ABSTRACT

Title of Document: DEVELOPMENT OF REQUIREMENTS TO INCORPORATE NEUROPHYSIOLOGICAL MEASURES IN HUMAN COMPUTER INTERFACE DESIGN

Colby Dean Raley, Master of Science in
Systems Engineering, 2005

Directed By: Mark Austin, Associate Professor, Systems Engineering

This project specifies requirements for testing platforms and facilities that will enable the use of neurophysiological data to help improve human computer-interfaces. The data used to generate these requirements was collected as part of an advanced human factors effort aimed at improving the usability of future releases of the Tactical Tomahawk Weapons Control System (TTWCS). Cognitive state was measured using electrocardiography (EKG), galvanic skin response (GSR), and electroencephalography (EEG), in addition to traditional measures using various subjective and psychological analyses.

This project demonstrated the value of neurophysiological measures into the Human Computer Interaction (HCI) design process, including increased objectivity of measures and consistency between measures. Simultaneous

neurophysiological and psychological measurements will enable researchers to better understand true usability of an interface and the requirements documented herein will enable such research.

DEVELOPMENT OF REQUIREMENTS TO INCORPORATE NEUROPHYSIOLOGICAL MEASURES IN HUMAN COMPUTER INTERFACE DESIGN

By

Colby Dean Raley

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:
Professor Mark Austin, Chair
Professor Kent Norman
Dr. James Baker

Dedication

Risk more than others think is safe,

Care more than others think is wise,

Dream more than others think is practical,

Expect more than others think is possible.

Acknowledgements

Many thanks to the engineering team at Lockheed Martin, especially Drs. Polly Tremoulet and Marianne Radding. I could not have completed this project without your help and support throughout the process.

Additional thanks to my team at DARPA for facilitating my project and allowing me the time to complete it in the face of many competing demands.

And finally, numerous thanks to my dear friends and colleagues for their assistance in developing and reviewing the final product throughout its many phases. *Peele*.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Objective

The objective of this effort is to specify system requirements to support a Human Computer Interaction (HCI) experiment that will evaluate the effectiveness of neurophysiological sensors in improving the usability of a human-computer interface. This system, once fully implemented, will enable researchers to use these tools to determine necessary interface improvements and to prove that quantitative measures of cognitive workload provide more useful information for improving the usability of an interface than subjective survey methods alone.

Description

The evaluation of human-computer interfaces is fraught with many inherent difficulties, including the determination of how much effort users must expend to perform tasks. Until now, the most effective method of measuring cognitive workload has been to either observe a user while they complete a task or to question them about their experiences (Nielsen & Mack, 1994). Users may be asked about their experience as they use a system (which can interrupt the process) or after completing a series of tasks or subtasks (which can make it difficult for them to remember what they have done). Additionally, users may feel pressure to indicate that a task was easier than it actually was (in order to seem more intelligent or in control) or they may rationalize their experience without realizing that they are doing so.

Neurophysiological methods will make this process more objective and consistent, will supplement existing measures with workload assessments, and will allow researchers to take several variables out of their experimental processes. Neurophysiological measures of workload are collected continuously by sensors that are already outfitted to a user and can therefore be obtained at exact intervals across many experiments with no interruption. Additionally, they cannot be skewed by participants or observers. Quantitative measures such as these are already in use in the research world, but they typically address such simple phenomena as isolated heart rate (HR) or galvanic skin response (GSR) measures, which give a gross level of excitement or arousal during the completion of a task (Noldus et. al., 1999). The more extensive incorporation of these measures could enable great advances in the effort of universal measures of utility of human-computer interfaces.

The pilot studies in this project simultaneously collected quantitative cognitive state measures and traditional psychological measures of the usability of an interface. The studies were a first attempt to include measures such as electroencephalography (EEG) and HR/GSR in the process to improve the interface for the next generation of Tactical Tomahawk Weapons Control System (TTWCS) software.

The pilot studies gave researchers significant insight into the required system improvements for useful incorporation of neurophysiological measures into the usability testing process. While traditional HCI experimentation is structured and rigorous, it utilizes an entirely different process and timescale than is required for neurophysiological measures.

These studies also gave a preview to the considerations for designing systems to support collection and analysis of objective measures of cognitive workload.

Chapter 2: Background

Human Computer Interaction

HCI is a field of study that aims to evaluate and improve the design of computational systems for their human users (Hewett, et. al., 2004). The field of HCI is a relatively young discipline that stems from human factors. Human factors, or ergonomics, aims to understand the interactions between humans and systems and uses various methods to optimize human well-being and system performance (IEA, 2000). HCI came to the forefront of research efforts as computer technology advanced and it became common for humans to interact with computers.

JCR Licklider (1960) laid the groundwork for this important work by calling for unprecedented cooperation between humans and machines. Two years later, Licklider became the director of the Information Processing Techniques Office at the Defense Advanced Research Projects Agency (DARPA), where his views on the ways that humans and computers should work together had a large impact on the development of advanced computing technology in Department of Defense (DoD) systems. The DoD was one of the first entities to incorporate computers into decision-making tasks, but the field of HCI has grown to include systems and applications in the commercial sector. Much of the field's pioneering work continues to be explored in the DoD domain and from there impacts commercial and personal products.

For example, in DARPA's Biocybernetics program, researchers explored the use of biologically-measurable signals, helped by real-time computer processing, to assist in the control of vehicles, weaponry, or other systems (Beatty, 1978). Performers on the program at the time were specifically investigating the use of EEG measures (specifically, the P-300 response) to measure workload (Isreal et. al., 1980). Significant work to further investigate the validity and utility of these measures continues, both in the DoD (Wilson, Lambert & Russell, 1999) and in academia (Gevins et. al., 1998).

HCI Design Process

Today, much HCI research focuses on User-Centered Design (UCD) principles, in which the user is the main focus throughout the design of a product or application (Righi, 2001). Figure 1 illustrates the process:

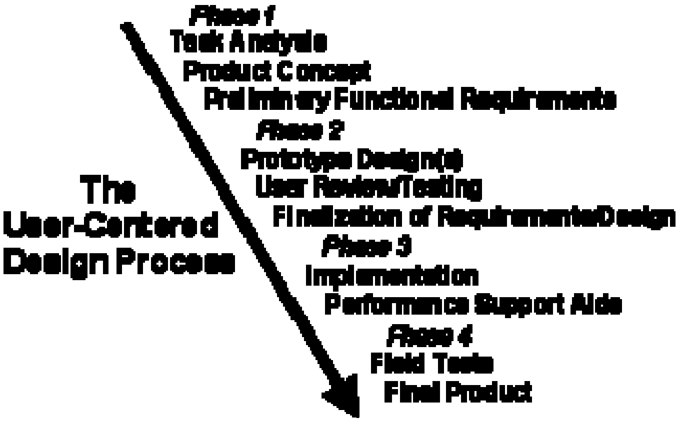


Figure 1: UCD process summary (UVA, 2005)

The TTWCS software examined in this project was in Phase 2 of the UCD process: User Review and Testing. Due to a renewed focus on usability by leadership developing the TTWCS, the entire design process relies on input from human factors personnel, and the user has been the primary focal point throughout software development. Following the review and testing, experimentation for Phases 3 and 4 of the UCD process has already been planned (TWCS SDA Team, 2005).

HCI Usability Testing

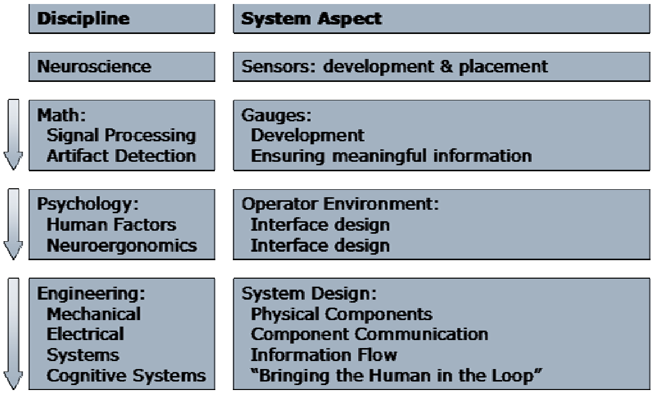
Traditional user workload measures (e.g. Situation Awareness Global Assessment Technique, or SAGAT and NASA's Task Load Index, or NASA TLX) are based upon observation of external signs of user performance and subjective reporting of a user's internal state after the performance of tasks (Endsley, 1998; Hart & Staveland, 1988). There is much debate over the best methods to use, in consideration of both the experimental subject and the researchers who look to elicit and analyze data (Norman & Panizzi, 2004). A tradeoff between the value of the data obtained by these methods and the time and effort to implement and analyze them must be completed. Often, the prototype development process is a fast-moving one, and suggestions for improvements must be made quickly or they have no chance of being implemented.

The addition of the neurophysiological measures will provide a measurement of the real-time cognitive state of the user while performing their task. This can provide more precise measurement and diagnostic information concerning defects in the interface and will improve the accuracy of measurements and increase the identification of defects and portions of the interface to improve.

Initial reporting from this project already shows promise for the incorporation of neurophysiological measures (Radding, Siegel & Russell, 2005).

Augmented Cognition

Augmented Cognition (AugCog) is an emerging field of research that is developing new technologies to noninvasively measure the cognitive state of humans and to use that state information to adapt closed-loop computational systems to humans' needs (Schmorrow & Kruse, 2004). AugCog systems are at the intersection of many different disciplines, including neuroscience, human factors, electrical engineering, mechanical engineering, and systems engineering, as illustrated in Figure 2.



Discipline	System Aspect
Neuroscience	Sensors: development & placement
Math: Signal Processing Artifact Detection	Gauges: Development Ensuring meaningful information
Psychology: Human Factors Neuroergonomics	Operator Environment: Interface design Interface design
Engineering: Mechanical Electrical Systems Cognitive Systems	System Design: Physical Components Component Communication Information Flow "Bringing the Human in the Loop"

Figure 2: Augmented Cognition enabling technology areas (Marhsall & Raley, 2004)

These technologies were initially investigated because of the continuing excess of information that people face every day, which is often rendered useless either because it is not essential to the user or because it is delivered at the wrong time (often too late). AugCog systems aim to dramatically increase the ratio of "good" information to "bad" information and enable end users to do better and more complete work faster (Raley et. al., 2004).

AugCog systems adapt the presentation of information and tasks to suit available cognitive resources using neurophysiological sensors and gauges. For example, if a user's spatial working memory is overloaded, critical incoming tasks may be presented verbally in order to maximize the capability of cognitive resources. Augmented Cognition goes a step beyond traditional HCI techniques to enable adaptation based upon not only the environment and tasks at hand, but also on the real-time assessment of operators' cognitive state (Raley et. al., 2004).

Sensors and Cognitive States

Augmented Cognitive researchers have experimented with and refined numerous sensing technologies, including:

- Direct Brain Measures
    - Electroencephalograpy (EEG) – electrical potentials in the brain
    - Functional Near Infrared imaging (fNIR) – oxygenated and deoxygenated hemoglobin in the blood
- Physiological Measures
    - Heart Rate (HR)
    - Electrocardiogram (EKG) – electrical potentials across the chest/heart
    - Pulse Oximetry – measures the oxygenation level of blood
    - Posture
    - Galvanic Skin Response (GSR) – moisture content in the skin

- Temperature
- Electrooculography (EOG) – electrical potentials in the muscles surrounding the eye
- Pupilometry – pupil diameter
- Gaze Tracking

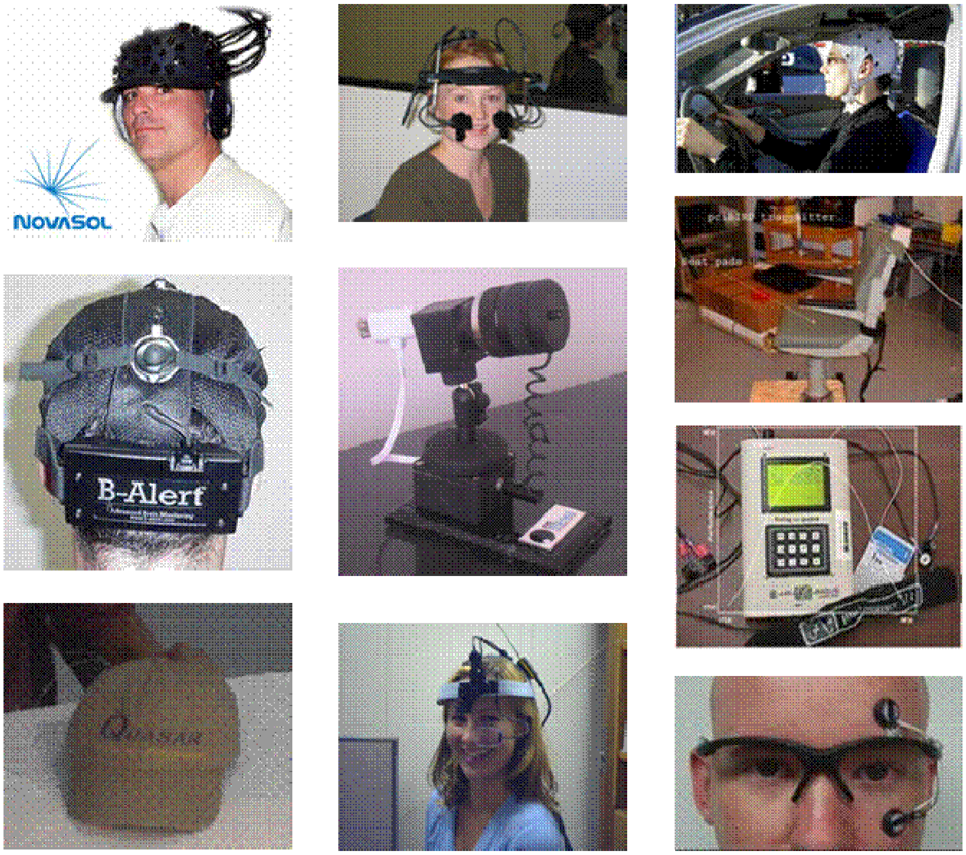


Figure 3: Various sensing technologies used in Augmented Cognition systems

Figure 3 highlights many of these sensing technologies. From the top left (going down by column): Archinoetics fNIR system, Advanced Brain Monitoring's EEG system, Quasar's non-contact electrode EEG prototype system, Eyetracking's eyetracker and pupilometry system, Eye Response Technology's ERICA eyetracker, a generic eyetracking system, DaimlerChrysler's EEG system, the University of Pittsburgh's posture sensing system, Clemson University's Arousal Meter system, and Quasar's current EOG system[1] (Kobus, Morrison, & Schmorrow, 2005).

Many of these sensing technologies have been integrated into cognitive state gauges, which give accurate classification of workload and other measures of cognitive state. Various research teams have investigated the feasibility of different sensors (by weighing their accuracy, operational relevance, and comfort/wearability), and the appropriate techniques by which to process sensor data into meaningful gauge information (St. John, Kobus, and Morrison, 2005). These gauges also take into account the neurological and physiological delays between a firing event in the body and the sensing of this event (Belyavin, 2005). These delays have been considered by many researchers in the field, and quantifiable values for the delays are known and able to be incorporated into AugCog systems (Gerson, Parra & Sajda, 2005).

The Lockheed Martin Advanced Technology Laboratories (LMATL) sensor system used in the pilot studies described here includes EEG, EKG, and GSR sensors (LMATL 2005). The EEG system gathered data from seven sensors and

---

[1] All of these companies have participated in DARPA's Improving Warfighter Information Intake Under Stress program. More information is available at www.augmentedcognition.org (last accessed 22 November 2005).

the physiological sensor suite gathered data from 3 EKG sensors and one GSR sensor pair.

The B-Alert brain-monitoring system, from Advanced Brain Monitoring (ABM), measures electrical potentials on the scalp and generates a record of the electrical activity of the brain. The sensors are easy to apply, utilizing only a small amount of conductive cream through the hair to make electrical contact. The system amplifies the EEG close to the sensors and monitors impedance online, ensuring that all users can obtain high-quality EEG data (Berka et. al., 2004).

The system also included the B-Alert software suite, which classifies the brain's electrical activity into validated measures of engagement, mental workload, and distraction/drowsiness, identifying and resolving artifacts in real time (Berka et. al., 2004). Using the software, the classifications are visually presented on a second-by-second basis and can be summarized across a recording session into different states of alertness, enabling researchers to easily interpret real-time EEG state measurements.

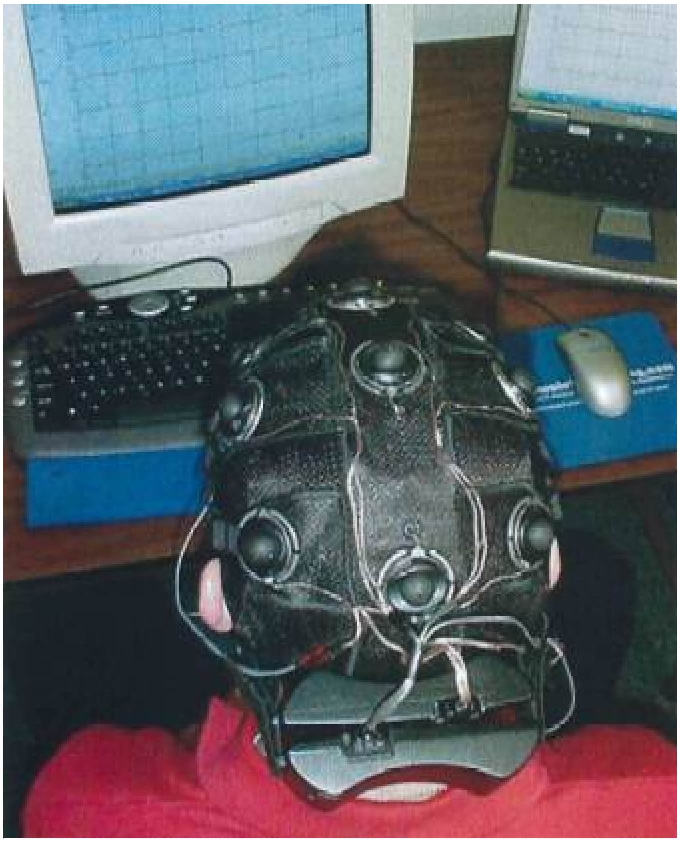


Figure 4: Sample configuration of ABM sensor suite

The physiological sensor system, comprised of Procomp hardware elements, included integrated EKG, which measures electrical potentials on the body surface and generates a record of the electrical currents associated with heart muscle activity and GSR, which measures change in the ability of the skin to conduct electricity.

The EKG and GSR sensors (shown in Figure 5) provide indirect measures of cognitive workload (Committee on Metabolic Monitoring for Military Field Applications, Standing Committee on Military Nutrition Research, 2004).

Figure 5: Sample Procomp System Components

The data from the B-Alert and physiological sensor system was interpreted using an LMATL-developed sensor data processing system, which buffers the measurements and outputs a meaningful range of measurements using a discriminant function analysis (LMATL, 2005).

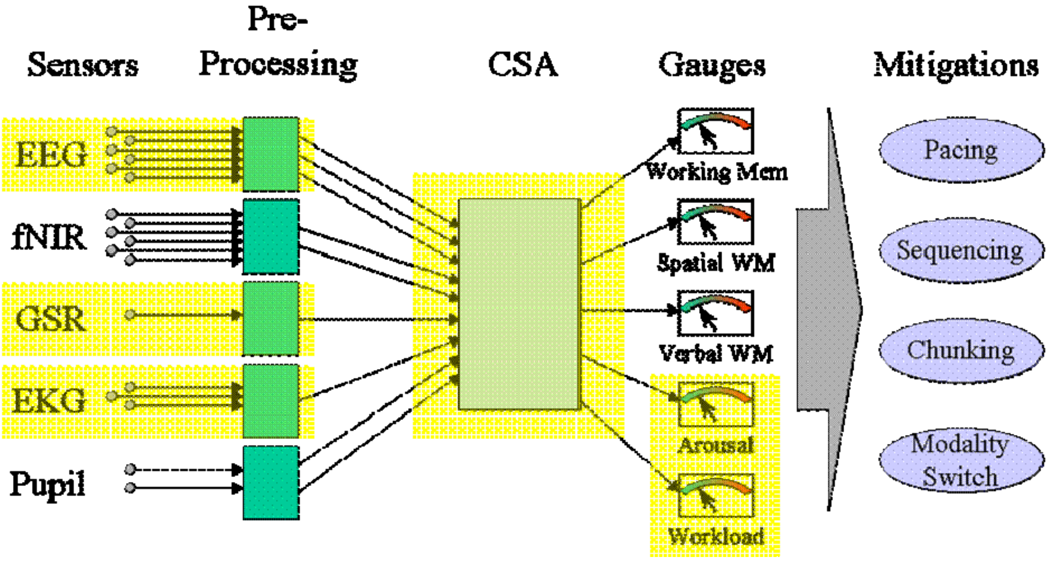


Figure 6: LMATL's Cognitive State Assessor Architecture

Figure 6 illustrates the Cognitive State Assessor (CSA) architecture used for a full Augmented Cognition system and the highlighted portions illustrate the parts of this system being used in the TTWCS HCI evaluation process. This data processing system allows researchers to get real-time measures of arousal and workload, rather than having to process the data offline to get these measures. The CSA uses a combination of sensor inputs and knowledge from an example set of sensor data (where users performed tasks where the workload was highly predicable) to understand cognitive workload. The physiological data from the arousal gauge[2] gives an indication of the stress that a user is experiencing due to a task and the neurophysiological data from the workload gauge gives a direct measure of cognitive workload as understood by electrical activity in the brain. These measures together give researchers a clear and quantitative understanding of workload.

System Architecture

A high-level system architecture for AugCog systems is illustrated in the following diagram (Marshall & Raley, 2004).

---

[2] It is important to consider arousal in the determination of cognitive workload and the prediction of performance. As Yerkes and Dodson determined in their seminal 1908 experiments, human performance peaks at an intermediate level of arousal – too low and a user is bored, too low and a user is too stressed to complete a task. Workload must complement these levels of arousal in order to achieve optimal performance.

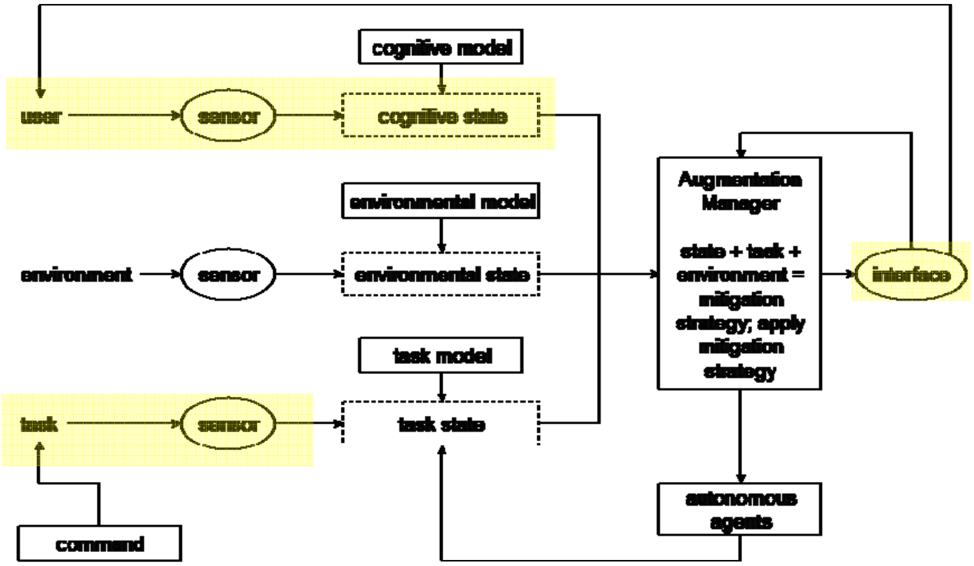


Figure 7: System level Augmented Cognition architecture

The arrows in Figure 7 indicate flows of information. The general flow is as follows: commands influence a task; sensors detect activity in the user, environment, and task; sensed and modeled information is combined to create real-time models of the user, environment, and task; state information and interface information is used by the Augmentation Manager to determine an appropriate information-bottleneck mitigation strategy; the Augmentation Manager impacts the interface, which communicates with the user and autonomous agents if necessary, which complete tasks that the user is too overloaded to complete. The highlighted portions of Figure 7 denote the system elements examined in this experiment.

Mitigation Strategies

In an AugCog system, mitigation strategies are used to increase overall performance while maintaining situational awareness and enabling a user to take on even more difficult tasks if a situation requires (Schmorrow & Kruse, 2004).

These mitigation strategies include methods such as pacing (or scheduling of tasks to ensure that they are completed in the most effective and efficient order); intelligent switching (or the maximization of information intake channels by presenting information spatially or verbally depending on cognitive state); and modality switching (or presenting information in different modalities if one is already full – i.e. presenting spatial information in a tactile way because the visual channel is already overloaded) (LMATL, 2005).

Mitigation strategies, though essential elements of closed-look Augmented Cognition systems, were not used in the experiments described here. This is because the goal of these experiments was not to close the loop or improve performance, but rather to demonstrate the use of AugCog sensing technologies to provide valuable information about the cognitive state of a user and thereby ascertain elements of the interface that should be improved.

Application Domains

AugCog systems aim to improve performance by enabling a cognitive closed feedback loop between the operator and an adaptive computer-based system. They anticipate human task loading and change information modalities (providing information through a different medium such as aural, spatial, or verbal) or offload tasks. They also augment users by assisting in task execution through pre-negotiated “crew coordination” either by task sharing between the human and machine or completely offloading certain tasks to machine automation (Kincses, 2005).

Much research in Augmented Cognition has been sponsored by DARPA’s Improving Warfighter Information Intake Under Stress program. Currently in Phase 3, the program is assessing automation techniques for maximizing warfighter performance and efficiency under stress, with an emphasis on prototype experimentation in conjunction with battle experiments. Increasingly complex scenarios with tactical and strategic decision-making tasks are being employed to stress the operators and systems under evaluation. These developments will pave the way for operational use and deployment of AugCog systems in the coming years (Schmorrow & Kruse, 2004).

AugCog technologies are also being used in other applications and domains. For example, functional near infrared (fNIR) imaging techniques are being used to provide a non-invasive brain-computer interface to patients with Amyotrophic Lateral Sclerosis (ALS). FNIR technologies allow them to communicate with the outside world when they would normally be living without communication with the outside world, in physically deteriorating bodies that house fully-functional minds (Rapoport et. al., 2005).

Parasuraman (2003) describes the need to consider brain function when doing work in human factors and HCI. In this vein, AugCog technologies are also being used in the improvement of human-computer interfaces. These technologies provide direct rather than inferred measures of cognitive workload, which allow researchers to conclusively and objectively determine which portions of an HCI are appropriate and/or need further improvement.

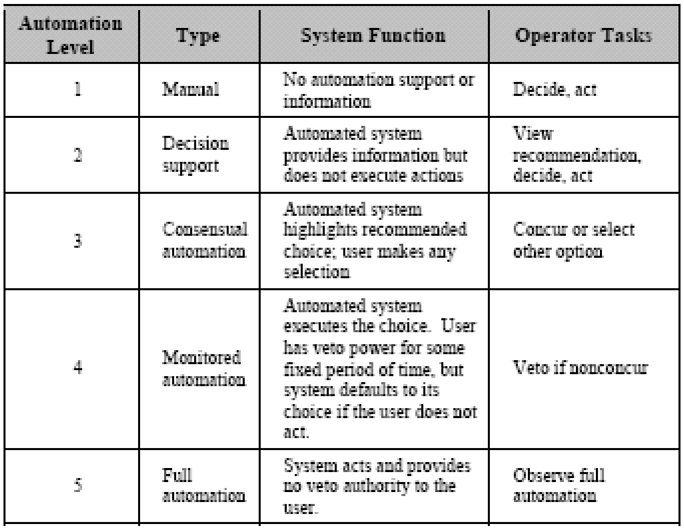
One platform for the Phase 3 demonstration of the Augmented Cognition technology is the Navy's Tactical Tomahawk Cruise Weapons System, a command and control suite that enables warfighters to launch and monitor reprogrammable missiles. AugCog researchers are working with the developers of this system to use quantitative measures of cognitive state to improve the HCI in this system (TWCS SDA Team, 2005).

Tactical Tomahawk Weapons Control System

Due to its impressive range and precision, the Tomahawk cruise missile is an extremely important part of the United States' military strategy (NAVLIB, 1993). There have been significant efforts throughout the Navy to improve further the Tomahawk technology with each system upgrade. The most recent upgrade, the Block-IV missile (or Tactical Tomahawk) is retargetable within a mission. This new capability, while dramatically increasing the potential efficacy of cruise missiles, also has the potential to drastically increase users' workloads. Now, their task will be extended from the original 'fire and forget' model to include monitoring of the mission as it progresses as well as the incorporation of incoming data and the prioritization of targets not just at an instant in time but throughout an entire mission (Willis, 2001).

The TTWCS is an up-to-date, expandable, reliable Weapons Control System (WCS) that enables operators to control Tactical Tomahawk missiles. It is being developed and delivered incrementally, using a tight spiral development plan, in an effort to minimize development time and get the best systems to the users in the shortest amount of time. Recently, there has been a growing focus on User UCD in this process, as older versions of the WCS required users to keep too many items in their working memory as they completed tasks. The newer versions of the interface strive to guide users through their tasks, rather than making it painful and laborious to complete their goals (Allen, 2004).

Due to the additional requirements levied by the new system capabilities, there is an increased need for automation in the missile launch and monitor process. As automation is introduced to improve speed and performance, care must be taken to maintain situational awareness, which is defined by Endsley (1998) as 'perception of the elements of the environment … the comprehension of their meaning, and the projection of their status in the near future.' Endsley's five levels of automation are highlighted in Figure 8 below.



Automation Level	Type	System Function	Operator Tasks
1	Manual	No automation support or information	Decide, act
2	Decision support	Automated system provides information but does not execute actions	View recommendation, decide, act
3	Consensual automation	Automated system highlights recommended choice; user makes any selection	Concur or select other option
4	Monitored automation	Automated system executes the choice. User has veto power for some fixed period of time, but system defaults to its choice if the user does not act.	Veto if nonconcur
5	Full automation	System acts and provides no veto authority to the user.	Observe full automation

Figure 8: Levels of automation (Endsley & Kiris, 1995)

Measures of situational awareness and performance for this experiment (TWCS SDA Team, 2005) include:

- Subjective cognitive workload as measured by the NASA-TLX
- Situational awareness as measured by probes being answered correctly and without experimenter assistance
- Task time as a measure of the amount of time taken to launch missiles
- Number of missiles launched on time
- Number of errors committed by the operator while conducting tasks associated with the scenario
- User satisfaction as measured by a questionnaire
- Workload as measured by neurophysiological sensors
- Attention requirements as measured by the neurophysiological sensors

This plan includes typical measures of situational awareness such as overall task performance, the ability to multitask successfully, and critical thinking (Campbell, Pharmer & Hildebrand, 2000). All of these measures have historically been taken through traditional means, which frequently interrupt users as they attempt to complete a task. The TTWCS design team has the specific goal for the HCI team to "develop TTWCS display concepts with a goal of minimizing mission execution time while maximizing operator situational awareness" (Willis, 2001). An appropriate interface can facilitate these goals simultaneously.

HCI Improvements for TTWCS

To facilitate this UCD, evaluations that measure user performance are continually being performed on the TTWCS prototypes under development. These evaluations occur at each stage of development to identify and propose improvements to alleviate defects that critically affect user performance.

The design of the updated interface begins with cognitive task analyses of the domain and the users (Cushing et. al., 2004), and includes the development of scenarios before work is started on actual interface. Following scenario development, the design of the interface and initial prototype development begins, and eventually, testing and analysis on subjects is initiated (Guerlain & Willis, 2001). This process is tightly coupled with the development of the system, and is therefore held to a tight spiral development plan as well, which enables interface designers to elicit and incorporate subject feedback early and frequently in the design process.

Recently, the HCI design team for TTWCS has initiated the inclusion of neurophysiological sensing data into their testing protocol, with the goal of taking more precise measures of a user's internal state as they complete a task. These neurophysiological measures, derived from research in Augmented Cognition, require additional system considerations to fully incorporate the neurophysiological measures with the existing measures. These considerations and requirements are addressed in Chapter 5: Requirements Generation.

Requirements Generation

Systems engineering is a rather loosely defined discipline concerned with ensuring the successful implementation of a desired system in any of various fields (INCOSE, 2003). The NASA Systems Engineering Handbook (Shishko & Chamberlain, 1995) includes system goal identification and quantification; design concept creation; design trading, selection and implementation; design verification, and post-implementation assessment as part of the process. The generation of requirements fulfills the need to quantify system goals – without formal requirements, large projects with disparate team members would not be able to effectively communicate or understand all of the design aspects necessary to create a successful system.

Many approaches to the specification of requirements have been identified in the literature (Sommerville & Sawyer, 1997; Krasner, 1985; Leite & Freeman, 1991; Davis, 1993; Jirotka & Goguen, 1994; Potts, Takahashi & Anton, 1994). All of these specification guidelines attempt to lay out the specific components of the requirements process, and can be summarized as a group to include:

- Elicitation – the determination of requirements by the systems engineer, through methods such as informal interviews, brainstorming sessions, and formal questionnaires
- Analysis – the conversion of the client's 'wish list' into attainable system attributes
- Specification – the documentation of analyzed requirements into a quantitative format that can be designed and measured against
- Verification & validation – the assurance that requirements are correctly specified (verification) and that a system that meets the listed requirements will also meet the client's needs (validation)
- Management – the continued maintenance of the requirements as the system matures and potentially changes (Lobo, 2004)

This project is primarily concerned with the analysis portion of requirements generation. The pilot studies for TTWCS HCI improvement provided an excellent opportunity for requirement elicitation, and the specification, verification & validation, and management tasks of the process are best left to the system designers. However, the incorporation of neurophysiological measures into the HCI improvements efforts left a gap in the process: the definition of system qualities necessary to incorporate this analysis into the HCI design process.

Requirements Generation Model

The Requirements Generation Model (RGM) proposed by Lobo (2004) provides systems engineers with a powerful tool to identify the phases of the requirements process, to create detailed explanations of all activities, and to facilitate future decomposition of requirements. Though the model does not fully refine the requirements or specify the methods required to complete activities, it does provide a complete framework for the analysis portion of the requirements generation process.

Attribute	**Description**
Activity Name	Name of the activity
Objective	Goal/aim of the activity
Action Points	Milestones to be achieved by the activity
Pre-condition	Conditions to be satisfied before activity commencement
Doer	Person conducting the activity
Participants	Participants in the activity
Input documents	Documents needed for the activity to begin
Output documents	Documents produced at the completion of the activity

Table 1: Requirements Generation Model (Lobo, 2004)

The activity model shown in Table 1: Requirements Generation Model (Lobo, 2004) is an ideal basis for specifying the requirements of a system that will enable the comparison of direct and traditional psychological measures of cognitive activity. An activity-based model allows us to incorporate lessons learned from the pilot studies, which are an invaluable tool at this stage of the project.

Other aspects of the RGM are appropriate for further refining the requirements developed herein, and can also expand to other portions of the requirements generation process.

Chapter 3: Experimental Plan for Pilot Studies

Method

Introduction

The objective of Lockheed Martin's HCI experiment was two-fold. The first was to evaluate new versions of the TTWCS prototype system to determine the usability of the HCI as compared to previous versions using operational users. The second was to incorporate neurophysiological measures into the evaluation process.

Training on neurophysiological sensing technologies for testing personnel who administered the experiments took place 6-7 June 2005 at Lockheed Martin Advanced Technology Labs in Cherry Hill, New Jersey. Following that training, a series of pilot studies was conducted at the testing lab at Lockheed Martin Integrated Systems and Solutions in Valley Forge, PA on 15 – 30 June 2005. This chapter outlines the events of those pilot studies. Several follow-on pilot and actual experiments in similar areas have been completed and more are being planned (Radding, Siegel & Russell, 2005).

In order to evaluate the TTWCS interface using both traditional and neurophysiological measures, subjects were asked to complete protocols using the TTWCS version 5.3 software. They were observed and asked questions to obtain a traditional psychological rating on the interface and were outfitted with non-invasive brain-sensing technology to obtain a quantitative rating. The basic order of events was:

- Consent and Background Questionnaire – 15 minutes
- Sensor Calibration – 1 hour
- Training – 2 hours
    - Concepts - 30 minutes
    - Practice - 90 minutes
- Task – 4 hours
    - Pre task scenario - 15 minutes
    - Tasking received to launch – 3 hours
    - Post Launch Execution - 30 minutes
    - Post task questionnaire - 15 minutes
- Debrief – 15 minutes
    - User Satisfaction Questionnaire - 5 minutes
    - Interview - 10 minutes

Consent and Background Questionnaire

During the consent and background questionnaire, subjects were asked to give some background information about their skills and experiences in areas related to the task at hand, and were briefed on their rights under the informed consent. This portion of the experiment was necessary to satisfy the Human Use research Institutional Review Board approval of the project.

The subjects selected for this pilot study were Lockheed Martin employees, some of whom had experience with WCSs. The demographics for this study were based more on availability than any specific characteristics, due to the pilot nature of the work. However, very specific subject selection guidelines (especially regarding experience) will be followed in the follow-on experimentation (TWCS SDA Plan, 2005).

Sensor Calibration

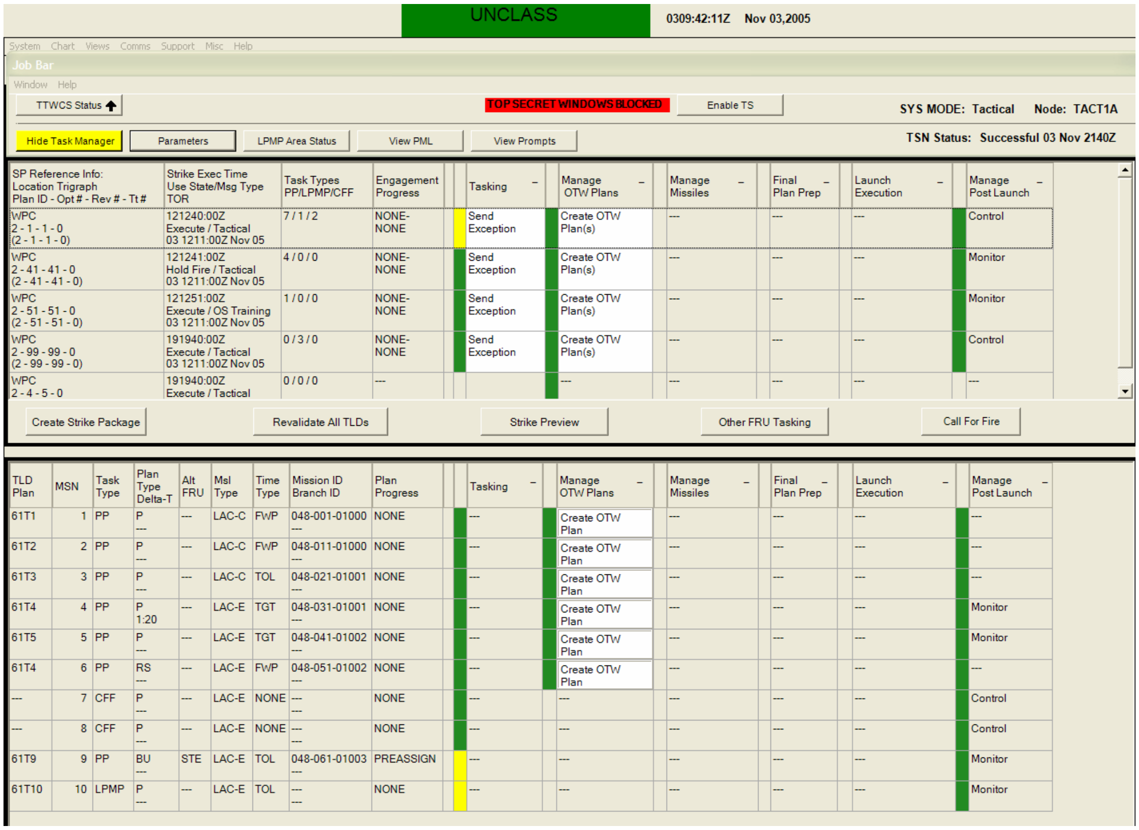
After the questionnaires, subjects were outfitted with the equipment necessary to obtain direct measures of cognitive state. The equipment used included an electroencephalography (EEG) system to measure electrical activity in the brain and a system that includes electrocardiography (EKG) and galvanic skin response (GSR) sensors. These devices, united in a sensor suite, provided the basis for a quantitative measurement of cognitive workload (LMATL, 2005).

The sensors were placed on the subjects by trained experimental personnel, according to manufacturer's instructions in order to get the best measurements possible – improper placement of electrodes can result in missing or unusable data. In order to allow the subjects unrestricted access to the TTWCS prototype, which requires use of a mouse and keyboard, the GSR sensors were placed on the toes, as is one of two typical placements (fingers or toes) defined by the manufacturers.

After placement, the sensors were calibrated to the subjects via simple resting and task periods, to get a baseline measurement of their resting and elevated workload and arousal levels. Calibration procedures have been minimized over time, to balance the value of specific settings for users against the time and effort required to calibrate the systems for individuals (LMATL, 2005; Belyavin, 2005).

Training

After the sensors were placed on the participants, they were given a short break, followed by a training session with engineers that were knowledgeable about the TTWCS version 5.3 software. They stepped through a launch scenario (highlighted in the next section) with the engineer standing by to answer questions or provide guidance when needed. This scenario was created by system engineers to thoroughly examine all aspects of the latest interface through user interaction, with a specific focus on the most changed and improved portions. The training time also provided the users with an opportunity to familiarize themselves with the interface and allowed researchers to monitor the physiological measures to ensure good data collection.



0309:42:11Z Nov 03,2005

SYS MODE: Tactical Node: TACT1A

TSN Status: Successful 03 Nov 2140Z

SP Reference Info: Location Trigraph Plan ID - Opt # - Rev # - Tt #	Strike Exec Time Use State/Msg Type TOR	Task Types PP/LPMP/CFF	Engagement Progress		Tasking		Manage OTW Plans	Manage Missiles	Final Plan Prep	Launch Execution		Manage Post Launch
WPC 2 - 1 - 1 - 0 (2 - 1 - 1 - 0)	121240:00Z Execute / Tactical 03 1211:00Z Nov 05	7 / 1 / 2	NONE-NONE	🟨	Send Exception	🟩	Create OTW Plan(s)	---	---	---	🟩	Control
WPC 2 - 41 - 41 - 0 (2 - 41 - 41 - 0)	121241:00Z Hold Fire / Tactical 03 1211:00Z Nov 05	4 / 0 / 0	NONE-NONE	🟩	Send Exception	🟩	Create OTW Plan(s)	---	---	---	🟩	Monitor
WPC 2 - 51 - 51 - 0 (2 - 51 - 51 - 0)	121251:00Z Execute / OS Training 03 1211:00Z Nov 05	1 / 0 / 0	NONE-NONE	🟩	Send Exception	🟩	Create OTW Plan(s)	---	---	---	🟩	Monitor
WPC 2 - 99 - 99 - 0 (2 - 99 - 99 - 0)	191940:00Z Execute / Tactical 03 1211:00Z Nov 05	0 / 3 / 0	NONE-NONE	🟩	Send Exception	🟩	Create OTW Plan(s)	---	---	---	🟩	Control
WPC 2 - 4 - 5 - 0	191940:00Z Execute / Tactical	0 / 0 / 0	---			🟩	---	---	---	---		---

TLD Plan	MSN	Task Type	Plan Type Delta-T	Alt FRU	Msl Type	Time Type	Mission ID Branch ID	Plan Progress		Tasking		Manage OTW Plans	Manage Missiles	Final Plan Prep	Launch Execution		Manage Post Launch
61T1	1	PP	P ---	---	LAC-C	FWP	048-001-01000 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	---
61T2	2	PP	P ---	---	LAC-C	FWP	048-011-01000 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	---
61T3	3	PP	P ---	---	LAC-C	TOL	048-021-01001 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	---
61T4	4	PP	P 1:20	---	LAC-E	TGT	048-031-01001 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	Monitor
61T5	5	PP	P ---	---	LAC-E	TGT	048-041-01002 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	Monitor
61T4	6	PP	RS ---	---	LAC-E	FWP	048-051-01002 ---	NONE	🟩	---	🟩	Create OTW Plan	---	---	---	🟩	---
---	7	CFF	P ---	---	LAC-E	NONE	--- ---	NONE	🟩	---		---	---	---	---	🟩	Control
---	8	CFF	P ---	---	LAC-E	NONE	--- ---	NONE	🟩	---		---	---	---	---	🟩	Control
61T9	9	PP	BU ---	STE	LAC-E	TOL	048-061-01003 ---	PREASSIGN	🟨	---		---	---	---	---	🟩	Monitor
61T10	10	LPMP	P ---	---	LAC-E	TOL	--- ---	NONE	🟨	---		---	---	---	---	🟩	Monitor

Figure 9: TTWCS v5.3 screenshot.

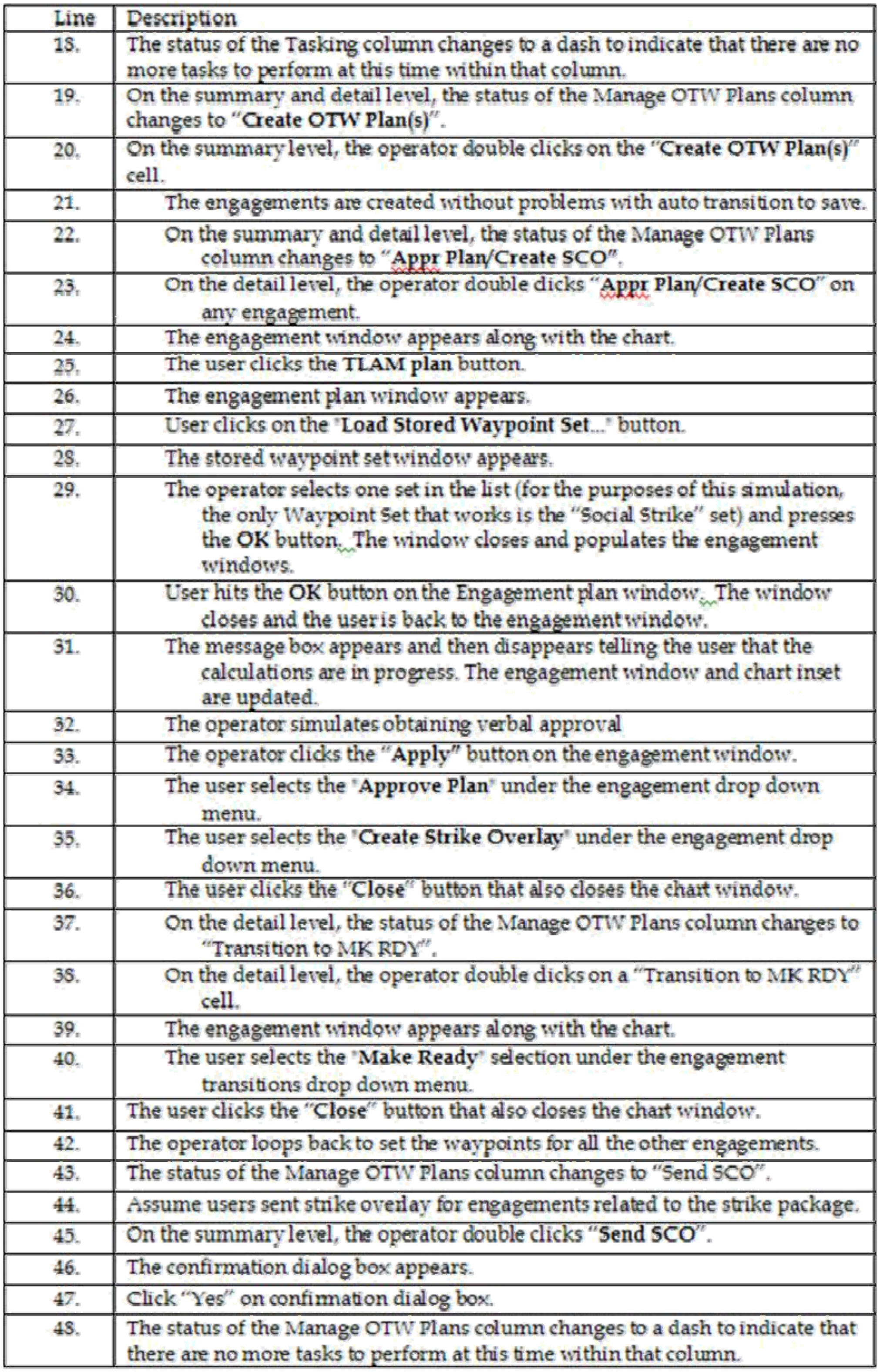
Task

After training, the subjects were given another short break and the session commenced. They followed a pre-defined scenario to achieve the launch goals presented to them, with the intermittent help of an engineer when required. The scenario guided users through the start up of the software, the selection of targets, and the initiation of missile launches (TWCS SDA Team, 2004).

Subjects performed their tasks on a Pentium-class computer with a dual-monitor display. The monitors were mounted one on top of another and the user interfaced with the system using a standard keyboard and mouse system similar to that used in existing Tomahawk systems (as depicted in Figure 10). The monitor configuration was selected to mirror the display arrangement in an operational environment.

Figure 10: TTWCS Interface Setup (L3 Communications, 2005).

The scenario was a pre-specified set of tasks, used for evaluating the interface as it was developed. Lines 18-48 of the scenario (displayed in Table 2) were used to evaluate the utility of the neurophysiological data. These lines were selected for analysis to allow time for a full review of the task completion by the subject during the debrief. The specific tasks were selected following careful review by a Subject Matter Expert (SME) because they are known to elicit a highly variable, but predictable workload.



Line	Description
18.	The status of the Tasking column changes to a dash to indicate that there are no more tasks to perform at this time within that column.
19.	On the summary and detail level, the status of the Manage OTW Plans column changes to "**Create OTW Plan(s)**".
20.	On the summary level, the operator double clicks on the "**Create OTW Plan(s)**" cell.
21.	The engagements are created without problems with auto transition to save.
22.	On the summary and detail level, the status of the Manage OTW Plans column changes to "**Appr Plan/Create SCO**".
23.	On the detail level, the operator double clicks "**Appr Plan/Create SCO**" on any engagement.
24.	The engagement window appears along with the chart.
25.	The user clicks the **TLAM plan** button.
26.	The engagement plan window appears.
27.	User clicks on the "**Load Stored Waypoint Set…**" button.
28.	The stored waypoint set window appears.
29.	The operator selects one set in the list (for the purposes of this simulation, the only Waypoint Set that works is the "Social Strike" set) and presses the **OK** button. The window closes and populates the engagement windows.
30.	User hits the **OK** button on the Engagement plan window. The window closes and the user is back to the engagement window.
31.	The message box appears and then disappears telling the user that the calculations are in progress. The engagement window and chart inset are updated.
32.	The operator simulates obtaining verbal approval
33.	The operator clicks the "**Apply**" button on the engagement window.
34.	The user selects the "**Approve Plan**" under the engagement drop down menu.
35.	The user selects the "**Create Strike Overlay**" under the engagement drop down menu.
36.	The user clicks the "**Close**" button that also closes the chart window.
37.	On the detail level, the status of the Manage OTW Plans column changes to "Transition to MK RDY".
38.	On the detail level, the operator double clicks on a "Transition to MK RDY" cell.
39.	The engagement window appears along with the chart.
40.	The user selects the "**Make Ready**" selection under the engagement transitions drop down menu.
41.	The user clicks the "**Close**" button that also closes the chart window.
42.	The operator loops back to set the waypoints for all the other engagements.
43.	The status of the Manage OTW Plans column changes to "Send SCO".
44.	Assume users sent strike overlay for engagements related to the strike package.
45.	On the summary level, the operator double clicks "**Send SCO**".
46.	The confirmation dialog box appears.
47.	Click "Yes" on confirmation dialog box.
48.	The status of the Manage OTW Plans column changes to a dash to indicate that there are no more tasks to perform at this time within that column.

Table 2: Evaluated Tasks in TTWCS v5.3 (TWCS SDA Team, 2004).

The experiments investigating the usefulness of the neurophysiological measures of workload included four independent measures of cognitive workload for each subject:

1) Expert review of interface and tasks predicting times of high and low workload
2) Naïve observation of subject noting times that subject appears to be highly taxed or very relaxed
3) Cognitive gauge values
4) Subject coding of video collected during testing

Measures 1-3 were taken before and during the task completion and measure 4 was taken during the debriefing portion of the experiment.

The expert review of the interface and tasks was completed before the experiments took place. The SME for this experimentation is working with Lockheed Martin Integrated Systems & Solutions and the Navy's Program Management Authority code 282 to develop the next generation of human-computer interfaces for TTWCS. Version 5.3 was tested in this protocol the supporting SME is the primary human factors scientist developing this system, therefore she is particularly well-suited to predict parts of the task that should cause a high or low workload. Ideally, this prediction would be made by a team of SMEs who are well-versed in the TTWCS software, but for the purposes of this pilot study, the specified SME's experience with the system, through all phases of its development, provided the necessary expertise.

Specific tasks of the scenario were rated on a scale from 1-7, with one being extremely low expected workload and seven being extremely high. Four is a neutral score.

The naïve observation of the tasks was completed during the task through simple observation. Several naïve observers, who are well-versed in the observation of subjects to determine workload but not familiar with the TTWCS software, watched subjects as they completed the scenario and rated their apparent workload, by time, on a scale from 1-7. It is important that the naïve observer be truly naïve, with respect to the software, so that they can provide an unbiased opinion of how hard a subject is working at any given point in time. It is also important that they remain naïve throughout the testing process, so that they do not develop biases regarding which portions of the task are more difficult than others.

The quantitative cognitive state gauge values were collected automatically throughout the experimentation: the values are a real-time representation of cognitive workload as calculated using data from the EKG, GSR, and EEG. This process is described in the Background section.

Debrief

After the tasks were completed, subjects were given a short break and then returned to answer questions about their perceived workload during the scenario. During the task completion, subjects were recorded from behind (with no identifying features on the screen) from two angles to facilitate the review of both their actions and the on-screen changes. During the debrief, 20-30 minutes of video was queued up for relevant interaction and within those 20-30 minutes, questions regarding workload (on a Likert scale) were asked regarding specific epochs:

- Epochs of expert-identified hi/low workload
- Epochs of naïve-observer identified hi/low workload
- Epochs of self-identified hi/low workload
- Several pseudo-randomly selected epochs where no source identified hi/low workload (control)

The goal of this design was to compare each of these qualitative measures with the self-ratings. The self-ratings could already be compared to the qualitative (neurophysiological) measures via timing.

Results & Analysis

To specifically address the comparison of neurophysiological and traditional psychological measures of cognitive workload, subjects were monitored as they completed lines 18-40 of the TTWCS v5.3 launch scenario tasks (these lines are listed in the methods section). Since tasks 1-17 were very short, the subjects were monitored for the first 13-20 minutes of the task, or until they completed through line 48.

Each subject's task completion was monitored in four ways that are analyzed here:

1. Predictions of workload using a subject matter expert (SME)
2. Neurophysiological measures of workload
3. Measures of workload using a naïve observer
4. Measures of workload using a subject's self-evaluation

Neurophysiological Measures

Neurophysiological measures included arousal, or how engaged a subject is, and workload, or how much effort a subject is putting forth. The B-Alert system quantified EEG measurements in real time, identifying and decontaminating the signal of artifacts such as eye blinks, electromyography, amplifier saturation, and movement (Berka, 2004). As discussed in the background, these EEG measurements are processed, in conjunction with the physiological sensor suite to produce measures of arousal and workload.

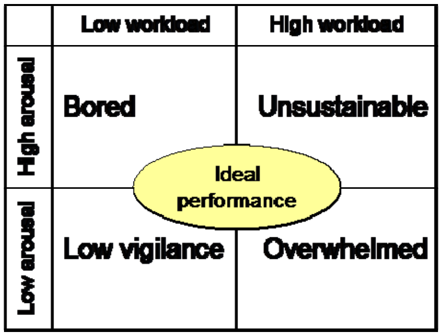


Figure 11: Ideal performance is at the intersection of workload and arousal.

Figure 11 illustrates the impacts of arousal and workload on performance – both aspects of cognitive state affect the ability of users to complete tasks.

Traditional Psychological Measures

All indirect measures of cognitive state were recorded on a Likert Scale from 1-7, with a score of one being extremely low workload, seven being extremely high workload, and four being neutral workload.

The project SME classified each task according to difficulty, where a more difficult task was understood to cause a higher workload and increased arousal. The SME had extensive experience with the protocol, as well as with initial user experiences with the protocol, and was therefore able to accurately predict difficulty. The predictions were done by task rather than time, so while they looked extremely consistent in the by-task listing, they were spaced out unevenly over time.

A naïve observer was asked to rate the subject on her perceived workload as she completed the prescribed tasks. The observer made a note of the current time and perceived workload whenever they noticed something significant in the subject's actions. This enabled the observations to be time-stamped easily, but they were infrequent and inconsistent (with respect to time) throughout the session. Additionally, since the ratings were completed by hand, they were not precisely mapped to times, as the observer was only able to record their thoughts so frequently and with a limited precision.

The self-evaluation was conducted using video screening during the debriefing portion of the experiment. After giving the subjects a brief opportunity to rest following the task, they were asked to review the task they had just completed, with the help of video data if needed, and were asked to rate their level of workload while completing different aspects of the task. This method was totally task-based, so the measures were not very tightly coupled with time.

In order to mitigate the time-scale difficulties, all data was reduced to a (1 measure) / (10 seconds) scale. Additionally, all measurements were normalized to enable a comparison of the neurophysiological and psychological scores:

$$normalized\_value = \left(\frac{current\_value - average\_value}{average\_value} \right)$$

Measurement Method Comparison

This study illustrates the need for a consistent and time-synced approach to collecting neurophysiological and traditional psychological data, which will be outlined further in the requirements section.

Chapter 4: Discussion

Preliminary Findings

Comparison by Subject

The following graphs highlight the initial results of comparing neurophysiological and traditional psychological measures. There are five major ratings shown:

- Normalized Arousal – values from arousal gauge
- Normalized Workload – values from workload gauge
- Normalized User Analysis Difficulty Rating – values from the user's self-evaluation during the debrief
- Normalized Subjective Analysis Difficulty Rating – values from the naïve observation of the user during task completion
- Normalized Expert Analysis Difficulty Rating – values from the expert's difficulty prediction of specific tasks

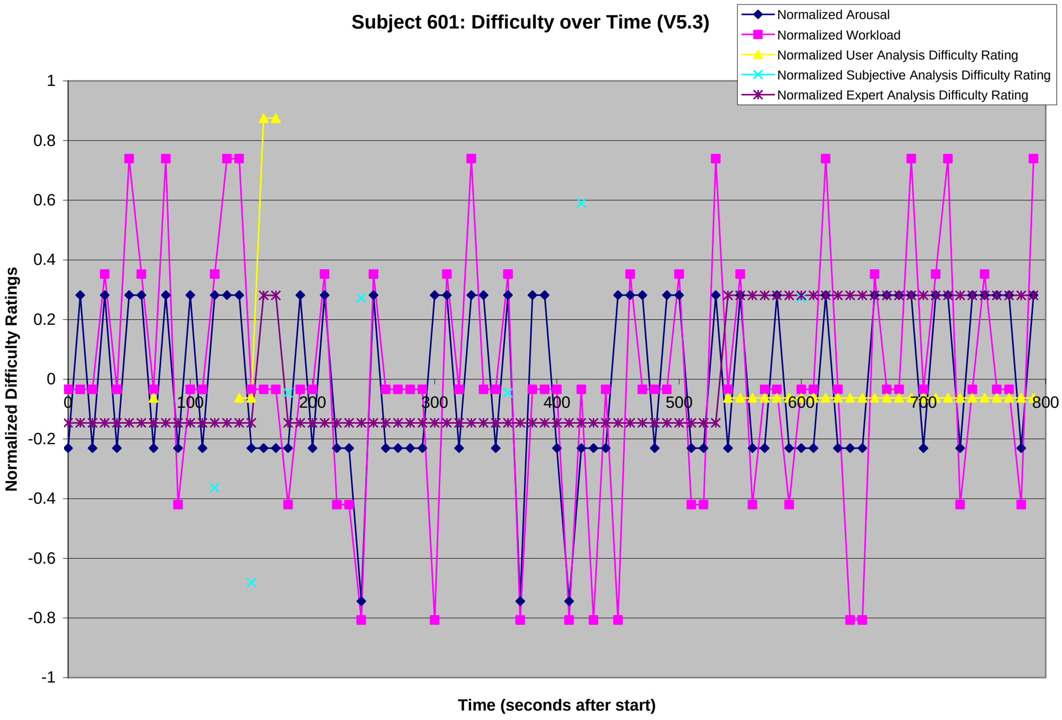


Figure 12: Comparison of effort level measurements (Subject 601)

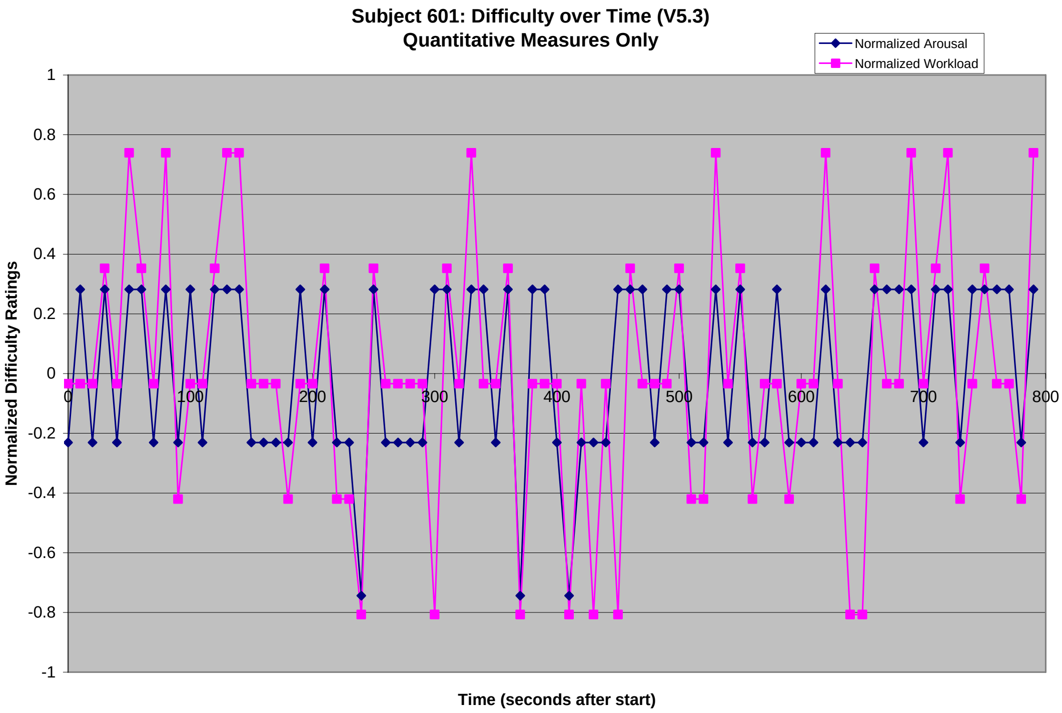


Figure 13: Neurophysiological measurements (Subject 601)

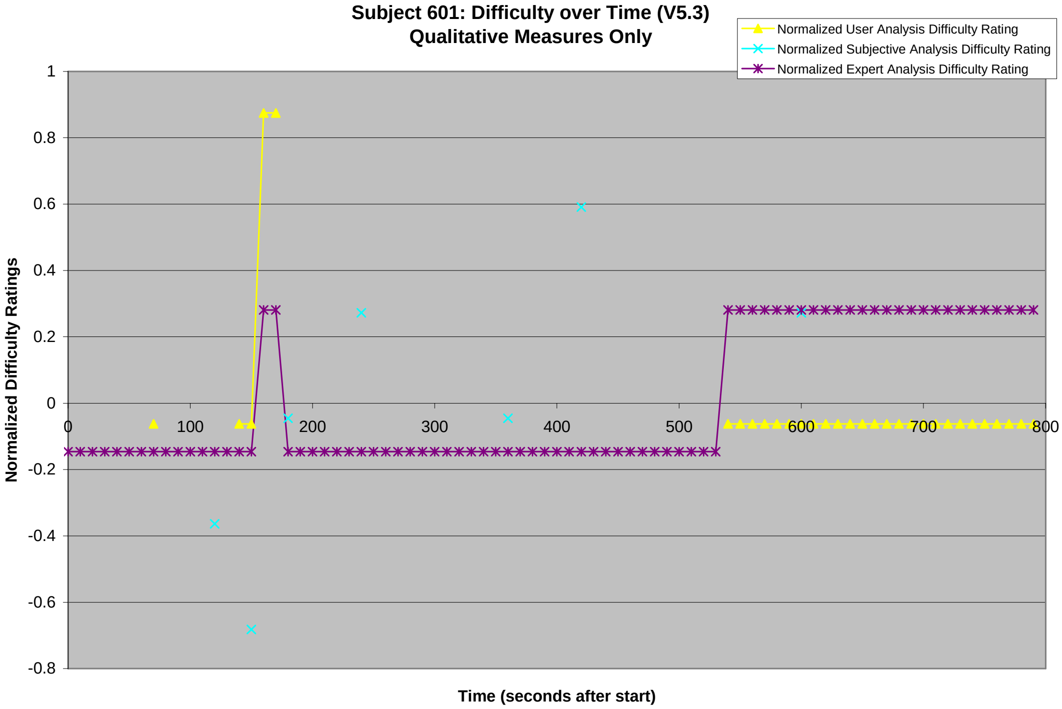


Figure 14: Psychological measurements (Subject 601)

In this example (Subject 601), the neurophysiological measures of workload and arousal match consistently and the expert prediction and self-evaluation measures exhibit moderate correlation. The subjective analysis does not align with either of the other measures.

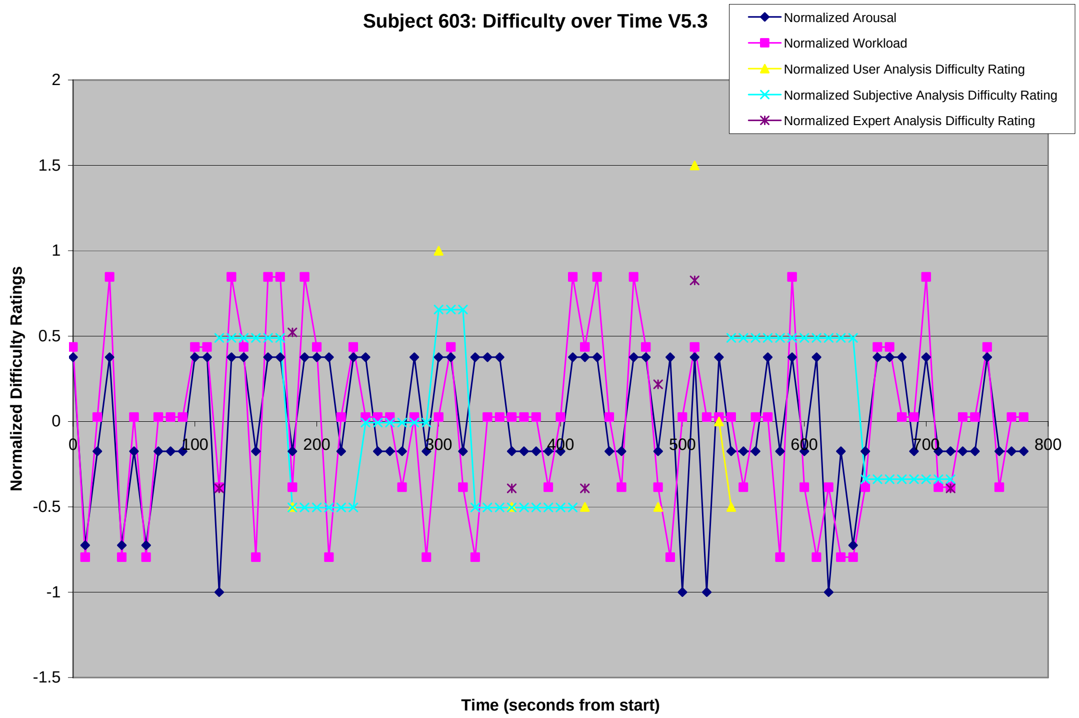


Figure 15: Comparison of effort level measurements (Subject 603)

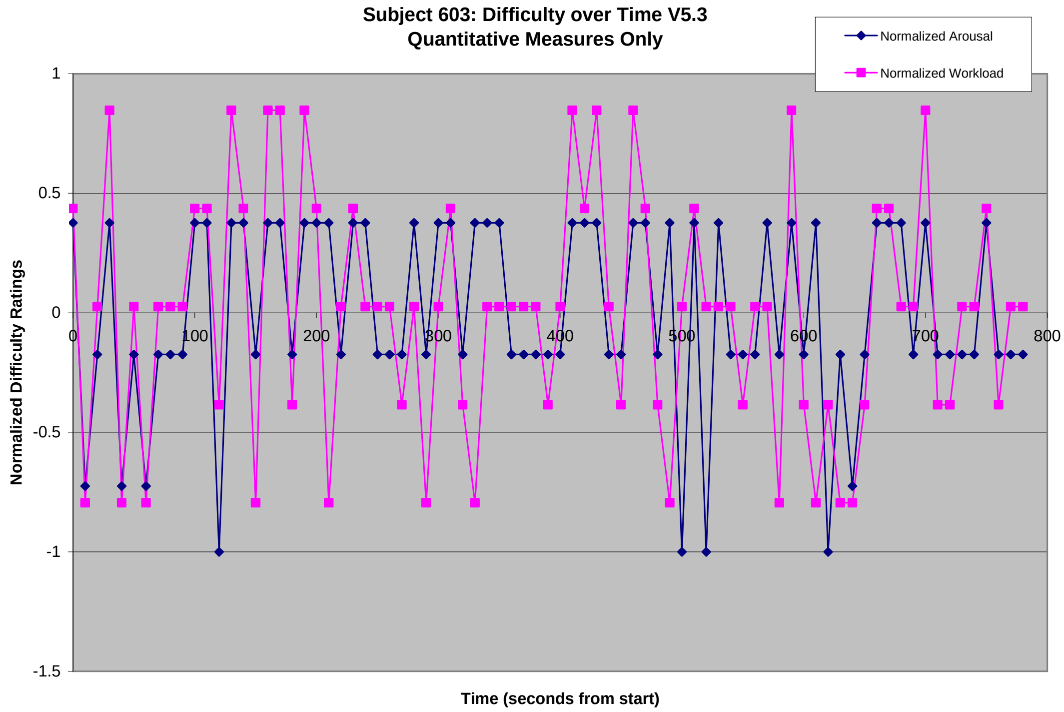


Figure 16: Neurophysiological measurements (Subject 603)

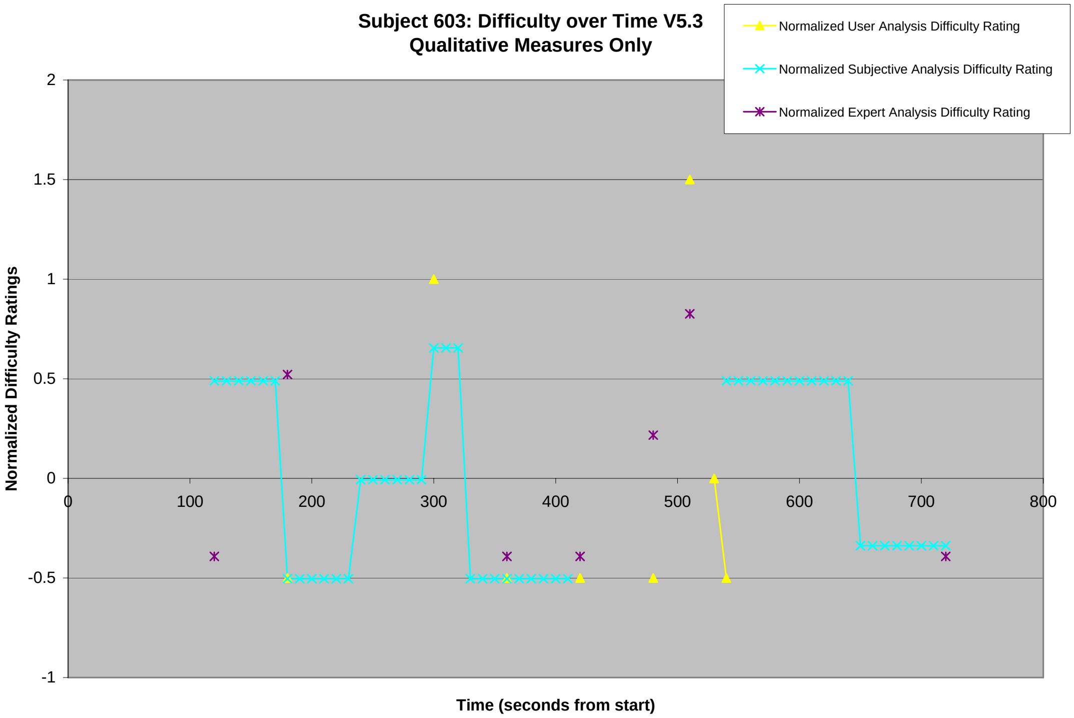


Figure 17: Psychological measurements (Subject 603)

In this example (Subject 603), the neurophysiological measures of workload and arousal again match consistently and all of the psychological measures correlate.

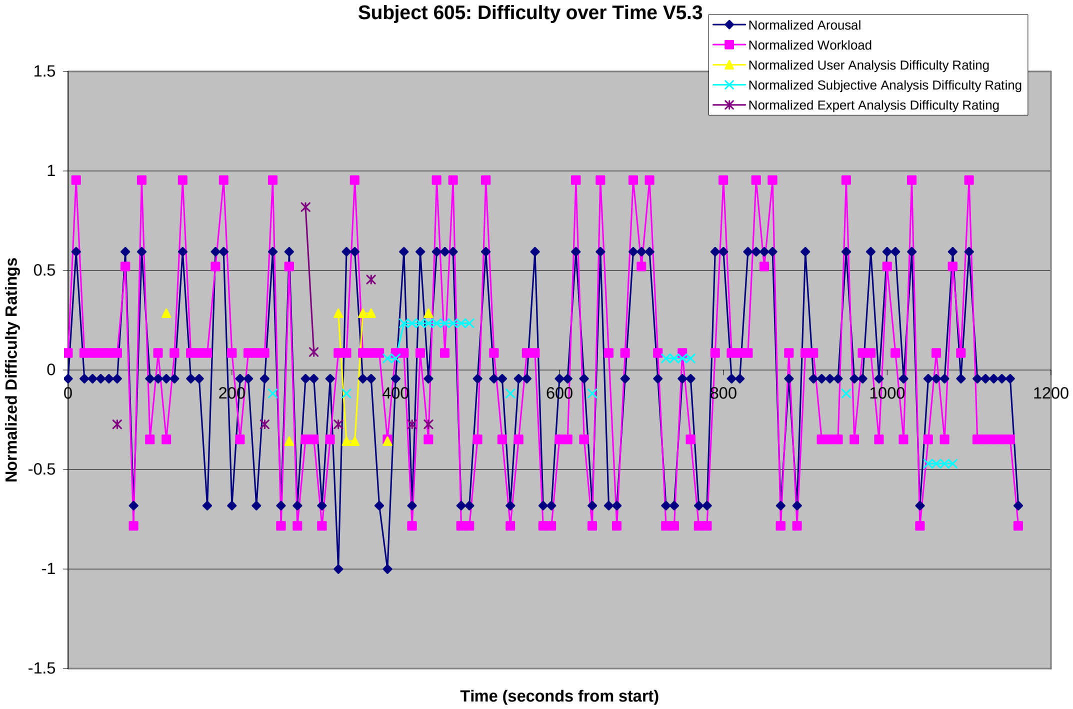


Figure 18: Comparison of effort level measurements (Subject 605)

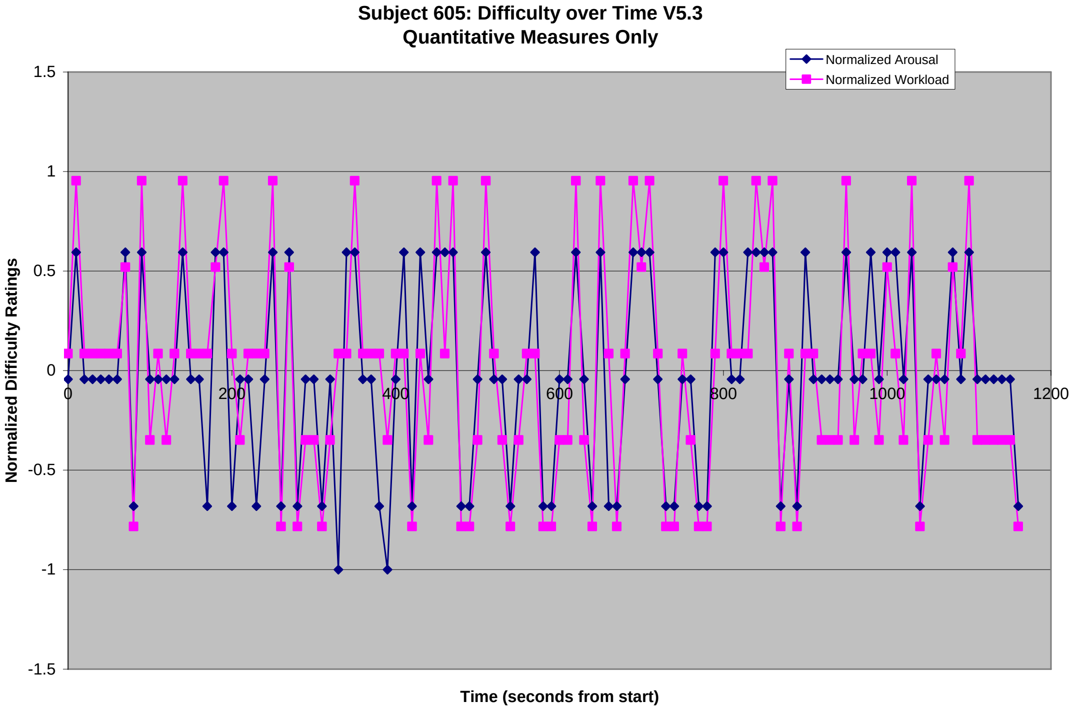


Figure 19: Neurophysiological measurements (Subject 605)

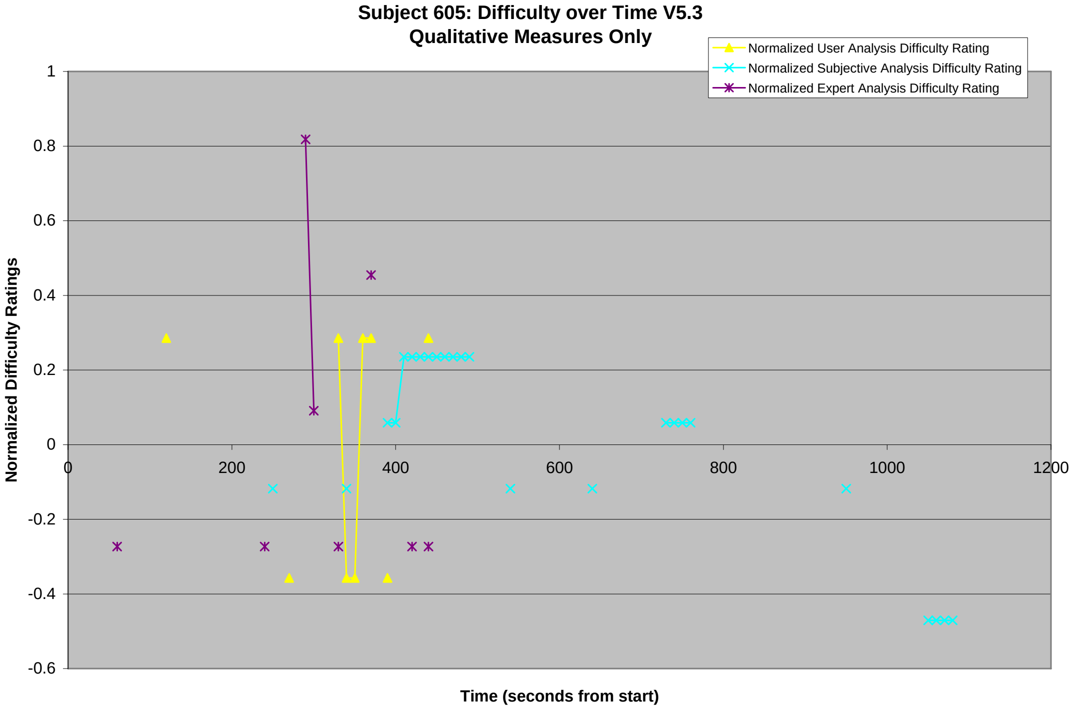


Figure 20: Psychological measurements (Subject 605)

In this example (Subject 605), the neurophysiological measures of workload and arousal again match consistently but the psychological measures are very infrequent and do not demonstrate consistency.

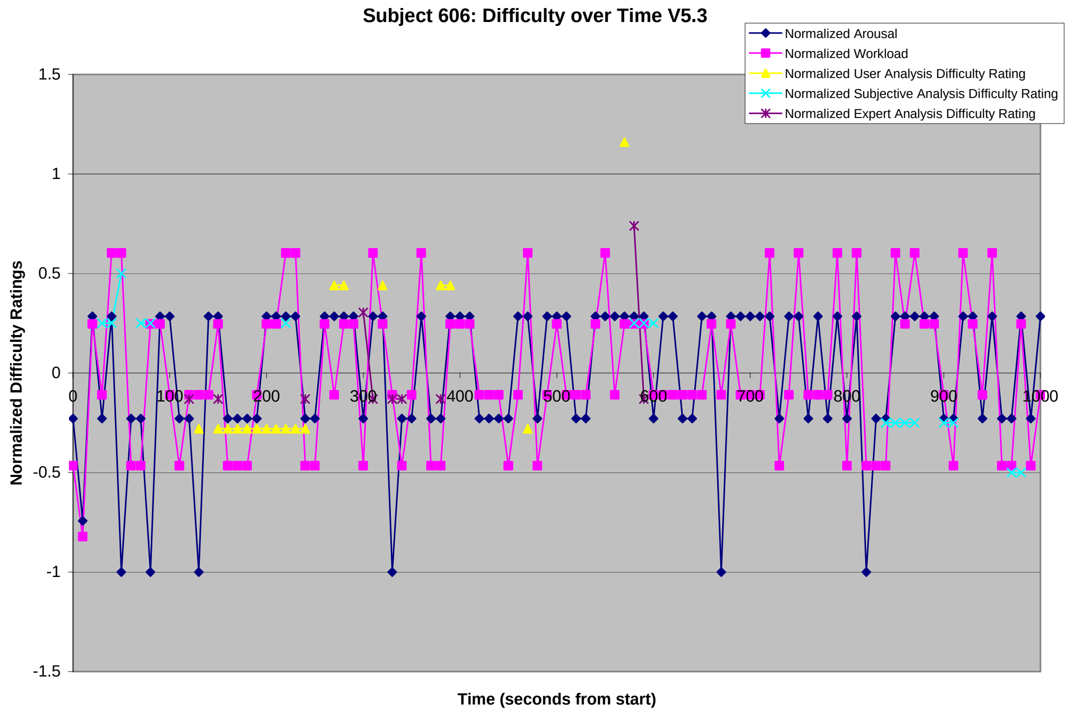


Figure 21: Comparison of effort level measurements (Subject 606)

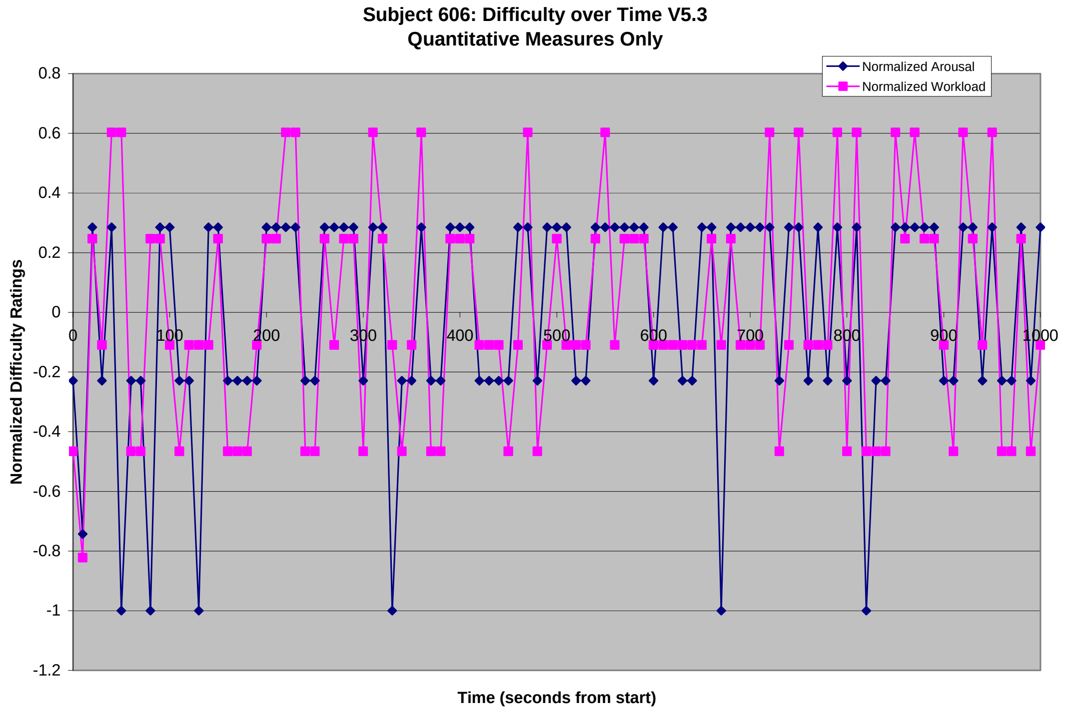


Figure 22: Neurophysiological measurements (Subject 606)

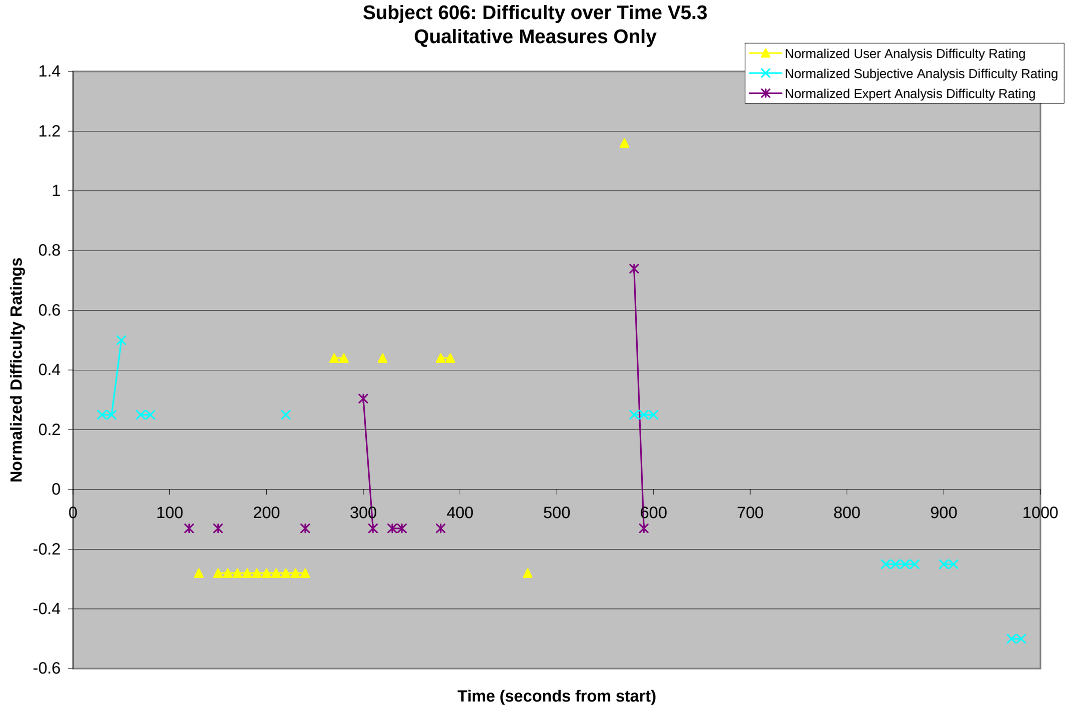


Figure 23: Psychological measurements (Subject 606)

In this example (Subject 606), the neurophysiological measures of workload and arousal again match consistently and the psychological measures show a relationship.

Comparison Summary

These comparisons show the value added by utilizing neurophysiological measures of workload and arousal. These measures can provide the researcher with many more data points and more objectivity than traditional psychological measures. The results from these pilot studies indicate correlation among measures, but the time differences between them prohibit the calculation and assertion of significance.

An initial comparison between the normalized arousal, workload, and expert analysis difficulty ratings yields similar conclusions to those obtained by visual inspection of the graph (as shown in Table 3).

Subject	Pearson's R between Arousal and Workload	Pearson's R between Arousal and Expert Analysis	Pearson's R between Workload and Expert Analysis
601	0.6454	0.0646	0.0374
603	0.5543	0.3917	0.1616
605	0.7929	0.3872	-0.0412
606	0.4635	0.2601	0.1047

Table 3: Data Correlation using Pearson's R.

With additional data points (as can be obtained by following the requirements specified in Chapter 5), analysis including the subjective evaluation (from the naïve observer), self-evaluation, and comparison between subjects can be completed on this type of data.

Lessons Learned

Importance of time-stamping

The most important system attribute that was overlooked in the neurophysiological sensor integration pilot studies was the time-stamping, or time-syncing of the equipment being used. Though this potential issue was generally considered by researchers, it was not fully investigated by the team before the experiments took place. Therefore, the computational and observational systems were synchronized manually, by comparing the times on various systems as called out by researchers and participants and several other measures were taken by hand. This level of specification is inadequate for the neurophysiological sensors, as they take multiple measurements each second. The capability for this time-stamping should be built into the systems being used.

Personnel selection and training

In the absence of complete training documentation, the selection of appropriate personnel for sensor application, system configuration, task completion, and observation is essential. Finding the ideal person with a complete understanding of the task at hand is quite difficult and can be mitigated by producing thorough training guidance that indicates the tasks that must be performed and how exactly to perform them.

Planning

Most of the things learned and to be changed in future research instantiations boil down to general planning issues. On one hand, the need for planning is precisely what drives the implementation of pilot studies – these studies force researchers, designers, and participants to interact in an experimental environment and force detailed thinking about the actual implementation of the planned experiments. However, much work and delays could be prevented with a little mental engineering before the process starts. These pilot studies would have been much more useful if they could have been used to refine the final experiments and tweak the system setup to allow for ideal data collection, rather than to provide a forum for initial trouble-shooting.

Future Research

Given the nature of this project, considerable future research efforts are anticipated. The recommendations included here highlight some of the important areas of study to consider, beyond just the incorporation of the requirements specified in the next chapter.

Deeper analysis with cleaner data

Once the time-synchronization issues (and other requirements as documented in Chapter 5) have been alleviated, further research can be done to compare the measures from the neurophysiological and traditional usability testing techniques. These comparisons can be approached from the point of view of validating one measure or another, or determining the added value of direct cognitive workload measures in conjunction with traditional methods.

Incorporating additional measurement techniques

Additional sensing technologies can be added to the sensor suite to glean further information about cognitive workload as a user completes her tasks. Sensors such as eye/gaze tracking and pupilometry, mouse tracking, and fNIR provide different information than the current sensing technologies, and could provide researchers with important details about cognitive workload.

Incorporating real-time questioning based on quantitative workload measures

If the system could produce a real-time report of when the cognitive workload gauge crossed high and low threshold values, participants could also be specifically asked about these epochs during the procedure (see Debrief, page

40). This would add another cross-reference between the neurophysiological and traditional measures of workload, giving additional information on the efficacy of the neurophysiological measures.

Automate Training

Many of the participants in the data collection process (the subject, the naïve observer, the expert, the sensor operator, the computer operator, etc…) require training about how they should complete their tasks. Incorporating automated training would both speed and give consistency to the process of educating these participants on their roles. Automation of the subject's training could be difficult due to the rapidly-changing aspects of interface prototypes, but all of the other roles in particular, should be consistent enough to rely on an automated training package.

Closing the loop

Existing Augmented Cognition systems utilize the mitigation strategies discussed in the background chapter to close the loop with the human operator. Closing the loop enables the computational system to adapt to the user in real time rather than forcing the user to adapt to the computational system. Though the goal of incorporating neurophysiological measures into the TTWCS design process is to improve the interface design, not improve performance, a closed-loop system could also help in this application.

If HCI designers developed multiple interface options, a closed-loop evaluation system could enable users to test different interfaces and provide researchers with feedback on several different scenarios in one testing environment. Additionally, AugCog technologies could allow researchers to maintain usability testers at a peak level of performance by maintaining an optimal balance of arousal and workload, to ensure that the system being tested was being evaluated under ideal circumstances.

Chapter 5: Requirements Generation

In performing the pilot studies for this experiment, it became clear that several requirements for incorporating neurophysiological measures of the effectiveness of HCIs had not been addressed. The researchers performing the experiments have been working in human factors and related fields for many years and were quite well-versed in the requirements for traditional evaluation. However, the added complexity of real-time neurophysiological measures (and the comparison of those measures with the traditional psychological measures) demands additional consideration.

High-Level Requirements

Requirements Generation Model

As discussed in the background chapter, Lobo (2004) specifies an activity-based model for generating requirements. As applied to this system, the model gives us the following:

Activity Name	**EEG measurement**
Objective	Understand electrical activity in the brain during a task to evaluate cognitive state
Action Points	Measure electrical potentials at 6-9 scalp locations at least 1x/second
Pre-condition	1. Subject is outfitted with EEG sensor 2. Sensor is properly connected to computational system that will analyze data 3. EEG system is calibrated for individual user
Doer	Subject; PI
Participants	Subject; PI

Input Documents	1. Training guidance on the application of the sensor 2. Training guidance on the connection of the sensor to the computational system
Output Documents	Record of EEG data sent to computational system

Activity Name	**EKG measurement**
Objective	Understand electrical activity in the heart during a task to evaluate cognitive state
Action Points	Measure electrical activity in the muscles surrounding the heart at least 1x/second
Pre-condition	1. Subject is outfitted with EKG sensor 2. Sensor is properly connected to computational system that will analyze data
Doer	Subject; PI
Participants	Subject; PI
Input Documents	1. Training guidance on the application of the sensor 2. Training guidance on the connection of the sensor to the computational system
Output Documents	Record of EKG data sent to computational system

Activity Name	**GSR measurement**
Objective	Understand moisture content of the skin during a task to evaluate cognitive state
Action Points	Measure skin conductance at least 1x/second
Pre-condition	1. Subject is outfitted with GSR sensor 2. Sensor is properly connected to computational system that will analyze data
Doer	Subject; PI
Participants	Subject; PI
Input Documents	1. Training guidance on the application of the sensor 2. Training guidance on the connection of the sensor to the computational system
Output Documents	Record of GSR data sent to computational system

Activity Name	**Naïve observer measurement**
Objective	Objectively document the focus, attention, and effort of a subject on elements of the task
Action Points	Note workload on a Likert scale (1-7) on a precise timescale

Pre-condition	1. Observer understands workload-indicating factors that a subject may showcase 2. Observer understands rating scale 3. Observer is outfitted with a mechanism to enable precise time-stamping of ratings
Doer	Subject; naïve observer
Participants	Subject; naïve observer
Input Documents	1. Training guidance on the exhibition of workload 2. Training guidance on the Likert scale
Output Documents	List of time-stamped psychological (rated 1-7) observations of behavior

Activity Name	**SME measurement**
Objective	Predict portions of the task that will be easy or difficult
Action Points	Predict workload on a Likert scale (1-7) by task
Pre-condition	1. SME is provided with a list of tasks that the subject will complete 2. SME understands the rating scale
Doer	SME
Participants	SME
Input Documents	1. List of tasks 2. Training guidance on the Likert scale
Output Documents	List of event-stamped psychological (rated 1-7) predictions of behavior

Activity Name	**Self-measurement**
Objective	Subjectively document the focus, attention, and effort of a subject on elements of the task
Action Points	Note workload on a Likert scale (1-7) on a precise timescale
Pre-condition	1. Subject understands rating scale 2. Subject is outfitted with a mechanism to enable precise time-stamping of ratings
Doer	Subject; debriefer
Participants	Subject; debriefer
Input Documents	Training guidance on the Likert scale
Output Documents	List of time-stamped psychological (rated 1-7) observations of behavior

Activity Name	**Process raw data**
Objective	Convert data from the neurophysiological measures into meaningful information about cognitive state
Action Points	Cognitive state measures must be outputted

Pre-condition	1. Raw data must be available to the computational system 2. Methods to convert the data to measures, or gauges, must be implemented in the system
Doer	Computational system
Participants	Computational system
Input Documents	Explanation of methods to create cognitive state gauges
Output Documents	Time-stamped, interpreted measures of cognitive state

Activity Name	**Compare neurophysiological and traditional psychological measures**
Objective	Understand the benefits and drawbacks of each type of measure
Action Points	Precise comparison by task and time relating the various measures of cognitive state
Pre-condition	1. All measures must be correlated precisely 2. All measures must be normalized
Doer	Computational system; PI
Participants	Computational system; PI
Input Documents	Explanation of normalization methods
Output Documents	Comparison of measures; can be in several forms

Activity Name	**Workload-inducing task completion**
Objective	Elicit arousal and workload to be measured by neurophysiological and traditional psychological means
Action Points	Elicit varying levels of arousal and workload
Pre-condition	1. Software for presenting the task must be accessible 2. Space for completing the task must be accessible 3. Hardware for completing the task must be accessible 4. Subject must have an understanding of the task 5. IRB approval for task completion and measurement
Doer	Subject
Participants	Subject
Input Documents	1. Training guidance for the task 2. Documentation of IRB approval
Output Documents	Time-stamped record of actions involved in the task

Most of the requirements generated by this model stem from the pre-conditions and input and output documents.

Training Requirements

1. Training guidance on the application of the sensor exists
2. Training guidance on the connection of the sensor to the computational system exists
3. Training guidance on the Likert scale/ratings system exists
4. Subject understands the task
5. Observer understands workload-indicating factors that a subject may showcase
6. Rating scale is understood

Sensor Requirements

7. Subject is properly outfitted with various sensors
8. Sensors are properly connected to the computational system
9. If necessary, sensors are calibrated for subject

Computational System Requirements

10. Mechanisms to enable precise time-stamping of ratings exist
11. Sensor data is time-stamped
12. Methods to convert the data to measures, or gauges, must be implemented in the system
13. Computational system can receive and store data from the sensors
14. Computational system can create a time-stamped record of actions involved in the task

Background/Supporting Requirements

15. A list of actions/steps associated with the task exists
16. Software for presenting the task must be accessible
17. Space for completing the task must be accessible
18. Hardware for completing the task must be accessible
19. Institutional Review Board (IRB) approval for task completion and measurement
20. Method for normalizing data exists

These are largely self-explanatory (i.e. "Training guidance for this task") though they require a more detailed approach to satisfy all of the needs of the system. However, several of the requirements generated by this model are complex.

Complex Requirements

Of the 20 requirements listed here, at least six are very complex:

- IRB approval for task completion and measurement
- Software for presenting the task must be accessible
- Subject must have an understanding of the task
- Computational system can create a time-stamped record of actions involved in the task
- Mechanisms to enable precise time-stamping of ratings exist
- Methods to convert the data to measures, or gauges, must be implemented in the system exist

IRB Approval

This process can be a long and arduous one, and is often enough to significantly delay the progress of a project involving humans. Human subject use is defined by the US Department of Health and Human Services (2005) as “a living individual about whom an investigator conducting research obtains data through intervention or interaction with the individual, or identifiable private information.” Any project involving human subject use must have approval for study from an IRB before proceeding, particularly if the funding organization sponsoring the research requires the approval. Obtaining IRB approval can seem like a mere administrative exercise, but since it governs even the start of data collection and must be addressed well in advance. Additionally, researchers should ensure that all researchers participating in the experiment have approval from their respective organizations before research begins.

Software Development

The software used for this testing plays a non-trivial role in the experimentation. Usually, the entire purpose of the HCI testing is to improve a software package, and an engineering team has spent months or even years improving the functionality, and hopefully the interface, of their product. The earlier the human factors design team can be involved with the product, the better, as late suggestions and changes to the software package can be expensive and time-consuming. The changes that come about during human testing of the software can be minimized by involving the HCI design team in the process early. Additionally, pilot studies with initial versions or even storyboards of the final product can give researchers ideas about which portions should be improved before the design is finalized.

Task Understanding

It is difficult to ensure that a subject understands the task at hand as a requirement for the experiment. However, subjects must have enough familiarity with the task to complete it on their own and to think through the processes involved in completing the task. Otherwise, the ratings regarding cognitive state lose their meaning – if an instructor has to guide a subject through the process step-by-step, the subject is not thinking or acting as they normally would and measurements about their level of arousal/workload will be inadequate. It is still helpful to perform pilot testing with subject who do not completely understand the task but they must have moved beyond novice level to get meaningful information.

Time-stamping of actions

The software needs to provide a time-stamped record of the actions taken by the user. An observer can try to document precisely what a subject is doing at a specific point in time, but for the neurophysiological measures proposed here to be accurate, the actions that the subject is taking must be correlated down to the second, or even more precisely. Once the provision for action time-stamping is made, individual actions must be translated into tasks or steps. For example, it is easy enough for a computer to understand that the subject moved the mouse to certain coordinates in on a screen, but the important information, to the experimenter, is what the subject was doing when they were moving or clicking the mouse – were they requesting data on a particular missile? Were they confirming and action box request? Were they opening a different view of the system? A history or trace file of the user interface actions will correlate the subject’s tasks/steps with time.

Time-stamping of ratings

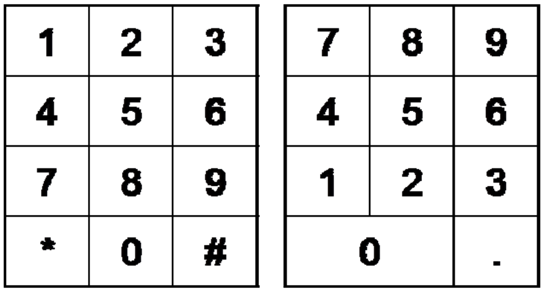
All of the ratings of cognitive state taken must be time-stamped and accurately correlated with each other. The four types of ratings (SME prediction, naïve observation, cognitive state gauge values, and subject self-evaluation) all need to be linked together, and the only possible way to do this is by time. The driving factor for the time-stamping is the neurophysiological data collection, which is collected several times per second. Therefore, any ratings that are going to be compared to that collection must have an extremely accurate timescale.

SME predictions, necessarily, are done by task, which is why the tasks need to be time-stamped as indicated in the previous section. This also requires that the computer on which the task is being completed and the computer that is recording the neurophysiological measures be on the exact same timescale. Additionally, there needs to be a clear marking somewhere of when the subject started the task (or the zero time-stamp). Both the sensor/gauge system (due to calibration needs) and the task system (due to software configuration needs) must begin running well before the actual task begins, so there must be a way to mark ‘zero time’ on all of the equipment.

The naïve observation also needs to be time-stamped precisely, not just by marking down times and making a note of perceived workload. This method is imprecise and is too time-consuming to allow an observer to make as frequent observations as are needed. One solution would be to link a naïve observation system consisting of only a numerical keypad, such as is shown in Figure 24, to the same system that is getting the cognitive state data. This way, a naïve observer could merely punch in their score of workload (from 1-7) and it would be automatically associated with a time in the task.

Figure 24: Sample numerical keypad for naïve observer input.

One caveat with the use of a numerical keyboard is the potential confusion between the "calculator-style" layout and the "phone-style" layout (as compared in Figure 25). Care should be used to ensure that the naïve observer does not confuse the number placement.



1	2	3
4	5	6
7	8	9
*	0	#

7	8	9
4	5	6
1	2	3
0		.

Figure 25: Illustration comparing "phone-style" numerical keypad layouts (left) to "calculator-style" numerical keypad layouts (right).

Finally, the subject self-evaluation needs to be tied to the other measures. Having a subject review a video on an unlinked system and rate different portions of their task at different load levels is not sufficient – their data also needs to be comparable to the neurophysiological measures, and on a unified timescale. This sort of general feedback (i.e. "task #21 was mildly challenging, I'd give it a score of five.") may be useful to the traditional analysts, but it is not sufficient for comparison with the neurophysiological measures. For this comparison, a setup similar to the one proposed for the naïve observer should be used. In this case, the video that the subject reviews should be on or linked to a system that can correlate the video time-stamp with the software being evaluated and the cognitive state gauges. That system could also be linked to a numerical keypad that would allow the subject to just punch a button and have that score automatically linked to a particular portion of the task. This system would also speed the processing of data, as all of the rating details could be downloaded into one spreadsheet or database. As the number of subjects grows, the data processing time for each additional piece of information becomes increasingly important.

Cognitive state gauges

The cognitive state gauges that convert the raw cognitive state information from the EEG, EKG, and GSR sensors are extremely important for the whole evaluation to function in real-time. Raw data has extensive artifacts and must be processed offline before it can be used and even data that is processed provides extraneous and unrelated information to the task at hand. The cognitive state gauges used in this system and Augmented Cognition systems provide experimenters with meaningful information about cognitive state that can be used to quantify improvements in HCI design. For more on the development of cognitive state gauges, see the Background chapter.

Detailed Requirements

Each of the requirements specified in the "High-Level Requirements" section involves additional detailed specifications. These specifications are highlighted here.

Training Requirements

Training guidance on the application of the sensor exists. Each type of sensor incorporated into the testing protocol must have extensive and precise instructions for application. Many of the researchers who do usability testing may not be used to the hardware specifications of these instruments and many instruments are extremely sensitive to placement (i.e. an EEG sensor may collect totally wrong data if improperly placed), connection (certain parts of the hardware may have to be initiated before others), and other aspects. The instructions should be easy to follow, and should be briefed to the researchers before the experiments begin, so they have time to understand and ask any necessary questions.

Training guidance on the connection of the sensor to the computational system exists. As anyone who uses a computer can attest, sometimes even the order of booting up a system, connecting hardware, and initializing software can dramatically impact the performance of system components. Therefore, clear guidance on the method and order of connecting sensor hardware to the computational processing system must be available to and learned by the researchers.

Training guidance on the Likert scale/ratings system exists. Clear guidance regarding the rating systems to be used must be developed and/or presented to experiment participants. Even though some participants may have existing understanding of rating systems, universal guidance helps to ensure that everyone is starting with the same level/nature of understanding.

Subject must have an understanding of the task. The subject performing the task needs to have at least a working-level understanding of their task in order to elicit accurate ratings of cognitive workload caused by the task. If a subject is merely following directions or listening to an instructor, they will not appear to be experiencing changes in cognitive state and they will not accurately represent the difficulty of various aspects of a task in an operational environment.

Observer understands workload-indicating factors that a subject may showcase. It is important that such observers understand some of the attributes and actions that a participant may exhibit while experiencing varying levels of stress.

Rating scale is understood. It is essential that anyone performing ratings on tasks (an SME, naïve observer, or subject) has an understanding or the ratings they are going to use regarding difficulty. Many researchers are familiar with the Likert scale of ratings, but it may seem counterintuitive to a subject to think of 'four' as a neutral score. Care must be exercised to instruct the participants on the meaning of their ratings.

Sensor Requirements

Subject is properly outfitted with various sensors. Researchers must follow guidance regarding the placement and connection of neurophysiological sensors.

Sensors are properly connected to the computational system. Researchers must follow guidance regarding the connection of the neurophysiological sensors to the computational processing system.

If necessary, sensors are calibrated for subject. Most neurophysiological sensors must be calibrated for and adjusted to individual subjects. Complex gauges can require additional calibration time (St. John, Kobus, Morrison, 2003). Many sensors require an additional short calibration time before each session – this calibration can usually be completed while other aspects of the experiment are being readied.

Computational System Requirements

Mechanisms to enable precise time-stamping of ratings exist. Ratings from a naïve observer or self-evaluation ratings from a user must have accurate time-stamps in order to correlate them with the timing of task aspects and the ratings that are associated with those tasks.

Sensor data is time-stamped. Data from the cognitive state gauges must be time-stamped and correlated with all other computational elements in the system. This is the only way to accurately compare data from all aspects of the system.

Methods to convert the data to measures, or gauges, must be implemented in the system. While raw data regarding cognitive state is useful to many researchers, that data must be converted into more meaningful information (or gauges) in order to be utilized by the system in real time.

Computational system can receive and store data from the sensors. The computational processing system must be able to receive real-time information from the neurophysiological systems. The cognitive state gauges usually reside on the computational system, so the connection must be consistent and have a low latency to allow real-time operation.

Computational system can create a time-stamped record of actions involved in the task. The system on which the software being tested resides must create a time-stamped record of the tasks that a user completes. Many software packages exist for documenting this kind of interaction (Noldus et. al, 1999), and it is the only way to ensure that the task-related ratings will correlate with the other time-based ratings of cognitive state.

Background and Supporting Requirements

A list of actions or steps associated with the task exists. The TTWCS design team specified the steps required to complete their desired actions (TWCS SDA Team, 2004), but the options available in the actual interface made it difficult to track exactly what the user was doing. In many instances, the obscurity of such directions makes them difficult for the user to follow. Ideally, the list of steps required by a user should be both clear enough to follow with little direction and restrictive enough that researchers can elicit the desired actions of users.

Software for presenting the task must be accessible. The software design team should be aware of and held to code-drop deadlines and specifications. The organization of usability testing requires significant planning and support, and cannot take place unless the software to be tested is ready.

Space for completing the task must be accessible. An appropriate environment in which the user can complete a task must be supplied. The desired environment may mimic the operational environment in which the product will eventually be used, or it may be designed to minimize interruptions and distractions in order to allow the user to fully concentrate on the task at hand. In either case, the desired environment (and purpose of the environment) should be considered well before the usability testing date, should be clearly documented, and should be implemented with discipline.

Hardware for completing the task must be accessible. Appropriate hardware can also have an impact on the user's environment while testing. Again, researchers must decide on the level of fidelity required in order to get good usability data from participants, as more reality at a higher cost is not always better (Cohn, Stripling & Kruse, 2005).

IRB approval for task completion and measurement. The IRB approval process can be long and arduous, and can prevent an otherwise well-formed experiment from proceeding. Obtaining IRB approval should be one of the first things that a researcher examines, at least 3-4 months before the planned start of experimentation.

Method for normalizing data exists. The data collected by the sensors is raw, unformatted, and not related to much of the traditional, psychological measures taken during usability testing. There must be a methodology for (1) cleaning and filtering the raw sensor data into useable gauge information regarding cognitive state, and (2) determining a meaningful comparison between the neurophysiological and psychological cognitive workload measures. The studies here used a simple normalization between -1 and 1 for all data values, but that method can only provide a comparison of the change in values over time. Additional information can be gleaned from the data if a more meaningful normalization structure is implemented.

Requirements Summary

Most of the requirements discussed here are easily implemented once identified, though the more complex requirements necessitate additional consideration and planning. The next steps in working with these requirements are to extend them into numerical specifications and to complete the steps necessary to meet the requirements.

Chapter 6: Conclusions

There are two major discoveries resulting from this research: an initial demonstration of the value added by incorporating these measures into traditional HCI improvement efforts and the importance of requirements in the inclusion of neurophysiological measures. These discoveries are extremely valuable to researchers who are aiming to incorporate quantitative measures of cognitive workload measures into their HCI improvement efforts.

Applicability of Quantitative Cognitive State Measures

Initial measures comparing neurophysiological and traditional psychological measures of cognitive workload were illustrated here. Some of the greatest added values of neurophysiological measures are:

- Increased objectivity of measurements (neither researchers' nor participants' individual or time-based view changes can skew the data)
- Increased consistency of measurements (measurements are taken consistently over time and do not depend on who is taking them)
- Additional information availability (measurements taken with neurophysiological sensors provide information that is not available from traditional questioning and measurements)

The results indicated here illustrate some correlation between neurophysiological and traditional psychological measures of cognitive state, but more importantly, show the value of incorporating these neurophysiological measures into the HCI improvement process.

Importance of Requirements

The initial goal of the TTWCS pilot studies was to compare neurophysiological and traditional psychological measures of cognitive workload in a semi-operational setting. It became apparent, through the course of experimentation, that the experimental setup was not sufficient to enable validation of the new neurophysiological-sensor-based quantitative measures.

The documentation of requirements before experimentation is absolutely necessary, and the list of high-level and detailed requirements specified here provides researchers with the tools to initiate the work in this area.

Glossary

ABM	Advanced Brain Monitoring (www.b-alert.com)
ALS	Amyotrophic Lateral Sclerosis, or Lou Gehrig's Disease
AugCog	Augmented Cognition
DARPA	Defense Advanced Research Projects Agency
DoD	Department of Defense
EEG	Electroencephalography
EOG	Electrooculography
ERICA	Eye-gaze Response Interface Computer Aid
fNIR	Functional near infrared
G2SEBoK	Guide to Systems Engineering Body of Knowledge
GSR	Galvanic skin response
HCI	Human Computer Interaction
HR	Heart rate
IEA	International Ergonomics Association
INCOSE	International Council on Systems Engineering
IRB	Institutional Review Board
LMATL	Lockheed Martin Advanced Technology Laboratories
NASA TLX	NASA Task Load Index
NASA	National Aeronautics and Space Administration
NAVLIB	Naval Public Library
SAGAT	Situation Awareness Global Assessment Technique
SDA	System Design Architecture

SME	Subject Matter Expert
TTWCS	Tactical Tomahawk Weapons Control System
TWCS	Tomahawk Weapons Control System
UCD	User-Centered Design
UVA	University of Virginia
WCS	Weapons Control System

Bibliography

Allen, S. (2004) Interface Design and Team Structure in a Collaborative Environment. University of Virginia Master's Thesis.

Beatty, J. "Concluding Report: ARPA Biocybernetics Project." 1978. Written for contract number N00014-70-C-0030.

Belyavin, A. (2005). Construction of appropriate gauges for the control of Augmented Cognition systems. *Proceedings of the 1st International Conference on Augmented Cognition*, Las Vegas, NV, 22-27 July 2005.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human-Computer Interaction*, 17(2), 151-170.

Campbell, G.E., Pharmer, J.A., & Hildebrand, G. (2000). Measuring the impact of human-centric design on human performance in a combat information center. URL http://www.manningaffordability.com/s&tweb/PUBS/Measuring-HCD-CIC/Measuring-HCD-CIC.htm. Last accessed 15 November 2005.

Card, S.K., Moran, T.P., Newell, A. (1983). The psychology of human-computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohn, J., Kruse, A, and Stripling, R. (July 2005). Investigating the Transition from Novice to Expert in a Virtual Training Environment Using Neuro-Cognitive Measures.

In D. Schmorrow (Ed) Foundations of Augmented Cognition, Lawrence Erlbaum Associates, New Jersey.

Committee on Metabolic Monitoring for Military Field Applications, Standing Committee on Military Nutrition Research (2004). Monitoring Metabolic Status: Predicting Decrements in Physiological and Cognitive Performance. National Academies Press, Washington, DC, 212.

Cummings, M. (2003). Designing decision support systems for a revolutionary command and control domains. University of Virginia Doctoral Dissertation.

Cummings, M. and Guerlain, S. (2003). The tactical Tomahawk conundrum: designing decision support systems for revolutionary domains. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*.

Cushing, J., Allen, S. and Guerlain, S. (2004). Measuring the effectiveness of an overview display for mission planning. *IEEE 2004 Conference on Systems, Man and Cybernetics*, The Hague, Netherlands, pp. 1533-1537.

Davis, A. (1993). Software Requirements: Objects, Functions, & States, Prentice-Hall, Upper Saddle River, New Jersey.

Endsley, M. (1988). Situation awareness global assessment technique (SAGAT). In Proceedings of the National Aerospace and Electronics Conference (NAECON) (pp.789-795). New York: IEEE.

Endsley, M., & Kiris, E. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37 (2), 381-394.

Gerson, A., Parra, L., & Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage*. Manuscript under revision.

Gevins, A., Smith, M.E., Leong, H., McEvoy, L, Whitfield, S., Du, R. and Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors* , 40, 79-91.

Guerlain, S. and Willis, R (2001). The tactical tomahawk weapons control system: operator interface design project. 2001 Human Systems Integration Symposium.

Hart, Sandra G., & Staveland, Lowell E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: Elsevier Science Publishers B.V. (North-Holland).

Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strong and Verplank (2004). ACM SIGCHI Curricula for Human-Computer Interaction, available at http://sigchi.org/cdg/index.html. Last accessed 20 November 2005.

International Council on Systems Engineering (INCOSE) (2003). Guide to Systems Engineering Body of Knowledge (G2SEBoK), available at http://g2sebok.incose.org. Last accessed 20 November 2005.

International Ergonomics Association (IEA) (2000). The Discipline of Ergonomics, available at http://www.iea.cc/ergonomics. Last accessed 20 November 2005.

Israel, J.D., Wickens, C.D., Chesney, G.L., Donchin, E. (1980). The Event Related Brain Potential as an Index of Display-Monitoring Workload. *Human Factors*, 22, pp. 214-217.

Jirotka, M. and Goguen, J. A. (1994). Introduction. In M. Jirotka and J. A. Goguen, editors, Requirements Engineering Social and Technical Issues, pages 1-13. Academic Press, Great Britain.

Jones, C. and Guerlain, S. (2004). Measuring the effectiveness of a "decision matrix" for missile retargeting decisions. *2004 IEEE Systems and Information Engineering Design Symposium*. Charlottesville, VA.

Kincses, W.E. (July 2005). DaimlerChrysler's Closed Loop Integrated Prototype: Current Status and Outlook. In D. Schmorrow (Ed) Foundations of Augmented Cognition, Lawrence Erlbaum Associates, New Jersey.

Kobus, D., Morrison, J. and Schmorrow, D. (2005). DARPA Improving Warfighter Information Intake Under Stress - Augmented Cognition Phase 3 Concept Validation Experiment (CVE) Analysis Report. To be published as a Space and Naval Warfare Systems Center San Diego Technical Report, to be available at (http://www.spawar.navy.mil/sti/publications/pubs/trindex.html).

Krasner, H. (1985). Requirements dynamics in large software projects, a perspective on new directions in the software engineering process. *Proceedings of IFIP*, pages 211-216, New York, Elsevier.

L3 Communications (2005). Product Listing. Available at http://www.l-3com.com/rcc/products/. Last accessed 05 December 2005.

Licklider, JCR (March 1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, volume HFE-1, pages 4–11.

Leite, J.C.S.P. and Freeman, P.A. (1991). Requirements validation through viewpoints resolution. *IEEE Transactions on Software Engineering*, 17(12):1253-1269.

Lobo, L. (2004). Analysis and Evaluation of Methods for Activities in the expanded Requirements Generation Model (xRGM). Virginia Tech Computer Science Department Masters Thesis.

Lockheed Martin Advanced Technology Laboratory (LMATL) (2005). DARPA Improving Warfighter Information Intake Under Stress - Augmented Cognition Phase 3 Concept Validation Experiment (CVE) Analysis Report for the Lockheed Martin ATL Team. To be published as a Space and Naval Warfare Systems Center San Diego Technical Report, to be available at (http://www.spawar.navy.mil/sti/publications/pubs/trindex.html).

Mao, J., Vredenburg, K., Smith, P., Carey, T. (March 2005). The State of User-Centered Deisgn Practice. *Communications of the ACM*, v 48, n 3, p 105-109.

Mayank V., Kositsyna N., and Austin M.A. (March 2004). "Requirements Engineering and the Semantic Web: Part II. Representation, Management and Validation of Requirements and System-Level Architectures," ISR Technical Report 2004-14, University of Maryland, College Park, MD., (p. 72).

Marshall, L. and Raley, C. (December 2004). Platform-Based Design of Augmented Cognition Systems. Project Report for ENSE623 course at the University of Maryland ISR.

Navy Public Affairs Library (NAVLIB), (1993). Tomahawk Cruise Missile Fact Sheet. Published by DOD Public Affairs Office; Program Executive Office for the Cruise Missile Project and Unmanned Aerial Vehicles Joint Project; Washington, DC 20361-1014; (202) 692-0565, available at http://www.chinfo.navy.mil/navpalib/weapons/missiles/tomahawk/facts.txt. Last accessed 22 November 2005.

Shishko, R., and Chamberlain, R. (1995). NASA Systems Engineering Handbook. NASA systems engineering handbook. NASA Jet Propulsion Laboratory, xii. Available at http://ldcm.gsfc.nasa.gov/library/Systems_Engineering_Handbook.pdf. Last accessed 20 November 2005.

Nielsen, J. and Mack, R.L. (eds.), (1994). Usability Inspection Methods, New York: John Wiley.

Noldus, L., Kwint, A., ten Hove, W., and Derix, R. (1999). Software tools for collection and analysis of observational data. In: Human-Computer Interaction: Communication, Cooperation and Application Design. *Proceedings of the 8th International Conference on Human Computer Interaction*, Vol, 1 (Eds. Bullinger, H.J. and Ziegler, J.), pp. 1114-1118.

Norman, K.L. & Panizzi, M. (April 2004). Levels of Automation and User Participation in Usability Testing. LAP-2004-01, HCIL-2004-17.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theoretical Issues in Ergonomics Science*, vol. 4, nos. 1–2, 5–20.

Potts, C., Takahashi, K., and Anton, A.I. (1994). Inquiry-based requirements analysis. *IEEE Software*, 11(2):21-32.

Radding, M., Siegel, S., and Russell, C. (2005). HCI (DTOF) Project Phase 1 Report. Produced for Lockheed Martin and the US Navy PMA-282.

Raley, C., Stripling, R., Kruse, A., Schmorrow, D., Patrey, J. (September 2004). Augmented Cognition Overview: Improving Information Intake Under Stress. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Conference*, New Orleans, LA.

Rapoport, E., Downs, T., Proffitt, D., Downs, J.H. (July 2005). First Investigations of an fNIR-Based Brain-Computer Interface. In D. Schmorrow (Ed) Foundations of Augmented Cognition, Lawrence Erlbaum Associates, New Jersey.

Sangiovani-Vincentelli, A. (2002). Defining Platform-based Design, *EE Design*, March 5, 2002.

Schmorrow, D. D. and Kruse, A. (2004). Augmented Cognition. In W. Bainbridge (Ed) Berkshire Encyclopedia of Human-Computer Interaction: When science fiction becomes science fact. (pp 54-59) Berkshire Publishing Group.

Selberg S.A. and Austin M.A. (May 2003). "Requirements Engineering and the Semantic Web," ISR Technical Report 2003-20, University of Maryland, College Park, MD., (p. 62).

Sommerville, I. and Sawyer, P. (1997). Viewpoints: principles, problems and a practical approach to requirements engineering. *Annals of Software Engineering, special issue on software requirements engineering*, 3:101-130.

St. John, M., Kobus, D.A., Morrison, J.G. (December 2003). DARPA Augmented Cognition Technical Integration Experiment (TIE), available at http://handle.dtic.mil/100.2/ADA420147. Last accessed 20 November 2005.

TWCS SDA Team (March 2005). TWCS SDA Tactical Tomahawk Weapons Control System (TTWCS) Human Computer Interface (HCI) Test and Evaluation Plan, Revision 0, Lockheed Martin Document # TTWCS-05-1739.

TWCS SDA Team (2004). Tactical Tomahawk Weapons Control System (TTWCS) V5.3 Scenario Script. Produced for Lockheed Martin and the US Navy PMA-282.

United States Department of Health and Human Services (2005). Code of Federal Regulations Title 45. Public Welfare Part 46. Protection of Human Subjects. Available at http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm. Last accessed 20 November 2005.

UVA Department of Systems and Information Engineering HCI (2005). HCI User-Centered Design Process, available at http://www.sys.virginia.edu/hci/usercentereddesign.asp. Last accessed 20 November 2005.

Vredenburg, K., Isensee, S., Righi, C. (2001). User-Centered Design: An Integrated Approach. Prentice Hall PTR, New Jersey.

Willis, R. (2001): Tactical Tomahawk weapon control system user interface analysis and design. University of Virginia Master's Thesis.

Wilson, G.F., Lambert, J.D., and Russell, C.A. (1999). Performance enhancement with real time physiologically controlled adaptive aiding," *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, vol. 3, pp. 61–64.

Wilson, W.M. (1997). Writing Effective Requirements Specifications, Software Technology Conference, Utah, April 1997, available at http://satc.gsfc.nasa.gov/support/STC_APR97/write/writert.html. Last accessed 20 November 2005.

Wood, R., Marajl, B., Lee, C.M., and Reyes, R. (2002). Short-term heart rate variability during a cognitive challenge in young and older adults. *Age and Ageing*; 31: 131-135.

Yerkes, R.M. & Dodson, J.D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

This page intentionally left blank.

Curriculum Vitae

Colby Dean Raley
3601 Wilson Blvd., Suite 500, Arlington, VA 22201 ~ 571-218-4310, craley@sainc.com
701 Pennsylvania Ave NW #1013, Washington, DC 20004 ~ cdraley@gmail.com

Education

University of Maryland, College Park, MD
Master of Science in Systems Engineering, GPA 3.8, expected December 2005

Bachelor of Science in Biological Resources Engineering, GPA 3.81, May 2001

- Graduated *Cum Laude*, University Honors Citation
- Biological Resources Engineering Outstanding Junior & Senior (1999-2001)
- Tau Beta Pi Engineering Honor Society, Phi Kappa Phi Honor Society, Golden Key National Honor Society

Related Experience

Strategic Analysis, Inc., Arlington, VA
Technical Consultant, November 2001 – present

- Evaluate technical performance, milestones and metrics; track collaborations and demonstrations through trip reports and write-ups; oversee public release of technical material; draft milestones, metrics including revision of programmatic and congressional language; evaluate white papers, proposals; locating new sources of funding; identify innovative performers; prepare and review programmatic documentation material; and review, evaluate, and manage proposal submissions
- Manage team correspondence, calendars, and data; collect Quarterly Reports, Quad charts, and one page program descriptions; organize disparate information into usable formats; and track funding activities
- Develop presentations; coordinate large-scale international technical meetings; manage program and meeting websites including design guidance
- Supervise development of program newsletters, websites, and various publications; manage summer internship program and coordinate various activities between government and contractors.

Vaxim Incorporated, Rockville, MD
Laboratory Technician, January – August 2001

- Assisted with every aspect of business and experimentation in this small, startup biotechnology company.
- Created company hierarchy flowcharts and floor plan layouts, oversaw ordering of chemicals and laboratory supplies, trained new technicians, searched, categorized, and organized patents and journal articles, helped set up an animal facility, worked confidential experiments and documents, and performed other tasks as needed.

University of Maryland, College Park, MD
Teaching Fellow and Tutor, August 2000 – May 2001

- Worked with the Engineering Department teaching an introductory engineering design course. Responsible for supervising group activities, holding office hours, and website maintenance.
- Worked in the Athletic Department and Math Success Program as a tutor for biology, engineering, and math courses.

Walter Reed Army Institute of Research, Silver Spring, MD
Laboratory Technician, August 1999 – May 2000

- Researched in the Neuropharmacology and Molecular Biology Department of the Neuroscience Division.
- Performed sodium channel research in neuronal cells using PCR analysis, fluorescent assays, gel electrophoresis, cell culturing, and other methods.

Selected Publications

- Raley, C., Marshall, L. (2005, July) Modular Design for Augmented Cognition Systems. In D. Schmorrow (Ed) Foundations of Augmented Cognition, Lawrence Erlbaum Associates, New Jersey.
- Schmorrow, D., Kruse, A., Raley, C. & Schmidt, E. (2004). Quantifying Human Information Processing. [CD-ROM]. Proceedings from 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, 622.
- Schmorrow, D., Raley, C., Kruse, A. & Schmidt, E. (2004). Augmented Cognition: A Neural Approach to Human Computer Interaction. [CD-ROM]. Proceedings from 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, 792.
- Schmorrow, D., Raley, C., Kruse, A. & Schmidt, E. (2004). Augmented Cognition: Brining The Human In the Loop. [CD-ROM]. Proceedings from 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, 1210.
- Schrauf, M., Kincses, W., Schmorrow, D., Raley, C. (2004) Toward the Perfect Operation of Complex Systems. In D. Vincenzi, M. Mouloua, P. Hanock (Eds.), Human Performance, Situation Awareness and Automation: Vol. 1. Current Research and Trends (pp. 253-257). Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.
- Diethe, T., Dickson, B., Schmorrow, D., Raley, C. (2004) Toward an Augmented Cockpit. In D. Vincenzi, M. Mouloua, P. Hanock (Eds.), Human Performance, Situation Awareness and Automation: Vol. 1. Current Research and Trends (pp. 253-257). Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.
- Graeber, D., Milham, L., Stanney, K., Cohn, J., Schmorrow, D., Raley, C. (2004). Adaptive Multimodal Interfaces in Training and Operational Environments. In D. Vincenzi, M. Mouloua, P. Hanock (Eds.), Human Performance, Situation Awareness and Automation: Vol. 2. Current Research and Trends (pp. 107-111). Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.

- Muller, P., Schmorrow, D., Raley, C. (2004). Military Training and Simulation for the Nintendo Generation. In D. Vincenzi, M. Mouloua, P. Hanock (Eds.), Human Performance, Situation Awareness and Automation: Vol. 2. Current Research and Trends (pp. 195-198). Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.
- Raley, C., Stripling, R., Kruse, A., Schmorrow, D., Patrey, J. (2004, September). Augmented Cognition Overview: Improving Information Intake Under Stress. Proceedings of the Human Factors and Ergonomics Society 48th Annual Conference, New Orleans, LA.
- Raley, C., Schmorrow, D. (2004, May). Biocybernetics to Superior Aviation: A Historical Perspective on Cognition Maximizing Research for the Aviator. Proceedings of the Aerospace Medical Association 75th AsMA Annual Scientific Meeting: Frontiers in Aerospace Medicine. Not yet published.
- Schmorrow, D., Raley, C., Cohn, J. & Stripling, R. (2003, August). InfoBionics: Nano-Enhanced Information Performance. Proceedings of the Sixth International Conference on Integrated Nano/Micro/Biotechnology for Space and Medical Applications.
- Stripling, R., Raley, C., Schmorrow, D. & Cohn, J. (2003, July). Infobionics: Exploring a Potential Concept for the Human-Computer Interface. Proceedings of the 7th World Multi-Conference on Systemics, Cybernetics, and Informatics.
- Raley, C. (2002, March). BioResearch at DARPA and Related Federal Research. Strategic Analysis Technical Report # SA-004.

Selected Presentations

- Raley, C. (2005, July) Modular Design for Augmented Cognition Systems. Paper presented at HCI International, Las Vegas, NV.
- Schmorrow, D., Patrey, J., Cohn, J., Bahr, G., Arnold, R., Fatolitis, P., Crooks, C., Raley, C. & Kruse, A. (2004, November). Augmented Cognition. Poster session presented at the 110th Annual Meeting of The Association of Military Surgeons of the United States (AMSUS), Denver, CO.
- Raley, C., Schmorrow, D., & Kruse, A. (2004, October). Augmented Cognition: Bringing the Human in the Loop. Program review presented at the 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, Philadelphia, PA.
- Raley, C., Schmorrow, D., & Kruse, A. (2004, October). Augmented Cognition: A Neural Approach to Human Computer Interaction. Program review presented at the 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, Philadelphia, PA.
- Raley, C., Schmorrow, D., Kruse, A., & McBride, D. (2004, October). Quantifying Human Information-Processing Capabilities. Program review presented at the 2004 Biomedical Engineering Society (BMES) Annual Fall Meeting, Philadelphia, PA.

- Raley, C., Schmorrow, D. (2004, May). Biocybernetics to Superior Aviation: A Historical Perspective on Cognition Maximizing Research for the Aviator. Presented at the Aerospace Medical Association 75th AsMA Annual Scientific Meeting: Frontiers in Aerospace Medicine.
- Schmorrow, D., Raley, C. (2004, May). Augmented Cognition. Presented in the Potential Applications of Cognition Augmenting Technologies for Assessing Aviator Readiness Panel at the Aerospace Medical Association 75th AsMA Annual Scientific Meeting: Frontiers in Aerospace Medicine.
- Kruse, A., Schmorrow, D., Raley, C. (2004, May). Novel Portable Medical Devices for Performance Monitoring and Casualty Care. Poster presented at the Aerospace Medical Association 75th AsMA Annual Scientific Meeting: Frontiers in Aerospace Medicine.
- Schmorrow, D., Raley, C., et. al. (2004, March). Technology and Today's Warfighter: From Simulation and Training to Operational Environments. Panel Discussion chaired at Second Annual Human Performance, Situation Awareness and Automation Technology Conference, Daytona Beach, FL.
- Schmorrow, D., Raley, C., Marshall, L. (2004, March). Toward a Cognitive Cockpit. Poster presented at Second Annual Human Performance, Situation Awareness and Automation Technology Conference, Daytona Beach, FL.
- Stripling, R., Raley, C., Schmorrow, D. & Cohn, J. (2003, July). Infobionics: Exploring a Potential Concept for the Human-Computer Interface. Concept brief presented at the 7th World Multi-Conference on Systemics, Cybernetics, and Informatics, Orlando, FL.
- Raley, C. & Schmorrow, D. (2003 June). User Modeling, Augmented Cognition, and Cognitive Systems in IPTO. Program brief presented at the User Modeling 2003 Annual Conference, Johnstown, PA.
- Raley, C. & Schmorrow, D. (2003, June). Agents and the Interoperability of Heterogeneous Systems. Technical brief presented to Texas Society of Professional Engineers, South Padre Island, TX.
- Raley, C. & Schmorrow, D. (2003, January). DARPA, IPTO, and Infobionics. Program brief presented at MIT Attention and Memory in Wearable Interfaces Workshop, Boston, MA.
- Raley, C. & Schmorrow, D. (2002, November). Augmented Cognition Architecture. Program brief presented to the Human Engineering Test and Evaluation Sub-group at the Department of Defense Human Factors and Engineering Technical Advisory Group Meeting, Alexandria, VA.

Professional Affiliations

- Biomedical Engineering Society
- Cognitive Science Society
- National Defense Industrial Association