human Factors Data for Enhanced Analytics**

Mr. Steve Dorton, Sonalysts

We are continually evaluating human-machine teams and interfaces with a variety of Human Factors Engineering (HFE) methods and metrics. While it is relatively simple to identify the ideal method(s) of investigation for a given study, practical limitations with scheduling, budget, and other logistical issues often force HFE/Human Systems Integration (HSI) practitioners to settle for less. By using smaller sample sizes or less detailed methods we lose fidelity, reliability, and validity in our analyses, as well as our ability to generalize findings to larger populations and other applications- reducing the total value of our efforts. To mitigate these issues, a methods and metrics metadata construct is being developed to enable longitudinal collection and synthesis of data from HFE/HSI studies and analyses. The notional construct includes metadata components about test settings (test systems, test location, and participants), the methods employed, and the metrics used in the study. By expressing research via metadata, and developing a standardized approach to analysis, DoD HFE practitioners will have increased abilities to triangulate metrics, make inference across demographics, and extend generalizability of findings across different contexts by collaborating as a community and leveraging prior work. Notional results of the metadata construct will be presented, along with results of an example study triangulating a metric across two domains (usability and performance), and some challenges for practical DoD-wide implementation.

1635 - 1655 **Development of a Novel Test Methodology for Assessing the Impact of Clothing and Individual Equipment (CIE) on Weapon Compatibility**

Mr. Jay McNamara, Natick Soldier Research, Development, and Engineering Center

An important compatibility/ interoperability issue for any item of CIE is the Soldier's ability to accurately engage targets with their primary weapons system without compromise of performance. Developers of CIE, including personal protective equipment (PPE), must consider the impact the equipment has on the User or Warfighter's ability to perform their mission and to protect them. Due to changing threats during recent conflicts the US Military has a desire to increase the area of coverage and level of protection achieved from our PPE, but this has been at the detriment of the Warfighter's mobility and interferes with marksmanship.

Previous research has identified interference and incompatibility between the weapon and the CIE. Specific issues have been identified with body armor, the bulk of the soft armor at the shoulder tends to block the shooters pocket, frequently keeping the Warfighter from being able to obtain a steady weapon position and making it difficult for Warfighters to easily and properly sight their weapon (Garlie, Mitchell, & Guerra, 2010; Mitchell, 2013). The majority of this data has been collected via observational data and subjective Soldier ratings. Objective measures have been collected using live shooting ranges, but there are increased costs, safety concerns and complexity associated with these methods.

The goal of this project was to develop a methodology that could be utilized in a laboratory or field setting, without the need for a live fire range. Additional quantitative and subjective measures would be

collected to determine compatibility/ interoperability of CIE marksmanship performance. A methodology capable of obtaining operationally relevant, measurable data will assist in the identification of restrictions encountered when performing tasks associated with target acquisition and engagement.

Over the past 2 years, the NSRDEC Human Factors team, using a Noptel MilTrainer weapon simulator system, has began working to develop and validate a test methodology for quantifying the effects of equipment on weapon compatibility (i.e., ability to quickly and accurately engage targets) using that weapon simulator. One of the benefits of the Noptel simulator system is that it can be operated both indoors and outdoors, and is portable and robust enough to be used in field test scenarios.

Data has been collected from two different evaluations utilizing similar methodology, and additional data collection efforts are planned. One evaluation assessed the impact of select Body Armor Protection Levels (BAPLs) on marksmanship performance. The second evaluation assessed the impact of the M40 Chemical Biological protective mask on marksmanship performance. The novel test methodology proposed, utilizes two different series of experimental scenarios to evaluate the impact of CIE on marksmanship. The first scenario consists of a single target located 5m directly in front of the test participant. The target is scaled such that when placed 5m from the participant, it represents a full size E-type silhouette at 75m. For the one-target scenario the test participant engages this target from the standing unsupported, kneeling unsupported, and prone unsupported firing positions. In each of these positions the test participant fires five series of five shots each (25 shots total), lowering the weapon to the low ready position in between series. For the second scenario the test participant engaged a total of five targets, firing one shot at each target. The five targets consist of a single target directly in front of the participant, two targets located 50° to the right and left of the center target, and two targets at high angles relative to the participant. All of the targets are placed 5m away from the participant, and are scaled such that they represent a full size E-type silhouette at 75m. For the five-target scenario the test participant engages the targets from the standing unsupported and kneeling unsupported firing positions.

Marksmanship performance for the one-target scenario was assessed with two variables generated from the Noptel system data – Euclidian-distance (or E-distance) and Bull-Distance (B-distance). The E-distance (or variable error, VE) measures the inconsistency (or consistency) in the outcome. This measurement is the averaged distance from the center of the shot series to each shot. If VE is small, the outcomes are consistent and located together closely. Another way to envision VE is “shot cluster size” or how tightly grouped the five shots in a shot series are.

The B-distance, or total variability (TV), measures how successfully/accurately the TPs performed. This measurement is the average distance of the five-shot series to the bull’s-eye. If the shot outcomes are close to the center of the target, TV is small. In other words, this variable is “shot accuracy.”

For the five-target scenario, an additional three timing variables were also assessed – Total Time Between Shots, Aiming Time, and Transition Time. The Total Time Between Shots measures the time from one shot to the next. Aiming Time measures the time that the participant is aiming at the target prior to firing. Transition Time measures the time required to transition from one target to the next, and is calculated by subtracting the Aiming Time from the Total Time Between Shots.

All five of the dependant measures were statistically compared across the test conditions to determine whether any performance differences in these variables existed across test conditions as well as firing positions (standing/kneeling/prone). Statistically significant differences in both E-distance and B-distance have been found across firing positions.

In general, the participants performed better (closer to the target) in the prone firing position than in the standing or the kneeling firing positions, and better in the kneeling firing position than in the standing firing position. However, the participants transitioned between targets significantly faster in the standing firing position than in the kneeling firing position. For the BAPL comparison evaluation, statistically significant differences in E-distance and B-distance have not been found across BAPL conditions. However, significant differences have been found between BAPL conditions for transition time. Similarly, for the M40 CB mask evaluation, no significant differences were found in terms of E-distance and B-distance, but participants were significantly slower in transitioning between targets when wearing the CB mask than when they were not wearing a mask.

Due to NSRDEC's mission, the primary focus of this effort has been on quantifying the effects of CIE on the marksmanship task. However, this simulator system, test methodology, and lessons learned may be used to optimize training requirements and define the relationship of training requirements to mission-level capabilities. The team plans to continue assessing the viability of the test methodology, including adding variants that could decrease independent differences and variability of individual shooters (e.g., looking at expert versus novice shooter, introduction of optics) and development of new timing related human performance metrics.

References:

Garlie, T., Mitchell, K.B., and Guerra, R. (2010) Anthropometric and Human Factors Evaluation of a Body Armor Vest with Females: A Report of Methodology. Proceedings of the Personal Protective Equipment Conference 2010. A paper presented at the Technical Support Working Group (TSWG) PPE Conference 2010. 29 November – 03 December 2010. Fort Lauderdale, Florida.

Mitchell, K.B. (2013) Standard Methodology for Assessment of Range of Motion While Wearing Body Armor. Natick Soldier Research, Development, and Engineering Center (NSRDEC); Natick, MA. NATICK/TR-13/033.

1700 - 1720 **Large scale usability evaluation of biometric technologies.**

Dr. Yevgeniy Sirotin, Scitor Corporation; Mr. Jacob Hasselgren, Scitor Corporation; Ms. Elise Rivera, Scitor Corporation

Biometric technologies are gaining wide acceptance as a means of securely establishing individual identity. Common biometric modalities include recognition of fingerprints, faces, and irises. However, introducing biometric technologies into existing human systems can have significant impact on system usability metrics: efficiency, effectiveness, and user satisfaction. We are involved in conducting a large-scale study aimed at evaluating the usability impact of different biometric technologies prior to integration into a human system. To do this, we are performing usability evaluations with hundreds of human subjects in realistic use-case scenarios at a dedicated test facility while recording subject demographics, video, audio, fine-scale timing information, as well as assessing satisfaction. Leveraging this comprehensive dataset, we show that different biometric modalities and collection methods are associated with unique behavioral, performance, and demographic challenges. Further, we show how system performance can be improved by adjusting components linked with identified usability problems. This approach to testing and evaluation may have broader application to other situations employing biometrics to meet security requirements.

SubTAG: Tech Soc/Industry

Chair: Barbara Palmer Co-Chair: Steve Merriman
Tuesday, 5 May 2015

1 **SAE International, G-45 Human Systems Integration Committee Update.**

Mr. Steve Merriman, Boeing

G-45 committee accomplishments for the past 8 years will be summarized. Current activities, including completion of revision to the HADAD-M DID and participation in the HSI Standard DOD working group will be presented. Future plans will be described.

2 **OSD Sponsored Human Systems Integration (HSI) Working Group Update**

Owen Seely, NAVSURWEPCEN Dahlgren, VA

The HSI Standard working group will be described, including the background and phase I accomplishments. Current status and near-term plans will be presented and discussed, including the development and distribution of a survey of HSI practitioners and the decision as to whether the standard should be industry or DOD based.

3 **Acquisition Program Human Systems Integration: An Air Force HSI Emerging Approach**

Ms. Julie Naga, Booz Allen Hamilton

The Air Force's Human Systems Integration Program has deployed Human Systems Integration Analysts in several Air Force Major Commands, who ensure that human elements are included in early acquisition documents. This brief will address the Air Force Research Lab/Human Systems Integration Directorate's evolution of acquisition program HSI assessment during the strategic rollout of HSI support to select Air Force program offices.

4 **Human Performance Data: The Delicate Balance Performance Optimization and Privacy**

Dr. Elizabeth Biddle, The Boeing Company

The use of various data sources / types (e.g., physiology, activity, preference and environment) to evaluate and improve and individual and/or team's performance is rapidly becoming a part of our daily routine. Individuals interested in getting into better physical shape are gravitating towards gadgets such as fitbittm to monitor their physical activity and/or nutritional intake and provide recommendations for reaching physical fitness goals. Applications such as AT&T's Navigator monitor an individual's transportation behavior and recommend optimal routes. Society is embracing social network sites such as Facebook, Twitter and Instragram to let "friends" know what they're thinking, doing and seeing throughout the day.

This is a vast amount of data that is personal, and potentially confidential or private, to an individual that is being uploaded to the "cloud" or sent over a variety of wireless networks. There are periodic news stories regarding concerns of privacy being violated by one of the many social networking sites or personalized monitoring technologies, but these fade away while the amount of personal data transmitted continues to increase. Is it that there has not yet been an incident of a serious enough nature to cause society to pause and take account of the data they so readily disclose? Is the desire to communicate with society or make self-improvement so much greater than an interest in privacy of certain data that the importance or notion of individual privacy is fading away? If society or technology

developers do not think of how to balance privacy with improved capability and performance, will privacy become a luxury of a past society?

The defense industry has also embraced big data analytics as a means of better understanding and supporting human performance needs. The term Quantified Warrior (Blackhurst, Gresham & Stone, 2012) describes a vision in which analysis of multi-dimensional human performance data can be leveraged to provide optimized Warfighter support. The vision of Personal Assistants for Learning (PALs), coined by the Advanced Distributed Learning Co-Laboratory, are automated assistants that use a variety of data sources to diagnose student needs and recommend tailored instructional activities (http://www.adlnet.gov/next-generation-learner/anthropology-of-pals/). These two examples represent the future of human performance optimization, and requiring detailed and perhaps, personal data – and understanding of that data. The types of data include neurophysiological, physiological, personality, performance history and demographic data. The capabilities provide fine-tuned diagnosis of human performance needs and automated, personalized human performance support that include on-demand performance support tools and tailored training interventions.

These applications have demonstrated effectiveness in improving human performance and perhaps even saving lives. However, what happens to all of the data used to drive these on-demand, adaptive performance support technologies? Is the data logged and stored in a data warehouse? Is it made available for other uses, such as evaluating promotions or other career advancement decisions? For example, suppose a performance support tool determines that a pilot consistently does not obtain the recommended duration of sleep required for flight. Should this information be presented to his Instructor with recommendation for grounding? Further, would this data then go into his performance record? Another scenario to consider, an individual enters the military with a desire for a specific occupation and perhaps even has related experience in which they performed well. However, the individual's personality and physiological data are not consistent with the average for a specific occupation. Should the individual not be allowed to continue with this occupation, even if it is the individual's passion and performance results are average – or above average?

The purpose of this presentation is to initiate a dialogue within the Human Factors and Ergonomics community on the topic of privacy protection strategies for personal data used in various human performance optimization technologies. Discussion points will include types of data that are considered personal information and methods to protect this data. The presentation will present examples of data protection strategies such as data encryption and data access and storage processes.

SubTAG: Unmanned systems

Chair: Ajoy Muralidhar

Wednesday, 6 May 2015
Session: I

1230 - 1300 **Multimodal Communication for Dismounted Human-Robot Teams**

Dr. Daniel Barber, IST, UCF; Dr. Julian Abich, IST, UCF; Dr Lauren Reinerman-Jones, IST, UCF

The application of Human-Robot Interaction (HRI), within the military context, is evolving beyond the contemporary explicit control to one reflecting the characteristics and interactions of a responsive human teammate. Currently, teleoperation is the primary method of robot control by a Soldier, with communication dominated by visual displays. One major goal of many robotics programs is to advance current robotic systems "from tools to teammates" (Phillips, Ososky, Grove, & Jentsch, 2011) by

developing communication interfaces that reflect the ways in which humans communicate with each other. This paradigm shift will drive future Human Robot Team (HRT) communication technology, fostering a blend of modes to enhance team effectiveness while avoiding additional cognitive strain on the Soldier.

Multimodal communication (MMC) is the ability to transmit and receive information through more than one communication method or "modality" with the emphasis on using all five senses (Dumas, Lalanne, & Oviatt, 2009). The use of MMC within dismounted HRTs has attracted much attention within the HRI community as a way of meeting operational demands (Barber, Lackey, Reinerman-Jones, & Hudson, 2013; Barber, Reinerman-Jones, & Matthews, 2014). The appeal of MMC in HRTs is largely due to the natural, intuitive, and flexible modes in which Soldiers can communicate with robots. MMC input and output devices must not only be sensitive enough to capture the information transmitted from the Soldier, but must also be robust enough to withstand the continuum of military operations.

The U.S. Army Research Laboratory (ARL) established the Robotics Collaborative Technology Alliance (RCTA) to progress the state-of-the-art in robotics to enable mixed-initiative HRTs. In 2014, a Capstone Assessment was held to demonstrate the successful integration of each program thrust. The portion of the field assessment referred to here is on the usability of a prototype multimodal interface (MMI), which was developed to assess elements of MMC including the use of gestures and speech to communicate with a robot, and options for displaying information to a user (e.g., path of the robot, color, symbols). The results of this assessment will be used to inform designers on the usability of the device, to evaluate the impact of the implemented design principles, and to select optimal modes of interaction that support effective MMC within HRTs.

1300 - 1330 **Augmenting Robot Behaviors Using Physiological Measures of Workload State**

Ms Grace Teo, IST, UCF; Dr Lauren Reinerman-Jones, IST, UCF; Dr Gerald Matthews, IST, UCF

The evolution of robots from being teleoperated tools to responsive teammates requires a paradigm shift in the way humans and robots interface and communicate. Beyond explicit commands, robot teammates need to be able to interpret accurately naturalistic and intuitive forms of human communication such as speech and gestures. Robot teammates also need to be able to pick up implicit, but important, cues that reflect the human teammate's physiological and psychological state (e.g. workload). Such enhancement of the robot's sensing capabilities would enable it to formulate responses and behaviors that are better adapted to the dynamic operational environment. However, assessment of workload in this context must be minimally invasive and continuous, and so most subjective measures are precluded, because they require overt responses and disrupt the task at hand. In contrast, physiological measures of workload are less intrusive and allow continuous assessment. Therefore, physiological measures are one means for communicating state experience from a human to a robot, but a model that effectively classifies workload state based on physiological measures has yet to be developed. Some models exist, but they mostly feature one or two physiological measures, i.e., EEG and HRV. Given the large individual differences in physiological workload responses observed throughout literature, it is more judicious to include multiple physiological measures and develop a model of workload that is individualized. Such individualized models can potentially be more diagnostic and enable more personalized aid or support to be put in place.

A series of three experiments was planned to test the utility of psychophysiological monitoring as a strategy for advancing the value of robots as tools to teammates. Achievements of the first experiment included the development and validation of experimental tasks that can be manipulated to study the effects of different levels and types of external task load on operator workload and performance. Using

these tasks, the second experiment examined the effects of different task load manipulations on various workload measures (e.g., EEG, ECG, TCD, fNIR, Eye metrics). In the present experiment, the physiological measures will be used to drive a closed-loop system via an individualized workload model. The closed-loop system would enable the robot teammate to sense the human teammate's workload state (being a form of implicit communication between human and robot), allowing it to render better, context-sensitive support to its human teammate. The model development process will be the primary focus for this talk. Questions for discussion will be included.

1330 - 1400 **The Effects of Information Level on Human-Agent Interaction for Route Planning**

Ms. Julia Wright, Army Research Laboratory; Mr. Michael Barnes, ARL-HRED

The effects of level of information (LOI) and information display type on human decision making in the context of dismounted infantry navigation were evaluated. In two experiments, participants supervised a simulated dismounted soldier team from a remote location, choosing the team's route as they moved from checkpoint to checkpoint through a simulated urban environment. The participants' task was to ensure that the soldier team arrived at the final destination with an adequate amount of resources for successful mission completion. Experiment 1 examined how information about resource usage/requirements affected route selection decisions. Experiment 2 increased the amount of information from Experiment 1 by adding a robotic asset to the unit and adding its resource usage/requirements. Results show that decision time increased as the level of information increased, but did not predict mission success. One-third of participants in Experiment 2 did not complete their mission with the required amount of robot resources, although all completed their mission with the required minimum of human resources. As the amount of information increased, preference for specific information presentations began to vary. In the condition with the greatest amount of information available, participants displayed no clear consensus as to preferred information presentation, with many indicating they relied upon presentation types that were unsuitable for successful mission completion. These findings demonstrate that presentation of information is as important as the amount of information in decision-making tasks.

1400 - 1430 **The Effects of Agent Transparency on Human Interaction with an Autonomous Squad Member**

Mr. Anthony Selkowitz, Institute for Simulation and Training at UCF; Dr. Michael Boyce ; Mr. Shan Lakhmani; Dr. Daniel Barber; Dr. Jessie Chen

A human in-the-loop experiment was conducted examining the effects of presenting information, to support agent transparency, in the display for an autonomous robotic agent. The information supporting agent transparency was displayed in accordance with the Situation awareness-based Agent Transparency (SAT) model. Level of SAT display was a between subjects variable. The response variables examined were the operator's Trust in Automated Systems, Situation Awareness, and Workload. The results suggest that presenting information to support agent transparency aids the operator in properly calibrating trust. Differences in situation awareness were observed among conditions. No differences in workload based on SAT information displayed were observed. These results suggest that providing information to support agent transparency can have different effects based on the amount of information presented to the operator.

SubTAG: Safety/Survivability/Health Hazards

Chair: John Plaga Co-Chair: MAJ Jay Clasing

Tuesday, 5 May 2015
Session: I

1245 - 1310 **Using Biomechanics to Improve Prediction of Physical Training Outcome: Overuse Injury and Army Physical Fitness Test (APFT) Models**

Dr. Bryant Sih, L-3 Applied Technologies; Dr. Charles Negus, L-3 Applied Technologies

BACKGROUND: US Army Basic Combat Training (BCT) is a nine-week program designed to improve the fitness of recruits to meet future operational demands. This training can also lead to negative outcomes including musculoskeletal overuse injuries and failure to meet performance standards. Traditionally, statistical methods have been employed to predict BCT outcomes and identify individuals at risk of a negative outcome. However, these results have limited applicability; they only apply to the BCT scenario from which they were derived and do not account for biological changes such as increased fitness due to the training regimen and/or loss of strength from fatigue. Thus, the goal of this project was to develop biomechanical- and physiological-based (B-P) predictive models as tools to individually assess a Soldier's propensity for injury during BCT, as well as their potential to pass the running portion of the end of cycle Army Physical Fitness Test (EOC APFT).

METHODOLOGY: Retrospective data from military training cohorts and their training regimen were used as input to a set of B-P models established as important in the pathway leading to training adaptation and injury. These models account for overall fitness, muscle strength, physiological loading, and bone integrity on an individual-by-individual basis. Model outputs (individualized tibial strain and fitness level) were correlated to observed overuse injury and APFT outcome. Statistically-based logistical and linear regressions were also derived to compare with B-P model accuracy. RESULTS: We find that the overuse injury model is most accurate when the effect of training activities on both overall fitness as well as muscle fatigue during activities is accounted for (area under the receiver-operator curve, AUC of 0.65). This compares favorably with the statistical-based models (AUC ≈ 0.56). The APFT outcome model for running was also better, predicting both the EOC as well as intermediate APFTs (R2 range 0.55-0.59), when compared to linear regression models that used the same intrinsic input variables (R2 range 0.36-0.50).

CONCLUSIONS: We conclude that accounting for B-P changes due to training and fatigue are important factors in the prediction of BCT outcome. In addition, overuse injury modeling remains more challenging than predicting APFT outcome since “injury” is a more subjective outcome and encompasses a variety of etiologies, yet inclusion of B-P factors still improved injury model predictability. Nevertheless, the ability to identify a training sub-cohort whose injury rate and/or fitness level is significantly different than the nominal rate could lead to savings in training and medical costs. This project also highlights the value of modeling to account for changes in biomechanics and physiology with time, which may be applicable in the outcome prediction of other military-relevant scenarios.

1310 - 1335 **The Cognitive Effects of Long-Range and Rough-Water Transit on High-Speed Craft Operators**

Dr. H.C. Ganey, NSWC-Carderock; Dr. Peter Hancock, UCF; Dr. Justin Morgan, Battelle

It has been long known that the rigors associated with long range and rough water transits take a physical toll on high speed craft operators (see Ensign et al, 2000 & Myers et al, 2008). It has also been suspected that those transits take a mental or cognitive toll on the operators, as well. Considered in the

framework of the extended-U model (Hancock & Warm, 1989), it would be expected that cognitive effects would present well before physical symptoms, but this has not been studied in high-speed craft operations. This presentation will consist of a discussion of the historical findings of cognitive impacts resulting from environmental stressors similar to those found in high-speed craft operations, as well as a framework for a program of study. References Ensign, W., Hodgdon, JA., Prusaczyk, WK., Shapiro, D., and Lipton, M. (2000). A survey of self-reported injuries among special boat operators. Naval Health Research Center Report 00-48. DTIC No. ADA421234. Hancock, PA., & Warm, JS. (1989). A dynamic model of stress and sustained attention. Human Factors, 31, 519-537. Myers, SD., Dobbins, TD., Hall, B., Gunston, T., Holmes, S.R., King, K. and Dyson R. (2008). The effectiveness of shock mitigation technology in reducing motion induced fatigue in small high speed craft. Pacific 2008 International Maritime Conference.

1335 - 1400 **Medical Cost Avoidance Model**

LTC Jay Clasing, Army Institute of Public Health

The MCAM was originally developed in 1998 and updated in 2005 using Headquarters, Department of the Army Study funding. It was published in the Defense Safety Oversight Council Injury Prevention Report No. 12-HF-04MT-08 in December 2008 and published in the January 2010 edition of the American Journal of Preventive Medicine. The MCAM was updated in 2013 with data extracted from the Military Health System Management Analysis Reporting Tool (M2) database for fiscal years (FY) 2010, 2011, and 2012 along with Army Manpower Cost System (AMCOS) and Veteran Administration data for disability compensation. This web accessible MCAM application predicts medical and lost time costs using M2 direct care and purchased care medical costs data, Military Personnel Cost data, and Veterans Affairs (VA) Disability Compensation data. These data sources are used to calculate cost factors such as, average clinic costs, average daily hospital costs, and expected lost time, and fatality costs for active duty Army personnel. The Hazard Analysis tab of the MCAM application is based on the concept of hazard severity and hazard probability and uses the cost factors and materiel system information in algorithms developed to calculate costs and present the likely monetary impact of unabated medical and lost time injuries. The International Classification of Diseases (ICD)-9 Analysis Tab of this application is designed to present medical and lost time costs based on actual Army medical records. In addition, an estimate is provided for potential fatality based on incidence rates in the Army active duty population and VA disability compensation costs based on the prevalence in the Army veteran population.

1400 - 1420 **Auditory Level-Dependence and Nonlinearity in the Auditory Hazard Assessment Algorithm for Humans (AHAAH)**

Dr. Paul Fedele, U.S. Army Research Laboratory; Dr. Joel Kalb, ASARL

ABSTRACT The Auditory Hazard Assessment Algorithm for Humans (AHAAH) (http://www.arl.army.mil/ahaah) is an advance in the evaluation of hearing damage risk associated with impulsive noise. AHAAH applies pressure response dynamics of the external, middle, and inner ear, to bio-mechanically model the ear's physical response to impulsive sound, allowing the accurate determination of strain-induced fatigue occurring in the cochlea's organ of Corti. The Hearing Protector Module (HPM) included with version 2.1 of the AHAAH software includes both level-independent linear and level-dependent nonlinear hearing protectors. Nonlinear behaviors can arise in hearing protection applications in AHAAH. But even without nonlinear hearing protection, AHAAH will exhibit nonlinear behavior associated with nonlinearity in the middle ear. Nonlinearity in the middle ear has been

observed by Guinan Jr, J. J., & Peake, W. T. (1967) and calibrated for the human ear by Price, G. R., & Kalb, J. T. (1991). Nonlinear middle ear models have been described by Pascal, J., Bourgeade, A., Lagier, M., & Legros, C. (1998). Human test results produced by Johnson (1966, 1993, and 1997) were used to validate AHAAH's application of a nonlinear middle ear model by Price, G. R. (2007). Presented analyses results will show that middle ear nonlinearity means that the risk of hearing hazard is not predictable based on waveform energy (A-weighted or not) or waveform peak pressure. The risk of hearing hazard does not necessarily behave monotonically with any summary waveform characterization. Although the AHAAH may seem complex, it analyzes response to the full time-dependence of the waveform to accurately analyze hearing damage risk through the nonlinear elements of the human middle ear.

SubTAG: Personnel

Chair: Hector Acosta Co-Chair: LT Mike Natali, USN

Wednesday, 6 May 2015
Session: I

1240 - 1310 **The Aviation Selection Test Battery - E: Improving Greatness**

LT Michael Natali PhD, USN, Naval Aerospace Medical Institute

The Aviation Selection Test Battery – E (ASTB-E), released in December 2013, is the newest version of the selection tests used to select naval aviation candidates for the United States Navy, Marine Corps, and Coast Guard. The new version features a computer adaptive format (CAT) as well as an entirely new psychomotor skill assessment: the Performance Based Measures Test (PBM). These new additions should improve the test's validity and help the services find better qualified aviation candidates. Three recent studies examined the performance of these new additions to the ASTB. In the first study, the performance of the computer adaptive testing algorithm was examined in simulations versus operational samples. Characteristics of test length, item exposure frequency, and item bank usage were compared between the simulated and operational samples across subtests with item banks of varying size and quality. Operational tests typically required more questions to reach the standard error of measurement for termination criteria (i.e., longer test lengths), but item exposures for the most utilized items were lower. The second study examined practice effects and how they may differ with the new CAT format and PBM measure. Results showed smaller increases in scores on the cognitive ability measures for the new CAT format versus standard versions of assessments. However, for the PBM, practice effects were larger than what is typically seen in cognitive ability tests. The third study examined whether and to what degree video game, flight simulator, and hands-on-throttle-and-stick (HOTAS) experience contributed to the prediction of psychomotor-based selection test scores and subsequent flight training performance for a sample of student naval pilots. Results suggested that while video game and flight simulator experience were associated with higher performance on the psychomotor test, only experience with HOTAS devices while playing flight simulators significantly contributed to the prediction of flight training grades. The results of these three studies provide initial insight into the performance of the new components of the ASTB.

1310 - 1340 **Manned to Unmanned Aerial System (UAS) Knowledge, Skill, and Ability (KSA) Comparison**

Ms Jennifer Pagan, Naval Air Warfare Center Training Systems Division; Dr Randy Ashwood, Training Systems Division (NAWCTSD) AIR, Science & Te; CDR Henry Phillips PhD, Training Systems Division (NAWCTSD) AIR - Science & Technology

Currently, no empirically validated qualification standard exists for Unmanned Aerial System (UAS) operators (Howse, 2011). For example, some Naval and Air Force UAS platforms (e.g., Triton, Fire Scout, Predator) require their operators be winged aviators. This involves a $1 million investment per pilot and years of pilot training, in addition to mandatory, UAS platform specific training (Cohn, 2012). The Shadow UAS program, on the other hand, uses junior to mid-grade enlisted personnel with no aviation experience. The training program for Shadow pilots is 10 weeks long and approximately a third of the investment (about $347,000) of manned aviators (Cohn, 2012). While adapting a Shadow-like selection/training model could yield significant cost avoidance, thorough research is necessary to develop qualification and training standards that support identification of the most qualified people to operate UAS and who will be most likely to succeed in training and operations (i.e., select the right individuals capable of acquiring these UAS specific skill sets). These differences in standards may be driven more by the relative size and cost of different UAS platforms rather than by empirical comparison of the Knowledge, Skills, and Abilities (KSAs) underlying performance in each (Howse, 2011). This paper describes differences between KSAs required by manned and unmanned platforms, reasons underlying those observed differences, and implications of the observed trends for selection criteria, training requirements, and system design.

1340 -1410 **Individual Differences in Performance, Trust, and Stress during Multi-UAV Operation**

Dr Gerald Matthews, Institute for Simulation & Trng, U of Central Florida; Lauren Reinerman-Jones, University of Central Florida; Peter Chiu, University of Cincinnati; Gloria Calhoun, Air Force Research Laboratory; Heath Ruff, Infoscitex

Operating Unmanned Aerial Vehicles (UAVs) in a pilot or sensor operator role makes rather different cognitive and affective demands to traditional aircraft piloting. Future UAV operations that require the pilot to manage multiple vehicles will necessitate expertise in multi-tasking, handling workload transitions, and appropriate use of automation. UAV operation also imposes less physical stress than manned flight, but may impose additional cognitive stress associated with concerns over effective coping with workload, as shown in a recent psychophysiological study in our lab. To investigate the role of individual differences in UAV operation, we ran an experiment, supported by the Air Force Office of Scientific Research (AFOSR), using the Adaptive Levels of Autonomy (ALOA) simulation in which 101 participants were allocated to higher or lower workload conditions. The level of automation (LOA) was also manipulated. ALOA is a realistic simulation requiring control of three UAVs, and entailing various subtasks, including surveillance tasks. We investigated predictors of performance on the surveillance tasks, reliance on the automation of the tasks, and subjective stress response. Predictors included demographic factors including videogame experience, gender, and personality (Five Factor Model). We found that experience with videogames, especially action games, was associated with more accurate surveillance performance, greater reliance on automation, and greater subjective engagement. Gender differences in performance and reliance favoring men appeared to be a side-effect of greater male involvement in videogames. We also found that the personality factor of conscientiousness predicted greater reliance on automation. All these associations between predictors and outcomes were stronger in the high than in the low workload condition. We also examined associations between subjective stress dimensions and objective measures of performance and reliance. Distress was associated with lower accuracy, lower reliance, and greater neglect of targets to be surveilled. Associations between distress and performance impairment were stronger when workload was high. These findings are

consistent with previous studies suggesting that distress specifically disrupts executive control of attention and multitasking. Also, under high workload, lower task engagement was associated with greater neglect of the surveillance task, consistent with findings from fatigue research. These findings suggest the value to the Air Force of recruiting from non-traditional populations such as videogamers. Training in use of automation should be personalized according to the individual's gaming experience and personality. The impact of distress under high workload also reinforces the need to select for resilience under cognitive demands, or to train the operator in demand management.

1410 - 1425 **Way-ahead, Part 1: Toward Developing a Modern Personnel Selection-Classification Systems Taxonomy**

Dr Hector Acosta, HQ Air Force Recruiting Service; LT Michael Natali, Naval Aerospace Medical Institute

First of two 15-minute Open Sub-TAG discussions to address support for cooperatively evolving an up to date and forward-looking taxonomy for proactive use across HFE-TAG participating Services and Agencies. Open panel discussion will begin with a 5 minute groundwork presentation by the Sub-TAG chairs.

Session: II

1505 - 1535 **The Navy's ASVAB Validation/Standards Program**

Ms. Janet Held, NPRST (BUPERS-1)

The Navy's ASVAB Validation/Standards Program Janet D. Held Navy Personnel Research, Studies, and Technology (BUPERS-1) Abstract High failure rates in technical training are costly to both the Navy and Sailors. Navy incurs costs in having to reclassify failed Sailors to other training and also having to more intensely manage the disruptions in the continual Sailor supply chain making its way through training pipelines. Sailors incur costs in terms of career setbacks and, potentially, negative morale that can lead to demotivation, poor job performance, and subpar evaluations that impact promotions. Setting effective aptitude/ability standards for the Navy's 85 occupations (Ratings) is an important Military Personnel Selection and Classification (S&C) function that is intended to maximize the number of Sailors graduating from initial technical training, generally using the Armed Services Vocational Aptitude Battery (ASVAB). At the same time, however, the S&C function must address the Navy's ability to fill all Ratings without having to issue many if any large ASVAB score point waivers. This balance between training and recruiting needs is best assessed through careful attention to the more than several technical areas involved in setting effective aptitude/ability standards. Presentation topics include (a) estimating ASVAB criterion related validity coefficients, (b) estimating the negative impact on training graduation rates when having to lower an ASVAB cutscore, (c) the study of cognitive tests that have been demonstrated to increase ASVAB validity, but at the same time improve gender/ethnic representation thereby increasing aggregate Rating qualification rates, and (d) simulation applications used to estimate the impact of changing one Rating's ASVAB standard on the fill of ASVAB qualified Sailors across all Ratings.

1535 - 1605 **Way-ahead, Part 2: Toward Developing a Modern Personnel Selection-Classification Systems Taxonomy**

Dr Hector Acosta, HQ Air Force Recruiting Service; LT Michael Natali, Naval Aerospace Medical Institute

Second of two 15-minute Open Sub-TAG discussions to begin to define a plan to cooperatively evolve an up-to-date and forward-looking taxonomy for proactive use across HFE-TAG participating Services and Agencies. 2nd open panel discussion toward planning details after summary of prior discussion vectors by the Sub-TAG chairs.

1535 - 1605 **Beyond the Dichotomy of Pass/Fail Training Criteria in High-Attrition Air Force Specialties**

Dr Gregory Manley, HQ Air Force Personnel Center; Dr Mark Rose, Air Force Personnel Center; Dr Hector Acosta, HQ Air Force Recruiting Service

Complex multi-course training pipelines have historically made training performance criteria a challenge. Six hard-to-fill Special Operations/Combat Support (SO/CS) Air Force Specialties (AFSs) have high value and high attrition; and up until recently, have generated training performance data limited to overall dichotomous pass/fail outcomes. Using this dichotomous training outcome as a criterion for statistical prediction modeling limits choice of analysis to categorical data procedures such as logistic regression or discriminant analysis and may utilize less information than a criterion measuring the extent of training completion. Recently, statistical prediction models were developed for the six SO/CS specialties using elements from the Armed Services Vocational Aptitude Battery (ASVAB; Segal, 2004), the Tailored Adaptive Personality Assessment System (TAPAS; Drasgow, Stark, Chernyshenko, Nye, & Hulin, 2012), and the Physical Ability and Stamina Test (PAST) using a dichotomous pass/fail criterion. These six models were shown to be reliable and predictive, thus are used as operational selection models with corresponding cut scores. Recently, training outcome data have become available allowing the calculation of alternative "pipeline penetration" outcome scores, which measure the extent to which trainees progress through the lengthy training pipeline prior to failing or passing.

The purpose of the current effort is to assess the modeling precision of increasing training outcome resolution with a devised metric that expands the training criterion beyond the pass/fail dichotomy. This ordinal scoring scheme assigns integer scores ranging from 1 to 7, indicating the extent of training completion prior to failing. This metric takes advantage of the fact that all six specialties accumulate the vast majority of attrition (93% to 100%) across Basic Military Training (BMT) and exactly 2 other technical training courses. This pattern of attrition allows computation of a simple ordinal metric by assigning a score of "1" for Entry into Active duty with an assigned SO/CS AFS and 1-point added for achieving each of 6 subsequent training milestones associated with Entry into and Graduation from the specified 3 courses. This effort begins a program of research examining the underlying hypothesis that the extent of training completion before elimination/graduation ("pipeline penetration") adds meaningful information concerning trainee quality. The use of a more continuous criterion scale has the added benefit of potentially reducing adverse impact on minorities due to the selection model which can lead to greater diversity (Aguinis, Bommer, & Pierce, 1996).

References

Aguinis, H., Bommer, W.H., & Pierce, C.A. (1996). Improving the estimation of moderating effects by using computer-administered questionnaires. Educational and Psychological Measurement, 56, 1043-1047.

Drasgow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions (Technical Report 1311). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Segall, D. O. (2004). Development and evaluation of the 1997 ASVAB score scale (Technical Report No. 2004-002). Seaside, CA: Defense Manpower Data Center.

****

RQ1a:

Will a model developed from a 7-point criterion have more precision in predicting training pipeline completion than a model developed from a dichotomous pass/fail criterion for mulit-course training pipelines with high early attrition?

RQ1b:

To what degree will a model developed from a 7-point criterion have more precision in predicting training pipeline completion than a model developed from a dichotomous pass/fail criterion?

H1:

A model developed from a 7-point criterion will have more precision in predicting training pipeline completion than a model developed from a dichotomous pass/fail criterion.

1605 - 1635 **Examination of the Predictive Validity of the ASVAB for Several USAF Enlisted Training Specialties**

Dr. Thomas Carretta, 711th HPW/RHCI; Dr Gregory Manley, Air Force Personnel Center

The Armed Services Vocational Aptitude Battery (ASVAB) is used by all branches of the US Military for enlistment qualification and to assign qualified applicants to training specialties. The primary purpose of the current study was to examine the predictive validity of the US Air Force classification composites and the Armed Forces Qualification Test versus initial training performance. A secondary purpose was to determine whether switching from the current classification composite to another would improve prediction of training performance. The sample consisted of 117,232 enlisted personnel who attended training between 2006-2013. Data were available for 111 Air Force Specialties. High levels of predictive validity were observed for most training specialties. After correction for range restriction, the mean correlation between the current classification composite and training performance was .70, weighted by course sample size. Several instances were identified where the current classification composite for a training specialty was not the one with the highest predictive validity for that specialty. Other instances were observed where an alternate composite had similar validity to the operational composite, but would reduce adverse impact. Additional analyses of training content and qualification rates for sex and racial/ethnic groups for the current and alternate composite are needed to determine whether switching from the current composite to another is warranted.

SubTAG: Training: Medical/Healthcare Applications

Chair: Jen Pagan Co-Chair: Kelly Hale

Wednesday, 6 May 2015
Session: I

1505 - 1525 **Modeling & Simulation Applications in the Medical Arena**

Joseph Cohn, USN OSD OUSD

Recently, there has been renewed interest in applying modeling and simulation (M&S) technologies to the challenge of training today's – and tomorrow's – medical practitioners. Since the applications of medical training tools based on M&S technologies are vast, it is also important to acknowledge that many of the challenges that medical training technology developers and end-users will likely face, have in some form been addressed by training technology developers and end-users from other application domains. At the same time, medical domain – specific challenges will certainly be encountered as M&S training systems become more common place. Consequently, it is useful and necessary to review the extent to which existing simulation and training development principles and guidelines can be leveraged to drive medical M&S training development, and to consider what new principles and guidelines may need to be developed (not discussed) as medical M&S technologies become more commonplace in medical education and training simulation and modeling.

1525 - 1545 **Simulation for Healthcare Training**

Prof Michael McCauley PhD, Naval Postgraduate School; Dr. Joe Sullivan PhD, VA/Naval Postgraduate School

Modeling and Simulation for Healthcare Training The Naval Postgraduate School (NPS) and the Uniformed Services University of the Health Sciences (USUHS) have collaborated to develop and deliver a four-course "Certificate" program on the topic of modeling and simulation (M&S) for healthcare training. The program is currently in its third year and is aimed at physicians, nurses, and healthcare simulation center personnel. In this session, we will describe the topics and curriculum of the certificate program, the research proposals that have been developed as part of a Capstone project, and the lessons learned from the first two years of the program. Issues to be addressed include Instructional Design, fidelity, feedback, non-technical skills, and measures of training effectiveness.

1545 - 1605 **Surgical simulation: Addressing needs, gaps, and targeted competencies**

Dr. T. Robert Turner Jr., American College of Surgeons

This presentation will include a high-level overview of several current initiatives spearheaded by the American College of Surgeons (ACS) whereby needs and gaps in simulation-based surgical education and training will be formally identified and addressed via consortium. Within this framework, the primary objectives and challenges of an ACS consortium model will be presented. Additional discussion will highlight efforts to classify specific skill domains, simulation models, learner groups, training environments, and standards to enhance simulation training for both residents and practicing surgeons. The opinions expressed and statements made in this presentation reflect the presenter's personal observations and do not imply endorsement by or official policy of the American College of Surgeons.

1605 - 1625 **Alternative Approach to Establish Inter-Rater Concordance**

Ms. Catherine Strayhorn, IVIR Inc.

Background: Effective inter-rater reliability necessitates rigorous training of observer/controllers to a unified standard. Training with an established scoring protocol often proves time-consuming to achieve this goal. The University of Missouri's Combat Casualty Training Consortium developed a method to anchor observer/controllers utilizing a software application that facilitated the training through real-time feedback resulting in strong inter-rater reliability in a short period of time. Methods: The site teams

developed a “Steps Dictionary” with a precise definition of “Accomplished” for each decomposed clinical step. The Observer/Controllers then practiced rating with this scoring rubric as the Principal Investigator verbalized the actions of a student performing a lifesaving procedure. The Observer/Controllers were tested with electronic checklists in the simulation lab as a professional role-played a student performing the procedures with planned mistakes. The Principal Investigator created the Master scores as the Observer/Controllers scored the role-player. The application collected the scores, immediately generated a report that demonstrated the percentage of agreement with the Master scores, and discussions of individual and group performance followed. The Observer/Controllers continued the trials until the inter-rater reliability achieved 90%. The calculation of reliabilities was a simple averaged comparison of each evaluator's score against the Master score. Results: The three study sites achieved >90% inter-rater reliability related to 11 decomposed tasks. Months later, the Observer/Controllers were re-tested and scores suggested that individual Observer/Controllers consistently rated student performance throughout the study. Conclusion: The immediate feedback software application provided an efficient method for training Observer/Controllers to reliably rate student performance resulting in strong and sustained inter-rater reliability.

1625 - 1645 **Adapting Immersive Training Environments to Develop Squad Resilience Skills**

Dr. Joan Johnston, Army Research Laboratory HRED; Ms. Samantha Napier, ARL HRED; Mr. Bill Ross, Cogn Perf Gp

This presentation is an update to one given at HFETAG 68 in 2014. The United States Army defines readiness and resilience as tactically proficient Soldiers and highly adaptive problem solvers capable of overcoming challenges and making decisions with strategic consequences in ambiguous situations. To address the resilience training gap, the Squad Overmatch study in 2014 produced recommendations for employing immersive and live training strategies within the Stress Exposure Training (SET) framework. SET is a three-phase training method de-signed to provide information, skills training, and practice; with the goal of learning how to cope and perform while exposed to combat stressors. The potential for a wide range of Soldier experience levels in the pre-deployment training phase presents challenges to effectively structuring training within the SET framework. This presentation provides recommendations for adapting immersive environments depending on levels of leader and squad expertise that could optimize the SET experience.

1645 – 1700 **Guidance for the HFM-RTG 258 on the Impact of Military Life on Children from Military Families**

Dr. Paul Chatelier, Naval Postgraduate School

CAPT(R) Chatelier's presentation will cover the latest NATO HFE related activities being pursued by the NATO Alliance. It will cover the NATO organization that supports Science and Technology as well as that which supports Training. He will quickly drill down to some representative activities of possible interest to the HFETAG. The goal of this presentation is to (a) acquaint the HFE TAG with the breadth and depth of the international interests and (b) to describe how government representatives could get involved. Copies of the presentation will be available via the HFE TAG management team. He will save time for Q&A if needed.

SubTAG: Human Performance Measurement

Chair: LCDR Jeff Grubb Co-Chair: Rahel Rudd

Wednesday, 6 May 2015
Session: I

1505 - 1600 **Proposed Human Performance Markup Language SISO Standard Discussion**

Walker Alex, Aptima, Inc.

1600 - 1615 **FMRI Study of Landing Decision in Civilian Pilots**

Dr. Maheen Adamson, VA Palo Alto/Stanford Medical School; Dr Jerome Yesavage, VA Palo Alto/Stanford Medical School

The most common lethal accidents in General Aviation are caused by improperly executed landing approaches in which a pilot descends below the minimum safe altitude without proper visual references. To understand how expertise might reduce such erroneous decision-making, we examined relevant neural processes in pilots performing a simulated landing approach inside a functional MRI scanner. Pilots (aged 20–66) were asked to "fly" a series of simulated "cockpit view" instrument landing scenarios in an MRI scanner. The scenarios were either high risk (heavy fog–legally unsafe to land) or low risk (medium fog–legally safe to land). Pilots with one of two levels of expertise participated: Moderate Expertise (Instrument Flight Rules pilots, n= 8) or High Expertise (Certified Instrument Flight Instructors or Air-Transport Pilots, n = 12). High Expertise pilots were more accurate than Moderate Expertise pilots in making a "land" versus "do not land" decision (CFII: d' = 3.62±2.52; IFR: d' = 0.98±1.04; p<.01). Brain activity in bilateral caudate nucleus was examined for main effects of expertise during a "land" versus "do not land" decision with the no-decision control condition modeled as baseline. In making landing decisions, High Expertise pilots showed lower activation in the bilateral caudate nucleus (0.97±0.80) compared to Moderate Expertise pilots (1.91±1.16) (p<.05). These findings provide evidence for increased "neural efficiency" in High Expertise pilots relative to Moderate Expertise pilots. During an instrument approach the pilot is engaged in detailed examination of flight instruments while monitoring certain visual references for making landing decisions. The caudate nucleus regulates saccade eye control of gaze, the brain area where the "expertise" effect was observed. These data provide evidence that performing "real world" aviation tasks in an fMRI provide objective data regarding the relative expertise of pilots and brain regions involved in it.

1615 - 1630 **Systems-Based Metrics of Unit Performance Using Communications Data**

Dr. Arwen DeCostanza, Army Research Laboratory; Dr. Kara Orvis, Aptima, Inc.

Teamwork is imperative for success at every military echelon, from squads to division staffs.

Over 50 years of teams research has provided strong evidence of the criticality of both team processes

(e.g., mission analysis, coordination) and team states (e.g., cohesion, shared mental models) for team effectiveness (e.g., Kozlowski & Ilgen, 2006; LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). These concepts are also embraced and emphasized as critical to mission effectiveness in Army doctrine (e.g., US Army, 2012a, 2012b). In fact, team states such as cohesion, trust, and shared understanding are at the core of the six principles of mission command as stated in Army Doctrine Research Publication (ADRP) 6-0 (US Army, 2012b, p. 2). Given the importance of these concepts for mission-related

outcomes, the military needs a better way to assess and track how these states and processes emerge, ebb, and flow so they can intervene before existing or potential future problems put the mission at risk.

This research addresses the need to better understand the nature of team states and processes especially in the complex environment that defines the Army, and develop practical, scientifically-valid, unobtrusive real-time

assessments that robustly aggregate across available data sources and help track unit states and processes over time. In the last decade, the military has produced an exponentially growing amount of continuous data due to the rising use of mission command systems and other communication technology (e.g., data sensors) that can be utilized to provide automated metrics of collective performance. Since 2008, a large-scale collaborative effort

involving government organizations (Army Research Laboratory, Army Research Institute, DARPA, and Mission Command Battle Lab) and industry partners (Aptima, Inc.) has focused on utilizing existing systems-based data for unobtrusive, automatic, real-time measurement of team performance. This presentation will begin by providing an overview of a suite of software, ACCRUE (Automated Collaboration Collection and Relationship Understanding Environment), developed to 1) collect and organize communications data such as email, chat, and face-to-face interactions using Sociometric badges; 2) analyze these data for indicators of team performance; and 3) display the levels of these indicators in real-time. An overview of the Command Operations Dashboard, a user interface developed for training environments to support observer, coach, trainers (OCTs) with real-time qualitative and quantitative data on team performance, will also be provided.

Next, data collection efforts for validating systems-based indicators of team states and processes critical to performance (e.g. cohesion, trust, situation awareness, understanding commander's intent) will be described.

For data collections, ACCRUE has been employed at several large-scale Mission Command Battle Lab and Mission Command Training Program (MCTP) exercises (Division and Brigade levels) to collect, aggregate, and visualize unit communications data including email, chat, and face-to-face interactions via Sociometric badges. Survey-based data and observer ratings were also collected to validate systems-based indicators. Using the communications data, a number of indicators of specific states and processes including shared situation awareness, understanding of command intent (shared mental models), and cohesion were developed using network-based and text-based analytic techniques. A summary of results focused on the relationship between these communications-based indicators of critical team performance-related constructs and survey data will be discussed. This work builds on evidence that communications and other types of organizational data yield rich indicators of cohesion and other team states (e.g., Orvis, Duchon, McCormack, & DeCostanza, 2012).

This presentation will end with a discussion of next steps in performance metric development and validation using system-based indicators of team processes and states. Particular emphasis will be placed on (1) considering temporal issues related to the unit states and developing measures that include a temporal element, (2) developing measures that are specific to large multi-team systems, instead of small teams, and (3) task-based versus generic team-based measures. Next steps for the transition of research products and findings will also be addressed.

References

Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams.

Psychological Science in the Public Interest, 7(3), 77-124.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. Personnel

Psychology, 61(2), 273-307.

Orvis, K., Duchon, A., McCormack, R. K., & DeCostanza, A. H. (2012). Constructing systems-based measures of team-related constructs: Validation of a method. Presented at the 27th Annual Conference of the Society for

Industrial and Organizational Psychology (SIOP), San Diego, CA.

US Army. (2012a). Army Doctrine Publication (ADP) 5-0: The Operations Process. Fort Leavenworth, KS: United States Army Combined Arms Center.

US Army. (2012b). Army Doctrine Reference Publication (ADRP) 6-0: Mission Command. Fort Leavenworth, KS: United States Army Combined Arms Center.

1630 - 1645 **Translating Research on Training Effectiveness to Requirements for Simulation Acquisitions**

Dr. Gregory Goodwin, U.S. Army Research Laboratory

Despite numerous requirements and a plethora of how-to guides, manuals, and references for conducting training effectiveness analyses (TEAs), such analyses of simulation training systems are rarely done. There seem to be many reasons for this, but one of the biggest has to do with the cost and technical challenge of conducting well-designed TEAs. Most TEAs require data collectors and subject matter experts to be present during training events to record and code performance measures from numerous individuals across different treatment conditions. Often the training needs of the participating individuals and units require significant compromises in the experimental design and may even limit the kinds of data that can be collected. The data must then be organized, analyzed and interpreted by individuals with expertise in inferential statistics and human behavior. To reduce this high overhead for conducting a TEA, simulation-based training systems need to incorporate the ability to automatically assess performance and provide summaries of training effectiveness to instructors, leaders, and researchers. With advances in technologies such as inexpensive, network-ready sensors, adaptive training systems, and powerful handheld computers, research is needed to examine the ways in which these technologies can improve human performance measurement. In this paper, we identify and discuss some of the research challenges associated with these kinds of performance assessment capabilities. These challenges include developing measures that are interoperable with adaptive training systems, learning and resource management systems and that can be used in joint and combined-arms training events. Other challenges have to do with developing meaningful formative measures that can be used to diagnose the causes of performance errors rather than simple summative measures of success and failure. These challenges are discussed from the point of view of a research program examining automated performance measures for basic rifle marksmanship training. Finally, in this paper we discuss a strategy for developing near- and far-term requirements for performance assessments for simulation based training systems.

1645 - 1700 **Augmenting P-8A Part-Task Training Operations through the Integration of Intelligent, Speech-Activated Role Player Agents**

Dr. Brian Stensrud, Soar Technology; Charles Newton, Soar Technology; Beth Wheeler Atkinson, NAWCTSD; John Killilea, Stracon

A deficiency in existing P-8A Part-Task Training (PTT) systems is a lack of inter- and intra- crew communication. While this deficiency is not critical when practicing simple procedural tasks (e.g., deploying a sonobuoy ordinance), Naval missions increasingly involve communication among onboard crew and coordination with other platforms (e.g., coordination with nearby submarines and ships). Opportunities to practice mission-critical behavioral skills like Crew Resource Management (CRM) are paramount to warfighter readiness and success. Optimally, P-8A flight crews would train alongside other trainees or human role players; however, this is often a cost- and time-prohibitive route. A key requirement for affordable, persistent, on-demand training is the integration of synthetic role players with the following capabilities: (1) fully-integrated behavior models that incorporate all aspects of an unfolding situation, including maintaining shared situational information; (2) speech recognition and speech understanding capabilities that enable natural, free-flowing interactions between the trainee and their training environment; and (3) speech generation capabilities that allow synthetic actors to first dynamically compose and then verbally express their intentions and answer queries.

In this presentation, we will describe a current research effort to augment P-8A PTT operations through the integration of intelligent, speech-activated role player agents. This capability includes autonomous, interactive crewmembers capable of supporting and providing information to, and interacting with, the Tactical Coordinator (TACCO). The TACCO is a front-line Naval Flight Officer in the P-8A aircraft responsible for the tactical portion of a patrol or surveillance flight mission. The TACCO serves as the nexus of mission-level information and must organize functions across the entire flight. Additionally, the TACCO directs sensor operators, develops tactics with aircrew, and informs navigational staff of any mission concerns.

As inter-crew coordination is a key skill in P-8A operations, the synthetic crewmembers will employ novel speech recognition and understanding strategies to support unstructured speech interaction between the trainee and synthetic role players. We anticipate that this capability will help provide crucial skill transfer from simulated training to mission operations.

SubTAG: Unmanned systems

Chair: Ajoy Muralidhar

Thursday, 7 May 2015

0830 – 1100 **Trust in Autonomy Panel Session**

Dr. Lauren Reinerman-Jones, Panel Moderator

Panelists: Dr. James P. Bliss, Mr. Irwin Hudson, Dr. Joseph B. Lyons, Dr. Gerald Matthews and Dr. Jim Szalma

SubTAG: Design Tools & Tech

Chair: Michael Feary Co-Chair: Chelsey Lever

Thurday, 7 May 2015

1 **Step Height and Ramp Slope for Vehicle Ingress/Egress**

MS. Dawn Woods, Natick Soldier Research Development and Engineering Center

Military vehicles often employ ramps to allow Soldiers to quickly ingress and egress their vehicles. Traditionally the ramp length is defined by the height of the vehicle (from the bottom of the body of the vehicle to the top of the body of the vehicle). When the ramp is employed, it is rotated out and down to the ground. The slope of this ramp is therefore based its length (defined by the height of the vehicle body), and the height the vehicle body is above the ground. For vehicle designers who want to increase survivability in the event of an underbody blast, the first line of defense is distance – move the vehicle body further from the blast, which results in raising the vehicle body further above the ground. Once that distance begins increasing, the ramp can become quite steep. One way to mitigate the steepness is for the lower end of the ramp to stop at some distance above the ground, creating an initial step up. This will yield a shallower ramp, but at the cost of having to step up onto and down off of the ramp. Of particular interest is the trade-off between the height of the step, and the slope of the ramp. To explore this tradeoff an apparatus was built which allows for varying the height of the top end of the ramp, and the step height at the bottom of the ramp. Several step height/ramp slope combinations are available and were evaluated with Soldiers. This presentation would describe the apparatus, and present an overview of the experimental design used and resultant recommendations. The apparatus currently resides at the Natick Soldier Research, Development and Engineering Center (NSRDEC) in Massachusetts, and is available for government use.

2 **The Use of Human Models to Verify Anthropometric Accommodation**

Mr. Stephen Merriman, The Boeing Company; Mr. Jay Frank, The Boeing Company

Are the use of Human Modeling Systems (HMS) an acceptable method for verification of physical accommodation of maintenance tasks? Several attempts have been made in the past validate the anthropometric accuracy of digital Human Modeling Systems. This presentation describes a recent project with the goal of validating the anthropometric accuracy of the CATIA, Classic Jack and Vis Jack Human Modeling Systems, and also validating the accuracy of the models' biomechanical posturing. The validation plan and initial results will be discussed.

SubTAG: Controls & Displays

Chair: Marianne Paulsen

Thursday, 7 May 2015

0840 - 0910 **The Bridging Technique: Crossing Over the Modality Shifting Effect**

Dr. Thomas Alicia, NAWCTSD

Operator responsiveness to critical alarm/alert display systems must rely on faster and safer behavioral responses in order to ensure mission success in complex environments such as the operator station of an Unmanned Aerial System (UAS). An important design consideration for effective UAS interfaces is how to map these critical alarm/alert display systems to an appropriate sensory modality (e.g., visual or auditory) (Sarter, 2006). For example, if an alarm is presented during a mission in a modality already highly taxed or overloaded, this can result in increased response time (RT), thereby decreasing operator performance (Wickens, 1976). To overcome this problem, system designers may allow the switching of the alarm display from a highly-taxed to a less-taxed modality (Stanney et al., 2004). However, this

modality switch may produce a deleterious effect known as the Modality Shifting Effect (MSE) that erodes the expected performance gain (Spence & Driver, 1997). The goal of this research was to empirically examine a technique called bridging which allows the transitioning of a cautionary alarm display from one modality to another while simultaneously counteracting the Modality Shifting Effect.

Sixty-four participants were required to complete either a challenging visual or auditory task using a computer-based UAS simulation environment while responding to both visual and auditory alarms. An approach was selected which utilized two 1 (task modality) x 2 (switching technique) ANCOVAs and one 2 (modality) x 2 (technique) ANCOVA, using baseline auditory and visual RT as covariates, to examine differences in alarm response times when the alert modality was changed abruptly or with the bridging technique from a highly loaded sensory channel to an underloaded sensory channel. It was hypothesized that the bridging technique condition would show faster response times for a new unexpected modality versus the abrupt switching condition. The results indicated only a marginal decrease in response times for the auditory alerts and a larger yet not statistically significant effect for the visual alerts; results were also not statistically significant for the analysis collapsed across modality. Findings suggest that there may be some benefit of the bridging technique on performance of alarm responsiveness, but further research is still needed before suggesting generalizable design guidelines for switching modalities which can apply in a variety of complex human-machine systems.

0915 - 0945 **Workload Differences across Simulation Interfaces in the Nuclear Power Plant Domain**

Rebecca Leis, Institute for Simulation and Training; Lauren Reinerman-Jones

Safety-Critical research is concerned with human performance and error, as mistakes and oversights can lead to severe consequences. To discover the cause of these errors, investigators should examine operator states across domains and across task types within each domain (Mercado, 2014). One factor contributing to successful human performance is workload. Scientists have inspected the effects of operator workload for over 50 years, yielding numerous publications (Brannick, Salas, & Prince, 1997; Meshkati & Hancock, 2011). One Safety-Critical domain, Nuclear Power Plant (NPP) operation, requires the completion of multiple task types for day-to-day operations and each task type often involves different types of controls (light boxes, valves, gauges, etc.). Current on-line NPPs are mostly analog systems, built 50-60 years ago. These analog system controls and displays are often difficult to replace because the parts are no longer manufactured. As a result, digital displays are replacing some of the individual analog controls. However, the impact of these digital displays on operator workload, performance, and errors for each task type is unclear. Nonetheless, newly proposed power plants will be adopting fully-digital systems, while existing analog systems will continue to produce energy. Investigation of the impact digital systems have on operators is urgent as these systems are already in the process of being adopted. The need to evaluate the impact of both old and new systems (and the differences between the two) on operators' workload is necessary. It is often ineffective, expensive, and dangerous to measure operator workload in operational settings; thus, simulation techniques provide a safe alternative to real-life environments. It is important to utilize these techniques when investigating NPP operation. As such, researchers at the Institute for Simulation and Training (IST) have investigated these concerns using a simulated replica of a decommissioned NPP Main Control Room (MCR).

Previous experimentation explored operator workload differences across task types using a desktop simulator. This simulation included several stations with dual monitor systems, similar to the new digital AP1000 Pressurized Water Reactor. Multiple types of measurement techniques shown to be sensitive to changes in workload were utilized, such as Electrocardiogram (ECG) Electroencephalogram (EEG) and the NASA-Task Load Index (NASA-TLX). Results from this previous experimentation suggests that

workload was experienced during all task types, but that certain task types elicited different kinds of workload experiences. While new digital systems are being adopted, older analog systems are obtaining re-certification in order to continue running for the next 20 years. Thus, recent investigation explores interfaces that are more similar to existing analog displays and training systems. This line of investigation is necessary to determine if the inferences drawn from the experiment using the desktop display (resembling new digital displays) can be generalizable across both analog and digital NPP controls and displays, as both types of systems will be utilized in the near future. The present talk focuses on operator workload across task types within the NPP domain using various controls and displays. It includes the importance of identifying suitable workload measurement techniques for tasks within the NPP domain and cautions against generalizing measurement usage and appropriateness, particularly when interface and task types and elements differ.

0950 - 0920 **Elevation Errors in 3D Audio**

Dr Dennis Folds, Georgia Institute Of Technology

Three dimensional (3D) audio involves presentation of sounds over headphones that appear to come from an externalized source. Anecdotal evidence suggests that at least some listeners have a pronounced tendency to consistently perceive the virtual audio as elevated higher than intended. The present study was conducted to investigate and compare errors in perception of elevation for 3D and real sources. Eight subjects listened to sounds presented through loudspeakers (behind an acoustic curtain) and through headphones (as virtual sources), and reported their perceived elevation for each sound. Five elevations were used (two above, two below, and one at ear level.) The results show near perfect performance for perception of elevation for the loudspeaker condition. Perception of elevation was far less accurate for the virtual sources, and some subjects consistently perceived virtual sources higher than intended. When used in certain applications, like threat warning systems, these errors could result in incorrect responses and could conceivably lead to decreased survivability.

0955 - 1025 **HSI for COTS**

Mrs. Marianne Paulsen, PSE

SubTAG: Extreme Environments

Chair: Mihriban Whitman Co-Chair: Allison Mead

Friday, 8 May 2015

0840 - 1000 **HSI Challenges in Arctic Naval Operations**

Dr. Gordon Gattie, NSWC Dahlgren HSI

Task Force Climate Change released the United States Navy Arctic Roadmap for 2014 to 2030 in February 2014. This roadmap described the strategic importance of the Arctic Region, with steps required to meet the President's National Strategy for the Arctic Region, and the Secretary of Defense's Department of Defense Arctic Strategy, both published in 2013. The USN Arctic Roadmap identified high-level operational challenges for Arctic operations, nonexistent infrastructure, and large distances in the region. Although Arctic nations have peacefully co-existed through multinational organizations such as The Arctic Council, recent developments along Russian borders require the USN to improve

operational readiness in the Arctic Region. Following years of conflict, the USN finds itself with limited operational experience in cold weather environments, and will need to smartly invest limited resources directed to Arctic security. Although there is a growing recognition for increasing our presence in the region, Naval S&T investments for improving cold weather operations have been limited and Naval acquisition programs contain relatively few cold weather-related requirements. From a human systems integration perspective, HSI challenges include inexperience operating in cold weather environments, operators and maintainers working with equipment not designed for cold weather environments, and the need to constantly monitor exposure rates. However, the Arctic Region may prove useful as a testing ground for evaluating novel command-and-control concepts with greater levels of autonomy, increased coalition operations, and more resilient personnel and equipment. This brief will outline the strategic importance of the Arctic Region, HSI challenges for cold-weather Naval operations, and opportunities for personnel and technological development.

1000 - 1020 **In Vitro Testing Platform For Optical Wearable Fitness Bands, Smart Watches, and other Wearable Devices**

Mr. Casey Pirnstill, Texas A&M University; Professor John Hanks, Texas A&M University; Dr. Gerard Cote, Texas A&M University

Wearable smart bands, watches, and other devices have started implementing optical sensors to measure a variety of fitness and health related parameters such as heart rate, blood pressure, and caloric burn. According to an industry market research firm Canalys [1], around 1.6 million wearable smart band devices were shipped in the second half of 2013. Market trends projected by CEA forecast the total number of wearables in 2015 to reach 20 million units, with expected revenue to surpass $1.8 billion in health and fitness devices alone [2]. As usage starts to increase globally, such devices can offer capability for remote monitoring in the battlefield. The primary optical sensor on most of the current devices is a reflectance based photoplehysmography (PPG) based sensors used for measuring heart rate and hear rate variability (HRV). Although, many improvements have been made with these devices a common problem with these devices concerning heart rate monitoring is their susceptibility to motion artifact during physical activities and the low signal output that occurs with increasing change in skin tones. In this work, we report on the development of an in vitro testing platform evaluating current commercially available technologies to validate performance of currently available smart devices. This platform coupled with optical modeling to simulate the light tissues interactions provides a tool that allows for predicting sensor performance for rapid design iterations on future devices. Current wearables were tested and the point the devices fail at was determined. References: [1] http://www.canalys.com/newsroom/16-million-smart-bands-shipped-h2-2013 [2] http://www.ce.org/News/News-Releases/Press-Releases/2014/Record-Breaking-Year-Ahead-CEA-Reports-Industry-Re.aspx

1000 - 1020 **The Use of Exhaled Breath for the Identification of Hypoxia Biomarkers**

Dr Claude Grigsby, AFRL

Pilots have reported experiencing in-flight hypoxic-like symptoms since the inception of high-altitude aviation. Such an event poses an immediate threat to the pilot, aircraft and civilians on the ground. While pilots undergo extensive training to recognize and react to symptoms of hypoxia, research has shown that hypoxic episodes can affect performance prior to loss of consciousness. Furthermore, these effects can persist for hours following adequate O2 restoration. As a result, the need to monitor pilots,

in-flight, for the onset of hypoxic conditions is of great interest to the aviation community. We propose that exhaled breath is an appropriate non-invasive medium for monitoring pilot hypoxic risk through volatile organic (VOC) compound analysis. To identify changes in the VOCs produced during periods of reduced O2 levels, volunteers at the Naval Medical Research Unit-Dayton were exposed to simulated flight conditions, i.e. O2 levels found at elevated altitudes, using a modified flight mask interfaced with a reduced O2 breathing device. During the course of these test events, time series breath samples were collected from the mask, and analyzed by gas chromatography/mass spectrometry (GC/MS). Additionally, single time point breath samples were collected in 1L ALTEF bags prior to and immediately following the simulated flights. This study establishes an experimental means for monitoring changes in volatile organic compounds in response to hypoxic conditions, a computational workflow for compound analysis via the Metabolite Differentiation and Discovery Lab and MatLab© software and identifies potential volatile organic compound biomarkers of hypoxia exposure.

SubTAG: Cognitive Readiness

Chair: LT Joseph Geeseman Co-Chair: Dr. Laura Milham

Friday, 8 May 2015

1 Comparing the Effects of Transcranial Direct Current Stimulation (tDCS) and Caffeine on Performance and Mood During 30-Hours of Sustained Wakefulness

Ms. Lindsey McIntire, Infoscitex, Inc.; Dr. Andy McKinley, 711th HPW/RHCP; Mr. Justin Nelson, Infoscitex, Inc.; Mr. Chuck Goodyear, Infoscitex, Inc.

Arguably, one of the biggest threats to cognitive readiness is fatigue induced by long periods of sustained wakefulness. Caffeine is a common countermeasure use to combat sleep deprivation. However, the benefits of caffeine last only a few hours and lessen with chronic use. Therefore, finding a longer lasting and more reliable countermeasure is necessary for many career fields across the DoD. Previous research using transcranial direct current stimulation (tDCS), a form of non-invasive brain stimulation, has shown that tDCS can improve attention, working memory, vigilance, and response times. These cognitive abilities are also significantly affect by fatigue; therefore, we hypothesized that using 2mA of anodal tDCS to the prefrontal cortex for 30 minutes could remediate the deleterious effects of performance on sustained wakefulness comparatively to that of caffeine. Three groups of ten participants in each group received either active tDCS with placebo gum, caffeine gum with sham tDCS, or sham tDCS with placebo gum after 21 hours of continued wakefulness. Testing continued until participants had completed a total of 30 hours of sustained wakefulness. Participants completed a 30 minute vigilance task, 10 minute working memory task, and 10 minute psychomotor vigilance task (PVT) beginning at 1800 hours and continued every two hours throughout the night until 1100 the next day for a total of 9 sessions. At the end of each session participants also filled out 3 subjective mood questionnaires. Our results show that once the intervention was given at 0400 that tDCS boosted performance on the vigilance task and that benefit lasted until the conclusion of the study. Comparatively, the caffeine group only had a benefit at 0400 testing before returning to levels comparable to our control group for the duration of testing. The tDCS group performed significantly better (p=.05) than both groups during all sessions once the intervention was given for the vigilance task. tDCS and caffeine performed significantly better (p=.05) on latency in the short-term memory task and reaction time for the psychomotor vigilance task compared to the control group. Subjectively, the tDCS group reported feeling significantly (p=.05) less drowsy, less fatigued, and more energetic than the

caffeine and control groups. Our data suggests that tDCS could be a fatigue countermeasure that is more beneficial than caffeine because the effects last longer while improving mood.

2 Adaptation of Search Strategy and Decision Making from Veracity and Volume of Intelligence information within a Network

Dr. Katherine Gamble, Army Research Laboratory

Advances in information and network technology have transformed military capabilities, offering unprecedented operational advantages, such as an information-rich, broadly collaborative environment. Understanding the human constraints to information processing in networked operational environments is a pressing issue, as Soldiers routinely make life-and-death decisions under time pressure, whether in a command center or in the field. Intelligence information (i.e., intel) is often received with large volume and uncertain veracity. With limited time to act on the information, it is important to understand how use of this information changes in varied contexts. Intel is often transmitted from person to person within a network, so it is also important to understand how Soldiers communicate and share intel with one another to achieve optimal performance for a shared goal. Understanding how Soldiers manage a complex information environment and negotiate competing information and task priorities is critical. Access to too much information at once may result in information overload (Hiltz & Turoff, 1985). In the Army, Mission Command staff constantly interact with large amounts of data, and they must learn how to adjust to changing dimensions to effectively use the information. This study focuses on two issues of Big Data (Ohlhorst, 2013), volume and veracity, or the amount and accuracy of data or intel being received. Human performance is limited by cognitive capacity to both perceive and process information (Cowan et al., 2005), the value of which is subject to contextual or individual differences (Cowan, 2010). Because capacity limits and other environmental factors are usually out of a person's control, individuals must learn to adapt strategies to improve performance. Developing a consistent strategy for visual search (a visual routine; Ullman, 1984) can reduce cognitive demand, freeing mental resources for other tasks (Biggs, Cain, Clark, Darling, & Mitroff)

3 Cognitive Modeling for Personalized Training in Health Care and Medical Domains

Dr. Matthew Walsh, TiER1 Performance Solutions; Mr. Stu Rodgers, TiER1 Performance Solutions

Military education and training is facing a “perfect storm” (Global Horizons, 2013). Constrained resources, increased operations tempo, and the unprecedented complexity of mission requirements necessitate exploring new approaches to skill acquisition and sustainment. In response to these demands, we have developed a novel training method to improve skill acquisition and retention. The method involves shifting from the one-size-fits-all, calendar-driven approach to training that is standard in military domains and health care. In its place, we advocate using computational cognitive models of skill acquisition and retention to prescribe personalized training schedules based on the unique needs and abilities of the individual.

At the center of our approach is a cognitive model of skill acquisition and retention called the Predictive Performance Equation (PPE; Jastrzembski, Rodgers, & Gluck, 2009). PPE is instantiated as running computational software and predicts performance as a function of three factors: (1) total amount of practice, (2) elapsed time since the end of practice, and (3) spacing of practice over time. In general, performance increases with total amount of practice and decreases with elapsed time since the end of practice. The effects of the third factor, spacing, are more dynamic; massing practice across a short duration speeds acquisition, but distributing practice across a longer period of time enhances retention.

Over 100 years of psychological research have demonstrated the impact of these factors on performance (Anderson & Schunn, 2000). PPE accounts for the effects of these factors, and for several other important and widely replicated phenomena documented in the psychological literature on human memory.

To extend PPE to applied contexts, we have adopted a set of Bayesian statistical techniques to calibrate model parameters (e.g., learning rate and decay rate) based on individuals' historical performance data (MacKay, 2003). In this way, we can extrapolate individual differences in training history and ability to future performance predictions. These predictions can be used to quantify the costs, outcomes, and uncertainty associated with different training regimes. Additionally, these predictions can be used to prescribe individualized refresher training schedules.

In this talk, we will describe the theoretical development of PPE and our efforts to validate the model using datasets from experiments on human memory. We will then focus on recent applications of PPE to the field of virtual reality laparoscopic surgery and CPR. In addition to establishing the theoretical adequacy of PPE, these examples demonstrate the model's applied utility. The development of PPE has been supported by investments from government (Air Force Office of Scientific Research along with Air Force Research Laboratory) and commercial sources (Laerdal Medical along with the American Heart Association) during the past five years. The implications of this work extend to training in all DoD agencies, and to training in other non-military domains such as health care.

4 **Practical Operator Intelligent Tutor - Simple Key Loader: The Path Towards Effective and Affordable Intelligent Tutoring Systems**

Ms. Tashara Cooper, NAWCTSD

In recent years, the Navy has been exploring innovative ways of utilizing advanced technologies to deliver effective training solutions to the fleet. One such technology is intelligent tutoring systems. In 2013, an integrated product team at the Naval Air Warfare Center Training Systems Division undertook a proof-of-concept research effort to determine the plausibility of developing an intelligent tutoring training solution for the IT 'A' school populace. At our customer's request, we researched the feasibility of utilizing the Army Research Laboratory's Generalized Intelligent Framework for Tutoring (GIFT). The sponsor suggested the AN/PYQ-10 Simple Key Loader (a hand-held encryption device), that has widespread use across DoD agencies, as the target domain. For the novice watchstander, this device possesses complex use and common error characteristics that embody high consequences (e.g. loss of data due to improper shutdown). In this presentation, I will describe our pedagogical approach, the capabilities leveraged to develop the SKL training application, the specific strengths and challenges that GIFT brought to the effort, lessons learned and implications for future research.

SubTAG: Training

Chair: Jen Pagan Co-Chair: Kelly Hale

Friday, 8 May 2015
Session: II

0835 - 0855 **Simulated Environments for Cybersecurity Training**

Mr Eric Ortiz, IST, UCF; Dr Lauren Reinerman-Jones, ISt, UCF

In 2013, Eric Snowden was an employee of Booz Hamilton, a federal contractor for the National Security Agency (NSA). The NSA is a United States (U.S.) Government agency responsible for worldwide monitoring data and information for both counterintelligence and foreign intelligence needs. Snowden used his position as an infrastructure analyst to leak thousands of top-secret classified documents detailing the U.S. Government's global covert surveillance and eavesdropping undertakings to the public. Cybersecurity is a computer term regarding the detection, anticipation, and prevention of computer technologies and peripherals from damage, attack, or unauthorized access. These technologies include the monitoring of networks, programs, applications, and people. Cybersecurity can be viewed from both an offensive or defensive perspective involving maintaining and proactively assessing security vulnerabilities. The Snowden incident is considered one of the biggest leaks in NSA history and a prime of example of a cybersecurity breach on a national scale which threatened to expose U.S. secrets and intelligence. This incident identified the human threat as a contributing factor, and highlighted several weaknesses in the present state of U.S. cybersecurity affairs. In efforts to strengthen cyber defenses, a solid theoretical research foundation regarding cyber vulnerabilities is warranted. Building upon this foundation, training and experimentation can provide insight into current cybersecurity training methods and how they can be transitioned and implemented into future training regimens. The current training environments for cybersecurity vary in fidelity and sophistication. This includes traditional, self-paced methods involving primarily text-based informational pages followed by pages to test information retention. In addition, developers are beginning to utilize gaming features in cybersecurity training to better understand aspects of motivation and engagement.

0855 - 0925 **Tracer FIRE Cyberforensic Training Platform**

Mr. Benjamin Anderson, Sandia National Laboratories; Mr. Kevin Nauer, Sandia National Laboratories; Mr. Wellington Lee, Sandia National Laboratories; Mr. Jonathan McClain, Sandia National Laboratories; Dr. Robert Abbott, Sandia National Laboratories

Tracer FIRE, (Forensic and Incident Response Exercise), was originally created in 2009 by Sandia National Laboratories (SNL) and Los Alamos National Laboratories (LANL) to educate and train cyber security incident responders (CSIRs) using hands-on classroom training, followed by a competitive, team-oriented cyber security exercise. Under this program, several hundred CSIRs from the Department of Energy, other U.S. government agencies, and critical infrastructure organizations have been trained. In 2012, a strategic partnership was formed between the Tracer FIRE program and the human factors/cognitive research group at SNL to conduct research into the human factors related to cybersecurity. To support this new research focus, the Tracer FIRE environment was updated to support improved data collection of human performance measures during the exercise portion of the training. To ensure performance at Tracer FIRE would correlate with real-world performance, the technical challenges and overall exercise narrative are based on actual malware campaigns, such as Dragonfly. This paper describes the architecture of the Tracer FIRE environment for presenting technical challenges and supporting narrative information, which is based on large number of virtual machines controlled by a simulation engine; the tracking of team performance, including timing information and number of correct/incorrect answers; and a variety of activity metrics for each individual that can be tied to overall team activities.

0925 - 0945 **Adaptive Training for Combat Information Centers: Transitioning from Research Concepts to Technology Development**

Dr. James Pharmer, NAWC TSD; Dr. Laura Milham, NAWCTSD

Under the Capable Manpower Future Naval Capabilities program, the Office of Naval Research (ONR) sponsored the Adaptive Training for Combat Information Centers (ATCIC) project, which has focused on developing technologies to support instructors in observation, assessment, and providing feedback to individuals and teams during tactical scenario based training (SBT) exercises. This effort began with a four year Science and Technology (S&T) phase, in which the team researched, selected, and empirically demonstrated effective and efficient adaptive training technologies aimed at reducing workload on instructors executing SBT exercises. Per a technology transition agreement with the Littoral Combat Ship (LCS) program office (PMS 505), this effort has moved toward supporting the transition of the technologies into the acquisition of SBT tools supporting qualification and certification of LCS sailors and crews, respectively, at Surface Warfare Officers School and the LCS Training Facility. This presentation will provide an overview of the ATCIC program, the research and development of the adaptive training demonstration software, and the approach and challenges with transition of S&T tools into the acquisition of production training software. Lessons learned, future directions in adaptive training S&T, and extensibility to other domains will also be discussed.

0945 - 1005 **Competency-Based Training System to Advance Learning Performance**

Mr. Stu Rodgers, TiER1 Performance Solutions; Mr. Ryan Meyer, TiER1 Performance Solutions

Work environments with multiple operating platforms, massive amounts of data, continuous change, and disparate personnel create demands for expert proficiency that can outpace the ability to develop fully qualified personnel. The demand to accelerate knowledge and skill levels from novice to proficient is growing within both government and commercial training programs. To solve this problem, many disparate models have been created and learning experts have proposed how these models could be linked together to form lifelong competency-based adaptive learning solutions (Ostyn, 2005). To date, no one has successfully implemented the models into a complete and commercializable solution. Our research team concluded that a standardized competency-based adaptive learning approach is needed. Our approach incorporates an underlying competency framework that 1) determines a learner's proficiency as a function of evidence of performance, 2) selects and presents a training program from the full breadth of competency-aligned learning and assessment opportunities available to an individual, and 3) incorporates a system to dynamically generate and update an individual’s learning path as new assets are added and the leaner's proficiencies change. This presentation describes a multiple year research project to develop the methods, tools, and adoption strategy within the Air and Space Operations Center (AOC) Intelligence Community. Through SBIR funding from the Air Force Research Laboratory, our team developed a competency-based adaptive training system (CATS) by analyzing the AOC competency model and training and performance standards. For DoD training needs, CATS provides a comprehensive decision support tool to help training professionals align training opportunities to specific jobs. The design of CATS is intended to 1) facilitate a data-driven competency-centric learning lifecycle; 2) capture performance evidence and update proficiencies; 3) compare current proficiency levels with role expectations; and 4) support the development of personalized learning, assessment, remediation, curriculum design, and role-based training. This presentation also describes data collected during an operational CATS prototype pilot study demonstration at Davis-Monthan AFB AOC and through continuous interactions with subject matter experts.

1005 - 1025 **Cognitive Alignment with Performance Targeted Training Intervention Model: CAPTTIM**

Dr. Quinn Kennedy, Naval Postgraduate School; Dr. Ronald Fricker Jr., Naval Postgraduate School

We propose that the use of two simple behavioral measures, in conjunction with neurophysiological measures, can be used to create a training intervention that has the potential to provide: (1) real-time notification as to when a training intervention is needed, and (2) real-time information as to the type of training intervention that should be employed. The Cognitive Alignment with Performance Targeted Training Intervention Model (CAPTTIM) determines if a trainee's cognitive state is aligned or misaligned with actual performance. When misalignment occurs, it indicates that a training intervention is needed. Neurophysiological markers as captured by eyetracking and electroencephalography (EEG) can assist in determining why misalignment between cognitive state and performance occurred, leading to more effective and targeted training intervention. Because all measures are captured continuously in real time, this model has the potential to increase training efficiency and effectiveness in a variety of training domains. The model is illustrated with two case studies.

1025 - 1045 **Connecting System Capability, Environment, Warfighter, and Training Utilizing the Task Map Analysis Process**

Gail Nicholson, M.S., Naval Surface Warfare Center Crane Electro-Optics Displays

1045 - 1105 **Training Effectiveness Readiness; a Scale to Quantify the Confidence of Training Systems' Capabilities**

Roberto Champney, PhD, Design Interactive